RESEARCH ARTICLE • OPEN ACCESS

Coastal Ecosystem Classification Using Satellite-Based Machine Learning Approaches

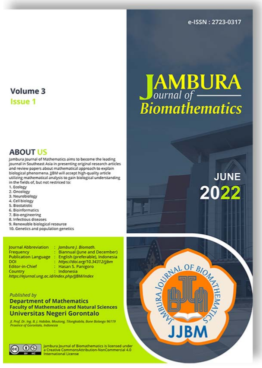
Giani Jovita Jane *et al.*

Volume 6, Issue 2, Pages 142–153, June 2025

Received 3 February 2025, Revised 18 March 2025, Accepted 20 June 2025, Published Online 30 June 2025

To Cite this Article : G. J. Jane *et al.*,“Coastal Ecosystem Classification Using Satellite-Based Machine Learning Approaches”, *Jambura J. Biomath*, vol. 6, no. 2, pp. 142–153, 2025, *https://doi.org/10.37905/jjbm.v6i2.30466*

JOURNAL INFO • JAMBURA JOURNAL OF BIOMATHEMATICS



Homepage	:	http://ejurnal.ung.ac.id/index.php/JJBM/index
Journal Abbreviation	:	Jambura J. Biomath.
Frequency	:	Quarterly (March, June, September and December)
Publication Language	:	English
DOI	:	https://doi.org/10.37905/jjbm
Online ISSN	:	2723-0317
Editor-in-Chief	:	Hasan S. Panigoro
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjbm/oai
Google Scholar ID	:	XzYgeKQAAAAJ
Email	:	editorial.jjbm@ung.ac.id

Jambura Journal of Biomathematics, Volume 6, Issue 2, Pages 142–153, June 2025 https://doi.org/10.37905/jjbm.v6i2.30466

Research Article

Check for updates

Coastal Ecosystem Classification Using Satellite-Based Machine Learning Approaches

Giani Jovita Jane[1], La Ode Alifatri[2], Etjih Tasriah[3], and Setia Pramana[1,*]

[1]*Politeknik Statistika STIS, Jl. Otista 64 C, Jakarta, Indonesia*
[2]*The National Research and Innovation Agency, Jl Pasir Putih Raya No.1, Jakarta, Indonesia*
[3]*BPS Statistics Indonesia, Jl Dr. Sutomo 6-8, Jakarta, Indonesia*

ARTICLE HISTORY
Received 3 February 2025
Revised 18 March 2025
Accepted 20 June 2025
Published 30 June 2025

KEYWORDS
Satellite imagery
Machine learning
Coastal area
Blue economy

ABSTRACT. *As an archipelagic country rich in natural resources, Indonesia has great marine economic potential. In order to maintain this economic potential in the long term, blue economy is needed as a concept in establishing development programs and public policies. One way to implement the concept is to arrange the ocean account, which framework implements the concept of a blue economy in the form of an environmental account. Ocean account can be considered to support the establishment of national policies and programmes of a country. Hence, accurate spatial data reflecting the current condition is essential for arranging this account. However, collecting such data can be costly and resource-intensive, making it challenging to ensure the availability of up-to-date and accurate information. In this context, alternative data sources could provide a viable solution. Previous research has successfully proven that machine learning modelling also Sentinel-1 and Sentinel-2 satellite imageries are capable for mapping coastal area, such as tidal and benthic areas. Therefore, this study attempts to classify coastal ecosystem of Karimunjawa National Park by utilizing Sentinel-1 and Sentinel-2 imagery and comparing the classification results of three machine learning methods, namely Random Forest (RF), Support Vector Classification (SVC), and Extreme Gradient Boosting (XGBoost), and analyse the ecosystem changes between 2020 and 2023. The result shows that RF gives the best result in performing classification for benthic areas achieving 0.77 and 0.78 in F1-score and Matthew's Correlation Coefficient (MCC), whereas SVC model succeed to achieve 0.83 in F1-score and MCC gives the best result for tidal areas. Furthermore, the coral reef and the seagrass area decreased by 6.524 km^2 and 1.39 km^2, respectively. Whereas, mangrove, built-up, and forest area show slight changes.*

1. Introduction

Blue economy comprises a range of economic sectors and related policies that together determine whether the use of ocean resources is sustainable [1]. Blue economy is seen as a strategy in managing marine resources. The concept has been introduced in Indonesia's governance since 2014. According to Law No. 32 of 2014 Section 1, Blue economy is an approach to improve sustainable marine management and conservation of marine and coastal resources and ecosystems in order to realise economic growth with principles such as community involvement, resource efficiency, waste minimisation, and multiple revenues. To implement this principle, Ministry of National Development Planning Agency (BAPPENAS) in co-operation with the Organisation for Economic Cooperation and Development (OECD) formulated the Blue Economy Development Framework for Indonesia's Economic Transformation [2]. As a member of Association of Southeast Asian Nations (ASEAN), one of the initiatives of the blue economy framework in ASEAN is the management of sustainable development and restoration of ecosystems such as forests, grasslands, mangroves and wetlands that protect, filter, store and regulate water, carbon and biodiversity [3].

One of the implementations of blue economy strategy is the development of ocean accounts. An ocean account is a system that allows us to measure not only economic achievements, but also environmental quality at the same time [4]. Basically, the preparation of environmental accounts follows the System of Environmental Economic Accounting (SEEA) framework agreed by the United Nations. The Indonesian government then implemented the SEEA guidelines in the Integrated System of Environmental and Economic Accounts (Sisnerling). Supporting data is needed in order to carry out the arrangement of the account, such as spatial data that depict the actual state of ecosystem. OECD encourages Indonesian government to collect more current and precise data on biodiversity to conserve marine ecosystem. However, collecting data requires substantial funding and adequate resources. This makes the availability of up-to-date and precise data cannot be guaranteed. Therefore, the existence of alternative data can be the answer to these problems.

The rapid development of technology allows the utilisation of big data as an alternative in field data collection. Several example of the big data potentials that have been implemented were recording the number of international visitor using MPD (Mobile Positioning Data), finding commuting pattern/flow using Twitter

*Corresponding Author.

Email : setia.pramana@stis.ac.id (S. Pramana)
Homepage : http://ejurnal.ung.ac.id/index.php/JJBM/index / E-ISSN : 2723-0317

Table 1. Details of the Variables Used

Variable Name	Data Source	Period	Classification Scheme
B2, B3, B4, and B8 band values	Sentinel-2 Level 2A SR	2020, 2021, 2023	Benthic Habitat Classification
Benthic Habitat Class Shapefile	National Research and Innovation Agency (BRIN)	2021	Benthic Habitat Classification
B2, B3, B4, B8, B5, B8A, B11, IBI, SAVI, MNDWI, and NBAI values	Sentinel-2 Level 2A SR	2019, 2020, 2023	Land Cover Classification
VV and VH band values	Sentinel-1 Level GRD	2019, 2020, 2023	Land Cover Classification
Land Cover Class Shapefile	Ministry of Environment and Forestry (KLHK)	2019-2020	Land Cover Classification

Statistics, and nowcasting food prices using crowdsourcing [5]. Another big data well-known for its ability to support field data collection is satellite image spatial data. The scope of satellite image information itself is quite broad, such as surface temperature, climate, atmosphere, weather, and earth observation. Several studies using satellite image spatial data have successfully classified coastal ecosystem areas. These studies classify coastal ecosystems that include tidal sites, such as mangrove forests, and benthic sites, such as coral reefs and seagrasses [6, 7]. Benthic itself means everything that is on the surface of the bottom of the water, both living and non-living things [8]. In particular, benthic studies using Sentinel-2 satellite imagery have begun to develop and are proving to provide good classification results [9–11]. Similar to benthic classification, the topic of mangrove classification has already produced a lot of research with good results, for example by combining Sentinel-2 and Sentinel-1 data [12–14].

In this classification, machine learning methods are commonly used. Misiuk and Brown state that supervised learning models, one type of learning technique in machine learning, have been widely used and proven to provide decent results for benthic classification [15]. Similar to benthic classification, the use of supervised learning methods has also been applied for a long time in studies that address mangrove coverage classification [16]. The previous review reference stated that Random Forest and Support Vector Classification algorithms have been widely used in coastal ecosystem since these methods perform well [6, 15]. Recently, a study by Nemani et al. [17] proved that Extreme Gradient Boosting algorithm's performance was able to surpass the performance of the previous two methods when classifying benthic assemblages. Thus, the authors intend to use all three algorithms and find the best algorithm that can classify benthic habitats. In the future, the result of coastal ecosystem classification would be useful in building ocean asset account since it measures the marine ecosystem as a wealth asset, specifically for ocean physical asset account.

This study is applied to a few islands in Karimunjawa National Park which has a rich coastal ecosystem consisting of coral reefs, seagrass, mangrove forests, and others, specifically to Karimunjawa island, Kemujan island, Menjangan Kecil island, and Menjangan Besar island [18]. The favourable geographical conditions make this region a much-visited coastal tourism area. This affects the available job choice that dominated in tourism also fishing and farming activities showing that coastal ecosystems have a significant impact on the economy of the Karimunjawa archipelago. In addition, the coastal area is protected included as a protected area by Ministry of Environment and Forestry (KLHK). Therefore, the coastal ecosystem classification is crucial in measuring the ecosystem assets not only for economic purposes but also for governance purposes. Previous coastal ecosystem study in Karimunjawa National Park results in a 94.17% overall accuracy using WorldView-2, a high resolution satellite, and 73.14% overall accuracy using Sentinel-2 [19, 20]. Both studies classify four kinds of benthic habitat categories. According to Varoquaux and Colliot, accuracy as a wide-used metrics would not suitable for imbalance classes [21]. However, there has been no research that covers the entire coastal area using Sentinel-2 satellite imagery and incorporating XGBoost as one of the candidate methods also incorporating other evaluation metrics that are more suitable for multiclassification classes. With the above concept in mind, this study covers coastal ecosystem classification and the changes in coastal ecosystem using satellite-based supervised learning modelling method in Karimunjawa National Park and covers tidal and coastal areas. In the future, the result would be useful to accommodate the arrangement of physical asset account.

2. Methods and Materials

2.1. *Scope of the research*

The research focuses on Karimunjawa National Park with reference year 2019-2021 and 2023. Geographically, Karimunjawa National Park located at the coordinates 5°40'39"- 5°55'00" South Latitude dan 110°05' 57"-110°31' 15" East Logitude [18]. Administratively, Karimunjawa National Park is located in Karimunjawa Islands, Jepara Regency, Central Java Province. Karimunjawa National Park is the only aquatic nature conservation consists of 27 small islands and several large islands [22]. The archipelago is known for its integrated marine tourism that excels regionally and internationally [23]. This conservation area has five ecosystem types, namely lowland tropical rainforest ecosystems, coastal forests, mangrove forests, seagrass and seaweed ecosystems, and coral reef ecosystems [24]. The natural conditions of the region influence the economic characteristics of the population. Most of the community runs businesses in the industry, including snorkeling,..., tour guide service, or homestay rental, while the rest still pursue a profession as a fisherman [25]. This proves that geographic condition plays a major role to the community's economy.

The types of data used in the research consist of satellite imagery and reference data. For benthic habitat classification

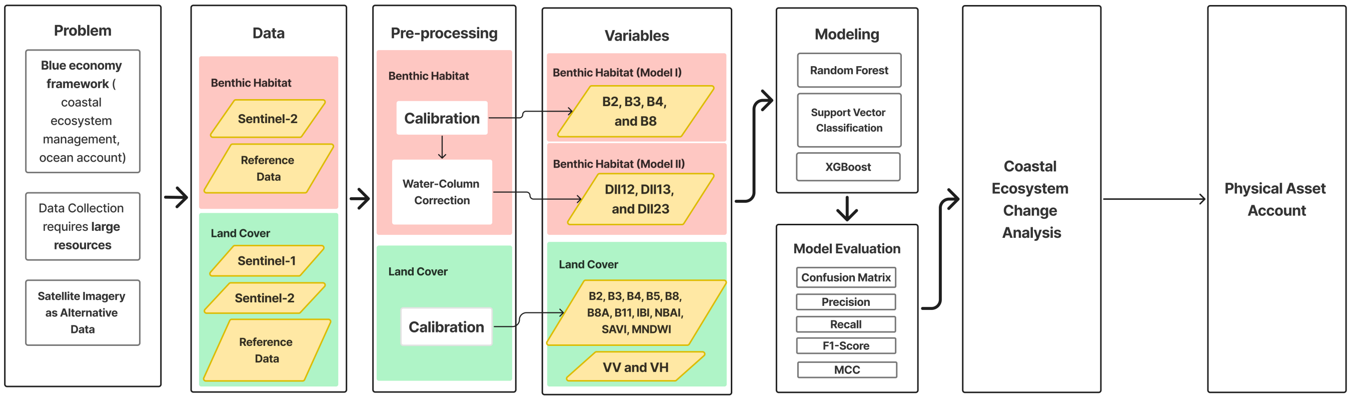


Figure 1. Research Framework

scheme, Sentinel-2 Level 2A Surface Reflectance (SR) and benthic habitat class shapefile were used as satellite imagery and reference data. Data taken from this image include B2 (blue band), B3 (green band), B4 (red band), and B8 (NIR band). The reference data used is 2021 benthic mapping data produced by the National Research and Innovation Agency (BRIN) which consists of five classes in it (coral, seagrass, macroalgae, sand, and mixed classes). Meanwhile, the land cover classification scheme used Sentinel-2 Level 2A SR and Sentinel-1 Ground Range Detected (GRD) as satellite imagery also land cover class shapefile as the reference data. Data taken from Sentinel-1 include Vertical transmit/Vertical receive (VV) and Vertical transmit/Horizontal receive (VH) bands whereas the data taken from Sentinel-2 include B2 (blue band), B3 (green band), B4 (red band), B8 (NIR band), B5 (Red Edge 1), B8 (NIR), B8A (Red Edge 4), B11 (SWIR 1), Modified Normalized Difference Water Index (MNDWI), Soil Adjusted Vegetation Index (SAVI), Index-based Built-up Index (IBI), and Normalized Built-up Area Index (NBAI) values. The reference data used is the 2019-2020 land cover data produced by the Ministry of Environment and Forestry (KLHK) which has four land cover classes in it (settlement class, airport, secondary dryland forest, and secondary mangrove forest). The satellite imageries products were obtained through Google Earth Engine (GEE). The data descriptions and sources used in this study are shown in Table 1.

2.2. Methods of data analysis

1. Random Forest (RF)
   Random Forest is a classification method composed of a collection of classifiers in the form of decision trees and letting the dominant class be the final class [26]. The general procedure is that for each tree classifier, a random training set is resampled from the original training set using bootstrap that leads to growing the tree and resulting in a tree classifier. Let $\hat{y}_i$ be the i-th predicted observation target, k be the number of trees, f_k be the k-th independent tree function, T as the training set, T_k as the bootstrap training set, and x, known as input, is the selected features in each tree. The tree classifiers are notated as follows:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x, T_k). \tag{1}$$

The tree is grown according to the classification and regression tree (CART) methodology, except for the pruning process. In growing tree, the node is split by selecting the split that decreases the impurity the most using only selected features and will stop once the elements contained in a node are fewer than the predetermined minimum elements [27]. Below is the formula of Gini impurity that is being utilized in this case [28]:

$$\text{Gini Impurity} = \frac{1}{2}\sum_i p_i(1-p_i) = \sum_{i<j} p_i p_j, \tag{2}$$

where:

p_i : probability of a data point with class label-i,
p_j : probability of a data point with class label-j,
i, j : class label.

Each generated tree then casts a vote in order to retrieve the final class for each training input. The constructed tree classifiers were then evaluated by calculating its error rate using the Out-Of-Bag samples, the left-out training set from bootstrap, to estimate the overall error of a classifier. In terms of coastal ecosystem classification, Random Forest has become one of the popular algorithms used in study cases. According to a study conducted by Traganos and Reinartz [29], the algorithm succeeded in classifying seagrass by yielding 96. 4% precision. Another study conducted by Wicaksono, et al. [19], yielded 88% precision in classifying benthic habitat. Whereas in the tidal area, Ghorbanian et al. [30] achieved 93% overall accuracy and 0.92 F1 score and Asy'ari et al. [31] achieved 96. 48% overall accuracy for classifying the mangrove ecosystem.

2. Support Vector Classification (SVC)
   SVC is a form of implementation of the Support Vector Machine (SVM) algorithm for classification tasks. SVM itself is an algorithm that works by finding the most optimal linear

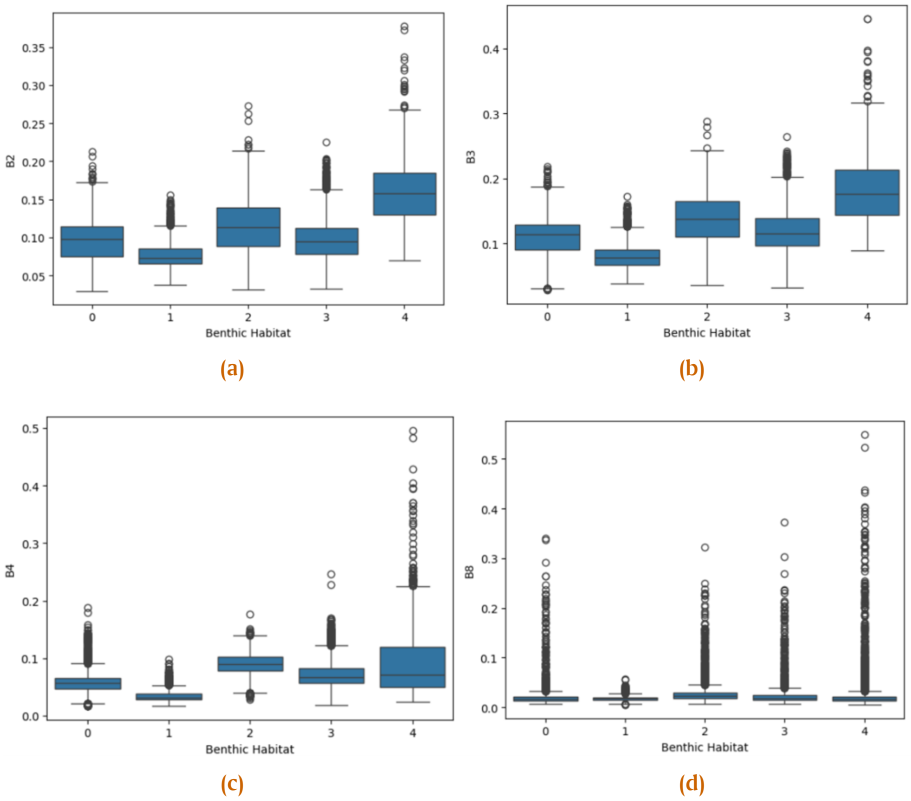


Figure 2. Box-plot of Sentinel-2 Surface Reflectance Bands (a) B2 band, (b) B3 band, (c) B4 band, and (d)B8 band (code 0 for mixed class, code 1 for coral reef class, code 2 for seagrass class, code 3 for macroalgae class, and code 4 for sand class)

separating hyperplane (decision boundary that separates one class from another) to carry out its task [32]. Here is the notation of the formula [33]:

$$h_{w,b}(x) = g(w^T x + b), \tag{3}$$

where w is a weight vector, namely W = $\{w_1, w_2, \cdots, w_n\}$, n is the number of attributes, and b referred as bias. This algorithm allows it to be applied to data with both linear and non-linear forms. In non-linear cases, the data needs to be transformed using kernel function first before performing the hyperplane search. Kernel function that is being utilized in this case is Gaussian radial basis function kernel. SVC, as an algorithm that is widely used, has given good performance for classifying coastal ecosystem in previous study. For instance, Wang et al. [34] conducted a research in classifying mangrove species that resulting in 79.63% overall accuracy, Zhao et al. [35] stated that the model performed well in classifying water class which closely related in mangrove habitat, Wicaksono et al. [36] proved SVM model achieved 73.23% overall result which was the highest result compared to others in a conducted research classifying benthic habitat, and Lazuardi et al. [11] succeeded in classifying coral reef and seagrass using SVM with 73.14% accuracy.

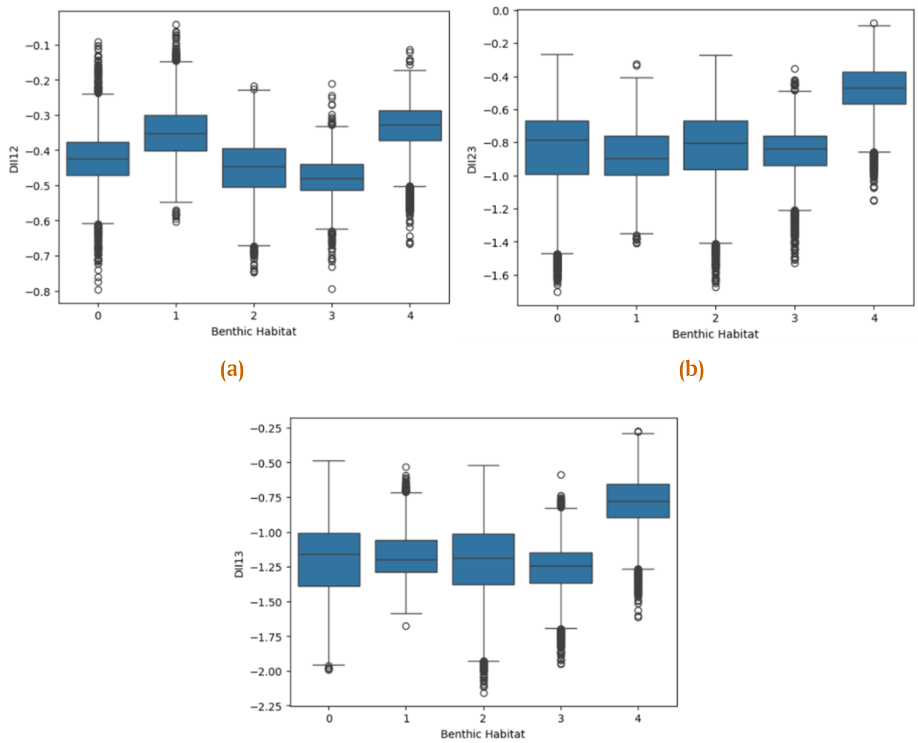
3. Extreme Gradient Boost (XGBoost)
XGBoost fundamentally consists of an ensemble of decision trees that are grown using CART algorithm. The difference is that the tree boosting system uses minimized regularized objective and additive function as parameters [37]. Notated below is the formula of the objective function [29, 30].

$$\text{Objective Function} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Upsilon T + \frac{1}{2}\lambda \sum_{T}^{j=1} w_j^2, \tag{4}$$

where:

$$\begin{aligned} n &: \text{data observation,} \\ l &: \text{differentiable loss function,} \\ y_i &: \text{actual target of the i-th instance,} \\ \hat{y}_i &: \text{predicted target of the i-th instance,} \\ \Upsilon &: \text{complexity cost,} \\ T &: \text{number of leaves,} \\ \lambda &: \text{lambda,} \\ w_i &: \text{score on the i-th leaf.} \end{aligned}$$

The algorithm uses shrinkage technique and feature subsampling technique in order to prevent overfitting. XGBoost equipped with advance split finding algorithms so that the model can find the best split in tree learning also processing big-size data and input data sparsity. This learning method accommodates optimal cache-aware prefetching algorithm, data compression, and additional techniques to improve out-of-core computation. These combined insights enable XGBoost to solve real-world-scale problems using minimal resources. There has not been many studies in classifying coastal ecosystem. A previous study conducted by Nemani,



(a) (b) (c)

Figure 3. Box-plot of Depth-Invariant Index Bands (a) DII12 band, (b) DII23 band, and (c) DII13 band (code 0 for mixed class, code 1 for coral reef class, code 2 for seagrass class, code 3 for macroalgae class, and code 4 for sand class)

et. al. [17] showed that XGBoost performed better compared to SVM and RF in classifying benthic assemblages with 61.67% accuracy.

3. Research Framework

This study focuses on spatial classification of coastal ecosystem and analysis change in the east region of Karimunjawa Islands, which consists of Karimunjawa island, Kemujan island, Menjangan Kecil island, and Menjangan Besar island, in 2020, 2021, and 2023 using three machine learning algorithms with the dependent variable derived from the reference data. For benthic classification scheme, the independent variables for the first model that will be used are the band values of Sentinel-2A SR and benthic shapefile. The band values comprise of B2 (blue band), B3 (green band), B4 (red band), and B8 (NIR band). The independent variables for the second model are the depth-invariant indexes resulting from water-column correction pre-processing. These variables are DII12 (combination of blue and green band), DII13 (combination of blue and red band), and DII23 (combination of green and red band). The period of extracted imagery is 2021 complying to the period of the reference data for training and validation data. As for the land cover classification scheme, the independent variables are the index and band values of Sentinel-2A SR and Sentinel-1 GRD also the land cover shapefile. These variables consist of VV and VH bands from Sentinel-1 also B2 (blue band), B3 (green band), B4 (red band), B8 (NIR band), B5 (Red Edge 1), B8 (NIR), B8A (Red Edge 4), B11 (SWIR 1), IBI, SAVI, MNDWI, and NBAI values from Sentinel-2A. Like the benthic classification scheme, the imagery in this scheme is extracted for 2019-2020 time period complying to the reference data. The 2019 imagery and reference data play the role in training step, while the 2020 imagery and reference data in testing step. For training model, the sampled training data retrieved from random sampling 10% of the original training data was used for both schemes. Not only to avoid overfitting, but also previous spatial studies have shown that there are no significant differences in model performance between using sampled training and entire training data [38, 39]. In benthic habitat classification scheme, six models are generated based on the imagery and methods. Both type of imagery, original and water-column corrected imagery, are trained using the three methods. Meanwhile, the land cover classification scheme generated three models based on the method only. Once the model from each chosen algorithm is built, the best model in each scheme would be chosen in the model evaluation process. If the best model has already been determined, it would be applied to the imagery in 2020 and 2023 for predicting the coastal ecosystem category, the dependent variable. The purpose of choosing these two periods is to analyse the coastal ecosystem change in three years interval and would be useful to give information or insights in arranging ocean account later. Hopefully, this study gives insight for the government on using satellite imagery as alternative data for managing coastal ecosystem and ocean account as the initiative effort in implementing blue economy framework. The research flow is explained in Figure 1.

4. Results and Discussions

Before discussing the process of training the model, the variables used were first explored in order to overview the char-

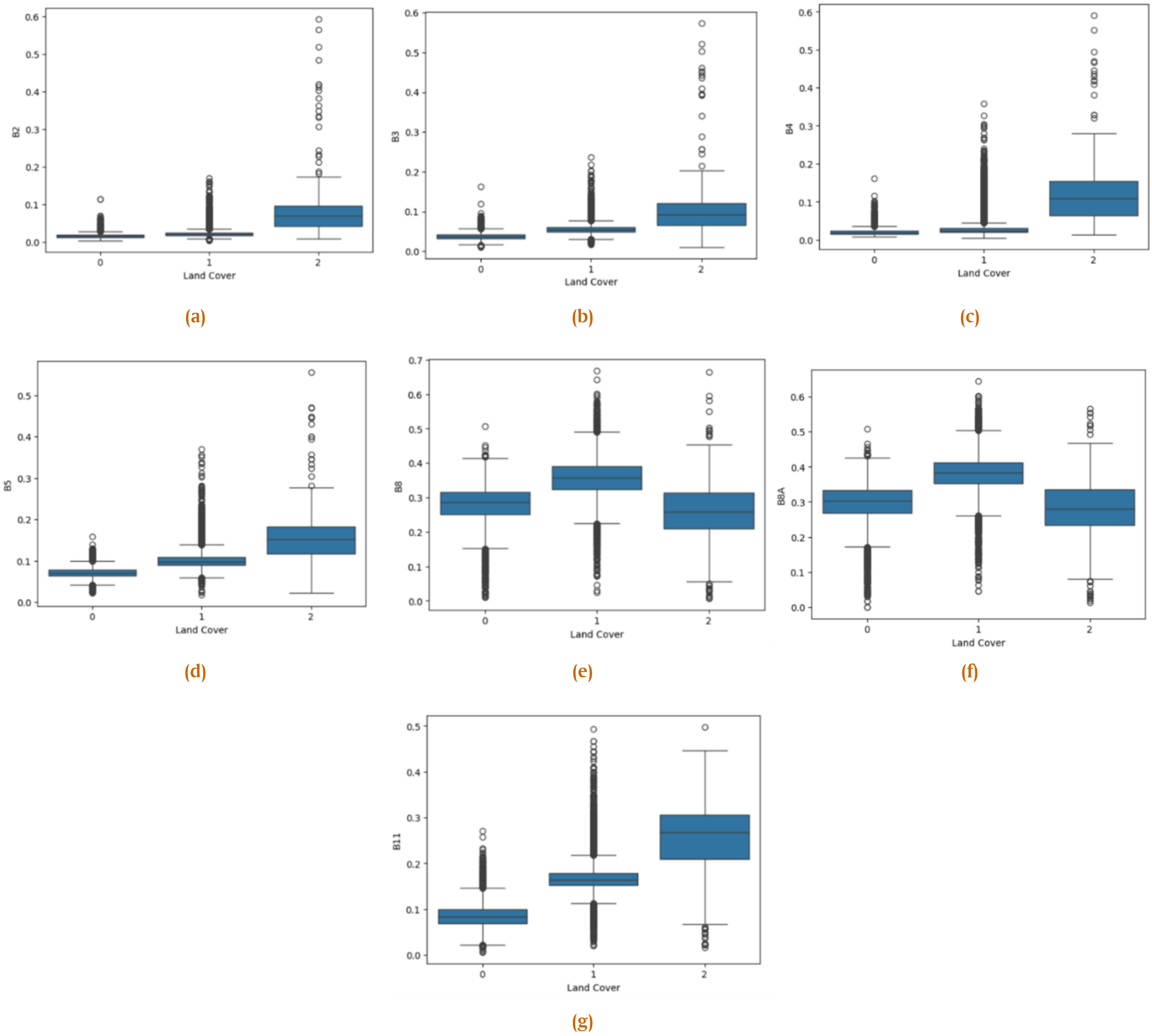


Figure 4. Box-plot of Sentinel-2 Surface Reflectance Bands (a) B2 band, (b) B3 band, (c) B4 band, (d) B5 band, (e) B8 band, (f) B8A band, and (g) B11 band (code 0 for mangrove class, code 1 for forest class, code 2 for built-up class)

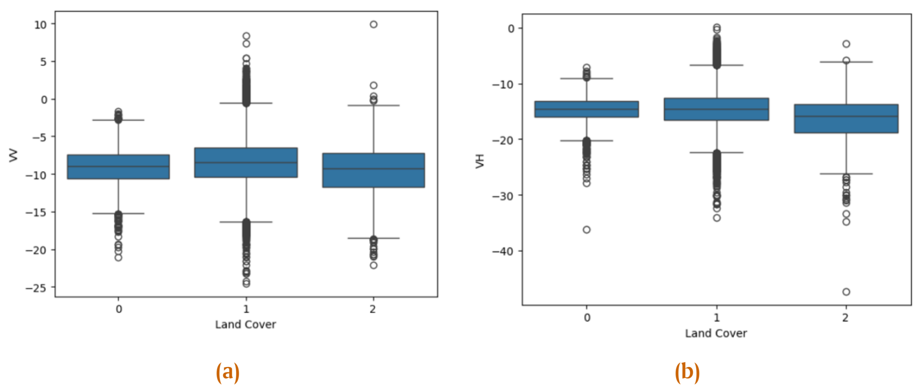


Figure 5. Box-plot of Sentinel-1 Ground Range Detected Images (a) VV polarisation and (b) VH polarisations (code 0 for mangrove class, code 1 for forest class, code 2 for built-up class)

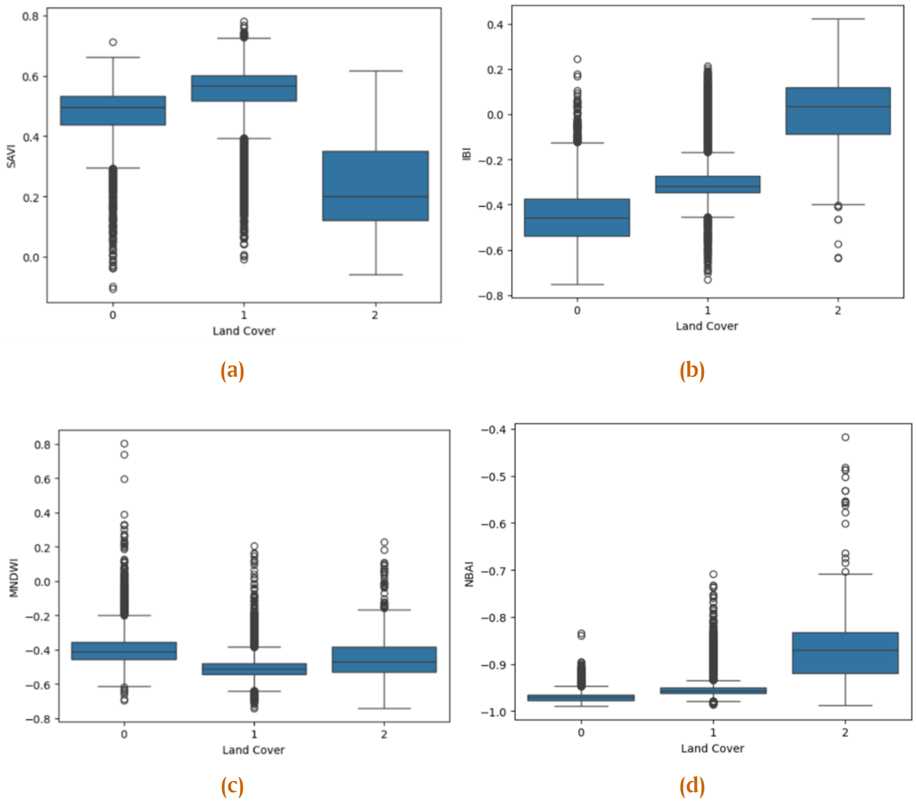


Figure 6. Box-plot of Indexes (a) Soil Adjusted Vegetation Index (SAVI), (b) Index-based Built-up Index (IBI), (c) Modified Normalized Difference Water Index (MNDWI), and (d) Normalized Built-up Area Index (NBAI). Code 0 for mangrove class, code 1 for forest class, code 2 for built-up class

acteristics of each variable between classes. The exploration is divided into benthic habitat and land cover classification since each scheme has their own variables.

4.1. Exploration of benthic habitat classifier variables

There are two models generated from the benthic habitat classification, namely the model with Sentinel-2 Surface Reflectance input image and the model with Sentinel-2 Depth-Invariant Index input image (Sentinel-2 Surface Reflectance image with water column correction). The model with Sentinel-2 Surface Reflectance input image uses visible band values (B2, B3, B4) and infrared band (B8). The band values used in classifying benthic habitats in this model are visible band (B2, B3, B4) and infrared band (B8). This is because the classification of shallow water objects is limited optically because the ability of wave penetration to take images is limited to the visible band only [40]. Meanwhile, other studies have also used the B8 value to classify other than the visible band [10]. According to Figure 2, several conclusions can be drawn about the characteristics of the predictor variables between the four classes.

In the B2 (Blue) and B3 (Green) values, the y-axis, the sand and coral classes can be distinguished from the other classes. The pixels value of coral class tends to be lower than the other classes, while the sand class' pixels value is the opposite to coral class. Other classes, such as the mixed class, seagrass class, and macroalgae class have intersected range of values with seagrass class that has the widest interval value among the three classes. As for the B4 (Red) value, the characteristics for coral class and sand class are still similar to the previous two band values while the other classes have slight difference. The mixed class and macroalgae class have similar value range while the seagrass class is slightly higher than the two classes. Another difference is shown from the sand class that has more outlier. Whereas for B8 (Near Infrared) band value, the value range rather smaller than the other band values so the difference between classes is difficult to spot. Also, each class has more outliers.

The coral class' outliers are the pixels that overly dark-coloured or slightly light-coloured compared to the most pixels in the class due to the class' pixels dominated with intense green colour. When investigating the imagery for sand class, turns out the outliers in sand class were the pixel that is brownish and dark. This is in line with the boxplot results since the chosen sand class observation were dominated by greenish and bluish pixels. Outliers for mixed class and seagrass classes have similarity that both classes tend to see dark or intense colour pixel as an observation that is not common. Whereas, macroalgae classes find lighter pixel or overly dark pixel is uncommon to the class. Coastline area, deep water transitional area, water area near mangrove, or water area near terrestrial tend to have outliers.

The model with Sentinel-2 DII input image uses three variables, namely band DII12 (Blue/Green), DII23 (Green/Red), dan DII13 (Blue/Red). Figure 3 below display the boxplot visualisation of each variable.

From Figure 3, the DII12 values of the mixed, seagrass,

Table 2. Selected Parameter for Machine Learning Model

Classification Scheme	RF Parameter	SVC Parameter	XGBoost Parameter
Benthic Habitat Classification (Surface Reflectance Bands Input)	$n_{\text{estimators}} = 100$	C =1000	tree_method = approx; max_depth = 12; min_child = 1; subsample = 0.2728; colsample_bynode = 0.5354; reg_lambda = 0.058
Benthic Habitat Classification (Depth-Invariant Index Bands Input)	$n_{\text{estimators}} = 300$	C =1000	tree_method = approx; max_depth= 5; min_child = 2; subsample = 0.4921; colsample_bynode= 0.8742; reg_lambda = 0.0109
Land Cover Classification	$n_{\text{estimators}} = 400$	C =1000	tree_method = approx; max_depth= 9; min_child = 1; subsample = 0.967; colsample_bynode= 0.602; reg_lambda = 0.422

and macroalgae classes have a similar range of data distribution. Meanwhile, coral and sand classes have little similarity in distribution between each other and both classes have higher values than other classes. Although the ranges of DII23 and DII12 values are different, the distribution patterns between classes are similar. In both box-plots, the mixed, coral, seagrass and macroalgae classes have similar data distribution characteristics to each other, while the sand class is the only class that differs. Overall, the outliers here are the very dark pixels. The class' outlier usually found in the area on the deeper water, either in the transition towards deep water area or deeper water area surrounded by shallow water areas, and darker water, e.g. water area near mangrove. In this case, sand class has distinctive outliers that not only darker pixels near mangrove or terrestrial area are considered outlier, but also the outlier can be found in coastline area with lighter pixels.

4.2. Exploration of land cover classifier variables

Land cover classification was performed using B2 (blue), B3 (green), B4 (red), B5 (Red Edge 1), B8 (NIR), B8A (Red Edge 4), and B11 (SWIR 1) values from Sentinel-2 imagery; VV and VH polarization from Sentinel-1 imagery; and MNDWI, SAVI, IBI, and NBAI. The use of B2, B3, B4, B5, B8A, and B11 values are stated as features that provide optimal and effective results in identifying mangrove forest objects [41]. The combination of B8A, B11, and B4 has the ability to identifying mangrove classes from other land cover classes since it is sensitive in detecting wetlands and the use of band values with a resolution of 10 meter (B2, B3, and B4 bands) improves the classification of mangrove ecosystem classes due to the higher resolution compared to other bands [30]. Meanwhile, the use of Sentinel-1 imagery with VV and VH polarization and Sentinel-2 imagery in classifying mangroves, according to previous research, can improve accuracy [12].

The visualization in Figure 4 shows that the values of B2 to B5 have similar patterns between classes. The built-up class here has overall higher values than the other two classes, while the forest class and mangrove class have a similar range of values. There is a similar pattern of B8 and B8A, where the forest class has a higher value followed by the mangrove class, while the built-up class has a lower value than the other two classes. The B11 pattern is most different from the other bands. The forest class has a higher value than the mangrove class and the built-up class has the highest value among the three classes. As for the box-plot of VV and VH variable in Figure 5, the three classes share a similar distribution of values, although the two variables have different ranges, making it difficult to distinguish one from the other.

Based on the visualization of Figure 6, the SAVI index between the three classes has a different distribution pattern. The forest class has the highest value, while the built-up class has the lowest value. This is in line with previous references that the higher the vegetation condition, the higher the SAVI value [42, 43]. The IBI index can illustrate different distribution patterns between the three classes. Mangrove classes tend to have low values and built-up classes tend to have high values. This value is in line with previous research which states that the IBI value in the built-up class is more than zero, while other classes are less than equal to zero [44]. The MNDWI index shows a slightly different distribution of values between the mangrove class and other classes. This class tends to have slightly higher values than the other classes with a range of values smaller than zero. For the NBAI index, this index can give a different pattern to the built-up class from the other classes. The developed class has a higher index value than the other classes.

Forest class' outliers usually caused by the appearance of darker green forest or lighter green forest and misclassification of other classes, such as road and inland water area. For the outliers in mangrove class, the pixel of darker appearance considered as an outlier, for instance darker green colour and lighter green colour. Whereas, the outliers found in built-up classes are usually the misclassification of forest, lighter pixels without, or other darker pixels, such as shrimp farms and airport runway.

4.3. Machine learning model building and its performance evaluation

In building model, the parameter needs to be inputted so the model can be trained using each algorithm with the extend of determined parameter. The search was conducted by dividing the data samples using 10-fold Stratified Cross Validation. Dividing the data using cross validation techniques in finding the best parameters can prevent overfitting compared to dividing the

Table 3. 10-Fold Stratified Cross Validation Habitat Benthic Models Using F1-Score Metric

	Input Image					
	Surface Reflectance	Depth-Invariant Index	Surface Reflectance	Depth-Invariant Index	Surface Reflectance	Depth-Invariant Index
	RF		SVC		XGBoost	
Overall	0.773	0.586	0.764	0.707	0.771	0.548
Std. Deviation	0.0067	0.0101	0.072	0.0088	0.01003	0.0092

Table 4. Precision, Recall, and F1-Score Selected Model

Evaluation Metrics		Mixed	Coral Reef	Seagrass	Macroalgae	Sand
Precision	Overall	0.624	0.944	0.817	0.642	0.845
	Std. Deviation	0.02154	0.00490	0.01616	0.02227	0.01432
Recall	Overall	0.523	0.949	0.867	0.579	0.893
	Std. Deviation	0.02369	0.00539	0.01735	0.02166	0.01269
F1-Score	Overall	0.570	0.948	0.841	0.610	0.868
	Std. Deviation	0.01897	0.004	0.00943	0.02	0.00748

data once. The search is conducted by defining the candidate parameter values to be tried first by defining a certain range of values. The parameter sought in RF modelling is the number of trees ($n_{\text{estimators}}$) with candidates of 100 to 500 trees with an interval of 100 additions. Based on previous research, benthic classifiers provide significant performance when the parameters are between 100 - 500 trees. For the SVC method, a value search was conducted on the C or regularization parameter. The search for the value of the parameter was carried out using a range of values between 0.01 to 1000 with a multiplication value of 10 between value candidates so that there were six candidate parameter values following previous research [36]. In training the XGBoost, the parameters that was tuned are tree_method (the method on constructing tree), max_depth (maximum depth of the tree), min_child weight (minimum sum of instance weight in a child), subsample (the ratio of subsample in training instances), colsample_bynode (the ratio of columns subsample every split), and reg_lambda (the L2 regularization term for weight) [45]. The parameter search is performed using the GridSearchCV function of the scikit-learn model and optuna framework. Based on the hyperparameter tuning, the best parameter of each model is described in Table 2.

Using the resulting parameters, a model was built. The next stage, model evaluation, is conducted to measure the performance of the model. This measurement is done using confusion matrix and other accuracy metrics. The metrics used are F1-score macro (arithmetic average F1-score of the entire class) and MCC. F1-score is a summary metric obtained from the calculation of the harmonic average of recall metric which takes into account the fraction of the correctly predicted class observations to the entire actual observations of the class and precision metric which takes into account the fraction of the correctly predicted class observations to the entire observations that are predicted to be the class [21]. Whereas, MCC, also another summary metric, is a correlation between the actual observation target and predicted observation target [46]. The higher the score of both metrics, the better the performance of the built model. The lowest value of the F1-score metric is 0 and the highest value is 1, while the lowest value of MCC metric is -1 and the highest is 1.

For the land cover model, the evaluation metrics were conducted with testing data, comparing the classification results with the actual data. Meanwhile, the benthic cover model was evaluated using the k-fold stratified cross-validation technique with k of 10. The measured metrics were then used as the basis for selecting the best model for each classification scheme. The results of the evaluation metrics of benthic habitat modelling are shown in Tables 3 and 4.

In the evaluation using F1-score as shown in Table 3, it can be concluded that the best modelling is produced by the RF algorithm with the surface reflectance input image yielding the highest F1-score and the lowest standard deviation among all algorithms and imageries. This shows that the model not only able to perform better classification, but also gives robust results. Therefore, it is decided that the RF algorithm with surface reflectance image is the best model. This result is in line with related research which states that surface reflectance input provides better modelling than depth-invariant index as the input of the model [20, 36]. Table 4 shows selected model evaluation metrics using the previous validation technique, 10-cross stratified cross-validation, of each class.

From Table 4, the model still has poor performance in the mixed and macroalgae classes based on the overall F1-score. The metric value in the mixed and macroalgae classes is still less than 0.8 and does not have the same consistent value as the coral, seagrass and sand classes. The mixed and macroalgae classes are difficult to distinguish well compared to other classes since the pixels of both classes are similar to each other. This indication comes from the box-plot visualization that shows that the both class pixel range noticeably intersects in B2, B3, and B4 band variables. Also, the model's performance on these classes is the least robust based on the standard deviation of each metrics.

Table 5. Modelling Evaluation of Data Testing

Model Algorithm	F1-Score			F1-Score Macro	MCC
	Mangrove	Forest	Built-Up		
RF	0.89	0.97	0.63	0.83	0.839
SVC	0.90	0.97	0.62	0.83	0.846

For land cover classification, the modelling results are then evaluated using training and testing data using a sample of 2020 satellite imagery. Table 5 displays the results of modelling evaluation of data testing using F1-score of each class, F1-score macro, and MCC metrics.

Based on Table 5, the best modelling is generated by the

SVC algorithm based on the F1-score macro and MCC metrics. Choosing the best model was decided using the SVC algorithm because it gave the best results on the testing data since the dataset was considered independent from the training data and it tests the universality of the model. Even though SVC outperform RF model only by 0.007 in MCC metrics. Based on the algorithm itself, the SVC works better for fewer-class classification. This is because the basic algorithm uses hyperplane to separate different classes. Table 6 shows classification report that compares the classification results with the actual class of the testing data.

Table 6. Precision, Recall, and F1-Score SVC Model Land Cover Classification

Class	Precision	Recall	F1-Score
Mangrove	0.89	0.91	0.9
Forest	0.96	0.97	0.97
Built-Up	0.92	0.46	0.62
F1-Score Macro			0.83

Based on the results of Table 6, overall, the model can classify well in mangrove and forest classes. The indication is based on the value shown in the table. The chosen model succeeded in classifying the land cover well, especially for forest class. It performs really well and able to exceed 0.8 in evaluation score. Unfortunately, for class Built-Up, the model is still not able to perform well in retrieving good recall score, which means it can only predict less than half of the overall true Built-Up class observations correctly. Interestingly, the same pattern was also found in RF algorithm. This indicates that the existence of outliers shown in box-plot earlier doesn't affect the model's performance in classifying the land cover. The poor score might be yielded since the sample amount of this class is fewer than the others despite having outliers.

4.4. Classification and analysis of coastal ecosystem change in 2023

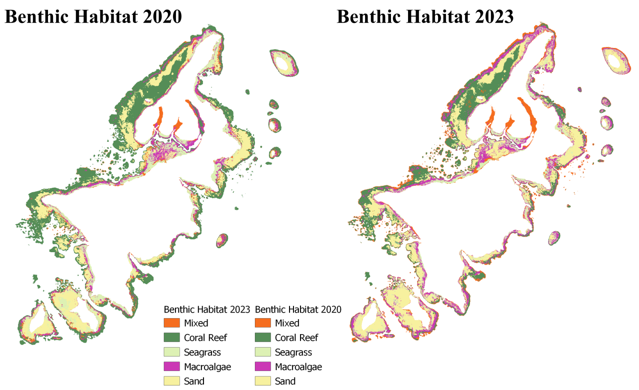


Figure 7. Mapping of Benthic Habitat Classification of 2020 and 2023

After obtaining the best model, the classification of coastal ecosystems was carried out in two different periods, namely the 2020 and 2023 periods. The chosen models were applied to the exported satellite imagery raster file and the classified raster is exported using rasterio library in python later will be displayed in Figures 7 and 8. The area of each class is then calculated from the classified raster for each period that will be shown in Tables 7 and 8. Figure 7 and Table 7 are the result of the benthic habitat classification scheme.

Table 7. Benthic Habitat Area in 2020 and 2023 According to Random Forest Classification Result

Class	Area in 2020 (km^2)	Area in 2023 (km^2)
Mixed	2.5240	5.6467
Coral Reef	15.0984	8.5743
Seagrass	5.4698	4.0798
Macroalgae	3.7490	6.1571
Sand	8.1204	10.4933
Total Area	34.9615	34.9512

Table 8. Benthic Habitat Area in 2020 and 2023 According to Random Forest Classification Result

Class	Area in 2020 (km^2)	Area in 2023 (km^2)
Mangrove	5.8707	5.2176
Forest	28.4984	29.0440
Built-Up	0.6321	0.7389
Total Area	35.0012	35.0005

From the results of benthic habitat classification in Table 7, there are changes in benthic habitat between the two periods. The mixed class experienced an expansion with an increase in the 2023 period of more than double compared to the 2020 period. Similar to the mixed class, the macroalgae class and sand class also both increased in area, although not as significant as the mixed class. Meanwhile, in the coral class, there was a decrease. The coral area decreased by 6.524 km^2 from the previous period. The seagrass class also decreased between the two periods. The seagrass class decreased by 1.39 km^2. However, it is important to note that these changes were analysed based on modelling results, so misclassification may occur when applying the model to both satellite imagery periods, especially in the seagrass and macroalgae classes. Figure 8 and Table 8 are the result of the land cover classification scheme.

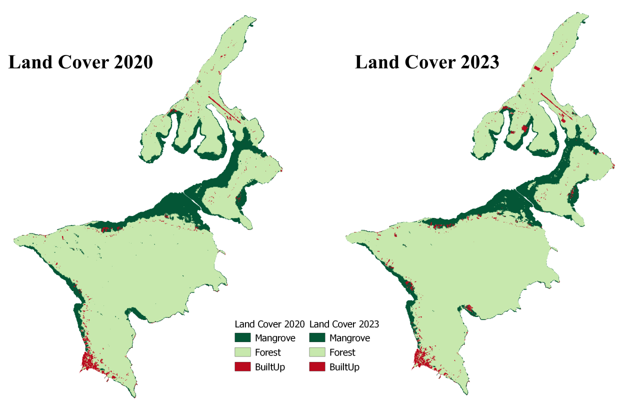


Figure 8. Mapping of Land Cover Classification Results of 2020 and 2023

Meanwhile, the results of the land cover classification in Table 8 show that there are changes in all three classes. The mangrove class decreased in area by 0.599 km^2. In contrast to the mangrove class, the forest class and the built-up class experienced expansion. The forest class expanded by 0.5456 km^2, while the built-up class expanded by 0.1068 km^2.The area of the classification results was then evaluated by comparing the area with the area of the reference data. The evaluation was only carried out on the land cover classification scenario by comparing the area of the classification results with the land cover testing data in 2020. The evaluation of land cover area is shown in Table 9.

Table 9. Comparison of Predicted Land Cover Area with Reference Data in 2020

Class	Predicted Area (km^2)	Reference Area (km^2)	Difference (km^2)	Absolute Difference Percentage(%)
Mangrove	5.8707	5.0577	0.8130	16.07
Forest	28.4984	29.1474	-0.6490	2.23
Built-Up	0.6321	0.7964	-0.1643	20.63

From the evaluation results, it can be concluded that there is an overestimation of the mangrove class, while the other classes experience a smaller estimation than the actual data. The built-up class provides the smallest difference between the two area values, while the mangrove class has the largest difference. However, when viewed in the absolute difference percentage column, the forest class has the smallest value and the built-up class has the largest value.

5. Conclusion

Based on the results and discussion in the previous section, there are several conclusions that can be drawn.

1. The one most common pixel characteristic is that the value of each class in each variable is various. Certain classes even have similar value distribution. However, after training and evaluating/validating the model, the model excellent performance is able to classify the category despite having outliers.
2. For land cover classification, it is shown that Support Vector Classifier algorithm slightly outperforms Random Forest algorithm. Unfortunately, the SVC model still performs poor for correctly predicting built-up classes due to the built-up class being the minority class in the locus. As for the imagery input in benthic habitat classification scheme, Sentinel-2, surface reflectance images performed better than images with depth-invariant index bands.
3. The best model was built with the Random Forest algorithm and surface reflectance as input. Even though the model is able to predict most of the classes, disappointing performance when classifying mixed class and macroalgae class because of the similar pixel characteristics between the two classes.
4. As for the result of change analysis, area changes of certain class are spotted between 2020 and 2023. There are significant changes in benthic habitat area. The biggest change (decrease) in benthic habitat area is coral reef. Meanwhile, there are slight differences of area in land cover classification between two periods. Mangrove class was decreasing by 0.599 km2 whereas the forest and built-up area was increasing by 0.546 km^2 and 0.107 km^2. These changes might be due to the model's misclassification or actual changes which need to be confirmed by ground-check validation.

Author Contributions. **Jane, G.J.**: data curation, validation, formal analysis, writing, and editing. **Alifatri, L.O.**: conceptualization, methodology. **Tasriah, E.**: methodology, investigation. **Pramana, S.**: conceptualization, methodology, writing—review and editing, supervision.

Acknowledgement. The authors would like to express gratitude to all individuals and institutions that supported this research through resources, guidance, or critical feedback.

Funding. This research received no external funding.

Conflict of interest. The authors declare that they have no competing interests or conflicts of interest to report on the present study.

Data availability. The satellite imagery datasets are available from Google Earth Engine (GEE).

Abbreviations.

RF	:	Random Forest.
SVC	:	Support Vector Classification.
XGBoost	:	Extreme Gradient Boosting.
MCC	:	Matthew's Correlation Coefficient.
OECD	:	Organisation for Economic Cooperation and Development.
BAPPENAS	:	Ministry of National Development Planning Agency.
ASEAN	:	Association of Southeast Asian Nations.
SEEA	:	System of Environmental Economic Accounting.
Sisnerling	:	Integrated System of Environmental and Economic Accounts.
MPD	:	Mobile Positioning Data.
GRD	:	Ground Range Detected.
VV	:	Vertical transmit/Vertical receive.
VH	:	Vertical transmit/Horizontal receive.
MNDWI	:	Modified Normalized Difference Water Index.
SAVI	:	Soil Adjusted Vegetation Index.
IBI	:	Index-based Built-up Index.
NBAI	:	Normalized Built-up Area Index.
KLHK	:	Ministry of Environment and Forestry.
BRIN	:	National Research and Innovation Agency.
GEE	:	Google Earth Engine.
CART	:	Classification and Regression Tree.
SVM	:	Support Vector Machine.

References

[1] U. Nations, "Blue economy definitions," 2024, https://www.un.org/regularprocess/sites/www.un.org.regularprocess/files/, Accessed on 27 August 2024.

[2] OECD, "Sustainable ocean economy country diagnostics of indonesia," 2021, https://www.oecd.org/en/publications/sustainable-ocean-economy-country-diagnostics-of-indonesia_9bc36234-en.html, Accessed on 11 July 2024.

[3] ERIA, "Asean blue economy framework," 2023, https://asean.org/wp-content/uploads/2023/09/ASEAN-Blue-Economy-Framework.pdf, Accessed on 27 August 2024.

[4] KKP, "Report of the ocean accounts development report in indonesia," 2022, https://oceanaccounts.atlassian.net/wiki/spaces/WD/pages/939884628/, Accessed on 6 June 2024.

[5] S. Pramana *et al.*, "Big data for government policy: Potential implementations of bigdata for official statistics in indonesia," in *2017 International Workshop on Big Data and Information Security (IWBIS)*, pp. 17–21, 2017. DOI:10.1109/IWBIS.2017.8275097

[6] K. Maurya, S. Mahajan, and N. Chaube, "Remote sensing techniques: mapping and monitoring of mangrove ecosystem—a review," *Complex and Intelligent Systems*, vol. 7, no. 6, pp. 2797–2818, 2021. DOI:10.1007/s40747-021-00457-z
[7] T. Nguyen *et al.*, "Mapping of coral reefs with multispectral satellites: A review of recent papers," *Remote Sensing*, vol. 13, no. 21, pp. 1–25, 2021. DOI:10.3390/rs13214470
[8] NOAA, "What is a benthic habitat map?," 2024, https://oceanservice.noaa.gov/facts/benthic.html, Accessed on 10 April 2024.
[9] A. Anas, V. Siregar, and S. Wouthuyzen, "Performance of mlh and svm algorithms in mapping macroalga habitats using satellite data in pannikiang island, south sulawesi," *MAJALAH ILMIAH GLOBE*, vol. 25, pp. 97–108, 11 2023.
[10] P. Wicaksono, M. A. Fauzan, and S. G. W. Asta, "Assessment of sentinel-2a multispectral image for benthic habitat composition mapping," *IET Image Processing*, vol. 14, no. 2, pp. 279–288, 2020. DOI:10.1049/iet-ipr.2018.8044
[11] W. Lazuardi, P. Wicaksono, and M. A. Marfai, "Remote sensing for coral reef and seagrass cover mapping to support coastal management of small islands," *IOP Conference Series: Earth and Environmental Science*, vol. 686, no. 1, p. 012031, 2021. DOI:10.1088/1755-1315/686/1/012031
[12] A. Sharifi, S. Felegari, and A. Tariq, "Mangrove forests mapping using sentinel-1 and sentinel-2 satellite images," *Arabian Journal of Geosciences*, vol. 15, no. 20, p. 1593, 2022. DOI:10.1007/s12517-022-10867-z
[13] K. Upakankaew *et al.*, "Discrimination of mangrove stages using multitemporal sentinel-1 c-band backscatter and sentinel-2 data—a case study in samut songkhram province, thailand," *Forests*, vol. 13, no. 9, p. 1433, 2022. DOI:10.3390/f13091433
[14] X. Liu *et al.*, "Large-scale high-resolution coastal mangrove forests mapping across west africa with machine learning ensemble and satellite big data," *Frontiers in Earth Science*, vol. 8, 2021. DOI:10.3389/feart.2020.560933
[15] B. Misiuk and C. J. Brown, "Benthic habitat mapping: A review of three decades of mapping biological patterns on the seafloor," *Estuarine, Coastal and Shelf Science*, vol. 296, p. 108599, 2024. DOI:10.1016/j.ecss.2023.108599
[16] T. D. Pham *et al.*, "Remote sensing approaches for monitoring mangrove species, structure, and biomass: Opportunities and challenges," *Remote Sensing*, vol. 11, no. 3, p. 230, 2019. DOI:10.3390/rs11030230
[17] S. Nemani *et al.*, "A multi-scale feature selection approach for predicting benthic assemblages," *Estuarine, Coastal and Shelf Science*, vol. 277, p. 108053, 2022. DOI:10.1016/j.ecss.2022.108053
[18] BTNKJ, *Panduan Pendidikan dan Penelitian di Taman Nasional Karimunjawa*, Karimunjawa: Departemen Kehutanan Direktorat Jenderal Perlindungan Hutan Dan Konservasi Alam Balai Taman Nasional Karimunjawa, 2025, Accessed on 13 September 2024.
[19] P. Wicaksono, P. A. Aryaguna, and W. Lazuardi, "Benthic habitat mapping model and cross validation using machine-learning classification algorithms," *Remote Sensing*, vol. 11, no. 11, p. 1279, 2019. DOI:10.3390/rs11111279
[20] W. Lazuardi *et al.*, "Coastal reef and seagrass monitoring for coastal ecosystem management," *International Journal of Sustainable Development and Planning*, vol. 16, no. 3, pp. 557–568, 2021. DOI:10.18280/ijsdp.160317
[21] G. Varoquaux and O. Colliot, "*Evaluating machine learning models and their diagnostic value*," New York: Machine Learning for Brain Disorders, pp. 601–630, 2023. DOI:10.1007/978-1-0716-3195-9_20
[22] A. Umardiono, "Pengembangan Obyek Wisata Taman Nasional Laut Kepulauan Karimun Jawa," *J.FISIP*, vol. 24, no. 4, pp. 192–201, 2013.
[23] M. S. Daniar, "Potensi Alam Dan Kepariwisataan Kepulauan Karimunjawa Jepara Provinsi Jawa Tengah Sebagai Medan Pengembangan Olahraga Rekreasi," vol. 53, no. 9, pp. 1689-1699, 2013.
[24] BTNKJ, "Statistik balai taman nasional karimunjawa tahun 2022," 2022, https://tnkarimunjawa.id/publikasi/dokumen, Accessed on 7 June 2024.
[25] I. Fajarini, D. Suryandari, and M. Ihlashu'amal, "Peningkatan perekonomian penduduk melalui pembentukan kelompok sadar wisata (pokdarwis) di desa kemojan kepulauan karimunjawa kabupaten jepara," in *Prosiding Seminar Nasional Kolaborasi Pengabdian Masyarakat UNDIP-UNNES 2019*, pp. 10–12, 2020.
[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32. DOI:10.1023/A:1010933404324
[27] H. I. Choi, "Lecture 10: Random forests," 2017.
[28] H. I. Choi, "Lecture 9: Classification and regression tree (cart)," 2017.
[29] D. Traganos and P. Reinartz, "Mapping mediterranean seagrasses with sentinel-2 imagery," *Marine Pollution Bulletin*, vol. 134, pp. 197–209, 2018. DOI:10.1016/j.marpolbul.2017.06.075
[30] A. Ghorbanian *et al.*, "Mangrove ecosystem mapping using sentinel-1 and sentinel-2 satellite images and random forest algorithm in google earth engine," *Remote Sensing*, vol. 13, no. 13, p. 2565, 2021. DOI:10.3390/rs13132565
[31] Asy'AriRahmat *et al.*, "Mapping mangrove forest distribution on Banten, Jakarta, and West Java Ecotone Zone from Sentinel-2-derived indices using cloud computing based Random Forest," *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management)*, vol. 12, no. 1, pp. 97–111, 2022. DOI:10.29244/jpsl.12.1.97-111
[32] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques (3rd ed.)*. Morgan Kaufmann Publishers, 2012, ISBN:9780123814791.
[33] A. Ng and T. Ma, "Cs229 lecture notes," 2023.
[34] D. Wang *et al.*, "Artificial mangrove species mapping using pléiades-1: An evaluation of pixel-based and object-based classifications with selected machine learning algorithms," *Remote Sensing*, vol. 10, no. 2, p. 294, 2018. DOI:10.3390/rs10020294
[35] Z. Zhao *et al.*, "Comparison of three machine learning algorithms using google earth engine for land use land cover classification," *Rangeland Ecology and Management*, vol. 92, pp. 129–137, 2024. DOI:10.1016/j.rama.2023.10.007
[36] P. Wicaksono *et al.*, "Sentinel-2 images deliver possibilities for accurate and consistent multi-temporal benthic habitat maps in optically shallow water," *Remote Sensing Applications: Society and Environment*, vol. 23, p. 100572, 2021. DOI:10.1016/j.rsase.2021.100572
[37] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016. DOI:10.1145/2939672.2939785
[38] J. Burger, H. J. Boonstra, and J. van den Brakel, "Effect of spatial scale, color infrared and sample size on learning poverty from aerial images," *Remote Sensing Applications: Society and Environment*, vol. 36, p. 101304, 2024. DOI:10.1016/j.rsase.2024.101304
[39] D. Abriha, P. K. Srivastava, and S. Szabó, "Smaller is better? unduly nice accuracy assessments in roof detection using remote sensing data with machine learning and k-fold cross-validation," *Heliyon*, vol. 9, no. 3, p. e14045, 2023. DOI:10.1016/j.heliyon.2023.e14045
[40] J. Goodman, S. Purkis, and J. Phinn, *Coral Reef Remote Sensing: A Guide for Mapping, Monitoring and Management*, 2013. ISBN 978-90-481-9291-5
[41] A. Purwanto and W. Asriningrum, "Identification of mangrove forests using multispectral satellite imageries," *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, vol. 16, no. 1, p. 63, 2019. DOI:10.30536/j.ijreses.2019.v16.a3097
[42] A. S. A. Nugraha and I. P. A. Citra, "Perbandingan metode soil adjusted vegetation index (savi) dan forest canopy density (fcd) untuk identifikasi tutupan vegetasi (kasus; area pembuatan jalan baru singaraja-mengwi)," *Jurnal Geografi Media Informasi Pengembangan dan Profesi Kegeografian*, vol. 18, no. 1, pp. 1–8, 2021. DOI:10.15294/jg.v18i1.25367
[43] A. A. Abubakar *et al.*, *Analysis of vegetation index (ndvi, savi, lai) on coffee productivity in bener meriah district, aceh province*, in *Proceedings of the International Conference on Educational Technology and Social Science (ICoETS 2023)*, pp. 39–45, 2024. DOI:10.2991/978-2-38476-200-2_9.
[44] J. Valdiviezo-N *et al.*, "Built-up index methods and their applications for urban extraction from sentinel 2a satellite data: discussion," *Journal of the Optical Society of America A*, vol. 35, no. 1, pp. 35–44, 2018. DOI:10.1364/JOSAA.35.000035
[45] XGB. Developerss, "Xgboost parameters," https://xgboost.readthedocs.io/en/stable/parameter.html, Accessed on 26 January 2025.
[46] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific reports*, vol. 14, no. 1,m p. 6086, 2024. DOI:10.1038/s41598-024-56706-x