JOURNAL OF
HYDRO-METEOROLOGY

Research Article

Comparison of random forest and extreme gradient boosting algorithms in land cover classification in Van Yen district, Yen Bai province, Vietnam

Dong Khuc Thanh[1], Dung Luong Ngoc[1]*, Hue Dang Dieu[1], Van Anh Tran[2]

[1] Ha Noi University of Civil Engineering; dungln@huce.edu.vn; dongkt@huce.edu.vn; huedd@huce.edu.vn
[2] Hanoi University of Mining and Geology; tranvananh@humg.edu.vn

*Corresponding author: dungln@huce.edu.vn; Tel.: +84–915157676

Received: 7 February 2025; Accepted: 5 March 2025; Published: 8 June 2025

Abstract: Land cover classification using remote sensing data plays a crucial role in resource management and environmental monitoring. This study compares the performance of Random Forest (RF) and Extreme Gradient Boosting (XGBoost) in land cover classification in Van Yen District, Yen Bai Province, Vietnam. The input data includes Sentinel-1 radar images, Sentinel-2 optical images, and a total of 7,214 sample points collected for model development using the Google Colab platform. The results indicate that both RF and XGBoost achieve high performance, with overall accuracy (OA) ranging from 94.8% to 96.3% and Kappa coefficients between 0.936 and 0.955. Notably, RF demonstrates greater stability and higher accuracy than XGBoost in both scenarios: using Sentinel-2 alone and combining Sentinel-2 with Sentinel-1. This study provides a scientific basis for selecting appropriate algorithms and data to improve land cover classification efficiency in the region.

Keywords: Land cover; Remote sensing images; Random Forest; Extreme gradient boosting.

1. Introduction

Monitoring and classifying land cover is one of the key applications of remote sensing and geographic information systems technologies, providing essential information for resource conservation, environmental management, and sustainable development [1–4]. The advancement of remote sensing technology and artificial intelligence has significantly improved the accuracy and practical applicability of land cover classification methods [3–5].

Numerous studies worldwide and in Vietnam have applied machine learning approaches to develop land cover classification models using remote sensing data. Among these approaches, ensemble models have proven effective in remote sensing image processing for various applications [6–8]. This technique combines multiple weaker machine learning models into a stronger one, reducing overfitting and enhancing classification performance. Ensemble learning commonly employs two techniques: Bagging (Bootstrap Aggregating) and Boosting. Previous studies have demonstrated the advantages of these methods in land cover classification tasks [9–11].

Among ensemble learning algorithms, Random Forest (RF) represents the Bagging approach, whereas Extreme Gradient Boosting (XGBoost) is a leading example of the Boosting technique. Both are highly regarded for their ability to handle complex data and their stability, particularly in land cover classification problems [6]. Although studies using remote sensing imagery and machine learning models for land cover mapping in Vietnam have made significant progress in recent years, there are still few studies conducted in the northern

mountainous region of Vietnam, where the terrain is complex and land cover types are highly mixed [9–14]. Moreover, there has been no comprehensive study comparing the effectiveness of RF and XGBoost for land cover classification in this region.

This study aims to compare the performance of RF and XGBoost in land cover classification using different remote sensing datasets. Van Yen district, a mountainous region in northern Vietnam with diverse terrain and land cover characteristics, was chosen as the study area. The findings will support the selection of suitable methods and datasets for land cover classification in the region while also providing a scientific basis for future research and applications in resource management and land cover monitoring.

2. Materials and Methods

2.1. Study area

Van Yen is a district in the northwest of Yen Bai province, located between 21º40’00” and 22º12’00” north latitude and 104º18’00” to 104º48’00” east longitude (Figure 1). The district shares its eastern boundary with Luc Yen and Yen Binh, while it is bordered by Van Chan to the west, Tran Yen to the south, and Van Ban and Bao Yen of Lao Cai province to the north. Covering a total natural area of 1,391.54 km², Van Yen is situated approximately 40 km north of Yen Bai’s provincial center. The district comprises 24 communes and one town. Mau A town serves as the district’s economic, political, cultural, and social center, strategically positioned along major transportation routes such as the Yen Bai - Lao Cai railway and the Noi Bai - Lao Cai expressway, fostering economic and social development while ensuring national defense and security [18].

The land cover characteristics of Van Yen are distinctly differentiated across its economic zones and natural landscapes. One of its most notable features is the dominance of forest and vegetation cover, particularly in highland communes and cinnamon-growing areas. Natural and plantation forests play a crucial role in environmental protection, ecosystem stability, and soil erosion prevention. Communes such as Phong Du Thuong, Xuan Tam, and Phong Du Ha have extensive cinnamon plantations, contributing to the local economy and forming a unique vegetation cover in the region. Van Yen experiences a high annual frequency of natural disasters, including landslides and floods. These events negatively impact land cover changes in the region, emphasizing the need for continuous land cover monitoring to ensure sustainable

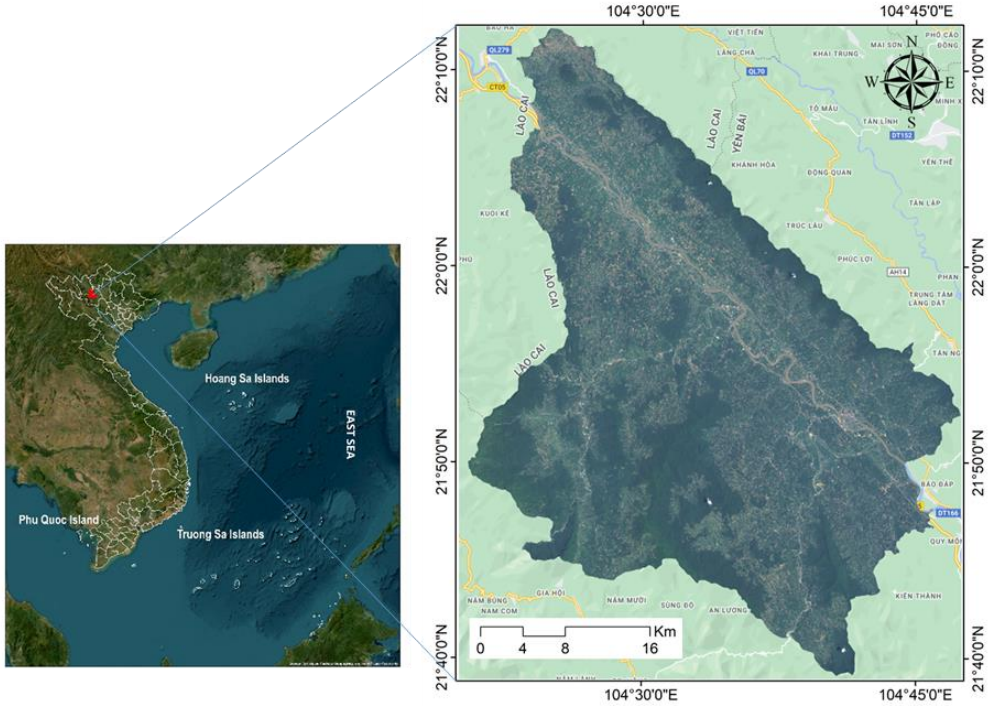

Figure 1. The study area in Van Yen district, Yen Bai province, Vietnam.

development. Additionally, the expansion of transportation networks and residential areas along the Red River and its tributaries in recent years has significantly altered the district's land cover characteristics [19].

2.2. Data collection

2.2.1. Remote sensing images

Sentinel is an Earth observation satellite program under The Copernicus program managed by ESA (European Space Agency), providing free data for environmental monitoring, resource management, and disaster prevention. Sentinel-1 utilizes C-band SAR (Synthetic Aperture Radar) technology, capturing day and night, all-weather, cloud-penetrating data offering spatial resolutions between 10 meters and 40 meters. It acquires different polarization combinations, such as VV and VH for land areas and HH and HV for maritime, ice, and snow regions. Sentinel-1 data have proven effective in monitoring floods, land subsidence, surface deformation, and ship tracking [20].

Sentinel-2 collects multispectral optical images across thirteen spectral bands spanning the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions. With a spatial resolution of up to 10m and a 5-day revisit cycle, Sentinel-2 is highly effective in monitoring vegetation, agriculture, water quality, wildfires, and land surface changes. Sentinel-1 provides cloud-penetrating radar data, capturing surface structure and moisture, while Sentinel-2 offers multispectral information for distinguishing land cover types. The integration of these complementary datasets harnesses their combined strengths to enhance classification accuracy, particularly in regions with complex terrain and heterogeneous land cover [21, 22]. Cloud masks are applied to select Sentinel-2 optical images with cloud cover below 10% and to remove pixels with extremely low reflectance values (< -30 dB) to reduce noise in Sentinel-1 imagery [21]. This study utilizes Sentinel data from 2022 for the study area, including VV and VH polarizations from Sentinel-1, and selected spectral bands from Sentinel-2: visible bands (2, 3, 4), NIR band (8), and SWIR bands (11, 12), with spatial resolutions of 10m and 20m [20].

2.2.2. Sample data

Table 1. Classification and description of characteristic land cover in the study area.

No.	Land Cover Type	Code	Description	Total sample
1	Forest	FO	Distributed in high mountainous areas, dark green, homogeneous.	940
2	Non-Forest Vegetation	VE	Light green or non-uniform green, or divided into large plots.	868
3	Bare Land	BA	Yellow, white, or light brown with high brightness.	1932
4	Built-Up	BU	White, light pink, or red, unevenly distributed in clusters in favorable terrain.	1413
5	Paddy Fields	RI	Green or brown depending on the season, divided into plots or distinct terraces.	717
6	Water	WA	Gray or blue, homogeneous, appearing in linear or area forms such as ponds and reservoirs.	1344

The study selected sample points from a 2022 field survey in accessible and stable areas, such as residential zones and paddy fields. Additionally, image interpretation was conducted using true-color composite images and Google Earth data captured at the same time as the remote sensing images. A total of 7,214 samples were collected, with 4,266 samples used for

training the model and 2,948 samples for accuracy assessment. Table 1 presents the sample data properties and the land cover classification system for the study area.

2.3. Calculation of spectral indices

The study utilizes spectral indices computed from the original bands of Sentinel-2 images. These spectral index layers are then stacked with Sentinel-1 and Sentinel-2 data layers for classification modeling (Figure 2).

- Normalized Difference Vegetation Index (NDVI):

NDVI is commonly applied spectral index for monitoring vegetation cover density under many conditions, based on the difference in spectral reflectance between the Red and Near-Infrared bands in vegetation [23]. It is calculated based on the Red and NIR bands, following the equation (1):

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

where NIR corresponds to Band 8 and Red corresponds to Band 4 of Sentinel-2 imagery.

- Normalized Difference Built-up Index (NDBI):

NDBI is a spectral index that enables high-accuracy mapping of built-up areas from optical satellite imagery, based on the differing reflectance properties of the Near-Infrared and Shortwave Infrared bands [24]. The NDBI equation (2) is:

$$\mathrm{NDBI} = \frac{(\mathrm{SWIR} - \mathrm{NIR})}{(\mathrm{SWIR} + \mathrm{NIR})} \quad (2)$$

where SWIR corresponds to Band 11 and NIR corresponds to Band 8 of Sentinel-2 imagery.

- Modified Bare Soil Index (MBI):

MBI is designed to differentiate bare soil, built-up land, and vegetation by analyzing reflectance in the NIR, SWIR1, and SWIR2 bands. This index provides a basis for distinguishing bare soil from other land cover types. It has been applied using Sentinel-2 spectral bands and has shown promising results in separating bare land from urban areas [25]. The MBI equation (3) is:

$$\mathrm{MBI} = \frac{(\mathrm{SWIR1} - \mathrm{SWIR2} - \mathrm{NIR})}{(\mathrm{SWIR1} + \mathrm{SWIR2} + \mathrm{NIR})} + \mathrm{f} \quad (3)$$

In this formula, SWIR1 refers to Band 11, SWIR2 to Band 12, and NIR to Band 8 of Sentinel-2 imagery, and f is an adjustment factor (f = 0.5).

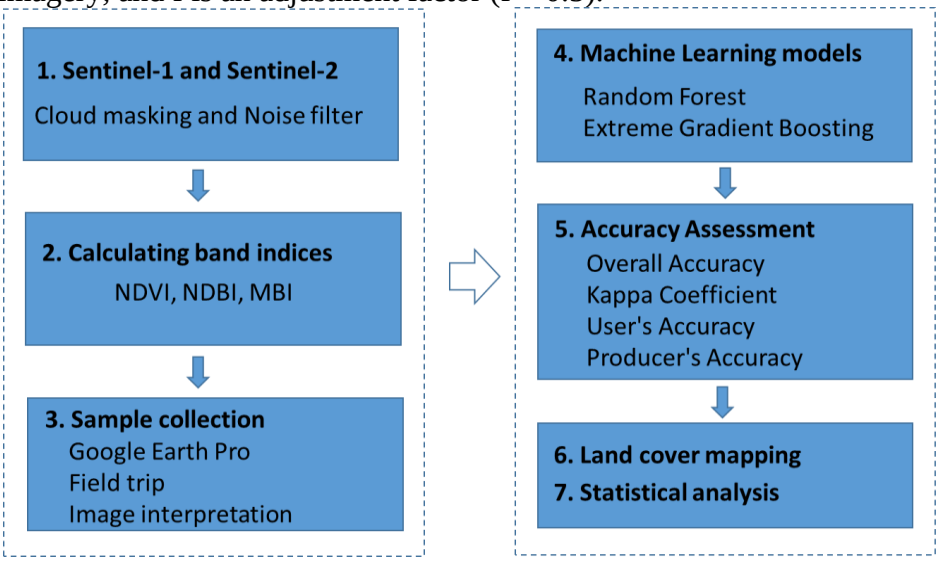

Figure 2. A flowchart of the study structure.

2.4. Machine leaning models

This study develops a land cover classification model using two machine learning techniques: Random Forest (RF) and Extreme Gradient Boosting (XGBoost), with each model comprising 100 decision trees. These models represent two distinct ensemble learning techniques. The classification model was implemented using the scikit-learn and XGBoost libraries on the Google Colab platform, applying different data combinations, including Sentinel-2 multispectral imagery and fused of Sentinel-2 and Sentinel-1 data (Table 2).

2.4.1. Random Forest (RF)

RF is a highly effective ensemble learning algorithm for classification tasks. It operates by constructing numerous decision trees using randomly sampled subsets of the training data, following the bagging technique. A key feature of RF is that the decision trees are trained independently and in parallel. The final classification result is determined through majority voting among the individual trees [26].

2.4.2. Extreme Gradient Boosting (XGBoost)

Among ensemble learning algorithms, XGBoost is recognized as a powerful Boosting technique. It builds a predictive model by iteratively combining less accurate decision trees, enhancing performance by reducing the loss function through gradient descent. XGBoost can be applied to both regression and classification tasks [27]. While RF employs Bagging by constructing independent decision trees and aggregating results through majority voting, XGBoost adopts a sequential learning process, where each tree is built to minimize the errors of the previous ones.

Table 2. Data and model used in the study.

No.	Code	Data	Model
1	Case 1 (C1)	Sentinel-2	RF
2	Case 2 (C2)	Sentinel-2	XGBoost
3	Case 3 (C3)	Sentinel-1, Sentinel-2	RF
4	Case 4 (C4)	Sentinel-1, Sentinel-2	XGBoost

2.5. Accuracy assessment

This study employs overall accuracy (OA), the Kappa coefficient, User's accuracy (UA), and Producer's accuracy (PA) to evaluate the classification model's performance [28]. Overall accuracy (OA) represents the proportion of correctly classified samples across the entire model, calculated as follows:

$$\mathrm{OA} = \frac{\text{Number of correct predictions} \times 100\%}{\text{Total number of predictions}} \tag{4}$$

The Kappa coefficient evaluates the consistency between model predictions and actual values while accounting for the possibility of random agreement. The coefficient ranges from 0 to 1, with higher scores reflecting better classification accuracy. The Kappa coefficient is calculated using the following formula:

$$\mathrm{Kappa} = \frac{P_o - P_e}{1 - P_e} \tag{4}$$

where P_o is the percentage of correctly classified samples; P_e is the proportion of chance agreement.

User's accuracy (UA) indicates the probability that a sample assigned to a particular class truly belongs to that class. Producer's accuracy (PA) represents the proportion of actual class samples that have been correctly classified. A high UA suggests that users can rely on the classification results for that class, whereas a high PA indicates that most samples

belonging to that class have been correctly identified. UA and PA are computed using the following formulas:

$$\mathrm{UA} = \frac{\text{Number correctly identified in a given map class}}{\text{Number claimed to be in that map class}} \quad (6)$$

$$PA = \frac{\text{Number correctly identified in a given map class}}{\text{Number actually in that reference class}} \quad (7)$$

3. Results and Discussion

3.1. Model accuracy assessment

The research results indicate strong performance of both RF and XGBoost machine learning models in land cover classifying land cover based on Sentinel-1 and Sentinel-2 satellite data. Overall, the tested models achieved high accuracy, with overall accuracy (OA) ranging from 94.8% to 96.3%, and Kappa coefficients between 0.936 and 0.955. Among the models, RF exhibited, RF demonstrated greater stability and higher accuracy than XGBoost in both data scenarios. Additionally, classification accuracy improved when combining Sentinel-2 multispectral imagery with Sentinel-1 radar observations.

Table 3. Accuracy assessment results of model C1.

	Confusion Matrix						**Accuracy**	
	BU	WA	FO	VE	BA	RI	UA	PA
BU	165	0	0	0	2	2	0.604	0.976
WA	0	604	0	0	0	0	1	1
FO	0	0	617	0	0	0	1	1
VE	11	0	0	600	0	1	1	0.980
BA	97	0	0	0	292	0	0.993	0.751
RI	0	0	0	0	0	557	0.995	1
OA = 96.2% and Kappa = 0.953								

Table 4. Accuracy assessment results of model C2.

	Confusion Matrix						**Accuracy**	
	BU	WA	FO	VE	BA	RI	UA	PA
BU	162	0	0	1	5	1	0.6353	0.959
WA	0	604	0	0	0	0	1	1
FO	0	0	617	0	0	0	1	1
VE	10	0	0	583	6	13	0.9983	0.953
BA	83	0	0	0	306	0	0.9653	0.787
RI	0	0	0	0	0	557	0.9755	1
OA =96.0% and Kappa = 0.951								

The models exhibited stable classification performance for water (WA), forest (FO), non-forest vegetation (VE), and paddy fields (RI), with user accuracy (UA) and producer accuracy (PA) values approaching or reaching 1 in all four cases. However, the classification accuracy of built-up (BU) and bare land (BA) was lower than for other land cover classes. This is due to the spectral similarity between these two land cover types. Furthermore, the shortwave infrared (SWIR) bands of Sentinel-2 have a resolution of 20 meters, which is relatively low, while these bands are primarily used for classifying BU and BA. Adding indices beyond NDBI and MBI or integrating higher-resolution data could be a promising approach to further improve the accuracy of these subclasses [26, 27]. RF outperformed XGBoost in both single Sentinel-2 data and the combined Sentinel-1 and Sentinel-2 dataset.

Moreover, RF improved classification accuracy for challenging classes such as BU, VE, and BA when integrating both data sources.

Table 5. Accuracy assessment results of model C3.

	Confusion Matrix						**Accuracy**	
	BU	WA	FO	VE	BA	RI	UA	PA
BU	162	0	0	0	5	2	0.6231	0.959
WA	0	604	0	0	0	0	1	1
FO	0	0	617	0	0	0	1	1
VE	16	0	0	592	0	4	1	0.967
BA	82	0	0	0	307	0	0.984	0.789
RI	0	0	0	0	0	557	0.9893	1
OA = 96.3% and Kappa = 0.955								

Table 6. Accuracy assessment results of model C4.

	Confusion Matrix						**Accuracy**	
	BU	WA	FO	VE	BA	RI	UA	PA
BU	157	0	0	1	9	2	0.6181	0.929
WA	0	604	0	0	0	0	1	1
FO	0	0	617	0	0	0	1	1
VE	6	0	0	566	5	35	0.9895	0.925
BA	91	0	0	5	293	0	0.9544	0.753
RI	0	0	0	0	0	557	0.9377	1
OA = 94.8% and Kappa = 0.936								

Among the four test scenarios, Case 3 (RF model with Sentinel-1 and Sentinel-2 combined) was the most optimal, achieving the highest overall accuracy (96.3%) and the highest Kappa coefficient (0.955). The findings highlight that using Sentinel-1 radar observations with Sentinel-2 multispectral imagery enhances land cover classification accuracy when using the RF model. This research underscores the usefulness for demonstrating the effectiveness of applying ensemble learning models and high-resolution satellite imagery in land cover classification in the Northern mountainous region of Vietnam. Although certain classification results have been achieved, limitations regarding artificial land cover classes will pose challenges when applying the approach to areas with complex urban cover.

3.2. Land cover map of study area

Based on the most accurate model (C3), a land cover map of Van Yen District was established (Figure 3). The results highlight that vegetation and forest areas dominate the district's land cover. Specifically, non-forest vegetation (VE) accounts for the largest proportion at 39.31% (59,522 ha), followed by forest (FO) at 28.79% (43,610.61 ha). Bare land (BA) covers a relatively large area of 23,301.03 ha, making up 15.38%, indicating land exploitation activities, shifting cultivation, or mountainous areas that have yet to be reforested. Paddy fields (RI) occupy 14,416.03 ha (9.51%), reflecting the significant role of agriculture in the local economic structure. Built-up (BU) has the smallest proportion, covering only 8,824.41 ha (5.82%). The development of certain residential areas may impact the natural ecosystem of the region. The spatial distribution map of land cover in Van Yen shows patches of bare land interspersed with natural forest areas and scattered built-up areas. The distribution of land cover types suggests that Van Yen District maintains a large proportion of forest and vegetation, but the considerable extent of bare land and cultivated

areas indicates ongoing land resource exploitation and agricultural-forestry development in the region. The land cover map built from the study is a valuable resource for local authorities in land-use planning, forest resource conservation, and sustainable development of residential areas.

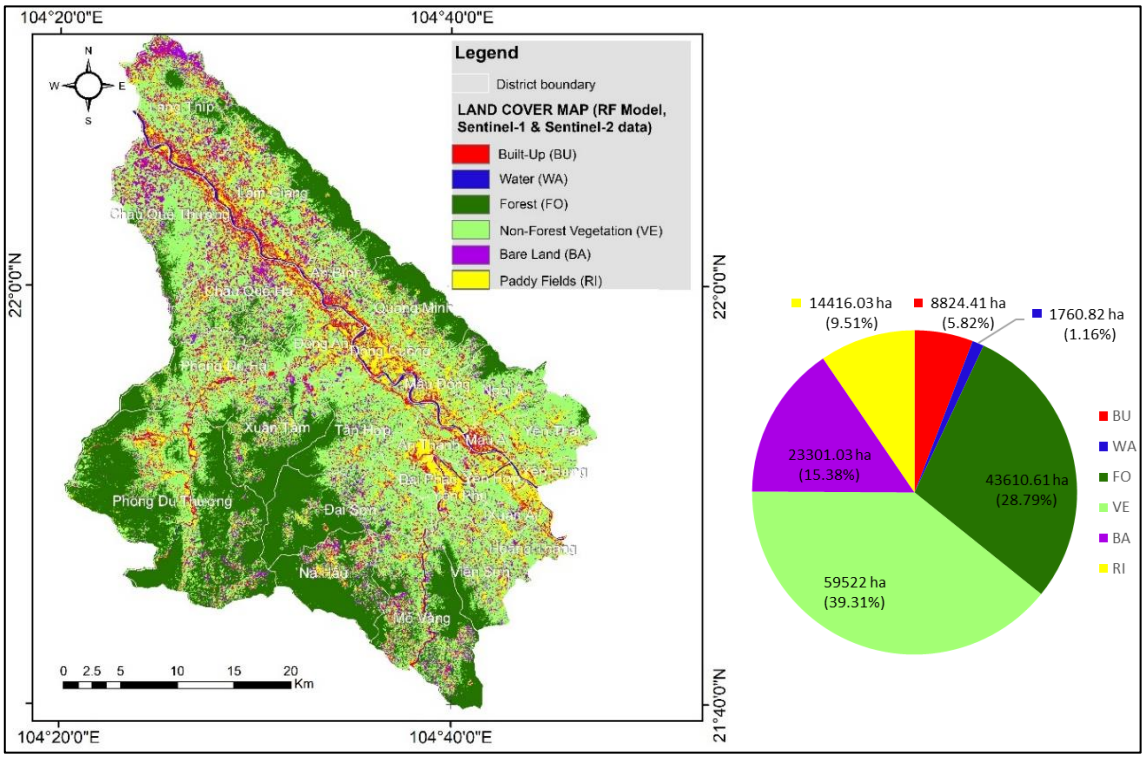

Figure 3. Land cover map of Van Yen district derived from the Random Forest model using Sentinel-1 and Sentinel-2 data.

4. Conclusion

This study successfully developed land cover classification models using Sentinel-1 radar and Sentinel-2 optical remote sensing data, employing the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models in Van Yen District, Yen Bai Province. Four model and data combinations were tested to evaluate their effectiveness. The results suggest that the most effective approach is the RF model, which integrates Sentinel-1 and Sentinel-2 datasets.

Both RF and XGBoost demonstrated high accuracy in classifying land cover from Sentinel remote sensing data, with overall accuracy (OA) ranging from 94.8% to 96.3% and Kappa coefficients between 0.936 and 0.955. However, RF exhibited greater stability and higher accuracy than XGBoost, particularly when integrating Sentinel-1 radar data with Sentinel-2 optical imagery. The bootstrap sampling and majority voting mechanism of RF enhances its noise resistance, which is especially useful when integrating radar data into the model. While XGBoost is renowned for its high performance and ability to detect complex patterns, it may be more vulnerable to noisy data. Specifically, the overall performance of the XGBoost model declined when radar data was added. The study highlights the potential of machine learning models and emphasizes the advantages of combining different remote sensing datasets to improve land cover classification accuracy.

Author contribution statement: Conceived and designed: D.L.N., D.K.T., V.A.T.; Analyzed and interpret the data: D.L.N., D.K.T.; Manuscript editing add: D.L.N., H.D.D.; Wrote the draft: D.K.T., V.A.T.; Reviewer and edit: D.L.N., D.K.T., V.A.T.

Acknowledgements: This research was funded by Hanoi University of Civil Engineering (HUCE) under grant number 04-2024/KHXD-TĐ.

Competing interest statement: The authors declare no conflict of interest.

References

1. Bai, X.; Sharma, R.C.; Tateishi, R.; Kondoh, A.; Wuliangha, B.; Tana, G. A detailed and high-resolution land use and land cover change analysis over the past 16 years in the Horqin sandy land, Inner Mongolia. *Math. Probl. Eng.* **2017**, *1*, 1316505.
2. Gibril, M.B.A.; Bakar, S.A.; Yao, K.; Idrees, M.O.; Pradhan, B. Fusion of RADARSAT-2 and multispectral optical remote sensing data for LULC extraction in a tropical agricultural area. *Geocarto Int.* **2017**, *32*, 735–748.
3. Shafizadeh-Moghadam, H.; Khazaei, M.; Alavipanah, S.K.; Weng, Q. Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors. *GIScience Remote Sens.* **2021**, *58*, 914–928.
4. Bui, D.Q.; et al. Landslide susceptibility prediction mapping with advanced ensemble models: Son La province, Vietnam. *Nat. Hazards* **2023**, *116*, 2283–2309.
5. Balha, A.; Mallick, J.; Pandey, S.; Gupta, S.; Singh, C.K. A comparative analysis of different pixel and object-based classification algorithms using multi-source high spatial resolution satellite data for LULC mapping. Earth Sci. *Informatics* **2021**, *14*, 2231–2247.
6. Saba, S.B.; Ali, M.; Turab, S.A.; Waseem, M.; Faisal, S. Comparison of pixel, sub-pixel and object-based image analysis techniques for co-seismic landslides detection in seismically active area in Lesser Himalaya, Pakistan. *Nat. Hazards* **2023**, *115*, 2383–2398.
7. Zaabar, N.; Niculescu, S.; Kamel, M.M. Application of Convolutional Neural Networks With Object-Based Image Analysis for Land Cover and Land Use Mapping in Coastal Areas: A Case Study in Ain Témouchent, Algeria. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5177–5189.
8. Adam, H.E.; Csaplovics, E.; Elhaja, M.E. A comparison of pixel-based and object-based approaches for land use land cover classification in semi-arid areas, Sudan. IOP Conf. Ser. *Earth Environ. Sci.* **2016**, *37*, 12061.
9. Tran, V.A.; Brovelli, M.A.; Ha, T.K.; Khuc, T.D.; Tran, N.D.; Tran, H.H.; Le, T.N. Land Subsidence Susceptibility Mapping in Ca Mau Province, Vietnam, Using Boosting Models. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 161.
10. Tran, V.A.; Khuc, T.D.; Ha, T.K.; Tran, H.H.; Le, T.N.; Pham, T.T.H; Nguyen, D.; Le, H.A.; Nguyen, Q.D. Land subsidence susceptibility mapping using machine learning in the Google Earth Engine platform. Intelligence of Things: Technologies and Applications, 2023, pp. 55–64.
11. Xu, S.; Zhao, Q.; Yin, K.; Zhang, F.; Liu, D.; Yang, G. Combining random forest and support vector machines for object-based rural-land-cover classification using high spatial resolution imagery. *J. Appl. Remote Sens*. **2019**, *13*, 14521.
12. Hoa, P.T.T.; Quang, V.N.; Nghi, L.T.; Phuong, D.T.N.; Hai, N.M. Research on the potential of applying Random Forest algorithm and Sentinel-2 to classify land cover in Quang Binh province on the Google Colab platform. *VN J. Hydrometeorol*. **2023**, *12*, 29–41.
13. Huong, N.T.T.; Trung, D.M.; Erkki, T.; Ronald, E.M. Land Use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods: A case study from Dak Nong, Vietnam. *Remote Sens.* **2020**, *12*, 1367.
14. Linh, D.Q.; Hai, L.T.; Khanh, D.Q. Applying machine learning in classifying and assessing LULC change in Hau Giang province. ISRM VietRock International Workshop. 2024.

15. Le, Q.T.; Dang, K.B.; Giang, T.L.; Tong, T.H.A; Nguyen, V.G.; Nguyen, T.D.L. Deep learning model development for detecting coffee tree changes based on Sentinel-2 imagery in Vietnam. *IEEE Access.* **2022**, *10*, 109097–109107.
16. Tran, X.B.; Vambol, V.; Luu, T.D. Assessing forest cover changes in Dak Lak province (Central Highlands of Vietnam) from multi-temporal Landsat data and machine learning techniques. *Ecol. Quest.* **2024**, *35*, 1–18.
17. Nguyen, T.P.; Nguyen, P.K.; Nguyen, H.N.; Tran, T.D.; Pham, G.T.; Le, T.H.; Le, D.H.; Nguyen, T.H.; Nguyen, V.B. Evaluation of statistical and machine learning models using satellite data to estimate aboveground biomass: A study in Vietnam tropical forests. *Forest Sci. Technol.* **2024**, *20*, 370–382.
18. Khuc, T.D.; Truong, X.Q.; Tran, V.A.; Bui, D.Q.; Ha, H.; Tran, T.H.M.; Pham, T.T.T.; Yordanov, V. Comparison of multi-criteria decision making, statistics, and machine learning models for landslide susceptibility mapping in Van Yen District, Yen Bai Province, Vietnam. *Int. J. Geoinformatics* **2023**, *19*, 33–45.
19. Tran, V.A.; Khuc, T.D.; Truong, X.Q.; Nguyen, A.B.; Phi, T.T. Application of potential machine learning models in landslide susceptibility assessment: A case study of Van Yen district, Yen Bai province, Vietnam. *Quat. Sci. Adv.* **2024**, *14*, 100181.
20. Arpitha, M.; Ahmed, S.A.; Harishnaika, N. Land use and land cover classification using machine learning algorithms in google earth engine. *Earth Sci. Inf.* **2023**, *16*, 3057–3073.
21. Schulz, D.; Yin, H.; Tischbein, B.; Verleysdonk, S.; Adamou, R.; Kumar, N. Land use mapping using Sentinel-1 and Sentinel-2 time series in a heterogeneous landscape in Niger, Sahel. ISPRS J. *Photogramm. Remote Sens.* **2021**, *178*, 97–111.
22. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land use land cover classification with U-net: Advantages of combining sentinel-1 and sentinel-2 imagery. *Remote Sens.* **2021**, *13*, 3600.
23. Viovy, N.; Arino, O.; Belward, A.S. The best index slope extraction (Bise): A method for reducing noise in NDVI time-series. *Int. J. Remote Sens.* **1992**, *13*, 1585–1590.
24. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594.
25. Nguyen, C.T.; Chidthaisong, A.; Diem, P.K.; Huo, L.Z. A modified bare soil index to identify bare land features during agricultural fallow-period in Southeast Asia using Landsat 8. *Land* **2021**, *10*, 231.
26. Noi, P.T.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 Imagery. *Sensors* **2018**, *18*, 18.
27. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611.
28. Vizzari, M. PlanetScope, Sentinel-2, and Sentinel-1 data integration for object-based land cover classification in Google Earth Engine. *Remote Sens.* **2022**, *14*, 2628.
29. Osgouei, P.E.; Kaya, S.; Sertel, E.; Alganci, U. Separating built-up areas from bare land in mediterranean cities using Sentinel-2A imagery. *Remote Sens.* **2019**, *11*, 345.
30. Zhou, Y.; Yang, G.; Wang, S.; Wang, L.; Wang, F.; Liu, X. A new index for mapping built-up and bare land areas from Landsat-8 OLI data. *Remote Sens. Lett.* **2014**, *5*, 862–871.