JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 2, April 2025, pp. 346-361

Analyzing Multiclass Land Cover and Spatial Point Patterns on Sentinel-2 Imagery Using Machine Learning and Deep Learning

Muna Faizatun Nabilah[1], Achmad Fauzan[1*]
[1]Department of Statistics, Universitas Islam Indonesia, Indonesia
achmadfauzan@uii.ac.id

Article History:
Received : 08-02-2025
Revised : 03-04-2025
Accepted : 03-04-2025
Online : 23-04-2025

Keyword:
Classification;
Sentinel-2;
Land Use;
Point Distribution Pattern.

ABSTRACT

Land cover conversion around educational centers, such as universities, is an inevitable consequence of increasing urban activity. The development of boarding houses, commercial zones, and other infrastructure often follows the expansion of academic institutions. To support sustainable spatial planning, early identification of land cover and analysis of spatial distribution patterns are crucial for zoning regulation and infrastructure management. This study focuses on classifying land cover and analyzing spatial patterns around Universitas Riau (UNRI) using Sentinel-2 satellite imagery with a 10-meter spatial resolution. The research applied a supervised classification approach, utilizing spectral bands—specifically Near-Infrared (NIR) and Short-Wave Infrared (SWIR)—as explanatory variables. The response variable was land cover, categorized into vegetation, non-vegetation, and water. Three machine learning models—Support Vector Machine (SVM), Naïve Bayes (NB), and Backpropagation Neural Network (BNN)—were compared based on overall accuracy and the Kappa coefficient. The models were trained and tested using a stratified 80-20 data split to ensure a balanced evaluation. Among the models, SVM demonstrated the highest accuracy, achieving an average of 91.15% in 2022 and 83.90% in 2023 with minimal variance, confirming its reliability for land cover classification. Spatially, non-vegetation areas were concentrated near major access routes and facilities, highlighting the influence of infrastructure development on land conversion. The study also identified potential growth zones within a 3–5 km radius from UNRI, emphasizing the need for anticipatory and sustainable land use policies. These findings support the formulation of spatial strategies aligned with Law No. 26 of 2007 on Spatial Planning and offer valuable insights for guiding urban development around academic hubs.

Crossref
https://doi.org/10.31764/jtam.v9i2.29683
This is an open access article under the CC–BY-SA license

A. INTRODUCTION

In Indonesia, especially in Pekanbaru, Riau University (UNRI) not only functions as an education centre but also plays a role in encouraging economic and social growth in the surrounding area. However, increasing activity in this area has triggered changes in land use, especially the conversion of vegetation areas into built-up land (Hapsary et al., 2021). These changes give rise to various spatial challenges, such as reduced green open space, increased surface temperatures, and decreased air quality. These conditions have the potential to affect the comfort and sustainability of the academic ecosystem, so research is needed to understand the impact of land change on the educational environment and formulate appropriate mitigation strategies.

This phenomenon is in line with various studies which show that changes in land use can have a negative impact on the environment and biodiversity. For instance, Thakrar (2024) in Environmental Science & Technology reported that deforestation and agricultural expansion contribute to significant air quality deterioration through increased emissions. Similarly, Galindo et al. (2022) in Ecosphere demonstrated that converting natural vegetation into agricultural land leads to measurable declines in species richness and ecosystem function. In addition, an article in *Diversity and Distributions* highlights how land-use changes, such as deforestation and habitat fragmentation, negatively impact local climatic conditions. These changes—such as increased temperatures and reduced humidity—can alter species composition and accelerate biodiversity loss (Williams & Newbold, 2020).

This land change is a major concern in sustainable spatial planning, especially in educational areas that are experiencing rapid infrastructure development. Therefore, data-based monitoring and analysis of land changes is a crucial step in supporting campus environmental planning and management. Remote sensing offers an effective solution in monitoring land change by enabling data collection without direct contact with the object being observed. This technology enables more extensive, efficient and sustainable monitoring, making it the main tool in analyzing changes in land cover and its spatial patterns (Elachi & Zyl, 2021; Lillesand et al., 2015) .

One of the high-resolution satellites widely used in environmental research is Sentinel-2. This satellite is designed to provide high-quality multispectral data to support environmental research and natural resource management (Immitzer et al., 2016; Mandanici & Bitelli, 2016). Sentinel-2 has a spatial resolution of up to 10 meters with superior multispectral capabilities. Additionally, its temporal resolution improved from 10 days to 5 days after the launch of Sentinel-2B, making it a reliable tool for monitoring land use changes (Agency, 2024; Phiri et al., 2020; Phiri & Morgenroth, 2017). With its open data access policy, Sentinel-2 enables efficient and sustainable land monitoring, thereby supporting various research and applications in remote sensing.

Several studies have utilized Sentinel-2 satellite imagery for various analyses, particularly in mapping and monitoring land changes. One such study was conducted Bayas et al. (2022), which highlighted the importance of Land Use Land Cover (LULC) analysis in detecting changes and monitoring natural resources. This study employed four classification algorithms: Naïve Bayes (NB), Classification and Regression Trees (CART), Gradient Tree Boost (GTB), and Random Forest (RF). Another study by Abbas et al. (2023) discussed the impact of accelerated urbanization on LULC changes and how remote sensing-based modeling can be used to understand historical change patterns and predict future change dynamics. Meanwhile, research by Zhang et al. (2021) highlighted the challenges in LULC classification using Sentinel-2 data. Additionally, this study evaluated the effectiveness of Bayesian parameter optimization in improving classification performance using the RF algorithm.

However, studies on the application of Sentinel-2 in educational areas are still rarely conducted, particularly in analyzing land distribution patterns around such areas. This study aims to classify land cover around the University of Riau (UNRI) and analyze spatial distribution patterns to support data-driven planning decisions. Referring to Law No. 26 of 2007 concerning Spatial Planning, the results of this study are expected to provide a significant contribution in supporting data-driven land management in rapidly developing regions. These findings offer critical insights for formulating spatial policies that promote balanced land use and sustainable urban development around educational centers.

B. METHODS

1. Machine Learning Based Classification

This study compares several classification models, namely: Support Vector Machine (SVM), Naïve Bayes (NB), and Backpropagation Neural Network (BNN). SVM is a supervised machine learning method designed to determine the optimal hyperplane to separate two classes of data in binary classification, based on machine learning theory (Li & Zhang, 2024; Urso et al., 2019). In multiclass classification, SVM applies the One-vs-All (OvA) method, where several hyperplanes (i) are constructed to separate one particular class from another. Each hyperplane assigns a label of +1 to the target class being processed, while the other classes are assigned a label of -1 (Krebs et al., 2024; Rifkin & Klautau, 2004). The OvA method simplifies the solution of multiclass classification problems by dividing them into several binary classification subproblems, thereby increasing efficiency in data processing (Alber et al., 2017; Shajari & Rangarajan, 2020). Suppose the training data is expressed as (x_i, y_i), where $x_i \epsilon \mathcal{R}^p$ is the input data and $y_i \epsilon \{-1,1\}$ is the class label of x_i, this problem can be solved by Equation (1).

$$\min \frac{1}{2} ||w^2|| \tag{1}$$

with the criteria,

$$y_i(w^i . x_i + b^i) \geq 1, for\ \forall_i \tag{2}$$

Naïve Bayes (NB) is a simple probabilistic prediction method, which applies the principle of Bayes' theorem and assumes that all features used for prediction are independent (Han et al., 2012). The main objective of the classification model using NB is to maximize the posterior probability of the target class $P(Y_j|X_i)$ based on the training data (Raschka, 2017). The posterior probability $(Y_j|X_i)$ can be written using Bayes' theorem presented in Equation (3).

$$P(Y_j|X_i) = \frac{P(Y_j)P(X_i|Y_j)}{P(X_i)} \tag{3}$$

X_i is the feature vector for the i-th instance, Y_j is the j-th target class, $P(X_i|Y_j)$ is the independent probability of class Y_j for all features in the vector X_i. The predicted class is determined based on the class Y_j that has the maximum posterior probability, which can be expressed in the notation form in Equation (4).

$$\text{predicted class label } (\widehat{Y}) = \arg\max P(Y_j)\, P(X_i|Y_j) \tag{4}$$

BNN is a machine learning algorithm designed for artificial neural networks with a feedforward multi-layers structure. Backpropagation is known as an efficient training method that updates the weights and biases of a network (Jayaraman et al., 2009). The steps involved in the backpropagation algorithm is given in Table 1.

Table 1. Steps of the backpropagation algorithm.

Algorithm		
Step 1		Determine the network architecture, initial weight initialization, and learning rate.
Step 2	Feed forward	The input unit receives the signal and passes it to the units in the hidden layer. The weighted sum and output are then calculated using the activation function.
Step 3	Backward propagation	Compute the error in the output unit based on the difference between the target and output values. The error is carried to the hidden layer.
Step 4		Calculate the error in the hidden units based on the contribution from the output layer, multiplied by the derivative of the activation function.
Step 5	Updating the weight and bias	Update the weights and biases for each output and hidden unit based on the computed error information and learning rate.
Step 6		Evaluate whether the stopping condition has been met. If not, repeat the process from steps 2 to 5 for several epochs or until convergence is achieved.

2. Confusion Matrix

Confusion matrix (CM) is an evaluation tool used to assess the performance of a classification model by comparing the model's prediction results to actual data. This matrix consists of four main components that represent the classification results, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) (Luque et al., 2019). Model performance can be evaluated using four main metrics which are Accuracy, Sensitivity, Specificity, and Area Under the Curve (AUC). The formulas of the four main metrics is obtained from Equations (5) to Equations (8).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \text{ x } 100\% \tag{5}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{ x } 100\% \tag{6}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \text{ x } 100\% \tag{7}$$

$$\text{AUC} = \frac{1 + \text{sensitivity}}{2} \text{ x } 100\% \tag{8}$$

The value interval of the four metric values is between 0 and 1 (0% to 100%), the closer to 100% the classification model formed, the better (Alwazy et al., 2025).

3. Points Distribution Analysis

Analysis of the distribution of life and activities of living things on the Earth's surface is an important aspect in understanding the distribution patterns and changes in land use (Cabernard et al., 2024). One approach used to evaluate spatial distribution patterns is quadrant analysis, which divides an area into a quadrant-shaped grid for systematic analysis (Wang & Dai, 2024). However, the main challenge in this method is the subjectivity in determining the size and number of quadrants, which can affect the significance of the analysis results (Hao et al., 2020). Therefore, an appropriate approach is needed to ensure accurate and well-interpretable results. In quadrant analysis, the two main methods that are often used are Variance Mean Ratio (VMR) and hypothesis testing (Aidi, 2013). VMR is a simple method for determining spatial distribution patterns. Based on the VMR value, the spatial point distribution pattern is categorized as follows: clustered if VMR > 1, regular if VMR < 1, and random if VMR = 1. The VMR value can be calculated using Equation (9).

$$\mathrm{VMR} = \frac{\mathrm{s}^2}{\bar{\mathrm{x}}} \tag{9}$$

m is the number of quadrants, N is the number of points, x_i is the number of points in cell i, $\bar{x}$ is the average number of points per cell, s^2 is the variance of the number of points per cell. In addition, you test statistics is employed as presented in Equation (10).

$$\mathrm{X}^2 = (\mathrm{m}-1)\mathrm{VMR} \tag{10}$$

Reject H_0 if $\mathrm{X}^2_{\text{statistic}} > \mathrm{X}^2_{\text{critical value}}$ (Aidi, 2013).

4. Data and Research Flowchart

The data samples used in this study were collected from Sentinel-2 satellite imagery of the area bordering UNRI. The sampling method was purposive sampling with the limitation being the availability of samples in each class. This research adopts a case study approach, focusing on land cover classification and spatial pattern analysis in the area surrounding the University of Riau (UNRI). The research stages began with determining the area of the research object, followed by a clipping process to focus the research area covering the area bordering UNRI. The research flowchart is presented in Figure 1.

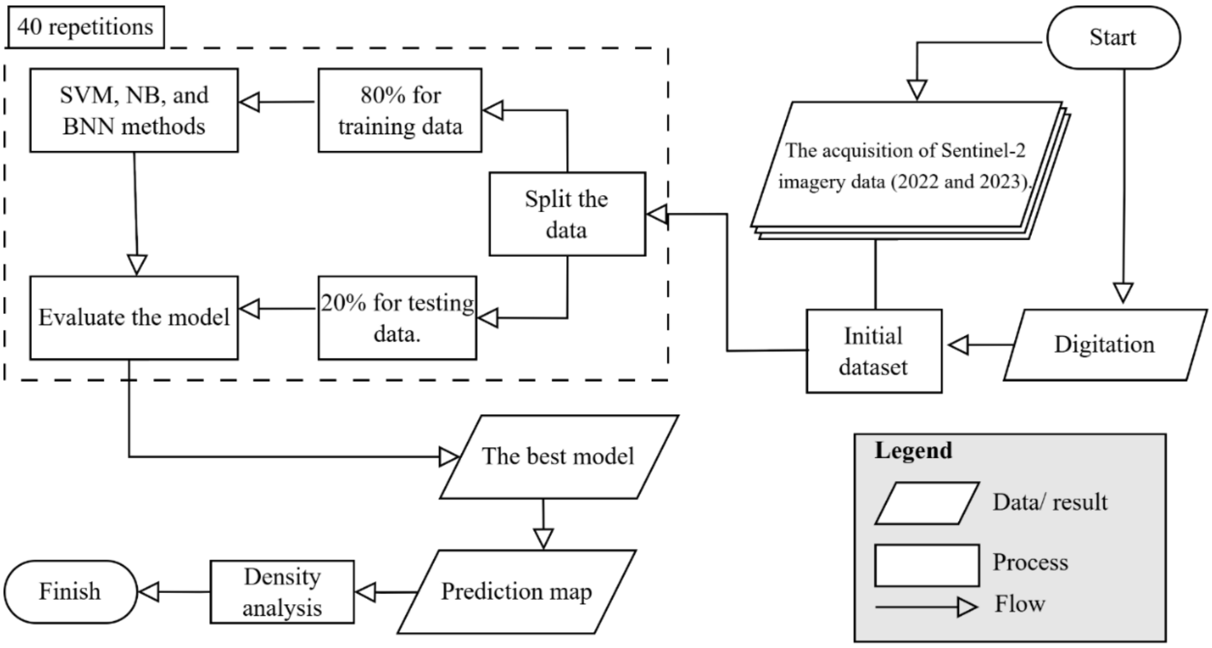


Figure 1. The research flowchart

The key spectral bands used as explanatory variables were Near-Infrared (NIR) and Short-Wave Infrared (SWIR), while the response variable was land cover, categorized into vegetation, non-vegetation, and water bodies. Points that fall into the vegetation category such as rice fields, forests, yards, or objects overgrown with plants. Non-vegetation objects include houses, roads, offices, or objects related to buildings. While water objects include reservoirs, rivers, or objects related to surface water bodies. The data was acquired on March 19, 2022 and April 13, 2023 under optimal atmospheric conditions without cloud interference. The resampling process was performed on the SWIR Band to adjust its spatial resolution to 10 meters to align with the highest resolution in Band 8. This adjustment is important because the 10-meter resolution allows for more detailed analysis of the earth's surface, especially in monitoring the development of urban areas and agricultural activities (Phiri et al., 2020).

After clipping and image processing, the extracted data on the NIR Band and SWIR Band were obtained. In addition, an analysis of the area that had been cut through the digitization process was also conducted to obtain more accurate spatial information. With this method, the study can provide a clearer picture of changes in land use around UNRI and its implications for the development of urban areas. Furthermore, the bands from satellite imagery were merged with the digitization results which we call the dataset. Table 2 presents an illustration of the dataset used.

Table 2. Illustration of Research Dataset

No	x (easting) mE	y (northing) mN	UTM Zone	Class	NIR	SWIR
1	763707.4	52570.47	47N	Water	0.2400	0.125768
2	763753.2	52145.12	47N	Water	0.3188	0.165858
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1002	764661.5	52049.63	47N	Non-vegetation	0.3932	0.19508
1003	765930.7	52722.76	47N	Non-vegetation	0.3452	0.178301
⋮	⋮	⋮	⋮	⋮	⋮	⋮

No	x (easting) mE	y (northing) mN	UTM Zone	Class	NIR	SWIR
2999	764208.8	52540.08	47N	Vegetation	0.2952	0.249660
3000	763940.8	52180.89	47N	Vegetation	0.2548	0.166384

In the third step, data resampling is performed using the holdout method, namely by randomly dividing the dataset into training data and test data. This resampling aims to optimize the model training process while evaluating its performance. The data division ratio is designed in such a way that the model can recognize complex patterns in the training data and provide an overview of accuracy when applied to new data outside of the scope of training (Joseph, 2022; Sivakumar et al., 2024). The fourth step is the repeated classification process using three machine learning models, namely SVM, NB, and BNN, each with specific parameter configurations. The SVM model uses a linear kernel with a cost parameter (C) of 1 to balance tolerance for misclassification and model complexity. The NB model assumes a Gaussian distribution, so it is suitable for continuous predictor variables. The BNN model has a 2-2-4-3 architecture with a sigmoid activation function in the hidden and output layers, and was trained with a learning rate of 0.1 to optimize convergence.

The selection of the best model is based on two main criteria, namely: (1) the smallest average difference in accuracy between training data and testing data, and (2) the smallest average difference in variance. The model with the best performance is then used to predict the entire area around UNRI in 2022 and 2023. Furthermore, a spatial distribution analysis of building objects around UNRI is conducted using Equations (9) and (10) to evaluate the distribution pattern of the classification results. This analysis aims to understand the distribution pattern of objects around UNRI and identify trends in land use changes based on the classification results obtained.

C. RESULTS AND DISCUSSION

1. Data Classification

The classification process often produces varying outputs due to random elements. Therefore, repetition is carried out to ensure the consistency of the performance of each classification result and to avoid the possibility that the results obtained are merely coincidence. This approach also helps minimize the impact of random fluctuations, provides a clearer picture of the model's ability to recognize data patterns, and strengthens model validation under various conditions (Athmaja et al., 2017; Caruana & Niculescu-Mizil, 2006). Through the repetition process, it is expected that the selected model will not only have high accuracy but also show consistency and be able to generalize reliably to new data (Varoquaux, 2018). Figures 2 and 3 present illustrations of several SVM model evaluation metrics on 2022 data. Several evaluation metric values of the SVM model for 2022 and 2023 are presented in Figure 2 and Figure 3. These visualizations provide a comprehensive overview of the model performance, including trends and variations over the analysis period.

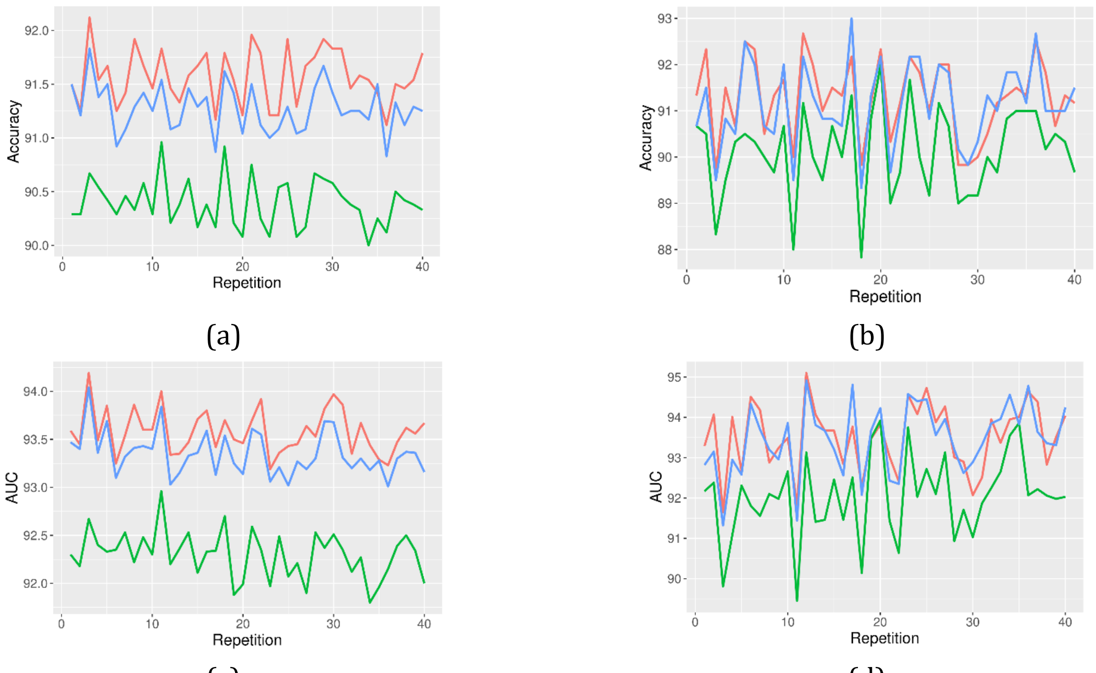


Figure 2. Line of the SVM model evaluation metrics for the year 2022, (a) accuracy of the training data, (b) accuracy of the training data, (c) AUC of the training data, and (d) AUC of the testing data.

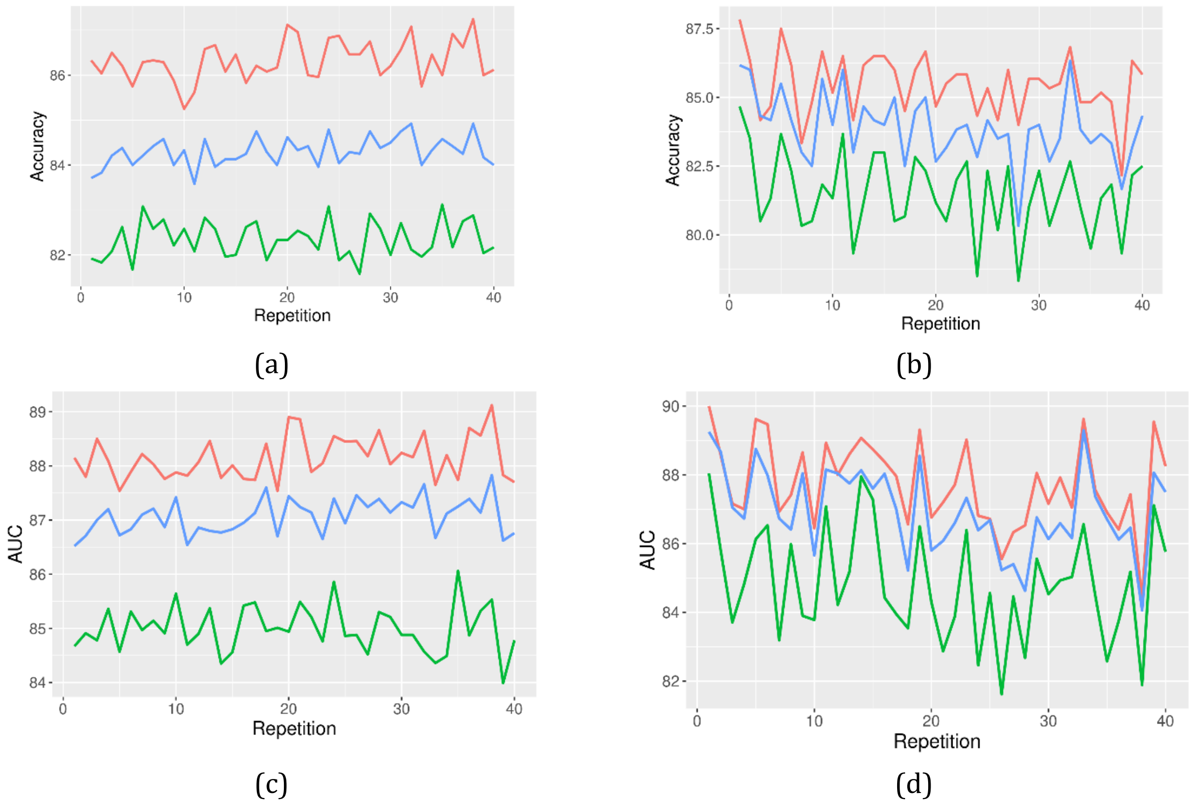


Figure 3. Line of the SVM model evaluation metrics for the year 2023, (a) accuracy of the training data, (b) accuracy of the training data, (c) AUC of the training data, and (d) AUC of the testing data.

Based on Figure 2 and Figure 3, the color red represents the BNN model, blue represents the SVM model, and green represents the NB model.

2. Method Comparison

After obtaining the classification results from each model, the next step is to select the model with the best performance. As explained in the methods section, model selection is based on two main criteria, the smallest average difference in accuracy between training data and test data and the smallest average difference in variance. This selection process is conducted through descriptive analysis of the repetition results of each classification method, with the main focus on the accuracy indicator. With this approach, the selected model is expected to not only have high accuracy but also show consistency in various data scenarios. A summary of the accuracy results from each model is presented in Table 3.

Table 3. Summary of accuracy values of each model.

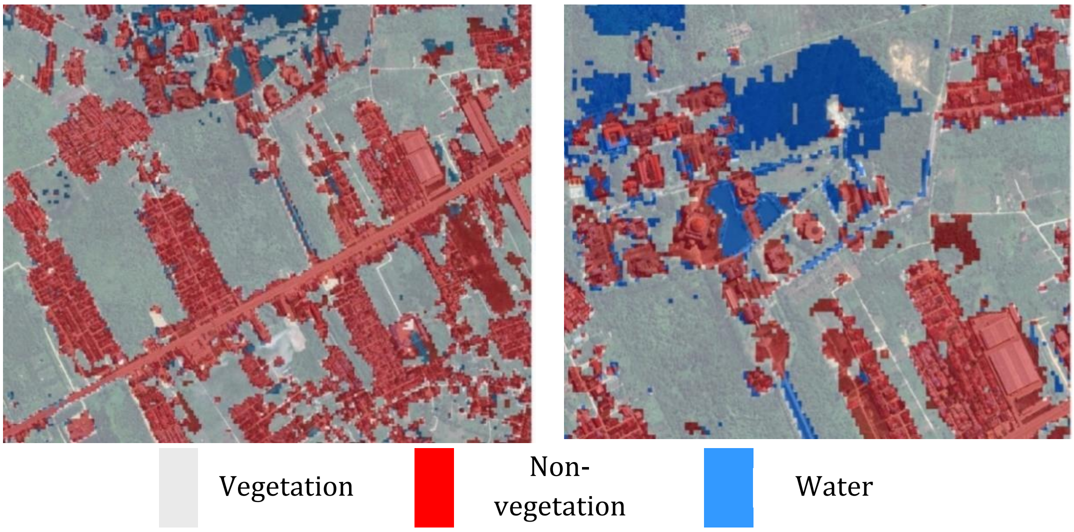
Model	2022				2023			
	Training		Testing		Training		Testing	
	Avg. Accuracy	Avg. Variance	Avg. Accuracy	Avg. Variance	Avg. Accuracy	Avg. Variance	Avg. Accuracy	Avg. Variance
SVM	91.28	0.05	91.15	0.80	84.30	0.10	83.90	1.48
NB	90.39	0.04	90.12	0.86	82.35	0.16	81.50	2.02
BNN	91.57	0.06	91.28	0.69	86.32	0.19	85.46	1.24

According to Table 3, in 2022, the average difference in accuracy between training data and testing data for the SVM model was 0.12, while for the NB and BNN models it was 0.28 and 0.29, respectively. Meanwhile, in 2023, the average differences in accuracy for the SVM, NB, and BNN models were 0.40, 0.85, and 0.87, respectively. Based on the comparative analysis of the average accuracy values obtained from each classification method, the BNN model showed the highest accuracy compared to other models. However, in the context of stability and generalization ability, SVM gave more significant results.

The SVM model recorded the smallest average accuracy difference, which was 0.12 on 2022 data and 0.40 on 2023 data. This difference indicates that SVM has a better ability to maintain consistent performance when applied to different datasets. In addition, the variance analysis shows that SVM has a relatively small variance difference compared to other classification methods. This indicates that SVM is more resistant to fluctuations in data, thus providing more stable and consistent results under various conditions. Model evaluation not only focuses on accuracy but also considers consistency in performance. A model with high accuracy but experiencing large fluctuations in different datasets may indicate instability and poor generalization ability (Athmaja et al., 2017). Therefore, in this study, SVM with high accuracy and low variance is considered as the most optimal classification method. The stability of SVM performance increases its reliability under various data conditions, making it a superior choice compared to other models in terms of generalization and consistency of results. The findings of this study are consistent with previous research, which reported that the SVM method outperforms both BNN and NB approaches (Erlin et al., 2014; Ifriza & Sam'an, 2021; Kaul & Raina, 2022; Millennianita et al., 2024). Therefore, this study is largely supported by existing literature while also offering a novel contribution in the context of spatial classification around the UNRI area, which has been relatively underexplored in prior studies.

3. Image Classification Results

The SVM model was then used to predict 835,152 data in 2022 and 322,320 data in 2023. Each data includes spatial information consisting of x (easting), y (northing) coordinates, Universal Transverse Mercator (UTM) zones, and prediction result classes. The UTM zone used in this study is zone 47N, which covers the area of Pekanbaru City, Riau Province. Geographically, the location is at coordinates 0°25'-0°45' North Latitude (LU) and 101°14'-101°34' East Longitude (BT). The classification outputs were visualized in the form of raster maps, as depicted in Figure 4. These visualizations demonstrate the spatial distribution of land cover classes for the years 2022 and 2023, providing a comprehensive overview of landscape dynamics within the study area. The predicted land cover is categorized into three main classes: surface water, vegetation, and non-vegetation. In the raster visualization, surface water areas—such as rivers, lakes, reservoirs, and swamps—are represented in blue. Vegetation, encompassing forests, grasslands, agricultural fields, and other green-covered areas, is shown in bright green. Conversely, the red color indicates non-vegetation classes, which include built-up areas, roads, infrastructure, and residential settlements.



(a) (b)

Figure 4. Predicted classification results, (a) year 2022, (b) year 2023 (source: data processing results).

Ability of the classification results, a visual comparison was conducted between the predicted raster outputs and the official basemap of the study area. This comparison revealed a high degree of spatial consistency, suggesting that the SVM model successfully captures land cover patterns that align closely with actual conditions observed in the field. Such validation is critical in spatial modeling, as it provides evidence of the model's robustness and applicability in real-world scenarios (Ghoggali & Melgani, 2009). Moreover, the integration of high-resolution spatial prediction and visual validation not only enhances the credibility of the model outputs but also highlights the model's potential in supporting decision-making processes related to land management and spatial planning. By offering accurate insights into land cover distribution and changes over time, the results of this study contribute meaningfully to environmental monitoring efforts and can serve as a valuable reference for sustainable urban development and land use policy formulation in Riau and its surrounding regions.

4. Density Analysis

Geoprocessing analysis using the buffer technique was conducted to identify areas based on distance from the central location. This technique allows measuring the spatial impact of land use and environmental factors at various distances, thus providing an effective approach in understanding the relationship between land use change, spatial distribution, and density (Su et al., 2015; Z. Zhang et al., 2021). Thus, buffer analysis provides in-depth insights into environmental impacts and density distribution within an area with a predetermined radius. In this study, buffers with radii of 2, 3, 4, and 5 km were formed around UNRI. The results of the analysis include the area within each radius and the percentage of points within that range. A summary of the results of the buffer analysis is presented in Table 4, which illustrates the spatial distribution of objects based on distance from the study center.

Table 4. Buffering analysis.

Dist. (km)	Area (km^2)	2022			2023		
		Non-vegetation (%)	Vegetation (%)	Water (%)	Non-vegetation (%)	Vegetation (%)	Water (%)
2	9628	59.37	37.34	34.02	53.72	34.35	12.04
3	25064	54.97	41.81	33.01	45.29	32.28	10.30
4	46675	50.26	45.33	28.65	31.82	23.69	7.072
5	74460	41.86	42.16	23.72	20.10	15.05	4.468

The analysis of land use patterns around UNRI, as presented in Table 4, indicates notable changes between the years 2022 and 2023. A particularly high concentration of non-vegetation land cover—representing built-up environments and impervious surfaces—was observed within the 2 km radius, reaching 59.37% in 2022 and slightly decreasing to 53.72% in 2023. This pattern reflects a persistent, albeit potentially shifting, trend of urbanization and infrastructure development concentrated around the university area. The elevated proportion of non-vegetation cover in the immediate vicinity emphasizes UNRI's role as a catalyst for urban expansion, likely driven by growing demands for housing, commercial establishments, and transportation infrastructure.

Concurrently, the proportion of vegetated areas showed a consistent decline across all buffer zones, with the steepest reductions occurring within the 2–3 km radius. Within 2 km, for instance, vegetation cover dropped from 37.34% in 2022 to 34.35% in 2023. This decrease suggests a progressive transformation of green spaces—such as woodlands, grasslands, and open fields—into urban land uses, which raises concerns over habitat loss, biodiversity decline, and reduced provision of ecosystem services near the campus.

A particularly alarming trend is the sharp decrease in surface water coverage. In the 2 km buffer, water bodies accounted for 34.02% of the area in 2022, but this figure plummeted to 12.04% in 2023. Such a drastic reduction may be attributed to factors such as land reclamation, urban infrastructure encroachment, or the desiccation of small wetlands and seasonal water features. As these water bodies are vital to agriculture, aquaculture, and maintaining hydrological balance, their loss could significantly undermine environmental sustainability and threaten the livelihoods of surrounding communities.

The rising prevalence of non-vegetated land cover also introduces a variety of socio-environmental issues, including increased traffic congestion, higher volumes of waste, deteriorating air and water quality, and growing strain on water supply systems. If left unaddressed, these pressures may degrade urban living conditions and further destabilize local ecosystems. Hence, comprehensive spatial planning and proactive environmental management are essential. The buffer-based geospatial analysis applied in this study provides a strategic framework for identifying zones at risk due to environmental stress and urban encroachment. These insights can support the formulation of spatial interventions that balance the need for development with the imperative of ecological protection. Striking this balance is crucial for promoting sustainable regional growth and enhancing the long-term well-being of local populations.

5. Points Distribution Analysis

In addition to conducting spatial classification, a key distinguishing feature of this study lies in its integration of density analysis, with particular emphasis on the non-built-up land class. This added analytical dimension not only enhances the interpretability of the classification results but also provides deeper insights into spatial distribution patterns that are often overlooked in conventional classification studies. The analysis of the distribution of service centers was accomplished using a geographic approach through quadrant modelling. This approach aims to describe the distribution pattern of locations based on the number of points and variations in quadrant size (Aidi, 2013). In this study, the analysis included the identification of 39,201 points in 2022 and 34,871 points in 2023, with variations in quadrant size between 9 and 400. These points represent the distribution of settlements or non-vegetation areas in the Simpang Baru Village area. The results of the quadrant analysis in both years show dynamics in the distribution pattern of non-vegetation areas, which can provide insight into the trend of settlement development in the study area. The data from the quadrant analysis in 2023 are presented in Table 5.

Table 5. Quadrant analysis of 2023 data.

Quadrant Dimensions	Mean	Variance	VMR	$X^2_{statistic}$	$X^2_{critical\ value}$
3 x 3	3874.556	1946730	502.4395	4188.9	15.50731
⋮	⋮	⋮	⋮	⋮	⋮
20 x 20	94.24595	4402.809	46.71617	878.7736	1.96

The results of the quadrant analysis shown in Table 5 describe the distribution of non-vegetation areas in Simpang Baru Village in 2023. This distribution pattern tends to be clustered, as indicated by the results of the Chi-square test which shows $X^2_{statistic} > X^2_{critical\ value}$ as stated in Equation (10). In addition, the Variance Mean Ratio (VMR) value which is consistently greater than 1 indicates a rejection of the null hypothesis (H_0) which means that the spatial distribution tends to be concentrated.

The clustering of non-vegetated areas is concentrated in areas with high accessibility, as shown in Figure 4. The highest concentrations are found along main roads, around UNRI, health facilities, and commercial zones. This reflects the significant influence of accessibility and the existence of essential services in driving urban growth. As the size of the quadrant increases, the VMR value decreases, indicating a development pattern that is increasingly spreading to the outskirts. These data indicate that areas within a radius of 3–5 km from UNRI, which currently still have lower density, have the potential to experience rapid growth in the future. The results of this analysis are very relevant in supporting urban planning that is oriented towards the development of new service centers in order to distribute development more evenly and reduce pressure on denser urban zones. These findings offer critical insights for formulating spatial policies that promote balanced land use and sustainable urban development around educational centers.

D. CONCLUSION AND SUGGESTIONS

This research identifies changes in land cover around the University of Riau (UNRI) and analyzes object distribution patterns using Sentinel-2 satellite imagery. Data were analyzed using three classification models: Support Vector Machine (SVM), Naïve Bayes (NB), and Backpropagation Neural Network (BNN). The analysis results show that the SVM model has the best performance in mapping land cover around the University of Riau (UNRI), with the highest average accuracy of 91.15% in 2022 and 83.90% in 2023. This model also has the smallest difference in average accuracy, so it is more reliable than other methods. Land distribution analysis reveals that within a 2 km radius of UNRI, the area of non-vegetation has decreased from 59.37% in 2022 to 53.72% in 2023, while green space has decreased from 37.34% to 34.35%. The same trend occurs within a radius of 3–5 km, with a significant reduction in vegetation and water bodies due to the expansion of built-up areas. The greatest decline occurred within a 4–5 km radius, where non-vegetation areas were reduced by more than 50%, while green spaces and water bodies experienced further degradation. In addition, the distribution of non-vegetated areas concentrated around main service centers and transportation routes indicates the intensification of development around the campus, as well as areas with high growth potential within a radius of 3–5 km which reflects the dynamics of land use changes that need to be managed sustainably.

This change in land cover emphasizes the need for more adaptive spatial planning policies to maintain a balance between development and environmental sustainability. Stricter zoning regulations are needed to control the conversion of green spaces into built-up areas, as regulated in Law no. 26 of 2007 concerning Spatial Planning. In addition, the development of green infrastructure, such as expanding urban forests and providing green open space of at least 20% of the total campus area, can help reduce the ecological impact of urbanization. As an academic institution, UNRI also plays a role in supporting environmental sustainability through implementing green campus policies, green space conservation, sustainable waste management, and the use of renewable energy. This step is in line with Law no. 32 of 2009 concerning Environmental Protection and Management, which emphasizes the importance of integrating environmental aspects in higher education policy and governance.

To increase the accuracy of land change analysis, it is recommended that further research use a multi-temporal approach with a longer time coverage (5–10 years) to understand change trends more comprehensively. The use of more sophisticated classification techniques, such as Convolutional Neural Networks (CNN) or Random Forest with hyperparameter optimization, can also improve mapping accuracy. In addition, the integration of spatial data with socio-economic indicators, such as population growth, changes in land prices, and local development policies, can provide deeper insights into urban dynamics and their impact on the environment.

By implementing this strategy, sustainable land management can be achieved, ecosystem balance can be maintained, and the development of more environmentally friendly areas can be realized.

ACKNOWLEDGEMENT

The author would like to thank the Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia for various assistances provided. Hopefully, the results of this study can provide significant contributions to the development of science.

REFERENCES

Abbas, H., Tao, W., Khan, G., Alrefaei, A. F., Iqbal, J., Albeshr, M. F., & Kulsoom, I. (2023). Multilayer perceptron and Markov Chain analysis based hybrid-approach for predicting land use land cover change dynamics with Sentinel-2 imagery. *Geocarto International*, *38*(1), 1–27. https://doi.org/10.1080/10106049.2023.2256297

Agency, E. S. (2024). *Sentinel-2*. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2

Aidi, M. N. (2013). *Konfigurasi Titik dalam Ruang Kajian Teoretis, Simulasi dan Kasus*. Institut Pertanian Bogor. https://repository.ipb.ac.id/jspui/bitstream/123456789/68973/1/Gabungan konfigurasi titik ruang.pdf

Alber, M., Zimmert, J., Dogan, U., & Kloft, M. (2017). Distributed optimization of multi-class SVMs. *PLoS ONE*, *12*(6), 1–18. https://doi.org/10.1371/journal.pone.0178161

Alwazy, A. S. H., Buyrukoğlu, G., Buyrukoğlu, S., & Baker, M. R. (2025). Evaluating machine learning and statistical learning techniques for cancer classification and diagnosis. *Iran Journal of Computer Science*, 1–20. https://doi.org/10.1007/s42044-025-00233-z

Athmaja, S., Hanumanthappa, M., & Kavitha, V. (2017). A survey of machine learning algorithms for big data analytics. *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 1–4. https://doi.org/10.1109/ICIIECS.2017.8276028

Bayas, S., Sawant, S., Dhondge, I., Kankal, P., & Joshi, A. (2022). Land Use Land Cover Classification Using Different ML Algorithms on Sentinel-2 Imagery. *Lecture Notes in Electrical Engineering*, *858*, 761–777. https://doi.org/10.1007/978-981-19-0840-8_59

Cabernard, L., Pfister, S., & Hellweg, S. (2024). Biodiversity impacts of recent land-use change driven by increases in agri-food imports. *Nature Sustainability*, *7*(11), 1512–1524. https://doi.org/10.1038/s41893-024-01433-4

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *ACM International Conference Proceeding Series*, *148*, 161–168. https://doi.org/10.1145/1143844.1143865

Elachi, C., & Zyl, J. (2021). *Introduction to the Physics and Techniques of Remote Sensing* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781119523048

Erlin, E., Rahmiati, R., & Rio, U. (2014). Two Text Classifiers in Online Discussion: Support Vector Machine vs Back-Propagation Neural Network. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *12*(1), 189. https://doi.org/10.12928/TELKOMNIKA.v12i1.1798

Galindo, V., Giraldo, C., Lavelle, P., Armbrecht, I., & Fonte, S. J. (2022). Land use conversion to agriculture impacts biodiversity, erosion control, and key soil properties in an Andean watershed. *Ecosphere*, *13*(3), e3979. https://doi.org/10.1002/ecs2.3979

Ghoggali, N., & Melgani, F. (2009). Automatic ground-truth validation with genetic algorithms for multispectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), 2172–2181. https://doi.org/10.1109/TGRS.2009.2013693

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier. https://doi.org/10.1016/C2009-0-61819-5

Hao, L., Qingdong, S., Imin, B., & Kasim, N. (2020). Methodology for optimizing quadrat size in sparse vegetation surveys: A desert case study from the Tarim Basin. *PLOS ONE*, *15*(8), e0235469. https://doi.org/10.1371/journal.pone.0235469

Hapsary, M. S. A., Subiyanto, S., & Firdaus, H. S. (2021). Analisis Prediksi Perubahan Penggunaan Lahan dengan Pendekatan Artificial Neural Network dan Regresi Logistik di Kota Balikpapan. *Jurnal Geodesi Undip*, *10*(2), 88–97. https://doi.org/10.14710/jgundip.2021.30637

Ifriza, Y. N., & Sam'an, M. (2021). Performance comparison of support vector machine and gaussian naive bayes classifier for youtube spam comment detection. *Journal of Soft Computing Exploration*, *2*(2), 93–98. https://doi.org/10.52465/joscex.v2i2.42

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, *8*(3), 166. https://doi.org/10.3390/rs8030166

Jayaraman, S., Thangaraj, V., & Esakkirajan, S. (2009). *Digital Image Processing* (1st ed.). Tata McGraw Hill Education.

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining*, *15*(4), 531–538. https://doi.org/10.1002/sam.11583

Kaul, A., & Raina, S. (2022). Support vector machine versus convolutional neural network for hyperspectral image classification: A systematic review. *Concurrency and Computation: Practice and Experience*, *34*(15), 1–35. https://doi.org/10.1002/cpe.6945

Krebs, R., Bagui, S. S., Mink, D., & Bagui, S. C. (2024). Applying Multi-CLASS Support Vector Machines: One-vs.-One vs. One-vs.-All on the UWF-ZeekDataFall22 Dataset. *Electronics*, *13*(19), 3916. https://doi.org/10.3390/electronics13193916

Li, Y., & Zhang, L. (2024). A Novel Loss Function-based Support Vector Machine for Binary Classification. *ArXiv*, 1–17. https://doi.org/10.48550/arXiv.2403.16654

Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation* (7th ed.). John Wiley & Sons.

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Mandanici, E., & Bitelli, G. (2016). Preliminary comparison of sentinel-2 and landsat 8 imagery for a combined use. *Remote Sensing*, *8*(12), 1–10. https://doi.org/10.3390/rs8121014

Millennianita, F., Athiyah, U., & Muhammad, A. W. (2024). Comparison of Naïve Bayes Classifier and Support Vector Machine Methods for Sentiment Classification of Responses to Bullying Cases on Twitter. *Journal of Mechatronics and Artificial Intelligence*, *1*(1), 11–26. https://doi.org/10.17509/jmai.v1i1.69959

Phiri, D., & Morgenroth, J. (2017). Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sensing*, *9*(9), 967. https://doi.org/10.3390/rs9090967

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

Raschka, S. (2017). Naive Bayes and Text Classification I - Introduction and Theory. *ArXiv*, 1–20. https://doi.org/10.48550/arXiv.1410.5329

Rifkin, R., & Klautau, A. (2004). In Defense of One-Vs-All Classification. *Journal of Machine Learning Research*, *5*, 101–141. https://dl.acm.org/doi/10.5555/1005332.1005336

Shajari, H., & Rangarajan, A. (2020). A Unified Framework for Multiclass and Multilabel Support Vector Machines. *ArXiv*, 1–27. https://doi.org/10.48550/arXiv.2003.11197

Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ Computer Science*, *10*, 1–17. https://doi.org/10.7717/peerj-cs.2245

Su, Z. H., Lin, C., Ma, R. H., Luo, J. H., & Liang, Q. O. (2015). Effect of land use change on lake water quality in different buffer zones. *Applied Ecology and Environmental Research*, *13*(3), 639–653. https://doi.org/10.15666/aeer/1303_639653

Thakrar, S. K., Johnson, J. A., & Polasky, S. (2024). Land-Use Decisions Have Substantial Air Quality Health Effects. *Environmental Science & Technology*, *58*(1), 381–390. https://doi.org/10.1021/acs.est.3c02280

Urso, A., Fiannaca, A., La Rosa, M., Ravì, V., & Rizzo, R. (2019). Data Mining: Classification and Prediction. In *Encyclopedia of Bioinformatics and Computational Biology* (Vol. 1, pp. 384–402). Elsevier.

https://doi.org/10.1016/B978-0-12-809633-8.20461-5

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, *180*(Part A), 68–77. https://doi.org/10.1016/j.neuroimage.2017.06.061

Wang, S., & Dai, L. (2024). Effect of Quadrat Shape on Spatial Point Pattern Performance of Haloxylon ammodendron. *Open Journal of Ecology*, *14*(01), 66–76. https://doi.org/10.4236/oje.2024.141004

Williams, J. J., & Newbold, T. (2020). Local climatic changes affect biodiversity responses to land use: A review. *Diversity and Distributions*, *26*(1), 76–92. https://doi.org/10.1111/ddi.12999

Zhang, T., Su, J., Xu, Z., Luo, Y., & Li, J. (2021). Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier. *Applied Sciences*, *11*(2), 543. https://doi.org/10.3390/app11020543

Zhang, Z., Zhang, F., Du, J., Chen, D., & Zhang, W. (2021). Impacts of land use at multiple buffer scales on seasonal water quality in a reticular river network area. *PLOS ONE*, *16*(1), 1–20. https://doi.org/10.1371/journal.pone.0244606