Automatic Detection and Identification of Floating Marine Debris Using Multispectral Satellite Imagery

Miguel M. Duarte and Leonardo Azevedo

***Abstract*— Floating plastic debris represent an environmental threat to the maritime environment as they drift the oceans. Developing tools to detect and remove them from our oceans is critical. We present an approach to detect and distinguish suspect plastic debris from other floating materials (i.e., driftwood, seaweed, sea snot, sea foam, and pumice) using Sentinel-2 data. We use extreme gradient boosting trained with data compiled from published works complemented by manual interpretation of satellite images. The method is trained with two spectral bands and seven spectral indices computed from the Sentinel-2 spectral bands. We consider three application scenarios. The first uses the database created under the scope of this work. While the classification achieved a 98% accuracy rate for suspect plastic debris, we acknowledge the need for ground-truth validation. The second, to enlarge the training dataset, uses synthetic data generated through a Wasserstein generative adversarial network. A supervised model trained exclusively with synthetic data successfully classified suspect plastic pixels with an accuracy of 83%. The third comprises an ensemble model that quantifies uncertainty about the predictions obtained with the classifier. We correctly classified 75% of the suspect plastic pixels. However, while the classification accuracy decreased, with the integration of uncertainty in the predictions, the number of misclassifications also significantly decreased when compared to the model with the highest accuracy of the previous scenarios. Due to the mixed band nature of the sensor and subpixel coverage of debris within a pixel, the application to other datasets might not be straightforward.**

***Index Terms*— Floating plastic debris, marine pollution, remote sensing, Sentinel-2.**

I. Introduction

GLOBAL plastic production has been steadily increasing in the last decades, reaching 380 million tons produced only in 2015, which represents around 190 times the value of 1950 [1]. The largest market sector for plastic resins is packaging [2]. While plastics are durable over time, 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 is biodegradable and only a small portion may be recycled or incinerated, they accumulate in landfills or the natural environment [3]. Consequently, in countries where waste management infrastructure is lacking, or poor, plastic waste is likely to enter water bodies and end up in the ocean.

Approximately, 65% of the synthetic polymers included in plastics have a lower density than seawater [1]. Hence, because of their durability, these buoyant objects tend to 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 79 000 tons of accumulated plastic [3], including not only macroplastics (plastic particles >5 mm in size), such as abandoned fishing nets, bottles, and containers, but also microplastics (plastic particles <5 mm in size), which usually result from the fragmentation of larger plastic items over time due to degradation resulting from the harshness of the ocean.

These plastic 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 might cause 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 the human food chain and might be detrimental to human health.

The impact of floating plastic debris is not limited to maritime wildlife. Marine plastics present a range of negative economic impacts. Beaumont et al. [6] estimate that the economic costs of marine plastic, as related to marine natural capital, are conservatively conjectured at between U.S. $3300 and U.S. $33 000 per ton of marine plastic per year, based on 2011 ecosystem service values and marine plastic stocks. For these reasons, industries, governments, and communities, especially in coastal regions, must take immediate action to prevent plastic waste from entering the hydrosphere. However, even

Manuscript received 3 January 2023; revised 5 April 2023, 21 May 2023, and 24 May 2023; accepted 5 June 2023. Date of publication 9 June 2023; date of current version 23 June 2023. This work was supported by Project diStributed AI systEM for mArine plastic debRis moniToring (SMART): the first edition's winner of the AI Moonshot Challenge, sponsored by the Portuguese Space Agency (PTSpace), in cooperation with Fundação para a Ciência e Tecnologia (FCT), Unbabel, European Space Agency (ESA), Agência Nacional de Inovação (ANI), and Web Summit. The work of Leonardo Azevedo was supported by Centro de Recursos Naturais e Ambiente (CERENA) under Project FCT-UIDB/04028/2020. *(Corresponding author: Leonardo Azevedo.)*

Miguel M. Duarte is with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal (e-mail: miguel.mendes.duarte@tecnico.ulisboa.pt).

Leonardo Azevedo is with CERENA/DER, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal (e-mail: leonardo.azevedo@tecnico.ulisboa.pt).

Digital Object Identifier 10.1109/TGRS.2023.3283607

if the world stopped generating plastic waste, macroplastics would persist on the ocean's surface for the next decades. Therefore, there is a prominent need for floating plastic debris to be detected, captured, and removed from the oceans.

Earth observation data have shown early promising results to detect marine plastic debris accumulations (e.g., [7], [8]). Satellites are a reliable data source, thanks to their spatial and temporal resolution and efficiency in covering extensive areas over time without human interaction and their cost-effectiveness. However, the automatic classification and identification of floating plastic debris from satellite data is not straightforward and is still currently a challenge. This objective depends on the characteristics of the sensor considered and the percentage of debris inside a given pixel for a specific spectral band (see [9], [10]).

In this work, we investigate the feasibility of using statistical learning classification algorithms in combination with satellite imagery to automatically detect and identify floating marine debris in coastal waters. Marine debris is a broad term that refers to any human-made or natural object that is floating at the sea surface or suspended in the water column and includes, for example, wood, seaweed, fishing nets, and plastic items. Our primary objective is to differentiate these floating debris, with a specific focus on distinguishing plastic debris from other debris that share similar spectral reflectance, including sea snot, sea foam, and pumice. In addition, we propose a way to augment the training data available, by creating synthetic data with deep generative models [i.e., generative adversarial networks (GANs)]. These synthetic data approximate the mean statistical spectra estimated from the multispectral instrument (MSI) of each type of debris considered as retrieved from the training dataset. These spectra do not correspond to true endmembers of each floating material due to the mixed band nature of the MSI and the percentage of debris inside a given pixel for a given band.

To acknowledge the need for independent and in situ validation of the detected plastic debris, we refer to them as "'suspect plastic"' throughout this article. This term also reflects the current limitations of the proposed approach and its extension to alternative datasets, and underscores the importance of further validation. The proposed classification algorithm is suitable for the dataset considered, but, since Sentinel-2 has mixed bands resolution and the ratio of debris inside a given pixel might be small, the applicability to other types of sensors and datasets might be limited [9], [10].

We start by introducing prior work on the detection of floating plastic debris using Earth observation data in Section II. In Section III, the characteristics of the satellites used to collect data and the features of the satellite imagery are described, as well as the steps taken to collect and process the data before classification. Additionally, it also describes the proposed approach to generate synthetic satellite pixels. Section IV introduces spectral indices, the classification algorithm, and the uncertainty quantification method proposed herein. Section V shows and discusses the results obtained under the several scenarios considered. In Section VI, we draw the main conclusions of this work and potential research paths for future work.

II. Related Work

In the past years, research on floating plastic debris detection and monitoring using data from ship-based visual surveys [11], unmanned aerial vehicles (UAVs) [12], numerical models [13], and cameras deployed at beaches [14] have revealed promising results. Despite the relative success of these methods, they do not provide an option for large-scale ocean monitoring as they are sparsely distributed in the ocean and represent local observations [15]. The use of Earth observation data has great potential to surpass these limitations and allow the automatic identification of floating plastic debris in extensive areas even though being affected by: physical and technical limitations, namely, cloud interference; atmospheric and sea-surface effects; the instrument's spatial resolution [16]; different band resolutions; different water–debris mixing; and band-to-band registration errors [9], [10].

In 2018, Topouzelis et al. [16] 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 the Copernicus Sentinel-2 satellites. Direct comparison of the UAV data with the Sentinel-2 satellite images led to the conclusion that the spectral reflectance of floating plastic positively correlates with the percentage pixel coverage of each target. These findings are in line with the numerical predictions of Hu [10]. The following year, the PLP 2019 [17] concluded that marine litter can be detected with at least 25% of the Sentinel-2 pixel covered in plastic. The potential causes for the misidentification of plastic pixels were also identified: clouds, shadows, vessels, fumes, sun glint, and bottom reflectance on the coastline and a mixture of the spectra from water and other debris. In PLP 2020, the same research group created large reference plastic targets to be deployed in the following year, during the PLP 2021. These targets were deployed in Gera Gulf, Greece, and, despite no studies published yet related to these last two PLPs, all information and main results are available online [18]. These studies represent a substantial source of information on what concerns marine plastic debris accumulations and satellite data. Recently, Kikaki et al. [19] published marine debris archive (MARIDA), a benchmark dataset to assess the potential of Sentinel-2 and automatic classification algorithms to detect floating debris in remote areas.

Similar to the PLPs, Themistocleous et al. [20] investigated the detection of a three-by-ten-meter artificial target made of water bottles. This target was deployed nearshore in Limassol, Cyprus. Seven spectral indices were examined to assess the ability to detect floating plastics from satellite images. Two new indices were proposed: the plastic index (PI) and the reversed normalized difference water index (RNDWI). The authors found that the target was easier to be detected in the near-infrared (NIR) wavelengths (Table I), and the PI was the most effective index in identifying floating plastic debris. However, when the PI was applied to the coast of Limassol, several misclassifications were reported, mainly related to boats with plastic surfaces. This application example illustrates the difficulties in obtaining a true spectral endmember of floating debris that can be generalized to other regions of the globe, other data sets, and sensors. Hu [9] addresses this question

TABLE I
SENTINEL-2 SPECTRAL BANDS, THEIR CENTRAL WAVELENGTHS, AND SPATIAL RESOLUTIONS

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

and shows that distinct nonalgae floating matters show relative flat reflectance spectral shapes in the vis–NIR range, so it appears difficult to separate them spectrally. Also, the remote detection of floating microplastics seems impossible based on the sensors' characteristics. Therefore, Hu [9] suggests that future research should focus on frontal zones or windrows that may aggregate microplastics. Furthermore, the identification of macroplastic debris appears possible from MSI, but additional work is required to modulate the pure spectral endmember of each material. Finally, after referring to the need for a more complete spectral library of various floating debris, this study recommends the development of more robust algorithms for automatic debris identification from Earth observation data.

Kikaki et al. [21] 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 responses of the pixels for that specific MSI with the ones reported in the literature. While the proposed method accuracy was not quantified, it provides a validated data library for future studies. This 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 normalized 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 corresponding mainly to plastic, 48 to seaweed, 60 to timber, 17 to spume, and 20 to seawater with data from PLP 2018 [16] and PLP 2019 [17]. The proposed approach correctly classified suspect plastic with an accuracy of 86%, whereas 3% of suspect plastic pixels were classified as seawater and 11% as sea foam. This study concludes that the spatial and spectral resolution of Sentinel-2 is sufficient to distinguish macroplastic accumulations from water and other floating debris. However, the percentage of misclassifications reflects the challenging nature of the task, as the spectra obtained depend on different factors that interact between themselves (e.g., the characteristics of the MSI installed in the sensor, the ratio between water and plastic debris, and the classification algorithm).

The previous studies boosted the amount of literature regarding the automatic identification of floating plastic debris from satellite imagery. Multiple scientific reports were published recently using machine learning algorithms along with both 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 [16], [17], [20] were considered to create 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, these studies highlight the need for more data related to marine litter to be collected globally and a better understanding of the true spectral response of different types of material. Nevertheless, the classification models that showed the best results rely on supervised classification methods, which highly depend on the amount and quality of the supplied training samples. Qi et al. [22] showed that preparing these samples is critical for the success of the predictions.

III. DATA

Building upon the studies mentioned in Section II, we use freely available satellite data products from the Sentinel-2 mission. This mission 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 (five days at the equator and two–three days at mid-latitudes) over all coastal waters up to 20 km from the shore. Each satellite has an MSI aboard that works passively, and its optical data are of high spatial resolution (10, 20, or 60 m, depending on the spectral band) (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 0–4095 potential light intensity values [23]. All these features make Sentinel-2 a preferential option for acquiring multispectral floating plastic data debris nearshore.

A. Data Preprocessing

Unlike UAVs data, where the atmospheric effects are dismissed because of the negligible path from the sensor to the target, 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 [24], which increases the water's reflectance. Therefore, land masking is necessary in maritime satellite studies since it removes unnecessary pixels that could be misinterpreted as floating materials while reducing the computational cost related to processing the image dataset.

To perform the atmospheric correction of the satellite images, we applied the dark spectrum fitting (DSF) algorithm from the atmospheric correction for OLI "lite" (ACOLITE) v.20210802.0 software [25]. 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 lookup table. Due to low atmospheric transmittance, bands 9 (B9) and 10 (B10) are excluded from the output of the corrections.

Land masks were created using the spectral index proposed by McFeeters [26]. The normalized difference water index (NDWI) is a mathematical formula that combines the third and eighth Sentinel-2 spectral bands [see (3)] 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 close to 0 allows the differentiation of water bodies from land and vessels depending on the size of the feature of interest and the spatial resolution of the sensors. However, sometimes, it identifies floating natural debris as a nonwater body, and although the same did not happen with plastic pixels, it is something to watch out for. Satellite images used in this work were preprocessed using the ESA open-source SNAP 8.0 software [27].

B. Data Acquisition

Complementary to most previous studies that focus on differentiating plastic debris from the water, we aim at distinguishing floating plastic debris from other maritime floating materials. Thus, data from seven different types of floating debris were collected to build a training dataset for our classifier. The nature of these data was confirmed by scientific reports, news articles, and social media posts. While we assume these data as in situ observations of different floating debris with sufficient size to be detected by the sensors, it is necessary to exercise caution in interpreting the results due to the need for ground-truth validation.

The spectral response of each material shown below represents a statistical average of the manually interpreted pixels given the nature of the floating material as well as the MSI used. These average spectra do not precisely represent the true endmember spectrum for each type of material. Hu [9] and Hu [10] discuss in detail the practical implications of estimating endmembers and the difficulties associated with mixed band resolutions and band-to-band registration errors.

All data were manually inspected in terms of their spectral responses (Table II). The data collected under the scope of this work are freely available through the Copernicus Open Access Hub [28], and the datasets used in this study are available at https://github.com/miguelmendesduarte/Floating-Marine-Debris-Data.

In the following, we describe the main characteristics of the data collected under the scope of this work for different floating debris considered herein: water, plastic, driftwood, seaweed, pumice, sea snot, and sea foam. It is, however, worthwhile to note that this description is valid for our dataset; discrepancies might be found as the spectral response depends on the floating debris-to-plastic ratio and the percentage of the pixel occupied by a given material, which will impact, for example, the sensor band-to-band registration. None of the spectra describe in the following represent a true endmember spectrum of the material considered.

C. 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. Of the 150 pixels, 121 are from Sentinel-2A, and the remaining are from Sentinel-2B images. Also, 25% of the water data are from shallower waters where the bottom of the ocean is visible from the original satellite images, thus resulting in brighter pixels. However, the reflectance of shallower waters is not considerably different from the deeper waters' reflectance. Therefore, there is no need to create two distinct categories, and all the data were grouped into a single class.

D. Plastic

Validated data related to floating plastic debris are scarce. We gathered 206 pixels of suspect plastic that were previously reported by scientific reports, news articles, or pictures on social media posts. Every pixel's spectral response was manually inspected and compared to the expected spectra in the literature [7], [29]. We followed a conservative approach, and data samples that did not meet the requirements were rejected and removed from the training dataset. From the 206 pixels, 102 were taken from Sentinel-2A images and 107 from Sentinel-2B imagery. Forty-two percent of the data, corresponding to 88 pixels, are from artificial plastic targets deployed in the ocean in the Gulf of Gera [18], Tsamakia beach [16], [17], and Limassol [20]. The remaining 58% result from observations and reports of floating plastic debris in the marine environment. In April 23, 2019, substantial quantities of plastic covered the Durban harbor, in South Africa, after a flood event [30]. 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 were collected from the work of Kikaki et al. [21] and their observations over the Bay Islands and Gulf of Honduras.

In our spectral library, the mean spectrum for plastic computed from the 206 individual pixels is characterized by two reflectance peaks, one centered at B3 and the other at B8, and one absorption peak centered at the fifth Sentinel-2

TABLE II
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	[20]	15/12/2018	Limassol, Cyprus	2A	0	4	0	0	0	0	0
	[16]	07/06/2018	Tsamakia beach, Greece	2A	0	1	0	0	0	0	0
	[17]	18/042019		2B	0	3	0	0	0	0	0
		18/05/2019		2B	0	2	0	0	0	0	0
	[18]	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
	[21]	09/10/2017	Caribbean Sea, Honduras	2A	0	49	0	0	0	0	0
		03/11/2016		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

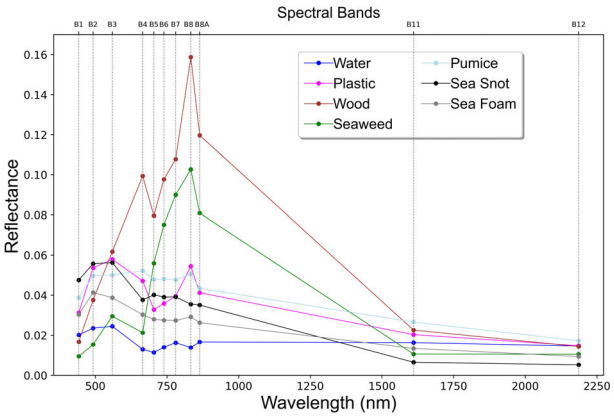

Fig. 1. Mean spectra computed from all pixels identified in this work after the atmospheric correction process. Despite the Sentinel-2 satellites' spectral bands having slightly different central wavelengths (Table I), this figure uses the same central wavelengths to facilitate interpretation. B9 (945 nm) and B10 (1375 nm) were removed in the atmospheric correction process.

spectral band (B5). Also, these plastic-like debris have higher reflectance values in all spectral bands compared to the water spectral response (Fig. 1).

E. 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 [18] allowed the collection of 62 pixels of driftwood on 13 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.

In our data library, the mean spectral response for driftwood, computed from the 62 pixels, is characterized by substantially more reflectance when compared to water or plastic (Fig. 1), and it has two main reflectance peaks at B4 and B8.

F. 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, large quantities of 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 between seaweed species, as considerable variations in the various seaweed reflectance are not expected (see [22]). One single Sentinel-2B image from October 2018 was used to collect 150 pixels of seaweed in the coastal waters of Accra, Ghana.

The mean seaweed's spectral response shows a sharp increase in reflectance in B4 (Fig. 1), followed by a fall in band 8A, being very distinct from the spectral responses of water, plastic, and driftwood.

G. Pumice

Pumice is a light-colored volcanic rock with a foamy appearance. It is formed when super-heated and highly pressurized molten rock (i.e., magma) is ejected from a volcano into the ocean waters 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 October 26, 2021, reveals thousands of bright pixels containing large quantities of floating pumice stone and was used to collect 31 098 pixels of this floating material.

In our library, the average spectrum of these pixels (Fig. 1) is close to the plastic mean response. However, in our data library, plastic presents an absorption trough in B5, which does not happen with pumice.

H. 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 June 6, 2021, showed thousands of pixels containing sea snot. From those, 26 403 pixels were selected.

Fig. 1 is an example that shows why Hu et al. [31] concluded that remote differentiation of sea snots and floating plastic debris using multiband sensors is problematic. Both materials have similar absolute reflectance values for the bands considered. The largest discrepancy happens for band B8, where plastics exhibit a reflectance peak. Also, there is a difference between the gradients between bands B3 and B4 and B4 and B5 for sea snot and plastic, respectively. The reflectance differences between these types of materials might be related to pixels with mixed materials (i.e., nonpurity pixels).

I. Sea Foam

The model proposed by Biermann et al. [7] showed some difficulties in distinguishing floating plastic from sea foam, bubbles, and froth. A Sentinel-2A image from Vigo Ria, Spain, was used to gather 2735 pixels of sea-foam accumulation.

In our data library, the mean sea-foam reflectance curve exhibits a similar pattern when compared to the one from plastic (i.e., the reflectance peaks and troughs are in the same bands) but with different absolute reflectance values (Fig. 1).

In summary, we collect 60 807 pixels in total. However, around 98% of these pixels are from pumice (51%), sea snot (43%), and sea foam (5%), culminating in an imbalanced dataset, which might bias the training process of the automatic classifier. Automatic classification models trained on imbalanced data sets usually have poor results when they need to generalize (i.e., classify unseen samples) since they cannot fully model the underlying patterns of the data and are prone to overfitting the majority class.

Multiple methods can be adopted to mitigate this problem and improve the performance of the automatic classification algorithm. One is to collect more data from the minority classes, which, in satellite data, might be challenging. Alternatively, models that penalize misclassifications from the minority classes more than the majority classes can be used. However, reaching the most suitable penalization values is difficult, might be subjective, and time-consuming. Another solution for imbalanced datasets is to resample the data (i.e., undersampling or oversampling). Undersampling balances the class distribution by removing samples from the most represented classes. Oversampling generates more observations (i.e., synthetic data) of the minority classes. This can be carried out, for example, with generative models (e.g., variational autoencoders, GANs, and synthetic minority oversampling technique (SMOTE) [38]).

In the application example shown herein, we used the random technique because of its implementation simplicity. Despite removing samples randomly, this method allows the user to select the ratio of each class in the data. After its execution, we confirmed that the statistical properties of each class were preserved. The training set distribution is shown in Table II and includes 156 pixels of suspect plastics, 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 water corresponds to 13.7% of the training data, with 98 pixels. There is still a considerable difference between the number of driftwood pixels and other classes as it is difficult to collect pixels associated with driftwood accumulations, as these are not common in the ocean. On the other hand, the mean spectral curve in our data library is very distinct from the others (Fig. 1), which facilitates its automatic detection.

J. Synthetic Data

Despite the efforts in gathering public domain samples of different floating materials, the number of validated satellite pixels obtained is relatively small for automatic classification algorithms. To overcome this limitation, we apply data augmentation methods. Data augmentation methods enable the classification 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 labeling data. In this work, applying classical techniques to 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 are not realistic for the floating debris considered. We opted to use

GANs [32] to create synthetic pixels that replicate patterns and features of the actual data, while adding variability to the dataset.

Briefly, a GAN comprises two deep adversarial neural networks: the generator and the discriminator. These networks are trained on a set of training samples. The generator's goal is to create samples that are indistinguishable from the training data. On the other hand, the discriminator tries to distinguish real data from the data generated by the first model (i.e., distinguish between true and fake data). Due to the widespread of this generative model in remote sensing and geosciences, we do not show a detailed description of both networks.

To avoid mode collapse (i.e., the generator produces a single type of output, usually close to the mean of the original data), we opted for a Wasserstein generative adversarial network (WGAN) [33]. A 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. 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 during the training stage. The loss function [see (1)], the Wasserstein distance, evaluates 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 [i.e., maximize (1)]. Inversely, the generator's goal is to minimize the distance between real and fake data scores [i.e., minimize (1)]. In the loss function [see (1)], f_ω acts as a critic and satisfies the Lipschitz constraint [see (2)], where the Lipschitz constant, K, represents the maximum value for the critic's gradients, 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)

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

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|, \quad \forall x_1, x_2. \tag{2}$$

WGAN was used to generate 50 000 pixels from each class (i.e., 350 000 pixels in total). We decided on this number of synthetic pixels after trial and error. During these experiments, the main objective was to ensure enough variability and represent the observed standard deviation within the ensemble. Since the water, plastic, and driftwood datasets were identified from images acquired by both Sentinel-2A and Sentinel-2B satellites, the synthetic pixels for these classes reproduce both sensors' characteristics. The remaining classes were generated based exclusively on a single sensor: Sentinel-2A for pumice and sea foam, and Sentinel-2B data for seaweed and sea snot. Fig. 2 shows the comparison of the class-dependent reflectance between original and synthetic data in each class.

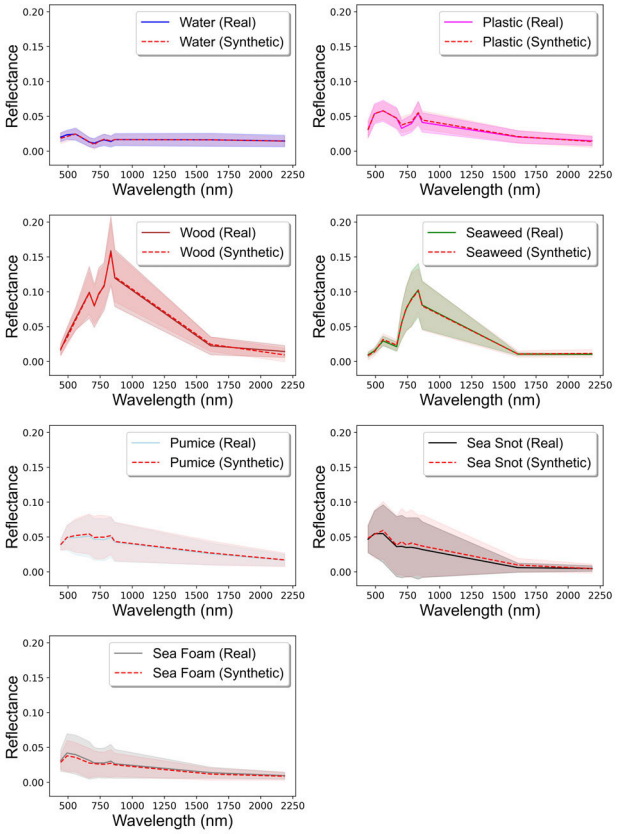

Fig. 2. Comparison between mean (lines) and standard deviation (shaded areas) of the spectral reflectance values from every class of real pixels and synthetic pixels generated from a WGAN (in red).

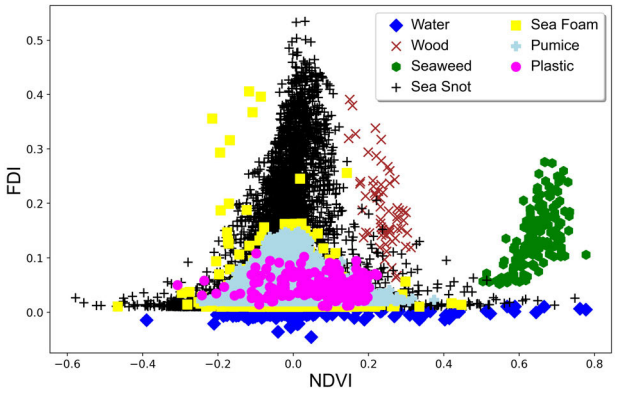

Fig. 3. Combination of the NDVI and the FDI of all real data collected under the scope of this work.

IV. Methodology

A. Spectral Indices

Biermann et al. [7] demonstrated that using the NDVI together with the FDI allows a distinct clustering of water, plastic, driftwood, and seaweed. In this 2-D domain, simple classification algorithms such as Naïve Bayesian classification can be applied successfully to automatically detect these floating materials from satellite images.

Nevertheless, using exclusively these indices does not enable a clear distinction between sea snot, sea foam, pumice, and suspect plastic (Fig. 3). This conclusion emphasizes the need for alternative combinations of spectral indices and bands that maximize the differences between these four classes. Including more variables for classification requires simultaneously advanced classification algorithms and a larger training dataset.

Besides the original spectral bands of the Sentinel-2 sensor, we compared 24 indices to verify which ones allow a distinct clustering of all the classes gathered in the data acquisition process: FDI, PI, NDVI, reversed normalized difference vegetation index (RNDVI), green normalized difference vegetation index (GNDVI), pan normalized difference vegetation index (PNDVI), NDWI, modified normalized difference water index (MNDWI), normalized difference moisture index (NDMI), normalized difference snow index (NDSI), water ratio index (WRI), normalized 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 in the following equations. In each equation, BX represents the reflectance value for the Sentinel-2 spectral band X, and in (4), λ_{BY} represents the central wavelength of the Sentinel-2 spectral band Y

$$\text{NDWI} = \frac{B3 - B8}{B3 + B8} \tag{3}$$

$$\text{FDI} = B8 - \left(B6 + (B11 - B6) \cdot \frac{\lambda_{B8} - \lambda_{B4}}{\lambda_{B11} - \lambda_{B4}} \cdot 10 \right) \tag{4}$$

$$\text{MNDWI} = \frac{B3 - B12}{B4 + B12} \tag{5}$$

$$\text{NDSI} = \frac{B3 - B11}{B3 + B11_2} \tag{6}$$

$$\text{WRI} = \frac{B3 + B4}{B8 + B12} \tag{7}$$

$$\text{MARI} = \frac{1}{B3} - \frac{1}{B5} \cdot B7 \tag{8}$$

$$\text{OSI} = \frac{B3 + B4}{B2}. \tag{9}$$

B. Extreme Gradient Boosting Classifier

Several supervised learning models were tested for the automatic classification of processed satellite images in a pilot area (e.g., random forest and k-nearest neighbor algorithm). Extreme gradient boosting (XGBoost) [34] was selected because of its relative easiness of implementation and the quality of the results obtained in these tests.

XGBoost is a tree-based ensemble machine learning algorithm trained using the boosting technique. During training, trees are grown sequentially so that each new tree corrects the classification errors of the previous one iteratively. Gradient descent [see (10)] establishes each new tree's parameters or weights.

The gradient descent algorithm is an optimization method used to minimize a function (i.e., the loss function) by iteratively computing the next point $[x_{(n+1)}]$ using the negative of the gradient at the current position $[-\nabla F(x_n)]$. The learning rate, γ, is the parameter with the most influence on the success of this method. It scales the gradient and, therefore, controls the step size. The steps should not be too big because the algorithm may not converge to the optimal point or too small since it will make the algorithm very slow. In summary, this method starts by selecting a starting point (x_0), computes the gradient at this point, and moves in its opposite direction based on its learning rate. It repeats these last two points until it converges, i.e., finds the optimal values where the function is minimum

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n), \quad n \geq 0. \tag{10}$$

XGBoost 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 regularizations to penalize the values of the weights. Both features aim to prevent the models from overfitting.

C. Uncertainty Quantification

XGBoost is a deterministic classification method and thus predicts a single solution for a given input data set. Due to the lack of spatial and spectral resolution of the satellite sensors, floating debris and water mixing, and band-to-band registration errors, there is uncertainty in the spectral responses obtained and consequently in the predictions. These uncertainties should be included in the final output of the predictor. Assessing uncertainty is critical in many Earth-related problems as it improves the information's reliability, which leads to better decision-making.

Uncertainty can be caused by the data (aleatoric uncertainty) and the model (epistemic uncertainty) [35]. In this study, we assess uncertainty through ensemble methods. Ensemble methods evaluate uncertainty based on the predictions of multiple models (i.e., ensemble members). These models are trained independently of each other using different techniques to increase their variety. Then the predictions' mean, variance, and standard deviation are computed to estimate the uncertainty. This approach has a high computational cost. However, ensemble methods were proven more reliable and applicable to real-world applications than alternative methods [35].

D. Outline

Fig. 4 summarizes the steps developed to reproduce the proposed classification model capable of detecting and distinguishing different types of floating materials. The first task is to collect in situ data (i.e., pixels with validated presence of floating materials). Second, the satellite images need to be preprocessed to remove the contribution of the atmosphere from the reflectance measured by the MSI. Also, a land mask needs to be applied to remove unnecessary pixels. 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 generating synthetic data to create samples of undersampled classes. Then, we compute the features that help separate different floating materials (i.e., spectral bands and indices). The classification model uses these features in its training process. Apart from the feature selection, the training of the classification

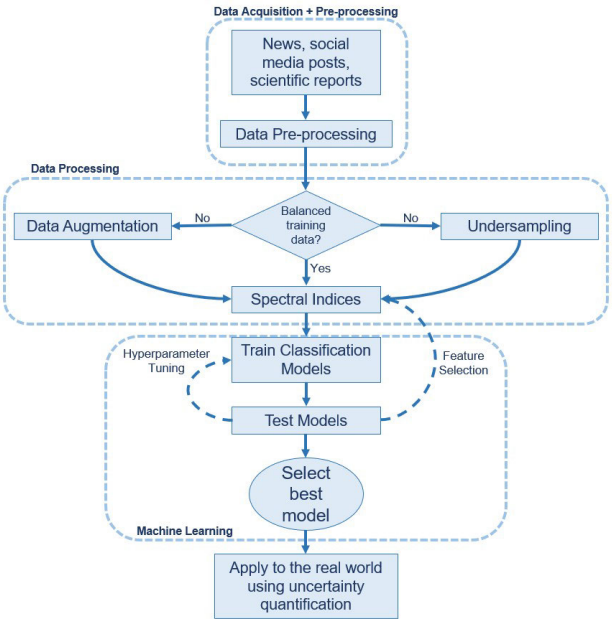

Fig. 4. Flowchart of the steps adopted to automatically identify floating debris on the ocean using satellite imagery.

model also involves hyperparameter tuning. Contrary to model parameters learned in the training process, hyperparameters define how the model is structured. Hyperparameter tuning comprises finding parameter values that optimize the model's performance. In the application example shown below, selecting the optimal combination of hyperparameters was based on the grid search algorithm [36]. This algorithm tests every combination of hyperparameters in a user-defined list and returns the one that has produced the best performance. This method has a high computational cost, but it is effective as long as the prior range of hyperparameters is within reasonable ranges. After training, the models are tested, and the one that delivers the best results is ready to generate predictions from Sentinel-2 images.

Finally, to improve the information provided by the model and help the decision-making process, we tested an uncertainty assessment method.

V. Results

A. Classification With Spectral Bands and Spectral Indices

We start by assessing the most relevant spectral bands and indices that boost the differences between the classes' reflectance with the proposed classification algorithm. XGBoost was trained with all spectral bands and indices described in Section IV-A, and the training and testing datasets are shown in Table II. The model achieved high accuracy (>90%) in every class except sea foam. However, the goal of this application example was to determine how each feature (i.e., spectral bands and indices) affects the overall accuracy of the model. This was accomplished through permutation importance [37]. This method focuses on answering one question: if one input feature, in this case, a spectral band or a spectral index, is randomly shuffled while all the other input feature stays 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

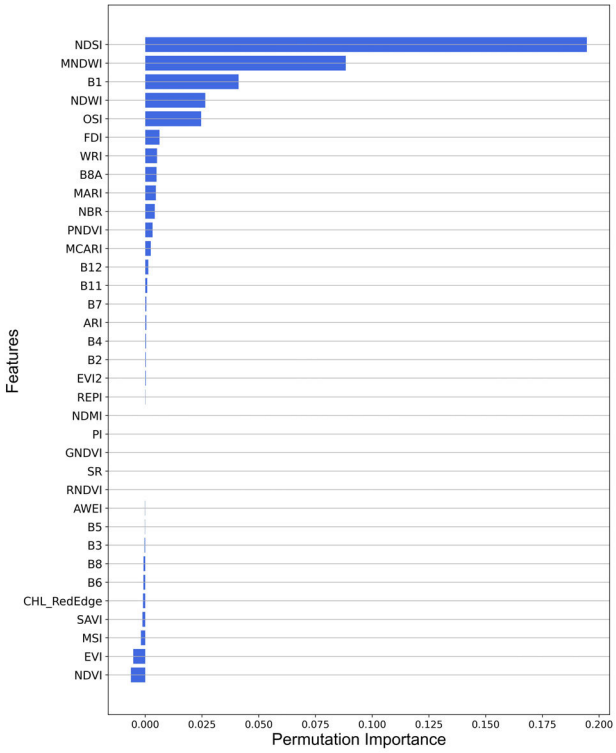

Fig. 5. Permutation importance of each feature in the XGBoost model trained with all spectral bands and the spectral index described in Section IV-A.

shuffling that feature. The results obtained from permutation importance are shown in Fig. 5.

With this information, the irrelevant features in the model (i.e., features with nil or negative permutation importance such as GNDVI, SR, NDMI, RNDVI, PI, B3, B8, B6, CHL, SAVI, MSI, EVI, or NDVI) were removed from the training process as they have little impact on the final predictions. The goal is to assess which combination of the remaining features maximizes the overall accuracy and minimizes the number of false plastic positives. Fig. 5 shows that the NDSI is the most important feature (i.e., with the highest permutation importance score), followed by the MNDWI and spectral band B1.

We summarize the results obtained in our application scenarios through a confusion matrix. A confusion matrix allows for quantitatively evaluating the performance of a classification algorithm. Each row of the matrix represents the samples in an actual class, while each column represents the samples in a predicted class. In this application example, since we are considering seven classes, the matrix is of size 7×7. In a confusion matrix, the diagonal elements denote correctly classified samples. On the other hand, all the nondiagonal elements represent the misclassified samples.

Fig. 6 shows the confusion matrix of a classification model. As an example, to study how the model performs with suspect plastic pixels, we focus on the second row and second column of the matrix. The diagonal element has the value 0.98, which tells us that 98% of the pixels presumably containing floating plastic were accurately labeled. The sixth element of the second row has the value 0.02, so we conclude that 2% of pixels expected to contain plastic were erroneously classified as sea snot (false negatives). Nondiagonal elements in the

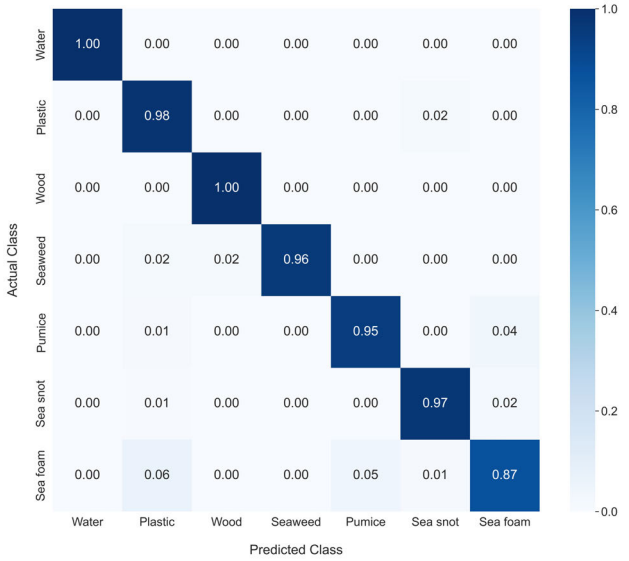

Fig. 6. Normalized confusion matrix of the XGBoost model trained with the nine features with the most permutation importance: band B1, band B8A, the NDSI, the MNDWI, the NDWI, the OSI, the FDI, the WRI, and the MARI. Results are shown rounded to two decimal points.

second column show us the false positives of plastics (i.e., the samples of other classes that were wrongly labeled as plastic).

Through interpreting the confusion matrices of several models trained with distinct features, we concluded that 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. The selected features comprise band B1, band B8A, the NDSI, the MNDWI, the NDWI, the OSI, the FDI, the WRI, and the MARI.

Fig. 6 shows the confusion matrix calculated from the predictions obtained by the XGBoost model with these nine input features. The model was initialized with a learning rate of 0.3, a maximum depth of a tree of 6, and a minimum child weight of 1. These hyperparameter values were chosen based on the results of the grid search algorithm. As a loss function, we selected the multiclass classification error rate, which is the ratio between the number of wrong predictions and all predictions. 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 labeled: 6% are predicted to be plastic and 5% are pumice. The model accurately predicts 98% of the suspect plastic pixels.

These results indicate that this model is ready to be applied in real-world conditions. Nevertheless, when inspecting the results, one must acknowledge that some seaweed, pumice, sea-snot, and mainly sea-foam pixels will be inaccurately labeled as plastic.

Despite the overall promising results of the proposed classification model, two aspects must be considered. First, 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, which affects further decisions regarding ocean clean-up. Therefore, efforts should be made to minimize

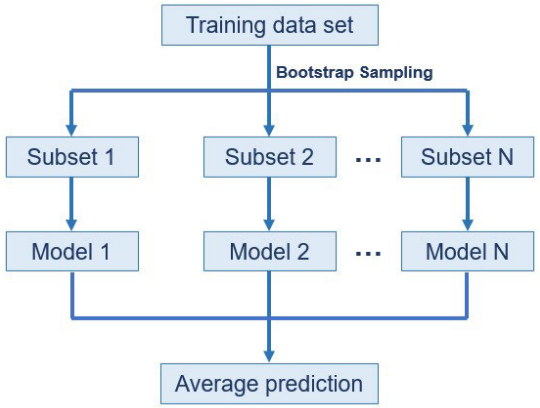

Fig. 7. Bagging-based algorithm with N classifiers. The training datasets of each model are generated through bootstrap sampling.

the number of false positives of plastic, knowing that it is impossible to achieve 100% accuracy in any classification problem. Second, predictions where the model shows little confidence because of aleatoric, 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, this implementation will probably result in a decrease in the number of pixels correctly labeled.

B. Classification With Uncertainty Estimation

To quantify uncertainty in the predictions, we developed a bagging-based algorithm with XGBoost classifiers (i.e., an ensemble of XGBoost models). Bagging, or bootstrap aggregating, is an ensemble technique that improves the stability and performance of automatic classification algorithms and helps to avoid overfitting. As we cannot fit multiple independent models in parallel due to the limited amount of data, we use bootstrap sampling to create random subsets of the training dataset. This method involves repeatedly drawing samples of data, with replacement, from the training dataset. Fig. 7 shows the proposed approach.

After testing, we used 20 parallel XGBoost models. To bring variety to the models, they were trained in different subsets of the training data generated by the bootstrap sampling method and initialized with distinct hyperparameter values. The models were initialized with a learning rate between 0.05 and 0.11, maximum depth of a tree between 2 and 7, and minimum child weight between 1 and 8. For each input sample, each model calculates its prediction. Then, the mean prediction's probability from all members and the standard deviation are computed. If the prediction has a standard deviation above 20 or a mean below 90%, the pixel is classified as uncertain. These thresholds were set after trial and error and can be considered conservative. Their selection should consider the final applicability of the classified images.

The results obtained with the ensemble model trained as previously described are shown in Fig. 8. As expected, the number of pixels correctly labeled decreased, compared to the previous model in every class except in seaweed, which is probably related to its singular spectral response within the training dataset. Around 23% of suspect plastic pixels were labeled uncertain, as well as 20% of pumice and 40% of sea

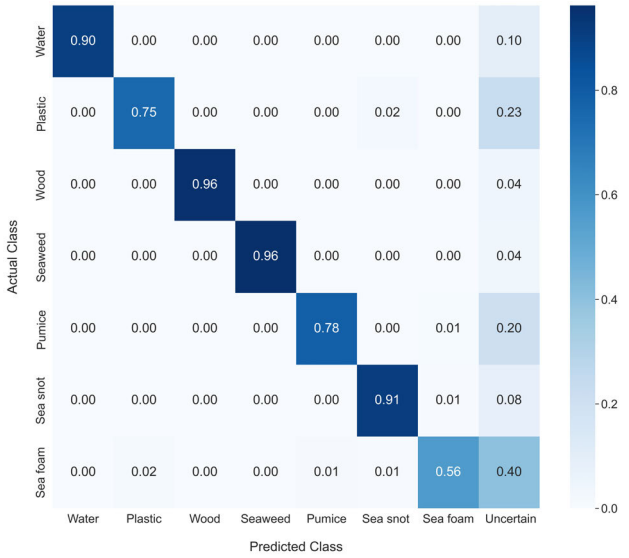

Fig. 8. Normalized confusion matrix of the ensemble model built with 20 XGBoost models trained with different data and different data sizes, using the nine features with the most permutation importance (Fig. 5). Predictions whose mean was below 90% or had a standard deviation above 20 were considered uncertain.

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.

To compare the results of both models, apart from the confusion matrix, we show real-world application examples. For this purpose, we use a Sentinel-2A multispectral image from July 31, 2021, in the Gulf of Gera, Greece. On this day, the Marine Remote Sensing Group from the University of the Aegean, Greece, performed another experience for the PLP 2021 [18]. 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.

Fig. 9 reveals that the first model can detect and correctly classify the wooden and plastic targets. However, some pixels on their borders are misidentified. Pixels around the wooden target are classified as plastic or seaweed, and some pixels on the plastic's borders are labeled as sea foam. These misclassifications are related to the quantity of floating material in those pixels. Therefore, if the floating material does not cover a certain percentage of a Sentinel-2 pixel, the model cannot accurately identify the material. Despite these misclassifications representing a low number of pixels in this case, if we scale up to larger satellite images, it might originate thousands of wrong predictions. Thus, as previously discussed, to accomplish the goal of providing meaningful information to the decision-makers, there is a need for a model that quantifies the predictions' uncertainty.

Fig. 10 shows the predictions of the ensemble model. It can correctly identify most pixels, but it classifies both targets' borders as uncertain (except for one pixel that is still labeled as seaweed). The ensemble model also identified the two previously misclassified nearshore pixels as uncertain. These results reveal the benefits of quantifying uncertainty in real-world conditions. Its ability to label some pixels as uncertain instead of providing misclassifications constitutes a substantial advantage compared to the previous classification model. Nevertheless, one must consider that despite the information given by this model is more trustworthy, small accumulations of floating materials may not be detected. Hence, there is a tradeoff between the output's reliability and the model's ability to identify small patches of floating materials. Fig. 11 shows the mean spectrum of all plastic data used to train the deep learning model (red line) and the individual spectra of pixels classified as suspect plastic by the model in Fig. 10 (colored lines). The most individual spectrum exhibits similar behavior to the mean reflectance of all plastic data, suggesting that the model correctly identified them as plastic debris. Furthermore, it clarifies why the model identified them as suspect plastics.

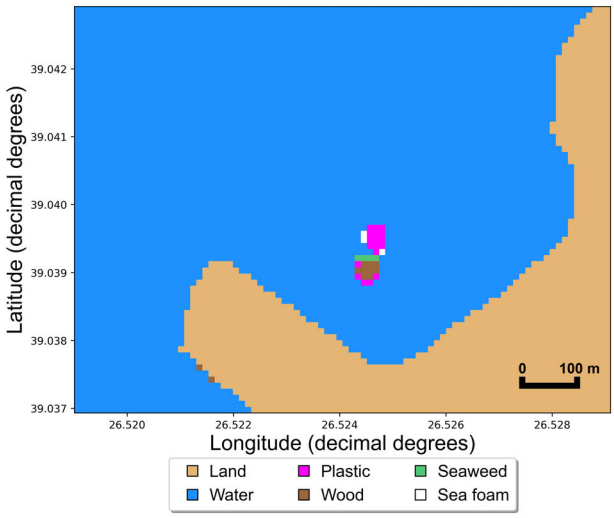

Fig. 9. Predictions from the XGBoost model that does not quantify uncertainty. The original data are from a Sentinel-2A image from July 31, 2021, in the Gulf of Gera, Greece.

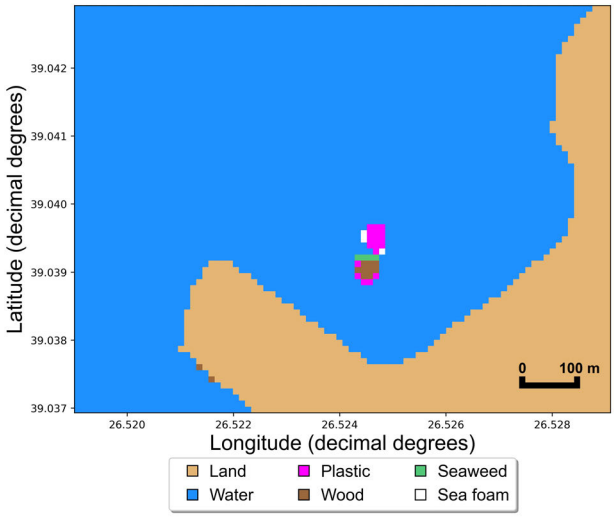

Fig. 10. Ensemble model's predictions. The original data are from a Sentinel-2A image from July 31, 2021, in the Gulf of Gera, Greece.

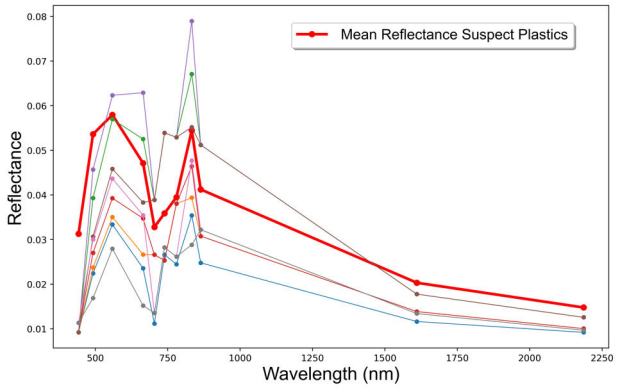

Fig. 11. Mean reflectance of all data classified as plastic used to train the model (red line) and the individual spectra of pixels identified as plastic by the model in Image 10 (shown in various colors).

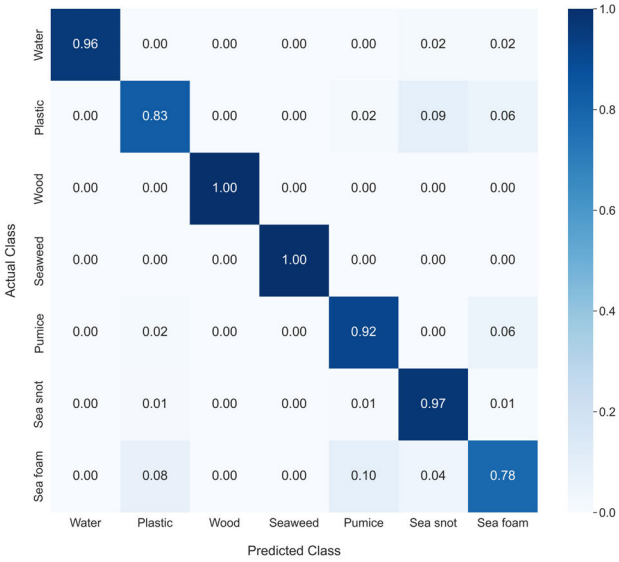

Fig. 12. Normalized confusion matrix of an XGBoost model trained with 50 000 synthetic pixels of each class.

C. Classification With Synthetic Data

We now show the benefits of including synthetic spectral data to increase the reliability and robustness of the proposed classification model. Theoretically, training a model with a larger balanced dataset with variability would make it more robust and less sensitive to outliers and mislabeled training data. Also, synthetic data samples may be a solution for the lack of data in this problem since some classes are hard to find.

We used 50 000 synthetic pixels from each class to create a balanced dataset. As discussed previously, we opted for this number of synthetic samples (350 000 in total) to ensure enough variability in the data. Fig. 12 displays the confusion matrix of an XGBoost model trained with these data and tested with the testing dataset described in Table II. The results show that synthetic data samples are a reliable solution to problems with a lack of data or data imbalance since the model achieved outstanding results. It accurately classified 83% of the suspect plastic pixels, and the number of false positives of plastic (i.e., pixels wrongly classified as plastic) is identical to the best model. On the other hand, this model shows more false

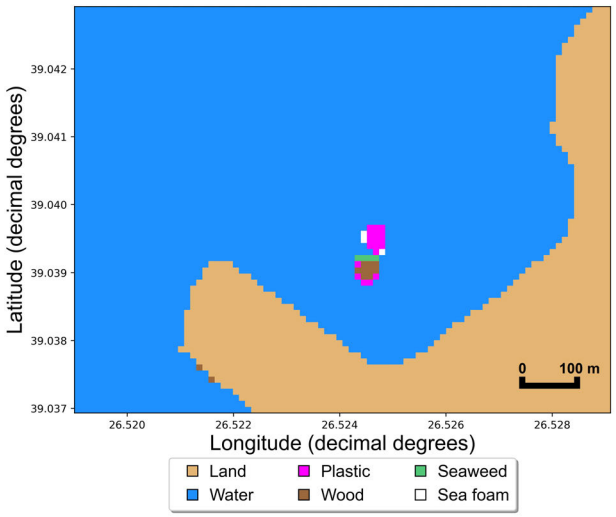

Fig. 13. Predictions from a Sentinel-2B image from September 4, 2021, in the Gulf of Gera, Greece. On this day, another experience for the PLP 2021 was performed. According to the authors, the sky was clear, but the sunlight intensity was lower than in previous experiments, and the weather was windy (8.5 m/s) [18]. They removed the wooden target and deployed some of it under the plastic target, simulating a mixed target that is closer to what is found in the ocean. The new target was in good condition, despite having some overworked points, and most of it was on the sea surface. The model successfully detected the plastic target. However, some of its pixels were considered uncertain, which is probably related to the presence of wood. Some uncertain pixels in the ocean may be related to sun glint or wave agitation. The plastic pixels near the shore are probably land pixels that were not removed in the land masking process.

negatives of plastic (i.e., plastic pixels wrongly classified as other classes) since 9% of plastic pixels were labeled as sea snot, 6% as sea foam, and 2% as pumice. All the other classes except sea foam are classified with accuracies above 90%, which proves that training a model with synthetic samples creates good results. Sea-foam pixels are correctly labeled 78% of the time, which shows that the model occasionally has trouble distinguishing this class from sea snot, pumice, and plastic. Despite the overall great results, we opted to use the ensemble model trained with real data not seen in the training or test datasets (i.e., in real-world conditions).

D. Monitoring the Ocean

Here, the ensemble model that quantifies uncertainty (Section V-B) is tested in real-world conditions, to assess its feasibility (Figs. 13–18).

In Figs. 13 and 14, most of the individual pixels classified as plastic exhibit a similar reflectance to that of plastic, which is a good indication that the model is detecting plastic debris. The standard deviation in the individual pixel spectra may be due to several factors, such as the presence of wood, sun glint, or wave agitation. Additionally, some of the pixels classified as plastic were near the shore, which could also contribute to the variability in their spectra.

The spectral analysis of the pixels identified as plastic in the Playa del Carmen image revealed a larger reflectance peak at B3 than expected (Figs. 15 and 16). This effect suggests that the plastic debris may have been mixed with other materials, such as seaweed or driftwood, which can affect the spectral

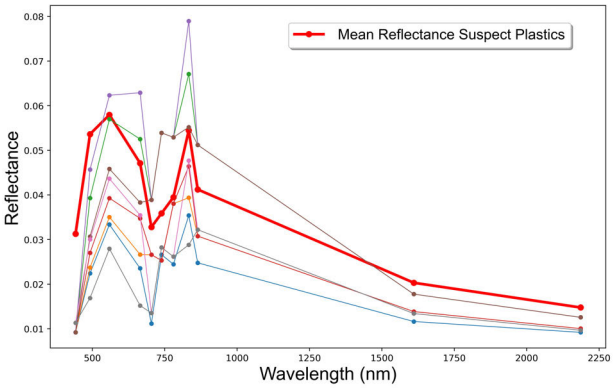

Fig. 14. Mean reflectance of all data classified as plastic used to train the model (red line) and the individual spectra of pixels identified as plastic by the model in Fig. 13 (shown in various colors).

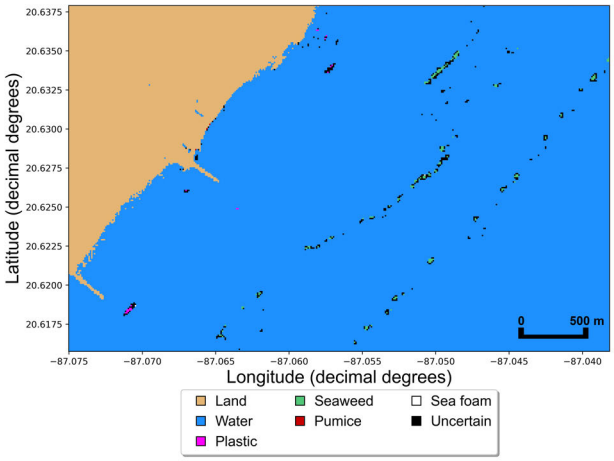

Fig. 15. Every year, beaches in Mexico suffer a massive invasion of Sargassum seaweed because of the water temperature, the increase of nutrients in the water, marine currents, and wind. Playa del Carmen is one of the most affected places. These predictions are from a Sentinel-2A image of Playa del Carmen, from June 15, 2022. The model only detected the seaweed accumulations that were further away from the shore. 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 they were mostly submerged. The model also classified a vessel that was not removed in the preprocessing phase as plastic. The model identified plastic in near shore pixels, suggesting that plastic pollution may also occur in these areas. These misclassificaitons highlight the challenges of accurately detecting and classifying plastic debris in satellite images.

signature of the plastic. Likewise, it is plausible that the materials in these pixels were not accounted for in the model's training process, explaining the deviation from the expected spectral signature. Furthermore, the model's misclassification of a vessel as plastic underscores the difficulty in distinguishing between plastic and other synthetic materials with similar spectral characteristics.

It is also important to note that detecting floating debris in controlled environments may differ from identifying debris in natural environments, which are more complex and may contain numerous materials with similar spectral properties (e.g., Figs. 17 and 18). These limitations highlight the need for further research to improve the model's accuracy in detecting plastic debris in real-world scenarios.

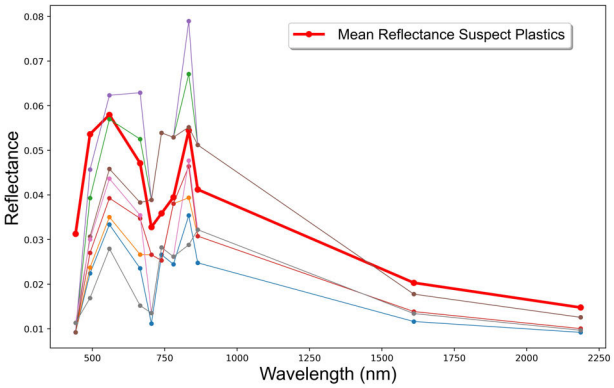

Fig. 16. Mean reflectance of all data classified as plastic used to train the model (red line) and the individual spectra of pixels identified as plastic by the model in Fig. 15 (shown in various colors).

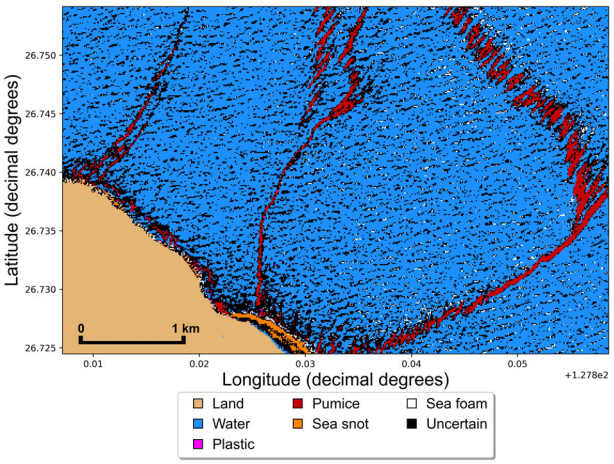

Fig. 17. Model's predictions from the Sentinel-2A image of Okinawa, Japan, from October 26, 2021, which was used to collect pumice pixels (Section III-B). None of the training data were collected from this specific subimage, so this is the first time the model is seeing these data. The model detects large accumulations of pumice and classifies their borders as uncertain. The remaining uncertain pixels may be due to ripples. Some pixels near the land that appear to be waves crashing on the shore were classified as sea snot instead of sea foam.

E. Limitations

The application examples shown herein exhibit most misclassifications in pixels close to the shore, where reflectance is usually higher because of the lower water depth, indicating 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 analyzing the model's predictions.

From our experiments, XGBoost was the deep classification tool with the highest performance. However, different training datasets might be more suitable for other classification algorithms. We stress the need for larger open-source databases with in situ validations to predict floating debris from Earth observation data more reliable.

In addition, in this work, we decided not to perform any spectral unmixing profile using ground-truth data. While this decision makes the classification task harder, it facilitates its potential applicability in real field conditions by removing a preconditioning data step. Notice that we are statistically classifying spectra for different floating materials considering

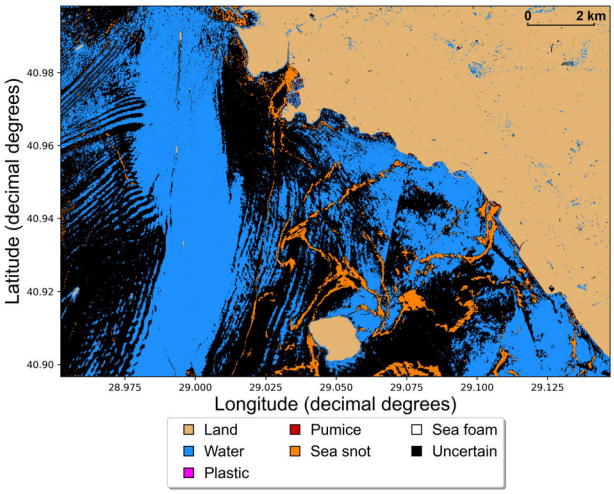

Fig. 18. Model's predictions based on a Sentinel-2B image from June 13, 2021, on the coast of Istanbul, where sea snot aggregations with hundreds of meters wide can be seen. There are also many pixels classified as uncertain because of different sea-snot depths in each pixel and the rough water conditions.

that they do not represent end members. The unmixing profile task could be performed for example using data acquired during the PLP [17], [18]. We use average spectra to train the classification algorithm. However, we do not handle uncertainties originated from different spatial resolutions of different spectral bands. This uncertainty layer originates band-to-band registration problems. This issue is discussed in detail [10].

Another source of uncertainty, which we tried to tackle with the proposed approach, is associated with the quantity of material in a pixel (i.e., floating debris-to-water ratio). The model regularly classifies the pixels of the image corresponding to the borders of the materials, where there is less quantity of the floating class, as uncertain. If not of enough size, above the spatial resolution of the sensor, floating debris accumulations will not be detected. 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 negatively affect the predictions. This usually happens in rivers, the primary conduits for plastic waste to the sea.

It is also worthwhile to mention that detecting plastic debris in a controlled environment (e.g., the PLP experiments in Greece) differs significantly from identifying plastics in a noncontrolled environment. The complex nature of coastal waters, including other floating materials and the effects of weather and lighting conditions, can make it challenging to accurately detect and classify these debris. As a result, a careful examination of the retrieved spectra is necessary.

Finally, the results obtained with the proposed automatic classification method hold exclusively for the data set considered using Sentinel-2 data. Alternative datasets and/or sensors would potentially result in the worst performance.

VI. Conclusion

A. Discussion

The main contributions of this work can be summarized as follows: 1) the data acquisition process allowed the collection of a large dataset related to suspect floating plastics, and other floating debris, in satellite imagery that is freely available to the community and 2) it shows that pixels suspected to contain floating plastic debris are distinguishable from five other classes of floating debris and water. The combination of features that proved most successful was using bands B1 and B8A, the NDSI, the MNDWI, the NDWI, the OSI, the FDI, the WRI, and the MARI. An XGBoost model trained with these features showed high overall accuracy. All the pixels of water and driftwood were correctly classified, 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. To reach this objective, we included uncertainty as part of the classification model using ensemble models. 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 revealed some limitations: clouds, shallower waters, suspended sediment, and rough sea conditions.

This study proposes an approach to generating synthetic data for deep learning classifiers due to the limited availability of validated spectral information and data on floating debris, especially plastic debris. We show the applicability of the proposed approach in real-case applications, where an XGBoost model is trained exclusively with data computed numerically. Training a model with synthetic data produces good results. However, as expected, the quality of the predictions decreases when compared to the XGBoost model trained with real data.

B. Future Work

One of the major 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. Furthermore, it is essential to validate the existing data through spectral analysis to confirm that they represent the considered debris accurately. This would ensure that the data used for training machine learning algorithms are reliable and can generalize to different regions and scenarios.

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 different atmospheric correction methods affect the detection of floating plastics. Finally, it would be relevant to assess if Sentinel-2 imagery can be used to detect floating debris in rivers since they are the main points of entry of plastic into the ocean.

While we deal with the identification of floating debris for a temporal snapshot (i.e., a single satellite image), coupling the temporal domain in this classification model (e.g., high-resolution Lagrangian ocean models) could allow us to model simultaneously the spatiotemporal evolution of the floating debris between the revisiting periods of the satellites to the study area, potentially accounting for the degradation of the material.

ACKNOWLEDGMENT

The authors acknowledge the fruitful discussions with Dr. Chuanmin Hu and the contributions from the anonymous reviewers that improved the quality of this article. The datasets used in this study are available at https://github.com/miguelmendesduarte/Floating-Marine-Debris-Data.

REFERENCES

[1] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Sci. Adv.*, vol. 3, no. 7, Jul. 2017, Art. no. e1700782.

[2] J. R. Jambeck et al., "Plastic waste inputs from land into the ocean," *Science*, vol. 347, no. 6223, pp. 768–771, Feb. 2015.

[3] L. Lebreton et al., "Evidence that the Great Pacific garbage patch is rapidly accumulating plastic," *Sci. Rep.*, vol. 8, no. 1, Mar. 2018, Art. no. 4666.

[4] L. Barboza, A. Cozar, B. Gimenez, T. L. Barros, P. Kershaw, and L. Guilhermino, "Macroplastics pollution in the marine environment," in *World Seas: An Environmental Evaluation*, vol. 3, 2nd ed., C. Sheppard, Ed. New York, NY, USA: Academic, 2018, 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, Jun. 2020, Art. no. 1900081.

[6] N. J. Beaumont et al., "Global ecological, social and economic impacts of marine plastic," *Mar. Pollut. Bull.*, vol. 142, pp. 189–195, May 2019.

[7] L. Biermann, D. Clewley, V. Martinez-Vicente, and K. Topouzelis, "Finding plastic patches in coastal waters using optical satellite data," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Apr. 2020.

[8] B. Basu, S. Sannigrahi, A. S. 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 Sens.*, vol. 13, no. 8, p. 1598, Apr. 2021.

[9] C. Hu, "Remote detection of marine debris using satellite observations in the visible and near infrared spectral range: Challenges and potentials," *Remote Sens. Environ.*, vol. 259, Jun. 2021, Art. no. 112414.

[10] C. Hu, "Remote detection of marine debris using Sentinel-2 imagery: A cautious note on spectral interpretations," *Mar. Pollut. Bull.*, vol. 183, Oct. 2022, Art. no. 114082.

[11] G. Suaria and S. Aliani, "Floating debris in the Mediterranean Sea," *Mar. Pollut. Bull.*, vol. 86, nos. 1–2, pp. 494–504, Sep. 2014.

[12] 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 Sens.*, vol. 12, no. 16, p. 2599, Aug. 2020.

[13] M. C. Sousa et al., "Modelling the distribution of microplastics released by wastewater treatment plants in Ria de Vigo (NW Iberian Peninsula)," *Mar. Pollut. Bull.*, vol. 166, May 2021, Art. no. 112227.

[14] S. Kako et al., "Sequential webcam monitoring and modeling of marine debris abundance," *Mar. Pollut. Bull.*, vol. 132, pp. 33–43, Jul. 2018.

[15] L. Fronkova, "Tackling marine litter in the Atlantic area, Deliverable 5.2—Monitoring the presence of marine litter in the marine environment," CleanAtlantic, 2019.

[16] K. Topouzelis, A. Papakonstantinou, and S. P. Garaba, "Detection of floating plastics from satellite and unmanned aerial systems (plastic litter project 2018)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 79, pp. 175–183, Jul. 2019.

[17] K. Topouzelis, D. Papageorgiou, A. Karagaitanakis, A. Papakonstantinou, and M. A. Ballesteros, "Remote sensing of sea surface artificial floating plastic targets with Sentinel-2 and unmanned aerial systems (plastic litter project 2019)," *Remote Sens.*, vol. 12, no. 12, p. 2013, Jun. 2020.

[18] Marine Remote Sensing Group. *Plastic Litter Project 2021*. Accessed: Dec. 18, 2022. [Online]. Available: http://plp.aegean.gr/

[19] K. Kikaki, I. Kakogeorgiou, P. Mikeli, D. E. Raitsos, and K. Karantzalos, "MARIDA: A benchmark for marine debris detection from Sentinel-2 remote sensing data," *PLoS ONE*, vol. 17, no. 1, Jan. 2022, Art. no. e0262247.

[20] K. Themistocleous, C. Papoutsa, S. Michaelides, and D. Hadjimitsis, "Investigating detection of floating plastic litter from space using Sentinel-2 imagery," *Remote Sens.*, vol. 12, no. 16, p. 2648, Aug. 2020.

[21] 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 Sens.*, vol. 12, no. 11, p. 1727, May 2020.

[22] L. Qi et al., "In search of floating algae and other organisms in global oceans and lakes," *Remote Sens. Environ.*, vol. 239, Mar. 2020, Art. no. 111659.

[23] European Space Agency. *Radiometric Resolutions—Sentinel-2*. Accessed: Dec. 18, 2022. [Online]. Available: https://sentinels.copernicus.eu/

[24] N. Hoepffner and G. Zibordi, "Remote sensing of coastal waters," in *Encyclopedia of Ocean Sciences*, 2nd ed., J. H. Steele, Ed. Oxford, U.K.: Academic, 2009, pp. 732–741.

[25] L. Rumora, M. Miler, and D. Medak, "Impact of various atmospheric corrections on Sentinel-2 land cover classification accuracy using machine learning classifiers," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 4, p. 277, Apr. 2020.

[26] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, May 1996.

[27] European Space Agency. *SNAP Download-STEP*. Accessed: Dec. 18, 2022. [Online]. Available: https://step.esa.int/

[28] European Space Agency. *Open Access Hub*. Accessed: Dec. 18, 2022. [Online]. Available: https://scihub.copernicus.eu/

[29] P. Tasseron, T. van Emmerik, J. Peller, L. Schreyers, and L. Biermann, "Advancing floating macroplastic detection from space using experimental hyperspectral imagery," *Remote Sens.*, vol. 13, no. 12, p. 2335, Jun. 2021.

[30] M. Aldersley. (2019). *South African Port Swamped in Plastic Waste and Debris After Floods*. Accessed: Dec. 18, 2022. [Online]. Available: https://www.dailymail.co.uk/

[31] 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 Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112842.

[32] I. J. Goodfellow et al., "Generative adversarial networks," 2014, *arXiv:1406.2661v1*.

[33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*.

[34] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[35] F. K. Gustafsson, M. Danelljan, and T. B. Schon, "Evaluating scalable Bayesian deep learning methods for robust computer vision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2019, pp. 318–319.

[36] P. Liashchynskyi and P. Liashchynskyi, "Grid search, random search, genetic algorithm: A big comparison for NAS," 2019, *arXiv:1912.06059*.

[37] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinf.*, vol. 8, no. 1, p. 25, Dec. 2007.

[38] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.