Global Journal of Engineering and Technology Advances

eISSN: 2582-5003

Cross Ref DOI: 10.30574/gjeta

Journal homepage: https://gjeta.com/

(Research Article)

Check for updates

Supervised land cover classification of Nueva Ecija using random forest in google earth

Bensi Papa Leonylyn *, Michael Enrile Bensi and Apple Grace Galang Oliveros

College of Information and Communications Technology, Nueva Ecija University of Science and Technology, Cabanatuan City, Nueva Ecija, Philippines.

Global Journal of Engineering and Technology Advances, 2024, 21(01), 115–118

Publication history: Received on 02 September 2024; revised on 12 October 2024; accepted on 15 October 2024

Article DOI: https://doi.org/10.30574/gjeta.2024.21.1.0186

Abstract

Land cover classification is essential for environmental monitoring, urban planning, and sustainable land use management. This study presents a supervised land cover classification in Nueva Ecija, Philippines, utilizing Google Earth Engine (GEE) and the Random Forest (RF) algorithm, applied to Landsat 8 imagery. A total of 1,523 samples were collected representing five land cover types: built-up areas, agricultural lands, water bodies, forests, and barren land. The classification achieved an overall accuracy of 87.69% with a Kappa coefficient of 0.841. Future work should explore the integration of seasonal imagery and topographic indices for improved performance. This methodology provides crucial insights for resource management and supports regional policy development in Nueva Ecija.

Keywords: Accuracy Assessment; Google Earth Engine; Land Cover Classification; Landsat 8; Nueva Ecija; Remote Sensing; Random Forest

1. Introduction

Land cover classification plays a crucial role in environmental monitoring, land use planning, resource management, and addressing global challenges like climate change. Accurate classification aids in assessing ecosystem changes, urban expansion, agricultural practices, and deforestation. Satellite imagery, such as Landsat, provides powerful tools for large-scale monitoring with high temporal and spatial resolution [1]. The integration of machine learning algorithms like Random Forest (RF) has significantly improved land cover classification accuracy [2].

Random Forest, a supervised machine learning algorithm, is well-suited for land cover classification due to its ability to handle large datasets, resist overfitting, and process high- dimensional data [3]. It has been shown to outperform traditional methods like Maximum Likelihood Classification and Decision Trees in satellite imagery classification [4]. These advancements strengthen the ability to differentiate complex land cover types with high accuracy [5].

This study aims to classify land cover in Nueva Ecija, a province in the Philippines, using Landsat 8 imagery and the RF algorithm. The region is primarily agricultural, with urban areas, forests, and water bodies. Understanding land cover distribution here is vital for sustainable management and environmental planning [6]. Supervised classification is crucial for monitoring climate change impacts, urban development, and resource management. This study uses cloud-masked Landsat 8 data from 2022–2023, applying the RF classifier to generate a land cover map and evaluate model performance through validation metrics.

* Corresponding author: Bensi Leonylyn P

2. Methodology

2.1. Data Acquisition

Landsat 8 imagery from the LANDSAT/LC08/C02/T1_L2 collection was utilized for this study, focusing on the province of Nueva Ecija in the Central Luzon region of the Philippines. The imagery was filtered to include scenes captured between March 2022 and June 2023, particularly during the summer months to reduce cloud cover impact [7,8]. A total of 1,523 samples were collected across five land cover classes: built-up areas, agricultural lands, water bodies, forests, and barren land (Table 1).

Table 1 Land Cover Classes and Corresponding Sample Counts

Land Cover Type	Sample Count
Built-up Areas	300
Agricultural Land	600
Water Bodies	150
Forests	323
Barren Land	150
Total	1,523

2.2. Cloud Masking and Composite Creation

Cloud masking was performed using the QA_RADSAT band from the Landsat 8 data, targeting cloud and cloud shadow pixels [9]. A composite image was created by applying a median reducer to the filtered images. Six reflectance bands (SR_B2, SR_B3, SR_B4, SR_B5, SR_B6, and SR_B7) were selected for subsequent classification processes.

2.3. Random Forest Classification

The Random Forest classifier was applied to the dataset using 1,523 samples, with 1,212 samples (approximately 80%) allocated for training and 311 samples (approximately 20%) reserved for testing. The classifier was configured to use 300 trees and three variables at each split. The six selected reflectance bands—SR_B2, SR_B3, SR_B4, SR_B5, SR_B6, and SR_B7—were used as input features for the classification model [2,5].

3. Results and Discussion

3.1. Land Cover Map

The land cover map generated by the Random Forest classifier effectively differentiated the five land cover types within the study area. Built-up areas were clearly identified in urban centers, demonstrating the model's ability to discern developed regions. Agricultural lands were predominantly mapped in rural zones, reflecting their extensive presence in the region [8]. Forested areas were accurately located in higher elevation zones, and water bodies were correctly identified along the province’s rivers and lakes. Figure 1 illustrates the classified land cover map.

3.2. Random Forest Performance

The Random Forest classifier demonstrated impressive performance, achieving a training accuracy of 99.83% and a validation accuracy of 87.69%. The confusion matrix for the validation data (Table 2) reveals that most land cover classes were predicted accurately, though some misclassifications occurred, particularly between agricultural areas and water bodies. The Kappa coefficient for the validation dataset was 0.841, indicating a substantial agreement between the predicted and actual land cover classes [10].

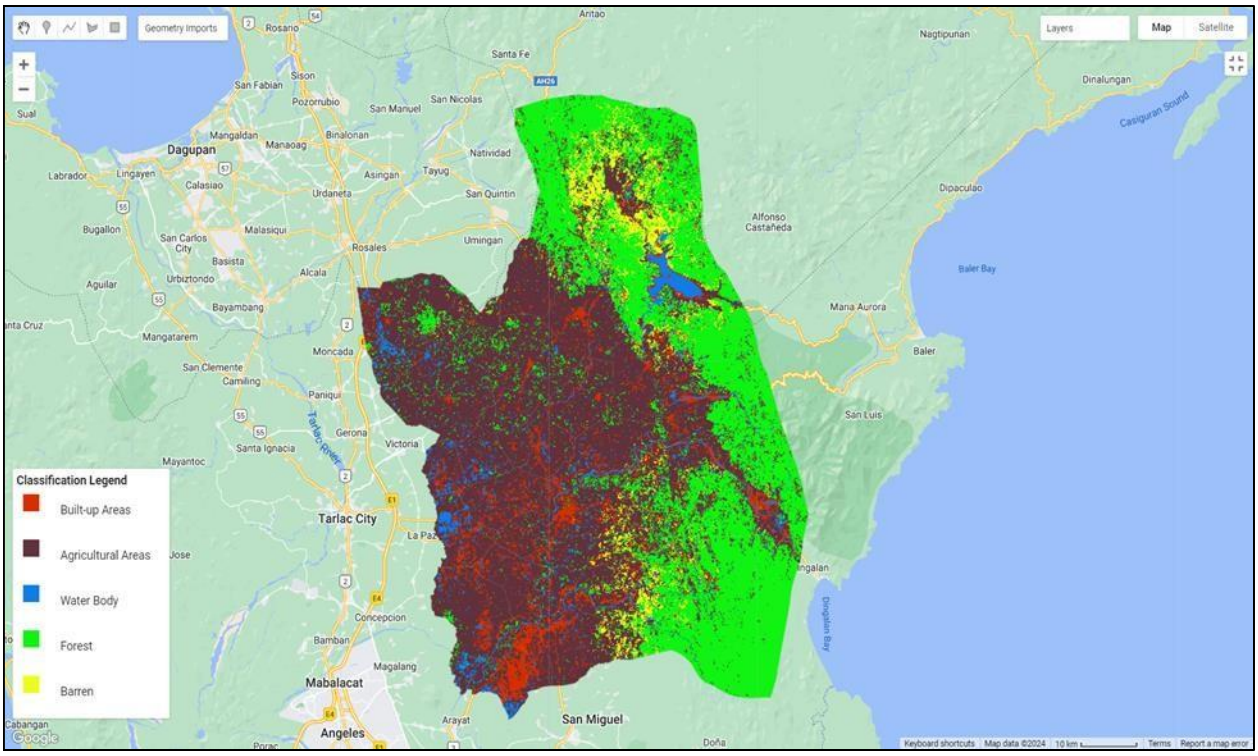

Figure 1 Land Cover Map of Nueva Ecija using Random Forest Classification

3.3. Interpretation

The classification results align with known land cover trends in Nueva Ecija, where agricultural areas are the predominant land cover type [12]. Incorporating additional data sources such as topographic indices or seasonal imagery could further enhance model performance, providing more nuanced insights and improving classification accuracy [13].

Table 2 Confusion Matrix for Validation Dataset

Class	Built- up	Agriculture	Water	Forest	Barren	Total	User's Accuracy (%)
Built-up Areas	60	5	2	1	2	70	85.71
Agricultural	6	150	3	5	2	166	90.36
Water Bodies	1	6	30	0	3	40	75
Forests	1	5	0	120	4	130	92.31
Barren Land	0	1	1	2	30	34	88.24
					Overall Accuracy		**87.69**

4. Conclusions

The application of the Random Forest algorithm for land cover classification in Nueva Ecija has yielded robust results, successfully distinguishing between built-up areas, agricultural lands, forests, and water bodies. The classifier's exceptional training accuracy of 99.83% and strong validation accuracy of 87.69% demonstrate its effectiveness in mapping land cover types in the region [5]. The high Kappa coefficient of 0.841 for the validation data further confirms substantial agreement between predicted and actual land cover classes, validating the model's reliability. Future research should consider integrating additional data sources, such as topographic indices and seasonal imagery, to enhance the model's accuracy and address misclassifications [13].

Compliance with ethical standards

Acknowledgments

The authors want to acknowledge Nueva Ecija University of Science and Technology for providing access to computing resources and research facilities. Thanks to Google Earth Engine for making remote sensing data readily accessible.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Hansen MC, Loveland TR. A review of large area monitoring of land cover change using Landsat data. Remote Sens Environ. 2012; 122:66–74.

[2] Breiman L. Random forests. Machine Learning. 2001;45(1):5–32.

[3] Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens. 2005;26(1):217–22.

[4] Gislason PO, Benediktsson JA, Sveinsson JR. Random forests for land cover classification. Pattern Recognit Lett. 2006;27(4):294–300.

[5] Pelletier C, Webb G, Petitjean F. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 2019;11(5):523.

[6] Phinn S, Roelfsema C, Dekker AG. Mapping coastal environments with hyperspectral remote sensing. Boca Raton: CRC Press; 2018.

[7] Zhu Z, Woodcock CE. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens Environ. 2012; 118:83–94.

[8] US Geological Survey. Landsat 8 (L8) Data Users Handbook. Sioux Falls, SD: US Geological Survey; 2016.

[9] Li P, Guo J, Xu J, et al. Integration of seasonal imagery for improving land cover classification using random forest algorithm. ISPRS J Photogramm Remote Sens. 2017; 132:63–73.

[10] Stehman SV, Foody GM. Accuracy assessment. In: Manual of Remote Sensing, Vol. 4. Hoboken: Wiley; 2018.

[11] Chen H, Liu D. Urban land cover classification with airborne hyperspectral data: A comparison of decision tree-based classifiers. J Appl Remote Sens. 2023;17(1):1–14.

[12] Gomez C, White JC, Wulder MA. Historical land cover classification with Landsat. Remote Sens Environ. 2020; 237:111466.

[13] Li Z, Eastman JR, Jing C. Spatial analysis of land use/land cover change. J Appl Geography. 2020; 121:102250