KADUNA JOURNAL OF GEOGRAPHY
2026, VOL. 8, No. 1, 565-578
https://doi.org/10.47514/kjg.2026.08.01.055

Research Article

Comparative Assessment of Machine Learning Classifiers for Land Use/Land Cover Mapping in a Tropical Savanna Landscape

Yasin Agono Awwal [a], Muhammad Isma'il [b], Sulaiman Lawan Ya'u [b], Lemuel Musa Maniyunda [b], Hussain Abba Ahmad [bs]

[a] Federal University Wukari, Nigeria. [b] Ahmadu Bello University, Nigeria

ABSTRACT

Accurate land use/land cover (LULC) information is essential for environmental monitoring, agricultural planning, and sustainable land management, particularly in rapidly changing tropical savanna landscapes. This study comparatively assessed the performance of Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Tree (CART) classifiers for LULC mapping in Sabon Gari and Zaria, Kaduna State, Nigeria, using Sentinel-2 spectral bands and derived spectral indices. Sentinel-2 imagery acquired between November 2024 and March 2025 was processed to generate ten spectral bands and four indices (NDVI, NDWI, NDBI, and BSI), which served as predictor variables. Classification was performed for six land cover classes, namely built-up areas, savanna vegetation, dense vegetation, farmland, water bodies, and rock outcrops. Model performance was evaluated using spatially independent validation data and confusion matrix-based metrics. The results revealed notable differences among the classifiers. RF achieved the highest classification accuracy, with an Overall Accuracy of 91.13% and a Kappa coefficient of 0.894, followed by CART (84.62%; κ = 0.832) and SVM (82.89%; κ = 0.809). Savanna vegetation (42.2%) and farmland (35.3%) were identified as the dominant land cover classes in the area, based on the RF output. Feature importance analysis showed that shortwave infrared and red-edge bands contributed most strongly to class discrimination, while derived spectral indices provided additional discriminatory power. Visual assessment further demonstrated that RF produced more spatially coherent and realistic land cover patterns than SVM and CART. The findings highlight the effectiveness of Sentinel-2 data for LULC mapping and indicate that RF performed most consistently among the evaluated classifiers in the heterogeneous tropical savanna environment of the study area.

ARTICLE HISTORY
Submitted 31 June 2026
Accepted 14 August 2026
Published 18 August 2026

GUEST EDITOR
A. M. Ahmed

KEYWORDS
Google Earth Engine; Machine learning; Polygon-Based Validation; Random Forest; Sentinel-2; Land use/land cover classification; Tropical savanna

1 Introduction

Accurate land use/land cover (LULC) information is fundamental for environmental monitoring, urban planning, agricultural management, and sustainable resource utilization (Gaur & Singh, 2023). Reliable LULC data support informed decision-making by providing insights into landscape dynamics, ecosystem health, and anthropogenic pressures on natural resources. This need has become increasingly important in tropical savanna regions, where rapid population growth, agricultural expansion, urbanization, and environmental degradation have accelerated landscape transformation over recent decades (Seyam et al., 2023; Awwal et al., 2025a). Consequently, timely and accurate mapping of land cover patterns has become essential for sustainable land management and environmental planning.

The increasing demand for spatially explicit LULC information has coincided with significant advances in satellite remote sensing technologies. Remote sensing offers a cost-effective means of monitoring large and often inaccessible areas while providing repeated observations over time (Seyam et al., 2023). Among the currently available Earth observation platforms, the Sentinel-2 mission has attracted considerable attention owing to its high spatial resolution, short revisit interval, and inclusion of red-edge and shortwave infrared bands that enhance land cover discrimination capabilities (Xiao et al., 2020). As noted by Phiri et al. (2020), the spectral richness of Sentinel-2 imagery has substantially improved the accuracy of LULC classification, particularly in heterogeneous environments where different land cover classes exhibit subtle spectral variations.

While the availability of high-quality satellite imagery has expanded opportunities for LULC mapping, the effectiveness of such efforts largely depends on the classification methods employed. Traditional classification approaches often struggle to capture the complex and nonlinear relationships that characterize remotely sensed data (Awwal et al., 2025a). In response, machine learning (ML) algorithms have emerged as powerful alternatives because of their ability to model intricate patterns within high-dimensional datasets. Among the most widely applied ML techniques are Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Tree (CART), each

Corresponding Author: Awwal, Y. A. awwalyasin313@gmail.com Federal University Wukari, Taraba State, Nigeria.

possessing distinct theoretical foundations and classification strategies. Previous studies have reported strong performance of RF in diverse landscapes (Jodhani et al., 2025; Awwal et al., 2025b), while SVM has demonstrated effectiveness in handling complex decision boundaries and high-dimensional feature spaces. Similarly, CART remains attractive because of its computational efficiency, interpretability, and ease of implementation (Tesfaye et al., 2024). However, despite their widespread use, the relative performance of these classifiers often varies according to landscape characteristics, spectral complexity, class heterogeneity, and training data structure (Arpitha et al., 2023). As a result, classifier performance cannot be generalized across environments and requires evaluation under specific ecological and geographical conditions.

This issue is particularly relevant in tropical savanna ecosystems, where the coexistence of cultivated lands, natural vegetation, built-up areas, exposed soils, and rock outcrops creates substantial spectral overlap among land cover classes (Awwal et al., 2025a). Although numerous LULC studies have been conducted in Nigeria and other tropical regions, comparatively few have systematically compared multiple ML classifiers using harmonized Sentinel-2 spectral datasets within the tropical savanna landscapes of northwestern Nigeria. Moreover, many existing studies such as Zhao et al. (2024) and Awwal et al. (2025a) employ pixel-level random validation procedures that may overestimate classification accuracy due to spatial autocorrelation between training and validation samples (Awwal et al., 2026). Spatially independent validation approaches based on polygon separation provide a more rigorous assessment of classifier generalization capability, yet their application remains relatively limited in regional LULC mapping studies.

The present study conducts a comparative assessment of RF, SVM, and CART classifiers for LULC mapping in the tropical savanna landscape of Sabon Gari and Zaria, Kaduna State, Nigeria, using Sentinel-2 spectral bands and derived spectral indices. Unlike many previous investigations, the study incorporates polygon-based holdout validation to ensure spatial independence between training and validation datasets. The study therefore contributes both methodological and practical insights by evaluating classifier suitability under savanna conditions while providing a robust assessment framework for regional LULC mapping. Specifically, the objectives were to (i) generate Sentinel-2-based LULC maps using RF, SVM, and CART classifiers, (ii) evaluate and compare classifier performance using spatially independent validation metrics, and (iii) determine the relative importance of spectral variables for LULC discrimination within the study area.

2 Materials and Methods

2.1 Study Area

The study was conducted in Sabon Gari and Zaria Local Government Areas of Kaduna State, northwestern Nigeria (Figure 1). The area lies within the Northern Guinea Savanna ecological zone and is characterized by a tropical continental climate with distinct wet and dry seasons (Abdulkadir et al., 2021). The landscape comprises a mosaic of agricultural land, built-up areas, natural vegetation, bare surfaces, rock outcrops, and water bodies, making it suitable for evaluating the performance of ML algorithms for LULC classification.

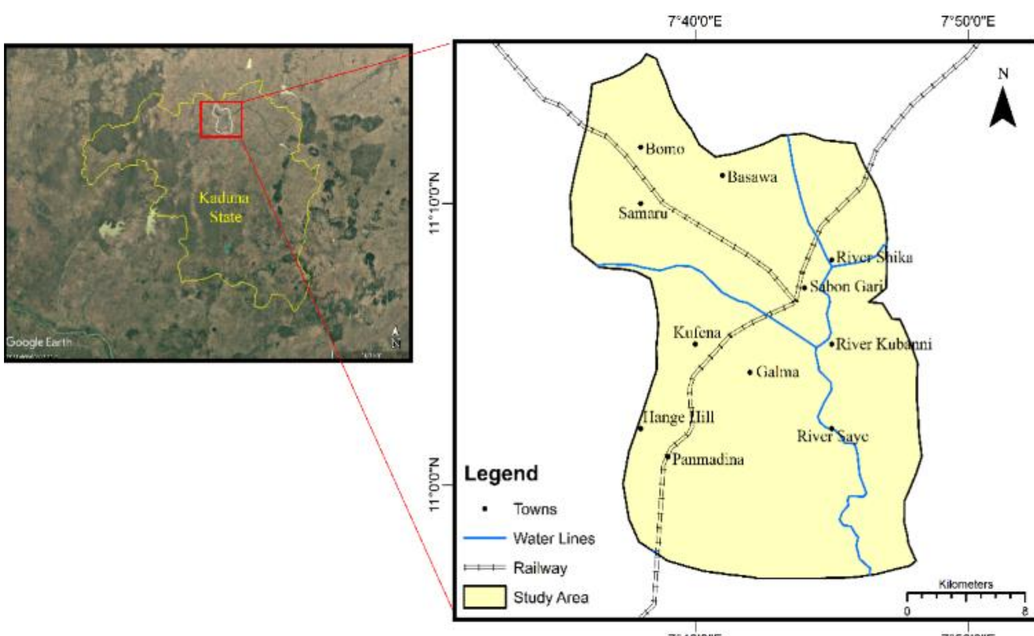

Figure 1: Map of Kaduna State Showing Location of Study Area

2.2 Satellite Data Acquisition and Pre-processing

Sentinel-2 Surface Reflectance Harmonized imagery was acquired from the Copernicus programme through the Google Earth Engine (GEE) platform. Images covering the period from November 2024 to March 2025 were selected to ensure adequate temporal representation while minimizing seasonal data gaps. Only scenes with cloud cover less than 10% were retained for analysis. All images were radiometrically scaled by dividing digital numbers by 10,000 to obtain surface reflectance values. A median composite image was subsequently generated to reduce residual cloud contamination and temporal noise. The composite image was clipped to the study area boundary and used for all classification analyses.

2.3 Spectral Feature Extraction

Land cover classification was based exclusively on spectral information derived from Sentinel-2 imagery. Ten spectral bands representing the visible, red-edge, near-infrared, and shortwave infrared regions were selected, and to enhance class separability, widely used spectral indices such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Wetness Index (NDWI), and Bare Soil Index (BSI) were computed. Table 1 summarizes variable sources utilized for the LULC classification. The final predictor dataset therefore consisted of fourteen variables comprising ten spectral bands and four spectral indices.

Table 1: Spectral predictor variables used for land use/land cover classification

Variable	Description	Band(s)/Formula	Resolution (m)	Source
Blue	Visible blue reflectance	Band 2 (490 nm)	10	Sentinel-2 MSI
Green	Visible green reflectance	Band 3 (560 nm)	10	Sentinel-2 MSI
Red	Visible red reflectance	Band 4 (665 nm)	10	Sentinel-2 MSI
Red 1	Red-edge reflectance	Band 5 (705 nm)	20	Sentinel-2 MSI
Red 2	Red-edge reflectance	Band 6 (740 nm)	20	Sentinel-2 MSI
Red 3	Red-edge reflectance	Band 7 (783 nm)	20	Sentinel-2 MSI
NIR	Near-infrared reflectance	Band 8 (842 nm)	10	Sentinel-2 MSI
Narrow NIR	Narrow near-infrared reflectance	Band 8A (865 nm)	20	Sentinel-2 MSI
SWIR 1	Shortwave infrared reflectance	Band 11 (1610 nm)	20	Sentinel-2 MSI
SWIR 2	Shortwave infrared reflectance	Band 12 (2190 nm)	20	Sentinel-2 MSI
NDVI	Normalized Difference Vegetation Index	$\frac{(B8-B4)}{(B8+B4)}$	10	Derived from Sentinel-2
NDWI	Normalized Difference Water Index	$\frac{(B3-B8)}{(B3+B8)}$	10	Derived from Sentinel-2
BSI	Bare Soil Index	$\frac{(B11+B4)-(B8+B2)}{(B11+B4)+(B8+B2)}$	10	Derived from Sentinel-2

Note: *B2, 3, 4, 8, 11 refer to Band 2, 3, 4, 8, and 11*

2.4 Reference Data Collection and Class Definition

Reference samples representing major land use and land cover categories were digitized using high-resolution imagery and local knowledge of the study area. For the purpose of this study, six LULC classes were defined for the classification process, namely built-up, savanna grassland, dense vegetation, farmland, water bodies, and wasteland or rock outcrops. Reference samples were converted into polygon features and assigned class labels. Spectral values corresponding to each sample point were extracted from the Sentinel-2 composite image at a spatial resolution of 10 m.

2.5 Polygon-Based Spatial Training and Validation Design

To ensure spatial independence between training and validation data and to minimize spatial autocorrelation effects, an object-based holdout validation strategy was adopted. All reference polygons were assigned class labels and then randomly partitioned at the polygon level using a 70:30 split. Specifically, 70% of polygons from each class were allocated to model training, while the remaining 30% were reserved exclusively for independent validation. This design ensured that all pixels extracted from a given polygon belonged to either the training or validation dataset, but never both, thereby eliminating spectral and spatial leakage between datasets (Kılıç, 2026).

2.6 Machine Learning Classification and LULC Map Generation

Three supervised machine learning algorithms were evaluated for LULC mapping, viz; Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Tree (CART). Random Forest is an ensemble learning algorithm that constructs multiple decision trees and combines their outputs through majority voting. The algorithm is known for its robustness to noise, ability to handle high-dimensional datasets, and resistance to overfitting (Breiman, 2001). In this study, the RF model was implemented using 100 trees within the GEE environment. Support Vector Machine is a non-parametric classification algorithm that identifies optimal decision boundaries between classes in a multidimensional feature space (Salleh et al., 2024). A radial basis function (RBF) kernel was employed with a gamma value of 0.5 and a cost parameter of 10. Lastly, the CART algorithm constructs a binary decision tree by recursively partitioning the predictor space based on impurity reduction criteria. CART provides an interpretable classification framework and serves as a useful benchmark against more complex ML approaches (Zafar et al., 2024). To ensure a controlled comparison among the classifiers, hyperparameter tuning was not performed; instead, fixed parameter settings were specified a priori for each algorithm. All classifiers were trained using the same predictor variables and training dataset. Each trained classifier was applied to the Sentinel-2 composite image to generate a classified LULC map of the study area. Classification outputs were produced at a spatial resolution of 10 m and exported for subsequent analysis and visualization.

2.7 Accuracy Assessment

The performance of the machine learning classifiers was evaluated using an independent validation dataset derived from polygons excluded during model training. Classification accuracy was assessed using confusion matrices generated for each classifier. The metrics employed included Kappa coefficient, Overall Accuracy (OA), User's Accuracy (UA), Producer's Accuracy (PA), and F1-score.

Overall Accuracy, which represents the proportion of correctly classified samples relative to the total number of validation samples, was computed as:

$$Overall\ accuracy = \frac{\sum_{i=1}^{n} T_{ii}}{N} \quad \text{Eq. (1)}$$

where T_{ii} denotes the number of correctly classified samples belonging to class i, n is the total number of classes, and N represents the total number of validation samples.

Producer's Accuracy (PA), which measures the probability that a reference sample is correctly classified (i.e., the ability of the classifier to correctly map actual ground truth samples), was computed as:

$$Producer's\ Accuracy = \frac{T_{ii}}{\sum_{j=1}^{n} T_{ij}} \quad \text{Eq. (2)}$$

where T_{ii} represents correctly classified samples in class i, while $\sum_{j=1}^{n} T_{ij}$ denotes the total number of reference samples belonging to class i.

User's Accuracy (UA), which indicates the probability that a sample classified into a given category actually represents that category on the ground, was computed as:

$$User's\ Accuracy = \frac{T_{ii}}{\sum_{j=1}^{n} T_{ji}} \quad \text{Eq. (3)}$$

where T_{ii} is the number of correctly classified samples in class i, and $\sum_{j=1}^{n} T_{ji}$ represents the total number of samples classified into class i.

The F1-score, which provides a harmonic mean of Producer's Accuracy (Recall) and User's Accuracy (Precision), was used to provide a balanced assessment of classification performance. It was computed as:

$$F1-score = 2 \times \frac{PA \times UA}{PA + UA} \quad \text{Eq. (4)}$$

Cohen's Kappa coefficient (κ) was employed to evaluate the degree of agreement between classified and reference pixels after accounting for agreement occurring by chance. It was estimated as:

$$k = \frac{p_0 - p_e}{1 - p_e} \quad \text{Eq. (5)}$$

where p_0 represents the observed agreement, equivalent to Overall Accuracy, while p_e denotes the probability of chance agreement, computed as:

$$p_e = \sum_{i=1}^{n} \left(\frac{(\sum_{j=1}^{n} T_{ij})(\sum_{j=1}^{n} T_{ji})}{N^2} \right) \quad \text{Eq. (6)}$$

Here, $(\sum_{j=1}^{n} T_{ij})$ and $(\sum_{j=1}^{n} T_{ji})$ correspond to the row and column totals, respectively, for class i, and N is the total number of validation samples. Higher κ values indicate stronger agreement between predicted and reference classifications, with values approaching 1.0 reflecting near-perfect classification performance.

2.8 Variable Importance Analysis

To assess the contribution of individual spectral variables, feature importance scores were extracted from the RF model. The importance values were used to identify the most influential spectral bands and indices for discriminating LULC classes within the tropical savanna environment.

3 Results and Discussion

3.1 Spectral Index Characteristics

The spatial distribution of the spectral indices derived from Sentinel-2 imagery is presented in Figure 2(a-d). The Normalized Difference Vegetation Index (NDVI) ranged from -0.66 to 0.86, indicating substantial variation in vegetation cover across the study area. Higher NDVI values were concentrated in areas characterized by dense vegetation and cultivated lands, whereas lower values were associated with built-up surfaces, exposed soils, water bodies, and rock outcrops.

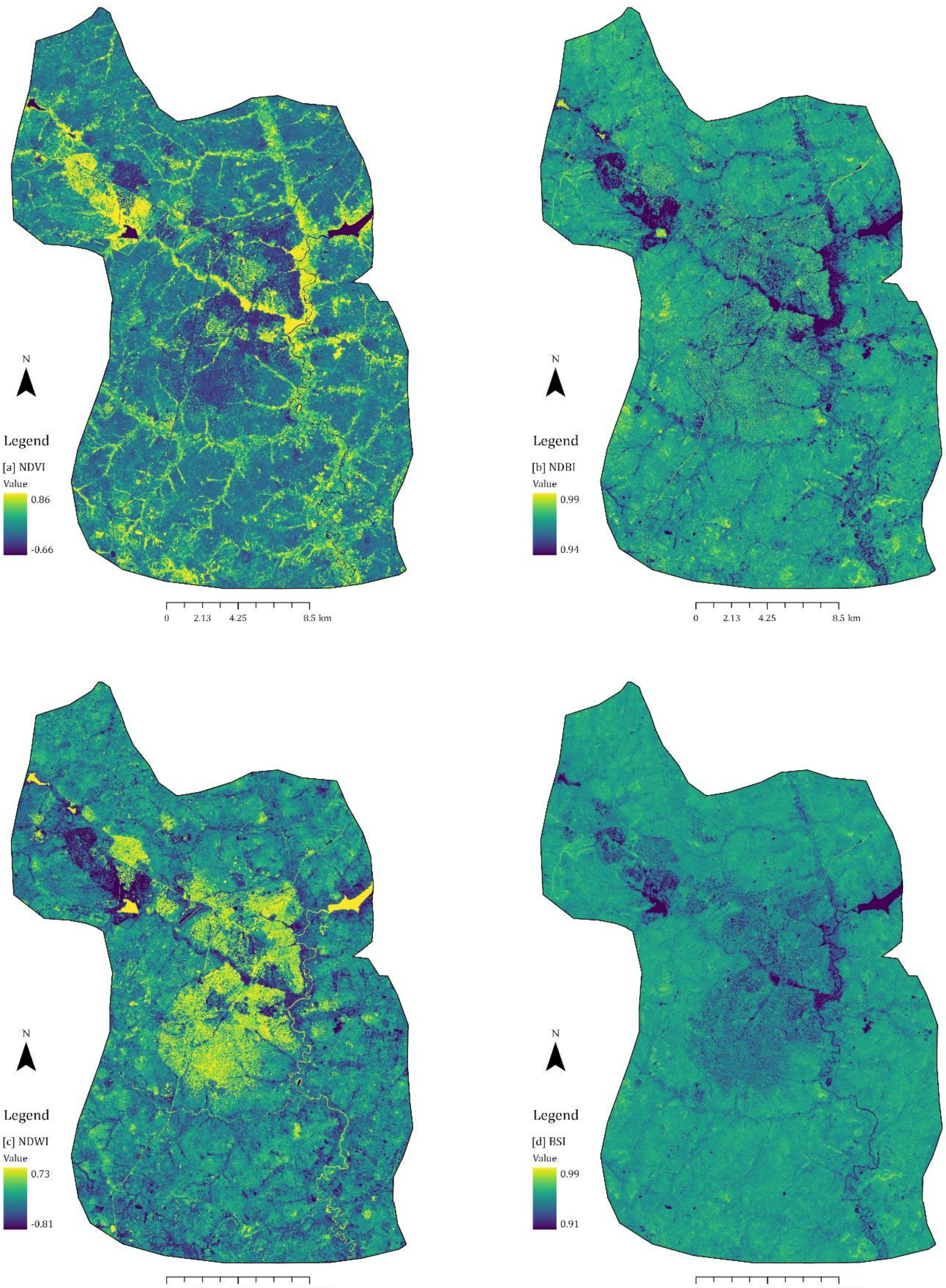

Figure 2: Spatial Distribution of Sentinel-2 Derived Spectral Indices (a) NDVI (b) NDBI (c) NDWI, and (d) BSI in the Study Area

The Normalized Difference Built-up Index (NDBI) exhibited values ranging from 0.94 to 0.99, reflecting the spatial distribution of urbanized and impervious surfaces. Normalized Difference Water Index (NDWI) varied between -0.81 and 0.73, with positive values delineating water bodies and moisture-rich environments, while negative values corresponded largely to terrestrial surfaces. The Bare Soil Index (BSI) ranged from 0.91 to 0.99, highlighting areas of exposed soil and sparsely vegetated surfaces within the landscape. Collectively, the spectral indices revealed considerable spatial heterogeneity, displaying complementary information capable of aiding discrimination between different LULC classes in the tropical savanna environment.

3.2 Land Use/Land Cover Maps from Individual Classifiers

The LULC maps generated using the CART, SVM, and RF classifiers are presented in Figure 3(a-c). Visual comparison of the classification outputs showed that all three algorithms successfully identified the major land use and land cover classes within the study area, including built-up areas, savanna vegetation, dense vegetation, farmland, water bodies, and rock outcrops. Although the overall spatial patterns were broadly similar among the classifiers, noticeable differences were observed in the delineation and extent of certain classes, particularly farmland, savanna vegetation, and rock outcrops (Figure 4).

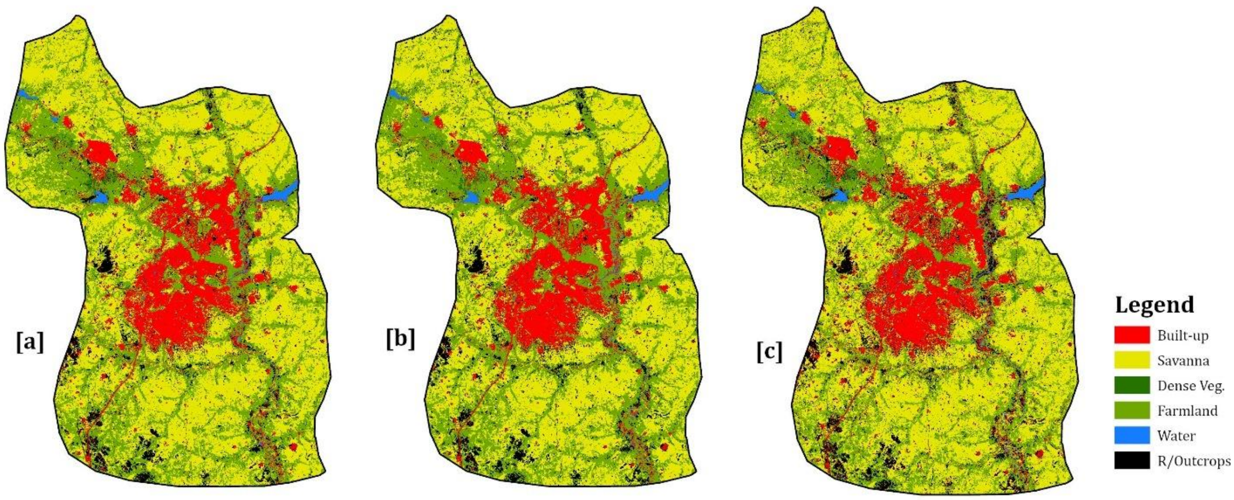

Figure 3: Land Use/Land Cover Maps Derived from CART, SVM, and Random Forest Classifiers in Sabon Gari–Zaria Area

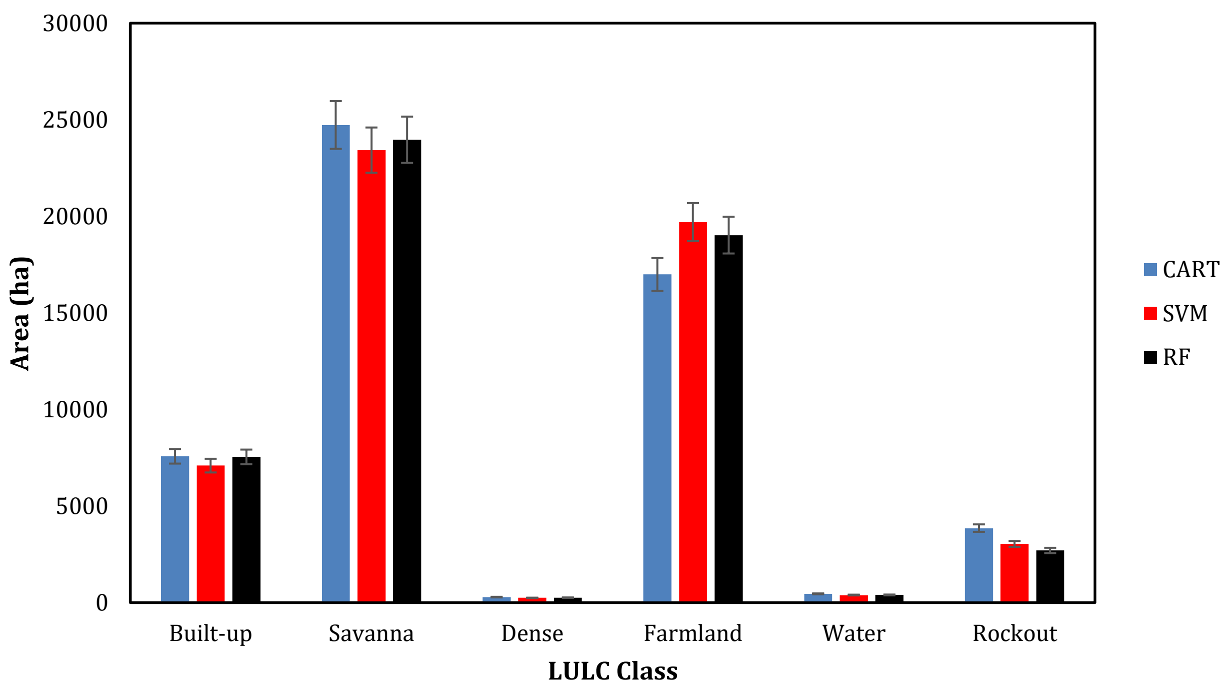

Figure 4: Comparative Area Statistics of Land Use/Land Cover Classes Derived from CART, SVM, and Random Forest Classifiers

The CART classifier produced estimates of 7,577.46 ha for built-up areas, 24,730.49 ha for savanna vegetation, 287.80 ha for dense vegetation, 16,991.89 ha for farmland, 457.95 ha for water bodies, and 3,859.04 ha for rock outcrops (Figure 4). The SVM classifier mapped 7,093.99 ha of built-up land, 23,429.62 ha of savanna vegetation, 245.88 ha of dense vegetation, 19,700.75 ha of farmland, 394.36 ha of water bodies, and 3,040.03 ha of rock outcrops. Similarly, the RF classifier estimated 7,546.30 ha of built-up land, 23,964.89 ha of savanna vegetation, 260.09 ha of dense vegetation, 19,028.90 ha of farmland, 401.21 ha of water bodies, and 2,703.24 ha of rock outcrops.

Across all classifiers, savanna vegetation constituted the dominant land cover type, occupying between 23,429.62 and 24,730.49 ha, followed by farmland, which ranged from 16,991.89 to 19,700.75 ha. Dense vegetation represented the least extensive class, accounting for less than 300 ha irrespective of the classifier used. Differences among the classifiers were most evident in the estimation of farmland and rock outcrop areas, whereas built-up areas, water bodies, and dense vegetation exhibited comparatively smaller variations.

3.3 Spatial Agreement Analysis

Spatial agreement between the RF classifier and the alternative classification approaches is illustrated in Figure 5. Comparison between RF and SVM (Figure 5a) revealed an agreement area of 45,775.25 ha and a disagreement area of 8,129.38 ha. In contrast, the RF–CART comparison yielded a larger agreement area of 48,964.57 ha and a smaller disagreement area of 4,940.06 ha (Figure 5b). These results indicate a higher level of spatial consistency between the RF and CART classification outputs than between RF and SVM.

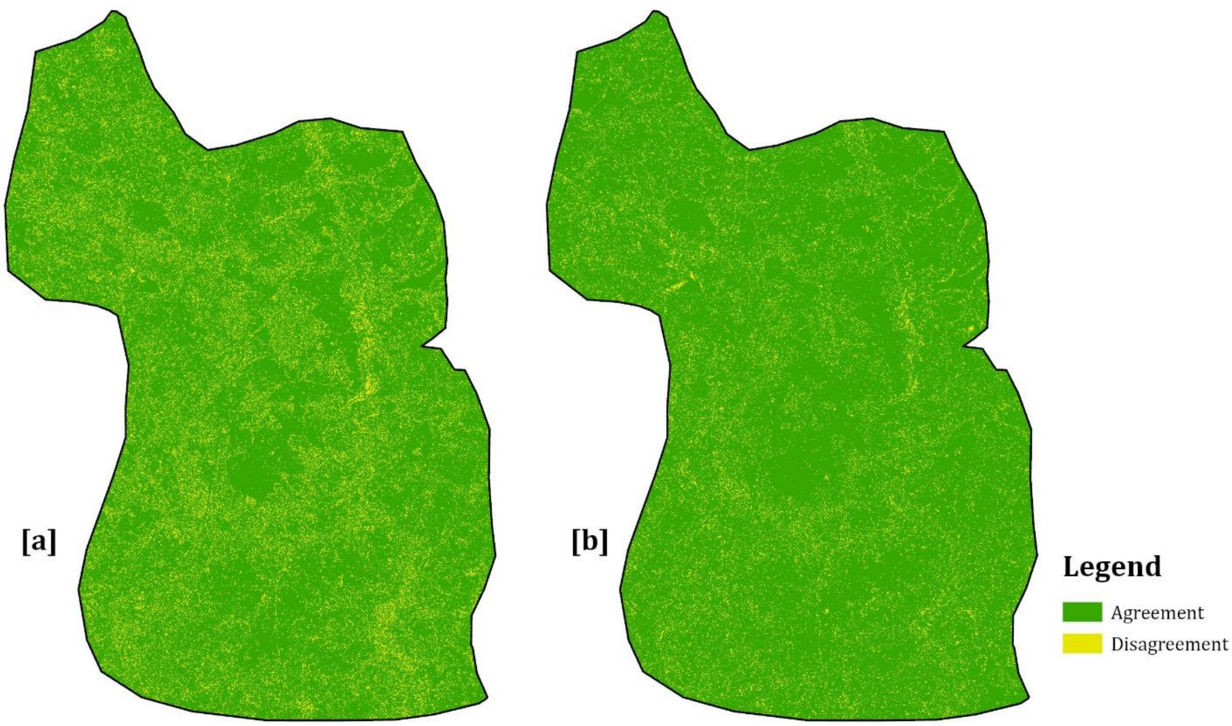

Figure 5: Spatial Agreement and Disagreement between RF and other Classifiers (SVM & CART)

Visual inspection of the agreement maps showed that disagreement was not concentrated in any single land-use or land-cover category. Instead, disagreement pixels occurred as dispersed patches throughout the study area, frequently appearing along class boundaries and transition zones. This pattern suggests that classification uncertainty was primarily associated with edge effects and spectrally mixed areas rather than systematic confusion involving a particular land cover class. However, the large extent of agreement relative to disagreement demonstrates a substantial degree of consistency among the machine learning classifiers in representing the spatial structure of the landscape.

3.4 Classification Accuracy Assessment

The classification accuracy metrics for the RF, SVM, and CART classifiers are presented in Table 2, while the comparative Overall Accuracy (OA) and Kappa coefficient (κ) values are illustrated in Figure 6. Considerable differences in classification performance were observed among the evaluated ML algorithms.

Table 2: Class-wise Accuracy Assessment of LULC Classification Using RF, SVM, and CART Models

LULC	RF			SVM			CART		
	PA	UA	F1	PA	UA	F1	PA	UA	F1
Built-up	92.58	93.25	92.91	88.33	88.66	88.50	88.16	88.66	88.41
Savanna	65.78	98.86	78.00	53.99	83.04	65.43	67.68	84.68	75.21
Dense Veg.	15.15	100.00	26.32	15.15	100.00	26.32	15.15	82.10	25.61
Farmland	99.85	85.48	91.87	96.78	81.83	88.68	98.10	84.72	90.92
Water	99.19	100.00	99.59	98.71	100.00	99.35	99.84	100.00	99.92
Rock Outcrops	99.47	99.47	99.47	97.34	96.71	97.02	97.61	97.21	97.41

Note: *RF – random forest; SVM – support vector machines; CART – classification and regression trees; PA – producer's accuracy (%); UA – user's accuracy (%); F1 – F1 score (%).*

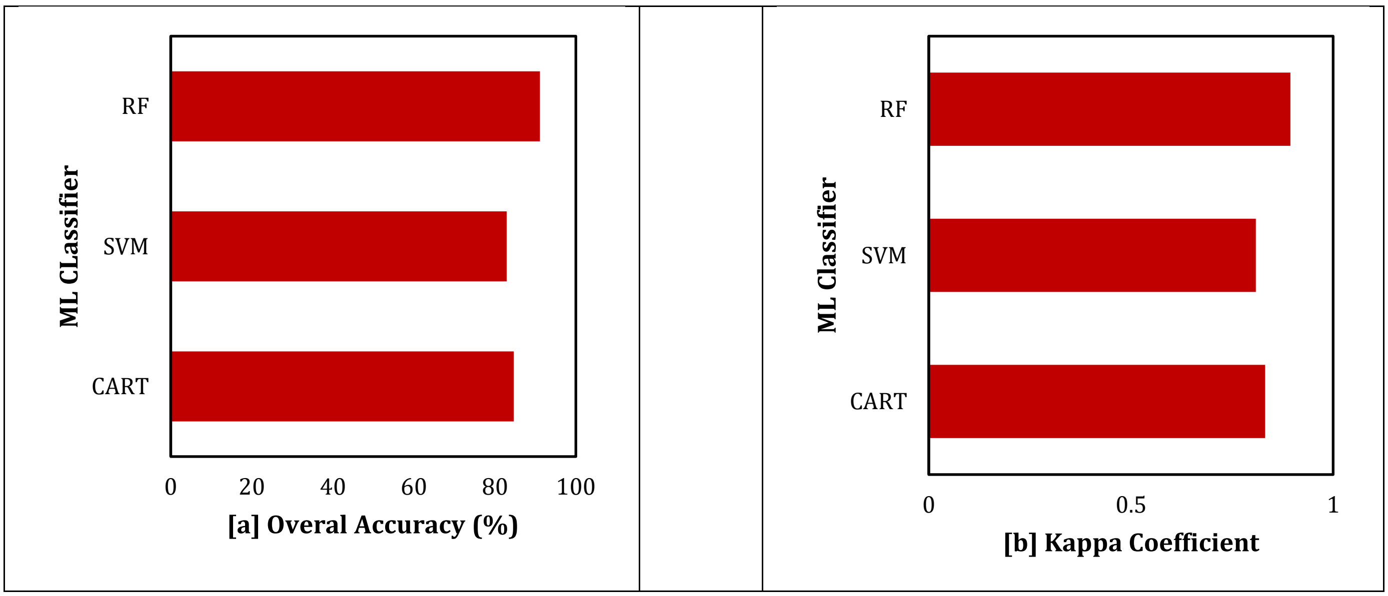

Figure 6: Comparison of (a) Overall Accuracy and (b) Kappa Coefficient for RF, SVM & CART

The RF classifier achieved the highest overall classification performance, producing an OA of 91.13% and a κ of 0.894 (Figure 6a-b). The CART classifier ranked second with an OA of 84.62% and a κ of 0.832, whereas the SVM classifier recorded the lowest OA (82.89%) and κ (0.809). At the class level, all classifiers performed exceptionally well in identifying water bodies and rock outcrops, with PA and UA values generally exceeding 97%. Farmland was also mapped with high reliability, particularly by the RF classifier, which achieved a PA of 99.85% and an F1-score of 91.87%. Built-up areas exhibited consistently high classification accuracies across all classifiers, although RF again produced the strongest performance. In contrast, Dense Vegetation represented the most challenging class, recording a PA of only 15.15% across all classifiers despite relatively high UA values. Savanna vegetation exhibited moderate classification performance, with PA values ranging from 53.99% for SVM to 67.68% for CART.

The results demonstrate the superior performance of the RF algorithm in discriminating LULC classes within the study area, as evidenced by its higher overall agreement statistics and stronger class-specific accuracy measures. Based on the accuracy assessment results, the RF classifier was selected as the optimal model for LULC mapping in the study area, outperforming both CART and SVM. Consequently, the RF-derived LULC map was considered the most reliable representation of the spatial distribution of land use and land cover classes within the study area (Figure 7).

A visual comparison of the classification outputs against high-resolution Google imagery in the vicinity of Shika Dam further supported the superiority of the RF classifier (Figure 8a-d). Although all three classifiers successfully identified the major land cover categories, the RF output exhibited greater spatial coherence and clearer delineation of landscape features. Class boundaries appeared more distinct, while the spatial arrangement of mapped features corresponded more closely with the reference imagery. In contrast, the SVM and CART outputs showed a comparatively fragmented classification pattern, with a greater occurrence of isolated pixels and localized inconsistencies along class boundaries. Differences among the classifiers were particularly evident in transitional areas where vegetation, farmland, and built-up surfaces occurred in proximity.

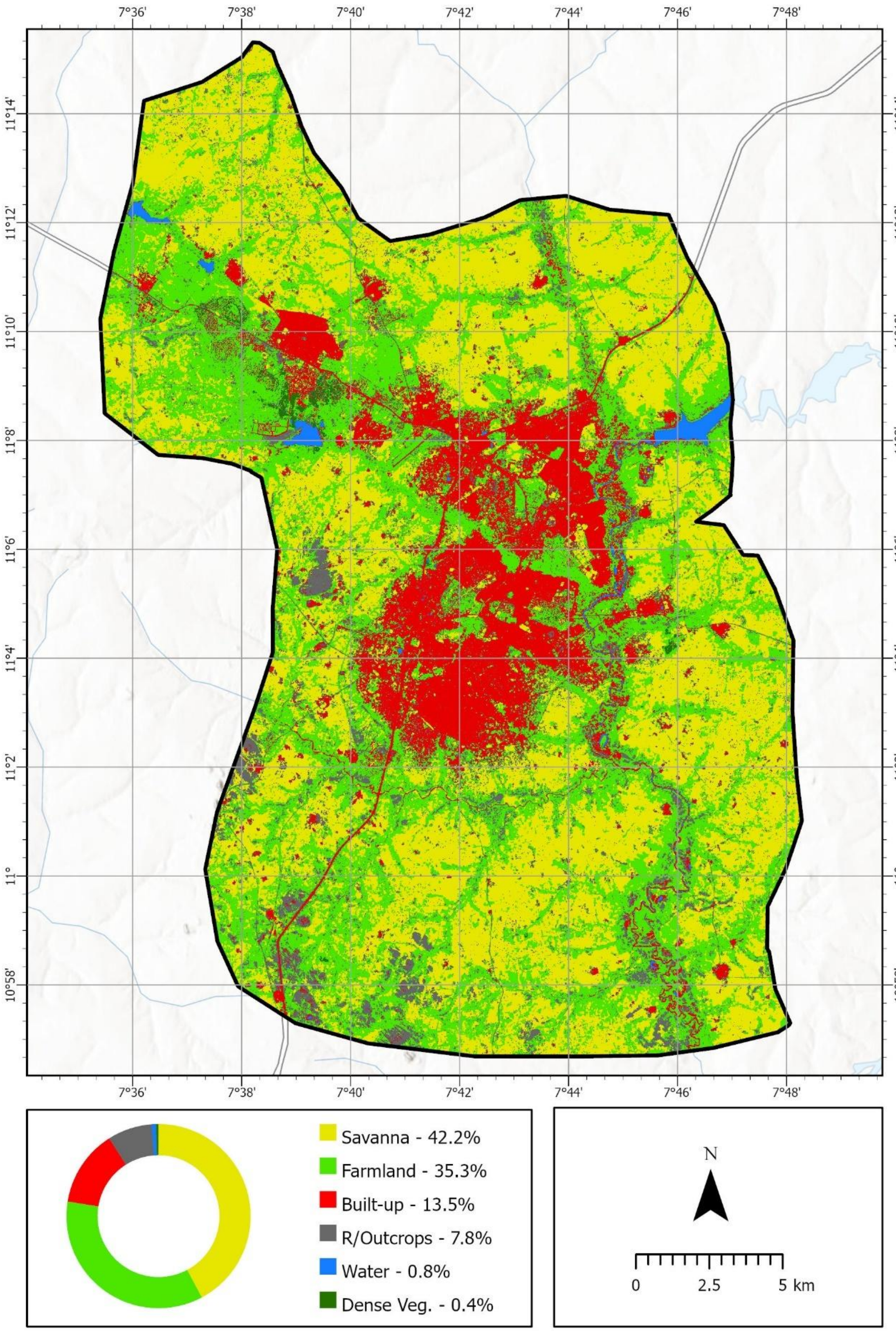

Figure 7: Final Land Use/Land Cover Map of the Study Area Derived from RF Classifier (2025)

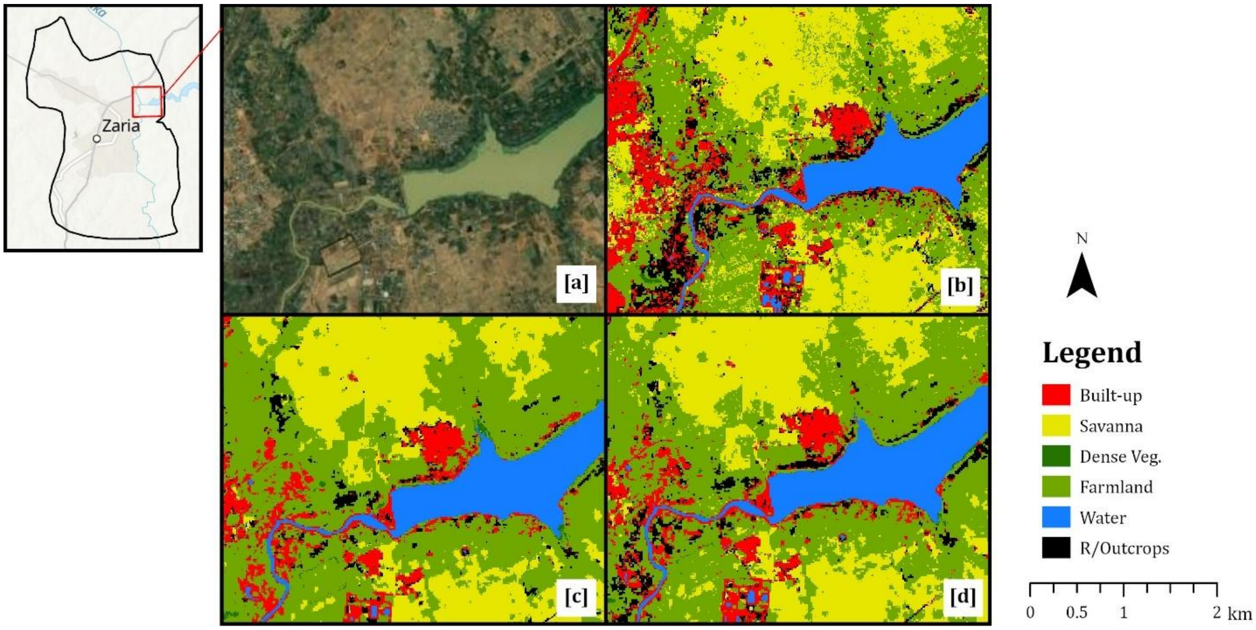

Figure 8: Spatial Comparison of Classifier Outputs with Google Earth Imagery in the Vicinity of Shika Dam showing (a) Google Imagery (b) RF Classification (c) SVM Classification (d) CART Classification

3.5 Feature Importance Analysis

The relative importance of predictor variables derived from the RF classifier is presented in Figure 9. The results revealed substantial differences in the contribution of spectral bands and indices to classification performance. The most influential variable was the shortwave infrared band B11, with an importance score of 327.42, followed by B12 (314.53), B8A (299.85), B5 (289.50), B6 (272.63), and B7 (268.30). Among the derived spectral indices, NDWI recorded the highest importance score (236.94), followed by BSI (223.12), NDVI (222.84), and NDBI (205.84). Although the spectral indices contributed meaningfully to classification performance, their importance scores were generally lower than those of the most influential spectral bands. The visible bands B2, B3, and B4 exhibited moderate importance, with values ranging from 207.58 to 229.75, while the broad near-infrared band B8 recorded the lowest importance score (198.36).

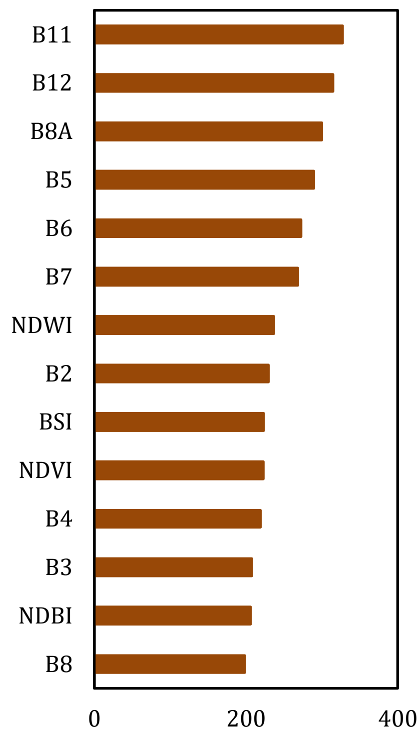

Figure 9: Random Forest Variable Importance Ranking for Sentinel-2 Spectral Bands and Derived Indices.

4 Discussions

The primary objective of this study was to evaluate the performance of three ML algorithms for LULC mapping in a tropical savanna environment using Sentinel-2 spectral information and derived indices. The results demonstrate that all three classifiers successfully captured the major land cover patterns of the Sabon Gari–Zaria landscape; however, substantial differences existed in their classification accuracy, spatial consistency, and representation of individual land cover classes. These findings confirm that classifier selection remains an important determinant of mapping quality, even when identical training data and predictor variables are employed.

4.1 Contribution of Spectral Variables to Land Cover Discrimination

The spectral index maps revealed considerable heterogeneity in vegetation condition, moisture status, built-up development, and bare surface exposure across the study area. Such variability is expected in tropical savanna ecosystems, where agricultural lands, natural vegetation, settlements, and exposed surfaces often occur in close spatial proximity (Koko et al., 2020). The observed NDVI range (-0.66 to 0.86) indicates the presence of both highly vegetated and sparsely vegetated surfaces, while the NDWI and BSI distributions highlight substantial contrasts in surface moisture and soil exposure. These spectral differences enhanced class separability and contributed to the generally high classification accuracies achieved by the models.

The feature importance analysis further demonstrates that classification performance was driven primarily by shortwave infrared (SWIR) and red-edge wavelengths. In particular, B11 and B12 emerged as the most influential variables, followed by B8A and the red-edge bands. This finding is consistent with previous studies that reported the superior capability of SWIR and red-edge regions for discriminating vegetation conditions, soil exposure, and moisture-related land cover characteristics in semi-arid and savanna environments (Xiao et al., 2020; Lin et al., 2020). According to Solgi et al. (2023), SWIR wavelengths are particularly sensitive to vegetation water content and soil moisture, making them effective for separating cropland, natural vegetation, and exposed surfaces. Similarly, the strong contribution of the red-edge bands may be attributed to their sensitivity to chlorophyll content and vegetation vigor, properties that are highly relevant for distinguishing savanna vegetation from cultivated fields and dense vegetation (Xiao et al., 2020).

Interestingly, the derived indices were less important than several of the original spectral bands. Although NDWI, BSI, and NDVI ranked among the more influential predictors, none surpassed the importance of B11, B12, or B8A. This pattern likely reflects the ability of RF to exploit the raw spectral information contained in individual bands directly, including nonlinear relationships and interactions among predictors, without requiring the information to be transformed into spectral indices. This suggests that while spectral indices enhance class discrimination by emphasizing specific land surface properties, substantial discriminatory information remained embedded within the original Sentinel-2 spectral bands. Similar observations have been reported by Tesfaye et al. (2024) and Zhao et al. (2024), who found that ML algorithms often extract greater predictive value from the complete spectral dataset than from a limited set of derived indices alone.

4.2 Differences in Land Cover Estimates Among Classifiers

Although the three classifiers produced broadly similar spatial patterns, notable differences were observed in the estimated extents of farmland, savanna vegetation, and rock outcrops. Such variations are not unexpected because each algorithm partitions feature space differently and responds uniquely to spectral overlap among classes (Zhao et al., 2024). Farmland exhibited the greatest variability, ranging from approximately 16,992 ha under CART to 19,701 ha under SVM. This disparity likely reflects the spectral complexity of agricultural landscapes within the study area, where cultivated fields may occur in different phenological stages and share spectral characteristics with surrounding savanna vegetation.

Similarly, the differences observed in rock outcrop estimates may be associated with spectral confusion between exposed rock surfaces, bare soils, and sparsely vegetated areas. In tropical savanna environments, these classes frequently exhibit overlapping reflectance characteristics, particularly during the dry season when vegetation cover is reduced (Obade & Lal, 2013). The relatively smaller differences observed for built-up areas and water bodies suggest that these classes possess more distinctive spectral signatures, allowing them to be consistently identified irrespective of classifier choice.

The dominance of savanna vegetation across all classification outputs is consistent with the ecological characteristics of the study area and aligns with previous land cover assessments conducted within the northern Guinea savanna zone of Nigeria (Akintuyi et al., 2021; Awwal et al., 2025a). Likewise, the limited extent of dense vegetation reflects the fragmented nature of closed-canopy vegetation in this predominantly agricultural and urbanizing landscape.

4.3 Spatial Agreement and Sources of Classification Uncertainty

The spatial agreement analysis provides additional insight into classifier behaviour beyond conventional accuracy statistics. Although RF achieved the highest overall performance, its agreement with CART (48,964.57 ha) exceeded its agreement with SVM (45,775.25 ha). This suggests that RF and CART produced more similar spatial representations of land cover despite differences in their overall accuracies. The dispersed nature of disagreement pixels is particularly noteworthy. Rather than being concentrated within specific land cover classes, disagreements occurred predominantly along class boundaries and transitional zones. Such patterns are characteristic of mixed-pixel effects and spatial heterogeneity, where a single pixel may contain multiple land cover components (He et al., 2023). Transition zones between farmland and savanna vegetation are especially susceptible to this phenomenon because vegetation cover often varies continuously rather than abruptly.

This finding indicates that classification uncertainty within the study area is primarily spatial rather than thematic. In other words, the classifiers generally agreed on the identity of major land cover classes but differed in the precise placement of boundaries. Similar patterns have been reported in savanna and agricultural mosaics where gradual ecological transitions create ambiguity in class assignment (Akintuyi et al., 2021).

4.4 Comparative Performance of RF, SVM, and CART

The accuracy assessment clearly identified RF as the best-performing classifier, achieving an Overall Accuracy of 91.13% and a Kappa coefficient of 0.894. These values indicate excellent classification performance and substantial agreement between classified and reference data (Usman & Awwal, 2025). The superiority of RF over SVM and CART has been widely documented in remote sensing literature (Chafiq et al., 2024; Tesfaye et al., 2024; Zafar et al., 2024), particularly for heterogeneous landscapes characterized by complex spectral relationships.

Several factors likely contributed to the superior performance of RF in this study. First, RF employs an ensemble learning strategy that combines predictions from numerous decision trees, thereby reducing variance and improving model stability (Cai et al., 2025). Second, the algorithm is capable of modelling complex nonlinear relationships among predictor variables without requiring assumptions regarding data distributions. Third, RF is relatively robust to noise, multicollinearity, and high-dimensional datasets, characteristics that are particularly advantageous when using multiple spectral bands and derived indices (Breiman, 2001).

In contrast, SVM produced the lowest overall performance despite its reputation as a powerful classifier. The comparatively lower accuracy may be linked to the heterogeneous nature of the study landscape and the spectral overlap among vegetation-related classes. While SVM performs exceptionally well when class boundaries are distinct, its performance can decline when classes exhibit substantial spectral mixing or when parameter selection is suboptimal (Zafar et al., 2024). The moderate performance of CART likely reflects the limitations of single-tree classifiers, which are generally more susceptible to overfitting and may not capture complex interactions among predictor variables as effectively as ensemble methods (Jodhani et al., 2025).

The visual comparison around Shika Dam further reinforces these quantitative findings. The RF output displayed greater spatial coherence, more realistic land cover patches, and clearer delineation of landscape features. Such characteristics are consistent with the smoothing effect often associated with ensemble-based classifiers, which tend to suppress isolated classification artefacts while preserving meaningful spatial patterns (Yassin et al., 2026). Conversely, the more fragmented appearance of the SVM and CART outputs suggests greater sensitivity to local spectral variability and boundary effects.

4.5 Class-Specific Performance and Classification Challenges

The class-level accuracy metrics provide additional insight into the strengths and limitations of the evaluated classifiers. Water bodies consistently achieved the highest accuracy across all models, with both PA and UA approaching or reaching 100%. This result is unsurprising because water possesses a highly distinctive spectral signature, particularly within the NIR and SWIR regions where reflectance values are markedly lower than those of terrestrial surfaces (Xiao et al., 2020).

Rock outcrops also exhibited exceptionally high classification accuracies. The strong performance of this class may be attributed to its unique spectral response and the effectiveness of SWIR-based variables in distinguishing exposed rock surfaces from surrounding land cover types. The importance ranking of B11 and B12 supports this interpretation (Yu et al., 2020). Farmland and built-up areas similarly achieved high accuracies, particularly under RF. The strong performance of these classes suggests that the combination of Sentinel-2 spectral bands and derived indices effectively captured the spectral properties associated with cultivated fields and urban surfaces. The contribution of NDWI, BSI, and NDBI likely enhanced discrimination by emphasizing moisture conditions, bare surfaces, and impervious features, respectively.

The most challenging class was dense vegetation, which recorded a PA of only 15.15% across all classifiers. This indicates that a large proportion of reference dense vegetation samples were omitted and assigned to other classes. Given the relatively small mapped extent of dense vegetation (<300 ha), this outcome may reflect class imbalance, whereby minority classes are underrepresented relative to dominant classes such as savanna vegetation and farmland (Zafar et al., 2024). Furthermore, dense vegetation patches within the study area are often small, fragmented, and embedded within broader vegetation mosaics, increasing the likelihood of mixed pixels and spectral confusion.

Savanna vegetation also exhibited comparatively lower PA values. This result likely reflects the transitional nature of savanna ecosystems, where vegetation structure and composition vary continuously across space. The spectral similarity between savanna vegetation and agricultural fields, particularly during active growing periods, may have contributed to increased confusion between these classes. Comparable challenges have been reported in other tropical savanna studies where agricultural and natural vegetation classes display overlapping spectral characteristics (Akintuyi et al., 2021; Awwal et al., 2026).

These findings indicate that while all three ML algorithms successfully mapped the principal land cover classes of the study area, RF provided higher accuracy, spatial coherence, and operationally reliable classification. The results therefore support the use of RF for Sentinel-2-based LULC mapping in tropical savanna landscapes characterized by heterogeneous land cover assemblages and complex vegetation–agriculture mosaics. However, these findings should be interpreted in light of several limitations. The use of a single Sentinel-2 composite may not fully capture seasonal variability, while reference data derived primarily from high-resolution imagery and local knowledge may introduce uncertainty in class delineation. Although polygon-level partitioning reduced spatial leakage, residual spatial autocorrelation may persist among nearby polygons. The use of spectral predictors alone, absence of hyperparameter tuning, and resampling of some 20-m bands to 10 m may also constrain the generalization of these findings.

5 Conclusion

This study evaluated the performance of RF, SVM, and CART algorithms for LULC mapping in the tropical savanna landscape of Sabon Gari and Zaria, Kaduna State, Nigeria, using Sentinel-2 spectral bands and derived indices. The results demonstrated that all three classifiers successfully identified the major LULC classes; however, their classification accuracies and spatial representations differed considerably. Among the evaluated models, RF achieved the highest classification performance, recording an OA of 91.13% and κ of 0.894, thereby outperforming both CART and SVM.

The classification outputs revealed that savanna vegetation and farmland constitute the dominant land cover types within the study area, while dense vegetation occupies only a limited proportion of the landscape. Spatial agreement analysis showed substantial consistency among the classifiers, with most disagreements occurring along class boundaries and transitional zones rather than within specific land cover classes. Feature importance analysis further indicated that shortwave infrared and red-edge bands contributed most strongly to class discrimination, highlighting the importance of moisture, vegetation condition, and soil exposure in differentiating tropical savanna land covers.

In conclusion, the findings confirm the suitability of Sentinel-2 data and ML approaches for LULC mapping in heterogeneous savanna environments and demonstrate the superior capability of the RF algorithm for producing accurate and spatially coherent classification outputs. The study therefore supports the use of RF for LULC mapping and environmental monitoring in tropical savanna landscapes with similar ecological characteristics. However, the findings should be interpreted within the spatial and methodological context of this study, particularly the relatively limited study extent, seasonal image composite, spectral-only predictor set, and predefined model parameters. Further evaluation across larger areas, multiple seasons, and independently collected field reference data would strengthen the transferability of the workflow to other savanna landscapes.

References

Abdulkadir, A., Mohammed, I., & Daudu, C.K. (2021). Organic Carbon in Tropical Soils: Current Trends and Potential for Carbon Sequestration in Nigerian Cropping Systems. In: Luetz, J.M., Ayal, D. (eds) Handbook of Climate Change Management. Springer, Cham. https://doi.org/10.1007/978-3-030-57281-5_307

Akintuyi, A. O., Fasona, M. J., Ayeni, A. O., & Soneye, A. S. O. (2021). Land use/land cover and climate change interaction in the derived savannah region of Nigeria. Environmental Monitoring and Assessment, 193(12), Article 848. https://doi.org/10.1007/s10661-021-09642-6

Arpitha, M., Ahmed, S. A., & Harishnaika, N. (2023). Land use and land cover classification using machine learning algorithms in Google Earth Engine. Earth Science Informatics, 16(4), 3057–3073. https://doi.org/10.1007/s12145-023-01073-w

Awwal, Y. A., Gani, A. T., & Akwanaki, G. L. (2025a). Spatiotemporal dynamics of land use/cover and its impacts on carbon stocks in a Nigerian savanna. Sustainable Geosciences: People, Planet and Prosperity, 1, 100008. https://doi.org/10.1016/j.susgeo.2025.100008

Awwal, Y. A., Gani, A. T., Young, A. C., & Stanley, T. N. (2025b). Soil property responses to land use dynamics in Wukari, Guinea Savanna of northeastern Nigeria. Dutse Journal of Pure and Applied Sciences, 11(4b), 137–147. https://dx.doi.org/10.4314/dujopas.v11i4b.13

Awwal, Y. A., Young, A. C., & Linus, M. (2026). Geospatial analysis of soil quality index in relation to land use/land cover and spectral indices in Wukari, Nigeria. Kuwait Journal of Science, 53(3), 100577. https://doi.org/10.1016/j.kjs.2026.100577

Breiman, L. (2001). Random forests. Machine Learning, 45, pp. 5–32. https://doi.org/10.1023/A:1010933404324

Cai, B., Camarcat, L., Formosa, N., & Quddus, M. (2025). A stacked ensemble model for traffic conflict prediction using emerging sensor data. Transportation Research Interdisciplinary Perspectives, 31, Article 101457. https://doi.org/10.1016/j.trip.2025.101457

Chafiq, T., Hmamou, M., Ouhammou, I., Azmi, R., & Kumar, M. (2024). Modelling change detection for unveiling urban transitions: Using machine learning algorithms and Sentinel-2 data in Larache City, Morocco. Modeling Earth Systems and Environment, 10(2), 1711–1725. https://doi.org/10.1007/s40808-023-01860-w

Gaur, S., & Singh, R. (2023). A Comprehensive Review on Land Use/Land Cover (LULC) Change Modeling for Urban Development: Current Status and Future Prospects. Sustainability, 15(2), 903. https://doi.org/10.3390/su15020903

He, D., Shi, Q., Xue, J., Atkinson, P. M., & Liu, X. (2023). Very fine spatial resolution urban land cover mapping using an explicable sub-pixel mapping network based on learnable spatial correlation. Remote Sensing of Environment, 299, Article 113884. https://doi.org/10.1016/j.rse.2023.113884

Jodhani, K.H., Patel, D., Madhavan, N., Gupta, N., Singh, S.K., Pandey, M., 2025. ML- based land use and land cover classification: assessing performance and predicting future changes. Journal of Hydrologic Engineering 30 (4). https://doi.org/10.1061/JHYEFF.HEENG- 6416.

Kılıç, M. (2026). Mapping and modelling land degradation vulnerability in a semi-arid region: a case study from Battalgazi District, Turkiye. PeerJ 14:e20606 https://doi.org/10.7717/peerj.20606

Koko, A. F., Yue, W., Abubakar, G. A., Hamed, R., & Alabsi, A. A. N. (2020). Monitoring and Predicting Spatio-Temporal Land Use/Land Cover Changes in Zaria City, Nigeria, through an Integrated Cellular Automata and Markov Chain Model (CA-Markov). Sustainability, 12(24), 10452. https://doi.org/10.3390/su122410452

Lin, R., Chen, H., Wei, Z., Li, Y., Zhang, B., Sun, H., & Cheng, M. (2022). Improved Surface Soil Moisture Estimation Model in Semi-Arid Regions Using the Vegetation Red-Edge Band Sensitive to Plant Growth. Atmosphere, 13(6), 930. https://doi.org/10.3390/atmos13060930

Obade, V. P., & Lal, R. (2013). Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). Catena, 104, 77–92. https://doi.org/10.1016/j.catena.2012.10.014

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. Remote Sensing, 12(14), 2291. https://doi.org/10.3390/rs12142291

Salleh, S. A., Khalid, N., Danny, N., Zaki, N. A. M., Ustuner, M., Latif, Z. A., & Foronda, V. (2024). Support Vector Machine (SVM) and object-based classification in earth linear features extraction: A comparison. Revue Internationale de Géomatique, 33, 183–199. https://doi.org/10.32604/rig.2024.050723

Seyam, M. M. H., Haque, M. R., & Rahman, M. M. (2023). Identifying the land use land cover (LULC) changes using remote sensing and GIS approach: A case study at Bhaluka in Mymensingh, Bangladesh. Case Studies in Chemical and Environmental Engineering, 7, Article 100293. https://doi.org/10.1016/j.cscee.2022.100293

Solgi, S., Ahmadi, S. H., & Seidel, S. J. (2023). Remote sensing of canopy water status of the irrigated winter wheat fields and the paired anomaly analyses on the spectral vegetation indices and grain yields. Agricultural Water Management, 280, Article 108226. https://doi.org/10.1016/j.agwat.2023.108226

Tesfaye, W., Elias, E., Warkineh, B., Tekalign, M., & Abebe, G. (2024). Modeling of land use and land cover changes using Google Earth Engine and machine learning approach: Implications for landscape management. Environmental Systems Research, 13(1), Article 31. https://doi.org/10.1186/s40068-024-00366-3

Xiao, C., Li, P., Feng, Z., Liu, Y., & Zhang, X. (2020). Sentinel-2 red-edge spectral indices (RESI) suitability for mapping rubber boom in Luang Namtha Province, northern Lao PDR. International Journal of Applied Earth Observation and Geoinformation, 93, Article 102171. https://doi.org/10.1016/j.jag.2020.102171

Yassin, W., Abdollah, M. F., Ismail, A., Kamis, N. H., Razak, S. F. A., & Joy, H. K. (2026). Multi-stage spatial temporal ensemble model with integrated learning methods for robust deepfake detection. Discover Computing, 29(1), Article 291. https://doi.org/10.1007/s10791-026-10093-1

Yu, X., Lu, D., Jiang, X., Li, G., Chen, Y., Li, D., & Chen, E. (2020). Examining the Roles of Spectral, Spatial, and Topographic Features in Improving Land-Cover and Forest Classifications in a Subtropical Region. Remote Sensing, 12(18), 2907. https://doi.org/10.3390/rs12182907

Zafar, Z., Zubair, M., Zha, Y., Fahd, S., & Nadeem, A. A. (2024). Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. The Egyptian Journal of Remote Sensing and Space Sciences, 27(2), 216–226. https://doi.org/10.1016/j.ejrs.2024.03.003

Zhao, Z., Islam, F., Waseem, L. A., Tariq, A., Nawaz, M., Islam, I. U., Bibi, T., Rehman, N. U., Ahmad, W., Aslam, R. W., Raza, D., & Hatamleh, W. A. (2024). Comparison of three machine learning algorithms using Google Earth Engine for land use land cover classification. Rangeland Ecology & Management, 92, 129–137. https://doi.org/10.1016/j.rama.2023.10.007