Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.5.331-348

***Corresponding author:**
Email address:
Lennt@hau.edu.vn

Received: 26/06/2026
Received in Revised Form: 14/10/2026
Accepted: 03/11/2026

Object-based LULC classification and urban dynamics analysis using machine learning on sentinel-2 imagery: A case study of selected wards in Hanoi, Vietnam

Le Thi Minh Phuong[1], Nguyen Thanh Len[2*], Le Huyen Trang[3]
[1]Head of Department of Science and Technology, Hanoi Architectural University, Hanoi 100000, Vietnam
[2]Faculty of Urban Infrastructure and Environmental Engineering, Hanoi Architectural University, Hanoi 100000, Vietnam; lennt@hau.edu.vn
[3]Intelligent Geosystems and Infrastructure Research Group, University of Transport Technology, Hanoi 100000, Vietnam

Abstract: This study investigates land use/land cover (LULC) classification and urban spatial dynamics in three wards: Thanh Xuan, Khuong Dinh, and Phuong Liet, Hanoi, Vietnam, using object-based image analysis (GEOBIA) integrated with machine learning on Sentinel-2 imagery. Sentinel-2 images acquired in 2015, 2020, and 2025 were preprocessed, segmented using the Simple Non-Iterative Clustering (SNIC) algorithm, and used to extract spectral, textural, and shape features at the object level. Four classifications, algorithms including Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Tree Boosting (GTB) were applied and compared. All models achieved satisfactory accuracy; however, RF outperformed the others and showed the most consistent validation results, with overall accuracy ranging from 82.62% to 87.35% and Kappa coefficients from 0.793 to 0.831. The RF-based results indicate that built-up land is the dominant class and has expanded significantly, increasing from 91.04 ha in 2015 to 100.88 ha in 2025, while water bodies, shrubland, agricultural land, and bare land declined. These findings demonstrate the effectiveness of GEOBIA combined with machine learning for analyzing LULC changes in complex urban environments.

Keywords: GEOBIA, Land use/land cover (LULC), Sentinel-2, SNIC, Machine learning, Random Forest, Urban dynamics, Hanoi.

1. Introduction

Rapid urbanization in Hanoi, particularly within the inner-city areas, has significantly altered land use structures, most notably through the expansion of built-up land at the expense of green spaces, agricultural land, and bare land. In this context, monitoring land use/land cover (LULC) changes plays a crucial role in land resource management, urban planning, and environmental assessment [1, 2]. In addition, Hanoi is characterized by high building density, fragmented spatial patterns, and a high degree of heterogeneity among land cover types, making accurate LULC classification particularly challenging.

In urban remote sensing studies, pixel-based

classification methods often face limitations due to small object sizes, spectral similarity, and spatial heterogeneity [3, 4]. To address these challenges, Geographic Object-based Image Analysis (GEOBIA) has been widely adopted, as it integrates spectral, textural, and shape information at the object level [3, 5]. With Sentinel-2 imagery (10 m spatial resolution), this approach offers strong potential for improving classification accuracy in complex urban environments [2, 6]. Accordingly, this study employs the Simple Non-Iterative Clustering (SNIC) image segmentation algorithm and extracts spectral, GLCM-based textural, and shape features as inputs for machine learning models [7, 8].

In addition to methodological selection, comparing classification algorithms is essential for identifying the most suitable model for the study area. Four machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Tree Boosting (GTB) were applied to evaluate LULC classification performance across three time points (2015, 2020, and 2025). The results provide a basis for identifying the most reliable model and analyzing LULC change trends associated with urbanization [9-12]

Although previous studies have widely applied machine learning and object-based approaches for LULC classification, several limitations remain in densely urbanized environments. First, many studies have focused primarily on pixel-based classification or single-date analyses, which are often insufficient for capturing the spatial complexity and temporal dynamics of highly fragmented urban landscapes [13, 14]. Second, comparative assessments among machine learning algorithms under object-based frameworks remain limited, particularly for Sentinel-2 imagery in rapidly urbanizing areas of Vietnam [15]. In addition, the integration of spectral, textural, and shape features derived from SNIC segmentation has not been comprehensively evaluated across multiple temporal periods in the context of inner-city land transformation [13, 16]. As a result, the relative effectiveness and stability of different machine learning models for urban LULC mapping over time remain insufficiently understood.

From a theoretical perspective, this study contributes to the growing body of GEOBIA-based urban remote sensing research by demonstrating the applicability of object-based machine learning classification in heterogeneous tropical urban environments. The study also provides a comparative framework for evaluating the performance consistency of RF, SVM, CART, and GTB models across different temporal datasets. Furthermore, by integrating SNIC segmentation with multi-feature extraction from Sentinel-2 imagery, the research highlights the importance of object-level spatial characteristics in improving classification reliability for complex urban landscapes. These findings contribute to methodological discussions on urban LULC mapping and provide a transferable approach for similar rapidly urbanizing cities in developing countries. Based on this framework, the study has two main objectives: (i) to evaluate the effectiveness of object-based classification combined with machine learning using Sentinel-2 data, and (ii) to analyze LULC changes in the study area from 2015 to 2025, thereby revealing the dynamics of urban expansion in inner Hanoi.

2. Materials and Methods

2.1. Data Used

a) Study Area

The study area comprises three wards: Thanh Xuan, Khuong Dinh, and Phuong Liet, located in Hanoi, Vietnam, covering a total area of approximately 11.11 km^2 (3.77 km^2 for Thanh Xuan, 3.59 km^2 for Khuong Dinh, and 3.75 km^2 for Phuong Liet). Situated in the southwestern part of the inner city, the area benefits from a well-developed transportation network and is adjacent to major roads and densely populated residential

zones. The study area is located between 20°58'-21°01' N latitude and 105°48'-105°50' E longitude (Fig. 1). The official administrative boundaries of Vietnam, including its mainland, sea areas, and islands (Hoang Sa and Truong Sa archipelagos), were obtained from the Department of Survey, Mapping and Geographic Information Vietnam, Ministry of Natural Resources and Environment. Given the complex urban spatial structure, small object sizes, diverse shapes, and similar spectral characteristics among land cover types, the study area is well suited for GEOBIA. This approach enables the integration of spectral, textural, and shape features, thereby improving the discrimination of land use classes in highly fragmented urban environments.

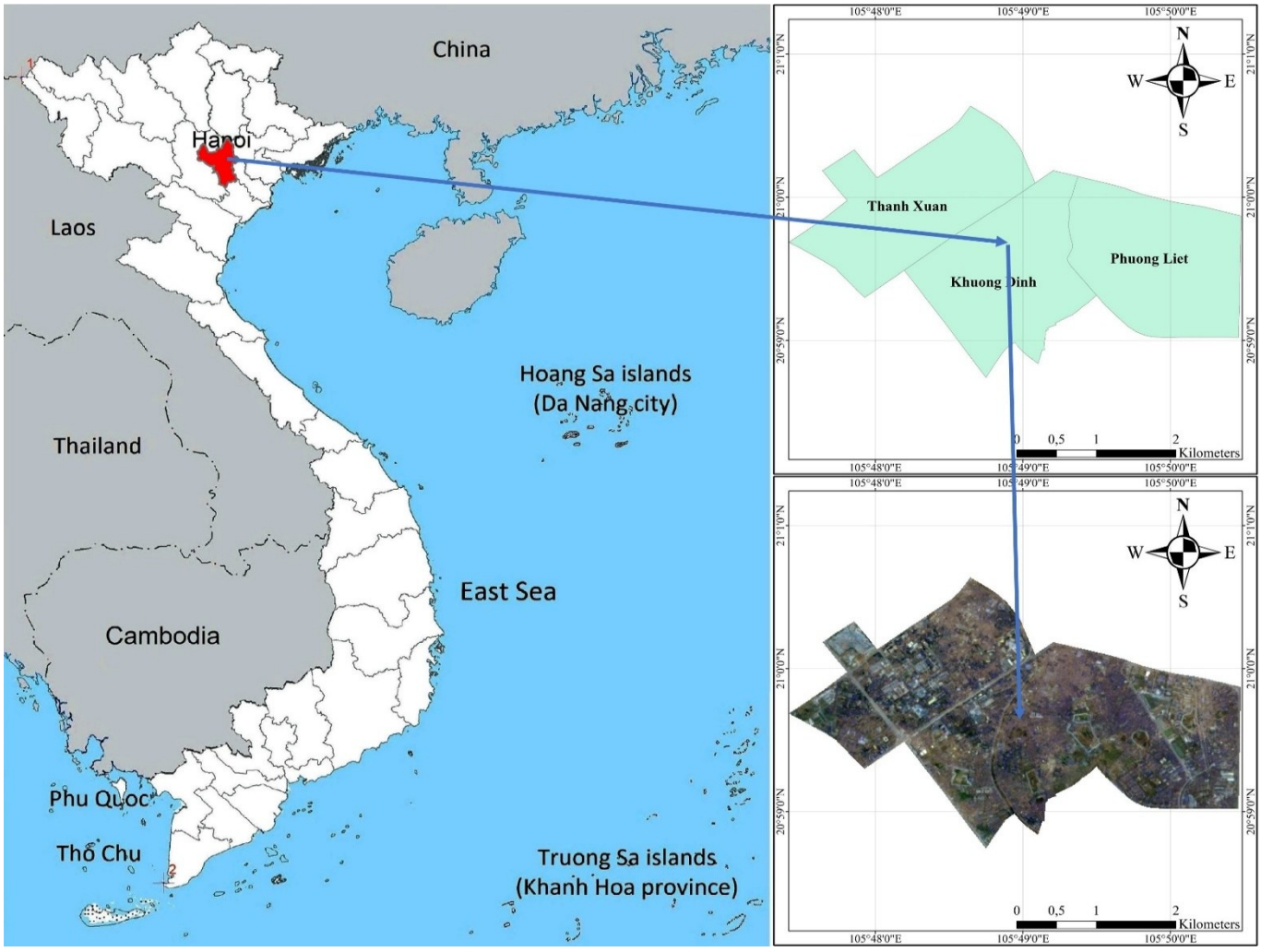


Fig. 1. Location map of the study area (Administrative boundaries were obtained from the Vietnam National Spatial Data Infrastructure, Ministry of Natural Resources and Environment).

b) Data

This study integrates remote sensing data, spatial data, and reference samples to support object-based LULC classification and urban change analysis. Sentinel-2 surface reflectance imagery (COPERNICUS/S2_HARMONIZED product) acquired over the study area was utilized as the primary data source for multi-temporal land cover classification across the three target years (2015, 2020, and 2025). Specifically, Sentinel-2 images acquired on 17 August 2015, 6 August 2020, and 20 August 2025 were selected based on low cloud cover (<20%) and consistent seasonal acquisition windows during the dry season to minimize atmospheric and moisture variations, with detailed metadata summarized in Table 1 (Fig. 2). To ensure high data quality, an automated cloud and shadow masking pipeline was implemented on

Google Earth Engine. Cloud screening was performed using the sentinelCloudScore algorithm with a cloud score threshold (cloudThresh) of 30, incorporating spectral thresholds on blue, cirrus, and coastal aerosol bands alongside normalized difference indices (NDMI and NDSI). Cloud shadows were dynamically projected and masked based on solar azimuth and zenith angles across multiple height intervals (200 m to 10,000 m) combined with dark pixel filtering (irSumThresh = 0.35). Surface reflectance bands were rescaled by a factor of 0.0001 to convert digital numbers to standard surface reflectance values (0-1). Daily mosaics were generated to eliminate orbit overlaps, and a median reducer was applied to construct clean, cloud-free seasonal image composites for each reference year. Furthermore, historical high-resolution imagery from Google Earth and extensive field verification data were collected to establish reference sample sets, ensuring strong temporal appropriateness and representation for 2015, 2020, and 2025.

Table 1. Technical specifications of the Sentinel-2 imagery used in the study

Metadata Attribute	2015 Dataset	2020 Dataset	2025 Dataset
Satellite / Sensor	Sentinel-2A / MSI	Sentinel-2A / MSI	Sentinel-2B / MSI
Acquisition Date	August 17, 2015	August 06, 2020	August 20, 2025
Spatial Resolution	10 m		
Cloud Cover (%)	2.5%	1.8%	3.1%
Spectral Bands Utilized	Blue (B2), Green (B3), Red (B4), NIR (B8), SWIR1 (B11)		
Study Area Coverage	Thanh Xuan, Khuong Dinh, Phuong Liet wards (11.11 km^2)		

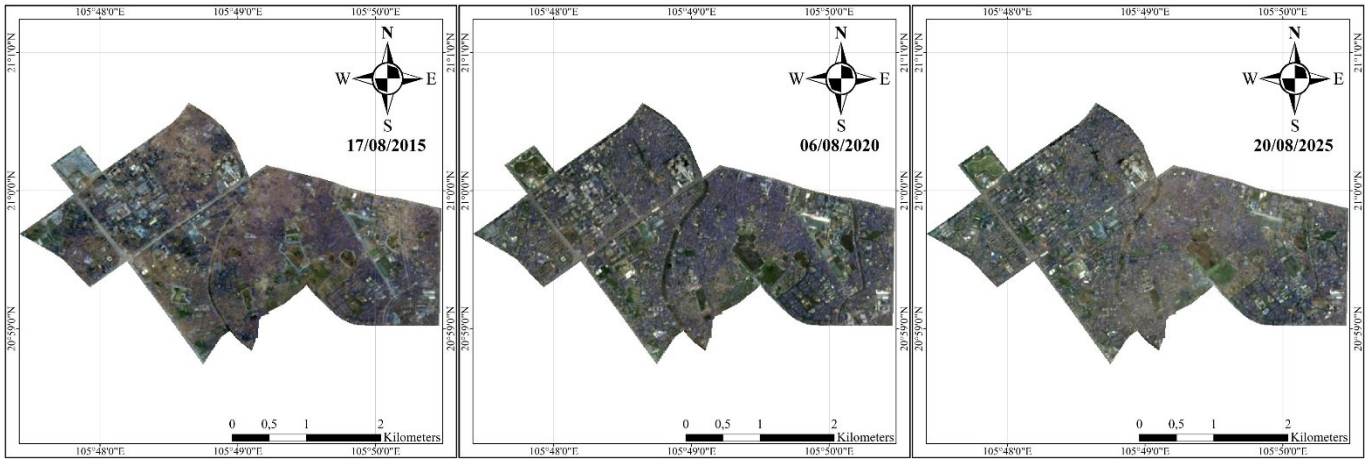


Fig. 2. Sentinel-2 imagery of the study area for the years 2015, 2020, and 2025

Table 2. Distribution of training and validation samples across LULC classes (2015, 2020, and 2025)

LULC Class	Training Samples (70%)	Validation Samples (30%)	Total Samples
Water	56	24	80
Shrubland	49	21	70
Agricultural land	49	21	70
Built-up land	77	33	110
Bare land	49	21	70
Total	280	120	400

Training and validation samples were generated for each target year (2015, 2020, and 2025) from multiple sources, including high-resolution imagery available on Google Earth, expert knowledge, and field verification. Samples were digitized as points or polygons representing

each LULC class and distributed across the study area using a stratified spatial sampling strategy. To ensure high quality, total sample sizes ranged from 70 to 110 samples per class, yielding a total of 400 reference samples per year (Table 2). For each respective year, the reference dataset was independently partitioned into training (70%, 280 samples) and validation (30%, 120 samples) subsets. Class sample sizes were proportioned to reflect the spatial representation of each land cover class, with higher sample density allocated to the dominant and heterogeneous Built-up class.

To guarantee the statistical independence of the validation dataset and prevent spatial autocorrelation and spatial information leakage between neighboring objects, training and validation samples were spatially segregated by applying a minimum spatial separation distance (buffer zone) of 100 m between training and validation objects. Furthermore, image segmentation was executed before sample partitioning, ensuring that individual segmented polygons were strictly assigned to either the training or the validation subset, but never shared across both sets. This spatial restriction eliminates adjacency effects and ensures a robust and truly independent evaluation of model performance

For object-based classification, input variables included both original spectral bands and derived features. Specifically, spectral bands B2 (blue), B3 (green), B4 (red), and B8 (near-infrared) were used. To enhance class separability between heterogeneous urban cover types, three standard spectral indices were calculated from the Sentinel-2 surface reflectance bands: the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Built-up Index (NDBI), and the Normalized Difference Water Index (NDWI) [17-19]. The mathematical formulations and input bands utilized are defined as follows:

- NDVI was computed using the near-infrared (B8) and red (B4) bands:

$$\mathrm{NDVI} = \frac{\mathrm{B8\ (NIR)} - \mathrm{B4\ (Red)}}{\mathrm{B8\ (NIR)} + \mathrm{B4\ (Red)}} \quad (1)$$

- NDBI was derived using the short-wave infrared (B11) and near-infrared (B8) bands:

$$\mathrm{NDBI} = \frac{\mathrm{B11\ (SWIR1)} - \mathrm{B8\ (NIR)}}{\mathrm{B11\ (SWIR1)} + \mathrm{B8\ (NIR)}} \quad (2)$$

- NDWI was calculated using the green (B3) and near-infrared (B8) bands:

$$\mathrm{NDWI} = \frac{\mathrm{B3\ (Green)} - \mathrm{B8\ (NIR)}}{\mathrm{B3\ (Green)} + \mathrm{B8\ (NIR)}} \quad (3)$$

These indices effectively accentuate vegetation vigor, built-up infrastructure, and surface water bodies, respectively, thereby improving feature representation for the subsequent machine learning classification models.

Table 3. LULC classes

LULC Class	Description
Water	Rivers, lakes, ponds, and other surface water bodies within the study area
Shrubland	Low vegetation, shrubs, and scattered urban green spaces
Agricultural land	Cropland and remaining cultivated areas within the urban region
Built-up land	Residential areas, buildings, infrastructure, and other urban surfaces
Bare land	Unused land, vacant lots, and areas under construction

In addition, textural features were derived from the gray-level co-occurrence matrix (GLCM), including Contrast, Entropy, Angular Second Moment (ASM), and Correlation, to capture spatial variability and surface homogeneity. Specifically, the GLCM was calculated based on a grayscale image extracted from Sentinel-2 bands using a sliding window size of 3×3 pixels, with gray levels quantized to 32 levels to optimize computational efficiency. Texture statistics were computed at a pixel distance of d = 1 pixel and averaged across four directional angles (0°, 45°, 90°, and 135°) to

ensure rotation invariance. To bridge pixel-level measures with the object-based framework, pixel-level GLCM values within each segmented object were aggregated by calculating their mean values. Furthermore, shape features such as area, perimeter, and compactness were extracted from segmented objects, contributing to improved classification performance in complex urban environments.

2.2. Methods used

a) Theoretical background of machine learning models and accuracy assessment metrics

RF: RF is an ensemble machine learning algorithm proposed by Breiman (2001). The method constructs multiple decision trees using bootstrap samples randomly drawn from the training dataset and combines their predictions through a majority voting mechanism. By aggregating the outputs of numerous decision trees, RF effectively reduces overfitting and improves model generalization performance. RF is capable of handling high-dimensional datasets, nonlinear relationships, and noisy data without requiring assumptions regarding data distribution. Due to these advantages, RF has been widely applied in remote sensing image classification, LULC mapping, and environmental monitoring.

SVM: SVM is a supervised learning algorithm developed by Cortes and Vapnik (1995). The fundamental principle of SVM is to identify an optimal hyperplane that maximizes the margin between different classes in the feature space. For datasets that are not linearly separable, SVM employs kernel functions to transform the original data into a higher-dimensional feature space, where class separation becomes more feasible. Common kernel functions include linear, polynomial, and radial basis function (RBF) kernels. SVM has demonstrated strong performance in remote sensing applications due to its ability to handle complex nonlinear classification problems and achieve high accuracy even with limited training samples.

CART: CART is a decision-tree-based algorithm introduced by Breiman et al. (1984). CART recursively partitions the dataset into increasingly homogeneous subsets by selecting the optimal splitting variables according to criteria such as the Gini Index or Entropy. For classification tasks, the algorithm generates a binary tree structure in which each internal node represents a decision rule, and each terminal node corresponds to a class label. CART is easy to interpret, computationally efficient, and capable of modeling nonlinear relationships. However, it may be prone to overfitting when the tree becomes excessively complex.

GTB, also known as Gradient Boosting Machine (GBM), is an ensemble learning technique proposed by Friedman (2001). Unlike RF, which builds decision trees independently, GTB constructs trees sequentially, where each new tree is trained to reduce the residual errors generated by the previous trees. The learning process is based on gradient descent optimization, which iteratively minimizes a predefined loss function. By combining multiple weak learners into a strong predictive model, GTB can effectively capture complex nonlinear relationships and interactions among predictor variables. Due to its high predictive capability and robustness, GTB has been widely used in remote sensing classification, environmental modeling, and spatial prediction studies.

Accuracy assessment metrics:

Overall Accuracy (OA) represents the proportion of correctly classified samples relative to the total number of validation samples [20]. It is one of the most widely used indicators for evaluating classification performance.

$$OA = \frac{\sum_{i=1}^{r} x_{ii}}{N} \quad (4)$$

where:

x_{ii} is the number of correctly classified samples in

class i.

r is the total number of classes.

N is the total number of validation samples.

OA values range from 0 to 1 (or 0-100%), with higher values indicating better classification accuracy.

Kappa Coefficient, proposed by Cohen (1960), measures the agreement between classified results and reference data while accounting for agreement occurring by chance [21].

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (5)$$

where:

P_o is the observed agreement (OA).

P_e is the expected agreement due to random chance.

Kappa values closer to 1 indicate higher classification accuracy and stronger agreement with the reference data.

b) Methods

To provide a clear methodological foundation, several key concepts and algorithms utilized in this study are briefly defined as follows: GEOBIA is an image analysis framework that groups contiguous pixels into spectrally homogeneous spatial entities (objects) based on spectral, spatial, and contextual attributes. SNIC is a non-iterative superpixel segmentation algorithm that partitions an image into compact, spatially contiguous regions by initializing seeds on a regular grid and expanding them deterministically based on distance and spectral similarity. GLCM is a statistical texture analysis method that characterizes image spatial texture by calculating how often pixel pairs with specific values occur within a local moving window. Hyperparameters refer to external configuration variables set prior to training (e.g., number of trees in RF) that govern learning architecture, while a Confusion Matrix is a cross-tabulation table comparing ground-truth reference samples against model predictions to derive accuracy metrics including OA, Kappa coefficient, Precision, Recall, and F1-score.

The research workflow was designed to effectively integrate GEOBIA with machine learning algorithms for LULC classification and urban change analysis in the study area. In addition, the performance of four classification algorithms, namely RF, SVM, CART, and GTB, was evaluated and compared. The overall workflow is illustrated in Fig. 3.

Initially, Sentinel-2 satellite imagery was collected and preprocessed to ensure data quality for subsequent analysis. The preprocessing steps included cloud and cloud-shadow masking, normalization of surface reflectance values, clipping to the study area boundary, and generation of composite images for each time period. Selecting images acquired under similar seasonal and atmospheric conditions helped minimize spectral inconsistencies between years.

Subsequently, reference sample datasets were independently constructed for each study year (2015, 2020, and 2025) using high-resolution imagery from Google Earth and expert knowledge. Samples representing each LULC class were digitized and spatially distributed across the study area to ensure geographic representativeness. For each temporal dataset, the samples were randomly partitioned into two independent subsets: 70% for model training and 30% for accuracy validation. The four machine learning algorithms were then trained and validated separately for each respective year, ensuring a consistent and comparable evaluation design across all three time points.

GEOBIA was implemented through an image segmentation process. In this study, the SNIC algorithm was applied to segment the imagery into homogeneous objects in terms of spectral and spatial characteristics. The segmentation process was governed by three key parameters: superpixel seed size, compactness, and connectivity. Based on preliminary tuning and the spatial

characteristics of the densely built urban environment in Hanoi, which features narrow residential blocks and fragmented land cover patches, the superpixel size was set to 10 pixels (corresponding to a 100 m² spatial domain at 10 m resolution), compactness was set to 1.0 to balance spectral and spatial weighting, and connectivity was set to 8. To maintain consistency and ensure objective comparability across the multi-temporal datasets, this identical set of SNIC segmentation parameters was uniformly applied to all Sentinel-2 imagery from 2015, 2020, and 2025.

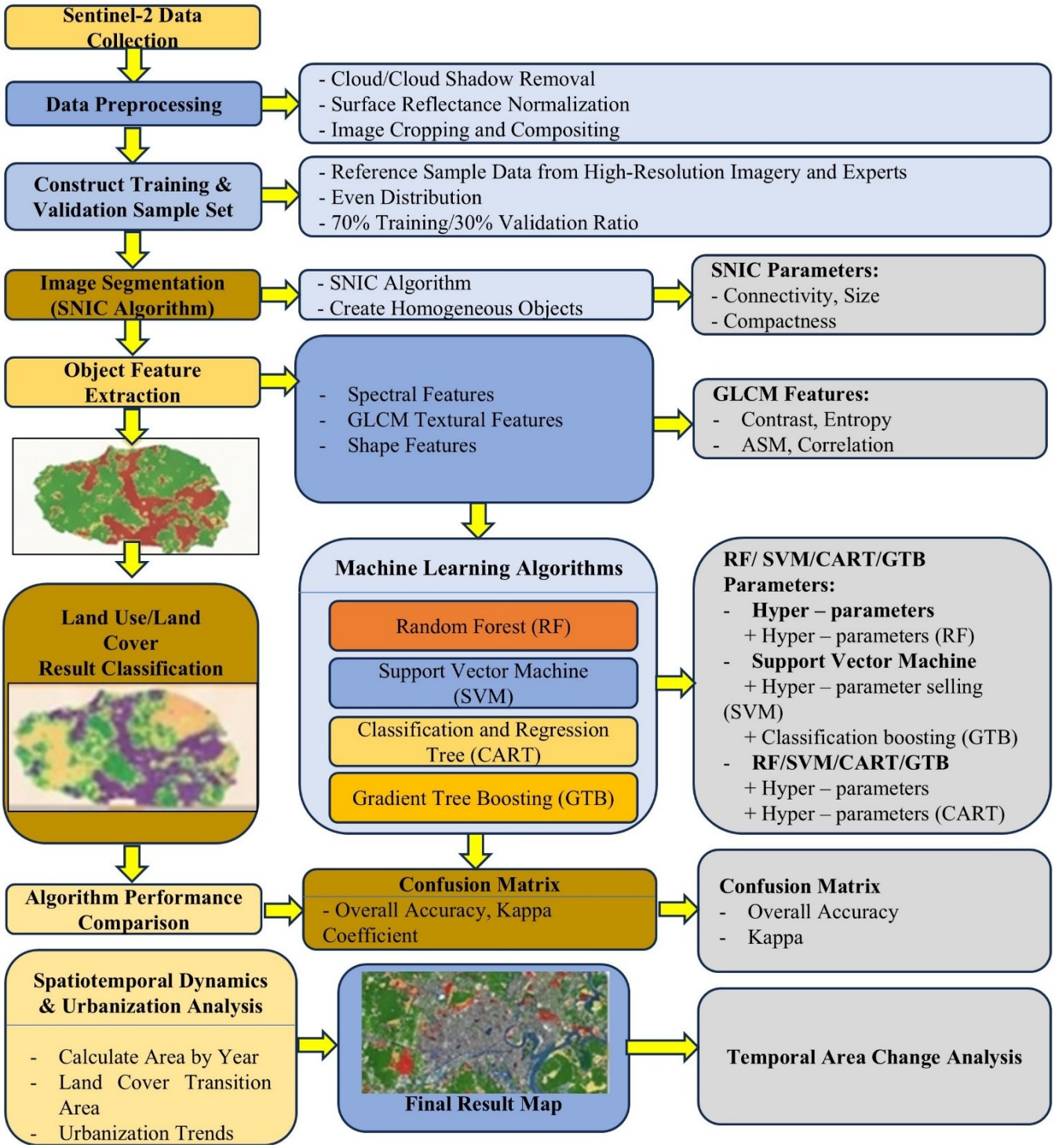


Fig. 3. Overall workflow of the study

Following segmentation, object-based features were extracted to support classification. These features include: (i) spectral features, calculated as the mean values of spectral bands

within each object; (ii) textural features derived from the GLCM, such as Contrast, Entropy, Angular Second Moment (ASM), and Correlation; and (iii) shape features, including area, perimeter, and compactness.

Based on the extracted features, four machine learning algorithms were employed for LULC classification and performance comparison. RF, an ensemble learning method, is capable of handling high-dimensional data and reducing overfitting [22]. SVM with a radial basis function (RBF) kernel was used to address nonlinear classification problems [4, 11]. CART is a simple and interpretable decision tree model, although it may suffer from overfitting if not properly controlled [10]. GTB, a boosting-based method, combines multiple weak learners to improve overall model accuracy [12]. To optimize classification performance and ensure model reproducibility, specific hyperparameters were configured for each machine learning algorithm based on preliminary experimentation, as summarized in Table 4.

Table 4. Summary of hyperparameters configured for machine learning classifiers

Classifier Algorithm	Hyperparameter Name	Configured Value
RF	Number of Trees	200
SVM	Kernel Type	RBF
	Gamma	1
	Cost	30
CART	Maximum Depth	3
GTB	Number of Trees	100
	Shrinkage	0.1
	Sampling Rate	0.7
	Max Nodes	50

Classification accuracy was evaluated using a confusion matrix, from which OA and the Kappa coefficient were derived. The use of an independent validation dataset ensured an objective assessment of model performance [23-25]. Finally, based on the classification results for each study year, the area of each LULC class was calculated and compared over time. These results were used to analyze temporal trends in land use change, thereby highlighting the process of urban expansion and spatial transformation within the study area.

3. Results and discussion

Figs. 4-7 present the LULC classification results of the study area using four machine learning algorithms, namely RF, SVM, CART, and GTB. The classification outputs indicate that all models successfully identified the five LULC classes, including water, shrubland, agricultural land, built-up land, and bare land. A visual comparison of the classification maps across the three time points (2015, 2020, and 2025) reveals that built-up land is the dominant class in terms of spatial distribution and exhibits a clear expansion over time. This trend reflects the rapid urbanization occurring in the study area. In contrast, water bodies, shrubland, agricultural land, and bare land occupy smaller and more fragmented areas, typically distributed in transitional zones or undeveloped spaces within the urban fabric. In terms of spatial patterns, built-up land is highly concentrated in the urban core areas, whereas shrubland, agricultural land, and bare land are more sparsely distributed, mainly occurring in peripheral zones or as fragmented patches within residential areas. Water bodies are consistently identified across all classification models, primarily associated with ponds, lakes, and small drainage networks within the study area. Overall, the classification maps reveal a clear trend of urban expansion, characterized by the continuous growth of built-up land alongside a gradual decline in

natural and semi-natural surfaces.

To rigorously evaluate and compare the performance of the four machine learning algorithms across different time periods, training and validation procedures were independently conducted for each study year (2015, 2020, and 2025). For each temporal dataset, 70% of the reference samples were used for model training, while the remaining 30% were reserved for independent validation using OA and Kappa coefficient metrics. Table 5 and Table 6 present the training and validation accuracies for all four models across the three target years.

The training accuracy results presented in Table 2 indicate that all four machine learning algorithms achieved high performance in classifying LULC data. OA values ranged from 83.64% to 90.81%, while Kappa coefficients varied between 0.817 and 0.860, demonstrating strong agreement between the classified results and the reference data. Among the evaluated models, RF achieved the highest training accuracy, with OA reaching 90.81% and Kappa reaching 0.860 in 2015. GTB also showed strong performance, followed by SVM and CART.

In addition, Table 3 presents the validation accuracy assessment results of the RF, SVM, CART, and GTB algorithms based on the OA and Kappa indices for the years 2015, 2020, and 2025, thereby enabling a comparison of the performance of the machine learning models applied in this study.

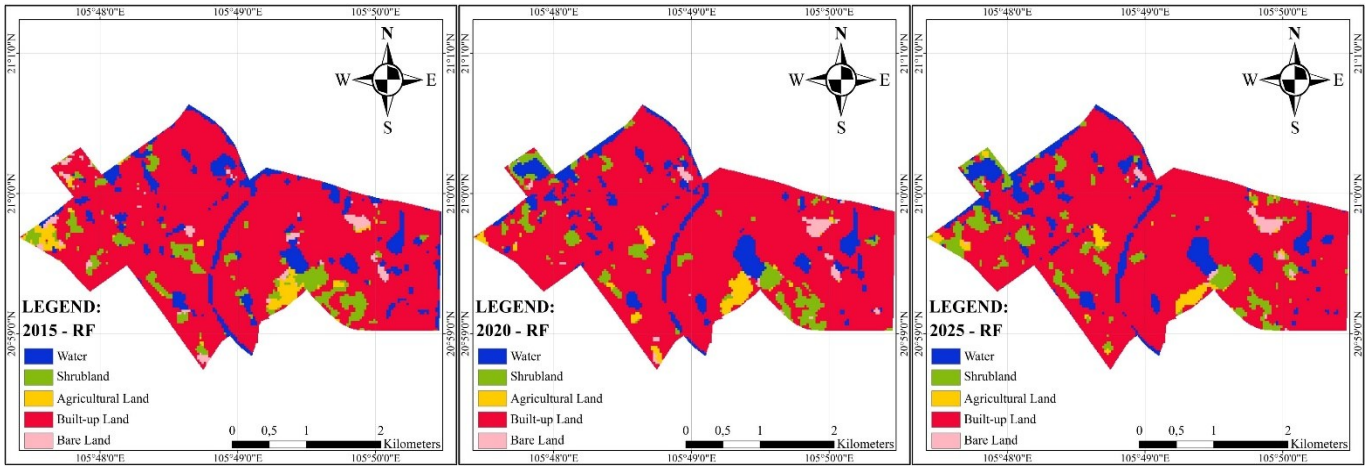


Fig. 4. LULC classification results of the study area using the RF algorithm

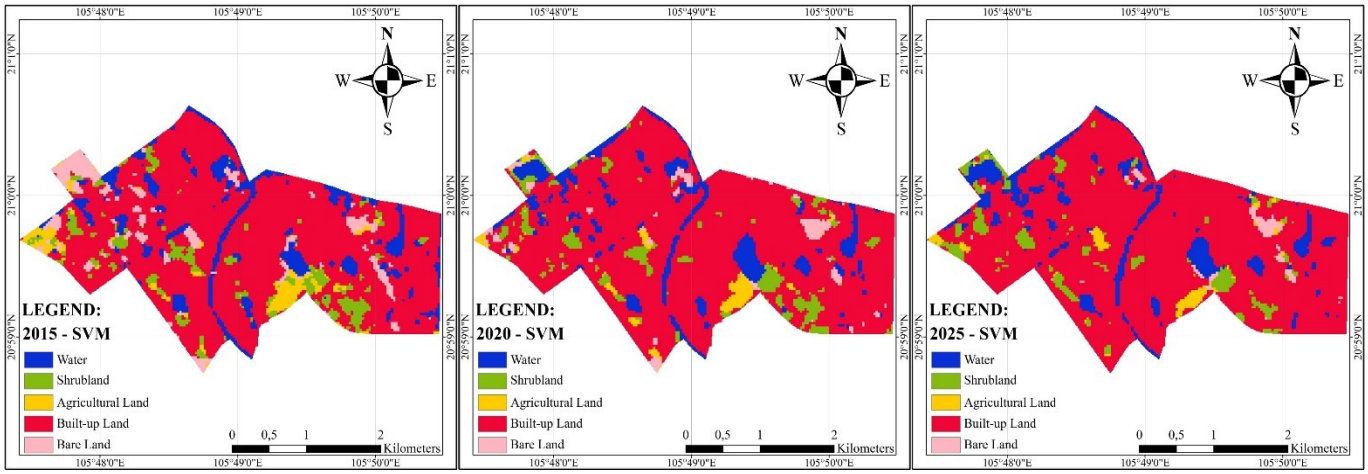


Fig. 5. LULC classification results of the study area using the SVM algorithm

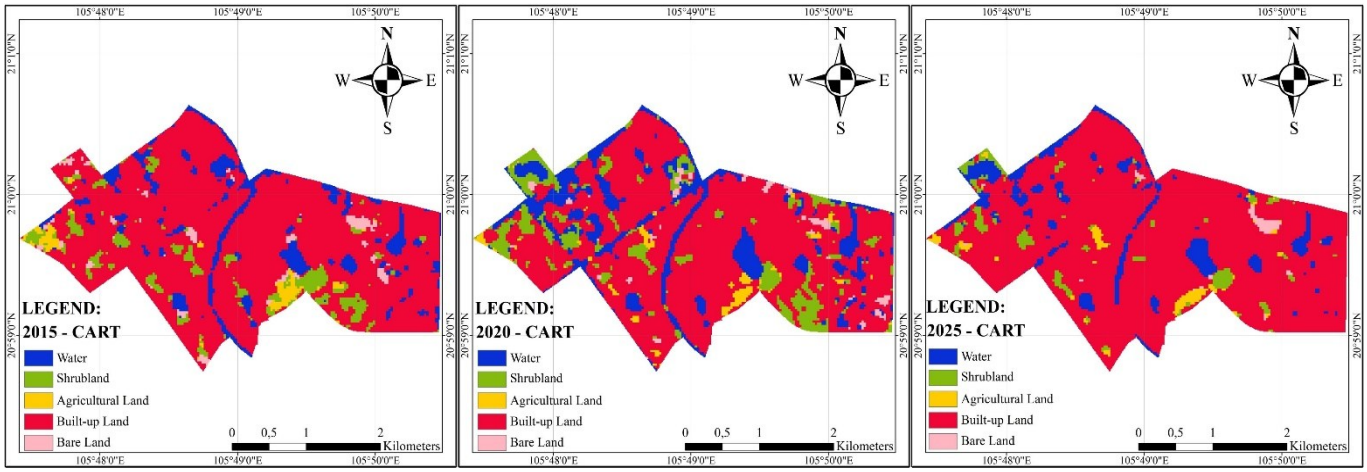


Fig. 6. LULC classification results of the study area using the CART algorithm

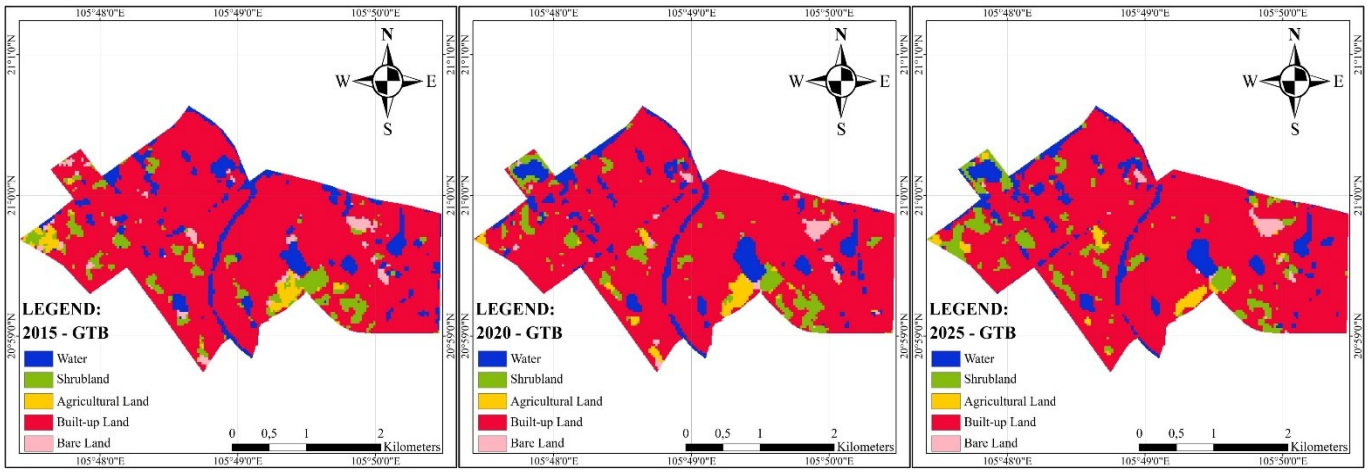


Fig. 7. LULC classification results of the study area using the GTB algorithm

Table 5. Training accuracy of four machine learning algorithms

Metric	RF			SVM			CART			GTB		
	2015	2020	2025	2015	2020	2025	2015	2020	2025	2015	2020	2025
OA (%)	90.81	87.81	85.84	86.03	82.85	85.8	84.45	86.71	83.64	88.38	87.71	84.34
Kappa Coefficient	0.86	0.856	0.84	0.842	0.822	0.838	0.829	0.845	0.817	0.856	0.855	0.826

Table 6. Validation accuracy of four machine learning algorithms

Algorithm	Metric	2015	2020	2025
RF	OA (%)	87.35	84.45	82.62
	Kappa Coefficient	0.831	0.793	0.818
SVM	OA (%)	82.75	79.76	82.6
	Kappa Coefficient	0.771	0.775	0.760
CART	OA (%)	81.25	83.45	80.51
	Kappa Coefficient	0.762	0.777	0.788
GTB	OA (%)	85.05	84.42	82.60
	Kappa Coefficient	0.799	0.834	0.818

The validation results indicate that all four algorithms: RF, SVM, CART, and GTB, achieved relatively high performance, with OA ranging from 79.76% to 87.35% and Kappa coefficients from

0.760 to 0.834. These values demonstrate good validation performance and confirm the capability of the models to effectively discriminate LULC classes within the study area. Overall, RF and GTB outperformed SVM and CART.

For RF, the highest validation accuracy was obtained in 2015 (OA = 87.35%, Kappa = 0.831). Although the OA slightly decreased in 2020 (84.45%) and 2025 (82.62%), the Kappa coefficient remained relatively high, even increasing to 0.818 in 2025. This indicates that RF maintains strong predictive consistency and robustness over time despite minor fluctuations in validation accuracy.

In contrast, SVM exhibited moderate validation performance, with OA ranging from 79.76% to 82.75% and Kappa values between 0.760 and 0.775. The lowest validation accuracy was observed in 2020 (OA = 79.76%), followed by a slight recovery in 2025. However, the overall performance of SVM remained less stable and inferior compared to RF.

CART showed comparable but slightly lower validation performance, with OA values ranging from 80.51% to 83.45% and Kappa from 0.762 to 0.788. Although CART achieved its highest OA in 2020, its performance fluctuated across years, indicating limited stability. Meanwhile, GTB demonstrated relatively strong validation performance, with OA between 82.60% and 85.05% and Kappa values from 0.799 to 0.834, making it competitive with RF, particularly in the earlier period.

A temporal comparison reveals that RF consistently achieved the highest or near-highest validation accuracy across all years. In 2015, RF clearly outperformed all other models. In 2020, RF and GTB showed nearly identical validation performance, while SVM lagged behind. By 2025, although all models experienced slight declines in validation accuracy, RF maintained the highest OA and a strong Kappa value.

Overall, RF and GTB demonstrated relatively high performance and consistency among the evaluated algorithms, with RF showing slightly superior stability across the three study years [22, 26]. These findings provide a reasonable empirical basis for selecting RF as the primary model for subsequent land use change analysis within the scope of this study.

Table 7. Class-specific validation metrics and confusion matrix for the RF model

LULC Class	Reference Total	Predicted Total	Precision (%)	Recall (%)	F1-Score (%)
Water	24	25	88	91.67	89.80
Shrubland	21	21	80.95	80.95	80.95
Agricultural land	21	22	82.61	90.48	86.36
Built-up land	33	35	91.43	87.88	89.62
Bare land	21	17	78.26	85.71	81.82
Overall Metrics	Total = 120	Total = 120	OA = 82.62%	Kappa = 0.818	-

To provide a more granular evaluation beyond OA and the Kappa coefficient, class-specific performance metrics, including Precision, Recall, and F1-score, were derived from the independent validation confusion matrix for the superior RF model. As detailed in Table 7, the RF classifier demonstrated robust performance across all five LULC classes. Built-up land achieved high accuracy (Precision of 91.43% and Recall of 87.88%), reflecting the effectiveness of object-based spectral and textural features in delineating dense urban structures from other cover types. Water bodies also exhibited high separability, yielding an F1-score of 89.80%. Minor spectral confusion was observed between Shrubland and Bare land due to similar spectral responses in transitional zones, yet both classes maintained acceptable F1-scores exceeding 80%. These

class-level metrics confirm the reliability and stability of the proposed GEOBIA-RF framework in mapping heterogeneous urban environments.

To quantitatively evaluate the individual contributions of the three feature groups, spectral, textural (GLCM), and shape features, an ablation analysis was conducted using the RF classifier on the 2025 dataset. As presented in Table 8, removing feature groups resulted in varying degrees of classification accuracy degradation, confirming that each group plays a vital role in capturing urban landscape complexity. The baseline model utilizing all feature combinations achieved the highest performance (OA = 82.62%, Kappa = 0.818). Omitting shape features led to a moderate decline in OA (ΔOA = -1.67%), demonstrating their utility in defining object boundaries and geometric properties of urban parcels. Removing GLCM textural features caused a more substantial drop (ΔOA = -4.22%), highlighting the importance of texture metrics in mitigating spectral confusion among heterogeneous materials (for example, distinguishing high-density residential roofs from bare soil or pavement). When relying solely on spectral features, accuracy dropped sharply to 74.85% (Kappa = 0.715). These ablation results quantitatively substantiate the necessity of integrating multi-feature sets within the GEOBIA framework to ensure reliable classification in complex urban environments.

To move beyond simple net area accounting and comprehensively understand the spatial conversion processes driving urban expansion, a land-cover transition matrix was generated for the 2015-2025 period (Table 9). The matrix reveals that the net expansion of built-up land by 9.84 ha (increasing from 91.04 ha in 2015 to 100.88 ha in 2025) was primarily fueled by the direct conversion of natural and semi-natural cover types within the urban transitional zones. Specifically, agricultural land experienced the most intensive transformation relative to its initial baseline, with approximately 2.30 ha converted into built-up infrastructure (representing over 65% of its 2015 extent). Furthermore, significant portions of water bodies (3.90 ha) and shrubland (3.85 ha) were encroached upon or filled for urban construction and residential development due to high land-use pressure in inner Hanoi. Conversely, transitions from built-up land back to other classes remained minimal, illustrating the predominant and irreversible nature of urban densification during this decade.

Based on the RF classification results (Table 10), built-up land is the dominant class and shows a continuous increase from 91.04 ha in 2015 to 98.10 ha in 2020 and 100.88 ha in 2025, representing a net increase of 9.84 ha over the study period. This class exhibits the most significant change, clearly reflecting the expansion of urban areas within the study region.

In contrast, all other LULC classes show a decreasing trend. Water bodies declined from 16.77 ha to 12.03 ha (a reduction of 4.74 ha), shrubland decreased from 11.13 ha to 8.20 ha, agricultural land dropped from 3.51 ha to 1.85 ha (approximately 47%), and bare land slightly decreased from 2.42 ha to 1.91 ha. These trends indicate that portions of natural, semi-natural, and unused land have been progressively converted into built-up areas.

The results indicate that the integration of GEOBIA with machine learning algorithms is well suited for urban environments with fragmented spatial structures, such as Thanh Xuan, Khuong Dinh, and Phuong Liet wards. The combination of spectral, textural, and shape features enhances the discrimination of classes with similar spectral responses, particularly built-up land, shrubland, and bare land. Given the 10 m spatial resolution of Sentinel-2 data, the use of segmented objects instead of individual pixels reduces salt-and-pepper noise and improves the spatial coherence of classification results. While this study did not conduct a direct benchmark comparison against a

traditional pixel-based classifier, theoretical considerations and previous literature suggest that the integration of GEOBIA helps reduce salt-and-pepper noise and improve spatial coherence in complex urban settings [3, 5]. However, further controlled comparative experiments are necessary to quantitatively verify the performance differences between object-based and pixel-based approaches using identical datasets.

Furthermore, the differences in classification performance among the algorithms reflect their underlying characteristics. RF and GTB achieved better results due to their ability to handle high-dimensional data, reduce noise sensitivity, and improve model stability through ensemble learning [9, 12]. In contrast, SVM and CART were more susceptible to spectral overlap and spatial heterogeneity, leading to lower and less consistent performance. Therefore, appropriate parameter tuning and model selection are essential to enhance the reliability of LULC classification in future studies [27, 28].

Table 8. Ablation analysis of feature group contributions on RF classification accuracy (2025 dataset)

Model Configuration	Feature Groups Included	OA(%)	Kappa Coefficient	Accuracy Drop (ΔOA %)
Full Model	Spectral + GLCM + Shape	82.62	0.818	-
Model A (No Shape)	Spectral + GLCM	80.95	0.792	-1.67
Model B (No GLCM)	Spectral + Shape	78.40	0.761	-4.22
Model C (Spectral Only)	Spectral bands & indices only	74.85	0.715	-7.77

Table 9. LULC transition matrix from 2015 to 2025 (unit: ha)

2015 \ 2025	Water	Shrubland	Agricultural land	Built-up land	Bare land	2015 Total
Water	11.25	1.12	0.20	3.90	0.30	16.77
Shrubland	0.35	6.50	0.15	3.85	0.28	11.13
Agricultural land	0.05	0.10	0.80	2.30	0.26	3.51
Built-up land	0.18	0.30	0.10	89.44	1.02	91.04
Share / Bare land	0.20	0.18	0.60	1.39	0.05	2.42
2025 Total	12.03	8.20	1.85	100.88	1.91	124.87

Table 10. Area of LULC classes based on RF classification (ha)

LULC Class	RF		
	2015	2020	2025
Water	16.77	13.88	12.03
Shrubland	11.13	7.88	8.20
Agricultural land	3.51	2.90	1.85
Build-up land	91.04	98.10	100.88
Bare land	2.42	2.11	1.91

The effectiveness of GEOBIA in this study was closely related to the spatial characteristics of the study area. Inner Hanoi is characterized by dense construction patterns, narrow residential blocks, mixed land use, and fragmented vegetation patches, which often generate high spectral confusion in medium-resolution satellite imagery [29]. In such conditions, relying solely on pixel spectral values may lead to misclassification between impervious surfaces, bare soil, and sparsely vegetated areas. By incorporating spatial context and object geometry, the object-based approach allows the classification process to better represent real-world urban structures. This is particularly important in rapidly urbanizing districts where land cover boundaries are irregular and

continuously changing over time.

The integration of GLCM-based textural features further improved class separability by capturing local spatial variability that cannot be fully represented through spectral information alone. Built-up areas in Hanoi typically exhibit heterogeneous textures due to variations in roofing materials, road networks, and building densities, while vegetation and water bodies generally present more homogeneous texture patterns [29, 30]. Consequently, texture metrics contribute significantly to reducing confusion among spectrally similar classes and enhancing classification robustness in complex urban environments.

Despite the satisfactory classification performance, several limitations should also be acknowledged. The 10 m spatial resolution of Sentinel-2 imagery may still be insufficient for detecting small urban features such as narrow roads, isolated trees, or small water bodies commonly found in densely populated wards. In addition, seasonal variations, shadow effects from high-rise buildings, and spectral similarity between construction materials and exposed soil may introduce classification uncertainty. Although the object-based framework reduces salt-and-pepper noise, segmentation quality remains highly dependent on parameter selection, which can influence object size and classification consistency across temporal datasets.

To highlight the novelty and methodological contributions of this research, its approach and findings are evaluated in comparison with previous remote sensing studies conducted in the Hanoi area. Earlier studies primarily utilized pixel-based algorithms or single-sensor datasets; for instance, Hang et al. (2017) mapped land cover in Hanoi using multi-temporal SAR imagery (Sentinel-1A), which was effective for surface roughness discrimination but remained prone to speckle noise and limited in optical feature representation [31]. More recently, Tung et al. (2023) monitored land cover changes across Hanoi during 2013-2023 using Google Earth Engine and pixel-based RF [32]. While their macro-level analysis captured broad urban growth trends, pixel-based classifications in highly heterogeneous inner-city wards often generate ‘salt-and-pepper’ artifacts and fail to account for spatial context and geometry.

In contrast, the key novelty of this study lies in establishing a fine-scale, object-based classification framework (GEOBIA) utilizing SNIC superpixel segmentation coupled with multi-feature extraction (integrating Sentinel-2 spectral bands, spectral indices, GLCM texture metrics, and object shape properties). By focusing on densely built inner-city wards (Thanh Xuan, Khuong Dinh, and Phuong Liet) over a 10-year span (2015-2025), this study effectively mitigates pixel-level spectral confusion among spectrally similar urban classes (e.g., bare soil, bright rooftops, and paved surfaces). Furthermore, unlike prior works, this research provides a comprehensive comparative evaluation across four machine learning classifiers (RF, GTB, SVM, CART) alongside quantitative ablation analysis, demonstrating that object-level texture and geometry significantly improve classification stability in complex, fragmented tropical urban landscapes.

From a broader perspective, the observed increase in built-up land and decline in vegetation-related classes reflect the continuing urban expansion process in inner Hanoi during the 2015-2025 period. These spatial transformations potentially contribute to environmental pressures, such as localized urban heating or reduced green space availability, highlighting the need for further detailed ecological investigations. Therefore, the integration of GEOBIA and machine learning provides a practical technical framework for LULC mapping, offering valuable spatial information to support sustainable urban planning and environmental monitoring in rapidly developing

cities.

4. Conclusions

This study successfully applied GEOBIA combined with machine learning algorithms to classify LULC and analyze spatial dynamics in Thanh Xuan, Khuong Dinh, and Phuong Liet wards, Hanoi, over the period 2015-2025. The results demonstrate that the proposed GEOBIA framework is effective in complex urban environments with fragmented spatial structures, as it integrates spectral, textural, and shape features to improve the discrimination of LULC classes with similar spectral characteristics. Among the tested algorithms, RF achieved relatively high and consistent performance across the study years, maintaining OA above 82% with strong Kappa values. GTB also showed competitive results, while SVM and CART exhibited lower and more variable accuracy. These findings highlight the utility of ensemble learning methods within the scope of this study. The LULC change analysis based on RF results indicates a clear trend of urban expansion in the study area. Built-up land increased continuously, whereas natural and semi-natural classes, including water bodies, shrubland, agricultural land, and bare land, declined. This pattern reflects ongoing urban transformation processes, underscoring the practical application of the approach for providing spatial information to support urban planning, while acknowledging that further multi-source environmental monitoring is required to fully assess specific ecological impacts. From a practical perspective, the quantitative findings of this study provide actionable baseline data and spatial insights for urban planners and local authorities in Hanoi. The continuous encroachment of built-up areas on surface water bodies, agricultural land, and urban green spaces underscores the urgent need for stringent urban growth boundary management and sustainable land-use zonation. Specifically, spatial monitoring results can support targeted policy interventions to preserve remaining blue-green infrastructure, optimize land-use efficiency in high-density inner-city wards, and mitigate urban flooding risks. Furthermore, the semi-automated GEOBIA-machine learning framework established in this study offers a transferable, cost-effective tool for continuous, multi-temporal monitoring of rapid urban dynamics in similar expanding metropolitan areas across Vietnam. Despite these promising results, several limitations remain, including the moderate spatial resolution of Sentinel-2 imagery (10 m), potential uncertainties in training sample selection, and the sensitivity of segmentation parameters. Future research should include direct benchmark comparisons between object-based and pixel-based classifiers to further substantiate these methodological advantages.

Declaration of Generative AI

The authors declare that no generative AI or AI-assisted technologies were used in the preparation, writing, or analysis of this manuscript.

References

[1]. M.C. Hansen, T.R. Loveland. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122, 66-74. https://doi.org/10.1016/j.rse.2011.08.024

[2]. D. Phiri, M. Simwanda, S. Salekin, V.R. Nyirenda, Y. Murayama, M. Ranagalage. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, 12(14), 2291. https://doi.org/10.3390/rs12142291

[3]. T. Blaschke. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2-16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

[4]. G. Mountrakis, J. Im, C. Ogole. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

[5]. M. Kucharczyk, G.J. Hay, S. Ghaffarian, C.H.

Hugenholtz. (2020). Geographic object-based image analysis: A primer and future directions. *Remote Sensing*, 12(12), 2012. https://doi.org/10.3390/rs12122012

[6]. M. Drusch, U.D. Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25-36. https://doi.org/10.1016/j.rse.2011.11.026

[7]. R. Achanta, S. Susstrunk. (2017). Superpixels and polygons using simple non-iterative clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4651-4660. IEEE.*

[8]. R.M. Haralick, K. Shanmugam, I. Dinstein. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), 610-621. https://doi.org/10.1109/TSMC.1973.4309314

[9]. L. Breiman. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[10]. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone. (1984). Classification and regression trees. *Taylor & Francis*. https://doi.org/10.1201/9781315139470

[11]. C. Cortes, V. Vapnik. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/BF00994018

[12]. J.H. Friedman. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

[13]. H.S. Nghiyalwa, M. Urban, J. Baade, I.P.J. Smit, A. Ramoelo, B. Mogonong, C. Schmullius. (2021). Spatio-temporal mixed pixel analysis of savanna ecosystems: A review. *Remote Sensing*, 13(19), 3870. https://doi.org/10.3390/rs13193870

[14]. Z. Shao, T. Cheng, H. Fu, D. Li, X. Huang. (2023). Emerging issues in mapping urban impervious surfaces using high-resolution remote sensing images. *Remote Sensing*, 15(10), 2562. https://doi.org/10.3390/rs15102562

[15]. M.A.M. Choudhury, G. Modica, S. Praticò, E. Marcheggiani. (2026). A scalable GEOBIA framework for urban landscape monitoring with Sentinel-2 data: A case study in Hue City, Vietnam. *Earth*, 7(2), 51. https://doi.org/10.3390/earth7020051

[16]. L. Liu, C. Zhang, W. Luo, S. Chen, F. Yang, J. Liu. (2022). New remote sensing image fusion for exploring spatiotemporal evolution of urban land use and land cover. *Journal of Applied Remote Sensing*, 16(3), 034527. https://doi.org/10.1117/1.JRS.16.034527

[17]. J.W. Rouse Jr., R.H. Haas, J.A. Schell, D.W. Deering. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium (Vol. 1, pp. 309-317). NASA Goddard Space Flight Center.* https://ntrs.nasa.gov/citations/19740022614

[18]. S.K. McFeeters. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425-1432. https://doi.org/10.1080/01431169608948714

[19]. Y. Zha, J. Gao, S. Ni. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583-594. https://doi.org/10.1080/01431160304987

[20]. J.R. Landis, G.G. Koch. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. https://doi.org/10.2307/2529310

[21]. J. Cohen. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. https://doi.org/10.1177/001316446002000104

[22]. M. Belgiu, L. Drăguţ. (2016). Random forest

in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[23]. R.G. Congalton. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46. https://doi.org/10.1016/0034-4257(91)90048-B

[24]. G.M. Foody. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185-201. https://doi.org/10.1016/S0034-4257(01)00295-4

[25]. P. Olofsson, G.M. Foody, M. Herold, S.V. Stehman, C.E. Woodcock, M.A. Wulder. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42-57. https://doi.org/10.1016/j.rse.2014.02.015

[26]. Y.-C. Tan, L. Duarte, A.C. Teodoro. (2024). Comparative study of random forest and support vector machine for land cover classification and post-wildfire change detection. *Land*, 13(11), 1878. https://doi.org/10.3390/land13111878

[27]. B. Hosseiny, A.M. Abdi, S. Jamali. (2022). Urban land use and land cover classification with interpretable machine learning—A case study using Sentinel-2 and auxiliary data. *Remote Sensing Applications: Society and Environment*, 28, 100843. https://doi.org/10.1016/j.rsase.2022.100843

[28]. T. Zhang, J. Su, Z. Xu, Y. Luo, J. Li. (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

[29]. D.T. Loan, G. Jian, N.T. Ha, Z. Guohua, N.T.T. Ha, H. Nguyen. (2026). Urban heat as a growing hazard: A spatiotemporal analysis of land-use/land-cover changes and land surface temperature in Ha Noi city, Viet Nam. *Natural Hazards*, 122(6), 228. https://doi.org/10.1007/s11069-025-07763-6

[30]. T.A. Nguyen, P.M.T. Le, T.M. Pham, H.T.T. Hoang, M.Q. Nguyen, H.Q. Ta, H.T.M. Phung, H.T.T. Le, L. Hens. (2019). Toward a sustainable city of tomorrow: A hybrid Markov-Cellular Automata modeling for urban landscape evolution in the Hanoi city (Vietnam) during 1990-2030. *Environment, Development and Sustainability*, 21(1), 429-446. https://doi.org/10.1007/s10668-017-0046-2

[31]. L.M. Hang, V.V. Truong, N.D. Duong, T.A. Tuan. (2017). Mapping land cover using multi-temporal Sentinel-1A data: A case study in Hanoi. *Vietnam Journal of Earth Sciences*, 39(4), 345-359. https://doi.org/10.15625/0866-7187/39/4/10730

[32]. D.T. Tung, D.T.T. Huyen, H.T. Thuy, T.M. Ngoc. (2023). Application of satellite images and artificial intelligence to monitor land cover changes in Hanoi area during 2013-2023 period. *VNU Journal of Science: Earth and Environmental Sciences*, 39(4), 102-112. (in Vietnamese). https://doi.org/10.25073/2588-1094/vnuees.4962