Comparison of Support Vector Machines (SVMs) and Maximum Likelihood Classification (MLC) for Nipa Palm (*Nypa fruticans*) Extent in East Coast of Sabah, Malaysia

Mckreddy Yaban, Mohd Arif Shahdan Bin Azmi, Muhamad Zulfazli Zakaria and Maizatuldura Mohd Isa

Malaysian Space Agency (MYSA)
Ministry of Science, Technology, and Innovation (MOSTI)
No. 13, Jalan Tun Ismail, 50480 Kuala Lumpur, Malaysia.

mckreddy@mysa.gov.my; arif@mysa.gov.my; zulfazli@mysa.gov.my; maizatuldura@mysa.gov.my

Abstract: Mangrove forests on the east coast of Sabah are very valuable ecosystems which have been identified as having high potential as carbon storage, coastal buffer zone, tourism values and the fishing industry. Nipa palm (*Nypa fruticans*) is one of the unique species in the mangrove forest that has its own special features as halophytic palm. Nipa forest are highly susceptible to various changes in forest cover especially aquaculture. However, there are no latest and reliable data related to the extent of Nipa Forest in the east coast of Sabah. Forest cover mapping in modern remote sensing has been easier by the use of high-resolution images, advanced image processing and geographical information system (GIS). This study will be conducted to determine the best mapping method to map the extent of Nipa Palm area. We will perform the classification using Support Vector Machine and Maximum Likelihood Classification method using high-resolution imagery from SPOT Satellite. The result showed Support Vector Machine (overall accuracy 90%) produced better than Maximum Likelihood (70%). The significance of this study is to identify the capabilities of these methods for sustainable forest management.

Keywords: Nipa Palm, Forest Cover, Support Vector Machine, Maximum Likelihood

1. INTRODUCTION

Nipa Palm (*Nypa fruticans*) is one of the most important resources for the survival of the people who live in a riparian or coastal area. The local community has historically managed Nipa Palm to produce food and building materials. Nipa Palm Forest area in east coast of Sabah have been decreased for various reasons ranging from the expansion of aquaculture, agriculture, road construction, urbanisation and the use of Nipa Palm for other various purposes. *Nypa fruticans* play an important role in ecological services as a buffer zone for coastal erosion and are habitats for reproductive and nursing aquatic animals.

Digital classification algorithms have been in use for landuse preparation since 1972 from Landsat satellite data (Townshend, 1992; Lu and Weng, 2007). Maximum Likelihood Classification (MLC) has become most popular and widely used in order to perform a land cover or forest cover with an acceptable rate of accuracy. The MLC classification is based on a parametric approach which implies the assumption of the selected signature classes within the normal distribution. Based upon the mean and covariance of the data this classifier generates decision surfaces (Srivastava et al., 2012). Some non-parametric based classification techniques have been used for extracting major classification as well as sub-classification with better accuracy (Kavzoglu and Reis, 2008). Among famous non-parametric classification techniques are Decision Trees, Fuzzy C-Mean, Artificial Neural Networks (ANN) and Support Vector Machines (SVMs). Support Vector Machines are based on statistical theory which is used for classification and regression problems (Vapnik, 1995). The classification accuracy produced by SVMs may show variation depending on the choice of the kernel function and its parameters (Kavzoglu and Colkesen, 2009). Performance of SVMs classification technique in tropical coastal area was done by Szuster et al. (2011). He also compared this method with the MLC and ANN techniques and has concluded that the SVMs is a better classifier for features that possess similar spectral signatures. Yu et al. (2012) has applied the SVMs algorithm for the automated lithological classification in a part of north-western India using ASTER imagery. He has also showed that SVM gives higher accuracy in comparison to MLC.

Figure 1: Nipa Palm area in Kuala Bonggaya, Sabah

2. STUDY AREA

The study area covers 586.78 km^2 of the Kuala Bonggaya & Kuala Labuk Forest Reserved, Sabah, Malaysia. Three major rivers flow through the study area, including the Klagan, Pimpim and Sapi rivers. The study extends from 5° 50' 7.6632" to 6° 9' 52.6032" latitude and 117° 22' 22.0296" to 117° 44' 26.9808" longitude, approximately 62 km from Sandakan, Sabah. This region is dominated by mangroves and nipa palm. However, there are also oil palm plantations in the west, as well as settlements in the southern part of the study area.

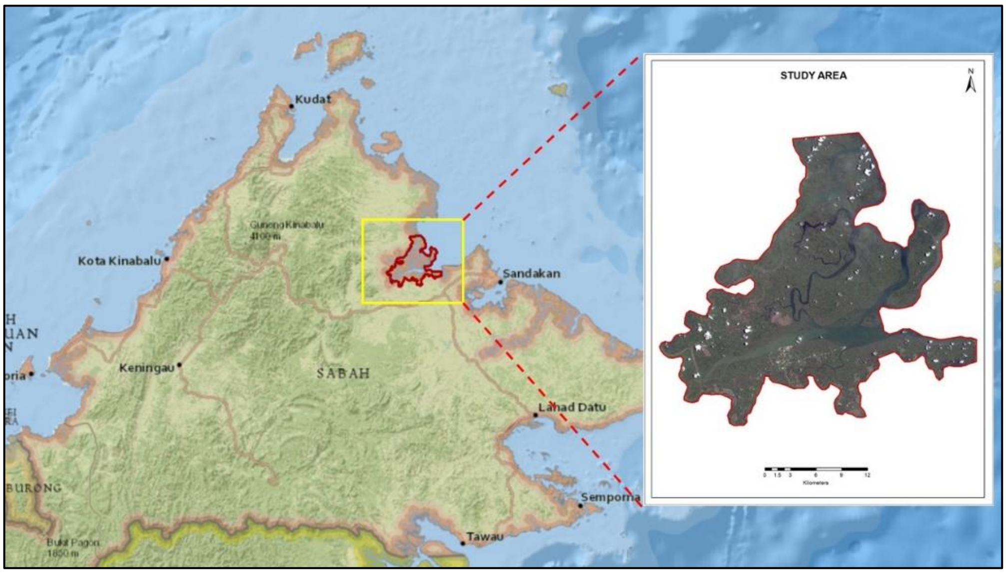

Figure 2: Location of the study area

3. MATERIAL AND METHODS

3.1 Data

SPOT7 satellite imagery with a resolution of 1.5m has been selected for the classification of forest cover. The acquisition date for the satellite imagery used in this study is 20 July 2021. Some corrections were made, such as radiometric and geometric corrections using raw imagery prior to land cover analysis. The swath width of this satellite image was 60 km by 60 km. Satellite image was subsets within the study area to obtain a small part of the image from a large data file. Sub-setting of the data for the study area also speeds up the processing time of the image, especially in the case of SVM image classification.

3.2 Method

Ten different sets of representative polygons of training samples were created in a random manner (Figure 4). Additionally, 150 reference points were collected through surveys in the study area for classification validation and accuracy assessment. For each classification class that represents the study area, the sampling area and reference data were generated using independently designed polygons and points. The class sampling used for both classification techniques is the same to reduce the inconsistency of the final forest cover classification results. SVMs and MLC techniques were applied to assess the accuracy of forest cover maps using corrected satellite images.

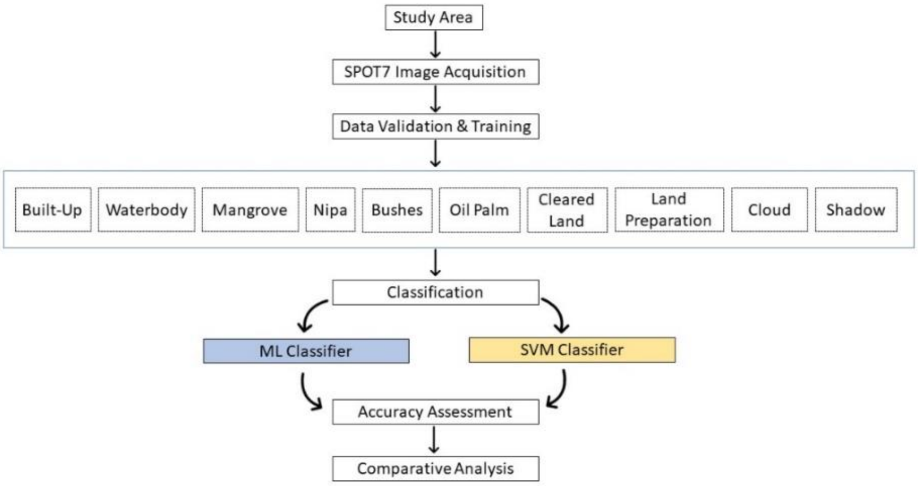

Figure 3: Flow chart methodology

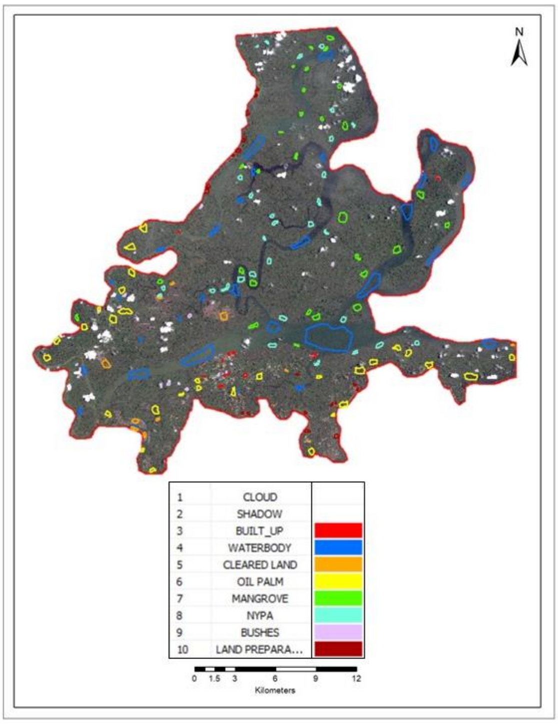

Figure 4: Map of 300 class samples

3.2.1 Support Vector Machines

SVMs are non-parametric supervised learning algorithms based on statistics. SVM is considered to be more transparent, easier to use, and often achieves better generalization comparing to the multilayer perceptron's neural networks (Yang et al., 2011). SVMs based approach to seeks a fit optimal hyperplane between the classes and since it uses only the training samples that rely on class distributions in feature space (support vectors), it may require only a small training sample (Dhriti et al., 2015). According to Szuster (2011), the algorithm finds an Optimal Separating Hyper-plane (OSH) between each pair of classes. This OSH is obtained by using training data from the study area. Hyperplanes are decision boundaries used to categorize data points. Therefore, the main purpose of SVM is to find out the OSH among all other separating hyper-planes which is achieved through an optimization problem by using quadratic programming methods and Lagrange multipliers (Szuster et al., 2011). All hyper-planes separate two classes but OSH minimizes the generalization error in classification by maximizing the margins (distance) between the two classes.

3.2.2 Maximum Likelihood Classification

MLC is based on estimation parameter of probabilistic model. The MLC method assumes that the statistics for each class in each band are normally distributed and calculates the likelihood that a particular pixel belongs to a specific class. The MLC classification is based on a parametric approach which assumes the selected signature classes in the normal distribution (Al-Ahmadi et al., 2009). The MLC method is based on the Bayes theory of decision-making in which the distribution of data in multidimensional space is linear or normal. The MLC method examines statistical data like covariance and variances of the selected class signatures when assigning values to the unknown pixels from represented class signature file. To calculate how relevant the cell is to the class, the statistical probability is calculated for each class. With prior specification of weightage each pixel has been identified to be a member of landcover class to which it has highest similarity under the rules (Srivastava et al., 2012).

3.2.3 Accuracy Assessment

The accuracy assessment is an important method that must be implemented to determine the accuracy of the classification that was performed. The determination of accuracy assessment points is carried out using the Stratified Random Sampling method to ensure that there is no human influence in the determination of each point used. Error matrix is considered as one of the common techniques for the measurement of thematic map accuracy (Foody, 2002). It is calculated by obtaining a sample from a particular class of a classified map and then the actual class is validated from the field (Congalton, 1991). The accuracy of each classification is expressed in the form of an error matrix.

Another accuracy indicator is the kappa coefficient. This is a measure of the comparison between the results of the classification and the values attributed. Kappa statistic is a measure of how closely the instances classified by the machine learning classifier matched the data labelled as ground truth, controlling for the accuracy of a random classifier as measured by the expected accuracy. If the kappa coefficient is 0, indicates that the classified image is not correlated with the reference image. The classified image and the ground truth image are totally identical when the Kappa coefficient equals to 1. The higher the kappa coefficient, the more accurate the classification is.

Overall accuracy is the probability that an individual will be correctly classified by a test; that is, the sum of the true positives plus true negatives divided by the total number of individuals tested (Anthony et al., 2015). Overall accuracy then gives the overall results of the error matrix. Kappa statistics are used to control for only those instances that may have been classified correctly by chance. It can be calculated using observed (total) accuracy and random accuracy (Ukrainski et.al, 2019).

Figure 5: Accuracy assessment points (550 point)

4. RESULT & ANALYSIS

Forest cover classification maps for the 1.5m resolution SPOT7 image were produced using MLC and SVMs classification techniques. Overall, there are 10 main classes of forest cover that have been mapped, namely built up, waterbody, cleared land, oil palm, mangrove, nipa, bushes, land preparation, cloud and shadow for each different method. Figure 6 shows the forest cover classification map of the study area with two algorithms, MLC and SVMs. However, the main focus of this study is to determine the ability of the classification methods chosen to extract the Nipa palm area.

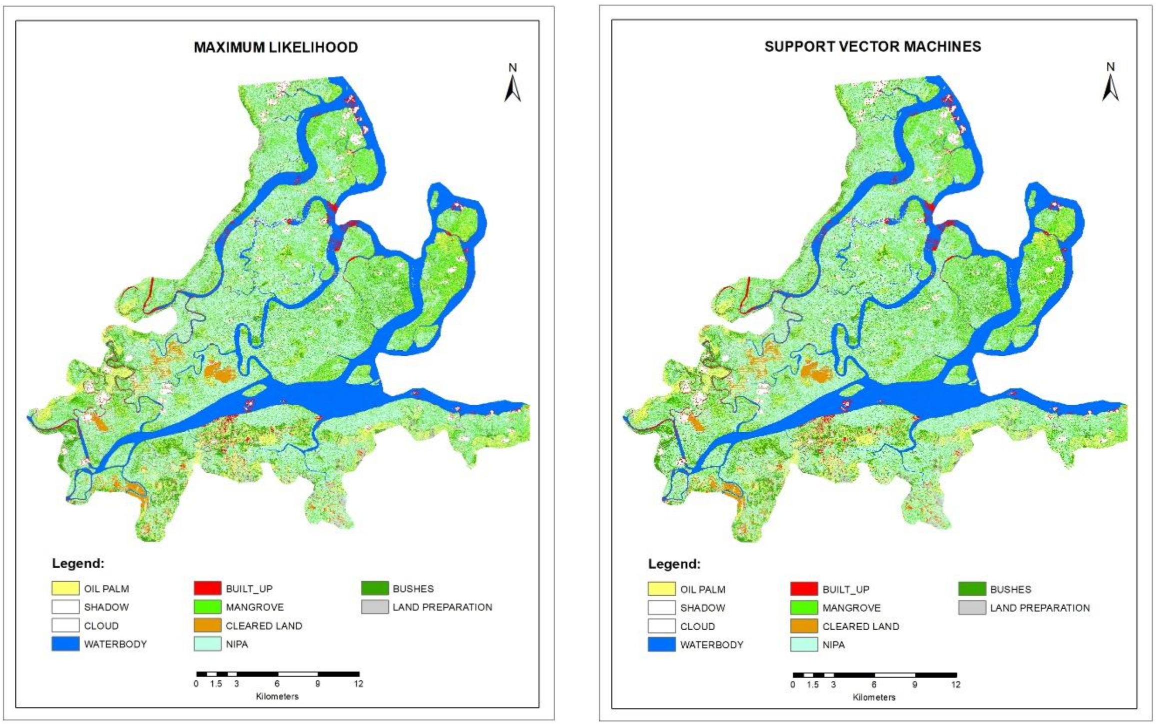

Figure 6: Forest cover maps generated using Maximum Likelihood and Support Vector Machine

A comparative analysis of MLC and SVMs classification reveals that SVMs classification gives a better result than MLC with an overall accuracy of 80.03% and kappa coefficient of 0.74, whereas with MLC it is 70.75% and 0.63 respectively. The SVM classifier identifies the classes more accurately compared to the ML classifier. (Table 1). The perfect User’s accuracy (100%) recorded from the cloud and waterbody class using SVMs. This research also discovered the highest producer’s accuracy and user’s accuracy for Nipa Palm class was in SVMs method, 85.28% and 89.36% respectively. The ML classifier only recorded user’s and producer’s accuracy of 79.29% and 74.44% for Nipa Palm class.

CLASS	MAXIMUM LIKELIHOOD		SUPPORT VECTOR MACHINES	
	Producer's Accuracy (%)	User's Accuracy (%)	Producer's Accuracy (%)	User's Accuracy (%)
NYPA	**79.29**	**74.44**	**85.28**	**89.36**
MANGROVE	57.33	78.18	75.21	85.44
WATERBODY	86.61	98.98	89.29	100.00
BUSHES	50.00	39.39	67.74	52.50
OIL PALM	60.38	61.54	59.18	69.05
CLEARED LAND	90.00	56.25	100.00	90.00

LAND PREPARATION	62.50	16.67	90.91	30.30
SHADOW	75.00	60.00	85.71	42.86
CLOUD	55.56	90.91	58.82	100.00
OVERALL ACCURACY	**70.75%**		**80.03%**	
KAPPA	**0.63**		**0.74**	

Table 1: Error matrix of classification result

The calculation of the area based on the number of pixels represented by each class is important in determining the capacity of each classifier used for estimating the extent of each class. The results of the areal statistics show that the Nipa class recorded a wider area which is 20,051.71 Ha using the SVM method compared to 19,157.70 Ha using the MLC method. The complete areal statistics of each forest cover class using MLC and SVMs methods is shown in figure 7.

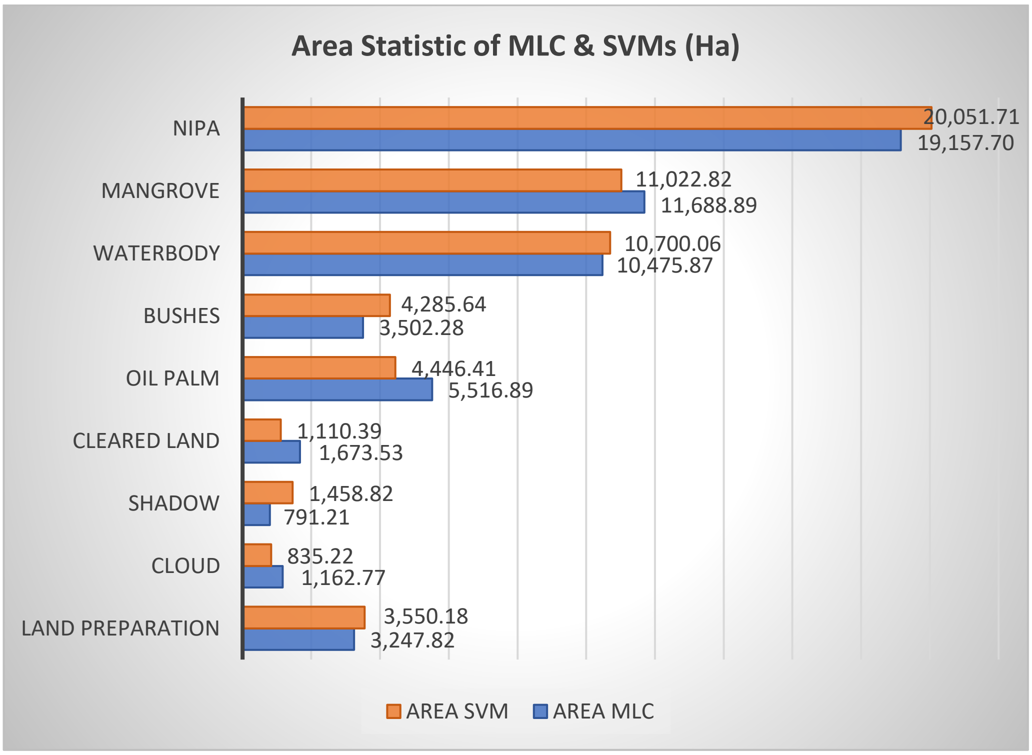

Figure 7: Area distribution of MLC and SVMs

5. CONCLUSION

In this study, MLC and SVM methods were used to determine the best method for mapping the nipa palm area in particular. This study indicates that the best classification results came from the SVM method under the linear kernel type, with the highest classification and user accuracy for Nipa Palm (89.36%). In terms of coverage area, the SVM method is able to map the Nipa palm area more realistically, which is 20,051.71 ha compared to the MLC method of only 19,157.70 ha. By using the SVM method, Nipa Palm area coverage can be mapped more realistically where the extent of the area can be estimated more accurately. This method proved to be potentially used to analyse nipa palm areas in other areas.

REFERENCES

Al-Ahmadi, F. and Hames, A. (2009). Comparison of four classification methods to extract land use and land cover from raw satellite images for some remote arid areas, kingdom of Saudi Arabia. Earth, 20, 167-191.

Anthony J. Alberg, Ji Wan Park, Brant W. Hager, Malcolm V. Brock, Marie Diener-West (2015). The Use of "Overall Accuracy" to Evaluate the Validity of Screening or Diagnostic Tests. Journal of General Internal Medicine, 19, 460-465.

Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37, 35−46.

Dhriti Rudrapal and Mansi Subhedar, 2015. Land Cover Classification using Support Vector Machine. *International Journal of Engineering Research and*, V4(09).

Foody, G.M. (2002). Status of land cover classification accuracy assessment. Remote Sensing of Environment, 80, 185−201.

Kavzoglu, T., Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geoinformation, 11, 352–359.

Kavzoglu, T., Reis, S. (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. GIScience and Remote Sensing, 45, 330–342.

Lu, D. and Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28, 823–870.

Srivastava, P.K., Kiran, G., Gupta, M., Sharma, N., Prasad, K. 2012. A study on distribution of heavy metal contamination in the vegetables using GIS and analytical technique. International Journal of Ecology & Development, 21: 89-99. (http://www.ceserp.com/cp-jour/index.php/ijed/article/view/3936).

Szuster, B.W., Chen, Q. and Borger, M. (2011). A comparison of classification techniques to support land cover and landuse analysis in tropical coastal zones. Applied Geography, 31, 525-532.

Tantus Piekkoontod, Bhumiphat Pachana, Karnjana Hrimpeng, Kitsanai Charoenjit. Assessments of Nipa Forest Using Landsat Imagery Enhanced with Unmanned Aerial Vehicle Photography. Applied Environmental Research, 42(3) (2020): 49-59.

Townshend, J.R.G. (1992). Land cover. International Journal of Remote Sensing, 13, 1319–1328.

Ukrainski, A. P., & Junior researcher at Regional federal centre of aerospace and ground monitoring of objects and natural resources at National Research University BelGU. (2019, February 26). *Classification accuracy assessment. confusion matrix method.* 50 North | GIS blog from Ukraine. Retrieved September 5, 2022, from http://www.50northspatial.org/classification-accuracy-assessment-confusion-matrix-method

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Yang, X. (2011). Parameterizing Support Vector Machines for Land Cover Classification. Photogrammetric Engineering & Remote Sensing, 77(1), 27–37. https://doi.org/10.14358/pers.77.1.27

Yu, L., Porwal, A., Holden, E.J. and Dentith, M.C. (accepted in 2011). Towards automatic lithological classification from remote sensing data using support vector machines. Computers & Geosciences, Article in Press.