Mapping Singapore's Land Use based on IPCC Guidelines:

A Framework-Based Training Guide for Effective Classification

Jacylin Jie Yi Tham\*1, Kim Hwa Lim2 and Soo Chin Liew3

1Associate Scientist, Centre for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore

S17 Level 2, Lower Kent Ridge Road, Singapore 119260

Email: [jacylint@nus.edu.sg](mailto:jacylint@nus.edu.sg)

2Deputy Director, Centre for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore

S17 Level 2, Lower Kent Ridge Road, Singapore 119260

Email: [crskhl@nus.edu.sg](mailto:crskhl@nus.edu.sg)

3Principal Research Scientist, Centre for Remote Imaging, Sensing and Processing (CRISP),

National University of Singapore

S17 Level 2, Lower Kent Ridge Road, Singapore 119260

Email: [scliew@nus.edu.sg](mailto:scliew@nus.edu.sg)

**KEY WORDS:** Land Use Classification, LULUCF, IPCC

**ABSTRACT:**  A comprehensive framework-based training guide designed to enhance the accuracy and efficiency of land use classification in Singapore is presented. The guide is created to align with the Intergovernmental Panel on Climate Change (IPCC)’s land use categories. This guide primarily utilises the capabilities of Google Earth Engine (GEE) to access extensive geospatial data and perform supervised classification on the land cover. The analysis is based on 10-m spectral bands of Sentinel-2 Harmonized collection, with cloud removal techniques applied and clipped to the boundary of Singapore. Training data for several machine learning classification algorithms is defined from two scenes of Sentinel-2 acquired in 2019 and 2020, by visually examining and labelling areas in these images to create corresponding feature collections. Results of a preliminary accuracy assessment show that Random Forest provides a higher overall accuracy of 80% to 83%. The validation accuracy was derived by calculating the confusion matrix of the trained model and then using it to compute the overall accuracy of the classification via Google Earth Engine. In conclusion, this guide aims to enhance classification accuracy of Singapore’s land use for measurement, reporting, and verification (MRV) of greenhouse gas (GHG) emissions from Singapore’s Land Use, Land-Use Change, and Forestry (LULUCF) sector.

# INTRODUCTION

Singapore is predominantly urbanised with a high population density, and thus, the planning of land use is vital to providing a long-term vision for the nation’s development (URA, 2023). As reported in Singapore’s Fourth Biennial Update (NEA, 2020), Singapore has experienced multiple shifts in land-use subcategories due to its dynamic nature within the last two decades. This can result in overestimations of greenhouse gas emissions due to exaggerated variations in carbon pools (NEA, 2020). Therefore, it is crucial that classifying land use is a seamless and transparent process to keep up to date with ever-changing landscapes and plans.

According to the Intergovernmental Panel on Climate Change (IPCC) guidelines for land classification, land is classified under the broad categories of Forest Land, Cropland, Grassland, Wetlands, Settlements, and Other Land. In Singapore, Forest Land includes some areas of Bukit Timah Nature Reserves, Central Catchment, and Singapore Botanic Gardens Rainforest. Regarding Cropland, Singapore has limited agricultural land due to the scarcity of land and urban focus, but it still exists mainly in the North and Northwest regions, such as the Lim Chu Kang area, where there are a few farms. The category Grassland is not relevant for Singapore, as lawns and grassland patches are located in between infrastructure, in urban parks, and stocked forests and are subsumed under the Forest Land category or under Settlements using specific emission factors for such low vegetation (NEA, 2018). The Wetlands category includes mangroves, reservoirs, natural rivers, and lakes such as Sungei Buloh Wetland Reserve. The Settlements include all developed land, including transportation infrastructure and human settlements of any size. The category Other Land is not occurring in Singapore but instead replaced with the ‘Other’ category to capture emissions and removals mainly from activities such as land reclamation projects.

By adopting IPCC guidelines, a framework that is globally accepted, data from different countries or regions can be compared and aggregated. It allows consistency and transparency to ensure accountability and help countries track progress over time and monitor the effects of land-use policies and interventions. In this report, we present a comprehensive framework-based training guide designed to enhance the accuracy and efficiency of land use classification in Singapore. Several machine learning algorithms, such as Classification and Regression Trees (CART), Random Forest (RF), and Support Vector Machines (SVM), are explored to compare and evaluate the performance of these algorithms in land classification.

# methods

## Classification Framework

This classification framework utilises Google Earth Engine (GEE), as its cloud-based processing allows for the quick processing of large datasets. This speed makes iterative analysis, experimentation with different classification algorithms, or validation processes more feasible. It also facilitates collaboration by allowing the sharing of scripts and results, leading to continuous improvements in classification techniques. Figure 1 summarises the whole flow chart of this framework.



Figure 1: Flow chart of the classification framework

GEE’s data catalogue allows easy retrieval of satellite imagery from the Sentinel-2 Harmonised collection. The analysis uses the spectral bands of B2, B3, and B4 with 10m resolution satellite imagery (Table 1) from the Sentinel-2 Harmonised collection, which was chosen for this analysis because it was recorded to have better classification performances than Landsat images (Bouslihim et al., 2022).

Thereafter, cloud removal techniques were applied and clipped to the boundary of Singapore. The remote sensing images were taken for 2019 and 2020, involving two Sentinel-2 scenes, and used as the classified images. Training data is defined from these images by visually examining and labelling areas to create corresponding feature collections, which are Forest Land, Cropland, Wetlands, Settlements, and Others. A total of 40 polygons were created for each category, so a total of 200 elements were used for this analysis. These datasets are then divided into training and validation sets of 70% and 30%, respectively, to assess the performance of the classification model. Machine learning algorithms such as Classification and Regression Trees (CART), Random Forest (RF), and Support Vector Machines (SVM) are explored. The guide provides insights into the strengths and limitations of each algorithm, enabling analysts to make informed choices based on their specific requirements.

Table 1: Sentinel-2A spectral bands

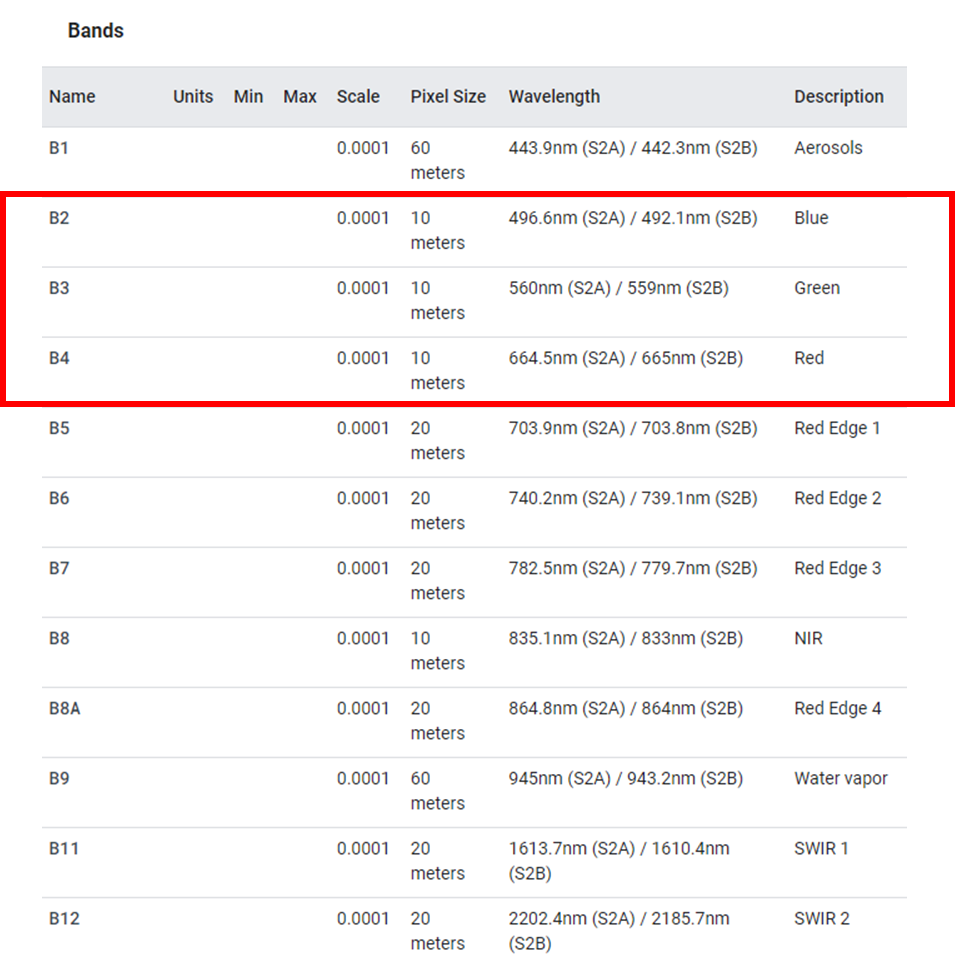


Table 1: Sentinel-2A spectral bands

Table 1: Sentinel-2A spectral bands

## Classification Algorithms

The three machine learning algorithms, such as Classification and Regression Trees (CART), Random Forest (RF), and Support Vector Machines (SVM), were used. CART is a decision tree-based method that segregates the dataset into subsets using a tree structure (Loh, 2011). For land classification, it examines satellite image pixel values and their attributes to determine decision rules, branching out until it classifies a land type. RF is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification (Belgiu & Drăguţ, 2016). It's highly accurate for land classification as it reduces overfitting by averaging results across numerous trees. SVM is a supervised learning algorithm that identifies hyperplanes in a multi-dimensional space that distinctly classify data points into land types (Mountrakis et al., 2011). It's effective for land classification, especially when the boundary between land types is clear and distinct.

# results

Table 2 shows the comparison of accuracy assessments for CART, RF, and SVM, based on the 2019 Sentinel 2 scene. CART has the highest training overall accuracy out of the three algorithms, at 99.99%, but an average validation overall accuracy of 75.23%. RF is performing well, with a high training overall accuracy of 98.35% and a high validation overall accuracy of 82.17%. The SVM model has a training accuracy of 75.4% and a validation accuracy of 74.87%. The validation accuracy was derived by calculating the confusion matrix of the trained model and then using it to compute the overall accuracy of the classification via Google Earth Engine. Tables 3, 4, and 5 show the error matrix of each classification and model to further analyse the classification accuracies.

Table 2: Accuracy assessment and comparison of CART, RF, SVM

|  |  |  |
| --- | --- | --- |
| Algorithms: | Training Overall Accuracy: | Validation Overall Accuracy: |
| CART | 0.9999733198153731 | 0.7523941532258065 |
| RF | 0.9835622306106605 | 0.82175 |
| SVM | 0.7540620581094422 | 0.7487399193548387 |

Table 3: Error matrix of CART classification

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Forest Land | Cropland | Wetlands | Settlements | Others | Total | Producer’s Accuracy |
| Forest Land | **5252** | 3 | 604 | 430 | 200 | 6489 | 80.93% |
| Cropland | 8 | **18** | 8 | 17 | 20 | 71 | 25.35% |
| Wetlands | 598 | 6 | **2341** | 61 | 486 | 3492 | 67.03% |
| Settlements | 406 | 34 | 78 | **2089** | 117 | 2724 | 76.68% |
| Others | 201 | 22 | 471 | 160 | **2242** | 3096 | 72.41% |
| Total | 6465 | 83 | 3502 | 2757 | 3065 | 15872 |  |
| User’s Accuracy | 81.23% | 21.68% | 66.84% | 75.77% | 73.14% |  |  |

Table 4: Error matrix of RF classification

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Forest Land | Cropland | Wetlands | Settlements | Others | Total | Producer’s Accuracy |
| Forest Land | **6063** | 2 | 231 | 151 | 148 | 6595 | 91.93% |
| Cropland | 2 | **9** | 7 | 31 | 29 | 78 | 11.53% |
| Wetlands | 583 | 0 | **2538** | 28 | 390 | 3539 | 71.71% |
| Settlements | 481 | 6 | 24 | **2111** | 109 | 2731 | 77.29% |
| Others | 118 | 1 | 412 | 99 | **2427** | 3057 | 79.39% |
| Total | 7247 | 18 | 3212 | 2420 | 3103 | 16000 |  |
| User’s Accuracy | 83.66% | 50% | 79.01% | 87.23% | 78.21% |  |  |

Table 5: Error matrix SVM classification

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Forest Land | Cropland | Wetlands | Settlements | Others | Total | Producer’s Accuracy |
| Forest Land | **6173** | 0 | 167 | 50 | 99 | 6489 | 95.13% |
| Cropland | 16 | **0** | 6 | 31 | 18 | 71 | 0% |
| Wetlands | 925 | 0 | **1777** | 235 | 555 | 3492 | 50.88% |
| Settlements | 763 | 0 | 17 | **1847** | 97 | 2724 | 67.80% |
| Others | 228 | 0 | 616 | 165 | **2087** | 3096 | 67.40% |
| Total | 8105 | 0 | 2583 | 2328 | 2856 | 15872 |  |
| User’s Accuracy | 76.16% | 0% | 68.79% | 79.33% | 73.07% |  |  |

# discussion

Metrics such as overall accuracy and class-specific accuracy are used to evaluate the model's effectiveness. The results show that the CART model can create very complex trees that almost perfectly classify the training data but fail to generalise to new data. Hence, there is a gap between training and validation accuracy, dropping from 99.99% to 75.23%. The Random Forest classifier has high absolute accuracy but has a 16% drop in validation accuracy, which might suggest the model might be overfitting the training data, capturing its noise and irregularities. On the other hand, Support Vector Machines are able to generalise better as they have a similar percentage of training and validation accuracy. This suggests that it might perform more consistently on various unseen datasets, making it potentially more robust in diverse real-world scenarios.

As it is crucial that the classification map will be able to compute land use as accurately as possible, the random forest would be a better option in this study. The test was repeatedly done for a 2020 image scene and derived similar results in accuracy, showing consistency in the model for the Random Forest model.

Overall, the Random Forest model also demonstrated better accuracy for the Forest Land, Wetlands, Settlements, and Others. It is also noted that many of the Cropland were misclassified as Settlements or Others and showing low accuracy for all three models.

# CONCLUSIONS

In conclusion, the Random Forest model is deemed to be more suitable for Singapore’s land classification due to its higher accuracy as compared to the CART and SVM models. In the future, the study can be repeated and refined to address classification errors and improve the overall quality of the land use map. Some methods could be looking into post-classification refinement techniques such as majority filtering, object-based post-classification, and spatial analysis to enhance the accuracy of the land use classification by removing noise, smoothing boundaries. Thus, with the use of remote sensing, users can perform stratified random sampling or use other systematic approaches to validate land use classifications, enhancing the accuracy assessment process.

# REFERENCES

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. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>

Bouslihim, Y., Kharrou, M.H., Miftah, A. Attou, T., Bouchaou, L., & Chehbouni, A. Comparing Pan-sharpened Landsat-9 and Sentinel-2 for Land-Use Classification Using Machine Learning Classifiers. J geovis spat anal 6, 35 (2022). <https://doi-org.libproxy1.nus.edu.sg/10.1007/s41651-022-00130-0>

Loh, W. Y. (2011), Classification and regression trees. WIREs Data Mining Knowl Discov, 1: 14-23. <https://doi.org/10.1002/widm.8>

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), 247-259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>

National Environment Agency (NEA). (2018). Fourth National Communication and Third Biennial Update Report. <https://unfccc.int/sites/default/files/resource/067382541_Singapore-NC4-BUR3-1-Singapore%20Fourth%20National%20Communication%20and%20Third%20Biennia.pdf>

National Environment Agency (NEA). (2020). Singapore's Fourth Biennial Update Report. <https://www.nccs.gov.sg/files/docs/default-source/default-document-library/2020-singapore-fourth-biennial-report.pdf>

Urban Redevelopment Authority (URA). (2023). <https://www.ura.gov.sg/Corporate/Resources/Publications/Skyline/Skyline-issue11/What-is-the-Master-Plan#:~:text=The%20Master%20Plan%20is%20one,plot%20of%20land%20in%20Singapore>