

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2026-10-13

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Costa, T. S., Felício, J., Matos, S. A., Costa, J. R., Fernandes, C. A. & Fonseca, N. J. G. (2025). Monitoring plastic accumulations in a river environment using machine learning on Sentinel-1 SAR data. 2025 19th European Conference on Antennas and Propagation (EuCAP). IEEE. <https://doi.org/10.23919/EuCAP63536.2025.10999662>

Further information on publisher's website:

[10.23919/EuCAP63536.2025.10999662](https://doi.org/10.23919/EuCAP63536.2025.10999662)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Costa, T. S., Felício, J., Matos, S. A., Costa, J. R., Fernandes, C. A. & Fonseca, N. J. G. (2025). Monitoring plastic accumulations in a river environment using machine learning on Sentinel-1 SAR data. 2025 19th European Conference on Antennas and Propagation (EuCAP). IEEE. <https://doi.org/10.23919/EuCAP63536.2025.10999662>, which has been published in final form at <https://dx.doi.org/10.23919/EuCAP63536.2025.10999662>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Monitoring Plastic Accumulations in a River Environment Using Machine Learning on Sentinel-1 SAR Data

Tomás Soares da Costa*, João M. Felício*, Sérgio A. Matos*[†], Jorge R. Costa*[†],
Carlos A. Fernandes* Nelson J.G. Fonseca[‡]

*Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal

[†]Instituto Universitário de Lisboa (ISCTE-IUL), 2649-026 Lisbon, Portugal

[‡]European Space Agency, 2200 AG, Noordwijk, The Netherlands (now with Anywaves, 31000 Toulouse, France)
tomas.costa@lx.it.pt

Abstract—Plastic litter is a growing environmental issue. Microwave (MW) technology has the potential for an all-time and weather remote detection of floating macroplastics via satellites or airborne platforms. However, none of the early studies in the existing literature has addressed MW sensor requirements or the application of Machine Learning (ML) for automated monitoring with readily available data. In this study, we employ a supervised learning workflow to monitor, in a preliminary phase, floating macroplastic accumulation in a river environment utilizing polarimetric Sentinel-1 (S-1) SAR data. The combination of co-polarization (VV) and cross-polarization (VH), specifically VV-VH and VV+VH, resulted in detection accuracies exceeding 90%, without overfitting. Analysis of scattering behavior revealed that both VV and VH backscatter were sensitive to plastic patch size, with a linear relationship. This differs from the low-intensity scattering of river water. The findings highlight the importance of dual polarization for effective MW-based plastic monitoring missions.

Index Terms—detection, floating macroplastics, machine learning, marine litter, polarization, propagation, Sentinel-1, SAR.

I. INTRODUCTION

Over the last five years, a significant amount of research has emerged, exploring the different remote sensing methodologies for floating macroplastics detection [1]–[3]. The bulk of these studies focuses on detection strategies aboard satellite and airborne platforms using optical or infrared instruments. The availability of open-source marine debris datasets, e.g., Sentinel-2 (S-2) Imagery [4], [5], has further encouraged the adoption of Machine Learning (ML) methodologies to automate and improve detection across different environments [6], [7]. However, the use of these wavelengths comes with inherent limitations that can negatively affect detection performance, particularly when using satellites [8]. Challenges such as cloud cover, shadowing, sunglint, and adverse weather conditions can all impair detection.

Microwaves (MWs), on the other hand, are less affected by atmospheric conditions and have been widely used in remote sensing for various scientific and Earth observation missions. Recent preliminary studies [9]–[15] have demonstrated the potential of MW technology to detect floating macroplastics in

small to mid-scale setups, as well as using Synthetic Aperture Radar (SAR) aboard satellite platforms [1], [16]. Satellites provide the geographical reach necessary to implement an automatic plastic detection service. However, the lack of research on the sensor requirements for an optimal, fully automated detection workflow has prevented the consideration of MW instruments for continuous, all-day monitoring.

To perform systematic plastic pollution studies using available datasets such as those from Sentinel-1 (S-1) or TerraSAR-X satellites, simultaneous knowledge of detailed ground truth data, i.e., labeled, is required (e.g., patch location, debris type and size, concentration, environmental conditions, etc.). This is practically impossible to obtain in the fast-changing ocean environment. However, rivers have been identified as hotspots for plastic accumulation, as they act as pathways for transporting plastic pollution to the oceans [17]. Land constraints, easily identified through masking, play a key role in simplifying the labeling of SAR data when combined with high-resolution optical data. With labeled data, we can apply supervised learning to monitor plastic accumulation in rivers and, potentially, extend the ML methodology to sea and ocean environments.

In this study, we propose an automatic detection workflow that utilizes ML to process polarimetric S-1 SAR imagery data. In this preliminary analysis, the workflow is aimed at monitoring yearly floating macroplastic accumulation in a riverine environment between October 2022 and November 2023. By considering different polarizations, i.e., co-polarization (VV), cross-polarization (VH), and their combinations (e.g., ratio, sum, and difference), as features in the ML model, we provide essential insights into the sensor requirements for an MW instrument to ensure optimal detection. This represents a crucial step towards developing an all-time and weather detection system.

For classification, we use a Random Forest (RF) algorithm [18] as our ML model to categorize the data as Water (class ‘0’) or Plastic (class ‘1’). We only consider the two classes as distinguishing between other floating objects requires more data. RF is an ensemble learning method that constructs a

large number of decision trees based on training data. This classifier is also relatively robust against outliers and noise, which is particularly important given the spatial variability of SAR data. In addition to performing supervised classification, the RF algorithm allows us to determine the importance of different features. By analyzing feature importance, we can identify which of the polarimetric features contribute most to the classification results.

Lastly, after monitoring the quantity of plastic over the selected time frame using the proposed ML workflow, we further enhance our analysis by characterizing the polarimetric scattering behavior of the plastic accumulations, i.e., backscattering mechanisms, and comparing it with the scattering characteristics of the river water surface.

II. MATERIALS

A. Study Area

The study area is a segment of the Drina River, located in Visegrád, Bosnia & Herzegovina. This region was selected for monitoring due to reports of plastic accumulation near prominent dams in recent years. In Fig. 1, the extent of the study area is indicated by the blue rectangle.

B. Data Sets

Within the selected time frame of October 2022 to November 2023, we collected a time series of 60 S-1 B images with identical orbit configuration (same image geometry). The Sentinel-1 mission consists of two satellites (Sentinel-1A/B) equipped with C-band sensors (wavelengths $\lambda = 5.6$ cm), which allow for Earth surface monitoring at a six-day repeat interval. The S-1 B images utilized both VV and VH polarizations in Interferometric Wide Swath (IW) mode and were processed by ESA as Ground Range Detected High Resolution (GRDH) products. Ground truth data was obtained through visual interpretation and manual digitization of seven randomly selected high-resolution optical S-2 images. When comparing the SAR and optical data, no changes in the extent of plastic accumulation were observed within a two-day span.

As the study focuses on predicting plastic accumulation areas within the water surface, land areas were excluded from the analysis and methodology. A binary mask was applied to segment the river area in the images, excluding land before applying the ML workflow. For ground truth image dates during SAR acquisition, we labeled the pixels into two classes:

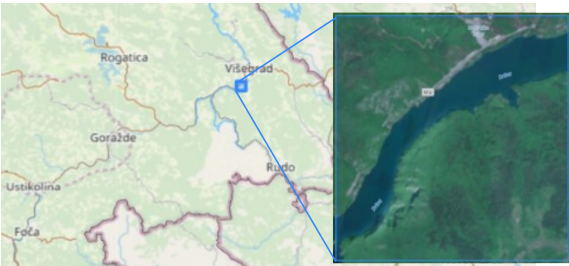


Fig. 1. Location map of the study area (blue rectangle) in Bosnia.

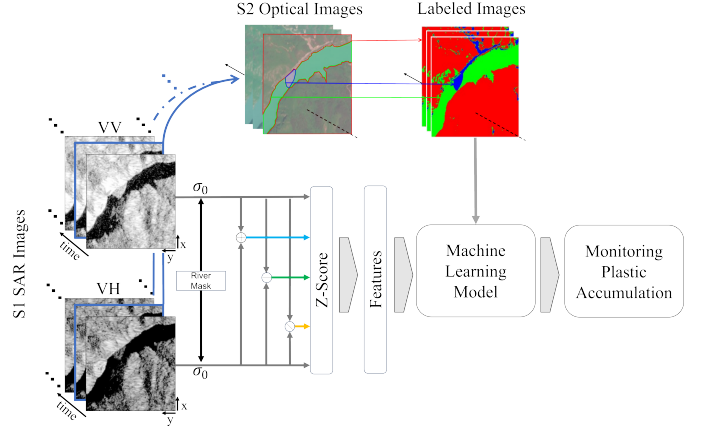


Fig. 2. Classification workflow for monitoring plastic accumulation in river.

‘0’ for Water and ‘1’ for Plastic. Due to the higher number of water pixels compared to plastic accumulation pixels, the image datasets are imbalanced.

III. METHODOLOGY

Fig. 2 presents the pixel-based classification workflow used to monitor plastic accumulation in the study area, utilizing the time series stack of S-1 imagery data. For conciseness, the S-1 image preprocessing steps are not depicted; however, they are essential for producing usable SAR imagery data and are detailed in Section III-A. The first step in our workflow involves normalizing and sequentially combining the VV and VH polarization as input features for our model (see Section III-B). Simultaneously, we apply the binary river mask previously mentioned to exclude land data from the classification. This is a crucial step in the workflow, as it simplifies plastic detection by converting it into a binary classification problem. As previously mentioned, we use the RF classifier as our ML model due to its robustness against outliers and noise, as well as its built-in feature importance estimation.

A. Image Preprocessing

Automated preprocessing of all S-1 images was performed using a Python script with the pyroSAR library [19]. This package provides a set of tools that utilizes Python to run the Sentinel Application Platform (SNAP) Graph Processing Tool (GPT). The preprocessing includes radiometric calibration to sigma naught (σ_0), range Doppler terrain correction using the SRTM digital terrain model, co-registration of the individual datasets, and speckle filtering. The co-registration of images is crucial for time series analysis, ensuring pixel-level accuracy so that backscatter values over time do not distort the interpretation of the desired classes. A spatially refined Lee filter with a window size of 7×7 was applied to reduce noise in the SAR data. This filter preserves image structure by smoothing homogeneous areas while retaining edges, as well as plastic-related temporal characteristics and patterns. These preprocessing steps were applied to both VV and VH polarizations, resulting in two independent time series. The outcome is a multi-temporal layer stack for each polarization.

B. Feature Extraction

Backscatter values are influenced by various environmental conditions, particularly wind, which can cause fluctuations in intensity, either increasing or decreasing within specific categories. This variability makes it challenging to directly compare absolute backscatter values over time. To ensure comparability, the backscatter values of each pixel are normalized across the time series. This normalization is performed using the Z-transform, or Z-score, which is computed using the following formula:

$$Z = (x - \mu) / \sigma. \quad (1)$$

In Eq. (1), x represents the pixel value of any image in the time series, while μ and σ denote the mean and standard deviation of the backscatter values for that pixel across the entire time series. The Z-transformed backscatter values allow for spatial comparability across the entire time series. By normalizing the backscatter changes, the Z-score, along with the respective polarizations (VV, VH), their ratio (VV/VH), and their combinations (VV + VH, VV - VH), constitute the key polarization features used in the subsequent classification step.

C. Binary classification

A quantitative analysis using all polarization features was carried out with supervised learning, employing the RF algorithm as our classifier. We trained and tested the ML model using five and two labeled images. During the training phase, we tuned the model's hyperparameters using a nested cross-validation (CV) approach [12] and grid search [18]. Once the optimal hyperparameters were determined, we evaluated the model's performance on the test images. We consider the typical classification metrics used in remote sensing to evaluate detection [20]: Overall Accuracy (OA), Kappa Coefficient (K), Producer and User Accuracy (PA and UA, respectively). These metrics are easily obtainable from the confusion matrix. Additionally, we assess the importance of each polarimetric feature using the RF algorithm, i.e., which polarization (or combination) has the highest contribution to the classification results.

To further enhance this polarization analysis, we repeat the training phase and the sequential evaluation on the test set using the RF model but with fewer polarization features. Specifically, we consider using the individual VV or VH features and the two highest contributing features from the previous analysis. Comparing output metrics consolidates the identification of the essential polarimetric information for the detection.

IV. DISCUSSION AND RESULTS

A. Classification Results

Fig. 3a shows a comparison of the output metrics for different numbers of selected features. Considering all five polarizations features, we obtain an OA and Kappa coefficient of 99.1% and 92%, respectively. For the Water class, the

Features Used		Single		Double	Quintuple
Metrics (%)	Class	VV	VH	VV-VH; VV+VH	VV; VH; VV/VH; VV-VH; VV+VH
Overall Accuracy		88.7	87.7	99.0	99.1
Kappa Coefficient		14.4	2.3	91.0	92.0
Producer Accuracy	0	93.7	93.4	99.5	99.6
	1	21.3	10.0	90.0	91.2
User Accuracy	0	94.2	93.4	99.3	99.4
	1	14.4	9.8	93.3	93.9

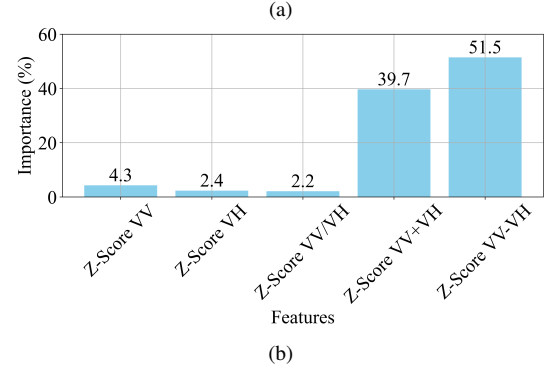


Fig. 3. Classification output from the ML workflow. (a) Summary of output metrics considering different polarimetric input features; (b) Feature importance considering all the polarimetric information as input features.

producer's accuracy (PA) and user's accuracy (UA) are 99.6% and 91.2%, respectively. For the Plastic class, the PA is 99.4% and the UA is 93.9%. Despite having an imbalanced dataset, these results demonstrate that plastic accumulation is effectively detected when using all polarimetric information. Combining different polarizations helps amplify effects such as varying sensitivities to cluster size, potentially enhancing detection accuracy through mathematical transformations of the data. This amplification was quantified using the RF algorithm, resulting in excellent classification accuracy for both classes.

Fig. 3b, illustrates the feature importance (see Section III-C) derived from the RF algorithm for the classification task using all polarimetric information as features. The polarimetric combinations VV-VH and VV + VH contribute the most, accounting for 51.5% and 39.7% of the feature importance, respectively. However, the highest contributing feature does not always guarantee the best classification accuracy. Therefore, an additional analysis is conducted to determine whether using a single feature (VV or VH) or the two most important features (VV-VH and VV + VH) would yield similar classification performance.

In the aforementioned Fig. 3a, we also observe the output metrics for both the single and double feature cases. For the single feature scenarios (VV and VH), we observe an OA of 88.7% and 87.7%, respectively. While these values are still quite good, the disparity in PA and UA for classes '0' (Water) and '1' (Plastic) - with approximately 93.5% accuracy for Water and only 14% for Plastic - along with a low Kappa coefficient, indicate slight overfitting to the Water class.

This suggests that the RF model struggles to separate plastic pixels. Still, comparing the two, VV produced better results. Regarding the double feature case (VV-VH and VV + VH), the OA remains high at 99%, with no evidence of overfitting from the complementary metrics. These findings underscore the importance of using dual polarization (vertical and horizontal) in an MW instrument to obtain optimal detection.

B. Time Series Monitoring

Building on the excellent classification accuracy obtained in the previous analysis, we now focus on monitoring plastic accumulation in the study area over the defined time period using all polarimetric information as features. The trained RF model incorporating all the polarimetric features is applied to predict the number of plastic pixels over time. Fig. 4 presents the predicted number of plastic pixels in the SAR image time series, along with the normalized backscatter variation per m^2 in VV and VH polarizations for the two possible classes.

The analysis of SAR time series data reveals a distinct pattern: a noticeable increase in backscatter values for the Plastic class during periods of higher plastic accumulation, i.e., a linear relationship. This behavior contrasts with the typical backscatter in a river environment, where low backscatter

values are observed. In these cases, the smooth water surface causes most of the emitted sensor energy to be reflected away due to specular reflection. However, when floating macroplastics accumulate, they create a rough surface, which results in higher backscatter values. In the next subsection, we will further investigate the scattering differences of the two classes.

C. Scattering Analysis

We begin by visualizing the distribution and separability of the training data for the two classes (Water and Plastic) in relation to the polarization features, using histograms. Fig. 5 shows the histogram distributions of the training data for the Water class (blue) and the Plastic class (red) across the different features. Clear separability between the two classes is observed for the VV + VH and VV - VH feature combinations. However, there is significant overlap between the histograms for the VV/VH feature, as well as between the individual VH and VV features. Among these, the overlap is more pronounced in the VH feature. These results are consistent with the previous classification outcomes on the test data (see Section IV-A).

From the histograms, it is evident that the VV/VH feature is less effective for identifying plastic accumulation. This can be attributed to the similar sensitivity of VV and VH to backscattering mechanisms, causing the backscatter values to rise or fall together in response to plastic accumulation. This pattern was already observed in Fig. 4, where the temporal backscatter trends for both polarizations were similar and correlated with the number of plastic pixels.

By examining the VV and VH histograms and the classification results using VV and VH as standalone features (refer to Subsection IV-A), we can see that the Plastic class is more separable and reliable when using the VV polarization. The increase in backscatter in the VV polarization is primarily associated with specular backscattering, which indicates that single-bounce effects are dominant. This suggests that specular backscattering plays a more significant role. However, the contribution from the VH polarization indicates that depolarizing effects from the rougher surface are also present, confirming the existence of higher-order scattering, particularly double-bounce effects. Therefore, the plastic accumulation increase backscatter intensity by embedding both surface and volumetric scattering effects in the MW response. This means that, while the plastic patch produces an increase on surface roughness on the accumulation area, the individual low permittivity plastic elements can still be interpreted as indentations in the surface [9].

Given the contrasting behavior of the "darker" water pixels, these two contributors to backscatter intensity—specular and higher-order scattering—are crucial for detecting floating macroplastic accumulations. This is why the difference and sum of the two polarizations, which capture these effects, are the most important and reliable features for accurate classification in this environment.

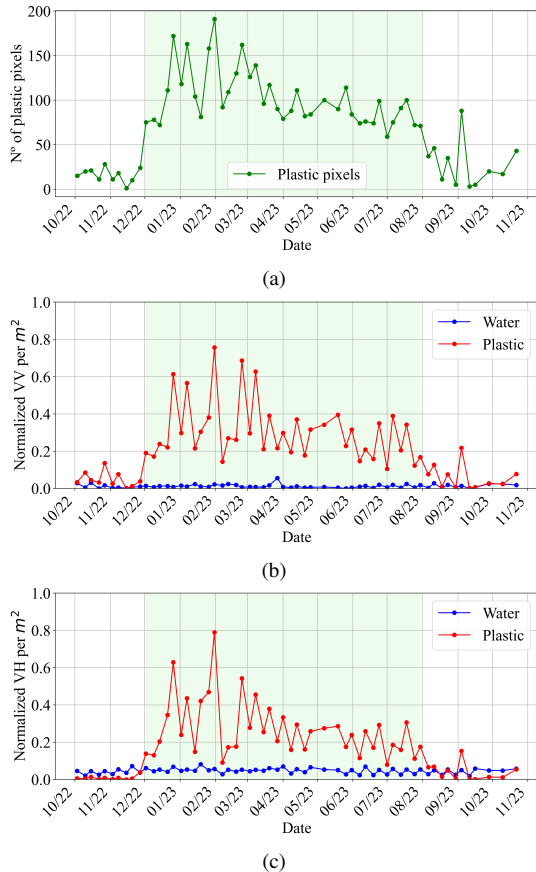


Fig. 4. Monitoring plastic accumulation. (a) Predicted number of plastic pixels in the SAR image time series (in light green, period with highest plastic accumulation); and temporal variation of the normalized backscattering: (b) VV; (c) VH, for the two possible classes.

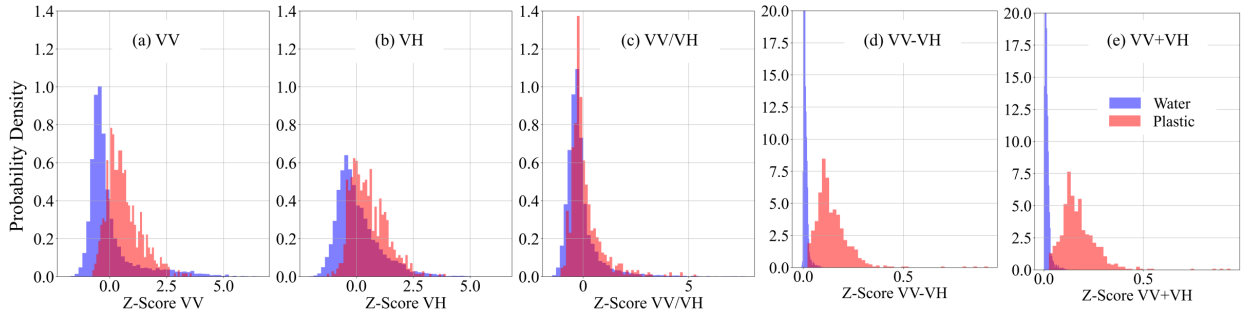


Fig. 5. Normalized histograms of polarimetric features of ground truth data: (a) VV; (b) VH; (c) VV/VH; (d) VV-VH; and (e) VV+VH.

V. CONCLUSIONS AND FUTURE WORK

The reported study demonstrated that ML can be effectively used to automate the monitoring of plastic accumulations in river environments using S-1 SAR images. We addressed a significant gap in the literature by investigating the use of ML algorithms and identifying MW sensor requirements for plastic detection. The results show that optimal detection is obtained through a combination of polarimetric features, specifically VV-VH and VV+VH, indicating that dual polarization is essential for MW instruments aboard platforms to conduct all-time and weather plastic monitoring missions.

In the study area, we established a ground truth dataset and employed a supervised learning workflow to detect floating macroplastics. By utilizing a combination of polarimetric information, we obtained detection accuracies exceeding 90% on the imbalanced test set without overfitting. Further investigation into the scattering behavior revealed a linear relationship between VV and VH backscatter values and the size of the plastic patches. Additional analysis of the co- and cross-polarization components revealed that the presence of floating macroplastics increases backscatter intensity by incorporating surface and volumetric scattering effects into the MW response. Specifically, the plastic patch response is characterized by a combination of specular and higher-order scattering, which enables detection amidst the "darker" water surface.

A follow-up work will focus on the requirements for applying ML algorithms to detect floating macroplastics in deep-sea environments, and the comparison with sensors that use MW wavelengths with higher resolutions.

ACKNOWLEDGMENT

This work was supported in part by Fundação para a Ciência e a Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) and Fundo Europeu de Desenvolvimento Regional (FEDER)–PT2020 Partnership Agreement under Project 10.54499/UIDB/50008/2020 and in part by ESA under Contract no. 2-1806/21/NL/GLC/ov. Lastly, this work was additionally funded by FCT through research scholarship UI/BD/151090/2021.

REFERENCES

- [1] Salgado-Hernanz, Paula M., et al. "Assessment of marine litter through remote sensing: recent approaches and future goals." *Marine Pollution Bulletin* 168 (2021): 112347.
- [2] Goddijn-Murphy, Lonneke, et al. "Emerging Technologies for Remote Sensing of Floating and Submerged Plastic Litter." *Remote Sensing* 16.10 (2024): 1770.
- [3] de Fockert, Anton, et al. "Assessing the detection of floating plastic litter with advanced remote sensing technologies in a hydrodynamic test facility." *Scientific Reports* 14.1 (2024): 25902.
- [4] Kikaki, Katerina, et al. "MARIDA: A benchmark for Marine Debris detection from Sentinel-2 remote sensing data." *PLoS one* 17.1 (2022): e0262247.
- [5] Topouzelis, Konstantinos, et al. "Plastic litter project 2019: Exploring the detection of floating plastic litter using drones and sentinel 2 satellite images." *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020.
- [6] Jia, Tianlong, et al. "Deep learning for detecting macroplastic litter in water bodies: A review." *Water Research* 231 (2023): 119632.
- [7] Basu, Bidroha, et al. "Development of novel classification algorithms for detection of floating plastic debris in coastal waterbodies using multispectral Sentinel-2 remote sensing imagery." *Remote Sensing* 13.8 (2021): 1598.
- [8] Danilov, Aleksandr, and Elizaveta Serdiukova. "Review of Methods for Automatic Plastic Detection in Water Areas Using Satellite Images and Machine Learning." *Sensors* 24.16 (2024): 5089.
- [9] Felfício, João M., et al. "Feasibility of Radar-based Detection of Floating Macroplastics at Microwave Frequencies." *IEEE Transactions on Antennas and Propagation* (2024).
- [10] Da Costa, Tomás Soares, et al. "Detection of low permittivity floating plastic sheets at microwave frequencies." *2023 17th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2023.
- [11] Da Costa, Tomás Soares, et al. "Feature Selection for Identifying Optimal Microwave Frequencies to Detect Floating Macroplastic Litter in C and X Bands." *2024 18th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2024.
- [12] Da Costa, Tomás Soares, et al. "Identifying Optimal Microwave Frequencies to Detect Floating Macroplastic Litter using Machine Learning." *International Journal of Microwave and Wireless Technologies* (2024).
- [13] Vala, Mário, et al. "Small-Scale Passive Millimetre-Wave Imaging Measurements for Marine Litter Detection at W-Band." *2024 18th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2024.
- [14] Vala, Mário, et al. "On the feasibility of using passive mm-wave imaging for marine litter detection at the w-band." *2023 17th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2023.
- [15] Serafino, Francesco, and Andrea Bianco. "X-Band Radar Detection of Small Garbage Islands in Different Sea State Conditions." *Remote Sensing* 16.12 (2024): 2101.
- [16] Davaasuren, Narangerel, et al. "Detecting microplastics pollution in world oceans using SAR remote sensing." *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018.
- [17] Stokral, Maryna, et al. "River export of macro- and microplastics to seas by sources worldwide." *Nature Communications* 14.1 (2023): 4842.
- [18] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32.
- [19] Truckenbrodt, John, et al. "pyroSAR—A Framework for Large-Scale SAR Satellite Data Processing." *Proceedings of the Big Data from Space, Munich, Germany* (2019): 19-20.
- [20] Fung, Tung, and Ellsworth LeDrew. "For change detection using various accuracy." *Photogramm. Eng. Remote Sens* 54.10 (1988): 1449-1454.