ADVANCED DATA FUSION METHODS TO IMPROVE WETLAND CLASSIFICATION USING MULTI-SOURCE REMOTELY SENSED DATA

AARON J. JUDAH

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN EARTH AND SPACE SCIENCE
ENGINEERING YORK UNIVERSITY TORONTO, ONTARIO

January 2023

Abstract

The goal of this research was to improve wetland classification accuracy and the reduction of classification errors and uncertainty by fully exploiting multi-source remotely sensed (Landsat-5/8, Sentinel-1/2, and RADARSAT), and ancillary (Digital Elevation model) data and image metrics using advanced data analysis techniques. This PhD research was done on a holistic approach in three phases: 1. Explorations of data type selections and significance in support of wetland classification. 2. The development of a hierarchically-based classification approach to best exploit the data identified and characterized through the first study. 3. The development of an ensemble classifier incorporating the aforementioned developments with Dempster-Shafer theory in order to reduce errors and streamline computations.

The first phase explored the most effective data features (multispectral image bands, image metrics, and ancillary data), and metrics or families of data features in support of wetland classification from a study area in Northern Ontario, Canada. Land covers examined were Open Fen, Treed Fen, Open Bog, Treed Bog, Cleared areas, Swamps, Grassy areas, and Dense Coniferous Forests. It was found that wetlands were best classified using the NDVI (Normalized Difference Vegetative Index) calculated from optical imagery obtained in the spring months, radar backscatter coefficients, surface temperature, and ancillary data such as surface slope, computed through either a Random Forest (RF) or Support Vector Machine (SVM) classifier. It was also found that preselection of features using Log-normal or RF variable importance analysis was an effective way of identifying low quality features and to a lesser extent features which were of higher quality. This work was also able to produce a wetland land cover map with an accuracy of 87.51% - an improvement from the ~82% typical of similar datasets and landcover types.

In the second phase a more effective approach to classify the aforementioned features in order to fully utilize the discriminant power of those features was explored. This was done through two hierarchically based classification strategies. The first focused on splits between treed and non-treed wetlands, while the second on splits between Fens, Bogs, and Swamps. The second hierarchically based RF classification methodology produced the most accurate classification result (91.94%). The hierarchically based approaches also improved classification accuracies for low quality data, as defined through feature analysis, when compared to a non-hierarchical classifier.

The third phase focused on how to better exploit broad class separations and to reduce the propagation of errors and uncertainty which cascades through the classification hierarchy. Three distinct classifiers were designed to distinguish individual or compound wetland categories using RF classification. They were determined, in large part, to best use the available features to maximize classification accuracy. The results from these classifiers were integrated using Dempster–Shafer theory (D–S theory). The developed method was tested on data collected from a study area in Northern Alberta, Canada. The data utilized were Landsat-8 and Sentinel-2 (multi-spectral), Sentinel-1 (synthetic aperture radar—SAR), and digital elevation model (DEM). Classification of Fen, Bog, Marsh, Swamps, and Upland resulted in an overall accuracy of ~93% an improvement of 5% when compared to a traditional classification method. It was noted that, with the traditional method, some pixels were misclassified with a high level of confidence (>85%). Such misclassification was reduced by ~10% by the proposed method. Results also showed that some features important in separating compound wetland classes were not considered important using the traditional method based on the RF feature selection mechanism. When used in the proposed method, these features increased the classification accuracy.

The major contribution of this research was the improvement of classification accuracy and the reduction of classification errors and uncertainty through use of multiple classifiers, designed to best exploit broad class separations, through selected data features computed within a D-S framework.

Dedication

Dedicated to my late brother Dr. David Judah, DVM, who has and will always inspire me to be a better scientist and a better man.

Acknowledgements

I would like to extend my sincere gratitude to all of the people who supported and helped me over the years in order to bring this research to completion, in particular my supervisor Prof. Baoxin Hu. I would like to thank her for her clear guidance, insightful suggestions, and challenging discussions that helped in the successful publication of the three papers which make up the core of this research, and helped to develop me into a better scientist.

I am also indebted to Prof. Jianguo Wang and Prof. Costas Armenakis, my supervisory committee members, for their scientific suggestions, valuable insights, and encouragements throughout this research.

I would also like to acknowledge my friends and family for their love and support throughout the years which helped to motivate me and to keep me going. And finally, I would like to acknowledge my extended military family, who not only tolerated my eccentricities, and listened with measured and patient interest as I would describe my thesis work but who also taught me humbleness, grit, perseverance, strength, and also inspired me to be part of something greater than myself. Without all of you none of this would have been possible.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ABMI	Alberta Biodiversity Monitoring Institute
AI	Artificial Intelligence
CNN	Convolution Neural Network
DEM	Digital Elevation Model
DN	Digital Number
D-S	Dempster Shafer
EVI	Enhanced Vegetation Index
K-NN	K Nearest Neighbor
LN	Log Normal
MIR	Mid-Infrared Region
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-infrared
NIRv	Near-Infrared Reflectance Vegetation Index
OvO	One Versus One
RF	Random Forest
RGB	Red Green Blue
SAR	Synthetic Aperture Radar
SLC	Single Look Complex
SVM	Support Vector Machine

Chapter 1 - Introduction

Wetlands play key roles in regional and global environments and are critically linked to major issues such as climate change, wildlife habitat health, and biodiversity. More specifically, wetlands play important roles in flood mitigation, water quality protection, and global carbon and methane cycles, acting as buffers against droughts, protecting coastlines from rising tides and storms, and being responsible for sediment retention [Blaustein et al 1994; Dahl, 2000; Kaplan and Avdan 2018; Mahdavi, et al 2018; U.S. Fish and Wildlife 2002; Finlayson and Davidson 1999, Tiner et al 2015]. In addition, nearly one-half of plant and animal species listed as endangered by the U.S. Fish and Wildlife Service are wetland dependent [U.S. Fish and Wildlife 2002], and wetland loss is arguably the largest factor for the cause of global amphibian declines [Blaustein et al 1994]. North American and global wetland losses are estimated to be on the order of 50% since the early 1700s [Dahl 2000; Finlayson and Davidson 1999]. The importance of wetland conservation is well-established as a matter of national and international public policy. In this vein, accurately mapping and monitoring wetlands and their changes in a timely and repeatable manner are of utmost importance. Remotely sensed imagery, and associated ancillary data provides researchers with a means to achieve these goals.

Maps of wetlands created with medium resolution (10 m–30 m) [Amani et al 2019; Guo and Meng 2017; Hu et al 2021; Hu et al 2019; Jiao et al 2019; Judah and Hu 2019, 2021; Mirmazloumi et al, 2021; Ozesmi and Bauer 2002; Ramsey and Laine 1997; Yuan et al 2022], high resolution (9 m -1.5 m) [Guo and Meng 2017; Mahdianpari et al 2017; Masoumi et al 2017; Tian et al 2016], and very high resolution (1.5 m>) [Kaplan and Avdan 2018; Koma et al 2021; Guo and Meng 2017; Mahdavi, et al 2018; Wang et al 2014; Wang et al 2022] remotely sensed imagery has provided some levels of success in creating

suitable wetland maps. From my literature review, over time, I noted a progressive increase in the overall classification accuracy of the wetland maps, and an increase in the types of wetlands classified – more narrow and sophisticated wetland class definitions. Additionally, for all studies there was an overall trend towards increasing the number and diversity of imagery sources, and the incorporation of ancillary data, such as elevation maps and field samples [Bhatt 2022; Kumar 2018; Gallant 2015; In-Young et al 2019; Millard and Richardson 2013] in an effort to better characterize those more difficult to classify wetland classes, and in an effort to increase classification accuracies. Finally, as the field of remote sensing of wetlands and landcovers has progressed, there has been a trend towards the incorporation of more computationally complex and sophisticated classification methodologies and algorithms. As before, the intent of this is to create more sophisticated classification maps with greater overall classification accuracies. However, as I outline in more detail in the next chapter, for this PhD work I did note the two following important issues with wetland classification. (1) All studies have challenges with regards to the characterization, identification, and selection of the most impactful and effective data features in support of wetland classification. (2) There is a need to determine how best to utilize the available remotely sensed data in order to best exploit that data's unique characteristics. The focus of this dissertation research was to address these issues, through the development of innovative classification approaches utilizing multi-source, multi-temporal data for the improvement of the classification of various wetlands. This was done in a holistic approach which structured my work in three phases, which built from one another.

(1) The characterization, identification, and selection of the most impactful and effective data features in support of wetland classification.

Effective data features are part of the fundamental basis for any successful classification using remotely sensed imagery, and the characterization and selection of

those data features for wetland classification would be the first and critical step towards improving it.

(2) Development of hierarchical classification strategies to best exploit the available remotely sensed, and ancillary data.

Once those features or families of features have been identified, determining the best way to classify them, to fully utilize the discriminant power of those features, would be the natural next step.

(3) Development of an advance data fusion approach to best utilize available remotely sensed data, with the intent on improving accuracy and reducing errors and uncertainty (in my case the quantitative measure of variability with a measurement or computational result).

Finally, the refinement of the aforementioned developed methodology in order to reduce errors and uncertainty would be a last step to improve classification accuracies and confidence in those results. As a point of clarity, confidence corresponds to the likelihood that the classification result would be replicated under similar computational conditions.

It should also be noted that for each phase the developed classifier(s) were unique, and each phase had its own unique data set. This is motivated the examination of feature selection, to some extent, across all phases of this PhD work. While commonalities between all data sets and classifiers was observed nuanced differences between data sets and classifiers drove the need to do this type of examination for each chapter.

Following the structure outlined above, this PhD research was conducted through those outlined phases, which culminated in the development of a classification methodology which best used the available remotely sensed features in order to maximize the impact of that information, to maximize classification accuracy and reduce errors and uncertainty – the last phase. Three journal publications were generated from those phases.

This dissertation is organized into 6 chapters. In the first chapter (this section), a general introduction and overview of the field of remote sensing techniques of wetlands is provided. Furthermore, in this chapter I present my research motivations and goals.

In chapter two the literature review of the state of the art of the field is presented.

The developed data feature characterization and selection techniques and corresponding classification results are presented in chapter three. The techniques were evaluated in a study area in Northern Ontario where the most significant features for wetland classification, for that study area, were identified. Classification results indicated improvements over what has been reported in the popular literature and are further discussed in that chapter.

In Chapter four several hierarchical classification approaches are described and the results and discussions are provided. The developed techniques are innovative in that they leverage broad class separations to facilitate efficient and accurate classifications. In order to execute these hierarchical approaches, in part, I utilize the techniques developed in chapter three.

In Chapter five, I present work where I used data fusion techniques to combine multiple classifiers to improve accuracies and reduce the number of misclassified high confidence data – an innovation for the field. This is achieved through building from the work described in chapters three and four and incorporating D-S theory to create the final classified maps. The results of this study are discussed in this chapter as well.

Lastly, Chapter six provides conclusions and significant contributions of the research. Future considerations for further improving the overall classification of wetlands are discussed, as well as additional applications for this work, and other research directions.

This dissertation research generated the following publications:

1. Judah, Aaron, and Baoxin Hu. 2019. "The Integration of Multi-source Remotely-Sensed Data in Support of the Classification of Wetlands." *Remote Sensing* 11 (13):28. doi: 10.3390/rs11131537.
2. Judah, Aaron, and Baoxin Hu. 2021. "The Integration of Multi-Source Remotely Sensed Data with Hierarchically Based Classification Approaches in Support of the Classification of Wetlands." *Canadian journal of remote sensing* 48 (2):158-81. doi: 10.1080/07038992.2021.1967732.
3. Judah, Aaron, and Baoxin Hu. 2022. "An Advanced Data Fusion Method to Improve Wetland Classification Using Multi-Source Remotely Sensed Data" Sensors 22, no. 22: 8942. https://doi.org/10.3390/s22228942

Chapter 2 - Literature Review

In this chapter I will present a review of the popular literature I examined in support of this dissertation research. The field of remote sensing of wetlands is vast and this necessitated the narrowing of the review of the popular literature to a more appropriate subset. Given the focus of this study was the development of advanced classification methodologies and the locations of the study areas were located in Northern Ontario, and Northern Alberta, this motivated my review to focus on the classification of medium and high resolution imagery (standard imagery products used in the classification of similar or identical wetlands and land covers from these area) and the most advanced classification and regression techniques (more colloquially referred to as data science algorithms) used in the classification of wetlands. This chapter is broken into three sections: 1. Imagery and data features used for wetland classification, 2. Methods for wetland classification based on multi-source remotely sensed data, and 3. Deep learning methods and their use in the remote sensing of wetlands.

2.1. Imagery and Data Features Used for Wetland Classification

The field of remote sensing of wetlands is a subset of land cover classification. As with all land cover classification, as a general trend, wetland mapping using Landsat TM imagery is common and standard approach. It is found to have good class separation when one class dominated the classification area (>30 m^2), but not when mixtures of wetlands types were of the same order as the sensor resolution [Ozesmi et al., 2002]. Additionally, for these Landsat derived maps, accuracy levels varied depending on the choice of datasets and the techniques used. Generally, higher accuracies (~63 to 92%) were observed if advance regression techniques (RF or K-NN) were used, combined with additional datasets, such as LiDAR or RADAR imagery [Amarsaikhan, D., & Douglas, T., 2004; Bwangoy et al., 2010; Dubeau et al., 2017; Eisavi et al, 2015; Huang et al., 2014; Gala et al., 2012; Klemas et al., 2013; Masoumi et al., 2017; Mirmazloumi et al, 2021; Tian et al, 2016; Yuan et al 2022]. When used alone with advance classification techniques Landsat imagery produced wetland maps with classification accuracies between 40 and 83% [Gallant, 2015; Ozesmi et al. 2002; Ramsey et al. 1997; Wang et al 2014]. For all studies, with finer class definitions, lower classification accuracies were observed, and in some cases aggregation of similar wetland classes are sometimes necessary in order to produce a product with desirable accuracies [Davranche, et al, 2010; Wright & Gallant, 2007]. For all of these studied individual image bands, in particular those centered on the green, red, and near infrared bands have been found to be significant for wetland classification. This has been found to be driven by the effectiveness of these bands at identifying and measuring variations in vegetation associated with wetlands. These variations are physically manifested as vegetation physical activity via chlorophyl, and unique vegetative colour and

texture as measured through remotely sensed imagery. [Amani et al. 2017a; Asner 1998; Cochrane 2000; Schmidt and Skidmore 2003; Thenkabail et al. 2004; Kamaruzaman and Kasawani 2007; Kaplan and Avdan 2018; Mutanga, and Rugege 2010]. Also, the Normalized Difference Vegetation Index (NDVI), and the Normalized Difference Water Index (NDWI) have been found to be a popular and significant data feature in many wetland classification studies. The success of NDVI and NDWI has been driven by not only its ability to effectively measure vegetation and water differences (both significant in wetland detection) but the detection of water bodies with relatively high phytoplankton content [Weichelt et al. 2011; Wen et al. 2014]. For some studies, the classification process with TM imagery is aided through the incorporation of ancillary data such as elevation maps and field samples [Kushwaha et al., 2000; Millard et al., 2013; Ozesmi et al., 2002, Wright & Gallant, 2007]. Physically, properties such as the identification of flat areas (in the case for wetlands with flooded areas) and areas with slopes (in the case of certain types of fens) would be identified from the aforementioned data. Additionally, more exotic ancillary data such as fire history [Rogan et al. 2003], and valley bottom flatness (VBF) [ABMI 2021] has been explored, for the same reasons. It should be noted that surface temperature, while a readily available Landsat product, is not commonly used in the classification of wetlands, mainly due to its relatively low spatial resolution. Due to its relatively low spatial resolution (compared with optical satellite imagery), surface temperature is utilized to differentiate surface cover types with a large difference in temperature, such as separating roads and buildings from vegetation. However, given the recent advances in machine learning, I contend that a smaller difference in surface temperature may be able to aid in the classification process of surface cover types. As a result, surface temperature was exploited in this study. Physically, the mechanism driving these smaller differences in surface temperature between different vegetative or other landcovers can be linked to differences in evapotranspiration driven by difference in

vegetative activity and water content. Since many wetland species have overlapping spectral reflectance at peak biomass (Schmidt & Skidmore, 2003), researchers have employed multitemporal imagery in the classification process of TM imagery [Coll et al., 2010; Davranche et al.,2010; Eisavi et al, 2015; Huang et al., 2014; Kaplan and Avdan 2018; Mirmazloumi et al, 2021; Ozesmi et al., 2002]. This aids in characterizing landcovers through seasonally driven variations. Other studies have approached this problem by incorporated RADAR or LiDAR based measurements with Landsat TM imagery to aid in their classification methodologies. As previously states resulting classification accuracies range from ~63% to 93%. The incorporation of RADAR or LiDAR data aids in characterizing structural differences in landcovers such as dense vegetation canopy compared to grasses.

It is also worth noting that in many studies, there is relatively little justification for choice in features used in classification, with trial and error being a common approach [Amarsaikhan, D., & Douglas, T., 2004; Bwangoy et al., 2010; Davranche et al.,2010; Tian, et al, 2016; Rapinel, et al 2015]. It is also still common practice to test all possible features or parameters to determine the most optimal set of inputs [Dubeau et al., 2017; Eisavi et al, 2015; Huang et al., 2014; Gala et al., 2012; Klemas et al., 2013]. This is not a desirable strategy as that this is both crude and time consuming. Furthermore, the performances of machine learning algorithms depend strongly on inputs used for classification, which could explain, partially, why there is no clear consensus on their relative performance of different algorithms to one another [Adam et al., 2010; Mahdavi et al., 2018; Maxwell et al., 2018; Rundquist et al., 2001]. However, the use of a set of features that are significant among the land covers of concern undoubtedly aid in the pursuit of superior classification accuracy. This is especially important given that the advances in remote sensing technology make an enormous amount of data readily available. A key remaining challenge in land cover classification lies in how to extract the best or

most relevant information from a huge amount of data in an efficient and logical way [Eisavi et al., 2015; Guo et al., 2017; Mahdavi et al., 2018; Maxwell et al., 2018]. Considering all of these factors, there is a strong need to determine which data features are best suited for identifying wetlands. Furthermore, quantifying the quality of these features can help provide a better understanding of how accuracy and error propagates through different types of analysis [Hu et al., 2021]. In support of this, one goal of this PhD research was to investigate the significance of different combinations of features and feature types through various analysis and classification methodologies, with the intent of determining which features were the most significant in the classification process of wetlands for my study area, classification approach, and which approaches are best suited in determining those features.

In an effort to better understand trends, common themes, and the edge of knowledge of this field of study, I have produced Table 2-1. It summarizes 15 key papers published in the last three years [ABMI 2021; Amani et al. 2019; Deng et al. 2022; Huaxin et al. 2022; Hosseiny et al. 2022; Hu et al. 2019; Juan et al. 2022; Koma et al. 2021; Jiao et al. 2019; Liu et al. 2022; Mao et al 2020; Mahdianpari et al. 2019; Meng et al. 2019; Munizaga et al. 2022; O'Neil et al. 2020; Pouliot et al. 2019; Yuan et al. 2022], highlighting the title of the paper, year of publication, main classification method, study area, cover types explored, data used, # of features used, and overall classification accuracy.

Table 2-1. List of recent papers on wetland classification, summarizing the title of the paper, year of publication, main classification method, study area, cover types explored, data used, # of features used, and overall classification accuracy.

#	Ref. Author	Paper Title	Year	Classification Method	Study Area	Cover Types	Data Used	# of Features	Accuracy
1	Hosseiny et al. 2022	WetNet: A Spatial-Temporal Ensemble Deep Learning Model for Wetland Classification Using Sentinel-1 and Sentinel-2	2022	CNN	Newfoundland Canada	Bog, Fen, Marsh, Swamp, Shallow water, Urban, Deep water, Upland	Sentinel-1, Seninel-2	60	95.80%
2	Juan et al. 2022	Mapping Coastal Wetlands Using Satellite Imagery and Machine Learning in a Highly Urbanized Landscape	2022	RF	Concepcion, Chile	Water Bodies, Native Forests, Beaches and Dunes, Bush, Wetlands, Built-up areas, Plantations, Agricultural lands, Young Plantation, Bare Land, Burnt Areas	RapidEye, DEM	50	88%
3	Deng et al. 2022	Comparison of Multi-Class and Fusion of Multiple Single-Class SegNet Model for Mapping Karst Wetland Vegetation Using UAV Images	2022	RF	Huixian National Wetland Park	Karst water, Karst grass, Karsy broad-leaved forest, Karst aquatic flora, other types	UAV - 1-inch 20MP CMOS camera	5	87.34%
4	Huaxin et al. 2022	Object-Based Multigrained Cascade Forest Method for Wetland Classification Using Sentinel-2 and Radarsat-2 Imagery	2022	object-based multigrained cascade forest (OGCF)	Wuyuer River National Natural Reserve, China	Carex marsh, Phragmites marsh, Paddy field, Surface water, Residential area, Road, Upland field, Forest, Medow	Sentinel-2, Radarsat-2	54	88.20%
5	Yuan et al. 2022	Multi-Resolution Collaborative Fusion of SAR, Multispectral and Hyperspectral Images for Coastal Wetlands Mapping	2022	RF, SVM	Yellow River Estuary, China	Artificial trench, Pothole, Construction, Culture Pond, Ocean, Reed, Salt marsh, Paddy field, Dry land, Spartina alterniflora, Suaeda salsa, Bare land	ZY-1 02D, GaoFen-5B, Sentinel-2, Sentinel-1	492	94.11%, 89.64
6	ABMI 2021	ABMI Wetland Inventory – Technical Documentation ABMI Geospatial Center, March, 2021	2021	CNN	Northern Alberta	Bog, Fen, Marsh, Swamp, Upland	Sentinel-1, Sentinel-2, Advanced Land Observing Satellite (ALOS) Digital Surface Model (DSM)	14	85.20%

7	Komza et al. 2021	Classifying wetland-related land cover types and habitats using fine-scale lidar metrics derived from country-wide Airborne Laser Scanning	2020	RF	Lauwersmeer, Netherlands	Open Water, Bare Ground, Pioneer vegetation, Grassland (low), Grassland (high), Land reed structurally rich (dense), Land reed structurally rich (open), Land reed structurally poor (dense), Land reed structurally poor (open), Water reed (dense), Water reed (open), Transition between reed and shrub, Shrub low (dense), Shrub low (open), Shrub mid-hight (dense), Shrub mid-height (open), Shrub height (dense), Shrub height (open), Forest (dense), Forest (open), Anthropogenic objects	Lidar - AHN2	31	82%
8	Mao et al 2020	National wetland mapping in China: A new product resulting from object-based and hierarchical classification of Landsat 8 OLI images	2020	Decision Trees	Geographic extent of China	Inland swamp, Inland Marsh, Lake, River, Coastal swamp, Coastal marsh, Lagoon, Estuarine water, Tidal flat, Shallow marine water, Reservoir/pond, Canal/channel, Salt pan, Aquaculture pond	Landsat-8	48	95.1
9	O'Neil et al. 2020	Deep learning Using Physically-Informed Input Data for Wetland Identification	2020	CNN	Virginia, USA	Open Water, Impervious, Barren, Forest, Tree, Shrub/Scrub, Harvested/Distrubed, Turfgrass, Pasture, Cropland, Wetlands	LiDAR DEM, National Agriculture Imagery Program (NAIP) aerial imagery	539	91%
10	Amani et al. 2019	Canadian Wetland Inventory using Google Earth Engine: The First Map and Preliminary Results.	2019	RF	Geographic extent of Canada	Bog, Fen, Marsh, Swamp, Shallow water, Deep Water, Forest, Grassland, Cropland, Barren	Landsat-8	11	71%, 66%, 63%
11	Hu et al. 2019	Hyperspectral Coastal Wetland Classification Based on a Multiobject Convolutional Neural Network Model and Decision Fusion	2019	CNN	Yellow River Estuary, China	Reed, Mixed Tamarix-S, Water, Spartina, Tidal flat	Compact High Resolution Imaging Spectrometer (CHRIS)	52	82.11%
12	Jiao et al. 2019	A Hierarchical Classification Framework of Satellite Multispectral/Hyperspectral Images for Mapping Coastal Wetlands	2019	RF	Yellow River Estuary and Yancheng coastal wetland, China	Sea, River, Buildings, Dry land, Paddy field, Ditch, Aquaculture, Pond, Mudflat, Marsh, Vegetation, Bare land, Salt marsh	Landsat-5, Landsat-7, Landsat-8, Sentinel-2, GF-5	123	93.46
13	Mahdianpari et al. 2019	The First Wetland Inventory Map of Newfoundland at a Spatial Resolution of 10 m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform	2019	RF	Newfoundland Canada	Bog, Fen, Marsh, Swamp, Shallow water, Deep Water, Upland, Urban/Bare land	Sentinel-1, Sentinel-2	57	88.37%

14	Pouliot et al. 2019	Assessment of Convolution Neural Networks for Wetland Mapping with Landsat in the Central Canadian Boreal Forest Region	2019	CNN	Northern Alberta	Bog, Fen, Marsh, Swamp, Open water, Upland	Landsat-5 and Landsat-8	50	90.00%
15	Meng et al. 2019	Lake Wetland Classification Based on an SVM-CNN Composite Classifier and High-resolution Images Using Wudalianchi as an Example	2019	SVM-CNN	Wudalianchi Nature Reserve, China	Forest, Meadow, Cultivated Land, Lake, Bare land, Limestone, Forest, Swamp, Aquatic vegetation, Volcanic rock	Gaofen 2 (GF-2)	64	92.50%

2.2. Methods for Wetland Classification Based on Multi-Source Remotely Sensed Data

In this section I will present the literature I reviewed in support of understanding different classification methodologies related to the classification of wetlands and other landcovers with remotely sensed imagery. It is broken into three sections: 1. Remotely-Sensed Data With Hierarchically Based Classification, 2. Feature-Level Fusion for Classification of Remotely Sensed Data, and 3. Decision-Level Fusion for Classification of Remotely sensed Data. This is a broad field, but I focused on the most cutting-edge areas that had the most promise for areas of investigation. Hierarchical classification techniques have found some level of popularity and success with the classification of remotely sensed imagery, driven by the identification of splits between class types and data clustering. With the wide availability of large sets of multi-source remotely sensed data, data fusion techniques have also become very popular in the classification of those data. This is partially driven by advances in the field of data science resulting in the development of new techniques and the advancement of computational power which has made these techniques more accessible. Accordingly, data fusion at the feature level and decision level is the focus of this study. In the below sections I will expand in more detail key areas of interest in those fields of study related to this research.

2.2.1. Remotely-Sensed Data With Hierarchically Based Classification

Most of the existing wetland classification methods consider all the wetland classes of interest together and then select the most effective features to identify them. Often a

large number of features are employed or for the sake of efficacy, a subset of features from a larger set is selected, often by trial and error or user skill [Bwangoy et al 2010; Davranche et al 2010; Frohn et al 2015; Gallant 2015; Klemas 2013; Ozesmi and Bauer 2002; Tian et al 2016]. In order to improve upon these approaches and to improve classification accuracies and computational efficiencies, some researchers have utilized advanced regression algorithms, executed with hierarchical approaches using various tactics. Some studies have used a single data source, and executed it through a hierarchical approach, of ever increasing class details [Jiao et al 2019; Koma el al 2020; Maoa et al 2020; Munoz 2019], while others have incorporated multiple data sources, executed with similarly setup hierarchies or hierarchies based on progressively increasing resolution of those imagery sources or through their unique features (i.e. radar or ancillary data) [Jiao et al 2019; Mahdianpari et al 2017, 2018, 2020]. These studies have produced accuracies ranging from to 71 to 95% [Jiao et al 2019; Koma el al 2020; Mahdianpari et al 2017, 2020; Mohammadimanesh et al 2018, 2020; Munoz 2019].

Part of my studies [Judah and Hu 2019] had a strong focus on feature selection to facilitate higher classification accuracies. That study showed that when contrasting land covers through select features gross statistical values, broad class separations were possible. Those results implied that a hierarchical classification strategy could be an effective and efficient approach for classification of that dataset. Furthermore, that study showed that through analysis, an optimal subset of features could be identified in order to most effectively classify those land covers, for each level of the hierarchy, which results in increased accuracy and more efficient computations. With these factors in mind I contended that different data features, and or datasets can be tailored, through analysis, to be used at different stages of classification, within a hierarchy, in order to create superior or more consistent results, when compared to previous studies which have relied strictly on the resolution or characteristics of those inputs to drive splits in those classification

hierarchies [Jiao et al 2019; Koma el al 2020; Mahdianpari et al 2017, 2020; Mohammadimanesh et al 2018, 2020; Munoz 2019].

2.2.2. Feature-Level Fusion for Classification of Remotely Sensed Data

With the wide availability of remotely sensed data, (such as hyper and multispectral sensors, and synthetic aperture radar sensors), data fusion techniques (such as probability theory, evidence theory, fuzzy logic and neural networks methods), have been utilized to improve the classification of remotely sensed imagery [Bo et al 2016; Bui and Mucsi 2022; Chen et al 2015; Hu et al 2019; Imani and Ghassemian 2020; Jia et al 2021; Rasti et al 2017; Zhong 2016]. With feature level fusion, an improved fused image is generated by combining at least two independent data sources, through a mathematical framework such as a Convolutional Neural Network (CNN) [Li et al 2017]. The most notable example is the pan-sharpened image by combining the low-spatial-resolution multi-spectral imagery and the high-spatial resolution panchromatic imagery. In this context of image classification, individual data sources are not explicitly analyzed and potentially not fully utilized with the pixel-level fusion methods. Feature-level fusion involves concatenating sets of features before the classification process. The feature-level fusion is the most commonly used due to its simplicity and demonstrated success [Chen et al 2015; Li et al 2017]. Fusion of optical and SAR data is the most common form of feature level fusion used in wetland mapping and monitoring [Wang et al. 1997; Grenier et al. 2007; Henderson and Lewis 2008; Koch et

al. 2012; Lin and Yue 2014; Hong et al. 2015; Amani et al. 2017]. However, the high dimensionality in the feature space that results from feature-level fusion, even after feature reduction efforts, is likely to be a concern for applications where the size of training samples is small [Maxwell et al 2018]. In addition, features derived from different data sources are usually treated equally by most classifiers (such as RF methods), even though some of the data sources may be more reliable than others [Hi et al 2021].

2.2.3. Decision-Level Fusion for Classification of Remotely Sensed Data

Decision-level fusion involves the merging of decisions from multiple classifiers either with different features or using different classification methods. In contrast to feature-level fusion, for decision-level fusion each data source is analyzed separately, and the uncertainty and imprecision associated with each data source can be measured and considered in the fusion process. The challenge with decision-level fusion lies in the selection of propositions for each data source and effective ways to combine the decisions. Y. Hu et al. [2019] proposed a multi-objective CNN decision-fusion classification method for coastal wetlands with hyperspectral images and found that the decision fusion classification method based on fuzzy membership achieved an overall accuracy of 82.11%. Meng et al. [2019] constructed a hybrid classifier using the decision fusion of CNN and SVM to perform lake wetlands mapping and produced overall classification accuracy of over 90%. Deng et al. [2022] used the maximum probability method to fuse multiple one-class SegNet models

and explored the identification ability of the fused models for karst wetland vegetation, demonstrating that the classification performance of the fused models was better than that of the single SegNet model and achieved over 87% overall accuracy. The above-mentioned studies indicated that the fusion of different classification models can compensate for their respective shortcomings and achieve higher accuracy than any single model. Nevertheless, current studies usually build fusion models only using one CNN algorithm, and there is a lack of research examining the classification performance of fusion models with different CNN algorithms for wetland vegetation. Meanwhile, previous studies focused on a single decision fusion strategy and did not sufficiently consider the applicability and effect of different decision fusion strategies on wetland vegetation classification.

2.3. Deep Learning Methods for Classification of Remotely Sensed Data

Deep learning (DL) is attracting substantial attention in cover type classification using multi-source remotely sensed data including but not limited to wetland classification [Hosseiny et al 2021; Hu et al 2019; O'Neil et al 2020; Jian et al 2019; Yuan et al 2022; Sun et al 2022]. DL classification is inspired by biological learning structures where for primate visual systems, the brain is organized in "deep architecture", where perception is represented in multiple levels of abstraction. Instead of relying on manually engineered features, DL techniques are able to automatically learn informative representations of raw input data with multiple levels of abstraction, either through supervised or unsupervised

styles of classification. Such learned features have achieved success by being used in many machine vision tasks [Krizhevsky et al 2012; Li et al 2017; Ronan el at 2008; Xu et al 2015].

These developed deep learning methods can also be categorized as pixel-level, feature-level, and decision-level fusion. Most of them implement feature-level fusion. The issues discussed earlier on pixel-level and feature-level fusion for classification also apply to those based on deep learning. Recent studies examined the ability of deep learning models and, in particular, the CNN to improve the accuracy of wetland classification [Li et al 2017; Mohammadimanesh 2018; Pouliot 2019; Rezaee et al. 2018]. Generally, CNN models are more common in remote sensing applications than are other deep learning methods such stacked autoencoder (SAE) and deep belief network (DBN) [Li et al 2017 ; Russakovsky et al., 2015; Xia et al., 2016]. This is because CNNs have the capability to maintain spatial relationships in different processing levels, as spatial filtering takes place in each layer without flattening data to a vector form [Li et al 2017; Mahdianpari 2018; Russakovsky et al., 2015; Xia et al., 2016]. Other more advanced DL models are also being developed which could be useful in the field of classification of remotely sensed imagery. In particular the 3D feature fusion framework proposed by Zhu et al [2021] to extract spectral-spatial features via 3D morpholigical profiles, 3D LBPs and 3D gabor surface features may help in the development of more efficient DL classification models for remotely sensed imagery.

Nevertheless, results have shown classification accuracies from DL models are not dramatically different from those using other classification techniques [Li et al 2017] but

have shown more flexibility and adaptability with larger datasets when compared to more conventional classification techniques. With that said, decision-level fusion methods including the one proposed in this study can be used together with deep learning networks.

Also, a number of fundamental challenges remain within the field of DL of remotely sensed imagery. DL techniques generally require very large datasets in terms of available features and training data, and very large computational resources. Additionally, in order to best leverage those datasets often a large number of labeled samples are required, more than traditional classifiers [Li et al 2017]. Finally, the classification of remotely sensed imagery with DL techniques is unique compared to DL with medical images or facial recognition in that RS images are more complex and have larger variations. Remotely sensed images may involve various types of objects, which are also different in size, color, location and rotation, which can complicate the classification and evaluation process.

It should also be mentioned that at the time of this writing artificial intelligence (AI) models are becoming increasingly more popular both inside and outside academia. With the advent of ChatGPT and other AI inspired models, academic researchers are responding to these developments in an effort to better understand how this disruptive technology will affect academic research [Zhu et al. 2023]. Within the context of the research presented in this thesis the ChatGPT style of AI, in principle, could have aided in framing research questions or summarizing large sets of academic articles but in my estimation would not have provided value when it comes to the development of the innovations presented here. The main factor driving this is that ChatGPT style of AI models rely very heavily on

training of language-based data sets where it draws correlations. This means that it will be naturally limited by what has already been presented in the academic literature [Ray 2023], and would not necessarily be a source of innovative ideas beyond what has already been presented. Also, it is worth noting that the advance regression and deep learning models mentioned in this section and throughout this thesis are the core algorithms driving these popular AI models, with the key innovation being the execution of those models on extremely large language bases datasets. In that vein this style of deep learning or AI learning is already an active area of research [Hosseiny et al 2021; Hu et al 2019; Jian et al 2019; Li et al 2017; Mohammadimanesh 2018; O'Neil et al 2020; Pouliot 2019; Rezaee et al. 2018; Sun et al 2022; Russakovsky et al., 2015; Yuan et al 2022; Xia et al., 2016].

Chapter 3 - The Assessment of Data Features and Metrics and Holistic Methodologies in Support of Wetlands Classification

A portion of the research in this chapter is accepted in the following journal paper:

[1]Judah, A., and B. X. Hu. 2019. "The Integration of Multi-source Remotely-Sensed Data in Support of the Classification of Wetlands." *Remote Sensing* 11 (13):28. doi: 10.3390/rs11131537.

---

[1]I would like to thank the publisher MDPI and Prof. Hu who have granted me permission to reuse this article in my dissertation.

In this chapter I will present methodologies which I developed to classify wetlands (Open Bog, Treed Bog, Open Fen, Treed Fen, and Swamps) from multi-source remotely sensed data using advanced classification algorithms. The data utilized included multispectral optical and thermal data (Landsat-5) and Radar imagery from RADARSAT-2 and Sentinel-1. The goals were to determine the best way to combine the aforementioned imagery to classify wetlands, and determine the most significant data features. SVM, in some cases, produced results of comparable or better accuracy than RF. My work also showed that the use of surface temperature (an untraditional feature choice) could aid in the classification process if the image is from an abnormally warm spring. This study found that wetlands were best classified using the NDVI calculated from optical imagery obtained in the spring months, radar backscatter coefficients, surface temperature, and ancillary data such as surface slope, computed through either an RF or SVM classifier. It was also found that preselection of features using Log-normal or RF variable importance analysis was an effective way of identifying low quality features and to a lesser extent features which were of higher quality.

3.1. Study Area and Data Used

The study area, located at approximately $49^0$31’25.34N, $80^0$43’37.04W, was chosen because of the availability of satellite and other geo-spatial data. Figure 3-1 illustrates the study area from a geographic perspective, and a Landsat-5 and aerial imagery perspective.

Landsat-5, RADARSAT-2, and Sentinel-1 imagery were the primary image sources used in this study. The Landsat-5 series of sensors collect multispectral optical imagery with a spatial resolution of 30 m by 30 m and thermal imagery at 120 m by 120 m [Irons et al 2003]. As a point to note, when creating layer stacks of these images for analysis, the

lower resolution (120 m by 120 m) temperature-based images were resampled to 30 m by 30 m.

The RADARSAT-2 imagery product used was a C-band, Wide Fine, SLC (Single Look Complex), quad-polarization image with a spatial resolution of 5.2 m by 7.7 m [RADARSAT-2 Product Format Definition 2011]. However, the features (such as entropy and alpha) derived from the original RADARSAT-2 imagery had a spatial resolution of 12.5 m by 12.5 m. The final step with preparing the RADARSAT-2 imagery was to resample it to 30 m by 30 m to match the resolution of the Landsat-5 imagery. For Sentinel-1 imagery (C-band), the product used was the duel-polarization imagery, and had a resolution of 5 m by 20 m [European Space Agency 2018]. As with the RADARSAT-2 imagery, the Sentinel-1 imagery was resampled to 30 m by 30 m in order to facilitate ease of analysis with the other imagery products. The final imagery product used in this study was the aerial imagery with four channels (590–675 nm, 500–650 nm, 400–580 nm, 675–850 nm) and with a very high resolution (0.4 m by 0.4 m) [Ministry of Natural Resources – Ontario 2012]. It was used for closer examinations of training and validation sites as identified by Ministry of Natural Resources surveys of the area.

Finally, a digital elevation map (DEM) of the study area taken from the Canadian Digital Surface Model [Natural Resources Canada Map Information Branch 2016] at the spatial resolution of 30 m by 30 m and an associated DEM derived slope were used. In total, five different Landsat-5 images, two different RADARSAT-2, and three Sentinel-1 images were collected. Table 3-1 summarizes the dates and types of imagery that were collected for this study.

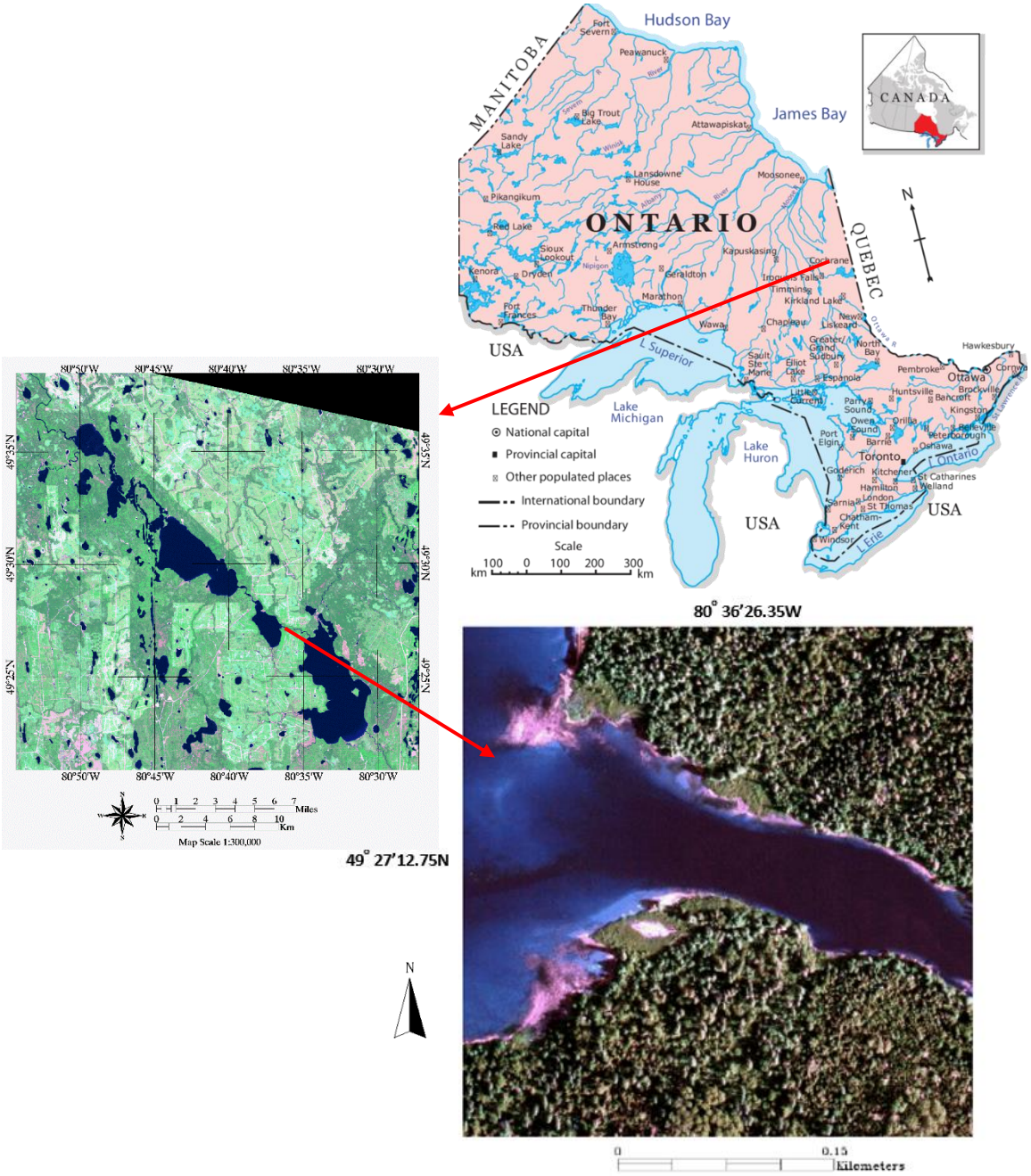

Figure 3-1. The study area from a geographic perspective, and from Landsat-5 and from aerial imagery.

Table 3-1. Summary of remotely-sensed imagery collected for this study.

	Imagery	Season	Date	Level of Processing	Accessed From
1.	Landsat-5	Spring-1	18-May-2010	Level 1G	United States Geological Survey (USGS)
2.	Landsat-5	Spring-2	15-May-2009	Level 1G	USGS
3.	RADARSAT-2	Summer-1	29-June-2014	Level 1—Single Look Complex (SLC)	Canadian Space Agency (CSA)
4.	Landsat-5	Summer-2	24-July-2011	Level 1G	USGS
5.	Sentinel-1	Summer-1	01-August-2015	Level 1—SLC	European Space Agency (ESA) - Sentinel
6.	Sentinel-1	Summer-2	25-August-2015	Level 1—SLC	ESA - Sentinel
7.	RADARSAT-2	Fall-1	10-October-2014	Level 1—SLC	CSA
8.	Landsat-5	Fall-1	09-October-2010	Level 1G	USGS
9.	Landsat-5	Fall-2	10-November- 2010	Level 1G	USGS
10.	Sentinel-1	Fall-1	18-September- 2015	Level 1—SLC	ESA - Sentinel

During covariance analysis of my datasets, it was discovered that inter-season Landsat-5 images were strongly correlated with one another. In an effort to promote better data independence, only a single Landsat-5 image for a particular season was chosen; the Landsat-5 image that produced the highest classification accuracy was selected. The selected Landsat-5 images for testing were Spring-1, Summer-2, and Fall-1, with the Summer-1 and Fall-1 images being selected from the Sentinel-1 images.

Eight different land covers were classified in this study. These land covers were Open Fen, Treed Fen, Open Bog, Treed Bog, Dense Coniferous Forest, Swamps, Grassy Areas, and Cleared Areas. Open Fens are non-treed Grassy areas, with open pools of water. Fens are peat-covered sloping plains or channels with very high water tables and with surface carpets of brown mosses and associated Sphagnum. The average depth to the water table, even in a dry season, is usually less than 20 cm [The Canadian Wetland Classification System 1997]. Treed Fens are fens, as described above, with dense shrubs and tamarack trees. In Northern Ontario, Treed Fens are usually dominated by Black Spruce (Picea

mariana). Treed Fens occur generally throughout the province but most extensively in the Hudson Bay-James Bay Lowlands [The Canadian Wetland Classification System 1997]. Bogs are peat-covered plains or peat-filled depressions with a high water table and a surface carpet of mosses dominated by Sphagnum. In flat or level Bogs, the water may remain at the surface throughout the spring and summer months. Open Bogs that may have a partial cover of stunted trees occur generally throughout the province of Ontario, Canada, but also exist very extensively in the Hudson Bay-James Bay area in Northern Ontario [The Canadian Wetland Classification System 1997]. Treed Bogs are bogs with a low to high density of tree cover. It was expected for there to be some degree of overlap between densely Treed Bog and Sparse Conifer Forest. Treed Bogs are typically dominated by Black Spruce trees. Treed Bogs exist in many parts of the province of Ontario, Canada, but extensively in the Hudson Bay-James Bay Lowlands area in Northern Ontario [The Canadian Wetland Classification System 1997]. Dense Coniferous Forests are large continuous forested areas, composed of at least 80 percent of coniferous species. Dense Coniferous Forest exists throughout the province of Ontario, Canada [The Canadian Wetland Classification System 1997]. Coniferous and deciduous Swamps occur along rivers, and lakes and are characterized by a range of moisture conditions and plant species such as cattails, grasses, and shrubs. The Swamps in Northern Ontario can also have a sparse presence of trees, both coniferous and deciduous [The Canadian Wetland Classification System 1997]. Grassy areas are flat open areas covered almost entirely of grass, colloquially known as meadows or fields. Some of these areas are older cleared areas that are regenerated and are almost entirely covered by tall grasses [The Canadian Wetland Classification System 1997]. Cleared areas are forested areas that are harvested, and are undergoing regeneration. Characterized by very young trees, open areas, low to medium height grasses, shrubs, and bare soil. These areas are generally dry and the soil is of poor nutrient content [The Canadian Wetland Classification System 1997].

3.2. Feature Significance Analysis

The objective of feature significance analysis was to quantify the statistical differences and similarities between land covers for a given feature. The intent of doing this analysis was to aid in determining which features and feature combinations would be desirable when classifying the selected land covers. To accomplish this, two strategies were used. They were the log-normal distance and RF predictor importance value.

In the field of data science, feature selection is an active area of research and is often broken into two branches – supervised and unsupervised [Hopf and Reifenrath 2021, Solario-Fernandez et al 2019]. Supervised methods require a labeled classification output while unsupervised methods use unlabeled data. In both methodologies features are either selected or dropped from either classification performance as specified by user input or from some form of statistical analysis which determines the significance of those features. It is also generally understood that supervised methods are generally more effective in selecting features when compared to unsupervised methods [Hopf and Reifenrath 2021]. When selecting methodologies for feature selection I was interested in exploring the data set from both a classification performance standpoint and a statistical standpoint. From a classification performance standpoint this necessitated selecting a supervised feature selection methodology. Given RF already produces a variable or feature significance value it was natural to incorporate it. Alternatives to this include Information Gain, Chi-Square Test, and Fisher Score. I decided against using these methods because in my opinion, they created extra complexity and possibly noise to the analysis, and would also not perform as effectively as the built in feature significance value calculated from RF, which directly measures feature performance within the classifier. When selecting a statistically based selection method I was interested in highlighting similarities and differences between spectral distributions. Methods such as Kolmogorov-Smirnov and Jensen-Shannon

Divergence are commonly used method in this regard [Poe et al 2005]. However, I decided against using these methods due to the fact that weighting influences outputs, and in my case I was simply looking at spectral similarity or dissimilarity. This drove the selection of the log-normal method.

The log-normal distance is a purely statistically determined value, while the predictor importance value is determined through an iterative exploration of the dataset with an RF classification scheme. By using these two different approaches, it provides me with contrasting statistical perspectives on the dataset and features, which in turn should affect classification results.

The first strategy, given a single feature with multiple land covers, was to measure the log-normal distance between land covers for that feature [Jensen, 2004]. The log-normal distance, in this case, measures the statistical similarity between two sets of data for a given measure where larger values imply dissimilarity between sets, when compared to smaller values. This is defined by:

$$D_{LN}(p,q) = \frac{1}{4} ln\left(\frac{1}{4}\left(\frac{\sigma_p^2}{\sigma_q^2} + \frac{\sigma_q^2}{\sigma_p^2} + 2\right)\right) + \frac{1}{4}\left(\frac{\left(\mu_p - \mu_q\right)^2}{\sigma_p^2 + \sigma_q^2}\right) \qquad (3\text{-}1)$$

where D_{LN} is the log-normal distance between the two classes, σ_p^2 is the variance of the p-th distribution, μ_p is the mean of the p-th distribution and p, q are two different class distributions. With this relationship, we can see that if the distributions are equal to one another, the first term will reduce to $\ln(1) = 0$ and the second term will reduce to $\frac{1}{4}\left(\frac{(0)^2}{\sigma_p^2+\sigma_q^2}\right) = 0$, which will sum to 0 – which mathematically is logical since the distributions are equal to one another. As a corollary, if the distributions are dissimilar in their variance, the first term will be non-zero and if the means are different the second term will be non-zero, with both terms increasing in value as those increase in dissimilarity. As an example, given two land covers, measured by features A, and B, if the log-normal

distance between land covers as measured by A was larger than B, it would imply that A is of a higher quality, compared to B. In other words, A would be a better feature to classify those land covers from one another. For instance given eight land covers classified this would corresponded to 28 unique combinations of land cover pairs to have their log-normal distance calculated for a given input feature. When those results were averaged together, an overall quality factor was produced for that feature. This strategy was executed on all input features.

The second strategy was based on the performance of features when utilized in an RF classification scheme. During the classification process with RF, a predictor importance value can be calculated for each feature input, for that given classification scheme. The predictor importance value was computed by summing changes in MSE due to splits on every predictor and dividing the sum by the number of branch nodes for that tree, averaged over all trees. These calculations are done on all input features, with larger values implying a feature is more important based on its impact on changes to the mean squared error. The objective here is to estimate a single features importance compared to the rest of the input features, using this metric. To accomplish this I ran a series of classification tests (N) where, for each test, a given feature was excluded for that test. In that way, for a given feature, when averaged over its N-1 tests, a metric for how important that feature was when compared to its peers can be computed. The use of predictor importance with the RF classification methodology is a standard approach to evaluate the performance of individual input from a classification result.

3.3. Methodology

The analysis for this study was carried out in five phases. In the first phase, individual samples for each land cover type were identified through Forest Resources of Inventory (FRI) data [Ministry of Natural Resources – Ontario 2012] and they were separated into two subsets for training and evaluation, respectively. In the second phase, the remotely sensed imagery was processed, georeferenced, and prepared for analysis. Relevant features were extracted in this phase as well. In the third phase, feature selection was carried out. Features derived in the second phase were analyzed using the log-normal distance, and an RF generated feature importance parameter based on the sum of changes to the mean squared error (MSE). In the fourth phase, various classification schemes were performed and the classification results were evaluated. For RF classified results, a corresponding 'confidence value' and a corresponding 'confidence map' were produced. Finally, for the fifth phase the best performing classification scheme was used to classify a test area to explore the functionality of that scheme and to provide a visual representation of a classified area. In the following, these phases are described in more detail.

3.3.1. Defining Training and Evaluation Areas

Training and evaluation areas were identified using ground survey data collected during the summers of 2011–2014 by the Ministry of Natural Resources in support of forest inventory resource management [Ministry of Natural Resources – Ontario 2012] and aerial imagery also collected for the Ministry of Natural Resources, as part of its internal inventory and records. Oftentimes, areas were cross referenced with one another for added verification. For the ground survey data, survey areas were defined by 100 m–200 m square

areas where generally 3–4 GPS points are taken to define the extents of those areas. Surveying of those areas followed the Ontario Forest Resource Inventory Calibration Plot Specifications guide [Ministry of Natural Resources – Ontario 2012]. Table 3-2 summarizes the sizes of the training areas (in pixels), and their corresponding evaluation sets. The evaluation and training were sets taken from separate areas to eliminate spatial correlation, which was observed in initial testing, illustrated in Figure 3-2.

Table 3-2. Number of pixels used for different classes in training and testing.

Class	Number Assigned to Class	Number of Pixels in Training Set	Number of Pixels in Validation Set
Open Fen	Class 1	839	624
Treed Fen	Class 2	766	526
Open Bog	Class 3	802	517
Treed Bog	Class 4	868	563
Dense Con. Forest	Class 5	933	672
Swamps	Class 6	881	580
Grassy Areas	Class 7	2541	1331
Cleared Areas	Class 8	3184	2147

The number of pixels for each study area was determined by the size of land cover plots identified through the ground survey data. I attempted to have approximately 60% of the identified pixels be part of the training set, with the remaining 40% be part of the validation set. Based on the boundaries of these land cover plots, a set of contiguous pixels were selected for that individual land cover.

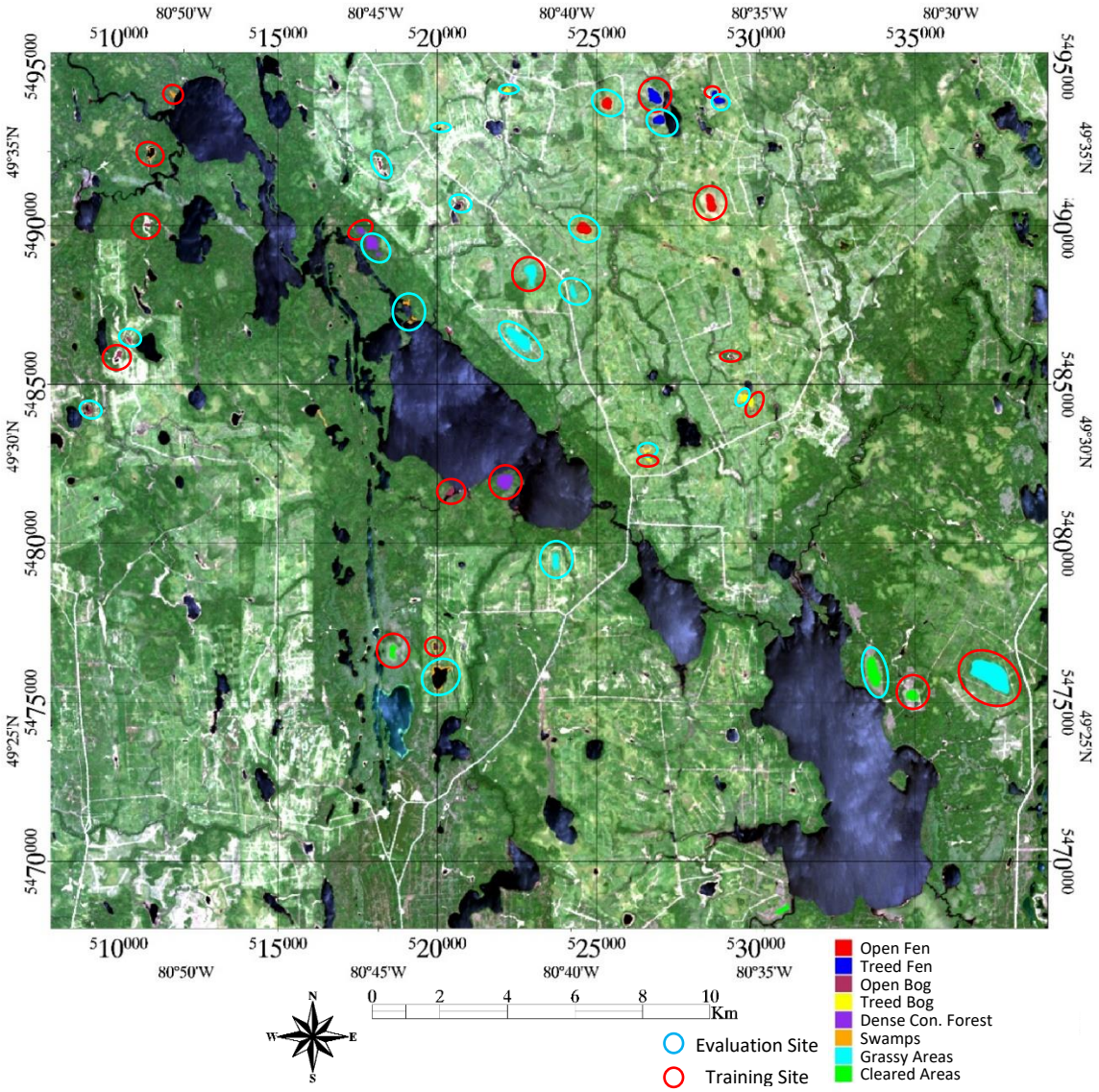

Figure 3-2. True color Landsat-5 image of study area with training (red circles) and evaluation (light blue circles) highlighted for individual land covers.

3.3.2. Image Preprocessing and Feature Selection

For this study, six different image indices or metrics were used: NDVI (The Normalized Difference Vegetation Index), NDWI (The Normalized Difference Water Index), Albedo, Surface Temperature, Alpha, and Entropy. These image metrics were selected due to the fact that they are all popularly used metrics in the analysis of multi-spectral and radar imagery, with the addition of Surface Temperature due to my intuition

that it might prove to be useful when incorporated into the correct classification strategy. Additionally, the DEM, and DEM derived slope were also incorporated into the classification of imagery. DEM and DEM derived slope were selected to determine the role geographic features play in the classification process. For instance, it is known that some species of Fens prefer to grow in slopes. All Landsat-5 imagery used was Level 1G, which are both radiometrically and geometrically corrected. NDVI, NDWI, Albedo, and Surface Temperature were calculated using Landsat-5 based imagery, which, through its multispectral measurements, provides a spectral representation of a surface, for multiple wavelength ranges.

NDVI is a popular vegetation index sensitive to leaf area index, coverage, and pigment content of vegetation canopies vegetative activity photoactivity [Sellers 1985, Richards 1999]. NDVI is defined as:

$$\text{NDVI} = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{RED}} \tag{3-2}$$

where ρ_{NIR} and ρ_{RED} are the reflectances in the near infrared and red band, respectively. NDWI works on a similar principle to NDVI, but is designed to be sensitive to water content rather than to photosynthetic activity. NDWI is defined as:

$$\text{NDWI} = \frac{\rho_{GREEN} - \rho_{MIR}}{\rho_{GREEN} + \rho_{MIR}} \tag{3-3}$$

where ρ_{GREEN} and ρ_{MIR} are the reflectance in the green and middle infrared band (MIR), respectively. In his paper describing NDWI, [Gao 1996] mentions that the green and MIR bands are located in the high reflectance plateau of vegetation canopies; the absorption by vegetation liquid water near the green band is negligible, but weak liquid absorption at MIR is present. Canopy scattering enhances the water absorption and as a result NDWI is sensitive to changes in liquid water content of vegetation canopies. Gao [1996] also argues that the effect of atmospheric aerosol scatter effects in the MIR region are weak; NDWI is

less sensitive to atmospheric-optical depth compared with NDVI. Due to its success in many applications, NDWI is a standard layer product for the Moderate Resolution Imaging Spectroradiometer the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor [Hubanks et al 2008].

Surface albedo is a measure of reflectivity from a surface, which takes on a value from 0 (absorption) to 1 (complete reflectance). A standard approach in determining the surface albedo using Landsat-5 imagery is through a numerically determined relationship described by Liang et al. [1999]. Liang describes albedo α using Landsat-5 TM imagery through the following equation:

$$\alpha = 0.356\alpha_1 + 0.130\alpha_3 + 0.373\alpha_4 + 0.085\alpha_5 + 0.072\alpha_7 - 0.0018 \qquad (3\text{-}4)$$

where in (3-4 the subscript on each α represents a band number in a Landsat-5 TM image. Note that band 6 and the panchromatic band are not present in ((3-4). Studies by Liang found that the removal of band 6 and the panchromatic band did not significantly attenuate results, implying that they are redundant variables, thus suggesting the use of the aforementioned five bands [Liang et al 2000]. A similar expression is created for Landsat-8 imagery where the band indexes are different but cover the same spectral ranges.

The first step in determining the surface temperature for an individual pixel from the Landsat-5 imagery was to calculate the surface radiance from Band 6 (Thermal Infrared). The following equation was used to convert the digital number (DN) of Band 6 into spectral radiance [Landsat 8 data Users Handbook, 2015]:

$$L_\lambda = 0.0370588 \times DN + 3.2 \qquad (3\text{-}5)$$

The next step was to convert the spectral radiance to the brightness temperature (i.e., blackbody temperature) under the assumption of uniform emissivity as shown in (3-6) [Landsat 8 data Users Handbook, 2015]:

$$T_B = \frac{K_2}{ln\left(\frac{K_1}{CV_{R2}} + 1\right)} \tag{3-7}$$

where T_BT_B is the blackbody temperature in kelvin, CV_{R2} is the radiance ($Wm^{-2}sr^{-1}um^{-1}$) at the surface; and K_1 = 607.76 $Wm^{-2}sr^{-1}um^{-1}$ and K_2 = 1260.56 K, which are numerically determined constants [Coll et al., 2010].

During initial examinations of the test imagery, it was noted that surface temperature when plotted against NDVI via a scatter plot, produced several well-defined clusters. These clusters could then be used to quickly classify the source image into two separate classes

Figure 3-3. This helped motivate the exploration of the role that temperature could play in the wetland classification process. Surface temperature is generally not used in the classification of land covers due to its low resolution. However, we contend that with advanced classification methodologies and the needs of specific land cover types, such as wetlands, surface temperature could play a role in improving classification accuracies for this application.

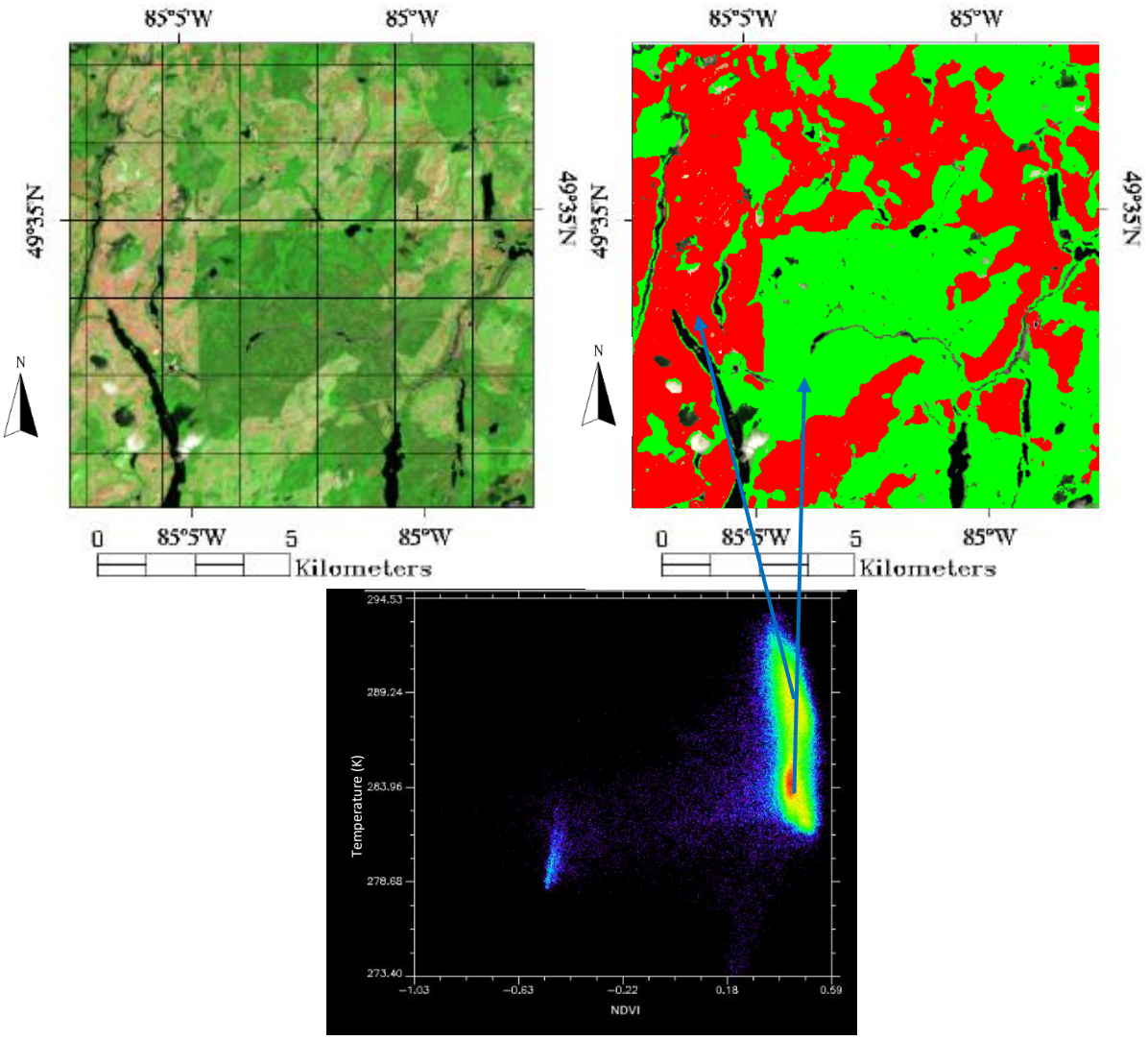

Figure 3-3. Top left: true color image of test area. Bottom: scatter plot of surface temperature versus NDVI. Top Right: Classified image based on the two main clusters produced from the bottom image. Classes are cleared areas (red) and grassy areas with small plants (green).

Alpha and Entropy were calculated from RADARSAT-2 imagery. The RADARSAT-2 imagery used in the study was the Level 1-Single Look Complex (SLC) imagery product. For the RADARSAT-2 images, the Alpha and Entropy values were determined through the European Space Agency software called PolSARpro v4.0 [(ESA, 2008]. PolSARPro also provided the means to initially process the raw RADARSAT-2 images into georeferenced images which could be inputted into other software suites such as ENVI 5.0 [Harris Geospatial, 2009] and Matlab r2016b [Mathworks, 2016]. Given a quad polarized radar image, the backscattered and polarized signal can be decomposed into roll invariant

parameters. Two of which are used frequently in the analysis of RADAR imagery, and are used in the analysis of the RADARSAT-2 imagery here are Alpha ($\bar{\alpha}$)) and Entropy (H). $\bar{\alpha}$ is a measure of the reflected angle of the radar signal, which physically is determined by the angle of incident, surface roughness, and dielectric constant of the reflecting surface [(Pottier et al., 1997]. From a physical standpoint, Entropy can be thought of as a measure of the degree of disorder from the measured reflected quad-polarization radar signal [(Pottier et al., 1997].

From a physical standpoint, $\bar{\alpha}$ provides the nature or the type of dominate scattering mechanism for a given scatter [(Pottier et al., 1997, Pottier et al., 1999]. The scattering nature of a given target can vary among three different categories: isotropic odd bounce ($\bar{\alpha} = 0^o$), dipole or volume bounce ($\alpha = 45^o$), or isotropic even bounce ($\alpha = 90^o$) [Pottier et al., 1997]. Figure 3-4 illustrates a physical interpretation of the alpha scattering mechanism. Scattering from a flat surface will result in $\alpha \simeq 0^o$, scattering from a surface dominated by random scattering medium with cylindrical geometry (such as branches or needles) will result in $\alpha \simeq 45^o$, surfaces which result in double or 'even' bounce scattering events, such as those provided by isolated dielectric and metallic dihedral scatters result in α values closer to 90^o.

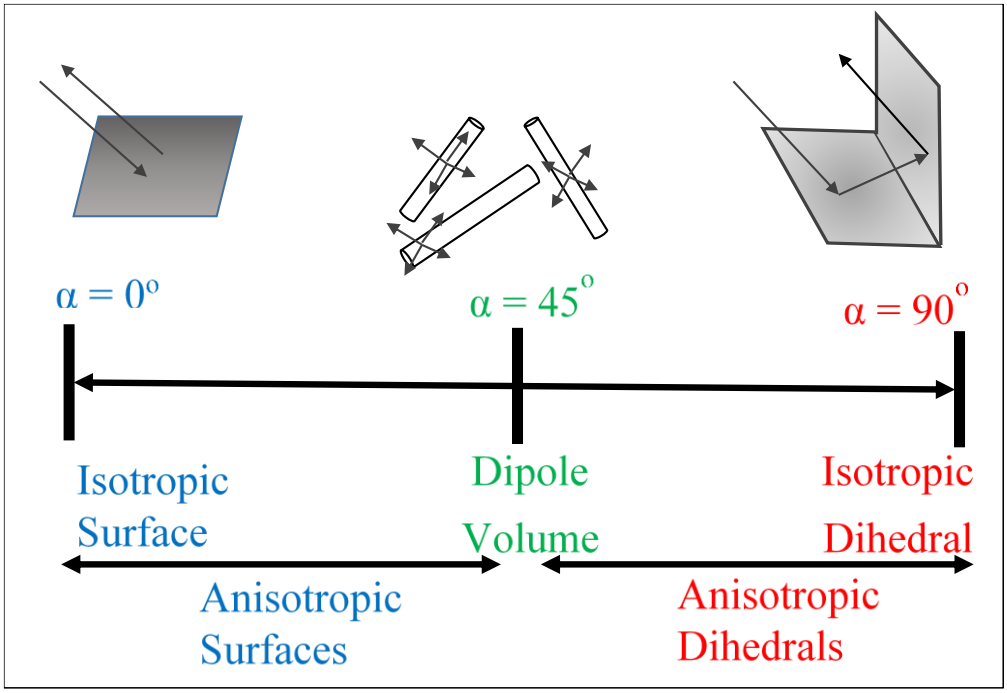

Figure 3-4. Illustration of the dominant scattering mechanisms as expressed by the mean alpha scattering angle. Scattering ranging from α = 0 to α = 90.

For the Sentinel-1 imagery, it was put through a similar georeferencing process as the RADARSAT-2 imagery.

In total, each land cover had a data set corresponding to 48 individual layers, with each layer representing a unique feature: either a spectral band value, an image metric, radar metric or value, digital elevation point, or a slope derived from the digital elevation. These individual features are summarized in Table 3-3. This parsed data will be known as the Master Data Set from hereon.

Table 3-3. Image Input features used during this study and their associated variable index.

Index	Feature Name	Index	Feature Name	Index	Feature Name	Index	Feature Name
1	B1 Reflect. Spring 1	**13**	B1 Reflect. Summer 2	**25**	B1 Reflect. Fall 1	**37**	Sentinel VV - Summer 1
2	B2 Reflect. Spring 1	**14**	B2 Reflect. Summer 2	**26**	B2 Reflect. Fall 1	**38**	Sentinel VH - Summer 1
3	B3 Reflect. Spring 1	**15**	B3 Reflect. Summer 2	**27**	B3 Reflect. Fall 1	**39**	Sentinel VV - Summer 2
4	B4 Reflect. Spring 1	**16**	B4 Reflect. Summer 2	**28**	B4 Reflect. Fall 1	**40**	Sentinel VH - Summer 2
5	B5 Reflect. Spring 1	**17**	B5 Reflect. Summer 2	**29**	B5 Reflect. Fall 1	**41**	Sentinel VV - Fall 1
6	B6 Reflect. Spring 1	**18**	B6 Reflect. Summer 2	**30**	B6 Reflect. Fall 1	**42**	Sentinel VH - Fall 1
7	B7 Reflect. Spring 1	**19**	B7 Reflect. Summer 2	**31**	B7 Reflect. Fall 1	**43**	Digital Elevation Map (DEM)
8	Temp Spring 1	**20**	Temp Summer 1	**32**	Temp Fall 1	**44**	DEM-Slope
9	Temp Spring 2	**21**	Temp Summer 2	**33**	Temp Fall 2	**45**	Entropy Fall
10	NDVI Spring 1	**22**	NDVI Summer 2	**34**	NDVI Fall 1	**46**	Alpha Fall
11	NDWI Spring 1	**23**	NDWI Summer 2	**35**	NDWI Fall 1	**47**	Entropy Summer
12	Albedo Spring 1	**24**	Albedo Summer 2	**36**	Albedo Fall 1	**48**	Alpha Summer

3.3.3. Classification and Feature Selection

The core of this research was the analysis of the master data set utilizing advanced data regression and classification techniques. These techniques have been applied and adapted to multiple fields such as remote sensing, finance, and spam filtering. For these purposes, I trained a classifier using data drawn from my study area, for a given set of features, which then classified a separate set of data, again drawn from the study area, using the same set of features, and then evaluated that classification result and based its producer accuracy and kappa value, which provides an assessment of the resulting accuracy when compared to chance. A higher kappa value implies a higher quality result. For this research, four popular techniques were selected. They are Naïve Bayes, K-NN, SVM, and RF. However, for some of these studies I did not use all of these classifiers. This is clarified in those studies' respective chapters. These techniques are described in more detail below.

The Naïve Bayes classifier assigns observations to the most probable class by estimating the probability densities of the training classes. Classification of an observation is completed by estimating the probability for each class, and then assigning the observation to the class yielding the maximum posterior probability. Unless a probability threshold is incorporated, all inputs are classified [Hastie et al, 2008].

The K-NN classification algorithm operates by finding a group of k objects in a training set that are closest, in feature space, to a provided test object, and bases the assignment of a classification label on the predominance of a particular class in this neighborhood [Haste el al, 2008, Hall et al, 2008]. To classify an unlabeled object, the distance, in feature space, of this object to each labeled object is computed. The K nearest neighbors of the unlabeled object are identified and the class labels of these K nearest neighbors are then used to predict the class label of the object.

SVM is a binary classification methodology that separates classes by fitting a hyperplane between two sets of data. The optimization of this fitting is determined by "maximum-margin hyperplane" that divides a group of points such that each point distance from the hyperplane is maximized [Hastie et al, 2008, Kecman et al, 2005]. Even though this methodology is binary in nature, it can be used in to classify multiple classes through an adoption of a one versus one (OvO) classification strategy. I adopted this strategy in this study. In an SVM-OvO classification strategy, n classes are parsed into n(n-1)/2 binary classifiers—essentially an ensemble classification method.

The RF classifier is an ensemble learning method and operates by constructing a multitude of decision trees with the ultimate class of a given input determined by the mode of the classes from those decision trees [Biau and Scornet 2015; Breiman 2001; Kecman et al 2005]. With RF, the diversity of the decision trees is accomplished by making them grow from different training data subsets created through bagging or bootstrap aggregating [Breiman, 2001]. RF lends itself well to parallelization and investigating the nuances of

large datasets. As a result, RF has become one of the most successful and widely implemented data mining methodologies to date [Biau 2015; Breiman 2001]. For this reason, it was chosen as the main classification methodology for this research. Finally, the two main input parameters needed to run the RF classifier were the number of trees and the depth or complexity of those trees. Choosing too few trees results in lower accuracies, while choosing too many trees results in no accuracy gain for extra computations. Additionally, choosing a tree depth that is too shallow tends to produce trees that underfit, while choosing trees that are too deep will overfit the data. In order to determine the right settings for my data, I utilized a built-in Matlab function that will optimize these features given an RF input, as a function. From these experiments, I determined to choose 150 trees to "grow" and have a p-value of 0.05 as the minimum value for the curvature test, which is utilized with the RF classifier to determine when to terminate a split. Using this type of technique to determine RF input parameters is considered to be a standard approach [Biau and Scornet 2015; Loh 2002].

3.3.4. Classification and Evaluation

Once the features were selected based on the training data sets, they were used in the classification for the test set drawn from my study area for visualization purposes and to explore the functionality of the classifier. Given that RF classifies an unknown pixel via a majority voting criteria, in addition to the class category, a confidence value was also calculated for each pixel. The confidence value represented the percentage of the votes the chosen class represented with a higher value representing a higher confidence for result.

3.4. Results

3.4.1. Results – Data feature and Metric Significance

Given a feature and eight land cover classes, 28 unique combinations of land cover pairs were created with an associated log-normal distance. By averaging these results together, an average log-normal distance for that feature was obtained. A larger value implied that feature could play a more significant role in the classification of those land covers compared with features which had lower log-normal values. Additionally, given the 48 features and eight land cover classes, using the RF computed predictor importance values, executed with the strategy described in Section 2.3, the importance of a given feature could be determined. Like the log-normal values, larger importance values implied that a given feature was more valuable in the classification process, and when utilized, would produce more accurate results. The results of the feature importance computations are summarized in Table 3-4.

Table 3-4. (Top) Summary of data feature quality analysis averaged over all land cover types. Image metrics, image bands, radar parameters, DEM and slope sorted from largest to smallest log-normal distance, with a longer distance implying a higher quality. (Bottom) Summary of data feature importance values computed through and averaged over 47 unique RF tests.

Log-Normal Distance

Bands	Distance	Metrics	Distance	Radar and DEM	Distance
B4 Reflect. Summer 2	1.876	NDVI Summer 2	1.826	DEM	1.311
B7 Reflect. Spring 1	1.473	Temp Summer 2	1.507	Entrophy Fall	0.556
B5 Reflect. Spring 1	1.187	NDVI Spring 1	1.400	Alpha Fall	0.530
B6 Reflect. Spring 1	1.160	Temp Spring 1	1.374	DEM-Slope	0.426
B4 Reflect. Reflect. Fall 1	1.140	Temp Summer 1	1.370	Alpha Summer	0.133
B7 Reflect. Summer 2	1.069	Temp Spring 2	1.355	Entrophy Summer	0.101
B2 Reflect. Fall 1	0.965	Albedo Summer 2	1.339	Sentinel VH - Summer 1	0.047
B6 Reflect. Summer 2	0.932	Temp Fall 2	1.220	Sentinel VV - Summer 1	0.043
B4 Reflect. Reflect. Spring 1	0.910	NDVI Fall 1	1.219	Sentinel VH - Fall 1	0.030
B1 Reflect. Fall 1	0.889	Temp Fall 1	1.210	Sentinel VV - Fall 1	0.030
B7 Reflect. Fall 1	0.884	NDWI Spring 1	1.086	Sentinel VH - Summer 2	0.027
B2 Reflect. Summer 2	0.881	Albedo Fall 1	0.866	Sentinel VV - Summer 2	0.025
B5 Reflect. Summer 2	0.874	Albedo Spring 1	0.863		
B6 Reflect. Fall 1	0.861	NDWI Summer 2	0.859		
B3 Reflect. Summer 2	0.853	NDWI Fall 1	0.607		
B2 Reflect. Spring 1	0.823				
B1 Reflect. Spring 1	0.774				
B3 Reflect. Fall 1	0.712				
B3 Reflect. Spring 1	0.679				
B1 Reflect. Summer 2	0.658				
B5 Reflect. Fall 1	0.647				
Average	0.964	Average	1.207	Average	0.272
Standard Deviation	0.282	Standard Deviation	0.298	Standard Deviation	0.370

Random Forest Determined Predictor Importance

Bands	Distance	Metrics	Distance	Radar and DEM	Distance
B1 Reflect. Fall 1	1.083	NDVI Fall 1	1.654	DEM	1.350
B5 Reflect. Spring 1	0.945	NDVI Summer 2	1.450	DEM-Slope	0.810
B4 Reflect. Summer 2	0.891	Temp Spring 1	1.262	Entrophy Fall	0.800
B1 Reflect. Summer 2	0.827	NDVI Spring 1	1.261	Alpha Summer	0.770
B4 Reflect. Fall 1	0.821	Temp Spring 2	1.246	Entrophy Summer	0.748
B7 Reflect. Summer 2	0.799	Temp Summer 1	1.139	Alpha Fall	0.588
B7 Reflect. Fall 1	0.745	NDWI Fall 1	1.126	Sentinel VH - Summer 2	0.471
B2 Reflect. Summer 2	0.725	Temp Fall 1	0.980	Sentinel VV - Fall 1	0.430
B2 Reflect. Fall 1	0.717	NDWI Spring 1	0.883	Sentinel VV - Summer 2	0.394
B7 Reflect. Spring 1	0.700	Temp Summer 2	0.865	Sentinel VH - Fall 1	0.376
B6 Reflect. Spring 1	0.682	Temp Fall 2	0.848	Sentinel VV - Summer 1	0.334
B1 Reflect. Spring 1	0.665	NDWI Summer 2	0.655	Sentinel VH - Summer 1	0.302
B4 Reflect. Spring 1	0.620	Albedo Summer 2	0.605		
B3 Reflect. Summer 2	0.600	Albedo Spring 1	0.583		
B2 Reflect. Spring 1	0.597	Albedo Fall 1	0.435		
B3 Reflect. Spring 1	0.588				
B6 Reflect. Summer 2	0.563				
B6 Reflect. Fall 1	0.557				
B5 Reflect. Fall 1	0.544				
B5 Reflect. Summer 2	0.502				
B3 Reflect. Fall 1	0.356				
Average	0.692	Average	0.999	Average	0.614
Standard Deviation	0.161	Standard Deviation	0.335	Standard Deviation	0.288

From the results in Table 3-4, it is noted that the features calculated from multispectral imagery, on average, were of the highest quality and importance, with traditional metrics such as NDVI performing well, when measured by both the log-normal and RF determined predictor importance values. In addition, the features calculated from data acquired in the spring and summer, was of a higher quality compared to the fall according to the log-normal results. However, according to the RF determined predictor importance values, there was no clear preference among the data acquired in different seasons; the metrics associated with fall, summer, and spring all ranked highly. It is also noted that surface temperature, traditionally a feature not associated with wetland land cover classification, was ranked fairly high by both feature analysis methodologies. These results also implied that the collected surface temperature data, despite its low resolution, was of a high enough quality that it could be useful for land cover classification. This was an unexpected result but also was in line with some of my early classification experiments, which showed that temperature could be useful in some circumstances. Finally, the features derived from Radar data and DEM were of a significantly lower quality compared with those derived from optical and thermal data based on the log-normal method. Similar results were obtained using the RF method. However, the difference (in magnitude) was not as large. Among the features from the Radar data and DEM, several of them, namely, DEM, slope, the entropy in the fall season, and the alpha in the summer season, were ranked similar by both methods.

3.4.2. Results – Classification

The classification results from the four classification methods and the 225 feature tests were computed on a desktop computer equipped with an AMD Ryzen 5 26000 Six-Core Processor with 32 gigabytes of RAM, analyzed, and ranked. From these 225 tests, the top

20 and bottom 20 results were extracted, and overall statistics for these tests, for each classification technique was calculated. These results are presented and summarized in Table 3-5 and the table in

Appendix 2.

Table 3-5. Summary of classification accuracy of different methodologies averaged over all classification strategies.

	Average	Standard Deviation	Max	Min	Ave. Kappa Value	Difference from Top to 25th Percentile	Difference from Top to 50th Percentile	Computation Time (s)
Random Forest Classifier	0.7127	0.0838	0.8751	0.3608	0.7341	0.1101	0.1528	2314
Support Vector Machine	0.6528	0.0785	0.8004	0.3979	0.7171	0.0891	0.1424	838
K-Nearest Neighbor Classifier	0.6210	0.0791	0.7746	0.2217	0.7218	0.1002	0.1407	364
Naive Bayes	0.5913	0.0643	0.6998	0.3105	0.7025	0.0567	0.0913	26

From Table 3-5, RF on average produced the most accurate results given all inputs scenarios, followed by SVM, K-NN and Naïve Bayes. It is also worth noting that the highest ranked test, one produced by RF was some 7 percent higher than its closest rival. Additionally, average Kappa values are consistent and are of a magnitude which imply that classification results are of a good agreement between producer and user accuracy. According to the results in Appendix 2A, the effects of input features on the classification accuracies varied among classification techniques and there were no clear set of metrics which consistently outperformed others. However, for individual classification techniques, it would appear that there was a performance preference for certain feature inputs. Parsing this further, I can generalize for each classification methodology the preferred input features which produced the highest classification results. These results are summarized in Table 3-6.

Table 3-6. Summary of best overall performing input features for individual classification methodologies.

	Random Forest	Support Vector	K-Nearest Neighbor	Naive Bayes

	Classifier	Machine	Classifier	
Image Reflectance	X	X	X	X
NDVI	X		X	X
NDWI			X	
Albedo				
Surface Temperature	X	X	X	X
RADARSAT-2 – Alpha and Entropy	X	X		
Sentinel-1 - Backscatter		X		
DEM		X		X
DEM - Slope	X		X	
Spring				
Summer				
Fall				
All season in Combination	X	X	X	X

A common theme from Appendix 2A and Table 3-6 was a preference for incorporating all seasons, surface temperature, and radar-based images into the analysis for the best performing classification methodologies (RF and SVM). Table 3-6 also shows that for all classification methodologies image reflectance data from all seasons, used in combination, is a high performer. Furthermore, among the image metrics, NDVI performs well, for three of the four classifiers. It is also noteworthy that NDWI only performed well with one classifier and surface albedo was not found to be a significant feature. It is also noted that there was a correlation between the number of features and the overall classification accuracy. More features generally resulted in higher classification accuracy; however, the highest ranked tests for all classification methodologies did not contain the most features. Additionally, one may note some other interesting peculiarities with the results presented in Appendix 2A and Table 3-6. An expected result was to see that feature inputs, selected due to their high quality or importance, would result in higher classification accuracies compared to results from inputs selected by a holistic approach. However, this was found to not always be true. With the exception of the K-NN Classifier, of remaining classifiers, the

vast majority top ranked tests were tests determined through a holistic approach. This is counter intuitive, and expanded on further in the discussion section.

For the bottom ranks results, I summarize the common features and themes in Table 3-7.

Table 3-7. Summary of the lowest performing input feature combinations for individual classification methodologies.

	Random Forest Classifier	Support Vector Machine	K-Nearest Neighbor Classifier	Naive Bayes
Image Reflectance				
NDVI	Fall data only	Fall data only	Fall data only	Spring and Fall data
NDWI	Fall data only	Fall data only	Fall data only	Spring and Fall data
Albedo	Fall data only	Fall data only	Fall data only	Spring and Fall data
Surface Temperature	Fall data only			
RADARSAT-2		Fall data only	Fall data only	
Sentinel-1			Fall data only	
DEM				X
DEM—Slope				X
Spring				
Summer				
Fall	X	X	X	X
All seasons in Combination				

From Table 3-7, it is noted that the worst performing results were from image metrics taken from falls scenes. This was common among all classification methodologies and classification structures. This was not unexpected given that during the fall scenes vegetation activity and temperature variations would be at a minimum, making it difficult to discern one land cover from another. Additionally, it is noted that the classification tests with the poorest accuracy were all tests from the worst performing features as measured from feature analysis. In fact, the lowest quality or the least important feature combinations were consistently in the bottom 30 percent of all tests—the expected result. However, it was noted that for the K-NN tests the bottom 50 percent of tests were all tests determined

through feature analysis, rather than the holistic approach, which was not always true for the other classification methodologies.

Regarding the best overall classification performance, the RF classification methodology using image bands, radar, slope, and surface temperature from multiple seasons, produced the best classification result (87.51%). Intuitively this was in line with the operation of the RF classifier which exceled when using large datasets, and when provided with similar inputs, RF generally outperformed other classification methodologies. However, it is worth mentioning that the OvO application of SVM produced results which also outperformed the other classification methodologies by a margin between 3-6% for averaged results. I explore these results further in the discussion section.

Additionally, to better examine the best performing classification result I present its corresponding confusion matrix in Table 3-8.

Table 3-8. Confusion matrix of the best performing classification test (RF test #77).

	Grassy Area	Treed Bog	Cleared Area	Treed Fen	Open Bog	Open Fen	Swamps	Coniferous Forest	Producer Accuracy	User Accuracy
Grassy Area	582	0	0	39	0	0	0	0	0.937	0.909
Treed Bog	56	247	7	5	0	0	0	0	0.784	0.969
Cleared Area	0	0	219	0	0	157	0	0	0.582	0.969
Treed Fen	2	8	0	299	0	0	12	0	0.932	0.872
Open Bog	0	0	0	0	794	84	0	0	0.904	0.951
Open Fen	0	0	0	0	41	341	0	0	0.893	0.580
Swamps	0	0	0	0	0	6	3109	4	0.997	0.985
Coniferous Forest	0	0	0	0	0	0	35	1142	0.970	0.997

When examining Table 3-8, I note that Cleared Areas and Open Fens have the biggest discrepancy. In fact, its producer accuracy is 58.2%. If this result could be improved to be more comparable with the other classification results, it could produce an even stronger classification result. Additionally, as a comparison, I examined the confusion matrix of the worst performing result in Table 3-9.

Table 3-9. Confusion matrix of the worst performing classification test (Naïve Bayes test #225).

	Grassy Area	Treed Bog	Cleared Area	Treed Fen	Open Bog	Open Fen	Swamps	Coniferous Forest	Producer Accuracy	User Accuracy
Grassy Area	206	26	85	56	0	51	167	30	0.332	0.516
Treed Bog	63	50	27	29	0	27	89	30	0.159	0.211
Cleared Area	9	15	40	15	19	28	162	88	0.106	0.145
Treed Fen	50	32	16	28	7	12	131	45	0.087	0.130
Open Bog	2	7	9	3	387	27	373	70	0.441	0.385
Open Fen	20	10	56	17	72	55	106	46	0.144	0.154
Swamps	10	49	27	29	485	105	1871	539	0.601	0.533
Coniferous Forest	39	48	16	38	35	51	614	340	0.288	0.286

When examining Table 3-9, it can be immediately see the contrast in classified results compared to Table 3-9. For all land covers, there is a great deal of misclassification, with some results producing an almost even distribution across all land covers (no better than guesswork). It is noted that the best classified land covers are Swamps, with a producer classification accuracy of ~60%. The worst performing land cover (Treed Fen) has a producer accuracy of ~8.7%. It can also be noted that five of the land covers have a producer accuracy below 30%

As a final examination of the classification results, I examined the average rank of a given test input averaged over the four classification methodologies. The objective was to determine a given classification inputs overall performance across all of the given classification methodologies. I also calculated the standard deviation for that given classification test across the different classification methodologies, in order to gain a sense of the spread of the distribution of those classification ranking results. A scatter plot of these results is presented in Figure 3-5.

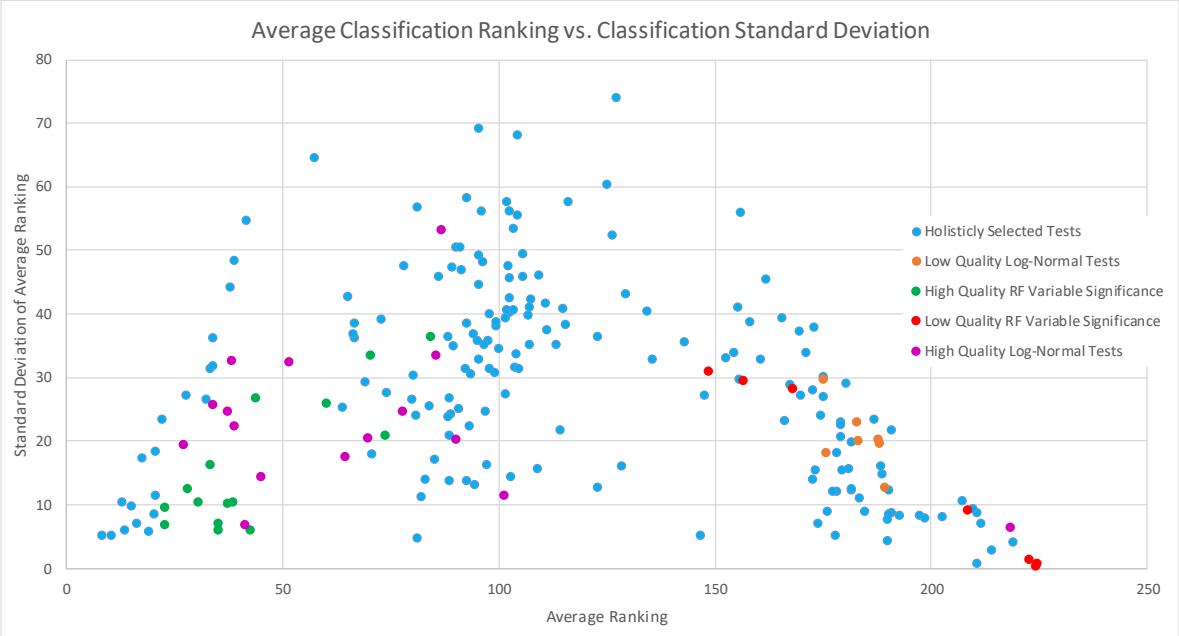

Figure 3-5. Scatter plot of average classification ranking versus classification standard deviation for each test. Results are broken down by classification input selection type.

Given that the highest ranking a test could achieve would be 1 and the lowest ranking a test could achieve would be 225 (the total number of tests conducted), I can interpret Figure 3-5 by noting that the highest quality results would be at the origin and the lowest quality results would be further down the x-axis and up the y-axis. When examining Figure 3-5 we note that results with the highest accuracy had lower spreads compared to results which were of lower quality; however, results of the lowest quality had similarly tight spreads with their distributions. These results imply that tests which produced the highest accuracies would tend to be similarly accurate across classification strategies, and alternatively, classification tests which were of lower accuracy would be of similarly lower accuracy across different classification strategies. Moreover, for tests that were of average accuracy, have large variations in accuracies across classification methods. Finally, according to Figure 3-5 feature inputs selected by their performance from feature significance analysis were generally of a higher accuracy and lower deviation when compared to results selected by a holistic approach, and alternatively, feature inputs indicated to be of lower quality and significance produce consistently lower accuracy results across all classification strategies.

As a test to explore the functionality of a classification scheme produced from this study, the most accurate classification scheme, produced through an RF classifier (test #77), was adapted to classify a test area from the main study area. This test area was chosen such that it did not contain any data drawn from the training or validation data and appeared to contain wetlands of varying types (identified through visual interpretation). The classification inputs were image bands, Temp, Radar (Alpha), and slope from all seasons, with open water, such as rivers, lakes, etc. masked out of the test image. Figure 3-6 contains three images, which are typical of the output from this classification scheme. Figure 3-6A), is a true color Landsat-5 image of a test area, Figure 3-6B) is the actual classification result. Figure 3-6C) is a ‘confidence map’ of the classification result, where 0

indicates low confidence and 1 high confidence. When examining Figure 3-6C) it should be noted that cleared areas, roads, shrubs and grass, were of low confidence, while wet areas or dense wooded areas were of high confidence.

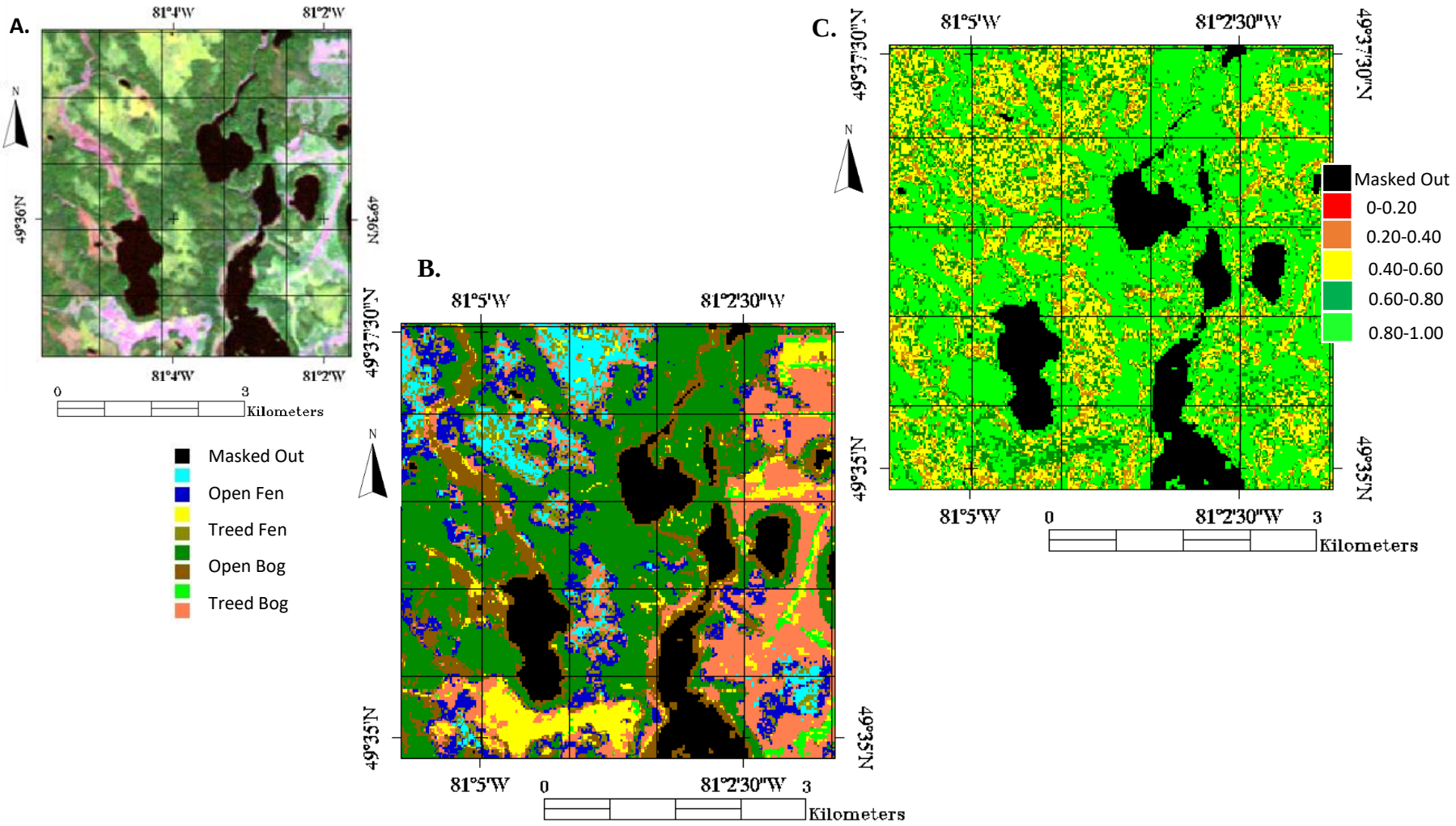

Figure 3-6. Output from classification result. (A) True color Landsat-5 image from a given test area. (B) Resulting classification result. (C) Confidence map of the classification result.

3.5. Discussion

From the feature importance analysis, I generally found that features which were ranked highly from this analysis correlated to higher ranked classification results. However, I noted that this performance varied among classifiers. The K-NN classifier benefited the most from selecting input features from feature analysis—more than half of the top 20 ranked classification results were all from tests derived from feature analysis. Alternatively, most of the worst ranked tests as produced from the K-NN classifier were from the lowest ranked features. Delving further into these results, I note that K-NN operates by finding a group of k-objects that were closest to a provided test object—in essence, its distance in some defined feature space. In that way, this algorithm would both benefit and be disadvantaged more by numerical similarities or differences in its inputs, compared to the other classification methodologies used, which, arguably, use a more gross statistical examination of the datasets or negates these issues through a more thorough examination of the datasets.

RF, the closest to K-NN's from a mathematical and algorithmic standpoint, had only three out of 20 of its top ranked tests coming from tests created from selected inputs from feature analysis, as opposed to 11 for K-NN. However, I do note that for the top 25 percent of tests classified by RF close to half of these tests were tests determined by feature selection analysis. This implies that while the highest ranked tests for RF might be selected through a holistic methodology, overall, selecting inputs from feature analysis is beneficial but not as beneficial when compared to the K-NN classifier. I reason that these differences could be accounted for by several factors which broadly differentiate how RF classifies a dataset from K-NN. Given that the log-normal feature analysis methodology provides a somewhat gross statistical interpretation of the inputs, and assumes that the data is not bi-modally distributed, it would not explore these

subsets within the data, if present, which could otherwise be helpful in the classification process when inputted into an RF classifier. Furthermore, even with the RF determined feature importance values, this style of analysis, while it utilized the RF classifier, I implementation of it still provided a somewhat gross perspective on the performance of these features. It means that the higher performance of a given feature when used in conjunction with other features was not examined from my testing. This could explain why, for RF, the highest accuracy tests were holistically determined tests rather than tests determined through feature analysis. However, the top quartile of tests were still highly represented by tests determined through feature analysis, implying that feature selection, overall, did provide value in the selection of sets of features for an RF based classifier, but in this context also did not provide the most accurate results. Furthermore, like the K-NN classifier, for the RF classifier, the poorest quality or least significant features all performed poorly, as expected.

For the SVM produced results, I note that out of the top 20 tests, only one was from features selected through feature significance analysis. This test was a hybrid test of the top 20 features and was 5 percent less accurate than the top result. However, I also note that for the top-quartile of tests some 26% of those tests were represented by tests selected by feature significance analysis, implying that tests determined by feature significance analysis could produce higher quality results for SVM. Additionally, I note that the tests created through the selection of features via RF feature importance produced, overall, better results compared to results determined by Log-normal distance analysis. This is an unexpected result, given that SVM operates by fitting a hyper-plane between inputs. By this measure, inputs which were further statistically separated should be of more significance, and thus higher accuracy. I speculate that the higher sensitivity to RF importance determined inputs was related to the fact that the SVM

was executed via an OvO approach. In this way, the SVM classifier was being executed in an ensemble fashion, not unlike the RF classifier, where it was likely that some of the 'trees' being grown in the RF classifier would be very similar to the ensemble results produced by the SVM. In other words, features and feature combinations which were significant to RF would also be significant to execution of SVM.

From the Naive-Bayes classification results, I note that from the top 20 ranked feature tests only three were from tests derived by feature significance analysis. Examining these results further, we note that the distribution of feature analysis derived tests were more even compared to the other three classification methodologies with higher quality or significant feature tests ranked in the top half of tests and lower quality or less significant feature tests ranked in the bottom half of tests. Given that Naïve-Bayes classifies through a Gaussian based probabilistic methodology, it would be expected that feature combinations determined through Log-normal analysis would produce the most accurate results, which was not the case. However, I note that the difference between the top ranked classification result and the 25^{th} percentile test was only ~5 percent, and the difference between the top ranked result and the bottom 50^{th} percentile result was only ~9 percent. This implies that the Naïve-Bayes results were closer in distribution and less sensitive to feature inputs but still benefited from the application of feature analysis, just not as dramatically as the other classification methodologies.

When examining all of these classification results from a more gross perspective in the form of Figure 3-5, I note that feature analysis both aided in determining which features can benefit and can be detrimental to classification. When examining both ends of the scatter plot, I note that it trends towards a decrease in distribution of standard deviation. This implies that for high and low ranked tests, the features used in those tests, generally perform the same across all

classification methodologies. Further to that, feature combinations that were predicted to do poorly, did perform poorly across all classification methodologies. Furthermore, feature combinations that were predicted to perform well generally produced higher accuracies, with consistency across all classification methodologies. It is also worth noting that high quality and low-quality feature selections were all ranked in either the top half or bottom half of the distribution, respectively, which implies that my selection methodology is working as designed. Finally, feature combinations that produce mediocre classification results also had large variability between classification methods, which implies that this style of analysis and selection does not have the same level of impact on average results compared to high or low performing results. As an overall take-away from Figure 3-5 , I assert that feature significance analysis could aid in identifying which features can both aid and be detrimental to classification, with the identification of lower quality features and feature combinations showing the strongest relationship across all classification methodologies.

Exploring the most successful features in more detail, I note that the addition of surface temperature, RADAR features, and DEM derived attributes, to the features derived from optical images, overall, increased classification accuracy. The most accurate classification results were generated from using optical data from more than one season and the addition of surface temperature and RADAR features. For individual seasons, classification using the data from the spring and summer season generally outperformed that using the fall season. When considering only individual seasons, classification using the data from the spring season usually produced better classification results than the summer and fall season. I speculate that this was due in part to the increase in vegetative driven spectral overlap seen during the summer months, and the slowing and decay of vegetative activity during the fall. It should also be noted that the 2010

spring season, for the study area, was abnormally warm. Temperature records from the area indicated that the air temperature for that particular image, at collection time, was over 300 K, 5-8 degrees higher than historical seasonal averages [Environment Canada 2017] and the recorded surface temperatures, in some cases, was well over 300 K, about 8-10 degrees warmer than temperatures recorded from the 2009 Landsat-5 image from a similar time of the year. I speculate that these higher temperatures and the incomplete seasonal growth aided in classification by further separating class differences for the spring scene. To explore the temperature results further, I produce Figure 3-7. From Figure 3-7 for the spring scene, I note that Grassy and Cleared areas had some of the lowest temperatures recorded, which was counter intuitive. The expected result would be that Grassy and Cleared areas would be higher in temperature compared to wetlands due to lower moisture content, and thermal inertia. However, if I consider that the vegetation was still developing and the land was still warming from the winter months, this could account for some of these observed differences in the distributions of land cover temperature. Furthermore, for the summer season we noted that temperatures for coniferous forests and Swamps had the lowest temperatures. For coniferous forests, the lower temperature could be attributed to the evapotranspiration effect produced in the needles of trees and leaves of other vegetation in that area. Similarly, Swamps, would have an equally profound evapotranspiration effect from their aquatic plant life, and the very high water content of the land cover which would cause the areas to be naturally cooler than dry land. Grassy and cleared areas measured the highest temperatures. These higher temperatures could be contributed by the relatively low water content compared to the aforementioned land covers, which resulted in lower evapotranspiration, and thermal inertia. The lower evapotranspiration produced lower latent cooling of the surface and the lower thermal inertia resulted in the land cover warming

more quickly compared to the relatively moister wetland land covers. Fen and bog land covers were ranked in the mid-range of summer temperatures, which might be driven by the relatively higher water content compared to the Grassy and Cleared areas which resulted in higher thermal inertia and slower heating and lower comparable temperatures.

Overall, despite its low resolution, temperature showed itself to be a feature which could be used to increase classification accuracy when used in conjunction with other features, with temperature based class differences found to be both physical and logical.

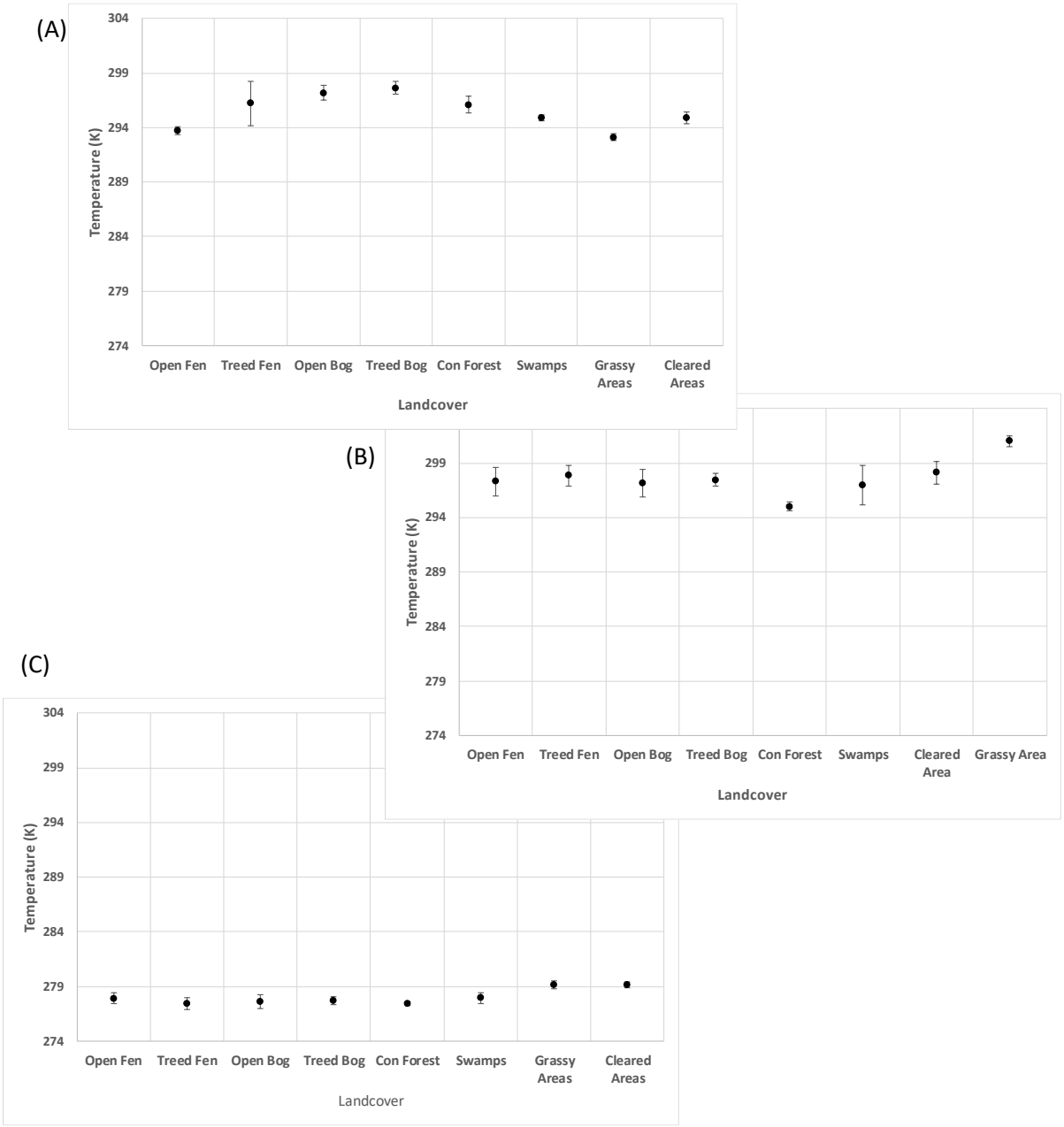

Figure 3-7. Plots of land covers versus temperature, Error bars represent 1 standard deviation. (A) Spring-1, (B) Summer, (C) Fall-1.

Regarding the addition of Radar features to the classification process, addition of RADARSAT-2, and/or Sentinel-1 imagery to Landsat-5 imagery was shown to improve overall classification accuracies by 2-6%, when compared to an input lacking those measurements, when using an RF classifier. Furthermore, using a combination of different seasons and features

produced the higher accuracies across all classification methodologies and schemes. For instance, given only the spring Landsat-5 data, when classified in an RF classifier, produced an accuracy of ~72%. When spring data was used in conduction with data from the summer and fall, in an RF classifier, the classification accuracy jumped ~81%—a 9% increase. Examining these results from a more physical standpoint, it was noted that since the intensity and scatter of the Radar signal is dependent on structural features of the measured surface, treed areas would have different scattering profiles compared to wetland types which do not have large and tall vegetative structures. The addition of these measurements would enhance the depth of the input dataset and thus the overall accuracy of the classification result. Moreover, it was found that DEM and DEM derived slope were significant features in the separation of wetlands from non-wetland classes. We speculate that this was driven by the fact that wetlands were generally flatter, due to the collection of water, when compared to other land cover types where terrain could vary significantly.

Examining the classification results from an overall perspective, I would like to note that during preliminary testing, training and validation sites were chosen randomly, from a pixel standpoint, from a base set and it was found that classification accuracies were, in some cases, over 98 percent as produced from some RF Classification tests. It was suspected that this extremely high accuracy was caused by the random sampling masking spatially driven differences from the training and evaluation sets, in effect the methodology was “over fitting” the dataset. This phenomenon of overfitting is a common and well known within the data science field. Furthermore, I suspect that this phenomenon was responsible for the very high classification accuracies presented in some papers utilizing these styles of algorithms to classify remotely-sensed imagery [Mellor et al 2012; Ming et al 2016; Tian et al 2016]. With

classification methodologies such as RF, the training sets are “learned” thoroughly. If the training and validation sets both have similar spatial representation, it is possible to achieve very high accuracies which may not necessarily be representative of true accuracies if given inputs from similar but spatially different areas. This has motivated me to use spatially separated training and evaluation data sets, which has reduced the overall accuracy of my results, but I believe is now producing results which are more representative of results which would be produced when these classifiers are applied to other study areas—the ultimate goal of this research. However, it should be noted that results produced by Naïve Bayes were not significantly affected by these spatial correlations. This represents how Naïve Bayes uses a more gross statistical representation of the training data compared to RF, SVM–OvO, and K-NN methods.

From an overall performance standpoint, the RF Classification methodology outperformed all other classification methodologies. RF classification, while more computationally intense compared to the other classification methods used in this study, outperformed its closest competitor by 8 percent. Additionally, upon closer examination of the best performing classification result (RF test #77 - Table 3-8), it was noted that the classification of cleared areas did rather poorly (producer accuracy of 58.2%). This also resulted in a poor user accuracy of Open Fens (58.0%) as illustrated in Table 3-8 . Despite this the classification of the rest of the land covers performed very well. I speculate that the misclassification between Cleared Areas and Open Fens lays within the image reflectance and spectral overlap between the two land covers. Upon further examination, it would appear that both Cleared Areas and Open Fens are very similar, spectrally, for both the Spring and Summer season. In particular, bands 2-4 tightly match one another. I suspect that this is likely the cause of the misclassification. Improving the

classification of Cleared Areas from Open Fens would further improve the classification accuracy and this could be accomplished through examining other classification schemes where cleared areas were classified more successfully. By comparing and contrasting the feature inputs used, I may be able to identify an even more superior set of inputs. When examining the worst performing classification tests (Naïve Bayes test #225 - Table 3-9), I noted that the most accurately classified land cover only had a producer classification accuracy of ~60%. The worst performing land cover (Treed Fen) has a producer accuracy of some ~8.7%. I also note that five of the land covers (Treed Bog, Cleared Areas, Treed Fen, Open Fen, and Coniferous Forests) have producer accuracies below 30%—essentially guesswork. Similar results are reflected in the corresponding user accuracy. For this test I noted that the features used are as the worst performing features as defined by the RF predictor importance analysis—3 of the 4 features are relatively noisy Sentinel-1 images and the other is a fall Band 3 image. Upon closer examination the statistical overlap between all land covers, for these features, is substantial, which indicates that this is the possible cause for this low level of classification accuracy across all land covers. In this case, there is not much which can be done to improve these results. However, what can be gleaned from this test is that these features truly are of poor quality.

When ranking the classification methods, overall, from most to least accurate, among all input features, it yielded (1) RF, (2) SVM, (3) K-Nearest Neighbours, 4) Naïve Bayes. Moreover, from an overall standpoint, RF classification results consistently outperformed all other classification methods, for all feature inputs. However, it is worth noting that in many cases the SVM and K-NN classification strategy produced results that were much closer in accuracy to the RF methodology when compared to Naïve-Bayes. As mentioned previously, one distinction between the RF, SVM, and the K-NN classification strategies compared to the Naïve-Bayes

strategy, was that they more thoroughly investigate subsets within the input training set, and are ensemble learning methods which do not operate on calculating gross statistics on the input datasets, at the cost of computation time. It is also worth mentioning, again, that from a mathematical perspective, RF regression and K-NN could be viewed as being part of similar mathematical families [Lin and Jeon 2002], which implies that they would interpret a given dataset in a similar fashion.

As a final note with respect to the accuracy of the classification results it is generally understood that co-registration of multi-source multi-spectral imagery can introduce a variation in final classification accuracy of several percent, depending on a multitude of factors. [Foody 2002; Husak et al. 1999; Loveland et al. 2000]. It has been noted in those previous studies that factors such as the overall accuracy and precision of the co-registration between images, the length of the perimeter between classes, and the overall area of the image, can play a role in how final classification accuracies can vary. In that regard a variation of several percent in the final classification accuracy can be expected due to those factors.

When considering how my work can be expanded upon, I note that this research would benefit from the addition of images from other years, and from other image sources. As a general principle, all of the classification methodologies used would benefit from additional data and data sources. To further develop my work, the addition of Lansat-8 and Sentinel-2 data (both now readily available) would be beneficial. However, it is worth noting that some of the best classification tests produced during this study already have very high accuracies and will probably not show vast improvement by the addition of more data. I speculate that the addition of more images in the form of Landsat-8 and Sentinel-2 images would provide more certainly with the variable significance analysis results, and possibly improve the accuracies of the worse

performing tests. However, the addition of large time-series of SAR data would be interesting. I speculate that through a large addition of SAR data more seasonal and structural features would become evident through the classification results. Furthermore, when considering how this work can be adapted to other study areas, I note that northern hemisphere temperate forests are all very similar in structure and vegetation distribution. The work done with this research should be sufficiently general with only minor local considerations from the study site. The methodologies and results produced in this study should be able to be applied without much difficulty to other northern hemisphere temperate study areas, in Canada or other parts of the world. Applying this work to tropical environments would likely be less compatible given the difference in vegetation density, vegetation types, and the lack of large seasonal variations with that vegetation. However, if given the appropriate datasets, the study methodology used here should be able to produce similar variable analysis and classification results, which would be an interesting contrast to our work.

3.6. Conclusions

A large focus of this study was the analysis and selection of features in order to facilitate the successful classification of the selected land covers from the test area. It was found that analysis of features using gross statistical analysis in the form of the Log-normal distance and an iterative regression approach in the form of the RF predictor importance value were an effective means of identifying which features were of high quality and should be used in classification and also which features were of low quality and should be either ignored or removed from classification. However, it was noted that while this style of analysis was effective across all classification methodologies in identifying low quality features, when it came to identifying the highest quality

features it was not as consistent. I suspect that these performance differences are driven by fundamental differences in how the log-normal distance is calculated (gross statistical measure, with no provisions for identifying multi-modal features) compared to the RF predictor importance value, which is iterative and explores subsets within a given dataset. Give these differences in feature analysis and the differences in how each classification technique analyzes a given dataset, the likely cause of this is discrepancy. It was also found that this analysis aided K-NN the most in identifying features, with its best performing tests being mostly represented by tests determined through this analysis (17 of its top 20 tests were determined by feature selection). For the other classification methodologies, results generally showed that the features determined by this analysis produced high accuracies but they did not produce the best results. Those results were produced by the input features determined through a holistic approach, with the best performing tests (RF test #77) produced an overall accuracy of 85.71%. The exact reason why holistically determined tests have performed so much better than quantitatively determined tests is unknown but should be further explored in future work. However, as a general trend I contend that applying this methodology to RF, SVM, and Naïve-Bayes especially provided value in determining lower quality features (features common in the bottom performing 20 tests), which could then be excluded from analysis to both speed up analysis time and ensure that results are more likely to be of a higher quality and accuracy. Moreover, I contend that with further development and study, this feature selection methodology could be refined such that it could produce selections of features, which would result in the highest classification accuracies.

When considering the classification results from a feature standpoint, my work has shown that the use of surface temperature, despite its low resolution, could be used to better classify wetlands in the study area in Northern Ontario, in particular if the temperature measurement was

from an abnormally warm, spring season. From this study I have estimated that the inclusion of surface temperature has increased overall classification accuracies between 2-3% for the highest performing tests. Additionally, the addition of RADARSAT-2, and or Sentinel-1 imagery to Landsat-5 imagery was shown to improve overall classification accuracies. It was also found that the data acquired in the fall season, if used solely as the classification input, consistently produced the poorest classification results. From these results I have also determined, through common trends, and data feature significance between tests, that data features such as NDVI, DEM, DEM-Derived slope, Surface temperature from abnormally warm spring months, and Band 3 (Green) and Band 4 (Red) from Landsat-8, can be generalized for other wetland study areas, and should be explored when evaluating and characterizing those areas.

Finally, from this study my analysis showed that the data used allowed for broad class separations (wetland-non versus wetland, treed wetland versus non-treed wetland), which implied that a hierarchical classification strategy could be an effective and efficient approach to the classification of wetlands. In order to explore this, further testing and development of these models should be undertaken. Additionally, further examinations of my results which would explore, and assign more quantifiable physical explanations to these results, and features should be carried out. Furthermore, optimum classification conditions for wetlands, and the ultimate limits that this style of analysis can produce should be explored. This is a challenging proposition but one that is worthwhile. This will not only provide a framework for wetland classification which can be used as a product but will also provide a level of expectation when it comes to the ultimate accuracy that this style of analysis can produce. This in turn will aid in determining the next steps required to achieve the next level of accuracy or detail.

Chapter 4 - Integration of Multi-source Remotely-Sensed Data With Hierarchically Based Classification Approaches in Support of the Classification of Wetlands

A portion of the research in this chapter is accepted in the following journal paper:

[1]Judah, Aaron, and Baoxin Hu. 2022. "The Integration of Multi-Source Remotely Sensed Data with Hierarchically Based Classification Approaches in Support of the Classification of Wetlands." *Canadian journal of remote sensing* 48 (2):158-81. doi: 10.1080/07038992.2021.1967732.

---

[1]I would like to thank the publisher Taylor & Francis, and Prof. Hu who have granted me permission to reuse this article in my dissertation.

In this chapter I will present the two hierarchical approaches I developed to classify wetlands (Open Bog, Treed Bog, Open Fen, Treed Fen, and Swamps) from remotely sensed data using a RF and a Niave-Bayes classification algorithms. The data utilized included multispectral optical and thermal data (Landsat-5, and Landsat-8), Radar imagery (Sentinel-1), and Digital Elevation Model (DEM). The goals were to determine the best way to combine imagery to classify wetlands, through hierarchically based classification approaches to produce more accurate and efficient classification maps compared to a standard classification approach. Classification algorithms used were RF, and Naïve Bayes. The hierarchically based approaches also improved classification accuracies for low quality data, as defined through data feature analysis, when compared to a non-hierarchical classifier. The hierarchical approaches also produced a significant increase in classification accuracy for the Naïve Bayes classifier vs the standard approach while not significantly increasing computation time - comparable in accuracy to the RF tests for around 20% the computational effort. The preselection of data features using Log-Normal or RF variable importance analysis, within the framework of a hierarchical classification strategy, was also very effective at identifying low quality data features and data features which were of higher quality which resulted in higher classification accuracies and lower computational costs.

4.1. Methodology

The analysis for this study was carried out in five phases. In the first phase, individual samples for each land cover type were identified through Forest Resources of Inventory (FRI) study data [Ministry of Natural Resources Canada 2012]. They were then separated into two subsets - training and evaluation. In the second phase, the remotely sensed imagery was

processed, georeferenced, and prepared for analysis. Relevant features were extracted in this phase as well. In the third phase, feature selection was carried out. Features derived in the second phase were analyzed using the Log-Normal distance, and an RF generated feature importance parameter based on the sum of changes to the mean squared error (MSE). In the fourth phase, various classification approaches executed through two hierarchical approaches were performed and the classification results were evaluated. Finally, for the fifth phase the best performing classification approach was used to classify a test area to explore the functionality of that approach and to provide a visual representation of a classified area. In the following sections these phases are described in more detail.

4.1.1. Defining Training and Evaluation Areas

Training and evaluation areas used were the same sites identified in section 3.2.1. Table 4-1 summarizes the sizes of the training areas (in pixels), and their corresponding evaluation sets.

Table 4-1. Number of pixels used for different classes in training and testing.

Class	Number Assigned to Class	Number of Pixels in Training Set	Number of Pixels in Validation Set
Open Fen	Class 1	1643	830
Treed Fen	Class 2	1377	752
Open Bog	Class 3	1105	634
Treed Bog	Class 4	983	617
Dense Con. Forest	Class 5	933	672
Swamps	Class 6	881	580
Grassy Areas	Class 7	2541	1331
Cleared Areas	Class 8	3184	2147

The number of pixels for each study area was determined by the size of land cover plots identified through the ground survey data. I attempted to have approximately 60% of the identified pixels be part of the training set, with the remaining 40% be part of the validation set. Based on the boundaries of these land cover plots, a set of contiguous pixels were selected for that individual land cover.

4.1.2. Image Processing and Feature Selection

For this study, six different image indices or metrics were used: NDVI, NDWI, Albedo, speckle filtered Radar reflectance, and Surface Temperature. These image metrics were selected in order to characterize vegetative activity (NDVI, NDWI), surface radiometric absorption (albedo), and surface structure measured via scattering (Radar). Surface temperature, due to my past study [Judah and Hu 2019], has shown that it can be useful if incorporated into the correct classification strategy. Additionally, DEM, and DEM derived slope were also incorporated into the classification of imagery. Slope was calculated from the DEM using the ENVI 5.6 topographic modeling function with a 3x3 window. DEM and DEM derived slope were selected to determine the role geographic features play in the classification process.

All Landsat imagery used was Level 1G, which is both radiometrically and geometrically corrected.

Surface albedo is a measure of reflectivity from a surface, which takes on a value from 0 (absorption) to 1 (complete reflectance). A standard approach in determining the surface albedo using Landsat-5 imagery is through a numerically determined relationship described by Liang et al. [1999], and in section 3.2.2. Liang describes albedo α using Landsat-5 TM however Liang's

expression was adapted for Landsat-8 imagery where the band indexes were changed to cover the same spectral ranges.

Surface temperature was calculated for individual pixels from the Landsat-5 and Landsat-8 imagery using the standard methodology outline in the Landsat 8 (L8) Data Users Handbook [Ihlen and Zanter 2019], which is also applicable to Landsat 5 imagery with interchanging of the appropriate bands.

For Sentinel-1 imagery, its calibrated Level-1 Single Look Complex (SLC) imagery layers were used. This product contains two polarized bands, VV and VH [European Space Agency 2018]. Once processed into equal X and Y pixel resolution geotiff it had a resolution of 12.5 m by 12.5 m. After processing it was georeferenced to the Landsat imagery. In order to reduce noise inherent throughout the images a standard speckle filtering algorithm was used. The Enhanced Frost speckle filter from PCI Geomatica with a 5x5 pixel window was used to filter all of the Sentinel-1 imagery. The window size was chosen based on optimizing the Log-Normal and RF significance results of the filtered imagery. A window less than 5x5 pixels and a window greater than 5x5 pixels were found to produce lower quality results, thus a 5x5 pixel window was optimal.

In total, each land cover had a data set corresponding to 238 individual layers, with each layer representing a unique feature: either a spectral band value, an image metric, radar value, digital elevation point, or a slope derived from the digital elevation. These individual features are summarized in Appendix 3.

4.1.3. Feature and Metric Significance Analysis

The objective of feature significance analysis was to quantify the statistical differences and similarities between land covers for a given feature. The intent of doing this analysis was to aid in determining which features and feature combinations would be desirable when classifying the selected land covers. To accomplish this, as with the previous chapter, two standard strategies were used. They were the Log-Normal distance and RF predictor importance value.

The first strategy, given a single feature with multiple land covers, was to measure the Log-Normal distance between land covers for that feature [Jensen 2004]. The Log-Normal distance, which has been previously described in section 3.2.3.

The second strategy was the RF predictor importance value, which again was previously described in section 3.2.3.

I ran a series of 238 classification tests where, for each test, a given feature was excluded from that test. In that way, for a given feature, when averaged over its 237 tests, a metric for how important that feature was when compared to its peers can be computed. The use of predictor importance with the RF classification methodology is a standard approach to evaluate the performance of individual input from a classification result [Biau and Scornet 2015].

Initial results of feature significance analysis showed that the significance of individual features can vary a great deal but it was decided to extract and use the complete set of features available in order to fully explore the discriminant power that each data set offers in the identification of wetland classes. In the discussion section I explore some of the peculiarities that I observed from the utilization of the entire data set.

4.1.4. Feature and Metric Selection and Classification

The core of this research was the classification of the collected imagery utilizing several hierarchical approaches executed with advanced data regression techniques. The hierarchal approaches utilize classifiers to segregate the data in a hierarchal fashion as opposed to training the classifier with all feature inputs at the same time. The rationale behind a hierarchal approach was to first separate broad classes and to then separate those results into more refined classes, using two different approaches. The key difference between the two approaches lay in the separation of wetlands in the final steps. What motivated the difference in the final steps was the recognition that the final separation could be parsed by either approaching it from a structural standpoint (Case-1) or a species standpoint (Case-2) Figure 4-1 provides an outline of these approaches. Both approaches began with an initial separation between wetland and non-wetland classes.

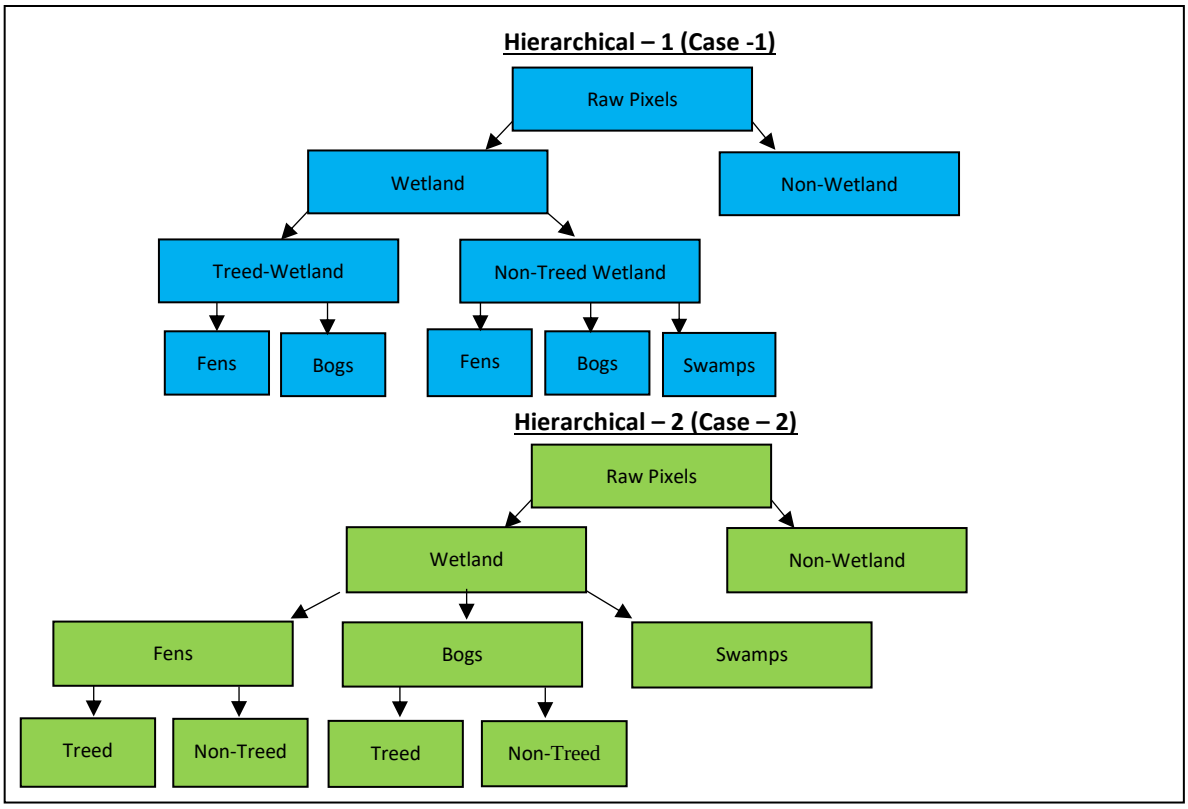

Figure 4-1. Illustration of the proposed hierarchical classification approaches.

For Case – 1 the second layer of separation breaks down wetland classes between treed and non-treed wetland classes. For Case-2 the second separation involves splitting the wetland class into fens, bogs, and swamps. Finally, Case-2 then separates fens and bogs into treed fens and treed bogs. In order to execute these hierarchical classification approaches feature inputs were selected by either feature significance analysis or through a holistic approach (described below in this section). Results from the hierarchical approaches and those from the traditional classification approaches were then compared and contrasted to determine performance differences in input features based on classification accuracies. The determination and selection of feature inputs for the hierarchical approaches used a "tailored" approach. With this tailored approach, for each split in the hierarchy, a selected set of features were used. These feature inputs were selected such that it would result in the maximum accuracy for that particular split.

These feature combinations, for each split, were determined by ranking the features for that split using the techniques described in the feature selection analysis section. For instance, the Log-Normal distance was between the board wetland and non-wetland classes for individual features. From these results the features were then ranked. This process was repeated for each split. Once this was established, using the training set, each split was then run through one of the classifiers in two feature decreasing inputs, from lowest ranked features to highest ranked features, to see which feature combination produced the highest accuracy output. The results from this produced a ranked list of features, based on classification accuracy, for that particular split, for that classifier – the feature combination which provided the highest accuracy was then selected for that split. This process was repeated for all splits for both hierarchical classification approaches. When executing the hierarchical approach, feature inputs were drawn from these results, and executed with matching rank indexes. For instance, the top ranked feature inputs, for all splits, were executed for a given test, and then the second ranked tests for feature inputs, for all splits, were executed, and so on.

For all of the aforementioned hierarchical classification approaches I used each of the classifiers – RF or Naïve Bayes.

The training data was analyzed and classified using the previously mentioned hierarchical classification approaches and for comparison purposes, using a standard classifier with feature inputs listed in Appendix 6. These features inputs were determined and assembled through a number of different methods. The first method was to select groups of feature inputs with a "holistic" approach. This involved selecting groups of features based on similarities or contrast in type (bands or metrics), similarities or contrasts in time (the same or different seasons) and combinations thereof. Additionally, combinations of features were selected from a physical or

structural standpoint in order to take into account seasonal variability in vegetation and structural differences in land covers which could be parsed by the classification approaches through the incorporation of features like Radar and DEM derived values. Using this holistic approach, 153 different sets of input features were created. The next set of input features was selected by examining the results from the feature significance analysis. Based on the overall ranking of those features, the top 50 percent of features were selected, in 5 feature ranked increments, as inputs. Also, feature inputs based on the bottom 50 percent ranked features, in similar decreasing feature ranking increments were selected as inputs. Additionally a hybrid combination of the top 10 to 60 percent of features were selected based on selecting a combination of surface reflectances from bands, image indices, and DEM derived features, in order to emulate the holistic approach but with a more quantitative background. In order to execute this, for instance, for the top 10 percent of features with the hybrid approach, the top three surface reflectances from bands, the top image indices and the top Radar or DEM or DEM Slope features were selected, for a total of 5 or 10 percent of available features. This approach was repeated until six different hybrid combinations reflecting the top 10 to 60 percent of features were created. Finally, the bottom ranked 25 percent of features were grouped together to form a hybrid of bottom ranked features from the bottom 16 to the bottom four features in two feature, decreasing, increments. The aforementioned feature selection strategies were executed for both the Log-Normal distance and RF determined feature importance values. In total, 312 unique tests were devised.

4.2. Study Area and Data Used

The study area is the same area described in section 3.1, illustrated in Figure 3-1. Landsat-5, Landsat-8, and Sentinel-1 imagery were the primary image sources used in this study. Landsat-5, imagery is previously described in section 3.1. The Landsat-8 series of sensors collect multispectral optical imagery with a spatial resolution of 30 m by 30 m and thermal imagery at 120 m by 120 m [U.S. Geological Survey Landsat Program 2002]. As a point to note, when creating layer stacks of these images for analysis, the lower resolution (120 m by 120 m) temperature-based images were resampled to 30 m by 30 m. The Landsat-8 series of sensors collect multispectral optical imagery with a spatial resolution of 30 m by 30 m across all spectral bands [U.S. Geological Survey Landsat Program 2019].

Sentinel-1 imagery (C-band), the product used was dual-polarization imagery, and had a resolution of 5 m by 20 m [European Space Agency 2018]. The Sentinel-1 imagery was resampled to 12.5 m by 12.5 m in order to facilitate ease of analysis with the other imagery products. More on this further in this section. The final imagery product used in this study was the aerial imagery previously described in section 3.1. It was used for closer examinations of training and validation sites as identified by Ministry of Natural Resources surveys of the area.

Finally, a DEM of the study area taken from the Canadian Digital Surface Model previously described in section 3.1. In total, three different Landsat-5 images, nine different Landsat-8 images, and 46 different Sentinel-1 images were collected. Table 4-2 summarizes the dates and types of imagery that were collected for this study.

Table 4-2. Summary of remotely-sensed imagery collected for this study.

	Imagery	Season	Date	Level of Processing	Accessed From
1.	Landsat-5	Spring	18-May-2010	Level 1G	United States Geological Survey (USGS)
2.	Landsat-5	Summer	24-July-2011	Level 1G	USGS
3.	Landsat-5	Fall	09-October-2010	Level 1G	USGS
4.	Sentinel-1	Summer	01-August-2015	Level 1—SLC	European Space Agency (ESA) - Sentinel
5.	Sentinel-1	Summer	25-August-2015	Level 1—SLC	ESA - Sentinel
6.	Sentinel-1	Fall	18-September- 2015	Level 1—SLC	ESA - Sentinel
7.	Landsat-8	Summer	05-September-2015	Level 1G	USGS
8.	Landsat-8	Fall	21-September-2015	Level 1G	USGS
9.	Landsat-8	Summer	22-August-2016	Level 1G	USGS
10.	Landsat-8	Summer	24-July-2017	Level 1G	USGS
11.	Landsat-8	Summer	25-August-2017	Level 1G	USGS
12.	Landsat-8	Summer	10-September-2017	Level 1G	USGS
13.	Landsat-8	Spring	25-May-2018	Level 1G	USGS
14.	Landsat-8	Summer	25-Jun-2018	Level 1G	USGS
15.	Landsat-8	Summer	12-August-2018	Level 1G	USGS
16.	Sentinel-1	Spring	20-May-16	Level 1—SLC	ESA - Sentinel
17.	Sentinel-1	Spring	13-Jun-16	Level 1—SLC	ESA - Sentinel
18.	Sentinel-1	Summer	14-Jul-16	Level 1—SLC	ESA - Sentinel
19	Sentinel-1	Summer	19-Jul-16	Level 1—SLC	ESA - Sentinel
20	Sentinel-1	Fall	30-Sep-16	Level 1—SLC	ESA - Sentinel
21	Sentinel-1	Fall	05-Oct-16	Level 1—SLC	ESA - Sentinel
22	Sentinel-1	Spring	02-Jun-17	Level 1—SLC	ESA - Sentinel
23	Sentinel-1	Spring	14-Jun-17	Level 1—SLC	ESA - Sentinel
24	Sentinel-1	Spring	26-Jun-17	Level 1—SLC	ESA - Sentinel
25	Sentinel-1	Summer	03-Jul-17	Level 1—SLC	ESA - Sentinel
26	Sentinel-1	Summer	08-Jul-17	Level 1—SLC	ESA - Sentinel

27	Sentinel-1	Summer	15-Jul-17	Level 1—SLC	ESA - Sentinel
28	Sentinel-1	Summer	20-Jul-17	Level 1—SLC	ESA - Sentinel
29	Sentinel-1	Summer	27-Jul-17	Level 1—SLC	ESA - Sentinel
30	Sentinel-1	Summer	01-Sep-17	Level 1—SLC	ESA - Sentinel
31	Sentinel-1	Summer	06-Sep-17	Level 1—SLC	ESA - Sentinel
32	Sentinel-1	Fall	18-Sep-17	Level 1—SLC	ESA - Sentinel
33	Sentinel-1	Fall	07-Oct-17	Level 1—SLC	ESA - Sentinel
34	Sentinel-1	Fall	12-Oct-17	Level 1—SLC	ESA - Sentinel
35	Sentinel-1	Spring	23-May-18	Level 1—SLC	ESA - Sentinel
36	Sentinel-1	Spring	04-Jun-18	Level 1—SLC	ESA - Sentinel
37	Sentinel-1	Spring	09-Jun-18	Level 1—SLC	ESA - Sentinel
38	Sentinel-1	Spring	16-Jun-18	Level 1—SLC	ESA - Sentinel
39	Sentinel-1	Spring	21-Jun-18	Level 1—SLC	ESA - Sentinel
40	Sentinel-1	Spring	28-Jun-18	Level 1—SLC	ESA - Sentinel
41	Sentinel-1	Summer	10-Jul-18	Level 1—SLC	ESA - Sentinel
42	Sentinel-1	Summer	15-Jul-18	Level 1—SLC	ESA - Sentinel
43	Sentinel-1	Summer	22-Jul-18	Level 1—SLC	ESA - Sentinel
44	Sentinel-1	Summer	27-Jul-18	Level 1—SLC	ESA - Sentinel
45	Sentinel-1	Summer	08-Aug-18	Level 1—SLC	ESA - Sentinel
46	Sentinel-1	Summer	15-Aug-18	Level 1—SLC	ESA - Sentinel
47	Sentinel-1	Summer	20-Aug-18	Level 1—SLC	ESA - Sentinel
48	Sentinel-1	Summer	01-Sep-18	Level 1—SLC	ESA - Sentinel
49	Sentinel-1	Summer	08-Sep-18	Level 1—SLC	ESA - Sentinel
50	Sentinel-1	Fall	13-Sep-18	Level 1—SLC	ESA - Sentinel
51	Sentinel-1	Fall	02-Oct-18	Level 1—SLC	ESA - Sentinel
52	Sentinel-1	Fall	07-Oct-18	Level 1—SLC	ESA - Sentinel
53	Sentinel-1	Fall	14-Oct-18	Level 1—SLC	ESA - Sentinel
54	Sentinel-1	Spring	18-May-19	Level 1—SLC	ESA - Sentinel
55	Sentinel-1	Spring	23-May-19	Level 1—SLC	ESA - Sentinel
56	Sentinel-1	Spring	30-May-19	Level 1—SLC	ESA - Sentinel

57	Sentinel-1	Summer	04-Jun-19	Level 1—SLC	ESA - Sentinel
58	Sentinel-1	Summer	11-Jul-19	Level 1—SLC	ESA - Sentinel

As a final note regarding the imagery used in this study, in order to preserve the information from the higher resolution Sentinel-1 imagery, all images were resampled to 12.5 m by 12.5 m resolution when layer stacked together. 12.5 m by 12.5 m resolution was chosen in order to match the resolution of the corrected and processed Sentinel-1 imagery.

Eight different land covers were classified in this study. These land covers were Open Fen, Treed Fen, Open Bog, Treed Bog, Dense Coniferous Forest, Swamps, Grassy Areas, and Cleared Areas, previously described in section 3.1.

4.3. Results

4.3.1. Results - Feature Significance

Given a feature and its corresponding Log-Normal distance a larger value implies that feature could play a more significant role in the classification of those land covers compared with features which had smaller Log-Normal values. Additionally, given the 238 features and eight land cover classes, using the RF computed predictor importance values, executed with the strategy described in Section 3.3.3, the importance of a given feature could be determined. Like the Log-Normal values, larger importance values implied that a given feature was more valuable in the classification process, and when utilized, would produce more accurate results. The results of the top 25% of feature importance computations are summarized in Table 4-3 and Table 4-4. The complete set of feature importance values are summarized in Appendix 4 and Appendix 5.

Table 4-3. Summary of top 25% of data feature quality analysis averaged over all land cover types. Image metrics, image bands, radar parameters, DEM and slope sorted from largest to smallest Log-Normal distance, with a longer distance implying a higher quality. ■ indicates a spring month, ■ indicates a summer, month, ■ indicates a fall month.

Log-Normal Distance							
Bands	**Distance**	**Metrics**		**Distance**	**Radar and DEM**	**Distance**
B4 Reflect. IMG-7	1.916	NDVI	IMG-7	2.276	Nill	
B4 Reflect. IMG-10	1.595	NDVI	IMG-9	2.199		
B4 Reflect. IMG-11	1.480	NDVI	IMG-12	2.068		
B7 Reflect. IMG-1	1.478	NDVI	IMG-11	2.032		
B4 Reflect. IMG-12	1.478	NDVI	IMG-10	2.000		
B4 Reflect. IMG-9	1.478	NDWI	IMG-2	1.890		
B4 Reflect. IMG-2	1.450	NDVI	IMG-8	1.725		
B4 Reflect. IMG-13	1.411	Temp 1	IMG-10	1.705		
B5 Reflect. IMG-1	1.402	Albedo	IMG-7	1.665		
B2 Reflect. IMG-7	1.336	NDVI	IMG-13	1.639		
B7 Reflect. IMG-10	1.281	NDWI	IMG-3	1.632		
B4 Reflect. IMG-3	1.253	Albedo	IMG-2	1.580		
B6 Reflect. IMG-1	1.230	Temp 2	IMG-10	1.534		
B7 Reflect. IMG-7	1.230	Albedo	IMG-13	1.528		
B4 Reflect. IMG-8	1.230	Temp 1	IMG-11	1.519		
B3 Reflect. IMG-7	1.206	Albedo	IMG-10	1.513		
B7 Reflect. IMG-2	1.189	NDWI	IMG-1	1.475		
B2 Reflect. IMG-10	1.188	Temp 2	IMG-11	1.426		
B2 Reflect. IMG-9	1.159	Temp 1	IMG-12	1.369		
B7 Reflect. IMG-12	1.159	Albedo	IMG-11	1.344		
B1 Reflect. IMG-7	1.159	Temp 2	IMG-12	1.340		
B2 Reflect. IMG-13	1.154	NDVI	IMG-3	1.310		
B6 Reflect. IMG-7	1.115	NDVI	IMG-2	1.298		
B6 Reflect. IMG-10	1.094	Temp 2	IMG-2	1.264		
B3 Reflect. IMG-9	1.053	Temp 1	IMG-9	1.249		
B4 Reflect. IMG-1	1.048	Temp 1	IMG-7	1.237		
		Albedo	IMG-12	1.231		
		NDVI	IMG-1	1.225		
		Temp 2	IMG-13	1.181		
		Temp 2	IMG-7	1.178		
		Temp 2	IMG-9	1.174		
		Temp 1	IMG-13	1.140		
		Albedo	IMG-3	1.111		
		Albedo	IMG-1	1.091		
Average	1.299	Average		1.745	Average	Nill
Standard Deviation	0.194	Standard Deviation		0.257	Standard Deviation	Nill

Table 4-4. Summary of top 25% of data feature quality analysis averaged over all land cover types. Image metrics, image bands, radar parameters, DEM and slope sorted from largest to smallest RF predictor importance value. Larger values implying a higher quality. ■ indicates a spring month, ■ indicates a summer, month, ■ indicates a fall month.

Random Forest Determined Predictor Importance							
Bands	**Distance**	**Metrics**		**Distance**	**Radar and DEM**	**Distance**
B4 Reflect. IMG-7	1.145	NDVI	IMG-7	0.788	DEM	1.072
B4 Reflect. IMG-10	1.130	NDVI	IMG-9	0.779	DEM-Slope	0.759
B4 Reflect. IMG-11	1.081	NDVI	IMG-12	0.704		
B7 Reflect. IMG-1	1.012	NDVI	IMG-11	0.699		
B4 Reflect. IMG-12	0.809	NDVI	IMG-10	0.669		
B4 Reflect. IMG-9	0.804	NDWI	IMG-2	0.621		
B4 Reflect. IMG-2	0.803	NDVI	IMG-8	0.575		
B4 Reflect. IMG-13	0.774	Temp 1	IMG-10	0.575		
B5 Reflect. IMG-1	0.770	Albedo	IMG-7	0.570		
B2 Reflect. IMG-7	0.769	NDVI	IMG-13	0.566		
B7 Reflect. IMG-10	0.767	NDWI	IMG-3	0.557		
B4 Reflect. IMG-3	0.745	Albedo	IMG-2	0.555		
B6 Reflect. IMG-1	0.680	Temp 2	IMG-10	0.548		
B7 Reflect. IMG-7	0.659	Albedo	IMG-13	0.547		
B4 Reflect. IMG-8	0.648	Temp 1	IMG-11	0.544		
B3 Reflect. IMG-7	0.633	Albedo	IMG-10	0.535		
B7 Reflect. IMG-2	0.633	NDWI	IMG-1	0.535		
B2 Reflect. IMG-10	0.631	Temp 2	IMG-11	0.533		
B2 Reflect. IMG-9	0.600	Temp 1	IMG-12	0.528		
B7 Reflect. IMG-12	0.595	Albedo	IMG-11	0.522		
B1 Reflect. IMG-7	0.588	Temp 2	IMG-12	0.495		
B2 Reflect. IMG-13	0.556	NDVI	IMG-3	0.486		
B6 Reflect. IMG-7	0.536	NDVI	IMG-2	0.473		
B6 Reflect. IMG-10	0.531	Temp 2	IMG-2	0.471		
B3 Reflect. IMG-9	0.528					
B4 Reflect. IMG-1	0.524					
B2 Reflect. IMG-3	0.523					
B6 Reflect. IMG-2	0.520					
B6 Reflect. IMG-12	0.508					
B1 Reflect. IMG-10	0.491					
B7 Reflect. IMG-11	0.487					
B7 Reflect. IMG-3	0.486					
B5 Reflect. IMG-2	0.476					
B5 Reflect. IMG-9	0.473					
Average	0.674	Average		0.578	Average	0.916
Standard Deviation	0.187	Standard Deviation		0.086	Standard Deviation	0.157

From the results in Table 4-3 and Table 4-4 it is noted that the features calculated from multispectral imagery, on average, were of the highest quality and importance, with traditional metrics such as NDVI performing well, when measured by both the Log-Normal and RF determined predictor importance values. In addition, the features calculated from data acquired in the spring and summer, were of a higher quality, on average, compared to the fall according to the Log-Normal and RF importance values. It is also noted that surface temperature, traditionally a feature not associated with wetland land cover classification, was ranked fairly high by both feature analysis methodologies for some images. These results also implied that the collected surface temperature data, despite its low resolution, was of a high enough quality that it could be useful for land cover classification. These results are consistent with a previous study I conducted with other images [Judah and Hu 2019]. Finally, it was noted that Radar, DEM, and slope were largely not present in the top 25% of features. When examining these features more closely in Appendix 4 and Appendix 5 these features were, on average, of a lower quality compared with those derived from optical and thermal data based on the Log-Normal method and the RF method. However, the difference (in magnitude) was not as large. Among the features from the Radar data and DEM, slope ranked higher than most of the sentinel-1 images by both Log-Normal and RF important values. With regards to Sentinel-1 images, there was no clear consensus which season was more significant than others and the very best Sentinel-1 images only ranked at the same level as some of the middle-ranked image bands and metrics according to the RF importance values. From the Log-Normal analysis, the Sentinel-1 images were of a magnitude similar to the bottom quarter of image bands and metrics for all images.

4.3.2. Results – Classification

The classification results for the various methods and inputs were computed on a desktop computer equipped with an AMD Ryzen 5 2600 Six-Core Processor with 32 gigabytes of RAM, analyzed, and ranked. From these tests, the basic statistical features for these tests, for each classification technique were calculated. These results are presented and summarized in Table 4-5 and Table 4-6. Table 4-5 A), using the standard classifier, RF, on average produced the most accurate results given all inputs scenarios. Also, the highest accuracy result, for all tests, for all classification approaches was produced by the RF classifier using the Case - 2 Classification approach (RF Heir. – 2 Test #1). Additionally, average Kappa values are consistent and are of a magnitude which imply that classification results are in agreement between producer and user accuracy. From Table 4-5 B) and C) the hierarchical classification approaches, produce the most accurate results. It is worth noting that tests produced by Naïve Bayes, using the hierarchical approaches, produced dramatic performance increases at the top end – for Case - 2 close to 12%. This performance increase is an unexpected result, especially given how generally, tests produced using Naïve Bayes, showed much poorer performance when compared to RF produced tests for the same sets of inputs.

Table 4-5. Summary of classification accuracy of different methodologies averaged over all classification strategies. A) Standard classification method, B) Case - 1 classification method, C) Case - 2 classification method.

A)

	Average	Standard Deviation	Max	Min	Ave. Kappa Value	Difference from Top to 25th Percentile	Difference from Top to 50th Percentile	Computation Time (s)
Random Forest	0.6400	0.1643	0.8006	0.2728	0.6871	0.0319	0.0777	5322
Naïve Bayes	0.5993	0.1255	0.7845	0.2751	0.6212	0.0719	0.1371	61

B)

	Average	Standard Deviation	Max	Min	Ave. Kappa Value	Difference from Top to 25th Percentile	Difference from Top to 50th Percentile	Computation Time (s)
Random Forest	0.6042	0.2494	0.8720	0.2632	0.6267	0.0566	0.1491	3180
Naïve Bayes	0.7556	0.0961	0.8680	0.4217	0.7697	0.0205	0.0947	302

C)

	Average	Standard Deviation	Max	Min	Ave. Kappa Value	Difference from Top to 25th Percentile	Difference from Top to 50th Percentile	Computation Time (s)
Random Forest	0.6076	0.2388	0.9194	0.2182	0.6010	0.0925	0.1815	2892
Naïve Bayes	0.5607	0.2129	0.9019	0.1547	0.5568	0.1817	0.2626	628

When examining these results in more detail in

Appendix 7, I note some common themes. Preselection of variables via Log-Normal and RF importance analysis is predictive of both high and low classification accuracy – high quality features, grouped together, corresponding to higher accuracy and low quality features, grouped together, corresponding to lower accuracies. This is the expected result. Another common theme is that the number of features used does not necessarily correspond to classification success. Some of the highest classification accuracy tests used a lower number of high quality, pre-selected features, compared to tests where more features were used. This is an unexpected result. In many instances, for advance regression algorithms, increasing the number of features generally results in more accurate and consistent results. In this case, we note a number of instances where the addition of more features results in lower accuracies, compared to results where less features were used. This was true for all classification methodologies and approaches. It was also noted that in many cases, the use of holistically chosen variable combinations produced some of the best results and choosing feature combinations via preselection techniques did not always correspond to superior classification accuracies.

When examining

Appendix 7 B), I note that for the bottom 10% of tests, selection of low quality features through Log-Normal and RF importance analysis, and the use of fewer features correlates strongly to low classification accuracy. This was true for all classification methods and approaches. In Table 4-6 when examining the average classification accuracies for the top and bottom ranked tests I note that for the top 10% of ranked tests, the hierarchical approaches provided an overall performance increase in classification accuracies, especially for the Naïve Bayes classification method, which generally does not perform as well as its ensemble classification method counterparts. When examining the bottom 10% of tests I note that for the hierarchical classification approaches I see a performance increase when compared to a traditional classifier.

Inspired by these results and to test the functionality of the hierarchical classification approaches, the bottom 10% of classification inputs, as determined by the standard classification results, were used as an inputs in the hierarchical classification approaches. In this case, for each split in the hierarchy I used the same input throughout. The results of these tests are summarized in Table 4-6 and Table 4-7.

Table 4-6. Average classification accuracies for top 10% and bottom 10% of classification tests for all methodologies and approaches.

	Top 10% Classification Accuracy Ranked Tests - Averaged			Bottom 10% Class. Accuracy Ranked Tests – Averaged		
	Standard	**Hier.-1**	**Hier.-2**	**Standard**	**Hier.-1**	**Hier.-2**
Random Forest	0.7881	0.8478	0.8941	0.2540	0.2938	0.2541
Naïve Bayes	0.7493	0.8604	0.8783	0.2910	0.5790	0.2216

Table 4-7. Results of the bottom 10% of classification inputs from the standard classification method, run through the hierarchical classification approaches.

	Bottom 10% Classified Inputs - Maximum Values			Bottom 10% Classified Inputs - Minimum Values		
	Standard	**Hier.-1**	**Hier.-2**	**Standard**	**Hier.-1**	**Hier.-2**

Random Forest	0.2821	0.3558	0.3399	0.2010	0.2638	0.2223
Naïve Bayes	0.3526	0.3893	0.3859	0.1892	0.1212	0.2118

From Table 4-7 I note that for these tests that there was a performance increase with these poorer performing tests. I note that for the maximum values produced through these tests, 3-7% increase and to the minimum values a 3-6% increase. A cursory experiment where I executed the same classification scheme but with the top 10 percent of tests, as determined by the standard classification approaches results, did not see the same performance increases, further validating the need for tailoring inputs for each split in the hierarchy.

To better examine the best performing classification result I examine its corresponding confusion matrix in Table 4-8 A). As a comparison the confusion matrix from the most accurate non-hierarchical classification is presented in Table 4-8 B).

When examining Table 4-8 A) I note that Grassy Areas and Treed Bog have the largest discrepancy, however, at 81.6% the classification accuracy of the Treed Bog is still relatively high. When examining Treed Bog more closely I see that the classifier conflates it the most with Open Fens. When examining Table 4-8 B) I also see that Grassy Areas was the best performing result and Open Bogs performed the worst. However, again, at 79.9% the classification accuracy is still relatively high. Examining Open Bogs more closely I see it is conflating its results the most with Swamps which are similar in structure to bogs based on the lack of tree cover and high water content. As a comparison, and to examine how the different classification approaches can affect classification results for the same input, I examine several confusion matrixes of the worst performing classification result, which was also common among all classification methods and approaches (Bottom 5 LN – Determined Variables). Table 4-9 presents these results. When examining Table 4-9 A), I can immediately see the contrast in results compared to Table 4-8.

When examining Table 4-9 A) I note that for all land covers misclassification between other land covers is very high and the classification accuracies are low – close to a random assignment. When examining Table 4-9 B) I note a more structured distribution of values, higher classification accuracies, and misclassification within dry and wetland land covers as the largest source of errors. While Table 4-8 B) is an improvement over the results from Table 4-9 A) (~7% improvement) it shows more structure indicating that the hierarchical based approach is parsing the inputted data in a more physically accurate fashion and it is better at maximizing the potential of an inputted dataset. Finally, it is worth noting that the results from Table 4-9 B) are from the Naïve Bayes classification method which generally does not perform as well as ensemble methods, such as its contemporaries in this study.

Table 4-8. A) Confusion matrix for the most accurate classification test (RF Hierarchical - 2 test #1). B) Confusion matrix for the most accurate non-hierarchical classification results (RF test #131)

	Open Fen	Treed Fen	Open Bog	Treed Bog	Coniferous Forest	Swamps	Grassy Area	Cleared Area	Producer Accuracy	User Accuracy
Open Fen	523	3	5	293	0	0	0	0	0.851	1.000
Treed Fen	0	696	8	6	0	0	0	0	0.980	0.996
Open Bog	0	0	377	4	2	89	0	0	0.799	0.889
Treed Bog	0	0	28	348	0	0	10	0	0.902	0.532
Coniferous Forest	0	0	0	0	784	0	74	0	0.914	0.966
Swamps	0	0	6	3	15	463	1	0	0.949	0.839
Grassy Area	0	0	0	0	0	0	3016	0	1.000	0.969
Cleared Area	0	0	0	0	11	0	10	1136	0.982	1.000

B)

	Open Fen	Treed Fen	Open Bog	Treed Bog	Coniferous Forest	Swamps	Grassy Area	Cleared Area	Producer Accuracy	User Accuracy
Open Fen	582	52	0	190	0	0	0	0	0.706	0.696
Treed Fen	53	645	0	11	0	1	0	0	0.909	0.802
Open Bog	0	0	67	0	0	402	0	3	0.142	0.221
Treed Bog	11	1	30	289	0	0	55	0	0.749	0.421
Coniferous Forest	0	0	0	3	783	3	71	0	0.911	0.956
Swamps	0	1	0	0	31	455	1	0	0.932	0.424
Grassy Area	0	0	0	4	9	22	3018	0	0.989	0.892
Cleared Area	0	0	0	0	5	0	22	1130	0.977	0.993

Table 4-9. A) Confusion matrix of the worst performing classification test (Bottom 5 LN – Determined Variables – Standard Naïve Bayes Classification). B) Bottom 5 LN – Determined Variables – run through the Hierarchical - 2 classification approach with the Naïve Bayes Classifier.

	Open Fen	Treed Fen	Open Bog	Treed Bog	Coniferous Forest	Swamps	Grassy Area	Cleared Area	Producer Accuracy	User Accuracy
Open Fen	101	74	32	63	132	24	275	123	0.123	0.126
Treed Fen	95	71	37	42	65	19	264	117	0.100	0.096
Open Bog	50	36	46	27	29	55	168	61	0.097	0.114
Treed Bog	50	34	11	28	45	26	141	51	0.073	0.064
Coniferous Forest	76	83	38	33	158	21	310	141	0.184	0.181
Swamps	24	28	55	19	22	160	137	43	0.328	0.372
Grassy Area	288	308	130	157	310	93	1218	549	0.399	0.409
Cleared Area	117	107	53	71	112	32	467	198	0.171	0.154

B)

	Open Fen	Treed Fen	Open Bog	Treed Bog	Coniferous Forest	Swamps	Grassy Area	Cleared Area	Producer Accuracy	User Accuracy
Open Fen	344	346	95	1	0	38	0	0	0.417	0.309
Treed Fen	325	267	82	1	2	35	8	1	0.370	0.307
Open Bog	160	78	107	0	20	127	68	7	0.189	0.238
Treed Bog	123	106	30	82	0	45	4	0	0.131	0.500
Coniferous Forest	0	0	0	0	221	0	605	34	0.257	0.345
Swamps	112	44	135	0	12	197	65	6	0.345	0.446
Grassy Area	0	0	0	0	264	0	2674	111	0.877	0.609
Cleared Area	0	0	0	0	122	0	966	69	0.060	0.303

As a final examination of the classification results, I examined the average rank of a given test input averaged over classification methodologies, by its standard deviation. The objective was to determine a given classification inputs overall performance across all of the given classification methodologies. A scatter plot of these results is presented Figure 4-2.

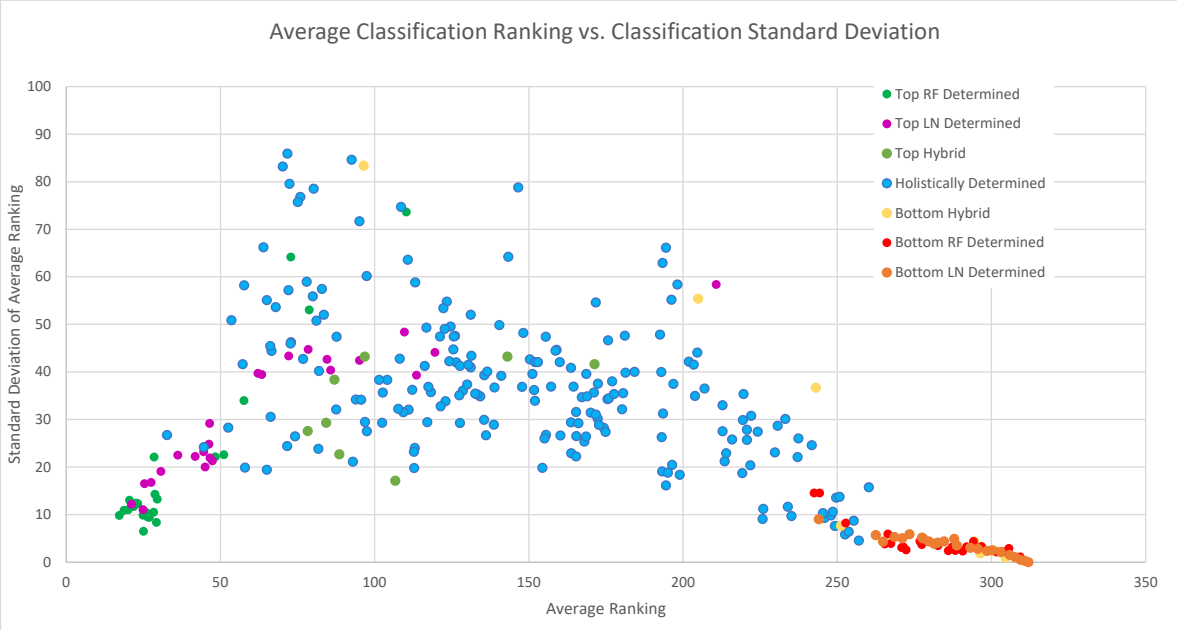

Figure 4-2. Scatter plot of average classification ranking versus classification standard deviation for each test. Results are broken down by classification input selection type.

Given that the highest ranking a test could achieve would be 1 and the lowest ranking a test could achieve would be 312 (the total number of tests conducted), I interpret Figure 4-2 by noting that the highest quality results would be at the origin and the lowest quality results would be further down the x-axis and up the y-axis. When examining Figure 4-2, I note that results with the highest accuracy had lower spreads compared to results which were of lower quality; however, results of the lowest quality had similarly tight spreads with their distributions. These results are consistent with the results in the previous chapter which resulted in a paper I published [Judah and Hu 2019]. These results imply that tests which produced the highest accuracies would tend to be similarly accurate across classification strategies, and alternatively,

classification tests which were of lower accuracy would be of similarly lower accuracy across different classification strategies. Moreover, for tests that were of average accuracy would have large variations in accuracies across classification methods. Finally, according to Figure 4-2 feature inputs selected by their performance from feature significance analysis were generally of a higher accuracy and lower standard deviation when compared to results selected by a holistic approach, and alternatively, feature inputs indicated to be of lower quality and significance produce consistently lower accuracy results across all classification strategies.

When examining the individual features for these tests I noted that features of lower quality as defined by both the Log-Normal distance and RF predictor importance values were more consistent represented in lower accuracy tests across all classifiers and classification approaches, when compared to the set of features which constituted the highest accuracy tests. I also note that features determined to be of high quality via RF predictor importance values generally out performed those determined by Log-Normal distance, across classifiers and classification approaches. My interpretation is that this has to do with how the RF predictor importance analysis investigates subsets within a dataset. This is further expanded in the discussion section.

When examining the results from the hierarchical classification approaches I noted that variable ranking did not always correspond to its ranking with regards to classification accuracy – highest ranked variable selections did not always correspond to the highest rank classification results. In order to better visualize this I created a similar plot to Figure 4-2, where for a given set of inputs for the hierarchical classification tests, their corresponding results, across the two hierarchical approaches, and across the two classifiers were averaged together and the standard deviation was calculated. When plotted together I produce Figure 4-3.

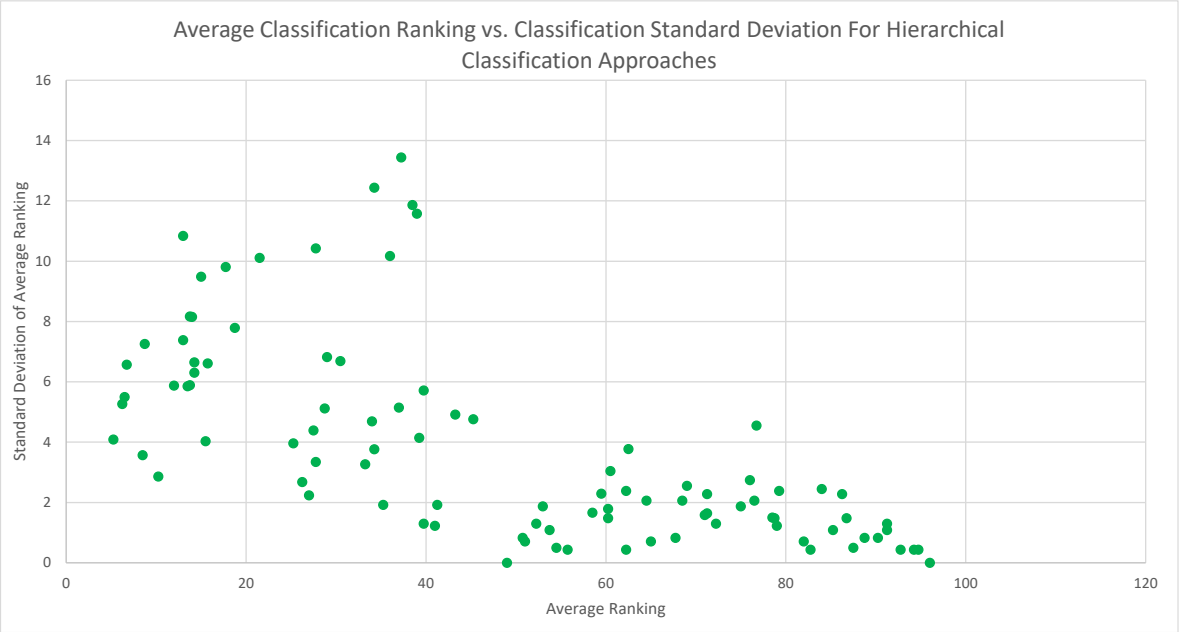

Figure 4-3. Average classification ranking for a given input across hierarchical approaches, and classifiers vs. standard deviation.

When examining Figure 4-3 I note that the lowest ranked tests (those with the worst performing feature inputs) are more consistent with one another, compared to the higher ranked tests which show more variation. However, it would appear that for the highest ranked tests, the variation decreases. The cone like structure in the distribution of values I observed in Figure 4-2 is not as pronounced as in Figure 4-3 but shows some of the same characteristics.

With regards to individual sets of features and when examining the best performing hierarchical classification test (RF Case - 1 Test #1) I noted that there were some common themes among all levels of the hierarchy but differences did exists throughout and became more nuanced further down the hierarchy. For the first broad class separation I noted that NDVI, DEM and surface temperature were of significant importance. This echoed earlier experiments where I saw that using these three features alone could result in broad class separations. At the second level of the hierarchy I noted that the addition of individual bands from Landsat-8 and Sentinel-1

imagery. This makes sense given the separation is focused on separating treed to non-treeded landcovers – a structural difference which could be better detected through radar data. For the final level in the hierarchy I noted that different Landsat-8 bands were of significance and while similar features were listed the order of significance was different. Also, for the splitting of non-treed wetlands, the use of NDWI was of significance while the splitting of treed fens and bogs was not. I speculate that this is driven by the larger water content of swamps vs the other landcovers.

Table 4-10. Summary of the most significant features used in different levels of the hierarchy.

Hierarchy #	Split in Hierarchy	Most Significant Features
1	Wetland vs Non-Wetland	NDVI, DEM, Surface Temp, Band 3 and 5 Landsat-8
2	Treed vs Non-Treed Wetland	DEM, NDWI, NDVI, Surface Temp, Band 2 and 5 Landsat-8, Sentinel-1
3 - A	Treed - Fens and Bogs	Band 1 and 3 Landsat-8, NDVI, DEM, Surface temp
3 - B	Non-Treed - Fens, Bogs, swamps	DEM, Band 2 and 1 Landsat-8, Surface Temp, NDWI

As a test to explore the functionality of a classification approach produced from this study, the most accurate classification approach, produced through an RF Case - 1 Test #1, was adapted to classify a test area from the study area. This test area was chosen such that it did not contain any data drawn from the training or validation data and appeared to contain wetlands of varying types (identified through visual interpretation). Open water, such as rivers, lakes, etc. masked out of the test image. Figure 4-4 contains two images, which are typical of the output from this classification approach Figure 4-4 A), is a true color Landsat-8 image of a test area, Figure 4-4 B) is the actual classification result.

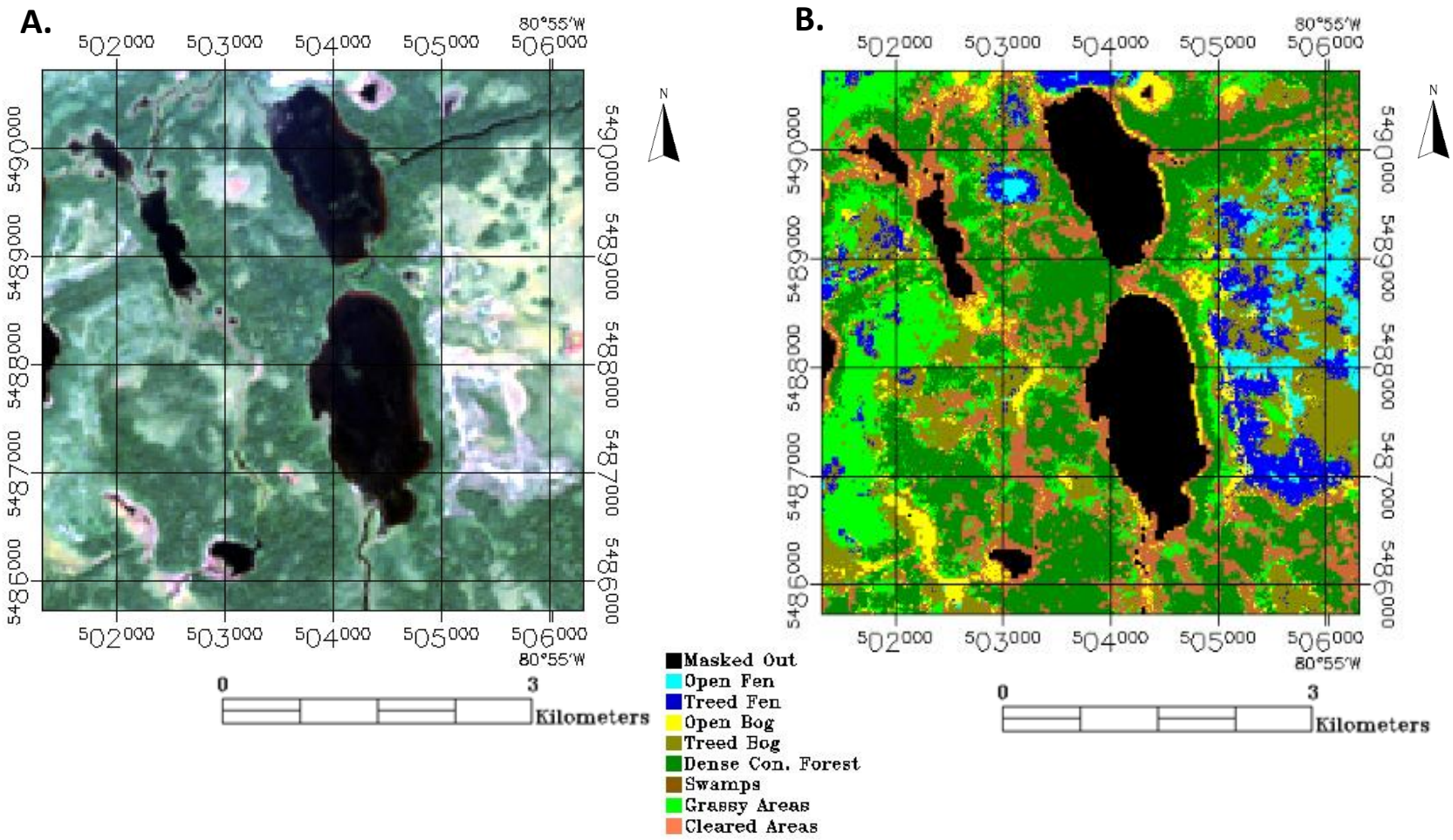

Figure 4-4. Output from classification result. (A) True colour Landsat-5 image from a given test area. (B) Resulting classification result.

4.4. Discussion

4.4.1. Feature Significance Analysis

When examining the results from the feature significance analysis section I note that features taken from spring and summer months generally showed high significance when compared to features taken from fall months, and that image metrics scored higher than image bands or radar bands. This is consistent with the previous chapter [Judah and Hu 2019], which used a different imagery set, and other studies from the field [Bwangoy et al 2010; Davranche et al 2010; Debeau et al 2017; Eisavi et al 2015; Gallant 2015; Masoumi et al 2017; Ozesmi and Bauer 2002; Ramsey and Laine 1997; Tian et al 2016]. The mechanisms I suspect driving this I expand on in the previous chapter. From this analysis it was also noted that images #14 and 15 showed lower quality compared to the others, even though they were from the summer months. As upon closer examination it was discovered that several of the training and evaluation sites had sparse, hard to detect, cloud cover in those images which would explain the poor quality of those features. Also, it was noted that surface temperature could be a significant feature despite its low resolution. Again, this was consistent with the work in my previous chapter [Judah and Hu 2019]. Finally, when examining the Sentinel-1 radar results I note its lower quality when compared to all of the other features. From a ranking standpoint the highest quality Sentinel-1 image is of comparable quality to middle-ranked Landsat-8 images. This is likely due to the natural noisiness of the image even after speckle reduction filtering. However, it was noted throughout the classification results that the inclusion of the lower quality Sentinel-1 images improved classification accuracies, especially with ensemble classification methods which are

effective at investigating structured subsets within a given dataset which could help negate the noise issues with the Sentinel-1 images.

Drawing from the feature significance analysis into the classification results, I generally found consistency between this dataset and the dataset from chapter 3 [Judah and Hu 2019] - Features which were ranked highly from this analysis correlated to higher classification accuracies. In total over 60% of the top 20 tests from all standard classification tests were derived through feature significance analysis with the remaining determined through holistically determined tests - some of which outperformed the tests determined through feature significance analysis. Finally, classification analysis of the lower ranked features showed that use in a hierarchically based classifier produced significant performance increase when compared to a traditional classifier.

4.4.2. Classification – Top Performing Tests

Examining the top ranked classification tests, the hierarchical classifier provided the most accurate results with a clear performance increase. It was noted that the Naïve Bayes classifier had the largest performance increase. However, when examining these results more closely it is worth noting that some of the most accurate tests came from holistically determined tests, using a traditional classifier. While the hierarchical classification approaches also did provide a clear performance increase, in many cases the traditional classification method using holistically determined inputs provided accurate results. In some cases the holistically determined inputs, run through a traditional classifier, outperformed many inputs determined through variable selection in a hierarchical classifier. It was also notable that from the hierarchical classification results the bottom ranked tests were more accurate when compared to the standard classification methods.

In some cases, I note a 1-6% classification accuracy increase. Furthermore, experiments where poor performing tests from a traditional classifier were executed with the hierarchical approach saw performance increases. These tests imply that when only poor or noisy data is available it could benefit more by being executed with a hierarchically based classifier. When examining these results more closely via a confusion matrix, for the least accurate test, I note how for each land cover, the misclassification between other land covers is high – close to an even distribution. However, when the same test is run though a hierarchical classification approach, I note the more structured distribution between land covers and the greater degrees of misclassification lays between land covers types (wetlands and dry land covers). My interpretation is that when these inputs are run through the hierarchical classification approach by forcing the classifier to first focus on producing broad class splits it has better utilized the limited amount of data by focusing on those boarder class differentiations, rather than finer class separations which also might not be possible with those same data. A similar result was observed with Berhane et al 2018. However, it is worth noting that the hierarchical splits with Berhane were determined via an apriori computational accuracy study, and my hierarchical splits were determined by class structures.

To better understand the performance of the most accurate classification result from the hierarchical approaches (RF Case - 2 test #1) I compare and contrast its confusion matrix to the most accurate non-hierarchical classification test's (RF test #131) confusion matrix. I note that in Table 4-8 A), the classification performance of the Open Bog class is the lowest for that test and that its largest classification error is with its conflation with swamps. When examining the confusion matrix in Table 4-8 B) I also note a conflation between open bog and swamps, which also results in the lowest classification accuracy among land covers, however I note a stark

performance difference between the two methods (79.9% vs 14% accuracy). Considering that the Case-2 classification approach performs a broad split between Fens, Bogs and Swamps, it is possible that the open water aspect of Open Bogs would make it susceptible to miss-classification with swamps, however, regardless of this factor the hierarchical classification approach still produced far better results when compared to the standard classification method, indicating that allowing the classifier to focus on splits rather than general classification is advantageous. Furthermore, finer study and refinement of the inputs for the Case-2 classification approach in order to further reduce the misclassification of bogs to swamps will likely increase the classification accuracy to the levels of the other land covers for that test.

When considering the best performing tests with regards to individual landcover performance I noted that bogs and fens were more difficult to classify when compared to swamps (the other wetland class). When examining these results more closely I noted that conflation between treed and non-treed wetlands (in the case of bogs and fens) was a source of increasing errors. This also makes me speculate that defining the training and evaluation data sites as provided by MNR likely had the same difficulties, even with expert interpreters. Also, it is worth noting that landcover classes such as cleared areas and dense coniferous forests encompass a large amount of species which results in broad and complex spectral signatures which are harder to characterize, which results in lower classification accuracies when compared to simpler classes such as swamps.

When examining which features sets which created the most successful results I note that this was dependent on both the classifier which was used and the classification approach which was executed. For instance, for the RF Classifier using its standard approach, tests created using

feature significance analysis made up 55% of its top 20 tests, while for the Naïve Bayes classifier using a standard approach 75% of its top 20 tests were from feature significance analysis.

Upon closer examination of the hierarchical results I note that given the structural differences in the splits between the two hierarchical approaches the resulting inputs performed best when I examine their individual performance in differentiating between wetland types. For instance, the input of the Top 70 RF importance ranked variables was better at differentiating between Treed Wetland and Non-Treed Wetlands, as opposed to the input of All Metrics, and Surface Temperature for all seasons, while the input of All Metrics, and Surface Temperature for all seasons, was better at differentiating between Fens, Bogs, and Swamps. Considering that the quality of the features is determined by their direct relationship between land covers classes it would make sense that input features determined in this fashion would perform the best in the Case-1 classification approach, given that the ultimate split in the hierarchy is directly between wetland land covers classes, as opposed to the Case-2 classification approach where the penultimate split is between a more broad class differentiation between Fens, Bogs, and Swamps, with the ultimate split between treed vs non-treed wetlands. By studying the quality relationship between features for these splits rather than their differences via land covers it is possible that the top ranked classification results will be more consistently represented by inputs determined through variable significance analysis rather than through holistic methods.

4.4.3. Classification – Gross Analysis

When examining the classification results from a more gross standpoint in Figure 4-2, I note that feature analysis both aided in determining which features can benefit and can be detrimental

to classification, which is also consistent with results from chapter 3 [Judah and Hu 2019]. When examining both ends of the scatter plot, I note that it trends towards a decrease in distribution of standard deviation, again this is consistent with results from chapter 3 [Judah and Hu 2019] and the interpretation of those results - for high and low ranked tests, the features used in those tests, generally perform the same across all classification methodologies, feature combinations that were predicted to do poorly, did perform poorly across all classification methodologies, and feature combinations that were predicted to perform well generally produced higher accuracies, with consistency across all classification methodologies.

When a similar plot is created for the hierarchical classification approaches in the form of Figure 4-3, in theory, use of the best feature inputs, regardless of the approach, should yield classification results with matching ranking – the top set of selected inputs should produce the most accurate classification result etc. However, my results do not reflect this. I also note that Figure 4-3 produces a similar structure to that of Figure 4-2 but it is not as pronounced. Examining Figure 4-3 more closely, the results imply that the lowest ranked tests (those with the worst performing feature inputs) are more consistent with one another, compared to the higher ranked tests which show more variation across classifiers and classification approaches. For the highest ranked tests, the variation is lower when compared to the averaged ranked tests indicating a convergence of the distribution but again, it is not as pronounced as the lowest ranked tests. This is consistent with the results from Figure 4-2. When examining the highest ranked tests more closely in the form of Table 4-11 I note that for the Case - 2 classification approach the input ranking matches the classification ranking more closely when compared to the Case - 1 classification approach. The structural difference between the Case - 1 and the Case - 2 classification approach lays with its break down of wetlands. For the Case - 1 approach the

wetland hierarchical breakdown begins between treed vs non-treed wetlands, while Case - 2 approach the wetland hierarchical breakdown begins with fens, bogs, and swamps.

Table 4-11. Top 10 feature inputs for hierarchical classification methods resulting classification rankings.

	Hierarchical - 1		Hierarchical - 2	
	RF	Naïve Bayes	RF	Naïve Bayes
Rank 1 Inputs	17	8	1	1
Rank 2 Inputs	12	5	2	2
Rank 3 Inputs	16	4	3	3
Rank 4 Inputs	15	1	4	5
Rank 5 Inputs	21	2	6	6
Rank 6 Inputs	27	14	7	7
Rank 7 Inputs	31	12	5	4
Rank 8 Inputs	9	3	13	9
Rank 9 Inputs	10	15	8	8
Rank 10 Inputs	34	17	10	10

It would appear that by first breaking down wetlands via species rather than structural differences (treed vs non-treed) are better preserving the quality impact of the selected features. My interpretation is that when inputs are used in conjunction with one another, these inputs did not consistently produce their anticipated level of classification accuracy due in part from the fact that the results from each split is dissimilar enough from the training sets that these discrepancies begin to emerge. A further examination of these results and the exact statistical factors driving these differences will likely lead to more refined feature selection and consistency with classification results.

As with the results from chapter 3 [Judah and Hu 2019] I also noted that the RF generally performed better, across all classification approaches, compared to the probabilistic approach used by the Naïve Bayes classifier. However, there were some other peculiarities with these results. It was noted that the best results from the Case-2 approach executed with a Naïve Bayes produced some of the best classification results – outperforming many of the best classification

results across all classification methods and approaches, at a much lower computational cost. This is a surprising result given how Naïve Bayes results are generally of lower accuracy when compared to those produced with ensemble methods. My interpretation is that some of the factors driving these high performing results are from the way the hierarchical approach mimics the structure of many of the ensemble methods. Also, the pre-selection of features for best performance, within splits in the hierarchy, is similar to how many ensemble methods operate, where majority voting (in the case of RF) or closeness of features in feature space partially determines outcomes. In the case of the hierarchical approach, feature selection and optimization are done before hand, with the Naïve Bayes classifier executing the classification hierarchy. These results imply that, for these purposes, it could be possible to tune a Naïve Bayes classifier to perform at the same level or higher, when compared to some of my best performing RF results, with much lower computation resources. Also, it was noted that there was no clear relationship between the number of variables used and the accuracy of tests. As a general trend, more variables resulted in more accurate results but when examining the most accurate results I noted that there was no clear pattern or trend regarding accuracy. It was observed that there were many instances where feature inputs determined through feature significance analysis, with a lower number of features out performed feature inputs which had a larger number of highly significant features. This is counter intuitive and I hypothesize that noise and or error considerations are the likely cause of this discrepancy and should be investigated further.

Finally, when examining features of the lowest quality, it was found that when inputted in a hierarchical classifier, classification results increased by up to 7 percent when compared to a traditional classifier. My interpretation of these results is that a hierarchical approach of focusing on the separation of broad to more refined classes allows the classifier to maximize the potential

of a low quality dataset by placing inputs into broader classes first, and then focusing on more refined class designations later, reducing the chances of early misclassification. With a higher quality dataset, finer class separations can be done more easily at higher levels of analysis due to the availability of more higher quality information.

4.5. Future Work

When considering how my work can be expanded upon, I note that this research would benefit from the addition of more images from other years, and from other image sources - as a general principle this is true for all classification methodologies. To further develop my work, the addition of additional LANDSAT-8 and Sentinel-2 data, and higher accuracy radar based images would be beneficial. When considering my results in the context of previously published work, others have found that having more continuous and diverse sets of multispectral imagery can lead to better classification results at finer class resolutions [Jiao et al 2019; Ramsey and Laine 2017] as well, finer and more band diverse Radar imagery, especially when used in large amounts can further increase accuracy in finer class separations by providing more nuanced structural information of the land covers by providing better differentiation of plant type and species. [Mahdianpari et al 2017; Mohammadimanesh et al 2018, 2018-2].

Also, a natural next step would be to perform similar classification analysis but with an object orientated approach. For this study I decided to use a pixel based approach in order to capture detailed wetland class features (especially since wetlands generally are of a small footprint). However, by examining land covers via an object orientated approach this would allow me to utilize other features more effectively such as texture and shape. Furthermore, an object orientated approach could arguably be more suited to the incorporation of satellite

imagery by generalizing segmented areas which could in turn reduce noise, and draw out features of those images.

However, it is worth noting that some of the best classification tests produced during this study already have very high accuracies and will probably not show vast improvements by the addition of more data. I speculate that the addition of more images would provide even greater certainly with the variable significance analysis, and possibly improve the accuracies and reduce variances of the average performing classification tests. For some of the peculiar characteristics of the hierarchical results, further development and refinement of those methods via additional feature significance analysis, tailored to each approach in order to create more accurate and consistent results should be investigated. Also, further investigations should be done to better understand why some holistically determined tests outperformed tests created through feature significance analysis. By better understanding these differences I hope to create a more complete way of pre-selecting features in order to create efficient and accurate classification results. Finally, when considering how this work can be adapted to other study areas, I note that northern hemisphere temperate forests are all very similar in structure and vegetation distribution. The work done with this research should be sufficiently general with only minor local considerations from the study site. The methodologies and results produced in this study should be able to be applied without much difficulty to other northern hemisphere temperate study areas, in Canada or other parts of the world. Applying this work to tropical environments would likely be less compatible given the difference in vegetation density, vegetation types, and the lack of large seasonal variations with that vegetation. However, if given the appropriate datasets, the study methodology used here should be able to produce similar variable analysis and classification results, which would be an interesting contrast to this work.

4.6. Conclusions

Two hierarchically based classification approaches, for wetlands, were developed based on starting with broad class separations and moving towards increasingly refined classes. The approaches differed in how they initially separated wetlands by either creating splits between treed vs non-treed wetlands (Case-1) or separating wetlands into the classes of Fens, Bogs, and Swamps (Case-2). For both of these hierarchically based approaches classification accuracy increases were observed when compared to a traditional classifier. Case-2 produced the most accurate classification results where I note an overall performance gain for the top 10% of results for both the Naïve Bayes and the RF classifier when compared to a traditional classifier. I suspect that this performance difference is driven from breaking down wetlands via species rather than structural differences (treed vs non-treed) which is better at preserving the quality impact of the selected features. For the most accurate classification result (91.94%), for all tests, a result from RF, executed through the Case-2 approach, outperformed the most accurate result from the traditional classification method by some 11%. It is also worth noting that Naïve Bayes classification results, using the Case - 2 approach, showed a great improvement to its accuracy, in one case ~12% increase when compared to the highest accuracy produced through the traditional approach. Further examinations of the top ranked test from the hierarchical approach showed that with further tuning of feature inputs it could be possible to further increase classification accuracy.

A notable result from the hierarchical classification results was that bottom ranked tests were of a higher accuracy when compared to the traditional classification methods for the same sets of features. In some cases I note a 2-6% classification accuracy increase. Furthermore I noted the peculiar result in that when the same test, from the standard classification method,

executed with the same classifier, is performed using the hierarchical approach I saw an increase in accuracy (3-6% increase). These are not insignificant performance increases which implies that when low quality data is available it could benefit from being executed with a hierarchically based classifier.

Further examinations of some of the more peculiar aspects of these results and efforts should be made to assign more quantifiable physical explanations of those peculiarities should be carried out. This is a challenging proposition but one that is worthwhile. This will possibly lead to a generalized framework for wetland classification which can be used as a product but will also provide a level of expectation when it comes to the ultimate accuracy that this style of analysis can produce. This in turn will aid in determining the next steps required to achieve the next level of accuracy or detail.

Chapter 5 - Advanced Data Fusion Method Using Multi-Source Remotely Sense Data for Improving Wetland Classification Accuracy and Reducing Misclassification Errors

A portion of the research in this chapter is accepted in the following journal paper:

[1]Judah, Aaron, and Baoxin Hu. 2022. "An Advanced Data Fusion Method to Improve Wetland Classification Using Multi-Source Remotely Sensed Data" *Sensors* 22, no. 22: 8942. https://doi.org/10.3390/s22228942

---

[2]I would like to thank the publisher MDPI and Prof. Hu who have granted me permission to reuse this article in my dissertation.

The motivation for the research presented in this chapter is to improve wetland (Fen, Bog, Marsh, Swamps) classification accuracy and confidence in those classification results in part from the reduction of high confidence misclassified pixels – error minimization I identified as an area of improvement for the field. The methodology I developed uses remotely sensed, and ancillary data processed through multiple RF derived classifiers, fused through D-S Theory in what I describe as an Ensemble Classifier. The data utilized included multispectral optical and thermal data (Landsat-8 and Sentinel-2), Radar imagery (Sentinel-1), and DEM. More specific goals were to determine the best way to combine imagery, in the context of being classified using three different RF based classifiers (a standard classification approach, and two classifiers focused on broad class separations), using D-S theory, in order to best utilize the available data to produce more accurate and efficient classification maps compared to a standard classifier. The Ensemble Classifier used a RF classifier to execute the classifications and was found to increase classification accuracy when compared to a standard classification approach using similar data for similar land covers. The Ensemble Classifier also reduced the number of high confidence - This provides increased confidence in classification results. This study also found that classifiers focused on broad class separations required a different number of features in order to maximize classification accuracy, and that a holistic method to selecting features was more effective than a strictly analytical methodology for feature selection – something echoed in my previous studies.

5.1. Introduction

This study is unique in two aspects considering the proposed method and analysis. For most existing methods related to ensemble classifiers and/or decision-level fusion, multiple classifiers are employed to deal with identical sets of classes [Bui et al 2022; Jia et al 2021; Guo 2019; Feng et al 2021]. This research fully leveraged the diversified features derived from available multi-source remotely sensed data to be used in individually designed classifiers with unique class propositions. Furthermore, prior knowledge on wetland cover types and remotely sensed data used was also utilized in the selection of features for each classifier, in addition to the data-driven machine learning approaches that were commonly and exclusively used in most classification methods. One might argue that hierarchical classification methods effectively utilized features to separate different categories of cover types. However, the uncertainty associated with the classifiers in the hierarchy was not addressed [Judah and Hu et al 2021] and was difficult to be accounted for in the lower parts of the hierarchy. The D–S theory used in this study provides an effective means to consider the uncertainty in each classifier. Similarly, to the argument on rule-based past-processing applied to classification maps, the advantage of this method in in its handling of uncertainty and avoiding the selection of thresholds in any rule-based methods. In this study, detailed analysis of the nature of misclassification was also carried out, which was lacking in the literature [Bo et al 2016; Bui et al 2022; Jia et al 2021; Guo 2019; Feng et al 2021].

In most classification publications, only accuracy-related measures were reported. However, it is also important to examine the uncertainty of the classification results and the nature of the misclassification [Hu et al 2021]. It is common that the uncertainty is high for some of the misclassified pixels. However, for some misclassified pixels, the uncertainty might be low,

which is more problematic in the utilization of the classification results. The same is true for the correctly classified pixels but with high uncertainty. In my previous studies [Judah and Hu 2019, 2021], where a random forest (RF) classifier was executed on a concatenating set of features derived from multi-source remotely sensed data, up to 30% of misclassified pixels were classified with a confidence level of greater than 85%. The results also showed that, for about 40% of the misclassified pixels, the correct class (ground truth) was found to be the second choice, and, in some cases, the difference in the posterior probability between the top two categories was only ~0.05. The method used in [Judah and Hu 2019, 2021] is commonly used for cover type classification using multi-source remotely sensed data, as discussed more in detail. An apparent conclusion is that the aggregated features may not be able to reliably separate the land-cover types of interest. However, this does not mean that these cover types cannot be separated by those features. With the aggregation, some features important to separate certain cover types might be masked by others. In that vein, classification may be improved if those features are utilized differently, in order to best exploit their discriminant power within the context of a particular separation of classes or class types.

In order to address the aforementioned issued and best exploit the discriminate power of individual or families of features, an ensemble classifier using a feature and decision-level fusion framework was developed. Leveraging prior knowledge and all available data in the study area, three classifiers were first designed to reliably distinguish individual or compound classes among the five cover types (fen, swamp, marsh, bog, and upland), executed in parallel with one another using a RF classifier, and the results from these classifiers were then combined according to the D–S theory. The base of this ensemble classifier was the commonly used (also referred to as the traditional method) feature-based fusion method (Classifier #1) for the classification of the five

individual classes (fen, swamp, marsh, bog, and upland). As discussed later, with the traditional method, some features known to have high discriminant powers in separating some broad classes (such as wetland and upland) are often not selected using automatic feature selection methods. This may lead to some confusion between wetland and upland classes due to the absence of these features. To overcome this problem, two additional classifiers (Classifiers #2 and #3) were designed to classify compound cover types. For Classifier #2, two broad cover types were classified—wetland (fen, swamp, marsh, bog) vs. dry land covers (upland). For Classifier #3 the focus was on separating more structured land covers (swamp and upland) to less structured land covers (fen, bog, and marsh). Due to the uncertainty expected from any classification method, the D–S theory was employed to combine the results from these classifiers.

Figure 5-1 outlines the overall workflow for my approach, and the details are described below.

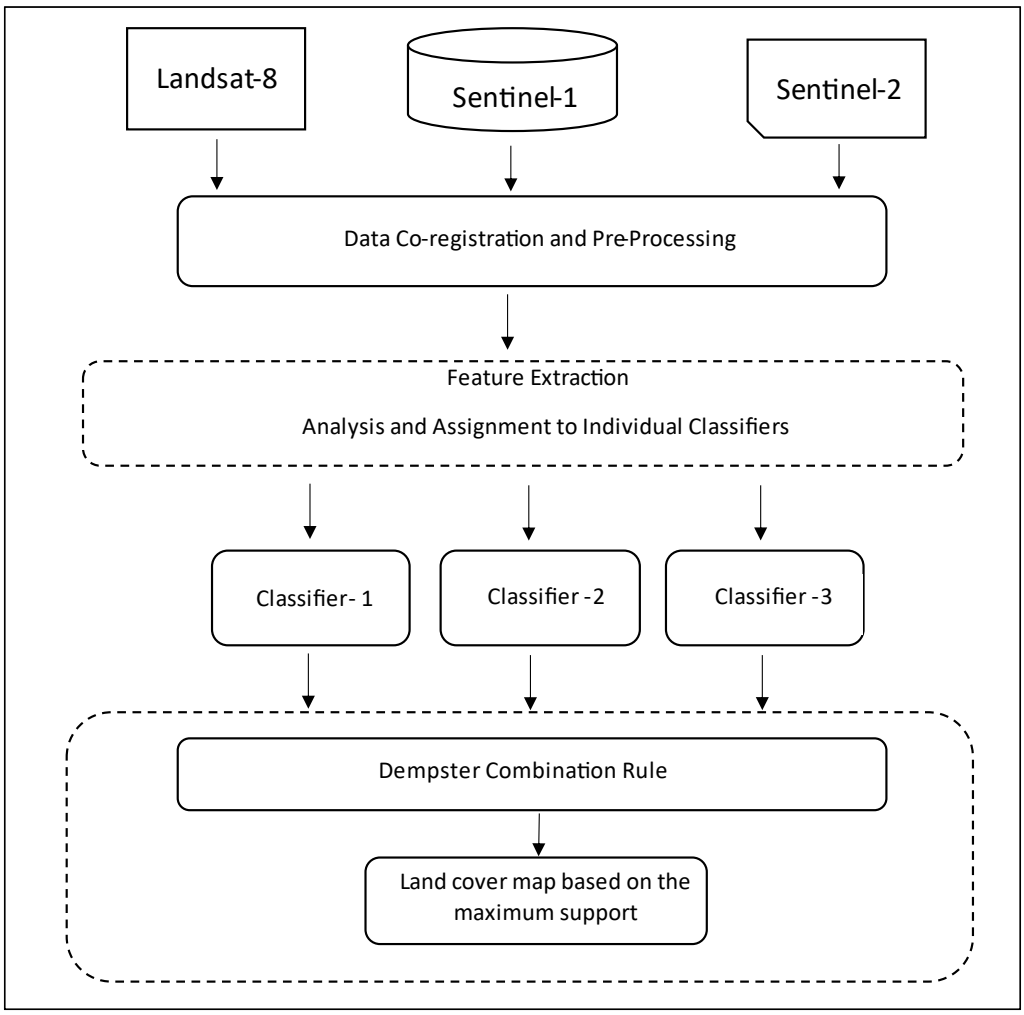

Figure 5-1. Workflow of the proposed ensemble classifier combining the results of three different classifiers based on D–S theory.

The remainder of the chapter is structured as follows: in Section 2, the study area and images used are described; the methodology including data processing, feature extraction and selection, and the developed ensemble classifier is documented in Section 3; results and a discussion are presented in Sections 4 and 5, respectively; in Section 6, conclusions and future work are provided.

5.2. Study Area and Images Used

The study area was selected from a location in Northern Alberta due to the availability of the Alberta Biodiversity Monitoring Institute (ABMI) wetland inventory [ABMI 2017]. Figure 5- illustrates the rough approximate area of interest. This wetland inventory comprises five different land cover classes (fens, bogs, marshes, swamps, and upland), identified and mapped out using photo interpreted data [ABMI 2017]. The ABMI wetland inventory is parsed out in individual study areas throughout Northern Alberta, Canada, each approximately 21 km^2 in size. For this study, 10 sites, as shown in Figure 5-, were selected because of the domination by wetland cover types. Collection and analysis for the photo data were completed in 2016 [ABMI 2017].

The land-cover classes identified in the ABMI wetland inventory (bogs, fens, marshes swamps, and upland) and their detailed descriptions can be found in [ABMI 2017]. Below, the characteristics of these cover types more relevant to remote sensing data interpretation are summarized. Bogs are hydrologically isolated peatlands, receiving water from precipitation only. They are stagnant, with low nutrient availability, and support low biological diversity. Bogs typically have a low water table, appearing dry at the surface.

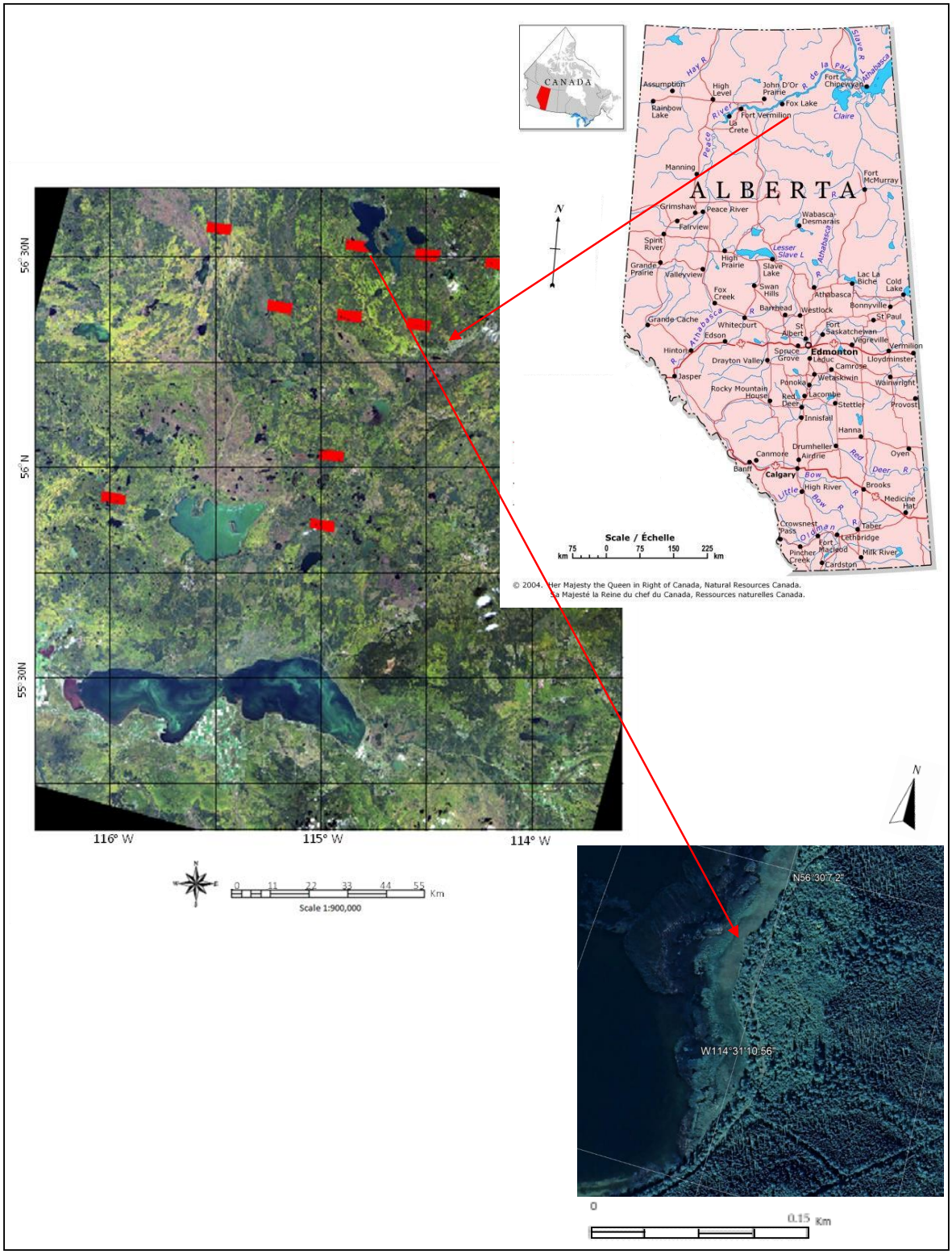

Figure 5-2. The study area from a geographic perspective together with a Landsat-8 True-Color image (RGB Bands 4, 3, 2) and aerial imagery. Individual study areas are highlighted as red polygons, drawn from the ABMI wetland inventory dataset on the Landsat-8 image.

Fens are also peatlands, but hydrologically connected. Fens can be nutrient-poor or -rich, depending on nutrient input from water sources. Fens often have high water tables and connect wetland systems over great distances. Marshes are mineral wetlands. They exhibit a variable water table, which can vary throughout the season. Marshes receive water from a combination of ground water, runoff, and precipitation, as well as through connecting streams. They are periodically dry, with nutrient-rich soil, promoting the growth of a diverse range of emergent, grass-like vegetation. Swamps are considered mineral wetlands, although they may also exist as peatlands in some cases, with woody plant cover that comprises more than 25% of the total area. Swamps receive water from a combination of ground water, runoff, and precipitation. Water movement ranges from stagnant to dynamic. Swamps typically represent transition zones between other wetlands and non-wetland areas, known as uplands, and support high biological diversity.

The upland class is a broad non-wetland class created by the ABMI in order to encompass non-wetland land covers such as grassy areas, cleared areas, and sparse and dense forests of various species. This includes upland deciduous, mixed-wood, and coniferous stands (age classes combined), grassy areas, and shrub areas [ABMI 2017].

For individual study areas shown in Figure 5-, ground-truth data were provided in the ABMI dataset. The number of labeled pixels for each study area was determined from the size of land-cover plots identified by ariel imagery and ground survey data as per the ABMI wetland maps. As shown by the example of areas identified as swamp in Figure 5-3, the labeled pixels were clustered in areas. In the selection of training and validation data, the groupings of pixels were maintained. On average, approximately 64% of the identified pixels were used for training, with

the remainder used for validation. These pixel and land-cover assignments are summarized in Table 5-1.

Table 5-1. Land-cover class assignment and the number of pixels contained in the training and validation set.

Class	Number Assigned to Class	Number of Pixels in Training Set	Number of Pixels in Validation Set
Fen	1	288,343	156,102
Bog	2	36,637	14,479
Marsh	3	25,309	23,416
Swamp	4	109,490	91,510
Upland	5	2,314,364	636,441

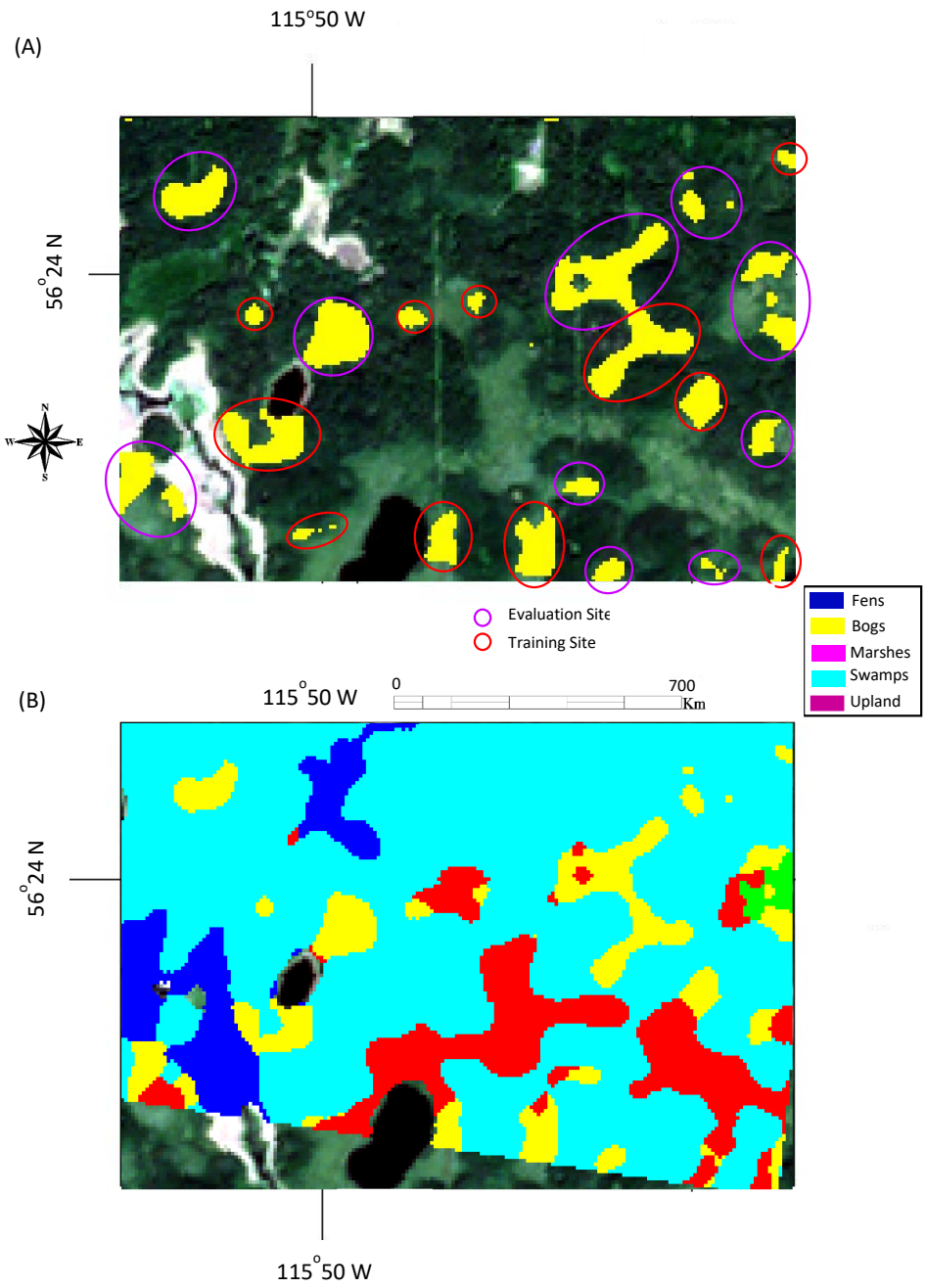

Figure 5-3. Example of a study area highlighting the individual evaluation and training sets for swamps (**A**) and its corresponding cover type map (ground truth) (**B**).

Landsat-8, Sentinel-2, and Sentinel-1 imagery represented the primary image sources used in this study. Landsat-8 and Sentinel-1 imagery are previously described in sections 3.1 and 4.1. Attempts were made to acquire imagery close to 2016–2017 to match the collection dates of the

aerial imagery used to create the ABMI dataset. However, additional images from other dates were also collected in order to create a more robust dataset. It should also be noted that Sentinel-1 imagery coverage of the study area was not available until 2017.

The Sentinel-2 imagery used was the Level 2A bottom-of-atmosphere reflectance in the cartographic geometry imagery product. These images have a resolution of 10 m by 10 m and contain four bands. These bands are centered on 492.4 nm, 559.8 nm, 664.6 nm, and 832.8 nm—blue, green, red, and near-infrared (NIR) respectively [European Space Agency 2021]. Sentinel-2 imagery was chosen due to its availability, higher resolution compared to Landsat-8, and spectral bands which are useful in characterizing both vegetation and water.

Lastly, a DEM of the study area taken from the Canadian Digital Surface Model [Natural Resources Canada Map Information Branch. 2016] at a spatial resolution of 30 m by 30 m with an associated DEM derived slope was used.

In total, three Landsat-8 images, seven Sentinel-2 images, and four Sentinel-1 images were collected. It should also be noted that the Landsat-8 and Sentinel-1 imagery was resampled to 10 m by 10 m in order to facilitate ease of analysis with the other imagery products.

Table 5-2 summarizes the dates and types of imagery that were collected for this study.

Table 5-2. Summary of satellite imagery collected for this study.

Image ID	Imagery	Season	Date	Level of Processing	Accessed From
#1	Landsat-8	Summer	27 July 2015	Level 1G	United States Geological Survey (USGS)
#2	Landsat-8	Fall	15 September 2016	Level 1G	USGS
#3	Landsat-8	Fall	10 September 2020	Level 1G	USGS
#4	Sentinel-2	Fall	17 September 2017	Level 2A	European Space Agency (ESA)—Sentinel
#5	Sentinel-2	Summer	11 August 2017	Level 2A	ESA—Sentinel
#6	Sentinel-2	Summer	2 September 2018	Level 2A	ESA—Sentinel
#7	Sentinel-2	Fall	29 September 2020	Level 2A	ESA—Sentinel
#8	Sentinel-2	Summer	28 June 2021	Level 2A	ESA—Sentinel
#9	Sentinel-2	Summer	1 July 2021	Level 2A	ESA—Sentinel
#10	Sentinel-2	Summer	28 July 2021	Level 2A	ESA—Sentinel
#11	Sentinel-1	Summer	12 August 2018	Level 1—SLC	ESA—Sentinel
#12	Sentinel-1	Summer	27 July 2019	Level 1—SLC	ESA—Sentinel
#13	Sentinel-1	Fall	19 September 2020	Level 1—SLC	ESA—Sentinel
#14	Sentinel-1	Summer	9 August 2021	Level 1—SLC	ESA—Sentinel

5.3. Methodology

In this study, an ensemble classifier using a feature- and decision-level fusion framework was developed. Leveraging prior knowledge and all available data in the study area, three classifiers were first designed to reliably distinguish individual or compound classes among the five cover types (fen, swamp, marsh, bog, and upland), executed in parallel with one another using a RF classifier, and the results from these classifiers were then combined according to the D–S theory. The base of this ensemble classifier was the commonly used (also referred to as the traditional method) feature-based fusion method (Classifier #1) for the classification of the five

individual classes (fen, swamp, marsh, bog, and upland). As discussed later, with the traditional method, some features known to have high discriminant powers in separating some broad classes (such as wetland and upland) are often not selected using automatic feature selection methods. This may lead to some confusion between wetland and upland classes due to the absence of these features. To overcome this problem, two additional classifiers (Classifiers #2 and #3) were designed to classify compound cover types. For Classifier #2, two broad cover types were classified—wetland (fen, swamp, marsh, bog) vs. dry land covers (upland). For Classifier #3 the focus was on separating more structured land covers (swamp and upland) to less structured land covers (fen, bog, and marsh). Due to the uncertainty expected from any classification method, the D–S theory was employed to combine the results from these classifiers.

As previously presented Figure 5-1 outlines the overall workflow for my approach, and the details are described below.

5.3.1. Features and Their Derivation

Furthermore, 184 candidate features were derived and are summarized in Table 5-3. The calculations and analyses performed to produce these features are described below.

For this study, 11 different types of remotely sensed features were used. They were vegetation indices, surface albedo, and textual measures derived from multi-spectral imagery, surface temperature from the thermal bands of multi-spectral imagery, backscatter coefficients and derived features from SAR imagery, and DEMs and features derived from DEMs. These features were selected in order to characterize vegetative activity, water content, radiometric absorption, horizontal structure and roughness, water content of surface objects, and topography. It is worth mentioning that textual features derived from Sentinel-2 imagery were selected due to

their success in the classification of land covers in the popular literature [Huaxin et al. 2022; Jiao et al. 2019; Juan et al. 2022; wu et al. 2021] and from my own observations. This is further expanded upon in Sections 4 and 5. Surface temperature, from my past study [Judah and Hu 2019] highlighted in chapter 3, was shown to be useful in classifying wetland types.

Specifically, the vegetation indices used included the normalized difference vegetation index (NDVI), the enhanced vegetation index (EVI), and the near-infrared reflectance vegetation index (NIRv). NDVI is a popular and standard vegetation index sensitive to leaf area index, coverage, pigment content of vegetation canopies, and vegetative photoactivity [Huete et al 2002; Lin et al 2019; Rocha and Shaver 2009]. EVI is defined as

$$\mathrm{EVI} = \frac{2.5 \times (\mathrm{R_{B8}} - \mathrm{R_{B4}})}{\mathrm{R_{B8}} + 6 \times \mathrm{R_{B4}} - 7.5 \times \mathrm{R_{B2}} + 1}, \tag{5-1}$$

where $\mathrm{R_{B2}}$, $\mathrm{R_{B4}}$, and $\mathrm{R_{B8}}$ are the reflectance at spectral bands 2, 4, and 8, from Sentinel-2 imagery, respectively. EVI was not calculated using Landsat-8 imagery due to early tests which found that EVI using Sentinel-2 imagery was of a much greater significance during classification. EVI has been shown to be effective in characterizing vegetation features such as leaf area index, temporal changes in vegetative activity and resolving vegetation differences from areas which have complex background surface reflectance [Huete et al 2002; Lin et al 2019; Rocha and Shaver 2009]. NIRv, a near-infrared reflectance vegetation index, is defined as

$$\mathrm{NIRv} = \frac{\mathrm{R_{B8}} - \mathrm{R_{B4}}}{\mathrm{R_{B8}} + \mathrm{R_{B4}}} \times \mathrm{R_{B8}}. \tag{5-2}$$

Its success in characterizing vegetation in a mixed pixel environment and low leaf areas has been reported in the literature [Badgley et al 2017]. Again, NIRv was calculated for Sentinel-2

imagery only due to its larger significance when compared to NIRv calculated with Landsat-8 imagery. NDWI works on a similar principle to NDVI but is designed to be sensitive to water content rather than to photosynthetic activity. NDWI has been previously described in section 3.2.2. Due to its success in many applications, NDWI is a standard layer product for the Moderate Resolution Imaging Spectroradiometer the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor [Hubanks et al 2008].

Surface albedo was calculated using the standard approach defined by Liang et al. [1999] and described in section 3.2.2.

Surface temperature was calculated for individual pixels from Landsat-8 imagery using the standard methodology from the Landsat-8 (L8) Data Users Handbook [Landsat 8 data Users Handbook, 2015].

The textural features were derived from Sentinel-2 imagery, due to its relatively higher spatial resolution in comparison with that of Landsat 8. The three texture features (mean, variance, and entropy) were calculated within a window size of 4 × 4 pixels for the four Sentinel-2 imagery bands, using the standard software suites in ENVI 5.6, and they are defined in Equations (5-3) - (5-5). This window size was determined empirically.

Mean texture is defined as:

$$\text{Mean(M)} = \sum_{i=0}^{N_g-1} iP(i) \tag{5-3}$$

Where N_g is the number of distinct grey levels in the quantized image, and $P(i)$ is the probability of each pixel value [Hall-Beyer 2021].

Variance texture is defined as:

$$\text{Variance} = \sum_{i=0}^{N_g-1} (i - M)^2 P(i) \tag{5-4}$$

Where N_g is the number of distinct grey levels in the quantized image, and $P(i)$ is the probability of each pixel value and M is the mean as defined in (X) [Hall-Beyer 2021].

Entropy texture is defined as:

$$\text{Entropy} = -\sum_{i=0}^{N_g-1} P(i) * lnP(i) \tag{5-5}$$

In these equations, Ng is the number of distinct gray levels in the quantized image, P(i) is the probability of the occurrence of each grey level, and M is the mean as defined in Equation (5-3) [Hall-Beyer 2021]. In this study, Ng was determined automatically by ENVI through the available quantization range of the imagery.

The backscatter coefficients in VV and VH, denoted as σ_{vv} and σ_{vH}, respectively, were obtained from the calibrated Level 1 SLC product of Sentinel-1 [European Space Agency 2018]. In order to reduce noise, the enhanced frost speckle filter from PCI Geomatica with a 5 × 5 pixel window was used to filter all Sentinel-1 imagery. The window size was chosen on the basis of empirical analysis. After processing, the Sentinel-1 imagery was georeferenced to the Sentinel-2 imagery.

From the Sentinel-1 imagery, I also produced an adaption of the quad-polarization of the SAR vegetation index (RVI) proposed by Periasamy [2018], i.e., the dual-polarization SAR vegetation index (DPSVI), defined as

$$\text{DPSVI} = \frac{\sigma_{vv} + \sigma_{vH}}{\sigma_{vv}}. \tag{5-6}$$

This index has been found to be a significant feature in separating different types of crops and from separating land covers of high vegetation water content from land covers better characterized by dry biomass.

Additionally, the DEM, DEM-derived slope, and valley bottom flatness (VBF) were used. Slope was calculated from the DEM using the ENVI 5.6 topographic modeling function with a 3 × 3 window. The DEM and DEM-derived slope were selected to determine the role geographic features play in distinguishing wetland classes. For instance, it is known that some species of fens prefer to grow on slopes. VBF was calculated using the open-source GIS software suite System for Automatic Geoscientific Analysis (SAGA) using the processed DEM data as previously described. VBF measures the degree of valley bottom flatness at multiple scales [Gallant et al 2003]. Large flat valleys are typical of landscapes for wetlands, once open water has been masked from the data. Experiments were conducted while varying slope thresholds, where it was found that a slope threshold of 17 produced the most significant VBF feature. VBF has been found to be a very significant feature in the classification of wetlands from the ABMI dataset, as reported by the Alberta Biodiversity Monitoring Institute [ABMI 2021].

Table 5-3. Features used during this study and their associated variable index. Reflection is shortened to "Reflect." and Sentinel is shortened to "Senti." M, V, and E correspond to the mean, variance, and entropy texture, respectively. The number at the end of each feature name refers to the image ID in Table 5-2.

Index	Name	Index	Name	Index	Name	Index	Name
1	B1 Reflect. #1	47	B3 Senti. 2 #6	93	B4-M-Senti. 2 #6	139	B2-E-Senti. 2 #4
2	B2 Reflect. #1	48	B4 Senti. 2 #6	94	B1-M-Senti. 2 #7	140	B3-E-Senti. 2 #4
3	B3 Reflect. #1	49	B1 Senti. 2 #7	95	B2-M-Senti. 2 #7	141	B4-E-Senti. 2 #4
4	B4 Reflect. #1	50	B2 Senti. 2 #7	96	B3-M-Senti. 2 #7	142	B1-E-Senti. 2 #5
5	B5 Reflect. #1	51	B3 Senti. 2 #7	97	B4-M-Senti. 2 #7	143	B2-E-Senti. 2 #5
6	B6 Reflect. #1	52	B4 Senti. 2 #7	98	B1-M-Senti. 2 #8	144	B3-E-Senti. 2 #5
7	B7 Reflect. #1	53	B1 Senti. 2 #8	99	B2-M-Senti. 2 #8	145	B4-E-Senti. 2 #5
8	NDVI #1	54	B2 Senti. 2 #8	100	B3-M-Senti. 2 #8	146	B1-E-Senti. 2 #6
9	NDWI #1	55	B3 Senti. 2 #8	101	B4-M-Senti. 2 #8	147	B2-E-Senti. 2 #6
10	Albedo #1	56	B4 Senti. 2 #8	102	B1-M-Senti. 2 #9	148	B3-E-Senti. 2 #6
11	Temp1 #1	57	B1 Senti. 2 #9	103	B2-M-Senti. 2 #9	149	B4-E-Senti. 2 #6
12	Temp2 #1	58	B2 Senti. 2 #9	104	B3-M-Senti. 2 #9	150	B1-E-Senti. 2 #7
13	B1 Reflect. #2	59	B3 Senti. 2 #9	105	B4-M-Senti. 2 #9	151	B2-E-Senti. 2 #7
14	B2 Reflect. #2	60	B4 Senti. 2 #9	106	B1-M-Senti. 2 #10	152	B3-E-Senti. 2 #7
15	B3 Reflect. #2	61	B1 Senti. 2 #10	107	B2-M-Senti. 2 #10	153	B4-E-Senti. 2 #7
16	B4 Reflect. #2	62	B2 Senti. 2 #10	108	B3-M-Senti. 2 #10	154	B1-E-Senti. 2 #8
17	B5 Reflect. #2	63	B3 Senti. 2 #10	109	B4-M-Senti. 2 #10	155	B2-E-Senti. 2 #8
18	B6 Reflect. #2	64	B4 Senti. 2 #10	110	B1-V-Senti. 2 #4	156	B3-E-Senti. 2 #8
19	B7 Reflect. #2	65	Senti. VV-#11	111	B2-V-Senti. 2 #4	157	B4-E-Senti. 2 #8
20	NDVI #2	66	Senti. VH-#11	112	B3-V-Senti. 2 #4	158	B1-E-Senti. 2 #9
21	NDWI #2	67	Senti. VV-#12	113	B4-V-Senti. 2 #4	159	B2-E-Senti. 2 #9
22	Albedo #2	68	Senti. VH-#12	114	B1-V-Senti. 2 #5	160	B3-E-Senti. 2 #9
23	Temp1 #2	69	Senti. VV-#13	115	B2-V-Senti. 2 #5	161	B4-E-Senti. 2 #9
24	Temp2 #2	70	Senti. VH-#13	116	B3-V-Senti. 2 #5	162	B1-E-Senti. 2 #10
25	B1 Reflect. #3	71	Senti. VV-#14	117	B4-V-Senti. 2 #5	163	B2-E-Senti. 2 #10
26	B2 Reflect. #3	72	Senti. VH-#14	118	B1-V-Senti. 2 #6	164	B3-E-Senti. 2 #10
27	B3 Reflect. #3	73	DEM	119	B2-V-Senti. 2 #6	165	B4-E-Senti. 2 #10
28	B4 Reflect. #3	74	Slope	120	B3-V-Senti. 2 #6	166	EVI Senti. 2 #4
29	B5 Reflect. #3	75	NDVI Senti. 2 #4	121	B4-V-Senti. 2 #6	167	NIRv Senti. 2 #4
30	B6 Reflect. #3	76	NDVI Senti. 2 #5	122	B1-V-Senti. 2 #7	168	EVI Senti. 2 #5
31	B7 Reflect. #3	77	NDVI Senti. 2 #6	123	B2-V-Senti. 2 #7	169	NIRv Senti. 2 #5
32	NDVI #3	78	NDVI Senti. 2 #7	124	B3-V-Senti. 2 #7	170	EVI Senti. 2 #6
33	NDWI #3	79	NDVI Senti. 2 #8	125	B4-V-Senti. 2 #7	171	NIRv Senti. 2 #6
34	Albedo #3	80	NDVI Senti. 2 #9	126	B1-V-Senti. 2 #8	172	EVI Senti. 2 #7
35	Temp1 #3	81	NDVI Senti. 2 #10	127	B2-V-Senti. 2 #8	173	NIRv Senti. 2 #7
36	Temp2 #3	82	B1-M-Senti. 2 #4	128	B3-V-Senti. 2 #8	174	EVI Senti. 2 #8
37	B1 Senti. 2 #4	83	B2-M-Senti. 2 #4	129	B4-V-Senti. 2 #8	175	NIRv Senti. 2 #8
38	B2 Senti. 2 #4	84	B3-M-Senti. 2 #4	130	B1-V-Senti. 2 #9	176	EVI Senti. 2 #9
39	B3 Senti. 2 #4	85	B4-M-Senti. 2 #4	131	B2-V-Senti. 2 #9	177	NIRv Senti. 2 #9
40	B4 Senti. 2 #4	86	B1-M-Senti. 2 #5	132	B3-V-Senti. 2 #9	178	EVI Senti. 2 #10
41	B1 Senti. 2 #5	87	B2-M-Senti. 2 #5	133	B4-V-Senti. 2 #9	179	NIRv Senti. 2 #10
42	B2 Senti. 2 #5	88	B3-M-Senti. 2 #5	134	B1-V-Senti. 2 #10	180	Senti. DPSVI-#11
43	B3 Senti. 2 #5	89	B4-M-Senti. 2 #5	135	B2-V-Senti. 2 #10	181	Senti. DPSVI-#12
44	B4 Senti. 2 #5	90	B1-M-Senti. 2 #6	136	B3-V-Senti. 2 #10	182	Senti. DPSVI-#13
45	B1 Senti. 2 #6	91	B2-M-Senti. 2 #6	137	B4-V-Senti. 2 #10	183	Senti. DPSVI-#14
46	B2 Senti. 2 #6	92	B3-M-Senti. 2 #6	138	B1-E-Senti. 2 #4	184	VBF-10

As a final note regarding the imagery and features used in this study, in order to preserve the information from the higher-resolution Sentinel-2 imagery, all images were resampled to 10 m by 10 m resolution when layers were stacked together.

5.3.2. Feature Selection

As mentioned earlier, three classifiers were designed in this study, and two feature selection methods were employed. For Classifier #1 (see details in the next section) where all cover types were identified, the built-in feature selection mechanism in the RF classification was used. This was to fully utilize the abovementioned extracted features and maximize their discriminant power in the classification. For Classifiers #2 and #3, where broad cover types (compounds of cover types) were considered, the feature selection was conducted on the basis of prior knowledge and experimentation. This was applied to maintain the independency in the features used for these three classifiers to avoid any bias in the fusion process. Furthermore, it was also noted in a previous study [Judah and Hu 2019] that a subset of an analyzed and ranked set of features could be outperformed, in a classification setting, by a set of features selected through a holistic approach. This is further expanded upon in Section 5.4.

The RF importance value is determined through an iterative exploration of the dataset [Biau and Scornet 2015]. It is computed by summing changes in the percentage increase in mean squared error (MSE) due to splits on every predictor and dividing the sum by the number of branch nodes for that tree, averaged over all trees. These calculations are performed on all input features, with larger values implying that a feature is more significant. Additionally, it was observed from my previous work presented in chapter 3 and in Judah and Hu 2019, that there was a plateau in classification accuracy once a specific number of data features was reached. With the increase in the number of features in a given classification, it was likely that the redundance among those features was increased, implying that there is a ceiling to the classification accuracy for a given dataset. Furthermore, any noise and confliction among a large number of features might negatively affect the classification accuracy. Keeping the

aforementioned in mind, when selecting sets of features, I was cognizant of identifying the appropriate number of features in order to avoid redundancy and noise. This is further expanded upon in Section 5.4.

5.3.3. The Ensemble Classification Method Based on the D–S Theory

My previous investigation presented in Chapter 4 [Judah Hu 2020] showed that, in the context of wetland classification, in a hierarchical framework, certain features can separate and classify a group of wetland types more effectively and more reliably when compared to a group focused on distinguishing individual types. One disadvantage of a hierarchical framework lies in the fact that the misclassification in the higher hierarchy is propagated to the subsequent levels of classification. To address this issue, three classifiers with different propositions were designed and carried out first, and their results were then combined according to the D–S theory. In this way, the uncertainty associated with each classifier was considered.

In Classifier #1, individual wetland cover types were considered. The classification propositions were fen, bog, marsh, swamp, and upland. This classifier type is commonly used; thus, it was taken as the baseline method for comparison. For Classifier #2, two broad cover types were classified—wetland vs. dry land covers. This classifier would utilize features which excel at identifying moisture, and flat structural features in pixels such as water indices, SAR backscatter coefficients, and DEM and its derivatives. For Classifier #3, the focus was on separating more structured land covers (swamp and upland) from less structured land covers (fen, bog, and marsh). For Classifier #3 the use of SAR features and textural features was leveraged given their performance advantages in those areas.

As mentioned in the previous section, for Classifier #1, a suitable set of features were selected using the RF feature selection method. For Classifiers #2 and #3, feature selection was conducted on the basis of prior knowledge in the separation of the two broad classes. The number of features used for each classifier was determined by observing when the classification accuracy was no longer improved from additional features – this asymptotic behaviour determined the number of input features for that classifier.

The RF classifier is an ensemble, supervised, machine learning algorithm. It operates by constructing multitudes of decision trees with the ultimate class of a given input determined by the majority vote from those decision trees [Biau and Scornet 2015; Breiman 2001; Kecman et al 2005]. With RF, diversification of the decision trees is accomplished by developing those trees from various subsets created through bagging or bootstrap aggregating from the training data [Breiman 2001]. RF lends itself well to parallelization and computational streamlining for investigating the nuances of large datasets. This has led RF to become one of the most successful and widely implemented machine learning algorithms to date [Biau 2015; Breiman 2001]. RF generally requires two main input parameters: the number of trees to grow and the depth or complexity of those trees (*p*-value). More trees generally result in higher classification accuracies but at greater computational costs. However, at some point, increasing the number of trees no longer increases classification accuracy. Similarly, choosing a tree depth that is too shallow tends to produce trees that underfit, whereas choosing trees that are too deep will overfit the data. A total of 150 trees were used as determined through experimentation. A *p*-value of 0.05 was determined using the curvature test, which is utilized with the RF classifier to determine when to terminate a split in a decision tree. The aforementioned techniques used to

determine RF input parameters is considered to be a standard approach [Biau and Scornet 2015; Loh 2002].

For the results generated from RF classification, not only was the class assignment for each pixel generated, but also the posterior probability, which was treated as the mass function within the framework of D–S theory.

As shown in Figure 5-1, the three classifiers mentioned earlier were first computed using the RF classification method; their results were then combined using the Dempster rule of combination under the D–S framework. The final classification was produced by assigning a given pixel to the class with the maximum mass function. As part of the analysis of the final classification result, comparisons were made to examine changes in land-cover assignment, and to see how the number of high-confidence misclassified pixels changed from the standard classifier (Classifier #1).

The D–S theory is a general framework for reasoning with uncertainty. It allows the user to combine evidence from different sources and arrive at a degree of belief (a mass function) that takes into account all of the available, independent, sources of evidence. Given that the RF classifier works on a majority voting principle, one product it produces is a confidence value for each of the possible outcomes based on the percentage of votes. I utilized this confidence value as a measure of belief in that outcome in the context of the D–S framework. When executing computations with the D–S rule, for each of the classifiers, I treat it as a proposition in the D–S framework. The D–S rule states that:

$$m(A) = \frac{\sum_{B_1 \cap \ldots \cap B_n = A} m_1(B_1) \ldots m_n(B_n)}{1 - K}, K = \sum_{B_1 \cap \ldots \cap B_n = \emptyset} \prod_{i=1}^{n} m_i(B_i) \qquad (5\text{-}7)$$

where $m(A)$ is the mass function of a proposition A after considering n pieces of evidence (in this case, the different classifiers), $m_i(B_i)$ is the mass function in the proposition B_i supported by the i-th piece of evidence, and K is known as the total conflict factor [Dempster 1967]. It should be noted that D-S rule has been criticized by Pearl [1997] in that it is misleading to interpret belief functions as representing either "probabilities of an event" or "the confidence one has in the probabilities assigned to various outcomes" or "degrees of belief in a proposition". Instead, belief functions represent the probability that a given proposition is provable from a set of other propositions, to which probabilities are assigned. While I recognize Pearl's concerns I contend that given the quantitative nature in which the probability for proposition is calculated (voting percentages from RF) these values do represent actual probabilities and is not within the realm of holistic speculation that Pearl is likely referencing. Furthermore, the physical nature of each proposition and the data they represent adds credence as we are not referencing an event per say but rather an object for which the direct measurements correspond to its existence. In that regard D-S rule is applicable here and provides value.

As shown in Figure 5-1, the three classifiers mentioned earlier were first computed using the RF classification method; their results were then combined using the Dempster rule of combination under the D–S framework. The final classification was produced by assigning a given pixel to the class with the maximum mass function. As part of the analysis of the final classification result, comparisons were made to examine changes in land-cover assignment, and to see how the number of high-confidence misclassified pixels changed from the standard classifier (Classifier #1).

5.4. Results

Table 5-4 summarizes the features selected for Classifier #1 resulting from the feature selection method described in Section 5.2.2, as well as those for Classifiers #2 and #3. On the basis of these features, the classification accuracies were 87.5%, 88.3%, and 89.5% for Classifiers #1, #2, and #3, respectively.

Table 5-4. Features used in each classifier, as determined though my analysis in order to maximize classification accuracy. Index refers to the image index from Table 5-3.

Classifier #1		Classifier #2		Classifier #3	
Index	**Name**	**Index**	**Name**	**Index**	**Feature Name**
1	B1 Reflect. #1	180	Senti. DPSVI-#11	65	Senti. VV-#11
2	B2 Reflect. #1	181	Senti. DPSVI-#12	66	Senti. VH-#11
3	B3 Reflect. #1	182	Senti. DPSVI-#13	67	Senti. VV-#12
4	B4 Reflect. #1	183	Senti. DPSVI-#14	68	Senti. VH-#12
7	B7 Reflect. #1	184	VBF-10	69	Senti. VV-#13
15	B3 Reflect. #2	21	NDWI #2	70	Senti. VH-#13
16	B4 Reflect. #2	33	NDWI #3	71	Senti. VV-#14
19	B7 Reflect. #2			72	Senti. VH-#14
20	NDVI #2			73	DEM
184	VBF			92	B3-M-Senti. 2 #6
23	Temp1 #2			108	B3-M-Senti. 2 #10
127	B2 -V- Senti.2 #8			115	B2-V-Senti. 2 #5
				123	B2-V-Senti. 2 #7

When examining Table 5-4 I can note that, for Classifier #1, there were a broad mix of features from different sources and image types. For Classifiers #2 and #3 (broad class separations) a more limited and specific set of features were used. It can be further noticed that the features employed in Classifiers #2 and #3 were mostly excluded from the feature sets automatically selected for Classifier #1. This demonstrated that important features that could be used to distinguish compound cover types would not be employed at all using the traditional feature-level fusion method. Additionally, having different groups of features utilized in these classifiers indicated independence in their classification results, which is important within the framework of the D–S theory.

Classification maps generated by the proposed ensemble method for two selected tested areas dominated by upland and wetland are shown in Figure 5-4 and Figure 5-5, respectively, together with the true-color composite of Sentinel-2 imagery and ground-truth maps. The misclassification pixels by the traditional method and those that were corrected by the proposed ensemble method are highlighted in Figure 5-4D and 4-5D. Observing Figure 5-4A,B and 4-5A,B it can be observed that the classification results generated using the ensemble method were consistent with the ground-truth maps and visual observations. It can be further noted that the misclassification using the traditional method (Classifier #1) was clustered in the upland area (Figure 5-4C) and fen area (Figure 5-5C), both in locations with high spatial variation, and the majority of the misclassification pixels were corrected by the addition of Classifiers #2 and #3 (the ensemble method).

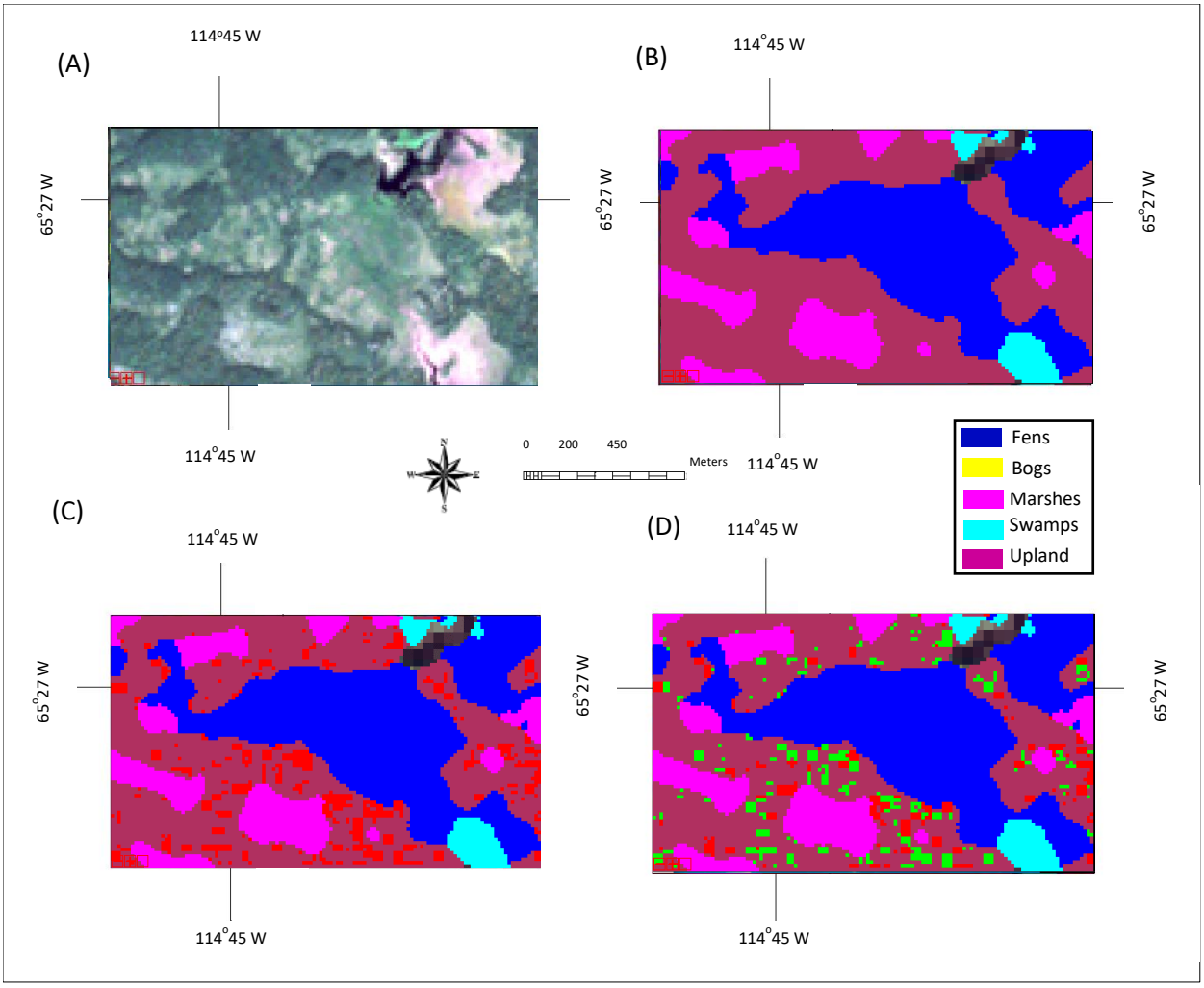

Figure 5-4. Classification result of a test area dominated by upland: (A) true-color composite of a Sentinel-2 image; (B) ground-truth-based classification map; (C) classification map using the proposed method. Misclassified pixels highlighted in red; (D) Classification map using Classifier #1, where the misclassified pixels, which were corrected using the proposed method, are highlighted in green.

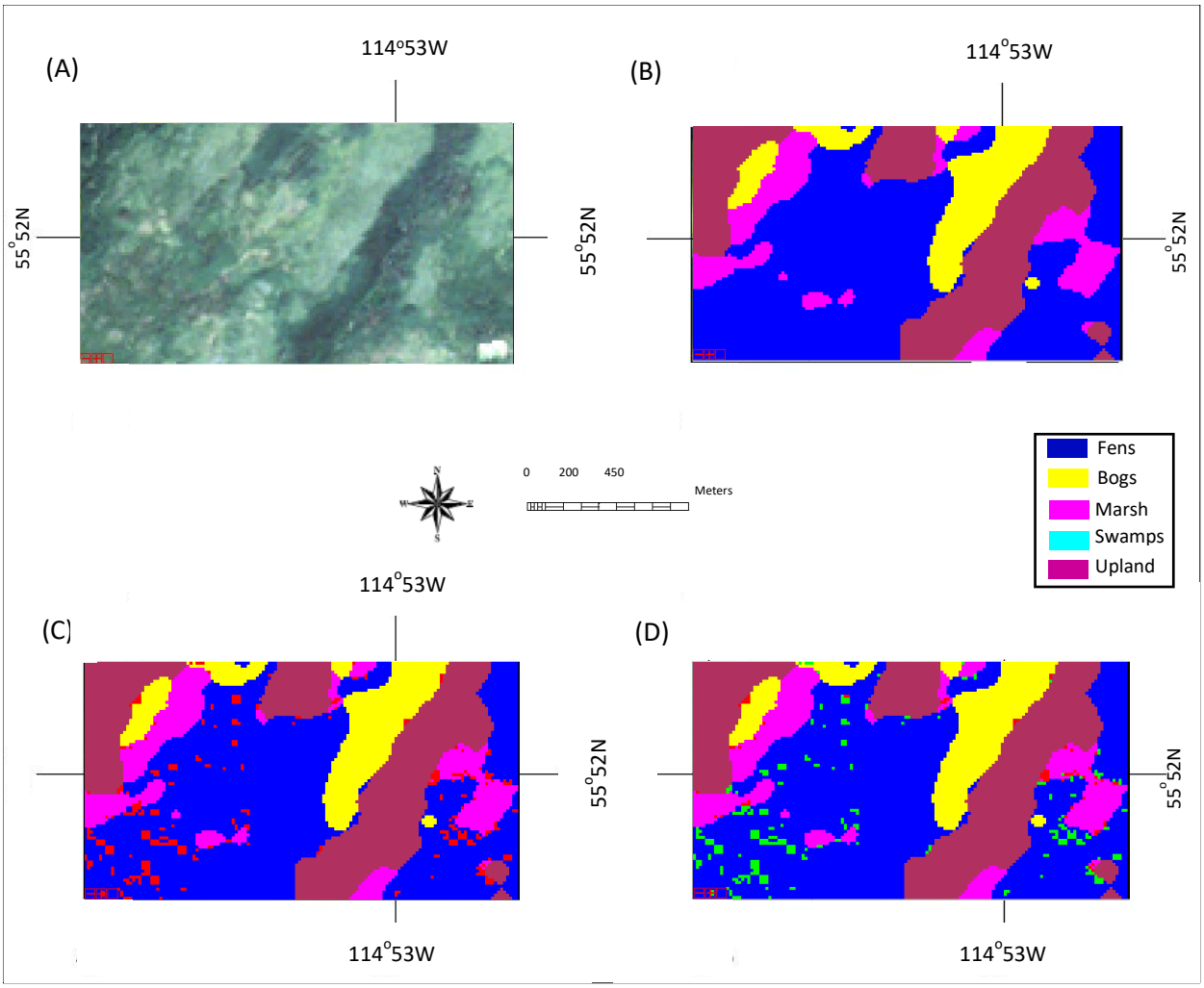

Figure 5-5. Classification result of a test area dominated by wetland: (A) true-color composite of a Sentinel-2 image; (B) ground-truth-based classification map; (C) classification map using Classifier #1 where misclassified pixels are highlighted in red; (D) classification map using Classifier #1, where the misclassified pixels, which were corrected using the proposed method, are highlighted in green.

In addition to the visual assessment of the classification results, quantitative analysis was carried out and the confusion matrices for the traditional method and the Ensemble Classifier are shown in Figure 5-4 and Figure 5-5, respectively. In this section, I will focus on observations from these results, while detailed discussion will be provided in the discussion section.

Table 5-5. Confusion matrix of the traditional method (Classifier #1). Rows represent the classification, while columns represent the reference.

	Fen	Bog	Marsh	Swamp	Upland	Producer Accuracy	User Accuracy
Fen	134,845	2212	1753	13,691	3601	0.8638	0.7354
Bog	948	13,186	0	277	68	0.9106	0.7505
Marsh	1502	1	21,182	468	263	0.9045	0.7595
Swamp	721	40	117	8100	173	0.8851	0.1422
Upland	45,327	2130	4836	34,388	549,760	0.8638	0.9925
					Overall accuracy	0.875	

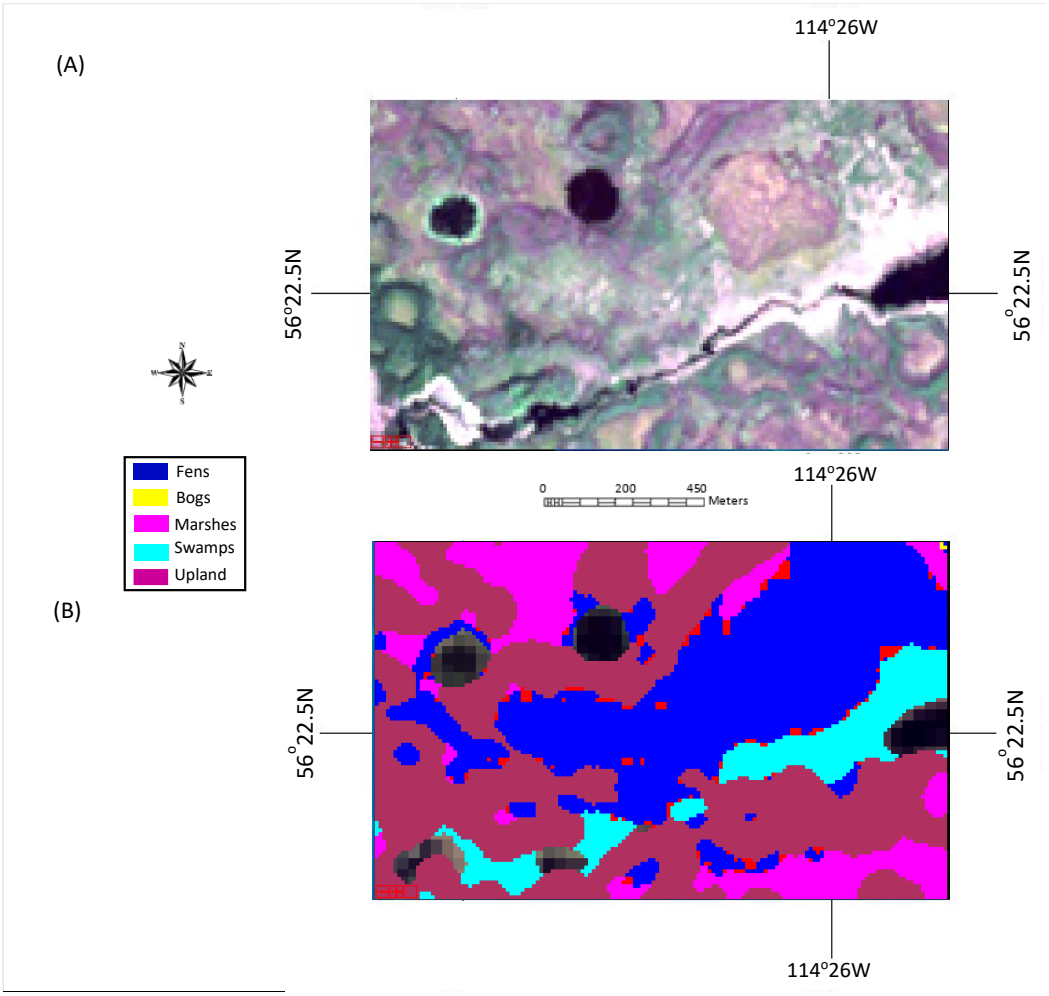

Figure 5-6. (A) Test area containing the misclassified pixels with high confidence: (A) true-color Sentinel-2 image; (B) classification map using the proposed method with the misclassified pixels highlighted in red.

Table 5-6. Confusion matrix of the proposed ensemble classifier. Rows represent the classification, while columns represent the reference.

	Fen	Bog	Marsh	Swamp	Upland	Producer Accuracy	User Accuracy
Fen	145,390	2132	1674	5508	1398	0.9315	0.8267
Bog	668	13,628	0	175	8	0.9412	0.8234
Marsh	822	0	22,200	290	104	0.9481	0.8142
Swamp	534	24	83	8462	48	0.9247	0.3203
Upland	28,445	766	3310	11,986	591,934	0.9301	0.9974
					Overall accuracy	0.935	

In general, the ensemble classifier incorporating the three classifiers together based on the D–S theory resulted in an increase in the classification accuracy from 87.5% (the traditional method, Classifier #1) to 93.5%. Upon closer examination of the results using the traditional method (Table 5-5), it can be noted that the producer accuracy was high and fairly uniform across all land covers. However, the user accuracy was lower; in particular, it was the lowest at 14% for swamp. When examining the results using the proposed method (Table 5-6), it can be noted that the user accuracy for swamps was increased by ~0.18 to 0.32. In addition, it can be observed that the upland land cover was misclassified the most in terms of the number of raw pixels and as a percentage of pixels misclassified. This might be due to the broad nature of the upland class. As a point of note, when compared to the classification result from the previous chapter from the study area in northern Ontario, the results from Table 5-5 and Table 5-6 are more uniform in their performance. I speculate that this is due to the broader nature of the classes when compared to the more narrow classes of the previous study.

Checking the misclassified pixels, it can be noted that, for some, the support (mass function) for the "wrong" cover type was very strong (over 0.85), indicating a high confidence for the class

assignment. However, it was observed that there was a reduction in the number of high-confidence misclassified pixels from 26222 to 23,588—a reduction of ~10% using the proposed method (Table 5-7). These results show that the addition of two classifiers with compound classes through the ensemble classifier provided value in increasing the accuracy and decreasing the number of the incorrectly classified pixels with high confidence.

Table 5-7. Number of the misclassified pixels with high confidence and their land cover assignments for the traditional and proposed methods.

	Fen	Bog	Marsh	Swamp	Upland
High conf. misclassified Pixels—Classifier #1	9714	604	321	414	20,167
High conf. misclassified Pixels—the proposed method	6608	505	265	348	15,370

To further examine the improvement in individual land-cover classification provided by the ensemble classifier, tables to show changes in the pixel assignments for each cover type were generated (Table 5-8 and Table 5-9).

It can be noted from these tables that the majority of misclassified pixels, across all classes, which were reclassified by the ensemble classifier, were moved to the upland class. Of additional note, a large number of pixels originally assigned to swamps were moved to other classes, including the upland class. This movement in the assignment of pixels would also explain the large increase in user accuracy for swamps by the proposed ensemble classifier. Among these misclassified pixels with their assignment changes, some of them were classified correctly using the proposed method, while some were still misclassified, and the correct class had the second strongest support from the evidence. However, for some in the latter group, the classification uncertainty was high. That is, for these pixels, the largest mass function was not significantly

different from the second largest one (difference between 0.05–0.10), leading to large uncertainty class assignment. These pixels were also summarized, as shown in Table 5-9.

Table 5-8. Matrix showing the assignments using the proposed ensemble classifier in comparison with Classifier #1 (the traditional method) for all pixels with changes in class assignment. Columns are the land covers that a misclassified pixel was first assigned to from Classifier #1. The rows correspond to the land cover that a pixel was assigned to by the ensemble classifier.

		Final Land Cover				
		Fen	**Bog**	**Marsh**	**Swamp**	**Upland**
Initial Land cover	Fen	0	9	11	60	4918
	Bog	0	0	0	1	204
	Marsh	0	0	0	2	336
	Swamp	15,394	868	614	0	18,180
	Upland	1206	69	134	43	0

Table 5-9. Matrix showing pixel assignments by the proposed ensemble classifier in comparison with Classifier #1 (the traditional method) for a subset of the pixels shown in Table 5-7. For the pixels shown here, the correct class had the second largest support from the evidence, but the largest and second largest mass functions were similar. Columns are the land covers that a misclassified pixel was first assigned to from Classifier #1. The rows correspond to the land cover that a pixel was assigned to by the ensemble classifier.

		Final Land Cover				
		Fen	**Bog**	**Marsh**	**Swamp**	**Upland**
Initial Land cover	Fen	0	1	2	28	4903
	Bog	0	0	0	0	204
	Marsh	0	0	0	2	333
	Swamp	7125	114	182	0	17,482
	Upland	708	22	34	10	0

5.5. Discussion

5.5.1. Discussion – Feature Significance and Selection

In this study, the selection of features for Classifier #1 followed a standard data-driven machine learning methodology, which is commonly used. The features for Classifiers #2 and #3 were manually selected, following a holistic approach, similar to that presented Chapter 4 and a previous paper I published [Judah and Hu 2019]. From a holistic standpoint, I selected families of features which, by design, were best suited for class separation sought for each classifier, while ensuring the independence of these classifiers required by the D–S theory. The design of Classifiers #2 and #3 in terms of class propositions was to fully utilize the available datasets. It was observed that, for Classifier #1, most features selected were from optical imagery. For instance, backscatter coefficients and related indices from SAR imagery and water indices from optical imagery were known and, thus, identified for Classifier #2 (separating wetlands from upland covers), while backscatter coefficients from SAR imagery and textural features from optical imagery were identified for Classifier #3 (separating structured from less structured land covers). Through feature analysis and experimentation, I was able to determine a set of features which maximized the classification accuracy for those classifiers. As an interesting note, in Chapter 4 [Judah and Hu 2021], I reported that there were many instances where a set of holistically determined features actually produced more accurate classification results when compared to sets of features selected through quantitative analysis. In this study, I also observed the same phenomenon when determining feature inputs for Classifiers #2 and #3. These results also confirmed my belief (briefly mentioned in Section 1) that simple feature-level fusion for classification using multi-source remotely sensed data might underutilize some features. In that

regard this motivates the need to more formally explore this phenomenon and to create a quantitative structure which can effectively and consistently integrate knowledge-based and data-driven feature information to select features.

From an imagery standpoint, it was noted that there was no clear correlation between the collection dates of the imagery and their significance. Intuitively, imagery closer to 2016 (the collection date of the aerial imagery used to create the ABMI plots) should be of greater significance but this was generally not the case. Landsat imagery from 2015 and 2016 was more significant when compared to the collection from 2020, while, for both Sentinel-1 and -2, there was no clear correlation. This might indicate that features of these cover types exhibited in Sentinel-2 SAR imagery were not highly dynamic. It was also suspected that factors such as atmospheric attenuation and inter-year variations in water levels were due in part in driving these differences.

By exploring images, it was also noted that classification accuracies from inputs strictly drawn from Landsat-8 images produced accuracies which were higher when compared to classification experiments where inputs were strictly drawn from Sentinel-2 images. This was counter-intuitive. It would be expected that inputs with higher resolution would result in higher classification accuracies. However, upon further analysis, it was noted that the land covers considered in this study were broader when compared to other land cover maps which have more narrow class definitions [The National Wetlands Working Group 1997]. In previous studies, by virtue of data availability, land covers such as fens and bogs were parsed further into treed and non-treed versions of those land covers. For those datasets, the higher-resolution imagery might have provided the expected accuracy increases; however, with this ABMI dataset with broader classes, it is suspected that the high spatial resolution of the Sentinel-2 images might have

introduced more variability among cover types, which made the classification it more difficult. Lastly, during these experiments, it was noted that the classification accuracy when using only individual datasets was some 5–8% lower when compared to classification accuracies from multi-source remotely sensed data, which is consistent with the literature.

5.5.2. On Misclassified Pixels

The core of this study was the development of the ensemble classifier in an effort to increase classification accuracies while also reducing the number of the incorrectly classified pixels with high confidence. The prevalence of misclassified pixels of high confidence (>85% certainty in assignment) and misclassified pixels which had the correct land cover class as the second highest ranked land cover was noted in this study with Classifier #1 (the traditional method). As shown by the results, these issues were overcome by adding two classifiers in the proposed ensemble classifier to a certain extent. Examining the mis-classified pixels using the proposed method, it was noticed that the three classifiers were not always in agreement with one another, as shown in Table 5-10. It was further noticed that the misclassified pixels with high confidence were located at the transition zones between cover types, as shown in Figure 5-6. This intuitively and physically makes sense since the transition from one wetland cover type to another is fuzzy in nature [Dronova 2015], and often implies that these pixels are mixed pixels of multiple landcover types. This would be especially pronounced with the lower resolution Landsat-8 imagery.

Table 5-10. Land-cover breakdown of the misclassified pixels where two classifiers disagreed or all disagreed.

	Fen	Bog	Marsh	Swamp	Upland
Two Disagree	6817	150	235	523	27,402
All Disagree	752	8	76	1	7330

In addition to the transition zones where these classifiers tended to conflict with each other, pixels with disagreement among classifiers were also within in the areas with high variability according to a visual examination, as shown in Figure 5-7. This would drive variations in features, which in turn could then contribute to the variability in the outcomes of the different classification propositions. Additional information may be needed to further solve this confliction.

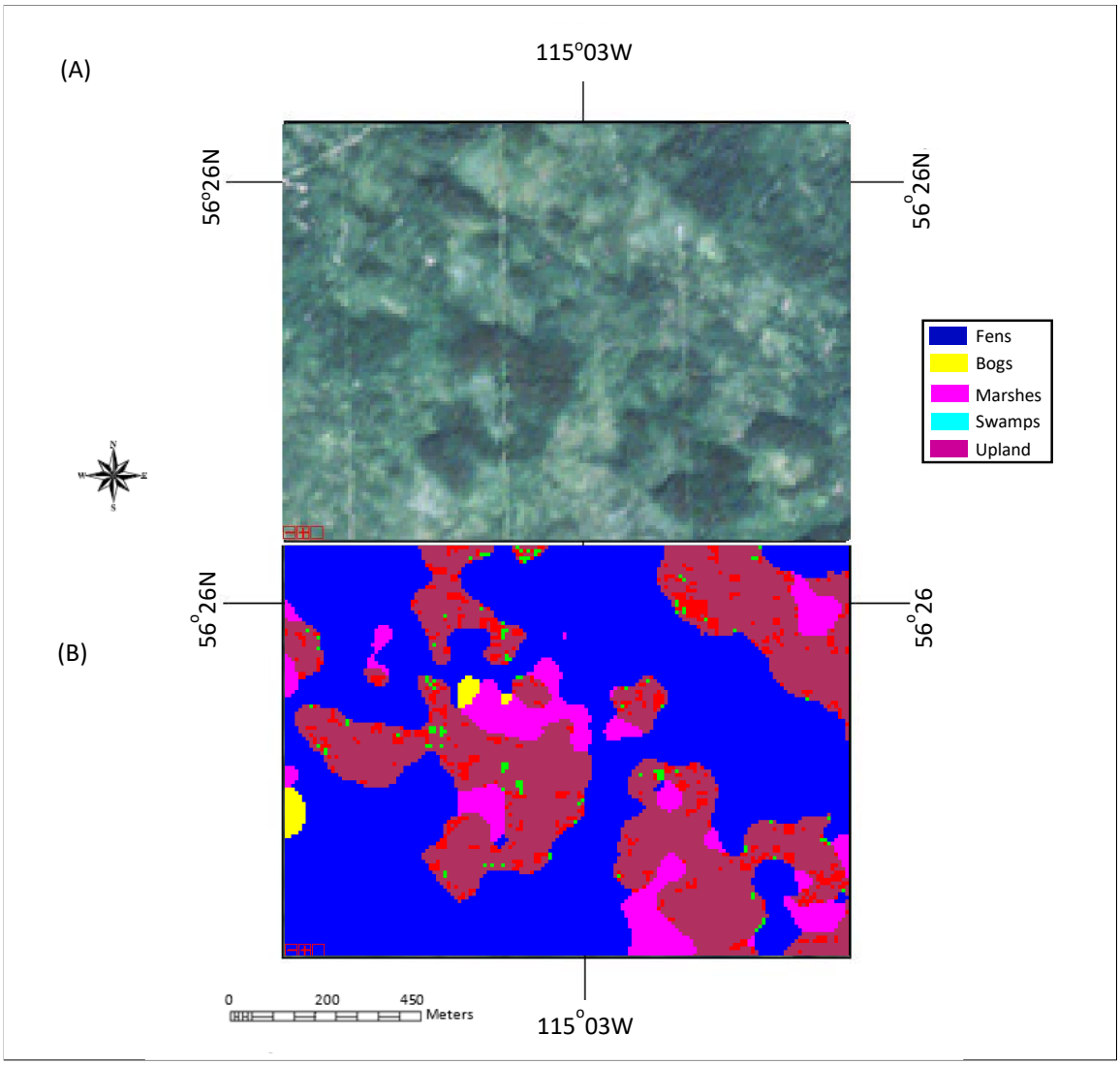

Figure 5-7. (A) True-color image of test area from Sentinel-2 imagery; (B) classification map showing pixels where two classifiers disagree or all classifiers disagree, as highlighted in red and green, respectively.

The misclassification involving upland may also be due to the fact that the upland class was very broad and encompassed a great deal of different land-cover types, which led to large variations in the selected features for it. As an example, Classifier #3 was used to classify structured and nonstructured cover type. Upland was included in the structured class, considering the domination of trees and shrubs in this class. However, there were also nonstructured cover types in this class. To mitigate this, I attempted to split the upland class to two subclasses during

the decision-level fusion process according to the D–S theory. However, there was no real improvement in the results (not shown). The best strategy was to separate upland to different categories, which was not attempted due to the lack of training samples for detailed upland cover types.

5.5.3. On the Proposed Ensemble Classifier

The overall classification accuracy of the proposed method was 0.93. When compared to other studies, it was noted that this accuracy was greater or comparable with those obtained for land-cover classification using multi-source remotely sensed data [Jai et al 2021; Ghosh et al 2014; Liu and Gao 2022; Useya and Chem 2018]. In addition, the proposed method was less complex than some of these studies. It should be stressed that I could not find classification studies of the study area which used a decision- or feature-level fusion framework for direct comparison in the literature; furthermore, all of the comparable studies I found used different datasets or combinations thereof, for both the land-cover maps and the remotely sensed imagery used. However, it can be noted that the ABMI conducted its own classification studies of its own dataset using Landsat-5 and Landsat-8 imagery, with an RF classifier; the classification accuracies were around 0.8–0.85 [ABMI 2021].

A direct comparison was carried out in this study with the traditional classification method based on feature-level fusion using multi-source remotely sensed data (Classifier #1). Results were presented and discussed in the previous sections. The improvement of the proposed method over the traditional method relied on its effective utilization of available datasets and features. Features that can be used to separate certain cover types might be excluded by considering all

cover types together, such as the features derived from SAR imagery and DEM. The inclusion of these features otherwise excluded in Classifiers #2 and #3 led to an increase in in user accuracy of the swamp class by ~18%. It was also noted that, in Classifier #1, the impact of the SAR imagery was lower when compared to it being utilized in a classifier focused on broader class separations. When combined in the ensemble classifier, the value of this imagery was better utilized.

In total, the proposed ensemble classifier provided a framework to effectively utilize the best available data in order to support wetland classification. In this study, while an RF classifier was employed, other classifiers could be utilized. I experimented with support vector machine (SVM) and naïve Bayes classifiers, where I found that the overall accuracies were generally lower (by ~5–8%), but the computation times were greatly reduced, compared to similar RF tests, in some cases by over 80%. The additional classification accuracy gained by using RF was obtained at a considerable computational cost.

The proposed ensemble classifier combined the strengths of various types of remotely sensed data in the differentiation of wetland cover types. This same principle could be applied to the classification of other cover types. The three classifiers were designed parallelly and independently, even though the same classifier (RF) was used. The idea of designing two classifiers to classify broader cover types was inspired by the hierarchical classification methods including my own work. As mentioned earlier, with hierarchical classification methods, the errors/uncertainties in the higher hierarchies are often not considered in the lower ones; thus, error propagation is the biggest problem. With the proposed method in this study, the uncertainty associated with these classifiers was explicitly considered under the framework of the D–S theory and, thus, solved the error propagation problem in hierarchical classification. In the same

vein, this study expands the literature on the utilization of the D–S theory. Even though the D–S theory is powerful conceptually, its application is not trivial, especially in the determination of the mass functions, including the selection of propositions of non-zero mass functions. As mentioned in Section 1, in most studies based on the D–S theory, identical sets of classes were often employed for different classifiers [Bo et al 2016; Jai et al 2021; Guo 2019; Feng 2021]. Different prepositions were considered for the three classifiers in this study, and they were selected according to prior knowledge of the wetland cover types and remotely sensed data. Not only did this result in higher classification accuracies but it also provides a framework for future work where I can more easily explore subclasses, class overlaps, and unknown classes. In that regard using other classification methods or strategies could be useful in expanding the capabilities of the ensemble method. To explore subclasses, deep learning methodologies could be helpful in that regard. This could also be accomplished with expanding the dataset (to help buttress the deep learning algorithm) and incorporating an unsupervised or semi-supervised approach to identify those subclasses. This is an active area of research with deep learning methodologies [Ezugwu et al. 2022]. I also previously explored the use of semi-supervised methods in the form of mean shift to identify clusters with the intent of parsing subclasses but found the computational loading to be too great for the resources I had. By moving to a deep learning framework with a supercomputer, this could be mitigated and would be a worthwhile endeavor.

The usage of prior knowledge in the designing of Classifiers #2 and #3 was also one of the disadvantages of the proposed approach. In this study, the categories and features were selected manually. This may not be practical for studies dealing with a large number of cover types.

Ideally, a knowledge-based automatic approach would be preferable. This will be pursued in future work.

Lastly, the ensemble classifier in its current form has not been successful in dealing with cover types with great diversity such as the upland class. Even though it is ideal to separate such cover types into different several classes during the training process, it might not be realistic due to the difficulty in the selection of training samples. Initial experiments where I tried to separate two subclasses during the decision-level fusion process did not show accuracy increases or effective or consistent class separation. Future versions of this classifier will have to address this.

5.5.4. Conclusions and Future Work

An ensemble classification methodology combining three classifiers based on the D–S theory was developed and tested on a study area in Northern Alberta. Classifier #1 was a traditional feature-level fusion method for classification using multi-source remotely sensed data where all land-cover classes were classified together. The other two classifiers were focused on compound cover types. With Classifier #2, wetland cover types (fen, bog, marsh, and swamp) and dry land covers (upland) were considered, whereas, with Classifier #3, the focus was on the separation of less structured land covers (fen, bog, and marsh) and more structured ones (swamp and upland). Features used for classification were determined using the analysis of RF feature significance for Classifier #1 and through a more holistic approach for Classifiers #2 and #3. Use of a holistic approach for feature selection was not traditional; however, on the basis of prior knowledge and experimentation, I was able to select a set of features for Classifiers #2 and #3 which produced high accuracy when compared to a strict feature significance analysis approach. This also mimicked the results observed in past studies [Judah and Hu 2021]. Once each

classifier was computed, those results were combined using the Dempster's combination rule. Results showed that the proposed ensemble classifier increased the classification accuracy from 0.88 to 0.93, compared with the traditional classification method (Classifier #1). Additionally, it was noted that there was a reduction of ~10% in the number of the misclassified pixels with high confidence, which provides additional assurance in the quality of the classification results, something which is generally not explored in this style of research.

The proposed approach provided a framework to intelligently utilize available remotely sensed data for wetland classification, which could be employed for other cover type classification. Incorporating data-driven machine learning and knowledge-based holistic methods, different propositions were designed; thus, different features were selected for these three classifiers to maximize their discriminant powers in the classification of these wetland cover types (individually or in combination). As detailed in the discussion, this made this framework unique compared with most studies based on D–S theory reported in the literature. In addition, compared with hierarchical classification methods, the proposed ensemble classifier's advantages were enhanced by selecting different features to classify different classes, while its weaknesses were addressed by explicitly taking into account the uncertainties of different classifiers.

Even though the holistic knowledge-based method was successful in the design of Classifiers #2 and #3, prior knowledge could be utilized in a more explicit and automatic fashion, enabling the proposed method to be employed as a general framework in wider applications. This will be endeavored moving forward. With the current approach, advanced features derived from the available datasets will be further explored, and more classifiers will be added. Additional testing will be also carried out by expanding the study area to the remaining

parts of the ABMI wetland inventory. Other data sources, such as RADARSAT-2 and LiDAR images, will be considered.

Chapter 6 - Conclusions and Future Considerations

6.1. Conclusions

Development and widespread access to multispectral remotely sensed imagery, and other ancillary data provide researchers with a means to quickly and effectively characterize and monitor landcovers. However, many challenges remain with regards to the methodologies used to achieve this, in particular, the characterization, identification, and selection of the most impactful and effective data features in support of wetland classification. There is a need to determine how best to utilize the available remotely sensed data in order to best exploit that data's unique characteristics in order to maximize classification accuracy and reduce errors. With the newer data sources and classification techniques becoming available, this last point will be an ongoing need for the field. Hence, this PhD research aimed to increase overall classification accuracy and to reduce the number of high confidence misclassified pixels through the development of advance classification techniques. This was achieved through a holistic approach executed in three phases, which built upon one another, leading to my ultimate goal. These phases were: (1). the characterization, identification, and selection of the most impactful and effective data features, and ancillary data in support of wetland classification, (2). The development of hierarchically based classifiers in order to best exploit the discriminant power of the aforementioned data features broad class separations observed in the first phase, in order to maximize classification accuracy, (3). The further refinement in the execution of wetland classification, using three different RF based classifiers (a standard classification approach, and two classifiers focused on broad class separations), using D-S theory, in order to best utilize the

available data to produce more accurate and efficient classification maps compared to a standard classifier and to reduce the propagation of errors and uncertainty within a developed classification framework. Through these studies I was able to produce knowledge and techniques which could be utilized by the scientific community to more efficiently and effectively produce improved wetland classification maps using remotely sensed imagery and ancillary data. Furthermore, my hope was that this work would inspire parallel applications and studies in other fields which use similar remotely sensed data, in those applications, and in the larger data science community where many of these tools and techniques are still under active development.

The major conclusion of this work was that a RF classification methodology combining multiple classifiers through D-S analysis could increase the peak classification accuracy when compared to a standard classification approach. Furthermore, the RF classification methodology consistently outperformed all other classification methodologies when it came to the classification of wetland, which is likely due to the algorithms ability to explore nuances within a complex data set. Given the board and complex spectral nature of wetlands this would be an appropriate match for its classification needs and I would recommend the use of RF classifier for the classification of wetlands. Additionally, it was noted that there was a reduction in the in the number of high confidence misclassified pixels, and the number of pixels where the second highest ranked land cover was the correct land cover assignment. It was also concluded that this methodology showed promise and has motivated me to apply the developed techniques to other study areas and datasets to further explore the functionality of the developed methodologies. The success of this research is due in part from the development and characterization of data feature analysis techniques and the development of advanced image classification methodologies of wetland landcovers. An overall take away from this work, in my estimation, the best strategy in

order to classify and characterize wetlands using remotely sensed imagery is to approach it using the methodology developed in the third paper. In broad terms, developing multiple classifiers, leveraging specific and broad class separations where each classifier is designed to best exploit the most impactful data features form that classifier, fused with a D-S framework is most effective at classifying landcover and in particular wetlands, based on the testing from the study area I used in Northern Alberta. These conclusions will help contribute to the larger scientific community by providing knowledge and techniques which can help produce higher classification accuracies, and reductions in the number of misclassified pixels which in turn helps to increase confidence in those results.

This dissertation is of scientific value due to the aforementioned conclusions and contributions which have aided in the overall advancement and characterization and classification of wetlands, and also helped to advance the still developing field of data science. The novelty and other major contributions can be summarized as follows:

(1) The use of surface temperature (an untraditional feature choice) could aid in the classification process and the identification of wetlands, when combined with NDVI calculated from optical imagery obtained in the spring months, radar backscatter coefficients, and ancillary data such as surface slope, computed through either an RF or SVM classifier (Section 3.4).

The use of surface temperature, despite its low resolution, could be used to better classify wetlands from the study area in Northern Ontario, in particular if the temperature measurement was from an abnormally warm spring season. Additionally, the addition of RADARSAT-2, and or Sentinel-1 imagery to Landsat-5/8 imagery was shown to improve overall classification accuracies. It was also found that the data acquired in the fall season,

if used solely as the classification input, consistently produced the poorest classification results.

(2) Preselection of features using Log-normal or RF variable importance analysis was an effective way of identifying low quality features and to a lesser extent features which were of higher quality. However, it was noted that a holistic approach to feature selection, aided through feature analysis, produced the most accurate classification tests (Section 3.5).

It was noted that preselection analysis of features was effective across all classification methodologies in identifying low quality features. When it came to identifying the highest quality features it was not as consistent. Furthermore, there was some inconsistency in data feature rankings determined by the log-normal distance method vs the RF predictor importance method. I suspect that these performance differences are driven by fundamental differences in how the log-normal distance is calculated (gross statistical measure, with no provisions for identifying multi-modal features) compared to the RF predictor importance value, which is iterative and explores subsets within a given dataset. Given these differences in feature analysis and the differences in how each classification technique analyzes a given dataset, is the likely cause of this discrepancy. It was also found that this analysis aided K-NN the most in identifying features, with its best performing tests being mostly represented by tests determined through this analysis (17 of its top 20 tests were determined by feature selection).

As a general trend I contend that applying this methodology to RF, SVM, and Naïve-Bayes especially provided value in determining lower quality features (features common in the bottom performing 20 tests), which could then be excluded from analysis to both speed up analysis time and ensure that results are more likely to be of a higher quality and accuracy.

For the other classification methodologies, results generally showed that the features determined by this analysis produced high accuracies, but they did not produce the best results. Those results were produced by the input features determined through a holistic approach, with the best performing tests (RF test #77 from Phase 2) produced an overall accuracy of 85.71%. The exact reason why holistically determined tests have performed so much better than quantitatively determined tests is unknown but should be further explored in future work. However, as a general trend I contend that applying this methodology to RF, SVM, and Naïve-Bayes especially provided value in determining lower quality features (features common in the bottom performing 20 tests), which could then be excluded from analysis to both speed up analysis time and ensure that results are more likely to be of a higher quality and accuracy.

(3) Development of methodologies to classify wetlands from remotely sensed multispectral data using advanced classification algorithms through two hierarchical approaches, which leveraged broad class separations. [Section 4.2.4]

The two hierarchical classification approaches differed in how they initially separated wetlands by either creating splits between treed vs non-treed wetlands (Case-1) or separating wetlands into the classes of Fens, Bogs, and Swamps (Case-2). For both of these approaches classification accuracy increases were observed when compared to a traditional classifier. Case-2 produced the most accurate classification results where I note an overall performance gain for the top 10% of results for both the Naïve Bayes and the RF classifier when compared to a traditional classifier. I suspect that this performance difference is driven from breaking down wetlands via species rather than structural differences (treed vs non-treed) which is better at preserving the quality impact of the selected features. For the

most accurate classification result (91.94%), for all tests, a result from RF, executed through the Case-2 approach, outperformed the most accurate result from the traditional classification method by some 11%. It is also worth noting that Naïve Bayes classification results, using the Case - 2 approach, showed a great improvement to its accuracy, in one case ~12% increase when compared to the highest accuracy produced through the traditional approach.

A notable result from the hierarchical classification results was that bottom ranked tests were of a higher accuracy when compared to the traditional classification methods for the same sets of features. In some cases I noted a 2-6% classification accuracy increase. A notable result from the hierarchical classification results was that bottom ranked tests were of a higher accuracy when compared to the traditional classification methods for the same sets of features.

(4) Development of a methodology to classify wetlands from remotely sensed data using multiple RF derived classifiers, fused through a D-S Theory (Section 5.3.3).

Three different classifiers were utilized and combined with D-S analysis. The first classifier was a standard classification approach where all land cover classes were classified together (Classifier #1). The other two classifiers focused on broad class splits to separate land covers. Classifier #2 focused on the classification of wet (fen, bog, marsh, swamp) to dry land covers (upland), and Classifier #3 focused on non-treed and less structured wetlands (fen, bog, marsh) to more structured land covers (swamps and upland). Feature inputs were determined using RF feature significance analysis for Classifier #1 and a holistic approach to select an optimized set of features for Classifier #2 and #3. Once each Classifier was computed those results were combined using D-S theory. Results show

that the combination of the three Classifiers D-S could increase the peak classification accuracy of the standard classification approach by ~5%. Additionally, it was noted that there was a reduction of ~10% in the number of high confidence misclassified pixels. These results imply that applying D-S analysis in the context of classifying wetland land covers for the study area is working as intended.

6.2. Future Considerations

The results from the developed methodologies demonstrated promise in terms of improving classification accuracies and reducing high confidence misclassified data of wetland maps using remotely sensed imagery and ancillary data but further improvements and areas of investigation remain. I present the following future considerations.

(1) Further examinations of feature significance results to explore and assign more quantifiable physical explanations to why certain features were more significant than others. Furthermore, optimum classification conditions for wetlands, and the ultimate limits that this style of feature analysis can produce should be explored. This is a challenging proposition but one that is worthwhile. This will not only provide a framework for wetland classification which can be used as a product but will also provide a level of expectation when it comes to the ultimate accuracy that this style of analysis can produce. This in turn will aid in determining the next steps required to achieve the next level of accuracy or detail.

(2) When considering how my hierarchically developed classification work can be expanded upon, I note that this research would benefit from the addition of more images from other years, and from other image sources - as a general principle this is true for all classification

methodologies. To further develop this work, additional LANDSAT-8 and Sentinel-2 data, and higher accuracy radar-based images would be beneficial. When considering these results in the context of previously published work, others have found that having more continuous and diverse sets of multispectral imagery can lead to better classification results at finer class resolutions [Jiao et al 2019; Ramsey and Laine 2017] as well, finer and more band diverse Radar imagery, especially when used in large amounts can further increase accuracy in finer class separations by providing more nuanced structural information of the land covers by providing better differentiation of plant type and species. [Mahdianpari et al 2017; Mohammadimanesh et al 2018, 2018-2].

(3) Performing similar hierarchical classification analysis but with an object orientated approach would also be another natural next step. For this study I decided to use a pixel-based approach in order to capture detailed wetland class features (especially since wetlands generally are of a small footprint). However, by examining land covers via an object orientated approach this would allow me to utilize other features more effectively such as texture and shape. Furthermore, an object orientated approach could arguably be more suited to the incorporation of satellite imagery by generalizing segmented areas which could in turn reduce noise, and draw out features of those images.

(4) For some of the peculiar characteristics of the hierarchical results, further development and refinement of those methods via additional feature significance analysis, tailored to each approach in order to create more accurate and consistent results should be investigated. Also, further investigations should be done to better understand why some holistically determined tests outperformed tests created through feature significance analysis. By better understanding these differences I hope to create a more complete way of pre-selecting

features in order to create efficient and accurate classification results. Finally, when considering how this work can be adapted to other study areas, I note that northern hemisphere temperate forests are all very similar in structure and vegetation distribution. The work done with this research should be sufficiently general with only minor local considerations from the study site. The methodologies and results produced in this study should be able to be applied without much difficulty to other northern hemisphere temperate study areas, in Canada or other parts of the world. Applying this work to tropical environments would likely be less compatible given the difference in vegetation density, vegetation types, and the lack of large seasonal variations with that vegetation. However, if given the appropriate datasets, the study methodology used here should be able to produce similar variable analysis and classification results, which would be an interesting contrast to this work.

(5) From the work using D-S analysis expanding out the dataset to the remaining parts of the ABMI wetland inventory in order to buttress the training sets would be a natural progression, as well as incorporating other sensors such as RADARSAT-2 and Lidar images to provide additional observations.

(6) Investigating further into the detection and characterization of subclasses within the ABMI wetland dataset. This study has already shown indications of subclass structures which can be detected but further investigation and characterization needs to be done. This can be accomplished through further ground or aerial imagery analysis of test areas to identify subclasses and to then see how that compares to classification results where subclasses were detected.

(7) Applying this developed methodology to other study areas with narrower and more classes would be a natural progression. Given the success with utilizing D-S with the same or similar image types, but with different observation times, would be a worthwhile endeavor in order to determine the ultimate limits of this style of analysis and to see how the success of this work can be replicated in different areas and if possible, with different imagery sources.

References

Adam, E., O. Mutanga, and D. Rugege. 2010. “Multispectral and Hyperspectral Remote Sensing for Identification and Mapping of Wetland Vegetation: A Review.” *Wetlands Ecology and Management* 18: 281–296. doi:10.1007/s11273-009-9169-z.

Alberta Biodiversity Monitoring Institute. 2021, “ABMI Wetland Inventory – Technical Documentation ABMI Geospatial Center, March, 2021” *University of Alberta Edmonton, AB, Canada*

Amani, M., S. Mahdavi, M. Afshar, B. Brisco, W. M. Huang, S. M. J. Mirzadeh, L. White, S. Banks, J. Montgomery, and C. Hopkinson. 2019. "Canadian Wetland Inventory using Google Earth Engine: The First Map and Preliminary Results." *Remote Sensing* 11 (7):20. doi: 10.3390/rs11070842.

Amarsaikhan, D., and T. Douglas. 2004. "Data fusion and multisource image classification." *International journal of remote sensing* 25 (17):3529-39. doi: 10.1080/0143116031000115111.

Asner, G. P. 1998. “Biophysical and Biochemical Sources of Variability in Canopy Reflectance.” Remote Sensing of Environment 64: 234–253. doi:10.1016/S0034-4257(98)00014-5.

Badgley, Grayson, Christopher B. Field, and Joseph A. Berry. 2017. "Canopy near-infrared reflectance and terrestrial photosynthesis." *Science advances* 3 (3):e1602244-e. doi: 10.1126/sciadv.1602244.

Berhane, T. M., C. R. Lane, Q. S. Wu, B. C. Autrey, O. A. Anenkhonov, V. V. Chepinoga, and H. X. Liu. 2018. "Decision-Tree, Rule-Based, and Random Forest Classification of High-Resolution Multispectral Imagery for Wetland Mapping and Inventory." *Remote Sensing* 10 (4):26. doi: 10.3390/rs10040580.

Biau, G., and E. Scornet. 2016. "A random forest guided tour." *Test* 25 (2):197-227. doi: 10.1007/s11749-016-0481-7.

Brisco, B.. 2015. “Remote Sensing of Wetlands: Applications and Advances”. Natural Resources Canada.

Blaustein, A. R., D. B. Wake, and W. P. Sousa. 1994. "Amphibian Declines - Judging Stability, Persistence, and Susceptibility of Populations to Local and Global Extinctions." *Conservation Biology* 8 (1):60-71. doi: 10.1046/j.1523-1739.1994.08010060.x.

Bo, Chunjuan, Huchuan Lu, and Dong Wang. 2016a. "Hyperspectral Image Classification via

JCR and SVM Models With Decision Fusion." *IEEE geoscience and remote sensing letters* 13 (2):177-81. doi: 10.1109/LGRS.2015.2504449.

———. 2016b. "Hyperspectral Image Classification via JCR and SVM Models With Decision Fusion." *IEEE geoscience and remote sensing letters* 13 (2):177-81. doi: 10.1109/LGRS.2015.2504449.

Bourgeau-Chavez, L., S. Endres, M. Battaglia, M. E. Miller, E. Banda, Z. Laubach, P. Higman, P. Chow-Fraser, and J. Marcaccio. 2015. "Development of a Bi-National Great Lakes Coastal Wetland and Land Use Map Using Three-Season PALSAR and Landsat Imagery." *Remote Sensing* 7 (7):8655-82. doi: 10.3390/rs70708655.

Breiman, L. 2001. "Random forests." *Machine Learning* 45 (1):5-32. doi: 10.1023/a:1010933404324.

Bui, Dang Hung, and László Mucsi. 2022. "Comparison of Layer-stacking and Dempster-Shafer Theory-based Methods Using Sentinel-1 and Sentinel-2 Data Fusion in Urban Land Cover Mapping." *Geo-spatial information science* ahead-of-print (ahead-of-print):1-14. doi: 10.1080/10095020.2022.2035656.

Bwangoy, J. R. B., M. C. Hansen, D. P. Roy, G. De Grandi, and C. O. Justice. 2010. "Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices." *Remote Sensing of Environment* 114 (1):73-86. doi: 10.1016/j.rse.2009.08.004.

Ceron, C. N., A. M. Melesse, R. Price, S. B. Dessu, and H. P. Kandel. 2015. "Operational Actual Wetland Evapotranspiration Estimation for South Florida Using MODIS Imagery." *Remote Sensing* 7 (4):3613-32. doi: 10.3390/rs70403613.

Chen, Guangyi, and Shen-En Qian. 2009. "Denoising and dimensionality reduction of hyperspectral imagery using wavelet packets, neighbour shrinking and principal component analysis." *International journal of remote sensing* 30 (18):4889-95. doi: 10.1080/01431160802653724.

Chen, Shengbo, Juliana Useya, and Hillary Mugiyo. 2020. "Decision-level fusion of Sentinel-1 SAR and Landsat 8 OLI texture features for crop discrimination and classification: case of Masvingo, Zimbabwe." *Heliyon* 6 (11):e05358. doi: 10.1016/j.heliyon.2020.e05358.

Chen, Wen-Sheng, Xiuli Dai, Binbin Pan, and Taiquan Huang. 2015. "A novel discriminant criterion based on feature fusion strategy for face recognition." *Neurocomputing (Amsterdam)* 159:67-77. doi: 10.1016/j.neucom.2015.02.019.

Cochrane, M. 2000. “Using Vegetation Reflectance Variability for Species Level Classification of Hyperspectral Data.” International Journal of Remote Sensing 21: 2075–2087. doi:10.1080/01431160050021303.

Coll, C., J. M. Galve, J. M. Sanchez, and V. Caselles. 2010. "Validation of Landsat-7/ETM+

Thermal-Band Calibration and Atmospheric Correction With Ground-Based Measurements." *IEEE transactions on geoscience and remote sensing* 48 (1):547-55. doi: 10.1109/TGRS.2009.2024934.

Dahl, Thomas E. 2000. *Status and trends of wetlands in the conterminous United States 1986 to 1997*. Washington, D.C: U.S. Dept. of the Interior, U.S. Fish and Wildlife Service.

Davranche, A., G. Lefebvre, and B. Poulin. 2010. "Wetland monitoring using classification trees and SPOT-5 seasonal time series." *Remote Sensing of Environment* 114 (3):552-62. doi: 10.1016/j.rse.2009.10.009.

Dempster, Arthur P. "Upper and Lower Probabilities Induced by a Multivalued Mapping." In, 57-72. Berlin, Heidelberg: Springer Berlin Heidelberg.

Deng, Tengfang, Bolin Fu, Man Liu, Hongchang He, Donglin Fan, Lilong Li, Liangke Huang, and Ertao Gao. 2022. "Comparison of Multi-Class and Fusion of Multiple Single-Class SegNet Model for Mapping Karst Wetland Vegetation Using UAV Images." *Scientific reports* 12, no. 1 : 13270–13270. doi:10.1038/s41598-022-17620-2

Dubeau, Pierre, Douglas King, Dikaso Unbushe, and Lisa-Maria Rebelo. 2017. "Mapping the Dabus Wetlands, Ethiopia, Using Random Forest Classification of Landsat, PALSAR and Topographic Data." *Remote sensing (Basel, Switzerland)* 9 (10):1056. doi: 10.3390/rs9101056.

European Space Agency. 2021. "Sentinel-2 Products Specification Document." European Space Agency: Paris, France.

European Space Agency. 2018. "Sentinel-1-observation scenario—planned acquisitions—ESA." https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario (accessed on 05 February 2018)

Eisavi, V., S. Homayouni, A. M. Yazdi, and A. Alimohammadi. 2015. "Land cover mapping based on random forest classification of multitemporal spectral and thermal images." *Environmental Monitoring and Assessment* 187 (5):14. doi: 10.1007/s10661-015-4489-3.

Ezugwu, Absalom E., Abiodun M. Ikotun, Olaide O. Oyelade, Laith Abualigah, Jeffery O. Agushaka, Christopher I. Eke, and Andronicus A. Akinyelu. 2022. "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects." *Engineering applications of artificial intelligence* 110:104743. https://doi.org/10.1016/j.engappai.2022.104743.

Feizizadeh, Bakhtiar. 2018. "A Novel Approach of Fuzzy Dempster-Shafer Theory for Spatial Uncertainty Analysis and Accuracy Assessment of Object-Based Image Classification." *IEEE geoscience and remote sensing letters* 15 (1):18-22. doi: 10.1109/LGRS.2017.2763979.

Feng, Tianjing, Hairong Ma, and Xinwen Cheng. 2021. "Land-cover classification of high-

resolution remote sensing image based on multi-classifier fusion and the improved Dempster–Shafer evidence theory." *Journal of applied remote sensing* 15 (1):014506-. doi: 10.1117/1.JRS.15.014506.

Finlayson, C.M., and Davidson, N.C., 1999. *Global review of wetland resources and priorities for wetland inventory: Summary report.* Supervising Scientist: Canberra, Australia

Foody, Giles M. 2002. "Status of land cover classification accuracy assessment." *Remote sensing of environment* 80 (1):185-201. https://doi.org/10.1016/S0034-4257(01)00295-4.

Frisk, J. 2010. *Guidance for the Preparation of ESTR Products—Land Classification Approach. Canadian Biodiversity: Ecosystem Status and Trends 2010.* Technical Thematic Report No. 3. Canadian Councils of Resource Ministers, Canadian Councils of Resource Ministers: Ottawa, ON, Canada

Frohn, R. C., B. C. Autrey, C. R. Lane, and M. Reif. 2011. "Segmentation and object-oriented classification of wetlands in a karst Florida landscape using multi-season Landsat-7 ETM+ imagery." *International Journal of Remote Sensing* 32 (5):1471-89. doi: 10.1080/01431160903559762.

Frohn, R. C., and N. Chaudhary. 2008. "Multi-scale Image Segmentation and Object-Oriented Processing for Land Cover Classification." *Giscience & Remote Sensing* 45 (4):377-91. doi: 10.2747/1548-1603.45.4.377.

Furtado, Luiz Felipe de Almeida, Thiago Sanna Freire Silva, and Evlyn Márcia Leão de Moraes Novo. 2016. "Dual-season and full-polarimetric C band SAR assessment for vegetation mapping in the Amazon várzea wetlands." *Remote sensing of environment* 174:212-22. doi: 10.1016/j.rse.2015.12.013.

Gala, T. S., and Assefa M. Melesse. 2012. "Monitoring prairie wet area with an integrated LANDSAT ETM+, RADARSAT-1 SAR and ancillary data from LIDAR." *Catena (Giessen)* 95:12-23. doi: 10.1016/j.catena.2012.02.022.

Gallant, A. L. 2015. "The Challenges of Remote Monitoring of Wetlands." *Remote Sensing* 7 (8):10938-50. doi: 10.3390/rs70810938.

Gallant, John C., and Trevor I. Dowling. 2003. "A multiresolution index of valley bottom flatness for mapping depositional areas." *Water resources research* 39 (12):1347-n/a. doi: 10.1029/2002WR001426.

Gao, B. C. 1996. "NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space." *Remote Sensing of Environment* 58 (3):257-66. doi: 10.1016/s0034-4257(96)00067-3.

Ghosh, Aniruddha, Richa Sharma, and P. K. Joshi. 2014. "Random forest classification of urban

landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion." *Applied geography (Sevenoaks)* 48:31-41. doi: 10.1016/j.apgeog.2014.01.003.

Guo, Baofeng. 2019a. "Entropy-Mediated Decision Fusion for Remotely Sensed Image Classification." *Remote sensing (Basel, Switzerland)* 11 (3):352. doi: 10.3390/rs11030352.

———. 2019b. "Entropy-Mediated Decision Fusion for Remotely Sensed Image Classification." *Remote sensing (Basel, Switzerland)* 11 (3):352. doi: 10.3390/rs11030352.

Guo, Meng, Jing Li, Chunlei Sheng, Jiawei Xu, and Li Wu. 2017. "A Review of Wetland Remote Sensing" *Sensors* 17, no. 4: 777. https://doi.org/10.3390/s17040777S

Grenier, M., A.-M. Demers, S. Labrecque, M. Benoit, R. A. Fournier, and B. Drolet. 2007. "An Object-Based Method to Map Wetland Using RADARSAT-1 and Landsat ETM Images: Test Case on Two Sites in Quebec, Canada." *Canadian Journal of Remote Sensing* 33: S28‐S45. doi:10.5589/m07-048.

Hall, Peter, Byeong U. Park, and Richard J. Samworth. 2008. "Choice of Neighbor Order in Nearest-Neighbor Classification." *The Annals of statistics* 36 (5):2135-52. doi: 10.1214/07-AOS537.

Hastie, T., Tibshirani, R., & Friedman, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction. 2nd ed.* pp. 210-201. New York: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction. 2nd ed.* pp. 423-425. New York: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction. 2nd ed.* pp. 590-592. New York: Springer.

Henderson, F. M., and A. J. Lewis. 2008. "Radar Detection of Wetland Ecosystems: A Review." *International Journal of Remote Sensing* 29: 5809–5835. doi:10.1080/ 01431160801958405.

Hopf, Konstantin, and Sascha Reifenrath. 2021. "Filter Methods for Feature Selection in Supervised Machine Learning Applications -- Review and Benchmark." https://doi.org/10.48550/arxiv.2111.12140.

Hong, S.-H., H.-O. Kim, S. Wdowinski, and E. Feliciano. 2015. "Evaluation of Polarimetric SAR Decomposition for Classifying Wetland Vegetation Types." *Remote Sensing* 7: 8563–8585. doi:10.3390/rs70708563.

Hosseiny, Benyamin, Masoud Mahdianpari, Brian Brisco, Fariba Mohammadimanesh, and Bahram Salehi. 2022. "WetNet: A Spatial-Temporal Ensemble Deep Learning Model for Wetland Classification Using Sentinel-1 and Sentinel-2." *IEEE transactions on*

geoscience and remote sensing 60:1-14. doi: 10.1109/TGRS.2021.3113856.

Hu, Baoxin, Qian Li, and G. Brent Hall. 2021. "A decision-level fusion approach to tree species classification from multi-source remotely sensed data." *ISPRS Open Journal of Photogrammetry and Remote Sensing* 1:100002. doi: 10.1016/j.ophoto.2021.100002.

Hu, Yabin, Jie Zhang, Yi Ma, Jubai An, Guangbo Ren, and Xiaomin Li. 2019. "Hyperspectral Coastal Wetland Classification Based on a Multiobject Convolutional Neural Network Model and Decision Fusion." *IEEE geoscience and remote sensing letters* 16 (7):1110-4. doi: 10.1109/LGRS.2018.2890421.

Hu, Yabin, Jie Zhang, Yi Ma, Xiaomin Li, Qinpei Sun, and Jubai An. 2019. "Deep learning classification of coastal wetland hyperspectral image combined spectra and texture features: A case study of Huanghe (Yellow) River Estuary wetland." *Acta oceanologica Sinica* 38 (5):142-50. doi: 10.1007/s13131-019-1445-z.

Huang, Chengquan, Yi Peng, Megan Lang, In-Young Yeo, and Greg McCarty. 2014. "Wetland inundation mapping and change monitoring using Landsat and airborne LiDAR data." *Remote sensing of environment* 141:231-42. doi: 10.1016/j.rse.2013.10.020.

Huaxin, Liu, Jiang Qigang, Ma Yue, Yang Qian, Shi Pengfei, Zhang Sen, Tan Yang, et al. 2022a. "Object-Based Multigrained Cascade Forest Method for Wetland Classification Using Sentinel-2 and Radarsat-2 Imagery." *Water (Basel)* 14 (82):82. doi: 10.3390/w14010082.

———. 2022b. "Object-Based Multigrained Cascade Forest Method for Wetland Classification Using Sentinel-2 and Radarsat-2 Imagery." *Water (Basel)* 14 (82):82. doi: 10.3390/w14010082.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." *Remote sensing of environment* 83 (1):195-213. doi: 10.1016/S0034-4257(02)00096-2.

Husak, G.J.; Hadley, B.C.; McGwire, K.C.1999. “Landsat thematic mapper registration accuracy and its effects on the IGBP validation.” *Photogramm. Eng. Remote Sens* 65, 1033–1039.

Ihlen, V., & Zanter, K. 2019. *Landsat 8 (L8) Data Users Handbook*. Version 5.0, Sioux Falls, South Dakota: Department of the Interior U.S. Geological Survey.

Imani, Maryam, and Hassan Ghassemian. 2020a. "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges." *Information fusion* 59:59-83. doi: 10.1016/j.inffus.2020.01.007.

———. 2020b. "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges." *Information fusion* 59:59-83. doi: 10.1016/j.inffus.2020.01.007.

Irons, J. R., N. J. Speciale, J. D. McCuistion, J. G. Masek, B. L. Markham, J. C. Storey, D. E.

Lencioni, R. E. Ryan, and Ieee. 2003. "Data specifications for the Landsat Data Continuity Mission." *Igarss 2003: Ieee International Geoscience and Remote Sensing Symposium, Vols I - Vii, Proceedings: Learning from Earth's Shapes and Sizes*:1335-7.

Irons, James R., Nicholas J. Speciale, J. Douglas McCuistion, Jeffrey G. Masek, Brian L. Markham, James C. Storey, Donald E. Lencioni, and Robert E. Ryan. 2003. "Data specifications for the Landsat data continuity mission." *IEEE International Geoscience and Remote Sensing Symposium Proceedings* 2003:1335-7.

Jensen, J.R. 2004. *Introductory Digital Image Processing: A Remote Sensing Perspective.* Upper Saddle River, NJ, USA: Prentice Hall

Jia, Sen, Zhangwei Zhan, Meng Zhang, Meng Xu, Qiang Huang, Jun Zhou, and Xiuping Jia. 2021a. "Multiple Feature-Based Superpixel-Level Decision Fusion for Hyperspectral and LiDAR Data Classification." *IEEE transactions on geoscience and remote sensing* 59 (2):1437-52. doi: 10.1109/TGRS.2020.2996599.

———. 2021b. "Multiple Feature-Based Superpixel-Level Decision Fusion for Hyperspectral and LiDAR Data Classification." *IEEE transactions on geoscience and remote sensing* 59 (2):1437-52. doi: 10.1109/TGRS.2020.2996599.

———. 2021c. "Multiple Feature-Based Superpixel-Level Decision Fusion for Hyperspectral and LiDAR Data Classification." *IEEE transactions on geoscience and remote sensing* 59 (2):1437-52. doi: 10.1109/TGRS.2020.2996599.

Jiao, L. L., W. W. Sun, G. Yang, G. B. Ren, and Y. N. Liu. 2019a. "A Hierarchical Classification Framework of Satellite Multispectral/Hyperspectral Images for Mapping Coastal Wetlands." *Remote Sensing* 11 (19):21. doi: 10.3390/rs11192238.

Jiao, Leilei, Weiwei Sun, Gang Yang, Guangbo Ren, and Yinnian Liu. 2019b. "A Hierarchical Classification Framework of Satellite Multispectral/Hyperspectral Images for Mapping Coastal Wetlands." *Remote sensing (Basel, Switzerland)* 11 (19):2238. doi: 10.3390/rs11192238.

Juan, Munizaga, García Mariano, Ureta Fernando, Novoa Vanessa, Rojas Octavio, and Rojas Carolina. 2022. "Mapping Coastal Wetlands Using Satellite Imagery and Machine Learning in a Highly Urbanized Landscape." *Sustainability (Basel, Switzerland)* 14 (9):5700. doi: 10.3390/su14095700.

Judah, A., and B. X. Hu. 2019. "The Integration of Multi-source Remotely-Sensed Data in Support of the Classification of Wetlands." *Remote Sensing* 11 (13):28. doi: 10.3390/rs11131537.

Judah, Aaron, and Baoxin Hu. 2021. "The Integration of Multi-Source Remotely Sensed Data with Hierarchically Based Classification Approaches in Support of the Classification of Wetlands." *Canadian journal of remote sensing* 48 (2):158-81. doi: 10.1080/07038992.2021.1967732.

Kamaruzaman, J., and I. Kasawani. 2007. “Imaging Spectrometry on Mangrove Species Identification and Mapping in Malaysia.” *WSEAS Trans Biol Biomed* 8: 118–126.

Kaplan, Gordana, and Ugur Avdan. 2018. "Monthly Analysis of Wetlands Dynamics Using Remote Sensing Data" *ISPRS International Journal of Geo-Information* 7, no. 10: 411. https://doi.org/10.3390/ijgi7100411

Kecman, V., Huang, T.-M., Vogt, M. 2005. *“*Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance.” In *Support Vector Machines: Theory and Applications. By* Lipo, W., 255-274. Berlin, Germany: Springer.

Klemas, V. 2013. "Remote sensing of emergent and submerged wetlands: an overview." *International Journal of Remote Sensing* 34 (18):6286-320. doi: 10.1080/01431161.2013.800656.

Koch, M., T. Schmid, M. Reyes, and J. Gumuzzio. 2012. “Evaluating Full Polarimetric C- and L-Band Data for Mapping Wetland Conditions in a Semi-Arid Environment in Central Spain.”*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5: 1033–1044. doi:10.1109/JSTARS.2012.2202091.

Koma, Z., A. C. Seijmonsbergen, and W. D. Kissling. 2021. "Classifying wetland-related land cover types and habitats using fine-scale lidar metrics derived from country-wide Airborne Laser Scanning." *Remote Sensing in Ecology and Conservation* 7 (1):80-96. doi: 10.1002/rse2.170.

Kushwaha, S. P. S., R. S. Dwivedi, and B. R. M. Rao. 2000. "Evaluation of various digital image processing techniques for detection of coastal wetlands using ERS-1 SAR data." *International Journal of Remote Sensing* 21 (3):565-79. doi: 10.1080/014311600210759.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012. “Imagenet classification with deep convolutional neural networks.” *Paper presented at International Conference on Neural Information Processing Systems*, Doha, Qatar, 1097–1105.

Li, Shutao, Xudong Kang, Leyuan Fang, Jianwen Hu, and Haitao Yin. 2017. "Pixel-level image fusion: A survey of the state of the art." *Information fusion* 33:100-12. doi: 10.1016/j.inffus.2016.05.004.

Li, Y., Zhang, H., & Shen, Q. 2017. “Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network.” *Remote Sensing*, 9(1), 67.

Liang, S. L., A. H. Strahler, and C. Walthall. 1999. "Retrieval of land surface albedo from satellite observations: A simulation study." *Journal of Applied Meteorology* 38 (6):712-25. doi: 10.1175/1520-0450(1999)038<0712:rolsaf>2.0.co;2.

Liang, Shunlin. 2001. "Narrowband to broadband conversions of land surface albedo I:

Algorithms." *Remote sensing of environment* 76 (2):213-38. doi: 10.1016/S0034-4257(00)00205-4.

Lin, Shangrong, Jing Li, Qinhuo Liu, Longhui Li, Jing Zhao, and Wentao Yu. 2019. "Evaluating the Effectiveness of Using Vegetation Indices Based on Red-Edge Reflectance from Sentinel-2 to Estimate Gross Primary Productivity." *Remote sensing (Basel, Switzerland)* 11 (11):1303. doi: 10.3390/rs11111303.

Lin, Yi, and Yongho Jeon. 2006. "Random Forests and Adaptive Nearest Neighbors." *Journal of the American Statistical Association* 101 (474):578-90. doi: 10.1198/016214505000001230.

Lin, Y., and C. Yue. 2014. “China’s New National Rules on Wetland Protection.” doi:10.2139/ssrn.2517481.

Liu, Shuai, and Mulan Gao. 2022. "Decision Fusion Using Similarity-weighted JCR and Mid-level Features based ELM for Hyperspectral Image Classification with Limited Training Samples." *International journal of remote sensing* 43 (3):873-93. doi: 10.1080/01431161.2021.2022238.

Loh, W. Y. 2002. "Regression trees with unbiased variable selection and interaction detection." *Statistica Sinica* 12 (2):361-86.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant. 2000. "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data." *International journal of remote sensing* 21 (6-7):1303-1330. https://doi.org/10.1080/014311600210191.

Mahdavi, S., B. Salehi, J. Granger, M. Amani, B. Brisco, and W. M. Huang. 2018. "Remote sensing for wetland classification: a comprehensive review." *Giscience & Remote Sensing* 55 (5):623-58. doi: 10.1080/15481603.2017.1419602.

Mahdianpari, M., B. Salehi, F. Mohammadimanesh, S. Homayouni, and E. Gill. 2019. "The First Wetland Inventory Map of Newfoundland at a Spatial Resolution of 10 m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform." *Remote Sensing* 11 (1):27. doi: 10.3390/rs11010043.

Mahdianpari, Masoud, Bahram Salehi, Mohammad Rezaee, Fariba Mohammadimanesh, and Yun Zhang. 2018. "Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery" *Remote Sensing* 10, no. 7: 1119. https://doi.org/10.3390/rs10071119

Mandianpari, M., B. Salehi, F. Mohammadimanesh, and M. Motagh. 2017. "Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery." *Isprs Journal of Photogrammetry and Remote Sensing* 130:13-31. doi: 10.1016/j.isprsjprs.2017.05.010.

Mathworks. 2016. “Global Optimization Toolbox: User's Guide (r2016b).” Available online: https://www.mathworks.com/help/pdf_doc/gads/gads_tb.pdf (accessed on 10 November 2016).

Mao, Dehua, Zongming Wang, Baojia Du, Lin Li, Yanlin Tian, Mingming Jia, Yuan Zeng, Kaishan Song, Ming Jiang, and Yeqiao Wang. 2020b. "National wetland mapping in China: A new product resulting from object-based and hierarchical classification of Landsat 8 OLI images." *ISPRS journal of photogrammetry and remote sensing* 164:11-25. doi: 10.1016/j.isprsjprs.2020.03.020.

Markham, B. L., J. C. Storey, D. L. Williams, and J. R. Irons. 2004. "Landsat sensor performance: history and current status." *IEEE transactions on geoscience and remote sensing* 42 (12):2691-4. doi: 10.1109/TGRS.2004.840720.

Masoumi, F., T. Eslamkish, A. A. Abkar, M. Honarmand, and J. R. Harris. 2017a. "Integration of spectral, thermal, and textural features of ASTER data using Random Forests classification for lithological mapping." *Journal of African Earth Sciences* 129:445-57. doi: 10.1016/j.jafrearsci.2017.01.028.

Masoumi, Feizollah, Taymour Eslamkish, Ali Akbar Abkar, Mehdi Honarmand, and Jeff R. Harris. 2017b. "Integration of spectral, thermal, and textural features of ASTER data using random forests classification for lithological mapping." *Journal of African earth sciences (1994)* 129:445-57. doi: 10.1016/j.jafrearsci.2017.01.028.

Maxwell, Aaron E., Timothy A. Warner, and Fang Fang. 2018. "Implementation of machine-learning classification in remote sensing: an applied review." *International journal of remote sensing* 39 (9):2784-817. doi: 10.1080/01431161.2018.1433343.

Mellor, A.; Haywood, A.; Jones, S.; Wilkes, P. 2012. “Forest Classification Using Random Forests With Multi Source Remote Sensing and Ancillary GIS Data.” In Proceedings of the 16th Australasian Remote Sensing and Photogrammetry Conference Proceedings, Melbourne, Australia, 27–28 August 2012.

Meng, Xiangrui, Shuqing Zhang, and Shuying Zang. 2019. "Lake Wetland Classification Based on an SVM-CNN Composite Classifier and High-resolution Images Using Wudalianchi as an Example." *Journal of coastal research* 93 (sp1):153-162. https://doi.org/10.2112/SI93-022.1.

Millard, K., and M. Richardson. 2013. "Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR-SAR fusion using a random forest classifier." *Canadian Journal of Remote Sensing* 39 (4):290-307. doi: 10.5589/m13-038.

Ministry of Natural Resources—Ontario. 2012. *Ontario Forestry Resources Inventory Calibration Plot Specifications and associated data sets, Published March, 2012* Queen's Printer for Ontario.

Ming, Dongping, Tianning Zhou, Min Wang, and Tian Tan. 2016. "Land cover classification using random forest with genetic algorithm-based parameter optimization." *Journal of applied remote sensing* 10 (3):035021-. doi: 10.1117/1.JRS.10.035021.

Mirmazloumi, S. Mohammad, Armin Moghimi, Babak Ranjgar, Farzane Mohseni, Arsalan Ghorbanian, Seyed Ali Ahmadi, Meisam Amani, and Brian Brisco. 2021. "Status and Trends of Wetland Studies in Canada Using Remote Sensing Technology with a Focus on Wetland Classification: A Bibliographic Analysis" *Remote Sensing* 13, no. 20: 4025. https://doi.org/10.3390/rs13204025

Miyamoto, M., K. Kushida, K. Yoshino, T. Nagano, Y. Sato, and Ieee. 2003. "Evaluation of multispatial scale measurements for monitoring wetland vegetation, Kushiro wetland, JAPAN: Application of SPOT images, CASI data, airborne CNIR video images and balloon aerial photography." *Igarss 2003: Ieee International Geoscience and Remote Sensing Symposium, Vols I - Vii, Proceedings: Learning from Earth's Shapes and Sizes*:3275-7.

Mohammadimanesh, F., B. Salehi, M. Mahdianpari, M. Motagh, and B. Brisco. 2018. "An efficient feature optimization for wetland mapping by synergistic use of SAR intensity, interferometry, and polarimetry data." *International Journal of Applied Earth Observation and Geoinformation* 73:450-62. doi: 10.1016/j.jag.2018.06.005.

Mohammadimanesh, F., B. Salehi, M. Mandianpari, M. Motagh, and Ieee. 2018. "A New Hierarchical Object-Based Classification Algorithm for Wetland Mapping in Newfoundland, Canada." *Igarss 2018 - 2018 Ieee International Geoscience and Remote Sensing Symposium*:9233-6.

Munoz, D. E., J. R. Cissell, and H. Moftakhari. 2019. "Adjusting Emergent Herbaceous Wetland Elevation with Object-Based Image Analysis, Random Forest and the 2016 NLCD." *Remote Sensing* 11 (20):20. doi: 10.3390/rs11202346.

Mwita, E., G. Menz, S. Misana, M. Becker, D. Kisanga, and B. Boehme. 2013. "Mapping small wetlands of Kenya and Tanzania using remote sensing techniques." *International Journal of Applied Earth Observation and Geoinformation* 21:173-83. doi: 10.1016/j.jag.2012.08.010.

Natural Resources Canada Map Information Branch. 2016. "Canadian Digital Elevation Model Product Specifications." Government of Canada. Last modified 01 April 2013. http://ftp.geogratis.gc.ca/pub/nrcan_rncan/elevation/cdem_mnec/doc/CDEM_product_specs.pdf (accessed on 2 March 2016).

O'Neil, Gina L., Jonathan L. Goodall, Madhur Behl, and Linnea Saby. 2020. "Deep learning Using Physically-Informed Input Data for Wetland Identification." *Environmental modelling & software : with environment data news* 126:104665. doi: 10.1016/j.envsoft.2020.104665.

Ozesmi, Stacy L., and Marvin E. Bauer. 2002. "Satellite remote sensing of wetlands." *Wetlands Ecology and Management* 10 (5):381-402. doi: 10.1023/a:1020908432489.

Pearl, Judea. 1997. *Probabilistic reasoning in intelligent systems : networks of plausible inference, Morgan Kaufmann series in representation and reasoning*. San Francisco, Calif: Morgan Kaufmann Publishers.

Periasamy, Shoba. 2018. "Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1." *Remote sensing of environment* 217:537-49. doi: 10.1016/j.rse.2018.09.003.

Poe, Gregory L., Kelly L. Giraud, and John B. Loomis. 2005. "Computational Methods for Measuring the Difference of Empirical Distributions." *American journal of agricultural economics* 87 (2):353-365. https://doi.org/10.1111/j.1467-8276.2005.00727.x.

Pouliot, Darren, Rasim Latifovic, Jon Pasher, and Jason Duffe. 2019. "Assessment of Convolution Neural Networks for Wetland Mapping with Landsat in the Central Canadian Boreal Forest Region" *Remote Sensing* 11, no. 7: 772. https://doi.org/10.3390/rs11070772

Ramsey, E. W., and S. C. Laine. 1997a. "Comparison of landsat thematic mapper and high resolution photography to identify change in complex coastal wetlands." *Journal of Coastal Research* 13 (2):281-92.

———. 1997b. "Comparison of landsat thematic mapper and high resolution photography to identify change in complex coastal wetlands." *Journal of Coastal Research* 13 (2):281-92.

Rapinel, Sébastien, Laurence Hubert-Moy, and Bernard Clément. 2015. "Combined use of LiDAR data and multispectral earth observation imagery for wetland habitat mapping." *ITC journal* 37:56-64. doi: 10.1016/j.jag.2014.09.002.

Ray, Partha Pratim. 2023. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope." *Internet of Things and Cyber-Physical Systems* 3:121-154. https://doi.org/10.1016/j.iotcps.2023.04.003.

Rasti, Behnood, Pedram Ghamisi, and Richard Gloaguen. 2017. "Hyperspectral and LiDAR Fusion Using Extinction Profiles and Total Variation Component Analysis." *IEEE transactions on geoscience and remote sensing* 55 (7):3997-4007. doi: 10.1109/TGRS.2017.2686450.

"Researchers at Jilin University Publish New Data on Science and Technology (Decision-level fusion of Sentinel-1 SAR and Landsat 8 OLI texture features for crop discrimination and classification: case of Masvingo, Zimbabwe)." 2020. *Science Letter*:1446-.

Rezaee, Mohammad, Masoud Mahdianpari, Yun Zhang, and Bahram Salehi. 2018. "Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery." *IEEE journal of selected topics in applied earth observations and remote sensing* 11 (9):3030-9. doi: 10.1109/JSTARS.2018.2846178.

Rocha, Adrian V., and Gaius R. Shaver. 2009. "Advantages of a two band EVI calculated from solar and photosynthetically active radiation fluxes." *Agricultural and forest meteorology* 149 (9):1560-3. doi: 10.1016/j.agrformet.2009.03.016.

Ronan Collobert, J. W. & Weston, J. 2008. "A unified architecture for natural language processing: Deep neural networks with multitask." *Paper presented at Proceedings of the 25th International Conference on Machine*, New York, USA, 160–167.

Rogan, John & Miller, Jennifer & Stow, Douglas & Franklin, Janet & Levien, Lisa & Fischer, Chris. 2003. "Land-Cover Change Monitoring with Classification Trees Using Landsat TM and Ancillary Data." *Photogrammetric Engineering & Remote Sensing* 69. 10.14358/PERS.69.7.793.

Rundquist, Donald C., Sunil Narumalani, and Ram M. Narayanan. 2001. "A review of wetlands remote sensing and defining new considerations." *Remote sensing reviews* 20 (3):207-26. doi: 10.1080/02757250109532435.

Schmidt, K., and A. Skidmore. 2003. "Spectral Discrimination of Vegetation Types in a Coastal Wetland." Remote Sensing of Environment 85: 92–108. doi:10.1016/S0034-4257(02)00196-7.

Sellers, P. J. 1985. "Canopy Reflectance, Photosynthesis and Transpiration." *International Journal of Remote Sensing* 6 (8):1335-72. doi: 10.1080/01431168508948283.

Shunlin, Liang, A. H. Strahler, and C. Walthall. 1999a. "Retrieval of Land Surface Albedo from Satellite Observations: A Simulation Study." *Journal of applied meteorology (1988)* 38 (6):712-25. doi: 10.1175/1520-0450(1999)0382.0.CO;2.

———. 1999b. "Retrieval of Land Surface Albedo from Satellite Observations: A Simulation Study." *Journal of applied meteorology (1988)* 38 (6):712-25. doi: 10.1175/1520-0450(1999)0382.0.CO;2.

Sentinel-1 Product Specification. 2018. Available online: https://sentinel.esa.int/documents/247904/1877131/Sentinel-1-Product-Specification (accessed on 5 February 2018).

Sellers, P. J. 1985. "Canopy reflectance, photosynthesis and transpiration." *International Journal of Remote Sensing* 6 (8):1335-72. doi: 10.1080/01431168508948283.

Solorio-Fernández, Saúl, Carrasco-Ochoa J. Ariel, and José Fco Martínez-Trinidad. 2020. "A review of unsupervised feature selection methods." *The Artificial intelligence review* 53 (2):907-948. https://doi.org/10.1007/s10462-019-09682-y.

Sun, W.; Ren, K.; Meng, X.; Yang, G.; Xiao, C.; Peng, J.; Huang, J. MLR-DBPFN: A Multi-Scale Low Rank Deep Back Researchion Fusion Network for Anti-Noise Hyperspectral and Multispectral Image Fusion. *IEEE Trans. Geosci. Remote Sens.* 2022, *60*, 5522914. https://doi.org/10.1109/TGRS.2022.3146296.

Thenkabail, P. S., E. A. Enclona, M. S. Ashton, and B. Van Der Meer. 2004. "Accuracy Assessments of Hyperspectral Waveband Performance for Vegetation Analysis Applications." *Remote Sensing of Environment* 91: 354–376. doi:10.1016/j.rse.2004.03.013.

Tian, S. H., X. F. Zhang, J. Tian, and Q. Sun. 2016. "Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China." *Remote Sensing* 8 (11):14. doi: 10.3390/rs8110954.

U.S. Fish and Wildlife Service. 2002."National Wetlands Inventory: A Strategy for the 21st Century." *Department of the Interior, Fish and Wildlife Service: Washington, DC, USA*

Useya, Juliana, and Shengbo Chen. 2018a. "Comparative Performance Evaluation of Pixel-Level and Decision-Level Data Fusion of Landsat 8 OLI, Landsat 7 ETM+ and Sentinel-2 MSI for Crop Ensemble Classification." *IEEE journal of selected topics in applied earth observations and remote sensing* 11 (11):4441-51. doi: 10.1109/JSTARS.2018.2870650.
———. 2018b. "Comparative Performance Evaluation of Pixel-Level and Decision-Level Data Fusion of Landsat 8 OLI, Landsat 7 ETM+ and Sentinel-2 MSI for Crop Ensemble Classification." *IEEE journal of selected topics in applied earth observations and remote sensing* 11 (11):4441-51. doi: 10.1109/JSTARS.2018.2870650.

Wang, J., J. Shang, B. Brisco, and R. J. Brown. 1997. Comparison of Multidate RADAR and Multispectral Optical Satellite Data for Wetland Detection in the Great Lakes Region. Proceedings of Geomatics in the era of RADARSAT.

Wang, Y., J. Knight, L. P. Rampi, R. Cao, and Ieee. 2014. "MAPPING WETLAND CHANGE OF PRAIRIE POTHOLE REGION IN BIGSTONE COUNTY FROM 1938 YEAR TO 2011 YEAR." *2014 Ieee International Geoscience and Remote Sensing Symposium (Igarss)*:4. doi: 10.1109/igarss.2014.6947451.

Weichelt, H., P. Rosso, A. Marx, S. Reigber, K. Douglass, and M. Heynen. 2011. "White Paper–The RapidEye Red Edge Band." *RapidEye White Papers*: 1–6.

Wen, X., Z. Zhou, B. Chen, Z. Li, and X. Tang. 2014. "Research on the Features of Chlorophyll-A Derived from Rapideye and EOS/MODIS Data in Chaohu Lake." In 35th International GIScience & Remote Sensing 657 Symposium on Remote Sensing of Environment (ISRSE35), IOP Publishing. Beijing. April 21–26.

Wenzhi, Liao, Aleksandra Pizurica, Rik Bellens, Sidharta Gautama, and Wilfried Philips. 2015a. "Generalized Graph-Based Fusion of Hyperspectral and LiDAR Data Using Morphological Features." *IEEE geoscience and remote sensing letters* 12 (3):552-6. doi: 10.1109/LGRS.2014.2350263.

———. 2015b. "Generalized Graph-Based Fusion of Hyperspectral and LiDAR Data Using Morphological Features." *IEEE geoscience and remote sensing letters* 12 (3):552-6. doi: 10.1109/LGRS.2014.2350263.

Wright, C., and A. Gallant. 2007. "Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data." *Remote Sensing of Environment* 107 (4):582-605. doi: 10.1016/j.rse.2006.10.019.

Wu, Zhenjiang, Jiahua Zhang, Fan Deng, Sha Zhang, Da Zhang, Lan Xun, Mengfei Ji, and Qian Feng. 2021. "Superpixel-Based Regional-Scale Grassland Community Classification Using Genetic Programming with Sentinel-1 SAR and Sentinel-2 Multispectral Images." *Remote sensing (Basel, Switzerland)* 13 (20):4067. doi: 10.3390/rs13204067.

Xu, L., A. Krzyzak, and C. Y. Suen. 1992. "Methods of combining multiple classifiers and their applications to handwriting recognition." *IEEE transactions on systems, man, and* cybernetics 22 (3):418-35. doi: 10.1109/21.155943.

Xu, Y., Du, J., Dai, L. R., & Lee, C. H. 2015. A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 23(1), 7–19.

Yuan, Yi, Xiangchao Meng, Weiwei Sun, Gang Yang, Lihua Wang, Jiangtao Peng, and Yumiao Wang. 2022. "Multi-Resolution Collaborative Fusion of SAR, Multispectral and Hyperspectral Images for Coastal Wetlands Mapping" *Remote Sensing* 14, no. 14: 3492. https://doi.org/10.3390/rs14143492

Zhong, Zisha, Bin Fan, Kun Ding, Haichang Li, Shiming Xiang, and Chunhong Pan. 2016. "Efficient Multiple Feature Fusion With Hashing for Hyperspectral Imagery Classification: A Comparative Study." *IEEE transactions on geoscience and remote sensing* 54 (8):4461-78. doi: 10.1109/TGRS.2016.2542342.

Zhu, Jun-Jie, Jinyue Jiang, Meiqi Yang, and Zhiyong Jason Ren. 2023. "ChatGPT and Environmental Research." *Environmental science & technology*. https://doi.org/10.1021/acs.est.3c01818.

Appendix

Appendix 1. Summary of test inputs for classification comparisons. Tests determined by the holistic approach (blue text), tests determined by Log-normal values (green text), and tests determined by RF determined predictor importance values (purple text). For tests in green and purple, features are represented by their respective index number ref. Table 3-3

1	Band 1-7 Reflectances (Spring)
2	Band 1-7 Reflectances (Summer)
3	Band 1-7 Reflectances (Fall)
4	Band 1-7 Reflectances, Temp 1 (Spring)
5	Band 1-7 Reflectances, Temp 1 (Summer)
6	Band 1-7 Reflectances, Temp 1 (Fall)
7	Band 1-7 Reflectances, Temp 2 (Spring)
8	Band 1-7 Reflectances, Temp 2 (Summer)
9	Band 1-7 Reflectances, Temp 2 (Fall)
10	Band 1-7 Reflectances, Radar-VV (Summer)
11	Band 1-7 Reflectances, Radar-VH (Summer)
12	Band 1-7 Reflectances, Radar-VV (Summer2)
13	Band 1-7 Reflectances, Radar-VH (Summer2)
14	Band 1-7 Reflectances, Radar-VV (Fall)
15	Band 1-7 Reflectances, Radar-VH (Fall)
16	Band 1-7 Reflectances, DEM (Spring)
17	Band 1-7 Reflectances, DEM (Summer)
18	Band 1-7 Reflectances, DEM (Fall)
19	Band 1-7 Reflectances, Slope (Spring)
20	Band 1-7 Reflectances, Slope (Summer)
21	Band 1-7 Reflectances, Slope (Fall)
22	Band 1-7 Reflectances, DEM, Slope (Spring)
23	Band 1-7 Reflectances, DEM, Slope (Summer)
24	Band 1-7 Reflectances, DEM, Slope (Fall)
25	Band 1-7 Reflectances, Radar - (Alpha - Summer)

26	Band 1-7 Reflectances, Radar - (Entropy - Summer)
27	Band 1-7 Reflectances, Radar - (Alpha - Fall)
28	Band 1-7 Reflectances, Radar - (Entropy - Fall)
29	Band 1-7 Reflectances, Temp, Radar-VV (Summer)
30	Band 1-7 Reflectances, Temp, Radar-VH (Summer)
31	Band 1-7 Reflectances, Temp, Radar-VV (Summer2)
32	Band 1-7 Reflectances, Temp, Radar-VH (Summer2)
33	Band 1-7 Reflectances, Temp, Radar-VV (Fall)
34	Band 1-7 Reflectances, Temp, Radar-VH (Fall)
35	Band 1-7 Reflectances, Temp2, Radar-VV (Summer)
36	Band 1-7 Reflectances, Temp2, Radar-VH (Summer)
37	Band 1-7 Reflectances, Temp2, Radar-VV (Summer2)
38	Band 1-7 Reflectances, Temp2, Radar-VH (Summer2)
39	Band 1-7 Reflectances, Temp2, Radar-VV (Fall)
40	Band 1-7 Reflectances, Temp2, Radar-VH (Fall)
41	Band 1-7 Reflectances, Temp, Radar-VV (Summer), DEM
42	Band 1-7 Reflectances, Temp, Radar-VH (Summer), DEM
43	Band 1-7 Reflectances, Temp, Radar-VV (Summer2), DEM
44	Band 1-7 Reflectances, Temp, Radar-VH (Summer2), DEM
45	Band 1-7 Reflectances, Temp, Radar-VV (Fall), DEM
46	Band 1-7 Reflectances, Temp, Radar-VH (Fall), DEM
47	Band 1-7 Reflectances, Temp2, Radar-VV (Summer), DEM
48	Band 1-7 Reflectances, Temp2, Radar-VH (Summer), DEM
49	Band 1-7 Reflectances, Temp2, Radar-VV (Summer2), DEM
50	Band 1-7 Reflectances, Temp2, Radar-VH (Summer2), DEM
51	Band 1-7 Reflectances, Temp2, Radar-VV (Fall), DEM
52	Band 1-7 Reflectances, Temp2, Radar-VH (Fall), DEM
53	Band 1-7 Reflectances, Temp, Radar-VV (Summer), Slope
54	Band 1-7 Reflectances, Temp, Radar-VH (Summer), Slope
55	Band 1-7 Reflectances, Temp, Radar-VV (Summer2), Slope
56	Band 1-7 Reflectances, Temp, Radar-VH (Summer2), Slope

57	Band 1-7 Reflectances, Temp, Radar-VV (Fall), Slope
58	Band 1-7 Reflectances, Temp, Radar-VH (Fall), Slope
59	Band 1-7 Reflectances, Temp2, Radar-VV (Summer), Slope
60	Band 1-7 Reflectances, Temp2, Radar-VH (Summer), Slope
61	Band 1-7 Reflectances, Temp2, Radar-VV (Summer2), Slope
62	Band 1-7 Reflectances, Temp2, Radar-VH (Summer2), Slope
63	Band 1-7 Reflectances, Temp2, Radar-VV (Fall), Slope
64	Band 1-7 Reflectances, Temp2, Radar-VH (Fall), Slope
65	Band 1-7 Reflectances, Temp, Alpha (Summer), Slope
66	Band 1-7 Reflectances, Temp, Entropy (Summer), Slope
67	Band 1-7 Reflectances, Temp, Alpha (Fall), Slope
68	Band 1-7 Reflectances, Temp, Entropy (Fall), Slope
69	All Band Reflectances All Seasons
70	All Band Reflectances, Radar(VV) All Seasons
71	All Band Reflectances, Temp1, All Seasons
72	All Band Reflectances, Slope, All Seasons
73	All Band Reflectances, Radar (VV), Temp1, All Seasons
74	All Band Reflectances, Radar (VV), Temp1 Slope, All Seasons
75	All Band Reflectances, Radar(VH) All Seasons
76	All Band Reflectances, Radar (VH), Temp1, All Seasons
77	All Band Reflectances, Radar (VH), Temp1, Slope, All Seasons
78	All Band Reflectances, Radar (VV,VH), Temp1, Slope, All Seasons
79	All Band Reflectances, Alpha, All Seasons
80	All Band Reflectances, Temp, Alpha, All Seasons
81	All Band Reflectances, Entropy, All Seasons
82	All Band Reflectances, Temp, Entropy, All Seasons
83	All Band Reflectances, Temp2, All Seasons
84	All Band Reflectances, Radar (VV), Temp2, All Seasons
85	All Band Reflectances, Radar (VV), Temp2 Slope, All Seasons
86	All Band Reflectances, Radar (VH), Temp2, All Seasons
87	All Band Reflectances, Radar (VH), Temp2, Slope, All Seasons

88	All Band Reflectances, Radar (VV,VH), Temp2, Slope, All Seasons
89	NDVI, NDWI, Albedo (Spring)
90	NDVI, NDWI, Albedo (Summer)
91	NDVI, NDWI, Albedo (Fall)
92	NDVI, NDWI, Albedo, Temp 1 (Spring)
93	NDVI, NDWI, Albedo, Temp 1 (Summer)
94	NDVI, NDWI, Albedo, Temp 1 (Fall)
95	NDVI, NDWI, Albedo, Temp 2 (Spring)
96	NDVI, NDWI, Albedo, Temp 2 (Summer)
97	NDVI, NDWI, Albedo, Temp 2 (Fall)
98	NDVI, NDWI, Albedo, Radar-VV (Summer)
99	NDVI, NDWI, Albedo, Radar-VH (Summer)
100	NDVI, NDWI, Albedo, Radar-VV (Summer2)
101	NDVI, NDWI, Albedo, Radar-VH (Summer2)
102	NDVI, NDWI, Albedo, Radar-VV (Fall)
103	NDVI, NDWI, Albedo, Radar-VH (Fall)
104	NDVI, NDWI, Albedo, DEM (Spring)
105	NDVI, NDWI, Albedo, DEM (Summer)
106	NDVI, NDWI, Albedo, DEM (Fall)
107	NDVI, NDWI, Albedo, Slope (Spring)
108	NDVI, NDWI, Albedo, Slope (Summer)
109	NDVI, NDWI, Albedo, Slope (Fall)
110	NDVI, NDWI, Albedo, Alpha (Summer)
111	NDVI, NDWI, Albedo, Alpha (Fall)
112	NDVI, NDWI, Albedo, Entropy (Summer)
113	NDVI, NDWI, Albedo, Entropy (Fall)
114	NDVI, NDWI, Albedo, DEM, Slope (Spring)
115	NDVI, NDWI, Albedo, DEM, Slope (Summer)
116	NDVI, NDWI, Albedo, DEM, Slope (Fall)
117	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer)
118	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer)

119	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2)
120	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2)
121	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall)
122	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall)
123	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer)
124	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer)
125	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2)
126	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2)
127	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall)
128	NDVI, NDWI, Albedo, Temp2, Radar-VH (Fall)
129	NDVI, NDWI, Albedo, Temp, Alpha (Summer)
130	NDVI, NDWI, Albedo, Temp, Entropy (Summer)
131	NDVI, NDWI, Albedo, Temp, Alpha (Fall)
132	NDVI, NDWI, Albedo, Temp, Entropy (Fall)
133	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer), DEM
134	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer), DEM
135	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2), DEM
136	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2), DEM
137	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall), DEM
138	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall), DEM
139	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer), DEM
140	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer), DEM
141	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2), DEM
142	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2), DEM
143	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall), DEM
144	NDVI, NDWI, Albedo, Temp2, Radar-VH (Fall), DEM
145	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer), Slope
146	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer), Slope
147	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2), Slope
148	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2), Slope
149	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall), Slope

150	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall), Slope
151	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer), Slope
152	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer), Slope
153	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2), Slope
154	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2), Slope
155	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall), Slope
156	NDVI, NDWI, Albedo, Temp, Alpha (Summer), Slope
157	NDVI, NDWI, Albedo, Temp, Entropy (Summer),Slope
158	NDVI, NDWI, Albedo, Temp, Alpha (Fall),Slope
159	NDVI, NDWI, Albedo, Temp, Entropy (Fall),Slope
160	NDVI, NDWI, Albedo, Temp2, Radar-VH (Fall), Slope
161	NDVI, NDWI, Albedo All Seasons
162	NDVI, NDWI, Albedo, Radar(VV) All Seasons
163	NDVI, NDWI, Albedo, Temp1, All Seasons
164	NDVI, NDWI, Albedo, Slope, All Seasons
165	NDVI, NDWI, Albedo, Radar (VV), Temp1, All Seasons
166	NDVI, NDWI, Albedo, Radar (VV), Temp1, Slope, All Seasons
167	NDVI, NDWI, Albedo, Radar (VV,VH), Temp1, Slope, All Seasons
168	NDVI, NDWI, Albedo, Radar(VH) All Seasons
169	NDVI, NDWI, Albedo, Temp2, All Seasons
170	NDVI, NDWI, Albedo, Radar (VH), Temp1, All Seasons
171	NDVI, NDWI, Albedo, Radar (VH), Temp1, Slope, All Seasons
172	NDVI, NDWI, Albedo, Alpha, Temp1, All Seasons
173	NDVI, NDWI, Albedo, Entropy, Temp1, Slope, All Seasons
174	NDVI, NDWI, Albedo, Temp2, All Seasons
175	NDVI, NDWI, Albedo, Radar (VV), Temp2, All Seasons
176	NDVI, NDWI, Albedo, Radar (VV), Temp2 Slope, All Seasons
177	NDVI, NDWI, Albedo, Radar (VH), Temp2, All Seasons
178	NDVI, NDWI, Albedo, Radar (VH), Temp2, Slope, All Seasons
179	NDVI, NDWI, Albedo, Radar (VV,VH), Temp2, Slope, All Seasons
180	NDVI, NDWI, Albedo, Radar (VV,VH), Temp2, Slope, Alpha, Entropy, All Seasons

181	Top 5 [16,22,21,7,10]
182	Top 10 [16,22,21,7,10,8,20,9,24,43]
183	Top 15 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6]
184	Top 20 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18]
185	Top 25 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17]
186	Top 30 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17]
187	Top 35 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17,36,12,30,23,15]
188	Top 40 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17,36,12,30,23,15,2,1,27,3,13]
189	Top 45 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17,36,12,30,23,15,2,1,27,3,13,29,35,45,46,44]
190	All Variables
191	Hybrid Top 5 [16,22,21,7,22,43]
192	Hybrid Top 10 [16,22,21,7,22,43,6,28,19,21,45]
193	Hybrid Top 15 [16,22,21,7,22,43,6,28,19,21,45,26,18,4,10,46]
194	Hybrid Top 20 [16,22,21,7,22,43,6,28,19,21,45,26,18,4,10,46,25,31,14,8,44]
195	Hybrid Top 25 [16,22,21,7,22,43,6,28,19,21,45,26,18,4,10,46,25,31,14,8,44,17,30,15,20,48]
196	Hybrid Top 30 [16,22,21,7,22,43,6,28,19,21,45,26,18,4,10,46,25,31,14,8,44,17,30,15,20,48,2,1,27,9,47]
197	Bottom 16 [39,40,41,42,37,38,47,48,44,46,45,35,29,13,3,27]
198	Bottom 14 [39,40,41,42,37,38,47,48,44,46,45,35,29,13]
199	Bottom 12 [39,40,41,42,37,38,47,48,44,46,45,35]
200	Bottom 10 [39,40,41,42,37,38,47,48,44,46]
201	Bottom 8 [39,40,41,42,37,38,47,48]
202	Bottom 6 [39,40,41,42,37,38]
203	Bottom 4 [39,40,41,42]
204	Top 5 [34,32,43,9,8]
205	Top 10 [34,32,43,9,8,21,20,33,25,22]
206	Top 15 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35]
207	Top 20 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19]
208	Top 25 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26]
209	Top 30 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26,7,6,1,23,4]

210	Top 35 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26,7,6,1,23,4,24,15,2,3,46]
211	Top 40 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26,7,6,1,23,4,24,15,2,3,46,12,18,30,29,17]
212	Top 45 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26,7,6,1,23,4,24,15,2,3,46,12,18,30,29,17,40,36,41,39,42]
213	Hybrid Top 5 [25,5,16,34,43]
214	Hybrid Top 10 [25,5,16,34,43,13,28,19,32,44]
215	Hybrid Top 15 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45]
216	Hybrid Top 20 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45,7,6,1,21,47]
217	Hybrid Top 25 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45,7,6,1,21,47,4,15,2,33,40]
218	Hybrid Top 30 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45,7,6,1,21,47,4,15,2,33,40,3,18,30,22,41]
219	Bottom 16 [2,3,46,12,18,30,29,17,40,36,41,39,42,27,37,38]
220	Bottom 14 [2,3,46,12,18,30,29,17,40,36,41,39,42,27]
221	Bottom 12 [2,3,46,12,18,30,29,17,40,36,41,39]
222	Bottom 10 [2,3,46,12,18,30,29,17,40,36]
223	Bottom 8 [2,3,46,12,18,30,29,17]
224	Bottom 6 [2,3,46,12,18,30]
225	Bottom 4 [2,3,46,12]

Appendix 2. Summary of top 20 (A) and bottom 20 (B) classification inputs for various classification schemes. Blue highlighted tests were determined through a holistic approach to feature selection, while the green and purple highlighted tests were selected through log-normal and RF importance value analysis, respectively.

A)

	Random Forest Classifier	Producer Accuracy	Kappa Value
1	All Band Reflectances, Radar (VH), Temp1, Slope, All Seasons	0.8781	0.8841
2	All Band Reflectances, Temp1, All Seasons	0.8766	0.8839
3	All Band Reflectances, Temp, Alpha, All Seasons	0.8714	0.8769
4	All Band Reflectances, Temp, Entropy, All Seasons	0.8671	0.8766
5	All Band Reflectances, Radar (VV,VH), Temp1, Slope, All Seasons	0.8574	0.8753
6	All Band Reflectances, Entropy, All Seasons	0.8515	0.8753
7	Top 40 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17,36,12, 30,23, 15,2,1,27,3,13]	0.8492	0.8734
8	All Band Reflectances, Radar (VV,VH), Temp2, Slope, All Seasons	0.8468	0.8718
9	All Band Reflectances, Radar (VV), Temp1 Slope, All Seasons	0.8467	0.8712
10	All Band Reflectances, Alpha, All Seasons	0.8466	0.8688
11	All Band Reflectances, Radar(VV) All Seasons	0.8454	0.8659
12	All Band Reflectances All Seasons	0.8453	0.8654
13	Hybrid Top 20 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45,7,6,1,21,47]	0.8450	0.8649
14	All Band Reflectances, Temp2, All Seasons	0.8447	0.8618
15	All Band Reflectances, Radar (VH), Temp1, All Seasons	0.8445	0.8617
16	All Band Reflectances, Radar(VH) All Seasons	0.8445	0.8587
17	All Band Reflectances, Radar (VH), Temp2, Slope, All Seasons	0.8436	0.8576
18	All Band Reflectances, Slope, All Seasons	0.8394	0.8532
19	All Band Reflectances, Radar (VV), Temp2 Slope, All Seasons	0.8346	0.8531
20	Top 45 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26,7,6,1,23,4,24,15, 2,3, 46,12,18,30,29,17,40,36,41,39,42]	0.8328	0.8511
	Nearest Neighbor Classifier	**Producer Accuracy**	**Kappa Value**
1	Top 20 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18]	0.7746	0.8540

2	Top 35 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17,36,12, 30,23,15]	0.7668	0.8512
3	Top 25 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17]	0.7662	0.8470
4	Top 30 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17]	0.7643	0.8499
5	Top 15 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6]	0.7475	0.8370
6	All Band Reflectances, Temp1, All Seasons	0.7425	0.8265
7	All Band Reflectances, Temp2, All Seasons	0.7368	0.8237
8	Top 10 [16,22,21,7,10,8,20,9,24,43]	0.7321	0.8250
9	Band 1-7 Reflectances, DEM (Spring)	0.7220	0.8189
10	Top 15 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35]	0.7215	0.8220
11	Hybrid Top 15 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45]	0.7210	0.8268
12	NDVI, NDWI, Albedo, DEM (Summer)	0.7117	0.8022
13	Band 1-7 Reflectances, Temp 1 (Spring)	0.7117	0.8074
14	Hybrid Top 30 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45,7,6,1,21,47,4,15,2,33,40,3,18,30,22,41]	0.7114	0.8125
15	All Band Reflectances All Seasons	0.7094	0.7967
16	Top 40 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17,4,25,31,14,17,36,12, 30,23, 15,2,1,27,3,13]	0.7084	0.8148
17	Top 20 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19]	0.7082	0.8179
18	NDVI, NDWI, Albedo, Temp1, All Seasons	0.7054	0.8077
19	All Band Reflectances, Radar (VH), Temp1, Slope, All Seasons	0.7023	0.8052
20	NDVI, NDWI, Albedo, Temp2, All Seasons	0.7021	0.8039
	Support Vector Machine	**Producer Accuracy**	**Kappa Value**
1	All Band Reflectances, Radar(VH) All Seasons	0.8004	0.8675
2	All Band Reflectances, Radar(VV) All Seasons	0.7985	0.8650
3	All Band Reflectances, Alpha, All Seasons	0.7964	0.8592
4	All Band Reflectances, Entropy, All Seasons	0.7910	0.8604
5	All Band Reflectances All Seasons	0.7883	0.8596
6	Band 1-7 Reflectances, DEM (Spring)	0.7871	0.8608
7	Band 1-7 Reflectances, DEM, Slope (Spring)	0.7808	0.8576

8	All Band Reflectances, Slope, All Seasons	0.7778	0.8557
9	Band 1-7 Reflectances, Temp2, Radar-VH (Fall), DEM	0.7708	0.8300
10	Band 1-7 Reflectances, Temp2, Radar-VV (Fall), DEM	0.7694	0.8264
11	All Band Reflectances, Radar (VV), Temp2, All Seasons	0.7605	0.8238
12	All Band Reflectances, Radar (VH), Temp1, All Seasons	0.7575	0.8192
13	Hybrid Top 20 [25,5,16,34,43,13,28,19,32,44,31,14,26,9,45,7,6,1,21,47]	0.7575	0.8340
14	All Band Reflectances, Radar (VV), Temp2 Slope, All Seasons	0.7552	0.8151
15	All Band Reflectances, Radar (VV), Temp1, All Seasons	0.7541	0.8050
16	All Band Reflectances, Temp1, All Seasons	0.7528	0.8040
17	All Band Reflectances, Temp2, All Seasons	0.7516	0.8193
18	All Band Reflectances, Radar (VH), Temp2, All Seasons	0.7501	0.8156
19	All Band Reflectances, Radar (VV,VH), Temp2, Slope, All Seasons	0.7499	0.8046
20	All Band Reflectances, Temp, Entropy, All Seasons	0.7486	0.8147
	Naive Bayes	**Producer Accuracy**	**Kappa Value**
1	All Band Reflectances, Temp, Entropy, All Seasons	0.6998	0.7931
2	All Band Reflectances, Temp, Alpha, All Seasons	0.6960	0.7912
3	All Band Reflectances, Radar (VV,VH), Temp1, Slope, All Seasons	0.6902	0.7897
4	All Band Reflectances, Temp2, All Seasons	0.6891	0.7895
5	All Band Reflectances, Radar (VH), Temp2, All Seasons	0.6886	0.7893
6	All Band Reflectances, Radar (VV,VH), Temp2, Slope, All Seasons	0.6875	0.7900
7	All Band Reflectances, Radar (VV), Temp2, All Seasons	0.6870	0.7885
8	All Band Reflectances, Radar (VV), Temp1, All Seasons	0.6861	0.7869
9	All Band Reflectances, Temp1, All Seasons	0.6860	0.7854
10	All Band Reflectances, Radar (VH), Temp1, All Seasons	0.6855	0.7858
11	All Band Reflectances, Radar (VH), Temp2, Slope, All Seasons	0.6851	0.7880
12	All Band Reflectances, Radar (VH), Temp1, Slope, All Seasons	0.6850	0.7858
13	All Band Reflectances, Radar (VV), Temp1 Slope, All Seasons	0.6841	0.7860
14	All Band Reflectances, Radar (VV), Temp2 Slope, All Seasons	0.6832	0.7870
15	Top 25 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19,48,47,31,14,26]	0.6783	0.7841
16	All Band Reflectances, Entropy, All Seasons	0.6713	0.7507

17	Top 20 [34,32,43,9,8,21,20,33,25,22,5,16,11,10,35,13,28,44,45,19]	0.6698	0.7766
18	All Band Reflectances, Radar(VV) All Seasons	0.6671	0.7497
19	All Band Reflectances, Alpha, All Seasons	0.6670	0.7492
20	Top 25 [16,22,21,7,10,8,20,9,24,43,33,34,32,5,6,28,11,19,26,18,4,25,31,14,17]	0.6648	0.7741

B)

	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
206	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2), Slope	0.6080	0.6519
207	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall)	0.6079	0.6518
208	NDVI, NDWI, Albedo, Slope (Fall)	0.6071	0.6498
209	NDVI, NDWI, Albedo, Temp 1 (Fall)	0.6061	0.6484
210	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer)	0.6057	0.6470
211	NDVI, NDWI, Albedo, Entropy (Fall)	0.6045	0.6383
212	NDVI, NDWI, Albedo, Temp, Alpha (Fall)	0.5990	0.6232
213	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2)	0.5975	0.6148
214	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2)	0.5963	0.6049
215	NDVI, NDWI, Albedo, Temp 2 (Summer)	0.5960	0.5995
216	NDVI, NDWI, Albedo, Alpha (Fall)	0.5914	0.5951
217	NDVI, NDWI, Albedo, Slope (Spring)	0.5820	0.5950
218	NDVI, NDWI, Albedo, Radar-VV (Fall)	0.5735	0.5922
219	NDVI, NDWI, Albedo, Radar-VH (Fall)	0.5590	0.5909
220	NDVI, NDWI, Albedo (Spring)	0.5531	0.5897
221	NDVI, NDWI, Albedo (Fall)	0.5529	0.5896
222	Top 5 [16,22,21,7,10]	0.5088	0.5878
223	Bottom 8 [2,3,46,12,18,30,29,17]	0.4917	0.5182
224	Bottom 6 [2,3,46,12,18,30]	0.3853	0.3840
225	Bottom 4 [2,3,46,12]	0.3638	0.3621
	Nearest Neighbor Classifier	**Producer Accuracy(%)**	**Kappa Value**
206	NDVI, NDWI, Albedo, Alpha (Summer)	0.5264	0.6151

207	Top 5 [16,22,21,7,10]	0.5244	0.5932
208	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall), Slope	0.5227	0.5974
209	NDVI, NDWI, Albedo, Alpha (Fall)	0.5186	0.6042
210	NDVI, NDWI, Albedo, Entropy (Fall)	0.5170	0.6112
211	NDVI, NDWI, Albedo, Slope (Fall)	0.5170	0.5873
212	NDVI, NDWI, Albedo (Spring)	0.5152	0.6115
213	NDVI, NDWI, Albedo (Fall)	0.5065	0.6106
214	NDVI, NDWI, Albedo, Radar-VH (Fall)	0.4974	0.5996
215	NDVI, NDWI, Albedo, Radar-VV (Fall)	0.4972	0.5920
216	Bottom 12 [2,3,46,12,18,30,29,17,40,36,41,39]	0.4898	0.5852
217	Bottom 8 [39,40,41,42,37,38,47,48]	0.4892	0.5780
218	Bottom 10 [39,40,41,42,37,38,47,48,44,46]	0.4606	0.5493
219	Bottom 12 [39,40,41,42,37,38,47,48,44,46,45,35]	0.4555	0.5393
220	Bottom 16 [39,40,41,42,37,38,47,48,44,46,45,35,29,13,3,27]	0.4522	0.5276
221	Bottom 14 [39,40,41,42,37,38,47,48,44,46,45,35,29,13]	0.4383	0.5223
222	Bottom 10 [2,3,46,12,18,30,29,17,40,36]	0.4012	0.4746
223	Bottom 4 [2,3,46,12]	0.2697	0.1994
224	Bottom 8 [2,3,46,12,18,30,29,17]	0.2515	0.1716
225	Bottom 6 [2,3,46,12,18,30]	0.2217	0.1039
	Support Vector Machine	**Producer Accuracy**	**Kappa Value**
206	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer), Slope	0.5342	0.5561
207	Bottom 10 [2,3,46,12,18,30,29,17,40,36]	0.5336	0.6677
208	NDVI, NDWI, Albedo, Temp, Entropy (Summer)	0.5332	0.5821
209	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer), Slope	0.5326	0.5586
210	NDVI, NDWI, Albedo, Temp, Alpha (Summer)	0.5325	0.5825
211	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2), Slope	0.5296	0.5569
212	NDVI, NDWI, Albedo, Entropy (Fall)	0.5286	0.6454
213	NDVI, NDWI, Albedo, Radar-VH (Fall)	0.5283	0.6403
214	NDVI, NDWI, Albedo, Temp 1 (Summer)	0.5258	0.5760
215	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2), Slope	0.5231	0.5821

216	NDVI, NDWI, Albedo, Alpha (Fall)	0.5214	0.6317
217	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2)	0.5164	0.5700
218	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2), Slope	0.5150	0.5483
219	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2)	0.5149	0.5726
220	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2), Slope	0.5132	0.5661
221	NDVI, NDWI, Albedo (Spring)	0.5015	0.5869
222	Top 5 [16,22,21,7,10]	0.4731	0.5243
223	Bottom 8 [2,3,46,12,18,30,29,17]	0.4355	0.5539
224	Bottom 6 [2,3,46,12,18,30]	0.4122	0.4473
225	Bottom 4 [2,3,46,12]	0.3979	0.4453
	Naive Bayes	**Producer Accuracy**	**Kappa Value**
206	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall), DEM	0.5167	0.6325
207	Bottom 4 [39,40,41,42]	0.5151	0.6520
208	Bottom 10 [2,3,46,12,18,30,29,17,40,36]	0.5128	0.6762
209	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall), DEM	0.5074	0.6247
210	NDVI, NDWI, Albedo, Entropy (Fall)	0.5065	0.6320
211	NDVI, NDWI, Albedo, DEM, Slope (Summer)	0.5040	0.6539
212	NDVI, NDWI, Albedo (Fall)	0.4990	0.6225
213	NDVI, NDWI, Albedo, DEM (Fall)	0.4918	0.6216
214	NDVI, NDWI, Albedo, DEM, Slope (Fall)	0.4913	0.6268
215	NDVI, NDWI, Albedo, Alpha (Fall)	0.4867	0.6177
216	NDVI, NDWI, Albedo, Temp 2 (Spring)	0.4726	0.6478
217	NDVI, NDWI, Albedo, Temp 1 (Spring)	0.4643	0.6424
218	NDVI, NDWI, Albedo, DEM, Slope (Spring)	0.4638	0.6160
219	NDVI, NDWI, Albedo, DEM (Spring)	0.4634	0.6055
220	Bottom 8 [2,3,46,12,18,30,29,17]	0.4573	0.6102
221	NDVI, NDWI, Albedo, Slope (Spring)	0.4337	0.5761
222	Top 5 [16,22,21,7,10]	0.4198	0.4514
223	NDVI, NDWI, Albedo (Spring)	0.4033	0.5437
224	Bottom 6 [2,3,46,12,18,30]	0.3432	0.3808

225	Bottom 4 [2,3,46,12]	0.3105	0.3502

Appendix 3. Image Input features used during this study and their associated variable index. Reflection shortened to "Reflect." and Sentinel-1 to "Senti." .

Index	Feature Name	Index	Feature Name	Index	Feature Name	Index	Feature Name
1	B1 Reflect. #1	**61**	B5 Reflect. #8	**121**	B5 Reflect. #13	**181**	Senti. VV - #30
2	B2 Reflect. #1	**62**	B6 Reflect. #8	**122**	B6 Reflect. #13	**182**	Senti. VH - #30
3	B3 Reflect. #1	**63**	B7 Reflect. #8	**123**	B7 Reflect. #13	**183**	Senti. VV - #31
4	B4 Reflect. #1	**64**	NDVI #8	**124**	NDVI #13	**184**	Senti. VH - #31
5	B5 Reflect. #1	**65**	NDWI #8	**125**	NDWI #13	**185**	Senti. VV - #32
6	B6 Reflect. #1	**66**	Albedo #8	**126**	Albedo #13	**186**	Senti. VH - #32
7	B7 Reflect. #1	**67**	Temp1 #8	**127**	Temp1 #13	**187**	Senti. VV - #33
8	NDVI #1	**68**	Temp2 #8	**128**	Temp2 #13	**188**	Senti. VH - #33
9	NDWI #1	**69**	B1 Reflect. #9	**129**	B1 Reflect. #14	**189**	Senti. VV - #34
10	Albedo #1	**70**	B2 Reflect. #9	**130**	B2 Reflect. #14	**190**	Senti. VH - #34
11	Temp1 #1	**71**	B3 Reflect. #9	**131**	B3 Reflect. #14	**191**	Senti. VV - #35
12	Temp2 #1	**72**	B4 Reflect. #9	**132**	B4 Reflect. #14	**192**	Senti. VH - #35
13	B1 Reflect. #2	**73**	B5 Reflect. #9	**133**	B5 Reflect. #14	**193**	Senti. VV - #36
14	B2 Reflect. #2	**74**	B6 Reflect. #9	**134**	B6 Reflect. #14	**194**	Senti. VH - #36
15	B3 Reflect. #2	**75**	B7 Reflect. #9	**135**	B7 Reflect. #14	**195**	Senti. VV - #37
16	B4 Reflect. #2	**76**	NDVI #9	**136**	NDVI #14	**196**	Senti. VH - #37
17	B5 Reflect. #2	**77**	NDWI #9	**137**	NDWI #14	**197**	Senti. VV - #38
18	B6 Reflect. #2	**78**	Albedo #9	**138**	Albedo #14	**198**	Senti. VH - #38
19	B7 Reflect. #2	**79**	Temp1 #9	**139**	Temp1 #14	**199**	Senti. VV - #39
20	NDVI #2	**80**	Temp2 #9	**140**	Temp2 #14	**200**	Senti. VH - #39
21	NDWI #2	**81**	B1 Reflect. #10	**141**	B1 Reflect. #15	**201**	Senti. VV - #40
22	Albedo #2	**82**	B2 Reflect. #10	**142**	B2 Reflect. #15	**202**	Senti. VH - #40
23	Temp1 #2	**83**	B3 Reflect. #10	**143**	B3 Reflect. #15	**203**	Senti. VV - #41
24	Temp2 #2	**84**	B4 Reflect. #10	**144**	B4 Reflect. #15	**204**	Senti. VH - #41
25	B1 Reflect. #3	**85**	B5 Reflect. #10	**145**	B5 Reflect. #15	**205**	Senti. VV - #42
26	B2 Reflect. #3	**86**	B6 Reflect. #10	**146**	B6 Reflect. #15	**206**	Senti. VH - #42
27	B3 Reflect. #3	**87**	B7 Reflect. #10	**147**	B7 Reflect. #15	**207**	Senti. VV - #43
28	B4 Reflect. #3	**88**	NDVI #10	**148**	NDVI #15	**208**	Senti. VH - #43
29	B5 Reflect. #3	**89**	NDWI #10	**149**	NDWI #15	**209**	Senti. VV - #44
30	B6 Reflect. #3	**90**	Albedo #10	**150**	Albedo #15	**210**	Senti. VH - #44
31	B7 Reflect. #3	**91**	Temp1 #10	**151**	Temp1 #15	**211**	Senti. VV - #45
32	NDVI #3	**92**	Temp2 #10	**152**	Temp2 #15	**212**	Senti. VH - #45
33	NDWI #3	**93**	B1 Reflect. #11	**153**	Senti. VV - #16	**213**	Senti VV - #46
34	Albedo #3	**94**	B2 Reflect. #11	**154**	Senti. VH - #16	**214**	Senti VH - #46
35	Temp1 #3	**95**	B3 Reflect. #11	**155**	Senti. VV - #17	**215**	Senti. VV - #47
36	Temp2 #3	**96**	B4 Reflect. #11	**156**	Senti. VH - #17	**216**	Senti. VH - #47
37	Senti. VV #4	**97**	B5 Reflect. #11	**157**	Senti. VV - #18	**217**	Senti. VV - #48
38	Senti. VH - #4	**98**	B6 Reflect. #11	**158**	Senti. VH - #18	**218**	Senti. VH - #48
39	Senti. VV - #5	**99**	B7 Reflect. #11	**159**	Senti. VV - #19	**219**	Senti. VV - #49
40	Senti. VH - #5	**100**	NDVI #11	**160**	Senti. VH - #19	**220**	Senti. VH - #49
41	Senti. VV - #6	**101**	NDWI #11	**161**	Senti. VV - #20	**221**	Senti. VV - #50
42	Senti. VH - #6	**102**	Albedo #11	**162**	Senti. VH - #20	**222**	Senti. VH - #50
43	DEM	**103**	Temp1 #11	**163**	Senti. VV - #21	**223**	Senti. VV - #51
44	DEM-Slope	**104**	Temp2 #11	**164**	Senti. VH - #21	**224**	Senti. VH - #51
45	B1 Reflect. #7	**105**	B1 Reflect. #12	**165**	Senti. VV - #22	**225**	Senti. VV - #52
46	B2 Reflect. #7	**106**	B2 Reflect. #12	**166**	Senti. VH - #22	**226**	Senti. VH - #52
47	B3 Reflect. #7	**107**	B3 Reflect. #12	**167**	Senti. VV - #23	**227**	Senti. VV - #53
48	B4 Reflect. #7	**108**	B4 Reflect. #12	**168**	Senti. VH - #23	**228**	Senti. VH - #53
49	B5 Reflect. #7	**109**	B5 Reflect. 12	**169**	Senti. VV - #24	**229**	Senti. VV - #54

50	B6 Reflect. #7	**110**	B6 Reflect. #12	**170**	Senti. VH - #24	**230**	Senti. VH - #54
51	B7 Reflect. #7	**111**	B7 Reflect. #12	**171**	Senti. VV - #25	**231**	Senti. VV - #55
52	NDVI #7	**112**	NDVI #12	**172**	Senti. VH - #25	**232**	Senti. VH - #55
53	NDWI #7	**113**	NDWI #12	**173**	Senti. VV - #26	**233**	Senti. VV - #56
54	Albedo #7	**115**	Albedo #12	**174**	Senti. VH - #26	**234**	Senti. VH - #56
55	Temp1 #7	**116**	Temp1 #12	**175**	Senti. VV - #27	**235**	Senti. VV - #57
56	Temp2 #7	**117**	Temp2 #12	**176**	Senti. VH - #27	**236**	Senti. VH - #57
57	B1 Reflect. #8	**118**	B1 Reflect. #13	**177**	Senti. VV - #28	**237**	Senti. VV - #58
58	B2 Reflect. #8	**119**	B2 Reflect. #13	**178**	Senti. VH - #28	**238**	Senti. VH - #58
59	B3 Reflect. #8	**120**	B3 Reflect. #13	**179**	Senti. VV - #29		
60	B4 Reflect. #8	**121**	B4 Reflect. #13	**180**	Senti. VH - #29		

Appendix 4. Summary of data feature quality analysis averaged over all land cover types. Image metrics, image bands, radar parameters, DEM and slope sorted from largest to smallest Log-Normal distance, with a longer distance implying a higher quality. ■ indicates a spring month, ■ indicates a summer, month, ■ indicates a fall month.

Log-Normal Distance								
Bands	**Distance**	**Metrics**		**Distance**	**Radar and DEM**	**Distance**		
B4 Reflect. IMG-7	1.916	NDVI	IMG-7	2.276	DEM	1.369		
B4 Reflect. IMG-10	1.595	NDVI	IMG-9	2.199	DEM-Slope	0.392		
B4 Reflect. IMG-11	1.480	NDVI	IMG-12	2.068	Sentinel VH - IMG-42	0.284		
B7 Reflect. IMG-1	1.478	NDVI	IMG-11	2.032	Sentinel VH - IMG-44	0.278		
B4 Reflect. IMG-12	1.478	NDVI	IMG-10	2.000	Sentinel VH - IMG-45	0.259		
B4 Reflect. IMG-9	1.478	NDWI	IMG-2	1.890	Sentinel VH - IMG-43	0.255		
B4 Reflect. IMG-2	1.450	NDVI	IMG-8	1.725	Sentinel VH - IMG-22	0.251		
B4 Reflect. IMG-13	1.411	Temp 1	IMG-10	1.705	Sentinel VH - IMG-23	0.247		
B5 Reflect. IMG-1	1.402	Albedo	IMG-7	1.665	Sentinel VH - IMG-26	0.228		
B2 Reflect. IMG-7	1.336	NDVI	IMG-13	1.639	Sentinel VH - IMG-50	0.220		
B7 Reflect. IMG-10	1.281	NDWI	IMG-3	1.632	Sentinel VH - IMG-46	0.205		
B4 Reflect. IMG-3	1.253	Albedo	IMG-2	1.580	Sentinel VH - IMG-47	0.200		
B6 Reflect. IMG-1	1.230	Temp 2	IMG-10	1.534	Sentinel VH - IMG-48	0.184		
B7 Reflect. IMG-7	1.230	Albedo	IMG-13	1.528	Sentinel VH - IMG-49	0.181		
B4 Reflect. IMG-8	1.230	Temp 1	IMG-11	1.519	Sentinel VH - IMG-34	0.178		
B3 Reflect. IMG-7	1.206	Albedo	IMG-10	1.513	Sentinel VH - IMG-31	0.177		
B7 Reflect. IMG-2	1.189	NDWI	IMG-1	1.475	Sentinel VH - IMG-38	0.176		
B2 Reflect. IMG-10	1.188	Temp 2	IMG-11	1.426	Sentinel VH - IMG-33	0.172		
B2 Reflect. IMG-9	1.159	Temp 1	IMG-12	1.369	Sentinel VH - IMG-30	0.172		
B7 Reflect. IMG-12	1.159	Albedo	IMG-11	1.344	Sentinel VH - IMG-41	0.164		
B1 Reflect. IMG-7	1.159	Temp 2	IMG-12	1.340	Sentinel VH - IMG-24	0.161		
B2 Reflect. IMG-13	1.154	NDVI	IMG-3	1.310	Sentinel VV - IMG-42	0.159		
B6 Reflect. IMG-7	1.115	NDVI	IMG-2	1.298	Sentinel VH - IMG-51	0.158		
B6 Reflect. IMG-10	1.094	Temp 2	IMG-2	1.264	Sentinel VV - IMG-49	0.158		
B3 Reflect. IMG-9	1.053	Temp 1	IMG-9	1.249	Sentinel VV - IMG-38	0.157		
B4 Reflect. IMG-1	1.048	Temp 1	IMG-7	1.237	Sentinel VV - IMG-33	0.154		
B2 Reflect. IMG-3	1.043	Albedo	IMG-12	1.231	Sentinel VH - IMG-37	0.151		
B6 Reflect. IMG-2	0.995	NDVI	IMG-1	1.225	Sentinel VV - IMG-51	0.150		
B6 Reflect. IMG-12	0.986	Temp 2	IMG-13	1.181	Sentinel VH - IMG-28	0.148		
B1 Reflect. IMG-10	0.974	Temp 2	IMG-7	1.178	Sentinel VH - IMG-57	0.143		
B7 Reflect. IMG-11	0.963	Temp 2	IMG-9	1.174	Sentinel VH - IMG-16	0.139		
B7 Reflect. IMG-3	0.960	Temp 1	IMG-13	1.140	Sentinel VH - IMG-27	0.139		
B5 Reflect. IMG-2	0.959	Albedo	IMG-3	1.111	Sentinel VH - IMG-53	0.139		
B5 Reflect. IMG-9	0.946	Albedo	IMG-1	1.091	Sentinel VH - IMG-25	0.139		
B5 Reflect. IMG-10	0.929	Temp1	IMG-2	1.044	Sentinel VV - IMG-32	0.135		
B3 Reflect. IMG-10	0.929	Albedo	IMG-8	0.986	Sentinel VV - IMG-30	0.134		
B6 Reflect. IMG-3	0.929	Albedo	IMG-9	0.985	Sentinel VH - IMG-29	0.132		
B6 Reflect. IMG-8	0.913	Temp 2	IMG-8	0.963	Sentinel VH - IMG-40	0.130		
B5 Reflect. IMG-11	0.908	NDWI	IMG-7	0.957	Sentinel VH - IMG-39	0.130		
B7 Reflect. IMG-8	0.908	Temp2	IMG-3	0.956	Sentinel VV - IMG-45	0.129		
B3 Reflect. IMG-11	0.901	NDWI	IMG-10	0.949	Sentinel VH - IMG-21	0.128		
B6 Reflect. IMG-11	0.897	Temp 1	IMG-1	0.927	Sentinel VV - IMG-53	0.125		
B5 Reflect. IMG-12	0.883	Temp 1	IMG-1	0.901	Sentinel VV - IMG-41	0.122		
B3 Reflect. IMG-2	0.883	Temp 1	IMG-8	0.897	Sentinel VV - IMG-39	0.122		
B5 Reflect. IMG-7	0.869	NDWI	IMG-9	0.887	Sentinel VV - IMG-31	0.121		
B2 Reflect. IMG-8	0.865	NDWI	IMG-12	0.867	Sentinel VH - IMG-56	0.120		
B1 Reflect. IMG-13	0.855	Temp 1	IMG-3	0.866	Sentinel VH - IMG-58	0.120		
B2 Reflect. IMG-1	0.852	NDWI	IMG-11	0.857	Sentinel VV - IMG-52	0.120		

B2 Reflect.	IMG-12	0.850	NDWI	IMG-8	0.832	Sentinel VH - IMG-17	0.119
B6 Reflect.	IMG-9	0.846	NDWI	IMG-13	0.829	Sentinel VV - IMG-43	0.119
B2 Reflect.	IMG-2	0.841	Temp 1	IMG-14	0.769	Sentinel VV - IMG-50	0.119
B5 Reflect.	IMG-13	0.822	NDWI	IMG-14	0.392	Sentinel VH - IMG-36	0.117
B1 Reflect.	IMG-3	0.818	NDWI	IMG-15	0.344	Sentinel VV - IMG-21	0.113
B3 Reflect.	IMG-3	0.809	Temp 2	IMG-14	0.273	Sentinel VH - IMG-32	0.112
B6 Reflect.	IMG-13	0.809	Temp 2	IMG-15	0.113	Sentinel VV - IMG-54	0.111
B1 Reflect.	IMG-9	0.809	NDVI	IMG-14	0.107	Sentinel VV - IMG-58	0.111
B7 Reflect.	IMG-9	0.790	Albedo	IMG-15	0.105	Sentinel VV - IMG-29	0.110
B1 Reflect.	IMG-1	0.789	Temp 1	IMG-15	0.105	Sentinel VV - IMG-24	0.107
B7 Reflect.	IMG-13	0.777	NDVI	IMG-15	0.092	Sentinel VV - IMG-57	0.105
B3 Reflect.	IMG-8	0.757	Albedo	IMG-14	0.085	Sentinel VV - IMG-40	0.105
B3 Reflect.	IMG-13	0.753				Sentinel VH - IMG-55	0.104
B3 Reflect.	IMG-1	0.752				Sentinel VV - IMG-16	0.102
B3 Reflect.	IMG-12	0.752				Sentinel VV - IMG-48	0.102
B5 Reflect.	IMG-8	0.752				Sentinel VV - IMG-35	0.099
B5 Reflect.	IMG-3	0.735				Sentinel VV - IMG-22	0.098
B1 Reflect.	IMG-8	0.727				Sentinel VH - IMG-35	0.095
B2 Reflect.	IMG-11	0.727				Sentinel VH - IMG-4	0.094
B1 Reflect.	IMG-12	0.716				Sentinel VH - IMG-54	0.094
B1 Reflect.	IMG-2	0.705				Sentinel VV - IMG-34	0.093
B1 Reflect.	IMG-11	0.705				Sentinel VH - IMG-52	0.092
B1 Reflect.	IMG-15	0.705				Sentinel VH - IMG-18	0.091
B1 Reflect.	IMG-14	0.705				Sentinel VV - IMG-27	0.090
B7 Reflect.	IMG-15	0.704				Sentinel VH - IMG-20	0.090
B5 Reflect.	IMG-15	0.698				Sentinel VV - IMG-44	0.090
B3 Reflect.	IMG-15	0.682				Sentinel VV - IMG-55	0.089
B4 Reflect.	IMG-14	0.655				Sentinel VV - IMG-36	0.089
B7 Reflect.	IMG-14	0.569				Sentinel VV - IMG-4	0.088
B5 Reflect.	IMG-14	0.565				Sentinel VV - IMG-25	0.084
B2 Reflect.	IMG-15	0.550				Sentinel VV - IMG-19	0.080
B6 Reflect.	IMG-15	0.549				Sentinel VH - IMG-6	0.079
B6 Reflect.	IMG-14	0.539				Sentinel VH - IMG-19	0.077
B4 Reflect.	IMG-15	0.504				Sentinel VV - IMG-37	0.074
B3 Reflect.	IMG-14	0.474				Sentinel VV - IMG-6	0.067
B2 Reflect.	IMG-14	0.329				Sentinel VV - IMG-56	0.064
						Sentinel VV - IMG-26	0.063
						Sentinel VV - IMG-46	0.063
						Sentinel VV - IMG-47	0.062
						Sentinel VV - IMG-20	0.061
						Sentinel VV - IMG-28	0.061
						Sentinel VV - IMG-18	0.059
						Sentinel VV - IMG-17	0.059
						Sentinel VV - IMG-5	0.058
						Sentinel VH - IMG-5	0.049
						Sentinel VV - IMG-23	0.048
Average		0.789	Average		0.945	Average	0.144
Standard Deviation		0.406	Standard Deviation		0.289	Standard Deviation	0.141

Appendix 5. Summary of data feature quality analysis averaged over all land cover types. Image metrics, image bands, radar parameters, DEM and slope sorted from largest to smallest RF predictor importance value. Larger values implying a higher quality. ■ indicates a spring month, ■ indicates a summer, month, ■ indicates a fall month.

Random Forest Determined Predictor Importance								
Bands		**Distance**	**Metrics**		**Distance**	**Radar and DEM**		**Distance**
B4 Reflect. IMG-7	■	1.145	NDVI	IMG-7	0.788	DEM		1.072
B4 Reflect. IMG-10	■	1.130	NDVI	IMG-9	0.779	DEM-Slope		0.759
B4 Reflect. IMG-11	■	1.081	NDVI	IMG-12	0.704	Sentinel VH - IMG-42	■	0.412
B7 Reflect. IMG-1	■	1.012	NDVI	IMG-11	0.699	Sentinel VH - IMG-44	■	0.376
B4 Reflect. IMG-12	■	0.809	NDVI	IMG-10	0.669	Sentinel VH - IMG-45	■	0.364
B4 Reflect. IMG-9	■	0.804	NDWI	IMG-2	0.621	Sentinel VH - IMG-43	■	0.363
B4 Reflect. IMG-2	■	0.803	NDVI	IMG-8	0.575	Sentinel VH - IMG-22	■	0.354
B4 Reflect. IMG-13	■	0.774	Temp 1	IMG-10	0.575	Sentinel VH - IMG-23	■	0.339
B5 Reflect. IMG-1	■	0.770	Albedo	IMG-7	0.570	Sentinel VH - IMG-26	■	0.338
B2 Reflect. IMG-7	■	0.769	NDVI	IMG-13	0.566	Sentinel VH - IMG-50	■	0.338
B7 Reflect. IMG-10	■	0.767	NDWI	IMG-3	0.557	Sentinel VH - IMG-46	■	0.335
B4 Reflect. IMG-3	■	0.745	Albedo	IMG-2	0.555	Sentinel VH - IMG-47	■	0.323
B6 Reflect. IMG-1	■	0.680	Temp 2	IMG-10	0.548	Sentinel VH - IMG-48	■	0.320
B7 Reflect. IMG-7	■	0.659	Albedo	IMG-13	0.547	Sentinel VH - IMG-49	■	0.318
B4 Reflect. IMG-8	■	0.648	Temp 1	IMG-11	0.544	Sentinel VH - IMG-34	■	0.318
B3 Reflect. IMG-7	■	0.633	Albedo	IMG-10	0.535	Sentinel VH - IMG-31	■	0.317
B7 Reflect. IMG-2	■	0.633	NDWI	IMG-1	0.535	Sentinel VH - IMG-38	■	0.315
B2 Reflect. IMG-10	■	0.631	Temp 2	IMG-11	0.533	Sentinel VH - IMG-33	■	0.315
B2 Reflect. IMG-9	■	0.600	Temp 1	IMG-12	0.528	Sentinel VH - IMG-30	■	0.313
B7 Reflect. IMG-12	■	0.595	Albedo	IMG-11	0.522	Sentinel VH - IMG-41	■	0.312
B1 Reflect. IMG-7	■	0.588	Temp 2	IMG-12	0.495	Sentinel VH - IMG-24	■	0.312
B2 Reflect. IMG-13	■	0.556	NDVI	IMG-3	0.486	Sentinel VV - IMG-42	■	0.311
B6 Reflect. IMG-7	■	0.536	NDVI	IMG-2	0.473	Sentinel VH - IMG-51	■	0.311
B6 Reflect. IMG-10	■	0.531	Temp 2	IMG-2	0.471	Sentinel VV - IMG-49	■	0.309
B3 Reflect. IMG-9	■	0.528	Temp 1	IMG-9	0.470	Sentinel VV - IMG-38	■	0.307
B4 Reflect. IMG-1	■	0.524	Temp 1	IMG-7	0.467	Sentinel VV - IMG-33	■	0.306
B2 Reflect. IMG-3	■	0.523	Albedo	IMG-12	0.462	Sentinel VH - IMG-37	■	0.306
B6 Reflect. IMG-2	■	0.520	NDVI	IMG-1	0.462	Sentinel VV - IMG-51	■	0.302
B6 Reflect. IMG-12	■	0.508	Temp 2	IMG-13	0.456	Sentinel VH - IMG-28	■	0.298
B1 Reflect. IMG-10	■	0.491	Temp 2	IMG-7	0.455	Sentinel VH - IMG-57	■	0.297
B7 Reflect. IMG-11	■	0.487	Temp 2	IMG-9	0.444	Sentinel VH - IMG-16	■	0.296
B7 Reflect. IMG-3	■	0.486	Temp 1	IMG-13	0.439	Sentinel VH - IMG-27	■	0.295
B5 Reflect. IMG-2	■	0.476	Albedo	IMG-3	0.427	Sentinel VH - IMG-53	■	0.295
B5 Reflect. IMG-9	■	0.473	Albedo	IMG-1	0.422	Sentinel VH - IMG-25	■	0.295
B5 Reflect. IMG-10	■	0.456	Temp 1	IMG-2	0.420	Sentinel VV - IMG-32	■	0.292
B3 Reflect. IMG-10	■	0.455	Albedo	IMG-8	0.416	Sentinel VV - IMG-30	■	0.292
B6 Reflect. IMG-3	■	0.451	Albedo	IMG-9	0.412	Sentinel VH - IMG-29	■	0.291
B6 Reflect. IMG-8	■	0.447	Temp 2	IMG-8	0.398	Sentinel VH - IMG-40	■	0.289
B5 Reflect. IMG-11	■	0.446	NDWI	IMG-7	0.394	Sentinel VH - IMG-39	■	0.287
B7 Reflect. IMG-8	■	0.446	Temp 2	IMG-3	0.387	Sentinel VV - IMG-45	■	0.285
B3 Reflect. IMG-11	■	0.443	NDWI	IMG-10	0.374	Sentinel VH - IMG-21	■	0.285
B6 Reflect. IMG-11	■	0.442	Temp 1	IMG-1	0.373	Sentinel VV - IMG-53	■	0.284
B5 Reflect. IMG-12	■	0.442	Temp 1	IMG-1	0.355	Sentinel VV - IMG-41	■	0.283
B3 Reflect. IMG-2	■	0.440	Temp 1	IMG-8	0.341	Sentinel VV - IMG-39	■	0.283
B5 Reflect. IMG-7	■	0.437	NDWI	IMG-9	0.341	Sentinel VV - IMG-31	■	0.281
B2 Reflect. IMG-8	■	0.428	NDWI	IMG-12	0.337	Sentinel VH - IMG-56	■	0.275
B1 Reflect. IMG-13	■	0.424	Temp 1	IMG-3	0.325	Sentinel VH - IMG-58	■	0.275
B2 Reflect. IMG-1	■	0.421	NDWI	IMG-11	0.318	Sentinel VV - IMG-52	■	0.274

B2 Reflect.	IMG-12	0.421	NDWI	IMG-8	0.306	Sentinel VH - IMG-17	0.273
B6 Reflect.	IMG-9	0.417	NDWI	IMG-13	0.304	Sentinel VV - IMG-43	0.273
B2 Reflect.	IMG-2	0.413	Temp 1	IMG-14	0.278	Sentinel VV - IMG-50	0.272
B5 Reflect.	IMG-13	0.395	NDWI	IMG-14	0.263	Sentinel VH - IMG-36	0.271
B1 Reflect.	IMG-3	0.384	NDWI	IMG-15	0.258	Sentinel VV - IMG-21	0.271
B3 Reflect.	IMG-3	0.381	Temp 2	IMG-14	0.211	Sentinel VH - IMG-32	0.270
B6 Reflect.	IMG-13	0.381	Temp 2	IMG-15	0.181	Sentinel VV - IMG-54	0.269
B1 Reflect.	IMG-9	0.376	NDVI	IMG-14	0.178	Sentinel VV - IMG-58	0.266
B7 Reflect.	IMG-9	0.371	Albedo	IMG-15	0.173	Sentinel VV - IMG-29	0.266
B1 Reflect.	IMG-1	0.369	Temp 1	IMG-15	0.164	Sentinel VV - IMG-24	0.264
B7 Reflect.	IMG-13	0.362	NDVI	IMG-15	0.146	Sentinel VV - IMG-57	0.264
B3 Reflect.	IMG-8	0.359	Albedo	IMG-14	0.096	Sentinel VV - IMG-40	0.263
B3 Reflect.	IMG-13	0.358				Sentinel VH - IMG-55	0.262
B3 Reflect.	IMG-1	0.353				Sentinel VV - IMG-16	0.260
B3 Reflect.	IMG-12	0.353				Sentinel VV - IMG-48	0.257
B5 Reflect.	IMG-8	0.350				Sentinel VV - IMG-35	0.256
B5 Reflect.	IMG-3	0.348				Sentinel VV - IMG-22	0.255
B1 Reflect.	IMG-8	0.346				Sentinel VH - IMG-35	0.255
B2 Reflect.	IMG-11	0.337				Sentinel VH - IMG-4	0.255
B1 Reflect.	IMG-12	0.335				Sentinel VH - IMG-54	0.255
B1 Reflect.	IMG-2	0.334				Sentinel VV - IMG-34	0.252
B1 Reflect.	IMG-11	0.331				Sentinel VH - IMG-52	0.252
B1 Reflect.	IMG-15	0.326				Sentinel VH - IMG-18	0.250
B1 Reflect.	IMG-14	0.323				Sentinel VV - IMG-27	0.249
B7 Reflect.	IMG-15	0.304				Sentinel VH - IMG-20	0.249
B5 Reflect.	IMG-15	0.303				Sentinel VV - IMG-44	0.249
B3 Reflect.	IMG-15	0.302				Sentinel VV - IMG-55	0.249
B4 Reflect.	IMG-14	0.302				Sentinel VV - IMG-36	0.249
B7 Reflect.	IMG-14	0.301				Sentinel VV - IMG-4	0.248
B5 Reflect.	IMG-14	0.296				Sentinel VV - IMG-25	0.247
B2 Reflect.	IMG-15	0.278				Sentinel VV - IMG-19	0.245
B6 Reflect.	IMG-15	0.278				Sentinel VH - IMG-6	0.240
B6 Reflect.	IMG-14	0.274				Sentinel VH - IMG-19	0.238
B4 Reflect.	IMG-15	0.273				Sentinel VV - IMG-37	0.238
B3 Reflect.	IMG-14	0.270				Sentinel VV - IMG-6	0.237
B2 Reflect.	IMG-14	0.262				Sentinel VV - IMG-56	0.233
						Sentinel VV - IMG-26	0.226
						Sentinel VV - IMG-46	0.219
						Sentinel VV - IMG-47	0.215
						Sentinel VV - IMG-20	0.211
						Sentinel VV - IMG-28	0.210
						Sentinel VV - IMG-18	0.210
						Sentinel VV - IMG-17	0.205
						Sentinel VV - IMG-5	0.203
						Sentinel VH - IMG-5	0.182
						Sentinel VV - IMG-23	0.175
Average		0.491	Average		0.440	Average	0.295
Standard Deviation		0.198	Standard Deviation		0.160	Standard Deviation	0.154

Appendix 6. Summary of test inputs for classification comparisons. Tests determined by Log-Normal values and by RF determined predictor importance values (blue text). Tests determined by the hybrid approach (green text). Tests determined by the holistic approach (purple text).

1	Top 120 LN - Determined Variables
2	Top 115 LN - Determined Variables
3	Top 110 LN - Determined Variables
4	Top 105 LN - Determined Variables
5	Top 100 LN - Determined Variables
6	Top 95 LN - Determined Variables
7	Top 90 LN - Determined Variables
8	Top 85 LN - Determined Variables
9	Top 80 LN - Determined Variables
10	Top 75 LN - Determined Variables
11	Top 70 LN - Determined Variables
12	Top 65 LN - Determined Variables
13	Top 60 LN - Determined Variables
14	Top 55 LN - Determined Variables
15	Top 50 LN - Determined Variables
16	Top 45 LN - Determined Variables
17	Top 40LN - Determined Variables
18	Top 35 LN - Determined Variables
19	Top 30 LN - Determined Variables
20	Top 25 LN - Determined Variables
21	Top 20 LN - Determined Variables
22	Top 15 LN - Determined Variables
23	Top 10 LN - Determined Variables
24	Top 5 LN - Determined Variables
25	Bottom 120 LN - Determined Variables
26	Bottom 115 LN - Determined Variables
27	Bottom 110 LN - Determined Variables
28	Bottom 105 LN - Determined Variables
29	Bottom 100 LN - Determined Variables
30	Bottom 95 LN - Determined Variables
31	Bottom 90 LN - Determined Variables
32	Bottom 85 LN - Determined Variables
33	Bottom 80 LN - Determined Variables
34	Bottom 75 LN - Determined Variables
35	Bottom 70 LN - Determined Variables
36	Bottom 65 LN - Determined Variables
37	Bottom 60 LN - Determined Variables
38	Bottom 55 LN - Determined Variables
39	Bottom 50 LN - Determined Variables
40	Bottom 45 LN - Determined Variables
41	Bottom 40 LN - Determined Variables
42	Bottom 35 LN - Determined Variables
43	Bottom 30 LN - Determined Variables
44	Bottom 25 LN - Determined Variables
45	Bottom 20 LN - Determined Variables
46	Bottom 15 LN - Determined Variables

47	Bottom 10 LN - Determined Variables
48	Bottom 5 LN - Determined Variables
49	Top 120 RF - Determined Variables
50	Top 115 RF - Determined Variables
51	Top 110 RF - Determined Variables
52	Top 105 RF - Determined Variables
53	Top 100 RF - Determined Variables
54	Top 95 RF - Determined Variables
55	Top 90 RF - Determined Variables
56	Top 85 RF - Determined Variables
57	Top 80 RF - Determined Variables
58	Top 75 RF - Determined Variables
59	Top 70 RF - Determined Variables
60	Top 65 RF - Determined Variables
61	Top 60 RF - Determined Variables
62	Top 55 RF - Determined Variables
63	Top 50 RF - Determined Variables
64	Top 45 RF - Determined Variables
65	Top 40RF - Determined Variables
66	Top 35 RF - Determined Variables
67	Top 30 RF - Determined Variables
68	Top 25 RF - Determined Variables
69	Top 20 RF - Determined Variables
70	Top 15 RF - Determined Variables
71	Top 10 RF - Determined Variables
72	Top 5 RF - Determined Variables
73	Bottom 120 RF - Determined Variables
74	Bottom 115 RF - Determined Variables
75	Bottom 110 RF - Determined Variables
76	Bottom 105 RF - Determined Variables
77	Bottom 100 RF - Determined Variables
78	Bottom 95 RF - Determined Variables
79	Bottom 90 RF - Determined Variables
80	Bottom 85 RF - Determined Variables
81	Bottom 80 RF - Determined Variables
82	Bottom 75 RF - Determined Variables
83	Bottom 70 RF - Determined Variables
84	Bottom 65 RF - Determined Variables
85	Bottom 60 RF - Determined Variables
86	Bottom 55 RF - Determined Variables
87	Bottom 50 RF - Determined Variables
88	Bottom 45 RF - Determined Variables
89	Bottom 40 RF - Determined Variables
90	Bottom 35 RF - Determined Variables
91	Bottom 30 RF - Determined Variables
92	Bottom 25 RF - Determined Variables
93	Bottom 20 RF - Determined Variables
94	Bottom 15 RF - Determined Variables
95	Bottom 10 RF - Determined Variables
96	Bottom 5 RF - Determined Variables

97	Hybrid Top 25 Percent - RF
98	Hybrid Top 20 Percent - RF
99	Hybrid Top 15 Percent - RF
100	Hybrid Top 10 Percent - RF
101	Hybrid Top 25 Percent - LN
102	Hybrid Top 20 Percent - LN
103	Hybrid Top 15 Percent - LN
104	Hybrid Top 10 Percent - LN
105	Hybrid Bottom 25 Percent - RF
106	Hybrid Bottom 20 Percent - RF
107	Hybrid Bottom 15 Percent - RF
108	Hybrid Bottom 10 Percent - RF
109	Hybrid Bottom 25 Percent - LN
110	Hybrid Bottom 20 Percent - LN
111	Hybrid Bottom 15 Percent - LN
112	Hybrid Bottom 10 Percent - LN
113	All Bands, Radar (VV), Temp 1, Slope, All Seasons
114	All Bands, Radar (VH), Temp 1, Slope, All Seasons
115	All Bands, Radar (VV,VH), Temp 1, Slope, All Seasons
116	All Bands, Radar (VV), Temp 1, All Seasons
117	All Bands, Radar (VH), Temp 1, All Seasons
118	All Bands,Temp 1, All Seasons
119	All Bands, Radar(VV) All Seasons
120	All Bands, Radar(VH) All Seasons
121	All Bands All Seasons
122	All Bands, Radar (VV), Temp 2, Slope, All Seasons
123	All Bands, Radar (VH), Temp 2, Slope, All Seasons
124	All Bands, Radar (VV,VH), Temp 2, Slope, All Seasons
125	All Bands, Radar (VV), Temp 2, All Seasons
126	All Bands, Radar (VH), Temp 2, All Seasons
127	All Bands, Slope, All Seasons
128	All Bands,Temp 2, All Seasons
129	Bands 1-7, Temp2, DEM, Spring
130	Bands 1-7, Temp2, Slope, Spring
131	Bands 1-7, Temp2, Slope, Fall, Radar (VV)
132	Bands 1-7, Temp2, Slope, Fall, Radar (VH)
133	Bands 1-7, DEM (Summer)
134	Bands 1-7 (Spring and Summer), Temp2 (Spring and Summer), Slope, Radar (VV - Fall)
135	Bands 1-7 (Spring and Summer), Temp2 (Spring and Summer), Slope, Radar (VH - Fall)
136	All Metrics, Radar (VV), Temp 1, Slope, All Seasons
137	All Metrics, Radar (VH), Temp 1, Slope, All Seasons
138	All Metrics, Radar (VV,VH), Temp 1, Slope, All Seasons
139	All Metrics, Radar (VV), Temp 1, All Seasons
140	All Metrics, Radar (VH), Temp 1, All Seasons
141	All Metrics,Temp 1, All Seasons
142	All Metrics, Radar(VV) All Seasons
143	All Metrics, Radar(VH) All Seasons
144	All Metrics All Seasons
145	All Metrics, Radar (VV), Temp 2, Slope, All Seasons
146	All Metrics, Radar (VH), Temp 2, Slope, All Seasons

147	All Metrics, Radar (VV,VH), Temp 2, Slope, All Seasons
148	All Metrics, Radar (VV), Temp 2, All Seasons
149	All Metrics, Radar (VH), Temp 2, All Seasons
150	All Metrics, Slope, All Seasons
151	All Metrics,Temp 2, All Seasons
152	Metrics 1-7, Temp2, DEM, Spring
153	Metrics 1-7, Temp2, Slope, Spring
154	Metrics 1-7, Temp2, Slope, Fall, Radar (VV)
155	Metrics 1-7, Temp2, Slope, Fall, Radar (VH)
156	Metrics 1-7, DEM (Summer)
157	Metrics 1-7 (Spring and Summer), Temp2 (Spring and Summer), Slope, Radar (VV - Fall)
158	Metrics 1-7 (Spring and Summer), Temp2 (Spring and Summer), Slope, Radar (VH - Fall)
159	All variables
160	Band 1-7 (Spring)
161	Band 1-7 (Summer)
162	Band 1-7 (Fall)
163	Band 1-7, Temp 1 (Spring)
164	Band 1-7, Temp 1 (Summer)
165	Band 1-7, Temp 1 (Fall)
166	Band 1-7, Temp 2 (Spring)
167	Band 1-7, Temp 2 (Summer)
168	Band 1-7, Temp 2 (Fall)
169	Band 1-7, Radar-VV (Summer)
170	Band 1-7, Radar-VH (Summer)
171	Band 1-7, Radar-VV (Summer2)
172	Band 1-7, Radar-VH (Summer2)
173	Band 1-7, Radar-VV (Fall)
174	Band 1-7, Radar-VH (Fall)
175	Band 1-7, DEM (Spring)
176	Band 1-7, DEM (Summer)
177	Band 1-7, DEM (Fall)
178	Band 1-7, Slope (Spring)
179	Band 1-7, Slope (Summer)
180	Band 1-7, Slope (Fall)
181	Band 1-7, DEM, Slope (Spring)
182	Band 1-7, DEM, Slope (Summer)
183	Band 1-7, DEM, Slope (Fall)
184	Band 1-7, Temp, Radar-VV (Summer)
185	Band 1-7, Temp, Radar-VH (Summer)
186	Band 1-7, Temp, Radar-VV (Summer2)
187	Band 1-7, Temp, Radar-VH (Summer2)
188	Band 1-7, Temp, Radar-VV (Fall)
189	Band 1-7, Temp, Radar-VH (Fall)
190	Band 1-7, Temp2, Radar-VV (Summer)
191	Band 1-7, Temp2, Radar-VH (Summer)
192	Band 1-7, Temp2, Radar-VV (Summer2)
193	Band 1-7, Temp2, Radar-VH (Summer2)
194	Band 1-7, Temp2, Radar-VV (Fall)
195	Band 1-7, Temp2, Radar-VH (Fall)
196	Band 1-7, Temp, Radar-VV (Summer), DEM

197	Band 1-7, Temp, Radar-VH (Summer), DEM
198	Band 1-7, Temp, Radar-VV (Summer2), DEM
199	Band 1-7, Temp, Radar-VH (Summer2), DEM
200	Band 1-7, Temp, Radar-VV (Fall), DEM
201	Band 1-7, Temp, Radar-VH (Fall), DEM
202	Band 1-7, Temp2, Radar-VV (Summer), DEM
203	Band 1-7, Temp2, Radar-VH (Summer), DEM
204	Band 1-7, Temp2, Radar-VV (Summer2), DEM
205	Band 1-7, Temp2, Radar-VH (Summer2), DEM
206	Band 1-7, Temp2, Radar-VV (Fall), DEM
207	Band 1-7, Temp2, Radar-VH (Fall), DEM
208	Band 1-7, Temp, Radar-VV (Summer), Slope
209	Band 1-7, Temp, Radar-VH (Summer), Slope
210	Band 1-7, Temp, Radar-VV (Summer2), Slope
211	Band 1-7, Temp, Radar-VH (Summer2), Slope
212	Band 1-7, Temp, Radar-VV (Fall), Slope
213	Band 1-7, Temp, Radar-VH (Fall), Slope
214	Band 1-7, Temp2, Radar-VV (Summer), Slope
215	Band 1-7, Temp2, Radar-VH (Summer), Slope
216	Band 1-7, Temp2, Radar-VV (Summer2), Slope
217	Band 1-7, Temp2, Radar-VH (Summer2), Slope
218	Band 1-7, Temp2, Radar-VV (Fall), Slope
219	Band 1-7, Temp2, Radar-VH (Fall), Slope
220	All Bands All Seasons
221	All Bands, Radar(VV) All Seasons
222	All Bands, Temp1, All Seasons
223	All Bands, Slope, All Seasons
224	All Bands, Radar (VV), Temp1, All Seasons
225	All Bands, Radar (VV), Temp1 Slope, All Seasons
226	All Bands, Radar(VH) All Seasons
227	All Bands, Radar (VH), Temp1, All Seasons
228	All Bands, Radar (VH), Temp1, Slope, All Seasons
229	All Bands, Radar (VV,VH), Temp1, Slope, All Seasons
230	All Bands, Temp2, All Seasons
231	All Bands, Radar (VV), Temp2, All Seasons
232	All Bands, Radar (VV), Temp2 Slope, All Seasons
233	All Bands, Radar (VH), Temp2, All Seasons
234	All Bands, Radar (VH), Temp2, Slope, All Seasons
235	All Bands, Radar (VV,VH), Temp2, Slope, All Seasons
236	NDVI, NDWI, Albedo (Spring)
237	NDVI, NDWI, Albedo (Summer)
238	NDVI, NDWI, Albedo (Fall)
239	NDVI, NDWI, Albedo, Temp 1 (Spring)
240	NDVI, NDWI, Albedo, Temp 1 (Summer)
241	NDVI, NDWI, Albedo, Temp 1 (Fall)
242	NDVI, NDWI, Albedo, Temp 2 (Spring)
243	NDVI, NDWI, Albedo, Temp 2 (Summer)
244	NDVI, NDWI, Albedo, Temp 2 (Fall)
245	NDVI, NDWI, Albedo, Radar-VV (Summer)
246	NDVI, NDWI, Albedo, Radar-VH (Summer)

247	NDVI, NDWI, Albedo, Radar-VV (Summer2)
248	NDVI, NDWI, Albedo, Radar-VH (Summer2)
249	NDVI, NDWI, Albedo, Radar-VV (Fall)
250	NDVI, NDWI, Albedo, Radar-VH (Fall)
251	NDVI, NDWI, Albedo, DEM (Spring)
252	NDVI, NDWI, Albedo, DEM (Summer)
253	NDVI, NDWI, Albedo, DEM (Fall)
254	NDVI, NDWI, Albedo, Slope (Spring)
255	NDVI, NDWI, Albedo, Slope (Summer)
256	NDVI, NDWI, Albedo, Slope (Fall)
257	NDVI, NDWI, Albedo, DEM, Slope (Spring)
258	NDVI, NDWI, Albedo, DEM, Slope (Summer)
259	NDVI, NDWI, Albedo, DEM, Slope (Fall)
260	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer)
261	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer)
262	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2)
263	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2)
264	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall)
265	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall)
266	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer)
267	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer)
268	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2)
269	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2)
270	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall)
271	NDVI, NDWI, Albedo, Temp2, Radar-VH (Fall)
272	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer), DEM
273	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer), DEM
274	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2), DEM
275	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2), DEM
276	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall), DEM
277	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall), DEM
278	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer), DEM
279	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer), DEM
280	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2), DEM
281	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2), DEM
282	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall), DEM
283	NDVI, NDWI, Albedo, Temp2, Radar-VH (Fall), DEM
284	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer), Slope
285	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer), Slope
286	NDVI, NDWI, Albedo, Temp, Radar-VV (Summer2), Slope
287	NDVI, NDWI, Albedo, Temp, Radar-VH (Summer2), Slope
288	NDVI, NDWI, Albedo, Temp, Radar-VV (Fall), Slope
289	NDVI, NDWI, Albedo, Temp, Radar-VH (Fall), Slope
290	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer), Slope
291	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer), Slope
292	NDVI, NDWI, Albedo, Temp2, Radar-VV (Summer2), Slope
293	NDVI, NDWI, Albedo, Temp2, Radar-VH (Summer2), Slope
294	NDVI, NDWI, Albedo, Temp2, Radar-VV (Fall), Slope
295	NDVI, NDWI, Albedo, Temp2, Radar-VH (Fall), Slope
296	NDVI, NDWI, Albedo All Seasons

297	NDVI, NDWI, Albedo, Radar(VV) All Seasons
298	NDVI, NDWI, Albedo, Temp1, All Seasons
299	NDVI, NDWI, Albedo, Slope, All Seasons
300	NDVI, NDWI, Albedo, Radar (VV), Temp1, All Seasons
301	NDVI, NDWI, Albedo, Radar (VV), Temp1, Slope, All Seasons
302	NDVI, NDWI, Albedo, Radar (VV,VH), Temp1, Slope, All Seasons
303	NDVI, NDWI, Albedo, Radar(VH) All Seasons
304	NDVI, NDWI, Albedo, Temp2, All Seasons
305	NDVI, NDWI, Albedo, Radar (VH), Temp1, All Seasons
306	NDVI, NDWI, Albedo, Radar (VH), Temp1, Slope, All Seasons
307	NDVI, NDWI, Albedo, Temp2, All Seasons
308	NDVI, NDWI, Albedo, Radar (VV), Temp2, All Seasons
309	NDVI, NDWI, Albedo, Radar (VV), Temp2 Slope, All Seasons
310	NDVI, NDWI, Albedo, Radar (VH), Temp2, All Seasons
311	NDVI, NDWI, Albedo, Radar (VH), Temp2, Slope, All Seasons
312	NDVI, NDWI, Albedo, Radar (VV,VH), Temp2, Slope, All Seasons

Appendix 7. Summary of top 20 (A) and bottom 20 (B) classification inputs for the standard classification approach. Top 20 (C) and bottom 20 (D) classification inputs for the Case – 1 Hierarchical classification approach. Top 20 (E) and bottom 20 (F) classification inputs for the Case – 2 Hierarchical classification approach.

A)

Standard			
	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
1	All Bands, Radar (VH), Temp 2, Slope, All Seasons	0.8006	0.8543
2	Bands 1-7, Temp2, Slope, Fall, Radar (VV)	0.7960	0.8388
3	All Bands, Radar (VV), Temp 1, Slope, All Seasons	0.7926	0.8489
4	All Bands,Temp 2, All Seasons	0.7923	0.8466
5	All Bands, Radar (VH), Temp 1, Slope, All Seasons	0.7900	0.8509
6	Top 55 RF - Determined Variables	0.7896	0.8290
7	All Bands, Radar(VH) All Seasons	0.7890	0.8439
8	All Bands, Radar (VV,VH), Temp 2, Slope, All Seasons	0.7879	0.8431
9	Top 75 RF - Determined Variables	0.7873	0.8272
10	Top 60 RF - Determined Variables	0.7869	0.8268
11	All Bands, Radar (VV), Temp 2, Slope, All Seasons	0.7865	0.8431
12	Top 120 RF - Determined Variables	0.7864	0.8292
13	All Bands All Seasons	0.7860	0.8419
14	Top 35 RF - Determined Variables	0.7848	0.8285
15	Top 95 RF - Determined Variables	0.7848	0.8254
16	Top 115 RF - Determined Variables	0.7848	0.8277
17	Top 120 LN - Determined Variables	0.7844	0.8263
18	All Variables	0.7843	0.8283
19	All Bands, Radar (VV,VH), Temp 1, Slope, All Seasons	0.7842	0.8427
20	Top 40RF - Determined Variables	0.7837	0.8258
	Naïve Bayes	**Producer Accuracy**	**Kappa Value**
1	Top 50 RF - Determined Variables	0.7845	0.8210
2	Top 55 RF - Determined Variables	0.7841	0.8189
3	Top 70 RF - Determined Variables	0.7812	0.8184
4	Top 65 RF - Determined Variables	0.7790	0.8149
5	Top 60 RF - Determined Variables	0.7765	0.8132
6	All Metrics,Temp 2, All Seasons	0.7778	0.8197

7	All Metrics,Temp 1, All Seasons	0.7707	0.8158
8	Top 80 RF - Determined Variables	0.7675	0.8090
9	All Metrics, Radar (VV), Temp 2, Slope, All Seasons	0.7706	0.8080
10	Top 75 RF - Determined Variables	0.7718	0.8140
11	Top 110 RF - Determined Variables	0.7659	0.8150
12	Top 45 RF - Determined Variables	0.7669	0.8176
13	Top 100 RF - Determined Variables	0.7667	0.8107
14	Metrics 1-7 (Spring and Summer), Temp2 (Spring and Summer), Slope, Radar (VH - Fall)	0.7692	0.8103
15	Metrics 1-7 (Spring and Summer), Temp2 (Spring and Summer), Slope, Radar (VV - Fall)	0.7700	0.8106
16	Top 95 RF - Determined Variables	0.7680	0.8083
17	All Metrics, Radar (VV), Temp 2, All Seasons	0.7654	0.8097
18	Top 120 RF - Determined Variables	0.7619	0.8078
19	Top 90 RF - Determined Variables	0.7646	0.8087
20	Top 35 RF - Determined Variables	0.7632	0.8022

B)

Standard			
	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
293	Hybrid Bottom 15 Percent - RF	0.4318	0.2675
294	Bottom 65 LN - Determined Variables	0.4317	0.2642
295	Hybrid Bottom 10 Percent - LN	0.4304	0.3663
296	Bottom 20 RF - Determined Variables	0.4298	0.2308
297	Bottom 55 LN - Determined Variables	0.4283	0.2569
298	Bottom 10 RF - Determined Variables	0.4234	0.2280
299	Bottom 60 LN - Determined Variables	0.4186	0.2208
300	Bottom 15 RF - Determined Variables	0.4179	0.2817
301	Bottom 50 LN - Determined Variables	0.3993	0.2629
302	Bottom 45 LN - Determined Variables	0.3921	0.2121
303	Bottom 40 LN - Determined Variables	0.3884	0.2188
304	Hybrid Bottom 10 Percent - RF	0.3811	0.2117
305	Bottom 35 LN - Determined Variables	0.3683	0.2078
306	Bottom 30 LN - Determined Variables	0.3335	0.1818
307	Bottom 25 LN - Determined Variables	0.3160	0.1762
308	Bottom 5 RF - Determined Variables	0.2966	0.1206

309	Bottom 20 LN - Determined Variables	0.2928	0.1371
310	Bottom 15 LN - Determined Variables	0.2857	0.1641
311	Bottom 10 LN - Determined Variables	0.2778	0.1314
312	Bottom 5 LN - Determined Variables	0.2728	0.1082
	Naïve Bayes	**Producer Accuracy**	**Kappa Value**
293	Bottom 50 LN - Determined Variables	0.3573	0.3735
294	Hybrid Bottom 15 Percent - LN	0.3572	0.3724
295	Bottom 30 RF - Determined Variables	0.3464	0.3697
296	Bottom 45 LN - Determined Variables	0.3455	0.3623
297	Bottom 25 RF - Determined Variables	0.3451	0.3567
298	Bottom 40 LN - Determined Variables	0.3408	0.3451
299	Hybrid Bottom 15 Percent - RF	0.3403	0.3435
300	Bottom 35 LN - Determined Variables	0.3383	0.3367
301	Hybrid Bottom 10 Percent - LN	0.3374	0.3304
302	Bottom 20 RF - Determined Variables	0.3354	0.3263
303	Bottom 15 RF - Determined Variables	0.3307	0.3227
304	Bottom 30 LN - Determined Variables	0.3270	0.3088
305	Bottom 10 RF - Determined Variables	0.3253	0.3020
306	Hybrid Bottom 10 Percent - RF	0.3153	0.2927
307	Bottom 25 LN - Determined Variables	0.3137	0.2863
308	Bottom 20 LN - Determined Variables	0.3034	0.2755
309	Bottom 5 RF - Determined Variables	0.2982	0.2555
310	Bottom 15 LN - Determined Variables	0.2957	0.2531
311	Bottom 10 LN - Determined Variables	0.2835	0.2304
312	Bottom 5 LN - Determined Variables	0.2751	0.2098

C)

Hierarchical - 1			
	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
1	Rank_18_Impact	0.8720	0.8732
2	Rank_17_Impact	0.8594	0.8537
3	Rank_20_Impact	0.8587	0.8568
4	Rank_16_Impact	0.8574	0.8628

5	Rank_22_Impact	0.8591	0.8684
6	Rank_21_Impact	0.8565	0.8817
7	Rank_27_Impact	0.8572	0.8623
8	Rank_13_Impact	0.8515	0.8705
9	Rank__8_Impact	0.8487	0.8524
10	Rank__9_Impact	0.8432	0.8515
11	Rank__2_Impact	0.8439	0.8783
12	Rank_15_Impact	0.8443	0.8491
13	Rank_47_Impact	0.8385	0.8351
14	Rank__4_Impact	0.8410	0.8740
15	Rank_42_Impact	0.8401	0.8379
16	Rank__3_Impact	0.8345	0.8577
17	Rank__1_Impact	0.8335	0.8571
18	Rank_46_Impact	0.8345	0.8414
19	Rank_45_Impact	0.8254	0.8348
20	Rank_34_Impact	0.8243	0.8504
	Naïve Bayes	**Producer Accuracy**	**Kappa Value**
1	Rank__6_Impact	0.8680	0.8859
2	Rank__7_Impact	0.8651	0.8834
3	Rank__5_Impact	0.8686	0.8826
4	Rank_13_Impact	0.8605	0.8807
5	Rank_14_Impact	0.8608	0.8786
6	Rank__8_Impact	0.8570	0.8807
7	Rank_26_Impact	0.8544	0.8732
8	Rank_10_Impact	0.8595	0.8797
9	Rank_20_Impact	0.8547	0.8751
10	Rank_12_Impact	0.8554	0.8787
11	Rank_17_Impact	0.8582	0.8766
12	Rank_29_Impact	0.8551	0.8744
13	Rank_35_Impact	0.8522	0.8745
14	Rank__9_Impact	0.8521	0.8733
15	Rank_31_Impact	0.8564	0.8764
16	Rank__4_Impact	0.8504	0.8742
17	Rank_24_Impact	0.8543	0.8727

18	Rank_23_Impact	0.8512	0.8743
19	Rank_15_Impact	0.8526	0.8769
20	Rank__2_Impact	0.8506	0.8759

D)

Hierarchical - 1			
	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
77	Rank_72_Impact	0.3510	0.3737
78	Rank_80_Impact	0.3490	0.3374
79	Rank_79_Impact	0.3505	0.3438
80	Rank_76_Impact	0.3439	0.3406
81	Rank_78_Impact	0.3494	0.3519
82	Rank_82_Impact	0.3450	0.3462
83	Rank_83_Impact	0.3397	0.3251
84	Rank_87_Impact	0.3366	0.3008
85	Rank_85_Impact	0.3360	0.3179
86	Rank_84_Impact	0.3265	0.3130
87	Rank_86_Impact	0.3190	0.2960
88	Rank_89_Impact	0.3200	0.2921
89	Rank_88_Impact	0.3208	0.3034
90	Rank_90_Impact	0.3062	0.2835
91	Rank_92_Impact	0.3042	0.2798
92	Rank_91_Impact	0.3016	0.2766
93	Rank_93_Impact	0.2891	0.2604
94	Rank_95_Impact	0.2744	0.2187
95	Rank_94_Impact	0.2653	0.2372
96	Rank_96_Impact	0.2632	0.2162
	Naïve Bayes	**Producer Accuracy**	**Kappa Value**
77	Rank_75_Impact	0.6716	0.6905
78	Rank_81_Impact	0.6548	0.6949
79	Rank_78_Impact	0.6563	0.6493
80	Rank_84_Impact	0.6440	0.6413

81	Rank_80_Impact	0.6478	0.6908
82	Rank_82_Impact	0.6361	0.6749
83	Rank_83_Impact	0.6338	0.6307
84	Rank_72_Impact	0.6361	0.5623
85	Rank_85_Impact	0.6283	0.6391
86	Rank_86_Impact	0.6292	0.6385
87	Rank_89_Impact	0.6288	0.6654
88	Rank_90_Impact	0.6165	0.6310
89	Rank_91_Impact	0.6087	0.5498
90	Rank_87_Impact	0.6139	0.6362
91	Rank_88_Impact	0.6079	0.6254
92	Rank_93_Impact	0.6046	0.6547
93	Rank_92_Impact	0.5953	0.5860
94	Rank_94_Impact	0.5865	0.5185
95	Rank_95_Impact	0.5575	0.5685
96	Rank_96_Impact	0.4217	0.4104

E)

Hierarchical - 2			
	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
1	Rank__1_Impact	0.9194	0.9272
2	Rank__2_Impact	0.9151	0.9352
3	Rank__3_Impact	0.9053	0.9218
4	Rank__4_Impact	0.8945	0.9135
5	Rank__7_Impact	0.8921	0.9141
6	Rank__5_Impact	0.8912	0.9066
7	Rank__6_Impact	0.8884	0.9128
8	Rank__9_Impact	0.8810	0.9026
9	Rank_16_Impact	0.8606	0.8887
10	Rank_10_Impact	0.8610	0.8898
11	Rank_12_Impact	0.8611	0.8759
12	Rank_14_Impact	0.8591	0.8815
13	Rank__8_Impact	0.8516	0.8977

14	Rank_11_Impact	0.8499	0.8891
15	Rank_15_Impact	0.8522	0.8814
16	Rank_17_Impact	0.8486	0.9017
17	Rank_21_Impact	0.8471	0.8799
18	Rank_20_Impact	0.8458	0.8625
19	Rank_19_Impact	0.8427	0.8829
20	Rank_25_Impact	0.8391	0.8798
	Naïve Bayes	**Producer Accuracy**	**Kappa Value**
1	Rank__1_Impact	0.9119	0.9413
2	Rank__2_Impact	0.9024	0.9354
3	Rank__3_Impact	0.8975	0.9362
4	Rank__7_Impact	0.8986	0.9262
5	Rank__4_Impact	0.8919	0.9278
6	Rank__5_Impact	0.8766	0.9216
7	Rank__6_Impact	0.8593	0.9120
8	Rank__9_Impact	0.8305	0.8934
9	Rank__8_Impact	0.8313	0.8957
10	Rank_10_Impact	0.8211	0.8887
11	Rank_11_Impact	0.8073	0.8807
12	Rank_12_Impact	0.8043	0.8799
13	Rank_13_Impact	0.7882	0.8739
14	Rank_15_Impact	0.7774	0.8617
15	Rank_14_Impact	0.7766	0.8657
16	Rank_16_Impact	0.7685	0.8610
17	Rank_20_Impact	0.7662	0.8448
18	Rank_17_Impact	0.7586	0.8476
19	Rank_18_Impact	0.7561	0.8491
20	Rank_19_Impact	0.7486	0.8447

F)

Hierarchical - 2			
	Random Forest Classifier	**Producer Accuracy**	**Kappa Value**
77	Rank_78_Impact	0.3629	0.3514
78	Rank_80_Impact	0.3501	0.3283
79	Rank_79_Impact	0.3465	0.3336
80	Rank_77_Impact	0.3428	0.3348
81	Rank_81_Impact	0.3311	0.3167
82	Rank_83_Impact	0.3120	0.2952
83	Rank_82_Impact	0.3137	0.2897
84	Rank_85_Impact	0.3087	0.2613
85	Rank_87_Impact	0.3062	0.2698
86	Rank_84_Impact	0.3015	0.2767
87	Rank_89_Impact	0.2863	0.2390
88	Rank_90_Impact	0.2842	0.2405
89	Rank_86_Impact	0.2768	0.2374
90	Rank_88_Impact	0.2726	0.2339
91	Rank_92_Impact	0.2643	0.2174
92	Rank_91_Impact	0.2573	0.2113
93	Rank_93_Impact	0.2531	0.1966
94	Rank_94_Impact	0.2392	0.1855
95	Rank_95_Impact	0.2204	0.1461
96	Rank_96_Impact	0.2182	0.1340
	Naïve Bayes	**Producer Accuracy**	**Kappa Value**
77	Rank_77_Impact	0.3310	0.2652
78	Rank_78_Impact	0.3241	0.2504
79	Rank_80_Impact	0.3174	0.2455
80	Rank_79_Impact	0.3120	0.2389
81	Rank_82_Impact	0.3077	0.2385
82	Rank_81_Impact	0.2974	0.2102
83	Rank_83_Impact	0.2906	0.1960
84	Rank_84_Impact	0.2777	0.1794
85	Rank_86_Impact	0.2743	0.1920

86	Rank_87_Impact	0.2698	0.1777
87	Rank_85_Impact	0.2630	0.1754
88	Rank_89_Impact	0.2557	0.1483
89	Rank_90_Impact	0.2424	0.1521
90	Rank_92_Impact	0.2447	0.1425
91	Rank_88_Impact	0.2391	0.1329
92	Rank_91_Impact	0.2335	0.1239
93	Rank_93_Impact	0.2221	0.1395
94	Rank_94_Impact	0.2014	0.1000
95	Rank_95_Impact	0.1929	0.1045
96	Rank_96_Impact	0.1547	0.0277