Chelonian Conservation And Biology

Vol. 19 No. 1 (2024) | https://www.acgpublishing.com/ | ISSN - 1071-8443

DOI: doi.org/10.18011/2024.01(1).1577-1585

APPLYING THE GOOGLE EARTH ENGINE CLOUD COMPUTING PLATFORM IN ENVIRONMENTAL RESEARCH: FORECASTING FUTURE LAND USE AND LAND COVER MAP

Yen Hai Hoang[1,], Dong Phuong Nguyen[2], Hoa Thi Nguyen[2], Quoc Phi Nguyen[2], Ngoc Thi Tran[2]

[1]Mining and Geological Technology Consulting and Implementation JSC.

[2] Hanoi University of Mining and Geology

No.18 Vien Street - Duc Thang Ward- Bac Tu Liem District - Hanoi- Vietnam

[*] Email: Hoangyendcbk59@gmail.com

Summary

Hanoi City is Vietnam's cultural, political, and economic centre, where dense population and rapid urbanisation pose many challenges to planning effective land use and sustainability. The study used a real-time satellite image data set for land cover classification Dynamic World V1 data to analyse land cover changes for Hanoi city from 2018 to 2022 combined with high-resolution digital elevation data and using random forest algorithms to predict future land cover classification. The results show a significant increase in Crops and Built areas in the future and a substantial decrease in the city's agricultural land, trees, and bare land. The model has a reasonably high classification accuracy of about 74.2% and a Kappa coefficient of 0.6688. The research results have also shown the ability to apply cloud computing on the Google Earth Engine tool platform to forecast land use cover changes. This is a scientific basis that provides information to help managers come up with long-term strategies for sustainable urban development.

Keywords: *Google Earth Engine, Cloud Computing, Land cover, Land use*

1. Introduction

Hanoi, the capital of Vietnam, is a large city with rapid development in Southeast Asia. However, the population explosion, robust urbanization process, and the expansion of urban areas, industrial areas and transportation infrastructure have changed the regional spatial structure and posed significant challenges in the city's land use planning and management. Changes in land use, the increase in concrete in urban areas of Hanoi city and the decrease in water surface land, green space and vegetation cover are the leading causes of the decline of regional climate regulation, adding to the urban heat island phenomenon and environmental problems. Therefore, land use and land cover (LULC) change research will be necessary in the context of the increasing pressure on urbanization in Hanoi.

Remote sensing images have long played an essential role in monitoring and assessing land use changes in particular and environmental changes in general (Dang et al., 2017), because, in reality, the investigation and assessment process encounters many problems difficult because of the scope of research, funding, as well as the ability to continuously update new data to evaluate fluctuations. The research of Yu and Gong (Yu & Gong, 2012) has identified challenges that traditional geographic information science faces, including the preprocessing and mining of extensive geographic data and the need for techniques that require time to study geographical processes. Significant technological advances are needed to solve these problems, and Google Earth Engine (GEE) is an effective tool for solving these issues. Google Earth Engine is not only a powerful cloud computing platform that provides a huge data warehouse with satellite image data from Landsat, Sentinel and other spatial datasets but also allows users to easily access, process, analyse and calculate these data (Le Minh & Bakaeva, 2023). Using cloud computing tools combined with machine learning models, especially Random Forest, has proven to be highly effective in predicting and classifying environmental data. Random Forest is an ensemble algorithm that uses multiple decision trees for prediction, helping to minimise prediction error and increase accuracy *(Breiman, 2001). Applying Random Forest in LULC forecasting provides the ability to accurately analyze complex land-use data changes, thanks to the ability to process many different inputs and create detailed forecasts. In particular, Random Forest can handle large and diverse datasets, which is suitable for the complex nature of LULC data (Belgiu & Drăguţ, 2016).*

This article aims to introduce and evaluate the application of the Google Earth Engine cloud computing platform in forecasting future land use and land cover maps for Hanoi city. In this study, the authors used the Random Forest model to forecast LULC changes to 2030, providing an overview of land use potential and trends, thereby helping to support the land use planning process. Plan and manage land resources effectively and sustainably.

2. Research methods

Remote sensing technology is increasingly developing strongly; high-resolution satellite images have contributed to solving problems in climate change, natural resource management, climate change assessment, impact of surface cover and land use purposes. Using remote sensing image sources combined with cloud computing allows for an overview and quick-to-update assessment.

The study used the following datasets:

- The satellite image data set for land cover classification Dynamic World V1 is a 10m near-real-time (NRT) Land Use/Land Cover (LULC), which includes information for land cover types. Dynamic World V1 classifies land cover into nine distinct classes: 1) Water, 2) Trees, 3) Grass, 4) Flooded Vegetation, 5) Crops, 6) Shrub and Scrub, 7) Built Area, 8) Bare Ground, 9) Snow and Ice (Brown et al., 2022). For the study area of Hanoi city, the layer of soil cover label information used is 8 types of land cover (1-8).

Table 1. Typical land covers in Hanoi

№	Typical land covers	Value	Color
1	Water	1	
2	Trees	2	
3	Grass	3	
4	Flooded_vegetation	4	
5	Crops	5	
6	Shrub and scrub	6	
7	Built	7	
8	Bare	8	

- The NASA Shuttle Radar Topography Mission (SRTM) Digital Elevation 30m dataset is a valuable resource for a broad range of applications, offering high-resolution elevation data with global coverage (Farr et al., 2007).

Satellite images, after collection and preprocessing, are classified to create LULC maps for the years 2018, 2020 and 2022 with 8 types of land cover corresponding to a specific value and color, shown according to the following code:

```
// Land cover parameter
var values = [1, 2, 3, 4, 5, 6, 7, 8];
var palette = [
  '419BDF', '397D49', '888053', '7A87C6', 'E49635', 'DFC35A', 'C4281B', 'A59B8F'
];
var names = [
  'Water', 'Trees', 'Grass', 'Flooded_vegetaion', 'Crops','Shrub and crub', 'Built', 'Bare'
];

// Show legend
legend(palette, values, names);

// Land cover dictionary for visualization
var lulcDict = {
  'LULC_class_palette': palette,
  'LULC_class_values': values,
  'LULC_class_names': names
};

// Land cover data for 2020 and 2022
var lulc2018 = ee.Image('projects/ee-projecthanoi/assets/Lulc2018');
var lulc2020 = ee.Image('projects/ee-projecthanoi/assets/Lulc2020');
var lulc2022 = ee.Image('projects/ee-projecthanoi/assets/Lulc2022');
```

The sample data set is randomly stratified for the research area and divided into 2 data sets: 80% of the sample points are used to train the machine learning model, and the remaining 20% of the sample points are used. Used to verify model quality, evaluate accuracy, and Kappa constant. After creating the Random Forest model, each decision tree result in the set will vote for the most common class and produce a classification result (Figure 1). The model will be created based on the classification of factors with the most impact index; this shows the role and influence of each parameter variable selected in the machine learning training of the model on the prediction results (Teluguntla et al., 2018).

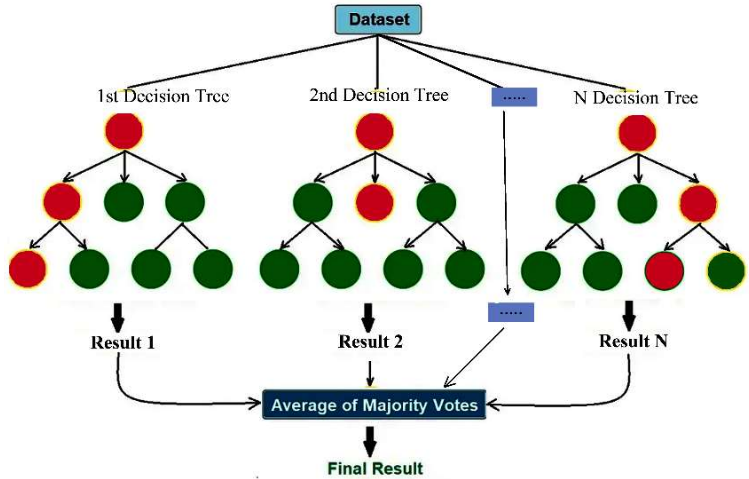
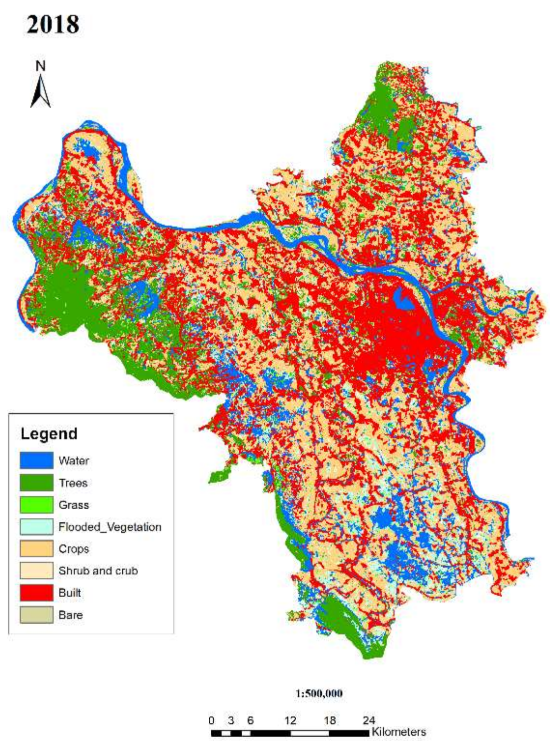


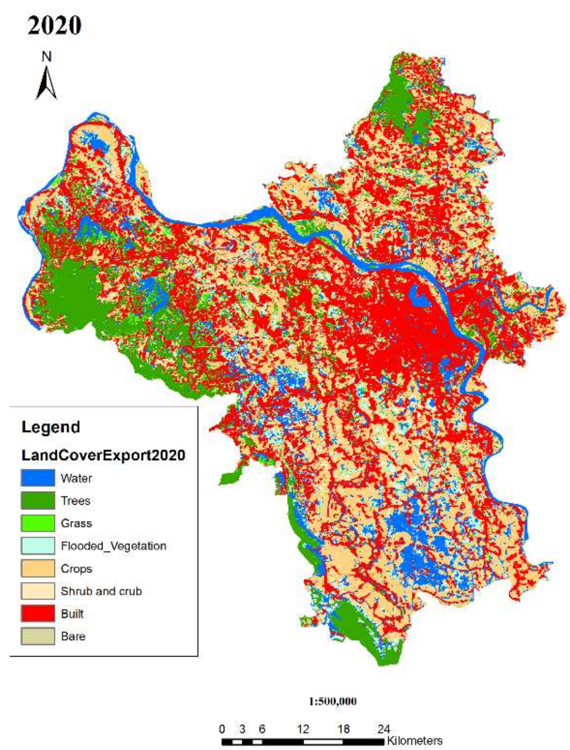
Figure 2. Using Random Forest (RF) algorithm

3. Results and discussion

3.1. Map of land use and land cover over the years in Hanoi city

Figure 3 shows the change in the area of land use and land cover (LULC) of Hanoi City from 2018 to 2022. The area of water surface land has a slight increase over the years the area The area of the Flooded vegetation, Bare, and Trees coatings will decrease in 2022, replaced by an increase in the area of the Crops and Built coatings, as shown in table 2.





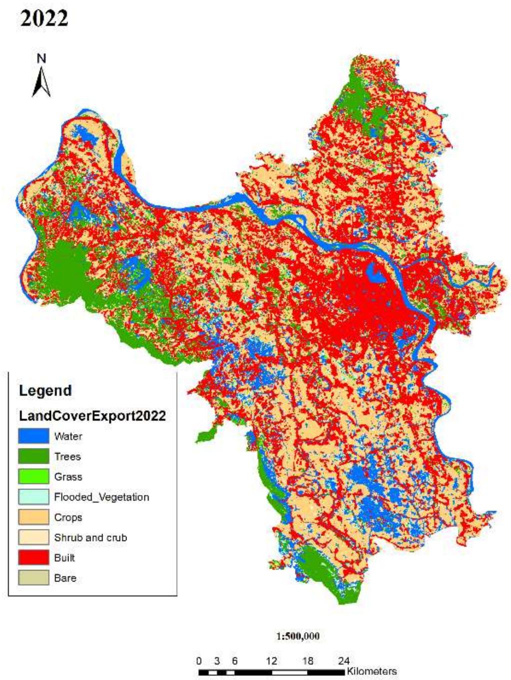


Figure 3. Changes LULC in Hanoi city over the years from 2018 to 2022

3.2. Establishment of a map predicting LULC in Hanoi city in 2030

The data set built for the machine learning model with the random forest algorithm includes selected variable layers, including the land use/land cover maps for years 2018, 2020 and 2022, images representing the changes LULC between 2018 - 2020 and 2020 - 2022, and digital elevation map Through the results for the accuracy index and Kappa, the study chose to apply a machine learning model using the random forest algorithm with 20 decision trees; the input data is the 2020 land use map (starting point), 2022 land use map (ending point), images representing the changes LULC of 2022 year and digital elevation map. By dividing the data into 80% training and 20% matching testing, the model has 149009 samples for training and 37588 data samples for validation and performance testing. This balance helps build a robust model and evaluate its generalizability. Model evaluation results showed that the overall accuracy was 74.2%, and the Kappa index was 0.6688 (Table 1).

Table 1. Results of confusion matrix for classifying LULC in Hanoi city

LULC	Water	Trees	Grass	Flooded Vegetation	Crops	Shrub & scrub	Built	Bare	UA (%)
Water	**4778**	51	10	0	960	0	471	0	*76.2*
Trees	80	**3303**	2	0	574	9	554	2	*73.0*

Grass	98	32	**1085**	10	442	2	203	1	*57.9*
Flooded Vegetation	105	52	9	**616**	392	9	137	0	*46.7*
Crops	535	309	16	3	**8423**	19	1981	7	*74.6*
Shrub & scrub	48	184	4	0	290	**686**	319	3	*44.7*
Built	239	205	1	0	961	17	**8192**	2	*85.2*
Bare	12	13	1	0	206	2	98	**825**	*71.3*
PA(%)	*81.1*	*79.6*	*96.2*	*97.9*	*68.8*	*92.2*	*68.5*	*98.2*	

According to the random forest model, the confusion matrix table of each soil cover type shows that most component accuracy is over 60%, except for the shrub and scrub, Flooded vegetation, and Grass cover layers, which have UA of 44.7%, 46.7% and 57.9%. Among all the types of coatings mentioned above, the water surface is the characteristic type that is very easy to distinguish on images with high accuracy, and this is entirely similar to the study of Li et al. (Li et al., 2023).

From the results of analyzing statistical parameters in evaluating the model's performance, the study applied the random forest model to predict the change in the land cover area of Hanoi city until 2030; the results are shown in Table 2 and Figure 4.

Table 2. LULC change over the years and forecast for 2030

Year / LULC	2018	2020	2022	2030
Water	36,917	37,302	37,507	32,546
Trees	38,796	41,774	38,031	32,804
Grass	8,770	10,230	6,829	1,996
Flooded_vegetaion	14,376	14,245	7,608	2,152
Crops	122,492	109,014	117,642	133,880
Shrub and scrub	7,779	8,035	8,460	1,729
Built	103,135	111,339	117,344	126,233
Bare	1,120	1,071	947	428

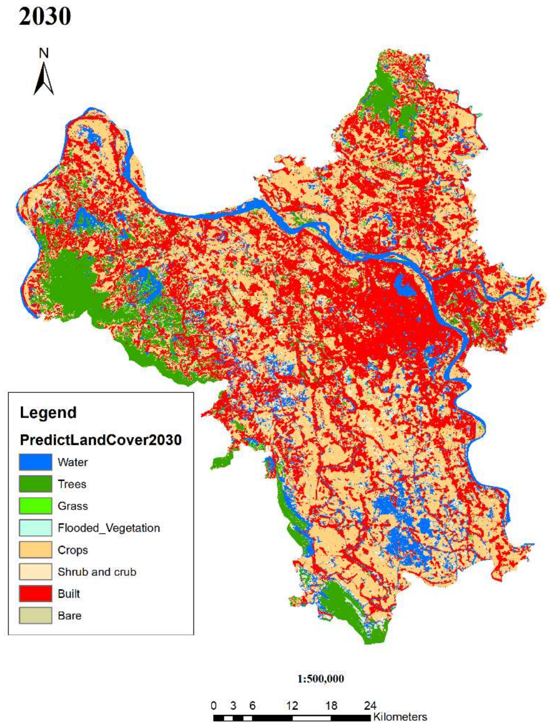


Figure 3. Forecast of land use/land cover in Hanoi city in 2030

4. Conclusion

This study uses 8 initial data layers representing the characteristics of each type of land cover used and the terrain elevation data layer in the Hanoi city area to build a change forecast map LULC on a random forest machine learning model. The results obtained from the Random Forest algorithm show a high accuracy rate of 74.2%, and the Kappa constant is 0.6688, shown through the comparison of accurate statistical parameters when evaluating the performance of the training and test dataset.

The research results have also shown the ability to apply cloud computing on the Google Earth Engine tool platform to forecast land use cover changes. This is a scientific basis that provides information to help managers and policymakers come up with long-term strategies for sustainable urban development.

Acknowledgements: The work presented in this paper is based on research support by the project: "Research and application of simulation model on the impact of the heat island effect on the atmosphere in urban areas of Hanoi city and proposed mitigation solutions" under contract code B2022-MDA-12. The authors are highly grateful for this support.

Conflict of interest

The authors declare that they do not have any conflict of interest.

References

1. Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing, 114*, 24-31.

2. Breiman, L. (2001, 2001/10/01). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

3. Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., & Ilyushchenko, S. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data, 9*(1), 251.

4. Dang, N. D., Dao, N. L., & Trinh, L. H. (2017, 12/01). Study on the change of land surface temperature in Thanh Hoa city in the period of 2000 – 2017 using Landsat thermal infrared data. *Journal of Geodesy and Cartography*(34), 26-32. https://doi.org/10.54491/jgac.2017.34.255

5. Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., & Roth, L. (2007). The shuttle radar topography mission. *Reviews of geophysics, 45*(2).

6. Le Minh, T., & Bakaeva, N. (2023, 06/27). A Technique for Generating Preliminary Satellite Data to Evaluate SUHI Using Cloud Computing: A Case Study in Moscow, Russia. *Remote Sensing, 15*, 3294. https://doi.org/10.3390/rs15133294

7. Li, Z., Chen, X., Qi, J., Xu, C., An, J., & Chen, J. (2023, 2023/08/09). Accuracy assessment of land cover products in China from 2000 to 2020. *Scientific Reports, 13*(1), 12936. https://doi.org/10.1038/s41598-023-39963-0

8. Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., Yadav, K., & Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS journal of photogrammetry and remote sensing, 144*, 325-340.

9. Yu, L., & Gong, P. (2012, 2012/06/01). Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives. *International Journal of Remote Sensing, 33*(12), 3966-3986. https://doi.org/10.1080/01431161.2011.636081