RESTORE+

A demonstration of national-scale thematically-detailed land cover mapping using automated methodology, cloud computing, and crowdsourcing in Indonesia

Algorithm Theoretical Basis Document

Technical Report

Supported by:

Federal Ministry for the Environment, Nature Conservation and Nuclear Safety

based on a decision of the German Bundestag

IIASA World Agroforestry Centre INPE ipea Instituto de Pesquisa Econômica Aplicada UN environment WCMC WWF MCC EDF WRI INDONESIA LSE

A demonstration of national-scale thematically-detailed land cover mapping using automated methodology, cloud computing, and crowdsourcing in Indonesia

Algorithm Theoretical Basis Document

International Institute for Applied Systems Analysis
Vienna, July 2022

The RESTORE+ project is implemented by the International Institute for Applied Systems Analysis (IIASA), World Agroforestry Centre (ICRAF), Brazil National Space Research Agency (INPE), Brazil Institute for Applied Economic Research (IPEA), UN Environment-World Conservation Monitoring Centre (UNEP-WCMC), World Resources Institute (WRI) Indonesia, World Wildlife Fund (WWF) Indonesia, Mercator Research Institute on Global Commons and Climate Change (MCC), Environment Defense Fund (EDF) and London School of Economics (LSE) Grantham Research Institute on Climate Change and the Environment.

The project is part of the International Climate Initiative (IKI). The German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) supports this initiative on the basis of a decision adopted by the German Bundestag.

Author:
Hadi, International Institute for Applied Systems Analysis

Publication Editing and Layouting:
Charlotte Kottusch, Enrico Confienza

Required Citation:
Hadi (2022). *A demonstration of national-scale thematically-detailed land cover mapping using automated methodology, cloud computing, and crowdsourcing in Indonesia*. RESTORE+ Technical Report, Vienna.

Table of Contents

1. Overview

This report describes the methodology to produce the national-scale land cover/land use map datasets for the given requirements of thematically-detailed national-scale restoration assessment in Indonesia under the RESTORE+ project. The methodology uses publicly available satellite datasets, Google Earth Engine cloud computing (Gorelick et al. 2017) as well as integrates crowdsourcing (citizen science) methodological component for generating training data and accuracy assessment data. The mapping source code (Earth Engine Javascript API) is openly available (link).

The methodology is designed to meet the following requirements:

- Thematic resolution. The classification scheme/typology (Table 1.1) was designed together with country partners to achieve the appropriate level of details for restoration assessment at national scale, and is compatible with existing classification schemes within the country.
- Spatial resolution. A spatial resolution of 100 meter (1 hectare) was determined, which is considered sufficient for the national scale restoration assessment, as well as matches the available reference land cover/land use map (henceforth, "country reference map"), which are the main source of the training data for classification, and which are readily used in the restoration analysis.
- Temporal resolution. A temporal resolution of annual i.e., 1-year time scale was determined. The current resulting map datasets were demonstrated for one year i.e., 2018.
- Integrates crowdsourced training data. To integrate the crowdsourced training data, two classification scenarios were carried out, one with all classes relying on training data from the reference map (henceforth, "detailed-legend classification"), and another with the simplified-legend classes crowdsourced to the general public (henceforth, "simplified-legend classification"). This was done to avoid "contaminating" the potentially higher-quality training data obtained via crowdsourcing (human verification) with the more uncertain training data derived directly from the reference map, as well as for practical reason i.e., timing of which the data became available at the time. The crowdsourced classes were Undisturbed Forest, Logged Over Forest, Oil Palm Monoculture, Tree Based Not Oil Palm, Cropland, Shrub, and Grass/Savanna. Details of the crowdsourcing campaign and dataset will be available in a separate publication (under review).
- Explicit account of resulting map thematic (classification) uncertainty, as well as modular design for easy updates of class-specific training data. A key design of the classification methodology was the decomposition of the otherwise multiclass classification into separate binary (i.e., one-versus-others) classification scenarios, and producing class probability for each land cover/land use class as output. These layers of class probability were then combined by incorporating expert rules informed by feedbacks from local experts, to arrive at the final most-confident class label, with the associated per-pixel uncertainty. The per-pixel uncertainty layers include the "least confidence" uncertainty (calculated as one minus the class probability of the "primary classification" i.e., most-likely or classified highest-probability class label), and the "margin of confidence" (Monarch 2021) uncertainty (calculated as the difference between the class probability of the "primary classification" and the class

probability of the "secondary classification", the latter being the second-most-likely or second-highest-probability class label (Brown et al. 2020).

The map datasets generated from the present methodological demonstration include:

- Pre-processed covariates/features mosaics
- Training data (from reference maps, and from crowdsourcing campaign with the participation of the general public in Indonesia)
- Validation/accuracy assessment data (from expert workshops with local experts in Indonesia)
- Predicted class probability maps for each class
- Preliminary final most-confident class label, with the application of expert rules, for "primary classification" and "secondary classification", as well as the per-pixel classification uncertainty

In addition, an interactive web-based application ([Google Earth Engine app]) to collect training data with the guidance of on-the-fly classification results was built to facilitate training data collection in a collaborative approach engaging local experts from different regions following the approach of Souza et al. 2020.

Table 1.1 The land cover/land use classification scheme required for the map, and the definition of the classes, as provided by the local experts.

Class ID ("Detailed Class")	Class name (Indonesian)	Class name (translated)	Class definition (Indonesian)	Class definition (translated)
1	"Hutan lahan kering primer"	Undisturbed Dryland Forest	"Tutupan hutan alami dengan kanopi yang rapat (>80%), spesies yang sangat beragam dan basal area yang relative tinggi. Secara mudah, hutan ini diindikasikan tidak adanya jalan logging. Pada citra satelit, diindikasikan dengan nilai index vegetasi dan band infrared yang tinggi, dan band tampak (visible) yang rendah."	"Natural forest cover with dense canopy (>80%), very diverse species, and relatively high basal area. They are indicated by the absence of logging roads. In the satellite image, they are indicated by high values of vegetation index and infrared band reflectance, and low values of visible bands reflectance."
2	"Hutan lahan kering sekunder"	Logged-Over Dryland Forest	"Tutupan hutan alam dengan kerapatan pohon yang bervariasi (30% - 80%) yang telah mengalami gangguan aktivitas manusia maupun aktivitas alam lainnya yang biasanya dicirikan oleh adanya jalan logging maupun bekas tebangan."	"Natural forest cover with varying tree density (30% - 80%) that has experienced disturbances from human activities or other natural activities, usually characterized by the presence of logging roads or logging signs/marks."
3	"Hutan mangrove primer"	Undisturbed Mangrove Forest	"Tutupan hutan yang didominasi oleh pohon bakau yang berlokasi pada pesisir pantai dan tidak pernah mengalami penebangan maupun aktivitas manusia lainnya."	"Forest cover dominated by mangrove ('bakau') trees, located on the coast, and has never experienced logging or other human activities "
4	"Hutan mangrove sekunder"	Logged-Over Mangrove Forest	"Tutupan hutan yang didominasi oleh pohon bakau yang telah mengalami degradasi berlokasi di sekitar pesisir pantai dan pernah mengalami penebangan	"Forest cover dominated by mangrove ('bakau') trees that has experienced degradation, located around the coast, and

			maupun aktivitas manusia lainnya.”	has experienced logging or other human activities.”
5	“Hutan rawa primer”	Undisturbed Swamp Forest	“Tutupan vegetasi alami yang berada pada lahan basah yang tergenang sementara maupun permanen, tidak pernah mengalami penebangan masalalu ataupun pengaruh aktivitas manusia yang biasanya dicirikan dengan mudah tidak adanya jalan logging, parit, maupun kanal.”	“Natural vegetation cover that is situated on wetlands that are temporarily or permanently inundated, that has never experienced past logging or impacts of human activities that is usually characterized by the absence of logging roads, ditches, or canals.”
6	“Hutan rawa sekunder”	Logged-Over Swamp Forest	“Tutupan vegetasi alami yang berada pada lahan basah yang tergenang sementara maupun permanen, yang telah mengalami penebangan masalalu ataupun pengaruh aktivitas manusia yang biasanya dicirikan dengan mudah dengan adanya jalan logging, parit, maupun kanal.”	“Natural vegetation cover that is situated on wetlands that are temporarily or permanently inundated, that has experienced past logging or impacts of human activities that is usually characterized by the presence of logging roads, ditches, or canals.”
7	“Agroforestri”	Agroforestry	“Tutupan vegetasi yang terdiri dari campuran komoditas perkebunan, buah-buahan, pohon berkayu, maupun tanaman semusim lainnya yang tumbuh secara bersama-sama dalalam satu waktu atau rotasi di dalam satu lahan.”	“Vegetation cover consisting of a mixture of estate/plantation commodities, fruits, woody trees, or other seasonal plants/crops that grow together at one time or rotation within one land area/field.”
8	“Hutan tanaman”	Plantation Forest	“Tutupan vegetasi pohon berkayu baik kayu lunak maupun keras yang ditanam secara monokultur untuk tujuan komersial. Hutan tanaman dalam skala luas biasanya dijalankan oleh pemegang konsesi, sedangkan dalam skala kecil biasanya dikelola oleh masyarakat local. Komoditasnya biasanya akasia, eucalyptus, pinus, jati, sengon, damar, dan lain-lain.”	“Vegetation cover of woody trees, both softwood and hardwood, which are planted in monoculture for commercial purpose. Plantation Forest in a large scale are usually run by concession holders, whereas in a small scale usually managed by local communities. The commodities are usually acacia, eucalyptus, pine, teak, sengon, resin, and others.”
9	“Karet monokultur”	Rubber Monoculture	“Tutupan vegetasi yang didominasi hampir 100% oleh tanaman karet dalam satu lahan.”	“Vegetation cover that is dominated almost 100% by rubber trees within one area/field.”
10	“Kelapa sawit monokultur”	Oil Palm Monoculture	“Tutupan vegetasi yang didominasi hampir 100% oleh tanaman kelapa sawit dalam satu lahan.”	“Vegetation cover that is dominated almost 100% by oil palm trees within one land area/field.”
11	“Monokultur lain”	Other Monoculture	“Tutupan vegetasi yang didominasi hampir 100% oleh tanaman perkebunan lain selain	“Vegetation cover that is dominated almost 100% by monoculture/plantation crops,

			karet dan kelapa sawit dalam satu lahan."	other than rubber or oil palm, within one land area/field."
12	"Rerumputan/padang rumput/savanna"	Grass or Savanna	"Tutupan vegetasi yang didominasi oleh tutupan rumput yang tumbuh secara alami.	"Vegetation cover dominated by grass cover that grows naturally."
13	"Semak belukar"	Shrub	"Tutupan vegetasi yang didominasi oleh vegetasi bukan pohon dengan ketinggian tidak lebih dari 5-6 m, biasanya hasil dari perladangan berpindah, atau bekas tebangan yang telah terdegradasi, maupun suksesi alami dengan tegakan pohon yang belum terbentuk yang berumur 2-3 tahun."	"Vegetation cover dominated by non-tree vegetation with a height of not more than 5-6 m, usually the result of shifting cultivation, or ex logging areas that have been degraded, or natural succession with yet-to-form tree stands that are 2-3 years old."
14	"Lahan pertanian"	Cropland	"Tutupan vegetasi yang didominasi oleh tanaman padi (sawah), atau tanaman palawija maupun hortikultura (pertanian lahan kering)."	"Vegetation cover dominated by rice (paddy), or "palawija" or horticultural plants/crops (dryland agriculture)". "Palawija" crops: secondary food crops i.e. usually grown after main crop i.e. rice, on dryland. Examples are groundnuts, maize, cassava, soybean, and roots/pulses. Horticultural crops: fruits, vegetables, herbs, medicinal and ornamental plants.
15	"Permukiman"	Settlement	"Tutupan yang didominasi oleh bangunan."	"Surface cover dominated by buildings."
16	"Lahan terbuka"	Cleared Land	"Tutupan lahan yang hampir 100% tanpa tutupan vegetasi."	"Land cover that is almost 100% without vegetation cover."
17	"Tubuh air"	Waterbody	"Tutupan yang didominasi oleh air, biasanya direpresentasikan dengan sungai, danau, waduk, tambak, dll."	"Surface cover dominated by water, usually represented as rivers, lakes, reservoirs, ponds, etc."

2. Classification Process

The classification process consists of design of classification scenarios to fit the mapping data product and mapping methodological requirements, training data preparation and collection, satellite remote sensing data preparation and feature engineering (including ancillary datasets), supervised classification model calibration (training) to produce class probability maps, assembly of class probability maps into final most-confident class label (and per-pixel classification uncertainty measures) based on expert rules and integration of crowdsourcing data, and a preliminary validation (accuracy assessment) of the final detailed-legend land cover/land use map.

Two classification scenarios were carried out, one for all detailed classes with the training data from the “country reference map” (“detailed-legend classification”), and one for a simplified class scheme with training data from crowdsourcing (“simplified-legend classification”). In the “detailed-legend classification”, as the available-to-use “country reference map” was produced for 2010, thus necessitating to use satellite data available for 2010 for temporally-consistent classification model training, followed by application of the trained model for the target demonstration year namely 2018. Further, due to the complex class typology required and the landscape heterogeneity in the vast land of the country, as recommended by local experts, the classification model calibration and application were applied separately for seven regions based on landscape characteristics namely Sumatera; Kalimantan; Java Madura Bali; Sulawesi; Nusa; Maluku; and Papua. An exception was for Rubber Monoculture class, for which an alternative training data source was used (Chapter 2.1), precluding model training and application by region. Model spatial stratification by using additionally ecoregion did not produce clearly better results in early experiments.

In the “simplified-legend classification”, in addition to the crowdsourced land cover/land use classes (Undisturbed Forest, Logged Over Forest, Oil Palm Monoculture, Tree Based Not Oil Palm, Cropland, Shrub, and Grass/Savanna), to create a spatially-complete map, classes were added to the classification scenario namely Waterbody, Settlement, and Cleared Land. The crowdsourcing activities interpret very high (sub-meter) spatial resolution image (VHSR) chips acquired in 2018, and thus classification model training and application were both based on satellite data for 2018. The classification model training and application was applied for the whole country i.e., not stratified by region as in the “detailed-legend classification” to ensure adequate quantity, representativeness, and class-balance (allocation) in the training data.

2.1 Training Data

The training data for the supervised classification was collated from our country partner i.e. generated from their land cover product (which mapped various classes but targeted specifically agroforestry), as well as from crowdsourcing initiatives (details in separate publication under review). For the “detailed-legend classification”, from the “country reference map”, 10000 training sample points (pixels) per class, per region, and per binary classification scenario (Chapter 2.3) were generated. The “detailed-legend” training sample is available as Earth Engine feature collection assets. To minimize class label error in sampled training points, the “country reference map” was beforehand filtered to remove pixels with patch size smaller than 20th percentile patch size for the same class and region, as well as pixels where tree cover loss was detected between 2010 and 2018 based on Global Forest Change dataset (Hansen et al. 2013). In addition, as using training data from the "country reference map" was observed to produce substantially overestimated area for Rubber Monoculture, alternative training data for Rubber Monoculture was obtained based on WRI Tree Plantation database (Harris, Goldman, and Gibbes 2019).

For the "simplified-legend classification", the crowdsourced training data were quality-filtered. There are different ways to filter crowdsourcing data, and in this version, different filters were tested in terms of number of annotations (unique-annotators) of either 3, 4, 5 or 6, and in terms of the weighted majority score threshold of 0.5 or 0.8 (0 being consensus "no" answer by the crowd i.e. rejecting the class label, 1 being consensus "yes" i.e. confirming the class label of the given sample, the weights being the annotator's reliability score calculated based on expert agreement i.e. their agreement with expert-annotated control set (details in separate publication under review)). Based on inspecting the total area and spatial distribution of the classes countrywide, minimum number of annotations of six, and weighted majority score threshold of 0.8 were determined to filter the crowdsourced data for high-quality training sample. To complete the class scheme, training data for Waterbody, Settlement, and Cleared Land, were obtained, respectively, from the JRC Global Surface Water layer (Pekel et al. 2016), Facebook High Resolution Settlement layer (CIESIN 2016), and the "country reference map". The "simplified-legend" training sample is available as Earth Engine feature collection assets.

2.2 Classification Input Feature Space

As the "country reference map" was available for 2010, satellite remote sensing datasets from Landsat and ALOS-PALSAR archive were the primary covariates (features) datasets.

For multispectral optical Landsat data, all images of Landsat 5 [L5] Thematic Mapper (TM), Landsat 7 [L7] Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 [L8] Operational Land Imager (OLI) 30m surface reflectance product (Collection 1) produced by NASA and USGS and available in Google Earth Engine were used. In prioritizing spatial coverage completeness using annual observations, all scenes with any amount of scene-level cloud cover spanning the whole one-year period were used, which amount to a total of 253 path-rows, representing 5978 Landsat images (2423 ETM+ images, 3555 OLI images) for 2018, and 2040 Landsat images (405 TM images, 1635 ETM+ images) for 2010, were processed in Google Earth Engine platform. Image pre-processing were applied namely edge masking (5.5 km), clouds and cloud shadows masking (using accompanying Landsat QA i.e., FMask), BRDF correction and terrain correction (Poortinga et al. 2019). OLI sensor data were harmonized with respect to TM and ETM+ sensors (Roy et al. 2016). Spectral indices were derived from the surface reflectance, namely Normalized Difference Vegetation Index (NDVI = (NIR - red) / (NIR + red)) (USGS n.d.), Normalized Difference Water Index (NDWI = (NIR - SWIR1)/(NIR + SWIR1)) (Sinergise n.d.), Normalized Burn Ratio (NBR = (NIR - SWIR2)/(NIR + SWIR2)) (USGS n.d.), Soil-Adjusted Vegetation Index (SAVI = (1 + L) × (NIR - red)/(NIR + red + L); L = 0.9) (USGS n.d.), and Enhanced Vegetation Index 2 (EVI2 = (2.5 × (NIR - red)/(NIR + 2.4 × red + 1)) (USGS n.d.). Spectral indices based on fractional cover from spectral mixture analysis were tested and were found to have minor importance (classifier's feature importance) in early experiments, which however might suggest the need for locally-retrieved endmembers. The per-pixel time series were then temporally-reduced into median and different percentile values (Table 2.1), the latter to capture temporal signature potentially helpful for classes with intra-annual reflectance dynamic. Compositing with smaller time window e.g., seasons or months were found to result in excessive no data areas due to the persistent cloud cover in the region. In total, 36 features (bands) were derived from Landsat data, including the original reflectance bands. Texture features derived from the panchromatic band were attempted but judged to be visibly too noisy to be useful in automated classification.

Table 2.1 Input feature space for land cover/land use classification model: radiometric features

Platform/sensor	Year	Features	Remark about generated asset
Landsat 5 (TM), Landsat 7 (ETM+), Landsat 8 (OLI)	2010, 2018	Annual median and percentiles (10th, 25th, 75th, 90th) of red, green, NIR, SWIR1, and SWIR2 surface reflectance; annual median of blue surface reflectance; annual median of spectral indices (NDVI, NDWI, NBR, SAVI, and EVI2)	Spatial resolution 30m; values were multiplied by scaling factor 10000.
ALOS PALSAR	2010, 2017	Data readily available as annual mosaic; HH backscatter, HV backscatter, derived HH/HV backscatter ratio, HH/HV texture features (contrast, variance, entropy, correlation, sum average)	*Texture*: To minimize data size, contrast and variance values were rescaled based on estimated min and max, then multiplied by 65535; Entropy was multiplied by 10000; Correlation was multiplied by 10000 add added by 20000.
Sentinel-1 (not used in final classification in the current version)	2018	Annual percentiles (10th, 20th, 30th, 40th, 60th, 70th, 80th, 90th) of VV and VH backscatter; seasonal (rainy or dry) mean and median of VV and VH backscatter; texture features (contrast, variance, entropy, correlation, sum average) applied to VV/VH backscatter ratio	*Annual* and *seasonal composites/features*: values were multiplied by scaling factor 100 *Texture*: same scaling and offset were applied as for ALOS PALSAR texture above. *Annual* and *seasonal composites/features* were spatially resampled to Landsat 30m grid. *Texture features* were computed at 10m.

For L-band radar ALOS-PALSAR data, the global 25m PALSAR/PALSAR-2 mosaic available in Google Earth Engine was used. No data, radar layover, and radar shadowing values were removed based on the QA layer. HH/HV backscatter ratio was derived from the HH and HV backscatter. Texture features (GLCM) were derived from the HH/HV ratio, with neighbourhood size of 11-by-11 pixels. In total, 8 features were derived from this dataset.

Sentinel-1 C-band radar data were tested but in the present investigation were found to not provide clear improvements in the final land cover/land use map considering all detailed-legend classes, and thus not included in the current version. However, their use for a targeted area and specific land cover types remains for future investigations, and the available Sentinel-1 features datasets are described here. Sentinel-1 data available in Google Earth Engine were used. Scenes with instrument mode IW and descending pass, with acquisition in the target mapping demonstration year i.e., 2018 were used. Minor preprocessing to remove edges and stripes were applied. Speckle filtering was not applied, as multitemporal averaging helps to reduce speckle noise. Two types of compositing were performed namely annual compositing and seasonal (rainy season and dry season) compositing. Both composites were resampled to Landsat grid (30m). The annual compositing provides temporal features in terms of 10th, 20th, 30th, 40th, 60th, 70th, 80th, and 90th percentile values from the temporal distribution. Monthly compositing was found to result in considerable missing data in some parts at national scale. The seasonal compositing reduces the data into their temporal mean and median. Both VV and VH backscatters were used, and VV/VH backscatter ratio was derived. Finally, texture features (GLCM) were derived from VV/VH backscatter dry-season median composite, at 10m resolution, with neighbourhood size of 11-by-11 pixels. In total, 29 features were derived from Sentinel-1.

In the expert interpretation workshop with the local experts, it was observed that the expert interpreters relied on other diagnostics than visual appearance of the sample in the image, namely location of the sample with respect to roads (in particular to determine logged-over forests versus undisturbed forests (see examples), rivers, and coasts. As part of the efforts to design automated features that represent the local expert interpreters human heuristics, a set of ancillary (non-radiometric) features (Table 2.2) were included in the feature space. The ancillary features include topographic variables (SRTM elevation, and derived slope and aspect), distance to main roads (), distance to all roads (based on combined Open Street Map roads and road map in national topographic database(BIG n.d.)), distance to human settlement (High Resolution Settlement Layer by Facebook Research (CIESIN 2016)), distance to rivers (based in river map in national topographic database (BIG n.d.)), and distance to coast (based on Natural Earth coastline map (Natural Earth n.d.)). It is recognized that a possible issue is that these ancillary maps are typically not updated i.e. the ancillary features are year-invariant, however these features were considered critical for classes in the detailed-legend that likely could not be classified solely based on spectral separability (as indicated by the visual interpretation process demonstrated by the local experts mentioned above).

Altogether, a total of 53 features were used in classification. Feature selection was not performed as tests using BORUTA random forest feature selection algorithm led to all or almost all features deemed important in each binary classification scenario (Chapter 2.3) and possible impact of label noise in training sample. A special case was for the binary classification scenario with Rubber Monoculture training sample from WRI Tree Plantation database, as the training sample is not spatially exhaustive (not all plantations are possibly mapped) unlike the case of sampling training sample from the country reference map, it was necessary to use only radiometric features (not including ancillary features) for this particular binary classification scenario.

Table 2.2. Input feature space for land cover/land use classification model: ancillary features

Feature/covariate	**Reference data source**
Elevation	SRTM: USGS/SRTMGL1_003 (Earth Engine data catalog)
Slope	SRTM:. USGS/SRTMGL1_003 (Earth Engine data catalog)
Aspect	SRTM:. USGS/SRTMGL1_003 (Earth Engine data catalog)
Distance to main roads	National topographic database (RBI)
Distance to all roads	National topographic database (RBI) and Open Street Map
Distance to river	National topographic database (RBI)
Distance to settlement	National topographic database (RBI); Facebook High Resolution Settlement Layer
Distance to coast	Natural Earth coastline

2.3 Classification Strategy and Algorithm

For the classification strategy, building a classifier to separate all the detailed-legend classes at once faces the challenge of unclear class separability (decision boundary) due to substantial intra-class landscape heterogeneity and thus intra-class spectral variability higher than inter-class variability, unavoidable label noise/errors in the training data from imperfect reference map and imperfect crowdsourced labels, and residual noise (from imperfect image preprocessing e.g., cloud masking) in the features. Early experiments with hierarchical classification (classifying four classes namely Forest, Tree Based, Not Tree Based, and Not Vegetation at first level, then classifying detailed classes for each of the four classes at higher levels) were judged to provide substantial errors at first level which would then propagate (remain uncorrected) to the classification of the higher level legends. Therefore, the approach of decomposing the multiclass classification problem into binary (one-vs-others) classification for each class, as demonstrated in (Saah et al. 2020), was adopted. This binary classification approach also has the advantage of being modularized to allow easy update of any class individually as more/better training data or/and map for that class or/and image preprocessing and engineered features or/and expert rules becomes available, as well as facilitates custom integration of the per-class probability maps into custom multiclass legend (typology) as demonstrated in (Saah et al. 2020).

For the classification algorithm, Random Forest algorithm available in Google Earth Engine was used. Based on early classification experiments, to maximize test set accuracy the hyperparameters of the Random Forest classifier were set as follows: number of decision trees to create equals 100, number of variables per split equals one-third of the total number of features, and minimum leaf population (i.e., the number of samples at a node to stop further splitting and make that node) equals 5. The output of the Random Forest classification is the class probability in each binary classification scenario (per class and per region for “detailed-legend” classification, per class for “simplified-legend” classification). The class probability is the proportion of trees in the forest that have chosen this label.

The final most confident hard class label is subsequently determined using “Winner-Takes-All” strategy (Duan et al. 2003) which assigns it to the class with the largest predicted probability (related to decision function value). Note here the probability is from separate binary classifiers and thus strictly speaking does not represent multiclass probability; it is however considered acceptable as the interest here is in the class which the classifier is most confident about in separating that class from the rest of the classes (i.e., of importance is the rank, instead of absolute value of confidence). This was done partly as multiclass probability output in Google Earth Engine platform was not available at the time of the study, which is the case now and hence remains for future investigation.

To integrate class probability maps from the “detailed-legend classification” and “simplified-legend classification”, a sequence of integration rules was defined. Expert rules are typically necessary in automated classification of a complex land cover typology (Buchhorn et al. 2020). The rules were determined based on preliminary feedback on the map, from local experts regarding apparent systematic misclassification of certain classes, and the classes in the “simplified-legend classification” that were assessed to be acceptably reliable (through visual examination of the spatial abundance and distribution of the classes, and comparison of the class area with respect to reference map). The expectation is that the “simplified-legend classification” is more accurate because the labels of the training sample are verified by humans i.e., the crowd annotators, provided that the class is relatively easy to be visually interpreted by likely inexperienced interpreters. The “simplified-legend” classified map then was used to correct the “detailed-legend” classified map. The highest probability class (primary classification) and the second highest probability class (secondary classification) from the “detailed-legend classification” (training data from “country reference map”) alone, although methodologically more straightforward, were assessed to be not acceptably reliable. For areas superimposed with most likely (maximum probability) class in the “simplified-legend classification” incorporating crowdsourced training data, no secondary classification was currently assigned as the “simplified-legend classification” was done for coarse classes and therefore the next highest probability class was not appropriate for the “detailed-legend” map intended.

The current version of integration rules are as follows. The integration rules were applied separately for the highest probability class (primary classification), and for the second highest probability class (secondary classification).

- Rule for forest classes: It was observed that some large forest areas were misclassified as Plantation Forest. To correct this, “Simplified-legend” predicted “forests” (“undisturbed forest” and “logged over forest” classes were merged) which overlays with GLAD Primary Humid Tropical Forest 2001 (Turubanova, Tyukavina, and Hansen 2018) were considered to be high-confidence forest prediction. Pixels in “detailed-legend classification” that were not predicted as forest class but fall in the high-confidence forest map were then corrected as the forest class (class id 1 to 6 in “detailed legend”) which has the highest predicted probability.
- Rule for Oil Palm Monoculture: Oil Palm Monoculture in “simplified-legend” classified map was superimposed on top of the “detailed-legend” classified map. Oil Palm Monoculture pixels in the “detailed-legend” classified map that were not Oil Palm Monoculture in the “simplified-legend” classified map were replaced with the next highest probability class (“detailed-legend classification”), excluding from the possible class the classes for which expert rules were applied (i.e., Oil Palm Monoculture, natural forest classes (class id 1 to 6 in detailed legend), Cropland, Settlement, and Waterbody.
- Rule for Cropland: similar to the rule for Oil Palm Monoculture above, but for Cropland.
- Rule for wetland forest classes: wetland forest pixels predicted in the “detailed-legend” classified map which fall outside the Wetlands landform (link) had their labels replaces with the next most probable class (“detailed-legend classification”), excluding from the possible class the classes for which expert rules were applied (i.e., Oil Palm Monoculture, natural forest classes (excluding wetland forest), Cropland, Settlement, and Waterbody.
- Rule for Waterbody: similar to the rule for Oil Palm Monoculture and Cropland above, but for Waterbody prediction in “simplified-legend” classified map.
- Rule for Settlement: similar to the rule for Oil Palm Monoculture, Cropland, and Waterbody above, but for Settlement prediction in the “simplified-legend” classified map.

In the absence of large sample size of ground truth labels to create spatial accuracy map (which is a true indicator of classification quality), the class probability is an alternative to provide a measure of classification uncertainty (or confidence) on a per-pixel basis (Canters 1997), which is useful in that it depicts the spatial structure of the uncertainty. Two common measures of uncertainty namely “least confidence” and “margin of confidence” (Monarch 2021). It has been shown that the classifier confidence is related to classification accuracy for high confidence values (Inglada et al. 2017). “Least confidence” uncertainty was calculated as one minus the class probability of the “primary classification” i.e., most-likely or classified highest-probability class label), and the “margin of confidence” uncertainty was calculated as the difference between the class probability of the “primary classification” and the class probability of the “secondary classification”, the latter being the second-most-likely or second-highest-probability class label. Note that the classification uncertainty, and in turn confidence, here was based on binary classifier confidence and not from a single probability distribution with all the classes as possible outcome considered, and thus they should be treated qualitatively, and that the “margin of confidence” is more reliable than “least confidence”; in future works the classification uncertainty based on a single multiclass probability distribution constructed from the binary classification probabilities (Duan et al. 2003) is to be investigated.

3. Practical Considerations, and Future Improvements

The presented methodology serves as a pilot demonstration of national-scale land cover/land use mapping in Indonesia targeting highly detailed typology required for the purpose of restoration assessment under the RESTORE+ project, with a methodology designed to employ cloud computing (Google Earth Engine platform) and crowdsourced training data and to be modularized, as well as to produce per-pixel classification uncertainty. A main limitation has been high quality (low rate of label errors) training sample datasets that satisfy the class legend requirements and are representative of the whole mapping domain i.e., the large classification region, with the vast spatial heterogeneity in natural landscape features and land management/use practices. Thus, it should be acknowledged that the full potential of the methodology likely has not been realized, and further improvements (iterations) to this first version of the results are needed as a stepwise approach, which are hopefully facilitated after the datasets, code, and tool (app) are made available. Particularly, the machine learning feedback aided interactive training data digitization tool (link) can be helpful to efficiently collect training sample polygons targeting certain problematic classes or areas of systematic misclassification, which is best done in a collaborative manner with a network of local/regional land cover experts. These training data can then be used to general an updated version of class probability for the classes or/and areas of interest.

Given the challenge for machine learning to map the detailed legend in the present work, among others, and that validation/accuracy assessment of the map is ongoing, caution needs to be exercised in the fit-for-purpose-ness usage of the current version of the result. It is recommended to use local reference data to estimate unbiasedly the class area and their uncertainties (confidence interval) in a targeted area to account for unavoidable misclassification errors, following good practices (Olofsson et al. 2014). In the absence of reference data for the area of interest, the predicted class probability may be used to estimate the class area and uncertainties via Monte Carlo simulation procedure in which multiple realizations of land cover/land use maps were generated (Canters 1997).

Considering the method of determining the final most-confident class label, based on the class probability predicted by the set of binary classifiers, a simple approach of considering the maximum probability (thus highest-ranked in likelihood) class as the final hard class label was chosen. That is, the trained classifier is most confident about that class in their separation from the rest of the classes. A better approach, requiring local reference data for a targeted area, would be to train a secondary supervised classifier such as decision tree to learn the optimal mapping between the class probabilities and the ground truth hard class labels. In the absence of local reference data, the decision tree to combine the class probabilities (sequence of rules with the classes and thresholds on their class probability values) into final hard class label can be determined via expert-defined rules incorporating local knowledge and visual examination (Saah et al. 2020). Reference information of class area such as government statistics can also be used in calibrating the class probability threshold.

A special remark of caution is needed for the important separation between undisturbed forest classed and the logged-over forest classes. It was observed from the local-expert interpretation activities, that in inferring whether a forest sample is undisturbed or logged, the signs of human activities needed to be searched within a large radius from the forest sample location, and the class determination decision was augmented by the local experts' knowledge of the legal status of the forest estate (which map is unavailable/inaccessible). As a fixed or variable-within-strata (regions) criterion such as distance to roads could not be determined in local-experts group discussion, here the undisturbed vs logged-over forests classification relied on a data mining approach, where the machine learning model learns the implicit rules (pertaining to covariates such as distance to roads) in the country reference map in which the map encodes the human rules of the local expert who supervised the map production. Logging practices are expected to vary, and here the implicit rules were learned separately for each of the seven classification regions; more detailed spatial stratification of the model might be necessary, however the extent the logging practices vary in space is not known well to inform the stratification. Note however the spectral response may play a role in discriminating between logged-over forests and undisturbed forests.

The crowdsourced training and validation data is documented, evaluated, and to be released in a separate publication (under review). The data contains additional label quality indicators including the annotator's control set agreement, inter-annotator agreement, and intra-annotator agreement, which were used to estimate the label uncertainty. Experiments on how to best utilize the crowdsourced training data for downstream impact on classifier quality, through different quality filtering (trade-off between label quantity and geographical diversity on one hand, and label overall noise level on the other hand), remains to be investigated.

Opportunities to improve the map also exist in the image pre-processing and feature engineering steps. In the current version, the Landsat compositing time window was limited to one calendar year, which was partly for computational reason and for design towards annual update frequency; combining data from adjacent years (3-year window; observations can weighted by year) and filtering out scenes with majority cloud cover, should result in "cleaner" (from residual unmasked clouds and shadows) composite image features. Testing the methodology for later years when Sentinel-2 surface reflectance product became available (e.g., from 2020 onwards) can be of interest as the temporal density of the observations will be much higher, however cloud masking is a major challenge with Sentinel-2, and the training sample needs to be filtered for land cover change. In a collaborative setting, the Landsat composites can be made spatially optimized, such as by experimenting through the composite explorer (Landsat and Sentinel-2) app (link). To note is that Landsat Collection 1 used in the current version is currently deprecated and thus the methodology should migrate to use the improved Landsat Collection 2 also readily available in Google Earth Engine platform. Sentinel-1 ARD with additional important radiometric terrain normalization step now possible to be generated in Google Earth Engine platform (Mullissa et al. 2021) presents an untapped opportunity. In the postprocessing step, a simple 3-by-3 pixels spatial window majority filter was applied, a more refined (requiring local landscape knowledge for it to be justified) postprocessing by class-specific or/and region-specific land cover patch minimum size filter can be recommended. Nevertheless, it is foreseen that improving the training sample (as was demonstrated in the use of WRI Tree Plantation database for Rubber Monoculture training data; or the more reasonable spatial distribution of cropland probability based on crowdsourced training data as compared to training data labels directly from the reference map) for all the required classes, or/and the expert rules in integrating the class probability layers, would be the most impactful effort in terms of improving map accuracy, to be given highest priority in future iterations of the product.

References

BIG. n.d. “Indonesia Geospatial Portal.” Pusat Pengelolaan Dan Penyebarluasan Informasi Geospasial Badan Informasi Geospasial (BIG). Accessed July 19, 2022. https://tanahair.indonesia.go.id/portal-web.

Brown, Jesslyn F., Heather J. Tollerud, Christopher P. Barber, Qiang Zhou, John L. Dwyer, James E. Vogelmann, Thomas R. Loveland, et al. 2020. “Lessons Learned Implementing an Operational Continuous United States National Land Change Monitoring Capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) Approach.” *Remote Sensing of Environment* 238 (July): 111356. https://doi.org/10.1016/j.rse.2019.111356.

Buchhorn, Marcel, Myroslava Lesiv, Nandin Erdene Tsendbazar, Martin Herold, Luc Bertels, and Bruno Smets. 2020. “Copernicus Global Land Cover Layers-Collection 2.” *Remote Sensing* 12 (6): 1–14. https://doi.org/10.3390/rs12061044.

Canters, Frank. 1997. “Evaluating the Uncertainty of Area Estimates Derived from Fuzzy Land-Cover Classification.” *Photogrammetric Engineering and Remote Sensing* 63 (4): 403–14.

CIESIN. 2016. “High Resolution Settlement Layer.” Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). 2016. https://ciesin.columbia.edu/data/hrsl/.

Duan, Kaibo, S. Sathiya Keerthi, Wei Chu, Shirish Krishnaj Shevade, and Aun Neow Poo. 2003. “Multi-Category Classification by Soft-Max Combination of Binary Classifiers.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2709: 125–34. https://doi.org/10.1007/3-540-44938-8_13.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. “Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone.” *Remote Sensing of Environment*, Big Remotely Sensed Data: tools, applications and experiences, 202 (December): 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Hansen, Matthew, Peter V. Potapov, Rebecca Moore, Matt Hancher, Svetlana Turubanova, Alexandra Tyukavina, David Thau, et al. 2013. “Hansen Global Forest Change v1.9 (2000-2021) | Earth Engine Data Catalog.” Google Developers. 2013. https://developers.google.com/earth-engine/datasets/catalog/UMD_hansen_global_forest_change_2021_v1_9.

Harris, Nancy, Elizabeth Dow Goldman, and Samantha Gibbes. 2019. “Spatial Database of Planted Trees (SDPT Version 1.0),” November. https://www.wri.org/research/spatial-database-planted-trees-sdpt-version-10.

Inglada, Jordi, Arthur Vincent, Marcela Arias, Benjamin Tardy, David Morin, and Isabel Rodes. 2017. “Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series.” *Remote Sensing* 9 (1): 95. https://doi.org/10.3390/rs9010095.

Monarch, Robert (Munro). 2021. *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI*. *Manning Publications*. MEAP Versi. Manning Publications.

Mullissa, Adugna, Andreas Vollrath, Christelle Odongo-braun, Bart Slagter, Johannes Balling, Yaqing Gou, Noel Gorelick, and Johannes Reiche. 2021. “Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine,” 5–11.

Natural Earth. n.d. “Natural Earth » Blog Archive » Coastline - Free Vector and Raster Map Data at 1:10m, 1:50m, and 1:110m Scales.” Accessed July 19, 2022. https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/.

Olofsson, Pontus, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. 2014. “Good Practices for Estimating Area and Assessing Accuracy of Land Change.” *Remote Sensing of Environment* 148: 42–57. https://doi.org/10.1016/j.rse.2014.02.015.

Pekel, Jean-Francois, Andrew Cottam, Noel Gorelick, and Alan S. Belward. 2016. “JRC Global Surface Water Mapping Layers, v1.3 | Earth Engine Data Catalog.” Google Developers. 2016. https://developers.google.com/earth-engine/datasets/catalog/JRC_GSW1_3_GlobalSurfaceWater.

Poortinga, Ate, Karis Tenneson, Aurélie Shapiro, Quyen Nquyen, Khun San Aung, Farrukh Chishtie, and David Saah. 2019. “Mapping Plantations in Myanmar by Fusing Landsat-8, Sentinel-2 and Sentinel-1 Data along with Systematic Error Quantification.” *Remote Sensing* 11 (7): 1–19. https://doi.org/10.3390/rs11070831.

Roy, D. P., V. Kovalskyy, H. K. Zhang, E. F. Vermote, L. Yan, S. S. Kumar, and A. Egorov. 2016. “Characterization of Landsat-7 to Landsat-8 Reflective Wavelength and Normalized Difference Vegetation Index Continuity.” *Remote Sensing of Environment* 185: 57–70. https://doi.org/10.1016/j.rse.2015.12.024.

Saah, David, Karis Tenneson, Ate Poortinga, Quyen Nguyen, Farrukh Chishtie, Khun San Aung, Kel N. Markert, et al. 2020. “Primitives as Building Blocks for Constructing Land Cover Maps.” *International Journal of Applied Earth Observation and Geoinformation* 85 (September 2019): 101979. https://doi.org/10.1016/j.jag.2019.101979.

Sinergise, Sentinel-Hub by. n.d. “NDWI Normalized Difference Water Index.” Sentinel-Hub Custom Scripts. Accessed July 19, 2022. https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndwi/.

Souza, Carlos M., Julia Z. Shimbo, Marcos R. Rosa, Leandro L. Parente, Ane A. Alencar, Bernardo F.T. Rudorff, Heinrich Hasenack, et al. 2020. “Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine.” *Remote Sensing* 12 (17). https://doi.org/10.3390/RS12172735.

Turubanova, Svetlana, Alexandra Tyukavina, and Matthew C. Hansen. 2018. “Primary Humid Tropical Forests | Earth Engine Data Catalog.” Google Developers. 2018. https://developers.google.com/earth-engine/datasets/catalog/UMD_GLAD_PRIMARY_HUMID_TROPICAL_FORESTS_v1.

USGS. n.d. “Landsat Enhanced Vegetation Index | U.S. Geological Survey.” Accessed July 19, 2022a. https://www.usgs.gov/landsat-missions/landsat-enhanced-vegetation-index.

———. n.d. “Landsat Modified Soil Adjusted Vegetation Index | U.S. Geological Survey.” Accessed July 19, 2022b. https://www.usgs.gov/landsat-missions/landsat-modified-soil-adjusted-vegetation-index.

———. n.d. “Landsat Normalized Burn Ratio | U.S. Geological Survey.” Accessed July 19, 2022c. https://www.usgs.gov/landsat-missions/landsat-normalized-burn-ratio.

———. n.d. “Landsat Normalized Difference Vegetation Index | U.S. Geological Survey.” Accessed July 19, 2022d. https://www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index.