Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms

ABSTRACT

Oil palm plays a pivotal role in the ecosystem, environment, economy and without proper monitoring, uncontrolled oil palm activities could contribute to deforestation that can cause high negative impacts on the environment and therefore, proper management and monitoring of the oil palm industry are necessary. Mapping the distribution of oil palm is crucial in order to manage and plan the sustainable operations of oil palm plantations. Remote sensing provides a means to detect and map oil palm from space effectively. Recent advances in cloud computing and big data allow rapid mapping to be performed over large a geographical scale. In this study, 30 m Landsat 8 data were processed using a cloud computing platform of Google Earth Engine (GEE) in order to classify oil palm land cover using non-parametric machine learning algorithms such as Support Vector Machine (SVM), Classification and Regression Tree (CART) and Random Forest (RF) for the first time over Peninsular Malaysia. The hyperparameters were tuned, and the overall accuracy produced by the SVM, CART and RF were 93.16%, 80.08% and 86.50% respectively. Overall, the SVM classified the 7 classes (water, built-up, bare soil, forest, oil palm, other vegetation and paddy) the best. However, RF extracted oil palm information better than the SVM. The algorithms were compared and the McNemar's test showed significant values for comparisons between SVM and CART and RF and CART. On the other hand, the performance of SVM and RF are considered equally effective. Despite the challenges in implementing machine learning optimisation using GEE over a large area, this paper shows the efficiency of GEE as a cloud-based free platform to perform bioresource distributions mapping such as oil palm over a large area in Peninsular Malaysia.

Keyword: Cloud computing; Image classification; Landsat; Machine learning; Oil palm