since 1961

Baltica

BALTICA Volume 36 Number 1 June 2023: 1–12
https://doi.org/10.5200/baltica.2023.1.1

Lithuanian river ice detection and automated classification using machine-learning methods

Linas Bevainis*, Martynas Bielinis, Agimantas Česnulevičius, Artūras Bautrėnas

Bevainis, L., Bielinis, M., Česnulevičius, A., Bautrėnas, A. 2023. Lithuanian river ice detection and automated classification using machine-learning methods. *Baltica*, *36 (1)*, 1–12. Vilnius. ISSN 1648-858X.

Manuscript submitted 24 May 2022 / Accepted 10 January 2023 / Available online 15 April 2023

Abstract. In regions susceptible to river freezing and flooding, river ice detection is a priority. Localization of ice jams and ice drift zones could mean a faster and better response to possible flooding areas, and classification of river ice could help better predict freezing and thawing conditions that hinder the use of commercial and recreational river transport. As many freezing-prone rivers are located in regions with short winter days and common cloud cover, the use of optical sensors can be very limited, therefore, the use of Synthetic Aperture Radar (SAR) – a microwave imaging radar – is more applicable. In this article, Sentinel-1 SAR C-band imagery is used to create derivate texture rasters, which are analyzed, compared with known optical imagery and then considered for river ice detection and discrimination. These results are compared in terms of their effectiveness for river ice discrimination, and the most useful methods are selected. The chosen methods are then compared in an experimental machine-learning model capable of detecting and classifying ice and water. Various machine-learning approaches (both classical and deep-learning) are considered and compared, and the best models are selected. The purpose of this research is to analyze the capability of texture rasters, calculated from a gray-level co-occurrence matrix (GLCM), to discriminate river ice. Texture rasters have recently been applied for river ice classification by de Roda Husman et al. (de Roda Husman *et al.* 2021), but included only three metrics. This research aims to expand on this knowledge by comparing eight metrics instead of three, as well as including an experiment with a deep-learning model. The results demonstrate that in machine-learning experiments, only one texture measure out of eight (GLCM Mean calculation) is able to discriminate river ice better than discrimination from a standard SAR backscatter intensity image (the baseline).

Keywords: *Sentinel-1; SAR; deep learning; classification methods*

Linas Bevainis (linas.bevainis@gf.vu.lt), https://orcid.org/0000-0001-9860-4440*
Martynas Bielinis (martynas.bielinis@icloud.com)
Agimantas Česnulevičius (algimantas.cesnulevicius@gf.vu.lt), https://orcid.org/0000-0003-3912-2403
Artūras Bautrėnas (arturas.bautrenas@gf.vu.lt)
Vilnius University, M.K. Čiurlionio 21/27, Vilnius, Lithuania
**Corresponding author*

INTRODUCTION

Countries in colder regions of the world regularly deal with a serious problem that restricts water activities in winter, i.e. river freeze-up. In Europe, there exist a vast number of wide rivers that could be commercially utilized for river transport, but seasonal river freeze-up hinders their exploitation. Modern SAR (Synthetic Aperture Radar) satellites can monitor the Earth in any weather or daylight conditions, and if they were used for river ice research, information on regularly fed river ice condition could allow us to accelerate response to a changing situation and monitor areas where ice is building up or thawing.

While the use of SAR seems straightforward, such methods bring new problems. Microwave imaging is sensitive to surface roughness and characteristics, and some frozen areas can look nearly identical to open water. Moreover, both ice and water features carry a lot of additional spectral information that needs to be identified and, if possible, used in classification.

The goal of this experiment is to evaluate the use of SAR for discrimination of water and ice in rivers. The tasks of this experiment are three-fold: 1) to analyze uncertainty in microwave backscatter that appears in certain ice zones, by finding a textural measure that would eliminate it; 2) to evaluate the discrimination of ice and open water by using texture analysis; and 3) to demonstrate automatic ice and water discrimination by using machine learning models.

LITERATURE REVIEW

Synthetic Aperture Radar

During winter season, remote sensing in north-eastern Europe is limited due to a constant cloud cover and short daylight hours. Traditional optical and infrared sensors cannot penetrate a thick cloud cover or work under low levels of daylight, which severely limits their potential for most monitoring tasks. Instead, a more useful solution is to use the Synthetic Aperture Radar (SAR). First described and tested in the 1950s (Sherwin *et al*. 1962), SAR works by moving a conventional microwave imaging radar along a platform, simulating a much larger aperture. Because of radar resolution being directly proportional to the ratio of wave length and antenna size (Fraceschetti, Lanari 1999), the radar can produce a higher-resolution image of the surface (Moreira *et al.* 2013). The active nature of SAR and the ability of microwave radiation to pass through a cloud cover allow constant observation of phenomena in areas otherwise impenetrable by visible-light and infrared sensors.

SAR imagery and ice discrimination

Contrary to traditional sensors, SAR imagery is not easily interpretable. Instead of collecting visible light or infrared radiation emitted or reflected from ground objects, SAR itself sends microwave radiation, what is called "active sensing", and registers the return signal (backscatter) scattered back from the ground (Chen 2016). Even after extensive product pre-processing, the end result is only a microwave echo of the surface. Therefore, interpreting a SAR product requires understanding how different surfaces and ground objects reflect microwave radiation.

For ice detection, the backscatter signal intensity depends on many factors, most important of which are surface roughness and the dielectric properties of ice (Carsey 1992). Sharp, protruding and variously orientated ("rough") areas of ice jams generate and reflect more and return a strong, bright signal, while areas covered with young, smooth ice can look nearly identical to open water, as signals pass through the ice and generate a weak response reflected from the ice-water interface, or plainly reflect less microwave radiation in the general direction of the radar. Open water itself has a very weak microwave response, primarily stemming from its low roughness, and therefore appears very dark in a SAR image. However, wave action on the water surface increases the roughness and can contribute to a stronger return signal (Gulácsi, Kovács 2020). Besides, thawing conditions mean more liquid water on the snow cover and ice surface, and since backscatter intensity depends on the dielectric constant, the response also becomes weaker (Hallikainen, Winebrenner 1992; Le Toan 2007).

Imagery analysis

Due to the similarity of backscatter intensity between some ice types and water, pure visual analysis of bands is not sufficient. However, as some researchers point out, no ice cover is perfectly smooth and a stronger signal response can be detected from cracks or at the ice-water interface. Such patterns could be found using texture analysis, for example, by creating a gray-level co-occurrence matrix from the input band and calculating Haralick features (Haralick *et al*. 1973). Some studies have already attempted to use this method and found favourable results (de Roda Husman *et al.* 2021). Moreover, other studies have tried using polarimetric decomposition to discriminate between surface and volume scatters, albeit they involved sea-ice and more versatile quad-polarized data (Scheuchl *et al.* 2021; Shokr, Dabboor 2020). Furthermore, summer (open-water) data has been found to be useful as a comparison tool for mapping sea-ice cover (Winsvold *et al.* 2018). However, the aforementioned study of river ice classification using Haralick features (de Roda Husman *et al.* 2021) only used three texture measures. This research attempts to expand on this knowledge by including more comparable Haralick features. Besides, this approach to river ice classification is novel to the region in question, as similar studies have not yet been done in Lithuania. The only study of Sentinel-1 dual-polarization data usage for river ice classification in the surrounding region was conducted by Łoś and Pawłowski in 2017 (Los, Pawlowski 2017).

Machine learning

The full potential of any ice discrimination study can only be unlocked by feeding the data into a ma-

chine learning model for automated classification. The amount of collected data and the size of areas of interest are large; therefore, modern automated classification solutions would be beneficial in detecting ice, discriminating between ice types or, at the very least, segmenting surface types for manual labelling. Traditional machine learning methods, such as Support Vector Machine (SVM) (Boser *et al.* 1992), Random Forest or k-means clustering (MacQueen 1967) have been used for similar classification tasks before (Chu, Lindenschmidt 2016; de Roda Husman *et al.* 2021; Sobiech, Dierking 2013; Weber *et al.* 2003) and their usefulness can be tested. Yet another interest is in determining the effectiveness of novel deep-learning methods that use context-aware operations such as convolution (Yamashita *et al.* 2018) to successfully analyze complex imagery. Modern deep learning architectures such as PSPNet (Zhao *et al.* 2016) and UNet (Ronneberger *et al.* 2015) have been demonstrated to successfully segment imagery, and some studies have already attempted to use them for remote sensing tasks (Singh *et al.* 2019). The use of deep-learning technology for classification using Haralick features is a novel approach to this problem.

METHODOLOGY

Study area

To assess the usefulness of river ice detection in Lithuania, we have chosen the Nemunas River as a study area. During the selection of imagery, the entire length of the river in Lithuania was taken into account, excluding the Kaunas Reservoir, where Nemunas is dammed by the Kaunas Hydroelectric Power Plant and lake ice phenomena can be observed (see Fig. 1).

After visual analysis of both SAR and optical imagery, more specific regions of interest were later selected and extracted from the original extent.

Imagery

The Sentinel-1 imagery library as available through Alaska Satellite Facility's Vortex database (National Aeronautics and Space Administration 2022) was analyzed. In total, from 2014 to 2021, 96 rasters were found where the Nemunas was frozen in at least one area. The winter season of 2019–2020 was excluded

Fig. 1 Map of the study area, with parts of the Nemunas River used in the study marked in red

Table 1 Selected SAR imagery with overlapping optical data

#	Sentinel Mission	Orbit Number	Product ID	SAR Sensing Time (Local)	Optical Imagery Sensor	Optical Sensing Time (Local)
1.	S1A	3878	7BD0	25.12.2014 18:11	Landsat 8	25.12.2014 09:25
2.	S1A	9573	A9F5	20.01.2016 06:51	Landsat 8	20.01.2016 09:31
3.	S1A	15275	609D	14.02.2017 06:43	Sentinel-2 L1C	14.02.2017 11:41
4.	S1A	26001	3D33	19.02.2019 18:20	Sentinel-2 L1C	19.02.2019 11:40
5.	S1B	25729	428A	23.02.2021 06:43	Sentinel-2 L1C	23.02.2021 11:40

due to abnormally high temperatures and no river ice cover (Lithuania Hydrometeorological Service 2020). All SAR images were captured in the Interferometric Wide Swath mode, with VV/VH polarizations, and downloaded in the GRD-HD mode. Both ascending and descending orbits were used, as the features in question have a small height difference (very rough ice jam conditions would contribute to the largest height difference).

In order to define ice cover conditions, optical imagery was used by downloading it from Creodias (CloudFerro 2022). For any sensing day of any SAR image of interest, supplementary Sentinel-2 or Landsat 8 imagery with no cloud cover over the river is needed, and only 5 SAR images were found to have overlapping optical imagery with good atmospheric conditions (see Table 1).

The optical sensing time for rasters where river freeze-up conditions were abundant (#3 and #5) was roughly 5 hours after SAR sensing time, which provided sufficiently similar results and easily distinguishable ice forms. The #1 and #4 rasters had ~7 hours of sensor time difference, which provided less precise comparisons but nevertheless was sufficient to assess the general state of river freeze-up. In addition, Landsat 8 optical imagery has a much lower spatial resolution than Sentinel-2 and was therefore also used only for the general assessment of river conditions.

Each SAR product was pre-processed in ESA SNAP software in this sequence:

- Orbit File Application
- Thermal Noise Removal
- Calibration with sigma band output
- Terrain Correction
- Scale Conversion from Linear Units to dB (decibels)

Speckle Filtering was omitted due to resulting loss of image texture information and spatial detail (Filipponi 2019).

For each image pair, the entire length of the river was visually assessed and regions of interest were selected. The regions were selected based on the existence of clearly visible ice conditions (complete freeze-up, ice jams and gaps of open water), as well as in frozen areas where SAR imagery looked visually similar to water (i.e., problematic regions of young/thermal ice) (see Figs 2–9).

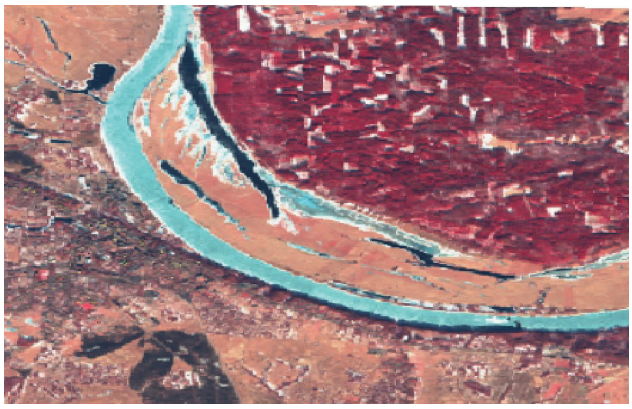

Fig. 2 Clearly identifiable ice cover in Sentinel-2

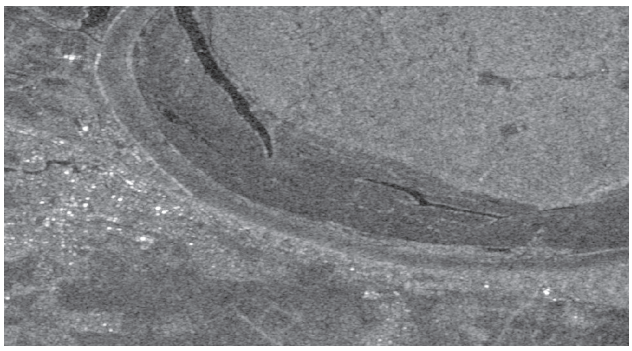

Fig. 3 Clearly identifiable ice cover in corresponding VV SAR imagery

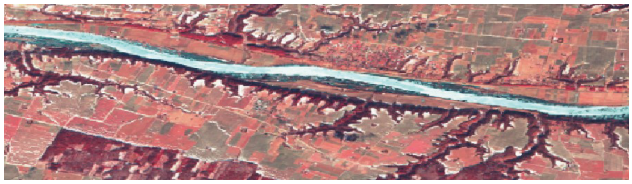

Fig. 4 Rough ice cover (ice jam locations) visible in Sentinel-2

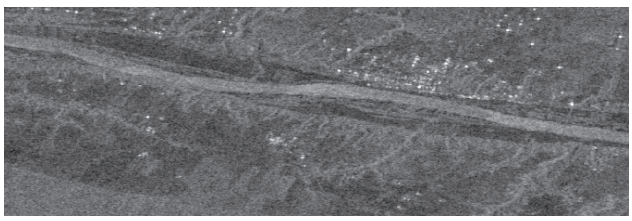

Fig. 5 Rough ice cover (ice jam locations) is visible in corresponding VV SAR imagery as brighter zones

Haralick feature calculation

Polarization intensity images are useful for the classification of such phenomena as ice jams or rugged ice areas, as these "bright" regions are easily separable from a low-intensity "dark" response of open water. However, young flat-surfaced thermal ice, as well as high amounts of meltwater over the ice surface create a backscatter response that is visually similar to that of open water (Schwaizer 2017). Therefore, backscatter intensity alone is insufficient. A method proposed by Gauthier et al. (Gauthier *et al.* 2006) and successfully applied for river ice classification by de Roda Husman et al. (de Roda Husman *et al.* 2021) is the use of image texture information.

The method used in this study is the calculation of gray level co-occurrence matrix (GLCM) – a matrix of image pixel value pair counts (The Mathworks Inc.

Fig. 6 Water with possible frazil ice cover in Sentinel-2

Fig. 7 Water with possible frazil ice cover is visible in VV SAR imagery

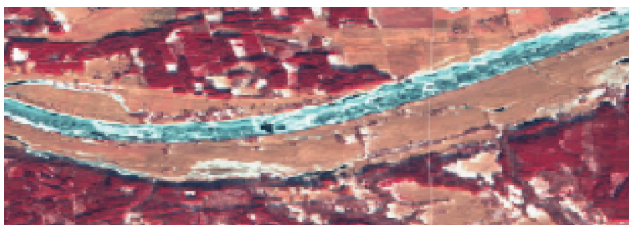

Fig. 8 Clearly visible ice cover in Sentinel-2

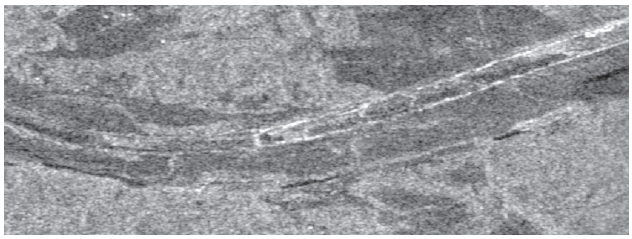

Fig. 9 Ambiguously visible ice cover in VV SAR imagery, where certain ice zones appear much darker. Ice cracks are visible and could be a visual indicator of ice cover

2022) and the subsequent use of the matrix for the calculation of statistical metrics, also known as Haralick features (Haralick *et al.* 1973).

For this study, the Orfeo Toolbox plugin for QGIS was used to calculate the features (Centre national d'études spatiales 2021). The following parameters were used:

- Computation step: 2
- X/Y Radius: 5
- X/Y Offset: 3
- Histogram bins: 12

To compare separate images, common minimum and maximum image values must be defined, taking into account all individual rasters that will be used in subsequent calculations. For VH polarization imagery, the maximum value is 21 dB (decibels) and the minimum is -68 dB; for VV polarization imagery, the maximum value is 35 dB and the minimum is -62 dB. These values were chosen as the common minimum and maximum.

The QGIS plugin allows for the calculation of 29 features in total, but only 8 were chosen for this study as not all are comparable.

The QGIS OTB plugin uses a variant of the GLCM, called the GLCIL (Grey Level Co-occurrence Indexed List), which applies an indexing method to the GLCM to improve the speed of the calculations. The resulting features are still Haralick features, but the approach is more fitting to a computationally demanding task such as the one in this study (Clausi, Jernigan 1998).

The QGIS OTB plugin uses the following formulas to calculate the selected Haralick features from the GLCM:

μ – weighted pixel mean
σ – weighted pixel variance
$g(i, j)$ – frequency of the pixel pair
i, j – pixel pair indices

$$GLCM\ energy = \sum_{i,j} g(i, j)^2 \quad (1)$$

- Equation 1. Energy: describes texture uniformity. Higher values correspond to a more uniform texture, while lower values correspond to a more varied texture (OTB Team 2018).

$$GLCM\ entropy = -\sum_{(i,j)} g(i,j)\log_2 g(i,j) \quad or\ 0,\ if\ g(i,j) = 0 \qquad (2)$$

- Equation 2. Entropy: describes texture randomness. Higher values correspond to a more chaotic texture, while lower values correspond to a more orderly texture (OTB Team 2018).

$$GLCM\ correlation = \sum_{i,j} \frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2} \qquad (3)$$

- Equation 3. Correlation: describes the correlation of a pixel with its neighbourhood. Higher values correspond to a higher correlation, while lower values correspond to a lower correlation (OTB Team 2018).

$$GLCM\ IDM = \sum_{i,j} \frac{1}{1+(i-j)^2} g(i,j) \qquad (4)$$

- Equation 4. Inverse difference moment (IDM): describes texture homogeneity. Higher values correspond to a lower pixel-to-pixel contrast in the neighbourhood, while lower values correspond to a higher pixel-to-pixel contrast in the neighbourhood (OTB Team 2018).

$$GLCM\ mean = \sum_{(i,j)} ig(i,j) \qquad (5)$$

- Equation 5. Mean: describes the mean pixel-pair occurrence in the neighbourhood (OTB Team 2018).

$$GLCM\ variance = \sum_{(i,j)} (i-\mu)_2 g(i,j) \qquad (6)$$

- Equation 6. Variance: describes the pixel-pair occurrence dispersion in the neighbourhood (OTB Team 2018).

$$GLCM\ Variance\ of\ Diff. = var\left(g_{x-y}(i)\right) \qquad (7)$$

- Equation 7. Variance of Differences: x, y – *gray levels*, describes the variance of gray level differences in the matrix (OTB Team 2018).

$$GLCM\ Entropy\ of\ Diff. = -\sum_i g_{x-y}(i)\log\left(g_{x-y}(i)\right) \qquad (8)$$

- Equation 8. Entropy of Differences: x, y – *gray levels*, describes the entropy of gray level differences in the matrix (OTB Team 2018).

Several regions of interest were selected from the optical-SAR image pairs that included open water zones, rough ice cover, smooth ice cover and problematic areas (areas with hard-to-distinguish classes in intensity images). For each of these regions, texture calculation was done using the aforementioned *Calculate Haralick Textures* tool with the described parameters. The selected common minimum and maximum values make sure that the resulting texture metrics are comparable. An example of the visual appearance of the texture metrics is presented below (see Figs 10–12).

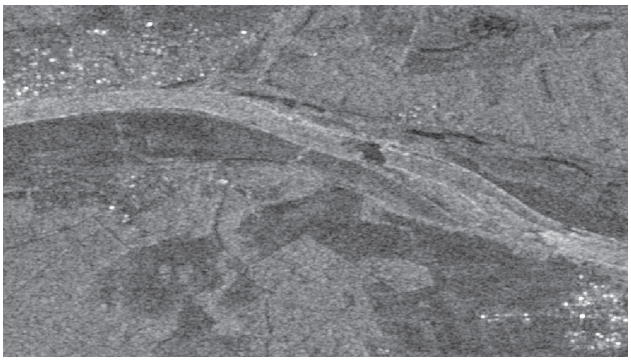

Fig. 10 Backscatter intensity raster in VV

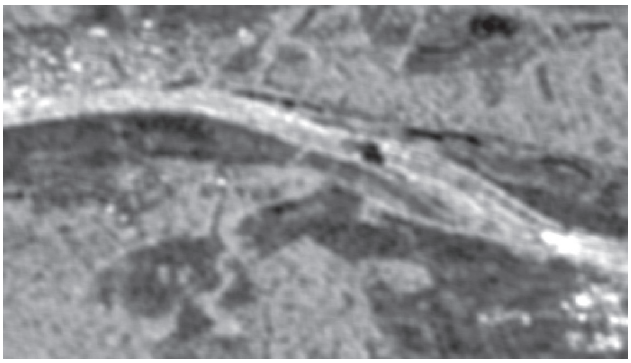

Fig. 11 GLCM Mean texture raster derived from VV

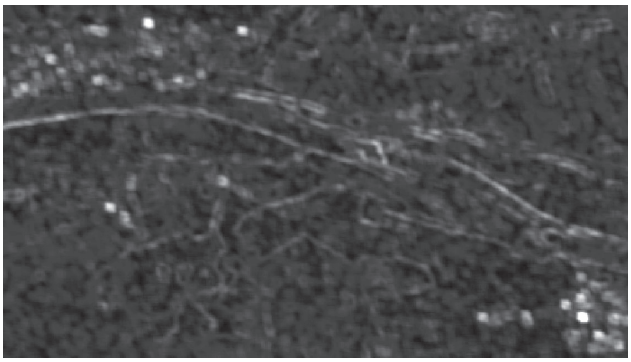

Fig. 12 GLCM Variance texture raster derived from VV

Feature analysis

To determine the usefulness and applicability of various texture metrics, several zones were selected from the regions of interest, with each zone covering the exact boundaries of one feature of interest, based on its extent in the texture metric image (see Table 2). For each feature zone of each texture metric, the mean and the values at one standard deviation from the mean were calculated.

Table 2 Analysis zones

Analysis Zones	Zone Description
W_1, W_2, W_3	Open water area
C_1, C_2	Ice-covered area with visible cracks
I_1, I_2, I_3, I_4, I_5, I_6	Ice-covered area
J_1, J_2	Ice-covered area with a very rough ice cover (ice jam)

The calculated zones are then compared with each other. The goal is to determine which of the metrics have the lowest amount of value overlap within 1 standard deviation range between water and ice features. A tally calculation is used to count:

- how many ice zones (out of 10 possible) overlap in their standard deviation value ranges with at least one water zone. Such overlaps show that water and ice are not perfectly separable and problems might occur in determining the class of an area from the values;
- how many ice zones (out of 10 possible) have their value medians within the standard deviation value range of the water zone. This shows probable critical overlap of values and could cause significant separability problems.

These two tally counts allow for a quick, non-visual evaluation of the metrics. At the end of this part of the analysis, the most separable metrics are determined, which will be later used in machine learning.

Machine learning

After the evaluation and selection of viable texture metrics, the metrics are used as single-band rasters in machine learning. In order to gather more samples for machine learning models, a new merged raster was created from extracted subsets of other rasters, so that various classes and ice conditions could be present in the training image.

Two classes were defined for machine learning: water (only open-water areas, including problematic areas that appeared visually ambiguous but were manually confirmed to be water) and ice (only ice-covered areas, including areas of rough ice, smooth ice and other ice types). For each class, around 30–50 samples were taken from various rasters. Sample classes were equal by total sample area.

Three machine learning models were chosen for this task – Support Vector Machine (SVM), Random Trees, and *k* Nearest Neighbour. These models were chosen due to their prevalence in modern remote sensing/GIS software and overall popularity as segmentation methods. However, it would be imprudent to claim that these models are the best for this particular task, and a more in-depth study comparing a larger selection of machine learning models in this task would be beneficial. Each model was trained on the class samples, and the model was then applied on a raster that included areas of rough and smooth ice, as well as ice jams and open water areas. The inferred result is a two-class raster. The test raster is also manually classified into two classes based on visual analysis and, more importantly, comparison with overlapping optical data. The parameters of these models are provided below (see Tables 3–5).

Table 3 Parameters used for the k-NN model

Parameters of the k-Nearest Neighbour model (non-segmented raster as input)	
Number of Classes	2
Max Number of Samples per Class	1000 (default)
K-Nearest Neighbours	1 (default)

Table 4 Parameters used for the RT model

Parameters of the Random Trees model (non-segmented raster as input)	
Maximum Number of Trees	50
Maximum Depth of Tree	30
Max Number of Samples per Class	1000

Table 5 Parameters used for the SVM model

Parameters of the Support Vector Machine model	
Maximum Number of Samples per Class	500
SVM Type	C-Support Vector Classification (c-svc)
Kernel Type	Radial basis function (rbf)
Gamma	0.5

The inference result and the manually-classified ground truth are then compared to assess model accuracy. The evaluation is performed by randomly sampling accuracy assessment points from the ground truth and matching them with the inference result. A confusion matrix is then calculated from the assessment points, and three measures of accuracy are calculated – *User's Accuracy, Producer's Accuracy,* and *Kappa index* (ESRI 2022). *User's Accuracy* measures the accuracy based on false positives, while *Producer's Accuracy* measures the accuracy based on false negatives. *Kappa index* computes the overall accuracy of the model based on model accuracy and calculated random accuracy, derived from class classification probabilities (ENVI 2019). *Kappa* ranges from -1 to 1. According to the interpretation proposed by Mary L. McHugh, a Kappa value of 1.0 would be considered a perfectly accurate score, while negative values indicate that the accuracy is performing below the random accuracy achieved by random guessing. Kappa values below 0.60 are to be considered unreliable (low accuracy), values in the 0.60–0.79 range are of moderate accuracy, and values in the 0.80–0.90 range are of strong accuracy (McHugh 2012).

Besides traditional machine learning methods, a more advanced deep learning model was tested. For this experiment, the Pyramid Scene Parsing Network (PSPNet) was developed by Zhao, Shi et al. 2016 (Zhao *et al.* 2016). This model uses a convolutional neural network to extract a feature layer and applies the pooling of different sizes, each being of higher

resolution (higher bin size) and therefore capable of capturing more fine detail. After additional processing, all these separate pooling features are concatenated with the original convolution feature layer to form a single feature that is once again convoluted. The novelty of this method is the inclusion of pooling at varied sizes (hence the "pyramid"), which allows the model to capture context information at various levels (Zhao *et al.* 2016). In some cases, PSPNet has demonstrated a better capability to segment the image with fine spatial detail and small features (Zhao *et al.* 2016) than UNet or other methods and was therefore chosen for this study; however, future analysis of river ice segmentation with a comparison of various deep learning models would be beneficial in determining the most fitting model. For the parameters used with the PSPNet Classifier, see below (Table 6).

To train the model, a fully classified river raster was used. The model was then fitted for 100 epochs, and the resulting model metrics were analyzed.

For the deep learning model evaluation, *precision*, *recall* and F_1 were used for assessment (Equations 9–11).

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \quad (9)$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad (10)$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \quad (11)$$

RESULTS

Feature analysis results

The charts below show the median value and the values at one standard deviation (σ) from the mean for each analysis zone of each calculated Haralick feature, as well as both polarization intensity images in dB units. Because backscatter intensity is the base for other texture calculations, the separability of classes (or lack thereof) present in the intensity image is notably echoed in its derivative textural calculations.

The analysis (see Table 7, Figs S1–S19) confirms that VH polarization is much worse in separating water from ice than VV polarization – all but three of VH metrics were fully overlapping. The best metric in VH polarization was GLCM Mean, but even it had 7 overlapping zones out of 10, as well as 7 ice zone medians within water σ range.

VV polarization, as expected, returned better results. However, only one metric (GLCM Mean) was better than the baseline backscatter intensity. However, it returned favourable results – only 1 σ range overlap and no medians within water σ range.

Table 6 Parameters used for the PSPNet Classifier

Parameters of the PSPNet Classifier	
Backbone CNN Model	resnet50
Pyramid Sizes	1, 2, 3, 6
Epochs	100
Learning Rate	Automatically deduced from the Learning Rate Finder (Fastai 2022)

Table 7 Texture metric separation analysis

Metric	σ range overlap (overlapping ice zones/10)	Ice median in water σ range (problematic ice zones/10)
VH Correlation	10	10
VH Entropy of Differences	10	10
VH Variance of Differences	10	10
VH Energy	10	10
VH Entropy	10	10
VH IDM	10	10
VH Backscatter Intensity	8	7
VH Mean	7	7
VH Variance	10	6
VV Correlation	10	10
VV Entropy of Differences	8	5
VV Variance of Differences	8	5
VV Energy	7	5
VV Entropy	8	5
VV IDM	7	4
VV Backscatter Intensity	3	1
VV Mean	1	0
VV Variance	8	5

In order to more fully assess the textural metrics, the following metrics were chosen for tests using machine learning methods: VV Energy, VV IDM, VV Backscatter Intensity, VV Mean.

Machine learning results

Backscatter intensity accuracies were of rather weak accuracy, with the highest Kappa values achieved using SVM (Kappa value of 0.5876). The results from this test will be used as a baseline to determine whether the texture measures perform better or worse than pure backscatter image classification (see Table 8).

Machine learning tests demonstrate that both IDM and Energy metrics cannot classify the image correctly and their Kappa index is close to random, while that of SVM test of GLCM IDM is at -0.0988 – a very low result as well (see Tables 9–10).

The baseline test of backscatter intensity demonstrates that only GLCM Mean could surpass the

Fig. 13 GLCM Mean + Intensity image

Fig. 14 GLCM Mean SVM result. Blue – water; gray – ice

baseline Kappa score. The highest Kappa value was achieved with the k-Nearest Neighbours method – 0.7167, which is of moderate accuracy. The results also show that the model had a higher user accuracy for water and a higher producer accuracy for ice. This shows that models were less likely to attribute a pixel to water if it was ice, but more likely to falsely classify water as ice. This means that the models were less confident about classifying pixels as water (see Fig. 14, Table 11).

It is not clear from the results which of the three machine learning solutions is best – it varies from metric to metric. However, the GLCM Mean – the most accurate metric – was most accurately inferred using k-Nearest Neighbours (k-NN).

To additionally understand the effect that metrics have on classification, a composite raster was created with all 18 metrics, including both backscatter intensities. The results were better than the baseline, but worse than using GLCM Mean itself. The higher-than-baseline Kappa values might be because of the effect that more accurate metrics, such as VV Mean and VV backscatter, have on the overall classification (see Table 12).

An additional raster was created using only two most-promising rasters – GLCM Mean and VV backscatter (see Fig. 13). The returned results show that only Random Trees managed to return a better Kappa score than the corresponding Kappa scores from other metrics tested separately. Even so, the returned Kappa score was still of moderate accuracy (see Table 13).

Overall, it can be concluded that VV GLCM Mean returns a better classification result than standard VV backscatter, but significant classification difficulties persist and prevent more accurate inference because even in the best-case configurations, Kappa score is in the 0.70–0.74 range.

Table 8 VV Backscatter Intensity results

VV Backscatter Intensity k – Nearest Neighbours		
	User's Accuracy, %	Producer's Accuracy, %
Water	85.45	66.43
Ice	65.83	85.12
Kappa		0.4979
VV Backscatter Intensity Random Trees		
	User's Accuracy, %	Producer's Accuracy, %
Water	82.03	74.20
Ice	69.83	78.60
Kappa		0.5202
VV Backscatter Intensity Support Vector Machine (SVM)		
	User's Accuracy, %	Producer's Accuracy, %
Water	92.79	68.20
Ice	68.97	93.02
Kappa		0.5876

Table 9 VV GLCM Energy results

VV Energy k – Nearest Neighbours		
	User's Accuracy, %	Producer's Accuracy, %
Water	59.31	60.78
Ice	46.63	45.12
Kappa		0.0592
VV Energy Random Trees		
	User's Accuracy, %	Producer's Accuracy, %
Water	57.44	68.20
Ice	44.44	33.49
Kappa		0.0174
VV Energy Support Vector Machine (SVM)		
	User's Accuracy, %	Producer's Accuracy, %
Water	58.15	73.14
Ice	46.48	30.80
Kappa		0.0401

Table 10 VV GLCM IDM results

VVIDM k – Nearest Neighbours		
	User's Accuracy, %	Producer's Accuracy, %
Water	59.57	68.20
Ice	48.28	39.07
Kappa		0.0744
VVIDM Random Trees		
	User's Accuracy, %	Producer's Accuracy, %
Water	61.85	54.42
Ice	48.19	55.81
Kappa		0.1004
VVIDM Support Vector Machine (SVM)		
	User's Accuracy, %	Producer's Accuracy, %
Water	51.50	42.40
Ice	38.49	47.44
Kappa		-0.0988

Table 11 VV GLCM Mean results

VV Mean k – Nearest Neighbours		
	User's Accuracy, %	Producer's Accuracy, %
Water	100	74.56
Ice	74.91	100
Kappa		0.7167
VV Mean Random Trees		
	User's Accuracy, %	Producer's Accuracy, %
Water	100	73.85
Ice	74.39	100
Kappa		0.7092
VV Mean Support Vector Machine (SVM)		
	User's Accuracy, %	Producer's Accuracy, %
Water	100	71.02
Ice	72.39	100
Kappa		0.6791

Table 12 Composite results

Composite of all metrics k – Nearest Neighbours		
	User's Accuracy, %	Producer's Accuracy, %
Water	84.96	73.85
Ice	70.63	82.79
Kappa		0.5549
Composite of all metrics Random Trees		
	User's Accuracy, %	Producer's Accuracy, %
Water	100	74.56
Ice	74.91	100
Kappa		0.7167
Composite of all metrics Support Vector Machine (SVM)		
	User's Accuracy, %	Producer's Accuracy, %
Water	100	69.61
Ice	71.43	100
Kappa		0.6642

Table 13 VV GLCM Mean + VV Intensity results

VV Mean + VV Backscatter Intensity k – Nearest Neighbours		
	User's Accuracy, %	Producer's Accuracy, %
Water	89.23	81.98
Ice	78.57	86.98
Kappa		0.6808
VV Mean + VV Backscatter Intensity Random Trees		
	User's Accuracy, %	Producer's Accuracy, %
Water	90.74	86.57
Ice	83.33	88.37
Kappa		0.7440
VV Mean + VV Backscatter Intensity Support Vector Machine (SVM)		
	User's Accuracy, %	Producer's Accuracy, %
Water	100	71.02
Ice	72.39	100
Kappa		0.6791

Table 14 Deep Learning results

	Pyramid Scene Parsing Network, 100 epochs, VV Intensity + VV GLCM Mean	
	Water	Ice
Precision	0.9340	0.7014
Recall	0.4076	0.8256
F_1	0.5675	0.7584

The PSPNet deep learning model confirms machine learning results – a 100-epoch model with training data made up of 668 features that were, after data augmentation, transformed into 9.506 individual features did perform similarly to previous machine learning models and shows similar patterns. Water classification has a very high precision (few ice pixels classified as water) but a low recall (many missed water pixels). The overall F_1 score is mediocre – 0.5675. Meanwhile, ice classification had a more balanced result of precision and recall, and an overall F_1 score of 0.7584 (see Table 14).

It is imperative to mention that deep learning models require a very large amount of data and a long period of training in order to be useful. The data used here had only 668 real features that were rotated to artificially increase (augment) the feature count to 9.506. Proper deep learning models could require tens or hundreds of thousands individual features before they can be considered for actual real-world deployment, and not just for experimentation and testing. This deep learning test is only meant to assess the extent to which this simple dataset can be fitted, and the resulting scores demonstrate that the model could classify ice with relative success, while water was met with certain difficulties.

The present higher error rate of water classification could also be a result of a low sample rate of water features. It is possible that water features have a much broader range of possible values, and additional

samples (more unique and manually confirmed water rasters) might be needed for the model to increase its classification accuracy.

CONCLUSIONS

Haralick feature analysis demonstrated and confirmed textural analysis results of de Roda Husman (de Roda Husman *et al.* 2021), i.e. VV polarization GLCM Mean feature had better class separability than VV polarization backscatter intensity. No other analyzed feature had better separability than the baseline VV polarization backscatter intensity from which the Haralick features were derived.

Machine learning tests also demonstrated that the GLCM Mean single-band raster achieves more favourable classification results (Kappa score 0.6791–0.7167) than the baseline VV polarization backscatter intensity (Kappa score 0.4979–0.5876). A composite raster of all analyzed features did return a slightly better result (Kappa score 0.5549–0.6642) than the baseline but fell short of the accuracy of GLCM Mean. It is important to note that even the best results are still in the 0.70–0.80 Kappa score range, which signifies results of moderate accuracy (McHugh 2012). Deep learning tests returned similar results to classical machine learning and demonstrated a prevailing problem – lower accuracy of water classification.

The results demonstrate that the ambiguity of water and ice backscatter intensity is not entirely eliminated but is decreased with the use of GLCM Mean metric derived from VV polarization. Additional tests with more unique and manually classified rasters are needed, and possibly a new (or ensemble) method is required to produce a composite raster that could lower the remaining ambiguity even further. An experimental test demonstrates that the use of deep learning to increase classification accuracy has a great potential but needs significantly more features within a broader range of possible values before the model can be properly deployed.

ACKNOWLEDGMENTS

The authors of the article would like to sincerely thank two anonymous reviewers for their valuable comments and recommendations.

REFERENCES

Boser, B.E., Guyon, I.M., Vapnik, V.N. 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, 144–52.

Carsey, F. D. 1992. *Microwave Remote Sensing of Sea Ice*. American Geophysical Union.

Chen, K. 2016. Chapter 2: SAR Models. In: *Principles of Synthetic Aperture Radar Imaging: A System Simulation Approach*. Boca Raton: CRC Press.

Chu, T., Lindenschmidt, K. 2016. Integration of Space-Borne and Air-Borne Data in Monitoring River Ice Processes in the Slave River, Canada. *Remote Sensing of Environment 181*, 65–81. https://doi.org/10.1016/j.rse.2016.03.041

Clausi, D.A., Jernigan, M. (eds) 1998. A Fast Method to Determine Co-Occurrence Texture Features. *IEEE transactions on geoscience and remote sensing 36 (1)*, 298–300.

CloudFerro. 2022. Creodias Data Finder. Retrieved (https://finder.creodias.eu/)

De Roda Husman, S., van der Sanden, J.J., Lhermitte, S., Eleveld, M.A. 2021. Integrating Intensity and Context for Improved Supervised River Ice Classification from Dual-Pol Sentinel-1 SAR Data. *International Journal of Applied Earth Observation and Geoinformation 101*, 102359. https://doi.org/10.1016/j.jag.2021.102359

Filipponi, F. 2019. Sentinel-1 GRD Preprocessing Workflow. In: *The 3rd International Electronic Conference on Remote Sensing*, 5.

Fraceschetti, G., Lanari, R. 1999. Chapter 1: Fundamentals. In: *Synthetic Aperture Radar Processing*, 16–18. Boca Raton: CRC Press.

Gauthier, Y., Weber, F., Savary, S., Jasek, M., Paquet, L., Bernier, M. 2006. A Combined Classification Scheme to Characterise River Ice from SAR Data. *A Combined Classification Scheme to Characterise River Ice from SAR Data 5 (1)*, 77–88.

Gulácsi, A., Kovács, K. 2020. Sentinel-1-Imagery-Based High-Resolution Water Cover Detection on Wetlands, Aided by Google Earth Engine. *Remote Sensing 12 (10)*, 1614. https://doi.org/10.3390/rs12101614

Hallikainen, M., Winebrenner, D. 1992. The Physical Basis for Sea Ice Remote Sensing. In: Carsey, F.D. (ed.) *Microwave Remote Sensing of Sea Ice*, 29–46. American Geophysical Union.

Haralick, R., Shanmugam, M.K., Dinstein, I. 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6)*, 610–21. https://doi.org/10.1109/TSMC.1973.4309314

Le Toan, T. 2007. *Introduction to SAR Remote Sensing*.

Los, H., Pawlowski, B. 2017. The Use of Sentinel-1 Imagery in the Analysis of River Ice Phenomena on the Lower Vistula in the 2015–2016 Winter Season. In: *2017 Signal Processing Symposium (SPSympo)*, 1–5. IEEE.

MacQueen, J. 1967. Some Methods for Classification and Analysis of Multivariate Observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1, 281–97. Oakland, CA, USA.

McHugh, M.L. 2012. Interrater Reliability: The Kappa Statistic. *Biochemia Medica 22 (3)*, 276–82.

Moreira, A., Prats-Iraola, I., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P. 2013. A Tutorial on Synthetic Aperture Radar. *IEEE Geosci-*

ence and Remote Sensing Magazine 1 (1), 6–43. https://doi.org/10.1109/MGRS.2013.2248301
OTB Team. 2018. Textural Features Extraction. In: *OTB Cook Book Documentation,* 111–12.
Ronneberger, O., Fischer, P., Brox, T. 2015. *U-Net: Convolutional Networks for Biomedical Image Segmentation*.
Scheuchl, B., Caves, R., Cumming, I., Staples, G. 2021. Automated Sea Ice Classification Using Spaceborne Polarimetric SAR Data. In: *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)*. Vol. 7, 3117–19. IEEE.
Schwaizer, G. 2017. *SAR/Optical Applications to Ice and Snow*. 172.
Sherwin, C.W., Ruina, J.P., Rawcliffe, R.D. 1962. Some Early Developments in Synthetic Aperture Radar Systems. *IRE Transactions on Military Electronics MIL 6(2)*, 111–115. https://doi.10.1109/IRET-MIL.1962.5008415
Shokr, M., Dabboor, M. 2020. Observations of SAR Polarimetric Parameters of Lake and Fast Sea Ice during the Early Growth Phase. *Remote Sensing of Environment 247*, 111910. https://doi.org/10.1016/j.rse.2020.111910
Singh, A., Kalke, H., Loewen, M., Ray, N. 2019. *River Ice Segmentation with Deep Learning*. https://doi.org/10.1109/TGRS.2020.2981082
Sobiech, J., Dierking, W. 2013. Observing Lake- and River-Ice Decay with SAR: Advantages and Limitations of the Unsupervised k -Means Classification Approach. *Annals of Glaciology 54 (62)*, 65–72. https://doi.org/10.3189/2013AoG62A037
Weber, F., Nixon, D., Hurley, J. 2003. Semi-Automated Classification of River Ice Types on the Peace River Using RADARSAT-1 Synthetic Aperture Radar (SAR) Imagery. *Canadian Journal of Civil Engineering 30 (1)*, 11–27. https://doi.org/10.1139/l02-073
Winsvold, S.H., Kääb, A., Nuth, Ch., Andreassen, L.M., van Pelt, W.J.J., Schellenberger, T. 2018. Using SAR Satellite Data Time Series for Regional Glacier Mapping. *The Cryosphere 12 (3)*, 867–90. https://doi.org/10.5194/tc-12-867-2018
Yamashita, R., Nishio, M., Gian Do, R.K., Togashi, K. 2018. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights into Imaging 9 (4)*, 611–29. https://doi.org/10.1007/s13244-018-0639-9
Zhao, H., Shi, J., Qi, X., Wang, X., Jiaya, Jia. 2016. Pyramid Scene Parsing Network. *ArXiv ID*, 1612.01105.

Internet sources:

Centre national d'études spatiales. [National Centre for Space Studies]. 2021. Orfeo Toolbox. Retrieved (https://www.orfeo-toolbox.org/)
ENVI. 2019. Calculate Confusion Matrices. *L3Harris Geospatial*. Retrieved December 20, 2022 (https://www.l3harrisgeospatial.com/docs/CalculatingConfusionMatrices.html%0Ahttps://www.harrisgeospatial.com/docs/CalculatingConfusionMatrices.html)
ESRI. 2022. Compute Confusion Matrix. Retrieved December 20, 2022 (https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/compute-confusion-matrix.htm)
Fastai. 2022. Callbacks.Lr_finder. Retrieved December 20, 2022 (https://fastai1.fast.ai/callbacks.lr_finder.html)
Liew, S.Ch. 2001. Microwave Frequency. *SAR Imaging – Frequency, Polarisation and Incident Angle, University of Singapore*. Retrieved February 20, 2022 (https://crisp.nus.edu.sg/~research/tutorial/freqpol.htm)
National Aeronautics and Space Administration. 2022. Alaska Satellite Facility. Retrieved (https://search.asf.alaska.edu/#/)
Sik-Ho, T. 2018. Review: PSPNet – Winner in ILSVRC 2016 (Semantic Segmentation / Scene Parsing). *Towards Data Science*. Retrieved December 20, 2022 (https://towardsdatascience.com/review-pspnet-winner-in-ilsvrc-2016-semantic-segmentation-scene-parsing-e089e5df177d)
The Mathworks Inc. 2022. Texture Analysis Using the Gray-Level Co-Occurrence Matrix (GLCM). Retrieved March 12, 2022 (https://uk.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html)

Supporting Online Material
Figures S1-S19