ONLINE JOURNAL OF BIOLOGICAL SCIENCES

RESEARCH ARTICLE

Assessing Supervised Classification Algorithms for Mangrove Detection and Differentiation from Other Forest Types in Paiton District, Probolinggo Regency

Karuniawan Puji Wicaksono[1*], Fathor Rohman[2], Yasushi Mitsuda[3]

[1]*Department of Agronomy, Universitas Brawijaya, Jl. Veteran, Ketawanggede, Malang City, East Java, Indonesia*
[2]*Department of Soil, Universitas Brawijaya, Jl. Veteran, Ketawanggede, Malang City, East Java, Indonesia*
[3]*Department of Forest Environment and Sustainability Sciences, University of Miyazaki, Gakuen-kibanadai-Nishi-1-1, Miyazaki, 889-2192, Japan*

*Corresponding Author: karuniawan.fp@ub.ac.id

Abstract: Remote sensing provides an effective approach for monitoring forest cover dynamics, yet distinguishing mangroves from other forest types remains challenging in heterogeneous coastal landscapes. This study compares six supervised classification algorithms: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Mixture Discriminant Analysis (MDA), Support Vector Machine (SVM), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) for mangrove detection in Paiton, Indonesia, using Landsat 8 imagery (2015–2025). Predictor variables included NDVI and MNDWI, with model performance evaluated via the Kappa coefficient and F1-score. Feature importance was analyzed using Mean Decrease Accuracy (RF) and Gain (XGBoost) to identify the most influential predictors. The results show that SVM achieved the highest average Kappa (0.93). RF and XGBoost demonstrated the most stable performance, with consistently high F1-scores and low standard deviations (0.05). In contrast, LDA and QDA showed lower and less stable accuracy. Overall, RF and XGBoost are recommended for multitemporal LULC classification, as they effectively utilize vegetation indices to separate mangroves from spectrally similar forest types with high reliability. Performance importance analysis indicates that NDVI plays a dominant role by capturing vegetation structure and seasonal dynamics, while MNDWI contributes to mangrove differentiation through tidal-driven hydrological characteristics despite its lower importance.

Keywords: Land Use Land Cover, Mangrove, Forest Classification, Landsat 8, LDA, QDA, MDA, SVM, RF, Multi-Temporal Analysis

Received: 18-03-2026 | **Revised:** 21-06-2026 | **Accepted:** 15-07-2026 | **DOI:** 10.3844/ojbsci.2026.26.03.067

Introduction

Remote sensing technology has significantly improved the ability to observe, analyze, and monitor forest cover dynamics over large areas. Satellite imagery provides a cost-effective, repeatable, and spatially extensive approach for detecting environmental changes across different forest ecosystems [1]. Numerous studies have demonstrated the usefulness of

© 2026 Karuniawan Puji Wicaksono et al. This is an open access article published by Science Publications and distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

satellite data for monitoring deforestation, forest degradation, and Land Use or Land Cover (LULC) changes. For instance, Landsat imagery has been widely used to assess long-term changes in tropical forest cover, while advanced classification techniques have enabled more precise differentiation between land cover categories. Despite these developments, achieving highly accurate forest cover classification remains challenging in heterogeneous landscapes. In areas such as Paiton, East Java, mangroves coexist with agroforestry systems and planted forests within a relatively small spatial extent, creating spectral similarities that complicate the reliable identification of mangrove vegetation from other forest types.

Various classification algorithms have been applied to improve forest cover monitoring from satellite imagery. Traditional statistical approaches, including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Multiple Discriminant Analysis (MDA), have been used due to their interpretability and relatively low computational cost. More recently, machine learning techniques such as Support Vector Machines (SVM), Random Forests (RF), and eXtreme Gradient Boosting (XGBoost) have become increasingly popular for their capacity to capture complex and non-linear relationships in high-dimensional data. Previous studies have reported that machine learning approaches often produce higher classification accuracy than traditional statistical methods when applied to remote sensing datasets [2, 3]. However, many studies primarily evaluate overall land cover classification performance, while limited attention has been given to the specific capability of the algorithms to accurately detect Mangroves and distinguish them from other forest types that exhibit similar spectral characteristics.

This limitation highlights the need for research that focuses not only on general classification accuracy but also on the effectiveness of algorithms in identifying Mangroves within mixed forest landscapes. Accurate discrimination between mangroves and other forest formations is essential because mangrove ecosystems provide critical ecological services, including coastal protection, carbon sequestration, and habitat provision. In regions where mangrove areas are interspersed with other forests, classification errors can lead to significant uncertainty in estimating mangrove extent and monitoring ecosystem changes. Therefore, evaluating the capability of different algorithms to detect Mangroves and separate them from other forest categories represents an important methodological challenge in remote sensing-based forest monitoring.

This study addresses this issue by comparing the performance of six supervised classification methods: LDA, QDA, MDA, SVM, RF, and XGBoost in monitoring forest cover using satellite imagery in Paiton District, Probolinggo Regency. Particular emphasis is placed on assessing the ability of each algorithm to accurately identify Mangroves and distinguish them from agroforestry and planted forest classes. By conducting a systematic comparison focused on mangrove detection performance, this research contributes to improving methodological approaches for mangrove ecosystem mapping. The findings are expected to provide practical guidance for selecting appropriate classification algorithms that support more reliable mangrove monitoring and inform sustainable forest management strategies in coastal landscapes.

Materials and Methods

Study Area

The research was conducted in Paiton District, Probolinggo Regency, East Java, Indonesia (Fig. 1). This region is located in UTM 49S. This region is dominated by a mosaic of land cover, including Mangroves, agroforestry, plantations, and non-forest areas. The presence of several forest types makes this region relevant for testing classification models to distinguish forest classes with spectral similarities.

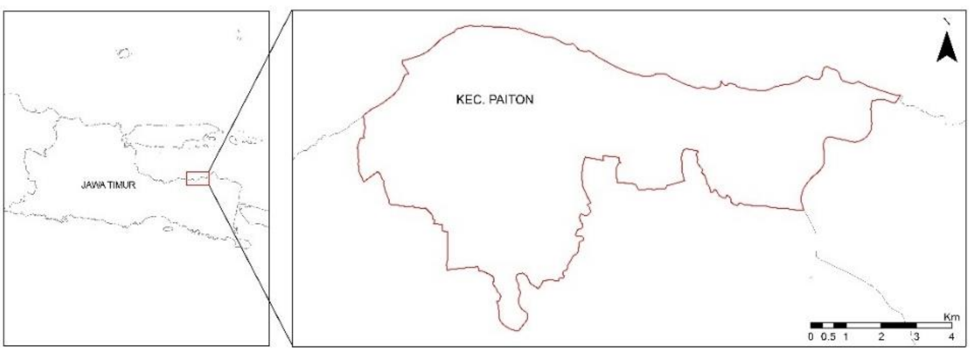

Fig. 1: Location map of Paiton Regency, Probolinggo District, East Java

DOI: 10.3844/ojbsci.2026.26.03.067

Data

Satellite Image

The primary dataset used in this study is Landsat 8 Collection 2 Level-2 Surface Reflectance imagery. This dataset was selected based on temporal consistency across observation years. Landsat 8 also provides reliable long-term data availability for time-series analysis.

The observation years for this study were 2015, 2017, 2019, 2021, 2023, and 2025. For each observation year, imagery was selected during two phases of the dry season: the beginning and end of the dry season. The beginning of the dry season occurs in April-May, while the end is in July-August. This approach was used to capture the phenological response of vegetation and variations in moisture conditions that influence the separation between various forest and non-forest types. The selected Landsat imagery had cloud interference of <10%. Next, the imagery was processed with cloud masking and cloud shadows to minimize atmospheric interference.

The Landsat 8 imagery data were processed to derive two spectral indices. These indices include the Normalized Difference Vegetation Index (NDVI) and the Modified Normalized Difference Water Index (MNDWI). NDVI and MNDWI were used as predictor variables in the LULC classification (Table 1). The entire workflow was performed within the Google Earth Engine platform, which ensures consistent processing of large-scale datasets.

Table 1: Spatial indices used as predictor variables

Indices	Equation	Reference
NDVI	$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	{4}
MNDWI	$MNDWI = \frac{\rho_{Green} - \rho_{SWIR}}{\rho_{Green} + \rho_{SWIR}}$	[5]

The two spectral indices were further divided into several groups, consisting of:

1. NDVI at the beginning of the dry season (NDVI_pre)
2. NDVI at the end of the dry season (NDVI_post)
3. NDVI difference between the beginning and end of the dry season (NDVI_diff)
4. MNDWI at the beginning of the dry season (MNDWI_pre)
5. MNDWI at the end of the dry season (MNDWI_post)

The selection of NDVI and MNDWI was guided by the ecological characteristics of the study area and the objective of evaluating classification algorithm performance using a consistent and interpretable predictor set. NDVI is a well-established indicator of vegetation vigour, canopy density, and biomass, making it suitable for distinguishing Mangroves, agroforestry systems, plantation forests, and non-forest areas [4]. MNDWI was incorporated to capture differences in moisture and surface water conditions, which are particularly relevant for coastal and mangrove ecosystems [5].

Additional spectral indices, including the Enhanced Vegetation Index (EVI) and Soil Adjusted Vegetation Index (SAVI), were considered. However, these indices were not included because the primary objective of this study was to compare classification algorithm performance under a standardized predictor framework rather than to optimize feature selection. Restricting the predictor set to NDVI and MNDWI reduced redundancy among vegetation-related variables and facilitated a more direct comparison of algorithm behaviour. Previous studies have also demonstrated that combining vegetation and water-related indices can effectively improve land-cover classification performance in tropical environments [6].

Data Sample

Reference data was obtained through a ground check, resulting in 180 sample points. This data was grouped into four LULC classes: Mangrove, Agroforestry, Plantation Forest, and Non-Forest (Fig. 2). This reference data served as the basis for identifying the characteristics of each class in the study area. Using reference data and high-resolution imagery, samples were collected for each observation year (Table 2). This dataset was used to perform LULC classification and accuracy assessment for each algorithm.

DOI: 10.3844/ojbsci.2026.26.03.067

Fig. 2: Close-up of the four LULC classes identified within the study area. Mangrove (A), Agroforestry (B), and Plantation Forest (C) exhibit distinct vegetation structure, canopy density, and phenological characteristics. Non-Forest areas (D–F) consist of built-up land, agricultural fields, and water bodies

Table 2: The distribution of data samples across LULC classes

Class	Year					
	2015	2017	2019	2021	2023	2025
Mangrove	31	31	31	31	31	31
Agroforestry	45	45	45	45	45	45
Planted Forest	120	120	120	120	120	120
Non-Forest	196	196	196	196	196	196
Total	392	392	392	392	392	392

Algorithms

This study used six supervised classification algorithms. The six classification algorithms were selected to represent different methodological families commonly used in remote sensing classification. LDA and QDA represent classical parametric discriminant approaches, MDA represents an extension capable of modeling multimodal class distributions, SVM represents margin-based machine learning methods, and RF and XGBoost represent ensemble tree-based learning. This combination enables comparison between traditional statistical classifiers and modern machine-learning approaches under identical predictor variables and training datasets. This strategic combination also enables a comprehensive comparison across interpretability, flexibility, and temporal consistency, specifically targeting the critical challenge of differentiating mangroves from other spectrally similar forest types in Paiton's dynamic landscape. While deep learning architectures have demonstrated state-of-the-art performance, they were not selected due to the limited sample size and the use of spectral indices rather than full multispectral imagery. Each Landsat image from each observation year was processed using all algorithms with the same sample dataset. The performance of each algorithm was then compared in LULC classification, with particular attention to distinguish forest types exhibiting similar spectral characteristics (Fig. 3).

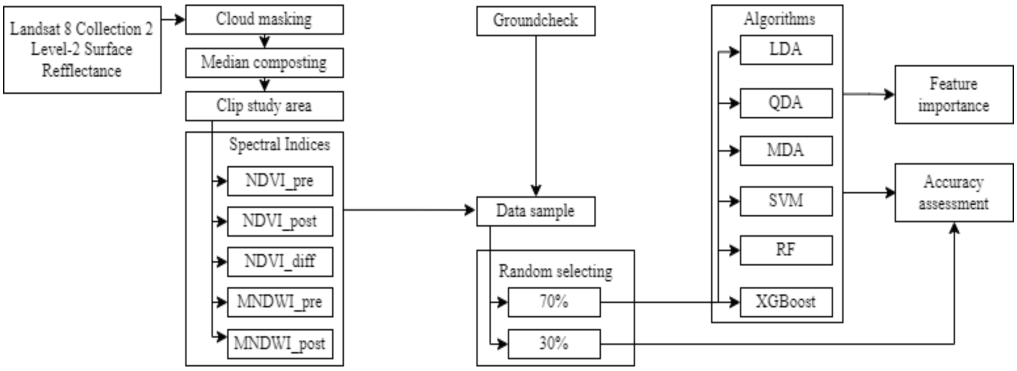

Fig. 3: The research workflow includes Landsat 8 image preprocessing, extraction of spectral indices, integration of ground-check data, and a comparison of the performance of supervised algorithms (LDA, QDA, MDA, SVM, RF, and XGBoost)

Linear Discriminant Analysis (LDA) is a statistical classification method that projects data into a lower-dimensional space by maximizing the ratio of between-class variance to within-class variance [7]. This algorithm assumes that all classes have the same covariance matrix. LDA is used as a baseline model to evaluate the ability of the linear approach to distinguish forest classes with spectral overlap. Quadratic Discriminant Analysis (QDA) is an extension of LDA that eliminates the assumption of equal covariance between classes. Each class is modeled with a different covariance matrix [8]. This allows for a non-linear decision boundary. This approach is used to capture more complex spectral variance between forest types. Mixture Discriminant Analysis (MDA) is an extension of the traditional discriminant model where each class is modeled as a mixture of several Gaussian distributions [9]. This allows for more flexible decision boundaries than linear discriminant models. MDA can capture internal class variation and produce better separation between land cover classes.

Support Vector Machine (SVM) is a machine learning algorithm that constructs an optimal hyperplane to differentiate classes based on support vectors and margins [10]. SVM is known to be effective in handling data with complex class separations and limited sample sizes. In this study, SVM was used to test the model's ability to differentiate forest types with high spectral similarity. Random Forest (RF) is an ensemble learning-based supervised machine learning algorithm that randomly constructs multiple decision trees and combines the results through a majority voting mechanism [11]. This algorithm is capable of handling non-linear relationships between predictor variables and is relatively stable against noise and overfitting. RF is used as the primary benchmark due to its consistent performance across various LULC classification studies. eXtreme Gradient Boosting (XGBoost) is a supervised machine learning algorithm based on gradient boosting that constructs an ensemble of decision trees sequentially to improve predictive performance. Each tree is developed to minimize the residual errors of the previous trees by optimizing a predefined loss function using gradient descent. The algorithm incorporates regularization techniques that control model complexity and reduce the risk of overfitting while maintaining high predictive accuracy [12]. Due to its ability to model complex non-linear relationships and efficiently process large datasets, XGBoost has been widely applied in remote sensing studies, particularly for land use and land cover classification and prediction.

Accuracy Assessment

Accuracy evaluation was performed using a confusion matrix with a hold-out validation approach. Each observation year, the data pool was randomly divided into 70% training data and 30% testing data. This division was applied consistently across all algorithms to ensure fair performance comparisons.

Classification performance was evaluated using Kappa Coefficient, Precision, Recall, and F1 Score. Kappa is used to measure the level of agreement between the classification results and the reference data, taking into account random fit. Precision indicated how accurate the models' predictions were for a given class. Recall measures the model's ability to correctly identify all samples that truly belong to that class. F1 score combines Precision and Recall into a single metric, providing a balanced measure of classification performance. All values were calculated using the following equations:

$$Kappa = \frac{\left(N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i}\right)}{N^2 - \sum_{i=1}^{r} (x_{i+} * x_{+i})} \quad (1)$$

Where r is the number of rows and columns in the confusion matrix, x_{ii} is the number of pixels that are correctly classified, x_{i+} is the marginal total for row i, and x_{+i} is the marginal sum for column *i*.

$$Precision — = TP/(TP + FP) \quad (2)$$

$$Recall — = TP/(TP + FN) \quad (3)$$

$$F1\ score = \frac{2\times(Precision\times Recall)}{(Precision+Recall)} \quad (4)$$

Where *TP* is positive data that is correctly predicted as positive, *FP* is negative data that is incorrectly predicted as positive, *FN* is positive data that is incorrectly predicted as negative, and *TN* is negative data that is correctly predicted as negative.

Evaluation results are compared across algorithms to assess classification performance. Comparisons are also conducted annually to identify model stability. This approach reveals the most effective model for supporting forest monitoring, particularly in Paiton District, Probolinggo Regency.

Feature Importance Assessment

Feature importance assessment was conducted to quantify the contribution of each predictor variable in the classification model. This analysis was performed using the built-in feature importance functions available in the RF and XGBoost algorithms. Each variable was assigned an importance score based on its contribution to improve model performance during training. Feature importance in RF is determined by the mean decrease in impurity, which reflects the reduction in prediction error during node splitting [13]. Feature importance in XGBoost is calculated based on gain, representing the contribution of each feature to improving model performance at each split [13]. The resulting scores were used to identify the most influential factors. These scores also support the interpretation of the classification results.

Results and Discussion

Spectral Confusion and Distinction: Mangrove vs. Other Forest Classes

The spectral analysis of the five predictor variables reveals a clear hierarchical distinction between stable and dynamic land cover classes. The absolute NDVI values (Fig. 4) show significant spectral overlap among the three vegetated classes, specifically Mangrove, Plantation, and Agroforestry, with values ranging broadly between 0.40 and 0.75. Agroforestry exhibits remarkable stability (0.60 to 0.75) across both pre- and post-dry seasons, confirming its resilient and structurally intact canopy. Conversely, Mangrove and Plantation display pronounced temporal fluctuations, reflecting growth cycles, harvesting activities in plantations, or sensitivity to tidal and climatic events in mangroves. This convergence in absolute greenness values creates a fundamental spectral ambiguity. Vegetation vigor alone is insufficient to reliably separate mangroves from other forest types, particularly during periods of active disturbance or recovery. The MNDWI analysis introduces a partial physical distinction. Agroforestry consistently occupies negative values (-0.25 to -0.30), reflecting dry woody biomass and minimal free surface water. In contrast, both Mangrove and Plantation occupy positive MNDWI ranges (0.00 to 0.20), with mangroves showing slightly elevated moisture signatures due to tidal inundation. Despite this, the overlap between Mangrove and Plantation in the positive MNDWI domain confirms that hydrological indices alone are also insufficient to fully disentangle these two classes.

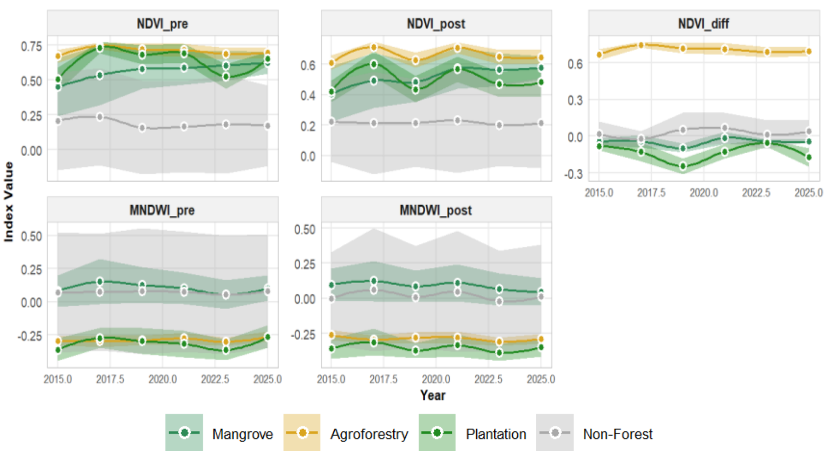

Fig. 4: Temporal trajectories of five spectral indices from 2015 to 2025 used as predictor variables

Algorithm Performance Evaluation for Mangrove Detection and Differentiation

Performance evaluation of the classification algorithms relies on the Kappa coefficient to measure predictive consistency across the study period. Table 3 presents the annual Kappa values for each algorithm from 2015 to 2025, including the mean and standard deviation for each model. These metrics provided a comprehensive overview of the model reliability and their stability in handling data variations. While Table 3 summarizes the overall agreement, Table 4 provides a more granular assessment by presenting the F1-scores for each algorithm across the four land cover classes (Mangrove, Agroforestry, Plantation, and Non-Forest) throughout the study period. This class-specific breakdown is essential to identify which algorithms perform reliably for mangrove detection specifically, rather than merely excelling in overall accuracy.

Table 3: Performance evaluation of LDA, QDA, MDA, SVM, RF, and XGBoost algorithms based on Kappa coefficients from 2015 to 2025

Year	Kappa					
	LDA	QDA	MDA	SVM	RF	XGBoost
2015	0.76	0.67	0.78	0.80	0.83	0.84
2017	0.81	0.87	0.94	0.96	0.94	0.85
2019	0.97	0.94	0.97	0.97	0.95	0.94
2021	0.87	0.87	0.94	0.96	0.94	0.97
2023	0.74	0.67	0.78	0.87	0.90	0.88
2025	0.96	0.97	0.97	1.00	0.97	0.92
Mean	0.85	0.83	0.90	0.93	0.92	0.90
Standard deviation (S)	0.10	0.13	0.09	0.08	0.05	0.05

Table 4: F1 score results of LDA, QDA, MDA, SVM, RF, and XGBoostfor Mangrove, Agroforestry, Plantation Forest, and Non-Forest classification from 2015 to 2025

Land Cover Class	Year	LDA	QDA	MDA	SVM	RF	XGBoost
Mangrove	2015	0.76	0.7	0.83	0.8	0.86	0.91
	2017	0.6	0.87	0.87	0.87	0.83	0.85
	2019	1	1	1	1	1	1
	2021	0.7	0.8	0.92	0.96	0.86	1
	2023	0.92	0.96	0.96	1	0.96	1
	2025	0.96	0.97	1	1	0.96	1
Agroforestry	2015	1	1	1	1	1	1
	2017	1	1	1	1	1	0.83
	2019	1	1	1	1	1	0.89
	2021	1	1	1	1	1	1
	2023	1	1	1	1	1	0.96
	2025	1	1	1	1	1	0.84
Plantation Forest	2015	0.83	0.77	0.83	0.85	0.87	0.86
	2017	0.85	0.88	0.9	0.94	0.96	0.91
	2019	0.97	0.94	0.97	0.97	0.95	0.94
	2021	0.89	0.88	0.96	0.96	0.96	0.97
	2023	0.78	0.74	0.78	0.86	0.91	0.88
	2025	0.96	0.96	0.99	1	0.99	0.92
Non-Forest	2015	0.82	0.75	0.84	0.87	0.88	0.89
	2017	0.91	0.89	0.91	0.94	0.95	0.92
	2019	0.98	0.95	0.98	0.98	0.96	0.98

	2021	0.95	0.94	0.95	0.98	0.96	0.98
	2023	0.79	0.68	0.84	0.91	0.93	0.92
	2025	0.98	0.97	0.99	1	0.98	0.99

The parametric models, LDA and QDA, consistently underperformed relative to the ensemble methods. LDA achieved a mean Kappa of 0.85, while QDA achieved 0.83, accompanied by the highest standard deviations (0.10 and 0.13, respectively). These limitations are rooted in their fundamental assumptions. LDA requires normally distributed data and homogeneous covariance matrices across classes [7]. QDA estimates separate covariance matrices for each class, which increases flexibility but also heightens sensitivity to data variability and sample size [8]. When these assumptions are violated, as is common in heterogeneous coastal landscapes where mangrove, plantation, and agroforestry exhibit overlapping spectral signatures, the linear or quadratic decision boundaries fail to disentangle entangled class distributions. This is empirically demonstrated in Fig. 4, where the NDVI_pre and NDVI_post panels show severe spectral convergence among Mangrove, Plantation, and Agroforestry in 2015 and particularly 2023. The consequence is directly observable in the F1-score fluctuations for the Mangrove class (Table 4). LDA dropped to 0.60 in 2017 and 0.70 in 2021; QDA dropped to 0.70 in 2015 and 0.80 in 2021. Meanwhile, Agroforestry consistently achieved perfect scores due to its distinct and stable spectral characteristics. Similar findings have been reported by [14], who observed that LDA displayed low class accuracy (<50%) in categories with spectral overlap, while homogeneous classes exceeded 90% accuracy. further demonstrated that QDA (Kappa = 0.736) underperformed compared to ensemble models such as RF (Kappa = 0.747) in heterogeneous classification tasks, reinforcing the vulnerability of parametric discriminants to data variation and mismatched statistical assumptions. The confusion matrix (Table 5) confirms these limitations: LDA showed substantial confusion between Plantation Forest and Non-Forest (19 errors) and between Mangrove and Plantation Forest (10 errors), while QDA exhibited the highest overall error, predominantly misclassifying Plantation Forest as Non-Forest (33 errors). These error patterns directly mirror the spectral overlap observed in Fig. 4, where plantation canopies, particularly after harvesting cycles, exhibit NDVI values that converge with non-vegetated areas.

MDA, which models each class as a mixture of Gaussian distributions, offered modest improvements over LDA and QDA. It achieved a mean Kappa of 0.90. This aligns with [9], who demonstrated MDA's ability to correct misclassified samples and adjust decision boundaries to complex data structures. However, MDA still exhibited notable performance drops in 2015 and 2023 (Kappa = 0.78 in both years). This reflects its sensitivity to changes in data structure and its requirement for sufficient and representative data to estimate mixture parameters reliably [15]. In 2015 and 2023, where spectral overlap among Mangrove, Plantation, and Agroforestry was most severe (Fig. 4), MDA's mixture-based boundaries became unstable, leading to reduced class separability. Table 4 shows that MDA maintained relatively stable F1-scores for Mangrove (0.83–1.00), yet its confusion matrix (Table 5) still recorded six Mangrove misclassifications as Non-Forest and 14 Plantation-to-Non-Forest errors. This confirms that even more flexible parametric approaches remain constrained by the intrinsic quality and structure of the data.

SVM achieved the highest mean Kappa (0.93), reflecting its capacity to construct optimal hyperplanes that maximize separation margins between classes [10, 16]. This mechanism is particularly effective for distinguishing vegetation types with overlapping spectral characteristics, such as mangroves and plantation forests. However, the perfect Kappa (1.00) and F1-scores achieved by SVM in 2025 require careful interpretation. As demonstrated in Fig. 4, 2025 exhibited exceptional spectral divergence across all five indices, with substantially narrowed standard deviation ribbons and clear separation between class trajectories. Under these unusually favorable data conditions, SVM's margin-maximization mechanism produced flawless classification. This phenomenon is not unique to our study; similar observations have been reported in remote sensing literature, where SVM can achieve overly optimistic accuracy when the feature space is well-separated and training data are limited or imbalanced [17]. Since this ideal separability is absent in other years, such as 2015 and 2023, this perfect score should be recognized as data-driven overfitting rather than evidence of SVM's universal superiority. The confusion matrix (Table 5) supports this interpretation: SVM achieved zero Mangrove-to-Non-Forest errors and minimal Mangrove-to-Plantation errors (only two), but its flawless 2025 performance coincides precisely with the anomalous spectral conditions shown in Fig. 4. The reliance on SVM for operational monitoring would therefore risk inconsistent performance when spectral overlap occurs, as SVM is known to be sensitive to the nature of the dataset and class imbalance [17]. This reinforces the importance of temporal stability over peak accuracy for long-term mangrove monitoring applications.

The ensemble models, RF and XGBoost, demonstrated the most stable performance across the study period. RF achieved a mean Kappa of 0.92, while XGBoost achieved 0.90. Both models recorded the lowest standard deviations (0.05

each). This stability is directly attributable to their algorithmic architectures. RF builds multiple decision trees using randomly selected samples and predictor variables, combining their results through majority voting [11]. This strategy reduces the influence of noise and minimizes the effect of spectral variability, allowing RF to produce stable decision boundaries across different observation periods. The ensemble nature of RF means that even if individual trees are confused by spectral overlap in a particular year, the aggregated vote across hundreds of trees remains robust, preventing the dramatic performance drops seen in parametric models. XGBoost constructs decision trees sequentially through gradient boosting, where each new tree corrects the errors of its predecessors. Regularization mechanisms control model complexity and reduce overfitting [12, 18]. The built-in regularization in XGBoost prevents the model from becoming too complex and overfitting to the noise or specific spectral anomalies of a single year, which is why it maintains consistency even when data conditions vary. Both models effectively exploit the stable discriminating signals identified in Fig. 4. By recursively partitioning the feature space and prioritizing these robust signals, RF and XGBoost maintain high F1-scores for the Mangrove class (0.83 to 1.00) even during years of spectral overlap, such as 2023 (Table 4). The confusion matrix (Table 5) further confirms their superiority. RF recorded zero Mangrove-to-Non-Forest errors and only one Mangrove-to-Plantation error. XGBoost recorded two Mangrove-to-Non-Forest errors and only one Mangrove-to-Plantation error. These minimal error rates, particularly for the Mangrove class, demonstrate that ensemble models effectively leverage the complementary strengths of NDVI and MNDWI to overcome the spectral ambiguity that confounds parametric classifiers. This robustness is consistent with previous studies.

Table 5: Summary of misclassification counts for each algorithm across the observation years (2015–2025)

Actual	Predicted	2017	2019	2021	2023	2025	Total
			LDA				
Agroforestry	Plantation Forest	0	0	0	0	0	0
Mangrove	Non-Forest	1	0	1	1	0	4
Non-Forest	Mangrove	1	0	0	1	0	6
Non-Forest	Plantation Forest	1	1	1	5	1	14
Plantation Forest	Mangrove	5	0	5	0	0	10
Plantation Forest	Non-Forest	6	1	3	0	0	19
			QDA				
Agroforestry	Plantation Forest	0	0	0	0	0	0
Mangrove	Non-Forest	1	0	0	0	0	2
Non-Forest	Mangrove	1	0	0	1	0	7
Non-Forest	Plantation Forest	1	1	0	1	0	8
Plantation Forest	Mangrove	1	0	4	0	0	5
Plantation Forest	Non-Forest	8	4	6	0	0	33
			MDA				
Agroforestry	Plantation Forest	0	0	0	0	0	0
Mangrove	Non-Forest	1	0	2	1	0	6
Non-Forest	Mangrove	1	0	0	0	0	3
Non-Forest	Plantation Forest	1	0	0	9	0	17
Plantation Forest	Mangrove	1	0	0	0	0	1
Plantation Forest	Non-Forest	6	2	0	0	0	14
			SVM				
Agroforestry	Plantation Forest	0	0	0	0	0	0
Mangrove	Non-Forest	1	0	0	0	0	1
Non-Forest	Mangrove	1	0	0	0	0	5
Non-Forest	Plantation Forest	1	1	0	7	0	17
Plantation Forest	Mangrove	1	0	1	0	0	2
Plantation Forest	Non-Forest	3	1	0	0	0	7

			RF				
Agroforestry	Plantation Forest	0	0	0	0	0	0
Mangrove	Non-Forest	1	0	0	0	0	1
Non-Forest	Mangrove	2	0	2	1	1	9
Non-Forest	Plantation Forest	0	2	0	4	1	12
Plantation Forest	Mangrove	0	0	1	0	0	1
Plantation Forest	Non-Forest	2	2	0	0	0	9
			XGBoost				
Agroforestry	Plantation Forest	0	24	0	0	4	28
Mangrove	Non-Forest	2	0	0	0	0	2
Non-Forest	Mangrove	1	2	0	0	0	5
Non-Forest	Plantation Forest	2	3	0	5	1	17
Plantation Forest	Mangrove	0	0	0	0	0	0
Plantation Forest	Non-Forest	2	0	0	0	0	6

Feature Importance and Key Predictors for Mangrove Differentiation

The feature importance analysis, performed using built-in metrics from RF (Mean Decrease Accuracy) and XGBoost (Gain), identified which spectral variables most effectively supported mangrove differentiation across the study period (Fig. 5). The ranking of predictors exhibited a distinct pattern between the two models. In RF, NDVI_diff emerged as the most influential predictor (mean = 81.87), substantially outperforming NDVI_pre (mean = 33.70), which ranked second. In contrast, XGBoost ranked NDVI_pre highest (mean = 0.3315), with NDVI_diff as a close secondary contributor (mean = 0.2810). Despite this rank swap between the two algorithms, NDVI_diff consistently remained among the top two predictors in both models, confirming that seasonal vegetation dynamics captured by the difference between pre- and post-dry season NDVI played a critical and stable role in distinguishing Mangroves from other land cover classes. NDVI_pre also contributed substantially to overall classification performance, particularly for separating vegetated from non-vegetated areas, although its relative importance compared to NDVI_diff varied depending on the algorithmic architecture. The consistent top-tier presence of NDVI_diff across both RF and XGBoost reinforces its robustness as the most reliable spectral discriminator for separating evergreen mangroves from deciduous plantations in heterogeneous coastal landscapes.

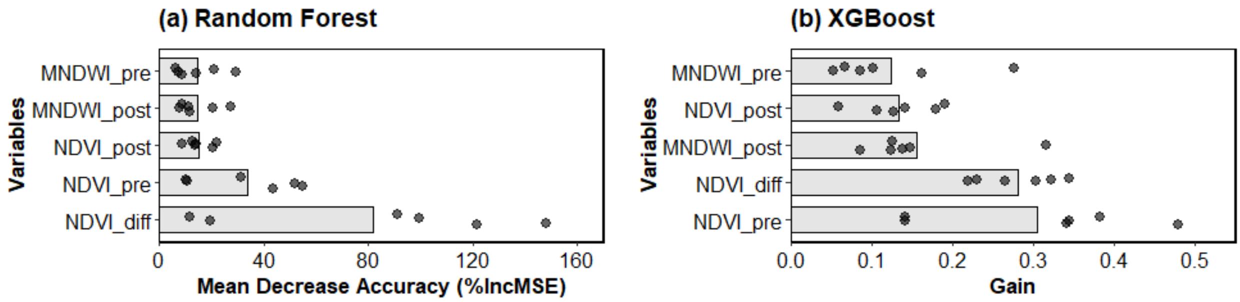

Fig. 5: Feature importance analysis across years (2015-2025): (a) RF based on Mean Decrease Accuracy and (b) XGBoost based on Gain score

The ecological significance of these importance scores becomes evident when examining the spectral behavior of these indices across the four land cover classes in Paiton District (Fig. 4). NDVI_pre provides a strong baseline for separating non-forest areas from all forest classes, capturing the fundamental difference in chlorophyll activity and canopy density between vegetated and non-vegetated surfaces. Mangrove, Agroforestry, and Plantation Forest consistently exhibit high NDVI_pre values, typically ranging from 0.60 to 0.75, reflecting dense and photosynthetically active canopies. However, this similarity also reveals a critical limitation: NDVI_pre alone cannot effectively differentiate among the three forest types due to overlapping spectral responses. The NDVI_pre trajectories for Mangrove and Plantation Forest converge in 2015, while Mangrove and Agroforestry converge in 2023 (Fig. 4), creating an intrinsically ambiguous feature space where absolute greenness provides no discriminative information. As reviewed by [11]. NDVI remains the most robust single index for broad mangrove detection due to its high chlorophyll sensitivity, but its integration with other indices is necessary to mitigate the effects of canopy saturation and tidal water interference that are prevalent in coastal mangrove ecosystems.

The seasonal vegetation dynamics captured by NDVI_diff directly address this limitation and, more importantly, provide the most reliable discriminant for mangrove differentiation across all years. Plantation Forest exhibits a more pronounced seasonal change compared to Agroforestry and Mangrove, reflecting its deciduous or semi-deciduous characteristics with distinct dry season senescence. In contrast, Mangrove maintains relatively stable NDVI values across seasons due to its evergreen nature and access to tidal water, which buffers it against seasonal drought stress. Agroforestry, which often combines woody perennials with crops, occupies an intermediate phenological position. This divergence in seasonal behavior means that NDVI_diff effectively separates Mangrove from Plantation Forest across all years, even when their absolute NDVI_pre values are nearly identical. This explains why NDVI_diff consistently ranks among the top predictors in both RF and XGBoost and in RF, it substantially outperforms NDVI_pre (mean importance of 81.87 versus 33.70). The superior performance of NDVI_diff lies in its ability to capture phenological differences that are independent of interannual variations in absolute greenness, making it a more robust and temporally stable predictor than NDVI_pre for separating mangroves from other forest types. However, NDVI_diff still leaves some ambiguity between Mangrove and Agroforestry. In years such as 2023, the NDVI_diff trajectories of these two classes converged (Fig. 4), indicating that both are relatively resistant to seasonal drought, which makes them phenologically similar and thus difficult to distinguish using seasonal signals alone. This residual ambiguity is precisely where MNDWI provides its critical supplementary value.

The MNDWI variables play a supplementary but essential role in resolving this remaining ambiguity. Despite their lower overall importance compared to NDVI-based indices (mean RF importances of 14.39 and 14.43 for MNDWI_pre and MNDWI_post; XGBoost means of 0.124 and 0.156), the MNDWI indices provide the critical hydrological signature that distinguishes Mangrove from Agroforestry. Mangrove exhibits consistently higher MNDWI values compared to Agroforestry and Plantation Forest across all years. This difference is driven by regular tidal inundation, which alters surface moisture and water presence, producing dynamic spectral responses that are not captured by vegetation indices alone [19, 20]. The distinct MNDWI signature is particularly valuable in years where NDVI values converge between Mangrove and Agroforestry, such as 2023, as it offers an independent spectral dimension for class separation. NDVI_post, which captures late dry season vegetation conditions, also contributes moderately (RF mean = 15.22; XGBoost mean = 0.167) by complementing the seasonal dynamics represented by NDVI_diff and providing end-of-season canopy status information. The integration of these three complementary signals NDVI_pre for vegetation vigor, NDVI_diff for phenological dynamics, and MNDWI for hydrological conditions collectively explains the superior and stable performance of RF and XGBoost. These ensemble models, through their recursive partitioning or gradient boosting mechanisms, automatically learn to prioritize the most appropriate variable for each specific classification challenge. When NDVI_pre overlap occurs, the models switch to NDVI_diff; when phenological signals are ambiguous, they rely on MNDWI. In contrast, parametric models such as LDA and QDA, which depend on global linear or quadratic boundaries across the entire feature space, cannot dynamically adjust to these varying spectral conditions.

Conservation Significance of Accurate Mangrove Differentiation

Paiton District, Indonesia, exemplifies the complex land-use dynamics of Indonesian coastal landscapes, where mangroves coexist with agroforestry and plantation forests. This spatial complexity creates spectral overlap that complicates accurate mapping using remote sensing. The methodological framework developed in this study directly addresses this challenge by identifying RF and XGBoost as the most reliable algorithms for separating mangroves from spectrally similar forest types, with consistently high F1-scores (0.83 to 1.00) for the Mangrove class. This finding provides conservation practitioners with an operational tool for producing defensible mangrove extent maps in heterogeneous landscapes.

The impact of this research is particularly significant for blue carbon accounting and biodiversity conservation. Mangrove ecosystems store substantially more carbon per unit area than terrestrial forests, with stocks reaching up to 1,000 Mg C per hectare [21, 22]. Misclassification of mangroves as plantations or agroforestry leads to underestimation of carbon stocks and flawed emissions estimates, which undermines REDD + initiatives and climate finance mechanisms. By demonstrating that RF and XGBoost maintain stable performance across years with severe spectral overlap (e.g., 2023), this study ensures that mangrove carbon monitoring can be conducted consistently over time. The ensemble models' ability to minimize confusion between Mangrove and Plantation Forest (RF: Only one error; SVM: Two errors) ensures that conservation priority areas and habitat connectivity assessments are based on reliable land cover maps.

The multi-index approach (NDVI_pre, NDVI_diff, and MNDWI) combined with RF and XGBoost offers an operational framework adaptable to Indonesia's dynamic coastal landscapes. Ultimately, the ability to consistently separate mangroves from spectrally similar forest types is not merely a technical achievement. It is a foundational step toward more effective conservation, sustainable resource management, and climate resilience in one of the world's most biodiversity-rich regions.

Conclusion

This study successfully identified the performance of RF and XGBoost as the most reliable supervised classification algorithms for detecting and differentiating mangroves from spectrally similar forest types in Paiton District. While SVM achieved the highest peak accuracy (Kappa = 0.93), its perfect score in 2025 reflected data-driven overfitting rather than algorithmic superiority, confirming its poor temporal stability. In contrast, RF and XGBoost delivered consistently high F1-scores for the Mangrove class (0.83 to 1.00) with the lowest standard deviations (0.05), effectively leveraging NDVI_diff for phenological separation and MNDWI for hydrological distinction. These ensemble models outperformed parametric classifiers (LDA and QDA) by maintaining stability even during severe spectral overlap, such as in 2023. The integration of RF and XGBoost with the multi-index approach (NDVI_pre, NDVI_diff, and MNDWI) provides a practical framework for reliable mangrove mapping, supporting blue carbon assessments, biodiversity conservation, and Indonesia's national restoration targets. Therefore, it is recommended to utilize RF and XGBoost for long-term, multi-temporal mangrove monitoring where temporal stability and class-specific accuracy are prioritized over peak performance.

Acknowledgment

The author would like to express sincere gratitude to Universitas Brawijaya for the institutional support provided throughout the research process.

Funding Information

The authors have not received any financial support or funding to report.

Author’s Contributions

Karuniawan Puji Wicaksono: Provided overall academic supervision and scientific direction during the research process and manuscript preparation, including critical evaluation, editing, and final refinement of the manuscript.
Fathor Rohman: Conducted the experimental work, performed data analysis, and prepared the initial draft of the manuscript.
Yasushi Mitsuda: Contributed academic supervision and scientific guidance during the research activities and the development of the manuscript.

Ethics

The data presented in this study are available on request from the corresponding authors.

References

1. Iheaturu, C. J., Curatola Fernández, G. F., Wingate, V. R., Akinyemi, F. O., Okolie, C. J., & Ifejika Speranza, C. (2026). Remote sensing of tropical forest recovery : A review and decision-support framework for multi-sensor integration. Remote Sensing of Environment, 335, 115257. https://doi.org/10.1016/j.rse.2026.115257
2. Das, M., Ghosh, S. K., Chowdary, V. M., Mitra, P., & Rijal, S. (2022). Statistical and Machine Learning Models for Remote Sensing Data Mining — Recent Advancements. Remote Sensing, 14(8), 1906. https://doi.org/10.3390/rs14081906
3. Dzulhijjah, D. A., & Martinez-Bejar, R. (2025). Comparative Analysis of Machine Learning Models for Sentinel-2 based Classification of the Bornean Heath Forest. Engineering, Technology & Applied Science Research, 15(2), 21937–21943. https://doi.org/10.48084/etasr.10173
4. Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.
5. Xu, H. (2007). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27(14), 3025–3033. https://doi.org/10.1080/01431160600589179

6. Mulkal, Rizkya, P., Aflah, N., & Muchlis. (2024). Exploring The Role of Spectral Indices on Improving Land Cover Classification Accuracy Based on Sentinel-2 Satellite Imagery in Banda Aceh City, Indonesia. Indonesian Journal of Environmental Sustainability, 2(2), 96–107. https://doi.org/10.22373/ijes.v2i2.5957
7. Cruz-Ramos, C., Garcia-Salgado, B. P., Reyes-Reyes, R., Ponomaryov, V., & Sadovnychiy, S. (2021). Gabor Features Extraction and Land-Cover Classification of Urban Hyperspectral Images for Remote Sensing Applications. Remote Sensing, 13(15), 2914. https://doi.org/10.3390/rs13152914
8. Lan, H., Stewart, K., Sha, Z., Xie, Y., & Chang, S. (2022). Data Gap Filling Using Cloud-Based Distributed Markov Chain Cellular Automata Framework for Land Use and Land Cover Change Analysis : Inner Mongolia as a Case Study. Remote Sensing, 14(3), 445. https://doi.org/10.3390/rs14030445
9. Bolatov, A., Legg, A., Melnykov, I., Nurlanuly, A., Tezekbayev, M., & Assylbekov, Z. (2025). Overspecified Mixture Discriminant Analysis : Exponential Convergence, Statistical Guarantees, and Remote Sensing Applications. ArXiv Preprint (ArXiv:2510.27056), 1–35. https://doi.org/10.48550/arXiv.2510.27056
10. Tamirat, H., Argaw, M., & Tekalign, M. (2023). Support vector machine-based spatiotemporal land use land cover change analysis in a complex urban and rural landscape of Akaki river catchment, a Suburb of Addis Ababa, Ethiopia. Heliyon, 9(11), e22510. https://doi.org/10.1016/j.heliyon.2023.e22510
11. Hosseiny, B., Abdi, A. M., & Jamali, S. (2022). Urban land use and land cover classification with interpretable machine learning – A case study using Sentinel-2 and auxiliary. Remote Sensing Applications: Society and Environment, 28, 100843. https://doi.org/10.1016/j.rsase.2022.100843
12. Wagle, N., Acharya, T. D., Kolluru, V., Huang, H., & Lee, D. H. (2020). Multi-Temporal Land Cover Change Mapping Using Google Earth Engine and Ensemble Learning Methods. Applied Sciences, 10(22), 8083. https://doi.org/10.3390/app10228083
13. Pensa, R. G., Crombach, A., Peignier, S., & Rigotti, C. (2025). Explaining Random Forest and XGBoost with Shallow Decision Trees by Co-clustering Feature Importance. Machine Learning, 114(12), 287. https://doi.org/10.1007/s10994-025-06932-9
14. Cao, C.-Y., Li, M.-T., Deng, Y.-J., Ren, L., Liu, Y., & Zhu, X.-H. (2024). Joint Sparse Local Linear Discriminant Analysis for Feature Dimensionality Reduction of Hyperspectral Images. Remote Sensing, 16(22), 4287. https://doi.org/10.3390/rs16224287
15. Qu, L., & Pei, Y. (2024). A Comprehensive Review on Discriminant Analysis for Addressing Challenges of Class-Level Limitations, Small Sample Size, and Robustness. Processes, 12(7), 1382. https://doi.org/10.3390/pr12071382
16. Chowdhury, Md. S. (2024). Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use / cover classification of urban setting. Environmental Challenges, 14, 100800. https://doi.org/10.1016/j.envc.2023.100800
17. Alizadeh, J., Bogdan, M., Classen, J., & Fricke, C. (2021). Support Vector Machine Classifiers Show High Generalizability in Automatic Fall Detection in Older Adults. Sensors, 21(21), 7166. https://doi.org/10.3390/s21217166
18. Illarionova, S., Tregubova, P., Shukhratov, I., Shadrin, D., Efimov, A., & Burnaev, E. (2024). Advancing forest carbon stocks' mapping using a hierarchical approach with machine learning and satellite imagery. Scientific Reports, 14(1), 21032. https://doi.org/10.1038/s41598-024-71133-8
19. Huang, K., Yang, G., Sun, W., Fu, B., Chen, C., Meng, X., Feng, T., & Wang, L. (2024). The phenology and water level time-series mangrove index for improved mangrove monitoring. International Journal of Applied Earth Observation and Geoinformation, 134, 104188. https://doi.org/10.1016/j.jag.2024.104188
20. Huang, X., Fu, Y., Ding, H., Tang, G., Ma, P., Liu, L., Xue, Y., Wu, S., & Chen, Y. (2025). Inter-annual changes and growth trends mapping of mangrove using Landsat time series imagery. GIScience & Remote Sensing, 62(1), 2480422. https://doi.org/10.1080/15481603.2025.2480422
21. Alongi, D. M. (2020). Global Significance of Mangrove Blue Carbon in Climate Change Mitigation. Sci, 2(3), 67. https://doi.org/10.3390/sci2030067
22. Donato, D. C., Kauffman, J. B., Murdiyarso, D., Kurnianto, S., Stidham, M., & Kanninen, M. (2011). Mangroves among the most carbon-rich forests in the tropics. Nature Geoscience, 4(5), 293–297. https://doi.org/10.1038/ngeo1123