PROC. INTERNAT. CONF. SCI. ENGIN.
Volume 4, February 2021
Page 229-234

ISSN 1504607797
E-ISSN 1505707533

Accuracy Assessment of Land Use/Land Cover Classification Data from Sentinel-2 & ASTER Imagery Interpretation using Unsupervised Classification Algorithm

Labisa Wafdan
Geodetic Engineering Department, Faculty of Engineering, Gadjah Mada University,
Jl. Grafika Bulaksumur No.2 Daerah Istimewa Yogyakarta 55281, Indonesia, Tel. +62-274-520226.
E-mail: labisawafdan240@gmail.com

Abstract. Accurate land use/land cover information is important for various spatial planning decision making. Remote sensing is an effective mapping technique such as those depicting land cover as it provides a map-like representation of the Earth's surface that is spatially highly consistent. This study compared the classification accuracies of land cover/land use maps created from Sentinel-2 and ASTER imagery with the Kalimanah Sub-district as a research area. Both images are clustered into 52 spectral clusters using Learning Vector Quantization (LVQ) and K-means unsupervised classification algorithm. Each spectral cluster from each image was assigned into four land use/land cover classes, i.e. urban, agricultural, forest, and barren land. 240 data references were generated from Google Earth imagery as the sample data set is compared with the classification maps that is being assessed. With the kappa analysis approach, error matrices are made based on the same data references for each of the two images to assess the classification quality and to find out the best imagery that yields the most accurate land use/land cover data. Overall accuracy of LVQ algorithm for the Sentinel-2 and ASTER imageries was 78.33% and 69.17%, respectively; while the kappa coefficient of LVQ algorithm for the Sentinel-2 and ASTER imageries were 0.71 and 0.55, respectively. In different circumstances, overall accuracy of K-means algorithm for the Sentinel-2 and ASTER imageries were 81.25% and 72.68%, respectively; while the kappa coefficient of K-means algorithm for both imageries were 0.74 and 0.61, respectively. At the 95% confidence level, for both LVQ and K-means classification algorithms, image classification accuracies of Sentinel-2 dataset are better than the ASTER dataset. Thus, Sentinel-2 imagery provides better accuracy than ASTER imagery in land use/land cover classification data from any unsupervised classification algorithms.

Keywords: remote sensing, land use/land cover classification, unsupervised classification algorithm, accuracy assessment.

Abbreviations: Overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), error of comission (EC), error of omission (EO).

Running title: Accuracy Assessment of Land Use/Land Data.

INTRODUCTION

Remote sensing is an attractive source of thematic maps such as those depicting land cover as it provides a map-like representation of the Earth's surface that is spatially continuous and highly consistent, as well as available at a large range of spatial and temporal scales. Thematic mapping from remotely sensed data is typically based on image classification. This may be achieved by either visual or computer-aided analysis. Acquired information from images remotely can be a valuable tool for a variety of resource-based applications (Foody, 2002). Governments are also trying to recognize the value of cooperating in the development of land use/land cover data resources that can be used in disaster response and recovery across jurisdictions. Because of their capacity to provide comparable data over a large number of administrative units, remote sensing can be a valuable source of information for regional responses to growth and natural disasters, and for policy and decision making (Board & National Research Council, 2003).

The classification of land use/cover data may be one that seeks to group cases by their relative spectral similarity or using the unsupervised classification method (Foody, 2002). The unsupervised procedures are applied in two separate steps. According to Lillesand *et al.* (2015), the fundamental difference between these techniques is that supervised classification involves a training step followed by a classification step. In the unsupervised approach, the image data are first classified by aggregating them into the natural spectral groupings, or clusters, present in the scene. The classifier identifies that clusters present in the image data. Then the image analyst determines the land cover identity of these spectral groups by comparing the classified image data to ground reference data.

Accuracy assessment or validation is a significant step after the classification in the processing of remote sensing data because a productive utilization of geodata is only possible if the quality of the data is known (Rwanga and Ndambuki, 2017). There are many reasons for performing an accuracy assessment. Perhaps the simplest reason is curiosity the desire to know how good a map you have made. Also, the satisfaction gained from this knowledge to increase the quality of the map information by identifying and correcting the sources of errors. Also, if the information derived from the remotely sensed data is to be used in some decision-making process, then some measure of its quality must be known. Finally, it is more and more common that some measure of accuracy is included in the contract requirements of many mapping projects. Therefore, valid accuracy is not only useful but may be required (Congalton and Green, 2019). This research aims to compare the accuracy of the land use/cover classification from Sentinel-2 and ASTER imagery using the unsupervised classification method especially using K-means and Learning Vector Quantization (LVQ) algorithms.

MATERIALS AND METHODS

Study Area

This research is located in Kalimanah District, Purbalingga Regency, Central Java. This study uses one image for each Sentinel-2 and ASTER satellites. Both imageries will be classified into four land use/land cover classes, i.e. built-up, agricultural, forest, and barren land. ASTER imagery was obtained from Global Visualization Viewer (GloVis, https://glovis.usgs.gov), while Sentinel-2 imagery was obtained from European Space Agency (ESA) Copernicus portal (https://scihub.copernicus.eu). Sentinel-2 imagery was acquired on May 22, 2017, while ASTER imagery was acquired on May 17, 2017. Google Earth imagery also was acquired as the reference or ground truth data set is compared with the classification maps that are being assessed. The equipment used consists of hardware and software as follows:

a. Hardware
   1. Lenovo laptops with specifications: Intel® Core ™ i5-8250U CPU@1.60 GHz (2 CPUs), ~ 1.80 GHz RAM 4.00 GB
b. Software
   1. QGIS 3.14
   2. Chrome browser
   3. Microsoft Word 2013
   4. Microsoft Excel 2013

Procedures

Pre-processing

The first step is the pre-processing, i.e. radiometric correction, band composite, geometric correction, and subsetting process on each of Sentinel-2 and ASTER imageries. This pre-processing step is done by Semi-Automatic Classification Plugin 5.3.6.1 in QGIS 3.14 Software. The value recorded in an image includes the radiation of the earth's surface and the atmosphere's radiation. To achieve the actual surface values, radiometric calibration must be applied. The DOS1 method is based on the properties of the image. Band compositing was performed on both of the imageries by combining the 3-2-1 bands, so the imageries will show the natural color. According to Lillesand *et al.* (2015), raw digital images usually contain geometric distortions because of some atmospheric refraction, relief displacement, and non-linearities in the sweep of a sensor's IFOV (Instantaneous Field of View). Geometric correction intends to compensate for the distortions introduced so that the corrected image will have the highest practical geometric integrity. Geometric correction uses nearest neighbor resampling method and 3rd order polynomial transformation. Subsetting was used to reduce the spatial extent of an image, so the image will cover only on Kalimanah District.

Processing

The second step is the processing or classification process, i.e. clustering, labeling, and reclassifying process. The classification process involves unsupervised classification algorithms. In the clustering process, the classifiers involve algorithms that examine the unknown pixels in an image and aggregate them into several classes or spectral clusters based on the natural groupings (Lillesand *et al.*, 2015). Both imageries are clustered into 52 spectral clusters using K-means and Learning Vector Quantization (LVQ) classifier method on Google Earth Engine. Each spectral cluster from both imageries is labeled or determined with four desired land use/land cover classes that previously mentioned based on the reference data, i.e. Google Earth, Sentinel-2, and ASTER imageries. Then, each labeled clusters was reclassified to unify that 52 labeled spectral classes into only 4 classes. Without considering the relationship between spatial and spectral properties.

Post-Processing

The second step is post-processing, i.e. sieving and accuracy assessment. To get more accurate the map of classification result, in the remote sensing processing are also used such assumptions are not valid in region having fuzziness, which occurred due to the presence of mixed pixels which consist heterogeneous properties for more than one class (Kaura and Bansal, 2018). Thus, every pixel image classification techniques would probably cause salt-and-pepper effect (Su, 2016). It is caused by high isolated local spatial heterogeneity between neighboring pixels, so single pixels may not represent real conditions on the ground. Remove that isolated pixels with a sieving process will enhance the classified image and obtained more accurate classification maps (Bakr *et al.*, 2019). Accuracy assessment determines the quality of a map created from remotely sensed data by comparing the ground truth data and the classified imageries. 240 data references were generated from Google Earth imagery as the sample data set is compared with the classification maps that are being assessed, i.e. Sentinel-2 and ASTER imageries. Error matrices are made based on that same data references for each of those images to calculate the elements of overall accuracy, user's accuracy, producer's accuracy, error of comission, error of omission, and kappa coefficient. The definitions of those elements are:

1. Overall accuracy (OA) is calculated by summing the number of correctly classified values and dividing by the total number of values.
2. User's accuracy (UA) is the probability that a pixel was predicted to be in a certain class is that class.
3. Producer's accuracy (PA) is the probability that a value in a class was classified correctly.
4. Error of comission (EC) is the proportion of a pixel that was predicted to be in a class but do not belong to that class.
5. Error of omission (EO) is the proportion of observed pixel on the ground that are not classified on the map
6. The kappa coefficient (KC) is a measure of the difference between the actual agreement between reference data and an automated classifier and the

chance agreement between the reference data and a random classifier (Lillesand *et al.*, 2015). The value of kappa was expressed by:

$$\hat{k} = \frac{\text{PO-PE}}{\text{1-PE}}$$

Where $\hat{k}$ is a kappa coefficient, PO is an overall proportion of observed agreement or overall accuracy, and PE is an overall proportion of agreement.

To determine whether Sentinel-2 imagery significantly yielded better accuracy than ASTER imagery at the 95% confidence level, Z-test was performed on the Kappa coefficient of each confusion matrix. In this research, 4 error matrices will be made for 4 land use/land cover classification maps, i.e. two classification maps from each Sentinel-2 & ASTER imagery which was acquired from the K-means clustering method and two classification map from each Sentinel-2 & ASTER imagery which was acquired from LVQ clustering method. The significance test on an error matrix is expressed by:

$$Z_1 = \frac{\hat{k}}{\sqrt{\widehat{\text{var}}(\hat{k})}}$$

where Z_1 is a standard normal deviate, $\hat{k}$ is a kappa coefficient, and var $\hat{k}$ is a variance value. The variance value of an error matrix is the square of the standard error. A standard error value can be express as:

$$\text{SE}(\hat{k}) = \frac{\text{SD}(\hat{k})}{\sqrt{\text{N}}}$$

where SE $\hat{k}$ is a standard error, SD $\hat{k}$ is a standard deviation, *N* is a total number of observations included in a matrix. Then, the standard deviation value of a confusion matrix as follows:

$$\text{SD}(\hat{k}) = \sqrt{\frac{\text{PO(1-PO)}}{(\text{1-PE})^2}}$$

where SD $\hat{k}$ is a standard deviation, PO is an overall proportion of observed agreement or overall accuracy, and PE is an overall proportion of agreement. The test statistic for testing whether Sentinel-2 imagery significantly yielded better accuracy than ASTER imagery are significantly different is expressed by:

$$Z_{12} = \frac{|\hat{k}_1\text{-}\hat{k}_2|}{\sqrt{\widehat{\text{var}}(\hat{k}_1) + \widehat{\text{var}}(\hat{k}_2)}}$$

where Z_{12} is a standard normal deviate, $\hat{k}_1$ is a kappa coefficient for an error matrix, $\hat{k}_2$ is a kappa coefficient for the other error matrix, var $\hat{k}_1$ is a variance value for an error matrix, and var $\hat{k}_1$ is a variance value for the other error matrix.

At the 95% (99%) confidence level, the critical value would be 1.96. For a single confusion matrix test, a value of the Z statistic > 1.96 means the result is significant (i.e., better than random) at the 95% confidence level. For a test between two confusion matrices, a value of the Z statistic > 1.96 means the results are significantly different, i.e., one method outperformed the other (Zhou *et al.*, 2018).

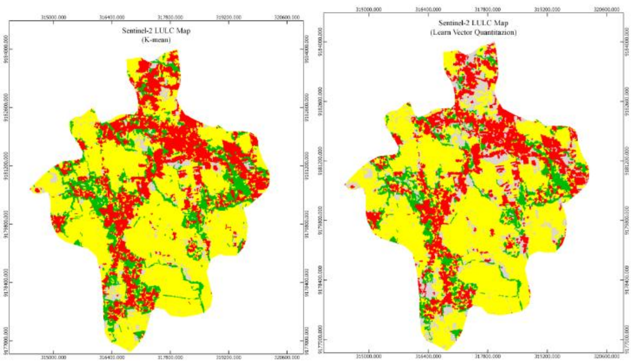
Data Analysis

Two map assessment/analysis methods will be based on the analytical statistical technique of kappa analysis. The Kappa analysis is a discrete multivariate technique used in accuracy assessment for statistically determining whether the accuracy of map classification which was acquired from Sentinel-2 and ASTER imageries with K-means and LVQ algorithms are significantly different. The result of performing a Kappa analysis is a KHAT statistic, which can also be used as another measure of agreement or accuracy. This measure of agreement is based on the difference between the actual agreement in the error matrix. The KHAT statistic is similar to the more familiar χ^2 analysis. The error matrices will also give the user's accuracy, producer's accuracy, the error of commission, and error of omission values. These four components will be analyzed for the relationship between the producer's accuracy and error of omission, the relationship between the user's accuracy and error of commission, and the meaning of each of these components. The assessment of the maps is involving the consideration of the error matrices and kappa analysis with the descriptive statistics method.

RESULTS AND DISCUSSION

Land Use/Land Cover Maps

The result was images with groups/classes of pixels each represented by a different color. (Yadahalli *et al.*, 2018). Red, yellow, green, and grey are representing built-up, agricultural, forest, and barren land, respectively.



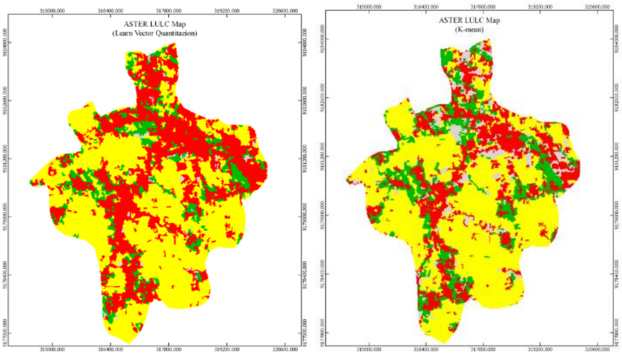


Figure 6. The maps classification results.

Kappa Analysis Result

Overall accuracy of LVQ algorithm for the Sentinel-2 and ASTER imageries was 78.33% and 69.17%, respectively; while the kappa coefficient of LVQ algorithm for the Sentinel-2 and ASTER imageries were 0.71 and 0.55, respectively. In different circumstances, overall accuracy of K-means algorithm for the Sentinel-2 and ASTER imageries was 81.25% and 72.68%, respectively; while the kappa coefficient of K-means algorithm for the Sentinel-2 and ASTER imageries were 0.74 and 0.61, respectively. Z values for each of Sentinel-2 and ASTER classification maps which were involved from K-means and LVQ clustering methods are 2.77 and 2.83, respectively. For both LVQ and K-means classification algorithms, image classification accuracies of the Sentinel-2 and ASTER datasets are significantly different because both of the z values are greater than 1.96.

*Table 4. **Kappa analysis results from the Sentinel-2 imagery.***

	K-means Algorithm				LVQ Algorithm			
	UA (%)	**PA (%)**	**EC (%)**	**EO (%)**	**UA (%)**	**PA (%)**	**EC (%)**	**EO (%)**
Built-up	84.06	84.06	15.94	15.94	90.00	85.71	10.00	14.29
Agricultural	78.48	77.50	21.52	22.50	85.00	62.96	15.00	37.04
Bare land	50	66.67	50.00	33.33	46.67	96.55	53.33	3.45
Forest	95.31	87.14	4.69	12.86	91.67	82.09	8.33	17.91

*Table 5. **Kappa analysis results from the ASTER imagery.***

	K-means Algorithm				LVQ Algorithm			
	UA (%)	**PA (%)**	**EC (%)**	**EO (%)**	**UA (%)**	**PA (%)**	**EC (%)**	**EO (%)**
Built-up	79.71	74.32	20.29	25.68	66.99	93.24	33.01	6.76
Agricultural	73.86	80.25	26.14	19.75	68.04	81.48	31.96	18.52
Bare land	45.16	70.00	54.84	30.00	75.00	28	25.00	72.00
Forest	75.00	60.00	25.00	40.00	77.78	43.08	22.22	56.92

Discussion

Unsupervised classification is a method which examines a large number of unknown pixels and divides into several classes based on natural groupings in the image value. Unsupervised classification does not require analyst-specified training data. The classes that result from the unsupervised classification are spectral classes which are based on natural grouping of the image values, the identity of the spectral class will not be initially known, must compare classified data from reference data to determine the identity and information values of the spectral classes (Yadahalli and Bellakki, 2018).

Based on the kappa analysis results, a remote sensing analyst can conclude two images quality whether the classification results of two different images have different or the same accuracy. This consideration was based on the kappa analysis, especially on the z-test. The

knowledge of the overall accuracy and kappa coefficient gives a rough estimation that the Sentinel-2 image has better accuracy than the ASTER image. Because of the higher overall accuracy and kappa coefficient values will not always indicate better accuracy. The accuracy of the two images can be on the same accuracy. With the z test, the level of accuracy can be found more precisely. Z values for both classification maps which were acquired from each K-means and LVQ clustering methods are 2.77 and 2.83, respectively. The image classification accuracies of the Sentinel-2 and ASTER datasets are significantly different because both of the z values are greater than 1.96. Thus, Sentinel-2 map classification has better accuracy than the ASTER map classification.

Considering the overall accuracy and kappa values, it is assumed that spatial resolution contributed to the map's accuracy. The Sentinel-2 classification which has a spatial resolution of 10 m has better accuracy than the ASTER image which has a spatial resolution of 15 m. The higher the resolution of an image, the wider the range of spectral values. The wider the range of spectral values, the more various colors are displayed. If the color diversity is high, the grouping of the spectral values will be difficult. High color diversity in a pixel causes the image to have mixed pixels. Mixed pixels occur when a sensor's IFOV includes more than one land use/land cover feature from the earth's ground, so its pure spectral responses of specific features are mixed. A classifier will not be possible to clearly/correctly identify that pixel into a class of land use/land cover. Therefore, a pixel that has the same spectral response are combined into one class. To minimize the mixed pixel effect, the sieve filters were applied where polygons smaller than the size of 8 × 8, pixels are merged with the largest neighboring of the pixel's polygon. Figure 2 shows an example of the salt-and-pepper effect. The red pixel (built-up) shows that pixel was built-up, but it is actually agricultural. After the sieving process, the red pixels turned into yellow (agricultural).

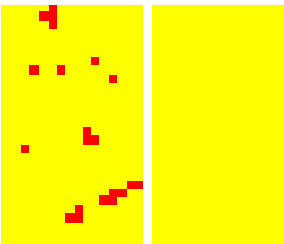


Figure 7. *The Example of salt-and-pepper effect. Before (left side) and after (right side) the sieving process.*

The labeling process also contributed to the map's accuracy. Labeling mistakes occur because a poorly analyst to match the spectral clusters with the reference data correctly. For example, a built-up is classified by K-means algorithm as cluster 52. However, at a different location, bare land is classified as cluster 52 too. If cluster 52 is labeled as the built-up, the built-up will include the bare land (commission error) and the built-up land pixels will lack its pixels (omission errors). If cluster 52 is labeled as built-up, the built-up will include the bare land (commission error) and the built-up pixel will lack its pixels (omission error).

Errors of classifiers in the classification pixels or analyst errors in the labeling clusters process will be causes errors of commission and errors of omission. If a pixel is not included in a proper land use/land cover class, it will cause errors of omission. On the other hand, if too many pixels are classified into a not proper land use/land cover class, it will cause errors of commission. The smaller the error of omission and error of commission value, the bigger the producer's accuracy and user's accuracy, respectively. For example, in the Table 1, the built-up have biggest UA and PA values (UA = 84.46%, PA = 84.06%) because of the small EC and OC values (EC = 15.94%, OC = 15.94%), while bare land have biggest UA and PA values (UA = 50.00%, PA = 66.67%) because of the big EC and OC values (EC = 50.00%, OC = 33.33%).

Table 1 indicates an overall accuracy of 81.25% on the Sentinel-2 map which was acquired from K-means algorithm. However, the forest results in a good producer's accuracy of 95.31%, and its user's accuracy was 87.14%. Thus, the results of land classification on the sentinel-2 imagery obtained from the k-means algorithm have more adequate accuracy to classify the forests. Although the overall accuracy of this map claimed was 81%, the producer's accuracy of the bare land was 66.67% and its user's accuracy was only 50%. That is, even though 66.67% of the bare land areas have been identified as bare land, only 50% of the areas identified as "bare land" actually be bare land. The only highly reliable category associated with this classification from both a producer's and a user's perspective is bare land.

The K-means algorithm has bigger overall accuracy (81.25% for the Sentinel-2 dataset and 72.68% for the ASTER) and a kappa coefficient (0.71 for the Sentinel-2 dataset and 0.63 for the ASTER) values than the overall accuracy (78.33% for the Sentinel-2 dataset 69.17% for the ASTER) and the kappa coefficient (0.74 for the Sentinel-2 dataset and 0.55 for the ASTER) values from the LVQ algorithm. Therefore, K-means algorithm has better accuracy then the LVQ algorithm to classify the land use/land cover for Sentinel-2 and ASTER images.

CONCLUSION

As a result of the study, overall accuracy of LVQ algorithm for the Sentinel-2 and ASTER imageries was 78.33% and 69.17%, respectively; while the kappa coefficient of LVQ algorithm for the Sentinel-2 and ASTER imageries were 0.71 and 0.55, respectively. In different circumstances, overall accuracy of K-means algorithm for the Sentinel-2 and ASTER imageries was 81.25% and 72.68%, respectively; while the kappa coefficient of K-means algorithm for both imageries were 0.74 and 0.61, respectively. At the 95% confidence level, for both LVQ and K-means classification algorithms, image classification accuracies of Sentinel-2 dataset are significantly different than the ASTER dataset. Sentinel-2 imagery provides better accuracy than ASTER imagery in land use/land cover classification data from any unsupervised classification algorithms.

ACKNOWLEDGEMENTS

The author would like to thank Allah SWT, who has given all His blessings. Also, the author thanks the author's parents who have given attention, love, enthusiasm, endless prayer, and various research support facilities, so that the author can complete this research.

REFERENCES

Bakr, N., & Afifi, A. A. (2019). Quantifying land use/land cover change and its potential impact on rice production in the Northern Nile Delta, Egypt. Remote Sensing Applications: Society and Environment, 13, 348-360.

Board, S. S., & National Research Council. (2003). Using Remote Sensing in State and Local Government: Information for Management and Decision Making. National Academies Press.

Congalton, R. G., & Green, K. (2019). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC Press, Boca Raton.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. Remote sensing of environment, 80(1), 185-201.

Kaura, S., & Bansal, R. K. (2018). Mixed Pixel Classification by Using Hybridization of Evolutionary Method With Neural Networks. International Journal of Applied Engineering Research, 13(8), 5736-5744.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). Remote Sensing and Image Interpretation. John Wiley & Sons, Hoboken.

Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. International Journal of Geosciences, 8(04), 611.

Su, T. C. (2016). A filter-based post-processing technique for improving homogeneity of pixel-wise classification data. European Journal of Remote Sensing, 49(1), 531-552.

Yadahalli, G. S., & Bellakki, M. A. (2018). Accuracy Assessment of Supervised and Unsupervised Classification using Landsat-8 Imagery of D-7 Shahapur Branch Canal of UKP Command Area Karnataka, India. Int. J. Curr. Microbiol. App. Sci, 7(7), 205-216.

Zhou, Z., Ma, L., Fu, T., Zhang, G., Yao, M., & Li, M. (2018). Change Detection in Coral Reef Environment Using High-Resolution Images: Comparison of Object-Based and Pixel-Based Paradigms. ISPRS International Journal of Geo-Information, 7(11), 441.