AUTOMATED LAND COVER MAPPING IN INSULAR SOUTHEAST ASIA USING MACHINE LEARNING

Aik Song CHIA*, Chenghua SHI, Kim Hwa LIM, Soo Chin LIEW
Centre for Remote Imaging Sensing and Processing (CRISP), National University of Singapore
10 Lower Kent Ridge Road, Blk S17 Level 2, Singapore 119076
Email: {crscas, crssc, crskhl, scliew}@nus.edu.sg

KEY WORDS: Land Cover Classification, Machine Learning, Landsat, Sentinel, Google Earth Engine

ABSTRACT: The per-humid insular Southeast Asia region has seen extensive land cover change in the past 30 years. Regularly updated land cover maps are crucial for continual tracking of these complicated changes and their effects. Such timely maps will greatly facilitate biomass estimation and measurement of deforestation which have potentially far-ranging impacts on climate change and the global environment. In this study, we present a method for automated generation of land cover maps via machine learning (ML). This approach enables automated production of 30m resolution maps of insular Southeast Asia with optical (Landsat 7,8) and radar (Sentinel-1) satellite data. We have previously generated land cover maps of insular Southeast Asia for 2015, using a manual rule-based decision tree method. To generate land cover maps for 2020, we trained a random forest classifier using the 2015 map as labels for the input training satellite data from 2015, and then applied the trained classifier to the 2020 satellite data to generate the 2020 map. This ML classification was implemented on the Google Earth Engine (GEE) platform. Accuracy assessment was done in ten sampling regions chosen for their landscape diversity. A total of 1000 random reference points in these areas were compared with very high-resolution images acquired from Google Earth as ground truth. Class-wise accuracies range from 66 to 97%, with an overall accuracy of 80%.

1. INTRODUCTION

The per-humid insular Southeast Asia region has seen extensive land cover change due to rapid development in the past 30 years (Miettinen *et al*., 2011). One of the more damaging changes is the continuing deforestation in the region which has potentially far-ranging impacts on climate change due to the carbon released into the atmosphere (Van der Werf *et al*., 2009). Regularly updated land cover maps are crucial for continual tracking of these complicated changes and their effects.

To facilitate periodic updating of regional land cover maps, we present in this paper a method for automated generation of land cover maps via machine learning (ML). We have applied this approach to produce a 30m resolution land cover map of insular Southeast Asia for the year 2020 (Fig. 1), using optical (Landsat 7,8) and radar (Sentinel-1) satellite data.

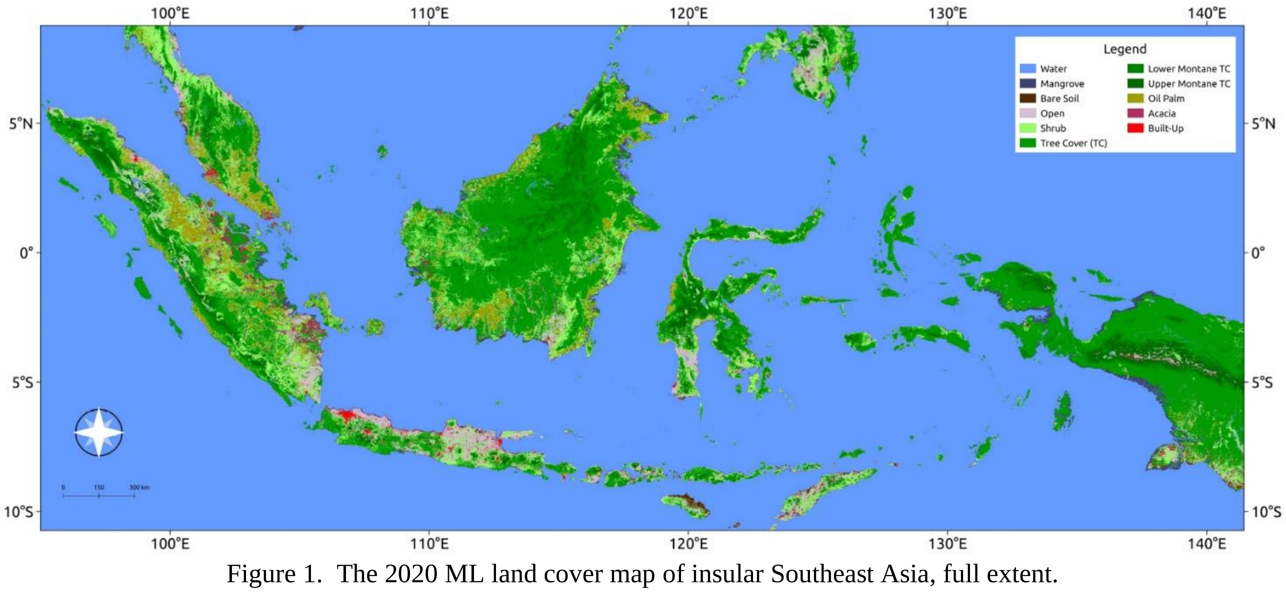

Figure 1. The 2020 ML land cover map of insular Southeast Asia, full extent.

We had previously generated a land cover map of Southeast Asia for 2015, using a manual rule-based decision tree (Miettinen *et al*., 2019). The rules of the decision tree were determined manually by an extensive visual examination of spectral indices and radar backscatter values in different land cover types. While the decision tree runs automatically once created, if we were to apply the same method to other time periods, even a small change in the datasets used requires a laborious re-specification of the parameters of the decision tree rules.

To generate the land cover map for 2020, we trained a ML classifier using the 2015 map as labels for the input training satellite data from 2015, and then applied the trained ML classifier to the 2020 satellite data to generate the 2020 map. Similar approaches to ML classification using training samples derived from existing land cover products have been implemented in the literature (Wessels *et al*., 2009; Zhang *et al*., 2021), but using solely optical satellite data. Our example, which utilized a combination of radar and optical satellite data, provides us the additional ability to extract plantation classes (oil palm and acacia) which are of particular interest in the Southeast Asia region.

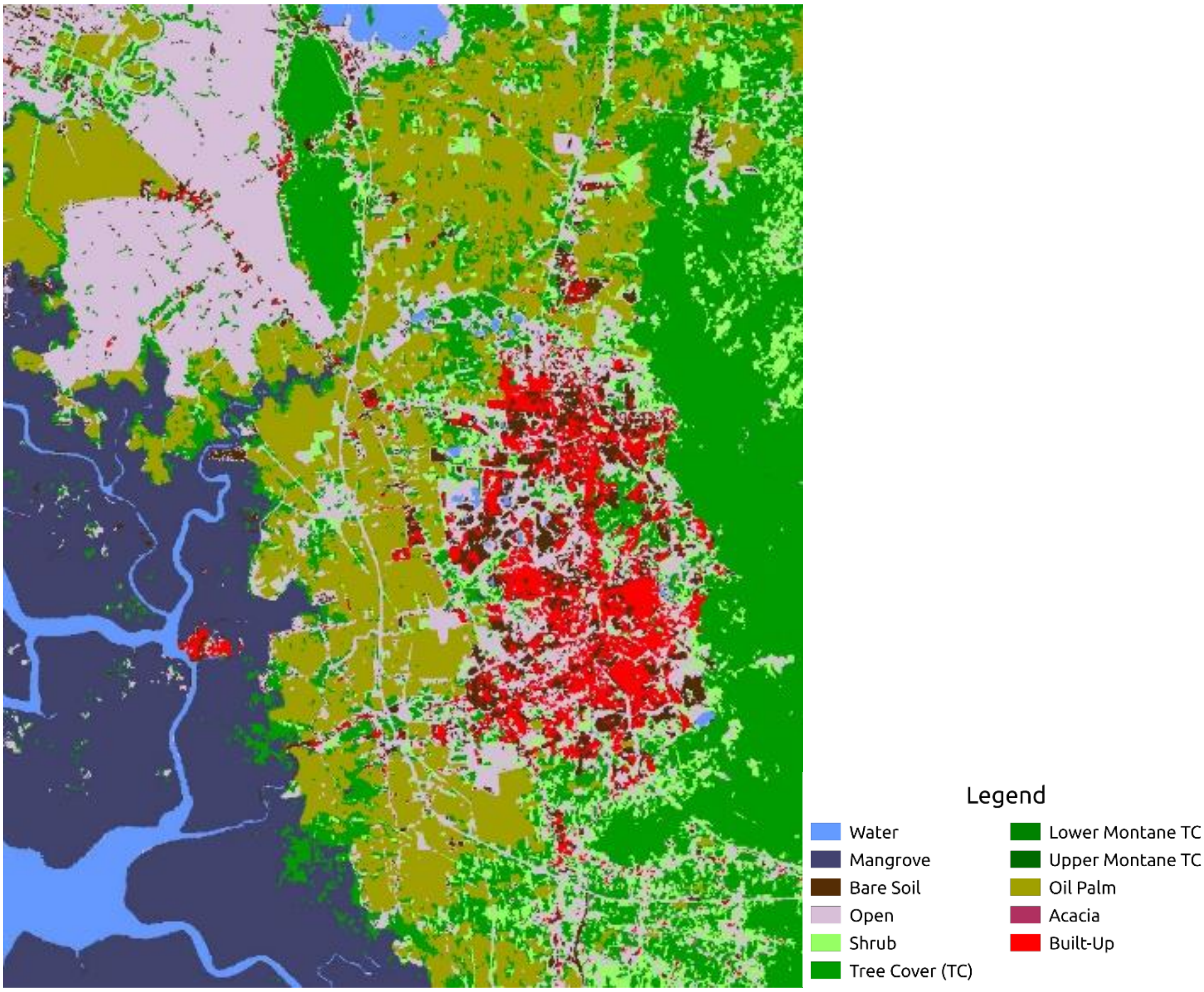

Figure 2. Sample from the 2020 ML land cover map at a larger scale, in Perak, Malaysia, showing the various land cover classes in more detail.

2. MATERIALS AND METHODS

We used a similar set of multi-sensor satellite data as in the 2015 decision tree land cover map, covering the same area of interest (AOI): 95°E - 141.5°E, 11°S - 9°N (see full extent in Fig. 1). The input variables for the ML training thus comprises optical (Landsat 7 and 8) and radar (Sentinel-1) data, as well as a set of auxiliary datasets which include digital terrain model (DTM) data derived from the Shuttle Radar Topography Mission (SRTM) (Farr *et al*., 2007) plus delineated regions denoting certain prior knowledge (i.e., areas where the *Acacia* class is known to occur, and areas where the oil palm class is known to be absent). The list of the input datasets is shown in Table 1.

Table 1. List of input datasets for ML training

Variable	Dataset	Description
B3-B7	Landsat 7&8 (1-yr composite)	Median surface reflectance of cloud-free pixels
VV, VH	Sentinel-1 (1-yr composite)	Low pass filtered backscatter (dB)
Elevation	SRTM	Low pass filtered elevation (m)
Slope	SRTM	Low pass filtered steepness (deg)
Acacia	Manual delineation	Binary mask of areas where *Acacia* is mapped
OilPalm	Manual delineation	Binary mask of areas where oil palm is not mapped

There are however two main changes from the original datasets used for the 2015 map: (a) the omission of ALOS PALSAR-2, as the 2020 dataset had gaps and also exhibited swath-edge errors, and (b) the Landsat datasets provided in GEE have been updated to Collection 5. We determined from testing that omission of the PALSAR-2 dataset did not have a significant impact on the ML classification.

There are a total of 11 land cover classes, following the ones identified and used in the 2015 land cover map. The ML classification was implemented on the Google Earth Engine (GEE) platform (Gorelick *et al*., 2019). We partitioned the AOI into 11 regions and separately for each region, trained the GEE random forest classifier on the above training variables, using the 2015 map as training labels. The trained classifier was then run on 2020 datasets to generate the land cover maps in each region for 2020. The individual land cover maps for each region were then mosaicked together to produce the full 2020 land cover map.

3. ACCURACY ASSESSMENT

Accuracy assessment was carried out in ten sampling regions. These ten regions were chosen for their landscape diversity, as well as the availability of very high-resolution imagery from Google Earth (Google, 2023) in the same year as the composite period for the Landsat and Sentinel-1 data.

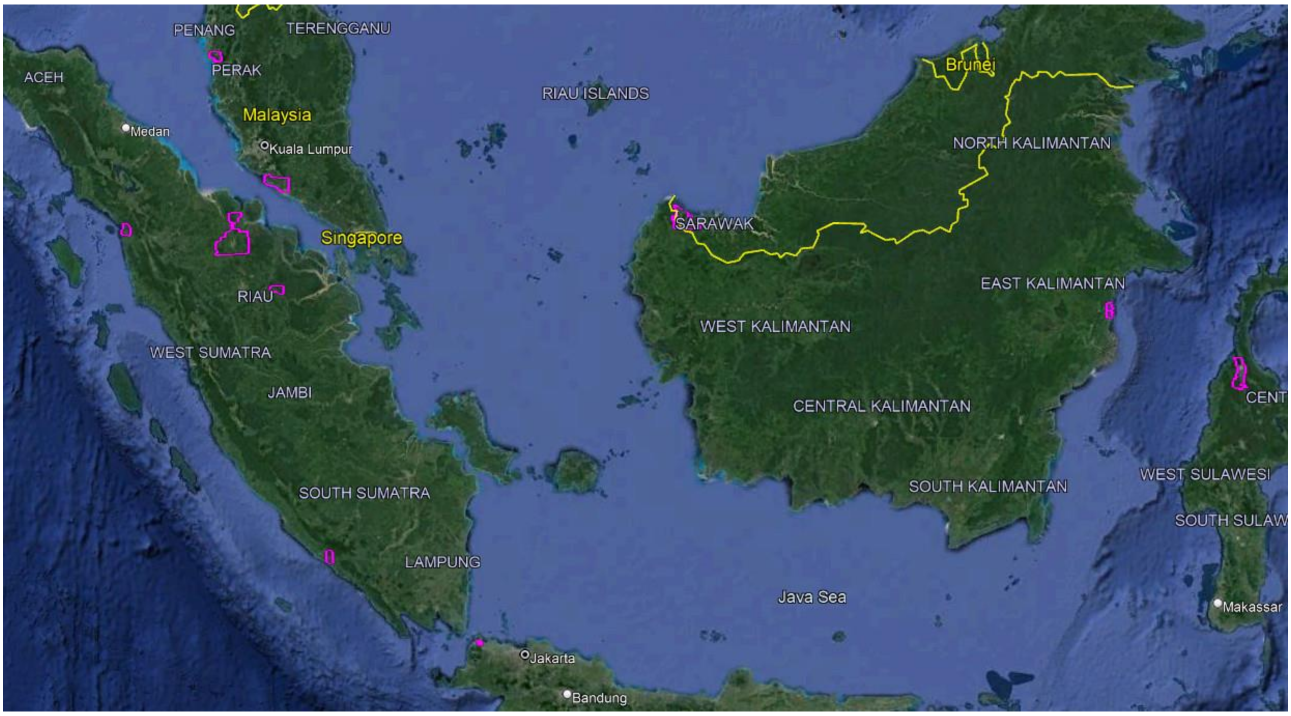

Figure 3. Ten sampling regions scattered throughout the land cover map AOI, marked out in pink.

A total of 1000 random reference points in these areas were compared with very high-resolution images acquired from Google Earth as ‘ground truth’. For simplicity, all the ‘Tree cover’ classes regardless of the elevation were considered as one ‘Tree cover’ class during the accuracy assessment, reducing the total number of classes to nine. The confusion matrix for this accuracy assessment is shown in Table 2 below:

Table 2. Accuracy assessment confusion matrix

		Reference										
		Water	Mangrove	Bare soil	Open	Shrub	Tree cover	Oil palm	*Acacia*	Built-up	**Total**	**User's acc.**
Map	Water	55	0	1	1	0	1	0	0	0	58	94.8
	Mangrove	2	38	0	7	4	3	0	0	0	54	70.4
	Bare soil	1	0	37	6	0	4	0	0	8	56	66.1
	Open	1	0	7	121	28	8	4	3	2	174	69.5
	Shrub	1	0	1	7	149	18	4	0	0	180	82.8
	Tree cover	0	0	0	3	29	232	23	0	0	287	80.8
	Oil palm	0	0	0	0	5	2	91	0	0	98	92.9
	Acacia	0	1	0	1	0	1	0	45	0	48	93.8
	Built-up	0	0	8	1	0	0	0	0	36	45	80.0
	Total	60	39	54	147	215	269	122	48	46	1000	
	Prod. Acc.	91.7	97.4	68.5	82.3	69.3	86.2	74.6	93.8	78.3		80.4

Class-wise accuracies range from 66.1 to 97.4%, with an overall accuracy of 80.4%. In particular, the plantation classes (oil palm and *acacia*) exhibited significantly high user's accuracies of above 90%. Apart from these accuracy values which can be considered acceptable for a fully automated land classification approach, the generated land cover maps are also visually informative at a high level of spatial detail (see Fig. 2).

4. CONCLUSION

We have presented a fully automated method of generating land cover maps in a specified AOI trained on a separate land cover map from a different time period covering that AOI. Applied to the insular Southeast Asia region, this allows us to extend our previously generated 2015 land cover map to provide a continued automated mapping of insular Southeast Asia with reasonable accuracy. These periodically generated maps will be useful in biomass estimation and understanding land cover change and deforestation in this region.

References:

Farr, T.G., Rosen, P.A., Caro, E. *et al.*, 2007. The shuttle radar topography mission. *Reviews of Geophysics*, 45(2).

Google, 2023. *Google Earth*. Available at https://earth.google.com (Latest access: 15th Sep 2023).

Gorelick, N., Hancher, M., Dixon, M. *et al.*, 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, pp.18-27.

Miettinen, J., Shi, C. and Liew, S.C., 2011. Deforestation rates in insular Southeast Asia between 2000 and 2010. *Global Change Biology*, 17(7), pp. 2261-2270.

Miettinen, J., Shi, C. and Liew, S.C., 2019. Towards automated 10–30 m resolution land cover mapping in insular South-East Asia. *Geocarto International*, 34(4), pp.443-457.

Van der Werf, G.R., Morton, D.C., DeFries, R.S. *et al.*, 2009. CO_2 emissions from forest loss. *Nature Geoscience*, 2(11), pp.737-738.

Wessels, K., Van den Bergh, F., Roy, D. *et al.*, 2009. Rapid land cover map updates using change detection and robust random forest classifiers, *Remote Sensing*, 8(11), p.888.

Zhang, X., Liu, L. and Chen, X., 2021. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery, *Earth System Science Data*, 13(6), pp.2753-2776.