Article

Assessing Derawan Island's Coral Reefs Over Two Decades: A Machine Learning Classification Perspective

Muhammad AL Fadio Ummam , Masita Dwi Mandini Manessa * , Anisya Feby Efriana , Jarot Mulyo Semedi

Posted Date: 21 November 2023

doi: 10.20944/preprints202311.1287.v1

Keywords: Spatial and Temporal Distribution; Coral Reef; Random Forest; Support Vector Machine; Classification and Regression Tree.

Article

Assessing Derawan Island's Coral Reefs over Two Decades: A Machine Learning Classification Perspective

Muhammad AL Fadio Ummam [1], Masita Dwi Mandini Manessa [2,*], Anisya Feby Efriana [3] and Jarot Mulyo Semedi [4]

1 Department of Geography, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok 16424; muhammad.al95@ui.ac.id
2 Department of Geography, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok 16424; manessa@ui.ac.id.
3 Department of Geography, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok 16424; anisya.feby@ui.ac.id
4 Department of Geography, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok 16424; semedi@ui.ac.id
* Correspondence: manessa@ui.ac.id

Abstract: This study provides a comprehensive assessment of Derawan Island's coral reefs over two decades (2003, 2011, and 2021) from a machine learning classification perspective. Employing non-parametric algorithms like Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Tree (CART), our analysis primarily focused on spatial and temporal changes in coral habitats. RF emerged as the most accurate method, demonstrating an accuracy of 71.43% with Landsat 9, 73.68% with Sentinel 2, and 78.28% with Multispectral Aerial Photos. We found that the accuracy of our classification results was significantly influenced by the geographic resolution, as well as the quality of field and satellite/aerial image data. Through spatial clustering, the coral habitats exhibited an Nearest Neighbor Index (NNI) value of 0.8727, indicating specific patterns of distribution. The analysis revealed a decrease in coral reef extent from 2003 to 2011, shrinking to 16 hectares with varying densities, followed by a slight area increase but with more heterogeneous densities between 2011 and 2021. This study not only highlights the dynamic nature of coral reef habitats over two decades but also underscores the critical role of machine learning in environmental monitoring and conservation efforts.

Keywords: spatial and temporal distribution; coral reef; random forest; support vector machine; classification and regression tree

---

1. Introduction

Indonesia is the largest archipelagic country in the world, with a total of 17,000 islands in 2021, according to an agreement on the Island Data Coordination Follow-up Meeting led by the Geospatial Information Agency [1]. Besides that, Indonesia has a coastline of 99,093 km2, a land area of 2.012 million km2, and a sea area of 5.8 million km^2 [2]. Based on data from the Ministry Maritime Affairs and Fisheries, Indonesia has 450 coral reef species out of 700 species in the world [3]. This shows that Indonesia has enormous marine tourism potential, so it can be an opportunity for the Indonesian people to improve their welfare and introduce Indonesia abroad through Indonesia's maritime potential. Marine tourism is a recreational activity carried out with objects and attractions originating from seascapes and coastal landscapes (coastal seascapes) [4]. One of Indonesia's marine tourism destinations, which is a favorite of domestic and foreign tourists, especially world-class divers, is the beauty of the coral reefs in the Derawan Islands.

Derawan Island, one of the Derawan Archipelago, was designated a UNESCO World Heritage Site in 2005 [5]. The main attraction of Derawan Island is the charm of its underwater tourism with the coral reef ecosystem of Derawan Island, which is known to be so beautiful that it rivals the beauty of the coral reefs of the Raja Ampat Islands [6]. Having a well-maintained ecosystem, Derawan Island is home to the natural habitat for various types of protected turtles, like hawksbill turtles and green turtles. According to the Directorate of Coastal and Small Island Utilization, coral reefs are a group of coral animals that have a symbiotic relationship with a type of algae called zooxanthellae. Coral reefs are dynamic ecosystems with rich biodiversity and high levels of productivity [7]. Coral reefs generally live on the coast or in shallow sea areas that are still exposed to sunlight at a depth of approximately 50 meters below sea level. However, several types of coral reefs can live without needing light, but these coral reefs do not have symbiosis with zooxanthellae and do not form coral [8]. In the growth of coral reefs, some factors influence the growth of coral reefs, such as tolerance limits for coral reefs are 21.7–29.6 °C for temperature, 28.7–40.4 psu for salinity [9], these systems can have water clarities ranging from reasonably clear oceanic-type to high-attenuation regimes with complex optical characteristics, sufficient water movement and currents, waters free from pollution, and a hard and compact substrate for coral reefs to attach to [10]. Ecologically, coral reefs are a place for marine biota to find food and provide shelter. Physically, coral reefs protect beaches and shallow water ecosystems from marine abrasion caused by sea waves [11]. Apart from this, coral reefs are also an ecosystem that is vulnerable to environmental change, but these changes can increase in line with the increasing number of tourists and community activities on the coast of Derawan Island.

The challenges of coral reef ecosystems are damage to coral reefs, High levels of pollution, overfishing, harmful fishing methods, and global warming are all significant contributors to coastal damage [12]. Referring to East Kalimantan Governor Regulation Number 60 of 2019 concerning RZWP3K for the Derawan Islands and Surrounding Waters, Derawan Island is the area with the highest damage to coral reefs compared to other areas in the Derawan Islands region [13], which refers to the results of the 2011–2012 Manta Tow survey on the coral around the island. Derawan shows HCL cover of 11-25% and 26-50%, and HCL cover <10% is found mainly on coral reefs near Derawan Island. Based on the data described, it is known that the role of coral reefs is very important in maintaining coastal ecosystems. By looking at the current condition of coral reefs on Derawan Island, a study is needed to monitor the condition of coral reef habitats in a time series on Derawan Island. Monitoring is carried out using remote sensing technology through the use of multispectral imagery. According to [14], remote sensing is a science that can be used to obtain information on objects, areas, or phenomena through the analysis of data obtained without making direct contact with the object being studied. Remote sensing data can be used to identify coastal typologies, one of which is coral reef habitat. Multispectral image data can record the condition of coral reef habitats because it contains visible light, which has a wavelength between 0.4 μm and 0.7 μm. Visible light can penetrate the water column to a depth of 20 meters so that objects at that depth can be recorded [15].

Satellite images that will be used in research on the spatial distribution of coral reef habitat include Landsat 7 ETM and Sentinel-2. Apart from that, in assessing the level of capability of the machine learning classification algorithm for spatial resolution, three spatial resolutions were used: low resolution from Landsat 9 satellite imagery, medium resolution from Sentinel-2 satellite imagery [16], and high resolution from Multispectral Aerial Photography [17]. In the process of monitoring changes in coral reef habitat, a classification stage is needed to identify benthic habitats in the coastal area of Derawan. The classification used is a non-parametric classification algorithm such as Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Tree (CART) [18]. This classification algorithm will be used to determine the level of precision and accuracy of the machine learning classification algorithm for satellite images that have low resolution, medium resolution, and high resolution. Through these stages, an algorithm will be selected that has the highest accuracy, which will be used in conducting mapping studies of the spatial distribution of coral reef habitats.

Remote sensing technologies with machine learning algorithms have been widely used to study changes in coral reefs. In previous research "*Benthic Habitat Mapping Model and Cross Validation Using Machine Learning Classification Algorithm*"[19] regarding the development of benthic habitat models using machine learning classification and applying classification models in several research areas, This research uses in situ integration data on benthic habitats and uses WorldView-2 imagery to parameterize machine learning algorithms, like Random Forest, Classification Tree Analysis, and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 November 2023 doi:10.20944/preprints202311.1287.v1

Support Vector Machine. The results obtained in previous research show that the Random Forest algorithm is more accurate than other algorithms, with a maximum accuracy of 94.17% for 4 classes and 88.54% for 14 classes. Meanwhile, in other research, coral habitat mapping using Sentinel-2 and Landsat 8 Oli imagery can be used [16] and also multispectral aerial photography can be used for habitat mapping. This is possible because Sentinel-2, Landsat 8 Oli, and multispectral aerial photography all have wavelengths between 400 and 700 nm [17]. Nevertheless, this research was carried out to determine the spatial distribution pattern of changes in coral reef habitat in the coastal area of Derawan Island, Berau Regency, East Kalimantan in 2003, 2011, and 2021. As well as assessing the influence of spatial resolution and the level of capability of machine learning classification algorithms in mapping the distribution of reef habitat coral.

In conclusion, while there have been significant advancements in the use of remote sensing technologies and machine learning algorithms for coral reef studies, there remains a gap in comprehensive, long-term analysis, particularly in the context of Derawan Island. Previous studies, such as the 'Benthic Habitat Mapping Model' and others utilizing Sentinel-2, Landsat 8 Oli, and multispectral aerial photography, have demonstrated the effectiveness of these technologies in habitat mapping. However, they have often focused on shorter time frames or specific types of coral habitats. This research seeks to bridge this gap by employing a detailed, long-term study (2003, 2011, and 2021) to understand the spatial distribution and change patterns of coral reef habitats in the Derawan Island coastal area. Furthermore, it aims to assess the impact of varying spatial resolutions and the efficacy of different machine learning classification algorithms in mapping these coral habitats, providing a more nuanced understanding of their temporal dynamics and contributing to more effective conservation strategies.

2. Materials and Methods

2.1. Location

The Derawan Islands are an archipelago located in Berau Regency, East Kalimantan, which has three sub-districts, like Derawan Island, Maratua, and Biduk Biduk. The Derawan Islands have an island that was named a World Heritage site by UNESCO in 2005, namely Derawan Island [5]. The coronation was based on the beautiful panoramic views of the beaches and underwater waters of Derawan Island, where this beauty attracted the attention of world-class diving lovers, to the point that it was nicknamed Pristine Island because of its natural beauty, which is still preserved [20].

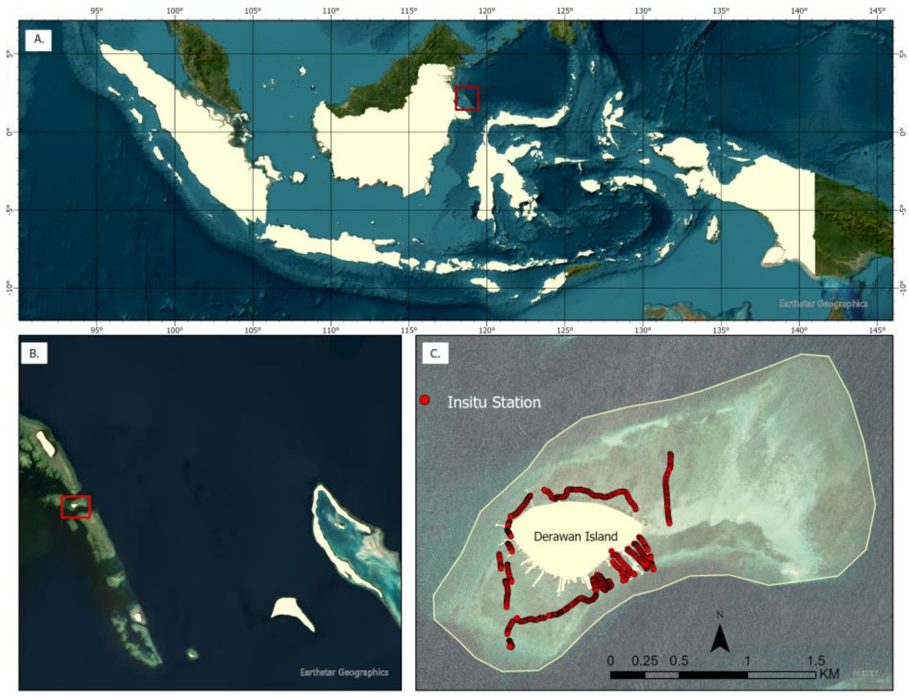

Figure 1. Research Area a.) Indonesia; b.) Derawan Archipelago; c.) Insitu Station in Derawan Island.

Based on data from the NOAA Satellite and Information Service, it appears that sea surface temperatures in East Kalimantan in 2002–2003 ranged between 15°C and 17°C, with bleaching warning conditions in April, June, and November 2002, as well as in April and November 2003. Then, from 2010 to 2011, the sea surface temperature ranged between 15 and 25 degrees Celsius, with a significant change in temperature from 15 degrees in April to 25 degrees in July. This happened for 10 months, which resulted in a level 2 coral bleaching warning, but in 2011, sea surface temperatures tended to be stable. Furthermore, in 2021–2022, the sea surface temperature will range between 15°C and 21°C, whereas in 2020 there will be a coral bleaching warning in May–July with a temperature of 18°C and 21°C. Meanwhile, in 2021, sea surface temperature conditions will only receive coral bleaching warnings in April, June, and October–November [21].

2.2. Data

This study relied on in-situ data, specifically multispectral aerial photography collected on Derawan Island. In situ data gathering includes not just UAV photos but also data from Underwater Photo Transects (UPT). The secondary data used in this study is Landsat-7 satellite imagery from 2003 and 2011, Landsat-9 imagery from 2022, and Monday-2 imagery from 2021 and 2022. Table 1 categorizes the utility of the image data used.

Table 1. Data used on research.

Data	Resolution Spatial	Year	Purpose of Data
Landsat 7	30 m	2003 and 2007	Time Series Mapping
Landsat 9	30 m	2022	Testing the Capabilities of ML Classification Algorithms
Sentinel -2	10 m	2021 and 2022	Time Series Mapping and Testing the Capabilities of ML Classification Algorithms
Multispectral Aerial Photography	8 cm	2021	Testing the Capabilities of ML Classification Algorithms

2.2.1. Landsat-7

Landsat 7 is one of the missions launched by NOAA, NASA, and the USGS. On April 15, 1999, it was launched from Vandenberg Air Force Base in California. The Landsat 7 satellite has an Enhanced Thematic Mapper Plus (ETM+) sensor, which is an advance over the sensors used by earlier Landsat series. In terms of band composition, Landsat 7 Enhanced Thematic Mapper Plus (ETM+) has eight spectral bands with spatial resolutions of 30 meters for Bands 1–7 and 15 meters for Band 8 (panchromatic). All bands in Landsat 7 can collect one of two gain settings (high or low) to increase radiometric sensitivity and dynamic range, while Band 6 collects high and low gain for all data [22].

2.2.2. Landsat-9

The Landsat 9 satellite was launched from Vandenberg Air Force Base in California on September 27, 2021, using a United Launch Alliance Atlas V 401 rocket. The Operational Land Imager 2 (OLI-2) sensor was aboard Landsat 9. According to Brian Markham [23], Landsat 9 contains an OLI-

2 sensor with 9 bands and a TIRS-2 sensor with 2 bands, comparable to Landsat 8. Band 1 collects ultra-blue waves with wavelengths ranging from 0.435nm to 0.451nm and a resolution of 30 meters. Bands 2, 3, and 4 may collect visible waves, particularly blue, green, and red waves, with a resolution of 30 meters in the wavelength range 0.452-0.512 nm, 0.533-0.590 nm, and 0.636-0.673 nm, respectively. Band 5 captures NIR waves with a wavelength range of 0.851-0.879 nm and a resolution of 30 meters. Bands 6 and 7 may catch SWIR waves with wavelengths ranging from 1,566-1,651 nm to 2,107-2,294 nm at a resolution of 30 meters. Band 8 is a band that combines numerous waves to produce a clear black-and-white image; the waves have a wavelength range of 0.503-0.676 nm with a resolution of 15 meters. Band 9 is a band that can catch cirrus cloud objects because of its shortwave range, precisely 1,363-1,384 nm, and a resolution of 30 meters.

2.2.3. Sentinel-2

The Copernicus program's Sentinel-2 satellite was launched in June 2015. Sentinel-2 images have a spatial resolution of 10, 20, and 60 m and two types of bands: one that can catch visible and near-infrared (VNIR) waves and one that can record short wave infrared (SWIR). This satellite can employ 12 different types of bands, 9 of which are in the VNIR band and 3 of which are in the SWIR band [24].

2.2.4. Multispectral Aerial Photography

The capacity of multispectral aerial photography to sharpen the contrasts in hue between two or more objects is its advantage. Tone sharpening can be employed in multispectral aerial pictures for visual observation without alterations, visual observation with reshoots, and additive color blending with viewing tools [14], With a wave range of 0.4 m–0.5 m, the blue channel can capture the reflection of waves that hit the water; hence, it is commonly used for bathymetry, detecting water turbidity, and other water variables. Because the green channel with a wave range of 0.5 m–0.6 m can capture the reflection of plant chlorophyll and the amount of chlorophyll in healthy and sick plants differs, the difference can be easily seen. Make use of this channel. With a wave range of 0.5 m–0.6 m, the Red Channel can be utilized to discriminate between vegetation and non-vegetation objects. The near-infrared (NIR) channel, which has a wave range of 0.7 m–1.1 m, can be used to identify the age of vegetation, the location of glacier layers, and the water content of plants. The SWIR channel can be utilized to identify aerosol and water vapor components, mineral kinds, soil components, and fire potential in dry locations [25].

2.3. Methodology

The data processing has been separated into two parts, the first of which is testing the capabilities of the machine learning classification algorithm on the accuracy and precision of benthic habitat classification produced using Landsat 9 satellite imagery in 2022, Sentinel-2 satellite imagery in 2022, and Multispectral Aerial Photography in 2021. The second part is time-series mapping of the island's coral reef habitat using an algorithm that has the best accuracy based on the results of Machine Learning Classification Algorithm testing using Landsat 7 satellite imagery in 2003 and 2011, as well as Sentinel-2 satellite imagery in 2021. In-situ data is taken directly during field observations in the form of Underwater Photo Transect data and secondary data obtained from the USGS via the Google Earth Engine platform including surface reflectance satellite images of Landsat 7 ETM, Landsat 9, and Sentinel-2.

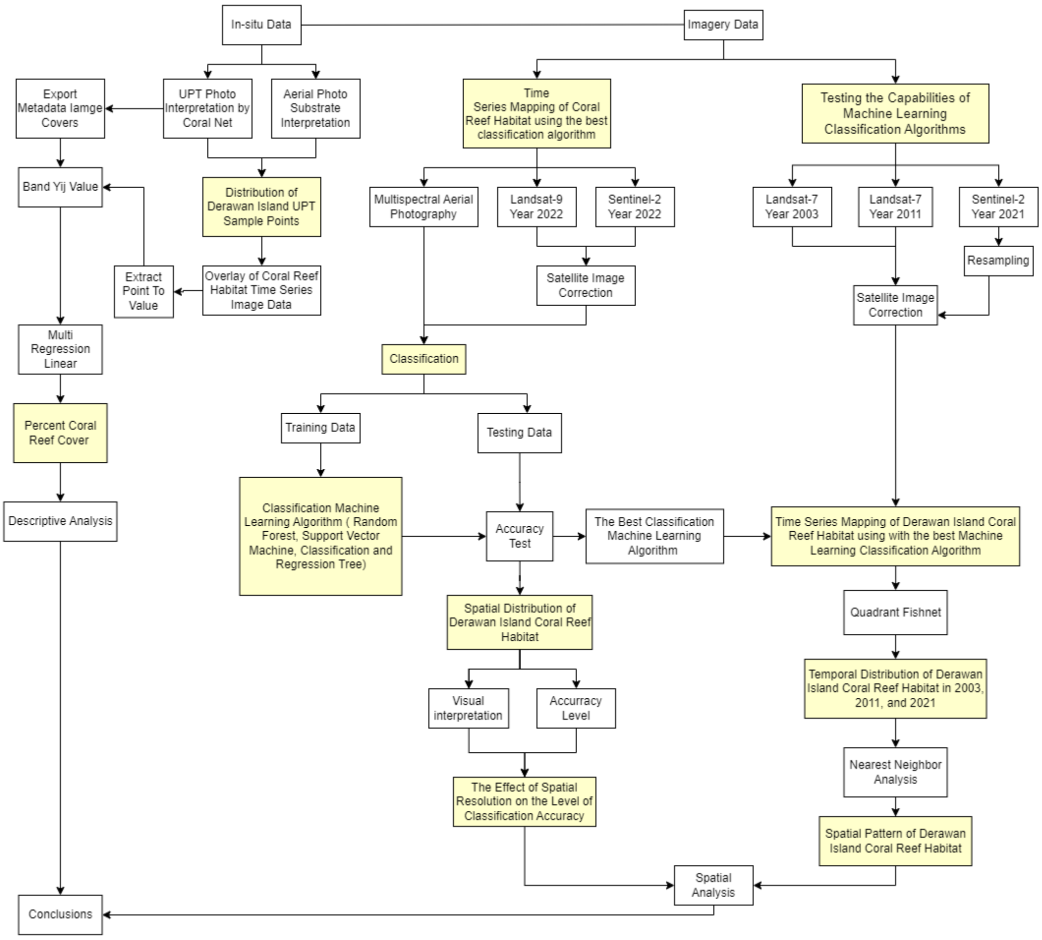

Figure 2. Methodology of research.

Data processing begins with recapitulating and validating Underwater Photo Transect (UPT) data, which is used to compute the percentage of coral reef ecosystem cover. To display the cover image of the UPT photo data, the UPT data is processed using the Coral Net platform. Then, using R software, numerous linear regression computations are performed to collect regression data and reflectance values from objects. The regression data will then be used to determine the percentage of cover for the coral reef habitat, and the reflectance data will be utilized to create raster data on the density of the coral reef habitat. Regression computations in ArcGIS Pro utilizing image reflectance data and Yij band values received from Google Earth Engine are required when creating coral reef habitat density data. The Yij Band was then composited using a raster calculator to create raster data on the density of the coral reef environment.

Following that, secondary data processing is performed on the Google Earth Engine platform to process low-, medium-, and high-resolution satellite image data. The next step is to correct the satellite image, which includes water surface correction and water column adjustment using the Lyzenga algorithm [26], also known as the Depth Invariant Index. Following the rectification of the satellite image, the image is masked to discern between deep sea, shallow sea, and land. Then, benthic habitat classification was performed using three machine learning classification algorithms, namely random forest (RF), support vector machine (SVM), and classification and regression tree (CART), on low-resolution satellite images in the form of Landsat 9 satellite images in 2022, medium-resolution images in the form of Sentinel-2 satellite images in 2022, and high-resolution images in the form of multispectral aerial photos in 2021. The classification findings were then subjected to an accuracy test using the confusion matrix approach so that each classification had its level of accuracy and Kappa value. The most accurate machine learning classification system will then be utilized to map changes in the area of Derawan Island's coral reef habitat. The next step is to rectify satellite pictures used to map changes in coral reef habitat, such as Landsat 7 in 2003 and 2011, and Sentinel-2 in 2021. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 November 2023 doi:10.20944/preprints202311.1287.v1

corrections performed include water surface and water column modifications using the Lyzenga method. Following that, classification is performed using the machine learning classification algorithm with the highest accuracy attained in the previous stage. The results of the 2022 satellite image classification were then validated using the confusion matrix method and the Kappa value.

The following step is to use ArcGIS Pro to build a map showing the pattern of changes in the temporal distribution of Derawan Island's coral reef ecosystem from 2003 to 2021. To create a map of temporal distribution change patterns, the classification results from the Sentinel-2 satellite image were first resampled to 30 meters to offer a balanced study of the resolution of the Landsat 7 ETM+ satellite image. The fishnet tool was then used to generate a quadrant area, which serves as an imaginary line to assist researchers in analyzing patterns of temporal changes in the coral reef ecosystem. Following the formation of the quadrant region, continue evaluating the spatial distribution pattern generated using nearest neighbor analysis to acquire the nearest neighbor index value. The final stage is data analysis using a descriptive statistical approach, to observe changes in the coral reef habitat of Derawan Island over time by comparing the results of coral reef mapping and the spatial distribution density of coral reef habitat in the years 2003, 2011, and 2021.

2.4. Machine Learning Classification

2.4.1. Random Forest

Random forest classification is a classification based on a collection of decision trees to significantly increase the accuracy of pattern recognition[27]. Random forest classification uses a hill-climbing search strategy to find a decision tree that will be the basis for classifying data samples accurately and without errors based on the training data provided[28]. The structure of the random forest classification algorithm is divided into several levels of nodes, namely the root node, branch node, and leaf node. Each class is created using a random vector from samples independently, and each decision tree will provide calculations on the most dominant class units to classify certain classes according to the training data[29].

2.4.2. Support Vector Machine

Support vector machines (SVM) were developed by Boser, Guyon, and Vapnik in 1992 [30] during the Annual Workshop on Computational Learning Theory. SVM is a technique for finding a hyperplane that can separate two sets of data from two different classes. SVM is a technique for making predictions in two cases, namely classification and regression. The basic concept of SVM is a combination of computational theories that have existed in previous years. The hyperplane is the best dividing line between the two classes, which can be found by measuring the hyperplane margin and looking for the maximum point[31]. The margin is the distance between the hyperplane and the closest pattern from each class, which is called the support vector. The largest margin can be found by maximizing the distance value between the hyperplane and its closest point.

2.4.3. Classification and Regression Tree

Classification and Regression Tree (CART) is an algorithmic development of the decision tree technique developed by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. The development carried out is in the form of using a binary recursive partitioning algorithm [32]. CART will produce a classification tree if the response variable has a categorical scale and a regression tree if the response variable is continuous data. The aim of using CART is to carry out classification analysis in the sector of response variables, whether nominal, ordinal, or continuous. The CART method is divided into two methods, namely the classification tree method and the regression tree method. If the dependent variable has a categorical type, CART will produce a classification tree. Meanwhile, if the dependent variable is numeric or continuous, CART will produce a regression tree.

The results of benthic habitat classification using Landsat 7 satellite imagery in 2003 on Derawan Island shown in Figure 7 show that the coastal area of Derawan Island in 2003 was dominated by seagrass areas with an area of 119.33 ha and covered 34.13% of the benthic habitat on Derawan Island

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 November 2023 doi:10.20944/preprints202311.1287.v1

in 2003. Meanwhile, with an area of 111 ha and a percentage of 31.78%, the dominating coral reef substrate is on the rim area of Derawan Island. Table 2, which presents a tabulation of the benthic habitat area of Derawan Island in 2003, reveals that the mixed zone only has a size of 6.6 ha and a percentage of 1.91%.

Table 2. Percentage of benthic habitat area on Derawan Island in 2003.

No	Class	Wide (Ha)	Percentage (%)
1	Mixed	6.67	1.91%
2	Coral	111.12	31.78%
3	Sand/Rubble	112.5	32.18%
4	Seagrass	119.33	34.13%

3. Results

3.1. Highest Overall Accuracy Validation Capability of Machine Learning Classification Algorithms

The accuracy test results on Landsat 9 images and Sentinel-2 images showed that Random Forest had the best level of accuracy, with accuracy values of 71.43% and 73.68%, respectively, based on the results of benthic habitat classification processing in 2022 on Landsat 9, Sentinel-2, and Multispectral Aerial Photography satellite images. Multispectral aerial photography, on the other hand, demonstrates that the support vector machine classification technique is the best, with an accuracy value of 78.28%.

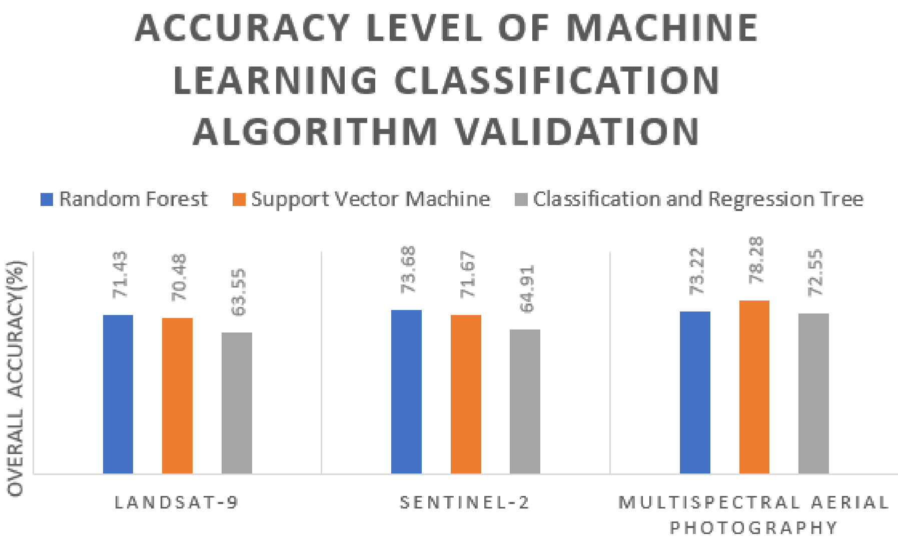

Figure 3. Accuracy level of the classic machine learning algorithm.

Random Forest algorithm classification results on Landsat 9 and Sentinel-2 satellite photos are more relevant and acceptable than the Support Vector Machine method classification results on multispectral aerial photographs. This is because the classification results in multispectral aerial photography have a significant bias, and field data validation reveals that the substrate classification is significantly off in several places. Figure 4, depicts two examples of bias on Derawan Island's coastline areas.

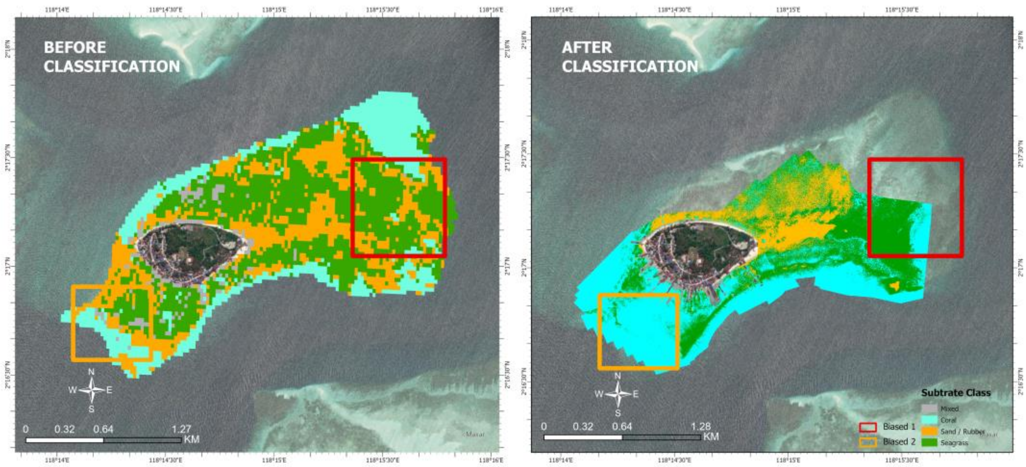

Figure 4. Comparison of Multispectral Aerial Photos before classification and after classification.

According to Figure 4, the sand substrate is detected as a coral reef substrate in the Red Box area, while the seagrass substrate is detected as a coral reef substrate in the Orange Box area. This is due to the resolution disparity between multispectral aerial photography and point data collecting techniques such as handheld GPS. The resolution discrepancy is related to the difficulties of obtaining field sample point data in the center of seawater, as well as the existence of ocean wave forces, which can cause shifts in field sample points. As a result, the ensuing classification results are biased, resulting in classification mistakes, and the categorization of benthic habitats generated using multispectral aerial photography data cannot be used to explain the observed occurrences.

3.2. The Effect of Spatial Resolution on the Level of Classification Accuracy

In general, the interpretation of object appearances in Landsat 9 image data and Sentinel-2 image data is almost identical, and in one image capture area, on average, only 1 to 3 objects are identified, and differences or delineations between one object and another object can only be seen well at a low scale, so the detail of object interpretation is low. This is due to the low spatial resolution of the Landsat 9 satellite image and the medium spatial resolution of the Sentinel-2 satellite image, both of which have spatial resolutions of 10 meters and 30 meters, respectively, where one pixel of image data represents a 10 x 10 or 30 x 30 meters area.

The results of object interpretation using photos obtained by multispectral aerial photography are substantially more variable than the findings from Landsat 9 and Sentinel-2 satellite data, as evidenced by one of the object interpretation data captures. By detecting more things and noticing differences between them, there can be a more accurate picture of what the objects in the field look like. That could be happening because the multispectral aerial photography's ground sampling distance (GSD) is getting close to 8 cm/px. According to the results of comparing the object interpretations of the three image data sets above, there are considerable changes in spatial resolution that affect how well objects are classified using the resulting classification approach. High spatial resolution data is able for data analysis needing fine detail since it can increase the degree of class variation and delineation of collected objects, resulting in the categorization and interpretation of exceedingly different data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 November 2023 doi:10.20944/preprints202311.1287.v1

Figure 5. Comparison of the level of accuracy of random forest classification with image data.

3.3. Spatial Pattern of Distribution of Derawan Island Coral Reef Habitat

Based on the identification of spatial patterns using Nearest Neighbor Analysis, the results demonstrate that the Nearest Neighbor Index (NNI) value obtained in Figure 5.8 is 0.8727, indicating that the NNI value achieved is 1. This demonstrates that the spatial structure of the coral reef habitat developed on Derawan Island is concentrated in the edge area, where the coral reef habitat is in places that are adjacent to each other and tend to cluster.

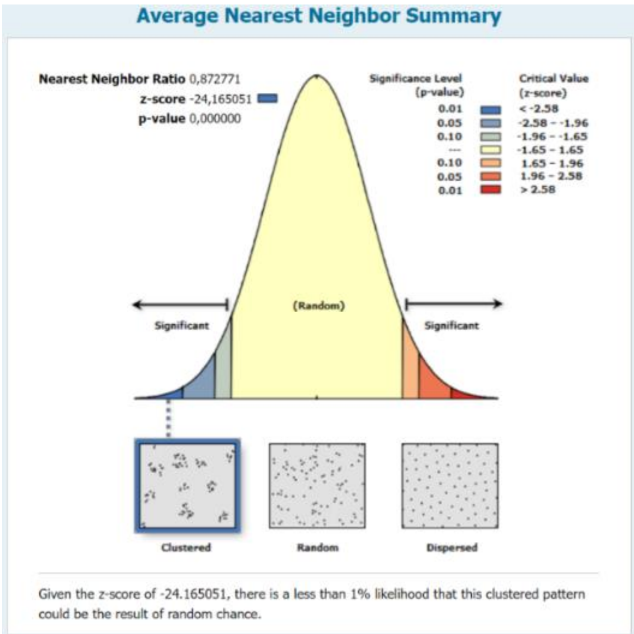

Figure 6. Nearest neighbor analysis results.

4. Discussion

4.1. Spatial Distribution and Habitat Density of Derawan Island Coral Reefs

Our evaluation revealed that the Random Forest classification algorithm exhibited the highest accuracy, a result attributable to its proficiency in handling large-scale data. In this study, the benthic habitat is categorized into four types: mixed, coral reefs, seagrass, and sand or rubble. Subsequent to this classification, we conducted coral reef habitat mapping using data from four satellite images: Landsat 7 imagery from 2003 and 2011, and Sentinel-2 imagery from 2021. This mapping forms the basis for an in-depth analysis of the regional and temporal variations in the coral reef habitat of Derawan Island, focusing on changes in size and composition over the studied periods

 doi:10.20944/preprints202311.1287.v1

The results of the benthic habitat classification using Landsat 7 satellite imagery in 2003 on Derawan Island, as shown in Figure 7, are based on the data processing that was done. These results indicate that the coastal area of Derawan Island in 2003 was dominated by seagrass areas, with an area of 119.33 ha and covering 34, 13% of the benthic habitat on Derawan Island in 2003. The Derawan Island rim area, which has a 111 Ha area and a percentage of 31.78%, is home to the majority of the coral reef substrate. With a total area of 112.5 ha and a proportion of 32.18%, the sand/rubble area comes in second. Table 2, which presents a tabulation of the benthic habitat area of Derawan Island in 2003, reveals that the mixed area only has a size of 6.6 ha and a percentage of 1.91%.

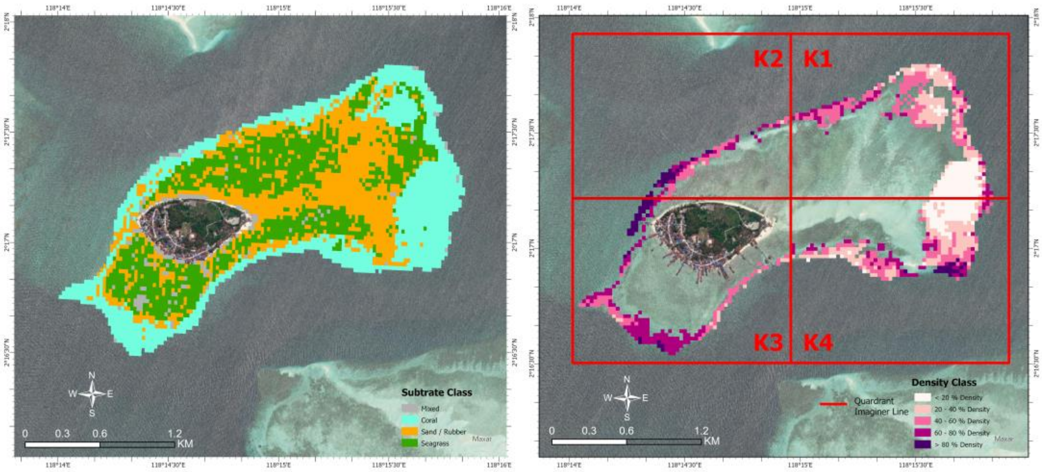

Figure 7. Distribution of benthic habitat on Derawan Island in 2003.

According to the data processing done, Figure 8's results of the benthic habitat classification using Landsat 7 satellite imagery in 2011 on Derawan Island reveal that the coastal area of Derawan Island was still largely dominated by seagrass beds in 2011, with an area of 175 ha and a percentage of 50.20%. The Derawan Island rim area, which has a 95.8 ha area and a percentage of 27.40%, is where the majority of the coral reef substrate is located. Afterward, the coral reefs suffered a 15% to 20% decline, leaving 95.8 ha of the eastern section with a decline. Then the sand/rubble area also experienced a decrease in habitat area, reaching more than 30 ha, and the mixed class only had a total area of 10.6 ha with a percentage of 3%, as shown in Table 3, which shows a tabulation of the benthic habitat area of Derawan Island in 2011.

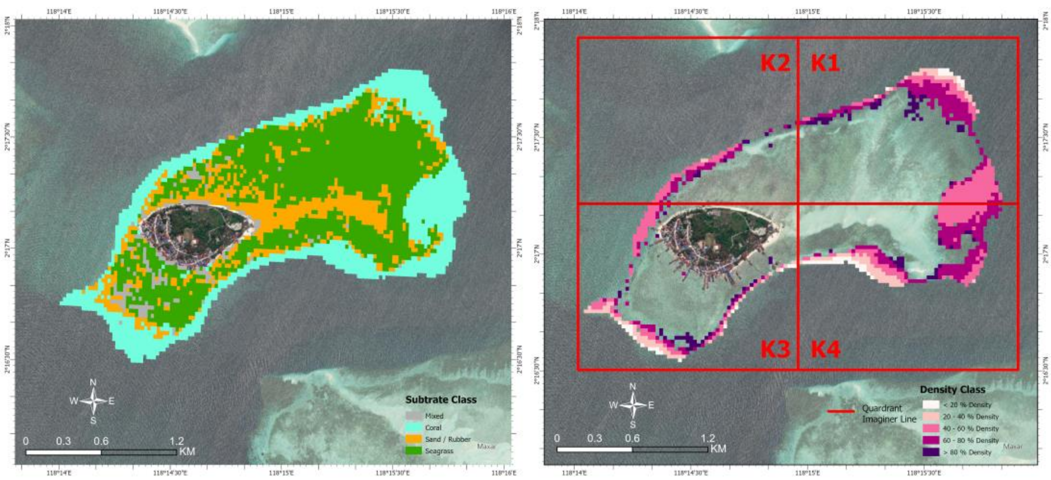

Figure 8. Distribution of benthic habitat on Derawan Island in 2011.

Table 3. Percentage of benthic habitat area on Derawan Island in 2011.

No	Class	Wide (Ha)	Percentage (%)
1	Mixed	10.63	3.04%
2	Coral	95.81	27.40%
3	Sand/Rubble	67.67	19.36%
4	Seagrass	175.52	50.20%

Based on the data processing, the conclusions of the benthic habitat classification were obtained from 2021 Sentinel-2 satellite photographs on Derawan Island, as shown in Figure 9. The classification that is formed is resampled by pixels first to 30 meters so that it can be relevant in analyzing changes in the area with Landsat 7 ETM+ satellite imagery. According to the results of the benthic habitat classification, the coastline area of Derawan Island will still be dominated by seagrass beds in 2021, but with an increase in area of 167.6 hectares and a percentage of 48.02%. Meanwhile, the coral reef substrate also experienced an increase in the area of 16 ha with a total area of 110.5 ha with a percentage of 31.68%, where there was an increase in the eastern region of Derawan Island. Then class sand/rubble substrate shows an area of 56.1 ha with a percentage of 13.24%. Then the mixed substrate only has a total area of 14.7 ha with a percentage of 4.40%, as shown in Figure 9 and Table 4, which gives information on a tabulation of the benthic habitat area of Derawan Island in 2021.

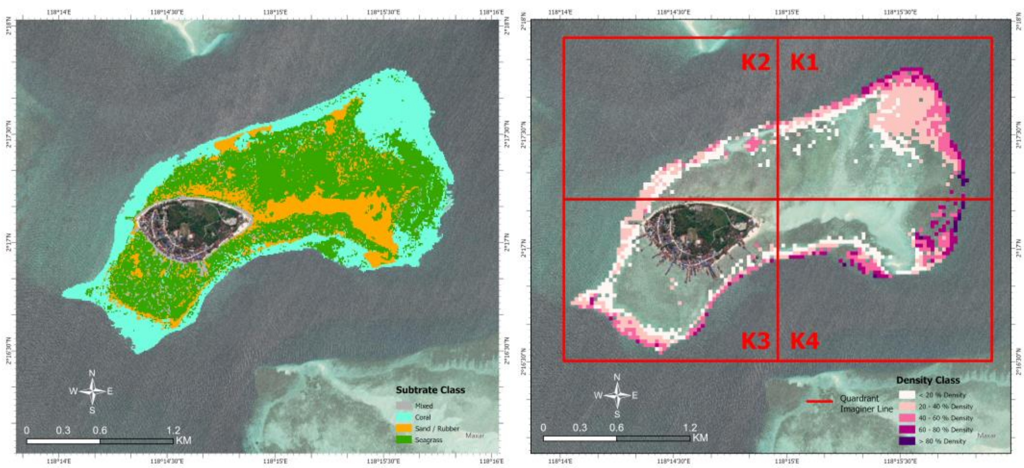

Figure 9. Distribution of benthic habitat on Derawan Island in 2021.

Table 4. Percentage of benthic habitat area on Derawan Island in 2021.

No	Class	Wide (Ha)	Percentage (%)
1	Mixed	14.77	4.40%
2	Coral	110.57	31.68%
3	Sand/Rubble	57.09	13.24%
4	Seagrass	167.58	48.02%

It is known that the Random Forest classification algorithm on Sentinel-2 satellite images provides the level of accuracy in benthic habitat classification based on the accuracy test results given in Figure 10. The accuracy test value for benthic habitats has a minimum accuracy value of 60%, according to SNI 7716:2011, established by the National Standardization Agency. With a validation accuracy value of 64.21% and a validation Kappa value of 0.44, the accuracy test results demonstrate that accuracy is satisfactory. The coral reef substrate generated in the accuracy validation has a producer accuracy of 52.63% and a user accuracy of 83.33%. In the accuracy validation, the accuracy level of the sand/rubble substrate formed was 75% with a producer's accuracy and 54.55% with a user's accuracy. Furthermore, accuracy validation was used to determine the accuracy of the seagrass substrate, with a producer's accuracy value of 62.5% and a user's accuracy value of 83.33%. Then the mixed substrate has a level of accuracy that is formed in accuracy validation with a producer's accuracy value of 75% and a user's accuracy of 21.43%.

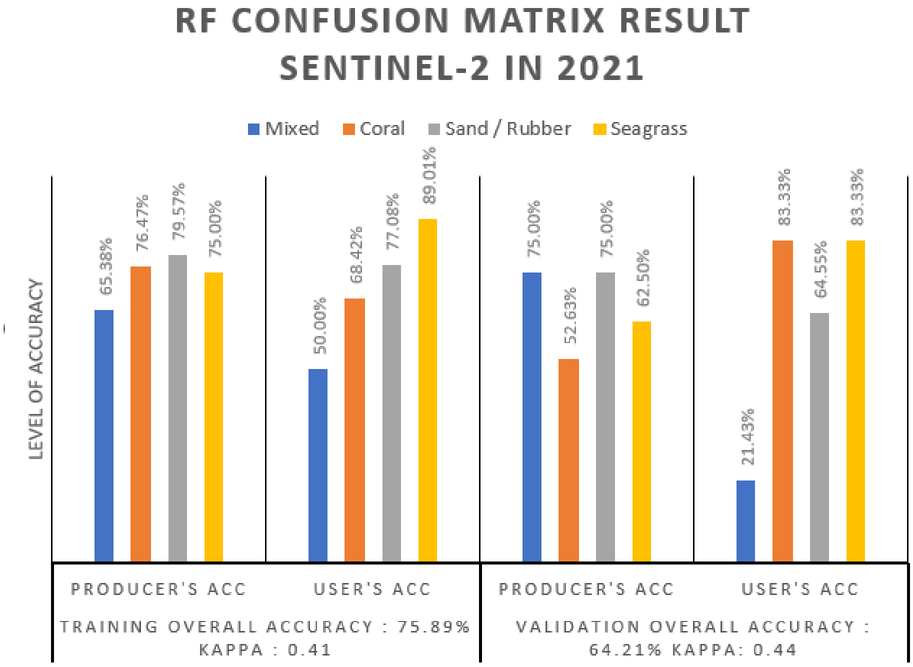

Figure 10. Random Forest Confusion Matrix Result Sentinel-2 in 2021.

4.2. Temporal Pattern of Changes in Derawan Island Coral Reef Habitat in 2003, 2011, and 2021

According on Figure 11, Pattern of Change in the Spatial Distribution of Benthic Habitat on Derawan Island in 2003–2021, it is shown that in analyzing the pattern of change in the distribution of benthic habitat, an imaginary line is used in the form of a quadrant area, which is divided into four areas with regional characteristics, such as Quadrant 1 and Quadrant 4, which are areas that are more influenced by natural factors, while Quadrant 2 and Quadrant 3 are more influenced by human activity factors, with each quadrant area being 200 ha. The division of the reference area using imaginary lines aims to make it easier to analyze patterns of changes in the distribution of benthic habitat on Derawan Island from 2003 to 2021.

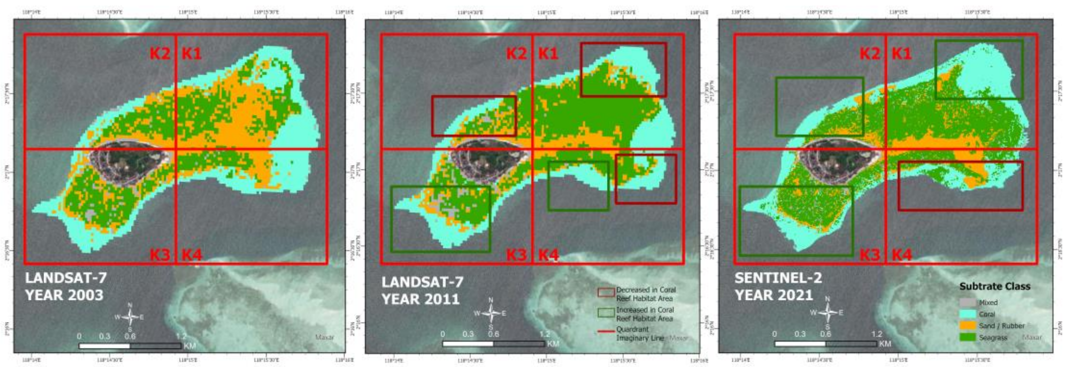

Figure 11. Changes in the benthic habitat of Derawan Island in 2003, 2011, 2021.

According to the conclusions drawn from the findings of the Derawan Island benthic habitat classification, the complete distribution of coral reef habitat is in the reef crest zone, also known as the edge area. The green polygon symbol represents places where the coral reef habitat area has increased, whereas the red polygon symbol represents areas where the coral reef habitat area has decreased. From 2003 to 2011, there was a significant deterioration of coral reef habitat in Quadrants 1 and 4, with an area drop of 15.3 hectares. Natural factors such as current movements and sea surface

doi:10.20944/preprints202311.1287.v1

temperature have a greater influence on changes in Quadrants 1 and 4. Based on the current data, it is clear that the current moves from east to west, causing changes in sea surface temperature over time. Then, from 2003 to 2021, the Quadrant 3 and Quadrant 2 areas experienced a decrease in a small portion of the distribution of coral reef habitat; the decrease occurred on the reef slope, although the Quadrant 3 and Quadrant 2 areas are areas closer to Derawan Island and human activities with gentle slopes.

Meanwhile, from 2011 to 2021, in general, there was an increase in coral reef habitat in Quadrant 1, Quadrant 3, and Quadrant 2, where the condition of coral reef habitat has begun to return to normal with a lower density, and there are several coral areas located outside the barrier. in the period 2011–2021. This increase occurred entirely in the reef crest, or tuber, which allows many types of coral to grow, increasing the distribution of Derawan Island's coral reef habitat. The increase in the area of coral reef habitat that occurred in the period 2011–2021 was 14.76 ha, and this value resembles the area of coral reef habitat on Derawan Island in 2003. However, only in the Quadrant 4 region does the area of coral reef habitat continue to decrease. If we look at the regional characteristics in Quadrant 4, changes in coral reef habitat tend to be influenced by natural factors, namely current movements and changes in sea surface temperature. From statistical data on the temporal distribution of benthic habitat obtained from data processing, it can be seen that a reduction in distribution continues to occur on sand and rubble substrates. Meanwhile, the distribution of coral reef substrates from 2011 to 2021 continues to experience an increase in area of up to 15 ha as well as mixed substrates of up to 4 ha. Then, for seagrass substrate, there was an increase in 2003–2011, but a decrease occurred in the period 2011–2021, reaching 8 ha.

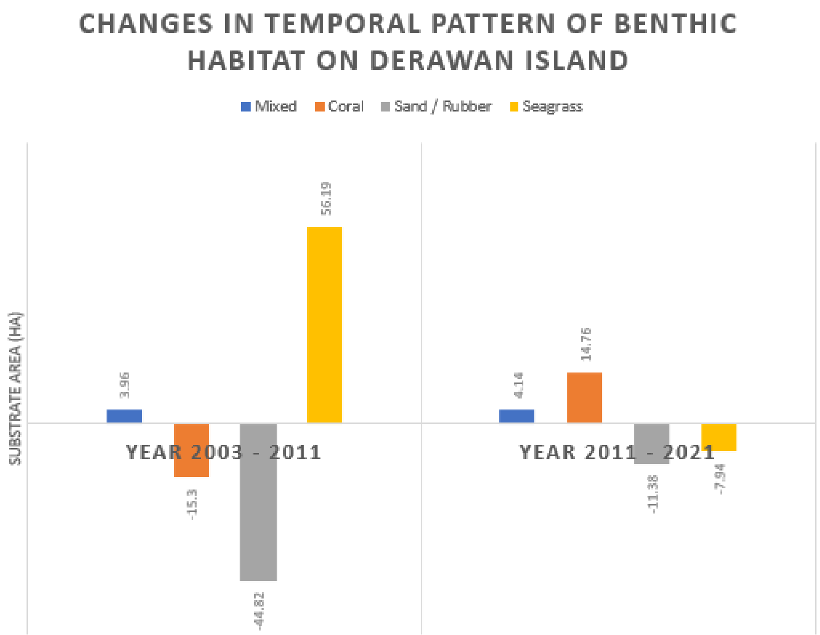

Figure 12. Graph of changes in the temporal distribution of benthic habitat on Derawan Island in 2003, 2011, 2021.

5. Conclusions

The study's analysis of the spatial and temporal distribution of coral reef environments on Derawan Island establishes the superiority of the Random Forest classification method. This conclusion is based on the accuracy of Random Forest in classifying the benthic ecosystem, where it notably surpassed the Support Vector Machine and Classification and Regression Tree algorithms. Furthermore, our findings confirm that the spatial resolution used, along with the quality of field and satellite/multispectral aerial image data, significantly influences classification accuracy.

In terms of spatial patterns, a clustered formation was observed along the edges of Derawan Island. Temporally, the study indicates a decline in both the area and density of coral reef habitats across all regions of the reef crest from 2003 to 2011. However, from 2011 to 2021, there was an overall increase in the area of coral reef habitats throughout Derawan Island, except in Quadrant 4. Despite this increase in area, the density of the coral reef habitats showed a general decline across all regions.

Author Contributions: Concept: MDMM and JMS; Data Collection: MDMM, MAFU, and AFE, Data Processing: MDMM, and MAFU; Writing and original draft preparation: MDMM and MAFU; Writing—review and editing: MDMM and AFE

Data Availability Statement: The datasets are unavailable due to privacy and ethical restrictions.

Acknowledgments: This research was funded by the University of Indonesia through the International Indexed Publication Grant (PUTI) Q2 with the contact number on the grant NKB-751/UN2.RST/HKP.05.00/2023 entitled Multi-level Spatial Analysis for Time Series Coral Reef Density Mapping.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. K. K. dan Perikanan, "Jumlah Pulau Indonesia," 2021. https://kkp.go.id/djprl/p4k/page/4270-jumlah-pulau.
2. M. Ramdhan and A. Taslim, "Aplikasi Sistem Informasi Geografis dalam Penilaian Proporsi Luas Laut Indonesia," *J. Ilm. Geomatika*, vol. 19, no. 2, p. 141, 2013.
3. Dewan Perwakilan Rakyat RI, "Academic Manuscript of the Bill Concerning Conservation of Biological Resources and Their Ecosystems," p. 173, 2017.
4. F. Chafid, *Dasar-dasar manajemen kepariwisataan alam*. Yogyakarta, 1995.
5. UNESCO, "World Heritage," *Conv. Concering Protrction World Curtural Nat. Herit.*, vol. 41, no. 09, pp. 41-5055-41–5055, 2005, doi: 10.5860/choice.41-5055.
6. Kementerian Pariwisata dan Ekonomi Kreatif RI, "Eksotisme Kepulauan Derawan, Surga Tersembunyi yang Menawan," 2022. https://www.indonesia.travel/id/id/destinasi/kalimantan/derawan-archipelago/eksotisme-kepulauan-derawan-surga-tersembunyi-yang-menawan.
7. S. J. Brandl *et al.*, "Coral reef ecosystem functioning: eight core processes and the role of biodiversity," *Front. Ecol. Environ.*, vol. 17, no. 8, pp. 445–454, Oct. 2019, doi: 10.1002/fee.2088.
8. L. N. T. Bara'langi', S. Yusuf, C. Rani, A. A. A. Husain, J. Tresnati, and A. Tuwo1, "Zooxanthellae Density in Different Zone and Life Form in Inner and Outer Zone of Spermonde Islands," *J. Ilmu Kelaut. SPERMONDE*, vol. 7, no. 1, pp. 27–35, 2021.
9. Y. Guan, S. Hohn, and A. Merico, "Suitable Environmental Ranges for Potential Coral Reef Habitats in the Tropical Ocean," *PLoS One*, vol. 10, no. 6, p. e0128831, Jun. 2015, doi: 10.1371/journal.pone.0128831.
10. Russell, Dierssen, and Hochberg, "Water Column Optical Properties of Pacific Coral Reefs Across Geomorphic Zones and in Comparison to Offshore Waters," *Remote Sens.*, vol. 11, no. 15, p. 1757, Jul. 2019, doi: 10.3390/rs11151757.
11. S. A. Saputra, I. P. Yuda, and F. Zahida, "Keanekaragaman dan penutupan terumbu karang di pantai Pasir Putih, Jawa Timur," 2016.
12. O. D. Elisha and M. J. Felix, "Destruction of coastal ecosystems and the vicious cycle of poverty in Niger Delta Region," *J. Glob. Agric. Ecol.*, vol. 11, no. 2, pp. 7–24, 2021.
13. K. Perairan, D. A. N. Pulau, and P. Kecil, "Identifikasi dan pemetaan zonasi kawasan konservasi perairan dan pulau - pulau kecil (kkp3k) kabupaten kutai kartanegara."
14. T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation, 7th Edition*. Wiley, 2015.
15. C. D. Prawoto and H. Hartono, "Pemetaan Habitat Bentik dengan Citra Multispektral Sentinel-2A di Perairan Pulau Menjangan Kecil dan Menjangan Besar, Kepulauan Karimunjawa," *J. Bumi Indones.*, vol. 7, no. 3, 2018.
16. J. D. Hedley *et al.*, "Coral reef applications of Sentinel-2: Coverage, characteristics, bathymetry and benthic mapping with comparison to Landsat 8," *Remote Sens. Environ.*, vol. 216, no. October 2017, pp. 598–614, 2018, doi: 10.1016/j.rse.2018.07.014.
17. J. Brodie, L. V. Ash, I. Tittley, and C. Yesson, "A comparison of multispectral aerial and satellite imagery for mapping intertidal seaweed communities," *Aquat. Conserv. Mar. Freshw. Ecosyst.*, vol. 28, no. 4, pp. 872–881, Aug. 2018, doi: 10.1002/aqc.2905.
18. C. Zhang, "Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 213–223, Jun. 2015, doi: 10.1016/j.isprsjprs.2014.06.005.
19. P. Wicaksono, P. A. Aryaguna, and W. Lazuardi, "Benthic Habitat Mapping Model and Cross Validation Using Machine-Learning Classification Algorithms," *Remote Sens.*, vol. 11, no. 11, p. 1279, May 2019, doi: 10.3390/rs11111279.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 November 2023 doi:10.20944/preprints202311.1287.v1

20. "Ministry of Tourism and Creative Economy," 2022. https://www.indonesia.travel/id/id/destinasi/kalimantan/derawan-archipelago/eksotisme-kepulauan-derawan-surga-tersembunyi-yang-menawan.
21. NOAA, "Pacific Climate Update Coral Bleaching Heat Stress Analysis and Seasonal Guidance through September 2022," 2022. https://coralreefwatch.noaa.gov/satellite/analyses_guidance/pacific_cbts_ag_20220531.php.
22. M. D. King and S. Platnick, "The Earth Observing System (EOS)," in *Comprehensive Remote Sensing*, Elsevier, 2018, pp. 7–26.
23. B. L. Markham *et al.*, "Landsat 9: status and plans," Sep. 2016, p. 99720G, doi: 10.1117/12.2238658.
24. H. Park, J. Choi, N. Park, and S. Choi, "Sharpening the VNIR and SWIR bands of Sentinel-2A imagery through modified selected and synthesized band schemes," *Remote Sens.*, vol. 9, no. 10, pp. 1–20, 2017, doi: 10.3390/rs9101080.
25. Z. Wu *et al.*, "User needs for future Landsat missions," *Remote Sens. Environ.*, vol. 231, p. 111214, Sep. 2019, doi: 10.1016/j.rse.2019.111214.
26. L. D. R., "Passive remote sensing techniques for mapping water depth and bottom features.," *Appl. Opt.*, vol. 17, p. 379, 1978.
27. L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
28. A. D. Purwanto, A. Ibrahim, A. Ulfa, E. Pawati, and A. Supriyono, "Pengembangan Model Identifikasi Habitat Bentik Menggunakan Pendekatan Segmentasi Object-Based Image Analysis (OBIA) dan Algoritma Machine Learning (Studi Kasus: Pulau Pari, Kepulauan Seribu)," *Pus. Ris. Kelaut.*, vol. Vol 17, No, 2022.
29. A. D. Kulkarni and B. Lowe, "Random Forest Algorithm for Land Cover Classification," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 4, no. 3, pp. 58–63, 2016.
30. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, Jul. 1992, pp. 144–152, doi: 10.1145/130385.130401.
31. E. Byvatov and G. Schneider, "Support vector machine applications in bioinformatics.," *Appl. Bioinformatics*, vol. 2, no. 2, pp. 67–77, 2003.
32. R. J. Lewis, D. Ph, and W. C. Street, "An Introduction to Classification and Regression Tree (CART) Analysis," *2000 Annu. Meet. Soc. Acad. Emerg. Med.*, no. 310, p. 14p, 2000, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.4103&rep=rep1&type=pdf.