Science on Natural Resources and Environment
Journal homepage: tapchikhtnmt.hunre.edu.vn

APPLICATION OF MACHINE LEARNING WITH OBJECT-BASED IMAGE ANALYSIS FOR LAND USE AND LAND COVER MAPPING IN DAK NONG UNESCO GLOBAL GEOPARK, VIETNAM

Nguyen Thi Thao Van[1], Pham Van Manh[1], Bui Thi Thuy Dao[2]
[1]VNU University of Science, Vietnam National University
[2]Hanoi University of Natural Resources and Environment, Vietnam

Received 11 October 2022; Accepted 28 November 2022

Abstract

Land use and land cover (LULC) information is a fundamental component of environmental research related to urban planning, green infrastructure sustainability and natural hazards assessment. In particular, remote sensing technology has demonstrated a powerful capacity for LULC modelling with a corresponding increase in sensor number and type. The integration of many algorithms into the object - based classification method to create different sets of machine learning algorithms has proven very effective for image feature extraction from satellite data. In this study, the Random Trees model was used in combination with object - based image analysis (OBIA) to map LULC in Dak Nong UNESCO Global Geopark, Vietnam, using Landsat imagery data for the period from 2005 to 2022. The accuracy results show an overall accuracy (OA) of 83.97 % (2005), 85.38 % (2015) and 86.75 % in 2022 while the results of the Kappa coefficient were 0.82 (2005), 0.83 (2015) and 0.84 (2022). Accordingly, it was concluded that the method proposed here is useful for LULC detection and can be applied to other areas with similar characteristics. The derived maps can also inform as a document to UNESCO and national - level decision making.

Keywords: Dak Nong UNESCO Global Geopark; Random Trees (RT); Land use and land cover (LULC) mapping; Machine learning; Object - Based Image Analysis (OBIA); Remote sensing.

Corresponding author. Email: bttdao@hunre.edu.vn

1. Introduction

Land use and land cover (LULC) are separate but related concepts. Land cover refers to the surface cover on the ground, such as vegetation, urban infrastructure, water, bare land,…while the term land use describes the human use of the land for economic and social purposes. The change in Land use and land cover results from complex interactions between humans and the natural environment [1]. Land use and land cover change over time can significantly impact ecosystem processes, biological cycles and biodiversity [2]. Monitoring and evaluation regularly are essential and necessary in planning and managing to ensure environmental sustainability.

Nowadays, remote sensing data are increasingly used for monitoring LULC, including changes in forest land, industrial cropland and construction land in many related studies [3]. Medium and high spatial resolution remote sensing image data always dominate and image processing techniques are also constantly developed to serve this purpose. It is necessary to mention the development of algorithms for image enhancement, image mixing and LULC classification applied to data with medium and high spatial resolution [4, 5] the increasing population drives the spatiotemporal land use and land cover change (LULCC).

With high spatial resolution remote sensing images, the object - based classification method has high accuracy and clear classification results and is better than traditional pixel - based classification methods. On aspects such as multi - scale image analysis, minimizing the appearance of small object changes and confusion and detecting changes in overlay/using objects better soil [6, 7]. Therefore, object - based classification is often used to extract information about spectral features, shape and spatial relationships of the object for classification.

Recently, there has been an increasing interest in machine learning algorithms used for classification. Integration of many algorithms in the object - based classification method to create different sets of machine learning algorithms. These algorithms have been proven to be very effective in classifying satellite images [8]. Based on performance comparisons in many studies, the Random Trees algorithm shows high accuracy compared to current supervised learning algorithms [9, 10]. This is an attribute classification algorithm developed by Leo Breiman at the University of California, Berkeley [11]. Random forest is applied for classification purposes and other tasks by building decision trees on different samples. Each decision tree itself will go from the top down according to the conditional nodes to get the predictions and based on the majority votes of predictions; it predicts the final output.

In this study, Landsat satellite image data and the Random Trees classification algorithm were used to classify and evaluate changes in land use and land cover in the Dak Nong UNESCO Global Geopark area.

2. Material and methods

2.1. Study area

Dak Nong Geopark covers 4,760 km^2 and stretches over 1/6 districts of Dak Nong province, namely the districts of Krong No, Cu Jut, Dak Mil, Dak Song, Dak Glong and Gia Nghia town. On July 7^{th}, 2020, Dak Nong Geopark was recognized by UNESCO as a Global Geopark, with 65 geological heritage sites, of which seven are world - class geological heritage sites. Dak Nong Global Geopark converges all typical values in terms of geological, archaeological, cultural and biodiversity.

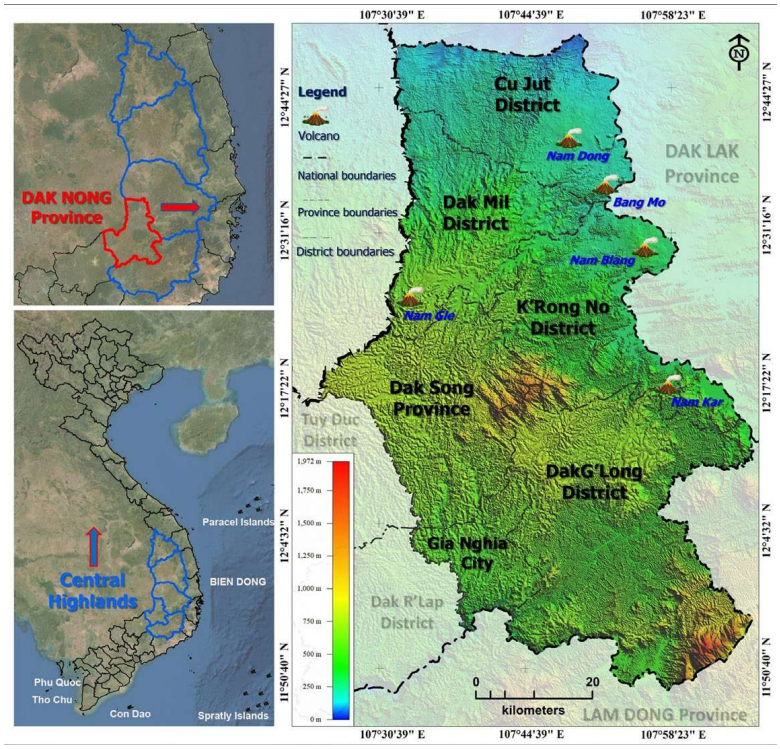

Figure 1: Location map of the study area

2.2. Data and classification method using Random Trees algorithm

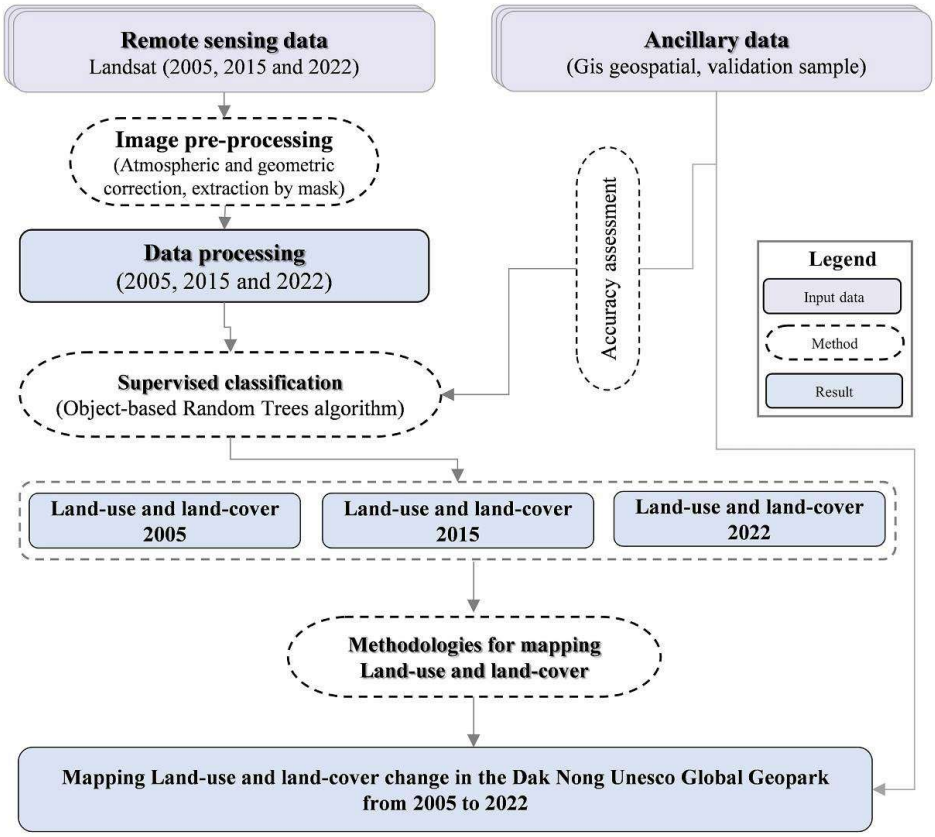

Figure 2: Flowchart of the methodology

Figure 2 illustrates the overall flowchart used in this study. Technically, this framework consists of three key components: (i) Preprocessing of satellite data; (ii) Land use and land cover (LULC) Classification into Residential area (RA), Natural Forest (NF), Plantation Forest (PF), Growing industrial crops (GIC), *Agricultural land (AGRL),* Water surface (WS) and other lands (OL); and (iii) Assessment of LULC change detection.

2.2.1. Data sources and pre - processing

The study used Landsat multi - temporal remote sensing data collected from the US Geological Survey website (https://earthexplorer.usgs.gov/). Selected data with less than 10 % cloud cover for the period from 2005 to 2022: Landsat-7 (2005), Landsat-8 (2015) and Landsat-9 (in 2022) (Table 1). Atmospheric/ radiation effects were removed through the COST (Cosine of the Solar Zenith Angle) method. These images were projected into the WGS-84/ UTM zone 48N projection and ensured DEM achieved a geometric correction accuracy of approximately ± 0.5 pixels. These data were finally spectrally equalized to the same time point to eliminate environmental differences in spectral irradiance values for variability assessment purposes.

Table 1. Sensor characteristics of multi - spectral temporal images ancillary data

Acquisition year	Type of data used	Spectral mode	Spectral resolution	Spatial resolution/ Scale
2005	Landsat-7	Multispectral	0.45 - 0.52 (µm) 0.52 - 0.60 (µm) 0.63 - 0.69 (µm) 0.76 - 0.90 (µm)	30 - m
		Panchromatic	0.52 - 0.90 (µm)	15 - m
2015	Landsat-8	Multispectral	0.45 - 0.51 (µm) 0.53 - 0.59 (µm) 0.64 - 0.67 (µm) 0.85 - 0.88 (µm)	30 - m
		Panchromatic	0.52 - 0.90 (µm)	15 - m
2022	Landsat-9	Multispectral	0.45 - 0.52 (µm) 0.52 - 0.60 (µm) 0.63 - 0.68 (µm) 0.84 - 0.88 (µm)	30 - m
		Panchromatic	0.50 - 0.68 (µm)	15-m
2021	GIS database	N/A	N/A	1:10,000

2.2.2. Object-based classification method using random trees algorithm

(1) Image segmentation: The process of segmenting objects is done based on customizing parameter values of shape, compactness and scale - the important factors that directly affect the size of each interpreted object. The data is merged into a single data set (including 2005, 2015 and 2022 images). After many tests, the image fragmentation results have selected the parameters Scale (20), Shape (0.85) and Compactness (0.65) to become the optimal parameters for minimizing confusion between objects on the image.

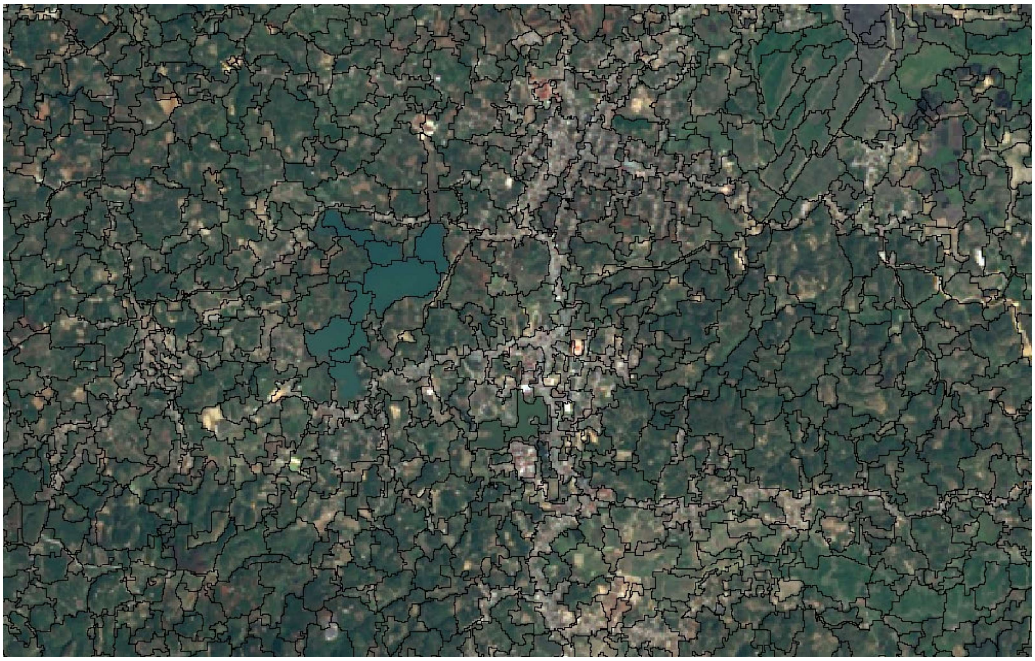

Figure 3: Experimental site selection of parameters during image segmentation

(2) Using Random Trees machine learning algorithm: Based on segmented data, conduct sampling and classification using Random Trees machine learning algorithm. Sample data were selected by the method of random sampling with replacement (Bootstrapping) and LULC was divided into seven classes (Table 2): Residential area (RA), Agricultural land (AGRL), Water surface (WS), Natural Forest (NF), Growing industrial crops (GIC), Plantation Forest (PL) and other lands (OLS).

Table 2. Types of LULC in Dak Nong UNESCO Global Geopark

LULC	Abbreviation	Describe
Residential area	RA	Includes all man-made coating surfaces built-up area, such as residential, buildings, commercial, industrial, or transportation infrastructure.
Natural Forest	NF	Forest areas are less affected by humans and natural disasters; The structure of the forest is still relatively stable.
Growing industrial crops	GIC	The land is used to grow perennial crops as raw materials for industrial production; Industrial crops such as (coffee, rubber, pepper and cashew,…)
Plantation Forest	PL	Forests are formed by new plantings of people human on unforested land without forests; Replanted forests clavation of natural forests; Replanting or regeneration after from forest exploited palatalisation.
Agricultural land	AGRL	Land used for arable purposes, including areas of Seasonal crops (especially rice and maize), vacant agricultural land during the tillage period with sparsely vegetated patches. agricultural land that is vacant during the tillage stage with patches, or sparse vegetation.

LULC	Abbreviation	Describe
Water surface	WS	Lakes, streams, canals and man-made bodies of water have different shapes (usually normally irregular) shapes. Ponds and water streams are also listed in this land use/cover category.
Other lands (empty land)	OLS	The area with vegetation has area with less than 1/3 of the total surface area, bare surface land, mostly barren arid land with a thin layer of soil, sandy or rocky soil, or other soil types.

The algorithm builds decision trees from the training sets (Bootstrap) and then divides recursively based on the best - selected attribute. The results from decision trees can be different because each decision tree is built of a randomly selected subset of the data.

$$f_n^M(x) = \frac{1}{M}\sum_{i=1}^{M} f_n^i(x) \quad (1)$$

Where $f_n^M(x)$ is the predicted value of decision tree *i* at query point *x*. *M* is the number of decision trees generated and n is the number of training samples. After aggregating the results from the decision tree (Decision tree), the algorithm will rely on majority voting to produce the final classification result.

$$f_n^M(x) = \begin{cases} 1, & \frac{1}{M}\sum_{i=1}^{M} f_n^i(x) > 1/2 \\ 0, & \end{cases} \quad (2)$$

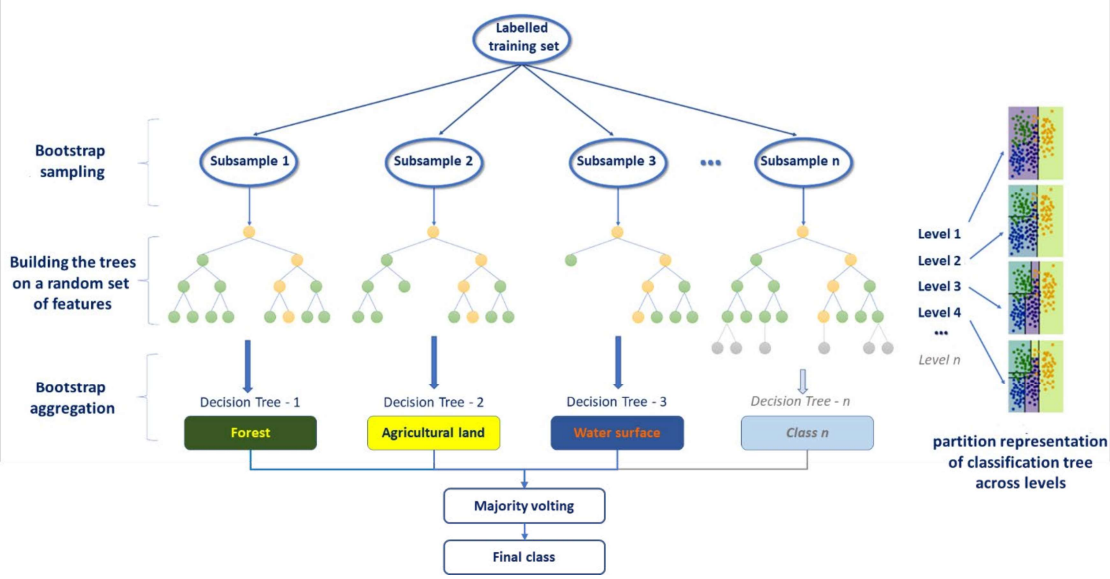

Figure 4: Simplified diagram of a random tree

(3) Manual editing and evaluate the accuracy of the classification results

The process of evaluating the classification results is carried out through the statistical process of the number of confusion points between the individual LULC. It is assumed that the subjects of the test samples were evenly distributed throughout the study area and tested using the overall accuracies (OA) and the Kappa coefficient (K). In addition, to increase the accuracy of the classification results, the discrete fragments are further screened and compared with the decoding key points on the image to correct some confusion between the classes.

3. Result and discussion

The LCLU classifications were conducted by using the object-based

Random Trees classifier algorithms (Fig.5). The proposed approach produced overall accuracies (OA) of 83.97 %, 85.38 % and 86.75% for LULC classifications of 2005, 2015 and 2022 images respectively while the Kappa coefficient was 0.82 (2005), 0.83 (2015) and 0.84 (2022). The overall accuracy (in percentage) for all the classes is shown in Table 3. In general, the overall classification accuracy of most classes is over 80%, especially WS with the highest accuracy, followed by NF, AGRL, GIC, NF and RA. The accuracy of OLS is lower than other land use classes because of confusion in the classification process.

Table 3. Accuracy of classification results of each class (Unit: %)

LULC / Years	Water surface	Natural forest	Plantation forest	Growing industrial crops	Residential area	Agricultural land	Other lands
2005	89,44	87,52	81,65	82,52	80,20	87,54	78,43
2015	91,76	88,83	82,49	85,78	81,50	88,68	78,62
2022	93,85	90,15	85,18	86,19	83,07	89,09	79,69
Medium	91,68	88,83	83,11	84,83	81,59	88,43	78,91

The process of assessing the variability of land use objects on the image is determined by comparing multi-temporal land use information. Based on using initial time data as a basis, the subsequent classification results are used to analyse the change of land use objects. In this way, the overall accuracy of the land use classification is achieved at a higher level than usual. The object-based classification approach focuses only on locations where the segmentation changes, improving the volatility assessment process in accuracy and efficiency; Especially in urban land areas (where there is a big change in land use).

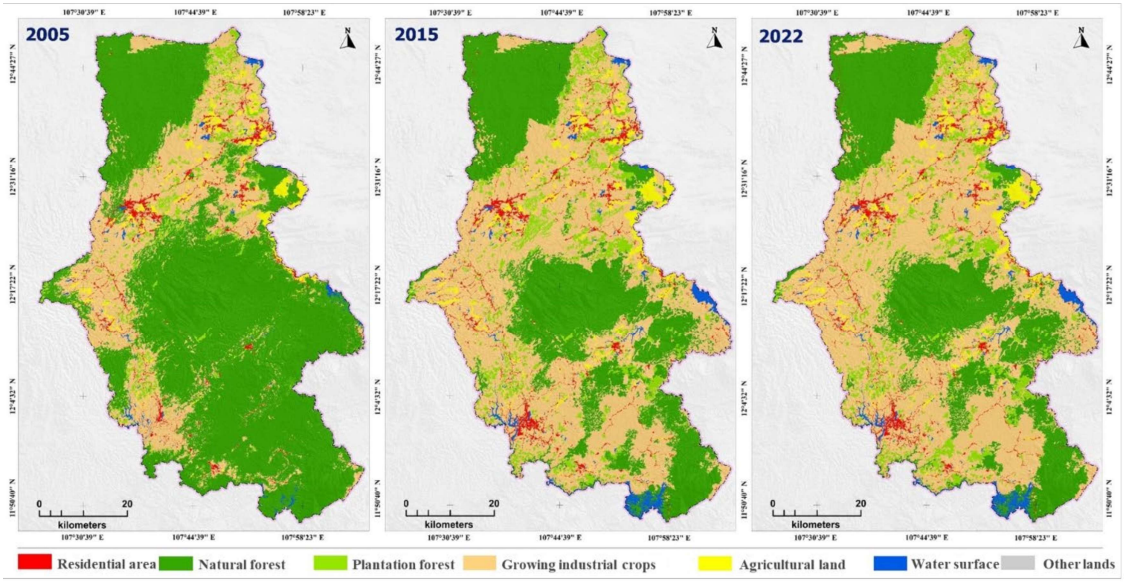

Figure 5: Land use/cover classification results for the period 2005 - 2022

The LULC classification results will be the input for the calculation of LULC volatility and will be analysed in the following three periods: (i) Period (2005 - 2015); (ii) Period (2015 - 2022); (iii) Phe entire period (2005 - 2022).

(i) Period (2005 - 2015): The area of growing industrial crops (GIC) increased by 108,468.25 ha, converted mainly from the area of natural forest (108,012.77 ha), the area of different types of land. The remainder is insignificant (ha). The conversion matrix between LULC types is shown in Table 4.

Table 4. Matrix of area conversion between types of LULC in the period of 2005 - 2015

Matrix converts (ha)		LULC 2015							
		AGRL	**GIC**	**NF**	**OLS**	**PF**	**RA**	**WS**	**Total**
LULC 2005	AGRL	11,714.8	147.19	0.41	1.65	0.92	4.69	0.96	11,790.61
	GIC	65.89	123,912.27	18.67	11.96	216.98	1,438.00	30.29	126,062.06
	NF	3,175.25	108,012.77	163,416.70	432.43	18,129.74	2,297.62	4,679.05	300,223.56
	OLS	25.67	195.86	1.27	401.47	105.35	83.78	11.77	825.17
	PF	3.22	87.96	3.16	2.01	20,826.39	4.92	0.52	20,928.18
	RA	3.09	6.29	2.60	1.32	2.94	7,889.06	0.61	7,915.90
	WS	0.48	8.19	1.63	1.29	0.42	0.40	8,186.47	8,198.89
	Total	**14,988.40**	**232,378.52**	**163,444.44**	**852.12**	**39,282.74**	**11,708.46**	**13,289.68**	**475,957.65**

(ii) Period (2015 - 2022): The area of growing industrial crops (GIC) continued to increase by 16,997.68 ha converted from the NF (11,801.48 ha), PF (4,159.83 ha) and 1036.37 ha from the remaining land area (GIC, AGRL, RA, WS and OLS). The area conversions between LULC types are shown in Table 5.

Table 5. Matrix of area conversion between types of LULC in the period 2015-2022

Matrix converts (ha)		LULC 2022							
		AGRL	**GIC**	**NF**	**OLS**	**PF**	**RA**	**WS**	**Total**
LULC 2015	AGRL	14,734.6	190.14	5.65	1.62	13.10	56.74	6.56	*14,988.43*
	GIC	308.84	230,563.47	24.40	10.74	654.44	1,359.14	28.85	*232,379.88*
	NF	74.13	11,801.48	149,264.82	3.27	3,192.06	218.60	39.62	*164,443.98*
	OLS	43.85	211.86	34.44	316.93	94.10	132.33	40.87	864.39
	PF	44.92	4,179.83	69.37	6.14	38,914.49	73.05	35.21	*39,283.02*
	RA	3.69	8.33	7.13	0.29	33.40	14,027.78	5.63	*14,708.25*
	WS	4.08	74.04	47.52	3.63	4.20	4.11	4,152.15	*4,289.72*
	Total	**15,191.15**	**246,521.15**	**149,886.34**	**342.62**	**38,775.79**	**15,871.74**	**9,368.88**	**475,957.65**

(iii) Period (2005 - 2022): In summary, the GIC have increased rapidly from 122,369.51 ha (2005) to 246,516.68 ha (in 2022) over this period. The table of the area conversion matrix of LULC types in the Dak Nong UNESCO Global Geopark (Table 6) shows that the increased RA was obtained from the NF (2,755.82 ha), GIC (2,291.83 ha), OLS (156.3 ha) transferred. Meanwhile, the area of natural forest (NF) has decreased sharply. This shows that the local forest allocation policy is not appropriate. In addition, illegal deforestation has resulted in a serious decrease in the natural forest area.

Table 6. Area conversion matrix between types of LULC in the period 2005 - 2022

Transformation Matrix (ha)		LULC 2022							
		AGRL	GIC	NF	OLS	PF	RA	WS	Total
LULC 2005	AGRL	11,542.5	127.60	1.92	1.12	9.65	51.04	6.53	*11,740.43*
	GIC	302.83	122,769.51	15.07	6.22	497.70	2,291.83	37.24	*126,060.41*
	NF	3,213.77	120,885.41	149,846.13	155.45	19,782.95	2,755.82	4,711.32	*301,220.86*
	OLS	48.58	298.77	5.64	163.32	101.82	156.30	48.11	*822.55*
	PF	72.75	2,461.25	5.10	0.36	18,355.27	152.97	10.43	*20,928.14*
	RA	16.60	8.11	6.34	0.24	25.64	10,461.97	3.72	*10,965.62*
	WS	3.78	43.04	4.06	2.71	2.43	1.69	4,139.59	*4,197.29*
	Total	**15,190.81**	**246,516.68**	**149,884.28**	**329.42**	**38,775.47**	**15,871.63**	**9,366.95**	**475,957.65**

4. Conclusion

The purpose of this study has been to LULC classification using object - based Random Trees classifier algorithms in Dak Nong UNESCO Global Geopark by Landsat images. The result of OA values obtained was over 80 %. Furthermore, the Kappa coefficient and index-based analysis also show variations in the accuracy of each LULC classifier. Regarding classification accuracy, the Random Trees algorithm performed better for water surface and forest land cover. This study can be useful to change the detection of LULC through time in Dak Nong UNESCO Global Geopark and similar regions for monitoring and management purposes.

REFERENCES

[1]. K. S. Sealey, P.-M. Binder and R. K. Burch (2018). *Financial credit drives urban land-use change in the United States.* Anthropocene, vol. 21, p. 42 - 51, Mar. 2018. Doi: 10.1016/j.ancene.2018.01.002.

[2]. M. J. Santos et al., (2021). *The role of land use and land cover change in climate change vulnerability assessments of biodiversity: A systematic review*. Landsc. Ecol., vol. 36, no. 12, p. 3367 - 3382, Dec. 2021. Doi: 10.1007/s10980-021-01276-w.

[3]. Md. J. Faruque et al., (2022). *Monitoring of land use and land cover changes by using remote sensing and GIS techniques at human-induced mangrove forests areas in Bangladesh*. Remote Sens. Appl. Soc. Environ., vol. 25, p. 100699, Jan. 2022. Doi: 10.1016/j.rsase.2022.100699.

[4]. E. A. Alshari and B. W. Gawali (2021). *Development of classification system for LULC using remote sensing and GIS.* Glob. Transit. Proc., vol. 2, no. 1, p. 8 - 17, Jun. 2021. Doi: 10.1016/j.gltp.2021.01.002.

[5]. A. Rahman et al., (2020). *Performance of different machine learning algorithms on satellite image classification in rural and urban setup*. Remote Sens. Appl. Soc. Environ., vol. 20, p. 100410, Nov. 2020. Doi: 10.1016/j.rsase.2020.100410.

[6]. V.-M. Pham, S. Van Nghiem, Q.-T. Bui, T. M. Pham and C. Van Pham (2019). *Quantitative assessment of urbanization and impacts in the complex of Hue Monuments, Vietnam.* Appl. Geogr., vol. 112, p. 102096, Nov. 2019. Doi: 10.1016/j.apgeog.2019.102096.

[7]. Q.-T. Bui, M. V. Pham, Q.-H. Nguyen, L. X. Nguyen and H. M. Pham (2019). *Whale optimization algorithm and adaptive neuro - fuzzy inference system: A hybrid method for feature selection and land pattern classification*. Int. J. Remote Sens., vol. 40, no. 13, p. 5078 - 5093, Jul. 2019. Doi: 10.1080/01431161.2019.1578000.

[8]. A. E. Maxwell, T. A. Warner and F. Fang (2018). *Implementation of machine - learning classification in remote sensing:*

An applied review. Int. J. Remote Sens., vol. 39, no. 9, p. 2784 - 2817, May 2018. Doi: 10.1080/01431161.2018.1433343.

[9]. P. Du, A. Samat, B. Waske, S. Liu and Z. Li (2015). *Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features*. ISPRS J. Photogramm. Remote Sens., vol. 105, p. 38 - 53, Jul. 2015. Doi: 10.1016/j.isprsjprs.2015.03.002.

[10]. B. E. Lefulebe, A. Van der Walt and S. Xulu (2022). *Fine - scale classification of urban land use and land cover with planet scope imagery and machine learning strategies in the city of Cape town, South Africa*. Sustainability, vol. 14, no. 15, p. 9139, Jul. 2022. Doi: 10.3390/su14159139.

[11]. D. C. Duro, S. E. Franklin and M. G. Dubé (2012). *Multi - scale object - based image analysis and feature selection of multi - sensor earth observation imagery using random forests*. Int. J. Remote Sens., vol. 33, no. 14, p. 4502 - 4526, Jul. 2012. Doi: 10.1080/01431161.2011.649864.