Detecting Marine Debris Using Sentinel-2 Satellite Images (Case Study: Kuta Beach, Bali)

F F Nasir[1], R Kurniawan[1]

[1] Politeknik Statistika STIS, Jl. Otto Iskandardinata No. 64C, Jakarta, Indonesia

*Corresponding author's email: 222112030@stis.ac.id, robertk@stis.ac.id

Abstract. Plastic waste pollution in the oceans remains a global problem. Kuta Beach is one of Bali's tourist destinations that has been affected by plastic waste pollution. This is not in line with the 14th SDGs, which is to prevent and reduce marine debris pollution. However, the marine debris monitoring process carried out by the Ministry of Environment and Forestry requires officers to conduct direct monitoring in the field, which incurs higher costs. Therefore, satellite imagery can be an alternative option for more effective and efficient marine debris detection. This study aims to detect marine debris on Kuta Beach using machine learning algorithms, namely Random Forest (RF), XGBoost, and LightGBM. This study uses the Marine Debris Archive (MARIDA) dataset, which has marine debris labels, and Sentinel-2 images of Kuta Beach from 2019–2023. The LightGBM algorithm provided the best performance in detecting marine debris with an F1-score of 95.16%. The area detected as marine debris on Kuta Beach in 2019–2023 was 500 m^2, 0 m^2, 100 m^2, 300 m^2, and 400 m^2, respectively. Based on these results, marine debris is generally detected around the coastline, particularly in the southern area of Kuta Beach, which is located near a shopping center.

Keyword: marine debris, satellite imagery, machine learning.

1. Introduction

One of the biggest causes of environmental degradation is plastic waste [1]. The use of single-use plastics has increased 20-fold during the past 50 years [2]. However, appropriate waste management has not kept pace with this situation [3]. Accordingly, between 19 and 23 million metric tons of plastic waste wind up damaging the environment, especially aquatic areas [3], [4]. One item that is hard to break down is plastic. It is estimated that plastic waste in the ocean will take 292 years to fully degrade [5].

Indonesia is one of the countries that contributes significantly to plastic waste pollution in the world [6], [7]. Due to urbanization, high population density, and a variety of coastal activities, urban coastal areas are particularly susceptible to marine debris pollution [8]. One place that is susceptible to this issue is Bali. Bali has become a destination that attracts many tourists. It is recorded that the number of tourists visiting Bali in 2024 will reach 6.31 million foreign tourist visits [9] and 10.12 million domestic tourist visits [10]. The possibility of environmental pollution is increased by the high level of tourism, particularly along Bali's shore [11], [12]. This is reinforced by the results of marine debris monitoring conducted by the Indonesian Ministry of Environment and Forestry in 2021, which showed that there were 12.82 grams of plastic waste per square meter on the beaches of Badung Regency [13]. Based on this data, plastic waste is the most dominant type of marine debris. In addition, the results of marine debris monitoring in 2023 show that Badung Regency, Bali, is the location with the highest abundance of microplastics in seawater, reaching 91.22 particles/m^3 [14]. Based on the results of this monitoring, marine debris on the beaches of Badung Regency, including Kuta Beach, remains a challenge.

Ocean contamination from marine debris can have detrimental effects on a number of factors. Ocean pollution from plastic debris can harm marine ecosystems and disturb the lifestyles of organisms [15], [16], [17]. The economy may be impacted as well. Plastic waste pollution can reduce fishermen's catches and decrease the tourist appeal of polluted beaches [18]. As a result, the United Nations (UN) has made this issue one of its global priorities under the Sustainable Development Goals (SDGs). One of the targets in goal 14 of the SDGs is to prevent and reduce various types of marine pollution, including marine debris.

Accurate detection and monitoring of marine debris in the ocean remains a challenge to this day [19]. The use of remote sensing with geospatial data using a machine learning approach has been developed to enable more efficient monitoring [19], [20]. The limited availability of actual datasets for detecting marine debris using satellite imagery led previous researchers to develop an open-access dataset called the Marine Debris Archive Dataset (MARIDA) [21]. MARIDA provides annotations on Sentinel-2 imagery at the pixel level, which can be used as a benchmark for marine debris detection. The spectral behavior captured by satellite images can be used to distinguish marine debris from other objects [22]. Previous study used a combination of several spectral composite indices, including the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Floating Algae Index (FAI), Floating Debris Index (FDI), Shadow Index (SI), Bare Soil Index (BSI), and NRD [21]. In this case, the FDI index is used to detect plastic waste floating in the ocean. In addition, there is a study that developed the Adjusted Plastic Index (API), which is capable of distinguishing plastic waste mixed with other land cover [20].

Although it has not been well investigated, the use of machine learning algorithms to identify marine debris in the ocean holds a lot of promise [19]. Thus, the purpose of this project is to use machine learning techniques using Sentinel-2 satellite image data and the publicly available MARIDA dataset to identify marine debris in the Kuta Beach area of Bali. To find out if there was marine debris on Kuta Beach, three classification algorithms were employed: Random Forest (RF), Extreme Gradient-Boosting Machine (XGBoost), and Light Gradient-Boosting Machine (LightGBM). These algorithms are used because they have been proven to perform well in detecting plastic waste using satellite imagery [20], [23], [24]. This study also looked at characteristics that are crucial for differentiating marine debris from other items. This study contributes by integrating various spectral indices derived from Sentinel-2 imagery with an machine learning model to identify marine debris in tropical coastal environments. Previous studies have mostly used the Floating Debris Index (FDI) with the Normalized Difference Vegetation Index (NDVI) to detect the presence of marine debris [19], [21], [22]. Therefore, this study attempts to add the Adjusted Plastic Index (API) feature, which is considered better at separating marine debris from vegetation and land cover [20].

2. Methodology

2.1. Study area

2.1.1. Case study area

The case study area for this research is Kuta Beach, located in Badung Regency, Bali. This research classifies marine debris detected in the Kuta Beach area from 2019 to 2023 at the pixel level. The image of Kuta Beach for each year has 65,268 pixels covering the area from the water to the coastline. The area covers the coastal region and waters of Kuta Beach, as shown in Figure 1.

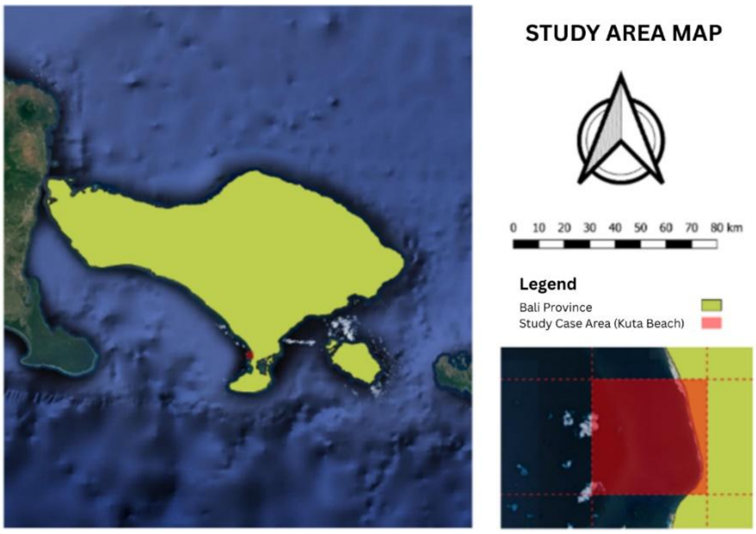


Figure 2. Map of Kuta Beach study area

2.1.2. Supporting study area

The training and testing sets were using publicly available data from the Marine Debris Archive (MARIDA). As indicated in Table 1, MARIDA offers 837,377 annotated pixels categorized into 15 classes. Due to their comparable spectral behavior, the Wakes, Cloud Shadows, Waves, and Mixed Water classes were combined into the Marine Water class in this study, aggregating the MARIDA dataset classes into 11 classes [21]. MARIDA was collected from 2015 to 2021 in eleven countries, namely Honduras, Guatemala, Haiti, Santo Domingo, Vietnam, South Africa, Scotland, Indonesia, the Philippines, South Korea, and China. Bali is one of the locations covered in the MARIDA dataset, so the characteristics of marine debris found on Kuta Beach can be represented. There is so much variance in the supporting regional data, which is dispersed over numerous countries, that machine learning models have more to learn during the data training process.

Table 1. MARIDA dataset thematic classes.

Class	Description	Number of pixels
Marine Debris	Floating polymers, such as plastics, and a mixture of man-made waste	3,399
Dense Sargassum	Dense floating Sargassum macroalgae	2,797
Sparse Sargassum	Sparse floating Sargassum macroalgae	2,357
Natural Organic Material	Vegetation and wood	864
Ship	Sailing and anchored vessels	5,803
Clouds	Clouds including thin clouds	117,400
Marine Water	Clear water	129,159
Sediment-Laden Water	Brown-colored discharges from high-sediment rivers	372,937

Foam	Foam captured near riverbanks or areas where waves break the coast	1,225
Turbid Water	Turbid waters close to coastal areas	157,612
Shallow Water	Coral reefs and submerged vegetation in coastal waters	17,369
Waves	Waves	5,827
Cloud Shadows	Cloud shadows	11,728
Wakes	Wakes and waves from a sailing vessel	8,490
Mixed Water	Water close to materials that float	410
	Total	837,377

2.2. Data sources

2.2.1. Case study area data

The Google Earth Engine (GEE) platform provided the Sentinel-2 satellite imaging data for the Kuta Beach region. Sentinel-2 was chosen because it is freely accessible and its spectral bands has been proven capable of detecting marine debris [22], [25]. To provide maximum detail in marine debris detection, the images used have the highest spatial resolution of Sentinel-2, which is 10 meters. After being prepared in shapefile format, the Kuta Beach region of interest (ROI) was loaded into the GEE code editor. To reduce the impact of clouds, a cloud masking function was applied to the imagery acquired using GEE. To further address the potential for noise in the data, the median approach was used to create annual aggregate statistics for the 2019–2023 timeframe.

2.2.2. Supporting study area data

The MARIDA dataset was obtained from the Zenodo online repository. MARIDA was collected over a period of seven years, from 2015 to 2021. MARIDA has developed a marine debris label obtained by acquiring Sentinel-2 imagery with ground-truth observations and literature studies. Experts in image interpretation used information from waste reports, Sentinel-2 satellite photos, very high-resolution Planet satellite images, and the spectral behavior of ocean objects to annotate the MARIDA dataset.

2.3. Modelling

2.3.1. Model development

This study uses three algorithms with the ensemble method, namely RF [26], XGBoost [27], and LightGBM [28]. These algorithms are capable of improving model performance by combining the prediction results from several estimators [29]. The three algorithms were chosen because they have been proven to perform well in marine debris detection cases.

In combining data, RF uses the bagging algorithm and then uses decision trees to train each combined group [30]. Each tree is trained on a smaller, randomly selected subset of data so that random forests are able to handle large, high-dimensional data. Based on previous research, RF has been proven to perform well in various cases, including in the case of marine debris detection [19], [20], [21], [31], [32], [33].

XGBoost performs gradient boosting incrementally by improving the error generated in the previously built model. Thus, XGBoost has good scalability in all scenarios [27]. XGBoost was also used on synthetic plastic debris data with Generative Adversarial Networks (GANs), yielding results that showed the model was able to distinguish plastic waste from other features [23].

LightGBM uses gradient-based one-sided sampling and exclusive feature bundling, which speeds up the data training process [28]. This makes LightGBM more efficient in terms of memory, cost, and data training time [34]. In the case of plastic waste detection, LightGBM has proven to be efficient and provides high accuracy [24].

Model development using RF and LightGBM was built using class weighting and without class weighting. Class weighting was used to handle imbalanced data, which gives weights that are inversely proportional to class frequency. This weighting ensures that prediction errors in minority classes are calculated as greater in the loss calculation, thereby helping the model to not ignore classes that rarely appear. Meanwhile, the XGBoost model has been proven to perform well in cases of imbalanced data [35]. The XGBoost model does not have a class_weight parameter, so the sample_weight parameter is used to weight each sample based on the confidence level in the labeling.

In maximizing the performance of each model, this study conducted parameter optimization using Bayesian optimization (BO). In BO, evaluation points are selected using an acquisition function to maximize the model [36]. In this way, BO ensures that optimization does not only concentrate on areas with good results but also explores areas that have not been studied much in order to find optimal solutions. Selection of the best model based on the evaluation metrics of each model. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics using macro-averaging. Macro-averaging is used in this study so that model performance could be assessed fairly without being influenced by classes with a larger number of samples [37]. The best model obtained was then applied for implementation in the Kuta Beach study area. For clarity, the research workflow is shown in Figure 2.

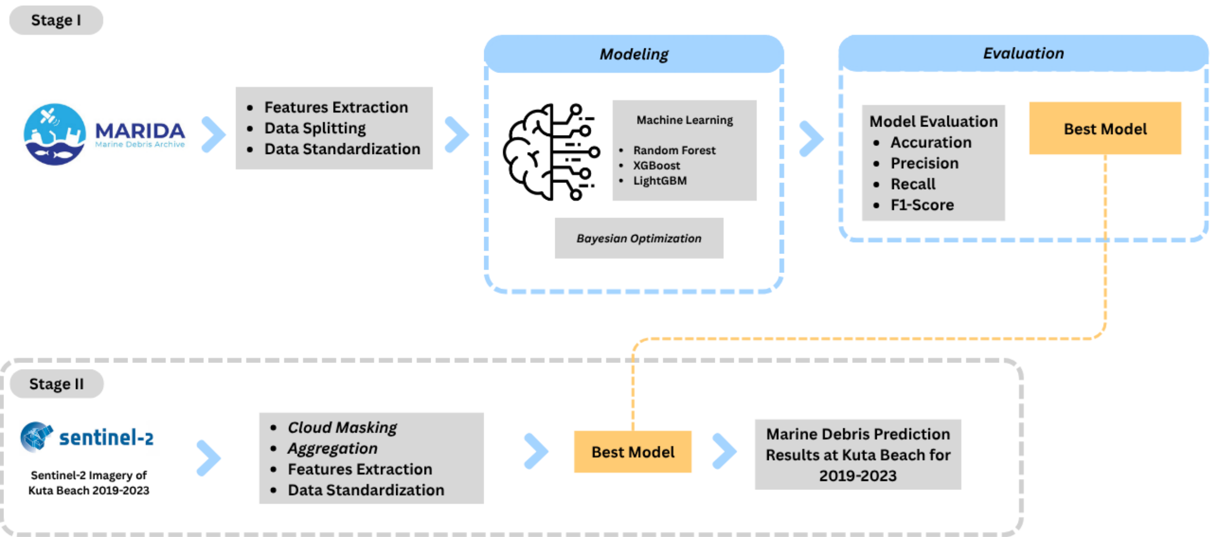


Figure 3. Research workflow diagram

2.3.2. Features

In detecting marine debris, this study uses the Floating Debris Index (FDI) [22] and the Adjusted Plastic Index (API) [20]. In addition, other indices are used to distinguish between marine debris and other objects in the ocean. The Normalized Difference Vegetation Index (NDVI), Floating Algae Index (FAI), and Near Infrared-Red Difference (NRD) are used to detect vegetation and macroalgae. The Normalized Difference Water Index (NDWI) and Normalized Difference Moisture Index (NDMI) are used to identify water features. The Shadow Index (SI) and Bare Soil Index (BSI) are used to separate bright objects on the earth's surface. In addition, the values of the Sentinel-2 spectral bands are also used to maximize the model in separating existing classes, including bands 1 to 10 [21].

3. Result

3.1. Model development results

After finding the optimal parameter values, each model was trained and then compared based on accuracy, precision, recall, and F1-score values. Table 2 shows the model performance on the training set and testing set without additional handling on RF and LightGBM for imbalance cases. In both the training data and testing data, the RF algorithm provided the best evaluation values compared to other algorithms.

Table 2. Model performance on training and testing data without additional handling for data imbalance.

Data	Model	Evaluation Metrics (%)			
		Accuracy	Precision	*Recall*	F1-score
Training	**Random Forest**	**99.81**	**99.34**	**98.24**	**98.78**
	XGBoost	98.76	95.28	92.10	93.58
	LightGBM	90.95	78.69	71.83	73.80
Testing	**Random Forest**	**99.01**	**95.07**	**92.46**	**93.68**
	XGBoost	98.47	93.21	89.98	91.49
	LightGBM	90.86	76.45	70.74	72.40

Table 3 shows the model performance after additional handling of imbalanced data using class weighting on RF and LightGBM. The results show that LightGBM provides the best performance. This is because LightGBM has better scalability, making it suitable for large datasets.

Table 3. Model performance on training and testing data without additional handling for data imbalance.

Data	Model	Evaluation Metrics (%)			
		Accuracy	Precision	*Recall*	F1-score
Training	Random Forest	99.74	97.65	99.68	98.61
	XGBoost	98.76	95.28	92.10	93.58
	LightGBM	**99.77**	**99.58**	**99.87**	**99.72**
Testing	Random Forest	98.92	92.67	93.89	93.19
	XGBoost	98.47	93.21	89.98	91.49
	LightGBM	**99.18**	**94.82**	**95.53**	**95.16**

In both training and testing data, LightGBM, the model with the greatest performance, displayed high values. To detect possible overfitting in the model, a check was performed using a learning curve. To provide more reliable model generalization, stratified 5-fold cross-validation was used to generate the learning curve. The accuracy learning curve against training size from LightGBM is displayed in

Figure 3. According to the chart, the training score's accuracy is extremely high and keeps rising as the volume of training data increases. Meanwhile, the accuracy of the cross-validation score is above 0.95. This accuracy value is slightly lower than the training score. However, the gap in accuracy between the training score and the cross-validation score is not significant. This indicates that the LightGBM model is still capable of making good predictions on new data. Therefore, the LightGBM learning curve plot does not show any severe overfitting and demonstrates the model's ability to generalize on new data.

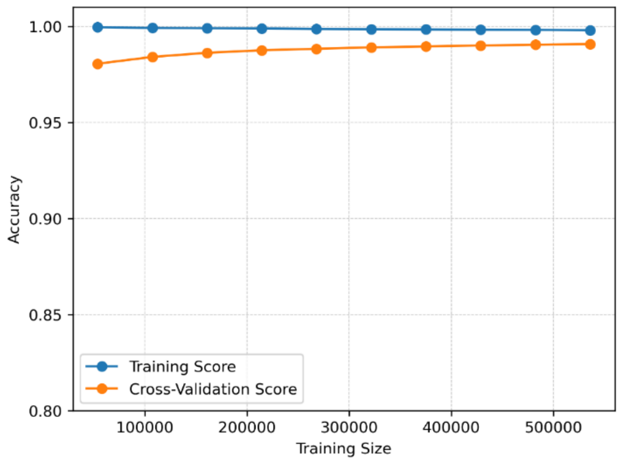


Figure 4. Learning curve of LightGBM

The model was also assessed using each current class in order to ascertain how well it predicted each class. The LightGBM model's performance for each class in the training and testing data is displayed in Tables 4 and 5, respectively. These two tables demonstrate how effectively the LightGBM model predicts each current class in both training and testing data. With an F1-score of 92.62%, the LightGBM model successfully predicted the marine debris class in the testing set.

Table 4. LightGBM performance by class on training data.

Class	Evaluation Metrics (%)		
	Precision	Recall	F1-score
Clouds	99,22	99,23	99,23
Dense Sargassum	100	100	100
Foam	99,49	100	99,75
Marine Debris	**99,02**	**100**	**99,51**

Marine Water	99,43	99,31	99,37
Natural Organic Material	100	100	100
Sediment-Laden Water	100	100	100
Shallow Water	99,91	100	99,95
Ship	98,49	100	99,24
Sparse Sargassum	99,84	100	99,92
Turbid Water	100	100	100

Table 5. LightGBM performance by class on testing data.

Class	Evaluation Metrics (%)		
	Precision	Recall	F1-score
Clouds	98,08	98,05	98,06
Dense Sargassum	96,63	97,32	96,97
Foam	85,34	92,65	88,85
Marine Debris	**91,17**	**94,12**	**92,62**
Marine Water	98,31	98,12	98,21
Natural Organic Material	92,07	87,28	89,61
Sediment-Laden Water	100	99,99	100
Shallow Water	98,14	98,85	98,49
Ship	90,41	90,18	90,30
Sparse Sargassum	93,10	94,48	93,78
Turbid Water	99,82	99,83	99,82

LightGBM model classification errors can be seen based on the confusion matrix. Figure 4 shows that the LightGBM model is best at distinguishing the Sediment-Laden Water class from other classes. This is in line with the high evaluation score for the Sediment-Laden Water class. For the Marine Debris class, the model was able to correctly predict 640 pixels. The model still made classification errors in several classes. For the Marine Debris class, there were 62 false positives, most of which originated from the Ship and Natural Organic Material classes. In addition, there were 40 false negatives, some of which originated from the Ship class. A total of 15 pixels with the Ship class were predicted as Marine Debris, and 13 pixels with the Marine Debris class were predicted as Ship. This is because there is a similarity in spectral behavior between the Marine Debris class and the Ship class.

Figure 5. Confusion matrix of LightGBM

In machine learning modeling, there are variables that play a more dominant role in the model that is formed. The feature importance value shows how much a variable contributes to distinguishing the 11 classes predicted by the model. Based on Figure 5, NDWI is the most influential variable in distinguishing the existing classes with a gain importance value of 2489121.7. The gain importance value shows how much NDWI contributes to reducing uncertainty (impurity) in each split of the decision tree built by the LightGBM model. The higher the gain value, the more frequently and effectively the variable is used by the LightGBM model to separate data into different classes. This means that NDWI is the feature that provides very significant information for the model in distinguishing marine debris classes from other objects such as seawater and surrounding coastal objects.

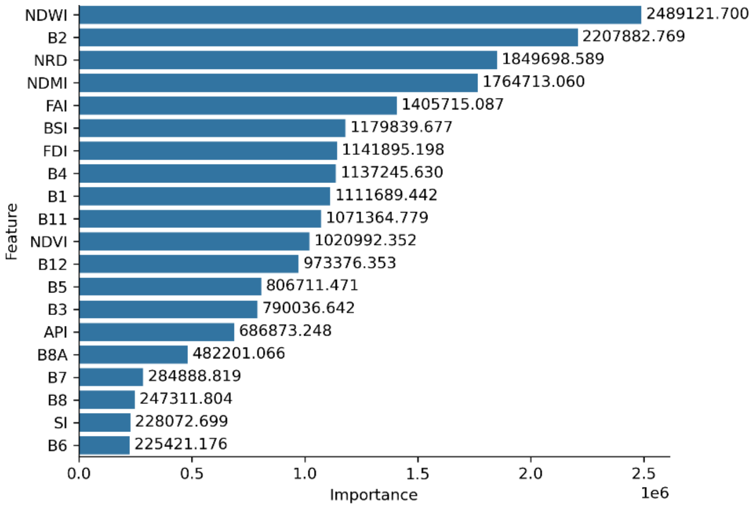
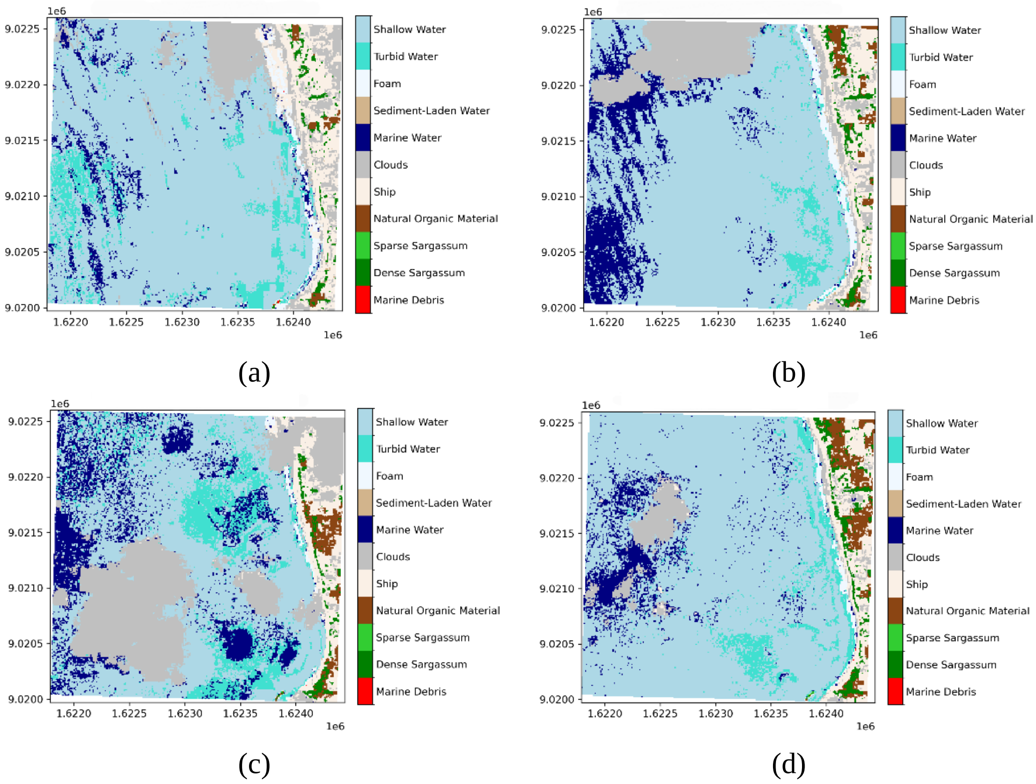
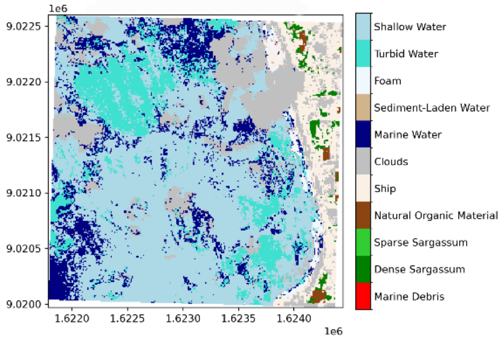


Figure 6. Feature importance value of each feature in LightGBM

3.2. *Implementation of machine learning model in the Kuta Beach area*

Based on the best model, LightGBM, predictions were made for plastic waste in the Kuta Beach area for 2019–2023. A visualization of the prediction results for Kuta Beach is shown in Figure 6.





(e)

Figure 7. Visualization of predicted results for Kuta Beach in 2019 (a), 2020 (b), 2021 (c), 2022 (d), and 2023 (e)

Based on Figure 6, it can be seen that the prediction results for each year are dominated by natural objects. In 2019–2023, the classes most frequently detected by the model were water classes, namely Shallow Water, Marine Water, and Turbid Water. Although the prediction results were dominated by natural features, the model was able to detect the presence of plastic waste in the Kuta Beach area. The number of pixels predicted for Kuta Beach in each class is shown in Table 6.

Table 6. Number of pixels predicted based on class and year.

Class	Year				
	2019	2020	2021	2022	2023
Clouds	6704	8865	15717	3480	10594
Dense Sargassum	658	1060	692	1292	641
Foam	1155	1345	381	560	776
Marine Debris	**5**	**0**	**1**	**3**	**4**
Marine Water	3084	7094	9256	4519	8171
Natural Organic Material	328	866	1134	1890	341
Sediment-Laden Water	49	92	9	34	56
Shallow Water	43952	39982	28134	46626	31401
Ship	5529	4734	3637	4156	5432
Sparse Sargassum	1	0	1	11	0
Turbid Water	4599	2026	7102	3493	8648

Figure 7 shows a bar chart of the number of pixels detected as marine debris on Kuta Beach in 2019–2023. There were 5 pixels detected as marine debris on Kuta Beach in 2019. This means that in 2019, there were approximately 500 m^2 of Kuta Beach area polluted by plastic waste. In 2020, the model did not detect the presence of pixels predicted to be marine debris on Kuta Beach. This is in line with the policy related to the implementation of Large-Scale Social Restrictions (PSBB) during the COVID-19 pandemic. The PSBB policy reduced tourism activity on Kuta Beach, which may have led to a decrease in the amount of waste originating from tourism activities. In 2021, the number of marine debris pixels detected on Kuta Beach increased again to 1 pixel or 100 m^2. The increase in the number of marine debris pixels detected from 2020 to 2021 may be due to the recovery of community activities in the Kuta Beach area. In 2022 and 2023, there was a consecutive increase in the number of marine debris pixels detected, namely 3 pixels (300 m^2) and 4 pixels (400 m^2), respectively.

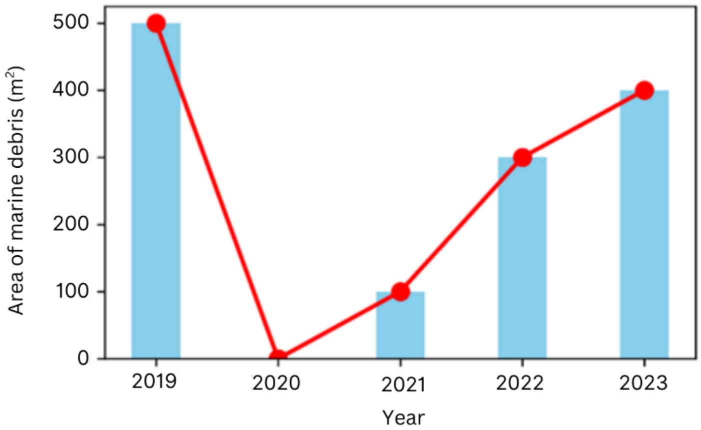


Figure 8. Area of marine debris detected on Kuta Beach from 2019 to 2023

4. Discussion

Based on the detection results in the case study area, the area of marine debris on Kuta Beach in 2019–2023 ranged from 100 m^2 to 500 m^2. These findings are consistent with the results of previous study which showed that plastic waste accounted for 26.6% of the total composition of waste polluting Kuta Beach [38]. This finding is also in line with previous research with a broader context, in which Indonesia was identified as one of the main contributors of plastic waste to the oceans globally [39]. This provides evidence that the problem of plastic waste on Kuta Beach is a consistent problem and a long-term issue that has not been fully resolved. However, because Sentinel-2 has a spatial resolution of 10 m, debris spots smaller than a full pixel can still be detected through subpixel response, where the debris partially covers a pixel but can still be distinguished spectrally from the surrounding water. Similar to previous studies, marine debris pixels often contain a mixture of water and debris material, indicating that detection can occur at the subpixel level. In other words, debris occupying approximately 30–55% of the pixel area can still be identified through its characteristic spectral response [22].

This study shows that in 2020, no marine debris was detected on Kuta Beach. These results are thought to be a direct impact of the COVID-19 pandemic and the accompanying global restrictions on activities. The phenomenon of clean coastal areas due to lockdowns did not only occur on Kuta Beach. Research by Okuku et al. (2021) found a similar effect, showing that limited community movement during the COVID-19 pandemic led to a significant decrease in the amount of marine debris on beaches in Kenya [40]. This is also in line with research by Sari et al. (2022), which shows that the amount of waste at tourist sites is directly proportional to the number of tourist visits [41].

This study demonstrates the potential of Sentinel-2 imagery and the use of the open-access MARIDA dataset in identifying marine debris along coastal areas such as Kuta Beach. This approach offers a cost-effective and easily accessible alternative to very high-resolution data, while maintaining reasonable accuracy in classifying various coastal features. This study attempted to use the API [20] spectral index. However, the API did not show a particularly high feature importance value.

In this study, Sentinel-2 satellite imagery with a spatial resolution of 10 m was used. Therefore, the detection results were less capable of detecting plastic waste that was very small in size. Further research related to marine debris detection is recommended to add satellite image data sources in model training using data that has more knowledge in detecting marine debris. In addition, to ensure that marine debris detection is carried out more optimally, further research can use images with much higher spatial resolution (e.g., WorldView-3 or PlanetScope) or UAV data.

5. Conclusion

Based on the results of the research conducted, the area detected as marine debris on Kuta Beach in 2019–2023 was 500 m^2, 0 m^2, 100 m^2, 300 m^2, and 400 m^2, respectively. Based on these results, marine debris is generally detected around the coastline, particularly in the southern area of Kuta Beach, which is located near a shopping center. The use of the open-access MARIDA dataset based on Sentinel-2 shows good performance that can be applied in various locations, including the Kuta Beach area. The results of marine debris detection at Kuta Beach from 2019 to 2023 can be a more efficient alternative for identifying locations on the beach that are more likely to be polluted by marine debris. Thus, this study can support relevant parties such as the Ministry of Environment and Forestry, environmental agencies, tourism area managers, and environmental communities in designing more targeted and efficient marine debris mitigation strategies in coastal areas.

References

[1] S. L. Wright and F. J. Kelly, “Plastic and Human Health: A Micro Issue?,” *Environ. Sci. Technol.*, vol. 51, no. 12, pp. 6634–6647, 2017, doi: 10.1021/acs.est.7b00423.

[2] T. R. Walker and L. Fequet, “Current trends of unsustainable plastic production and micro(nano) plastic pollution,” *Trends Anal. Chem.*, vol. 160, 2023, doi: 10.1016/j.trac.2023.116984.

[3] M. G. Kibria, N. I. Masuk, R. Safayet, H. Q. Nguyen, and M. Mourshed, “Plastic Waste: Challenges and Opportunities to Mitigate Pollution and Effective Management,” *Int. J. Environ. Res.*, vol. 17, no. 1, Feb. 2023, doi: 10.1007/s41742-023-00507-z.

[4] C. J. Rhodes, “Plastic pollution and potential solutions,” *Sci. Prog.*, vol. 101, no. 3, pp. 207–260, 2018, doi: 10.3184/003685018X15294876706211.

[5] X. Zhang and X. Peng, “How long for plastics to decompose in the deep sea?,” *Geochemical Perspect. Lett.*, vol. 22, pp. 20–25, 2022, doi: 10.7185/geochemlet.2222.

[6] L. C. M. Lebreton, J. Van Der Zwet, J. W. Damsteeg, B. Slat, A. Andrady, and J. Reisser, “River plastic emissions to the world’s oceans,” *Nat. Commun.*, vol. 8, pp. 1–10, 2017, doi: 10.1038/ncomms15611.

[7] L. J. J. Meijer, T. van Emmerik, R. van der Ent, C. Schmidt, and L. Lebreton, “More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean,” *Sci. Adv.*, vol. 7, no. 18, pp. 1–13, 2021, doi: 10.1126/sciadv.aaz5803.

[8] N. Yoshioka, M. Era, and D. Sasaki, “Towards integration of climate disaster risk and waste management: A case study of urban and rural coastal communities in the Philippines,” *Sustain.*, vol. 13, no. 4, pp. 1–16, 2021, doi: 10.3390/su13041624.

[9] BPS, *International Visitor Arrivals Statistics 2024*. 2025.

[10] BPS, “Number of Domestic Visitor to Bali by Month, 2004-2024.” [Online]. Available: https://bali.bps.go.id/id/statistics-table/1/MjkjMQ==/kunjungan-wisatawan-domestik-ke-bali-per-bulan-2004-2018.html

[11] K. L. Chong, “The side effects of mass tourism: the voices of Bali islanders,” *Asia Pacific J. Tour. Res.*, vol. 25, no. 2, pp. 157–169, 2020, doi: 10.1080/10941665.2019.1683591.

[12] I. M. W. Widyarsana, M. A. Priyanka, and L. A. Devianto, “Solid Waste Analysis and Processing Potential in the Tourism Sector: Case Study in Nusa Dua, South Kuta, Bali,” *Indones. J. Urban Environ. Technol.*, vol. 5, no. 2, pp. 181–192, 2022, doi: 10.25105/urbanenvirotech.v5i2.13538.

[13] KLHK, "Performance Report 2021," 2022.
[14] Kementerian Lingkungan Hidup Dan Kehutanan, "Performance Report 2024," 2024.
[15] D. Brennecke, K. Knickmeier, I. Pawliczka, U. Siebert, and M. Wahlberg, *Marine Mammals A Deep Dive into the World of Science*. Springer, 2023. doi: 10.1007/978-3-031-06836-2.
[16] S. C. Gall and R. C. Thompson, "The impact of debris on marine life," *Mar. Pollut. Bull.*, vol. 92, no. 1–2, pp. 170–179, 2015, doi: 10.1016/j.marpolbul.2014.12.041.
[17] A. Isdianto *et al.*, "The Occurrence of Marine Debris and Its Impacts on Coral Reefs in the Sempu Island Nature Reserve, Malang, Indonesia," *J. Ecol. Eng.*, vol. 25, no. 9, pp. 70–80, 2024, doi: 10.12911/22998993/190514.
[18] S. Arabi and A. Nahman, "Impacts of marine plastic on ecosystem services and economy: State of South African research," *S. Afr. J. Sci.*, vol. 116, no. 5–6, pp. 1–7, 2020, doi: 10.17159/sajs.2020/7695.
[19] S. Sannigrahi, B. Basu, A. S. Basu, and F. Pilla, "Development of automated marine floating plastic detection system using Sentinel-2 imagery and machine learning models," *Mar. Pollut. Bull.*, vol. 178, May 2022, doi: 10.1016/j.marpolbul.2022.113527.
[20] A. D. Sakti *et al.*, "Identification of illegally dumped plastic waste in a highly polluted river in Indonesia using Sentinel-2 satellite imagery," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-32087-5.
[21] 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 January, Jan. 2022, doi: 10.1371/journal.pone.0262247.
[22] 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, 2020, doi: 10.1038/s41598-020-62298-z.
[23] M. M. Duarte and L. Azevedo, "Automatic Detection and Identification of Floating Marine Debris Using Multispectral Satellite Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, doi: 10.1109/TGRS.2023.3283607.
[24] N. Taggio *et al.*, "A Combination of Machine Learning Algorithms for Marine Plastic Litter Detection Exploiting Hyperspectral PRISMA Data," *Remote Sens.*, vol. 14, no. 15, Aug. 2022, doi: 10.3390/rs14153606.
[25] 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 Obs. Geoinf.*, vol. 79, no. January, pp. 175–183, 2019, doi: 10.1016/j.jag.2019.03.011.
[26] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: doi.org/10.1023/A:1010933404324.
[27] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[28] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Neural Information Processing Systems*, 2017. [Online]. Available: https://github.com/Microsoft/LightGBM.
[29] G. Kunapuli, *Ensemble Methods for Machine Learning*. Manning Publication, 2022.
[30] Q. Xu and J. Yin, "Application of Random Forest Algorithm in Physical Education," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/1996904.
[31] T. Acuña-Ruz *et al.*, "Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications," *Remote Sens. Environ.*, vol. 217, pp. 309–322, Nov. 2018, doi: 10.1016/j.rse.2018.08.008.
[32] Á. Pérez-García, T. H. M. van Emmerik, A. Mata, P. F. Tasseron, and J. F. López, "Efficient plastic detection in coastal areas with selected spectral bands," *Mar. Pollut. Bull.*, vol. 207, Oct. 2024, doi: 10.1016/j.marpolbul.2024.116914.
[33] K. Sasaki, T. Sekine, L. J. Burtz, and W. J. Emery, "Coastal Marine Debris Detection and Density Mapping With Very High Resolution Satellite Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 6391–6401, 2022, doi: 10.1109/JSTARS.2022.3193993.
[34] B. Quinto, *Next-generation machine learning with spark: Covers XGBoost, LightGBM, Spark NLP, distributed deep learning with keras, and more*. Apress Media LLC, 2020. doi: 10.1007/978-1-4842-5669-5.
[35] S. B. S. Lai, N. H. N. B. M. Shahri, M. B. Mohamad, H. A. B. A. Rahman, and A. Bin Rambli, "Comparing the performance of adaboost, xgboost, and logistic regression for imbalanced data," *Math. Stat.*, vol. 9, no. 3, pp. 379–385, 2021, doi: 10.13189/ms.2021.090320.
[36] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization," Jan. 01, 2016, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/JPROC.2015.2494218.
[37] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 820–836, 2024, doi: 10.1162/tacl_a_00675.
[38] A. Attamimi, N. P. Purba, S. R. Anggraini, S. A. Harahap, and S. Husrin, "Investigation of Marine Debris in Kuta Beach, Bali Semeidi Husrin Ministry of Marine Affairs and Fisheries," in *Proceedings of Environmental Engineering and Water Technology, Integrated Water System and Governance, Water Science and Engineering*, 2015. [Online]. Available: https://www.researchgate.net/publication/336486249

[39] J. R. Jambeck *et al.*, “Plastic waste inputs from land into the ocean,” *Science (80-.).*, vol. 347, no. 6223, pp. 768–771, Feb. 2015, doi: 10.1126/science.1260352.

[40] E. Okuku *et al.*, “The impacts of COVID-19 pandemic on marine litter pollution along the Kenyan Coast: A synthesis after 100 days following the first reported case in Kenya,” *Mar. Pollut. Bull.*, vol. 162, no. September, p. 111840, 2021, doi: 10.1016/j.marpolbul.2020.111840.

[41] M. M. Sari, I. W. K. Suryawan, B. S. Ramadan, I. Y. Septiariva, and S. Notodarmojo, “Marine Debris Management in the Parangtritis Beach Tourism Area, Yogyakarta during Covid-19 Pandemic,” *Nat. Environ. Pollut. Technol.*, vol. 21, no. 3, pp. 1183–1190, 2022, doi: 10.46488/NEPT.2022.v21i03.023.