Introduction to the course

Artificial Intelligence and Data Mining Methods in Ecology

University of Tuscia – Viterbo, July 21–25, 2025

Alex Falcon
Beatrice Portelli
University of Udine

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

About us

Alex Falcon falcon.alex@spes.uniud.it
Beatrice Portelli portelli.beatrice@spes.uniud.it

Post-doc Research Fellows at AILab – University of Udine

- DMIF – Department of Mathematics, Computer Science and Physics
- DI4A – Department of Agricultural, Food, Environmental and Animal Sciences

About you?

- What is your background? (BSc/MSc)
- What is your current research topic?
- Have you ever collaborated with people from CompSci / EnvSci?

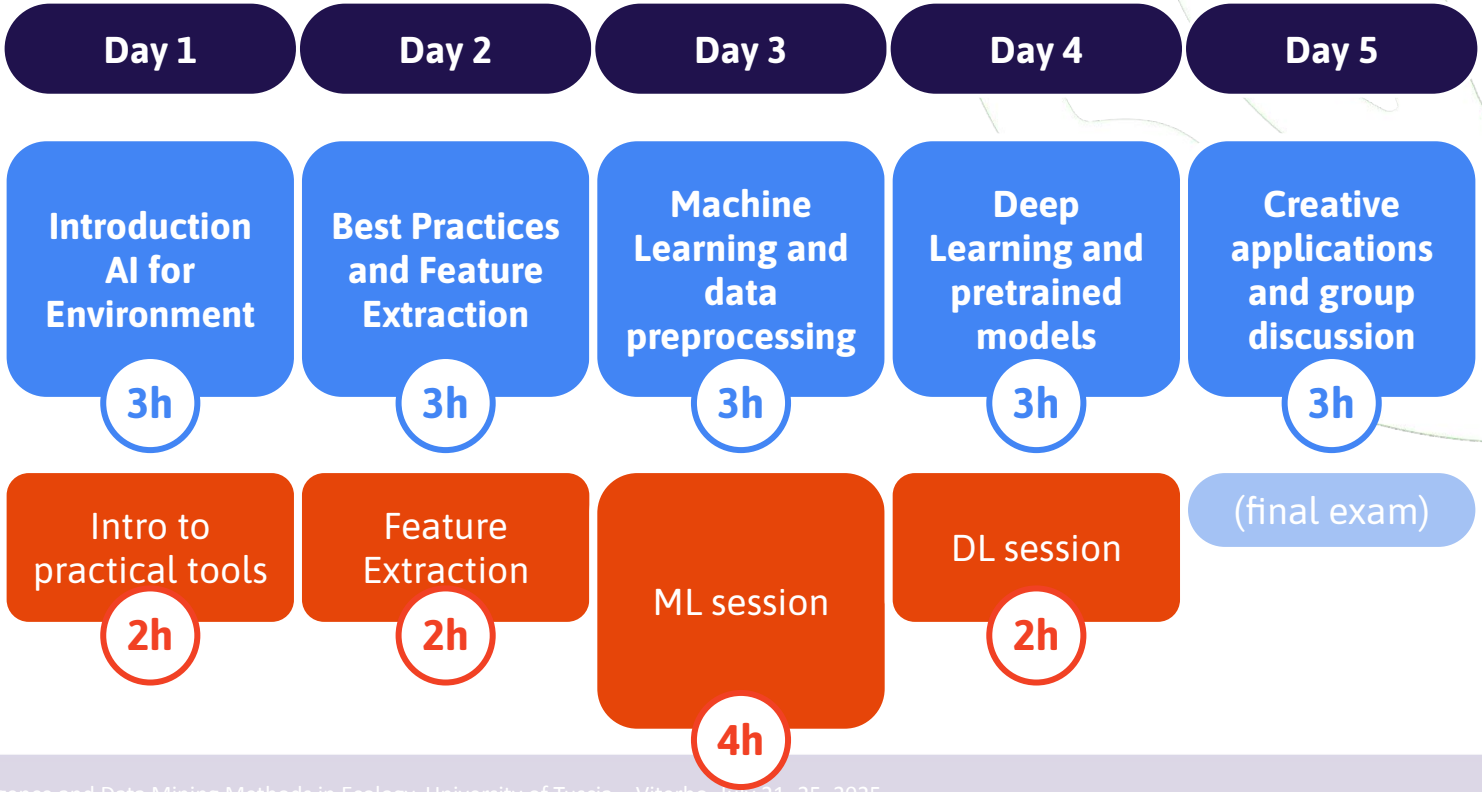
About you?

- What is your background? (BSc/MSc)
- What is your current research topic?
- Have you ever collaborated with people from CompSci / EnvSci?
- Have you ever used statistical models in your research?
- Have you ever used machine learning models in your research?
- Have you ever used machine deep learning models in your research?

About you?

- What is your background? (BSc/MSc)
- What is your current research topic?
- Have you ever collaborated with people from CompSci / EnvSci?
- Have you ever used statistical models in your research?
- Have you ever used machine learning models in your research?
- Have you ever used machine deep learning models in your research?
- Have you ever coded? (In which language?)
- Have you ever written in python?
- Do you know about the following python libraries?
  opencv (cv2), pandas, scikit-learn (sklearn), pytorch (torch)

About the course!



Day 1

Why AI for Environment?

Why Are We Talking About AI in Environmental Sciences?

Why AI?
AI is everywhere

AI is everywhere

AI is everywhere

Professional environments

finance

industry

health

others

General public

language translation

image generation

audio/video generation

others

AI is an umbrella term

Machine Learning - ML

- statistical models and algorithms
- **hand-crafted** features
- needs experts
- relies on **clean** data
- can work on **small** datasets
- **white box** models

AI is an umbrella term

Machine Learning - ML

- statistical models and algorithms
- **hand-crafted** features
- needs experts
- relies on **clean** data
- can work on **small** datasets
- **white box** models

Deep Learning - DL

- complex relationships and patterns
- **automated** feature extraction
- less expert intervention
- can deal with **noisy** data
- needs **large** datasets
- **black box** models

AI is…
not so widespread in environmental sciences!

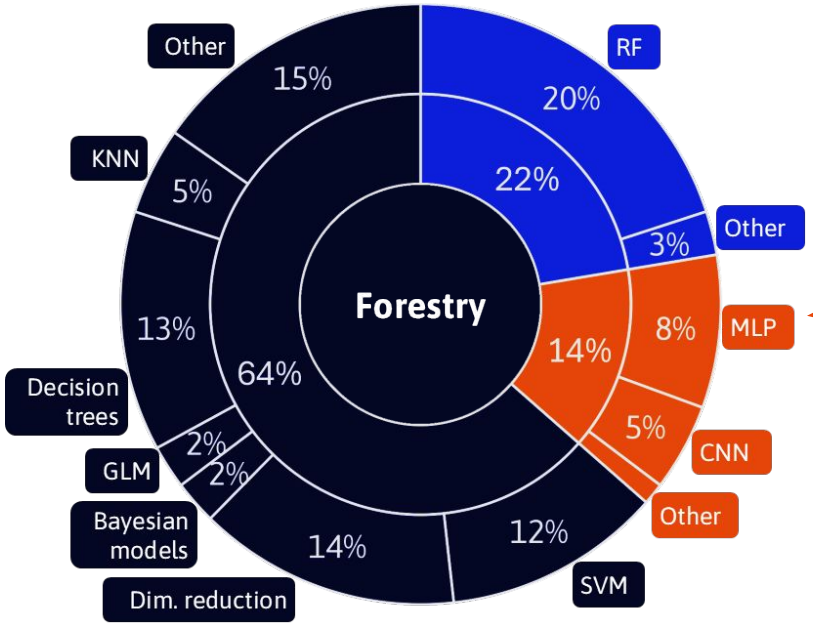
According to multiple survey papers on ML and DL applications in

- Forestry
- Agriculture
- Wastewater management

AI is…

not so widespread in environmental sciences!

(especially DL)



ML – Ensemble

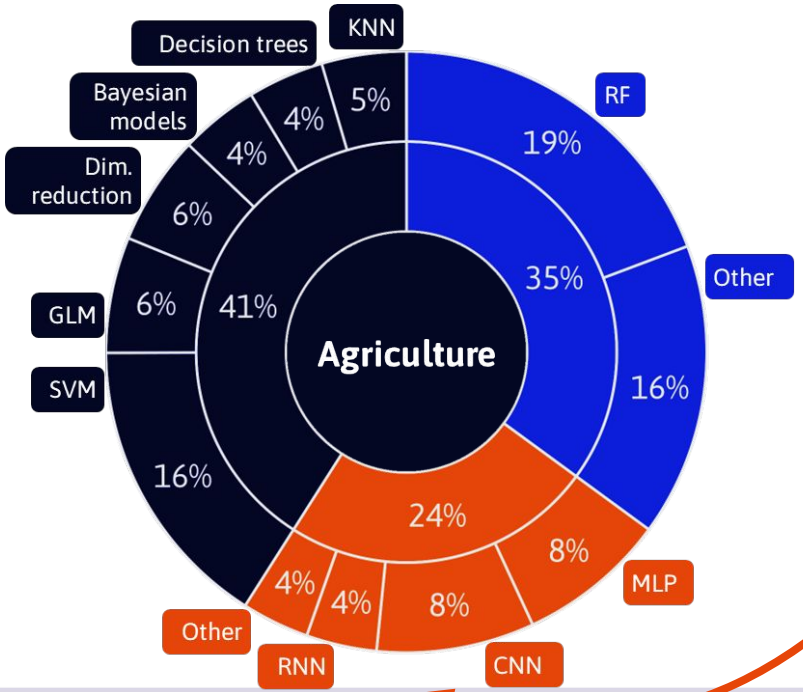
ML – Other

DL

AI is…

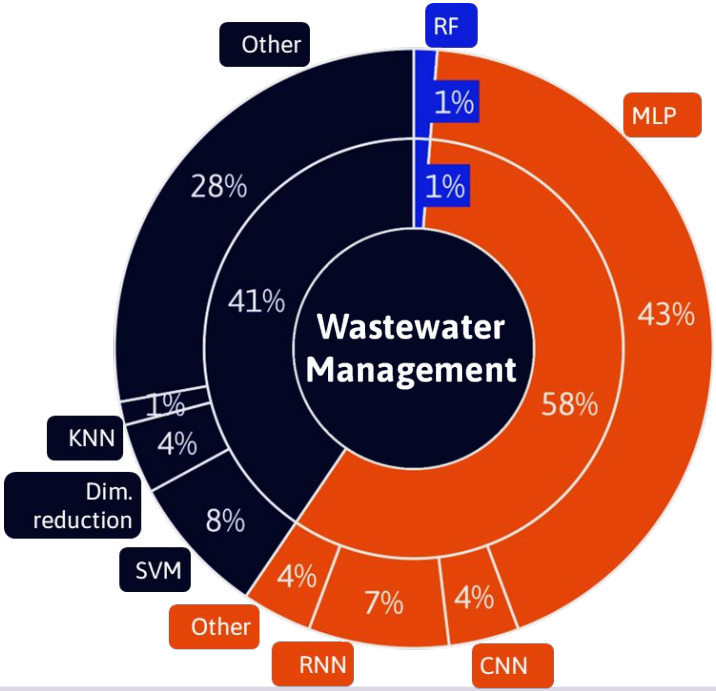
not so widespread in environmental sciences!

(especially DL)



- ML – Ensemble
- ML – Other
- DL

AI is…
not so widespread in environmental sciences!



This is a little bit of an outlier...

- Data is more structured, standardized, homogeneous
- Data collection form sensors, easy to get large time series
- Easier to produce labeled data using long-term observations or controlled trials

AI is…
not so widespread in environmental sciences!

- Handwritten and manually updated Excel databases are still going strong
- → **small** datasets are still the norm

AI is…
not so widespread in environmental sciences!

- Handwritten and manually updated Excel databases are still going strong
- → **small** datasets are still the norm
- Agronomists go to the field and quantify plant coverage “by hand”

AI is…
not so widespread in environmental sciences!

- Handwritten and manually updated Excel databases are still going strong
- → **small** datasets are still the norm
- Agronomists go to the field and quantify plant coverage “by hand”
- Vineyard monitoring mostly done by hand

AI is…
not so widespread in environmental sciences!

- Handwritten and manually updated Excel databases are still going strong
- → **small** datasets are still the norm
- Agronomists go to the field and quantify plant coverage “by hand”
- Vineyard monitoring mostly done by hand
- Nutritional value estimation is rarely used by diet experts

AI is…
not so widespread in environmental sciences!

- Handwritten and manually updated Excel databases are still going strong
- → **small** datasets are still the norm
- Agronomists go to the field and quantify plant coverage “by hand”
- Vineyard monitoring mostly done by hand
- Nutritional value estimation is rarely used by diet experts

What about you? Have you seen AI-related applications in your field?

Why EnvSci?
Big questions, urgent challenges

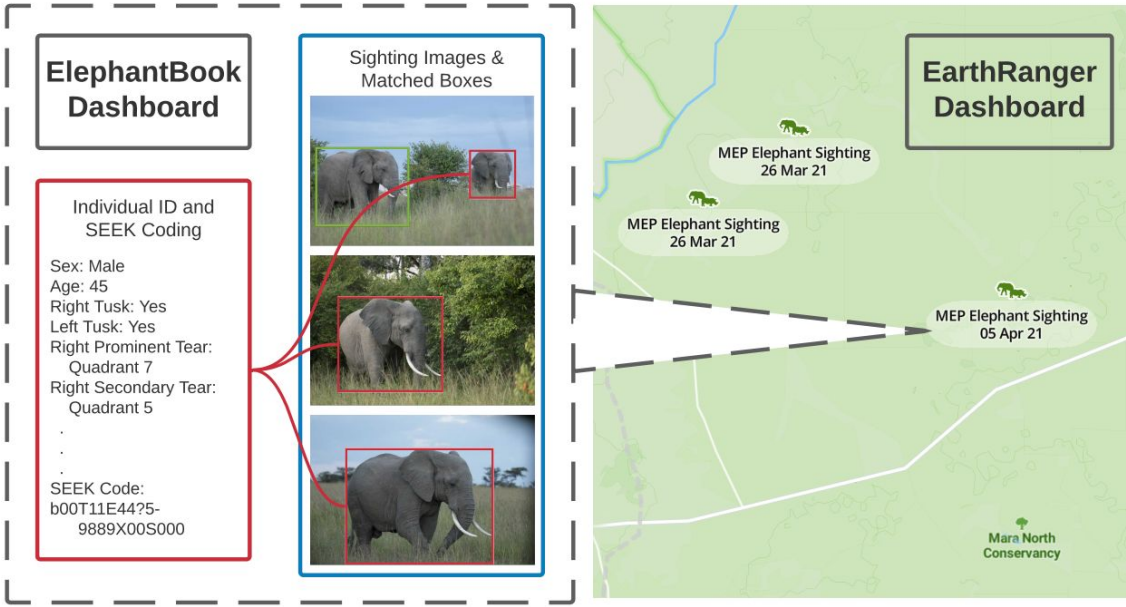
Environmental sciences face complex, urgent problems, like:

- 🌍 Biodiversity loss and conservation
- 🦠 Disease spread and pandemics
- 🌾 Sustainable agriculture
- 🌡️ Climate change monitoring
- 📡 Data deluge from satellites, sensors, field surveys...

These challenges demand new tools for understanding and action

Some examples from the literature

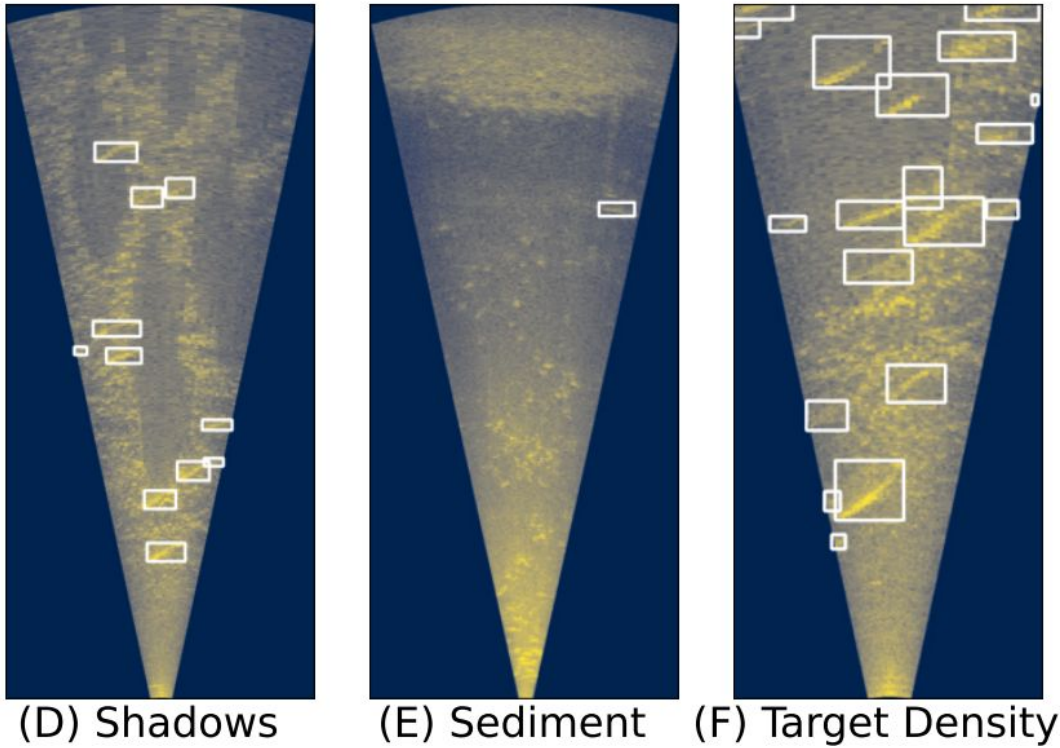
- Re-identifying instances of specific endangered species
  - Elephants, here
- Useful for biodiversity monitoring, conservation efforts



Kulits et al. *ElephantBook: A Semi-Automated Human-in-the-Loop System for Elephant Re-Identification.* SIGCAS 2021

Some examples from the literature

- Counting of salmon instances in Alaskan rivers
  - Using sonar data
  - Very challenging
- Note this is otherwise done *manually* by field technicians
- Useful for biodiv and migration monitoring



Kay et al. The Caltech Fish Counting Dataset: A Benchmark for Multiple-Object Tracking and Counting. ECCV 2022

Some examples from the literature

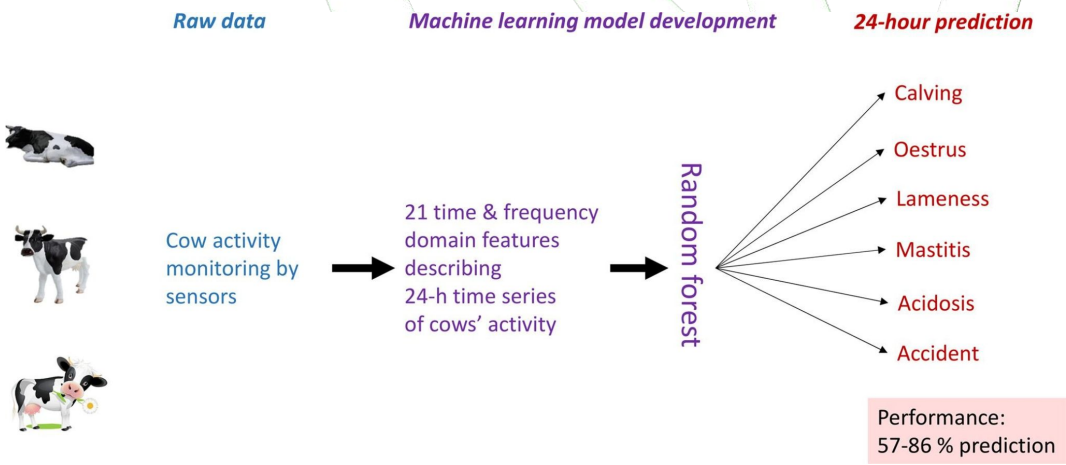
- Detection of diseases in leafs
  - Can be performed on drones → kind-of automatic
- Useful for epidemic management, possibly more sustainable agriculture

Disease	Diseased Image	Causative agent	Symptoms
Powdery Mildew		Fungus	White powdery spots on both sides of leaves, yellowing and wilting leaves
Anthracnose		Fungus	Small, water-soaked lesions become sunken and dark, lesions are pink spore masses, leaf distortion, and curling
Angular Leaf Spot		Bacteria	Angular, water-soaked lesions with yellow halo, lesions later become necrotic leaf curling and distortion
Fusarium Wilt		Fungus	Yellowing of leaves, and small necrosis is appeared

Jafar et al. Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. Front Plant Sci 2024

Some examples from the literature

- Identification of pathological, reproductive, or stress conditions in cows
- Useful for better management of cows
  - farm animals in general



Lardy et al. Discriminating pathological, reproductive or stress conditions in cows using machine learning on sensor-based activity data. Comp Ag 2023

Big questions, urgent challenges

The same, on a smaller scale, goes for your PhD research questions:

- How can I quickly estimate the biodiversity of an habitat?
- How can I estimate the effect of agricultural practices on yield?
- How can I leverage thousands of images from drones or satellites?
- How can I find hidden patterns in my data?

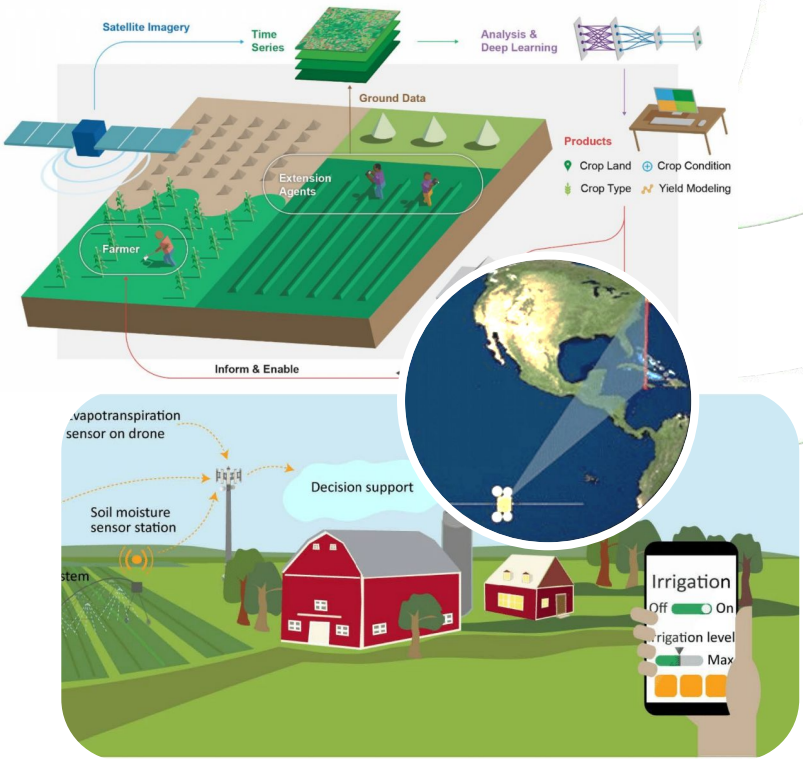
📌 These are **data-driven** questions… and data is now everywhere!

Data - Once rare, now everywhere

- In the past collecting data was an important bottleneck

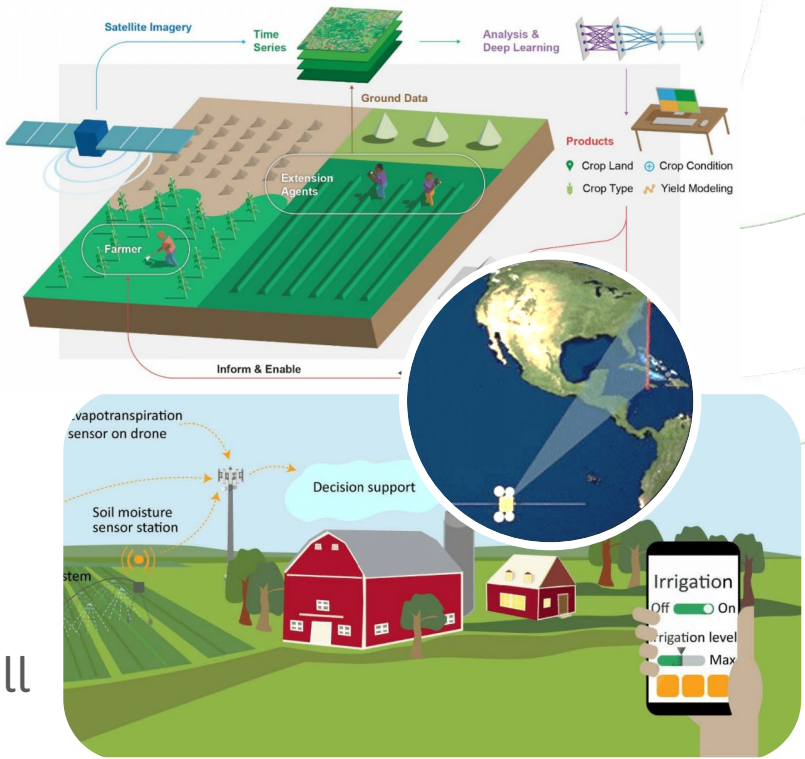
Data - Once rare, now everywhere

- In the past collecting data was an important bottleneck
- Now we are flooded with data!
  - 🌐 IoT
  - 🛰️ satellites
  - 📱 apps
  - 🧪 labs
  - 🌱 field sensors
  - ...



Data - Once rare, now everywhere

- In the past collecting data was an important bottleneck
- Now we are flooded with data!
  - 🌐 IoT
  - 🛰️ satellites
  - 📱 apps
  - 🧪 labs
  - 🌱 field sensors
  - …
- But it's complex, noisy, incomplete
- What we need: ways to make sense of it all



Data - Overarching Problem

Several issues in collecting and processing data

https://rewildingeurope.com/news/camera-trap-studies-support-wildlife-comeback-in-the-rhodope-and-velebit-mountains/
https://connectedconservation.foundation/news/can-satellite-imagery-and-ai-modernise-wildlife-surveys-in-africa/

Data - Overarching Problem

Several issues in collecting and processing data:

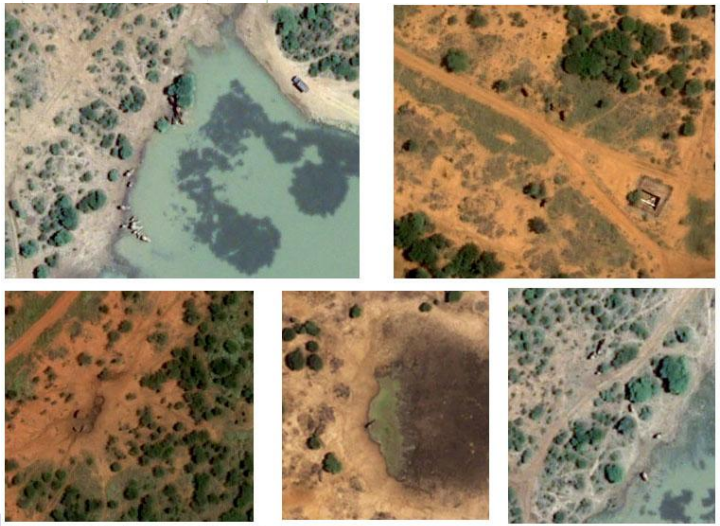
- High heterogeneity in *format* and *perspective*
  - Numbers, images, videos, audio, 3D reconstructions, ...

https://rewildingeurope.com/news/camera-trap-studies-support-wildlife-comeback-in-the-rhodope-and-velebit-mountains/
https://connectedconservation.foundation/news/can-satellite-imagery-and-ai-modernise-wildlife-surveys-in-africa/

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
  - Numbers, images, videos, audio, 3D reconstructions, …
  - e.g. biodiversity monitoring → camera traps, drones imagery, acoustic monitoring, citizen science (photos, surveys, …)



All images: Pléiades Neo © Airbus DS 2022

https://rewildingeurope.com/news/camera-trap-studies-support-wildlife-comeback-in-the-rhodope-and-velebit-mountains/
https://connectedconservation.foundation/news/can-satellite-imagery-and-ai-modernise-wildlife-surveys-in-africa/

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
  - Images/videos with varying degrees of resolution, framerate/shutter speed
    - E.g. gigapixel images in medicine fields *vs* low res crops of small fishes in the wild
  - Different format = different storage and computational requirements

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
  - Typical example: sensors that “suddenly” give wrong readings (e.g. → absurd values)
    - Eg temperatures, humidity in the soil, etc

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
  - Typical example: sensors that “suddenly” give wrong readings (e.g. → absurd values)
    - Eg temperatures, humidity in the soil, etc
  - Subjectivity for some information
    - E.g.: subtle differences in phenological stages that may confuse even an expert
    - E.g.: errors in identifying the species of a fungus, animal, ...

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
- The worst of them all: most of the data is raw and unlabeled

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
- The worst of them all: most of the data is raw and unlabeled
    - Correct labels often require **experts** → expensive, slow, difficult to scale

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
- The worst of them all: most of the data is raw and unlabeled
    - Correct labels often require **experts** → expensive, slow, difficult to scale
        - Attempts to fix with crowdsourcing → cheaper, yet lower quality
        - Auto-annotation, Web-scraped data → even cheaper, even worse quality

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
- The worst of them all: most of the data is raw and unlabeled
    - Correct labels often require **experts** → expensive, slow, difficult to scale
        - Attempts to fix with crowdsourcing → cheaper, yet lower quality
        - Auto-annotation, Web-scraped data → even cheaper, even worse quality
        - Generative AI… interesting, yet likely unreliable and difficult to validate

Data - Overarching Problem

Several issues in collecting and processing data:

- High heterogeneity in *format* and *perspective*
- From *very small* to potentially *very large* samples
- Incomplete data, noisy or even wrong
- The worst of them all: most of the data is raw and unlabeled
    - Correct labels often require **experts** → expensive, slow, difficult to scale
        - Attempts to fix with crowdsourcing → cheaper, yet lower quality
        - Auto-annotation, Web-scraped data → even cheaper, even worse quality
        - Generative AI… interesting, yet likely unreliable and difficult to validate
    - Incoherent and subjective → human bias

How AI can help

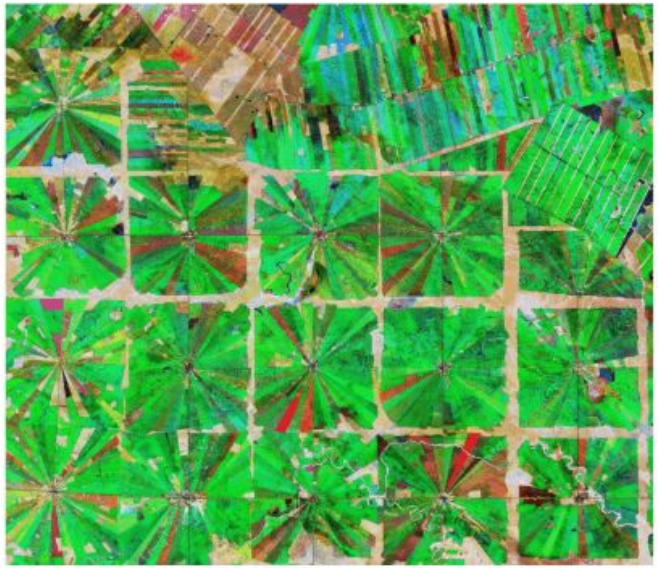
- Distilling raw data
- Optimizing complex systems
- Improving predictions
- Accelerating scientific discovery
- Approximating simulations

How AI can help

- Distilling raw data
  - Getting actionable information from large amounts of raw data
  - ML areas: Computer vision, natural language processing

  Examples

  - Mapping deforestation and carbon stock
  - Gathering data on building footprints/heights
  - Evaluating coastal flood risk



- Optimizing complex systems
- Improving predictions
- Accelerating scientific discovery
- Approximating simulations

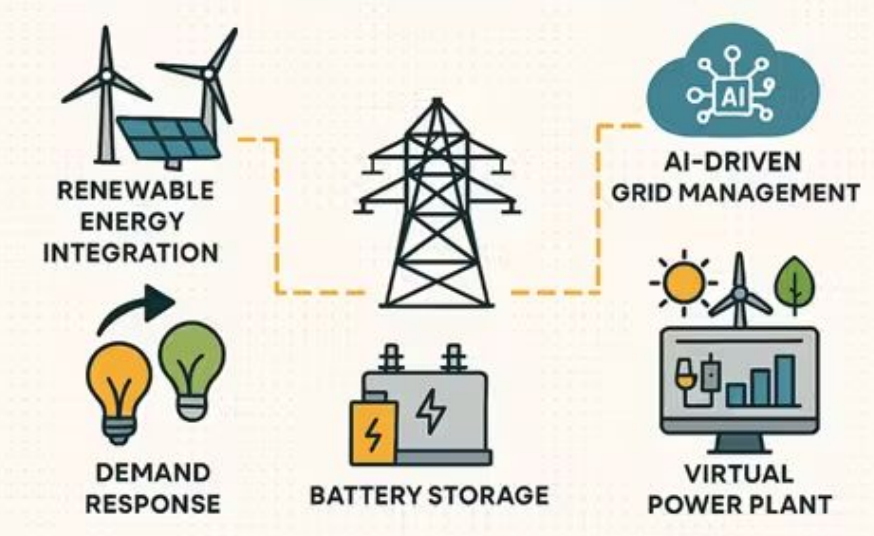
How AI can help

- Distilling raw data
- Optimizing complex systems
  - Improving efficient operation of complex automated systems
  - ML areas: Optimization, control, reinforcement learning

  Examples

  - Design irrigation schedules or routes
  - Demand response in electrical grids

- Improving predictions
- Accelerating scientific discovery
- Approximating simulations



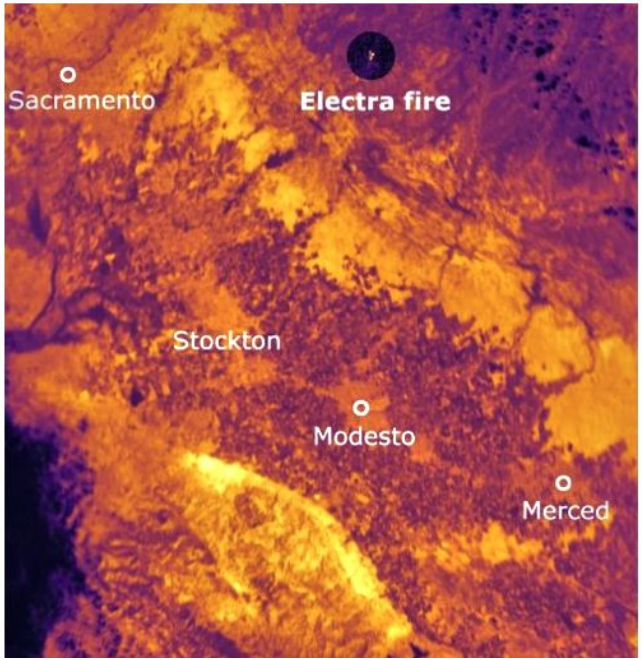
How AI can help

- Distilling raw data
- Optimizing complex systems
- Improving predictions
  - Forecasts and time series predictions
  - ML areas: time series analysis, computer vision

  Examples

  - Forecasting crop yield
  - Forecasting species decline
  - Pest and disease early warning system

- Accelerating scientific discovery
- Approximating simulations

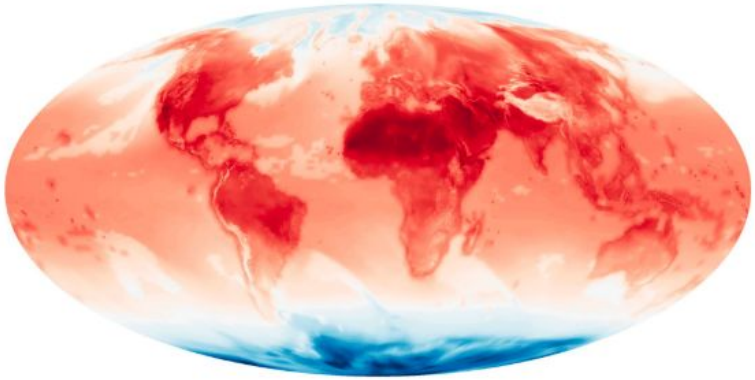


How AI can help

- Distilling raw data
- Optimizing complex systems
- Improving predictions
- Accelerating scientific discovery
  - Suggesting experiments in order to speed up the design process
  - ML areas: generative models, active learning, reinforcement learning, graph neural networks
- Examples
  - Identify candidate materials for batteries
  - Suggest new uses for existing drugs
- Approximating simulations

How AI can help

- Distilling raw data
- Optimizing complex systems
- Improving predictions
- Accelerating scientific discovery
- Approximating simulations
  - Accelerating time-intensive (physics-based) simulations
  - ML areas: physics-informed ML, computer vision, interpretable ML
- Examples
  - Superresolution of predictions from climate models
  - Simulating crop yield predictions based on plant treatments



How AI can help

- Distilling raw data
- Optimizing complex systems
- Improving predictions
- Accelerating scientific discovery
- Approximating simulations

Overall: Think of AI as a toolbox: not one tool, but many which *may* be useful to speedup some aspect of your research.

Examples of AI in Action

In the literature

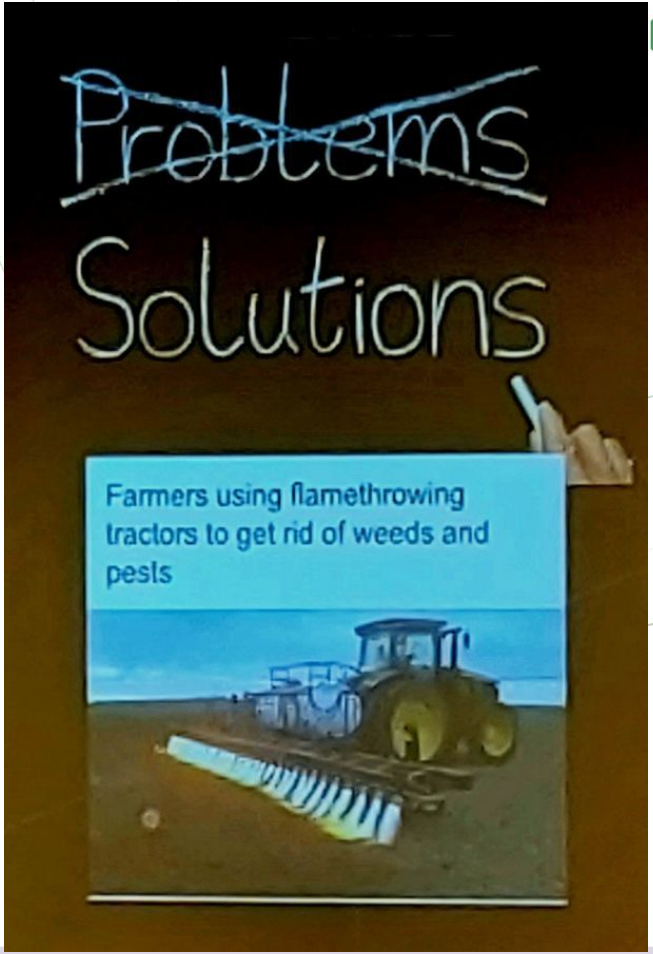
- 🚜 Detecting weeds (and killing them) live with a precision robot
- 🛰️ Field anomaly detection with satellite data
- 🐟 Fish fry health status classification
- 🐄 Bull sperm morphology evaluation

Our case studies

- 🛰️ Estimating forest carbon storage from satellite image using ML/DL
- 🌱 Estimating soil composition and carbon stock using DL models
- 🍇 Grapevine phenological stage estimation
- 🍝 Food nutritional values estimation

Case studies (in the world)

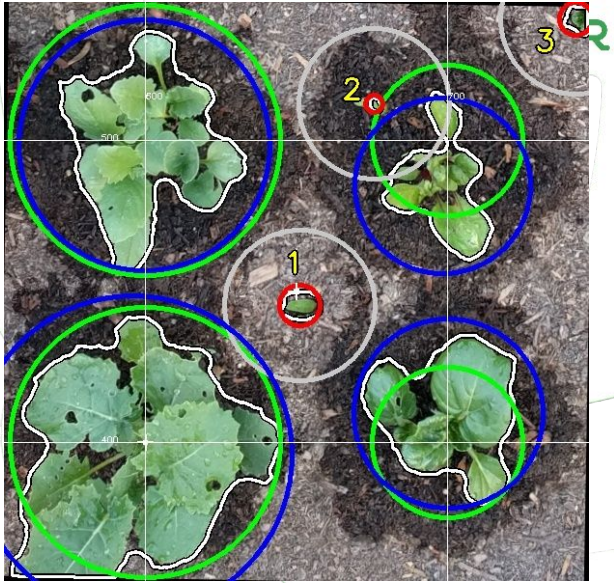
Weed detection in the field



Case studies (in the world)

Weed detection in the field

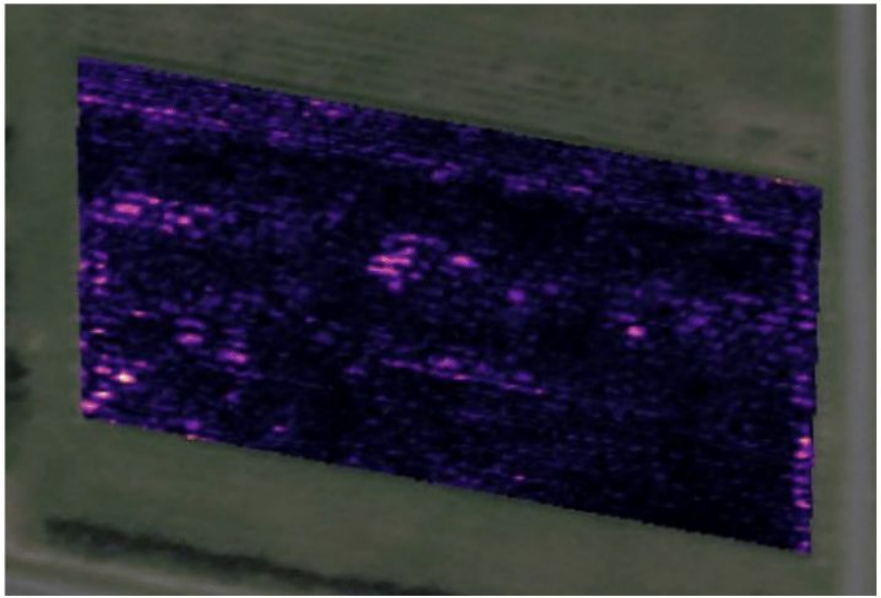
- Computer vision system
  trained to distinguish young crops from weeds
- Integrated with a precision robot, which can burn (laser) weeds



Case studies (in the world)

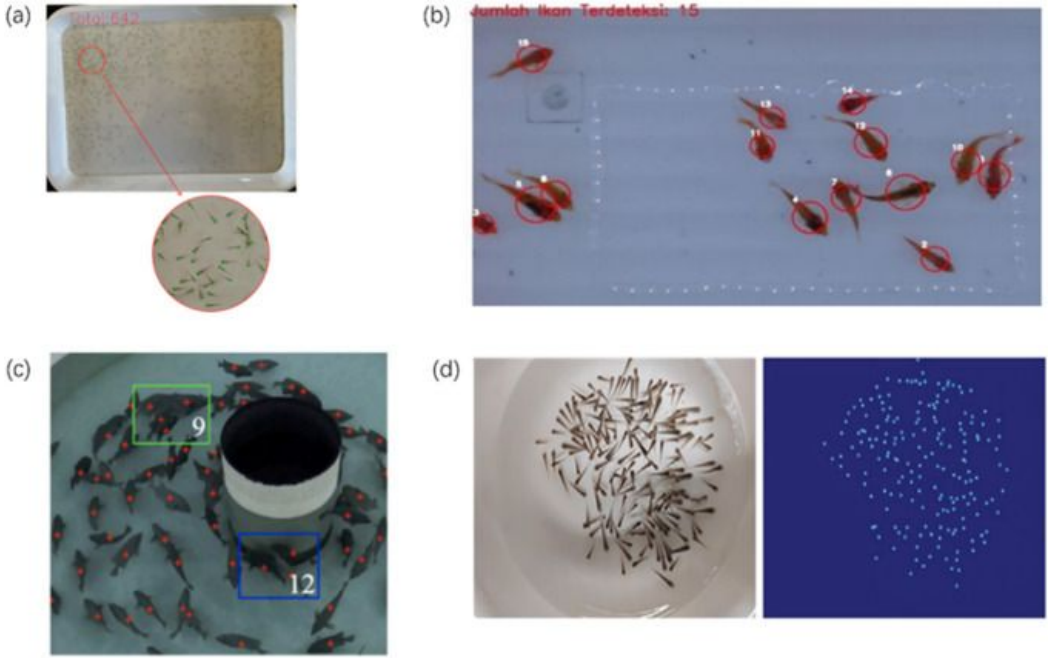
Detection of anomalous field conditions

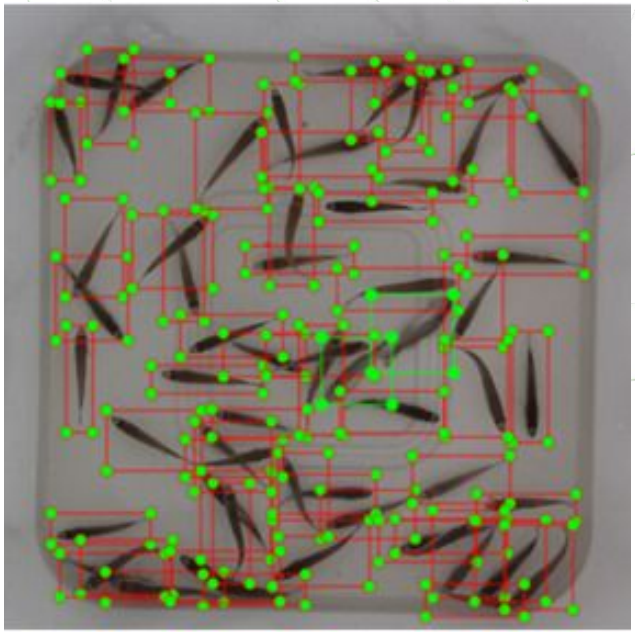
Downy mildew disease



Case studies (in the world)

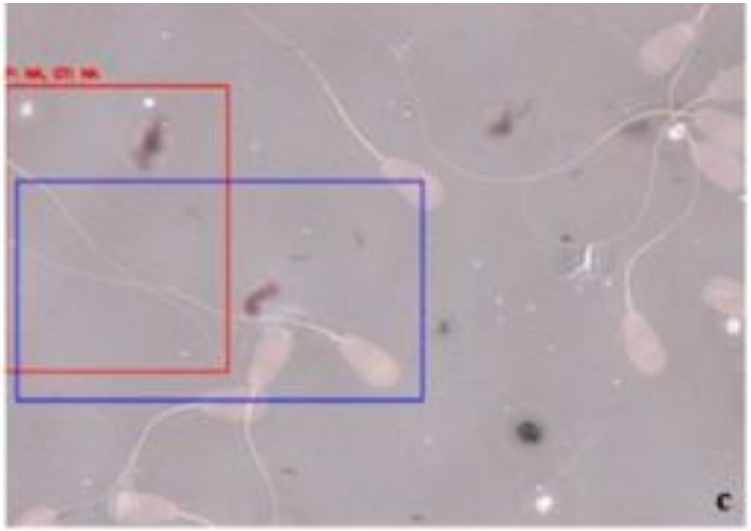
Fish fry health status estimation

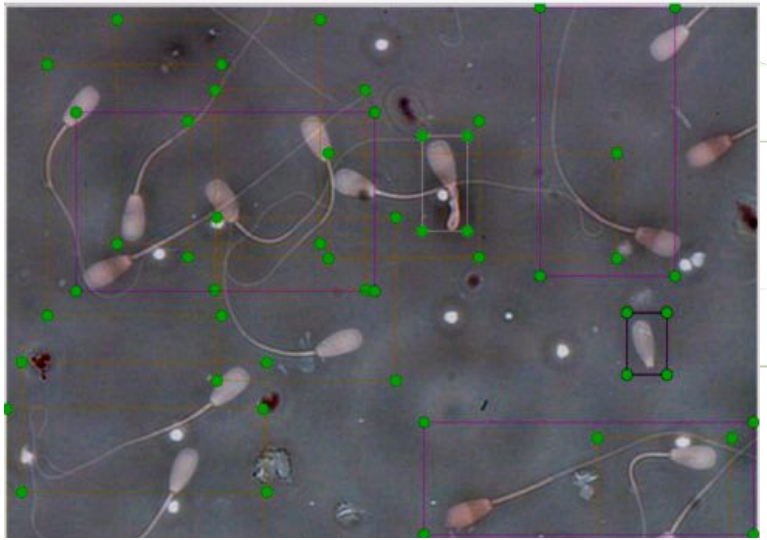




Case studies (in the world)

Bull sperm morphology evaluation





Case studies (involving us)

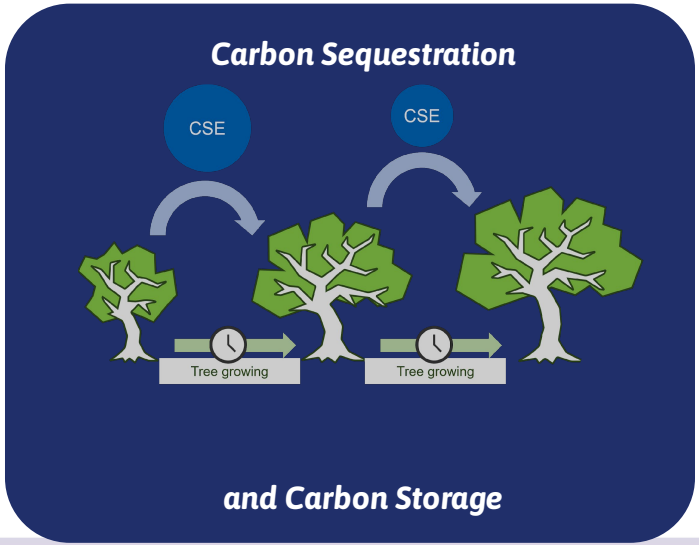
In collaboration with:

- DI4A – *Department of Agricultural, Food, Environmental and Animal Sciences*
  University of Udine
- DAFNAE – *Department of Agronomy, Food, Natural resources, Animals and Environment*
  University of Padova
- TESAF – *Department of Land, Environment, Agriculture and Forestry*
  University of Padova
- DMED – *Department of Medicine*
  University of Udine
- *Department of Clinical Sciences and Community Health*
  University of Milan
- *IRCCS Foundation*
  Ospedale Maggiore Policlinico, Milan

Case studies (involving us)

Carbon storage/sequestration in forests

- **Objective**: estimating how much carbon is sequestered (CSE) and stored (CS) in the forests of Friuli Venezia Giulia



Case studies (involving us)

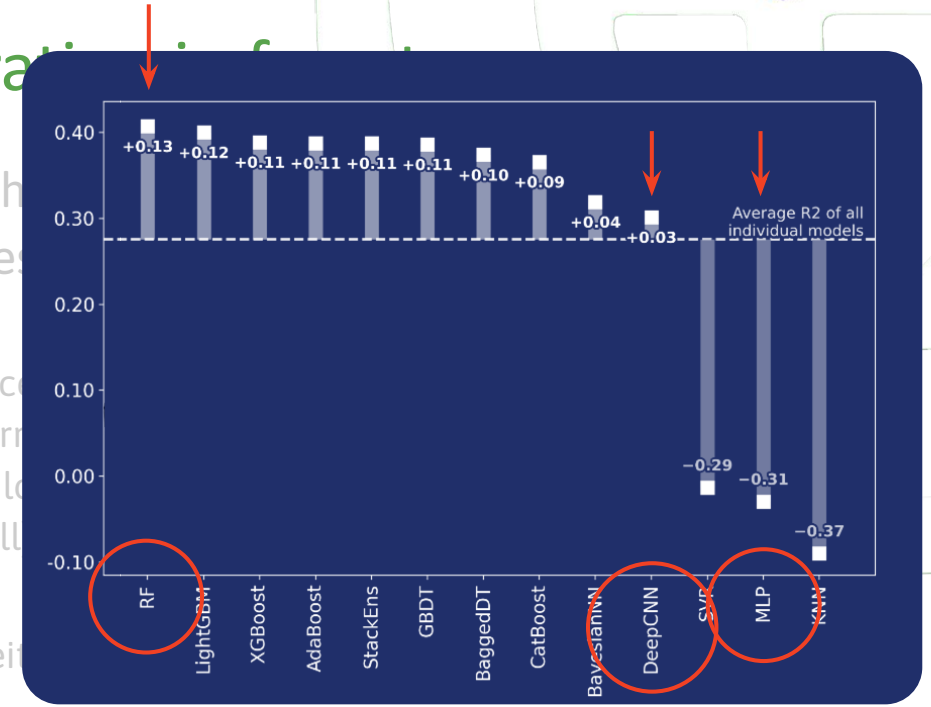
Carbon storage/sequestration in forests

- **Objective**: estimating how much carbon is sequestered (CSE) and stored (CS) in the forests of Friuli Venezia Giulia
- **Input**
  - satellite images (and spectral indices derived from them: NDVI, NDII, ...),
  - geomorphological data (digital terrain model),
  - CHM (canopy height model) from local LiDAR scans,
  - climatic data (temperature, rainfall)
- **Methods**
  - Various ML and DL models using either tabular data or image data

Case studies (involving us)

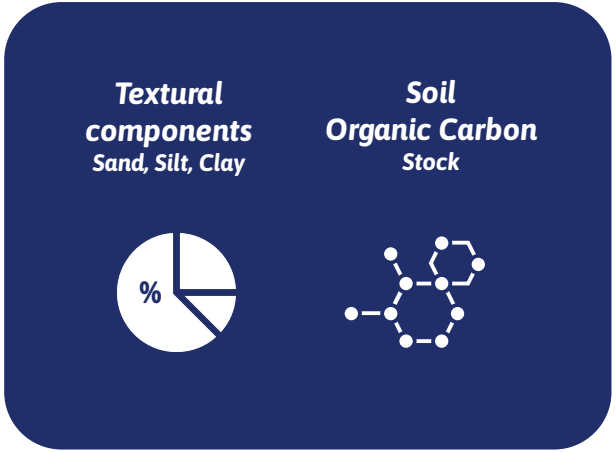
Carbon storage/sequestra

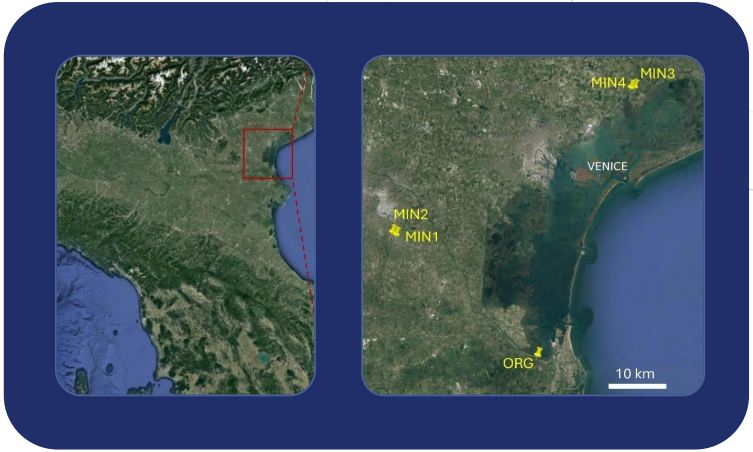
- **Objective**: estimating how much
  (CSE) and stored (CS) in the fores
- **Input**
  - satellite images (and spectral indic
  - geomorphological data (digital ter
  - CHM (canopy height model) from l
  - climatic data (temperature, rainfall
- **Methods**
  - Various ML and DL models using eit
- **Specific challenges / findings**
  - “only” 279 annotated samples (measurements on the field)
  - the most important input feature (CHM) is the most expensive to acquire
  - DL models performed way worse than classical ML models (too few data points)



Soil texture and soil organic carbon estimation

- **Objective**: estimating the composition of the soil (sand, silt, clay) and the SOC stock (soil organic carbon) in different fields





Case studies (involving us)

Soil texture and soil organic carbon estimation

- **Objective**: estimating the composition of the soil (sand, silt, clay) and the SOC stock (soil organic carbon) in different fields
- **Input**
  - EMI sensor → apparent electrical conductivity
  - Gamma-ray spectrometer → radionuclide counts
    Total counts, Potassium (40K), Uranium (238U), Thorium (232Th)
- **Methods**
  - ML and DL models using tabular data

Case studies (involving us)

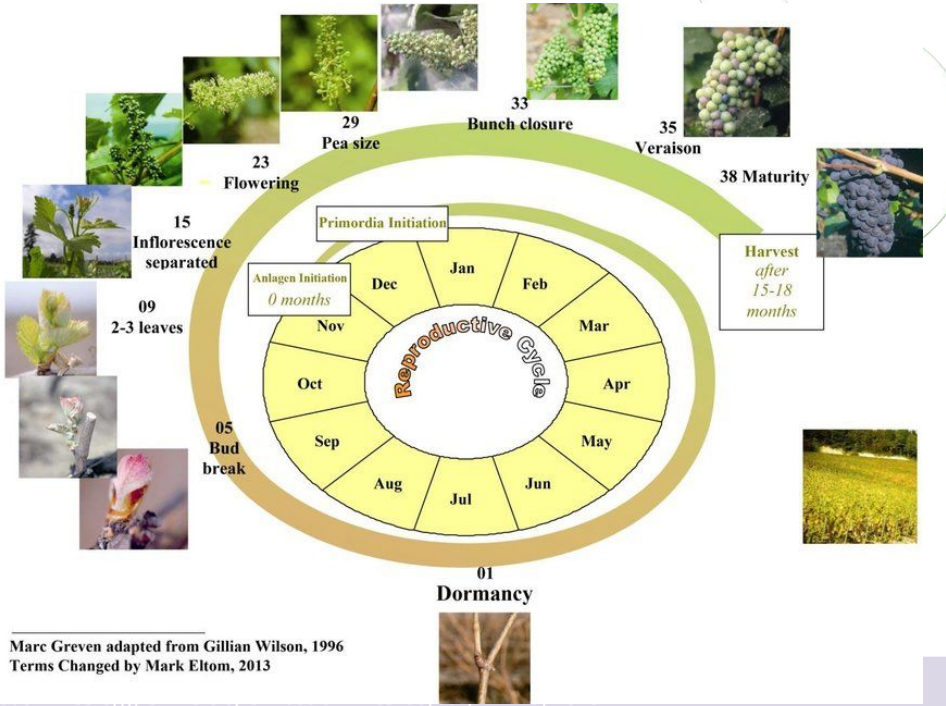
Soil texture and soil organic carbon estimation

- **Objective**: estimating the composition of the soil (sand, silt, clay) and the SOC stock (soil organic carbon) in different fields
- **Input**
  - EMI sensor → apparent electrical conductivity
  - Gamma-ray spectrometer → radionuclide counts
    Total counts, Potassium (40K), Uranium (238U), Thorium (232Th)
- **Methods**
  - ML and DL models using tabular data
- **Specific challenges / findings**
  - “only” 354 annotated samples (measurements on the field)
  - DL models performed better than ML models
  - DL and ML models focused on the same inputs and DL models were explainable (enough)

Case studies (involving us)

Grapevine phenological stages

- **Objective**: estimating the phenological stage (GPHS) by local climatic data



Case studies (involving us)

Grapevine phenological stages

- **Objective**: estimating the phenological stage (GPHS) by local climatic data
- **Input**
  - Daily avg/max/min temperature
  - Daily rainfall
  - Cumulated rainfall
  - Days with temperature >30/35°
  - Growing degree days (based on avg > 10°)
- **Methods**
  - Several ML models: RF, GBDT, CatBoost, ...

Case studies (involving us)

Grapevine phenological stages

- **Objective**: estimating the phenological stage (GPHS) by local climatic data
- **Input**
  - Daily avg/max/min temperature
  - Daily rainfall
  - Cumulated rainfall
  - Days with temperature >30/35°
  - Growing degree days (based on avg > 10°)
- **Methods**
  - Several ML models: RF, GBDT, CatBoost, ...
- **Specific challenges / findings**

Case studies (involving us)

Grapevine phenological stages

- **Objective**: estimating the phenological stage (GPHS) by local climatic data
- **Input**
  - Daily avg/max/min temperature
  - Daily rainfall
  - Cumulated rainfall
  - Days with temperature >30/35°
  - Growing degree days (based on avg > 10°)
- **Methods**
  - Several ML models: RF, GBDT, CatBoost, ...
- **Specific challenges / findings**
  - GPHS assessed by experts on field → **few** labels (11%)

T_{min}	T_{max}	...	CDR	GPHS
-1.4	4.0	...	7.2	53
-0.6	4.1	...	0.1	53
-4.4	4.1	...	0.0	
-2.5	7.6	...	0.0	
-1.0	2.6	...	0.0	55
-0.7	3.9	...	0.0	
...	...	...	...	
0.6	2.6	...	2.1	
0.7	4.2	...	24.4	
2.9	7.1	...	12.1	57
4.9	10.2	...	0.1	
6.3	12.3	...	1.9	

Case studies (involving us)

Grapevine phenological stages

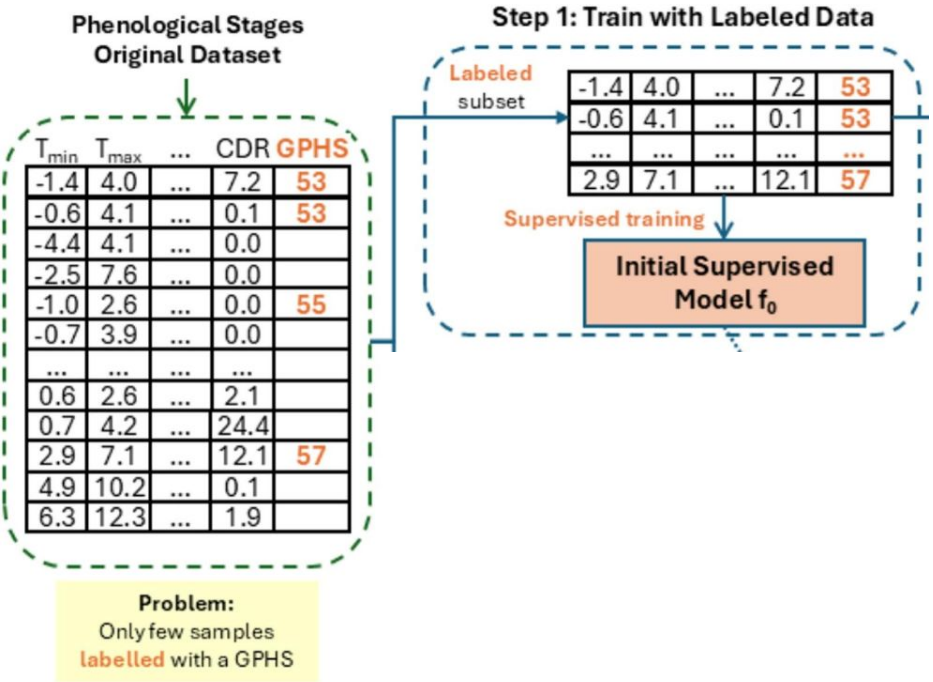
Phenological Stages Original Dataset

T_{min}	T_{max}	...	CDR	GPHS
-1.4	4.0	...	7.2	53
-0.6	4.1	...	0.1	53
-4.4	4.1	...	0.0	
-2.5	7.6	...	0.0	
-1.0	2.6	...	0.0	55
-0.7	3.9	...	0.0	
...	...	...	...	
0.6	2.6	...	2.1	
0.7	4.2	...	24.4	
2.9	7.1	...	12.1	57
4.9	10.2	...	0.1	
6.3	12.3	...	1.9	

Problem:
Only few samples **labelled** with a GPHS

Case studies (involving us)

Grapevine phenological stages

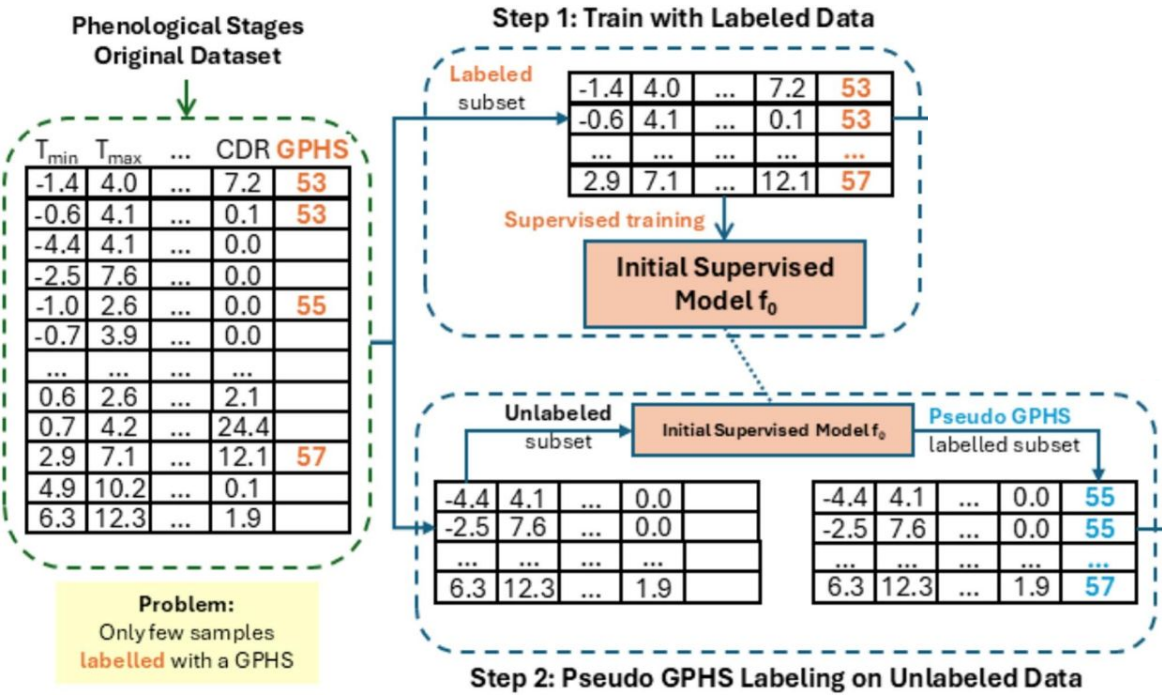


T_{min}	T_{max}	...	CDR	GPHS
-1.4	4.0	...	7.2	53
-0.6	4.1	...	0.1	53
-4.4	4.1	...	0.0	
-2.5	7.6	...	0.0	
-1.0	2.6	...	0.0	55
-0.7	3.9	...	0.0	
...	...	...	...	
0.6	2.6	...	2.1	
0.7	4.2	...	24.4	
2.9	7.1	...	12.1	57
4.9	10.2	...	0.1	
6.3	12.3	...	1.9	

-1.4	4.0	...	7.2	53
-0.6	4.1	...	0.1	53
...	...	...	...	...
2.9	7.1	...	12.1	57

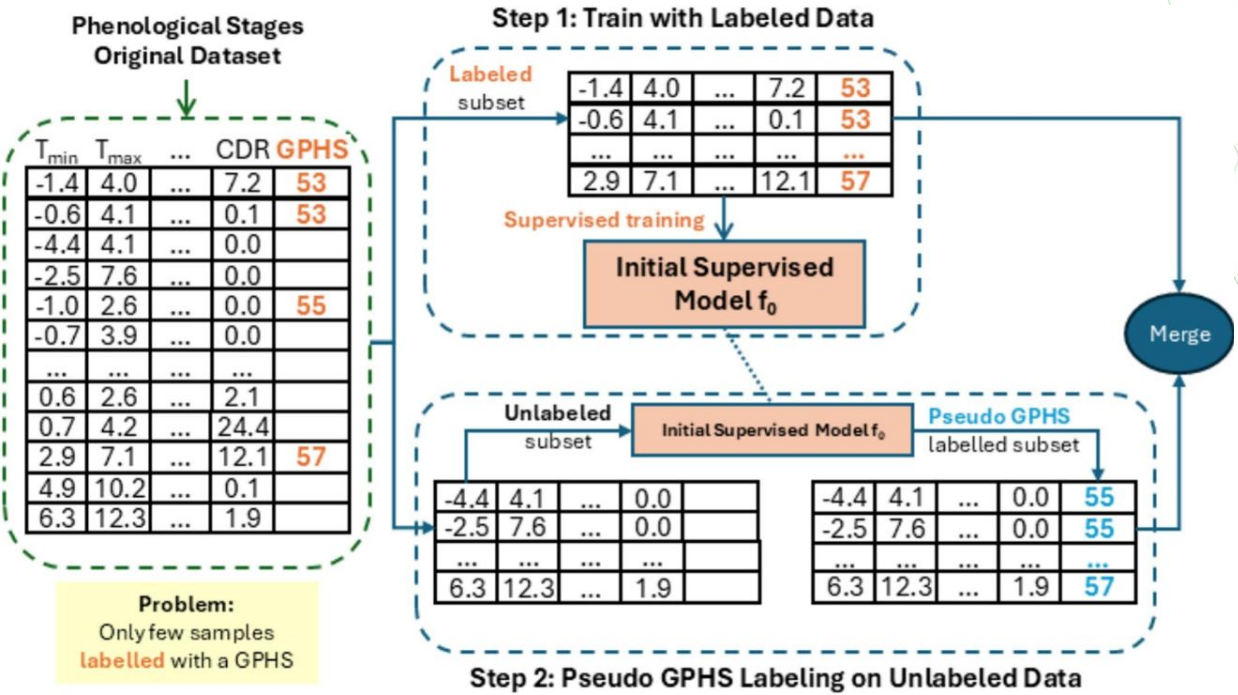
Case studies (involving us)

Grapevine phenological stages



Case studies (involving us)

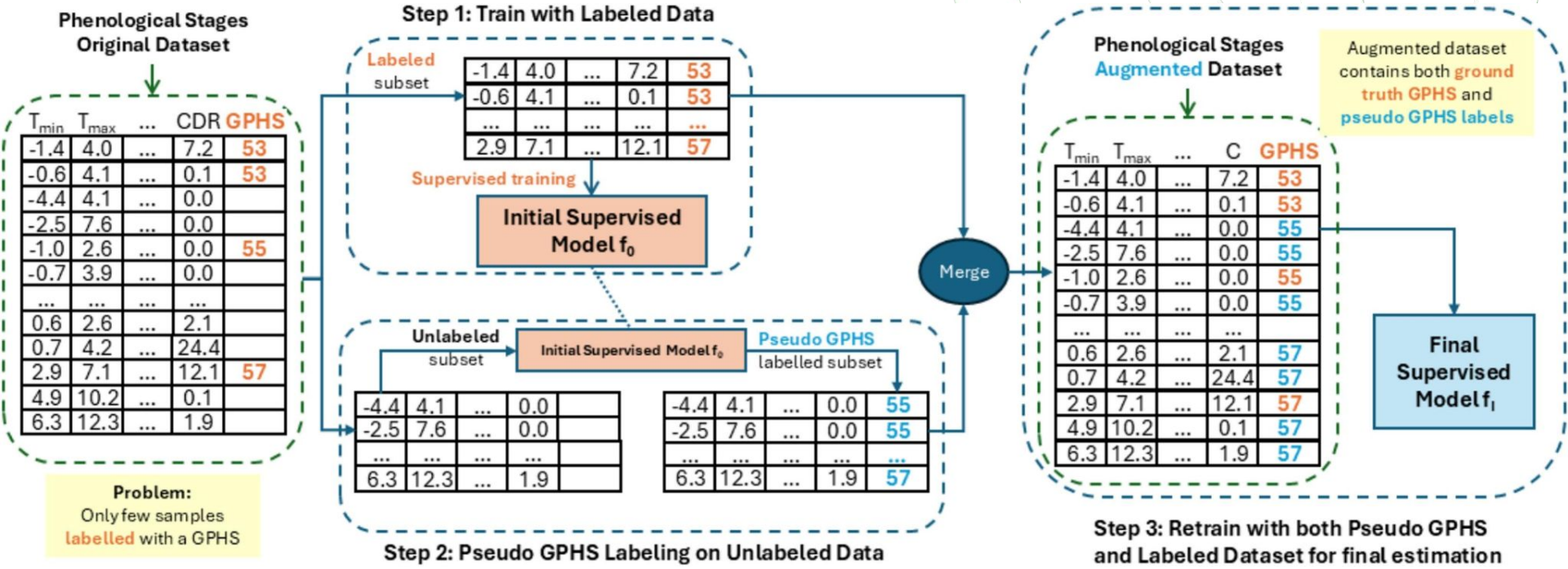
Grapevine phenological stages



T_{min}	T_{max}	...	CDR	GPHS
-1.4	4.0	...	7.2	53
-0.6	4.1	...	0.1	53
-4.4	4.1	...	0.0	
-2.5	7.6	...	0.0	
-1.0	2.6	...	0.0	55
-0.7	3.9	...	0.0	
...	...	...	...	
0.6	2.6	...	2.1	
0.7	4.2	...	24.4	
2.9	7.1	...	12.1	57
4.9	10.2	...	0.1	
6.3	12.3	...	1.9	

Case studies (involving us)

Grapevine phenological stages



Grapevine phenological stages

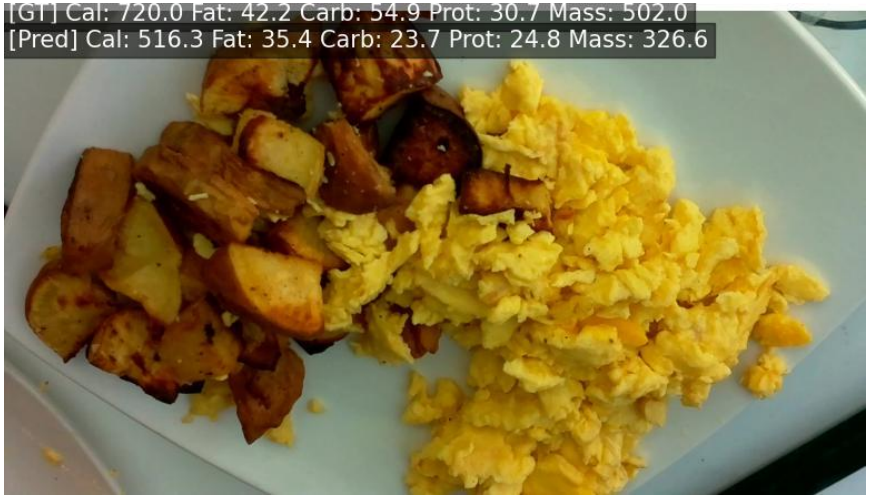
- **Objective**: estimating the phenological stage (GPHS) by local climatic data
- **Input**
  - Daily avg/max/min temperature
  - Daily rainfall
  - Cumulated rainfall
  - Days with temperature >30/35°
  - Growing degree days (based on avg > 10°)
- **Methods**
  - Several ML models: RF, GBDT, CatBoost, ...
- **Specific challenges / findings**
  - GPHS assessed by experts on field → **few** labels (11%)
  - Pseudo-labels help improving performance

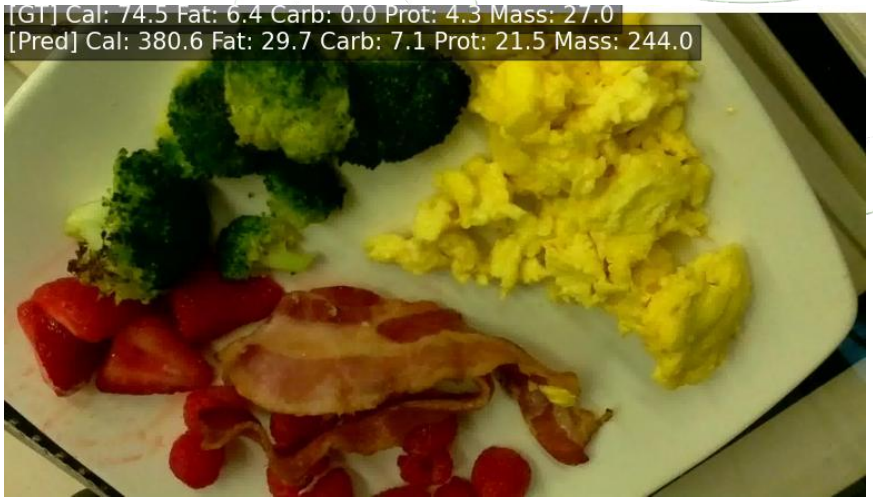
T_{min}	T_{max}	...	CDR	GPHS
-1.4	4.0	...	7.2	53
-0.6	4.1	...	0.1	53
-4.4	4.1	...	0.0	
-2.5	7.6	...	0.0	
-1.0	2.6	...	0.0	55
-0.7	3.9	...	0.0	
...	...	...	...	
0.6	2.6	...	2.1	
0.7	4.2	...	24.4	
2.9	7.1	...	12.1	57
4.9	10.2	...	0.1	
6.3	12.3	...	1.9	

Case studies (involving us)

Food nutritional info estimation

- **Objective**: Predicting nutritional values (energy, carbs, …) from pictures





Case studies (involving us)

Food nutritional info estimation

- **Objective**: Predicting nutritional values (energy, carbs, …) from pictures
- **Input**
  - Food image
- **Methods**
  - Deep learning models to automatically extract features
  - Multilayer perceptrons for nutritional values estimation
- **Specific challenges / findings**
  - **Many**:
    - Depth is difficult to grasp
    - Occlusions
    - Some ingredients are “invisible” (e.g. oil, melt butter, …)
    - “Mixed” dishes (soups, salads)

Case studies (involving us)

Food nutritional info estimation

- **Objective**: Predicting nutritional values (energy, carbs, ...) from pictures
- **Input**
  - Food image
- **Methods**
  - Deep learning models to automatically extract features
  - Multilayer perceptrons for nutritional values estimation
- **Specific challenges / findings**
  - **Many**:
    - Depth is difficult to grasp
    - Occlusions
    - Some ingredients are “invisible” (e.g. oil, melt butter, ...)
    - “Mixed” dishes (soups, salads)
  - Encouraging results (e.g. median abs err 40-50 kcal)
    - Improvable: e.g. depth estimator for additional features

Case studies (involving us)

Crop segmentation for plant coverage

- **Objective**: Estimate the plant coverage
  - Note this means: identifying the pixels covered by each plant, separating the plants correctly
- **Input**
  - Images taken from about 1m

Case studies (involving us)

Crop segmentation for plant coverage

Case studies (involving us)

Crop segmentation for plant coverage

Crop segmentation for plant coverage

Label	vegetation	ground
vegetation	70.69	26.97
ANGAR	0.02	
GALPA	0.58	
VERHE	0.65	
DIGSA	0.62	
CONAR	6.12	
Lentil	40.88	
SORHA	0.00	
TAROF	0.00	
MENAR	16.19	

Case studies (involving us)

Crop segmentation for plant coverage

- **Objective**: Estimate the plant coverage
- **Input**
  - Images taken from about 1m
- **Methods**
  - Large deep learning models pretrained on “generic” data (e.g. cars, tools, landscapes, …)
  - Smaller DL models, with patch-level aggregation

Case studies (involving us)

Crop segmentation for plant coverage

- **Objective**: Estimate the plant coverage
- **Input**
  - Images taken from about 1m
- **Methods**
  - Large deep learning models pretrained on “generic” data (e.g. cars, tools, landscapes, ...)
  - Smaller DL models, with patch-level aggregation
- **Specific challenges / findings**
  - **Very few** pixel-level labeled masks (40-80)
  - Yet large: 16 Megapixels (16 * $\mathbf{10^6}$ pixels... DL models often trained with images of <50 * $\mathbf{10^3}$ px)
    - → several thousands of 50k-pixels “patches”

Case studies (involving us)

Crop segmentation for plant coverage

- **Objective**: Estimate the plant coverage
- **Input**
  - Images taken from about 1m
- **Methods**
  - Large deep learning models pretrained on “generic” data (e.g. cars, tools, landscapes, ...)
  - Smaller DL models, with patch-level aggregation
- **Specific challenges / findings**
  - **Very few** pixel-level labeled masks (40-80)
  - Yet large: 16 Megapixels (16 * $\mathbf{10^6}$ pixels... DL models often trained with images of <50 * $\mathbf{10^3}$ px)
    - → several thousands of 50k-pixels “patches”
  - Preliminary results:
    - Students annotating images with AI-assistance **improve** over time
    - Counting pixels as a proxy → *very similar* to expert results
    - It is feasible to perform the task **fully automatically**

Limitations, risks, and when not to use it

- ⚠️ AI doesn’t replace domain knowledge… It **complements** it!
- Some problems may be **too small** or **don’t need** AI
- AI models can be:
  - Hard to interpret (black box)
  - Data-hungry
  - Biased if trained on biased data
  - Fragile in the face of noise or change
- Use AI critically, not blindly

AI is not a silver bullet

Become an informed AI user and thinker

Goals for the course:

- Understand what AI/ML can and cannot do
- Learn how to preprocess and explore data
- Apply basic ML workflows to real environmental data
- Know how to evaluate models properly
- Think critically about data, ethics, and impact

🧭 Whether you'll use AI directly or collaborate with data scientists
the goal is to ask better questions, build better tools, and do better science!

THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani