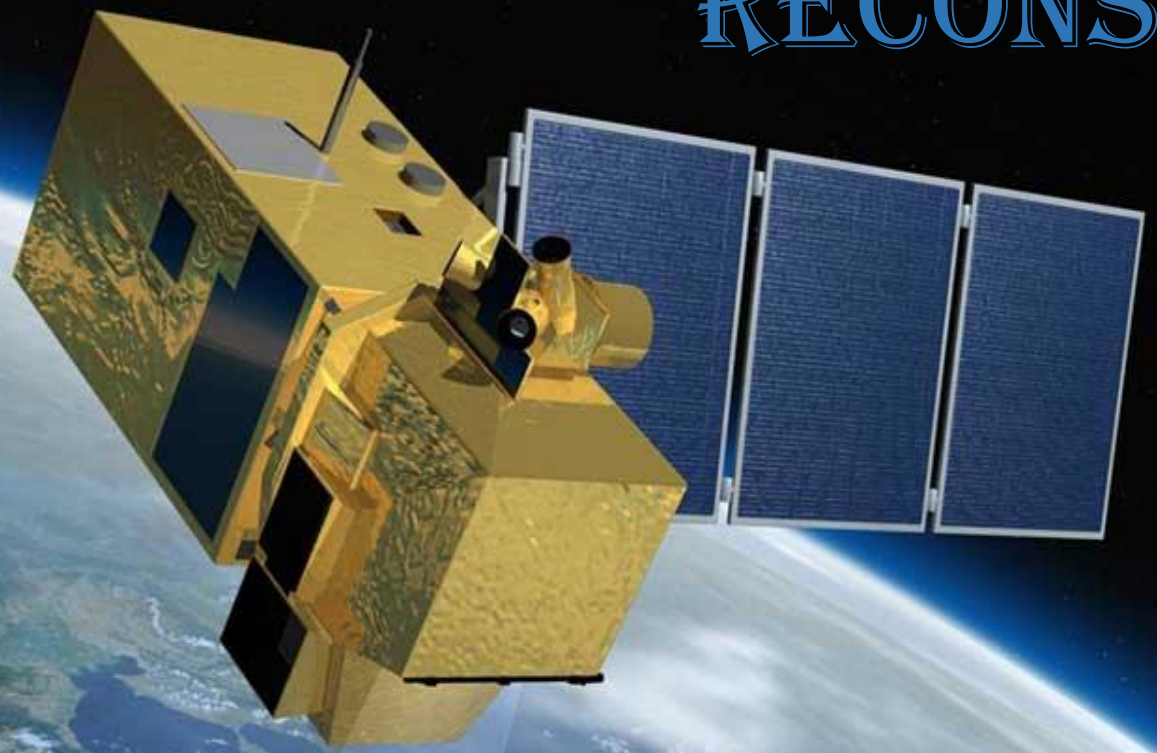
FLAGSHIP 2.6.3: AI ALGORITHM FOR (SATELLITE) IMAGING RECONSTRUCTION



REPORT FOR WP6 MEETING, 20/02/2024

A. Tricomi[1,2,3], G. Piparo[1], G. A. Anastasi[2], E. Tramontana[2], V. Strati[4]

1. INFN Sezione di Catania
2. Università degli studi di Catania
3. Centro Siciliano di Fisica Nucleare e Struttura della Materia (CSFNSM)
4. Università degli studi di Ferrara

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

OVERVIEW OF THE FLAGSHIP 2.6.3

The flagship activities are divided into three distinct but interconnected working groups:

- Identification of objects from aerial images using a deterministic learning algorithm. Currently used for identification of photovoltaic panels (**Emiliano and his group**).
- Study of vineyard diseases using of airborne images (drones and UAVs) by clustering techniques (**Virginia and her group**).
- Analysis of satellite images (Copernicus constellation) using Machine Learning techniques to identify vineyard diseases and/or wildfire segmentation/predictions (**Alessia, Giuseppe and Gialex**)

IDENTIFICATION OF OBJECTS FROM AERIAL IMAGES WITH DETERMINISTIC ALGORITHM

- The activities of Emiliano's group are focused on the development of a **deterministic algorithm** for object identification and segmentation of aerial images.

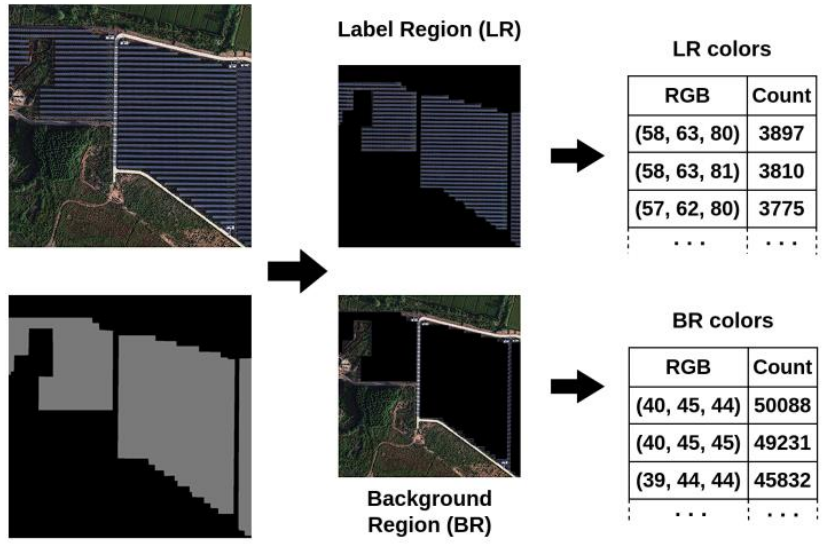
- Currently, the algorithm has been used to identify photovoltaic panels (PV) based on their characteristic colors. **The work has been published**: https://doi.org/10.3390/technologies11060174.

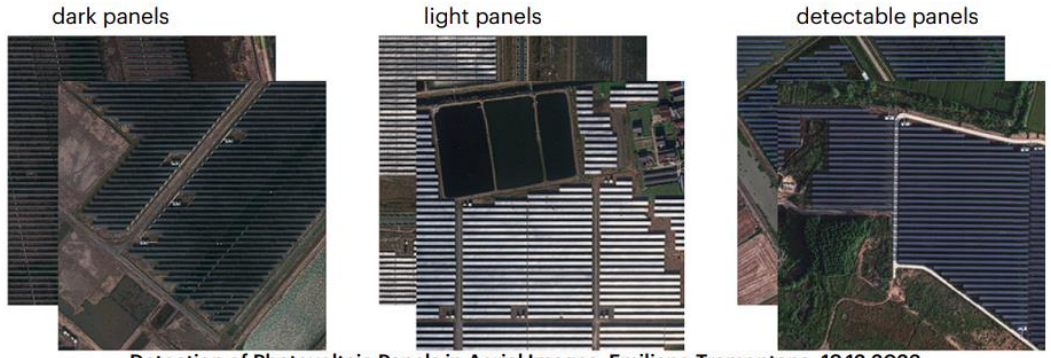
- The importance of this work lies in the possibility of identifying PV and estimating the **amount of green energy production** expected in a community/region.

- The use of a deterministic algorithm as opposed to ML methods gives as benefits **explainability, speed of execution and no need for large datasets**.

CLASSIFICATION OF PV IMAGES AND COLOR EXTRACTION

- The first phase of the pipeline is the **classification of PV images for colour extraction**.
- Colors found in both labeled regions and the background are deemed PV panel colors if more prevalent in labeled regions (LR).
- The subset of LR colours that are also frequent in the background (BR) were removed.
- Images with panels having **too many dark pixels** (lightness < 0.2) and a **high quantity of saturated pixels** (saturation > 0.33) are removed.



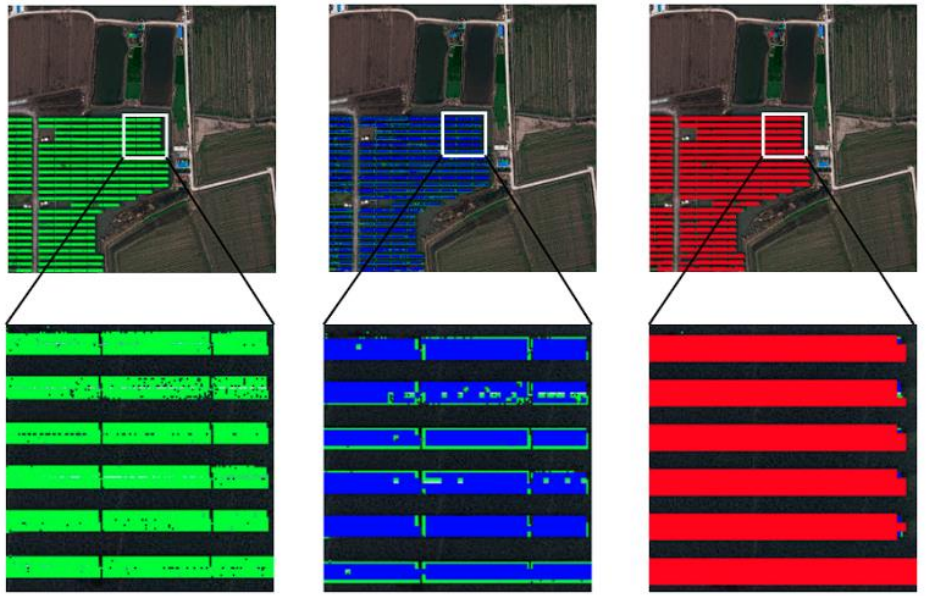


Detection of Photovoltaic Panels in Aerial Images, Emiliano Tramontana, 19.12.2023

IDENTIFICATION OF PV

The outcome is a collection of RGB color values characteristic of PV panels, which are subsequently **searched for in an unknown image** (green pixels in the image below)

Detection of Photovoltaic Panels in Aerial Images, Emiliano Tramontana, 19.12.2023



For the refinement of the results A 5x5 tile (blue) shifts across the image, turning entirely blue **if over 75% of its pixels are green**; the same applies to an 11x11 tile (red)

RESULTS

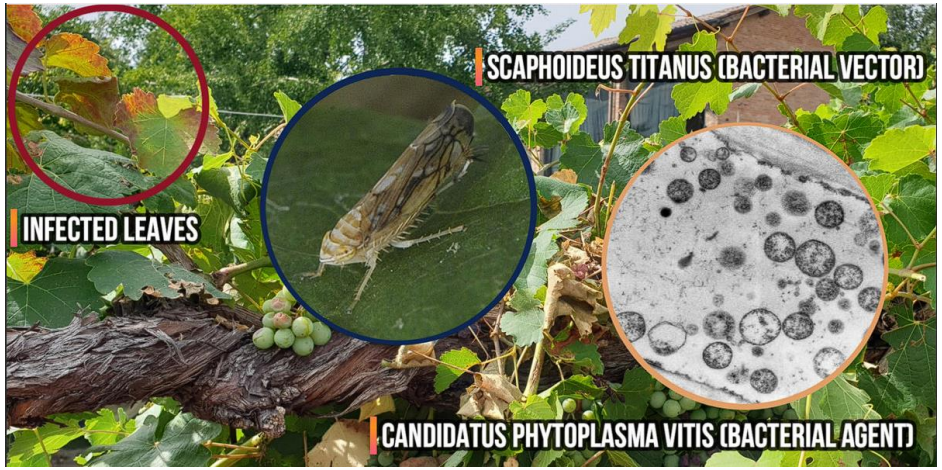
Dataset	Size	Accuracy	Precision	Recall	F1 Score	IoU
Ground	458	0.963	0.918	0.901	0.897	0.833
Cropland	146	0.971	0.931	0.938	0.931	0.877
Grassland	42	0.952	**0.953**	0.881	0.912	0.844
SalineAlkali	53	**0.982**	0.913	**0.958**	0.934	0.878
Shrubwood	77	0.975	0.934	0.956	**0.944**	**0.897**
WaterSurface	140	0.974	0.915	0.928	0.917	0.852
max in [30]		0.981	0.960	0.903	0.931	0.877

Phase	Average Execution Time per Image
Image Classification	2.42 s
cPV Colours Extraction	1.63 s
PVs Detection (3 passes)	9.25 s
Image Denoising	3.16 s

Detection of Photovoltaic Panels in Aerial Images, Emiliano Tramontana, 19.12.2023

STUDY OF VINEYARD DISEASES FROM AIRBORNE IMAGES

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing



- Flavescence dorée is a vineyard disease caused by a phytoplasma that **is devastating crops across Europe**.
- This study aims to quantify the incidence of the disease using airborne images to identify, **through AI techniques**, one of the main symptoms of the disease, namely **reddening of the leaves**.

THE RADGYRO

- Engine: 1.2 liter turbo – 90 kW
- Payload: 150 kg
- Fuel tank: 90 liter
- Fuel: regular gasoline
- Space for take off: < 70 m
- Range of flight: < 3.5 hours
- Range of investigation: ~ 50 km² / h
- Easy to move without disassemble

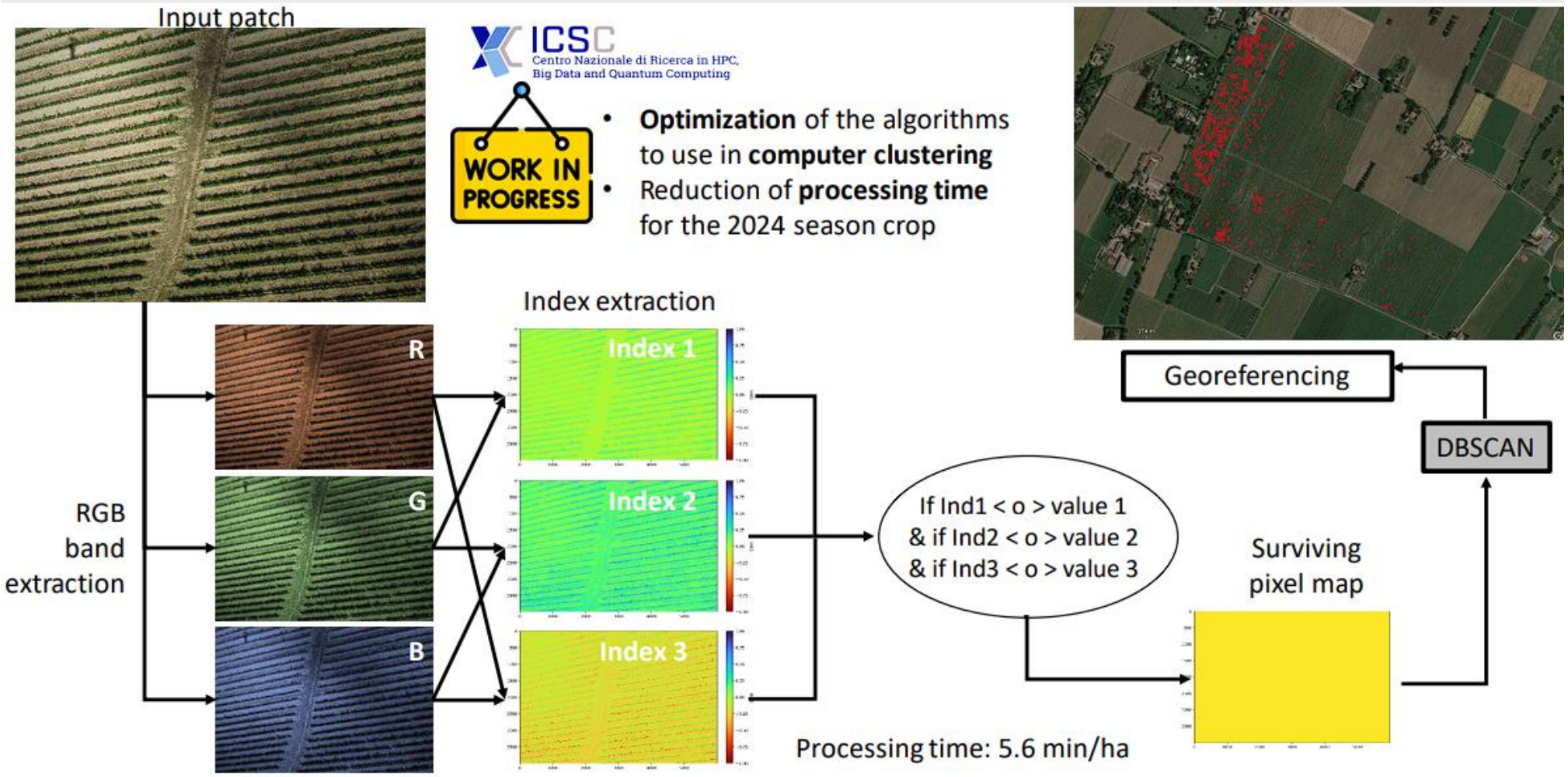
Sensors in automatic and synchronized acquisition

- RGB phorogrammteric cameras
- Multispectral camera (infrared)
- Thermal camera
- Radioactivity detectors
- Radar altimeter
- GPS + electronics
- Remote control

Three simple ideas

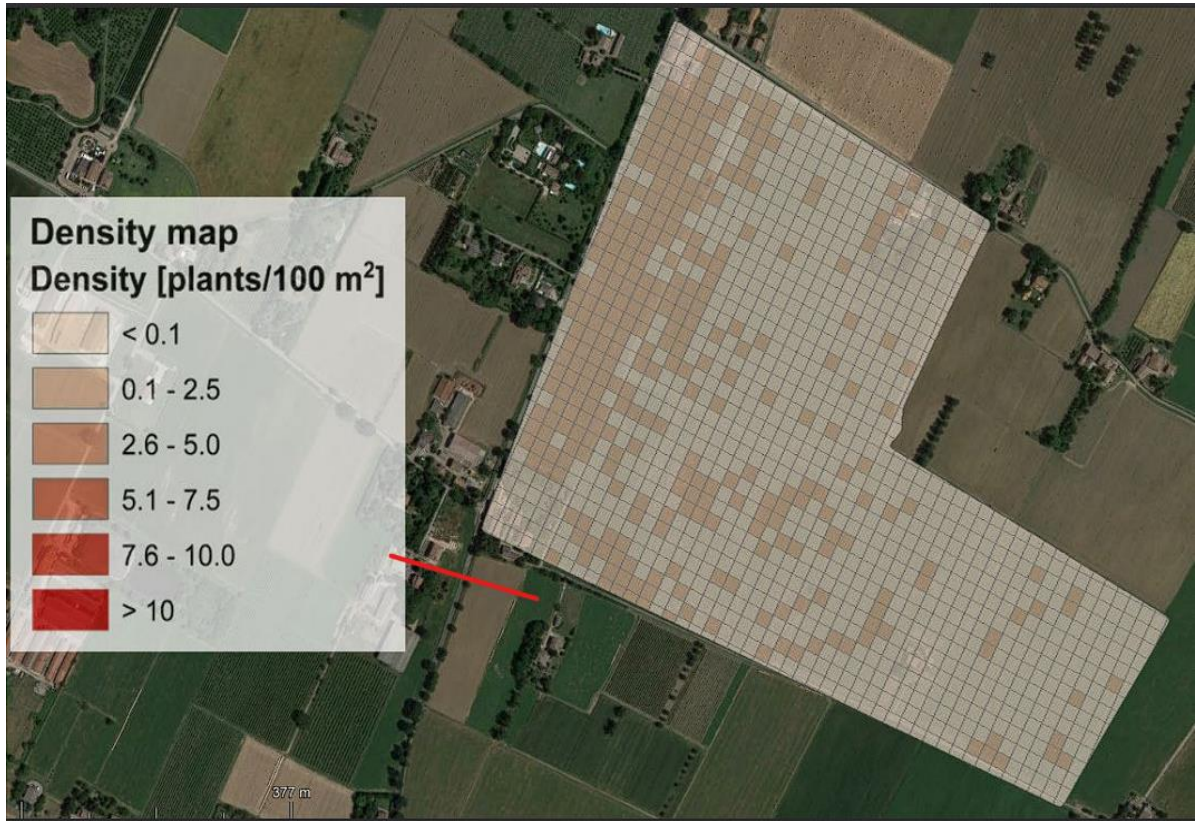
- Agility
- Compactness
- Affordability

ANALYSIS PIPELINE

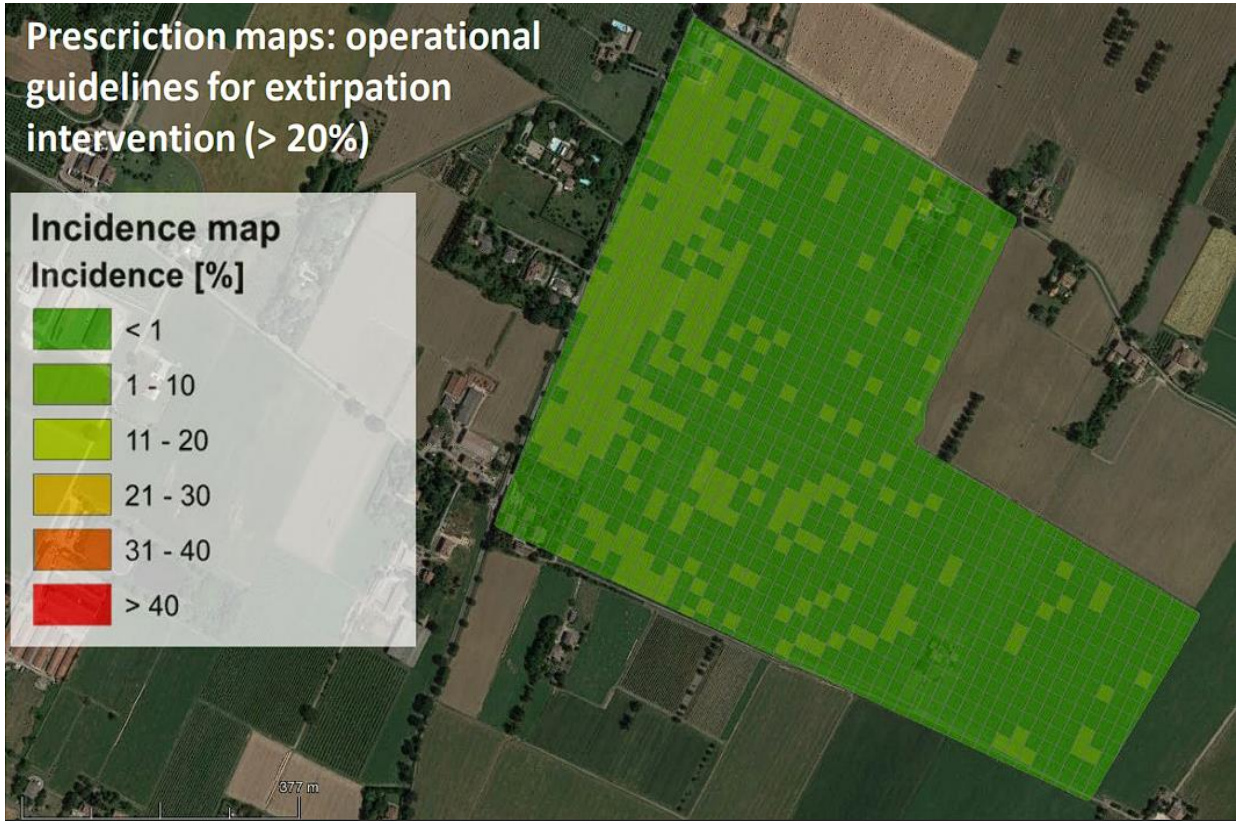


RESULTS

DENSITY MAP

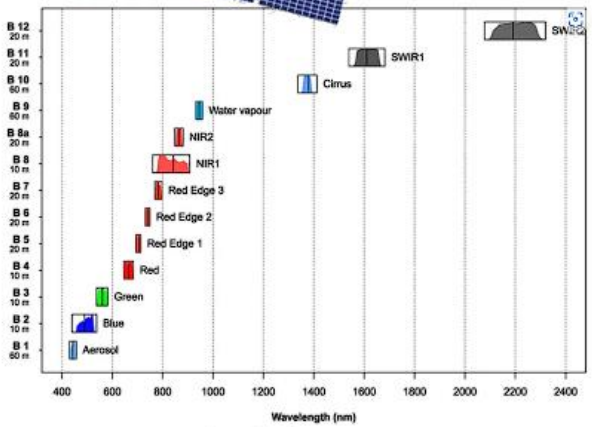


INCIDENCE MAP



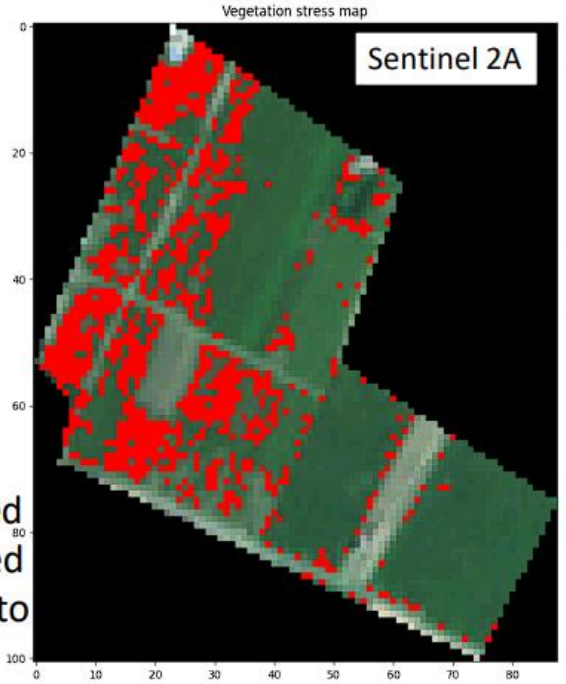
STUDY OF VINEYARD DISEASES FROM AIRBORNE+SATELLITE IMAGES

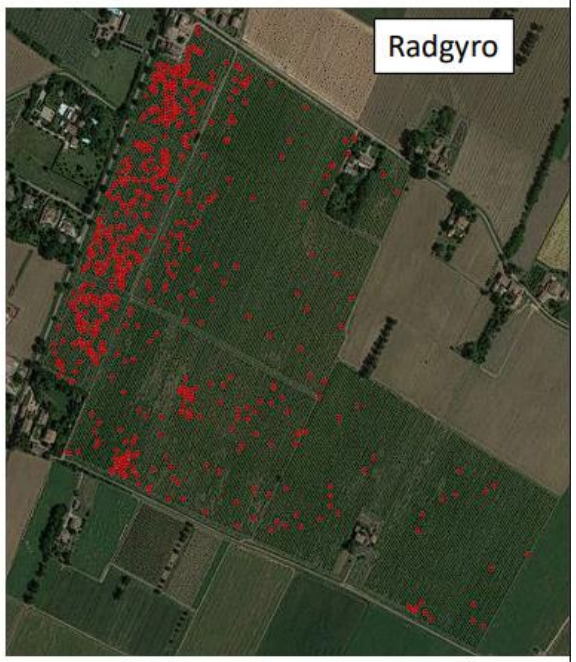
What next: use of satellite imagery



- **Sentinel-2A** is a satellite of the Copernicus constellation specifically designed for **vegetation monitoring** and **natural disaster management.**
- It acquires images in various spectral bands (spatial resolution: 10 m), allowing the calculation of many **vegetation indices.**

- First reproducibility tests were carried out using **clustering techniques** based on indices of vegetation susceptible to **flavescence dorée symptoms.**





ANALYSIS OF SATELLITE IMAGES USING MACHINE LEARNING TECHNIQUES

- In some of the discussions at the Spoke 2 meeting, **the difficulty of managing the downloading and manipulation of satellite data** for the realization of datasets using the available frameworks emerged.
- **We decided to develop a library that would serve as a unified framework for the preparation of datasets used for training AI techniques, and possibly integrate these techniques within it.**
- Development is already underway and the first version of the code is already available on github (currently in private form): https://github.com/gpiparo2/AgriSentinel.
- **In addition to being useful internally, the development of this library could be a valuable support for the completion of certain KPIs.**


```
"MCARI": {
    "calculate": true,
    "description": "Modified Chlorophyll Absorption in Reflectance Index",
    "formula": "((RedEdge1-RED)-0.2*(RedEdge1-GREEN))*(RedEdge1/RED)"
},
"MCARI1": {
    "calculate": false,
    "description": "Modified Chlorophyll Absorption in Reflectance Index 1",
    "formula": "1.2*(2.5*(NIR1-RED)-1.3*(NIR1-GREEN))"
},
"MCARI2": {
    "calculate": false,
    "description": "Modified Chlorophyll Absorption in Reflectance Index 2",
    "formula": "1.5*(2.5*(NIR1-RED)-1.3*(NIR1-GREEN))/sqrt(((2.0*NIR1+1)**2)-(6*NIR1-5*sqrt(RED))-0.5)"
},
"NDRE": {
    "calculate": true,
    "description": "Normalized Difference Red Edge Index",
    "formula": "(NIR1-RedEdge1)/(NIR1+RedEdge1)"
},
"NDVI": {
    "calculate": true,
    "description": "Normalized Difference Vegetation Index",
    "formula": "(NIR1-RED)/(NIR1+RED)"
},
```

```
# Credentials configuration
config=SHConfig("peppe")

# AgriSentinel object creation
agri_sentinel = AgriSentinel(config=config, data_folder="prova/data", processed_data_folder="prova/data_off", images_folder="prova/images")

# Definition of the area of interest via a GeoJSON object
geojson_path='geojson_archive/vineyard_area.geojson'
with open(geojson_path) as f:
    geojson = json.load(f)

# Name of the area
area_name = "FormigineVineyard"

#Downloading Images
agri_sentinel.download_images(date_from=datetime(2023, 8, 1), date_to=datetime(2023, 8, 31), interval_days=5, location_geojson=geojson, mosaicking_order="leastCC", resolution=10)
```

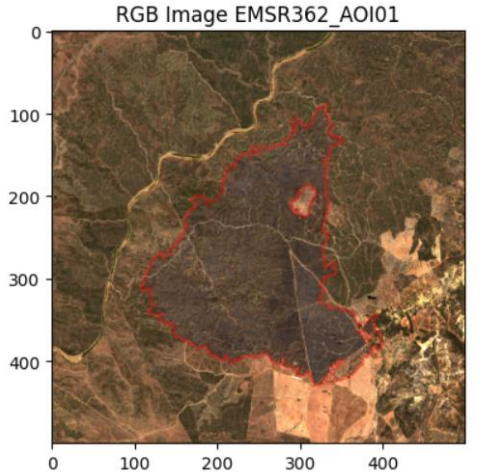
```
# Manipulation of downloaded data
agri_sentinel.manipulate_data(area_name=area_name, config_file="configuration/indices_config.json")
```

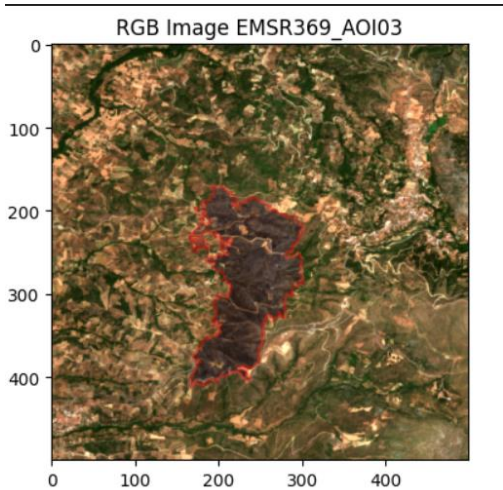
```
# Visualisation of Results
agri_sentinel.visualize_data(area_name, bands='RGB', crop_mask_path=geojson_path, crop=True)
```

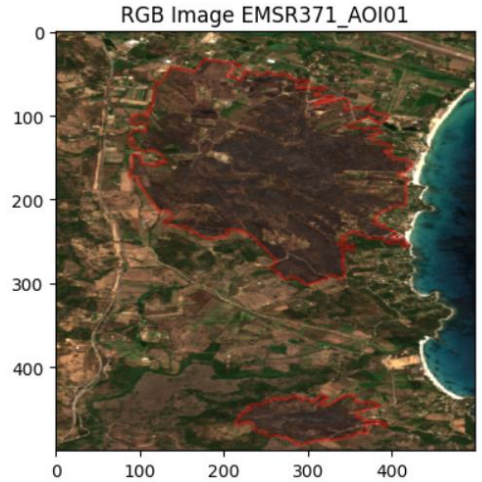
```
# Preparation of MS and VI arrays
ms_array=agri_sentinel.prepare_spectral_bands_dataset_array(area_name, crop_mask_path=geojson_path, crop=True)
vi_array=agri_sentinel.prepare_vegetation_indices_dataset_array(area_name, crop_mask_path=geojson_path, crop=True)
```

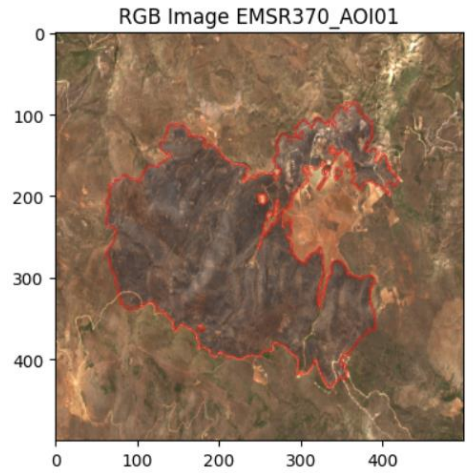
WILDFIRE DATASET FOR BURNT AREA DELIMITATION WITH CNN

- Using the AgriSentinel library and historical fire information taken from the COPERNICUS Emergency Management Service (List of EMS Rapid Mapping Activations | COPERNICUS EMERGENCY MANAGEMENT SERVICE), **we created a test dataset for delimiting burnt areas.**
- The dataset consists (currently) of 85 500x500 pixel images, and was used to train and test a simple Convolutional Neural Network, **which gave fairly good results**.
- Some spectral bands and vegetation indices (NDVI, NBR, NDWI) were used as input.
- **There is a lot of room for improvement by using more images, more features, more complex and deeper models**. The preparation of the dataset for training and testing also plays an important role, as does cloud management.
- **The most interesting objective, apart from the delimitation of burnt areas and the prediction of fires by means of risk maps. is much more complex and it may be necessary to add data from other sources and to exploit time sequences.**





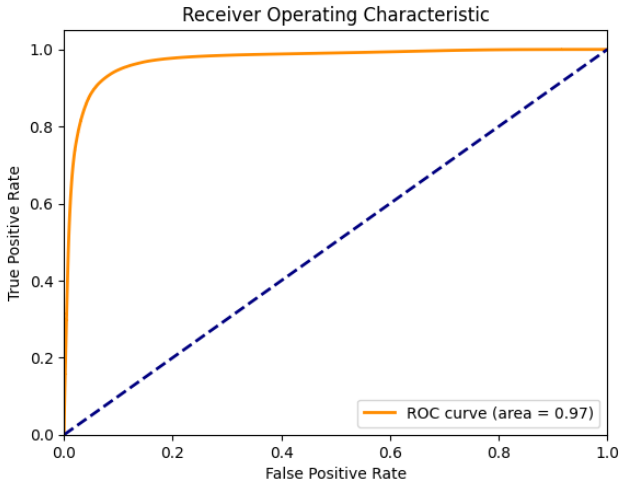


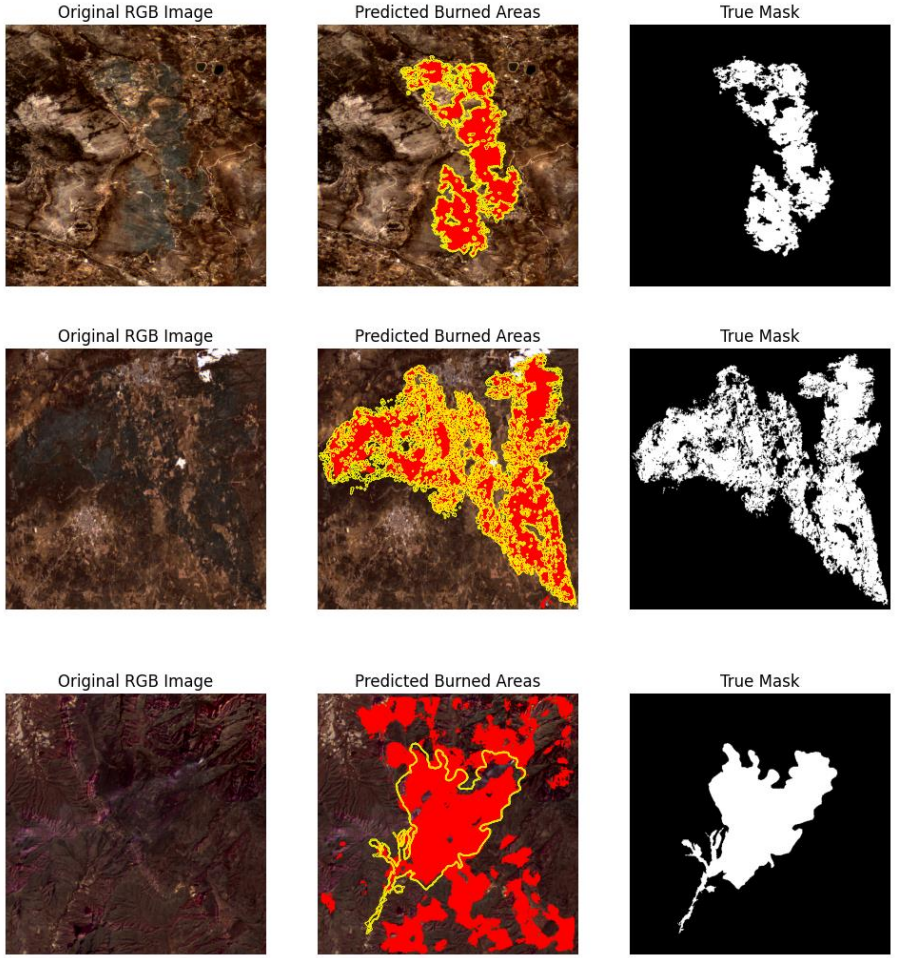


PRELIMINARY RESULTS

CNN USED FOR THIS WORK

```python
def create_model(input_shape):
    inputs = Input(shape=input_shape)

    # Downsample
    c1 = Conv2D(16, (3, 3), activation='relu', padding='same')(inputs)
    c1 = Dropout(0.3)(c1)
    p1 = MaxPooling2D((2, 2))(c1)

    c2 = Conv2D(32, (3, 3), activation='relu', padding='same')(p1)
    c2 = Dropout(0.3)(c2)
    p2 = MaxPooling2D((2, 2))(c2)

    # Bottleneck
    bn = Conv2D(64, (3, 3), activation='relu', padding='same')(p2)

    # Upsample
    u1 = UpSampling2D((2, 2))(bn)
    concat1 = Concatenate()([u1, c2])
    c3 = Conv2D(32, (3, 3), activation='relu', padding='same')(concat1)
    c3 = Dropout(0.3)(c3)

    u2 = UpSampling2D((2, 2))(c3)
    concat2 = Concatenate()([u2, c1])
    c4 = Conv2D(16, (3, 3), activation='relu', padding='same')(concat2)
    c4 = Dropout(0.3)(c4)

    outputs = Conv2D(1, (1, 1), activation='sigmoid', padding='same')(c4)

    model = Model(inputs=[inputs], outputs=[outputs])
    model.compile(optimizer=Adam(learning_rate=0.001), loss='binary_crossentropy', metrics=['accuracy'])

    return model


model = create_model(input_shape=(500, 500, 13))
history = model.fit(X_train, y_train, epochs=50, batch_size = 5, validation_data=(X_test, y_test))
```





MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS7 (TO BE COMPLETED NEXT MONTH)

1. M1-M6 (corresponding to MS7): Survey of the State-of-the-Art; tracking of R&D technologies to be used; selection of datasets for use cases (at least one).
   - ➢ D1: report on technologies to be used, selection of at least one test dataset.

MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS7 (TO BE COMPLETED NEXT MONTH)

1. M1-M6 (corresponding to MS7): Survey of the State-of-the-Art; tracking of R&D technologies to be used; selection of datasets for use cases (at least one).
   - ➢ D1: report on technologies to be used, selection of at least one test dataset.

The report will focus on the different technologies used by the three working groups, namely:

- **Deterministic identification algorithm .**
- **Clustering methods (k-Means and DBSCAN).**
- **Convolutional NN and eventually LSTM.**

MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS7 (TO BE COMPLETED NEXT MONTH)

1. M1-M6 (corresponding to MS7): Survey of the State-of-the-Art; tracking of R&D technologies to be used; selection of datasets for use cases (at least one).
   - ➢ D1: report on technologies to be used, selection of at least one test dataset.

The report will focus on the different technologies used by the three working groups, namely:

- **Deterministic identification algorithm .**
- **Clustering methods (k-Means and DBSCAN).**
- **Convolutional NN and eventually LSTM.**

Satellite datasets based on Copernicus satellite constellation **can be indicated as use case test datasets** (it can be also be distributed for the next KPIs). Also, PV and airborne can be indicated as well, **but I think they cannot be distributed (?)**

MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS8 (TO BE COMPLETED BY JUNE 2024)

- TAR6.11 [MS report for UC2.6.3, including a first experimentation with data sources and algorithms, demonstration on the feasibility of choices]. **KPI**: [Intermediate report provided; first code repository available].

MS9 (TO BE COMPLETED BY OCTBER 2024)

- TAR6.14 [MS report for UC2.6.3]. **KPI**: [Intermediate report provided].

MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS8 (TO BE COMPLETED BY JUNE 2024)

- TAR6.11 [MS report for UC2.6.3, including a first experimentation with data sources and algorithms, demonstration on the feasibility of choices]. **KPI**: [Intermediate report provided; first code repository available].

The results obtained by Emiliano and Virginia groups **can already be indicated as feasible** (publications have been made or are being made). We hope this will also be the case for the Sentinel data trials. **Surely the first experimentations will have already been carried out.**

MS9 (TO BE COMPLETED BY OCTBER 2024)

- TAR6.14 [MS report for UC2.6.3]. **KPI**: [Intermediate report provided].

MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS8 (TO BE COMPLETED BY JUNE 2024)

- TAR6.11 [MS report for UC2.6.3, including a first experimentation with data sources and algorithms, demonstration on the feasibility of choices]. **KPI**: [Intermediate report provided; first code repository available].

Sentinel's library for downloading, manipulating and visualising satellite images **is already uploaded to a private repository (**gpiparo2/AgriSentinel (github.com)**)**. For the MS8 report, it will be improved and updated with ML analysis methods for the images. The repository can be made public at the end of the project for final KPIs.

The results obtained by Emiliano and Virginia groups **can already be indicated as feasible** (publications have been made or are being made). We hope this will also be the case for the Sentinel data trials. **Surely the first experimentations will have already been carried out.**

MS9 (TO BE COMPLETED BY OCTBER 2024)

- TAR6.14 [MS report for UC2.6.3]. **KPI**: [Intermediate report provided].

MILESTONES AND KPI OF FLAGSHIP 2.6.3

MS10 (TO BE COMPLETED BY AUGUST 2025)

- TAR6.10 [MS report for UC2.6.3: Implementation of the selected technology(ies); test and validation on selected dataset(s). Proof-of-Concept deployment]. **KPI**: [Final report provided; release of the developed code on public repository];

KPIs

KPI ID	Description	Acceptance threshold
KPI2.6.1.1	Publications	2
KPI2.6.1.2	Presentations at conferences	2
KPI2.6.1.3	Publicly available Code repositories	1
KPI2.6.1.4	Use case Test Datasets	1

Current status

1* + 1 to be prepared (Virginia's Group)

1 to be held in April 2024 (Virginia's abstract accepted**)

1 to be made public before the end of the project (AgriSentinel Library+ML methods)

1 to be prepared and made public (?) before the end of the project (Sentinel datasets)

*D. Marletta, A. Midolo, E. Tramontana, Detecting Photovoltaic Panels in Aerial Images by Means of Characterising Colours, Technologies 2023, 11(6), 174; https://doi.org/10.3390/technologies11060174

**Strati et al., Airborne surveys for the detection of Flavescence Doreé in vineyards
EGU General assembly 2024 - 14-19 aprile - Vienna
Session: BG9.1 – Remote Sensing applications for the biosphere, https://www.egu24.eu/

COMPUTATIONAL RESOURCES REQUEST FOR UC2.6.3

For the future activities of the flagship we have requested the following computational resources:

- CPU time on Galileo-100 (200k/300k core hours), to be used mainly by the Virginia group for future intensive analysis of airborne data.
- GPU time on Leonardo-Booster (10k/15k core hours), to be used both from Emiliano group and for the final tests of ML methods developed for the analysis of satellite data.
- Temporary disk space (500 Gb) and Analysis disk space (15 Tb) for all three working groups.
- A virtual machine with CPU (64 cores, 2 Gb of ram each) and GPU (4 GPUs) for the development of ML techniques for the analysis of satellite data. Temporary disk space (500 Gb) and Permanent disk space (1.5 Tb).