NRC Publications Archive
Archives des publications du CNRC

Identification of the best method for detecting surface water in Sentinel-2 multispectral satellite imagery

Kirby, Katelyn; Ferguson, Sean; Rennie, Colin D.; Cousineau, Julien; Nistor, Ioan

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.rsase.2024.101367
Remote Sensing Applications: Society and Environment, 36, C, 2024-10-05

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=4ecf6443-4de1-417e-8df3-0bfdffc6f08c
https://publications-cnrc.canada.ca/fra/voir/objet/?id=4ecf6443-4de1-417e-8df3-0bfdffc6f08c

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada
Conseil national de recherches Canada

Remote Sensing Applications: Society and Environment 36 (2024) 101367

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Identification of the best method for detecting surface water in Sentinel-2 multispectral satellite imagery

Katelyn Kirby[a,*], Sean Ferguson[b], Colin D. Rennie[a], Julien Cousineau[b], Ioan Nistor[a]

[a] *Department of Civil Engineering, University of Ottawa, 161 Louis Pasteur, Ottawa, ON, Canada, K1N 6N5*
[b] *Ocean, Coastal and River Engineering Research Centre, National Research Council Canada, 1200 Montreal Rd, Ottawa, ON, K1A 0R6, Canada*

ARTICLE INFO

Keywords:
Surface water mapping
Remote sensing classification
AWEI
NDWI
Sentinel-2 multispectral satellite imagery
Earth observation

ABSTRACT

Surface water maps are useful in a variety of disciplines from climate change analysis to water resource management. Multispectral satellite imagery can be used to derive such surface water maps using a variety of image processing methods. The medium resolution Sentinel-2 multispectral satellite imagery catalogue is currently used extensively for surface water mapping. The quality and accuracy of these maps produced from Sentinel-2 imagery can vary greatly depending on the method applied to classify the image pixels into land or water. Thus far, there has not been a consensus on which method produces the highest accuracy surface water maps, warranting a direct comparison to assess these methods in a wide range of geographic settings. Here we show that among some of the most commonly applied surface water mapping methods (NDWI, MNDWI, AWEI_SH, AWEI_NSH, AWEI_BOTH, SVM, RT, MLC, and KNN) that no single method produced the most accurate maps across the four locations studied, but AWEI_NSH performed the best overall across the four locations, and SVM was the best performing machine learning technique. Rather, each method's performance was shown to depend on the objects present in the image (e.g., built-up, shadows, vegetated shorelines, narrow waterbodies, etc.) and how successfully the method was able to classify those objects properly. This is in-line with current understanding of spectral index methods' performance, and we provide recommendations to aid remote sensing data users in choosing a suitable method based on their image's characteristics. Using these recommendations, we hope that the quality of surface water maps derived from multispectral satellite imagery will be improved for all disciplines that utilize such data by allowing users to choose the method that is best fit to the application.

Nomenclature

AWEI	Automated water extraction index
KNN	K-nearest neighbour
LWIC	Land water image classification
MLC	Maximum likelihood classifier
MNDWI	Modified normalized difference water index

---

* Corresponding author.
E-mail address: kkirb087@uottawa.ca (K. Kirby).

https://doi.org/10.1016/j.rsase.2024.101367
Received 16 February 2024; Received in revised form 18 September 2024; Accepted 30 September 2024

NDWI	Normalized difference water index
RT	Random trees
SVM	Support vector machine
S2/MSI	Sentinel-2 Multispectral Instrument

1. Background and research need

Remotely sensed satellite imagery is a valuable source of data to understand the impact of humans on the environment and the impact of the environmental processes on humans at a moment or over a period of time. Satellite imagery can be used to investigate a variety of anthropogenic activities and their consequences as well as environmental phenomena, including deforestation, agricultural intensification, urbanization, flooding, droughts, natural fires, and volcanic activity. Remotely sensed satellite data are generally processed via classification methods into Land Cover and Land Use (LCLU) data and are valuable resource to produce large Earth Observation (EO) datasets and Satellite Image Time Series (E. D. Chaves et al., 2020). A subcategory of LCLU data are remotely sensed satellite data that are classified into just two categories: land and water pixels. This is done to closely analyze the relationship between humans and the Earth's surface water. Surface water mapping or land-water image classification (LWIC) of satellite imagery is a popular research topic and has included detection of dam surface water (Yilmaz et al., 2023), floods (Ireland et al., 2015), rice crops (Boschetti et al., 2014), coastal lagoons (Senel et al., 2020), wetlands (Pena-Regueiro et al., 2020), and lakes (Roach et al., 2012; Thati et al., 2021). Additionally, land-water classified satellite data have been used to track changes in surface water extent (Kseňak et al., 2022) or shoreline deviation (Alcaras et al., 2022; Manaf et al., 2016) for climate change purposes. However, successful classification of remotely sensed satellite data into land and water pixels is a process that can be impacted by numerous factors (Khatami et al., 2016), and many methods to perform LWIC exist with varying levels of accuracy. Thus, it is the goal of this study to test commonly used LWIC methods to identify which method performs best overall and to identify what characteristics or objects within the image impact the quality of the surface water maps produced by each method.

Most applications of remote sensing satellite data classification have been developed for moderate spatial resolution instruments (e.g., the Moderate Resolution Imaging Spectroradiometer or MODIS at 250 × 250 m) (Bolton et al., 2020; E. D. Chaves et al., 2020). More recently, there has been heavy emphasis on utilizing the medium-resolution (e.g., 30 × 30 m) Landsat image series, specifically Landsat-5 Thematic Mapper (L5/TM), Landsat-7 Enhanced Thematic Mapper (L7/ETM), and Landsat-8 Operational Land Imager (L8/OLI), for EO applications, as medium-resolution data are more suitable to detect changes in human-environment systems at regional scales (Chen et al., 2015; Huang et al., 2020). The Landsat resources cover a timespan from 1984 to present and have a repeat cycle of 16 days. In contrast, similar products from the Sentinel-2 MultiSpectral Instrument (S2/MSI) cover a timespan from 2015 to present and have a repeat cycle of 5 days. This shorter time span of available imagery may explain why, comparatively, less image classification research and application has been conducted using S2/MSI, despite the fact that several S2/MSI bands have a finer spatial resolution than the Landsat resources. For example, S2/MSI bands 2, 3, 4, and 8 have a spatial resolution of 10m compared to L8/OLI which has a spatial resolution of 30m (Department of the Interior and U.S. Geological Department of the Interior, U.S. Geological Survey, 2021; ESA, 2023).Table. 1

LWIC relies on the principle that water pixels and non-water pixels will have different reflectance values. For multispectral imagery (L8/OLI and S2/MSI), each band of the imagery represents a different wavelength of light (see Table 2 for S2/MSI bands and comparable L8/OLI bands), and water more readily absorbs certain wavelengths of light than others. In particular, light in the near infrared (NIR) and shortwave infrared (SWIR) wavelengths are more readily absorbed by water than land. This difference in absorption can be enhanced through spectral indexes or classical machine learning methods, to name a few. The simplest LWIC method is Density Slicing where a single band of the imagery is used, and a pixel value threshold is selected to classify image pixels into the land category or the water category (Bijeesh and Narasimhamurthy, 2020). Multiple slices or intervals can be applied to classify the image pixels into

Table 1
Study locations and characteristics.

Location	Area [km^2]	Date of high-resolution image	Source and resolution of high-resolution image	Date of S2/MSI image	River flow rate[a] [m^3/s]	Mean river width [m]	Surficial Geology [b]
Red River (RedR)	33	2019/06/30	PLÉIADES 1 50 cm	2019/06/30	171	80	Silt and clay, locally containing stone
Sylvia Grinnell River (SGR)	39	2018/07/21	WorldView-2 50 cm	2018/07/21	178	120	Till blanket; diamicton; thick and continuous
Rivière-Rouge (RougeR)	25	2019/09/29	TripleSat 80 cm	2019/09/21	61	50	Sand and gravel; variable thickness
Fraser River (FR)	30	2020/09/08	KOMPSAT-3 70 cm	2020/09/07	2110	75	Colluvial deposits, undifferentiated; blocks, rubble with sand and silt with minor clay

[a] River flow information from Water Survey of Canada (Government of Canada, 2023).

[b] Surficial Geology from the Gridded Soil Landscapes of Canada (Agriculture and Agri-Food Canada; Science and Technology Branch; Government of Canada, 2016)

Table 2
S2/MSI imagery resolutions, wavelengths, and comparable L8/OLI bands (Department of the Interior and U.S. Geological Department of the Interior, U.S. Geological Survey, 2021; ESA, 2023).

Band	Resolution [m]	Central wavelength [μm]	Description	L8/OLI equivalent band (central wavelength)
1	60	0.443	Coastal & aerosol	–
2	10	0.490	Blue	1 (0.482)
3	10	0.560	Green	2 (0.561)
4	10	0.665	Red	3 (0.656)
5	20	0.705	Visible and NIR	–
6	20	0.740	Visible and NIR	–
7	20	0.783	Visible and NIR	–
8	10	0.842	Visible and NIR	4 (0.865)
8a	20	0.865	Visible and NIR	–
9	60	0.940	SWIR	–
10	60	1.375	SWIR	–
11	20	1.610	SWIR	5 (1.609)
12	20	2.190	SWIR	7 (2.201)

categories corresponding to the likelihood of the pixel representing water or land (Bijeesh and Narasimhamurthy, 2020). Instead of using a single band of imagery, a spectral index can be used to make use of multiple bands of imagery to better differentiate between land and water. Common spectral index-based LWIC methods include the Water Ratio Index (WRI) (Shen and Li, 2010), the Normalized Difference Water Index (NDWI) (McFeeters, 1996), the Normalized Difference Vegetation Index (NDVI) (Goward et al., 1991), the Modified Normalized Difference Water Index (MNDWI) (Xu, 2006), and the Automated Water Extraction Index Shadow (AWEI_SH) and No Shadow (AWEI_NSH) (Feyisa et al., 2014). Spectral index-based methods can only be applied to multispectral imagery. The general categories of common classical machine learning based methods include Support Vector Machine (SVM) (Cortes and Vapnik, 1995), Random Trees (RT) or Random Forest (RF) (Breiman, 2001), Maximum Likelihood Classifier (MLC), K-nearest Neighbour (KNN) (Cover and Hart, 1967), K-means Clustering Algorithm (Hartigan and Wong, 1979), Artificial Neural Network (ANN) (Schalkoff, 1997), and Iterative Self-Organizing Data Analysis Technique (ISODATA) (Ball and Hall, 1965). Many more LWIC methods exist beyond those stated (Bijeesh and Narasimhamurthy, 2020), and countless more exist when considering the combination of spectral index-based methods and machine learning based methods (e.g., applying machine learning methods to an image that has had a spectral index applied to it). Studies that have compared the success of these methods at LWIC have produced diverging results. Many studies have found that NDWI produced superior results in their study location compared to other indexing methods (Alcaras et al., 2022; Aroma and Raimond, 2021; Doña et al., 2016; Mukherjee and Samuel, 2016; Pena-Regueiro et al., 2020; Zhang et al., 2022), while some studies specify NDWI only performed best under specific conditions (e.g., it performs best in high mountains (Wajiha Khalid et al., 2021) or shallow waters (Yilmaz et al., 2023)) or even that NDWI performed poorly overall compared to other methods (Gudelj et al., 2018; Ireland et al., 2015; Senel et al., 2020). Other studies have demonstrated that AWEI ((Feyisa et al., 2014; Senel et al., 2020)), SVM ((Feizizadeh et al., 2023; Li et al., 2022; Manaf et al., 2016)), or RF ((Gudelj et al., 2018; Otukei and Blaschke, 2010; Singh and Pandey, 2021; Yilmaz et al., 2023)) produced the most accurate LWIC data in their study location. Additional observations include that MNDWI performed best in clear water (Sun et al., 2012), that MLC performed best in scene with built up environment (i.e., urban infrastructure) (Sun et al., 2012), that AWEI, SVM, and RF performed best in hilly, flat regions (Wajiha Khalid et al., 2021), and that ISODATA was the best performing method in the study area (Thati et al., 2021). These examples illustrate the range and inconsistency of findings from previous research focused on assessing suitability of LWIC methods. These inconsistencies may be linked to site- or condition-specific factors that influence the performance and suitability of various LWIC methods.

Some research gaps are present that have not been fully addressed by existing literature comparing the performance of various LWIC methods. First, it is uncommon for the LWIC methods to be compared in multiple locations with different objects and characteristics present in the imagery. Therefore, sensitivity to various site- and condition-specific factors such as vegetation cover, soil types, built-up, and the presence of shadows is not well understood. In addition, the spectral index-based methods that have been evaluated in past studies were often applied using a single threshold, often selected using fundamental thresholding algorithms such as Otsu's method. In some studies, a threshold of zero was chosen (Sun et al., 2012; Wei et al., 2020), and, rather than testing the success of the method at detecting water, this tested the stability of the spectral based-index's threshold around the number zero. Most often, the Otsu global threshold selection algorithm was applied (Ireland et al., 2015; Kseňak et al., 2022; Li et al., 2022; Thati et al., 2021), or it was not explicitly stated how thresholding was conducted. Therefore, the degree to which individual LWIC methods were optimized prior to comparison with other LWIC methods is often unclear or not fully considered in past research. Threshold selection sensitivity across multiple spectral indices and across multiple image locations, in the interest of optimizing the accuracy of classified land and water pixels, has been a rarely studied subject. However, the comparison of the performance of various automated thresholding algorithms has been explored thoroughly in past research at a single waterbody location (Karaman, 2021; Sekertekin, 2021). This research found that the minimum thresholding method produced the most accurate classification at their study site (Karaman, 2021; Sekertekin, 2021), although Otsu's method is generally the most widely applied automated threshold selection algorithm. Additionally, error analysis of LWIC methods has sometimes been performed with medium resolution imagery (Boschetti et al., 2014; Sekertekin, 2021; Wei et al., 2020), so small differences in classification results caused by the different LWIC methods compared could not be observed at a fine resolution. Finally, image classification results have often been assessed by evaluating the classification accuracy at randomly selected points in the image (i.e., the kappa coefficient approach) (Doña et al., 2016; Wajiha Khalid et al., 2021; Mansaray et al., 2020;

Pena-Regueiro et al., 2020; Yilmaz et al., 2023; Zhang et al., 2022) or in-situ measured shoreline points (Karaman, 2021). However, these approaches can limit the observations that are able to be made regarding the image characteristics that cause misclassifications (e.g., wet sand, certain types of infrastructure or vegetation, etc.) compared to when performing validation using high resolution imagery.

Due to these aforementioned research gaps, this study attempts to.

1. Determine if there is an ideal method to perform LWIC of satellite imagery across multiple locations and to
2. Identify the characteristics of images that can cause misclassifications among these commonly applied LWIC methods.

This is the first study that this degree of comparison was performed for surface water mapping, as nine methods were compared across four diverse images with varied image characteristics across Canada. To ensure fair comparison, each image was prepared to the same level of pre-processing (Section 3.1), and each method was optimized by testing the stability of the thresholds for spectral index-based classifiers (Section 3.2). The optimized thresholds were also tested against the thresholds selected via Otsu's method, to understand the ability of Otsu's method at selecting a suitable threshold. Additionally, the effects of resampling the 20 m multispectral bands of S2/MSI imagery to 10 m were evaluated. The success of each method was tested using two baselines: the critical success index (CSI, Section 3.3) and the root mean squared error of river widths (Section 3.4). It is hoped that this work will be used to guide individuals to use the LWIC methods and optimization techniques best suited for their image of interest in order to create high quality surface water maps.

2. Study location and data acquisition

The S2/MSI imagery evaluated in this paper was open-source and downloaded through Copernicus Open Access Hub. S2/MSI imagery was chosen for analysis in this paper as it is used extensively in surface water mapping applications. It provides a higher resolution than the Landsat series, although the Landsat series has greater temporal coverage. Because of the similarities between these two multispectral imagery catalogues (see Table 2 for similar band wavelengths between the two products), the findings in this paper are expected to also be useful for the Landsat series. The S2/MSI imagery is available in two levels of processing: top of atmosphere

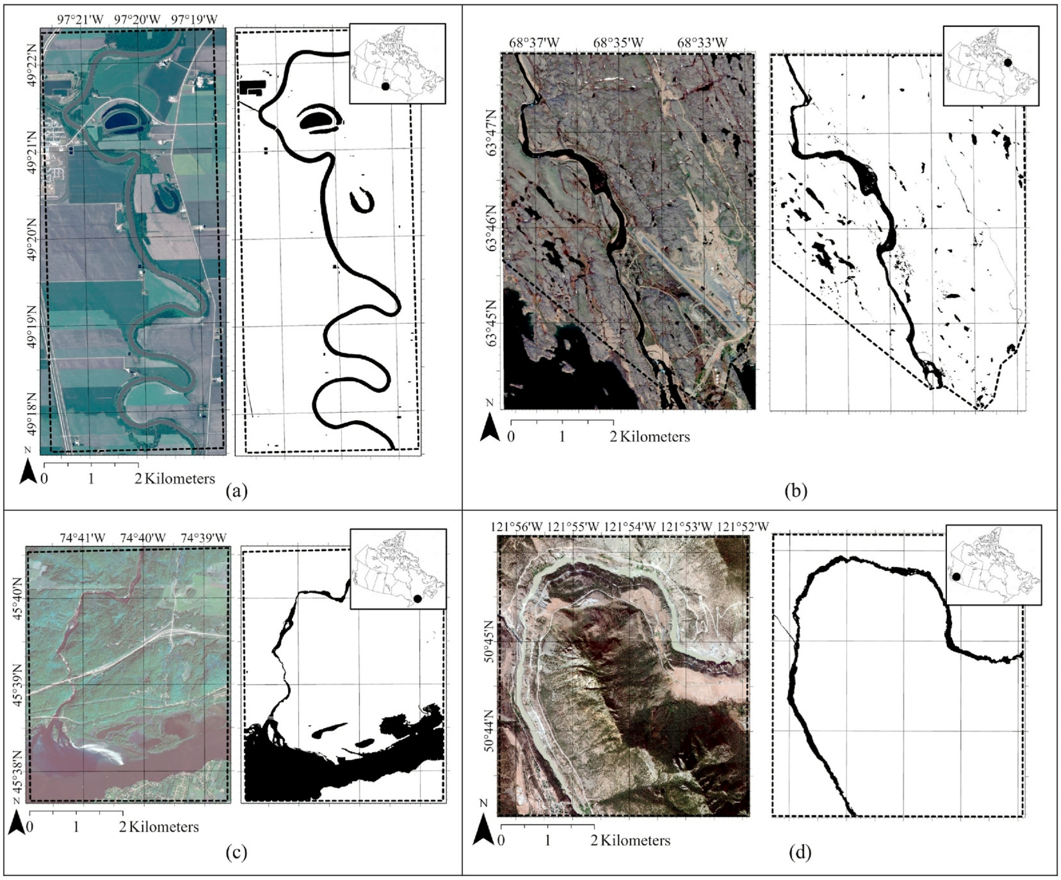

Fig. 1. Study locations and water polygons derived from high resolution imagery for validation purposes: (a) Red River (RedR), (b) Sylvia Grinnell River (SGR), (c) Rivière-Rouge (RougeR), and (d) Fraser River (FR). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(TOA) in Level 1C and bottom of atmosphere (BOA) in Level 2A. Level 2A products are only available for dates 2019 to present and Level 1C is available for 2015 to present.

Four locations in Canada were chosen for analysis based on the following considerations in an attempt to diversify the image characteristics as much as possible: soil, vegetation, built-up, depth and turbidity of water, high albedo surfaces, presence of shadow, presence of snow, and seasonality of imagery. The four locations are shown in Fig. 1 and are the Red River (RedR) near Morris, Manitoba, in UTM zone 14N, the Sylvia Grinnell River (SGR) near Iqaluit, Nunavut, in UTM Zone 19N, the Rivière-Rouge (RougeR) and Ottawa River near L'Orignal, Ontario, in UTM Zone 18N, and the Fraser River (FR) near Xaxli'p First Nation, British Columbia, in UTM Zone 10N. The image for the RedR location was taken during a high flow event (50th percentile flow is 66), the image for the SGR location was taken during a high flow event (50th percentile flow is 58), the image for the RougeR location was taken during a mid-flow event (50th percentile flow is 71), and the image for the FR location was taken during a high flow event (50th percentile flow is 1010). Because all study locations were selected within Canada, there may be a lack of generalizability to the study. However, the RougeR and FR locations are situated approximately 3500 km apart in the longitudinal direction, and the RougeR and SGR locations are situated approximately 2100 km apart in the latitudinal direction. Thus, the breadth of image landcover, vegetation, and climates are varied between the four image areas. Table A1 in the supplementary material lists all landcover classes included in the image areas. Additionally, the effects of seasonal variation (e.g., additional hydrological conditions at each location) were not explored in this work as a single hydrological condition was analyzed for each image area.

The Landcover of Canada dataset (Canada Centre for Remote Sensing; Natural Resources Canada; Government of Canada, 2015) (Table A1 in supplementary material) was used to ensure the locations were varied enough to draw conclusions about the influence of image characteristics on the success of LWIC. Snow and ice are included in the SGR image as the river is located in the Arctic, and shadows are present in the FR image as the river is located in a canyon. Snow/ice and shadows are often difficult to discern from liquid water in satellite imagery. Some other characteristics that were captured in the various locations were: different turbidity in waterbodies (all four locations), many small bodies of water (SGR), uniform river size (RedR), narrow river size and rapids (RougeR and SGR), built-up (all four locations), and river confluences (RedR, RougeR, FR). Table A1 summarizes the landcover classes that are represented in each of the four locations. Although detailed quantitative data related to the turbidity and water quality (e.g., salinity, sediment concentration, chl-a) for each of the waterbodies in the four locations was not available, quantitative observations and generalized qualitative data were available from personal experience, field data collection at these locations (Kirby et al., 2023;

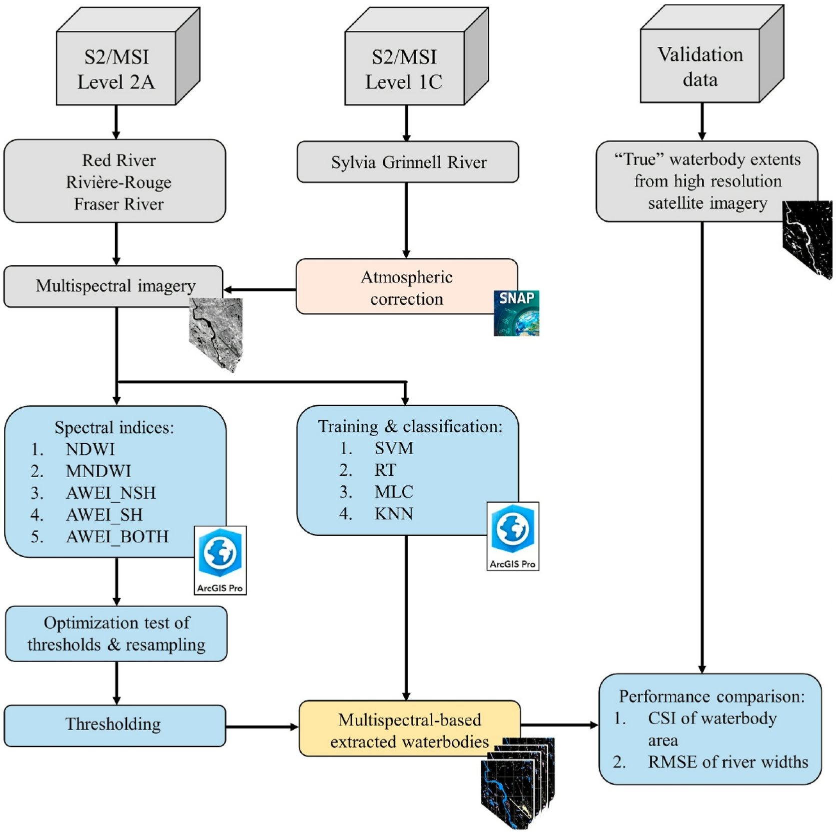

Fig. 2. Methodology framework. Grey boxes represent processing inputs, yellow boxes represent intermediate processing products, orange boxes represent image pre-processing steps, and blue boxes represent main processing steps. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Venditti et al., 2014), and historical measurements from the Water Survey of Canada (Government of Canada, 2024). The Red River in the RedR image location has high turbidity and low salinity. The average suspended sediment concentration across 1956 to 1961 in the Red River was 303 mg/L. The Sylvia Grinnell River in the SGR image location has very low turbidity (nearly clear) and low salinity. Suspended sediment data was not available for the Sylvia Grinnell River. The Rivière-Rouge and Ottawa River in the RougeR image location have medium turbidity and low salinity. The average suspended sediment concentration across 1977 to 1979 in the Ottawa River was 37 mg/L. The Fraser River in the FR image location has high turbidity and low salinity. The average suspended sediment concentration across 1965 to 1979 in the Fraser River was 125 mg/L.

To perform comparison analysis, validation data were generated from high resolution imagery (shown in Fig. 1). Water polygons were manually drawn from the high-resolution images, and it is assumed that there are negligible errors associated with these water extents. The S2/MSI images were chosen to be as close to the high-resolution imagery date as possible with cloud-free conditions. The dates of all the images are presented in Table 1, as well as the high-resolution imagery sources and resolutions. Level 2A S2/MSI imagery was used for the RedR, RougeR, and FR sites. For the SGR site, only Level 1C S2/MSI imagery was available for the date of interest.

3. Methods

3.1. Image preparation

No pre-processing was required for the Level 2A S2/MSI images which are already corrected to BOA reflectance. The Level 1C S2/MSI image for the SGR site was converted from TOA to BOA reflectance using the Sen2Cor tool (Main-Knorn et al., 2017) in the SeNtinel Application Platform (SNAP). The Sen2cor algorithm begins by applying a series of threshold tests to identify clouds, snow, and other image features. The algorithm then uses these features' reflective properties to calculate atmospheric and surface parameters for future processing steps. Then atmospheric correction is completed based on aerosol and ozone values related to the image's geographic location. The pre-processing and main processing steps for the S2/MSI images are shown in Fig. 2.

3.2. Image processing and optimized thresholding

3.2.1. Spectral indices

For multispectral imagery, LWIC was conducted using both spectral index-based and classical machine learning-based methods, as illustrated in Fig. 2. The spectral indices applied were NDWI, MNDWI, AWEI_NSH, and AWEI_SH. The equations for these indices are shown below where the subscripts of each wavelength correspond to the bands presented in Table 2. These indices are also applicable for use with bands of similar wavelengths for Landsat 8 imagery, and thus it is likely that this study's findings are also applicable to the Landsat imagery catalogue. The processing was completed using ArcGIS Pro Version 3.03.

$$NDWI = \frac{\rho_{green} - \rho_{NIR}}{\rho_{green} + \rho_{NIR}} \tag{1}$$

$$MNDWI = \frac{\rho_{green} - \rho_{SWIR1}}{\rho_{green} + \rho_{SWIR1}} \tag{2}$$

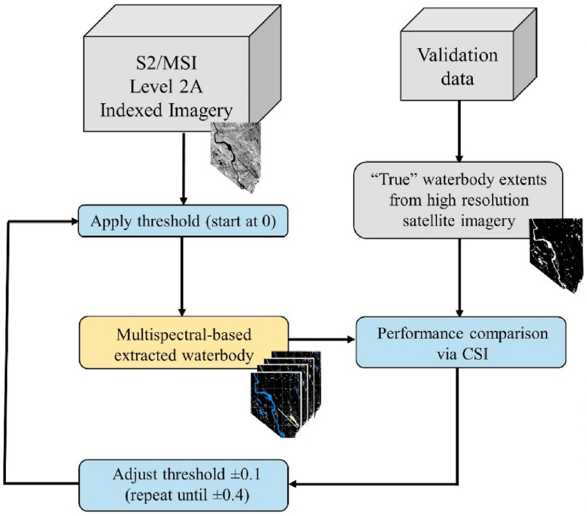

Fig. 3. Methodology of manual threshold adjustment procedure. Grey boxes represent processing inputs, yellow boxes represent intermediate processing products, and blue boxes represent main processing steps. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

$$AWEI\_NSH = 4 \times \left(\rho_{green} - \rho_{SWIR1}\right) - \left(0.25 \times \rho_{NIR} + 2.75 \times \rho_{SWIR2}\right) \tag{3}$$

$$AWEI\_SH = \rho_{green} + 2.5 \times \rho_{blue} - 1.5 \times \left(\rho_{NIR} + \rho_{SWIR1}\right) - 0.25 \times \rho_{SWIR2} \tag{4}$$

where ρ_{blue} is the reflectance of the blue band (band 2 in S2/MSI), ρ_{green} is reflectance of the green band (band 3 in S2/MSI), ρ_{NIR} is the reflectance of the NIR band (band 8 in S2/MSI), ρ_{SWIR1} is the reflectance of the first SWIR band (band 11 in S2/MSI), and ρ_{SWIR2} is the reflectance of the second SWIR band (band 12 in S2/MSI). An additional parameter, AWEI_BOTH, was generated by merging the waterbody areas detected by both AWEI_NSH and AWEI_SH.

3.2.2. Resampling

As shown in Table 2, bands 2, 3, 4, and 8 (blue, green, red, NIR) have a pixel resolution of 10m, but the two SWIR bands (11 and 12) have pixel resolution 20m. The indices MNDWI, AWEI_NSH, and AWEI_SH include one or more SWIR band. Thus, the impact of resampling the SWIR band from 20m to 10m on the LWIC results was also studied for each location. The resampling algorithm used in this study was the cubic resampling algorithm. It resamples by performing a cubic convolution, and the updated cell value is calculated by fitting a curve through the 16 nearest cell centres. This approach was suitable as the raster data used in the study was continuous.

3.2.3. Thresholding procedure

For each image and each spectral index-based classifier, the threshold selected to distinguish water from land was optimized manually to: (1) be able to compare the results of each method while optimized in order to diminish the influence of the threshold selection on the success of the methods and (2) to compare against the threshold selected by Otsu's method. The methodology of the manual threshold adjustment is present in Fig. 3.

By comparing the manually determined optimized threshold against the Otsu threshold, the reliability of Otsu's method was tested. Otsu's method of determining an optimal threshold consists of minimizing the intra-class variance between the background and foreground classes by utilizing the zeroth and the first order cumulative moments of the grey-level histogram (Otsu, 1979). To manually determine the optimal threshold for NDWI, MNDWI, AWEI_SH, AWEI_NSH, and AWEI_BOTH, a sensitivity test for the optimal threshold was performed by varying the threshold by 0.1 intervals from +0.4 to −0.4. As such, the following thresholds were tested for each spectral index: 0.4, 0.3, 0.2, 0.1, 0, −0.1, −0.2, −0.3, and −0.4. The success of the resulting classified image produced by each threshold was then tested using the CSI (see Equation (5)). The threshold with the highest CSI was the optimized threshold and was used in future comparisons.

3.2.4. Machine learning algorithms

The classical machine learning-based LWIC methods compared in this paper are SVM, RT (or RF), MLC, and KNN. The classical machine learning tools were tested using the default user settings within the ArcGIS Pro environment, as described herein. This enabled comparison of machine learning tools that are readily available for processing multiple images over large domains to identify waterbodies. SVM is based on statistical learning theory (Smola and Schölkopf, 2004) which is used to locate boundaries of different classes. The SVM algorithm applied in this study used nonlinear radial basis function with a 2S grid and optimal parameter pairs (C, gamma) for the kernels used (Li et al., 2022). Additionally, it does not assume that the samples follow a normal distribution. The default value of 1000 training samples per class was used for the SVM technique. RT (or RF) first uses training samples to generate a discriminant function and then builds branches according to differing values. The process is repeated between lower nodes and branches in each subset until the classification tree is formed (Li et al., 2022). The random forest algorithm developed by Leo Breiman and Adele Cutler (Breiman, 2001) was used, and the function is resistant to overfitting. Default numbers of 50 maximum trees, 30 maximum tree depth, and 1000 samples per class were used for the RT function. MLC is a parametric classifier based on the Gaussian probability density function model, which essentially calculates the possibility that each pixel can be included into a particular class (Paola and Schowengerdt, 1995). KNN classification, the input for the training is the "k" closest examples in the samples. The training samples are assigned to their respective classes via the training process (Cover and Hart, 1967). Default values of 1 neighbour and 1000 samples per class were used for the KNN technique.

For each study location, the high-resolution imagery was divided into two halves. One half was used to train each machine learning method. The remaining half was used to evaluate the results of the LWIC. Therefore, the training/validation split for each location was 50/50. The spatial split of each of the four high-resolution images between training and validation was performed in an attempt to include the unique image characteristics (e.g., built up, high albedo surfaces, shadows, snow/ice, etc.) in both the training and validation samples.

3.3. Accuracy analysis

3.3.1. Comparison on the basis of waterbody area

Error analysis was performed by comparing the predicted waterbody area to the actual waterbody area (i.e., the waterbody area manually delineated from high-resolution satellite imagery), rather than comparison based on randomly selected points. The advantage of this comparison over randomly selected points is that many observations across the entire image's extent can be made regarding the characteristics in the images that cause misclassification, rather than observations that could be made only at the locations of the randomly selected points. Additionally, this method ensures that small differences between the results of the LWIC methods are captured. Error analysis was performed using the CSI, shown below:

$$CSI = \frac{A}{A+B+C} \tag{5}$$

where A are the correct predictions made by the model (in km2), or hits, B are the overpredictions by the model (in km2), or false alarms, and C are the underpredictions by the model (in km2), or misses (Lim and Brandt, 2019). The CSI is commonly used as a verification statistic for flood modelling accuracy (Sreekumar et al., 2022). Supplementary to these definitions for flood modelling, A, B, and C will be defined as the matching area (MA), the omission error (OE), and the commission error (CE) for the purposes of satellite imagery classification herein. In short, OE is where water was predicted at locations where there is no water in reality, and CE is where land was predicted and where there was no land in reality.

3.3.2. Comparison on the basis of waterbody boundary

In addition to evaluating the success of LWIC methods at detecting the correct waterbody area, analysis of the widths of the rivers extracted was performed to evaluate the success of the LWIC methods at detecting the waterbody boundaries. This method has been used in floodplain modelling accuracy assessment as well (Lim and Brandt, 2019; Sreekumar et al., 2022). River centerlines were generated using the true river extents and any flow split (caused by islands, sandbars, etc.) had an associated centerline. Perpendicular transects were generated along the river centerlines at 100m intervals. The transects were clipped to the true waterbody extent for the measured river width. The transects were also clipped to the extracted waterbody extent for the estimated river width. The two widths (measured and estimated) were compared using root mean squared error (RMSE) and mean absolute error (MAE), shown below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(w_i - \widehat{w}_i\right)^2}{N}} \tag{6}$$

$$MAE = \frac{\sum_{i=1}^{N}\left|w_i - \widehat{w}_i\right|}{N} \tag{7}$$

where N is the number of transects or widths, w_i is the width measured from the water polygons derived from S2/MSI imagery, and $\widehat{w}_i$ is the width measured from the water polygons derived from the high-resolution satellite imagery.

4. Results

4.1. Threshold optimization, assessment of Otsu's method, and effect of resampling

The manually determined optimized thresholds (MT) used to distinguish water from land using each index method on S2/MSI data are illustrated in Table 3. Across the four study locations, the methods that produced the smallest range in optimal thresholds were AWEI_SH and AWEI_BOTH. Both had a range of 0.2. The other methods' threshold ranges were not much higher at 0.3. A smaller range in optimal threshold is desirable when processing many tiles of imagery, as the optimal threshold would be less likely to vary from image to image. The optimal thresholds for AWEI_SH, AWEI_NSH, and AWEI_BOTH trended negative, while the optimal thresholds for NDWI and MNDWI were closer to zero.

The reliability of Otsu's method in selecting the most optimal threshold varied across the images and LWIC methods. Fig. 4 shows the thresholds computed using Otsu's method (Otsu's threshold or OT), the manually determined optimized thresholds (MT), and the histogram of pixel values for each LWIC method at each study location. The hollow points along the histogram curves indicate the thresholds determined via Otsu's method, and the solid points indicate the thresholds determined manually through optimization. The legends at the top right of each graph show the difference between the OT and MT. The OT was the closest to the MT when using the MNDWI spectral index, where the mean absolute difference (represented by MAE) between the OT and MT across the four images was the smallest. The OT was consistently lower than the MT when using both the AWEI_NSH and AWEI_SH spectral indices. The OT had largest difference from the MT when using the AWEI_NSH spectral index.

When grouping the comparison of OT to MT by location rather than by LWIC method, the following average differences between

Table 3
Optimized (manually determined) thresholds for each spectral index method and each of the four locations, and the impact of resampling on the success of LWIC.

Location (Water area)	Metric	NDWI	MNDWI	AWEI_SH	AWEI_ NSH	AWEI_ BOTH
Red River (2.34 km^2)	MT Resample[a]	−0.1 –	−0.1 Yes	−0.2 Yes	−0.4 Yes	−0.2[b] Yes
Sylvia Grinnell River (2.11 km^2)	MT Resample[a]	0 –	−0.1 No	−0.1 Yes	−0.4 Yes	−0.1[b] Yes
Rivière-Rouge (5.18 km^2)	MT Resample[a]	−0.2 –	−0.2 No	−0.1 No	−0.2 No	−0.1[b] No
Fraser River (1.15 km^2)	MT Resample[a]	0.1 –	0.1 Yes	0 Yes	−0.1 Yes	0[b] Yes
Threshold Range	–	−0.2 to 0.1	−0.2 to 0.1	−0.2 to 0	−0.4 to −0.1	−0.2 to 0

[a] Resample indicates the impact of resampling on CSI. "Yes" indicates that a higher CSI was achieved when the image was resampled to a finer resolution.
[b] These values represent the selected threshold for both AWEI_SH and AWEI_NSH.

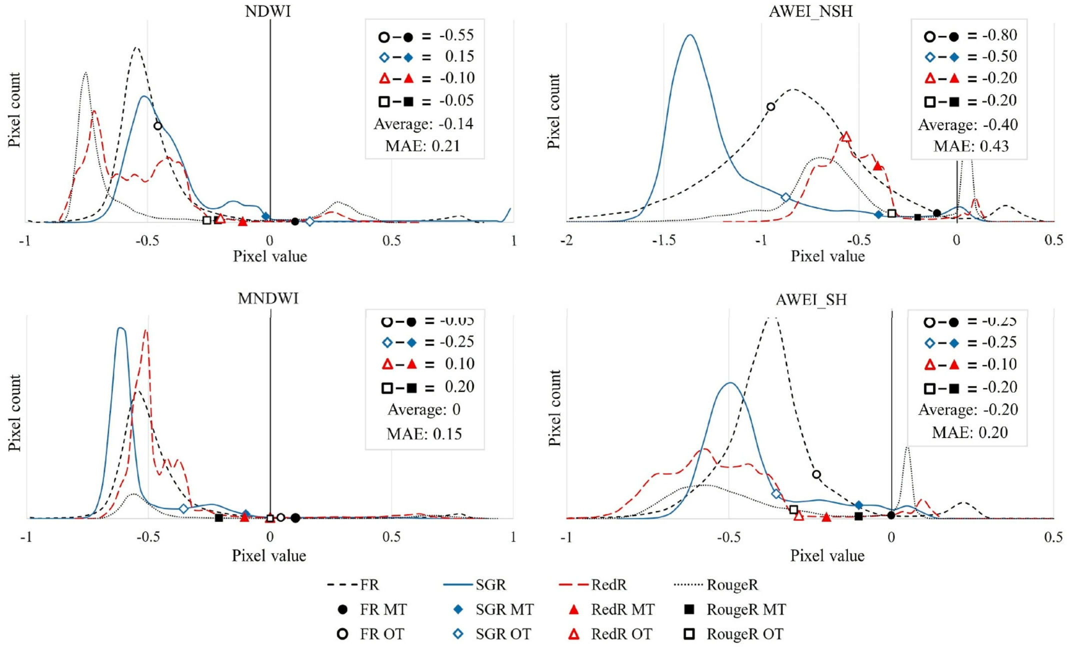

Fig. 4. Manually determined optimal threshold between land and water classes compared to the threshold determined via Otsu's method.

OT and MT were calculated. For the RedR location, the average difference was −0.08 (−0.10, 0.10, −0.20, and −0.10 for NDWI, MNDWI, AWEI_NSH, and AWEI_SH, respectively). For the SGR location, the average difference was −0.21 (0.15, −0.25, −0.50, and −0.25), which was the second highest difference by location. For the RougeR location, the average difference was −0.06 (−0.05, 0.20, −0.20, and −0.20), which was the lowest difference by location. For the FR location, the average difference was −0.41 (−0.55, −0.05, −0.80, and −0.25), which was the highest difference by location and the difference between the OT and MT was consistently negative (i.e., OT was consistently higher than MT). Of the four locations, the FR location contained the most shadow due to the canyon walls. Additionally, the FR image's histogram had some of the largest separation of water and non-water pixels peaks (see the black dashed histograms in Fig. 4). This large separation in histogram peaks, with the water peak (i.e., the peak in the positive pixel value range) amplitude being relatively small, may have created difficulties with Otsu's method selecting a suitable threshold. The threshold optimization results for all four locations are presented in Tables A2–A6 in the supplementary material.

Resampling improved the results in the RedR, SGR, and FR images, but not in the RougeR image, as shown in Table 3. Resampling did not improve classification when the resulting extracted waterbody shorelines were not smoothed by this process. Rather, the densification of the pixels at the shorelines created unorganized, scattered, and misclassified water and non-water pixels that did not accurately represent the physical shoreline. The results of testing the effects of resampling at all four locations are presented in Tables A3–A6 in the supplementary material.

4.2. *Land water image classification of S2/MSI imagery – waterbody area basis*

The best performing method of LWIC across the four sites was not consistent. Upon using the CSI metric to perform the comparison based on total waterbody area, MLC produced the best results for both the RedR and SGR locations. AWEI_SH produced the best results for the RougeR location. NDWI produced the best results for the FR location. Overall, SVM performed well at all four sites. MNDWI was not the best performing method at any of the four sites. The OE, CE, MA, and CSI results are presented in Fig. 5, and the order of the methods' success in each image location are presented in Table 4. Upon summing all four CSI values resultant from each method and listing the methods from highest summed CSI to lowest summed CSI, the order of most successful to least successful method was as follows: AWEI_NSH (325.3%), AWEI_BOTH (319.4%), AWEI_SH and NDWI (318.2%), SVM (298.1%), MNDWI (297.2%), RT (280.9%), MLC (243.7%), and KNN (177.4%).

The best performing index for the RedR area (shown in Fig. 6) was AWEI_BOTH, and all AWEI indices produced a CSI at least 5% higher than MNDWI or NDWI. The best performing classical machine learning based method was MLC, and MLC slightly outperformed AWEI_BOTH. KNN performed the worst overall with a large amount of scatter, creating a high CE. Following the organization of Fig. 6, the results of each method are as follows. NDWI performed worst of the index methods due to high OE. In the threshold optimization exercise, NDWI was not able to capture the oxbow lakes as water without misclassifying the built up as water and thus increasing the CE and decreasing the CSI. MNDWI had similar issues, but with misclassifying the shoreline pixels along the river as land, leading to higher OE. AWEI_SH was able to detect most pixels correctly as water without misclassifying the built up as water, as indicated by the small OE and CE. AWEI_NSH produced a slightly higher CE than AWEI_SH because some land shoreline pixels were misclassified as water. The AWEI_BOTH result was very similar to AWEI_SH with similar OE and CE. Overall, MLC, SVM, and RT produced good results.

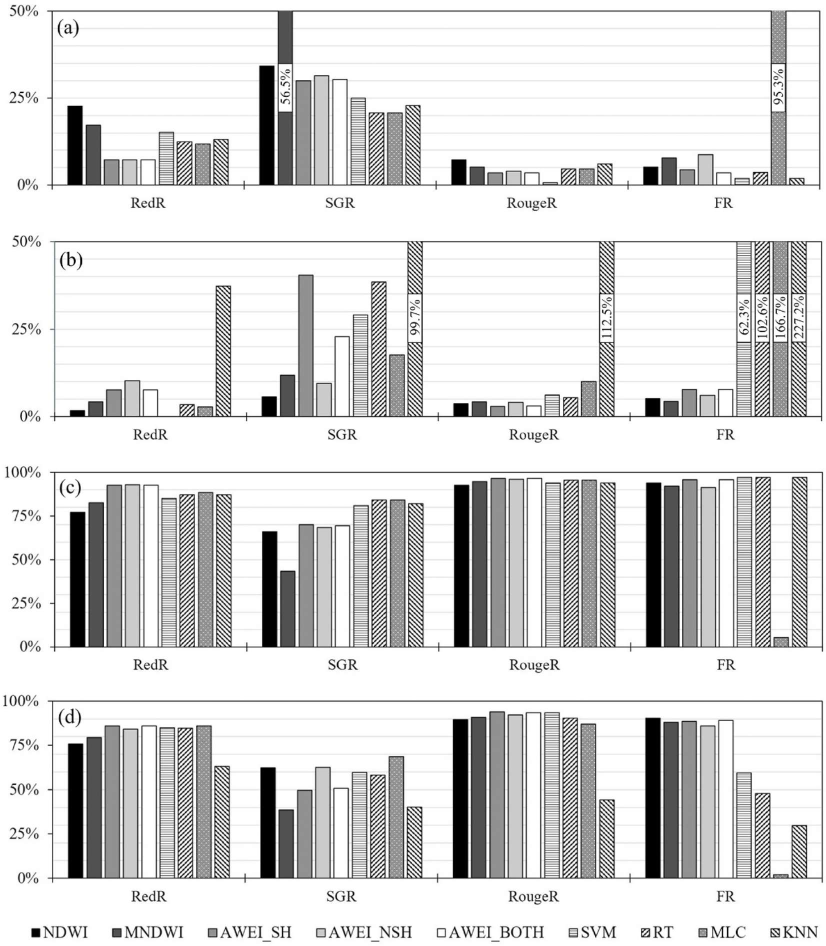

Fig. 5. Classification accuracy of LWIC methods on S2/MSI imagery based on (a) OE [%], (b) CE [%], (c) MA [%], and (d) CSI [%]. The OE, CE, and MA are normalized with the total waterbody area, and the values of the OE and CE outside of the 0%–50% range are shown.

Table 4
Order of classification success for each method in each image location with 1 being best and 9 being worst.

Image Location	1	2	3	4	5	6	7	8	9
RedR	MLC	AWEI_BOTH	AWEI_SH	SVM	RT	AWEI_NSH	MNDWI	NDWI	KNN
SGR	MLC	AWEI_NSH	NDWI	SVM	RT	AWEI_BOTH	AWEI_SH	KNN	MNDWI
RougeR	AWEI_SH	AWEI_BOTH	SVM	AWEI_NSH	MNDWI	RT	NDWI	MLC	KNN
FR	NDWI	AWEI_BOTH	AWEI_SH	MNDWI	AWEI_NSH	SVM	RT	KNN	MLC

The best performing index for the SGR image area (Fig. 7) was AWEI_NSH because it produced a lower CE compared to other indices, due to the correct classification of the runway. The best performing classical machine learning based method was MLC which produced the highest CSI by 10% compared to the other classical machine learning based methods. MLC also outperformed AWEI_NSH. The worst performing method was MNDWI. Following the organization of Fig. 7, the results of each method are as follows. NDWI was a close second to the best index method because it produced the lowest CE. MNDWI produced the lowest CSI because of high OE and low MA while also misclassifying the built-up as water, resulting in high CE. The CE from the AWEI_SH method was also high. The AWEI_NSH method produced much better results than AWEI_SH or AWEI_NSH because it properly classified the runway. The AWEI_BOTH results were affected by the inclusion of AWEI_SH, which increased the CE and CSI. For the classical machine learning based methods, all methods except MLC had high CE. Additionally, KNN produced a lot of speckle, resulting in further increased CE.

The best performing index method for the RougeR image area (Fig. 8) was AWEI_SH. The best performing classical machine learning based method was SVM, but AWEI_SH performed best overall. The method that performed the most poorly was KNN, as it produced speckle. Following the organization of Fig. 8, the results of each method are as follows. NDWI was the worst performing index method, due to high OE and CE. MNDWI also produced some CE, reducing the success of this method. AWEI_SH, AWEI_NSH, and AWEI_BOTH all performed well with varying success at extracting the location of the shoreline. SVM performed well at extracting the

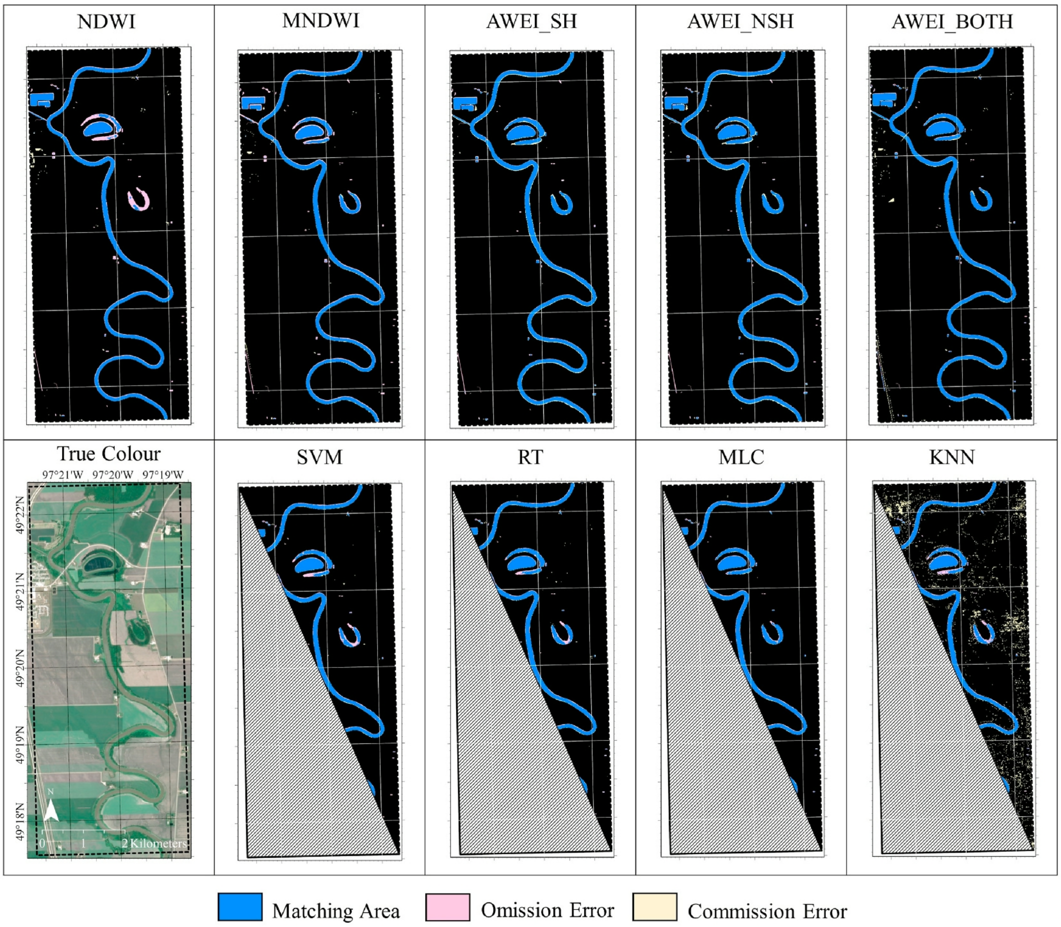

Fig. 6. Red River S2/MSI imagery classification results. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

shoreline of both the Ottawa River and Rivière-Rouge in the validation area. RT performed well both at shoreline extraction and at producing a low CE. MLC shoreline extraction was acceptable, but SVM and RT produced better results. Additionally, MLC produced some misclassifications of the highway pixels. KNN performed the worst of all methods because of the speckle that was produced.

The best performing index method for the FR image area (Fig. 9) was NDWI, but all spectral index-based methods were similarly successful. The best performing classical machine learning based method was SVM, but the indices performed better overall for this image. The worst performing method was MLC, which did not detect the river pixels as water, creating a high OE, and misclassified the shadows as water, also creating a high CE. SVM and RT produced classifications with high CE due to the shadows, and KNN produced high CE due to shadows as well as speckle.

4.3. Land water classification of S2/MSI imagery – waterbody boundary basis

River widths at 100m intervals extracted from the high-resolution validation data and extracted from the LWIC imagery were compared to quantify the success of each method at extracting the waterbody shorelines. The results are presented in Fig. 10.

In this analysis, AWEI_SH and AWEI_NSH were not tested separately and AWEI_BOTH was used instead to represent the success of the AWEI indices. The results between the most optimal method in terms of total waterbody area and in terms of waterbody boundary were not consistent across the four study locations. For the RedR area, the best performing method in terms of total waterbody area was MLC, whereas on a boundary-basis, the best performing method was AWEI_BOTH. AWEI_BOTH produced more CE in the RedR image, due to misclassifying buildings as water. This built-up was not included in the MLC validation area. AWEI_BOTH overall extracted the shoreline more accurately. Because of the uniformity of the river, all methods performed well using this metric. For the SGR area, the best performing method for both area- and boundary-based performance assessments was MLC; however, as shown with the CSI indicator in the waterbody area analysis, AWEI_NSH may have outperformed AWEI_BOTH in the SGR image. Thus, AWEI_NSH may have outperformed MLC on the waterbody boundary-basis if it had been tested separately from AWEI_BOTH. MLC had the lowest OE and second lowest CE, because the method captured most small waterbodies and did not misclassify the runway as water. Both the shorelines of the small waterbodies and the river were well defined by the MLC method. For the RougeR area, the best performing method based on waterbody area was AWEI_SH, but the best performing method based on the waterbody boundary was RT. RT had

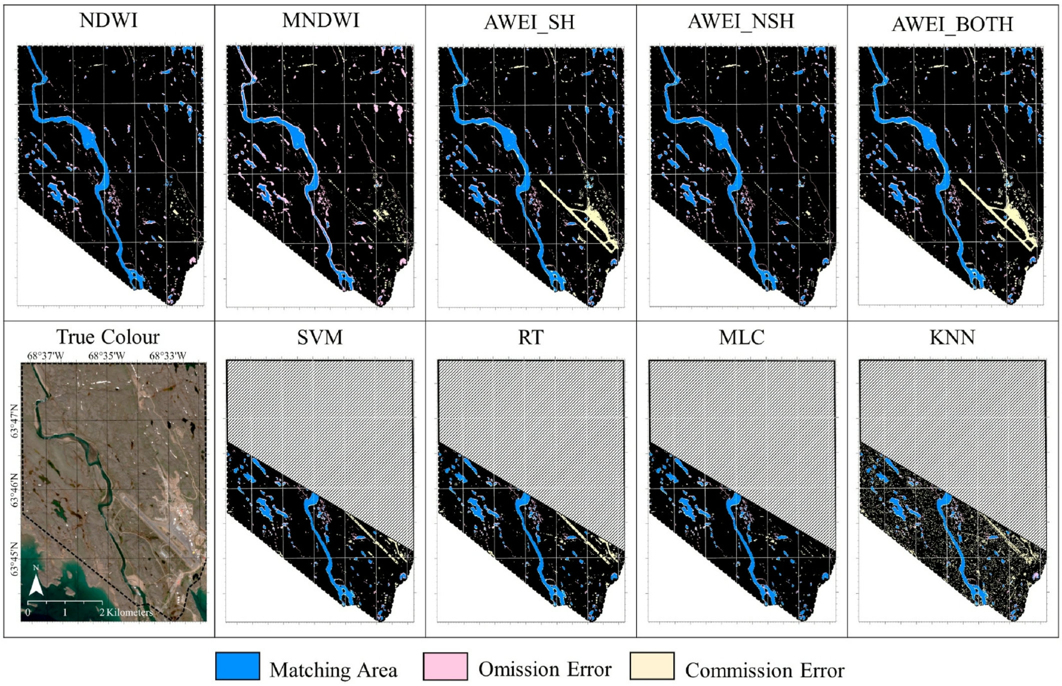

Fig. 7. Sylvia Grinnell River S2/MSI imagery classification results.

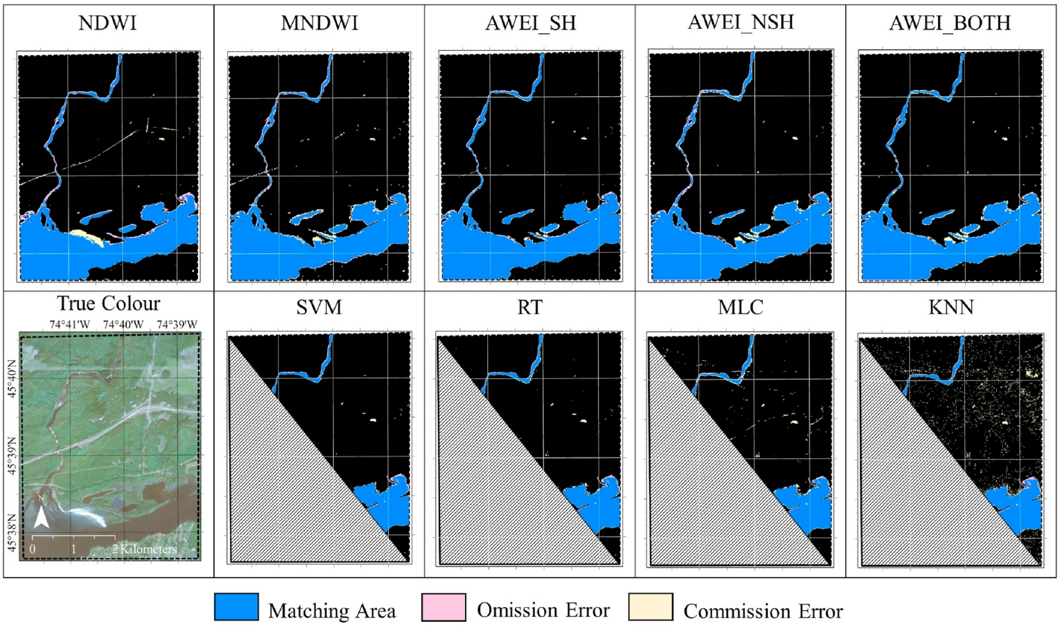

Fig. 8. Rivière-Rouge S2/MSI imagery classification results.

less OE and CE at the shorelines, and all classical machine learning-based methods performed better than the spectral index-based methods using the waterbody boundary basis. This may be due to the training area including more narrow areas of the Riviere Rouge and more wet sand and shallow vegetated shorelines on the Ottawa River in the training section of the image than the validation section. For the FR area, the best performing method for the waterbody area basis was NDWI while the best performing method for the waterbody boundary basis was AWEI_BOTH. All index methods extracted more accurate shorelines than the classical machine learning-based methods in the FR image, which had trouble with differentiating the shadows from water in the high relief areas along the river.

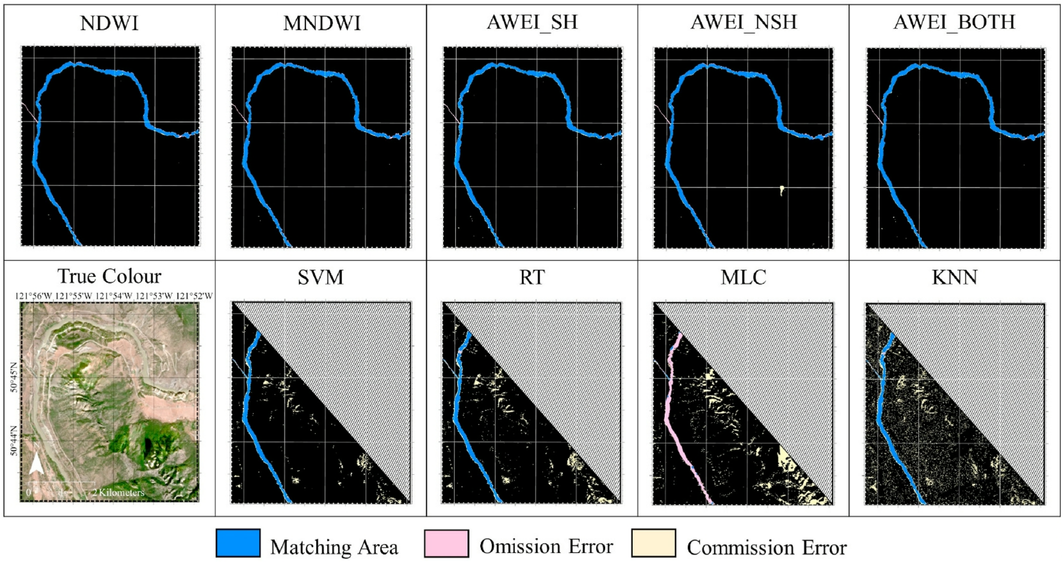

Fig. 9. Fraser River S2/MSI imagery classification results.

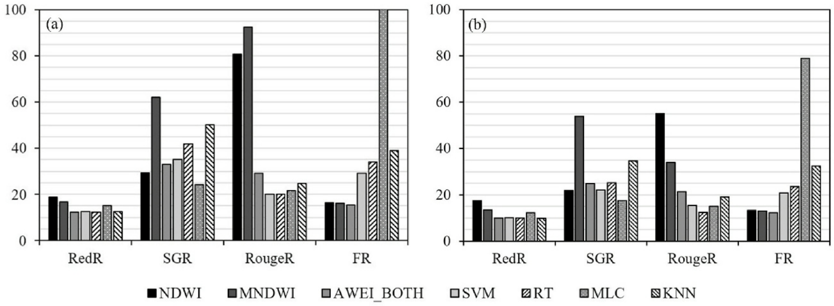

Fig. 10. Classification accuracy of multispectral imagery based on (a) RMSE of river width [m] and (b) MAE of river width [m].

It is shown that the purpose of the LWIC can influence the best performing method. The best performing method of extracting the most accurate waterbody area in an image can be different than the best performing method of extracting the waterbody boundary. This will be the case if one method is more successful at minimizing the CE by correctly classifying non-water objects as water. This is because the CSI metric represents all of the CE and OE in the entire image, whereas the RMSE of the river width represents the CE and OE at the shoreline. The second metric would not include shadows, snow/ice, or any other difficult-to-classify non-water objects outside of the river area. Thus, the RMSE of river width gives a better measure of which method extracts the shoreline most accurately.

5. Discussion

5.1. Impacts of method optimization

When comparing the threshold derived from Otsu's method (OT) and the threshold derived from manual optimization (MT), it was found that the MAE was smallest for MNDWI between the four images. This suggests that Otsu's method generally produces a close-to-optimized threshold for MNDWI-based LWIC. For AWEI_NSH and AWEI_SH, the difference between OT and MT was consistently negative (i.e., OT was lower than MT). Thus, if OT was used with AWEI_NSH or AWEI_SH in previous work, it is more likely that the threshold was not optimized. This may have skewed previous results to indicate that the AWEI indices performed more poorly than other indices, such as MNDWI, if thresholds were estimated using Otsu's method. Likewise, previous studies have illustrated that the success of LWIC methods is dependent on suitable threshold selection, and automated threshold selection algorithms vary in their success to select a suitable threshold (Karaman, 2021; Sekertekin, 2021). The analysis of manually optimized thresholds and automated threshold selection algorithms could be further refined by iteratively evaluating other threshold values within the range of tested values at smaller intervals (e.g., iteratively adjust and evaluate threshold values in increments of 0.05 or 0.01, as opposed to 0.1). Previous literature had found that resolution increase by fusing imagery from 10m to 5m improved classification accuracy (Manaf et al., 2016), but this is likely image and location dependent. The result of resampling the resolution of the 20m bands (SWIR1 or band 11 and SWIR2 or band 12) to 10m to match bands 2, 3, 4, and 8 was inconsistent between images. Adjustment of the training and validation ratio for the machine learning-based LWIC methods may improve the quality of the surface water maps produced if more training data is available. A sensitivity test was performed with SVM by varying the training/validation ratio from 50/50 to 75/25. Additionally, the images included in the training data were adjusted such that the RedR, SGR, and FR images were used to train the model and the model was validated on the RougeR image. The resulting surface water map produced a slightly lower CSI (93.6%–91.2%), and thus a slightly lower quality surface water map. The total percentage of MA increased, but both the OE and CE also increased. From this test, it cannot be understood if the change in the CSI is due to the change in the training/validation ratio or the change in the content of the training data. This example illustrates the need for further evaluation of training and validation ratios and improved understanding about the relationship between the quantity and content of training data and the resulting quality of surface water maps.

5.2. Influence of imagery characteristics on surface water maps

Because of the method used to quantify the error of each LWIC method (i.e., CSI and RMSE of widths instead of the kappa coefficient approach), the characteristics and objects in the images that caused difficulties for successful LWIC could be observed. A summary of these findings is illustrated in Table 5.

The effect of built-up being present in an image was inconsistent across the LWIC methods tested. Overall, the method that performed best at differentiating built-up from water across the images was AWEI_NSH, which was able to successfully classify the runway as land in the SGR image as shown in Fig. 11, the town in the SGR image also shown in Fig. 11, the highway in the RougeR image, and the town in the RedR image.

AWEI_SH could not differentiate the runway in the SGR image from water but did not struggle with the highway in RougeR or town in RedR. NDWI struggled with town buildings in the RedR image and the SGR image, and the highway in the RougeR image, but not the runway in the SGR image. The effect of town buildings in the RedR image could not be studied for the classical machine learning-based methods. In the SGR image, the MLC had the least amount of error caused by built up but, conversely, had the most amount of the highway in the RougeR image misclassified as water.

Shallow shorelines with vegetation caused issues for all methods, but could be minimized most with AWEI_SH, which performed well in the RougeR image with both the Ottawa River and Rivière-Rouge and the oxbow lakes in the RedR image, as shown in Fig. 12. The effect of shallow shores with vegetation was not consistent across the indices: NDWI and MNDWI produced increased OE while AWEI_NSH produced increased CE in areas where shallow, vegetated shores were present. MLC produced the least amount of error caused by shallow vegetated shores in the RedR image, while the other classical machine learning-based methods had increased OE caused by the vegetated shorelines. None of the classical machine learning-based methods had difficulty extracting the vegetated shoreline in the RougeR image, likely because the methods were adequately trained on areas that had shallow vegetation.

Wet sand was able to be differentiated from water with all spectral index-based methods, except for NDWI, which misclassified areas with wet sand as water, increasing the CE. Wet sand was not present in any other image than the RougeR image on the Ottawa River shoreline. Thus, the findings could not be compared across imagery to understand if NDWI consistently misclassifies areas with wet sand. Additionally, the wet sand area was not included in the validation area for the classical machine learning-based methods.

Table 5
Summary of objects and characteristics present in the four images and the recommended (✓) or not recommended (✗) methods to support accurate LWIC when these objects and characteristics are present.

Image Characteristic	NDWI	MNDWI	AWEI_SH	AWEI_NSH	SVM	MLC	RT	KNN
Built-up				✓				
Shallow Vegetated Shorelines			✓			✓		
Wet Sand	✗							
High Albedo Surfaces	✓					✓		
Mountain or Cliff Shadows					✗	✗	✗	✗
Shadows on Water				✓				
Narrow Waterbodies		✗	✓					

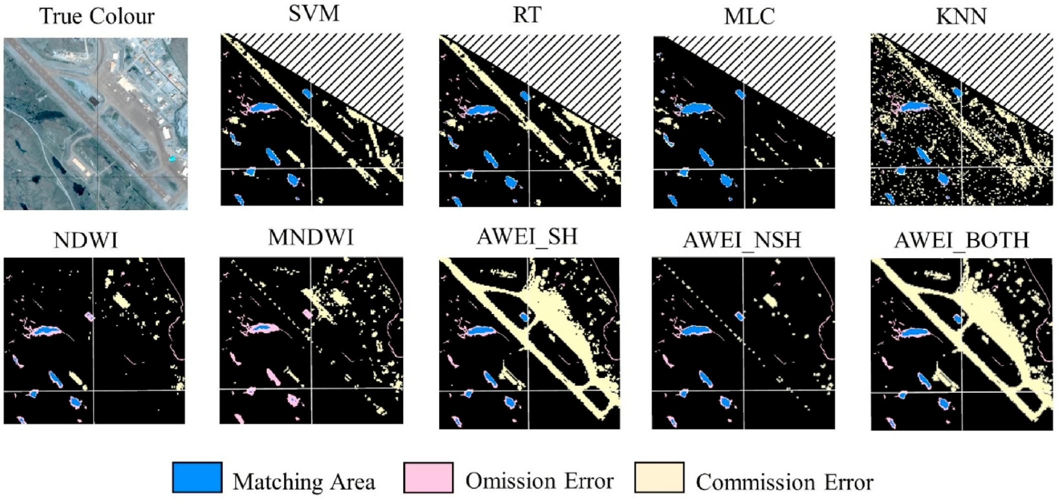

Fig. 11. Built-up classification results from the SGR image area.

High albedo surfaces, especially snow and ice shown in Fig. 13, caused issues for all LWIC methods tested. Snow was present in only the SGR image, so it could not be compared across sites for consistency. NDWI had the least amount of error caused by snow/ice surfaces in the SGR image compared to the other spectral index-based methods, but still was not able to completely differentiate snow and ice from open surface water. The MLC method was able to reduce the amount of error caused by snow and ice surfaces, which is consistent with the MLC results of other high albedo surfaces in the image (town buildings).

Dark shadow, caused by cliffsides, caused errors for all methods where the shadows were covering the water surface. In these instances, the water was misclassified as land. This finding indicates that shadows on water surfaces caused by dense clouds will cause misclassification by OE as well. AWEI_NSH produced the least amount of OE when cliff shadows were present on the waterbody, but it was also the only spectral index-based method that had a portion of mountain shadow that caused CE in the same FR image. No other images included significant shadowed areas, so this could not be compared across images. All classical machine learning-based methods produced some CE due to mountain and cliff shadows in the FR images as shown in Fig. 14, even though the methods were trained on the opposite side of the image that also included shadows, though fewer shadows were present on the training side of the image. The MLC method was not able to detect the waterbody in the FR image. This is likely because the MLC tool in ArcGIS Pro may only consider the class means of the "primary" band of imagery (i.e., NIR in this study). When a second band of imagery (i.e., green in this study) is added, the class means do not appear to be adjusted. Thus, if the second band of imagery changes the class means significantly, wide-spread misclassification occurs, as illustrated in Fig. 14.

Narrow waterbodies, particularly in the RougeR image illustrated in Fig. 15, are inherently difficult for LWIC of S2/MSI imagery due to pixel resolution constraints. AWEI_SH comparatively was able to detect more of the narrow locations of the Rivière-Rouge in the RougeR image than the other spectral index-based methods, though all methods created disjointed river water polygons at locations with river widths smaller than 40m.

This is also the case for the Sylvia Grinnell River in the SGR image, although NDWI also performed adequately to detect the narrow areas (sub-40m width) of the river. MNDWI performed poorly at extracting the narrow waterbodies in both the RougeR and SGR images, as illustrated by the RMSE and MAE. MLC produced disjointed waterbody polygons at narrow locations of rivers in the SGR image, as well as misclassification of rapids as non-water pixels, but the other classical machine learning-based methods did not have these issues. This was also the case for the RougeR image in which the MLC struggled the most to produce continuous water polygons at the narrow locations, but all classical machine learning-based methods did struggle with this in the RougeR image.

The different types of soil and vegetation across the four images did not seem to influence the success of the LWIC methods. The different types of water colour and turbidity also did not seem to influence the LWIC results across the images. When the thresholds were optimized for the spectral index-based methods, the methods were just as successful at LWIC for high turbidity waters (in the RedR image, averaging 303 mg/L) as they were for low turbidity waters (in the RougeR image, averaging 37 mg/L). Turbid waters differ from green and clear water with a higher spectral reflectance in the NIR band. Bare soil also has its highest spectral reflectance in the NIR band; however, the reflectance of bare soil is much higher than turbid water (Sun et al., 2012). Thus, turbid water is not difficult to differentiate between other image characteristics, such as bare soil, when the NIR band is included in the classification. Additionally, green water, or water containing chlorophyll, differs from clear water with its highest spectral reflectance in the NIR band. Green water and terrestrial vegetation have low red reflectance and high NIR reflectance, but green water has much lower reflectance of the NIR band compared to vegetation (Sun et al., 2012). Thus, green water is also not difficult to distinguish from other image characteristics, such as vegetation, if the NIR bands is included in the classification.

A limitation of this study consisted of the fact that temporal changes at each study location, and potential impacts on the quality of surface water maps produced, were not evaluated in this study. Future research focused on assessing LWIC methods under different

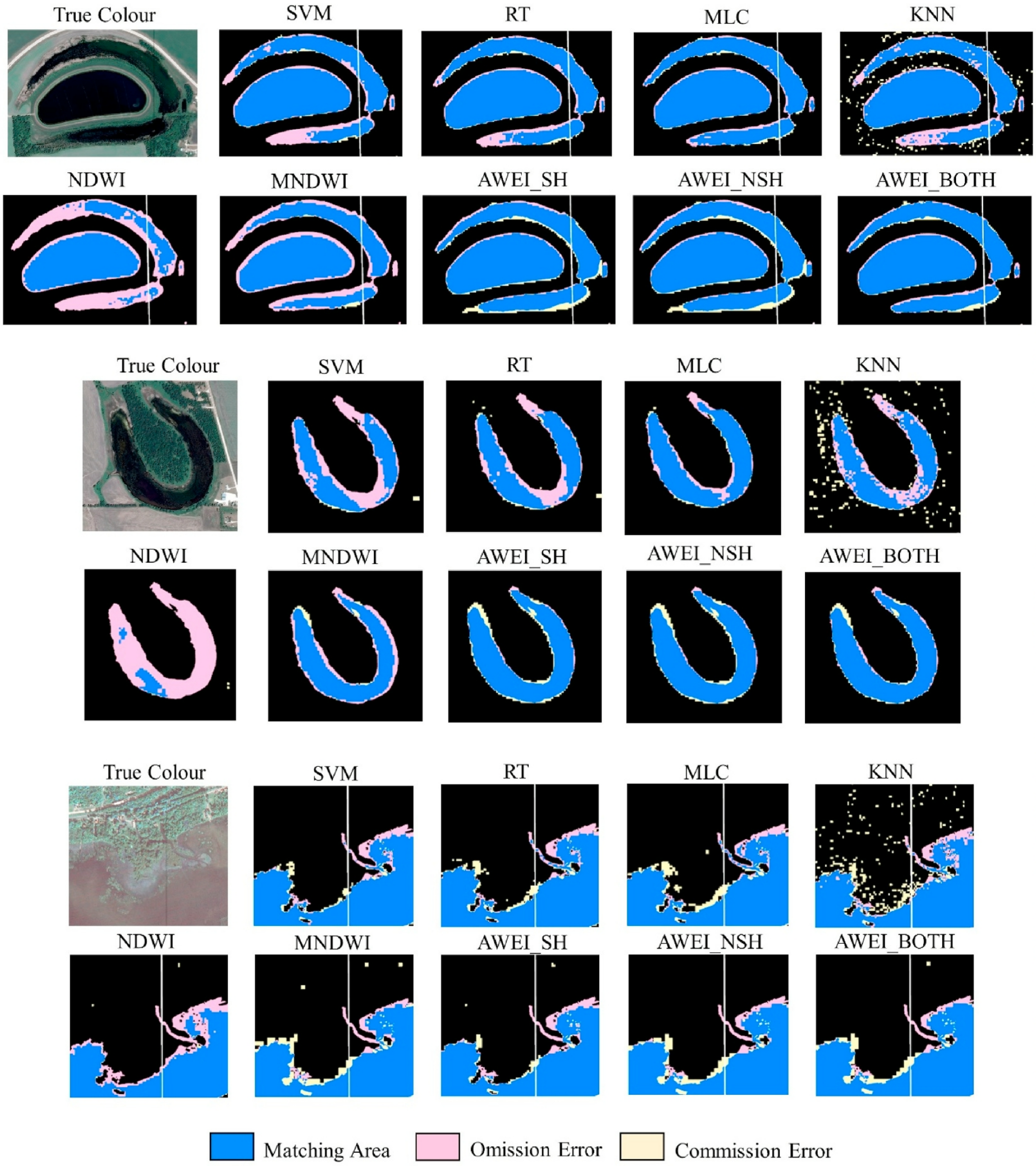

Fig. 12. Shallow vegetated shoreline classification results in the RedR (top and middle) and RougeR (bottom) image areas.

seasonal conditions would provide valuable context to further inform LWIC strategy.

5.3. *Comparison to other studies' findings*

To understand how this study's findings align with the current understanding of LWIC methods' success, the order of most to least successful LWIC method for this study and others is compared in Table 6. A general trend exists for the machine learning LWIC methods where SVM is the most successful with RT, or random forest (RF) or decision tree (DT), second and MLC third. This trend is also illustrated in our study findings when considering the mean CSI across the four image areas. Similar comparison-based studies did not evaluate and compare the KNN algorithm. When comparing spectral index methods to machine learning methods, our study found the spectral index methods to outperform the machine learning methods. However, it has been shown that an increased number of training samples may improve the output of machine learning models beyond that of spectral index methods, specifically NDWI (Ireland et al., 2015). However, at a lower number of training samples, NDWI was shown to outperform SVM (Ireland et al., 2015). This highlights the fact that machine learning techniques are only useful if a good amount of training data is available, and successful global application of one trained model to other imagery may vary.

The trend for most successful to least successful spectral index-based methods is less clear than it is for machine learning LWIC methods. NDWI out-performed MNDWI in three studies, and NDWI was outperformed by MNDWI in a different three studies. Fewer

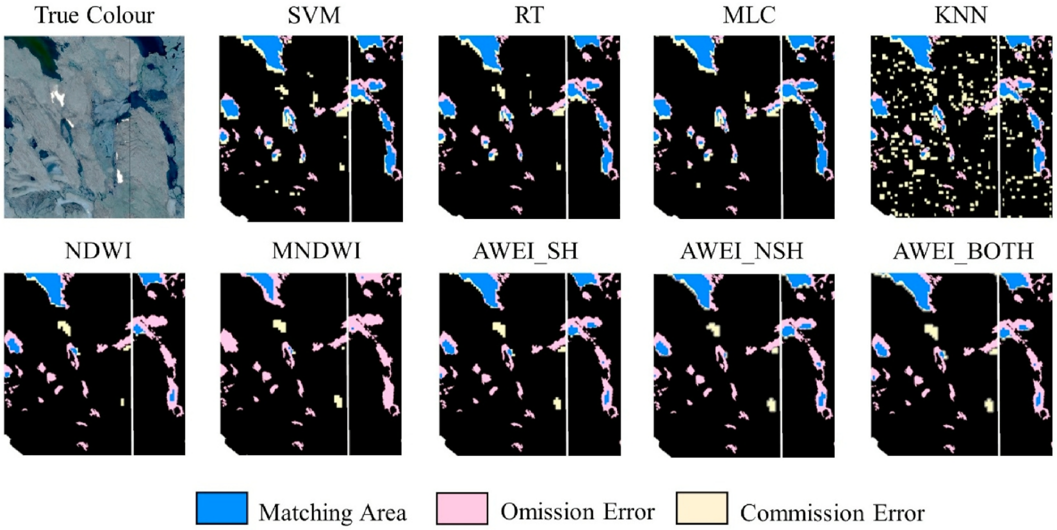

Fig. 13. Snow and ice classification results from the SGR image area.

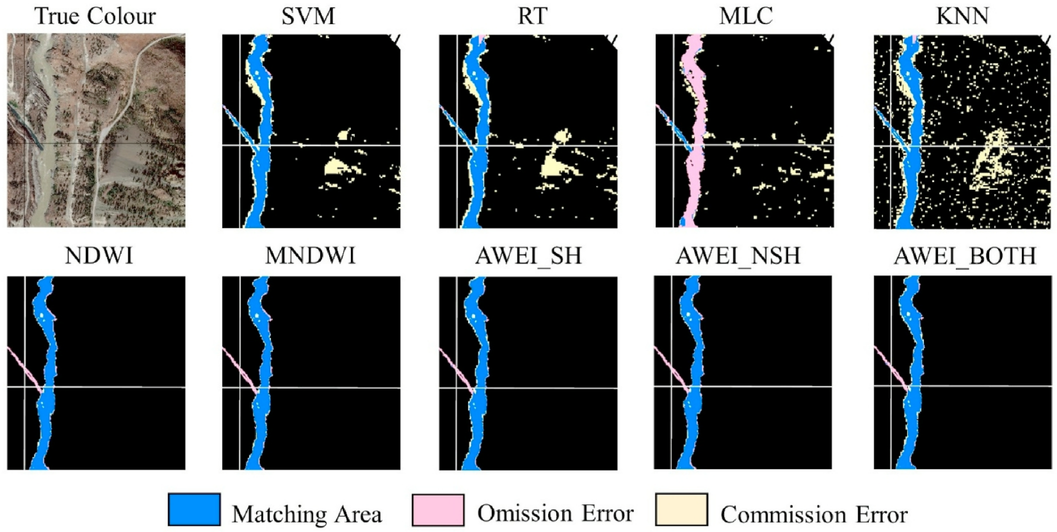

Fig. 14. Shadow classification results from the FR image area.

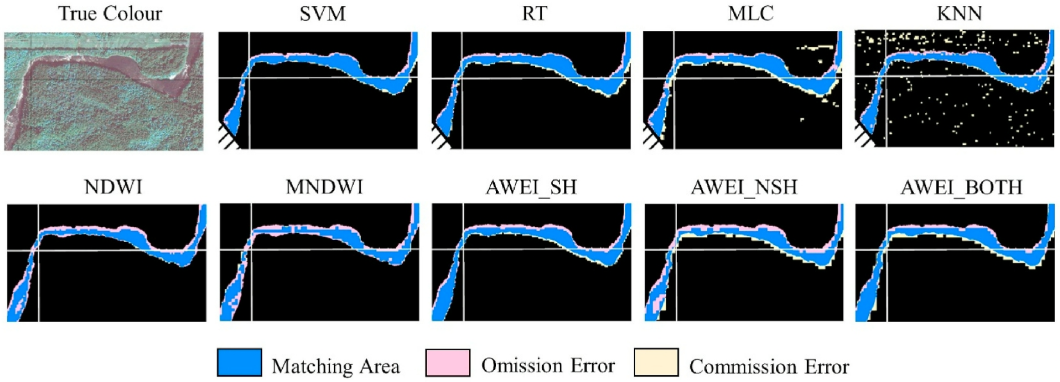

Fig. 15. Narrow waterbody classification results from the RougeR image area.

Table 6

Order of classification success for this study compared to other studies. All acronyms for other study methods are included in the appendix in Table A7. The methods in dark red are the methods explored in this study.

Study	Best	→	→	→	→	→	→	→	→	→	→	→	→	Worst
This study ** [1]	**AWEI_NSH**	AWEI_BOTH	**AWEI_SH**	**NDWI**	**SVM** NIR+Green	**MNDWI**	**RT** NIR+Green	**MLC** NIR+Green	KNN NIR+Green	-	-	-	-	-
Alcaras et al. (2022) [1]	**NDWI**	**MNDWI**	RGR	NDVI	EVI	NRR	-	-	-	-	-	-	-	-
Aroma and Raimond (2021) * [a] [2]	**NDWI**	FCM Band 5	K-mean Band 5	ARKFCM Band 5	K-means Band 6	ARKFCM Band 6	SLIC-SUPER	FCM Band 6	**MNDWI**	ARKFCM Band 3	FCM Band 3	K-means Band 3	-	-
Dona et al. (2016) [2]	GP Band 4	**(1)** Band 4 **(2)** Mahalanobis Bands 3,4 **(3) SVM** Cubic Bands 3,4	**(1)** ISODATA Bands 3,4 **(2)** ANN Bands 1-5,7 **(3) SVM** M Gauss Bands 3-5	**SVM** Linear Bands 3-5	**SVM** Course Bands 3-5	**(1) MLC** Bands 3,4 **(2) SVM** Linear Bands 1-5,7 **(3) SVM** Course Bands 1-5,7	**(1) MNDWI** **(2) NDWI** (McFeeters)	**MLC** Bands 4,5	**(1)** I_CEDEX **(2)** Band 5 **(3) SVM** Quadratic Bands 3-5 **(4) SVM** F Gauss Bands 3-5	Minimum Distance Bands 4,5	**NDWI** (Gao)	**(1)** K-means Bands 4,5 **(2)** K-means Band 4	**(1)** NDVI **(2)** SAVI	Parallelpiped All Combinations
Pena-Reguerio (2020) * [1]	**NDWI**	B_Blue	**AWEI_SH**	**MNDWI**	**AWEI_NSH**	CEDEX	RE-NDWI	-	-	-	-	-	-	-
Zhang et al. (2022) * [2]	**(1) NDWI** **(2)** CIWI	NIR Band	-	-	-	-	-	-	-	-	-	-	-	-
Hafiza Wajiha Khalid et al. (2021) * [a] [2]	LBWEI	**NDWI**	NWI	WRI	**MNDWI**	NDMI	**AWEI_SH**	-	-	-	-	-	-	-
Yilmaz et al. (2023) [a] [1] [2]	**RF** (Vis + NIR)	**RF** (Vis + NIR + **NDWI**)	**RF** (Vis + NIR + **MNDWI**)	**MLC** (Vis + NIR + **NDWI**)	**MLC** (Vis + NIR + **MNDWI**)	**MLC** (Vis + NIR)	K-means (Vis + NIR + **NDWI**)	K-means (Vis + NIR)	K-means (Vis + NIR + **MNDWI**)	-	-	-	-	-
Ireland et al. (2015) [2]	rkFDA RBF	**SVM** RBF	rkFDA Linear	**SVM** Linear	**NDWI**	-	-	-	-	-	-	-	-	-
Feyisa et al. (2014) ** [2]	**AWEI**	**MNDWI**	**MLC**	-	-	-	-	-	-	-	-	-	-	-
Senel et al. (2020) [1] [2]	**(1) AWEI_NSH**, Landsat **(2) AWEI_NSH**, Sentinel	**(1) MNDWI**, Sentinel **(2) MLC**, Sentinel	**NDWI**, Sentinel	**(1) MNDWI**, Landsat **(2) AWEI_SH**, Landsat **(3) AWEI_SH**, Sentinel **(4)** TCWI, Sentinel	**(1) NDWI**, Landsat **(2)** TCWI, Landsat	**MLC**, Landsat	-	-	-	-	-	-	-	-
Feizizadeh et al. (2023) [a] [2]	**SVM**	**RF**	CART	-	-	-	-	-	-	-	-	-	-	-
Li et al. (2022) * [2]	**SVM**	CART	Object-Oriented	Deep Learning	**NDWI**	-	-	-	-	-	-	-	-	-
Manaf et al. (2016) [3]	**SVM**	Neural Network	**MLC**	Mahalanobis Distance	Minimum Distance	Parallelpiped	-	-	-	-	-	-	-	-
Otukei and Blaschke (2010) [a] [2]	**DT**	**MLC** (All Bands)	**SVM** (Reduced Bands)	**SVM** (All Bands)	**MLC** (Reduced Bands)	-	-	-	-	-	-	-	-	-
Singh and Pandey (2021) [a] [2]	**RT**	**SVM**	**MLC**	-	-	-	-	-	-	-	-	-	-	-
Sun et al. (2012) [2]	New pixel-based method	**MLC** + NIR	**MNDWI**	New integrated method	**NDWI**	-	-	-	-	-	-	-	-	-
Thati et al. (2021) * [a] [2]	**MNDWI** + ISODATA	**MNDWI** + Otsu	**MNDWI** + K-means	**MNDWI** + Threshold segmenting	-	-	-	-	-	-	-	-	-	-

* Study completed with multiple image areas.

** Study completed with multiple image areas with a variety of landcover characteristics.

a Study completed in multiple seasonal conditions or different years.

1 Study completed with Sentinel-2.

2 Study completed with Landsat.

3 Study completed with SPOT.

studies compared AWEI_SH and AWEI_NSH than NDWI and MNDWI with only four studies examining either AWEI_SH or AWEI_NSH and 14 studies examining either NDWI or MNDWI. Additionally, a direct comparison to some of the studies that varied bands of imagery (Doña et al., 2016) and employed combinations of machine learning and spectral index methods (Yilmaz et al., 2023) is not possible. The range in the trend most to least successful spectral index methods is likely due to the range of image characteristics, imagery pre-processing, and thresholding techniques applied in each of the studies. For example, five studies in Table 6 employed Otsu's method, six studies adjusted the threshold manually (as performed in this study), two studies did not mention the threshold procedure used, one study used default thresholds (e.g., 0 for NDWI and MNDWI), and two studies used a variety of machine learning algorithms (e.g., RF, MLC, K-means, etc.). As shown in our analysis comparing the Otsu derived threshold to the manually adjusted threshold, and in the works by Sekertekin (2021) and Karaman (2021), the choice of threshold selection algorithm not only can determine the success of the method, but the combination of certain spectral indices with certain thresholding algorithms are more likely to produce accurate surface water maps than others. For example, in this study Otsu's method was more likely to produce an optimized threshold for MNDWI than it was for AWEI_SH or AWEI_NSH.

It is also important to note that this study's order of most to least successful LWIC methods represents an average over the four image areas. For example, although AWEI_NSH was the most successful overall, it was not the highest performing method in any of the individual study images. This suggests that a single image area cannot determine which LWIC is most successful overall, and rather, the success of each LWIC method depends on the imagery characteristics that each LWIC method is more likely to properly or improperly classify, as illustrated in Table 5.

6. Conclusions and recommendations

Five methods of spectral index-based LWIC and four methods of classical machine learning-based LWIC on S2/MSI imagery were compared across four diverse locations in Canada. The methods and imagery were chosen because of their regular use in remote sensing applications, especially in surface water detection. The purpose of this comparison was to provide recommendations of the best performing method of LWIC for future surface water detection research. The following list outlines the main conclusions of this work.

- The spectral index **AWEI_NSH** was the overall most successful method of LWIC across the four study locations. **SVM** was the most successful machine learning method of LWIC across the four study locations, though it was outperformed by most of the spectral indices.
- Threshold selection is important. The threshold optimization exercise performed in this study showed that optimizing thresholds is important before comparing the results of LWIC methods against each other.
- The most successful LWIC method was not consistent across the sites and the order of most to least successful LWIC methods across the sites was also not consistent. This illustrates that each method's success is dependent on the characteristics of the image area.
- It was not consistent across the four study locations whether spectral index-based methods or classical machine learning-based methods produced the highest accuracy surface water maps.
- Certain image characteristics (i.e., built-up, shallow vegetated shorelines, wet sand, high albedo surfaces, shadowy areas, and narrow waterbodies) were more likely to cause classification errors. For correctly classifying **built-up** as land pixels, AWEI_NSH produced the most successful LWIC. For **shallow vegetated shorelines**, AWEI_SH or MLC were both successful at classifying the vegetated water as water pixels. For images including **wet sand**, NDWI is not recommended, but this was not tested across multiple images, so this conclusion is not as strong comparatively. For **high albedo surfaces,** such as **snow and ice**, NDWI and MLC were able to minimize the amount of snow and ice classified as liquid surface water but not completely. For **shadowy images**, spectral index-based methods overall perform successfully and better than classical machine learning-based methods. For **narrow waterbodies**, AWEI_SH was able to correctly detect more of the shoreline than the other methods.
- The most optimal method of detecting surface water overall in the image may not be the most successful at detecting the most representative waterbody boundary. Thus, it is important to consider what the purpose of the LWIC is when selecting an optimized approach.

The following is the future recommended work based on the limitations and findings of this study.

- Analysis of the four locations in different seasonal conditions was not performed; thus, in future work it may be interesting to understand if the best performing method can change in different images at the same location.
- Sensitivity testing on the training/validation ratio of the classical machine learning methods was performed only for one additional case beyond the 50/50 spatial split applied in this study. This ratio is an important consideration when performing surface water mapping, as the results can vary based on the training data available. Thus, the training/validation ratio of classical machine learning methods should be tested and considered in future works.

Funding statement

This work was supported by the Natural Sciences and Engineering Research Council of Canada [Canada Graduate Scholarship –Doctoral (CGM-D) scholarship to Kirby and Discovery Grants to Rennie and Nistor], by the Office of Energy Research and Development (OERD) of Natural Resources Canada, and the National Research Council of Canada (NRC) Collaborative Science, Technology and Innovation Program (CSTIP).

Ethical statement for *remote sensing applications: Society and environment*

-This material is the authors' own original work, which has not been previously published elsewhere.
-The paper is not currently being considered for publication elsewhere.
-The paper reflects the authors' own research and analysis in a truthful and complete manner.
-The paper properly credits the meaningful contributions of co-authors and co-researchers.
-The results are appropriately placed in the context of prior and existing research.
-All sources used are properly disclosed (correct citation).
-All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

CRediT authorship contribution statement

Katelyn Kirby: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sean Ferguson:** Writing – review & editing, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Colin D. Rennie:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Julien Cousineau:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition. **Ioan Nistor:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Katelyn Kirby reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Colin Rennie reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Ioan Nistor reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Colin Rennie reports financial support was provided by Office of Energy Research and Development (OERD) of Natural Resources Canada. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rsase.2024.101367.

References

Agriculture and Agri-Food Canada; Science and Technology Branch; Government of Canada, 2016. Gridded soil Landscapes of Canada - open government portal [WWW Document]. Sentin.-2 L2A Data. URL. https://open.canada.ca/data/en/dataset/cb29b370-3639-4645-9ef9-b1ef131837b7, 11.30.22.

Alcaras, E., Amoroso, P.P., Figliomeni, F.G., Parente, C., Prezioso, G., 2022. Accuracy evaluation of coastline extraction methods in remote sensing: a smart procedure for SENTINEL-2 images. Presented at the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 13–19. https://doi.org/10.5194/isprs-archives-XLVIII-4-W3-2022-13-2022.

Aroma, R.J., Raimond, K., 2021. Investigation on spectral indices and soft classifiers-based water body segmentation approaches for satellite image analysis. J. Indian Soc. Remote Sens. 49, 341–356. https://doi.org/10.1007/s12524-020-01194-5.

Ball, G.H., Hall, D.J., 1965. ISODATA, A NOVEL METHOD OF DATA ANALYSIS AND PATTERN CLASSIFICATION.

Bijeesh, T.V., Narasimhamurthy, K.N., 2020. Surface water detection and delineation using remote sensing images: a review of methods and algorithms. Sustain. Water Resour. Manag. 6. https://doi.org/10.1007/s40899-020-00425-4.

Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. Remote Sens. Environ. 240, 111685. https://doi.org/10.1016/j.rse.2020.111685.

Boschetti, M., Nutini, F., Manfron, G., Brivio, P., Nelson, A., 2014. Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems. PLoS One 9, e88741. https://doi.org/10.1371/journal.pone.0088741.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Canada Centre for Remote Sensing; Natural Resources Canada; Government of Canada, 2015. 2015 land cover of Canada - open government portal [WWW Document]. URL. https://open.canada.ca/data/en/dataset/4e615eae-b90c-420b-adee-2ca35896caf6, 11.30.22.

Chen, Jun, Chen, Jin, Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2015. Global land cover mapping at 30m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27. https://doi.org/10.1016/j.isprsjprs.2014.09.002.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964.

Department of the Interior, U.S. Geological Survey, 2021. Landsat 4-7 Collection 2 (C2) Level 2 Science Product (L2SP) Guide (No. LSDS-1618 Version 4.0).

Doña, C., Chang, N.-B., Caselles, V., Sánchez, J.M., Pérez-Planells, L., Bisquert, M.M., García-Santos, V., Imen, S., Camacho, A., 2016. Monitoring hydrological patterns of temporary lakes using remote sensing and machine learning models: case study of La Mancha Húmeda Biosphere Reserve in Central Spain. Remote Sens 8. https://doi.org/10.3390/rs8080618.

E. D. Chaves, M., C. A. Picoli, M., D. Sanches, I., 2020. Recent applications of landsat 8/OLI and sentinel-2/MSI for land use and land cover mapping: a systematic review. Remote Sens 12, 3062. https://doi.org/10.3390/rs12183062.

ESA, 2023. Sentinel-2 L2A [WWW Document]. Sentin.-2 L2A Data. URL. https://docs.sentinel-hub.com/api/latest/data/sentinel-2-l2a/, 3.28.23.

Feizizadeh, B., Omarzadeh, D., Kazemi Garajeh, M., Lakes, T., Blaschke, T., 2023. Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine. J. Environ. Plan. Manag. 66, 665–697. https://doi.org/10.1080/09640568.2021.2001317.

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: a new technique for surface water mapping using Landsat imagery. Remote Sens. Environ. 140, 23–35. https://doi.org/10.1016/j.rse.2013.08.029.

Government of Canada, 2023. Historical hydrometric data - water level and flow - environment Canada [WWW Document]. Sentin.-2 L2A Data. URL. https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html, 1.8.24.

Government of Canada, 2024. Sediment data search - water level and flow - environment Canada [WWW Document]. URL. https://wateroffice.ec.gc.ca/search/sediment_e.html, 9.17.24.

Goward, S.N., Markham, B., Dye, D.G., Dulaney, W., Yang, J., 1991. Normalized difference vegetation index measurements from the advanced very high resolution radiometer. Remote Sens. Environ. 35, 257–277. https://doi.org/10.1016/0034-4257(91)90017-Z.

Gudelj, M., Gasparovic, M., Zrinjski, M., 2018. Accuracy analysis of the inland waters detection. https://doi.org/10.5593/sgem2018V/1.5/S02.025.

Hartigan, J.A., Wong, M.A., 1979. Algorithm as 136: a K-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830.

Huang, C., Zhang, C., He, Y., Liu, Q., Li, H., Su, F., Liu, G., Bridhikitti, A., 2020. Land cover mapping in cloud-prone tropical areas using sentinel-2 data: integrating spectral features with ndvi temporal dynamics. Remote Sens. Basel Switz 12, 1163. https://doi.org/10.3390/rs12071163.

Ireland, G., Volpi, M., Petropoulos, G.P., 2015. Examining the capability of supervised machine learning classifiers in extracting flooded areas from landsat tm imagery: a case study from a mediterranean flood. Remote Sens 7, 3372–3399. https://doi.org/10.3390/rs70303372.

Karaman, M., 2021. Comparison of thresholding methods for shoreline extraction from sentinel-2 and landsat-8 imagery: extreme lake salda, track of mars on Earth. J. Environ. Manage. 298, 113481. https://doi.org/10.1016/j.jenvman.2021.113481.

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. 177, 89–100. https://doi.org/10.1016/j.rse.2016.02.028.

Kirby, K., Rennie, C.D., Cousineau, J., Ferguson, S., Nistor, I., 2023. Impacts of seasonal flow variation on riverine hydrokinetic energy resources and optimal turbine location – case study on the Rivière Rouge, Québec, Canada. Renew. Energy. https://doi.org/10.1016/j.renene.2023.04.067.

Kseňak, Ľ., Pukanská, K., Bartoš, K., Blišťan, P., 2022. Assessment of the usability of SAR and optical satellite data for monitoring spatio-temporal changes in surface water: bodrog river case study. Water Switz 14. https://doi.org/10.3390/w14030299.

Li, J., Ma, R., Cao, Z., Xue, K., Xiong, J., Hu, M., Feng, X., 2022. Satellite detection of surface water extent: a review of methodology. Water 14, 1148. https://doi.org/10.3390/w14071148.

Lim, N.J., Brandt, S.A., 2019. Are feature agreement statistics alone sufficient to validate modelled flood extent quality? A study on three Swedish rivers using different digital elevation model resolutions. Math. Probl Eng. 2019, e9816098. https://doi.org/10.1155/2019/9816098.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for sentinel-2, in: image and signal processing for remote sensing XXIII. Presented at the Image and Signal Processing for Remote Sensing XXIII, pp. 37–48. https://doi.org/10.1117/12.2278218. SPIE.

Manaf, S.A., Mustapha, N., Sulaiman, M.N., Husin, N.A., Hamid, M.R.A., 2016. Comparison of classification techniques on fused optical and sar images for shoreline extraction: a case study at northeast coast of peninsular Malaysia. J. Comput. Sci. 12, 399–411. https://doi.org/10.3844/jcssp.2016.399.411.

Mansaray, L.R., Wang, F., Huang, J., Yang, L., Kanu, A.S., 2020. Accuracies of support vector machine and random forest in rice mapping with Sentinel-1A, Landsat-8 and Sentinel-2A datasets. Geocarto Int. 35, 1088–1108. https://doi.org/10.1080/10106049.2019.1568586.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432. https://doi.org/10.1080/01431169608948714.

Mukherjee, N., Samuel, C., 2016. Assessment of the temporal variations of surface water bodies in and around Chennai using landsat imagery. Indian J. Sci. Technol. 9. https://doi.org/10.17485/ijst/2016/v9i18/92089.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66. https://doi.org/10.1109/TSMC.1979.4310076.

Otukei, J.R., Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. Int. J. Appl. Earth Obs. Geoinformation, Supplement Issue on "Remote Sensing for Africa – A Special Collection from the African Association for Remote Sensing of the Environment (AARSE) 12, S27–S31. https://doi.org/10.1016/j.jag.2009.11.002.

Paola, J.D., Schowengerdt, R.A., 1995. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. IEEE Trans. Geosci. Remote Sens. 33, 981–996. https://doi.org/10.1109/36.406684.

Pena-Regueiro, J., Sebastiá-Frasquet, M.-T., Estornell, J., Aguilar-Maldonado, J.A., 2020. Sentinel-2 application to the surface characterization of small water bodies in wetlands. Water 12, 1487. https://doi.org/10.3390/w12051487.

Roach, J.K., Griffith, B., Verbyla, D., 2012. Comparison of three methods for long-term monitoring of boreal lake area using Landsat TM and ETM+ imagery. Can. J. Remote Sens. 38, 427–440. https://doi.org/10.5589/m12-035.

Schalkoff, R.J., 1997. Artificial Neural Networks. McGraw Hill. https://doi.org/10.5555/541158.

Sekertekin, A., 2021. A Survey on global thresholding methods for mapping open water body using sentinel-2 satellite imagery and normalized difference water index. Arch. Comput. Methods Eng. 28, 1335–1347. https://doi.org/10.1007/s11831-020-09416-2.

Senel, G., Dogru, A., Göksel, Ç., 2020. Exploring the potential of landsat-8 OLI and sentinel-2 MSI data for mapping and monitoring enez dalyan lagoon. Desalination Water Treat. 177, 330–337. https://doi.org/10.5004/dwt.2020.24802.

Shen, L., Li, C., 2010. Water body extraction from Landsat ETM+ imagery using adaboost algorithm. In: 2010 18th International Conference on Geoinformatics. Presented at the 2010 18th International Conference on Geoinformatics, pp. 1–4. https://doi.org/10.1109/GEOINFORMATICS.2010.5567762.

Singh, G., Pandey, A., 2021. Evaluation of classification algorithms for land use land cover mapping in the snow-fed Alaknanda River Basin of the Northwest Himalayan Region. Appl. Geomat. 13, 863–875. https://doi.org/10.1007/s12518-021-00401-3.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Sreekumar, U., Rennie, C., Mohammadian, A., Nistor, I., Lovitt, J., Zhang, Y., 2022. Modelling the transport of tailings after Mount Polley tailings dam failure using multisource geospatial data. Can. J. Civ. Eng. 49, 1870–1884. https://doi.org/10.1139/cjce-2021-0596.

Sun, F., Sun, W., Chen, J., Gong, P., 2012. Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery. Int. J. Remote Sens. 33, 6854–6875. https://doi.org/10.1080/01431161.2012.692829.

Thati, J., Ari, S., Agrawal, K., 2021. Extraction of glacial lakes in the Himalayan region using landsat imagery. In: New Paradigms in Computational Modeling and its Applications, pp. 241–254. https://doi.org/10.1016/B978-0-12-822133-4.00004-9.

Venditti, J.G., Rennie, C.D., Bomhof, J., Bradley, R.W., Little, M., Church, M., 2014. Flow in bedrock canyons. Nature 513, 534–543.

Wajiha Khalid, Hafiza, Rao, M., Khalil, Zahid, Qureshi, Muhammad Ateeq, 2021. Evaluating spectral indices for water bodies extraction in western Tibetan Plateau. Egypt. J. Remote Sens. Space Sci. Online 24, 619–634. https://doi.org/10.1016/j.ejrs.2021.09.003.

Wei, X., Xu, W., Bao, K., Hou, W., Su, J., Li, H., Miao, Z., 2020. A water body extraction methods comparison based on fengyun satellite data: a case study of poyang lake region, China. Remote Sens 12, 1–21. https://doi.org/10.3390/rs12233875.

Xu, H., 2006. Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033. https://doi.org/10.1080/01431160600589179.

Yilmaz, O.S., Gulgen, F., Balik Sanli, F., Ates, A.M., 2023. The performance analysis of different water indices and algorithms using sentinel-2 and landsat-8 images in determining water surface: demirkopru dam case study. Arab. J. Sci. Eng. https://doi.org/10.1007/s13369-022-07583-x.

Zhang, Q., Xiao, H., Wu, D., Wang, P., Guo, F., Tao, C., 2022. Comparative Study of Water-Body Extraction Methods Based on Landsat8 Remote Sensing Images 12506, 125063K. https://doi.org/10.1117/12.2662521.