TÉCNICO LISBOA

Automatic Detection of Floating Marine Debris Using Multi-spectral Satellite Imagery

Miguel Mendes Duarte

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor: Prof. Leonardo Azevedo Guerra Raposo Pereira

Examination Committee

Chairperson: Prof. João Manuel de Freitas Xavier
Supervisor: Prof. Leonardo Azevedo Guerra Raposo Pereira
Members of the Committee: Prof. Maria Margarida Campos da Silveira
Dr. Renato Paulo dos Santos Mendes

June 2022

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

This dissertation would not have been possible without the contribution of some special people.

I would like to express my gratitude to my supervisor, Professor Leonardo Azevedo. Firstly, for allowing me to take part in Project SMART, the first edition's winner of the AI Moonshot Challenge, promoted by the Portuguese Space Agency. It was a pleasure to meet and present my work to many talented people and be involved in the same project as them. Then, for his availability, guidance and precise suggestions and tips.

I profoundly thank my parents, Rui and Odete, for supplying me with the tools needed to chase my goals. I'll be forever grateful for the unconditional support and encouragement throughout the years. A special thanks to Joana for always making me feel confident in my capabilities and for the incredible support. I would also like to express my deepest gratitude to my grandparents, Maria, António and Bertilde, for being always available, and to my sister, Mariana, for the funny moments. This thesis is also yours and I hope you are proud of it.

Lastly, I would like to acknowledge all my friends, colleagues, and professors that helped me grow as a person over the years.

Abstract

Marine plastic pollution represents a maritime environmental emergency that needs to be addressed. Floating plastic debris must be detected, captured, and removed from the ocean, in order to preserve such a fragile ecosystem. In this work, it is shown that floating plastic debris are not only detectable but also distinguishable from other floating materials, such as driftwood, seaweed, sea snot, sea foam, and pumice, in optical data from the European Space Agency (ESA) Sentinel-2 satellites, using a supervised learning method trained with data compiled from published works and complemented by some manual interpretation of satellite images. The proposed model, an Extreme Gradient Boosting (XGBoost) trained with seven spectral indices and two spectral bands, successfully classified 98% of the pixels that contained floating plastic debris in coastal waters. Additionally, due to the need for more floating plastic data in the training dataset, synthetic data were generated through a Wasserstein Generative Adversarial Network (WGAN). A supervised model trained only with synthetic data successfully classified plastic pixels with an accuracy of 91%. Finally, to build a system that provides reliable results when applied in real-world conditions, an ensemble model that quantifies uncertainty was created. This novel approach correctly classified 79% of the plastic pixels. However, the number of misclassifications decreased significantly compared to the model with the highest accuracy, making it the best option to monitor the ocean.

Keywords

Marine Pollution; Floating Plastic Debris; Sentinel-2; Remote Sensing.

Resumo

A poluição dos mares por plástico representa uma emergência ecológica e é um problema que tem de ser abordado e resolvido brevemente. Detritos de plástico flutuantes têm de ser detetados, recolhidos e removidos do oceano, de modo a preservar este frágil ecossistema. Este trabalho demonstra que plásticos são não só detetáveis, como também distinguíveis de outros materiais flutuantes, tais como madeira, algas, muco marinho, espuma e pedra-pomes, em imagens dos satélites Sentinel-2, da Agência Espacial Europeia (ESA), através de um método de aprendizagem supervisionada. O modelo aqui proposto, um Extreme Gradient Boosting (XGBoost), treinado com sete índices e duas bandas espectrais, classificou corretamente 98% dos pixéis que continham plástico. Além disso, devido à escassez de dados de plástico flutuante, foram gerados dados sintéticos a partir de uma Rede Adversária Generativa do tipo Wasserstein (WGAN). Um modelo supervisionado treinado apenas com dados sintéticos classificou corretamente pixéis de plástico com uma precisão de 91%. Por fim, de modo a conceber um sistema que forneça resultados confiáveis quando aplicado em condições do mundo real, foi criado um modelo que quantifica a incerteza nas previsões. Esta nova abordagem classificou acertadamente 79% dos pixéis de plástico. Contudo, o número de previsões erradas diminuiu consideravelmente quando comparado com o modelo que apresenta a maior precisão, o que torna esta alternativa a melhor para monitorizar o oceano.

Palavras Chave

Poluição Marinha; Detritos de Plástico Flutuantes; Sentinel-2; Deteção Remota.

Contents

List of Figures

List of Tables

Acronyms

ACOLITE Atmospheric Correction for OLI 'lite'

ANN Artificial Neural Network

AOI Area of Interest

ARI Anthocyanin Reflectance Index

AWEI Automated Water Extraction Index

Adam Adaptive Moment Estimation

C2RCC Case 2 Regional Coast Colour

CHL Red-Edge Chlorophyll Red-Edge Index

CNN Convolutional Neural Network

DSF Dark Spectrum Fitting

ESA European Space Agency

EVI2 Enhanced Vegetation Index 2

EVI Enhanced Vegetation Index

EXP Exponential Extrapolation

FAI Floating Algae Index

FDI Floating Debris Index

GAN Generative Adversarial Network

GNDVI Green Normalised Difference Vegetation Index

iCOR Image correction for atmospheric effects

KNN K-nearest Neighbour

Leaky ReLU Leaky Rectified Linear Unit

MARI Modified Anthocyanin Reflectance Index

MCARI Modified Chlorophyll Absorption Reflectance Index

MI Moisture Index

MNDWI Modified Normalised Difference Water Index

MSI Multi-Spectral Instrument

NBR Normalised Burn Ratio

NDMI Normalised Difference Moisture Index

NDSI Normalised Difference Snow Index

NDVI Normalised Difference Vegetation Index

NDWI Normalised Difference Water Index

NIR Near-infrared

OBIA Object Based Image Analysis

OSI Oil Spill Index

PET Polyethylene Terephthalate

PI Plastic Index

PLP Plastic Litter Project

PNDVI Pan Normalised Difference Vegetation Index

REPI Red Edge Position Index

RMSprop Root Mean Square Propagation

RNDVI Reversed Normalised Difference Vegetation Index

RVI Ratio Vegetation Index

SAM Spectral Angle Mapper

SAVI Soil-Adjusted Vegetation Index

SGD Stochastic Gradient Descent

SNAP Sentinel Application Platform

SR Simple Ratio

SST Sea Surface Temperature

SVM Support Vector Machine

SWIR Short-Wave InfraRed

SWI Sentinel-2 Water Index

Sen2Cor Sentinel 2 Correction

Tanh	Hyperbolic Tangent
UAV	Unmanned Aerial Vehicle
WGAN	Wasserstein Generative Adversarial Network
WRI	Water Ratio Index
WSI	”Windrows As Proxies” Spectral Index
XGBoost	Extreme Gradient Boosting

1

Introduction

Contents

The ocean makes up one of the most important Earth's natural resources. It covers about 71% of the Earth's surface and holds 97% of the world's water. Oceans control the weather and climate since they absorb most of the sun's radiation and distribute the heat around the globe through their currents, compensating for the uneven distribution of radiation throughout the Earth's surface. That is why the system of ocean currents is known as Global Conveyor Belt. However, this "conveyor belt" is not only transporting heat but also plastic debris, representing a threat to marine ecosystems, which include not merely sea creatures but also entire countries whose economies rely on healthy coastal and ocean resources.

1.1 Motivation

The first fully synthetic plastic, i.e., containing no molecule found in nature, results from Belgian chemist Leo Baekeland's work, in 1907. This new material, Bakelite, proved to be durable, heat resistant and ideal for mass production, which led major chemical companies to invest in the research and development of new polymers. Its benefits quickly became obvious because of their versatile, inexpensive, strong, corrosion-resistant, durable, with high thermal and electrical insulation properties [6]. It was not until the 1940s and 1950s, however, that mass production of everyday plastic items began [7], marking the beginning of the "Plastic Era" [8]. Since then, global plastic production has been steadily increasing, reaching 380 million tonnes produced only in 2015, which represents around 190 times the value in 1950 [9]. The largest market sector for plastic resins is packaging [10], which means that most of these products are designed for immediate disposal. Thus, plastic makes up a significant percentage of all solid waste generated and, since none of the commonly used plastics are biodegradable and only a small portion may be recycled or incinerated, they accumulate, rather than decompose, in landfills or the natural environment [11]. In 2015, an estimated 20% of global plastic waste was recycled, 25% was incinerated and the remaining 55% was discarded [9], meaning that in many countries where waste management infrastructure is lacking, plastic waste enters water bodies.

Approximately 65% of the synthetic polymers produced have a lower density than seawater [9]. Thus, because of their durability, these buoyant objects accumulate on the ocean's surface and travel worldwide through currents. The first reports of plastic pollution in the oceans are from the early 1970s [12]. Since then, hundreds of articles showing plastics on shorelines of remote islands, beaches and even in arctic sea have been published, but the most well-known proof of substantial plastic accumulations in the ocean is in the North Pacific Gyre. The so-called Great Pacific Garbage Patch is estimated to comprise almost 79000 tonnes of plastic [11], including not only macroplastics ($>$ 5 mm), such as abandoned fishing nets, bottles, and containers but also microplastics ($<$ 5 mm), which usually result from the fragmentation of larger plastic items, because of physical mechanisms, namely weathering because

of UV radiation, mechanical (wave action), thermal and chemical action [7].

Plastic litter affects marine ecosystems in multiple ways. Its most visible effect on marine life is the entanglement of organisms, such as birds, turtles, mammals, and fish, in marine debris. Entanglement often results in death by drowning, suffocation, or strangulation. If not instantly fatal, it causes injuries and wounds, leading the animal to starvation through reduced feeding efficiency and making it difficult to escape predators [8]. Many marine organisms mistake plastic for food and ingest it. Ingestion of plastic can cause lacerations in the digestive system, can cause a false sense of satiety leading to poor physical development, and its retention in the digestive system has potential negative consequences for reproduction and growth [7]. Since animals carry these debris in their bodies, plastic is already part of the food chain, and it might affect human health. The durability and buoyancy of plastics also allow the transportation of species to ecosystems where they are not native [7], so the open ocean is no longer a physical and biological barrier to the dispersal of coastal marine species, as they can survive and reproduce for years, leading to self-sustaining coastal communities on the high seas [13]. For example, in the six years following the 2011 East Japan tsunami, at least 289 coastal marine species have travelled to the United States on massive rafts of pollution [14]. This phenomenon, plastic rafting, can reduce habitats for native species and carry diseases, threatening biological diversity. Finally, marine plastics present a range of negative economic impacts. A study estimated that the economic costs of marine plastic, as related to marine natural capital, are conservatively conjectured at between $3300 and $33,000 per tonne of marine plastic per year, based on 2011 ecosystem service values and marine plastic stocks [15]. It also makes beaches aesthetically unappealing to users, which reduces the number of visitors, resulting in tourism revenue losses, as several studies have documented. After a period of heavy rainfall in July 2011, a large amount of marine debris washed up on the beaches of Geoje Island, South Korea, leading to a 63% reduction in visitors, representing a revenue loss of around US$29 - 37 million [16]. A similar study carried out at two Brazilian subtropical beaches estimated losses up to US$8.5 million per year [17]. In terms of fisheries, plastic litter can reduce catches and damage vessels.

The priority to reduce the amount of plastic pollution in the ocean should be to stop plastic waste from entering the hydrosphere. The industry could design products that avoid unnecessary plastic usage or allow reusability and recyclability. Governments could tax non-recyclable products to discourage their purchase and should cooperate globally to regulate the primary sources of plastics, namely, industrial, and domestic products [7]. Raising community awareness on the consequences of plastic pollution could empower more people and organizations to take action. Finally, improving waste management facilities, mainly in low-income countries, would be the measure with the most beneficial effects. However, even if the world stopped generating plastic waste, macroplastics would persist on the ocean's surface for many more years. Therefore, these plastics must be detected, captured, and removed from the oceans. The first step, detection, provides an understanding of the quantity of plastic in a specified area and

how it drifts through time, enabling the design of policies to tackle the problem. There are three main techniques to locate and monitor floating plastic. The first one is through numerical models or ocean particle tracking models. These models are based on ocean currents and provide knowledge of how plastics enter the ocean, how they spread and for how long they have been there. The second technique is in situ observations and refers to samples or measurements collected at various locations that can validate models' results. For example, a study conducted at Vigo Ria, using a particle tracking model, examined the distribution and accumulation areas of microplastics released by wastewater treatment plants, and validated the results using in situ observations [4]. Lastly, the use of earth observation data to detect marine plastic debris accumulations is a recent area of research, but some early studies show promising results. Satellites provide a reliable source of data thanks to their efficiency to cover extensive areas over time without human interaction and their cost-effectiveness.

The motivation previously identified exposes the importance of identifying plastic accumulations in the ocean using accurate and effective methods. This work investigates the combination of machine learning and satellite imagery to detect floating plastic debris in coastal waters.

1.2 Objectives

The main challenge addressed in this work is to create an automatic method that can detect floating debris on satellite images. However, these debris can be made of materials with a similar reflectance to floating plastics. For example, a recent study shows that remote differentiation of sea snots and marine plastic debris using satellites may be difficult because of their spectral similarity [18]. Hence, the first aim of this thesis is to create a machine learning model that not only detects floating debris, such as plastic and sea snot but also can distinguish them with a high degree of certainty. Therefore, this work differs from the previous studies since its goal is to differentiate plastic from water and from debris that show similar spectral reflectance, such as sea snot, sea foam, and pumice.

Data are the most significant part of a machine learning model since they allow the training and testing of models. Therefore, to have reliable results, it is crucial to collect enough high quality data. However, satellite data of floating plastics are scarce, as almost every study in this area highlights [19, 20]. Thus, the second objective of this thesis is to create a dataset of multiple classes of floating materials and make it available for everyone, which will hopefully lead to considerable progress in this area. The third goal of this work is to investigate if artificially expanding the size of the training set by creating synthetic data can improve the results. This technique is called data augmentation and its use is suitable when the initial dataset is too small. Finally, the last objective is to apply the model that achieved the best results in the testing phase to some areas during a reasonable time frame to monitor the evolution of the floating debris.

1.3 Outline

This thesis is is organized as follows: in chapter 2, a review of the state-of-the-art models that detect floating plastics in satellite imagery is conducted. Chapter 3 discusses the reasons for the choice of the Sentinel-2 satellite from the Copernicus Programme. Additionally, it focuses on the steps taken to get the data and process it. Furthermore, it displays how spectral indices, which are mathematical combinations of spectral reflectance from two or more wavelengths, can provide additional variables of interest. It also introduces the supervised learning algorithms and the tools to perform data augmentation used under the scope of this thesis. Finally, it assesses if applying an uncertainty quantification method improves the model's results by decreasing the number of misclassifications. Chapter 4 reveals the results obtained using different models and the effects of using synthetic data. Furthermore, it shows the application of the best model to certain areas and reveals its weaknesses. Chapter 5 is the final chapter, where conclusions referring to this thesis are outlined, together with some topics for future work. Appendix A presents a case study where the best model was applied in combination with a particle-tracking model and appendix B shows the results of the several classifiers tested in this thesis.

2

Literature Review

In the past few years, research on plastic detection and monitoring using data from ship-based visual surveys [21], Unmanned Aerial Vehicle (UAV) [22], numerical models [4], and cameras deployed at beaches [23] have revealed promising results. Despite the relative success of these methods, they do not provide an option for monitoring larger spatial scales [24]. Data from ship-based visual surveys and numerical models have inherent uncertainty because of the different survey, model, or sampling methods used. UAVs are limited in terms of flight time and payload capabilities, as well as in terms of automation potential [25]. Although having an appropriate spatial resolution, fixed cameras in coastal areas have a narrow spatial range and aerial surveys using occupied aircraft, despite solving the spatial range problem, are bounded by the acquisition frequency and are exceedingly costly. Therefore, the use of satellites is the method with the most potential in this area, even though being affected by some physical and technical limitations, namely cloud interference, atmospheric and sea-surface effects, and the instrument's spatial resolution. This chapter reviews relevant studies regarding the detection of floating plastic debris utilizing satellite imagery. The conclusions from each study reveal the rapid evolution in this area: from using satellite imagery to identify oceanic features that can lead to the detection of litter, to applying machine learning and spectral indices to distinguish different floating marine debris.

Picher et al. [26] performed a marine debris survey, in the Gulf of Alaska, in the summer of 2003, using satellite imagery to locate convergent areas where ghost nets are likely to accumulate, followed by aircraft observations to spot individual nets. In the first place, areas of probable debris accumulation were selected by researching historical studies and analysing wind and current information. Then, those areas were monitored using satellite imagery to observe oceanic features indicative of ocean convergence. These features were identified using Sea Surface Temperature (SST), chlorophyll and altimeter products, from six different satellites, mostly non-open access. Finally, crewed instrumented aircraft flew over the areas of ocean convergence and observed individual debris objects and documented the general distribution of debris. Despite not observing debris using satellite products, this study shows that satellites can identify oceanic features leading to ghost net detection.

Aoyama [27] used high-resolution satellite images to develop methods to identify marine debris drifting in the Sea of Japan (East Sea), in 2012. Initially, 20 pixels were chosen randomly from an Area of Interest (AOI) of a true colour image from the WorldView-2 satellite in a coastal area of the Tsuruga Peninsula. Despite the spatial resolution of 2 m, marine debris could not be confirmed in the AOI. However, by plotting the spectral response of the 20 pixels using the blue, the green, the red, and the Near-infrared (NIR) spectral bands of the WorldView-2 satellite, several pixels with anomalous spectral characteristics were observed, so they were likely to contain an anthropogenic materials. After that, by using 2D scatter diagrams and then Spectral Angle Mapper (SAM), which is a classification method for directly comparing image spectra to a known spectrum (average spectrum of pixels in the AOI), several suspected marine debris pixels were extracted from the 9207 pixels in the AOI. To validate these results,

a WorldView-3 satellite image of a fixed fishery net off the coast of the Tsuruga Peninsula was used. Scatter diagrams and the SAM method appear to classify successfully buoy pixels but with many false positives. Accuracy was not specified.

Topouzelis et al. [28] created the Plastic Litter Project (PLP) 2018 to explore the feasibility of detecting plastics in the aquatic environment using data acquired from UAVs and the open-access Copernicus Sentinel-2 satellites. Three ten-by-ten-meter targets made of plastic bottles, plastic bags and plastic fishing nets were created and deployed at least 30 meters away from the coast of Tsamakia beach of Mytilene on Lesvos Island, Greece. Very high-resolution images were generated from aerial survey using multiple sensors integrated into the UAV and, in all of them except the thermal image, the targets were brighter than water. The three targets were also detected in the Sentinel-2 true colour image after performing the atmospheric correction using Atmospheric Correction for OLI 'lite' (ACOLITE). Direct comparison of the UAV data with the Sentinel-2 satellite image allowed a percentage calculation of plastic coverage in the satellite image, which led to the conclusion that spectral reflectance of floating plastic positively correlates with the percentage pixel coverage of each target. It also enabled the identification of spectral differences between plastics and the surrounding water pixels. The following year, Topouzelis et al. [19] published a second study on detecting artificial floating plastic targets - PLP 2019. This edition's goal was to better simulate near-real conditions and examine the limitations of identifying floating plastic using Sentinel-2 satellite images. Two new types of targets were created with plastic bottles, plastic bags, and natural debris: six five-by-five-meter size targets and two with a one-by-five-meter size, which allowed to make several configurations and simulate varied sizes and shapes. UAV data was used together with Sentinel-2 data, enabling the calculation of the percentage coverage of plastic litter in a pixel through Object Based Image Analysis (OBIA) and image overlay. Reverse spectral unmixing was used to get the spectral signature of the plastic targets made of Polyethylene Terephthalate (PET), which was then used to perform a matched filtering processing on the satellite images to classify the pixels containing concentrations of plastic litter. Depending on the matched filtering thresholds, marine litter can be detected with a PET abundance fraction of at least as low as 25%, however, there were also many false positives. The causes for false positives were also identified: clouds, shadows, vessels, fumes, sun glint and bottom reflectance on the coastline. In 2020, the Marine Remote Sensing Group from the University of the Aegean worked in the Plastic Litter Project 2020 towards the creation of large reference targets to provide data for the scientific community. In 2021, during the PLP 2021, those targets were deployed in Gera Gulf, Lesvos Island, Greece. Despite no studies published yet related to these last two PLPs, all information is available online [29], which represents a substantial source of marine plastic debris data.

Similarly to the PLPs, Themistocleous et al. [30] investigated the detection of floating plastic litter using a Sentinel-2 image and UAVs data of a three-by-ten-meter artificial target made of water bottles

placed in the sea near Old Port in Limassol, Cyprus. Seven spectral indices were examined and two new were developed: the Plastic Index (PI) and the Reversed Normalised Difference Vegetation Index (RNDVI). The study found that the plastic target was easier to detect in the NIR wavelengths and the PI was the most effective in identifying the plastic target. However, when the PI was applied to the coast of Limassol, several false positives were reported, mainly related to boats with plastic surfaces.

Fronkova [24] assessed a method for detecting the artificial plastic targets deployed in the PLP 2018 [28] using the difference in reflectance between bands 8 and 12 of the Sentinel-2. The Normalised Difference Water Index (NDWI) was used for land masking. The targets were detected successfully, but with many misclassifications (the number of false positives was not quantified in the work).

Kikaki et al. [31] investigated the capability of satellite sensors in detecting marine plastic debris over the Bay Islands and Gulf of Honduras between 2014 and 2019. Multiple sources of satellite data were used, namely Sentinel-2, Landsat-8, and Planet Labs. In situ data were collected through vessels and diving expeditions. The detection of plastic litter was performed manually by comparing the spectral signatures of the pixels with the ones reported in the literature and based on the debris size, pattern, and their association with specific river discharges. It is not possible to assess the accuracy of this method. However, it provides data that future studies can use and validate. The study highlights the need for automated machine learning algorithms capable of plastic pollution detection.

Biermann et al. [20] studied the detection of floating plastic debris using Sentinel-2 imagery and assessed the capability to distinguish plastic from other floating debris, such as timber and seaweed. To collect floating macroplastics data, the authors searched for published studies, news articles and even social media posts, which allowed them to gather 53 Sentinel-2 pixels of floating plastic. Atmospheric correction was performed using the ACOLITE open-access software. Two spectral indices were used: the Normalised Difference Vegetation Index (NDVI) and the newly developed Floating Debris Index (FDI), which is a modification of the Floating Algae Index (FAI). When FDI and NDVI were examined together, all the floating materials studied (seawater, seaweed, timber, plastic, sea foam and pumice) showed distinct clustering. This allowed them to use the Naïve Bayes classifier, where the spectral signatures and the two indices were used as features. Finally, the model was tested using data from PLP 2018 [28] and PLP 2019 [19] and it correctly classified plastic with an accuracy of 86%, whereas 3% of plastics were classified as seawater and 11% as sea foam. This study proves that the spatial and spectral resolution of Sentinel-2 is sufficient for macroplastics to be distinguishable from natural sources of floating debris and seawater itself when using the FDI together with the NDVI.

The previous study boosted floating plastic debris research. Multiple scientific reports published afterwards use machine learning algorithms along with the NDVI and the FDI. Basu et al. [32] used two supervised and two unsupervised classification algorithms to detect floating plastic in coastal waters. Five Sentinel-2 images from the experimental campaigns in Greece [19, 28] and Cyprus [30] were con-

sidered to calibrate the supervised models and estimate model efficiency, which resulted in 59 pixels with floating plastics. From the available Sentinel-2 spectral bands after the atmospheric correction, a combination of six bands and the NDVI and FDI were selected to develop the models. The supervised classification outperformed the unsupervised clustering algorithms. The best model had an accuracy of 96.7%. On the other hand, Ciappa [33] compared the spectral signature of unknown floating material found offshore Hawaii (presumed plastic) with known types of oceanic patches, specifically "sargassum" macroalgae and mixed vegetation materials. His study concluded that Sentinel-2 pixels containing floating vegetation did not show distinct clustering from the ones with plastic, using both the NDVI and the FDI.

Mifdal et al. [34] employed a deep learning predictor (a Convolutional Neural Network (CNN)) that learned the spatial characteristics of floating objects from data manually selected from multiple coastal areas using the FDI and NDVI indices. The model could detect the geometrical shapes of floating objects, however, with a high number of false positives (not quantified in the work), which led to the conclusion that the CNN architecture is not suitable for a stand-alone detection of floating objects. Arias et al. [35], using 40 Sentinel-2 images, aimed at prototyping an operational processor capable of identifying filaments of floating marine debris but without performing atmospheric correction. An index that uses spectral bands removed in the atmospheric correction step was developed: "Windrows As Proxies" Spectral Index (WSI). Using the novel index, they detected accumulations of debris of land origin, which are a mix of wood, dried leaves, vegetation, and actual litter. Nevertheless, it was not possible to distinguish the different materials.

Despite the differences, every study highlights the need for more plastic data collected globally. The models that showed the best results rely on supervised classification methods, which are highly dependent on the supplied training samples.

3

Data and Methodology

Contents

3.1 Summary

This section summarizes the necessary steps needed to accomplish the objectives established in chapter 1. The first task is to check the news, scientific reports, or even social media posts to collect in situ data (section 3.4). Secondly, the satellite images need to be pre-processed to remove the contribution of the atmosphere from the reflectance measured by the multi-spectral instrument, and a land mask needs to be applied to remove pixels that are not relevant for the study (section 3.3). The next step is to ensure that the data that will be used to train the classification models is balanced, i.e., contain an almost equal number of pixels from each class and represents the actual spectral signatures. This can be done by undersampling the most represented classes in the training set or by generating synthetic data (section 3.6). Then, combining the reflectance values of different spectral bands by using spectral indices may help in separating the different floating classes (section 3.5). Classification models will use these features in the training process and, after tuning the hyperparameters and testing the models, the one with the best results can be applied to the real world (section 3.7). However, in order to remove uncertain estimates that deliver false information regarding the floating material in a pixel, an uncertainty quantification method will also be used, improving the information given by the model and helping the decision-making process (section 3.8). The flowchart in Figure 3.1 outlines all these steps.

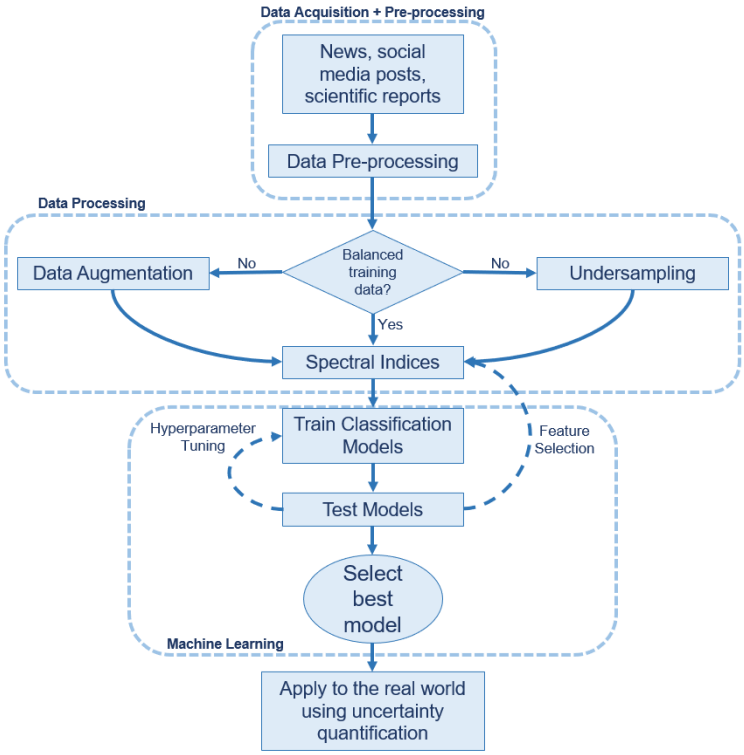

Figure 3.1: Flowchart with the steps adopted to detect plastic and other floating debris on the ocean using satellite imagery.

3.2 Satellite

This work uses freely available satellite data products from the Sentinel-2 mission, which comprises a constellation of two identical satellites, Sentinel-2A and Sentinel-2B, developed and operated by the European Space Agency (ESA) under the Copernicus Programme. It provides systematic coverage (5 days at the equator and 2 to 3 days at mid-latitudes) over all coastal waters up to 20 km from the shore. Each satellite has a Multi-Spectral Instrument (MSI) aboard that works passively (i.e., it measures the sunlight reflected from the Earth). Its optical data is of high spatial resolution (10 m, 20 m, or 60 m, depending on the spectral band). To detect floating plastic, the most crucial feature of the instrument besides the spatial resolution is its radiometric resolution, which is the instrument's capacity to distinguish differences in light intensity or reflectance and typically ranges from 8 to 16 bits. The greater the total number of discrete signals that the sensor can record (spectral bands), the greater the radiometric resolution and the more accurate the image is. Sentinel-2 has 13 spectral bands that range from the visible and NIR to the Short-Wave InfraRed (SWIR), shown in Table 3.1, allowing for a 12-bit radiometric resolution and enabling the image to be acquired over a range of 0 to 4095 potential light intensity values [36]. All these features, as well as being used by most studies whose goal is to detect and monitor floating debris, make Sentinel-2 a preferential option for acquiring multi-spectral floating plastic data.

Table 3.1: Sentinel-2 spectral bands, their central wavelengths, bandwidths, and spatial resolutions. Table adapted from [5].

Spectral Band	Sentinel-2A		Sentinel-2B		Spatial Resolution (m)	
	Central wavelength, λ (nm)	Bandwidth (nm)	Central wavelength, λ (nm)	Bandwidth (nm)		
B1	442.7	21	442.2	21	60	
B2	492.4	66	492.1	66	10	
B3	559.8	36	559.0	36	10	Visible
B4	664.6	31	664.9	31	10	
B5	704.1	15	703.8	16	20	
B6	740.5	15	739.1	15	20	
B7	782.8	20	779.7	20	20	
B8	832.8	106	832.9	106	10	NIR
B8A	864.7	21	864.0	22	20	
B9	945.1	20	943.2	21	60	
B10	1373.5	31	1376.9	30	60	
B11	1613.7	91	1610.4	94	20	SWIR
B12	2202.4	175	2185.7	185	20	

3.3 Data Pre-processing

Unlike UAVs data, where the atmospheric effects are not considered because of the negligible path from the sensor to the observation sensor, satellite images require atmospheric correction before being used to extract data and information. Atmospheric correction methods aim at removing the contribution of the atmosphere from the reflectance measurements taken by the satellite sensor [37] that can influence and change the results [38]. The atmospheric correction of coastal waters is challenged by the presence of continental aerosols, bottom reflectance, and adjacency of land [39], which significantly raises the water's reflectance. Land mask helps to remove pixels that the proximity to the coast may spoil and, at the same time, removes the land pixels, which are not essential for ocean studies, decreasing the computational power needed to process the data. The following section reveals the atmospheric correction method used in this work and the land mask techniques considered.

3.3.1 Atmospheric Correction

ESA makes Sentinel-2 products available for free as Top of Atmosphere reflectance values in cartographic geometry (Level-1C product) [40]. To obtain Bottom of Atmosphere reflectance images (Level-2A product) based on the Level-1C product, various atmospheric correction algorithms can be used such as Image correction for atmospheric effects (iCOR) [41], Case 2 Regional Coast Colour (C2RCC) [42], Sentinel 2 Correction (Sen2Cor) [43], or ACOLITE [44]. Topouzelis et al. [28] demonstrated that both ACOLITE and Sen2Cor are suitable for the atmospheric correction of products used to detect floating plastics. Recently, besides Level-1C products, ESA started distributing Level-2A products processed by the Sen2Cor algorithm. However, this method has been found ineffective and performed poorly over the open ocean and coastal waters [32], so it has been recommended not to use it for sea and open ocean areas [45]. Therefore, this work uses ACOLITE v.20210802.0 [38] to atmospherically correct all the Sentinel-2 images.

ACOLITE performs atmospheric correction by using the Dark Spectrum Fitting (DSF) algorithm [46–50] by default, but it can also be configured to use the Exponential Extrapolation (EXP) algorithm [51–53]. Overall, the DSF achieves better results than the EXP algorithm, especially in the blue spectral region [54]. This method assumes that the atmosphere is homogeneous (i.e., the atmospheric reflectance is constant within the image), and the scene contains pixels with zero or very close to zero surface reflectance in at least one of the sensor bands (dark pixels). The spectral signature of the dark pixels, or dark spectrum, is then used to determine the best fitting combination of the spectral band and aerosol model for the atmospheric correction. With the most appropriate combination selected, the parameters required for the "path-corrected" reflectance computation are then chosen from a look-up table. Due to low atmospheric transmittance, band 9 (B9) and band 10 (B10) are excluded from the outputs. As all the

atmospheric correction methods, DSF shows some limitations. For example, since it selects the darkest pixels, the ones affected by sun glint are avoided, and the glint signal will still be present in the resulting surface reflectance.

3.3.2 Land Mask

Land masking is a necessary step in hydrological satellite studies since it removes unnecessary pixels that could be mistaken for floating materials and reduces the computational power needed to process an entire image. The open-source Sentinel Application Platform (SNAP) 8.0 software [55], used to visualize the satellite data, provides an automatic land mask that uses the reflectance values of the selected spectral bands (land pixels have higher reflectance than water pixels). However, as can be seen in Figure 3.2, it has some limitations and does not mask all land pixels. Another alternative is to manually draw polygon regions to mask out the land.

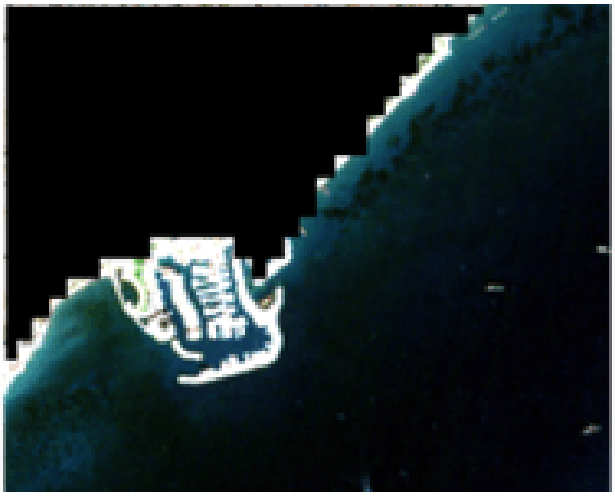

Figure 3.2: Land mask (in black) based on the reflectance values of all spectral bands, using SNAP software in Sentinel-2A image of port of Limassol, Cyprus, on the 15^{th} of December 2018.

McFeeters [56] proposed the NDWI, which is a mathematical formula that combines the third and eighth Sentinel-2 spectral bands to delineate open water features and enhance their presence in remotely sensed digital imagery. The NDWI is dimensionless and varies between -1 and 1, depending on the quantity of water in the pixel. Therefore, by setting a threshold it is possible to differentiate water bodies from land and vessels, as in Figure 3.3. Jiang et al. [57] also proposed a spectral index for performing land masking, the Sentinel-2 Water Index (SWI), that combines the fifth and the eleventh Sentinel-2 spectral bands. Both indices were compared in several Sentinel-2 images of coastal waters and, overall, the NDWI method showed better results, so it was the chosen method for this work. However, in some cases, it identified floating natural debris as a non-water body and, although the same never happened with plastic pixels, it is something to watch out for.

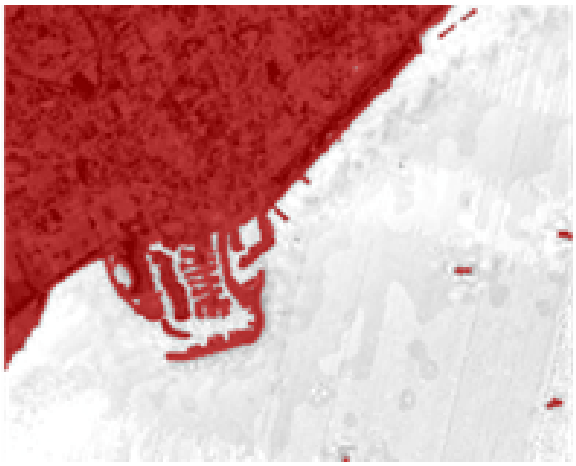

Figure 3.3: Land mask (in red) using the NDWI method, with threshold = 0.1, in Sentinel-2A image of port of Limassol, Cyprus, on the 15^{th} of December 2018.

3.4 Data Acquisition

The first step towards creating a machine learning model that detects plastic floating in the ocean is to gather data. Unlike most previous studies that only collect water and plastic data, this work focuses on distinguishing plastic from other floating materials with similar spectral signatures. Thus, data from seven classes were collected. All the data used are confirmed by scientific reports, news articles, or social media posts (in situ data), followed by a manual inspection of the spectral responses. Both Sentinel-2A and Sentinel-2B were used, and their data are freely available through the Copernicus Open Access Hub [58]. Despite the satellites' spectral bands having slightly different central wavelengths, the spectral signatures' figures use the same central wavelengths to facilitate interpretation. Satellite images were visualized using the ESA open-source SNAP 8.0 software [55] and Jupyter Notebook was used to process the data, plotting graphs, and implementing algorithms.

3.4.1 Water

Every day, the Sentinel-2 constellation gathers millions of pixels of ocean water. In this work, 150 pixels of ocean water were collected equally from two distinct locations: the Caribbean Sea and the Gulf of Gera, in fifteen different days. From the 150 pixels, 121 are from the Sentinel-2A and the remaining from the Sentinel-2B. Also, 25% of the water data, corresponding to 30 pixels, are from waters where the bottom of the ocean is visible, resulting in brighter pixels. This suggests that different water depths match different spectral reflectance. Reflectance is the ratio between the amount of light leaving the pixels to the amount of light striking them. Waters closer to land usually have shallower depth, so the ocean floor reflects sunlight, in contrast to waters far from the shore, where most of the light is absorbed. Figure 3.4 confirms the assumption since shallower waters exhibit higher reflectance in all spectral bands.

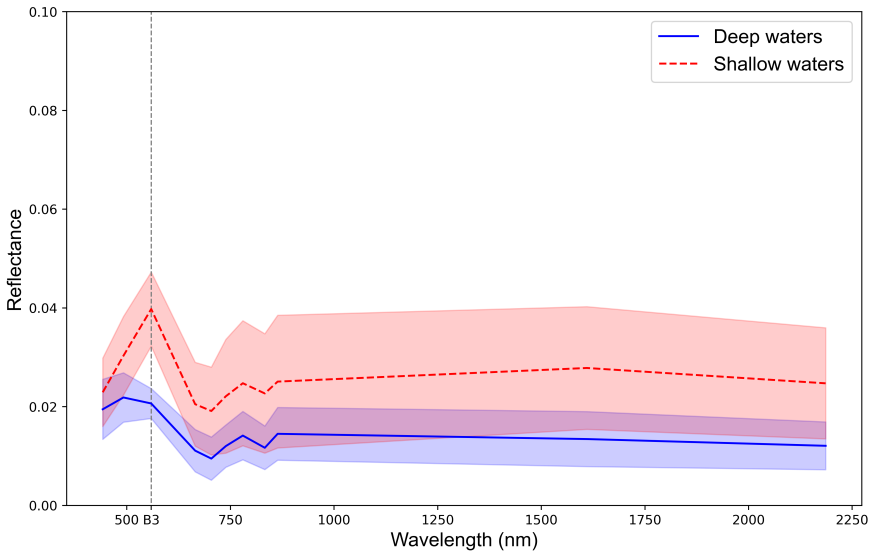

Figure 3.4: Spectral reflectance (mean - line, and standard deviation - shaded area) of Sentinel-2 water pixels of different depths after atmospheric correction.

Overall, the spectral signatures are identical except in the third band, where shallow depth water reflectance doubles the one from the deep water. By comparing these spectral signatures with the ones in [20], it is safe to assume that the reflectance of shallower waters is not high enough for it to be mistaken for another material, and it is not so different from the deeper waters' reflectance. Therefore, there is no need to create two distinct categories and all the data are grouped into a single class (water) with a spectral signature shown in Figure 3.5.

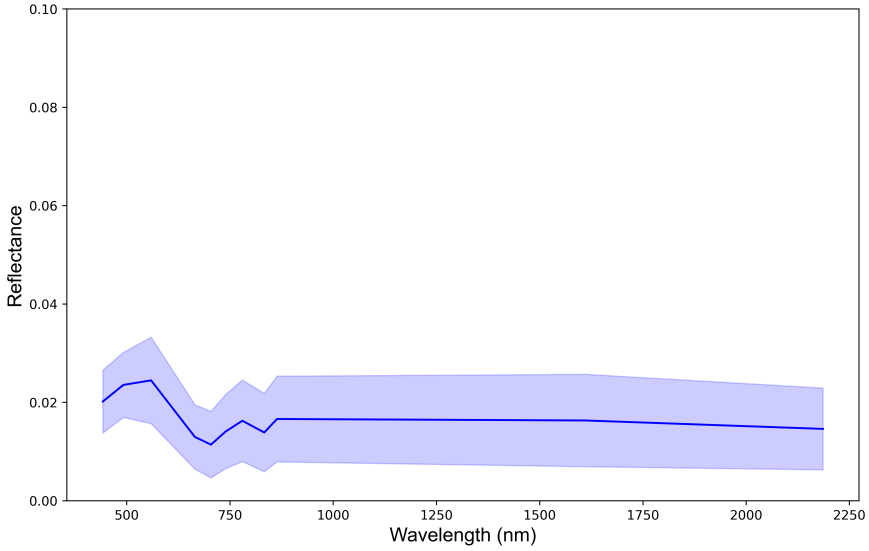

Figure 3.5: Spectral reflectance (mean - line, and standard deviation - shaded area) of all water pixels after atmospheric correction.

3.4.2 Plastic

As discussed in the previous chapter, floating plastic data are scarce. This work gathered 206 pixels of plastic that are confirmed by scientific reports, news articles or pictures on social media posts (in situ data). Every pixel's spectral response was manually inspected and compared to the expected spectral signature in the literature [20,59], and the ones that did not meet the requirements were rejected. Therefore, many pixels identified as suspected plastic in scientific reports were discarded. From the 206 pixels, 102 were taken from Sentinel-2A images and 107 from Sentinel-2B imagery. Around 42% of the data, corresponding to 88 pixels, are from artificial plastic targets deployed in the ocean in the Gulf of Gera [29] (Figure 3.6), Tsamakia beach [19, 28], and Limassol [30]. The remaining 58% result from observations and reports of plastic floating in the marine environment. On the 23^{rd} of April 2019, substantial quantities of plastic covered the Durban harbour, in South Africa, after a flood event [60]. The debris eventually washed out to the sea, and a Sentinel-2 image from the following day allowed the detection of 72 pixels with spectral reflectance similar to plastic. The remaining pixels result from the work of Kikaki et al. [31] and their observations over the Bay Islands and Gulf of Honduras.

Figure 3.6: Artificial plastic target deployed on the 10^{th} of August 2021 for the PLP 2021 [1].

The plastic spectral signature stemming from the data reveals the differences to the water pixels. From Figure 3.7, it is noticeable that plastic shows two reflectance peaks, one centred at the third spectral band and the other at the eighth, and one absorption peak centred at the fifth Sentinel-2 spectral band. It is also clear that plastic has higher reflectance values in all spectral bands compared to the water spectral signature.

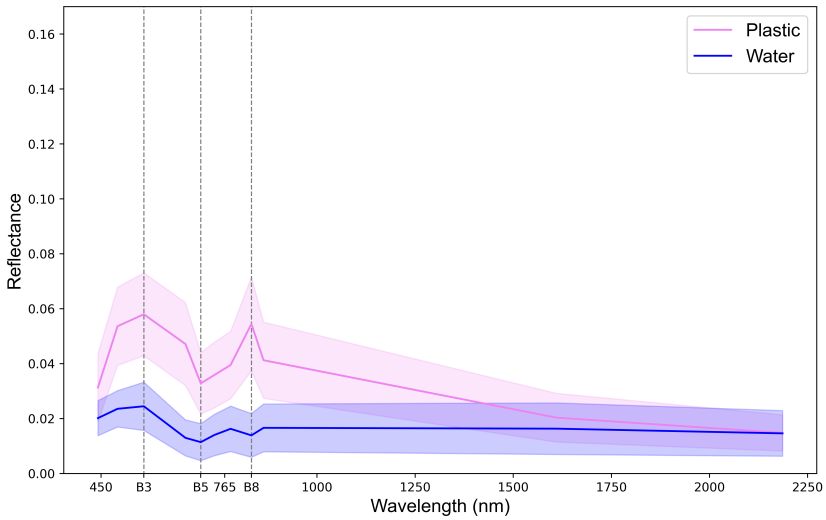

Figure 3.7: Spectral reflectance (mean - line, and standard deviation - shaded area) of plastic pixels after atmospheric correction.

3.4.3 Driftwood

Driftwood is wood that has been washed into the ocean through the action of natural occurrences such as winds or flooding, or because of logging. Its study is of great interest in many research fields. For example, in geomorphology, knowing the accumulation rates of wood in rivers may help in creating measures concerning the maintenance of watercourses and assist in risk management [61]. However, it is challenging to find these pixels in Sentinel-2 images since it is not common to exist significant accumulations of driftwood.

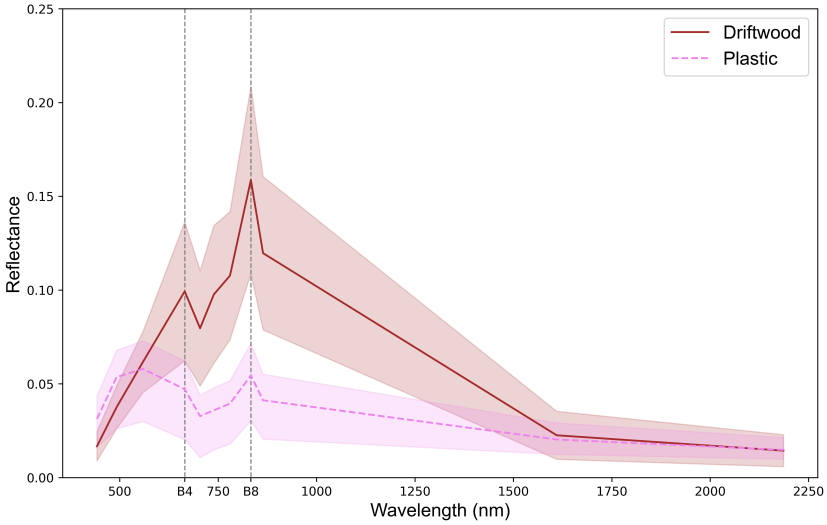

Figure 3.8: Spectral reflectance (mean - line, and standard deviation - shaded area) of pixels of driftwood after atmospheric correction.

PLP 2021 [29] allowed the collection of 62 pixels of driftwood on thirteen different days since they deployed a wooden target that simulates natural driftwood. Around 55% of these pixels were taken from Sentinel-2A images and the remaining from Sentinel-2B. Driftwood shows substantially more reflectance when compared to water or plastic (Figure 3.8), and it has two reflectance peaks in the fourth and eighth Sentinel-2 spectral bands.

3.4.4 Seaweed

Seaweed is the common name for countless species of marine plants and algae that grow in the ocean [62]. There are several types of seaweed, but the most prevalent is the Sargassum or brown algae, which floats in large masses and even inspired the name of a region in the Atlantic Ocean, the Sargasso Sea. Its presence in the ocean is essential since it provides nutrients and shelter for many marine organisms, but too much seaweed can be harmful. Substantial accumulations of seaweed may block sunlight, preventing the seagrass below from growing and, when decomposing, its organic matter removes oxygen from the water. This work does not focus on differentiating the distinct species of seaweed, as its goal is to discriminate floating debris, and considerable variations in the various seaweed reflectance are not expected. One Sentinel-2B image from October 2018 was used to collect 150 pixels of seaweed in the coastal waters of Accra, Ghana. The seaweed's spectral signature (Figure 3.9) coincides with the literature [20, 59] since it presents a sharp increase in reflectance in the fourth Sentinel-2 spectral band, followed by a fall in the band 8A, being very distinct from the spectral responses of water, plastic and driftwood. Also, the standard deviation reveals that there is not much dispersion in the data relative to the mean before the reflection peak.

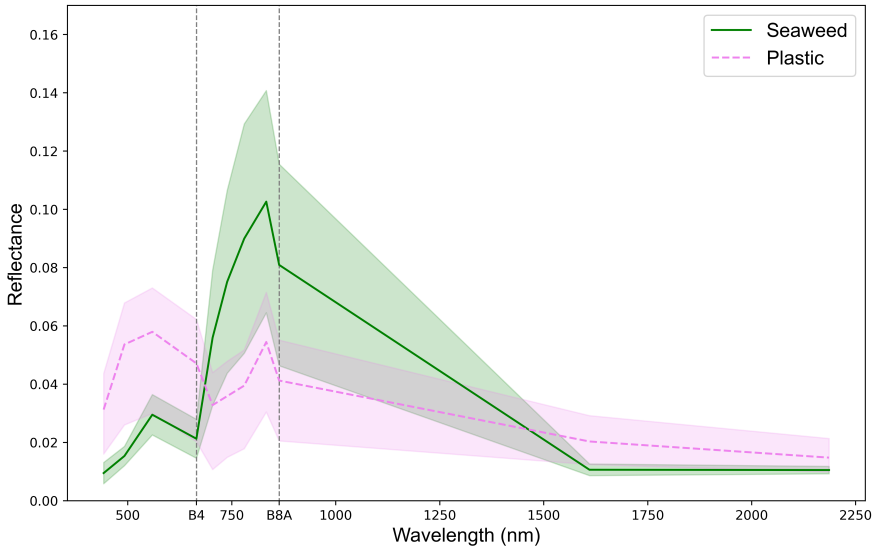

Figure 3.9: Spectral reflectance (mean - line, and standard deviation - shaded area) of pixels of seaweed after atmospheric correction.

3.4.5 Pumice

Pumice is a light-coloured volcanic rock with a foamy appearance. It is formed when super-heated and highly pressurized molten rock, magma, is powerfully ejected from a volcano and rapidly cools down, which commonly happens in underwater eruptions. Pumice is so light that it may float on water for years, potentially forming gigantic floating islands - pumice rafts (Figure 3.10). These pumice rafts are considered a danger to navigation since they can cause damage to cargo vessels and mess up with radar signals. For example, tankers carry thousands of tonnes of oil and, if damaged from a collision, can provoke a massive environmental disaster. Hence, information on the location and course of pumice rafts can be valuable for the shipping industry.

Figure 3.10: A cleaning vessel removes pumice stones off the coast of Nago, Okinawa [2].

In October 2021, a large underwater volcanic eruption spewed massive amounts of floating pumice stones that littered coastlines in Okinawa, Japan, damaging dozens of fishing vessels and forcing a large percentage to remain stuck at ports. A Sentinel-2A image from 26 October 2021 reveals thousands of bright pixels containing floating pumice stone and was used to collect 31098 pixels of this floating material. Pumice's reflectance values (Figure 3.11) are close to the plastic mean spectral signature. However, plastic presents an absorption peak in the fifth spectral band, which does not happen with pumice. Pumice's standard deviation reveals a lot of dispersion relative to the mean reflectance, which may be a consequence of different floating depths in different pixels. Pixels where pumice floats on the ocean's surface will have higher reflectance values than ones where pumice is slightly submerged.

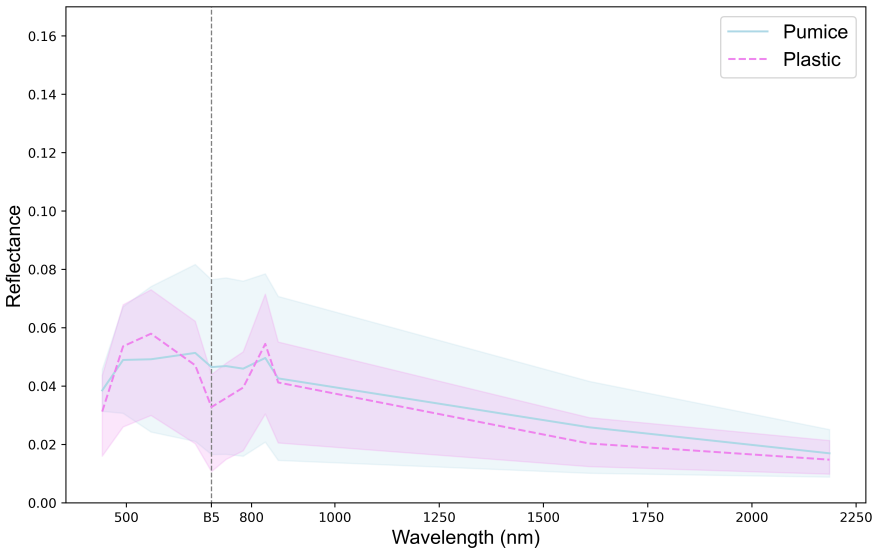

Figure 3.11: Spectral reflectance (mean - line, and standard deviation - shaded area) of pixels of pumice and plastic after atmospheric correction.

3.4.6 Sea Snot

Marine mucilage, also known as sea snot, is a thick slimy organic substance that floats on the ocean. It forms when algae are overloaded with nutrients because of global warming and water pollution that results from industrial waste dumped into the seas. Warmer and slower-moving waters also increase the production of sea snot and allow its accumulation. Marine mucilage surge poses severe threats to public health since it contains bacteria, transports diseases, and has adverse economic and environmental consequences. This substance harshly affects the fishing industry as it clogs fishing nets, removes oxygen from the water and limits sunlight from reaching marine ecosystems, killing sea creatures. There are several reports of sea snot outbreaks in the last few years, however, none of them reaches the level of the one in the Marmara Sea, Turkey, in 2021 (Figure 3.12). Short-term countermeasures include laying barriers on the sea surface and collecting the substance. On the other hand, long-term countermeasures comprise improving wastewater treatment and imposing fines on companies that dump industrial waste in the ocean.

Figure 3.12: Sea snot patches floating on the Marmara Sea in Bursa, Turkey, on the 11^{th} of June 2021 [3].

One Sentinel-2B image from the Marmara Sea, on the 6^{th} of June 2021, showed thousands of pixels containing sea snot. From those, 26403 pixels were selected. By examining Figure 3.13, it is clear why Hu et al. [18] concluded that remote differentiation of sea snots and marine debris using multi-band sensors is problematic. The two classes show a similar mean spectral reflectance, except in the eighth Sentinel-2 spectral band, where plastics have a reflectance peak. In contrast to plastic, sea snot's standard deviation reveals a lot of dispersion, which probably results from different floating depths in different pixels, just like pumice.

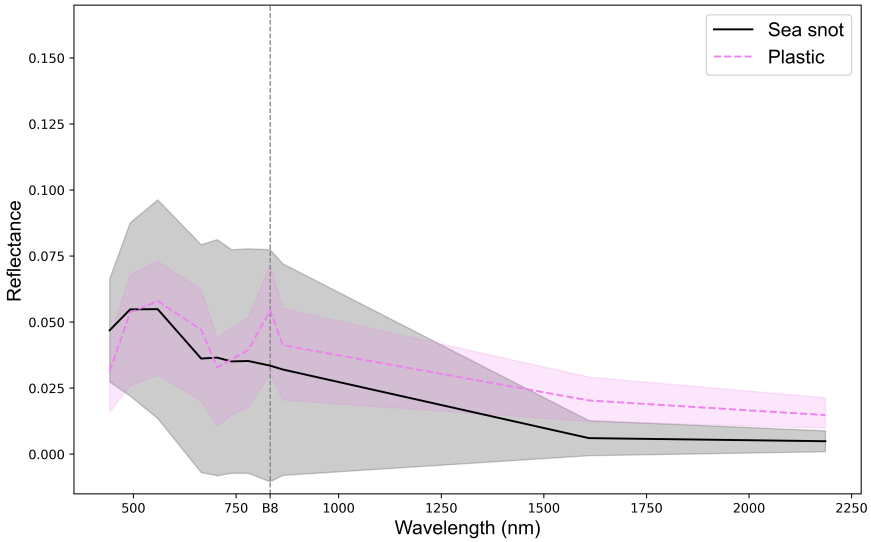

Figure 3.13: Spectral reflectance (mean - line, and standard deviation - shaded area) of pixels of sea snot and plastic after atmospheric correction.

3.4.7 Sea Foam

The model from Biermann et al. [20] showed some difficulties in distinguishing plastic from sea foam, bubbles, and froth, so this group of substances was included in this study. A Sentinel-2A image from Vigo Ria in Galicia, Spain, was used to gather 2735 pixels and study sea foam's spectral behaviour (Figure 3.14). Sea foam presents a small reflectance peak in the early spectral bands and another one in the eighth Sentinel-2 band, just like the plastic mean spectral signature. These features, adding to the relatively high standard deviation, suggest that sea foam might be confused with plastic.

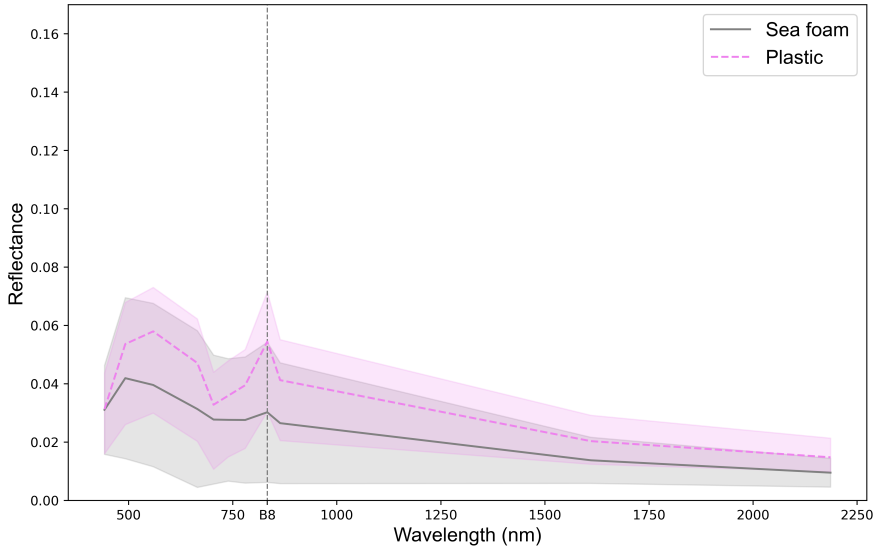

Figure 3.14: Spectral reflectance (mean - line, and standard deviation - shaded area) of pixels of sea foam and plastic after atmospheric correction.

3.4.8 Data Overview

With all data gathered from common floating classes such as seaweed, sea foam and driftwood, and from substances that, despite not being so common, can provoke harmful environmental consequences such as sea snot and pumice, it is possible to compare all spectral signatures to plastic (Figure 3.15). Both driftwood and seaweed have very high reflectance in the eighth Sentinel-2 spectral band, making them distinguishable from all other materials. Driftwood has a reflectance peak where seaweed has an absorption peak, so they are also distinct from each other. As expected, water has lower reflectance than all other classes thanks to its high heat capacity. On the other hand, separating plastic from pumice, sea snot, and sea foam based on spectral responses is challenging. Therefore, using spectral indices, which are mathematical equations that combine values from two or more wavelengths enhancing spectral features that were not visible initially, is appropriate.

The acquisition of data allowed the collection of 60807 pixels in total. However, around 98% of these pixels are from pumice (51%), sea snot (43%), and sea foam (5%), meaning that the dataset is unbalanced. Training a machine learning model with this data distribution causes issues since the algorithm will be biased towards the majority classes, i.e., it is prone to overfitting the most represented classes and won't understand the underlying patterns of the data, leading to poor results. Consequently, there is a need to balance the training dataset.

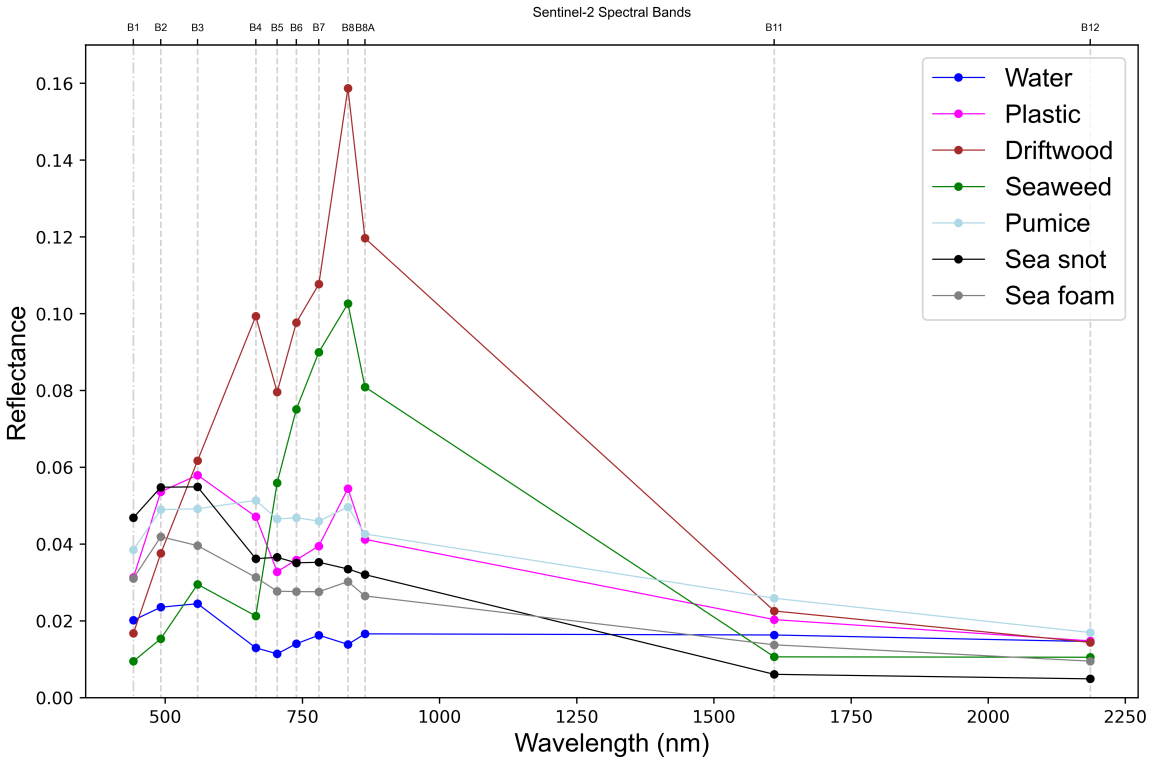

Figure 3.15: Spectral signatures derived from the mean reflectance of all data after atmospheric correction.

There are several methods to balance a dataset. The most trivial would be collecting more data from the minority classes but finding more plastic or driftwood pixels is not achievable. Oversampling the minority classes by creating synthetic pixels is also an option. Undersampling is probably the most direct strategy. This method selects only some data from the majority classes, using close to as many pixels as the minority classes have, but maintaining the classes' probability distributions. The training set distribution can be seen in Table 3.2. To maintain each class characteristics, the training set includes 156 plastic pixels, corresponding to around 22% of the data. Both pumice and sea foam classes have 105 pixels (14.7%), sea snot has 114 pixels (16%), driftwood has 39 pixels (5.5%), seaweed has 97 pixels (13.6%) and, finally, water corresponds to 13.7% of the training data, with 98 pixels. The remaining pixels are grouped in the testing set and are used to evaluate the machine learning models' results. The table also shows the dataset used to train the model that will perform predictions in real-world conditions.

Table 3.2: All data collected to train and test the machine learning models proposed in this work.

					Number of pixels						
Type	Source	Date (dd/mm/yyyy)	Location	Sentinel	Water	Plastic	Wood	Seaweed	Pumice	Sea snot	Sea foam
Artificial	[30]	15/12/2018	Limassol, Cyprus	2A	0	4	0	0	0	0	0
	[28]	07/06/2018	Tsamakia beach, Greece	2A	0	1	0	0	0	0	0
	[19]	18/042019		2B	0	3	0	0	0	0	0
		18/05/2019		2B	0	2	0	0	0	0	0
	[29]	11/06/2021	Gulf of Gera, Greece	2A	10	9	0	0	0	0	0
		21/06/2021		2A	8	9	6	0	0	0	0
		26/06/2021		2B	8	9	9	0	0	0	0
		01/07/2021		2A	4	4	4	0	0	0	0
		06/07/2021		2B	4	4	3	0	0	0	0
		11/07/2021		2A	4	4	4	0	0	0	0
		16/07/2021		2B	4	4	4	0	0	0	0
		21/07/2021		2A	4	5	4	0	0	0	0
		26/07/2021		2B	3	3	2	0	0	0	0
		31/07/2021		2A	5	4	6	0	0	0	0
		05/08/2021		2B	4	4	4	0	0	0	0
		10/08/2021		2A	6	6	6	0	0	0	0
		25/08/2021		2B	6	6	6	0	0	0	0
		30/08/2021		2A	5	7	4	0	0	0	0
Real	[20]	24/04/2019	Durban, South Africa	2B	0	72	0	0	0	0	0
		31/10/2018	Accra, Ghana	2B	0	0	0	150	0	0	0
	[31]	09/10/2017	Caribbean Sea, Honduras	2A	0	49	0	0	0	0	0
	[31]	03/11/2016	Caribbean Sea, Honduras	2A	75	0	0	0	0	0	0
	News/Social Media	26/10/2021	Okinawa, Japan	2A	0	0	0	0	31098	0	0
	News/Social Media	06/06/2021	Marmara Sea, Turkey	2B	0	0	0	0	0	26403	0
	Observation	20/09/2016	Vigo, Spain	2A	0	0	0	0	0	0	2735
	Total				150	209	62	150	31098	26403	2735
	Training set				98	156	39	97	105	114	105
	Testing set				52	53	23	53	30993	26289	2630
	Training set for the final model				150	209	62	150	150	150	150

3.5 Spectral Indices

Spectral indices are mathematical combinations of spectral reflectance values from two or more wavelengths which display the relative abundance of features of interest [63]. Using spectral indices to detect and differentiate floating materials in the ocean through multi-spectral data gains relevance since some classes, such as plastic, pumice, sea snot, and sea foam, have similar spectral signatures (Figure 3.15). These indices can highlight the slight differences in spectral reflectance for the different materials, allowing a better separation. Spectral bands 9 and 10 cannot be used to compute indices since their values were removed in the atmospheric correction process for all data.

Biermann et al. [20] showed that the NDVI, which is commonly used for agricultural applications since it quantifies the vegetation's greenness and density, has distinct values for different materials. Water, sea foam, driftwood, and pumice occupy separate NDVI ranges. The study also showed that combining the NDVI with the recently developed FDI allows distinct clustering of materials, such as floating plastic, water, seaweed, and driftwood. Themistocleous et al. [30] performed an identical study to examine which indices can distinguish plastic from water. Two new indices were developed: the RNDVI and the PI. Seven already established indices were also studied: the NDWI, the Water Ratio Index (WRI), NDVI, the Automated Water Extraction Index (AWEI), the Modified Normalised Difference Water Index (MNDWI), the Normalised Difference Moisture Index (NDMI), and the Simple Ratio (SR). This investigation concluded that the PI provides a clearer separation between plastic and water than the other indices. This work differs from the previous studies since its goal is not only to differentiate plastic from water but also from materials that show similar spectral reflectance, such as sea snot, sea foam, and pumice. Consequently, the combination of NDVI with FDI, or the use of PI and RNDVI may not be enough. Hence, besides the use of the individual spectral bands, 24 different indices will be compared to verify which ones allow a distinct clustering of all the classes gathered in the data acquisition process. Their equations are shown below. In each equation, BX represents the reflectance value for the Sentinel-2 spectral band X, and λ_{BY} represents the central wavelength of the Sentinel-2 spectral band Y.

The FDI [20] is based on the FAI, developed for the Landsat satellite's spectral instruments, and combines the values of three Sentinel-2 spectral bands:

$$FDI = B8 - (B6 + (B11 - B6) \cdot \frac{\lambda_{B8} - \lambda_{B4}}{\lambda_{B11} - \lambda_{B4}} \cdot 10) \tag{3.1}$$

PI [30] was designed for the specific task of identifying plastic floating on the ocean in satellite imagery. It uses the fourth and eighth Sentinel-2 bands, which are two of the spectral bands with the most spatial resolution:

$$PI = \frac{B8}{B4 + B8} \tag{3.2}$$

The NDVI is usually applied to measure vegetation health since healthy vegetation reflects more in band 8 (higher reflectance values) and absorbs more in band 4 (lower reflectance values). This formula generates values between -1 and 1:

$$NDVI = \frac{B8 - B4}{B4 + B8} \tag{3.3}$$

The RNDVI uses the same bands as the NDVI but changes the operation in the numerator:

$$RNDVI = \frac{B4 - B8}{B4 + B8} \tag{3.4}$$

Instead of using band 4, the Green Normalised Difference Vegetation Index (GNDVI) uses band 3 whose wavelength is associated with green light:

$$GNDVI = \frac{B8 - B3}{B3 + B8} \tag{3.5}$$

The Pan Normalised Difference Vegetation Index (PNDVI) combines the spectral values of bands 2, 3, and 4, and executes the same operation as the NDVI:

$$PNDVI = \frac{B8 - (B2 + B3 + B4)}{B8 + (B2 + B3 + B4)} \tag{3.6}$$

The NDWI [56] is strongly related to the pixel water content. As shown previously, this index is also used to differentiate the ocean from land (land mask). It uses the third and eighth Sentinel-2 spectral bands:

$$NDWI = \frac{B3 - B8}{B3 + B8} \tag{3.7}$$

The MNDWI is a variation of the previous index. Instead of using band 8, it uses band 12:

$$MNDWI = \frac{B3 - B12}{B4 + B12} \tag{3.8}$$

Just like the couple of previous indices, the NDMI is used to determine water content in a pixel. It is the ratio between the difference and the sum of radiations in the NIR (band 8) and SWIR (band 11):

$$NDMI = \frac{B8 - B11}{B8 + B11} \tag{3.9}$$

The Normalised Difference Snow Index (NDSI) is related to the presence of snow in a pixel. It is the normalised difference between band 3 (green spectral band) and band 11 (SWIR):

$$NDSI = \frac{B3 - B11}{B3 + B11} \tag{3.10}$$

The WRI [64] uses four spectral reflectance values. It is the ratio between two visible spectral bands (green and red) and the eighth and twelfth bands:

$$WRI = \frac{B3 + B4}{B8 + B12} \tag{3.11}$$

The Normalised Burn Ratio (NBR) is commonly used to highlight burnt areas and measure burn severity. The formula is similar to the NDVI except that it uses both NIR and SWIR spectral bands:

$$NBR = \frac{B8 - B12}{B8 + B12} \tag{3.12}$$

The AWEI [65] was introduced to improve classification accuracy in areas that include shadow and dark surfaces that other methods often failed to label. It uses the reflectance values of four Sentinel-2 spectral bands:

$$AWEI = 4 \cdot (B3 + B12) - (0.25 \cdot B8 + 2.75 \cdot B11) \tag{3.13}$$

The SR, or Ratio Vegetation Index (RVI), is the ratio between the NIR and red spectral bands:

$$SR = \frac{B8}{B4} \tag{3.14}$$

Gitelson et al. [66] proposed the Anthocyanin Reflectance Index (ARI) to estimate anthocyanin content in leaf tissues, which is based on the green (band 3) and red-edge (band 5) spectral responses of the pixels:

$$ARI = \frac{1}{B3} - \frac{1}{B5} \tag{3.15}$$

The Modified Anthocyanin Reflectance Index (MARI) adds band 7 to the previous equation:

$$MARI = \frac{1}{B3} - \frac{1}{B5} \cdot B7 \tag{3.16}$$

The Chlorophyll Red-Edge Index (CHL Red-Edge) uses the ratio between the NIR and Red-Edge spectral bands:

$$CHLRed - Edge = \frac{B8}{B5} - 1 \tag{3.17}$$

Red Edge Position Index (REPI) uses the reflectance values of four spectral bands and is sensitive to changes in chlorophyll concentration in a pixel:

$$REPI = 700 + 40 \cdot \frac{\frac{B4+B7}{2} - B5}{B6 - B5} \tag{3.18}$$

The Enhanced Vegetation Index (EVI), as the NDVI , is used to quantify vegetation health. It uses the NIR , red and blue spectral bands:

$$EVI = 2.5 \cdot \frac{B8 - B4}{B8 + 6 \cdot -7.5 \cdot B2 + 1} \tag{3.19}$$

The Enhanced Vegetation Index 2 (EVI2) has the same goal as the EVI but does not use the blue band:

$$EVI2 = 2.4 \cdot \frac{B8 - B4}{B8 + B4 + 1} \tag{3.20}$$

The Modified Chlorophyll Absorption Reflectance Index (MCARI) measures the depth of chlorophyll absorption and is sensitive to changes in chlorophyll concentrations. It uses three Sentinel-2 spectral bands:

$$MCARI = B5 - B4 - 0.2 \cdot (B5 - B3) \cdot \frac{B5}{B4} \tag{3.21}$$

The Moisture Index (MI) is the ratio between the SWIR and NIR bands. As the NDMI, it determines the water content in a pixel:

$$MI = \frac{B11}{B8} \tag{3.22}$$

The Soil-Adjusted Vegetation Index (SAVI) is a modification of the NDVI to minimize the brightness influence. L represents the amount of green vegetation expected to be found in a pixel (L=1 where there's no vegetation cover, and L=0 where there's a high vegetation cover, which equals the NDVI formula). In this work, L=0.5:

$$SAVI = \frac{B8 - B4}{B4 + B8 + L} \cdot (L + 1) \tag{3.23}$$

The Oil Spill Index (OSI) uses the Sentinel-2 visible bands and was developed to reveal oil spills in the marine environment:

$$OSI = \frac{B3 + B4}{B2} \tag{3.24}$$

3.6 Data Augmentation

Machine learning models highly depend on the quantity and quality of the training data. Data augmentation is a set of techniques to artificially increase the size of the datasets by generating new data points from existing data. These techniques include making slight changes to the data. For example, rotating an image creates several new images, or creating new data from scratch based on the characteristics of the original data. Data augmentation methods enable the models to learn from a variety of data that could not be gathered in the data acquisition step, making them more robust, and reducing the time-consuming process of collecting and labelling data. In this work, making minor changes in the original data, such as rotating, cropping, zooming or grayscaling is not possible, and slightly changing the values of the spectral bands may create spectral responses that do not represent any floating class. Therefore, in this work, Generative Adversarial Networks (GANs) were used to generate synthetic pixels since they can learn and replicate patterns and features of the actual data. A GAN is a type of neural network (section 3.6.1) that generates new data. This chapter describes the two methods used to create data (sections 3.6.2 and 3.6.3) and shows the spectral signatures of the synthetic data.

3.6.1 Neural Networks

Neural networks, or Artificial Neural Networks (ANNs), are a subset of machine learning inspired by a study published in 1943 by Warren McCulloch and Walter Pitts [67]. The study described brain functions in abstract terms and showed that simple elements combined in a neural network can have immense computational power. ANNs are composed of multiple layers of interconnected units, also known as nodes or neurons. ANNs have three layer types: the input layer that receives the data as input, the hidden layers, and the output layer that directly outputs a prediction. If, in a hidden layer, each neuron connects to all the others in two adjacent layers, its layer is fully connected (also known as a dense layer) and, when an ANN has more than one hidden layer, it is called a Deep Neural Network. Figure 3.16 shows the structure of a Deep Neural Network with fully connected layers.

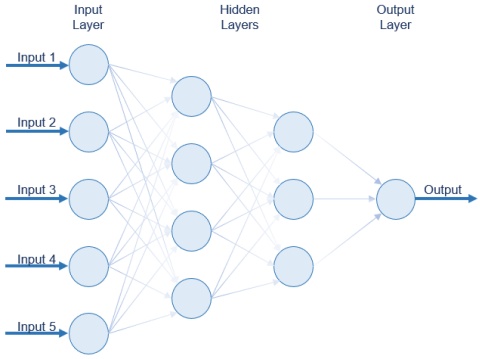

Figure 3.16: Architecture of a Deep Neural Network.

In a neural network, each node is composed of input data, weights to determine the importance of an input variable, a threshold to quantify the output, and an output. All inputs are multiplied by their respective weights and summed. This value then passes through an activation function, which determines the node's output and, if the output exceeds the threshold, the node is activated, passing data to the next layer. Passing data from one layer to the next defines a neural network as a feedforward network. Three of the most common types of activation functions are the Sigmoid, the Hyperbolic Tangent (Tanh) and the Leaky Rectified Linear Unit (Leaky ReLU). These functions add non-linearity to a machine learning model. The Sigmoid function (Eq. 3.25), also known as the logistic function, takes any real value and maps it between 0 and 1. If the value of x goes to positive infinity, then the predicted value of f(x) is 1. If, on the other hand, the value of x goes to negative infinity, then f(x) becomes 0. Therefore, if the outcome of the sigmoid function is higher than 0.5, the node is activated, otherwise, it is not. The Tanh activation function (Eq. 3.26) is similar to the Sigmoid function, having the same 'S' shape. However, instead of mapping the outputs between 0 and 1, it takes any real value and maps it between -1 and 1. Finally, the Leaky ReLU activation function (Eq. 3.27), is based on a ReLU but has a slight slope for negative values instead of being flat like the ReLU. The slope coefficient, a, is determined before the training process. This activation function is popular when training GANs. The three functions are graphically represented in Figure 3.17.

$$f_1(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.25}$$

$$f_2(x) = tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.26}$$

$$f_3(x) = max(\alpha \cdot x, x) \tag{3.27}$$

Eventually, the goal of the neural network is to minimize a cost function. The cost function represents the average loss produced by the model, which quantifies the difference between the produced and the expected outcome. When there are only two classes, binary cross-entropy is commonly used as a cost function. It measures the difference between computed probabilities and actual probabilities for predictions with only two possible classes. The greater the cross-entropy loss, the further away the predictions are from the true labels [68]. Equation 3.28 shows the binary cross-entropy formula, where y is the true label (1 for one class and 0 for the other class), and p(y) is the predicted label for all N points.

$$L_p(w) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i)) \tag{3.28}$$

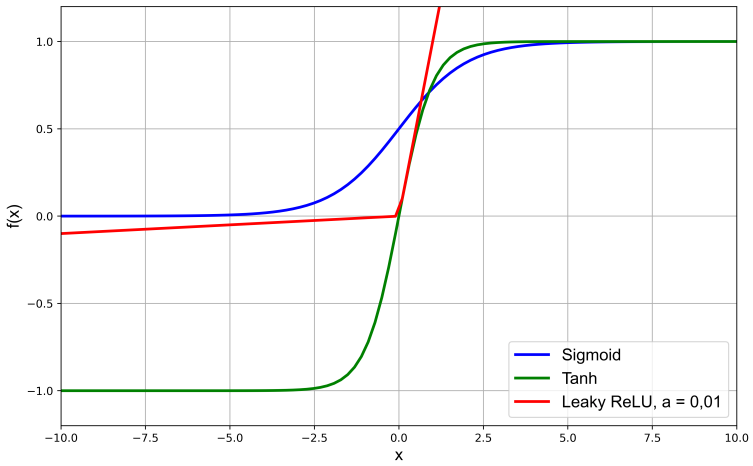

Figure 3.17: Sigmoid, Tanh and Leaky ReLU activation functions. The Leaky ReLU function's slope coefficient is 0.01.

Finally, to compute the contribution of each parameter to the loss and update the weights of each variable, several optimization algorithms can be used. One of the most common is the stochastic gradient descent, which maintains a single learning rate (α) for all weight updates and does not change during the training process. The Adaptive Moment Estimation (Adam) optimization algorithm is an extension of the stochastic gradient descent and has become the most popular optimizer for GAN implementations since, in most cases, it generally achieves better results in faster computation time and requires fewer parameters for tuning.

3.6.2 Generative Adversarial Networks

In 2014, Goodfellow et al. [69] designed a class of machine learning methods called GANs. As the word generative suggests, the purpose of this class of machine learning techniques is to generate realistic data with the same statistics as the training set. It comprises two simultaneously trained models: the Generator and the Discriminator. There is a competitive dynamic between the two models that make up the GAN structure. The Generator's goal is to create samples that are indistinguishable from the training data. On the other hand, the Discriminator tries to distinguish authentic data from the synthetic data generated by the first model. Finally, the word networks points out the class of machine learning models used to create the Generator and the Discriminator: neural networks (section 3.6.1) [68].

To begin the process of creating synthetic data, the Generator starts by receiving a vector of random numbers as input and outputs fake data. Its goal is to generate data that is indistinguishable from the training data, so that the Discriminator cannot tell the difference between fake and authentic data. The Discriminator takes as input an example from the training data, which in this case corresponds to a pixel of a floating class, or an example created by the Generator and outputs the probability that the input is real. The Discriminator's loss is used to iteratively tune both networks through backpropagation and its weights are updated to maximize the probability of correct prediction, on the other hand, the Generator's weights are updated to maximize the probability of misclassification by the Discriminator. Thus, the networks have conflicting objectives, meaning that when one gets better, the other gets worse, i.e., the Generator gains correspond to Discriminator losses and vice versa. This situation is known as a zero-sum game. Zero-sum games reach a point where neither player, in this case, networks, can improve their situation by changing their actions, called Nash equilibrium. GANs achieve equilibrium (converge) when the Generator's outputs are indistinguishable from the training data, and the Discriminator can at best randomly guess if the input is authentic or synthetic. However, in practice, it's almost impossible to find Nash equilibrium for GANs [68]. Figure 3.18 shows the architecture of a GAN and the relationships between its elements.

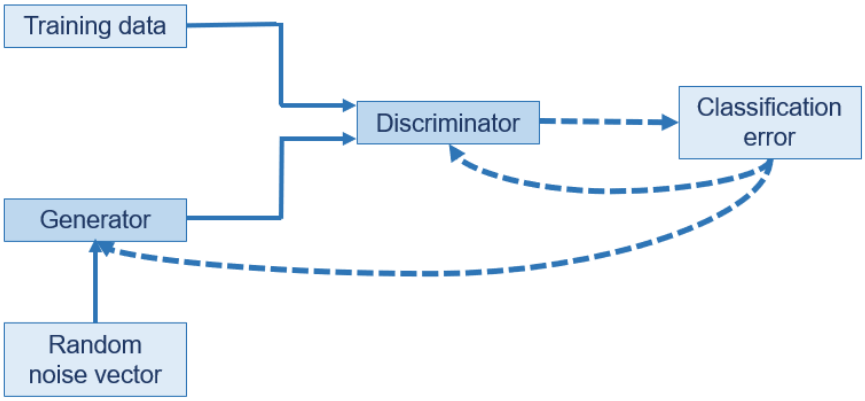

Figure 3.18: GAN architecture diagram.

In this work, both the Generator and the Discriminator are ANNs and have six hidden fully connected layers with different numbers of nodes. All hidden layers use the Leaky ReLUs activation function, allowing a small positive gradient that prevents the gradients from fading away, enabling better training outcomes. The Generator's input layer takes the noise vector, which has 100 values sampled from a standard normal distribution (0 mean and unit standard deviation). Its output layer has 11 values, since it creates pixels that have 11 different spectral band values and uses the Tanh activation function. As discussed previously in this section, this activation function scales the output values to the range [-1, 1]. On the other hand, the Discriminator's output layer uses the sigmoid activation function, ensuring that the output value is between 0 (100% probability the pixel is synthetic) and 1 (100% probability the pixel is real). Binary cross-entropy is used as the loss function and, to optimize each network the Adam optimization algorithm is applied. In each iteration, the Discriminator network is trained while keeping the Generator's parameters fixed, then the Generator is trained while the Discriminator's parameters are constant, which maintains both networks at the same level during the training process. There is no absolute method to determine the number of iterations, or the number of samples processed in each iteration (batch size), so they are set through trial and error. The data created in this work through a GAN uses 500 iterations and has a batch size with half the size of the training data.

The resulting synthetic data generated from the GAN is shown in Figure 3.19. In total, 14000 pixels were generated, equally distributed (2000 pixels from each class). If the class had training data from both Sentinel-2 satellites (water, plastic, and driftwood), 1000 pixels were generated from each satellite. Overall, for every class except pumice and sea snot, the synthetic data mean reflectance is similar to the original reflectance. Pumice synthetic data has higher reflectance in all spectral bands, although having an identical spectral behaviour. Sea snot synthetic pixels are probably the worst, since spectral band values are far from the original data, due to the large standard deviation of the original dataset. The generated plastic pixels resemble the real ones in the mean but have slightly lower standard deviation. The remaining classes (water, driftwood, seaweed, and sea foam) have low standard deviation, which

reveals one common problem when using GANs: mode collapse. Mode collapse happens when the Generator only produces a single type of output, usually close to the mean of the original data, since that type of data can ”fool” the Discriminator. Other problem that arises during the training of the GAN is the slow speed of convergence. Therefore, another solution to generate data is needed.

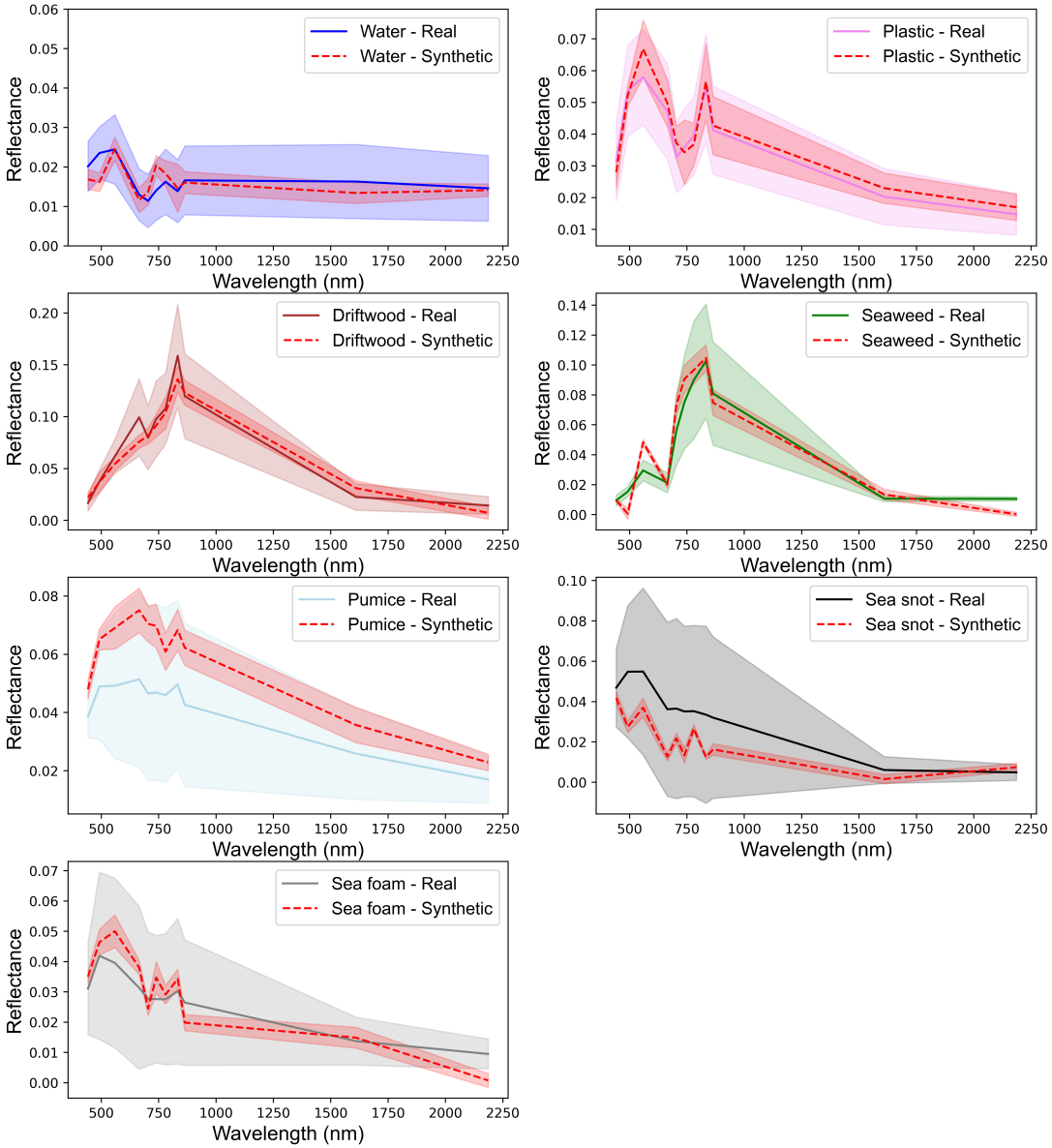

Figure 3.19: Comparison between spectral reflectance (mean - line and standard deviation - shaded area) of real pixels and synthetic pixels (in red), from every class, generated from a GAN.

3.6.3 Wasserstein Generative Adversarial Networks

In 2017, an alternative method for generating synthetic data using GANs was introduced [70]. The Wasserstein Generative Adversarial Network (WGAN), uses an alternative way of training the Generator network to better approximate the generated data distribution to the training dataset and offers higher stability in the training process. Sometimes it takes more time to train than a GAN. However, it usually achieves more efficient results and is less sensitive to the model architecture and choice of hyperparameters. Instead of using a Discriminator to predict the probability of the input being real or fake, it uses a Critic that scores the "realness" or "fakeness" of the data, which, by using an improved loss function, provides a clearer stopping criteria.

The implementation of a WGAN requires some changes to the previous architecture. The main difference between both types of GANs is the loss function. WGAN uses the earth mover's distance, or Wasserstein distance, as the loss function to evaluate the distance between the distribution of the training data and the generated one. The Critic's goal is to maximize the distance between real and synthetic data scores - maximize Eq. 3.29, since it needs to distinguish the data properly. Inversely, the Generator's goal is to minimize the distance between real and fake data scores - minimize Eq. 3.29, because a lower distance means that the synthetic data is closer to the real data. Therefore, this loss function is much more informative, providing significant gradients to both the Critic (the Discriminator in GANs) and the Generator. In Eq. 3.29, function f_ω acts as a Discriminator and satisfies the technical constraint shown in Eq. 3.30, m is the batch size, ω are the parameters of the critic, function g_θ acts as a Generator, θ are the parameters of the generator and z corresponds to the latent space (noise vector).

$$\frac{1}{m}\sum_{i=1}^{m} f_\omega(x^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} f_\omega(g_\theta(z^{(i)})) \tag{3.29}$$

$$|f(x_1) - f(x_2)| \leq |x_1 - x_2|, \forall x_1, x_2 \tag{3.30}$$

The Critic model uses a linear activation function in the output layer (Eq. 3.31), instead of a sigmoid function, to not limit the output values. The labels for fake data and real data also change to be used in the cost function. In a GAN, real data were labelled as 1 and synthetic data as 0. On the other hand, in a WGAN, real data have label -1 and generated data have label 1.

$$f(x) = x \tag{3.31}$$

WGAN also requires gradient clipping for the Critic model (i.e., forcing the gradient values to a maximum or a minimum value if the gradient exceeds an expected range). This thesis uses the range [-0.01,0.01]. Unlike the GAN implementation, where the Generator and Discriminator are updated in equal amounts, the Critic is updated more often than the generator. In this work's implementation, the Critic updates five times for each time the Generator does so. Finally, instead of using the Adam optimization algorithm, the WGAN uses the Root Mean Square Propagation (RMSprop) with a small learning rate (0.00005 in this implementation).

The resulting synthetic data generated from the WGAN is shown in Figure 3.21. In total, 280000 pixels were generated, equally distributed (40000 pixels from each class). If the class had training data from both Sentinel-2 satellites (water, plastic, and driftwood), 20000 pixels were generated from each satellite. As the previous method, the WGAN was difficult to train since the fact that it converges does not always mean that the generated data are of good quality. For example, Figure 3.20 shows the generator and critic losses when training the WGAN with water pixels from Sentinel-2A. The WGAN only converged at around 1400 iterations and the resulting data is not perfect (Figure 3.21). The mean reflectance values of the synthetic pixels are similar to the real ones in most classes and have higher standard deviations compared to the previous generated data. Driftwood and seaweed generated classes are near perfect. In opposition, mean values of pumice and sea foam differ from the real pixels in some spectral bands. Synthetic water and plastic pixels show higher deviation than real data. In contrast, pumice, sea snot, and sea foam show lower standard deviation than the real pixels. The main difference between data generated by the WGAN to data generated by the previous method (GAN) is the fact that there is no sign of mode collapse. These data (Figure 3.21) show higher variability (i.e., higher standard deviation), which constitutes an advantage relative to the previous method since a model trained with them will be more prepared to predict the class of outliers. However, the tuning of this architecture takes some time and the data generation process has a high computational cost.

Figure 3.20: Generator and critic losses when training the WGAN with water pixels from Sentinel-2A.

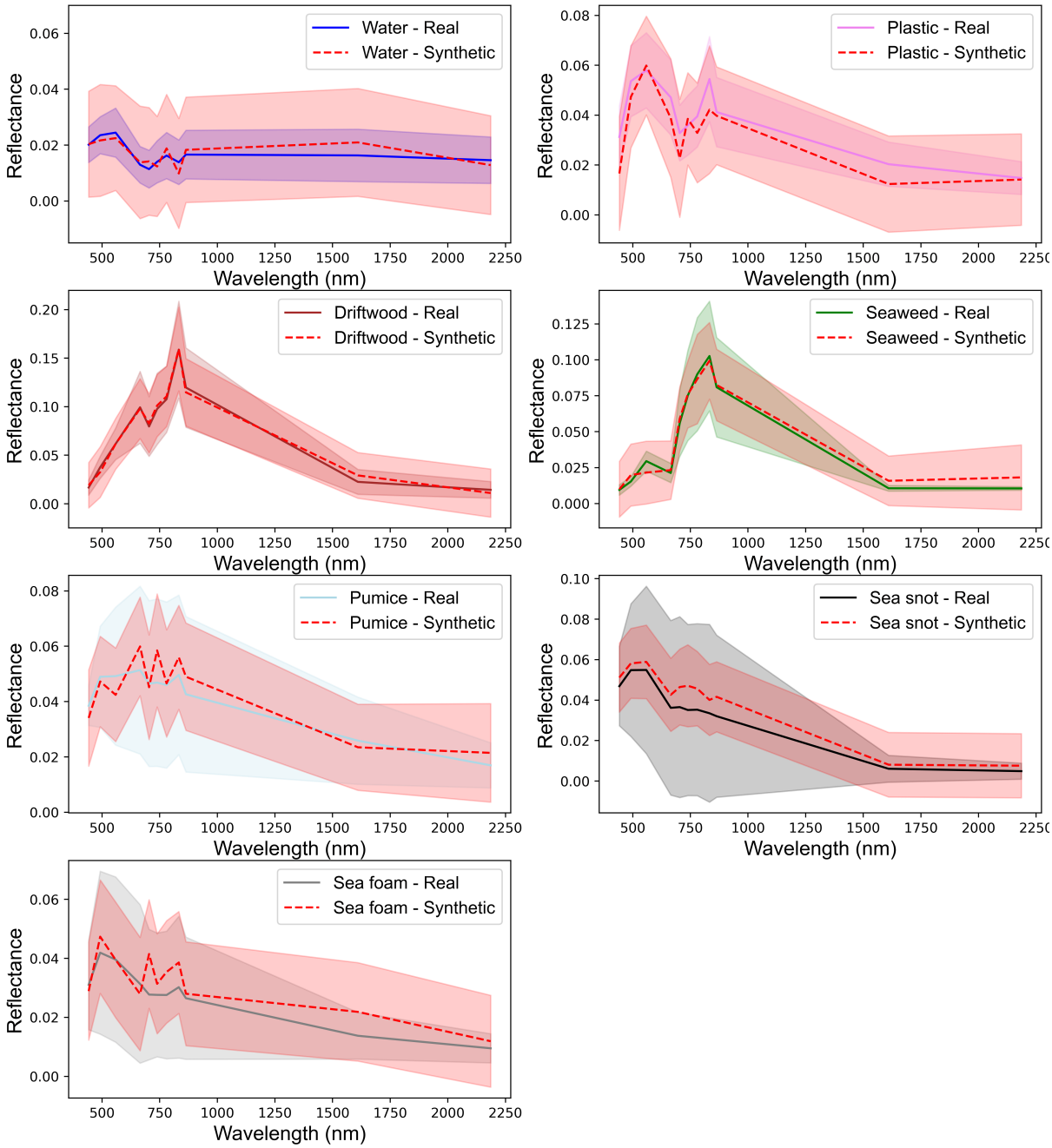

Figure 3.21: Comparison between spectral reflectance (mean - line and standard deviation - shaded area) of real pixels and synthetic pixels (in red), from every class, generated from a WGAN.

3.6.4 Synthetic Data Overview

Table 3.3 shows the synthetic data distribution and the training data used in each case. In total, 14000 pixels were created using a GAN and 280000 using a WGAN. In both cases, the data are equally distributed through all classes, meaning that each class represents around 14.29% of the datasets - balanced data. The pixels in the training data were atmospherically corrected. Therefore, the generated data do not have the ninth and tenth spectral bands. Data generated using the WGAN shows more variability than the one created using a GAN, which allows for better classification results, since it makes the models more robust. The next step is to study if using synthetic data to train the classification models can improve the results.

Table 3.3: All synthetic data generated using a GAN and a WGAN with the respective training data.

Class	Training data		Generated data	
	Sentinel	Number of pixels	Number of pixels (GAN)	Number of pixels (WGAN)
Water	2A	121	1000	20000
	2B	29	1000	20000
Plastic	2A	102	1000	20000
	2B	107	1000	20000
Driftwood	2A	34	1000	20000
	2B	28	1000	20000
Seaweed	2A	0	0	0
	2B	150	2000	40000
Pumice	2A	150	2000	40000
	2B	0	0	0
Sea snot	2A	0	0	0
	2B	150	2000	40000
Sea foam	2A	150	2000	40000
	2B	0	0	0
Total			14000	280000

3.7 Classification Algorithms

The primary goal of this study is to create a model that assigns each pixel of a satellite image with one of the seven classes available (water, plastic, driftwood, seaweed, pumice, sea snot, and sea foam). The algorithm that maps the input data into a specific class is called a classifier and, in this case, since the problem has more than two possible categories, it is called a multi-class classifier. There are three ways to train a classification algorithm, depending on the data available. In a supervised learning model the algorithm learns through an entirely labelled dataset. As data enter the model, it adjusts its weights until it predicts outcomes accurately. On the other hand, in unsupervised learning, the dataset is not labelled and, therefore, the model tries to find a structure in the data by extracting features and groups similar data together. Semi-supervised learning is a method that relies both on labelled and unlabelled data. In this work, since all the data are labelled and it is a classification task, it is used the supervised learning approach. Also, a previous study using unsupervised models to differentiate plastic from water did not achieve good results [32]. This section provides an overview of each supervised learning algorithm tested under the scope of this thesis and how they were evaluated.

3.7.1 Logistic Regression

Logistic regression is utilized to solve classification problems with binary outputs (binary classification) and cannot be used for classification tasks with more than two classes (multi-class classification) without modification. It predicts the probability of an output by fitting the data to a logistic or sigmoid function. If the output's value is greater than a threshold value it is assigned a label, else it is assigned another label. One approach to adapt the logistic regression to multi-class classification problems is to split the problem into multiple binary subproblems and apply the algorithm. An alternate procedure is to change the model to support the prediction of several classes by changing the loss function and the output's shape - Multinomial Logistic Regression.

3.7.2 Support Vector Machine

Support Vector Machine (SVM) is another popular supervised learning model used for classification tasks. The goal of this algorithm is to find a hyperplane in an N-dimensional space, being N the number of features, with the maximum margin that separates and allows the classification of the data points. So, a class is attributed to each data point according to the side of the hyperplane where it stays. The position of the hyperplane depends on the data points that are closer to it, called support vectors. The margin's distance is maximized using the hinge loss as the loss function. Basu et al. [32], using an SVM, achieved an accuracy in the range [96.9,98.4]% in differentiating plastic from water.

3.7.3 Support Vector Machine with Stochastic Gradient Descent

The Stochastic Gradient Descent (SGD) is an optimization algorithm that can be used to minimize the loss function of linear classifiers such as the Logistic Regression and the Support Vector Machine. In this work, this algorithm is used to minimize the loss function of an SVM. SGD is a stochastic approximation of the Gradient Descent algorithm. Gradient Descent is an iterative optimization algorithm that minimizes a function by travelling down its slope until it reaches the minimum. In each iteration, it computes the gradient of the whole dataset. The convergence of this algorithm significantly depends on the learning rate. Larger learning rates make the algorithm run faster. However, it might miss the minimum. On the other hand, small steps impact the time it takes to converge. Gradient Descent is especially slow on large datasets. The Stochastic Gradient Descent focuses on solving this problem. Instead of calculating the gradient of the whole dataset, it randomly selects a single sample to perform each iteration and computes the gradient, which significantly decreases the total time.

3.7.4 Gaussian Naïve Bayes

Naïve Bayes methods are a set of algorithms based on the Bayes' theorem that assume conditional independence between features, given the value of the class. A Naïve Bayes classifier considers that a particular feature in a class does not influence the presence of any other feature in the probability of a given outcome. One of the Naïve Bayes classifier types is the Gaussian Naïve Bayes. This method assumes that the class-conditional densities are normally distributed. Normal or Gaussian distribution is a probability distribution symmetric about the mean (i.e., with an equal number of values above and below the mean value). Despite their apparent simplicity, Naïve Bayes classifiers work well in real-world applications and require a small amount of training data to estimate their parameters. Biermann et al. [20] achieved 86% accuracy in distinguishing plastic from water, seaweed, sea foam, and timber using a Gaussian Naïve Bayes classifier.

3.7.5 K-nearest Neighbour

The K-nearest Neighbour (KNN) algorithm classifies data points based on their proximity to other data points. For each new data point, it computes its distance to the other data points, usually the Euclidian distance, and then it assigns the most frequent class of its k nearest neighbours. The value of k is specified by the user, but its optimal value highly depends on the data. Generally, a larger k suppresses the effects of noise, however, it makes the classification boundaries less distinct [71]. The simplicity and low processing time for smaller datasets make this algorithm popular. Nevertheless, as the dataset grows, the processing time also rises, making it less viable.

3.7.6 Random Forest

The Random Forest algorithm is based on the concept of ensemble learning. Ensemble learning is the process of combining multiple models, such as classifiers, to solve a problem. As the name implies, a Random Forest classifier is composed of multiple decision trees trained on various subsets of the given training set that predict the output's value. The final output is the one selected by most trees or the average output, which solves the overfitting problem that using just one decision tree brings and, therefore, improves the accuracy.

3.7.7 Extreme Gradient Boosting

Extreme Gradient Boosting, also known as XGBoost, is a tree-based ensemble machine learning algorithm (i.e., it comprises multiple combined decision trees), just as the Random Forest, but with a different training strategy. Random Forests are trained using the Bagging method. Bagging consists of random sampling subsets of the training data, fitting these subsets to various models (trees), and then the output corresponds to the one predicted by most trees or the predictions' average. On the other hand, Extreme Gradient Boosting (XGBoost) is trained using the Boosting technique. Boosting is an additive and sequential learning method where trees are grown sequentially (i.e., are added one at a time), so that each new tree, or weak learner, corrects the errors of the previous one in each iteration. The individual models are built on subsets of data that put more weight on instances with wrong predictions and high errors in the preceding weak learners. Each new tree's parameters, or weights, are established by the gradient descent algorithm, whose goal is to minimize the loss function of the ensemble model. Extreme Gradient Boosting is a specific implementation of the Gradient Boosting method. Two of the most important differences are that it computes the second-order gradients (i.e., second partial derivatives) of the loss function, which provides more information on how to reach the minimum of the loss function [72] and uses both L1 and L2 regularization to penalize the models. Both features prevent the models from overfitting. XGBoost has become widely used and is popular in machine learning competitions due to its highly accurate results. This was the classifier that showed the best results in the testing phase, compared to all the classification algorithms above. Therefore, the XGBoost was the chosen method for developing the final model.

3.7.8 Model Evaluation

For classification problems, the performance of a classifier is typically interpreted according to a confusion matrix (Figure 3.22). A confusion matrix is an N-by-N matrix, being N the number of different classes, whose generic element, Pij, is the probability of classifying an example of class i in class j. Each line of the confusion matrix represents a probability distribution. Therefore, it describes how the

data from that line's class is classified, providing information on what are the most common errors. Figure 3.22 shows a confusion matrix of a binary classification problem. In this work, since it is a multi-class classification problem, with seven distinct classes, the confusion matrix has seven rows and columns. The main goal is to create a model that has a high accuracy classifying plastic (i.e., the sum of the true positives (TP) with true negatives (TN) over the total number of samples is high and, at the same time, the number of false positives of plastic is low).

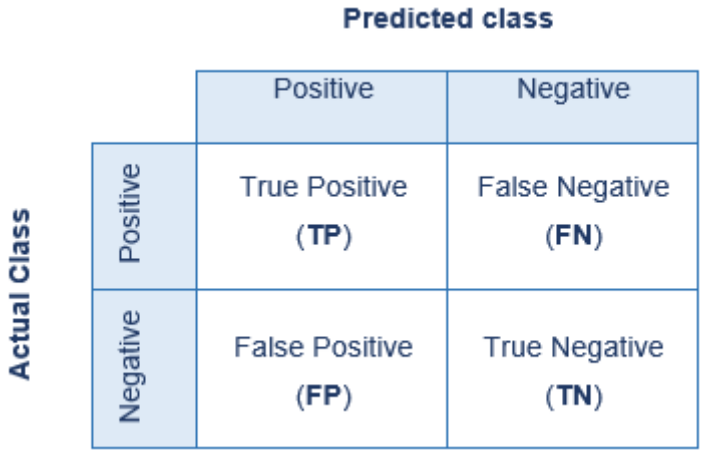

Figure 3.22: Confusion matrix of a binary classification problem.

3.8 Uncertainty Estimation

While machine learning grew into an essential part of multiple real-world applications, the predictions made by these models are uncertain. Uncertainty can be caused by the data (aleatoric or data uncertainty) and the model (epistemic or model uncertainty). According to Gawlikowski et al. [73] there are five key factors that contribute to uncertain predictions: the variability in real-world situations that do not exist in the training set (distribution shift); the errors and noise inherent to the measurements systems that can be provoked by the satellite's spatial and radiometric resolutions and errors caused by false labelling the pixels (label noise); there can be errors in the model structure that originate overconfident or under-confident predictions; a training dataset that suffers from imbalance or low coverage in some areas of the data distribution causes uncertainty; finally, unknown data are also a cause of uncertainty. For example, if there's a material floating in the ocean other than plastic, driftwood, seaweed, pumice, sea snot, or sea foam, the model's output suffers from uncertainty. If these factors are not considered when training and testing a model, their solutions have a high risk of failure in actual world deployment. Providing uncertainty estimates is not only important for a safe decision-making in high-risk fields, but also essential in areas where the data sources are highly inhomogeneous and labelled data is limited, such as in remote sensing [73]. Knowing the uncertainty of a prediction is crucial to prepare an action plan and, in this case, to create measures that aim at improving the management of marine ecosystems.

Currently, there are four main approaches to model the uncertainties specified above. Firstly, Deterministic Neural Networks can be used in two different ways: one single network is modelled and trained to quantify uncertainties, or two neural networks are trained, one to output the predictions and the other to predict the uncertainties on the first network. Compared to the other methods, the deterministic approach is more efficient. However, it relies on the prediction of a single network and, therefore, is very dependent on the network's structure and the training data. The second approach is using Bayesian Neural Networks. Since this method computes predictions based on the law of total probability, it can predict uncertainty. One of the simpler uncertainty prediction methods is the Test Time Augmentation. The first step is to generate multiple test samples from the real test samples using data augmentation techniques, then test the model with all the data and compute the predictive distribution in order to estimate uncertainty. This method does not alter the model and is simple to apply, but significantly depends on the data augmentation technique used. Finally, Ensemble Methods measure uncertainty based on the predictions of multiple models (ensemble members). The ensemble models are trained independently from each other using different techniques to increase their variety, and then, the mean, variance, and standard deviation of their predictions are computed to estimate the uncertainty. This approach has a high computational cost. However, ensemble methods were proven to be more reliable and applicable to real-world applications than the other methods [74].

The chosen strategy to quantify uncertainty in this work is the Ensemble Methods. The classification

model that achieves the best results in the testing phase will be selected to create several ensemble members. To bring variety into the ensembles and generate more reliable results, the models will have random initialization, will be trained with different data sizes (mini batches), and the training data will be randomly selected. Then, if a prediction's mean for a pixel is below 90% or its standard deviation above 20, the pixel's classification is considered uncertain. These thresholds were selected after some adjustments, based on the confusion matrix of the ensemble model.

4

Results

Contents

4.1 Classification with spectral bands

In this section, it is studied if training machine learning algorithms with spectral bands as features allows to distinguish plastic from the other classes. Figure 4.1 shows the mean and standard deviation for every class and for each spectral band. For example, in the first graph, B1, seaweed shows a clear separation from every class, except driftwood. This means that a model trained only with the first spectral band as a feature would distinguish seaweed from the other classes (except driftwood) with high accuracy. The second graph, B2, shows that in this spectral band, seaweed has a different signature than driftwood. Therefore, if a model was trained with B1 and B2 as features, it could distinguish seaweed from all the other classes with high accuracy. On the other hand, plastic, pumice, and sea foam show similar characteristics (especially the mean) in every spectral band. Consequently, it is expected that a model trained only with spectral bands reveals some problems in distinguishing plastic from pumice and sea foam.

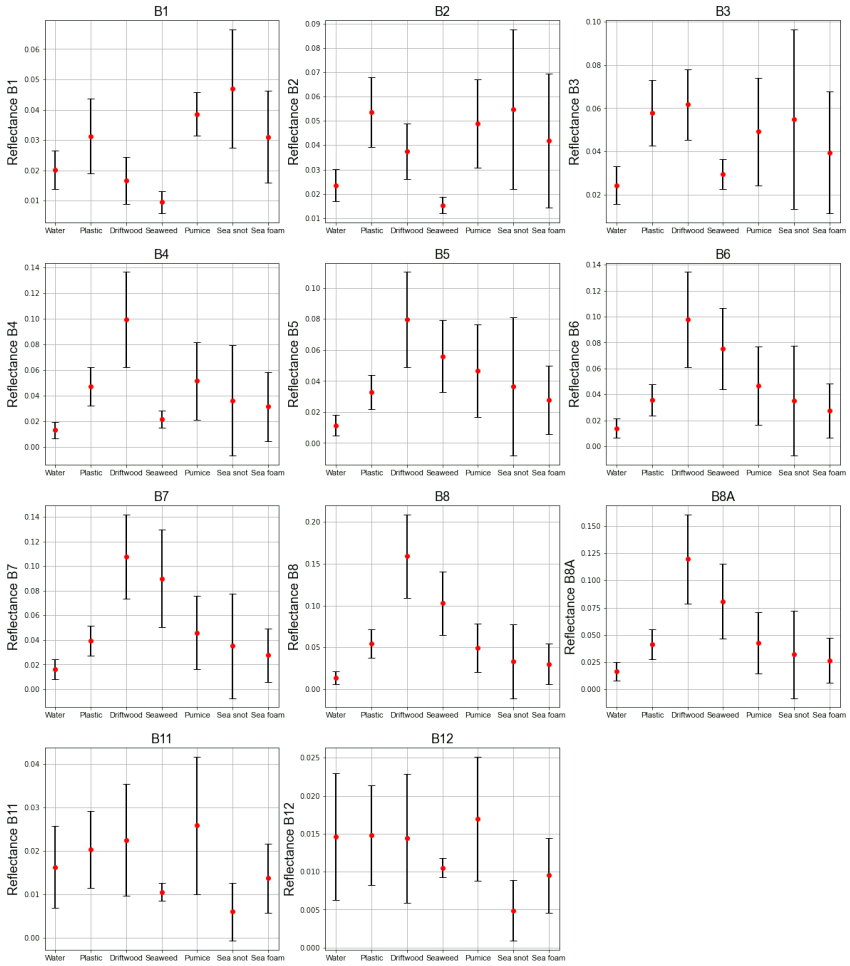

Figure 4.1: Spectral reflectance (mean in red and standard deviation in black) of each class, for each Sentinel-2 spectral band. Bands 9 and 10 have been removed in the atmospheric correction process.

Every classification method described in the previous chapter was trained with the ”training set for the final model” (Table 3.2), with all spectral bands as features and their hyperparameters were selected to maximize the accuracy and minimize the possibility of overfitting. The models that showed the best results were the Logistic Regression, the Random Forest, the K-nearest neighbour with K=1 (making it extremely sensitive to noise, outliers, and mislabelled data), and the XGBoost. From these, the XGBoost has a higher percentage of true positives for each class. However, the model that showed the least number of false positives in plastic was the Logistic Regression. Therefore, both methods are suitable candidates for creating the final model that will monitor the ocean in certain areas. The normalised confusion matrix of the XGBoost is shown in Figure 4.2. The results of every model except the XGBoost are shown in appendix B. The XGBoost model is capable of differentiating seaweed from all the other classes (i.e., all the seaweed pixels were classified correctly and none of the other classes' pixels were classified as seaweed). 96% of plastic pixels were classified accurately, but there are some false positives of plastic. 4% of water, driftwood, and pumice pixels were labelled as plastic, as well as 1% of sea snot and 8% of sea foam pixels. It is clear that the model has the most trouble when classifying sea foam pixels, since it has a spectral signature similar to plastic and pumice. Overall, the results are promising and represent an improvement to the study published by Biermann et al. [20]. However, the number of pixels wrongly classified as plastic must be addressed in the following models.

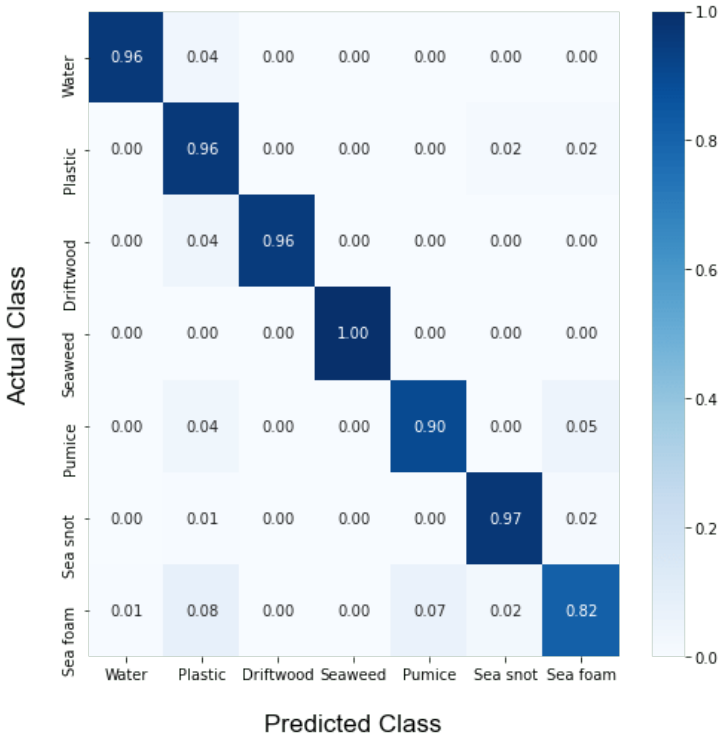

Figure 4.2: Normalised confusion matrix of the XGBoost model trained only with spectral bands.

4.2 Classification with spectral bands and spectral indices

As discussed previously, the aim of using spectral indices is to display the relative abundance of features of interest. In this case, the goal is to highlight characteristics in the materials' reflectance that allow their differentiation by the model. As Biermann et al. [20] demonstrated, using the NDVI together with the FDI allows a distinct clustering of water, plastic, driftwood, and seaweed (Figure 4.3). Nevertheless, using these indices does not enable a clear distinction between sea snot, sea foam, pumice, and plastic. This conclusion emphasizes the need for spectral indices that maximize the differences between these four classes.

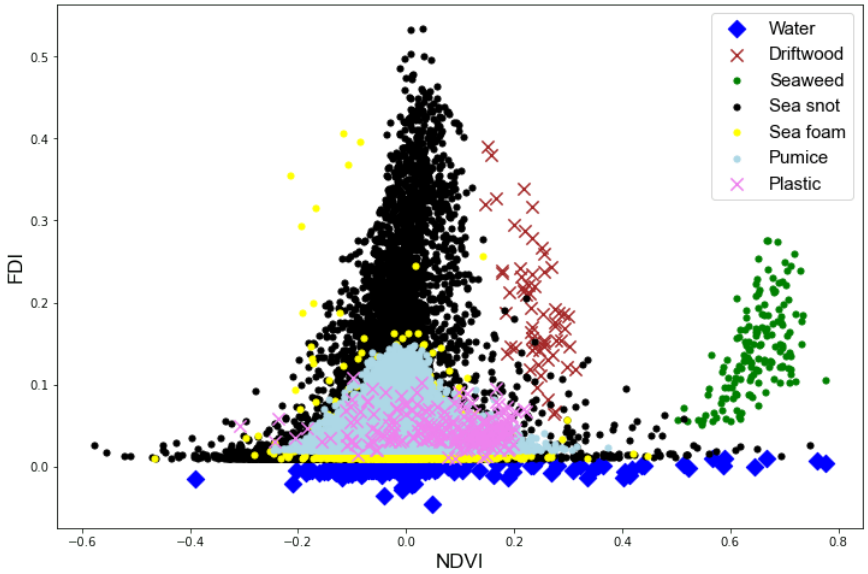

Figure 4.3: Combination of the NDVI and the FDI of all data.

In order to find the spectral indices that boost the divergences between the classes' reflectance, every classification method was trained with all spectral bands and all spectral indices described in the previous chapter: FDI (Eq. 3.1), PI (Eq. 3.2), NDVI(Eq. 3.3), RNDVI (Eq. 3.4), GNDVI (Eq. 3.5), PNDVI (Eq. 3.6), NDWI (Eq. 3.7), MNDWI (Eq. 3.8), NDMI (Eq. 3.9), NDSI (Eq. 3.10), WRI (Eq. 3.11), NBR (Eq. 3.12), AWEI (Eq. 3.13), SR (Eq. 3.14), ARI (Eq. 3.15), MARI (Eq. 3.16), CHL Red-Edge (Eq. 3.17), REPI (Eq. 3.18), EVI (Eq. 3.19), EVI2 (Eq. 3.20), MCARI (Eq. 3.21), MI (Eq. 3.22), SAVI (Eq. 3.23) and OSI (Eq. 3.24). The SVM and the Random Forest Classifier exceeded a reasonable amount of time to show results, therefore they were discarded since one of this work's goals is to find a model that is scalable to the real world and in real-world conditions the images have much more pixels, so their computational time would not be feasible. The SVM classifier's objective is to find a hyperplane in an N-dimensional space, being N the number of features, increasing the number of features directly impacts the computational time. As expected, the Random Forest classifier's main limitation is that it is

quite slow to create predictions.

The model that achieved the best results by a large margin was the XGBoost. All the other models except the Gaussian Naïve Bayes did not find patterns in the data that allowed them to distinguish the different classes (appendix B). The Gaussian Naïve Bayes classifier performed relatively well on every class except for sea foam (70%) and for plastic (74%). At the same time, the XGBoost accomplished an amazing overall accuracy. Once more, it classifies 96% of plastic pixels correctly, but reduced the number of pixels misclassified as plastic. This time, instead of seaweed, only water is completely differentiated from the other classes. Sea foam is again the class where the model has the most trouble labelling. 7% of sea foam pixels were labelled as plastic and 5% as pumice.

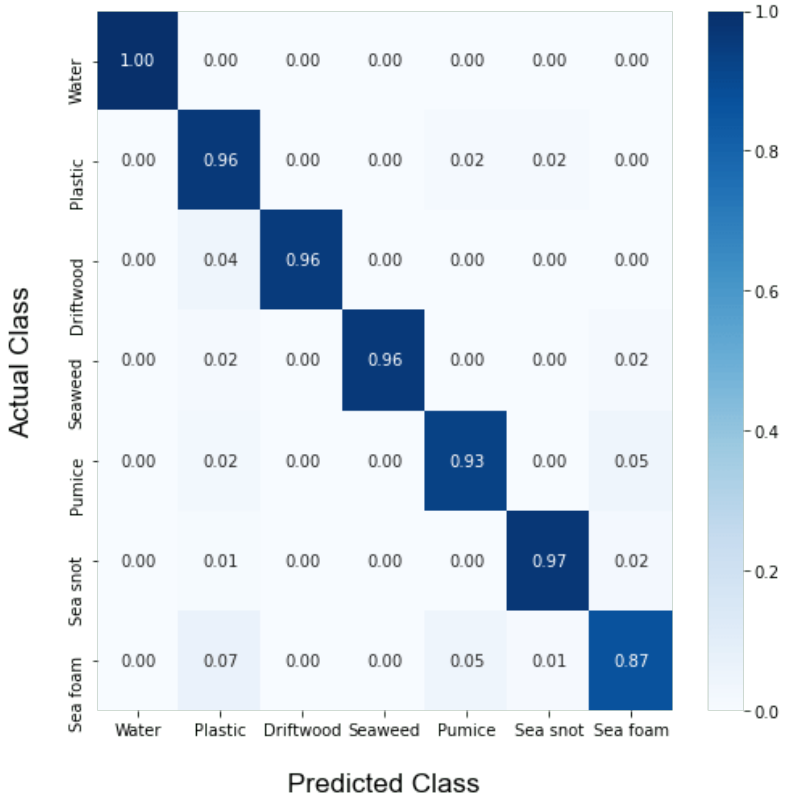

Figure 4.4: Normalised confusion matrix of the XGBoost model trained with spectral bands and all the spectral indices.

It is now possible to determine how each feature affects the overall accuracy of this model - feature importance. There are several ways to measure this metric, and one of them is using the permutation importance concept. Permutation importance focuses on answering one question: if one column of the testing set corresponding to one feature (in this case, a spectral band, or a spectral index) is randomly shuffled, while all the other columns stay intact, how would that affect the overall accuracy of the predictions? Therefore, the importance of a feature is measured by how much the loss function is affected by shuffling that feature's column. This model's feature importance is shown in Figure 4.5.

The NDSI stands out from all the other features with a great permutation importance and is followed by the MNDWI, B1 spectral band, the NDWI and the OSI. In contrast, some features reveal a negative permutation importance, meaning that the model's loss decreased when their columns were randomly shuffled. Surprisingly, one of them is the NDVI, which, apart from the FDI, was one of the most used features to detect floating plastic.

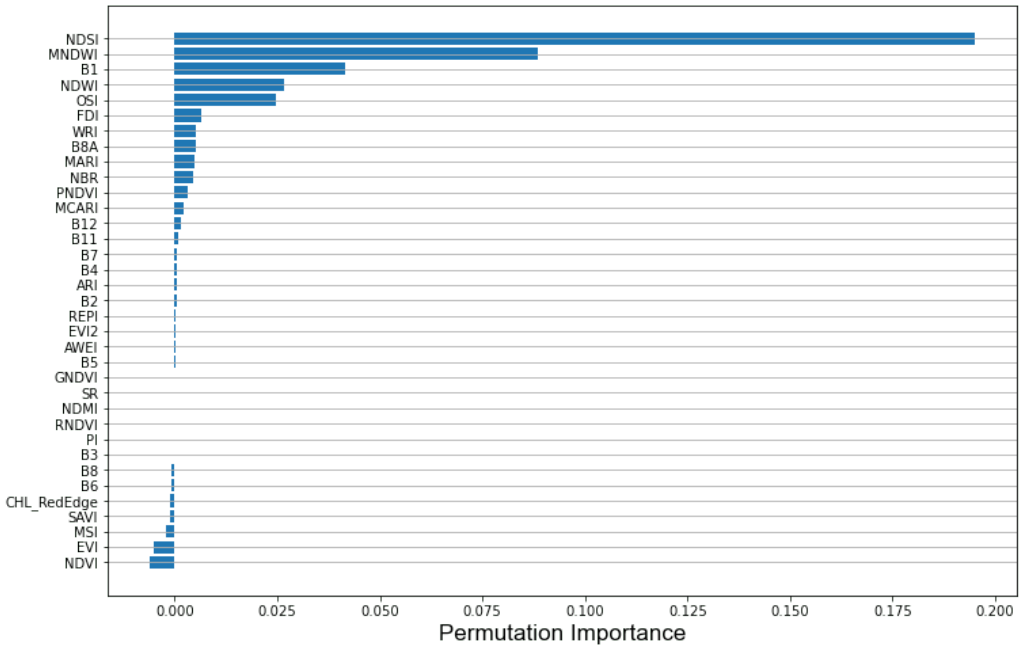

Figure 4.5: Permutation importance of each feature in the XGBoost model that was trained with all spectral bands and all spectral indices.

With this information, the irrelevant features in the model (i.e., features that showed a nil or negative permutation importance), can be removed from the training process. Then, the goal is to assess which combination of the remaining features (i.e., features that revealed a positive effect on the model's results), maximizes the overall accuracy and minimizes the number of false plastic positives. Training the XGBoost using only the nine features with the most permutation importance was the combination that achieved the best results in the testing phase (NDSI, MNDWI, B1, NDWI, OSI, FDI, WRI, B8A, and MARI). Figure 4.6 shows the normalised confusion matrix of this model, with all the numbers rounded to two decimal points. That's the cause for the sum of all values in the last line of the matrix (sea foam) being only 99%. Once again, water pixels are perfectly classified (100% true positives of water) and none of the other classes' pixels is classified as water (0% false positives of water). All the driftwood pixels are also correctly classified, but there are false positives. The model shows an accuracy above 95% for each class except for sea foam, as in the previous model. Around 13% of sea foam pixels are incorrectly labelled: 6% are predicted to be plastic and 5% pumice. The model accurately predicts 98% of plastic pixels, representing a 2% increase from the previous model. Also, the number of pix-

els incorrectly labelled as plastic decreased compared to the model trained with all features. Overall, these results indicate that a model trained with these data is ready to be applied in real-world conditions. However, when analysing the results, one must acknowledge that some seaweed, pumice, sea snot and mainly sea foam pixels will be inaccurately labelled as plastic.

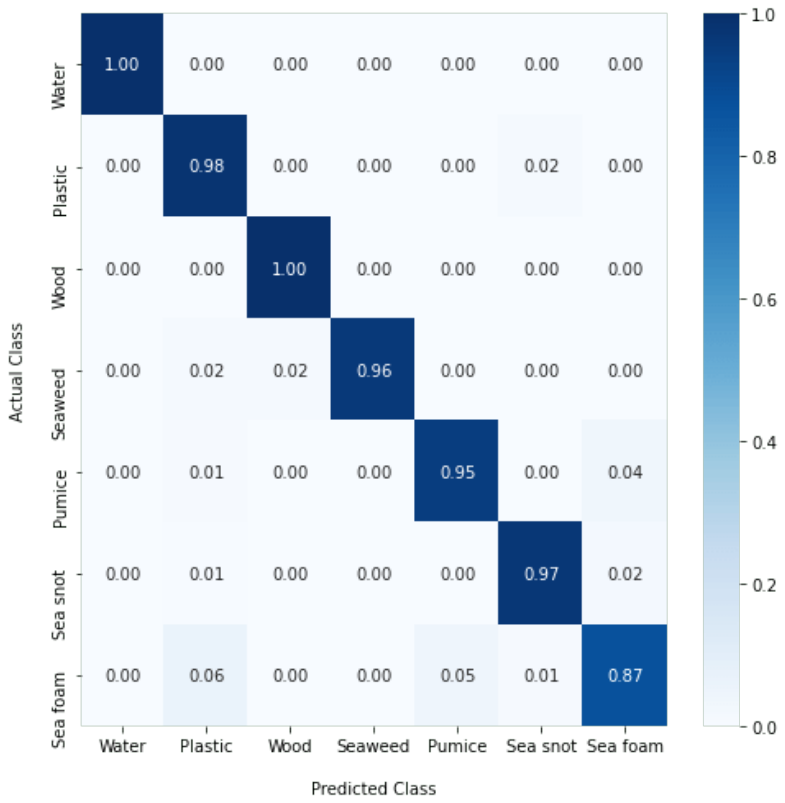

Figure 4.6: Normalised confusion matrix of the XGBoost model trained with the 9 features with most permutation importance.

4.3 Uncertainty quantification

Despite the overall good results of the previous model, two aspects must be considered. Firstly, the number of pixels wrongly classified as plastic (false positives of plastic). Although looking like a small number, a Sentinel-2 satellite image is composed of millions of pixels, so the model may wrongly predict thousands of pixels. Therefore, efforts should be made to minimize the number of false positives of plastic, knowing that it is impossible to achieve a 100% accuracy in any classification problem. Secondly, predictions where the model showed little confidence due to data uncertainty or epistemic uncertainty should be classified as another class - uncertain. This change may reduce the number of misclassified pixels and create more reliable results. However, it is also expected that implementing this modification decreases the number of pixels correctly labelled. So, just like in the GAN architecture, this situation represents a zero-sum game. On one hand, to increase the accuracy, the number of false positives of plastic increases. On the other hand, decreasing the number of false positives of plastic means decreasing the number of plastic pixels correctly classified.

To quantify uncertainty in the predictions and reach an equilibrium where the model has high accuracy in each class and a low number of false positives (Nash equilibrium), an ensemble model was created. This model is composed of twenty XGBoost models trained with different subsets of the training set, which provides variability to each prediction. For each input, the mean prediction's probability from all ensemble members and the standard deviation are computed. If the predictions' mean probability is below 90% or it has a standard deviation above 20, it is classified as uncertain. The results of the ensemble model are shown in Figure 4.7. As expected, the number of pixels correctly labelled decreased, comparing to the previous model in every class except in seaweed, which is probably related to its singular spectral signature. Around 21% of plastic pixels were labelled uncertain, as well as 16% of pumice and 30% of sea foam. Oppositely, few pixels of water, driftwood, seaweed, and sea snot were classified as uncertain, meaning that the features used to train the ensemble members allow a clear distinction between these classes. The number of pixels incorrectly classified also decreased. Now, only 2% of sea foam pixels were predicted to be plastic, representing a 4% drop from the previous results.

The best way to compare the results of both models, apart from the confusion matrix, is to use them in real-world conditions. For this purpose, a Sentinel-2A multi-spectral image from the 31^{st} of July 2021 in the Gulf of Gera, Greece, was used (Figure 4.8). On this day, the Marine Remote Sensing Group from the University of the Aegean performed another experience for the PLP 2021 [29]. They deployed two large artificial targets, one made of wood (lower target in the image), and one composed of plastic (upper target in the image), on the ocean.

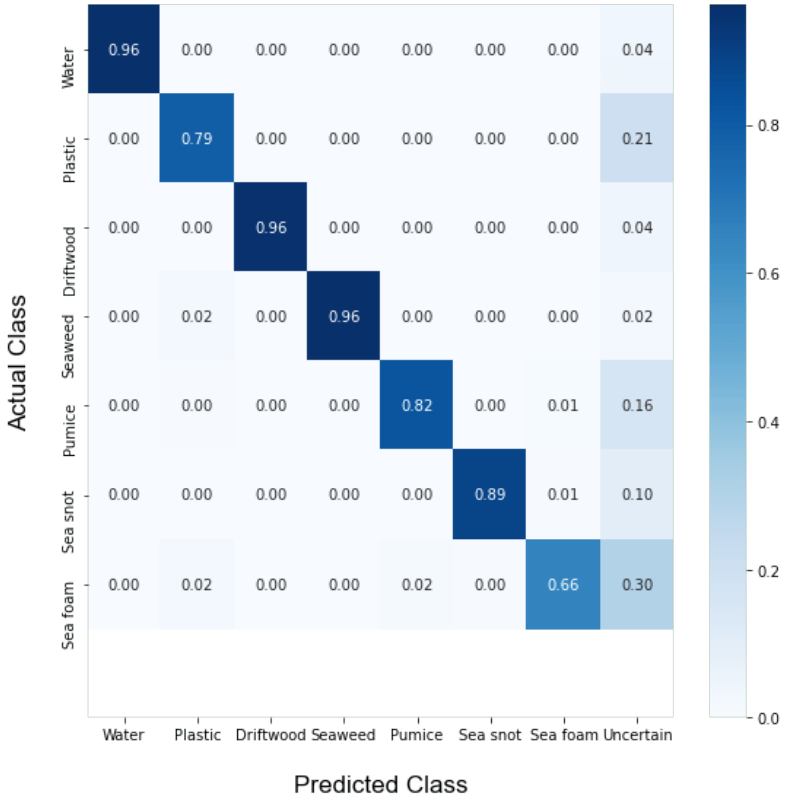

Figure 4.7: Normalised confusion matrix of the ensemble model built with 20 XGBoost models trained with different data and different data sizes, using the 9 best features. Predictions whose mean was below 90% or had a standard deviation above 20 were considered uncertain.

Figure 4.8: Sentinel-2A image of the Gulf of Gera, Greece, from the 31^{st} of July 2021, after the atmospheric correction process.

Both models were trained using a balanced dataset that had samples from both the training and the testing set. Data collected from the Sentinel-2A image previously shown were removed from the training data, meaning that both models never saw these data. The first model's predictions are shown in Figure 4.9.

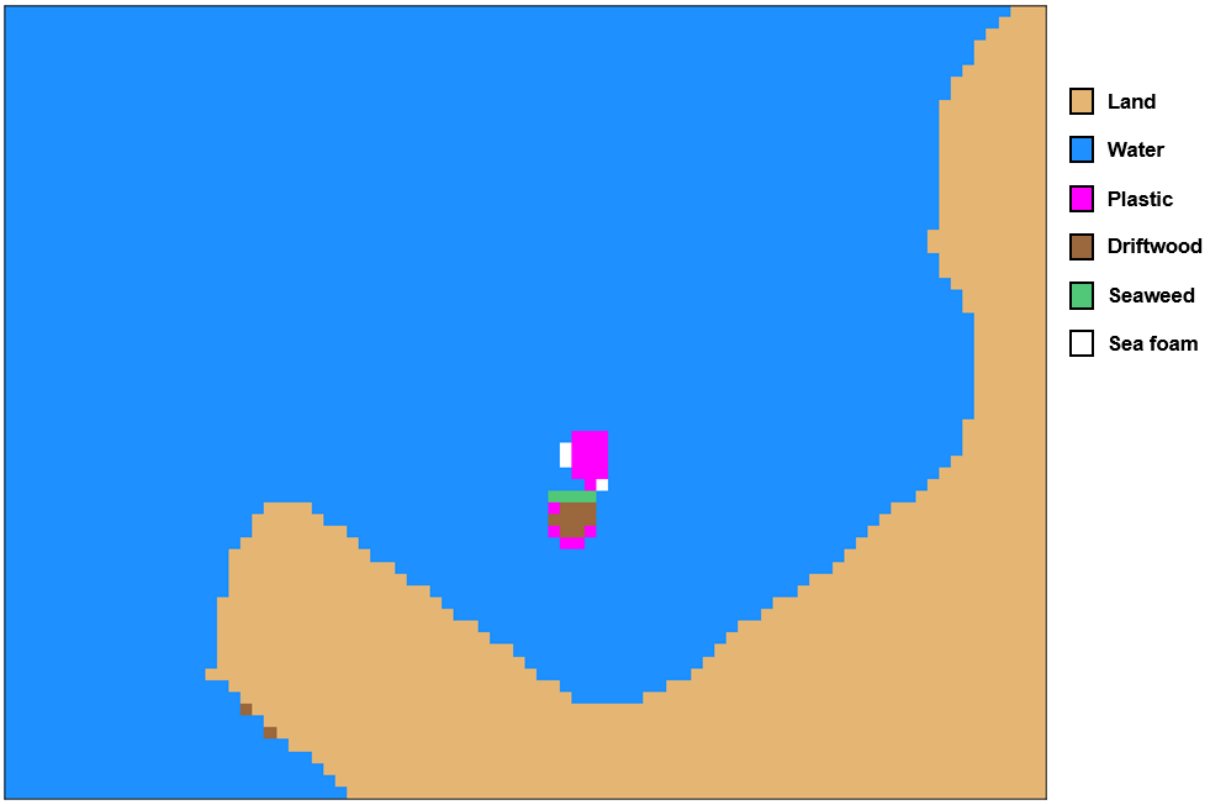

Figure 4.9: Predictions from the first model.

Results from Figure 4.9 reveal that the first model (XGBoost) is able to correctly classify most of the pixels. Both targets are detected and labelled with the right class. However, the targets' borders show some misclassifications. Some pixels around the wooden target are labelled as plastic and others as seaweed, and some pixels around the plastic target are labelled as sea foam. This indicates that when the pixel is not fully covered by a material, the probability of being misclassified is higher. There are also two pixels near the land that were classified as driftwood, which suggests that they should have been removed in the land masking process.

The ensemble model's results are shown in Figure 4.10. The model correctly predicted most of the pixels and labelled the borders of both targets as uncertain, except for one pixel that is still classified as seaweed. There are also some uncertain pixels near land that should have been removed by the land masking process. These results show that the ensemble model is the best option to deploy in the real world and the fact that it classifies some pixels as uncertain instead of labelling them incorrectly

constitutes a significant advantage compared with the previous model.

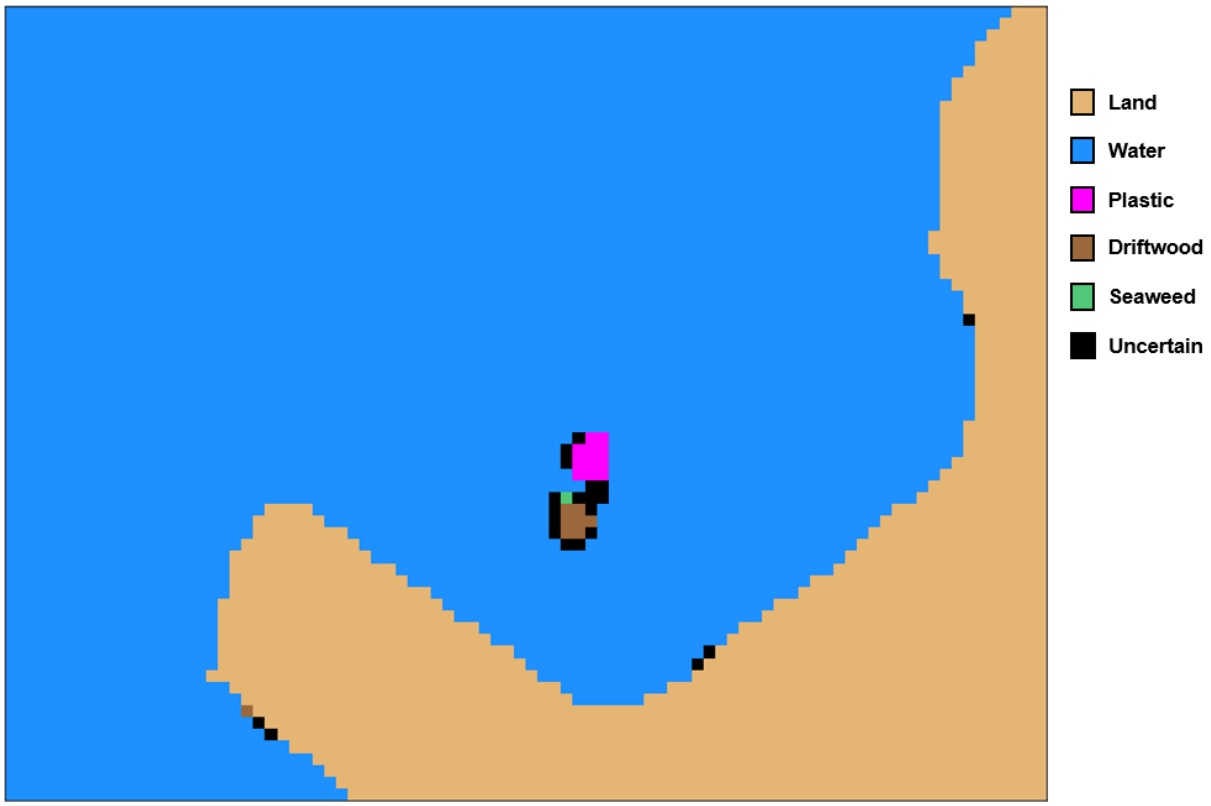

Figure 4.10: Ensemble model's predictions.

4.4 Synthetic data

Due to the low quantity of plastic and driftwood data available, one objective proposed by this study was to artificially increase the size of the datasets using data augmentation techniques. Theoretically, training a model with a larger balanced dataset with variability would make it more robust and less sensitive to outliers and mislabelled training data. The first method used to generate data was through a Generative Adversarial Network. However, some problems arised. Firstly, adjusting its parameters took a lot of time, then, sometimes it did not converge and, finally, it suffered mode collapse. Therefore, as expected, an XGBoost model trained with this dataset did not perform well. Its confusion matrix indicates that the model cannot classify data whose reflectance is far from the reflectance mean of any floating class. The second method evaluated to create artificial data was the Wasserstein GAN, that uses a different approach to train the network. This strategy produced data with closer spectral signatures to the original data and with higher variability. The normalised confusion matrix of an XGBoost model trained with these data is shown in Figure 4.11. The model struggles in differentiating water, pumice, and sea snot from

sea foam. It shows some success in classifying pixels from plastic, driftwood, and seaweed (classes that have the greater similarities with original data), but with many false positives. These results indicate that synthetic data generated from a WGAN have too much variability, so the model cannot find the characteristic patterns of water, pumice, sea snot, and sea foam. On the other hand, the classification results for plastic, driftwood and seaweed are promising, since they show that when synthetic data have characteristics close to the original data, they allow for acceptable results. However, this model is far from good enough to be deployed in real-world conditions.

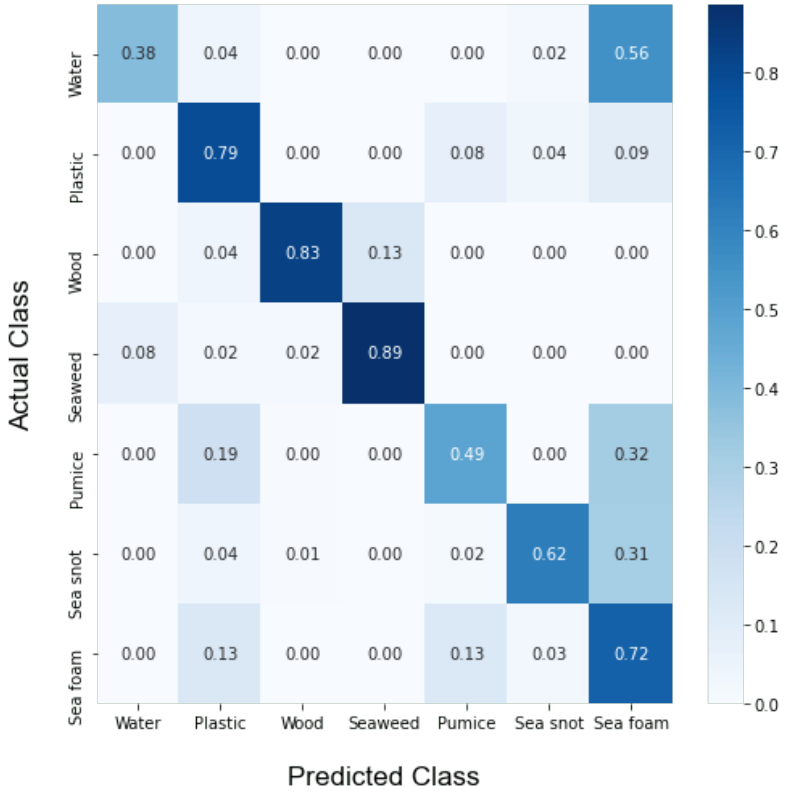

Figure 4.11: Normalised confusion matrix of an XGBoost model trained with 40000 synthetic pixels from each class (280000 pixels in total) generated from a WGAN.

4.4.1 Quality control method for synthetic data

There are two solutions to improve the previous model trained with synthetic data. The best one is generating new artificial data by changing the WGAN's hyperparameters until it produces data indistinguishable from the original, just like the driftwood and seaweed data. However, this process has a high computational cost. The second alternative is to filter these synthetic data samples, meaning that data that do not accurately represent a class are discarded.

The second approach was applied to the previously generated data, using the ensemble model that quantifies uncertainty as a filter. The generated data was used as a testing set. If a synthetic pixel was

classified as uncertain (i.e., the computed probability of being from one class was lower than 90%, or had a standard deviation higher than 20), the pixel was discarded. The pixel was also rejected if the ensemble model misclassified it (i.e., if it was a synthetic pixel from one class and the model classified it as other class). Therefore, in the end, there are only synthetic pixels that were correctly classified and were not considered uncertain by the ensemble model: 19342 water pixels, 10742 plastic pixels, 10933 driftwood pixels, 29484 seaweed pixels, 2834 pumice pixels, 8280 sea snot pixels, and 1434 sea foam pixels. To create a balanced dataset to train a model, only 1434 pixels from each class were chosen, since it was the lowest number of pixels from one class (sea foam) that passed the filtering. The confusion matrix of an XGBoost model trained with this subset is shown in Figure 4.12. The results improved as every class had an increase in the number of true positives and a decrease in false positives, which proves that a model trained only with synthetic data can achieve successful results. Plastic and sea foam are still the classes with the highest number of false positives, representing the biggest disadvantage compared to the model trained exclusively with authentic data.

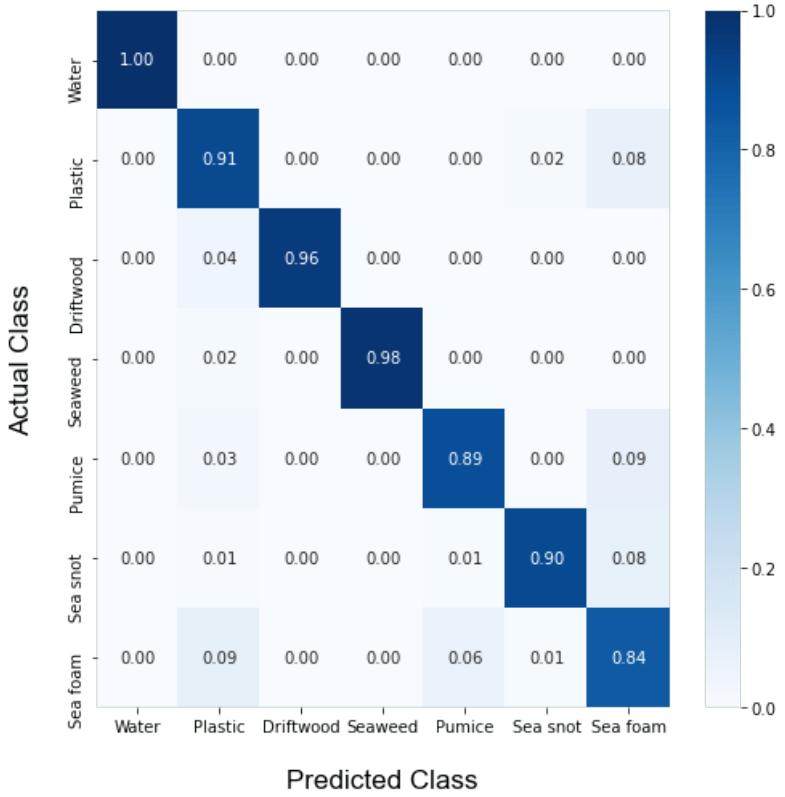

Figure 4.12: Normalised confusion matrix of an XGBoost model trained with 1434 synthetic pixels from each class (10038 pixels in total) generated from a WGAN and that passed the quality filter.

4.5 Using the best model to monitor the ocean

In this section, the best model (ensemble model with 20 ensemble members trained with different subsets of authentic data) is tested in real-world conditions, to assess its feasibility.

4.5.1 Plastic target mixed with wood - Gulf of Gera, Greece

A Sentinel-2B multi-spectral image from the 4^{th} of September 2021, in the Gulf of Gera, Greece, was used - Figure 4.13. On this day, the group from the University of Aegean performed another experience for the PLP 2021 [29]. They removed the wooden target and deployed some of it under the plastic target, simulating a mixed target which is closer to what is found in the ocean. This prevented the use of this image in the training dataset.

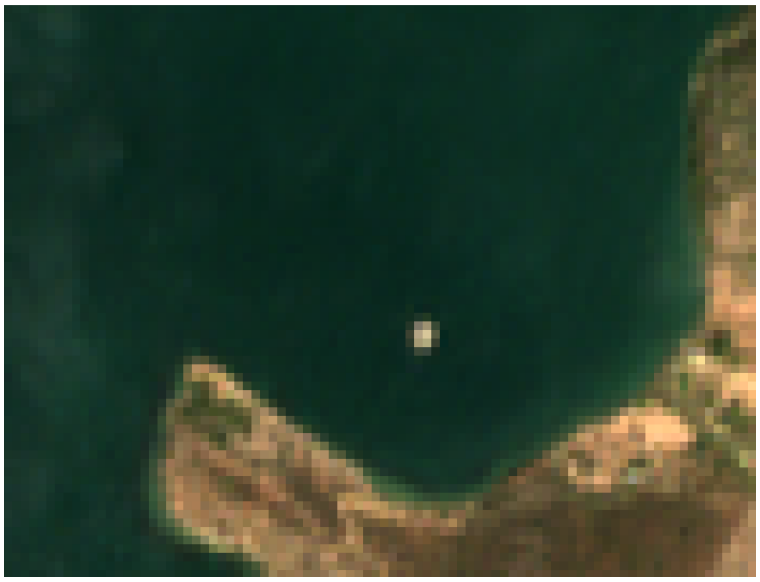

Figure 4.13: Sentinel-2B image of the Gulf of Gera, Greece, from the 4^{th} of September 2021, after the atmospheric correction process.

Figure 4.14 shows the outputs of the ensemble model. It successfully detected the plastic target. However, some of its pixels were considered uncertain, which is probably related to the presence of wood. There are also some pixels classified as plastic near the land that should have been removed in the land masking process, and some uncertain pixels in the open ocean that may be related to sun glint or wave agitation.

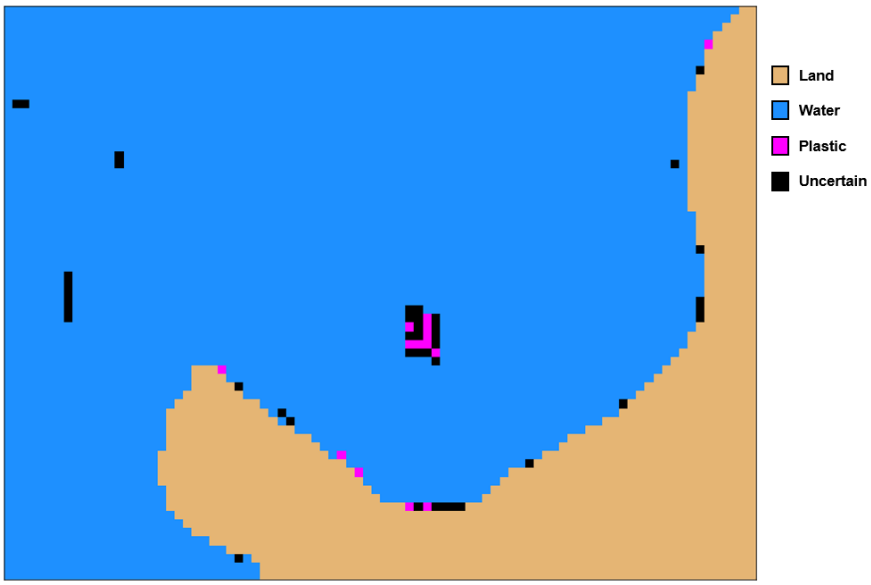

Figure 4.14: Model's predictions based on the previous Sentinel-2B image, from the 4^{th} of September 2021.

4.5.2 Seaweed - Playa del Carmen, Mexico

Every year, beaches in Mexico suffer a massive invasion of Sargassum seaweed because of the water temperature, the increase of nutrients in the water, marine currents, and wind. Playa del Carmen is one of the most affected places. An article from National Geographic on the 4^{th} of July 2021 exposed the problem [75]. So, one Sentinel-2B image from the following day, on this location, was used to test the ensemble model (Figure 4.15), where accumulations of a green material can be seen really close to land and also some agglomerations further away from the shore.

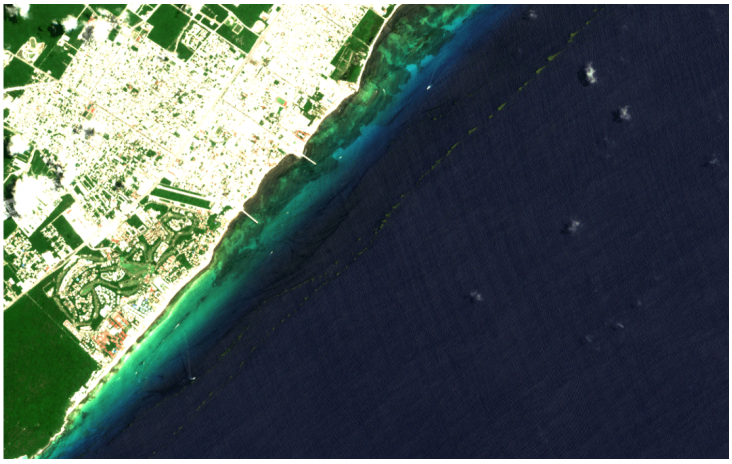

Figure 4.15: Sentinel-2B image of Playa del Carmen, Mexico, from the 5^{th} of July 2021, after the atmospheric correction process.

The model only detected the seaweed accumulations that were further away from the shore (Figure 4.16). Once again, the debris' borders were classified as uncertain, meaning that the predictions depend on the percentage of material in a pixel. Most of the seaweed accumulations that were near the shore were classified as water, probably because of the depth of the water that produced a higher reflectance in the pixels, so the model confused them with shallow waters.

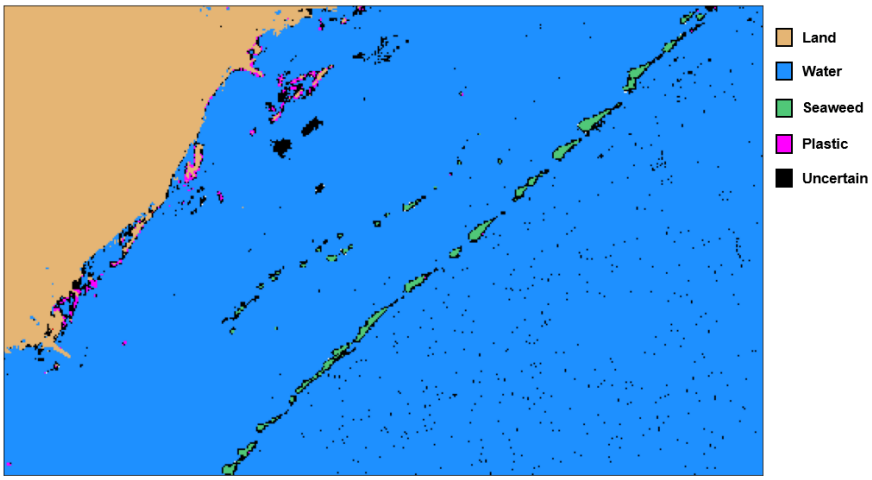

Figure 4.16: Model's predictions based on part of the previous Sentinel-2B image, from the 5^{th} of July 2021.

4.5.3 Pumice - Okinawa, Japan

As described in the data acquisition section, in October 2021, pumice originated from a submarine volcanic eruption invaded the Okinawa coastline. Here, the model is tested with the same Sentinel-2A image, but none of the training data were collected from this specific sub-image - Figure 4.17.

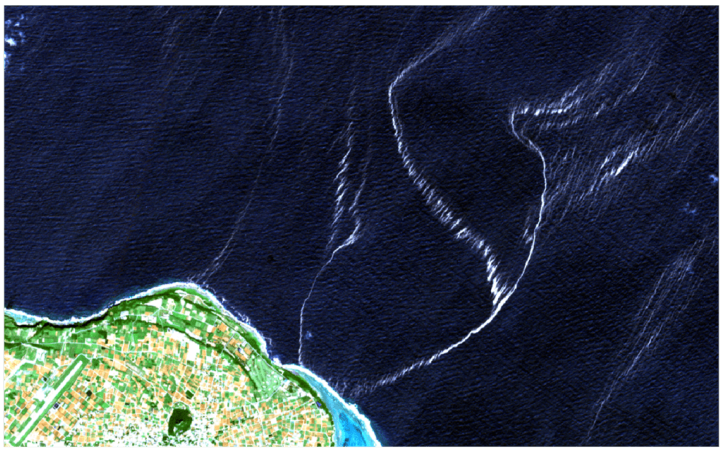

Figure 4.17: Sentinel-2A image of the coast of Le, Okinawa, Japan, from the 26^{th} of October 2021, after the atmospheric correction process.

The model detects the large accumulations of pumice and, once again, it classifies their borders as uncertain – Figure 4.18. There are more uncertain pixels than the previous predictions, which may be related to rough sea conditions. The model classified some pixels near the land that appear to be waves crashing on the shore as sea snot, instead of sea foam.

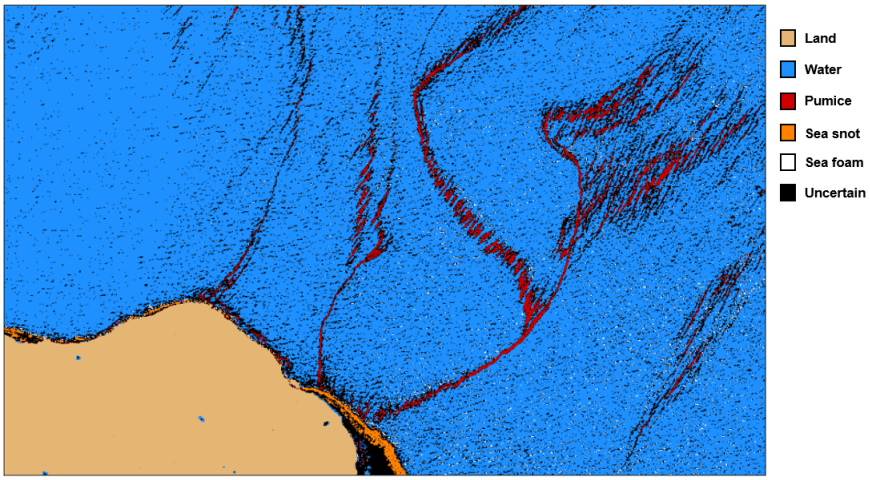

Figure 4.18: Model's results based on the previous Sentinel-2A image, from the 26^{th} of October 2021.

4.5.4 Sea snot - Marmara Sea, Coast of Istanbul, Turkey

Here, the model is used to study the evolution of sea snot accumulation on the coast of Istanbul. This phenomenon first appeared at the beginning of May 2021. Figure 4.19 shows a Sentinel-2B from June 2021, where sea snot aggregations with several metres wide can be seen. The image also indicates rough water conditions.

Figure 4.19: Sentinel-2B image of the coast of Istanbul, Turkey, from the 13^{th} of June 2021, after the atmospheric correction process.

The first predictions, shown in Figure 4.20, are based on a Sentinel-2A image from April 2021. The model does not detect any accumulation of any material. There are only some uncertain pixels caused by vessels or water movements.

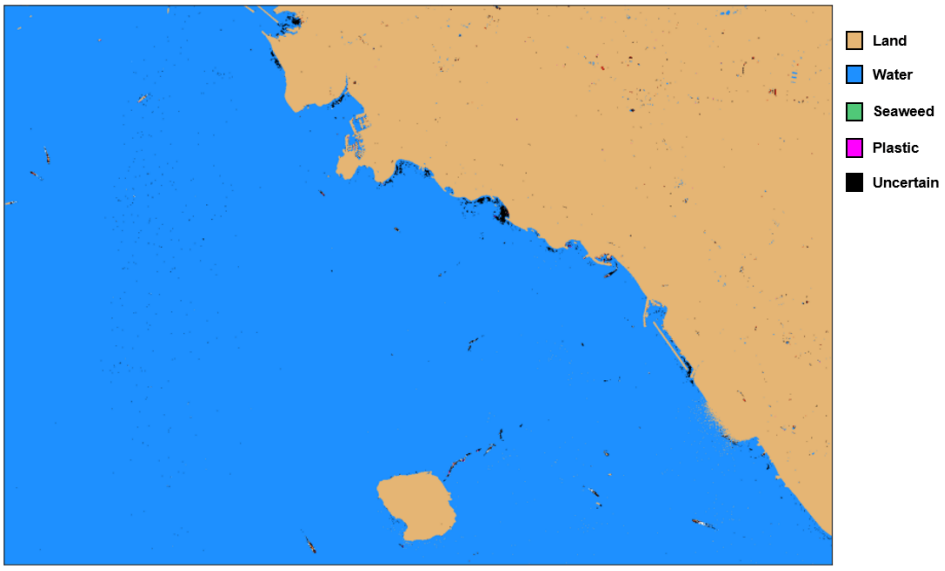

Figure 4.20: Model's predictions based on a Sentinel-2A image from the 2^{nd} of April 2021.

Contrastingly, predictions from May 2021, shown in Figure 4.21, indicate the presence of large sea snot accumulations. There are also many pixels classified as uncertain, which is probably related to the substance's depth on each pixel. Some pixels were inaccurately labelled as plastic.

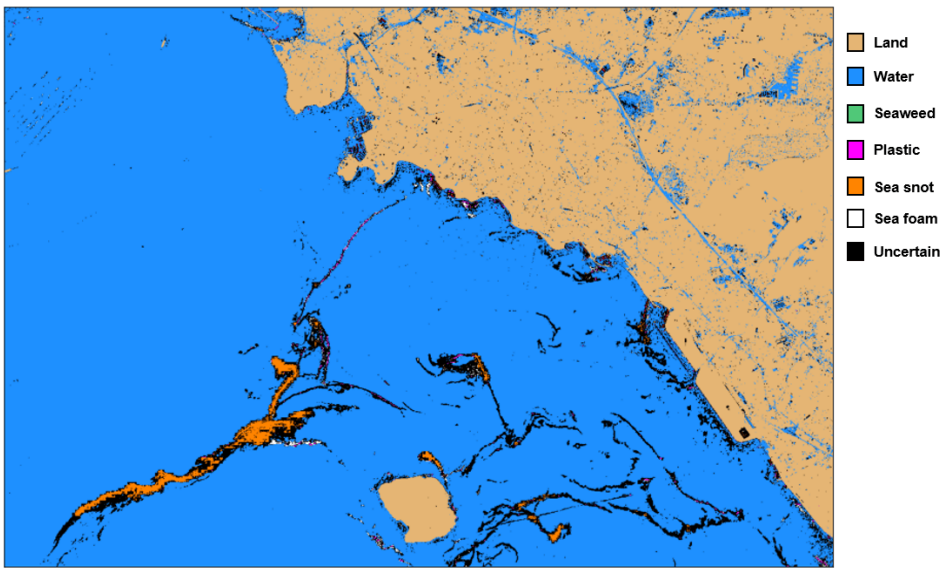

Figure 4.21: Model's predictions based on a Sentinel-2B image from the 14^{th} of May 2021.

Predictions from June 2021, in Figure 4.22, show even more sea snot accumulations than May 2021. There are also many pixels classified as uncertain, which has two causes. Firstly, different sea snot depths in each pixel. Then, as seen in Figure 4.19, the water shows a lot of movement.

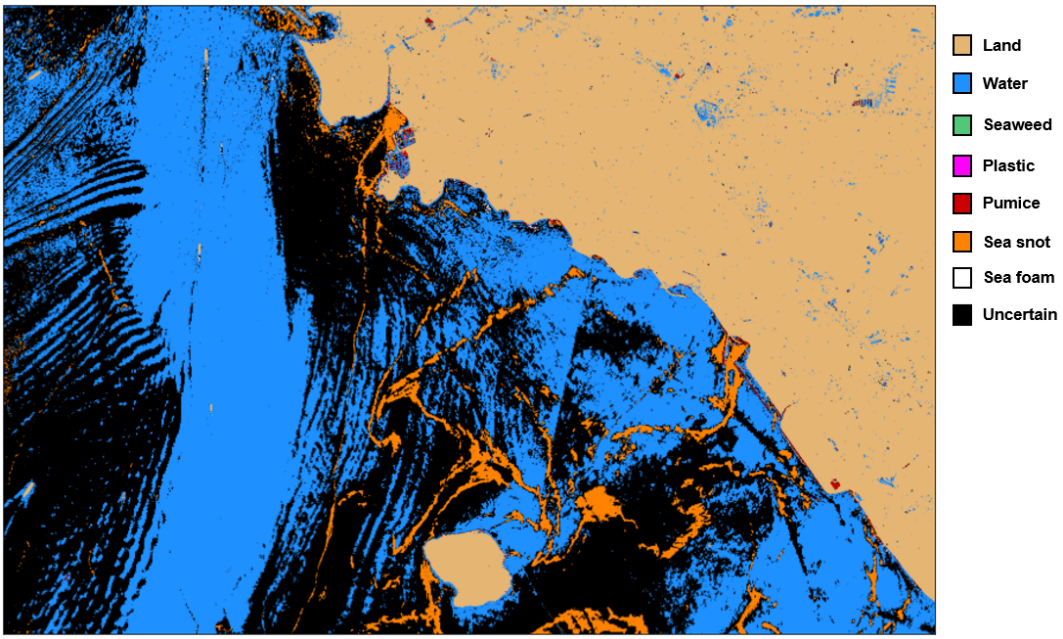

Figure 4.22: Model's predictions based on a Sentinel-2B image from the 13^{th} of June 2021.

4.5.5 Limitations

The previous results highlight some of the model's limitations. Most of the misclassifications happened in pixels very close to the shore, where reflectance is usually higher because of the lower water depth, which indicates that the model produces more reliable results in deeper waters. As expected, rough waters cause uncertainty in the predictions. Therefore, sea conditions should be considered when analysing the model's predictions. Another source of uncertainty is associated with the quantity of material in a pixel. Topouzelis et al. [19] concluded that for plastic to be detected on a Sentinel-2 pixel, it must cover at least as low as 25% of it. However, more data need to be collected and studied to know the specific percentage that should vary from one class to another. Here, the model regularly classifies the borders of the materials, where there is a less quantity of the floating class, as uncertain. There is also a common challenge in every study that uses satellite data: clouds. Clouds spoil the satellite data even if they are not dense, since they reflect sunlight, preventing the computation of predictions if above the area of interest. Finally, sediments in suspension also have a negative impact on the predictions. This usually happens in rivers, the primary conduits for plastic waste to the sea. The model was applied at the mouth of three rivers that contribute the most to plastic pollution in the ocean: the Pasig River in the

Philippines, the Yangtze River in China, and the Angke River in Jakarta, Indonesia (Figure 4.23), but it was not possible to achieve any conclusion because of the quantity of sediments in the water. As shown in Figure 4.24, the model cannot compute accurate predictions because of the quantity of sediments in the water.

Figure 4.23: Sentinel-2A image of the Angke River's mouth, Jakarta, Indonesia, from the 24^{th} of April 2021.

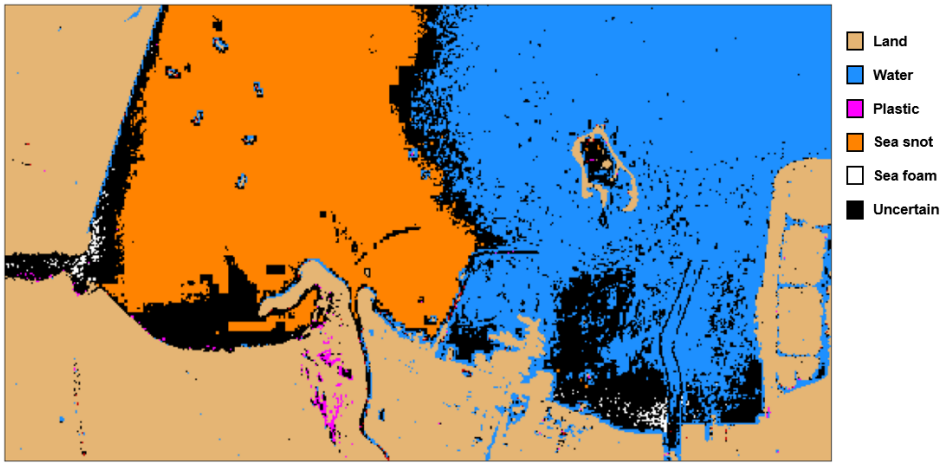

Figure 4.24: Model's predictions based on the previous Sentinel-2A image, from the 24^{th} of April 2021.

4.6 Discussion

The study's results constitute a breakthrough in this area for multiple reasons. Firstly, the data acquisition process allowed the collection of the largest dataset related to floating plastics in satellite imagery that was ever published (available at [76]). Then, it is also the first study that assesses and proves that plastic is distinguishable from five other classes of floating debris and water. Hu et al. [18] declared that remote differentiation of sea snot and marine debris using multi-spectral imagery is problematic due to their spectral similarity. An XGBoost model trained with only the Sentinel-2 spectral bands as features classified 96% of plastic pixels and 97% of sea snot pixels correctly, which proves that both classes are differentiable. In addition, this work also evaluates if using spectral indices as features improves the classification results. The answer is positive, but it depends on the indices used. Previous works used the spectral bands, the NDVI and the FDI as features to train their models and achieved significant results [20, 32]. Here, it is proven that plastics, driftwood, seaweed, and water form distinct clustering when using the NDVI combined with the FDI, allowing the differentiation. Nevertheless, this combination does not separate sea snot, sea foam, pumice, and plastic. The combination of features that proved most successful was using the NDSI, the MNDWI, spectral band B1, the NDWI, the OSI, the FDI, the WRI, spectral band B8A and the MARI. An XGBoost model trained with these features showed an amazing overall accuracy. All the pixels of water and driftwood were classified correctly, as well as 98% of plastic pixels, 96% of seaweed pixels, 95% of pumice pixels, 97% of sea snot pixels, and 87% of sea foam pixels. Despite the high accuracy values, for a model to be deployed in real-world conditions and provide meaningful information for the decision-makers, the percentage of misclassifications, in particular, the number of false positives of plastic labels, must be minimized. Therefore, it was used an uncertainty quantification strategy: an Ensemble method. The ensemble model achieved lower percentages of correct classifications, but also, and most importantly, it decreased the number of misclassifications. Deploying the model in real-world conditions confirmed the good results, but also showed some of its limitations: clouds, shallower waters, sediments in suspension, and rough sea conditions. The study also proved that training a model with synthetic data produces good results, however, not as good as the model trained with authentic data.

The model created in this study shows incredible potential. Its strengths can be maximized by using other systems in parallel. For example, a numerical model estimates the location of plastic debris, then the model proposed in this study checks if there's anything in that location and, finally, a marine robot or an unmanned aerial vehicle confirms the presence of debris, enabling the decision-makers to act and tackle the problem. The model can also study the evolution of a substance in a specific location, for instance, the expansion of sea snot in the Marmara Sea.

5

Conclusion

Contents

This thesis focused on creating a machine learning model that detects and distinguishes several types of floating materials on the ocean using Sentinel-2 satellite imagery. This work demonstrates that applying an uncertainty quantification method to the predictions creates more reliable results. Additionally, the thesis shows that lack of satellite data regarding floating plastics can be addressed with synthetic data, but with some limitations.

Chapter 3 introduced the satellite used and its characteristics, such as the spatial resolution, the radiometric resolution, and the temporal resolution. Also, the steps taken to pre-process its data were explained. First, ACOLITE software was used to perform the atmospheric correction to remove the contribution of the atmosphere from the reflectance values by using the Dark Spectrum Fitting algorithm, then the land mask process used the Normalised Difference Water Index to remove nonessential data. Furthermore, the spectral signatures of the seven classes collected in the data acquisition step were shown, as well as the spectral indices used, which are mathematical combinations of spectral bands to highlight features of interest. Also in chapter 3, neural networks were introduced to explain the concept of Generative Adversarial Networks. Generative Adversarial Networks were then used to generate synthetic data from all the classes. However, these data did not show variability due to one common problem in GANs: mode collapse. Therefore, an alternative method for generating data using GANs was introduced: Wasserstein Generative Adversarial Network. This method allowed to create 280000 pixels that showed spectral responses close to the original data. Moreover, in this chapter, seven different classification algorithms were described.

In chapter 4, the different classification algorithms were tested using only the spectral bands as features. Despite the best model reaching an accuracy of over 90% for every class except sea foam, some pixels were misclassified as plastic. The addition of spectral indices as features improved the results. Additionally, the selection of the nine most important indices and bands maximized the accuracies and reduced the number of misclassifications as plastic. The best model, an XGBoost trained with the NDSI, the MNDWI, spectral band B1, the NDWI, the OSI, the FDI, the WRI, spectral band B8A and the MARI, classified correctly 98% of the plastic pixels. This model was then used to create an ensemble model that quantifies uncertainty. Pixels where the ensemble model showed some distrust were labelled as uncertain, decreasing the number of misclassifications. The uncertainty quantification method allowed to minimize the number of false positives of plastic, while maintaining relatively high accuracies for all classes, meaning that this was the best model to apply in real-world conditions. The application of this model on several Sentinel-2 images revealed overall excellent results, but also showed its limitations. Most of its misclassifications occur in shallower waters, when there are clouds, sediments in suspension or rough sea conditions.

Furthermore, in chapter 4, the results of training the best algorithm with synthetic data revealed some limitations. However, after filtering the synthetic data, the model performed almost at the same level as

the model trained with real data, which indicates the potential of using synthetic data when authentic data are scarce.

5.1 Future work

One of the main limitations regarding the detection of floating plastic in satellite imagery is the lack of in situ data, since the best detection methods rely on supervised learning approaches. Therefore, there is a need for more plastic data to be collected globally, whether via artificial targets or natural occurrences. Future studies should focus on collecting data from other floating substances to determine if they have spectral signatures similar to plastic.

Future research should also focus on maximizing the qualities of this work's model by creating parallel systems. For example, numerical models could be used to indicate areas of study (appendix A). The same could be done by models that focus on the spatial characteristics of floating materials through satellite imagery (e.g. CNNs). Marine robots or unmanned aerial vehicles could also be deployed in those locations to confirm the model's classifications. Furthermore, it would be interesting to compare how the different atmospheric correction methods affect the detection of floating plastics. Finally, it would be relevant to assess if Sentinel-2 imagery can detect floating debris in rivers, since they are the main points of entry of plastic in the ocean.

Bibliography

[1] "Measurement #12 10.08.2021 – PLP," accessed 13-May-2022. [Online]. Available: http://plp.aegean.gr/2021/08/10/measurement-12-10-08-2021/

[2] "Okinawa's floating pumice mass likely to disappear by June, experts say," The Japan Times, accessed 13-May-2022. [Online]. Available: https://www.japantimes.co.jp/news/2022/02/14/national/okinawa-pumice-stones-decline/

[3] A. Taylor, "Photos: Turkey's Sea-Snot Disaster - The Atlantic," www.theatlantic.com, accessed 13-May-2022. [Online]. Available: https://www.theatlantic.com/photo/2021/06/photos-turkeys-sea-snot-disaster/619254/#img20

[4] M. C. Sousa, M. deCastro, J. Gago, A. S. Ribeiro, M. Des, J. L. Gómez-Gesteira, J. M. Dias, and M. Gomez-Gesteira, "Modelling the distribution of microplastics released by wastewater treatment plants in Ria de Vigo (NW Iberian Peninsula)," *Marine Pollution Bulletin*, vol. 166, 2021.

[5] "Radiometric - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online - Sentinel Online," sentinels.copernicus.eu, accessed 12-May-2022. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric

[6] R. C. Thompson, S. H. Swan, C. J. Moore, and F. S. vom Saal, "Our plastic age," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, pp. 1973–1976, 2009.

[7] I. E. Napper and R. C. Thompson, "Plastic Debris in the Marine Environment: History and Future Challenges," *Global Challenges*, vol. 4, no. 6, 2020.

[8] L. Barboza, A. Cózar, B. Gimenez, T. Lima Barros, P. Kershaw, and L. Guilhermino, "Macroplastics Pollution in the Marine Environment," in *World Seas: An Environmental Evaluation*, second edition ed., C. Sheppard, Ed. Academic Press, 2018, vol. 3, ch. 17, pp. 305–328.

[9] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Science Advances*, vol. 3, no. 7, 2017.

[10] J. R. Jambeck, R. Geyer, C. Wilcox, T. R. Siegler, M. Perryman, A. Andrady, R. Narayan, and K. L. Law, "Plastic waste inputs from land into the ocean," *Science*, vol. 347, no. 6223, pp. 768–771, 2015.

[11] L. Lebreton, B. Slat, F. Ferrari, B. Sainte-Rose, J. Aitken, R. Marthouse, S. Hajbane, S. Cunsolo, A. Schwarz, A. Levivier, K. Noble, P. Debeljak, H. Maral, R. Schoeneich-Argent, R. Brambini, and J. Reisser, "Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic," *Scientific Reports*, vol. 8, no. 4666, 2018.

[12] E. J. Carpenter and K. L. Smith, "Plastics on the Sargasso Sea Surface," *Science*, vol. 175, no. 4027, pp. 1240–1241, 1972.

[13] L. E. Haram, J. T. Carlton, L. Centurioni, M. Crowley, J. Hafner, N. Maximenko, C. C. Murray, A. Y. Shcherbina, V. Hormann, C. Wright, and G. M. Ruiz, "Emergence of a neopelagic community through the establishment of coastal species on the high seas," *Nature Communications*, vol. 12, no. 6885, 2021.

[14] J. T. Carlton, J. W. Chapman, J. B. Geller, J. A. Miller, D. A. Carlton, M. I. McCuller, N. C. Treneman, B. P. Steves, and G. M. Ruiz, "Tsunami-driven rafting: Transoceanic species dispersal and implications for marine biogeography," *Science*, vol. 357, no. 6358, pp. 1402–1406, 2017.

[15] N. J. Beaumont, M. Aanesen, M. C. Austen, T. Börger, J. R. Clark, M. Cole, T. Hooper, P. K. Lindeque, C. Pascoe, and K. J. Wyles, "Global ecological, social and economic impacts of marine plastic," *Marine Pollution Bulletin*, vol. 142, pp. 189–195, 2019.

[16] Y. C. Jang, S. Hong, J. Lee, M. J. Lee, and W. J. Shim, "Estimation of lost tourism revenue in Geoje Island from the 2011 marine debris pollution event in South Korea," *Marine Pollution Bulletin*, vol. 81, no. 1, pp. 49–54, 2014.

[17] A. P. Krelling, A. T. Williams, and A. Turra, "Differences in perception and reaction of tourist groups to beach marine debris that can influence a loss of tourism revenue in coastal areas," *Marine Policy*, vol. 85, pp. 87–99, 2017.

[18] C. Hu, L. Qi, Y. Xie, S. Zhang, and B. B. Barnes, "Spectral characteristics of sea snot reflectance observed from satellites: Implications for remote sensing of marine debris," *Remote Sensing of Environment*, vol. 269, 2022.

[19] K. Topouzelis, D. Papageorgiou, A. Karagaitanakis, A. Papakonstantinou, and M. Arias Ballesteros, "Remote Sensing of Sea Surface Artificial Floating Plastic Targets with Sentinel-2 and Unmanned Aerial Systems (Plastic Litter Project 2019)," *Remote Sensing*, vol. 12, no. 12, 2020.

[20] L. Biermann, D. Clewley, V. Martinez-Vicente, and K. Topouzelis, “Finding Plastic Patches in Coastal Waters using Optical Satellite Data,” *Scientific Reports*, vol. 10, p. 1–10, 2020.

[21] G. Suaria and S. Aliani, “Floating debris in the Mediterranean Sea,” *Marine Pollution Bulletin*, vol. 86, no. 1, pp. 494–504, 2014.

[22] G. Gonçalves, U. Andriolo, L. Gonçalves, P. Sobral, and F. Bessa, “Quantifying Marine Macro Litter Abundance on a Sandy Beach Using Unmanned Aerial Systems and Object-Oriented Machine Learning Methods,” *Remote Sensing*, vol. 12, no. 16, 2020.

[23] S. Kako, A. Isobe, T. Kataoka, K. Yufu, S. Sugizono, C. Plybon, and T. A. Murphy, “Sequential webcam monitoring and modeling of marine debris abundance,” *Marine Pollution Bulletin*, vol. 132, pp. 33–43, 2018.

[24] L. Fronkova, “Tackling Marine Litter in the Atlantic Area,” Clean Atlantic, 2019.

[25] K. Topouzelis, D. Papageorgiou, G. Suaria, and S. Aliani, “Floating marine litter detection algorithms and techniques using optical remote sensing data: A review,” *Marine Pollution Bulletin*, vol. 170, no. 112675, 2021.

[26] W. Pichel, T. Veenstra, J. Churnside, E. Arabini, K. Friedman, D. Foley, R. Brainard, D. Kiefer, S. Ogle, P. Clemente-Colón, and X. Li, “GhostNet marine debris survey in the Gulf of Alaska - Satellite guidance and aircraft observations,” *Marine pollution bulletin*, vol. 65, pp. 28–41, 2011.

[27] T. Aoyama, “Extraction of marine debris in the Sea of Japan using high-spatial-resolution satellite images,” in *Remote Sensing of the Oceans and Inland Waters: Techniques, Applications, and Challenges*, R. J. Frouin, S. C. Shenoi, and K. H. Rao, Eds., vol. 9878, International Society for Optics and Photonics. SPIE, 2016, pp. 213–219.

[28] K. Topouzelis, A. Papakonstantinou, and S. P. Garaba, “Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018),” *International Journal of Applied Earth Observation and Geoinformation*, vol. 79, pp. 175–183, 2019.

[29] M. R. S. Group, “Plastic Litter Project 2021,” accessed 13-May-2022. [Online]. Available: http://plp.aegean.gr/plastic-litter-project-2021/

[30] K. Themistocleous, C. Papoutsa, S. Michaelides, and D. Hadjimitsis, “Investigating Detection of Floating Plastic Litter from Space Using Sentinel-2 Imagery,” *Remote Sensing*, vol. 12, no. 16, 2020.

[31] A. Kikaki, K. Karantzalos, C. A. Power, and D. E. Raitsos, “Remotely Sensing the Source and Transport of Marine Plastic Debris in Bay Islands of Honduras (Caribbean Sea),” *Remote Sensing*, vol. 12, no. 11, 2020.

[32] B. Basu, S. Sannigrahi, A. Sarkar Basu, and F. Pilla, “Development of Novel Classification Algorithms for Detection of Floating Plastic Debris in Coastal Waterbodies Using Multispectral Sentinel-2 Remote Sensing Imagery,” *Remote Sensing*, vol. 13, no. 8, 2021.

[33] A. Ciappa, “Marine plastic litter detection offshore Hawai’i by Sentinel-2,” *Marine Pollution Bulletin*, vol. 168, no. 112457, 2021.

[34] J. Mifdal, N. Longépé, and M. Rußwurm, “Towards detecting floating objects on a global scale with learned spatial features using Sentinel-2,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, pp. 285–293, 2021.

[35] M. Arias, R. Sumerot, J. Delaney, F. Coulibaly, A. Cózar, S. Aliani, G. Suaria, T. Papadopoulou, and P. Corradi, “Mapping Windrows as Proxies for Marine Litter Monitoring from Space (WASP).” EGU General Assembly 2021, 2021.

[36] E. S. Agency, “Radiometric Resolutions - Sentinel-2,” accessed 13-May-2022. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric

[37] Q.-T. Bui, C. Jamet, V. Vantrepotte, X. Mériaux, A. Cauvin, and M. A. Mograne, “Evaluation of Sentinel-2/MSI Atmospheric Correction Algorithms over Two Contrasted French Coastal Waters,” *Remote Sensing*, vol. 14, no. 5, 2022.

[38] L. Rumora, M. Miler, and D. Medak, “Impact of Various Atmospheric Corrections on Sentinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers,” *ISPRS International Journal of Geo-Information*, vol. 9, no. 4, 2020.

[39] N. Hoepffner and G. Zibordi, “Remote Sensing of Coastal Waters,” in *Encyclopedia of Ocean Sciences (Second Edition)*, second edition ed., J. H. Steele, Ed. Oxford: Academic Press, 2009, pp. 732–741.

[40] I. Sola, A. García-Martín, L. Sandonís-Pozo, J. Álvarez Mozos, F. Pérez-Cabello, M. González-Audícana, and R. Montorio Llovería, “Assessment of atmospheric correction methods for Sentinel-2 images in Mediterranean landscapes,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 73, pp. 63–76, 2018.

[41] L. D. Keukelaere, S. Sterckx, S. Adriaensen, E. Knaeps, I. Reusen, C. Giardino, M. Bresciani, P. Hunter, C. Neil, D. V. der Zande, and D. Vaiciute, “Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: validation for coastal and inland waters,” *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 525–542, 2018.

[42] C. Brockmann, R. Doerffer, M. Peters, S. Kerstin, S. Embacher, and A. Ruescas, “Evolution of the C2RCC Neural Network for Sentinel 2 and 3 for the Retrieval of Ocean Colour Products in Normal

and Extreme Optically Complex Waters," in *Living Planet Symposium*, ser. ESA Special Publication, vol. 740, Aug. 2016, p. 54.

[43] 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*, L. Bruzzone, Ed., vol. 10427, International Society for Optics and Photonics. SPIE, 2017, pp. 37–48.

[44] "ACOLITE - Overview," accessed 13-May-2022. [Online]. Available: https://github.com/acolite

[45] M. Warren, S. Simis, V. Martinez-Vicente, K. Poser, M. Bresciani, K. Alikas, E. Spyrakos, C. Giardino, and A. Ansper, "Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters," *Remote Sensing of Environment*, vol. 225, pp. 267–289, 2019.

[46] Q. Vanhellemont and K. Ruddick, "Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications," *Remote Sensing of Environment*, vol. 216, pp. 586–597, 2018.

[47] Q. Vanhellemont, "Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives," *Remote Sensing of Environment*, vol. 225, pp. 175–192, 2019.

[48] ——, "Daily metre-scale mapping of water turbidity using CubeSat imagery," *Opt. Express*, vol. 27, pp. A1372–A1399, 2019.

[49] ——, "Sensitivity analysis of the dark spectrum fitting atmospheric correction for metre- and decametre-scale satellite imagery using autonomous hyperspectral radiometry," *Opt. Express*, vol. 28, no. 20, pp. 29 948–29 965, 2020.

[50] Q. Vanhellemont and K. Ruddick, "Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters," *Remote Sensing of Environment*, vol. 256, no. 112284, 2021.

[51] ——, "Turbid wakes associated with offshore wind turbines observed with Landsat 8," *Remote Sensing of Environment*, vol. 145, pp. 105–115, 2014.

[52] ——, "Advantages of high quality SWIR bands for ocean colour processing: Examples from Landsat-8," *Remote Sensing of Environment*, vol. 161, pp. 89–106, 2015.

[53] Q. Vanhellmont and K. G. Ruddick, "Acolite for Sentinel-2: Aquatic Applications of MSI Imagery," 2016.

[54] Q. Vanhellemont, “Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives,” *Remote Sensing of Environment*, vol. 225, pp. 175–192, 2019.

[55] European Space Agency, “SNAP Download - STEP,” accessed 13-May-2022. [Online]. Available: https://step.esa.int/main/download/snap-download/

[56] S. K. McFeeters, “The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features,” *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425–1432, 1996.

[57] W. Jiang, Y. Ni, Z. Pang, G. He, J. Fu, J. Lu, K. Yang, T. Long, and T. Lei, “A new index for identifying water body from Sentinel-2 satellite remote sensing imagery,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, pp. 33–38, 2020.

[58] European Space Agency, “Open Access Hub,” accessed 13-May-2022. [Online]. Available: https://scihub.copernicus.eu/

[59] P. Tasseron, T. van Emmerik, J. Peller, L. Schreyers, and L. Biermann, “Advancing Floating Macroplastic Detection from Space Using Experimental Hyperspectral Imagery,” *Remote Sensing*, vol. 13, no. 12, 2021.

[60] M. Aldersley, “South African port swamped in plastic waste and debris after floods,” Mail Online, 2019, accessed 13-May-2022. [Online]. Available: https://www.dailymail.co.uk/news/article-7154433/South-African-port-swamped-plastic-waste-debris-heavy-rain-flash-flooding.html

[61] S. Steelandt, D. Marguerie, N. Bhiry, and A. Delwaide, “A study of the composition, characteristics, and origin of modern driftwood on the western coast of Nunavik (Quebec, Canada),” *Journal of Geophysical Research: Biogeosciences*, vol. 120, pp. 480–501, 2015.

[62] N. O. US Department of Commerce and A. Administration, “What is seaweed?” Noaa.gov, 2019, accessed 13-May-2022. [Online]. Available: https://oceanservice.noaa.gov/facts/seaweed.html

[63] “Spectral Indices,” www.l3harrisgeospatial.com, accessed 13-May-2022. [Online]. Available: https://www.l3harrisgeospatial.com/docs/spectralindices.html

[64] L. Shen and C. Li, “Water body extraction from Landsat ETM+ imagery using adaboost algorithm,” in *18th International Conference on Geoinformatics*, 2010, pp. 1–4.

[65] G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud, “Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery,” *Remote Sensing of Environment*, vol. 140, pp. 23–35, 2014.

[66] A. A. Gitelson, O. B. Chivkunova, and M. N. Merzlyak, “Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves, doi = 10.3732/ajb.0800395,” *American Journal of Botany*, vol. 96, pp. 1861–1868, 2009.

[67] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943.

[68] J. Langr and V. Bok, *GANs in action: deep learning with Generative Adversarial Networks*. Shelter Island, New York, Manning Publications, 2019.

[69] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Networks,” 2014.

[70] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” 2017.

[71] “1.6. Nearest Neighbors,” scikit-learn, accessed 13-May-2022. [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html#nearest-neighbors-classification

[72] D. S. Elsinghorst, “Machine Learning Basics - Gradient Boosting & XGBoost,” Shirin’s playgRound, 2018, accessed 13-May-2022. [Online]. Available: https://shirinsplayground.netlify.app/2018/11/ml_basics_gbm/

[73] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, “A Survey of Uncertainty in Deep Neural Networks,” 2021.

[74] F. K. Gustafsson, M. Danelljan, and T. B. Schön, “Evaluating Scalable Bayesian Deep Learning Methods for Robust Computer Vision,” 2019.

[75] “Seaweed is spoiling Caribbean beaches. Can science and shovels solve the problem?” Travel, 2021, accessed 13-May-2022. [Online]. Available: https://www.nationalgeographic.com/travel/article/can-science-solve-the-seaweed-problem-on-mexican-beaches

[76] M. Duarte, “Automatic Detection of Floating Marine Debris Using Multi-spectral Satellite Imagery - Datasets,” accessed 23-May-2022. [Online]. Available: https://github.com/miguelmendesduarte/EO_data

A

Case Study - Vigo Ria, Spain

As discussed in chapter 5, future research should focus on maximizing the qualities of this thesis' model by creating parallel systems such as numerical models. Here, it is shown how combining the ensemble model with a particle tracking model can provide meaningful results and insights related to specific locations. Sousa et al. [4] studied the distribution and concentration of microplastics released by wastewater plants in the Vigo Ria from September 2016 to October 2016, using a particle-tracking model. The model was validated by comparing the results with in situ concentrations of microplastics detected in mussels collected in different areas in the estuary.

There are twelve available Sentinel-2A images from Vigo Ria during this period (Sentinel-2B was only deployed the following year, in 2017). However, only one of them was not affected by clouds. The satellite image of the AOI, from the 20^{th} of September 2016, is shown in Figure A.1. The red circles reveal the zones where the previous study found accumulations of plastic.

Figure A.1: Sentinel-2A image of Vigo Ria from the 20^{th} of September 2016, after the atmospheric correction process. The red circles indicate the areas where plastic accumulates, according to the numerical model of [4].

Figure A.2 shows the ensemble model's predictions. As expected, uncertainty increases the closer the pixels are to the shore. There are also some pixels that the land mask did not exclude and were classified as sea snot.

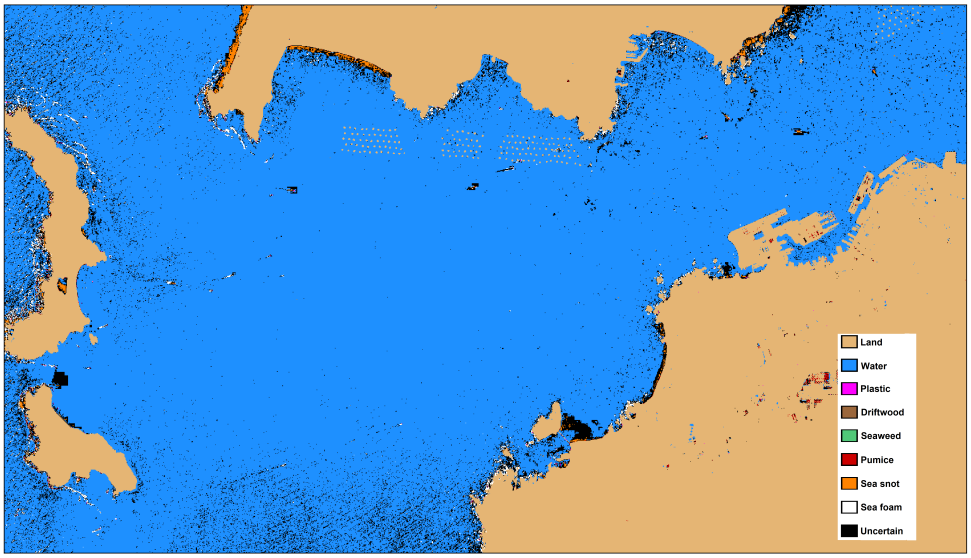

Figure A.2: Model's predictions based on the previous Sentinel-2A image, from the 20^{th} of September 2016.

Figure A.3 shows the satellite image from the first zone of interest (bottom right red circle in Figure A.1) and the model's classification results. There are no signs of a floating material in the image and the model did not classify any pixel as plastic.

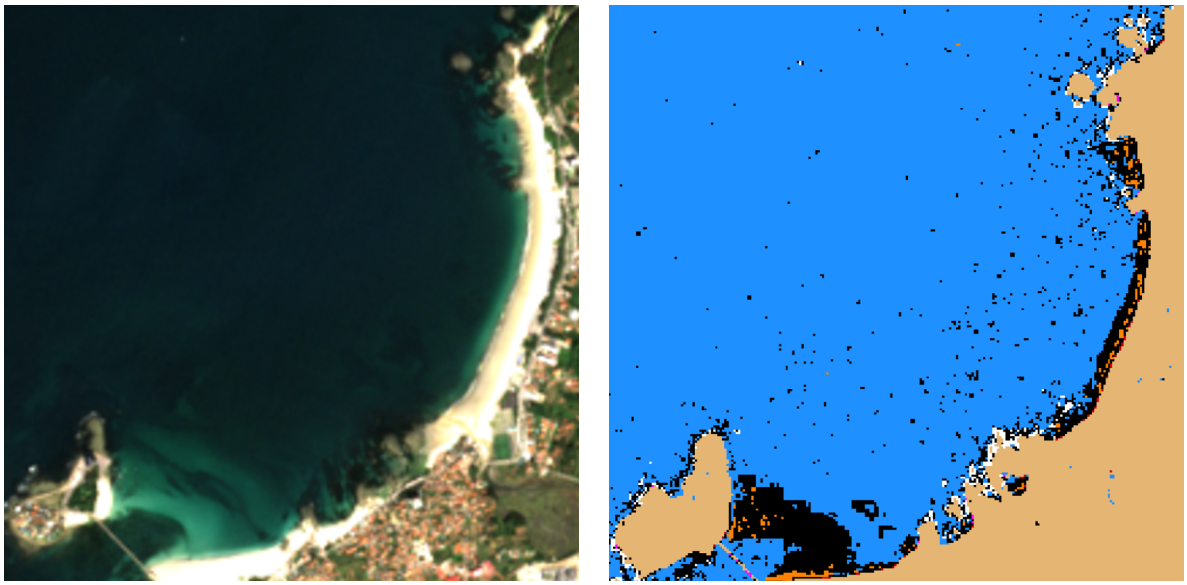

Figure A.3: Sentinel-2A image from the first zone of interest (left) and model's predictions (right).

Figure A.4 shows the satellite image from the second zone of interest (top right red circle in Figure A.1) and the model's classification results. There are no signs of a floating material in the Sentinel-2 image. The only pixel classified as plastic by the model is from a ship. In these predictions there is more uncertainty, which is probably related to the depth of the water.

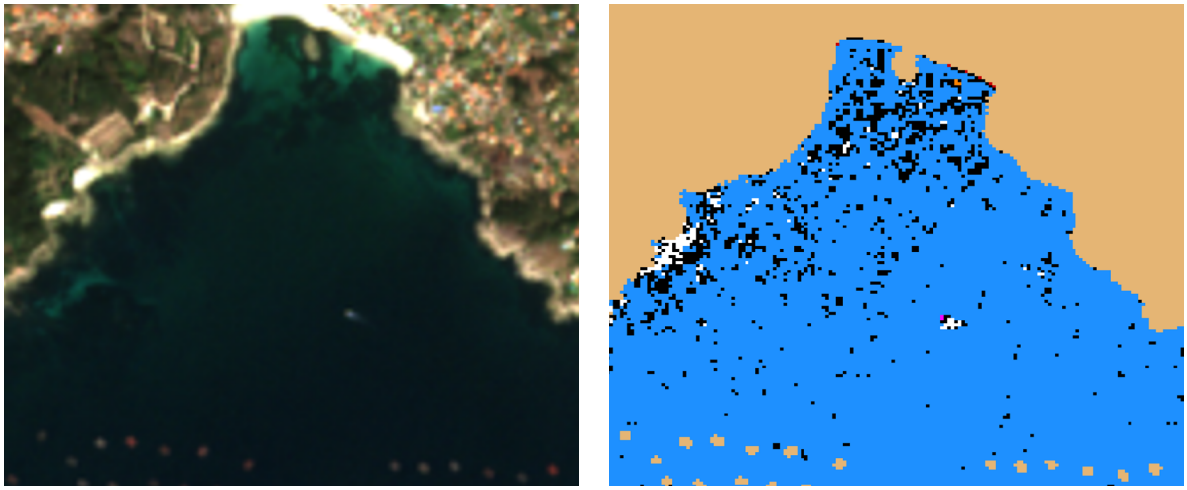

Figure A.4: Sentinel-2A image from the second zone of interest (left) and model's predictions (right).

Figure A.5 shows the satellite image from the third zone of interest (top left red circle in Figure A.1) and the model's classification results. The Sentinel-2 image shows two small floating materials near the shore. There are also several structures in the lower part of the image related to mussel farming. The structures were removed by the land mask. The model classified two pixels as plastic, where the original image shows floating materials. These results associated with the particle-tracking model [4] indicate that there is a strong chance of being floating plastic debris.

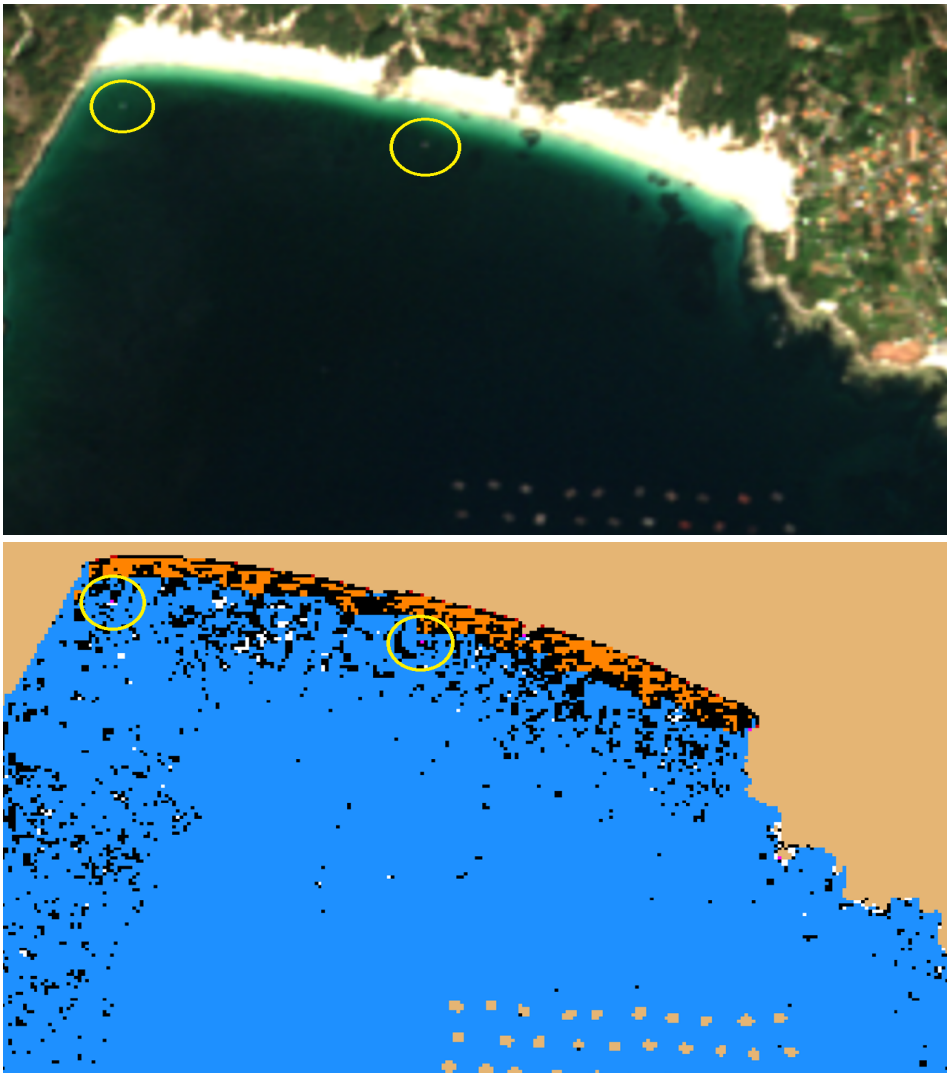

Figure A.5: Sentinel-2A image from the third zone of interest (upper image) and model's predictions (lower image). The yellow circles highlight pixels with higher reflectance than water.

Figure A.6 shows the satellite image from the third zone of interest (left red circle in Figure A.1) and the model's classification results. The Sentinel-2 image shows several pixels with higher reflectance than water near the shore. The model only classified two pixels as plastic. However, two adjacent pixels were also removed by the land mask, meaning that it probably is a small ship.

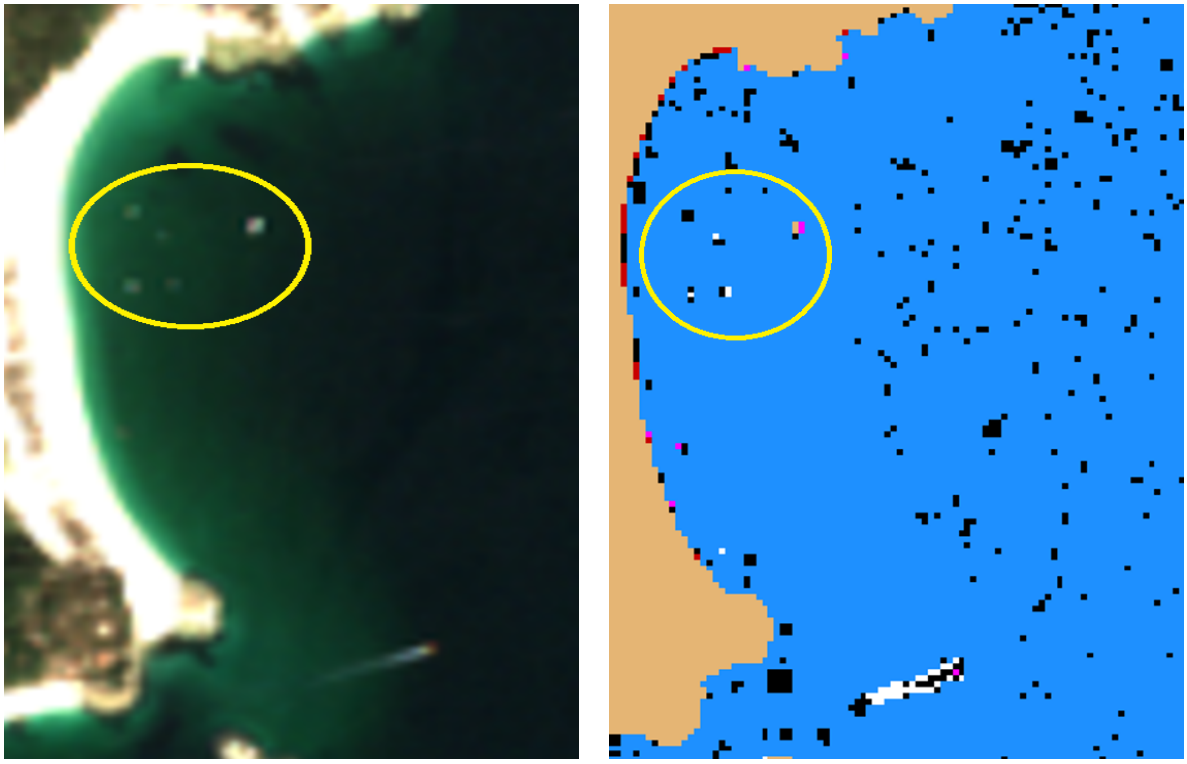

Figure A.6: Sentinel-2A image from the fourth zone of interest (left) and model's predictions (right). The yellow circle highlights pixels with higher reflectance than water.

These results highlight the potential of combining several plastic detection systems to maximize the quality of the results. This study found two pixels that have a high probability of containing floating plastic debris, in the third zone of interest. The next step should be confirming the results with in situ observations, which can be achieved by deploying marine robots or UAVs in this area.

B

Classification Results

B.1 Classifiers trained with spectral bands

This section shows the classification results of all methods tested, except the best one, XGBoost, which was already shown in the thesis (Figure 4.2). All the models were trained only with the Sentinel-2 spectral bands and the hyperparameters were selected to maximize the global accuracy. The models that showed the best results apart from the XGBoost were the Logistic Regression, the Random Forest and the K-nearest neighbour with K=1 (despite the overall good results, this K value makes the model extremely sensitive to noise, outliers, and mislabelled data). The Gaussian Naïve Bayes model has a lot of trouble classifying pixels of sea snot, sea foam, and pumice. Additionally, the Support Vector Machine wrongly classifies most of the sea foam and sea snot as water. Furthermore, the Support Vector Machine with Stochastic Gradient Descent shows good accuracies in every class except for sea foam pixels. It classified 21% of sea foam as water, 30% as plastic, and 40% as pumice.

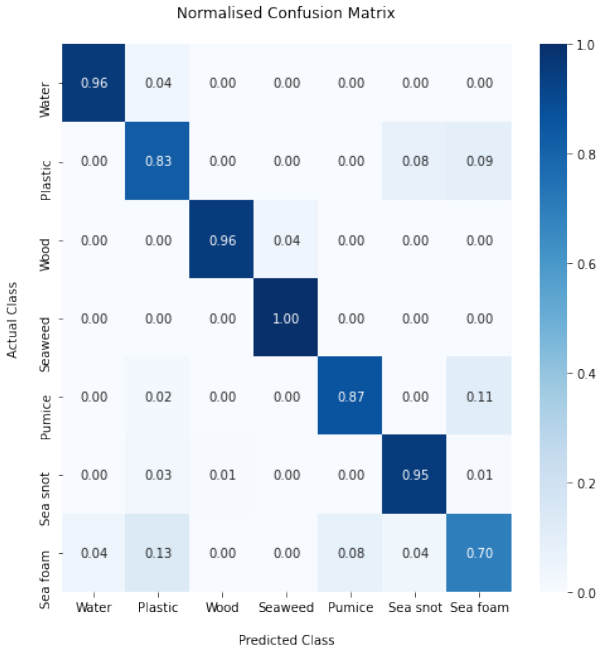

Figure B.1: Normalised confusion matrix of the Logistic Regression model trained with all spectral bands.

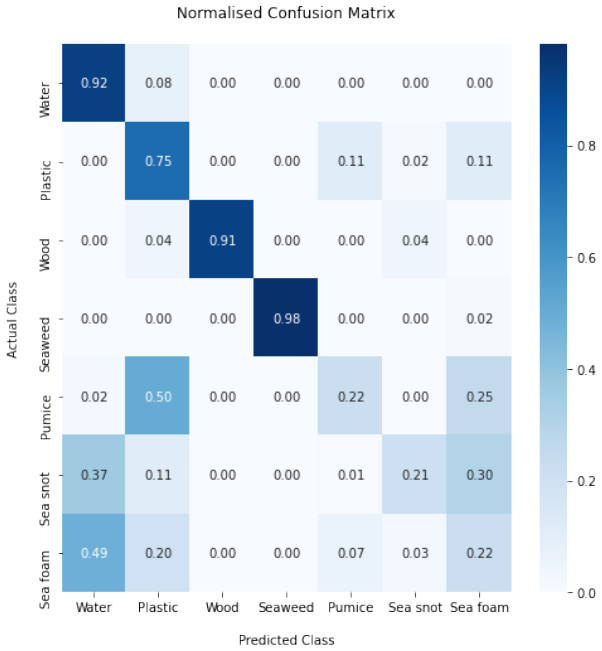

Figure B.2: Normalised confusion matrix of the Gaussian Naïve Bayes model trained with all spectral bands.

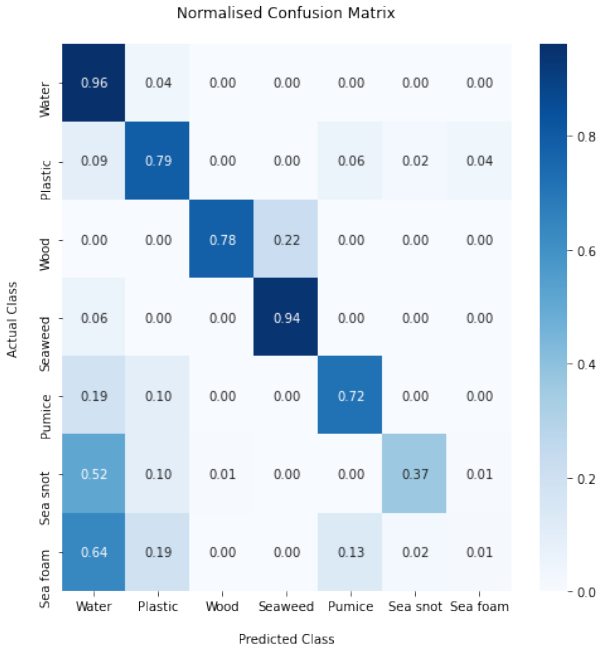

Figure B.3: Normalised confusion matrix of the Support Vector Machine model trained with all spectral bands.

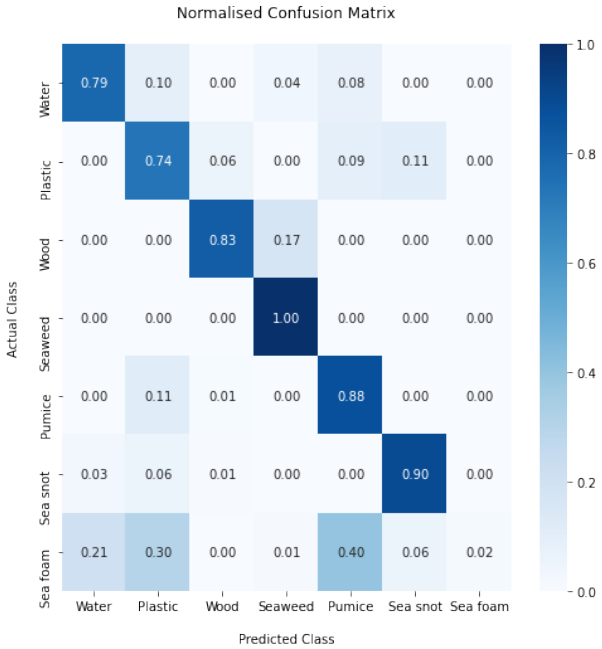

Figure B.4: Normalised confusion matrix of the Support Vector Machine with Stochastic Gradient Descent model trained with all spectral bands.

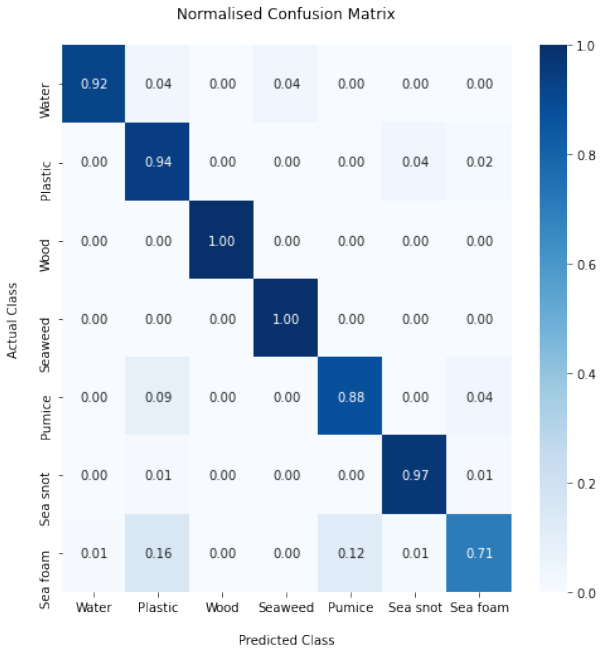

Figure B.5: Normalised confusion matrix of the Random Forest model trained with all spectral bands.

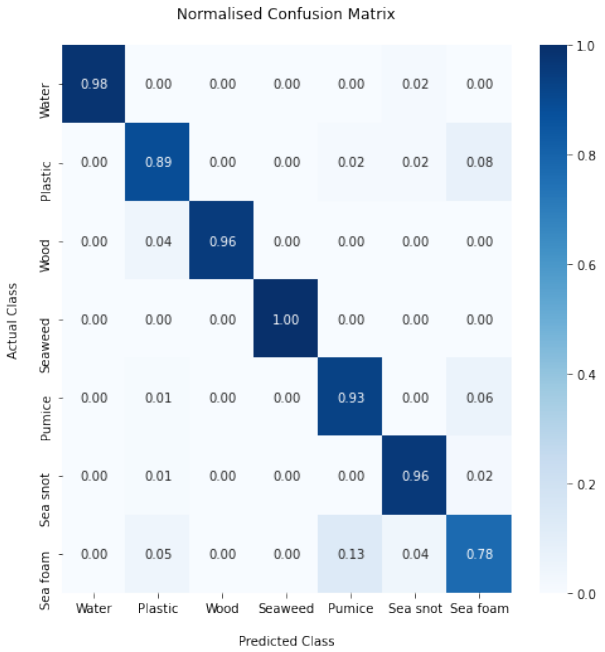

Figure B.6: Normalised confusion matrix of the K-nearest Neighbour model, with K=1, trained with all spectral bands.

B.2 Classifiers trained with spectral bands and spectral indices

Just like in the previous section (B.1), this one shows the classification results of all methods tested, except the best one, XGBoost, which was already shown in the thesis. All the models were trained with the Sentinel-2 spectral bands and all spectral indices described in the thesis (section 3.5). Their hyperparameters were selected to maximize the global accuracy. The only model that showed good results apart from the XGBoost was the Gaussian Naïve Bayes. All the others could not find a pattern in the data that allowed them to distinguish the different classes.

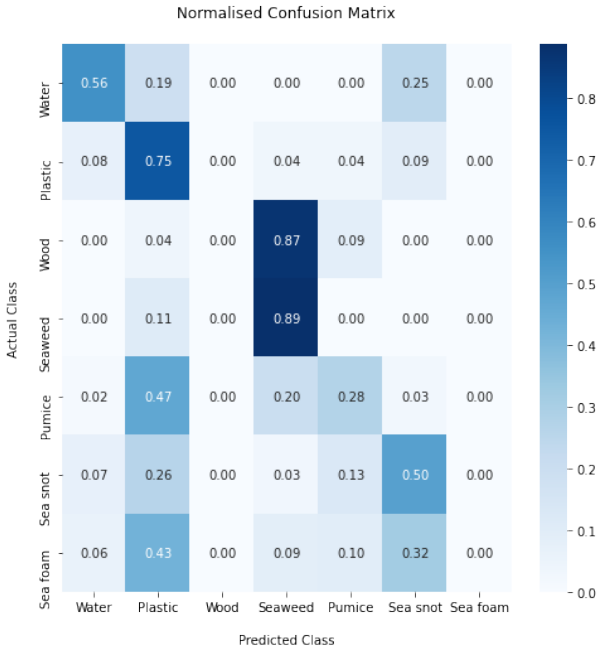

Figure B.7: Normalised confusion matrix of the Logistic Regression model trained with all spectral bands and all spectral indices.

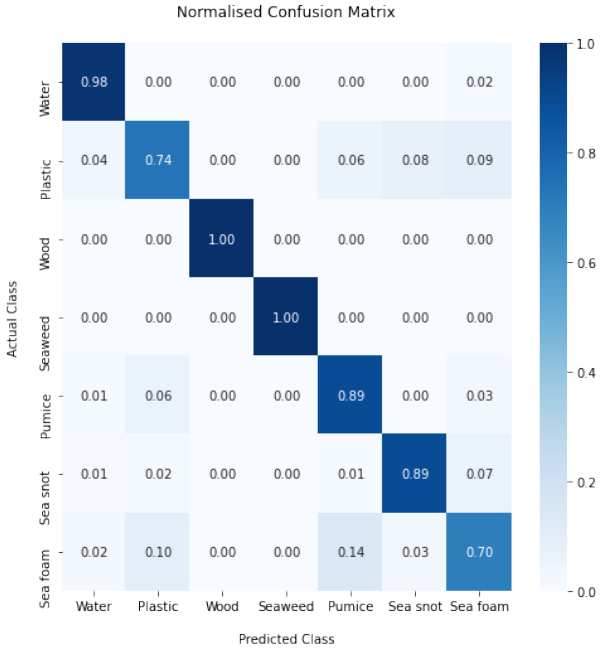

Figure B.8: Normalised confusion matrix of the Gaussian Naïve Bayes model trained with all spectral bands and all spectral indices.

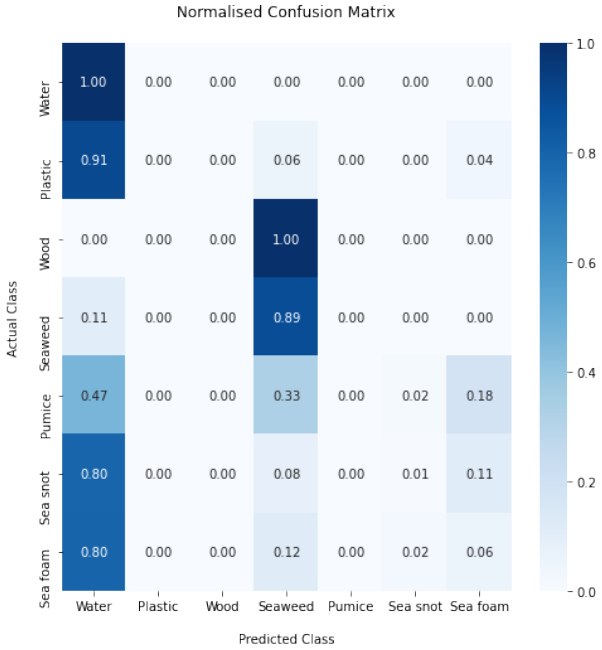

Figure B.9: Normalised confusion matrix of the Support Vector Machine with Stochastic Gradient Descent model trained with all spectral bands and all spectral indices.

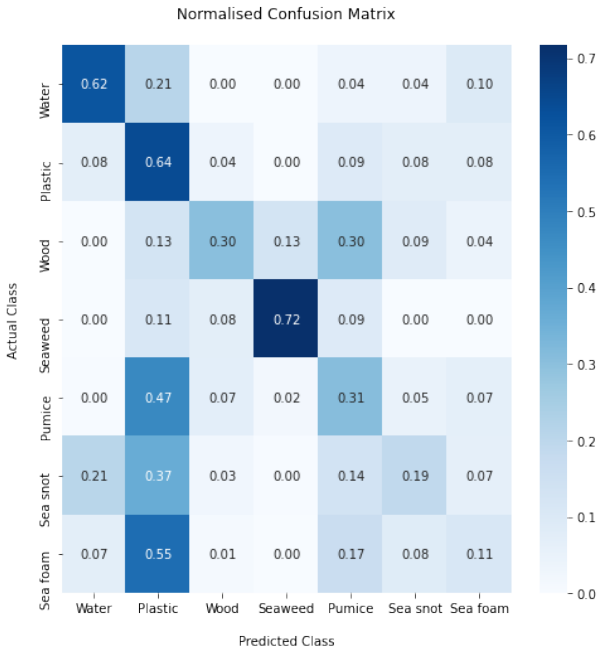

Figure B.10: Normalised confusion matrix of the K-nearest Neighbour model, with K=20, trained with all spectral bands and all spectral indices.