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

Miguel M. Duarte, Leonardo Azevedo

miguel.mendes.duarte@tecnico.ulisboa.pt, leonardo.azevedo@tecnico.ulisboa.pt

Instituto Superior Técnico, University of Lisbon

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.

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

I. INTRODUCTION

GLOBAL plastic production has been steadily increasing, reaching 380 million tonnes produced only in 2015, which represents around 190 times the value in 1950 [1]. The largest market sector for plastic resins is packaging [2], so 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 [3]. Therefore, in 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 [1]. Hence, because of their durability, these buoyant objects accumulate on the ocean's surface and travel worldwide through ocean currents. The most well-known proof of substantial marine plastic accumulations is in the North Pacific Gyre. The so-called Great Pacific Garbage Patch is estimated to comprise almost 79000 tonnes of plastic [3], including not only macroplastics (plastic particles > 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.

These debris affect marine ecosystems in multiple ways. One of its most visible effects is the entanglement of organisms, such as birds, turtles, mammals, and fish, often resulting 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 [4]. Many marine creatures mistake plastic for food and ingest it. Ingestion of plastic can cause lacerations in the digestive system, and its retention in the organism has potential negative consequences for reproduction and growth [5]. Since animals carry these debris in their bodies, plastic is already part of human's food chain, and it might affect human health. 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 [6]. For all these reasons, industries, governments and communities, especially coastal communities, must take immediate action to prevent plastic waste from entering the hydrosphere. 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.

Using earth observation data to detect marine plastic debris accumulations is a recent area of research, but some early studies show promising results [7], [8]. Satellites are a reliable data source thanks to their efficiency in covering extensive areas over time without human interaction and their cost-effectiveness. This work investigates how satellite imagery can be combined with machine learning to detect floating plastic debris in coastal waters automatically. It differs from previous studies since its primary goal is to differentiate floating plastic debris from water and from debris that show similar spectral reflectance, such as sea snot, sea foam, and pumice. It also proposes a way to augment the training data available using synthetic data created with deep generative models.

Section II introduces prior work on the detection of plastic in satellite imagery. Section III describes the satellite used to collect data. In section IV the steps taken to collect and preprocess the data are described. Additionally, it is shown how synthetic satellite pixels were generated. Section V introduces

spectral indices and the classification algorithm used in this work. Furthermore, it presents the uncertainty quantification method proposed herein. Section VI shows the results of every model tested, as well as some results in real-world conditions, and describes the limitations of the best model. In section VII conclusions are outlined besides potential research paths for future work.

II. Related Work

In the past few years, research on floating plastic debris detection and monitoring using data from ship-based visual surveys [9], unmanned aerial vehicles (UAVs) [10], numerical models [11], and cameras deployed at beaches [12] have revealed promising results. Despite the relative success of these methods, they do not provide an option for monitoring larger spatial scales [13]. 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.

In 2018, Topouzelis et al. [14] created the Plastic Litter Project (PLP) to explore the feasibility of detecting man-made plastic targets in the aquatic environment using data from UAVs and from the Copernicus Sentinel-2 satellites. Direct comparison of the UAV data with the Sentinel-2 satellite image led to the conclusion that spectral reflectance of floating plastic positively correlates with the percentage pixel coverage of each target. The following year, in the PLP 2019 [15], it was concluded that marine litter can be detected with at least 25% of the Sentinel-2 pixel covered in plastic. The potential causes for misidentification of plastic pixels were also identified: clouds, shadows, vessels, fumes, sun glint and bottom reflectance on the coastline. In the PLP 2020, the same research group created large reference plastic targets. In 2021, during the PLP 2021, those targets were deployed in Gera Gulf, Greece, and, despite no studies published yet related to these last two PLPs, all information is available online [16]. These studies represent a substantial source of marine plastic debris data.

Similarly to the PLPs, Themistocleous et al. [17] investigated the detection of floating plastic litter using a Sentinel-2 image and UAV data of a three-by-ten-meter artificial target made of water bottles placed in the sea near in Limassol, Cyprus. Seven spectral indices were examined and two new were developed: the Plastic Index (PI) and the Reversed Normalised Difference Water Index (RNDVI). The study found that the target was easier to detect in the Near-infrared (NIR) wavelengths, and the PI was the most effective index in identifying plastic. However, when the PI was applied to the coast of Limassol, several misclassifications were reported, mainly related to boats with plastic surfaces.

Kikaki et al. [18] investigated the capability of satellite sensors in detecting marine plastic debris over the Bay Islands and Gulf of Honduras between 2014 and 2019. 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. 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 methods capable of detecting marine plastic pollution.

Biermann et al. [7] assessed the capability to distinguish plastic from other floating debris, such as timber and seaweed, using Sentinel-2 imagery. Two spectral indices were used: the Normalised Difference Vegetation Index (NDVI) and the newly developed Floating Debris Index (FDI). When FDI and NDVI were examined together, all the floating materials studied (i.e., seawater, seaweed, timber, plastic, sea foam and pumice) showed distinct clustering. Then, they tested a Naïve Bayes classifier (trained with 53 pixels of plastic, 48 of seaweed, 60 of timber, 17 of spume, and 20 of seawater) with data from PLP 2018 [14] and PLP 2019 [15], and it correctly classified plastic with an accuracy of 86%, whereas 3% of plastic pixels were classified as seawater and 11% as sea foam. This study shows that the spatial and spectral resolution of Sentinel-2 is sufficient for macroplastics accumulations to be distinguishable from water and other floating debris.

The previous study boosted floating plastic debris research. Multiple scientific reports published afterwards use machine learning algorithms along with the NDVI and the FDI. For example, Basu et al. [8] used two supervised and two unsupervised classification algorithms to detect floating plastic in coastal waters. Five Sentinel-2 images from previous studies [14], [15], [17] were considered to build a dataset, which resulted in 59 pixels with floating plastics. Then, a combination of six spectral bands, the NDVI, and the FDI, were selected to develop the models. The supervised classification outperformed the unsupervised clustering algorithms. The best model had an accuracy of 96,7%.

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.

III. Satellite

TABLE I
Sentinel-2 spectral bands, their central wavelengths and spatial resolutions.

Spectral Band	Sentinel-2A Central wavelength, λ (nm)	Sentinel-2B Central wavelength, λ (nm)	Spatial Resolution (m)	
B1	442,7	442,2	60	Visible
B2	492,4	492,1	10	
B3	559,8	559,0	10	
B4	664,6	664,9	10	
B5	704,1	703,8	20	
B6	740,5	739,1	20	
B7	782,8	779,7	20	NIR
B8	832,8	832,9	10	
B8A	864,7	864,0	20	
B9	945,1	943,2	60	
B10	1373,5	1376,9	60	SWIR
B11	1613,7	1610,4	20	
B12	2202,4	2185,7	20	

Just like the studies mentioned in section II, 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, and their optical data is of high spatial resolution (10 m, 20 m, or 60 m, depending on the spectral band), as shown in Table I. Each MSI has 13 spectral bands that range from the visible and NIR to the Short-wave infrared (SWIR), 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 [19]. All these features make Sentinel-2 a preferential option for acquiring multi-spectral floating plastic data.

IV. DATA

A. 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 a correction method to remove the contribution of the atmosphere from the MSI measurements. Satellite data of coastal waters are also challenged by continental aerosols, bottom reflectance, and adjacency of land [20], which raises the water's reflectance. Therefore, land masking is necessary in maritime satellite studies since it removes unnecessary pixels that could be mistaken for floating materials and reduces the computational power needed to process the image.

This work uses the Dark Spectrum Fitting algorithm (DSF) from the Atmospheric Correction for OLI 'lite' (ACOLITE) v.20210802.0 software [21] to perform the atmospheric correction process. This method assumes that the atmosphere is homogeneous, and that the scene contains pixels with zero or very close to zero surface reflectance in at least one of the sensor bands (i.e., 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.

In this study, the land masks were created using the spectral index proposed by McFeeters [22]. The Normalised Difference Water Index (NDWI) 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. It varies between -1 and 1, depending on the quantity of water in the pixel. Therefore, setting a threshold (usually 0) allows the differentiation of water bodies from land and vessels. However, sometimes, it identifies floating natural debris as a non-water body and, although the same did not happen with plastic pixels, it is something to watch out for.

Satellite images were visualized using the ESA open-source SNAP 8.0 software [23].

B. Data Acquisition

In contrast with most previous studies that focus on differentiating plastic from water, this work focuses on distinguishing

TABLE II
ALL DATA COLLECTED TO TRAIN AND TEST THE MACHINE LEARNING MODELS PROPOSED IN THIS WORK.

Type	Source	Date (dd/mm/yyyy)	Location	Sentinel	Number of pixels						
					Water	Plastic	Wood	Seaweed	Pumice	Sea snot	Sea foam
Artificial	[17]	15/12/2018	Limassol, Cyprus	2A	0	4	0	0	0	0	0
	[14]	07/06/2018		2A	0	1	0	0	0	0	0
	[15]	18/042019	Tsamakia beach, Greece	2B	0	3	0	0	0	0	0
		18/05/2019		2B	0	2	0	0	0	0	0
	[16]	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	[7]	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
	[18]	09/10/2017	Caribbean Sea, Honduras	2A	0	49	0	0	0	0	0
	[18]	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 model deployed in real-world conditions					150	209	62	150	150	150	150

floating plastic debris from other floating materials with similar spectral signatures. Thus, data from seven classes were collected and confirmed by scientific reports, news articles, or social media posts (in situ data), followed by a manual inspection of the spectral responses (Table II). The data are freely available through the Copernicus Open Access Hub [24] and the datasets used in this study are available at https://github.com/miguelmendesduarte/EO_data.

- *Water*

Fifteen satellite images were used to collect 150 pixels of ocean water in two distinct areas: the Caribbean Sea and the Gulf of Gera. From the 150 pixels, 121 are from the Sentinel-2A and the remaining from the Sentinel-2B. Also, 25% of the water data are from shallower waters where the bottom of the ocean is visible, resulting in brighter pixels. However, the reflectance of shallower waters 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.

- *Plastic*

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. Every pixel's spectral response was manually inspected and compared to the expected spectral signature in the literature [7], [25], and the ones that did not meet the requirements were rejected. 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 [16], Tsamakia beach [14], [15], and Limassol [17]. The remaining 58% result from observations and reports of plastic floating in the marine environment. On the 23rd of April 2019, substantial quantities of plastic covered the Durban harbour, in South Africa, after a flood event [26]. 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. [18] and their observations over the Bay Islands and Gulf of Honduras. Plastic has two reflectance peaks, one centred at B3 and the other at B8, and one absorption peak centred at the fifth Sentinel-2 spectral band (B5). It is also clear that plastic has higher reflectance values in all spectral bands compared to the water spectral signature (Figure 1).

- *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. However, it is challenging to find these pixels in Sentinel-2 images since significant accumulations of driftwood are not common. PLP 2021 [16] allowed the collection of 62 pixels of driftwood on thirteen different days since they deployed an artificial wooden target. 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 1), and it has two reflectance peaks at B4 and B8.

- *Seaweed*

Seaweed is the common name for countless species of marine plants and algae that grow in the ocean [27]. Its presence in the ocean is essential since it provides nutrients and shelter for many marine organisms. Nevertheless, too much seaweed can be harmful, since it 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 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 presents a sharp increase in reflectance in B4 (Figure 1), followed by a fall in the band 8A, being very distinct from the spectral responses of water, plastic and driftwood.

- *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. 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 1) are close to the plastic mean spectral signature. However, plastic presents an absorption peak in B5, which does not happen with pumice.

- *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. There are several reports of sea snot outbreaks in the last few years, however, none of them in the level of the one in the Marmara Sea in 2021. One Sentinel-2B image from the Marmara Sea, on the 6th of June 2021, showed thousands of pixels containing sea snot. From those, 26403 pixels were selected. By examining Figure 1, it is understandable why Hu et al. [28] concluded that remote differentiation of sea snots and marine debris using multi-band sensors is problematic. The two classes only differ in B8, where plastics have a reflectance peak.

- *Sea foam*

The model from Biermann et al. [7] 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, Spain, was used to gather 2735 pixels of it. Sea foam presents a small reflectance peak in the early spectral bands and another one in B8, just like the plastic

mean spectral signature (Figure 1). These features, adding to its relatively high standard deviation, suggest that sea foam might be confused with plastic.

The process described in this section 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. Consequently, there is a need to balance the training dataset. Undersampling, which is probably the most direct strategy, was used. 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 II. 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. Table II also shows the dataset used to train the model that will perform predictions in real-world conditions.
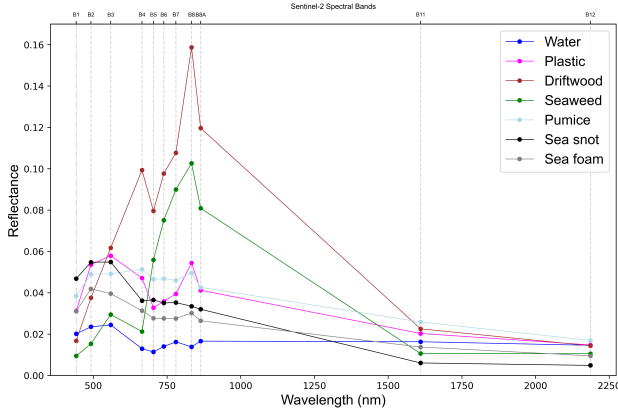

Fig. 1. Spectral signatures derived from the mean reflectance of all data after the atmospheric correction process. Despite the Sentinel-2 satellites' spectral bands having slightly different central wavelengths (Table I), this figure uses the same central wavelengths to facilitate interpretation. B9 (945 nm) and B10 (1375 nm) were removed in the atmospheric correction process.

C. Synthetic 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, Generative Adversarial Networks (GANs) [29] were used to generate synthetic pixels that replicate patterns and features of the actual data, and to assess if artificial datasets are a solution for the lack of floating plastic data. A GAN comprises two simultaneously trained models:

the Generator and the Discriminator. 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 data generated by the first model.

In total, 2000 pixels from each class were generated using a GAN. If the class had training data from both Sentinel-2 satellites (water, plastic, and driftwood), 1000 pixels were generated from each satellite. All the generated pixels showed low standard deviations, 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. To avoid these cases, it was used a Wasserstein Generative Adversarial Network (WGAN). The WGAN [30] 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. 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. In total, 280000 pixels were generated, equally distributed. If the class had training data from both Sentinel-2 satellites (water, plastic, and driftwood), 20000 pixels were generated from each satellite.

V. METHODOLOGY

A. Spectral Indices

As Biermann et al. [7] demonstrated, using the NDVI together with the FDI allows a distinct clustering of water, plastic, driftwood, and seaweed (Figure 2). Nevertheless, using exclusively 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.
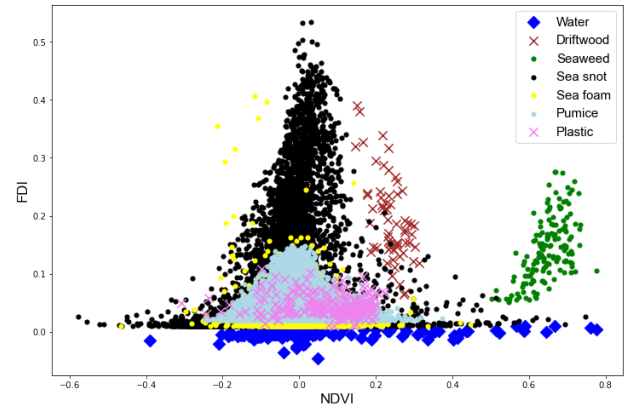

Fig. 2. Combination of the NDVI and the FDI of all real data.

Besides the spectral bands, 24 indices were compared to verify which ones allow a distinct clustering of all the classes gathered in the data acquisition process: Floating Debris Index (FDI), Plastic Index (PI), Normalised Difference Vegetation

Index (NDVI), Reversed Normalised Difference Vegetation Index (RNDVI), Green Normalised Difference Vegetation Index (GNDVI), Pan Normalised Difference Vegetation Index (PNDVI), Normalised Difference Water Index (NDWI), Modified Normalised Difference Water Index (MNDWI), Normalised Difference Moisture Index (NDMI), Normalised Difference Snow Index (NDSI), Water Ratio Index (WRI), Normalised Burn Ratio (NBR), Automated Water Extraction Index (AWEI), Simple Ratio (SR), also known as Ratio Vegetation Index, Anthocyanin Reflectance Index (ARI), Modified Anthocyanin Reflectance Index (MARI), Chlorophyll Red-Edge Index (CHL Red-Edge), Red Edge Position Index (REPI), Enhanced Vegetation Index (EVI), Enhanced Vegetation Index 2 (EVI2), Modified Chlorophyll Absorption Reflectance Index (MCARI), Moisture Index (MI), Soil-Adjusted Vegetation Index (SAVI), and Oil Spill Index (OSI). The equations of the most relevant indices in this study are shown below (Equations (1)-(7)). In each equation, BX represents the reflectance value for the Sentinel-2 spectral band X, and, in Equation (1), λ_{BY} represents the central wavelength of the Sentinel-2 spectral band Y.

$$FDI = B8 - \left(B6 + (B11 - B6) \cdot \frac{\lambda_{B8} - \lambda_{B4}}{\lambda_{B11} - \lambda_{B4}} \cdot 10\right) \quad (1)$$

$$NDWI = \frac{B3 - B8}{B3 + B8} \quad (2)$$

$$MNDWI = \frac{B3 - B12}{B4 + B12} \quad (3)$$

$$NDSI = \frac{B3 - B11}{B3 + B11} \quad (4)$$

$$WRI = \frac{B3 + B4}{B8 + B12} \quad (5)$$

$$MARI = \frac{1}{B3} - \frac{1}{B5} \cdot B7 \quad (6)$$

$$OSI = \frac{B3 + B4}{B2} \quad (7)$$

B. Extreme Gradient Boosting Classifier

Extreme Gradient Boosting, also known as XGBoost [31], is a tree-based ensemble machine learning algorithm and was the chosen method for developing the detection model. XGBoost is trained using the Boosting technique, which is an additive and sequential learning method where trees are grown sequentially, so that each new tree corrects the errors of the previous one in each iteration. 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 of the loss function, which provides more information on how to reach the minimum of the loss function and uses both L1 and L2 regularization to penalize the models. Both features

prevent the models from overfitting. XGBoost has become widely used in classification tasks and is popular in machine learning competitions due to its highly accurate results.

C. Uncertainty Quantification

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 uncertainty) and the model (epistemic uncertainty). The chosen strategy to quantify uncertainty in this work is through Ensemble Methods. Ensemble Methods measure uncertainty based on the predictions of multiple models (ensemble members), that are trained independently from each other using different techniques to increase their variety. 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 alternative methods [32].

To bring variety into the ensembles and generate more reliable results, the models (XGBoost) have random initialization, are trained with different data sizes (mini batches), and the training data are 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 (thresholds selected based on the results of the ensemble model, after some tests).
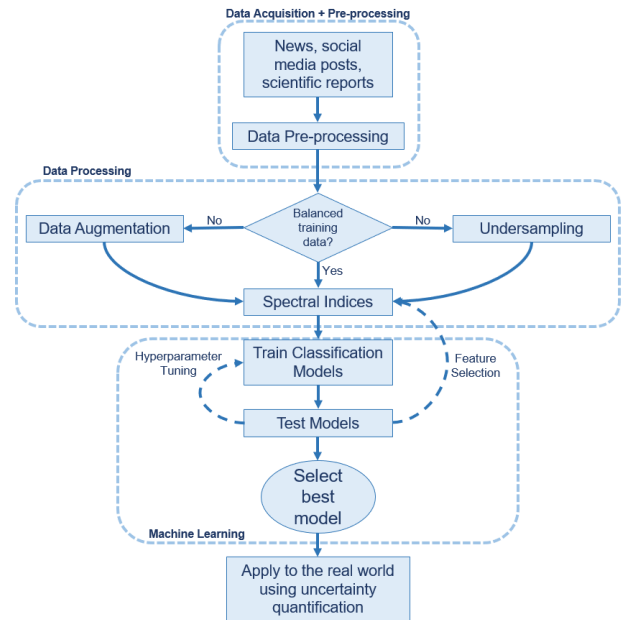
D. Outline

Fig. 3. Flowchart with the steps adopted to detect plastic and other floating debris on the ocean using satellite imagery.

The flowchart in Figure 3 summarizes the steps needed to create a model that can detect and distinguish floating plastic from the other classes studied. The first task is to collect in situ data. 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. The next step is to ensure that the data used to train the classification models is balanced. This can be done by undersampling the most represented classes in the training set or by generating synthetic data. Then, spectral indices may help in separating the different floating classes. 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. Finally, an uncertainty quantification method should be used to improve the information given by the model, helping the decision-making process.

VI. RESULTS

A. Classification with Spectral Bands and Spectral Indices

In order to find the spectral indices that boost the differences between the classes' reflectance, the XGBoost was trained with all spectral bands and all spectral indices described in section V-A. The training and testing datasets are shown in Table II.

The model achieved high accuracies (> 90%) in every class except sea foam. However, this experience's goal was to determine how each feature affects the overall accuracy of the model. This was done through 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 each feature is measured by how much the loss function is affected by shuffling that feature's column. The results are shown in Figure 4.
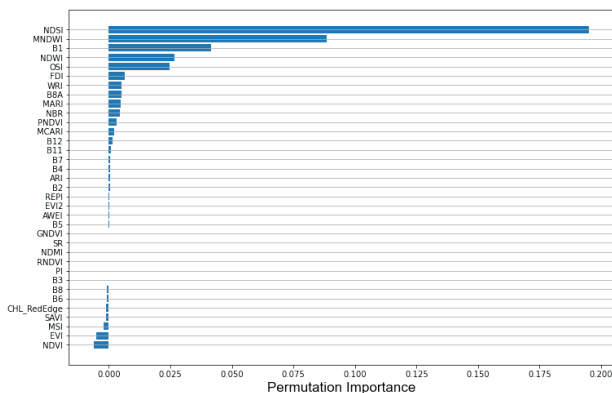

Fig. 4. Permutation importance of each feature in the XGBoost model trained with all spectral bands and every spectral index described in V-A.

With this information, the irrelevant features in the model (nil or negative permutation importance) can be removed from the training process. The goal is to assess which combination of the remaining features 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 (B1, B8A, the NDSI, the MNDWI, the NDWI, the OSI, the FDI, the WRI, and the MARI). Figure 5 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%. The model shows an accuracy above 95% for each class except for sea foam. Water pixels are perfectly classified and none of the other classes' pixels is classified as water. All the driftwood pixels are also correctly classified, but there are false positives. 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.

Overall, these results indicate that this model 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.
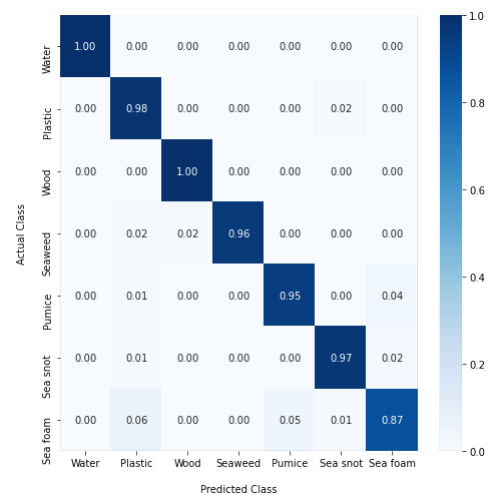

Fig. 5. Normalised confusion matrix of the XGBoost model trained with the 9 features with most permutation importance: B1, B8A, the NDSI, the MNDWI, the NDWI, the OSI, the FDI, the WRI, and the MARI. All the numbers are rounded to two decimal points.

B. Classification with Uncertainty Estimation

Despite the overall good results of the previous model, two aspects must be considered. Firstly, the number of pixels wrongly classified as plastic. Although looking like a small number, a Sentinel-2 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.

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, an ensemble model composed of twenty XGBoost models trained with different subsets of the training set was created. 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 6. 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.

Fig. 6. Normalised confusion matrix of the ensemble model built with 20 XGBoost models trained with different data and different data sizes, using the 9 features with the most permutation importance (Figure 4). Predictions whose mean was below 90% or had a standard deviation above 20 were considered uncertain.

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 31st of July 2021 in the Gulf of Gera, Greece, was used. On this day, the Marine Remote Sensing Group from the University of the Aegean performed another experience for the PLP 2021 [16]. They deployed two large artificial targets, one made of wood, and one composed of plastic, on the ocean.

Both models were trained with a balanced dataset and the pixels from this Sentinel-2 image were removed from the training dataset, so this is the first time both models are seeing these data.

Results from Figure 7 reveal that the first model 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.
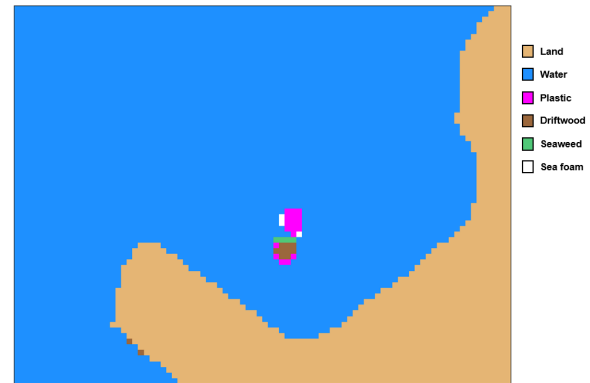

Fig. 7. Predictions from the model that does not quantify uncertainty. The original data are from a Sentinel-2A image from the 31st of July 2021 in the Gulf of Gera, Greece.

The ensemble model's results are shown in Figure 8. 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. 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.

Fig. 8. Ensemble model's predictions. The original data are from a Sentinel-2A image from the 31st of July 2021 in the Gulf of Gera, Greece.

C. Classification with Synthetic Data

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. However, the model trained with 280000 synthetic pixels generated from a WGAN did not achieve good results, indicating that these data have too much variability, so the model cannot find patterns that allow the differentiation of the several classes. To overcome this problem, the synthetic pixels were filtered through the ensemble model. If a synthetic pixel was classified as uncertain, the pixel was discarded. The pixel was also rejected if the ensemble model misclassified it.

After this process, only 1434 pixels from each class were chosen to create a balanced dataset. The confusion matrix of an XGBoost model trained with these data and tested with the testing dataset described in Table II is shown in Figure 9. The results prove 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.
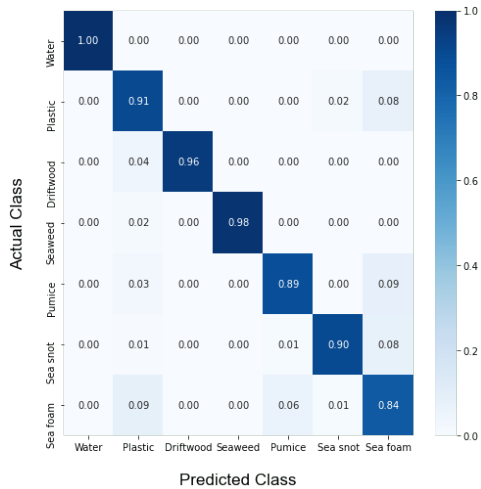
Fig. 9. Normalised confusion matrix of an XGBoost model trained with 1434 synthetic pixels from each class (10038 pixels in total) generated from a WGAN and that passed the filtering process.

D. Monitoring the ocean

Here, the best model (ensemble model with 20 ensemble members trained with different subsets of authentic data from the "training set for the model deployed in real-world conditions" - Table II, with the 9 features with the most permutation importance) is tested in real-world conditions, to assess its feasibility (Figures 10 anf 11).

E. Limitations

Most misclassifications happen 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, so 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. 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.
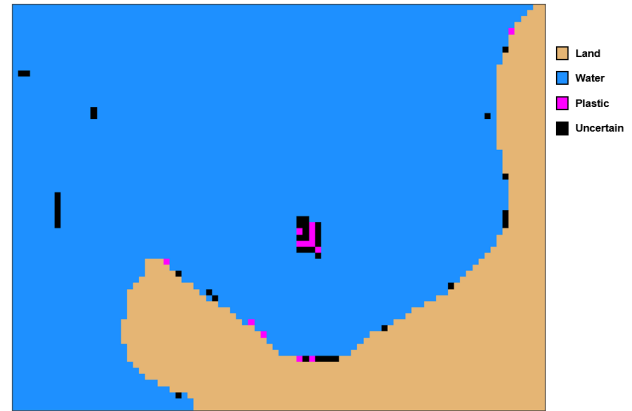

Fig. 10. Model's predictions from a Sentinel-2B image from the 4th of September 2021, in the Gulf of Gera, Greece. On this day, the group from the University of Aegean performed another experience for the PLP 2021 [16]. They removed the wooden target and deployed some of it under the plastic target, simulating a mixed target which is closer to what is found in the ocean. The model 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 uncertain pixels in the open ocean that may be related to sun glint or wave agitation.
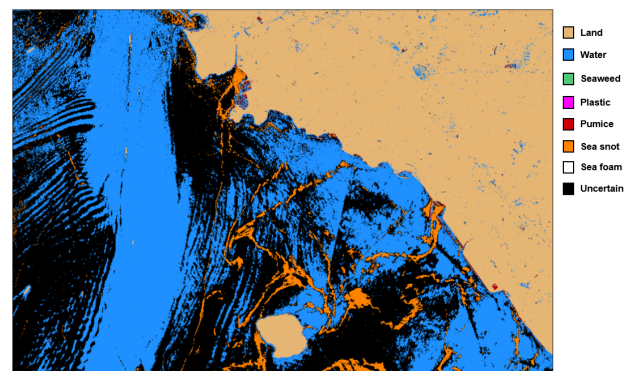

Fig. 11. Model's predictions based on a Sentinel-2B image from the 13th of June 2021, on the coast of Istanbul, where sea snot aggregations with several metres wide can be seen. There are also many pixels classified as uncertain because of the different sea snot depths in each pixel and the rough water conditions.

VII. CONCLUSION

A. 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. Then, it is also the first study that assesses and proves that plastic is distinguishable from five other classes of floating debris and water. The combination of features that proved most successful was using B1, B8A, the NDSI, the MNDWI, the NDWI, the OSI, the FDI, the WRI, and the MARI. An XGBoost model trained with these features showed a high 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 mean-

ingful information for the decision-makers, the percentage of misclassifications, in particular, the number of false positives of plastic labels, must be minimized. The ensemble model achieved lower percentages of correct classifications than the previous model, but also, and most importantly, it decreased the number of misclassifications. Deploying the model in real-world conditions confirmed the good results and showed some of its limitations: clouds, shallower waters, sediments in suspension, and rough sea conditions. The study also proves that training a model with synthetic data produces good results. However, not as good as the model trained with authentic data.

B. 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 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. The same could be done by models that focus on the spatial characteristics of floating materials through satellite imagery (e.g. Convolutional Neural Networks). Marine robots or UAVs 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.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] L. Barboza, A. Cozar, 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.

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

[6] N. J. Beaumont, M. Aanesen, M. C. Austen, T. Borger, 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.

[7] 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.

[8] 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.

[9] G. Suaria and S. Aliani, "Floating debris in the Mediterranean Sea," *Marine Pollution Bulletin*, vol. 86, no. 1, pp. 494-504, 2014.

[10] 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.

[11] M. C. Sousa, M. deCastro, J. Gago, A. S. Ribeiro, M. Des, J. L. Gomez-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.

[12] 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.

[13] L. Fronkova, "Tackling Marine Litter in the Atlantic Area," Clean Atlantic, 2019.

[14] 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.

[15] 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.

[16] M. R. S. Group, "Plastic Litter Project 2021," accessed 13-May-2022. [Online]. Available: http://plp.aegean.gr/

[17] 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.

[18] 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.

[19] E. S. Agency, "Radiometric Resolutions - Sentinel-2," accessed 13-May-2022. [Online]. Available: https://sentinels.copernicus.eu/

[20] N. Hoepffner and G. Zibordi, "Remote Sensing of Coastal Waters," in *Encyclopedia of Ocean Sciences*, Second Edition ed., J. H. Steele, Ed. Oxford: Academic Press, 2009, pp. 732–741.

[21]] 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.

[22] 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.

[23] European Space Agency, "SNAP Download - STEP," accessed 13-May-2022. [Online]. Available:https://step.esa.int/

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

[25] 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.

[26] 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/

[27] 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/

[28] 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.

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

[30] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017.

[31] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pp. 785–794, 2016.

[32] F. K. Gustafsson, M. Danelljan, and T. B. Schon, "Evaluating Scalable Bayesian Deep Learning Methods for Robust Computer Vision," 2019.