

MANGROVE CLASSIFICATION USING SUPPORT VECTOR MACHINES AND RANDOM FOREST ALGORITHM: A COMPARATIVE STUDY

F. Campomanes^a, A. V. Pada^a, J. Silapan^b

^aUniversity of the Philippines Cebu Phil-LiDAR 2, Gorordo Avenue, Lahug, Cebu City – (enzo.campomanesv, avspada)@gmail.com

^bUniversity of the Philippines Cebu, Gorordo Avenue, Lahug, Cebu City – jsilapan@hotmail.com

KEY WORDS: Mangroves, Support Vector Machine, Random Forest, LiDAR, OBIA

ABSTRACT:

Mangrove forest ecosystems fulfil a number of important functions like supporting the conservation of biological diversity by providing habitats, nurseries, and nutrients for animal species. In the Philippines, mangrove forests are declining due to the growth of aquaculture production. Mangrove forests are slowly being replaced by fishponds. An accurate inventory of what are left of these natural resources is important to know how we can conserve and manage them. This study aims to compare the performance of support vector machines (SVM) with random forest (RF) algorithm in automatically classifying mangrove forests using LiDAR data and orthophotographs in an object based approach. The site is a 36 sq. km. coastal area in Manapla, Negros Occidental, Philippines. Various derivatives were created from the LiDAR data like the pit-free canopy height model (CHM) and intensity. The CHM was used in contrast split segmentation to distinguish between ground and non-ground objects. Only the non-ground objects were segmented further knowing that majority of mangroves are tall. Inventory of short mangroves is not yet included in this study. The non-ground objects were further segmented using multiresolution segmentation with the CHM and RGB bands of the orthophoto using a scale of 15. The non-ground class was further separated into four classes namely: mangrove, built-up, other trees, and sugarcane. 120 training points and 30 validation points per class were collected by visual inspection using the orthophoto as reference. Several features of the training objects were computed from both the LiDAR and orthophoto derivatives and used for classification. SVM with radial basis function was used to classify the rest of the image and resulted in an overall accuracy of 95.83%. For mangroves, its precision and recall reached 83.33% and 100%, respectively. In the SVM classification, mangroves were confused with other trees and sugarcane. Another machine learning algorithm, random forest (RF) was used to classify the same area to compare their performance and accuracy. Using the same features, the RF classification achieved an overall accuracy of 99.1667%. For mangroves, the RF classification obtained 100% and 96.70% for its precision and recall, respectively. The RF classifier confused other trees with mangroves which caused the error. The accuracies from both machine learning algorithms show that the RF classifier performed better than the SVM classifier and further implies the potential of using RF in classifying mangroves in other areas.

1. INTRODUCTION

Mangrove forests are very important coastal resources because of the many benefits it gives us especially in our socio-economic development. It protects our coast from UV-B radiation, “green house” effects, as well as floods and sea level rise (Kathiresan, 2012). They are called the “rainforests of the seas” as they are the backbone of coastal and offshore food webs (Carter et al., 2015). Mapping of these mangroves is a vital input for mangrove management plans to be able to conserve these very important coastal resources.

There have been many approaches to mapping mangroves. Light detection and ranging (LiDAR) and optical remote sensing combined with the support vector machine (SVM) algorithm has been used to map and create a mangrove inventory (David, 2015; Pada, 2016). Other remotely sensed data like Landsat and ALOS-PALSAR have also been used to map mangroves with the random forest (RF) classification (Jhonnerie, 2015).

Comparisons of RF and SVM classification have been done before. RF was found to perform better than SVM in classifying electronic tongue data (Liu, 2013) and RF was also found to be steady and effective compared to SVM in predicting β -hairpin motifs in proteins (Jia, 2013). However, RF was outperformed by SVM in microarray-based cancer classification (Statnikov, 2008). For mangrove mapping using LiDAR and optical images, RF has not yet been used. Furthermore, a comparison of SVM and RF in mangrove mapping has not yet been done.

2. OBJECTIVES

This study aims to compare the performance of SVM and RF in classifying mangroves using LiDAR and optical imagery with an object based approach. It also aims to provide an alternative or maybe even better method to the currently popular method of using LiDAR and SVM approach.

3. DATA AND METHODS

3.1 Study Area

The study site that was selected for this study was an area in Manapla, Negros Occidental, Philippines where there are a lot of mangroves in the area based on local data and optical imagery.

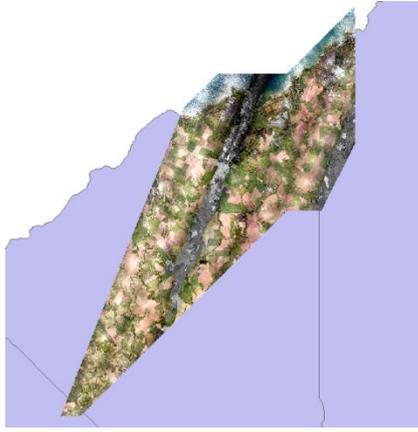


Figure 1. Orthophoto of the study site taken in June 2014. Intensity (from the LiDAR) is shown in areas where there is no orthophoto.

3.2 Data Used

LiDAR and orthophoto data was collected in June 2014 with a point density of 2 points/sq. meter at 1 meter resolution (UP TCAGP, 2015). Table 1 shows the details of the LiDAR and orthophoto data acquisition.

Data Source	Resolution	Date Acquired
LiDAR	1m	June 2014
Orthophoto	0.5m	June 2014

Table 1. Summary of data sources, spatial resolution, and acquisition date.

3.3 Training and Validation Points

Training and validation points were collected through visual interpretation with the optical imagery and other derivatives as reference. Only four (4) classes were collected from the non-ground objects namely builtup (Bu), tall sugarcane (SC3), other trees (OTr), and Mangroves (Mgr). A total of approximately 150 points per class were collected. By rule of thumb, 20% of the 150 points, which is 30 points, were used for validation and the rest for training. Table 3 shows the breakdown of points per class.

	Class	Number of Points
Training	Builtup	120
	Tall Sugarcane	116
	Other Trees	120
	Mangroves	119
Validation	Builtup	30
	Tall Sugarcane	30
	Other Trees	30
	Mangroves	30

Table 2. Breakdown of number of points per class.

The training points were then converted to training samples which were used in the classification phase of the methodology.

3.4 Image Segmentation

The whole process flow for segmentation was implemented using eCognition Developer (Trimble, 2014). The CHM was used in the contrast split segmentation in determining the ground and non-ground objects as most mangroves are known to be tall. Only the non-ground objects were further segmented

using the multiresolution segmentation algorithm (Trimble, 2014) with a scale of 15, shape of 0.3, compactness of 0.8, and putting more weight on the canny edge of the CHM and normalized difference vegetation index (NDVI).

3.5 Features used

Several features from the LiDAR data and optical imagery were utilized for both SVM and RF algorithms. Among these are the red, green, and blue bands from the orthophoto, digital terrain model (DTM), normalized digital surface model (nDSM), canopy height model (CHM), green-red vegetation index (GRVI). Table 2 shows all the 38 features used.

Statistical Method	Feature
Mean	Red, Green, Blue, nDSM, CHM, intensity, GRVI, DTM, number of returns,
Standard Deviation	Red, Green, Blue, nDSM, intensity, number of returns
GLCM Homogeneity	*
GLCM Entropy	*
GLCM Ang. 2 nd moment	*
GLCM Standard Deviation	*
GLDV Entropy	nDSM, intensity

Table 3. Summary of features used. Features with “*” signifies “Red, Green, nDSM, GRVI, intensity”.

3.6 Support Vector Machine Classification

SVM classification was done in eCognition with the use of the training samples using the mentioned features. Radial basis function was the desired kernel to be used due to the complexity of the training data and C parameter of the SVM was set to 1000 in order to lower bias and penalize misclassification heavily.

3.7 Random Forest Classification

The segmented training samples from eCognition were exported as a CSV with all 38 features and used as input in the random forest classification that was implemented in Weka 3.8 (Hall, 2009). Preprocessing was done to the training and test (remaining nonground objects) sets to convert them from CSV to ARFF which is the file format used by Weka. Using the training samples, a classification model was trained using RF with 100 iterations. The trained model was then used to classify the validation samples as well as the remaining test samples.

3.8 Accuracy Assessment

The same validation samples were used for both RF and SVM classification results. Accuracy assessment for the SVM result was done in eCognition by generating an error matrix based on samples. Accuracy assessment was done in Weka by using the trained RF model to classify the validation samples. 30 samples per class were classified for a total of 120 points for 4 classes. Precision, recall, and F-measure values were also calculated.

4. RESULTS AND DISCUSSION

4.1 SVM Classification

SVM was able to completely classify the rest of the image as implemented in eCognition. However, there was obvious misclassification of mangroves in the areas with missing orthophoto. This can be attributed to the fact that the RGB

values of these areas are zero. Objects classified as mangroves were also still found in areas far from the coast and rivers. Also, for areas with fishponds and water, there is misclassification of builtup. This can be because of the triangulation that happens in the CHM or DSM in areas with water.

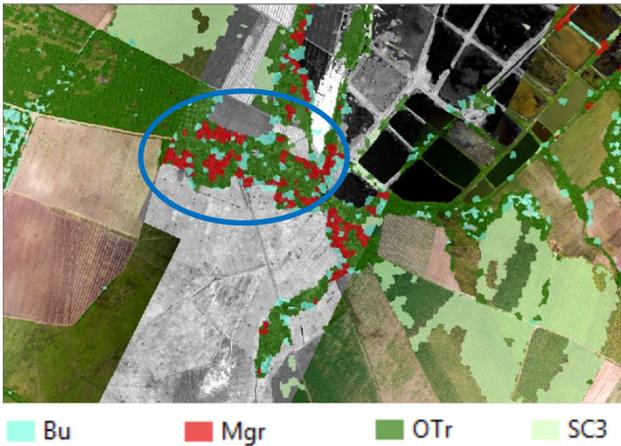


Figure 2. SVM classification result for subset area A. Builtup areas are misclassified as mangroves (encircled in blue).

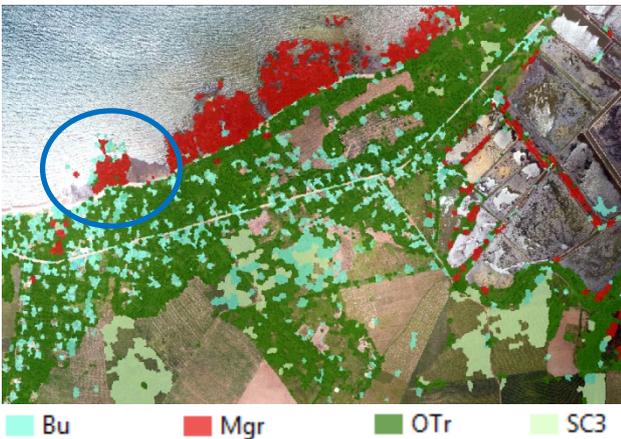


Figure 3. SVM classification result for subset area B. Mangroves are misclassified as builtup in the upper left area (encircled in blue).

Aside from the mentioned misclassifications, the SVM classifier was generally able to classify mangroves correctly. Of the 30 validation samples for mangroves, 25 were classified correctly, 4 were classified as “Other Trees” or “OTr”, and 1 was classified as “Tall Sugarcane” or “SC3”. For the 3 remaining classes, all validation samples were classified correctly. 115 samples were correctly classified and 5 were misclassified. With this, an overall accuracy of 95.83% was obtained.

Bu	Mgr	OTr	SC3	
30				Bu
	25			Mgr
	4	30		OTr
	1		30	SC3

Table 4. Confusion matrix for SVM classification result.

Class	Precision	Recall	F-Measure
Bu	1.00	1.00	1.00
Mgr	1.00	0.83	0.90
OTr	0.88	1.00	0.94
SC3	0.97	1.00	0.98
Mean	0.9625	0.9575	0.96

Table 5. Precision-Recall and F-Measure matrix for SVM classification result.

For the mangrove class, a precision of 1 was obtained while 0.83 was obtained for recall. An F-Measure value of 0.9 was calculated for mangroves. Taking into account all classes, generally higher values were obtained for precision, recall and F-measure at 0.9625, 0.9575, and 0.96, respectively. Recall and F-measure values obtained for the mangrove class are quite low considering that the primary goal of the study is to correctly classify mangroves.

4.2 RF Classification

The trained RF model implemented in Weka was able to classify all objects. Similar to the SVM result, areas with water and fishponds have confusion between builtup and mangroves because of the triangulation of the CHM or DSM in the area.

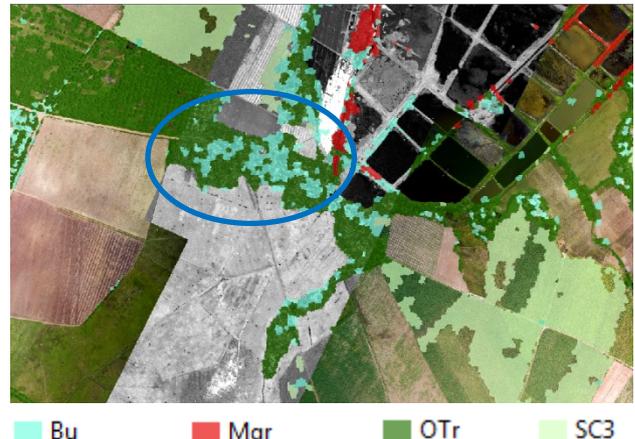


Figure 4. RF classification result for subset area A. Builtup areas are classified correctly, with no misclassification of mangroves.

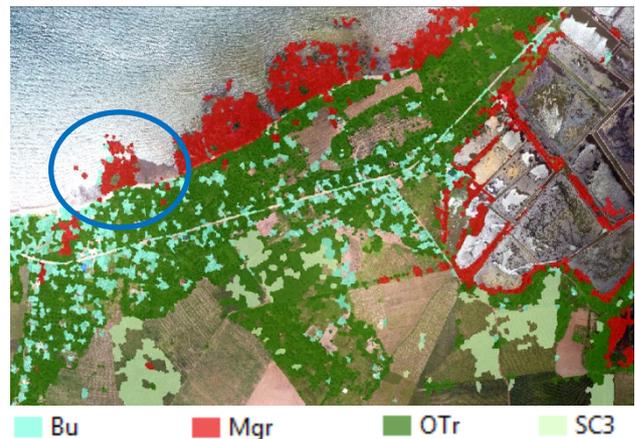


Figure 5. RF classification result for subset area B. Less misclassification of mangroves as builtup is evident in the area encircled in blue compared to the SVM result.

In figure 4, builtup areas are correctly classified and there is no more confusion with mangroves in the area. Even with the lack

of RGB information, the classifier was still able to correctly classify the objects. Figure 5 shows little to no misclassification of mangroves as builtup in the upper left portion of the image (encircled in blue). Also, the RF classification result had no errors with regards to mangroves in the non-coastal areas.

Of the 30 validation samples for mangroves, 29 were classified correctly and 1 was classified as “Other trees” or “OTr”. Overall, 119 of 120 samples were classified correctly. An overall accuracy of 99.17% was obtained. Table 6 shows the confusion matrix for the RF classification result.

Bu	Mgr	OTr	SC3	
30				Bu
	29			Mgr
	1	30		OTr
			30	SC3

Table 6. Confusion matrix for RF classification result.

Class	Precision	Recall	F-Measure
Bu	1.00	1.00	1.00
Mgr	1.00	0.97	0.98
OTr	0.97	1.00	0.98
SC3	1.00	1.00	1.00
Mean	0.992	0.992	0.992

Table 7. Precision-Recall and F-Measure matrix for RF classification result.

The precision and recall values for the mangrove class are 1 and 0.97, respectively. An F-Measure of 0.98 was also calculated for mangroves. Overall, the RF classifier achieved high accuracies at 0.992 for precision, recall, and F-measure.

4.3 Comparison of SVM and RF

Table 8 shows the precision, recall and F-measure values per class and for both SVM and RF.

Class	Precision		Recall		F-Measure	
	SVM	RF	SVM	RF	SVM	RF
Bu	1.00	1.00	1.00	1.00	1.00	1.00
Mgr	1.00	1.00	0.83	0.97	0.90	0.98
OTr	0.88	0.97	1.00	1.00	0.94	0.98
SC3	0.97	1.00	1.00	1.00	0.98	1.00
Mean	0.9625	0.992	0.9575	0.992	0.96	0.992

Table 8. Comparison of accuracies obtained by SVM and RF classifiers.

After comparing the performance of the SVM and RF classifiers based on their obtained accuracies, it was found that RF performed better in the mangrove class as well as the other classes. The RF classifier got higher values in precision, recall, f-measure, and also achieved a higher overall accuracy of 99.17% (119 of 120) compared to the 95.83% (115 of 120) that the SVM classifier obtained. This could be because of the characteristic of SVM to have difficulty in training a good model if there are a lot of training samples. For mangrove mapping, there is a need for a lot of training samples to be able to differentiate classes especially mangroves from other trees. RF, on the other hand, because of how they are constructed, are expected to handle a large number of training data.

5. CONCLUSION AND RECOMMENDATIONS

Based on the results, it can be said that mangroves can be classified using a trained RF model in an object based approach

and should be tested in a different study site to check its applicability for different data. Furthermore, based on the data and performance of both classifiers, it can be said that both the SVM and RF classifiers performed well in classifying mangroves but it should be noted that RF performed better than SVM in areas with no orthophoto. Also, RF performed better in terms of accuracy as it obtained higher values in precision, recall, F-measure, and overall accuracy.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the Department of Science and Technology (DOST) of the Philippines as the funding agency and the Department of Science and Technology – Philippine Council for Industry, Energy and Emerging Technology Research and Development (DOST-PCIEERD) as the monitoring agency.

7. REFERENCES

- Carter, H., Schmidt, S., Hiron, A., 2015. An International Assessment of Mangrove Management: Incorporation in Integrated Coastal Zone Management, *Diversity* 7, pp. 74-104.
- David, L. C., Ballado, A., 2015, Mapping Mangrove Forest from LiDAR Data using Object-Based Image Analysis and Support Vector Machine: The Case of Calatagan, Batangas, *Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, pp. 1-5.
- Hall M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I., 2009. The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, Vol. 11, Issue 1, pp. 10-18.
- Jhonnerie, R., Siregar, V., Nababan, B., Prasetyo, L. B., Wouthuyzen, S., 2015. Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and ALOS PALSAR Imageries. *Procedia Environmental Sciences* 24, pp. 215-221.
- Jia, S., Hu, X., Sun, L., 2013, The Comparison between Random Forest and Support Vector Machine Algorithm for Predicting β -Hairpin Motifs in Proteins. *Engineering* 5, pp. 391-395.
- Kathiresan K., 2012, Importance of Mangrove Ecosystem. *International Journal of Marine Science*, Vol. 2, No.10 pp. 70-89.
- Pada, A. V., Silapan, J., Cabanlit, M. A., Campomanes F., Garcia, J. J., 2016, Mangrove Forest Cover Extraction of the Coastal Areas of Negros Occidental, Western Visayas, Philippines using LiDAR Data, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B1*, pp. 73-79.
- Statnikov, A., Wang, L., Aliferis, C., 2008, A Comprehensive Comparison of Random Forests and Support Vector Machines for microarray-based cancer classification. *BMC Bioinformatics*.
- Trimble, 2014. eCognition Developer Reference Book, pp. 32-39.
- UP TCAGP, 2015, DREAM Lidar Data Acquisition and Processing for Lucena River Floodplain, Disaster Risk And Exposure Assessment for Mitigation Program, *DOST Grants-in-Aid Program*, 55 pp.