Comparison of Machine Learning Classifiers for Land Cover Mapping in Google Earth Engine

Nyein Soe Thwal [*1] Takaaki Ishikawa [*2] Hiroshi Watanabe [*1, 2]

[*1] Graduate School of Fundamental Science and Engineering, Waseda University
[*2] Global Information and Telecommunication Institute, Waseda University

1. Introduction

Rapid migration and urbanization can affect many environmental issues. Monitoring land cover is crucial for effective environmental management and urban planning. Land cover maps are usually produced by using satellite image classification approaches because of their extensive geographical coverage at an efficient cost. Even though the land cover classification can be performed by using different machine learning algorithms, the accuracy and the processing time of classification using remote sensing images still have limitations.

The Google Earth Engine (GEE) is a cloud-based platform that can freely access a large number of satellite imageries. It can also perform monitoring and measurement of changes in the earth's environment at a planetary scale. Land cover classification for a large area can be executed in GEE by using its provided classifiers for pixel-based classification due to the advantages of its computational speed. The proposed study aims to use seven machine learning classifiers from GEE and compares the classification results of the classifiers.

2. Materials and Methods

2.1 Image Collections and Pre-processing

Landsat-8 and Sentinel-2 image collections of 2017 from GEE are used for land cover mapping over the study region, Yangon. Image mosaicking and cloud removing for Yangon region are performed on the two satellite image collections. After performing image enhancement, some spectral indexes such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Buildup Index (NDBI) and SRTM 30m slope are combined to the input image as the additional features.

NDVI and NDBI can be calculated based on the band values of Landsat-8 and Sentinel-2 images according to equation (1) and (2):

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad (1)$$

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR} \qquad (2)$$

Where NIR and RED in Equation (1) represent the near infrared and red reflectance values. SWIR in Equation (2) stands for the shortwave infrared reflectance values.

2.2 Collection of Training Data

In this study, seven land cover categories (Shrubland, Bare land, Forest, Agriculture, Buildup area, Lake and River) have been collected with the pixel and object-based methods. True color composite and false color composite methods are used for the extraction of spectral and spatial features from Landsat-8 images based on the pixel-based approach. The spectral, spatial and texture features are extracted from Sentinel-2 images by using object-based segmentation method. Moreover, feature fusion is performed by combining with the GPS metadata and Open Street Map (OSM) data. There are total 7813 vectors for training and testing dataset.

2.3 Image Classification

Classification of multi-sensors satellite imagery is performed at per-pixel basic in this experiment. Among the multiple classification algorithms in GEE, only six different classifiers are used for the comparison because of the low accuracies and less popularity of the other classifiers. 12 spectral bands combination from Landsat-8 and Sentinel-2 such as coastal aerosol, blue, green, red, near infrared and shortwave infrared bands are used in the training stage of the classifier.

Land cover classification over Yangon region is performed in GEE by using Random Forest (RF), Classification and Regression Tree (CART), Multiclass Perceptron, GMO Max Entropy, Minimum Distance and Continuous Naive Bayes algorithms. Multiclass Perceptron classifier is trained with 65 epochs and the Mahalanobic distance from the class mean is applied as a metric value in minimum distance algorithm

Table 1 Overall classification accuracy and kappa statistic achieved by GEE classifiers

Classifier	Overall Accuracy (%)	Kappa Statistic
Random Forest	96.73	0.952
CART	95.09	0.927
Multiclass Perceptron	66.21	0.437
GMO Max Entropy	80.85	0.716
Minimum Distance	87.85	0.827
Continuous NaiveBayes	81.53	0.733

3. Experiment Results

The overall accuracy (OA) and kappa static obtained from the six classifiers are described in Table 1and the best performance was achieved for Random Forest at OA of 96.73% and kappa statistic of 0.952. Likewise, RF, CART yields OA of 95.09% and the kappa statistic of 0.927. Among the six classifiers, Multiclass Perceptron provides the lowest accuracy of 66.21%. The classification map of Yangon obtained from the six classifiers of GEE are shown in Fig. 1. Although the GEE platform provides a set of classification algorithms, the best accuracy for land cover mapping was achieved from the decision tree-based classifiers, namely RF and CART.

4. Conclusion

In this paper, we compare six different machine learning classifiers of Google Earth Engine. The experimental result shows that the best accuracy of land cover mapping is achieved by RF and CART.

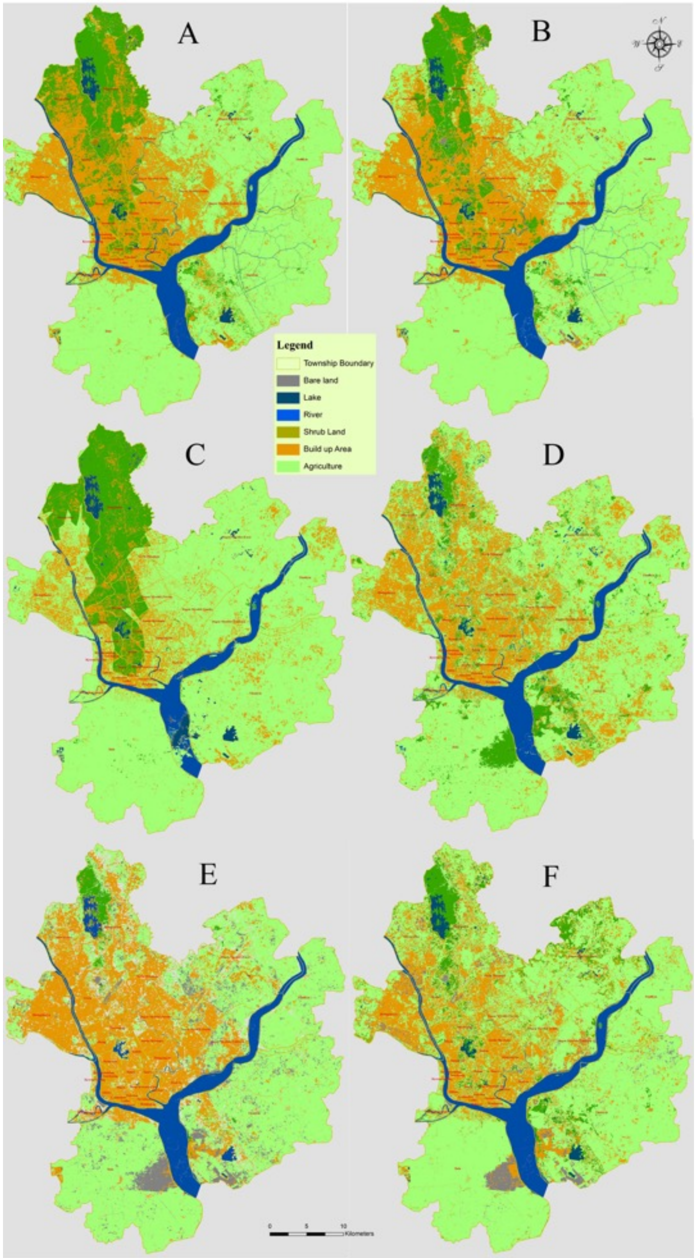

Fig. 1 Land cover maps of Yangon in 2017 generated by A RF, B CART, C Multiclass Perceptron, D GMO Max Entropy, E Minimum Distance and F Continuous NaiveBayes algorithms

REFERENCES

[1] Y. H. Tsai, D. Stow, H. L. Chen, R. Lewison, L. An, and L. Shi, "Mapping Vegetation and Land Use Types in Fanjingshan National Nature Reserve Using Google Earth Engine," Remote Sensing, Vol.10, No.6, pp.927-940, June 2018.

[2] A. Y. Shelestov, M. Lavreniuk, N. Kussul, S. V. Skakun, "Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping", Front. Earth Sci. 5:17. doi: 10.3389/feart.2017.00017.

[3] P. Thanh Noi, M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery", Sensors 2018, 18, 18; doi: 10.3390/s18010018

[4] OpenStreetMap, <http://www.openstreetmaps.org> (2018.10.20).