

Random Forest-Based Vegetation and Non-Vegetation Classification in Siak Regency Using Landsat 8 and Google Earth Engine

Faiza Aulia Supriatna¹⁾, Elvira Nurrahma²⁾, Inggit Lolita Sari³⁾, Emilia Roza⁴⁾,
Harry Ramza^{5)*}

^{1,2,4,5} Department of Electrical Engineering, Faculty of Industrial Technology and Informatics
Universitas Muhammadiyah Prof. DR. Hamka

³ Research Center for Geoinformatics, The National Research and Innovation Agency of
Indonesia (BRIN)

* Corresponding author : hramza@uhamka.ac.id

Abstract

Monitoring vegetation degradation is essential for supporting sustainable land resource management through accurate and continuous land cover assessment. This quantitative study classifies vegetation in Siak Regency, Riau Province, using Landsat 8 Surface Reflectance (Level-2) imagery from 2024 and the Random Forest (RF) algorithm in the Google Earth Engine (GEE) environment. Fifty purposively selected ground truth samples, categorized as vegetation and non-vegetation, were used. Spectral bands were combined with three vegetation indices, namely NDVI, SAVI, and EVI, to improve classification performance. The dataset was randomly split into 70% training data (33 samples) and 30% validation data (17 samples) to maintain reproducibility. The model achieved an Overall Accuracy of 88.24% and a Kappa coefficient of 0.757, indicating substantial agreement between predicted and reference classes. The integration of the three indices provided complementary spectral information, improving the model's ability to distinguish vegetation density, atmospheric influence, and soil background sensitivity. However, the small validation dataset and binary classification scheme limit the statistical reliability of the results and their comparability with multi-class studies. Overall, the findings show that combining Random Forest and Google Earth Engine offers an efficient and scalable approach to vegetation mapping. Future studies should use larger datasets, more robust validation strategies, and finer land cover classification schemes to improve reliability in tropical environmental monitoring.

Keywords- Landsat 8, Vegetation Index, *GEE*, *QGIS*, *Siak Regency*

1. INTRODUCTION

In recent decades, environmental security worldwide has been severely threatened by shifting land functions and the depletion of green cover. Data from the FAO State of the World's Forests 2024 emphasizes that forest loss continues at a worrisome pace, reaching an average annual decline of 10.2 million hectares between 2015 and 2020. This ongoing trend exacerbates global warming, accelerates the extinction of species, and destabilizes vital ecological functions. Indonesia, as a country with vast tropical forest areas, faces significant pressure due to land conversion for agriculture, plantations, and infrastructure development. Therefore, accurate and regular monitoring of vegetation conditions is of utmost importance. Remote sensing technology, particularly satellite imagery such as Landsat 8, has proven effective in providing consistent temporal and spatial data for land cover change analysis [1], [2], and the development of the Google Earth Engine computing platform has accelerated large-scale data analysis [3].

Although utilizing machine learning models within the Google Earth Engine (GEE) environment has transformed broad-scale mapping, current literature still exhibits notable methodological variations. For instance, the platform's capacity for processing long-term datasets was verified by Alshehri et al. [3]. Nonetheless, Nasiri et al. [4] pointed out that the choice of composite techniques such as selecting between median or percentile values can induce noticeable variations in mapping outputs, even when dealing with data from the exact same satellite instrumentation. Tassi et al. [5] further complicated this picture by showing that pixel-based and object-based classification within GEE produce divergent accuracy levels depending on landscape heterogeneity, indicating that algorithmic choice cannot be decoupled from spatial context. The Random Forest (RF) algorithm has emerged as a preferred tool among various classifiers because it effectively manages dense multi-spectral inputs without suffering

from overfitting issues [6]. Even so, research by Ewunetu and Abebe [7] underscores that the efficacy of RF relies heavily on the caliber of the training datasets and the chosen feature combination, rather than relying solely on the model's baseline architecture. This core operational dependency is often ignored by researchers who highlight outstanding overall performance metrics without deeply analyzing the underlying factors.

A deeper examination of the literature reveals a systematic gap: most RF-based classification studies optimize for overall accuracy while neglecting the discriminative contribution of individual spectral features under tropical conditions. Melichar et al. [8] confined their feature importance analysis to temporal indices rather than raw spectral bands in arid environments, limiting transferability to humid tropical forests where phenological contrast is minimal; similarly, Vizzari [9] demonstrated improved separability through multisource integration, yet exclusively within Mediterranean land cover types structurally incomparable to tropical spectral conditions. In the Indonesian context, studies by Adrian et al. [10] and Agaventa et al. [14] prioritized spatial change patterns and crop-specific mapping without evaluating spectral band combinations for native vegetation discrimination, while broader RF applications spanning urban mapping [11], vegetation type identification [12], mangrove phenology [13], river basin LULC analysis [15], and forest cover classification [16] consistently prioritize classification output over systematic input feature evaluation. This collective oversight is particularly consequential where spectral overlap between secondary forest, mixed agriculture, and shrubland constitutes the principal source of misclassification. Therefore, this study addresses this gap by systematically evaluating the contribution of Landsat 8 spectral bands and vegetation indices within the RF GEE framework to improve vegetation class discrimination under tropical spectral conditions.

2. METHODS AND DATA

2.1. Research Design

This study employs a supervised machine learning classification workflow to identify and map vegetation cover over Siak Regency, Riau Province, Indonesia, using Landsat 8 satellite imagery processed within the Google Earth Engine (GEE) cloud computing platform [3], [18]. The classification target comprises two classes: vegetation and non-vegetation, for the observation period of January–December 2024. The complete research workflow is presented in Figure 1.

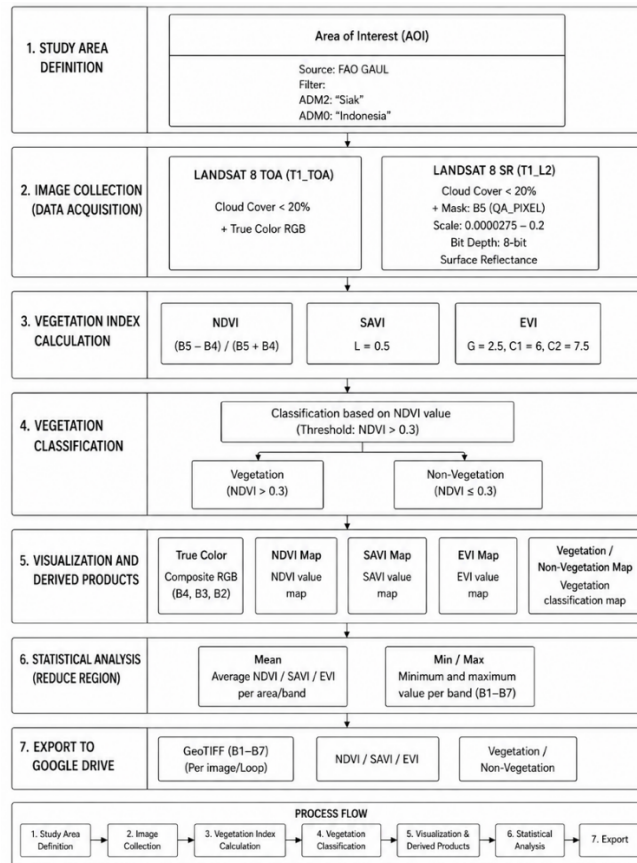


Figure 1. Research Framework

2.2. Study Area and Data Sources

The study area covers Siak Regency, spatially delineated using an administrative boundary shapefile obtained from Bappeda Kabupaten Siak, Provinsi Riau Figure 2. Two Landsat 8 image collections were utilized for distinct purposes.

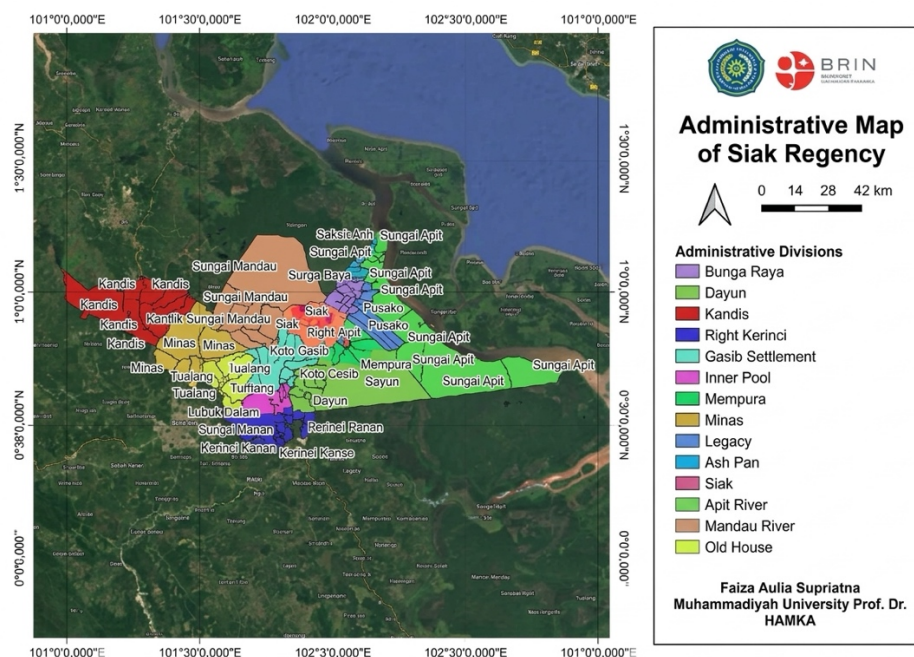


Figure 2. Administrative Boundaries of Siak Regency 2024

Since TOA imagery retains atmospheric influences that restrict its applicability for quantitative spectral analysis, only the Top-of-Atmosphere (TOA) reflectance dataset (LANDSAT/LC08/C02/T1_TOA) was used for true-color visualization [30]. Because it has undergone atmospheric correction and thus offers a more realistic picture of actual surface reflectance conditions, the Level-2 Surface Reflectance (L2) dataset (LANDSAT/LC08/C02/T1_L2) was chosen as the main dataset for all spectral analysis and categorization [17], [30]. Both collections were filtered to the Siak administrative boundary, the 2024 acquisition period, and a cloud cover threshold of less than 20%. Landsat 8 is equipped with eleven spectral bands spanning the electromagnetic spectrum from visible to thermal infrared wavelengths, as summarized in Table 1, and has been widely used in land cover mapping due to its radiometric consistency and multitemporal analysis capability [12], [17].

Table 1. Landsat 8 Spectral Bands

No	Band	Band Name	Wavelength (μm)	Resolution	Description
1.	Band 1	Coastal / Aerosol	0.433 - 0.453 μm	30 m	Coastal and aerosol studies
2.	Band 2	Visible blue	0.450 - 0.515 μm	30 m	Water body analysis and vegetation
3.	Band 3	Visible green	0.525 - 0.600 μm	30 m	Vegetation condition monitoring
4.	Band 4	Visible red	0.630 - 0.680 μm	30 m	Vegetation analysis, NDVI
5.	Band 5	Near-infrared	0.845 - 0.885 μm	30 m	Biomass estimation, NDVI
6.	Band 6	Short wavelength infrared	1.56 - 1.66 μm	30 m	Soil and vegetation moisture
7.	Band 7	Short wavelength infrared	2.10 - 2.30 μm	60 m	Burned areas, mineral detection
8.	Band 8	Panchromatic	0.50 - 0.68 μm	15 m	Image sharpening (pan-sharpening)
9.	Band 9	Cirrus	1.36 - 1.39 μm	30 m	Thin cloud detection
10.	Band 10	Long wavelength infrared	10.3 - 11.3 μm	100 m	Surface temperature
11.	Band 11	Long wavelength infrared	11.5 - 12.5 μm	100 m	Surface temperature

Source: U.S. Geological Survey (2021). Landsat 8 (L8) Data Users Handbook

2.3. Pre-Processing

All pre-processing was done inside GEE. To guarantee radiometric quality control, a pixel-quality cloud masking function was applied to the L2 collection using the QA_PIXEL band, masking pixels marked as cloud (Bit 3) and cloud shadow (Bit 4). To reduce inter-scene variability and residual ambient noise, a temporal median composite was created for each filtered capture. In compliance with Landsat Collection 2 Level-2 surface reflectance conversion protocols (USGS, 2021), radiometric calibration was then applied by multiplying pixel values by a scale factor of 0.0000275 and adding an offset of 0.2, producing calibrated surface reflectance values for Bands 1 through 7. Figure 3 displays the pre-processed L2 imagery (TOA true color).

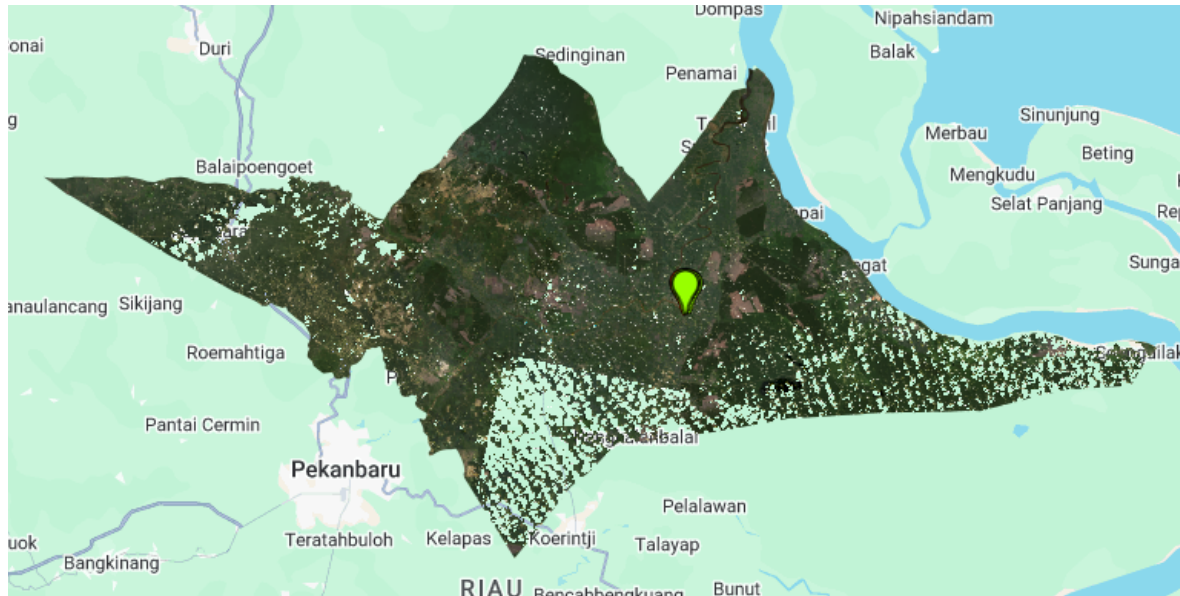


Figure 3. Pre Processing of Cloud Masking in Siak Regency 2024

2.4. Feature Extraction

Three vegetation indices were computed from the calibrated L2 bands to augment the spectral feature space and improve vegetation class separability [15], [28], [33]

The Normalized Difference Vegetation Index (NDVI) was calculated using Equation (1) to quantify vegetation greenness based on the contrast between near-infrared (NIR) and red band reflectance, with values ranging from 1 to 1 [32], [33].

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

The Enhanced Vegetation Index (EVI) was calculated using Equation (2) to enhance sensitivity in areas of dense vegetation while minimizing atmospheric and soil background effects [28], [34], using parameters $G=2.5$, $C_1=6$, $C_2=7.5$, and $L=1$.

$$EVI = 2.5 \frac{(NIR - RED)}{(NIR + 6 \cdot RED - 7.5 \cdot BLUE + 1)} \quad (2)$$

The Soil-Adjusted Vegetation Index (SAVI) was calculated using Equation (3) to reduce soil background influence, particularly relevant for the peatland-dominated landscape of Siak Regency [27], [34], using a soil adjustment factor $L=0.5$.

$$SAVI = \frac{(NIR - RED)}{(NIR + RED + L)} (1+L) \quad (3)$$

The final composite image used as classifier input comprised 10 features: seven calibrated spectral bands (B1–B7) and three vegetation indices (NDVI, SAVI, EVI). The integration of spectral bands and vegetation indices has been demonstrated to significantly improve classification performance compared to single-band input [15], [26].

2.5. Training Sample Collection and Data Split

Figure 4 illustrates the distribution of training samples used in the Random Forest vegetation classification process in Siak Regency. Training samples were collected through purposive visual interpretation of the Landsat 8 L2 true-color and vegetation index composites within the GEE interface. A total of 50 sample points were digitized, comprising 25 vegetation points (class = 1) selected from areas exhibiting consistently dense canopy characteristics, and 25 non-vegetation points (class = 0) selected from clearly identifiable water bodies, open land, roads, and settlements, distributed spatially across Siak Regency to ensure spatial representativeness. Spectral values from all 10 composite features were extracted at each sample point using the sampleRegions function at 30-meter scale.



Figure 4. Training Sample of Random Forest in Siak Regency 2024

The 50 samples were subsequently partitioned into training and validation subsets using a randomized column approach with a fixed random seed (seed = 123) to ensure reproducibility. Approximately 70% of samples ($n = 33$) were allocated for classifier training and 30% ($n = 17$) were retained as an independent validation set, consistent with the training-validation split strategy widely applied in remote sensing classification studies [19], [22], [23].

2.6. Random Forest Classification

The Random Forest (RF) algorithm, an ensemble technique that builds several decision trees using bootstrapped training subsets and combines their results via majority voting, is used in this study's supervised classification [6], [7]. Because RF can handle high-dimensional spectral data, is resistant to overfitting, and can quantify feature relevance, it has shown strong performance in remote sensing classification tasks [19], [20], and [21]. It was used to implement RF-based categorization. Classifier.smileRandomForest in GEE, set up using a fixed classifier seed (seed = 123) for reproducibility and 100 decision trees (numberOfTrees = 100). A binary vegetation map of Siak Regency was created by applying the learned model to the complete composite image after the classifier was trained on the 33 training samples utilizing all 10 input features [24], [25].

2.7. Accuracy Assessment

The classification performance was evaluated using an independent validation subset comprising 30% of the total samples ($n = 17$). A confusion matrix generated from this validation process was used to compute Overall Accuracy (OA), Kappa coefficient, Producer's Accuracy (PA), and User's Accuracy

(UA), as formulated in Equations (4)–(7) [22], [23], [32]. In addition to these external validation metrics, the study also considered the out-of-bag (OOB) error produced by the Random Forest algorithm as an internal cross-validation measure to further assess model consistency. Furthermore, feature importance analysis was conducted on the trained RF model to determine the relative contribution of each spectral band and vegetation index in distinguishing vegetation classes, thereby providing quantitative insight into the effectiveness of the selected input features in supporting the study objectives [26], [31].

Overall Accuracy :

$$\text{Overall Accuracy} = \frac{\sum \text{Diagonals}}{\text{Total Samples}} \quad (4)$$

Producer's Accuracy :

$$\text{Producer's Accuracy} = \frac{\text{Correct Class}}{\text{Total Reference}} \quad (5)$$

User's Accuracy :

$$\text{User's Accuracy} = \frac{\text{Correct Class}}{\text{Total Classified}} \quad (6)$$

Kappa :

$$\text{Kappa} = \frac{P_o - p_e}{1 - p_e} \quad (7)$$

2.8. Limitations

The primary limitation of this study concerns the relatively small sample size ($n = 50$), which constrained the independent validation set to 17 points following the 70:30 split. While the achieved overall accuracy of 88.24% and Kappa coefficient of 0.757 satisfy accepted remote sensing accuracy thresholds [19], [23], the limited validation sample size reduces the statistical robustness of the accuracy assessment. Additionally, the absence of systematic hyperparameter tuning such as optimization of variables per split (mtry) represents a methodological constraint. Future studies should employ a minimum of 100–200 samples per class with larger independent validation sets and systematic parameter optimization to produce more statistically reliable results.

3. RESULTS AND DISCUSSION

Table 2 presents the descriptive statistics of three vegetation indices NDVI, SAVI, and EVI derived from Landsat 8 imagery for Siak Regency in 2024.

Table 2. Landsat 8 Image Processing Results for Siak District and Siak Regency in QGIS 2024

Index	Region	Min	Max	Mean	Std. Dev.	Median	N Pixels
NDVI	Siak	-0.9991	1.0000	0.7742	0.1445	0.8243	7,632,081
	Regency						
SAVI	Siak	-0.1597	1.0935	0.4697	0.0987	0.4883	7,632,081
	Regency						
EVI	Siak	-0.1403	1.7868	0.4974	0.1145	0.5117	7,632,081
	Regency						

As shown in Table 2, the three indices reveal complementary rather than redundant distributional patterns. NDVI records the highest mean (0.7742) and median (0.8243), yet their divergence indicates a negatively skewed distribution dense vegetation dominates the landscape, but a non-trivial proportion of low-value pixels corresponding to water bodies, bare surfaces, and settlements suppresses the mean, exposing land-cover heterogeneity that aggregate statistics alone cannot capture. SAVI yields a lower yet more symmetrical distribution (mean: 0.4697; median: 0.4883) with the narrowest standard deviation (0.0987), reflecting not sparser vegetation but its soil-brightness correction mechanism producing more stable values for differentiating vegetation across heterogeneous terrain. EVI, despite a comparable mean to SAVI (0.4974), registers a wider standard deviation (0.1145) and the highest maximum value (1.7868), indicating greater sensitivity to high-biomass canopy complexity that both NDVI and SAVI tend to compress. Collectively, the three indices illuminate distinct yet complementary analytical dimensions, rendering their integrated interpretation substantively more informative than any single-index assessment.

The processing framework applied in this study Figure 5 follows a systematic six-stage workflow from study area delineation and Landsat 8 OLI/TIRS data acquisition through Band 5 (NIR) and Band 4 (Red) extraction, NDVI computation, and final variable assignment ensuring that the resulting spatial outputs are methodologically grounded and interpretively reliable.

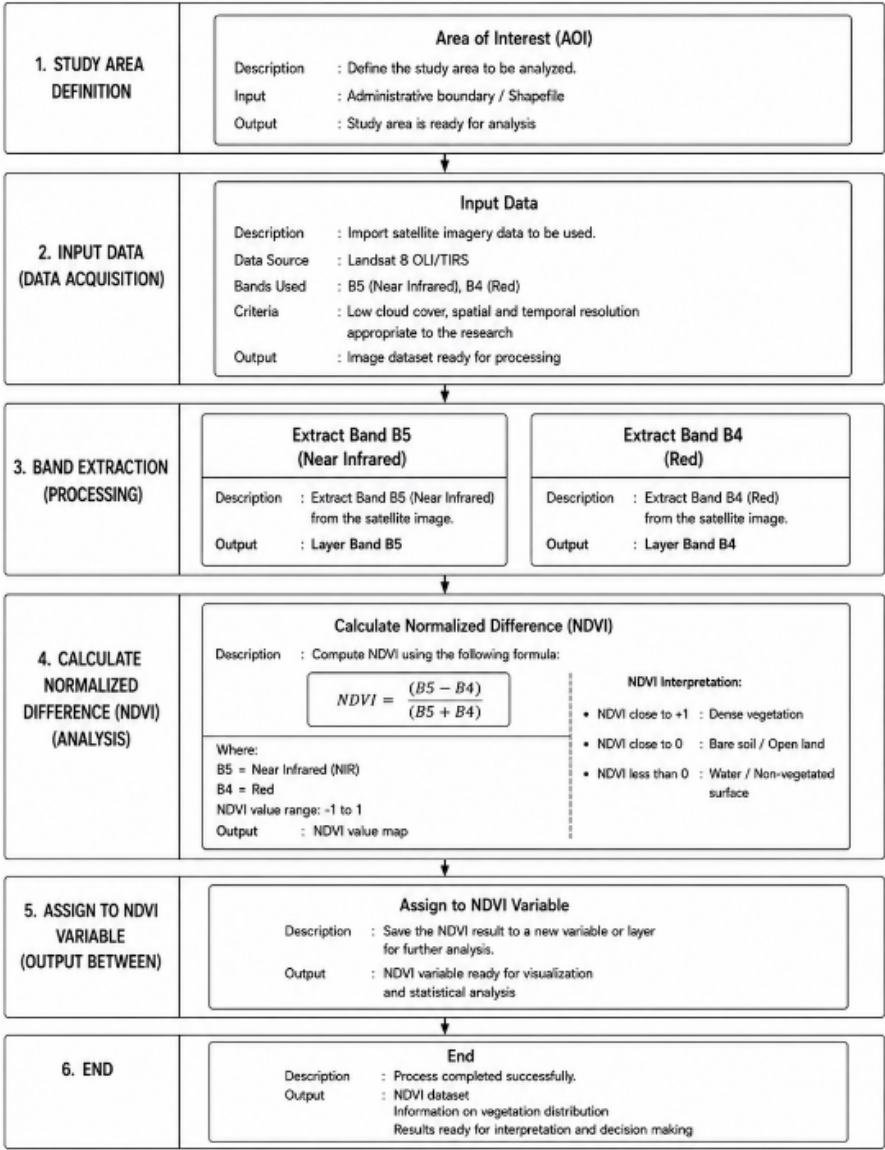


Figure 5. Framework of NDVI Classification 2024

The NDVI classification map derived from this framework Figure 6 reveals a spatially structured vegetation pattern that directly corresponds to differentiated land cover conditions across Siak Regency. High NDVI values (approaching 0.999954), represented by dominant dark green tones across the central and northern zones, are consistent with continuous canopy cover associated with dense peatland forest, mature oil palm plantations, and intact riparian vegetation all land cover types characterized by high near-infrared reflectance and elevated biomass density. Conversely, lower NDVI values (approaching -0.234146), visible as lighter and fragmented patches concentrated along the western and southern margins, correspond directly to identifiable land cover types confirmed through the Google Hybrid basemap: built-up settlement clusters, cleared agricultural land, and road infrastructure corridors. Elongated low-NDVI linear patterns specifically trace transportation routes, while dispersed low-value patches adjacent to the Siak River and its tributaries reflect the spectral suppression effect of water bodies and riparian open spaces on near-infrared reflectance. Comparatively, this spatial contrast between the high-NDVI interior and the low-NDVI periphery demonstrates that vegetation degradation in Siak Regency is not spatially random but structurally concentrated along infrastructure and settlement gradients establishing NDVI as an effective spatial proxy for detecting the boundary between ecologically intact zones and areas under active anthropogenic land use pressure.

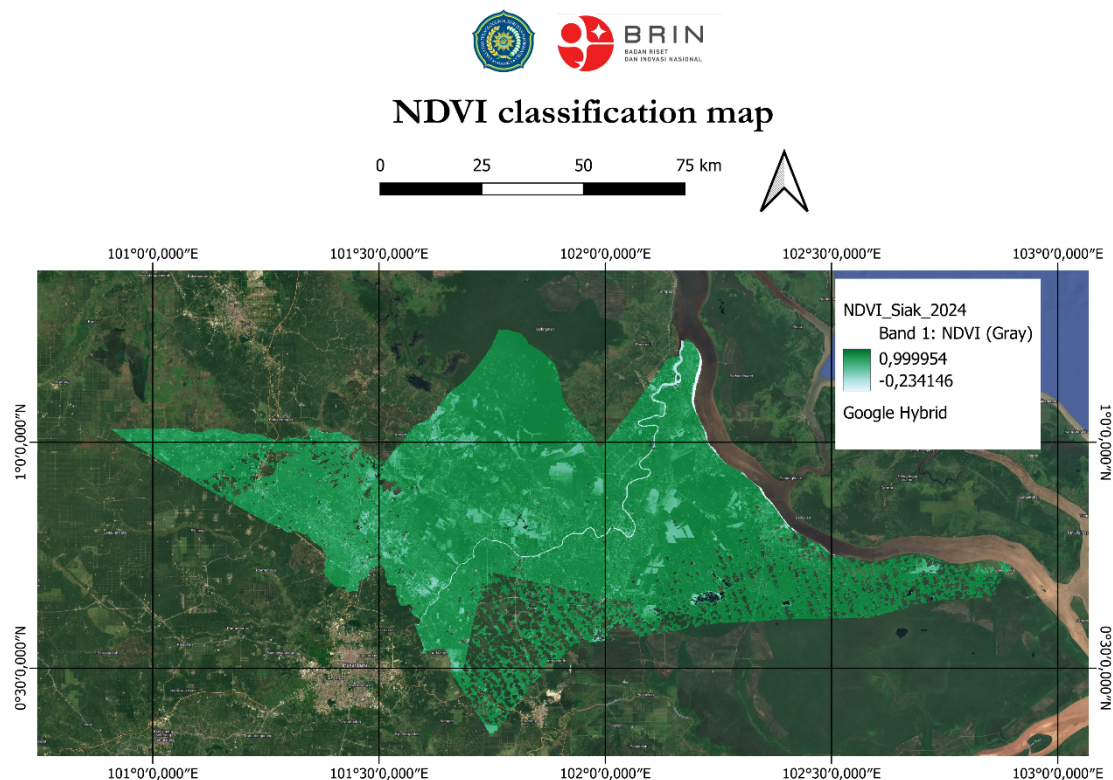


Figure 6. NDVI Classification Map of Siak District and Siak Regency 2024

The SAVI processing framework Figure 7 follows an identical sequential structure to the NDVI workflow utilizing Landsat 8 OLI/TIRS Band 5 (NIR) and Band 4 (Red) but introduces a soil brightness correction factor ($L = 0.5$) into the normalized difference formula, a methodological distinction that fundamentally differentiates SAVI's spatial output from NDVI despite sharing the same input bands.

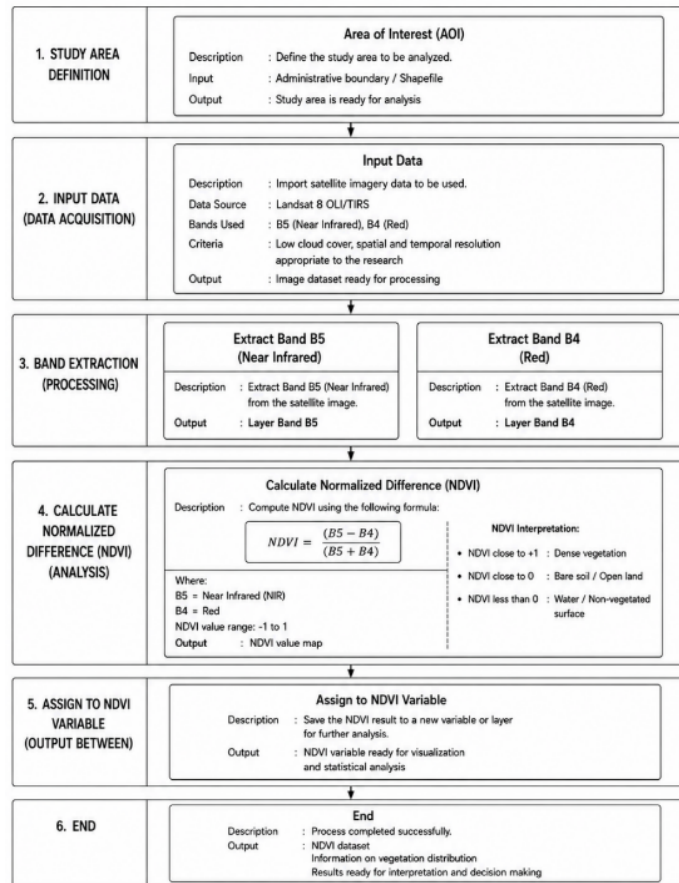


Figure 7. Framework of SAVI Classification 2024

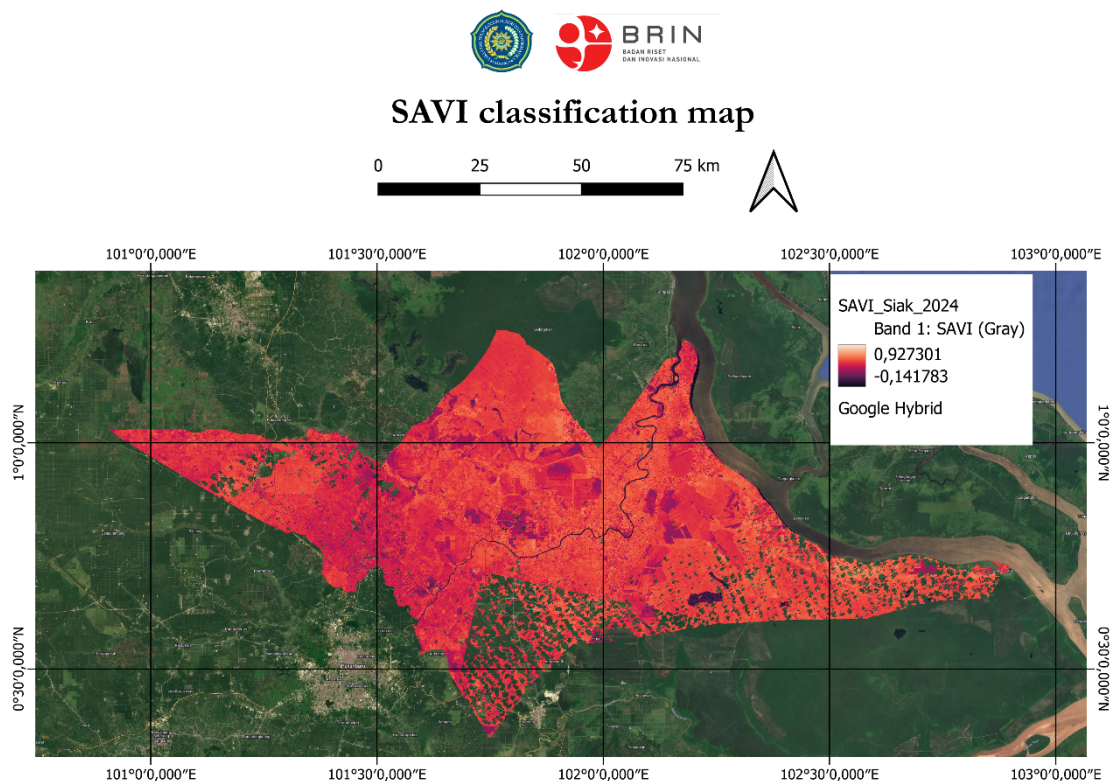


Figure 8. SAVI Classification Map of Siak District and Siak Regency 2024

The resulting SAVI categorization map (Figure 8) shows predominant red-to-orange tones throughout the northern and central zones, indicating high SAVI values (almost 0.927301) that are compatible with mature plantation canopy and deep peatland forest. When compared directly to the NDVI map, SAVI shows more internal variation within zones that NDVI identified as uniformly high, revealing sub-canopy variability in oil palm blocks and transitional peatland borders that NDVI suppresses because of exposed soil reflection. In contrast to NDVI, which tends to overstate vegetation density in these transitional zones, darker fragmented patches (approaching -0.141783) more accurately identify settlement clusters, cleared agricultural land, road corridors, and water-adjacent surfaces in the western and southern edges. In the varied land cover environment of Siak Regency, where soil background effects are ecologically significant, this spatial comparison demonstrates that SAVI offers analytically superior discriminating between degraded vegetation, bare soil, and open land.

The EVI processing framework Figure 9 distinguishes itself methodologically from both NDVI and SAVI by incorporating three spectral bands NIR, Red, and Blue alongside four calibrated constants (G=2.5, C1=6.0, C2=7.5, L=1.0), enabling simultaneous correction for both soil background reflectance and atmospheric aerosol interference a dual-correction capacity that neither NDVI nor SAVI possesses.

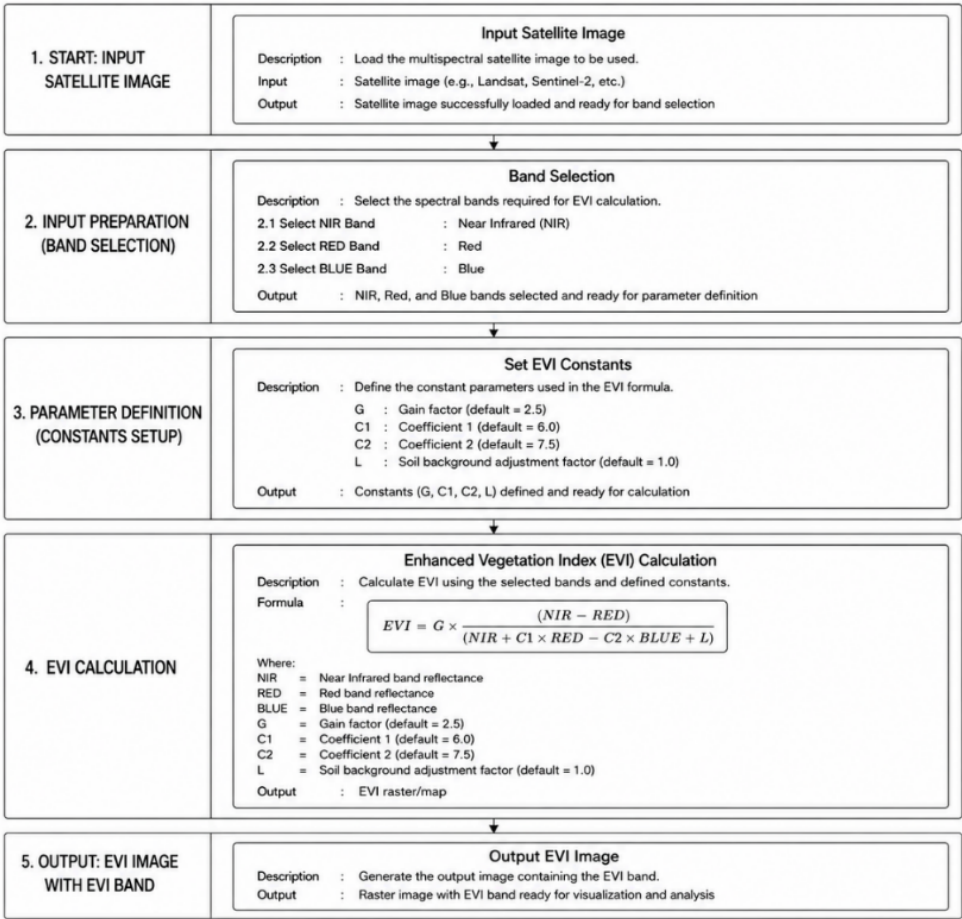


Figure 9. Framework of EVI Classification 2024

The central and northern zones of Siak Regency exhibit a predominant light-to-medium blue tone on the EVI classification map Figure 10, which corresponds to high EVI values (approaching 1.153047) that reflect dense, high-biomass canopy cover consistent with intact peatland forest and mature plantation stands. EVI shows better spatial gradation inside high-biomass areas, highlighting canopy structural variation that NDVI saturates under dense cover circumstances, in contrast to the NDVI map, which portrayed these same zones in uniform dark green. More importantly, settlement zones, cleared land, road corridors, and water-adjacent surfaces visible on the Google Hybrid basemap spatially correspond to darker blue and fragmented patches concentrated in the southern and western margins

(approaching -0.131331). These patches are consistent with SAVI's delineation but have a higher tonal contrast, indicating EVI's superior sensitivity in resolving the spectral boundary between high-biomass vegetation and degraded or non-vegetated surfaces. In high-density canopy zones where NDVI saturates and SAVI's soil correction is insufficient to discern biomass gradients, this three-index comparison indicates that EVI offers the most spectrally distinct depiction of Siak Regency's vegetation structure.

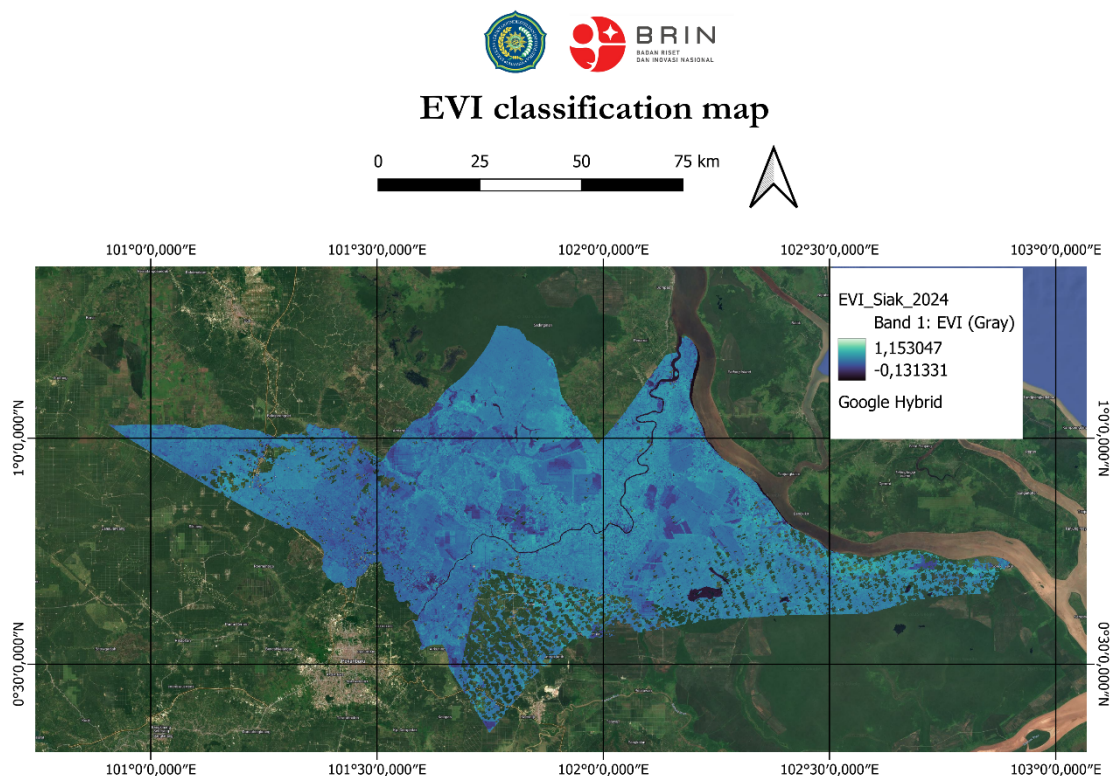


Figure 10. EVI Classification Map of Siak District and Siak Regency 2024

Critically, the spatial positioning of training samples directly corresponds to the low and high-value zones previously identified across all three vegetation indices red markers align with areas where NDVI, SAVI, and EVI consistently registered low values, while green markers occupy zones of high spectral vegetation response. This cross-index consistency strengthens the reliability of the training sample placement, as it confirms that sampled locations are spectrally stable across multiple indices rather than artifacts of a single index's sensitivity. However, the comparatively denser clustering of red markers in the western road-adjacent zones relative to the more dispersed green marker distribution suggests potential spatial imbalance in sample representation a condition that may introduce classification bias toward non-vegetated surfaces in transitional zones where vegetation density is moderate and spectrally ambiguous across NDVI, SAVI, and EVI outputs.

The Random Forest approach was used to classify the land cover after generating the vegetation indices (NDVI, SAVI, and EVI) and identifying the ROIs. The distribution of vegetation and non-vegetation classes in the study region was then shown by visualizing the classification findings.

The Random Forest classification framework Figure 11 operationalizes a machine learning workflow that is fundamentally distinct from the threshold-based approach characteristic of single-index NDVI classification. Rather than applying a fixed spectral cutoff to separate vegetation from non-vegetation, the framework initializes an ensemble classifier (numberOfTrees=100, seed=123) trained on ground truth samples using NDVI, SAVI, and EVI composite band values as predictor features a multi-feature input structure that directly leverages the complementary spectral dimensions established by the three indices. The trained classifier is subsequently applied across the full image composite via

`composite.classify()`, with model reliability assessed through Out-of-Bag (OOB) error estimation, providing an internal validation metric absent in any threshold-based classification approach.

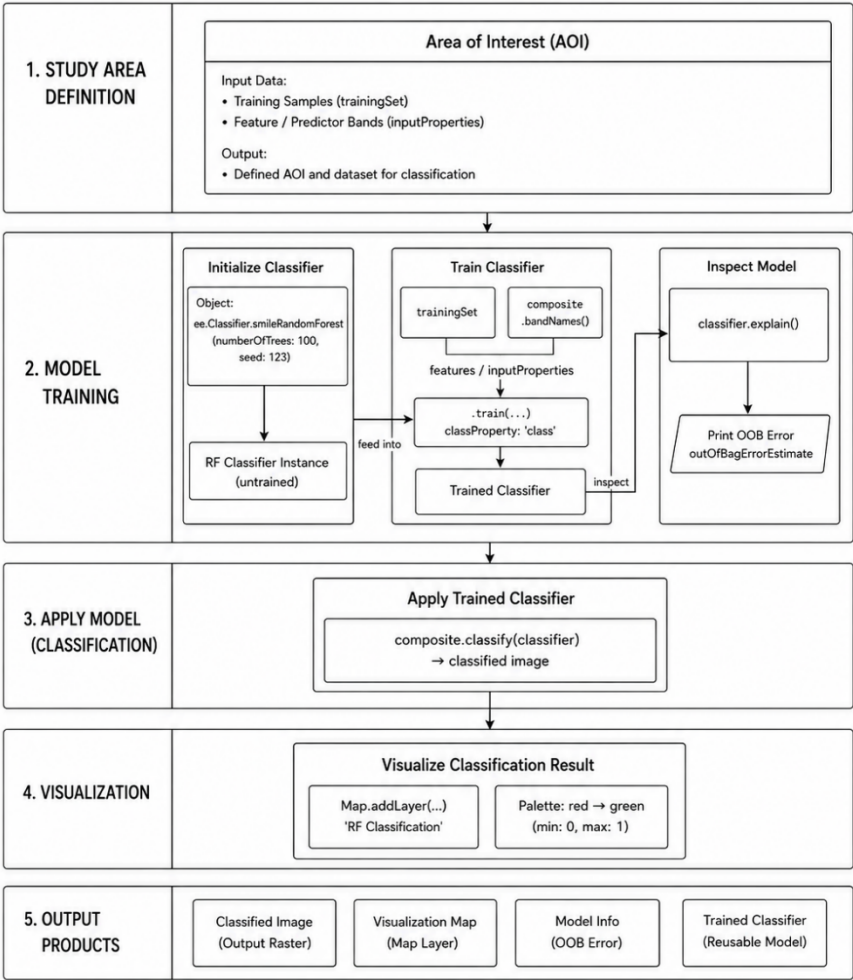


Figure 11. Flowchart of Random Forest Classification

Compared to the continuous high-value zones proposed by NDVI alone, the resulting Random Forest classification map (Figure 12) yields a binary spatial output Class 0 (red, non-vegetated) and Class 1 (green, vegetated), revealing a significantly more spatially fragmented and diverse vegetation border. While non-vegetated pixels (red) predominate in the western, southern, and peripheral margins, which are spatially consistent with settlement zones, road corridors, cleared agricultural land, and water-adjacent surfaces previously identified across NDVI, SAVI, and EVI maps, vegetated pixels (green) are concentrated in the central and northeastern regions of Siak Regency. Crucially, Random Forest effectively resolves sub-canopy land cover heterogeneity that individual vegetation indices, acting independently, would either overestimate or suppress, as evidenced by the high spatial fragmentation of green pixels within the regency boundary, especially in the central zone. This confirms the analytical superiority of ensemble classification over single-index thresholding in the heterogeneous land cover context of Siak Regency.

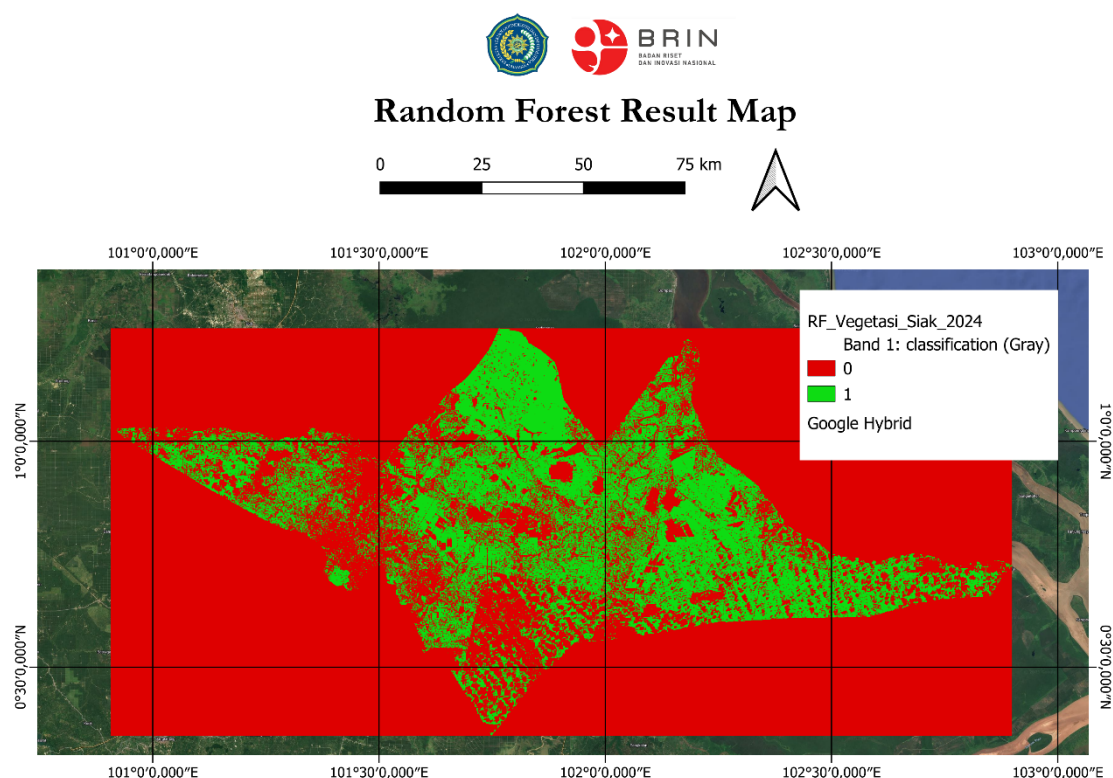


Figure 12. Random Forest Classification Results Siak District 2024

The Random Forest classification model was trained on 33 samples and validated against 17 independent samples drawn from a total of 50 ground truth points, with the 70/30 split applied via random column assignment (seed=123) to ensure reproducible partitioning. The resubstitution confusion matrix yielded a perfect Overall Accuracy of 100% and Kappa of 1.0 results that, while mathematically expected from training data self-evaluation, are analytically uninformative and should not be used as the primary performance indicator due to their inherent overfitting bias.

Table 3. Random Forest Algorithm for Vegetation and Non-Vegetation in Siak District and Siak Regency 2024

No	Class	Non-Vegetation	Vegetation	Total	User Accuracy
1	Non-Vegetation	9	1	10	90.00%
2	Vegetation	1	6	7	85.71%
3	Total	10	7	17	
4	Producer Accuracy	90.00%	85,71%		

The independent validation matrix provides a more statistically credible performance assessment: an Overall Accuracy of 88.24% (15/17), a Kappa coefficient of 0.757 falling within the "substantial agreement" category and class-specific accuracies of 90.00% Producer Accuracy and 90.00% User Accuracy for Non-Vegetation, alongside 85.71% for both Producer and User Accuracy for the Vegetation class. The OOB Error of 9.09% further corroborates moderate model reliability, as this internal ensemble metric operates independently of the validation split. Critically, however, these accuracy values must be interpreted with measured caution: a validation set of only 17 samples remains statistically limited for drawing broad generalizations about classification performance across the full spatial extent of Siak Regency, wherein a single misclassification as observed symmetrically in both

classes exerts a disproportionate influence on reported metrics. The model's performance is therefore best characterized as promising and indicative rather than conclusively robust, and future work should prioritize expansion of the validation dataset to strengthen the statistical reliability of the reported accuracy assessment.

4. CONCLUSION

This study classified vegetated and non-vegetated land cover in Siak Regency using a Random Forest model implemented in Google Earth Engine (GEE), integrating spectral bands from Landsat 8 OLI/TIRS with three complementary vegetation indices NDVI, SAVI, and EVI. The independent validation yielded an Overall Accuracy of 88.24% and a Kappa coefficient of 0.757, indicating substantial agreement between classification outputs and reference data. The multi-index input strategy demonstrably contributed to classification performance: NDVI provided general vegetation detection, EVI enhanced sensitivity toward high-biomass canopy and atmospheric variation, and SAVI reduced soil brightness interference in transitional and sparse vegetation zones a spectral complementarity that no single index could replicate independently.

However, these findings must be contextualized within the study's inherent limitations. The validation dataset comprised only 17 independent samples derived from a total of 50 ground truth points, a sample size that is statistically insufficient to support broad generalizations about model reliability across the full spatial and ecological heterogeneity of Siak Regency. Furthermore, the binary classification scheme restricted to vegetated and non-vegetated classes represents a significant simplification of the actual land cover complexity in the study area, which encompasses peatland forest, oil palm plantation, riparian vegetation, settlement, and bare soil as functionally distinct categories. Compared to multi-class classification studies utilizing larger validation datasets, the present model's accuracy figures should therefore be regarded as indicative rather than conclusive. Future research should expand both the validation sample size and the classification scheme to encompass more granular land cover typologies, thereby improving the statistical robustness and practical applicability of Random Forest-based vegetation mapping in tropical peatland landscapes such as Siak Regency.

5. ACKNOWLEDGEMENTS

The authors would like to express their deepest gratitude to Prof. Dr. Muhammad Rokhis Khomarudin, Head of the Geoinformatics Research Center, National Research and Innovation Agency, Republic of Indonesia, for his support and the opportunity to conduct research on remote sensing and satellite applications. They also extend their gratitude to all parties who cannot be mentioned individually for their support in implementing the Independent Learning/Independent Campus program (MBKM Program).

REFERENCES

- [1] H. Shafizadeh-Moghadam dkk., "Google Earth Engine for large-scale land use and land cover mapping," *GIScience & Remote Sensing*, vol. 58, no. 6, hlm. 914–928, 2021, <https://doi.org/10.1080/15481603.2021.1947623>.
- [2] A. M., S. A. Ahmed, dan N. Harishnaika, "Land use and land cover classification using machine learning algorithms in Google Earth Engine," *Earth Science Informatics*, vol. 16, hlm. 3057–3073, 2023, <https://doi.org/10.1007/s12145-023-01113-5>
- [3] B. Alshehri, Z. Zhang, dan X. Liu, "A review of Google Earth Engine for land use and land cover change analysis," *ISPRS International Journal of Geo-Information*, vol. 14, no. 11, hlm. 416, 2025, <https://doi.org/10.3390/ijgi14110416>
- [4] V. Nasiri dkk., "Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods," *Remote Sensing*, vol. 14, no. 9, hlm. 1977, 2022, <https://doi.org/10.3390/rs14091977>.

- [5] A. Tassi dkk., "Pixel- vs. Object-Based Landsat 8 Data Classification in Google Earth Engine Using Random Forest: The Case Study of Maiella National Park," *Remote Sensing*, vol. 13, no. 12, hlm. 2299, 2021, <https://doi.org/10.3390/rs13122299>
- [6] C. B. Pande dkk., "Characterizing land use/land cover change dynamics by an enhanced random forest machine learning model: a Google Earth Engine implementation," *Environmental Sciences Europe*, vol. 36, hlm. 79, 2024, <https://doi.org/10.1186/s12302-024-00901-0>
- [7] A. Ewunetu dan G. Abebe, "Integrating google earth engine and random forest for land use and land cover change detection and analysis in the upper Tekeze Basin," *Earth Science Informatics*, vol. 18, hlm. 227, 2025, <https://doi.org/10.1007/s12145-025-01750-y>
- [8] M. Melichar dkk., "Random Forest Classification of Multitemporal Landsat 8 Spectral Data and Phenology Metrics for Land Cover Mapping in the Sonoran and Mojave Deserts," *Remote Sensing*, vol. 15, no. 5, hlm. 1266, 2023, <https://doi.org/10.3390/rs15051266>.
- [9] M. Vizzari, "PlanetScope, Sentinel-2, and Sentinel-1 Data Integration for Object-Based Land Cover Classification in Google Earth Engine," *Remote Sensing*, vol. 14, no. 11, hlm. 2628, 2022, <https://doi.org/10.3390/rs14112628>
- [10] A. Adrian dkk., "Pola Spasial Perubahan Tutupan Lahan/Penggunaan Lahan Menggunakan Google Earth Engine di Kabupaten Majalengka," *Jurnal Pembangunan Wilayah dan Kota*, vol. 19, no. 4, hlm. 447–463, 2023, <https://doi.org/10.14710/pwk.v19i4.46254>.
- [11] S. Amini dkk., "Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series," *Remote Sensing*, vol. 14, no. 11, hlm. 2654, 2022, <https://doi.org/10.3390/rs14112654>
- [12] M. Aghababaei dkk., "Vegetation Types Mapping Using Multi-Temporal Landsat Images in the Google Earth Engine Platform," *Remote Sensing*, vol. 13, no. 22, hlm. 4683, 2021, <https://doi.org/10.3390/rs13224683>
- [13] M. A. P. Aji, M. Kamal, dan N. M. Farda, "Mangrove species mapping through phenological analysis using random forest algorithm on Google Earth Engine," *Remote Sensing Applications: Society and Environment*, vol. 30, no. 4, hlm. 100978, 2023, <https://doi.org/10.1016/j.rsase.2023.100978>
- [14] C. D. Agaventa dkk., "Pemanfaatan GEE dan Random Forest untuk pemetaan perkebunan jeruk," *METHODIKA*, vol. 11, no. 2, hlm. 52–57, 2025.
- [15] K. N. Loukika, V. R. Keesara, dan V. Sridhar, "Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India," *Sustainability*, vol. 13, no. 24, hlm. 13758, 2021, <https://doi.org/10.3390/su132413758>.
- [16] G. Aziz dkk., "Remote sensing based forest cover classification using machine learning," *Scientific Reports*, vol. 14, no. 1, hlm. 1-15, 2024, <https://doi.org/10.1038/s41598-023-50863-1>
- [17] D. K. Singh dkk., "Spatiotemporal Vegetation Variability and Linkage with Snow-Hydroclimatic Factors in Western Himalaya Using Remote Sensing and Google Earth Engine (GEE)," *Remote Sensing*, vol. 15, no. 21, hlm. 5239, 2023, <https://doi.org/10.3390/rs15215239>
- [18] L. Yang dkk., "Google Earth Engine and Artificial Intelligence (AI): A Comprehensive Review," *Remote Sensing*, vol. 14, no. 14, hlm. 3253, 2022, <https://doi.org/10.3390/rs14143253>.
- [19] S. Gülci, M. Wing, dan A. E. Akay, "Land Use and Land Cover (LULC) Mapping Accuracy Using Single-Date Sentinel-2 MSI Imagery with Random Forest and Classification and Regression Tree Classifiers," *Geomatics*, vol. 5, no. 3, hlm. 29, 2025, <https://doi.org/10.3390/geomatics5030029>.
- [20] H. Yuliana dan Z. C. H. Rizqiana, "Analisis Metode Klasifikasi Pemetaan Tutupan Lahan (Land Cover) di Area Kota Bandung Menggunakan Algoritma Random Forest Pada Google Earth Engine," *EPSILON: Journal of Electrical Engineering and Information Technology*, vol. 22, no. 2, hlm. 63-72, 2024, <https://doi.org/10.55893/epsilon.v22i2.119>
- [21] H. Al Fathoni, "PEMETAAN TUTUPAN LAHAN MENGGUNAKAN RANDOM FOREST PADA PLATFORM GOOGLE EARTH ENGINE: STUDI KASUS IKN," dalam *Prosiding Seminar Nasional Teknologi dan Rekayasa (TECHNOPEX)*, 2025, hlm. 1-8.
- [22] R. Wumu dkk., "Analisis Spasial Sawah Kabupaten Sidenreng Rappang Menggunakan Data Landsat-8 dengan Metode Random Forest," *Journal of Geomatics Engineering, Technology, and Science (GETS)*, vol. 2, no. 1, hlm. 36-40, 2023, <https://doi.org/10.51967/gets.v2i1.31>
- [23] M. G. Tuğaç, F. F. Şimşek, dan H. Torunlar, "Classification of Agricultural Crops with Random Forest and Support Vector Machine Algorithms Using Sentinel-2 and Landsat-8 Images," *International Journal of Environment and Geoinformatics (IJEGEO)*, vol. 11, no. 3, hlm. 106-118, 2024, <https://doi.org/10.30897/ijegeo.1479116>

- [24] S. Mateen dkk., "Billion Tree Tsunami Forests Classification Using Image Fusion Technique and Random Forest Classifier Applied to Sentinel-2 and Landsat-8 Images: A Case Study of Garhi Chandan Pakistan," *ISPRS International Journal of Geo-Information*, vol. 12, no. 1, hlm. 9, 2022, <https://doi.org/10.3390/ijgi12010009>
- [25] H. Shokati dkk., "Random Forest-Based Soil Moisture Estimation Using Sentinel-2, Landsat-8/9, and UAV-Based Hyperspectral Data," *Remote Sensing*, vol. 16, no. 11, hlm. 1962, 2024, <https://doi.org/10.3390/rs16111962>.
- [26] X. Peng dkk., "A Comparison of Random Forest Algorithm-Based Forest Extraction with GF-1 WFV, Landsat 8 and Sentinel-2 Images," *Remote Sensing*, vol. 14, no. 21, hlm. 5296, 2022, <https://doi.org/10.3390/rs14215296>
- [27] D. X. Song dkk., "Estimation and validation of 30 m fractional vegetation cover over China through integrated use of Landsat 8 and Gaofen 2 data," *Science of Remote Sensing*, vol. 6, hlm. 100058, 2022, <https://doi.org/10.1016/j.srs.2022.100058>.
- [28] F. Sivrikaya dkk., "Forest fire risk mapping with Landsat 8 OLI images: Evaluation of the potential use of vegetation indices," *Ecological Informatics*, vol. 79, hlm. 102461, 2024, <https://doi.org/10.1016/j.ecoinf.2024.102461>
- [29] Y. Zhao dan Z. Zhu, "ASI: An artificial surface Index for Landsat 8 imagery," *International Journal of Applied Earth Observations and Geoinformation*, vol. 107, hlm. 102703, 2022, doi: 10.1016/j.jag.2022.102703.
- [30] E. Rahimi dan C. Jung, "Evaluating the applicability of landsat 8 data for global time series analysis," *Frontiers in Remote Sensing*, vol. 5, hlm. 1492534, 2024, <https://doi.org/10.3389/frsen.2024.1492534>
- [31] H. Farhadi dkk., "Near Real-Time Flood Monitoring Using Multi-Sensor Optical Imagery and Machine Learning by GEE: An Automatic Feature-Based Multi-Class Classification Approach," *Remote Sensing*, vol. 16, no. 23, hlm. 4454, 2024, <https://doi.org/10.3390/rs16234454>
- [32] S. A. Nazhifah, Maulyanda, A. Putri, dan U. Kiftiyani, "Comparative Evaluation of NDVI-Based Vegetation Classification Using Rule-Based Approach and Random Forest Models," *Cyberspace: Jurnal Pendidikan Teknologi Informatika*, vol. 9, no. 2, hlm. 22-32, 2025, <https://doi.org/10.22373/cjpti.v9i2.21556>
- [33] B. P. Kumar, K. R. Babu, B. N. Anusha, dan M. Rajasekhar, "Geo-environmental monitoring and assessment of land degradation and desertification in the semi-arid regions using Landsat 8 OLI / TIRS, LST, and NDVI approach," *Environmental Challenges*, vol. 8, hlm. 100578, 2022, <https://doi.org/10.1016/j.envc.2022.100578>
- [34] N. Mzid dkk., "Evaluation of Agricultural Bare Soil Properties Retrieval from Landsat 8, Sentinel-2 and PRISMA Satellite Data," *Remote Sensing*, vol. 14, no. 3, hlm. 714, 2022, <https://doi.org/10.3390/rs14030714>