Accuracy assessment of several classification algorithms with and without hue saturation intensity input features on object analyses on benthic habitat mapping in the Pajenekang Island Waters, South Sulawesi

T Pragunanti1, B Nababan2, H Madduppa2 and D Kushardono3

1Marine Technology Study Program, Graduate School IPB University, Bogor
2Department of Marine Technology, Faculty of Fisheries and Marine Science, IPB University, Bogor
3Remote Sensing Center, National Aeronautics and Space Agency, Jakarta

*e-mail: turissa_pragunanti@apps.ipb.ac.id

Abstract. Object-based image analysis (OBIA) method has been proven to improve the accuracy value on benthic habitat mapping. The purpose of this study was to assess the accuracy of several classification algorithms on benthic habitat mapping based on OBIA method with and without input feature of Hue Saturation Intensity (HSI) in the Pajenekang island waters, South Sulawesi, Indonesia. Sentinel-2A satellite imagery with 10 m² spatial resolution acquired on 3 September 2018 was used in this study. During OBIA analyses, we segmented the object into 5, 10, and 15 classes and treated each of them with input features of mean+ratio vs mean+ratio+hue saturation intensity (HSI). We later classified the benthic habitat by applying several classification algorithms such as the Bayes, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Decision Tree (DT). The results showed that the Bayes algorithm produced highest accuracy of 78.35% within 10 segmentation classes and input features of mean+ratio+HSI followed by the KNN of 71.13% with 5 segmentation classes and input features of mean+ratio+HSI. The addition of HSI input features into OBIA analyses increased the accuracy of benthic habitat classification mapping of 4.13% with the Bayes classification algorithm.

1. Introduction
Seagrass ecosystem plays an important role in a habitat benthic and coastal ecosystem. Spatial information on seagrass distribution is important because it is related to the diversity of shallow water habitat ecosystems and the survival of associated marine biota. Seagrass are also one of the primary producers in the ocean for food chain. Growth, morphology, abundance and primary productivity of seagrass in a water are generally determined by the availability of phosphate, nitrate and ammonium nutrients [1].

Spatial information on seagrass distribution is very important in monitoring and management seagrass growth, health, and abundance. However, spatial information on seagrass ecosystems in Indonesian waters, especially on the Pajenekang island waters is very limited. On the other hand, various human activities are trampling seagrasses, laying the ship's body, oil waste from passing boats, and muddy waters which can cause damage to seagrass ecosystem [2]. Pajenekang Island is one of the islands of the Spermonde archipelago of South Sulawesi, data and research on the distribution of...
seagrass in the area is still limited specifically on the spatial information on seagrass health, abundance, and coverage.

Along with the development of remote sensing technology and increasing the resolution of satellite image data, it is necessary to apply various methods of image extraction and classification methods to obtain higher precision and accurate information. In general, the image extraction method has two approaches namely pixel based and object based. Based on some previous researches conducted for mapping the composition of benthic habitats, object-based image extraction techniques (OBIA) showed more effective results compared to pixel-based techniques [3][4] because they did not only pay attention to the pixel scale but also were able to define object classes based on aspects spectral and spatial aspects simultaneously [5].

The application of several classification algorithms, optimization of segmentation scales and input features is one way to obtain optimum accuracy. Naive Bayes and K-Nearest Neighbor are well-known classification methods with good levels of accuracy. Many studies have been conducted relating to the classification method. Generally, Bayes and KNN methods that produce good accuracy (> 50%) in the field of land mapping [6] while in benthic habitat mapping are generally used SVM algorithm or Bayes, and KNN. [7] In mapping benthic habitats on Wangi-Wangi island waters using Bayes, KNN, SVM, and DT algorithms, SVM algorithm produced better accuracy than other algorithms. In this study, we conducted classification of benthic habitat based on object segmentation and several classification algorithms in Pajenekang island waters.

The purposes of this study were to (1) map the distribution of seagrass habitats using the OBIA method applying several classification algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Bayesian, Decision Tree (DT) with and without HSI input feature, (2) determine the accuracy classification result applying input features of mean+ratio vs mean+ratio+HSI (Hue Saturation Intensity) with the various classification algorithms.

2. Material and methods

Study area of this research was conducted on benthic habitat which contain seagrass ecosystem of Pajenekang island waters, South Sulawesi (Figure 1). The equipment used in this study consisted of computers that were supported by several data processing software and satellite image data analyses. Equipment used during field observations were Garmin 62S Handheld GPS, quadrant transects measuring 1 x 1 m², underwater cameras, and related stationeries. We used Sentinel Level 2A satellite image data with a spatial resolution of 10 m² acquired on 3 September 2018. The sampling point area was marked by a green circle around the island waters.

Figure 1. The coastal waters of Pajenekang island in Pangkep Regency, South Sulawesi.
2.1. **Field data collection**

Field data collection of shallow water habitat data were conducted using the stratified random sampling method. Transect photo technique was taken perpendicularly from above with transect size 1x1 m². Data collection in the field were carried out on transects with zigzag design around the study area. For this study, we collected 196 sampling points field data that would be used for classification and accuracy test.

2.2. **Atmospheric and geometric correction**

The atmospheric correction method used was Dark object Subtraction (DOS) with the following formula:

\[ Rc = Rs – Rsi, \]

where \( Rc \) = Object reflectance value which is atmospherically corrected, \( R \) = Object reflectance value before correction, \( Rsi \) = \( \text{Mean Rw} – (2 \times \text{Standard Deviation Rw}) \), and \( Rw \) = spectral value which is considered as an offset [8].

The geometric correction was carried out using a transformation method, which was adjusted to the UTM projection system and the field control point or GCP (Ground Control Point) which coordinates were determined from coordinating the coordinates on the field via GPS.

2.3. **Image classification**

The classification method used in this study was OBIA. The OBIA method was a classification approach process that considers the spectral and spatial aspects of objects. This method went through a segmentation process in grouping adjacent pixels of the same quality. Generally, the OBIA method went through two stages, namely image segmentation and classification in each segment [9].

The segmentation process in this study used a multiresolution segmentation (MRS) algorithm by using several trial and error scale values (trial and error), namely scales 5, 10, and 15. The optimization of the segmentation scale in the segmentation process was carried out to obtain the value of the segmentation scale that can produce the best result. The results of the application of some segmentation scale values indicated that the segmentation scale parameters greatly affected the process of forming objects in the image both the number and shape of objects. [4] had proven that the effect of the scale of segmentation can affect the shape, size, and number of objects produced. In addition to the segmentation process, the image classification process was also performed by testing several satellite image classification algorithms such as Bayes, K-Nearest Neighbor (KNN), Super Vector Machine (SVM), and Decision Tree (DT). The four algorithms were compared to get the best accuracy in mapping the habitat benthic.

In this study, we also compared a number of input feature variables in the classification using the OBIA method to obtain optimum accuracy results. Mean, ratio, and HSI were used as input features. We compared input features between mean+ratio vs mean+ratio+HSI.

Mean, ratio, and HSI were part of image objects that had spectral values, shapes and hierarchical characteristics in each band that were used as sources of information for classifying certain class objects [10]. Mean returns the average feature value of all objects from the selected domain or band. Ratio is an input feature made using a comparison between a combination of two selected bands (for example, band 1/2 or 1/3). While HSI is an image transformation that can strengthen the spectral response of an object and distinguish it from other water-based substrates [11].

2.4. **Assessment Accuracy**

Accuracy test was carried out on the classified images so that the accuracy values of the applied algorithms can be determined. In general, the error matrix (error matrix) was an accuracy test that was usually used in testing data from remote sensing classification results [12]. In addition, kappa coefficient values can also be used to illustrate the value of agreement between field data and the results of the satellite image classifications.

The accuracy test formula used [12] confusion matrix that results in Overall Accuracy (OA), User Accuracy (UA), Producer Accuracy (PA), Kappa coefficient, and Z statistic values. The accuracy test can be calculated with the following equation:
\[ OA = \frac{\sum_{i=1}^{k} n_{ii}}{n} \]  

(1)

\[ PA = \frac{n_{ij}}{n+j} \]  

(2)

\[ UA = \frac{n_{ii}}{n+i} \]  

(3)

\[ k \text{ is the number of rows in the matrix, } n \text{ is the number of observations, } n_{ii} \text{ is the number of observations in the } i \text{ column and the } i \text{ row and } n_{jj} \text{ is the number of observations in the } j \text{ column and } j- \text{row. Whereas the calculation of Kappa coefficient values is in the range of 0-1 and is smaller than the overall accuracy value which can be calculated using the following formula:} \]

\[ K = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{ii} x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{ii} x_{+i})} \]  

(4)

which according to [13] \( r \) is the number of rows in the matrix, \( x_{ii} \) is the number of observations of row \( i \) and column \( i \), and \( x_{+i} \) and \( x_{+i} \) is the total margin of row \( i \) and column \( i \), while \( N \) is the total number of observations (overall accuracy).

3. Result and discussion

3.1. Classification scheme

Based on the field observations, we produced 4 classes for the shallow water benthic habitat i.e., seagrass, coral, sand, and rubble. The benthic habitat component was arranged based on the dominant benthic habitat component in the field. [14] explained that the determination of benthic habitat classification scheme was based on the dominant cover of benthic habitat components obtained from field observations on quadratic transects both visually and with the help of quadratic transect photos.

3.2. Comparison algorithm and the application of HSI

During OBIA analyses, we segmented the image within several scales by trial and error until we found the maximum accuracy. Based on the trial and error studies, we found that the object segmentation scales were 5, 10, and 15. Within this OBIA analyses we also added some input features such as mean+ratio vs mean+ratio+HSI. Furthermore, we applied several classification algorithms (The Bayesian, SVM, KNN, and DT) to classify the benthic habitat into 4 (four) classes (seagrass, coral, sand, and rubble) and determined their accuracies.

Based on the application of the above segmentation scales, input features, and the application of the classification algorithms on the Sentinel satellite image, the results showed that the Bayesian Algorithm with 10 segmentation scales and input features of mean+ratio+HSI produced the highest accuracy value for classification benthic habitat map of 78.35% (Figure 2, 3). The KNN algorithm produced second best accuracy value of 71.13% with 5 segmentation scales followed by SVM and DT both with accuracy values of 70.10% with 10 segmentation scales (Figure 3). The highest accuracy value produced by the Bayesian algorithm can be explained because the Bayes algorithm is an algorithm that performs welding based on simple probabilities and is designed to be used on the grounds that between one class and other classes are not interdependent (independent) [15] [16] [17] also explained that the benefits of classifying objects by using the Bayes algorithm can reduce the error value when the data set is large and proven to have high accuracy and speed when applied to a large number of data sets. In this study, the number of field observations was 196 sampling points so and this was in accordance with the Bayesian algorithm that can be suitable for large amounts of data.
Figure 2. Accuracy results with input feature of mean+ratio of several classification algorithms.

Figure 3. Accuracy results with input feature of mean+ratio+HSI of several classification algorithms.

For KNN algorithm, it classified based on the closest distance from the classified object. According to [18] the Nearest Neighbor algorithm (K-Nearest Neighbor or K-NN) is an algorithm that classifies based on the proximity of the location (distance) of a data with other data. Therefore, when applied to the KNN classification algorithm, it produced higher accuracy on the 5-segmentation scale that formed small segmented objects adjacent to one another.

When viewed from the classification of seagrass distribution maps produced, the classification map that applies the KNN algorithm has a large training data. The advantage of KNN is that it is very compatible with training data that has a lot of noise and is effective if the training data is large [17].

The results of this study also showed the difference in accuracy classification map between input features of HSI and without HSI. Accuracy results with input feature of HSI produced the highest accuracy value (78.35%) with The Bayes algorithm compared with without input features of HSI of 74.22% with the Bayes algorithm (Figure 3) HSI (Hue Saturation Intensity) is one of the image transformations that can strengthen the spectral response of an object and can distinguish it from other substrate waters [11].

The application of HSI was thought to strengthen the spectral reflection, so that the spectral value captured by the object can be clearer. Previous researchers [11] explained that compared with other image transformations such as Lyzenga and PC 1, the transformation of HSI was able to strengthen or accentuate spectral reflection in shallow water ecosystems such as coral reef ecosystem. This was certainly not much different when applied to seagrass which was also a part of shallow water ecosystem.

From these results, it can be proven that the Bayes was the best classification algorithm for benthic habitat mapping in Pajenekang island waters. Meanwhile, some factors affecting the accuracy of benthic
habitat classification map in this study could be influenced by the complexity of the benthic habit in this case seagrass found in the study site and the number of observation points, the similarity factor of spectral values according to [14] cannot be avoided by the classification algorithm, especially the benthic habitat class which was arranged by several benthic habitat classes in the observation location, and the high complexity of benthic habitat as well as the mismatch between GPS accuracy and spatial resolution of images [19].

The results of this study also showed that the benthic habitat distribution on Pajenekang island was almost homogeneously distributed in certain areas, specifically for seagrass ecosystems on the Pajenekang island waters. Five species of seagrass were found in the study region i.e., *Enhalus acoroides*, *Thalassia hemprichii*, *Cymodocea rotundata*, *Halophila ovalis*, and *Halinerule uninervis* which were scattered in a colony. [20] mentioned that for tropical waters like Indonesia, seagrass beds were more dominant to grow with colonies consisting of several species (mix species) in a certain region. Unlike the temperate or cold regions which are mostly dominated by one species of seagrass. The distribution of seagrass varied greatly depending on the beach topography and tidal patterns.

The Bayes classification results at optimum accuracy using the input feature implementation with and without HSI showed that the dynamics of addition or reduction of area in the four benthic habitat classes. The area of seagrass, coral, sand, and rubble in sequence before the application of HSI was 41.49 ha, 27.52 ha, 18.02 ha, and 1.36 ha, respectively. After HSI treatment, the areas of seagrass, coral, sand, and rubble were 41.25 ha, 27.74 ha, 18.12 ha, and 1.28 ha respectively (Figure 4 and 5). Changes in area before and after the application of HSI indicated that there was a strengthening or protrusion in the spectral reflection of benthic habitat classes that had been applied by HSI that provide certain information in the classification process. Meanwhile, for KNN and DT classification results there were no change in the area before and after the application of his.

Some researchers used the SVM classification algorithm in mapping benthic habitats of shallow marine waters including seagrass. According to [21] in the field of remote sensing, better accuracy values was produced by the SVM algorithm compared to other classification techniques. However, in this study after experimenting with several classification algorithms including SVM, the best accuracy value was obtained by applying the Bayes algorithm followed by KNN algorithm.

![Figure 4. Benthic habitat map classification by applying Bayes algorithm, input feature mean+ratio+HSI on several segmentation scales: (a) 5; (b) 10; (c) 15.](image-url)
Figure 5. Benthic habitat map classification by applying Bayes algorithm, input feature mean+ratio on several segmentation scales: (a) 5; (b) 10; (c) 15.

4. Conclusion
Hue saturation intensity (HSI) played an important factor to enhance the accuracy value of benthic habitat mapping of 4.13% with the Bayes classification algorithm. The Bayesian algorithm produced highest accuracy of 78.35% within 10 segmentation classes and input features of mean+ratio+HSI followed by the KNN of 71.13% with 5 segmentation classes and input features of mean+ration+HSI.

5. References
[1] Green E P and Short F T 2003 World Atlas of Seagrasses (University of California Press)
[2] Fähruddin M, Yulianda F and Setyobudiandi I 2017 Kerapatan dan penutupan lamun di pesisir Desa Bahoi, Sulawesi Utara Jurnal Ilmu dan Teknologi Kelautan Tropis 9 1 375-383
[3] Phinn S R, Roelfsema C M and Mumby P J 2011 Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs International Journal Remote Sensing 33 3768-3797
[4] Wahidin N, Siregar V P, Naban B, Jaya I and Wouthuyzend S 2015 Object-based image analysis for coral reef benthic habitat mapping with several classification algorithms Procedia Environmental Sciences 24 222-227
[5] Danoedoro P 2012 Pengantar Penginderaan Jauh Digital (Indonesia: Andi Offset) 398 p
[6] Khotimah N and Istiawan D 2018 Perbandingan algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour untuk prediksi lahan kritis di Kabupaten Pemalang The 7th University Research Colloquium 2018 (Indonesia: STIKES PKU Muhammadiah Surakarta, Pemalang)
[7] Mastu L K 2018 Pemetaan Habitat Bentik Berbasis Objek Menggunakan Citra Unmanned Aerial Vehicle (UAV) dan Satelit Sentinel-2 di Perairan Pulau Wangi-Wangi Kabupaten Wakatobi (Indonesia:Institut Pertanian Bogor)
[8] Dewa R P and Danoedoro 2017 The effect of image radiometric correction on the accuracy of vegetation canopy density estimate using several Landsat-8 OLI’s vegetation indices: A case study of Wonosari area, Indonesia Proceeding of IOP Conference Series: Earth and Environmental Science 54 012046
[9] Zhang C 2014 Applying data fusion techniques for benthic habitat mapping ISPRS Journal of Photogrammetry and Remote Sensing 104 213-223
[10] Trimble 2014 Ecognition developer: user guide Munchen, Germant: Trimble Germany GmbH, 262 p
[11] Helmi M, Hartoko A, Herkiki S, Munasik and Wouthuyzen 2011 Analisis respon spektral dan ekstraksi nilai spektral terumbu karang pada citra digital multispektral satelit ALOS-AVNIR di Perairan Gugus Pulau Pari, Kepulauan Seribu, Jakarta Buletin Oseanografi Marina 1 120-136
[12] Congalton R G and Green K 2009 Assessing the accuracy of remotely sensed data principles and practices (second edition) CRC Press: Taylor & Francis Group France 183p
[13] Whiteside T G, Boggs G S and Maier S W 2011 Comparing object-based and pixel-based classification for mapping savannas International J. of Applied Earth Observation and Geoinformation 13 6 884-893
[14] Mastu L K, Nababan B and Panjaitan J P 2018 Pemetaan habitat bentik berbasis objek menggunakan citra Sentinel-2 di perairan Pulau Wangi-Wangi, Kabupaten Wakatobi J. Ilmu dan Teknologi Kelautan Tropis 10 2 381-396
[15] Putri R E, Suparti and Rahmawati R 2014 Perbandingan metode klasifikasi Naive Bayes dan K-Nearest Neighbour pada analisis data status kerja di Kabupaten Dema tahun 2012 Jurnal Gaussian 3 4 831-838
[16] Berry I H and Browne M 2006 Lecture Notes in Data Mining(USA: World Scientific)
[17] Han J and Kamber M 2006 Data Mining Concepts and Techniques, Second Edition (California: Morgan Kaufman)
[18] Prasetyo E 2012 Data Mining Konsep dan Aplikasi Menggunakan MATLAB (Indonesia: ANDI Yogyakarta)
[19] Anggoro A, Siregar V P, Agus S B 2017 Multiscale classification for geomorphic zone and benthic habitats mapping using OBIA method in Pari Island Jurnal Penginderaan Jauh 14 2 89-93
[20] Azkab M H 2006 Ada apa dengan lamun Majalah Semi Populer Oseana 31 3 45-55
[21] Mountrakis G, Im J and Ogole C 2011 Support vector machines in remote sensing: A review ISPRS J. Photogramm. 66 3 247-259