# Transferability of Deep Learning Models for Land Use/Land Cover Classification

MAXIMILIAN HELL[1], MELANIE BRANDMEIER[1] & ANDREAS NÜCHTER[2]

*Abstract: Deep learning models in remote sensing are often trained once for benchmarking their results and not further applied to new domains or newer data. In this study, we test five previously developed DeepForest model variations on new data for land use and land cover classification. The models were pre-trained for this task on a multi-modal and –temporal data set from 2018 covering parts of the Amazon rainforest. The data comprises a twelve-month time-series of Sentinel-1 SAR data combined with a single Sentinel-2 multispectral scene. Classification maps from the MapBiomas Brazil project are used as label data to assess classification performance. The DeepForest classification models are able to classify the test scene from 2020 with up to 83.84% overall accuracy, producing reliable land use and land cover maps.*

## 1 Introduction

New deep learning models for land use and land cover classification are published frequently, mostly without any further studies of their potential transferability to new satellite data or other areas. In a previous study, we trained five variations of our newly proposed DeepForest model (CHERIF et al. 2022) for land use and land cover classification in the Amazonas region of Brazil. The models leverage on the multi-modal and –temporal aspects of the input satellite imagery, consisting of a multi-spectral Sentinel-2 image and a twelve-month synthetic aperture radar (SAR) Sentinel-1 time-series. One major challenge for continuous landcover mapping is the transfer of the process to new data. The goal of developing the DeepForest models was to provide architectures that are capable of generalizing well, not just on validation data but also on new, large datasets. Thus, these models will be tested in the same study area, but on more recent satellite data and their respective labels. Testing the transfer of these models to the new data will show if they learned a generalized representation and are able to be further employed in a land use and land cover classification workflow or whether larger amounts of data and/or better labels are necessary to achieve the ultimate goal of an automatized workflow.

## 2 Materials

### 2.1 Study Area

The study area is located in the region Amazônia Legal in Brazil. This socio-geographic region comprises nine of Brazil's 26 states and contains the Amazonas basin and rainforest. Testing the performance of the trained deep learning models was performed on a mosaic of satellite data

[1] Technische Hochschule Würzburg-Schweinfurt (THWS),
E-Mail: [maximilian.hell, melanie.brandmeier]@thws.de
[2] Julius-Maximilians-Universität Würzburg, E-Mail: andreas.nuechter@uni-wuerzburg.de

 DOI: 10.24407/KXP:1841078182

covering an area of 35,800 km$^2$ in the state of Mato Grosso, shown in Fig. 1. The satellite data was processed as described in Sec. 2.2.

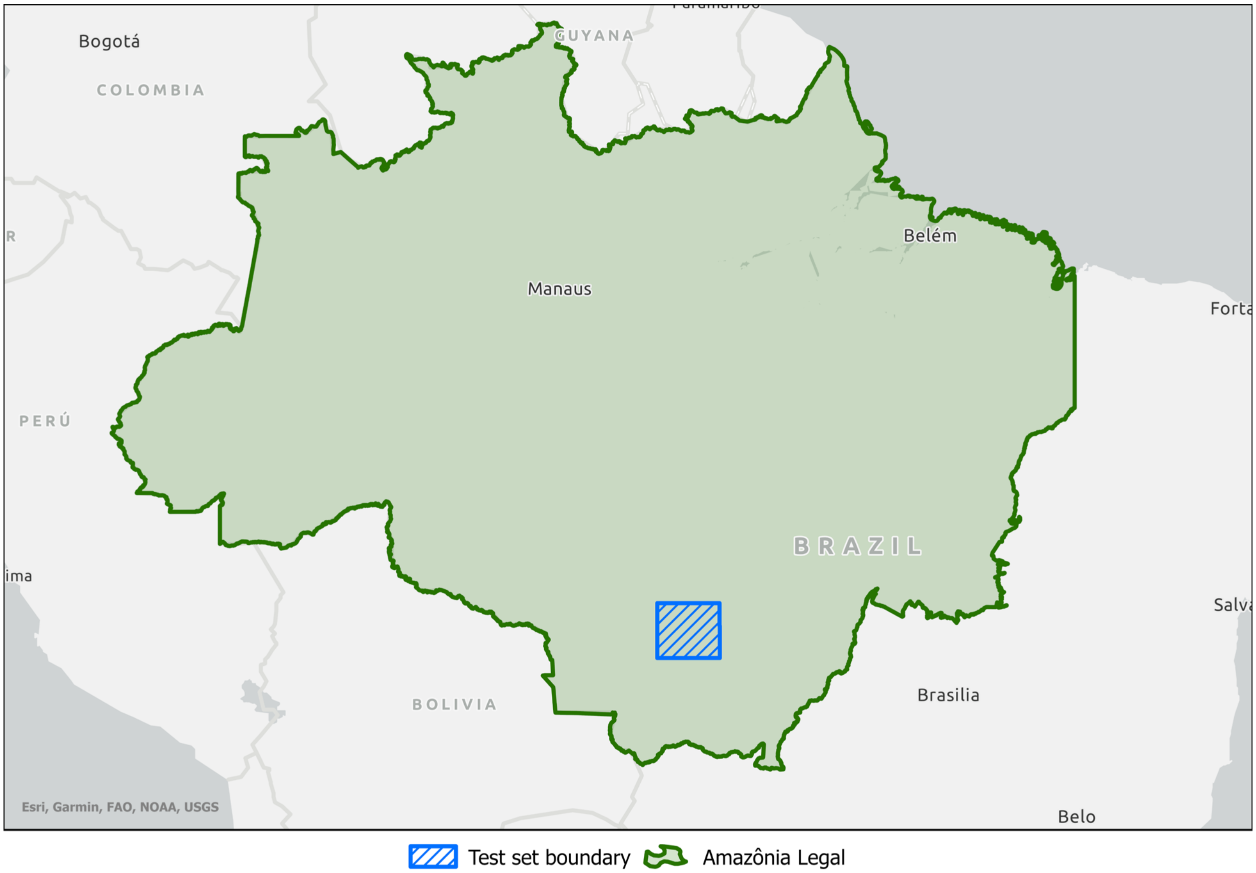


Fig. 1: Study area in the Amazon basin in Brazil. The footprint of the used satellite data mosaic is marked in blue.

## 2.2 Data

The satellite data used for the land use and land cover classification was preprocessed in the same way as described in the prior study (CHERIF et al. 2022). The data is comprised of a stack of a 12-month Sentinel-1 time-series and a single image of a Sentinel-2 scene. All data was captured in the year 2020, as this was the year with the most current label data available.

The Sentinel-1 SAR time-series contains twelve captures, one from each month, in VV+VH dual polarization, resulting in 24 bands. All scenes were captured in interferometric wide swath (IW) mode and acquired as Level-1 GRD products. The imagery was then further preprocessed using ESA's SNAP software (ESA 2023). The correction workflow (FILIPPONI 2019) comprises the following steps:

1. Precise orbit correction
2. Thermal noise removal
3. Calibration
4. Speckle Filtering (using the Refined Lee filter)
5. Conversion to the dB scale

The Sentinel-2 scene was chosen to have minimal cloud coverage and with an acquisition time in the middle of the year (June—August). In total, four image tiles were used, captured in the same pass of the satellite and spatially intersecting the Sentinel-1 time-series. The imagery was acquired from the Copernicus Hub as Level-2A product (Bottom-of-Atmosphere reflectance) with most of the atmospheric influences corrected. This product consists of 13 bands, of which only 10 are used. The three omitted bands have a pixel spacing of 60m×60m and are primarily utilized in atmospheric applications. The remaining bands with a pixel spacing of 10m×10m and 20m×20m were resampled to match the highest resolution of 10m×10m using the nearest neighbor method. All 34 bands, composed of Sentinel-1 bands and 12 Sentinel-2 bands, were then stacked and mosaicked to represent one image in a geodatabase using ArcGIS Pro (ESRI 2023).

As ground-truth labels, we used data from the MapBiomas project (MAPBIOMAS 2022). The project generates land use and land cover classification maps based on Landsat data. The published maps have a spatial resolution of 30m×30m, like the underlying satellite data. They publish iterations of their classification techniques as collections. We used Collection 6, as this was the most current version at the beginning of the project. The most recent year for a classification map within this collection is 2020. This collection uses a classification scheme, which comprises 26 different land use and land cover classes. However, only 20 of these classes are present in the Amazonas region.

The DeepForest models used for the transfer study were trained with label data from MapBiomas Collection 4. Data from this collection is provided up to the year 2018. Furthermore, the classification scheme changed from Collection 4 to Collection 6 with no apparent backwards compatibility, especially in the agricultural usage classes. The label map had to be reclassified to match the labels output by the models. To extract a class remapping scheme, the Collection 6 classification map from 2018 was compared to Collection 4 from the same year, further described in Section 3.2.

# 3 Methodology

## 3.1 Deep learning models

The models used in this transfer study are pre-trained variations of the DeepForest models (CHERIF et al. 2022). These models were trained on a combination of a Sentinel-1 time-series and a single Sentinel-2 capture, as described in Section 2.2. The training satellite data was captured in the year 2018 and the labels from the MapBiomas project are within Collection 4 of the same year.

The models are grouped according to their respective data fusion approach: The DeepForest-1 family of models uses an early fusion approach and contains three variations: DF1a, DF1b, and DF1c. In these models, both input data modalities are processed together and in parallel. The DeepForest-2 models, DF2a and DF2b, use two different data streams in the network to process the SAR time-series and optical data separately. The learned representations are then fused for the final classification step. All models use convolutional long-short-term memory [ConvLSTM] (SHI et al. 2015) as a defining building block to efficiently learn the spatio-temporal relationships of the SAR time-series data.

## 3.2 Transfer of the classification scheme

The MapBiomas Collection 4 used to train the original models provided classification maps up to the year 2018. Based on this scheme, the pre-trained models are able to differentiate between thirteen land use and land cover classes: *Forest Formation*; *Savanna Formation*; *Forest Plantation*; *Wetland*; *Grassland*; *Other non Forest Formation*; *Pasture*; *Annual and Perennial Crop*; *Semi-perennial Crop*; *Urban Infrastructure*; *Other Non-Vegetated Areas*; *River, Lake and Ocean*; and *Mining*.

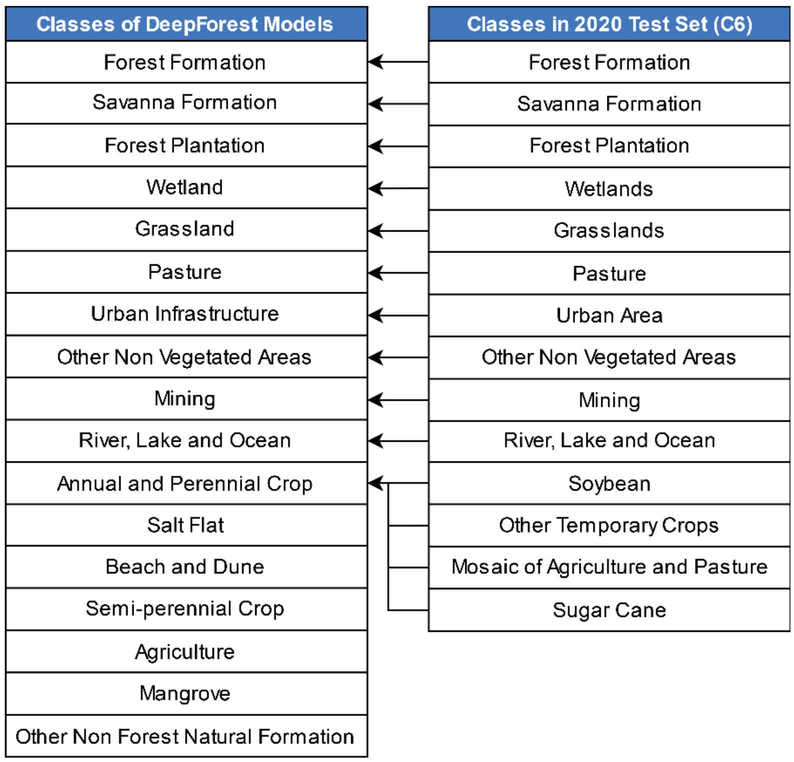


Fig. 2: Remapping from the classification scheme of Collection 6 to Collection 4 of the land use and land cover classes present in the new test set.

However, the data used for our transfer study was captured throughout 2020. The latest MapBiomas iteration at the start of the data processing was Collection 6 with an altered classification scheme. Thus, some re-mapping of the classes to the older scheme had to be conducted prior to inference and accuracy assessment. New (sub-)classes were introduced to the more recent collection and some classes were completely changed. Most significant was the change of the agricultural classes. Collection 4 used the distinction between *Annual and Perennial Crop* and *Semi-perennial Crop*. In contrast, Collection 6 distinguishes between *Perennial Crop* and *Temporary Crop*, with three and four further subclasses, respectively. These two macro and seven micro classes had to be reclassified into the two agricultural classes of Collection 4. The label maps covering the test area from 2018 of both collections were compared, to achieve this mapping from one scheme to the other. The *Soybean* class, contained in the *Temporary Crop* macro class of Collection 6, matched almost fully (98.76% of all pixels) with the *Annual and Perennial Crop* class. The class *Other temporary Crops* was not as unambiguously (67.38%)

ascribed to this class. *Sugar Cane* (also within in the *Temporary Crop* class) and *Mosaic of Agriculture and Pasture* were also mapped to this class. It is also noticeable that the class *Wetland* of Collection 6 is not present in this area in Collection 4, although this class exists in that scheme. The majority of the *Wetland* pixels (53.17%) map back to the *Savannah Formation* class. Figure 2 shows the remapping of the classes present in the test set to the classification scheme of Collection 4, on which the DeepForest models are classifying.

### 3.3 Inference on the test data

All five pre-trained model variations were used to classify the test scene (Fig. 1). The resulting maps were then compared pixel-wise to the label data from MapBiomas Collection 6 remapped to Collection 4, as described in Sec. 3.2. These label maps were registered to the satellite data and upsampled from 30m×30m using nearest neighbor sampling to match the pixel spacing of 10m×10m of the test data. All analysis was performed within ArcGIS Pro to make the best use of the mosaic dataset.

## 4 Results

In the original study (CHERIF et al. 2022) all models reached overall accuracies (OA) of at most 74.4% (DF1c) on the test set. On the new data set, all models performed better with respect to the OA. DF1c reached the highest score, with 83.84% of all pixels classified correctly. All models perform at least 8.9 percentage points higher than on the originally trained data, as shown in Tab. 1.

Tab. 1: Overview of the overall accuracies on the test set of the previous study and the new data set

| Model | OA in CHERIF et al. 2022 [%] | OA on the new test set 2020 [%] |
|---|---|---|
| *DF1a* | 74.3 | 81.57 |
| *DF1b* | 72.9 | 82.20 |
| *DF1c* | 74.4 | 83.84 |
| *DF2a* | 70.9 | 79.85 |
| *DF2b* | 69.0 | 80.84 |

However, three of the thirteen classes the models were trained on are missing in the new dataset. These classes are *Mining*, *Other non Forest Formation*, and *Semi-perennial Crops*. The classes *Forest Formation* and *Annual and Perennial Crop* are the majority land cover classes, with 32.4% and 37.2% of the test data set, respectively.

A qualitative assessment of the results by visual inspection should be conducted, when producing classification maps. Fig. 3 shows the ground truth labels of the MapBiomas project compared to the resulting map of the DF1c model, which achieves the highest overall accuracy in the quantitive assessment. The classification of *Annual and Perennial Crop* and *Pasture* is very similar to the ground truth. Some of the small *Savanna Formation* patches are missing in the classification. The DF1c model classifies much less area as *Urban Infrastructure* (dark red) than given as ground truth. When comparing this to the satellite image in Fig. 4 the model classifies the underlying land use and land cover classes well.

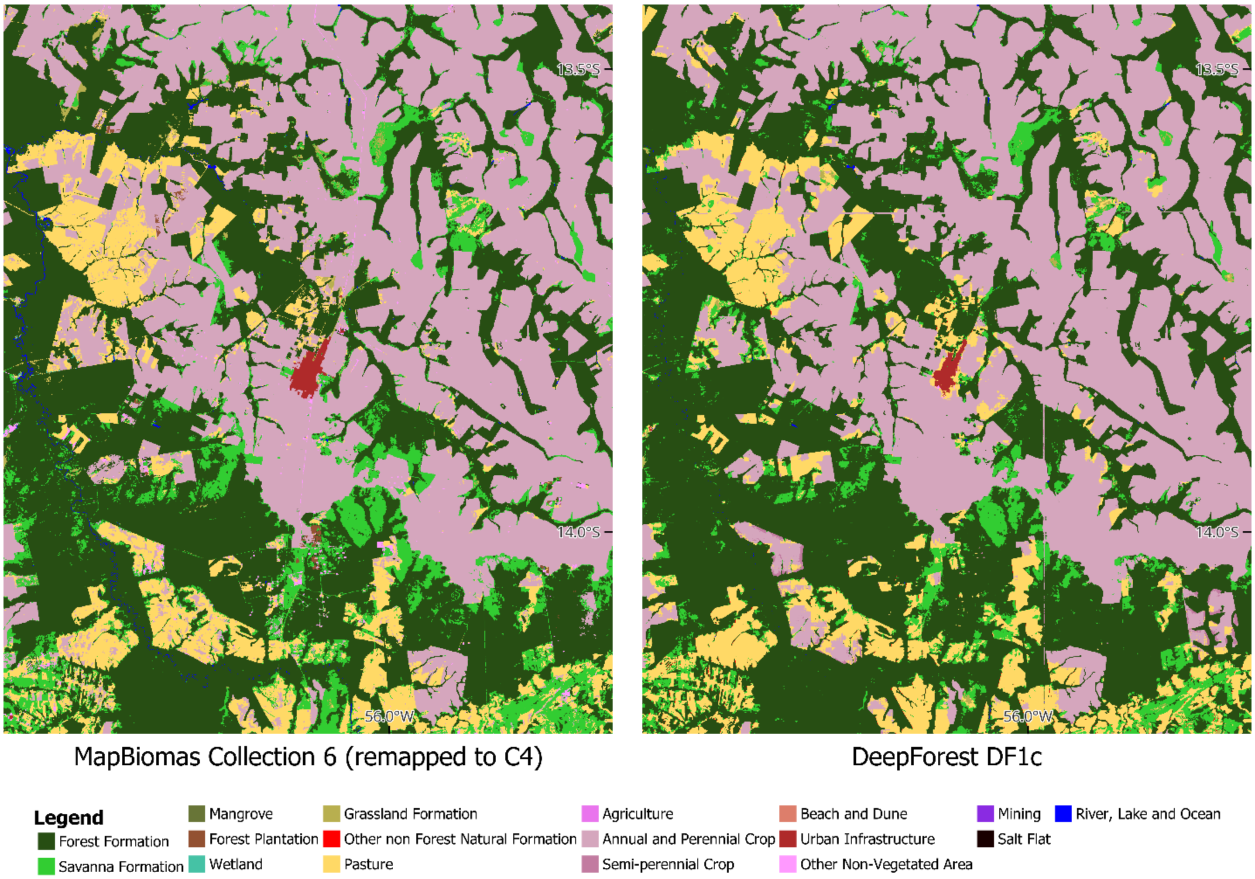


Fig. 3: The resulting land use and land cover map produced by the DF1c model (right) compared to the ground truth of MapBiomas (left).

Figure 4 shows a satellite base image, as well as one of the models from the DeepForest-2 late fusion family (DF2a) with the same spatial extent. The classified map shows a huge overlap with the ground truth in the agricultural and *Forest Formation* classes. However, the urban settlement in the center of the scene seems to be not fully captured. The model produces smaller patches of contiguously classified areas. When comparing this to the satellite image, the model seems to underestimate the size of the urban area and misclassifies parts of it as *Pasture* and *Annual and Perennial Crop*. The model also doesn't fully capture the patches of *Savanna Formation* within the *Forest Formation* south of the urban area. Although, when comparing this to the satellite base map in Fig. 4 it is not clear if these areas are truly savanna instead of forest.

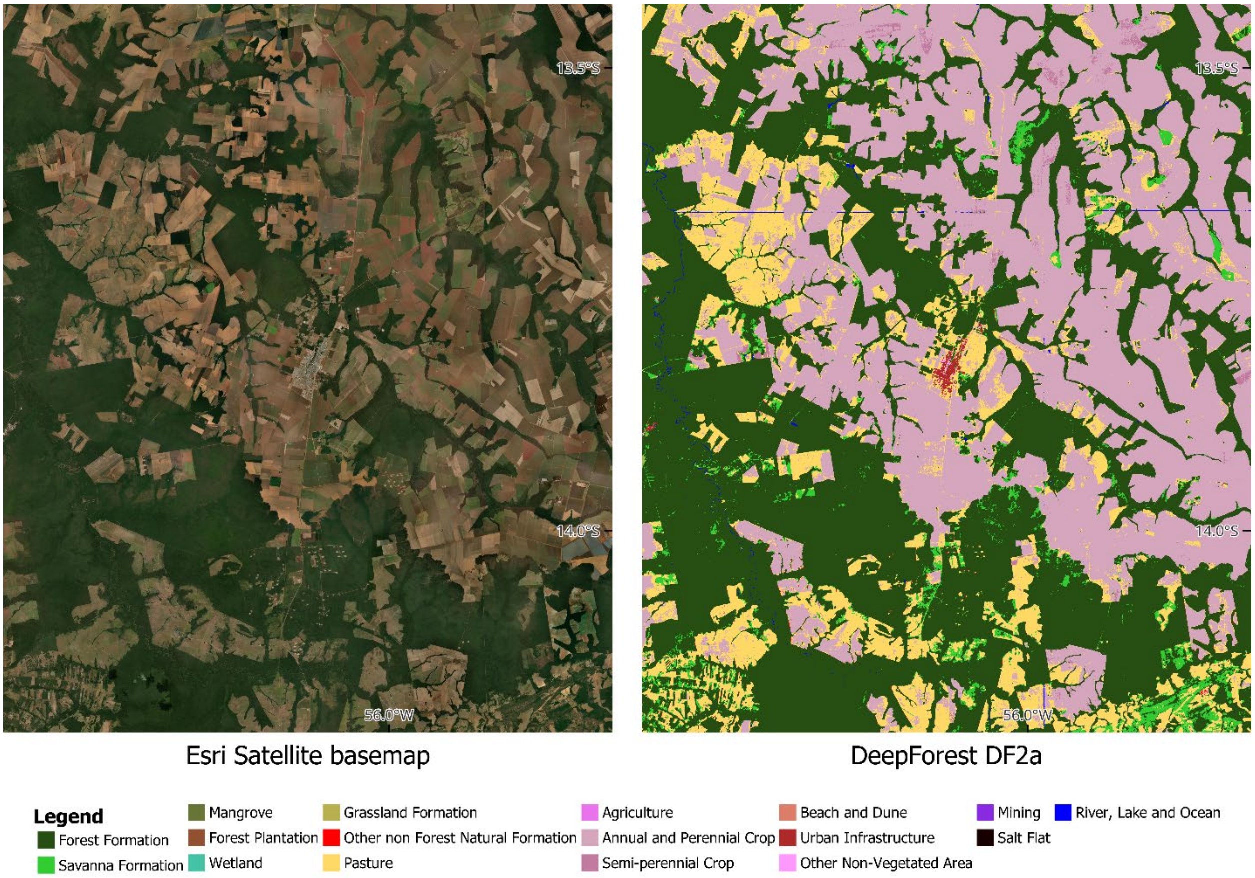


Fig. 4: The resulting land use and land cover map produced by the DF2a model (right) compared to the Esri Satellite basemap (left).

# 5 Discussion

The results, surprisingly, show some improvement in classification accuracies. However, this improvement is probably due to the underrepresented classes in the original dataset as well as class imbalance in the test dataset. Two major classes make up more than two thirds of the scene (*Forest Formation*: 32.4%; *Annual and Perennial Crop*: 37.2%) while the other widely distributed classes, *Savanna Formation* and *Pasture,* cover 18.1% and 10.1% of the scene, respectively. This imbalance is present in the whole Amazon basin, where *Forest Formation* covers 43.7% of the whole area. Together with the other dominating classes: *Pasture*, *Grassland*, and *Savanna Formation*. These four classes define over 80% of the Amazon basin. Although the models were also trained on this imbalanced data set, a weighting factor was added to the loss function to account for the imbalance.

Another factor contributing to the improvement in overall accuracies might be the change in the classification scheme and the used methodology to derive the classes in the MapBiomas project itself. It can be argued that the new scheme better represents landcover classes and is more accurate than previous collections and thus reduces the error produced by inaccurate labels. To further investigate these effects, we will train the models on larger, new datasets with the new label collection and compare results to the present transfer study. The qualitative assessment with the produced land use and land cover maps shows consensus with the true classes when visually comparing these with a higher resolution satellite base map and, thus, highlight the potential of these models for automated classification.

## 6 Literature

CHERIF, E., HELL, M. & BRANDMEIER, M., 2022: DeepForest: Novel Deep Learning Models for Land Use and Land Cover Classification Using Multi-Temporal and -Modal Sentinel Data of the Amazon Basin. Remote Sensing, **14**(19), 5000, https://doi.org/10.3390/rs14195000.

ESA, 2023: SNAP – STEP. https://step.esa.int/main/toolboxes/snap/, letzter Zugriff 09.01.2023.

ESRI, 2023: ArcGIS Pro | 2D, 3D & 4D GIS Mapping Software. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview, letzter Zugriff 25.01.2023.

FILIPPONI, F., 2019: Sentinel-1 GRD Preprocessing Workflow. International Electronic Conference on Remote Sensing, Proceedings, **18**(1), 11, https://doi.org/10.3390/ECRS-3-06201.

MAPBIOMAS, 2022: MapBiomas Brasil. https://mapbiomas.org/en, letzter Zugriff am 17.03.2022.

SHI, X., CHEN, Z., WANG, H., YEUNG, D.-Y., WONG, W. & WOO, W., 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, **28**.