Support Vector Machine for Land Cover Classification using Lidar Data

M I Hariyono¹,², Rokhmatuloh¹, M P Tambunan¹, R S Dewi²

¹Department of Geography, Faculty of Mathematics and Natural Science, University of Indonesia, Depok City, 16424, Indonesia
²Geospatial Information Agency (BIG), Jl. Raya Jakarta-Bogor KM.46 Cibinong, 16911, Indonesia

Email : moch.irwan@ui.ac.id, mochamad.irwan@big.go.id

Abstract. The Lidar technology is widely used in various studies for mapping needs. In this study was to extract land cover using Lidar data by incorporating a support vector machine (SVM) approach. The study was located in the city of Lombok, Nusa Tenggara Barat. Image extraction was performed on single wavelength Lidar data to produce intensity and elevation (Digital Surface Model) features. Feature extraction of Lidar data was implemented by using a pixel-based approach. The extracted features used as an attribute for training data to generate the SVM prediction model. The prediction model to predict the types of land cover in the study area such as buildings, trees, roads, bare soil, and low vegetations. For accuracy assessment purposes, we used topographic map available in shapefile format as the reference map and estimated the accuracies of the resulted classifications. In this study, land cover classification used combination bands which improved the overall accuracy by approximately 20%. The use of the intensity data in this band combination was the reason for the increasing accuracy.

1. Introduction
Currently, many studies use Lidar data as the main data, this data can be used for various applications, such as topographic mapping, DEM modeling [1], building extraction [2,3] vegetation [4] and others. In simple terms, in many researches, Lidar data was used to classify ground and non-ground. Furthermore, Lidar data can be useful for land cover classification [5,6,7,8]. Lidar has high accuracy and resolution for both vertical and horizontal directions. From the Lidar data, features are obtained in the form of height and intensity. The intensity data provides good information in land cover classification processes [9,10] The height features commonly used for classification are DSM, nDSM, DTM and Slope. In this research, we used intensity data from Lidar point cloud. Based on previous study conducted by Hui et al. [11], land cover classification method by combining intensity data with height data provided good results. Hui stated that intensity data has advantages compared to multispectral remote sensing data by avoiding shadows that usually exist in multispectral data [11].

In literature, researches by using Lidar intensity data are increasingly being developed, and carried out in various combinations to improve the accuracy of land cover classification results by combining intensity data with other features [12]. Another study by Yan et al. [13] conducted an evaluation before and after radiometric correction. In their study, land cover homogeneity was also incorporated so that it
helped in increasing feature extraction or similarity of land cover surfaces which was further increased the accuracy of the classification results after radiometric correction.

Various methods have been used for land cover classification processes, one of which is machine learning support vector machine (SVM). According to Noi and Kappas [14] this method shows a better accuracy compared to random forest (FR) method, and K-Nearest Neighbor (kNN). SVM has capabilities that have been tested and have a major role in classification processes [15]. Another study using SVM for land cover classification with a combination of Lidar data and ortho image shows excellent accuracy results from the combination of these data [16]. The purpose of this study was to extract land cover using Lidar data by incorporating a support vector machine (SVM) approach.

2. Methodology
2.1 Study Area
The study area was located in a part of the village of Tanjung Karang, Mataram, Nusa Tenggara Barat (NTB), Indonesia. This area was selected because the area has various land cover features on the ground including buildings, vegetation, and roads. Several datasets were used in this study including: a) Lidar data (Figure 1a); b) orthophoto (Figure 1b) and c) topographic map (RBI) provided by Geospatial Information Agency (BIG, 2017) at scale 1: 5000. The point cloud density of the Lidar data was approximately 10 points per square meters (ppm). The aerial photo was acquired at the same time with the Lidar data which was recorded in 2016. Figures 1a-b represent the Lidar and the ortho-rectified aerial photos (orthophoto) of the study area.

![Figure 1](image1.png)  
(a) Lidar data  
(b) Orthophoto data

**Figure 1.** Study area that has various land cover feature presented by using: a). Lidar, and b). Orthophoto

2.2 Data Preparation
The data sets used in this research were prepared by converting Lidar point cloud into raster images with pixel size equals to 15 cm. Lidar data were filtered to obtain ground and non-ground information, the filtering process was carried out using Global Mapper software. New images (bands) were created namely: DTM (digital terrain model), DSM (digital surface model), nDSM (normalized digital surface model), and intensity. In this step, DSM, DTM and intensity data were extracted by using tools provided in Global Mapper software. DSM was extracted from non-ground data while DTM was extracted from ground data. Following these, we extracted nDSM by calculating the differences between DSM and DTM data. Figures 2a-c illustrate the produced Lidar raster images (intensity, DSM, and nDSM).

Furthermore, to evaluate the influence of intensity data on the classification results, seven band combinations were created consisting of one, two, and three band combinations, namely: 1) DSM; 2)
intensity; 3) nDSM; 4) DSM and intensity; 5) DSM and nDSM; 6) Intensity and and nDSM; 7) DSM, Intensity and nDSM. We used these seven band combinations as the input image when performing classification.

![Figure 2](image_url)

**Figure 2.** The visualisation of a) Intensity, b) DSM, c) nDSM

### 2.3 Data Classification

Support vector machine is a supervised algorithm that requires training data before applying the model. For the study area, we identified five land cover classes, namely buildings, trees, roads, bare soil, and low vegetations. SVM was then performed on seven band combinations, and statistical assessment was conducted for the training data. We evaluated the classification accuracies obtained when performing SVM and added more training data if required.

In this research, we used Radial Basis Function (RBF) when performing SVM. From the previous study, the use of these functions affect the efficiency of the classification when performing SVM and increase the accuracy [17], besides using two influential parameters namely cost (C) and gamma (γ) [18].

For accuracy assessment purposes, we used topographic map available in shapefile format as the reference map and estimated the accuracies of the resulted classifications by performing error matrix. We calculated the overall accuracy values based on confusion matrix principle. The calculation provides for a specific test between the classification results and the conditions at ground level [19].
3. Results and Discussion

Table 1 presents the overall accuracy resulted from the experiment when implementing SVM for land cover classification of seven combination of bands. From the results, we can see that the overall accuracies by using DSM and intensity are less than 50%. Meanwhile, by combining both DSM and intensity data as the input images improved the results to 68%. Furthermore, using nDSM does not improve the accuracy. Nevertheless, combining DSM, nDSM and intensity data can improvement in the overall accuracy of the classification results. We obtained the overall accuracies for about 75-77%. In conclusion, by using three band combinations produced the best performance of SVM classification by 77.9%.

Figures 3a-g show the classification results for each band combination when we performed SVM in image classifications. For visualization, we compared the classification results with the reference map in Figure 3h. In Figure 3a, only three classes could be identified namely building, trees and low vegetation. Meanwhile in Figure 3b, we can see from the classification result that only vegetation classes such as trees and low vegetation were classified. It might be due to higher intensity value of vegetation classes in the NIR range [12]. Furthermore, in Figure 3c, building features could be identified instead of vegetation classes. By adding more bands, the classification results produced increase accuracy and more features were identified (see Figures 3d-g). From these experiments, the errors in classification results may be due to lack of training data for each classes or maybe due to the similarity in intensity values between land cover classes.

Table 1. The overall accuracy results when performing SVM by using seven band combinations

| No. | Band Combination          | Overall accuracy |
|-----|---------------------------|------------------|
| a.  | DSM                       | 45.9 %           |
| b.  | Intensity                 | 42.7 %           |
| c.  | nDSM                      | 48.8 %           |
| d.  | DSM and Intensity         | 68.1 %           |
| e.  | DSM and nDSM              | 75.2 %           |
| f.  | Intensity and nDSM        | 70.8 %           |
| g.  | DSM, Intensity and nDSM   | 77.9 %           |
Figure 3. The results land cover classifications by performing SVM using seven band combinations: a) DSM, b) Intensity, and c) nDSM. d) DSM and Intensity, e) DSM and nDSM), f) Intensity and nDSM, g) DSM, Intensity and nDSM, compared with h) the reference map.
4. Conclusion
In this study, we elaborated intensity image extracted from Lidar data for land cover classification by using a machine learning algorithm, SVM. The results of the land cover classification show that the classification using individual band (DSM, nDSM, Intensity) has an overall accuracy of less than 50%. Then the classification with a combination of two bands improved the results of accuracy by approximately 20% compared with using individual band. By using combination DTM and nDSM, we only obtained 75.2% accuracy value, however by including the intensity band for combination, we can improve the accuracy of land cover classification up to 77.9%.

Acknowledgments
We thank Geospatial Information Agency of Indonesia for all the materials that have been provided as support for this research.

References
[1] Hariyono M I and Windiastuti R 2018 Classification of LiDAR Data to Generate Digital Terrain Model 39th Asian Conf. Remote Sens 1743–1754
[2] Meng X, Wang L, and Currit N 2008 Morphology-based Building Detection from Airborne Lidar Data ASPRS 75 437–442
[3] Miliaris G and Kokkas N 2007 Segmentation and object-based classification for the extraction of the building class from LIDAR DEMs Comput. Geosci 33 1076–1087
[4] Nardinocchi C, Forlani G, and Zingaretti P 2003 Classification and filtering of laser data Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci 79–86
[5] Mallet C, Bretar F, Roux M, Soergel U, and Heipke C 2011 Relevance assessment of full-waveform lidar data for urban area classification ISPRS J. Photogramm. Remote Sens. 66 S71–S84
[6] Chen Z and Gao B 2014 An Object-Based Method for Urban Land Cover Classification Using Airborne Lidar Data IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 4243–4254
[7] Morsy S, Shaker A, and El-Rabbany A 2017 Multispectral LiDAR Data for Land Cover Classification of Urban Areas Sensors 17 5 958–979
[8] Pan S, Guan H, Yu Y, Li J and Peng D 2019 A Comparative Land-Cover Classification Feature Study of Learning Algorithms: DBM, PCA, and RF Using Multispectral LiDAR Data IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 4 1314–1326
[9] Song J, Han S, Yu K, and Kim Y 2002 Assessing The Possibility Of Land-Cover Classification Using Lidar Intensity Data Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 34 3B 259–262
[10] Flood M 2001 Lidar Activities and Research Priorities in The Commercial Sector Int. Arch. Photogramm. Remote Sens. XXXIV-3/W4 22–24
[11] Hui L, Di L, Xianfeng H, and Deren L 2008 Laser Intensity Used In Classification Of Lidar Point Cloud Data IGARSS 2 1140–1143
[12] Shaker A and El-ashmawy N 2012 Land cover information extraction using lidar data Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXIX-B7 XXII ISPRS Congress167–172
[13] Yan W Y, Shaker A, Habib A and Kersting A P 2012 Improving Classification Accuracy of Airborne LiDAR Intensity Data by Geometric Calibration and Radiometric Correction ISPRS J. Photogramm. Remote Sens. 67 1 35–44
[14] Noi P T and Kappas M 2017 Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery Sensors 18 18–37
[15] Shao Y and Lunetta R S 2012 Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points ISPRS J.
Photogramm. Remote Sens. 70 78–87
[16] Tseng Y H, Wang C K, Chu H J and Hung Y C 2015 Waveform-based point cloud classification in land-cover identification Int. J. Appl. Earth Obs. Geoinf. 34 1 78–88
[17] Qin R A 2015 Mean Shift Vector-Based Shape Feature for Classification of High Spatial Resolution Remotely Sensed Imagery IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8 5 1974–1985
[18] Jozdani S E, Johnson B A and Chen D 2019 Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use Land Cover Classification Remote Sensing 11 1713–1736
[19] Foody G M 2002 Status of land cover classification accuracy assessment Remote Sensing of Environment 80 1 185-201