# Deep Learning Projects

## Introduction

The goal of these projects is to design a deep learning methodology to solve a task on a provided dataset. This involves:

- Implementing methods to load and preprocess the data
- Implementing and training several relevant models and/or one model with several parameter sets
- Computing relevant metrics and analyzing the results
- Visualizing results

**Choose one of the projects** proposed in the next sections.

Specific expected outcomes are defined for each project, and the following criteria are used for **evaluation**:

| **Criterion** | **Explanation** | **Points** |
|---|---|---|
| **Method** | - Relevant choice of methods<br>- Several methods are compared<br>- Decisions are justified | 5 |
| **Reproducibility, code** | - Provided code can be run easily<br>- A readme file with instructions is provided | 2 |
| **Evaluation of results** | - Several relevant metrics are computed over train / val splits<br>- Computational complexity (e.g. training time, inference time) is assessed | 3 |
| **Report** | - Report is clearly written<br>- Graphics are readable and complete (e.g. axis titles)<br>- Results are analysed<br>- Discussion in terms of scientific (domain) output<br>- Limitations are clearly discussed | 2 |
| **TOTAL** | | **12** |

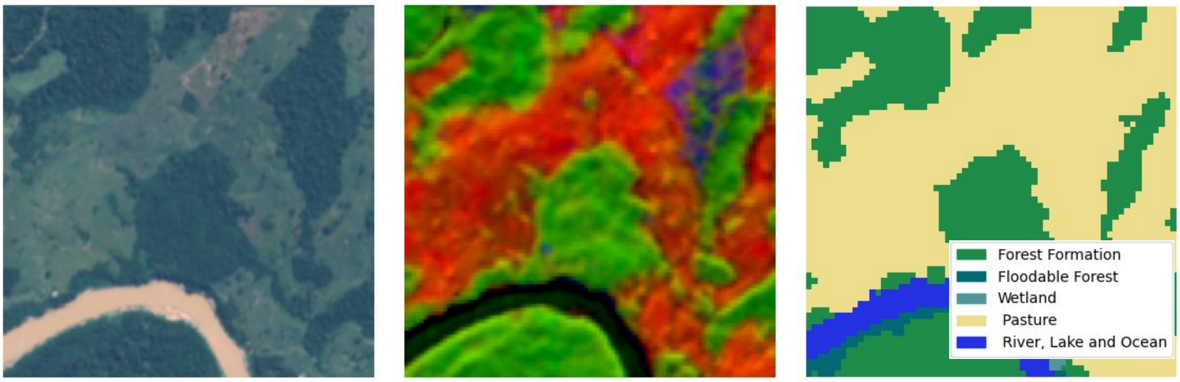
Table 1: Evaluation criteria

# 1 Comparison of satellite images and learned embeddings for land cover mapping in the Brazilian Amazon

**Task.** Accurate land cover information in the Brazilian Amazon is essential for monitoring deforestation, land use dynamics, and ecosystem health. While land cover maps are typically derived from satellite imagery, learned embeddings are emerging as a promising alternative. Embeddings are representations learned from large Earth observation datasets that contain high-level features and might reduce the need for large labeled datasets.
In this project, you will compare models trained on Sentinel-2 images with models using the AlphaEarth (AE) embeddings [1] for land cover mapping in the Brazilian Amazon.

**Data.** The dataset consists of 5000 training samples and 1000 test samples. Each sample consists of an image patch, AE embedding patch, and a ground truth mask covering the same area. You can see an example in Figure 1. All data comes at 10m resolution and corresponds to the year 2020. Download the dataset here.

- **Images.** Sentinel-2 cloud-free multi-spectral mosaic images averaged over the calendar year, all bands upsampled to a common 10m resolution.
- **AE embeddings.** Each pixel corresponds to a feature vector of size 64, containing semantic information extracted by the AlphaEarth Foundations model for.
- **Labels.** Per-pixel land cover map. A .csv file is provided with mapping from the pixel values to class names. Values of 0 correspond to 'nodata' values.




a) b) c)

Figure 1: a) Sentinel-2 image; b) AE embeddings (3 random channels shown); c) ground truth land cover map;

**Expected outcome.**

- Design and implement a training pipeline for a deep learning model to classify land cover from Sentinel-2 images; this includes selecting appropriate model architecture, loss function, hyperparameters search, etc.
- Explore different approaches for modeling land cover from AE embeddings - implement and train at least 1 deep learning and 1 non-deep learning model;
- Optionally - design and implement a deep learning approach to combine both Sentinel-2 images and AE embeddings
- Compute and discuss appropriate performance metrics; compare results obtained with different approaches

**Challenges.**

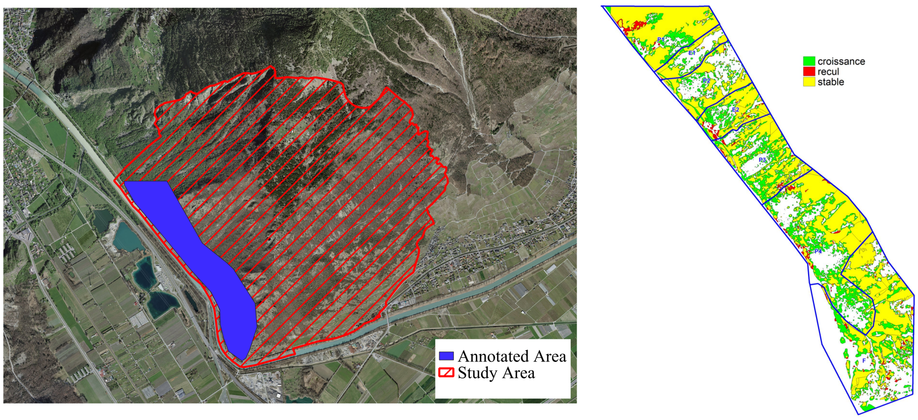
- You will need to design a suitable model architecture for working with AE embeddings.

- Sentinel-2 images have 12 input bands; most standard deep learning models have been developed with natural 3-band images in mind; you may need to adapt an existing architecture to serve your purposes.
- Class imbalance - some classeso are much more common than others; this is a common problem in machine learning which can have significant impact on the performance of your model if not addressed, particularly on less common classes.

**References** [1] Brown, Christopher F., et al. "AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data." arXiv preprint arXiv:2507.22291 (2025).

## 2 Changes in brush cover on Mont Rosel from 1992 to 2023

**Task.** The project aims to detect and map the evolution of brush vegetation on Mont Rosel near Martigny between 1992 and 2023. This site, part of the protected grasslands of the Follatères, contains grasslands of high ecological value, but these are increasingly threatened by the spread of shrubs and trees. Monitoring this process is essential for biodiversity conservation and for planning future management. Building on earlier manual labels obtained from photo-interpretation, you will develop a deep learning model to automatically detect vegetation cover from aerial images. By comparing results across the two time periods, you will generate a change-detection map.




(a) (b)

Figure 2: (a) Perimeter of the overall study area (red) and the manually annotated reference area (blue), obtained during previous photo-interpretation work. (b) Results of vegetation change analysis within the annotated area: green = regions where vegetation has grown, red = regions where trees have recessed, and yellow = regions that remained stable between 1992 and 2023. The goal of the project is to develop a deep learning model to extend this analysis to the entire study area.

**Data.** You can download the data here. The project relies on two sets of aerial orthophotos: black-and-white images from 1992 and RGB images from 2023. The dataset is divided into two parts: (1) an annotated area, and (2) the rest of the study area.

- For the annotated area, the dataset includes 146 RGB tiles of size $500 \times 500$ and at $0.1$m resolution from 2023, and 146 black-and-white tiles of size $100 \times 100$ at $0.5$m resolution from 1992, together with two corresponding label sets. The two sets of images span the same area and are named consistently. The labels are binary masks, where 1 indicates the presence of bushes and 0 their absence. This annotated subset will be used for model development, and you will split it into training, validation, and test images.
- For the unannotated part of the study area, two larger sets of 1991 RGB images for 2023 and 1991 black-and-white images for 1992 are available. These will serve for inference, allowing the trained model to generate a complete vegetation change map across the full study area. All images are georeferenced.

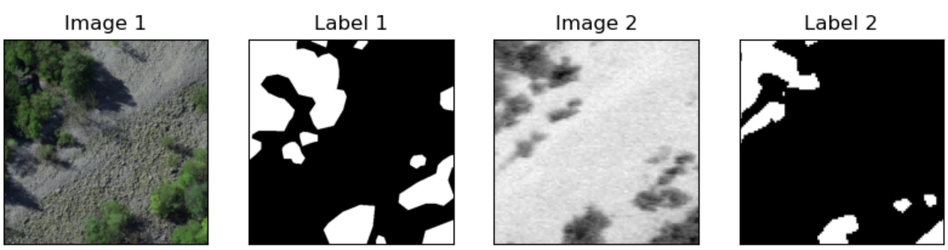



Figure 3: Example from the annotated dataset.

**Expected Outcome.**

- Design and implement two deep learning models tailored for segmenting bushes and vegetation in the two image datasets (1992 and 2023).
- Explore one (expected) or more (optional) of the following directions
  - Experiment with various model architectures and training strategies, particularly those addressing the challenge of limited dataset size. Evaluate model performance using appropriate quantitative metrics.
  - Error analysis and uncertainty quantification: Conduct an analysis of failure cases to identify conditions under which the models produce unreliable predictions. Consider integrating uncertainty estimation techniques to generate uncertainty maps alongside the segmentation outputs.
- Apply the best-performing models to the complete unlabeled image dataset for both years. Generate vegetation cover maps for the entire study area (analogous to Figure 2) and deliver them in `.shp` format. Include visual examples of these results in the final report.

**Challenges.**

- Limited dataset size. The annotated dataset is very limited in size (146 images for both modalities). While this reduces computational requirements and training time, it introduces a high risk of overfitting. You will have to implement appropriate strategies to mitigate this effect and ensure that the models generalize well beyond the annotated subset.
- Tile Management. The study area is divided into multiple georeferenced tiles. After model inference on the unlabeled dataset, these tiles must be recombined to produce a single, coherent vegetation map of the entire site for both years in `.shp` format. *Hint: This step can be performed with python's library* `rasterio` *and* `QGIS`.

**References.** Documents provided by the Commission des Follatères are included with the dataset. These materials, written in French, are not required for understanding or completing the project but are supplied to offer additional context for interested students.

## 3 Species distribution modeling with multimodal satellite and environmental data

**Task.** Species distribution models play a key role in conservation by predicting how suitable different locations are for species based on their environmental conditions. Different types of data, called modalities, can be used to capture these conditions. Climatic tabular variables provide essential information on factors such as temperature and precipitation. Satellite time series data help detect key seasonal patterns and extreme events that influence species distributions. Finally, satellite imagery offers spatial context, revealing characteristics of the surrounding environment, such as vegetation cover or human presence.
In this project, you will integrate and combine these three data modalities to predict the distribution of 342 plant species across Europe.

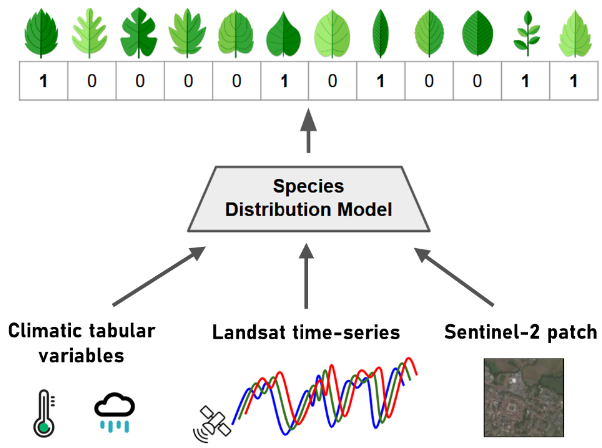



Figure 4: Project overview

**Data.** The data used in this project comes from the GeoPlant dataset [1]. We provide a reduced version of the dataset to allow you to train models more quickly. You can download the data here. It consists of 5,000 training samples (each representing data for a single location) and 1,000 test samples. For each sample, you are provided with:

- A presence–absence vector for 342 species, where 1 indicates that the species was observed and 0 indicates absence. Note that the species identities are anonymized, but they correspond to plant species recorded in vegetation plots.
- 19 tabular climatic variables, representing statistical summaries of temperature and precipitation. A description of the corresponding variables is available here. Geolocation data is also provided, which can be used to generate prediction maps.
- Quarterly Landsat time series spanning 10 years (2008–2017) for four spectral bands (RGB + NIR).
- Sentinel-2 image patches centered on the sample location at a 10m spatial resolution. Only the RGB bands are provided for simplicity.

All data indices are aligned across modalities. The objective is to predict the suitability (i.e., presence) of each species based on the provided environmental variables.

**Challenges.**

- Designing effective strategies to integrate multiple data modalities.
- Multi-label classification with significant class imbalance.

**Expected Outcome.**

- Develop a functional pipeline that produces species suitability predictions by leveraging all three data modalities. Evaluate its performance using the mean area under the ROC curve (AUC) across species, along with other relevant metrics, and critically assess its limitations as well as potential areas for improvement.

- Explore one (expected) or more (optional) of the following directions:
  - Experiment with and compare different model architectures and hyperparameters for each modality, and discuss the results.
  - Investigate different approaches to address class imbalance in the species data.
  - Explore and compare methods for assessing the importance or contribution of each modality.
  - Create informative visualizations that geographically map species distributions predicted by your model and use them to understand the sources of model error.

**References.** [1] Picek, L., Botella, C., Servajean, M., Leblanc, C., Palard, R., Larcher, T., Deneu, B., Marcos, D., Bonnet, P. and Joly, A., 2024. Geoplant: Spatial plant species prediction dataset. Advances in Neural Information Processing Systems

## 4 Hurricane Damage Detection with Deep Learning

**Task.** After natural disasters such as hurricanes, rapid and reliable assessment of damaged areas is essential for coordinating response and recovery efforts. In this project, you will develop a deep learning model to automatically classify satellite image patches as damaged or undamaged based on post-hurricane observations. The task focuses on evaluating how well deep learning models can detect hurricane-induced damage patterns from optical imagery.
While achieving high accuracy is important, emergency response systems require models that are not only effective but also *trustworthy*. A model that confidently predicts the wrong outcome can be more harmful than one that acknowledges its limitations. Therefore, beyond building a classifier, you will evaluate the model's **calibration**, that is, how well predicted probabilities reflect actual correctness. A well-calibrated model should, for instance, be correct about 70% of the time when it assigns 70% confidence to its predictions. This project thus aims to explore not only what the model sees, but also how accurately its confidence aligns with reality.
For a short introduction to model calibration, see this tutorial.

(a) Damaged (b) Undamaged

Figure 5: Comparison of damaged vs undamaged buildings from satellite imagery.

**Data.** The dataset is based on the version provided by Cao et al. [1] and can be downloaded here.

- **Images:** RGB satellite patches of areas in Texas after Hurricane Harvey, captured by the GeoEye-1 satellite.
- **Classes:** *damage* and *no damage*.
- **Splits:** 19,000 training images (unbalanced), 2,000 validation images, and 2,000 test images (both balanced).
- **Data quality.** The validation and test sets include visually ambiguous or lower-quality samples compared to the training set, representing more realistic and noisy post-disaster conditions.

**Challenges.**

- High intra-class variability, "damage" can manifest in many ways (e.g., roof loss, flooding, vegetation destruction).
- Some image patches are visually ambiguous, and differences in lighting or sensor conditions introduce additional noise.
- Domain shift between training and test events.

**Expected Outcome.** Students are expected to develop a deep learning model for detecting hurricane-related damage from satellite imagery and evaluate its performance using standard classification metrics such as accuracy, and F1-score. Beyond these metrics, students should assess how well the model's predicted probabilities correspond to actual outcomes, its *calibration*, and explore ways to improve it where appropriate. This may involve analyzing reliability diagrams or related visualizations to examine whether the model tends to be over- or under-confident. The goal is to develop a calibrated model whose confidence values can be meaningfully interpreted, supporting more reliable decision-making in emergency response settings.

**References.** [1] Cao, Quoc Dung, and Youngjun Choe. "Detecting damaged buildings on post-hurricane satellite imagery based on customized convolutional neural networks." IEEE Dataport 1.206 (2018): 1-19.

## 5 Mapping Swiss Ecosystems from Aerial Images and Environmental Variables

**Task.** This project aims to develop a deep learning framework for predicting ecosystem categories in Switzerland by jointly leveraging remote sensing imagery and environmental variables.
The goal of this study is to evaluate the contribution of remote sensing images to ecosystem classification, compare their predictive performance with environmental variables, and investigate how to combine both data sources effectively. The aim is also to identify information provided by environmental variables that is not captured by remote sensing, but is relevant for ecosystem mapping.

**Data.** The data spread across Switzerland and comprises 16’925 locations with:

- One aerial image of 100x100m at a 50cm spatial resolution with RGB bands from the swisstopo product swissIMAGE [1].
- 48 numerical variables from SWECO25 [2], a raster database for ecological research in Switzerland, covering 6 thematic: land use/land cover, bioclimatic and edaphic (soil) properties, population, geology, hydrology and vegetation. The variables have been standardised (mean=0,std=1).
- Ecosystem label extracted from the EUNIS (European Nature Information System)[3] framework, spanning 17 categories.

The dataset is already split into training (60%), validation (10%) and test (30%) sets following a geographic split to avoid spatial autocorrelation. Results are expected to be computed on the test set. Data are available *here [4]* .

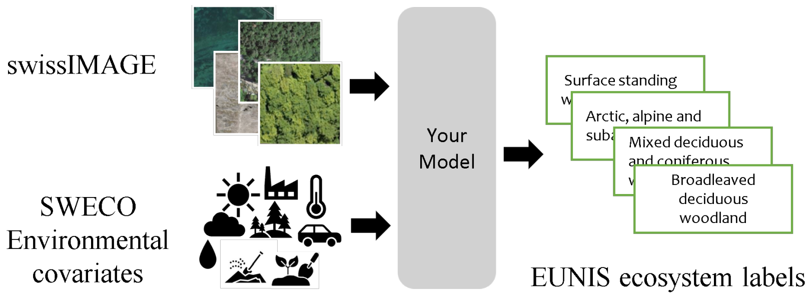



Figure 6: General pipeline for predicting ecosystems from aerial images and environmental variables.

**Expected Outcome.**

- Build a training pipeline with the appropriate model architecture: i.e.starting from an appropriate pretrained backbone, choosing a loss function, searching for the best hyperparameters, etc. Compute and discuss appropriate performance metrics on EUNIS ecosystem predictions.
- Compare the performance of models trained on the different subsets of variables (ablation study by thematic group of variables: impact of bioclimatic variables, of geology, etc.) and critically analyse which variables are useful for predicting the ecosystem. Discuss an effective combination of input (images, sweco variables) to reach the best performance.
- Highlight and interpret interesting results to support your analysis, for example, with figures or maps.

**References.**

[1] swissIMAGE orthophotos : `https://www.swisstopo.admin.ch/en/orthoimage-swissimage-10`
[2] Külling, N., et al., (2024). ***SWECO25: a cross-thematic raster database for ecological research in Switzerland***. Scientific Data, 11(1), 21.
[3] Chytrý, M., et al., (2020). ***EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats.*** Applied Vegetation Science, 23(4), 648-675.
[4] Drive to dataset : `https://filesender.switch.ch/filesender2/?s=download&token=cf23c578-fdae-443f-aba4-c05b819b79d5`