Classification and change detection of Sabah mangrove forest using decision-tree learning technique

N A Kamaruddin¹ and F Shigeo²

¹Faculty of Bioresources & Food Industry, Universiti Sultan Zainal Abidin, 22200 Besut, Terengganu, Malaysia
²Graduate School of Global Environmental Studies, Kyoto University Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

nurulkamaruddin@unisza.edu.my

Abstract. The objective of this study is to determine the potential of decision tree-learning technique to classify and detects the changes of the Sabah mangrove forest area. The study area was conducted in the Mengkabong mangrove forest which is located on the west coast of Sabah. The multi-temporal of Landsat series (TM, ETM+, and OLI_TIRS) with five years interval data from 1990 and 2013 were used in this study. The results show that the use of decision-tree learning technique integrated with multi-temporal Landsat series and GIS data can be effective in delineating spatial and temporal change of the Sabah mangrove forest. The selection of suitable attributes from spectral features of Landsat data, topographic data and GIS database has promoted the high accuracy of the mangrove classification result with 90.8%. 40 hectares of Mengkabong mangrove were reduced from 1990 to 2013 and the fragmentation was obvious. In conclusion, the decision-tree learning technique was successfully classified and detects the changes of mangrove forest in the Mengkabong area.

1. Introduction

A recent study has demonstrated that decision-tree learning technique, one of the most popular machine learning approaches may be accurate and efficient for land cover classification based on remotely sensed data [1]. The decision-tree learning technique is capable of producing classification rules directly from the training data without human intervention. Besides, the method does not depend on assumptions of value distribution or the independence of variables [2].

The decision-tree learning technique is applicable incorporating with ancillary GIS data because these data usually have various value distributions and may be highly correlated [3]. The rule sets acquired from the decision-tree learning technique can be applied to the classification of multi-temporal satellite images, and the results can be compared due to the identical rule set used. The decision-tree learning technique can give more advantage rather than using traditional methods for monitoring the changes of mangrove forest from time series of remote sensing data [1, 4, 5].

Mangroves are vital components of the coastal ecosystems worldwide, but they are under threat from expanding human settlement, an explosion of commercial aquaculture, the exploitation of wood fuel and the impact of climate change. Such threats led to increasing demand for detailed mangrove maps to measure the extent of deterioration of the mangrove ecosystem. This phenomenon also treated to the Sabah’s mangrove. In Malaysia, Sabah is the largest of mangrove forest distributions with
320,000 ha in 2010 [6]. However, it is difficult to produce a detailed mangrove map mainly because mangrove forest is very difficult to access. Thus, remote sensing technology provides a genuine alternative to the traditional field-based method of mangrove mapping and monitoring. Due to its many advantages, such as being cost-effective, time-saving, and providing access to long-term data, the application of remote sensing technology to mangrove studies is already well established. However, some the advanced remote sensing applications with the capability of using low-cost satellite data remain unexplored for mangrove mapping, change detection, and monitoring. Furthermore, the use of satellite technology for studying Sabah’s mangrove regions remains poorly developed.

Therefore, the objective of this study is to determine the capability of the decision tree-learning technique to classify and detects the changes of mangrove forest land cover in the Sabah area integrating of multi-temporal Landsat data series (TM, ETM+, and OLI_TIRS). The rule set derived from the satellite data of 2013 to the date of 1990, 1995, 2000 and 2005 to detect changes of the mangrove forests over the period. We expected that the decision-tree method would be able to improve the performance mangrove forest classification with multi-temporal Landsat data and GIS ancillary data.

2. Method

2.1. Study area

The Mengkabong mangrove forest is located on the west coast of Sabah, Malaysia. This mangrove forest area is dominated by Rhizophora apiculata species which is healthy and dense mangrove. Conversions in aquaculture area and land reclamation have been a major factor for degradation of Mengkabong mangrove forest area.

2.2. Materials

The Landsat satellite series (TM, ETM+, and OLI_TIRS) with five years interval data from 1990, 1995, 2000, 2005, 2010 and 2013 were downloaded freely from US Geological Survey (USGS) website (http://earthexplorer.usgs.gov/). The reference data were included in field data, topographic data and vegetation map.

2.3. Image analysis

All the Landsat data series were analyzed for pre-processing analysis using Environment for Visualizing Images (ENVI) 5.1 software and MS-Excel-2010. The pre-processing analyses were covered radiometric correction, creating a multispectral image, sub-setting image, gap-filling analysis, cloud masking and histogram analysis which were to produce a good quality of images.

2.4. Decision-tree learning classification technique

2.4.1. Features extraction and determining attributes

All the Landsat satellite series data were processed for feature extraction using the tasseled cap transformation (TCT) index [7-8], band ratio, the spectral reflectance of bands 4, 5 and 7, and Normalize Vegetation Index (NDVI). The TCT index was extracted the greenness, brightness, and moisture features from the related six bands (1–5, 7) and (2–7) of TM, ETM+, and OLI_TIRS, respectively. The band ratios 5/4 and 3/5 [9] were used to discriminate between mangrove and non-mangrove area. The NDVI was calculated from the following equation:

\[ NDVI = \frac{NIR - RED}{NIR + RED} \]  

Where Red and NIR refer to the spectral reflectance measurement acquired in the red and near-infrared wavelength, respectively.

2.4.2 Generating decision-tree learning classification

The decision-tree learning classification was analyzed using the ENVI 5.1 software. The attribute rules are presented in Table 1 which were
comprised spectral reflectance attributes values of Landsat bands (4, 5, 7, greenness, moisture, and NDVI), elevation attributes (DEM), and distance attributes (distance to the coastline).

Table 1. Attributes used for the decision-tree learning technique

| Attributes                  | Acquisition method                        | Value range          |
|-----------------------------|-------------------------------------------|----------------------|
| Spectral reflectance        | NDVI and band ratio                       | -1 to 1              |
| Greenness and moisture      | TCT index                                 | -1 to 1              |
| Digital elevation model     | DEM Image in ARC/INFO Grid                | 0-500 integer        |
| Distance to coastline       | Eucuccu Distance of ARC/INFO Grid         | 0-20km               |

2.5. M-statistic

The M-statistic was validated to identify the accuracy of the features selection. It quantitatively assesses the separability of two classes regarding mean distance and standard deviation [10]. The formula for normalizing mean distance follows:

\[
M = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}
\]

(2)

Where \(M\) is the normalized mean distance and \(\mu_1\) and \(\mu_2\) are the means for classification feature of two samples with two different types of objects. \(\sigma_1\) and \(\sigma_2\) refer to the standard deviations of the classification feature of two samples with different types of objects. A greater separability is preferable. Figure 1 shows a flowchart of the overall decision-tree learning classification using the multi-temporal Landsat series (1990–2013) data.

Figure 1. Decision-tree learning technique classification protocol
2.6. Classification assessment
The confusion matrix method, integrated with ground-truth data from the field survey, topographic map, and Google Earth, were used for the multi-temporal data accuracy analysis.

3. Results

3.1. Decision-tree learning classification
A decision-tree learning classification technique was classified seven land cover classes of Mengkabong area which are covered mangrove, built-up, water-vegetation, secondary forest, grassland, bare soil and open water. The result of $M$-statistic values in separating among the land cover classes are shown in Table 2.

Table 2. $M$-statistic values between mangrove area and non-mangrove areas

| Data          | Classification features | Terrestrial vegetation area | Mangrove area | Water-vegetation-mixed pixel | Built-up area | Water       |
|---------------|-------------------------|----------------------------|---------------|------------------------------|---------------|-------------|
| 1990TM        | Greenness               | 0.83                       | 0.93          | 1.93                         | 2.41          | 1.72        |
|               | Moisture                | 2.59                       | 3.26          | 0.47                         | 1.26          | 0.18        |
|               | NDVI                    | 1.56                       | 1.52          | 1.28                         | 2.90          | 4.39        |
|               | Bands 3/5               | 2.27                       | 2.32          | 0.67                         | 0.03          | 1.56        |
|               | Band 4                  | 2.22                       | 2.69          | 1.69                         | 0.26          | 0.05        |
| 1995TM        | Greenness               | 0.63                       | 0.98          | 1.95                         | 2.12          | 4.71        |
|               | Moisture                | 2.47                       | 3.22          | 0.13                         | 1.25          | 0.17        |
|               | NDVI                    | 1.48                       | 1.48          | 1.24                         | 2.98          | 4.37        |
|               | Bands 3/5               | 1.58                       | 1.67          | 0.93                         | 4.98          | 0.71        |
|               | Band 4                  | 2.18                       | 2.60          | 1.48                         | 0.23          | 0.04        |
| 2000TM        | Greenness               | 0.62                       | 0.86          | 1.98                         | 2.10          | 4.77        |
|               | Moisture                | 2.36                       | 3.18          | 0.41                         | 1.23          | 0.15        |
|               | NDVI                    | 1.38                       | 1.38          | 1.21                         | 2.97          | 4.35        |
|               | Bands 3/5               | 1.48                       | 1.58          | 0.95                         | 4.99          | 0.73        |
|               | Band 4                  | 2.12                       | 2.48          | 1.34                         | 0.21          | 0.05        |
| 2005ETM+      | Greenness               | 0.64                       | 0.78          | 2.18                         | 2.08          | 4.79        |
|               | Moisture                | 2.28                       | 3.13          | 0.42                         | 1.21          | 0.13        |
|               | NDVI                    | 1.36                       | 1.35          | 1.23                         | 2.95          | 4.35        |
|               | Bands 3/5               | 1.46                       | 1.48          | 0.97                         | 5.01          | 0.69        |
|               | Band 4                  | 2.1                        | 2.36          | 1.26                         | 0.18          | 0.03        |
| 2010ETM+      | Greenness               | 0.72                       | 0.75          | 2.21                         | 2.11          | 4.81        |
|               | Moisture                | 2.73                       | 3.18          | 0.43                         | 1.23          | 0.11        |
|               | NDVI                    | 1.37                       | 1.33          | 1.21                         | 2.97          | 4.31        |
|               | Bands 3/5               | 1.35                       | 1.35          | 0.95                         | 5.05          | 0.71        |
|               | Band 4                  | 1.98                       | 2.28          | 1.2                          | 0.16          | 0.02        |
| 2013OLI_TIRS | Greenness               | 0.71                       | 0.78          | 2.23                         | 2.13          | 4.97        |
|               | Moisture                | 2.75                       | 3.22          | 0.41                         | 1.21          | 0.13        |
|               | NDVI                    | 1.35                       | 1.31          | 1.2                          | 2.93          | 4.42        |
|               | Bands 3/5               | 1.25                       | 1.28          | 0.93                         | 5.07          | 0.69        |
|               | Band 4                  | 1.95                       | 2.54          | 1.18                         | 0.14          | 0.03        |
Based on the result, only the $M$-statistic values of moisture, greenness and Band 4 were large. Thus the separability between terrestrial vegetation, mangrove, and water–vegetation mixed pixels were better. The mangroves class in all the data recorded the largest $M$-statistic value of moisture, with a highest value of 3.26 in 1990 and 3.22 in 2013. Mangrove forests are known to be hydrophytic, a term that the plants grow in water or very wet soil. Therefore, the background of the environment reflects the moisture of the mangrove canopy [11]. Moisture is an indicator of water bodies and moisture content in the vegetation canopy [12]. The mangrove vegetation shows the largest $M$-statistic value of Band 4 in all the data (1990-2013). The highest value was recorded in the years 1990, 1995 and 2013 and its show of high of mangrove distributions in that year.

The result of spectral profile classification in Figure 2 produced significant differences in spectral profile characteristics of each land cover class. Mangrove has high reflectance value in the Band 4 due to the high of moisture content in the mangrove leaves [13]. Thus, the Band 4 is useful to discriminate mangrove with others vegetation plants. The attributes of greenness and the moisture of TCT indexes were used to recognize of mangrove classes, and if moisture between -0.04 and 0 and greenness was greater than 0.6 at the same time, the pixel belongs to mangrove. Both of these TCT indexes are very useful for mangrove identification [14, 15].

![Figure 2](image1.png)

**Figure 2.** Spectral characteristics of mangrove and non-mangrove land cover types using Landsat data series (TM, ETM+, and OLI_TIRS)

![Figure 3](image2.png)

**Figure 3.** Spectral characteristics of mangrove and non-mangrove land cover types using NDVI value
The result of NDVI attribute analysis presented in Figure 3 was used to differentiate the different classes of vegetation and non-vegetation classes. The negatives of NDVI values were represented the water, built-up and bare soil areas. The mangrove and secondary forest in this study shows a high of NDVI values which are 0.08 and 0.06, respectively. The previous research suggested that the standard of high NDVI values from 0.4 to 1 corresponded to the density of vegetated area [16]. This can be supported that the Mengkabong mangrove forest was dominated by Rhizophora apiculata species, which is a healthy and dense mangrove. Thus, the thresholds of NDVI values used in this study were used to identify the different vegetation mangrove classification and promoted on high accuracies classes.

A total of 519 validation samples of mangrove and 4443 validation samples of non-mangrove which included water, secondary forest, water-vegetation, bare soil and built-up area were used for this analysis. The commission of error in mangrove identification was very small (7.5% in 0.90 of kappa coefficients) resulted from adequately considering the particular growth environment of mangrove and extracted moisture and greenness feature. The decision-tree classification in this study was demonstrated that selection of good attributes for mangrove classification was promoted on high accuracies of the result.

3.2. Results of change detection
The details of the detected changes of mangrove forest land cover in the study area are presented in Table 3.

| Classification features | 1990(ha) | 1995(ha) | 2000(ha) | 2005(ha) | 2010(ha) | 2013(ha) |
|-------------------------|---------|---------|---------|---------|---------|---------|
| Open water              | 1531.26 | 1555.20 | 1593.81 | 1616.67 | 1549.80 | 1536.84 |
| Mangroves               | 1145.16 | 1024.20 | 945.72  | 872.82  | 997.20  | 1185.12 |
| Secondary forest        | 846.90  | 826.20  | 780.12  | 627.21  | 518.40  | 499.14  |
| Built-up                | 921.42  | 1037.34 | 1142.19 | 1186.83 | 1309.05 | 1048.50 |
| Bare soil               | 342.00  | 253.17  | 198.72  | 266.85  | 246.33  | 270.54  |
| Grassland               | 132.75  | 238.77  | 277.38  | 371.52  | 324.18  | 407.16  |
| Water-Veg.              | 62.19   | 46.80   | 43.74   | 39.78   | 36.72   | 34.38   |

The details of the detected changes of mangrove forest and other types of land cover in the study area are presented in Table 4.6. The total area of Mengkabong is 4981.68 ha. According to the results of this study, mangrove forests were distributed extensively in the Mengkabong area in 1990, when they had an area of 1145.16 ha (almost 25% of the total area). However, there was a significant decrease from 1990 to 2000 (1145.16 ha to 945.72 ha), and a slight increase from 2010 to 2013 (997.20 ha to 1185.12 ha). The total area of mangrove forest declined (a reduction of 40 ha from 1990 to 2013) and fragmentation was obvious. The field investigation showed that most of the mangrove forest in the Mengkabong area had been lost. Most of the mangrove areas have been converted to shrimp ponds and housing settlements. According to Department of Fisheries Sabah, shrimp pond activity in this area has occurred since 2000 [17].
4. Conclusion
In summary, this study successfully classified mangrove and non-mangrove areas in the Mengkabong area by integrating the decision-tree learning technique with multi-temporal Landsat series and GIS ancillary data. The selection of suitable attributes, from the spectral features of Landsat data, and topographic data (DEM and distance to the coastline) from the GIS database, for the mangrove classification promoted the high accuracy of the result of 90.8%.

References
[1] Li P, Wu X, Hu X, Liang Q and Gao Y 2010 *Applied Artificial Intelligence* **24** 680-710
[2] Quinlan J R 1993 *C4.5 Programs for Machine Learning* (San Mateo CA: Morgan Kaufmann)
[3] Jensen J R 1995 *Introductory Digital Image Processing. A Remote Sensing Perspective* 2 Ed. (Englewood Cliffs: Prentice-Hall)
[4] Colstoun E C B, Story M H, Thompson C, Commissio K, Smith T G and Irons J R 2003. *Remote Sensing of Environment* **85** 16-27
[5] Horsness P W, Pebesma V E J and Schot P P 2002 *Ecological Modelling* **154** 93-101
[6] Sabah Forestry Department 2012 *Mangrove Forest Management & Restoration* p 223-226
[7] Zhan X, Sohlberg R A, Townshend J R G, DiMiceli C, Carroll M L and Eastman J C 2002 *Remote Sensing of Environment* **83** 336-350.
[8] Ali Baig M H, Zhang L, Shuai T and Tong Q 2014 *Remote Sensing Letters* **5** 423-31
[9] Green E P, Clark C D, Mumby P J, Edwards A J and Ellis A C 1998 *International Journal of Remote Sensing* **19** 935-56
[10] Zhang X H 2011 *Fourth International Conference on Information and Computing* p130-2
[11] Ajithkumar T T, Thangaradjou T and Kannan L 2008 *Journal of Environmental Biology* **29** 785-88
[12] Crist E P, Laurin R and Cicone R C 1986 IGARSS ’86 Symposium, Zurich, Switzerland, 8-11 September 1986 ESA Publ. Division p 1465-70
[13] Rahman M M 2013 *International Journal of Geomatic and Geoscience* **4** 36-46
[14] Razali S M and Nuruddin A A 2012 *African Journal of Agricultural Research* **7** 1901-09
[15] Zhang C, Kovacs J M, Wachowiak M P and Flores-Verdugo F 2013 *Remote Sensing* **5** 891-908
[16] Green E P, Mumby P J, Edwards A J, Clark C D and Ellis A C 1997 *Aquatic Botany* **58** 11-19
[17] Sabah Department of Fisheries. 2012. Series of 23: 223-226. Fisheries Statistics. Annual Report, 1: 14-16.