Evaluation of Machine Learning Algorithms for Lake Ice Classification from Optical Remote Sensing Data

by

Yuhao Wu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2020

AUTHOR'S DECLARATION

Abstract

The topic of lake ice cover mapping from satellite remote sensing data has gained interest in recent years since it allows the extent of lake ice and the dynamics of ice phenology over large areas to be monitored. Mapping lake ice extent can record the loss of the perennial ice cover for lakes located in the High Arctic. Moreover, ice phenology dates, retrieved from lake ice maps, are useful for assessing long-term trends and variability in climate, particularly due to their sensitivity to changes in near-surface air temperature. However, existing knowledge-driven (threshold-based) retrieval algorithms for lake ice-water classification that use top-of-the-atmosphere (TOA) reflectance products do not perform well under the condition of large solar zenith angles, resulting in low TOA reflectance. Machine learning (ML) techniques have received considerable attention in the remote sensing field for the past several decades, but they have not yet been applied in lake ice classification from optical remote sensing imagery. Therefore, this research has evaluated the capability of ML classifiers to enhance lake ice mapping using multispectral optical remote sensing data (MODIS L1B (TOA) product).

Chapter 3, the main manuscript of this thesis, presents an investigation of four ML classifiers (i.e. multinomial logistic regression, MLR; support vector machine, SVM; random forest, RF; gradient boosting trees, GBT) in lake ice classification. Results are reported using 17 lakes located in the Northern Hemisphere, which represent different characteristics regarding area, altitude, freezing frequency, and ice cover duration. According to the overall accuracy assessment using a random k-fold cross-validation (k = 100), all ML classifiers were able to produce classification accuracies above 94%, and RF and GBT provided above 98% classification accuracies. Moreover, the RF and GBT algorithms provided a more visually accurate depiction of lake ice cover under challenging conditions (i.e., high solar zenith angles, black ice, and thin cloud cover). The two tree-based classifiers were found to provide the most robust spatial transferability over the 17 lakes and performed consistently well across three ice seasons, better than the other classifiers. Moreover, RF was insensitive to the choice of the hyperparameters compared to the other three classifiers. The results demonstrate that RF and GBT provide a great potential to map accurately lake ice cover globally over a long time-series.

Additionally, a case study applying a convolution neural network (CNN) model for ice classification in Great Slave Lake, Canada is presented in Appendix A. Eighteen images acquired during the the ice season of 2009-2010 were used in this study. The proposed CNN produced a 98.03% accuracy with the testing dataset; however, the accuracy dropped to 90.13% using an independent (out-of-sample) validation dataset. Results show the powerful learning performance of the proposed CNN with the testing data accuracy obtained. At the same time, the accuracy reduction of the validation dataset indicates the overfitting behavior of the proposed model. A follow-up investigation would be needed to improve its performance.

This thesis investigated the capability of ML algorithms (both pixel-based and spatial-based) in lake ice classification from the MODIS L1B product. Overall, ML techniques showed promising performances for lake ice cover mapping from the optical remote sensing data. The tree-based classifiers (pixel-based) exhibited the potential to produce accurate lake ice classification at a large-scale over long time-series. In addition, more work would be of benefit for improving the application of CNN in lake ice cover mapping from optical remote sensing imagery.

Acknowledgments

I wish to express my deepest appreciation to my thesis supervisor Dr. Claude Duguay for his patient and expert guidance over the past three years (from undergraduate to Master's). I am extremely grateful for all his contributions of time, immense ideas, and continued encouragement to support my M.Sc. research as well as the numerous opportunities that he gave me ranging from conferences, workshops, and involvement in a number of projects. I would also like to thank Dr. Linlin Xu for his valuable insights, suggestions and discussions. I would like to acknowledge Dr. Grant Gunn for his participation on my thesis committee.

I am also indebted to all group colleagues at "Team Duguay" for stimulating discussions. Meanwhile, I mourn the loss of our colleague and friend, Marzieh Foroutan. Additionally, a special thanks to all my friends I have met during my time at the University of Waterloo. In particular, the CNU squad, including Junqian Wang, Liuyi Guo, and Yue Zhao, has actively supported me as a family since my first day at Waterloo. My gratitude goes to all those friends outside of academics during this journey as well. A special gratitude also goes to Mitch Wang for his guide, support, and encouragement. I must thank Jane Russwurm, a writing specialist, who has expertly and patiently taught me a variety of writing knowledge and skills.

Last but not the least, I would like to pay gratitude to my parents for their selfless love, endless encouragement, and sincere faith. Without them, none of this could happen.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMSR	Advanced Microwave Scanning Radiometer
AO	Arctic Oscillation
AOI	Areas of interest
BPNN	Back-propagation neural network
CLIMo	Canadian Lake Ice Model
CNN	Convolutional neural network
Conv	Convolutional layer
CV	Cross-validation
DA	Discriminant Analysis
DT	Decision tree
ECV	Essential climate variable
ENSO	El Ninõ/Southern Oscillation
GBT	Gradient boosting trees
GEOS	Geostationary Operational Environmental Satellite
GLCM	Gray Level Co-occurrence Matrix
HDF	Hierarchical Data Format
IRGS	Iterative region growing using semantics
KNN	K-Nearest Neighbors
lr	Learning rate
MA	Mean accuracy
MAE	Mean absolute error
ML	Machine Learning
MLR	Multinomial logistic regression
MODIS	Moderate Resolution Imaging Spectroradiometer
mtry	The number of variables available to a split
NAO	North Atlantic Oscillation
NDSI	Normalized difference snow index

NIR	Near infrared
NP	North Pacific
ntree	The number of trees
OA	Overall accuracy
PBVI	Permutation-based variable importance
PCA	Principal Component Analysis
PDO	Pacific Decadal Oscillation
PNA	Pacific North American
Pool	Max-pooling layer
ReLU	Rectified linear unit
RF	Random forest
RMSE	Root mean square error
SD	Standard deviation
SSM/I	Special Sensor Microwave/Imager
SVM	Support vector machine
SWIR	Shortwave infrared
SZA	Solar zenith angle
TIR	Thermal infrared
TOA	Top-of-the-atmosphere
TP	Tibetan Plateau
VIIRS	Visible Infrared Imaging Radiometer Suite
VIS	Visible

Chapter 1

General Introduction

1.1 Motivation

Lakes occupy approximately 2% of the Earth's landscape (Brown and Duguay, 2010), and a total of about 3.3% of the land surface above latitude 58°N is seasonally ice covered (Duguay et al., 2015). Hence, lake ice is a major component of the cryosphere due to its large areal coverage in the high latitude regions. The extent and duration of lake ice cover have wide-ranging socio-economic impacts such as navigation, winter transportation, resource development, and distribution of drinking water (Benson et al., 2012; Brown and Duguay, 2010). In addition, lakes provide habitat for several floral and faunal species. The presence of lake ice has a significant effect on the composition and abundance of aquatic species (Livingstone, 1997). Lake ice also affects water-column oxygen concentration and water temperature by limiting heat and gas exchanges with the atmosphere. A reduction in the length of ice cover seasons may facilitate greater emissions of microbial methane (Greene et al., 2014), which could further accelerate climate warming due to its role as a potent greenhouse gas. The spread of human-made pollutants (e.g. perfluorinated chemicals) is also influenced by the presence and absence of lake ice (Veillette et al., 2012; Wrona et al., 2016).

Several European studies have revealed strong relationships between lake ice phenology and large-scale teleconnections, especially with atmospheric oscillation patterns such as the North Atlantic Oscillation (NAO) (Blenckner et al., 2004; George, 2007; Karetnikov and Naumenko, 2008; Korhonen, 2006). In Canada, Bonsal et al. (2006) show strongest links between the Pacific-related indices (El Ninõ/Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), the Pacific North American (PNA) pattern, and the North Pacific (NP) index) and ice dates over western Canada, particularly break-up dates. The impact of the NAO and the Arctic Oscillation (AO) is found to be generally less coherent over regions of Canada (Bonsal et al., 2006). Thus, variability and trends in lake ice during freeze-up and break-up can be useful indicators of climate change and variability. Additionally, the interactions of energy between atmosphere-water-ice occur during lake ice formation, growth, and decay. The processes of energy transition can significantly affect the magnitude and timing of evaporation and precipitation rates in lake-rich and surrounding regions. Therefore, accurate estimation of lake ice cover is important for improving numerical weather forecasting in regions occupied by lakes (Brown and Duguay, 2010). Overall then, lake ice observations are useful for many biological, ecological and socio-economic purposes.

In practice, some larger lakes do not form a complete ice cover (e.g. the Laurentian Great Lakes). On the other hand, some lakes located in the Arctic do not completely melt their ice cover in some years (i.e. perennially ice-covered lakes) (Latifovic and Pouliot, 2007). A more recent study, however, suggests that these lakes may be transitioning from perennially ice-covered to seasonally ice-covered such as Lake Hazen on Ellesmere Island, Canada (Surdu et al., 2016). Hence, mapping ice cover extent/area is important for climate monitoring at high latitudes, and this can be best achieved using satellite remote sensing data.

1.2 Significance of proposed research

With the surface-based lake ice network having decreased dramatically over the last three decades (Duguay et al., 2006), the use of remote sensing has become the most logical means to establish a large-scale observational network of lake ice. Optical remote sensing products provide, at present, extensive multispectral data available for lake ice cover mapping. Moderate Resolution Imaging Spectroradiometer (MODIS) products from NASA’s Terra (2000-present) and Aqua (2002-present) satellite platforms have gained popularity for mapping lake ice cover and determining ice phenology dates (freeze-up, break-up, and ice duration) because they can provide a near twenty-year record of Earth observations at a daily temporal resolution.

Various knowledge-driven (threshold-based) methods have been developed and applied on MODIS products to retrieve lake ice and examine changes in lake ice phenology. For example, studies by Gou et al. (2017), Qi et al. (2019), and Šmejkalová et al. (2016), applied threshold-based approaches to detect lake ice and monitor ice phenology events using MODIS radiance or reflectance imagery. Besides the radiance and reflectance products, numerous studies employed MODIS snow products, produced using the normalized difference snow index (NDSI), to determine lake ice phenology (Brown and Duguay, 2012; Cai et al., 2019; Kropáček et al., 2013; Murfitt and Brown, 2017). However, the knowledge-driven algorithms may not provide adequate classification results under complex conditions. Specifically, lakes located in high-latitude regions display lower top-of-the-atmosphere (TOA) reflectance in the visible-infrared spectral range during the ice freeze-up period, due to low solar illumination (large solar zenith angle). Thus, existing threshold-based retrieval algorithms for lake ice-water classification using optical remote sensing data do not perform well under such conditions.

Machine learning (ML) approaches have been applied in many studies of ice retrieval from remote sensing imagery (Han et al., 2016; Leigh et al., 2014; Su et al., 2015). The majority of these studies have employed microwave satellite data for sea ice classification. Hence, it is imperative to understand the performance of ML to lake ice classification from optical and microwave remote sensing imagery. Three aspects need to be carefully considered when applying ML for remote sensing classification. First, variable selection and importance measurement allow further understanding of the underlying classification processes by classifiers and improve their performance. Second, hyperparameter selection must be conducted to exploit the full capability of a classifier instead of evaluating simplex classification derived with only one set of hyperparameters. Finally, since remote sensing measurements over lake ice can vary in time (e.g. daily and seasonally between freeze-up and break-up) and in space (e.g. from one lake or lake region to another), the temporal and spatial transferability of classifiers should also be examined.

1.3 Research objectives

The overall objective of this research is to evaluate the performance of machine learning (ML) classifiers for lake ice classification from optical remote sensing data. This thesis assesses different ML approaches for lake ice cover mapping using MODIS data, thereby proposing an optimal classifier for lake ice cover mapping. Specifically, three sub-objectives are to: 1) determine the optimal combination of input variables, including variable importance, to obtain robust classification results; 2) perform hyperparameter selection and examine the sensitivity of the change in the hyperparameters; and 3) investigate the suitability of ML classifiers for lake ice cover mapping across different regions of the globe and over a relatively long time period (ca. 20 years).

1.4 Thesis structure

This thesis has been written following the manuscript format where a paper is included as an individual chapter. Chapter 1 provides the motivation of this research and addresses the research objectives. Chapter 2 covers background knowledge regarding lake ice, ice classification of remote sensing, and remote sensing classification using ML. Chapter 3 contains the paper and is entitled “Mapping Lake Ice Cover from MODIS Using Machine Learning Approaches”. A short paper titled “Lake Ice Classification from MODIS TOA Reflectance Imagery Using A Convolutional Neural Network: A Case Study of Great Slave Lake, Canada” is included in Appendix A. It has been submitted for publication in *Proceedings of the 2020 IEEE International Geoscience and Remote Sensing Symposium* and is to be presented at the related meeting in July 2020. Finally, Chapter 4 provides conclusions and recommendations for future work.

Chapter 2

Background

This chapter is divided into three main sections designed to present the background knowledge relevant to this research. First, the mechanisms of lake ice formation and decay are addressed, followed by a review of recent trends in lake ice phenology. Then, the need for lake ice mapping by remote sensing is also argued in the first section. Section 2.2 presents an extensive introduction to optical remote sensing of ice aimed at discussing previous remote sensing methods for lake ice retrieval and highlighting current classification challenges. A description of the MODIS L1B top-of-the-atmosphere (TOA) reflectance product, the main input data used in this research, is covered in section 2.3. Finally, a review on the application of machine learning (ML) in remote sensing classification is provided to show the potential of ML algorithms for lake ice classification; algorithms that are then tested in the research manuscript included in section 2.4.

2.1 Lake Ice

2.1.1 Lake Ice Phenology

Lake ice phenology is the term used to define the stages of ice formation and decay, and the duration of ice cover on lakes. Freeze-up covers the period between initiation of ice formation on a lake surface until the time of complete ice cover, occurring in the fall and winter months (Brown and Duguay, 2010). Break-up, which is basically the progress of ice disintegration, refers to the period from the beginning of ice melt until the entire lake becomes completely ice-free, occurring in spring to summer depending on geographical location (Brown and Duguay, 2010). Ice season is defined as ice cover duration from the first day of ice presence to the day of complete ice disappearance (Kang et al., 2012).

The formation, growth, and decay of ice are affected significantly by the surplus and deficit of the energy balance at the lake surface (Williams, 1965). The energy available for ice formation and decay is influenced by three factors: heat exchange between lake and atmosphere, the heat stored in the lake, and heat import from inflows of water (Williams, 1965). The exchanges with the atmosphere are governed by climatic factors (e.g., air temperature, precipitation, wind, and radiation), whereas the amount of heat storage is mainly controlled by non-climate factors such as lake morphometry (area and depth) (Brown and Duguay, 2010).

Air temperature, in summer and fall, is the dominant climatic control on the timing of lake ice freeze-up (Williams, 1965). With more heat absorbed by the lake during ice-free months, ice formation can be delayed and vice versa. A study by Bonsal et al. (2006) indicates that freeze-up dates were delayed by up to 5 days in Western Canada during the warm anomalous years of El Niño events in the 1950 to 1999 period. When air temperatures drop in the fall, heat is lost at the surface, which causes vertical convection between the cooler denser water of the surface layer and the warmer water of the underlying layer (Brown and Duguay, 2010). With freshwater reaching its maximum density at 4℃, the lake surface cooling hampers convective overturning resulting in ice formation (Jeffries et al., 2005). Once surface water temperature has reached the freezing point, black ice, which is known as congelation ice, forms downward from the surface, generating layers of vertically orientated c-axis (column-like) ice crystals (Jeffries et al., 2005). When the weight of snow (the snow mass) is large enough to depress the ice below its hydrostatic level, water will seep through cracks in the congelation ice and wet the base of the snowpack which will result in slush formation. Slush freezing, afterward, will form snow ice, often referred to as white ice. Snow accumulation on the ice surface can also slow down the ice growth rate because it insulates the lake, thereby also reducing heat loss (Sturm et al., 1997). Additionally, during the freeze-up period, wind not only can promote the mixing of cooler water and warmer water which may delay the initial formation of skim ice (Soja et al., 2014; Williams, 1965), but it can also break the skim ice, which first forms, to delay the formation of a solid ice cover (Brown and Duguay, 2010). Another climatic variable that has been linked to ice formation and growth is cloud cover. The presence of cloud cover can lead to lower air temperatures, thus accelerating ice formation and growth, by reflecting incoming shortwave radiation away from the ice cover (Brown and Duguay, 2010; Wang et al., 2016). Conversely, Brown and Duguay (2010) indicate that clouds can trap longwave radiation to slow down ice growth by causing a warmer atmosphere, especially at night.

The process of lake ice break-up is significantly dominated by temperature via heat gain from the atmosphere and solar radiation (Brown and Duguay, 2010; Williams, 1965). The break-up period can last a few days to several weeks after the 0℃ isotherm date is reached (Duguay et al., 2006). Moreover, ice break-up shows a stronger temporal coherence with changes in air temperature, compared to ice freeze-up timing which is also strongly influence by lake morphometry (described below). Jeffries and Morris (2007) show that a ± 1°C change in air temperature results in a ± 1.86 days change in break-up dates for Alaskan ponds. Duguay et al. (2006) indicate significantly earlier break-up occurring over Canadian lakes due to climate warming in recent decades, but did not observe an apparent pattern of changes in freeze-up dates. Furthermore, ice/on-ice snow melt can be affected by the radiative processes. The albedo of the exposed surface (ice, snow, ponding water) controls the amount of solar radiation absorbed and heat available for melting (Brown and Duguay, 2010). Albedo falls typically in the range of 0.70 to 0.90 as fresh snow accumulates on the ice, and drops to 0.28-0.54 as the presence of water increases (Heron and Woo, 1994; Howell et al., 2009; Petrov et al., 2005). Snow cover increases the albedo of the lake surface to hinder the lake from absorbing heat, resulting in a delay in the timing of break-up (Michel et al., 1986). Once the temperature rises, the melt of snow and ice occurs on the lake surface, exposing the underlying darker ice (for congelation ice) (Heron and Woo, 1994). The change of crystal orientation in the surface layer reduces the albedo so that more solar radiation is absorbed by the lake surface, meaning that more heat is available for melting (Brown and Duguay, 2010; Heron and Woo, 1994). In addition to the radiative actions, wind also affects the break-up event since the mechanical process will lead to ice disintegration and the formation of large cracks (Williams, 1965).

The timing of lake ice phenology is additionally impacted by non-climatic factors including lake morphometry, lake elevation, and water inflow to the lake. Lake morphometry, linked to factors such as lake depth, area, volume, and fetch, determines the amount of heat storage in the water body that affects the time needed for the lake to lose heat and eventually freeze (Brown and Duguay, 2010; Korhonen, 2006). Deeper lakes can accumulate more heat during ice-free seasons (i.e. summer and fall) due to their large thermal inertia (Choiński et al., 2015).

Moreover, lake fetch, which is the longest distance over the lake surface that can generate wind-driven waves, influences the event of ice formation (Jeffries et al., 2012). At the initial timing of ice formation, the average bulk temperature on small lakes is around 2 to 3℃, which is higher than the bulk temperature on large lakes (lower than 1℃) (Jeffries et al., 2005; Scott, 1964). The relationship between lake elevation and lake ice phenology has been described in a few studies (Brown and Duguay, 2010; Livingstone et al., 2010; Williams and Stefan, 2006). For example, Livingstone et al. (2010) reported ice cover duration increasing by 10.2 days per 100 m altitude on alpine lakes, whose elevations range from 1,581 to 2,157 meters above sea level, located in the Tatra Mountains, Poland. Water generated from rivers or land runoff affects the ice events of break-up and freeze-up by breaking ice cover mechanically (Brown and Duguay, 2010; Howell et al., 2009; Williams, 1965).

2.1.2 Recent Trends in Lake Ice Phenology

The timing and duration of lake ice events are the main metrics sensitive to weather and climate conditions, thus ice phenology can be considered as a powerful indicator of climate change and variability (Duguay et al., 2014, 2006; Mishra et al., 2011). Trends in lake ice phenology over the Northern Hemisphere have attracted public interest recently for studies of climate change at large spatial scales. A study by Magnuson et al. (2000), examining the changes in the freshwater ice events from 1846 to 1995, shows a trend of 5.8 days later per decade for ice freeze-up and 6.5 days earlier per decade for ice break-up around the Northern Hemisphere. Benson et al. (2012) performed a linear regression analysis of ice phenology variables using the dataset of 75 Northern Hemisphere lakes over the period of 1855-2005. Their study shows 0.3-1.6 days per decade trend towards later for freeze-up, 0.5−1.9 days per decade trend for earlier for break-up, and 0.7−4.3 days per decade trend towards shorter ice duration. A recent study by Du et al. (2017) reveals trends towards later ice-on dates in 43 of 73 study lakes located in the Northern Hemisphere among the period of 2002-2015, also demonstrating an increasingly shorter ice cover season. Meanwhile, they also indicate a latitudinal pattern of the changing trends. Specifically, lakes at higher latitudes (> 60° N) are more likely to experience trends of earlier ice break-up and shorter ice seasons compared to lakes at lower latitudes (<

50° N) (J. Du et al., 2017). A prediction study on future trends of lake ice phenology by Dibike et al. (2012) shows that freeze-up will shift later by 5-20 days and break-up will shift earlier by 10-30 days around the Northern Hemisphere by 2040-2079 when compared to the baseline period of 1960 – 1999. The above analyses, as well as others, have indicated that warming climate conditions have led to the earlier occurrence of lake ice break-up dates broadly over the Northern Hemisphere.

Several recent studies also present trends of lake ice phenology over more specific regions. Choiński et al. (2015) describe trends in later formation of complete ice cover and earlier disappearance of ice over 18 polish lakes between 1961 and 2010. Brown and Duguay (2011) performed simulations of lake ice phenology over the North America Arctic region applying the Canadian Lake Ice Model (CLIMo). The simulation results reveal a 10–30 day reduction for the ice cover duration by 2041-2070, compared to the period of 1961–1990. Cai et al. (2019a) analyze the characteristics of ice phenology over 58 lakes on the Tibetan Plateau from 2001 to 2017. This study indicates that the freeze-up events have been delayed and ice cover duration has become shorter over the majority of the study lakes (Cai et al., 2019a). Furthermore, changing patterns of ice phenology associated with later freeze and shorter ice cover duration have been found in other studies on lakes located on the Tibetan Plateau (Gou et al., 2017; Qi et al., 2019) and Northeast China (Yang et al., 2019).

2.1.3 The Need for Monitoring Lake Ice by Remote Sensing

Diverse methods have been proposed for monitoring lake ice phenology, including surface-based (government agencies or volunteers) and satellite-based networks (Brown and Duguay, 2010, 2012; Duguay et al., 2014; Jeffries and Morris, 2007). Nationally, surface-based networks have mainly been supported by operations at meteorological or hydrological stations (Brown and Duguay 2010). However, the stations are distributed unevenly and sparsely with limited records available to support studies over remote regions. In addition, a cutback in surface-based observational networks has unfortunately occurred globally since the 1980s (Duguay et al., 2006; Key et al., 2007). The volunteer-based networks require volunteer observers to collect data for lake ice with digital camera imagery (Dyck, 2007; Jeffries and

Morris, 2007). Monitoring lake ice from the volunteer-based networks is challenging for two reasons. First, since lake ice can freeze and melt multiple times in winter and spring, definitive freeze-up and break-up dates are difficult to determine (Futter, 2003). Another reason is that long-term and wide-scale monitoring requires a lot of volunteers who will hold various definitions for freeze-up and break-up (Futter, 2003). Surface-based networks operated by volunteers provide quite limited measurements depending on the human experience and the observer scope.

In contrast, remote sensing data provide large-scale and objective observations ranging from individual lakes to regional or even global scale. Hence, mapping lake ice from remote sensing data is vital to assess lake ice phenology over large spatial scales and over increasingly longer time periods.

2.2 Optical Remote Sensing of Lake Ice

2.2.1 Optical Properties of Lake Ice

Lake ice is a relatively translucent material with an intricate structure and complex optical properties. The term “optical” refers to the portion of the electromagnetic spectrum of the wavelength (shortwave) range of radiation from the sun (roughly 0.25 to 2.50 μm). Understanding the reflection, absorption, and transmission of shortwave radiation by lake ice is useful to drive lake ice classification from optical remote sensing imagery. The optical properties of lake ice vary with ice type, the bubble content of ice, and incident radiation.

Due to various physical structures, lake ice exhibits a variety of optical characteristics. A pilot study by Bolsenga (1969) investigated the broadband albedo (0.3~3.0 µm) of various types of snow-free ice over the Great Lakes at solar elevations ranging from 32° to 40°. Albedo ranged from 0.10 for clear ice to 0.46 for snow ice. Refrozen slush ice and brash ice also showed high albedos above 0.40, which are slightly higher than that of pancake ice (0.31) and slush curd ice (0.32). In a subsequent study by Bolsenga (1983), the spectral reflectance signatures of lake ice types from 0.340 to 1.100 µm were examined. Figure 2-1 shows selected results from Bolsenga (1983) (i.e., brash ice, refrozen slush ice, and clear ice). Overall, the spectral reflectance of lake ice types is quite uniform across the visible spectrum. There is a slight rise from 0.340 to 0.550 µm and a faint decline from 0.550 to 0.700 µm, forming a wave peak at around 0.550 µm; however, a rapid decrease of reflectance (approx. 0.20 – 0.30) occurs from 0.700 to 1.100 µm (Bolsenga, 1983). Likewise, the measured surface reflectance of clear ice was around 0.10 across the spectrum. Other ice types (i.e., snow ice and refrozen slush ice) have remarkably high reflection compared to clear ice. Maslanik and Barry (1987) analyzed mean digital counts of different ice and open water types recorded by Landsat Thematic Mapper (TM) channels 1-4. The results show that snow-free black ice with low reflectance was not distinguishable clearly from turbid water in the spectral range of the four channels. Additionally, congelation ice presents very high radiation transmittance from 0.77 to 0.89 in the 0.400 – 0.700 µm range (Bolsenga, 1981), accompanying low reflection.

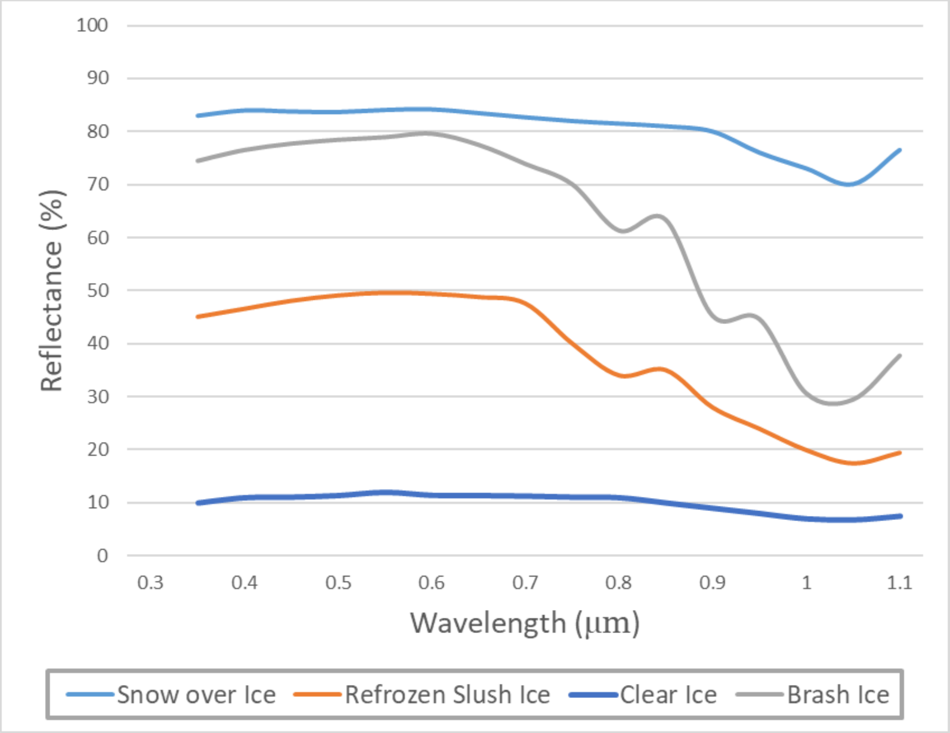
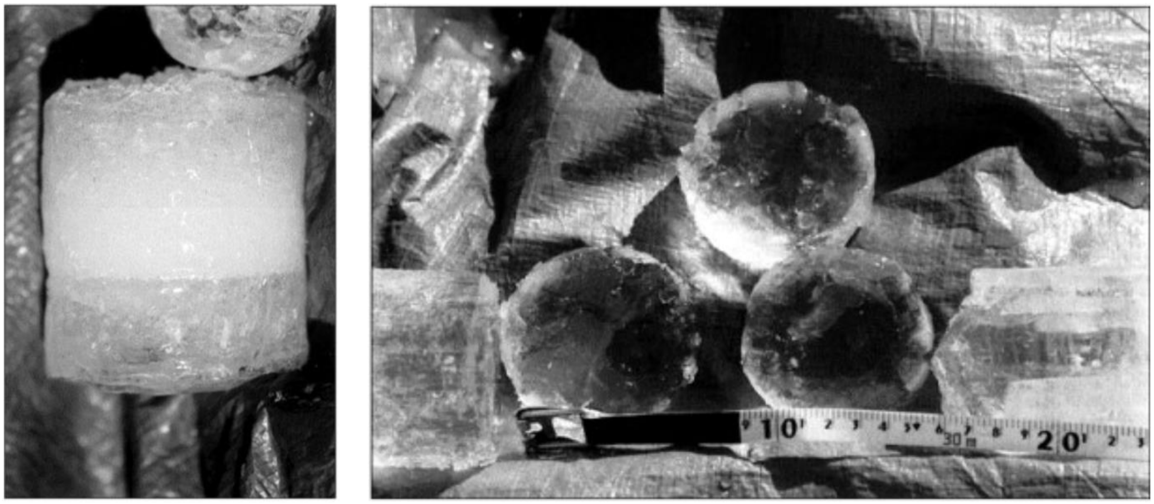


Figure 2-1 Spectral reflectance factor of lake ice in VIS and NIR range. Figure reproduced based on observations by Bolsenga (1983).

As shown in Figure 2-2, snow ice (a.) has a white appearance, and clear ice (b.) is transparent visually. Snow ice consists of numerous spherical bubbles since air dissolved in the surface saturated water cannot be incorporated into the ice crystal lattice (Mullen and Warren, 1988). However, the concentration of bubbles in congelation ice is generally low for water bodies (i.e. excluding shallow Arctic/sub-Arctic lakes). In the case of the shortwave spectrum, the scattering of light by bubbles increases the albedo over the spectral range since specular reflection exits alone (Mullen and Warren, 1988). In addition to the effect of bubbles, as mentioned previously, the crystal orientation can also affect the albedo of ice. Heron and Woo (1994) found a significant decrease of albedo from 0.45 to 0.20 during the break-up period. The removal of the c-axis vertical ice resulted in exposure of the underlying c-axis horizontal crystals. Laboratory sea ice observations reveal that the reflectance of ice increases as ice thickness increases (Perovich, 1979; Perovich et al., 1998).

Besides the effect of ice type, snow cover can also result in spectral variations of remote sensing imagery. Snow accumulation on lake ice occurs during the ice season. Snow-covered ice overall demonstrates very high reflectivity compared to snow-free ice, and thereby snow cover on top of the ice is clearly distinguishable from snow-free ice (Maslanik and Barry, 1987). According to the study by Bolsenga (1983), as shown in Figure 2-1, the reflectance values for ice are significantly lower than those of snow over ice. The variation of snow reflectance is highly dependent on grain size; specifically, snow reflectance decreases with an increase in grain size.

Optical remote sensing instruments mounted on satellites typically measure solar radiation reflected by the Earth with narrow fields of view. Hence, parameters affecting incident radiation also influence the reflectance of ice. The daily variability of ice albedo is high due to changing solar elevation (Bolsenga, 1977; Leppäranta et al., 2010). With solar zenith angle increasing, the diffuse radiation flux increases, resulting in the attenuation of the radiation flux of incident sunlight at the surface (Coakley, 2003). Therefore, top-of-the-atmosphere (TOA) reflectance is very low due to the lack of solar radiation reflected by the surface.



a. b.

Figure 2-2 Types of ice present on shallow sub-Arctic lakes, Churchill, Manitoba: (a.) snow (white) ice; (b.) clear (bubble-free) ice. Source: Duguay et al. (2002).

2.2.2 Reviews of Lake Ice Cover Retrieval Approaches

Various knowledge-driven, threshold-based, methods have been developed to retrieve lake ice from optical remote sensing imagery. The main idea of the knowledge-driven algorithms is to develop generic rules using inference from empirical observations.

The Moderate Resolution Imaging Spectroradiometer (MODIS) snow products in collection 5 (C5) and 6 (C6) also include retrieved lake ice. The MODIS snow product was developed using MODIS Level 1B (TOA), MODIS Cloud Mask products and geolocation fields. In the C5 product, snow cover was classified by a set of decision rules with the thresholds (Riggs et al., 2006) shown in equation 1.

$$\begin{gathered} NDSI = \frac{(Band\ 4 - Band\ 6)}{(Band\ 4 + Band\ 6)} \geq 0.4 \\ Band\ 2 > 0.11 \\ Band\ 4 > 0.1 \end{gathered} \tag{1}$$

where Band 2: reflectance at 0.865 μm; Band 4: reflectance at 0.555 μm; Band 6: reflectance at 1.640 μm. Lake ice is detected in the C5 product using the same criteria as for snow on land and a lake mask. The Normalized Difference Snow Index (NDSI) serves as the basis for the MODIS snow product. The ice phenology retrieved by the MODIS snow C5 product from 2000 – 2011 for Quebec, Canada is comparable to the simulation derived by the 1-D Canadian Lake Ice Model (Brown and Duguay, 2012). Furthermore, the snow C5 imagery was found to present lake ice conditions similar to that of the Interactive Multi-sensor Snow and Ice Mapping System (IMS) product, produced from multiple remote sensing sources (Brown and Duguay, 2012). Chen et al. (2018) employed the MODIS snow C5 product to calculate the daily fraction of ice cover on alpine lakes of the Tibetan Plateau (TP) to detect the change of ice phenology from 2002 to 2015. Likewise, Cai et al. (2019a) applied the snow products from Aqua and Terra to determine ice phenology over eight lakes on the TP using the same calculation method of ice fraction. The accuracy of the MODIS snow cover products has been validated against other sources (i.e., Landsat imagery, AMSR-E/2, SSM/I), and shows varying agreements (Cai et al., 2019a). Moreover, using the MODIS snow C5 products, Murfitt and Brown (2017) investigated short term trends in lake ice phenology for Ontario and Manitoba in Canada. The validation results against in-situ data show an average mean absolute error (MAE) of 9 days for both ice-on and ice-off dates (Murfitt and Brown, 2017). In addition to the daily snow product, the MODIS 8-day composite snow products in C5 have been used to monitor lake ice phenology over the TP region (Kropáček et al., 2013).

Compared to the snow C5 product, the MODIS snow C6 product presents a NDSI value higher than 0 in each pixel (Riggs and Hall, 2015). Moreover, a number of data screens have been developed as filter criteria to further identify snow (Riggs and Hall, 2015). When any pixel with a valid NDSI value fails on one of the screens, the pixel is flagged as no snow or uncertain snow (Riggs and Hall, 2015). The MODIS snow C6 product has been applied to determine lake ice phenology using band thresholds on the TP (Qiu et al., 2019), Northeast China (Yang et al., 2019), and Xinjiang, China (Cai et al., 2019b). However, the classification by the MODIS snow product shows confusion between clouds and ice/snow due to the occurrence of omission when the MODIS cloud mask misclassifies regions of snow/ice as a certain cloud. Similarly, the retrieval algorithm using NDSI was applied using the Visible Infrared Imaging Radiometer Suite (VIIRS) to perform lake ice classification with bias range from 0.25% to 3.2% in comparison with the AMSR2 product (Dorofy et al., 2016).

Besides the MODIS snow products, the MODIS surface reflectance product has been employed commonly to examine lake ice phenology. Šmejkalová et al. (2016) built daily surface-reflectance time series for 2000–2013 from MODIS band 2 (0.865 μm) data to determine ice phenology in the Arctic region. The results present a root-mean-square-error (RMSE) of 6.16 days in comparison to in-situ data; the authors also indicate that cloud cover and low sun angle during freeze-up significantly affect the quality of the estimation (Šmejkalová et al., 2016). A threshold-based approach using MODIS bands 1 (0.645 μm) and 2 (0.865 μm) was developed to discriminate lake ice and open water, thus identifying ice phenology events on the TP with combing the MODIS surface temperature products (Gou et al., 2017). Furthermore, the MODIS snow product has been utilized with surface reflectance bands 3-5 to calculate the ice fraction in the study area (Gou et al., 2017). In this research, a pixel is classified as ice only if all three sources indicate ice, resulting in cautious perdition (Gou et al., 2017). Likewise, Qi et al. (2019) developed another threshold-based algorithm using MODIS surface reflectance bands 1 and 2 for ice monitoring on Qinghai Lake. However, instead of applying two bands respectively, the difference between bands 1 and 2 was used as a criterion of the algorithm to label pixels as lake ice ($\text{Band 1} - \text{Band 2} > 0.028$); another criterion was a threshold of band 1 only ($\text{Band 1} > 0.05$) (Qi et al., 2019). Additionally, Zhang and Pavelsky (2019) used dynamic thresholds of MODIS band 2 to discriminate lake ice based on the size of lakes.

The MODIS top-of-the-atmosphere (TOA) reflectance product was applied by Reed et al. (2009) to provide a more accurate lake ice classification by manual interpretation as compared to various digital analysis techniques. The TOA reflectance product has also been combined with MODIS sea surface temperature data to identify the occurrence of water-clear-of-ice by a threshold-based approach for Lake Baikal, Russia (Nonaka et al., 2007). Making use of Geostationary Operational Environmental Satellite GOES imagery, Dorofy et al. (2016) identified two lake ice classes (i.e., thick ice, gray ice) in the Great Lakes region by a threshold-based approach developed using Mid-Infrared Sea and Lake Ice Index (MISI), incorporating reflectance of the mid-infrared and visible bands. They indicated that MISI considers the spectral differences between thick ice and the relatively darker, gray ice so that it is useful for distinguishing the two types of lake ice (Dorofy et al., 2016).

The knowledge-driven algorithms have generally exploited variations of the optical properties of ice, water, and cloud to define specific spectral criteria to classify these features. NDSI has not only been employed for the development of the MODIS snow product, but also utilized as the main means to perform lake ice estimation from optical remote sensing data. Additionally, the ability of ice to reflect high radiation of NIR wavelengths (MODIS band 2) to the atmosphere was found as a useful characteristic for detecting lake ice from open water, having therefore been used for lake ice investigations frequently (Oke, 1987; Svacina et al., 2014). However, the varying agreements of the MODIS snow products and validation data indicate the limitation of lake ice classification from the products. Moreover, most studies using the MODIS reflectance products to classify lake ice employed the MODIS cloud mask, known as MOD35, to filter cloudy pixels. However, previous assessments of the MODIS snow and cloud products have shown confusion between ice and cloud (Hall and Riggs, 2007; Leinenkugel et al., 2013; Tekeli et al., 2005). The confusion significantly affects the quality of lake ice cover estimations.

2.2.3 Challenges for Lake Ice Classification

A series of knowledge-driven algorithms have been developed to monitor lake ice in many regions. However, the existing knowledge-driven algorithms have difficulties dealing with complex conditions to provide highly accurate classification results. For example, the condition of high solar zenith angles in high-latitude regions results in lower TOA reflectance over lakes in the visible-infrared spectral range during the ice freeze-up period. Thus, threshold-based retrieval algorithms for lake ice-water classification using TOA satellite data do not perform well under such a condition. Šmejkalová et al. (2016) indicated that freeze-up dates for high latitude lakes are difficult to identify due to problems with high solar zenith angles. High solar zenith angles prevent the solid classification of snow/ice cover, which is important when studying freeze-up at northern latitudes (Brown and Duguay, 2012). Moreover, traditional threshold-based approaches for cloud detection still face severe challenges over ice-covered areas in the Arctic and sub-Arctic (Chen et al., 2018). As pointed out earlier, the confusion between ice and cloud exists in the MODIS cloud product, therefore introducing uncertainty in lake ice maps (Hall and Riggs, 2007; Leinenkugel et al., 2013; Tekeli et al., 2005). The classification challenge due to the low reflectance contrast of black ice with turbid water has been indicated in numerous studies regarding ice monitoring (Bolsenga, 1983; Liu et al., 2016; Mullen and Warren, 1988; Perovich, 1979). Additionally, the threshold-based approaches significantly rely on particular sensors that are not directly applicable to data obtained with other sensors. This is especially true for cloud detection algorithms because these algorithms strongly depend on the thermal bands. However, unlike MODIS, the majority of optical remote sensing data with high spatial resolution, do not have sufficient thermal bands. Therefore, the threshold-based algorithms lack generalization ability to transfer to mutiple sensors (Hagolle et al., 2010; Roy et al., 2010).

2.3 MODIS Level 1B Product

The MODIS sensors, onboard the Terra and Aqua satellites, launched in 1999 and 2002 respectively, scan the majority of the entire Earth's surface every day. The two satellites are in a Sun-synchronous orbit at a 705 km altitude, such as Terra at a 10:30 AM equatorial crossing time, descending node, and Aqua at a 1:30 PM equatorial crossing time, ascending node (Wolfe, 2006). MODIS takes measurements with a whiskbroom electro-optical instrument to provide scans in the along-track direction (Wolfe, 2006). Since launch, both MODIS instruments have been delivering near-continuous observations and yielding scientific and environmental products useful to study the Earth's system of atmosphere, land, ocean, and cryosphere (Xiong et al., 2015). As shown in the last section, the MODIS products have gained popularity for delineating lake ice cover by the threshold-based algorithms.

The MODIS Level 1B (L1B) product, the main input data product used in the ML algorithms tested in this thesis, is comprised of calibrated Earth view data of 36 spectral bands ranging from 0.41 to 14.4 µm, stored in three HDF files corresponding to three spatial resolutions. The MODIS 36 spectral bands are bands 1 and 2 at a nadir spatial resolution of 250 m, bands 3-7 at a nadir spatial resolution of 500 m, and all other bands at a nadir spatial resolution of 1 km. Specifically, bands 1-19 and 26 are the reflective solar bands, and bands 20-25 and 27-31 are the thermal emissive bands. The L1B calibrated data include TOA reflectance for the reflective solar bands, and radiances for both the reflective solar and thermal emissive bands.

The now twenty-year record of MODIS data from the Terra satellite (2000-present) is useful for investigating changes in lake ice phenology over a relatively long time period. The MODIS L1B product allows monitoring lake ice conditions at a high temporal resolution (daily). Moreover, the 36 bands provide sufficient spectral information of the Earth surface helpful to map lake ice extent. The powerful capability of Earth coverage and the highest spatial resolution of 250 m are allow for lake ice mapping at the global scale.

2.4 Machine Learning in Remote Sensing

2.4.1 Review of Machine Learning for Ice Classification

The overall challenge for lake ice mapping using optical imagery is to identify two features with similar optical properties, such as thin cloud against ice, black ice against water, turbid water against ice. Machine learning (ML) techniques, known as data-driven algorithms, are generally able to model complex class signatures with a variety of input variable data. Hence, ML classification has received considerable attention for the past several decades and researchers in the field of remote sensing are increasingly applying these classifiers for ice detection.

Several ML models have been developed for ice retrieval from Synthetic Aperture Radar (SAR) imagery. In order to provide sea ice observations with a high level of confidence for data assimilation of a climate change prediction system, Komarov and Buehner (2017) proposed a technique using logistic regression (LR) for automated detection of ice and open water using RADARSAT-2 ScanSAR images. Three input features computed from SAR and wind speed data are used in the proposed LR model. A rigorous probability threshold of 0.95 was adopted to determine ice pixels, producing 79.41% ice-class accuracy. A further study by

Komarov and Buehner (2018) introduced an adaptive probability thresholding approach to improve this technique. The authors applied sequential RADARSAT-2 imagery to examine the performance of the improved technique and indicated that 98.93% of collected sea ice observations were retrieved correctly (Komarov and Buehner, 2019). Shen et al. (2017) compared the performance of six ML classifiers for sea ice classification from Cryosat-2 data. Random forest (RF) achieved the best performance (about 90% classification accuracy), followed by Support Vector Machine (SVM), back propagation neural network (BPNN), and Bayesian, with K nearest-neighbor (KNN) performing the worst (Shen et al., 2017). SVM has been tested for detection of multiple sea ice types using RADARSAT-2 imagery with above 86% classification accuracy (Liu et al., 2015). The SVM model requires three input variables, i.e., HH, HV, and the gray-level co-occurrence matrix (GLCM) feature, which is also an approach to extract textural features (Liu et al., 2015). Due to the limited number of SAR bands, Han et al. (2017) generated 12 texture features from KOMPSAT-5 applied on a RF model, producing a 99.24% overall accuracy of ice detection in the Chukchi Sea. In addition to extracting spatial patterns with GLCM, an image segmentation technique, named iterative region growing using semantics (IRGS), was employed for sea ice retrieval from single and dual polarization SAR imagery (Leigh et al., 2014; Ochilov and Clausi, 2012). IRGS is able to minimize the impact of the incidence angle variations through conducting segmentation separately on smaller polygons (Ochilov and Clausi, 2012). Furthermore, Leigh et al. (2014) combined IRGS and SVM results using 28 textural features from dual polarization SAR imagery to map sea ice with an overall classification accuracy of 96.42%. Wang et al. (2018) applied the IRGS technique but with manual labeling on ice mapping in Lake Erie, producing an overall accuracy of 89.5% from RADARSAT-2 scenes.

Besides SAR data, optical satellite imagery has been used to perform ice classification using ML techniques. Han et al. (2018) proposed a framework combining active learning and transductive SVM for sea ice detection. The framework achieved retrieval results of above 90% accuracy from Landsat-8 and above 85% accuracy from EO-1 (Han et al., 2018). Similar to the SAR applications, Su et al. (2015) derived texture features (GLCM) and surface temperature from MODIS images for ice detection in the frozen Bohai Bay, China. SVM was adopted to identify sea ice and open water pixels, producing an 87.13% overall accuracy. Tom et al. (2018) applied SVM to obtain an accuracy range from 75 to 100% on ice-water classification within four lakes located in Switzerland using the MODIS TOA reflectance product. Research by Barbieux et al. (2018) employed the decision tree (DT) technique to optimize a threshold-based algorithm of ice classification using radiometric indexes and TOA reflectance bands on five different lakes from the Landsat-8 OLI multispectral data. The range of accuracy varied from 93 to 97% among the selected study zones except for one showing an 84.4% accuracy (Barbieux et al., 2018). Moreover, MODIS, AMSR-E and SSM/I data were cobined to monitor landfast sea ice in the Antarctic using DT and RF (Kim et al., 2015). The accuracy assessment shows comparable results between RF (94.77%) and DT (93.09%) (Kim et al., 2015).

Overall (Table 2-1), the majority of studies applying ML techniques on ice detection are based on microwave remote sensing images, for instance RADARSAT-2 (Komarov and Buehner, 2017; Leigh et al., 2014; Liu et al., 2015), TerraSAR-X (Han et al., 2016), and Cryosat-2 data (Shen et al., 2017). Nevertheless, in fact, optical remote sensing imagery provides abundant spectral information on Earth's surface for ice detection. Additionally, only a few studies (Barbieux et al., 2018; Tom et al., 2018; Wang et al., 2018) have explored the feasibility and performance of ML models in lake ice mapping. Actually, the retrieval of lake ice using ML algorithms from optical remote sensing data has received much less attention than for sea ice. Hence, an examination of the capability of ML algorithms for lake ice cover mapping is very much a new topic that merits investigation.

Table 2-1 Studies on ice classification by ML approaches.

Study	Objective	Data	Algorithm	Result
Barbieux *et al.* (2018)	**Lake ice** classification (5 lakes in Europe and North America)	Landsat 8 OLI	DT	Five testing areas: 84.40% to 97.30 %
Han *et al.* (2017)	**Sea ice** classification (Chukchi Sea)	KOMPSAT-5 HH-Pol. EW	RF	Overall accuracy: 99.24%
Han *et al.* (2018)	Multiple **sea ice** classification (Baffin, Liaodong, Bohai Bays)	EO-1 Landsat-8	TSVM	Three testing areas: 87 % - 97 %
Kim *et al.* (2015)	Landfast **sea ice** classification (Antarctic)	MODIS IST/AMSR-E	DT RF	Overall accuracy of DT: 93.09% Overall accuracy of RF: 94.77%
Komarov and Buehner (2017, 2018, 2019)	**Sea ice** classification (Labrador Sea & Baffin Island)	RADARSAT-2 Dual-Pol ScanSAR	LR	Ice classification accuracy:79.41% (2017), 88.23% (2018), 98.93% (2019)
Leigh *et al.* (2014)	**Sea ice** classification (Alaskan coast)	RADARSAT-2 Dual-Pol ScanSAR	IRGS/SVM	Ice classification accuracy: 98.21% Water classification accuracy: 92.72%
Liu *et al.* (2015)	Multiple **sea ice** classification (Beaufort Sea)	RADARSAT-2 Dual-Pol ScanSAR	SVM	Overall accuracy of two testing areas: 91.74%, 91.43%
Shen *et al.* (2017)	Multiple **sea ice** classification (Arctic)	Cryosat-2	6 classifiers	RF achieved the best, followed by SVM, BPNN, Bayesian, with KNN the worst
Su *et al.* (2015)	**Sea ice** classification (Bohai Bay)	MODIS TOA (L1B)	SVM	Overall accuracy: 84.73%
Tom et al. (2018)	**Lake ice** classification (4 Swiss lakes)	MODIS TOA (L1B) VIIRS TOA (L1B)	SVM	Four testing lakes (MODIS): 99.50%~100% Four testing lakes (VIIRS): 99.30%~100%
Wang *et al.* (2018)	**Lake ice** classification (Lake Erie)	RADARSAT-2 Dual-Pol ScanSAR	IRGS	Overall accuracy: 90.4%

2.4.2 Applications of Machine Learning

The examination of variable importance and variable selection has become an apparent need in remote sensing applications using ML techniques. In the studies reviewed in the previous section, the comparison of the impact of different input variables on the algorithms has been the main object of analysis. Specifically, the tree-based algorithms provide a measurement of the relative importance values of input variables to the final model. The importance values have been used to present their attribute usage to the classifications as shown by Han et al. (2017) and Kim et al. (2015). Similar to the importance calculation by the tree-based models, permutation-based variable importance (PBVI)- another variable importance measurement- was implemented to evaluate individual variable importance to ML models (Shen et al., 2017; Xu et al., 2014). In addition to the variable importance measurement, Su et al. (2015) examined the accuracy performance of several input variable combinations for SVM, therefore identifying the most useful input variables. Additionally, a forward feature search approach proposed by Guyon and Elisseeff (2003) has been applied to extract the useful SAR textural features for SVM (Leigh et al., 2014). The objective of the variable importance measurement and selection is manifold: (a) to avoid overfitting and improve algorithm performance; (b) to provide faster and more cost-effective variables; (c) to allow a better insight of the underlying processes that generated the data (Guyon and Elisseeff, 2003).

In addition to variable selection, hyperparameter selection is necessary for exploiting the full capacity of a classifier for a given retrieval purpose. Moreover, lack of testing of different combinations of the hyperparameters could result in bias or improper perceptions of different classifiers (Shih et al., 2019). Shih et al. (2019) indicate the importance of testing the combinations of the two hyperparameters for SVM by their review of two articles. Specifically, the article by Foody and Mathur (2004) only compared the crop classifications of several Gamma values with only one given Cost value for SVM. Another research by Maxwell et al. (2018) presents that SVM derived by the optimal hyperparameter (Gamma and Cost) combination was found to have the highest overall accuracy for all testing datasets among all tested classifiers (RF, KNN, DT, ANN, etc.). Moreover, Mountrakis et al. (2011), in a review article of SVM with remote sensing, concluded that over-small and over-large hyperparameters may lead to overfitting or underfitting. Thus, the selection of SVM hyperparameters should be implemented with a trial-and-error approach when new data are introduced (Mountrakis et al., 2011).

The overfitting behavior is a major challenge in implementing ML classification from remote sensing data. With a model suffering from overfitting, the accuracy from a training dataset is far higher than that of the validation dataset. When overfitting occurs, the training dataset is usually too small or biased to represent the actual data distribution and variation. Nonetheless, in practice, the characteristics of retrieved features in remote sensing variables can differ significantly from one location to another. Additionally, features at the same location can present varying characteristics in remote sensing variables at different temporal steps due to the impact of climatic conditions and intra-annual changes (Karpatne et al., 2016). Meanwhile, the actual distribution of remote sensing observations for a ground feature is typically unknown. The ML models are prone to overfitting due to the presence of such heterogeneity in remote sensing observations across space and time. Therefore, the examination of spatial and temporal transferability of models is essential when performing classification over large spatial-scales in long-term timescales. Waske and Braun (2009) separated the whole study area into different clusters in terms of the spatial and temporal relationship, and afterward performed cross-validations across the spatial and temporal data clusters to examine the transferability of RF for land cover mapping.

Feature engineering, known as feature extraction, is able to capture domain and meaningful information from raw data via data mining techniques or prior knowledge. For example, the GLCM feature, a textural analysis, has been employed for ice detection in numerous SAR applications using ML classifiers (Leigh et al., 2014; Liu et al., 2015; Zhang et al., 2019). The optimal window size of GLCM has been generally exploited in those studies. Besides spatial features, radiometric features have been adopted in ML applications. Barbieux et al. (2018) proposed a new radiometric index (developed using Red, NIR, and SWIR bands from Landsat 8 imagery) for ice detection based on the understanding of ice optical properties. Subsequently, DT was applied to compute the optimal threshold of the index used to classify ice and water (Barbieux et al., 2018). To obtain efficient and promising retrieval, Discriminant Analysis

(DA) and Principal Component Analysis (PCA) have been applied with ML classifiers on multispectral or hyperspectral imagery (Ishida et al., 2018; Kang et al., 2014). These techniques can reduce high dimensional remote sensing data to extract the most useful and informative features by developing a new feature space.

Model comparison has been performed to indicate the optimal ML classifier in comparative studies of remote sensing classification (Cracknell and Reading, 2014; Shen et al., 2017; Xu et al., 2014). Instead of showing a classification performance by a single classifier, the comparative studies provide various analytical aspects to present a comprehensive evaluation of the classifier capability. Until now, no previous study has involved a comparison with classification results of different ML classifiers for lake ice mapping from optical remote sensing imagery. Therefore, an examination of the capability of classifiers through multi-tier comparison is needed on this topic.

Chapter 3

Mapping Lake Ice Cover from MODIS Using Machine Learning Approaches

3.1 Introduction

Lakes cover approximately 2% of the Earth’s land surface (Brown and Duguay, 2010), and a total of about 3.3% of the land surface above latitude 58°N is seasonally ice covered (Duguay et al., 2015). Lakes play a significant role in local/regional weather and climate at high latitudes. Two-way energy interactions (feedbacks) between the atmosphere-water-ice have an impact of regional weather and climate as well as the timing of lake ice formation (freeze-up) and melt (break-up), and the duration of ice growth, which are referred to as lake ice phenology (Kang et al., 2012). Ice phenology dates and ice duration are known to be particularly sensitive to changes in near-surface air temperatures. Several studies have documented trends and variability in lake ice phenology in response to climate (e.g. Duguay et al., 2006; Brown and Duguay, 2010; Howell et al., 2009; Kang et al., 2012) and changes in large-scale atmospheric teleconnection patterns; the North Atlantic Oscillation (NAO) in Europe (Blenckner et al., 2004; George, 2007; Karetnikov and Naumenko, 2008; Korhonen, 2006) and Pacific-related indices (El Ninõ/Southern Oscillation, the Pacific Decadal Oscillation, the Pacific North American pattern, and the North Pacific index) and, to a lesser extent, NAO/Arctic Oscillation in North America (e.g. Bonsal et al., 2006). Unfortunately, a cutback in the ground-based observation networks that formed the basis for documenting changes in ice cover has occurred globally since the 1980s (Duguay et al., 2006; Key et al., 2007). Hence, satellite remote sensing has assumed a greater role in recent years for the monitoring of lake ice cover, an essential climate variable (ECV) (GCOS, 2016).

Moderate Resolution Imaging Spectroradiometer (MODIS) products from NASA’s Terra (2000-present) and Aqua (2002-present) satellites have gained popularity for monitoring lake ice cover since they provide Earth observations over ca. 20 years and with at least daily temporal resolution. Knowledge-driven (threshold-based) algorithms have been developed to retrieve lake ice from MODIS top-of-atmosphere (TOA) and surface reflectance products. This type of algorithm relies on variations of the spectral signature of ice, water, and clouds to define thresholds to classify these features. Several studies (Gou et al., 2017; Qi et al., 2019; Riggs and Hall, 2015; Riggs et al., 2006; Šmejkalová et al., 2016; Zhang and Pavelsky, 2019) have applied threshold-based methods to retrieve lake ice cover and monitor ice phenology events using MODIS radiance or reflectance data. In addition to MODIS radiance and reflectance products, MODIS Terra/Aqua snow products, produced from the normalized difference snow index (NDSI), have been used to determine the presence of lake ice and lake ice phenology (Brown and Duguay, 2012; Cai et al., 2019; Kropáček et al., 2013; Murfitt and Brown, 2017). The main idea behind knowledge-driven algorithms is to develop generic rules using inference from empirical observations. However, such algorithms may fail to provide robust classification results under complex conditions. For example, lakes located in high-latitude regions exhibit lower TOA reflectance in the visible-infrared spectral range during the ice freeze-up period due to low solar illumination (i.e. large solar zenith angles). Existing threshold-based retrieval algorithms for lake ice-water classification using optical remote sensing data do not perform well under such condition. Šmejkalová et al. (2016), for example, indicate that freeze-up dates for high latitude lakes are difficult to determine due to problems with high solar zenith angles. This is one of the reasons as to why most studies using data from MODIS tend to focus on the break-up period instead; a time of the year when low solar illumination does not present a significant issue. Moreover, traditional threshold-based cloud detection approaches still face severe challenges over ice-covered areas in the Arctic and sub-Arctic (Chen et al., 2018). Finally, threshold-based algorithms develop for one sensor lack generalization ability for direct transfer to other sensors (Hagolle et al., 2010; Roy et al., 2010).

Machine learning (ML), also known as data-driven, approaches have become increasingly popular in the remote sensing community due to their ability to learn complex representations in the data and achieve excellent gains in classification accuracy. Su et al. (2015) employed a support vector machine (SVM) algorithm to detect sea ice cover from MODIS images in Bohai Bay, China, and achieved an overall accuracy of 87 %. Shen et al. (2017) compared five ML classifiers for sea ice detection using Cryosat-2 SAR data and found a random forest (RF) algorithm to provide the best overall classification accuracy at 89.15 %. Other studies (Han et al,, 2017; Han et al., 2016; Han et al., 2018; Komarov and Buehner, 2017; Liu et al., 2015) have proposed valuable applications to sea ice mapping from remote sensing data using machine learning algorithms. In contrast to the knowledge-driven retrieval algorithms that rely on the physical concepts of spectroscopy, ML approaches provide a mathematical rigorous way to extract information from reference data. Hence, machine learning models can learn and delineate complex class signatures under various conditions. To date, the retrieval of lake ice cover using ML algorithms from remote sensing data has received much less attention than for sea ice. Tom et al. (2018) applied a SVM for ice-water classification on four lakes in Switzerland using the MODIS TOA reflectance product and achieved ranging from 75% to 100%. The study conducted a binary classification of ice and water with cloud masking provided by the MODIS cloud product (MOD35) instead of performing a multiclass classification with cloud cover. The low quality of the MODIS cloud product over ice cover in some cases (Hall and Riggs, 2007; Leinenkugel et al., 2013; Tekeli et al., 2005) can introduce uncertainty in the classification.

The primary aim of this study is to investigate the capability of ML algorithms in multiclass feature (i.e. ice, water, and cloud) extraction of large northern lakes using MODIS Terra L1B TOA data. The classification performance of multinomial logistic regression (MLR), support vector machine (SVM), random forest (RF), and gradient boosting trees (GBT) is evaluated and compared. Specific objectives are to: 1) find the optimal input variables (bands) to obtain best classification results and rank the importance of each variable; 2) examine the impact of internal hyperparameters on classification accuracy for the four classifiers with the best variable selection 3) to compare the performance of classifiers based on statistical and visual assessments; and 4) assess the spatial and temporal transferability of the classifiers.

3.2 Data and methods

3.2.1 Data and study area

The MODIS instrument aboard NASA’s Terra satellite, launched in 1999, has been providing data since 2000. The instrument can view the majority of the entire Earth’s surface every day, acquiring data in 36 spectral bands from visible to thermal infrared wavelengths. The MODIS

Level 1B calibrated radiances product (MOD02), Collection 6, provides TOA reflectance data stored in three separate files as a function of spatial resolution: MOD02QKM (250 m, bands 1-2), MOD02HKM (500 m, bands: 3-7), and MOD021KM (1 km, bands 8-36). The algorithm proposed by Trishchenko et al. (2006) was applied to downscale MODIS bands 3-7 (500 m) to the same grid resolution as bands 1-2 (250 m).

Seventeen lakes (Table 3-1, Figure 3-1) distributed across the Northern Hemisphere (Eurasia and North America), which present different characteristics regarding area, latitude, altitude, freezing frequency, and ice-on duration, were selected to collect sample data (i.e. sets of pixels for training and validation of classifiers). For each lake, one image during ice freeze-up and one image during break-up were chosen if available for each of three ice seasons (2002-2003, 2009-2010, 2016-2017) over the length of the MODIS/Terra record. We used false color RGB composites (R: band 2, G: band 2, B: band 1) at 250 m grid resolution as reference images to manually collect areas of interest (AOI) with labels (lake ice, open water, and cloud). In total, 54 images (20 from freeze-up and 34 from break-up periods) were sampled which resulted in 276,003 pixels that were relatively evenly distributed in number between images.

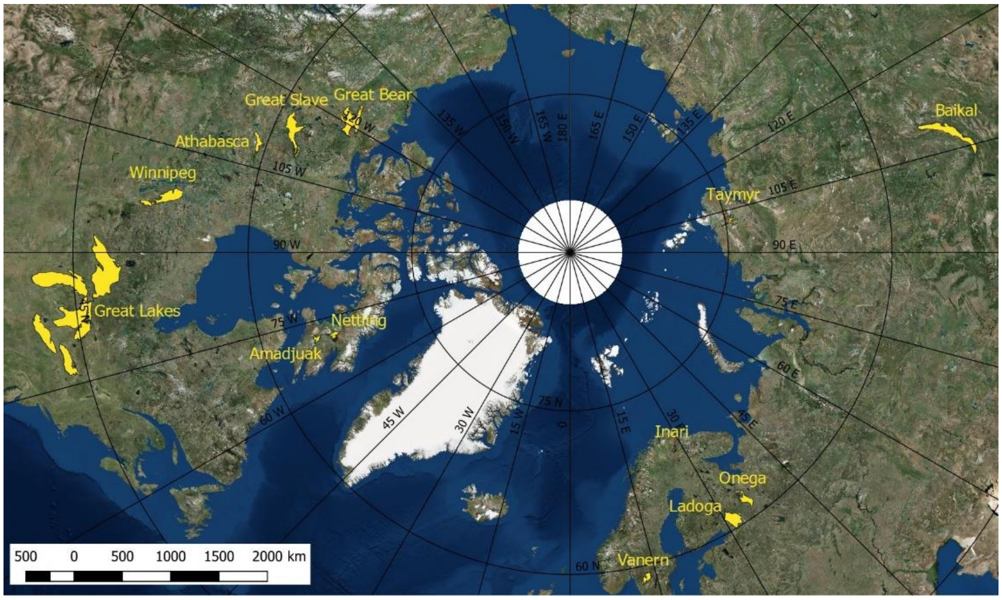


Figure 3-1 Lakes in the study shown in WGS 84/Arctic Polar Stereographic projection

Table 3-1 List of lakes selected for this study.

No.	Lake	Country	Latitude (°)	Longitude (°)	Elevation (m a.s.l.)	Area (km^2)
1	Amadjuak	Canada	64.925	-71.149	113	3,115
2	Athabasca	Canada	59.424	-109.340	213	7,900
3	Baikal	Russia	53.525	108.207	456	31,500
4	Erie	Canada/USA	42.209	-81.246	174	25,821
5	Great Bear	Canada	66.024	-120.610	186	31,153
6	Great Slave	Canada	61.579	-114.196	156	28,568
7	Huron	Canada/USA	44.918	-82.455	176	59,570
8	Inari	Finland	69.048	27.876	118	1,040
9	Ladoga	Russia	60.830	31.578	5	18,135
10	Michigan	USA	43.862	-87.093	177	58,016
11	Nettilling	Canada	66.420	-70.280	30	5,542
12	Onega	Russia	61.750	35.407	35	9,890
13	Ontario	Canada/USA	43.636	-77.727	75	19,009
14	Superior	Canada/USA	47.945	-87.320	183	82,367
15	Taymyr	Russia	74.538	101.639	6	4,560
16	Vanern	Sweden	58.880	13.220	44	5,650
17	Winnipeg	Canada	52.421	-97.677	217	23,750

3.2.2 Variable selection and variable importance measurement

Four input band configurations (Table 3-2) were evaluated to find the optimal input variables for lake ice, water, and cloud classification from the ML classifiers. All configurations include solar zenith angle (SZA) as an additional band to cope with low TOA reflectance, which is prevalent at higher latitudes during the freeze-up period. The ability of snow-covered ice and ice to reflect a significant amount of radiation at visible and near-infrared (NIR) wavelengths is a useful characteristic for discriminating lake ice from open water (Oke, 1987; Svacina et al., 2014). Reference images used for visual collection of samples therefore consisted of false color RGB images produced from a combination of red (band 1, 0.645μm) and NIR (band 2, 0.858μm) bands. Hence, the first configuration was a 3-band combination with red, NIR, and SZA as input bands. In addition to these three bands, blue (band 3, 0.469μm) and green (band 4, 0.555μm) bands are helpful for distinguishing ice from open water in areas of lakes with high turbidity that can result in similar NIR reflectance values between the two classes. Thus, the second band configuration tested used five bands (blue, green, red, NIR, and SZA). Shorter wavelengths are more affected by atmospheric aerosols (mainly in the visible spectrum), whereas shortwave infrared (SWIR) wavelengths show enhanced atmospheric transparency which facilitates the discrimination between clouds and ice. SWIR2 (band 6, 1.640μm) and SWIR3 (band 7, 2.130μm) are introduced into the third, 7-band, configuration. Since stripe noise remains in SWIR1 (band 5, 1.240μm) in the MODIS Terra product (Wang et al., 2011), we excluded this band from the study. Hence, the 7-band configuration contains spectral bands over visible, NIR, and SWIR wavelengths. Finally, Metsämäki et al. (2015) showed that MODIS thermal infrared (TIR) bands 20 (3.750μm), 31 (11.030μm), and 32 (12.055μm) are useful for detecting clouds. Hence, the fourth configuration examined consisted of 10 bands (seven bands from the third configuration plus three thermal bands).

Table 3-2 MODIS band configurations.

Configuration	Spectral bands
3-band	Red + NIR + SZA
5-band	Red + NIR + Green + Blue + SZA
7-band	Red + NIR + Green + Blue + SWIR2 + SWIR3 + SZA
10-band	Red + NIR + Green + Blue + SWIR2 + SWIR3 + Band20 + Band31 + Band32 + SZA

The determination of variable importance is of interest to many practitioners since it allows for a better understanding of variable contribution and classifier underlying process. However, it is challenging to quantify the importance of a variable in a complicated classification model developed from multiple variables since the evaluation scheme could affect the performance of variable importance. Moreover, the variables that significantly contribute to a particular model may be useless for other models, and vice versa. We therefore implemented the permutation-based variable importance (PBVI) approach to measure the importance of each individual variable. This approach has previously been applied to multiple classification problems (Xu et al., 2014). The idea of this approach is to measure the degree of accuracy degradation when the tested variable is not available. The first step of the approach is to replace the tested variable with random noise and subsequently compute the cross-validation (CV) accuracy. The k-fold (k=100) CV method was applied to calculate accuracy, and each variable was permutated 10 times. Then, the importance was calculated by the reduction in CV accuracy. To normalize the output, the importance value was divided by the largest CV accuracy reduction value to represent the relative importance of each variable. Compared to the univariate importance measurement, PBVI takes into account the interaction amongst covariate of a variable in the context of others in the evaluation of variable importance. Hence, this approach can provide a useful measurement of the variable importance for the four classifiers evaluated herein.

3.2.3 Machine learning algorithms

Four ML algorithms were evaluated with the four configurations described above: multinomial logistic regression (MLR), support vector machine (SVM), random forest (RF), and gradient boosting trees (GBT). The characteristics of each ML algorithm are described below.

Logistic regression is used as an approach to develop a model of the log odds of binary class probabilities as a linear function of one or more explanatory variables (Murphy, 2013). Then, the model can inversely compute the probability of each class using the explanatory variables of a given unknown sample. Multinomial logistic regression (MLR) is an extension of logistic regression applied to multiple response variables. One of the response variables is designated as the baseline class. In this manner, the probability of membership in the different classes is related to the probability of membership in the baseline class. The optimal values of the function parameters are computed using the training data. The MLR probability estimate for each class falls within a range from 0 to 1, resulting in a realistic probability surface. The maximum probability among the classes is the predicted class for an unknown sample.

Another algorithm is support vector machine (SVM). SVM's basic idea is to determine support vectors to build an optimal boundary separating the given observations in terms of classes (Burges, 1998; Vapnik, 1998; Weston and Watkins, 1999). The distance from the support vectors to a hyperplane is known as the margin. SVM, in its simplest form, is a linear binary classifier that labels a given sample using a hyperplane in the original input space. However, to solve the inseparability problem in the original space, SVM maps multidimensional data into an enlarged feature space to build a hyperplane using a kernel function (e.g., polynomial, radial basis, sigmoid). Since the radial basis function (RBF) kernel has a promising ability in non-linear classification, the RBF kernel was adopted in this research. Additionally, we applied the one-vs-one scheme to handle the multiclass problem. SVM is sometimes called a soft margin classifier because training samples could lie on the incorrect side of the hyperplane, thereby creating a violation. The model hyperparameter, *Cost*, is a regularization constant controlling the violation degree. Another model hyperparameter, *Gamma*, is the kernel width of RBF.

Random forest (RF), an ensemble approach, integrates decision trees developed by bagging samples to improve the limitations of the single-tree structure (Breiman, 2001). The bagging creates randomly several subsets from training samples with replacement, i.e., a sample can be collected several times in the same subset whereas other samples are probably not selected in this subset. Subsequently, each data subset is used to train a decision tree. For building a single tree, a random sample with a number of variables is chosen as split candidates from all variables. The number of variables available to a split is one of key RF hyperparameters, denoted as *mtry*. For the whole RF model, the number of trees (*ntree*) is defined *a priori* to develop various independent classifier outputs. The final class of each unknown sample is assigned by the majority vote of all outputs from the trees.

Gradient boosting trees (GBT) is another ensemble classifier inspired by the boosting technique developed by Freund and Schapire (1996). In contrast to RF, GBT applies the entire training dataset on classification rather than resampling partial samples. The training samples are initially assigned equal weights in the first iteration to develop the first tree, and afterwards the weights are altered based on the fitting performance to the training dataset. Misclassified samples in the previous iteration are assigned a higher weight in subsequent iterations. Each tree is also given a weight based on the fitting error. The final class of an unknown observation is assigned by computing the output of all trees multiplied by their weights. The term, gradient, is associated with iterative functional gradient descent algorithms used to optimize cost functions. Similar to RF, the hyperparameters of GBT, as a tree-based classifier, include number of variables available to a split (*mtry*) and number of trees (or iterations) (*ntree*). Moreover, an additional hyperparameter, learning rate (*lr*), controls overfitting in the range between 0 and 1 via shrinkage. The higher *lr* drives a faster learning process, and vice versa.

All four classifiers were implemented using the scikit-learn package in Python (Pedregosa et al., 2011). The functions invoked for classifier development are presented in Table 3-3. The table also shows the testing values of each hyperparameter to examine the sensitivity of classifiers to the classification accuracy. Generally, MLR does not have user-defined hyperparameters. However, the scikit-learn package provides several parameters to define MLR functions. The two parameters, *solver* and *niter*, concern the convergence performance of functions, which influence significantly the classification performance of the MLR classifier. *solver* is the algorithm used in the optimization problem (see details in the user guide (Scikit-learn, 2020)). *niter* is the maximum number of iterations to solve the optimization problem. For implementing ML classifiers, the internal parameter setting could be erratic when applying a big dataset with high variability. Thus, it is meaningful to investigate the effect of hyperparameters changes to classifier performance. The results of this sensitivity study appear in section 3.3.2.

Table 3-3 Classifier functions of the scikit-learn package and their hyperparameters.

Classifier	Function	Hyperparameter	
		Name	Testing values
MLR	LogisticRegression	*solver*	newton-cg, lbfgs, sag, saga
		niter	100, 500, 1000, 3000, 5000
SVM	svm.SVC	*Cost*	0.1, 1, 10, 100
		Gamma	0.01, 0.1, 1, 2, 3
RF	RandomForestClassifier	*mtry*	2, 4, 6
		ntree	50, 100, 500, 1000, 2000, 3000
GBT	GradientBoostingClassifier	*mtry*	2, 4, 6
		ntree	50, 100, 500, 1000, 2000, 3000
		lr	0.1, 0.05, 0.01

3.2.4 Cross-validation strategies

In machine learning, using the same dataset for both model training and validation can produce over-optimistic assessments of model performance. The assessment method should guarantee that the data used to validate models are independent from the data used to train models. In this research, therefore, we applied three cross-validation (CV) strategies to provide a comprehensive comparison of classifiers for global lake ice mapping over long time-series. The first one is the random k-fold CV, which has been employed to obtain bias-reduced accuracy measurement of classifiers with remote sensing data (Hand, 1997). This CV strategy randomly separates a data set into k subsamples of equal size. Subsequently, of the k subsamples, one single subsample is retained to validate the classifier developed using k-1 subsamples. The validation process is repeated k times until each subsample has been used once as a testing data. The overall classification accuracy is averaged over the accuracy of all k subsamples. The 100-fold CV was used in this research and the results are presented in section 3.3.3.

We also implemented spatial and temporal CV strategies to enhance the interpretability of the classifiers' capability to cope with spatiotemporal heterogeneity in data. Spatial CV, also named leave-location-out CV, has been applied to assess the performance of classifiers in previous studies (Gasch et al., 2015; Ho et al., 2014; Meyer et al., 2016). In spatial CV, models are repetitively built by leaving the data from one location or a group of locations out and using the remaining one for model validation. We grouped the study areas into 11 clusters (lake or set of lakes falling into a particular lake region) shown in Table 3-4. The spatial CV accuracy can, therefore be considered a valuable performance indicator to examine a classifier's ability to cope with spatiotemporal issues. Similar to the spatial CV, the temporal CV approach separates the data based on time steps. Subsequently, a subsample of one-time step is removed, and the model is trained on the subsamples of the reminding time steps. For this CV strategy, we selected three ice seasons (2002-2003, 2009-2010, 2016-2017) separated by seven years across the full Terra/MODIS record. One reason for selecting such a spread was to assess the stability of the classifiers' performance over time which could be affected, for example, by MODIS sensor degradation. Results of the spatial and temporal transferability of the four classifiers are described in section 3.3.4.

Table 3-4 The clusters for spatial CV.

Clusters	Lakes
AN	Nettilling Lake, Amadjuak Lake
Ath	Lake Athabasca
Bai	Lake Baikal
GBL	Great Bear Lake
GLs	Lake Ontario, Lake Superior, Lake Huron, Lake Erie, Lake Michigan
GSL	Great Slave Lake
Ina	Lake Inari
OL	Lake Onega, Lake Ladoga
Tay	Lake Taymyr
Van	Lake Vanern
Win	Lake Winnipeg

3.3 Results and discussion

3.3.1 Comparison of variables combinations

The effect of changes in the number of input spectral bands on overall accuracy was assessed. As Figure 3-2 shows, the overall accuracy improves for each classifier along with the number of input bands used. The accuracy increases from the 5-band to the 7-band configuration in each classifier (i.e. MLR: 8.38%; SVM: 6.62%; RF: 5.65%; GBT: 1.59%). However, applying the 10-band configuration into the classifiers leads to results comparable to the 7-band configuration. Wieland et al. (2019), for example, have shown a slight improvement in accuracy for cloud detection when TIR bands are added into the input feature space of a convolutional neural network. Taking into account the potential for transferability of the classifiers across multispectral remote sensing datasets that do not provide TIR bands (e.g., Sentinel-2, Worldview1-3, Gaofen-1&2), we decided to retain the 7-band configuration instead of the 10-band configuration that includes the TIR bands in the subsequent steps of this research. The difference between the minimum and maximum accuracy (3- to 10-band configuration) is above 8% for MLR, SVM, and RF. However, for GBT it is only a 2.81%

difference in accuracy over the four configurations. It is especially noticeable that GBT with only the 3-band configuration was still able to produce a 96.05% accuracy, which is superior to the best performance of MLR trained with the 10-band configuration. Thus, GBT is least influenced by the input band configurations among the proposed models.

Figure 3-3 shows the permutation-based variable importance (PBVI) measurement derived from the 7-band configuration. NIR and SWIR-3 bands have very high PBVI values for all classifiers. Additionally, SWIR-2 has higher PBVI values in three of the four classifiers (SVM, RF, and GBT). The two tree-based models (RF and GBT) depict the same pattern of variable importance, where NIR and two SWIR bands are the most dominant variables in the classification, and the visible bands (red, green, and blue) show very small importance values. However, both MLR and SVM achieved predominant usage of red and green bands. Interestingly, none of the four classifiers yielded a higher PBVI value on the SZA. This may be due to the fact that overall few of the sampled images and only a few lakes, particularly during the freeze-up period, are affected by low TOA reflectance due to high SZA.

The major accuracy improvement from the 5-band configuration to the 7-band configuration indicates that SWIR bands significantly add to the classification accuracy. This is also revealed in the variable importance measurement results of Figure 3-3. Previous studies on image classification from optical remote sensing data have shown that adding SWIR bands into ML models can increase classification accuracy due to more accurate cloud detection than with Visible-NIR bands alone (Chai et al., 2019; Chen et al., 2018; Wieland et al., 2019). In addition, in terms of the water-ice classification, the accuracy produced by the 7-band configuration did not outperform that obtained with the 5-band configuration. This is because the NIR band dominates in the discrimination between ice and open water consequent to the ice's ability to reflect more radiation in this part of the spectrum (Brown and Duguay, 2012; Jönsson and Eklundh, 2004; Nonaka et al., 2007; Šmejkalová et al., 2016). High PBVI values also indicate the significance of the NIR band in all classifiers for lake ice mapping with multispectral remote sensing data. Meanwhile, given the high sensitivity of NIR reflectance to the presence of lake ice, previous studies using knowledge-driven (threshold-based) approaches have widely employed the MODIS NIR band alone or in combination with other bands and indices to detect ice phenology dates.

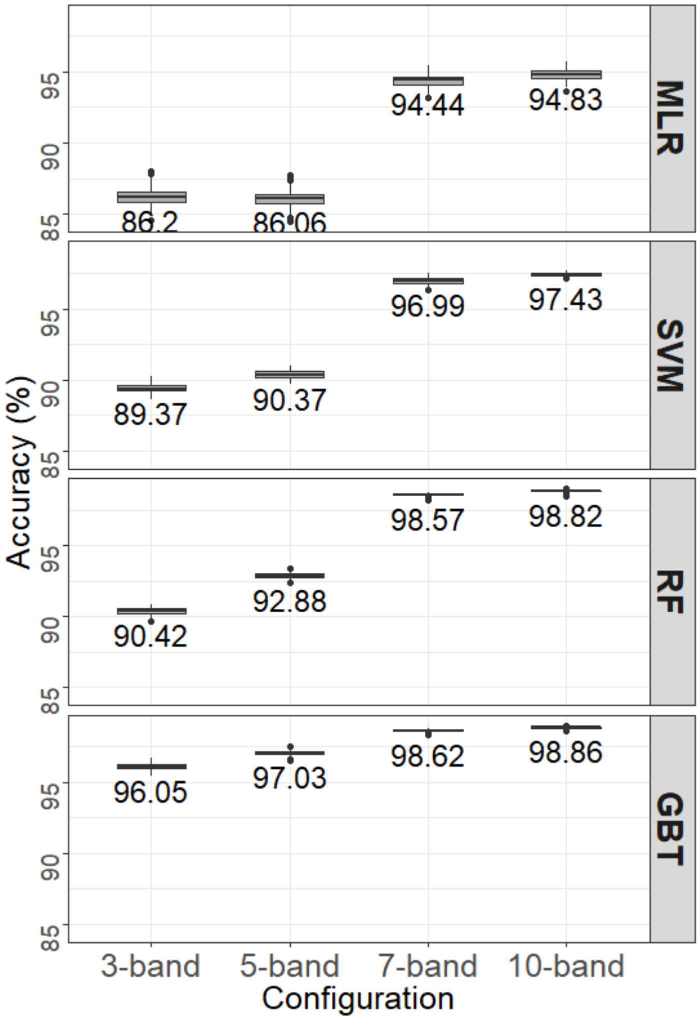


Figure 3-2 Comparison of classification accuracies (%) obtained with different band configurations across classifiers.

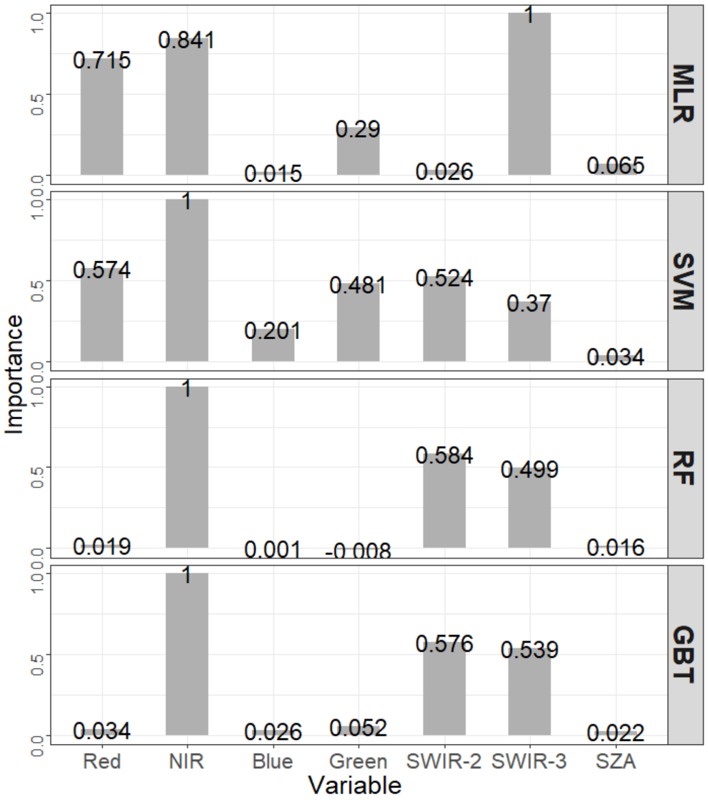


Figure 3-3 Comparison of permutation-based variable importance for the input bands across classifiers based on 7-band configuration.

3.3.2 Sensitivity to classifier hyperparameters

We employed a 100-fold cross-validation (CV) strategy to examine the impact of internal hyperparameters on classification accuracy for the four classifiers with the 7-band configuration. Figure 3-4 shows results of the sensitivity study of classifier hyperparameters. According to Figure 3-4-a, the MLR classification performance is sensitive to *solver*. The *newton-cg* solver can produce stable classification accuracies (94.01% - 94.44%) over the tested iterations. The other three solvers require a large number of iterations to produce high CV accuracy. However, even under high iterations, they still perform worse than the *newton-cg* solver.

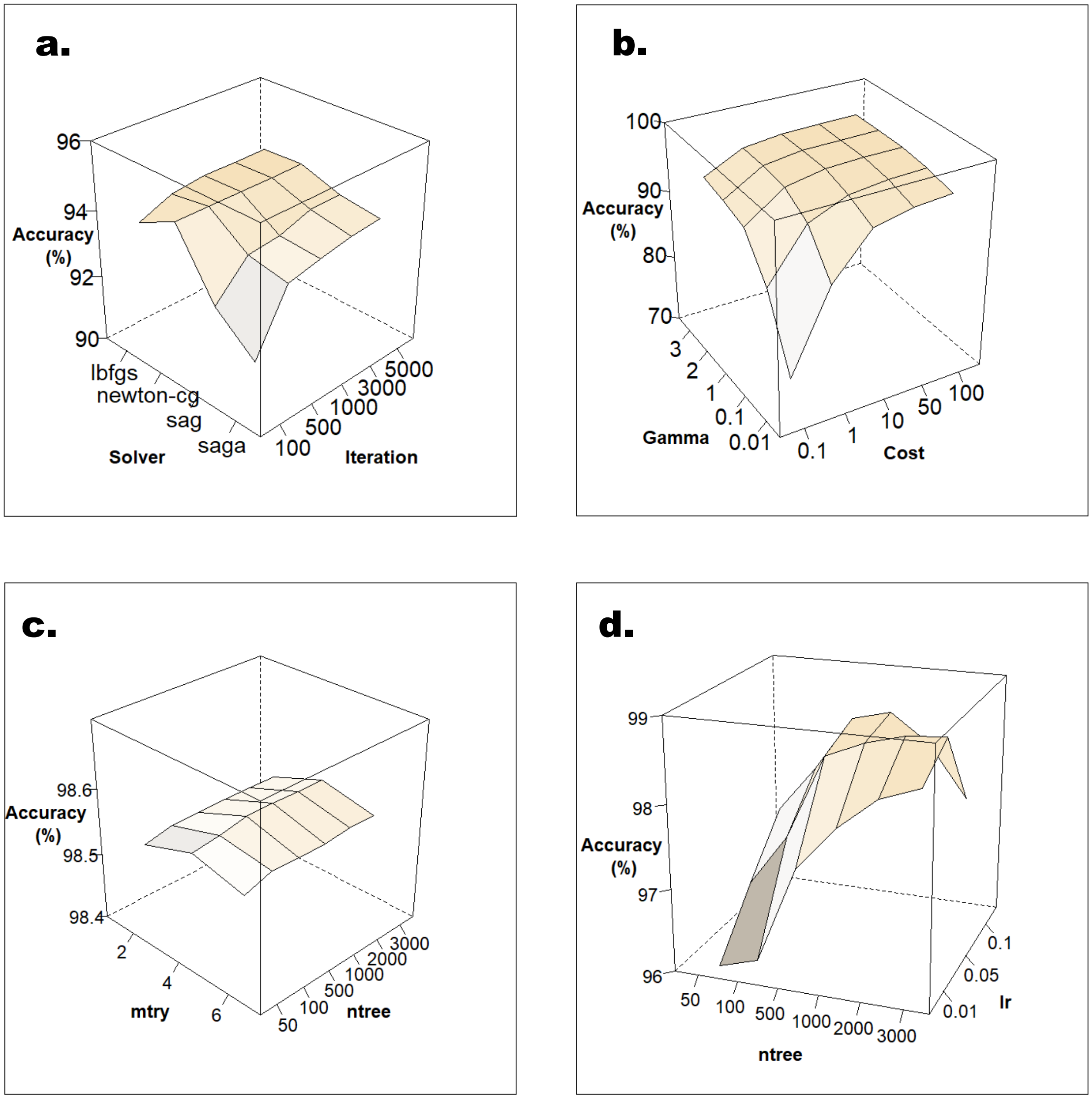


Figure 3-4 Comparison of classification accuracies (%) as a function of classifier hyperparameters based on 7-band configuration. (a) MLR, (b) SVM, (c) RF, and (d) GBT.

Cost and Gamma are two key hyperparameters of SVM with the radial basis function (RBF) kernel. Cost controls the margin width to trade-off misclassified samples in order to generate a robust classifier. A small value for Cost leads to a broader margin with high training errors, while a large value results in a severe margin identical to the hard margin. Gamma is a free hyperparameter of the radial basis function. Figure 3-4-b shows that low Cost and Gamma values lead to low accuracy, whereas increasing the values of the two hyperparameters result in higher classification accuracy. The accuracy range obtained is between 75.23% and 96.99% over all hyperparameter combinations. The results indicate that the performance of SVM is highly sensitive to both hyperparameters. Additionally, previous studies have shown that Gamma is less effective than Cost on classification performance; furthermore, the hyperparameters are also sensitive to the size and variation of training data (Huang et al., 2002; Kavzoglu and Colkesen, 2009). Mountrakis et al. (2011), in a review article on SVM in remote sensing, conclude that over-small and over-large hyperparameters may lead to over-fitting or over-smoothing. Thus, the selection of SVM hyperparameters should be implemented with a trial-and-error approach when new data are introduced (Mountrakis et al., 2011).

Concerning RF, we tested the sensitivity of *ntree* and *mtry* parameters to classification accuracy. As Figure 3-4-c shows, as *ntree* increases RF can reach higher classification accuracies, particularly as the number of trees increases from 50 to 500. Random variable selection is helpful to minimize the tree-based model's bias when a few variables are overused to develop the splitting nodes (Breiman, 2001). Regarding *mtry*, random 4 variables for a split achieves the best performance in each value of *ntree* compared to the other two testing values of *mtry*. The square root of the entire variable number has usually been used to obtain optimal classifier performance (Breiman, 2001; Ghosh et al., 2014; Rodriguez-Galiano et al., 2012). However, in this experiment, the accuracy only varies by 0.015% with different *ntree*. In fact, the number of trees does not significantly affect the performance of RF, as long as the number is sufficiently large. In other words, the accuracy tends to be stable once a sufficient number of trees has been reached (Maxwell et al., 2018). That 50 trees are sufficient to stabilize the classification accuracy has been reported in couple of recent studies (Ghimire et al., 2012; Shi and Yang, 2016), whereas Rodriguez-Galiano et al. (2012) demonstrate that 100 trees are needed. Belgiu and Drăgu (2016) list numerous articles applying RF with 5000 trees to achieve promising classification results. Hence, the optimal *ntree* is likely case specific. We therefore investigated six testing values for *ntree* that are commonly used in practice. Taking computing resources into account, *ntree* has been set to 500 to process large-scale data. In addition, the different values of *mtry* only resulted in 0.031% accuracy difference that, nonetheless, is larger than that of *ntree*. Similar findings have been reported by others whereby the classification accuracy has been shown to be more sensitive to *mtry* than to *ntree* (Du et al., 2015; Ghosh et al., 2014; Kulkarni and Sinha, 2012) . However, herein, changes in *mtry* still lead to a marginal difference in accuracy (< 0.05%). Hence, the impact of changes in these two hyperparameters was quite limited for RF.

With respect to GBT, the sensitivity of three hyperparameters (i.e. *mtry*, *ntree*, *lr*) was examined. As Table 3-2 shows, *ntree* has six testing values, and each of *mtry* and *lr* has three testing values, resulting in 54 hyperparameter combinations. The accuracy ranges from 95.01% to 98.62% over all combinations. *ntree* and *lr* led to 3.72% and 3.76% accuracy differences, respectively, whereas *mtry* resulted in a lower accuracy difference (1.78%). Hence, the accuracy was highly related to *ntree* and *lr*. Figure 3-4-d shows the accuracy as a function of *ntree* and *lr* at *mtry* of four. Under smaller *ntree* (< 1000), GBT with faster *lr* (0.1) can yield relatively more accurate classification results, whereas when using large *ntree* (2000, 3000), 0.1 *lr* was not the optimal parameter for GBT. Moreover, overall, GBT appears to require large *ntree* to produce high accuracy. Similar to the studies by Freeman et al. (2015) and Filippi et al. (2014), shrinking learning rate (lower *lr*) requires a higher iteration (*ntree*) to produce preferable classification using GBT; however, overlarge iterations (*ntree*) appear to result in overfitting under fast learning rates (Elith et al., 2008). Generally, compared to faster *lr* values, slower values can shrink the contribution of each tree more to help the classifier to produce reliable estimated responses. However, it is noticeable in our experiment that GBT with 0.05 *lr* rather than the lowest *lr* (0.01), under 2000 and 3000 *ntree*, was able to produce the best performance. Hence, the optimal pair of *lr* and *ntree* is flexible for different datasets.

3.3.3 Statistical and visual accuracy assessments

Figure 3-5 shows the accuracies achieved by the four classifiers with the 7-band configuration using a 100-fold CV. The four classifiers, overall, were able to produce high classification accuracies above 94%. MLR performs worst with 94.44% overall accuracy (OA), followed by the SVM with 96.99%. The two tree-based classifiers (RF and GBT) considerably outperform the two function-based classifiers (MLR and SVM) and reach comparable classification performances (RF: 98.57%; GBT: 98.62%). Based on boxplots, the results of RF and GBT show less variance of OA. Furthermore, the performance of the two tree-based classifiers is only strong when examining class-specific accuracies (Figure 3-5). All class accuracies (cloud, water, ice) produced by RF and GBT are above 98.00% and more stable with less than 0.30% accuracy difference compared to other classifiers. More specifically, MLR yielded a cloud detection accuracy of only 90.49% which is 7.92% lower than its water classification accuracy, resulting in a poorer OA. SVM provided a moderate ability in every aspect of the classification amongst the classifiers. Additionally, the quality of all classifiers is highest for the water class accuracy, which is above 98.00%.

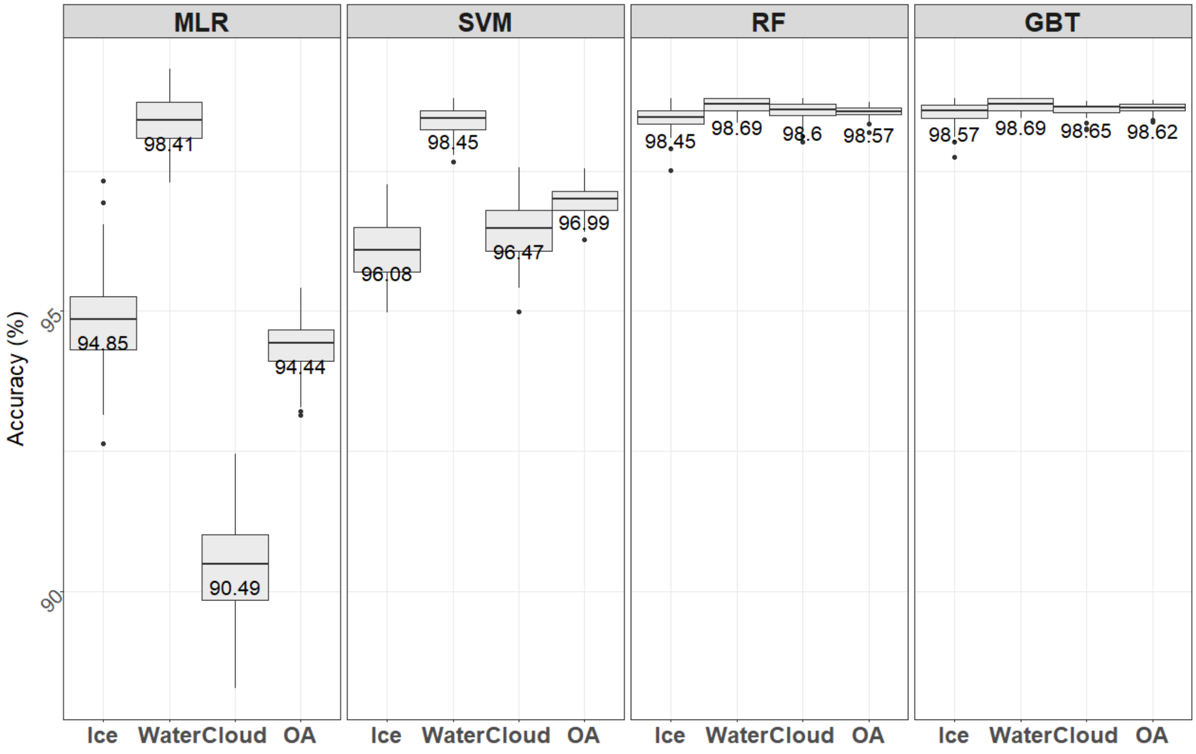


Figure 3-5 Comparison of accuracies (%) obtained using random 100-fold CV across classifiers for the ice, water and cloud classes individually, and overall (OA).

The visual assessment of the classification results confirms the findings from the above statistical assessment. Figure 3-6, Figure 3-7, and Figure 3-8 show examples of lake ice maps during the break-up period produced by the four classifiers. Overall, the two tree-based classifiers were able to discern more accurately lake ice cover compared to MLR, and SVM achieved a moderate performance. As shown in Figure 3-6, the classifiers can all detect accurately the majority of lake ice cover. However, MLR underestimates the ice edge, whereas RF, GBT, and SVM provide more precise ice edge delineation. Figure 3-7 illustrates a noticeable classification drawback of MLR during the advanced stage of ice melt. MLR misclassified decaying ice mostly as water, while the other three classifiers discriminated the two classes accurately according to the false color composite image. Another error source of MLR arises from the confusion between cloud and ice. As Figure 3-8 shows, thin cloud cover highlighted by the black rectangle was misclassified as ice by MLR, while RF and GBT could retrieve the cloudy area correctly. Despite a more accurate detection by SVM compared to MLR, it cannot delineate features with as good details in contrast to the two tree-based classifiers. For example, compared to the lake ice maps produced by RF and GBT, SVM slightly underestimates cloud cover in the area highlighted by the red rectangle. Moreover, in the same area, SVM cannot retrieve the ice edge and floe as accurately as the tree-based classifiers can. The visual inspection supports the moderate classification capability of SVM revealed in the statistical assessment. Figure 3-9 presents an example of lake ice mapping by the classifiers during the freeze-up period under high SZA (84-85 degrees). SVM and MLR were unable to classify cloud cover accurately under such illumination condition. MLR underestimates the black (snow-free) and grey ice highlighted by the black rectangle. In contrast, the tree-based classifiers perform quite well in detecting this ice under low solar illumination conditions.

During the break-up period, lake ice can appear from bright (white ice or snow-covered ice) to dark (snow-free black ice or surface melt/ponding on the ice surface). In the early melt stage, white ice with high reflectance in the visible spectrum is gradually exposed once snow cover has been melted (Jeffries et al., 2005). On the other hand, black ice appears darker in the visible bands against the water background (Jeffries et al., 2005). The four classifiers can all discriminate grey/white ice from open water accurately (Figure 3-6); however, MLR failed to detect black ice (Figure 3-7). The poorer classification performance of MLR is likely the result of overlap in the input variable space between features. The linear boundary of MLR was limited its discrimination of open water from black ice due to their similar VIS-NIR-SWIR reflectance. It also resulted in the confusion between ice and thin cloud (Figure 3-8). By contrast, the RBF kernel allowed the SVM to produce a non-linear separation plane via enlarging the input variable space; thus, the classifier can distinguish two features with similar spectral signatures at VIS-NIR-SWIR. In the case of RF and GBT, both bagging and boosting techniques are of help to tree-based classifiers in discriminating features with high spectral variability. In addition, random variable selection for the splitting nodes also improve the classification of the three features where the relationships between variables are complicated. Similar to the statistical assessment, RF and GBT produced comparable lake ice maps for different cases. The four classifiers performed comparably well for open water detection since water shows low and relatively stable reflectance at VIS-NIR-SWIR wavelengths. During the freeze-up period, mapping of lake ice can be problematic due to large solar zenith angles, especially in the Arctic. The reflectance of all surface types can be very low under such condition. Moreover, during that period, new ice often forms as black ice free of snow presence which can result in extremely low reflectance. For example, in Figure 3-9, the reflectance of ice in the NIR band is lower than 0.05, and cloud cover also shows low reflectance values. However, under this extreme case, RF and GBT were still able to provide a strong ability to detect ice.

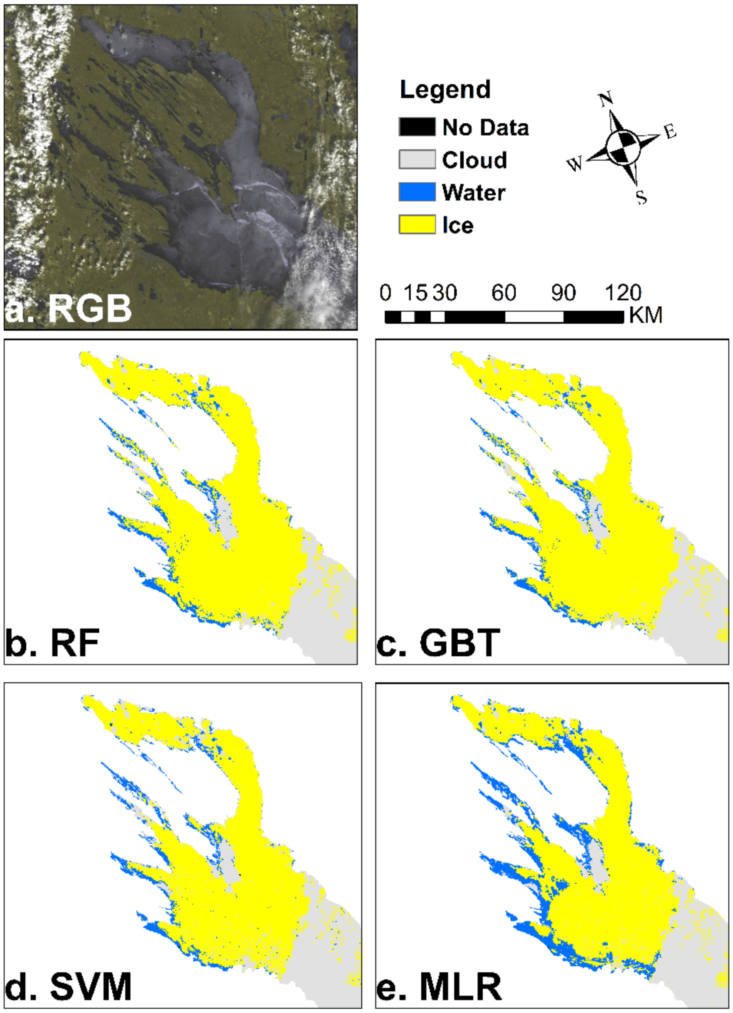


Figure 3-6 Ice maps of Lake Onega (Russia) during break-up (11 May 2003, UTC 09:25) produced by the four classifiers. (a) RGB false color composite, (b) RF, (c) GBT, (d) SVM, and (e) MLR.

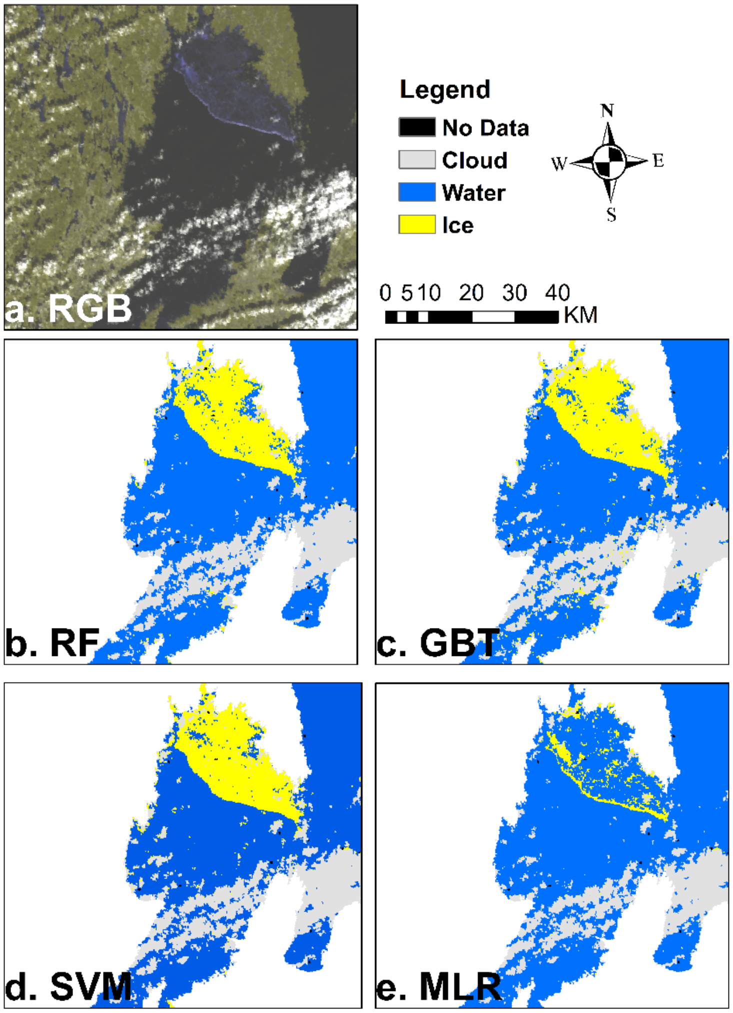


Figure 3-7 Ice maps of Lake Vänern (Sweden) during break-up (30 March 2003, UTC 10:30) produced by the four classifiers. (a) RGB false color composite, (b) RF, (c) GBT, (d) SVM, and (e) MLR.

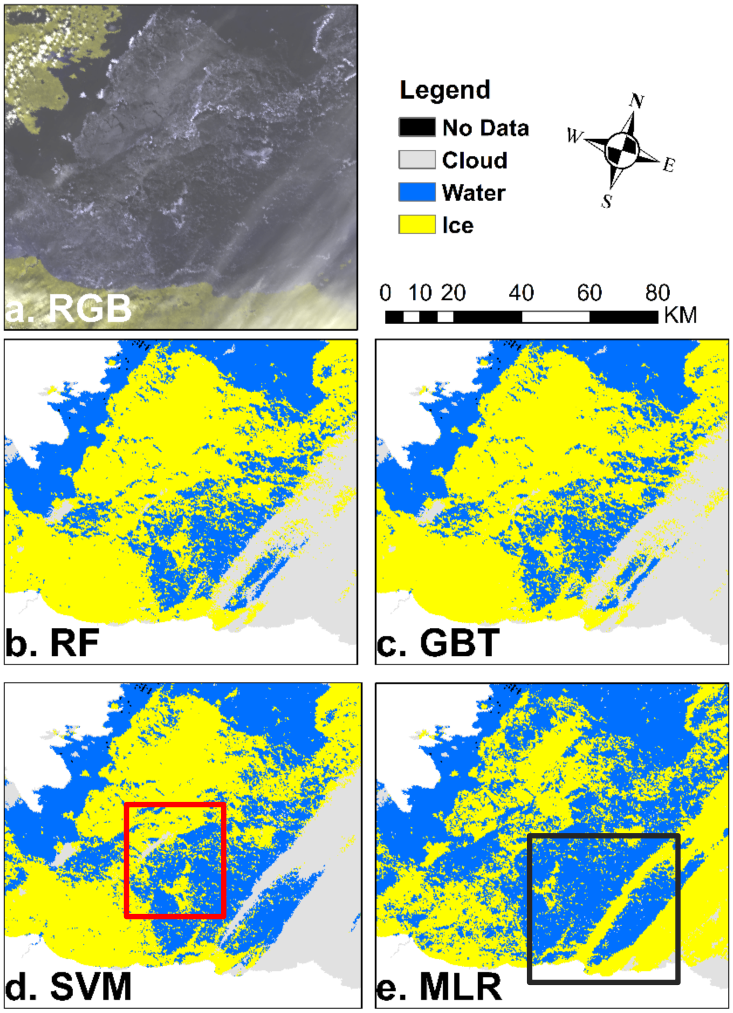


Figure 3-8 Ice maps of Great Slave Lake (Canada) during break-up (5 June 2003, UTC 19:15) produced by the four classifiers. (a) RGB false color composite, (b) RF, (c) GBT, (d) SVM, and (e) MLR.

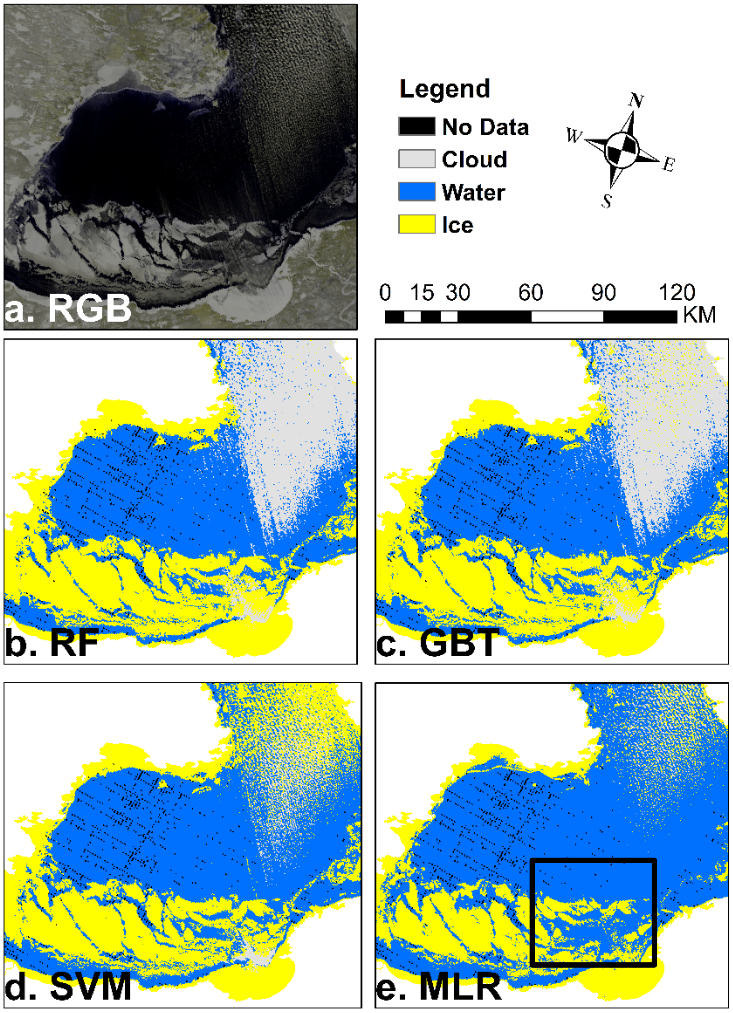


Figure 3-9 Ice maps of Great Slave Lake (Canada) during freeze-up (2 December 2009, UTC 18:50) produced by the four classifiers. (a) RGB false color composite, (b) RF, (c) GBT, (d) SVM, and (e) MLR. The black pixels correspond to no data (NaN value in input spectral bands).

3.3.4 Spatial and temporal transferability assessments

Table 3-5 summarizes the accuracy assessment with regards to spatial transferability of the four classifiers. RF (mean accuracy (MA) = 95.64%, standard deviation (SD) = 0.0447) and GBT (MA = 95.26 %, SD = 0.0478) appear to be the most robust classifiers in terms of the spatial transferability, followed by MLR (MA = 90.98%, SD = 0.0861) and SVM (MA = 79.36%, SD = 0.1244). RF and GBT achieved the best performance for six and four spatial lake clusters, respectively, and mostly produced above 95% classification accuracy over all clusters. The accuracy of SVM ranges from 63.17% to 98.03% with 0.1244 SD, and this classifier yielded below 80% accuracy in half of all lake clusters. Thus, SVM particularly suffered from the spatial variation of training and testing data.

A performance pattern similar to that of spatial transferability was also obtained with the temporal transferability evaluation (Table 3-6). RF and GBT provided above 95% MA and a stable accuracy (SD less than 0.02) over the three ice seasons examined, resulting in less variance of the accuracy. SVM performed the worst with 83.00% MA and 0.0312 SD. MLR achieved a moderate performance of 93.21% MA and 0.0227 SD. Hence, RF and GBT provide the most stable classification behavior over ice seasons.

Table 3-5 Accuracy assessment using spatial CV for lake clusters across classifiers. MA: mean accuracy, SD: standard deviation. The maximum accuracy in each cluster is bold.

Clusters	MLR	SVM	RF	GBT
AN	97.07%	85.03%	**99.70%**	99.29%
Ath	69.54%	63.17%	**91.20%**	86.44%
Bai	96.33%	72.70%	95.66%	**97.06%**
GBL	96.46%	98.03%	**99.60%**	99.02%
GLs	97.94%	64.54%	98.54%	**98.62%**
GSL	80.36%	64.80%	93.92%	**94.22%**
Ian	92.62%	77.86%	**99.70%**	98.80%
OL	93.37%	90.54%	**97.13%**	96.45%
Tay	**90.74%**	90.07%	85.20%	86.44%
Van	91.45%	92.01%	**93.57%**	93.31%
Win	94.94%	74.20%	97.79%	**98.22%**
MA	**90.98%**	**79.36%**	**95.64%**	**95.26%**
SD	**0.0861**	**0.1244**	**0.0447**	**0.0478**

Table 3-6 Accuracy assessment using temporal CV in the clusters across classifiers. MA: mean accuracy, SD: standard deviation. The maximum accuracy in each ice year is bold.

Ice Year	MLR	SVM	RF	GBT
2002-2003	91.85%	81.31%	**93.91%**	93.37%
2009-2010	95.83%	86.60%	96.36%	**96.40%**
2016-2017	91.95%	81.10%	96.19%	**95.67%**
MA	**93.21%**	**83.00%**	**95.49%**	**95.15%**
SD	**0.0227**	**0.0312**	**0.0137**	**0.0158**

For lake ice mapping over large spatial and long temporal scales, ML models should be able to handle spatiotemporal heterogeneity in satellite datasets. Our results are in general agreement with previous studies (Micheletti et al., 2014; Ruß and Brenning, 2010) to the effect that ML models do not perform as well when applying spatial and temporal CV compared to using random k-fold CV alone. The two tree-based classifiers (RF and GBT) were able to produce mean classification accuracies above 95% in both spatial and temporal CV; a ca. 3% decrease in performance from random k-fold CV. MA produced by SVM with the spatial and temporal CV, in contrast, dropped by more than 13% compared to the results using random k-fold CV. Thus, SVM is considerably sensitive to the spatial and temporal variations of the training and validation data. In a study on the mapping of maximum air temperature from MODIS, RF was found to slightly outperform SVM with respect to the spatial CV (Ho et al., 2014). Findings from our study indeed suggest that tree-based classifiers are better candidates than SVM and MLR for global lake ice mapping from relatively long historical data records such as the one available from MODIS Terra (2000-present).

3.4 Conclusion

We conducted a comprehensive assessment of four machine learning algorithms in multiclass feature (i.e. ice, water, and cloud) extraction of large northern lakes using MODIS Terra L1B TOA data. Results from k-fold CV reveal that MLR is the least promising classifier, particularly for cloud and ice cover compared to SVM, RF and GBT. SVM, on the other hand, is found to be less consistent in terms of spatial and temporal transferability; its spatial (79%) and temporal (83%) CV accuracies differ greatly from the random k-fold CV (97%) accuracy. Moreover, SVM is over-sensitive to the change of hyperparameter sets. RF and GBT did better than MLR and SVM in all aspects of the study. The two tree-based classifiers performed similarly in terms of overall and class specific accuracies as well as in spatiotemporal transferability; however, they showed differences in two aspects. First, compared to GBT, the performance of RF is less sensitive to the choice of the hyperparameters. However, changes of input band configurations (three to 10 bands) were not expressively influential on GBT results. GBT could potentially be applied to optical data from other satellite platforms with less spectral bands, as revealed from its overall performance of 96% with only three MODIS bands, and high spatial resolutions for lake ice mapping on smaller lakes (e.g. Landsat, Sentinel-2).

Overall, results show the strong potential of RF for global lake ice mapping using TOA reflectance data from MODIS and other satellite platforms that offer similar band configurations such as Sentinel-3 (OLCI/SLSTR synergy).

Chapter 4

General Conclusion

4.1 Summary

To date, remote sensing is the most efficient and reliable means to achieve lake ice observations, which are useful for many biological, ecological and socio-economic applications. MODIS imagery has become indispensable data for the development of a global lake ice network across long time-series. However, previous studies applying threshold-based algorithms in lake ice classification from MODIS products have face some difficulties in correctly identifying ice, particularly under high solar zenith angles, certain cloud cover conditions and clear (black) ice. While machine learning (ML) techniques have recently been employed for sea ice classification and microwave remote sensing applications, ML has not yet been applied in lake ice classification from optical remote sensing data. Hence, the overall objective of this study was to exploit the capability of ML algorithms to enhance lake ice classification using the MODIS Level 1B product.

Chapter 3 presented a comparative study of the performance of four ML classifiers in lake ice classification from MODIS L1B imagery. Overall, all four algorithms produced above 94% accuracy classification. According to the visual examination, Random Forest (RF) and Gradient Boosting Tree (GBT) can overcome a number of challenges (i.e. black ice, high solar zenith angles), thereby performing satisfactory classification. Despite showing comparable performance with RF and GBT in terms of the visual assessment, Support Vector Machine (SVM) is too sensitive to the parameterization to conduct large-scale lake ice cover mapping. Conversely, RF was the most insensitive to the change of the hyperparameters. The experiment results also showed that Multinomial Logistic Regression (MLR) was less powerful for providing accurate classification compared to the other three classifiers. For the transferability examination, RF and GBT outperformed other two classifiers. The two function-based classifiers (MLR and SVM) generate a separating boundary to classify remote sensing observations as a label. However, according to the results, the separating boundary shows weak classification capability of lake ice (MLR) and unstable transferability (SVM) in spatial and temporal aspects. On the contrary, in the case of RF and GBT, the two ensemble techniques (bagging and boosting) are of help to tree-based classifiers in discriminating features with high spectral variability and coping with the spatial and temporal heterogeneity of remote sensing data. Additionally, the MODIS reflectance bands at Visible-NIR-SWIR wavelengths were found to be the most useful and efficient input variable combination for the four algorithms.

In summary, this research demonstrated the potential of ML algorithms to perform lake ice classification at a global scale across a long time-series. The tree-based classifiers (RF and GBT) provided the most promising results.

4.2 Limitations and Recommendations for Future Work

The primary limitation of this research is the limited study dataset. This research focused on lake ice classification during the ice freeze-up and break-up periods. However, the performance of the classifiers during the completely ice-free (open water) and full-ice periods was not investigated. Due to the lack of "training" samples over these two periods, the classifiers may fail under some severe cases. For example, high algal bloom concentration and extreme high sediment (even almost dried-out lakes) during the ice-free season (summer), could result in extreme high reflectance of open water, which is homogenous to the ice reflectance. Additionally, during the complete freeze over period, extremely high TOA values over snow/ice might be misclassified as cloud cover. Besides, the study lakes in this research are large at the global scale (the minimum is 1,040 km^2). To determine whether the classifiers are as promising for the mapping of lake ice on smaller lakes than the one investigated herein, more studies are needed. Therefore, in follow-up investigations, a larger sample from the two periods (ice-free and fully ice covered) and small-scale lakes would need to be collected and input into the classifiers to examine their performance.

Although the tree-based classifiers produced above 98% classification accuracy using the random K-fold CV, they still yielded lower accuracy (about 95%) via the temporal and spatial CV strategies. The accuracy degradation indicates the effect of temporal and spatial variation on classifier performance. Recent studies of transfer learning for remote sensing have proposed the use of domain adaptation (Bruzzone and Marconcini, 2009; Matasci et al., 2015), covariate shift (Persello and Bruzzone, 2014) or multi-task learning (Leiva-Murillo et al., 2013) techniques to tackle the variation problems to enhance the transferability of pixel-based classifiers. In future investigations, it is recommended that these techniques would be applied alongside the classifiers to enhance the performance of lake ice cover mapping.

Additionally, since the MODIS Terra sensor has missing acquisitions/detections from time to time, the input band configuration of classifiers lacks complete reflectance values for the classification. Therefore, the produced lake ice cover maps present absence of labels (NoData). Interpolation approaches could be applied to tackle this issue so that the produced maps would more comprehensively capture lake ice spatial conditions.

Recently, convolutional neural network (CNN), which is a state-of-the-art approach, has been shown to be successful for sea ice classification and the estimation of ice concentration from SAR remote sensing imagery (Wang et al. , 2016; Zhang et al., 2019). The retrieval of ice cover from optical remote sensing imagery using a CNN has been rarely performed. In particular, until now, no previous study had applied CNN in lake ice classification from MODIS imagery. A case study of lake ice classification from MODIS TOA reflectance imagery using a CNN model in Great Slave Lake, Canada, was presented in Appendix A. The proposed CNN produced a 98.03% accuracy with the testing dataset; however, accuracy dropped to 90.13% using an independent (out-of-sample) validation dataset. Overfitting apparently occurred in this case. A collection of a larger sample from more years and other lake sites would improve the representativeness of training data. Meanwhile, some advanced techniques of label-preserving transformations, such as rotation and flipping, could be employed to enlarge training sample augmentation artificially. Additionally, a more comprehensive study and evaluation of other CNN architectures and configurations, such as the number of layers, kernel size, and patch size, would be beneficial.

Appendix A.

Lake Ice Classification from MODIS TOA Reflectance Imagery Using A Convolutional Neural Network: A Case Study of Great Slave Lake, Canada

I. Introduction

Lake ice cover is highly responsive to changes in weather and climate as shown in (Duguay et al., 2003, 2006). Ice phenology dates associated with freeze-up and break-up, and ice duration are useful for assessing trends and variability in climate, particularly due to their sensitivity to changes in near-surface air temperature (Howell et al., 2009; Kang et al., 2012). Lake ice mapping from satellite remote sensing data allows for the investigation of ice phenology over large areas, and provides an alternative for filling gaps of sparse ground-based networks of lake ice observations globally (Duguay et al., 2006). The Moderate Resolution Imaging Spectroradiometer (MODIS), aboard NASA's Terra and Aqua satellite platforms, provides a continuous stream of daily Earth surface records available for the monitoring of lake ice dynamics. However, it remains challenging to perform lake ice-water classification well using existing knowledge-driven (threshold-based) retrieval algorithms that use top-of-the-atmosphere (TOA) reflectance data, particularly under the condition of large solar zenith angles resulting in low TOA reflectance. Recently, convolutional neural network (CNN), which is a state-of-the-art approach, has been shown to be successful for sea ice classification and the estimation of ice concentration from SAR remote sensing imagery (Wang et al., 2016; Zhang et al., 2019). A CNN is comprised of a stack of alternating convolution layers and pooling layers, followed by a number of fully connected layers. Compared to traditional neural networks, CNN enforces weight sharing and local connectivity between layers (Lecun et al., 1998). The retrieval of lake ice cover from MODIS TOA reflectance imagery using a CNN has not been yet performed. Hence, this study designed a CNN architecture applied to lake ice cover mapping using Great Slave Lake, Canada, as a case study.

II. Study area and data

This research was carried at Great Slave Lake (GSL). The lake is located in the Northwest Territories, Canada, within the Mackenzie River Basin (Figure A-1). GSL has a surface area of 28.6 × 10^3 km^2. It is bounded in the east-west direction by longitudes 108° and 116° W, and lies between 63° to 67° N in the south-north direction.

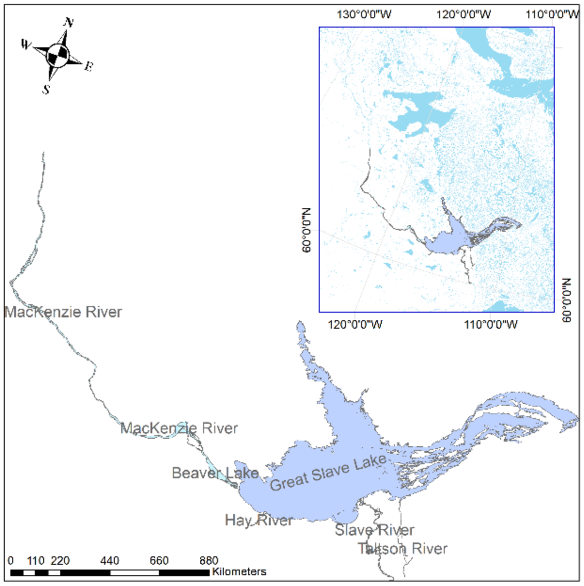


Figure A-1 The location of Great Slave Lake, Canada

The MODIS instrument onboard Terra, launched in 1999, has been delivering data since 2000. The sensor scans the majority of the entire Earth's surface every day, recording observations in 36 spectral bands from visible to thermal infrared wavelengths. The MODIS Level 1B product (MOD02) provides top-of-the-atmosphere (TOA) reflectance data stored in three separate files based on spatial resolution; MOD02QKM (250m: bands 1-2), MOD02HKM (500m: bands: 3-7), and MOD021KM (1km: bands 8-36).

In this study, five images during the freeze-up period and 13 images from the break-up period were acquired for GSL over the ice season of 2009-2010. Sample pixels were collected manually from MODIS false color composite images (R: Band 2; G: Band 2; B: Band 1). The sample classes were labelled as lake ice, open water, and cloud cover. We then separated the 18 images into two groups. Group A (13 images) was used for CNN model training and testing. The samples of group A were randomly split into training data (70%) and testing data (30%). Group B (five images) was used for validation.

III. Methodology

i. Preprocessing

The MOD02 product records Earth observations in a 5-min orbital swath format including latitude and longitude without projection. Thus, the images were projected into an equirectangular projection. All MOD02 bands were resampled onto this grid using the nearest neighbor method.

Since the input of CNN is an image patch, we used each sample pixel as a central point, generating a three-dimensional matrix of n × 11 × 11 where n corresponds to the number of input spectral bands. The width and height of patches are 11 pixels. Each patch was labelled based on the class of the central sampled pixel.

ii. Input band configurations

In order to investigate the optimal input bands, we tested four band configurations as shown in Table A-1.

Table A-1 MODIS band configuration.

Configuration	Spectral bands
3-band	Red, NIR, SZA
5-band	Red, NIR, Green, Blue, SZA
7-band	Red, NIR, Green, Blue, SWIR2, SWIR3, SZA
10-band	Red, NIR, Green, Blue, SWIR2, SWIR3, Band20, Band31, Band32, SZA

All configurations include solar zenith angle (SZA) to address low reflectance, particularly relevant during the freeze-up period in fall at this high-latitude lake location. In addition to SZA, the 3-band configuration contained the red and near infrared (NIR) bands that provide observations at the highest spatial resolution (250 m) in the MOD02 product. Moreover, the NIR band has been applied to retrieve ice frequently since the surface reflectance of ice and snow-covered lake ice is relatively high in the NIR (Brown and Duguay, 2012; Nonaka et al., 2007; Šmejkalová et al., 2016). The 5-band configuration additionally includes MODIS green and blue bands, which are helpful in identifying open water with high sediment concentration in suspension. Two shortwave infrared bands at 500 m resolution were introduced to the 7-band configuration. Additionally, the 10-band configuration includes three more thermal infrared (TIR) bands.

iii. CNN Architecture

The architecture of the CNN used this study is shown in Table A-2. It consists of five convolutional layers (Conv), two max-pooling layers (Pool), and seven fully connected layers (FC). The input patch is a 3-D matrix of n × 11 × 11. In convolutional layers, a number of 2-D filters (kernels) of size 3×3 are applied to the input patch, producing intermediate image patches which are processed in the next layer. Rectified Linear Unit (ReLU) is used as the activation function in the convolutional layers. ReLU can enhance nonlinearity by reassigning zero to negative values outputted by the previous layer (Gonzalez, 2007). The max-pooling layers subsample the intermediate image patches by 2×2 windows, thus selecting the maximum value in the windows. The fully connected layer, which functions is identical to the basic neural network, has a 1-D vector of neurons. Each neuron in a fully connected layer is linked to all the neurons of its preceding layer. The last fully connected layer using the softmax function computes the probabilities of the three classes. The workflow of the proposed CNN was implemented using Pytroch with NVIDIA T4 GPUs in Python 3.7.

Table A-2 Architecture of the proposed CNN.

Layer	Structure
Input patch	n × 11 × 11
Conv1	64 × n × 3 × 3 stride 1, pad 1, ReLU 64 × 11 × 11
Conv2	64 × n × 3 × 3 stride 1, pad 1, ReLU 64 × 11 × 11
Pool1	2 × 2 stride 2, pad 0, Max 64 × 5 × 5
Conv3	128 × 64 × 3 × 3 stride 1, pad 1, ReLU 12 × 5 × 5
Conv4	128 × 64 × 3 × 3 stride 1, pad 1, ReLU 128 × 5 × 5
Conv5	256 × 128 × 3 × 3 stride 1, pad 1, ReLU 256 × 5 × 5
Pool2	2 × 2 stride 2, pad 0, Max 256 × 2 × 2
FC1	512 × 256 × 2 × 2, Linear 512 × 1
FC2	256 × 512, Linear 256 × 1
FC3	128 × 256, Linear 128 × 1
FC4	64 × 128, Linear 64 × 1
FC5	48 × 64, Linear 48 × 1
FC6	16 × 48, Linear 16 × 1
FC7	16 × 512, Softmax 3 × 1

Note: Each row for a given convolutional or pooling layer corresponds to: the kernel configuration (top row), the layer configuration (middle row) and the dimension of the output (bottom row). (e.g. there are 64 filters of n × 3 × 3 in Conv1 that used the input patch of n × 11 × 11 with a stride of 1 and a pad of 2; the dimension of the output is 64 × 11 × 11)

IV. Results and discussion

i. Band Configuration Comparison

As shown in Table A-3, the testing accuracy improves along with an increase in the number of input bands used.

Table A-3 Testing accuracy of band configurations.

Configuration	Testing accuracy
3-band	90.23 %
5-band	93.64 %
7-band	98.03 %
10-band	98.18 %

The 3-band configuration produced the lowest (testing) classification accuracy at 90.23%. The accuracy increases by 3.41% using the 5-band configuration compared to the 3-band configuration. Furthermore, the accuracy is improved significantly from the 5-band configuration to the 7-band configuration. Applying the 10-band configuration on the proposed CNN led to a result comparable to the 7-band configuration. Therefore, for assessing the transferability of the proposed CNN to an independent set of MODIS images, the study only applied the 7-band configuration (i.e. without including TIR bands) for validation.

ii. Testing and Validation Accuracy Comparison

Classification results for two examples from the validation dataset are shown in Figure A-2 for the break-up and freeze-up periods, respectively. It can be seen that the performance of the proposed CNN is visually accurate overall. In both figures, the majority of ice and water pixels were retrieved correctly. The CNN produced a spatially smooth distribution of water-ice, which is reasonable since CNN with deep networks extracts more abstract texture features resulting in predictions that are less sensitive to pixel-based spectral values.

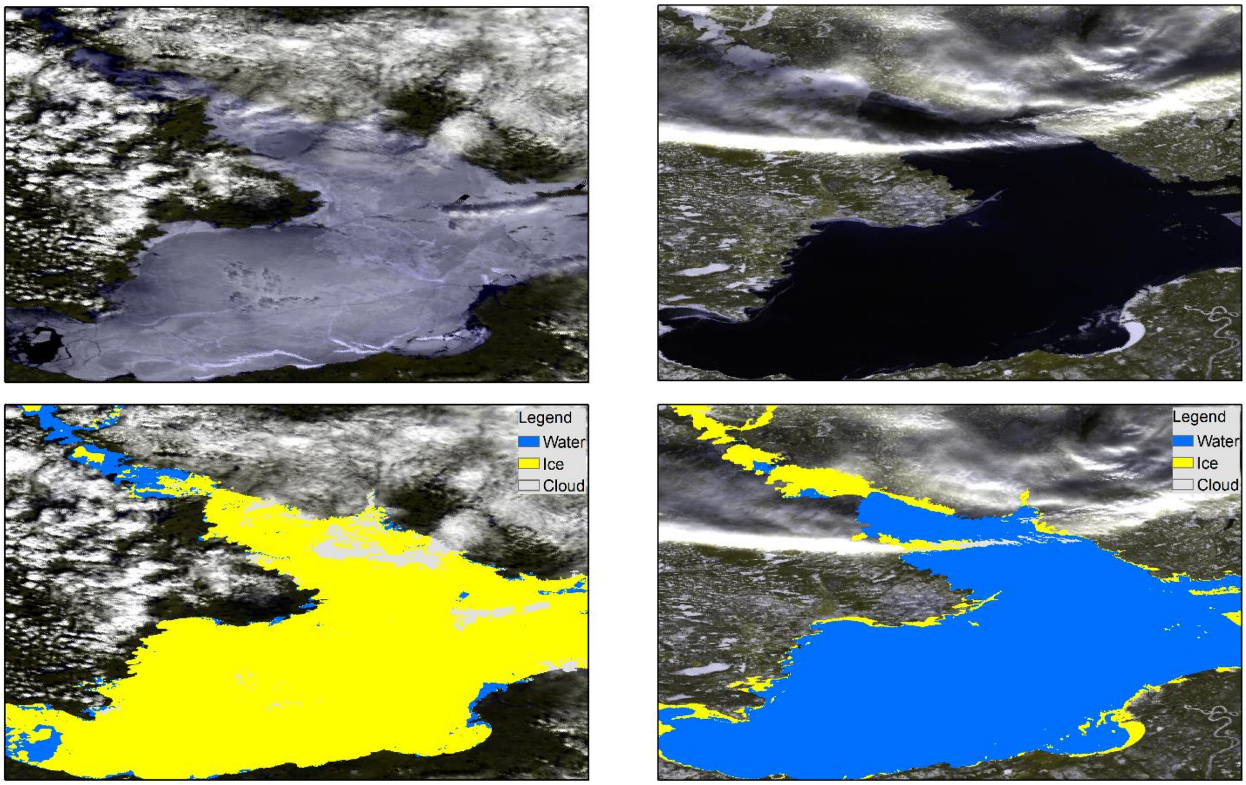


Figure A-2 Lake ice cover maps produced by the processed CNN. Left: Example during break-up period (15 May 2010, UTC 2000; top: RGB composite image from MOD02 product bands 1 and 2; bottom: Lake ice map from CNN); Right: Example during freeze-up period (15 November 2009, UTC 1945; top: RGB composite image from MOD02 product bands 1 and 2; bottom: Lake ice map from CNN).

However, the accuracy of the validation dataset produced by the trained CNN with the 7-band configuration is 90.13%, which is considerably less than the 98.03% testing accuracy. As seen in Figure A-2 left, the model somewhat underestimated thin cloud cover. Likewise, in Figure A-2 right (close to the top section of the lake), the edge of cloud cover was misclassified as ice and the cloud shallow over ice was misclassified as water.

The different results between the testing and validation data indicate the overfitting of the CNN in the two datasets. Nevertheless, the high testing accuracy demonstrates the powerful learning and classification capability of the CNN given that the training and testing data are independent and identically distributed. Hence, the variation between testing and validation data makes the CNN prone to overfitting.

V. Conclusions and Future Work

This research investigated the classification performance of a CNN in mapping lake ice cover from MODIS TOA reflectance imagery. The input variable configuration of SZA, visible, near-infrared and shortwave infrared bands produced optimal classification results. The proposed CNN model performed well when assessed with the testing dataset with 98.03% accuracy; accuracy which dropped to 90.13% using an independent (out-of-sample) validation dataset.

To further reduce overfitting, we plan to collect a larger sample from more years and other lake sites as to increase the diversity of the training data. On the other hand, performing label-preserving transformations, such as rotation and flipping, enlarges training sample augmentation artificially. Additionally, a more comprehensive study and evaluation of other CNN architectures and configurations, such as the number of layers, kernel size, and patch size, is required.

References

Barbieux, K., Charitsi, A., & Merminod, B. (2018). Icy lakes extraction and water-ice classification using Landsat 8 OLI multispectral data. *International Journal of Remote Sensing*, *39*(11), 3646–3678. https://doi.org/10.1080/01431161.2018.1447165

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: a review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Benson, B. J., Magnuson, J. J., Jensen, O. P., Card, V. M., Hodgkins, G., Korhonen, J., … Granin, N. G. (2012). Extreme events, trends, and variability in Northern Hemisphere lake-ice phenology (1855-2005). *Climatic Change*, *112*(2), 299–323. https://doi.org/10.1007/s10584-011-0212-8

Blenckner, T., Järvinen, M., & Weyhenmeyer, G. A. (2004). Atmospheric circulation and its impact on ice phenology in Scandinavia. *Boreal Environment Research*, *9*(5), 371–380.

Bolsenga, S. J. (1969). Total albedo of Great Lakes ice. *Water Resources Research*, *5*(5), 1132–1133. https://doi.org/10.1029/WR005i005p01132

Bolsenga, S. J. (1977). Short note: preliminary observations on the daily variation of ice albedo. *Journal of Glaciology*, *18*(80), 517–521. https://doi.org/10.3189/s0022143000021171

Bolsenga, S. J. (1981). Radiation transmittance through lake ice in the 400-700m range. *Journal of Glaciology*, *27*(95), 57–66. https://doi.org/10.3189/s0022143000011229

Bolsenga, S. J. (1983). Spectral reflectances of snow and fresh-water ice from 340 through 1 100 nm. *Journal of Glaciology*, *29*(102), 296–305. https://doi.org/10.1017/S0022143000008352

Bonsal, B. R., Prowse, T. D., Duguay, C. R., & Lacroix, M. P. (2006). Impacts of large-scale teleconnections on freshwater-ice break/freeze-up dates over Canada. *Journal of Hydrology*, *330*(1–2), 340–353. https://doi.org/10.1016/J.JHYDROL.2006.03.022

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

https://doi.org/10.1023/A:1010933404324

Brown, L. C., & Duguay, C. R. (2010). The response and role of ice cover in lake-climate interactions. *Progress in Physical Geography*, *34*(5), 671–704. https://doi.org/10.1177/0309133310375653

Brown, L. C., & Duguay, C. R. (2011). The fate of lake ice in the North American Arctic. *Cryosphere*, *5*(4), 869–892. https://doi.org/10.5194/tc-5-869-2011

Brown, L. C., & Duguay, C. R. (2012). Modelling lake ice phenology with an examination of satellite-detected subgrid cell variability. *Advances in Meteorology*, *2012*, 1–19. https://doi.org/10.1155/2012/529064

Bruzzone, L., & Marconcini, M. (2009). Toward the automatic updating of land-cover maps by a domain-adaptation svm classifier and a circular validation strategy. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(4), 1108–1122. https://doi.org/10.1109/TGRS.2008.2007741

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167. https://doi.org/10.1023/A:1009715923555

Cai, Y., Ke, C. Q., Li, X., Zhang, G., Duan, Z., & Lee, H. (2019). Variations of lake ice phenology on the tibetan plateau from 2001 to 2017 based on modis data. *Journal of Geophysical Research: Atmospheres*, *124*(2), 825–843. https://doi.org/10.1029/2018JD028993

Cai, Y., Ke, C. Q., Yao, G., & Shen, X. (2019). MODIS-observed variations of lake ice phenology in Xinjiang, China. *Climatic Change*. https://doi.org/10.1007/s10584-019-02623-2

Chai, D., Newsam, S., Zhang, H. K., Qiu, Y., & Huang, J. (2019). Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. *Remote Sensing of Environment*, *225*(September 2018), 307–316. https://doi.org/10.1016/j.rse.2019.03.007

Chen, N., Li, W., Gatebe, C., Tanikawa, T., Hori, M., Shimada, R., … Stamnes, K. (2018). New neural network cloud mask algorithm based on radiative transfer simulations. *Remote Sensing of Environment*, *219*, 62–71. https://doi.org/10.1016/j.rse.2018.09.029

Choiński, A., Ptak, M., Skowron, R., & Strzelczak, A. (2015). Changes in ice phenology on polish lakes from 1961 to 2010 related to location and morphometry. *Limnologica*, *53*, 42–49. https://doi.org/10.1016/J.LIMNO.2015.05.005

Coakley, J. A. (2003). Reflectance and albedo, surface. In *Encyclopedia of Atmospheric Sciences* (pp. 1914–1923). Elsevier. https://doi.org/10.1016/B0-12-227090-8/00069-5

Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers and Geosciences*, *63*, 22–33. https://doi.org/10.1016/j.cageo.2013.10.008

Dibike, Y., Prowse, T., Shrestha, R., & Ahmed, R. (2012). Observed trends and future projections of precipitation and air temperature in the Lake Winnipeg watershed. *Journal of Great Lakes Research*, *38*, 72–82. https://doi.org/10.1016/j.jglr.2011.04.005

Dorofy, P., Nazari, R., Romanov, P., & Key, J. (2016). Development of a mid-infrared sea and lake ice index (MISI) using the GOES imager. *Remote Sensing*, *8*(12). https://doi.org/10.3390/rs8121015

Du, J., Kimball, J. S., Duguay, C. R., Kim, Y., & Watts, J. D. (2017). Satellite microwave assessment of Northern Hemisphere lake ice phenology from 2002 to 2015. *Cryosphere*, *11*(1), 47–63. https://doi.org/10.5194/tc-11-47-2017

Du, P., Samat, A., Waske, B., Liu, S., & Li, Z. (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 38–53. https://doi.org/10.1016/j.isprsjprs.2015.03.002

Duguay, C. R., Bernier, M., Gauthier, Y., & Kouraev, A. (2015). Remote sensing of lake and river ice. *Remote Sensing of the Cryosphere*, (2015), 273–306.

https://doi.org/10.1002/9781118368909.ch12

Duguay, C. R., Brown, L. C., Kang, K. K., & Kheyrollah Pour, H. (2015). Lake ice. In J. Blunden & D. S. Arndt (Eds.), *State of the Climate in 2014* (Vol. 96, pp. S144–S145). Bulletin of the American Meteorological Society.

Duguay, C. R., Flato, G. M., Jeffries, M. O., Ménard, P., Morris, K., & Rouse, W. R. (2003). Ice-cover variability on shallow lakes at high latitudes: Model simulations and observations. *Hydrological Processes*, *17*(17), 3465–3483. https://doi.org/10.1002/hyp.1394

Duguay, C. R., Prowse, T. D., Bonsal, B. R., Brown, R. D., Lacroix, M. P., & Ménard, P. (2006). Recent trends in Canadian lake ice cover. *Hydrological Processes*, *20*(4), 781–801. https://doi.org/10.1002/hyp.6131

Duguay, C. R., Pultz, T. J., Lafleur, P. M., & Drai, D. (2002). RADARSAT backscatter characteristics of ice growing on shallow sub-Arctic lakes, Churchill, Manitoba, Canada. *Hydrological Processes*, *16*(8), 1631–1644. https://doi.org/10.1002/hyp.1026

Dyck, M. G. (2007). Community monitoring of environmental change: college-based limnological studies at Crazy Lake (Tasirluk), Nunavut. *Arctic*, *60*(1), 55–61. https://doi.org/10.14430/arctic265

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Filippi, A. M., Güneralp, İ., & Randall, J. (2014). Hyperspectral remote sensing of aboveground biomass on a river meander bend using multivariate adaptive regression splines and stochastic gradient boosting. *Remote Sensing Letters*, *5*(5), 432–441. https://doi.org/10.1080/2150704X.2014.915070

Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(6), 1335–1343. https://doi.org/10.1109/TGRS.2004.827257

Freeman, E. A., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2015). Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Canadian Journal of Forest Research*, *46*(3), 323–339. https://doi.org/10.1139/cjfr-2014-0562

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *International Conference on Machine Learning* (pp. 148–156). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.6252

Futter, M. N. (2003). Patterns and trends in southern ontario lake ice phenology. *Environmental Monitoring and Assessment*, *88*(1), 431–444. https://doi.org/10.1023/A:1025549913965

Gasch, C. K., Hengl, T., Gräler, B., Meyer, H., Magney, T. S., & Brown, D. J. (2015). Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. *Spatial Statistics*, *14*, 70–90. https://doi.org/10.1016/j.spasta.2015.04.001

GCOS. (2016). *The global observing system for climate: implementation needs*. Geneva. GCOS-200, 325 pp, https://doi.org/10.13140/RG.2.2.23178.26566

George, D. G. (2007). The impact of the North Atlantic Oscillation on the development of ice on Lake Windermere. *Climatic Change*, *81*(3), 455–468. https://doi.org/10.1007/s10584-006-9115-5

Ghimire, B., Rogan, J., Galiano, V., Panday, P., & Neeti, N. (2012). An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. *GIScience and Remote Sensing*, *49*(5), 623–643. https://doi.org/10.2747/1548-1603.49.5.623

Ghosh, A., Sharma, R., & Joshi, P. K. (2014). Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion. *Applied Geography*, *48*, 31–41. https://doi.org/10.1016/j.apgeog.2014.01.003

Gonzalez, T. F. (2007). ImageNet classification with deep convolutional neural networks.

Handbook of Approximation Algorithms and Metaheuristics, 1–1432. https://doi.org/10.1201/9781420010749

Gou, P., Ye, Q., Che, T., Feng, Q., Ding, B., Lin, C., & Zong, J. (2017). Lake ice phenology of Nam Co, Central Tibetan Plateau, China, derived from multiple MODIS data products. *Journal of Great Lakes Research*, *43*(6), 989–998. https://doi.org/10.1016/j.jglr.2017.08.011

Greene, S., Walter Anthony, K. M., Archer, D., Sepulveda-Jauregui, A., & Martinez-Cruz, K. (2014). Modeling the impediment of methane ebullition bubbles by seasonal lake ice. *Biogeosciences*, *11*(23), 6791–6811. https://doi.org/10.5194/bg-11-6791-2014

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(7–8), 1157–1182. https://doi.org/10.1162/153244303322753616

Hagolle, O., Huc, M., Pascual, D. V., & Dedieu, G. (2010). A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, *114*(8), 1747–1755. https://doi.org/10.1016/j.rse.2010.03.002

Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, *21*(12), 1534–1547. https://doi.org/10.1002/hyp.6715

Han, H., Hong, S. H., Kim, H. cheol, Chae, T. B., & Choi, H. J. (2017). A study of the feasibility of using KOMPSAT-5 SAR data to map sea ice in the Chukchi Sea in late summer. *Remote Sensing Letters*, *8*(5), 468–477. https://doi.org/10.1080/2150704X.2017.1285501

Han, H., Im, J., Kim, M., Sim, S., Kim, J., Kim, D. J., & Kang, S. H. (2016). Retrieval of melt ponds on arctic multiyear sea ice in summer from TerraSAR-X dual-polarization data using machine learning approaches: A case study in the Chukchi Sea with mid-incidence angle data. *Remote Sensing*, *8*(1). https://doi.org/10.3390/rs8010057

Han, Y., Li, P., Zhang, Y., Hong, Z., Liu, K., & Wang, J. (2018). Combining active learning and transductive support vector machines for sea ice detection. *Journal of Applied Remote Sensing*, *12*(02), 1. https://doi.org/10.1117/1.jrs.12.026016

Hand, D. J. (1997). *Construction and Assessment of Classification Rules*. Wiley. Retrieved from https://books.google.ca/books?id=NRfvAAAAMAAJ

Heron, R., & Woo, M.-K. (1994). Decay of a high Arctic lake-ice cover: observations and modelling. *Journal of Glaciology*, *40*(135), 283–292. https://doi.org/10.1017/S0022143000007371

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., & Henderson, S. B. (2014). Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, *154*, 38–45. https://doi.org/10.1016/j.rse.2014.08.012

Howell, S. E. L., Brown, L. C., Kang, K. K., & Duguay, C. R. (2009). Variability in ice phenology on Great Bear Lake and Great Slave Lake, Northwest Territories, Canada, from SeaWinds/QuikSCAT: 2000-2006. *Remote Sensing of Environment*, *113*(4), 816–834. https://doi.org/10.1016/j.rse.2008.12.007

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Ishida, H., Oishi, Y., Morita, K., Moriwaki, K., & Nakajima, T. Y. (2018). Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions. *Remote Sensing of Environment*, *205*, 390–407. https://doi.org/10.1016/J.RSE.2017.11.003

Jeffries, M. O., & Morris, K. (2007). Some aspects of ice phenology on ponds in central Alaska, USA. *Annals of Glaciology*, *46*, 397–403. https://doi.org/10.3189/172756407782871576

Jeffries, M. O., Morris, K., & Duguay, C. R. (2012). Floating ice: lake ice and river ice. In R. S. Williams & J. G. Ferrigno (Eds.), *Satellite Image Atlas of Glaciers of the World – State of the Earth's Cryosphere at the Beginning of the 21st Century: Glaciers, Global Snow Cover, Floating Ice, and Permafrost and Periglacial Environments* (pp. A381–A424).

U.S. Geological Survey Professional Paper 1386-A. Retrieved from https://pubs.usgs.gov/pp/p1386a

Jeffries, M. O., Morris, K., & Kozlenko, N. (2005). Ice characteristics and processes, and remote sensing of frozen rivers and lakes. In *Remote Sensing in Northern Hydrology: Measuring Environmental Change* (pp. 63–90). American Geophysical Union (AGU). https://doi.org/10.1029/163GM05

Jönsson, P., & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences, 30*(8), 833–845. https://doi.org/10.1016/j.cageo.2004.05.006

Kang, K. K., Duguay, C. R., & Howell, S. E. L. (2012). Estimating ice phenology on large northern lakes from AMSR-E: Algorithm development and application to Great Bear Lake and Great Slave Lake, Canada. *The Cryosphere*, *6*(2), 235–254. https://doi.org/10.5194/tc-6-235-2012

Kang, X., Li, S., & Benediktsson, J. A. (2014). Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(6), 3742–3752. https://doi.org/10.1109/TGRS.2013.2275613

Karetnikov, S. G., & Naumenko, M. A. (2008). Recent trends in Lake Ladoga ice cover. *Hydrobiologia*, *599*(1), 41–48. https://doi.org/10.1007/s10750-007-9211-1

Karpatne, A., Jiang, Z., Vatsavai, R. R., Shekhar, S., & Kumar, V. (2016). Monitoring land-cover changes: a machine-learning perspective. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 8–21. https://doi.org/10.1109/MGRS.2016.2528038

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, *11*(5), 352–359. https://doi.org/10.1016/j.jag.2009.06.002

Key, J., Drinkwater, M., & Ukita, J. (2007). *Cryosphere Theme Report*. Retrieved from http://igos-cryosphere.org/docs/cryos_theme_report.pdf

Kim, M., Im, J., Han, H., Kim, J., Lee, S., Shin, M., & Kim, H. C. (2015). Landfast sea ice monitoring using multisensor fusion in the Antarctic. *GIScience and Remote Sensing*, *52*(2), 239–256. https://doi.org/10.1080/15481603.2015.1026050

Komarov, A. S., & Buehner, M. (2017). Automated detection of ice and open water from dual-polarization radarsat-2 images for data assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(10), 5755–5769. https://doi.org/10.1109/TGRS.2017.2713987

Komarov, A. S., & Buehner, M. (2018). Adaptive probability thresholding in automated ice and open water detection from RADARSAT-2 images. *IEEE Geoscience and Remote Sensing Letters*, *15*(4), 552–556. https://doi.org/10.1109/LGRS.2018.2806189

Komarov, A. S., & Buehner, M. (2019). Improved retrieval of ice and open water from sequential RADARSAT-2 Images. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(6), 3694–3702. https://doi.org/10.1109/TGRS.2018.2886685

Korhonen, J. (2006). Long-term changes in lake ice cover in Finland. *Hydrology Research*, *37*(4–5), 347–363. https://doi.org/10.2166/nh.2006.019

Kropáček, J., Maussion, F., Chen, F., Hoerz, S., & Hochschild, V. (2013). Analysis of ice phenology of lakes on the Tibetan Plateau from MODIS data. *The Cryosphere*, *7*(1), 287–301. https://doi.org/10.5194/tc-7-287-2013

Kulkarni, V. Y., & Sinha, P. K. (2012). Pruning of random forest classifiers: a survey and future directions. In *2012 International Conference on Data Science & Engineering (ICDSE)* (pp. 64–68). IEEE. https://doi.org/10.1109/ICDSE.2012.6282329

Latifovic, R., & Pouliot, D. (2007). Analysis of climate change impacts on lake ice phenology in Canada using the historical satellite data record. *Remote Sensing of Environment*, *106*(4), 492–507. https://doi.org/10.1016/j.rse.2006.09.015

Lecun, Y., Bottou, L., Bengio, Y., & Ha, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, (November), 1–46. https://doi.org/10.1109/5.726791

Leigh, S., Wang, Z., & Clausi, D. A. (2014). Automated ice-water classification using dual polarization SAR satellite imagery. *IEEE Transactions on Geoscience and Remote*

Sensing, *52*(9), 5529–5539. https://doi.org/10.1109/TGRS.2013.2290231

Leinenkugel, P., Kuenzer, C., & Dech, S. (2013). Comparison and enhancement of MODIS cloud mask products for Southeast Asia. *International Journal of Remote Sensing*, *34*(8), 2730–2748. https://doi.org/10.1080/01431161.2012.750037

Leiva-Murillo, J. M., Gomez-Chova, L., & Camps-Valls, G. (2013). Multitask remote sensing data classification. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 151–161. https://doi.org/10.1109/TGRS.2012.2200043

Leppäranta, M., Terzhevik, A., & Shirasawa, K. (2010). Solar radiation and ice melting in Lake vendyurskoe, Russian Karelia. *Hydrology Research*, *41*(1), 60–62. https://doi.org/10.2166/nh.2010.122

Liu, H., Guo, H., & Zhang, L. (2015). SVM-based sea ice classification using textural features and concentration from RADARSAT-2 dual-pol scansar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(4), 1601–1613. https://doi.org/10.1109/JSTARS.2014.2365215

Liu, Y., Key, J., & Mahoney, R. (2016). Sea and freshwater ice concentration from VIIRS on Suomi NPP and the future JPSS satellites. *Remote Sensing*, *8*(6), 1–20. https://doi.org/10.3390/rs8060523

Livingstone, D. M. (1997). Break-up dates of Alpine lakes as proxy data for local and regional mean surface air temperatures. *Climatic Change*, *37*(2), 407–439. https://doi.org/10.1023/A:1005371925924

Livingstone, D. M., Adrian, R., Blenckner, T., George, G., & Weyhenmeyer, G. A. (2010). Lake ice phenology. In G. George (Ed.), *The Impact of Climate Change on European Lakes* (pp. 51–61). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-2945-4_4

Magnuson, J. J., Robertson, D. M., Benson, B. J., Wynne, R. H., Livingstone, D. M., Arai, T., … Vuglinski, V. S. (2000). Historical trends in lake and river ice cover in the Northern Hemisphere. *Science*, *289*(5485), 1743–1746.

https://doi.org/10.1126/science.289.5485.1743

Maslanik, J. A., & Barry, R. G. (1987). Lake ice formation and breakup as an indicator of climate change : potential for monitoring using remote sensing techniques. *IAHS Proceedings of the Vancouver Symposium*, (168), 10.

Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L., & Tuia, D. (2015). Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(7), 3550–3564. https://doi.org/10.1109/TGRS.2014.2377785

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Metsämäki, S., Pulliainen, J., Salminen, M., Luojus, K., Wiesmann, A., Solberg, R., … Ripper, E. (2015). Introduction to GlobSnow snow extent products with considerations for accuracy assessment. *Remote Sensing of Environment*, *156*, 96–108. https://doi.org/10.1016/j.rse.2014.09.018

Meyer, H., Katurji, M., Appelhans, T., Müller, M. U., Nauss, T., Roudier, P., & Zawar-Reza, P. (2016). Mapping daily air temperature for Antarctica based on MODIS LST. *Remote Sensing*, *8*(9), 1–16. https://doi.org/10.3390/rs8090732

Michel, B., Ashton, G. D., Beltaos, S., Davar, K., Frederking, R., & Gerard, R. (1986). Hydraulics. In G. D. Ashton (Ed.), *River and Lake Ice Engineering*. Water Resources Publications.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., & Kanevski, M. (2014). Machine learning feature selection methods for landslide susceptibility mapping. *Mathematical Geosciences*, *46*(1), 33–57. https://doi.org/10.1007/s11004-013-9511-0

Mishra, V., Cherkauer, K. A., Bowling, L. C., & Huber, M. (2011). Lake ice phenology of small lakes: Impacts of climate variability in the Great Lakes region. *Global and*

Planetary Change, *76*(3–4), 166–185. https://doi.org/10.1016/J.GLOPLACHA.2011.01.004

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: a review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Mullen, P. C., & Warren, S. G. (1988). Theory of the optical properties of lake ice. *Journal of Geophysical Research*, *93*(D7), 8403. https://doi.org/10.1029/JD093iD07p08403

Murfitt, J., & Brown, L. C. (2017). Lake ice and temperature trends for ontario and manitoba: 2001 to 2014. *Hydrological Processes*, *31*(21), 3596–3609. https://doi.org/10.1002/hyp.11295

Murfitt, Justin, & Brown, L. C. (2017). Lake ice and temperature trends for Ontario and Manitoba: 2001 to 2014. *Hydrological Processes*, *31*(21), 3596–3609. https://doi.org/10.1002/hyp.11295

Murphy, K. P. (2013). *Machine learning : a probabilistic perspective*. Cambridge, Mass. [u.a.]: MIT Press.

Nonaka, T., Matsunaga, T., & Hoyano, A. (2007). Estimating ice breakup dates on Eurasian lakes using water temperature trends and threshold surface temperatures derived from MODIS data. *International Journal of Remote Sensing*, *28*(10), 2163–2179. https://doi.org/10.1080/01431160500391957

Ochilov, S., & Clausi, D. A. (2012). Operational SAR sea-ice image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(11 PART1), 4397–4408. https://doi.org/10.1109/TGRS.2012.2192278

Oke, T. R. (1987). *Boundary Layer Climates* (2nd Editio). London: Routledge. https://doi.org/10.4324/9780203407219

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: machine learning in python. *Journal of Machine Learning Research, 12,* 2825–2830.

Perovich, D. K. (1979). The optical properties of young sea ice. *Washington University, Dept. Atmospheric Sciences, Scientific Report, 17*(May).

Perovich, D. K., Roesler, C. S., & Pegau, W. S. (1998). Variability in Arctic sea ice optical properties. *Journal of Geophysical Research: Oceans*, *103*(C1), 1193–1208. https://doi.org/10.1029/97JC01614

Persello, C., & Bruzzone, L. (2014). Active and semisupervised learning for the classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(11), 6937–6956. https://doi.org/10.1109/TGRS.2014.2305805

Petrov, M. P., Terzhevik, A. Y., Palshin, N. I., Zdorovennov, R. E., & Zdorovennova, G. E. (2005). Absorption of solar radiation by snow-and-ice cover of lakes. *Water Resources*, *32*(5), 496–504. https://doi.org/10.1007/s11268-005-0063-7

Qi, M., Yao, X., Li, X., Duan, H., Gao, Y., & Liu, J. (2019). Spatiotemporal characteristics of Qinghai Lake ice phenology between 2000 and 2016. *Journal of Geographical Sciences*, *29*(1), 115–130. https://doi.org/10.1007/s11442-019-1587-0

Qiu, Y., Xie, P., Leppäranta, M., Wang, X., Lemmetyinen, J., Lin, H., & Shi, L. (2019). MODIS-based daily lake ice extent and coverage dataset for Tibetan Plateau. *Big Earth Data*, *3*(2), 170–185. https://doi.org/10.1080/20964471.2019.1631729

Reed, B., Budde, M., Spencer, P., & Miller, A. E. (2009). Integration of MODIS-derived metrics to assess interannual variability in snowpack, lake ice, and NDVI in southwest Alaska. *Remote Sensing of Environment*, *113*(7), 1443–1452. https://doi.org/10.1016/j.rse.2008.07.020

Riggs, G. A., & Hall, D. K. (2015). *MODIS snow products collection 6 user guide*.

Riggs, G. A., Hall, D. K., & Salomonson, V. V. (2006). *MODIS snow products user guide to collection 5*.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104.

https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rongbin Wang, Chao Zeng, Li, P., & Shen, H. (2011). Terra MODIS band 5 Stripe noise detection and correction using MAP-based algorithm. In *2011 International Conference on Remote Sensing, Environment and Transportation Engineering* (pp. 8612–8615). IEEE. https://doi.org/10.1109/RSETE.2011.5964181

Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., … Zhang, C. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment*, *114*(1), 35–49. https://doi.org/10.1016/j.rse.2009.08.011

Ruß, G., & Brenning, A. (2010). Data mining in precision agriculture: management of spatial information. In E. Hüllermeier, R. Kruse, & F. Hoffmann (Eds.), *Computational Intelligence for Knowledge-Based Systems Design* (pp. 350–359). Berlin, Heidelberg: Springer Berlin Heidelberg.

Scikit-learn. (2020). *Scikit-learn user guide*. Retrieved from https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf

Scott, J. T. (1964). *A Comparison of the Heat Balance of Lakes in Winter*.

Shen, X., Zhang, J., Zhang, X., Meng, J., & Ke, C. (2017). Sea ice classification using cryosat-2 altimeter data by optimal classifier-feature assembly. *IEEE Geoscience and Remote Sensing Letters*, *14*(11), 1948–1952. https://doi.org/10.1109/LGRS.2017.2743339

Shi, D., & Yang, X. (2016). An assessment of algorithmic parameters affecting image classification accuracy by random forests. *Photogrammetric Engineering and Remote Sensing*, *82*(6), 407–417. https://doi.org/10.14358/PERS.82.6.407

Shih, H. chien, Stow, D. A., & Tsai, Y. H. (2019). Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. *International Journal of Remote Sensing*, *40*(4), 1248–1274. https://doi.org/10.1080/01431161.2018.1524179

Šmejkalová, T., Edwards, M. E., & Dash, J. (2016). Arctic lakes show strong decadal trend in earlier spring ice-out. *Scientific Reports*, *6*(November), 1–8. https://doi.org/10.1038/srep38449

Soja, A. M., Kutics, K., Maracek, K., Molnár, G., & Soja, G. (2014). Changes in ice phenology characteristics of two Central European steppe lakes from 1926 to 2012 - influences of local weather and large scale oscillation patterns. *Climatic Change*, *126*(1–2), 119–133. https://doi.org/10.1007/s10584-014-1199-8

Sturm, M., Holmgren, J., König, M., & Morris, K. (1997). The thermal conductivity of seasonal snow. *Journal of Glaciology*, *43*(143), 26–41. https://doi.org/10.3189/S0022143000002781

Su, H., Wang, Y., Xiao, J., & Yan, X. H. (2015). Classification of MODIS images combining surface temperature and texture features using the support vector machine method for estimation of the extent of sea ice in the frozen Bohai Bay, China. *International Journal of Remote Sensing*, *36*(10), 2734–2750. https://doi.org/10.1080/01431161.2015.1041619

Surdu, C. M., Duguay, C. R., & Prieto, D. F. (2016). Evidence of recent changes in the ice regime of lakes in the Canadian High Arctic from spaceborne satellite observations. *Cryosphere*, *10*(3), 941–960. https://doi.org/10.5194/tc-10-941-2016

Svacina, N. A., Duguay, C. R., & King, J. M. L. (2014). Modelled and satellite-derived surface albedo of lake ice - part II: Evaluation of MODIS albedo products. *Hydrological Processes*. https://doi.org/10.1002/hyp.10257

Tekeli, A. E., Akyürek, Z., Arda Şorman, A., Şensoy, A., & Ünal Şorman, A. (2005). Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey. *Remote Sensing of Environment*, *97*(2), 216–230. https://doi.org/10.1016/J.RSE.2005.03.013

Tom, M., Kälin, U., Sütterlin, M., Baltsavias, E., & Schindler, K. (2018). Lake ice detection in low-resolution optical satellite images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*(2), 279–286. https://doi.org/10.5194/isprs-annals-IV-2-279-2018

Trishchenko, A. P., Luo, Y., & Khlopenkov, K. V. (2006). A method for downscaling MODIS land channels to 250-m spatial resolution using adaptive regression and normalization. *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VI*, *6366*(October 2006), 636607. https://doi.org/10.1117/12.689157

Vapnik, V. (1998). The Support Vector Method of Function Estimation. In J. A. K. Suykens & J. Vandewalle (Eds.), *Nonlinear Modeling: Advanced Black-Box Techniques* (pp. 55–85). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-5703-6_3

Veillette, J., Muir, D. C. G., Antoniades, D., Small, J. M., Spencer, C., Loewen, T. N., … Vincent, W. F. (2012). Perfluorinated chemicals in meromictic lakes on the northern coast of Ellesmere Island, High Arctic Canada. *Arctic*, *65*(3), 245–256.

Wang, J., Duguay, C. R., Clausi, D. A., Pinard, V., & Howell, S. E. L. (2018). Semi-automated classification of lake ice cover using dual polarization RADARSAT-2 imagery. *Remote Sensing*, *10*(11). https://doi.org/10.3390/rs10111727

Wang, L., Scott, K. A., Xu, L., & Clausi, D. A. (2016). Sea ice concentration estimation during melt from dual-pol sar scenes using deep convolutional neural networks: a case study. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4524–4533. https://doi.org/10.1109/TGRS.2016.2543660

Wang, X., Liu, Y., & Bao, Q. (2016). Impacts of cloud overlap assumptions on radiative budgets and heating fields in convective regions. *Atmospheric Research*, *167*, 89–99. https://doi.org/10.1016/J.ATMOSRES.2015.07.017

Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(5), 450–457. https://doi.org/10.1016/j.isprsjprs.2009.01.003

Weston, J., & Watkins, C. (1999). Support vector machines for multi-class pattern recognition. *Proceedings of the 7th European Symposium on Artificial Neural Networks (ESANN-99)*, (April), 219–224.

Wieland, M., Li, Y., & Martinis, S. (2019). Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment*, *230*(May). https://doi.org/10.1016/j.rse.2019.05.022

Williams, G. P. (1965). Correlating freeze-up and break-up with weather conditions. *Canadian Geotechnical Journal*, *2*(4), 313–326. https://doi.org/10.1139/t65-047

Williams, S. G., & Stefan, H. G. (2006). Modeling of lake ice characteristics in north america using climate, geography, and lake bathymetry. *Journal of Cold Regions Engineering*, *20*(4), 140–167. https://doi.org/10.1061/(ASCE)0887-381X(2006)20:4(140)

Wolfe, R. E. (2006). MODIS geolocation. In J. J. Qu, W. Gao, M. Kafatos, R. E. Murphy, & V. V Salomonson (Eds.), *Earth Science Satellite Remote Sensing: Vol. 1: Science and Instruments* (pp. 50–73). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-37293-6_4

Wrona, F. J., Johansson, M., Culp, J. M., Jenkins, A., Mård, J., Myers-Smith, I. H., … Wookey, P. A. (2016). Transitions in Arctic ecosystems: ecological implications of a changing hydrological regime. *Journal of Geophysical Research: Biogeosciences*, *121*(3), 650–674. https://doi.org/10.1002/2015JG003133

Xiong, X., King, M. D., Salomonson, V. V, Barnes, W. L., Wenny, B. N., Angal, A., … Link, D. O. (2015). Moderate resolution imaging spectroradiometer on Terra and Aqua missions. In *Optical Payloads for Space Missions* (pp. 53–89). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118945179.ch3

Xu, L., Li, J., & Brenning, A. (2014). A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery. *Remote Sensing of Environment*, *141*, 14–23. https://doi.org/10.1016/j.rse.2013.10.012

Yang, Q., Song, K., Wen, Z., Hao, X., & Fang, C. (2019). Recent trends of ice phenology for eight large lakes using MODIS products in Northeast China. *International Journal of Remote Sensing*, *40*(14), 5388–5410. https://doi.org/10.1080/01431161.2019.1579939

Zhang, L., Liu, H., Gu, X., Guo, H., Chen, J., & Liu, G. (2019). Sea ice classification using TerraSAR-X ScanSAR data with removal of scalloping and interscan banding. *IEEE*

Journal of Selected Topics in Applied Earth Observations and Remote Sensing, *12*(2), 589–598. https://doi.org/10.1109/JSTARS.2018.2889798

Zhang, S., & Pavelsky, T. M. (2019). Remote sensing of lake ice phenology across a range of lakes sizes, ME, USA. *Remote Sensing*, *11*(14), 1–13. https://doi.org/10.3390/rs11141718