Proceedings of the 1st Faculty of Industrial Technology International Congress
International Conference
Bandung, Indonesia, October 9-11, 2017
ISBN 978-602-53531-8-5

Investigation of Classification Algorithm for Land Cover Mapping in Oil Palm Area Using Optical Remote Sensing

Anggun Tridawati* and Soni Darmawan
Department of Geodesy Engineering, Institut Teknologi Nasional (Itenas), Bandung - INDONESIA
* Corresponding author e-mail: angguntridawati.30@gmail.com

Abstract

Technological development on globalization era enable the application of remote sensing technology in access speed and accuracy for mapping land cover based on image classification. The objective on this study is to investigate the utilization of remote sensing imagery of Landsat 8 Operational Land Imager (OLI) for the classification of land cover in oil palm area. Methodology consists of collecting of Landsat 8 OLI, radiometric and geometric correction, making the Region of Interest (ROI) based on class are used in the multispectral classification process to know the land cover in the form of oil palm area, accuracy analysis, and mapping land cover classification with the highest accuracy and kappa coefficient. The algorithm used for classification of land cover types to make class are Maximum Likelihood, Minimum Distance, and Support Vector Machine (SVM) by using various bands on Landsat image and added Normalized Difference Vegetation Index (NDVI). The results in this study show that Support Vector Machine is the best Algorithm of three classification algorithms using all bands on Landsat 8 OLI image with overall accuracy of 96,21% (kappa coefficient 0,9041), Maximum Likelihood Algorithm with overall accuracy of 89,53% (kappa coefficient 0,7713), and Minimum Distance Algorithm with overall accuracy of 84,83% (kappa coefficient 0,6799).

Keywords: Support Vector Machine, Kappa Coefficient, Supervised Classification, NDVI, Landsat 8 OLI

1. Introduction

Oil palm is the world's most productive oil seed and an increasingly important agricultural product for tropical countries around the world (Butler et al., 2009). Malaysia and Indonesia fulfill 85% of Crude Palm Oil (CPO) of the world (Ministry of Agriculture, 2010). Therefore, there is an urgent need to monitor oil palm expansion in the region to provide a better estimate of smallholder plantation activities. This information is an important input to evaluate the success of development policies and the impact of these activities (I.K. Nooni, 2014). The rapidly evolving technological developments of this era of globalization allow the application of remote sensing methods in the classification of land cover in oil palm plantations. This technique is considered important and effective in monitoring land cover because of its ability to provide information of spatial diversity on the surface of the earth quickly, widely, precisely and easily (Sampurno, 2016). Oil palm mapping in tropical environments is challenging due to its landscape nature, cost of image data acquisition, and classification selection (I.K. Nooni, 2014). Remote sensing data began to be widely used to derive the information needed in environmental planning activities at local, regional, and global scales (Johannes et al., 2003 in Jensen 2005). Remote sensing has several methods developed to gain information on land cover. One widely used method is the digital multispectral classification based on statistical analysis (Hamdir, 2014). The classification of land cover or land use based on statistical pattern recognition techniques applied to remote multispectral sensing data is one of the most widely used methods in the information gathering process (Narumalani et al., 2002 in Jensen, 2005). The most common method of classification is the supervised classification method. Several algorithms in the supervised classification are Maximum Likelihood, Minimum Distance, and Support Vector Machine as done by Sampurno (2014) classification of land cover with Maximum Likelihood algorithm resulted accuracy of 99.61%. Setyowati (2015) classified land cover using Maximum likelihood algorithm to produce accuracy of 93.5%, while according to Eko (2012) Support Vector Machine algorithm is the best algorithm in classifying land cover in mangrove area with of 77.93% accuracy. However, little research has been done to map palm oil related land cover. Therefore, researchers continue to seek the best classification algorithm to further improve the accuracy of classification, especially in areas of oil palm plantations.

2. Methodology

This study is concentrated on the utilization of remote sensing technology by utilizing Landsat 8 OLI (Operational Land Imager) Satellite Image for classification of land cover in oil palm plantation area. Starting from the data collection is LANDSAT 8 OLI (Operational Land Imager) satellite imagery for the year of July 2013 and Bing Satellite imagery for December 2013. The next step is radiometric correction using MODTRAN 4 method which is the name FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) available on ENVI software and geometric correction of Landsat 8 OLI imagery with Bing satellite imagery as a reference. Making Region of Interest (ROI) based on class that will be used in the classification process to know land cover in the study area in the form of oil palm, non-oil palm (settlements, road, and empty land), water, forest and cloud cover. The classification algorithm used are Maximum Likelihood, Minimum Distance, and Support Vector Machine (SVM) algorithm. Then assessment accuracy by displaying the error matrix (Confusion Matrix) for each classification algorithm. **I**n general, the methodology in this study can be seen in Figure 1.

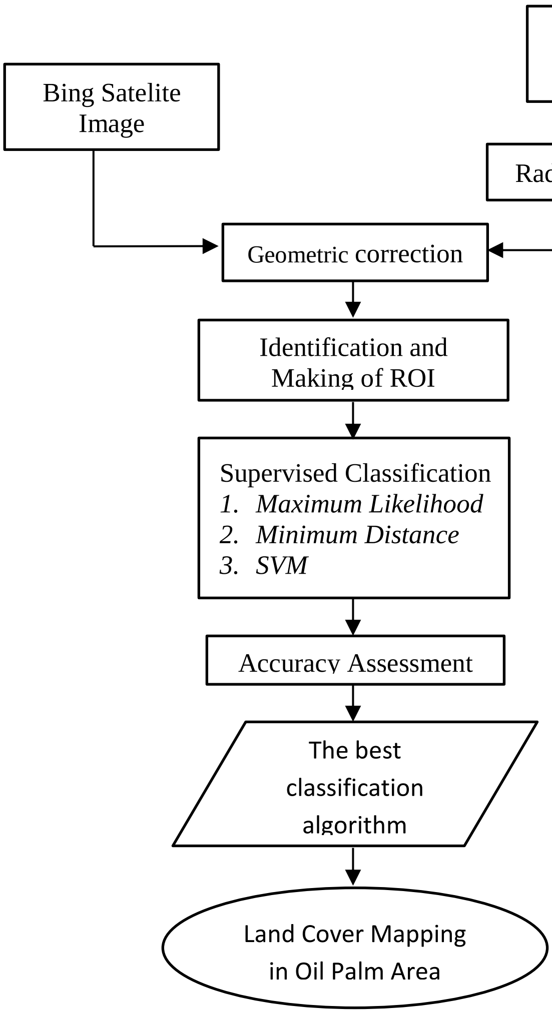
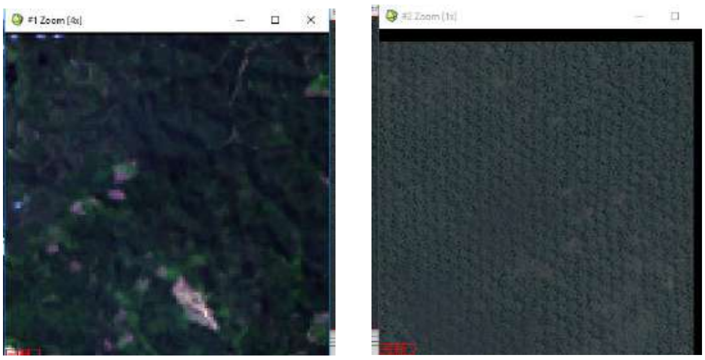


Fig. 1: Methodology on this study

2.1. Identification and Making of Region of Interest

Identification (interpretation) is done to see the difference of each class visually and then made the manufacture of ROI. This identification is done using the data from Bing satellite imagery that has a better spatial resolution of 0.6 m from Landsat 8 OLI image (30 m spatial resolution) so that the land cover is visible. For example, the appearance of oil palm trees with forests that have similarities with Bing Satellite Imagery will be seen oil palm appearance because it is planted with regular pattern. This process is done by displaying the image in a row. As in Figure 2.



a. Landsat Satellite Image b. Bing Satellite Image

Fig. 2: Identification of Oil Palm Plantation

Oil palm plantation have patterns and textures that are different from other objects located in the area. When compared to the surrounding objects, this oil palm plantation has a pattern that is in groups and lined regularly, so that when viewed from the top has a pattern that is very orderly and neat and make the process of classification multi spectral easier. The making of Region of Interest (ROI) or grouping of pixels on each object is done to retrieve statistical information of land cover classes. The retrieval of statistical information on Landsat imagery is done by determining all sample areas of each land cover class manually with the Bing satellite imagery for each class of 3x3 pixels.

2.2 Multispectral Classification

The purpose of the classification is to produce land cover map in oil palm plantation area. The method used in this classification is the supervised classification method of Maximum Likelihood, Minimum Distance, and Support Vector Machine (SVM) algorithms using the existing bands approach on Landsat 8 OLI and added with NDVI (Normalized Difference Vegetation Index) vegetation index. Classification is based on oil palm, non-oil palm (settlements, road, and empty land), water, forest and cloud cover. The result of the multispectral classification of the three classification algorithms is the land cover map in the area of oil palm plantations.

2.2.1 Maximum Likelihood

The Maximum Likelihood algorithm will classify a pixel on a particular land cover class by calculating a probable log function (Ahmad and Quegan, 2012). Mathematically the log function formula is probably written as follows:

$$g_i(\omega) = \ln P(\omega|i) = \frac{1}{2}(\omega - m_i)^T E_I{}^{-1}(\omega - m_i) - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|E_i| \quad (1)$$

where $g_i(\omega) = \ln P(\omega|i)$ is a possible log function, ω is the characteristic vector of an x pixel to be classified, m_i is the average vector of the class, E_i is the covariance matrix of the class, $E_I{}^{-1}$ is the inverse of the covariance matrix and N is the number of land cover classes. This method has good performance for remote sensing, because it is suitable for the object class whose gray distribution approaches the Gaussian model (Duda, 1973).

- Result of Maximum Likelihood Algorithm

The result of this algorithm can be seen in Figure 3.

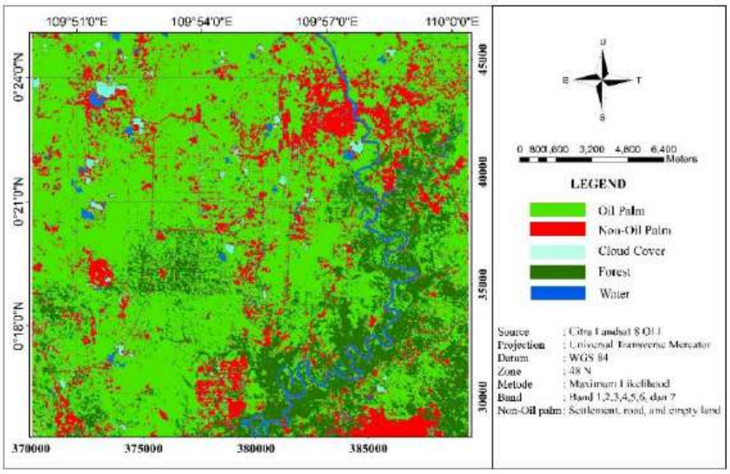


Fig. 3: Land Cover Classification Map of Maximum Likelihood

In the Gaussian Maximum Likelihood Classification method, in addition to the average variables also pay attention to the variance of the vector characteristics of objects in the class. This information is useful for knowing the distribution of each class of the measured variable.

- Accuracy Assessment

Accuracy assessment is showed by confusion matrix using kappa coefficient, that is comparing the result of classification obtained from each image with reference image. The selection of these coefficients is based on a consistency of judgment that considers the producer's accuracy and user's accuracy aspects. The kappa coefficient value has a range of 0 to +1, in the mapping process of land cover / land use an acceptable accuracy of 85%, or 0.85 (Anderson, 1976). The confusion matrix can be seen in Table 1.

Table 1: Confusion Matrix of Maximum Likelihood

	Class	ROI Ground Check							
		Non-oil palm	Forest	Palm	Water	Cloud cover	Total	Product accuracy	Error omission
Klasifikasi	Non-oil palm	270	1	41	2	0	314	85,99	14,01
	Forest	0	181	208	1	0	390	46,41	53,59
	Palm	7	7	1784	0	0	1798	99,22	0,78
	Water	2	0	3	60	0	65	92,31	7,69
	Cloud cover	0	0	7	0	90	97	92,78	7,22
	Total	279	189	2043	63	90	266	Overall accuracy	0,89572
	Product accuracy	96,77	95,77	87,32	95,24	100			
	Error omission	3,23	4,23	12,68	4,76	0		Kappa	0,7713

The result of accuracy assessment of multispectral classification with the actual condition as a whole (Overall Accuracy) is 89,53% (kappa coefficient 0,77) that indicates the process of this multispectral classification, using Maximum Likelihood algorithm can avoid interpretation error of 89.53%, so the possibility to make a mistake of only 10.47%. The accuracy is strongly influenced by the selection of samples and also the number of samples selected. For example, in the forest classification that has an accuracy of 95.77% in which of the total 189 existing samples there is 1 sample that goes into the non-oil palm class, and 7 samples into the oil palm class. This is because the similarity of form and value of Digital Number is almost the same in the oil palm and non-oil palm classes.

2.2.2 Minimum Distance

Pixel grouping is based on the shortest distance to the average vector of a class, so no unclassified pixels (John, 2012). If each pixel in the image is represented in the vector of character u, then one way to determine the membership of u is to put it into the nearest class of distance u. The distance between the u characteristic vector and the j j class pattern vector (mj) can be calculated using the Euclidean distance with the formula:

$$D_j(u) = |u - m_j| \quad (2)$$

Where j=1,2,…,M and $|a| = (a^T a)^{1/2}$ is *norm Euclidean.* The pattern vector u is assigned to the class kj if $D_J\ (u)$ is the closest distance compared to $D_p(u)$, 1< p < M, with p≠j. This method is known as a geometric classification (Hudson, 1987).

- Result of Minimum Distance Algorithm

The result of this algorithm can be seen in Figure 4.

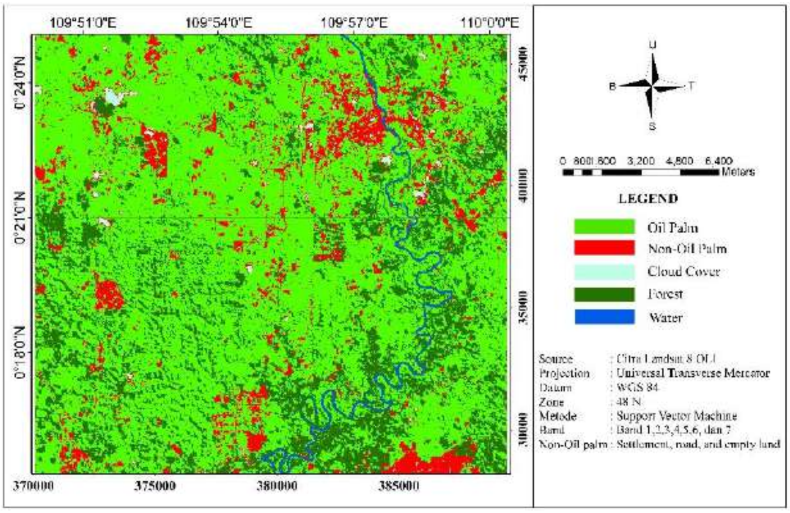


Fig. 4: Land Cover Classification Map of Minimum Distance

- Accuracy Assessment

Accuracy assessment is showed by confusion matrix using kappa coefficient, The result of accuracy assessment of multispectral classification with actual condition as a whole (Overall Accuracy) of 84.83% and Kappa coefficient shows 0.68. This indicates in this multispectral classification process, using Minimum Distance algorithm can avoid interpretation error equal to 84,83% so the possibility to make mistake only equal to 15,17%. Table of accuracy assessment can be seen in Table 2.

Table 2: Confusion Matrix of Minimum Distance

	Class	ROI Ground Check							
		Non-oil palm	Forest	Palm	Water	Cloud cover	Total	Product accuracy	Error omission
Klasifikasi	Non-oil palm	251	0	78	4	11	344	72,97	27,03
	Forest	6	177	267	3	0	453	39,07	60,93
	Palm	19	12	1698	1	0	1730	98,15	1,85
	Water	0	0	0	55	0	55	100	0
	Cloud cover	3	0	0	0	79	82	96,34	3,66
	Total	279	189	2043	63	90	2664	Overall accuracy	0,848348
	Product accuracy	89,96	93,65	83,11	87,3	87,78			
	Error omission	10,04	6,35	16,89	12,7	12,22		Kappa	0,6806

Based on table 2 in the forest classification that has an accuracy of 93.65% of which of the total 189 samples there are 12 samples entered into the palm class. This is because class retrieval with Minimum Distance algorithm is based on the closest distance, because in the study area of oil palm area there are also many forests.

2.2.3 Support Vector Machine

Support Vector Machine (SVM) is a learning system that uses hypothetical space in the form of linear functions in a high-dimensional feature space (Feature Space), trained with learning algorithms based on optimization theory by implementing learning bias derived from the theory of statistical learning (Sembiring , 2007). The main purpose of the Support Vector Machine classification is to find the best hyperplane function with the maximum hyperplane margin measurement (Muhammad, 2014) can be written with the formula:

$$f(x) = sign(\sum y_i a_i{}^0 K(X_i, X) - b^0) \quad (3)$$

Where X_i is vector of data (pixels), X_j is vector on training sample data, and y_i is kernel parameters.

- Result of Support Vector Machine

The result of this algorithm can be seen in Figure 5.

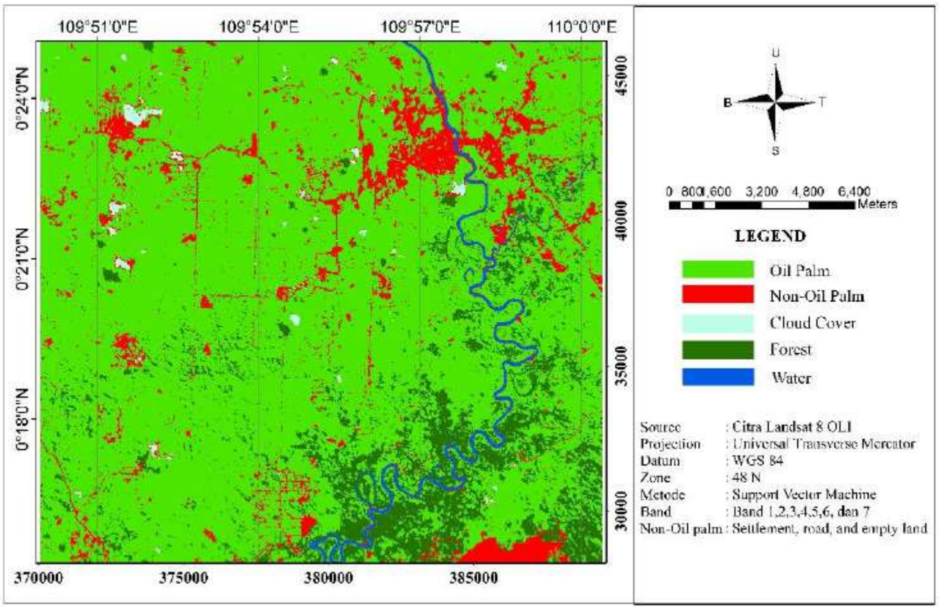


Fig. 4: Land Cover Classification Map of Support Vector Machine

- Accuracy Assessment

Accuracy assessment is showed by confusion matrix using kappa coefficient, The result of accuracy assessment of multispectral classification with the actual condition as a whole (Overall Accuracy) is 96.21% while the classification accuracy for each land cover has different percentage. Kappa shows the number 0.90. Table of accuracy assessment can be seen in Table 3.

Table 3: Confusion Matrix of Support Vector Machine

	Class	ROI Ground Check							
		Non-oil palm	Forest	Palm	Water	Cloud cover	Total	Product accuracy	Error omission
Klasifikasi	Non-oil palm	250	0	5	4	3	262	95,42	4,58
	Forest	0	176	43	3	0	222	79,28	20,72
	Palm	29	13	1995	0	1	2038	97,89	2,11
	Water	0	0	0	56	0	56	100	0
	Cloud cover	0	0	0	0	86	86	100	0
	Total	279	189	2043	63	90	2664	Overall accuracy	96,21%
	Product accuracy	89,61	93,12	97,65	88,89	95,56			
	Error omission	10,39	6,88	2,35	11,11	4,44		Kappa	0,9041

Based on the results in Table 3, the Support Vector Machine algorithm is the algorithm that produces the highest accuracy compared to the Maximum Likelihood algorithm and Minimum Distance algorithm. The kappa coefficient shows that the SVM algorithm reduces classification errors compared to Maximum Likelihood and Minimum Distance. The results of this study indicate the sensitivity of the SVM method in mapping oil palm plantations in heterogeneous environments in terms of overall accuracy, user accuracy, and product accuracy. According to Mathur (2004) the SVM classification basically takes input from the training data and predicts the input given, which outside the class forms the input by linking the training data assigned to each pixel in the image. It then operates to find the separation limit that is between the class pairs by marking each pixel as part of the class, based on the input. So the SVM algorithm classification is the most sensitive algorithm to the specified training size and the SVM algorithm has also demonstrated its potential to identify and map the distribution of oil palm in a heterogeneous environment.

2.2.2 Bands Combination Analysis

Bands used in this classification can be seen in Table 4. Based on the classification that has been done the use of all bands on Landsat 8 gives the best accuracy results.

Table 4: Combination Bands

Metode Klasifikasi	Parameter	Saluran				
		23456	234567	123456	1234567	1234567+NDVI
Maximum Likelihood	Accuracy	85,21%	87,01%	86,82%	89,53%	85,59%
	Kappa coefficient	0,6963	0,7274	0,7217	0,7713	0,7033
Minimum Distance	Accuracy	84,72%	84,61%	84,83%	84,83%	84,72%
	Kappa coefficient	0,6782	0,6772	0,6799	0,6806	0,6756
Support Vector Machine	Accuracy	94,97%	95,91%	95,35%	96,21%	95,12%
	Kappa coefficient	0,8739	0,8967	0,8828	0,9041	0,8776

The multispectral classification assumes that each object can be distinguished from other objects based on its spectral value. Based on the results of this study, the composition of the band that produces the highest accuracy is to use all bands in Landsat 8 image that are band 1,2,3,4,5,6, and 7 because more and more use of band, hence more yield more object This representation is reinforced by Danoedoro (2012) who says that the more bands used, the more accurate it is to provide accuracy.

According to Wielicki (1985), the condition of resistance can affect the overall reflectance value of an optical image. The Landsat satellite is a passive satellite that relies on the sun as an energy source. The condition of resistance will affect the amount of solar radiation that reaches the object on earth. Band 1 is a coastal / aerosol designed for aerosol monitoring closely related to band 2, visible blue. Aerosol is an airborne particle, since band 1 can be used to analyze the effect of atmospheric disturbance (aerosols), by reducing band 1 automatically the spectral information used for classification will decrease so the accuracy value will also decrease.

The NDVI value is a value for knowing the greenery of the leaves with excellent infrared wavelengths as the beginning of the division of the vegetation area. Because the optical properties of chlorophyll are so characteristic that chlorophyll absorbs the red spectrum and reflects strongly in the infrared spectrum. But whenever the addition of information is not effective because of the redundant information on multispectral bands. Such as NDVI, made from the Red and NIR bands are also used as input in the classification so that information is repeated. Therefore band bands 1,2,3,4,5,6,7 and NDVI have accuracy values smaller than band 1,2,3,4,5,6, and 7.

Classification by eliminating band 7, using bands 1,2,3,4,5,6 also has a smaller accuracy value than using all bands, because band 7 has a function can be used to detect the stress of dryness of plants and describe the burned area and fire-burning vegetation, and also sensitive to thermal radiation emitted by fire can be used to detect active fires, especially at night. Therefore, reducing band 7 gives a lower accuracy.

3. Conclusion

The classification of land cover in the oil palm plantation area of the supervised classification method of the Support Vector Machine (SVM) algorithm using all bands 1,2,3,4,5,6 and 7 on Landsat 8 OLI image is the best algorithm with overall accuracy of 96,21% (kappa coefficient 0,90) in producing land cover map in oil palm plantation area.

4. References

Ahmad A, Quegan S. 2012. Analysis of Maximum Likelihood Classification on Multispectral Data. Applied Mathematical Sciences. 129 (64):25–36.

Anderson, J. H., E., Roach J.T., & R. Wittmer. (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data.Geological Survey Professional Paper 964. Washington: United States Government Printing Office.

Butler, R. A., L. P. Koh, and J. Ghazoul. (2009). “REDD in the Red: Palm Oil Could Undermine Carbon Payment Schemes.” Conservation Letters 2 (2): 67–73. doi:10.1111/j.1755263X.2009.00047.x.

Danoedoro, Projo. (2012). Pengantar Penginderaan Jauh Digital. Yogyakarta: ANDI Yogyakarta.

Duda, R.O., P.E. Hart. (1973). Pattern Classification and Scene Analysis, New York: Wiley

Eko, Rudi. (2012). Klasifikasi Penutupan Lahan Hutan Mangrove Di Kecamatan Buduran, Kabupaten Sidoarjo, Provinsi Jawa Timur, Dengan Citra Satelit Terrasar-X High Resolution. Skripsi Program Sarjana pada Fakultas Pertanian IPB. Bogor.

Hamdir, Ardian. (2014). Studi Perbandingan Klasifikasi Multispektral Maximum Likelihood Dan Support Vector Machine Untuk Pemetaan Penutup Lahan. Kartografi dan Pengindraan Jauh. Universitas Gajah Mada.

Hudson, William D. (1987), Digital Classification of Landsat Multispectral Scanner data – An Introduction, Research Report Natural Resources From The Michigan State University Agricultural Experiment Station East Lansing, No 483.

I.K. Nooni, A.A. Duker, I. Van Duren, L. Addae-Wireko & E.M. Osei Jnr. (2014). Support vector machine to map oil palm in a heterogeneous environment. International Journal of Remote Sensing Vol. 35, No. 13, 4778–4794.

Kementrian Pertanian. (2010). Outlook Komiditi Kelapa Sawit. Pusat dan Sistem Informasi Pertanian.

Mathur A. (2004). A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing.

Muhammad, Teguh. (2014). Kajian Beberapa Metode Klasifikasi Citra Digital Terhadap Data Penginderaan Jauh. Fakultas Matematika Dan Ilmu Pengetahuan Alam. Institut Pertanian Bogor.

Narumalani, D.C. Rundquist, R.L. Perk, J. Schalles, K. Hutchins, dan J. Keck. (2002). Classifying and Mapping General Coral-Reef Structure Using Ikonos Data. Photogrammetric Engineering & Remote Sensing. Vol. 68, No. 12, hal: 1297-1305

Sampurno, Rizky. (2016). Klasifikasi Tutupan Lahan Menggunakan Citra Landsat 8 Operational Land Imager (Oli) Di Kabupaten Sumedang. Fakultas Teknologi Industri Pertanian, Universitas Padjadjaran

Setyowati, Heratania. (2015). Aplikasi Citra Spot-6 Berbasis Transformasi spektral Untuk Estimasi Produksi Kelapa Sawit. Kartografi dan Pengindraan Jauh Universitas Gajah Mada.

Wielicki, Bruce. (1985). Cumulus Cloud Properties Derived Using Landsat Satellite Data: Journal of Climate and Applied Meteorology vol 25 no 3.

Yohannes. (2012). Sistem Pengindraan Jauh Teknik Survey dan Pemetaan. Diktat Kuliah. Lampung: Universitas Lampung.