VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i2.7281

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

UNVEILING SPATIAL PATTERNS OF LAND CONVERSION THROUGH MACHINE LEARNING AND SPATIAL DISTRIBUTION ANALYSIS

Mufida Fauziah Faiz[1]; Achmad Fauzan[1*]

Statistics Study Program[1]
Faculty of Mathematics and Natural Science, Universitas Islam Indonesia, Sleman, Indonesia[1]
https://statistics.uii.ac.id/[1]
achmadfauzan@uii.ac.id[*]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Kayu Agung District in Ogan Komering Ilir (OKI) Regency, South Sumatra, has undergone rapid population growth, resulting in notable land-use transformations. This study examines land-use change dynamics from 2019 to 2023 and identifies their spatial distribution using satellite imagery. Satellite imagery classification was performed using three machine learning algorithms—K-Nearest Neighbors (KNN), Naïve Bayes, and Logistic Regression—with KNN achieving the highest accuracy. Spatial analysis employing the Variance-to-Mean Ratio (VMR) revealed that land-use changes are spatially clustered, indicating concentrated land conversion in specific areas. These findings emphasize potential environmental risks, including declining green open spaces and increasing urban pressure. The study contributes by integrating machine learning and spatial statistical analysis (VMR) as a comprehensive framework for understanding land-use conversion, providing scientific insights to support adaptive spatial planning and the achievement of Sustainable Development Goal (SDG) 11: Sustainable Cities and Communities.*

Keywords: *land use change, machine learning, Sustainable Development Goals (SDGs), Variance Mean Ratio (VMR),*

Intisari— *Kecamatan Kayu Agung di Kabupaten Ogan Komering Ilir (OKI), Sumatera Selatan, mengalami pertumbuhan penduduk yang pesat sehingga mendorong terjadinya perubahan penggunaan lahan yang signifikan. Penelitian ini bertujuan untuk menganalisis dinamika perubahan penggunaan lahan selama periode 2019–2023 serta mengidentifikasi pola persebaran spasialnya menggunakan citra satelit. Klasifikasi citra satelit dilakukan dengan tiga algoritma machine learning, yaitu K-Nearest Neighbors (KNN), Naïve Bayes, dan Regresi Logistik, di mana metode KNN menunjukkan akurasi tertinggi. Analisis spasial menggunakan Variance-to-Mean Ratio (VMR) menunjukkan bahwa perubahan penggunaan lahan bersifat spatially clustered, yang berarti konversi lahan terkonsentrasi di area tertentu. Temuan ini menegaskan adanya potensi risiko lingkungan, seperti berkurangnya ruang terbuka hijau dan meningkatnya tekanan perkotaan. Penelitian ini berkontribusi melalui integrasi metode machine learning dan analisis spasial (VMR) sebagai pendekatan komprehensif untuk memahami dinamika konversi lahan, serta memberikan dasar ilmiah bagi perencanaan tata ruang adaptif guna mendukung pencapaian Tujuan Pembangunan Berkelanjutan (SDG) 11: Kota dan Permukiman yang Berkelanjutan.*

Kata Kunci: *Perubahan penggunaan lahan, pembelajaran mesin, Variance Mean Ratio (VMR), Pembangunan Berkelanjutan (SDGs),*

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7281

INTRODUCTION

Ogan Komering Ilir (OKI) Regency is one of the administrative regions in South Sumatra Province, covering an area of 17,025.87 km^2. The regency is subdivided into 18 districts, 314 villages, and 13 sub-districts, with Kayu Agung serving as its capital. According to data from Statistics Indonesia (BPS) in 2024, the population of OKI Regency reached 807,09 inhabitants, with a population growth rate of 1.82% [1]. The dynamics of population growth indicate an increasing trend that requires further in-depth examination. Therefore, a comprehensive analysis is essential to better understand the demographic development in OKI Regency [2].

Kayu Agung District is one of the most densely populated areas in Ogan Komering Ilir Regency. Data from Statistics Indonesia (BPS) show an increase in population in this district from 75,976 people in 2020 to 76,939 in 2021, and further to 77,970 in 2022 [3]. This population growth has directly contributed to the rising demand for land for housing, public facilities, and infrastructure. Consequently, land-use conversion has occurred, particularly the transformation of agricultural land and green open spaces into residential and commercial areas.

Land-use change not only influences spatial planning but also produces substantial ecological and social consequences. Proper management is therefore essential to maintain environmental sustainability while enhancing community well-being. Conceptually, land use represents human intervention on the Earth's surface, reflecting dynamic interactions between socio-economic activities and environmental conditions. Although related to *land-cover change*, which describes the physical alteration of the Earth's surface, *land-use change* in this study specifically refers to the transformation of land functions driven by human activities [4]. Population growth is among the main drivers of such transformations, as the increasing demand for space inevitably leads to spatial conversion for residential, commercial, and infrastructural purposes [5]. Consequently, examining the dynamics of land-use change is fundamental to formulating effective regional planning and sustainable development strategies.

Several previous studies have highlighted issues related to land-use change in various regions. Hakim et al. (2021) modeled the prediction of land-use change in Makassar City using the Multi-Layer Perceptron Neural Network (MLPNN) combined with a Markov chain, revealing that built-up areas are projected to dominate approximately 80% of the total area by 2031 [6]. Achmad et al. (2024) investigated land-cover transformation in Central Aceh and projected a decline in carbon stocks by 2039 as a result of forest conversion [7]. Sari et al (2021). evaluated the accuracy of land-cover change maps through automated digital processing and visual interpretation, and identified challenges in distinguishing between natural forests and plantation forests[8].

Another study by Kelly-Fair et al. (2022) linked land-cover change to the achievement of the Sustainable Development Goals (SDGs) in Semarang, demonstrating that urbanization and conversion into rubber plantations can trigger conflicts among different SDGs [9]. Meanwhile, Basheer et al. (2022) compared the performance of several classification algorithms—such as SVM, Random Forest, and CART—in mapping land-use/land-cover (LULC) change using Landsat, Sentinel, and Planet imagery, emphasizing the relevance of these methods in improving land-change detection accuracy [10]. Similarly, Venter et al. (2022) compared three high-resolution global LULC datasets (Dynamic World, World Cover, and Esri Land Cover) and concluded that selecting an appropriate dataset should be aligned with the spatial context and research objectives [11].

While these studies highlight urbanization, carbon loss, and classification performance in other regions, limited research has focused on the spatial clustering of land-use change in Ogan Komering Ilir (OKI) Regency, particularly in Kayu Agung District. The present study addresses this gap by analyzing the spatial and temporal dynamics of land-use change in Kayu Agung from 2019 to 2023. Methodologically, this research integrates machine-learning-based classification (K-Nearest Neighbors, Naïve Bayes, and Logistic Regression) with the Quadrant Method implemented through the Variance-to-Mean Ratio (VMR) to identify spatial clustering patterns. This integrated framework not only facilitates the detection of land-use change intensity and direction but also enhances understanding of its spatial distribution characteristics.

Overall, this study contributes by combining machine learning and spatial statistical analysis into a comprehensive framework for examining land-use change dynamics. The findings are expected to support sustainable spatial planning and policy formulation in alignment with the Sustainable Development Goal (SDG) 11: Sustainable Cities and Communities.

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i2.7281

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

MATERIALS AND METHODS

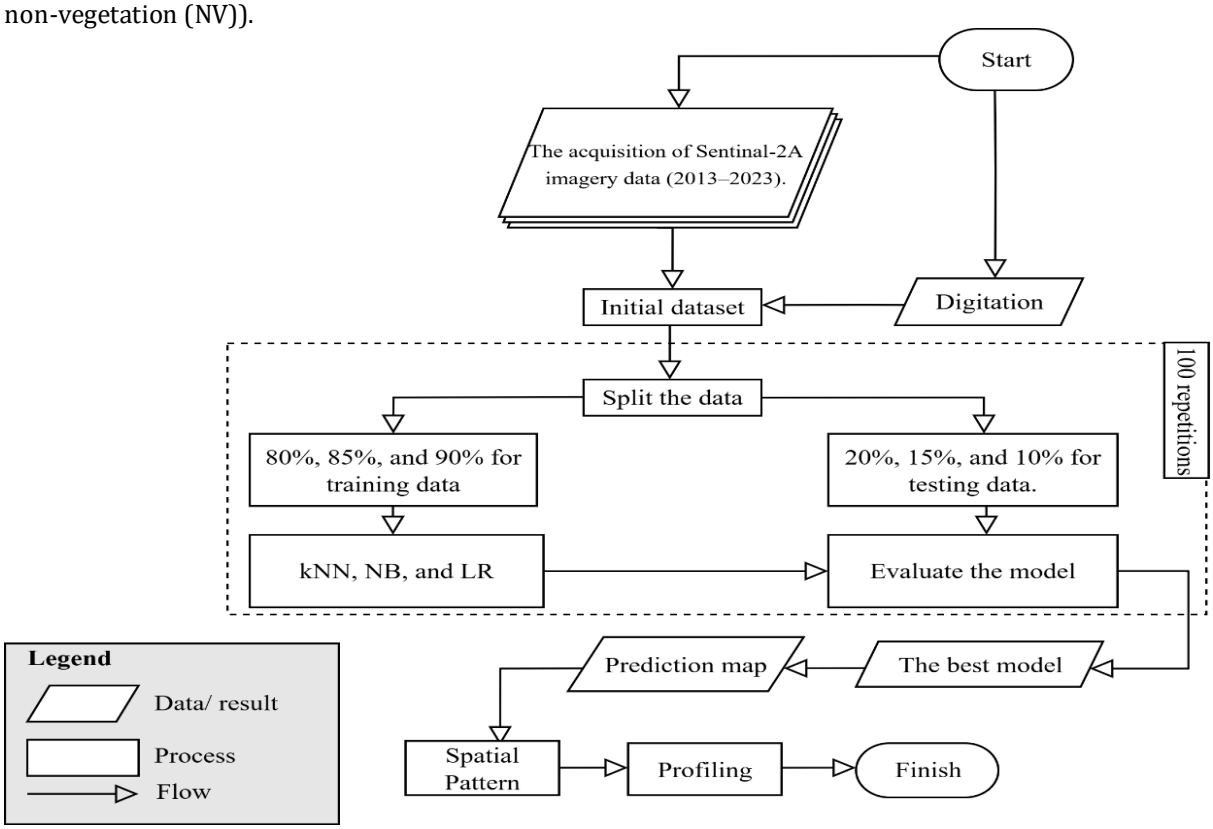
A. Data and Sources of Information

The data used in this study consist of Sentinel-2A satellite imagery covering the Kayu Agung District for the period 2019–2023. The imagery was obtained through the Google Earth Engine platform. The dataset is in raster format with a spatial resolution of 10 meters, meaning that each pixel represents a real-world area of 10 × 10 meters [12]. Subsequent digitization was carried out using the Quantum Geographic Information System (QGIS) software. Digitization refers to the process of converting information from analog or image form into digital data that can be processed by a computer [13]. In this stage, polygons were created to represent the Kayu Agung District area. The study was limited to the Kayu Agung sub-district to explore local land-use patterns in detail, but this spatial scope may limit the generalization of the results. The digitization process was categorized into two main classes, namely vegetation (V) and non-vegetation (NV). For each polygon, 10 random points were generated. These points were then assigned attributes that included location information, coordinates (longitude and latitude), and the designated class (0 = vegetation (V), 1 = non-vegetation (NV)).

B. Materials and Methods

The dataset, once obtained and preprocessed, was subsequently split using the hold-out method [14]. In this approach, the dataset was divided into two subsets: training data and testing data. Three partitioning schemes were applied, namely 90%:10%, 85%:15%, and 80%:20%. The overall research workflow is illustrated in Figure 1.

The training data from each partition were classified using three machine learning methods: k-Nearest Neighbors (Knn), Naïve Bayes (NB), and Logistic Regression (LR). The Knn method operates by comparing new data with previously observed instances. The principle of Knn is straightforward: it identifies a specified number of *k* nearest neighbors of a new data point based on a chosen distance metric [15], [16]. This distance is used to measure the similarity between the new instance and the training data. Once the distances between the new data point and all training samples are calculated, Knn selects the *k* closest instances as references and assigns the class of the new data point according to the majority class among these nearest neighbors.



Source : (Research results, 2025)

Figure 1. Research Methodology Flowchart

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7281

Naïve Bayes (NB) is a classification method based on the concept of probability. It operates by estimating the likelihood that a new data point belongs to a particular class, using the frequency patterns and value combinations of the training data [17]. For example, if there are two classes, $m = 1$ and $m = 0$, NB calculates the posterior probability—that is, the probability that a data point with characteristics *x* belongs to class *m*. This posterior probability is generally expressed as shown in Equation (1).

$$p_m = P(Y = m|x) = \frac{P(x|Y = m)\ P(Y = m)}{P(x)} \quad (1)$$

The probability of an instance x occurring within a particular class $Y = m$, denoted as $P(x|Y = m)$, is calculated by multiplying the conditional probabilities of each predictor variable. This is possible because NB relies on the fundamental assumption that all predictor variables are mutually independent. Based on this assumption, the general form of $P(x|Y = m)$ can be expressed as presented in Equation (2).

$$P(x|Y = m) = \prod_{i=1}^{p} P(x_i|Y = m) \quad (2)$$

In the Naïve Bayes method, the data to be predicted are represented as a predictor vector $x = (x_1, x_2, \dots, x_p)$, where p denotes the number of predictor variables used in the model. Once the posterior probabilities for each class (p_m) are calculated, the next step is to determine the final class by comparing these probability values. A new data point is then classified into the class with the highest probability [15], [16]. Subsequently, Logistic Regression (LR) was employed because its regression coefficients can be directly interpreted as the influence of predictor variables on the likelihood of a data point belonging to a specific category [18]. LR links the predictor variables $X_1, X_2, \dots, X_p$with the response variable Y, which consists of two classes (0 and 1). This relationship is formulated as shown in Equation (3) [19].

$$\pi_i = P(Y = 1|x) = \frac{\exp\left(\sum_{j=1}^{p} \beta_j X_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{p} \beta_j X_{ij}\right)} \quad (3)$$

$$\Leftrightarrow \text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \sum_{j=1}^{p} \beta_j X_{ij}$$

Here, β_j represents the parameter value (model coefficient) for the j^{th} variable, with $j = 1, 2, \dots, p$, where p denotes the number of predictor variables. Classification is performed by comparing the probability value π_i against a threshold θ. If $\pi_i > \theta$, the data point is classified into class 1; otherwise, if $\pi_i \leq \theta$, it is assigned to class 0. Although this threshold can be flexibly chosen between 0 and 1, the commonly used value is 0.5 [20].

To ensure consistency of the results, each classification method was executed 100 times by randomly resampling the training and testing data in each partitioning scenario. For the Knn method, simulations were conducted using three variations of the number of nearest neighbors, namely k = 3, 5, and 7. In total, 1,500 iterations were performed, consisting of 900 iterations from the Knn method and 600 iterations from LR and NB. Upon completion of the repetitions, the next step was to evaluate the performance of each method. The initial stage of this evaluation involved constructing a confusion matrix. An illustration of the confusion matrix is presented in Table 1 [21], [22].

Table 1. Confusion matrix

	Positive Prediction	Negative Prediction
Actual Positive (P)	True Positive (TP)	False Negative (FN)
Actual Positive (N)	False Positive (FP)	True Negative (TN)

Source : (Research results, 2025)

Based on Table 1, True Positive (TP) refers to the condition in which positive data are correctly identified as positive, while False Positive (FP) occurs when negative data are incorrectly classified as positive. False Negative (FN) arises when positive data are misclassified as negative, and True Negative (TN) represents the condition in which negative data are correctly identified as negative. Thus, TP and TN indicate correct predictions, whereas FP and FN represent misclassifications.

The performance of the models was evaluated using several classification metrics, including accuracy, precision, recall, F1-score, specificity, Area Under the Curve (AUC), error rate, and Matthews Correlation Coefficient (MCC). The formulas for each metric are presented in Equations (4) to (11) [23], [24].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (8)$$

$$AUC = \int_0^1 TPR\left(FPR^{-1}(x)\right)dx \quad (9)$$

$$\text{Error Rate} = 1 - \text{Accuracy} \quad (10)$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (11)$$

Values closer to 1 in the metrics of accuracy, precision, recall, F1-score, specificity, and AUC indicate better model performance, as 1 represents perfect classification. In contrast, values closer to 0 in the error rate metric are preferable, since 0 indicates no misclassifications. Regarding MCC, a value of 1 denotes perfect prediction, a value of 0 indicates performance equivalent to random guessing, and a value of −1 reflects that all predictions are incorrect. Overall, high values (approaching 1) across most metrics signify good model performance, whereas low values (approaching 0, or negative in the case of MCC) indicate poor performance.

Once the evaluation metrics were obtained from all iterations, the next step was to calculate the average classification accuracy. The comparison of these average values served as the basis for identifying the most suitable classification method to analyze the dynamics of land-use change in the study area. Subsequently, the raster data were converted into sample points using Rstudio software, and the results were stored in CSV format. This dataset was then used as the foundation for the classification process and land-use change analysis.

The subsequent step involved land-cover prediction. Predictions for the 2019–2023 period were carried out using the best classification method identified from the previous evaluation stage. This prediction process integrated two data sources, namely Sentinel-2A imagery and the digitized dataset. Land-cover change analysis was then conducted by comparing the conditions in 2019 and 2023. The analysis focused on two main types of change: (1) Vegetation (V) to Non-vegetation (NV), representing the conversion of green areas into built-up or non-green areas, and (2) Non-vegetation to Vegetation, reflecting the reconversion of non-green areas back into vegetated land. The results of this analysis provide insights into the dynamics of land-use conversion in Kayu Agung District during the study period.

The classification results were subsequently followed by the identification of spatial distribution patterns using the Quadrant Method [25]. The Quadrant Method is a planar approach in which the study area is divided into grids of equal-sized cells, referred to as quadrants, with the number of points in each cell being random. These quadrants are generally square-shaped. The underlying hypothesis aims to determine whether the points are distributed in a regular, clustered, random, or non-random pattern [26]. The purpose of this method is to analyze the distribution of points in space in order to identify the tendency of their spatial patterns. In principle, there are three possible types of spatial point distribution: (1) random pattern, in which points are scattered without any specific order; (2) regular spatial pattern, in which points are relatively evenly spaced from one another; and (3) clustered spatial pattern, in which points are concentrated in specific groups within certain areas. By applying the Quadrant Method, the distribution of points obtained from classification and digitization can be analyzed, allowing the spatial patterns of land-use change in Kayu Agung District to be more clearly identified. The identification of point distributions using the Quadrant Method can be based on the values of VMR and χ^2. The calculation of VMR is formulated in Equation (12) [27].

$$VMR = \frac{s^2}{\bar{x}} \qquad s^2 = \frac{\sum_{i=1}^{m}(x_i-\bar{x})^2}{(m-1)}, \bar{x} = \frac{N}{m} \quad (12)$$

m denote the number of cells, N the total number of points, x_i the number of points in the i^{th} cell, $\bar{x}$ the average number of points per cell, and s^2 the variance of the number of points per cell. The interpretation criteria for the Variance-to-Mean Ratio (VMR) can be described as follows: (1) VMR = 0 indicates a uniform or systematic distribution of points, (2) VMR = 1 indicates a random distribution, and (3) VMR > 1 indicates a clustered distribution [28]. Although relatively simple, the VMR offers a practical and efficient means of identifying general spatial distribution tendencies. It enables rapid detection of whether a spatial phenomenon exhibits randomness, regularity, or clustering, without the need for complex spatial weight matrices or intensive computation. Consequently, VMR serves

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7281

as an effective preliminary diagnostic tool for spatial pattern analysis. The determination of spatial distribution patterns can also be statistically tested using the chi-square (χ^2) approach, as formulated in Equation (13) [29].

$$\chi^2 = (m-1) \times VMR \quad (13)$$

In testing spatial distribution patterns using the chi-square approach, the null hypothesis (H_0) states that the distribution of points is random or follows a Poisson distribution. The χ^2 test statistic follows a chi-square distribution with degrees of freedom equal to $n-1$, where n denotes the number of observation units. The calculated χ^2 value is then compared with the critical values of the chi-square distribution at a specified significance level, for example $\alpha = 0.05$. If the calculated χ^2 is smaller than the lower critical value, the distribution pattern is classified as uniform. If the calculated χ^2 lies between the lower and upper critical values, the distribution is considered random. Conversely, if the calculated χ^2 exceeds the upper critical value, the point distribution pattern is identified as clustered [30].

RESULTS AND DISCUSSION

A. Spatial Classification Results

As in the research methodology, the first stage involved data preprocessing, which began with downloading satellite imagery, digitization, and the compilation of satellite data with the digitization results. The downloaded Sentinel-2A imagery, particularly the red and near-infrared bands, was combined with the digitization output. An illustration of the integration between the digitized data and the satellite imagery is presented in Table 2.

Table 2. Data Illustration

UTM Zone	x	y	Class	B4	B8
48S	480768	9631319	0	0.072	0.180
48S	480797	9631325	0	0.056	0.144
48S	480815	9631315	0	0.066	0.316
48S	480554	9631244	1	0.074	0.144
48S	480562	9631243	1	0.073	0.188
48S	480553	9631253	1	0.114	0.209

Source : (Research results, 2025)

Table 2 presents an illustration of the extracted data from Sentinel-2A satellite imagery in the study area, using the UTM coordinate system, zone 48S. Each row represents a sample point identified by its coordinates (x, y) and classified into two categories: class 0 for vegetation and class 1 for non-vegetation. In addition to positional and class information, the table also includes reflectance values from two key Sentinel-2A bands, namely Band 4 (Red) and Band 8 (Near Infrared/NIR). These two bands play an essential role in distinguishing vegetation from non-vegetation, as healthy vegetation typically exhibits low reflectance in the red band but high reflectance in the NIR band. For instance, the point with coordinates (480768, 9631319), classified as vegetation, shows a reflectance value of 0.072 (B4) and 0.180 (B8), whereas the point with coordinates (480553, 9631253), classified as non-vegetation, records values of 0.114 (B4) and 0.209 (B8).

In this stage, 1,530 random points were generated from the digitization process. These points were then subjected to data splitting with training and testing proportions of 80%:20%, 85%:15%, and 90%:10%. The training datasets from each split were analyzed using machine learning methods. The applied models included: (1) kNN with *k = 3, 5,* and *7*; (2) Naïve Bayes (NB); and (3) Logistic Regression (LR). Each model and partitioning scheme was repeated 100 times. The evaluation results (both training and testing) for all repetitions are presented in Tables 3 and 4.

Table 3. Evaluation Results on Training Data

No	Method	Proportion	Accuracy	Precision	Recall	F1-Score	Specificity	AUC	Error Rate	MCC
1	kNN (k=3)	80%:20%	0.969	0.975	0.963	0.969	0.975	0.969	0.031	0.938
2			0.966	0.971	0.958	0.964	0.973	0.965	0.034	0.931
⋮			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100			0.968	0.980	0.956	0.968	0.981	0.968	0.032	0.937
101		85%:15%	0.971	0.977	0.965	0.971	0.977	0.971	0.029	0.942
⋮			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
200			0.970	0.976	0.963	0.970	0.977	0.970	0.030	0.940
201		90%:10%	0.971	0.975	0.967	0.971	0.975	0.971	0.030	0.941
⋮			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
300			0.969	0.978	0.959	0.969	0.978	0.969	0.031	0.938
301	kNN (k=5)	80%:20%	0.962	0.967	0.958	0.963	0.967	0.962	0.038	0.925
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
600		90%:10%	0.970	0.972	0.967	0.970	0.972	0.970	0.031	0.937
601	kNN (k=7)	80%:20%	0.960	0.979	0.944	0.961	0.978	0.960	0.040	0.921

No	Method	Proportion	Accuracy	Precision	Recall	F1-Score	Specificity	AUC	Error Rate	MCC
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
900		90%:10%	0.964	0.979	0.950	0.964	0.980	0.965	0.036	0.929
901	NB	80%:20%	0.935	0.933	0.933	0.933	0.936	0.935	0.065	0.869
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1200		90%:10%	0.925	0.928	0.918	0.923	0.932	0.925	0.075	0.850
1201	LR	80%:20%	0.945	0.948	0.940	0.944	0.950	0.945	0.055	0.891
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1500		90%:10%	0.942	0.948	0.934	0.941	0.950	0.942	0.058	0.884

Source : (Research results, 2025)

Table 4. Evaluation Results (on Testing Data)

No	Method	Proportion	Accuracy	Precision	Recall	F1-Score	Specificity	AUC	Error Rate	MCC
1	kNN (k=3)	80%:20%	0.967	0.972	0.959	0.966	0.975	0.967	0.033	0.935
2			0.967	0.964	0.976	0.970	0.957	0.967	0.033	0.934
⋮			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100			0.958	0.981	0.939	0.959	0.979	0.959	0.043	0.916
101		85%:15%	0.948	0.937	0.954	0.946	0.942	0.948	0.052	0.896
102			0.965	0.974	0.957	0.965	0.974	0.965	0.035	0.931
⋮			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
200			0.944	0.939	0.947	0.943	0.940	0.944	0.057	0.887
201		90%:10%	0.954	0.971	0.932	0.951	0.975	0.953	0.046	0.909
⋮			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
300			0.961	1.000	0.926	0.962	1.000	0.963	0.039	0.925
301		80%:20%	0.948	0.934	0.948	0.941	0.948	0.946	0.052	0.894
⋮	kNN (k=5)	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
600		90%:10%	0.941	0.969	0.899	0.932	0.976	0.945	0.059	0.916
601		80%:20%	0.951	0.947	0.940	0.944	0.959	0.951	0.049	0.900
⋮	kNN (k=7)	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
900		90%:10%	0.941	0.961	0.924	0.942	0.960	0.941	0.059	0.883
901		80%:20%	0.892	0.913	0.872	0.892	0.913	0.893	0.108	0.785
⋮	NB	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1200		90%:10%	0.928	0.941	0.929	0.935	0.927	0.928	0.072	0.855
1201	LR	80%:20%	0.935	0.949	0.925	0.937	0.945	0.934	0.065	0.870
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1500		90%:10%	0.961	0.963	0.963	0.963	0.959	0.961	0.039	0.921

Source : (Research results, 2025)

From the classification results, the average values of the evaluation metrics were calculated for each method. As an illustration, the mean values for the testing data are presented in Table 5

Table 5. Average Evaluation Metrics on Testing Data

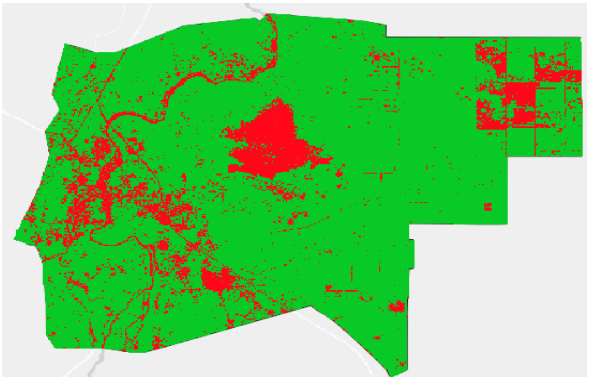
	80% : 20%					85% : 15%					90% : 10%				
	kNN (k=3)	kNN (k=5)	kNN (k=7)	NB	LR	kNN (k=3)	kNN (k=5)	kNN (k=7)	NB	LR	kNN (k=3)	kNN (k=5)	kNN (k=7)	NB	LR
Accuracy	0.96	0.95	0.95	0.92	0.95	0.96	0.95	0.95	0.92	0.95	0.96	0.96	0.96	0.93	0.95
Precision	0.96	0.96	0.96	0.93	0.95	0.96	0.96	0.96	0.92	0.95	0.96	0.96	0.96	0.93	0.95
Recall	0.95	0.94	0.94	0.91	0.94	0.95	0.95	0.95	0.92	0.94	0.96	0.95	0.95	0.92	0.94
F1-Score	0.96	0.95	0.95	0.92	0.94	0.96	0.95	0.95	0.92	0.94	0.96	0.95	0.95	0.92	0.94
Specificity	0.96	0.96	0.96	0.93	0.95	0.96	0.96	0.96	0.93	0.95	0.96	0.96	0.96	0.93	0.95
AUC	0.96	0.95	0.95	0.92	0.95	0.96	0.95	0.95	0.92	0.95	0.96	0.96	0.96	0.93	0.95
Error Rate	0.04	0.05	0.05	0.10	0.06	0.04	0.05	0.05	0.08	0.05	0.04	0.05	0.05	0.07	0.05
MCC	0.91	0.90	0.91	0.84	0.89	0.92	0.91	0.91	0.84	0.89	0.92	0.91	0.91	0.85	0.89

Source : (Research results, 2025)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7281

This method achieved evaluation scores with accuracy, precision, recall, and specificity of 96%, which falls into the category of very good performance, while the error rate was 4% and the MCC reached 92%. After identifying the best-performing model, classification was applied to the entire land cover of Kayu Agung District for the 2019–2023 period. The predicted land-cover changes were then visualized in raster format to facilitate spatial interpretation. As an illustration, the classification result of land cover in 2023 is presented in Figure 2.



Source : (Research results, 2025)
Figure 2. Land-Cover Classification Result in 2023

In Figure 2, the green color represents the vegetation class, while red represents the non-vegetation class. In this study, the vegetation class includes areas of rice fields, plantations, forests, and planted green trees, whereas the non-vegetation class covers residential areas, open land, and roads. Figure 3 provides an illustration of the classification output to facilitate the interpretation of the results.

Source : (Research results, 2025)
Figure 3. Illustration of Classification.

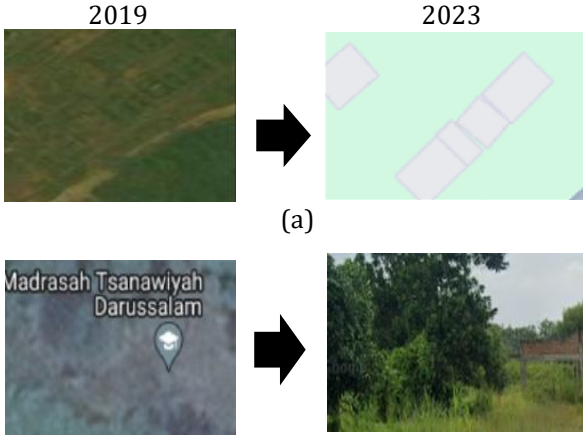
Similar to the previous explanation, Figure 3 clarifies the classification results, with green representing vegetation objects and red representing non-vegetation objects. Furthermore, the classified outputs for each year were used to calculate the area of land-cover change between classes. The results of the land-cover change area calculations from 2019 to 2023 are presented in Table 6.

Table 6. Predicted Land-Cover Area

Year	Vegetation (m^2)	Non-Vegetation (m^2)
2019	195677000	28856100
2020	198475800	26057300
2021	202914400	21618700
2022	203552400	20980700
2023	191368000	33165100

Source : (Research results, 2025)

Quantitatively, the results presented in Table 6 indicate that the vegetated area decreased by approximately 4.31 million m^2 (≈ 430.9 ha) between 2019 and 2023, corresponding to a 1.9-percentage-point reduction in the total land-cover proportion (from 87.15 % to 85.23 %). Conversely, non-vegetated areas expanded by the same magnitude, reflecting a consistent trend of urban expansion and the decline of green open spaces. This net loss of vegetation—averaging about 107 ha per year—suggests increasing ecological pressure and a potential reduction in ecosystem functions such as micro-climate regulation, surface-runoff control, and habitat provision. If this linear trend continues, an additional 5.39 million m^2 (≈ 539 ha) of vegetated land could disappear over the next five years, underscoring the urgency of implementing adaptive land-use and environmental management strategies. As an illustration, evidence of land-cover change—both from vegetation to non-vegetation and from non-vegetation to vegetation—can be observed in the calculated areas of change presented in Visualization 4.



Source : (Research results, 2025)
Figure 4. Land-Cover Change: (a) Vegetation to Non-Vegetation, (b) Non-Vegetation to Vegetation

Figure 4(a) illustrates land-use change from vegetation to non-vegetation. An area that was previously green (as shown using the Bing base map) was converted into a residential zone. In contrast, Figure 4(b) depicts land-use change from non-vegetation to vegetation. An area that was formerly occupied by the MTS Darussalam building (visible in the Google Satellite base map) was transformed into a yard or green area (vegetation), as displayed in Google Maps. These changes provide concrete evidence of the dynamic nature of land-cover transformations in the study area.

B. Identification of Spatial Patterns of Land-Cover Change

Based on the classification results for 2019 and 2023, the spatial distribution patterns of land-cover change over this interval were identified. Figure 5 presents the changes from vegetation (V) to non-vegetation (NV) and vice versa. Figure 5 visualizes the areas of change between 2019 and 2023. Yellow indicates regions with no change (areas classified as vegetation in 2019 that remained vegetation in 2023, and likewise for non-vegetation). Purple represents land-cover change from vegetation to non-vegetation, while blue indicates the reverse, namely change from non-vegetation to vegetation. The distribution of land-cover change in areas that experienced transitions (purple and blue) was then analyzed using the Quadrant Method.

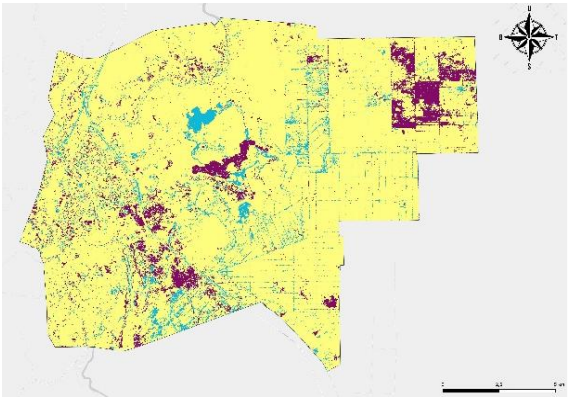
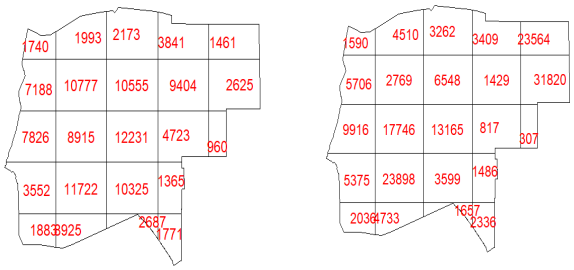


Figure 5. Land-Cover Change from 2019 to 2023

In this method, changes from vegetation to non-vegetation and vice versa were examined separately. The initial step involved converting the raster-based areas into point data, after which the VMR value was calculated using Equation (12). As an illustration, Figure 6 presents the grid used for identifying point distributions in the context of vegetation-to-non-vegetation (and vice versa) changes.



(a) (b)

Source : (Research results, 2025)

Figure 6. Illustration of the Number of Points in Each Grid: (a) Vegetation to Non-Vegetation, (b) Non-Vegetation to Vegetation

Figure 6 illustrates the number of grids (m) for each type of land-cover change, with a total of 6 × 4 = 24 grids. The value in each grid indicates the number of points that experienced change within that grid. Based on this illustration, the VMR values are presented in Equations (13) and (14).

$$VMR_{v\to nv} = \frac{s^2}{\bar{x}} = \frac{\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2}{(m-1)}}{\bar{x}} = \frac{(2336-7646.26)^2+\cdots+(2002-7646.26)^2}{(24-1)} = \frac{7647.26}{} = 8383.277$$

$$VMR_{nv\to v} = \frac{s^2}{\bar{x}} = \frac{\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2}{(m-1)}}{\bar{x}} = \frac{(2441-5593.13)^2+\cdots+(436-5593.13)^2}{(24-1)} = \frac{5593.13}{} = 3679.898$$

Equation (13) represents the VMR value of point distributions from vegetation-to-non-vegetation changes, whereas Equation (14) represents the change from non-vegetation to vegetation, with a total of 24 grids. Since the VMR value varies with the number of grids, calculations were performed using different grid combinations. In this study, a range of grid sizes, from smaller to larger, was applied. The results of the VMR calculations and their statistical tests from these combinations are presented in Table 7.

Table 7. VMR Values and Test Statistics with Combined Grid Sizes

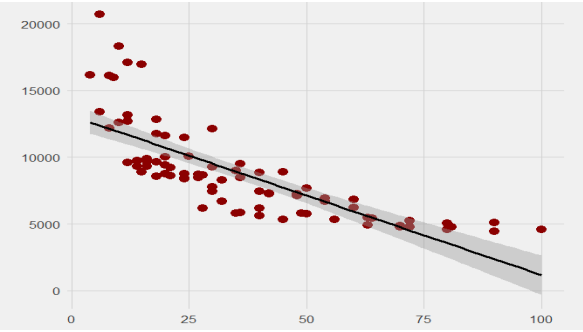
No	Point Distribution	m	VMR	Test Statistics	Critical Value
1	V ke NV	4	16172.11	19791.49	7.81

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7281

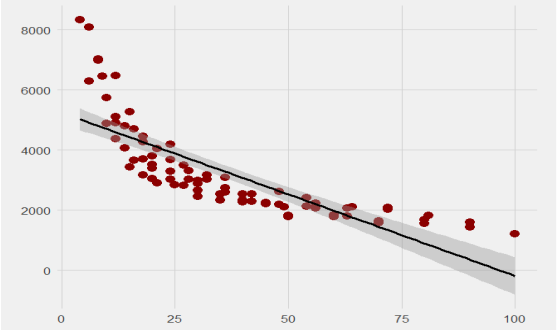
No	Point Distribution	m	VMR	Test Statistics	Critical Value
2		6	20719.24	55044.17	11.07
⋮		⋮	⋮	⋮	⋮
99		90	5083.47	282091.80	112.02
100		100	4583.20	282330.30	123.23
101	NV ke V	4	8332.44	12532.08	7.81
102		6	6297.09	15736.91	11.07
⋮		⋮	⋮	⋮	⋮
199		90	1597.28	78458.50	112.02
200		100	1212.46	66474.16	123.23

Source : (Research results, 2025)

When visualized, the calculated VMR values based on Table 7 are presented in Figure 7. Based on Table 7, the VMR values for both vegetation-to-non-vegetation and non-vegetation-to-vegetation changes, obtained from different grid combinations, are substantially greater than the threshold, indicating that the points are clustered (VMR > 1). Similar results are observed for the χ^2 test statistic, which is greater than its critical value (χ^2 table). This finding is further reinforced by the visualization in Figure 7. In Figure 7, the X-axis represents the grid combinations (ranging from 2 to 100), while the Y-axis indicates the corresponding VMR values. The red dots show the calculated VMR values, whereas the linear line represents the relationship between the number of grids and the VMR values. Taken together, Table 7 and Figure 7 demonstrate that regardless of the grid combinations applied, the conclusion remains consistent: the distribution of points (both from V to NV and vice versa) exhibits a clustered pattern.



(a)



(b)

Source : (Research results, 2025)

Figure 7. VMR Values: (a) Vegetation to Non-Vegetation, (b) Non-Vegetation to Vegetation

The analysis results indicate that land conversion from vegetation to non-vegetation and vice versa tends to form clustered distribution patterns (VMR > 1 and $\chi^2_{calculate} > \chi^2_{table}$). This pattern is consistently evident in Table 7 and the visualization in Figure 7, reflecting the existence of spatial pressure centers. Such clustering is in line with the findings of Qiao et al. (2025), who emphasized the importance of identifying critical zones to support sustainable ecological practices through simulations of land-use change patterns. The concentration of land conversion in specific areas may reduce the ecosystem's capacity to provide essential services, such as climate regulation, water absorption, and habitat provision. Similarly, Li et al. (2025) using the "Production–Living–Ecological Space" approach, demonstrated that ecological space conversion leads to a significant decline in ecosystem service values. These findings highlight the urgent need for mitigation strategies to address the spatially concentrated impacts of land-use change.

The clustered spatial pattern of land conversion has direct implications for the global agenda of SDG 11 (Sustainable Cities and Communities). Cardenas-Ritzert et al. (2024) emphasized that the resolution of land-use data significantly affects the assessment of SDG indicator 11.3.1 on the rate of urban expansion, making accurate monitoring of land-use change essential [34]. These findings indicate that controlling land conversion and improving the quality of spatial data are key elements in realizing inclusive, safe, resilient, and sustainable cities and settlements.

Thus, the identified spatial patterns not only contribute to the advancement of scientific knowledge but also hold strategic significance for the formulation of integrated spatial planning policies grounded in spatial data. The tendency of land conversion to cluster should be anticipated

through adaptive and inclusive spatial planning that integrates data accuracy, legal certainty in policy, and the principles of sustainable development. Integrating research findings with strengthened regulations and SDG reporting mechanisms will enhance the synergy between development needs and efforts toward long-term environmental sustainability.

Overall, the identified spatial patterns not only contribute to the advancement of scientific knowledge but also carry strategic significance for the formulation of integrated spatial planning policies grounded in spatial data. The clustering tendency of land conversion should be addressed through adaptive and inclusive spatial planning that incorporates data accuracy, legal certainty in policymaking, and the principles of sustainable development. Integrating these research findings with strengthened regulations and SDG reporting mechanisms will enhance the synergy between development needs and efforts toward long-term environmental sustainability.

CONCLUSION

Based on the research objectives and findings, this study successfully examined the dynamics of land-use change in Kayu Agung District from 2019 to 2023 using satellite imagery. Among the applied machine learning methods, the K-Nearest Neighbors (KNN) algorithm produced the most accurate classification results. The analysis showed a clear shift from vegetation to non-vegetation areas, driven by increasing development pressure and growing land demand. Spatial analysis using the Quadrant Method revealed that land-use change is not evenly distributed but tends to be spatially clustered, indicating that land conversion is concentrated in specific areas. These findings highlight the importance of controlling land conversion through adaptive and data-driven spatial planning policies. Identifying clustered spatial patterns can help prioritize areas that require stricter monitoring and regulation. Overall, this study provides valuable insights into local land-use change dynamics and supports sustainable spatial planning aligned with national policies and Sustainable Development Goal (SDG) 11 on Sustainable Cities and Communities. Future research should integrate social, economic, and spatial policy variables as key drivers of land-use change to achieve a more holistic understanding of spatial dynamics. Combining the Variance-to-Mean Ratio (VMR) with advanced spatial autocorrelation measures such as Moran's I or Getis-Ord Gi* may further strengthen spatial validation and pattern interpretation. Additionally, the application of advanced machine learning methods, such as Random Forest, Support Vector Machine (SVM), or Deep Learning, could enhance classification performance and ensure more accurate and reliable analytical outcomes.

REFERENCE

[1] BPS-Statistics of Ogan Komering Ilir Regency, *Ogan Komering Ilir Regency in Figures*, vol. 21. BPS-Statistics of Ogan Komering Ilir Regency, 2025.

[2] A. Andikaputra, "Analisis potensi ekonomi dan kependudukan Kabupaten Ogan Komering Ilir," *Jurnal Paradigma Ekonomika*, vol. 17, no. 2, pp. 439–450, 2022, doi: 10.22437/jpe.v17i2.16428.

[3] BPS-Statistics of Ogan Komering Ilir Regency, "Kota Kayu Agung Subdistrict in Figures," 2022.

[4] H. Long, Y. Zhang, L. Ma, and S. Tu, "Land Use Transitions: Progress, Challenges and Prospects," *Land (Basel)*, vol. 10, no. 9, p. 903, Aug. 2021, doi: 10.3390/land10090903.

[5] A. Bikis, M. Engdaw, D. Pandey, and B. K. Pandey, "The Impact of Urbanization on Land Use Land Cover Change using Geographic Information System and Remote Sensing: A Case of Mizan Aman City Southwest Ethiopia," *Sci Rep*, vol. 15, no. 1, p. 12014, Apr. 2025, doi: 10.1038/s41598-025-94189-6.

[6] A. M. Y. Hakim, S. Baja, D. A. Rampisela, and S. Arif, "Modelling land use/land cover changes prediction using multi-layer perceptron neural network (MLPNN): a case study in Makassar City, Indonesia," *International Journal of Environmental Studies*, vol. 78, no. 2, pp. 301–318, 2021, doi: 10.1080/00207233.2020.1804730.

[7] A. Achmad, I. Ramli, S. Sugiarto, I. Irzaidi, and A. Izzaty, "Assessing and Forecasting Carbon Stock Variations in Response to Land Use and Land Cover Changes in Central Aceh, Indonesia," *International Journal of Design and Nature and Ecodynamics*, vol. 19, no. 2, pp. 465–475, Apr. 2024, doi: 10.18280/ijdne.190212.

[8] I. L. Sari, C. J. Weston, G. J. Newnham, and L. Volkova, "Assessing accuracy of land cover change maps derived from automated digital processing and visual interpretation in tropical forests in indonesia," *Remote Sens (Basel)*, vol. 13, no. 8, Apr. 2021, doi: 10.3390/rs13081446.

[9] M. Kelly-Fair *et al.*, "Analysis of Land Use and Land Cover Changes through the Lens of SDGs

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7281

in Semarang, Indonesia," *Sustainability (Switzerland)*, vol. 14, no. 13, Jul. 2022, doi: 10.3390/su14137592.

[10] S. Basheer *et al.*, "Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques," *Remote Sens (Basel)*, vol. 14, no. 19, Oct. 2022, doi: 10.3390/rs14194978.

[11] Z. S. Venter, D. N. Barton, T. Chakraborty, T. Simensen, and G. Singh, "Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover," *Remote Sens (Basel)*, vol. 14, no. 16, Aug. 2022, doi: 10.3390/rs14164101.

[12] F. Gao, M. Anderson, and R. Houborg, "Impacts of Spatial and Temporal Resolution on Remotely Sensed Corn and Soybean Emergence Detection," *Remote Sens (Basel)*, vol. 16, no. 22, p. 4145, Nov. 2024, doi: 10.3390/rs16224145.

[13] R. Verma, "Spatial Data Capture in GIS: A Review," *International Journal of Multidisciplinary Research and Growth Evaluation*, no. 5, pp. 228–233, 2024.

[14] G. Mezzadri, T. Laloë, F. Mathy, and P. Reynaud-Bouret, "Hold-out strategy for selecting learning models: Application to categorization subjected to presentation orders," *J Math Psychol*, vol. 109, p. 102691, Aug. 2022, doi: 10.1016/j.jmp.2022.102691.

[15] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.

[16] R. K. Halder, M. N. Uddin, Md. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J Big Data*, vol. 11, no. 1, p. 113, Aug. 2024, doi: 10.1186/s40537-024-00973-y.

[17] S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, *Data Mining dengan R: Konsep Serta Implementasi*. Bogor: IN MEDIA, 2018.

[18] D. Dey *et al.*, "The Proper Application of Logistic Regression Model in Complex Survey Data: A Systematic Review," *BMC Med Res Methodol*, vol. 25, no. 1, p. 15, Jan. 2025, doi: 10.1186/s12874-024-02454-5.

[19] A. Ostovar, D. D. Davari, and M. Dzikuć, "Determinants of Design with Multilayer Perceptron Neural Networks: A Comparison with Logistic Regression," *Sustainability (Switzerland)*, vol. 17, no. 6, Mar. 2025, doi: 10.3390/su17062611.

[20] B. Sartono and H. Dharmawan, *Pemodelan prediktif berbasis pohon klasifikasi*. Bogor: IPB Press, 2023.

[21] M. S. Sandeep, K. Tiprak, S. Kaewunruen, P. Pheinsusom, and W. Pansuk, "Shear strength prediction of reinforced concrete beams using machine learning," *Structures*, vol. 47, pp. 1196–1211, Jan. 2023, doi: 10.1016/j.istruc.2022.11.140.

[22] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.

[23] D. Chicco and G. Jurman, "A statistical comparison between Matthews correlation coefficient (MCC), prevalence threshold, and Fowlkes–Mallows index," *J Biomed Inform*, vol. 144, p. 104426, Aug. 2023, doi: 10.1016/j.jbi.2023.104426.

[24] J. C. Obi, "A comparative study of several classification metrics and their performances on data," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, Feb. 2023, doi: 10.30574/wjaets.2023.8.1.0054.

[25] S. Wang and L. Dai, "Effect of Quadrat Shape on Spatial Point Pattern Performance of Haloxylon ammodendron," *Open J Ecol*, vol. 14, no. 01, pp. 66–76, 2024, doi: 10.4236/oje.2024.141004.

[26] J. Lochman, "The Spatial Distribution of Sustainable Gastronomy: A Case Study of Tourism in Prague," *Tourism Recreation Research*, vol. 48, no. 5, pp. 693–709, Sep. 2023, doi: 10.1080/02508281.2021.1949676.

[27] G. S. Fivash *et al.*, "Increasing spatial dispersion in ecosystem restoration mitigates risk in disturbance-driven environments," *Journal of Applied Ecology*, vol. 59, no. 4, pp. 1050–1059, Apr. 2022, doi: 10.1111/1365-2664.14116.

[28] C. Santos, "How the Variance-to-Mean Ratio Behaves in Relation to the Basic Principles for Inequality Measures," *WSEAS Transactions on Business And Economics*, vol. 21, pp. 2354–2362, Nov. 2024, doi: 10.37394/23207.2024.21.194.

[29] P. M. B. Cahusac, "Likelihood Ratio Test and the Evidential Approach for 2 × 2 Tables," *Entropy*, vol. 26, no. 5, p. 375, Apr. 2024, doi: 10.3390/e26050375.

[30] J. M. Blochberger, "Characterization of Beech Bark Disease on Camels Hump Mountain, Vermont," 2025.

[31] X. Qiao *et al.*, "Simulating land use change for sustainability: a case study of the northern slope of Tianshan Mountains," *Int J Digit Earth*, vol. 18, no. 1, Dec. 2025, doi: 10.1080/17538947.2025.2524055.
[32] Z. Li, B. Zhang, J. Luo, L. Yang, C. Liu, and Q. Deng, "Research on the synergistic evolution of land use transformation and ecosystem service value in the Anning River Basin," *Sci Rep*, vol. 15, no. 1, p. 28430, Aug. 2025, doi: 10.1038/s41598-025-12657-5.
[33] O. S. E. Cardenas-Ritzert, J. C. Vogeler, S. Shah Heydari, P. A. Fekety, M. Laituri, and M. R. McHale, "Effects of Land Use Data Spatial Resolution on SDG Indicator 11.3.1 (Urban Expansion) Assessments: A Case Study Across Ethiopia," *Sustainability*, vol. 16, no. 22, p. 9698, Nov. 2024, doi: 10.3390/su16229698.
[34] The United Nations, "The Sustainable Development Goals Report: Special Edition Towards a Rescue Plan for People and Planet," United Nations, Department of Economic and Social Affairs Statistics Division, Jul. 2025. Accessed: Aug. 25, 2025. [Online]. Available: https://unstats.un.org/sdgs/report/2025/.