APPLIED COMPUTER SCIENCE

Applied Computer Science, vol. 20, no. 4, pp. 100–116
doi: https://doi.org/10.35784/acs-2024-42

Submitted: 2024-09-12 | Revised: 2024-11-13 | Accepted: 2024-11-20

Keywords: acoustic parameters, single-beam echosounder, mapping accuracy, Random Forest, Support Vector Machine

Baigo HAMUNA [0000-0002-0706-2496]*, *Sri PUJIYATI* [0000-0003-4049-4589]**, *Jonson Lumban GAOL* [0000-0001-8908-3161]**, *Totok HESTIRIANOTO* [0000-0002-1636-4525]**

CLASSIFICATION AND PREDICTION OF BENTHIC HABITAT BASED ON SCIENTIFIC ECHOSOUNDER DATA: APPLICATION OF MACHINE LEARNING ALGORITHMS

Abstract

This study aims to map three main benthic habitats (coral, seagrass, and sand) in Kapota Atoll (Wakatobi, Indonesia) using a single-beam echosounder (SBES) Simrad EK15. The acoustic data were processed using Sonar5-Pro software. Eight acoustic parameters were used as input for the classification and prediction of benthic habitats, including depth (D), five acoustic parameters of the first echo (BD, BP, AttSv1, DecSv1, and AttDecSv1), and cumulative energy of the second and third echoes (AttDecSv2 and AttDecSv3). The classification and prediction process of benthic habitats uses two machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), in XLSTAT Basic+ software. The study results show that 49 combinations of acoustic parameters produce benthic habitat maps that meet the minimum accuracy standards for benthic habitat mapping (≥60%). Using eight acoustic parameters produces a more accurate benthic habitat map than using only two main SBES parameters (DecSv1 and AttDecSv2 parameters or E1 and E2 in the RoxAnn system indicating the roughness and hardness indices). The RF and SVM algorithms produce benthic habitat maps with the highest accuracy of 79.33% and 78.67%, respectively. Each acoustic parameter has a different importance for the classification of benthic habitats, where the order of importance of each acoustic parameter in the overall classification follows the following order: AttDecSv2 > D > DecSv1 > BD > AttDecSv3 > AttSv1 > AttDecSv1 > BP. Overall, using more acoustic parameters can significantly improve the accuracy of benthic habitat maps.

* Cenderawasih University, Faculty of Mathematics and Natural Science, Department of Marine Science and Fisheries, Indonesia
** IPB University, Faculty of Fisheries and Marine Sciences, Department of Marine Sciences and Technology, Indonesia, sripu@apps.ipb.ac.id

1. INTRODUCTION

Developments in benthic habitat mapping have resulted in various approaches, data types, technologies, and models that can be used to understand and map the distribution patterns of biotic and abiotic components on the seafloor (Misiuk & Brown, 2024). Various hydroacoustic instruments for seafloor habitat mapping have developed rapidly with varying degrees of effectiveness (Anderson et al., 2008; Brown et al., 2011; Pijanowski & Brown, 2022; Wölfl et al., 2019). Scientific echosounders, such as single-beam echosounders (SBES), have been reported as reliable instruments for detecting objects in the water column (Manik et al., 2014; Moszynski & Hedgepeth, 2000; Pujiyati et al., 2022) and for classifying and mapping the seafloor (Henriques et al., 2015; Lee & Lin, 2018; McLaren et al., 2019; Reshitnyk et al., 2014; Sánchez-Carnero et al., 2023; Solikin et al., 2018; Vassallo et al., 2018). SBES have become standard instruments in recent decades due to their affordable cost (Anderson et al., 2008; Fajaryanti & Kang, 2019; Sánchez-Carnero et al., 2023) and standard data processing procedures (Anderson et al., 2008).

The accuracy level in the classification process and spatial mapping of benthic habitats and seabed substrates is a fundamental problem. The selection of classification methods is an important factor in improving map accuracy (Shao et al., 2021). Various studies have used multiple methods to classify and map seabed habitats or substrates from SBES instruments, such as Principal Component Analysis (PCA) (Bartholomä et al., 2020; Bravo & Grant, 2020; Fajaryanti & Kang, 2019; Freitas et al., 2011; Reshitnyk et al., 2014) and clustering analysis (Henriques et al., 2015; Lee & Lin, 2018; Poulain et al., 2011; Reshitnyk et al., 2014; Riegl & Purkis, 2005). Most of the classifications using PCA and clustering analysis are based on the working principles of the software used, namely the QTC View system and RoxAnn system.

Many studies have used machine learning classification methods to classify seabed habitats based on acoustic backscatter strength and characteristics. The use of various machine learning algorithms to classify hydroacoustic data has increased significantly and has replaced manual interpretation (Misiuk & Brown, 2024), and has shown more accurate benthic habitat classification results in several studies (Gumusay et al., 2018; Misiuk & Brown, 2024; Shao et al., 2021).

Maps of benthic habitats and seabed substrates produced from SBES data have a very variable level of accuracy, depending on the number of acoustic parameters used as input in the classification process (Bejarano et al., 2010; Sánchez-Carnero et al., 2023; Shao et al., 2021). This study uses eight acoustic parameters, namely the first echo to the third echo acoustic parameters from the seabed. The assumption is that each type of seabed benthic will have different third echo acoustic characteristics, as in the first and second echo, and may play a role as an input parameter in the classification process. Until now, no study has defined the role of acoustic parameters in the third echo from the seabed and used it as an input parameter in classifying benthic habitats or seabed substrates. Therefore, this study aims to: (1) map benthic habitats (coral, seagrass, and sand) using various combinations of first echo to third echo acoustic parameters and using two machine learning algorithms for classification and prediction of benthic habitats, namely Random Forest (RF) and Support Vector Machine (SVM); and (2) determine the order of importance of acoustic parameters in the classification process and their influence on increasing the accuracy of benthic habitat maps.

2. METHODOLOGY

2.1. Acoustic data recording and ground-truth data collection

An acoustic survey to record acoustic backscatter energy from the seabed was conducted at Kapota Atoll, located in the waters of Wakatobi National Park, Wakatobi Regency, Indonesia (Figure 1). The hydroacoustic instrument used was the SBES Simrad EK15 operating at a frequency of 200 kHz. A total of 46 acoustic survey track with an east-west orientation in the eastern part of Kapota Atoll (Figure 1c). The transducer was placed on the ship's starboard side at a depth of 0.5 m. The shipping speed during the acoustic survey was 3.5 to 4.5 knots, which is the ideal ship speed range for acoustic surveys (Sánchez-Carnero et al., 2023; Shao et al., 2021).

The ground-truth data collected represented three main types of benthic habitats in shallow waters: coral, sand, and seagrass. A total of 211 ground-truth data points were obtained, consisting of 176 data points collected right on the acoustic track and 35 data points not on the acoustic track (Figure 1c). Garmin 65s multi-band GPS was used to determine the coordinate position of each ground-truth data point. A total of 61 data points along the acoustic track were used as training data, where the acoustic parameter values in each training data could be determined by finding the closest point between the training data point and the acoustic data ping point. Furthermore, a 10 m buffer was created around the training data points, and several acoustic data pings were taken as training data. The similarity of benthic habitat types between the training data and acoustic data visually using high-resolution imagery (SPOT-6 imagery) was considered when using the acoustic data pings in the 10 m buffer as training data. This process resulted in 506 acoustic data as training data used in the classification process.

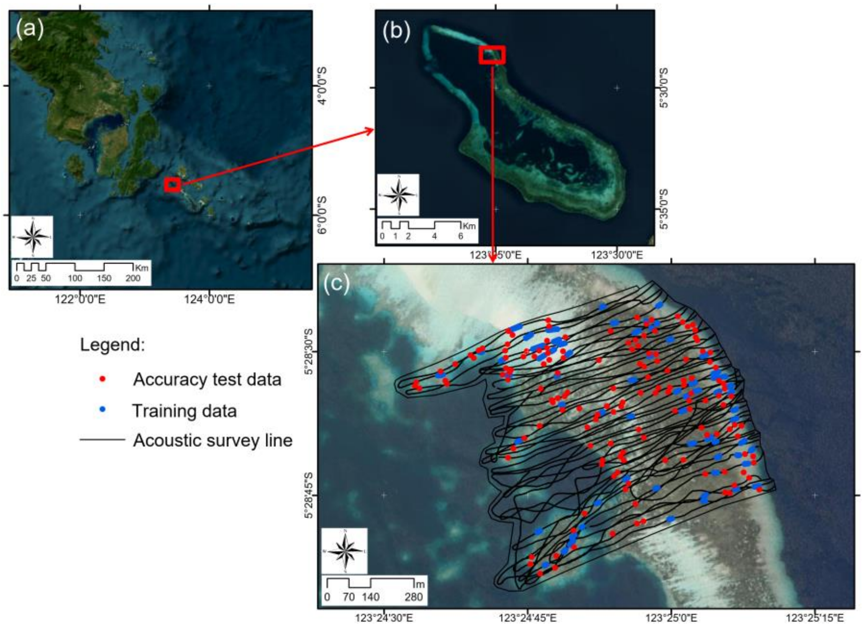

Fig. 1. Map of the study area: (a) waters of Wakatobi Regency (Indonesia), (b) Kapota Atoll, and (c) acoustic survey area

2.2. Acoustic data recording and ground-truth data collection

Acoustic data was processed using Sonar5-Pro software (Balk & Lindem, 2015). Sonar5-Pro is a tool that performs well in extracting acoustic data up to the third echo from the seabed. *Bottom Detection* tools are used to detect the seabed automatically (Balk & Lindem, 2015; Hilgert et al., 2016). The *Microphyte Top Line* algorithm is used to detect the seabed surface so that vegetation growing on the seabed surface is detected below the seabed surface line. The threshold value to determine the seabed surface is -40 dB (Shao et al., 2021).

Each acoustic data ping is used as a single sample size in the classification process. The acoustic parameter values of the first echo to the third echo can be exported using Sonar5-Pro. The acoustic parameters used in the classification and prediction process include depth (D), bottom detection (BD; backscattered energy on the seabed surface), bottom peak (BP; echo peak from the seabed), attack phase energy or the energy of the first part on the first echo (AttSv1), decay phase energy or the energy of the last part on the first echo (DecSv1), and cumulative energy in the first echo (AttDecSv1), second echo (AttDecSv2), and third echo (AttDecSv3). An illustration of the seabed echoes showing the eight acoustic parameters is presented in Figure 2. The roughness and hardness energy of the seabed in the RoxAnn system are different from Sonar5-Pro (Hilgert et al., 2016; Poulain et al., 2011). The energy at the tail of the first echo shows the roughness index (E1) by the RoxAnn system and the DecSv1 energy by Sonar5-Pro, which is obtained directly from Sonar5-Pro. The entire second echo energy shows the hardness index (E2) by the RoxAnn system and are assessed from the AttDecSv2 energy in Sonar5-Pro.

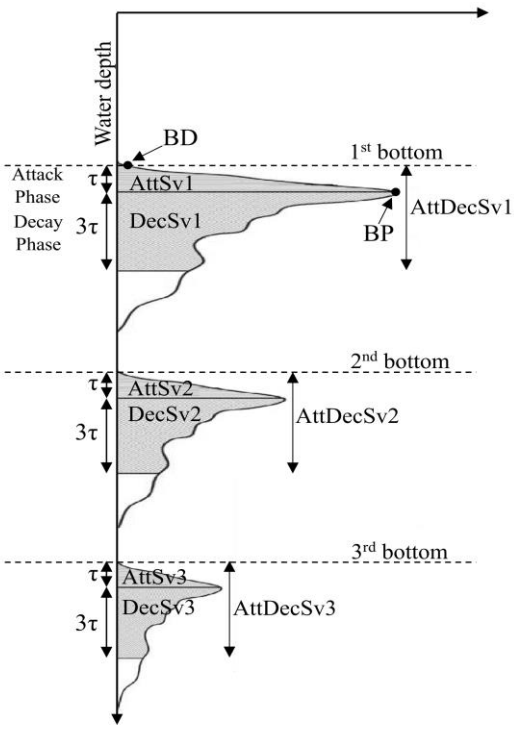

Fig. 2. Illustration of the division of the first, second, and third seabed echoes, as well as the acoustic parameters used as classification input

Mathematically, the $AttSv_x$ and $DecSv_x$ values of the first, second and third echo can be calculated using equations (Balk & Lindem, 2015):

$$AttSv_x = 10\log\left(\frac{1}{N_A}\sum_{i=AI_1}^{AI_2}(10^{Sv_1/10}\right) \quad (1)$$

$$DecSv_x = 10\log\left(\frac{1}{N_D}\sum_{i=DI_1}^{DI_2}(10^{Sv_1/10}\right) \quad (2)$$

where: x – the echoes from the first, second or third bottom,
N_A – number of attack samples (8),
N_D – number of decay samples (24),
AI_1 – bottom index.

AI_2, DI_1, and DI_2 can be calculated using equations (Balk & Lindem, 2015):

$$AI_2 = Bottom\ index + Attack\ samples \quad (3)$$

$$DI_1 = Bottom\ index + Attack\ samples + 1 \quad (4)$$

$$DI_2 = Bottom\ index + Attack\ samples + 1 + Decay\ samples \quad (5)$$

The values of AttDecSv1, AttDecSv2, and AttDecSv3 can be calculated from the values of $AttSv_x$ and $DecSv_x$ at each echo using the equation (Balk & Lindem, 2015; Poulain et al., 2011):

$$AttDecSv_x = 10\log\left(\frac{1}{N_A+N_D}\left(N_A\times10^{\frac{Att_i}{10}} + N_D\times10^{\frac{Dec_i}{10}}\right)\right) \quad (6)$$

In determining the AttDecSv value, the first attack sample is not used when integrated into the DecSv value, so the *ND* value used is 23 (Poulain et al., 2011).

2.3. Training data classification and benthic habitat prediction

Classification of 506 labelled acoustic data (training data) and prediction of benthic habitats were performed using XLSTAT Basic+ software (Lumivero, 2023). The results of the training data classification were used to predict observation data (16447 unlabeled acoustic data) into certain benthic habitat classes based on the values of their acoustic parameters, which were used as input variables in the classification process. There were three schemes of acoustic parameter combinations in the classification and prediction process of benthic habitats (Table 1), namely using only the first echo acoustic parameter (14 combinations), a combination of the first and second echo acoustic parameters (10 combinations), and a combination of the first, second, and third echo (10 combinations).

Tab. 1. Combination of acoustic parameters for the classification process

Combination	Acoustic parameters
First echo	
A1	AttSv1-DecSv1
A2	BD-AttSv1-DecSv1
A3	BP-AttSv1-DecSv1
A4	BD-BP-AttSv1-DecSv1
A5	D-AttSv1-DecSv1
A6	D-BP-AttSv1-DecSv1
A7	D-BD-AttSv1-DecSv1
A8	D-BP-DecSv1-AttDecSv1
A9	D-BD-DeSv1-AttDecSv1
A10	D-AttSv1-DecSv1-AttDecSv1
A11	D-BD-BP-AttSv1-DecSv1
A12	D-BD-AttSv1-DecSv1-AttDecSv1
A13	D-BP-AttSv1-DecSv1-AttDecSv1
A14	D-BD-BP-AttSv1-DecSv1-AttDecSv1
First and second echoes	
B1	DecSv1-AttDecSv2
B2	D-DecSv1-AttDecSv2
B3	BP1-DecSv1-AttDecSv2
B4	D-BP-DecSv1-AttDecSv2
B5	BD-DecSv1-AttDecSv2
B6	D-BD-DecSv1-AttDecSv2
B7	BD-BP-DecSv1-AttDecSv2
B8	D-BD-BP-DecSv1-AttDecSv2
B9	D-BD-BP-AttSv1-DecSv1-AttDecSv2
B10	D-BD-BP-AttSv1-DecSv1-AttDecSv1-AttDecSv2
First, second, and third echoes	
C1	DecSv1-AttDecSv2-AttDecSv3
C2	D-DecSv1-AttDecSv2-AttDecSv3
C3	BP1-DecSv1-AttDecSv2-AttDecSv3
C4	D-BP-DecSv1-AttDecSv2-AttDecSv3
C5	BD-DecSv1-AttDecSv2-AttDecSv3
C6	D-BD-DecSv1-AttDecSv2-AttDecSv3
C7	BD-BP-DecSv1-AttDecSv2-AttDecSv3
C8	D-BD-BP-DecSv1-AttDecSv2-AttDecSv3
C9	D-BD-BP-AttSv1-DecSv1-AttDecSv2-AttDecSv3
C10	D-BD-BP-AttSv1-DecSv1-AttDecSv1-AttDecSv2-AttDecSv3

The machine learning algorithms used for the classification and prediction process of benthic habitats are the RF and SVM. The RF algorithm is one of the algorithms that implement multiple decision trees that use a combination of tree predictors, where each tree depends on the value of an independent random sample vector (Belgiu & Drăguţ, 2016; Breiman, 2001). This study uses the Bagging (bootstrap aggregating) method for benthic habitat classification (Breiman, 1996). In the Bagging method, classification trees are built from different bootstrap samples, modifying predictions and building a diverse collection of predictors. Accuracy increases proportionally with the number of predictors at a certain level

(Gislason et al., 2006). For classification optimization in this study, the scales applied to the minimum node size, minimum son size, and maximum depth tree are 2, 1 and 25, respectively. The number of trees applied is 500, as recommended by Belgiu and Drăguţ (2016).

The SVM algorithm is a non-parametric technique developed from statistical learning theory (Vapnik, 1999). The SVM algorithm exploits a model based on the concept of margin maximization, where the algorithm will select a hyperplane that separates the data set into two different classes by maximizing the distance between the hyperplane and the nearest observation from the training data set (Cortes & Vapnik, 1995). In this study, the kernel function used for benthic habitat classification is the Linear Kernel. For classification optimization in this study, the scales used for the regularization parameters and tolerance values used are 1 and 1×10^{-3}, respectively.

2.4. Spatial mapping of benthic habitats and accuracy assessment

Kriging interpolation technique is used to interpolate the results of benthic habitat predictions from acoustic data into spatial maps (McLaren et al., 2019). Interpolation of benthic habitat prediction results was carried out using ArcMap 10.8.1 software. A mapping accuracy assessment was carried out to determine the accuracy of the benthic habitat map using 150 independent test data presented in a confusion matrix. There are three categories of accuracy used, namely producer accuracy (PA), user accuracy (UA), and overall accuracy (OA) (Congalton & Green, 2019).

2.5. Important parameter analysis

Parameter importance analysis aims to determine each acoustic parameter's level and order of importance to the accuracy of model prediction in the RF algorithm. The advantage of the RF algorithm is that the Out-Of-Bag (OOB) sample for each tree can be used to obtain a measure of the importance of each variable (Breiman, 2001), both for the overall classification results (Nemani et al., 2022; Shao et al., 2021; Stephens & Diesing, 2014) and for each classification class (Hasan et al., 2012; Sklar et al., 2024). In this study, parameter importance analysis was carried out to determine the important parameters in the combination of acoustic parameters with the highest accuracy.

2.6. Statistical analysis

The statistical analysis used was Analysis of Variance (ANOVA) and Z-test using XLSTAT Basic+ software (Lumivero, 2023). The ANOVA analysis aims to test the significance level of acoustic backscatter energy similarities between benthic habitat classes using the Tukey HSD (Honestly Significant Difference) method with a 5% confidence interval ($\alpha = 0.05$).

3. RESULTS AND DISCUSSION

3.1. Acoustic backscatter of training data

Figure 3 shows the variation of the acoustic backscatter energy of the first echo from the training data for the three types of benthic habitats. Overall, the energy of the acoustic parameters in the first echo (BD, BP, DecSv1, and AttDecSv1) from coral is higher than that of sand and seagrass vegetation, except for AttSv1 from sand and seagrass, which is higher than coral. However, there is no significant difference in two acoustic parameters between coral and sand, namely BD (p-value 0.41) and BP (p-value 0.37). Likewise, the energy for AttSv1 and DecSv1 between sand and seagrass are not significantly different (AttSv1, p-value 0.87; DecSv1, p-value 0.37).

In contrast to the first echo, the acoustic backscatter energy of the second and third echoes from the sand was higher than that of coral and seagrass (Figure 4). However, there was no significant difference in the acoustic backscatter energy for the AttSv2 parameter of the three benthic types in the second echo (sand vs. coral, p-value 0.872; sand vs. seagrass, p-value 0.754; coral vs. seagrass, p-value 0.971), while the other two acoustic parameters were significantly different (p-value <0.05). Likewise, in the third echo, there was no significant difference in the acoustic backscatter energy between coral and seagrass (AttSv3, p-value 0.404; DecSv3, p-value 0.985; AttDecSv3, p-value 0.952).

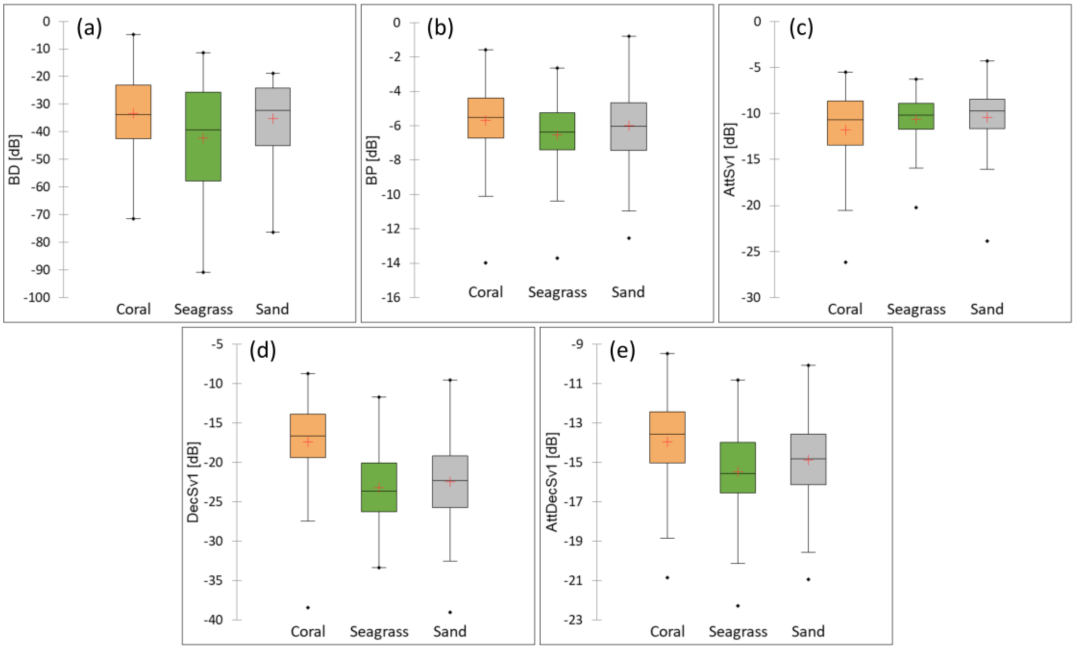

Fig. 3. Boxplot: variation of acoustic parameter energy in the first echo of the training data: (a) BD, (b) BP, (c) AttSv1, (d) DecSv1, and (e) AttDecSv1

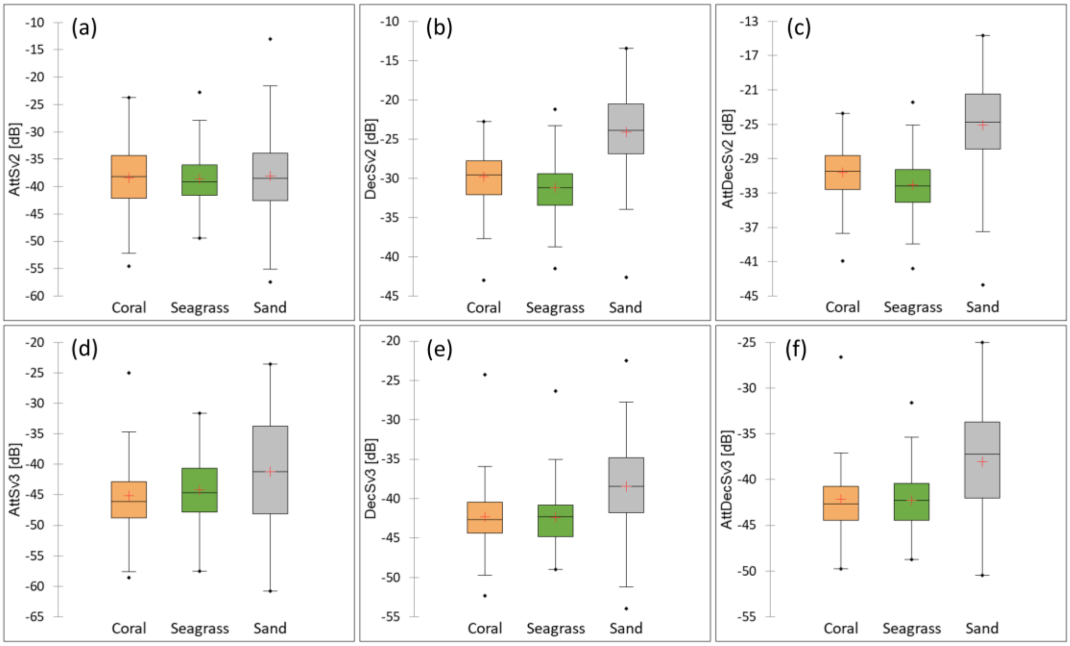

Fig. 4. Boxplot: variation of acoustic parameter energy in the second echo (top) and third echo (bottom) of the training data: (a) AttSv2, (b) DecSv2, (c) AttDecSv2, (d) AttSv3, (e) DecSv3, and (f) AttDecSv3

3.2. Map accuracy of each combination of acoustic parameters

The OA accuracy of benthic habitat maps for 68 combinations of acoustic parameters ranged from 48.67% to 79.33% (Figure 5). The OA accuracy when applying the RF algorithm ranges from 53.33% to 79.33% while using the SVM algorithm ranges from 48.67% to 78.67%. The highest OA accuracy was obtained from a combination of eight acoustic parameters (combination: C10), which was 79.33% and 78.67% when using the RF and SVM algorithms, respectively. A total of 49 combinations of acoustic parameters produced benthic habitat maps that met the minimum accuracy standards for benthic habitat mapping (≥60%) (Green et al., 2000). Five combinations of acoustic parameters of the first echo (combinations: A1, A3, A5, A8, and A13) did not meet the mapping standards when using the RF algorithm, and all combinations of the first echo acoustic parameters (14 combinations) did not meet the standards when using the SVM algorithm. This study's results indicate that more acoustic parameters can improve map accuracy. This is also shown in the study of Shao et al. (2021), where the use of seven acoustic parameters from the SBES instrument can significantly improve map accuracy compared to using only two main parameters from the SBES instrument (roughness and hardness index: E1 and E2 in RoxAnn system or DecSv1 and AttDecSv2 in this study), as in the RoxAnn system. Likewise, several studies using the multi-beam echosounder (MBES) instrument significantly increased map accuracy by adding derived parameters from the primary acoustic parameters as classification inputs (McLaren et al., 2019; Stephens & Diesing, 2014).

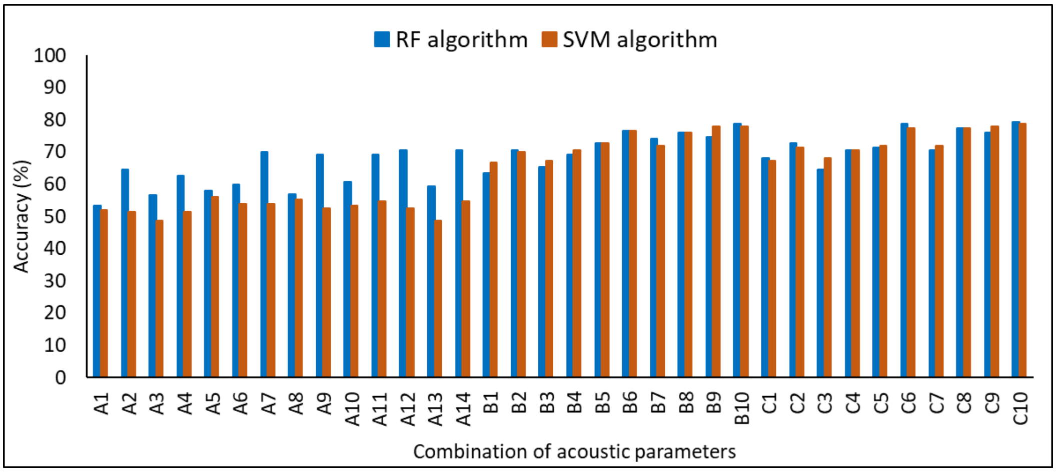

Fig. 5. Accuracy of benthic habitat maps from various combinations of acoustic parameters using different algorithms

The results of this study show that the RF algorithm performs better in producing benthic habitat maps with higher accuracy than the SVM algorithm. These results are similar to the results of a study conducted by Shao et al. (2021), where the map accuracy of the RF algorithm is higher than the SVM, Generalized Discriminant Analysis, and Decision Tree (DT) algorithms for classifying the seabed (substrate and vegetation) from SBES data. Similarly, Stephens & Diesing (2014) reported that classification tree-based algorithms (RF and DT algorithms) produced the highest accuracy for seabed sediment mapping from MBES data compared to SVM, k-Nearest Neighbor (k-NN), and Neural Networks algorithms. Hasan et al. (2012) also reported that the performance of the RF algorithm is better than the SVM, Quick Unbiased Efficient Statistical Tree (QUEST) and Maximum Likelihood Classifier (MLC) algorithms in classifying three substratum classes, but not for five biota classes where the accuracy of the SVM algorithm is more accurate. Even Diesing et al. (2014) evaluated the performance of several classification methods, where the machine learning-based method using the RF algorithm produced higher accuracy than object-based classification, the geo-statistics approach, and manual classification. Overall, the machine learning method provides better results (Gumusay et al., 2018), and this also applies to optical remote sensing data where the RF algorithm offers high accuracy for benthic habitat mapping (Misiuk & Brown, 2024; Nguyen et al., 2021).

3.3. Spatial distribution of benthic habitats

A spatial distribution map of benthic habitat from SBES data in the study area is presented in Figure 6. Both algorithms mapped seagrass more dominantly at depths <2.5 m, especially near the outer side of the atoll (eastern part of Kapota Atoll). Corals were also distributed along the outer and inner sides of the atoll, but they were also found to be associated with seagrass vegetation. Meanwhile, sand at depths <2.5 m was predominantly found in the northern part of the study area, and little was mapped in the seagrass and coral areas. At a depth range of 2.5–5 m, corals dominate along the outer and inner sides of the atoll, while sand is dominant on the inner side of the atoll to shallower waters, especially in the northern part of the study area. Meanwhile, seagrass was increasingly mapped very little with increasing water depth, and only a little was mapped at depths of 2.5–5 m. Only coral and sand were found at the 5–10 m depth range, 10-20 m and >20 m, where the RF algorithm mapped more coral than the SVM algorithm. Coral and sand are more commonly found on the inner side of the atoll at a depth of 5–10 m. Meanwhile, in the depth range of 10–20 m and >20 m, the coral on the outer side of the atoll is only mapped with a small spatial extent.

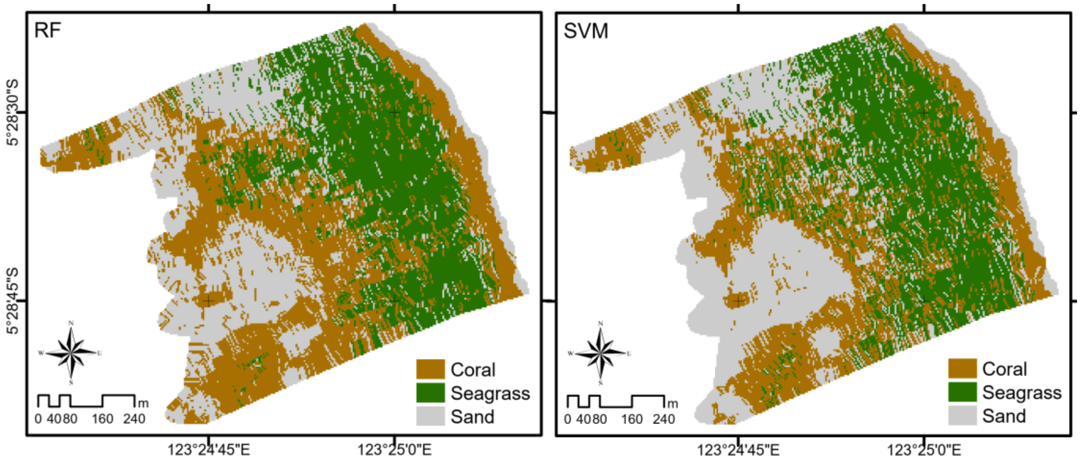

Fig. 6. Spatial map of benthic habitat using a combination of eight acoustic parameters (combination: C10) and two different classification algorithms: RF algorithm (left) and SVM algorithm (right)

Based on the confusion matrix in Table 2, no ideally classified and mapped benthic habitats were found using either the RF or SVM algorithms. Both classification algorithms can map corals with PA accuracy >80%, even the PA accuracy of the RF algorithm reaches 96.00%. The PA accuracy for sand mapping is relatively low, especially when using the RF algorithm. This is likely due to the composition of the sand benthic class in this study, which consisted of sand beds and a mixture of sand and small-sized rubble. The composition of mixed sediment is sensitive to roughness elements at half the acoustic wavelength so that it can be separated from coarse sediment based on acoustic backscattering strength (Goff et al., 2004). The presence of a small portion of gravel in the sediment causes a significant increase in acoustic backscattering, where the gravel content will dominate the acoustic backscattering response. Several studies have also shown that acoustic backscattering strength is significantly correlated with the grain size of seabed sediments (Hamuna et al., 2018; Hilgert et al., 2016; Huang et al., 2018). Therefore, rubble is estimated to impact acoustic backscatter energy in the composition of the sand and rubble mixture, so it has a relatively similar energy to the large rubble included in the corals class in this study. However, the PA accuracy of sand still meets the minimum standards for shallow water benthic habitat mapping (≥60%) (Green et al., 2000).

The three classes of benthic habitat (coral, sand, and seagrass) mapped in this study are the main benthic habitats in shallow waters. The number of classes significantly differs from the number of benthic habitat classes in Kapota Atoll mapped from Sentinel-2A imagery, with five homogeneous benthic classes and four mixed benthic classes (Hamuna et al., 2023). The number of benthic classes in this study is relatively similar to the results of the studies of McLaren et al. (2019), who mapped three benthic classes (coral class and two classes of submerged vegetation), Hamouda et al. (2019) who mapped three substrate classes (hard substrate class, and two sand classes), and Bejarano et al. (2010) who mapped four benthic classes (*Gorgonian* community, sand patch class, and two coral classes).

Tab. 2. Confusion matrix of benthic habitat mapping using two different classification algorithms on a combination of eight acoustic parameters (combination: C10)

Benthic classes	Seagrass	Sand	Coral	UA (%)	OA (%)
RF algorithm					
Seagrass	42	3	1	91.30	79.33
Sand	2	29	1	90.63	
Coral	8	16	48	66.67	
PA (%)	80.77	60.42	96.00		
SVM algorithm					
Seagrass	42	5	4	82.35	78.67
Sand	4	35	5	79.55	
Coral	6	8	41	74.55	
PA (%)	80.77	72.92	82.00		

Compared with the mapping accuracy of several previous studies, the highest OA accuracy obtained in this study is considered comparable to the accuracy of several studies that mapped benthic habitats, substrates, and vegetation using SBES instruments, such as in the studies conducted by Bejarano et al. (2010), McLaren et al. (2019), Reshitnyk et al. (2014), Riegl & Purkis (2005), and Shao et al. (2021) with maximum accuracies of 61%, 88.8%, 80.7%, 66%, and 80%, respectively. The five studies also produced several benthic classes comparable to this study (3–6 classes). A larger number of benthic classes (seven classes) with an accuracy level of 80.6% were successfully mapped in the study by (Henriques et al., 2015). Meanwhile, McIntyre et al. (2018) and Sánchez-Carnero et al. (2023) successfully mapped substrate and aquatic vegetation separately with mapping accuracy of >80% (substrate) and >70% (aquatic vegetation). However, it should be noted that some of these studies used cross-validation methods for assessing map accuracy, which can cause bias (Wadoux et al., 2021) and can produce over-estimated accuracy (Henriques et al., 2015).

3.4. Important acoustic parameters in RF algorithm

The eight acoustic parameters used as classification input have different levels of importance in the classification of benthic habitats (Figure 7). The higher the mean decrease accuracy, the higher the level of importance of the parameter in the benthic habitat classification process. The order of importance of each acoustic parameter in classifying the entire benthic habitat and for each benthic class is different. Figure 7a shows the level of importance of each acoustic parameter in the classification of the whole of the benthic habitat with a mean decrease accuracy ranging from 4.07 to 67.46. The mean decrease accuracy for the classification of each benthic class ranges from -1.16 to 23.92 for the classification of the coral class (Figure 7b), 2.75 to 54.75 for the classification of the seagrass class (Figure 7c), and 1.11 to 56.18 for the classification of the sand class (Figure 7d). Five acoustic parameters have the highest importance in benthic habitat classification as indicated by the mean decrease accuracy values and should be considered as input parameters of SBES acoustic data, namely AttDecSv2, D, DecSv1, BD, and AttDecSv3. However, only four acoustic parameters (AttDecSv2, D, DecSv1, and BD) are consistently highly important when applied to the benthic habitat type level classification. Only the AttSv1 parameter negatively influences coral classification, so it cannot be used as an input parameter for coral thematic mapping.

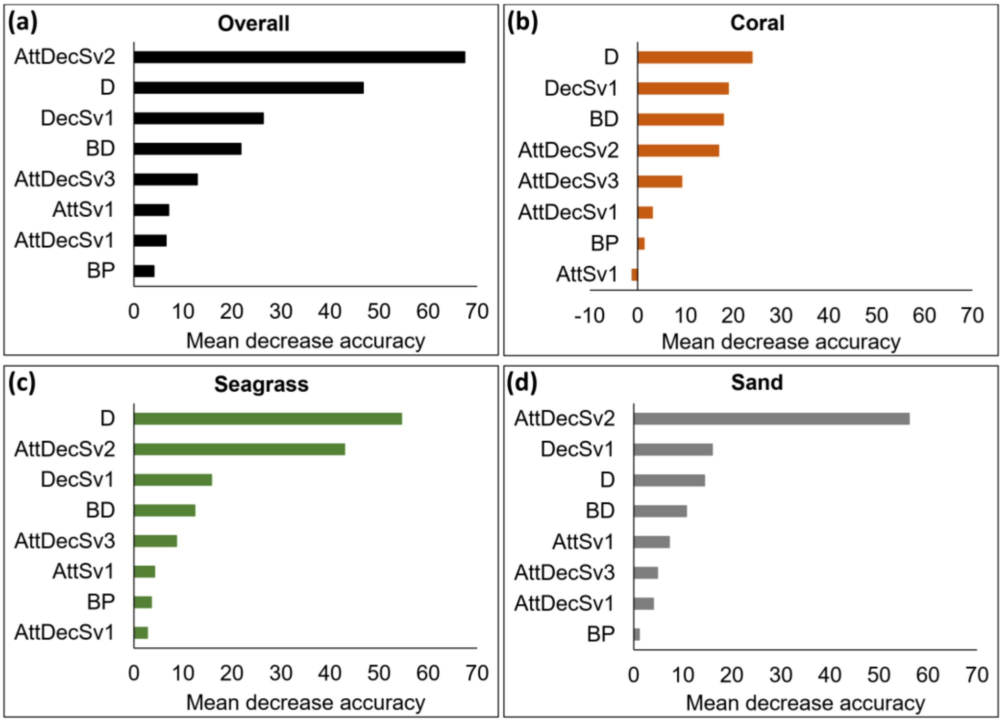

Fig. 7. Important acoustic parameters in the benthic habitat classification using the RF algorithm

AttDecSv2 (E2 in RoxAnn system) and DecSv1 (E1 in RoxAnn system) parameters as the main acoustic parameters in the SBES system are essential in classifying the seabed (Hamilton, 2001; Penrose et al., 2006). For the importance level in this study's overall classification, the hardness index (AttDecSv2) is more effective for high-accuracy classification than the seabed roughness (DecSv1). However, AttDecSv2 is very susceptible to conditions during field data acquisition (ship speed, changes in depth or slope and sea surface conditions) (Sánchez-Carnero et al., 2023). In general, the roughness and hardness of the seabed are different from aquatic vegetation, which shows that the three types of benthic can be distinguished based on their roughness and hardness levels.

In this study, parameter D (water depth) is the second most important parameter for overall classification and is the most important acoustic parameter for sand and seagrass classification. Parameter D is a counterweight to the decrease in acoustic backscatter energy with increasing depth. Several studies have shown that water depth is the most important parameter that significantly influences the results of seabed classification (Henriques et al., 2015; Nemani et al., 2022; Sánchez-Carnero et al., 2023; Shao et al., 2021). In the study of (Shao et al., 2021), parameter D plays an important role in vegetation classification because the presence of vegetation tends to decrease with increasing depth. The BD parameter indicates the energy at the beginning of one acoustic pulse length that hits the seabed (the object's surface). The BD parameter is of higher importance for overall classification and coral classification.

Another important acoustic parameter that should be considered as input in the classification process is AttDecSv3. The AttDecSv3 parameter is the cumulative energy of the third echo, such as AttDecSv2 in the second echo. Although the AttDecSv3 parameter has never been discussed in all seabed acoustic studies, the results of this study show its importance in the classification of benthic habitats. The AttDecSv3 parameter plays a more significant role in the classification of corals and seagrasses than in the classification of sand. The energy of the AttDecSv3 parameters of corals and seagrasses is significantly different from sand, so it will be easy to discriminate them.

4. CONCLUSSION

In this study, 49 combinations of acoustic parameters from 68 combinations can produce benthic habitat maps (coral, sand, and seagrass) with accuracy that meets the minimum accuracy standards for benthic habitat mapping (≥60%). The use of eight SBES acoustic parameters as classification input parameters can improve the accuracy of benthic habitat mapping with an overall accuracy of 79.33% when applying the RF algorithm and 78.67% when applying the SVM algorithm. This accuracy is higher than using only two main acoustic parameters from SBES, which are usually applied in the RoxAnn system (combination of B1; DecSv1 and AttDecSv2 parameters or E1 and E2 in the RoxAnn system indicating the roughness and hardness indices) with an accuracy of 66.67% when using the SVM algorithm and 63.33% when using the RF algorithm. Each acoustic parameter is of different importance level for benthic habitat classification. The order of importance of the eight acoustic parameters used in the overall classification of benthic habitats is AttDecSv2 > D > DecSv1 > BD > AttDecSv3 > AttSv1 > AttDecSv1 > BP. A different order of importance of acoustic parameters was obtained when classifying each benthic habitat class, namely D > DecSv1 > BD > AttDecSv2 > AttDecSv3 > AttDecSv1 > BP > AttSv1 for the classification of coral classes, D > AttDecSv2 > DecSv1 > BD > AttDecSv3 > AttSv1 > BP > AttDecSv1 for the classification of seagrass classes, and AttDecSv2 > DecSv1 > D > BD > AttSv1 > AttDecSv3 > AttDecSv1 > BP for the classification of sand classes. Overall, map accuracy can be significantly increased when two or more acoustic parameters are added simultaneously as input into the RoxAnn system classification, which uses only two main acoustic parameters (E1 or DecSv1 and E2 or AttDecSv2).

Author contributions

Baigo HAMUNA: Conceptualization, Methodology, Data curation, Visualization, Validation, Writing-Original draft preparation. Sri PUJIYATI: Conceptualization, Methodology, Supervision, Funding acquisition, Writing-Reviewing and Editing. Jonson Lumban GAOL: Conceptualization, Supervision, Writing-Reviewing and Editing. Totok HESTIRIANOTO: Conceptualization, Methodology, Supervision.

Acknowledgments

This research was financed by Directorate General of Higher Education, Research, and Technology at the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia through the Doctoral Dissertation Research grant in 2024 (Grant Number: 027/E5/PG.02.00.PL/2024 and 22165/IT3.D10/PT.01.03/P/B/2024).

Conflicts of interest

The authors declare that he has no conflicts of interest.

REFERENCES

Anderson, J. T., Van Holliday, D., Kloser, R., Reid, D. G., & Simard, Y. (2008). Acoustic seabed classification: current practice and future directions. *ICES Journal of Marine Science*, *65*(6), 1004-1011. https://doi.org/10.1093/icesjms/fsn061

Balk, H., & Lindem, T. (2015). *Sonar4 and Sonar5-Pro post processing systems: operator manual version 6.0.3*. University of Oslo. https://www.scribd.com/document/477760502/SonarX-Manual-v603-2014-12-30-pdf

Bartholomä, A., Capperucci, R. M., Becker, L., Coers, S. I. I., & Battershill, C. N. (2020). Hydrodynamics and hydroacoustic mapping of a benthic seafloor in a coarse grain habitat of the German Bight. *Geo-Marine Letters*, *40*(2), 183-195. https://doi.org/10.1007/s00367-019-00599-7

Bejarano, S., Mumby, P. J., Hedley, J. D., & Sotheran, I. (2010). Combining optical and acoustic data to enhance the detection of Caribbean forereef habitats. *Remote Sensing of Environment*, *114*(11), 2768-2778. https://doi.org/10.1016/j.rse.2010.06.012

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bravo, F., & Grant, J. (2020). Benthic habitat mapping and sediment nutrient fluxes in a shallow coastal environment in Nova Scotia, Canada. *Estuarine, Coastal and Shelf Science*, *242*, 106816. https://doi.org/10.1016/j.ecss.2020.106816

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123-140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324

Brown, C. J., Smith, S. J., Lawton, P., & 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*(3), 502-520. https://doi.org/10.1016/j.ecss.2011.02.007

Congalton, R. G., & Green, K. (2019). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third Edition (3rd ed.)*. CRC Press.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273-297. https://doi.org/10.1007/BF00994018

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., & Dove, D. (2014). Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, *84*, 107-119. https://doi.org/10.1016/j.csr.2014.05.004

Fajaryanti, R., & Kang, M. (2019). A preliminary study on seabed classification using a scientific echosounder. *Journal of the Korean Society of Fisheries Technology*, *55*, 39-49. https://doi.org/10.3796/KSFOT.2019.55.1.039

Freitas, R., Ricardo, F., Pereira, F., Sampaio, L., Carvalho, S., Gaspar, M., Quintino, V., & Rodrigues, A. M. (2011). Benthic habitat mapping: Concerns using a combined approach (acoustic, sediment and biological data). *Estuarine, Coastal and Shelf Science*, *92*(4), 598-606. https://doi.org/10.1016/j.ecss.2011.02.022

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294-300. https:/doi.org/10.1016/j.patrec.2005.08.011

Goff, J. A., Kraft, B. J., Mayer, L. A., Schock, S. G., Sommerfield, C. K., Olson, H. C., Gulick, S. P. S., & Nordfjord, S. (2004). Seabed characterization on the New Jersey middle and outer shelf: correlatability

and spatial variability of seafloor sediment properties. *Marine Geology*, *209*(1-4), 147-172. https://doi.org/10.1016/j.margeo.2004.05.030
Green, E. P., Mumby, P. J., Edwards, A. J., & Clark, C. D. (2000). *Remote sensing: handbook for tropical coastal management*. UNESCO Pub.
Gumusay, M., Bakırman, T., Tüney Kızılkaya, I., & Aykut, N. (2018). A review of seagrass detection, mapping and monitoring applications using acoustic systems. *European Journal of Remote Sensing*, *52*(1), 1-29. https://doi.org/10.1080/22797254.2018.1544838
Hamilton, L. (2001). *Acoustic Seabed Classification Systems*. DSTO Aeronautical and Maritime Research Laboratory.
Hamouda, A., Soliman, K., El-Gharabawy, S., & Nassar, M. (2019). Comparative study between acoustic signals and images for detecting seabed features. *Egyptian Journal of Aquatic Research*, *45*(2), 145-151. https:/doi.org/10.1016/j.ejar.2019.03.002
Hamuna, B., Dimara, L., Pujiyati, S., & Natih, N. (2018). Correlation of substrate fraction percentage with acoustic backscattering strength from single beam echosounder detection. *AACL Bioflux*, *11*, 1343-1351.
Hamuna, B., Pujiyati, S., Gaol, J., & Hestirianoto, T. (2023). Spatial distribution of benthic habitats in Kapota Atoll (Wakatobi National Park, Indonesia) using remote sensing imagery. *Biodiversitas Journal of Biological Diversity*, *24*(7). https://doi.org/10.13057/biodiv/d240706
Hasan, R. C., Ierodiaconou, D., & Monk, J. (2012). Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sensing*, *4*(11), 3427-3443. https://www.mdpi.com/2072-4292/4/11/3427
Henriques, V., Guerra, M. T., Mendes, B., Gaudêncio, M. J., & Fonseca, P. (2015). Benthic habitat mapping in a Portuguese Marine Protected Area using EUNIS: An integrated approach. *Journal of Sea Research*, *100*, 77-90. https://doi.org/10.1016/j.seares.2014.10.007
Hilgert, S., Kiemle, L., Fuchs, S., & Wagner, A. (2016). Investigation of echo sounding parameters for the characterisation of bottom sediments in a sub-tropical reservoir. *Advances in Oceanography and Limnology*, *7*(1). https://doi.org/10.4081/aiol.2016.5623
Huang, Z., Siwabessy, P. J., Heqin, C., & Nichol, S. (2018). Using multibeam backscatter data to investigate sediment-acoustic relationships. *Journal of Geophysical Research: Oceans*, *123*(7), 4649-4665. https://doi.org/10.1029/2017JC013638
Lee, W. S., & Lin, C. Y. (2018). Mapping of tropical marine benthic habitat: Hydroacoustic classification of coral reefs environment using single-beam (RoxAnn™) system. *Continental Shelf Research*, *170*, 1-10. https://doi.org/10.1016/j.csr.2018.09.012
Lumivero. (2023). *XLSTAT statistical and data analysis solution*. https://www.xlstat.com/en
Manik, H., Mamun, A., & Hestirianoto, T. (2014). Computation of single beam echo sounder signal for underwater objects detection and quantification. *International Journal of Advanced Computer Science and Applications*, *5*(5). https://doi.org/10.14569/IJACSA.2014.050514
McIntyre, K., McLaren, K., & Prospere, K. (2018). Mapping shallow nearshore benthic features in a Caribbean marine-protected area: assessing the efficacy of using different data types (hydroacoustic versus satellite images) and classification techniques. *International Journal of Remote Sensing*, *39*(4), 1117-1150. https://doi.org/10.1080/01431161.2017.1395924
McLaren, K., McIntyre, K., & Prospere, K. (2019). Using the random forest algorithm to integrate hydroacoustic data with satellite images to improve the mapping of shallow nearshore benthic features in a marine protected area in Jamaica. *GIScience & Remote Sensing*, *56*(7), 1065-1092. https://doi.org/10.1080/15481603.2019.1613803
Misiuk, B., & Brown, C. J. (2024). Benthic habitat mapping: A review of three decades of mapping biological patterns on the seafloor. *Estuarine, Coastal and Shelf Science*, *296*, 108599. https://doi.org/10.1016/j.ecss.2023.108599
Moszynski, M., & Hedgepeth, J. B. (2000). Using single-beam side-lobe observations of fish echoes for fish target strength and abundance estimation in shallow water. *Aquatic Living Resources*, *13*(5), 379-383. https://doi.org/https://doi.org/10.1016/S0990-7440(00)01087-1
Nemani, S., Cote, D., Misiuk, B., Edinger, E., Mackin-McLaughlin, J., Templeton, A., Shaw, J., & Robert, K. (2022). A multi-scale feature selection approach for predicting benthic assemblages. *Estuarine, Coastal and Shelf Science*, *277*, 108053. https:/doi.org/10.1016/j.ecss.2022.108053
Nguyen, T., Liquet, B., Mengersen, K., & Sous, D. (2021). Mapping of coral reefs with multispectral satellites: A review of recent papers. *Remote Sensing*, *13*(21), 4470. https://doi.org/10.3390/rs13214470

Penrose, J., Siwabessy, P. J., Gavrilov, A., Parnum, I., Hamilton, L., Bickers, A., Brooke, B., Ryan, D., & Kennedy, P. (2006). *Acoustic Techniques for Seabed Classification*. CRC for Coastal Zone, Estuary & Waterway Management.
Pijanowski, B., & Brown, C. (2022). Grand challenges in acoustic remote sensing: Discoveries to support a better understanding of our changing planet. *Frontiers in Remote Sensing*, *2*. https://doi.org/10.3389/frsen.2021.824848
Poulain, T., Argillier, C., Gevrey, M., & Guillard, J. (2011). Identifying lakebed nature: Is it feasible with a combination of echosounder and Sonar5-pro? *Advances in Oceanography and Limnology*, *2*(1), 49-53. https://doi.org/10.1080/19475721.2011.565803
Pujiyati, S., Hamuna, B., Rohilah, Hisyam, M., Srimariana, E. S., & Natih, I. N. M. (2022). Distributions of environmental parameters and Plankton's volume backscattering strength at Yos Sudarso Bay, Jayapura, Indonesia. *Egyptian Journal of Aquatic Research*, *48*(1), 37-44. https://doi.org/https://doi.org/10.1016/j.ejar.2021.08.001
Reshitnyk, L., Costa, M., Robinson, C., & Dearden, P. (2014). Evaluation of WorldView-2 and acoustic remote sensing for mapping benthic habitats in temperate coastal Pacific waters. *Remote Sensing of Environment*, *153*, 7-23. https://doi.org/10.1016/j.rse.2014.07.016
Riegl, B. M., & Purkis, S. J. (2005). Detection of shallow subtidal corals from IKONOS satellite and QTC View (50, 200 kHz) single-beam sonar data (Arabian Gulf; Dubai, UAE). *Remote Sensing of Environment*, *95*(1), 96-114. https://doi.org/10.1016/j.rse.2004.11.016
Sánchez-Carnero, N., Rodríguez-Pérez, D., Llorens, S., Orenes-Salazar, V., Ortolano, A., & García-Charton, J. A. (2023). An expeditious low-cost method for the acoustic characterization of seabeds in a Mediterranean coastal protected area. *Estuarine, Coastal and Shelf Science*, *281*, 108204. https://doi.org/10.1016/j.ecss.2022.108204
Shao, H., Kiyomoto, S., Kawauchi, Y., Kadota, T., Nakagawa, M., Yoshimura, T., Yamada, H., Acker, T., & Moore, B. (2021). Classification of various algae canopy, algae turf, and barren seafloor types using a scientific echosounder and machine learning analysis. *Estuarine, Coastal and Shelf Science*, *255*, 107362. https://doi.org/10.1016/j.ecss.2021.107362
Sklar, E., Bushuev, E., Misiuk, B., Morissette, G., & Brown, C. (2024). Seafloor morphology and substrate mapping in the Gulf of St Lawrence, Canada, using machine learning approaches. *Frontiers in Marine Science*, *11*. https://doi.org/10.3389/fmars.2024.1306396
Solikin, S., Manik, H., Pujiyati, S., & Susilohadi, S. (2018). Measurement of bottom backscattering strength using single-beam echosounder. *Journal of Physics: Conference Series*, *1075*, 012036. https://doi.org/10.1088/1742-6596/1075/1/012036
Stephens, D., & Diesing, M. (2014). A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *PLOS ONE*, *9*(4), e93950. https://doi.org/10.1371/journal.pone.0093950
Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988-999. https://doi.org/10.1109/72.788640
Vassallo, P., Bianchi, C. N., Paoli, C., Holon, F., Navone, A., Bavestrello, G., Cattaneo Vietti, R., & Morri, C. (2018). A predictive approach to benthic marine habitat mapping: Efficacy and management implications. *Marine Pollution Bulletin*, *131*(Part A), 218-232. https://doi.org/10.1016/j.marpolbul.2018.04.016
Wadoux, A. M. J. C., Heuvelink, G. B. M., de Bruin, S., & Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, *457*, 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692
Wölfl, A.-C., Snaith, H., Amirebrahimi, S., Devey, C. W., Dorschel, B., Ferrini, V., Huvenne, V. A. I., Jakobsson, M., Jencks, J., Johnston, G., Lamarche, G., Mayer, L., Millar, D., Pedersen, T. H., Picard, K., Reitz, A., Schmitt, T., Visbeck, M., Weatherall, P., & Wigley, R. (2019). Seafloor mapping – The challenge of a truly global ocean bathymetry. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00283