Performance of Different Classifiers for Marine Habitat Mapping using Side Scan Sonar and Object-Based Image Analysis

Raihanah Rusmadi [1] and Rozaimi Che Hasan[1, 2] *

[1]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.
[2]Center for Coastal and Ocean Engineering (COEI), Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.
*Email: rozaimi.kl@utm.my

Abstract. Acoustic sonar techniques have been one of the successful underwater mapping alternatives for identifying the seafloor features. The integration between the technique and classification analysis can produce detail map of the seafloor. Among these sonar technologies, side-scan sonar (SSS) is one of the tools for underwater mapping that can provide high spatial resolution seafloor mosaic which is presented in greyscale level. However, before it can be used for the coral reef marine habitat mapping, it is essential to properly assess its performance and quantify the amount of information that can be extracted. The objective of this study is to determine the accuracy of habitat maps derived using side scan sonar data, Object-based Image Analysis (OBIA) and five different classifier algorithms; Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbour (k-NN), Decision Tree, and Bayes. This study utilized side-scan sonar model Klein system 3000 which operated at 100kHz combined with video data that was conducted in shallow water (depth > 10m). First, eight (8) texture layers were derived from side scan sonar mosaic using GLCM technique. Then, the GLCM layers of texture features were reduced using Principal Component Analysis (PCA) and analysed to seek for the most contributed texture layers. A total of 80 samples were derived which consist of four (4) classes; coral, sand, silt and mud. The result shows that the Support Vector Machine (SVM) method produced the highest accuracy which is 81.25% followed by k-Nearest Neighbours (k-NN), Random Forest (RF), Decision Tree and Bayes (68.75%, 66.25%, 57.5% and 45% respectively). The used of OBIA with SSS data offers a promising method to map marine habitats for a better understanding of spatial distribution and monitoring habitat changes in the future.
Keywords: Side scan sonar, GLCM, OBIA classification

1. Introduction

Coral reefs are rich in biodiversity and very sensitive to any disturbance caused by human activities and natural phenomenon. The importance to conserve their sustainability leading to the designation of marine protected areas. To achieve this goal, conventional technique such as diving, and point sampling is not feasible to be used for large areas. Full coverage mapping is needed to produce a detail, high spatial resolution marine habitat map for better management [1,2].

Underwater acoustic technique has become an effective way for seafloor mapping. The concept of sound transmission through water column with specific frequencies ensonifies a wide range area of the seafloor for shipwreck detection, sea floor drilling for oil production and can produce detail map of seafloor which is potential for habitat mapping. Among the acoustic techniques, side-scan sonar (SSS) can be used to produce high-quality seafloor mosaics [3], and effectively can cover large seafloor area [4]. It works when the sound pulses transmitted from the transducer through the water and return back to the receiver. To link these mosaics with habitat types, ground truthing is needed and proper analysis is required [5]. The SSS mosaic is represented in greyscale mosaic where the tones of color are closely related to backscatter intensity[6]. High spatial resolution mosaic of SSS is often useful for texture analysis that could be used to differentiate and classify sediment types [7,8] and some are used for coral reef detection and mapping [9-11].

To produce detail habitat map from SSS mosaic, a classification technique is needed. Many classification techniques are available where this is mostly divided into pixel-based, and object-based image analysis. The first method has been used extensively in terrestrial remote sensing with common problem where it produces salt and pepper effect in the classification maps. OBIA has been well-developed for terrestrial [12], however there is sparse study in marine environment [13]. The main goal of this study is to compare the performance of five (5) classification techniques to be used with OBIA method for marine habitat mapping using SSS data. First, we generate texture layers from the SSS and use Principle Component Analysis (PCA) to evaluate the contribution of each layers. Then, we used and compare five (5) different supervised classification using OBIA approach to produce benthic habitat maps. Finally, the performance of all classifiers is discussed in detail with respect to the accuracies achieved.

2. Methods

The study area is located at Labuan Marine Park which is consists of three island which are Kuraman Island, Rusukan Besar Island and Rusukan Kecil Island (Figure 1).

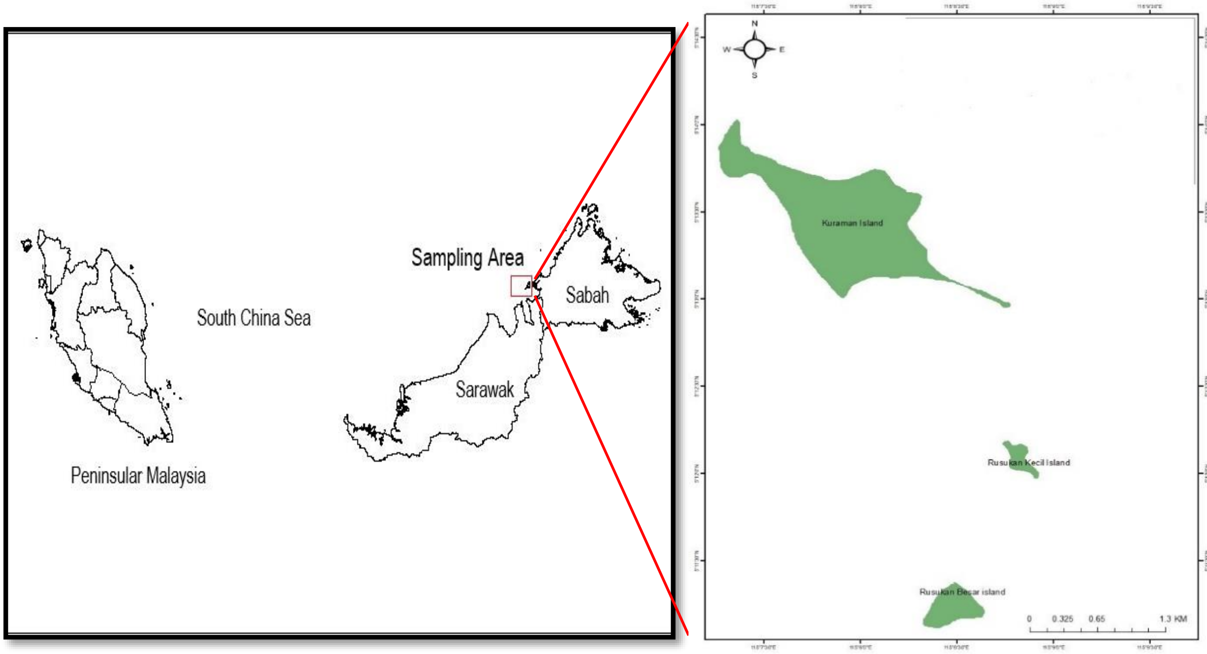

Figure 1. The location of the study site.

This park was designated as a part of Malaysia's marine park in 1994. The depth range of the study area is from ~0.3 m to ~29m. The dominant structure is silt sand, coral, rocky and the major biological habitat which is macroalgae, seagrass, and coral [14]. Apart from that, this marine park has become a tourist's attraction by its beautiful long white beach and the coral reefs for snorkelling activities and nearby shipwreck areas [15].

The acoustic surveyed using SSS was conducted from 6 to 24 of April 2017 using a Klein system 3000 SSS side mounted SSS, with an operating frequency of 100 kHz. A Wide Area Differential Global Positioning System (DGPS) receiver was used to provide position for the SSS. During the survey, a single beam echosounder was also deployed simultaneously from the same survey vessel to record the depth at the area. The raw data from SSS was processed in SonarWiz version 7 software to correct for the slant range and produced greyscale intensity mosaic at 1m spatial resolution.

Ground truthing activities through SCUBA diving, snorkelling and sediment grab were successfully conducted in July and September 2017 respectively including benthic sampling which is coral video transect (CVT), sediment grab and coral survey (point line transect). The sample sediment was collected by using ponar grab and samples were stored in zip-locked plastic bag. The samples of sediment were sieved using Particle Size Analysis (PSA) to extract the mean value of phi (Φ) which indicates the size of particle diameter. Finally, all samples were classified into four major classes consists of coral, sand, silt and mud for further analysis. Note that all ground truth samples were randomly separated into two sets; 1) to run the supervised classification, 2) for accuracy assessment.

The grey-level co-occurrence matrix [16] also known as grey-level spatial dependence matrix which is by far the most widely used approach in remote sensing for image texture features. This technique is used to compute second-order statistical textural features. There are several different textural features extracted based on this matrix. For this study, 8 (eight) textural layers were derived such as contrast, correlation, dissimilarity, energy, entropy, homogeneity, sum average and cluster shade with a window size of 50×50.

Then, Principal Component Analysis (PCA) approach for data reduction was conducted using SPSS. PCA computes a reduced set of new, linearly independent variables, called principal components (PCs) that account for most of the variance of the original variables. The PCs are a linear combination of the original variables. The PCA was based on correlation matrix, implying that the Kaiser-Guttman criterion could be applied. This means that PCs with eigen values larger than 1 were preserved as meaningful components for the analysis. The PCs were the input variables for the cluster analysis. This method is used to determine which texture layers have the most contributions to the total variance of each rotated PC, and to calculate the correlations between different texture layers with each PC. Results from this will provide a broad understanding of which texture layers are much more important.

Object based image analysis includes segmentation and classification process. First, image segmentation was conducted using eCognition software version 7 which segmenting or grouping pixel based on spectral and spatial properties [17]. Segmentation process is crucial and base of OBIA classification and its affect the accuracy of final product [18]. For this study, these parameters were used; scale 100, shape 0.1 and compact 0.7, which were initially tested and visually validated by the authors. Validation was made by visual inspection of the results from the segmentation (i.e. if the polygon consists of homogenous pixels/regions). Secondly, five supervised classification algorithms; Support Vector Machine (SVM), Random Forest, Nearest Neighbour, Decision Tree and Bayes methods were applied. SVM is a method to detect the hyperlane in an N-dimensional space (N- the number of features) that distinctly classifies the data points [19] and in this study kernel linear C is set as 2. Random Forest (RF) is an ensemble classifier that consists of many decision trees an output the class that is the mode of the classes output by individual trees. While, Bayes is a method in which take the parameters as random variables with known prior distribution. All parameters for each classifier were set as default except for NN, the *k* was set as three (3) as mentioned by Ma *et al.* [20]. Decision Tree is a classified method by repetitively subdividing it according to a tree-defined the decision framework and a class label is assigned to each observation according to the leaf node into which the observation falls [21]. Accuracy assessment of object-based mosaic analysis was produced using error matric and by calculating the Kappa statistic [22]. The result of the comparison were also presented in user's accuracy and producer's accuracy [23]. User's accuracy is the total number of the object rightly classified in each class per total number of samples that were classified in that class. While, producer's accuracy is the total number of the corrected classified object per total number of samples derived as reference in that class.

3. Results and Discussion

Acoustic mapping technique using side-scan sonar (SSS) was able to produce seafloor mosaic (Figure 2). Figure 2 shows the brightness of the greyscale mosaic which indicates the characteristics of the seafloor features. The high greyscale area can be predicted as rough and coarse area which expected as coral or rock. While the low greyscale areas indicate smooth textural properties of sedimentary. These seafloor features (coral and rocks) give a high reflectance while smooth area such as sand and mud give a low reflectance [8].

Table 1 shows the results from the PCA analysis where two (2) principal components (factors) have been identified (eigenvalue >1), explaining 81.3% of the data variance. Particularly, the first and the second factor explain 63.745% and 17.552% respectively of the total variance. The main GLCM variables that contributed to the highest variance of the PCA can be described in Table 2 which are Dissimilarity (PCA1 0.993%) and Correlation (PCA2 0.76%). These texture layer resulted from principal component analysis with the percentage eigenvalue more than 1%.

Table 1. The total variance explained from the Principal Component Analysis (PCA) analysis.

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	5.100	63.745	63.745	5.100	63.745	63.745	5.099
2	1.404	17.552	81.298	1.404	17.552	81.298	1.407
3	.757	9.466	90.763				
4	.567	7.089	97.852				
5	.158	1.972	99.824				
6	.012	.155	99.979				
7	.001	.019	99.998				
8	.000	.002	100.000				

Figure 2. The mosaic of SSS in Labuan Marine Park.

Table 2. The component matrix extracted from 2 components. The highest value in PC 1 and PC2 are highlighted in bold.

	Component	
Texture Layer	1	2
Cluster Shade	.497	-.524
Contrast	.952	-.088
Correlation	-.103	**.760**
Dissimilarity	**.993**	-.025
Energy	-.935	-.099
Entropy	.985	.068
Homogeneity	-.986	-.020
Sum Average	.367	.727

Accuracy assessment were performed for each of the habitat map where 80 samples were derived which consist of marine habitat classes such that coral, sand, silt and mud. The accuracy statistics for the fives (5) classification algorithms was presented in Figures 3 to 6. SVM achieved the highest accuracy with 81% (kappa coefficient 0.75) while the lowest accuracy was achieved by Bayes classifier (45% accuracy, kappa coefficient 0.27). Among these five algorithms, SVM was the only method that could be regarded as moderate classifier [19]. SVM is a powerful classifier and capable to handle high-dimensional dataset as it based on kernel method [24-26]. SVM can separate data belonging to either of one class as compare to k-NN which is identifying objects based on the closest training samples in the feature space. Random forest is a decision tree-based ensemble classifier. While, RF consists of a combination of decision trees in which each decision tree contributes single vote to assign the most frequent class to an input vector. For user and producer's accuracy (Figures 5 and 6), coral shows minimum misclassification from all the classifier algorithms used as compared to other classes. This is mainly due to the higher and unique reflectance caused by corals as opposed to the other classes that can be seen from the original mosaic. The habitat map derived from SVM can be described in Figure 7.

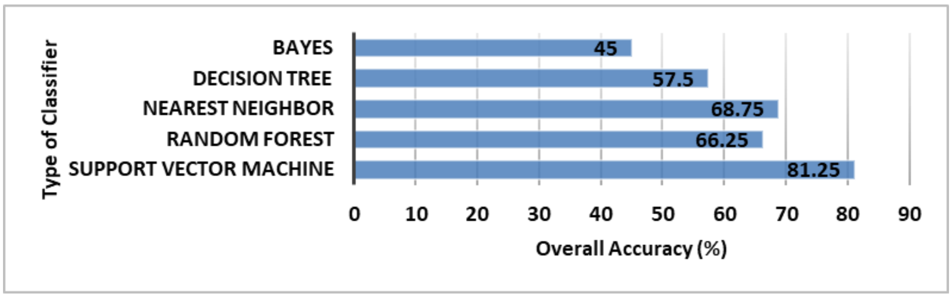

Figure 3. The Overall accuracy of five (5) classification algorithms.

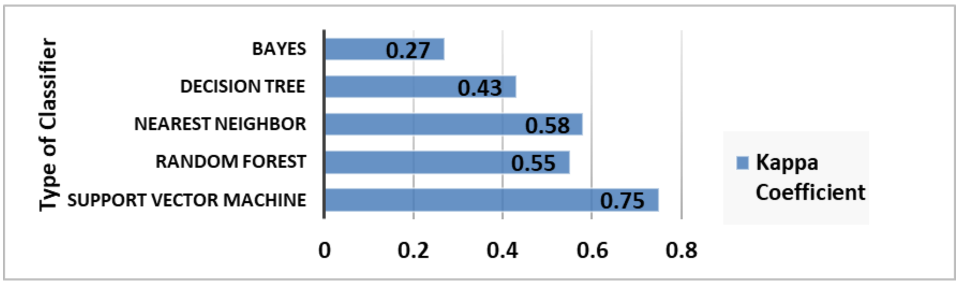

Figure 4. The Kappa Coefficient of five (5) classification algorithms.

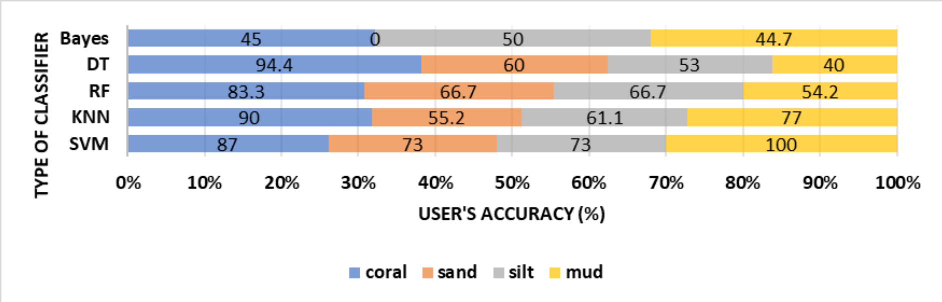

Figure 5: The User's accuracy of five (5) classification algorithms.

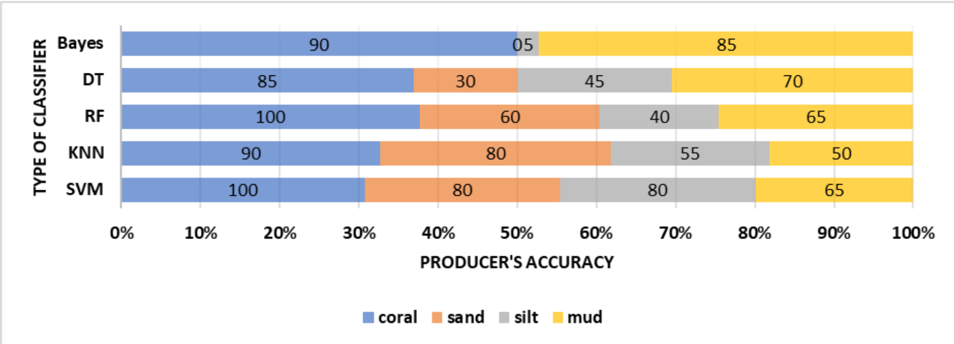

Figure 6: The Producer's accuracy of five (5) classification algorithms.

This paper aimed to create a classification map of marine habitat of Labuan Marine park by performing and comparing method from different supervised classifier algorithms. This study included analysis to select the most contributed layer using PCA that later could be used in OBIA classification method using different classifier algorithms from eCognition software. Accuracy of object-based image classification can be improved by running different segmentation parameter settings [27,28]. This is always determined by experimental trial-and-error to find the optimal value [29,30]. However, this was not the focus of this study and the selection of parameter is done using visual interpretation of segmentation results. It is expected that the accuracy of all classifiers tested in this study might be improved if detail segmentation parameters were completed. Based on Figure 15, most of the corals are found near to the shoreline. This is due to the fact that the corals live in clear and shallow water where lots of sunlight for photosynthesis through their symbiotic algae [31]. This is also supported by the study from Mustajap *et al.* [14] which stated that major sediments such as silt and sand are dominant classes, while coral is located near to the shoreline.

effective ocean protection *Marine Policy* **91** 11-3

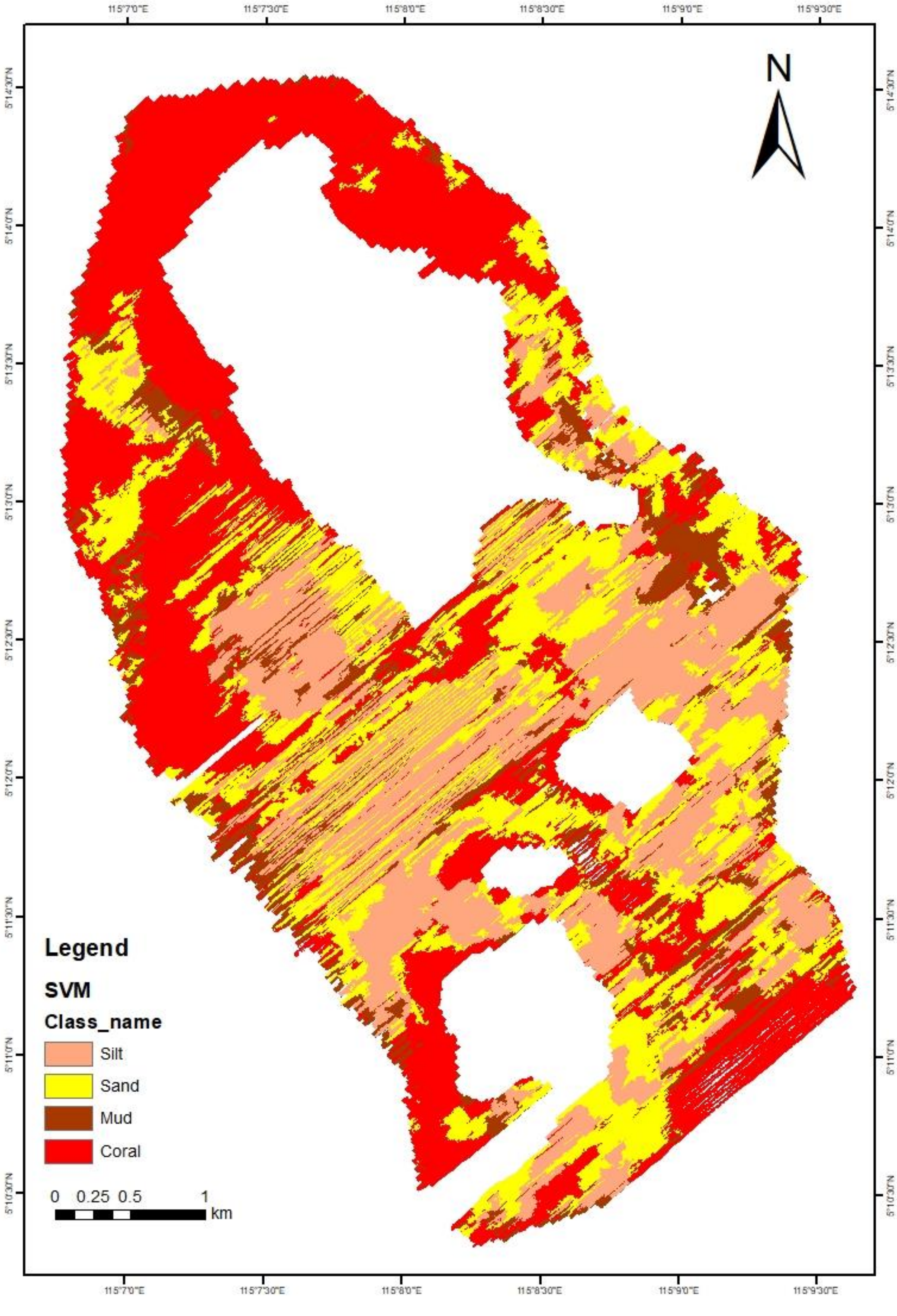

Figure 7. Habitat maps derived from Support Vector Machine (SVM).

Most of the habitat maps derived in this study shows high level of noise which also appear in the original SSS mosaic. We speculate that this is due to the way SSS was operated (i.e. SSS was side mounted in this study). Normally, SSS system is towed close to the seabed where good quality of seafloor image can be recorded [32]. However, this approach is not feasible to be used when working in shallow water region such as in Labuan Marine Park. Due to this, side mounted SSS was used where it could cover larger area, but with lower quality mosaic. As the depth increased, water column (i.e. gap of the nadir region) became larger and prone to noises that could give distortion to the mosaic quality.

4. Conclusion

In this paper, we have tested five different supervised classifiers with OBIA to produce habitat map using side-scan sonar and ground truth data. GLCM Dissimilarity and GLCM Correlation texture layers found to be more important for the classification process. In terms of accuracy, SVM performs the best compared to other classifiers. The method to produce high spatial resolution marine habitat maps in this study might be useful for the establishment of a baseline information for seabed habitat monitoring.

Acknowledgements

The authors would like to thank the Ministry of Education (MOE) and Universiti Teknologi Malaysia for providing fund to carry out this research (vot. No. 19H32 *Determination of potential automated classification method for shallow water coral reef habitat mapping*). Special thanks to project team members from *Institut Oseanografi dan Sekitaran* (INOS), Universiti Malaysia Terengganu and from the National Hydrographic Centre (NHC).

References

[1] Sala E, Lubchenco J, Grorud-Colvert K, Novelli C, Roberts C and Sumaila U R 2018 Assessing real progress towards effective ocean protection *Marine Policy* **91** 11-3
[2] Smith W, Marks K and Schmitt T 2017 Airline flight paths over the unmapped ocean *EOS* **98**
[3] Burguera A and Oliver G 2016 High-resolution underwater mapping using side-scan sonar *PloS one* **11** e0146396
[4] Kaeser A J, Litts T L and Tracy T W 2013 Using Low-Cost Side-Scan Sonar For Benthic Mapping Throughout The Lower Flint River, Georgia, Usa *River Research and Applications* **29** 634-44
[5] Herbert R J and Saunders J 2017 Management of marine ecosystems *Marine and Coastal Resource Management: Principles and Practice* 148
[6] de Barros Pereira T R, Vital H and da Silva A G A 2017 Classification of textural distribution of seabed using hydroacoustics data: Case study of Ponta Negra beach, Natal/RN-Brazil. In: *2017 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics)*: IEEE) pp 1-6
[7] Hamill D, Buscombe D and Wheaton J M 2018 Alluvial substrate mapping by automated texture segmentation of recreational-grade side scan sonar imagery *PloS one* **13** e0194373
[8] Blondel P 2010 *The handbook of sidescan sonar*: Springer Science & Business Media)
[9] Collier J and Humber S 2007 Time-lapse side-scan sonar imaging of bleached coral reefs: a case study from the Seychelles *Remote Sensing of Environment* **108** 339-56
[10] Degraer S, Moerkerke G, Rabaut M, Van Hoey G, Du Four I, Vincx M, Henriet J-P and Van Lancker V 2008 Very-high resolution side-scan sonar mapping of biogenic reefs of the tube-worm Lanice conchilega *Remote Sensing of Environment* **112** 3323-8
[11] Bollinger M A and Kline R J 2017 Validating Sidescan Sonar as a Fish Survey Tool over Artificial Reefs *Journal of Coastal Research* **33** 1397-407
[12] Blaschke 2010 Object based image analysis for remote sensing *ISPRS journal of photogrammetry and remote sensing* **65** 2-16
[13] Lucieer V L 2008 Object-oriented classification of sidescan sonar data for mapping benthic marine habitats *International Journal of Remote Sensing* **29** 905-21
[14] Mustajap F, Saleh E, Madin J and Hamid S A 2015 Marine habitat mapping at Labuan Marine Park, Federal Territory of Labuan, Malaysia *Ocean Science Journal* **50** 291-8

[15] Malaysia D o M P 2017, December 19 Labuan.
[16] Haralick R M, Shanmugam K and Dinstein I 1973 Textural Features for Image Classification *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3** 610-21
[17] Benz U C, Hofmann P, Willhauck G, Lingenfelder I and Heynen M 2004 Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information *ISPRS Journal of Photogrammetry and Remote Sensing* **58** 239-58
[18] Lippitt C D, Coulter L L, Freeman M, Lamantia-Bishop J, Pang W and Stow D A 2012 The effect of input data transformations on object-based image analysis *Remote Sensing Letters* **3** 21-9
[19] Congalton R G and Green K 2002 *Assessing the accuracy of remotely sensed data: principles and practices*: CRC press)
[20] Ma C-M, Yang W-S and Cheng B-W 2014 How the parameters of k-nearest neighbor algorithm impact on the best classification accuracy: In case of parkinson dataset *Journal of Applied Sciences* **14** 171-6
[21] Friedl M A and Brodley C E 1997 Decision tree classification of land cover from remotely sensed data *Remote Sensing of Environment* **61** 399-409
[22] Congalton R G and Mead R A 1983 A quantitative method to test for consistency and correctness in photointerpretation *Photogrammetric Engineering and Remote Sensing* **49** 69-74
[23] Van Genderen J, Lock B and Vass P 1978 Remote sensing: statistical testing of thematic map accuracy *Remote Sensing of Environment* **7** 3-14
[24] Mountrakis G, Im J and Ogole C 2011 Support vector machines in remote sensing: A review *ISPRS Journal of Photogrammetry and Remote Sensing* **66** 247-59
[25] Eugenio F, Marcello J and Martin J 2015 High-resolution maps of bathymetry and benthic habitats in shallow-water environments using multispectral remote sensing imagery *IEEE Transactions on Geoscience and Remote Sensing* **53** 3539-49
[26] Zhang C 2015 Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem *ISPRS journal of photogrammetry and remote sensing* **104** 213-23
[27] Wahidin N, Siregar V P, Nababan B, Jaya I and Wouthuyzen S 2015 Object-based Image Analysis for Coral Reef Benthic Habitat Mapping with Several Classification Algorithms *Procedia Environmental Sciences* **24** 222-7
[28] Benfield S, Guzman H, Mair J and Young J 2007 Mapping the distribution of coral reefs and associated sublittoral habitats in Pacific Panama: a comparison of optical satellite sensors and classification methodologies *International Journal of Remote Sensing* **28** 5047-70
[29] Munyati C 2018 Optimising multiresolution segmentation: delineating savannah vegetation boundaries in the Kruger National Park, South Africa, using Sentinel 2 MSI imagery *International Journal of Remote Sensing* **39** 5997-6019
[30] Kavzoglu T and Tonbul H 2017 A Comparative study of segmentation quality for multi-resolution segmentation and watershed transform. In: *2017 8th International Conference on Recent Advances in Space Technologies (RAST)*: IEEE) pp 113-7
[31] Goreau T F, Goreau N I and Goreau T J 1979 Corals and coral reefs *Scientific American* **241** 124-37
[32] Grządziel A, Felski A and Wąż M 2015 Experience with the use of a rigidly-mounted side-scan sonar in a harbour basin bottom investigation *Ocean Engineering* **109** 439–43