MACHINE LEARNING-BASED MANGROVE LAND CLASSIFICATION ON WORLDVIEW-2 SATELLITE IMAGE IN NUSA LEMBONGAN ISLAND

Aulia Ilham[1]* and Marza Ihsan Marzuki[2]
[1]Program Studi Oseanografi, Fakultas Ilmu dan Teknologi Kebumian, Institut Teknologi Bandung,
Jl. Ganesha No. 10, Bandung 40132, Indonesia.
[2]Badan Riset dan Sumber Daya Manusia, Kementerian Kelautan dan Perikanan,
Jl. Pasir Putih 1, AncolTimur, Jakarta 14430, Indonesia.
*e-mail: aulia.ilham27@gmail.com

Received: 7 November 2017; Revised: 25 November 2017; Approved: 26 December 2017

Abstract. Machine learning is an empirical approach for regressions, clustering and/or classifying (supervised or unsupervised) on a non-linear system. This method is mainly used to analyze a complex system for wide data observation. In remote sensing, machine learning method could be used for image data classification with software tools independence. This research aims to classify the distribution, type, and area of mangroves using Akaike Information Criterion approach for case study in Nusa Lembongan Island. This study is important because mangrove forests have an important role ecologically, economically, and socially. For example is as a green belt for protection of coastline from storm and tsunami wave. Using satellite images Worldview-2 with data resolution of 0.46 meters, this method could identify automatically land class, sea class/water, and mangroves class. Three types of mangrove have been identified namely: *Rhizophora apiculata*, *Sonnetaria alba*, and other mangrove species. The result showed that the accuracy of classification was about 68.32%.

Keywords: *clustering, machine learning, remote sensing data*

1 INTRODUCTION

Remote sensing system consists of data collection (image) by the sensor, followed by the initial processing of image, analysis and extraction of information to produce thematic maps that will be further utilized by the user for various purposes (Wicaksono 2009). In General, remote sensing systems can be distinguished by the energy source used, the recording mode, the wavelength spectrum region, and the type of platform that is used as the basis of sensor placement. One of the remote sensing utilization is for multi-temporal analysis of mangrove area, which effective to monitor changes in the condition of mangrove forests (Suk-ueng *et al.* 2017).

Machine Learning is an empirical approach to regression or classification (supervised or unsupervised) on a nonlinear system (David *et al.* 2015) and excellent for processing large amounts of data. Machine Learning began to be developed since the 1950s, where in the early stages only conducted by a simple algorithm. According to Rhee (2016), machine learning method has a better advantage than the interpolation method for classification and regression. Machine learning approach (statistical approach) is so prominent in analyzing a complex system that has wide observations (Ashkezari 2016).

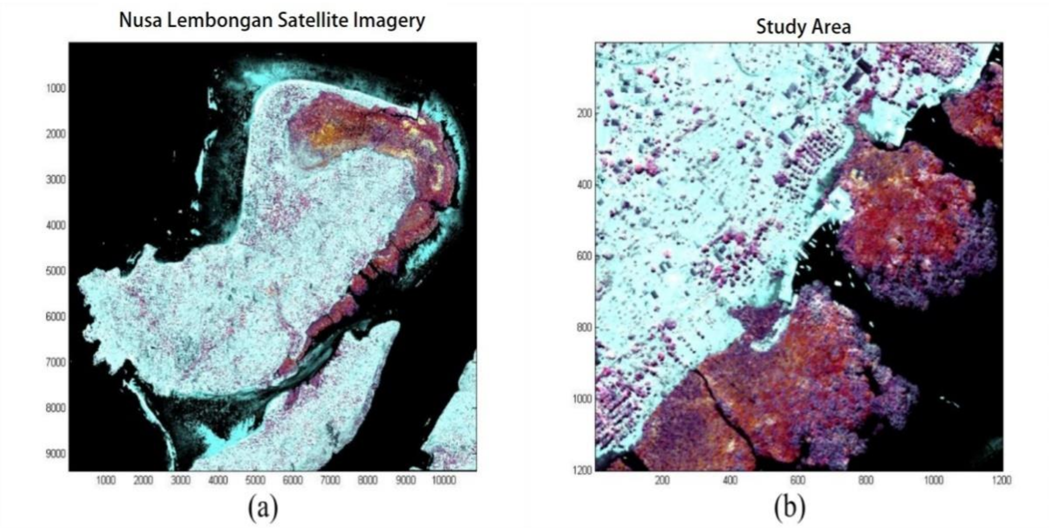

Figure 2-1: Worldview-2 satellite imagery. (a) Nusa Lembongan Island, and (b) The study area

To perform the data processing from unsupervised to be supervised can be done using algorithm of Akaike Information Criterion (AIC). The AIC model is the best choice for doing linear data processing (Bracher *et al.* 2015). AIC is used to perform processing with statistical approaches on large data. Hosseini *et al.* (2015) said this AIC can be used to achieve high predictive power from statistical models, which requires a substantial set of training and spatial data.

Mangrove forests in Nusa Penida sub-district are mostly concentrated on the northern side of Nusa Lembongan. The existence of mangroves along the coast of an island has an important role as a green belt or a protector against tsunami waves or hurricanes (Nuryani 2011). To manage mangrove areas in Nusa Lembongan and Nusa Ceningan effectively, basic information on the extent of mangrove forests, mangrove species that grow in the area, and fauna that live in the mangroves are necessary (Marthen *et al.* 2010).

This study aims to classify the determination of the distribution, type and extent of the mangrove area using the AIC approach on the satellite image data Worldview-2 in the Nusa Lembongan Island area. The assumption used is the study area is that selected sample location could represent the total area extent. The AIC approach is used as a reference to determine the number of existing classes resulted by unsupervised clasification as new features, followed by supervised classification with the Gaussian Mixture Model (GMM) approach. The level of accuracy is analyzed using Cohen's calculation Kappa.

2 MATERIALS AND METHODOLOGY

This research used satellite imagery data Worldview-2 of 2013 over Nusa Lembongan Island, Bali. The research sites include Nusa Lembongan Island and part of Nusa Ceningan Island as shown in Figure 2-1.a. The area of interest in this study is an area of study that can represent mangrove, terrestrial, and ocean vegetation as shown in Figure 2-1.b.

Satellite imagery Worldview-2 itself is usually used to perform spatial analysis as mapping, land use, and some needs requiring spectral data. This satellite is equipped with a high-resolution sensor that is 0.46 meters and 8 multispectral bands (Figure 2-2) which are capable to acquire image data in wide coverage of 1 million km^2 in a day.

The research method is to classify the land cover types using high resolution image data. This land classification can distinguish between land, sea, and mangrove. The mangrove land itself can later be differentiated by type and compared with previous studies that processed the same data with GIS method. An example of mangrove classification using GIS method obtained by Gaëlle *et al.* (2013) is shown in Figure 2-3.

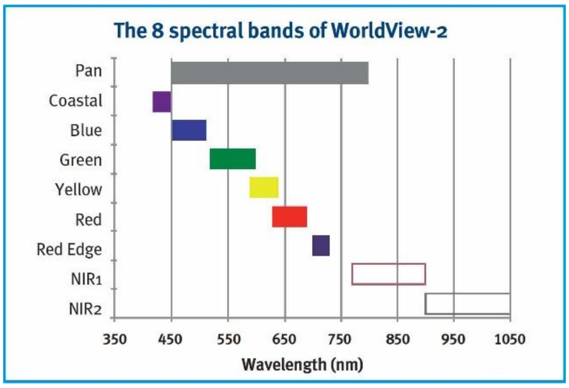

Figure 2-2: Wavelength for each bands (source: satimagingcorp.com)

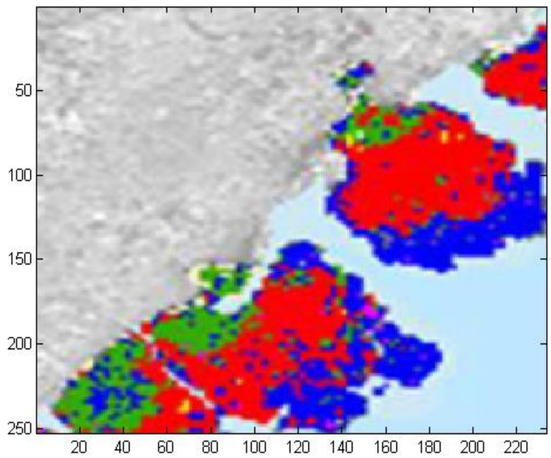

Figure 2-3: Mangrove classification using GIS method (Source: Gaëlle *et al.* 2013)

The data validation used the results of previous research conducted by Gaëlle *et al.* (2015), which also did the mangrove land classification in Nusa Lembongan Island using GIS method. In the area of interest, Figure 2-3, which corresponds to the current research site, the previous study revealed that there were 3 dominant mangrove species, namely: *Sonneratia alba* (blue), *Rhizophora apiculata* (red), and other types of mangroves (green). And there are also land and water/sea that have not been taken into account into the classification. Previous studies have classified mangrove land using the Worldview-2 satellite image data by matching field data results, so the results can be more accurate.

In this research, image data is processed by statistical learning approach that we developed using MATLAB software. The method used is with the AIC to determine the number of clusters of the image data. By having a collection of statistical data, AIC can help to determine the best quality model for certain data without guidance. According to Parviainen *et al.* (2013), AIC works by examining the sampled or whole data repeatedly to determine the parameters (clusters) in order to obtain the best results for the whole data. The AIC formulation used in this study is as shown in formula (1).

$$AIC = -2 \log [\pounds (\theta \mid y) + 2k] \qquad (2\text{-}1)$$

Then we used the GMM approach to conduct a supervised classification with the number of classes according to the calculation results from AIC. GMM is a probabilistic model to represent the existence of a parameter against the parameters as a whole, without using the data set previously calculated by the AIC. The function of this approach is used to calculate the probability of each data in each cluster which will then be grouped by the largest probability value.

3 RESULTS AND DISCUSSION

Data processing conducted by AIC approach has resulted 14 clusters. The cluster number is derived from the process of calculating the parameter determination, which value is generated from the data input processed is taken from the smallest value among the input numbers.

After clustering, then the data is merged according to the existing class in the reference for the same result, from the 14 clusters, there are 2 clusters 2 and 6 which cannot represent any classification because of the number of pixels and unspecified spatial depiction. In Table 3-1 it is seen that clusters 1 and 3 go into class 1 i.e Mangrove Other Types.

Then clusters 4 and 5 go into *Sonneratia alba* (SA) class, as well as clusters 7,8,10, and 12 belong to the *Rhizophora apiculata* (RA) class. In addition to be distinguished from the types of mangrove, the subsequent classes are differentiated to land and water/sea each consisting of clusters 9 and 11. While there are several clusters that are classified into other classes because they do not represent any class of clusters 2 and 6. Thus, clustering obtained from AIC can be considered to function as a new feature that can help ease the process of classification to be performed by GMM rather than its original feature, the spectral values of 8 bands of Worldview-2 image.

The incorporation of these clusters is carried out using supervised classification by utilizing spatial depictions that can represent classes in validation data.

3.1 Mangrove Land Classification

From unstructured land classification process based on calculation of AIC over study area, we obtained 14 clusters of mangrove land covers as shown in Figure 3-1.

Table 3-1: Distribution of clusters in each class

Cluster (pixel)	*Class (pixel)*					
	OM	*SA*	*RA*	*Water/Ocean*	*Land*	*Other*
1	13067					
2						4119
3	27775					
4		10160				
5		19498				
6						14719
7			131330			
8			52582			
9				65768		
10			38628			
11					27337	
12			42258			
13		39837				
14		76923				

In Figure 3-1, there are 14 clusters of classification without AIC supervised and GMM supervised classification. The details of each class are shown in Table 3-2. Among all clusters, there is 1 green cluster representing the water class, 1 orange cluster representing the land class, and 10 other color clusters representing the mangrove class. While cluster 1 (dark blue) and cluster 14 (dark red) do not represent the five classes, so these two clusters could be ignored and we used only 12 clusters for further analysis.

By combining several clusters it could produce the same pattern with reference results. The number of generated classes is divided into 5 land cover types, namely *Sonneratia alba*, *Rizhopora apicullata*, other types of mangrove, land, and water/sea, as shown in Figure 3-2.

In the area of interest it can be seen that mangrove land is presented by green, light blue, and dark blue color. While the yellow and orange color represent water/sea and land with the largest percentage of area is 26% and 33% of the total area respectively. The extent of mangrove distinguished by type has an area of OM 3.5%, SA 17%, and RA 20.5%. Comparing with the reference, the results obtained by this research are not fully similar because we used statistical approach in the calculation process. So that some points can be read as mangrove by the system and the resulting image is purely based on the value of each pixel that has not done smoothing process data. To see the relationship between the two methods, then calculate the accuracy value of AIC-GMM method to reference.

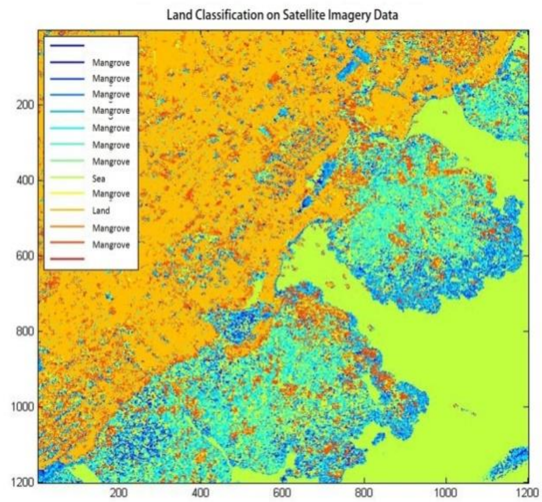

Figure 3-1: Supervised Field Classification Results Used AIC with 14 Clusters

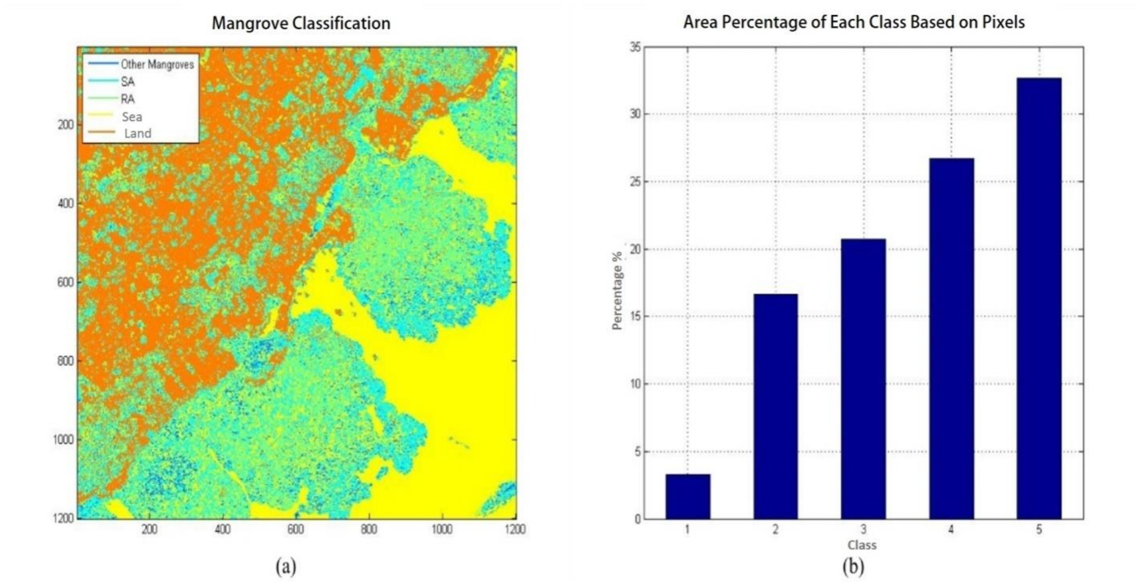

Figure 3-2: Results of land classification by type of mangrove. (a) types of mangroves, and (b) percentage of each class

Table 3-2: Accuracy table from Cohen's Kappa coefficient calculation

Classification	*1*	*2*	*3*	*4*	*5*	*Total*
OM	**14125**	26275	31600	7800	15200	95000
SA	4800	**88600**	28900	30150	1550	154000
RA	4450	33950	**178825**	21325	1700	240250
Water/Sea	1825	14875	7725	**228875**	4200	257500
Land	9625	75712.5	90787.5	12775	**501750**	690650
Total	34825	239412.5	337837.5	300925	524400	**1437400**
Agreement	14125	88600	178825	228875	501750	1012175
By Chance	2301.6384	15823.144	22328.205	19888.601	34658.411	95000

Kappa	**0.6832353**

Accuracy of the results can be calculated using Cohen's Kappa Coefficient method by considering all factors contained in each classification column. The calculation results can be seen in Table 3-2. The calculation is done to analyze the influence of AIC-GMM calculation with comparison in previous study using field data validation.

The accuracy resulted by Cohen's Kappa coefficient calculation is of 68.32% comparing with the reference result. The accuracy result is good enough for the AIC-GMM method that applies supervised/unsupervised classification simply by using pixel values as statistical data. There is a difference in the calculation process of each component, as we found in other types of mangrove classes having pixel values which were quite different from the appropriate classes. This may cause decreasing percentage of the accuracy.

3.2 Discussion

In this study we used the AIC method to determine the number of clusters from satellite imagery over the study area that were accounted as the new features in the further supervised and unsupervised classification processes. After obtaining the number of classes from the satellite image, then the selection of each pixel value to be classified into the appropriate class. GMM method has been successfully used to handle limited test data across multiple applications, including in remote sensing (Davari 2017). The method used a statistical learning approach for clustering. The output of the process is shown in the processed image that has been classified according to their respective classes.

AIC is used as a link between unsupervised and supervised processing. AIC is functioned to generate of new features to replace the original features. The new feature will be used in classification using GMM. Determination of the cluster number of spatial data is quite difficult to do. So this could be done by the statistical approach using AIC. The processing result is then used as the reference for the grouping of each data against the existing cluster. According to Liu, the advantage of using a combination between unsupervised and supervised is that the user can determine the parameters or the number of clusters that exist on the data quickly and minimize errors.

The AIC method could provide good result to classify mangrove over imagery data, both on large and small scales. For high resolution image data (Worldview-2 of 0.46 m resolution), the processing results showed almost the same density but different cluster readings. This is because in high resolution data, resulted image will be very detailed and the reading of values in each region will greatly affect the reflectance of the sensor. Image data classification is done for each pixel value, so the results of some pixels could differ from the result of other method. To obtain maximum results, it is necessary to conduct data smoothing into the results obtained, so the later results could be classified more clearly.

In this study, for validation we used is the data resulted by the previous works using GIS method. In the same study area, the results show similarity, so it can be said that the results obtained from processing using machine learning can be received well.

In the reference results, there are 3 mangrove types outside the study area, namely: *Rhizophora apiculata* (red), *Sonneratia alba* (blue), and other mangroves (green). The same result was also shown if we used the AIC method. The result of GIS method had better density when compared with machine learning method. This is because at the end process could smoothen the data so the areas could be distinguished clearly. The basic difference of both methods is on data processing. The AIC method is not limited by the tools provided by the processing software in conducting the classification, because the statistical approach is based on each pixel value. The model we applied in this research is the best choice for the data we used. Meanwhile, image data processing using other methods is usually limited by the tools of data processing software.

Using Cohen's Kappa we obtained the accuracy by considering the other components beside the main ones. This will produce different numbers if it is done with confusion matrix only. The advantage of image data processing method in this research is that we can identify spatial data without determining the parameters manually. The process can be conducted automatically using the GMM model. This method is also able to process large data with fast and efficient computing capabilities, which makes it suitable for processing high resolution satellite imagery. While the constraints are on the initial classification which is based on pixel spectral value, so the processing is not as good as GIS. In some cases, the AIC method still cannot distinguish water classes of sea water, pond water, or pond water, because the pixel values are similar. In addition, the texture variables is expected to provide a solution to solve the problem in water classification. This method could also perform big data processing which requires adequate device.

4 CONCLUSION

Based on the merger of several clusters from the AIC calculation, we obtained another type of mangrove classes consisting of clusters 1 and cluster 3; *Sonneratia alba* class consists of clusters 4,5, and 13; the *Rhizophora apiculata* class consists of clusters 7, 8, 10, and 12; while the water/sea and terrestrial classes consist of clusters 9 and 11 respectively. There were also two clusters that could not be identified because they did not represent any classes in clusters 2 and 6. Using the Akaike Information Criterion method, we obtained data accuracy calculations with Cohen's Kappa Coefficient which is equal to 68.32%. Akaike Information Criterion is one step ahead in image data processing, which is from unsupervised into supervised classification.

For better results in future research using AIC for supervised and unsupervised classification process, it is necessary to perform repetitive testing of the data samples. This testing is necessary, so that resulted parameters can be tested based on the actual data.

ACKNOWLEDGEMENT

Thanks to the Indeso Project which has provided the image of the Worldview-2 satellite and the Research and Human Resources Agency, the Ministry of Marine Affairs and Fisheries which supported this research, as well as the relevant parties contributing to the research implementation.

REFERENCES

Ashkezari MD, Hill CN, Follett CN, (2016), Oceanic Eddy Detection and Life Time for Ecast using Machine Learning Methods. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology.

Bracher A., Taylor MH, Taylor B., *et al.*, (2015). Using Empirical Orthogonal Functions Derived from Remote-Sensing Reflectance for the Prediction of Phytoplankton Pigment Concentrations. Ocean Science, Germany.

Davari AA, Christlein V., Vesal S., *et al.*, (2017), GMM Supervectors for Limited Training Datain Hyperspectral Remote Sensing Image Classication. Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.

Hosseini R., Newlands NK, Dean CB, *et al.*, (2015), Statistical Modeling of Soil Moisture, Integrating Satellite Remote-Sensing (SAR) and Ground-Based Data. Remote Sensing Journal.

Lary DJ, Alavi AH, Gandomi AH, *et al.*, (2015), Machine Learningin Geosciences and Remote Sensing. University of Texas.

Liu X., (2005), Supervised Classification and Unsupervised Classification, ATS 670 Class Project.

Parviainen M., Zimmermann NE, Heikkinen RK, *et al.*, (2013), Using Unclassified Continuous Remote Sensing Data to Improve Distribution Models of Red-Listed Plant Species. Biodivers Conserv.

Rhee J., Im J., Park S., (2016), Drought Forecasting Based on Machine Learning of Remote Sensing and Long-range Forecast Data, APEC Climate Center, Republic of Korea.

Suk-ueng K., Buranapratheprat A., Gunbua V., *et al.*, (2017), Applicationof Remote Sensing Technique for Mangrove Mapping at the Welu Estuary, Thailand. Chiangrai Rajabhat University.

Viennois G., Proisy C., Feret JB, *et al.*, (2015), Multitemporal Analysis of High-Spatial-Resolution Optical Satellite Imagery for Mangrove Species Mappingin Bali, Indonesia. France.

Welly M., Sanjaya W., (2010), Identifikasi Flora dan Fauna Mangrove Nusa Lembongan dan Nusa Ceningan. Coral Triangle Center.

Wicaksono P., Danoedoro P., Hartono, *et al.*, (2015), Mangrove Biomass Carbon Stock Mapping of the Karimun Jawa Islands using Multispectral Remote Sensing. ITT.

Widagti N., Triyulianti I., Manessa MDM, (2011), Changesin Density of Mangrove Forestin Nusa Lembongan, Bali. Institute for Marine Research and Observation.