HARNESSING MACHINE LEARNING TECHNIQUES FOR LARGE-SCALE MAPPING OF INLAND AQUACULTURE WATERBODIES IN BANGLADESH

By

Hannah Ferriby

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biosystems Engineering – Master of Science

2021

ABSTRACT

HARNESSING MACHINE LEARNING TECHNIQUES FOR LARGE-SCALE MAPPING OF INLAND AQUACULTURE WATERBODIES IN BANGLADESH

By

Hannah Ferriby

Aquaculture in Bangladesh has grown dramatically in an unplanned manner in the past few decades, becoming a major contributor to the rural economy in many parts of the country. National systems for the collection of statistics have been unable to keep pace with these rapid changes, and more accurate, up to date information is needed to inform policymakers. Using Sentinel-2 Top of Atmosphere Reflectance images within Google Earth Engine and ArcGIS platforms, we proposed six strategies for improving fishpond detection as the existing techniques seem unreliable. The study area is comprised of seven districts in south-west and south-central Bangladesh. The tested strategies include: 1) identification of the best period for image collection, 2) testing the buffer size for threshold optimization, 3) determining the best combination of image reducer and water-identifying indices, 4) introduction of a convolution filter to enhance edge-detection, 5) evaluating the impact of ground-truthing data on machine learning algorithm training, and 6) identifying the best machine learning classifier. Each enhancement builds on the previous one to develop a comprehensive improvement strategy called the *Enhanced Method* for fishpond detection. We compared the results of each improvement strategy to the known ground-truthing fishponds as the metric of success. We compared the precision, recall, and F1 score for machine learning classifiers to determine the quality of results. Among the studied methods, the Classification and Regression Trees performed the best. Overall, the proposed strategies enhanced fishpond area detection in all districts within the study area.

Copyright by
HANNAH FERRIBY
2021

ACKNOWLEDGMENTS

I am extremely grateful to Dr. Pouyan Nejadhashemi who guided me through both my research and my program overall. Dr. Nejadhashemi not only led this research and advised me academically, but he also has been a great mentor. I thank him for his encouragement and patience with me throughout this project. In addition, I would like to thank my committee, Dr. Nathan Moore and Dr. Narendra Das for their continual support this past year.

Many thanks to both Dr. Ben Belton and Dr. Mohammad Mahfujul Haque for their great expertise and crucial work for this project.

I would also like to thank all of my lab mates for always supporting me and challenging me to be better; Anna Raschke, Dr. J. Sebastian Hernandez-Suarez, Josué Kpodo, Ian Kropp, Enid Banda, Mervis Chikafa, Kieron Moller, Shashank Mohan, Nicolas Fernandez, and Dr. Eeswaran Rasu. You all made my time at MSU really enjoyable and exciting – even over Zoom!

I cannot begin to express how thankful I am for the support of my family, especially my mother. You three always keep me sane. I would also like to thank Matt for always encouraging me to do my best. You have really been my rock the last year and a half.

This thesis is made possible by the support of the United States Agency for International Development (USAID) Grant No. 193900.312455.12B.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

AOI	Areas of Interest
AWEI	Automated Water Extraction Index
CART	Classification and Regression Trees
CRESDA	China Center for Resources Satellite Data and Application
EM	Electromagnetic energy
FAO	Food and Agriculture Organization of the United Nations
FtF	Feed the Future
GDP	Gross domestic product
GEE	Google Earth Engine
GIS	Geographic Information Systems
GPS	Global Position Systems
GRDH	Ground range detected high-resolution
GT	Ground Truth(ing)
H	Horizontal polarization
IP	Image processing
IW	Interferometric wide-swath
JRC	Joint Research Center
k-NN	*k*-Nearest Neighbor
LIFDCs	Low-income food-deficit countries
LR	Logistic Regression
LST	Land surface temperature
LWIR	Long-wave infrared

MM	MNDWI Mask
MNDWI	Modified Normalized Difference Water Index
MWIR	Mid-wave infrared
NDWI	Normalized Difference Water Index
NIR	Near infrared
OBF	Object-based Feature
OLI	Operational Land Imager
RF	Random Forest
SAR	Synthetic aperture radar
SNAP	Sentinel Application Platform
SVM	Support Vector Machine
SWIR	Short-wave infrared
TOA	Top of Atmosphere
TOPSAR	Terrain observation with progressive scans SAR
USAID	United States Agency for International Development
USGS	United States Geological Survey
UV	Ultraviolet
V	Vertical polarization
WPF	Water Presence Frequency
WRI	Water Ratio Index
WSA	Water Surface Area

1 INTRODUCTION

There is an increasing importance placed on inland aquaculture and fisheries due to a growing demand for fish and a stagnating production in capture fishing (Ottinger et al., 2016). Fish is a main source of nutrients for many people around the world and especially in Bangladesh (Heck et al., 2010). The increase in demand for fish is expanding aquaculture, putting pressure on croplands necessary to produce rice, Bangladesh's primary food source, creating a tension for space that needs to be addressed (Yu et al., 2018; Hashem et al., 2014). Meanwhile, there are little to no studies on the current production and potential of aquaculture in Bangladesh to inform policymakers and researchers (Shamsuzzaman et al., 2017).

Bangladesh was the fifth-highest aquaculture-producing country in the world (Subasinghe, 2017). However, aquaculture in Bangladesh has evolved dramatically in the last thirty years (Hernandez et al., 2018). A few highlights of this period include: 1) the amount of fish produced skyrocketed with 94% of the production for Bangladeshi consumption, 2) the farmed fish market expanded from 124,000 tons to 2 million tons, and 3) the number of people involved in the aquaculture value chain tripled, and 4) freshwater aquaculture production increased by 167% from 2001 to 2017, surpassing capture fishery (Hernandez et al., 2018; Sattar, 2019). Therefore, there is a need to better understand this evolving industry.

There are two primary methods for determining the extent of fishponds. The first method utilizes ground-truthing surveys. The surveys would provide accurate information on fishponds throughout the region of interest, but would be costly and time-consuming (Rhodes et al., 2015). The alternative to ground-truthing surveys is remote sensing to identify the specific locations in which aquaculture occurs. Remote sensing has been used to help with aquaculture management,

environmental monitoring, and aquaculture extent (Anand et al., 2020, Ottinger et al., 2016). Remote sensing can analyze large areas in much less time than it would take to conduct a survey. However, due to the complexity and variety in types of aquacultural production in Bangladesh, identifying a reliable technique that can be applied for large-scale studies can be challenging as previous studies only focused on small regions (Hashem et al., 2014, Yu et al., 2020, Huda et al., 2010).

This study aims to identify bodies of water throughout the region of interest in Bangladesh and to determine if they are being used for aquaculture using satellite imagery. This process can lead to creating an aquaculture identification method that does not rely on ground-truthing surveys.

The specific objectives of this study area to:

- Identify the best methods for locating waterbodies throughout the study region.
- Examine the existing methods for identifying fishponds and propose new strategies to improve the detection techniques.
- Evaluate the extent of aquaculture productions throughout the region of interest.

2 LITERATURE REVIEW

2.1 Global Aquaculture

2.1.1 Capture Fisheries versus Aquaculture

Aquaculture is comprised of breeding, rearing, and harvesting of fish and aquatic products within freshwater, sea water, brackish water, or inland saline water (FAO, 2020c). Aquaculture can be divided into three main categories: marine, brackish, and freshwater (US Department of Commerce, n.d.). Aquaculture systems can also be divided into different farming styles: water-based systems (e.g., cages and pens, onshore/offshore), land-based systems (e.g., rainfed ponds, irrigated or flow-through systems, tanks, and raceways), recycling systems (e.g., high control enclosed systems), and integrated farming systems (e.g., livestock-fish, agriculture, and fish dual-use aquaculture and irrigation ponds) (Funge-Smith, Simon; Phillips, 2001). Capture fisheries encompass all the aquatic animals killed, caught, trapped or collected in freshwater, brackish water, and/or marine water (United Nations et al., 2003).

Due to the accessibility of aquaculture, especially inland, its continual rise in production makes up for the decline in the availability of capture fisheries (Ottinger et al., 2018). Aquaculture contributed 42% of the global fish supply in 2012 and will eclipse capture fishery production by 2030 (Lam, 2016). Despite the global rise in production from aquaculture, a large portion of the fish produced and traded in low-income regions still comes from capture fisheries (Thilsted et al., 2016). Together, these two sectors support the livelihoods of 10-12% of the world population (Lam, 2016). Global food security, income distribution, and ecological sustainability all rely on the balance between capture fisheries and aquaculture (Lam, 2016). Figure 1 shows the trends in global aquaculture, both marine and inland, and capture fisheries, both marine and inland.

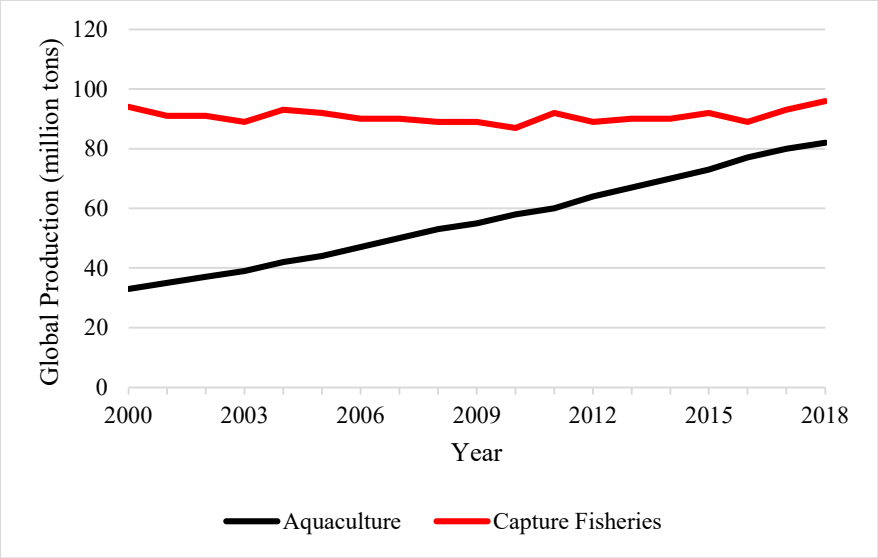

Figure 1. Global production in million tons of aquaculture and capture fisheries from 2000 to 2018 (FAO, 2020c).

2.1.2 Aquaculture by Water Type

There are three main types of environments used in aquaculture: freshwater (fishponds, fish pens, fish cages), brackish water (fishponds in coastal areas), and marine (fish cages or substrates) (FAO, 1989). Most of the world's aquaculture occurs in freshwater systems (Ross et al., 2013). Inland aquaculture produces nearly double what marine aquaculture does. In 2016, inland aquaculture produced 51.4 million tons of product, while coastal produced 28.7 million tons, excluding aquatic mammals, crocodiles, alligators, and aquatic plants (FAO, 2018). The gap in fish production between inland and coastal (marine) aquaculture has been growing significantly each year (FAO, 2018). It is important to note that inland fisheries are more accessible to even the poorest of people (FAO, 2018).

Taking into account all aquatic organisms, freshwater, and marine habitats contribute nearly the same to global aquaculture production (FAO, 2014). In 2014, freshwater aquaculture

produced 46.3 million tons, while marine aquaculture produced 47.4 million (FAO, 2014). Table 1 looks at the increasing trends in all aquaculture from 2010 to 2014.

Table 1. Global aquaculture production in million tons of all aquatic organisms by environment type from 2010 to 2018 (FAO, 2020a).

Environment	Year				
	2010	2012	2014	2016	2018
Freshwater	35.4	38.9	43.6	47.1	50.4
Brackishwater	5.4	6.4	7.6	8.7	8.6
Marine	37.1	42.8	48.4	52.4	55.6

2.1.3 Global Aquaculture Market

Global demand for fish, crustaceans, and mollusks is rising yearly (Ottinger et al., 2018). Global aquaculture in 2015 produced 106 million tons, worth approximately US$163 billion (Subasinghe, 2017). Aquaculture's role in global fish production increased from 13% to 45% between 1990 and 2015 (Ottinger et al., 2018). According to the Food and Agriculture Organization of the United Nations (FAO), fish and fish production products, such as aquatic plants, sponges, fats and oils, are the most traded food goods in the world (Subasinghe, 2017). At least 75% of the global fish and fishery products are entering international markets (Subasinghe, 2017). Aquaculture products contribute a critical role in the global food system, providing approximately 3 billion people with a minimum of 15% of their animal protein intake (Charles et al., 2010). 19.3 million people were involved in aquaculture during 2016, making it a substantial line of employment (FAO, 2018).

During the period between 1961 and 2013, the growth in the global supply of fish is larger than the rate of population growth for the same period (Ahmed et al., 2019). Fish consumption

increased per capita from 9.9 kg in the 1960s to 18.9 kg in 2010 (Ottinger et al., 2016). Despite this massive average growth, the growth in low-income food-deficit countries (LIFDCs) was much smaller, from 3.5 to 7.6 kg during the same period (Subasinghe, 2017). LIFDCs make up 15 of the 21 countries with the highest per capita inland fish production (FAO, 2018). With a rapidly growing world population and an annual loss of 5-7 million ha of farmland, there is a pressure on aquaculture to fill the gap in food production (Ahmed et al., 2019).

Capture fisheries around the world see stagnating production, yet with growing population and demand, obtaining aquatic food is a major concern for global food security (Ottinger et al., 2016). The current largest region of aquaculture production is Asia, followed distantly by the Americas, Europe, and Africa (Subasinghe, 2017). Table 2 breaks up global aquaculture production data from 2015 to the different regions of the world.

Table 2. Aquaculture production in 2015 by region (adapted from Subasinghe 2017)

Region	Production (tons)	Percentage of World Production
Africa	1,772,391	2.3%
America	3,273,376	4.3%
Asia	68,432,034	89.4%
Europe	2,875,159	3.8%
Oceania	188,066	0.2%

2.1.4 Aquaculture and the Environment

While the benefits to individual livelihoods and potentially global food security are great, aquaculture is a cause of environmental degradation and biodiversity loss (Ottinger et al., 2017). The environmental harm from aquaculture is due to the use of pesticides and other chemicals, as well as the discharging of untreated wastewater (Ottinger et al., 2017). Sustainability certification programs were created to address the negative environmental impacts and potentially increase

efficiency, but the programs were aimed at higher-value sectors (Charles et al., 2010). If aquaculture's production keeps growing at its current pace, the environmental effects could be detrimental (Ahmed et al., 2019).

2.2 Aquaculture in Bangladesh

2.2.1 Aquaculture Output in Bangladesh

Aquaculture in Bangladesh has shifted dramatically in the last three decades (Hernandez et al., 2018). A few major changes occurred during that time period: the amount of fish produced skyrocketed with 94% of the production for domestic consumption, the farmed fish market grew from 124,000 tons to almost 2 million tons, and the number of people involved in the aquaculture value chain tripled (Hernandez et al., 2018).

Bangladesh was the fifth-highest aquaculture producing country in the world in 2015 with 2.1 million tons and a growth rate of 5.3% from the previous year (Subasinghe, 2017). Inland waters in Bangladesh alone produced over 1 million tons in 2015 (FAO, 2018). Aquaculture produced 55% of the country's fish production in 2014 (Jahan et al., 2015). Freshwater aquaculture production increased by 167% from 2001 to 2017; surpassing capture fishery as the main source of fish in Bangladesh (Sattar, 2019). Figure 2 shows the change in aquaculture production in freshwater and brackishwater from 1980 to 2018.

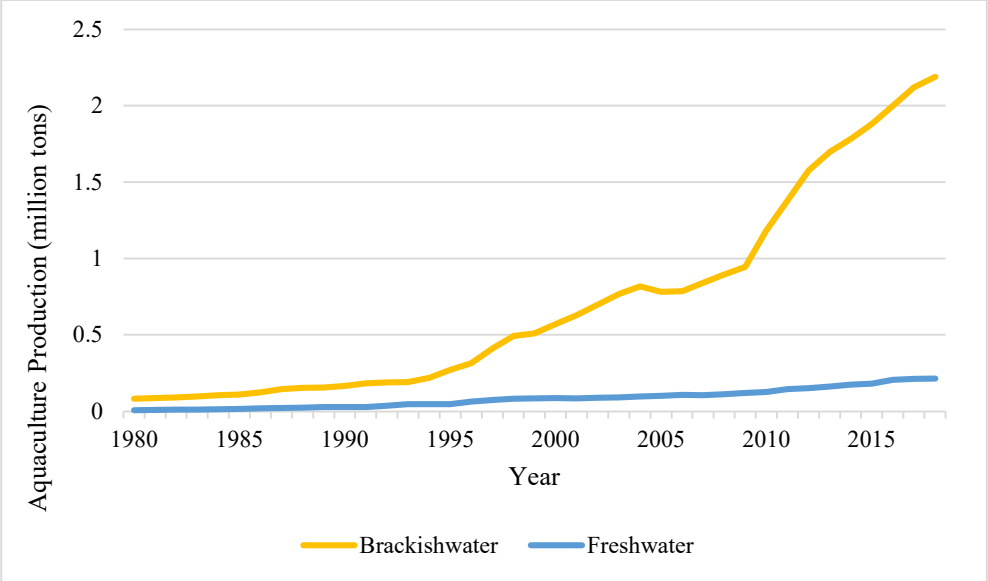

Figure 2. Aquaculture production in Bangladesh by environment type from 1980 to 2018 (FAO, 2020b)

Fish makes up two-thirds of the animal protein consumed and one-quarter of the agricultural gross domestic product (GDP) of Bangladesh (Sattar, 2019). Approximately 12 million people are connected with fisheries in Bangladesh, of which 1.4 million people are purely involved in fishery work (Shah, 2003).

2.2.2 Aquaculture Types and Technologies

There are two types of aquaculture in Bangladesh: freshwater and coastal, with inland pond culture being the mainstay (Islam et al., 2019). Bangladesh has 4 million ha of inland open waterbodies (rivers, lakes, ponds) and 0.7 million ha of closed waterbodies (man-made ponds) (Sattar, 2019). Pond farming in Bangladesh produces carp, cichlids, and catfish, which contributes 80% of the total recorded aquaculture in the country (Islam et al., 2019). Coastal aquaculture in Bangladesh is comprised of mostly shrimp and prawn farming in ghers (Islam et al., 2019).

Integrating fishponds with crops have many advantages including increased diversification, intensification, improved natural resource efficiency, increased productivity, and increased sustainability (Karim & Little, 2018). That is why, pond aquaculture has experienced the fastest growth in Bangladesh (Rashid et al., 2019). There are approximately 1.3 million fishponds in Bangladesh, covering 0.151 million ha (FAO, 2020b). Of the 0.151 million ha, 55% is cultured, 29% is culturable, and 16% is unused (FAO, 2020b). Most of the waterbodies used for aquaculture are operated by a single individual and contain water year-round (Jahan et al., 2015). There are different waterbody types used in Bangladesh: homestead pond, gher, commercial pond, beel, and rice-fish plot (Jahan et al., 2015). The descriptions of these waterbodies are presented in Table 3. For simplification, all types of aquaculture waterbodies will be referred to as fishponds.

Table 3. Descriptions of the five different types of waterbodies used in aquaculture in Bangladesh, adapted from (Jahan et al., 2015).

Aquaculture Waterbody Type	Definition
Homestead Pond	A pond located near a homestead that is used for numerous domestic purposes. These are usually small in size.
Gher	A type of rice field in southern Bangladesh that is deepened to hold fish and/or crustaceans. The excess soil is used to created dykes around the pond that can also be used to grow crops. Rice is not necessarily grown concurrently with the fish.
Commercial Pond	A pond dedicated to the year-round production of fish. Most of the fish grown are sold.
Beel	These are formed in low-lying lands after floodwaters have recessed or after heavy rains. These ponds are large and can be made suitable for fish by enclosing it with dykes.
Rice-Fish Plot	A type of rice field in northern Bangladesh that is deepened to hold fish. The excess soil is used to create dykes to prevent fish from escaping. Rice is grown with fish or in consecutive seasons.

2.3 Introduction to Remote Sensing

Remote sensing is the acquisition and measurement of data on specific properties of phenomena, objects, or materials through the use of a recording device not in physical contact with what is being observed (Khorram et al., 2012). Remote sensing if often grouped with image processing (IP), geographic information systems (GIS), and Global Positioning System (GPS) to create geospatial science (Khorram et al., 2016). Remote sensing, in an environmental context,

involves sensors recording electromagnetic (EM) energy that emanates from the areas or objects being observed on Earth (Khorram et al., 2012). Remote sensing uses wavelengths along the electromagnetic spectrum to view targets. Remote sensing can be small-scale, individual projects looking only at one small section, to large-scale, remote sensing via planes or space-borne sensors. In the following sections, we will discuss the different EM spectrum sections that sensors use, different types of sensors, and the types of resolutions used to describe sensors.

Space-borne sensors are broken up into three groups; visible, infrared, and microwave (JianCheng et al., 2012). Remote sensing in the visible wavelength range is commonly used for aerial photographs (Natural Resources Canada, 2015). Infrared remote sensing detects infrared radiation in the form of heat (CRISP & National University of Singapore, 2001). Microwave remote sensing relies on the different physical parameters that govern the microwave range to obtain new, useful data on targets (Woodhouse, 2017). All three are discussed further in sections 1.4.1 to 1.4.3. Table 4 provides examples of active satellites and their sensor types.

Table 4. List of active space-borne platforms and their sensor types.

Satellite Name	Sensor Type(s)
GeoEye-1	Optical - Panchromatic and multispectral imaging (1.5.2) (UCSB, 2020)
WorldView-1	Optical - Panchromatic and multispectral imaging (1.5.2) (ESA, 2020d)
ALOS-2	Radar - SAR (1.5.1) (JAXA, 2020)
Landsat-8	Optical - Thermal infrared sensor, radiometer for imaging (USGS, 2020)
CBERS-4	Optical - Panchromatic, multispectral imaging (both visual and infrared) (ESA, 2020a)
Sentinel-1A/1B	Radar - SAR (1.5.1) (ESA, 2020b)
Terra	Optical - Radiometer, spectroradiometer (NASA, 2020c)

There are two different modes of sensing: passive and active. Passive sensors measure the energy that is naturally available, while active sensors provide a source of energy for illumination of the target to be investigated (Natural Resources Canada, 2015). Lists of different passive and active sensors are in Table 5.

Table 5. List of different passive and active sensors (NASA EarthData, 2020).

Type of Sensor	Specific Sensor	Description
Active	Laser altimeter	Measures altitude from Earth's surface using lidar
	Lidar	Calculates distance by timing laser pulses to reflected light from pulses
	Radar	Calculates distance by timing backscattered radiation from microwave radiation emissions
	Ranging Instrument	Measures distance between sensor and target
	Scatterometer	Uses radar to determine surface wind speed and direction
	Sounder	Measures vertical distributions of atmospheric conditions (precipitation, temperature, humidity, cloud composition)
Passive	Accelerometer	Measures acceleration (translational or angular)
	Hyperspectral Radiometer	Detects narrow spectral bands in visible, near-infrared, and mid-infrared sections of electromagnetic spectrum
	Imaging Radiometer	Provides two-dimensional array of pixels
	Radiometer	Measures intensity of electromagnetic radiation in some bands within the spectrum
	Sounder	Measures vertical distributions of atmospheric conditions (temperature, pressure, and composition)
	Spectrometer	Measures and analyzes incident electromagnetic radiation
	Spectroradiometer	Intensity of radiation in multiple wavelength bands

The data from remote sensing can be described by four types of resolution: spatial, spectral, temporal, and radiometric. Spatial resolution refers to the ground area captured by a single pixel in an image. Spectral resolution refers to a sensor's ability to store and detect different wavelengths and is represented by the width of wavelength interval or number of spectral channels. The temporal resolution is the amount of time for a sensor to revisit the same geographic location.

Lastly, the radiometric resolution is the sensor sensitivity to brightness (Khorram et al., 2012). Table 6 lists examples of different active satellites with their spatial and temporal resolutions.

Table 6. Examples of active satellites and their respective spatial resolution and sensor(s) (Ottinger et al., 2016).

Satellite Name	Spatial Resolution (m)	Temporal Resolution (day)
GeoEye-1	0.41 – 1.65	2.1 – 8.3
WorldView-1	0.5 – 1.8	3.7
ALOS-2	1 – 100	14
Landsat-8	15 – 30	16
CBERS-4	20 – 64	5
Sentinel-1A/1B	5 – 80	12
Terra	15 – 90	16

2.4 Space-borne Sensors

2.4.1 Visible Light Remote Sensing

Remote sensing using the visible light range in the EM spectrum mainly refers to satellite imagery. This is the simplest and oldest remote sensing technique. Images taken in the visible light range are variants of blue, green, and red light – just as human eyes do. Visible light images show the target the same way our eyes would (Natural Resources Canada, 2015). Objects that appear "whiter" have high albedo, meaning it reflects more radiation than "darker" objects (NSIDC, 2020). Cloud cover and daytime hours can impact what the sensor is able to detect. Disaster

management, weather forecasting, urban planning, archeology, and environmental assessment are all examples of applications of satellite imagery (PBS, 2007).

2.4.2 Infrared Remote Sensing

The infrared region of the EM spectrum covers from 0.7 μm to 100 μm, making the range 100 times larger than the visible portion. Infrared can be broken up into two different categories: reflected infrared and thermal or emitted infrared. The reflected infrared is used for remote sensing in ways very similar to the visible section. Reflected infrared covers 0.7 μm to 3.0 μm. Thermal infrared is essentially the radiation that is emitted from the target, or the Earth's surface (Natural Resources Canada, 2015). The wavelength emitted depends on its temperature (CRISP & National University of Singapore, 2001). Clouds also emit and reflect infrared radiation, preventing satellites from obtaining data from the surface of the Earth (NSIDC, 2020).

2.4.3 Microwave Remote Sensing

Microwave interactions are typically governed by different physical parameters than other wavelengths along the EM spectrum (Woodhouse, 2017). They range from 1mm to 1 m in length (Gade & Stoffelen, 2019). Microwave sensing can see through obstructive weather conditions (Ellowitz, 1992) and pass through the top layer of soil (Woodhouse, 2017) in a way that other wavelengths cannot due to their length. Thermal emission can also be observed by passive microwave sensors, meaning that they do not rely on background sources like the Sun (Woodhouse, 2017). Microwave sensors are very responsive to different forms of water. They are able to make observations on soil moisture, vegetation water content, and snow cover (JianCheng et al., 2012). Table 7 shows the different bands of microwave sensing that are used and their respective characteristics. When discussing microwave sensing, the band is used to describe the wave attributes.

Table 7. Frequently used microwave wavelength bands, frequency bands, and their common names (Gade & Stoffelen, 2019).

Band	Wavelength band (cm)	Frequency band (GHz)
L	15.0 - 30.0	1.0 - 2.0
S	7.5 - 15.0	2.0 - 4.0
C	3.8 - 7.5	4.0 - 8.0
X	2.5 - 3.8	8.0 - 12.0
Ku	1.6 - 2.5	12.0 - 18.5
K	1.3 - 1.6	18.5 - 24.0
Ka	0.8 - 1.3	24.0 - 40.0
V	0.4 - 0.8	40.0 - 75.0
W	0.3 - 0.4	75.0 - 110.0

Common microwave sensors include radiometers (passive), altimeters, scatterometers, and synthetic aperture radar (SAR) (active). A few examples of satellites with microwave sensors include TerraSAR-X, ALOS/PALSAR, RADARSAT-2, and COSMO-SkyMed (JianCheng et al., 2012).

There are also disadvantages to microwave sensing, such as long antennas are needed to obtain spatial resolutions appropriate for large areas due to long microwave wavelengths (Woodhouse, 2017). Active microwave setups, such as SAR, tend to be the heaviest, largest, and most power-consuming Earth-observing satellites (Woodhouse, 2017).

2.5 Remote Sensing for Aquaculture

Remote sensing has the potential to provide data for aquaculture management, including site selection, mapping (Anand et al., 2020), environmental monitoring, and aquaculture inventory (Ottinger et al., 2016). The first step to the outcomes above is using data from remote sensing to locate under-utilized waterbodies to promote aquaculture production (Anand et al., 2020).

Generally, three types of products are used for aquaculture structures detection: SAR imagery, medium-resolution multispectral imagery, and high- and very-high-resolution multispectral and panchromatic imagery (Ottinger et al., 2016).

There is a wide range of space-borne sensors and instruments that can be used for aquaculture observation. GeoEye-1 is the satellite with the lowest spatial resolution at 0.41 – 1.65 m, followed by WorldView-1 and WorldView-2/3, which each have 0.5 and 1.8 m range, respectively. Some satellites, such as ALOS, Envisat, and ERS-1/2, would work for aquaculture observation, but lack the appropriate spatial resolution needed for small-scale details. These satellites have spatial resolutions of 30 – 1000 m (Ottinger et al., 2016).

2.5.1 Synthetic Aperture Radar

SAR is a technique for creating fine-resolution images from a radar system (NASA, 2020a). The term "synthetic" refers to the processing method of backscattered waves to improve the azimuthal resolution, which allows for smaller spatial resolutions (Woodhouse, 2017). The wavelength range for SAR is from a few to tens of centimeters (NASA, 2020a). The resolution of an imaging radar depends heavily on the antenna length; longer antennas provide finer resolution. This creates issues when using SAR on a satellite because an antenna would need to be kilometers long. To address this, the radar's real aperture is synthetically enhanced by using the Doppler shift of the backscattered signal (Gade & Stoffelen, 2019). As the SAR moves along its path, it sweeps the antenna's calculated length across the ground while continuously transmitting pulses and receiving the backscattered signal (NASA, 2020a).

Both Figure 3 and Figure 4 depict the Doppler shift but as two different representations. At point "A" in Figure 3, the backscattered signal causes a positive Doppler shift at "a", a zero Doppler shift at "b", and a negative Doppler shift at "c". This phenomenon can be referred to as

the Doppler History, which the end result of is a much longer aperture (Gade & Stoffelen, 2019). SAR systems use pulse compression, a technique where the echo of a pulse is matched with its original signal creating a shortened pulse of higher energy (Gade & Stoffelen, 2019). The graphical representation of this can be seen in Figure 4. Equation (1 shows how to calculate the beam size on the ground/target (NASA, 2020b):

$$\text{Distance to Target} \times \frac{\text{Wavelength}}{\text{Antenna Dimension}} = \text{Beam Width at Target} \qquad (1)$$

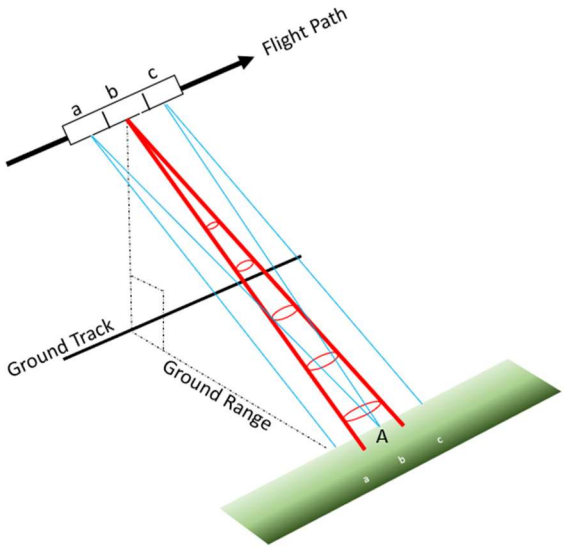

Figure 3. Illustration of the SAR Doppler technique (adapted from Gade & Stoffelen, 2019)).

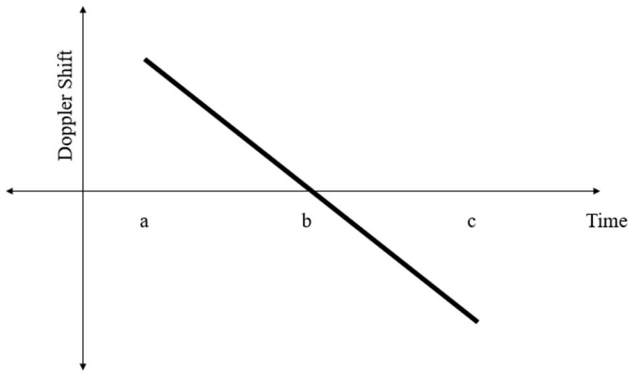

Figure 4. SAR backscattered Doppler shift (adapted from Gade & Stoffelen, 2019)).

Spaceborne SAR systems will typically have a wavelength in the C or X bands. A SAR system sends either horizontally (H) or vertically (V) polarized pulses. They then receive either H or V polarized return signals. SAR systems are characterized by how they send and receive signals. For example, a SAR system that sends horizontally polarized signals and receives horizontally polarized signals would be named HH (Travaglia et al., 2004). There are three SAR acquisition modes: Stripmap, ScanSAR, and Spotlight (Gade & Stoffelen, 2019). Stripmap is the classic acquisition mode in which the radar stays at a constant angle. ScanSAR mode involves multiple parallel swaths being scanned by the radar, resulting in one wide swath. Lastly, spotlight mode continuously changes the direction of the radar so that a single spot on Earth's surface is imaged for a longer time (Gade & Stoffelen, 2019).

The unique SAR system does have some complications. SAR images are impacted by noise, or speckling, created by constructive and destructive interference between the backscattered energy. This causes the value of the pixel to either increase or decrease, creating random bright and dark spots. It can be useful to apply speckle reducing procedures to better SAR images (Travaglia et al., 2004).

2.5.2 Multispectral and Panchromatic Imagery

Panchromatic images are taken using the UV (100 nm – 400 nm) and visible portions (400 nm – 700 nm) of the EM spectrum (Natural Resources Canada, 2015). The resulting images are in black-and-white (Khorram et al., 2012). Multispectral imaging involves simultaneously imaging in multiple wavelength bands to analyze targets (Coffey, 2012). A multispectral system normally provides a mixture of visible, near infrared (NIRS), short-wave infrared (SWIR), mid-wave infrared (MWIER), and/or long-wave infrared (LWIR) bands (Coffey, 2012). Multispectral imaging can be thought of as layers of panchromatic images, where each layer corresponds to a specific portion of the EM spectrum (Khorram et al., 2012). A benefit of multispectral imagery is much less subjective than aerial photography due to the higher information content (Chu, 2020).

2.6 Remote Sensing Water Indices for Small Waterbody Location

Inland waterbodies play a major role in contributing both food and livelihood security in rural areas (Anand et al., 2020). Small waterbodies include ponds, small lakes, low-order streams, ditches, and springs, but it is an ambiguous term, nonetheless. Ponds are standing waters varying from 2 to 5 ha in area and can be permanent or seasonal. Small lakes constitute standing water greater than 5 ha, but smaller than 100 ha (Biggs et al., 2017). There is no agreed-upon definition for a small stream, but the width of a small stream could vary from <3 to 6 m. Ditches are man-made channels which typically (a) are horizontal, (b) have linear boundaries and often turn at right angles, and (c) show little influence of the natural landscape contours. Lastly, springs are fixed places where groundwater emerges to the surface (Biggs et al., 2017).

In order to discover impacts of the number and extent of ponds on climate variability, Al Sayah et al. (2020) researched the French Claise watershed. The study used LANDSAT images, varying in time, to derive a Normalized Difference Water Index (NDWI). An NDWI map is used

to find and delineate water surfaces on the visible green light and near-infrared radiations. NDWI maps are widespread and have been used in applications as fine as mapping swimming pools. The NDWI map that was created had an accuracy of 85.74% for pond count and 75% accuracy for pond spatial allocation. In addition to the NDWI, the land surface temperature (LST) index was extracted from each LANDSAT map to show how pond numbers impacted the climate of the region. An LST index can be used to study the temperature measured on Earth's surface and is known for being one of the most suitable accurate indicators for studying climatic variabilities. The study reached the conclusion that pond-less zones had higher average temperatures than zones with ponds.

In addition to NDWI, the Automated Water Extraction Index (AWEI), Modified Normalized Difference Water Index (MNDWI), and Water Ratio Index (WRI) (Zeng et al., 2019) are also widely used indices for detecting waterbodies. Each of these indices is a way of extracting or pinpointing water in a target area. Each index requires different bands of the EM spectrum. The AWEI is an index formulated to effectively eliminate all non-water pixels in an image and improve the accuracy by removing shadow pixels (Feyisa et al., 2014). The MNDWI is advantageous in using to reduce and potentially remove built-up land noise as well as pinpoint water. Because of this, it is suitable for extracting water in an area with a background dominated by built-up land areas (Xu, 2006). The WRI was created to compensate for water's high reflectance in green and red bands compared to NIR and medium infrared (Gautam et al., 2015).

Anand et al. (2020) used high-resolution Cartosat 1 PAN and IRS ResourceSAT LISS IV merged imagery as well as Sentinel 2 Multi-Spectral Imagery to determine the Spatio-temporal water spread and effective water spread area for aquaculture in Chhattisgarh, India. The study used data from the years 2016 to 2018. The ortho-corrected images provided radiometric measurements

in top of atmosphere (TOA) reflectance. The study used an NDWI to create a Water Surface Area (WSA) map from each individual image. Equation (2) shows how the green (X_{Green}) and NIR (X_{NRI}) bands of the Sentinel 2A/2B MSI images are used to create the NDWI (McFeeters, 1995, Gautam et al., 2015). Green and NIR were chosen because of how they highlight the difference between water and land. Aquatic vegetation within a waterbody is differentiated from water. This also can happen in silty waters.

$$\text{NDWI} = \frac{X_{\text{Green}} - X_{\text{NIR}}}{X_{\text{Green}} + X_{\text{NIR}}} \tag{2}$$

To aggregate the NDWI water bitmaps, the study used Water Presence Frequency (WPF). The WPF is represented in Equation (3) where WPF_j is WFP of jth pixels in a time period; I_j is jth pixel having water in the selected NDWI images; n is the number of images (Anand et al., 2020). The WPF values range from 0 to 100%, where below 66% WPF concluded dry bed area and over 66% concluded water.

$$\text{WPF}_j = \frac{\sum_{i=1}^{n} I_j}{n} * 100 \tag{3}$$

The satellite data (e.g., Cartostat 1 PAN, IRS ResourceSAT LISS IV, and Sentinel-2 MSI) provided an approach for establishing the number of small waterbodies required for a certain production potential in standard production conditions. The study first created a waterbody boundary layer vector file, chose the appropriate time interval (in this case June to September due to monsoons), then extracted the surface water using NDWI and WPF before generating a composite seasonal water surface area map. To obtain accuracy, the study used field verification. Chhattisgarh is a landlocked state in central India which has large numbers of waterbodies. The study mapped over 120,000 waterbodies, creating a total area of 202,000 ha. Of the waterbodies

mapped, 97% had an area less than 5 ha. The study concluded that the methods used could be replicated in other areas with poor *in-situ* data.

SAR data is beneficial in addition to other forms of remote sensing (David Ballester-Berman et al. 2018). The Travaglia et al. (2004) study used SAR data because of its all-weather capabilities and because the backscatter allows for identification and separation of features. The aim of the study was to map and inventory coastal aquaculture and fisheries in Lingayen Gulf, the Philippines, using data from ERS-2 and RADARSAT-1. The ERS-2 satellite has a quasi-polar orbit, meaning that the descending orbit is approximately opposite to the ascending orbit. The angle between the scanning directions of two ERS SAR images is 152.7 degrees. Rough water conditions cause the mean backscattering coefficient of fishponds to increase on C-band images acquired in VV polarization than in HH. The ERS SAR has VV polarization while the RADARSAT SAR has HH. Using both SAR systems, a more accurate image is created. An analysis of the two images identifies all features in the target area. The results of the SAR imaging were compared to a database of topographic maps to define the accuracy of the study. The comparison concluded the study had an accuracy of 95% for identifying fishponds.

Platforms like Google Earth Engine (GEE) make mapping aquaculture ponds on a national scale much more efficient (Duan, Li, Zhang, Chen, et al., 2020). Duan et al. (2020) integrated spectral, spatial characteristics and morphological operations to create a decision tree classifier. The decision tree was used to extract the aquaculture pond regions along the Chinese coastal zone with an accuracy of 96%. The data came from the Landsat 8 Surface Reflectance Tier 1 dataset from GEE. The decision tree classifier has five steps: creating boundaries of potential aquaculture pond regions, extraction of waterbodies, extraction of intensive ponds, extraction of extensive ponds, and post-processing. Understanding the typical environment for ponds, e.g., low-lying

planes of coastal areas, helps determine where potential aquaculture pond regions are. The AWEI distinguished between non-water and water range the low-lying coastal zones, extracting the large waterbodies (ocean, lakes, rivers, extensive ponds). Equation (4 shows how the AWEI is determined, where ρ represents the reflectance value of spectral bands of Landsat 8: band 3 (green), band 5 (NIR), band 6 (SWIR 1), and band 7 (SWIR 2) (Feyisa et al., 2014, Wicaksono & Wicaksono, 2019).

$$\mathrm{AWEI}_{\mathrm{nsh}} = 4 * (\rho_{\mathrm{band3}} - \rho_{\mathrm{band6}}) - (0.25 * \rho_{\mathrm{band5}} + 2.75 * \rho_{\mathrm{band7}}) \quad (4)$$

Feyisa et al. (2014) introduced an additional Equation (5 to the AWEI that removes shadow pixels that were not removed in the previous equation where band 1 refers to the blue section of the EM spectrum.

$$\mathrm{AWEI}_{\mathrm{sh}} = \rho_{\mathrm{band1}} + 2.5 * \rho_{\mathrm{band2}} - 1.5 * (\rho_{\mathrm{band4}} + \rho_{\mathrm{band5}}) - 0.25 * \rho_{\mathrm{band7}} \quad (5)$$

To extract intensive aquaculture ponds and small waterbodies from what was identified by the AWEI, Duan et al. (2020) used an MNDWI proposed by Xu (2006). The MNDWI is shown in Equation (6, where the middle infrared (MIR) band replaces the NIR band.

$$\mathrm{MNDWI} = \frac{X_{\mathrm{Green}} - X_{\mathrm{MIR}}}{X_{\mathrm{Green}} + X_{\mathrm{MIR}}} \quad (6)$$

The MNDWI produces three results: water will have greater positive values than what the NDWI results in as it absorbs more MIR light than NIR, built-up land will have negative values, and vegetation and soil will have negative values as soil reflects MIR light more than NIR (Xu, 2006).

The WRI depends on the spectral reflectance of the green, red, NIR, and MIR bands. Equation (7 shows the WRI equation (Mohsen et al., 2018). A WRI value greater than 1 represents water.

$$\mathrm{WRI} = \frac{X_{\mathrm{Green}} + X_{Red}}{X_{NIR} + X_{\mathrm{MIR}}} \quad (7)$$

Mohsen et al. (2018) used a combination of the NDWI and WRI to extract water features from satellite images of Lake Burullus in Egypt. WRI requires the MIR band, a band that many satellites do not measure (Mukherjee & Samuel, 2016). The Mohsen et al. (2018) study did not report the accuracy of the methods but instead conducted a statistical analysis using the Mann-Kendall test on whether Lake Burullus was decreasing in size from the year 1972 to 2015. Using the water data from the WRI and NDWI, Mohsen et al. (2018) concluded that the lake had lost approximately half of its surface area in the given time frame.

While each index mentioned does extract water data from a target area, they are not all the same. Choosing which index or indices to use depends on a few factors, including, but not limited to, what data is available, the topography of the target area, and the objective of the study. Each index requires different bands in the EM spectrum. If that data is not available for the specific target area, then a different index must be used. The AWEI is useful in areas where steep topographic changes can create shadows in the images but may not be needed in flat areas. The MNDWI is useful for eliminating noise in an image. Many studies either used the indices in tandem to create the best image or used multiple indices separately to compare the results of each.

2.7 Remote Sensing Aquaculture Applications

2.7.1 SAR Applications

Prasad et al. (2019) used SAR data from Sentinel-1 satellite to assess the coastal aquaculture in India from September 2014 to June 2017. The SAR instrument on Sentinel-1 operates at a C-band frequency of 5.5 GHz. The study used all available VH polarized SAR data in the interferometric wide-swath (IW) mode and the ground range detected high-resolution

(GRDH) format. In the IW mode, three sub-swaths were captured using terrain observation with progressive scans SAR (TOPSAR). TOPSAR is a form of ScanSAR imaging where the antenna beam is switched cyclically between multiple adjacent sub-swaths to obtain data (ESA, 2020c). This Sentinel SAR method resulted in a swath width of 250 km and 5 m by 20 m spatial resolution.

Prasad et al. (2019) methods can be explained in five steps: preprocessing, calculation of a temporal median layer, topographic masking, segmentation, and spatial analysis. All data was obtained from the free and open-source Sentinel Application Platform (SNAP). Preprocessing involved removing thermal noise, converting intensity values, and correcting terrain distortions. The temporal median layer was calculated in the VH polarization at the pixel level. VH polarization was used because it is slightly better for differentiating land and water areas because of its more pronounced bimodal distribution of backscatter values. The temporal median layer was averaged over time, greatly diminishing speckle noise. Aquaculture and rice fields can be differentiated despite often having the same features because the aquaculture ponds will appear much darker in the temporal median layer because of their year-round water presence. The topographic masking excluded mountain areas and rough terrain since aquaculture typically occurs in flat environments. The target area was then broken into 11 approximately equal sub-regions for segmentation, then the Euclidian intensity pixel distance criterion was applied to see if two pixels belonged to the same segment. The final step involved using object-based image filtering with aquaculture pond characteristics inputs to identify ponds. After validating the results with reference datasets, the methods led to an accuracy of 97% for identifying aquaculture ponds.

Ottinger et al. (2017) used similar methods when using high-resolution Sentinel-1 SAR data to identify and map aquaculture ponds in China and Vietnam from September 2014 to September 2016. The target areas, more specifically, were the Mekong Delta, Red River Delta,

Pearl River Delta, and Yellow River Delta. High spatial resolution is needed for detecting certain pond features like embankments, levees, or dikes. The study also used Sentinel-1, but instead of data from both Sentinel-1A and -1B, Ottinger used Sentinel-1A dual-polarized (VV and VH) data in IW and Ground Range Detected High Resolution (GRDH). Sentinel-1B was not used because it finished its commissioning phase in September 2016. After *in situ* observations, more than 3000 aquaculture ponds were identified in the target areas. These would be used to calculate accuracy of the computer identified ponds. Only ponds within 20 km of the shoreline were selected for sampling.

The data for the Ottinger et al. (2017) study was also obtained from SNAP. The preprocessing included removal of thermal processing. After this, a radiometric calibration was performed, and then a terrain correction applied. Since many aquaculture ponds are enclosed, they have a very low backscatter value and distinct characteristics. To reduce speckling, the pixel-wise median was calculated, a process of comparing pixels to their surrounding pixels to remove outliers. Similarly to the previously discussed study, Ottinger et al. (2017) also calculated a temporal median image, terrain masking, and segmentation. However, unlike Prasad et al. (2019) Ottinger et al. (2017) performed an edge sharpening, finding that bilateral denoising and non-local means filter as the two best methods. Both methods successfully removed noise, but the non-local means filter did a better job of preserving detail. The average overall accuracy of their methods was 83%.

These two studies used similar methods and obtained high accurate results. The Prasad et al. (2019) study followed the Ottinger et al. (2017) study and had better results. Both studies looked at aquaculture in low-lying coastal areas, similar to the landscape in Bangladesh, which is the focus of our study.

2.7.2 Multispectral and Panchromatic Imaging Applications

An alternative to radar imaging is optical imaging, such as multispectral and/or panchromatic imaging. Virdis (2014) used high and very-high resolution panchromatic imaging, SPOT5 and Worldview-1 (both optical sensor satellites), and machine learning to detect shrimp farms in the Tam Giang-Cau Hai Lagoon in central Vietnam (70 km in length). To classify aquaculture shrimp farms, the study looked at panchromatic images from SPOT5 Level11A (5 m pixel size) and Worldview-1 Ortho-Ready Standard OR2A panchromatic images (0.5 m pixel size).

The data collected first went through a geometric correction and spatial accuracy assessment. The images then were cropped to fit the two areas of interest (AOIs), had contrast-enhanced, and edge detection filtering. The study used a non-commercial software called SPRING to segment the image through clustering. The clusters were then compared to a previously created aquaculture reference database of the area to find accuracy. The SPOT5 images led to an accuracy of 84.7% and 93.2% for AOI1 and AOI2, respectively. The Worldview-1 images led to an accuracy of 90.6% band 95.7% for AOI1 and AOI2, respectively. The algorithm implemented through the SPRING software worked well with both SPOT5 and Worldview-1, especially in AOI2, where there was a higher contrast between ponds and embankments.

Zeng et al. (2019) used medium resolution multispectral images to extract aquaculture ponds from water surfaces around inland Liangzi Lake in China. The study took medium-resolution multispectral image data from Landsat-5 (Thematic Mapper, TM) and Landsat-8 (Operational Land Imager, OLI), as well as high-resolution observation data from GaoFen-1 (World Field of View, WFV). The data was obtained via the United States Geological Survey (USGS) and the China Center for Resources Satellite Data and Application (CRESDA). All images

chosen had to be cloud-free and from December to January for minimal vegetation coverage. The water pixels were extracted using the MNDWI and the NDWI. These indices were chosen after calculating numerous indices and their accuracy for each data source. The MNDWI and NDWI had the best accuracy. The MNDWI was applied to the OLI and TM data, while the NDWI was applied to GaoFen-1 data. The water pixels were then separated into segments and geometrical features were calculated. The feature vectors formed were used as inputs to Support Vector Machine (SVM) classifier, an algorithm used for classification, regression, and outlier detection. The SVM classifier detected whether the water pixels were natural water surfaces or aquaculture. The results were validated from previously digitized and labeled water types in the area. The user accuracy ranged from 91.8% to 99.2% depending on the area of the lake (either eastern or western Liangzi Lake) and which image sensor. The lowest accuracy, 91.8% was the TM sensor in eastern Liangzi Lake. The highest accuracy was the OLI sensor in western Liangzi Lake. The methods used in this study can be applied to identify inland aquaculture ponds around the world. Landsat satellites have been collecting data since 1972, meaning the archives have great potential to monitor historic changes in inland lake environments.

2.7.3 Table of Relevant Studies on SAR and Optical Remote Sensing

Table 8 lists relevant studies that utilized radar (SAR) and/or optical satellite imagery.

Table 8. List of relevant studies for fishpond detection using radar and optical satellite imagery.

Product	References	Country	Resolution		Algorithm	Prediction accuracy	Software
			Spatial	Temporal			
Synthetic aperture radar (SAR)							
ERS-2 (C band VV polarization), Radarsat-1 (C band HH polarization)	Travaglia et al., 2004	Philippines	12.5m (ERS-2) 6.25m (Radarsat-1)	Two images at different dates (ERS-2) A single image (Radarsat-1)	Visual interpretation	95%	Source: ESA Preprocessing: ERDAS, ArcGIS
Radarsat-1 (C band HH polarization)	(Liu et al., 2010)	China	6m (raw 4.6x5.1m)	Less than a month (three images total)	Object-based (multi-temporal segmentation). Parameters: scale, color, and shape.	83.1% kappa 0.81	Preprocessing: Radar Analysis Package - PCI; ERDAS Processing: Definiens
Sentinel-1A/ B (C band VV + VH polarization)	(Ottinger et al., 2017, 2018)	China, Vietnam	10m (after 5x1 multi-looking) (raw 5x20m) Ground Range Detected High Resolution (GRDH)	Less than two weeks (66 to 192 scenes in a two-year period)	Object-based mapping (temporal filtering, topographic masking, connected component segmentation)	Overall 84% (ranging from 80% to 88%) kappa 0.68 (ranging from 0.59 to 0.77)	Source: Sentinel Scientific Data Hub/Google Earth Engine Preprocessing: Sentinel Application Plattform (SNAP) Processing: GDAL tools, Orfeo Toolbox (OTB)
Sentinel-1A/B Single Look Complex (SLC)	(Ballester-Berman et al., 2018)	Spain, Norway	2.3x14m	One image per study site was used	Unsupervised Wishart classification	Not reported	Source: Alaska Satellite Facility's data portal Processing: Sentinel Application Platform (SNAP)
TerraSAR-X, Sentinel-1B	(Dumitru et al., 2018)	Albania, Greece	Not reported	Images for a single date were used	Images are tiled into patches. A primitive feature vector is extracted from each patch. The patches are then clustered and classified using a Support Vector Machine classifier and predefined labels from optical imagery.	From 80 to 95%	Not reported

Table 8 (cont'd)

Product	References	Country	Resolution		Algorithm	Prediction accuracy	Software
			Spatial	Temporal			
Sentinel-1A/B (C band VH polarization)	(Prasad et al., 2019)	India	10m	Imagery collected for a two-year period	Object-based mapping (temporal filtering, topographic masking, connected component segmentation)	Overall 90%	Source: Sentinel Scientific Data Hub/Google Earth Engine Preprocessing: Sentinel Application Plattform (SNAP) Processing: GDAL tools, Orfeo Toolbox (OTB)
Optical							
Landsat 4 and 5 imagery	(Kapetsky, 1987)	Zimbabwe	1:250,000 scale	Two images were employed (dry season, and end of rainy season)	Visual interpretation	43 (satellite) out of 77 (aerial photos) (56%)	Not reported
IRS 1D (panchromatic images) SPOT (multi-spectral images)	(De Graaf et al., 2000)	Bangladesh	6m (IRS 1D) 20m (SPOT)	Two images at different dates	Visual interpretation	65 to 75% of ponds larger than 1,000 m2 are detected.	Not reported
CORONA (panchromatic), SPOT-3 (multispectral), ERS-1 (SAR), SIR-C (SAR), X SAR, Landsat 5 (multi-spectral), IRS (panchromatic, multi-spectral)	(Huda et al., 2010)	Bangladesh	CORONA (2m) SPOT-3 (1:50,000) ERS-1 (12.5m) SIR-C (30m) X SAR (25m) Landsat 5 (30m) IRS (6m, 23m)	Up to two images of each platform along multiple years between 1972 and 2003	Visual interpretation	Not reported	Not reported
SPOT5 Level1A, Worldview-1	(Virdis, 2014)	Vietnam	5m (SPOT5) 0.5m (Worldview)	One image for each product	Image segmentation using basin-detection and region-growing techniques. Unsupervised clustering classification ISOSEG	> 95%	SPRING
Optical imagery	(Z. Yu, 2019)	Bangladesh	Not reported	Not reported	Algorithm incorporating spectral and spatial filtering on multi-temporal images.	Not reported	Google Earth Engine
IKONOS, Quickbird, Worldview-2, Worldview-3	(Gusmawati et al., 2017)	Indonesia	0.5 - 1m (panchromatic) 1.5 - 4m (multispectral)	Images from 2000 to 2015	Visual interpretation	> 80%	ArcGIS

Table 8 (cont'd)

Product	References	Country	Resolution		Algorithm	Prediction accuracy	Software
			Spatial	Temporal			
Sentinel-1A, Landsat 5/7/8 Surface Reflectance (Level 2)	(Stiller et al., 2019)	China	10m (Sentinel-1) 30m (Landsat)	66 to 174 scenes in a two-year period for Sentinel-1 Yearly cloud-free image composites between 1984-2016 using Landsat	For Sentinel-1, same as Ottinger et al. (2018) For Landsat: 1) Cloud and cloud shadow masking 2) NDWI calculation 3) Difference of 10th and 90th percentiles (to discard rice paddy fields) 4) Set threshold for land-water separation (trial and error, visual interpretation) 5) Comparison with Sentinel-1 detected ponds, water proportion over time is obtained 6) For each year, Sentinel-1 fishponds that had at least 1/3 of their surface are retained. 7) False positives are detected using the Global Surface Water (GSW) raster dataset.	Overall 89% kappa 0.78	Google Earth Engine
Landsat 5/8, Gaofen-1 Wide Field of View (WFV)	(Zeng et al., 2019)	China	30m (Landsat) 16m (Gaofen)	Up to three images from the end of December to January for each product	Water surface extraction is performed using NDWI, MNDWI, AWEI, NDVI, WRI, NDMI. Thresholds are determined using the Otsu method (Otsu, 1979). The most accurate water and non-water binary image is selected. Water segments are obtained, and for each one, geometric features (area, regularity, perimeter) are determined based on boundary tracing and contour-based regularity and curvature computations. A SVM classifier incorporating the geometric features is used to separate aquaculture from natural water surfaces.	Overall > 94% kappa > 0.8	Source: USGS, China Center for Resources Satellite Data and Application Preprocessing: FLAASH tool in ENVI 5.3

Table 8 (cont'd)

Product	References	Country	Resolution		Algorithm	Prediction accuracy	Software
			Spatial	Temporal			
Landsat 5/7/8	(Ren et al., 2019)	China	30m	Images for 5 years between 1984 and 2016	Updating approach using visual interpretation and automatic classification based on water surface extraction with NDWI and object-based classification (parameters: scale, shape, and compactness).	Overall > 87% kappa > 0.80	Source: USGS Preprocessing: FLAASH tool in ENVI 5.0 Processing: eCognition Developer 8.64
Sentinel-2A/B, Cartosat 1 PAN, IRS LISS IV	(Anand et al., 2020)	India	10m (Sentinel-2) 2.5m (Cartosat 1 PAN + IRS LISS IV)	Sentinel-2: Images for two time periods: February and May (selection based on agricultural operations). Cartosat-1 + IRS: A single merged image.	Visual interpretation (2.5m data) Water surface extraction using NDWI (Sentinel-2)	Not reported	Source: Copernicus open access hub Processing: ArcGIS
Landsat 5/8 Surface Reflectance Tier 1	(Duan, Li, Zhang, Liu, et al., 2020)	China	30m	Revisit time: 16 days Seven-time slices from 1988 to 2018, 53 year each, between April to October	Decision-tree classifier: 1) Non-water land elimination using a water extraction index 2) Intensive aquaculture ponds extraction using the MNDWI and the 8-neighborhood Laplacian operator. 3) Extensive aquaculture ponds identification using two shape indexes representing the regular shape of these ponds. 4) Merging different pond types)	Overall >91% kappa > 0.79	Google Earth Engine
Landsat 5/7/8, Sentinel-2A	(Al Sayah et al., 2020)	France	30m (Landsat) 10m (Sentinel-2A)	Multiple images for October	NDWI	Landsat: 86% (pond count) 75% (pond spatial allocation) Sentinel: 93% (pound count) 84% (pond spatial allocation)	Source: USGS Earth Explorer (Landsat)

2.8 Basic Machine Learning and its Applications in Detecting Aquaculture Farms

Machine learning is computer algorithms for translating human ways of learning into machines (Faul, 2019). Machine learning is founded on the need for a computer to convert data examples into knowledge (Kubat, 2017). It is a field of computer science that studies techniques for obtaining results to complex issues that are difficult to solve using conventional programming methods. Machine learning algorithms have provided solutions to very complex problems, from internet searching to speech recognition (Rebala et al., 2019). Benefits to machine learning include a reduction in programming time, ability to customize and scale products, and complete "unprogrammable" tasks (Google, 2020).

Machine learning generally uses two different techniques: supervised learning and unsupervised learning. Supervised learning trains a model on known input and output data to predict future outputs by using classification and/or regression. Classification is a technique that predicts discrete responses, meaning that it categorizes input data (e.g., is a pond is a fishpond or not). Different classification algorithms include SVM, Discriminant Analysis, Naïve Bayes, and Nearest Neighbor. Regression is a supervised learning technique that predicts continuous responses (e.g., temperature changes). Examples of different regressions are Linear Regressions, Support Vector Regression, Ensemble Methods, Decision Trees, and Neural Networks. Unsupervised learning is a model that finds hidden patterns or structures in input dataset. This method uses clustering, a method of grouping data together based on certain similarities or patterns. Different types of clustering methods include *k*-Means, *k*-Medoids, Hierarchical, Gaussian Mixture, Neural Networks, and Hidden Markov Model (MathWorks, 2016d). Table 9 and Table 10 describe some common supervised and unsupervised algorithms and when they are best used.

Table 9. Common supervised algorithms and their applications (MathWorks, 2016a).

Algorithm Name	Process	Applications
Logistic Regression	Creates a model that predicts the probability of a binary event	• Data is clearly separated by a single, linear boundary • Can be a baseline for more complex classifications
k Nearest Neighbor (*k*-NN)	Categorizes based on characteristics on nearest neighbors in a dataset	• Establishes benchmark learning rules • Memory usage is not important • Prediction speed is less of a concern
Support Vector Machine (SVM)	Classifies through discovering linear decision boundary that separates all data points from one group to another	• Data only has two classes • High-dimensional, nonlinearly separable data • Easy to interpret
Neural Network	Highly connected networks relate inputs to outputs	• Modeling non-linear systems • Data is available in increments • If there is a possibility of unexpected change in input data • Model does not need to be easily interpreted
Naïve Bayes	Assumes presence of a certain feature is unrelated to the presence of another feature	• Small datasets with many features • Easy to interpret • Model will encounter circumstances not covered in training data

Table 10. Common unsupervised algorithms and their applications (MathWorks, 2016b).

Algorithm Name	Process	Applications
k-Means	Divides data into k number of mutually exclusive clusters	• k is known • Fast clustering of large data sets
k-Medoids	Divides data into k number of mutually exclusive clusters, but cluster center must be a data point	• k is known • Fast clustering of categorical data • Scales to large datasets
Hierarchical	Creates nested sets of clusters into a binary, hierarchical tree	• Number of clusters unknown • Provides visualization for selection
Self-Organizing Map	Neural network that changes a dataset into a multidimensional map	• Provides data visualized in 2D or 3D • Deduces dimensionality of data by preserving the shape

Using machine learning occurs in steps. The first step is to access the proper data, then preprocess data (e.g., check for outliers or missing data points), derive certain features (i.e., turning raw data into information), train models using the features, iterate to find the best model, and then integrate the best model into a productive system (MathWorks, 2016c).

Machine learning techniques such as SVM has been used extensively for remote sensing classification as a supervised method and can classify aquaculture ponds from natural water surfaces. For example, Zeng et al. (2019) used the SVM classifier algorithm to identify aquaculture farms in an inland lake in China. Geometric features, like perimeter, area, and contour-based regularity, are examples of input features into an SVM. The Zeng et al. (2019) study used geometric features from satellite images as training datasets. These datasets enabled the algorithm to pick out aquaculture ponds from natural surface water with an accuracy of at least 91.8%. In

addition to supervised classification, machine learning can be unsupervised, meaning without being presented a dataset with a known outcome. In SNAP there is the Wishart classifier, which has been used to identify differences in aquaculture structures within SAR imagery (Ballester-Berman et al., 2018). WorldView-1 and SPOT5 imagery have been also used in partnership for fishpond classification using the unsupervised Isogeg classifier in the Spring software (Virdis, 2014).

2.9 Limitations of Remote Sensing Techniques in Detection of Small Waterbodies

Despite previous research, identifying small waterbodies or aquaculture ponds on a large scale using remote sensing data can be a very challenging task mostly due to:

- There is no universal index to use for extracting water from an image (Zeng et al., 2019)
- There are multiple machine learning algorithms could be used for a fishpond detection, but there is no grantee that any or all of them work for an area of interest (Virdis, 2014; Dumitru et al., 2018; David Ballester-Berman et al., 2018; Zeng et al., 2019)
- Inland aquaculture comes in many different forms (Jahan et al., 2015). Because of this, there are no apparent geometric features to provide for an algorithm. Some aquaculture ponds like homestead ponds and beels form naturally with the shape of the land. Ghers and rice-fish plots are aquaculture coinciding with agriculture, meaning the shape of the pond will take the shape of the rice field it is built into.
- There is no freely available high-resolution (< 5 m pixel size) imagery to improve current fishpond detection (Z. Yu et al., 2020).

2.10 Goals

There is currently a lack of understanding of the extent of aquaculture in Bangladesh. The goals of this thesis are to find appropriate remote sensing data, applying different manual and machine learning techniques to distinguish various aquaculture, and verify the manual and machine learning algorithms against ground-truthing. The result will be used to improve official statistics and to enhance the capacity for aquaculture production and regulations within Bangladesh.

3 HARNESSING MACHINE LEARNING TECHNIQUES FOR LARGE-SCALE MAPPING OF INLAND AQUACULTURE WATERBODIES IN BANGLADESH

3.1 Introduction

Aquaculture (farming fish and other aquatic animals) is growing at an extremely rapid pace, increasing around 6.5 times, from 13 million tonnes to 85 million tonnes over the past thirty years (FAO, 2020c), and is forecast to continue growing rapidly in coming decades (Kobayashi et al., 2015) Most of this growth has occurred in Asia, which accounts of 89% global aquaculture production (FAO, 2020c). Aquaculture farm expansion in Asia has been concentrated mainly in water abundant deltaic regions (Bernzen et al., 2021). The growth of aquaculture since the 1990s has been particularly rapid in Bangladesh, induced by demand caused by rising incomes, urbanization, a growing population, and declining supplies of fish from capture fisheries. This development has occurred in a largely unplanned and spontaneous way as farmers have converted rice paddies into fishponds that typically offer much higher returns than paddy cultivation (Belton et al., 2018). As a result, Bangladesh is now the 5th largest aquaculture producing country in the world (FAO, 2020c).

Aquaculture in Bangladesh has shifted dramatically from subsistence to commercial farming in the last three decades (Hernandez et al., 2018). During this time, the farmed fish market grew from 124,000 tons to 2.4 million tons, and the number of people involved in the aquaculture value chain tripled (Hernandez et al., 2018; Department of Fisheries, 2018). This dynamic growth has created a huge new industry that makes substantial contributions to the rural economy in many parts of the country, but the rapid emergence of the sector means that it performance is not well understood by policymakers and researchers (Jahan et al., 2015; Hernandez et al., 2018, Belton & Azad, 2012). Some observers contend that unplanned expansion of aquaculture is creating

competition for arable land necessary to produce rice, Bangladesh's primary food crop, creating pressures on limited space that need to be addressed (Yu et al., 2018; Hashem et al., 2014). Moreover, studies have suggested that national aquaculture statistics may not have kept pace with this rapid growth, meaning that production volumes might be underestimated (Belton & Azad, 2012).

Accurately determining the physical extent of aquaculture in Bangladesh can be approached in different ways. One is to conduct physical surveys. These would provide accurate information about the use of waterbodies throughout the region, but because of the large areas and populations involved, this approach would be highly resource-intensive (Rhodes et al., 2015). Using remote sensing to identify where aquaculture is occurring is an alternative approach with potential to provide data for aquaculture management, including site selection for potential farmers, mapping (Anand et al., 2020), environmental monitoring, and aquaculture inventory (Ottinger et al., 2016). Remote sensing is helpful for analyzing large areas and obtaining information quickly, but it is difficult to determine use from an identified body of water.

Generally, three types of remote sensing products are used for aquaculture structure detection: synthetic-aperture radar (SAR) imagery, medium-resolution multispectral imagery, and high- and very-high-resolution multispectral imagery (Ottinger et al., 2016). SAR is a technique for creating fine-resolution images from a radar system (NASA, 2020a). The term “synthetic” refers to the processing method of backscattered waves to improve the azimuthal resolution, which allows for smaller spatial resolutions (Woodhouse, 2017). Multispectral imaging refers to sensors that can simultaneously capture many wavelength bands across the electromagnetic spectrum (Coffey, 2012). Medium spatial resolutions refer to any images with pixel sizes of 10 to 30 m, whereas high and very high spatial resolutions refer to any images with pixel sizes of 30 cm to 5 m per pixel

(Earth Observing System, 2019). Ultimately, selecting a particular product mostly depends on trade-offs between cost, spatial resolution, temporal availability, and surface area coverage (Bello & Aina, 2014, Zoran et al., 2010).

Some current methods for utilizing SAR imagery to identify aquaculture structures include object-based mapping in China, Vietnam (Ottinger et al., 2018), and India (Prasad et al., 2019) and Wishart classifiers applied to coastal regions of Spain and Norway (Ballester-Berman et al., 2018). However, filtering to remove speckle noise in SAR imagery negatively impacts its spatial resolution (Z. Yu et al., 2020). For medium-resolution optical imagery, the current methods include object-based mapping and classification along China's coasts (Ren et al., 2019), utilizing the Normalized Difference Water Index (NDWI) to determine aquaculture value thresholds in Southern China (Stiller et al., 2019), and a combination of pixel selection and image segmentation to find aquaculture in northern Bangladesh (Z. Yu et al., 2020). Meanwhile, methods utilizing Landsat medium resolution multi-spectral imagery (Ren et al., 2019) do not work for identifying small (approximately 0.002 km^2) fishponds because of its 30 m spatial resolution. Finally, high-resolution optical imagery can be resource-intensive and expensive; however, it has been used for visual aquaculture interpretation in India (Anand et al., 2020) and Indonesia (Gusmawati et al., 2017).

This study aims to identify bodies of water throughout the study region and determine if their use is for aquaculture. This can lead to the creation of an aquaculture pond identification method that does not rely on tedious surveys for information and can locate drastically different shapes and sizes of fishponds. However, what makes Bangladesh so different from previous studies is that there are many types of fishponds, all with different characteristics. 'Commercial ponds' (purpose-built ponds producing fish exclusively for sale) and 'homestead ponds' (small multi-use backyard

ponds), usually contain water year-round but can vary considerably in size. Ghers (modified diked rice paddies), only contain water during certain months of the year and can vary widely in terms of size and extent of integration with field crops. Ghers can be as small as 0.002 km^2 to as large as 0.4 km^2, and tend to not have a standard shape, instead following the topography. In the remainder of the paper for simplicity, we refer to homestead ponds, commercial ponds, and ghers using the catchall term 'fishponds', except when emphasizing salient differences between these waterbody types.

Among the studied methods, the method outlined in Yu et al. (2020) was deemed to be the best fit for this study because 1) similar to this study, their method was also developed and tested in Bangladesh, although their research was conducted in the Natore district in northern Bangladesh, 2) both studied regions fall in the same agroecological zone (Department of Agricultural Extension, 2021), 3) the logic behind their method is supported by open access code and can be easily implemented to other regions, and 4) inputs required for their method are free while using other methods rely on high- and very-high-resolution multispectral imagery can cost thousands of dollars. Therefore, in this study, we hypothesized that the Yu et al. (2020) method could be employed to achieve the research goal. Building on the Yu et al. (2020) method, first, we examined its performance within the study area and then proposes several improvement strategies to 1) enhance water detection specifically for small water bodies, 2) advance the edge detection for fishponds, and 3) achieve better fishpond classification.

3.2 Materials and Methods

3.2.1 Study Area

The study area is comprised of seven districts in southwest and south-central Bangladesh, with a total area of 17,385 km^2 (Figure 5). These are Bagerhat, Barisal, Bhola, Gopalganj, Jessore, Khulna, and Satkhira. These districts were chosen because they are home to some of the highest concentrations of aquaculture farms in the country, including nearly all of Bangladesh's shrimp and prawn farms (Department of Fisheries, 2018). The districts are also in the USAID Feed the Future (FtF) Zone of Influence meaning they are being targeted to ensure long-term economic sustainability in farming (Feed the Future, 2021).

For this research, we focus on aquaculture in enclosed waterbodies (ponds and ghers), and exclude farming taking place in stocked natural waterbodies, pens, and cages (Department of Fisheries, 2018). The study region excludes the Sundarbans National Park in the southern parts of Satkhira, Khulna, and Bagerhat. The land use for the study region was calculated from data from GlobeLand30 (China Ministry of Natural Resources, 2020). The study region is made up of 62% cultivated land (9966 km^2), 23% artificial surfaces (3661 km^2), 13% water bodies (2136 km^2), and 2% wetlands (296 km^2). Forest, shrubland, grassland, and bare land each encompass less than 1% of the total land use (Figure A1).

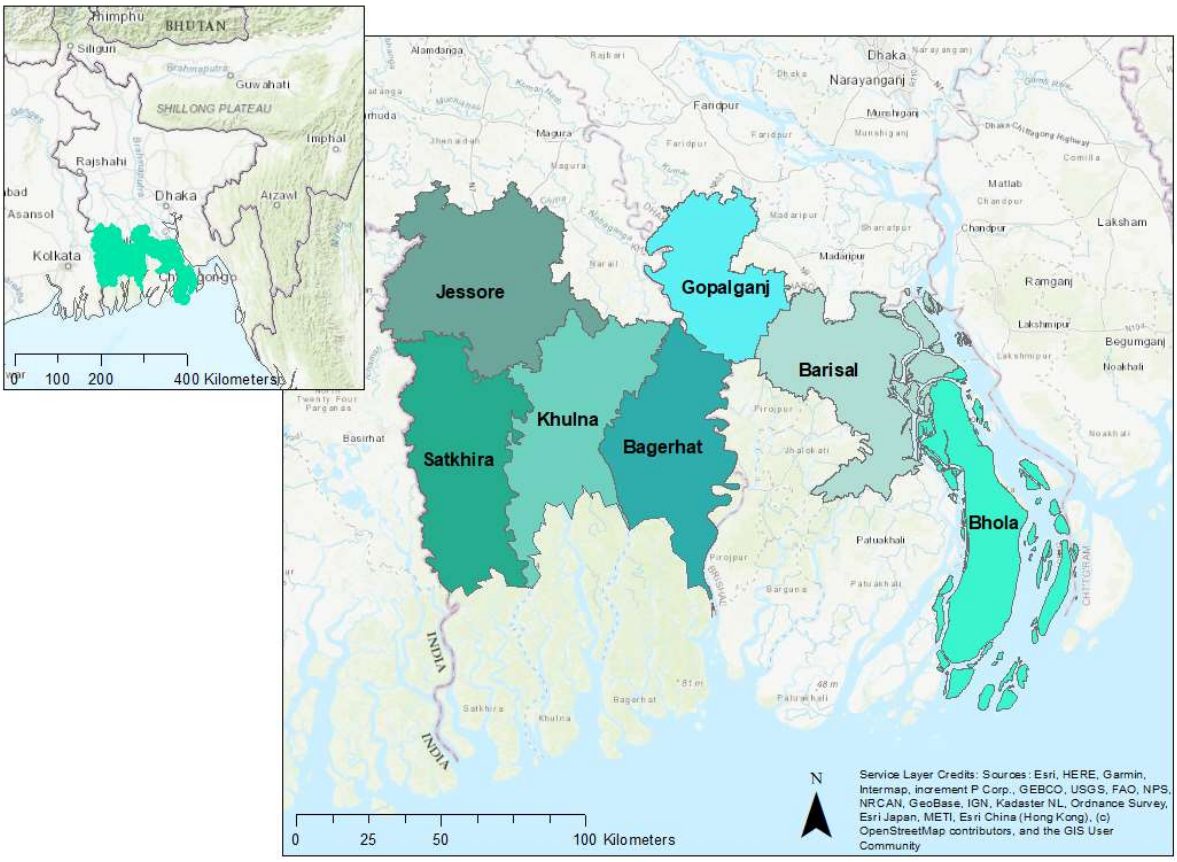

Figure 5. Study area is broken into seven individual districts.

The seven districts are primarily made up of cultivated land, artificial surfaces, water bodies, and wetlands (Table A1). Gopalganj, Bagerhat, and Jessore have the highest percentages of cultivated land (all over 65%). Satkhira has the highest percentage of water bodies at 39% - even higher than their cultivated land. Bhola is one of the few districts with wetlands as a land cover and has the highest percentage of wetlands with 13%.

The study area falls into three different climate subregions: the south-eastern zone, south-western zone, and south-central zone (Momtaz & Shameem, 2016). The south-eastern zone includes Bhola and the southern part of Barisal. This region sees high rainfall from May to September. Northern Barisal, Gopalganj, southern Khulna, and southern Satkhira fall within the south-central zone. This area has less rain on average than the south-eastern zone. The south-

western zone is comprised of northern Khulna, northern Satkhira, and Jessore and has the lowest levels of rainfall on average out of the three zones (Bangladesh Meteorological Department, 2021).

The study area has four seasons throughout the year that are described as winter (December to February), pre-monsoon (March to May), monsoon (June to September), and post-monsoon (October and November) (Momtaz & Shameem, 2016). The heavy rainfall of the monsoon season makes it challenging to obtain could-free satellite images during that season. The districts within the study region see anywhere from 1200 mm to 2800 mm of rainfall during the monsoon season (Bangladesh Water Development Board, 2019). The south-eastern zone has a slightly longer rainy season with higher precipitation from May to September (Bangladesh Meteorological Department, 2021).

3.2.2 Overview of the Base Method

As described in the Introduction section, the Yu et al. (2020) method was selected as a *Base Method* for detecting fishponds for the study area. In general, only a small number of changes (e.g., temporal and spatial extend of satellite imageries) to the original code were needed to fit the method for the study area. The *Base Method* comprises six main steps: data acquisition, water index calculation, initial mask, threshold optimization, majority vote, and fishpond identification (Figure 6).

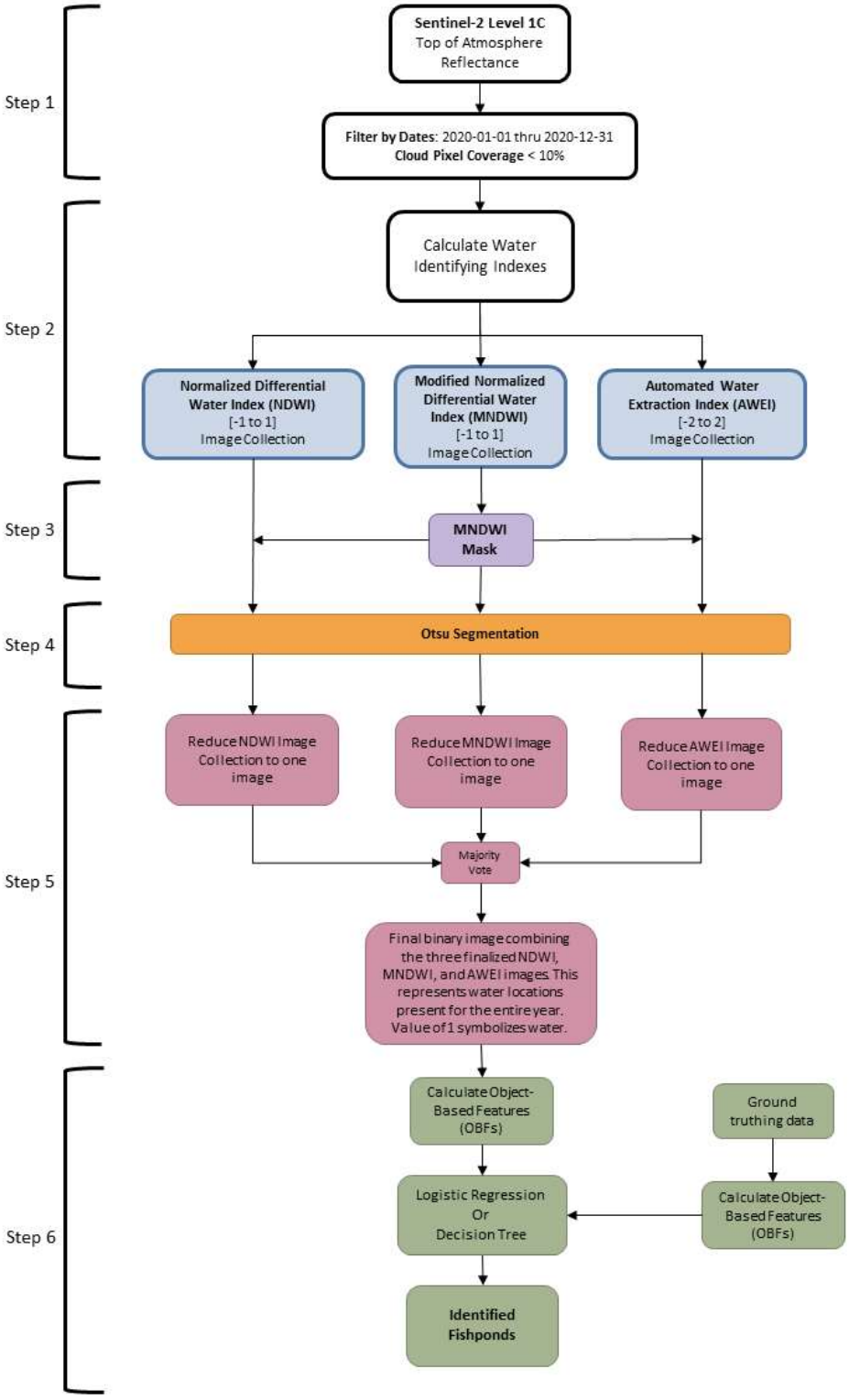

Figure 6. Flowchart outlining the major steps of the Yu et al. (2020) for fishpond identification.

Step 1. Data Acquisition: The process starts with combining data from Sentinel-2 Level-1C (Top of Atmosphere Reflectance) from 1/1/2020 to 12/31/2020. The images are then filtered for cloud coverage, keeping only those images with less than 10% cloud pixels. We obtained 311 images for our study region during 2020 that include any image that covers any part of our study region. This is one of the major differences with the original Yu et al. (2020) study as they only obtained four images for their region of interest during the study period of 2016. In addition, the size of our study region is just under 33 times larger than the Yu et al. (2020) study area. During the 2020 study period, no images were available between mid-June and the end of September due to the monsoon occurrence.

Step 2. Water Index Calculation: To identify where water is located within the study area, three different water indexes were used. These indexes analyze different wavelengths of reflected light to determine the presence of water. Using the images left after the cloud filter, we calculated the three water indexes: the Normalized Difference Water Index (NDWI) (S. K. McFeeters, 1995), the Modified NDWI (MNDWI) (Xu, 2006), and the Automated Water Extraction Index with no shadow (AWEI) (Feyisa et al., 2014).

Step 3. Initial Mask: Previous studies showed that MNDWI performed the best among the three water indices (Ji et al., 2009, Zhou et al., 2017, Jiang et al., 2014) and also recommended by Yu et al. (2020) method; the mask is created from this index. The MNDWI is filtered by the 'greater than' function. If an MNDWI pixel value is greater than 0, the function returns a value of 1. The filtered MNDWI images are reduced into one image using the *allNonZero* reducer command in the Google Earth Engine (GEE). The reducer assigns the pixel values of 0 or 1 based on all values obtained from every image in the collection at a specific location. If every value at that pixel location was a non-zero, then the final image has a pixel value of 1 at that location. If there is even

one 0 value in any image at that location, then the final image has a pixel value of 0. This creates an initial binary mask that is converted to polygons. The polygons are enlarged with a buffer of 5 pixels to capture values of pixels that surround bodies of water. The polygons are then rasterized, creating a new image called MNDWI Mask (MM). The MM is then applied to the NDWI, original MNDWI, and AWEI image collections.

Step 4. *Threshold Optimization*: This initiates a process called Otsu Segmentation which optimizes the threshold between foreground and background in images (water and not-water) (Otsu, 1979). The segmentation process creates a histogram of values from the locations the MM provided. The histogram should be bimodal with a buffer of 5 pixels, the two modes being water and non-water. The threshold between the two modes is found using the sum of squares. The Otsu Segmentation process occurs within each index's image collection separately.

Step 5. Majority Vote: The optimized NDWI, MNDWI, and AWEI image collections after Otsu Segmentation are then reduced into three images (one for each index) using the *allNonZero* reducer. This creates three separate images that show where each water index identified water in every image during the time period. The three reduced index images are then reduced to one image through the *Mode*/*Majority Vote* method. The mode method sees what value is present at that pixel location most often and assigns the final image that value. The final combined image depicts where water is in the majority of indexes and also where water is located in every image. This image is then converted into polygons and these polygons are given object-based features (OBFs). The OBFs include the Iso-Perimetric Quotient, Solidity, Patch Fractal Dimensions, Convexity, and Square Pixel Metric. The Convexity, Solidity, and Path Fractal Dimensions describe basic geometric characteristics of the pond and how complex it may be (Jiao et al., 2012). The Iso-Perimetric Quotient measures how similar an object is to compact shapes like circles (Q. Yu et al.,

2006). Lastly, the Square Pixel Metric analysis for the shape convexity and is similar to the Iso-Perimetric Quotient (Frohn, 2007).

Step 6. Fishponds Detection: Two machine learning techniques are utilized following the Yu et al. (2020) recommendations: Classification and Regression Trees (CART) and Logistic Regression (LR). These two algorithms were run in R Studio utilizing the caret package (Kuhn, 2008). The algorithms are trained by analyzing the calculated OBFs of the ground-truthing data. The base method determined algorithm performance by using five-fold cross-validation.

The *Base Method* used images from the GEE platform to determine what defined a fishpond visually. The fishponds were then manually digitized for their study area to train their algorithms. In addition, non-fishpond water bodies were identified by removing digitized lakes from the Tibetan Plateau.

3.2.3 Proposed Improvements

We identified Yu et al. (2020) method as having the most promising and applicable methods for our region of interest in southern Bangladesh, but with some modifications. From initial surveys, we found that many of the assumptions in the Yu et al. (2020) paper do not apply to the south-west and south-central regions of Bangladesh (Table 11).

Table 11. Base Method assumptions, shortcomings, and alternative approaches

Base Method assumptions	Shortcoming of the *Base Method* and alternative approach	Improvement number
Fishponds are filled with water year-round	Some fishpond types in Bangladesh may dry out for a portion of the year (e.g., Homestead ponds). Other fishponds may be planted with rice for part of the year (e.g., Gher). Therefore, assuming all fishponds are filled with water all year is not correct. Instead of focusing on areas that have water for an entire year, we should focus on time periods when each type of fishpond holds water.	1, 3
Fishponds are surrounded by non-water	Bangladesh's landscape is very diverse, and in many areas, different types of land use can be found around fishponds such as trees, rice paddies, buildings. In some regions, fishponds are very close with very narrow boundaries (< 10 m apart, which is the highest resolution of imageries used here). This assumption impacts the buffer size for the MM.	2, 4
Fishponds are easy to detect visually from images. Non-fishponds water bodies from another region were selected for algorithm training.	There is a high probability that small fishponds cannot be detected correctly through visual observation of satellite imageries. The preferred method is to use ground-truthing fishpond collections from the study region that are diverse in size, shape, type, and surrounding areas.	5
CART and LR are the preferred machine learning techniques for this application	Support Vector Machine and Random Forest are identified as promising methods (Maxwell et al., 2018) for differentiating between fishponds and non-fishponds classes.	6

In order to address these shortcomings, the following changes were proposed to improve the *Base Method* (1) identify a period of time that all types of fishponds are filled with water, (2) modify the buffer size to improve differentiating fishponds from their surroundings, (3) change the image reducer and water index combination to enhance waterbody detection, (4) introduce a convolution filter for edge detection, (5) utilize Support Vector Machine (SVM) and Random

Forest (RF) as two additional machine learning techniques, and (6) utilize ground-truthing data for better algorithm training.

Improvement 1: As described earlier, the shortcoming of the *Base Method* in assuming year-round fishponds impoundment is not valid in our study region as many fishponds are dried for a certain part of the year. In addition, using many images for the entire year for a large study area, such as the one here, can significantly increase the computational time or prohibit the implementation of the *Base Method* for a large study area.

In order to address this shortcoming, we identified a period of time in which all types of fishponds are filled with water. For this period, two sets of analyses were performed. First using all images with less than 10% cloud cover over the span of one month. Secondly, in order to further reduce the computational needs for data processing in our large study region, a single image for each district during the month of interest was selected for the second set of analyses.

Using images for a single month, relevant satellite images were obtained for the period of October 11th to November 13th of 2020 instead of the full year. Therefore, the total number of images was reduced from 311 for the whole year observation to 43 for one month observation. For the second scenario, one single day during the month was selected. Since our study area is very large compared to the *Base Method*, each district has a different day that falls between October 11th and November 13th. The images chosen were based on visual inspection and clarity, such as cloud coverage and contrast between water and non-water. The days chosen for each district are October 13, 2020 for Barisal, October 28, 2020 for Bagerhat and Gopalganj, November 5, 2020 for Jessore and Satkhira, and November 7, 2020 for Bhola and Khulna.

Improvement 2: The *Base Method* assumes that fishponds are surrounded by non-water pixels, but this is not the case in some of the districts in our study region. The Satkhira and Khulna districts,

for example, have fishponds and agricultural fields that are very close together with dikes and buffers that are smaller than 10 m. We tested buffer sizes of 0, 1, 3, and 5 pixels to identify the best size as the buffer in the *Base Method* is a very important step in creating the MM. The MM is a binary raster that is one of the inputs into the Otsu Segmentation function. Where the MM has a value of 1, the Otsu Segmentation function creates a histogram of index values at those locations. The buffer selects an additional region around the areas where the MNDWI has identified water that is assumed to be land. The histogram created by Otsu should be bimodal, where one peak represents water and the other represents non-water. Having a bimodal histogram makes the threshold values from Otsu more accurate. The potential issue with a buffer of 5 pixels is that it assumes that the water bodies identified by MNDWI during the MM process are surrounded by 50 m of land, trees, or buildings. This is not the case in many of the districts in our study region. In areas where aquaculture ponds are being used, they tend to be very close together, with only a dike separating them. Dikes are much smaller than 50 m; some are even smaller than the 10 m pixel that we can see. Fishponds surrounded by other water bodies could change the histogram from bimodal to unimodal since most pixels identified by the MM are water. We wanted to see if the buffer size impacts the histogram and threshold value from the Otsu Segmentation. If that threshold impacts how much water we identify and whether they have clear boundaries for a better classification.

Improvement 3: The *Base Method* functions under the assumption that water is present in fishponds throughout the year and in every image. To implement this assumption, they utilize an all-or-nothing reducer called *allNonZero* and combine all three water index images. We know that some fishponds are drained for maintenance or may not be visible due to an imaging error (e.g., sensor issues, platform issues, or angle issues). With the all-or-nothing style of reducer, the *Base*

Method removes too many potential fishponds and under detects. Therefore, we tested an image reducer that would take the value of what is present in the majority of images instead of all.

The *allNonZero* reducer in the *Base Method* labels final image pixels with a value of 0 (or non-water) if even one image does not have water at that location. This is a very strict reducer that eliminates too much water during the pre-processing and establishes a lower bound estimate. It does not take into account potential cloud interference or the periods of time when ponds may be dried or have rice within them. The *allNonZero* reducer is used two times in the *Base Method*. The first instance is in creating the MM when combining all the MNDWI images. The second instance is when combining the calculated index images into one summary image for each index (Step 5: Majority Vote). The first instance of the *allNonZero* reducer remains the same since we want very strict water locations to create the MM. As an improvement, we replaced the second instance of the *allNonZero* reducer with the *Mode* reducer. The *Mode* reducer, instead of requiring water to be present for every image requires that water be present for most of the images. The *Mode* reducer allows for more water pixels to be identified in the final classification.

The *Base Method* combines images from all three water-identifying indexes together to create one final summary image. We assumed this muddles the shape and location of identified water bodies, decreasing the overall amount of detected water. To improve upon this, we tested four different index combinations to evaluate what performed best in each district. The combinations are the *Base Method*, AWEI individually, NDWI individually, and MNDWI individually. This test will show whether combining the three indexes negatively impacts the fishpond identification results or filters out likely non-fishpond water bodies.

Improvement 4: Increasing the number of water pixels, whether from reducing the time frame, changing the buffer, or using the *Mode* reducer, will cause many water bodies to merge, forming

much larger polygons. This problem is exacerbated by the issue of dikes and other dividers commonly being less than 10 m wide, much smaller than the best resolution images that are used in this study. These large polygons have a much higher chance of being labeled as a non-fishpond by the classifier because of their size and odd shapes. To try to reduce the size of these polygons, we introduced a Laplacian 5×5 convolution filter to smooth the images (Gao et al., 2018). The convolution filter was applied to the best index or index combination image from Improvement 3 of each district and enhances the differences between water and non-water. The convolution filter is used in ArcGIS Pro 2.4. We chose Laplacian 5×5 as the filter type since it visually performed better than the other filter types (e.g., Smoothing 3×3, Smoothing 5×5, Laplacian 3×3).

After the convolution filter is used, we put the smoothed index image through Otsu Segmentation with the MM mask. The process follows the same as the previously improved *Base Method*. This improvement aims to provide the Otsu Segmentation process with a clearer index image than the one provided by the *Base Method*.

Improvement 5: The Base Method assumes that fishponds are easy to detect visually from standard RGB images, but that may depend more on when the images were captured and what surrounds the fishpond. Many fishponds in Bangladesh are hard to detect visually because they are surrounded by trees or many rice paddies and they have different fishpond layouts. In the study region, four types of fishponds were identified that include gher with and without rice, commercial ponds, and homestead ponds. All four types of fishponds vary in average size, typical shape, and period with water, all highlighted in Table 12. Gher without rice is, on average, much larger than all the other fishpond types. Homestead ponds and gher without rice can have round and irregular shapes, whereas gher with rice and commercial ponds typically are rectangular. The period in which each of these fishpond types is filled with water varies as well.

Commercial and homestead ponds are typically permanent ponds that have water year-round. Gher can have water or not because of their design to cultivate both fish and crops throughout the year. All of these differences between fishpond types make it very difficult to locate a diverse range of fishponds visually from satellite images.

Table 12. Average size, shape, and period with water for gher with and without rice, commercial ponds, and homestead ponds.

Characteristic	Gher with rice	Gher without rice	Commercial pond	Homestead pond
Average size (m^2)	1620-2020	4047	1620	810
Typical Shape	Rectangular	Rectangular, but could have curved edges	Rectangular	Round
Period with water	May - December	February - December	Year-round	Year-round
Predominant farming system	Freshwater prawn, fish, rice, and vegetables	Shrimp and fish	Fish	Fish

The *Base Method* utilizes historical GEE images to visually find and trace fishponds. The traced fishponds are not directly used for training, but instead, they used the shape of the polygon of water identified at the traced locations through the *Base Method* for training. For their non-fishpond data, they manually trace water bodies from a Tibetan plateau for algorithm training using the Joint Research Centre (JRC) Yearly Water Classification dataset that has 30-m spatial resolution (Pekel et al., 2016). Tibet is north of Bangladesh and has a very different elevation and climate.

In comparison, for this study, we used 1,728 ground-truthing polygons provided by surveyors in Bangladesh to validate the *Base Method's* performance and the proposed

improvements. The ground-truthing data consists of 996 fishponds and 741 non-fishpond waterbodies (Table A2 and Figure A2). Using ground-truthing data additionally makes validation results reliable instead of trusting that the waterbody chosen visually was a fishpond or not. The ground-truthing data was created by manually drawing borders around the known fishponds and non-fishponds in GEE. This results in the training dataset having smoother edges than what the water identification process will be. This is because the water identification process takes the shape of pixel groupings that appear blocky with hard edges. For this simplification of shape to not impact the results, we used the 'Simplify Polygon' function in ArcGIS to simplify the polygons identified through Steps 1-4.

Improvement 6: The two machine learning algorithms used in the *Base Method* are CART and LR due to their wide use and applications in addition of being transcribed into GEE. In addition to these algorithms, SVM and RF were identified as promising methods and also recommended by Maxwell et al. (2018), considering many factors such as training data, requirements, and computational cost. SVM was selected because it is useful for finding an optimal boundary between two classes (Maxwell et al., 2018) and performs reliably when trained with a smaller dataset (Ramezan et al., 2019). RF is a larger combination of many decision trees and is easy to optimize, so it is reasonable to compare its results with the CART method (Shi & Yang, 2016).

3.2.4 Evaluation Criteria for Comparing Fishpond Detection Algorithms

To compare the performance of the different fishpond detection algorithms, we used several criteria, including:

1) The number of ground-truthing fishponds that are correctly identified,
2) The percentage of the ground-truthing area that is correctly identified by the classifier
3) The number of fishponds classified,

4) The recall: Recall is the ratio of true positive fishponds to the total of all known fishponds, whether true positive or false negative (Equation (8)).
5) The precision: Precision is the ratio of true positive fishponds to all identified fishponds, whether true positive or false positive (Equation (9)).
6) The F1 score: The F1 score (Equation (10)) is the harmonic mean of the recall and precision and is used to provide a more insightful characteristic of performance than the arithmetic mean (Sasaki, 2007).

The range for recall, precision, and F1 score is 0.0 to 1.0, with 1.0 being the highest score for all of them.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$\text{F1 score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{10}$$

To determine whether the proposed improvements significantly impacted the results, we compared the areas that are correctly identified between the *Base Method* and the proposed improvement method against the ground-truthing fishpond areas. The comparison was made using a confidence interval for the mean relative error between the identified area and the known ground-truthing area (Abramowitz & Stegun, 1972).

3.3 Results and Discussion

3.3.1 Base Method Performance Evaluation within the Study Area in Detecting Waterbodies and Fishponds

The performance of the *Base Method* was evaluated based on 1) how well it identified water bodies and 2) how well it classified those water bodies as fishponds. Figure 3 shows the percentage of the ground-truthing area that are correctly identified at pre-classifier (waterbody identification) and post-classifier (fishpond identification) stages. First, concerning detecting water areas within the known fishponds, the highest detection was in Bagerhat and had 9.4% of ground-truthing fishpond area overlap. The second highest was Gopalganj with 5.9%, but the rest of the districts saw less than 1% ground-truth fishpond area detection. Second, regarding the applicability of classifiers to differentiate between fishponds and non-fishponds, the LR method was identified more ground-truth fishponds correctly than CART. The performance post-classifier will always be the same or less than the waterbody identification since the classifiers are only applied to the areas identified as water. Bagerhat had the highest correct classification out of the seven districts. Most of the other districts had correct classifications under 1%, except Gopalganj had around 5% classification after the LR classifier (Table A3). The CART classification true positives were lower than the LR, with the highest being Bagerhat's 5% ground-truth area identification. In Bagerhat, the *Base Method* using the LR classifier correctly identified 16 ground-truthing fishponds out of 235 (9% of the total ground-truthing fishpond area) (Figure 7). Having said that, still the performance of the *Base Method,* even in the Bagerhat, is low.

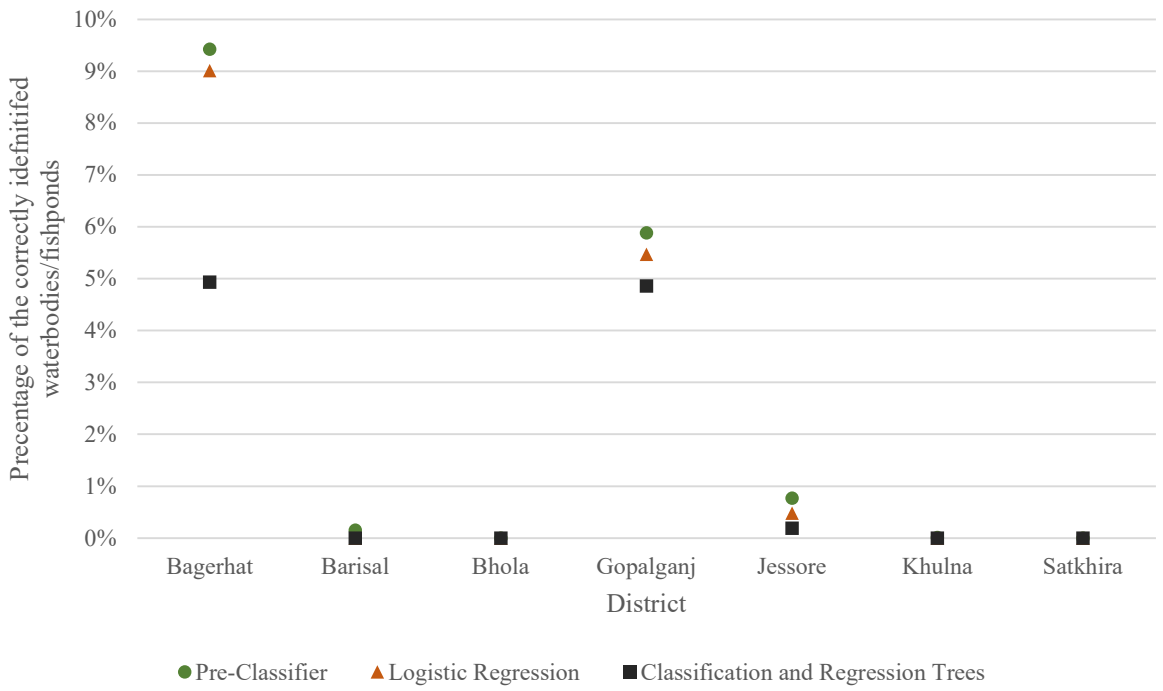

Figure 7. Percentage of the ground-truthing area that were correctly identified as waterbodies, classified as fishpond using logistic regression method, and classified using the classification and regression trees method for different districts within the study area.

Meanwhile, the Bagerhat district is the only one of the seven in which the results are significant enough to visualize. Therefore, in Figure 8, we visually compared the performance of the classifiers that were used in the *Base Method*. While CART correctly identifies fewer ground-truthing fishponds (Figure 8C), LR over-classifies water bodies as fishponds (Figure *8*B). Also notable in Figure 8 are the large polygons that are identified as water and how those polygons are classified by LR and CART. LR tends to classify the large polygons as water, whereas CART is more conservative and classifies them as non-fishponds. The polygons are much larger than the ground-truthing fishponds in the area, and in some cases, overlap the ground-truthing fishponds. These large polygons make individual fishpond detection difficult since they typically encompass multiple fishponds or several agricultural lands (e.g., rice paddies).

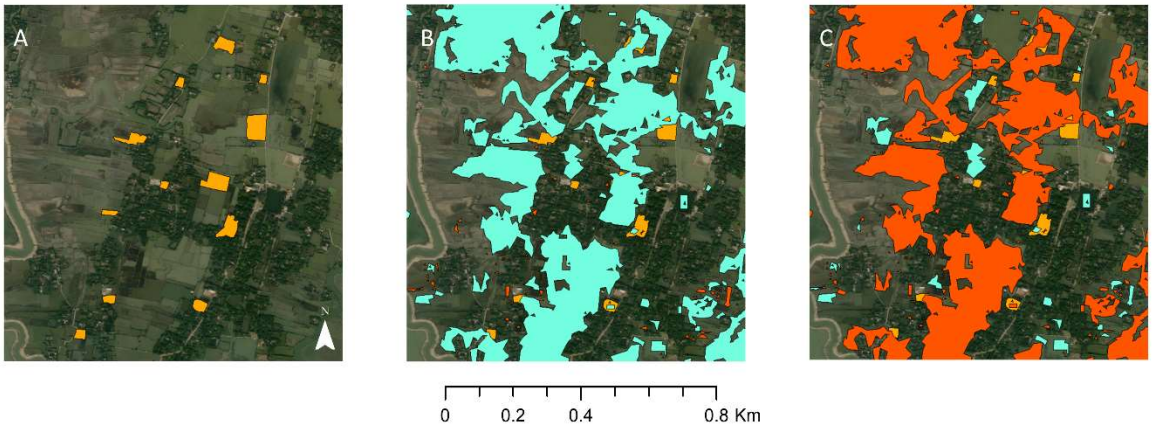

Figure 8. Example of the Base Method results in Bagerhat. Orange polygons represent ground-truthing fishponds. Blue polygons are classified as fishponds. Red polygons are classified as non-fishponds. (A) Ground-truthing fishponds. (B) Logistic Regression classification results. (C) Classification and Regression Trees classification results. (Centroid of the pictures above: 22° 35' 59.4096" N, 89° 34' 57.414" E)

The threshold optimization (Step 4) was created to pinpoint the water bodies most likely to be fishponds. For the small study area that the *Base Method* was applied, this technique works very well, especially since the layout of their study area has clear divided water bodies. However, even broken into the seven districts in our study area, the pixel selection technique does not perform to the same standard. The underlying assumptions of the *Base Method* are not valid for our study area and must be adjusted to make accurate fishpond identification possible.

Overall, the *Base Method* is under-identifying water bodies before the classifier is even applied. In addition, the *Base Method*'s performance in detecting fishponds is very low, the highest of all districts being 9%, and can be enhanced significantly. In the following sections, we will evaluate the effectiveness of proposed improvements on waterbody and fishpond detections in a very large and diverse study area.

3.3.2 Identifying the Best Period of Image Collection for Detecting Fishponds (Improvement 1)

We tested one year, one month, and one day time periods for image collection to evaluate how the number of images impacts the threshold optimization results. The one year period combined information for 311 images while one month lowered the number of images to 43. These totals are for the study area as a whole and not for each specific district. Table 13 compares the waterbody identification results from each time period. Rows 1, 3, and 5 highlight the ground-truth fishpond area percentage identified and how it improves with the shorter time for every district. The district of Bhola performed the worst out of all districts with 0% water identification for both year and month periods and 3% for the day. The poor performance in Bhola could be explained by the low numbers of ground-truthing data (27 locations) we have for the district. Generally, the fewer ground-truthing fishponds we have, the harder it is to identify any of them specifically. Bagerhat had the best water identification during the one-month period with 52% of the ground-truthing fishpond area identified, but Jessore had the best performance in a single day with 91% area identification. The districts of Barisal, Gopalganj, Khulna, and Satkhira all saw significant improvement over the *Base Method*, but performed worse than the Bagerhat and Jessore values (Table A4-Table A10). All seven districts improved water identification when we changed the time period from a year to either a month or a day. The single-day images had the highest results out of the three periods.

Table 13. The percentage of ground-truthing fishponds area and total ground-truthing fishponds correctly identified in all districts using different lengths for image processing (i.e., the entire year, one month, one day).

Period of image collection	Performance criteria	District						
		Bagerhat	Barisal	Bhola	Gopalganj	Jessore	Khulna	Satkhira
One year	Percentage of GT* fishpond area identified pre-classifier	9.4%	0.2%	0%	6%	1%	0.01%	0%
	GT* fishponds identified	25 of 235	1 of 168	0 of 27	7 of 77	13 of 113	1 of 163	0 of 208
One month	Percentage of GT* fishpond area identified pre-classifier	52%	4.3%	0%	22%	44%	36%	66%
	GT* fishponds identified	139 of 235	8 of 168	0 of 27	44 of 77	70 of 113	71 of 163	75 of 208
One Day	Percentage of GT* fishpond area identified pre-classifier	67%	16.3%	3%	36%	91%	67%	85%
	GT* fishponds identified	182 of 235	28 of 168	1 of 27	66 of 77	99 of 113	112 of 163	115 of 208

*Ground truthing

Changing the period from the full 2020 year to just October 11 to November 13, 2020 proved to be effective in identifying more water than the *Base Method*. The *Base Method* utilizes the *allNonZero* reducer in GEE to identify water that is present year-round. A longer time period means more images to compare and more chances for water to not be present. All districts saw an improvement in water detection from the *Base Method* and from the month period when using an image for one day. Figure 9 shows the drastic difference in water detection between the three periods, but also how small boundaries between ponds and agriculture can blend together to form large polygons. Figure 9 also compares a location in which we had good water detection, Jessore, to an area with very poor water detection, Bhola. The method performs much better when using a

shorter time frame with fewer images. The best period proved to be the single day. We do not expect that the day chosen will drastically impact results, but the image must have low cloud coverage (e.g., less than 10%) and it must be during a time period when we are fully confident fishponds contain water. Because of these guidelines and Bangladesh's climate, we are only able to obtain a few images that fit these standards. To ensure that the best time period is a day, we used both single day and single month for Improvements #2 and #3.

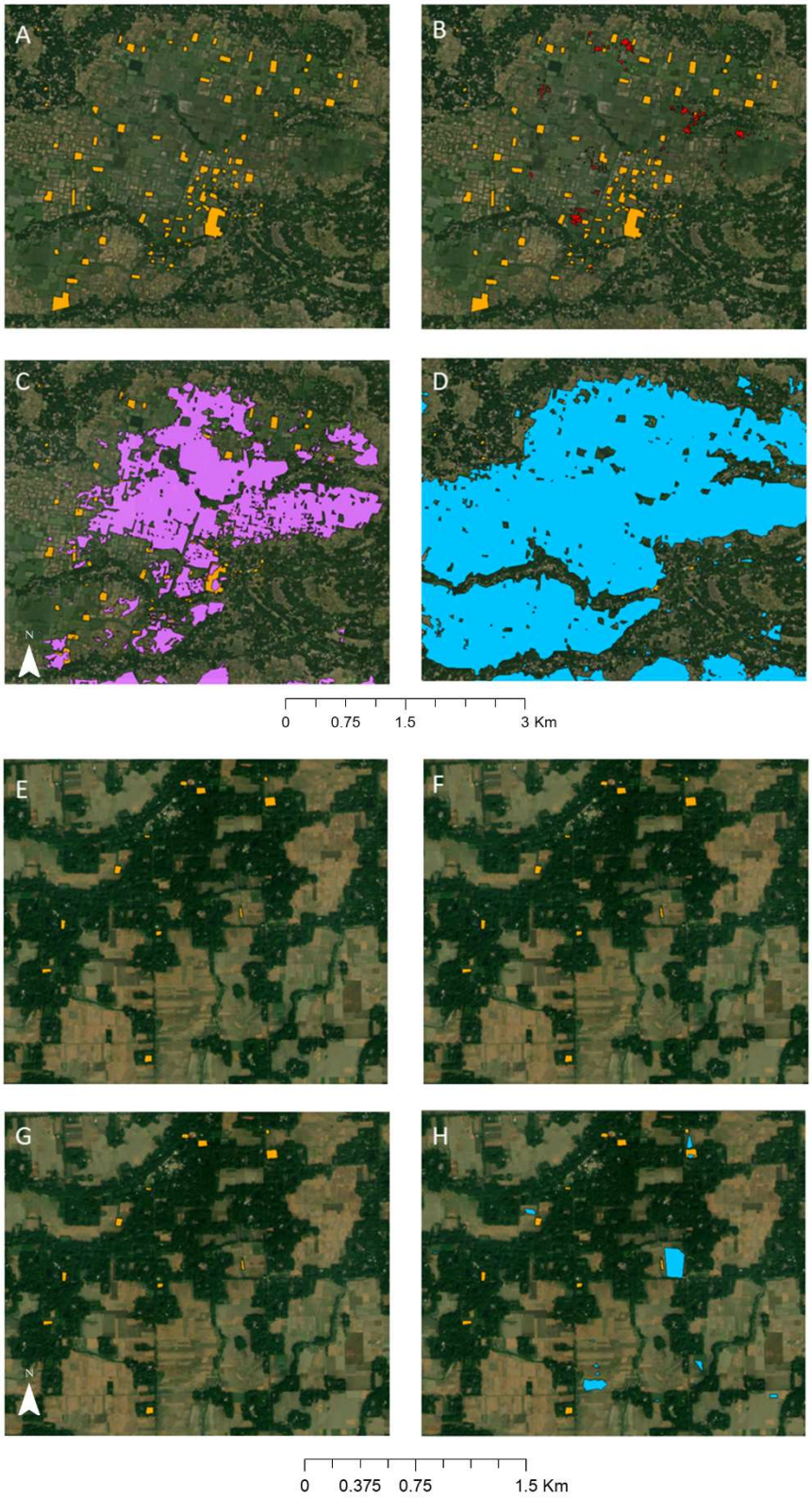

Figure 9. Visual comparison of the three-time periods in southern Jessore and southern Bhola. (A) Jessore ground-truthing fishponds (orange). (B) Jessore ground-truthing fishponds and full

Figure 9 (cont'd)
year water (red). (C) Jessore ground-truthing fishponds and one-month water (purple). (D) Jessore ground-truthing fishponds and single-day water (blue). (Centroid: 22° 57' 49.9248" N, 89° 16' 46.3188" E). (E) Bhola ground-truthing fishponds (orange). (F) Bhola ground-truthing Fishponds and full year water (red). (G) Bhola ground-truthing Fishponds and one-month water (purple). (G) Bhola ground-truthing Fishponds and single-day water (blue). (Centroid: 22° 10' 7.0464" N, 90° 41' 27.546" E)

3.3.3 Testing the Buffer Size for Threshold Optimization (Improvement 2)

The buffer size impacts the amount of water identified by the Threshold Optimization process. The larger the buffer, the more pixel values are included in determining the threshold value between water and non-water. Having the best possible threshold will significantly improve upon water and fishpond identification. We tested buffer sizes of 5, 3, 1, and 0 pixels within the MM for both single day and one month time periods. The water identification pre-classifier for Bagerhat with a buffer size of 5 was 67% for one day, but that dropped to 58% with a buffer size of 0 (Table A11). This was the same trend across all districts for both time frames except for Bhola (Table A11-Table A17). None of the attempted methods have resulted in any success in Bhola higher than the 3% of ground-truthing area identified pre-classifier achieved from the 5-pixel buffer for a single day. The results showed that the buffer of five pixels performed better than any other size buffer for both time periods.

Figure 10 shows the histogram of NDWI values at the MM location for Khulna. The final results for each district are from using the *Base Method*, in which all three indices are combined. Khulna was chosen for the example since it had a noticeable change from the buffer size and NDWI is one of the most used water indices. Khulna saw a large decrease in ground-truth water identification from 67% to 59% when changing the 5-pixel buffer to 0-pixel for a single day. Khulna saw a similar decrease from 36% to 27% for 5-pixel to 0-pixel buffer for the month period. The histograms for 0- and 1-pixel buffers have large peaks between -0.2 and -0.1 that represent non-water pixels. As the buffer size increases, the second peak around 0.3 to 0.4 increases

considerably. With NDWI, it is commonly interpreted that values above 0.3 are water bodies and values between 0 and 0.3 may be water bodies (Stuart K. McFeeters, 2013). These histograms show that with the increased buffer size, water pixels become more prominent. The 5-pixel buffer may work best because it includes more pixels in total. The more pixels included in the histogram, the more reflective the histogram is to the land use in the district. Additionally, the 5-pixel buffer works better in districts with small fishpond boundaries because it provides more opportunities for endmember pixels, or pure water pixels, to be accounted for.

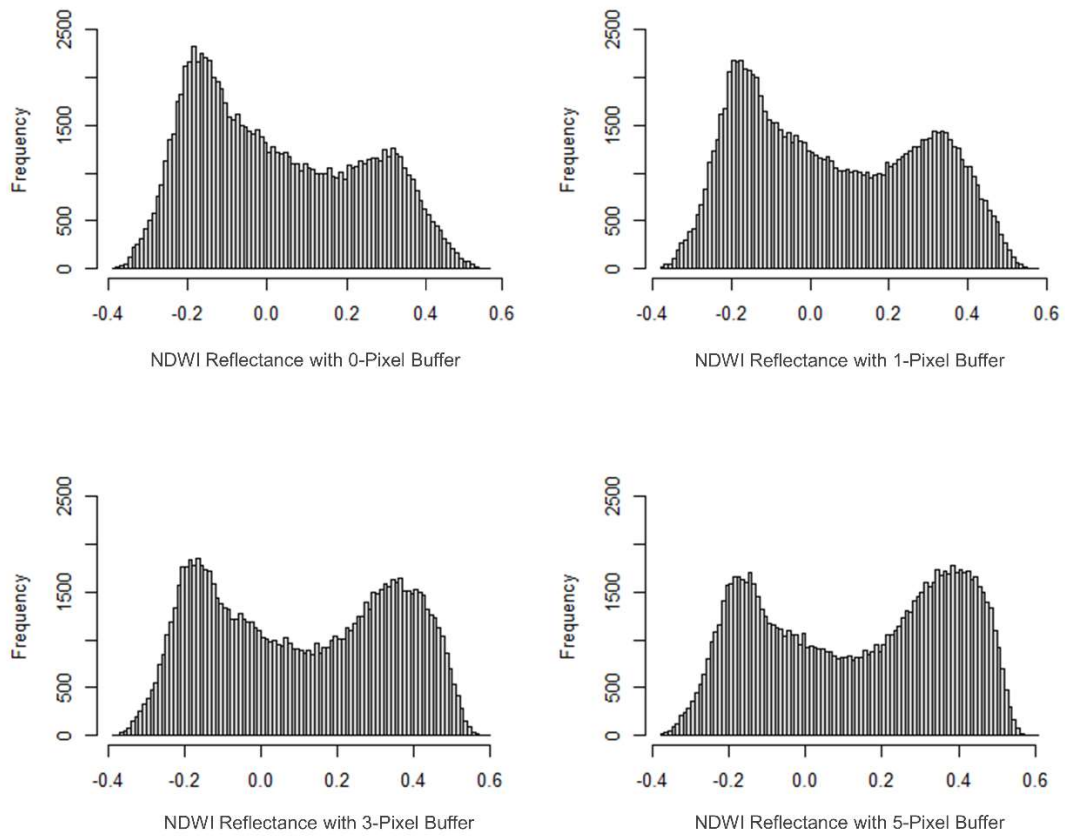

Figure 10. NDWI reflectance values for MM in Khulna for a single-day image. Top left to bottom right: 0-pixel buffer, 1-pixel buffer, 3-pixel buffer, 5-pixel buffer

3.3.4 Determining the Combination of Image Reducer and Water-Identifying Index to Improve Waterbody and Fishpond Detection (Improvement 3)

Both the image reducer and water-identifying indices impact the amount of water identified for classifying. The *Base Method* resulted in too few of the ground-truthing fishponds being identified. Therefore, we hypothesized that determining the best combination of image reducer and indices will significantly improve the overall detections. To increase the amount of water identified, we tested two image reducers (i.e., *Mode*and *allNonZero*) and four different water-identifying index combinations (i.e., AWEI, MNDWI, NDWI, and combination of AWEI; MNDWI; and NDWI). The *Mode* reducer increased the area percentage of ground-fishponds identified for all districts (Table A18-Table A24) compared to the *allNonZero* reducer (Table A25-Table A31). The exception to this is Bhola, in which most of the results were 0% except for the NDWI index in which the Mode reducer identified 7%.

Figure 11 shows the comparison of the combined indexes, AWEI, MNDWI, and NDWI in correctly identifying the ground-truth fishpond areas for each district. The combination of all three indexes that represents the *Base Method* did not perform the best for any district. Figure 11 also shows that the MNDWI by itself also does not perform the best for any district (Table A32-Table A38).

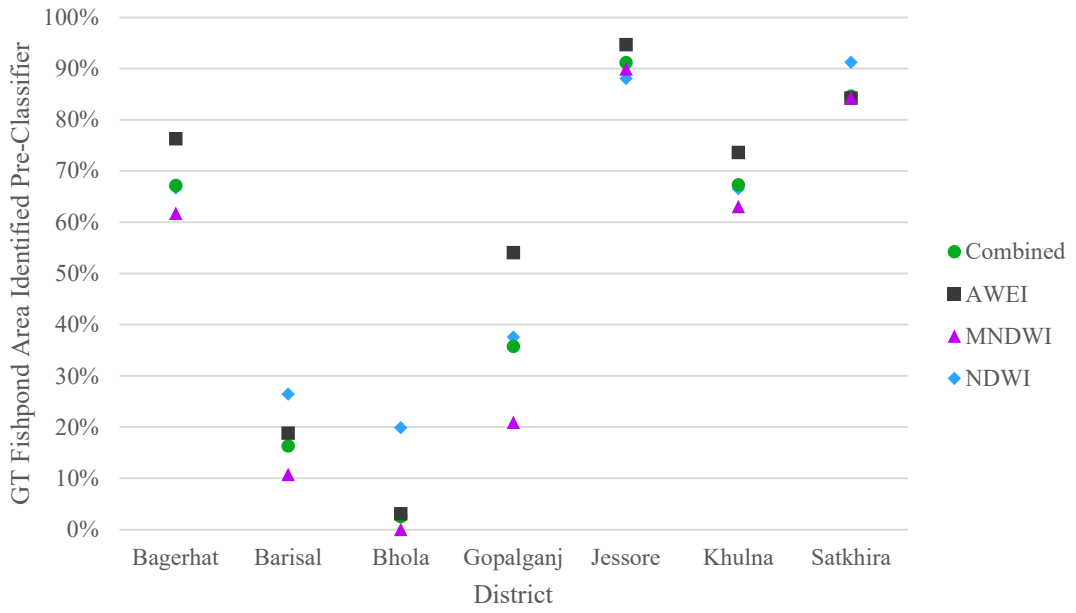

Figure 11. Ground-truthing (GT) fishpond area identified pre-classifier per district for the combined index, only AWEI, only MNDWI, and only NDWI results using the single day images

Testing the image reducer type proved again that the single day time period was one of the best improvements. Testing the *Mode* reducer on the month-long period resulted in less water identification than the single day (Table A18-Table A24), but more than the same period using the *allNonZero* reducer (Table A25-Table A31). Meanwhile, the *Mode* reducer did improve water identification, but since single-day images only have utilized one image, the reducer has no impact on its performance.

3.3.5 Implementing Edge Detection with a Convolution Filter to Improve Fishpond Boundary Detection (Improvement 4)

Considering the results from *Improvement 1* through *3*, the best combination of time of image collection, buffer size for the threshold optimization, and water index for each district was identified. Two major combinations are as follows: (1) single day, 5-pixel buffer, with AWEI index for Bagerhat, Gopalganj, Jessore, and Khulna and (2) single day, 5-pixel buffer, with NDWI index for Barisal, Bhola, and Satkhira (Table A39). Since the best combinations for every district entail

using the single day image, the reducer type is no longer relevant. A potential reason why MNDWI does not perform to the same standards as NDWI and AWEI is that MNDWI was created to mainly differentiate water from developed land (Wicaksono & Wicaksono, 2019), which is not the case within our study area. In fact, the majority of the representative fishponds (ground-truth) are not located near significant developed land. Our preliminary literature review also supported this assumption that is most likely the case with the majority of fishponds in the country.

In the next step, we apply a Laplacian 5×5 convolution filter for edge detection to identify the best image combination for each district. The purpose of introducing edge detection is to begin with an enhanced image that highlights sharp differences in pixel values. The difference in pixel values shows where boundaries are located around different plots of land. The convolution filter emphasized the differences in the select index images to identify the boundaries around water. The spectrum of index values widens after the filter has been applied, allowing for differences to become more apparent numerically. For example, the original range of NDWI values in Barisal is -0.50 to 0.65. After using the filter, the range lengthens from -14.83 to 8.95. Visually, the filter highlights boundaries but does make many of the pixels become a very similar grey color (Figure 12B).

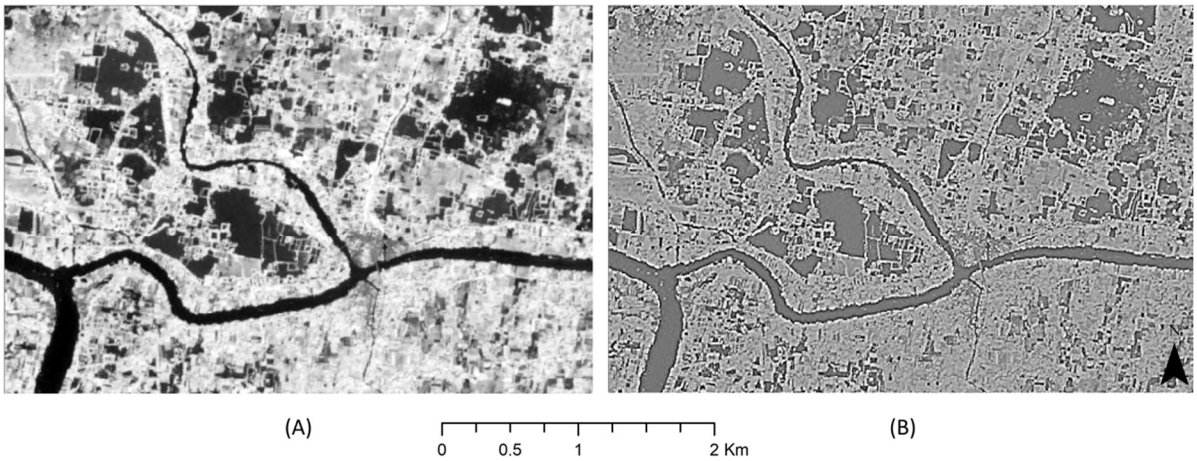

Figure 12. NDWI image in western Barisal (A) before and (B) after the Laplacian 5×5 convolution filter was applied (Centroid: 22° 51' 38.37" N, 90° 5' 53.28" E)

The results of applying the Laplacian 5×5 convolution filter were mixed; however, every district saw a large increase in the overall number of fishponds identified, with some now seeing over 100,000 fishponds (Table A40). Bagerhat, Bhola, Gopalganj, Jessore, and Khulna had all their ground-truthing fishponds identified before classification. Jessore performed the best with edge detection seeing all 113 ground-truthing fishponds and 91.5% of the area correctly identified before any classifier was used. Despite Jessore's high performance, the water body identification pre-classifier decreased with adding edge-detection from 95% to 91.5%. Satkhira experienced a similar result from adding edge-detection with the ground-truth area identified decreasing from 91% to 82%. Oppositely, Gopalganj, Barisal, and Bhola all increased their ground-truthing fishpond area identification at the pre-classifier stage from 54% to 68%, 26% to 58%, and 20% to 57%, respectively. The same district divide was seen with the post-classifier results as well. Barisal and Bhola saw an increase in their post-classifier area identification for both LR and CART. The ground-truthing fishpond area percentage for Jessore decreased from 94.3% to 91.4% with LR and from 3% to 0% with CART using edge-detection. Satkhira saw a slight increase from 0.3% to 5% for the CART classifier, but its LR dropped from 93% to 56% ground-truth area identification.

3.3.6 Evaluating the Impact of Ground-Truthing Data on Machine Learning Training (Improvement 5)

To train machine learning classifiers, the *Base Method* manually traced suspected fishponds from historical satellite images. The water bodies they traced may have only been representative of one type of fishpond because homestead ponds and gher with rice are very difficult to identify as fishponds. To improve the machine learning classifier training, we used detailed ground-truthing data for all four types of ponds described previously. In addition to this, we had ground-truthing data for non-fishpond water bodies that would provide juxtaposing training data. Table 14 compares the performance of the machine learning classifiers when they are trained with ground-truthing data versus historical imagery that was used in the Yu et al. (2020). The recall significantly increased from 0.538 to 0.989 for LR and from 0.495 to 0.827 for CART. This shows that the ground-truthing data is decreasing the number of false negatives. The precision score decreased from 0.788 for LR and 0.773 for CART to 0.594 and 0.738, respectively, when using the ground-truthing data for machine learning training. A decrease in precision implies an increase in the number of false positives. The increases in recall and F1 score show that utilizing ground-truthing data improves the performance of both the LR and CART classifiers, but the decrease in precision increases the number of false positives compared to the *Base Method*.

Table 14. Comparing statistics from Yu et al. (2020) machine learning training with historical imagery to our ground-truthing data training.

Training method	Performance Criteria	Logistic Regression	Classification and Regression Trees
Yu et al. (2020) Historical imagery	Precision	0.788	0.773
	Recall	0.538	0.495
	F1 Score	0.640	0.604
Ground-truthing data	Precision	0.594	0.738
	Recall	0.898	0.827
	F1 Score	0.715	0.780

3.3.7 Adding Random Forest and Support Vector Machine Classifiers to Determine the Best Classifier for the Data (Improvement 6)

In addition to LR and CART, RF and SVM are widely recognized classifiers and can be used for many different research applications that require the differentiation of two or more classes (Agmalaro et al., 2021), Yan et al., 2021, Iordache et al., 2020). To improve the classification results, we added RF and SVM since they are identified as promising classifiers (Maxwell et al., 2018). Overall, the LR and SVM classifiers identified more ground-truthing fishponds both in terms of area percentage and total number than the CART and RF (Table A41 and Table A42). Figure 13 shows how the results of the RF and CART are similar to each other as well as how similar the results from LR and SVM are. The district with the best performance is Jessore, with both LR and SVM classifying 91% of the ground-truthing fishpond area and all 113 ground-truthing fishponds correctly. Jessore is also where there is the most drastic difference in performance by classifiers was observed. This could be due to the predominant presence of large water polygons in Jessore. LR and SVM tend to classify the large polygons are fishponds, whereas RF and CART do not. The RF and CART classifiers in Jessore performed the worst out of any district with 3% ground-truthing fishpond area identification and 3 of 113 fishponds correctly identified. Introducing two additional classifier types yielded similar results to the *Base Method*.

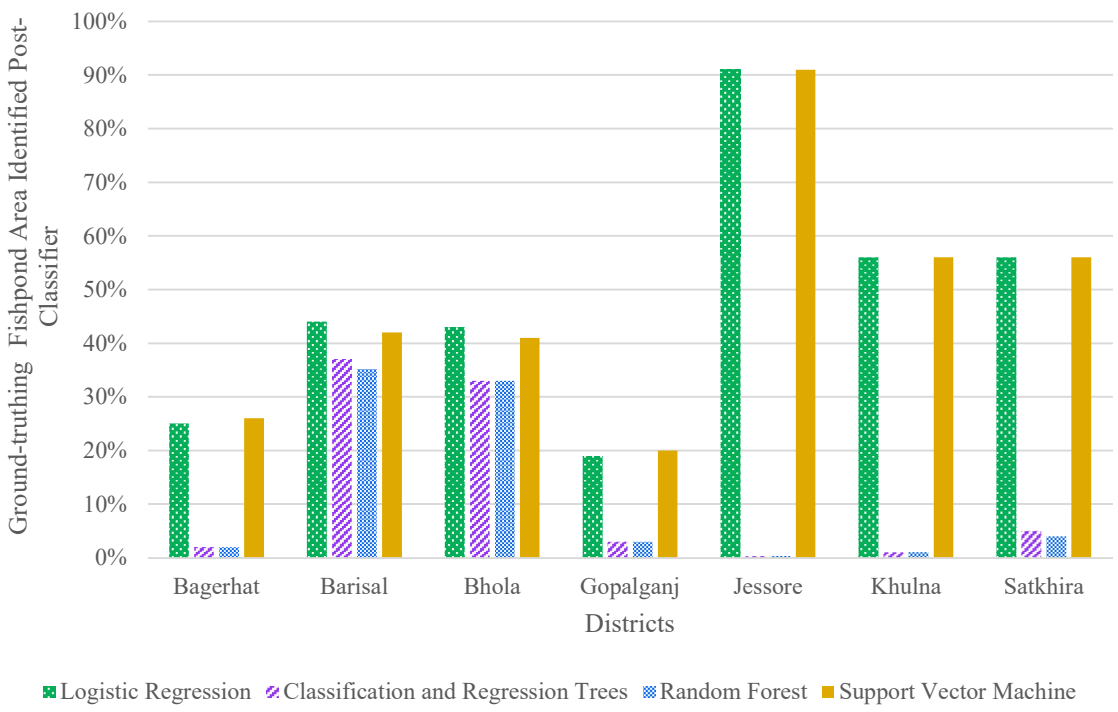

Figure 13. Ground-truthing fishpond area percentage identified by each classifier type.

Table 15 compares the precision, recall, and F1 scores for each of the four classifier types we used. The F1 scores are all very similar to each other, with the range from lowest value (0.706) to highest (0.780) being only 0.074. The LR recall and precision results are interesting because it has the worst precision, but the best recall meaning the LR is very good at preventing false negatives, but the worst at preventing false positives. RF has the exact opposite results with the highest precision and the lowest recall. Both LR and RF have lower F1 scores. CART and SVM fall between LR and RF for precision and recall, but because of their consistency, they have the two higher F1 scores. From these scores, we can conclude that LR is over-classifying fishponds and RF is under-classifying fishponds. When looking at the results in Figure 13, we can conclude that since SVM performed very similarly to LR, it must also be over-classifying. In addition, since RF typically identifies less than CART and RF under-classifies, CART can be the best final option.

Table 15. Comparing the training performance of the four classifiers on the validation ground-truthing data.

Performance Criteria	Logistic Regression	Classification and Regression Trees	Random Forest	Support Vector Machine
Precision	0.594	0.738	0.784	0.720
Recall	0.898	0.827	0.645	0.822
F1 Score	0.715	0.780	0.706	0.768

The total number of classified fishponds for each district also supports the conclusion that CART is the best middle ground (Table A41 and Table A42). Figure 14 shows the classifier results from LR, CART, RF, and SVM in eastern Gopalganj as compared to the ground-truth fishponds in that location. Since the data shows that LR is over-classifying, we can conclude SVM must also be since SVM classifies 3 to 5 times as many fishponds as LR does. This conclusion is also backed by the fact that LR and SVM are very similar classifiers, so it is realistic that they would behave similarly (Zhang et al., 2003). The CART classifier numbers always fall below LR and above RF levels. This could also support our previous finding that CART provides the best classification out of the four.

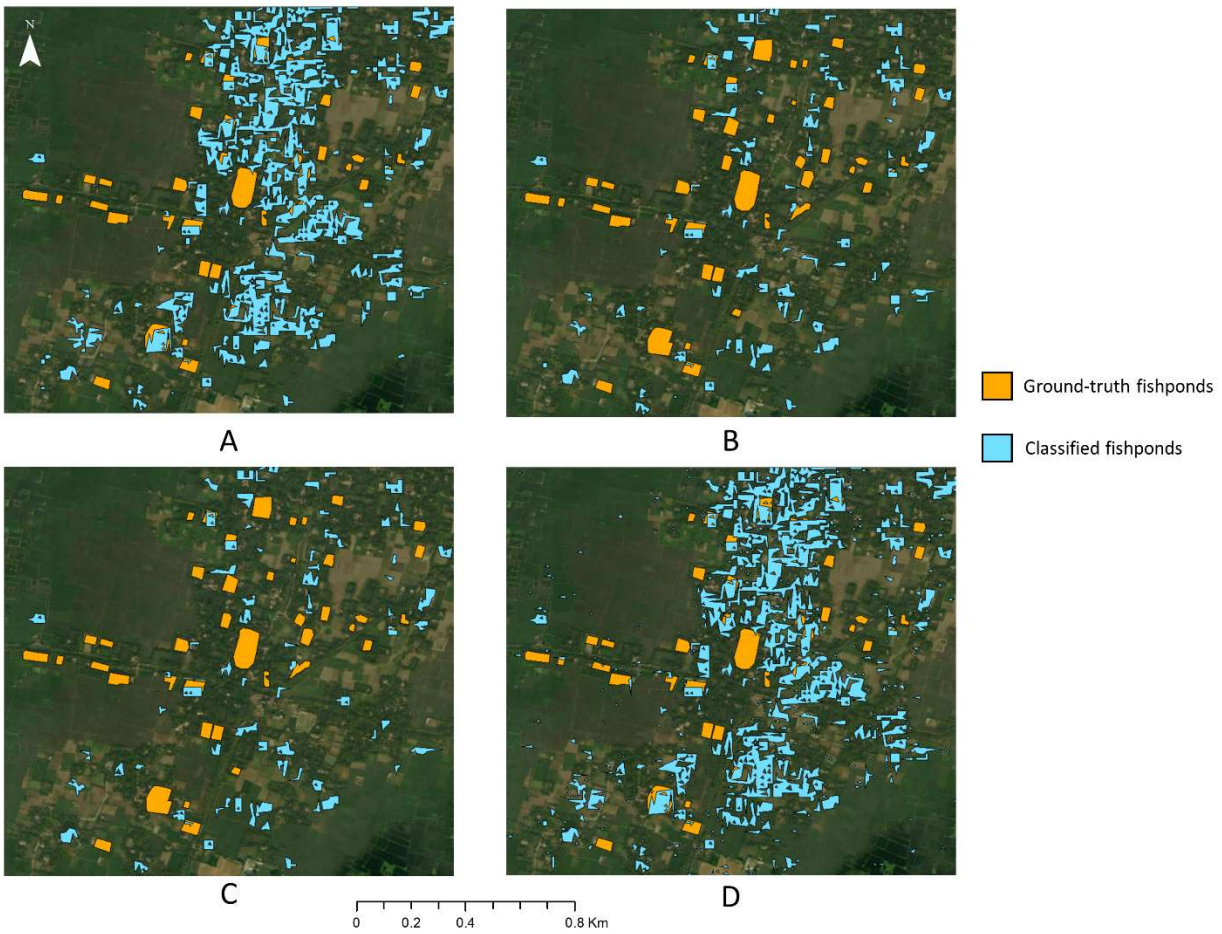

Figure 14. Comparison example of classified fishponds in eastern Gopalganj using (A) Logistic Regression, (B) Classification and Regression Trees, (C) Random Forest, and (D) Support Vector Machine. Ground truthing fishponds are in orange and classified fishponds are in blue. (Centroid: 23° 0' 17.9136" N, 90° 3' 46.8936" E).

3.3.8 The Overall Enhancement in Fishpond Detection as Results of the Proposed Improvements based on Medium-Resolution and High-Resolution Imagery

We calculated the mean relative error for four different scenarios to determine whether the six proposed improvements considerably enhanced the capability of the *Base Method* for detecting fishponds. The scenarios that are considered here are the *Base Method* before classification, the *Enhanced Method* (considering all improvements) before classification, the *Base Method* using LR, and the *Enhanced Method* using CART. Table 16 outlines these statistics. The mean relative difference closest to 0 is the *Enhanced Method* without classifier (lower and upper limits of -34.4%

and -31.3%). This means that the *Enhanced Method* is regularly 31-34% under detecting the ground-truthing area. The other three scenarios had mean relative differences very close to -100%. The *Enhanced Method* CART mean relative difference was better than for the *Base Method* pre- and post-LR. This shows that the *Enhanced Method* without classification considerably improves upon the *Base Method* water identification.

Table 16. The lower and upper limits for the mean relative difference of the Based and Enhanced Methods performances before and after a machine learning classifier.

Scenarios	Mean relative difference	
	Lower limit	Upper limit
Base Method before classifier	-99.4%	-98.5%
Enhanced Method before classifier	-34.4%	-31.3%
Base Method with Logistic Regression	-99.7%	-97.7%
Enhanced Method with Classification and Regression Trees	-91.5%	-88.6%

Having access to high-resolution imagery can improve the results of the *Enhanced Method* significantly. A WorldView-2 image was obtained for an area of 45 km^2 in Jamalnagar, Satkhira to analyze the performance of the *Enhanced Method*. The high-resolution image was used to identify the boundaries around all land plots and intersected those boundaries with the final *Enhanced Method* for Satkhira. Figure 15 compares the boundary identification from the WorldView-2 image, the *Base Method* water identification, the *Enhanced Method* water identification, and the intersect of the WorldView-2 boundaries and the *Enhanced Method*. We applied the CART classification to the intersect of WorldView-2 and the *Enhanced Method* to compare it to the *Enhanced Method* fishpond identification alone. There are 47 ground-truth

fishponds in this region. The *Enhanced Method* correctly identified 9 of the 47 whereas the intersect correctly identified 29 of the 47 fishponds. In addition, the area of fishponds correctly identified was doubled with the introduction of the high-resolution boundaries. The improvement in fishpond and area identification in one small region shows that having access to high-resolution imagery would refine the results of the *Enhanced Method* even further. In summary, using high-resolution imagery would significantly improve boundary detection around fishponds and provide more accurate water locations.

Figure 15. Impacts of high-resolution imagery in Jamalnagar, Satkhira. (A) High-resolution boundary identification, (B) Base Method water identification, (C) Enhanced Method water

Figure 15 (cont'd)
identification, and (D) Intersect of Enhanced Method and high-resolution boundaries. (Centroid: 22° 34' 57.6624" N, 89° 12' 12.1536" E)

3.3.9 Land Use and Ground Truth Fishpond Characteristics to Explain Trends in Results

All seven districts within the study region are very different from each other. This makes creating one over-arching method very difficult. We looked at the land use surrounding the ground-truth fishponds to see if there was a trend in districts that performed better/worse than others. To do this, we tested buffer sizes of 10 m, 50 m, and 100 m around ground-truth fishponds in each district separately (Table A43-Table A45). The predominant land use types in all districts were cultivated land, water bodies, and artificial surfaces. Bhola and Barisal had the lowest percentage of ground-truthing fishpond area identified before the machine learning classifier. These two districts had the highest percentage of artificial surfaces, with Bhola seeing 73% and Barisal having just under 50% with the 100 m buffer. All other districts had artificial surfaces as 26% or less of their land use at 100 m. Water and artificial surfaces may have very similar reflectance values (Worden & de Beurs, 2020). In this instance, since there was a very high percentage of artificial surfaces and a very low percentage of water, the indexes may be identifying artificial surfaces instead of water during Otsu Segmentation.

The median size ground-truthing fishpond also varied significantly by district (Table A46). The district with the smallest median size ground-truthing fishpond was Bhola (713 m^2), followed by Barisal (833 m^2), then Satkhira (1252 m^2). All other districts had median ground-truthing fishponds larger than 2,200 m^2. The combination of having the least and smallest ground-truthing fishponds may be why the improved methods performed the worst in Bhola. Barisal had more

ground truth fishponds with 168, but they were additionally much smaller than other districts. This could explain why the *Enhanced Method* performed worse in Barisal than other districts.

3.4 Conclusion

Aquaculture is becoming increasingly more important in regions that relies heavily on fish for food, but capture fisheries are producing less and less. Identifying where aquaculture is occurring is critical in understanding the industry, how it operates, and the areas it can improve in. With this information, we can estimate total production, inform policy, or analyze trends in aquaculture locations.

This study showed that the approach taken for identifying fishponds must be tailored to the characteristics of the fishponds, the fishponds' surrounding areas, and the satellite data being used. The best approach for this is to test the time period for image capture, the buffer size for image optimization with Otsu Segmentation, the combination of water-identifying indexes, the image reducer type, and the machine learning classifier type. Utilizing edge-detection techniques and high-resolution imageries can improve the overall waterbody detection. For south-west and south-central Bangladesh, limiting the number of images for a specific time period results in an increased number of waterbodies detected. One reason for this is that additional errors can be introduced when combining multiple images together if an image does not capture the period with water. The buffer size for threshold optimization of water and non-water had an enormous impact on the results. The 5-pixel buffer performed the best for our study region. This proves that including more pixels in image thresholding results in better thresholds altogether, especially when the shape and size of the fishponds change throughout the year. Of the three water indices, NDWI and AWEI performed the best for the region, but this varies on location and land use/cover characteristics. The *Mode* reducer did allow for more waterbodies to be identified than the *allNonZero* image

reducer when working with more than one image for the study region. The image reducer type, in the end, did not matter for the *Enhanced Method* since one image is usually selected for fishpond detection. The training dataset for machine learning plays an important role in identifying the best classifier. Having diverse and representative ground-truthing fishponds will improve the machine learning classifiers.

Some of the challenges that were observed in this study include 1) size of the study area: due to the size of the study area and the limitation of the GEE platform computational power, many of the tasks proposed in this study could not be executed for the entire region. One way to address this issue was to divide the study area into seven districts and apply various tasks separately for each district; 2) threshold optimization: in the original *Base Method*, one threshold value for water detection was obtained for the study region; however, our results showed that this approach does not work in a large region. Therefore, in this study, we optimized the threshold values based on the water index values at the district level. One alternative is to train the system based on the region's physiographical and climatological characteristics instead of its political boundaries. Using agroclimotological zone is one alternative but is not a good representation of the fishpond distribution and attributes; and 3) high-resolution imagery: Our study showed that using high-resolution imagery can significantly improve the overall performance of the *Enhanced Method* in detecting fishponds. Having high-resolution imagery would be very beneficial in identifying the boundaries between fishponds and breaking up any large polygons. As the cost for high-resolution imagery decreases, these six strategies will become increasingly more useful and 4) rainy season: The rainy and monsoon season in Bangladesh is from June through mid-October. This limits the availability of high-quality images with low cloud coverage. In fact, our study showed that the viable images for this study are mainly available from January to May and October to December.

One solution can be the combined usage of SAR and optical imaging in identifying small waterbodies. Ultimately, verification of the *Enhanced Method* in other regions is crucial to examine the reliability of the proposed method.

4 OVERALL CONCLUSION

Aquaculture in Bangladesh is growing very rapidly and plays an important role in addressing agriculture and food security. Therefore, mapping and classifying fishponds is essential for economic purposes and understanding how land use is changing. Better statistics on aquaculture come from two different methods: surveys or remote sensing. Surveys are time-consuming and costly to be a viable option for the large-scale study. Meanwhile, utilizing remote sensing data can be a good alternative as it can cover a large area and can be implemented in a platform such as GEE that is available for free. However, the products from remote sensing do not have the same accuracy as the survey and should be calibrated based on ground-truthing data.

The research found that the process used to identify fishponds must be specific to the region where it is applied, as the complexity of the region does not allow for the development of the universal fishpond detection algorithm. Therefore, we developed six strategies for identifying fishponds that can build off each other and can be applied broadly. The findings from each of these strategies are presented below:

(1) When determining the best time-period for image identification, it was concluded that the fewer images analyzed, the higher the fishpond identification was. This relies on knowing when the typical fishpond has water and the climate of the region to determine the best single image to use.

(2) To optimize the threshold between water and non-water, using a larger buffer size around waterbodies yielded higher fishpond identification.

(3) Regardless of the region of study, several water-identifying indexes should be tested to determine whether a single index or combination of indices should be selected to detect existing waterbodies.

(4) In general, detecting fishpond shapes in Bangladesh can be challenging as the boundary widths around fishponds are generally smaller than the resolution of the optical imagery. Therefore, utilizing edge-detection techniques can increase the number of waterbodies identified and can be used to differentiate between land plots.

(5) Performing a limited but targeted ground-truthing survey can help the overall performance of machine learning techniques.

(6) Due to spatial and temporal variabilities of fishpond types and shapes in Bangladesh, not a single machine learning technique could be identified as the best. Therefore, it is recommended to test different machine learning classifiers to ensure the selection of the most robust technique for different regions (e.g., district).

5 FUTURE RESEARCH RECOMMENDATIONS

This research provided six strategies to improve fishpond identification in the south-west and south-central Bangladesh. However, due to the limitations of using medium-resolution imagery, the results cannot be further improved as the boundary of the fishponds in the region is smaller, in many cases, than the resolution of the satellite imagery. Therefore, additional research should be mainly focused on fishpond boundary detection. Below, more specific recommendations are provided for future studies:

- Apply the six improvement strategies in different Southeast Asian countries with similar fishpond prevalence as this study to evaluate the overall robustness of the proposed strategies.
- Adjust the proposed improvement strategies with high-resolution imagery to better detect fishpond shapes and characteristics.
- Explore the applicability of other satellite imagery and datasets as inputs to the proposed fishpond detection strategies. Among the existing imageries, SAR seems to be promising as it has been used for detecting waterbodies. In addition, the problem with cloudiness is not impacting the quality of the imagery as radar sensors can see through clouds.
- Improve the threshold optimization through better identification of fishpond zones (e.g., land use/land cover) within a study area. As the utilized boundary for this study (i.e., district) does not capture the spatial variabilities of fishpond settings.

APPENDIX

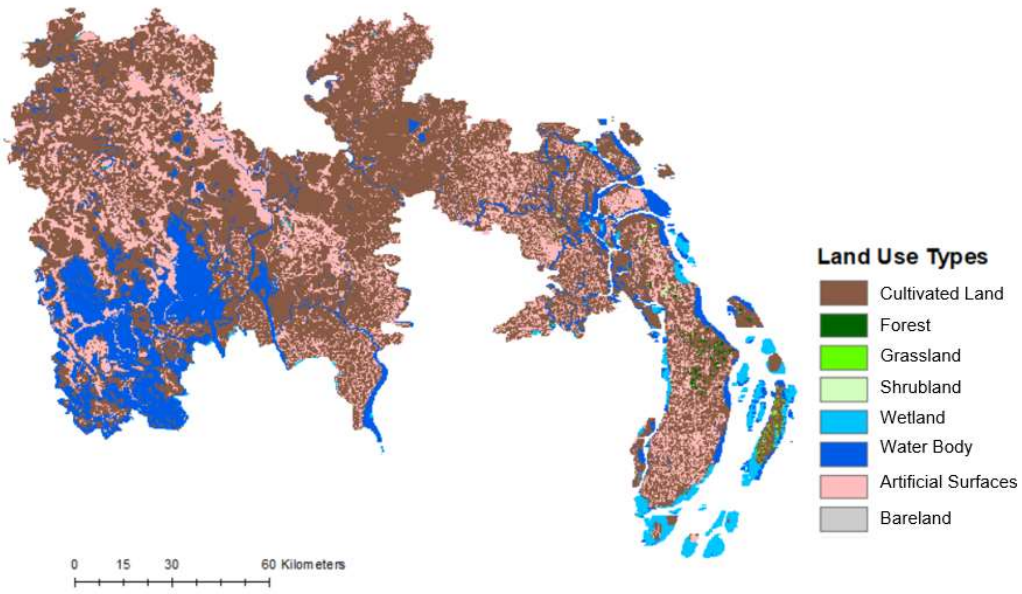

Figure A1. Land use map of the study region (adapted from region (China Ministry of Natural Resources, 2020))

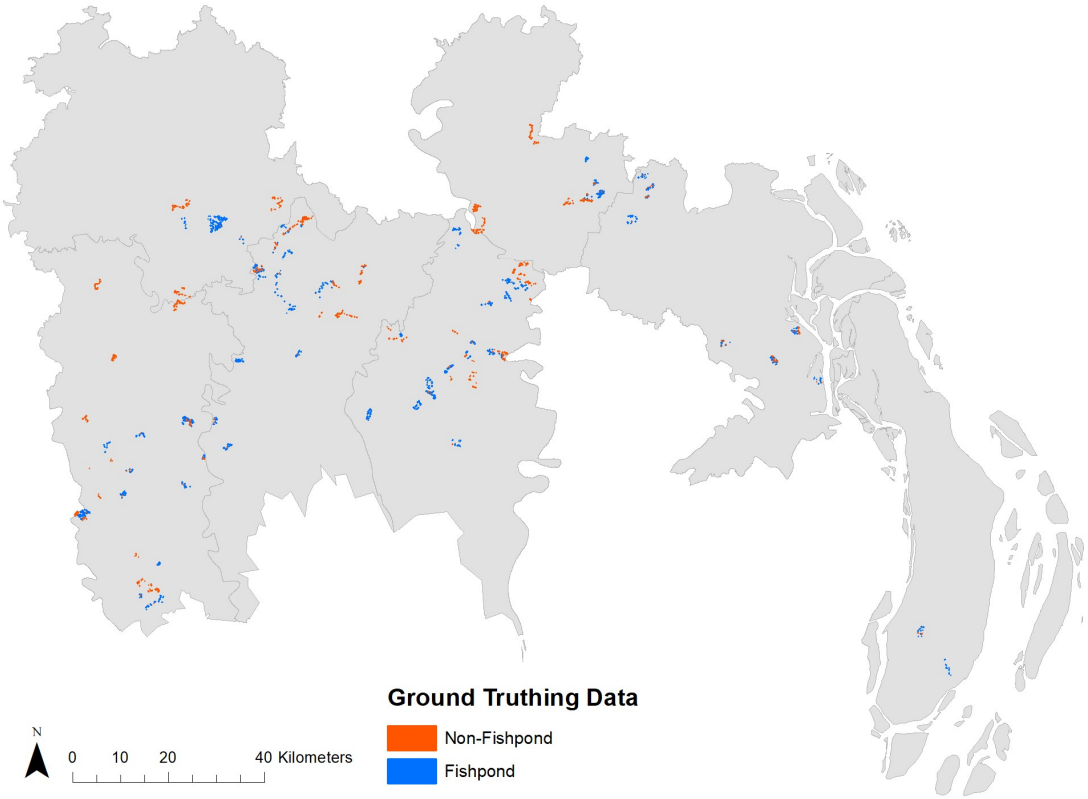

Figure A2. Distribution of Ground Truthing Data locations by type.

Table A1. Land cover as percentage of area for each district within the study region (China Ministry of Natural Resources, 2020)

Land Cover	Percentage of area within each District						
	Bagerhat	Barisal	Bhola	Gopalganj	Jessore	Khulna	Satkhira
Cultivated Land	69.9	60.7	52.5	84.1	67.4	56.0	38.2
Forest	0	0.1	2.0	0	0	0	0
Grass Land	0	0.1	1.0	0	0	0	0
Shrub Land	0	0	0.3	0	0	0	0
Wetland	0.4	0.7	13.3	0.1	0.2	0.4	0.1
Water Body	6.3	9.2	7.0	2.2	3.0	22.6	39.0
Artificial Surfaces	23.4	29.1	23.8	13.6	29.4	21.0	22.7
Bareland	0	0.1	0.1	0	0	0	0

Table A2. Numbers of ground-truthing data fishponds and non-fishponds by district.

District	Number of fishponds	Number of non-fishponds
Bagerhat	235	144
Barisal	168	32
Bhola	27	2
Gopalganj	77	131
Jessore	113	111
Khulna	163	146
Satkhira	208	175

Table A3. Base method comparison for all seven districts.

District	Bagerhat		Barisal		Bhola		Gopalganj		Jessore		Khulna		Satkhira	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**
Percentage of GT*** fishpond area identified pre-classifier	9%	9%	0.2%	0.2%	0%	0%	6%	6%	0.8%	0.8%	0%	0%	0%	0%
GT*** fishponds identified pre-classification	25 of 235	25 of 235	1 of 168	1 of 168	0 of 27	0 of 27	7 of 77	7 of 77	13 of 113	13 of 113	1 of 163	1 of 163	0 of 208	0 of 208
Percentage of GT*** fishpond area identified post-classifier	9%	5%	0%	0%	0%	0%	6%	5%	0.5%	0.2%	0%	0%	0%	0%
GT*** fishponds identified post-classification	16 of 235	5 of 235	0 of 168	0 of 168	0 of 27	0 of 27	3 of 77	2 of 77	4 of 113	2 of 113	0 of 163	0 of 163	0 of 208	0 of 208

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A4. Time period comparison for Bagerhat.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	9%	9%	52%	52%	67%	67%
GT*** fishponds identified pre-classification	25 of 235	25 of 235	139 of 235	139 of 235	182 of 235	182 of 235
Percentage of GT*** fishpond area identified post-classifier	9%	5%	51%	19%	66%	16%
GT*** fishponds identified post-classification	16 of 235	5 of 235	117 of 235	49 of 235	159 of 235	51 of 235

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A5. Time period comparison for Barisal.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	0.2%	0.2%	4%	4%	16%	16%
GT*** fishponds identified pre-classification	1 of 168	1 of 168	8 of 168	8 of 168	28 of 168	28 of 168
Percentage of GT*** fishpond area identified post-classifier	0%	0%	4%	4%	13%	12%
GT*** fishponds identified post-classification	0 of 168	0 of 168	3 of 168	4 of 168	11 of 168	10 of 168

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A6. Time period comparison for Bhola.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	0%	0%	0%	0%	3%	3%
GT*** fishponds identified pre-classification	0 of 27	0 of 27	0 of 27	0 of 27	1 of 27	1 of 27
Percentage of GT*** fishpond area identified post-classifier	0%	0%	0%	0%	3%	3%
GT*** fishponds identified post-classification	0 of 27	0 of 27	0 of 27	0 of 27	1 of 27	1 of 27

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A7. Time period comparison for Gopaglanj.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	6%	6%	22%	22%	36%	36%
GT*** fishponds identified pre-classification	7 of 77	7 of 77	44 of 77	44 of 77	66 of 77	66 of 77
Percentage of GT*** fishpond area identified post-classifier	6%	5%	19%	11%	30%	23%
GT*** fishponds identified post-classification	3 of 77	2 of 77	19 of 77	6 of 77	36 of 77	22 of 77

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A8. Time period comparison for Jessore.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	1%	1%	44%	44%	91%	91%
GT*** fishponds identified pre-classification	13 of 113	13 of 113	70 of 113	70 of 113	99 of 113	99 of 113
Percentage of GT*** fishpond area identified post-classifier	0.5%	0.2%	43%	6%	91%	3%
GT*** fishponds identified post-classification	4 of 113	2 of 113	59 of 113	9 of 113	90 of 113	7 of 113

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A9. Time period comparison for Khulna.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	0%	0%	36%	36%	67%	67%
GT*** fishponds identified pre-classification	1 of 163	1 of 163	71 of 163	71 of 163	112 of 163	112 of 163
Percentage of GT*** fishpond area identified post-classifier	0%	0%	33%	9%	65%	5%
GT*** fishponds identified post-classification	0 of 163	0 of 163	60 of 163	19 of 163	95 of 163	15 of 163

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A10. Time period comparison for Satkhira.

Time Frame	Year		Month		Day	
Classifier Type	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	5	5	5	5
Percentage of GT*** fishpond area identified pre-classifier	0%	0%	66%	66%	85%	85%
GT*** fishponds identified pre-classification	0 of 208	0 of 208	75 of 208	75 of 208	115 of 208	115 of 208
Percentage of GT*** fishpond area identified post-classifier	0%	0%	66%	1%	84%	1%
GT*** fishponds identified post-classification	0 of 208	0 of 208	72 of 208	9 of 208	103 of 208	14 of 208

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A11. Buffer size comparison for Bagerhat for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART**	LR*	CART**	LR*	CART***	LR*	CART**	LR*	CART***	LR*	CART***	LR*	CART***	LR*	CART***
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT**** fishpond area identified pre-classifier	52%	52%	50%	50%	45%	45%	42%	42%	67%	67%	66%	66%	61%	61%	58%	58%
GT**** fishponds identified pre-classification	139 of 235	139 of 235	136 of 235	136 of 235	111 of 235	111 of 235	95 of 235	95 of 235	182 of 235	182 of 235	180 of 235	180 of 235	157 of 235	157 of 235	141 of 235	141 of 235
Percentage of GT**** fishpond area identified post-classifier	51%	19%	49%	17%	44%	0%	41%	13%	66%	16%	65%	15%	60%	13%	57%	15%
GT**** fishponds identified post-classification	117 of 235	49 of 235	107 of 235	48 of 235	87 of 235	0 of 235	80 of 235	39 of 235	159 of 235	51 of 235	156 of 235	50 of 235	138 of 235	46 of 235	131 of 235	44 of 235

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A12. Buffer size comparison for Barisal for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART*	LR*	CART*	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT*** fishpond area identified pre-classifier	4.3%	4.3%	3%	3%	1%	1%	1%	1%	16%	16%	13%	13%	10%	10%	6%	6%
GT*** fishponds identified pre-classification	8 of 168	8 of 168	7 of 168	7 of 168	4 of 168	4 of 168	2 of 168	2 of 168	28 of 168	28 of 168	22 of 168	22 of 168	12 of 168	12 of 168	8 of 168	8 of 168
Percentage of GT*** fishpond area identified post-classifier	4%	4%	2%	3%	1%	0%	0%	1%	13%	12%	12%	8%	9%	6%	5%	4%
GT*** fishponds identified post-classification	3 of 168	4 of 168	3 of 168	4 of 168	1 of 168	0 of 168	0 of 168	1 of 168	11 of 168	10 of 168	11 of 168	6 of 168	9 of 168	3 of 168	4 of 168	3 of 168

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A13. Buffer size comparison for Bhola for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT*** fishpond area identified pre-classifier	0%	0%	0%	0%	0%	0%	0%	0%	3%	3%	1%	1%	0%	0%	0%	0%
GT*** fishponds identified pre-classification	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	1 of 27	1 of 27	1 of 27	1 of 27	0 of 27	0 of 27	0 of 27	0 of 27
Percentage of GT*** fishpond area identified post-classifier	0%	0%	0%	0%	0%	0%	0%	0%	3%	3%	0%	0%	0%	0%	0%	0%
GT*** fishponds identified post-classification	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	1 of 27	1 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A14. Buffer size comparison for Gopalganj for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT*** fishpond area identified pre-classifier	22%	22%	20%	20%	13%	13%	9%	9%	36%	36%	35%	35%	32%	32%	30%	30%
GT*** fishponds identified pre-classification	44 of 77	44 of 77	42 of 77	42 of 77	24 of 77	24 of 77	13 of 77	13 of 77	66 of 77	66 of 77	65 of 77	65 of 77	62 of 77	62 of 77	55 of 77	55 of 77
Percentage of GT*** fishpond area identified post-classifier	19%	11%	18%	13%	8%	0%	9%	6%	30%	23%	30%	23%	28%	21%	28%	18%
GT*** fishponds identified post-classification	19 of 77	6 of 77	19 of 77	9 of 77	8 of 77	0 of 77	7 of 77	3 of 77	36 of 77	22 of 77	38 of 77	21 of 77	36 of 77	20 of 77	37 of 77	15 of 77

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A15. Buffer size comparison for Jessore for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT*** fishpond area identified pre-classifier	44%	44%	38%	38%	22%	22%	18%	18%	91%	91%	91%	91%	89%	89%	86%	86%
GT*** fishponds identified pre-classification	70 of 113	70 of 113	62 of 113	62 of 113	43 of 113	43 of 113	35 of 113	35 of 113	99 of 113	99 of 113	97 of 113	97 of 113	91 of 113	91 of 113	90 of 113	90 of 113
Percentage of GT*** fishpond area identified post-classifier	43%	6%	37%	6%	21%	0%	17%	6%	91%	3%	90%	1%	88%	1%	86%	1%
GT*** fishponds identified post-classification	59 of 113	9 of 113	55 of 113	13 of 113	33 of 113	0 of 113	26 of 113	10 of 113	90 of 113	7 of 113	90 of 113	5 of 113	88 of 113	6 of 113	85 of 113	5 of 113

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A16. Buffer size comparison for Khulna for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART*	LR*	CART*	LR*	CART*	LR*	CART*	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART*
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT*** fishpond area identified pre-classifier	36%	36%	32%	32%	29%	29%	27%	27%	67%	67%	66%	66%	62%	62%	59%	59%
GT*** fishponds identified pre-classification	71 of 163	71 of 163	66 of 163	66 of 163	60 of 163	60 of 163	58 of 163	58 of 163	112 of 163	112 of 163	110 of 163	110 of 163	103 of 163	103 of 163	98 of 163	98 of 163
Percentage of GT*** fishpond area identified post-classifier	33%	9%	30%	8%	27%	0%	0%	7%	65%	5%	65%	6%	61%	5%	56%	9%
GT*** fishponds identified post-classification	60 of 163	19 of 163	56 of 163	19 of 163	48 of 163	0 of 163	0 of 163	15 of 163	95 of 163	15 of 163	93 of 163	16 of 163	90 of 163	17 of 163	84 of 163	21 of 163

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A17. Buffer size comparison for Satkhira for both month and single day.

Time frame	Month		Month		Month		Month		Day		Day		Day		Day	
Classifier type	LR*	CART*	LR*	CART*	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART**	LR*	CART*
Buffer Size	5	5	3	3	1	1	0	0	5	5	3	3	1	1	0	0
Percentage of GT*** fishpond area identified pre-classifier	66%	66%	64%	64%	61%	61%	60%	60%	85%	85%	84%	84%	84%	84%	83%	83%
GT*** fishponds identified pre-classification	75 of 208	75 of 208	73 of 208	73 of 208	69 of 208	69 of 208	67 of 208	67 of 208	115 of 208	115 of 208	108 of 208	108 of 208	101 of 208	101 of 208	94 of 208	94 of 208
Percentage of GT*** fishpond area identified post-classifier	66%	1%	64%	2%	61%	1%	60%	5%	84%	1%	84%	1%	83%	1%	83%	1%
GT*** fishponds identified post-classification	72 of 208	9 of 208	69 of 208	9 of 208	65 of 208	7 of 208	62 of 208	10 of 208	103 of 208	14 of 208	101 of 208	13 of 208	91 of 208	6 of 208	91 of 208	6 of 208

* Logistic Regression
** Classification and Regression Trees
***Ground truthing

Table A18. *Mode* reducer applied for one month period in Bagerhat comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	64%	64%	74%	74%	57%	57%	62%	62%
GT[c] fishponds identified pre-classification	171 of 235	171 of 235	183 of 235	183 of 235	145 of 235	145 of 235	182 of 235	182 of 235
Percentage of GT[c] fishpond area identified post-classifier	63%	19%	73%	15%	56%	22%	60%	19%
GT[c] fishponds identified post-classification	141 of 235	59 of 235	174 of 235	37 of 235	129 of 235	57 of 235	142 of 235	51 of 235

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A19. *Mode* reducer applied for one month period in Barisal comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	12%	12%	14%	14%	8%	8%	17%	17%
GT[c] fishponds identified pre-classification	17 of 168	17 of 168	20 of 168	20 of 168	11 of 168	11 of 168	40 of 168	40 of 168
Percentage of GT[c] fishpond area identified post-classifier	11%	9%	13%	11%	8%	7%	13%	13%
GT[c] fishponds identified post-classification	12 of 168	6 of 168	15 of 168	10 of 168	7 of 168	5 of 168	14 of 168	13 of 168

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A20. *Mode* reducer applied for one month period in Bhola comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR [a]	CART [b]	LR [a]	CART [b]	LR [a]	CART [b]	LR [a]	CART [b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	0%	0%	0%	0%	0%	0%	7%	7%
GT[c] fishponds identified pre-classification	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	5 of 27	5 of 27
Percentage of GT[c] fishpond area identified post-classifier	0%	0%	0%	0%	0%	0%	0%	0%
GT[c] fishponds identified post-classification	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A21. *Mode* reducer applied for one month period in Gopalganj comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR [a]	CART [b]	LR [a]	CART [b]	LR [a]	CART [b]	LR [a]	CART [b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	30%	30%	48%	48%	14%	14%	33%	33%
GT[c] fishponds identified pre- classification	58 of 77	58 of 77	62 of 77	62 of 77	22 of 77	22 of 77	61 of 77	61 of 77
Percentage of GT[c] fishpond area identified post-classifier	24%	19%	46%	30%	10%	9%	26%	20%
GT[c] fishponds identified post- classification	25 of 77	14 of 77	49 of 77	26 of 77	10 of 77	7of 77	26 of 77	16 of 77

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A22. *Mode* reducer applied for one month period in Jessore comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	88%	88%	94%	94%	85%	85%	85%	85%
GT[c] fishponds identified pre- classification	95 of 113	95 of 113	98 of 113	98 of 113	87 of 113	87 of 113	98 of 113	98 of 113
Percentage of GT[c] fishpond area identified post-classifier	87%	1%	94%	3%	84%	0%	84%	3%
GT[c] fishponds identified post- classification	88 of 113	5 of 113	94 of 113	6 of 113	84 of 113	2 of 113	87 of 113	10 of 113

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A23. *Mode* reducer applied for one month period in Khulna comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	65%	65%	69%	69%	61%	61%	64%	64%
GT[c] fishponds identified pre- classification	104 of 163	104 of 163	107 of 163	107 of 163	93 of 163	93 of 163	115 of 163	115 of 163
Percentage of GT[c] fishpond area identified post-classifier	64%	7%	67%	6%	60%	9%	62%	6%
GT[c] fishponds identified post- classification	93 of 163	18 of 163	96 of 163	17 of 163	88 of 163	21 of 163	94 of 163	14 of 163

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A24. *Mode* reducer applied for one month period in Satkhira comparing water-identifying index combinations.

Index Type	Combined		AWEI[d]		MNDWI[e]		NDWI[f]	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	84%	84%	84%	84%	82%	82%	86%	86%
GT[c] fishponds identified pre- classification	100 of 208	100 of 208	102 of 208	102 of 208	93 of 208	93 of 208	126 of 208	126 of 208
Percentage of GT[c] fishpond area identified post-classifier	83%	1%	83%	1%	82%	1%	86%	1%
GT[c] fishponds identified post- classification	92 of 208	6 of 208	95 of 208	9 of 208	88 of 208	6 of 208	112 of 208	6 of 208

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A25. *allNonZero* reducer applied for one month period in Bagerhat comparing water-identifying index combinations.

Time frame	Month	Month	Month	Month	Month	Month	Month	Month
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	52%	52%	60%	60%	48%	48%	52%	52%
GT[c] fishponds identified pre-classification	139 of 235	139 of 235	156 of 235	156 of 235	122 of 235	122 of 235	157 of 235	157 of 235
Percentage of GT [c] fishpond area identified post-classifier	51%	19%	60%	14%	47%	18%	50%	14%
GT [c] fishponds identified post-classification	117 of 235	49 of 235	134 of 235	46 of 235	108 of 235	50 of 235	115 of 235	49 of 235

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A26. *allNonZero* reducer applied for one month period in Barisal comparing water-identifying index combinations.

Time frame	Month	Month	Month	Month	Month	Month	Month	Month
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	4.3%	4.3%	6%	6%	3%	3%	6%	6%
GT[c] fishponds identified pre-classification	8 of 168	8 of 168	10 of 168	10 of 168	6 of 168	6 of 168	18 of 168	18 of 168
Percentage of GT[c] fishpond area identified post-classifier	4%	4%	6%	5%	2%	2%	4%	5%
GT[c] fishponds identified post-classification	3 of 168	4 of 168	5 of 168	4 of 168	3 of 168	3 of 168	5 of 168	6 of 168

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A27. *allNonZero* reducer applied for one month period in Bhola comparing water-identifying index combinations.

Time frame	Month	Month	Month	Month	Month	Month	Month	Month
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	0%	0%	0%	0%	0%	0%	0%	0%
GT[c] fishponds identified pre-classification	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27
Percentage of GT[c] fishpond area identified post-classifier	0%	0%	0%	0%	0%	0%	0%	0%
GT[c] fishponds identified post-classification	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27	0 of 27

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A28. *allNonZero* reducer applied for one month period in Gopalganj comparing water-identifying index combinations.

Time frame	Month	Month	Month	Month	Month	Month	Month	Month
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	22%	22%	38%	38%	10%	10%	25%	25%
GT[c] fishponds identified pre-classification	44 of 77	44 of 77	56 of 77	56 of 77	14 of 77	14 of 77	53 of 77	53 of 77
Percentage of GT[c] fishpond area identified post-classifier	19%	11%	35%	26%	7%	3%	21%	16%
GT [c] fishponds identified post-classification	19 of 77	6 of 77	38 of 77	24 of 77	8 of 77	3 of 77	22 of 77	12 of 77

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A29. *allNonZero* reducer applied for one month period in Jessore comparing water-identifying index combinations.

Time frame	Month	Month	Month	Month	Month	Month	Month	Month
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	44%	44%	53%	53%	24%	24%	66%	66%
GT[c] fishponds identified pre-classification	70 of 113	70 of 113	74 of 113	74 of 113	41 of 113	41 of 113	86 of 113	86 of 113
Percentage of GT[c] fishpond area identified post-classifier	42.8%	6%	52.3%	10%	23.0%	4%	65.1%	10%
GT[c] fishponds identified post-classification	59 of 113	9 of 113	71 of 113	9 of 113	38 of 113	10 of 113	74 of 113	12 of 113

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A30. *allNonZero* reducer applied for one month period in Khulna comparing water-identifying index combinations.

Time frame	Month		Month		Month		Month	
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	36%	36%	43%	43%	33%	33%	31%	31%
GT[c] fishponds identified pre-classification	71 of 163	71 of 163	79 of 163	79 of 163	67 of 163	67 of 163	64 of 163	64 of 163
Percentage of GT[c] fishpond area identified post-classifier	33%	9%	42%	43%	32%	8%	28%	10%
GT[c] fishponds identified post-classification	60 of 163	19 of 163	69 of 163	79 of 163	62 of 163	21 of 163	47 of 163	17 of 163

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A31. *allNonZero* reducer applied for one month period in Satkhira comparing water-identifying index combinations.

Time frame	Month		Month		Month		Month	
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	66%	66%	69%	69%	63%	63%	62%	62%
GT[c] fishponds identified pre-classification	75 of 208	75 of 208	75 of 208	75 of 208	73 of 208	73 of 208	80 of 208	80 of 208
Percentage of GT [c] fishpond area identified post-classifier	66%	1%	69%	3%	63%	1%	61%	2%
GT [c] fishponds identified post-classification	72 of 208	9 of 208	72 of 208	6 of 208	70 of 208	8 of 208	72 of 208	9 of 208

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A32. Water-identifying index combination comparison for Bagerhat using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	67%	67%	76%	76%	62%	62%	67%	67%
GT[c] fishponds identified pre-classification	182 of 235	182 of 235	190 of 235	190 of 235	155 of 235	155 of 235	197 of 235	197 of 235
Percentage of GT[c] fishpond area identified post-classifier	66%	16%	76%	13%	61%	16%	65%	16%
GT[c] fishponds identified post-classification	159 of 235	51 of 235	181 of 235	30 of 235	138 of 235	50 of 235	164 of 235	54 of 235

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A33. Water-identifying index combination comparison for Barisal using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	16.3%	16.3%	19%	19%	11%	11%	26%	26%
GT[c] fishponds identified pre-classification	28 of 168	28 of 168	28 of 168	28 of 168	12 of 168	12 of 168	61 of 168	61 of 168
Percentage of GT[c] fishpond area identified post-classifier	13%	12%	15%	15%	10%	6%	20%	17%
GT[c] fishponds identified post-classification	11 of 168	10 of 168	13 of 168	12 of 168	8 of 168	3 of 168	21 of 168	16 of 168

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A34. Water-identifying index combination comparison for Bhola using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	3%	3%	3%	3%	0%	0%	20%	20%
GT[c] fishponds identified pre-classification	1 of 27	1 of 27	2 of 27	2 of 27	0 of 27	0 of 27	10 of 27	10 of 27
Percentage of GT[c] fishpond area identified post-classifier	3%	3%	3%	0%	0%	0%	12%	9%
GT[c] fishponds identified post-classification	1 of 27	1 of 27	1 of 27	0 of 27	0 of 27	0 of 27	4 of 27	3 of 27

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A35. Water-identifying index combination comparison for Gopalganj using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	36%	36%	54%	54%	21%	21%	38%	38%
GT[c] fishponds identified pre-classification	66 of 77	66 of 77	67 of 77	67 of 77	38 of 77	38 of 77	69 of 77	69 of 77
Percentage of GT[c] fishpond area identified post-classifier	30%	23%	50%	28%	15%	15%	31%	23%
GT[c] fishponds identified post-classification	36 of 77	22 of 77	56 of 77	26 of 77	20 of 77	13 of 77	37 of 77	22 of 77

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A36. Water-identifying index combination comparison for Jessore using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	91.2%	91%	95%	95%	90%	90%	88%	88%
GT[c] fishponds identified pre-classification	99 of 113	99 of 113	102 of 113	102 of 113	91 of 113	91 of 113	103 of 113	103 of 113
Percentage of GT [c] fishpond area identified post-classifier	90.7%	3%	94.3%	3%	89.8%	2%	87.6%	4%
GT [c] fishponds identified post-classification	90 of 113	7 of 113	95 of 113	7 of 113	89 of 113	5 of 113	92 of 113	7 of 113

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A37. Water-identifying index combination comparison for Khulna using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	67%	67%	74%	74%	63%	63%	67%	67%
GT[c] fishponds identified pre-classification	112 of 163	112 of 163	117 of 163	117 of 163	103 of 163	103 of 163	122 of 163	122 of 163
Percentage of GT[c] fishpond area identified post-classifier	65%	5%	73%	6%	62%	5%	56%	7%
GT[c] fishponds identified post-classification	95 of 163	15 of 163	109 of 163	19 of 163	94 of 163	15 of 163	88 of 163	18 of 163

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A38. Water-identifying index combination comparison for Satkhira using single day time period.

Time frame	Day	Day	Day	Day	Day	Day	Day	Day
Index Type	Combined	Combined	AWEI[d]	AWEI[d]	MNDWI[e]	MNDWI[e]	NDWI[f]	NDWI[f]
Classifier type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5
Percentage of GT[c] fishpond area identified pre-classifier	85%	85%	84%	84%	84%	84%	91%	91%
GT[c] fishponds identified pre-classification	115 of 208	115 of 208	114 of 208	114 of 208	100 of 208	100 of 208	159 of 208	159 of 208
Percentage of GT[c] fishpond area identified post-classifier	84%	1%	84%	2%	84%	1%	91%	0%
GT[c] fishponds identified post-classification	103 of 208	14 of 208	103 of 208	14 of 208	93 of 208	6 of 208	133 of 208	6 of 208

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Modified Normalized Difference Water Index
[f] Normalized Difference Water Index

Table A39. Best results of the index tests for each of the seven districts.

District	Bagerhat		Barisal		Bhola		Gopalganj		Jessore		Khulna		Satkhira	
Water Index	AWEI[d]		NDWI[e]		NDWI[e]		AWEI[d]		AWEI[d]		AWEI[d]		NDWI[e]	
Image date	10/28/2020		10/13/2020		11/7/2020		10/28/2020		11/5/2020		11/7/2020		11/5/2020	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]
Buffer Size	5	5	0	0	5	5	5	5	5	5	5	5	5	5
Percentage of GT[c] Fishpond area identified pre-classifier	76%	76%	26%	26%	20%	20%	54%	54%	95%	95%	74%	74%	91%	91%
GT[c] Fishponds Identified	190 of 235	190 of 235	61 of 168	61 of 168	10 of 27	10 of 27	67 of 77	67 of 77	102 of 113	102 of 113	117 of 163	117 of 163	159 of 208	159 of 208
Percentage of GT[c] Fishpond area identified post-classifier	76%	13%	20%	17%	12%	9%	50%	28%	94.3%	3%	73%	6%	91%	0.3%
GT[c] Fishponds Identified After Classification	181 of 235	30 of 235	21 of 168	16 of 168	4 of 27	3 of 27	56 of 77	26 of 77	95 of 113	7 of 113	109 of 163	19 of 163	133 of 208	6 of 208

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Normalized Difference Water Index

Table A40. Results from adding the Laplacian 5×5 convolution filter for each district.

District	Bagerhat		Barisal		Bhola		Gopalganj		Jessore		Khulna		Satkhira	
Water Index	AWEI[d]		NDWI[e]		NDWI[e]		AWEI[d]		AWEI[d]		AWEI[d]		NDWI[e]	
Image date	10/28/2020		10/13/2020		11/7/2020		10/28/2020		11/5/2020		11/7/2020		11/5/2020	
Classifier Type	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]	LR[a]	CART[b]
Buffer Size	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Percentage of GT[c] Fishpond area identified pre-classifier	76%	76%	58%	58%	57%	57%	68%	68%	91.5%	91%	78%	78%	82%	82%
GT [c] Fishponds Identified	235 of 235	235 of 235	160 of 168	160 of 168	27 of 27	27 of 27	77 of 77	77 of 77	113 of 113	113 of 113	163 of 163	163 of 163	206 of 208	206 of 208
Percentage of GT[c] Fishpond area identified post-classifier	25%	2%	44%	37%	43%	33%	19%	3%	91.4%	0%	56%	1%	56%	5%
GT[c] Fishponds Identified After Classification	131 of 235	40 of 235	90 of 168	83 of 168	16 of 27	14 of 27	30 of 77	10 of 77	113 of 113	3 of 113	134 of 163	239 of 163	138 of 208	56 of 208
# of Hits	131	40	90	83	16	14	30	10	113	3	134	23	138	56
# of Classified Fishponds	94794	66043	219955	185758	115902	98263	47046	33129	123786	83825	97045	66392	155887	126110
# of GT[c] Fishponds	235	235	168	168	27	27	77	77	113	113	163	163	208	208

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Ground truthing
[d] Automated Water Extraction Index
[e] Normalized Difference Water Index

Table A41. Comparison of LR, CART, RF, and SVM results when applied to ground-truthing fishpond data for Bagerhat, Barisal, Bhola, and Gopalganj during one day with that district's best performing water-identifying index.

District	Bagerhat				Barisal				Bhola				Gopalganj			
Water Index	AWEI[f]				NDWI[g]				NDWI[g]				AWEI[f]			
Image Date	10/28/2020				10/13/2020				11/7/2020				10/28/2020			
Classifier Type	LR[a]	CART[b]	RF[c]	SVM[d]	LR[a]	CART[b]	RF[c]	SVM[d]	LR[a]	CART[b]	RF[c]	SVM[d]	LR[a]	CART[b]	RF[c]	SVM[d]
Buffer Size	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Percentage of GT[e] Fishpond area identified pre-classifier	76%	76%	76%	76%	58%	58%	58%	58%	57%	57%	57%	57%	68%	68%	68%	68%
GT[e] Fishponds Identified	235 of 235	235 of 235	235 of 235	235 of 235	160 of 168	160 of 168	160 of 168	160 of 168	27 of 27	27 of 27	27 of 27	27 of 27	77 of 77	77 of 77	77 of 77	77 of 77
Percentage of GT[e] Fishpond area identified post-classifier	25%	2%	2%	26%	44%	37%	35%	42%	43%	33%	33%	41%	19%	3%	3%	20%
GT[e] Fishponds Identified Post-Classifier	131 of 235	40 of 235	37 of 235	153 of 235	90 of 168	83 of 168	76 of 168	103 of 168	16 of 27	14 of 27	14 of 27	17 of 27	30 of 77	10 of 77	10 of 77	38 of 77
# of Hits	131	40	37	153	90	83	76	103	16	14	14	17	30	10	10	38
# of Classified Fishponds	94794	66043	58466	426858	219955	185758	154127	666932	115902	98263	80723	446448	47046	33129	28794	243441
# of GT[e] Fishponds	235	235	235	235	168	168	168	168	27	27	27	27	77	77	77	77

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Random Forest
[d] Support Vector Machine
[e] Ground truthing
[f] Automated Water Extraction Index
[g] Normalized Difference Water Index

Table A42. Comparison of LR, CART, RF, and SVM results when applied to ground-truthing fishpond data for Jessore, Khulna, and Satkhira during one day with that district's best performing water-identifying index.

District	Jessore				Khulna				Satkhira			
Water Index	AWEI[f]				AWEI[f]				NDWI[g]			
Image Date	11/5/2020				11/7/2020				11/5/2020			
Classifier Type	LR[a]	CART[b]	RF[c]	SVM[d]	LR[a]	CART[b]	RF[c]	SVM[d]	LR[a]	CART[b]	RF[c]	SVM[d]
Buffer Size	5	5	5	5	5	5	5	5	5	5	5	5
Percentage of GT[e] Fishpond area identified pre-classifier	91.5%	91%	91.5%	91%	78%	78%	78%	78%	82%	82%	82%	82%
GT[e] Fishponds Identified	113 of 113	113 of 113	113 of 113	113 of 113	163 of 163	163 of 163	163 of 163	163 of 163	206 of 208	206 of 208	206 of 208	206 of 208
Percentage of GT[e] Fishpond area identified post-classifier	91.4%	0%	0.3%	91%	56%	1%	1%	56%	56%	5%	4%	56%
GT[e] Fishponds Identified Post-Classifier	113 of 113	3 of 113	3 of 113	113 of 113	134 of 163	23 of 163	23 of 163	136 of 163	138 of 208	56 of 208	51 of 208	147 of 208
# of Hits	113	3	3	113	134	23	23	136	138	56	51	147

Table A42 (cont'd)

# of Classified Fishponds	123786	83825	74365	667161	97045	66392	58987	477444	155887	126110	104324	389366
# of GT[e] Fishponds	113	113	113	113	163	163	163	163	208	208	208	208

[a] Logistic Regression
[b] Classification and Regression Trees
[c] Random Forest
[d] Support Vector Machine
[e] Ground truthing
[f] Automated Water Extraction Index
[g] Normalized Difference Water Index

Table A43. Land use as percent of total area for 10-meter buffer around ground truth fishpond locations for each district.

Land Use	District						
	Bagerhat	Barisal	Bhola	Gopalganj	Jessore	Khulna	Satkhira
Cultivated Land	79	46	21	67	90	63	6
Forest	0	0	0	0	0	0	0
Grass Land	0	0	0	0	0	0	0
Shrub Land	0	0	0	0	0	0	0
Wetland	0	0	0	0	0	0	0
Water Body	8	0	0	0	1	19	82
Artificial Surfaces	13	54	79	33	9	18	12
Bareland	0	0	0	0	0	0	0
Unknown	0	0	0	0	0	0	0

Table A44. Land use as percent of total area for 50-meter buffer around ground truth fishpond locations for each district.

Land Use	District						
	Bagerhat	Barisal	Bhola	Gopalganj	Jessore	Khulna	Satkhira
Cultivated Land	80	46	26	68	86	61	10
Forest	0	0	0	0	0	0	0
Grass Land	0	0	0	0	0	0	0
Shrub Land	0	0	0	0	0	0	0
Wetland	0	0	0	0	0	0	0
Water Body	7	0	0	0	1	16	69
Artificial Surfaces	13	54	74	32	13	23	21
Bareland	0	0	0	0	0	0	0
Unknown	0	0	0	0	0	0	0

Table A45. Land use as percent of total area for 100-meter buffer around ground truth fishpond locations for each district.

Land Use	District						
	Bagerhat	Barisal	Bhola	Gopalganj	Jessore	Khulna	Satkhira
Cultivated Land	78	50.4	27	71	82	59	12
Forest	0	0	0	0	0	0	0
Grass Land	0	0	0	0	0	0	0
Shrub Land	0	0	0	0	0	0	0
Wetland	0	0	0	0	0	0	0
Water Body	7	0.1	0	0	2	15	63
Artificial Surfaces	15	49.5	73	29	16	26	25
Bareland	0	0	0	0	0	0	0
Unknown	0	0	0	0	0	0	0

Table A46. Median ground truth fishpond size for each district.

Statistic	District						
	Bagerhat	Barisal	Bhola	Gopalganj	Jessore	Khulna	Satkhira
Median Ground truth fishpond size (m^2)	2503	833	713	2237	3169	2972	1252

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramowitz, M., & Stegun, C. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (9th ed., p. 14). Dover.

Agmalaro, M. A., Sitanggang, I. S., & Waskito, M. L. (2021). Sentinel 1 Classification for Garlic Land Identification using Support Vector Machine. *9th International Conference on Information and Communication Technology*, 440–444.

Ahmed, N., Thompson, S., & Glaser, M. (2019). Global Aquaculture Productivity, Environmental Sustainability, and Climate Change Adaptability. *Environmental Management*, *63*(2), 159–172. https://doi.org/10.1007/s00267-018-1117-3

Al Sayah, M. J., Nedjai, R., Abdallah, C., & Khouri, M. (2020). On the use of remote sensing to map the proliferation of aquaculture ponds and to investigate their effect on local climate, perspectives from the Claise watershed, France. *Springer Nature Switzerland*. https://doi.org/10.1007/s10661-020-08250-0

Anand, A., Krishnan, P., Kantharajan, G., Suryavanshi, A., Kawishwar, P., Raj, U., Rao, C. S., Choudhury, S. B., Manjulatha, C., & Babu, D. E. (2020). Assessing the water spread area available for fish culture and fish production potential in inland lentic waterbodies using remote sensing: A case study from Chhattisgarh State, India. *Remote Sensing Applications: Society and Environment*, *17*. https://doi.org/10.1016/j.rsase.2019.100273

Ballester-Berman, D. J., Sanchez-Jerez, P., & Marino, A. (2018). Detection of aquaculture structures using Sentinel-1 data. *Proceedings of the European Conference on Synthetic Aperture Radar, EUSAR*, *2018-June*, 888–891. https://servicio.pesca.mapama.es/acuivisor/

Bangladesh Meteorological Department. (2021). *Maproom: Seasonal Climate Analysis*. Maproom. http://datalibrary.bmd.gov.bd/maproom/Climatology/Climate_Analysis/seasonal.html?resolution=0.05&YearStart=2020&YearEnd=2020&seasonStart=Jan&seasonEnd=Dec&var=Rain&yearlyStat=Mean

Bangladesh Water Development Board. (2019). *Summary of Rainfall in Bangladesh for the Year 2017 & 2018*.

Bello, O. M., & Aina, Y. A. (2014). Satellite Remote Sensing as a Tool in Disaster Management and Sustainable Development: Towards a Synergistic Approach. *Procedia - Social and Behavioral Sciences*, *120*, 365–373. https://doi.org/10.1016/j.sbspro.2014.02.114

Belton, B., & Azad, A. (2012). The characteristics and status of pond aquaculture in Bangladesh. *Aquaculture*, *358–359*, 196–204. https://doi.org/10.1016/j.aquaculture.2012.07.002

Belton, B., Bush, S. R., & Little, D. C. (2018). Not just for the wealthy: Rethinking farmed fish consumption in the Global South. *Global Food Security*, *16*, 85–92.

https://doi.org/10.1016/J.GFS.2017.10.005

Bernzen, A., Pritchard, B., Braun, B., Belton, B., & Rigg, J. (2021). Guest Editorial: Geographies of engagement, livelihoods and possibility in South and Southeast Asian deltas. *Singapore Journal of Tropical Geography*, *42*(2), 197–202. https://doi.org/10.1111/SJTG.12373

Biggs, J., von Fumetti, S., & Kelly-Quinn, M. (2017). The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. In *Hydrobiologia* (Vol. 793, Issue 1, pp. 3–39). Springer International Publishing. https://doi.org/10.1007/s10750-016-3007-0

Charles, H., Godfray, J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. In *New Series* (Vol. 12, Issue 5967).

China Ministry of Natural Resources. (2020). *GlobeLand30*. http://www.globeland30.org/defaults_en.html?type=data&src=/Scripts/map/defaults/En/browse_en.html&head=browse

Chu, D. (2020). Remote Sensing of Land Use and Land Cover in Mountain Region. In *Remote Sensing of Land Use and Land Cover in Mountain Region*. https://doi.org/10.1007/978-981-13-7580-4

Coffey, V. C. (2012, April 1). *Multispectral Imaging Moves into the Mainstream*. Optics & Photonics. https://www.osa-opn.org/home/articles/volume_23/issue_4/features/multispectral_imaging_moves_into_the_mainstream/

CRISP, & National University of Singapore. (2001). *Infrared Remote Sensing*. https://crisp.nus.edu.sg/~research/tutorial/infrared.htm

De Graaf, G., Khan, G., Omar, M., Lubna, F., Abdullah, Y., & Mamun, A. (2000). *Environment and GIS Support Project for Water Sector Planning EGIS-II FISH-GIS An Introduction to the Use of Geographical Information Systems and Remote Sensing in Fisheries Monitoring, Analysis and Management*. www.Nefisco.org

Department of Agricultural Extension. (2021). *Agro-Ecological Zones -AEZs Maps*. https://www.bamis.gov.bd/en/page/aezs-maps/

Department of Fisheries. (2018). *Yearbook of Fisheries Statistics of Bangladesh 2017-18*. www.fisheries.gov.bd

Duan, Y., Li, X., Zhang, L., Chen, D., Liu, S., & Ji, H. (2020). Mapping national-scale aquaculture ponds based on the Google Earth Engine in the Chinese coastal zone. *Aquaculture*, *520*. https://doi.org/10.1016/j.aquaculture.2019.734666

Duan, Y., Li, X., Zhang, L., Liu, W., Liu, S., Chen, D., & Ji, H. (2020). Detecting spatiotemporal

changes of large-scale aquaculture ponds regions over 1988–2018 in Jiangsu Province, China using Google Earth Engine. *Ocean and Coastal Management*, *188*, 105144. https://doi.org/10.1016/j.ocecoaman.2020.105144

Dumitru, C. O., Schwarz, G., & Datcu, M. (2018). Earth Observation Semantics and Data Analytics for Coastal Environmental Areas. *International Journal on Advances in Software*, *11*(3), 390–399. http://www.iariajournals.org/software/2018,

Earth Observing System. (2019, December 4). *Satellite Data: What Spatial Resolution Is Enough For You?* Earth Observing System. https://eos.com/blog/satellite-data-what-spatial-resolution-is-enough-for-you/

Ellowitz, H. I. (1992). An introduction to microwave remote sensing. *Microwave Journal*, *35*(9), 69. https://go-gale-com.proxy1.cl.msu.edu/ps/i.do?p=ITOF&u=msu_main&id=GALE%7CA13412901&v=2.1&it=r&sid=summon

ESA. (2020a). *CBERS - Second Generation* . https://earth.esa.int/web/eoportal/satellite-missions/c-missions/cbers-3-4

ESA. (2020b). *Sentinel-1 - Missions* . https://sentinel.esa.int/web/sentinel/missions/sentinel-1

ESA. (2020c). *TOPSAR Processing - Sentinel-1 SAR Technical Guide - Sentinel Online*. https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-algorithms/topsar-processing

ESA. (2020d). *WorldView-1*. https://directory.eoportal.org/web/eoportal/satellite-missions/v-w-x-y-z/worldview-1

FAO. (1989). *4. AQUACULTURE METHODS AND PRACTICES: A SELECTED REVIEW*. http://www.fao.org/3/t8598e/t8598e05.htm

FAO. (2014). *A-2 World aquaculture production by culture environment*. http://www.fao.org/fishery/docs/STAT/summary/a-2.pdf

FAO. (2018). *WORLD FISHERIES AND AQUACULTURE*. www.fao.org/publications

FAO. (2020a). *FIGIS - Fisheries Statistics - Aquaculture*. http://www.fao.org/figis/servlet/TabLandArea?tb_ds=Aquaculture&tb_mode=TABLE&tb_act=SELECT&tb_grp=COUNTRY

FAO. (2020b). *National Aquaculture Sector Overview - Bangladesh*. http://www.fao.org/fishery/countrysector/naso_bangladesh/en

FAO. (2020c). The State of World Fisheries and Aquaculture. In *The State of World Fisheries and Aquaculture 2020*. FAO. https://doi.org/10.4060/ca9229en

Faul, A. C. (2019). A Concise Introduction to Machine Learning. In *A Concise Introduction to*

Machine Learning. Chapman and Hall/CRC. https://doi.org/10.1201/9781351204750

Feed the Future. (2021, January). *Bangladesh*. https://www.feedthefuture.gov/country/bangladesh/

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). *Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery*. https://doi.org/10.1016/j.rse.2013.08.029

Frohn, R. C. (2007). The use of landscape pattern metrics in remote sensing image classification. *Https://Doi.Org/10.1080/01431160500212229*, *27*(10), 2025–2032. https://doi.org/10.1080/01431160500212229

Funge-Smith, Simon; Phillips, M. J. (2001). *Aquaculture Systems and Species*. http://www.fao.org/3/AB412E/ab412e07.htm

Gade, M., & Stoffelen, A. (2019). An Introduction to Microwave Remote Sensing of the Asian Seas. In V. Barale & M. Gade (Eds.), *Remote Sensing of the Asian Seas* (pp. 81–101). Springer, Cham. https://doi.org/10.1007/978-3-319-94067-0_4

Gao, J., Du, Z., Shi, Z., Xu, Z., Cao, Q., & Tang, R. (2018). Switching impulse noise filter based on Laplacian convolution and pixels grouping for color images. *Signal, Image and Video Processing*, *12*(8), 1523–1529. https://doi.org/10.1007/S11760-018-1308-7

Gautam, V. K., Gaurav, P. K., Murugan, P., & Annadurai, M. (2015). Assessment of Surface Water Dynamicsin Bangalore Using WRI, NDWI, MNDWI, Supervised Classification and K-T Transformation. *Aquatic Procedia*, *4*, 739–746. https://doi.org/10.1016/J.AQPRO.2015.02.095

Google. (2020). *Introduction to Machine Learning | Machine Learning Crash Course*. Google. https://developers.google.com/machine-learning/crash-course/ml-intro

Gusmawati, N., Soulard, B., Selmaoui-Folcher, N., Proisy, C., Mustafa, A., Le Gendre, R., Laugier, T., & Lemonnier, H. (2017). *Surveying shrimp aquaculture pond activity using multitemporal VHSR satellite images-case study from the Perancak estuary, Bali, Indonesia*. https://doi.org/10.1016/j.marpolbul.2017.03.059

Hashem S., Akter T., Salam M. A., H. M. T. (2014). Aquaculture planning through Remote Sensing Image Analysis and GIS tools in Northeast region, Bangladesh. *International Journal of Fisheries and Aquatic Studies*, *1*(5), 127–136. www.fisheriesjournal.com

Heck, J. E., Nieves, J. W., Chen, Y., Parvez, F., Brandt-Rauf, P. W., Howe, G. R., & Ahsan, H. (2010). Protein and amino acid intakes in a rural area of Bangladesh. *Food and Nutrition Bulletin*, *31*(2), 206. https://doi.org/10.1177/156482651003100203

Hernandez, R., Belton, B., Reardon, T., Hu, C., Zhang, X., & Ahmed, A. (2018). The "quiet revolution" in the aquaculture value chain in Bangladesh. *Aquaculture*, *493*(June 2017), 456–468. https://doi.org/10.1016/j.aquaculture.2017.06.006

Huda, K. M. S., Atkins, P. J., Donoghue, D. N. M., & Cox, N. J. (2010). Small water bodies in Bangladesh. *Area*, *42*(2), 217–227. https://doi.org/10.1111/j.1475-4762.2009.00909.x

Iordache, M.-D., Mantas, V., Baltazar, E., Lewyckyj, N., & Souverijns, N. (2020). Application of Random Forest Classification to Detect the Pine Wilt Disease from High Resolution Spectral Images. *International Geoscience and Remote Sensing Symposium*, *776026*, 4489–4492.

Islam, M. M., Barman, A., Kundu, G. K., Kabir, M. A., & Paul, B. (2019). Vulnerability of inland and coastal aquaculture to climate change: Evidence from a developing country. *Aquaculture and Fisheries*, *4*(5), 183–189. https://doi.org/10.1016/j.aaf.2019.02.007

Jahan, K., Belton, B., Ali, H., Dhar, G., & Ara, I. (2015). Aquaculture technologies in Bangladesh: An assessment of technical and economic performance and producer behavior. In *WorldFish*.

JAXA. (2020). *ALOS-2 Project / PALSAR-2*.

Ji, L., Zhang, L., & Wylie, B. K. (2009). Analysis of dynamic thresholds for the normalized difference water index. *Photogrammetric Engineering and Remote Sensing*, *75*(11), 1307–1317. https://doi.org/10.14358/PERS.75.11.1307

JianCheng, S., Yang, D., JinYang, D., LingMei, J., LinNa, C., KeBiao, M., Peng, X., WenJian, N., Chuan, X., Qiang, L., ChenZhou, L., Peng, G., Qian, C., YunQing, L., Jing, C., AnQi, W., HeJia, L., & YinHui, W. (2012). Progresses on microwave remote sensing of land surface parameters. *Sci China Earth Sci*, *55*(7), 1052–1078. https://doi.org/10.1007/s11430-012-4444-x

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J., & Xiao, T. (2014). An Automated Method for Extracting Rivers and Lakes from Landsat Imagery. *Remote Sensing 2014, Vol. 6, Pages 5067-5089*, *6*(6), 5067–5089. https://doi.org/10.3390/RS6065067

Jiao, L., Liu, Y., & Li, H. (2012). Characterizing land-use classes in remote sensing imagery by shape metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, *72*, 46–55. https://doi.org/10.1016/J.ISPRSJPRS.2012.05.012

Kapetsky, J. M. (1987). Satellite remote sensing to locate and inventory small water bodies for fisheries management and aquaculture development in Zimbabwe. *CIFA Occas. Pap.*

Karim, M., & Little, D. C. (2018). The impacts of integrated homestead pond-dike systems in relation to production, consumption and seasonality in central north Bangladesh. *Aquaculture Research*, *49*(1), 313–334. https://doi.org/10.1111/are.13462

Khorram, S., Koch, F. H., van der Wiele, C. F., & Nelson, S. A. C. (2012). Introduction. In *Remote Sensing* (pp. 1–15). https://doi.org/10.1007/978-1-4614-3103-9_1

Khorram, S., van der Wiele, C., Koch, F., Nelson, S., & Potts, M. (2016). *Principles of Applied Remote Sensing*. https://doi.org/10.1007/978-3-319-22560-9

Kobayashi, M., Msangi, S., Batka, M., Vannuccini, S., Dey, M. M., & Anderson, J. L. (2015). Fish to 2030: The Role and Opportunity for Aquaculture. *Aquaculture Economics and Management*, *19*(3), 282–300. https://doi.org/10.1080/13657305.2015.994240

Kubat, M. (2017). An Introduction to Machine Learning. In *An Introduction to Machine Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-319-63913-0

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Lam, M. E. (2016). The Ethics and Sustainability of Capture Fisheries and Aquaculture. *Journal of Agricultural and Environmental Ethics*, *29*(1), 35–65. https://doi.org/10.1007/s10806-015-9587-2

Liu, K., Ai, B., & Wang, S. (2010). *Fishpond change detection based on short term time series of RADARSAT images and object-oriented method Reconstruction of green open space in perspective of social equity: assessment, mechanism and optimization View project Environment Remote sensing, View project Fish-pond Change Detection Based on Short Term Time Series of RADARSAT Images and Object-oriented Method*. https://doi.org/10.1109/CISP.2010.5647643

MathWorks. (2016a). *Applying Supervised Learning*.

MathWorks. (2016b). *Applying Unsupervised Learning*.

MathWorks. (2016c). *Getting Started with Machine Learning*.

MathWorks. (2016d). *Introducing Machine Learning*.

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

McFeeters, S. K. (1995). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Https://Doi.Org/10.1080/01431169608948714*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

McFeeters, Stuart K. (2013). Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach. *Remote Sensing 2013, Vol. 5, Pages 3544-3561*, *5*(7), 3544–3561. https://doi.org/10.3390/RS5073544

Mohsen, A., Elshemy, M., & Zeidan, B. A. (2018). Change detection for Lake Burullus, Egypt using remote sensing and GIS approaches. *Environmental Science and Pollution Research*, *25*(31), 30763–30771. https://doi.org/10.1007/s11356-016-8167-y

Momtaz, S., & Shameem, M. I. M. (2016). The Research Setting. *Experiencing Climate Change in Bangladesh*, 41–68. https://doi.org/10.1016/b978-0-12-803404-0.00004-1

Mukherjee, N. R., & Samuel, C. (2016). Assessment of the Temporal Variations of Surface Water Bodies in and around Chennai using Landsat Imagery. *Indian Journal of Science and Technology*, *9*(18), 1–7. https://indjst.org/articles/assessment-of-the-temporal-variations-of-surface-water-bodies-in-and-around-chennai-using-landsat-imagery

NASA. (2020a). *Get to Know SAR – NASA-ISRO SAR Mission (NISAR)*. https://nisar.jpl.nasa.gov/mission/get-to-know-sar/overview/

NASA. (2020b). *SAR | NASA Earth Science Disasters Program*. https://disasters.nasa.gov/instruments/sar

NASA. (2020c, July 29). *About Terra | Terra*. https://terra.nasa.gov/about

NASA EarthData. (2020). *Remote Sensors | Earthdata*. https://earthdata.nasa.gov/learn/remote-sensors

Natural Resources Canada. (2015). *Fundamentals of Remote Sensing*. https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/earthsciences/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf

NSIDC. (2020, April 3). *Remote Sensing*. https://nsidc.org/cryosphere/seaice/study/infrared_remote_sensing.html

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern*, *SMC-9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Ottinger, M., Clauss, K., & Kuenzer, C. (2016). Aquaculture: Relevance, distribution, impacts and spatial assessments - A review. *Ocean and Coastal Management*, *119*, 244–266. https://doi.org/10.1016/j.ocecoaman.2015.10.015

Ottinger, M., Clauss, K., & Kuenzer, C. (2017). Large-scale assessment of coastal aquaculture ponds with Sentinel-1 time series data. *Remote Sensing*, *9*(5). https://doi.org/10.3390/rs9050440

Ottinger, M., Clauss, K., & Kuenzer, C. (2018). Opportunities and challenges for the estimation of aquaculture production based on earth observation data. *Remote Sensing*, *10*(7). https://doi.org/10.3390/rs10071076

PBS. (2007, October). *NOVA | Sputnik Declassified | What Satellites See (non-Flash) | PBS*. https://www.pbs.org/wgbh/nova/sputnik/sate-nf.html

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature 2016 540:7633*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Prasad, K. A., Ottinger, M., Wei, C., & Leinenkugel, P. (2019). Assessment of coastal aquaculture for India from Sentinel-1 SAR time series. *Remote Sensing*, *11*(3). https://doi.org/10.3390/rs11030357

Ramezan, C. A., Warner, T. A., & Maxwell, A. E. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, *11*(2). https://doi.org/10.3390/rs11020185

Rashid, S., Minot, N., & Lemma, S. (2019). Does a "Blue Revolution" help the poor? Evidence from Bangladesh. *Agricultural Economics*, *50*(2), 139–150. https://doi.org/10.1111/agec.12472

Rebala, G., Ravi, A., & Churiwala, S. (2019). *An Introduction to Machine Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-15729-6

Ren, C., Wang, Z., Zhang, Y., Zhang, B., Chen, L., Xi, Y., Xiao, X., Doughty, R. B., Liu, M., Jia, M., Mao, D., & Song, K. (2019). *Rapid expansion of coastal aquaculture ponds in China from Landsat observations during 1984-2016*. https://doi.org/10.1016/j.jag.2019.101902

Rhodes, C. J., Henrys, P., Siriwardena, G. M., Whittingham, M. J., & Norton, L. R. (2015). The relative value of field survey and remote sensing for biodiversity assessment. *Methods in Ecology and Evolution*, *6*(7), 772–781. https://doi.org/10.1111/2041-210X.12385

Ross, L. G., Telfer, T. C., Falconer, L., Soto, D., & Aguilar-Manjarrez, J. (2013). Site selection and carrying capacities for inland and coastal aquaculture. In *FAO Fisheries and Aquaculture Proceedings No. 21* (Issue December). http://web.a.ebscohost.com.proxy1.cl.msu.edu/ehost/ebookviewer/ebook?sid=6361a65c-22b2-4e69-afa9-22ca9df1e735%40sessionmgr4007&vid=0&format=EB

Sasaki, Y. (2007). *The truth of the F-measure* (Issue January 2007).

Sattar, M. N. (2019). *The Freshwater Aquaculture Revolution in Bangladesh: Impacts on Land, Water, and Livelihoods*.

Shah, M. S. (2003). Human resource development activities infisheries sector. *Fish Fortnight Compendium 2003*, 57–59.

Shamsuzzaman, M. M., Islam, M. M., Tania, N. J., Al-Mamun, M. A., Barman, P. P., & Xu, X. (2017). *Fisheries resources of Bangladesh: Present status and future direction*. Aquaculture and Fisheries. https://reader.elsevier.com/reader/sd/pii/S2468550X16300260?token=D4D1183AFAF673CC35884FDC579258AC3B92C6F79612827C172DF1BC8C94C66545D953E385C931B859370D634C2EF949&originRegion=us-east-1&originCreation=20210603182845

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

Stiller, D., Ottinger, M., & Leinenkugel, P. (2019). *Spatio-Temporal Patterns of Coastal Aquaculture Derived from Sentinel-1 Time Series Data and the Full Landsat Archive*. https://doi.org/10.3390/rs11141707

Subasinghe, R. (2017). *FAO Fisheries and Aquaculture Circular FIAA / C1140 (En) WORLD AQUACULTURE 2015 : A BRIEF OVERVIEW Fisheries and WORLD AQUACULTURE 2015 : A BRIE F OVERVIEW* (Vol. 1140, Issue January).

Thilsted, S. H., Thorne-Lyman, A., Webb, P., Bogard, J. R., Subasinghe, R., Phillips, M. J., & Allison, E. H. (2016). Sustaining healthy diets: The role of capture fisheries and aquaculture for improving nutrition in the post-2015 era. *Food Policy*, *61*, 126–131. https://doi.org/10.1016/j.foodpol.2016.02.005

Travaglia, C., Profeti, G., Aguilar Manjarrez, J., & Lopez, N. A. (2004). Mapping coastal aquaculture and fisheries structures by satellite imaging radar. *FAO. Fisheries Technical Paper*, *459*, 58. http://www.fao.org/documents/card/en/c/d8c9fd81-4655-55f9-b359-8107c6629622/

UCSB. (2020). *GeoEye-1 Gives Google Highest Resolution Imagery*. https://geog.ucsb.edu/geoeye-1-gives-google-highest-resolution-imagery/

United Nations et al. (2003). *Integrated Environmental and Economic Accounting 2003*.

US Department of Commerce, N. O. and A. A. (n.d.). *What is aquaculture?*

USGS. (2020). *Landsat 8*. https://www.usgs.gov/land-resources/nli/landsat/landsat-8?qt-science_support_page_related_con=0#qt-science_support_page_related_con

Virdis, S. G. P. (2014). An object-based image analysis approach for aquaculture ponds precise mapping and monitoring: A case study of Tam Giang-Cau Hai Lagoon, Vietnam. *Environmental Monitoring and Assessment*, *186*(1), 117–133. https://doi.org/10.1007/s10661-013-3360-7

Wicaksono, A., & Wicaksono, P. (2019). Geometric accuracy assessment for shoreline derived from NDWI, MNDWI, and AWEI transformation on various coastal physical typology in Jepara Regency using Landsat 8 OLI imagery in 2018. *Geoplanning: Journal of Geomatics and Planning*, *6*(1), 55–72. https://doi.org/10.14710/geoplanning.6.1.55-72

Woodhouse, I. H. (2017). Introduction to Microwave Remote Sensing. In *Introduction to Microwave Remote Sensing*. Taylor & Francis Group. https://doi.org/10.1201/9781315272573

Worden, J., & de Beurs, K. M. (2020). Surface water detection in the Caucasus. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102159. https://doi.org/10.1016/J.JAG.2020.102159

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Yan, W., Wang, R., Shengyi, Y., & Jianping, L. (2021). Tight Reservoirs Classification using Random Forest. *3rd International Conference on Intelligent Control, Measurement and*

Signal Processing and Intelligent Oil Field, *Icmsp*, 1–5. https://doi.org/101109

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing*, *72*(7), 799–811. https://doi.org/10.14358/PERS.72.7.799

Yu, Z. (2019, April 4). *Fish Ponds Detection in Bangladesh on Google Earth Engine: A case study of Singra Upazila*.

Yu, Z., Di, L., Rahman, M. S., & Tang, J. (2020). Fishpond mapping by spectral and spatial-based filtering on google earth engine: A case study in singra upazila of Bangladesh. *Remote Sensing*, *12*(17). https://doi.org/10.3390/RS12172692

Yu, Z., Di, L., Tang, J., Zhang, C., Lin, L., Yu, E. G., Rahman, M. S., Gaigalas, J., & Sun, Z. (2018, September 27). Land use and land cover classification for Bangladesh 2005 on google earth engine. *International Conference on Agro-Geoinformatics*. https://doi.org/10.1109/Agro-Geoinformatics.2018.8475976

Zeng, Z., Wang, D., Tan, W., & Huang, J. (2019). Extracting aquaculture ponds from natural water surfaces around inland lakes on medium resolution multispectral images. *International Journal of Applied Earth Observation and Geoinformation*, *80*, 13–25. https://doi.org/10.1016/j.jag.2019.03.019

Zhang, J., Jin, R., Yang, Y., & Hauptmann, A. G. (2003). Modified Logistic Regression: An Approximation to SVM and Its Applications in Large-Scale Text Categorization. *Twentieth International Conference on Machine Learning*.

Zhou, Y., Dong, J., Xiao, X., Xiao, T., Yang, Z., Zhao, G., Zou, Z., & Qin, Y. (2017). Open Surface Water Mapping Algorithms: A Comparison of Water-Related Spectral Indices and Sensors. *Water 2017, Vol. 9, Page 256*, *9*(4), 256. https://doi.org/10.3390/W9040256

Zoran, L. F. V., Ionescu Golovanov, C., & Zoran, M. A. (2010). Spectral characteristics and feature selection of satellite remote sensing data for land use/cover changes assessment in the Romanian northwestern Black Sea coastal area. *Earth Resources and Environmental Remote Sensing/GIS Applications*, *7831*, 78311P. https://doi.org/10.1117/12.864793