Applied Environmental Research
- RESEARCH FOR SUSTAINABLE PLANET -

Research Article

Applying Random Forest and Support Vector Machine for Land-use Classification in Phu Giao District, Vietnam

Trong Dieu Hien Le[1,*], Truong Vinh Linh[2]

[1] *Faculty of Resources & Environmental, University of Thu Dau Mot, 06 Tran Van On street, Thu Dau Mot City, Binh Duong, Vietnam*
[2] *Faculty of Information Technology, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao street, Go Vap District, Ho Chi Minh, Vietnam*

*Corresponding Email: dieuhien247@gmail.com

Abstract

Machine learning algorithms are currently widely used to classify satellite images to create surface maps of the earth. Support vector machines (SVMs) and random forests (RFs) are more effective ML algorithms and more accurate classifications than other methods. The aim of this study is to analyze the performance of these two algorithms in land-use and land cover classification. For this purpose, the Landsat 8 OLI satellite image freely provided by the United States Geological Survey (USGS) was used to classify land use in the Phu Giao District, Binh Duong Province, Vietnam, where forestland is being strongly converted into rubber plantations and cultivated land. The results revealed that the accuracies of the SVM model were 0.87 (overall accuracy) and 0.89 (Cohen's kappa), which are 2% lower than those of the optimal RF model. The land-use classification maps can be used as essential information in ecological and environmental management, such as natural habitats, urbanization and deforestation status, and species impact. Therefore, accurate and objective data/tools are extremely important.

ARTICLE HISTORY

Received: 25 Dec. 2024
Accepted: 14 May 2025
Published: 22 May 2025

KEYWORDS

Machine learning;
Random forest;
Support vector machine;
Landsat OLI 8

Introduction

Information about land cover (LC) is among the most important data for a variety of fields, such as environmental, ecological, and climate change studies, as well as resource management and monitoring [1,2]. The use of land cover and land use maps is the most effective method for recording and conveying information related to land use and land cover (LULC) data. LULC maps require consideration of many issues, including purpose, content, scale, input data types, and algorithms used. The land-use map can be divided into three categories on the basis of the extent of the area that it covers: local scale (covering a small area from 100-103 km^2), regional scale (104-106 km^2), and continent-to-global scale (>106 km^2) [3] or according to its spatial resolution: low (1 km), medium (1 km-100 m) and high resolution (<10 m) [4].'

In the past, traditional classification techniques, including supervised and unsupervised techniques, were widely applied to create land-use maps with different resolutions [1,5–6]. Nevertheless, because of their significant drawbacks, the use of these methods has begun to decrease, such as the Gaussian normal distribution, which rarely occurs in remote sensing data [7–8], and pixel clustering algorithms with similar spatial properties classified into a single class on the basis of some predefined criteria [9].

To overcome the constraints related to classification by traditional methods, new classification methods have been invented and utilized in recent years to obtain better classification performances [7–9]. Machine learning (ML) algorithms are among the most powerful and widely used classification algorithms [5,7,10–12].

The ML algorithms are nonparametric supervised techniques without any assumptions about the statistical distribution of entered datasets, in contrast to traditional methods [7–9]. There are numerous varieties of ML algorithms that have been rigorously applied in actual environments (nonresearch). Therefore, these methods, such as support vector machines (SVMs), random forests

https://doi.org/10.35762/AER.2025016

(RFs), and artificial neural networks (ANNs), are advanced. With respect to the overall performance of ML algorithms, many previous studies have confirmed that ML strategies yield greater accuracy outcomes than traditional parameter classification does, particularly for complex multivariate data [5, 7,13–16].

Although some studies have compared two widely adopted and powerful machine learning algorithms, RF and SVM, using different RS datasets for distinctive aims, the conclusions drawn have been inconsistent and contradictory. For example, [14,17–18] stated that SVM and RF obtained comparable classification accuracies, or, in other words, both were equally reliable, whereas [8, 12, 16,19–,20] claimed that SVM outperformed RF. In the assessment of each finding, the examination of previous studies [21–22] indicated that the RF was better than the SVM. Furthermore, these studies were carried out on a small, local scale with medium- or excessive-resolution satellite images as input data.

Research on land-use classification in Southeast Asia has evolved significantly over the past few decades, reflecting the region's diverse ecosystems, rapid urbanization, and agricultural expansion [23,24]. Early approaches relied heavily on field surveys, which were time-consuming and limited in scalability. By the late 20th century, the advent of satellite remote sensing, particularly Landsat imagery, enabled broader coverage and more systematic classification via supervised and unsupervised methods [25]. However, challenges such as cloud cover in tropical regions and mixed land-use patterns (e.g., agroforestry and fragmented urban areas) often reduce accuracy. More recently, advancements in machine learning, geographic information systems (GIS), and high-resolution datasets (e.g., Sentinel-2, LiDAR) have improved classification precision, incorporating spectral, temporal, and spatial dimensions [26–28]. Despite these innovations, inconsistencies persist owing to varying national classification systems, limited ground-truth data, and dynamic land-use changes driven by deforestation, plantations, and urban sprawl. Collaborative efforts between governments, researchers, and regional organizations (e.g., ASEAN) have sought to standardize methodologies, yet localized adaptations remain critical for capturing the complexity of Southeast Asia's landscapes.

Phu Giao, a district in Binh Duong Province, Vietnam, has a dynamic land-use structure characterized by a mix of agriculture (rubber, fruit orchards, and crop farms), forestry, and emerging industrial and residential zones. Rapid urbanization and industrialization, driven by Binh Duong's strategic location near Ho Chi Minh City and its economic growth policies, are key forces transforming land use. These changes are further broadened by infrastructure development, population growth, and investment in industrial parks. Every year, local authorities conduct field surveys to inventory the types of land use in the area. No studies have yet been conducted to classify land in the Phu Giao District via remote sensing and machine learning meth-ods. The main objective of this study is to evaluate the performance of these powerful machine learning algorithms in LULC classification. For this purpose, RF and SVM algorithms have been selected and applied to Landsat 8 OLI satellite image solutions to classify land use in the Phu Giao District, Binh Duong Province, Vietnam. The best model of each classification model (RF and SVM) was selected by testing the models generated throughout various values of the maximum influential parameters of every set of rules, which can be the wide variety of trees (ntree) and the wide variety of variables (mtry) in the RF and the cost function (cost) and gamma within the SVM.

Materials and methods

1) Study area

Phu Giao is a district located in northeast Binh Duong Province, approximately 70 km from Ho Chi Minh City. The east is Vinh Cuu District (Dong Nai), the west is Ben Cat District (Binh Duong), the south is Bac Tan Uyen District (Binh Duong), and the north is Dong Phu District (Binh Phuoc) (Figure 1). In equable climate conditions, including temperatures ranging from 26°C to 31°C, the average annual rainfall in the district is 1,947.7 mm, with an average of 163 rainy days per year and high humidity. The Phu Giao District is suitable for the planting of high-value industrial crops such as rubber, cashew, pepper, etc. Forest and agricultural land have been immediately decreasing due to the process of urbanization, which may cause some difficulties in land-use management [29].

2) Data and sample

2.1) Remote sensing data

Landsat 8 OLI remote sensing data are freely downloaded from the United States Geological Survey (USGS) website [30]. The eight spectral bands (from 1 to 7 and 9) of the Landsat 8 OLI image range from visible wave to shortwave infrared (SWIR) with a resolution of 30 m. Band 8 has a panchromatic wavelength with a resolution of 15 m. In this study, Landsat 8 OLI images with bands (2, 3, 4, 5, 7, 9, 10, and 11) taken on June 30, 2023, were used. This image is not affected by cloud cover.

Band combinations in remote sensing refer to the selection and merging of different spectral bands from satellite or aerial imagery to enhance the visibility of specific LULC features. Different band combinations help distinguish between vegetation, water, urban areas, and other land cover types. Examples include NIR-Red–Green (5-4-3) for vegetation, SWIR-NIR-Red (7-5-4) for urban areas, and NIR-SWIR-Green (5-6-3) for water. However, in this study, we used bands (2-7, 9, and 10) of multitemporal satellite imagery Landsat 8 as predictors in the models. We then compared the predictive power of the RF and SVM models for classification.

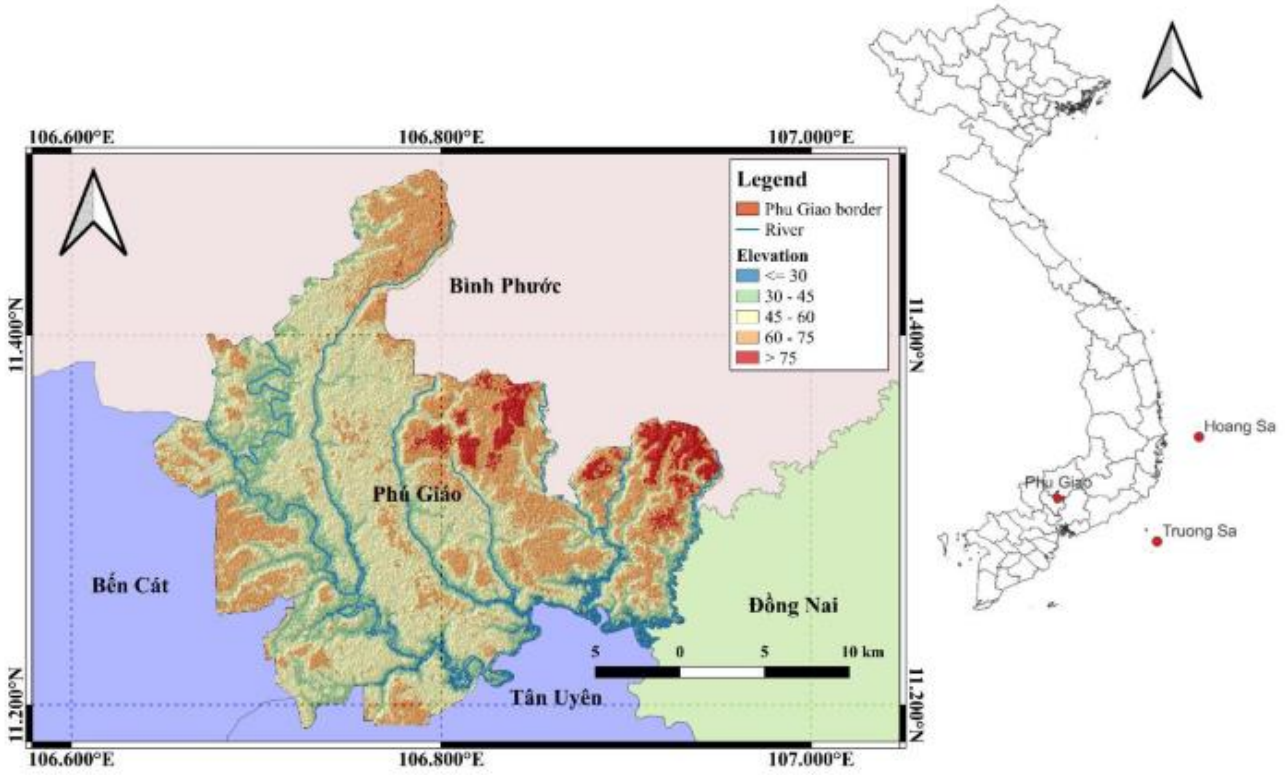


Figure 1 Satellite map of Phu Giao District.

Table 1 Landsat 8 OLI images used in this study

Image bands	Wavelength (micrometer)	Resolution (meter)	Date
Band 2 – Blue	0.45 – 0.51	30	30th June 2023
Band 3 – Green	0.53 - 0.59	30	(Path=125, Row=052)
Band 4 – Red	0.64 – 0.67	30	
Band 5 – Near infrared	0.85 – 0.88	30	
Band 6 – Infrared 1 (SWIR 1)	1.57 – 1.65	30	
Range 7 – Infrared 2 (SWIR 1)	2.11 – 2.29	30	
Band 9 – Cirrus	1.36 – 1.38	30	
Band 10 – Thermal Infrared 1	10.60 – 11.19	100	
Band 11 – Thermal Infrared 2	11.50 – 12.51	100	

2.2) Training data and test data

In this study, the six types of land use in the Phu Giao District are listed in Table 2. The training and testing data are based on field point samples and the local land-use plan map from 2022. This method has been widely applied and published in previous studies [6, 31–32]. The data used for training (70%) and testing (30%) for the classification models, including 655 samples and the number for each land-use type, are provided in Table 2.

3) Algorithms

3.1) Random forest (RF)

RF is a machine learning method introduced by Breiman [31] that allows for improved accuracy prediction and classification without overfitting data. RF is based on classification and regression trees (CARTs). The training and testing data are described in Table 2. We used repeated cross-evaluations to divide the training data into 10 datasets (k-fold=10) and 3 repetitions to determine the optimal parameters, including the number of randomly selected variables (mtry) and the number of trees (ntree) at each node.

Table 2 Number of samples used in training and testing for each soil classification

Type of land-use	Total number of samples	Training sample	Testing sample
Forestland	93	65	28
Urban land	125	87	38
Rubber land	115	80	35
Water surface	120	84	36
Bareland	100	70	30
Agricultural land	102	72	30
Total	655	458	197

In detail, the repeated cross-validation technique consists of randomly dividing the initial reference data into 10 groups, with group 1st as the validation group and groups 2nd-10th as the training group. The same procedure was repeated 3 times each time, with distinct groups used as the validation group and the rest of the groups used as the training group. Compared with other machine learning methods, RF models are insensitive to overfitting, can model nonlinear relationships, and have high predictive performance. Numerous previous studies have implemented RF classification to map land cover with high precision. In their initial assessment of the RF's performance for land cover classification in Spain, Rodriguez-Galiano et al. [33] obtained a higher overall accuracy of 91% and a kappa coefficient of 0.92. Using an RF classifier, Hayes et al. [34] created a high-resolution (1 m) land cover map of Wyoming, achieving a kappa coefficient of 0.79 and an overall accuracy of 81%. The RF classification method has also been used in recent studies to generate flood inundation maps [35–36].

3.2) Support vector machine (SVM) model

The concept of the SVM was introduced in 1979 by Vapnik [37]. SVM is a powerful tool for multidimensional, linear, and nonlinear data. SVMs are typically chosen for classification and regression analysis because of their high degree of customization. However, the critical parameter search procedure is more complicated than that of the RF. SVM is based on the principle of performing a "hyperplane" that divides layers into two separate parts [38–39]. In reference to kernel functions, there are numerous varieties, including sigmoid, linear, polynomial, and radial basis functions (RBFs). The most effective and often utilized method for image classification via remote sensing is the radial basis function [40]. Therefore, two parameters (Cost-C) and gamma ({γ), which are essential for regulating the SVM's performance in the radial basis function, were used in this study. SVMs are capable of classifying data linearly or nonlinearly. Nonlinear data are handled by the kernel function. Compared with previous studies [13, 41–42]. Relatively accurate classification can be achieved with SVM. Foody confirmed that accurate classifications may be obtained via a single multiclassification SVM [13]. They evaluated several classification techniques, such as SVM, feed-forward neural networks, decision trees, and discriminant analysis, and discovered that SVM had the highest accuracy. These results are consistent with those of a study by Shi and Yang [42], which indicated that the SVM outperforms the MLC in terms of quantitative accuracy. Additionally, Candale and Dixon [41] reported that the SVM outperforms other classification methods. Figure 2 shows a summary of the land cover classification results of the RF and SVM methods. All calculations and imaging in the study were performed in R version 4.1.2 [43].

4) Model variation

The best RF and VSM models are determined with test data during the model's internal accuracy validation. The factors that must be applied to evaluate the accuracy of the classification model include the error matrix, overall accuracy (OA), manufacturer accuracy (PA), user accuracy (UA) and the kappa coefficient [44]. The confusion matrix quantifies the similarity be-tween the model-classified samples and the reference data. The OA is the number of samples that are efficiently labeled, which is consistent with the pattern that enters the category. However, the OA tends to overestimate the overall performance. The kappa coefficient is used to assess the overall performance of the model. The PA is the percentage of efficiently labeled samples of a particular category. Moreover, the UA is the percentage of efficiently labeled samples, and the entire wide variety of samples are labeled into that category [45].

For algorithm comparison, we also applied the OA, UA, PA and kappa coefficient to determine the classification method that achieves better accuracy [11, 40–46]. Figure 2 shows the process of land-use classification applied in this study.

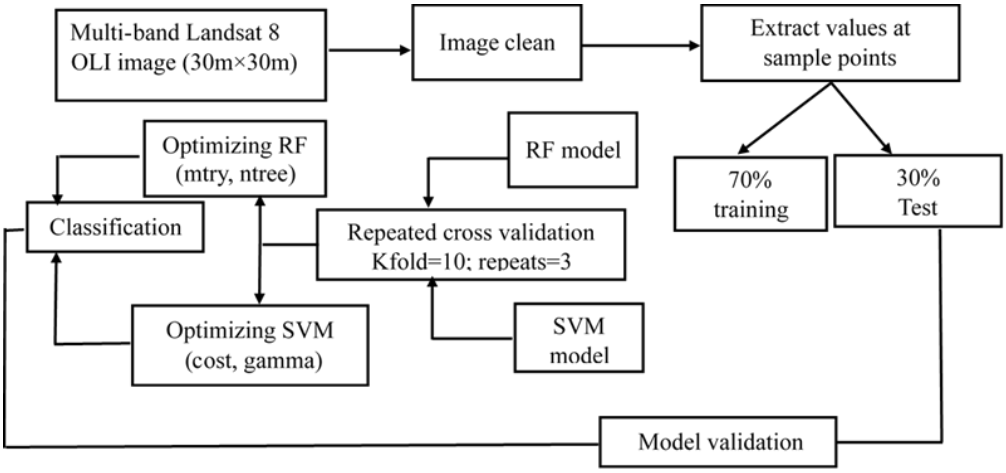


Figure 2 Diagram of the classification process of the land-use layer.

Results

1) Engineering features

Figure 3 (a-d) shows similar trends for all classification classes in bands 2-4. In detail, the value in bands 2-4 for barren land is the highest (ranging from 0.05-0.5), followed by urban (0.03-0.16), agricultural (0.02--0.15), water bodies (0.02-0.1), rubber (0.02-0.08) and forestland (0.02-0.1). The values of brands 5-9 are the lowest for the water surface, whereas the values of 10 and 11 are the lowest for barren land.

2) Adjustment parameter

In repeated cross-evaluation with Kfold = 10 and number of repetitions of 3, the initial reference sample is randomly divided into 10 groups to build each decision tree during the training process. The training samples are predicted from the model to evaluate the classification accuracy, and the out of bag error (OBB) is the lowest. The number of decision trees that need to be optimally constructed (ntree = 1,000) and the number of variables at each decision tree node split (mtry = 6) create the highest accuracy in the RF model (Figure 4). On the other hand, the parameters determining the optimal SVM model include Cost = 10 and gamma = 1.

3) Classification results

Table 3 shows the overall accuracy and kappa value generated by the land-use type classification via the RF and SVM models. In addition, errors are also presented for all classified layers. In general, the classification accuracy of both models is approximately 90%. The optimal SVM algorithm achieved overall accuracy (OA = 0.89), and the kappa value (0.87) was 2% lower than that of the RF algorithm (Table 3). Therefore, the two models create slightly different land cover maps (Figure 5). Furthermore, the map generated by the SVM seemed less accurate because of the misclassification between layers, a common phenomenon in interpreting satellite images, which is represented by the confusion matrix below. The reason may be that the multiclassifier combination process in the RF algorithm plays a crucial role. The combination of multiclass classification for decision making can be more accurate than single-class classification. The SVM model does not have this mechanism; as a result, the RF can solve the problem of multilayer classification.

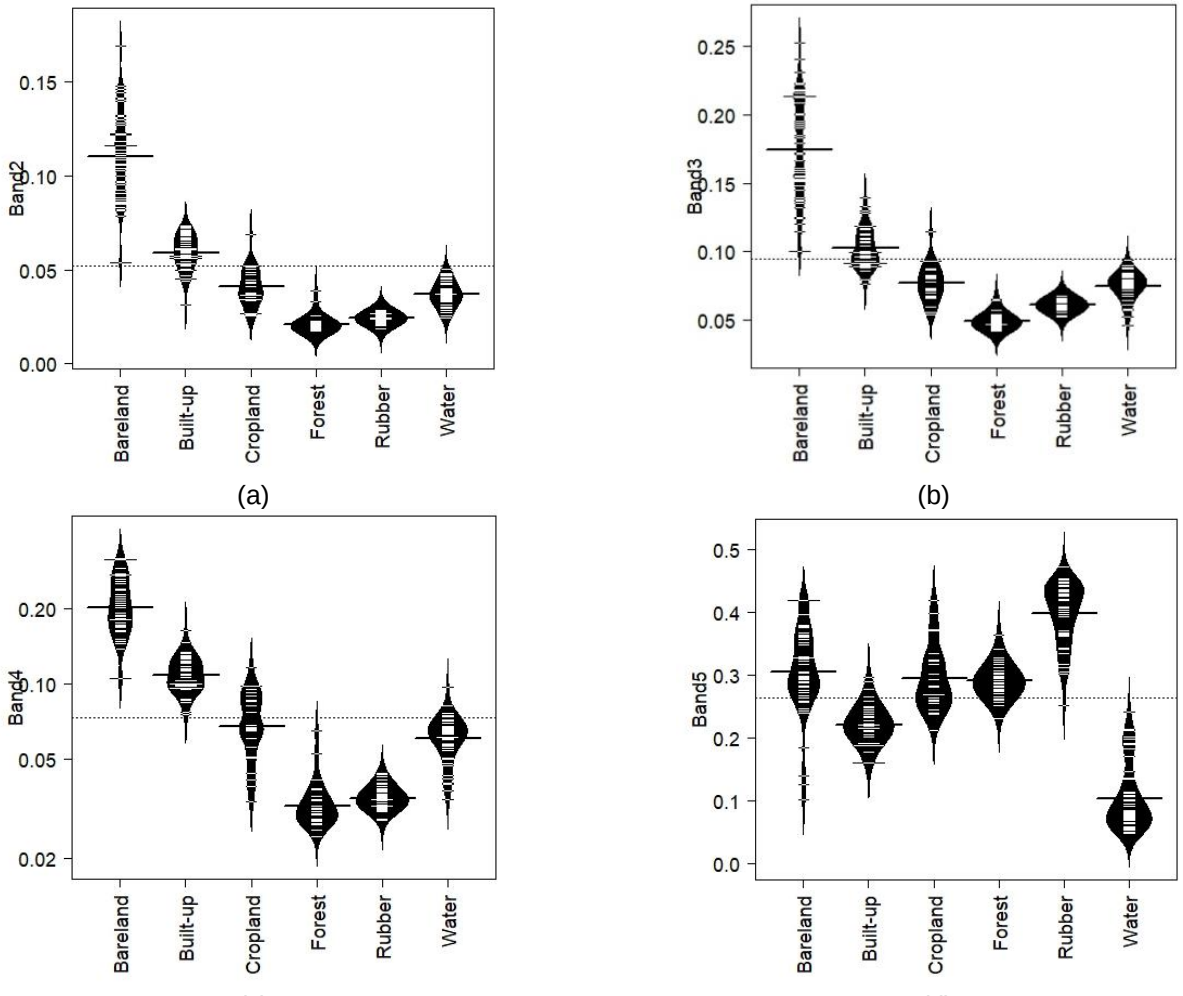
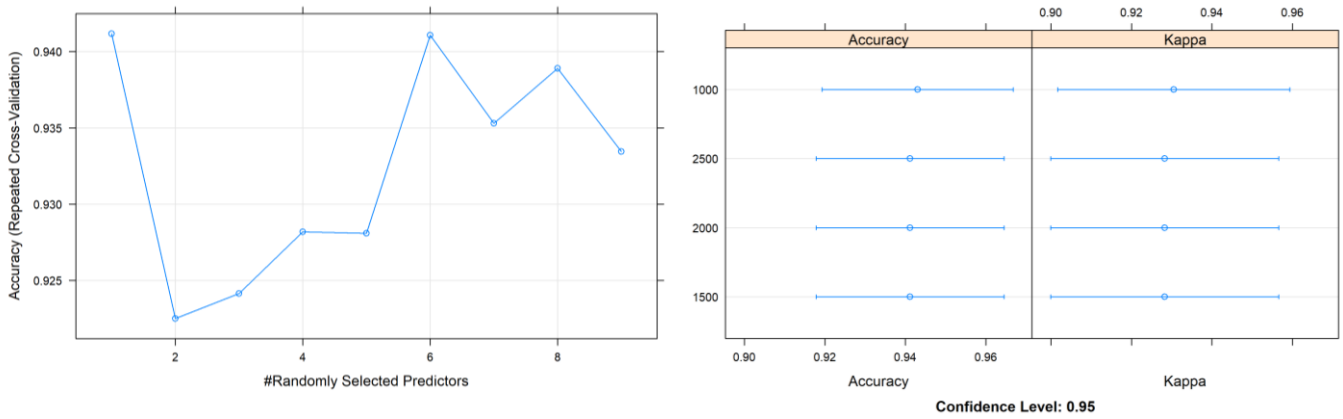
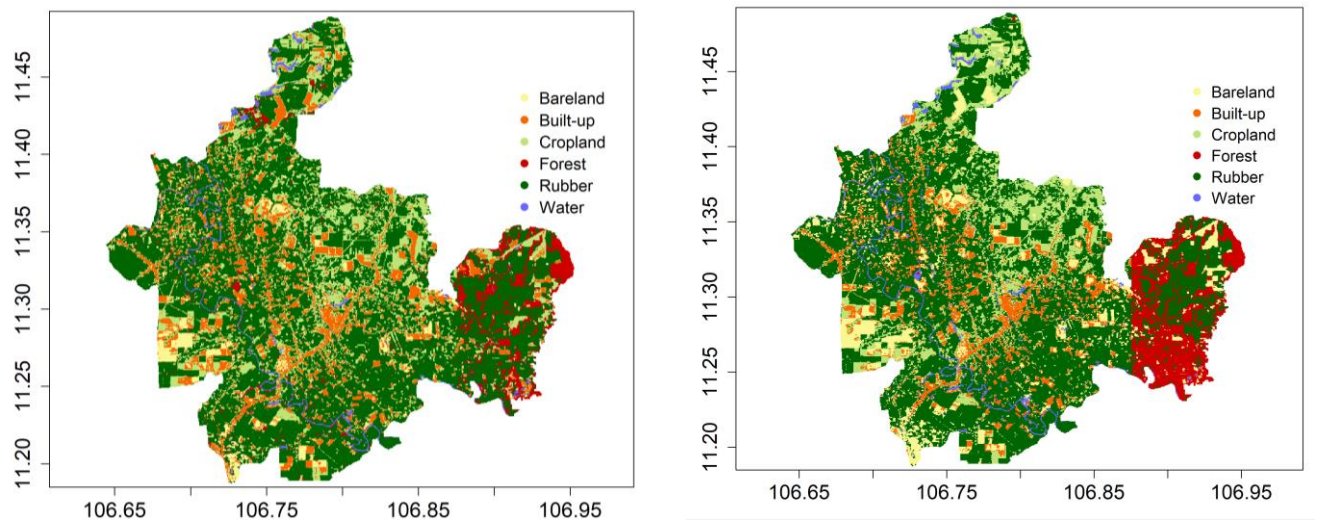


Figure 3 Variation of waveband in land-use types with band 2 (a), band 3 (b), band 4 (c), and band 5 (d) (see other band in the Supplementary Figure 1).



(a) (b)

Figure 4 Number of variables at each node (mtry) (a) number of trees and (b) variables randomly divided at each node of RF (ntree).



(a) (b)

Figure 5 Land cover generated by the best RF model (a) and the best SVM model (b).

Table 3 Number of samples used in training and testing for each soil classification

Confusion matrix in the RF model								
	Vacant	**Urban**	**Agricultural**	**Forest**	**Rubber**	**Water**	**Total**	**UA**
Vacant	27	2	1	0	0	0	30	90
Urban	2	36	3	0	0	0	41	87.8
Agricultural	1	0	25	1	0	1	28	89.3
Forest	0	0	1	24	1	0	26	92.3
Rubber	0	0	0	2	34	0	36	94.4
Water	0	0	0	1	0	35	36	97.2
Total	30	38	30	28	35	36	197	NA
PA	90	94.7	87.1	85.7	100	94.4	NA	91.9
Kappa	0.89							
Confusion matrix in the SVM model								
	Vacant	**Urban**	**Agricultural**	**Forest**	**Rubber**	**Water**	**Total**	**UA**
Vacant	26	3	1	0	0	0	30	86.7
Urban	4	33	1	0	0	0	38	86.8
Agricultural	0	2	27	1	0	1	31	87.1
Forest	0	0	1	23	1	2	27	85.2
Rubber	0	0	0	3	34	0	37	91.9
Water	0	0	0	1	0	33	34	97.1
Total	30	38	30	28	35	36	197	NA
PA	86.7	86.8	90	82.1	97.1	91.7	NA	89.3
Kappa	0.87							

Furthermore, Du et al. [47] applied SVM and RF algorithms to divide airborne visible/infrared imaging spectrometer (AVIRIS) images into nine land-use groups. The results show that the general classification accuracy and kappa coefficient of the RF and SVM algorithms are 95.1%, 94.3%, 0.94, and 0.93, respectively. Li et al. [48] used the Landsat thematic mapper dataset in Guangzhou city, China, to test 13 supervised and two unsupervised classification techniques, including the SVM and RF algorithms. With an accuracy of 0.917 for RF and an SVM of 0.891, the final results demonstrate that the RF algorithm has higher classification accuracy. On the basis of Lidar data, Qin et al. [49] classified the land cover of Zhangye city via the maximum likelihood classification (MLC), SVM, and RF algorithms. The results revealed that the RF model yielded the highest accuracy, with an overall classification accuracy of 91.82% and a kappa coefficient of 0.88, whereas those of the SVM model were 88.48% and 0.83, respectively. Compared with other machine learning methods, RF models are insensitive to overfitting, can model nonlinear relationships, and have high predictive performance. RF classification has been used in many previous studies because of its high accuracy in land cover mapping. Faeehem et al. [50] applied RF with an overall accuracy of 0.99 to determine LCLU changes over the past three decades (1990–2020) in an arid ecosystem in Pakistan. Svoboda et al. (2022) focused on the development of Sentinel-2 data and RF classification according to the LULUCF requirements on the cloud-based platform Google Earth Engine (GEE). The results obtained an accuracy classification with an overall accuracy = 89.1% and Cohen's kappa = 0.84. Kasahun and Legesse [51] compared three machine learning algorithms—ANN, SVM, and RF—for LULC classification via Google Earth images from the years 2006, 2014, and 2022 in Dilla town, Ethiopia. The results showed that the RF algorithm outperformed both the SVM and ANN algorithms, with an average OA of 0.97, a kappa of 0.98, a PA of 0.99, and a UA of 0.97. Rodriguez-Galiano et al. [21] initially evaluated the performance of the RF for classifying land cover in Spain and achieved a higher overall accuracy of 91% and a kappa coefficient of 0.92. Hayes et al. [22] prepared a high-resolution (1 m) land cover map in Wyoming via an RF classifier and obtained an overall accuracy of 81% and a kappa coefficient of 0.79. The RF classification method has also been used in recent studies to generate flood inundation maps [23–24].

Through the previous comparison, we determined that with the improvement in the resolution of the image data, the accuracy of RF and SVM classification also gradually improved. Under the same image resolution conditions, the accuracy of the RF model is better than that of the SVM model. There are other factors that affect the classification accuracy, such as the selection of the right parameters or the number of samples for each class in the training data. Some environmental and geographical variables, such as topographic features, elevation, slope, texture, the vegetable index, and soil properties, were omitted from our models [52–56]; therefore, our classification for this area may be biased. We chose the RF model with more accurate classification to classify land-use types for Phu Giao in 2023. The results of the study revealed that the total area of land-use types in Phu Giao is 536.74 km^2, of which the forest cover area is 11.5%, which is equivalent to 61.52 km^{2,} and the area of rubber plantations is 55.5%, which is equivalent to 297.69 km^2. In the LULC group without forest, cropland occupied the largest area (16.7% of the total area), equivalent to 89.61 km^2 (Figure 6).

4) Model limitations

Owing to the limited sample numbers for training, the sampling sites were spatially unevenly distributed, and some areas of Phu Giao lacked sampling sites, leading to an inadequate representation in the study region [57], particularly when classifying multitemporal satellite imagery from dates mismatched with field surveys. In agricultural land, for example, crop development throughout the year can alter spectral signatures [58]. Second, other input variables, such as the normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and normalized difference built-up index (NDBI), were omitted from our model; this issue is likely to be a general issue in statistical models.

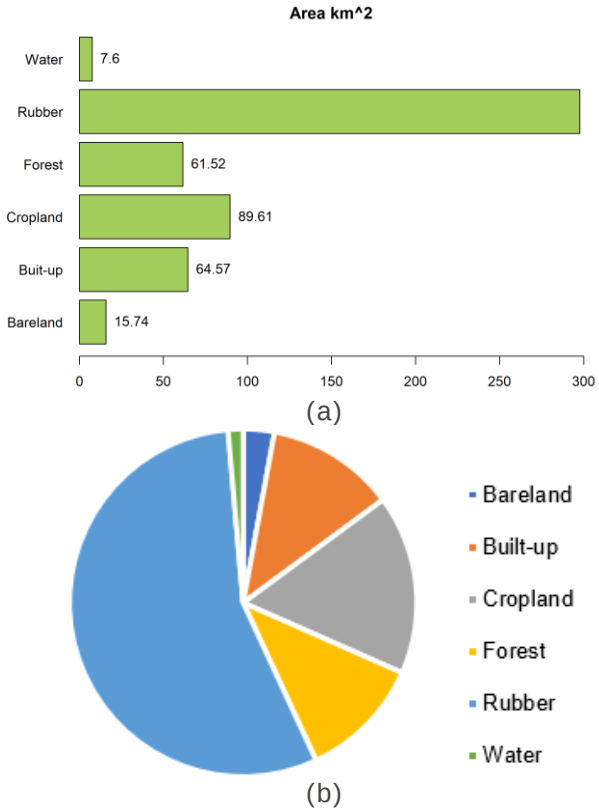


Figure 6 Area (a) and percentage (b) of land-use types in the Phu Giao district in 2023.

Conclusions

RF and SVM were applied to classify land use in the Phu Giao district, Binh Duong Province, Vietnam. The results of the study indicate that the SVM model has 2% lower accuracy than the RF model for mapping the studied area. In addition, the values of the image brands are necessary to add other variables, such as terrain (elevation and slope), the NDVI, NDWI, NDBI, etc. These factors can increase the performance of the models. Further studies should consider the use of these variables to investigate improvements in model performance.

Credit author statement

T.D.H.L contribution to data collection, data extraction; Analyze and interpret data, draft articles, and revise manuscript. T.V.L proposes and edits a manuscript. All authors have final approval for publication.

Conflict of interest

The authors have no conflict of interest with any individual or organization related to the research paper.

Acknowledgement

This research is funded by Thu Dau Mot University, Binh Duong Province, Vietnam under grant number DT.22.3-038. The research team would like to thank Thu Dau Mot University for funding this research.

References

[1] Tchuenté, A.T.K., Roujean, J.L., Jong, S.M.D. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. International Journal of Applied Earth Observation and Geoinformation, 2011, 13(2), 207–219.

[2] Latifovic, R., Olthof, I. Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. Remote Sensing of Environment, 2004, 90, 153–165.

[3] Cihlar, J. Land cover mapping of large areas from satellites: Status and research priorities. International Journal of Remote Sensing 21, 2000, 1093–1114.

[4] Feng, M., Li, X. Land cover mapping toward finer scales. Science Bulletin, 2020,65, 1604–1606.

[5] Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley. A., Huang. X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment, 2010, 114, 168–182.

[6] Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R. 2000 Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing, 2000, 21,

[7] Ghimire, B., Rogan, J., Miller, J. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. Remote Sensing Letters, 2010. 1, 45–54.

[8] Mountrakis, G., Im, J., Ogole, C. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66, 2011, 247–259.

[9] Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T. Land cover 2.0. International Journal of Remote Sensing, 2018, 39, 4254–4284.

[10] Otukei, J.R., Blaschke, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. International Journal of Applied Earth Observation and Geoinformation, 2010, 12, S27–S31.

[11] Waske, B., van der Linden, S., Benediktsson, J.A., Rabe, A., Hostert, P. Sensitivity of support vector machines to random feature selection in classification of Hyperspectral Data. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48, 2880–2889.

[12] Khatami, R., Mountrakis, G., Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. Remote Sensing of Environment, 2016, 177, 89–100.

[13] Foody, G.M., Mathur, A. A relative evaluation of multiclass image classification by support vector machines. IEEE Transaction on Geoscience and Remote Sensing, 2004 42, 1335–1343.

[14] Pal, M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 2005, 26, 217–222.

[15] Pal, M., Mather, P.M. Support vector machines for classification in remote sensing. International Journal of Remote Sensing, 2005, 26, 1007–1011.

[16] Maxwell, A.E., Warner, T.A., Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 2018, 39, 2784–2817.

[17] Dalponte, M., Ørka, H.O., Gobakken, T., Gianelle, D., Næsset, E. Tree Species Classification in Boreal Forests With Hyperspectral Data. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51, 2632–2645.

[18] Adam, E., Mutanga, O., Odindi, J., Abdel-Rahman, E.M. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers.

International Journal of Remote Sensing, 2014, 35, 3440–3458.

[19] Zhang, C., Xie, Z. 2013 Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques. Wetlands, 2013, 33, 233–244.

[20] 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.

[21] Shang, X., Chisholm, L.A. Classification of Australian native forest species using hyper-spectral remote sensing and machine-learning classification algorithms. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7, 2481–2489.

[22] Lawrence, R.L, Moran, C.J. The AmericaView classification methods accuracy comparison project: A rigorous approach for model selection. Remote Sensing of Environment, 2015, 170, 115 –120.

[23] Waiyasusri, K., Chotpantarat, S. Spatial evolution of coastal tourist city using the Dyna-CLUE Model in Koh Chang of Thailand during 1990–2050. ISPRS International Journal of Geo-Information, 2022, 11, 49.

[24] Waiyasusri, K., Vangpaisal, R., Chotpantarat, S. 2024 Climate and land use change impacts on groundwater recharge in Prachinburi–Sakaeo groundwater basin by integrating the CA–Markov model with the WetSpass Model. Earth Systems and Environment, 2024, 8, 1179–1206.

[25] Halder, A., Ghosh, A., Ghosh, S. Supervised and unsupervised land use map generation from remotely sensed images using ant based systems. Applied Soft Computing, 2011, 11, 5770–5781.

[26] Chaves, M.E.D., de Alves, M.C., Sáfadi, T., de Oliveira, M.S., Picoli, M.C.A., Simoes, R.E.O., Mataveli, G.A.V. Time-weighted dynamic time warping analysis for mapping interannual cropping practices changes in large-scale agro-industrial farms in Brazilian Cerrado. Science of Remote Sensing, 2021, 3, 100021.

[27] Li, Z., Chen, B., Wu, S., Su, M., Chen, J.M., Xu, B. Deep learning for urban land use category classification: A review and experimental assess-ment. Remote Sensing of Environment, 2024, 311, 114290.

[28] Fayaz, M., Nam, J., Dang, L.M., Song, H.K., Moon, H. Land-cover classification using deep learning with high-resolution remote-sensing imagery. Applied Sciences, 2024, 14, 1844.

[29] Cục Thống kê Bình Dương. 2024 Binh Duong Statistical yearbook 2023. Hà Nội Bình Dương : Thanh niên : Cục thống kê tỉnh Bình Dương, 2024.

[30] United States Geological Survey. EarthExplorer. USGS EasthExplorer's website, 2025. [Online] Available from: https://earthexplorer.usgs.gov/ [Accessed 4 April 2025].

[31] Breiman, L. Random forests. Machine Learning, 2001, 45, 5–32.

[32] Rabiet, M., Margoum, C., Gouy, V., Carluer, N., Coquery, M. Assessing pesticide concentrations and fluxes in the stream of a small vineyard catchment – Effect of sampling frequency. Environmental Pollution, 2010, 158, 737–748.

[33] Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sensing of Environment, 2012, 121, 93–107.

[34] Hayes, T., Usami, S., Jacobucci, R., McArdle, J.J. Using classification and regression trees (CART) and random forests to analyze attrition: Results from two simulations. Psychology and Aging, 2015, 30, 911–929.

[35] Ismanto, R.D., Fitriana, H.L., Manalu, J., Purboyo, A.A., Prasasti, I. Development of flood-hazard-mapping model using random forest and frequency ratio in Sumedang Regency, West Java, Indonesia. Geomatics and Environmental Engineering, 2023, 17, 129–157.

[36] Esfandiari, M., Jabari, S., McGrath, H., Coleman, D. Flood mapping using random forest and identifying the essential conditioning factors; A case study in Fredericton, New Brunswick, Canada. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3–2020, 609–615.

[37] Vapnik, V. 1999 The nature of statistical learning theory. Springer Science & Business Media.

[38] Melgani, F., Bruzzone, L. Classification of hyper-spectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42, 1778–1790.

[39] Qian, Y., Zhou, W., Yan, J., Li, W., Han, L. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. Remote Sensing, 2015, 7, 153–168.

[40] Yang, X. Parameterizing support vector machines for land cover classification. Photogrammetric Engineering & Remote Sensing, 2011, 77, 27–37.

[41] Candade, N., Dixon, D.B. 2004 Multispectral classification of Landsat images: A comparison of support vector machine and neural network classifiers. Proceedings of ASPRS Annual Meeting, 28 May 2004.

[42] Shi, D., Yang, X. Support vector machines for land cover mapping from remote sensor imagery.

In: Li, J., Yang, X. Eds., Monitoring and modeling of global changes: A geomatics perspective, Dordrecht, 2015, 265–279

[43] R Core Team. R: The R Project for statistical computing. Foundation for Statistical Computing, Vienna, Austria. 2021. [Online] Available from: https://www.r-project.org/ [accessed 1 December 2021].

[44] Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 1991, 37, 35–46.

[45] Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment, 2013, 129, 122–131.

[46] Rodriguez-Galiano, V.F., Chica-Rivas, M. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and digital terrain models. International Journal of Digital Earth, 2014, 7, 492–509.

[47] Du, P., Xia, J., Chanussot, J., He, X. Hyperspectral remote sensing image classification based on the integration of support vector machine and random forest. IEEE International Geoscience and Remote Sensing Symposium, 2012, 174–177.

[48] Li, X., Chen, W., Cheng, X., Wang, L. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. Remote Sensing, 2016, 8, 514.

[49] Qin, Y., Li, S., Vu, T.T., Niu, Z., Ban, Y. Synergistic application of geometric and radiometric features of LiDAR data for urban land cover mapping. Optics Express, 2015, OE 23, 13761–13775.

[50] Faheem, Z., Kazmi, J.H., Shaikh, S., Arshad, S., Noreena, Mohammed, S. Random forest-based analysis of land cover/land use LCLU dynamics associated with meteorological droughts in the desert ecosystem of Pakistan. Ecological Indicators, 2024, 159, 111670.

[51] Kasahun, M., Legesse, A. Machine learning for urban land use/ cover mapping: Comparison of artificial neural network, random forest and support vector machine, a case study of Dilla town. Heliyon, 2024, 10, e39146

[52] Cai, H., Zhang, S., Bu, K., Yang, J., Chang, L. Integrating geographical data and phenological characteristics derived from MODIS data for improving land cover mapping. Journal of Geographical Sciences, 2011, 21, 705–718.

[53] Fichera, C.R., Modica, G., Pollino, M. Land cover classification and change-detection analysis using multi-temporal remote sensed imagery and landscape metrics. European Journal of Remote Sensing, 2012, 45, 1–18.

[54] Xiao, H., Tang, Y., Li, H., Zhang, L., Ngo-Duc, T., Chen, D., Tang, Q. Saltwater intrusion into groundwater systems in the Mekong Delta and links to global change. Advances in Climate Change Research, 2021, 12(3), 342–352.

[55] Jin, Y., Liu, X., Chen, Y., Liang, X. Land-cover mapping using random forest classification and incorporating NDVI time-series and texture: A case study of central Shandong. International Journal of Remote Sensing, 2018, 39, 8703–8723.

[56] Zeferino, L.B., de Souza, L.F.T., do Amaral, C.H., Elpidio, I,F.F., de Oliveira, T.S. Does environmental data increase the accuracy of land use and land cover classification? International Journal of Applied Earth Observation and Geoinformation, 2020, 91, 102128.

[57] Tuia, D., Pasolli, E., Emery, W.J. Using active learning to adapt remote sensing image classifiers. Remote Sensing of Environment, 2011, 115, 2232–2242.

[58] Stehman, S.V., Foody, G.M. Key issues in rigorous accuracy assessment of land cover products. Remote Sensing of Environment, 2019, 231, 111199.