COMPARISON OF MACHINE LEARNING MODELS FOR LAND COVER CLASSIFICATION

Bambang H. Trisasongko[1,2], Dyah R. Panuju[1,2], Nur Etika Karyati[2], Rizqi I'anatus Sholihah[2]
[1]Department of Soil Science and Land Resources, Bogor Agricultural University, Bogor, Indonesia
[2]GITIIA/CRESTPENT, Bogor Agricultural University, Bogor, Indonesia
e-mail: trisasongko@apps.ipb.ac.id

Received:19.08.2022; Revised:29.08.2022; Approved:27.09.2022

Abstract. Land cover data remain one of crucial information for public use. With rapid human-associated land alteration, this information needs to be frequently updated. Remotely-sensed data provide the best option to construct land cover maps with numerous methods available in the literature. While disagreement exists to select the robust one, further exploration should be made to extend the understanding on the behavior of machine learners, in particular, for classification problems. This article discusses performance of pixel-based machine learning algorithms, frequently used in research or implementation. Five popular algorithms were evaluated to distinguish five rural land cover classes, i.e. built-ups, crops, mixed garden, oil palm plantations and rubber estates, from Sentinel-2 data. This research found that the benchmark, i.e., Classification and Regression Trees, was unable to differentiate woody vegetation, although the overall accuracy was moderate. This suggested that overall accuracy cannot be seen as the only measure for assessing the quality of the thematic output. Meanwhile, support vector machines and random forest competed to yield the highest accuracy and class detection capability, although the latter was in favor with 98% accuracy level. A newly developed model, like extreme gradient boosting, achieved a similar level of accuracy. This research implies that modern machine learning approaches would be invaluable for land cover classification; hence, access to these modeling toolkits is substantial.

Keywords: *artificial neural networks, classification and regression tree, extreme gradient boosting, random forests, Sentinel-2, support vector machines*

1 INTRODUCTION

Provision of land cover/land use map is crucial for numerous applications. In the modern world, this is achievable using publicly available remote sensing data. The legacy of Landsat sensors has been undisputed, providing decadal time series data covering most of Earth's surface. These have been implemented for various purposes, including general land cover mapping (Nasiri et al. 2022), forestry applications (Izadi et al. 2022), and crop investigation (Najafi et al. 2018). With expanding applications and the necessity to obtain a better spatial scale, one could consider Sentinel-2 Multispectral Imager (MSI).

This sensor has offered the possibility to exploit more spectral range, provision of the red edge band and more importantly, a frequent time of acquisition. These benefits allow in-depth investigations within contexts of agriculture (Zou et al. 2022), forestry (Xi et al. 2022) and land cover mapping (Yousefi et al. 2022). Accuracy levels are generally sufficient for mapping purposes. Nonetheless, further examination is needed to confirm previous findings.

A survey of the literature suggests that successful information extraction would also depend on the analysis, not merely the data ingest. To date, machine learning approaches have been largely employed. Numerous algorithms are available for classification or regression problems. Conventional approaches like Classification and Regression Trees (CART) or other monolithic tree algorithms are often used for simple cases.

Complex applications would likely require better constructed algorithms. Ensemble learning could be a better candidate for this problem. It allows an iterative learning process and summarizes the most acceptable outcome to a variety of datasets. To date, algorithms like random forests (RF), proposed by Breiman (2001) or support vector machines (SVM) by Vapnik (2000) have extensively been studied. Attempts to investigate both algorithms have been presented in the literature with a confronting conclusion. Some studies have found that SVM delivered a better output than RF (Shih et al. 2019, Abdi 2020). Nonetheless, the opposite inference has also been reported (Panuju et al. 2019, Adab et al. 2020, Panuju et al. 2021, Adugna et al. 2022). This confusion may be rooted from various sources. While examination to this was mostly about the data ingestion and the machine learning algorithms, the rarity of studies focusing on tuning parameters is evident in the literature.

Uncertain summary suggests that large scale comparison should be made to further investigate benefits and limitations on using certain machine learning algorithms. This article attempts to fill the gap by providing an assessment on the robustness of machine learners. The purpose of this research was twofold. The first was to measure the robustness of machine learners in discriminating woody vegetation over rural areas. Secondly, this study evaluated critical parameter settings of potential models.

2 METHODOLOGY

2.1 Test Site

Lebak regency, Indonesia, was selected as a research location, considering its setting as a hinterland for either Jakarta metropolitan or Serang-Cilegon industrial areas (Figure 2-1). With this position, Lebak possesses a tendency for land conversion, while its role in providing agricultural products remains important. Food crops have been planted throughout the regency, although they are concentrated in the northern part of the regency. Plantations have also been important to Lebak's economy, especially oil palm and rubber estates.

Land resources in Lebak regency suit a variety of plantations. Northern part of the test site is generally dominated by a flat and gentle slope. Soil type of this region is predominantly Latosols, Cambisols and Podzolic, according to Indonesian soil classification. Latosols are also found in hilly terrain in the mid part of the regency. In addition, a humid tropical environment sustains agriculture, with relatively constant temperature (26.9°-28.4°C) and at least 7 rainy days per month.

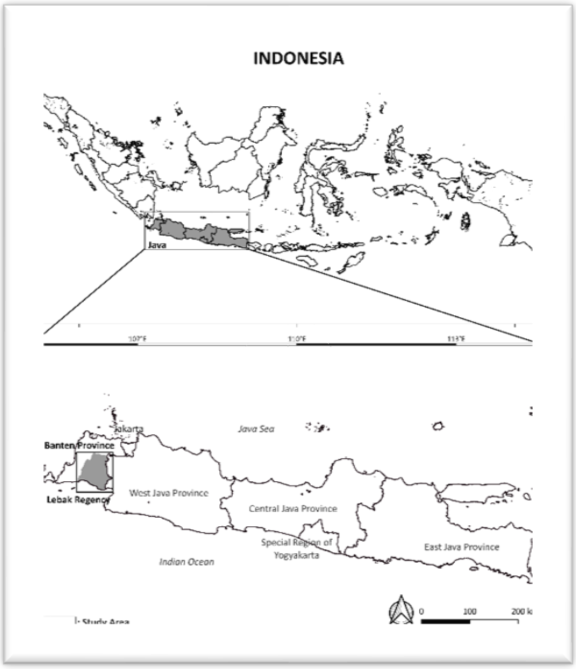

Figure 2-1: Research site, Lebak, Indonesia.

2.2 Data

This research relied on Sentinel-2 Multispectral Imager (MSI) as spatial data ingest. The data were downloaded from the European Space Agency (ESA) website, dated 8 July 2022. About 1347 samples were available for this research. This dataset was mainly rooted from previous 2015-2016 surveys and Google Earth imagery. Sample polygons derived from the latter were then reconfirmed through ground survey. The survey was conducted at the end of the month, covering the entire upper north of the regency. It covered a variety of land cover/land use, including built-up areas, waterbodies (basically rivers), agricultural crops (mainly rice fields), mixed gardens (fruits and small lumber trees) and plantations (oil palm and rubber). Figure 2-2 depicts field documentation.

2.3 Pre-processing and analysis

With the nature of level 2 data (bottom of atmosphere), Sentinel-2 pre-processing was straightforward. The data were subsetted covering only the test site, before resampled and reprojected according to baseline maps (EPSG 4326). Since partial areas were covered by clouds and shadows, masking was performed utilizing scene classification image taken from the original level 2 data. All preprocessing procedures were done in SNAP software from ESA.

The output of pre-processing procedures was then ingested into R statistical software. A code was written using RStudio editor to accommodate the whole process, starting from reading original SNAP's DIM format until model inversion into thematic map. This research implemented ten-fold cross validation during modeling, in order to minimize bias. Training and validation ratio was 75:25. All training data were analyzed using five algorithms. CART was implemented using 'rpart' package, while 'nnet' package was exploited in artificial neural networks (ANN) modeling. Current study focused on three contemporary machine learning approaches, i.e., RF, SVM and extreme gradient boosting (XGB), proposed by Chen and Guestrin (2016). Variety of RF approaches is available in R packages. Nonetheless, this research only explored commonly-used 'randomForest' package. For SVM, we evaluated 'kernlab', instead of 'e1071' package. Implementation of XGB was made using 'xgboost' package. Accuracy assessment was implemented using validation dataset, by examining overall accuracy and individual class-based error. The best performing model was then predicted into a raster map, using R 'raster' package.

2.3 Pre-processing and analysis

With the nature of level 2 data (bottom of atmosphere), Sentinel-2 pre-processing was straightforward. The data were subsetted covering only the test site, before resampled and reprojected according to baseline maps (EPSG 4326). Since partial areas were covered by clouds and shadows, masking was performed utilizing scene classification image taken from the original level 2 data. All preprocessing procedures were done in SNAP software from ESA.

The output of pre-processing procedures was then ingested into R statistical software. A code was written using RStudio editor to accommodate the whole process, starting from reading original SNAP's DIM format until model inversion into thematic map. This research implemented ten-fold cross validation during modeling, in order to minimize bias. Training and validation ratio was 75:25. All training data were analyzed using five algorithms. CART was implemented using 'rpart' package, while 'nnet' package was exploited in artificial neural networks (ANN) modeling. Current study focused on three contemporary machine learning approaches, i.e., RF, SVM and extreme gradient boosting (XGB), proposed by Chen and Guestrin (2016). Variety of RF approaches is available in R packages. Nonetheless, this research only explored commonly-used 'randomForest' package.

Figure 2-2: Field records, equipped with coordinates and some local identification. Top to bottom: settlements, rice agriculture with mixed garden in the background, oil palm plantation and rubber estate.

For SVM, we evaluated 'kernlab', instead of 'e1071' package. Implementation of XGB was made using 'xgboost' package. Accuracy assessment was implemented using validation dataset, by examining overall accuracy and individual class-based error. The best performing model was then predicted into a raster map, using R 'raster' package.

3 RESULTS AND DISCUSSION

3.1 Classification accuracy

Table 3-1 summarizes accuracies of studied machine learning algorithms. This research found that single, monolithic tree-based learning like CART did not fully achieve an acceptable outcome. Despite its overall accuracy being above 80%, the mapping benchmark (Panuju et al. 2020), detectabilities of mixed garden and rubber were severely low. This research suggested that with a sufficiently large number of predictors (i.e., Sentinel-2 20-m and 10-m bands), CART was unable to entirely distinguish woody vegetation (Figure 3-1).

Table 3-1: Overall and class-wise accuracies recorded from validation datasets.

Machine Learner	Accuracy (%)					
	Overall	**Settlement**	**Crops**	**Mixed Garden**	**Oil Palm**	**Rubber**
CART	86.3	93.3	88.9	50.0	94.2	50.0
ANN	93.9	96.3	96.4	69.5	97.1	85.5
RF	**97.6**	**98.2**	**98.3**	**86.4**	**98.9**	**96.6**
SVM	95.6	97.6	97.7	74.8	97.8	89.7
XGB	96.8	97.7	98.1	83.1	98.4	94.8

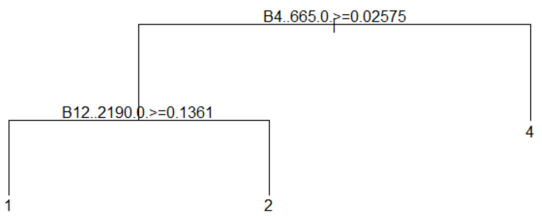

Figure 3-1: CART model. Codes: 1=Built-up; 2=Crops; 4=Oil Palm.

Distinctiveness of woody vegetation was clear; the case where red band (band 4, with central wavelength around 665 nm) plays a significant role. This spectrum is well-known as a good band source for classification, either as the raw band or in its combination with other spectral bands in vegetation indices. Discrimination of vegetation with low biomass was done using short-wave infrared (band 12; central wavelength about 2190 nm), although other NIR bands could serve similarly. Andrade et al. (2021) concluded that SWIR was one of important bands in tropical land cover classification using single date data. Similarity of land cover, however, remains the issue with the implementation of CART, or perhaps with similar monolithic tree learners.

The rests of the machine learning model successfully achieved very high overall accuracy. While recently being less popular, ANN could provide an acceptable accuracy. Complexity of mixed garden, however, remained problematic in ANN. As shown in Figure 2-2, mixed garden in the research site was composed from shrubs and woody vegetation, especially fruits and timber. As shown in Figure 3-2, ANN model successfully identified all land cover types.

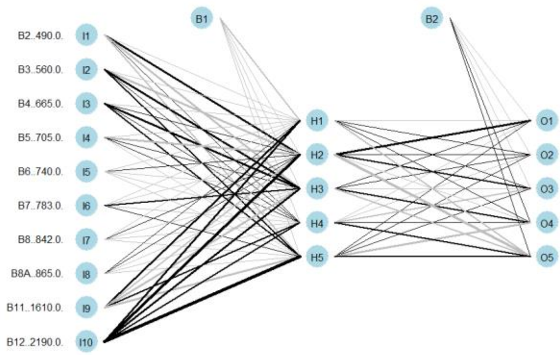

Figure 3-2: ANN model. Codes: 1=Built-up; 2=Crops; 3=Mixed garden; 4=Oil Palm; 5=Rubber.

This research suggested that only a few bands contributed to class separation. As previously indicated, red (band 4) and SWIR (band 12) spectra greatly contributed to the discrimination. Nonetheless, this network's structure informed that additional bands suited to classification problems as well, i.e., bands 2 (blue), 3 (green) and 11 (another SWIR). Interestingly, red edge (bands 5 to 7) lightly performed in the discrimination. It was expected that the provision of red edge spectrum could contribute to differentiating woody vegetation. This finding was fairly similar to a conclusion of Shafeian et al. (2021), suggesting that not all Sentinel-2 bands are robust in all cases of land cover mapping.

We note, however, that some studies have reported inversely. Heckel et al. (2020), for instance, described those contributions of red edge bands, along with SWIR and RGB bands as significant in the case of temperate forest. In line with this, Waśniewski et al. (2020) similarly found the contribution of red edge when applied in tropical forest monitoring.

This research found that RF and SVM produced similar levels of accuracy. While numerous publications favored SVM over RF (Abdi 2020), this study indicated that the accuracy of RF models surpassed the one yielded by SVM.

In SVM, radial basis function (RBF) kernel is often used, especially when classification problems are fairly complex. As separating woody vegetation has been challenging, this research did not attempt to employ a linear kernel, although the computation time is generally faster than other kernels. Robustness of RBF has also been reported elsewhere (Trisasongko et al. 2022).

Although Table 3-1 suggests that SVM was inferior to RF, the discrepancies are generally within 5%. It should be noted however, that a threshold of 85% accuracy has been suggested to deliver a useful thematic map (Panuju et al. 2019, Panuju et al. 2020). Complexity of mixed gardens was the root of confusion in SVM analysis, which is well understood since the variety of trees is substantially high.

Optimal SVM-RBF model in this research was achieved by setting the cost parameter to 1 and gamma=0.23. We noticed, however, that the range of accuracy, according to different cost settings, was less than 2%. This indicates that variation of accuracy was quite low, despite the best outcome being at cost=1.

Setting cost in SVM would be a complicated task (see also a discussion in Adugna et al. (2022)). Previous research have reported variably. While large cost setting (above 10) was discovered to be significant (Wan et al. 2021, Zagajewski et al. 2021, Adugna et al. 2022), smaller penalty regularization (the cost) values were found useful (Foody and Mathur 2004, Trisasongko et al. 2017). These suggest that large scale studies should be conducted to further examine this parameter in various environments.

While it has been introduced a while, limited studies have explored XGB in classifying land cover. This research found that its discrepancy to the outcomes of RF, either in overall or class-wise accuracies, was less than 3%. This suggests that XGB potentially contributes as an alternative to RF or SVM.

Like SVM, XGB model requires efforts to examine parameter settings. Accuracy made by XGB, i.e., 96.8%, was achieved by adjusting parameters. This research found the optimal column subsample was 0.8, similar to the ones achieved by Buthelezi et al. (2020), i.e., 0.94 - 0.96. A slightly lower parameter, ranging from 0.42 to 0.65, was used in temperate forest classification (Grabska et al. 2020).

Maximum depth was found optimal at 3, which was close to the finding of Pham et al. (2020) for mangrove's aboveground biomass estimation. The same publication also mentioned that zero was the ideal gamma parameter, which was supported by the finding of current study.

Rigorous conclusion is yet to be made. It is, therefore, important to note that access to larger models would be a great advantage to seek a better model and, in turn, a better model prediction. In terms of land cover mapping using remotely-sensed data, rural areas expose a great challenge. This is due to the complexity of land cover types with the domination of woody vegetation. With this challenging situation, finding an optimized model is usually sought through parameter tuning.

3.2 Tuning parameters

Default parameter setting for each machine learning algorithm could not be optimal for any case. For this reason, analysis of tuning parameters should be conducted. Figure 3-3 presents an experiment on the effect of mtry parameter to the accuracy. As shown, initial mtry value achieved high accuracy, suggesting the robustness of RF in its initial stage. We found that significant improvement of the accuracy was made up until about mtry=6. Setting mtry with a large number appeared unnecessary.

Meanwhile, tuning the number of trees is depicted in Figure 3-4. The figure indicates that a large number of trees is superfluous when numerous proxies are exploited. In this research, setting ntree parameter to 50 would be sufficient to reduce errors, either overall or class-wise errors. Setting a large tree number in RF consumes substantial computing time; hence, understanding the consequence of employing an excessive number of trees is pivotal in modeling.

This research indicated that mixed gardens were exceptionally responsive to tuning. By setting a larger number of trees in the forest, more than 10% improvement was made. It appears that complexity of this class requires more trees to resolve the discrimination. With that case, it is understood that a single, monolithic tree, such as CART, was unable to separate all targets. Second responsive class type was rubber estate. This was possibly due to the fact that the majority of rubber estates had been managed by farmers; some of which were considered jungle rubber.

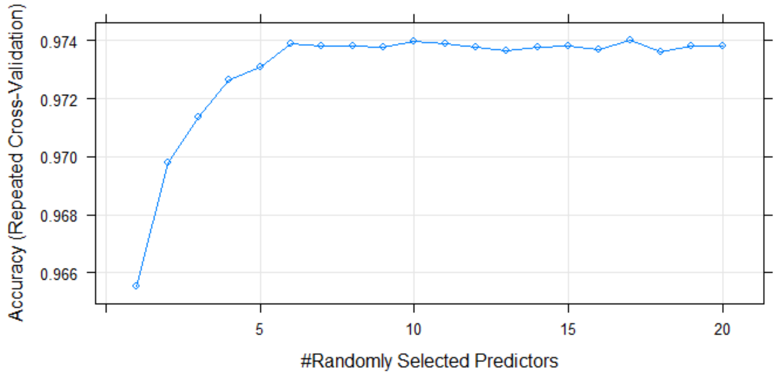

Figure 3-3: Tuning mtry parameter.

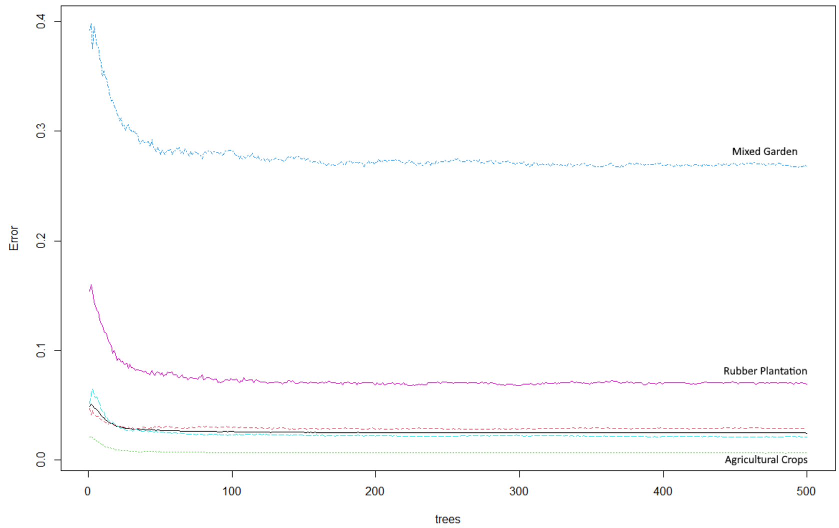

Figure 3-4: Tuning number of trees.

Predicted thematic map is presented in Figure 3-5. Large contribution to this thematic map was due to SWIR and red bands, or bands 12 and 4 of Sentinel-2 MSI, which is in line with aforementioned discussion about CART. In addition, another SWIR band (band 11) provided an assistance in the discrimination. This was consistent with the ANN model. Similar to the neural network model presented in Figure 3-2, red edge bands were found irresponsive during the discrimination.

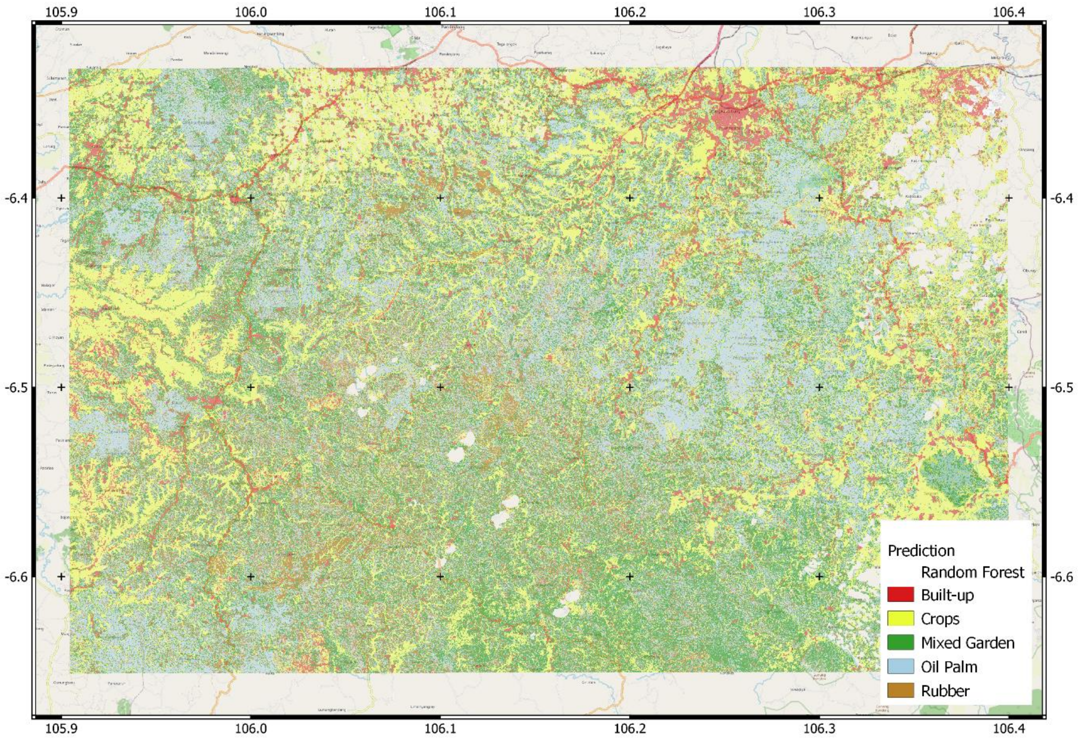

Figure 3-5: Classified map using random forest model, draped over OpenStreetMap data.

4 CONCLUSIONS

This research compares various machine learners to map and to verify the robustness of five popular learners in differentiating rural land cover classes. Employing Sentinel-2 imagery, classification and regression tree was incapable of separating woody vegetation. Meanwhile, four other classifiers were superior in terms of overall accuracy. The contribution of Sentinel-2 spectra was heterogeneous during classification; thus selection, transformation or other strategies are required. Tuning parameters informed the sensitivity of learners within a range of parameter values being explored. The process appears useful to generate the best accuracy which varies across environmental settings. Complexity of mixed garden seemed constraining machine learners to perform; yet tuning successfully improved the performance. Investigating unexplored tuning parameters in the future may better comprehend the sensitivity and robustness of learners towards complex settings. Moreover, the result advises to employ additional measures in assessing the quality of classification.

ACKNOWLEDGEMENTS

This research was supported by the Directorate General for Higher Education, Ministry of Education, Culture, Research and Technology through PDUPT research scheme, under contract 186/E5/PG.02.00.PT/2022. We extend our gratitude to field assistants for helping during surveys. We also acknowledge contributions of reviewers and editors for their careful reading, leading to a more improved manuscript.

REFERENCES

Abdi A M, (2020), Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GISci Remote Sens 57(1): 1-20. doi:10.1080/15481603.2019.1650447

Adab H, Morbidelli R, Saltalippi C, Moradian M, Ghalhari GAF, (2020), Machine learning to estimate surface soil moisture from remote sensing data. Water (Switzerland) 12(11): 1-28. doi: 10.3390/w12113223

Adugna T, Xu W, Fan J, (2022), Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sens 14(3). doi: 10.3390/rs14030574

Andrade J, Cunha J, Silva J, Rufino I, Galvão C, (2021), Evaluating single and multi-date Landsat classifications of land-cover in a seasonally dry tropical forest. Remote Sens Appl: Soc Environ 22: 100515. doi: 10.1016/j.rsase.2021.100515

Breiman L, (2001), Random forests. Mach Learn 45(1): 5-32. doi: 10.1023/A:1010933404324

Buthelezi MNM, Lottering RT, Hlatshwayo ST, Peerbhay K, (2020), Comparing rotation forests and extreme gradient boosting for monitoring drought damage on KwaZulu-Natal commercial forests. Geocarto Int 37(11): 3223-3246. doi: 10.1080/10106049.2020.1852612

Chen T, Guestrin C, (2016), XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA, ACM: 785-794. doi: 10.1145/2939672.2939785

Foody GM, Mathur A, (2004), A relative evaluation of multiclass image classification by support vector machines. IEEE Trans Geosci Remote Sens 42(6): 1335-1343. doi: 10.1109/TGRS.2004.827257

Grabska E, Frantz D, Ostapowicz K, (2020), Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. Remote Sens Environ 251: 112103. doi: 10.1016/j.rse.2020.112103

Heckel K, Urban M, Schratz P, Mahecha MD, Schmullius C, (2020), Predicting forest cover in distinct ecosystems: The potential of multi-source Sentinel-1 and -2 data fusion. Remote Sens 12(2): 302. doi: 10.3390/rs12020302

Izadi S, Sohrabi H, Khaledi MJ, (2022), Estimation of coppice forest characteristics using spatial and non-spatial models and Landsat data. J Spatial Sci 67(1): 143-156. doi: 10.1080/14498596.2020.1734110

Najafi P, Navid H, Feizizadeh B, Eskandari I, (2018), Object-based satellite image analysis applied for crop residue estimating using Landsat OLI imagery. Int J Remote Sens 39(19): 6117-6136. doi: 10.1080/01431161.2018.1454621

Nasiri V, Deljouei A, Moradi F, Sadeghi SMM, Borz SA, (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 Sens 14(9). doi: 10.3390/rs14091977

Panuju DR, Paull DJ, Griffin AL, (2020), Change Detection Techniques Based on Multispectral Images for Investigating Land Cover Dynamics. Remote Sens 12(11). doi: 10.3390/rs12111781

Panuju DR, Paull DJ, Griffin AL, Trisasongko BH, (2021) Mapping Rice Growth Stages Employing MODIS NDVI and ALOS AVNIR-2. In: Kumar P, Sajjad H, Chaudhary BS, Rawat JS, Rani M, (eds) Remote Sensing and GIScience: Challenges and Future Directions. Springer, Cham, p 185-203. doi: 10.1007/978-3-030-55092-9_11

Panuju DR, Paull DJ, Trisasongko BH, (2019), Combining binary and post-classification change analysis of augmented ALOS backscatter for identifying subtle land cover changes. Remote Sens. 11(1): 100. doi: 10.3390/rs11010100

Pham TD, Yokoya N, Xia J, Ha NT, Le NN, Nguyen TTT, Dao TH, Vu TTP, Pham TD, Takeuchi W, (2020), Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam. Remote Sens 12(8). doi: 10.3390/RS12081334

Shafeian E, Fassnacht FE, Latifi H, (2021), Mapping fractional woody cover in an extensive semi-arid woodland area at different spatial grains with Sentinel-2 and very high-resolution data. Int J Appl Earth Obs Geoinfo 105: 102621. doi: 10.1016/j.jag.2021.102621

Shih HC, Stow DA, Tsai YH, (2019), Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. Int J Remote Sens 40(4): 1248-1274. doi: 10.1080/01431161.2018.1524179

Trisasongko BH, Panuju DR, Griffin AL, Paull DJ, (2022), Fully Polarimetric L-Band Synthetic Aperture Radar for the Estimation of Tree Girth as a Representative of Stand Productivity in Rubber Plantations. Geographies 2(2): 173–185. doi: 10.3390/geographies2020012

Trisasongko BH, Panuju DR, Paull DJ, Jia X, Griffin AL, (2017), Comparing six pixel-wise classifiers for tropical rural land cover mapping using four forms of fully polarimetric SAR data. Int J Remote Sens 38(11): 3274-3293. doi: 10.1080/01431161.2017.1292072

Vapnik VN, (2000), The nature of statistical learning theory. Springer, New York.

Wan S, Yeh ML, Ma HL, (2021), An innovative intelligent system with integrated CNN and SVM: Considering various crops through hyperspectral image data. ISPRS Int J Geo-Info 10(4). doi: 10.3390/ijgi10040242

Waśniewski A, Hościło A, Zagajewski B, Moukétou-Tarazewicz D, (2020), Assessment of Sentinel-2 Satellite Images and Random Forest Classifier for Rainforest Mapping in Gabon. Forests 11: 941. doi: 10.3390/f11090941.

Xi Y, Tian J, Jiang H, Tian Q, Xiang H, Xu N, (2022), Mapping tree species in natural and planted forests using Sentinel-2 images. Remote Sens Lett 13(6): 544-555. doi: 10.1080/2150704X.2022.2051636

Yousefi S, Mirzaee S, Almohamad H, Dughairi AA, Gomez C, Siamian N, Alrasheedi M, Abdo HG, (2022), Image Classification and Land Cover Mapping Using Sentinel-2 Imagery: Optimization of SVM Parameters. Land 11(7). doi: 10.3390/land11070993

Zagajewski B, Kluczek M, Raczko E, Njegovec A, Dabija A, Kycko M, (2021) Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the krkonoše/karkonosze transboundary biosphere reserve. Remote Sens 13: 2581. doi: 10.3390/rs13132581

Zou X, Zhu S, Mõttus M, (2022), Estimation of Canopy Structure of Field Crops Using Sentinel-2 Bands with Vegetation Indices and Machine Learning Algorithms. Remote Sens 14: 2849. doi: 10.3390/rs1412284