INTRODUCTION TO MACHINE LEARNING

Core Concepts of ***Artificial Intelligence*** (AI) and ***Machine Learning*** (ML) in Earth Observation

Programming Python Fundamentals

16 January 2025

Session Objectives

01 Understand **what AI/ML means** in Earth Observation context

02 Learn some **end-to-end workflow** for ML projects

03 Distinguish **supervised and unsupervised learning**

What is AI/ML?

Definition

Data	→	any information that can be stored and processed
Data Science	→	Set of methods, processes, heuristics, and algorithms to extract insights from data
Big Data	→	extremely large amounts of data which traditional data processing systems fail to handle
Artificial Intelligence	→	study of intelligent agents or developing intelligent systems
Machine Learning	→	allow computer systems to learn from data without explicit programming

Machine Learning...

[...] is turning things

(data in a table format, images, texts, audios, videos)

into numbers and finding patterns in those numbers.

But how do we find those patterns?

The computer does this part using code and math!

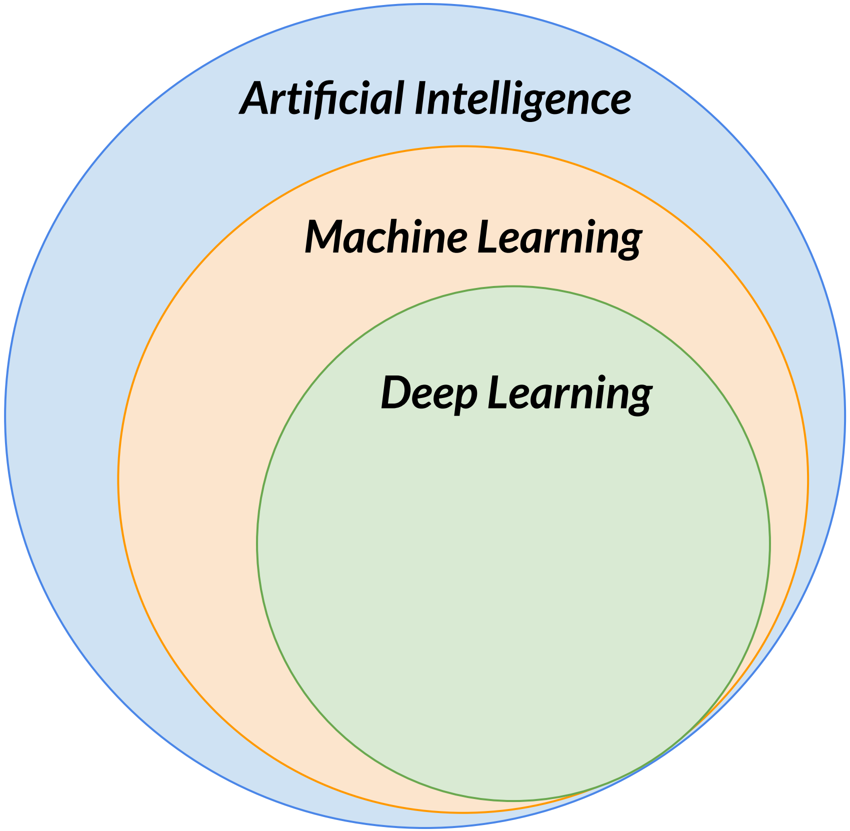
Artificial Intelligence

↓

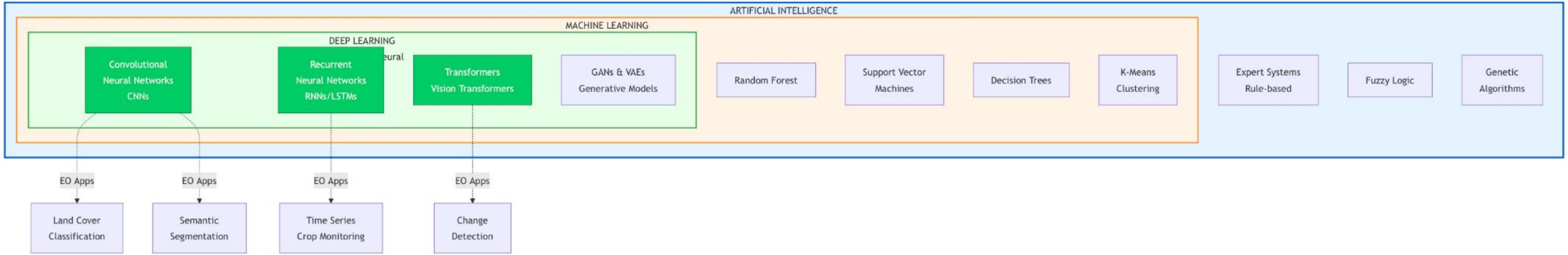
Machine Learning

↓

Deep Learning



AI, ML, and Deep Learning Hierarchy with EO Applications



Artificial Intelligence (AI): Broad field of making machines “smart”

Machine Learning (ML): Subset of AI where algorithms learn from data

Deep Learning (DL): Subset of ML using neural networks with many layers

Machine Learning in Simple Terms

Traditional Programming

- ➔ Programmer writes explicit rules
- ➔ Fixed logic
- ➔ Hard to handle complexity

Data + Rules → Results

Machine Learning

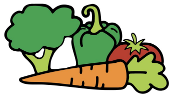
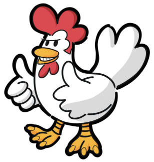
- ➔ Algorithm learns rules from examples
- ➔ Adaptive
- ➔ Handles complex patterns

Data + Results → Rules

Machine Learning in Simple Terms

Traditional Programming

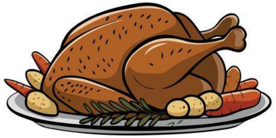
Data + Rules → Results

+

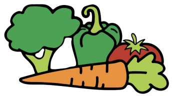
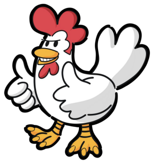
1. Cut vegetables
2. Season chicken
3. Preheat oven
4. Cook chicken for 30-min.
5. Add vegetables

→

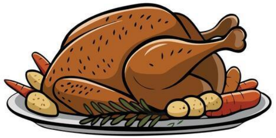


Machine Learning

Data + Results → Rules

+



→

1. Cut vegetables
2. Season chicken
3. Preheat oven
4. Cook chicken for 30-min.
5. Add vegetables

Why use Machine Learning?

Traditional Approach

- ➔ Manual interpretation
- ➔ Rule-based classification
- ➔ Simple thresholds
- ➔ Time-consuming
- ➔ Hard to scale

AI/ML Approach

- ➔ Automated pattern recognition
- ➔ Learn from examples
- ➔ Complex decision boundaries
- ➔ Fast processing
- ➔ Scalable to large areas

ML can process years of satellite data in hours!

ML in Earth Observation Context

Example: Forest VS Non-Forest

Traditional

IF NDVI > 0.6

THEN Forest

ELSE Non-Forest

- ➔ Simple but breaks easily

Machine Learning

- ➔ Show 1000 examples of forest pixels
- ➔ Show 1000 examples of non-forest
- ➔ Algorithm learns complex patterns
- ➔ Works in diverse conditions

Machine Learning is good for...

- ➔ Problems with long lists of rules
- ➔ Continually changing environments
- ➔ Discovering insights within large collections of data

Machine Learning is typically NOT good for...

- ➔ When you need explainability
- ➔ When traditional approach is a better option
- ➔ When errors are unacceptable
- ➔ When you don’t have much data

If you can build a **simple rule-based system** that doesn’t require machine learning, do that

~ Google’s #1 Rule on the Machine Learning Handbook

Machine Learning is a powerful tool...

[...] and its applications are progressing at a very fast rate.

Things might change in the future,
so it is always a good idea to keep an open mind.

The AI/ML Workflow for EO

End-to-End Machine Learning Workflow

Step 1: Problem Definition

Key Questions

- ➔ What exactly are we trying to achieve?
- ➔ What decisions will this support?
- ➔ What level of accuracy is needed?
- ➔ What resources are available?

Examples

- ➔ Map rice paddy extent
- ➔ Detect flooded areas after typhoon
- ➔ Classify land cover types
- ➔ Estimate crop yield
- ➔ Monitor deforestation

Clear problem definition = 50% of success

Step 2: Data Acquisition

Satellite Imagery

- ➔ Sentinel-1/2
- ➔ Landsat
- ➔ Planet
- ➔ High-resolution commercial
- ➔ Multiple dates/seasons

Ground Truth / Labels

- ➔ Field surveys
- ➔ GPS points
- ➔ Existing maps
- ➔ Photo interpretation
- ➔ Expert knowledge

Challenge: *Getting quality labels is often hardest part*

Step 3: Data Processing

For Satellite Imagery:

- ➔ Atmospheric correction (use Level-2A!)
- ➔ Cloud masking
- ➔ Geometric correction
- ➔ Radiometric calibration
- ➔ Co-registration (multiple sensors)
- ➔ Temporal compositing

"Garbage In, Garbage Out" *- preprocessing matters!*

Preprocessing Example:

Cloud Removal Before and After Comparison

Before Preprocessing:
- Clouds present
- Atmospheric haze
- Different acquisition dates

After Preprocessing:
- Clouds masked
- Atmospherically corrected
- Temporal composite created

Step 4: Feature Engineering

What are Features?

- ➔ Input variables for the model
- ➔ Derived from raw data
- ➔ Informative for the task

EO features

- ➔ Spectral bands (Blue, Red, NIR, etc.)
- ➔ Spectral indices (NDVI, NDWI)
- ➔ Texture measures
- ➔ Temporal statistics
- ➔ Topography (elevation, slope)

Deep Learning: *Often learns features automatically!*

Common EO Features

Feature Type	Examples	What They Capture
Spectral Bands	B2, B3, B4, B8	Reflectance at different wavelengths
Vegetation Indices	NDVI, EVI, SAVI	Vegetation health, density
Water Indices	NDWI, MNDWI	Water presence, moisture
Texture	GLCM variance, entropy	Spatial patterns
Temporal	Mean, std over time	Phenology, seasonality
Topographic	Elevation, slope, aspect	Terrain characteristics

Step 5: Model Selection & Training

Model Selection

Choose based on:

- ➔ Problem type (classification vs regression)
- ➔ Data size
- ➔ Interpretability needs
- ➔ Computational resources

Common EO Models

- ➔ Random Forest
- ➔ Support Vector Machines
- ➔ Convolutional Neural Networks
- ➔ U-Net (segmentation)
- ➔ Recurrent networks (time series)

Step 6: Validation & Evaluation

Why Validate?

- ➔ Ensure model generalizes
- ➔ Detect overfitting
- ➔ Compare different models
- ➔ Build confidence

Evaluation Metrics

- ➔ Overall Accuracy
- ➔ Confusion Matrix
- ➔ Precision & Recall
- ➔ F1-Score
- ➔ Kappa coefficient

Use independent test data - never validate on training data!

Step 7: Deployment

Deployment Options

- ➔ Generate full maps
- ➔ Near real-time monitoring
- ➔ Operational pipelines
- ➔ Decision support systems
- ➔ Web applications

Considerations

- ➔ Model retraining schedule
- ➔ Computational requirements
- ➔ User interface
- ➔ Data updates
- ➔ Maintenance plan

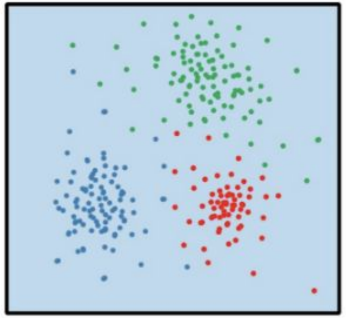
Workflow is Iterative

- **Poor validation?** → Go back to data acquisition or model selection
- **New data available?** → Retrain model
- **Requirements change?** → Redefine problem
- **Continuous improvement** is key

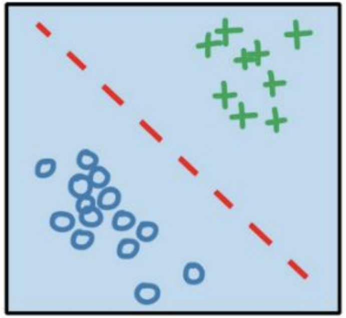
Types of Machine Learning

Main Machine Learning Paradigms

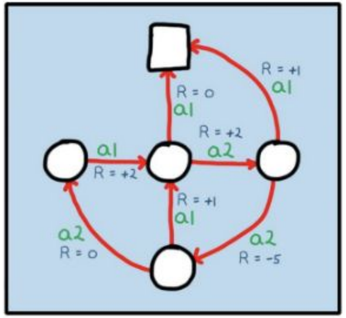
1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning



Unsupervised Learning



Supervised Learning



Reinforcement Learning

Supervised Learning

Definition

- ➔ Learning from **labeled data**
- ➔ Known input-output pairs
- ➔ Model learns mapping from inputs to outputs
- ➔ Like learning with an answer key

Requires ground truth labels for training

Types of Supervised Learning

CLASSIFICATION	REGRESSION
➔ Predict **categorical** labels ➔ Discrete classes ➔ Example outputs: “Forest”, “Water”, “Urban”	➔ Predict **continuous** values ➔ Numeric outputs ➔ Example outputs: 25.3 tons/hectare, 15.2° C

Supervised Learning Applications

CLASSIFICATION	REGRESSION
Land Cover Classification	*Biomass Estimation*
➔ Forest, agriculture, urban, water	➔ Predict tons of biomass per hectare
➔ Pixel-wise or object-based	➔ Important for carbon accounting
➔ Multi-class problem	➔ Uses SAR and optical data
Crop Type Mapping	*Crop Yield Prediction*
➔ Rice, corn, sugarcane	➔ Predict tons per hectare
➔ Seasonal patterns important	➔ Seasonal NDVI time series
➔ Supports agricultural planning	➔ Supports food security planning

Common Supervised Algorithms

Algorithm	Strengths	EO Applications
Random Forest	Handles high dimensions, robust	Land cover, crop classification
SVM	Effective in high dimensions	Binary classification, change detection
Neural Networks	Learns complex patterns	Image classification, segmentation
Decision Trees	Interpretable	Quick classifications
k-NN	Simple, non-parametric	Local classifications

Unsupervised Learning

Definition

- ➔ Learning from **unlabeled data**
- ➔ No known outputs
- ➔ Discover hidden patterns
- ➔ Like sorting without instructions

Useful for exploratory analysis and finding structure

Main Unsupervised Technique

Clustering

- ➔ **Group similar pixels** based on spectral characteristics
- ➔ Algorithm decides number of clusters (or you specify)
- ➔ **Analyst interprets** what each cluster means
- ➔ Example: “Cluster 3 looks like water, Cluster 7 looks like forest”

Unsupervised Learning Applications

Change Detection

- ➔ Cluster "before" and "after" images
- ➔ Identify changed areas
- ➔ No labels needed

Initial Exploration

- ➔ Quick overview of spectral classes
- ➔ Inform supervised approach
- ➔ Generate training samples

Anomaly Detection

- ➔ Find unusual pixels
- ➔ Potential forest disturbance
- ➔ Data quality issues

Supervised VS Unsupervised

Aspect	Supervised	Unsupervised
Labels	Required	Not needed
Accuracy	Generally higher	Lower, needs interpretation
Use Case	Precise classification	Exploration, pattern discovery
Effort	High (collecting labels)	Low (no labels)
Output	Predefined classes	Discovered clusters
Control	High (you define classes)	Low (algorithm decides groups)

Which to Choose?

Use Supervised When:

- ➔ You have ground truth labels
- ➔ Need specific classes
- ➔ Accuracy is critical
- ➔ Operational application

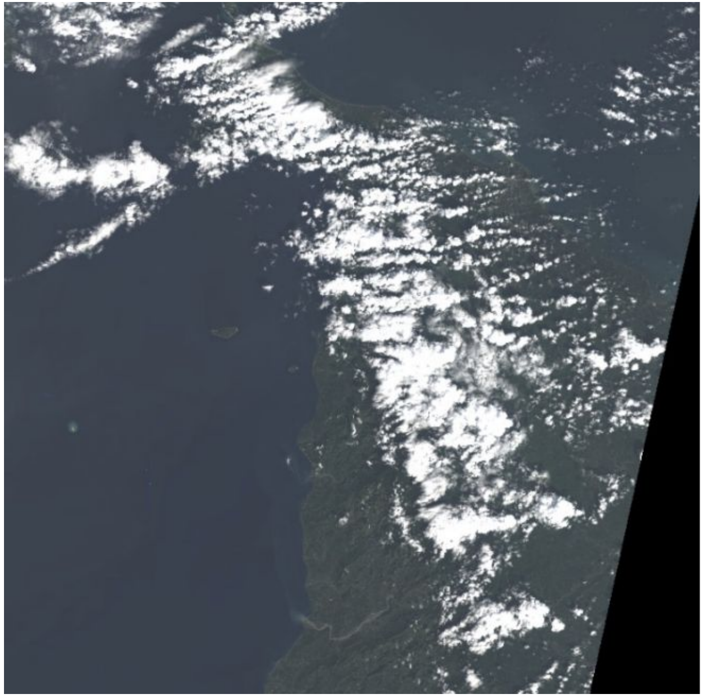
Use Unsupervised When:

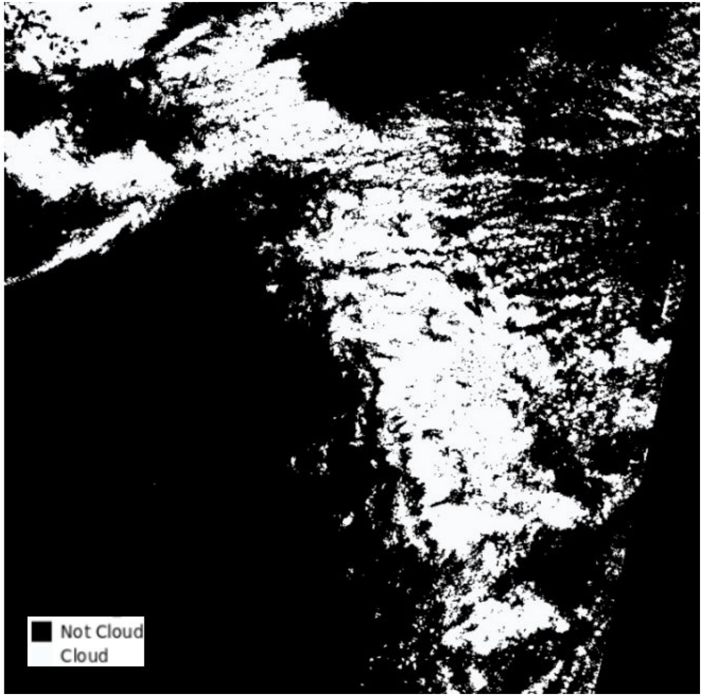
- ➔ No labels available
- ➔ Exploratory analysis
- ➔ Discovering unknown patterns
- ➔ Quick initial assessment

In practice: *Often combine both approaches!*

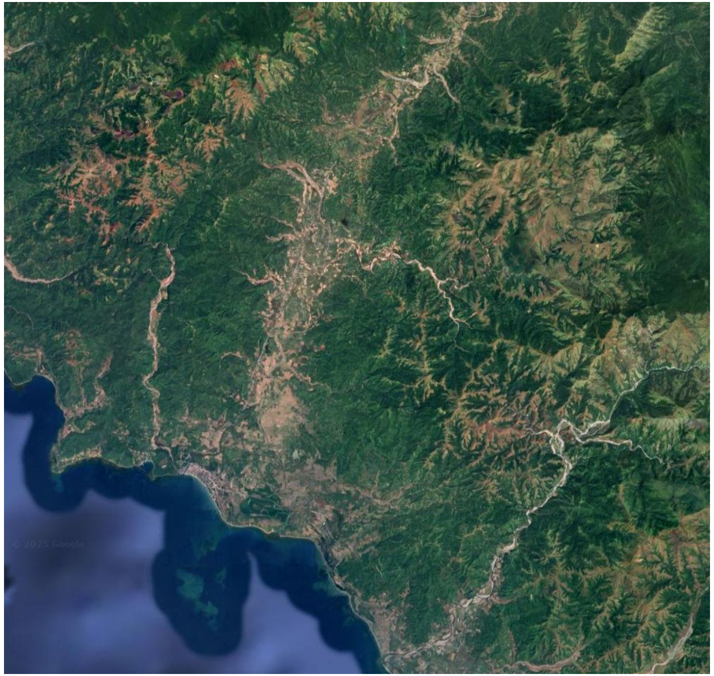
Sample Outputs

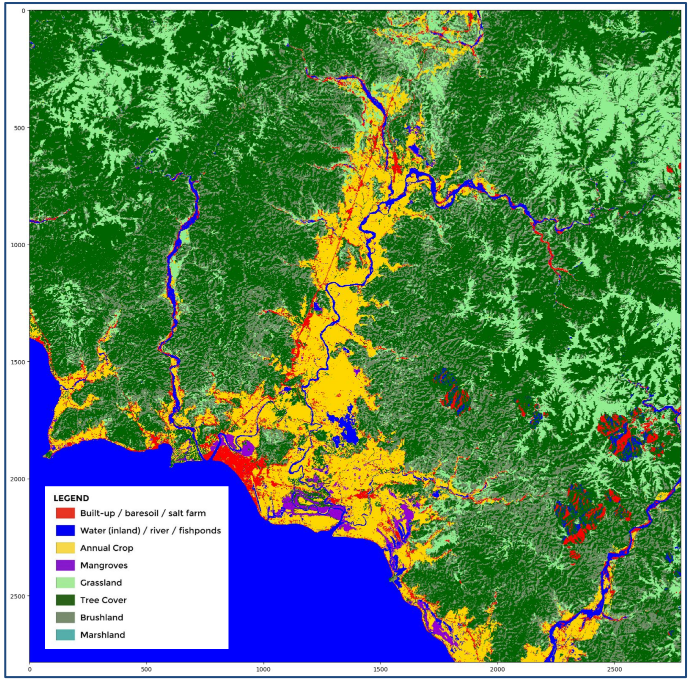
Cloud Detection





Land Cover Classification





Water Detection

THANK YOU

INTRODUCTION TO MACHINE LEARNING

*Core Concepts of **Artificial Intelligence** (AI) and **Machine Learning** (ML) in Earth Observation*

References:
CoPhil EO AI/ML Training Programme. (2025). Core Concepts of AI/ML for Earth Observation. https://skotsopoulos-del.github.io/cophil-training-v1.0/day1/sessions/session2.html
DOST Advanced Science and Technology Institute. (2023). Introduction to Artificial Intelligence [Slide deck].
DOST Philippine Institute of Volcanology and Seismology. (2025). Spatio-temporal Land Cover Mapping for Risk Assessments in Land Use Planning Using AI [Slide deck].

Programming Python Fundamentals

16 January 2025

Demonstration:

Access notebook here: