Geografia-Malaysian Journal of Society and Space 22 issue 2 (74-90)
 https://doi.org/10.17576/geo-2026-2202-05

Comparison of Random Forest and Support Vector Machine algorithms in urban land use and land cover classification

Nur Hakimah Asnawi, Lam Kuok Choy, Rosniza Aznie Che Rose

Geography Program, Centre for Development, Social and Environment Studies,
Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia

Correspondence: Lam Kuok Choy (email: lam@ukm.edu.my)

Received: 16 January 2026; Accepted: 29 April 2026; Published: 21 May 2026

Abstract

A precise and up-to-date land use and land cover (LULC) map is crucial for sustainable development planning and monitoring environmental change. Various machine learning (ML) algorithms are widely used to classify remote sensing data for mapping the Earth's surface, particularly Random Forest (RF) and Support Vector Machine (SVM). Although numerous comparative studies evaluating these two algorithms exist, their results remain inconsistent and vary across geographic regions and datasets. This study aims to evaluate the accuracy of ML algorithms of Random Forest (RF) and Support Vector Machine (SVM) for LULC classification in the Seremban and Port Dickson districts, Negeri Sembilan, Malaysia. LULC classification was conducted using the Google Earth Engine (GEE) platform, utilizing Landsat 5 TM imagery for the year 2010 and Landsat 8 OLI imagery for the year 2020. The findings revealed that the RF classifier outperformed the SVM classifier, achieving higher overall accuracy of 0.918 (2010) and 0.890 (2020) compared to 0.780 (2010) and 0.842 (2020) for SVM. RF also demonstrated a higher Kappa coefficient (0.8589 in 2010; 0.8093 in 2020) than SVM (0.6384 in 2010; 0.7318 in 2020). Based on the accuracy metrics, forest, agriculture and built-up areas exhibited lower classification accuracy due to mixed and complex LULC patterns, whereas water bodies were classified with high accuracy by both classifiers due to their distinct spectral characteristics. Overall, the RF algorithm demonstrates superior performance in mixed-class classification and more efficiently handles large volumes of medium-resolution images. Future research may explore hybrid models that combine RF with deep learning or other ML algorithms, alongside higher-resolution imagery and global datasets, to improve classification accuracy and broaden comparative perspectives.

Keywords: Google Earth Engine, kappa coefficient, Landsat satellite imagery, land use and land cover, Random Forest, Support Vector Machine

Introduction

Land use and land cover (LULC) changes are the most significant impacts of the urbanization process at the local and regional levels (Nuissl & Siedentop, 2021). These changes are driven by population growth, increasing socioeconomic development, and infrastructure development (Othman et al., 2021). As cities grow, natural landscapes, such as forest areas and agricultural

lands, are often converted and developed for settlement development and infrastructure construction, resulting in significant changes in LULC patterns (Ma et al., 2024). This situation is particularly pronounced in areas experiencing rapid population influx and inadequate land management policies, which require sustainable planning to balance urban growth with environmental preservation (Wu et al., 2011). Hence, monitoring urban growth and land use changes is important for the benefit of ecosystems and sustainable development planning.

Remote sensing technology and Geographic Information Systems (GIS) are powerful tools for obtaining accurate spatial information and detecting changes over large areas (Rane et al., 2023; Olokeogun et al., 2014). Spatial-temporal data, such as satellite imagery, enables the monitoring of land use dynamics at high temporal resolution and at a lower cost than conventional methods (El-Raey et al., 1995). Landsat satellite images are often used in LULC change detection studies due to their free accessibility, and long-term data record (Alam et al., 2020). Consequently, up-to-date and accurate LULC maps are crucial for planning, sustainable development, environmental monitoring and global change studies (Shapla et al., 2015; Dewan & Yamaguchi, 2009).

Google Earth Engine (GEE), a cloud-based geospatial computing platform, is used widely in LULC research, providing petabyte-scale datasets and scalable analytical capabilities (Alshehri et al., 2025; Liu et al., 2018). Additionally, GEE provides efficient large-scale analysis due to its high-performance computing capabilities without requiring local storage (Amani et al., 2020). As a free-access platform, GEE allows users to process datasets without downloading or installing additional software (Tamiminia et al., 2020). Moreover, GEE offers a wide range of built-in algorithms, including classification algorithms such as Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Tree (CART), k-Nearest Neighbors (k-NN), and other supervised classification methods (Amiren et al., 2024; Tesfaye et al., 2024; Sharma et al., 2023; Hasan et al., 2022; Hayashi et al., 2019). These features facilitate diverse geospatial applications (Alshehri et al., 2025), particularly in LULC change monitoring (Amini et al., 2022; Gautam & Rai 2022; Loukika et al., 2021).

The most robust methodology for generating LULC maps for monitoring spatial changes relies on the classification of remotely sensed imagery using various classification techniques or algorithms (Kavzoglu et al., 2018). The selection of an appropriate classifier is of critical importance for accurate LULC mapping. In recent years, machine learning algorithms have emerged as superior approaches, improving classification accuracy while reducing processing cost compared to traditional methods (Yuh et al., 2023; Zerrouki et al., 2019). The most widely used algorithms in LULC classification are Random Forest (RF) (Xu et al., 2024; Amini et al., 2022; Tokar et al., 2018; Ming et al., 2016) and Support Vector Machine (SVM) (Jozdani et al., 2019; Ustuner et al., 2015) algorithms.

Previous studies reported that RF algorithm, an ensemble classifier, consistently achieves high classification accuracy, robustness in handling complex terrain, ability to reduce the dimensionality of feature parameters effectively, and is adaptable to various types of remote sensing data (Xu et al., 2024; Ma et al., 2019; Ma et al., 2016; Tokar et al., 2018; Guan et al., 2012). In contrast, the SVM algorithm, a kernel-based classifier, performs well with small training samples (Thanh Noi & Kappas, 2017; Shang et al., 2011), handles complex datasets effectively (Zare et al., 2019) and offers flexibility through kernel selection (Kumar et al., 2024; Samardžić-Petrović et al., 2016). Although both algorithms demonstrate strong performance, their results are inconsistent across different studies (Avci et al., 2023). For examples, Mirjalalov et al. (2025), Amiren et al. (2024), Atef et al. (2023), Nazri et al. (2023) and Dabija et al. (2021) reported that SVM classifier outperforms the RF classifier in LULC classification. In contrast, studies by Bogale

et al. (2025), Thiyagarajan and Vijayalakshmi (2024), Zafar et al. (2024), and Adugna et al. (2022) revealed that RF classifiers are superior to SVM. While previous studies have compared RF and SVM algorithms, their relative performance remains context-dependent and varies across geographic regions and datasets.

Therefore, this study aims to evaluate and compare the performance of machine learning algorithms in LULC classification within a specific regional context. For this purpose, Random Forest (RF) and Support Vector Machine (SVM) algorithms were applied to Landsat satellite images on the GEE platform for the study area of Seremban and Port Dickson districts in Negeri Sembilan, Malaysia. To the best of our knowledge, region specific comparisons of these algorithms in this study area remain limited. In addition, this study examines the spatial and temporal dynamics of LULC over the 2010-2020 period, providing locally relevant information to support land management and urban planning.

Study area

The study area comprises the Seremban dan Port Dickson districts, located in Negeri Sembilan, with a total area of 153,674.11 hectares. Figure 1 shows the location of the Seremban and Port Dickson districts, within the administrative boundary of Negeri Sembilan. The study area is located adjacent to Klang Valley and lies within a radius of 60 km from Kuala Lumpur metropolitan city. Seremban and Port Dickson districts are part of the new growth corridor known as Malaysia Vision Valley (MVV), which aims to support development within the National Conurbation region. The establishment of MVV aligns with the Eleventh Malaysia Plan (RMK-11) 2016-2020, which emphasizes the development of a resilient region with balanced economic, social and environmental growth (PLANMalaysia, 2021). The topography of the area consists of relatively flat and gently sloping terrain near the coastline, with some hilly regions towards the eastern part of the Seremban district. According to the Census of Malaysia (2024), the population of the study area increased exponentially from 490,160 in 2000 to 847,900 in 2024 (Department of Statistics Malaysia, 2024). Seremban and Port Dickson districts have experienced significant urbanization spillovers from the Klang Valley, where rapid urban sprawl has occurred at varying rates. Since becoming part of the National Conurbation in 2016, this expansion has contributed to encroachments on vegetation and natural land cover (Aburas et al., 2018). Therefore, this study area was selected due to its rapid urban expansion and strategic location within a major growth corridor, making it highly suitable for analyzing LULC changes.

© 2026, e-ISSN 2682-7727 https://doi.org/10.17576/geo-2026-2202-05

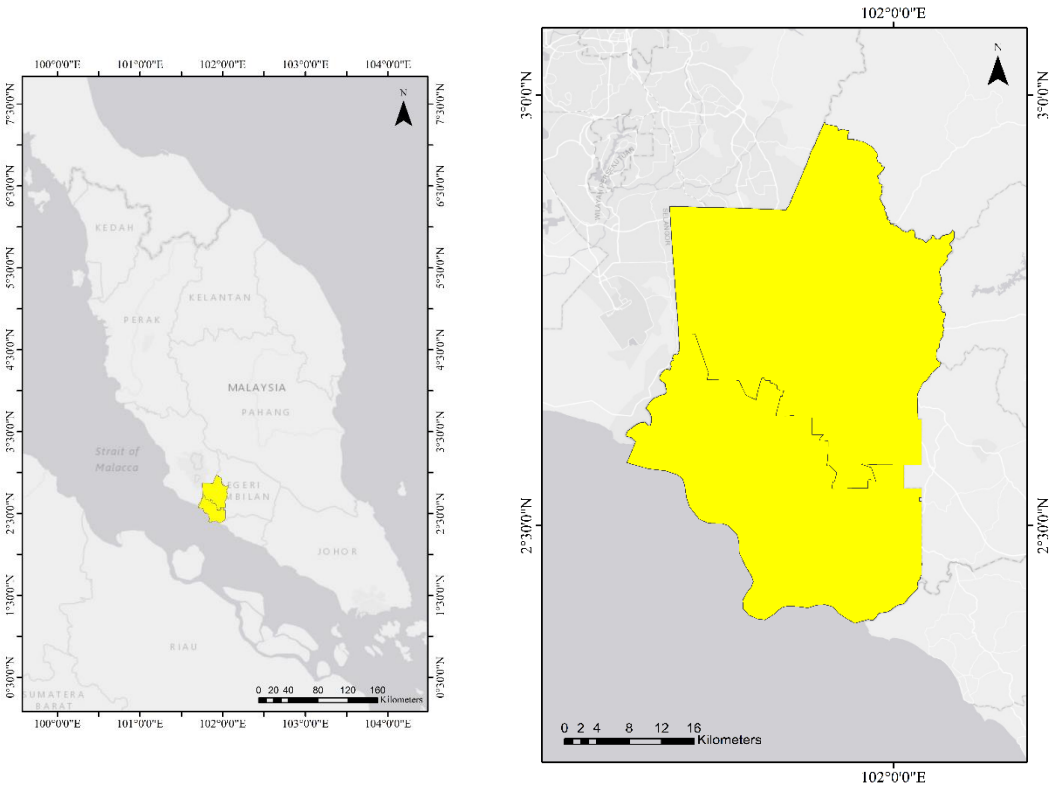


Figure 1. Location of Seremban and Port Dickson districts, Negeri Sembilan, Malaysia

Materials and method

Data sources and preparation

The workflow of the research method is shown in Figure 2. This research utilized Landsat 5 TM imagery from 2010 and Landsat 8 OLI imagery from 2020, both acquired from the Google Earth Engine (GEE) data catalog. The acquired Landsat imagery was processed on GEE platform. The spatial resolution of the images is 30 meters. The Landsat images were selected, followed by spatial filtering to the region of interest (ROI), temporal filtering, and cloud masking. A buffer area around the study area was drawn as the ROI to ensure the desired boundary was included by using the *ee.Geometry* tool in GEE JavaScript. The time span of imagery for 2010 is from January 1, 2009, to December 31, 2011, while for 2020, the imagery spanned from January 1, 2019, to December 31, 2021. These varying date ranges of Landsat images are needed to create a cloud-free median composite. Subsequently, a radiometric scaling process was performed to convert pixel values, referred to as digital numbers (DN), into reflectance values, thereby assigning physical meaning to the Landsat images. Lastly, RGB composition was applied to distinguish surface features for each study year.

© 2026, e-ISSN 2682-7727 https://doi.org/10.17576/geo-2026-2202-05

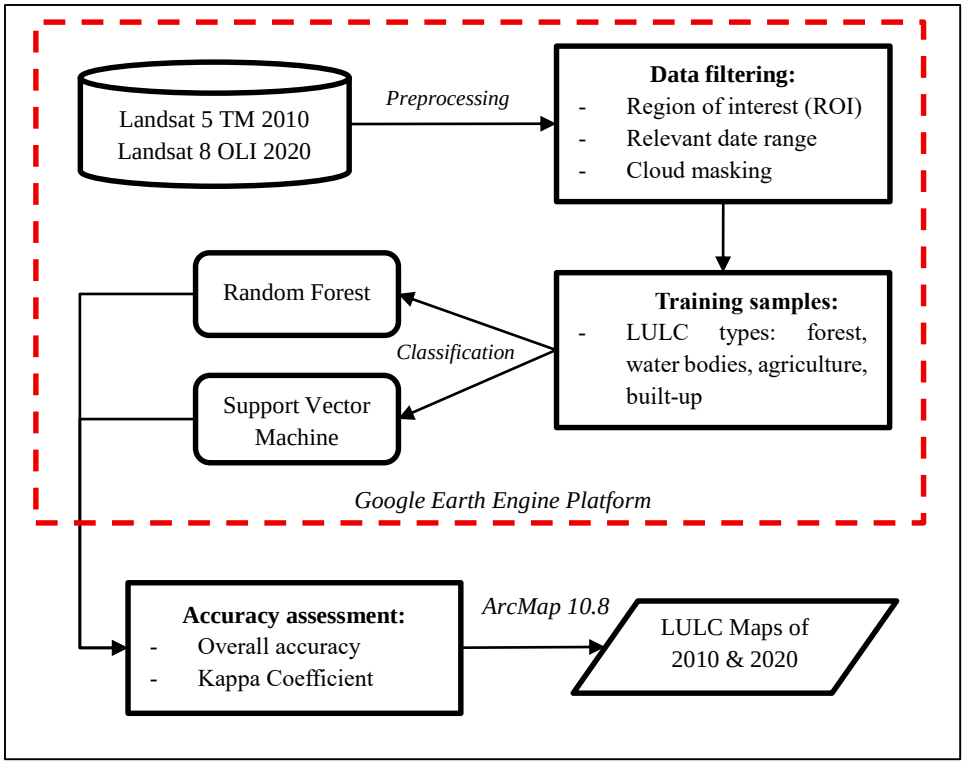


Figure 2. Flowchart of research method

Classification

In the LULC classification procedure, a set of training data with a specified number of pixels per class was created first. The LULC categories are divided into four main categories: forest, water bodies, agriculture, and built-up areas. Feature extraction using RGB composition was drawn in polygons. Table 1 shows the number of training samples collected for each LULC class. Then, the collected features or samples were grouped into four main categories. The pixel classification was carried out using the Random Forest (RF) and Support Vector Machine (SVM) algorithms on the GEE platform. The samples generated were used to train RF and SVM algorithms for LULC classification.

RF is an ensemble learning algorithm for machine learning models that use bootstrap techniques to build many single decision tree models (Mellor et al., 2013). Meanwhile, the principle of the SVM model is based on the optimal hyperplane to find the best boundary that separates two different classes (Samardžić-Petrović, 2017). In this research, the RF classifier parameters were set to 250 trees and 3 variable numbers using the tool *ee.Classifier.smileRandomForest()*. For the SVM classifier, the default setting of *e.Classifier.libsvm()* was employed on GEE. Subsequently, the trained classifiers were applied to classify the LULC of the four images.

Table 1. Numbers of training samples for each LULC classes

Land use and land cover classes	Training samples	
	2010	2020
Forest	13	13
Water bodies	6	6
Agriculture	30	23
Built-up	17	16

Accuracy assessment

The LULC maps produced by the RF and SVM classifiers were assessed and compared to identify the best model for land use classification. Accuracy assessment of classification images is necessary to ensure the reliability of findings obtained from remote sensing data and to evaluate the robustness of classification or distribution techniques (Panuju et al., 2020). In this research, an error matrix, also known as a confusion matrix, has been adopted. This method compares classified images with real-time reference data to calculate the user's accuracy, producer's accuracy, overall accuracy, and kappa coefficient (Wickham et al., 2017).

A confusion matrix was first constructed by generating a set of sampling points on the classified images. In this research, the stratified random sampling point technique was applied to create 500 sampling points, proportional to the total area of each land use class, using ArcMap 10.8 software for four classified maps. This stratified sampling approach ensures better representation of all land use types (Shetty et al., 2021). Then, the classified maps for each study year were compared with high-resolution Google Earth Pro images as ground truth data for assessing the accuracy of land use classification. Finally, an error matrix table was generated that contained the user's accuracy, producer's accuracy, overall accuracy and kappa coefficient values.

The user's accuracy is a commission error, in which pixels are incorrectly classified as a known class when they should have been classified as something else. In contrast, the producer's accuracy is an omission error, indicating how accurately the classification results meet the expectations of the creator. The kappa coefficient, also known as Cohen's kappa, is a crucial measurement method for accuracy assessment, evaluating the agreement between the classified map and the reference data (Maxwell & Warner., 2020). Cohen's kappa value ranges from 0 (complete disagreement) to 1 (perfect agreement). The Cohen's kappa is calculated using the kappa equation (k) as follows:

$$k = \frac{P_o - P_e}{1 - P_e}$$

k = kappa coefficient value
P_o = The proportion of observed agreement
P_e = The proportion of expected agreement

© 2026, e-ISSN 2682-7727 https://doi.org/10.17576/geo-2026-2202-05

Results

Accuracy assessment

The LULC map produced by different classifiers (RS and SVM) were assessed and compared to identify the best algorithm for land use classification. Landsat satellite-based Google Earth imagery for the years 2010 and 2020 were used as ground truth data for validation of LULC classification of both RF and SVM algorithms. In the accuracy assessment, four classified land use categories were considered: forest, water bodies, agriculture and built-up land. A set of total 500 sampling points on the classified images was generated using the stratified random sampling tool in ArcMap application for both classification algorithms. Then, the accuracy assessment results of LULC classification were derived from a confusion matrix which consists of producer's accuracy (PA), user's accuracy (UA), overall accuracy and kappa statistics. A confusion matrix defined the performance of the classification algorithms (Singh et al., 2021).

Based on the results of 2010 (Table 2), the confusion matrix table indicates minor misclassification occurred for agricultural and built-up land according to the UA value above 0.85 in both RF and SVM algorithms. Meanwhile, the water bodies category was classified 100% precisely in RF and SVM algorithms. However, the forest category was frequently misclassified as agricultural land , particularly using the SVM classifier where 67 samples (40.85%) were misclassified. Consequently, the classification of forests achieved a lower UA value of 0.5732 for SVM algorithm compared to RF algorithm with UA value of 0.8333 for the same land use category. This result indicates that the RF classifier performs very well, with most classes having 83% - 100% accuracy.

Table 2. Confusion matrix of land use classification using RF and SVM algorithms for year 2010

Algorithms	Land use & land cover	Forest	Water bodies	Agriculture	Built-up	Total	User's accuracy
Random Forest	Forest	**90**	1	17	0	108	0.8333
	Water bodies	0	**19**	0	0	19	1
	Agriculture	13	0	**276**	5	294	0.9388
	Built-up	2	1	2	**74**	79	0.9367
	Total	105	21	295	79	**500**	0
	Producer's accuracy	0.8571	0.9048	0.9356	0.9367	0	**0.918**
Support Vector Machine	Forest	**94**	1	67	2	164	0.5732
	Water bodies	0	**19**	0	0	19	1
	Agriculture	19	0	**220**	15	254	0.8661
	Built-up	1	0	5	**57**	63	0.9048
	Total	114	20	292	74	**500**	0
	Producer's accuracy	0.8246	0.95	0.7534	0.7703	0	**0.780**

© 2026, e-ISSN 2682-7727 https://doi.org/10.17576/geo-2026-2202-05

Regarding the year 2020 results (Table 3), the confusion matrix table exhibits the same pattern as the previous year where agricultural and built-up land categories achieved minor misclassification with the UA value more than 0.84 for both algorithms. Water bodies obtained a UA value of 1, indicating perfect classification by RF and SVM classifiers. Meanwhile, the classification of forest achieved the lowest accuracy by SVM classifier with a UA value of 0.744 due to its misclassification as agricultural land. In contrast, 91% of forest cover was classified correctly using RF algorithm for year 2020 indicating a well-performing classifier compared to SVM algorithm.

Table 3. Confusion matrix of land use classification using RF and SVM algorithms for year 2020

Algorithms	Land use & land cover	Forest	Water bodies	Agriculture	Built-up	Total	User's accuracy
Random Forest	Forest	**71**	0	7	0	78	0.9103
	Water bodies	0	**21**	0	0	21	1
	Agriculture	30	3	**273**	12	318	0.85856
	Built-up	1	0	2	**80**	83	0.9639
	Total	102	24	282	92	**500**	0
	Producer's accuracy	0.6961	0.875	0.9681	0.8696	0	**0.890**
Support Vector Machine	Forest	**93**	0	30	2	125	0.744
	Water bodies	0	**19**	0	0	19	1
	Agriculture	15	0	**251**	31	297	0.8451
	Built-up	0	1	0	**58**	59	0.9831
	Total	108	20	281	91	**500**	0
	Producer's accuracy	0.8611	0.95	0.8932	0.6374	0	**0.842**

Table 4 presents the overall accuracy and kappa coefficient of each algorithm. The RF classifier achieved the highest accuracy for two-year studies, with an overall accuracy value higher than 0.890 and a kappa coefficient value greater than 0.80. In contrast, the SVM classifier produced the lowest accuracy for 2010, with an overall accuracy of 0.780 and a kappa coefficient value of 0.6384, while the overall accuracy value was 0.842 and the kappa coefficient value of 0.7318 for land use classification of 2020. To summarize, the RF algorithm yielded better overall accuracy with a higher kappa coefficient value, compared to the SVM algorithm in both study years for land use classification.

© 2026, e-ISSN 2682-7727 https://doi.org/10.17576/geo-2026-2202-05

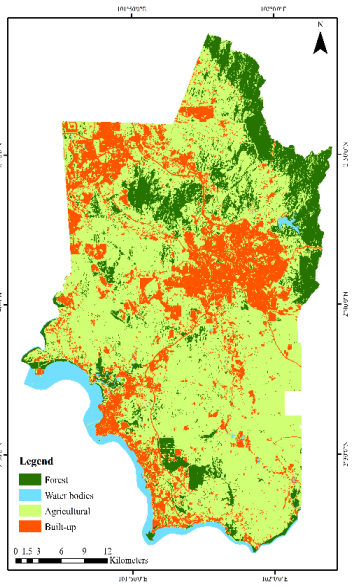
Table 4. Overall accuracy and kappa values of the LULC classification using RF and SVM algorithms

Year	Random Forest		Support Vector Machine	
	Overall accuracy	Kappa	Overall accuracy	Kappa
2010	0.918	0.8589	0.780	0.6384
2020	0.890	0.8093	0.842	0.7318

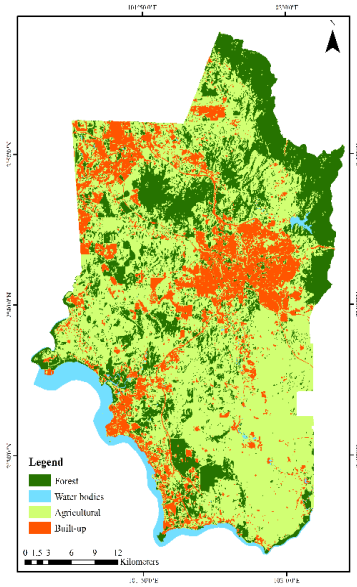
Comparison of land use classification

In this study, land use and land cover classification can be divided into four main categories: forest, water bodies, agriculture and built-up areas. Forest classifications consist of lowland and hill dipterocarp forests, dry inland forests and mangrove forests. Water bodies encompass the open surface area of water, including rivers, reservoirs, lakes, dams, ponds, mining areas and the sea along the coastline to include the reclamation project of overwater built-up areas classification. The agricultural land comprises oil palm plantations, rubber crops, orchards, and cash crops. Built-up areas include a wide range of developed land such as residential, commercial, industrial, infrastructures, facilities and utilities.

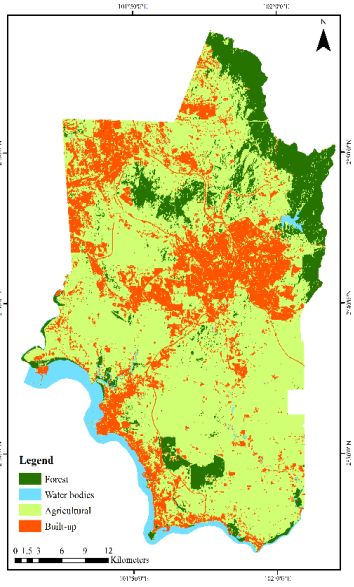
Land use maps of Seremban and Port Dickson districts in 2010 and 2020, classified by different classifiers are presented in Figure 3. The graphical LULC distribution of all the maps reveals a similar pattern for all LULC categories with varying intensity and density. Spatially, the dominant distribution of agricultural land is evident across all classified outputs of RF and SVM classifiers, which are scattered throughout the region. On the east side, the density of forest land is higher. Meanwhile, built-up areas are concentrated in the middle, northwest, and southwest sides of the study area. The major water bodies are clearly visible on the southwest and south sides of the study area. For the classified map of the SVM classifier for both study years, agricultural land is largely intermixed with forest and built-up areas, resulting in an exchange of areas between them in the maps. Meanwhile, the LULC map classified by RF classifier indicates that agricultural land and forest area are greatly mixed up in the classified map of 2020. From the four maps, misclassification occurred in areas where different features were either mixed up or in proximity to each other.



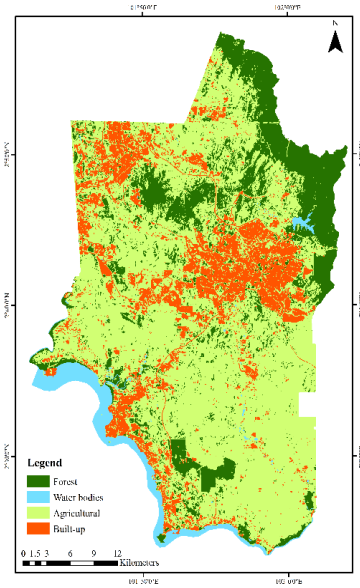
(1a) RF 2010



(1b) SVM 2010



(2a) RF 2020



(2b) SVM 2020

Figure 3. LULC classification results using RF and SVM classifier for study year 2010 and 2020

Based on Table 5, areas of the maps are calculated to compare the variation in the spatial extent of the different LULC classes. The results showed that the LULC of the study area is predominantly covered by agricultural land, with the highest total area recorded in both study years and algorithms. The SVM algorithm overclassified the forest cover with 51,914.56 hectares (33.15%) in 2010 and 39,631 hectares (25.01%) in 2020 compared to the RF algorithm. At the same time, the RF algorithm overclassified the agricultural land with 93,098.06 hectares (58.76%) in 2010 and 100,604.54 hectares (63.50%) in 2020 compared to the SVM algorithm. Besides, the RF algorithm measured the highest total area of built-up land, at 24,998.27 hectares (15.78%) in 2010 and 26,356.01 hectares (16.64%) in 2020, compared to the SVM algorithm. For the water bodies class, the RF algorithm has classified a slightly higher total area, at 6,008.04 hectares (3.79%) in 2010 and 6,664.20 hectares (4.21%) in 2020, compared to the SVM algorithm.

Since the RF classifier produced the most accurate classification result, the total area of LULC was taken into consideration (Table 5). The LULC pattern of study area revealed that the Seremban and Port Dickson districts are predominantly covered by agricultural land, accounting for more than 50% of the total area, followed by built-up area, forest cover and water bodies. The LULC change pattern in the study area displayed an increasing trend for agricultural land, water bodies, and built-up areas, except for forest cover, which showed a decline in total area between 2010 and 2020.

In general, the total area of LULC from Table 5 provides a quantitative comparison of LULC class distribution derived from RF and SVM classifiers for 2010 and 2020. This comparison not only highlights the consistency and differences between two classification approaches but also validates the selection of the RF classifier as the more reliable model based on its higher accuracy. Furthermore, the table supports the analysis of temporal LULC dynamics by clearly illustrating the dominance of agricultural land, the expansion of built-up areas and water bodies, and the decline in forest cover over the study period. These quantified changes form the basis for interpreting spatial trends in study area.

Table 5. Total area of land use and land cover classes in different algorithms

Land use	Random Forest				Support Vector Machine			
	2010 (ha)	%	2020 (ha)	%	2010 (ha)	%	2020 (ha)	%
Forest	34,323.629	21.67	24,810.41	15.66	51,914.56	33.15	39,631.00	25.01
Water bodies	6,008.041	3.79	6,664.20	4.21	5,963.78	3.81	6,191.61	3.91
Agriculture	93,098.063	58.76	100,604.54	63.50	78,882.94	50.37	94,038.39	59.35
Built-up	24,998.270	15.78	26,356.01	16.64	19,858.91	12.68	18,573.77	11.72
Total	**158,428.003**	**100**	**158,435.16**	**100**	**156,620.20**	**100**	**158,434.76**	**100**

Discussion

In this research, two types of algorithms, RF and SVM, were employed in LULC classification using the GEE platform for the study area. The overall accuracy and kappa coefficient results of the RF classifier are comparatively higher than those of the SVM classifier. The kappa result of the RF classifier is interpreted as an almost perfect agreement according to Landis and Koch's (1977) kappa coefficient scale interpretation. Unlike the RF classifier, the SVM classifier exhibited relatively unstable performance, with kappa results indicating a substantial level of interpretation.

According to the confusion matrix results, the forest cover was consistently misclassified as agricultural land, particularly by the SVM classifier. It was challenging to classify vegetation categories, as the different types have very similar spectral reflectance (Zhao, 2013). Besides, the slight misclassification between built-up areas and agriculture may be due to neighboring spatial features and the sparse vegetation surrounding the development landscape. Due to the complex and mixed built infrastructures in urban areas, such as small parks or gardens in front of houses, the presence of scattered large trees beside dwellings affects the spectral, geometrical and statistical properties of the image band, leading to misclassification (Manandhar et al., 2009; Nimbalkar et al., 2018). In addition, there is a high degree of heterogeneity common in urban areas, which can lead to misclassification in machine learning algorithms (Cai et al., 2019; Ouma et al., 2023). In contrast, the water bodies were classified perfectly by the RF and SVM classifiers for both study years. This was expected, as the spectral reflectance of water bodies is significantly different from that of other land use categories. Overall, the RF classifier achieved a higher accuracy in LULC classification, while the SVM classifier produced lower accuracy in this research. Hence, selecting the right classifier is crucial for improving land use and land cover classification accuracy performance.

The performance of machine learning algorithms might be affected by several factors. The size number of tree of the RF algorithm could influence the classification performance. On the other hand, SVM algorithm was sensitive to noisy datasets, which affect hyperparameter tuning. In addition, the number of training samples for each LULC class also could affect the classifiers' performance. To sum up, the output produced by both classifiers was significantly different.

Conclusion

In general, this study has employed geospatial techniques and remote sensing data as the primary research method to identify the most effective algorithm for LULC classification of the study area. A LULC map provides important information for the development and management of urbanized regions. The LULC maps will be more reliable and useful if the classification accuracy is more accurate with a good scale, which largely varies from algorithm to algorithm. LULC classification of urban areas is more difficult due to the high degree of spectral mixing and spatial heterogeneity.

In this research, the LULC map of the Seremban and Port Dickson districts is classified using the machine learning algorithms of RF and SVM. The results showed that the RF classifier outperformed the SVM classifier in LULC classification. The highest degree of misclassification occurred in forest cover and agricultural land due to the almost similar spectral reflectance, particularly by the SVM classifier. In conclusion, training data size and classification algorithm types are important to achieve higher classification accuracy.

This study has generated significant information about the LULC dynamics of Seremban and Port Dickson districts and their transformation during the last 10 years (2010-2020) using Landsat satellite images. While high spatial resolution imagery exists such as Sentinel-2, Landsat image provides more consistent long-term observation, which are essential for temporal LULC analysis. In addition, this study specifically focuses on locally validated LULC information and cannot be fully substituted by global datasets. Finally, this study could contribute to urban planning applications by providing locally validated LULC dynamics.

In summary, this study demonstrates the relevance of region-specific LULC analysis using Landsat data, particularly for capturing decadal-scale land dynamics that are essential for urban

planning and policy support. The selection of a 10-year study period (2010-2020) is consistent with established practices in remote sensing, providing a balanced timeframe to detect meaningful land transformations while maintaining data consistency. Furthermore, the comparative evaluation of RF and SVM classifiers offers valuable context-dependent insights, as classification performance varies across regions and datasets. While acknowledging the advantages of higher-resolution imagery and global LULC products, this study emphasizes the importance of locally validated, temporally consistent analyses. Future work could build on this foundation by testing additional machine learning algorithms and incorporating extended study periods alongside higher-resolution imagery, such as Sentinel-2 and global datasets, such as Dynamic World to improve classification accuracy and enhance spatial and comparative perspectives.

Acknowledgement

The authors gratefully acknowledge Universiti Kebangsaan Malaysia for financial support under Geran Universiti Penyelidikan (GUP-2024-066).

References

Aburas, M. M., Ho, Y. M., Ramli, M. F., & Ash'aari, Z. H. (2018). Monitoring and assessment of urban growth patterns using spatio-temporal built-up area analysis. *Environmental Monitoring and Assessment, 190*(3), 156.

Adugna, T., Xu, W., & Fan, J. (2022). Comparison of Random Forest and Support Vector Machine classifiers for regional land cover mapping using coarse resolution FY-3C images. *Remote Sensing*, *14*(3), 574.

Alam, A., Bhat, M. S., & Maheen, M. (2020). Using Landsat satellite data for assessing the land use and land cover change in Kashmir valley. *GeoJournal*, *85*(6), 1529-1543.

Alshehri, B., Zhang, Z., & Liu, X. (2025). A review of Google Earth Engine for land use and land cover change analysis: Trends, applications, and challenges. *ISPRS International Journal of Geo-Information, 14*(11), 416.

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Moghaddam, S. H. A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., & Brisco, B. (2020). Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 5326-5350.

Amini, S., Saber, M., Rabiei-Dastjerdi, H., & Homayouni, S. (2022). Urban land use and land cover change analysis using Random Forest classification of Landsat time series. *Remote Sensing, 14*(11), 2654.

Amiren, M., Nazhifah, S. A., Rusdi, M., & Misbullah, A. (2024). Comparison of Support Vector Machine and Random Forest methods on Sentinel-2A Imagery for land cover identification in Banda Aceh City using Google Earth Engine. *The Indonesian Journal of Computer Science, 13*(6).

Atef, I., Ahmed, W., & Abdel-Maguid, R. H. (2023). Modelling of land use land cover changes using machine learning and GIS techniques: A case study in El-Fayoum Governorate, Egypt. *Environmental Monitoring and Assessment*, *195*(6), 637.

Avcı, C., Budak, M., Yağmur, N., & Balçık, F. (2023). Comparison between Random Forest and Support Vector Machine algorithms for LULC classification. *International Journal of Engineering and Geosciences, 8*(1), 1-10.

Bogale, T., Degefa, S., Dalle, G., & Abebe, G. (2025). Machine learning-based analysis of land use and land cover trends in southeastern Ethiopia using Google Earth Engine. *Discover Sustainability, 6*(1), 878.

Cai, G., Ren, H., Yang, L., Zhang, N., Du, M., & Wu, C. (2019). Detailed urban land use land cover classification at the metropolitan scale using a three-layer classification scheme. *Sensors, 19*(14), 3120.

Dabija, A., Kluczek, M., Zagajewski, B., Raczko, E., Kycko, M., Al-Sulttani, A.H., Tardà, A., Pineda, L. and Corbera, J. (2021). Comparison of Support Vector Machines and Random Forests for corine land cover mapping. *Remote Sensing, 13*(4), 777.

Department of Statistics Malaysia (2024). Data catalogue.

Dewan, A. M., & Yamaguchi, Y. (2009). Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. *Applied Geography, 29*(3), 390-401.

El-Raey, M., Nasr, S. M., El-Hattab, M. M., & Frihy, O. E. (1995). Change detection of Rosetta promontory over the last forty years. *Remote Sensing, 16*(5), 825-834.

Gautam, L., & Rai, R. (2022). Land use and land cover change analysis using Google Earth Engine in Manamati watershed of Kathmandu district, Nepal. *The Third Pole Journal of Geography Education, 22,* 49-60.

Guan, H., Yu, J., Li, J., & Luo, L. (2012). Random Forests-based feature selection for land-use classification using LiDAR data and orthoimagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39,* 203-208.

Hasan, S. H., AL-Hameedawi, A. N., & Ismael, H. S. (2022). Supervised classification model using Google Earth Engine development environment for Wasit Governorate. *IOP Conference Series: Earth and Environmental Science, 961*(1), 012051.

Hayashi, T., Tamukoh, H., & Kubota, R. (2019). Modified hierarchical k-Nearest Neighbor method with application to land-cover classification. *International Workshop on Advanced Image Technology (IWAIT), 11049,* 746-749.

Jozdani, S. E., Johnson, B. A., & Chen, D. (2019). Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine algorithms for object-based urban land use/land cover classification. *Remote Sensing, 11*(14), 1713.

Kavzoglu, T., Tonbul, H., Yildiz Erdemir, M., & Colkesen, I. (2018). Dimensionality reduction and classification of hyperspectral images using object-based image analysis. *Journal of the Indian Society of Remote Sensing, 46*(8), 1297-1306.

Kumar, M. D., Bhavani, Y. L., Sahithi, V. S., Kumar, K. A., & Cheepulla, H. (2024). Analysing the impact of training sample size in classification of satellite imagery. 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Liu, X., Hu, G., Chen, Y., Li, X., Xu, X., Li, S., Pei, F., & Wang, S. (2018). High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sensing of Environment, 209*, 227-239.

Loukika, K. N., Keesara, V. R., & Sridhar, V. (2021). Analysis of land use and land cover using machine learning algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability, 13*(24), 13758.

Ma, H., Gao, X., & Gu, X. (2019). Random Forest classification of Landsat 8 imagery for the complex terrain area based on the combination of spectral, topographic and texture information. *Journal of Geo-information Science, 21*, 359-371.

Ma, T. Z., Teh, B. T., & Kho, M. Y. (2024). Land use change and ecological network in rapid urban growth region in Selangor region, Malaysia. *Scientific Reports, 14*(1), 16470.

Ma, Y., Jiang, Q., Meng, Z., Li, Y., Wang, D., & Liu, H. (2016). Classification of land use in farming area based on Random Forest algorithm. *Transactions of the Chinese Society for Agricultural Machinery, 47*(1), 297-303.

Manandhar, R., Odeh, I. O., & Ancev, T. (2009). Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sensing, 1*(3), 330-344.

Maxwell, A. E., & Warner, T. A. (2020). Thematic classification accuracy assessment with inherently uncertain boundaries: An argument for center-weighted accuracy assessment metrics. *Remote Sensing, 12*(12), 1905.

Mellor, A., Haywood, A., Stone, C., & Jones, S. (2013). The performance of Random Forests in an operational setting for large area sclerophyll forest classification. *Remote Sensing, 5*(6), 2838-2856.

Ming, D., Zhou, T., Wang, M., & Tan, T. (2016). Land cover classification using Random Forest with genetic algorithm-based parameter optimization. *Journal of Applied Remote Sensing, 10*(3), 035021-035021.

Mirjalalov, N., Teshaev, N., Safarov, E., Gerts, J., Mominov, A., & Pardaboyev, A. (2025). Comparative analysis of Random Forest and Support Vector Machine for LULC classification in Tashkent Region using Landsat-8 imagery. *InterCarto InterGIS, 31*(1), 519.

Nadzri, I. F. M., Khalid, N., Wahab, W. A., & Hashim, N. (2023). Analyzing the effectiveness of Support Vector Machine and Random Forest classifiers in delineating the green area. *IOP Conference Series: Earth and Environmental Science, 1217*(1), 012032.

Nimbalkar, P., Jarocinska, A., & Zagajewski, B. (2018). Optimal band configuration for the roof surface characterization using hyperspectral and LiDAR imaging. *Journal of Spectroscopy, 2018*(1), 6460518.

Nuissl, H., & Siedentop, S. (2021). Urbanisation and land use change. *Sustainable Land Management in a European Context, 8*, 75-99.

Olokeogun, O. S., Iyiola, K., & Iyiola, O. F. (2014). Application of remote sensing and GIS in land use/land cover mapping and change detection in Shasha forest reserve, Nigeria. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40*, 613-616.

Othman, A. G., Ali, K. H., Yin, I., Tan, M. L., & Jizan, N. H. M. (2021). Urbanization and land use changes in rural town: Guar Cempedak, Kedah. *Planning Malaysia, 19*(19).

Ouma, Y. O., Keitsile, A., Nkwae, B., Odirile, P., Moalafhi, D., & Qi, J. (2023). Urban land-use classification using machine learning classifiers: Comparative evaluation and post-classification multi-feature fusion approach. *European Journal of Remote Sensing, 56*(1), 2173659.

Panuju, D. R., Paull, D. J., & Griffin, A. L. (2020). Change detection techniques based on multispectral images for investigating land cover dynamics. *Remote Sensing, 12*(11), 1781.

PLANMalaysia. (2021). Rancangan Fizikal Negara Keempat.

Rane, N. L., Achari, A., Choudhary, S. P., & Giduturi, M. (2023). Effectiveness and capability of remote sensing (RS) and geographic information systems (GIS): A powerful tool for land use and land cover (LULC) change and accuracy assessment. *International Journal of Innovative Science and Research Technology, 8*(4), 4375-4392.

Samardžić-Petrović, M., Dragićević, S., Kovačević, M., & Bajat, B. (2016). Modeling urban land use changes using support vector machines. *Transactions in GIS, 20*(5), 718-734.

Samardžić-Petrović, M., Kovačević, M., Bajat, B., & Dragićević, S. (2017). Machine learning techniques for modelling short term land-use change. *ISPRS International Journal of Geo-information, 6*(12), 387.

Shang, K., Li, P., & Cheng, T. (2011). Land cover classification of hyperspectral data using composite kernel support vector machines. *Acta Scientiarum Naturalium Universitatis Pekinensis, 47*(1), 109-114.

Shapla, T., Park, J., Hongo, C., & Kuze, H. (2015). Agricultural land cover change in Gazipur, Bangladesh, in relation to local economy studied using Landsat images. *Advances in Remote Sensing, 4*(3), 214-223.

Sharma, V., Singh, A., Prasad, Y., Gupta, S., Choudhury, T., & Kotecha, K. (2023). Land Use and Land Cover Classification for Temporal Analysis on Ganjam District Region, Odisha Using Remote Sensing and Google Earth Engine. IEEE International Conference on ICT in Business Industry & Government (ICTBIG).

Shetty, S., Gupta, P. K., Belgiu, M., & Srivastav, S. K. (2021). Assessing the effect of training sampling design on the performance of machine learning classifiers for land cover mapping using multi-temporal remote sensing data and Google Earth Engine. *Remote Sensing, 13*(8), 1433.

Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. In. K. K. Singh, M. Elhoseny, A, Singh, & A. A. Elngar (Eds.), *Machine learning and the internet of medical things in healthcare* (pp. 89-111). Academic Press.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing, 164*, 152-170.

Tesfaye, W., Elias, E., Warkineh, B., Tekalign, M., & Abebe, G. (2024). Modeling of land use and land cover changes using Google Earth Engine and machine learning approach: Implications for landscape management. *Environmental Systems Research, 13*(1), 31.

Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and Support Vector Machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors, 18*(1), 18.

Thiyagarajan, G., & Vijayalakshmi, V. (2024). Classification of land cover using machine learning models in Landsat satellite data. 15th International Conference on Computing Communication and Networking Technologies (ICCCNT).

Tokar, O., Havryliuk, S., Korol, M., Vovk, O., & Kolyasa, L. (2018). Using Multitemporal and Multisensoral Images for Land Cover Interpretation with Random Forest Algorithm in the Prykarpattya Region of Ukraine. In *Conference on Computer Science and Information Technologies* (pp. 48-64). Springer International Publishing.

Ustuner, M., Sanli, F. B., & Dixon, B. (2015). Application of support vector machines for land use classification using high-resolution rapid eye images: A sensitivity analysis. *European Journal of Remote Sensing, 48*(1), 403-422.

Wickham, J., Stehman, S. V., Gass, L., Dewitz, J. A., Sorenson, D. G., Granneman, B. J., Poss, R.V., & Baer, L. A. (2017). Thematic accuracy assessment of the 2011 national land cover database (NLCD). *Remote Sensing of Environment, 191*, 328-341.

Wu, Y., Zhang, X., & Shen, L. (2011). The impact of urbanization policy on land use change: A scenario analysis. *Cities, 28*(2), 147-159.

Xu, J., Chen, C., Zhou, S., Hu, W., & Zhang, W. (2024). Land use classification in mine-agriculture compound area based on multi-feature random forest: A case study of Peixian. *Frontiers in Sustainable Food Systems, 7*, 1335292.

Yuh, Y. G., Tracz, W., Matthews, H. D., & Turner, S. E. (2023). Application of machine learning approaches for land cover monitoring in northern Cameroon. *Ecological Informatics, 74*, 101955.

Zafar, Z., Zubair, M., Zha, Y., Fahd, S., & Nadeem, A. A. (2024). Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Sciences, 27*(2), 216-226.

Zare, M., Behnia, N., & Gabriels, D. (2019). Assessment of land cover changes using Taguchi-based optimized SVM classification approach. *Journal of the Indian Society of Remote Sensing, 47*(1), 45-52.

Zerrouki, N., Harrou, F., Sun, Y., & Hocini, L. (2019). A machine learning-based approach for land cover change detection using remote sensing and radiometric measurements. *IEEE Sensors Journal, 19*(14), 5843-5850.

Zhao, Y. (2013). Chapter 4—Decision Trees and Random Forest. In. Zhao, Y (Eds.), *R and data mining* (pp. 27-40). Cambridge, MA, USA, Academic Press.