Technical description of the image processing and data collection pipelines

Oliveira, Artur Andre & Casagrande, Luan C. & Soler, Diego Pavan

1. ML/CV solution

Our solution can be divided into two parts: The fully automatic and semi-automatic approaches. Even though mapping forest vegetation in riparian zones seems to be a trivial problem, it is important to note that we are proposing a solution that should be capable of mapping forests in any place and under any conditions. That is, the generalization needed is high and, consequently, the process to compose a model/solution capable of achieving good results for it is complex.

The automatic solution was trained using data from multiple biomass and under different conditions. However, it is important to highlight that this solution is prone to errors. That is, free accounts will be able to use such models and achieve results, but our company doesn't guarantee the output quality. That is an opportunity for users to get used to our platform and better understand its potential.

This fully automatic solution will be used as a basis for our paying users as well, however, a quality assurance team will be part of the process (human in the loop) aiming to guarantee that the result is inside the minimum requirements. In addition, this team has access to a semi-automatic solution that recognizes specific patterns and can help to achieve better results for situations that are not being covered by our fully automatic models.

In addition to being an option to overcome cases where the model is not being able to generalize correctly, such a semi-automatic solution can be used to continuously improve our automatic solution, since it can be used to compose additional data for our datasets.

2. Dataset composition

The dataset is composed of 96 multispectral sentinel 2 rasters. We have used only Level-2A products since it provides bottom of atmosphere reflectance. Even though 13 spectral bands are available, only bands 2, 3, 4, and 8 (blue, green, red, and NIR) were used. We have selected these bands by their resolution since it might impact the potential of our solution.

Multiple riparian zones were selected aiming to cover a higher number of cases. A total area of 1521,29 km2 was labeled and was divided as follows:

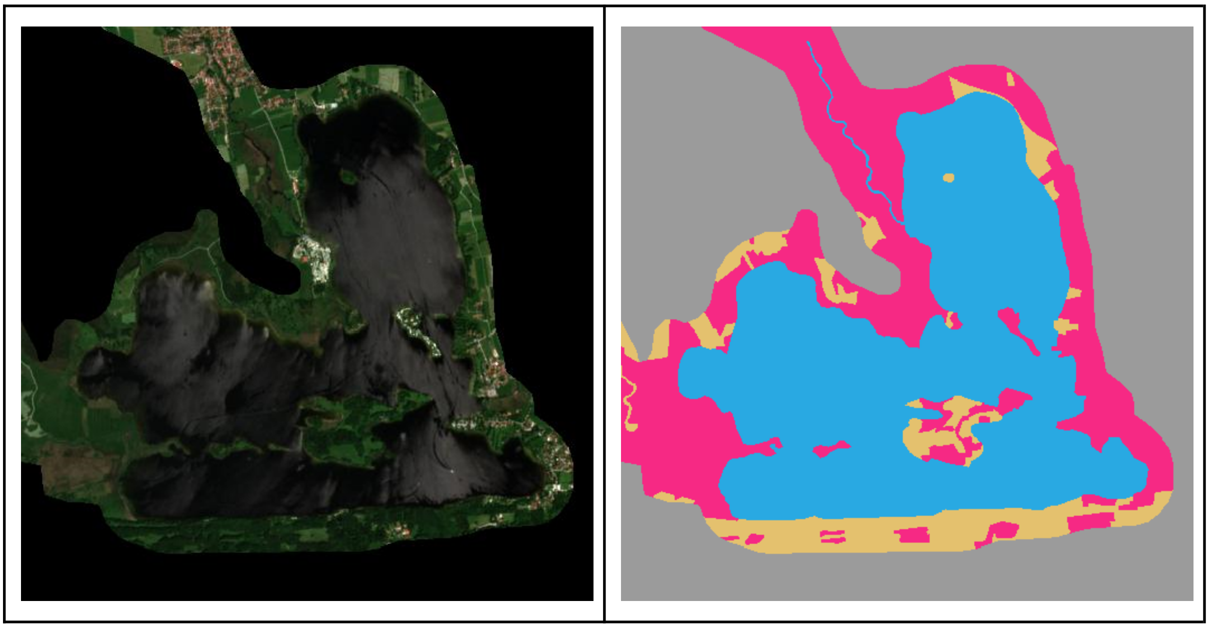
Table 1 - Dataset composition

	Area (km2)	**Number of Rasters**
Train	1204.35	73
Validation	208.14	15
Test	108.80	8

Two methodologies were used to compose our dataset: based on Copernicus Priority Area Monitoring product Riparian Zones (RZ Land Cover) [1] and based on our semi-supervised solution for regions covered by the previous product.

RZ Land Cover product provides a consistent and very-high resolution delineation and characterization of the Riparian Zones of major and medium-sized rivers for most of Europe and Turkey (i.e. the 38 EEA member and cooperating countries + UK), based on optical 2.5m spatial resolution satellite imagery from the ESA Data Warehouse [1]. However, the dataset was not externally validated and its reference is 2018. Consequently, a team of specialists visually inspected the polygons aiming to guarantee their accuracy.

The same approach was used for results achieved using the semi-supervised approach. Aiming to guarantee that the dataset was composed properly, each result was visually inspected by a team of specialists before its inclusion in the dataset. The following Figure has samples from the dataset composed:



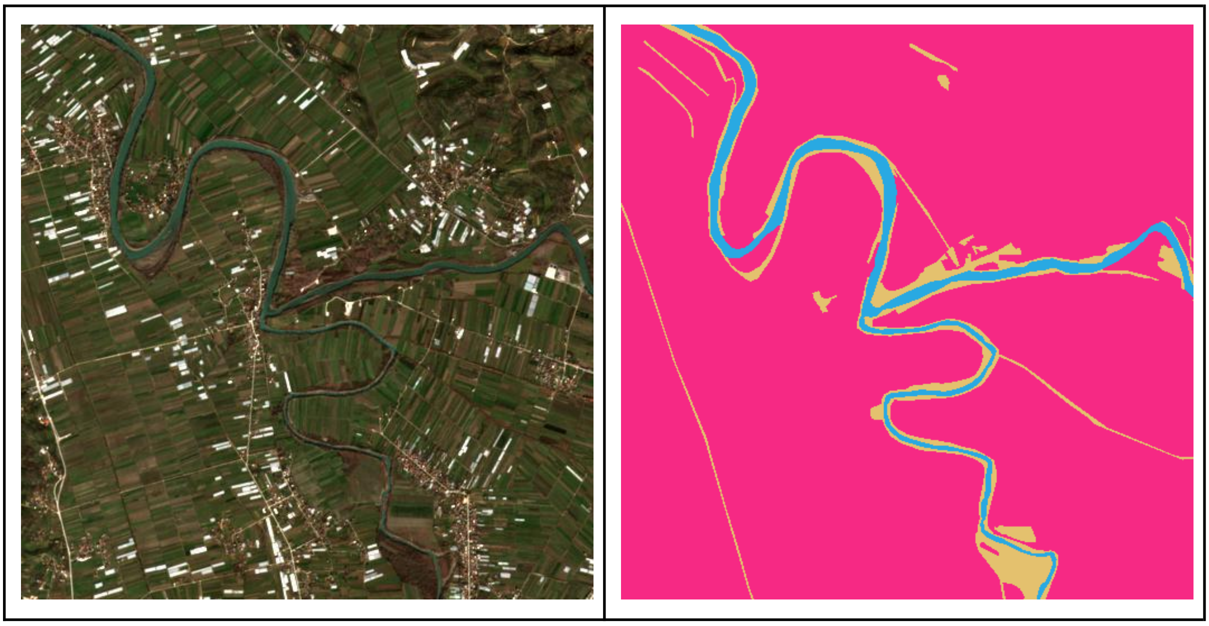


Figure 1 - Left column are input samples. Right column represents the expected output, where blue connected components represent the river, yellow the forestry, pink all other objects, and gray is outside the AOI.

3. Semi-supervised approach

We can abstract the semi-supervised approach using the following Figure:

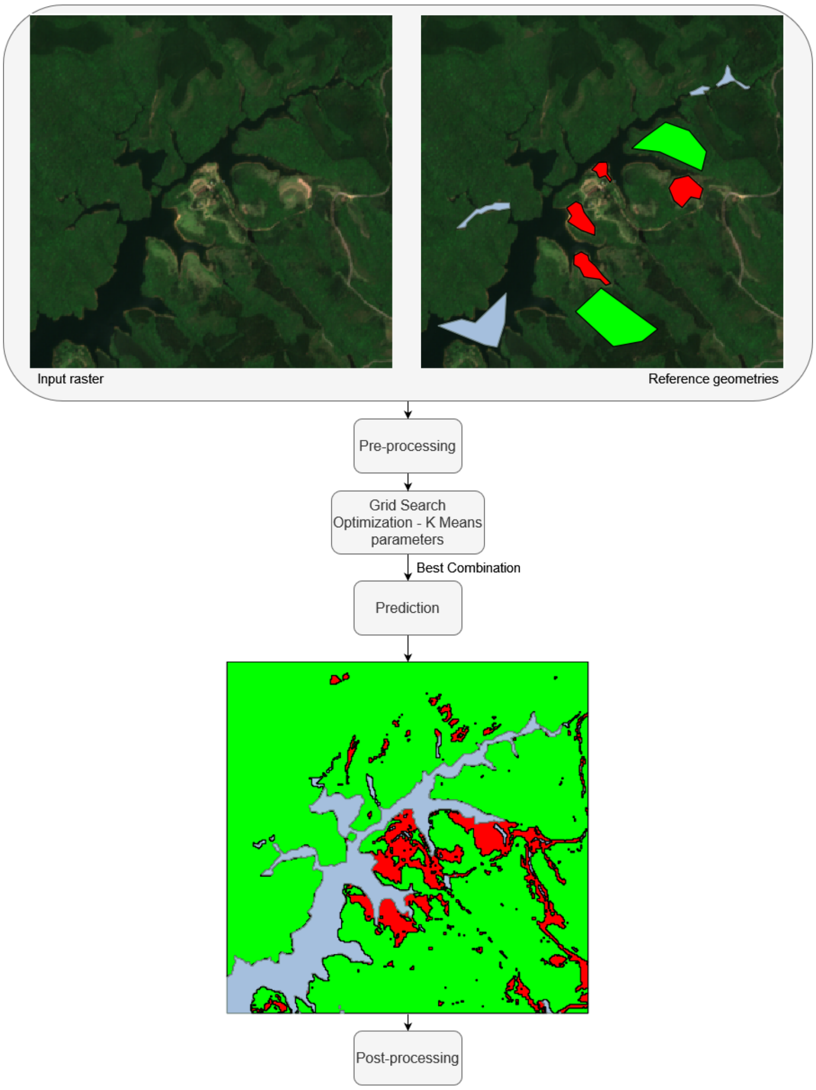


Figure 2 - Semi-supervised approach abstraction. Please note that we can abstract such a solution by four main steps: data input (raster and reference polygons), grid search optimization, k-means clustering prediction, and raster to geometry conversion.

The proposed solution uses the small amount of labeled data defined by our data labeling team or quality assurance team to find the proper set of vegetation indexes/color bands and k-means clustering parameters for such a problem. That is, it recognizes the pattern based on some small amount of information defined by our internal users and tries to find the best possible solution based on it.

These parameters are being selected based on Grid-search (GS). Such an algorithm, that is an exhaustive search algorithm, is very useful mainly when the search space is well defined. This algorithm consists of exhaustive searching through a manually defined subset of the hyperparameter space. The two most common types of this algorithm are the cartesian and the random grid search. In the first type, the algorithm tests each possible combination of hyperparameter values. In the second, the algorithm will sample uniformly from the set of all possible hyperparameter value combinations in the search space. In this work, the Cartesian type was used.

In this case, we are trying to find optimal parameters for a K-means algorithm. K-means is a classical unsupervised clustering ML approach, as described in []. The number of clusters (K) has an important impact on the final result, so this value is being optimized using GS. Two separated steps are conducted: the first one aiming to clusterize forestry against all other objects using NDVI and the second one to clusterize water based on NDWI, where:

$$NDVI = \frac{NIR-RED}{NIR+RED} \; (Equation\ 1)$$

$$NDWI = \frac{NIR-GREEN}{NIR+GREEN} \; (Equation\ 2)$$

Please note that water has priority over forestry and other objects. That is, if a pixel is classified as water and forestry, its final class is water.

4. Automatic approach

Several approaches based on classical machine learning and deep learning were tested aiming to find a solution suitable for this problem. The main problem identified was the generalization needed. That is, several approaches failed to generalize over multiple places. The best solution was achieved with a U-Net deep learning approach trained using the previously cited sequence of spectra.

Introduced in [1], Unet is an architecture for semantic segmentation, as is common to architecture for this problem, Unet consist of a Encoder, which contracting path, built using a sequence of 3x3 convolutions, where each one is followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling, each downsampling doubles the number of features channels. The encoder part of the architecture is used to extract features from input. The second part of the architecture is the Decoder path, which expands the dimension by applying upsampling followed by a 2x2 convolution that halves the number of feature channels. A new introduction on this architecture concatenates with the feature map from the Decoder path with the same dimension, which gives the characteristic U. In our problem, the network's output has 4 classes (void, water, others, vegetation) and input 4 bands (as described in the previous section). Please note that these rasters were normalized. Pytorch framework was used to implement such architecture.

Training

Before training, because of the size of satellite images, we cropped the input image in tiles. During tiling we only accepted tiles where the ground truth mask had any valid pixel, the tiles dimensions varied according to the shape of the water poligon.

During the training phase, we use the Adam Optimizer and Focal Loss as the loss functions. Since it is an unbalanced dataset, focal loss helped to overcome such a problem. In cases such as ours, where we don't have enough training image, data augmentation is a good practice. We have applied to following data augmentation techniques:

Table 2 - List of Augmentation

Augmentation Input	**Parameter**
Rotation	between +-90 degrees
Flip	Vertical and Horizontal

Learning rate decay factor of 0.5 was used after 5 epoches without any improvement with initial learning rate equal to 0.001

During inference, we break the satellite input image into tiles with fixed resolution, and apply our networks, after inference, we rebuild the image to the initial resolution, finalizing the process.

5. Results and discussion

Four pipelines were compared: First, our baseline result is a reproduction from the work proposed in [2] using decision trees, random forest, and Normal Bayes. The second approach used the same classical machine learning approaches, but added NDVI, NDWI, GNDVI, NRBI, SAVI, SR, and GVI aiming to improve pixel description. The third approach is based on the U-Net network, but using normalized red, green, and blue spectra only. The fourth step is the one described in the previous section. The following table describes the pixel-wise accuracy achieved with multiple approaches:

Table 2 - Pixel-wise accuracy achieved for the previously described four approaches

Baseline approach			Modified approach			Normalized R,G, B	Described approach
DT	NB	RF	DT	NB	RF	U-Net	U-Net
82.6%	34.7%	80.1%	85.4%	39.9%	85.4%	86.6%	90.9%

As described by the previous table, it is clear that deep learning approach with the NIR spectrum suppressed the classical approaches. These results were expected, since deep learning approaches can better generalize over large datasets.

Another important aspect to be highlighted is the NIR influence over the results. NIR light can reveal vegetation density differences because it can transmit through the top leaf layer and reflect off lower layers, then transmit back through the canopy to the sensor. NIR radiation can also reflect off soil in less dense, but closed, canopy areas, which will lower reflectance and reveal thinner areas of the canopy [3]. That is, it helps to describe smaller differences over the AOI.

The following tables describes the mean intersection over union (IoU) for each class:

Table 3 - Mean IoU per class

Class	IoU
Water	0.778
Forestry	0.686
Other	0.545

The following image shows One good case, one average, and one poor case:

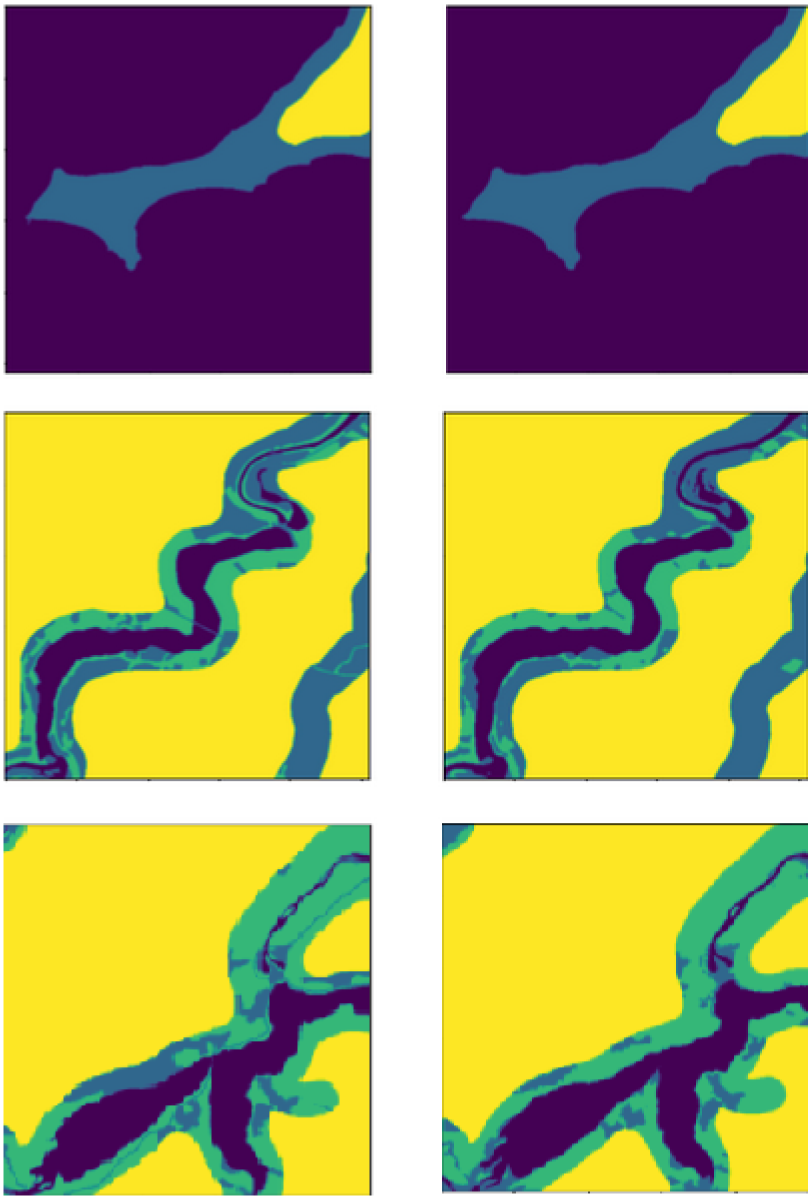


Figure 3 - First column is the reference and second column predicted results. Each image represents one case. Yellow regions represent NODATA. Dark blue represents water. Dark green other objects. Light green forestry.

As can be seen, the first case is almost 100% correct (99.5% accurate). Minor mistakes happened in the boundary between classes. The second case is an average case (88.3% accurate). Model failed mostly to precisely classify forestry x others. Even though visually the third case seems to be reasonable, it is 82.4% accurate. Again, the main problem identified was other x forestry class. It is clear that the model/dataset must be improved, but it is important to note that this model was trained to identify multiple bioma from different countries and showed a good potential.

6. References

[1]O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.” arXiv, 2015 [Online]. Available: https://arxiv.org/abs/1505.04597

[2]D. E. G. Furuya et al., “A Machine Learning Approach for Mapping Forest Vegetation in Riparian Zones in an Atlantic Biome Environment Using Sentinel-2 Imagery,” Remote Sensing, vol. 12, no. 24. MDPI AG, p. 4086, Dec. 14, 2020 [Online]. Available: http://dx.doi.org/10.3390/rs12244086

[3] Gisagmaps.com. 2022. *GIS Ag Maps - Why NIR for Vegetation*. [online] Available at: <https://www.gisagmaps.com/why-nir-for-vegetation/> [Accessed 25 April 2022].