Comparison of Landsat 8, Sentinel-2 and spectral indices combinations for Google Earth Engine-based land use mapping in the Johor River Basin, Malaysia

Zeng Ju[1], Mou Leong Tan[1], Narimah Samat[1], Chun Kiat Chang[2]

[1]GeoInformatic Unit, Geography Section, School of Humanities, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia;
[2]River Engineering and Urban Drainage Research Centre (REDAC), Universiti Sains Malaysia, Engineering Campus, Nibong Tebal 14300, Penang, Malaysia

Correspondence: Mou Leong Tan (email: mouleong@usm.my / mouleong@gmail.com)

Abstract

Accurate land use information is the basis for scientific research related to carbon cycle analysis, hydro-climatic modelling, soil degradation assessment, etc. It is also an indispensable basic information for local land management departments in land use planning and management. With development in big data and internet network, Google Earth Engine (GEE), a cloud-based computing platform allows users to perform satellite images processing more efficiently. This study aims to improve the land use mapping in a tropical region based on the GEE platform. Seven satellite images and indices combinations include Landsat 8 (C1), Sentinel-2 (C2), Landsat 8+Sentinel-2 (C3), Landsat 8+Indices (C4), Sentinel-2+Indices (C5), Landsat 8+Sentinel-2+Indices (C6), Normalized Difference Vegetation Index (NDVI)+Normalized Difference Water Index (NDWI)+Enhanced Vegetation Index (EVI)+Elevation (C7) were developed to evaluate the best combination for land use mapping in the Johor River Basin (JRB), Malaysia. The Random Forest (RF) algorithm was used to classify the land use land cover (LULC) with 222 training samples and 78 verification samples obtained through the Google Earth Pro higher resolution satellite images and field samplings. The results show that the overall accuracies of all the seven combinations are mostly more than 75%, ranging from 72% (C1) to 86% (C6). The findings show that adding of additional indices information before the land use classification helps to increase the overall accuracy significantly. For instance, the overall accuracy of the C6 combination is 14% higher than that of the Landsat 8 image solely. This study can act as a reference to effectively improve the land use mapping in cloud-prone tropical regions.

Keywords: Google Earth Engine, Image combination, Johor River Basin, Land Use, Malaysia, Random Forest

Introduction

Satellite imagery is one of the important data sources for producing land use maps to analyse the changes over time (Liu, 2015; Miettinen et al., 2016). Optical remote sensing data such as Landsat is the most commonly used data to analyse land use land cover (LULC) changes. The emergence of a new-generation of multispectral sensors in the Landsat 8 and Sentinel-2 provides opportunities for multi-temporal analysis of locations with frequent cloud cover problems (Mandanici & Bitelli, 2016).

Data fusion technology refers to the processing of data and information from different sources to obtain new information for decision-making (Pohl & van Genderen, 1998). In a given time, compared to single-band images, fusion products usually contain more information. Therefore, image data fusion technology or combination technology can be used to overcome or partially overcome the spatial-temporal defects of specific sensors (Poortinga et al., 2019).

Traditionally, users need to download satellite data through a data portal, perform data processing using a standalone software and then deliver to the end-users. With the development of Google Earth Engine (GEE), satellite data and even processing can be done easily through the online platform. The emergence of the GEE cloud platform provides a good support for solving the problem of how remote sensing users can efficiently select data of interest from massive amounts of remote sensing data. GEE includes a multi-source database, high-performance computing capabilities, as well as its online application programming interface and network-based interactive development environment for developers, so that it has powerful simulation and calculation results capabilities (Amani et al., 2020; Gorelick et al., 2017).Many scholars have used the advantages of GEE to carry out a lot of scientific research on remote sensing and geography from small urban areas to national scales (Li et al., 2020; Oliphant et al., 2019; Sidhu et al., 2018).

To analyse image combination on the effect of LULC classification, some scholars used GEE to analyse the aggregation of single sensor data at different time intervals. Phan et al. (2020) utilized the Landsat 8 images to generate eight different image compositions as input images to produce a land use map of Mongolia. They found median aggregation method can reduce data volume and get high accuracy. Praticò et al. (2021) applied Sentinel-2 data with different combination strategies to implement classification for a study area in Mediterranean Forest Habitats. The highest accuracy obtained from the Random Forest classifier, additional of summer imagery and three vegetation indices band, Liu et al. (2020) integrated the 2016-2018 time series data from Landsat and Sentinel-2 on the GEE platform to map China's annual cropping intensity, using an algorithm based on pixels and phenology, with the overall accuracy (OA) and the kappa coefficient of 93% and 0.84, respectively. In the United Kingdom, Carrasco et al., (2019) used Landsat 8, Sentinel-2, and Sentinel-1 data to generate data sets at different time intervals within a year to create land cover maps. They found that the combined data sets (Sentinel1 +Sentinel-2 or Sentinel1+Sentinel-2+Landsat 8) performed better than single sensor data set.

From the literature, it is found that selecting an appropriate image aggregation strategy to generate an image set according to the main land cover type and vegetation growth patterns in the study area is one of the key factors affecting the accuracy of LULC mapping. However, such assessment is still limited in tropical regions, including Malaysia. This study aims to improve LULC mapping in tropical basins by identifying the best satellite image fusion combination through the GEE platform. The specific objective is to assess the capability of seven different combinations of optical sensor images (Landsat 8 and Sentinel-2) and three indices through data fusion technology to generate the LULC map of the Johor River Basin (JRB) in 2019. The basin

is an important freshwater source to Malaysia and Singapore (Tan et al. 2015), was selected as the study site.

Method and study area

Study Area

The climate of the JRB is characterized by humid, hot and rainy climate all year round. The national temperature ranges from 23°C to 32°C (Shaharum et al., 2020; Tan et al., 2019). It is located between 1°30′-2°10′N north latitude and 103°20′-104°10′E east longitude. Major land use types of JRB are oil-palm and forests. The location map of the study area is shown in Figure 1, which shows the Landsat8 image of the basin on May 27,2019.
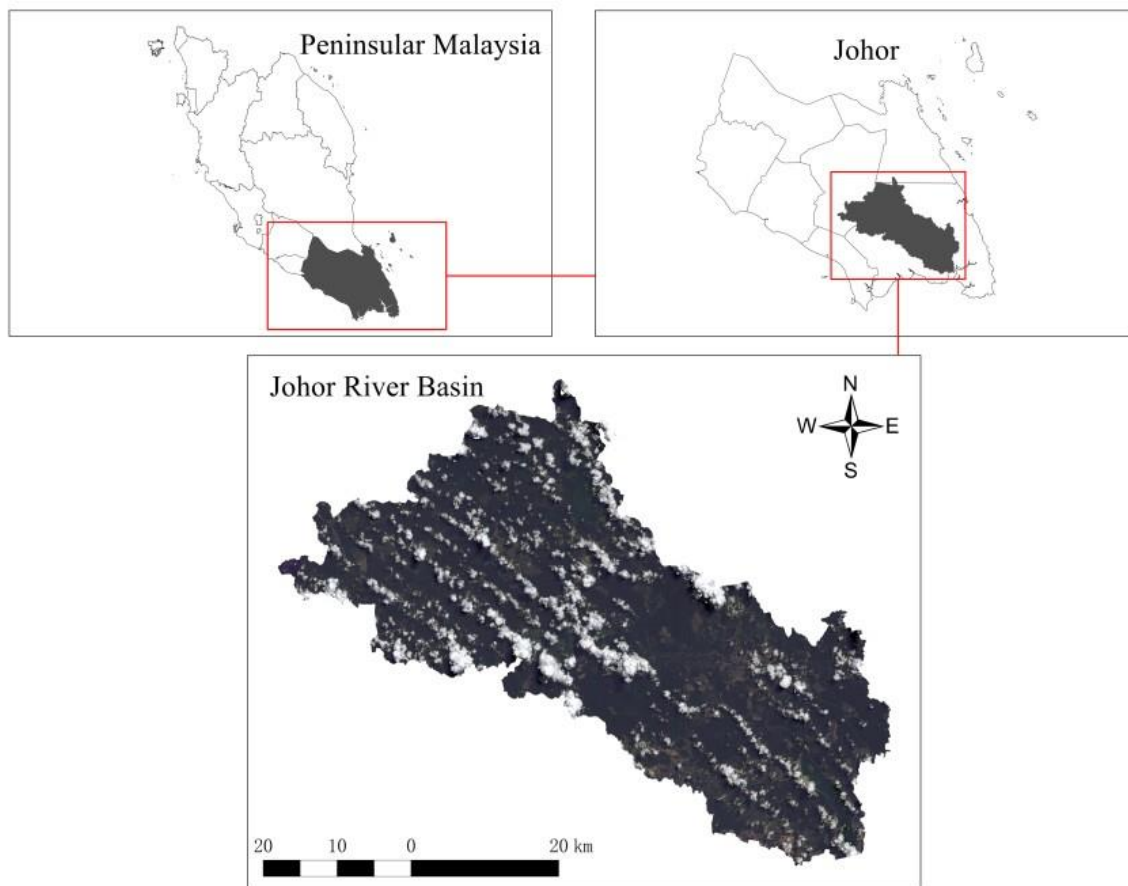

Figure 1. The Landsat 8 image of JRB on May 27, 2019 obtained through the GEE platform.

Methodology

The land use map generation process includes three main stages: data source selection, classification method selection, and accuracy evaluation. Suitable Landsat 8 and Sentinel-2 images were selected for pre-processing and analysis. The median index is derived from the time series

image data, which is used to replace the time series image for classification. Secondly, this research adopted a pixel-based supervised classification, the Random Forest (RF) was used to classify the selected images. Training samples were selected based on the GEE map visualization interface and field samplings to form the final training and verification samples. Finally, the classification results of the seven data combinations were evaluated. The accuracy evaluation methods include three popular indicators, the overall accuracy (OA), producer accuracy (PA), and consumer accuracy (CA) that extracted from the confusion matrix report. Besides that, the kappa coefficient of statistical methods (Gyamfi-Ampadu et al., 2020) was also used in this study. The specific research process is shown in Figure 2.
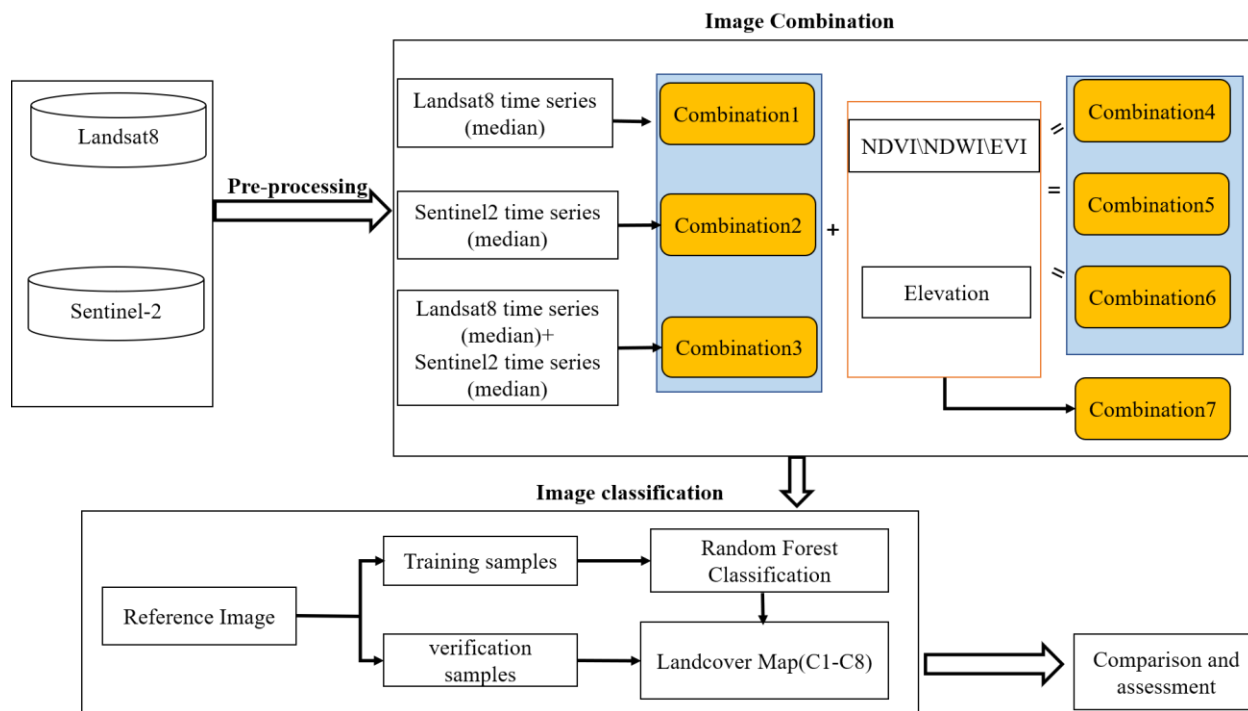

Figure 2. Research flow chart.

Satellite data

We selected the Landsat 8 and Sentinel-2 satellite data in 2019. These satellite images are freely extracted and accessed through the GEE code editor platform as shown in Figure 3. According to GEE's internal calculation and data management mechanism, users can handle different data resolution matching by limiting the output resolution of the analysis results. In this study, we have resampled the Sentinel 2 data to 30 meter to match with the resolution of the Landsat 8 data. The image attribute values are shown in Table 1. Through a preliminary cloud cover analysis, it is found that the study area has serious cloud cover and poor image quality. Combining images with a short time range result in many missing values. As a result, this increases the probability of cloud-covered polluted data in the map (Foga et al., 2017). In our study, the time interval is set to one year and the median is used to combine the images. In short, multiple images for one year are aggregated into one image.

Table 1. Parameters of Landsat 8 and Sentinel-2 images required for mapping.

Sensor	Band		Pixel size (m)	Wavelength (µ m)
Landsat 8	2	Blue	30	0.45-0.51
OLI	3	Green	30	0.53-0.59
	4	Red	30	0.64-0.67
	5	Near-infrared (NIR)	30	0.85-0.88
	6	Short-wave infrared 1(SWIR1)	30	1.57-1.65
	7	Short-wave infrared 2(SWIR2)	30	2.11-2.29
	11	Thermal infrared 2	30	11.50-12.51
Sentinel-2 MSI	2	Blue	10	0.49
	3	Green	10	0.56
	4	Red	10	0.665
	5	Red Edge1	20	0.704
	8	Near-infrared (NIR)	10	0.842
	11	Short-wave infrared 1(SWIR1)	20	1.61
	12	Short-wave infrared 2(SWIR2)	20	2.19
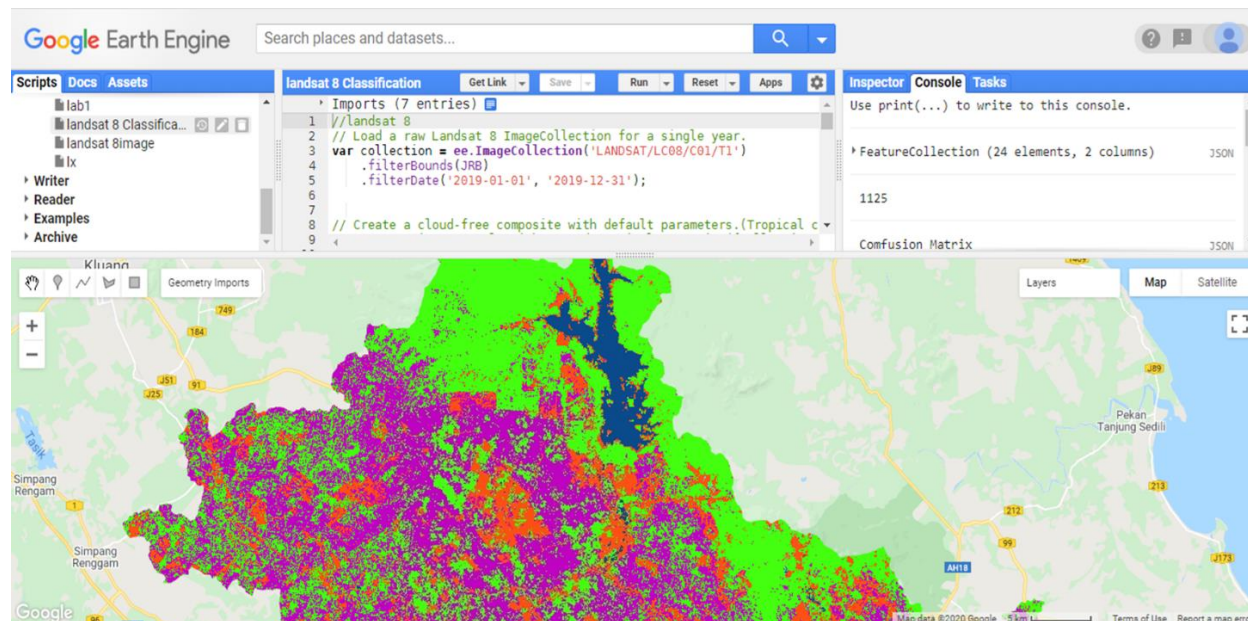
Figure 3. Google Earth Engine Javascript API

a. Landsat 8

Landsat 8 was developed by NASA and USGS in 2013. The satellite carries two sensors with an operational land imager (OLI) and a thermal infrared sensor (TIRS) (Roy et al., 2014). Approximately every two weeks, the entire surface of the earth is taken with a resolution of 30 meter. Products that can be directly called through the GEE platform include Surface Reflectance, Top of Atmosphere, and raw images. In our research, we selected seven bands of Landsat 8, including blue (B2), green (B3), red (B4), near-infrared (B5), shortwave-infrared 1 (B6), short-

infrared 2 (B7) and Thermal infrared 2 (B11). To create a simple cloudless Landsat 8 composite image, we used the simple composite method to select a subset of the scene at each location. Then, this image was converted to Top-of-Atmosphere Reflectance (TOA) reflectance by applying a simple cloud score and median of the least cloudy pixels.

b.   Sentinel-2

Sentinel-2 is a multi-spectral imaging satellite launched by the European Space Agency (ESA) in 2015. It is divided into two satellites 2A and 2B with a ground resolution of 10m, 20m, and 60m (van der Meer et al., 2014). After the two satellites enter the operational state at the same time, under complementary circumstances, the revisit period will be shortened from 10 days to 5 days. The GEE platform provides two types of products: Surface Reflectance and TOA. In this study, we used TOA, Multi-Spectral Instrument and Level-1C image. Sentinel-2 has a variety of spatial resolutions, this article selects four bands with 10m resolution, including blue (B2), green (B3), red (B4), and near-infrared (B8), and the resolution was selected as three bands of 20m, including red edge (B5), short-wave infrared 1 (B11), and short-wave infrared 2 (B12). The attribute values of Sentinel-2 images are shown in Table 1.

Image Pre-processing

In the data pre-processing stage, JRB was set as the area of interest, considering the images from 1 January 2019 to 31 December 2019. A total of 56 Landsat 8 and 142 Sentinel-2 raw images were collected and converted into TOA images using the simple combination operations. The images values were obtained by calculating the median. Finally, a cloudless image was obtained for further analysis. Figure 4 shows the true color composite images generated and displayed on the GEE platform.

Landsat8

Sentinel2
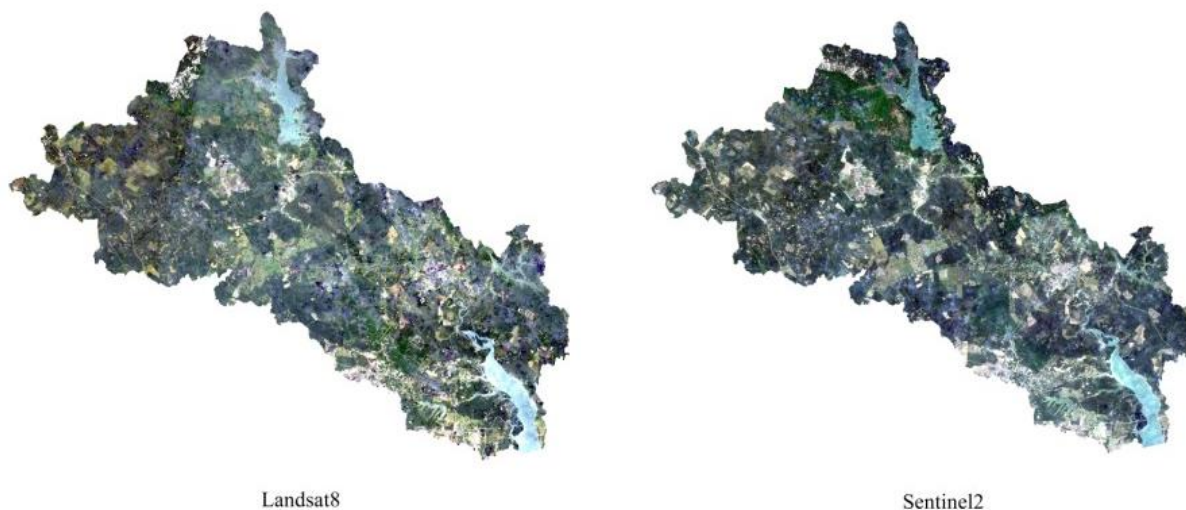
Figure 4. cloud-free composite images.

Additional indices

In this study, three spectral indices, Normalized Difference Vegetation Index (NDVI) (Aburas et al., 2015), Normalized Difference Water Index (NDWI) (Shaharum et al., 2020) and Enhanced Vegetation Index (EVI) (Poortinga et al., 2019) were calculated from Landsat 8 and Sentinel-2 combined images. NDVI, NDWI and EVI are widely used in land use classification (Nguyen et al., 2020; Scheffler et al., 2020; Vijith & Dodge-Wan, 2020). The specific formulas are shown in Table 2.

a.   NDVI

NDVI is often used to detect vegetation growth conditions and vegetation cover changes. The NDVI value is calculated by combining the red band and the NIR band in the image. The maximum value of NDVI value is 1, and the minimum value is -1 (Aburas et al., 2015). Negative values indicate clouds, water and snow that highly reflect visible light. Zero indicates rocks or bare soil with approximately equal NIR and RED, while positive values indicate vegetation coverage. The increase in coverage rate increases the NDVI value.

b.   NDWI

NDWI is a commonly used index for extracting water information in images. When extracting the open water body in the image, the NDWI value was calculated by the NIR and green bands construction equation developed by McFeeters (1996) to make the extraction effect better. However, when NDWI extracts more water bodies in the context of buildings, such as water bodies in cities, the extracted water body information will increase, which deviates greatly from the actual water body information (Xu, 2006).

c.   EVI

EVI is a vegetation index obtained after optimizing NDVI using the soil adjustment factor "L". It improves vegetation monitoring by canceling the coupling of canopy background signals and reducing atmospheric influence. It can be calculated using blue, red, and NIR bands. In areas with high biomass, EVI exhibits higher sensitivity to topographical conditions (Matsushita et al., 2007).

Table 2. Three Spectral Indices

Indices	Formula
NDVI	NDVI = (NIR – RED)/(NIR+RED)
NDWI	NDWI = (NIR – SWIR)/(NIR+SWIR1)
EVI	EVI = 2.5* (NIR – RED)/ (NIR+6.0 * RED – 7.5 * BLUE+1.0)

Training and validation sample data

The LULC types of the JRB were divided into six classes: forest, urban, water, oil-palm, bare-land, and forest-shrubs (Table 3). Forest-shrubs mainly includes degraded forests and forest

fragments located near human settlements. For each class, 30-50 training samples are selected to be suitable for land cover classification (Wulder et al., 2006). All samples are created from the GEE platform using a point format and manual visual interpretation (de Sousa et al., 2020; Praticò et al., 2021). In the study, 222 samples were selected to train the classifier, and 78 samples were selected to verify the classification map.

Table 3. Land cover type and sample description.

Code	Class	Training Samples(point)	Test Samples(point)
0	Forest	50	20
1	Urban	50	20
2	Water	50	20
3	Oil-palm	50	20
4	Bare-land	31	12
5	Forest-shrub	50	20
	Total	281	112

Classification and accuracy assessment

The RF algorithm is an integrated multi-decision tree classification method (Jin et al., 2018). The decision tree model in the RF is constructed by randomly extracting sample data and feature quantities, and the remaining data Compose external data for error testing (De Alban et al., 2018). Because the RF algorithm has the characteristics of high classification accuracy and good anti-noise performance, it is widely used in remote sensing image classification (Belgiu & Drăguţ, 2016). In this study, the number of trees in the RF classifier is set to 30 (Shaharum et al., 2020). According to previous research, it is a common practice to use OA, PA and CA to evaluate the accuracy (De Alban et al., 2018; Johnson & Iizuka, 2016). The size of these indicators directly reflects the classification accuracy. The accuracy of classification results can also be evaluated based on Kappa statistics (Jin et al., 2018; Li et al., 2020; Phan et al., 2020). In our research, both types of evaluation methods are used at the same time (Carrasco et al., 2019; Forkuor et al., 2018; Phan et al., 2020).

Seven combinations

As mentioned earlier, the image data used for classification were divided into seven combinations. C1 is formed by the spectral band of Landsat 8, C2 is formed by the spectral band of Sentinel-2, C3 is formed by the spectra of Landsat 8 and Sentinel-2 images, and only additional indices are used in C7. To test the influence of the index, additional indexes were added to C1, C2, and C3 to obtain C4, C5, and C6. Table 4 shows the details of different data combinations.

Table 4. Seven different image combination information.

Symbol	Name	Description	Bands
C1	Landsat 8	Use the median method to combine all Landsat 8 images in 2019 to form a combination.	7
C2	Sentinel-2	Use the median method to combine all Sentinel-2 images in 2019 to form a combination.	7
C3	Landsat 8 + Sentinel-2	the bands in the C2 image were added to the C1 image according to the band combination.	14
C4	Landsat 8 +indices	Four indices calculated from C1 were added to C1 as bands to form an image.	11
C5	Sentinel-2+indices	Four indices that were calculated from C2 were added to C2 as bands to form an image.	11
C6	Landsat 8+Sentinel-2+indices	Add all the bands of the C5 image were added to the C4 to form an image.	22
C7	NDVI\NDWI\EVI\Elevation	Combine the four indices calculated from C1 and C2 according to the band combination method to obtain images.	8

Results

Visualization comparison

i. Classification map comparison

The LULC classification results with a resolution of 30 meter generated by the seven combinations are shown in Figure 5. Main LULC of the basin are oil-palm and forest. Forest is dominated in the northern part of the basin, while urban land mainly concentrated on both sides of the Johor River. Forest-shrub is mainly found around the town. The land cover type patterns of all categories in the seven maps are basically the same, but the recognition degree of each category is slightly different.
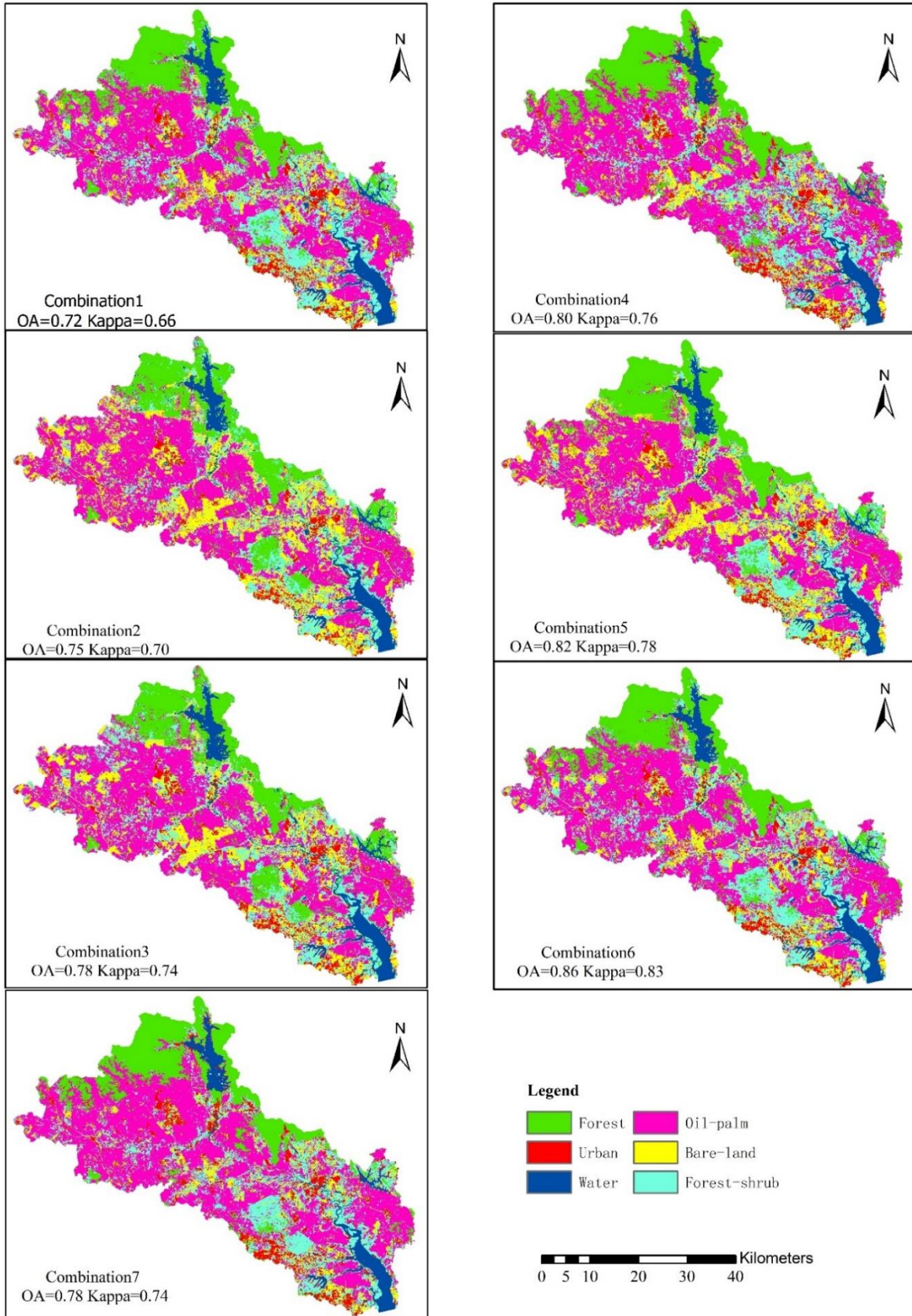
Figure 5. land cover classification maps

ii. Specific area result analysis

The highest classification accuracy of 0.86 is obtained by the C6 combination. During the field trip, the samples collected in the southern part of the JRB were 4 urban, 2 forest-shrub and 1 oil-palm. Comparing the classification map with the field trip data, it is found that the classification effect of continuous and large-area distribution features is better, such as water bodies and large area oil-palm. However, some low vegetation types are wrongly classified as bare-land as shown in Figure 6.
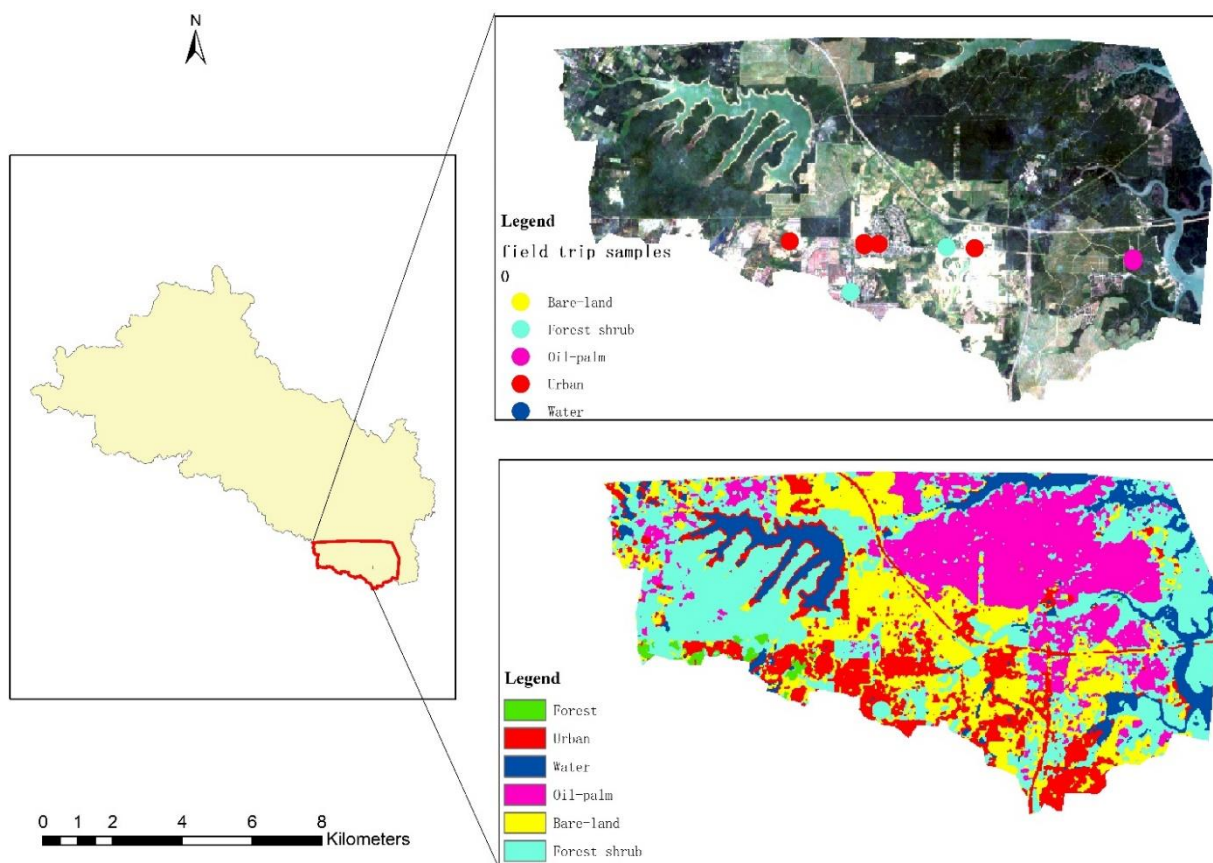
Figure 6. Comparison of sampling points with a classified land use map.

Accuracy Assessment

a. Comparison of data sources

The accuracy of the seven combinations is listed in Table 5. The overall accuracy difference between Landsat 8 (C1) and Sentinel-2(C2) is 3%. When the Landsat 8 and Sentinel-2 spectral bands (C3) are combined, the classification accuracy increased to 78%. The C6 combination has the highest OA and kappa of 86% and 0.83, respectively. However, the PA and CA obtained for different land cover types are quite difference. It can be seen from Table 5 that the seven combinations classified forest and water well, and the generated CA and PA values are between 0.75-1. When distinguishing oil-palm, there are big differences among seven combinations, with the C5, C6, and C7 combinations show a better accuracy. In terms of distinguishing urban, the

accuracy obtained for all combinations is similar. When distinguishing forest-shrub, the C5 and C6 combinations obtain a higher accuracy. Interestingly, all combinations have low accuracy in distinguishing bare-land.

Table 5. Accuracy assessment of land cover classification.

Dataset		Forest	Urban	Water	Oil-palm	Bare-Land	Forest-shrub	OA	Kappa
Combination1 (L8)	CA	0.95	0.77	0.75	0.63	0.54	0.61	0.72	0.66
	PA	0.95	0.65	0.90	0.63	0.58	0.55		
Combination2(S2)	CA	1	0.78	0.91	0.92	0.41	0.59	0.75	0.70
	PA	0.75	0.70	1	0.58	0.58	0.80		
Combination3(L8+S2)	CA	1	0.73	1	0.86	0.44	0.68	0.78	0.74
	PA	0.85	0.80	1	0.63	0.58	0.75		
Combination4(L8+Indices)	CA	1	0.80	0.91	0.85	0.57	0.65	0.80	0.76
	PA	1	0.80	0.95	0.58	0.67	0.75		
Combination5(S2+Indices)	CA	1	0.75	0.91	0.93	0.54	0.73	0.82	0.78
	PA	0.95	0.75	1	0.74	0.59	0.80		
Combination6(L8+S2+Indices)	CA	1	0.76	1	0.93	0.64	0.78	0.86	0.83
	PA	0.95	0.80	1	0.74	0.75	0.85		
Combination7(Indices)	CA	1	0.64	0.86	0.93	0.50	0.70	0.78	0.74
	PA	0.95	0.70	0.95	0.74	0.42	0.80		

b. Comparison of band combination

Table 5 shows that adding the additional indices to Landsat 8 can increase the classification accuracy from 75% to 77%. Meanwhile, adding an index band to the Sentinel-2 image able to increase the classification accuracy from 72% to 80%. As the additional indices increases, the accuracy from C3 to C6 increased from 78% to 86%. But if only the index is used for classification, the classification accuracy obtained by using only the spectral band is similar.

Discussion

The finding is consistent with other studies where combined images and indices can achieve higher classification accuracy than a single image (Carrasco et al., 2019; Phan et al., 2020). When making maps of 13 types of land cover types in the United Kingdom (Carrasco et al., 2019), the classification accuracy obtained using L8 and S2 images was 72.8%. When analysing the image combination of a single sensor in different periods (Phan et al., 2020), the combination of Landsat 8 and auxiliary variables obtained the highest classification accuracy of 89.80%. In this study, it can be seen from Table 5 that as the data in the image combination increases, the classification

accuracy gradually increases, and finally C6 obtains the highest OA of 86%. Our results are similar to those analysed in other tropical countries as well (De Alban et al., 2018; Li et al., 2020; Sarzynski et al., 2020). Li et al. (2020) used multiple data sources including Sentinel-2 and Landsat 8 to make LULC maps of the African continent. When all the data are combined into input data sources, the classification results and kappa coefficients are the highest (OA= 0.81, Kappa=0.76). A study in the Tanintharyi region of Myanmar shows that the method combining optical image and radar image has the highest classification accuracy (OA=0.94) (De Alban et al., 2018). A similar conclusion appeared when Sarzynski et al. (2020) drew the oil-palm map of Indonesian, and the combined image produced the highest classification result (OA=84%). However, this article only analyses the impact of optical remote sensing image combination on classification accuracy. The image combination strategy only considers the similar spectral bands of two sensors and four important index bands.

In the process of data combination, with the participation of the spectral index, the classification accuracy fluctuates greatly. When analysing the land cover classification of Mongolia, Phan et al., (2020) added 13 index variables to the image, and the OA increased by 4.1% to 7.7%. Since the main vegetation cover types in this watershed are oil-palm and forest, the additional indices such as NDVI, NDWI, EVI and elevation adding more specific information, so the OA increased by 7.2% to 8.1%. However, the accuracy improvement is very little only when using the combination of index bands to produce LULC maps. In certain cases, the accuracy rate may decrease. Carrasco et al. (2019) found that when the accuracy is lower than usual when only the derived spectral index band is used for classification.

Landsat 8 and Sentinel-2 images have similar spectral bands. Many scholars have compared and analysed the application of LULC based on these two images (Forkuor et al., 2018; Sothe et al., 2017; Topaloğlu et al., 2016). Normally, because Sentinel-2 has a higher spatial resolution, the OA obtained from this image is slightly higher than that of Landsat 8. For example, Topaloğlu et al. (2016) compared the accuracy of Landsat 8 and Sentinel-2 data used in Anatolian side of Istanbul land cover mapping, and found that the accuracy difference obtained is 2.5%. Forkuor et al. (2018) used Landsat 8 and Sentinel-2 to draw LULC maps of Burkina Faso, Africa, and found that the accuracy of image classification using Sentinel-2 images increased by 4%. Our finding is quite consistent with previous studies, where the accuracy of Sentinel 2 is about 3% better than Landsat 8 images in this study. When estimating the forest canopy coverage and leaf area index in the north, Korhonen et al., (2017) found that the red edge band may make Sentinel-2 perform better, but there is no systematic difference inaccuracy. Similar conclusions can be found from the classification results table of this article.

The overall accuracy of the Landsat 8 images (72% – 80%) is slightly lower than that of other tropical studies. For example, Johnson & Lizu (2016) and Shaharum et al. (2020) reported the accuracy of the Landsat 8 Images in Philippines and Johor, Malaysia as 72% to 84% and 82% to 89%, respectively. This may be due to the uneven distribution of the sample data collected during the verification phase of land classification results. In general, quality and quantity of selected samples may affect the accuracy of classification. However, field data collection become difficult due to the COVID-19 lockdown situation. In the absence of field survey data, Li et al. (2020) used higher resolution images or Google earth PRO to obtain training data or directly on GEE maps to select samples. The more field data collected for training samples, the higher accuracy of the satellite imagery classification results. However, when a large area is mapped, the cost of obtaining field data samples will increase. Using polygons to select samples has better sample homogeneity and higher classification accuracy (Tew et al., 2019). However, in the study

area, except for water bodies and oil-palm that are relatively homogeneous, cities, bare land, shrubs, etc. are scattered, and the samples are selected by selecting points, so the classification accuracy is lower compared with other good studies (Tew et al., 2019). Due to the different time of obtaining the images, when visual interpretation is performed, the characteristics of the newly planted oil-palm and the harvested oil-palm and some other low crops are similar in the images. This may cause misjudgments exist in the selection of samples, leading to the classification accuracy of bare land, forest shrubs and cities is low.

Conclusion

Under the GEE platform, the satellite images processing time can be greatly shortened. For example, we used less than 20 minutes to process a total of 56 Landsat 8 and 142 Sentinel-2 images for the Johor River Basin. In addition, the GEE cloud database platform helps users to store decades of data. Users do not need to download the satellite images one by one. Moreover, the cloud platform also provides users with high-performance computing capabilities. However, in tropical regions, due to the high cloud content, cloud removal will cause poor image quality. Therefore, this study aims to evaluate the impact of seven different combinations of Sentinel-2 and Landsat 8 images on LULC mapping.

The research results show that the overall accuracy of each combination is generally high, with the combined images with additional indices had the highest overall accuracy. The quality of the image is greatly affected by cloud cover and shadow. Since the area where the basin is located in a tropical area, the obtained images have a high cloud content. Hence, cloudless images can be obtained through combination with the GEE platform, but the quality of the images is greatly affected. In the follow-up study, we can consider combining the advantages of radar images that are not cloud-free images to further improve the accuracy of image classification.

Considering that the major LULC type of the JRB is oil-palm with the growth cycle of 4-5 years, the LULC type will not change much within one year. This study chooses the image data for one the year of 2019 without considering the growth cycle of other crops. Therefore, during the classification process, the seedling oil-palm was incorrectly classified as bare land, which led to an increase in the area of bare land. Further detailed classification will be made according to the growth pattern of crops. The classification accuracy is affected by the classifier. Future study should compare and analyse the use of other popular classifiers such as SVM in the basin.

Acknowledgement

References

Aburas, M. M., Abdullah, S. H., Ramli, M. F., & Ash'aari, Z. H. (2015). Measuring Land Cover Change in Seremban, Malaysia Using NDVI Index. *Procedia Environmental Sciences, 30*, 238-243.

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., . . . 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*.

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

Carrasco, L., O'Neil, A. W., Morton, R. D., & Rowland, C. S. (2019). Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine. *Remote Sensing, 11*(3), 288.

De Alban, J., Connette, G., Oswald, P., & Webb, E. (2018). Combined Landsat and L-Band SAR Data Improves Land Cover Classification and Change Detection in Dynamic Tropical Landscapes. *Remote Sensing, 10*(2), 306.

de Sousa, C., Fatoyinbo, L., Neigh, C., Boucka, F., Angoue, V., & Larsen, T. (2020). Cloud-computing and machine learning in support of country-level land cover and ecosystem extent mapping in Liberia and Gabon. *PLoS One, 15*(1), e0227438.

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Beckmann, T., . . . Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment, 194*, 379-390.

Forkuor, G., Dimobe, K., Serme, I., & Tondoh, J. E. (2018). Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience & Remote Sensing, 55*(3), 331-354.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27.

Gyamfi-Ampadu, E., Gebreslasie, M., & Mendoza-Ponce, A. (2020). Mapping natural forest cover using satellite imagery of Nkandla forest reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment, 18*, 100302.

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 of Remote Sensing, 39*(23), 8703-8723.

Johnson, B. A., & Iizuka, K. (2016). Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. *Applied Geography, 67*, 140-149.

Korhonen, L., Hadi, Packalen, P., & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment, 195*, 259-274.

Li, Q., Qiu, C., Ma, L., Schmitt, M., & Zhu, X. (2020). Mapping the Land Cover of Africa at 10 m Resolution from Multi-Source Remote Sensing Data with Google Earth Engine. *Remote Sensing, 12*(4), 602.

Liu, L., Xiao, X., Qin, Y., Wang, J., Xu, X., Hu, Y., & Qiao, Z. (2020). Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine. *Remote Sensing of Environment, 239*, 111624.

Liu, P. (2015). A survey of remote-sensing big data. *Frontiers in Environmental Science, 3*, 45.

Mandanici, E., & Bitelli, G. (2016). Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use. *Remote Sensing, 8*(12), 1014.

Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-Density Cypress Forest. *Sensors, 7*(11), 2636-2651.

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425-1432.

Miettinen, J., Shi, C., & Liew, S. C. (2016). Land cover distribution in the peatlands of Peninsular Malaysia, Sumatra and Borneo in 2015 with changes since 1990. *Global Ecology and Conservation, 6*, 67-78.

Nguyen, L. H., Joshi, D. R., Clay, D. E., & Henebry, G. M. (2020). Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. *Remote Sensing of Environment, 238*, 111017.

Oliphant, A. J., Thenkabail, P. S., Teluguntla, P., Xiong, J., Gumma, M. K., Congalton, R. G., & Yadav, K. (2019). Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *International Journal of Applied Earth Observation and Geoinformation, 81*, 110-124.

Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *Remote Sensing, 12*(15), 2411.

Pohl, C., & van Genderen, J. L. (1998). Review article Multisensor image fusion in remote sensing concepts, methods and applications. *International Journal of Remote Sensing, 19*(5), 823-854.

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., & Saah, D. (2019). Mapping Plantations in Myanmar by Fusing Landsat-8, Sentinel-2 and Sentinel-1 Data along with Systematic Error Quantification. *Remote Sensing, 11*(7), 831.

Praticò, S., Solano, F., Di Fazio, S., & Modica, G. (2021). Machine Learning Classification of Mediterranean Forest Habitats in Google Earth Engine Based on Seasonal Sentinel-2 Time-Series and Input Image Composition Optimisation. *Remote Sensing, 13*(4), 586.

Roy, D. P., Wulder, M. A., Loveland, T. R., C.E, W., Allen, R. G., Anderson, M. C., . . . Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154-172.

Sarzynski, T., Giam, X., Carrasco, L., & Lee, J. S. H. (2020). Combining Radar and Optical Imagery to Map Oil Palm Plantations in Sumatra, Indonesia, Using the Google Earth Engine. *Remote Sensing, 12*(7), 1220.

Scheffler, D., Frantz, D., & Segl, K. (2020). Spectral harmonization and red edge prediction of Landsat-8 to Sentinel-2 using land cover optimized multivariate regressors. *Remote Sensing of Environment, 241*, 111723.

Shaharum, N. S. N., Shafri, H. Z. M., Ghani, W. A. W. A. K., Samsatli, S., Al-Habshi, M. M. A., & Yusuf, B. (2020). Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. *Remote Sensing Applications: Society and Environment, 17*, 100287.

Sidhu, N., Pebesma, E., & Câmara, G. (2018). Using Google Earth Engine to detect land cover change: Singapore as a use case. *European Journal of Remote Sensing, 51*(1), 486-500.

Sothe, C., Almeida, C., Liesenberg, V., & Schimalski, M. (2017). Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. *Remote Sensing, 9*(8), 838.

Tan, M. L., Chua, V. P., Li, C., & Brindha, K. (2019). Spatiotemporal analysis of hydro-meteorological drought in the Johor River Basin, Malaysia. *Theoretical and Applied Climatology, 135*(3), 825-837.

Tan, M. L., Ibrahim, A. L., Yusop, Z., Duan, Z., & Ling, L. (2015). Impacts of land-use and climate variability on hydrological components in the Johor River basin, Malaysia. *Hydrological Sciences Journal, 60*(5), 873-889.

Tew, Y. L., Tan, M. L., Samat, N., & Yang, X. I. A. O. Y. I. N. G. (2019). Urban Expansion Analysis using Landsat Images in Penang, Malaysia. *Sains Malaysiana, 48*(11), 2307-2315.

Topaloğlu, R. H., Sertel, E., & Musaoğlu, N. (2016). Assessment of Classification Accuracies of Sentinel-2 and Landsat-8 Data for Land Cover / Use Mapping. *International archives of the photogrammetry, remote sensing & spatial Information Sciences, 41*.

van der Meer, F. D., van der Werff, H. M. A., & van Ruitenbeek, F. J. A. (2014). Potential of ESA's Sentinel-2 for geological applications. *Remote Sensing of Environment, 148*, 124-133.

Vijith, H., & Dodge-Wan, D. (2020). Applicability of MODIS land cover and Enhanced Vegetation Index (EVI) for the assessment of spatial and temporal changes in strength of vegetation in tropical rainforest region of Borneo. *Remote Sensing Applications: Society and Environment, 18*, 100311.

Wulder, M. A., Franklin, S. E., White, J. C., Linke, J., & Magnussen, S. (2006). An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data. *International Journal of Remote Sensing, 27*(4), 663-683.

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing, 27*(14), 3025-3033.