Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Assessing the Accuracy of Random Forest in Mapping Urban Green Cover in Baguio City Using Sentinel-2 Imagery and Spectral Indices

Joan M. Peralta [1*]

[1] *Assistant Professor, College of Information Technology and Computer Science, University of the Cordilleras, Baguio City, Philippines.*
Email: jmperalta@uc-bcf.edu.ph, ORCID Id: https://orcid.org/0009-0002-7858-6286

ARTICLE INFO

Received: 30 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Urban green cover plays a pivotal role in sustainable urban development by providing environmental and socio-economic benefits. Accurate mapping of urban green cover is essential for developing urban greening strategies and managing urban green spaces in smart cities. Remote sensing, particularly Sentinel-2 imagery, offers high-resolution multispectral data suitable for vegetation analysis. Machine learning algorithms, such as Random Forest, have proven effective in classifying land cover, including urban green spaces. This research investigates the accuracy of Random Forest in mapping urban green cover in Baguio City, Philippines, utilizing Sentinel-2 imagery and spectral indices. The study utilized spectral indices, such as NDVI, SAVI, NDWI, and NDBI to train and validate the Random Forest model. The performance of the Random Forest classifier is evaluated using standard accuracy assessment metrics, such as overall accuracy, producer's accuracy, user's accuracy, F1-score, and Kappa coefficient. The Random Forest model proved to enhance the classification of urban green cover with 85.71%, 86%, and 86% for the overall accuracy, producer's accuracy for urban green cover, and consumer's accuracy for urban green cover, respectively. The F1-score of 0.92308 and Kappa coefficient of 0.7790927 showed that the Random Forest model accurately classified the urban green cover. By utilizing remote sensing and machine learning techniques, this research seeks to contribute to the development of accurate and up-to-date urban green cover maps. The findings shall provide valuable insights for urban planners and policymakers in Baguio City, enabling them to implement effective strategies for urban greening and sustainable urban development.

Keywords: Machine Learning, Random Forest, Remote Sensing, Sentinel-2 Imagery, Spectral Indices, Urban Green Cover Mapping, SDG 11 Access to Green Space.

INTRODUCTION

Urban green covers are essential for improving the quality of life in cities, as they offer a variety of economic, social, and environmental advantages [1], [2], [3], [4]. Urban green cover encompasses the urban green space or vegetative land in the urban environment, including a wide variety of vegetation types ranging from large parks and forests to small patches of trees and even individual trees [5], [6]. These urban green spaces are instrumental in biodiversity preservation [5], climate mitigation [7], air purification [8], and satisfactory human well-being. Effective urban planning, environmental management, and public health initiatives compel precise and current information regarding urban green spaces [9].

Many cities are currently striving to achieve sustainable development, and urban green spaces are integral in achieving sustainability [3], [5]. Achieving the United Nations Sustainable Development Goals, SDG11, targeting the provision of universal access to safe, inclusive, and accessible, green and public spaces is essential [2], [42]. Cities can enhance the quality of life for the residents, reduce pollution, and increase their resilience to climate change by promoting sustainable land use practices and incorporating green spaces.

Baguio City, located in the northern part of the Philippines, is currently in the process of transforming into a smart city. The objective of this initiative is to enhance the quality of life for its residents, as well as the city's efficacy and sustainability, by utilizing technology [10], [11]. Baguio's smart city development is centered on the following key areas such as smart governance, which is the integration of digital systems to optimize administrative procedures, increase transparency, and improve citizen engagement; smart infrastructure deals with the creation of intelligent infrastructure, including energy-efficient buildings, smart transportation systems, and waste management solutions; smart mobility is the promotion of sustainable transportation options, such as pedestrian-friendly infrastructure, bike-sharing programs, and electric vehicles; smart environment with the implementation of technology to monitor environmental conditions, manage natural resources, and encourage sustainability; and smart living deals with improving the quality of life for residents by utilizing smart services, including e-health, e-education, and smart home technologies. Baguio City endeavors to establish itself as a paradigm for sustainable urban development in the Philippines by adopting smart city initiatives [10], [11].

Smart cities are increasingly incorporating advanced technologies to enhance sustainability, efficiency, and viability. Smart green cities that are more resilient, sustainable, and equitable can be established by integrating smart technologies with urban green spaces [12]. The use of machine learning algorithms, in conjunction with remote sensing techniques, provides an efficient and scalable method for the evaluation and monitoring of urban green spaces [2]. Random forest, a machine learning algorithm, is well-suited for urban green space mapping due to its ability to manage intricate relationships between spectral indices and land cover classes [13], [14], [15]. In remote sensing techniques, the distinctive spectral signatures of various land cover categories, such as urban built-up areas, bare soil surfaces, water bodies, and urban green cover, can be captured by spectral indices that are derived from multispectral or hyperspectral imagery. These indices, combined with machine learning, offer significant insights into the health, density, and composition of vegetation [16].

This research paper explores the application of spectral indices and random forest machine learning to map urban green spaces in Baguio City. The objectives of this paper are as follows: (1) evaluate the effectiveness of various spectral indices in differentiating urban green spaces from other land cover categories in Baguio City; (2) develop a random forest model for classifying urban green spaces based on spectral index values; and (3) assess the random forest model's performance and accuracy in classifying urban green space. This research endeavors to advance urban green space mapping, investigate the potential applications of the model, and offer valuable insights to policymakers and urban planners in Baguio City establishing a more sustainable and resilient city by addressing these objectives.

RELATED WORK

Sustainable urban development, a paradigm that seeks to balance economic growth, social equity, and environmental protection [43], is a key driver for the integration of urban green spaces into urban planning. Urban green spaces play a vital role in promoting sustainable urban development by providing a variety of ecosystem services. To create an accurate map of urban green spaces, machine learning techniques can be employed. Machine learning is a subset of artificial intelligence that has become a powerful tool for analyzing large, complex data sets. The random forest model is a popular algorithm of machine learning that has been widely used in remote sensing technology. The random forest's ability to process high-dimensional data, robustness to noise, and interpretability make it a promising candidate for urban green space mapping.

The random forest is a powerful machine learning model well-suited for classifying urban green space due to its ability to handle high-dimensional data, robustness to noise, and capacity to provide insights into variable importance [13], [17], [18]. Random Forest operates as a machine learning method that constructs various decision trees, and each decision tree is trained on a different random subset of the data [19]. The final classification is determined by a majority vote among all the trees.

Many studies have confirmed the effectiveness of Random Forest for urban vegetation mapping. Reference [13] in Augsburg, Germany, successfully identified various types of green spaces with a 97% accuracy using Random Forest on Sentinel-2 imagery and explored the use of single spectral bands, vegetation indices, and a combination of both as input variables, finding that while all models performed well, the inclusion of vegetation indices significantly improve the accuracy.

Research in China further demonstrates the efficacy of Random Forest in urban green space mapping. A study mapping urban forests across the country using Sentinel-2 images and the Random Forest model on the Google Earth Engine platform achieved an overall accuracy of 92.30%. The study highlighted the importance of using appropriate spectral indices, such as NDBI, NDVI, and NDWI, as supplementary information for classification [14]. Another study in Beijing successfully classified urban green space using Random Forest and Sentinel-2 imagery, demonstrating the algorithm's ability to handle complex urban environments [9].

Further research has explored the integration of Random Forest with other data sources for enhanced classification accuracy. A study in Xuzhou, China, found that combining Sentinel-1 SAR and Sentinel-2 MSI data with Random Forest and an object-oriented approach classification improved the accuracy of aboveground biomass estimation for different urban vegetation types [15]. A study in Europe showed the importance of multi-temporal data having a significant impact on improving accuracy in urban vegetation mapping using Random Forest, which is a top performer in the study, specifically in handling large datasets and complex classification tasks in urban environments [20]. This further supports the effectiveness of Random Forest for handling the complexities of urban environments and multi-temporal data [6]. Another study's findings of using advanced methods for extracting meaningful features from both image and non-image data, capturing the complexity of urban land use, offered a promising approach for leveraging multi-source data to improve classification accuracy in urban land use mapping and inform urban planning efforts [21]. These studies emphasized the potential of multi-source data fusion for improved classification accuracy.

Comparative studies have consistently shown Random Forest's superior performance and consistently high accuracy in urban green space classification over other machine learning models, such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), and K-Nearest Neighbours (KNN). In China, [14] determined that the urban forest cover is between 10% and 20% by using the Random Forest machine learning algorithm on Google Earth Engine to classify urban forests from Sentinel-2 images. In Africa, [22] highlighted that Random Forest demonstrably outperformed Support Vector Machine in terms of accuracy, and significantly excelled in classifying mixed land cover classes, particularly well-suited for mapping large areas using coarse-resolution imagery. In one study [8], the Random Forest algorithm was found to be more accurate and faster than the K-Nearest Neighbours (KNN) algorithm for urban footprint extraction, showing overall accuracies of 82.08% and 77.89%, respectively. Another study found that the Random Forest model was better than both the SVM and Artificial Neural Network models for classifying urban land use, achieving a validation accuracy of 79.88% for Level I and 71.89% for Level II land use [23]. Finally, [15] showed that an object-based Random Forest classification had higher overall accuracy (86.59%) than KNN (82.68%) for identifying urban green space types, and also showed the highest classification accuracy with a value of 89.62% as its producer accuracy for coniferous forests.

While Random Forest has proven highly effective for classifying urban green space, research suggests the importance of careful parameter tuning and training data selection for optimal performance [24]. Studies [7], [25] also suggest that integrating spatial metrics and texture analysis with Random Forest can further enhance classification accuracy in urban environments. Similarly, [26] discussed evaluating hundreds of input parameter combinations and identifying the combination resulting in the highest accuracy based on overall accuracy and the Kappa index. [27] also investigates the effect of the number of trees in the Random Forest classifier, concluding that using more trees improves accuracy. Finally, the Random Forest classifier demonstrated robustness in handling high-dimensional datasets and its ability to provide insights into feature importance in classifying land cover [28].

MATERIALS AND METHODS

This section discusses the focused study area, the dataset used, and the workflow of the Random Forest model development starting with data preprocessing, feature extraction, and feature engineering. Figure 1 shows the workflow of developing the Random Forest model.

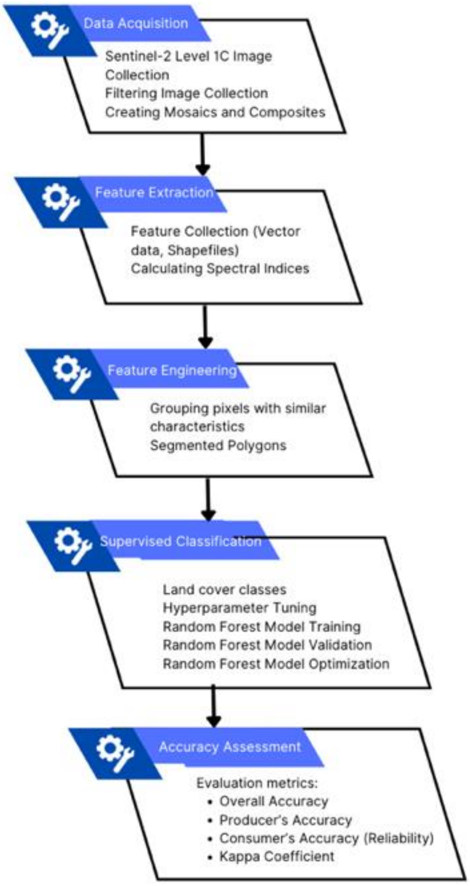
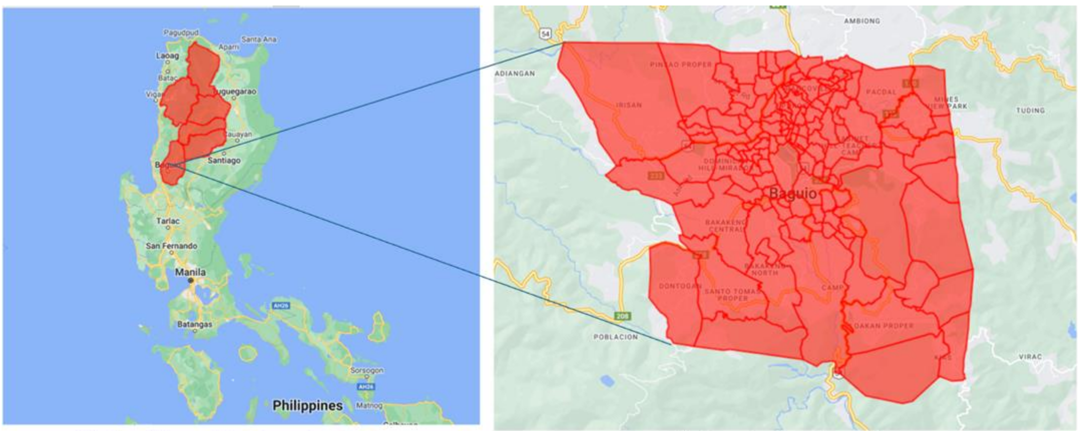


Figure 1. Workflow of Random Forest Model

A. Study Area

Baguio City, a highly urbanized city in the Philippines located at latitude 16.41639 and longitude 120.59306 [29], is selected as the study area. The city has a land area of 57.51 square kilometers and its population as determined by the 2020 Census was 366,358 [30]. The city is geographically situated within the province of Benguet and is part of the Cordillera Administrative Region (see Figure 2).



(a) Cordillera Administrative Region *(b) Baguio City Shapefile*

Figure 2. The Cordillera Administrative Region (a), located in the northern part of the Philippines, is home to the picturesque province of Benguet, which in turn boasts the charming chartered city of Baguio (b).

The city is ranked 12th in the 2024 Rankings of Highly Urbanized Cities in the Philippines by the Department of Trade and Industries [31] with a 48.1823 score in terms of economic dynamism, government efficiency, infrastructure, resiliency, and innovation. The built-up areas of Baguio City are continuously expanding as a result of the city's economic development, the thriving tourism industry, and the rapid growth of the urban population.

B. Dataset

In this study, the Harmonized Sentinel-2 MSI: Multispectral Instrument, Level 1C dataset, provided by the European Union through the European Space Agency and Copernicus, is utilized. The access and use of the Copernicus Sentinel Data and Service Information is regulated under the European Law. The dataset is a collection of harmonized sentinel-2 remote sensing data which is a wide-swath, high-resolution, multi-spectral imaging supporting Copernicus Land Monitoring studies, including monitoring of vegetation, soil and water cover, as well as of inland waterways and coastal areas [34]. The Sentinel-2 satellite series provides diverse spatial resolutions (10m, 20m, and 60m), enabling a wide range of applications, including atmospheric and geophysical corrections, vegetation analysis, and land classification. The 13 spectral bands captured by Sentinel-2, spanning the visible, near-infrared, and shortwave infrared regions, offer comprehensive data for detailed Earth surface analysis [24]. The study focused on the Sentinel-2 high-resolution multispectral satellite imagery of Baguio City, Philippines from January 1, 2023 to December 31, 2023.

C. Data Preprocessing

The study used Sentinel-2 imagery covering latitude 16.41639 and longitude 120.59306 for urban green cover classification. The Sentinel-2 Level 1C dataset has undergone radiometric and geometric corrections within a global orthogonal rectification and spatial registration [14] and is accessible via the Google Earth Engine image collection. Baguio City is characterized by frequent rainfall throughout the year, and due to its high elevation, the city is often enveloped in cloud cover, necessitating the implementation of cloud masking techniques. Cloud masking pertains to developing a dependable cloud masking algorithm to eradicate cloud-contaminated pixels from the Sentinel-2 imagery. Cloud detection is a necessary phase in satellite image processing to retrieve atmospheric parameters which includes classifying pixels into thick clouds, thin clouds, and the background [35]. An atmospheric correction using the quality band (QA60) from Sentinel-2 in Google Earth Engine was implemented to mask cloud and pick optimal images with low cloud cover. See Figure 3 for the cloud mask image of Baguio City.

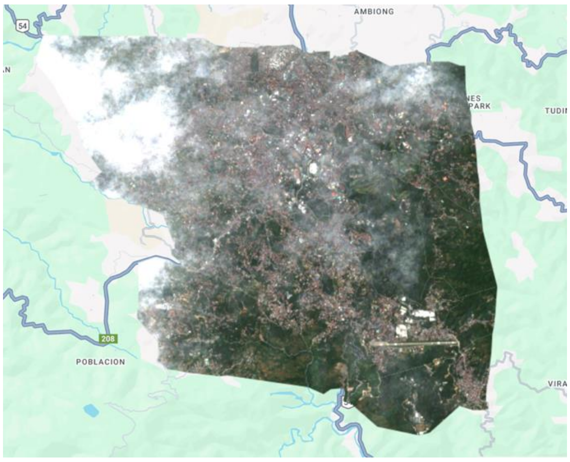


Figure 3. Cloud-covered Baguio City Satellite Imagery

D. Feature extraction

Relevant spectral and spatial features were extracted from the Sentinel-2 Level 1C imagery, such as vegetation indices, texture metrics, and elevation data for obtaining vegetation information. The use of vegetation indices is a valuable feature for identifying urban green cover by helping to reduce soil and atmospheric effects while enhancing the information present in single spectral bands, leading to a better understanding of vegetation characteristics by calculating NDVI, SAVI, NDBI, and MNDWI from Sentinel-2 spectral bands. Section IV-A provides a detailed discussion on NDVI, SAVI, NDBI, and MNDWI and how it affects feature engineering methods.

E. Feature engineering

New features may be created by combining existing features or applying mathematical transformations to improve classification performance. Pixels in the imagery are grouped together with similar characteristics. Replacing the isolated pixels with the value of neighboring pixels to have connected pixels then generates a smoothed, segmented polygon.

RESULTS

A. Effectiveness of Spectral Indices in Differentiating Urban Green Space

Spectral indices are mathematical combinations of reflectance values from various spectral bands that are employed to emphasize particular features of interest in a remote-sensing image [13]. These indices are intended to distinguish vegetation cover from other land cover categories and emphasize it in the context of green space mapping. The researcher chose the following spectral indices from the Sentinel-2 index database to distinguish urban green cover: Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Normalized Difference Built-Up Index (NDBI), and Modified Normalized Difference Water Index (MNDWI). The next section discusses the description of the selected indices as well as the equations to calculate them.

The Normalized Difference Vegetation Index (NDVI) is a remote sensing technique that utilizes light reflectance in the visible and near-infrared (NIR) spectra to assess vegetation quantity and health in a given area. It utilizes the disparity between near-infrared (NIR) and red reflectance to distinguish vegetation from other land cover categories. NDVI values generally span from -1 to 1, wherein elevated positive values signify denser and more robust vegetation, while lower or negative values denote barren soil or places devoid of vegetation [36]. Utilizing NDVI on urban imagery enables the identification and mapping of diverse green spaces, including parks, gardens, trees, and urban forests. NDVI may be associated with urban heat island phenomena. Regions with elevated NDVI values frequently experience reduced temperatures because of the cooling influence of vegetation [37]. Furthermore, NDVI can facilitate the identification of optimal sites for new green spaces and evaluate the effects of development initiatives on existing vegetation See Figure 4 (b) for the graphical calculation of NDVI with reference to the raw composite of Baguio City in Figure 4 (a).

The NDVI is calculated by

$$NDVI = \frac{NIR[Band8] - RED[Band4]}{NIR[Band8] + RED[Band4]} \tag{1}$$

where NIR[Band8] and RED[Band4] represent the near-infrared and the red band of the Sentinel-2A imagery, respectively.

The Soil-Adjusted Vegetation Index (SAVI) is a modified version of NDVI that is specifically designed to reduce the impact of soil luminosity on vegetation index calculations, particularly in regions with reduced vegetation cover or bare soil. SAVI includes a soil brightness correction factor that accounts for the influence of soil reflectance on the calculated vegetation index. This is particularly critical in urban settings, where soil variability can be substantial. By adjusting for soil brightness, SAVI can provide more precise estimates of vegetation cover and health in areas with mixed land covers, such as urban-agriculture interfaces or areas with exposed soil [38].

The SAVI is calculated by

$$SAVI = \frac{(1+L) * ((NIR[Band8] - RED[Band4]))}{(NIR[Band8] + RED[Band4] + L)} \tag{2}$$

where NIR[Band8] is the reflectance in the near-infrared range, RED[Band4] is the reflectance of red light, and L is the correction factor for soil luminance, which is typically set between 0.1 and 0.5. The researcher enhanced the accuracy of green space mapping and analysis by obtaining more reliable vegetation information in urban environments where soil conditions can vary substantially, through the use of SAVI. See Figure 4 (c) for the graphical calculation of SAVI with reference to the raw composite of Baguio City in Figure 4 (a).

The Normalized Difference Built-Up Index (NDBI) is a remote sensing technique used to identify and quantify built-up areas. By analyzing satellite images, NDBI capitalizes on the spectral differences between developed and undeveloped land to distinguish between built-up and non-built regions [39]. With an accuracy rate often exceeding 80%, NDBI is a valuable tool for assessing urban growth, land use changes, and population estimation in various geographical areas. By deducting NDBI from NDVI, one can isolate green zones that are unassociated with developed regions. NDBI can be utilized to observe alterations in urban land cover over time, offering critical insights for urban planning and development [40] [41].

The NDBI is calculated by

$$NDBI = \frac{NIR[Band8] - SWIR[Band11]}{NIR[Band8] + SWIR[Band11]} \tag{3}$$

where NIR [Band8] is the near-infrared reflectance and SWIR[Band11] is the short-wave infrared reflectance.

Utilizing NDBI enabled the researcher to acquire precise and dependable data regarding built-up regions in urban settings, facilitating many applications in urban planning, environmental management, and disaster response. See Figure 4 (d) for the graphical calculation of NDBI with reference to the raw composite of Baguio City in Figure 4 (a).

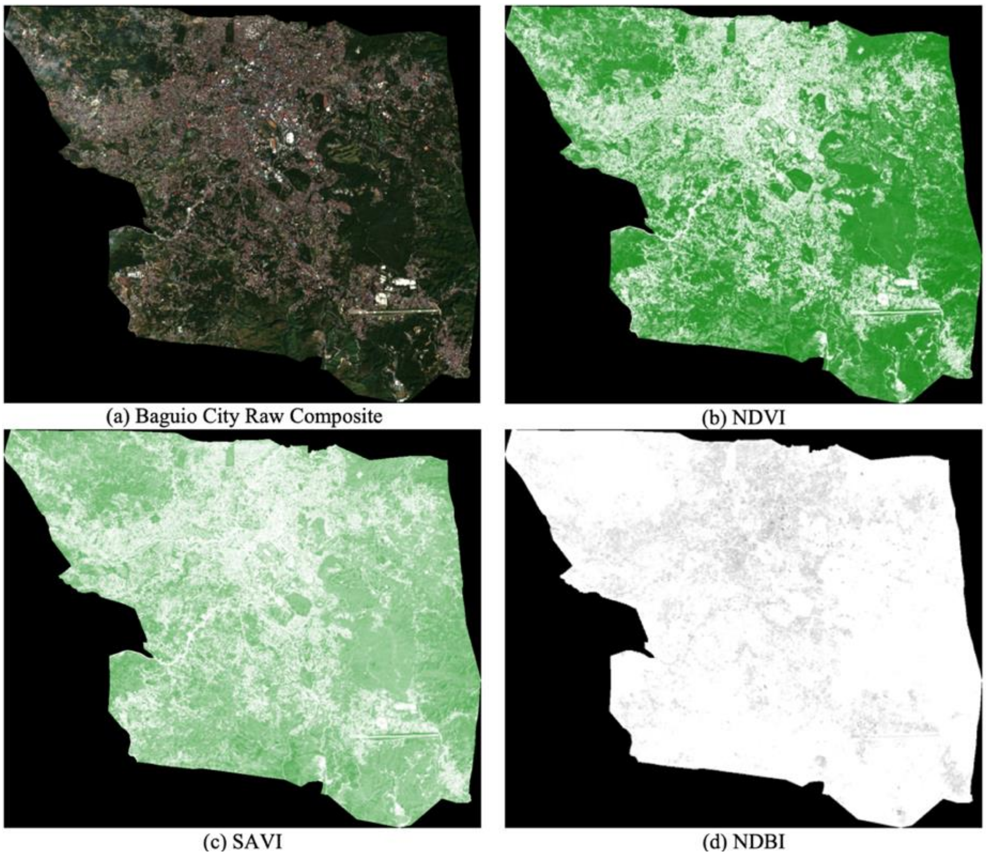
The Modified Normalized Difference Water Index (MNDWI) is a spectral index specifically formulated to augment the sensitivity of the Normalized Difference Water Index (NDWI) to water bodies, especially in urban settings where buildings and other structures may disrupt conventional water indices.

MNDWI is superior to NDWI in detecting water bodies in urban environments, including ponds, lakes, and irrigation canals, despite the presence of developed regions and various land cover types. By precisely identifying water bodies inside urban green spaces, MNDWI can facilitate the assessment of the ecological integrity of these regions, as water bodies are frequently essential elements of urban ecosystems. MNDWI can be utilized to track temporal variations in water bodies, offering critical insights for urban water resource management and planning.

The MNDWI is calculated by

$$MNDWI = \frac{GREEN[Band3] - NIR[Band8]}{GREEN[Band3] + NIR[Band8]} \tag{4}$$

where GREEN[Band3] is the green reflectance and NIR [Band8] is the near-infrared reflectance. Since MNDWI has less susceptibility to urban regions, the researcher calculated the result with enhanced accuracy in detecting water bodies within intricate metropolitan settings. MNDWI exhibited heightened sensitivity to shallow water bodies, which may be challenging to identify using conventional water indices. Utilizing MNDWI enabled the researcher to acquire more precise and dependable data regarding water bodies in urban green spaces, facilitating many applications in urban planning, environmental management, and water resource management. See Figure 4 (e) for the graphical calculation of MNDWI with reference to the raw composite of Baguio City in Figure 4 (a).



(a) Baguio City Raw Composite (b) NDVI

(c) SAVI (d) NDBI

(e) MNDWI

Figure 4 (a) A very high-resolution composite image of Baguio City from Google Earth is in reference with the visual calculation of the selected spectral indices such as **(b)** Normalized Difference Vegetation Index (NDVI), **(c)** Soil-Adjusted Vegetation Index (SAVI), **(d)** Normalized Difference Built-Up Index (NDBI), and **(e)** Modified Normalized Difference Water Index (MNDWI).

B. Random Forest Model Development

The initial step in the land cover classification process involved collecting training samples directly from the Google Earth Engine code editor using high-resolution imagery sourced from Google Maps. This method proved effective in generating high-quality classification samples for a specific area without requiring any additional training data. Using the raw composite imagery of Baguio City, the objective was to classify each pixel into one of four categories: built-up area, bare soil, water, or urban green cover. Using the drawing tools within the code editor, four distinct feature collections were created, each representing one of the aforementioned land cover classes. Each feature collection was assigned a numerical value (0, 1, 2, or 3) to denote its corresponding class. Subsequently, a Random Forest classifier was trained using these feature collections to construct a model that could classify all pixels in the image into one of the four classes. To implement the Random Forest classifier in Google Earth Engine, the Java statistical machine intelligence and learning engine (SMILE) library was utilized. The trained Random Forest classifier employed initially 10 trees, then hyper-tuned into 50 trees for the classification process. The number of trees parameter in the Random Forest model determined the number of decision trees that were combined to create the final classification. The Random Forest model categorized each pixel of the raw composite image of Baguio City into four classes: built-up area, bare soil, water, or urban green cover. The resulting classification map provided a color palette for each class such that built-up areas are in pink color, bare soil areas in yellow color, bodies of water in blue color, and the urban green cover in green color, as shown in Figure 5.

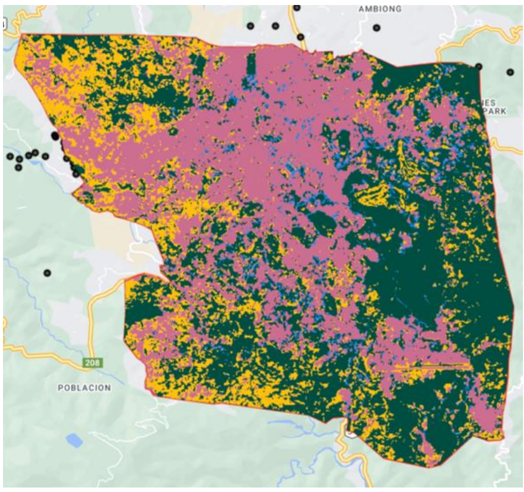


Figure 5. The Random Forest model categorizes each pixel of the raw composite image of Baguio City into four classes: built-up area (pink), bare soil (yellow), water (blue), or urban green cover (green).

C. Random Forest Model Classification Accuracy

To evaluate the model's performance, several accuracy metrics were computed. These include overall accuracy, which measures the overall correctness of the classification; producer's accuracy, which assesses the model's ability to correctly identify a particular class; consumer's accuracy, which measures the reliability of the model's predictions for a given class; F1-score, which provides a balanced measure of a model's performance by combining precision and recall; and the kappa coefficient, which evaluates the agreement between the predicted and actual classifications, accounting for random assignment

The confusion matrix in Table 1 provides a concise summary of the performance of a random forest model in the classification of land cover. The Random Forest model accurately classified 85.71% of the samples, as evidenced by the overall accuracy level of 0.85714. This is an accuracy that is relatively high, which implies that the model is performing well in general.

Table 1. Confusion Matrix

		Classification (Predicted)					
		0	1	2	3		
Ground Truth (Actual)	0	8	0	0	0	1	Producer's Accuracy
	1	2	3	0	0	0.60	
	2	0	0	1	0	1	
	3	1	1	0	12	0.86	
		0.73	0.75	1	1	0.85714	
		Consumer's Accuracy				Overall Accuracy	

For class-specific metrics, the producer's accuracy showed that the model accurately predicted 73% of the samples that were actually classified as built-up areas. Similarly, 75% of the samples that were genuinely bare soil areas were correctly classified. For the water areas, a value of 1.00 was computed indicating that all samples that were classified as water areas were accurately predicted. Lastly, the classification of 0.86 - 86% of the samples that were genuinely urban green cover was accurate. On the other hand, the consumer’s accuracy showed that 73% of the samples that were predicted as the built-up area were correctly classified, 75% of the samples predicted as bare soil area, all samples that were predicted to be in water area were correctly classified, and 86% of the samples that were predicted as urban green cover were correctly classified.

Furthermore, F1-score and Kappa coefficient should be considered dealing with the random forest model’s accuracy. The F1-score metric is particularly valuable when dealing with datasets where the classes are imbalanced, as it considers both false positives and false negatives, while the kappa coefficient quantifies the agreement between predicted and actual classifications, accounting for chance agreement. A kappa value of 0 indicates no agreement beyond random chance, while a value of 1 signifies perfect agreement.

The F1-score of 0.84211 suggests that the model is capable of accurately classifying built-up area while also maintaining a high level of recall and precision. A faultless F1-score for water area indicates that the model is capable of accurately classifying all instances of this class. Another high F1-score of 0.92308 indicates that urban green cover classification is performing well. Although the F1-score of 0.66667 for bare soil area is lower than that of built-up area and urban green cover, it still indicates a reasonable level of performance. This kappa coefficient value of 0.7790927 is relatively high, suggesting a strong agreement between the predicted and actual classifications which implies that the model is outperforming the random variable by a substantial margin.

DISCUSSIONS

The results of this study demonstrate the effectiveness of machine learning techniques for urban green space mapping in the context of sustainable urban development. NDVI, NDBI, SAVI, and MNDWI have been used to improve the accuracy of urban green space mapping using random forest models. These indices provide additional spectral information that helps distinguish between various land cover categories, such as vegetation, built-up areas, water bodies, and bare soils. By merging multiple indices, the model's ability to distinguish complex urban land cover types can be further improved. Therefore, integrating the spectral index and the traditional spectral bands as input features in the random forest model can significantly improve the accuracy of urban green space mapping. However, the optimal choice of an index may vary depending on the specific study area characteristics and satellite sensor capabilities. Therefore, it is important to experiment with various index combinations to determine the most efficient configuration for a given application.

The Random Forest model exhibits exceptional accuracy in classifying water areas and urban green cover, with water areas being classified with faultless precision. The model's ability to accurately predict and classify built-up and bare soil areas is supported by the fact that their producer's and consumer's accuracies are comparable. The model's effectiveness in classifying urban green cover is suggested by its relatively high overall accuracy.

However, the classification errors for built-up and bare soil areas could be addressed in order to accomplish further improvements, despite the model's overall impressive performance. This may entail the investigation of various feature engineering techniques, the modification n of hyperparameters, or the integration of supplementary data sources. If the class distribution is significantly imbalanced, it may be advantageous to implement strategies such as cost-sensitive learning or class balancing. In general, the random forest model exhibits optimistic performance in the mapping of urban green cover. The model's accuracy and effectiveness can be further improved by addressing the identified areas for enhancement.

The Random Forest model exhibits robust overall performance in the classification of urban green cover, as demonstrated by the high F1-scores and kappa coefficient. The model's capacity to precisely identify and prevent misclassification of built-up area, water area, and urban green cover is particularly evident. The model's capacity to classify bare soil area is indicated by a slightly lower F1-score, which implies that there may be room for advancement.

Nonetheless, it may be beneficial to investigate methods such as class balancing or modifying the hyperparameters of the random forest algorithm in order to improve the model's performance for bare soil areas. It may be beneficial to employ supplementary features or data sources that are more informative for bare soil areas and assess the model's efficacy in relation to various feature engineering techniques, including feature selection and normalization. In general, the F1-score and kappa coefficient values indicate that the random forest model is a highly effective model for urban green cover mapping. However, there is still room for further development, particularly in bare soil area classification.

While the Random Forest model performed well in this study, there are limitations to consider. The accuracy of the results depends on the quality of the remote sensing data and ground truth information. Factors such as cloud cover, sensor calibration, and the accuracy of field surveys can influence the overall performance of the model.

Thus, this study demonstrated the potential of Random Forest for accurate and efficient urban green space mapping in sustainable cities. The findings contribute to a better understanding of the role of urban green space in urban development and sustainability, providing valuable information for decision-makers and planners. Future research is needed to further refine urban green space mapping methods and explore their applications in addressing the challenges of urbanization and climate change.

CONCLUSIONS

This study successfully demonstrated the potential of machine learning techniques, particularly Random Forest, in accurately mapping urban green cover using Sentinel-2 imagery and spectral indices. By combining the strengths of remote sensing and machine learning, the researcher was able to generate a Random Forest model to be used to provide precise and up-to-date information on Baguio City's green spaces. The Random Forest model achieved high accuracy and precision in identifying and classifying urban green cover, providing valuable awareness into their distribution, extent, and characteristics. The results highlight the importance of urban green spaces in enhancing urban sustainability and resilience for the development of smart city in Baguio City, Philippines. The accurate mapping of these areas offers significant insights for urban planners, policymakers, and environmental scientists to make informed decisions regarding green space management, conservation, and expansion.

Future research could explore the use of advanced machine learning models, such as deep learning, for urban green space mapping. Incorporating additional data sources, such as LiDAR or drone imagery and hyperspectral data could enhance the accuracy and spatial resolution of urban green space maps. Furthermore, investigating the dynamics of urban green space and their response to climate change and urbanization patterns would provide valuable insights for sustainable urban planning. Additionally, future research directions may include developing effective methods for time-series analysis of urban green cover using remote sensing data, and quantifying the impact of urban green space loss in the ecosystem.

By capitalizing on the power of machine learning models and remote sensing technologies, coupled with sophisticated analytical techniques, will further enhance our capacity to effectively monitor, manage, and conserve urban green spaces, contributing to the development of sustainable and resilient cities.

Funding Statement:

There is no funding received for this article.

Data Availability:

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflict of Interest:

The author declares that there is no conflict of interest.

REFERENCES

[1] K. Abutaleb, M. F. Mudede, N. Nkongolo and S. W. Newete, "Estimating urban greenness index using remote sensing data: A case study of an affluent vs poor suburbs in the city of Johannesburg," *The Egyptian Journal of Remote Sensing and Space Science,* vol. 24, no. 3, pp. 343-351; https://doi.org/10.1016/j.ejrs.2020.07.002, 2021.

[2] S. S. Burrewar, M. Haque and T. U. Haider, "A Survey on Mapping of Urban Green Spaces within Remote Sensing Data Using Machine Learning & Deep Learning Techniques," *2023 15th International Conference on Computer and Automation Engineering (ICCAE),* pp. 30-34; https://doi.org/10.1109/ICCAE56788.2023.10111467, 2023.

[3] S. Shekhar and J. Aryal, "Role of geospatial technology in understanding urban green space of Kalaburagi city for sustainable planning," *Urban Forestry and Urban Greening,* vol. 46, p. 126450; https://doi.org/10.1016/j.ufug.2019.126450, 2019.

[4] M. Atasoy, "Monitoring the urban green spaces and landscape fragmentation using remote sensing: a case study in Osmaniye, Turkey," *Environmental Monitoring and Assessment,* vol. 190, no. 12, pp. 713; https://doi.org/10.1007/s10661-018-7109-1, 2018.

[5] S. Dinda, N. Das Chatterjee and S. Ghosh, "An integrated simulation approach to the assessment of urban growth pattern and loss in urban green space in Kolkata, India: A GIS-based analysis," *Ecological Indicators,* vol. 121, p. 107178; https://doi.org/10.1016/j.ecolind.2020.107178, 2021.

[6] A. R. Shahtahmassebi, C. Li, Y. Fan, Y. Wu, Y. Lin, M. Gan, K. Wang, A. Malik and G. A. Blackburn, "Remote sesning of urban green spaces: a review," *Urban Forestry and Urban Greening,* vol. 57, p. 126946; https://doi.org/10.1016/j.ufug.2020.126946, 2021.

[7] Z. Yin, W. Kuang, Y. Bao, Y. Dou, W. Chi, F. U. Ochege and T. Pan, "Evaluating the Dynamic Changes of Urban Land and Its Fractional Covers in Africa from 2000-2020 Using Time Series of Remotely Sensed Images on the Big Data Platform," *Remote Sensing,* vol. 13, p. 4288; https://doi.org/ 10.3390/rs13214288, 2021.

[8] I. Bhayunagiri and M. Saifulloh, "Urban footprint extraction derived from worldview-w satellite imagery by random forest and k-nearest neighbours algorithm," *IOP Conference Series: Earth and Environmental Science,* vol. 1200, pp. 012043; https://doi.org/10.1088/1755-1315/1200/1/012043, 2023.

[9] Q. Chen, C. Zhong, C. Jing, Y. Li, B. Cao and Q. Cheng, "Rapid Mapping and Annual Dynamic Evaluation of Quality of Urban Green Spaces on Google Earth Engine," *ISPRS International Journal of Geo-Information,* vol. 10, no. 10, p. 670; https://doi.org/10.3390/ijgi10100670, 2021.

[10] Baguio City Public Information Office, "Mayor Talks on Baguio's Smart City VIsion During ICT Conference," 16 March 2024. [Online]. Available: https://new.baguio.gov.ph/news/mayor-talks-on-baguios-smart-city-vision-during-ict-conference.

[11] Herald Express, "Baguio City aims to Transform to a “Smart City” by 2027," 21 April 2024. [Online]. Available: https://baguioheraldexpressonline.com/baguio-city-aims-to-transform-to-a-smart-city-by-2027/#:~:text= BAGUIO%20CITY%20%E2%80%93%20The%20local%20government,and%20sustainable%20quality%20of %20life..

[12] K. Gupta, K. Puntambekar, A. Roy, K. Pandey and P. Kumar, Smart Environment Through Smart Tools and Technologies for Urban Green Spaces. Smart Environment for Smart Cities, Springer, 2020, pp. 149-194.

[13] I. Ismayilova and S. Timpf, "Classifying Urban Green Spaces using a combined Sentinel-2 and Random Forest approach," *Agile: GIScience Series,* vol. 3, pp. 1-6; https://doi.org/10.5194/agile-giss-3-38-2022, 2022.

[14] Q. Duan, M. Tan, Y. Guo, X. Wang and L. Xin, "Understanding the Spatial Distribution of Urban Forests in China Using Sentinel-2 Images with Google Earth Engine," *Forests,* vol. 10, no. 9, p. 729; https://doi.org/ 10.3390/f10090729, 2019.

[15] J. Xiao, L. Chen, T. Zhang, L. Li, Z. Yu, R. Wu, L. Bai, J. Xiao and L. Chen, "Identification of Urban Green Space Types and Estimation of Above-Ground Biomass Using Sentinel-1 and Sentinel-2 Data," *Forests,* vol. 13, no. 7, p. 1077; https://doi.org/10.3390/f13071077, 2022.

[16] Y. Jin, X. Liu, Y. Chen and X. Liang, "Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong," *International Journal of Remote Sensing,* vol. 39, no. 23, pp. 8703-8723; https://doi.org/10.1080/01431161.2018.1490976, 2018.

[17] N. Kranjčić, D. Medak, R. Župan and M. Rezo, "Machine Learning Methods for Classification of the Green Infrastructure in City Areas," *ISPRS International Journal of Geo-Information,* vol. 8, no. 10, p. 463; https://doi.org/10.3390/ijgi8100463, 2019.

[18] D. Dobrinić, M. Gašparović and D. Medak, "Sentinel-1 and 2 Time-Series for Vegetation Mapping Using Random Forest Classification: A Case Study of Northern Croatia," *Remote Sensing,* vol. 13, no. 12, p. 2321; https://doi.org/10.3390/rs13122321, 2021.

[19] I. E. Ruiz Hernandez and W. Shi, "A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis," *International Journal of Remote Sensing,* vol. 39, no. 4, pp. 1175-1198; https://doi.org/10.1080/01431161.2017.1395968, 2017.

[20] M. Gašparović and D. Dobrinić, "Comparative Assessment of Machine Learning Methods for Urban Vegetation Mapping Using Multitemporal Sentinel-1 Imagery," *Remote Sensing,* vol. 12, no. 12, p. 1952; https://doi.org/ 10.3390/rs12121952, 2020.

[21] Z. Huang, H. Qi, C. Kang, Y. Su and Y. Liu, "An Ensemble Learning Approach for Urban Land Use Mapping Based on Remote Sensing Imagery and Social Sensing Data," *Remote Sensing,* vol. 12, no. 19, p. 3254; https://doi.org/10.3390/rs12193254, 2020.

[22] T. Adugna, W. Xu and J. Fan, "Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images," *Remote Sensing,* vol. 14, no. 3, p. 574; https://doi.org/10.3390/rs14030574, 2022.

[23] W. Mao, D. Lu, L. Hou and W. Yue, "Comparison of Machine-Learning Methods for Urban Land-Use Mapping in Hangzhou City, China," *Remote Sensing,* vol. 12, no. 17, p. 2817; https://doi.org/10.3390/rs12172817, 2020.

[24] T. Zhang, J. Su, Z. Xu, Y. Luo and J. Li, "Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier," *Applied Sciences,* vol. 11, no. 2, p. 543; https://doi.org/10.3390/ app11020543, 2021.

[25] I. Khosravi and S. K. Alavipanah, "A random forest-based framework for crop mapping using temporal, spectral, textural and polametric observations," *International Journal of Remote Sensing,* vol. 40, no. 18, pp. 7221-7251; https://doi.org/10.1080/01431161.2019.1601285, 2019.

[26] J. Svoboda, P. Stych, J. Lastovicka, D. Paluba and N. Kobliuk, "Random Forest Classification of Land Use, Land Use Change and Forestry (LULUCF) Using Sentinel-2 Data - A Case Study of Czechia," *Remote Sensing,* vol. 5, no. 14, p. 1189; https://doi.org/ 10.3390/rs14051189, 2022.

[27] S. Amini, M. Saber, H. Rabiei-Dastjerdi and S. Homayouni, "Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series," *Remote Sensing,* vol. 11, no. 14, p. 2654, 2022.

[28] Y. Cai, H. Lin and M. Zhang, "Mapping paddy rice by the object-based random forest method using time series Sentinel-1/ Sentinel-2 data," *Advances in Space Research,* vol. 64, no. 11, pp. 2233-2244; https://doi.org/10. 1016/j.asr.2019.08.042, 2019.

[29] "Baguio Geographic Coordinates," August 2024. [Online]. Available: https://www.geodatos.net/en/ coordinates/philippines/baguio.

[30] "PhilAtlas Baguio," August 2024. [Online]. Available: https://www.philatlas.com/luzon/car/baguio.html.

[31] DTI, "2024 Rankings of Highly Urbanized Cities," August 2024. [Online]. Available: https://cmci.dti.gov.ph/rankings-data.php?unit=Highly%20Urbanized%20Cities.

[32] Philippine Statistics Authority, "Urban Population of the Cordillera Administrative Region (2020 Census of Population and Housing)," September 2022. [Online]. Available: https://rssocar.psa.gov.ph/population-statistics/node/56230.

[33] R. Estoque and Y. Murayama, "Examining the potential impact of land use/cover changes on the ecosystem services of Baguio city, the Philippines: A scenario-based analysis," *Applied Geography,* no. 35, p. 316–326. 10.1016/j.apgeog.2012.08.006., 2012.

[34] European Union, European Space Agency, & Copernicus, *Harmonized Sentinel-2 MSI: Multispectral Instrument Level 1C [Data set]. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED#description,* 2022.

[35] N. Ghasemian and M. Akhoondzadeh, "Introducing two Random Forest based methods for cloud detection in remote sensing images," *Advances in Space Research,* vol. 62, no. 2, ISSN 0273-1177, pp. 288-303; https://doi.org/10.1016/j.asr.2018.04.030, 2018.

[36] J. Aryal, C. Sitaula and S. Aryal, "NDVI Threshold-Based Urban Green Space Mapping from Sentinel-2A at the Local Governmental Area (LGA) Level of Victoria, Australia," *Land,* vol. 11, no. 3, p. 351; https://doi.org/10.3390/land11030351, 2022.

[37] N. Colantino, "Insights in heat islands at the regional scalse using a data-driven approach," *City and Environment Interactions,* vol. 20, no. ISSN 2590-2520, p. 100124; https://doi.org/10.1016/j.cacint.2023.100124., 2023.

[38] R. R. Fern, E. A. Foxley and M. L. Morrison, "Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland," *Ecological Indicators,* vol. 94 Part 1, no. ISSN 1470-160X, pp. 16-21; https://doi.org/10.1016/j.ecolind.2018.06.029., 2018.

[39] M. Muhaimin, D. Fitriani, S. Adyatma and D. Arisanty, "Mapping build-up area density using normalized difference build-up index (ndbi) and urban index (ui) wetland in the city banjarmasin," *IOP Conference Series: Earth and Environmental Science,* vol. 1089, no. 1, pp. 012036; https://dx.doi.org/10.1088/1755-1315/1089/1/012036, November 2022.

[40] L. Nuraini, A. Nugraha, R. Yanti and L. Janah, "Comparison Normalized Dryness Built-Up Index (NDBI) with Enhanced Built-Up and Bareness Index (EBBI) for Identification Urban in Buleleng Sub-District," *Media Komunikasi FPIPS,* p. 10330; https://api.semanticscholar.org/CorpusID:249910330, 2022.

[41] T. A. Kebede, B. T. Hailu and K. V. Suryabhagavan, "Evaluation of spectral built-up indices for impervious surface extraction using sentinel 2A MSI iamgeries: A case of Addis Ababa City Ethiopia," *Environmental Challenges,* vol. 8, no. ISSN 2667-0100, p. 100568; https://doi.org/10.1016/j.envc.2022.100568., 2022.

[42] M. Jansson and Randrup, Thomas B, "Urban open space governance and management,"; New York, NY: Routledge, 2020.

[43] A. H. Qureshi, W S. Alaloul, and M. A. Musarat, "Sustainable development and urban design,"; *Encyclopedia of Renewable Energy, Sustainability and the Environment,* vol. 1, pp. 375-384; https://doi.org/10.1016/B978-0-323-93940-9.00066-9, 2024.