ENDAU ROMPIN NATIONAL PARK LAND USE LAND COVER CHANGES USING REMOTE SENSING APPROACH

Atif Abdullah[1] and Noordyana Hassan[1,2*]

[1]*Department of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.*
Author email: atif28300@gmail.com
[2]*Geoscience and Digital Earth Centre (INSTEG), Research Institute of Sustainable Environment (RISE), Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia*
**Corresponding author: noordyana@utm.my*

How to cite: A. Abdullah and N. Hassan, "Endau Rompin national park land use land cover changes using remote sensing approach," *Kurvatek*, vol. 8, no. 2, pp. 221-230, 2023. doi: 10.33579/krvtk.v8i2.4662 [Online].

Abstract - Identification of land use and land cover in forest areas can be challenging due to various land cover types within a forest can be similar, making it hard to differentiate between them using remote sensing approach. We hypothesized that random forest classification (RF) would outperform maximum likelihood (ML) in the classification of land use and land cover (LULC) in forest areas compared to maximum likelihood (ML). To verify this hypothesis, we conducted a comparative analysis, assessing the accuracy of RF and ML in the classification of LULC within the Endau-Rompin National Park (ERNP) region, utilizing Landsat 8 imagery. An accuracy assessment demonstrated that the RF classifier (overall accuracy: 92% (2013) 91% (2016) 79% (2022) with kappa coefficient: 0.843 (2013), 0.817 (2016) and 0.674 (2022), performed better than ML classifying land cover. Our results suggest that both methods are able to classify land cover of forest area, but RF is more accurate than ML. From the classification result of RF classification, we calculate the land cover changes of ERNP from 2013 to 2022. Results showed that there are small changes of forest area were found in ERNP. The total forest area decreases from 163250.089 ha to 144765.46 ha during 2013 to 2022. This finding suggests that the effectiveness of the protected area in mitigating deforestation in its surrounding regions may be somewhat limited, as indicated by the observed minor changes.

Keywords: land use and land cover change, protected area, remote sensing, maximum likelihood, random forest.

I. INTRODUCTION

In recent years, human activities have significantly impacted forest areas, resulting in a decline in global biodiversity. This decline can be attributed to extensive forest degradation and changes in (LULC) [1]. The alteration of LULC has been identified as a major contributor to global biodiversity change over the past few decades [2]. Within the field of remote sensing (RS), image classification plays a crucial role, as selecting the most suitable method to generate land cover information is a fundamental step in any study. This technique is widely employed to investigate changes in land use and land cover. Previous studies have utilized various methods to extract information about land use and land cover changes from satellite imagery, including supervised and unsupervised classification approaches [3]. Supervised classification, which encompasses techniques such as minimum distance, parallelepiped, and maximum likelihood classification, has been commonly used. Additionally, non-parametric methods like fuzzy classification, nearest- neighbor classification, and machine learning techniques such as random forest, support vector machines, and neural networks have been employed [4]. From previous studies, the Maximum Likelihood classification method is a well-known statistical decision criterion used for the analysis of satellite images. This method has been mostly applied for land cover classification and monitoring of land use changes, showing overall high accuracies. However, when applied to forest areas, this method may have difficulty distinguishing between different forest types due to the complex and diverse vegetation cover. Additionally, overlapping signatures, sensitivity to training data, and vulnerability to the distribution of categories in feature space are other weaknesses of this method in forest areas. Maximum likelihood classification may have difficulty distinguishing the pixels that come from different land cover classes but have very similar spectral properties [5][3]. This can lead to errors in classification, especially in areas where different land cover types overlap, such as forest edges.

Received November 4, 2023; Revised November 5, 2023; Accepted November 6, 2023

The random forest algorithm has gained significant popularity and widespread adoption in the field of land use/land cover classification, with a particular emphasis on applications related to forests [6][7]. Its utilization in this domain has been expanding rapidly, driven by its ability to handle the complex and heterogeneous nature of forest environments, and its capacity to provide accurate and reliable classification results [8]. Random Forest classification is a supervised machine learning algorithm that is constructed from decision tree algorithms [9]. The decision tree is a flowchart-like structure where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or a prediction. In the context of Random Forest classification, multiple decision trees are created during the training phase. Each decision tree is constructed using a subset of the training data and a random subset of features. This randomness injects diversity among the individual trees and helps to reduce overfitting [10][11].

In this study, the ability of RF and ML classifiers were explored in identification land use land cover of Endau-Rompin National Park (ERNP). The accuracy of these two classifiers on Landsat 8 imagery were compared and choose the best classifier to estimate the land use land cover changes over ERNP and its surrounding areas.

II. MATERIALS AND METHODS

A. Study Site

Taman Negara Endau Rompin, alternatively referred to as Endau-Rompin State Park, is a vast national park situated in the states of Pahang and Johor in Peninsular Malaysia. Encompassing a total area of approximately 87,000 hectares, it stands as one of the largest protected regions within the country. The park is divided into two sections, with the northern part designated as Endau-Rompin State Park and overseen by the Pahang state government, while the southern part falls under the management of the Johor National Parks Corporation (JNPC). Access to the park is facilitated through two entry points on the Johor side and one on the Pahang side. Renowned for its rich and distinctive ecosystem, Endau-Rompin National Park features diverse habitats, including lowland rainforests, freshwater wetlands, and peat swamp forests, supporting a wide array of wildlife species. Notably, the park holds great significance for conservation efforts and scientific exploration, being home to numerous threatened and endangered species. Moreover, it has been recognized as a UNESCO Biosphere Reserve [1].

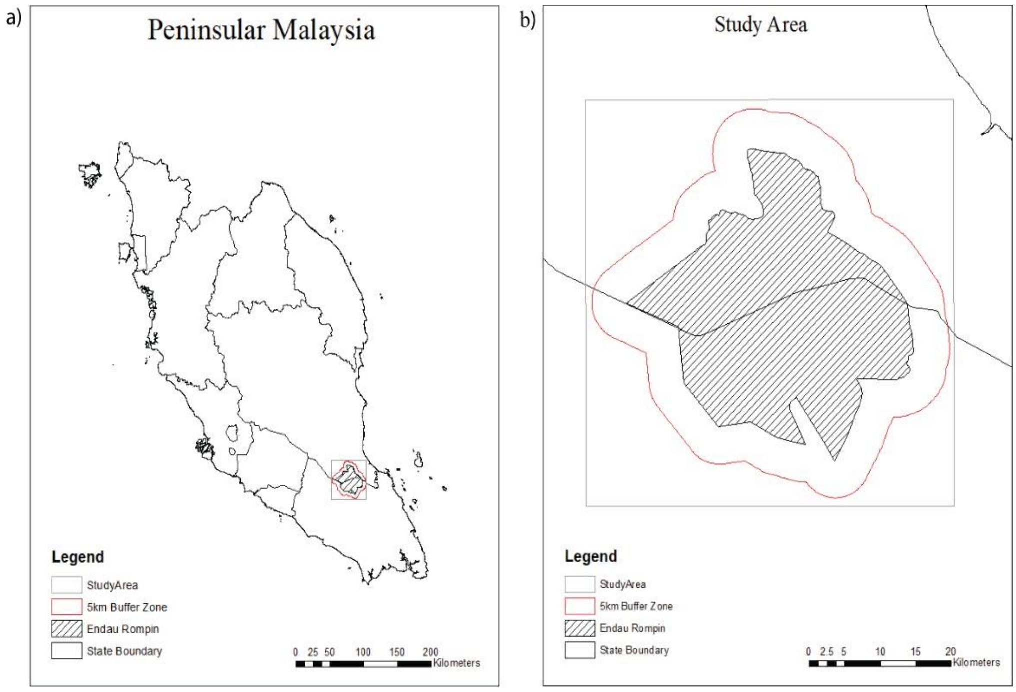

Figure 1. Maps of (a)location of ERNP in Peninsular Malaysia and (b) confined study area: Endau-Rompin National Park (ERNP)

B. Materials

1. Remotely sensed data

Multispectral Landsat 8 images were acquired for every three years period of time as the forest changes can be seen within that time frame (2013, 2016, and 2022) through the USGS website (earthexplorer.usgs.gov), ensuring that the images had cloud cover below 30% to maintain their quality. After obtaining the Landsat 8 images, additional preprocessing steps were conducted to enhance the quality and reliability of the images before performing the land use land cover classification. Following the acquisition of Landsat 8 images, additional preprocessing steps were undertaken to enhance the quality and reliability of the images before conducting the land use land cover classification. The preprocessing phase includes cloud masking, which involved the utilization of the QA_PIXEL Band file. The QA_PIXEL Band file contains quality statistics and cloud mask information gathered from the image data (USGS, n.d.). By analyzing this information, the cloud-affected pixels were accurately identified and removed from the images. The cloud masking process using the QA_PIXEL Band file was performed by applying thresholding techniques.

2. Supplementary data

One of the ancillary data used for the study area is the land use map of Johor State. This supplementary data is utilized to enhance and support the analysis of primary data. To assess the accuracy of our land use land cover classification, land use maps were used and were obtained from the Department of Agriculture for the years 2013 and 2016. These land use maps served as reference data, providing a reliable representation of the actual land use categories during those specific years.

3. Ground data collection

In addition, we collected XY-coordinates using a handheld GPS unit (Garmin GPSmap 60CSx) on May 24, 2023. We utilized these coordinates as reference data to evaluate the classification accuracy of the most recent dataset, which is 2022.

C. METHODOLOGY

1. Supervised classification

The supervised classification was employed by using the Maximum Likelihood and Random Forest algorithm on the Landsat images. The classification adopted were supervised classification method, using the maximum likelihood algorithm and random forest to classify the Landsat images (2013, 2016, 2019, and 2022) with different training sites selected. These methods extract information from a variety or multiple raster bands to complete the probabilities of group membership for each cell pixel in satellite images. Besides, it uses the sample areas of the same multispectral space to identify the object classes' characteristics. The same characteristics of the object colour will be assigned as one class.

2. Classifications' accuracy assessment

Confusion matrix was used to evaluate the produced classification output using a land use map and ground truth points as reference information. An accuracy assessment of the RF and ML classifications was conducted by comparing them to a land use map for the years 2013 and 2016, as well as ground-truth data for the year 2022. The assessment was based on reference data that accurately represent the true land cover conditions. To evaluate the errors of each classifier, confusion matrix approach was employed [12]. Both the user's accuracy and producer's accuracy were calculated for each classification method. The overall accuracy and kappa coefficient were used to determine the percentage of correctly classified classes and the level of agreement [13], respectively, as shown in the table. Accuracy assessment is a critical step in verifying the accuracy of classification results and identifying potential errors due to similarities in spectral responses among different classes [14]. The confusion matrix, which is a square table, was employed to analyze the accuracy of LULC images at various dates. This matrix captures the pixels of the image, with rows representing the classified categories and columns representing the ground truth. The off- diagonal cells indicate misclassified values or instances classified into incorrect categories [15].

Table 1. Training pixels and test pixels for each class for random forest classifications (RF) and Maximum Likelihood (ML)

Year	**2013**		**2016**		**2022**	
Classification	Training pixel	Test pixel	Training pixel	Test pixel	Training Pixel	Test Pixel
Forest	100	34	100	37	18	10
Agriculture	40	16	30	18	5	5
Built up / Bare Land	10	6	6	6	5	5
Water Body	5	5	5	5	5	5

3. Change Detection Analysis

From the RF classification results, a change detection analysis was performed by implementing post-change detection techniques. This analysis aimed to identify and assess the changes that occurred in the land cover over time. By comparing the classified maps for different years, areas of significant land cover change were identified and analysed. The post- change detection techniques involved comparing the pixel values and their corresponding classes between the different time periods. This analysis provided valuable insights into the dynamics and trends of land cover changes within the study area. It allowed for the identification of areas that underwent substantial transformations, such as deforestation, built up/bare land, or agricultural expansion.

III. RESULTS

A. Assessment on capability of random forest and maximum likelihood classifiers in identification of land use land cover of Endau-Rompin National Park

Based on the result of RF and ML classifiers in identification land use land cover of ERNP and its surrounding areas, LULC classification maps of ERNP were produced with forest areas for the years 2013, 2016, and 2022 by the RF algorithm (Figure 2) and ML algorithm (Figure 3). Overall accuracy and kappa statistics of both classifiers showing that RF performed better than ML (Table 2). Therefore, the RF classifier will be used in identification of Endau-Rompin National Park LULC changes.

Table 2. Accuracy assessment on both classifiers; random forest and maximum likelihood on Landsat 8

Classification	**RF**		**ML**	
	Overall accuracy (%)	**Kappa**	**Overall accuracy (%)**	**Kappa**
2013	92	0.843	86	0.752
2016	91	0.817	81	0.665
2022	79	0.674	66	0.443

Throughout the analyzed Landsat imagery (Table 3), forested areas consistently maintained their prominence as the predominant land cover, with agricultural land, built-up or cleared areas, and water bodies. Nevertheless, the LULC classification maps offer valuable observations regarding the evolving trends in land use and land cover within ERNP and its environs over the specified years. Table 3 showing the land cover in hectare over ERNP from 2016 until 2022.

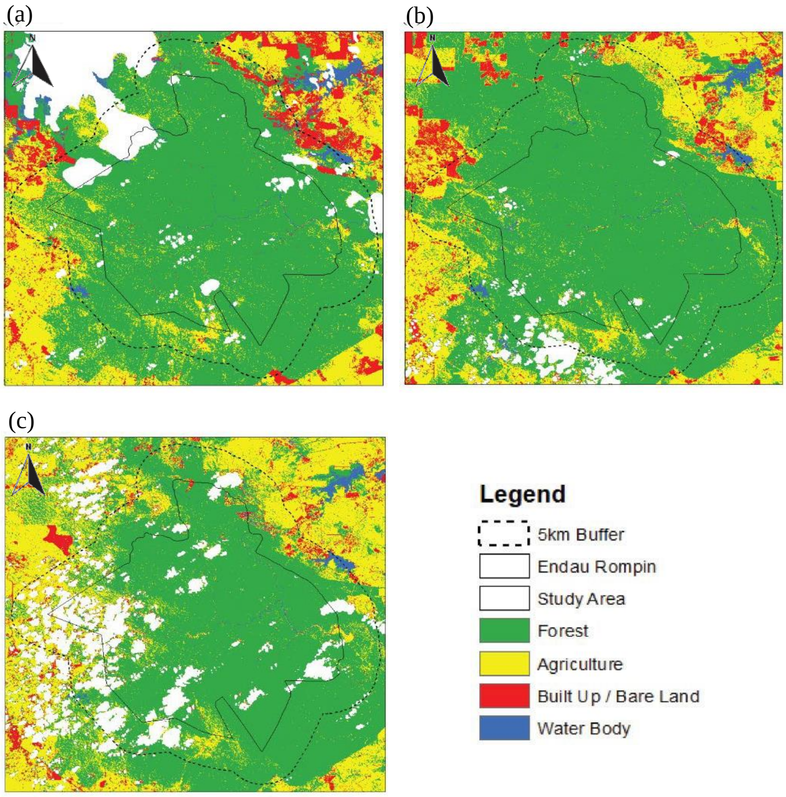

Figure 2. LULC maps for the years (a) 2013, (b) 2016, and (c) 2022 in the studied area by using RF classification.

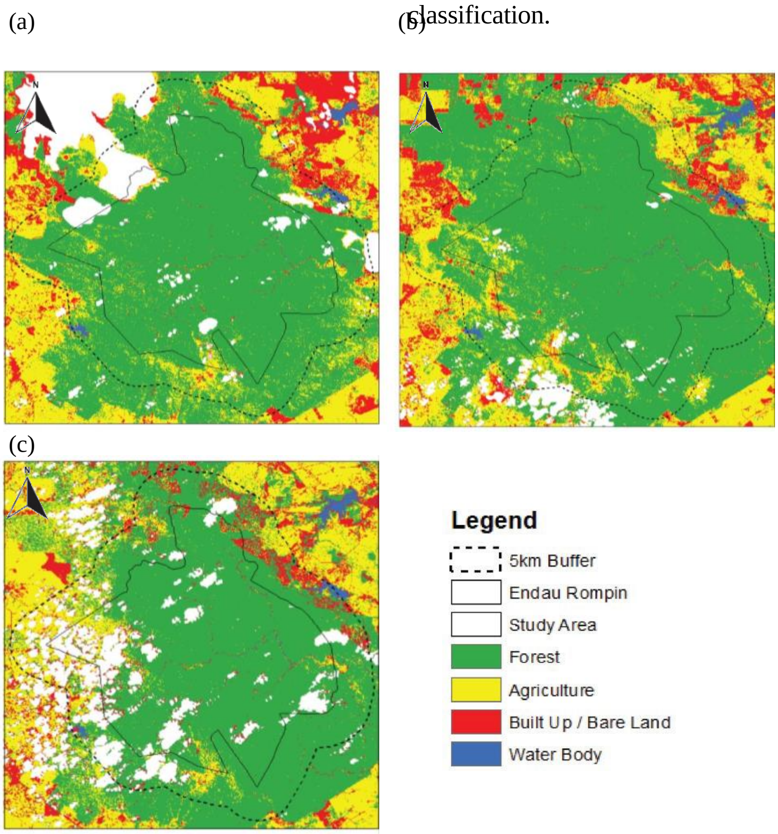

Figure 3. LULC maps for the years (a) 2013, (b) 2016, and (c) 2022 in the studied area by using Maximum Likelihood classification

Generally, forest land remained the dominant land cover over the studied period, followed by agricultural land, built-up or cleared land, and waterbody. From 2013 to 2022, small declines in forest area were found in ERNP and its surrounding area. Forest land covered more than 60 % of the total area and there was no significant change in forest area (Table 3). The changes in forest areas did not follow a consistent pattern. The forest area within ERNP decreased slightly from 67.28% in 2013 to 66.20% in 2016.

Table 3. Land-cover (in hectare and percentage) from 2013 to 2022.

Class	2013		2016		2022	
	Area (ha)	Percentage (%)	Area (ha)	Percentage (%)	Area (ha)	Percentage (%)
Forest	163250.089	67.28	169890.223	66.20	144765.46	61.19
Agriculture	54696.2625	22.54	59120.943	23.04	77639.58	32.82
Built up / Bare Land	18999.13039	7.83	25577.329	9.97	10526.22	4.45
Water Body	5712.470023	2.35	2026.651	0.79	3644.86	1.54

B. Analysis on Endau-Rompin National Park land use land cover changes

The change detection analysis provided a basis for understanding the drivers and impacts of land cover changes (Table 4 and Figure 4), which is crucial for effective land management and conservation strategies. The results of the change detection analysis indicate significant land cover transformations within the study period. From 2013 to 2016, the conversion of forest to agriculture covered an area of 13,871.17 hectares, while forest to build- up/bare land accounted for 5,516.75 hectares, and forest to water body encompassed 31.84 hectares. In the subsequent period from 2016 to 2019, the forest underwent further changes with 8,926.15 hectares converted to agriculture, 3,702.19 hectares transformed into built- up/bare land, and 801.32 hectares transitioning to water bodies. Additionally, from 2016 to 2019, the conversion of forest to agriculture expanded significantly, covering an area of 27,147.57 hectares, with 3,178.10 hectares converted to built-up/bare land, and 1,034.492 hectares transformed into water bodies. Significant conversion was observed, especially in the outskirt region. Small clearances of forest land and agricultural land occurred, and most of the cleared land was replaced by agricultural land. Only a minority of the cleared land experienced forest regrowth.

Table 4. Forest area changes for year (2013-2016), (2016-2019) and (2019-2022).

Class Changes	2013 - 2016	2019-2022
	Area (ha)	Area (ha)
Forest - Agriculture	13871.17	27147.57
Forest - Built Up / Bare Land	5516.75	3178.10
Forest - Water Body	31.84	1034.49
Total	19419.76	31360.15

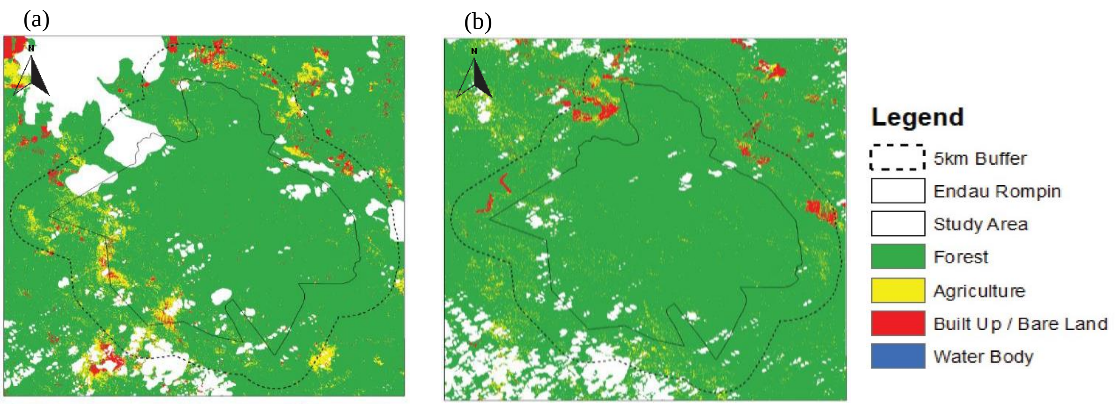

Figure 4. Land cover changes maps for year (a) 2013-2016, and (b)2019-2022

IV. DISCUSSIONS

Our overall results revealed significant changes in forest area in ERNP between 2013 to 2022 due to increases in agricultural land in the region. However, the result of land use land cover over the years shows a difference total size of the study area according to Table 2, from 242657.95 ha (2013), 256615.14 ha (2016), and 236576.11 ha (2022), this happened due to the number of clouds cover for every year are different make it difficult to produce a perfect LULC maps. Despite that, the percentage of every class for every year seems consistently maintained that can be seen in Table 2 as forested areas as the predominant land cover, with agricultural land, built-up or cleared areas, and water bodies.

The classification results demonstrate that RF outperforms ML in accurately classifying forest areas, as indicated by higher overall accuracy and Kappa coefficient values (Table 3 and appendix). This improvement is clear in our mapping analysis, as shown in Figure 2, and Figure 3. The RF classification achieved an impressive overall accuracy of 87% and a Kappa coefficient of 0.778, whereas the ML classification achieved a lower overall accuracy of 77% and a Kappa coefficient of 0.473. Additionally, the RF classification exhibited superior performance in detecting water bodies compared to ML, where water bodies were misclassified as built-up/bare land. This can be attributed to the robustness of the RF algorithm, which effectively handles noisy data and outliers, minimizing the risk of overfitting and enabling better generalization to new data [16]. Unfortunately, the accuracy assessment of the year 2019 was unable to conduct due to lack of ground data. However, the result for both classification in year 2022 seems to be a little bit lower than others due to the insufficient ground data and the area of data collected were smaller than the study area where the data were only collected in Peta area but the whole study area covered the whole area of ERNP.

The analysis of land change depicted in Table 2 and Figure 4 reveals a notable increase in the extent of agricultural and built-up/bare land areas. It is evident that forest areas have undergone significant transformations, being converted to both agricultural and built-up/bare land categories. The expansion of agricultural land, particularly the cultivation of oil palm, has been identified as the primary driver of increased forest loss in the vicinity of ERNP, aligning with the findings of previous research [1].

V. CONCLUSION

As a conclusion, RF classification outperforms ML classification (Table 3) in accurately mapping land use land cover in forested areas. However, it is crucial to emphasize the significance of image quality in minimizing classification errors, particularly in densely forested regions where class mixing may occur. The accuracy of classification outcomes greatly relies on the quality and resolution of the imagery utilized for analysis. Moreover, forest area loss in and surrounding ERNP from 2013 to 2022 was assessed in this study. The results indicated that ERNP itself exhibited a commendable level of protection against forest degradation. However, ERNP experienced slight decreases in forest cover primarily due to the expansion of agricultural activities, particularly the establishment of oil palm plantations. Despite these localized challenges, ERNP demonstrated effective forest conservation measures.

ACKNOWLEDGEMENTS

Authors were gratefully acknowledging Universiti Teknologi Malaysia for the financial support of UTM Fundamental Research Grant (Grant number Q.J130000.3852.22H69) and the authors sincerely thank the Johor National Park Corporation for the permission to conduct this research in Endau Rompin National Park.

REFERENCES

[1] Y. Sin Foo and S. Numata, "Deforestation and forest fragmentation in and around Endau-Rompin National Park, Peninsular Malaysia," *Tropics*, vol. 28, no. 2, pp. 23–37, 2019, doi: 10.3759/tropics.ms18-16.

[2] D. B. Kintz, K. R. Young, and K. A. Crews-Meyer, "Implications of land use/land cover change in the buffer zone of a national park in the tropical andes," *Environ. Manage.*, vol. 38, no. 2, pp. 238–252, 2006, doi: 10.1007/s00267-005-0147-9.

[3] R. A. Schowengerdt, "Remote sensing: models and methods for image processing," *Remote Sensing*, p. 560, 2007, doi: 10.1016/B978-012369407-2/50014-0.

[4] W. Cui, Z. Jia, X. Qin, J. Yang, and Y. Hu, "Multi-temporal satellite images change detection algorithm based on NSCT," *Procedia Eng.*, vol. 24, pp. 252–256, 2011, doi: 10.1016/j.proeng.2011.11.2636.

[5] G. Foody, "Hard and soft classifications by a neural network with a non-exhaustively defined set of classes," *International Journal of Remote Sensing*, vol. 23, no. 18, pp. 3853–3864, 2002. https://doi.org/10.1080/01431160110109570.

[6] M. E. Fagan *et al.*, "Mapping species composition of forests and tree plantations in Northeastern Costa Rica with an integration of hyperspectral and multitemporal landsat imagery," *Remote Sensing*, vol. 7, no. 5. pp. 5660–5696, 2015, doi: 10.3390/rs70505660.

[7] Y. Liu, W. Gong, X. Hu, and J. Gong, "Forest type identification with random forest using sentinel-1a, sentinel-2a, multi-temporal Landsat-8 and DEM Data," *Remote Sensing*, vol. 10, no. 6. 2018, doi: 10.3390/rs10060946.

[8] R. Cutler *et al.*, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.

[9] L. Guo, N. Chehata, C. Mallet, and S. Boukir, "Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 1, pp. 56–66, 2011, doi: https://doi.org/10.1016/j.isprsjprs.2010.08.007.

[10] T. G. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization," *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, 2000, doi: 10.1023/A:1007607513941.

[11] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[12] R. G. Congalton and K. Green, *assessing the accuracy of remotely sensed data: principles and practices*, 2005.

[13] G. M. Foody, "Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification," *Remote Sens. Environ.*, vol. 239, p. 111630, 2020, doi: https://doi.org/10.1016/j.rse.2019.111630.

[14] S. V Stehman and G. M. Foody, "Key issues in rigorous accuracy assessment of land cover products," *Remote Sens. Environ.*, vol. 231, p. 111199, 2019, doi: https://doi.org/10.1016/j.rse.2019.05.018.

[15] F. Ait El Haj, L. Ouadif, and A. Akhssas, "Monitoring land use and land cover change using remote sensing techniques and the precipitation-vegetation indexes in Morocco," *Ecol. Eng. Environ. Technol.*, vol. 24, no. 1, pp. 272–286, 2023, doi: 10.12912/27197050/154937.

[16] A. Balha and C. K. Singh, "Comparison of maximum likelihood, neural networks, and random forests algorithms in classifying urban landscape," in *Application of Remote Sensing and GIS in Natural Resources and Built Infrastructure Management*, V. P. Singh, S. Yadav, K. K. Yadav, G. A. Corzo Perez, F. Muñoz-Arriola, and R. N. Yadava, Eds. Cham: Springer International Publishing, pp. 29–38, 2022.

Appendix

2013 (RF)

Class Value	Forest	Agriculture	Built up / Bare Land	Water Body	Total	User Accuracy	Kappa
Forest	127	4	2	1	134	95%	0
Agriculture	3	40	1	2	46	87%	0
t up / Bare Land	1	0	15	0	16	94%	0
Water Body	1	1	1	7	10	70%	0
Total	132	45	19	10	206	0%	0
Producer Accuracy	96%	89%	79%	70%	0%	92%	0
Kappa	0	0	0	0	0	0	0.843

2013 (ML)

Class Value	Forest	Agriculture	Built up / Bare Land	Water Body	Total	User Accuracy	Kappa
Forest	113	4	6	4	127	89%	0
Agriculture	4	39	6	1	50	78%	0
t up / Bare Land	2	1	16	2	21	76%	0
Water Body	0	0	0	10	10	100%	0
Total	119	44	28	17	208	0%	0
Producer Accuracy	95%	89%	57%	59%	0%	86%	0
Kappa	0	0	0	0	0	0	0.752

2016 (RF)

Class Value	Forest	Agriculture	Built up / Bare Land	Water Body	Total	User Accuracy	Kappa
Forest	129	8	0	0	137	94%	0
Agriculture	6	40	2	0	48	83%	0
lt up / Bare Land	0	0	12	0	12	100%	0
Water Body	2	1	0	7	10	70%	0
Total	137	49	14	7	207	0%	0
Producer Accuracy	94%	82%	86%	100%	0%	91%	0
Kappa	0	0	0	0	0	0	0.817

2016 (ML)

Class Value	Forest	Agriculture	Built up / Bare Land	Water Body	Total	User Accuracy	Kappa
Forest	110	7	7	6	130	85%	0
Agriculture	10	39	0	0	49	80%	0
Built up / Bare Land	0	0	14	6	20	70%	0
Water Body	1	2	1	6	10	60%	0
Total	121	48	22	18	209	0%	0

Producer Accuracy	91%	81%	64%	33%	0%	81%	0
Kappa	0	0	0	0	0	0	0.665

2022 (RF)

Class Value	Forest	Agriculture	Built up / Bare Land	Water Body	Total	User Accuracy	Kappa
Forest	28	0	0	0	28	100%	0
Agriculture	3	7	0	0	10	70%	0
t up / Bare Land	5	0	1	4	10	10%	0
Water Body	0	0	0	10	10	100%	0
Total	36	7	1	14	58	0%	0
Producer Accuracy	78%	100%	100%	71%	0%	79%	0
Kappa	0	0	0	0	0	0	0.674

2022 (ML)

Class Value	Forest	Agriculture	Built up / Bare Land	Water Body	Total	User Accuracy	Kappa
Forest	26	2	0	0	28	93%	0
Agriculture	5	5	0	0	10	50%	0
t up / Bare Land	5	0	1	4	10	10%	0
Water Body	3	1	0	6	10	60%	0
Total	39	8	1	10	58	0%	0
Producer Accuracy	67%	63%	100%	60%	0%	66%	0
Kappa	0	0	0	0	0	0	0.443