Pixel-based forest classification of Sentinel-2 images using automatically generated datasets

Arminas Šidlauskas[1], Andrius Kriščiūnas[1]

[1] Kaunas University of Technology, K. Donelaičio str. 73, LT-44249 Kaunas, Lithuania

Abstract

Remote sensing tools are becoming popular in gathering information about forest area changes. The European Space Agency has launched multiple Sentinel satellites for land and marine monitoring. The Sentinel-2 (S2) satellite has great forest monitoring capabilities with its 13 high resolution bands. With the capabilities provided by this satellite, high accuracy pixel-based classification can be applied. In order to train a model that would be well suited to recognize forested areas from S2 images, a solid training dataset must be provided. In this study, two different information sources, Copernicus High Resolution Layers (HRL) and OpenStreetMap (OSM), were used to automatically create datasets. Models were trained and evaluated using the same artificial neural network architecture. After further analysis, it was noted that both OSM and HRL trained models yielded similar numerical evaluation results. Both models adjusted well to their data source classification and reached similar evaluation results of around 0.92 pixel accuracy. Upon further visual inspection, it was noted that OSM trained models created more false negative classifications identifying small forest patches and forest areas along rivers/lakes, HRL on the other hand created more false positives when identifying not only areas along rivers but rivers themselves as forest. All models failed to properly identify forest clearings in large forest areas, although HRL-trained models provided slightly better results.

Keywords

Forest classification, Sentinel-2 imagery, fully convolutional network, copernicus high resolution layers, openstreetmap

1. Introduction

Monitoring of forest areas is carried out on a continuous global and national scale. Field monitoring methods are not sufficient to monitor changes on a continuous basis, and there is a need to automate the process to achieve the highest possible accuracy. The use of remote sensing tools to monitor forest cover is increasing worldwide [1]. One of the major drivers for frequent forest monitoring is deforestation [2] and illegal logging [3].

The European Space Agency's Sentinel satellites are well suited for global forest observation. The main advantages of Sentinel satellites in forest monitoring are the long-term delivery of satellite imagery, global and frequent coverage, good data accessibility for the general public, and a wide variety of observation methods (radar, spectral bands) [4].

Sentinel-2 (S2) mission satellites provide 13 high resolution bands for land and sea monitoring. These bands and their combinations have already been used in various ways to classify forests [5, 6, 7, 8]. Reference [5] evaluates S2 capabilities to classify forest categories and European Forest types in the Mediterranean area, [6] evaluates the performance of dense S2 time series in forest species mapping in a challenging mountainous environment, [7] investigates the use of multi-temporal S2 data to identify tree species, [8] assesses the suitability of S2 data of typical land cover classifications (crop and forest). Often these

IVUS 2022: 27th International Conference on Information Technology
EMAIL: sidlauskasarminas@gmail.com (A. Šidlauskas); andrius.krisciunas@ktu.lt (A. Kriščiūnas)

classification tasks are completed using machine learning. In the case of referenced studies, a supervised random forest (RF) algorithm has been applied for pixel-based classification.

To train a precise model, a good dataset is required. Failure to prepare a precise dataset can result in an inaccurate classification model. Studies often use national data provided by forest/statistics agencies [5, 6, 7], which can then be further processed manually [6]. This data is provided in polygon form, polygons are then used to classify a certain area as a forest or specific forest species. Information about land use classification can also be received from OSM [9]. This information also includes forest polygons, similar to national data which is provided in polygon form. In other cases, Copernicus High Resolution Layers are used [10], these layers, which are provided in raster form, are then processed to act as pixel-based masks. The model accuracy in these papers varies from 83% to 95% when evaluating using pixel accuracy metrics.

Preparation of a precise dataset can take a long time if classification is done by hand or if institution data is used. The latter can have outdated/incomplete data which could severely restrict the ability to create a good dataset for certain areas. Additionally, different states may restrict access to this data. From this, the necessity of open access data, which could always be accessed and would be constantly updated, arises. In this work comparison of two open access data sources suitable for automatic ground-truth mask generation are investigated to evaluate their applicability to use directly for the selected machine learning model. Both HRL and OSM data sources are used as ground truths during evaluation. Accuracy has been tested using Copernicus S2 True Color Images (TCI). These images were collected in the summertime. The pixel-based classification was applied to a fully convolutional network (FCN) model with a resnet50 architecture.

2. Materials

2.1. Study area

Lithuania has been selected as the study area. The territory of Lithuania consists mostly of flatlands with lakes, swamps, and forests. Dominant species – pine, spruce, and birch. Lithuania covers an area of 65 300 km^2. The main reason for limiting the study area to one country is to avoid introducing new forest types during training and evaluation.

2.2. OpenStreetMap polygons

OpenStreetMap is a free editable geographic database of the world. During this research instance, data from the OSM database was taken for the year 2020. The database contains polygons of various areas – buildings, rivers, lakes, states, forests, etc. Forest polygons from the database can easily be converted to shapefile, geojson, or any other geospatial vector file format type. This is administrative information, meaning that if the database returns a polygon with a forest, it does not necessarily mean that there is a forest in that area, only that there should be a forest in that area. The opposite is true as well, small patches of forests might not be marked with polygons, which again introduces obscurity. Since OSM is massive in its scope, it is obvious that small inaccuracies are to arise and data will take longer to be updated. This becomes especially apparent with forest clearings which are officially marked as forest areas as shown in Figure 1.

Figure 1: Forest with clearing and mask generated from OSM polygons

2.3. Copernicus pan-European

The Copernicus pan-European HRL portfolio provides detailed land monitoring information including the HRL Forest layer. The approach to constructing the HRL Forest layer is based on a random forest classifier and is able to handle outliers to a certain context for forest classification problems achieving an accuracy of more than 98%. Unfortunately, implementation of such an approach requires intensive initial data preparation from different sources including the Sentinel missions and ancillary data sources like land-parcel identification systems (LIPIS), OSM data, and other local data sources. Respectively validation steps require semi-automatic validation steps [11]. HRL data is provided only every three years, while the last available forest coverage data is from 2018. The data provided by HRL on forest coverage can be received in raster files separated by European countries. This information can be used to create forest/non-forest pixel-based masks for training datasets which may be stated as valid information and used as ground truth labels during the periods of layer construction. Copernicus provides various three main forest layers – tree cover density (TCD), dominant leaf type (DLT), and forest type product (FTY). In this case, the TCD layer will be used.

2.4. Mosaic of the study area

The mosaic of the study area is a single raster image merged from multiple S2 images after they undergo preprocessing. Preprocessing includes cropping S2 images into small parts and merging them. Although a single S2 image already takes up only a part of the study area, it may contain clouds. Areas of image that contain clouds are unusable, no forest can be classified over them. Hence, there is a need to "remove" these clouds from the study area. The removal method infers cropping a single S2 image into small parts, then using a cloud mask (provided by S2) we ignore images that contain clouds. If all images in a given area contain clouds, select the image with the lowest amount of cloudiness.

In this study, mosaic was created from images from the 2018 summer period. This year's choice was motivated by the need to align it with the latest HRL data. Since S2 images are heavily impacted by clouds and shadows of clouds, priority has been placed to month with the lowest percentile of clouds in images. The month of June provided most images with a low distribution of clouds; hence, the study area is comprised of images from June. To create a single raster of the study area 21 S2 images have been used. Created mosaic of the study area is provided in Figure 2.

Figure 2: The study area of Lithuania. Image comprised of S2 data

2.5. Randomly generating points

One of the main advantages of automatically generating datasets is the ability to change the size of the dataset easily. Additionally, you can select specific areas of interest from which to generate datasets. Within these areas, points can be specified or they can be randomly selected. In the present case, points were generated randomly within the entire study area. Raster of the territory of Lithuania contains geocoordinates. Using these coordinates, boundaries of latitude and longitude can be extracted. These boundaries are then used to generate two random floating-point numbers, one for latitude, and the other for longitude. Two randomly generated numbers then make up a point. Then it can be calculated if the generated point is within the study area polygon. After generating the required number of random points inside the area of interest, these points can be used to crop out fixed size images from the study area. Using this method, a subset of random images can be created. The subsets are then used as the basis for new datasets. Selected points in the study area are presented in Figure 3.

Figure 3: Randomly generated points within the polygon of the study area. Image displays 1600 generated points

2.6. Generating datasets

In the scope of this paper, three random subsets of points were generated consisting of 800, 1600, and 3200 points respectively. Then for each point, an image sized 200x200 pixels is generated. The S2 TCI images have 10m spatial resolution. From this, a single image forms a square with a single side of 2000 meters, the image's area is $4km^2$. To create the datasets each subset of images is then duplicated, this is done so that mirrored datasets can be created, the only difference between these datasets is their masks. Every image in a dataset has a mask image. Mask images contain the classification of every pixel from the original image. From 800 randomly generated points, 2 datasets have been generated – 800 images and OSM masks and 800 images and HRL masks. Finally, each dataset was split into 9/10 training images and 1/10 validation images. In Figure 4 several examples are provided to explain the most noticeable differences among generated masks. In the first example, we can see that the OSM database provides a generalized forest area, which does not take into account any forest clearings, whereas HRL does. The second example shows that OSM fails to precisely identify forests along the river. The last example provides not a single larger forest area, but small patches of forests, and again OSM is at a detriment, lacking a substantial number of polygons to identify small forest patches.

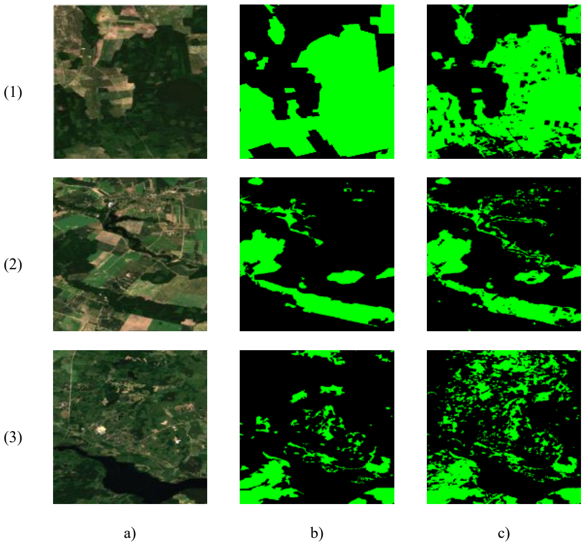

Figure 4: S2 image and generated forest (green) and non-forest (black) masks; a) True Color Image (TCI), 10m special resolution b) masks generated from OSM data c) masks generated from HRL data

2.7. Evaluation dataset

For evaluation, two new unique datasets are generated, one based on HRL data and the other on OSM data. These datasets were created using the same principle as the training datasets. A single dataset is made up of 200 images. After training all models will be additionally evaluated using these datasets, which means that both HRL and OSM will be regarded as ground truth during evaluation. The evaluation datasets were created to introduce new images that have not been processed by models and test their accuracy. Additionally, both datasets allow evaluating models against the same data, since during training they have their validation subset.

2.8. Training model

The main goal of this paper is to evaluate the differences between two pixel-based classification datasets. This means that during training the same model has to be used with all datasets. A fully convolutional network model has been selected with resnet50 architecture. The model distinguishes itself as fast, which is perfect when training multiple pixel-based classification models on different datasets.

3. Methods

3.1. Overview

To compare two different datasets and their precision, pixel-based classification will be performed. Figure 5 provides a general workflow.

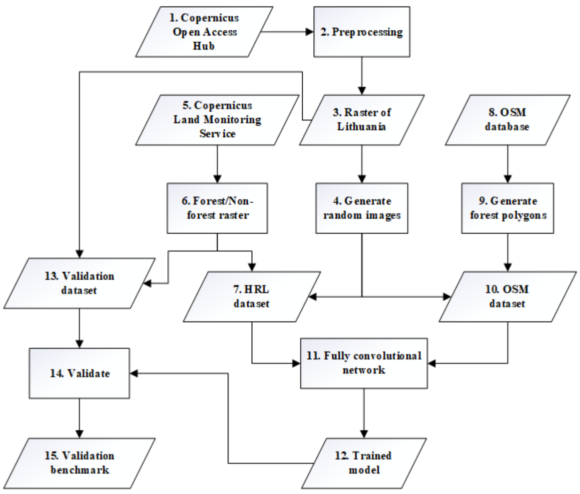

Figure 5: Workflow of automatically creating datasets and testing their performance

The workflow consists of:

1. Gather S2 images for the study area from Copernicus Open Access Hub.
2. Involves cloud removal and changing the coordinate system to WGS84.
3. Forming a cloudless single raster mosaic of the study area.
4. Generate a specified number of random images from the study area.
5. Gather HRL images from Copernicus Land Monitoring Service.
6. Convert pan-European raster into pixel-based forest/non-forest mask.

7. Generated a complete dataset from the HRL raster and a list of random images.
8. Gather forest polygons from the OSM database.
9. Generate a geojson format file that contains the required forest polygons.
10. Generate a complete dataset of OSM polygons and a list of random images.
11. Feed datasets to an FCN model.
12. Get a trained FCN model.
13. Generate a validation only dataset from a new list of random images in the study area and HRL raster.
14. Test trained FCN model accuracy against validation dataset.
15. Check the evaluation results.

3.2. Calculating accuracy

Accuracy during training and evaluation will be calculated using pixel accuracy and mean intersection over union (MIoU). Although pixel accuracy is a more common accuracy metric, it suffers when predicted images have a class imbalance. For example, an image consists of 100 pixels, 90 of which are non-forest pixels, and the rest are forest pixels. Then a trained model predicts that 100 pixels (the entire image) are non-forest. Pixel accuracy will be 90%. However, if we take intersection over union of forest and non-forest, we will have 0% and 90% accuracy, respectively. Then, if we calculate the mean of both classes, prediction accuracy drops to 45%. In this instance, mean intersection over union is a more accurate metric since datasets have images generated randomly, which can lead to a severe class imbalance in a single image. Both accuracy metrics are provided in the scope of this research.

Pixel accuracy equation:

$$a = \frac{TP}{TP + FP} \quad (1)$$

where TP – true positive pixels, FP – false positive pixels.

Mean intersection over union equation:

$$IoU = \frac{P \cap A}{P \cup A} \quad (2)$$

$$mIoU = \frac{IoU_{forest} + IoU_{non-forest}}{2} \quad (3)$$

where P – predicted pixels, A – actual pixels.

4. Results

4.1. Training results

Each model was trained for 1000 epochs with its own dataset. Validation masks were created from their own data source (OSM had its polygons, pan-European its raster). In Figure 6 we can see that pixel accuracy is generally similar across all datasets. MIoU, however, does vary more with OpenStreetMap. The lower MIoU can be attributed to inaccuracies of OSM. Validation data from this dataset could contain forested areas that are not marked as forest, thus impacting the validation results. Increasing the size of the training dataset also produced better overall validation results during training.

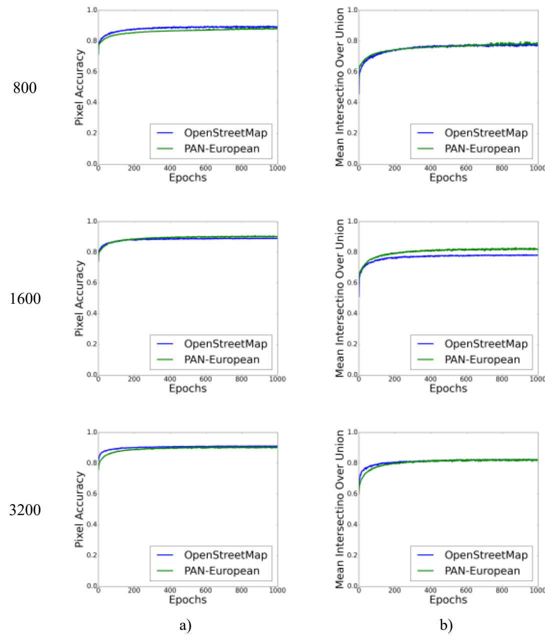

Figure 6: Validation results with different datasets a) using pixel accuracy b) using MIoU

4.2. Evaluation results

All trained models have been evaluated using two different datasets, one based on HRL data, and the other on OSM data. Table 1 provides evaluation results from the HRL based testing dataset, whereas Table 2 provides evaluation results from the OSM based dataset. Based on the results, it can be seen that both models have adjusted well to their training datasets. When evaluating HRL trained models with a newly created HRL evaluation dataset it performs better than the dataset trained with OSM data. However, when evaluation is done the other way around, OSM trained datasets to perform better. Additionally, it can be noted that models with larger training dataset sizes had slightly better accuracy, especially when validating against a dataset from the same source. Both models reach a similar accuracy ceiling of ~0.92 pixel accuracy and ~0.84 MIoU, when tested against their relative evaluation dataset. Based on these evaluation results, it cannot be stated that either HRL or OSM prove to be better sources for ground truth. Direct comparison of these results cannot be conducted with referenced papers, because different data is regarded as ground truth.

Table 1
HRL based evaluation results

Data source	Dataset size	Pixel accuracy	MIoU
HRL	800	0.891	0.798
HRL	1600	0.907	0.828
HRL	3200	0.917	0.843
OSM	800	0.844	0.717
OSM	1600	0.859	0.742
OSM	3200	0.854	0.734

Table 2
OSM based evaluation results

Data source	Dataset size	Pixel accuracy	MIoU
HRL	800	0.872	0.758
HRL	1600	0.861	0.741
HRL	3200	0.859	0.738
OSM	800	0.902	0.802
OSM	1600	0.910	0.819
OSM	3200	0.921	0.838

4.3. Noticeable differences

Although the evaluation results are very similar, certain differences can be identified by visually inspecting how models predict more edge cases. In Figure 7 we can see how the trained models compare. The first example shows that the OSM trained model ignores forest clearings while the HRL trained model recognizes clearings, albeit not very precisely. Both models still suffered heavy inaccuracies when they had to recognize forest clearings in large forested areas. Models would simply opt out to mark the entire area as forest and ignore clearings. The second example provides evidence of pan-European being better at recognizing forest areas along rivers. Since OSM rarely provides forest polygons for areas along rivers and lakes, HRL trained models become better at recognizing them. When it comes to small rivers, OSM models tend to completely ignore forest areas around rivers, while HRL trained models have a recurring issue of often identifying river itself as a forest. The last example shows how OSM has an issue with recognizing small forest patches. This is probably the most noticeable difference of all. On the other hand, pan-European is very good at identifying these patches, however it can at times identify larger areas that are no longer outside the bounds of small forest patches.

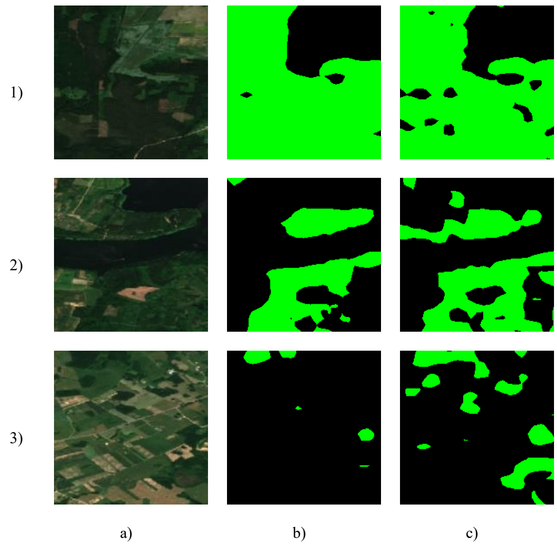

Figure 7: Examples of trained model classification a) S2 images, 10m spatial resolution b) classification, model trained with OSM data c) classification, model trained with HRL data

5. Conclusion

Six pixel-based forest/non-forest classification datasets were generated, three based on OSM data, and another three on HRL data, in order to evaluate the applicability of using open access data for dataset generation. All datasets were used to train a model that represents them. After training they were evaluated using additional evaluation datasets. Evaluation showed that both data sources yielded similar numerical accuracy results. Both data sources provided accurate data, that allowed models to reach ~0.92 pixel accuracy and ~0.84 MIoU, when evaluating with datasets from relative data source. During the evaluation, it was also noted that increasing the training dataset size increased the accuracy of the relative dataset evaluation. After additional visual inspection of edge cases, it was noted that models trained with OSM datasets tend to create a false negative classification of forest areas along rivers and small forest patches scattered in an area.

Models that were trained using HRL datasets were better at classifying forest clearings, forest areas along rivers and small forest patches scattered in an area. However, HRL trained models could provide false positive classification, identifying parts of the river as forest. Numerical differences between these two data sources proved to be negligible, one data source cannot be regarded as worse than the other. Although HRL data is produced only once every three years, visual inspections of generated dataset masks and trained model classified masks prove that it is better at detecting fine details in remote sensing images. Taking this into account, a pixel-based classification model can be trained using 2018 data, which can then be used to classify newer or older remote sensing data by year, which is especially important in HRL dataset case which is expensive to prepare and are provided only once every three years.

6. Data availability statement

Datasets that were generated during this research, both training and evaluation, together with complete study area and HRL raster of the study area can be found at https://zenodo.org/record/6548615 (accessed on 20 May 2022). OSM data can be found at https://planet.openstreetmap.org (accessed on 20 May 2022).

7. References

[1] M. K. Nesha et al., "An assessment of data sources, data quality and changes in national forest monitoring capacities in the Global Forest Resources Assessment 2005-2020," Environmental Research Letters, vol. 16, no. 5. IOP Publishing Ltd, May 01, 2021. doi: 10.1088/1748-9326/abd81b.

[2] M. A. Zambrano-Monserrate, C. Carvajal-Lara, R. Urgilés-Sanchez, and M. A. Ruano, "Deforestation as an indicator of environmental degradation: Analysis of five European countries," Ecological Indicators, vol. 90, pp. 1–8, Jul. 2018, doi: 10.1016/j.ecolind.2018.02.049.

[3] S. T. Thompson and W. B. Magrath, "Preventing illegal logging," Forest Policy and Economics, vol. 128. Elsevier B.V., Jul. 01, 2021. doi: 10.1016/j.forpol.2021.102479.

[4] Z. Malenovský et al., "Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land," Remote Sensing of Environment, vol. 120, pp. 91–101, May 2012, doi: 10.1016/j.rse.2011.09.026.

[5] N. Puletti, F. Chianucci, and C. Castaldi, "Use of Sentinel-2 for forest classification in Mediterranean environments," Annals of Silvicultural Research, vol. 42, no. 1, pp. 32–38, 2018, doi: 10.12899/ASR-1463.

[6] E. Grabska, P. Hostert, D. Pflugmacher, and K. Ostapowicz, "Forest stand species mapping using the sentinel-2 time series," Remote Sensing, vol. 11, no. 10, May 2019, doi: 10.3390/rs11101197.

[7] M. Persson, E. Lindberg, and H. Reese, "Tree species classification with multi-temporal Sentinel-2 data," Remote Sensing, vol. 10, no. 11, Nov. 2018, doi: 10.3390/rs10111794.

[8] M. Immitzer, F. Vuolo, and C. Atzberger, "First experience with Sentinel-2 data for crop and tree species classifications in central Europe," Remote Sensing, vol. 8, no. 3, 2016, doi: 10.3390/rs8030166.

[9] J. Estima and M. Painho, "Exploratory analysis of OpenStreetMap for land use classification," in GEOCROWD 2013 - Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, 2013, pp. 39–46. doi: 10.1145/2534732.2534734.

[10] A. Dostálová, M. Lang, J. Ivanovs, L. T. Waser, and W. Wagner, "European wide forest classification based on sentinel-1 data," Remote Sensing, vol. 13, no. 3, pp. 1–27, Feb. 2021, doi: 10.3390/rs13030337.

[11] European Environment Agency (EEA), "Copernicus Land Monitoring Service User Manual Consortium Partners," 2018. [Online]. Available: https://land.copernicus.eu/