Winter School 2025: Program

Harnessing Machine Learning for Breakthroughs in Plant and Environmental Sciences

WORKSHOPS, March 10-12th, 2025
Location: ETHZ, HG F26.1

Monday, March 10, 2025

9:00	**Introductory Lecture: Machine Learning Refresher** (Prof. Dr. Jan Dirk Wegner, UZH/ETHZ)
10:00	**Machine Learning for Predicting Ecosystem Fluxes** (Dr. Fabian Bernhard/ Prof. Dr. Benjamin Stocker, UNIBE) – Practical workshop
13:00	Lunch Break
14:00	**Plankton Classification: Opportunities, Methods and Pitfalls** (Dr. Marco Baity Jesi, EAWAG) – Practical workshop
18:00	End

Tuesday, March 11, 2025

08:30	**Remote Sensing: Satellite Data Selection and Preparation for Machine Learning Applications** (Dr. Jochem Braakhekke, UZH) Practical workshop
10:30	Break
10:45	**Plant Species Identification from Photos and Local Habitat Conditions** (Dr. Philipp Brun, WSL) – Practical workshop
12:45	Lunch Break
13:30	**Plant Species Identification from Photos and Local Habitat Conditions** (continuation)
Part 2 15:45	**Individual Tree Species and Health Detection Using Deep Learning Model** (Xia Zhongyu, ETHZ) - Practical workshop
18:00	End

Wednesday, March 12, 2025

08:30	**Image analysis from a 'phenotypers' perspective: The example of image-based wheat head detection to analyze heading dynamics in a diverse winter wheat genotype set** (PD. Dr. Andreas Hund & Dr. Lukas Roth, ETHZ) - Practical workshop
13:00	Lunch Break
14:00	**Plantseg 2.0: Powerful and User-Friendly Plant Tissue Segmentation** (Dr. Lorenzo Cerrone UZH) - Practical workshop
18:00	End

Thursday & Friday, March 13-14, 2025
March 13th in ETHZ, HG F30 (Audimax) / March 14th in ETHZ, HG D7.2

Symposium 2025: Harnessing Machine Learning for Breakthroughs in Plant and Environmental Sciences
(Program in online: https://www.plantsciences.uzh.ch/en/Outreach/Symposia/Symposium-2025.html)

Winter School 2025: Program

Harnessing Machine Learning for Breakthroughs in Plant and Environmental Sciences

Monday, March 10, 2025

Introductory Lecture: Machine Learning Refresher
(Prof. Dr. Jan Dirk Wegner, UZH/ETHZ)

ABSTRACT
This introductory lecture will provide a short recap of the most important elements of machine learning. To motivate the topic, several examples from a wide range of applicants are described. The basic machine learning workflow is presented, and basic techniques are discussed like k-fold cross-validation to avoid train-test split biases. Regression, classification, clustering and deep learning and the most important methods from an application perspective are presented. Linear regression, ridge regression, LASSO regression are discussed for regression. A summary of Random Forests is explained for classification while k-means and mean shift methods are explained for clustering. Deep learning is applicable for any of the above, regression, classification and clustering but emphasis will be on convolutional neural networks (CNNs). The lecture closes with a short summary and take-home message.

Machine Learning for Predicting Ecosystem Fluxes
(Prof. Dr. Benjamin Stocker, UNIBE) – Practical workshop

ABSTRACT
In this workshop, we use ecosystm flux data and parallel measurements of meteorological variables to model ecosystem gross primary production (the ecosystem-level CO2 uptake by photosynthesis). These data and prediction task is used to introduce fundamental methods of machine learning (data splitting, model training, random forest algorithm) and their implementations in R. After this course, you will

- Understand how overfitting models can happen and how it can be avoided.
- Implement a typical workflow using a machine learning model for a supervised regression problem.
- Evaluate the power of the model.
- Visualise results.

Technical requirements: Please install **R and R Studio**, Basic programming.
Alternative title: ml4ec - Machine Learning for Eddy Covariance
Course materials: https://stineb.github.io/ml4ec_workshop/

Plankton Classification: Opportunities, Methods and Pitfalls
(Dr. Marco Baity Jesi & Juan-Felipe Ruiz, EAWAG) – Practical workshop

ABSTRACT
Plankton are essential to aquatic ecosystems, forming the foundation of the food web and serving as indicators of environmental change. Traditional plankton monitoring methods are labor-intensive and often miss rapid community shifts, such as harmful algal blooms. Automated high-frequency imaging systems overcome these limitations, by enabling real-time observation of plankton organisms. The images generated by these systems must however be annotated. Given the extremely high volume of images, this is not feasible manually, and one usually resorts to deep learning classifiers.

Content
In this module, we will go through the basic steps that need to be followed to implement a plankton classifier. The module consists of 1 hour of theory, and 2 hours of hands-on work.

The theory session will be about good practices around machine learning classification, and will be

targeted to the hands-on session. The topics will be:

• Dataset curation and splitting
• Transfer learning
• Common solutions to common problems (e.g. class imbalance)
• Hyper-parameter tuning and creating a test harness
• Model selection and evaluation
• Dataset shift

In the hands-on session, the students will:

• Download a plankton dataset
• Do stratified splitting of the data into train/validation/test sets
• Download a pretrained deep neural network
• Perform hyper-parameter tuning (image resizing strategies, strategies to address class imbalance, learning schedules)
• Evaluate on the validation set and pick best model candidate
• Retrain the best model on train+validation set, and evaluate it on the test set
• Evaluate the models on the test set
• Calculate metrics
• Evaluate metrics in a deployment setting, and observe performance degradation related to dataset shift

Requirements
Being comfortable with python coding, and with basic shell scripting (bash).

Winter School 2025: Program

Harnessing Machine Learning for Breakthroughs in Plant and Environmental Sciences

Tuesday, March 11, 2025

Remote Sensing: Satellite Data Selection and Preparation for Machine Learning Applications
(Dr. Jochem Braakhekke, UZH) Practical workshop

ABSTRACT
Satellites provide a powerful means to monitor land and ocean processes, track environmental changes, and gain insights into our planet's dynamics. With a growing variety of freely accessible and commercial satellite data, researchers, agencies, and the public have more opportunities than ever to explore Earth observation (EO) data.
This workshop begins with an introduction to satellite data—what it is, how it differs from traditional image data, what datasets are currently available and what can we expect from upcoming missions relevant to environmental and plant sciences. We will also explore key considerations for using ML for EO data, showcasing real-world examples of ML-driven EO applications.
Participants will dive hands-on into the visually stunning world of EO data using different tools and browsers. The aim is to equip participants with the knowledge and tools to leverage EO data effectively.
Requirements:
Create a free account with the Copernicus Data Space Ecosystem to fully utilize the Copernicus Browser. Simply click Login and register here: https://browser.dataspace.copernicus.eu.
\
Workshop Format: The session is designed to be interactive, and your feedback is essential. Depending on participants' interests and know-how, we may split into no-code and Python groups to keep everybody on board.

Plant Species Identification from Photos and Local Habitat Conditions
(Dr. Philipp Brun & Manuel Popp, WSL) Practical workshop

ABSTRACT
We will explore how deep neural networks can be used to both identify plant species from images and to model species distributions. The workshop is structured into five practicals.

In the first practical, we investigate how well the identification of *Calamagrostis* – a genus of difficult-to-identify grass species - works using FlorID, a state-of-the-art plant identification service. We will make customized API requests and evaluate the responses.
In the second practical, we explore and prepare our own set of observational data on the distribution of *Calamagrostis*.
In the third practical, we train a deep neural network, a multilayer perceptron, that predicts the relative observation probabilities of the different *Calamagrostis* species based on local habitat conditions. In other words, we will create a model that gives us a prior expectation on which species is most likely observed at each location in Switzerland.
In practical number four, we evaluate the performance of our model and use it to make spatial predictions to an exemplary region in Switzerland.
Finally, in the fifth practical we assess ways to ensemble image-based and ecological probabilities and investigate to which extent the ecological predictions can improve identification success.

Requirements: basic Python skills and Jupyter Notebooks.
Libraries: PyTorch, torchvision, pandas, rasterio, pillow-timm (instalation will be guided on site).

Individual Tree Species and Health Detection Using Deep Learning Model
(Xia Zhongyu, ETHZ) - Practical workshop

ABSTRACT
Sustainable and effective forest management depends on timely and accurate inventory and monitoring of forest resources. Information on tree species and health conditions is particularly critical, as it forms the basis for assessing the ecological and economic value of forests. Traditional field surveys, while

valuable, are labour-intensive and costly. Over the past century, remote sensing has emerged as a powerful tool to enhance efficiency in forest monitoring. In the last decade, advancements in data accessibility, spatial resolution, and machine learning—particularly deep learning—have transformed forest inventory methods. Convolutional Neural Networks (CNNs) now enable the detection, mapping, and quantification of individual tree species and their health status from high-resolution aerial imagery in an automated and efficient manner. This innovation provides fine-grained, actionable information, facilitating sustainable forest management and advancing ecological research, making it an active field of study nowadays.

Introduction: a brief introduction to high-resolution aerial imagery and CNNs. I could present how their application in individual tree detection, and how we train the CNN model.
Hands-on Exercise: As training DL models is a time-consuming and computation-intensive process, it is not feasible to train models by students during the workshop. Instead, we can apply the pre-trained model on images, to have an understanding about how the model works. The exercise will be conducted on Google Colab.
Requirement: Students are expected to have a basic knowledge of python and remote sensing imagery.

Winter School 2025: Program

Harnessing Machine Learning for Breakthroughs in Plant and Environmental Sciences

Wednesday, March 12, 2025

Image analysis from a 'phenotypers' perspective: The example of image-based wheat head detection to analyze heading dynamics in a diverse winter wheat genotype set
(PD. Dr. Andreas Hund & Dr. Lukas Roth, ETHZ) - Practical workshop

ABSTRACT
Winter wheat is one of the world's most important crops for direct human consumption. Increasingly adverse climatic conditions threaten this global food resource. To accelerate the breeding of adapted varieties, high-throughput field phenotyping has become an important tool to better inform breeders' selection decisions. Heading date, the point in time when wheat heads become visible in the field, is an important physiological trait that is closely linked to the yield performance of plants. Number of wheat heads per area is another crucial component related to yield.
In this tutorial, you will use a pre-trained deep learning model to detect wheat heads in field phenotyping images. To validate the performance of the model for your specific use case, you will annotate and compile an independent test set. Finally, you will analyse the dynamics of head appearance to determine heading date.
In summary, you will learn how to use state-of-the-art deep learning models with proprietary and publicly available phenotyping image data to gain new physiological insights into crop growth and performance. Workshop will be done in Jupyter Notebook accessible via Moodle.

Workshop: PlantSeg 2.0: The Machine Learning Behind
(Dr. Lorenzo Cerrone UZH) - Practical workshop

ABSTRACT:
This workshop will introduce PlantSeg 2.0, the latest version of the popular tool for segmenting large volumetric images. Together, we will walk through the machine learning models powering PlantSeg, learn how to use its new napari-based GUI, select pre-trained networks, adjust parameters, correct results, and launch headless jobs.

After this introduction, the workshop will split into two tracks. Participants interested in applying PlantSeg to their own data can do so. At the same time, those keen to explore the underlying machine learning can follow a hands-on Python tutorial showing how the method works under the hood.

Thursday & Friday, March 13-14, 2025: Symposium 2025:

Harnessing Machine Learning for Breakthroughs in Plant and Environmental Sciences

Please see the program & abstracts online:
https://www.plantsciences.uzh.ch/en/Outreach/Symposia/Symposium-2025.html