Geografia-Malaysian Journal of Society and Space 19 issue 3 (1-16)
 e-ISSN 2682-7727 https://doi.org/10.17576/geo-2023-1903-01

Forest mapping in Peninsular Malaysia using Random Forest and Support Vector Machine Classifiers on Google Earth Engine

Farah Nuralissa Muhammad, Lam Kuok Choy

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

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

Received: 4 July 2023; Accepted: 28 August 2023; Published: 30 August 2023

Abstract

Forests play a crucial role in maintaining the balance of the global ecosystem by sustaining the interactions between living and non-living entities. Changes in forest areas encompass both growth and loss, often driven by development activities. Assessing forest cover and its changes is also a pivotal issue in forest management. Therefore, this study aims to investigate the performance of machine learning algorithms, namely Random Forest (RF) and Support Vector Machine (SVM), in mapping forest cover in Peninsular Malaysia. Landsat 5 TM and Landsat 8 OLI images were utilized to derive forest cover information. The classification process was automated using the remote sensing data management platform, Google Earth Engine (GEE). The accuracy assessment test using the Kappa coefficient resulted in a value of 0.7893 for the RF algorithm and 0.6328 for the SVM algorithm for the year 2010. Whereas, for the year 2020 the Kappa coefficient yielded 0.7475 for RF and 0.5893 for SVM. However, forest cover returned highest RF Kappa coefficient values of 0.875 (2010) and 0.8793 (2020), and SVM Kappa coefficient values of 0.8116 (2010) and 0.7313 (2020). The results implied that RF performed better in the land use classification compared to SVM. It is evident that this study can aid various stakeholders in assessing future plans and developments without compromising the environment.

Keywords: Google Earth Engine (GEE), Landsat, machine learning, Random Forest, stratified random, Support Vector Machine

Introduction

Forests worldwide are currently facing significant pressures due to deforestation and degradation. According to the FRA 2020 report, the global forest area in 2020 covered approximately 4.06 billion hectares or about 31%, but the distribution is not uniform across the globe. The rate of primary forest loss in tropical regions has been consistent in recent years. The loss of primary forests in tropical regions decreased by 11% in 2021 compared to 2020. Malaysia possesses extensive and rich forest areas with valuable natural resources. However, it is not immune to forest exploitation activities like other tropical forests in different countries. There are two main challenges in mapping vast forest cover: the processing of large datasets and the availability of cloud-free imagery. Therefore, Google Earth Engine (GEE), a cloud-based computing platform, can address these challenges in mapping extensive areas.

Google Earth Engine (GEE) is a cloud-based platform that provides users with free access to perform internet-based spatial data processing and analysis. GEE boasts a data repository in the petabyte range, containing various types of satellite images such as Landsat, Sentinel, MODIS, geophysical data (e.g. Digital Elevation Model (DEM), and climate data (e.g. CHIRPS, TRMM). This represents a modern scientific breakthrough as researchers, students, governments, and the public are encouraged to explore the applications of available data in GEE. Users can access GEE through an internet-based Application Programming Interface (API) and a web-based Interactive Development Environment (IDE) (Tamiminia et al., 2020; Gorelick et al., 2017). Furthermore, users don't need expertise in web programming or HyperText Markup Language to utilize GEE for various applications (Gorelick et al., 2017). GEE makes data processing more efficient and cost-effective in terms of time and expenses; for instance, it provides algorithms to generate cloud-free composite images, which are valuable for land-use studies.

Geospatial Big Data (GDB) is gaining global attention, and Google Earth Engine (GEE) is the latest platform capable of handling large-scale data processing for Remote Sensing and GIS. GDB refers to spatial datasets that exceed the current capacity of computing systems (Lee et al., 2015; Yang et al., 2017). GEE is a popular platform for processing massive geospatial data, facilitating scientific discoveries by providing users free access to a plethora of remote sensing datasets (Amani et al., 2019; Tamiminia et al., 2020). It's also a platform for performing planet-scale geospatial analysis with supercomputer capabilities for complex computations or large-scale processing tasks (Gorelick et al., 2017).

Remote sensing provides suitable data for forest area mapping and monitoring due to its high spatial and temporal resolution. Landsat data is frequently used for forest mapping and monitoring due to its adequate spatial resolution (30 m) and frequent temporal coverage (16 days) (FAO, 2010; Hansen et al., 2013; Gohar Ghazaryan, 2015; Nazarin Ezzaty, 2015; Nitze et al., 2015; Kanniah, Nazarin Ezzaty & Tuong, 2016; Nazarin Ezzaty & Kanniah, 2018; Fadli et al., 2019; Yantao et al., 2019; Osei, 2019; Xiaomei et al., 2019; Yang & Yunfeng, 2020; Potapov et al., 2022; Vahid et al., 2022). To date, RF is considered one of the most widely used algorithms for land cover classification using remote sensing data (Li et al., 2016; Jin et al., 2018; Millard & Richardson, 2015; Canovas-Garcia et al., 2017; Maxwell, 2019; Kelley et al., 2018; Teluguntla, 2018; Amani et al., 2019; Mellor et al., 2012; Jin et al., 2016; Lowe & Kulkarni, 2015; Pelletier et al., 2016; Basten, 2016; Gislason et al., 2006; Alper Akar, 2021). According to Mahdianpari et al. (2017) and Xia et al. (2017), the significant benefits of RF over the past two decades are its robustness to noisy and outlier-prone datasets. It performs well with high-dimensional and diverse sources of data and achieves higher accuracy compared to popular classifiers like SVM, kNN, or MLC in many applications (Rodriguez-Galiano & Chica-Rivas, 2012). It also enhances processing speed by selecting important variables (Schmidt et al., 2019). On the other hand, SVM is popular in remote sensing classification studies (Qian et al., 2015; Han, Pan & Devlin, 2018) and it is stated that SVM can be dealt with the classification of complex land use and land cover (Pretorius et al., 2016).

Various techniques, starting with visual interpretation of digital image classification, have been used for forest area mapping and monitoring (Berberoglu & Akin, 2009; Sirén & Brondizio, 2009; Forkuo & Frimpong, 2012; Hamdan et al., 2017; Huang et al., 2008; Sugumaran, 2001; Desclée et al., 2006; Rathinagiri et al., 2010). GIS approaches, such as GRASS, ArcGIS, and QGis, are often employed for forest mapping studies (Gautam, Shivakoti & Webb, 2004; Sau, 2013; Mellor et al., 2012; Qamer et al., 2016; Rio, Joko & Ratna, 2016; Wan Abdul Hamid, 2016). The Earth Resources Data Analysis System (ERDAS) software is also commonly used for forest mapping and change analysis (Tamaluddin et al., 2012; Prabhat Kumar Rai, 2013; Rio, Joko & Ratna, 2016; Wan Abdul Hamid, 2016; Solomon et al., 2018; Negassa, Mallie & Gemeda, 2020). Additionally, the CLASlite software, developed not only for forest cover detection but also for analyzing forest degradation and disturbances at different temporal scales using satellite images captured by various optical sensors (Asner, 2009;

Geografia-Malaysian Journal of Society and Space 19 issue 3 (1-16)
© 2023, e-ISSN 2682-7727 https://doi.org/10.17576/geo-2023-1903-01

Nazarin Ezzaty, 2015; Nazarin Ezzaty & Kanniah, 2018; Kanniah, Nazarin Ezzaty & Tuong, 2016). Large-scale forest mapping, such as across continents or globally, remains challenging. However, Google Earth Engine (GEE) has recently gained attention for large-scale remote sensing data processing on a global scale. Therefore, GEE has been used to map forest cover and changes (Gohar Ghazaryan, 2015; Xiaomei et al., 2019; Yantao et al., 2019; Fadli et al., 2019; Yang & Yunfeng, 2020; Fortin et al., 2020; Osei, 2019; Venkatappa et al., 2020; Shijuan et al., 2021). The aim of this study is to investigate the performance of machine learning algorithms to map the forest cover in Peninsular Malaysia. For this purpose, the Random Forest (RF) and Support Vector Machine (SVM) algorithms were selected while demonstrating the effectiveness of GEE in mapping vast areas at the national scale.

Methods

Peninsular Malaysia

Peninsular Malaysia or West Malaysia consists of 11 states and two Federal Territories, covering an area of 132,078 km^2 (Figure 1). The land use pattern in the Peninsular Malaysia region is divided into thirteen (13) land use types, including forests, agriculture, housing, water bodies, transportation, institutions and community facilities, vacant land, industry, open space and recreation, infrastructure and utilities, commercial, coastal, and mixed development. Forests are the predominant or dominant land use type in Peninsular Malaysia. The topography of Peninsular Malaysia encompasses areas of hills, mountains, rivers, and lakes. It also has gently sloping coastlines, dense and hilly forests. Peninsular Malaysia experiences an equatorial climate, characterized by hot and humid conditions throughout the year.

Figure 1. Peninsular Malaysia

Data

This study utilizes Landsat data from the Earth Engine data catalogue archive (https://developers.google.com/earth-engine/datasets/). The Landsat TM data used includes images from the year 2010, while the Landsat OLI data used covers images from the year 2020. Landsat images are employed for mapping the forest cover of Peninsular Malaysia due to their optimal spatial resolution and spectral bands that can discern land cover types. The resolution of Landsat images is 30 meters. For the year 2010, the data time period used spans from June 1, 2005, to May 5, 2012, with a total of 1237 images. For the year 2020, 876 images are used, covering the time period from January 1, 2019, to December 31, 2021. The number of images used for the study year varies based on the data time period required to generate the best cloud-free composite image for Peninsular Malaysia through the cloud masking and filling process, which will be discussed in the next subtopic. Additionally, this study also utilizes high-resolution (4.77 m) PlanetScope satellite imagery obtained through Norway's International Climate and Forest Initiative (NICFI) program and Google Earth Pro to aid in the validation process of the land cover classification result.

Mapping the forest cover of Peninsular Malaysia

Figure 2 shows the flowchart of the data analysis method employed in this study. The process of mapping the forest cover of Peninsular Malaysia is carried out within the Google Earth Engine (GEE) platform using JavaScript programming and involves Landsat data from the Earth Engine catalog. GEE functions through scripts. The Landsat TM and Landsat OLI TOA collections are accessed using the image collection function, and then the filter function is utilized to retrieve the data collection based on the selected year. The study uses the border of Peninsular Malaysia to filter the collection within the study area. At this stage, the filter function called for Landsat imagery collection for the whole of 2010 with a total of 108 images and for 2020 with a total of 294 images of the least clouds cover. Subsequently, the cloud masking and filling function is applied to obtain cloud-free Landsat composite images. This function is used to mask clouds and fill those areas with cloud-free imagery using various combinations of images from different time periods. For the year 2010, a total of 1237 images, spans from June 1, 2005 to May 5, 2012 were used to fill the masked cloud pixels. Whereas for the year 2020, a total of 876 images are used, beginning from January 1, 2019, to December 31, 2021. The basic concept for cloud masking involves assigning a value of '0' to cloud pixels. The cloud masking process involves identifying clouds from satellite images and masking them to obtain cloud-free images using the updateMask method. The updateMask method will replace pixel values between 0 and -1. Pixel values above 0 are retained. The *select()* function is used to extract the necessary bands from the image. The 'QA_RADSAT' band is chosen in this study. Next, the logical operator *eq()* (meaning 'equal') is used to produce a binary image where all pixel values that do not have 0 in the 'QA_RADSAT' band (no data) result in only a value of 1 in the generated image. The filling method is employed to fill in the areas that have been masked with cloud-free imagery using various combinations of median pixel values from images within different time periods.

The mapping of the forest cover of Peninsular Malaysia utilizes the RF and SVM classification algorithm from the JAVA library within GEE. RF has only two parameters: the number of trees in the forest (ntree) and the number of random variables (mtry) used in each tree. This study employs 250 trees (tree = 250) and mtry (3). RF is a non-parametric classification technique that utilizes bootstrap aggregation to combine the classification outcomes of multiple random decision trees to predict class labels (Breiman, 2001). The method requires training samples, where each pixel represents a land cover category. RF classification is used to classify data into land cover classes based on the defined training samples. Each random decision tree outcome is trained on a subset of training samples, which is the in-bag for internal cross-validation. Finally, the outcomes are combined to generate the land cover classification. In the SVM algorithm, a hyperlane is created and the data is divided into two classes. For the SVM classifier, the radial basic kernel function was selected and the hyperparameters for gamma and cost were selected respectively 0.5 and 10.

In this study, six land cover categories are defined: forest, water bodies, oil palm, rubber, built-up areas, and others. A dataset, represented as a FeatureCollection in GEE, is created for the training points. Subsequently, the RF algorithm is employed to generate multiple decision trees (forming a forest), each with slightly different randomness for training data and predictions. Each split in the decision trees is performed with a subset of training data and spectrum values. When the forest of decision trees is built and then applied to the images, each pixel is assigned labels from each tree, and the final label for that pixel is determined. Table 1 shows the number of training samples used for each land cover class.

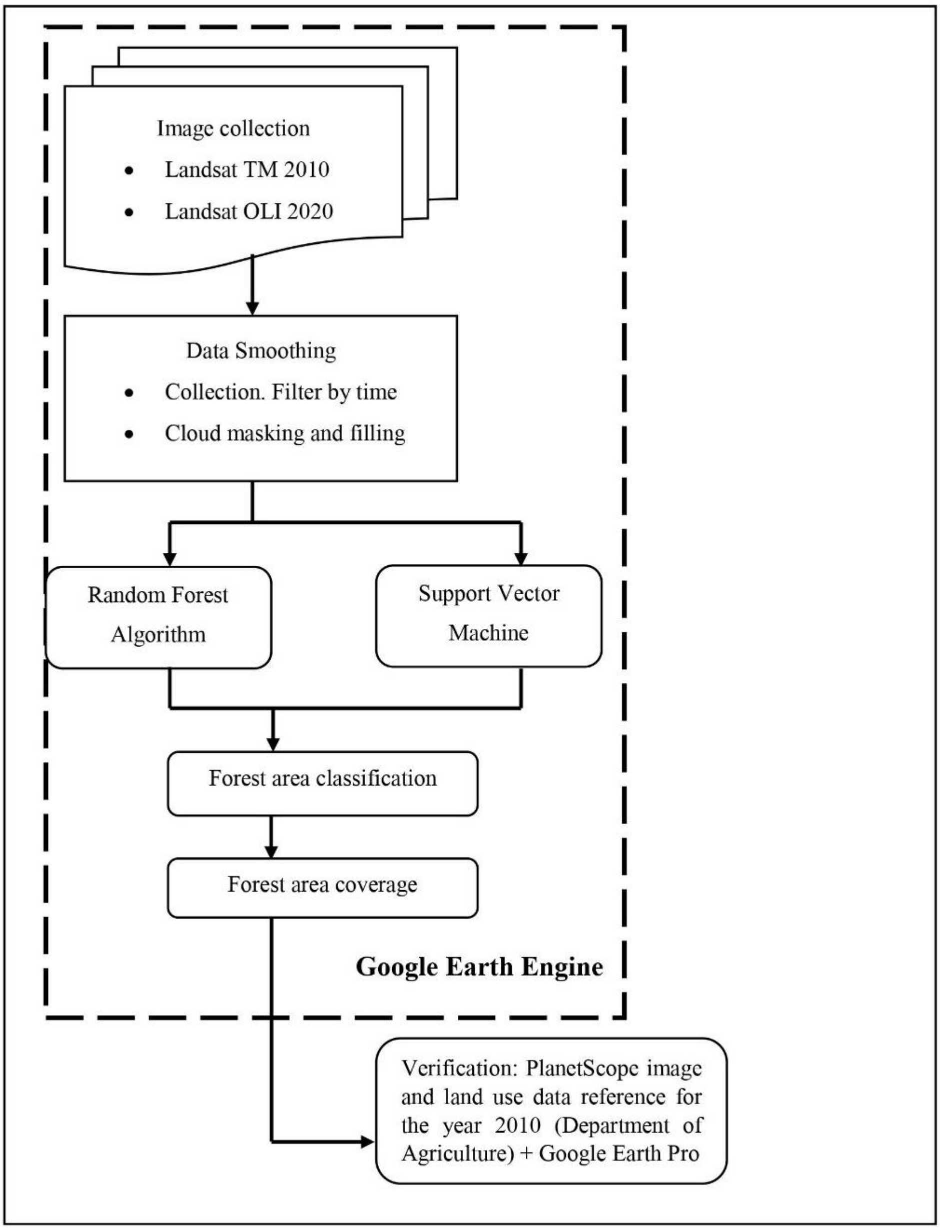

Figure 2. Flowchart of the data analysis method

Table 1. Number of training samples for each land cover class

Land cover	Training samples	
	2010	2020
Forest	20	20
Water bodies	9	12
Oil palm	29	27
Rubber	13	9
Built-up areas	8	30
Others	10	10

Accuracy assessment and data validation

Accuracy assessment is an essential part of land cover classification. It involves comparing the classification results with other data sources considered as ground truth. High-resolution satellite imagery data from PlanetScope provided through Norway's International Climate and Forest Initiative (NICFI) and Google Earth Pro have been used as ground truth data. The first step in the accuracy assessment process is to generate random sampling points. A total of 134 random sampling points were used for the year 2010, while 129 random sampling points were used for 2020 for classification using RF algorithm (Figure 3). As for SVM, a total of 125 random sampling points were used for the year 2010, while 122 random sampling points were used for 2020 (Figure 4). Commonly used sampling methods for accuracy assessment in land use and land cover (LULC) classification include simple random sampling, stratified random sampling, systematic sampling, and cluster sampling (Wagner & Stehman, 2015).

The sampling strategy employed in this study is stratified random sampling. Stratified random sampling generates a set of accuracy assessment points proportional to the area of each land cover class. This method was chosen due to its simplicity and comprehensibility (Cakir, Khorram & Nelson, 2006; Huang et al., 2010; Mayaux et al., 2006; Olofsson et al., 2011). Additionally, each land cover class is considered a stratum, and samples are randomly drawn from these strata, either based on proportional area, expected variance ratio, or an equal number across strata. This approach indirectly avoids biased sampling and reduces specific accuracy errors in classes with limited representation. Table 2 shows the number of random sampling points for each land cover class using RF algorithm. Meanwhile, Table 3 shows the number of random sampling points for each land cover class using SVM algorithm.

Table 2. Number of random sampling points for each land cover class using RF algorithm

Year	2010		2020	
Land cover	Area (ha)	Random sampling points	Area (ha)	Random sampling points
Forest	8,472,335.04	64	7,698,367.98	58
Water bodies	131,096.34	10	137,460.42	10
Oil palm	3,977,479.35	30	4,139,494.29	31
Rubber	95,523.39	10	19,385.55	10
Built-up areas	213,973.74	10	787,575.60	10
Others	385,717.77	10	492,039.27	10
Total	**13,276,125.63**	**134**	**13,274,323.11**	**129**

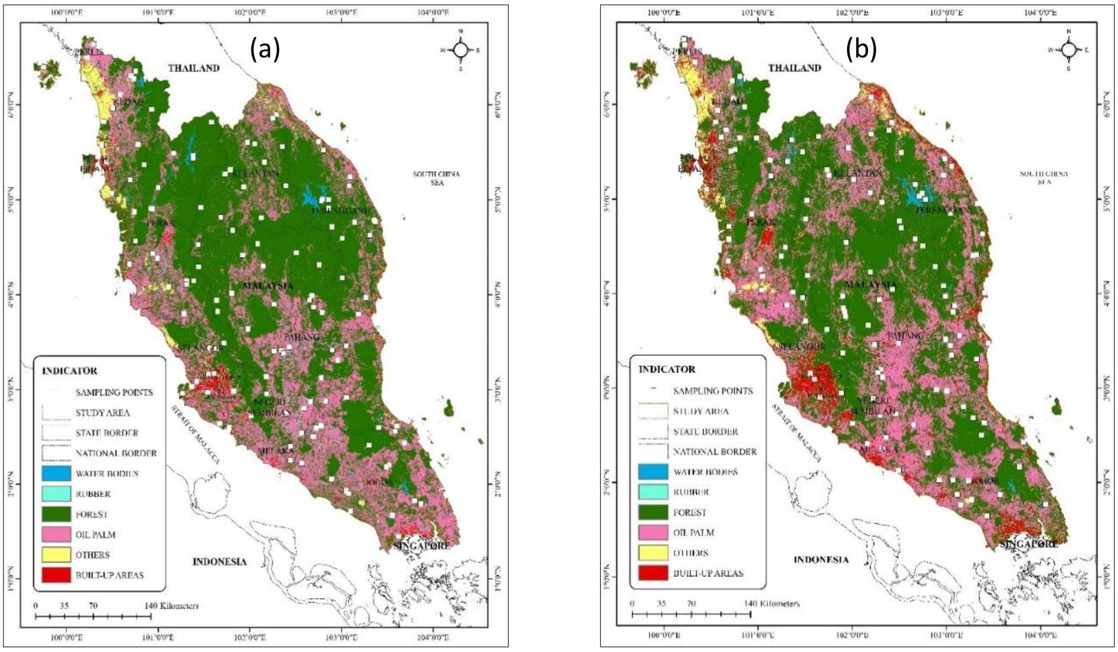

Figure 3. Sampling points for the valid determination of land classification using RF for the (a) 2010, (b) 2020 image

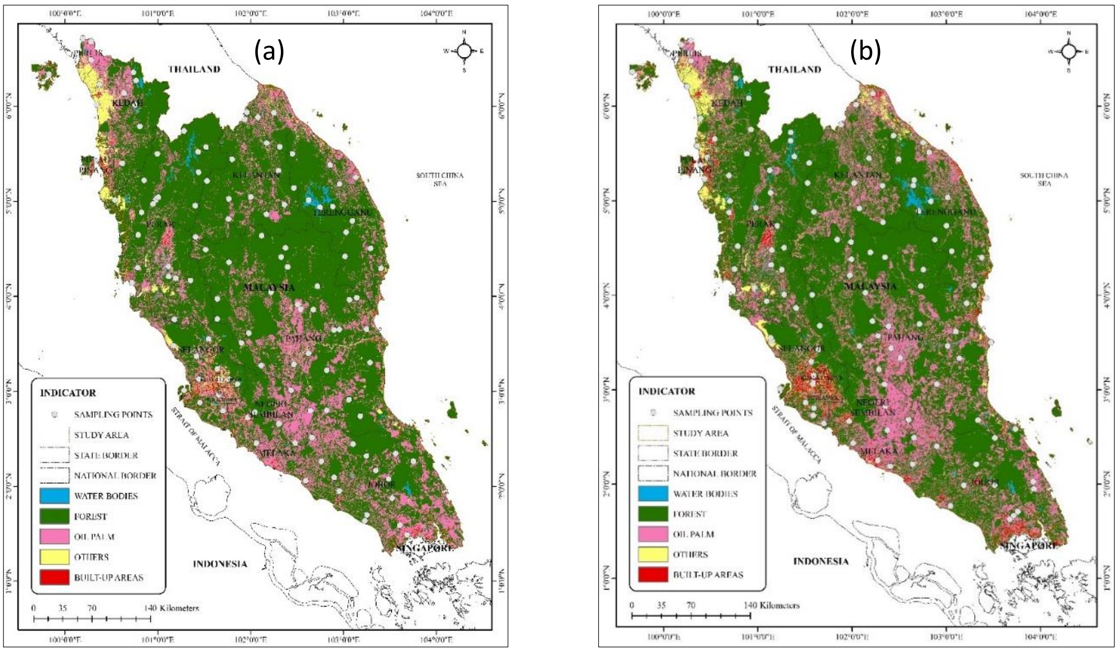

Figure 4. Sampling points for the valid determination of land classification using SVM for the (a) 2010, (b) 2020 image

Table 3. Number of random sampling points for each land cover class using SVM algorithm

Year	2010		2020	
Land cover	**Area (ha)**	**Random sampling points**	**Area (ha)**	**Random sampling points**
Forest	9,528,728.4	72	8,883,286.11	67
Water bodies	121,384.26	10	155,103.21	10
Oil palm	3,084,155.1	23	3,364,050.06	25
Rubber	-	-	-	-
Built-up areas	109,574.55	10	322,155.27	10
Others	432,229.95	10	549,741.24	10
Total	**13,276,072.3**	**125**	**13,274,335.89**	**122**

 e-ISSN 2682-7727 https://doi.org/10.17576/geo-2023-1903-01

Finally, this is followed by the process of calculating the confusion matrix, including omission and commission errors, to obtain the Kappa Coefficient and overall accuracy. The confusion matrix, or error matrix, is a quantitative method used to test the accuracy of classification results. It's a table that shows the agreement between land cover classification results and reference images (ground truth). Based on the confusion matrix, various accuracy metrics can be calculated, such as overall accuracy, producer's accuracy, user's accuracy, and the Kappa Coefficient. Overall accuracy represents the total accuracy of classification. Producer's accuracy refers to the probability that a specific land area is correctly classified, while user's accuracy refers to the probability that a pixel labeled as a certain class is indeed that class. The Kappa Coefficient indicates the classification accuracy. The range of the Kappa Coefficient is from 0 to 1, where 1 represents 100% accuracy. Therefore, a higher Kappa Coefficient indicates more accurate classification results. The computation for assessing image classification accuracy through the error matrix is carried out based on the Kappa equation (k) as follows:

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

k = Kappa Coefficient value
P_o = sum of the probability values from field inspections
P_e = sum of the expected probability values occurring by chance

Results and discussion

Accuracy assessment

A total of 134 random sampling points were generated on land use images for the year 2010, while 129 random sampling points were generated for 2020 for classification using the RF algorithm. As for classification using the SVM algorithm, a total of 125 random sampling points were generated on land use images for the year 2010, while 122 random sampling points were generated for 2020 using a stratified random sampling strategy. The RF classification indicated that for the year 2010, there were 64 sample points in forest land and 30 in oil palm. The number of sampling points for water bodies, rubber, built-up areas, and others were each 10. For the year 2020, there were 58 sample points in forest land and 31 in oil palm. The number of sampling points for water bodies, rubber, built-up areas, and others were each 10. The SVM classification using sampling points showed that for the year 2010, there were 72 sample points in forest land and 23 in oil palm. The number of sampling points for water bodies, built-up areas, and others were each 10. For the year 2020, there were 67 sample points in forest land and 25 in oil palm. The number of sampling points for water bodies, built-up areas, and others were each 10.

The accuracy testing for the RF classification using the Kappa coefficient indicated a value of 0.7893, which corresponds to an actual accuracy of 78.93 percent for the year 2010. For the year 2020, it was 0.7475 or 74.75 percent. These results demonstrate a high level of reliability in data interpretation through remote sensing image applications (Monserud & Leemans, 1992; Landis & Koch, 1977; Fleiss et al., 2013; Czaplewski, 1994). The user's accuracy values for forest land for the years 2010 and 2020 were very high, at 0.8750 and 0.8793, respectively. As for producer's accuracy values, they were 0.9180 for 2010 and 0.8361 for 2020. In conclusion, based on this assessment, there is confidence in using the land use data for further studies, allowing for subsequent research. Tables 3 and 4 show the accuracy assessment results for land use in 2010 and 2020 using the RF algorithm through a confusion matrix. The accuracy testing for the SVM classification using the Kappa coefficient indicated a value of 0.6328, which corresponds to an actual accuracy of 63.28 percent for the year 2010. For the year 2020, it was 0.5892 or 58.92 percent. These results indicate a moderate level of reliability in data interpretation through remote sensing image applications (Monserud & Leemans, 1992; Landis & Koch, 1977; Fleiss et al., 2013; Czaplewski, 1994). The user's accuracy values for forest land for the years 2010 and 2020 were very high, at 0.8116 and 0.7313, respectively. As for producer's accuracy values, they were 0.8615 for 2010 and 0.8750 for 2020. In conclusion, based on this assessment, there is confidence in using the land use data for further studies, allowing for subsequent research. Tables 5 and 6 show the accuracy assessment results for land use in 2010 and 2020 using the SVM algorithm through a confusion matrix.

Table 4. Land cover accuracy assessment for the year 2010 using RF algorithm through the confusion matrix

Land cover	Forest	Water bodies	Oil palm	Rubber	Built-up areas	Others	Total	User's accuracy	Kappa
Forest	56	0	7	1	0	0	64	**0.875**	
Water bodies	0	10	0	0	0	0	10	**1**	
Oil palm	5	0	21	1	3	0	30	**0.7**	
Rubber	0	0	2	8	0	0	10	**0.8**	
Built-up areas	0	0	0	0	9	1	10	**0.9**	
Others	0	0	0	0	0	10	10	**1**	
Total	61	10	30	10	12	11	134	0.0	
Producer's accuracy	**0.9180**	**1**	**0.7**	**0.8**	**0.75**	**0.9091**	0.0	**0.8507**	
Kappa									**0.7893**

Table 5. Land cover accuracy assessment for the year 2020 using RF algorithm through the confusion matrix

Land cover	Forest	Water bodies	Oil palm	Rubber	Built-up areas	Others	Total	User's accuracy	Kappa
Forest	51	0	7	0	0	0	58	**0.8793**	
Water bodies	0	10	0	0	0	0	10	**1**	
Oil palm	8	0	22	0	0	1	31	**0.7097**	
Rubber	1	0	3	6	0	0	10	**0.6**	
Built-up areas	1	1	0	0	7	1	10	**0.7**	
Others	0	0	0	0	0	10	10	**1**	
Total	61	11	32	6	7	12	129	0.00	
Producer's accuracy	**0.8361**	**0.9091**	**0.6875**	**1**	**1**	**0.8333**	0.00	**0.8217**	
Kappa									**0.7475**

Table 6. Land cover accuracy assessment for the year 2010 using SVM algorithm through the confusion matrix

Land cover	Forest	Water bodies	Oil palm	Rubber	Built-up areas	Others	Total	User's accuracy	Kappa
Forest	56	0	10	2	1	0	69	**0.8116**	
Water bodies	0	10	0	0	0	0	10	**1**	
Oil palm	7	0	15	2	0	0	24	**0.625**	
Rubber	0	0	0	8	0	0	0	**0**	
Built-up areas	1	0	0	0	8	1	11	**0.7273**	
Others	1	4	0	0	0	10	11	**0.5455**	
Total	65	14	25	4	9	11	125	0.0	
Producer's accuracy	**0.8615**	**0.7143**	**0.6**	**0**	**0.75**	**0.9091**	0.0	**0.76**	
Kappa									**0.6328**

Table 7. Land cover accuracy assessment for the year 2020 through using SVM algorithm the confusion matrix

Land cover	Forest	Water bodies	Oil palm	Rubber	Built-up areas	Others	Total	User's accuracy	Kappa
Forest	49	0	17	1	0	0	67	**0.7313**	
Water bodies	0	10	0	0	0	0	10	**1**	
Oil palm	5	0	12	5	0	3	25	**0.48**	
Rubber	0	0	0	0	0	0	0	**0**	
Built-up areas	0	0	0	0	10	0	10	**1**	
Others	2	1	0	0	0	7	10	**0.7**	
Total	56	11	29	6	10	10	122	0.00	
Producer's accuracy	**0.875**	**0.9090**	**0.4138**	**0**	**1**	**0.7**	0.00	**0.7213**	
Kappa									**0.5892**

Forest coverage of Peninsular Malaysia

The land use pattern resulting from the classification process of Landsat satellite images for the study area is divided into six (6) main land use types: forest, water bodies, oil palm, rubber, built-up areas, and others. Forest refers to land covering an area of more than 0.5 hectares with trees reaching a height of over 5 meters and a canopy cover exceeding 10%, or trees capable of reaching those thresholds 'in-situ' (PLANMalaysia). The forest categories analyzed in this study include dryland forest, hill dipterocarp forest, mangrove forest, and peat swamp forest. Water bodies are areas where water naturally or artificially covers the Earth's surface, encompassing rivers, lakes, reservoirs, ponds, and mining areas. Oil palm refers to areas cultivated with oil palm trees, while rubber pertains to areas cultivated with rubber trees. Built-up areas encompass land developed for housing, infrastructure and utilities, commercial, industrial, institutional, and community facilities. Other land use refers to areas used for farming, other crops such as coconut, pineapple, paddy, and others.

Although there are various land use types in this area, the discussion is focused solely on land cover involving forest areas. The study results using RF and SVM algorithms indicate that forests are the dominant land use in Peninsular Malaysia for the years 2010 and 2020 (Figure 5 & 6) and it can be observed that the forest area has decreased. The forest area using the RF algorithm shows that in 2010, it was 8,472,335.04 hectares, equivalent to 63.82 percent, and in 2020, it was 7,698,367.98 hectares, which is equivalent to 57.99 percent (Table 8). On the other hand, the forest area using SVM shows that in 2010, it was 9,528,728.4 hectares, equivalent to 71.77 percent, and in 2020, it was 8,883,286.11 hectares, which is equivalent to 66.92 percent (Table 9).

Table 8. The forest coverage area in Peninsular Malaysia for 2010 and 2020 using RF algorithm

Year	2010		2020	
Land cover	**Area (ha)**	**Percentage (%)**	**Area (ha)**	**Percentage (%)**
Forest	8,472,335.04	63.82	7,698,367.98	57.99
Water bodies	131,096.34	0.99	137,460.42	1.04
Oil palm	3,977,479.35	29.96	4,139,494.29	31.18
Rubber	95,523.39	0.72	19,385.55	0.15
Built-up areas	213,973.74	1.61	787,575.60	5.93
Others	385,717.77	2.91	492,039.27	3.71
Total	**13,276,125.63**	**100**	**13,274,323.11**	**100**

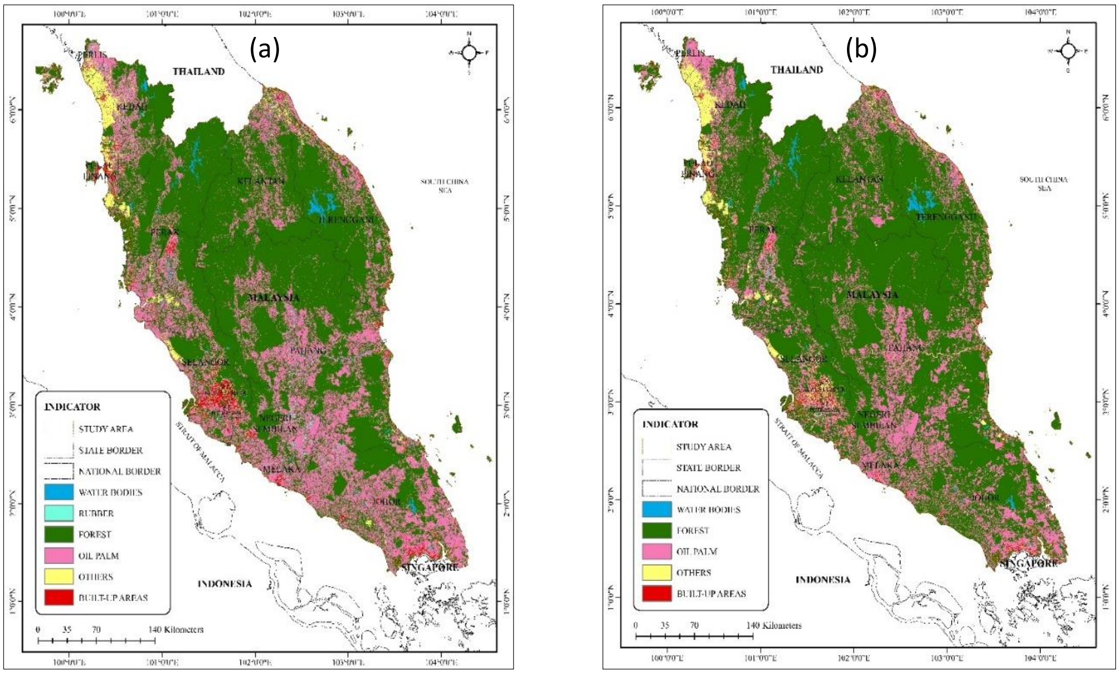

Figure 5. The forest coverage area in Peninsular Malaysia using (a) RF, (b) SVM for the year 2010

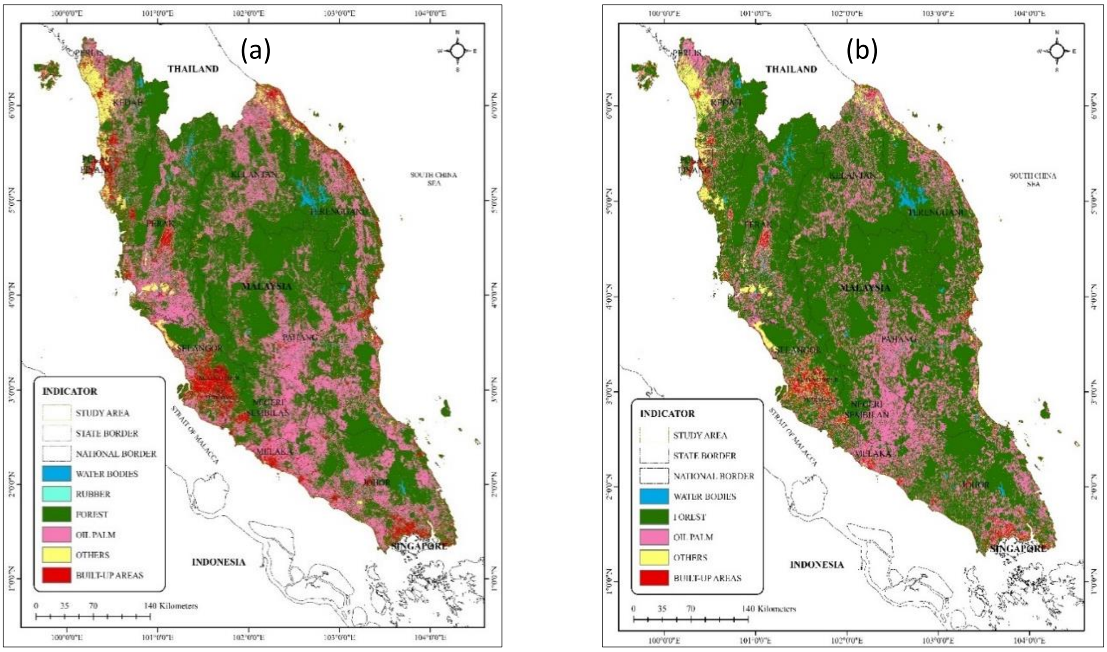

Figure 6. The forest coverage area in Peninsular Malaysia using (a) RF, (b) SVM for the year 2020

Table 9. The forest coverage area in Peninsular Malaysia for 2010 and 2020 using SVM algorithm

Year	2010		2020	
Land cover	**Area (ha)**	**Percentage (%)**	**Area (ha)**	**Percentage (%)**
Forest	9,528,728.4	71.77	8,883,286.11	66.92
Water bodies	121,384.26	0.91	155,103.21	1.17
Oil palm	3,084,155.1	23.23	3,364,050.06	25.34
Rubber	-	-	-	-
Built-up areas	109,574.55	0.83	322,155.27	2.43
Others	432,229.95	3.26	549,741.24	4.14
Total	**13,276,072.3**	**100**	**13,274,335.89**	**100**

Conclusion

Images of vast countries like Peninsular Malaysia are inevitably affected by cloud cover. Land cover also changes throughout the seasons, making it impractical to rely on a single date or monthly composite images for mapping forest coverage. The process of cloud masking and filling is essential to obtain cloud-free composite images for producing high-quality forest coverage maps. Therefore, this method is more accessible and feasible when conducted in GEE, as it doesn't require substantial costs and extensive time. The study's finding shows that the RF algorithm performs better compared to the SVM in large scale remote sensing image classification in the overall land use classification and in classify forest cover. Overall, the forest cover saw a reduction in 2020 compared to 2010. Malaysia must take the issue of deforestation and land use change seriously and with concern, as it is currently occurring at an alarming rate. These changes are detrimental to both the environment and human well-being. Consequently, proper management of forest development must be reassessed to ensure the needs of the present and future generations are met without compromising the environment.

References

Alper Akar. (2021). Improving the accuracy of random forest-based land-use classification using fused images and digital surface models produced via different interpolation methods. *Concurrency and Computation: Practice and Experience, 34*(6), e6787.

Amani, M., Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mohammad Javad Mirzadeh, S., White, L., Banks, S., Montgomery, J., & Hopkinson, C. (2019). Canadian Wetland Inventory using Google Earth Engine: The First Map and Preliminary Results. *Remote Sensing, 11*(7), 842.

Asner, G. P. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters, 4*(3), 034009.

Basten, K. (2016). *Classifying Landsat Terrain Images via Random Forests* (Bachelor dissertation, Radboud University).

Berberoglu, S., & Akin, A. (2009). Assessing different remote sensing techniques to detect land use-cover changes in the Eastern Mediterranean. *International Journal of Applied Earth Observation and Geoinformation, 11*(1), 46–53.

Cakir, H. I., Khorram, S., & Nelson, S. A. C. (2006). Correspondence analysis for detecting land cover change. *Remote Sensing of Environment*, *102*(3-4), 306–317.

Cánovas-García, F., Alonso-Sarría, F., Gomariz-Castillo, F., & Oñate-Valdivieso, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Computer Geosciences*, *103*, 1–11.

Czaplewski, R. L. 1994. Variance approximations for assessments of classification accuracy. Res. Pap. RM-316. U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO (29 pp.).

Desclée, B., Bogaert, P., & Defourny, P. (2006). Forest change detection by statistical object-based method. *Remote Sensing of Environment, 102*(1-2), 1–11.

Fadli, A. H., Kosugo, A., Ichii, K., & Ramli, R. (2019). Satellite-based monitoring of forest cover change in Indonesia using Google Earth Engine from 2000 to 2016. Journal of Physics: Conference Series volume 1317, The 3rd International Conference on Mathematics, Sciences, Education and Technology. Padang, Indonesia, 4-5 October.

FAO. (2010). Global Forest Resources Assessment 2010. Rome.

Fleiss, J. L., Levin, B., & Paik, M.C. 2013. *Statistical methods for rates and proportions, Third edition*. John Wiley & Sons.

Forkuo, E. K., & Frimpong, A. (2012). Analysis of forest cover change detection. *International Journal of Remote Sensing Applications*, *2*(4), 82–92.

Fortin, J. A., Cardille, J. A., & Perez, E. (2020). Multi-sensor detection of forest-cover change across 45 years. *Remote Sensing of Environment,* 238, 111266.

Gautam, P., Shivakoti, G., & Webb, E. (2004). Forest cover change, physiography, local economy and institutions in a Mountain Watershed in Nepal. *Environmental Management, 33*(1), 48–61.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forest for land cover classification. *Pattern Recognition Letters, 27*(4), 294-300.

Gohar Ghazaryan. 2015. *Analysis of temporal and spatial variation of forest* (Degree of Master dissertation, University of Muenster).

Gorelick, N., Hanche, M., Dixon, M., Ilyushchenko, S., Thaub, D., & Moore, R. (2017). Google Earth Engine - Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27. doi: 10.1016/j.rse.2017.06.031

Hamdan, O., Misman, M. A., & Kassim, A. R. (2017). Synergetic of PALSAR-2 and Sentinel-1A SAR Polar-imetry for retrieving aboveground biomass in dipterocarp forest of Malaysia. *Applied Sciences, 7*(7),675.

Han, X., Pan, J., & Devlin, A. T. (2018). Remote sensing study of wetlands in the Pearl River Delta during 1995–2015 with the support vector machine method. *Frontiers of Earth Science*, *12*(3), 521-531

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science, 342*(6160), 850-853.

Hansen, M. C., Potapov, P. V., Pickens, A. H., Tyukavina, A., Hernandez-Serna, A., Zalles, V., Turubanova, S., Kommareddy, I., Stehman, S. V., & Song, X. P. (2022). Global land use extent and dispersion within natural land cover using Landsat data. *Environmental Research Letters*, *17*(3), 034050.

Huang, C., Goward, S. N, Masek, J. G, Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, *114*(1), 183–198.

Huang, C., Song, K., Kim, S., Townshend, J. R. G., Davis, P., Masek, J. G., & Goward, S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, *112*(3), 970–985.

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: A case study of central Shandong. *International Journal Remote Sensing, 39*(23), 8703–8723.

Jin, Y., Sung, S., Lee, D. K., Biging, G. S., & Jeong, S. (2016). Mapping deforestation in North Korea using Phenology-Based Multi-Index and Random Forest. *Remote Sensing, 8*(12), 997. doi:10.3390/rs8120997

Kanniah, K. D., Nazarin Ezzaty, M. N., & Tuong, T. V. (2016). Forest cover mapping in Iskandar Malaysia using satellite data. The international archives of the photogrammetry, remote sensing and spatial information sciences, volume xlii-4/W1, International Conference on Geomatic and Geospatial Technology (GGT). Kuala Lumpur, Malaysia, 3–5 Oktober.

Kelley, L. C., Pitcher, L., & Bacon, C. (2018). Using Google Earth Engine to map complex shade-grown coffee landscapes in Northern Nicaragua. *Remote Sensing*, *10*(6), 952.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lee, J., & Kang, M. (2015). Geospatial big data: Challenges and opportunities. *Big Data Research,* 2(2), 74-81. doi: 10.1016/j.bdr.2015.01.003

Li, X., Chen, W., Cheng, X. & Wang, L. (2016). A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 Stereo satellite imagery. *Remote Sensing, 8*(6), 514.

Lowe, B., & Kulkarni, A. (2015). Multispectral image analysis using Random Forest. *International Journal on Soft Computing (IJSC), 6*(1), 1-14.

Maxwell, A. E., Strager, M. P., Warner, T. A., Ramezan, C. A., Morgan, A. N., & Pauley, C. E. (2019). Large-area, high spatial resolution land cover mapping using Random Forests, GEOBIA and NAIP orthophotography: Findings and recommendations. *Remote Sensing, 11*(12), 1409.

Mayaux, P., Eva, H., Gallego, J., Strahler, A. H., Herold, M., Agrawal, S., et al. (2006). Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing, 44*(7), 1728–1739.

Mellor, A., Haywood, S. J., & Wilkes, P. (2012). Forest classification using Random Forests with multisource remote sensing and ancillary GIS data. 16th Australian Remote Sensing and Photogrammetry Conference. Australia, Melbourne, 27-28 August

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing,7*(7), 8489–8515.

Monserud, R. A. & Leemans, R. 1992. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling, 62*(4), 275–293.

Nazarin Ezzaty, M. N. & Kanniah, K. D. (2018). Optical and radar remote sensing data for forest cover mapping in Peninsular Malaysia. *Singapore Journal of Tropical Geography, 40*(2), 272-290. doi:10.1111/sjtg.12274

Nazarin Ezzaty, M. N. (2015). Forest cover change in Peninsular Malaysia using satellite remote sensing data (Master dissertation, Universiti Teknologi Malaysia).

Negassa, M. D., Mallie, D. T., & Gemeda, D. O. (2020). Forest cover change detection using Geographic Information Systems and remote sensing techniques: A spatio temporal study on Komto Protected Forest Priority Area, East Wollega Zone, Ethiopia. *Environment System Research, 9*(1) https://doi.org/10.1186/s40068-020-0163-z

Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation, 34*, 136–46.

Olofsson, P., Kuemmerle, T., Griffiths, P., Knorn, J., Baccini, A., Gancz, V., Blujdea, V., Houghton, R. A., Abrudan, I. V., & Woodcock, C. E. (2011). Carbon implications of forest restitution in post-socialist Romania. *Environmental Research Letters, 6*(4), 045202.

Osei, J. D., Andam-Akorful, S. A., & Osei jnr, E. M. (2019). Long term monitoring of Ghana’s Forest Reserves. *Preprints*, 2019090016

Pelletiera, C., Valeroa, S., Inglada, J., Championb, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment, 187*, 156-168.

Potapov, P., Hansen, M. C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harriz, N., Song, X. P., Bagget, A., Kommareddy, I., & Kommareddy, A. (2022). The global 2000–2020 land cover and land use change dataset derived from the Landsat archive: first results. *Frontiers in Remote Sensing 3*.

Prabhat, K. R. (2013). Forest and land use mapping using remote sensing & Geographical Information System: A case study on model system. *Environmental Skeptics and Critics, 2*(3), 97-107.

Pretorius, L., Brown, L. R., Bredenkamp, G. J. & van Huyssteen, C. W. (2016). The ecology and classification of wetland vegetation in the Maputaland Coastal Plain, South Africa. *Phytocoenologia*, *46*(2), 125-139.

Qamer, F. M., Shehzad, K., Abbas, S., Murthy, M. S. R, Xi, C., Gilani, H., & Bajracharya, B. (2016). Mapping deforestation and forest degradation patterns in Western Himalaya, Pakistan. *Remote Sensing*, *8*(5), 385. doi:10.3390/rs8050385

Qamer, F. M., Shehzad, K., Abbas, S., Murthy, M. S. R, Xi, C., Gilani, H., & Bajracharya, B. 2016. Mapping deforestation and forest degradation patterns in Western Himalaya, Pakistan. *Remote Sensing*, *8*(5), 385. doi:10.3390/rs8050385

Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing, 7*(1), 153-168.

Rathinagiri, S., Mirunalini, M., Raj, J., Pugalanthi, V., Ravichandran, N., & Anand, V. D. (2010). Remote sensing and GIS based forest cover change detection study in Kalrayan Hills, Tamil Nadu. *Journal of Environmental Biology, 31*(5), 737–747.

Rio, M. F, Joko, N. R. & Ratna, H. (2016). Analisa perubahan penutupan lahan pada kawasan hutan lindung Gunung Naning Kabupaten Sekadau provinsi Kalimantan Barat. *Jurnal Hutan Lestari, 4*(4), 520-526.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Saralioglue, E., & Vatandaslar, C. (2022). Land use/land cover classification with Landsat and Landsat-9 satellite images: A comparative analysis between forest and agriculture dominated landscapes using different machine learning methods. *Acta Geodaetica et Geophysica*, *57*, 695-716.

Sau, A. A. W. T. (2013). *Quantifying forest degradation and deforestation using Geographic Information System (GIS): A case study in the three provinces, South Kalimantan, East Kalimantan and South-east Sulawesi, Indonesia* (Master dissertation, University of Canterbuy).

Schmidt, J., Marques, M. R., Botti, S., & Marques, M. A. (2019). Recent advances and applications of machine learning in solid-state materials science. *NPJ Computational Materials* 5(83), 1–36.

Shijuan, C., Curtis, E., Woodcock, E. L., Bullock, P. A, Paata, T., Siqi, P., & Pontus, O. (2021). Monitoring temperate forest degradation on Google Earth Engine using Landsat time series analysis. *Remote Sensing of Environment*, 265, 112648.

Sirén, A. H., & Brondizio, E. S. (2009). Detecting subtle land use change in tropical forests. *Applied Geography*, *29*(2), 201–211.

Solomon, N., Hishe, H., Annang, T., Pabi, O., Asante, I. K., & Birhane, E. (2018). Forest cover change, key drivers and community perception in Wujig Mahgo Waren Forest of Northern Ethiopia. *Land*, *7*(1), 32 doi:10.3390/land7010032

Sugumaran, R. (2001). Forest land cover classification using statistical and Artificial Neural Network approaches applied to IRS LISS−III sensor. *Geocarto International*, 16, 39–44.

Svoboda, J., Stych, P., Lastovicka, J., Paluba, D., & Kobliuk, N. (2022). Random Forest classification of land use, land-use change and forestry (LULUCF) using Sentinel-2 data - A case study of Czechia. *Remote sensing*, *14*(5), 1189.

Tamaluddin, S, Arif, D, Irwan, S. B., & Kuswibowo, N. (2012). Pemanfaatan citra satelit dalam mengidentifikasi perubahan penutupan lahan: Studi kasus Hutan Lindung Register 22 Way Waya Lampung Tengah. *Globe, 14*(2), 146–156

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 J. Photogramm. Remote Sensing, 164,* 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001

Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., Yadav, K., & Huete, A. (2018). A 30-m Landsat-derived cropland extent product of Australia and China using Random Forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sensing, 144,* 325–340.

Vahid, N., Azade, D., Fardin, M., Seyed, M. M. S., & Stelian, A. B. (2022). Land use and land cover mapping using Sentinel-2, Landsat-8 satellite images and Google Earth Engine: A comparison of two composition methods. *Remote Sensing*, *14*(9), 1977.

Venkatappa, M., Sasaki, N., Anantsuksomsri, S. & Smith, B. (2020). Applications of the Google Earth Engine and phenology-based threshold classification method for mapping forest cover and carbon stock changes in Siem Reap Province, Cambodia. *Remote Sensing*, *12*(18), 3110. doi:10.3390/rs12183110

Wagner, J. E., & Stehman, S. V. (2015). Optimizing Sample size allocation to strata for estimating area and map accuracy. *Remote Sensing of Environment*, *168*, 126–133. doi:10.1016/j.rse. 2015.06.027

Wan Abdul Hamid Shukri B. Wan Abd Rahman. (2016). Comparison results of forest cover mapping of Peninsular Malaysia using geospatial technology. *IOP Conf. Series Earth and Environmental Science, 37*(1), 012027 doi:10.1088/1755-1315/37/1/012027

Yang, C., Yu, M., Hu, F., Jiang, Y., & Li, Y. (2017). Utilizing cloud computing to address big geospatial data challenges. *Computers, environment and urban systems, 61*, 120-128.

Yang, H., & Yunfeng, Hu. (2020). Detecting forest disturbance and recovery in Primorsky Krai, Russia. *Remote Sensing*, *12*(1), 129; doi:10.3390/rs12010129

Yantao, G., Zhang, X., Long, T., Jiao, W., He, G., Yin, R., & Dong, Y. Y. (2019). China forest extraction based on Google Earth Engine. The international archives of the photogrammetry, remote sensing and spatial information sciences, volume xlii-3/w10, International Conference on Geomatics in the Big Data Era (ICGBD), Guilin, Guangxi. China, 15–17 November.

Zhang, X., Long, T., He, G., & Guo, Y. (2019). Gobal Forest Cover Mapping using Landsat and Google Earth Engine cloud computing. 8th International conference on Agro-Geoinformatics (Agro-Geoinformatics).