IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Marine Habitat Mapping using Multibeam Echosounder Survey and Underwater Video Observations: A Case Study from Tioman Marine Park

To cite this article: Muhammad Abdul Hakim Muhamad *et al* 2023 *IOP Conf. Ser.: Earth Environ. Sci.* **1240** 012006

IOP Conf. Series: Earth and Environmental Science 1240 (2023) 012006 doi:10.1088/1755-1315/1240/1/012006

Marine Habitat Mapping using Multibeam Echosounder Survey and Underwater Video Observations: A Case Study from Tioman Marine Park

Muhammad Abdul Hakim Muhamad[1*], Rozaimi Che Hasan[2*], Najhan Md Said[3], Mohd Shahmy Mohd Said[1], Raiz Razali[1]

[1] School of Geomatics Science and Natural Resources, College of Built Environment, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia

[2] Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

[3] Eastern Fleet Command Headquarters, Kota Kinabalu Naval Base, Teluk Sepangar, Kota Kinabalu, Sabah

abdulhakim@uitm.edu.my
rozaimi.kl@utm.my

Abstract. In recent years, there has been an increasing trend of utilizing high-resolution multibeam echosounder (MBES) datasets and supervised classification via machine learning to create marine habitat maps. The purpose of current study was threefold: (1) to extract bathymetric and backscatter derivatives from a multibeam dataset, (2) to measure the correlation between bathymetric and backscatter derivatives, and (3) to generate a marine habitat map using the Random Forest (RF). Tioman Marine Park (TMP), which is situated Southeast China Sea. MBES surveyed area are encompassed an area of 406 km² and served as the location for the study. Based on results and analysis, fourteen (14) derivative were derived from bathymetry map and backscatter mosaic. The second step involved integrating variables and a total of 152 of habitat ground-truth data were used, derived from underwater imageries, and sediment samples, into an RF model to generate a map of the marine habitat. Based on marine habitat map, six habitat classes including sand, rock, gravel and sand, coral rubble, coral and rock, and coral were classified. The distribution of coral habitat was found to be correlated with the depth of the bathymetry in the shallow water region. Therefore, the study has reached the conclusion that the integration between MBES derivatives, ground-truth data, and RF machine learning algorithm is an effective in classifying the distribution of marine habitats, specifically the coral habitat.

1. Introduction

Marine conservation efforts are heavily dependent on the availability of accurate and comprehensive data on marine habitats, particularly for vulnerable ecosystems like coral reefs [1]. Despite the richness of marine habitats, including corals, in Malaysia, the lack of comprehensive data collection using conventional techniques has resulted in an unfortunate situation where high-resolution spatial distribution information is not fully available. The data available to date have been primarily collected through traditional methods such as diving and transect lines, which are not as accurate and comprehensive as modern technologies. This lack of data is particularly apparent in Malaysia's marine parks, including the Tioman Marine Park (TMP) in Pahang.

In 1994, the TMP, comprising of nine islands, was established as a Marine Park through official proclamation. The area's significance lies not only in its contribution to the marine ecosystem but also in its economic value, owing to the abundance of diverse soft and hard coral species [2]. As a result, it is imperative to acquire baseline spatial mapping data for systematic conservation and planning approaches aimed at monitoring the condition and wellbeing of coral reef habitats. Thus, examining and evaluating the physical features of the seabed is crucial for managing coral reef and fisheries ecosystems and marine geology [3].

Marine habitat mapping has been greatly influenced by the development of advanced marine acoustic survey techniques in recent years, which have made it possible to produce highly precise seabed maps [4]. The application of a high-resolution acoustic approach, such as the multibeam echosounder system (MBES), plays a crucial role in predicting sediment and habitat types by providing comprehensive coverage of topography produced from bathymetry data, simultaneously sediment composition produced from intensity returns [5]. The backscatter data, intensity returns obtained through the use of MBES are considered a valuable dataset in acoustic analysis. This technique provides significant information about the scattering properties of different sediment types and holds great potential for remote identification of the seabed and classification of habitat types. MBES are widely utilized for this objective, primarily due to their technological features (such as wide coverage and high-resolution data) that exceed the capabilities of all other current systems, including single beam echosounders, side scan sonars, and Light Detection and Ranging (LiDAR) [6].

Over the past decades, several image analysis have been introduced to classify MBES data and create maps of marine habitats [7]. The classification methods exhibit considerable diversity concerning the data characteristics utilized for classification, including those obtained from MBES data, and the image classification that are employed [8]. Despite the multitude of techniques documented in literature, there is still a requirement for developing sturdy methods that provide better classification and prediction rates and yield more precise marine habitat maps.

The random forest (RF) algorithm was able to perform classification and regression tasks in the realm of machine learning [9]. The way it operates is by generating numerous decision trees, with single tree being trained on a random subset of the data. [10, 11]. When making a prediction, the algorithm aggregates the predictions of all the individual trees to arrive at a final prediction [12]. RF is often used in acoustic MBES applications, such as habitat mapping, because it can handle large amounts of data and can effectively model complex relationships between environmental variables and habitat types [13]. Using MBES with this algorithm able to create a habitat classification is an effective method for mapping benthic habitats. It also can offer significant insights for managing and conserving marine ecosystems.

The study aims to obtain a comprehensive acoustic sonar analysis of the seabed in TMP by utilizing the Multibeam Echosounder System (MBES). The study's objectives are as follows; (1) derive bathymetric and backscatter derivatives; (4) assess the correlation between MBES derivatives; and (3) develop a marine habitat map of TMP using acoustic data, ground-truth data, and machine learning techniques.

2. Method

2.1 Study Area

A total area of 406 square kilometers, study area is located in Tioman Marine Park, Pahang and comprises numerous islands, such as Tioman, Tulai, Sepoi, Labas, and Chebeh, with Tioman being the main island. Figure 1 provides a representation of the location of certain islands within the study area, with a range of up to 2 nautical miles from the shoreline.

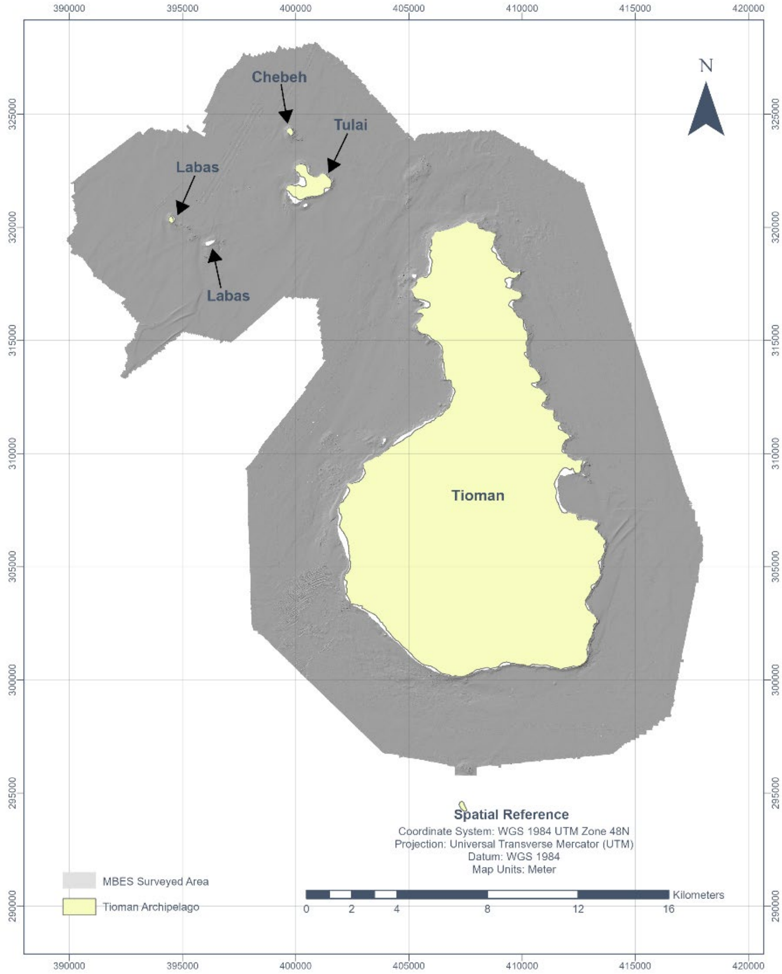

Figure 1. The study was conducted in Tioman Marine Park (TMP), which consists of five archipelagos including Tioman and four smaller islands, namely Tulai, Sepoi, Labas, and Chebeh.

2.2 Bathymetry and Backscatter Data: MBES Survey

From 8th June 2020 to 25th September 2020, a MBES survey was carried out within a 2 nautical mile radius from the Tioman coastline and neighboring small islands. The survey was conducted in collaboration with the National Hydrographic Centre (Malaysia) and was conducted on a local fishing vessel, as depicted in Figure 2 and Figure 3. The Kongsberg EM 2040C Single Head, in conjunction with Seafloor Information System (SIS) 4.3.2, was utilized to conduct the MBES survey. The MBES operates within a frequency range of 300 kHz to 350 kHz and has an angular coverage spanning from 120° to 128°. Two types of data were recorded during the acoustic survey, which includes the depth data and intensity return data (commonly referred to as backscatter data). The Veripos - Septentrio Satellite Navigation Positioning System was utilized for the MBES survey, which provides horizontal accuracy of less than 0.01 m using the signal RTK Primary Floating. Daily sound velocity observation was conducted using AML Oceanographic (SVP) during the acoustic survey, with observations taken three times a day (morning, noon, and evening) throughout the survey period. Tidal observations were carried out during the MBES survey by installing tide gauges at multiple locations, such as Marine Park jetty in Kampung Tekek, Kampung Juara jetty, and Tulai Island. Additionally, an established tide gauge station owned by the Department of Survey and Mapping Malaysia (DSMM) was also utilized during the survey.

Figure 2. Local fishing vessel used for MBES Survey.

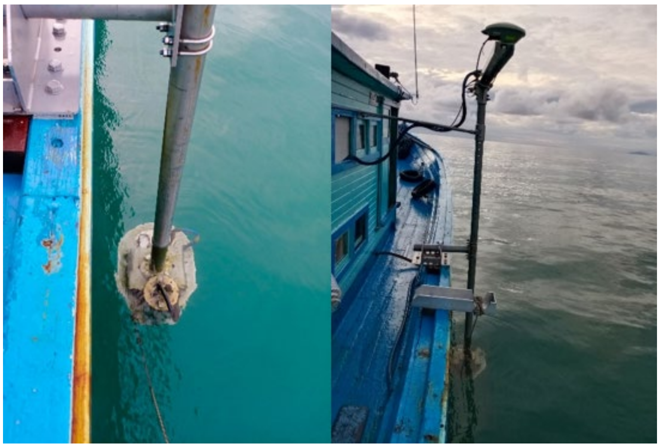

Figure 3. Pole and mounting for sonar head and the position of the sonar head and GNSS receivers.

2.3 Ground-Truth Data

Surveys to obtain ground-truth data was carried out using underwater imagery sampling technique from 25th to 30th September 2020 (Phase 1) and 17th to 19th November 2021 (Phase 2), with the sampling process being performed within the area that was surveyed by MBES. In conducting underwater imagery sampling, an illuminated underwater drop frame was used, which held recording cameras including GoPro Hero 4 and GoPro Hero 8. For Phase 1, 116 underwater imagery sampling points were selected using a stratified random sampling approach, while a total of 68 underwater imagery sampling points were selected for Phase 2. In obtaining photographic evidence of the seabed's characteristics, the video footages for each underwater imagery sampling points was captured were selected, analyzed, and categorized. The captured images were annotated to identify the presence of various seabed features including gravel and sand, sand, rock, coral rubble, coral and rock, and coral. In addition to the underwater imagery sampling, sediment sampling was carried out from 17th to 19th November 2021, within the area surveyed by MBES (as depicted in Figure 8). A sum of 28 sediment samples from the surface of the seabed were obtained using a Ponar grab. All the sediment samples are analysed to identify the sediment types using the GRADISTAT v9.1 software. During sediment grain size analysis, the GRADISTAT software assigns physical descriptions of textural groups and corresponding sediment names to each sample. The software utilizes the Method of Moments and Folk and Ward method to classify the sediment composition of each sample. A ground-truth catalogue was created in Microsoft Excel, which includes captured images and their corresponding details such as ID, coordinates, and seabed features. The catalogue is then exported as a CSV file. The CSV file is first transformed into a shapefile (.shp) format using ArcMap 10.5. Each sample is then converted into polygon features. In assessing the accuracy of the classified images, a train-test split approach is utilized. Specifically, 75% of the samples are used for the train split to generate supervised image classification output, while the remaining 25% of samples are allocated to the test split for validation purposes.

2.4. Data Processing and Analysis

2.4.1. MBES Dataset. After obtaining the raw multibeam dataset, it underwent processing with CARIS HIPS & SIPS software to remove blunders. Subsequently, the depth information was transformed into a gridded bathymetry surface (i.e., high-resolution bathymetric map) with a resolution of 1 meter per pixel. FMGT (FM Geocoder Toolbox) was utilized to process the backscatter data, which resulted in a high-resolution backscatter mosaic 32-bit with a resolution of 1 meter per pixel. This software was responsible for performing several corrections, such as geographical and radiometric corrections for each beam. The corrected beams was converted to a raster format with a resolution of 1 meter per pixel. Both outputs (i.e., bathymetric map and backscatter mosaic 32-bit) can now undergo a supervised image classification process or be utilized in generating other secondary layers such as texture layers.

2.4.2. Second-Derivatives of Bathymetric Map and Backscatter Mosaic 32-bit. The topography of the seabed was analysed using bathymetric maps. Two bathymetric derivatives, namely aspect and slope, were obtained using the Spatial Analyst tool, while the Ruggedness Index and Slope, Aspect, and Curvature tools were used to derive ruggedness index and curvature. Meanwhile, texture and sediment characterization of the seabed were also analysed using the backscatter mosaic 32-bit. Angular Range Analysis (ARA) was conducted using the ARA analysis tool in FMGT to characterize the backscatter data into two derivatives, namely ARA characterisation and ARA phi (Φ). In addition, the backscatter mosaic 8-bit was obtained by deriving it from the backscatter mosaic 32-bit. Using the Co-occurrence measure tool in ENVI software, texture features were derived using the Gray Level Co-Occurrence Matrix (GLCM) analysis method, which produced correlation, entropy, homogeneity, and mean values. The list of first-derivatives and second-derivatives that were derived is presented in Table 1. The list of first-derivatives and second-derivatives that were derived is presented in Table 1.

Table 1. First-derivatives (bathymetric map and backscatter mosaic 32-bit) and second-derivatives derived from bathymetric map and backscatter mosaic 32-bit.

First-Derivative	Second-Derivative	Spatial Resolution	Tool/Software
Bathymetric Map	Aspect	1 meter	Aspect tool (ArcGIS 10.8)
	Slope		Slope tool (ArcGIS 10.8)
	Ruggedness Index		Ruggedness Index tool (ArcGIS 10.8)
	Curvature		Curvature tools (ArcGIS 10.8)
Backscatter Mosaic 32-bit	Backscatter Mosaic 8-bit		ArcGIS 10.8
	ARA Characterisation		FMGT
	ARA Phi (Φ)		FMGT
	GLCM Correlation		FMGT
	GLCM Entropy		Co-occurrence Measures (ENVI Classic
	GLCM Homogeneity		Co-occurrence Measures (ENVI Classic
	GLCM Mean		Co-occurrence Measures (ENVI Classic

2.4.3. Correlation Analysis. The relationship between all derivatives was analysed using Pearson's correlation coefficient. During the model fitting process, derivatives that were highly correlated were excluded based on previous studies [14]. In accordance with Zuur, Ieno, Walker, Saveliev and Smith [15], derivatives having a correlation coefficient value greater than 0.5 were classified as possessing high correlation, whereas those with a correlation coefficient value less than 0.5 were utilized as predictors for model fitting, and conversely.. This approach is also commonly used in statistical modeling to reduce the number of predictors and avoid issues related to multicollinearity [16].

2.4.4. Supervised Image Classification. Random Forest Algorithm. The marine habitat map was generated using the machine learning algorithm, Random Forest (RF) and RStudio software. The development of models for intricate data classification, the R language platform's Caret package was utilized [17]. The process consisted of three phases, which involved preparing data, fitting the model, and performing the classification. During the initial stage, the ground-truth data was utilized as the data inputs. From this data, six distinct habitat classes were identified, which included gravel and sand, sand, rock, coral rubble, coral and rock, and coral (refer to Table 2). Following this, the data was tabulated and saved in a Comma-Separated Values file (.csv) format, which was then transformed into a polygon shapefile (.shp) format utilizing ArcMap 10.5. Additionally, to decrease the file size without compressing the original data, all derivatives from the ESRI ASCII raster (.asc) format for the MBES derivative were changed into the Tag Image File (.TIF) format.

Table 2. Six habitat classes were used in developing marine habitat map using RF algorithm.

	Habitat Class	Number of Sample
1	Gravel and Sand	37
2	Sand	44
3	Rock	40
4	Coral Rubble	3
5	Coral and Rock	11
6	Coral	47
	Total	182

During the second stage, in RStudio software, the *shapefile* function available in the *sdm* package was utilized to read the ground-truth data. Subsequently, the *responCol* function was used to convert ground-truth data into the *SpatialPolygonDataFrame* object. Concerning MBES derivatives, the *list.*files and *raster::stack* functions were used to list and stack all the derivatives. Following this, the stacked predictor was transformed into a *RasterBrick* object using the *brick* function. Meanwhile, the *list.files* and raster::stack functions were also utilized to list and stack all the MBES derivatives. The *data.frame* function was then used to overlay and extract pixel values from derivatives based on seascape feature data. The data that was extracted was later separated into two datasets (training and the testing datasets) using the *createDataPartition* function. Both datasets underwent the cleaning process to eliminate null data utilizing the *na.omit* function.

During the final stage of the process, the *train* function was utilized to fit and create a model object for supervised image classification. The *raster::predict* function was then used to generate a new raster that contained predictions from the fitted model object. To speed up computations, the *clusterR* functions from the *raster* package were employed, which support parallelization or multi-core computing for the *raster::predict* function. This function was effective in reducing computation time. The resulting raster with predictions was then exported into a binary image format (.TIF) using the *writeRaster* function. In validating the classified image, which was a marine habitat map, the *confusionMatrix* function was used to compute the overall accuracy and kappa statistics.

3. Result and Analysis

3.1. First-Derivatives. Bathymetric Map and Backscatter Mosaic. A bathymetry map was generated through the use of bathymetric data collected over a 406 km^2 area of the seabed within Tioman Marine Park. The study area's terrain was discovered to be intricate, with depths varying from 0 meters close to the islands' coastlines to a maximum of -70 meters in the western and northern regions of Tioman Island (as shown in Figure 9 and Figure 10). After processing the intensity return from beam, the raw amplitude data was successfully converted into reflectivity or intensity backscatter in decibel values that ranged from 10 to -70 decibels (dB) (refer to Figure 11). The shallow areas, located near the coastline, had mostly shown high intensity values, which can be attributed to the presence of rocks or coral cover. On the other hand, lower intensity values were generally observed in deeper areas, which may indicate sandy sediment such as clay or fine sand. In the north-west region of Tioman Island, in addition to the high intensity reflectance in the shallow areas, seabed features such as rocks in sandy areas were also observed. These features were mostly located around flat and plain sandy sediment areas, making them clearly visible.

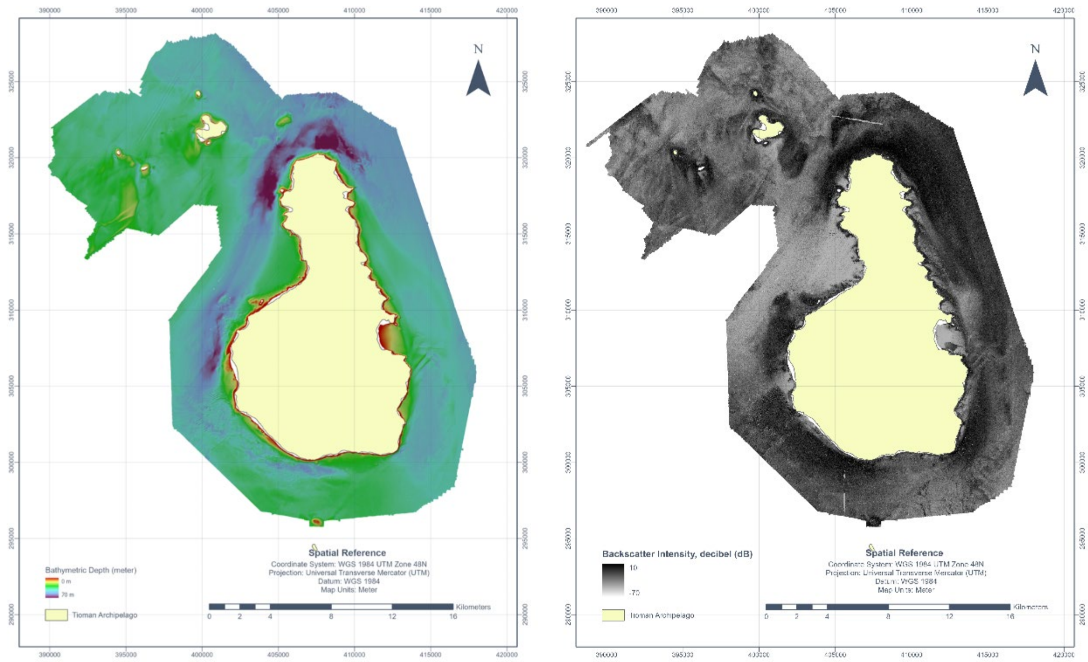

Figure 4. A bathymetric map (left) that represented underwater topography and a backscatter mosaic 32-bit (right) that represented acoustic signal intensity, created from a dataset collected by MBES in the Tioman Marine Park region.

3.2. Second-Derivatives. The results derived from the MBES dataset provide valuable insights into the underwater environment around the Tioman Marine Park area. The MBES dataset was used to create several derivatives, including the aspect, curvature, ruggedness index (Figure 5), 32-bit backscatter mosaic, 8-bit backscatter mosaic, ARA parameters, GLCM texture features (Figure 6). The bathymetric map derived from the MBES dataset provides a detailed representation of the underwater topography in the study area. Additionally, the backscatter mosaic, which represent the intensity of the returned acoustic signals, offer insights into the seabed substrate and texture. The 32-bit backscatter mosaic provides a more detailed representation of the acoustic signal intensity, while the 8-bit backscatter mosaic provides a more visually appealing representation. The other derivatives, including the aspect, curvature, and ruggedness index, provide further insights into the seabed morphology and geological structures. The phi and characterization, correlation, entropy, homogeneity, and mean were texture-based analyses that help to identify and classify different seabed substrates based on their acoustic signatures. These derivatives are crucial in studying the ecology and geology of the study area, as they provide valuable information about the seabed and the habitats that it supports.

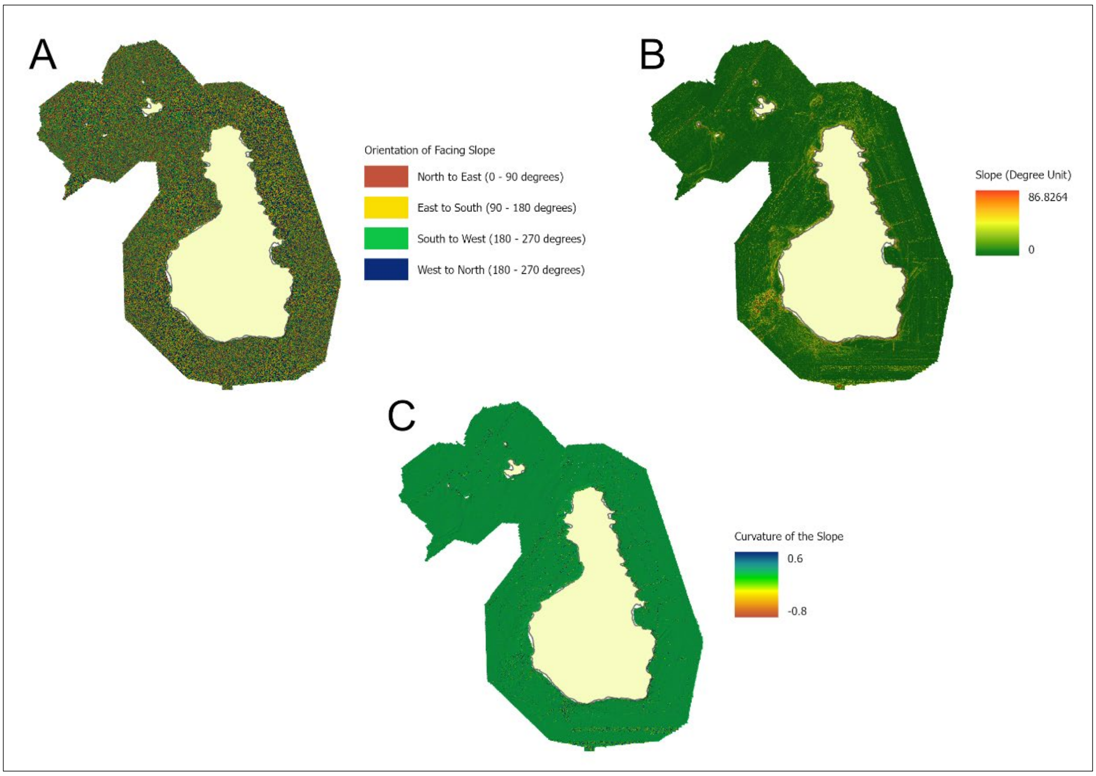

Figure 5. Bathymetric derivatives derived from bathymetric map. (a) aspect, (b) slope, and (c) curvature.

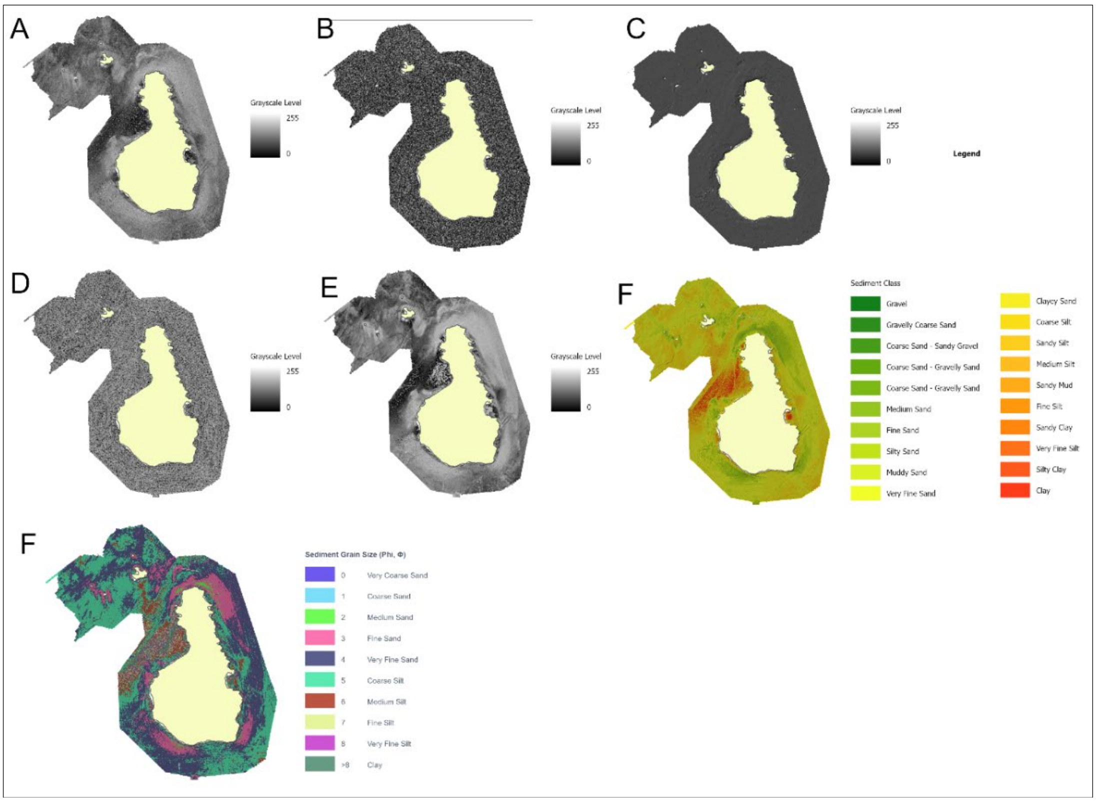

Figure 6. backscatter derivatives derived from backscatter mosaic 32-bit. (a) backscatter mosaic 8-bit, (b) correlation, (c) entropy, (d) homogeneity, (e) mean, (f) characterisation, and (g) phi.

3.3. Correlation between Derivatives. A weak correlation is characterized by an *r* value of less than 0.5, while a strong correlation is represented by an *r* value of 0.5 or greater. Weak correlation and strong correlation were found among several derivatives derived from MBES dataset, as shown in Figure 7 and Table 3. Derivatives such as aspect, bathymetric map, curvature, ruggedness index, backscatter mosaic 32-bit, backscatter mosaic 8-bit, correlation, entropy, and homogeneity were observed as weak correlation derivatives, while characterization, phi, and mean were observed as strong correlation derivatives. As a result, the weak correlation derivatives were selected as data input for the supervised image classification process.

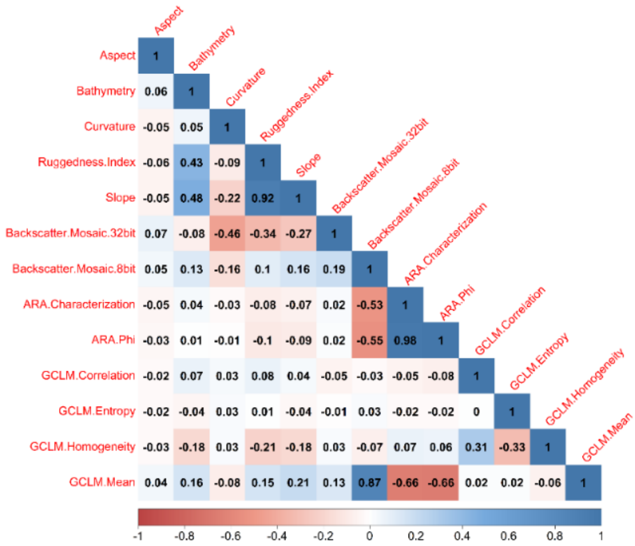

Figure 7. A correlogram was created to display the correlation matrix and highlight the correlations between derivatives.

Table 3. A weak correlation between MBES derivatives indicates that there is little relationship between them, while a strong correlation indicates a significant relationship.

Weak Correlation (*r* less than 0.5)	Strong Correlation (*r* greater than or Equal to 0.5)
Aspect	Slope
Bathymetric Map	ARA Characterization
Curvature	ARA Phi
Ruggedness Index	GLCM Mean
Backscatter Mosaic 32-bit	
Backscatter Mosaic 8-bit	
GLCM Correlation	
GLCM Entropy	
GLCM Homogeneity	

3.4. Marine Habitat Map. For accuracy of classified image, the statistical approach of accuracy assessments provides an overall quantitative assessment, indicating that the marine habitat map is an excellent and convincing reference tool for identifying marine habitat in the MBES surveyed area. The study revealed that the overall accuracy and kappa statistics were 98% and 0.88, respectively. Figure 8 displays the marine habitat map, which identifies six habitat classes including sand, rock, gravel and sand, coral rubble, coral and rock, and coral. Table 4 shows that sand is the predominant habitat class, covering 78.70% of the total surveyed area using MBES and clearly illustrated in Figure 8. Additionally, about 18.75% of the surveyed area, equivalent to approximately 74.66 km^2 out of 398.2 km^2, was classified as an area comprising of mixed gravel and sand. Habitat classes such as coral, rock, coral mixed with rock, and coral rubble were found to have a classification of less than 1.60%. Therefore, the analysis of the surveyed area using MBES, which covers the coastal waters of TMP area, reveals that sand and gravel are the dominant habitat types.

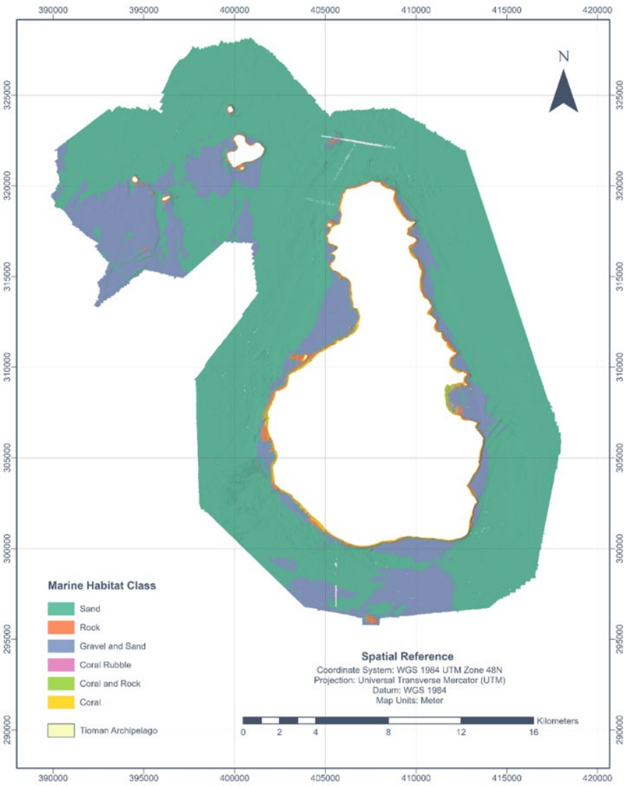

Figure 8. Marine habitat map derived using RF algorithm. Six habitat classes including sand, rock, gravel and sand, coral rubble, coral and rock, and coral were identified.

Table 4. The marine habitat map categorizes sand, rock, gravel and sand, coral rubble, coral and rock, and coral and provides their respective rank and percentage area.

Rank	Habitat Class	Percentage Area (%)
1	Sand	78.70
2	Gravel and Sand	18.75
3	Coral	0.78
4	Rock	1.60
5	Coral and Rock	0.15
6	Coral Rubble	0.02

4. Discussion

The incorporation of a dataset obtained through high-resolution multibeam echosounder (MBES) technology, ground-truth data, and RF algorithm to create a marine habitat map for the TMP in Malaysia's coastal areas was explored in this study. This research represents one of the first attempts to employ these techniques for this purpose. Based on marine habitat map illustrated that sand is the primary habitat covering of the MBES surveyed area. Areas consisting of mixed gravel and sand are also predominantly found around the MBES surveyed area. The habitat classes of coral, rock, mixed coral and rock, and coral rubble are found in small patches in this area. Most of coral habitats in this study were associated with sand and rock. Yap [18] has stated clearly that corals were discovered in close proximity to both sandy and bedrock regions, providing supporting evidence for obtained results. According to , coral and coral rubble were typically present in the shallow water areas along the Tioman coastline, which aligns with the findings of the previous study.

The data collection process used in the study involved the use of MBES to collect bathymetry and backscatter data. Bathymetry data refers to measurements of the depth of the seabed [19]. This data was used to extract information of the seabed, which can provide valuable information about underwater features such as ridges, canyons, and seamounts [20]. Meanwhile, backscatter data refers to measurements of the intensity of returned acoustic signals [21]. This data was used to obtained related information such as texture and composition of the seabed, as different types of sediments and rocks will reflect acoustic signals differently [22]. Previous study by Zakariya, Abdullah, Hasan and Khalil [23] created and utilized backscatter mosaic, bathymetric map, slope, and rugosity to establish different categories for characterizing the seabed. The outcome indicated that the sediment of the study area was accurately classified, since the overall accuracy and kappa coefficient were higher than 80% and 0.8, respectively. This study presents a chance to improve the precision of identifying the characteristics of seabed sediment in the MBES surveyed area. Recent study by Che Hasan, Md. Said and Khalil [24] utilized high-spatial resolution bathymetry and backscatter data, along with their related derivatives, to apply the RF algorithm to generate marine habitats around Redang Marine Park (RMP). This study accomplished the construction of comprehensive maps of benthic habitats at a spatial resolution of 1 meter, offering an understanding of how coral, fine sand, and coarse sand are distributed throughout this region. Through analysis, the significant variables for predicting these habitats were found to be bathymetry (and related data) and GLCM mean. This study's results highlight the possibility of integrating machine learning techniques with underwater sonar data to accurately map seabed habitats.

The RF algorithm is highly effective in handling datasets that are intricate and diverse, making it a robust machine learning algorithm [25]. This algorithm works by creating an ensemble of decision trees and combining their predictions to generate a final output [21]. RF algorithm has been widely used for habitat mapping in marine environments because it can handle large datasets with multiple variables [26], and it is able to detect non-linear relationships between environmental variables and habitat types [27]. The current study's findings indicate that the RF algorithm was employed to create a marine habitat map with a remarkable accuracy rate of 98%. This finding was supported by previous study [28], using a multi-frequency MBES dataset, the RF algorithm was employed to map habitats such as very fine sand, sand, and red algae in the Southern Baltic Sea. This study also utilized several supervised image classification algorithms, including classification and regression trees, random forest, support vector machine, and k-nearest neighbors. These algorithms were used to classify the coral reef habitats into seven categories, including very fine sand, sand, and algae, boulders, and sandy gravel, and gravelly sand. The results of the study showed that RF algorithm achieved a classification accuracy of over 85%. Another study by Wan, Qin, Cui, Yang, Yasir, Ma and Liu [29] used MBES dataset and RF algorithm to map the sediments in the sea around the United Kingdom. This study used a MBES to collect the acoustic data, which was then used to classify the sediment composition. According to this study's findings, the RF algorithm attained an accuracy rate of over 86%, making it the most effective algorithm. However, the recent study conducted by Nemani, Cote, Misiuk, Edinger, Mackin-McLaughlin, Templeton, Shaw and Robert [30] reported lower accuracy rates for classification when compared to other algorithms like extreme gradient boosting support vector machines.

The study's findings regarding the spatial distribution of marine habitats in Malaysia’s coastal water area have several implications for coral habitat conservation and management in the country. The marine habitat map created from this study could be used to prioritize areas for conservation efforts, plan conservation activities, monitor coral health, guide further research, and improve management effectiveness [31]. According to recent research, the world's oceans are facing unprecedented levels of stress from human activities, including overfishing, pollution, and climate change, which has led to an increased need for marine habitat mapping and conservation efforts [32]. The implication is that in recent times, protecting marine habitats has gained greater significance in order to alleviate the adverse effects of anthropogenic activities on the health and diversity of marine environments. A crucial component of these efforts is the development of accurate and comprehensive maps of marine habitats, which can help guide conservation and management actions.

5. Conclusion

This study used MBES dataset (i.e., bathymetry and backscatter), MBES derivatives, and RF supervised image classification approach to classify marine habitat including sand, rock, and gravel and sand, coral rubble, coral and rock, and coral in TMP. The coral habitat distribution was associated with the bathymetry depths that distributed within shallow water area. Consequently, this study concluded that that the RF machine learning algorithm is useful in categorizing the distribution of marine habitats, particularly coral habitat. Since this is an initial study conducted to develop marine habitat maps in the study area, further research is required in the future, which should incorporate various machine learning algorithms to create the marine habitat maps. Creating an accurate map of marine habitats is crucial for studying, conserving, and monitoring these habitats in Malaysia’s coastal water areas, particularly in the face of anthropogenic activities and the effects of climate change.

References

[1] Hughes T P, Anderson K D, Connolly S R, Heron S F, Kerry J T, Lough J M, Baird A H, Baum J K, Berumen M L and Bridge T C 2018 Spatial and temporal patterns of mass bleaching of corals in the Anthropocene *Science* **359** 80-3
[2] Saad S, Khodzori F A, Salleh F, Mohammad Noor N and Mukai Y 2016 Current status of coral reefs in Tioman Island, Peninsular Malaysia *Turkish Journal of Zoology* **41**
[3] Hedley J D, Roelfsema C M, Chollett I, Harborne A R, Heron S F, Weeks S, Skirving W J, Strong A E, Eakin C M, Christensen T R L, Ticzon V, Bejarano S and Mumby P J 2016 Remote Sensing of Coral Reefs for Monitoring and Management: A Review *Remote Sensing* **8** 118
[4] Brown C J, Todd B J, Kostylev V E and Pickrill R A 2011 Image-based classification of multibeam sonar backscatter data for objective surficial sediment mapping of Georges Bank, Canada *Continental Shelf Research* **31** S110-S9
[5] De Falco G, Tonielli R, Di Martino G, Innangi S, Simeone S and Parnum I 2010 Relationships between multibeam backscatter, sediment grain size and Posidonia oceanica seagrass distribution *Continental Shelf Research* **30** 1941-50
[6] Brown C J, Smith S J, Lawton P and Anderson J T 2011 Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques *Estuarine, Coastal and Shelf Science* **92** 502-20
[7] Che Hasan R, Rattray A, Monk J and Laurenson L 2011 *Applications of multibeam echosounder data and video observations for biological monitoring on the south east Australian continental shelf*
[8] Calvert J, Strong J A, Service M, McGonigle C and Quinn R 2015 An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data *ICES Journal of Marine Science* **72** 1498-513
[9] Mohapatra N, Shreya K and Chinmay A 2020 Optimization of the random forest algorithm. In: *Advances in Data Science and Management: Proceedings of ICDSM 2019*: Springer) pp 201-8

[10] Hu H, Feng C, Cui X, Zhang K, Bu X and Yang F 2023 A Sample Enhancement Method based on Simple Linear Iterative Clustering Superpixel Segmentation applied to Multibeam Seabed Classification *IEEE Transactions on Geoscience and Remote Sensing*
[11] Held P and Schneider von Deimling J 2019 New feature classes for acoustic habitat mapping—A multibeam echosounder point cloud analysis for mapping submerged aquatic vegetation (SAV) *Geosciences* **9** 235
[12] Diesing M and Stephens D 2015 A multi-model ensemble approach to seabed mapping *Journal of Sea Research* **100** 62-9
[13] Misiuk B, Diesing M, Aitken A, Brown C J, Edinger E N and Bell T 2019 A Spatially Explicit Comparison of Quantitative and Categorical Modelling Approaches for Mapping Seabed Sediments Using Random Forest *Geosciences* **9** 254
[14] Trzcinska K, Janowski L, Nowak J, Rucinska-Zjadacz M, Kruss A, von Deimling J S, Pocwiardowski P and Tegowski J 2020 Spectral features of dual-frequency multibeam echosounder data for benthic habitat mapping *Marine Geology* **427** 106239
[15] Zuur A, Ieno E N, Walker N, Saveliev A A and Smith G M 2009 *Mixed effects models and extensions in ecology with R*: Springer Science & Business Media)
[16] Powers D 2008 Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation *Mach. Learn. Technol.* **2**
[17] Kuhn M 2015 A Short Introduction to the caret Package *R Found Stat Comput* **1** 1-10
[18] Yap K 2017 An Attempt Using Hydro-Acoustic Approach for Rapid Coral Reef Mapping in Malaysia
[19] Long J, Zhang H and Zhao J 2023 A Comprehensive Deep Learning-Based Outlier Removal Method for Multibeam Bathymetric Point Cloud *IEEE Transactions on Geoscience and Remote Sensing* **61** 1-22
[20] Idárraga-García J and León H 2019 Unraveling the underwater morphological features of Roncador bank, archipelago of San Andrés, Providencia and Santa Catalina (Colombian Caribbean) *Frontiers in Marine Science* **6** 77
[21] Ji X, Yang B and Tang Q 2020 Seabed sediment classification using multibeam backscatter data based on the selecting optimal random forest model *Applied Acoustics* **167** 107387
[22] Lamarche G and Lurton X 2018 Recommendations for improved and coherent acquisition and processing of backscatter data from seafloor-mapping sonars *Marine Geophysical Research* **39** 5-22
[23] Zakariya R, Abdullah M A, Hasan R C and Khalil I 2018 *Engineering Applications for New Materials and Technologies,* ed A Öchsner (Cham: Springer International Publishing) pp 579-91
[24] Che Hasan R, Md. Said N and Khalil I 2022 Seafloor Habitat Mapping Using Machine Learning and Underwater Acoustic Sonar. In: *Computational Intelligence in Machine Learning,* ed A Kumar*, et al.* (Singapore: Springer Nature Singapore) pp 281-7
[25] Rawson A, Sabeur Z and Brito M 2022 Intelligent geospatial maritime risk analytics using the Discrete Global Grid System *Big Earth Data* **6** 294-322
[26] Xu W, Cheng H, Zheng S and Hu H 2021 Predicted Mapping of Seabed Sediments Based on MBES Backscatter and Bathymetric Data: A Case Study in Joseph Bonaparte Gulf, Australia, Using Random Forest Decision Tree *Journal of Marine Science and Engineering* **9** 947
[27] See K E, Ackerman M W, Carmichael R A, Hoffmann S L and Beasley C 2021 Estimating carrying capacity for juvenile salmon using quantile random forest models *Ecosphere* **12** e03404
[28] Janowski L, Trzcinska K, Tegowski J, Kruss A, Rucinska-Zjadacz M and Pocwiardowski P 2018 Nearshore benthic habitat mapping based on multi-frequency, multibeam echosounder data using a combined object-based approach: a case study from the rowy site in the Southern Baltic Sea *Remote Sensing* **10** 1983

[29] Wan J, Qin Z, Cui X, Yang F, Yasir M, Ma B and Liu X 2022 MBES Seabed Sediment Classification Based on a Decision Fusion Method Using Deep Learning Model *Remote Sensing* **14** 3708
[30] Nemani S, Cote D, Misiuk B, Edinger E, Mackin-McLaughlin J, Templeton A, Shaw J and Robert K 2022 A multi-scale feature selection approach for predicting benthic assemblages *Estuarine, Coastal and Shelf Science* **277** 108053
[31] Asmawi M Z, Daud M A R M and Abd Aziz A R 2020 SPATIAL MAPPING OF ENVIRONMENTAL SENSITIVE INDEX IN PULAU HUJUNG AND PULAU TENGAH, JOHOR *Journal of Architecture, Planning and Construction Management* **10**
[32] Halpern B S, Frazier M, Potapenko J, Casey K S, Koenig K, Longo C, Lowndes J S, Rockwood R C, Selig E R, Selkoe K A and Walbridge S 2015 Spatial and temporal changes in cumulative human impacts on the world's ocean *Nature Communications* **6** 7615

Acknowledgements
The National Hydrographic Centre (NHC), Royal Malaysian Navy is acknowledged by the authors for their assistance in collecting the acoustic data. The authors would also like to express their gratitude towards the team partners from Universiti Teknologi Malaysia Kuala Lumpur, Department of Fisheries Malaysia (DoF), and Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu.