Norges miljø- og biovitenskapelige universitet

Master's Thesis 2023 30 ECT
Faculty of Science and Technology

Seagrass and Seaweed Detection Approaches Using Remote Sensing and Google Earth Engine: A Comparative Analysis of Different Machine Learning Techniques

Sigrid Knag
Master of Science in Data Science

Acknowledgement

Firstly, I would like to thank my supervisor, Associate Professor Dr. Habil. Fadi Al Machot for his excellent guidance and feedback on this thesis, in addition to his presence, swift replies, good mood, and patience. I also have to applaud you for your consistency in hosting weekly regular meetings, good advice from the start, working on weekends, and caring for the students. Thank you.

Secondly, I would also like to thank Team Zostera for inspiring this thesis topic, as well as the team's guidance, data, and relevant information sent my way. A special thanks to Glenn Page, Founder/CEO of SustainaMetrix, COBALT, and the global lead of Team Zostera. Thanks for being a good mentor, providing professional feedback, believing in me, sharing your time and resources, and giving me a chance to work with your team and learn. Then, I would also like to give a special thanks to Lauren Hayden, a good teammate, and a fantastic Zoom partner, for providing this thesis with local knowledge of Casco Bay. I would also give thanks to Stefan Claesson for his great feedback reviewing this work.

Lastly, I would like to thank my friends for their support, patience, and time spent together at NMBU. Additionally, I thank my parents and brothers for their humor and kind hearts.

Abstract

Seagrasses and seaweed habitats contribute to crucial ecological services globally, from capturing carbon dioxide and supporting 20% of the world's largest fisheries to sustaining the small, but many coastal communities [1]. Across the globe, an alarming decline in their wild distribution has been recorded, attributed to climate change and direct pollution [2]. Current estimates of how much the loss is are uncertain and mapping and monitoring efforts are costly, data-intensive, and lack scalability. Thus, freely available data and software in remote sensing, coupled with Machine Learning (ML) are deemed important means to leverage existing mapping of seagrasses and seaweed spatial distribution [3, 4].

This thesis explored a free and scalable workflow by comparing three different ML techniques mainly on Overall Accuracy (OA) and Tau(e) in classifying seagrass, seaweed, and water. These are supervised, unsupervised, and semi-supervised learning (SSL) which used data from the satellite, Sentinel-2 Level-2A, applied to a novel area of study, from Biddeford Pool to Small Point at the Coast of Maine, United States of America. Results showed that the SSL achieved the highest OA of 76% and Tau(e) = 0.72 on the hard test, in line with previous work. To the best of the author's knowledge, this work contributes to the field of science by being the first in its field to use the geospatial analysis package 'geemap', along with the software Google Earth Engine, and SSL for classifying seagrass and seaweed. Through demonstration, this work shows the potential of free data in remote sensing, leveraged by ML to aid community monitoring in the environmental management of seagrass and seaweed. The results here can be considered as a starting point for further exploring the SSL paired with freely available data and community monitoring to lower costs, handle data scarcity, and scale up in the field of aquatic vegetation mapping and monitoring.

Contents

List of Figures

List of Tables

List of Acronyms

ISODATA Iterative Self-Organizing Data Analysis Technique 5

GCS Geographical Coordinate System XIII

WGS84 World Geodetic System 1984 XII

FAI Floating Algae Index 5

KD Kelp Difference 5

PCA Principal Component Analysis 5

SAV Submerged Aquatic Vegetation 5

ESA European Space Agency 12

PB Pixel-Based XII

OBIA Object Based Image Analysis XV

GEOBIA Geographical Object-Based Image Analysis XII

GEE Google Earth Engine XV

GC Google Colab 13

DN Digital Number 16

CNN Convolutional Neural Network 20

Chapter 1

Introduction

Seagrasses and seaweed are essential for the survival of countless organisms because of their crucial role in providing ecological services on a global scale. Half of the oxygen on Earth comes from seaweed and microscopic algae, thus they play a significant role in the oxygen cycle [8]. Additionally, the value of seagrass habitats to the largest fisheries in the world cannot be overstated. These habitats support 20% of the world's fisheries, worth about 28 billion USD [9].

Nevertheless, seaweed and seagrass are in danger, with climate change and pollution being the main causes. There is an urgent need for more accurate mapping and monitoring of the spatial distribution of seagrass and seaweed populations because the continued decline of these populations will undoubtedly have significant ecological and economic effects [2].

New developments in Machine Learning (ML) and Remote Sensing (RS) have been identified as promising approaches to solve this problem [4]. These state-of-the-art tools can be used by researchers to collect and analyze enormous amounts of data, giving them important new insights into the status of seagrasses and seaweed populations today. Scientists and decision-makers can take well-informed action that will aid in preserving and restoring these vital marine habitats by using accurate maps and timely monitoring.

This thesis presents background information on seagrasses, seaweed, and citizen science and emphasizes the proposed approach's significance in distinguishing between seagrass, seaweed, and water using satellite images with three different ML algorithms: A supervised with RF, an unsupervised using K-means, and a semi-supervised model using RF with pseudo labels. Moreover, this thesis tries to answer two research questions that will direct the investigation following contextual and background information. It will emphasize the broader implications of the research and its potential contributions to the field by elucidating the significance and purpose of the work.

1.1 Background

1.1.1 Seagrass

Seagrasses are flowering marine plants living underwater [10, 11]. Out of the four families and 72 species they comprise, this thesis will only consider *Zostera marina*, also known as eelgrass, as it is the dominant species of Casco Bay, the Area Of Interst (AOI). Seagrasses are found to be some of the most productive ecosystems. Their benefits are diverse, touching the economy, human health, biodiversity, and many of the Sustainable Development Goals of the United Nations from 2016 [12], with their protection and restoration seen as tools in the fight against climate change [1, 2, 13, 4, 14, 15]. Additionally, seagrasses, as well as mangroves, salt marshes, and coral reefs, are classified as Blue Carbon Ecosystem because of their ability to sequester large amounts of carbon dioxide in the ocean floor, which emphasizing their importance [16, 17, 18, 19, 20].

Mapping the spatial extent of seagrass is therefore of global concern, but current estimates are uncertain, ranging between 160,387 to 266, 562km^2 [2]. Moreover, since the 1940s, an estimated one-third of the seagrass population has been lost [21] and continues to decline on an annual basis [1], caused by anthropogenic stressors such as agriculture run-offs, waste, and sewage [22, 23, 24]. Thus, urgent action and higher accuracy are needed, and new technology in RS and ML are seen as viable solutions [2, 25, 26].

1.1.2 Seaweed

Seaweeds, unlike seagrasses, are not plants but algae and can be categorized as either brown, red, or green seaweed based on its apparent color. The dominant seaweed species in the AOI is *Ascophyllum nodosum*, a brown seaweed more commonly known as Rockweed, which will be the main seaweed species of focus in this work. However, the priority was broadened to include all species of brown, green, and red seaweed, as the only encountered dataset representing Rockweed in the area was considered insufficient.

Globally, seaweed are commercially grown as part of being an integrated part of Asian cuisine [27, 9]. While the commercial amount increases, wild seaweed are in decline for the same reasons as for seagrasses. This has economic and biological consequences for the local communities, small-scale harvesting businesses, and biodiversity relying on their resources [28, 9]. In Casco Bay, seagrass and seaweed often co-exist. Thus, a problem occurs in delineating areas for protecting seagrass from those of sustainable harvesting seaweed [26]. Data portals and regulatory institutions providing an overview of managerial boundaries need tools that can easily distinguish them accurately to avoid conflicting interests. Current efforts in the area exist but become too infrequent, costly, and labor-intensive [29, 30].

1.2 Motivation

The motivation for this thesis was to aid a local effort, Team Zostera [31], in the restoration and conservation of seagrass in Casco Bay, Maine, United States (U.S.). Experimental and exploratory research on RS combined with ML was performed to do so. The aim is to start with the local situation of Casco Bay but with a bigger goal of generalizing the workflow to other areas, as these efforts on seagrasses are global. Furthermore, many of these local initiatives have volunteers aiding the scientific work, performing what is called 'citizen science' but will be referred to here as 'community monitoring' as the term is more inclusive. Therefore, the goal is to create an accurate, easy, free, and scalable workflow that distinguishes seagrass from its surroundings, in addition to making the work here accessible to all[1]. As a starting point, distinguishing seagrass from seaweed and the surrounding waters is done in Casco Bay. This is because the dominant seagrass species, *Zostera marina*, co-exists with seaweed, especially *Ascophyllum nodosum*, and local stakeholders want to assess how accurately they could be distinguished using only freely available data and tools.

1.3 Community Monitoring

Community monitoring is defined in this thesis as "volunteers who engage in collecting and analyzing data for scientific research". Incorporating community monitoring was tied to the many volunteering efforts in the AOI concerning seagrass restoration in addition to literature emphasizing their potential in aquatic vegetation mapping [2]. Firstly, the potential is in the number of volunteers. Take the U.S. for example, reporting on 64 million people volunteering every year [32]. The second potential is the Internet's ability to leverage volunteering work through 'open source', 'open access', and 'crowdsourcing' [33, 34]. 'Open source' refers to source code anybody can access and modify [35]. 'Open access', refers to scientific literature, such as journals [36], while crowdsourcing refers to information, goods, or services anybody can obtain for free, like data or funding [37, 33, 38]. Also, with the Internet, the term 'community' is broadened to include any voluntary work online, hence connecting the local volunteer to a global network. All levels are needed for mapping seagrasses and seaweed as their decline is a global problem, while simultaneously, each place requires unique adaptation due to differences in environment. The local level must therefore be included, as locals, and especially indigenous communities, possess the best place-based knowledge and is starting to gain wider recognition [1]. Therefore, community monitoring is accounted for in this thesis, as an important dimension that can take part in seagrass and seaweed mapping as they are powerful in numbers and leveraged

[1]Code for this work can be accessed via:$https://shorturl.at/yHKL8$

through technology by connecting the local to the global level.

1.4 Research Questions

Motivated by the advancement in freely available technology and data in RS [39, 40], coupled with Artificial Intelligence (AI) [41] and a wish to aid community monitoring efforts [42, 43], this thesis will answer the following Research Question (RQ)s:

RQ1 [Accuracy] How accurate are the Machine Learning (ML) techniques (supervised, unsupervised, and semi-supervised) in classifying seagrass (specifically considered here is *Zostera marina*) from seaweed (mainly, *Ascophyllum nodosum*) using only freely available data and software?

RQ2 [Implications and Applications] What are the potential applications and implications of using ML for classification in Google Earth Engine for community monitoring and environmental management?

To answer these questions, a case study in Casco Bay and Saco Bay, Maine U.S. will be performed to aid recently launched local efforts in mapping seagrass [31]. This is done on the cloud software, Google Earth Engine and Google Colab, using free satellite data by Sentinel-2 Level 2A and spatial analysis performed with the Python programming language.

1.5 Related Works

A quick overview of related works is given in Table 1.1 and elaborated on in chronological order in the subsequent paragraph. Related works were found through search terms in Google Scholar, using keywords such as 'Sentinel 2', 'seagrass mapping', 'machine learning', 'GEE' (a cloud computing platform), and 'Seaweed', resulting in a total of 11 comparable studies (listed in Table 1.1). Two main criteria were used to deem a work as relatable; the work used Sentinel-2 data and combined it with ML to classify aquatic vegetation. Further similarities and dissimilarities in data collection, methodology, software, AOI, and results are described and compared to this study below.

The earliest related works are from 2018 [26, 44] as Sentinel-2, consisting of two identical satellites (Sentinel-2 A and Sentinel-2 B), was only first launched in 2015 and 2017 respectively [53]. Among similarities between them and this work are the goal of creating a scalable workflow to map seagrasses [26]. Other similarities are type of model, methods to derive features, and features such as visible light (the spectral bands B2, B3, B4), and Near-Infrared (NIR). Dissimilarities are in the AOI, additional spectral

Table 1.1: Overview of related works. The most essential processes of this thesis are included in the table. Method - Image classification method, either PB or OBIA. Resolution - Spatial resolution. Model - Either SML, SSL, or UML machine learning. Software - Considered if GEE is used, if not, other. AOI - Area of Interest by marine region [7]. Results - The best overall accuracy score, if not documented, F1-score is used.

Study	**Method**	**Resolution**	**Model**	**Software**	**AOI**	**Results**
[26]	PB	10m	SML	GEE	Mediterranean	72%
[44]	OBIA	10m	SML	other	Baltic Sea	72%
[45]	PB	10m	SML/UML	other	South Pacific	97.3%
[46]	PB	10m, 20m	SML	GEE	Global	88%
[47]	PB	10m	SML/UML	other	Western Indian	82%
[48]	PB	10m	SML	GEE	Western Indian	79%
[49]	PB	10m	SML	GEE	Western Indian	60.71%
[50]	PB	10m	SML	GEE	Baltic Sea	86%
[19]	PB	10m	SML/SSL	GEE	West Atlantic	70%
[51]	PB	10m,20m	SML	GEE	Mediterranean	72 %
[52]	PB	10m	SML	GEE	West Atlantic	97%

bands (e.g., B1-coastal aerosol and B11-SWIR1), methods of obtaining bathymetry data, other types of classes such as seagrass (*P. oceanica*) and non-seagrass, and models like support-vector machine (SVM). One of these works were also similar by being a master thesis [44], enabling one to compare scope, model, and method. The main differences lie in that it did not use GEE as the software, and performed no hard test like here.

In 2019, one study [45] tested both an SML (with commercial software ENVI) and the UML model, Iterative Self-Organizing Data Analysis Technique (ISODATA), using three and 20 classes respectively. The best results were achieved with the ISODATA and 20 classes, giving an OA of 97.3%. Thus, the main difference here is the use of commercial algorithms, pre-procession, the type of algorithms, classes, and AOI.

In 2020 four similar studies were found. The first one [46] differed from this thesis in spatial resolution, algorithm, AOI, and features such as the Floating Algae Index (FAI) and Kelp Difference (KD), but like this study performed a hard test which reached an OA of 88%. Secondly, a very similar study [47], differed only in the use of Principal Component Analysis (PCA) and type of classes: coral reef, seagrass and Submerged Aquatic Vegetation (SAV), sand, and deep waters. Reaching an OA of 82%. The third work [48], differed in that it did not use GEE as software, and specifically looked at detecting SAV vs non-SAV. However, initially they wanted to classify seagrass from seaweed, like in this work, but found it not feasible due to too similar spectral signature between them. Their OA for the detection of SAV was 79% with a soft test. The fourth work was like the other parallel publications in both method and software, only differing in the types of features and achieved OA of 62.27% [49].

In 2021 another three studies were published ([50, 19], and [51]). Their main simi-

larities to this work were the use of GEE, and testing RF as an SML model. The main difference were in the features, i.e., FAI, for [50], green/blue depth invariant indices and geomorphological data for [19] and the use of spectral bands B1 and B11 for [51]. Lastly, one study was found published in 2023 [52]. Unlike any of the previous related works and this, it tested the biggest number of features, achieving an OA of 97% with RF.

To summarize, all studies used the spectral features of visible light, NIR, and some version of bathymetry like this work. All except one used the PB method, an iImage classification method (see chapter 2 section 2.1.2). SML models were the most prevalent, followed by a few studies testing UMLs, and only a single one testing an SSL with a weakly supervised U-Net achieving an F1-score of 0.71% ([19]). GEE was the software used by the majority. The accuracy ranged from 60.27-97.3% with a median OA of 79%. Compared to this work, few were found to use both soft and hard tests, due to a lack of independent validation data from a different site - a common problem in the field of aquatic vegetation classification - more so for non-tropical zones than others [1]. Only on study [19] used an impressive five AOI for soft testing (training) and the sixth site for the hard test (testing). Comparably, this work used one site for soft testing and one for hard testing. Furthermore, while many used GEE, only this thesis used geemap, a Python package for interactive mapping with GEE, and an SSL model based on Random Forest and pseudo labels.

1.6 Contributions

This thesis contributes a comprehensive approach to classifying seagrass from seaweed and water in Casco Bay, Maine, utilizing ML and GEE. The proposed approach analyzes various methods and tools to improve the accuracy of seagrass and seaweed classification, which is essential for the monitoring and management of coastal ecosystems.

First, this thesis begins by classifying seagrass, seaweed, and water using ML and GEE techniques. This method permits the efficient and precise processing of massive amounts of satellite data, which can be difficult to manage using conventional methods. This thesis demonstrates how ML algorithms can effectively classify seagrass, which can aid in the monitoring and management of seagrass ecosystems.

Second, using Sentinel-2 L2A, this thesis evaluates a semi-supervised classification model for seagrass. This strategy has the potential to reduce the cost and time necessary to collect labeled data, a common challenge in remote sensing applications. Using Sentinel-2 L2A data also provides high temporal and spatial resolution, relative to other freely available satellites, which can enhance the classification accuracy of seagrass.

Thirdly, this thesis examines the 'community monitoring' role in seagrass and seaweed classification. Community monitoring uses volunteers to collect and report data, which can supplement conventional data sources. This thesis emphasizes the potential for community monitoring to increase the amount of labeled data available for seagrass and seaweed classification, thereby improving the accuracy of ML models.

This thesis underscores the significance of reproducible results which enables other researchers to replicate the results and expand upon the work presented in this thesis. Additionally, the availability of the code facilitates collaboration and knowledge exchange among researchers in the field.

This thesis utilizes GEE's geemap integration, first launched in 2020 [54]. This integration provides an intuitive interface for exploring and visualizing satellite data, which can facilitate the data analysis process for researchers and practitioners. This thesis uses this instrument to demonstrate how GEE's capabilities can be expanded and enhanced to support seagrass classification and other remote sensing applications.

1.7 Thesis Overview

This thesis presents a comprehensive exploration of RS with ML models to map and classify seagrass from seaweed and water. The subsequent chapters delve further into this thesis's goals, process, and contributions. Chapter 1 introduced and motivated the topic, in addition to examining existing literature informing this work. Chapter 2 focuses on theory, and provides definitions and foundation of claims, assumptions, and methods used in the experiment. In Chapter 3, the method is presented, from data collection, pre-processing of satellite data, features selection, training, validation, and testing of the MLs, to final predictions. Chapter 4 provides the empirical findings, comparing the models against different performance metrics to evaluate them in chapter 5. From the results derived, Chapter 5 discusses the results in relation to the research questions, and compares them to related works, highlighting both potentials and flaws of this thesis. Lastly, Chapter 6 concludes with what has been discussed and relates the findings of this thesis to the bigger picture, and proposes recommendations for future research. Together, these chapters aim to create a complete overview of this thesis research.

Chapter 2

Theory

The theoretical framework will be present and focus on three overarching concepts to understand their influence on this thesis. First, RS will be discussed in relation to the choice of mapping technique, and image classification method suitable for detecting aquatic vegetation. The second topic shortly describes the software and hardware used, as they are relevant for this thesis methodology and making results and workflow reproducible. Thirdly, a general overview of ML is described in relation to the field of data science, followed by the types of ML models in the field of classification to provide an understanding of the choice of ML models for this work. By examining RS, software, hardware, and ML, this thesis aims to build a solid foundation for the coming chapters on methodology, results, and discussion.

2.1 Remote Sensing

RS is a technique to collect data on a system from a distance [55]. In this thesis, the system is aquatic vegetation, and techniques of RS are distinguished by scale [56]. A distinction is made between broad-scale, meso-scale, and fine-scale depending on the area coverage of the sensor. Broad-scale includes air-born, optical sensors such as drones, airplanes, and satellites covering large grounds. Examples of meso-scale RS are on-the-ground/water equipment, such as a sonar attached to a boat, covering medium-scaled areas. Fine-scale is often performed by community monitoring volunteers, students, and scientists collecting data by hand. In this thesis, broad-scale with satellite Sentinel-2 L2A was used because of its free data, and high spatial and temporal resolution.

2.1.1 Broad-scale Mapping

Mapping and collecting data on aquatic vegetation are usually done through broad-scale mapping [57]. Globally, satellites are the most common [48], while at the national level,

it is airplanes. An example is the National Agriculture Imagery Program which maps the U.S., state by state, approximately every 3-5 year [58] by airplane. In Maine, this is enacted by the Maine Department of Environmental Protection (MaineDEP). With the last mapping of Casco Bay in August 2018 [59], and the next mapping in August 2022, with data released in March 2023. While mapping by airplane yield a high resolution, the temporal frequency of 5-years, considering the rapid decline of seagrass and seaweed, becomes too infrequent, labor intensive, and lacks scalability. Thus, satellites are an attractive addition [4].

Starting with the launch of Landsat-1 in the 1970s, both the number of satellites and their resolution have increased in quality manifolds [60]. Landsat-8, WorldView-2 (WV2), and Sentinel-2 are the most used satellites in seagrass and seaweed mapping. Landsat-8 has a spatial resolution of 30 meters, free data, a historical archive dating back to 1972 (good for time-series analysis), and its latest versions are equipped specially to monitor the environment (both land and sea) [4, 61, 1]. WV2 is a commercial satellite with a spatial resolution of $\sim$50cm [62]. Sentinel-2 has the highest spatial resolution of freely available data, 10 meters, and a temporal resolution of 5-10 days [63] (due to it's twin constellation) and was therefore chosen for this work.

Three main barriers should be considered when using satellite data. The spatial resolution, radiometric errors, and geometric errors. Spatial resolution, ranked as either high ($\leq 5m$), moderate ($5m - 30m$), or low ($30m+$) [4] plays a big role in classification, illustrated by Figure 2.1 (next page). It shows the same image of seagrass, taken with different spatial resolutions, from high (left) to low (right). In the first image (to the left) taken with a drone and a high resolution of 4cm, one can clearly distinguish seagrass from its surroundings. Progressing from left to right, the spatial resolution decrease, and the more pixelated or 'blurred' the image becomes, making it harder to distinguish the seagrass from its surroundings. Sentinel-2 is regarded as having a medium resolution and despite the visible blurriness seen in Figure 2.1, experts would recommend using it if costs were to be an issue [4]. Hence, experts in the field have expressed optimism about its potential in mapping aquatic vegetation given optimal environmental conditions, enough data, and sufficiently good processing. For these reasons, the satellite data of Sentinel-2 Level-2A were found suitable to use here.

The second barrier is radiometric errors. They can either be internal or external and cause brightness distortion in an image. Internal errors refer to errors by the instrument [64], seen as white or black pixel noise (also referred to as 'no data pixels' [65]). These pixel-values must be removed to obtain reliable information and to retrieve valid data from the satellite images. While the pixel's loss of information is irreversible, it can be interpolated with a new pixel-value approximated by the average or median values of neighboring pixels. This type of interpolation is only possible, assuming the first law

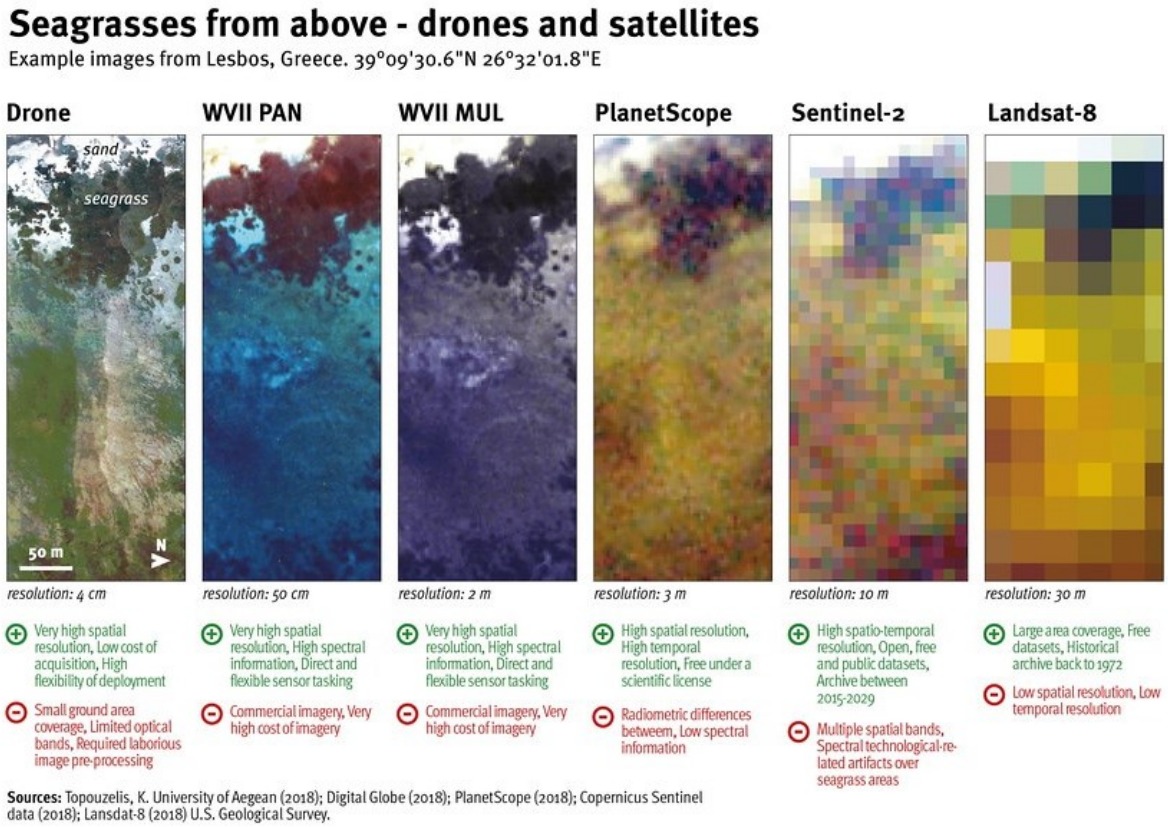
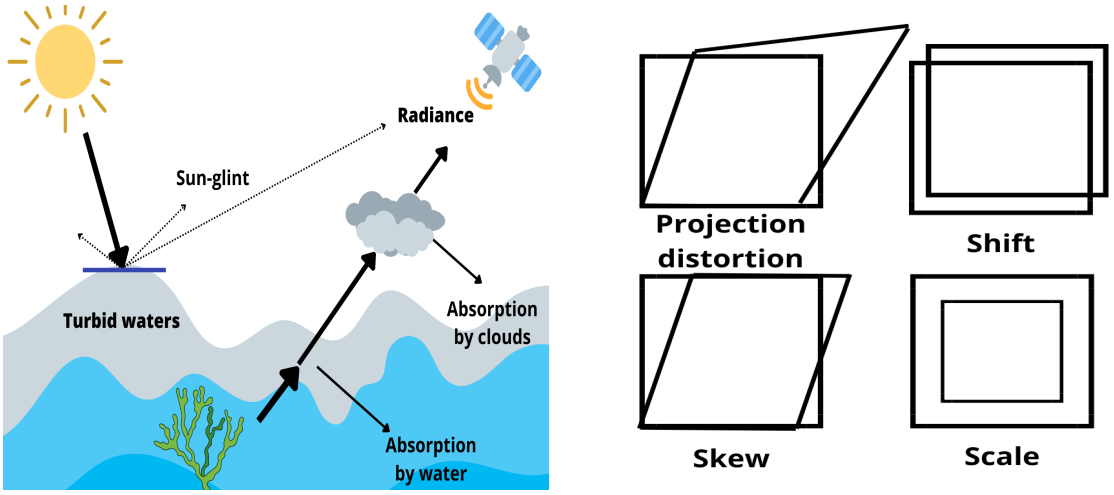


Figure 2.1: Comparing the spatial resolution of drones and satellites. From left to right. Drone, WorldView-2 (WVII) pan-sharpened (PAN), WVII Multispectral (MUL), Planetscope, Sentinel-2, Landsat-8. [5]

of geography holds: 'closer things are more similar than distant things'. Compared to internal errors, external errors are caused by the atmosphere interfering with the incoming reflectance from the object to the sensor. Figure 2.2(a) illustrates the external radiometric errors of a passive satellite, such as the Sentinel-2 used here. Between the sensor and the sensed object, there is a range of atmospheric (e.g., haze, fog, aerosol) and biophysical entities, i.e. water, in the case of this thesis, that will absorb parts of the emitted light from the sun, causing a discrepancy in the object's reflectance sent back to the sensor [66]. In Level-2A, atmospheric correction of surface reflectance has already been applied, removing haze in an image [67], leaving radiometric errors such as sun-glint and water-column correction to be removed manually. While sun-glint correction was applied in this thesis, the water-column correction was not, see chapter 5 on 'Thesis limitations'.



(a) External radiometric errors. Illustration of phenomenons who can obscure the incoming reflectance from the object being sensed.

(b) Geometric errors. Illustration of four potential errors that can be observed in the image.

Figure 2.2: Difference between relevant radiometric and geometric errors for this thesis.

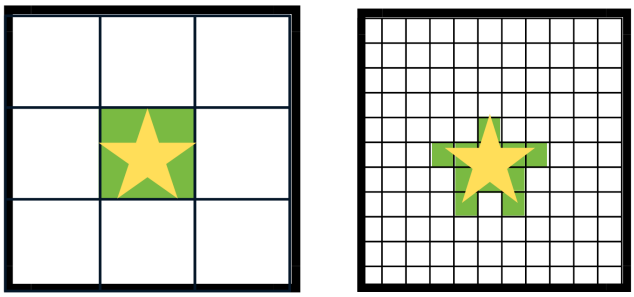
Lastly, the third barrier to be considered when using satellite data is geometric errors. Figure 2.2(b) exemplifies the effect geometric distortion has on an image, i.e. scale, skew, shift, and/ or projection distortion. The common denominator for all of them is that pixels end up deviating from their proper map location. Like radiometric errors, geometric errors can also be either external or internal. Internal errors are caused by the sensor systems themselves, or the rotation and curvatures of the earth [68]. Geometric external errors are mainly caused by changes in altitude and rotation of the aircraft in relation to the Earth. After addressing the cause of distortion, correction can be performed accordingly [68]. In the context of this thesis, only the internal geometric error of skew will be discussed, given the use of Sentinel-2 L2A which delivers orthoimages atmospherically corrected by the European Space Agency (ESA) [69]. The skew defect is caused by how the detectors in Sentinel-2 are stitched together for a 290km field of view with 10m spatial resolution [70], resulting in different viewing angles by its detectors. To correct for it, sun-glint correction should be sufficient [71] and is thus part of the pre-processing here.

2.1.2 Image classification method

The choice of image classification method matters. Two types exist, PB or OBIA, and will be explained in the context of this thesis. The PB method is the default and represents the scenario where the smallest entity in an image is a pixel. The OBIA method applies segmentation to the image, for instance, by clustering pixels into 'super-pixels' or 'objects'. A further distinction is made between OBIA and GEOBIA [72], which is the special case where objects are geographical, compared to any object. Considering this thesis, the term

GEOBIA will be used. Furthermore, choosing to use PB or GEOBIA for classification depends on the satellite *spatial resolution* and the object's *size* that are being classified [48].

Figure 2.3 exemplifies the effect size and spatial resolution have for the choice of image classification method. The figure shows an image of low 2.3(a) and high resolution 2.3(b) for the same object (a star). In Figure 2.3(a) the object is smaller than the pixel, and would thus not be detected by any super-pixels from the GEOBIA method, hence the PB method is preferred. In Figure 2.3(b), the image has a high resolution and a super-pixel could potentially capture more of its shape than individual pixels as the GEOBIA method accounts for similarity between neighboring pixels, while the PB method does not. That the PB method does not account for the similarity between neighboring pixels causes a common problem called salt-and-pepper noise, seen in an image as disconnected pixels, or in Figure 2.3, it would look like disconnected squares around the object, compared to the connected ones currently seen in 2.3(b).



(a) Low resolution. (b) High resolution

Figure 2.3: Difference between pixel-based (PB) (a) and object-based (OBIA) image classification (b) given different spatial resolutions for the same object (star). Green pixels in the back of the object illustrate what the respective methods capture of the it.

In general, literature show that using GEOBIA significantly improves classification in RS [72, 60, 4, 73, 74]. Of related work, only one ([44]) chose to use GEOBIA achieving acceptable results. The others did not, with the main reason being that they found Sentinel-2's spatial resolution to be too low, and the detectable objects too small to choose it as an image classification method. However on of the related works ([26]) who used the PB method proposed to use the GEOBIA method as a future endeavor to improve their current work. In this thesis, both the PB method and the GEOBIA method were used based on suggestions from literature to explore the GEOBIA [4], and literature on the size of seagrass and seaweed habitats in the AOI [29].

2.2 Software and Hardware

For later comparison, reproducible purposes, and community monitoring, the hardware and software are described. Of hardware, the properties of the computer will be described, while for software, GEE and Google Colab (GC) will be presented.

2.2.1 Hardware

This work was done on a Yoga 730-13IKB laptop (ideapad) - type 81CT. Processor with 8th generation intel® core™ i5-8250U, a memory of 8GB DDR4 2400, and Windows 11 home version 21H2 as the operating system. In relation to community monitoring, it's noted that Windows is a commercial operating system, however, the same workflow can be run from any computer with other, free types of operational system, such as Linux, as all software used are performed on the free cloud computing of GEE and GC.

2.2.2 Software: Google Earth Engine

Google Earth Engine (GEE) is openly accessible cloud computing software, made for fast processing and manipulation of satellite data, launched in 2010 [75]. Its cloud-based property enables it to store large quantities of data remotely by powerful computers that can be used and retrieved from anywhere online. This makes it especially useful in the field of RS, collecting an ever-increasing amount of data. Comparable commercial software to detect aquatic vegetation is ArcGIS, TerraSet, and PCI Geomatica as used in related studies such as [76, 77, 44]. Non-commercial software would be QGIS with R, used by related studies such as [44] and [48]. Compared to the other freely available software, GEE is advantageous in that it is a multifunctional platform, storing satellite data on the cloud to make data analysis a fast process, in addition to featuring a toolbox of algorithms, satellite datasets, and a code editor based on JavaScript. Documentation on further adaptations of GEE to Python exists, launched in 2020, enabled through a package called geemap, supporting an even wider selection of data analysis tools. Furthermore, while commercial software is often aimed to be used by experts, GEE and QGIS, as they are free and aim to be used by the general public are arguably more user-friendly [4, 75]. For these reasons, GEE with geemap is used in this work, with all parts of the process housed by GC.

2.2.3 Software: Google Colab

GC, also on the cloud like GEE, houses Jupyter Notebook, a coding environment available in the browser to anybody who wants to write and execute Python code [78], without requiring any downloads or installations, making it a suitable analysis tool for community monitoring. Moreover, through the geemap package, the workflow from fetching satellite data from GEE's data catalog, and pre-processing, to data analysis, can all be facilitated in the same space, inside GC, enabling a single continuous pipeline. Moreover, the creators of the geemap packages have provided a range of online tutorials paired with open-source code, that can be executed in GC [54], facilitating those without any pre-knowledge in coding to be able to learn and use the tools. Furthermore, 1 GPU can be borrowed from

GC. This enabled jobs on the computer to run in parallel (restricted to 12 hours at a time) speeding up the workflow and training of ML algorithms. On the flip side, unless the data process is well automated in GC, the runtime shuts itself off if one stays offline for too long to minimize unnecessary computational power. For each new runtime session, locally stored data have to be uploaded, libraries installed, access to GEE re-authorized, and session re-run, causing a slightly inconvenient time delay. However, this can be combated with good coding flow and planning. For these reasons, despite GC's inconvenient self-regulating runtime, its advantages (GPU, cloud-based, geemap-integration, ease) outweigh the downsides and are the reason GC was used as the interface for data analysis in this thesis.

2.3 Machine Learning

In the age of technology, ML permeates modern society and has shown great potential in the field of RS [41]. Examples of use cases range from everyday tasks like sorting e-mails to more specialized ones, such as classifying seagrass from seaweed. ML is just one of many sub-fields of AI, others at a similar level would be natural language processing, while a further specialization of ML would be DL [79]. An overview of these fields are shown in Figure 2.4.

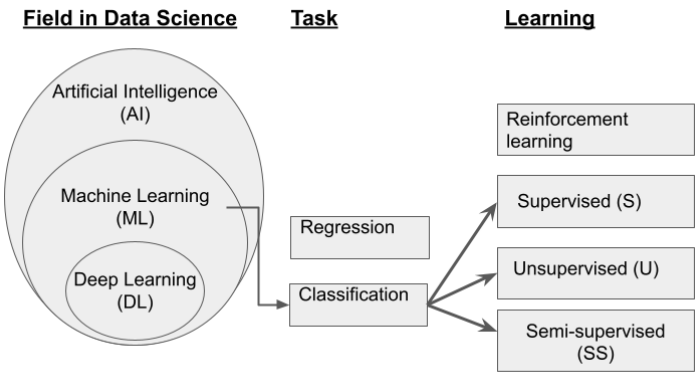


Figure 2.4: Data Science concepts. From left to right, The Field of Data Science comprises AI, the sub-field ML, which again contains DL. The tasks of ML are classification or regression. Classification can either be supervised, unsupervised, or semi-supervised learning.

AI is a technology that mimics the human brain's abilities in decision-making, and problem-solving. Two tasks it is most concerned with are called regression and classification [80]. Both are predictive modeling, but the former is used to find relationships between variables and predict continuous outcomes, e.g., the relationship between height and weight, while the latter predicts qualitative outcomes based on qualitative or quantitative data. For instance, given a satellite image, each pixel in that image contains a value, also called a Digital Number (DN), who's data can be used to classify key species such as seagrass, seaweed, and water form each other. Additionally, the rapid processing power and storing ability of modern hardware enables AI to process, and connect massive amounts of data, like the human brain, only more powerful (depending on the task), thus, ML in relation to RS shows great potential in reducing time, labor, and costs.

In Figure 2.4 the arrows show this thesis choice of field, task, and learning method used. Types of learning are mainly categorized as SML, UML, SSL, and reinforcement learning. This thesis explores the first three, thus only these will be explained. SML and UML will only shortly be described providing the necessary foundation to understand the SSL model. Shortly summarized, these models can mainly be distinguished by the data fed to them. If the input data contains labels, the ML is supervised, when it does not, it is unsupervised, and semi-supervised learning is the combination of SML and UML.

2.3.1 Supervised

The fundamental characteristics of the SML model are the use of labeled data and it's predictions for direct feedback [81]. The objective function of SML is to minimize loss, informed by the feedback. Used in both regression and classification problems, only classification is relevant and will be described here. Figure 2.5 illustrates SML "behind the scenes". For each input value, x (x $\in$ set X), the respective true class, y (y $\in$ set Y), is known. Thus, errors are kept track of by comparing the predicted class, $\hat{y}$ ($\hat{y} \in$ set $\hat{Y}$), with the true class, sending direct feedback to the model that it can use for "training" until satisfying results are acquired.

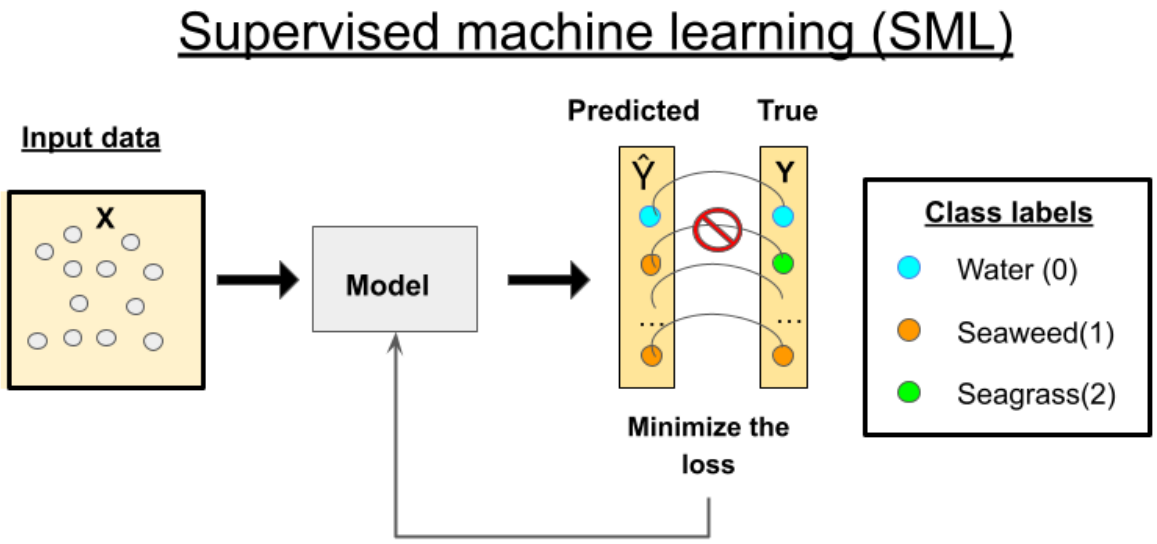


Figure 2.5: SML. Given labeled input data, the objective function of SML aims to minimize the loss by training the model on direct feedback from its predictions.

In the field of RS using ML on classification tasks, the most used classifiers are supervised. Examples are the Maximum Likelihood Classifier, RF, SVM, and K-Nearest Neighbor [82] where the best performers are RF and SVM [83, 4]. The benefits of this method are that, given enough good quality data, the classification task is straightforward and yields high accuracy and robustness [48, 19, 84, 85, 86] as misclassification can easily be identified by comparing the predicted class to the true class. Despite SML's many benefits, some challenge are presented in the case of seagrass and seaweed classification. Firstly, the requirement of enough labeled data is rarely fulfilled due to a lack of research in the field [4, 2]. Secondly, the requirement for high-quality data depends on the collectors, pre-processing phase, and analysts. Thus, there are many phases that can be affected by human error in addition to internal and external geometric and radiometric errors from a sensor or its surroundings as discussed in chapter 2.1. Consequently, if the data is not handled properly, these errors will propagate into the results, making them invalid.

2.3.2 Unsupervised

UML, in contrast to SML, is characterized by the use of solely unlabeled data, no feedback or predictions but instead finds underlying patterns [81]. Figure 2.6 illustrates this behavior of UML. Given a set of input values X, its objective function aims to cluster the data points by similarities and dissimilarities. As the UML often base themselves on clustering, they also go by the name *unsupervised clustering* or just *clustering* [87, 88, 89].

Given the above challenges presented for the SML, researchers in the field of RS are exploring the use of UML and SSL which are showing potential [45, 90, 25] in mapping aquatic vegetation as they do not need no training data at all, or, can work with a small amount. Another benefit of the UML is that they are mostly automated, needing few parameters to tune [91, 89], resulting in higher computational efficiency. Commonly used UML models in the field of aquatic vegetation mapping are, K-means (e.g., [60]), ISODATA, (e.g., [92, 60]), and agglomerative hierarchical clustering [73]. Overall, UML yields lower accuracy than SML. For example, three studies [76, 93], and [48] did a comparison of SML and UML in aquatic vegetation mapping, and found both that the accuracy provided by the UML was insufficient [93] and that the models yielded equally good results [76].

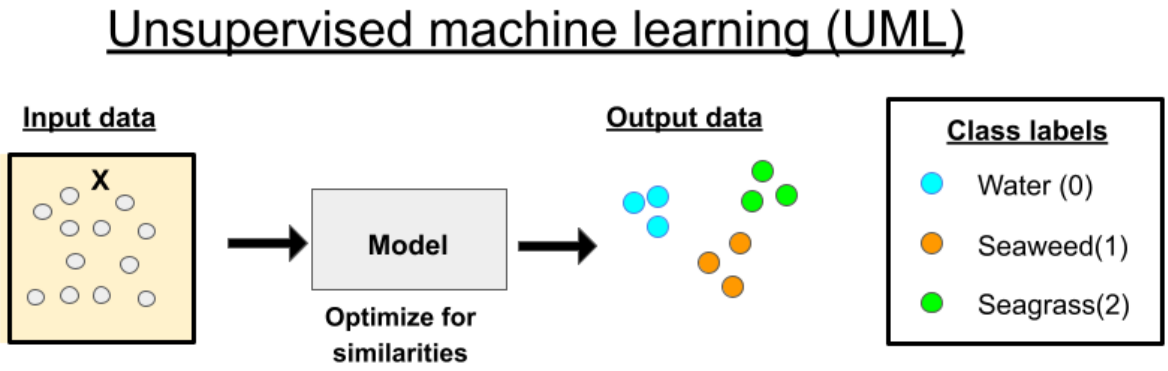


Figure 2.6: Illustration of UML behavior. Given an input set X to the model, its objective function is to group data points together based on some similarity metric. This outputs a set of clusters.

2.3.3 Semi-supervised

SSL, also referred to as 'weekly supervised learning' is the combination of SML and UML [94]. Several methods of SSL exists and in this thesis, the common method of self-training is employed [95]. To provide some intuition on how self-training works, this process is similar to how children learn. The child is given a small amount of labeled data by their parents (apple, dog, mom, dad, etc.) but they mostly learn by encountering a big amount of unlabeled data which they have to predict labels for. Thus, for a machine, self-training

means that a model is trained on a small amount of labeled data, followed by receiving a big amount of unlabeled data which it predicts labels for, then re-trains using both the small set of labeled data merged with the big amount of predicted labels. In the field of RS a big and frequent amount of unlabeled data is accumulated while only a limited amount of labeled data exists, thus being able to utilize the combination of the two is of intriguing inquiry [96, 97], especially in the domain of aquatic vegetation [98]. Therefore, classifying seagrass from water and seaweed provides a great opportunity to test the SSL technique.

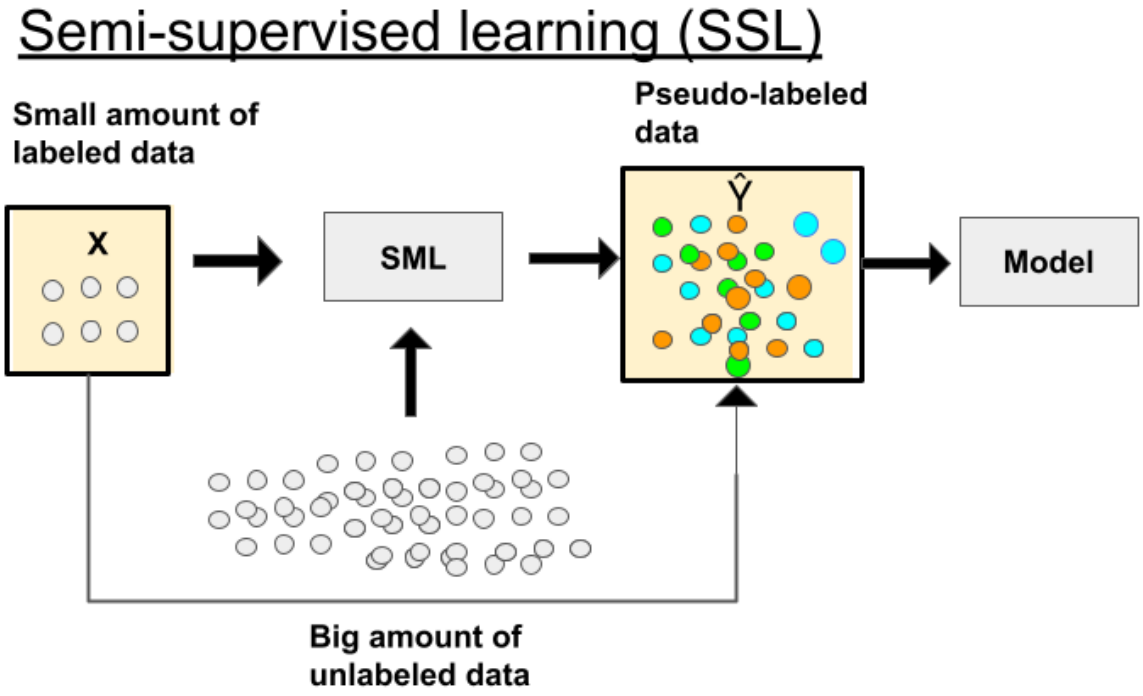


Figure 2.7: Illustrating the SSL workflow. First, an SML model is trained on a small set of labeled data. Then, it uses this training to predict the values of a large set of unlabeled data. The prediction produced are re-used together with true labels, making a pseudo-labeled dataset which the SSL model trains on to create the final SSL model.

There are three underlying assumptions that must be accounted for when using SSL to make use of unlabeled data [94]. First is the smoothness assumption, just like the first law of geography, it assumes points who lie close to each other in a given feature space are more similar and consequently label them as such. Second, is the cluster assumption, which considers that similar data points form clusters and dissimilar data points behave "repulsively". Lastly, the manifold assumption expects the data to lie close to the manifold but of a lower dimension. The manifold is a topological space, an approximation of the real Euclidean space, just with a smoother surface. This is necessary for the definitions of distances and densities to hold true which are based on the manifold. These assumptions, in addition to the background knowledge of SML and UML, provide the foundation of

how the SSL works.

Exhibited benefits of SSL are easy implementation, high effectiveness, and the need for few labeled data to train on [98]. The disadvantages are found to be weak performance [98], overfitting, and negative transfer [99]. However, from 2010 and onward, SSL in combination with DL has had many success stories [100, 25, 101, 102, 96].

Regarding the SSL classification of seagrass and seaweed, only a single related work was found ([19]), given the criteria of this thesis. However, when broadening the scope to include other types of RS data sources and deep learning, a deep Convolutional Neural Network (CNN) was found to achieve 99.99% accuracy in detecting seagrass [90], while another deep Fully Convolutional Neural Network (FCNN) achieved an accuracy of 81% in monitoring seagrass [25]. It is also worth noting that both methods used a much higher spatial resolution compared to this thesis' Sentinel-2, which makes their results not comparable. To summarize, despite more challenging implementation as a lack of standardized methods and software for SSL, it is recommended to be further developed it, especially within RS.

2.3.4 Models

In this thesis, the performance of three models was assessed and compared. The choice of models was based on the preceding theory and related work, in addition to what was feasible regarding time limits for this thesis. Considering the classification task, a multi-classification problem of seaweed, seagrass, and water, the chosen models were: RF for SML, K-means for UML, and a combination of RF and pseudo-labels for the SSL model. Many more models and learning techniques exists and have been used in benthic habitat mapping, such as transferred learning[98, 60, 83, 92]. Additionally, as demonstrated above, DL's entry into the field is mention-worthy considering the encouraging results [102, 103, 104, 74]. Nevertheless, none of these will be further discussed here given the scope of this thesis.

2.3.4.1 Random Forest

In the field of mapping aquatic vegetation, RF has proven to outperform previous state-of-the-art classification algorithms, i.e., Maximum Likelihood [86, 48, 105, 85]. RF is an ensemble of Decision Tree (DT)s. A DT consists of a root, decisions- and leaf nodes. Starting at the root, a dataset of features and samples is inputted. Features provide the splitting criteria at each node with the objective function to maximize Information Gain (IG). IG is the difference in purity of a parent's and child's node. A fully pure node is one where the samples are of the same class. When a node fulfills the purity criteria the tree stops growing, and the leaves represent the predicted classes. For the

RF, many configurations of DTs are created, simulating multiple scenarios. The average accuracy is calculated over them, which decreases the variance produced by every single tree, consequently making RF a robust algorithm [106]. Additionally, the RF has relatively few hyperparameters [48], making it easy to handle. Therefore, recommendations to use this algorithm are owed to its inherent traits of robustness and ease.

While RF poses many advantages it loses interoperability and becomes computationally expensive with increasing amounts of data. Loss of interoperability is a problem when one needs to understand the threshold values and features chosen by the RF as the splitting criteria, as they can give insight into how they affect the resulting output (e.g., accuracy). With lots of data, the visualization of RF will be messy and hard to interpret as all decisions will be displayed. High computational costs are rooted in the objective function. In RF's quest to maximize IG, it does this brute force, with no consideration for the number of splits, trees, or how deep each three should grow. In other words, if the purity-criteria is very strict, the algorithms will run until all possible combinations are tried, which can be time-consuming. For instance, if the samples have very similar characteristics, but belong to different classes, the tree might grow until it only has a single sample in each leaf. This can however be fixed, by tuning its hyperparameters [81].

Similar to related work, this thesis chose to use RF based on its documented advantages and considering the context in which it will be employed in. Of related works, many of them tested RF with the highest F1-score equal to 86% for a soft test [50], while for a hard test, an OA of 72% [107]. Their choice of using RF is due to the above-mentioned advantages, in addition to being robust to overtraining [108], making no assumptions on the data distributions fed to it, and needing less training data compared to the Maximum likelihood [48]. Furthermore, considering this thesis's use of freely available data, and the aim of making a easy repeatable workflow for community monitoring, the RF algorithm was chosen as the SML due to its simplicity, robustness, and high accuracy.

2.3.4.2 K-means

K-means is an unsupervised clustering algorithm. While the ISODATA is claimed to be a more appropriate algorithm as a UML in marine habitat classification [45], the K-means was chosen over ISODATA because of easier accessibility and show equally good results in similar studies [60, 109]. Given N number of data points, k elements in N are chosen as the centroids (or medoids) for each cluster. The remaining N-k data points are then assigned to each of the clusters based on the similarity *within* a group. How similar a point x is to k is given by the Euclidean distance metric between features. The objective function of K-means is therefore to minimize this distance, measured as the within-cluster sum of squared errors [81].

2.3.4.3 Random Forest and Pseudo labels

The same process of self-learning applies to RF using pseudo labels, like it would for any other model [94]. The process can be divided into two stages. The initial stage is supervised, where a small amount of labeled data is fed to an RF algorithm for training. The newly trained model goes over to the unsupervised stage, where it is fed a large amount of unlabeled data which it predicts labels for. The predicted labels and the true labels (from the initial supervised stage) are merged, forming a big pseudo-labeled dataset. This is then used to train a new RF algorithm.

2.3.4.4 SNIC

To perform the GEOBIA method in this thesis, the satellite image must first be segmented, hence the Simple Non-Iterative Clustering (SNIC) algorithm was applied. It was proposed in 2017, based on the previous state-of-the-art Simple Linear Iterative Clustering (SLIC) [110] and outperforms its predecessor with less memory, higher speed, and a simpler algorithm [111].

SNIC performs segmentation through clustering pixels into super-pixels by a five-dimensional color space called CIELAB. Here, pixels are clustered based on color and spatial distance. While SLIC used K-means to cluster, SNIC uses a priority queue. Its clustering can be described in three steps. First, seeds or centroids are initialized in a regular grid fashion in the five-dimensional color space. Second, the distance between each pixel to the centroid is measured and pixels are added to a priority queue in ascending order of distance. Third, super-pixels are created in a growing fashion by adding one pixel at a time from this priority queue, eventually creating an image consisting of objects. From these objects, additional features can be derived, such as width, length, perimeter, area, and number of clusters, as used in this thesis. Thus, the SNIC algorithm was chosen based on its state-of-the-art performance, in addition to already being integrated with GEE, making it easy to use for this work.

2.4 Model Evaluation

Model evaluation is the act of validating how well an ML model performs [81]. A difference between internal and external validation is made. Internal validation - onward referred to as 'validation', is done during the process and is used by the researcher to understand how the model may generalize on unseen data. This thesis used Cross-Validation (CV) for internal validation. External validation, and testing, are done on the unseen data and give the final evaluation of the model's performance. The metrics used were OA, F1-score, precision, recall, confusion matrix, and Tau(e) for external validation, chosen on the basis

that the results of this thesis could be compared to related work.

2.4.1 Model Evaluation on Training

In cases where data for training and validation are scarce, CV is a good internal validation metric to obtain reliable validation accuracies. It is not only the data, features, and choice of model that affect the accuracy but also how the data is split. To minimize the amplitudes of varying accuracies, CV, also called k-fold CV can be applied.

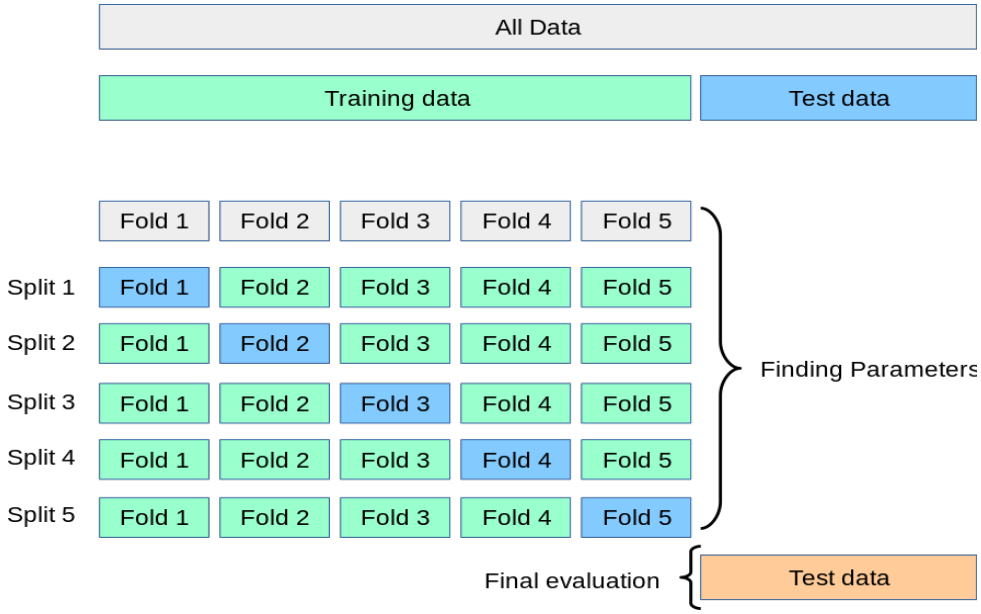


Figure 2.8: Illustration of a 5-fold Cross Validation (CV). All data represent the total amount of data. Data is split into training and a test dataset. The training data is further split into five folds. For each split, one fold is used for validation, while the remaining is for training. This is repeated over five splits. Through the training and validation, the optimized parameters are found, making a trained model. This trained model is then tested on the hold-out test data for a final evaluation [6].

To identify errors within the model and improve its generalizability, internal validation is important [112]. The CV process is shown in Figure 2.8. Given a dataset, with CV it is first partitioned into a training set and one test set. Then, the training set again split five times into five folds, where four out of five will be used for training and the remaining for validation. The model is trained and validated, such that each fold becomes the validation set once. For each split, respective validation accuracy is reported. After the training and validation on all folds are performed, the reported accuracies are averaged to provide a more representative performance of the model. Assuming the validation accuracy is satisfying, the model can be tested on an unseen dataset (test data) for external validation.

2.4.2 Model Evaluation on Testing

Model evaluation on external unseen data provides the ultimate summary of how well the algorithm performed. To answer the first RQ on effectiveness, OA, confusion matrix, Tau(e), precision, recall, and F1-score were used.

The OA also called 'accuracy', is a commonly used metric of related works as it gives an overall summary of a models performance and is the general agreement over a confusion matrix (see Figure 2.9) calculated as

$$\text{OA} = (TP + TN)/(TP + FP + TN + FN)$$

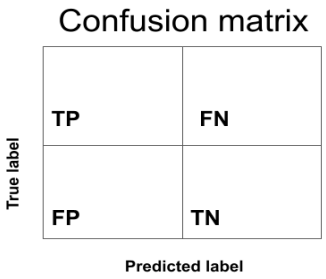


Figure 2.9: Confusion matrix. 'True Positive' (TP) contains the number of correctly classified positive samples and 'True Negative' does the same for the negative class. 'False Positive' (FP) reports samples incorrectly labeled as Positive, and like 'False Negative' (FN).

which gives numbers ranging between 0 to 1 [6]. An OA of 0 indicates that the classifier is not accurate, a value of 0.3 would indicate random classification for this work given the three classes used. Therefore, the model aims to score close to 1 which would indicate high overall accuracy. As a result, this thesis uses OA to be able to compare results to related works, and for summarizing each of the model's performances in a single metric. Despite the OA being a good summary metric, it can be misleading by itself, and hence additional metrics are necessary. Take this thesis issue, in which the main interest lies in accurately classifying seagrass from seaweed, and then from water. If all water samples were to be classified correctly and seagrass and seaweed at random this would result in an OA of 0.66, and hence can be misleading. Therefore, the confusion matrix, Tau(e), precision, recall, and F1-score will be used here to supplement the OA.

The confusion matrix (see Figure 2.9) adds necessary information to the OA about which classes are classified correctly and which ones are confused [113]. Each quadrant reports on the portion of a class that has fallen within each category, which is either True Positive (TP), correctly classified positive samples as the positive class, True Negative (TN), corresponds to the portion of correctly classified samples of the negative class, while False Negative (FN) and False Positive (FP) report on the portion of samples that have been incorrectly predicted as the opposite class. Figure 2.9 illustrates the 2x2 format of a confusion matrix, where classes correctly classified are seen along the diagonal from the top left to the lower right (TP and TN) and miss-classification are in the quadrants around it (FP and FN). Tau(e) is a complementary metric to the OA in that it accounts for chance, while many related works use the very similar metric Kappa, Tau(e) was chosen as it was found superior [114]. To account for chance, it is calculated as

$$Tau(e) = \frac{\text{OA} - p(chance)}{1 - p(chance)}$$

Where 'p' stands for 'probability'. It ranges from -1 to 1, where negative values (close to -1) indicate low agreement, while high values (close to 1) indicate high agreement that a class was drawn by chance or not, respectively. Furthermore, precision and recall are two metrics that aid in the understanding of how the classifier works and will in this thesis aid answering RQ2 (see chapter 4). Precision is calculated as

$$Precision = \text{TP}/(TP + FP)$$

and tells how trustworthy the model's predictions of the positive class are. While the recall is calculated as

$$Recall = \text{TP}/(TP + FN)$$

and signifies how many samples of the positive class the model was able to capture.

Thus, an ideal case outputs both high precision and recall, meaning that the model captures a lot of the samples and classifies them correctly. Furthermore, a score of <0.5 for either precision or recall is deemed 'not good', 0.5-0.8 is 'ok', while 0.8-0.9 is 'good' and >0.9 is 'very good' [115, 116, 117]. Combining precision and recall into a single metric is done to provide a harmonic mean of the two, this metric is called the F1-score [6, 81].

By validating the model's test results through multiple metrics, a throughout understanding of its performance is provided, enabling one to detect the model's shortcomings or identify its strengths. Therefore, this work used OA, complemented by Tau(e), confusion matrices, recall, precision, and F1-scores.

Chapter 3

Methodology

The methodology section aims to give a detailed overview of the process behind the SML, UML, and SSL classifications. From how the satellite images was pre-processed, how Ground Truth (GT) data was collected and overlaid to extract features, datasets created, to how the models were trained, validated, and eventually used to classify. The methodological workflow can be visually summarized in Figure 3.1.

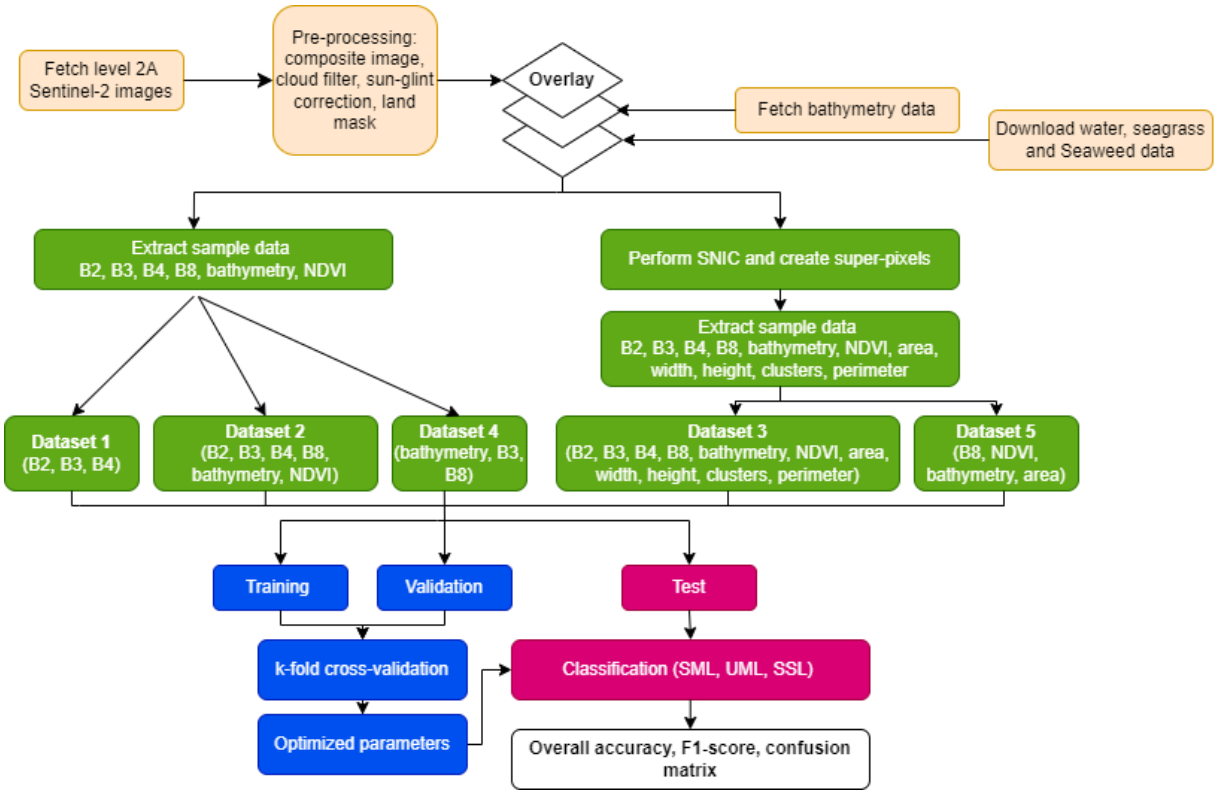


Figure 3.1: Workflow. From top to bottom. Orange represents data import. Green represents feature extraction. Blue represents the training and validation stage for SML and SSL. Purple, the testing and final classification outputting external validation metrics.

3.1 Study Site

From the motivation of this thesis, Casco Bay (green area in Figure 3.2) was chosen to perform the soft test on (training) due to local community monitoring efforts trying to map and restore eelgrass in the area [31]. To test the algorithms scalability to other areas, a neighboring Bay, Saco Bay (pink square in Figure 3.2), was chosen to perform the hard test on (testing) as it had available independent data on eelgrass, seaweed, and water, [118] and assumed similar site characteristics to that of Casco Bay because of its nearby location.

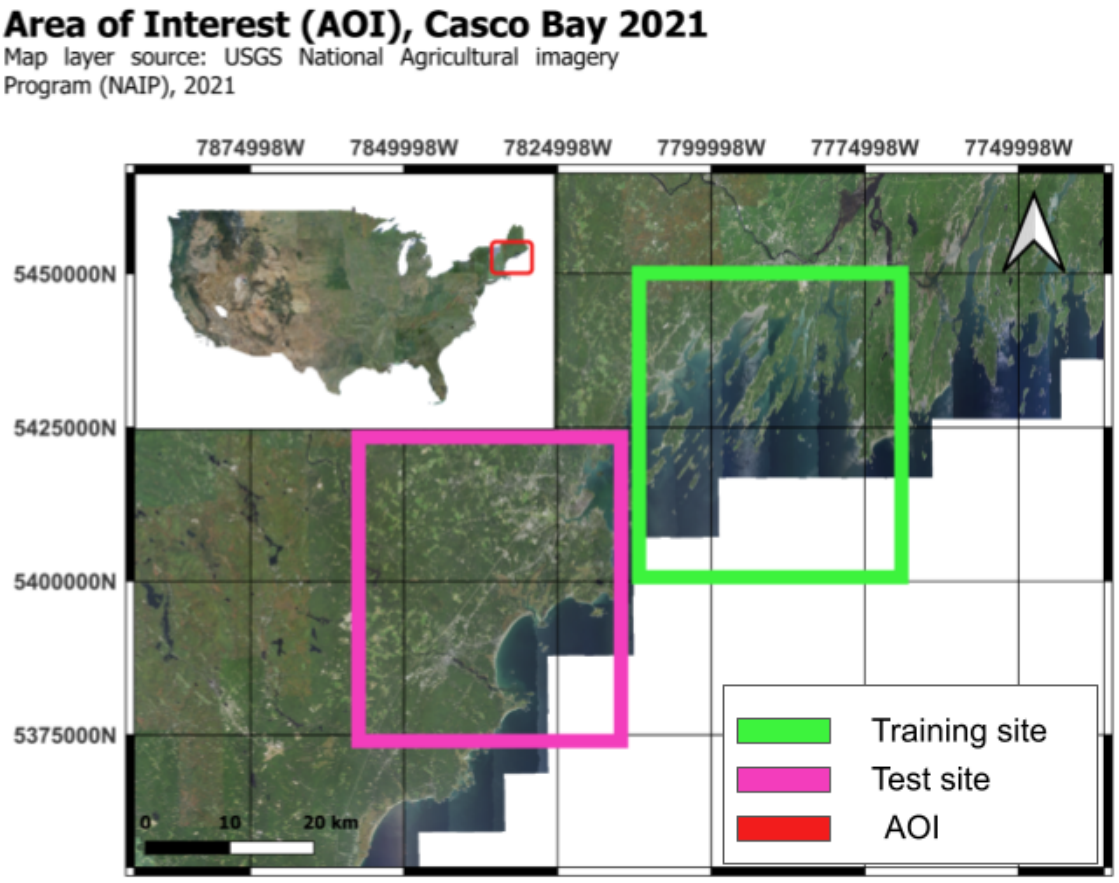


Figure 3.2: Casco Bay, the area of interest (AOI) located in the northeast of the U.S. (upper left corner). The red square in the image of the U.S. indicates the AOI. The purple square indicates the test site (Saco Bay), green square indicates the training site (Casco Bay). Projection: WGS84 Pseudo-Mercator

Casco Bay extends from Cape Elizabeth in the south to Cape Small in the north. The estuary is characterized by a rocky coastline, a large intertidal zone, mudflats, and islands. Tides play a dominant role in the ecosystem, rising and falling between 0-3 meters in a single day [119] and can therefore affect species distribution and how detectable they are.

Furthermore, the combination of soft sediments with 758 scattered, small islands creates a protected, ideal environment for both seaweed and eelgrass[120]. This is emphasized by the fact that the largest eelgrass extent ever recorded in Maine was in Casco Bay [121]. Mapping eelgrass in the area is of importance due to local stressors that can impact their distribution such as industrial activities, run-offs, development, and warming waters [122]. To summarize, Casco Bay is an inlet in the Gulf of Maine, housing both seagrass and seaweed, mainly in the form of eelgrass (*Zostera marina*) and Rockweed (*Aschophyllum nodosum*) [123].

3.2 Pre-processing

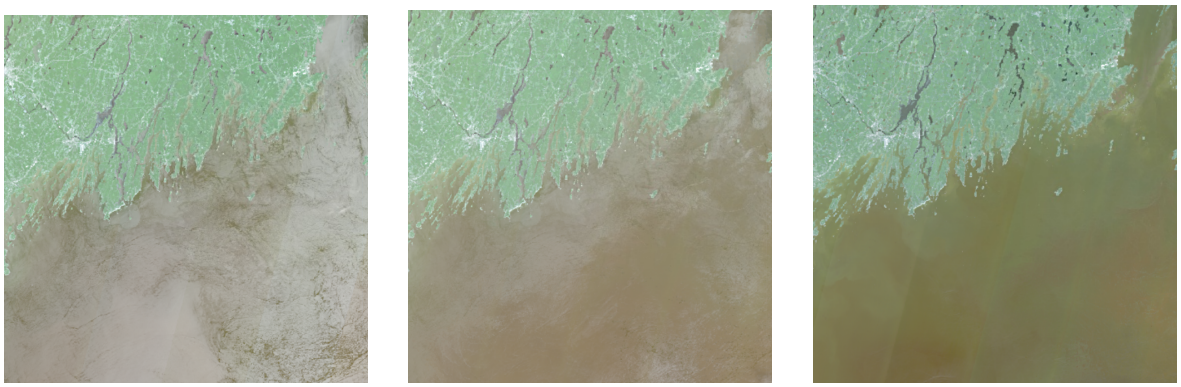
In this thesis, the aim was to assess how well ML models classify seagrass, seaweed, and water using only freely available satellite-derived data. In order to relate features from the image to field data, pre-processing of the satellite data is necessary. Image data from Sentinel-2 level-2A[1], was used here, retrieved from GEE's data catalog and was obtained around the period of maximum vegetation, June-September, as recommended by [56] for both the soft and hard test. The pre-processing consisted of external radiometric correction (minimizing cloud coverage, sun-glint correction) and applying a land mask.

Other typical pre-processing steps in the field of satellite data are atmospheric and water-column correction, converting DN values to spectral units and deep-water masks [3]. Since Sentinel-2 L2A data was used here, the atmospheric correction has already been applied through the Sen2Cor algorithm by ESA [124]. The water-column correction was neither applied (further discussed under chapter 5.3), instead it was assumed that the waters were clear. Converting DN-values to spectral units was not deemed necessary as we are only comparing Sentinel-2 L2A images from before 2022-01-25 [124, 125] against each other [70, 126]. Masking of deep waters was not done due to time constraints, leaving cloud removal, sun-glint correction, and land masking as described below.

3.2.1 Cloud removal

The maximum cloud coverage for images was set to 5%. This resulted in an image collection of four and five images for the training and test site respectively. Figure 3.3(a) shows the image with the least amount of clouds, Figure 3.3(b) show the resulting composite image after taking the median value of them which empirically reduced the amount of noise by the clouds. Another option was to use the minimum DN-values which produced the composite seen in Figure 3.3(c). Empirically it was found that the minimum values gave the least amount of noise, thus this technique was chosen to create the composite. While

[1]Sentinel-2, level-2A data can be found here: *https : //developers.google.com/earth − engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED* [Last accessed: 26.04.23]



(a) Single Sentinel-2 image with the least cloud coverage.

(b) Composite of Sentinel-2 images by median values.

(c) Composite of Sentinel-2 images by minimum values.

Figure 3.3: Cloud removal of Sentinel-2 images by reducing multiple images into a composite by median or minimum values.

a composite image decreases cloud coverage, it is more difficult to control confounding variables such as algae bloom, temporal changes, tides, wind speed, etc. Algae bloom could be detected as seagrass or seaweed by the algorithm and lead to over- or underestimate of the species. Temporal change in vegetation such as tides, wind speed, or season affect the amount of vegetation and how well it is detected by a sensor, also leading to additional misrepresentation of the amount of seagrass and seaweed. The state of the confounding factors can only be determined by the acquisition time the satellite image was taken. Given that a composite image is the result of a function that reduces multiple images into a single one by comparing all cells (pixels) of same location in an image and takes the minimum value out them to make the composite, this makes it harder to determine the state of external factors, such as the above-mentioned, which is good to take into account when assessing the algorithms resulting accuracies for the sake of not over- and underestimate the amount of seagrass and seaweed.

3.2.2 Sun-glint Correction

Another confounding factor that, unlike the above-mentioned ones, could be accounted for was sun-glint. When waters are not still, the sun's radiance on wavy and curvy waters can reflect back bright pixels, distorting the water-leaving signal [127, 128]. To correct for this, glint removal, also called 'deglinting' was applied to the composite image. Sun-glint correction is based on three assumptions [129]. First, the brightness (high reflections) in NIR is only caused by sun-glint and an 'ambient' NIR component. This leads to the second assumption, that the darkest and brightest pixel in an area of water with the same elevation, should have equal values. Thirdly, there is a linear relationship between the NIR and visible light bands. Considering these assumptions, the brightness sun-glint accounts

for can be calculated and corrected. This was done by taking a sample of pixel-values from the image where sun-glint was identified, preferably in a homogeneous area, thus a place of deep waters were chosen in this thesis. Furthermore, linear regression was performed in which the product representing the slope was subtracted from the minimum pixel-value of the image. This difference represented the brightness produced by sun-glint. The derived difference-value was subtracted from each pixel-value in all bands (NIR, R, G, and B) essentially removing the effect of sun-glint [130].

3.2.3 Land masking

Land masking was applied to remove irrelevant areas and avoid the misclassification of seagrass and seaweed. The threshold with the DN-values of the NIR band was used to discern land from water as it is easily reflected by land and absorbed by water surfaces. In this work, the threshold value was set to 500 based on similar work [131], the knowledge that high values in NIR represent vegetation or land characteristics while values from zero or below represents water [132], in addition to experimenting with different values.

3.3 Data

Labeled data on the three classes: seagrass, seaweed, and water are needed to train the SML and SSL algorithms. In benthic habitat classification, labeled data is referred to as in situ data or GT. These terms will be used interchangeably. Enough training data for SML and SSL depends on the dataset, but some general recommendations exist to increase the validity. Both [60] and [83] suggested 10-30 times the number of training samples per feature, per class. In this thesis, the datasets used contain either three, four, six, or eleven features. Thus, the minimum amount of training samples per dataset can be calculated as

$$Datasetsize = nr.features * (nr.classes) * (min[trainingsample])$$

The smallest dataset used in this thesis consisted of three features, then the equation yields: 3*3*10* = 90. Applying the same equation to the biggest training data shows that it requires 330 training samples. Hence, to obtain valid results, these recommendations provided a simple framework to follow for what amount of training data is needed.

3.3.1 Data Collection Method

GT on eelgrass was sourced from the MaineDEP [133], GT on seaweed was sourced from the Maine Geological Survey (MGS), while GT for the water was created with GEE's interactive tools.

GT of eelgrass data for training was captured by flight on June 16th and 17th, 2018 under ideal conditions [133] in Casco Bay. Then, experts labeled and verified the data by foot, boat, and photography. GT data for testing was also obtained through MaineDEP's collection of in situ data by airplane on the 21st of June 2019 [118] in Saco Bay. For this thesis, both datasets were downloaded locally as shapefiles. The training data consisted of 956 polygons, while the test data of 295.

Seaweed polygons, representing GT data, were downloaded from MGS [134] for both the training and testing site. The geology data were from the year 1976, with the latest update on the 30th of November 2018. While the dataset contained general data on geological substrates of the Maine coast, retrieving data only on seaweed required manual filtering. Data were filtered by the units "Fs" which gave 410 polygons representing seaweed community, and "F2" which yielded 72 polygons, representing seaweed-covered coarse flats. Originally, only *Ascophyllum nodosum* was the seaweed species of interest for this thesis. However, the category was broadened to all brown, red, and green, species of seaweed as this was the only GT data fulfilling the criteria on the recommended training sample size as mentioned under chapter 3.3. Thus, a total of 482 polygons, representing the seaweed was found for the entire coast of Maine and was further separated into individual sets to derive independent features for the soft and hard test, resulting in approximately 30 polygons for each place.

To obtain GT polygons representing water, tools within GEE were used for both the training and testing site. Polygons of water were stored locally as a shapefile, just like that of seaweed and seagrass. The validity of the water data is based on two assumptions. Firstly, water polygons were made where the literature suggests eelgrass or the seaweed are very unlikely to inhabit, such as deep waters [20]. Secondly, the water polygons created in shallow areas did not overlap the polygons of eelgrass or seaweed. This assumption was deemed weak as there could be other substrate types that are picked up instead. A better solution would be to create a category called "others", however, keeping with the initial assumption, while being aware of flaws, a total of five water polygons for each site were created. The polygon geometry obtained for the three classes provided the geographical location used to capture their features. This is done by overlaying different geographical layers, each layer containing a single feature.

3.3.2 Features Selection

Features are distinct characteristics of a class, enabling an algorithm to distinguish one from the other. Three of the datasets (1,2, and 4) had PB-derived features, while the remaining (3 and 5) had GEOBIA. From the composite image, available features were the Sentinel-2 L2A bands 1 - 13 representing different central wavelengths of the electromagnetic specter. Only the four bands, B2, B3, B4, and B8 were utilized due to their 10m resolution. B2-B4

represents the visual light, blue, green, and red, respectively, while B8, represents NIR. Figure 3.4 illustrate how these bands provide data and make up the composite image. In Geographical Information System (GIS), such as satellite images, data are layered on top of each other to connect spatial data and find relationships between them. In Figure 3.4 there are four sources of data, Band B2, B3, B4, and B8 and respective layers. The composite image collected the data from each of these layers in it's grid of cells (or pixles). Each cell contained a number of values, in the case of Figure 3.4 a single cell would contain four values, one for each of the four bands that would eventually represent the feature for seagrass, seaweed, and water.

Moreover, the composite image was further overlaid with bathymetry data (also retrieved from GEE's data catalog [2]), acquiring a fifth feature, and lastly with a layer representing the Normalized Difference Vegetation Index (NDVI). NDVI is derived from red and NIR, due to how the bands are absorbed and reflected by vegetation [135]. The cellular structure of a leaf strongly reflects NIR, while its chlorophyll (the green pigment in plants) strongly absorbs visible light such as the color red. Therefore, for vegetation who contains a lot of chlorophyll, the color red, will be strongly absorbed and NIR will be highly reflected, thus the NDVI will be high [136]. The index range between -1 and 1 with high values of ~ 1 indicating vegetation of high chlorophyll, 0-0.33 indicate vegetation of less chlorophyll content, while low values (around -1) signal objects on none chlorophyll, such as water. Thus, it is possible to indicate the presence or absence of vegetation.

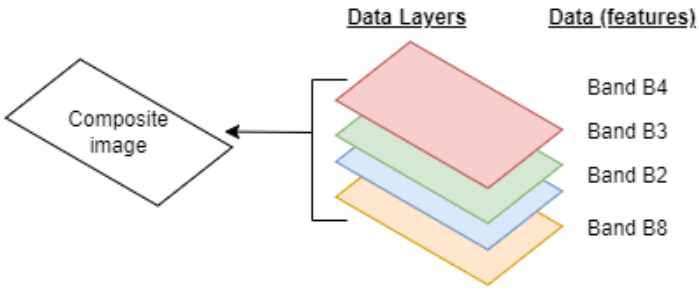


Figure 3.4: Data layers. Illustration of features data in Geographical Information Systems (GIS).

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red}$$

To obtain the six features (B2, B3 and B4, B8, bathymetry, and NDVI) polygons for

[2]Bathymetry data can be found here: *https : //developers.google.com/earth − engine/datasets/catalog/NOAA_NGDC_ETOPO*1 [Last accessed: 26.04.23]

the classes (water, seaweed, and seagrass) are overlayed as a final layer, providing the geometries ready for PB-derived sampling.

Within GEE, a maximum of 5000 sample points per geometry can be collected. In this work, the collection of polygons representing seagrass becomes one geometry, similar to seaweed and water, making a total of three geometries. Figure 3.5 shows how samples were taken given seaweed and seagrass geometries. Seagrass geometries are given as green polygons, while seaweed as brown. Points (green for seagrass and brown for seaweed) were randomly distributed within the respective polygons extracting data (B2, B3, B4, etc.) from the underlying pixels containing values representing features of the respective classes. Consequently, a total of 15 000 sample points were collected for both the soft and hard test.

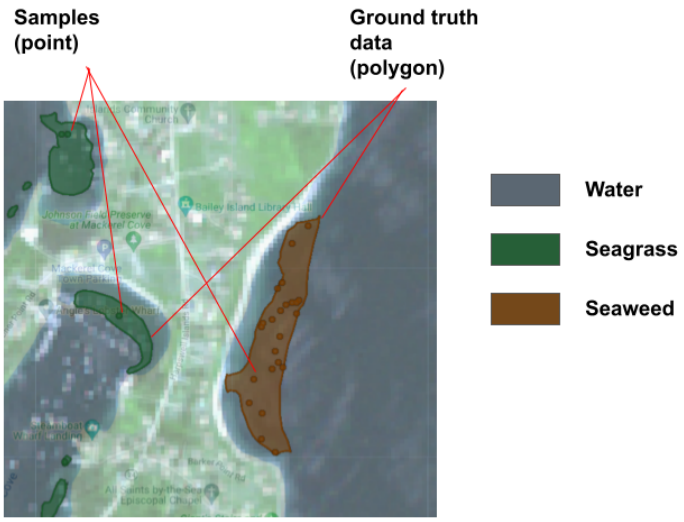
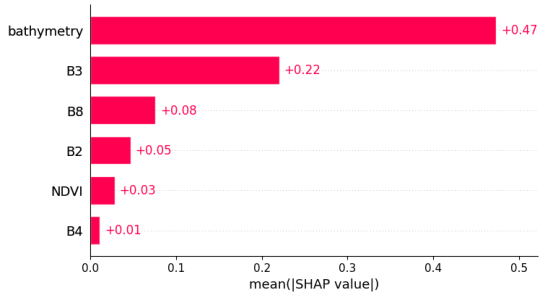


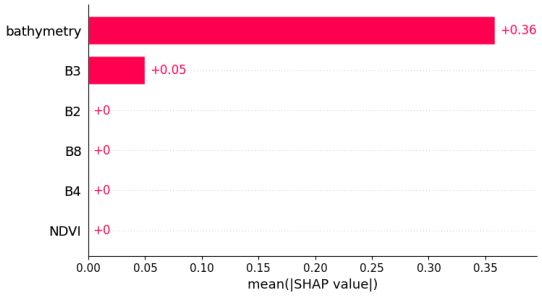
Figure 3.5: Illustration of how the sampling function in GEE works. The polygons represent the classes seagrass (green) and seaweed (brown), while the dots represent the samples containing feature values.

After all the samples from the PB method were extracted, datasets 1, 2, and 4 can be defined. Dataset 1 is the baseline dataset, comprising B2-B3. Dataset 2 uses all six features from the PB method: B2-B4, B8, NDVI, and bathymetry. Dataset 4 comprises the three features contributing the most to ML's prediction found using the SHAP algorithm on the SML and SSL models[137]. SHAP is a feature selection algorithm applied to a model that can detect the most influential features. Figure 3.6(a) is given as an example with dataset 2 for the SML and Figure 3.6(b) for the SSL model. In the case of dataset 2, the most influential features were found to be bathymetry and B3 for both the SML and SSL models.

Moving on, the GEOBIA is applied with the SNIC algorithm, and the super-pixels derive an additional five features: area, perimeter, clusters, height, and width, in addition to the mean value of B2, B3, B4, B8, bathymetry, and NDVI for each super-pixel. Dataset 3 comprises all the 11 features above, while dataset 5, like dataset 4, consists of the three most influential features of the GEOBIA method, also found with the SHAP algorithm. Figure 3.7(a) shows the results of applying SHAP to SML and Figure 3.7(b) to the SSL with GEOBIA-derived features. For the SML bathymetry, B3, NDVI, and area were found most impact-full, while for the SSL it was bathymetry, width, and clusters. The

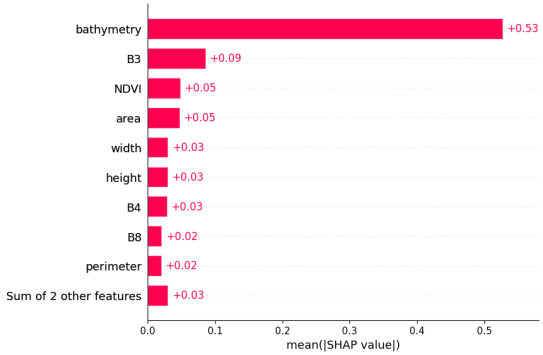


(a) Shap-values for SML model and dataset 2

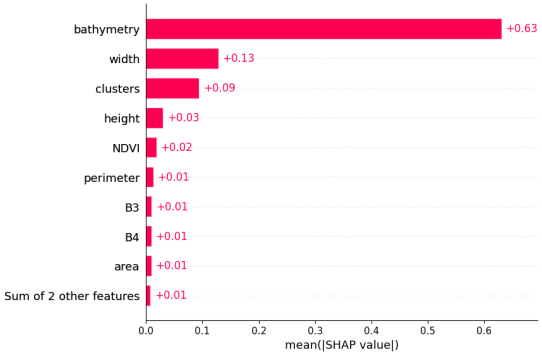


(b) Shap-values for SSL model and dataset 2

Figure 3.6: Feature contribution from the PB method



(a) Shap-values for SML model and dataset 3



(b) Shap-values for SSL model and dataset 3

Figure 3.7: Feature contribution from the GEOBIA method

choice of features was based on similar studies, bands of the highest spatial resolution of Sentinel-2, and other literature recommendations [4]. Thus a total of 5 datasets were obtained. Datasets 1, 2, and 4 with PB-derived features, and datasets 3, and 5 with features from the GEOBIA method. Different combinations of feature datasets were used to detect if some impacted the accuracy more than others. Moreover, the features from the PB were held separate from those derived with the GEOBIA method, as to see if a difference in the two image classification methods could be identified.

3.3.3 Data Cleaning and Inspection

To obtain representative and quality data, cleaning of the extracted samples was conducted. Cleaning involved removing duplicates, shuffling, and balancing the data, which resulted in a reduction from an original 15 000 to 720 samples for training and validating the SML and SSL. Duplicates need to be removed to reduce bias [138] and were identified through pair plots, and seeing suspiciously high training and validation accuracies. In the initial pair plots, samples were clustered densely, thus indicating a lot of overlapping samples or duplicates. Then, very high training and validation accuracies gave the second hint

that something was wrong. By removing duplicates from the original 15 000 data samples, 2817 samples remained.

Before balancing the data, it was shuffled. Shuffling the data decreased bias, given that data might have been stored in a certain order (e.g.: by class). Then, the data was balanced regarding the water, seaweed, and seagrass classes by finding the number of samples of the minority class. The minority class contained $\sim$ 240 samples, thus the 2817 samples were further reduced to 240x3 = 720 sample points. This led to each dataset being duplicate-free and balanced, containing 720 samples, 240 of them representing each class.

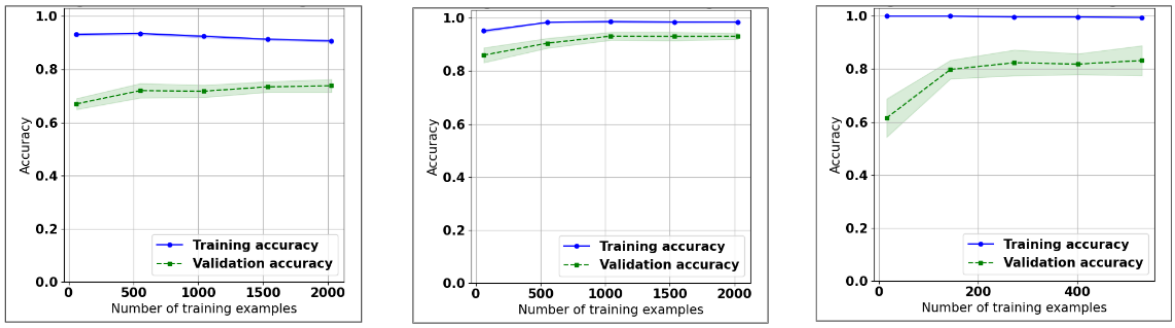
On a last note, through descriptive statistics (median, min, max, etc.) and box plots, negative digital numbers for the bands were also identified. Given that the Sentinel 2 images are coded in UIN16 means that the DN-values should range between 0-65535. Higher values would represent saturation, while negative values seem to be some sort of fill-values [139]. Ideally, these would have been removed [140] but were not done due to time constraints.

3.4 Classification

The aim of the classification chapter is to give a good overview of how the three algorithms were executed. After data was cleaned, balanced, and datasets derived, it is eventually ready to be used for classifying seagrass, seaweed, and water. The classification was done with SML employing the RF algorithm, SSL with RF and pseudo-labels, and UML with K-means. The training and validation of SML and SSL will be described, as well as how the K-means were assembled for the UML.

3.4.1 Supervised Classification of Seagrass and Seaweed

The supervised classification consists of three phases: training, validating, and testing. Training and validation were done using a learning curve, validation curve, hyperparameter tuning via grid search, and CV. Each dataset was split into 80% training and 20% for validation. A learning curve can be used to spot over- and underfitting [81]. For example, the learning curves of datasets 1, 2, and 3 from this work are given in Figure 3.8. From these plots, overfitting is identified in Figure 3.8(a) and3.8(c), while for Figure.3.8(b), overfitting becomes less prevalent when the number of samples is above 1000. Thus, by identifying over- or underfitting, the models can be optimized, aiming at a sufficient bias-variance trade-off.



(a) Dataset 1. (b) Dataset 2. (c) Dataset 3.

Figure 3.8: Illustration of learning curves for datasets 1, 2, and 3 for SML with RF. The x-axis represents the number of samples in a training set, while the y-axis represents accuracy.

In general, the SML and SSL models used here employed grid search, a hyperparameter tuning technique, given a search space, to find the optimal combination of hyperparameters for the RF. Hyperparameters are, unlike parameters, externally adjustable variables. While variables such as clusters and weights can be learned by the model itself, the number of trees in an RF or how compact clusters should be in +acSNIC can be controlled and adjusted to optimize the model. Therefore, hyperparameter tuning was performed. The default hyperparameters of sklearn's RF algorithm provided a good start for further tuning

[141]. Similar studies were also looked at for further guidance (i.e. [142, 143, 48]), but experimenting with different values was the main method used. Table 3.1 shows the types of hyperparameters inputs into the grid-search, the range of values for each to be tested, and the outputted value of the grid search, representing the combination of hyperparameters achieving the highest accuracy for SML.

Table 3.1: Hyperparameter tuning by grid search for RF.

Hyperparameters	**Input**	**Output**
trees	[10-40]	30
features	['sqrt', 'log2']	sqrt
max. tree depth	[2, 4, 10]	10
min. samples in a split	[2, 5]	2
min. samples in a leaf	[1, 2]	1
bootstrap	[True, False]	False

To find the most optimal combination of hyperparameters for RF, sklearn's grid-search algorithm was used. Given a set of hyperparameters and a range of values for each, the grid search goes through all possible combinations of them brute-force and outputs the value combination generating the highest accuracy. A drawback of this approach is that it takes time [81]. Thus, reducing the search space reduces training time.

A validation curve was used to reduce the search space, hence reducing training time. A validation curve can visualize how a hyperparameter affects the accuracy. Figure 3.9 illustrates this using the hyperparameter 'number of trees', where its effect on training and validation accuracy can be seen. The resulting validation plot was used to identify the intersection between the maximum number of trees needed while maintaining high accuracy. In the case of Figure 3.9, it shows that there is no need to explore more than 30 trees as the curve decrease and flattens after this. In general, the curve is very flat from the start (at 10 trees), thus if it is possible to trade off a small amount of accuracy for time efficiency, the number of trees could further be reduced. In this thesis, the number was not, as efficiency was not considered equally important as accuracy.

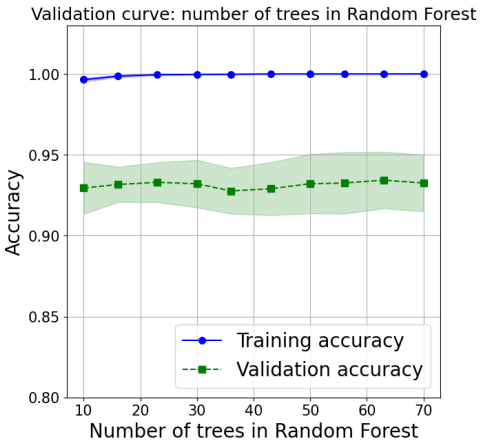


Figure 3.9: Validation curve. Example using dataset 2. The x-axis represents the increasing number of trees in RF, while the y-axis the corresponding accuracies.

Given the use of sklearn's grid-search algorithm, k-fold CV was internally performed. Based on the amount of data in this thesis, it was more important to have a bigger dataset for training and a smaller one for validation, thus CV with 10-folds was applied.

Furthermore, choosing the number of folds was also based on empirical evidence showing that an optimal number of splits range between five and ten [144, 145].

An optimized SML was obtained from tuning hyperparameters with the learning curve, validation curve, grid-search, and performing CV. The optimized model was then applied to the hold-out test data for Casco Bay (the soft test) and to the new unseen data from Saco Bay (the hard test). Using two test sites gives a more valid representation of how generalizable the model is to other areas and SML's results for both the soft and hard test are reported in chapter 4 and further discussed in chapter 5.

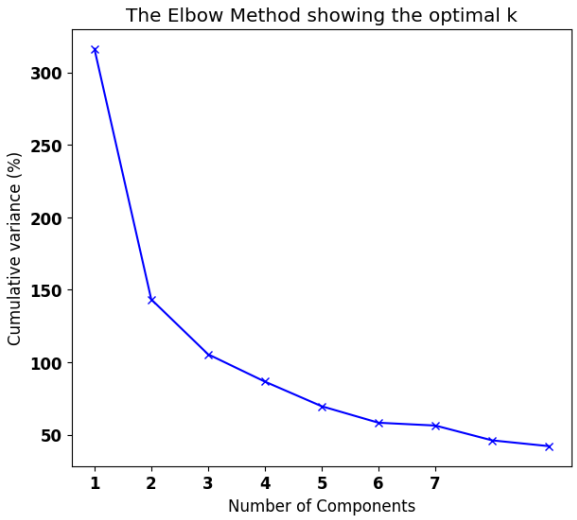
3.4.2 Unsupervised Classification of Seagrass and Seaweed

Unlike the SML, no training or validation was performed for the UML. Instead, the steps in classification with UML consisted of feature reduction with PCA simply followed by performing the soft test with the same balanced, cleaned data of 720 samples and then the hard test.

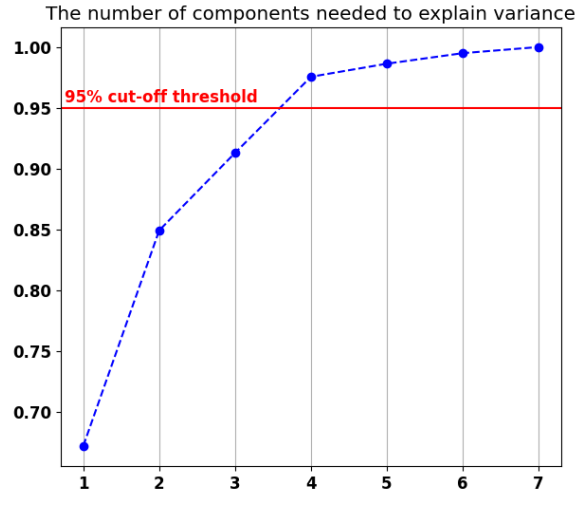
Before feeding the datasets to the UML, standardization and PCA were performed after recommendations by [81]. To identify how many components a dataset should be reduced to via the PCA, a graph for explained variance with a threshold of 95% was used to find the minimum amount of components needed to explain 95% of the variance. Given the criteria of keeping 95% of explained variance, three-four components should be used. Although the threshold was closer to four than three, the features were projected onto three principal components [146] to keep the ability for visualization.

For the K-means, two hyperparameters are available, the number of clusters and the number of iterations. Despite knowing the number of clusters (three) the elbow method was checked for alignment. Figure 3.10(a) showed a soft curve instead of an 'elbow' which is characterized by a sharp break or angle in the curve. Thus, from the plot it was concluded that two or three clusters could be reasonable, however, the plot does not clearly align with what is known. Given that the k is known, k in K-means is set to three. Furthermore, iterations are set. They tell how many times the algorithms shall recalculate the distance between every point and rearrange them into clusters. The number of iterations used was the default value of sklearn's K-means. To summarize, standardization was performed to combat high bias, plots were used to identify a good number of components for PCA used to reduce variance before the data was fed to the K-means algorithm. For the k-means k was known and set to three before all five datasets were fed to it for the soft and hard tests.

A note on the choice of the UML algorithm: From the literature, the first choice was to use the K-means clustering algorithm because of its simplicity, high performance, and low computational cost. However, after visually inspecting the data, K-means were not found suitable as the assumption of spherical clusters did not hold true for the data at hand.



(a) The Elbow method. X-axis marks the number of components, while the y-axis tracks the cumulative variance.



(b) Explained variance with a threshold of 95%

Figure 3.10: Elbow method and explained variance for dataset 2.

Thus, a Density-Based Spatial Clustering of Application with Noise (DBSCAN) algorithm was also tested, but neither of these was found suitable and performed even worse than the K-means [81], hence only the results of K-means were documented for comparison with SML and SSL.

3.4.3 Semi-supervised Classification of Seagrass and Seaweed

The SSL model used here was RF with pseudo labels and consisted of three phases, supervised, unsupervised, and semi-supervised.

The first supervised phase behaves like a reversed version of the SML with RF. Given a dataset, the data is partitioned into a 70:20:10 ratio, but instead of using the 70% for training and validation, the supervised stage of the SSL used the 20%. Furthermore, in the supervised stage the same grid-search and CV with 10-folds is applied to the SSL in a similar fashion as for the SML. After obtaining sufficient validation accuracy (see chapter 4 Figure 4.7(a)) from the small amount of data, an optimized RF model is produced.

In the second phase, a big amount of unlabeled data was predicted. From the input data, 70% was held-out and tagged as "unlabeled data". Using the optimized model from phase one, the labels of this big dataset were predicted. Next, the predicted labels from the unsupervised phase were merged with the true labels from the supervised, making the so-called 'pseudo-labels' and respective 'pseudo-data'. Out of the total amount of data, the pseudo data made up 90% of it (70+20) given the aim of producing more labeled data supposing that more data facilitates improved classification accuracy for the model. This 90% is thus used to train a new RF model.

Table 3.2: Overall accuracies of first and second phase in SSL with the optimized parameters from grid search. For each dataset, the first phase used 10% of the total data, while the second phase used 90%.

Dataset	**First phase**	**Second phase**
1	59.83%	48.65%
2	79.49%	81.08%
3	75.59%	68.92%
4	82.20%	82.43%
5	72.20%	68.92%

A comparison of the accuracy results from the first stage with the small dataset and the second stage with the big dataset is given in Table 3.2. From Figure 3.2 there is in general decrease in accuracy between the first supervised phase and the second semi-supervised. The exception lies with dataset 2 and 4. Because the result from the first phase and the second phase were reasonable, assuming self-learning with SSL for RF with pseudo labels were deemed to be OK and hence moved forward with to phase three.

Stage three trained a new RF model with the pseudo-data and pseudo-labels. Given that some of the labels are true and others are semi-true (the predicted labels), this makes a semi-supervised process. Here again, grid-search and CV with k-folds (as there is more data now) are used to find optimal hyperparameters for the SSL. After optimized hyperparameters are found, the remaining 10% of the data is used for the soft test.

Initially, a combination of K-means clustering and RF was supposed to make the SSL model, being a prominent solution in both [147] and [148]. However, after the poor results from the K-means alone, indicated that the underlying clustering assumption of SSL did not apply to the data on seagrass and seaweed, RF was used instead as it does not rely on clustering and knowing that it would be robust to noise introduced by the pseudo-labels.

Chapter 4

Experimental Results

Monitoring aquatic vegetation across the globe is deemed important to account for their ecosystem services, threats, and population trends [17].

Relevant to this study was assessing the combination of freely available satellite data coupled with ML, to aid community monitoring and explore the latest freely available technologies and tools within RS. An experiment of classifying seagrass from seaweed at Casco Bay and Saco Bay of Maine was used to compare three different ML models and found that both the SML and SSL comply with related work. From this, SSL with RF and pseudo labels, poses the potential to map aquatic vegetation with less labeled data than SML, solving a common problem in the field.

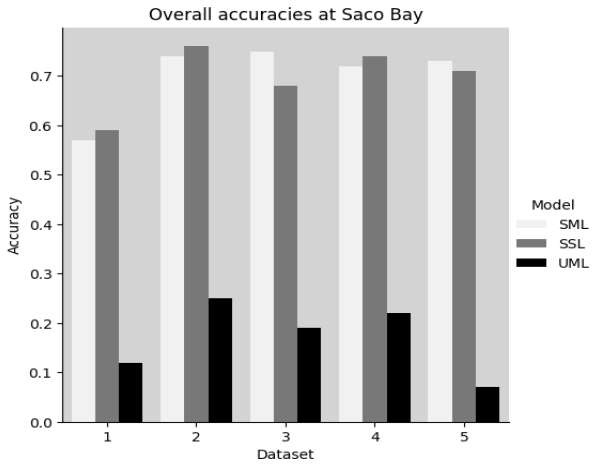


Figure 4.1: The overall accuracies for all models at Saco Bay

Figure 4.1 show the models' accuracy results from the hard test. From the bar plot, SSL achieves the highest accuracy but is closely followed by SML. The remaining chapter will provide results of the individual classifiers SML, SSL, and UML. The models' performances were evaluated using OA, F1-score, confusion matrix, precision, recall, and Tau(e) summarized in tables, box plots, line plots, and confusion matrices for all datasets at both sites (training and testing).

4.1 Supervised Classification Results

The results of the supervised classification are presented here. Evaluation metrics of SML with the soft test are summarized in Table 4.1. The accuracy ranged between 67-91% and the average F1-score from 64-89% which is in line with results obtained in related work. Their corresponding Tau(e) range between moderate to a nearly perfect agreement [149], indicating that the results are not likely to occur due to chance alone. Both precision and recall are moderate to high, hinting that the consequences of SML underestimating or overestimating a class are low. Only the seaweed class in dataset 1 had a distinctive low recall value, hence implying that the model is underestimating seaweed in this case.

Table 4.1: Results summarized from the SML model with Random Forest (RF) at Casco Bay (soft test). P = Precision. R = Recall. F1-score = individual class F1-score. Avg. F1-score = the Average F1-score for the given dataset. OA = Overall Accuracy for the given dataset. T = Tau(e) for the given dataset.

Class	P	R	F1-score	Avg. F1-score	OA	T
Dataset 1						
Water	0.74	0.88	0.8			
Seaweed	0.61	0.36	0.45	0.64	0.67	0.50
Seagrass	0.60	0.77	0.68			
Dataset 2						
water	0.97	0.98	0.98			
seaweed	0.87	0.77	0.82	0.86	0.87	0.80
seagrass	0.72	0.83	0.77			
Dataset 3						
water	1.00	0.96	0.98			
seaweed	0.74	0.82	0.78	0.85	0.86	0.79
seagrass	0.81	0.76	0.79			
Dataset 4						
water	0.98	0.98	0.98			
seaweed	0.90	0.83	0.86	0.89	0.91	0.86
seagrass	0.79	0.89	0.84			
Dataset 5						
water	0.98	0.98	0.98			
seaweed	0.70	0.67	0.68	0.80	0.81	0.72
seagrass	0.71	0.74	0.72			

Under- and overestimation, in addition to confusion between the classes, are further interpreted through confusion matrices. Figure 4.2 (a)-(e) shows the confusion matrices for the SML model of the five datasets. From the confusion matrices, water samples were evidently more classified correctly and less confused compared to both seagrass and seaweed, regardless of the feature combination. Dataset 1 is used as a reference dataset as it only contains features representing visible light (blue, green, and red), hence compared to dataset 1, the confusion between seagrass and seaweed decreases in 2 and further in 4, while for datasets 3 and 5 stayed stable or increased. Consequently, for the soft test, adding additional features to visible light impact the individual accuracies of seagrass and seaweed at Casco Bay. Furthermore, water was not found to be much over- or underestimated (dataset 1 being the exception), and overall, the matrices showed that seaweed was more classified wrongly as seagrass than vice versa.

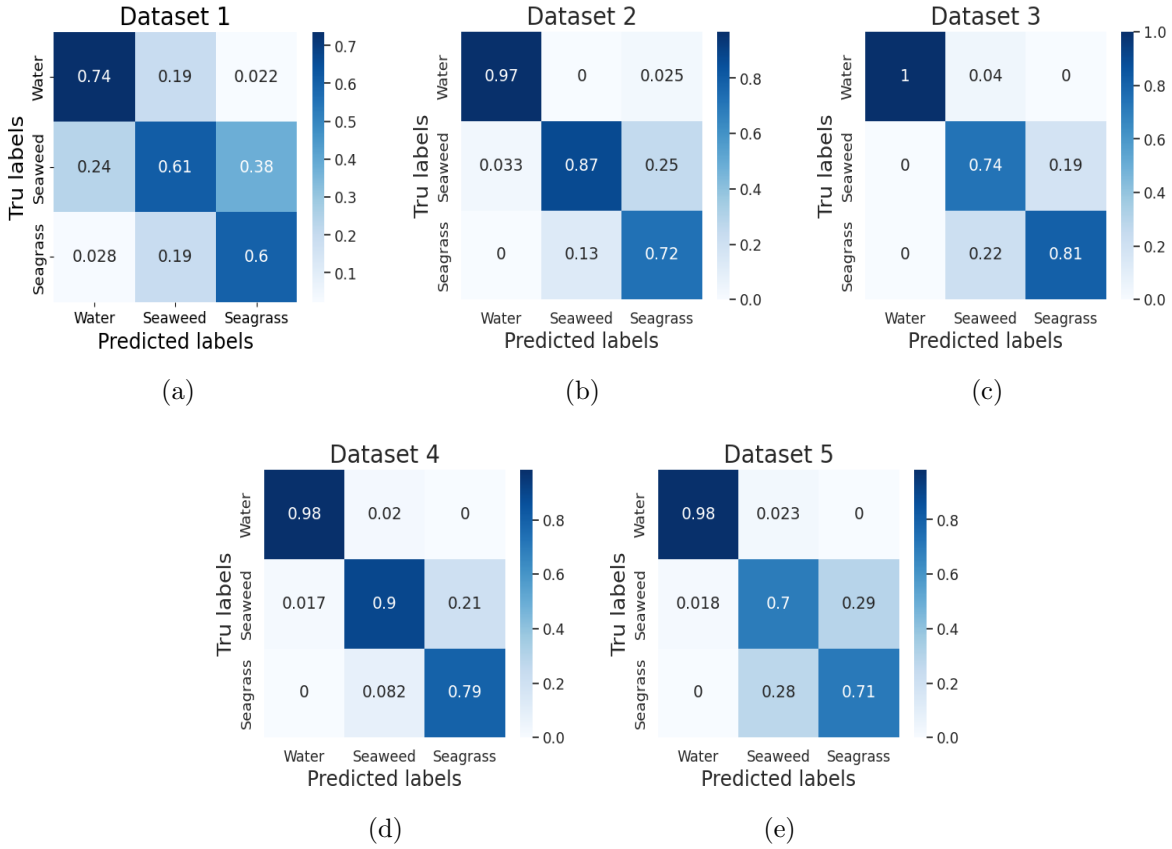


Figure 4.2: Confusion matrices from the soft test of SML. The individual accuracy scores for water, seaweed, and seagrass for datasets 1, 2, 3, 4, and 5 are given, obtained with data from Casco Bay.

The equivalent summarizing table over performance metrics for SML on the hard test can be seen in Table 4.2. Here, the accuracy ranges from 57-75% and the average F1-score from 55-74%, hence a decrease from the soft test to the hard one. Relating Tau(e) values are from 0.35, suggesting that the result could have emerged by chance, to 0.62, denoting good agreement. Again, compared to the soft test, precision and recall values are lower. Low precision can be found for seagrass and seaweed in dataset 1, otherwise, precision for seagrass and seaweed stays around in the range of 0.6-0.68 which is ok, while only high precision is found for the water class. A similar pattern is seen for recall values, i.e., high for water, while moderate for both seagrass and seaweed in the remaining datasets. Only very low for seaweed in dataset 1, the same finding as for the soft test. A last anomaly is the F1-score for seagrass in dataset 3 which was found rather high.

Table 4.2: Results summarized from the SML model with Random Forest (RF) at Saco Bay (hard test). P = Precision. R = Recall. F1-score = individual class F1-score. Avg. F1-score = the Average F1-score for the given dataset. OA = Overall Accuracy for the given dataset. T = Tau(e) for the given dataset.

Class	P	R	F1-score	Avg. F1-score	OA	T
Dataset 1						
Water	0.79	0.79	0.79			
Seaweed	0.37	0.26	0.31	0.55	0.57	0.348
Seagrass	0.50	0.64	0.56			
Dataset 2						
water	0.93	1.0	0.96			
seaweed	0.64	0.54	0.59	0.74	0.74	0.614
seagrass	0.64	0.68	0.66			
Dataset 3						
water	0.97	1.00	0.99			
seaweed	0.68	0.47	0.56	0.74	0.75	0.624
seagrass	0.61	0.78	0.88			
Dataset 4						
water	0.94	1.00	0.97			
seaweed	0.59	0.50	0.54	0.71	0.72	0.576
seagrass	0.60	0.63	0.63			
Dataset 5						
water	0.96	1.00	0.98			
seaweed	0.57	0.71	0.64	0.72	0.73	0.592
seagrass	0.65	0.47	0.55			

Figure 4.3 (a)-(e) (see next page) shows the confusion matrices of the SML model's hard test. Similar to the soft test, water is classified correctly, while seagrass and seaweed are equally confused with each other on all datasets.

The box-plot and line-plot in Figure 4.4 (see next page) summarizes the results of the supervised classifier. From the box plot, the highest training accuracies were acquired

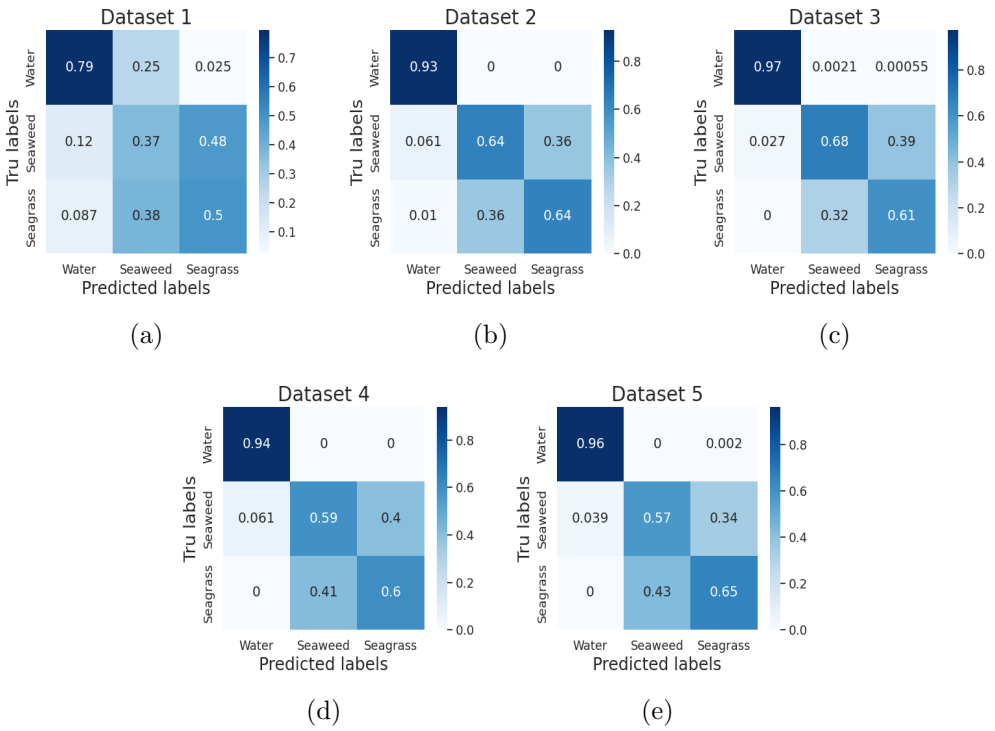
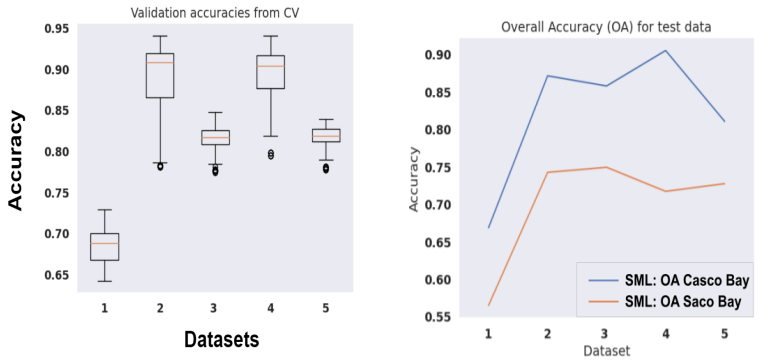


Figure 4.3: Confusion matrices from the hard test of SML. The individual accuracy scores for water, seaweed, and seagrass for datasets 1, 2, 3, 4, and 5 are given, obtained with data from Saco Bay.

with datasets 2 and 4, however, the estimates are rather uncertain given the wide boxes compared to datasets 3 and 5 with narrow ones. The line plot displays a similar trend to the box plot for the validation and testing accuracy scores. In general, the accuracies from the soft test dropped with the hard test.



(a) Box plot for datasets 1, 2, 3, 4, 5. The data represents validation accuracies obtained during CV with 10 folds.

(b) Line plots of OA from Casco Bay (blue line) and at Saco Bay (orange line).

Figure 4.4: Validation and test-accuracies for the SML model. Illustration of a box plot (a) and a line plot (b) of Overall Accuracies (OA) achieved with different datasets. The y-axis is shared by the plots.

4.2 Semi-supervised Classification Results

An overview of the soft test results is shown in Table 4.3 for the semi-supervised classifier. Accuracy range between 49-82%, average F1-score from 47-81%, and corresponding Tau(e) values between 0.23-0.74, meaning that the model's results, for example in dataset 1 bear little support, to good support for datasets 2 and 4. In general for precision and recall, precision is found low for seaweed, while recall is high, meaning that seaweed is overestimated. Combining the high recall and low precision for seaweed, the high recall and precision for water with the low recall for seagrass, indicate that seagrass is confused for seaweed. Thus, seaweed is overestimated while seagrass is underestimated in most of the datasets.

Table 4.3: Results summarized from the SSL model with Random Forest (RF) and pseudo labels at Casco Bay (soft test). P = Precision. R = Recall. F1-score = individual class F1-score. Avg. F1-score = the Average F1-score for the given dataset. OA = Overall Accuracy for the given dataset. T = Tau(e) for the given dataset.

Class	P	R	F1-score	Avg. F1-score	OA	T
Dataset 1						
Water	0.65	0.77	0.70			
Seaweed	0.35	0.19	0.25	0.47	0.49	0.23
Seagrass	0.38	0.59	0.47			
Dataset 2						
water	1.00	0.92	0.96			
seaweed	0.79	0.81	0.78	0.80	0.81	0.72
seagrass	0.65	0.65	0.65			
Dataset 3						
water	1.00	0.90	0.95			
seaweed	0.46	0.90	0.61	0.64	0.69	0.53
seagrass	0.75	0.25	0.38			
Dataset 4						
water	1.00	0.96	0.98			
seaweed	0.78	0.81	0.79	0.81	0.82	0.74
seagrass	0.65	0.65	0.65			
Dataset 5						
water	1.00	0.93	0.97			
seaweed	0.46	0.95	0.62	0.62	0.69	0.53
seagrass	0.80	0.17	0.28			

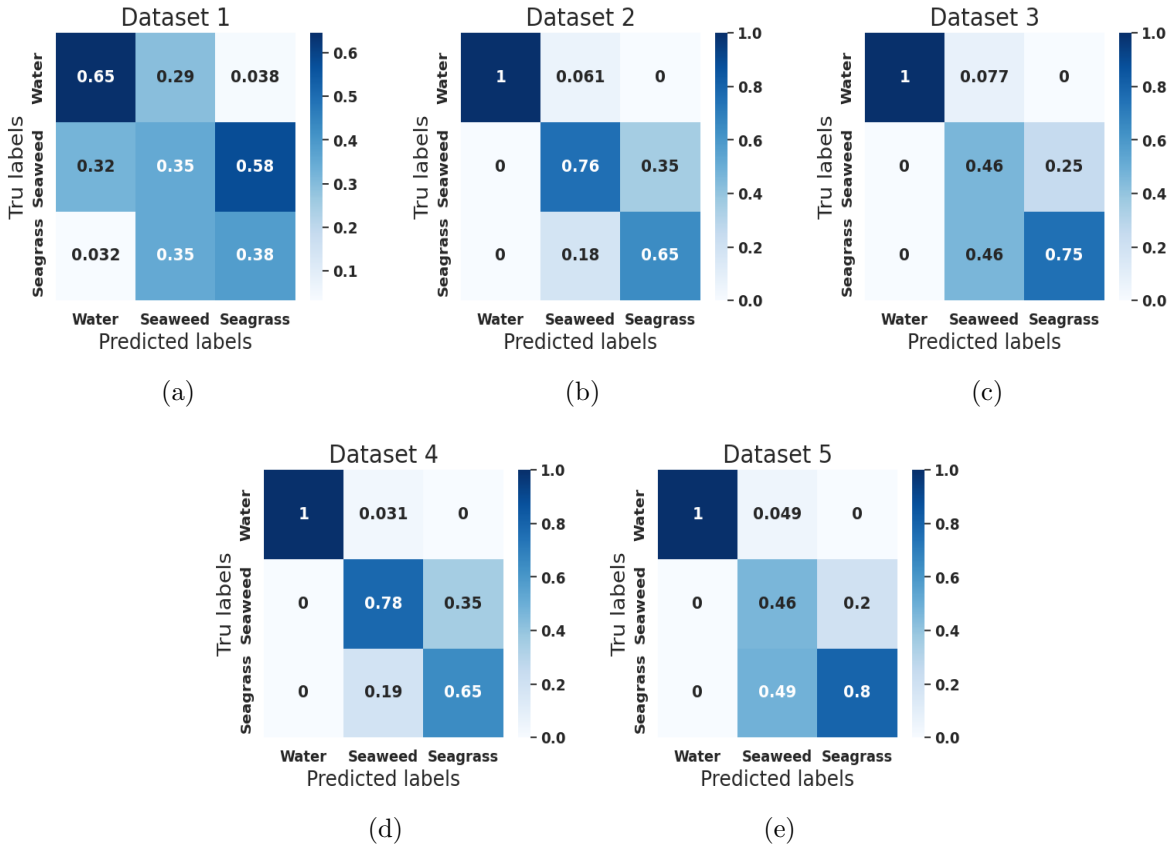


Figure 4.5: Confusion matrices from the soft test of SSL. The individual accuracy scores for water, seaweed, and seagrass for datasets 1, 2, 3, 4, and 5 are given, obtained with data from Casco Bay.

The confusion matrices in Figure 4.5 (a)-(e) are included to highlight the findings of over- and underestimation. For example, Figure 4.5(a) representing dataset 1 shows that both seagrass and seaweed resemble accuracy equal to chance ($\sim$ 30%) and that 30 % of the seaweed samples have been classified as water, while the remaining as seagrass. The seagrass class for the same dataset confused almost 60% of its samples for seaweed here, thus the SSL model has a hard time classifying. While the classification was good for the water class for all datasets, with a 100% accuracy it hints at overfitting, thus a drop in accuracy for the hard test was expected. Results improve for the remaining datasets, however, as Table 4.3 showed, seagrass and seaweed are confused with each other. An interesting finding is that seaweed is better detected in datasets 2 and 4, while seagrass is better detected in datasets 3 and 5.

An overview of the hard test results is shown in Table 4.4 for the semi-supervised classifier. From the soft to hard test, accuracy, the F1-score, and Tau(e) values drop as expected. Like previous findings in Table 4.1, 4.3, and 4.2, precision and recall show that seagrass and seaweed are confused with each other.

Table 4.4: Results summarized from the SSL model with Random Forest (RF) and pseudo labels at Saco Bay (hard test). P = Precision. R = Recall. F1-score = individual class F1-score. Avg. F1-score = the Average F1-score for the given dataset. OA = Overall Accuracy for the given dataset. T = Tau(e) for the given dataset.

Class	P	R	F1-score	Avg. F1-score	OA	T
Dataset 1						
Water	0.74	0.91	0.82			
Seaweed	0.46	0.61	0.48	0.57	0.59	0.23
Seagrass	0.51	0.34	0.41			
Dataset 2						
water	1.00	1.00	1.00			
seaweed	0.64	0.64	0.64	0.76	0.76	0.72
seagrass	0.64	0.64	0.64			
Dataset 3						
water	0.86	0.85	0.85			
seaweed	0.54	0.85	0.66	0.66	0.68	0.53
seagrass	0.76	0.33	0.46			
Dataset 4						
water	0.97	0.98	0.97			
seaweed	0.61	0.58	0.59	0.73	0.74	0.74
seagrass	0.62	0.65	0.64			
Dataset 5						
water	0.99	0.76	0.86			
seaweed	0.55	0.81	0.65	0.72	0.71	0.53
seagrass	0.75	0.56	0.64			

The related hard test confusion matrices are seen in Figure 4.6. Overall, they reflect the findings of over- and underestimation for Table 4.4. However, the same finding as in the confusion matrices for the soft test, they clearly highlight that seagrass is found better detected by the classifier in Figure 4.6(c) and 4.6(e) and seaweed is better detected in Figure 4.6(b) and 4.6(d).

To summarise the results of the SSL model, an expected drop in accuracy from the soft to hard test was seen, seagrass and seaweed are more confused with each other than the water class, and using datasets 2 and 4 increased detection of seaweed while using dataset 3 and 5 increase detection of seagrass. The box-plot and line-plot in Figure 4.7(a) and 4.7(b) respectively display the accuracy drop. From the validation accuracies in the box plot, the median is very similar for all models laying at around 90-95% accuracy, however with wide boxes for all, except dataset 4, they are more uncertain, indicating overfitting.

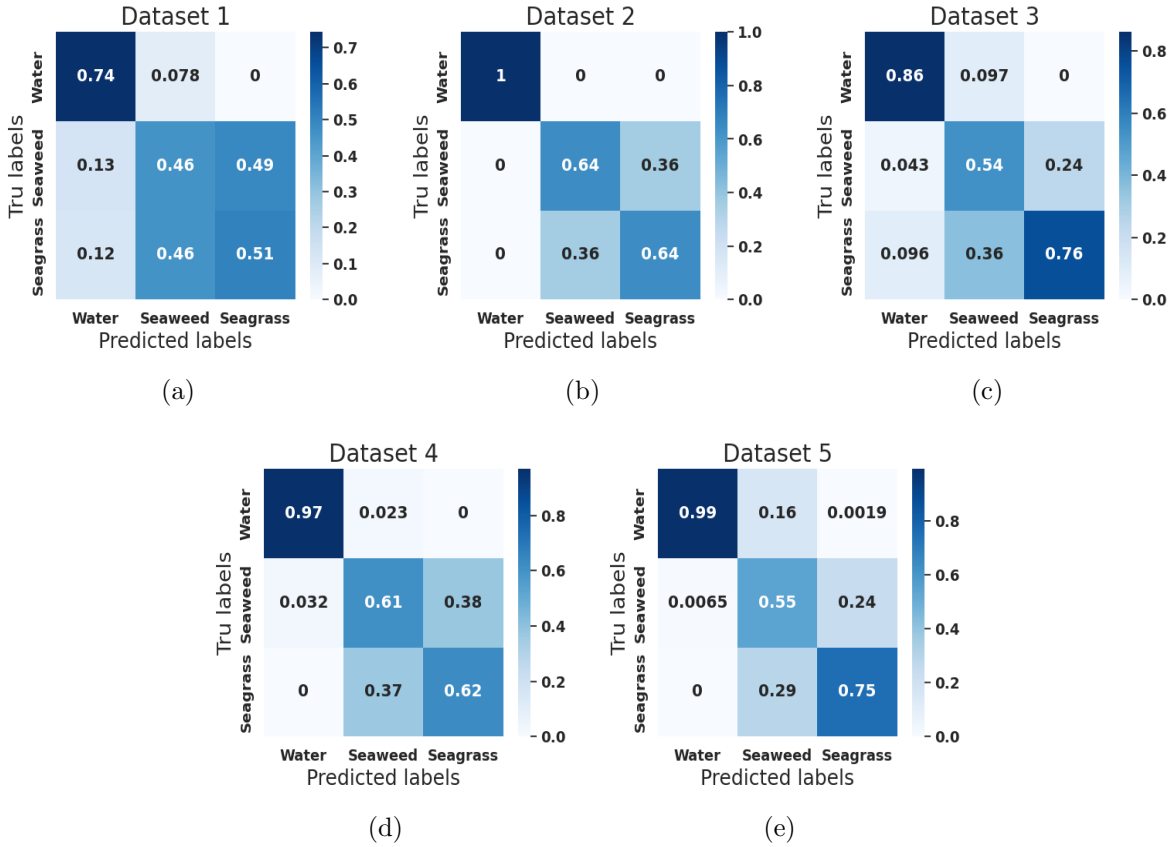
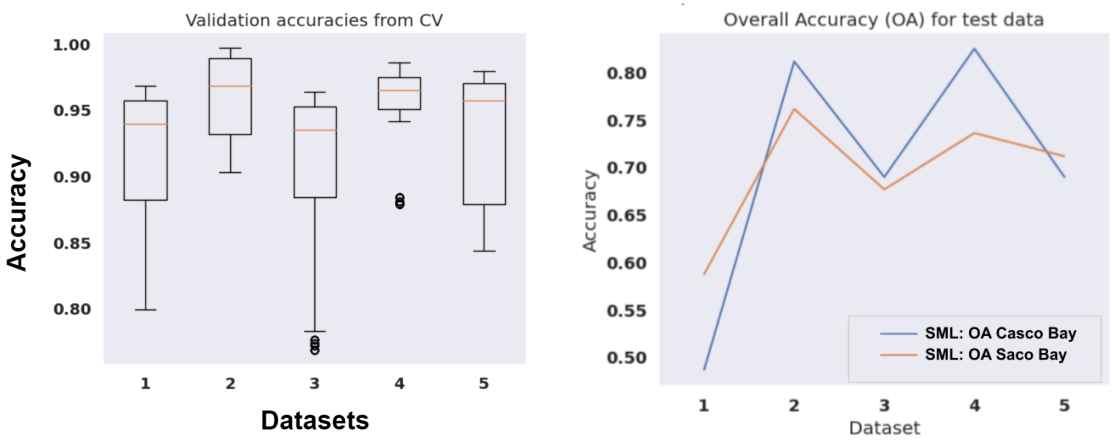


Figure 4.6: Confusion matrices from the hard test of SSL. The individual accuracy scores for water, seaweed, and seagrass for datasets 1, 2, 3, 4, and 5 are given, obtained with data from Saco Bay.



(a) Box plot for datasets 1, 2, 3, 4, 5. The data represents validation accuracies obtained during CV with 10 folds.

(b) Line plots of OA from Casco Bay (blue line) and at Saco Bay (orange line).

Figure 4.7: SSL results. Illustration of a box plot (a) and a line plot (b) of Overall Accuracies (OA) achieved with different feature combinations.

4.3 Unsupervised Classification Results

Out of all the classification algorithms, the UML yielded the worst results. Table 4.5 for its soft test shows that the accuracy range between 8-41%, F1-score from 7-40%, and low or negative Tau(e) values make the model resemble chance-like behavior. Furthermore, both precision and accuracy values are low for all datasets, showing high confusion between all three classes.

Table 4.5: Results summarized from the UML model with K-means at Casco Bay (soft test). P = Precision. R = Recall. F1-score = individual class F1-score. Avg. F1-score = the Average F1-score for the given dataset. OA = Overall Accuracy for the given dataset. T = Tau(e) for the given dataset.

Class	P	R	F1-score	Avg. F1-score	OA	T
Dataset 1						
Water	0.02	0.02	0.02			
Seaweed	0.29	0.50	0.37	0.22	0.23	- 0.16
Seagrass	0.49	0.18	0.26			
Dataset 2						
water	0.01	0.01	0.01			
seaweed	0.10	0.10	0.10	0.18	0.17	-0.24
seagrass	0.42	0.42	0.42			
Dataset 3						
water	0.55	0.41	0.47			
seaweed	0.33	0.60	0.42	0.40	0.41	0.12
seagrass	0.56	0.22	0.32			
Dataset 4						
water	0.00	0.00	0.00			
seaweed	0.00	0.00	0.00	0.07	0.08	-0.38
seagrass	0.21	0.24	0.22			
Dataset 5						
water	0.16	0.27	0.20			
seaweed	0.45	0.26	0.33	0.18	0.18	-0.24
seagrass	0.00	0.00	0.00			

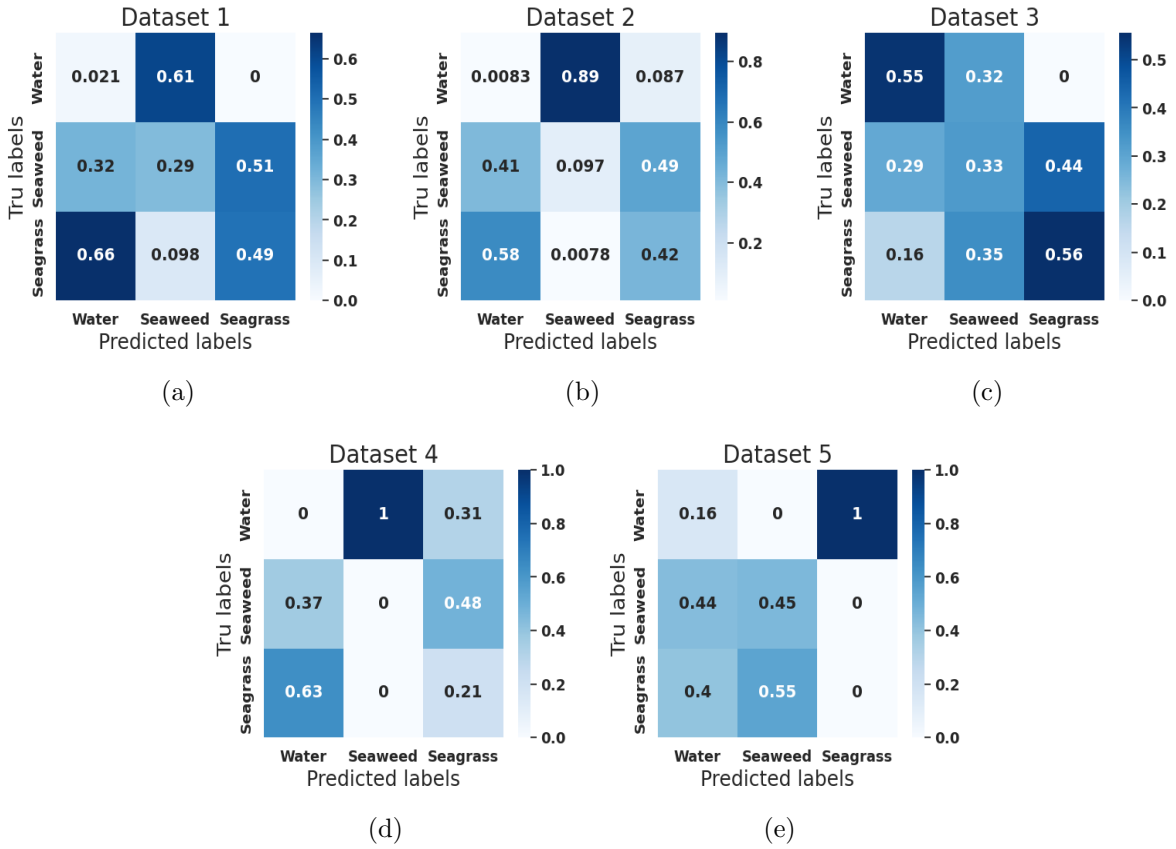


Figure 4.8: Confusion matrices from the soft test of UML. The individual accuracy scores for water, seaweed, and seagrass for datasets 1, 2, 3, 4, and 5 are given, obtained with data from the Casco Bay.

The confusion matrices in Figure 4.8 for the soft test support the findings in the above table. Ideally, the matrices would have a dark blue diagonal from top left to bottom right, indicating correct classification, however, no pattern is seen as the blue color is scattered. Moreover, the classes' individual accuracies are around 30%, echoing that the results resemble a random state.

For the hard test, results are poor here too. Table 4.6 (see next-next page) shows that the UML's accuracy range between 10-32%, F1-score from 7-25%, and with only negative Tau(e) values making the results, identical to the soft test, resembling random behavior. Also, both precision and accuracy values are low for all datasets, thus confusion is very present.

The confusion matrices for the hard test in Figure 4.9 neither shows the ideal confusion matrix. This makes it hard to identify any trends of confusion between the classes as they seem all equally confused. Condensing the findings of the UML, it was found highly confused, and resemble results equal to chance. Figure 4.10 (next-next page) emphasize this with low accuracy for both the soft and hard test, with little apparent patterns.

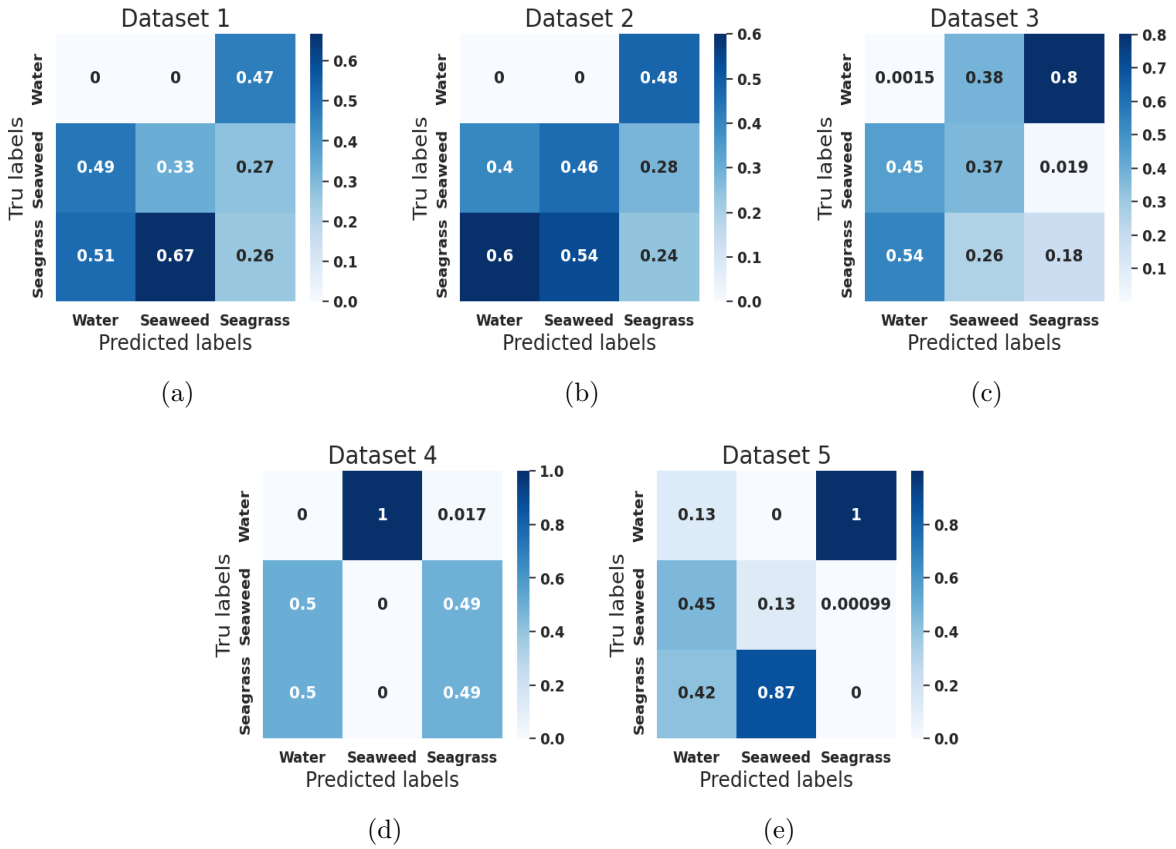


Figure 4.9: Confusion matrices from the hard test of UML. The individual accuracy scores for water, seaweed, and seagrass for datasets 1, 2, 3, 4, and 5 are given, obtained with data from Saco Bay.

Table 4.6: Results summarized from the UML model with K-means at Saco Bay (hard test). P = Precision. R = Recall. F1-score = individual class F1-score. Avg. F1-score = the Average F1-score for the given dataset. OA = Overall Accuracy for the given dataset. T = Tau(e) for the given dataset.

Class	P	R	F1-score	Avg. F1-score	OA	T
Dataset 1						
Water	0.00	0.00	0.00			
Seaweed	0.33	0.01	0.02	0.12	0.19	- 0.22
Seagrass	0.26	0.55	0.35			
Dataset 2						
water	0.00	0.00	0.00			
seaweed	0.46	0.39	0.42	0.25	0.30	-0.05
seagrass	0.24	0.51	0.33			
Dataset 3						
water	0.00	0.00	0.00			
seaweed	0.37	0.56	0.44	0.19	0.22	-0.17
seagrass	0.18	0.09	0.12			
Dataset 4						
water	0.00	0.00	0.00			
seaweed	0.00	0.00	0.00	0.22	0.32	-0.02
seagrass	0.49	0.97	0.65			
Dataset 5						
water	0.13	0.28	0.18			
seaweed	0.13	0.01	0.03	0.07	0.10	-0.35
seagrass	0.00	0.00	0.00			

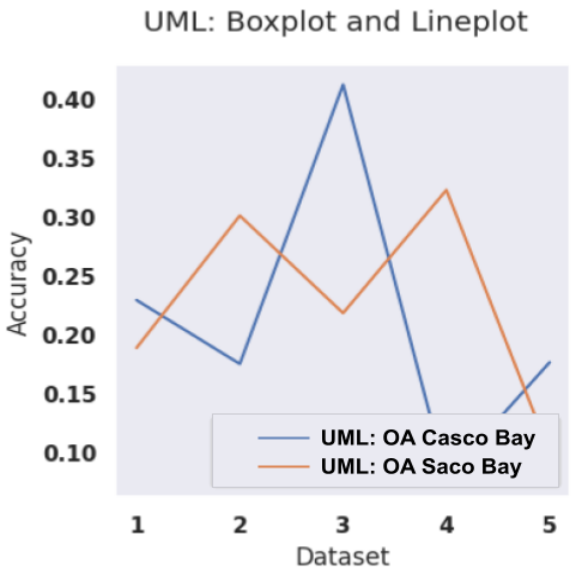


Figure 4.10: OA for Casco Bay (blue line) and Saco Bay (orange line)

Chapter 5

Discussion

This thesis explored a novel ML model, the potential of free data, and new RS technology in the field of seagrass and seaweed classification. The discussion will shed light on its most significant findings from the results, and review and compare them against related work. Thus, the RQs will be restated and answered, followed by a review of the methodology, and lastly, a section on general thesis limitations will be presented. The discussion facilitates a crucial part in evaluating this thesis's contributions to the field of environmental management of seagrass as seaweed. The research questions for this thesis were as follows:

RQ1 [Accuracy] How accurate is the machine learning (ML) algorithms (supervised, unsupervised, and semi-supervised) in classifying seagrass (*Zostera marina*) from seaweed (mainly, *Aschophyllum nodosum*) using only freely available data and software?

RQ2 [Implications and Applications] What are the potential applications and implications of using ML for classification in Google Earth Engine (GEE) for community monitoring and environmental management?

5.1 Exploring Research Questions 1

The experimental results answer RQ1. The main findings were:

- The accuracy of SML agress with related work
- SSL shows potential in classifying aquatic vegetation from water
- UML with K-means was not found a suitable algorithm

First up for discussion is the SML model results. The highest OA scores with the SML using RF achieved OA = 91% and Tau(e) = 0.86 for the soft test, while an OA of 75%

and respective Tau(e) value of 0.62 for the hard test. Given that these results aligned with related work in section 1.5 and previous literature [86, 105, 85, 47, 150, 68], they were partially expected [48], but still unexpected as it was applied to a new area with new data, additionally, compared to related work, the pre-processing was much poorer, thus making the results better than expected.

Comparably, similar studies have reported equal or better OA compared to this thesis's results for SML using the same RF. Distinguishing between studies using soft and/or hard tests further provides a more proper comparison between their results and this work to aid the understanding of *how* and *why* the observed accuracies here are as they are. Of related work, the highest OA for the soft test was 97%, while the highest accuracy for related works using a hard test (see [19, 26]) was 72%. This accuracy does don't deviate a lot from the observed highest 75% nor from the SML's average accuracy equal to 70%, thus the marginal difference between this works accuracies and related works can be attributed to chance. It should also be mentioned that the related works of [46] also used a hard test and achieved an OA of 88%, but did not employ RF, instead, an algorithm with a series of filter thresholds based on a tree-like structure. Given the exception of [46], the RF algorithm employed in this thesis aligns with previous related studies using the same RF algorithm.

Despite the SML's results being found to comply with related work, it cannot be concluded that the algorithm classifies seagrass, seaweed, and water well. This claim stems from how related work trying to do the same classification using RF failed. For example, [48] also intended to classify seagrass and seaweed, but found it not feasible due to their spectral signature being too similar. Instead, they ended up classifying SAV versus not-SAV. Likewise, this thesis of [44] took the same turn, and the same for the related works of [47]. This begs the question: *despite this work proven high OA and Avg. F1-score for the hard test, does it manage to distinguish seagrass from seaweed given this is the main goal?* The answer to this is: mostly no. The water class was the best-classified class and hence was the main contributor to the high OA and Avg. F1-score observed here, which can be misleading as already discussed in section 2.4.2. 70-75% in accuracy for the SML is not far from 66%, hence the results can be attributed to how the classifier well detects water but still confuses seagrass and seaweed for each other. Also, the individual F1-scores, pair-plot, confusion matrices, and spectral signature derived echo the claim and findings of [48, 44], and [47], that seagrass and seaweed are hard to separate. This also accounts also for the SSL model's findings, hence a more appropriate statement is to say that the SSL, which will be discussed shortly, shows potential in classifying *aquatic vegetation* from water.

However, before discussing the SSL's potential and its results, the anomaly found will be addressed. An exception was found in the individual F1-score of the seagrass class with

dataset 3 for SML (hard test), with a value of 88%, compared to all the others of F1-score ∼66% or lower. This could be a mistake and thus the experiment is recommended to be re-run, or it could signify that the feature combination used here enables a better detection of seagrass compared to the other feature combinations. Dataset 3 contained the bathymetry, B3, NDVI, area, width, height, B4, B8, perimeter, B2, and the number of clusters ordered by influence (from the SHAP algorithm), all GEOBIA-derived features which represent the average values for each super-pixels/object in the image. The contribution of the bathymetry to the high individual F1-score for seagrass can be reasoned from the literature on seagrass and seaweed habitats characteristics. Despite the species overlapping nature, differences related to bathymetry, B3, and NDVI can aid explain the observed outlier, mainly by habitat location and color. The dominant seaweed community in Maine consists of *Aschophyllum nodosum* which tolerate air exposure and rely on rocks for their habitat substrate. Thus, along the rocky coastline of the AOI, seaweed are assumed to be more prevalent than seagrass. Eelgrass on the other hand doesn't tolerate being equally much as seaweed above water, and additionally needs soft sediments to root in which is why they are assumed to be found further away from land. These habitat differences can aid in explaining why bathymetry becomes such an important contributor. While both species co-inhabit the same shallow waters as they need both water and sunlight for photosynthesis, there are still differences in their exact locations which seemingly the bathymetry feature captures. Furthermore, the importance of bathymetry was also mentioned by [19] and [4], while indirectly by [47, 48, 52], and [26] as they apply the feature.

Furthermore, considering both B3 and NDVI which represent green and vegetation respectively it makes sense that it is eelgrass and not seaweed that is being better detected as eelgrass are green while the dominant seaweed species have a characteristic brown color. However, a co-founding factor could be that the sensor is not picking up on seagrass, but rather on other types of green vegetation or green-looking waters given the fact that the satellite data is from the summertime when the biomass is peaking, and eutrophication leading to algae bloom, or other types of pollution which can cause seagrass to be overestimated. Lastly, considering that there are three other datasets that also use bathymetry, NDVI, and B3, why did dataset 3 do presumably better? There are two main differences between dataset 3 and the rest which are that it uses the GEOBIA-derived method and 5 additional features derived from this process. Given that the next most contributing features after NDVI are area and width, this can be attributed to the configuration eelgrass has. They are known for creating extensive meadows as they spread by cloning themselves via the rooting system [151], they are reasonably more clustered or dense than seaweed, which is not connected with each other. Thus, that dataset 3 uses GEOBIA-derived features seemingly aid the detection of seagrass.

Moving on to the second main finding for discussion, SSL with RF and pseudo labels

showed potential in mapping aquatic vegetation from water. As argued above, a distinction between classifying aquatic vegetation versus specifically seagrass and seaweed is made, given the low individual F1-scores compared to the OA, thus taking all the external evaluation metrics into account, it was not found that any of the classifier well-distinguished seagrass from seaweed, but rather aquatic vegetation from water. SSL with RF and pseudo labels obtained accuracy values, like the SML, within the range of general acceptance for the field of RS [152]. Its highest OA = 82% was achieved with dataset 4 on the soft test, while for the hard test, OA reached 76% with dataset 2. Potential explanations for the results obtained with the SSL are attributed to the model, features, image classification method, chance, and the size of training data. Acknowledging that there can be many reasons for the SSL achieving these results, only a few will be discussed based on the author's own thoughts and limited knowledge.

Generally, the RF algorithm is known for its high accuracy, and in the context of classifying seagrass, seaweed, and water its relatively good performance can be credited to its inherent flexibility. To support the previous statement a comparison between RF and K-means is given. Because RF is an ensemble of DTs, and each DT is a tree-structure based on random features which the RF aggregate over, making it flexible. However, the trade-off for flexibility is time, which is where K-means becomes empirically superior in this work. The K-means were found to converge fast, needed less hyperparameter tuning, and no labeled data, however, it is the type of data used that ultimately determines if the above-mentioned are advantages or rather a hinder. In this thesis, the objective function of K-means to optimize for similarity and dissimilarity using a distance metric seems to be a constraint. Figure 5.1 provides an example of how the seagrass and seaweed samples are almost identical and majorly overlap with each other, and hence, a more adaptive and modular algorithm is suited, such as the RF.

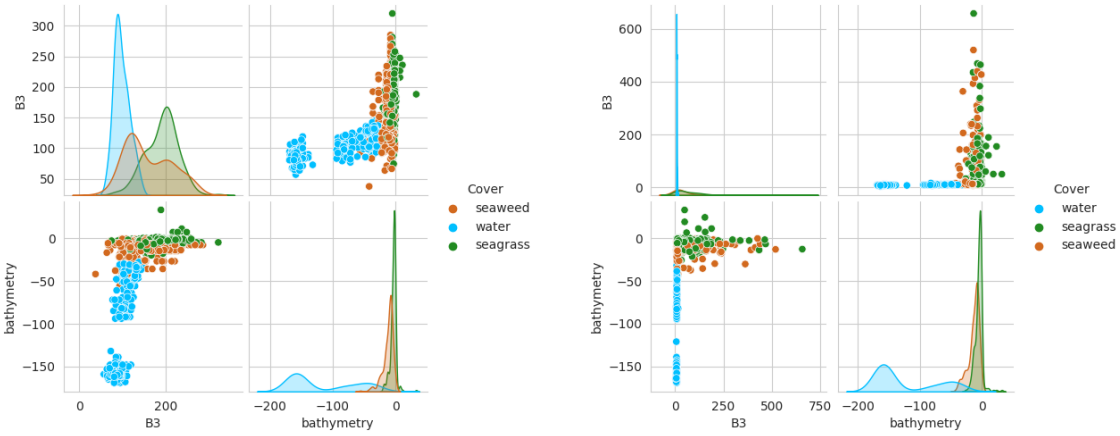
Moving on from looking at the inherent traits of RF, features contributing to the OA for the hard test were mainly found to be the PB-derived bathymetry and B3, given by the SHAP algorithm employed on SSL for dataset 2, and the same applies to dataset 4. These features are the same as discussed for the SML and are there not repeated here. Furthermore, the highest accuracies for the SSL model were only obtained with features from the PB image classification method. For the soft test, the difference in accuracy from datasets 2 and 4 to datasets 3 and 5 is greater than the difference between them in the hard test, however still greater. Explanations for this can either be that the PB works better with SSL than the GEOBIA method, or the accuracy can be attributed to the features/feature combination as all the four datasets are not equal, which disabled a proper comparison between PB and the GEOBIA method. Furthermore, as the difference in accuracies between the PB and GEOBIA method in the hard test are not very big, the difference can also be due to chance.

Another explanation for why the GEOBIA and PB do not differ much in the hard test can be attributed to how Sentinel-2's spatial resolution is in a grey zone between high and medium, or because the objects to be detected span around 100m^2, making them just equally detectable by the GEOBIA method as with the PB. Most of the related work used PB, while one did not. While the general literature on aquatic vegetation mapping finds GEOBIA to outperform PB [4, 153, 60], this was mostly for tropical regions with clear waters and high spatial resolution, such as commercial satellites with spatial of 0.6-2.4-meter, while - as [42] argue - that for turbid coastal waters, with patchy seagrass meadows, a PB approach is more suitable, as was the case for [154] and assuming, partially the case here.

Lastly, how the size of the training data can have affected SSL's result is discussed. Ultimately, SSL used a bigger dataset than the SML for training which can be the reason it achieved a better result than the SML. From the total amount of data, the SSL trained on 90%, while the SML only trained on 80%, and ideally the SSL should have trained on the same amount for a more acceptable comparison, thus this was a mistake made in this work. Hence, the results observed for the SSL are attributed to the size of the training data, choosing a flexible and robust algorithm against noise, the RF, features/feature combination, the PB method, and chance.

Comparing the results of SSL to related work could not be done as it, to the extent of the author's knowledge is not used in combination with Sentinel-2. However, similar literature that has used SSL with other RS sensors and algorithms to map seagrass are [25] and [90]. [25] reports an average class accuracy of 83.3% for a soft test, using a spatial resolution of 8cm (SONY ILCE-6000 camera) and an FCNN, a DL algorithm. Authors in [90] use a deep CNN, and report the best accuracy at 99.99% for a hard test, making the results more generalizable than the ones from [25]. However, unlike this work, they both use a much higher spatial resolution, and different models and tools which can explain the difference in accuracies.

Regarding the UML, the results proved poor. This can be due to the overlapping samples of all three classes as seen in Figure 5.1 and the very similar spectral signatures of seagrass and seaweed. Thus, the underlying problem lies in the wrong choice of UML classifier as K-means are preferred for spherical clusters, which was not seen here. This was speculated already in the phase where raw data was inspected, hence the choice of testing DBSCAN as well. However, this also performed poorly. A better choice might have been ISODATA, a much-used algorithm for classifying aquatic vegetation [45].



(a) Dataset 2. PB-derived samples

(b) Dataset 3. GEOBIA-derived samples

Figure 5.1: Pair plots of band B3, B8, and bathymetry. Seagrass is shown in green, water in blue, and seaweed in orange. Comparison of sample-values extracted with the PB and GEOBIA-method.

5.2 Exploring Research Question 2

The motivation behind this work was to explore the potential implications and applications ML in GEE has for community monitoring and environmental management. This was assessed through a soft test in Casco Bay and a hard test in the adjacent Bay, Saco Bay. In these AOIs, the performance of three ML models was compared on their classification mainly on their classification accuracy of seagrass, seaweed, and water. To answer RQ2, firstly, the findings of this work comply with related work in that GEE's main potential lies in its cloud computing power. Secondly, ML poses potential with the SSL model in solving a well-known issue in aquatic vegetation mapping: the lack of labeled data. This work found that SSL can work with less labeled data and still achieve similar and more robust results compared to the SML tested here. Third, this work also contributes to reproducible results as the methodology was based on freely available data, software, and the source code of this work can be found online, making this work applicable to community monitoring of seagrass and seaweed. Lastly, despite the above-mentioned potential, this work also found prevalent over- and underfitting in the results, which has implications for the further use of the methodology and the results.

GEE's cloud computing power affected the general workflow in speed, memory, and user-friendliness. Time and memory were saved by working with the satellite images on the cloud of GEE, compared to storing them locally. Regarding community monitoring, the cloud computing ability makes it possible to work with this work from any type of computer without requiring any 'fancy' hardware equipment for data processing or analysis. Furthermore, incorporating GEE into Python with the geemap library [155], made rather user-friendly, has enabled more people to participate [156]. Of related work, eight of

them used GEE as a tool and also attributed most of its potential to cloud computing characteristics. In addition to the above-mentioned benefits of GEE found here, other works highlight benefits such as automation, and up-scaling as the online method[1] of seagrass mapping by [3], low cost, and up-to-date data [52, 47].

ML poses a special application to the field of aquatic vegetation mapping, found here with the SSL model. This is given by how its results were within an acceptable range for aquatic vegetation mapping [152], and that the drop in accuracy from the soft to hard test was less for the SSL than for SML given by their line plots. Implications of these findings to community monitoring are how SSL can generalize better to new areas, and need less labeled data to train on compared to a fully supervised method, while still achieving equally high accuracy to it. Applications of this lie in the SSL and workflow can be further developed, eventually into an app like that of mapseagrass.org. However, this assumes that the observed accuracy is representative and correct, in addition to that the SSL model actually is as robust as found here. Conclusively, the SSL model should further be inspected and investigated for future work.

That this work is made freely accessible enables this further development of SSL and the workflow which is another implication of this work for community monitoring, free accessibility. This work and many of the related works also using GEE, provide code for their workflow through a link, taking one directly to the source code. By the thousands of volunteers across the globe, small code snippets of freely available data, knowledge, and computational power add up, posing the great potential for global-scale collaboration on seagrass and seaweed classification, which is a global-scale problematic, not to mention emerging other global-scale problems, such as climate change, which also is in need of global-scale collaborative, place-based knowledge from local to the global scale (!). Thus, freely accessible data, leveraged by the power of ML, cloud computing, and community monitoring arguably decrease the number of barriers to environmental management by providing reproducible results, the rapid development of existing work, and access to the knowledge needed for decision-making, crucial for the times ahead.

Lastly, despite the above-mentioned potential of this work, it is in need of considerable improvement before these can be realized, as overestimation and underestimation of seaweed and seagrass were prevalent in the experimental results. Possible explanations are everything from bad training data to mistakes in the pre-processing phase. The training data had a major time discrepancy and will be further discussed under section 5.3 on thesis limitation 5.3, while the atmospheric correction and sun-glint correction were sub-optimal, or the lack of water-column correction make the reflectance from seagrass and seaweed and water much more alike and thus result in confusion. Generally, in the field of aquatic mapping, over- and under-estimation is a big topic as their extent is often estimated to

[1] see https://mapseagrass.org/. Last accessed [07.05.2023]

monitor trends or they want to account for seagrass blue carbon storage [107]. Thus, many of the reasons for over-and underestimation that has been mentioned here are also mentioned in related works, in addition to factors such as turbidity and varying seagrass density [44]. Over- and under-estimation of seagrass and seaweed are important topics in the field of classification as the goal is to present stakeholders with informative data on the species distribution, trends, and behavior in order for them to take action to ensure their flourishing and restoration. Hence, using the predictions derived here could lead them to change current managerial boundaries between harvesting areas for seaweed and protected areas for seagrass, leading to over-harvesting or over-protection of the species. Therefore, to make use of SSL or the workflow here, substantial validation and further improvement it is strongly suggested before it could be used.

To summarize, it was found potential in the use of GEE, freely available data and software, SSL as an algorithm, and the work here is made accessible enabling others to build upon it. However, care needs to be taken when interpreting the results found here, as the communication of them to policymakers has implications for the environmental management of seagrass and seaweed. This has to do with the over- and underestimation of the two species. Underestimating the extent of seagrass and seaweed can signal urgency but decrease the size of protected areas. Overestimating seagrass can cause alarming decreases to go unnoticed. Regarding seaweed, overestimation of it where it co-inhabits with seagrass can lead to the over-harvesting of both species. Eventually, making both harder to restore which has both financial and ecological consequences. Therefore, before using the results and workflow here, it is advocated to do substantial verification of the results to make sure they are robust and trustworthy before they can be of any use.

5.3 Thesis limitations

Limitations of this thesis mainly focus on the lack of training sites (for the soft test), pre-processing (DN-conversion to spectral units, water-column correction), non-optimal GT data (time discrepancy), and limitations of community monitoring. Further limitations of this thesis will be briefly mentioned to acknowledge their presence, however, are not discussed due to time constraints. Many of the above-mentioned issues are correctable, given more time. Thus, an underlying cause of many of the above problems was restricted time.

In this work, only two sites were used for the study case. One for the soft test (training and validation) and one for the hard test (testing). While this is more than most related work had, having more training sites could have decreased the overfitting (drop in accuracy) of the SML model from the soft to hard test. This could have been done either by augmenting the satellite image or by cutting the image up into multiple smaller patches.

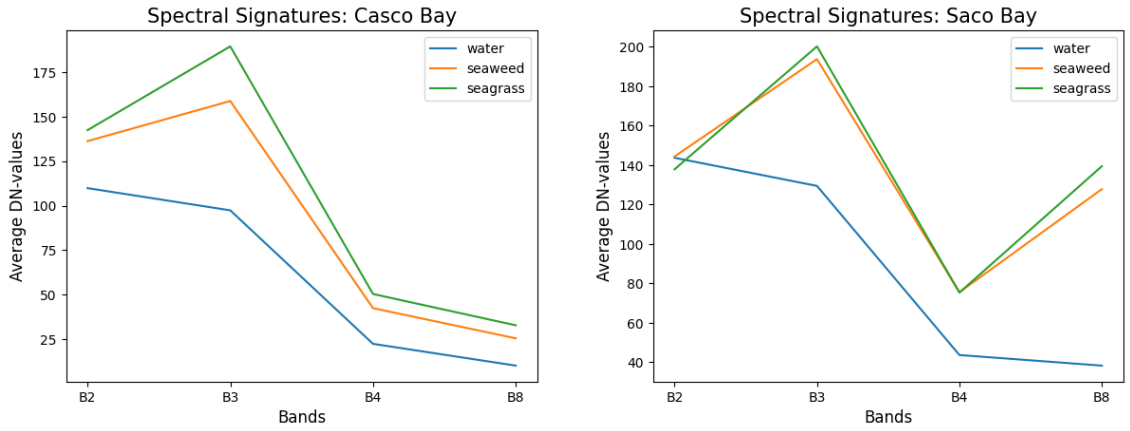
Code for this was developed (see [157]), but not implemented due to time constraints.

Lack of pre-processing is also a potential cause for why the SML overfitting on the training site's characteristics. Firstly, this thesis classification was done using raw DN values. According to [3], this is problematic as the image of one place can't be compared with another as the DN values are image specific. Consequently, any analysis, classification, and comparison of the two will be affected by the sun's location, viewing geometry, weather conditions, etc. The problem of using DN values for interpreting results was ignored here mainly due to the author's understanding of the level-2A product and time constraints. Sentinel-2 level-2A images used here have been atmospherically corrected by ESA using Sen2Cor, and it was assumed they had corrected it for the sun's location viewing geometry, etc. leading to the conclusion that it was OK to compare the results from the two different satellite images as they had been pre-processed the same way by ESA. The second reason for not converting DN values to radiance (giving them a spectral unit) was due to the fact that this issue was first discovered after analysis had been done, and was not able to redo the experiment due to time restrictions. Based on this, the author acknowledges using DN values could lead to misleading and inaccurate classification results and recommend for the future to redo the analysis after converting DN values to radiance.

Secondly, the water-column correction was another pre-processing step not included in the process, which could be the cause of the errors seen in the results. The correction was not applied as it was not known if the water clarity was consistent across the study sites given the use of composite images and lack of local knowledge by the author. One of the related studies ([48]) had experimented with water column correction but ended up not using it due to the above-mentioned reasons. The GT data used in this thesis was was done in June and according to local turbidity monitoring [158] the waters were on average below 10 nephelometric turbid unit (NTU which is considered "clear" [159], thus this thesis followed the approach as for the work by [48].

Furthermore, three limitations for GT data were found: time discrepancy in seagrass and seaweed GT, the multi-class problem for seaweed, and the validity of created water data. Firstly, both the seagrass and seaweed data are from before December 2018, while Sentinel-2 first starts covering Casco Bay from December 2018 and onward [160]. This results in a time discrepancy between when GT data was collected, and the satellite data used to gather sample data for classification. Figure 5.2(b) illustrates this time discrepancy where Figure 5.2(a) shows the spectral signature of samples derived from a satellite image a year after (such as in Casco Bay), while Figure 5.2 show an image captured in parallel with when the GT data was collected (such as at Saco Bay).

Therefore, to compensate for the time deviation, literature and talking to locals were done to give some idea if significant changes occurred in the area (e.g., storms, oil spills, etc.). Based on the literature read and local knowledge, storms do occur and can change



(a) Spectral signature of the training data from Casco Bay.

(b) Spectral signature of the test data from Saco Bay.

Figure 5.2: Spectral signature of water, seaweed, and seagrass for the six features B2, B3, B4, B8. Given in DN-values. Dataset 2 was was used for illustration

the configuration of eelgrass [120]. Related literature who were impacted by this was [48] who collected field data a few days after Hurricane Dorian passed through the area and had to perform additional quality control steps. Without such stressors, the general literature on seagrass and seaweed states that their configuration does not change much [151, 66, 161]. For this work, as few other options for GT data on eelgrass existed, this data was used for training, despite the known time discrepancy.

Secondly, using multiple species to represent seaweed is not optimal as their behavior can vary greatly and cause confusion in the spectral signature. Nonetheless, it was seen as the best option. Originally considered, was the ASCO dataset, a community monitoring project crowdsourcing data solely on *Aschophyllum nodosum* [162]. While this would have been a good demonstration of the power of community monitoring, it was not chosen mainly due to too few data points based on recommendations in subsection 3.3, hence this work also demonstrates the very common issue in benthic habitat mapping, the lack of in-situ data, as it became a direct and indirect limitation of this work. Additionally, the dataset contained some misrepresentations leading to less validity and mistrust of their data which is a general phenomenon for geographical information obtained by volunteers despite evidence that it can provide equally accurate data as professionals [163, 164, 165, 166, 167]. Nonetheless, the mistrust is understandable as illustrated by Figure 5.3.

Figure 5.3 exemplify why the ASCO dataset was deemed inadequate for training the ML algorithms. To the right side of the figure, an overview of the entire ASCO dataset is shown, while to the left are two boxes (A and B) containing examples of dislocated data points (nr.40 and nr. 50). Both data points, 40 and 50, should represent seaweed, however, the points are in the middle of a forest and sampling them would hence not capture features representing seaweed. Therefore, it is advised to be cautious when using

data from community monitoring efforts [168, 169]. For this thesis, the goal was therefore to find a method to monitor seagrass and seaweed satisfying the intersection between including volunteers and ensuring valid results. No data from community monitoring was used, however, all data, source code, and software used here were freely available, and the workflow created can thus be reused and further worked on and incorporate crowdsourced data from the ASCO dataset at a later point in time, contributing to future community monitoring.

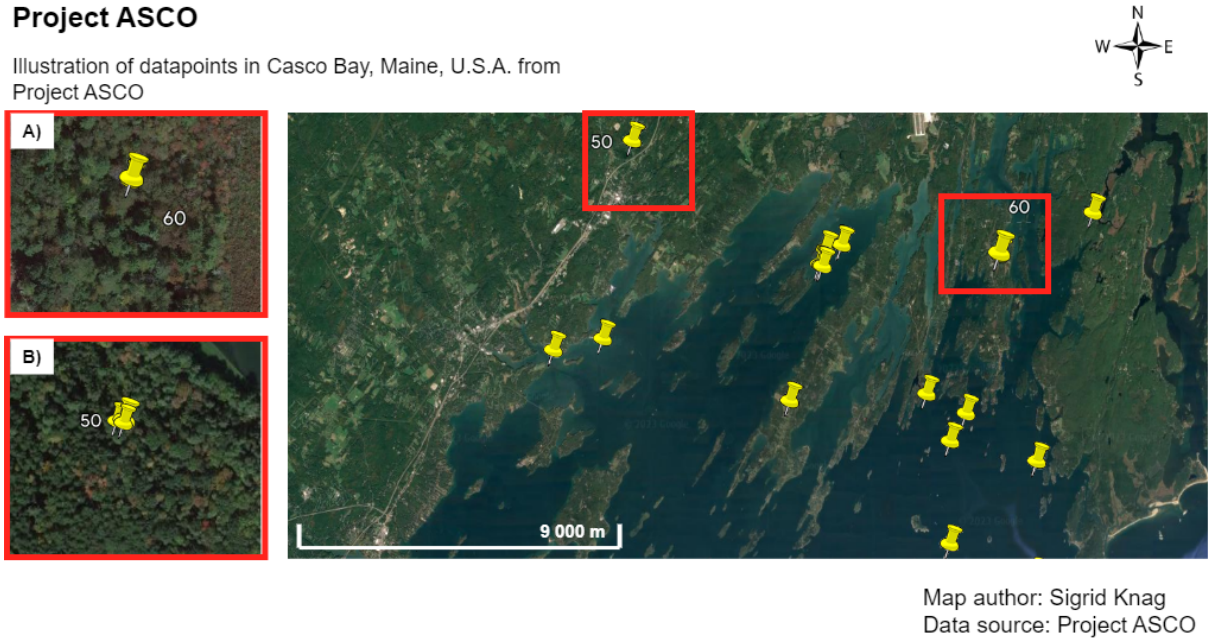


Figure 5.3: The ASCO dataset in Casco Bay. The right pane gives an overview of all 36 data points (yellow pins) in Casco Bay, Maine. In the left pane, boxes A) and B) display zoomed-in views of two dislocated points representing Rockweed. All data points are in GCS WGS84.

Thirdly, the validity of the created water data is questionable. This is because no on-the-ground sampling or labeling was done, rather, created remotely by a non-expert. Another considered option to obtain GT data on the water was to use the NDVI. NDVI is mainly used to detect vegetation by combining the Near-infrared (NIR) band with the red (B4) band which strongly reflects vegetation. As the NDVI ranges from -1 to 1, values around 1 indicate dense vegetation, while 0 can represent bare land or rocks, and values around -1 water. Thus, sampling the entire image and sorting out all values that are around -1 could give a dataset of water [135].

Lastly, other areas in which this thesis work falls short in both data collection, methodology, and results are briefly mentioned. From the data collection method, the few points from the ASCO dataset could have been incorporated into the training data used here. All external data sources, especially the ones on seaweed from MGS and bathymetry data

could have been investigated in more depth to increase their validity, currently, it was not done as the emphasis was more on creating a functional workflow and being able to produce some results instead of none. This makes the results less valid given the known phrase "shit in, shit out". However given a workflow that works, the data source could later be changed to produce better results. Furthermore, in pre-processing, a deep water mask was not applied, which could have been the reason why the image classifications deviated from the quantitative results and hence were not displayed. Additionally, the lack of removing negative values (image noise) could have impacted the results, and stripe removal did affect the image classification (results not shown). Regarding results, this thesis falls short on the evaluation of the results by not including producer and user metrics, which could have made it easier to see how well the model delineated specifically seagrass and seaweed, and as many related studies used this, it would have been easier to compare results on a more detailed level in regards to seagrass and seaweed. Many of them were explored, but not applied due to time constraints. Luckily, future work has the potential to improve on many of this thesis's limitations.

Chapter 6

Conclusion and Future Work

The first objective of this study was to assess the accuracy of three different ML models in classifying seagrass (specifically *Zostera Marina*), seaweed (mainly *Aschophyllum nodosum*), and water, using only freely available data and software. The second objective was to understand the potential implications and application of using GEE and ML in the field of RS and for community monitoring. Thus, this thesis's main contribution to the field of RS and aquatic vegetation mapping has been to explore a new SSL model in the field, the geemap integration of GEE to Python for analysis, the potential of freely available data and software, and shed light on community monitoring's role and potential in the field.

Three ML models, SML with RF, UML with K-means, and SSL with RF and pseudo labels, with GEE, were explored and performance compared through a case study in Casco Bay and Saco Bay of Maine, U.S. In classifying wild seagrass from seaweed. For validity, both a soft and a hard test were performed. Casco Bay provided a soft test (training), while Saco Bay, provided a hard test (testing). Motivated by aiding a community monitoring project in restoring seagrass (specifically considered here eelgrass) in this region, the workflow explored only freely available data, software, and analysis tools to make it reusable by them or other stakeholders in the field of aquatic vegetation mapping as seagrass and seaweed restoration are of global concern due to their many ecosystem benefits.

Conclusively, the main finding of this study was the SSL shows potential as a robust model in classifying aquatic vegetation from water at Saco Bay of Maine, achieving the highest OA of 76% and a Tau(e) = 0.72 (hard test) using a PB-derived method with the features B2-B4, NIR, NDVI, and bathymetry, in-line with related work and literature in RS [152]. The implications of ML integrated with GEE for community monitoring and environmental management were explored and found that the cloud computing capabilities of GEE leverage citizen science in saving memory and speed strive to become accessible to a wide audience, and most importantly, provide easy access to freely available satellite data and GIS tools for managing the environment. Further building upon this work and

advancing it makes it applicable. Especially since this work uses real-life data and applies it to a real-life problem, makes the findings here inform stakeholders on tools, data, and techniques that can be used and how. Thus it is seen as a starting point for future research.

Before the work here actually becomes applicable, further testing and validation of the results should be performed. While results are promising, care is recommended as the results show a tendency to over- and underestimate seagrass and seaweed in addition to confusing the species. Hence, a better description of the current results is that the ML model classifies aquatic vegetation from water, and not seagrass and seaweed in particular. Furthermore, the GT on both seagrass, seaweed, and water should be throughout examined for better validity as they are from different sources and times, which can impact the results. Therefore, further development of the current workflow is needed before it can become applicable. Nevertheless, optimistic that these barriers can be fixed and the fact that the workflow here is reproducible makes them manageable.

Arguably, this work's main contribution to the field of seagrass and seaweed mapping has been exploring the simple SSL model with RF and pseudo labels, which can score high on accuracy, use less labeled data compared to other conventional supervised ML models, and was found more robust than the SML model. Thus, the SSL shows potential in the field of RS for aquatic vegetation as it solves a common issue, lack of data. In the general field of science, this thesis's main contribution has been to produce open source code [157] and learn how to work with freely available data, software, and analysis tools. By itself, adding an extra code snippet to the big bucket of knowledge online does not make a difference, but when adding up the thousands of code snippets, coupled with information, and leverage by cloud computing power and ML for speed and accessibility, it adds up and does make a difference. Acquiring the skill of how to work with freely available tools is essential for the coming years ahead, as challenges such as seagrass and seaweed restoration, or the many complexities of climate change are global problems that need local-to-global solutions and collaboration. Conclusively, by exploring the potential of new technology such as ML and GEE in relation to community monitoring, the individual can find strength in numbers, leverage by tech to solve the problems of tomorrow *glocally*; by taking local action, that has global effects.

6.1 Future Work

Recommendations for future work are three-folded, firstly general recommendations are addressed, secondly, workflow improvements, and thirdly, how to obtain more valid and better qualitative and quantitative results are briefed. General recommendations are to explore the SSL model with dataset 2 as it achieved the highest and most robust results with the hard test. Additionally, further investigation of the GEOBIA by comparing for

instance dataset 2 with PB-derived features against a dataset x that has the exact same features, only derived with the GEOBIA method should be done to see if the results differed, which would tell if either the GEOBIA or PB method actually impact the results seen. Then, to improve the validity and accuracy of the current workflow, redo it and convert DN-values to values of spectral units [3], explore water-column correction, add more training sites by clipping one of the Sentinel-2 tiles into multiple regions or augment the image, apply pan-sharpening to derive more bands of 10m resolution from Sentinel-2 [170, 171, 172] and apply PCA to the SML and SSL techniques for feature selection instead of manually reducing them as recommended by [173, 174]. For more acceptable results, the training data can be improved by using the recently launched (March 2023) GT data on eelgrass from MaineDEP for Casco Bay which would remove the current time discrepancy. Furthermore, calculate the producer and user accuracy for seagrass and seaweed to assess how good the models are, specifically, detecting seagrass and seaweed compared to related work. Also, include more classes such as sand and bottom-habitat that are possible to GT, such as from the MGS dataset where seaweed data was sourced from as this will decrease the chance of seeing misleading OA. Lastly, to improve the visual result and their validity, deep water could be masked, and the resulting image could be compared with the current state-of-the-art platforms [3]. Regarding the quantitative results, the statistics should be looked into, and multiple configurations of the datasets should be applied and results re-run to obtain more results such that variance could be added to the OA score, making the analysis more rigorous.

Bibliography

[1] M. Potouroglou, G. Grimsditch, L. Weatherdon, and S. Lutz, "Out of the blue: The value of seagrasses to the environment and to people," United Nations Environment Programme, Nairobi, June 2020.

[2] R. Unsworth, L. C.-U. Cullen-Unsworth, B. H. Jones, and R. Lilley, "The planetary role of seagrass conservation," *Science*, vol. 377, no. 6606, pp. 609–613, 2022. [Online]. Available: https://www.science.org/doi/10.1126/science.abq6923

[3] T. Komatsu, M. Hashim, N. Nurdin, T. Noiraksar, A. Prathep, M. Stankovic, T. P. H. Son, P. M. Thu, C. Van Luong10, S. Wouthyzen11 *et al.*, "Practical mapping methods of seagrass beds by satellite remote sensing and ground truthing."

[4] S. Pittman, C. Roelfsema, D. Swanborn, B. Thapa, C. Say, A. Otaño C., S. Baez, and K. Jensen, *Outlining a methodological pathway to improve the global seagrass map*, 01 2022.

[5] L. Westerveld. (2020) Seagrasses from above - drones and satellites. [Online]. Available: https://www.grida.no/resources/13591

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] I. H. Organization, *Limits of oceans and seas*. International Hydrographic Organization, 1953, no. 23.

[8] L. R. Foundation, "Seaweed manifesto," 2020.

[9] FAO, "The state of world fisheries and aquaculture 2022. towards blue transformation," FAO, Rome, p. 226, 2022.

[10] K. R. Pedersen, P. R. Crane, and M. E. Friis, *Introduction to angiosperms*. Cambridge University Press, 2011.

[11] D. Alderton and S. A. Elias, *Angiosperms*. Academic Press, 2005.

[12] U. Nations, "The sustainable development goals 2016," eSocialSciences, Tech. Rep., 2016.

[13] B. M. Dewsbury, M. Bhat, and J. W. Fourqurean, "A review of seagrass economic valuations: Gaps and progress in valuation approaches," *Ecosystem Services*, pp. 68–77, Feb. 2016.

[14] W. M., D. C. M., C. T. J., O. R. J., D. W. C., O. S., C. A., F. J. W., H. K. L., A. R. Jr Hughes, K. G. A., K. W. J., S. F. T., and W. S. L., "Accelerating loss of seagrasses across the globe threatens coastal ecosystems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 30, p. 12377–12381, 2009. [Online]. Available: https://doi.org/10.1073/pnas.0905620106

[15] P. L. Erftemeijer and R. R. R. Lewis III, "Environmental impacts of dredging on seagrasses: a review," *Marine pollution bulletin*, vol. 52, no. 12, pp. 1553–1572, 2006.

[16] I. Normandeau Associates, "Southern maine seagrass mapping," 25 Nashua Road Bedford, NH 03110, 2022.

[17] IOC-UNESCO. (2023) Blue carbon. [Online]. Available: https://ioc.unesco.org/our-work/blue-carbon

[18] T. D. Pham, J. Xia, N. T. Ha, D. T. Bui, N. N. Le, and W. Takeuchi, "A review of remote sensing approaches for monitoring blue carbon ecosystems: mangroves, seagrasses and salt marshes during 2010–2018," *Sensors*, vol. 19, no. 8, p. 1933, 2019.

[19] R. Keay, I. Hassall, C. Fitzsimmons, and L. Castle. (2021) Collaborative seagrass mapping from sentinel-2 satellite imagery. [Online]. Available: https://medium.com/uk-hydrographic-office/collaborative-seagrass-mapping-from-sentinel-2-satellite-imagery-fd499edc75cd

[20] P. J. Hogarth, "44Seagrasses and Their Environment," in *The Biology of Mangroves and Seagrasses*. Oxford University Press, 05 2015. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780198716549.003.0003

[21] M. Waycott, C. M. Duarte, T. J. Carruthers, R. J. Orth, W. C. Dennison, S. Olyarnik, A. Calladine, J. W. Fourqurean, K. L. Heck Jr, A. R. Hughes *et al.*, "Accelerating loss of seagrasses across the globe threatens coastal ecosystems," *Proceedings of the national academy of sciences*, vol. 106, no. 30, pp. 12 377–12 381, 2009.

[22] D. G. Hole, P. Collins, A. Tesfaw, L. Barrera, M. B. Mascia, and W. R. Turner, "Make nature's role visible to achieve the sdgs," *Global Sustainability*, vol. 5, 2022.

[23] M. Land, W. Granéli, A. Grimvall, C. C. Hoffmann, W. J. Mitsch, K. S. Tonderski, and J. T. Verhoeven, "How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal? a systematic review," *Environmental Evidence*, vol. 5, no. 1, pp. 1–26, 2016.

[24] J. Vymazal, "Constructed wetlands for wastewater treatment: five decades of experience," *Environmental science & technology*, vol. 45, no. 1, pp. 61–69, 2011.

[25] B. Hobley, R. Arosio, G. French, J. Bremner, T. Dolphin, and M. Mackiewicz, "Semi-supervised segmentation for coastal monitoring seagrass using rpa imagery," *Remote Sensing*, vol. 13, no. 9, p. 1741, 2021.

[26] D. Traganos, B. Aggarwal, D. Poursanidis, K. Topouzelis, N. Chrysoulakis, and P. Reinartz, "Towards global-scale seagrass mapping and monitoring using sentinel-2 on google earth engine: The case study of the aegean and ionian seas," *Remote Sensing*, vol. 10, no. 8, p. 1227, 2018.

[27] L.-E. Rioux, L. Beaulieu, and S. L. Turgeon, "Seaweeds: A traditional ingredients for new gastronomic sensation," *Food hydrocolloids*, vol. 68, pp. 255–265, 2017.

[28] J. C. (FAO. (2021) Global status of seaweed production, trade and utilization. [Online]. Available: https://www.competecaribbean.org/wp-content/uploads/2021/05/Global-status-of-seaweed-production-trade-and-utilization-Junning-Cai-FAO.pdf

[29] S. Barker, ""eelgrass distribution in casco bay 2018"," Maine Department of Environmental Protection, 15 Little Pond Road, East Boothbay, Maine 04544, 2018.

[30] M. DEP. (2021) 2020 plan. [Online]. Available: https://www.cascobayestuary.org/wp-content/uploads/2022/03/Maine-DEP-Water-Quality-Eelgrass-Brewer-2-17-22.pdf

[31] G. Page, S. Matey, Y. Johnston, M. Schulte, L. Hayden, and S. Knag, "Bioregional learning journey launches seagrass conservation in gulf of maine: A case study of governance response to ecosystem change," *Social Innovations Journal*, vol. 15, 2022.

[32] Statista, "Number of volunteers in the united states from 2008 to 2017," 2020, [Online; accessed January 24, 2023]. [Online]. Available: https://www.statista.com/statistics/189297/number-of-volunteers-in-the-united-states-since-2003/

[33] M. Haklay, "Citizen science and volunteered geographic information: Overview and typology of participation," *Crowdsourcing geographic knowledge*, pp. 105–122, 2013.

[34] H. E. Roy, M. J. Pocock, C. D. Preston, D. B. Roy, J. Savage, J. Tweddle, and L. Robinson, "Understanding citizen science and environmental monitoring: final report on behalf of uk environmental observation framework," 2012.

[35] Open Source Org., "The open source defenition," 2006, [Online; accessed January 5, 2023]. [Online]. Available: https://web.archive.org/web/20070611152544/https://opensource.org/docs/osd

[36] Suber, Peter, "Open access overview," 2007, [Online; accessed January 24, 2023]. [Online]. Available: http://legacy.earlham.edu/~peters/fos/overview.htm

[37] Dvergsdal, Henrik, "crowdsourcing," 2019, [Online; accessed January 24, 2023]. [Online]. Available: https://snl.no/crowdsourcing

[38] A. F. Sveen, A. S. S. Erichsen, and T. Midtbø, "Micro-tasking as a method for human assessment and quality control in a geospatial data import," *Cartography and Geographic Information Science*, vol. 47, no. 2, pp. 141–152, 2020. [Online]. Available: https://doi.org/10.1080/15230406.2019.1659187

[39] K. L. Wilson, M. C. Wong, and E. Devred, "Comparing sentinel-2 and worldview-3 imagery for coastal bottom habitat mapping in atlantic canada," *Remote Sensing*, vol. 14, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/5/1254

[40] K. Eva, R. Christiaan, L. Mitchell, Z. Shihu, and P. Stuart, "Seagrass habitat mapping: how do landsat 8 oli, sentinel-2, zy-3a, and worldview-3 perform?" *Remote Sensing Letters*, pp. 686–695, 2018.

[41] L. Yang, J. Driscol, S. Sarigai, Q. Wu, H. Chen, and C. D. Lippitt, "Google earth engine and artificial intelligence (ai): A comprehensive review," *Remote Sensing*, vol. 14, no. 14, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/14/3253

[42] L. J. McKenzie, L. A. Langlois, and C. M. Roelfsema, "Improving approaches to mapping seagrass within the great barrier reef: From field to spaceborne earth observation," *Remote Sensing*, vol. 14, no. 11, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/11/2604

[43] M. Wyatt, B. Radford, N. Callow, M. Bennamoun, and S. Hickey, "Using ensemble methods to improve the robustness of deep learning for image classification in marine environments," *Methods in Ecology and Evolution*, vol. 13, no. 6, pp. 1317–1328, 2022. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13841

[44] J. Fethers, "Remote sensing of eelgrass using object based image analysis and sentinel-2 imagery," Ph.D. dissertation, University of Aalborg Aalborg, Denmark, 2018.

[45] F. Immordino, M. Barsanti, E. Candigliota, S. Cocito, I. Delbono, and A. Peirano, "Application of sentinel-2 multispectral data for habitat mapping of pacific islands: Palau republic (micronesia, pacific ocean)," *Journal of Marine Science and Engineering*, vol. 7, no. 9, 2019. [Online]. Available: https://www.mdpi.com/2077-1312/7/9/316

[46] A. Mora-Soto, M. Palacios, E. C. Macaya, I. Gómez, P. Huovinen, A. Pérez-Matus, M. Young, N. Golding, M. Toro, M. Yaqub *et al.*, "A high-resolution global map of giant kelp (macrocystis pyrifera) forests and intertidal green algae (ulvophyceae) with sentinel-2 imagery," *Remote Sensing*, vol. 12, no. 4, p. 694, 2020.

[47] D. Poursanidis, D. Traganos, L. Teixeira, A. Shapiro, and L. Muaves, "Cloud-native seascape mapping of mozambique's quirimbas national park with sentinel-2," *Remote Sensing in Ecology and Conservation*, vol. 7, no. 2, pp. 275–291, 2021.

[48] K. L. Wilson, M. C. Wong, and E. Devred, "Branching algorithm to identify bottom habitat in the optically complex coastal waters of atlantic canada using sentinel-2 satellite imagery," *Frontiers in Environmental Science*, vol. 8, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2020.579856

[49] C. B. Lee, "Development of a semi-analytical model for seagrass mapping using cloud-based computing and open sourced optical satellite data," Ph.D. dissertation, Julius-Maximilians-Universität Würzburg, 2020.

[50] F. Uhl, T. Græsdal Rasmussen, and N. Oppelt, "Classification ensembles for beach cast and drifting vegetation mapping with sentinel-2 and planetscope," *Geosciences*, vol. 12, no. 1, p. 15, 2022.

[51] C. B. Lee, D. Traganos, and P. Reinartz, "A full cloud-native dive into bioregional-scale seagrass mapping in the mediterranean using sentinel-2 multitemporal data," 2021.

[52] M. C. León-Pérez, A. S. Reisinger, and J. C. Gibeaut, "Spatial-temporal dynamics of decaying stages of pelagic sargassum spp. along shorelines in puerto rico using google earth engine," *Marine Pollution Bulletin*, vol. 188, p. 114715, 2023.

[53] L. Pascal, "Revisit and coverage," https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage.

[54] Q. Wu and K. Markert. (2021) Geopython 2021. [Online]. Available: https://geemap.org/workshops/GeoPython_2021/?h=geopython

[55] A. Chawla, A. Kumar, A. Warghat, S. Singh, S. Bhushan, R. K. Sharma, A. Bhattacharya, and S. Kumar, "Approaches for conservation and improvement of himalayan plant genetic resources," in *Advancement in crop improvement techniques*. Elsevier, 2020, pp. 297–317.

[56] D. Pierce and P. DiBona, "Eelgrass monitoring: Development of a citizen scientist monitoring method," 2018.

[57] T. KOMATSU, C. IGARARASHI, K.-I. TATSUKAWA, M. NAKAOKA, T. HIRAISHI, and A. TAIRA, "Mapping of seagrass and seaweed beds using hydro-acoustic methods," *Fisheries science*, vol. 68, no. sup1, pp. 580–583, 2002.

[58] USGS. How often is orthoimagery in the national map updated and what are the acquisition dates? [Online]. Available: https://www.usgs.gov/faqs/how-often-orthoimagery-national-map-updated-and-what-are-acquisition-dates

[59] M. C. Program, "Assessing various methods to map the health and extent of eelgrass beds," Maine Outdoor Heritage Fund, 2019.

[60] C. Li, J. Wang, L. Wang, L. Hu, and P. Gong, "Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery," *Remote sensing*, vol. 6, no. 2, pp. 964–983, 2014.

[61] M. Hossain, J. Bujang, M. Zakaria, and M. Hashim, "The application of remote sensing to seagrass ecosystems: an overview and future research prospects," *International Journal of Remote Sensing*, vol. 36, no. 1, pp. 61–114, 2015.

[62] European Space Agency. About worldview-2. [Online]. Available: https://earth.esa.int/eogateway/missions/worldview-2

[63] E. S. Agency, *Sentinel-2 User Handbook*. Luxembourg: Publications Office of the European Union, 2015. [Online]. Available: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

[64] J. A. Richards, *Error Correction and Registration of Image Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, pp. 39–74. [Online]. Available: https://doi.org/10.1007/978-3-642-88087-2_2

[65] S. M. E. team, "Data Quality Report Sentinel-2 MSI L2A," Tech. Rep., 2023.

[66] S. Selvaraj, B. S. Case, and W. L. White, "Effects of location and season on seaweed spectral signatures," *Frontiers in Ecology and Evolution*, vol. 9, p. 581852, 2021.

[67] E. S. A. (ESA). (2023) Level-2. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-2

[68] J. R. Jensen *et al.*, *Introductory digital image processing: a remote sensing perspective.* Prentice-Hall Inc., 1996, no. Ed. 2.

[69] E. Insider. (2016) What is orthorectified imagery? [Online]. Available: https://www.esri.com/about/newsroom/insider/what-is-orthorectified-imagery/

[70] SUHET, "Sentinel-2 user handbook," ESA, 2013.

[71] T. Harmel, M. Chami, T. Tormos, N. Reynaud, and P.-A. Danis, "Sunglint correction of the multi-spectral instrument (msi)-sentinel-2 imagery over inland and sea waters from swir bands," *Remote Sensing of Environment*, vol. 204, pp. 308–321, 2018.

[72] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 65, no. 1, pp. 2–16, 2010.

[73] Q. Wu, "2.07 - gis and remote sensing applications in wetland mapping and monitoring," in *Comprehensive Geographic Information Systems*, B. Huang, Ed. Oxford: Elsevier, 2018, pp. 140–157. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124095489104609

[74] T. Yamakita, F. Sodeyama, N. Whanpetch, K. Watanabe, and M. Nakaoka, "Application of deep learning techniques for determining the spatial extent and classification of seagrass beds, trang, thailand," *Botanica Marina*, vol. 62, pp. 291–307, 05 2019.

[75] R. Moore. (2010) Introducing google earth engine. [Online]. Available: https://blog.google/outreach-initiatives/sustainability/introducing-google-earth-engine/

[76] K. L. Wilson, M. A. Skinner, and H. K. Lotze, "Eelgrass (zostera marina) and benthic habitat mapping in atlantic canada using high-resolution spot 6/7 satellite imagery," *Estuarine, Coastal and Shelf Science*, vol. 226, p. 106292, 2019.

[77] J. Barrell, J. Grant, A. Hanson, and M. Mahoney, "Evaluating the complementarity of acoustic and satellite remote sensing for seagrass landscape mapping," *International Journal of Remote Sensing*, vol. 36, no. 16, pp. 4069–4094, 2015.

[78] E. Bisong and at al., *Building machine learning and deep learning models on Google cloud platform.* Apress, 2019.

[79] J. A. Bullinaria. (2005) The roots, goals and sub-fields of ai. [Online]. Available: https://www.cs.bham.ac.uk/~jxb/IAI/w2.pdf

[80] J. McCarthy, "What is artificial intelligence?" 2007.

[81] S. Raschka and V. Mirjalili, *Python Machine Learning Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*, 12 2019.

[82] B. Ferreira, R. G. Silva, and M. Iten, "Earth observation satellite imagery information based decision support using machine learning," *Remote Sensing*, vol. 14, no. 15, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/15/3776

[83] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817, 2018.

[84] P. Wicaksono, "Improving the accuracy of multispectral-based benthic habitats mapping using image rotations: the application of principle component analysis and independent component analysis," *European Journal of Remote Sensing*, vol. 49, no. 1, pp. 433–463, 2016. [Online]. Available: https://doi.org/10.5721/EuJRS20164924

[85] P. Wicaksono, P. A. Aryaguna, and W. Lazuardi, "Benthic habitat mapping model and cross validation using machine-learning classification algorithms," *Remote Sensing*, vol. 11, no. 11, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/11/1279

[86] B. Tolga and G. Mustafa Umit, "Assessment of machine learning methods for seagrass classification in the mediterranean," *Baltic Journal of modern computing*, vol. 8, pp. 315–326, 2012. [Online]. Available: https://avesis.yildiz.edu.tr/yayin/2337aa6f-31fc-4ef7-bf08-2b768cebd93e/assessment-of-machine-learning-methods-for-seagrass-classification-in-the-mediterranean

[87] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science," *Supervised and unsupervised learning for data science*, pp. 3–21, 2020.

[88] N. Grira, M. Crucianu, and N. Boujemaa, "Unsupervised and semi-supervised clustering: a brief survey," *A review of machine learning techniques for processing multimedia content*, vol. 1, pp. 9–16, 2004.

[89] T. Duda and M. Canty, "Unsupervised classification of satellite imagery: choosing a good algorithm," *International Journal of Remote Sensing*, vol. 23, no. 11, pp. 2193–2212, 2002.

[90] K. A. Islam, V. Hill, B. Schaeffer, R. Zimmerman, and J. Li, "Semi-supervised adversarial domain adaptation for seagrass detection using multispectral images in coastal areas," *Data Science and Engineering*, vol. 5, pp. 111–125, 2020.

[91] S. R. Phinn, C. M. Roelfsema, and P. J. Mumby, "Benthic cover map of Heron Reef derived from a high-spatial-resolution multi-spectral satellite image using object based image analysis," 2012, supplement to: Phinn, SR et al. (2012): Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. International Journal of Remote Sensing, 33(12), 3768-3797, https://doi.org/10.1080/01431161.2011.633122. [Online]. Available: https://doi.org/10.1594/PANGAEA.789968

[92] M. Hasmadi, H. Pakhriazad, and M. Shahrin, "Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data," *Geografia: Malaysian Journal of Society and Space*, vol. 5, no. 1, pp. 1–10, 2009.

[93] A. P. St-Pierre and P. Gagnon, "Kelp-bed dynamics across scales: Enhancing mapping capability with remote sensing and gis," *Journal of Experimental Marine Biology and Ecology*, vol. 522, p. 151246, 2020.

[94] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.

[95] X. J. Zhu, "Semi-supervised learning literature survey," 2005.

[96] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[97] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *Advances in neural information processing systems*, vol. 32, 2019.

[98] Y. He, J. Wang, C. Liao, B. Shan, and X. Zhou, "Classhyper: Classmix-based hybrid perturbations for deep semi-supervised semantic segmentation of remote sensing imagery," *Remote Sensing*, vol. 14, no. 4, p. 879, 2022.

[99] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[100] A. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," *Advances in neural information processing systems*, vol. 31, 2018.

[101] E. B. Alexakis and C. Armenakis, "Evaluation of semi-supervised learning for cnn-based change detection," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 829–836, 2021.

[102] K. A. Islam, "Deep learning approaches for seagrass detection in multispectral imagery," Ph.D. dissertation, Old Dominion University, 2021.

[103] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[104] P. Ganesh. Knowledge distillation : Simplified. Medium. [Online]. Available: https://towardsdatascience.com/knowledge-distillation-simplified-dd4973dbc764

[105] C. Zhang, D. Selch, Z. Xie, C. Roberts, H. Cooper, and G. Chen, "Object-based benthic habitat mapping in the florida keys from hyperspectral imagery," *Estuarine, Coastal and Shelf Science*, vol. 134, pp. 88–97, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0272771413004253

[106] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, R. Tibshirani, and J. Friedman, "Random forests," *The elements of statistical learning: Data mining, inference, and prediction*, pp. 587–604, 2009.

[107] D. Traganos, D. Poursanidis, N. Chrysoulakis, and P. Reinartz, "Mediterranean-wide seagrass mapping using sentinel-2 on google earth engine."

[108] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern recognition letters*, vol. 27, no. 4, pp. 294–300, 2006.

[109] H. Xie, X. Luo, X. Xu, H. Pan, and X. Tong, "Evaluation of landsat 8 oli imagery for unsupervised inland water extraction," *International Journal of Remote Sensing*, vol. 37, no. 8, pp. 1826–1844, 2016.

[110] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[111] R. Achanta and S. Susstrunk, "Superpixels and polygons using simple non-iterative clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[112] G. Varoquaux and O. Colliot, "Evaluating machine learning models and their diagnostic value," 2022.

[113] M. AI, "Understanding the metrics for better evaluation of your ai models," 2021.

[114] Z. Ma and R. L. Redmond, "Tau coefficients for accuracy assessment of classification of remote sensing data," *Photogrammetric Engineering and Remote Sensing*, vol. 61, no. 4, pp. 435–439, 1995.

[115] S. Allwright. What is a good auc score? (simply explained). [Online]. Available: https://stephenallwright.com/good-auc-score/

[116] J. Li. Precision - recall. [Online]. Available: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

[117] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[118] J. Lynam. (2021) Mainedep seagrass 2021 (south coast - elliot to cape elizabeth). [Online]. Available: https://maine.hub.arcgis.com/datasets/4cae3da79ed74b5ca78b8eeeac967d68_2/about

[119] MarineWeather.net. (2022) Portland, casco bay, me tides. [Online]. Available: https://marineweather.net/tide/portland-casco-bay

[120] t. E. Casco Bay Estuary Partnership, "State of casco bay," Casco Bay Estuary Partnership, University of Southern Maine, p. 52, 2021.

[121] H. A. Neckles, "Loss of Eelgrass in Casco Bay, Maine, Linked to Green Crab Disturbance," *Northeastern Naturalist*, vol. 22, no. 3, pp. 478 – 500, 2015. [Online]. Available: https://doi.org/10.1656/045.022.0305

[122] A. J. Pershing, M. A. Alexander, C. M. Hernandez, L. A. Kerr, A. Le Bris, K. E. Mills, J. A. Nye, N. R. Record, H. A. Scannell, J. D. Scott *et al.*, "Slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery," *Science*, vol. 350, no. 6262, pp. 809–812, 2015.

[123] M. S. Grant and M. D. of Marine Resources. (2013) Rockweed ecology, industry and management. [Online]. Available: https://www.maine.gov/dmr/sites/maine.gov.dmr/files/docs/brochure01-11-13.pdf

[124] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2cor for sentinel-2," in *Image and Signal Processing for Remote Sensing XXIII*, vol. 10427. SPIE, 2017, pp. 37–48.

[125] Harmonized sentinel-2 msi: Multispectral instrument, level-2a. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED

[126] radouxju, "Sentinel-2 reflectance calculation," 2017, [Online; accessed May 5, 2023]. [Online]. Available: https://gis.stackexchange.com/a/247323

[127] T. Kutser, E. Vahtmäe, and J. Praks, "A sun glint correction method for hyperspectral imagery containing areas with non-negligible water leaving nir signal," *Remote Sensing of Environment*, vol. 113, no. 10, pp. 2267–2274, 2009.

[128] B. Saltzman, "Satellite oceanic remote sensing; advances in geophysics. volume 27," 1985.

[129] J. Hedley, A. Harborne, and P. Mumby, "Simple and robust removal of sun glint for mapping shallow-water benthos," *International Journal of Remote Sensing*, vol. 26, no. 10, pp. 2107–2112, 2005.

[130] geografif. (2023) Sunglint correction (deglint) in google earth engine. [Online]. Available: https://www.youtube.com/watch?v=_V0X1LdSta4

[131] S. Carpenter, "Cme mapping workshop," 2023. [Online]. Available: https://code.earthengine.google.com/?accept_repo=users%2Fstcarp%2FCME_MappingWorksho

[132] N. M. Short, *The Landsat tutorial workbook: Basics of satellite remote sensing*. National Aeronautics and Space Administration, Scientific and Technical . . . , 1982, vol. 1078.

[133] J. Lynam, "MaineDEP - Eelgrass 2018 (Casco Bay Only)," https://hub.arcgis.com/datasets/maine::mainedep-eelgrass-2018-casco-bay-only/about.

[134] C. H. Halsted. Maine coastal marine geologic environments. [Online]. Available: https://mgs-maine.opendata.arcgis.com/datasets/maine::maine-coastal-marine-geologic-environments/about

[135] J. Jena, S. Misra, and K. Tripathi, "Normalized difference vegetation index (ndvi) and its role in agriculture," 11 2019.

[136] S. K. Jain and V. P. Singh, *Water resources systems planning and management*. Elsevier, 2003, vol. 1.

[137] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[138] M. Rehman and V. E., ""duplicate record detection for database cleansing."," vol. Second International Conference on Machine Vision. IEEE, 2009.

[139] Yale University, “”negative reflectance values”,” 2010, [Online; accessed May 5th, 2023]. [Online]. Available: https://surfaceheat.sites.yale.edu/sites/default/files/files/Negative%20Reflectance%20Values.pdf

[140] R. S. Paulino, V. S. Martins, E. M. Novo, C. C. Barbosa, L. A. de Carvalho, and F. N. Begliomini, “Assessment of adjacency correction over inland waters using sentinel-2 msi images,” *Remote Sensing*, vol. 14, no. 8, p. 1829, 2022.

[141] P. Probst, M. N. Wright, and A.-L. Boulesteix, “Hyperparameters and tuning strategies for random forest,” *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, vol. 9, no. 3, p. e1301, 2019.

[142] P. Wicaksono, P. A. Aryaguna, and W. Lazuardi, “Benthic habitat mapping model and cross validation using machine-learning classification algorithms,” *Remote Sensing*, vol. 11, no. 11, p. 1279, 2019.

[143] Y. Liu, W. Gong, X. Hu, and J. Gong, “Forest type identification with random forest using sentinel-1a, sentinel-2a, multi-temporal landsat-8 and dem data,” *Remote Sensing*, vol. 10, no. 6, p. 946, 2018.

[144] scikit-learn developers, “3.1. cross-validation: evaluating estimator performance,” 2022.

[145] S. M. Weiss and C. A. Kulikowski, *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. Morgan Kaufmann Publishers Inc., 1991.

[146] B. Mikulski. (2023) Pca-how to choose the number of components? [Online]. Available: https://www.mikulskibartosz.name/pca-how-to-choose-the-number-of-components/

[147] S. Soheily-Khah, P.-F. Marteau, and N. Béchet, “Intrusion detection in network systems through hybrid supervised and unsupervised machine learning process: A case study on the iscx dataset,” in *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 2018, pp. 219–226.

[148] M. Aamir and S. M. A. Zaidi, “Clustering based semi-supervised machine learning for ddos attack classification,” *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 4, pp. 436–446, 2021.

[149] Mapseagrass. (2020) Accuracy assesment. [Online]. Available: https://mapseagrass.org/index.php/methodologies-and-tools/accuracy-assessment

[150] P. M, *Computer processing of remotley sensed images: An introduciton.* Biddley Limited Publication,UK., 1991.

[151] S. H. Kim, J. Kim, S. Park, and K.-S. Lee, "Annual and perennial life history strategies of zostera marina populations under different light regimes," *Marine Ecology Progress Series*, vol. 509, pp. 1–13, 08 2014.

[152] P. Mumby, E. Green, A. Edwards, and C. Clark, "Coral reef habitat mapping: how much detail can remote sensing provide?" *Marine Biology*, vol. 130, pp. 193–202, 1997.

[153] D. Liu and F. Xia, "Assessing object-based classification: advantages and limitations," *Remote sensing letters*, vol. 1, no. 4, pp. 187–194, 2010.

[154] L. J. McKenzie, L. M. Nordlund, B. L. Jones, L. C. Cullen-Unsworth, C. Roelfsema, and R. K. F. Unsworth, "The global distribution of seagrass meadows," *Environmental Research Letters*, vol. 15, no. 7, p. 074041, 07 2020. [Online]. Available: https://dx.doi.org/10.1088/1748-9326/ab7d06

[155] Q. Wu, "geemap: A python package for interactive mapping with google earth engine," *Journal of Open Source Software*, vol. 5, no. 51, p. 2305, 2020.

[156] G. E. Engine. (2023) Google earth engine. [Online]. Available: https://earthengine.google.com/

[157] S. Knag, "Seagrass vs seaweed detection trail mt 2023," 4 2023. [Online]. Available: https://github.com/SigridKnag/Master-Thesis-2023/blob/41d43ec29a4a981542de3c94de187532f963027f/MasterThesis_2023_SK.ipynb

[158] F. of Casco Bay. Continuous monitoring stations. [Online]. Available: https://www.cascobay.org/our-work/science/continuous-monitoring-stations/

[159] D. Stream. Turbidity. [Online]. Available: https://datastream.org/en/guide/turbidity

[160] S. Barker, "Eelgrass distribution in casco bay 2018," 2018.

[161] A. C. Mathieson, E. J. Hehre, C. J. Dawes, and C. D. Neefus, "AN HISTORICAL COMPARISON OF SEAWEED POPULATIONS FROM CASCO BAY, MAINE," *Rhodora*, vol. 110, no. 941, pp. 1 – 102, 2008. [Online]. Available: https://doi.org/10.3119/06-23.1

[162] Seaweed observations in casco bay. ASCO (Assessing Seaweed via Community Observations). [Online]. Available: https://shorturl.at/lBM49

[163] M. Haklay, "How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets," *Environment and planning B: Planning and design*, vol. 37, no. 4, pp. 682–703, 2010.

[164] J.-F. Girres and G. Touya, "Quality assessment of the french openstreetmap dataset," *Transactions in GIS*, vol. 14, no. 4, pp. 435–459, 2010.

[165] T. K. Davies, G. Stevens, M. G. Meekan, J. Struve, and J. M. Rowcliffe, "Can citizen science monitor whale-shark aggregations? investigating bias in mark–recapture modelling using identification photographs sourced from the public," *Wildlife Research*, vol. 39, no. 8, pp. 696–704, 2012.

[166] M. V. Hoyer, N. Wellendorf, R. Frydenborg, D. Bartlett, and D. E. Canfield Jr, "A comparison between professionally (florida department of environmental protection) and volunteer (florida lakewatch) collected trophic state chemistry data in florida," *Lake and Reservoir Management*, vol. 28, no. 4, pp. 277–281, 2012.

[167] J. Gollan, L. L. de Bruyn, N. Reid, and L. Wilkie, "Can volunteers collect data that are comparable to professional scientists? a study of variables used in monitoring the outcomes of ecosystem rehabilitation," *Environmental management*, vol. 50, no. 5, pp. 969–978, 2012.

[168] M. M. Gardiner, L. L. Allee, P. M. Brown, J. E. Losey, H. E. Roy, and R. R. Smyth, "Lessons from lady beetles: accuracy of monitoring data from us and uk citizen-science programs," *Frontiers in Ecology and the Environment*, vol. 10, no. 9, pp. 471–476, 2012.

[169] T. Snäll, O. Kindvall, J. Nilsson, and T. Pärt, "Evaluating citizen-based presence data for bird monitoring," *Biological conservation*, vol. 144, no. 2, pp. 804–810, 2011.

[170] G. Kaplan, "Sentinel-2 pan sharpening—comparative analysis," *Proceedings*, vol. 2, no. 7, 2018. [Online]. Available: https://www.mdpi.com/2504-3900/2/7/345

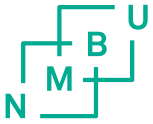
[171] A. A. Smadi, S. Yang, A. Mehmood, A. Abugabah, M. Wang, and M. Bashir, "Smart pansharpening approach using kernel-based image filtering," *IET Image Processing*, vol. 15, no. 11, pp. 2629–2642, 2021. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ipr2.12251

[172] G. Kaplan and U. Avdan. (2018) Sentinel-2 pan sharpening comparative analysis. [Online]. Available: https://sciforum.net/manuscripts/5158/slides.pdf

[173] P. Wicaksono, "Improving the accuracy of multispectral-based benthic habitats mapping using image rotations: the application of principle component analysis and

independent component analysis," *European Journal of Remote Sensing*, vol. 49, no. 1, pp. 433–463, 2016. [Online]. Available: https://doi.org/10.5721/EuJRS20164924

[174] T. D. Pham, J. Xia, N. T. Ha, D. T. Bui, N. N. Le, and W. Tekeuchi, "A review of remote sensing approaches for monitoring blue carbon ecosystems: Mangroves, seagrasses and salt marshes during 2010–2018," *Sensors*, vol. 19, no. 8, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/8/1933



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway