ISSN 2029-6991 / eISSN 2029-7009

VILNIUS TECH
Vilnius Gediminas Technical University

GEODESY and CARTOGRAPHY

2026
Volume 52
Issue 1
Pages 42–53
https://doi.org/10.3846/gac.2026.22678

UDC 528.94

PREDICTING LAND USE/LAND COVER CHANGES USING CA-ANN MODEL IN THE GIA LAM DISTRICT, HANOI CITY

Thanh Tung DANG, Anh Tuan PHAM

Land Administation Faculty, Hanoi University of Natural Resources and Environmet, Hanoi, Vietnam

Article History:
- received 14 November 2024
- accepted 04 March 2026

Abstract. Gia Lam is a district in the eastern part of Hanoi, Vietnam, undergoing rapid development and urbanization. Predicting land use/land cover changes is crucial for supporting policymakers in designing effective local development strategies. This study aims to forecast land use/land cover changes in Gia Lam District through 2028. Sentinel-2 optical satellite imagery serves as the primary data source, with the classification of five land use/land cover types at different times performed using the Random Forest (RF) algorithm, and forecast results generated using a Cellular Automata (CA) model combined with an Artificial Neural Network (ANN) model. The results indicate a significant expansion of built-up areas, projected to cover 68.92% of the natural area by 2028. On the contrary, the area under annual vegetation is expected to decline to just 7.71%. These findings provide valuable insights for researchers and land use planners.

Keywords: Gia Lam, Random Forest, Cellular Automata, Artificial Neural Network, land use/land cover, classification, Sentinel-2 satellite imagery.

Corresponding author. E-mail: ***dttung.qldd@hunre.edu.vn***

1. Introduction

Gia Lam District is located at the eastern gateway of Hanoi, the capital of Vietnam, positioned on the major transportation axis of the Northern Key Economic Region, connecting Hanoi with the provinces of Hai Phong, Quang Ninh, Hai Duong, Hung Yen, and Bac Ninh through an essential transportation network. Gia Lam is an area experiencing rapid socio-economic development, accompanied by changes in land use within the study area

Studies on Land Use/Land Cover (LULC) changes often utilize algorithms such as MLC, KNN, and typically rely on analytical tools provided by commercial software, which require certain fees and can be challenging to access for research projects with limited funding (Musleh & Jaber, 2021; Swetanisha et al., 2022). Currently, numerous studies use satellite imagery and machine learning algorithms to monitor LULC changes (Phuong et al., 2024; Hoan & Hoan, 2024). These studies have been widely applied globally, including in Europe, the Americas, Africa, Oceania, and Asia (Assede et al., 2023; Li et al., 2019; Marshall et al., 2018). The data used in these studies include freely available sources, such as Landsat and Sentinel satellite images, which have yielded useful results for monitoring LULC changes over specific time series in various research areas (Karra et al., 2021; Owojori & Xie, 2005). Besides free satellite imagery, these studies also utilize machine learning and deep learning algorithms that are increasingly accessible through free platforms like GEE (Kouassi et al., 2024).

The Cellular Automata (CA) model was first developed by Stanislaw Ulam in the 1940s while researching crystals, and John Von Neumann in the field of self-replicating systems (Von Neumann, 2017). By the 1950s, the CA model began being used in ecological system simulations Based on LULC fluctuation research results, researchers have also made future LULC fluctuation forecasts for periods from 5 to 10 years, 20 years, and even longer periods (Karimi et al., 2018; Liping et al., 2018). Recently, Artificial Neural Network (ANN) has been used by researchers in artificial intelligence fields such as image recognition. Modeled after biological characteristics, ANN is deployed to perform tasks like clustering, classification, and pattern recognition. ANN has demonstrated effectiveness in forecasting outputs when compared to statistical models in regression-based problems and time series forecasting (Astuty & Dimyati, 2024; Lin et al., 2011). These studies use models such as CA and ANN to calculate forecasts and also demonstrate the reliability of the forecast results (Baig et al., 2022; Islam et al., 2018). However, forecast results depend on various influencing factors such as terrain, population density, transportation, economic

development. Therefore, each area will have different LULC fluctuation rates and extents, especially in rapidly developing areas like Gia Lam District, Hanoi, Vietnam. The Markov chain model for forecasting LULC changes is considered a traditional method and has been applied in many published studies (Liang et al., 2021). However, the Markov chain model only provides temporary dynamics and does not consider control data in spatial change prediction (Halmy et al., 2015). To overcome this limitation, the CA model represents spatial and dynamic processes, where future changes depend on the spatial state of neighboring pixels (Rahman et al., 2017). The CA model based on ANN in Deep Learning is widely applied and has achieved good results in forecasting future LULC (Kouassi et al., 2024; Sajan et al., 2022).

Recent advances have further demonstrated the robustness of CA-ANN hybrids in rapidly urbanising contexts, particularly in Asian and other developing megacities. For instance, (Jain, 2024) applied CA-ANN with MOLUSCE in Delhi megacity (India), projecting significant Built-up Area expansion at the expense of cultivable land and vegetation, with strong validation (OA = 79.78%, Kappa = 0.7725). Similarly, in Lucknow City (India), CA-ANN models integrated with Landsat data forecasted Built-up Area reaching 29.99% by 2031, accompanied by significant Vegetation decline and intensified urban heat island effects (Khan & Khan, 2025). In Vietnam, coupled CA-ANN applications for Hanoi Capital have predicted LULC changes up to 2043, confirming persistent vegetation-to-built-up conversions driven by infrastructure and population growth (Thien et al., 2025). These studies highlight the effectiveness of CA-ANN in data-limited settings using open-source platforms like GEE and QGIS-MOLUSCE. This study applies the CA-ANN model to forecast LULC changes in Gia Lam District, Hanoi City, up to 2028. As the first application of this integrated approach in Gia Lam—one of Hanoi's most rapidly urbanising peri-urban districts—this research addresses a key gap in localised predictive modelling. Located at the eastern gateway of the capital and along a major transportation corridor in the Northern Key Economic Region, Gia Lam faces intense socio-economic pressure and accelerated land conversion, making spatially explicit forecasts highly necessary.

The results provide essential, evidence-based insights into future Built-up Area expansion, Vegetation loss, and Water Bodies trends, supporting sustainable land-use planning, urban development strategies, and local socio-economic decisions. Utilising open-source tools (Google Earth Engine, QGIS-MOLUSCE) and free data (Sentinel-2, DEM, WorldPop), the study offers a cost-effective, reproducible framework adaptable to similar peri-urban areas in Vietnam and beyond, aiding policy efforts to balance rapid urbanisation with environmental sustainability in Hanoi's metropolitan fringe.

2. Study area and data

2.1. Study area

The study area has a climate divided into two distinct seasons: the hot and humid season lasting from April to

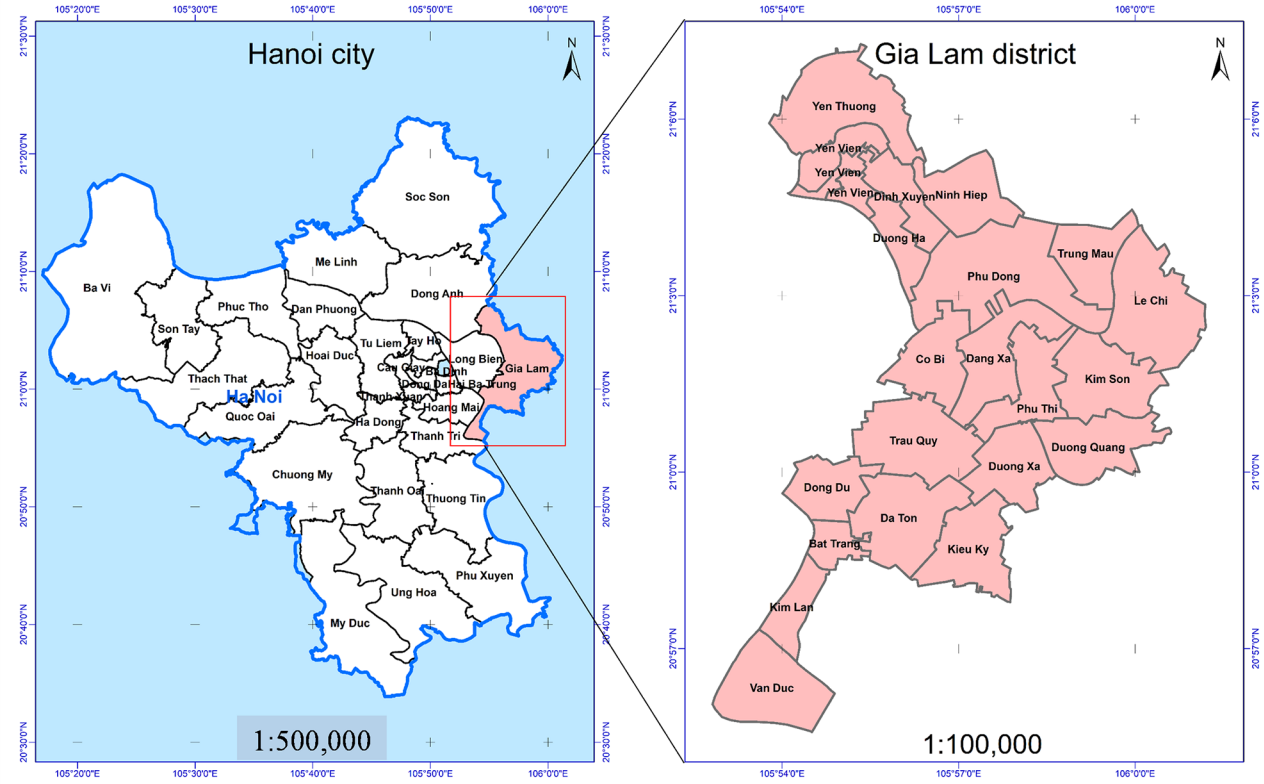

Figure 1. Location of study area

Table 1. Sentinel-2 parameters

Band number	Spatial sample distance (m)	Spatial sample distance (m) Central wavelength (nm)	Bandwidth (nm)	Radiance sensibility range Lmin < Lref < Lmax ($W.m^{-2}.sr^{-1}.\mu m^{-1}$)	SNR specification (at Lref)
2	10	490	65	11.5 < 128 < 615.5	154
3	10	560	35	6.5 < 128 < 559	168
4	10	665	30	3.5 < 108 < 484	142
8	10	842	115	1 < 103 < 308	174

October, and the dry season extending from November to March of the following year. Between these two seasons, there are transitional periods creating a four-season climate: spring, summer, autumn, and winter. Gia Lam District is part of the Red River Delta region, with low terrain and an average elevation ranging from approximately 4.5 m to 7.2 m. Gia Lam District covers an area of 114 km^2; the population is currently about 290,000 people, comprising 20 communes and 2 towns. Currently, Gia Lam is experiencing rapid urbanization, developing sustainably in a civilized and modern direction. Up to now, the district has no new poor households, 114 near-poor households have been reduced, the near-poor household rate is 0.58%. The study area has continued economic development, the economic structure has shifted in the right direction. The production value of the main economic sectors is estimated to increase by 11.80%, equal to 122.66% of the growth rate of the same period in 2023, continuing to build and develop district-level urban areas according to plan (Anh, 2022). Figure 1 illustrates the location of Gia Lam District and its spatial relationship to the principal neighbouring areas within Hanoi City.

2.2. Data using

The data used in this study includes Sentinel-2 satellite images, including Sentinel-2A and Sentinel-2B images, with each individual Sentinel satellite acquisition of a specific location every 10 days. The combination of Sentinel-2A and Sentinel-2B creates a Sentinel-2 satellite imagery set with a temporal resolution of 5 days. Sentinel-2 satellite images are stored at a 12-bit level, thus providing high radiometric resolution. This allows the ability to distinguish brightness levels in the image, with a potential range from 0 to 4095. The Sentinel-2 satellite imagery for the study area was selected to ensure clear quality, with cloud cover approximately 0.05%. With parameters such as high resolution, a relatively short image acquisition cycle (5 days), and a wide swath width of up to 290 km, which is significantly larger compared to Landsat swath widths of 185 km, and SPOT-5 of 120 km, these conditions indicate that free Sentinel-2 satellite imagery is a very suitable data source for this study. The basic parameters of Sentinel-2 satellite imagery include spectral bands with spatial resolutions ranging from 10 m to 60 m, an altitude of 786 km, and a local time at descending node of 10:30 hours (Suhet, 2015). Bands 2, 3, 4, and 8 are used for LULC classification of the study area. The detailed spectral and radiometric characteristics of the selected Sentinel-2 bands are summarised in Table 1.

Sentinel 2 level 2A images for the period 2018–2023 in Gia Lam district, Hanoi city are detailed in Table 2. These images were compiled in June each year, with very low cloud cover. In June 2018, the weather was cloudy and rainy, so to ensure the best possible conditions, the study used images compiled in July, with a cloud cover of 1.85%.

Table 2. Satellite images in the study area

Image acquisition time	Image type	Cloud cover (%)
07-2018	Sentinel-2A	1.85%
06-2023	Sentinel-2A	1.23%

In addition, Digital Elevation Model (DEM) data and population data were also used for the study. These data combined with the LULC classification results in 2018 and 2023 were used as input data to predict LULC changes in 2028.

3. Research methodology

3.1. Clasification method

3.1.1. Classification process

Based on satellite image data freely provided through the GEE platform, this study uses the Random Forest (RF) algorithm to classify LULC from Sentinel-2A satellite images. The Sentinel-2A data used in this study includes atmospherically corrected images with low cloud cover (Table 2) and is geometrically corrected to the WGS84 coordinate system. The image classification is conducted online on the GEE cloud computing platform. This study classifies images into five land use/land cover types: 1) Bare Land, 2) Water Bodies, 3) Perennial Vegetation, 4) Annual Vegetation, and 5) Built-up Area.

Sampling for training and classification is conducted for each LULC type. For Bare Land cover, sampling points include areas such as bare land, sand, alluvial areas, construction preparation sites, and newly leveled land. The Water Bodies cover includes ponds, lakes, rivers, streams, and canals. Perennial Vegetation includes areas with large, dense, perennial trees, while Annual Vegetation cover areas with crops, lawns, and fields containing seasonal vegetation. Built-up Area include houses, standalone buildings, and residential areas such as villages, apartment buildings, and high-rise structures. A total of 419 samples, evenly

distributed across the five LULC types, were collected to record information specific to each type and to train the machine learning model. Sample locations are evenly distributed based on the characteristics of each LULC type. Upon completing the training, the RF algorithm is used to classify the LULC types in the study area. The classification process follows the steps illustrated in Figure 2, including: Collecting Sentinel-2 data through the GEE platform; Filtering for images with minimal cloud cover; Sampling from specific locations for classification; Creating samples and training the machine learning model; Classifying land cover types using the RF algorithm; Collecting classification results; and Evaluating the accuracy of these results. Once accurate classification results are obtained, the study proceeds with analyzing and monitoring LULC changes in the study area for the 2018–2023 period.

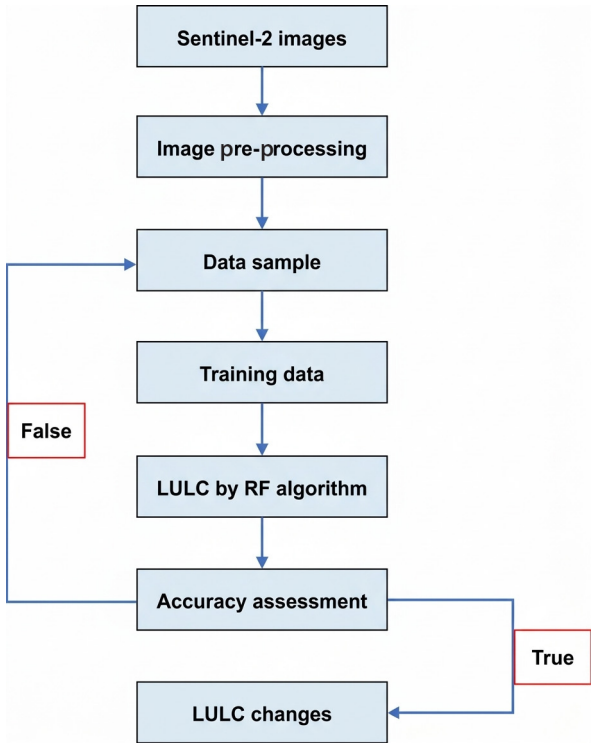

Figure 2. LULC classification flow chart

3.1.2. Classification algorithm

RF was introduced in 2001 by Leo Breiman, which is an ensemble machine learning algorithm that can integrate multiple decision trees and then form a forest (Breiman, 2001). The bagging method is used to generate training samples, and each selected feature is randomly drawn by replacing N (the size of the initial training set). Then, the final prediction result is obtained by combining multiple decision trees (Pal, 2005). The equation below makes the final classification decision as follows:

$$H(x) = \arg\max_Y \sum_{i=1}^{k} I\left(h_i(x) = Y\right), \quad (1)$$

where: $H(x)$ is the ensemble model, h_i is the classification model of a single decision tree, k is the elasticity coefficient; Y is the output variable (or target variable) and $I(\cdot)$ is the indicator function. The equation shows that the RF algorithm uses the majority of voting decisions to determine the final classification. The tuning parameter of the RF algorithm is the number of trees, and the number of trees can be selected according to the experience of the computational processor. Figure 3 presents a schematic representation of the RF algorithm, illustrating how multiple decision trees are aggregated through majority voting to generate the final classification output.

In this study, the RF classification algorithm was configured on GEE to predict 5 LULC classes (Bare Land, Water Bodies, Perennial Vegetation, Annual Vegetation, and Built-up Area) from Sentinel-2 satellite imagery, using the following hyperparameters: Number of trees (k in Equation (1)): 100, providing stable performance while ensuring computational efficiency; Maximum number of features considered at each split, which increases randomness and reduces correlation among individual trees, proving effective for multi-class classification problems with potential spectral overlap (e.g., between Annual Vegetation and Perennial Vegetation).

In Equation (1), the final prediction $H(x)$ is determined by majority voting across k = 100 individual decision trees $h\_i(x)$ (for i = 1 to 100). The indicator function $I(\cdot)$ counts the votes for each of the 5 classes, and the class Y receiving the highest number of votes is assigned. No additional weighting was applied beyond the standard majority voting mechanism. This configuration yielded high reliability, with overall accuracy (OA) and Kappa values presented more detail in the accuracy assessment section.

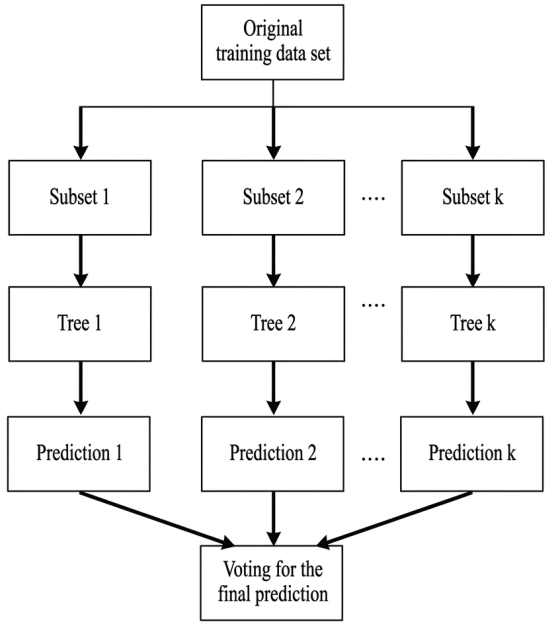

Figure 3. Illustration of RF algorithm

RF algorithm is used quite commonly in data classification. The RF algorithm is highly appreciated for its model accuracy (Liu et al., 2012). The main disadvantage of the RF algorithm is the large amount of computation, but the data processing time is considered fast (Liu & Wu, 2017).

3.1.3. Accuracy assessment method

The confusion matrix is used to evaluate the classification accuracy of LULC on the image. The confusion matrix is an important and popular method used to evaluate accuracy, which can describe the accuracy of classification and indicate the confusion between object classes. The basic statistics for the confusion matrix include the assessment of overall accuracy (OA) and the Kappa coefficient (Cohen, 1960). In which the Kappa coefficient has a value from 0.4 to 0.6 is considered to achieve average results, a value from greater than 0.6 to 0.8 is good and more than 0.8 to 1.0 is very good (Landis & Koch, 1977; Byrt et al., 1993). The evaluation of the accuracy of the classified images was carried out on GEE. In which, 70% of the samples were used for image classification and 30% of the samples were used for testing and evaluation. Specifically, 294 samples were used for classifying the cover classes and 125 samples were used for evaluating the accuracy of the image classification. Equations for calculating *OA* (2), and *Kappa* (3):

$$OA = \frac{TP + TN}{TP + FP + TN + FN}, \quad (2)$$

where: *TP* is True Positive, *FP* is False Positive, *TN* is True Negative, *FN* is False Negative.

$$Kappa = \frac{P_o - P_e}{1 - P_e} = 1 - \frac{1 - P_o}{1 - P_e}, \quad (3)$$

where: P_o is the observed agreement of the raters, P_e is the expected agreement of the raters.

3.2. Predictive method

The prediction of LULC change of the study area to 2028 (5-year cycle) is based on the CA model combined with the ANN model. In which, the CA model is a computational model based on a grid of cells, each cell can change state over time based on certain rules and the state of neighboring cells. The CA model was first developed by mathematicians John von Neumann and Stanislaw Ulam in the 1940s, completed by Arthur Walter Burks in 1966 (Von Neumann & Burks, 1966), used in the study of the evolution of urban land use patterns (White & Engelen, 1993). The CA model is widely used in simulating urban development based on input data such as population density, land use status, traffic infrastructure, and terrian. ANN model is a machine learning (ML) model consisting of interconnected "neurons" that process information from inputs to outputs through hidden layers (Rumelhart, Hinton, & Williams, 1986). ANN models help classify images into different classes, for example in object recognition, remote sensing analysis, and are also used in predicting land use, finance, energy, etc. The combination of CA-ANN brings the predictive ability of ANN and the locality of CA to simulate complex phenomena, especially those with spatial and temporal properties. For research on predicting urban development and land use needs, the role of CA-ANN is flexible, accurate, and effective (Yang et al., 2008). With this application, ANN can predict the development probability of each area in the city, while CA performs detailed simulation of surrounding areas. The above studies provide scientific foundation and evidence for the effectiveness and application of each model, as well as the benefits of

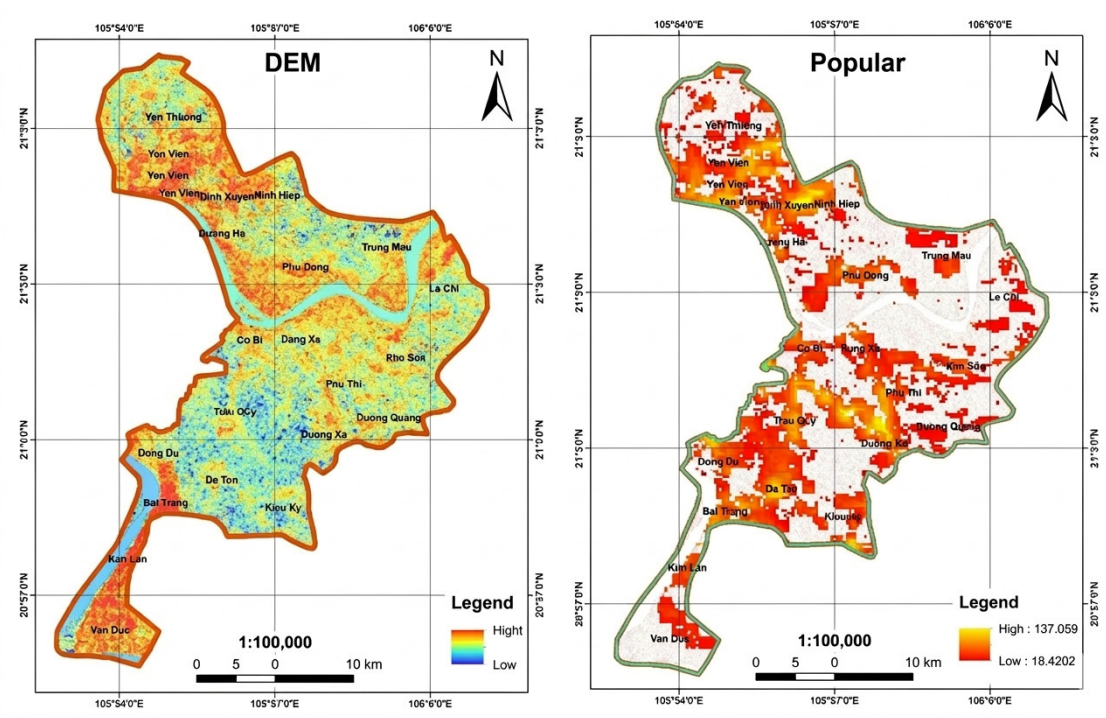

Figure 4. Digital Elevation Model and population data

combining models to improve accuracy and application in the fields of spatial simulation and classification. DEM data, population in 2023 are also included in the model along with land use data in 2018 and 2023 to predict LULC changes for the forecast year 2028. The DEM data was downloaded from NASA's Earthdata (https://earthdata.nasa.gov/) in raster format with a resolution of 30 m, then resampled to a resolution of 10 m to match the LULC data for 2018 and 2023 used in the study. Similarly, population data was downloaded from WorldPop (https://www.worldpop.org/) and used as input data for predicting LULC changes in 2028. Figure 4 illustrates the resampled Digital Elevation Model (DEM) and the 2023 population density data, which serve as key driving factors in the CA-ANN simulation.

The CA-ANN simulation was implemented using the MOLUSCE plugin in QGIS, which integrates ANN for modelling transition potential and CA for spatial allocation. The ANN component employed a multilayer perceptron with 10 hidden neurons, sigmoid activation function, and up to 1000 training iterations with early stopping applied when the error stabilised. Transition potential maps were generated based on observed LULC changes from 2018 to 2023, incorporating the selected drivers: resampled DEM (10 m resolution) and 2023 population density. A 5×5 Moore neighbourhood filter was used to capture local spatial dependencies and autocorrelation in land change processes.

Model validation followed the standard MOLUSCE protocol: the 2023 LULC map was simulated from 2018 inputs, and agreement was assessed through visual comparison of simulated versus observed maps and the overall Kappa statistic of 0.85. Pixel-level agreement confirmed reasonable performance for short-term (5-year) projections to 2028. These hyperparameters and setup align with common practices in MOLUSCE-based CA-ANN studies for urban/peri-urban LULC forecasting, balancing computational efficiency and predictive accuracy in data-limited settings.

4. Results and disscusion

4.1. Land use/land cover changes

The classified images clearly show that the majority of the LULC is Built-up Area, covering approximately 51.17% to 62.00% of the total natural area of the study area. The land cover with the least proportion is Bare Land, covering about 1.66% to 4.74%. The distribution results of these LULCs align with the natural characteristics, Built-up Area distribution, and socio-economic development situation in Gia Lam district, Hanoi city. The results also clearly indicate the rapid expansion of Built-up Area in Gia Lam District, Hanoi city, during the period from 2018 to 2023. Specifically, the classification results of Sentinel-2 satellite images show that the expansion rate of Built-up Area class reached 1.80% annually, increasing from 6058 ha (2018) to 7340 ha (2023). Meanwhile, the Bare Land and Annual Vegetation covers showed an average annual decrease of 0.51% and 0.76%, respectively, compared to the total natural area of Gia Lam district, Hanoi city.

The classification accuracy results were calculated from the confusion matrix, including overall accuracy (OA) and the Kappa coefficient, as shown in Table 3. The overall accuracy (OA) was 0.94, while the Kappa coefficient ranged from 0.88 to 0.89. The study results indicate that the post-classification image accuracy is very high, ensuring the reliability of the statistics and analysis of LULC in Gia Lam district, Hanoi city, during the period from 2018 to 2023. Figure 5 illustrates the classified LULC maps for 2018 and 2023, emphasising the spatial

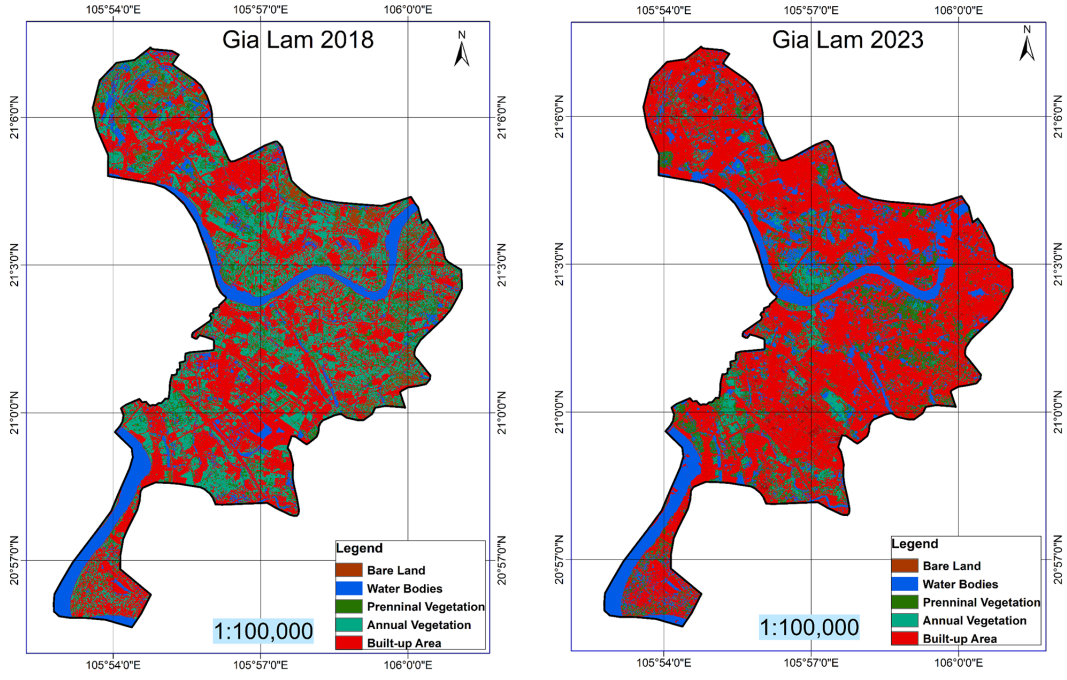

Figure 5. LULC map 2018 and 2023

distribution and the expansion of built-up areas over the five-year period.

Table 3. Image classification accuracy

Image acquisition time	OA	Kappa
2018	0.94	0.88
2023	0.94	0.89

Apart from the quality of the training samples, the classification accuracy also depends on the quality of the acquired images. In this study, the quality of the training samples and the images acquired over the years is relatively consistent, with very low cloud cover, clear images, and distinct physical information. Therefore, according to the evaluation results, all post-classification images from 2018 to 2023 are highly reliable. The results of the LULC classification accuracy assessment, with OA and Kappa values shown in Table 3, ensure reliability for subsequent tasks such as change analysis and LULC trend predicting.

The LULC change results for in the period 2018–2023 are also detailed in Figure 6 and Table 4. The conversion of Bare Land, Water Bodies, Perennial Vegetation, and Annual Vegetation to Built-up Area is 318.33 ha, 470.38 ha, 573.40 ha, and 1994.14 ha, respectively. Additionally, there is a significant conversion of Annual Vegetation to other land cover types: Bare Land, Water Bodies, Perennial Vegetation, and Built-up Area, corresponding to 56.07 ha, 628.9 ha, 704.17 ha, and 1994.14 ha, respectively.

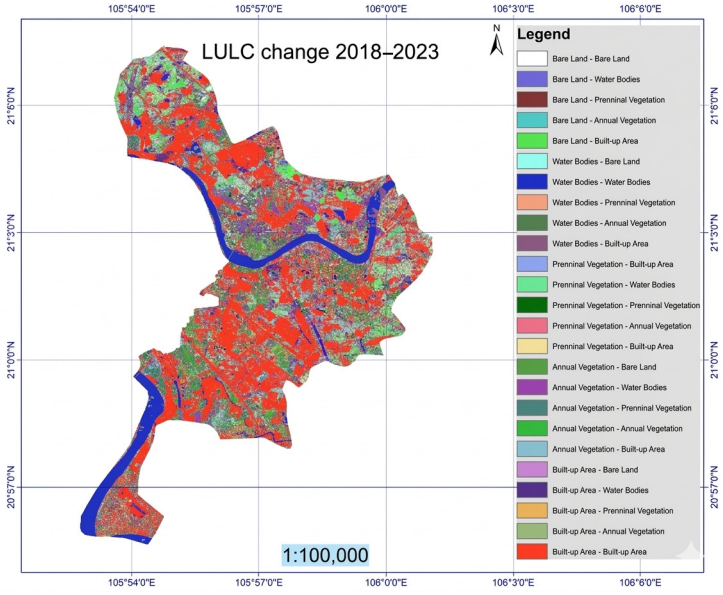

Figure 6. Map of LULC changes in the 2018–2023 period

Table 4. The change in LULC types 2018–2023 period

N°	LULC changes (2018–2023)	Area (ha)	% of total area	% of the from-class area in 2018
1	Bare Land – Bare Land	25.18	0.198%	5.37%
2	Bare Land – Water Bodies	40.5	0.318%	8.64%
3	Bare Land – Prenninal Vegetation	44.22	0.347%	9.44%
4	Bare Land – Annual Vegetation	40.28	0.316%	8.60%
5	Bare Land – Built-up Area	318.33	2.497%	67.95%
6	Water Bodies – Bare Land	36.37	0.285%	2.16%
7	Water Bodies – Water Bodies	1087.92	8.534%	64.60%
8	Water Bodies – Prenninal Vegetation	39.41	0.309%	2.34%
9	Water Bodies – Annual Vegetation	49.98	0.392%	2.97%
10	Water Bodies – Built-up Area	470.38	3.690%	27.93%
11	Prenninal Vegetation – Bare Land	19.31	0.151%	1.78%
12	Prenninal Vegetation – Water Bodies	117.37	0.921%	10.80%
13	Prenninal Vegetation – Prenninal Vegetation	227.33	1.783%	20.91%
14	Prenninal Vegetation – Annual Vegetation	149.73	1.175%	13.77%
15	Prenninal Vegetation – Built-up Area	573.4	4.498%	52.74%

End of Table 4

N°	LULC changes (2018–2023)	Area (ha)	% of total area	% of the from-class area in 2018
16	Annual Vegetation – Bare Land	56.07	0.440%	1.45%
17	Annual Vegetation – Water Bodies	628.9	4.933%	16.26%
18	Annual Vegetation – Prenninal Vegetation	704.17	5.524%	18.21%
19	Annual Vegetation – Annual Vegetation	484.38	3.800%	12.52%
20	Annual Vegetation – Built-up Area	1994.14	15.643%	51.56%
21	Built-up Area – Bare Land	74.4	0.584%	1.32%
22	Built-up Area – Water Bodies	409.34	3.211%	7.26%
23	Built-up Area – Prenninal Vegetation	350.73	2.751%	6.22%
24	Built-up Area – Annual Vegetation	258.79	2.030%	4.59%
25	Built-up Area – Built-up Area	4547.15	35.670%	80.62%

The summary of LULC changes for the period 2018–2023 is presented in Figure 7. The land cover classes show different rates of increase and decrease in area as a percentage of the total natural area of the study region. Specifically, Bare Land decreased by 2.02%, Water Bodies increased by 4.71%, Perennial Vegetation increased by 2.19%, Annual Vegetation decreased by 22.63%, and Built-up Area increased by 17.75%.

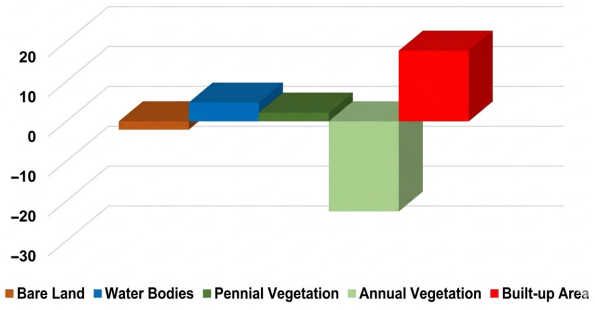

Figure 7. LULC area difference chart 2018–2023

4.2. Predicting land use/land cover changes

The MOLUSCE (Modules for Land Use Change Evaluation) model predicts future land use changes, often influenced by historical changes and impacting factors. This model is quite suitable for application in developing decision support systems for land use issues on a geographic information system (GIS) platform. Figure 8 illustrates the workflow diagram of the LULC change prediction process for 2028 based on the MOLUSCE model.

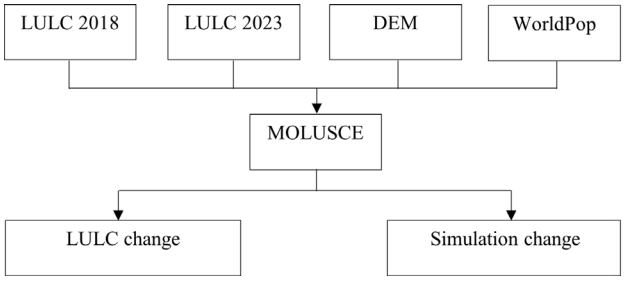

Figure 8. LULC change predict diagram 2028

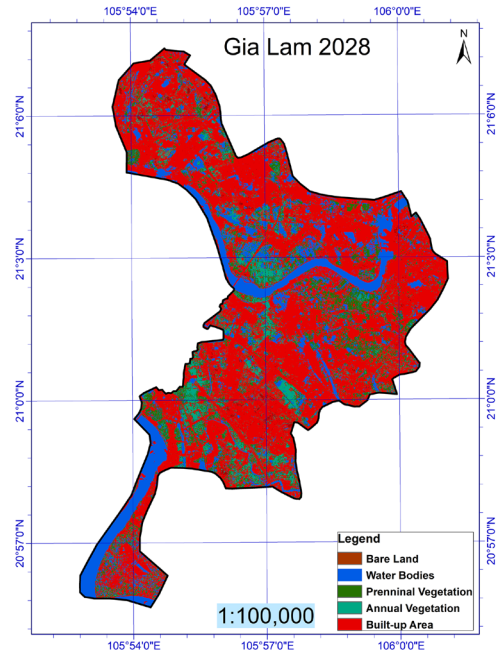

Figure 9. LULC predict map 2028

In QGIS, using the Molusce model also allows for the calculation of change areas. Figure 9 shows the predicted land cover map for 2028, based on the CA-ANN algorithms analyzed above and reference data. The predict results are detailed for each LULC for the year 2028.

According to the results calculated from the Molusce model, by 2028, the area of Bare Land will reach 21,133 ha, Water Bodies will be 228403 ha, Perennial Vegetation will be 136586 ha, Annual Vegetation will be 98316 ha, while Built-up Area will be 790340 ha. This indicates that the forecasted trend is a clear decrease in the area of Annual Vegetation, whereas the Built-up Area shows an increase of approximately 17.75% and 6.93% compared to 2018 and 2023, respectively. Table 5 summarises the projected percentage changes and area proportions for each LULC class in 2028 compared with 2018 and 2023.

From the results table on the area changes of land use cover forecasted for 2028 compared to 10 years earlier in 2018 and 5 years earlier in 2023, Gia Lam district, Hanoi

Table 5. Predicted area defference results to 2028

	Defference 2018–2028 (%)	Defference 2023–2028 (%)	Area in 2028 (%)
Bare Land	–2.018	1.065057	1.66
Water Bodies	4.706	8.262074	17.92
Prenninal Vegetation	2.187	1.815344	10.71
Annual Vegetation	–22.628	–18.0701	7.71
Built-up Area	17.752	6.92667	68.92

city, shows rapid urban development through the significant expansion of the Built-up Area cover. By 2028, the Built-up Area will account for approximately 68.92% of the total natural area of Gia Lam district. The study results on the Built-up Area cover changes from 2018 to 2023 and the forecast product for 2028 indicate that the expansion rate in the 5-year period from 2018 to 2023 was very high (17.77%), while the forecasted expansion rate for the next 5-year period from 2023 to 2028 is much lower (6.92%) compared to the previous 5 years. This result is also consistent with the general development trend and is largely dependent on land management and the actual land fund situation in Gia Lam district for Built-up Area development, which is not much compared to the total natural area. Figure 10 compares the spatial distribution of built-up areas in 2018, 2023, and the projected map for 2028, illustrating a continued but slowing trend of urban expansion.

4.3. Discussion

The study results on land use cover changes from 2018 to 2023 in Gia Lam District, Hanoi, show that the Water Bodies and Perennial Vegetation covers experienced less fluctuation compared to other covers. Specifically, the Water Bodies cover tended to decrease by about 0.59% per year, while the Perennial Vegetation cover showed a slight increase of about 0.06% per year. For the Bare Land cover in the study area, there was no clear pattern of fluctuation between years, with the 1.66% (2023), averaging a decrease of 0.51% per year from 2018 to 2023. The Annual Vegetation cover had an average annual decrease of 0.76% during the study period. The research results shown in Table 4 also indicate the main role of urbanization area expansion, which occurred across most land cover classes during the 2018–2023 period. The largest conversion to Built-up Area was from Annual Vegetation, with 51.56% (1,994.14 ha) of the 2018 Annual Vegetation class transitioning to Built-up Area. A similarly substantial proportion of Perennial Vegetation (52.74%; 573.40 ha) was converted to Built-up Area. The Bare Land exhibited the highest proportional conversion rate to Built-up Area, at 67.95% (318.33 ha), indicating near-complete incorporation of vacant or underutilized land into urban development. Water Bodies also experienced significant transformation, with 27.93% (470.38 ha) converted to Built-up Area, reflecting substantial infilling of rivers, ponds, and other aquatic features due to land reclamation and infrastructure projects. These conversion patterns, particularly the dominant transfer from vegetation classes (especially Annual Vegetation) to Built-up Area, are consistent with observed trends in Hanoi's peri-urban expansion, where agricultural and open lands have served as the primary sources for residential, commercial, and industrial growth. The high conversion rate from Bare Land to Built-up Area further underscores the opportunistic nature of urban development on readily available sites, while the decline in water body area highlights significant environmental trade-offs associated with urban expansion. The reasons for these increases and decreases, besides the actual land use situation varying each year, could also be due to the timing of image acquisition and image quality, which may introduce classification errors. Thus, it can be seen that over the 5 years of development, Gia Lam's development rate is relatively high, clearly evidenced by the rapid expansion of Built-up Area, residential areas, commercial buildings, etc., while most other LULC covers have decreased in area annually.

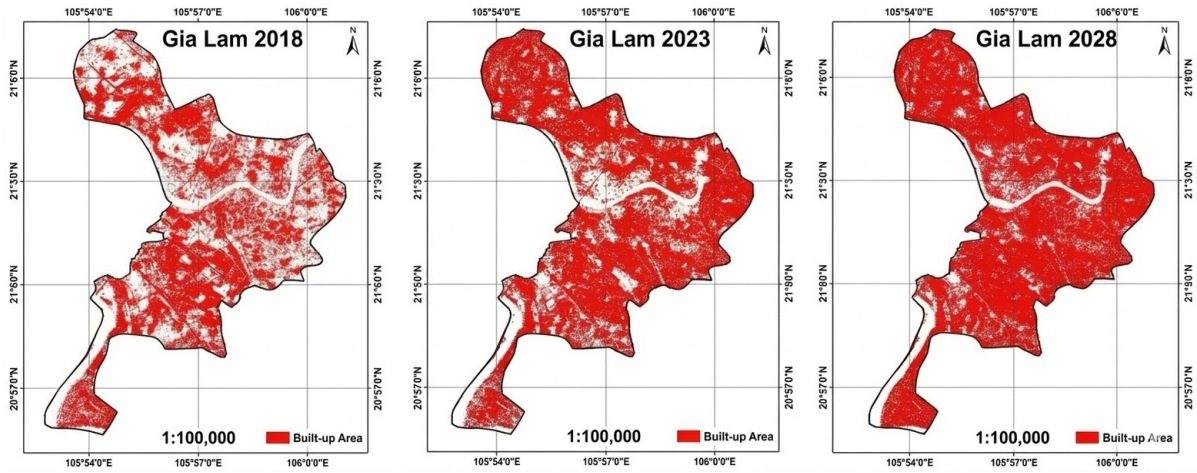

Figure 10. Bult-up Area map: current status 2018, 2023, predict 2028

The predict results for LULC changes in 2028 are clearly aligned with the urban growth rate of Gia Lam district, Hanoi city. The Water Bodies cover is also forecasted to increase, occupying up to 17.92% of the natural area. The Perennial Vegetation cover is expected to expand insignificantly, and the Bare Land cover is forecasted to have a very small expansion, accounting for 1.66% of the area in 2028. There are many reasons and different influencing factors, including human impact, climate change, and socio-economic development trends, leading to the land use change results as forecasted in this study. Notably, the results show a clear inverse correlation between the area changes of Built-up Area and Annual Vegetation in the study area. By 2028, the Annual Vegetation cover will have decreased by 22.63% compared to 10 years earlier (2018) and by about 18.07% compared to 5 years earlier (2023), occupying only about 7.71% of the total natural area of Gia Lam district, Hanoi city.

The exploitation and analysis of online satellite image data for land management are increasingly proving to be efficient, fast, and cost-effective. The research method has provided reliable results, and the analyzed data is often the most recent. These results are highly detailed, timely, and show high effectiveness for land use planning, LULC change monitoring, and forecasting. The application of artificial intelligence with the RF machine learning algorithm to classify LULC from Sentinel-2 satellite images in Gia Lam district, Hanoi city, has advantages such as speed, high automation capability, and the convenience of high-resolution data, which is easy to use, time-saving, and significantly cost-effective compared to other methods and types of materials.

This study uses a hybrid RF + CA-ANN approach, achieving high classification accuracy (OA = 0.94, Kappa = 0.89) and providing clear spatial simulations through the MOLUSCE plugin. However, several limitations remain. The model assumes stationarity in transition rules based on 2018–2023 patterns, which may not hold true under future policy shifts, climate change, or socio-economic transitions. Furthermore, the limited set of input variables (DEM, 2023 population data, and historical LULC) excludes factors such as road proximity or zoning regulations; this could potentially lead to an overestimation of urban expansion in areas where legal constraints on agricultural land protection are strictly enforced. Additionally, the seasonality of Sentinel-2 data acquisition may introduce minor biases in vegetation classification, although these were largely mitigated by consistent sensing dates and robust RF training.

Future research could incorporate additional spatial drivers, such as road networks and zoning layers, or utilize ensemble methods to address uncertainty and non-stationarity. Despite these constraints, this study provides a reliable, reproducible, and cost-effective framework for short-to-medium-term LULC forecasting in data-scarce peri-urban regions, offering valuable support for sustainable land-use planning in rapidly urbanizing districts like Gia Lam.

5. Conclusions

The research results have shown the LULC changes in Gia Lam district, Hanoi city from 2018 to 2023. During this 5-year period, there were changes in different areas of each type of LULC. After 5 years, the study area clearly showed an increase in the Built-up Area with an average growth rate of 1.80% per year. In contrast, the Annual Vegetation had an average annual decline during the study period of 0.76% compared to the total natural area in the study area. Thus, it can be seen that during the 5-year development period, the development speed of Gia Lam is relatively high, clearly due to the rapid development and expansion of residential areas, apartment buildings, and commercial housing areas.

In addition, the research product on predicting changes in LULC up to 2028 in Gia Lam district, Hanoi city also shows a similar trend to the 5-year change period from 2018 to 2023. However, the rate of change in area of the LULC is different for the predict period up to 2028. The results show that by 2028, the Built-up Area cover for 68.92% of the natural totals area, which is also a trend showing the suitability with the economic - social growth rate and urbanization of Gia Lam district, Hanoi city in the next 5 years.

The method of using artificial intelligence and RF machine learning algorithms combined with QGIS open source software has proven its effectiveness in analyzing, managing land use and monitoring the changes in LULC of Gia Lam district, Hanoi city in the period of 2018–2023. This is an effective solution, ensuring reliability, fast processing time, no cost for data and for renting and purchasing processing software. The research products are capable of serving the annual and periodic statistics and inventory work in LULC, supporting planning and monitoring LULC in near real time; Creating accurate and timely updated LULC maps; Capable of being used in monitoring and evaluating the status of land resources... helping managers to make decisions on land management quickly and effectively.

Acknowledgements

The authors would like to thank Hanoi University of Natural Resources and Environment for supporting this research, as well as all the contributing authors for their efforts toward the final version. We also express our gratitude to the reviewers and editors of Geodesy and Cartography for their valuable support. Additionally, we thank the European Space Agency (ESA) for providing free access to Sentinel-2 data under the Copernicus program, and Google for offering free access to the Google Earth Engine (GEE) cloud computing platform, WoldPop for providing population data, and National Aeronautics and Space Administration (NASA) for providing DEM data for the study area.

Disclosure statement

The author has no competing financial, professional, or personal interest in the other part.

References

Anh, D. (2022). *Gia Lam is always the leading unit in the district block with rapid urbanization speed*. https://thanglong.chinhphu.vn/gia-lam-luon-la-don-vi-dan-dau-khoi-huyen-co-toc-do-do-thi-hoa-nhanh-103220118163530724.htm

Assede, E. S., Orou, H., Biaou, S. S., Geldenhuys, C. J., Ahononga, F. C., & Chirwa, P. W. (2023). Understanding drivers of land use and land cover change in Africa: A review. *Current Landscape Ecology Reports*, *8*(2), 62–72. https://doi.org/10.1007/s40823-023-00087-w

Astuty, Y. I., & Dimyati, M. (2024). Prediction of land use/land cover change in Indonesia using the open source land cover dataset: A review. *Geodesy and Cartography*, *50*(2), 67–75. https://doi.org/10.3846/gac.2024.19285

Baig, M. F., Mustafa, M. R. U., Baig, I., Takaijudin, H. B., & Zeshan, M. T. (2022). Assessment of land use land cover changes and future predictions using CA-ANN simulation for selangor, Malaysia. *Water*, *14*(3), Article 402. https://doi.org/10.3390/w14030402

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*(5), 423–429. https://doi.org/10.1016/0895-4356(93)90018-V

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

Halmy, M. W. A., Gessler, P. E., Hicke, J. A., & Salem, B. B. (2015). Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. *Applied Geography*, *63*, 101–112. https://doi.org/10.1016/j.apgeog.2015.06.015

Hoan, P. X., & Hoan, P. H. (2024). Research on the process of automatically detecting terrain changes and geographic objects based on cloud computing technology. *Magazine of Geodesy – Catography*, *10*(03), 15–20.

Islam, K., Rahman, M. F., & Jashimuddin, M. (2018). Modeling land use change using cellular automata and artificial neural network: The case of Chunati Wildlife Sanctuary, Bangladesh. *Ecological Indicators*, *88*, 439–453. https://doi.org/10.1016/j.ecolind.2018.01.047

Karimi, H., Jafarnezhad, J., Khaledi, J., & Ahmadi, P. (2018). Monitoring and prediction of land use/land cover changes using CA-Markov model: A case study of Ravansar County in Iran. *Arabian Journal of Geosciences*, *11*, 1–9. https://doi.org/10.1007/s12517-018-3940-5

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). *Global land use / land cover with Sentinel 2 and deep learning* [Paper presentation]. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium. https://doi.org/10.1109/IGARSS47720.2021.9553499

Khan, D., & Khan, N. (2025). Modelling urban future: Integrating CA-ANN model for comprehensive understanding of land use, land cover changes, and temperature dynamics in Lucknow City, India. *Geology, Ecology, and Landscapes*, 1–26. https://doi.org/10.1080/24749508.2025.2524207

Kouassi, C. J. A., Qian, C., Khan, D., Achille, L. S., Kebin, Z., Omifolaji, J. K., Ya, T., & Yang, X. (2024). Land use land cover change mapping from Sentinel 1B & 2A imagery using random forest algorithm in Côte d'Ivoire. *Geodesy and Cartography*, *50*(1), 43–59. https://doi.org/10.3846/gac.2024.18724

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

Li, J., Chen, H., Zhang, C., & Pan, T. (2019). Variations in ecosystem service value in response to land use/land cover changes in Central Asia from 1995–2035. *PeerJ*, *7*, Article e7665. https://doi.org/10.7717/peerj.7665

Liang, Y., Hashimoto, S., & Liu, L. (2021). Integrated assessment of land-use/land-cover dynamics on carbon storage services in the Loess Plateau of China from 1995 to 2050. *Ecological Indicators*, *120*, Article 106939. https://doi.org/10.1016/j.ecolind.2020.106939

Lin, Y.-P., Chu, H.-J., Wu, C.-F., & Verburg, P. H. (2011). Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling–a case study. *International Journal of Geographical Information Science*, *25*(1), 65–87. https://doi.org/10.1080/13658811003752332

Liping, C., Yujun, S., & Saeed, S. (2018). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. *PloS ONE*, *13*(7), Article e0200493. https://doi.org/10.1371/journal.pone.0200493

Liu, Y., Wang, Y., & Zhang, J. (2012, September 14–16). *New machine learning algorithm*: *Random forest* [Paper presentation]. Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China. https://doi.org/10.1007/978-3-642-34041-3

Liu, Y., & Wu, H. (2017). *Prediction of road traffic congestion based on random forest* [Paper presentation]. 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China. https://doi.org/10.1109/ISCID.2017.216

Marshall, L., Biesmeijer, J. C., Rasmont, P., Vereecken, N. J., Dvorak, L., Fitzpatrick, U., Francis, F., Neumayer, J., Ødegaard, F., Paukkunen, J. P. T., Pawlikowski, T., Reemer, M., Roberts, S. P. M., Straka, J., Vray, S., & Dendoncker, N. (2018). The interplay of climate and land use change affects the distribution of EU bumblebees. *Global Change Biology*, *24*(1), 101–116. https://doi.org/10.1111/gcb.13867

Musleh, A. A., & Jaber, H. S. (2021). Comparative analysis of feature extraction and pixel-based classification of high-resolution satellite images using geospatial techniques. *E3S Web of Conferences*, *318*, Article 04007. https://doi.org/10.1051/e3sconf/202131804007

Owojori, A., & Xie, H. (2005). *Landsat image-based LULC changes of San Antonio, Texas using advanced atmospheric correction and object-oriented image analysis approaches* [Paper presentation]. 5th International Symposium on Remote Sensing of Urban Areas, Tempe, AZ.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Phuong, D. L., Thuy, H. T., & Hiep, D. N. (2024). Application of Artificial Intelligence to monitor changes in land use in the BacTu Liem District area, Hanoi, during the period 2019-2023. *Magazine of Geodesy – Cartography*, *10*(01), 22–29. https://zenodo.org/records/13218549

Rahman, M. T. U., Tabassum, F., Rasheduzzaman, M., Saba, H., Sarkar, L., Ferdous, J., Uddin, S. Z., & Zahedul Islam, A. Z. M. (2017). Temporal dynamics of land use/land cover change and

its prediction using CA-ANN model for southwestern coastal Bangladesh. *Environmental Monitoring and Assessment*, *189*, Article 565. https://doi.org/10.1007/s10661-017-6272-0

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Sajan, B., Mishra, V. N., Kanga, S., Meraj, G., Singh, S. K., & Kumar, P. (2022). Cellular automata-based artificial neural network model for assessing past, present, and future land use/land cover dynamics. *Agronomy*, *12*(11), Article 2772. https://doi.org/10.3390/agronomy12112772

SUHET. (2015). *Sentinel-2 user handbook* (Vol. 2 rev.). European Space Agency Agence spatiable europeenne (ESA).

Swetanisha, S., Panda, A. R., & Behera, D. K. (2022). Land use/land cover classification using machine learning models. *International Journal of Electrical Computer Engineering*, *12*(2), 2040–2046. https://doi.org/10.11591/ijece.v12i2.pp2040-2046

Thien, B. B., Alexsander, I. R., & Denis, K. O. (2025). *Prediction of future land use and land cover changes using a coupled CA-ANN model in Hanoi capital*, *Vietnam* [Paper presentation]. Problems of Coastal Area Management to Ensure Environmental Safety and Rational Environmental Management, Cham. https://doi.org/10.1007/978-3-031-90873-6_1

Von Neumann, J. (2017). The general and logical theory of automata. In *Systems research for behavioral science* (pp. 97–107). Routledge.

Von Neumann, J., & Burks, A. W. (1966). *Theory of self-reproducing automata*. University of Illinois Press.

White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*: *Economy and Space*, *25*(8), 1175–1199. https://doi.org/10.1068/a251175

Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Geosciences Computers*, *34*(6), 592–602. https://doi.org/10.1016/j.cageo.2007.08.003