FOLOU
D3.1 – Technical description of all the FOLOU Innovative solutions developed

Author(s):	Wouter Maes, EN-NEJJARY Driss
Version:	Version 3
Quality review:	Wouter Maes, UGent
Date:	27/06/2024
Dissemination level:	Public (PU)
Grant Agreement N°:	101084106
Starting Date:	01-01-2023
Duration:	18 months
Coordinator:	UVIC
Contact details:	wouter.maes@ugent.be driss.ennejjary@ugent.be

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table of Contents

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

List of Figures

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

List of Tables

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Abbreviations

API Application Programming Interface

ARIMA Autoregressive Integrated Moving Average

CPI Consumer Price Index

CV Coefficient of Variation

DGAC General Directorate for Civil Aviation

EFSA European Food Safety Authority

FAO Food and Agriculture Organization of the United Nations

FL Food Loss

GANs Generative Adversarial Networks

INE Spanish National Statistics Institute

MAPA Spanish Ministry of Agriculture, Fisheries and Food

MLP MultiLayer Perceptron

NDVI Normalized Difference Vegetation Index

NIR Near Infra-Red

ONNX Open Neural Network Exchange

R-CNN Region-based Convolutional Neural Networks

RF Random Forest

RGB Red Green Blue

SQL Structured Query Language

TTN The Things Network

UAV Unmanned Aerial Vehicle

UNILET French National Association for Canned and Frozen Vegetables

VAEs Variational Autoencoders

VIS Visible

WP Work Package

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Glossary

Anomaly Detection Neural Network This type of network identifies deviations from normal behaviour within datasets.

APSIM (Agricultural Production Systems sIMulator) A highly advanced simulator of agricultural systems which is used to model and simulate the biophysical processes in farming systems.

Blockchain A distributed ledger in which records are linked together in a secure way through cryptography.

Bounding Box A rectangular box used in machine vision and image processing that serves as a reference system for locating objects and defining their boundaries within an image.

Coefficient of Variation A statistical measure of the dispersion of data points in a data series around the mean.

Crop Modelling The use of computer-based models to simulate the growth and development of crops under varying environmental and management conditions.

Data Annotation The process of labelling or tagging data with information that makes it usable for machine learning.

Data Augmentation A technique used in machine learning to increase the diversity of data available for training models without actually collecting new data.

De-bayering Also known as demosaicking, is a process used to convert raw image data from a Bayer filter mosaic into a full-colour image

Detection Neural Network This type of network refers to an object detection model which is designed to identify and localize objects within an image or video.

DSSAT (Decision Support System for Agrotechnology Transfer) A software application program that comprises crop simulation models for over 42 crops and is used to facilitate decision-making.

Fertigation System A method of applying fertilizers soil amendments and other water-soluble products through an irrigation system.

Food Loss food loss definition based on FOLOU**.**

Generative Models Models In machine learning that generate new data instances resembling training data.

Image Annotation The process of manually or automatically adding metadata to an image.

Merkle Tree A data structure used in computer science and cryptography to efficiently summarize and verify the integrity of large sets of data.

MultiLayer Perceptron A Multilayer Perceptron (MLP) is a neural network with an input layer, one or more hidden layers, and an output layer, used for tasks like classification and regression, trained using backpropagation.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Multispectral Imagery An imaging technique that captures image data within specific wavelength ranges across the electromagnetic spectrum.

Normalization The adjustments made to data to enable its comparison in different scales or formats.

Object Detection A technology related to computer vision that deals with detecting instances of semantic objects.

Phenological Surveys Studies that involve observing the stages of plant development over time.

Polygon Annotation A method of image annotation that involves drawing polygons around objects within an image to create precise labels for machine learning and computer vision models.

Reflectance The measure of the proportion of light or other radiation that is reflected off a surface.

Supervised Learning A machine learning paradigm where the algorithm learns from labelled data.

Thermal Imagery Imaging that detects and records the infrared radiation (heat) emanating from objects.

YOLO (You Only Look Once) A state-of-the-art real-time object detection system that applies a single neural network to the full image which makes predictions with a single evaluation.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

This document presents a comprehensive technical description of the innovative technological solutions based mainly on remote sensing, AI and crop modelling, proposed in the WP3 of FOLOU project to quantify the food loss in primary stages across various sectors such as vegetables, fruits, and horticultural crops. It is structured to ensure uniformity and coherence, facilitating ease of comprehension and follow-up, as all solutions adhere to the same layout besides, it provides an overview of the progress for each challenge, detailing the stages of development, from initial data collection to prototype testing and model development.

The primary objective of WP3 is to develop and validate cost-effective and efficient technological tools that can replace traditional methods of quantifying food loss, which are often time-consuming and expensive. The tools developed are intended to enhance the accuracy and efficiency of food loss and production loss measurement on a large scale.

The document is organized into distinct challenges each targeting specific types of food losses and commodities:

1. Tractor-embedded video cameras for assessing food loss in vegetable crops (Dilepix-T3.1)
2. UAV-based high-resolution RGB for food loss assessment in apple (UGent-T3.2)
3. Repeated multispectral and satellite data for production loss assessment in potato (UGent-3.2 & T3.3)
4. Repeated multispectral and satellite data for production loss assessment in maize, corn, faba bean and sunflower (UNIVPM-T3.2 & T3.3)
5. Automated fish egg sorting using multispectral camera technology in trout aquaculture (UNIVPM-T3.4)
6. Blockchain technology implementation in mussel aquaculture (UNIVPM-T3.5)
7. Market demand tools from social networks (CIRCE-T3.6)

The project aims to produce robust, scalable, and easy-to-use tools for assessing food loss: Estimation of food and production losses, public blockchain to track food losses over the supply chain, and food prediction demand tool based on social networks. Hereunder an overview of the challenges :

Task ID	Organization	Commodity	Food Loss Category	Technology	Desired Outcome
T3.1	Dilepix	Vegetables (Cauliflower)	Pre-harvest, Harvest losses	Deep learning with tractor-based RGB cameras	Affordable tool to automatically measure yield and food losses in field
T3.2	UGent	Fruit (Jonagold apples)	Pre-harvest, Harvest losses	Deep learning with UAV-based RGB data	Methodology to estimate losses using UAVs in fruit orchards
T3.2 & T3.3	UGent	Vegetables (Potato)	Production losses	Crop growth model with multispectral and satellite data	Automated method to estimate potato

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

					production losses
T3.2 & T3.3	UNIVPM	Maize, Corn, Faba Bean, Sunflower	Production losses	Repeated multispectral imagery (UAV or satellite-based)	Automated method to estimate production losses using imagery
T3.4	UNIVPM	Aquaculture (Trout)	Production losses	ML with multispectral cameras for infected egg detection	Automated system for sorting trout fish eggs
T3.5	UNIVPM	Mussels	Production, Pre-harvest, Harvest losses	Blockchain for tracking losses	Blockchain-based platform to track and secure data on food losses
T3.6	CIRCE	Diverse (Grains, Fruits, Root Tubers, Meat, Fish)	Surplus	NLP and ML for social network data	Models to predict food consumption from social network messages

Table 1: Summary of Innovative Technologies Applied to quantify Food Losses in Primary Agricultural Production Under WP3.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Introduction

The purpose of WP3 is to develop technological tools to measure and estimate food losses at the primary production stage across various agricultural sectors, including vegetables, fruits, and horticultural crops leveraging the revolutionary advancements in remote proxy sensing, as well as the cutting-edge capabilities of AI. The final goal is to develop methods that could partially replace the time-consuming and expensive manual sampling needed today to quantify food losses in the primary sector.

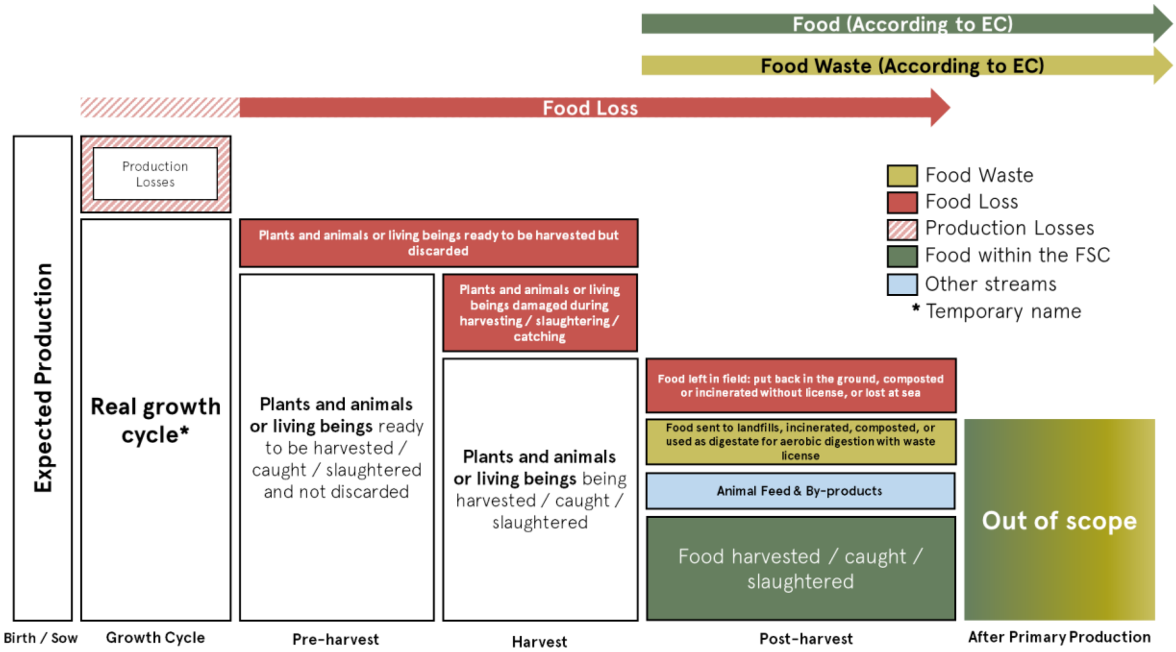

Figure 1: Different categories of food losses in the primary sector distinguished and measured within the FOLOU project

Within FOLOU, a definitional framework of food loss has been developed. Food losses are here categorized in three different categories (Figure 1):

- Pre-harvest losses: plants, animals or other living beings ready to be harvested/caught/slaughtered but discarded;
- Harvest losses: plants, animals or other living beings that are damaged during harvesting/ catching and are therefore discarded;
- Post-harvest losses: plants, animals or other living beings that are harvested, but then are not used for food or treated as a waste stream

Production losses can be defined as the difference between the attained yield – the yield when the crop/animal/other living being is ready to be harvested – and the attainable yield – the yield that could be expected from this crop/animal/other living being in the given conditions (Figure 2). In order to be consistent with EU regulations/approaches, production losses are not included int he definitional framework of FOLOU. However, production losses have a significant impact and when possible, will be assessed in FOLOU to have a comprehensive overview of its magnitude.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

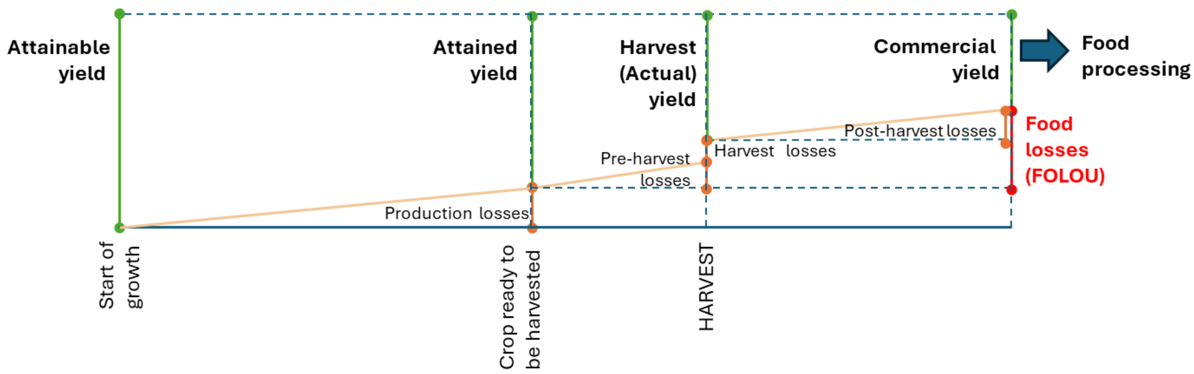

Figure 2: Schedule of production losses and food losses relative to the yield.

The technological solutions under development in WP3 typically do not cover all categories of food losses. Furthermore, some technological solutions under development in WP3 do focus on Production losses rather than on Food losses. Hereunder a table that give an overview of challenges and the category of loss.

Challenge	Type of loss
Tractor-embedded video cameras for assessing food loss in vegetable crops	Food loss
UAV-based high-resolution RGB for food loss assessment in apple	Food loss
Repeated multispectral and satellite data for production loss assessment in potato	Production loss
Repeated multispectral and satellite data for production loss assessment in maize, corn, faba bean and sunflower	Production loss
Automated fish egg sorting using multispectral camera technology in trout aquaculture	Production loss
Blockchain technology implementation in mussel aquaculture	Production loss & food loss
Market demand tools from social networks	Other

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

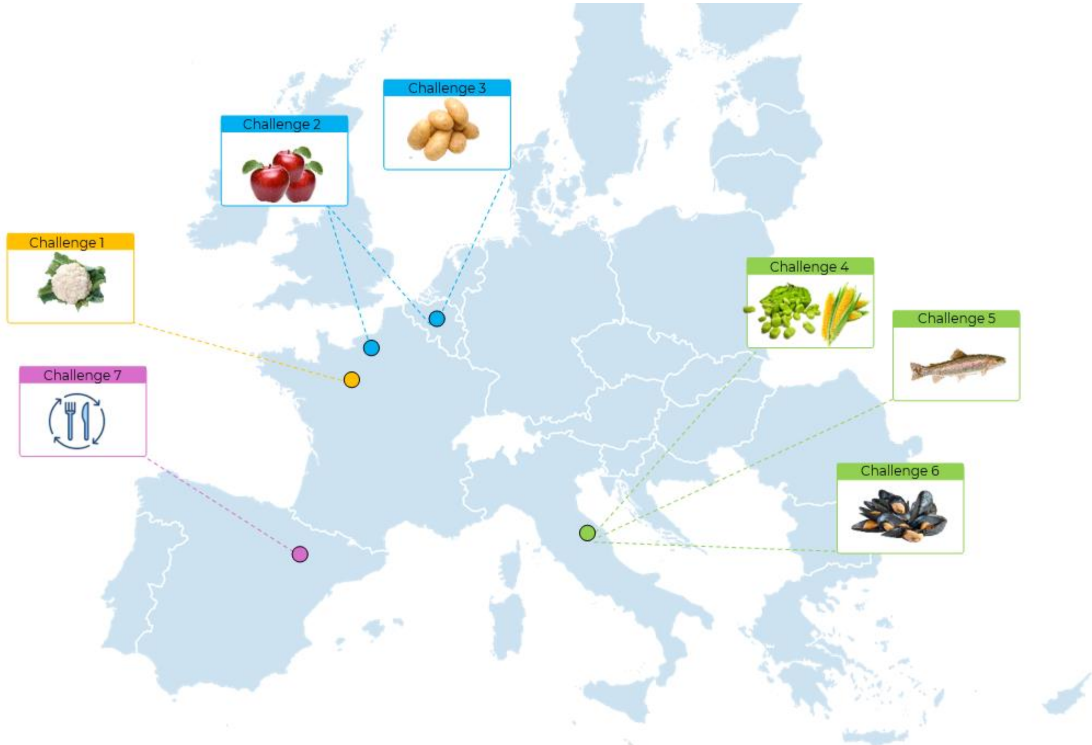

Figure 3: Geographic location of various collection sites.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

1. Tractor-embedded video cameras for assessing food loss in vegetable crops (Dilepix-T3.1)

1.1 Overview of challenge

Application of deep learning (AI) technology to measure pre-harvest and harvest losses using high resolution imagery acquired from tractor based RGB cameras.

Commodity: vegetables (tested for cauliflower)

Food loss category: Pre-harvest losses | Harvest losses

Technology: Deep learning based on tractor-based high-resolution RGB data

Current stage: Under development: first dataset collected, annotation performed, deep learning network under construction

Desired outcome: Affordable and easy-to-use tool to automatically acquire yield and food losses in the field

Cauliflower is highly valued in Europe for its nutritional benefits and culinary versatility, making it a staple in health-conscious and diverse culinary markets. Rich in vitamins and minerals, it supports various diets and can be utilized in numerous culinary forms. Europe, particularly countries like Spain, Italy, and France, is a significant producer of cauliflower, contributing substantially to the region's agricultural economy. For instance, Spain produced approximately 150,000 tons of cauliflower in 2020, while Italy and France produced around 130,000 tons and 110,000 tons, respectively. Despite its adaptability, cauliflower cultivation in Europe faces challenges such as pest infestations, diseases, and environmental stressors. The combination of remote sensing and artificial intelligence in agriculture is becoming crucial for overcoming challenges. These technologies facilitate accurate monitoring and management of crop health and environmental conditions, as well as aid in the estimation of food loss. This leads to improved yields, reduced losses, and the promotion of sustainable production practices.

1.2 Data Acquisition

Focusing on vegetable crops, the use case of cauliflower was chosen for two reasons:

- The adequate time of harvesting (autumn)
- Availability of the crop in France and Spain.

In collaboration with project partner Unilet, Dilepix accessed different cauliflower fields in Brittany, France, in Autumn 2023. Data acquisition has been performed at different stages, focusing on subsets of the entire plots:

- Pre-harvesting (about 2 to 3 weeks before harvesting).
- During harvesting (As the harvesting is done in several turns, this led to acquisition of partially harvested fields).
- Post-harvesting

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Figure 1.1. Clubroot disease which leads to growth issues.

Figure 1.2. Cauliflower not harvested on time (left) Cauliflower left after trials (right).

1.2.1 Measurements

The current solution is centred on RGB cameras embedded on tractors. However, due to the difficulties related to logistics and tractor manoeuvrability in the fields, initial footage was collected by using a drone (DJI Mavic 2 Pro) flying at low altitude.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Figure 1.3. DJI Mavic Pro 2 drone flying over a cauliflower field.

Multiple flights have been performed from mid-October to mid-November 2023:

- On 3 different fields in Brittany, France.
- Pre-harvesting, harvesting and post harvesting.
- With multiple flying heights (2m, 3m, 10m...).
- At several speeds (5km/h, 10km/h, more than 20km/h).
- With different recording settings (Aperture, Shutter, ISO...).

Flight speed has always been set according to the altitude, so that recorded data did not get blurry due to the motion speed. Recording settings have been set dynamically depending on the weather conditions.

Data source	Number of videos	Number of images
DJI Mavic 2 Pro	Over 60	About 100
Extra Equipment (e.g., smartphones, GoPro)	About 40	Around 150

Table 1.1. Overview of the acquired dataset.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

1.3 Data annotation

Dilepix has its own agronomy-oriented annotation tools. Data annotation was performed using the bounding box by marking each cauliflower manually. The data was recorded in video format, and then converted into still images. During the training phase, annotated images are extracted from the video on-the-fly. Semi-autonomous internal algorithms wered used to propagate those manual annotations on subsequent frames. Labelling occurred at the level of the entire image, so all plants in the images were annotated. A single category (Caulflower) was used for the annotations at this stage, but can be differentiated into quality levels in the future. An overview of the number of labels is provided in Table 1.2. This table also shows the amount of data between the training set and the testing dataset. To objectively evaluate the model's generalization performance, training and testing have been separated depending on where data were acquired, more precisely, from which plot. With this strategy, we kept one specific plot that the neural network has never seen for evaluation.

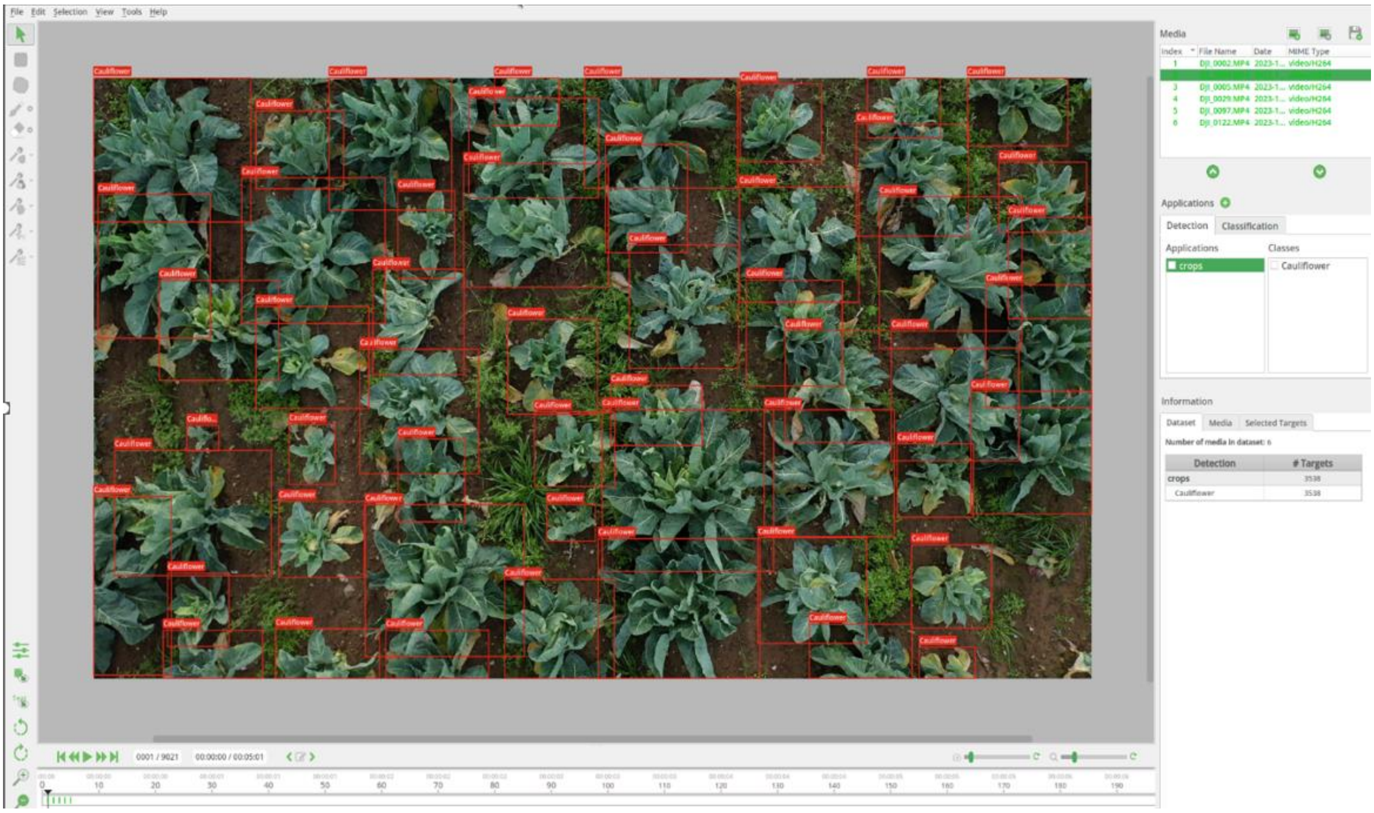

Figure 1.4. Dilepix annotation tool where each cauliflower is labelled with a bounding box.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Dataset	Quantity of Cauliflowers	Quantity of images
Annotated Cauliflowers (total)	4199	109
Training	*3538*	*87*
Evaluation	*661*	*22*
Cauliflower dataset after augmentation	18592	476

Table 1.2. Overview of the number of annotated images and individual cauliflower plants

1.4 Data augmentation

Data augmentation is particularly useful when the size of the training dataset is limited or when the dataset lacks diversity. Common techniques for data augmentation used in the project include:

- **Geometric transformations:** Rotations, translations, scaling, cropping, and flipping of images or data points.
- **Color and contrast adjustments:** Changing brightness, contrast, saturation, and hue levels of images.
- **Noise injection:** Adding random noise to images or data to simulate variations in real-world conditions.
- **Augmentation through generative models:** Generating synthetic data using generative models such as Variational Autoencoders (VAEs) or Generative Adversarial Networks (GANs).

These data augmentation techniques were applied to significantly increase the number of labels (Table 1.2).

1.5 The proposed method

To develop the method, first, a detection neural network will be developed that is capable to identify and localize objects within an image or video. Next, Dilepix plans to integrate different types of neural networks that could lead to different metric extraction, specifically, anomaly detection models with the capacity to identify patterns or instances that deviate significantly from normal behaviour within a dataset based solely on the characteristics of the input data. These models typically employ techniques such as autoencoders, variational autoencoders (VAEs), or generative adversarial networks (GANs) to learn a representation of normal data and detect deviations from this learned representation.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Networks are trained via the Dilepix technological stack based on common Deep Learning frameworks like Tensorflow or Pytorch. For performance purposes, trained models are then exported in formats that are more suited for an embedded usage.

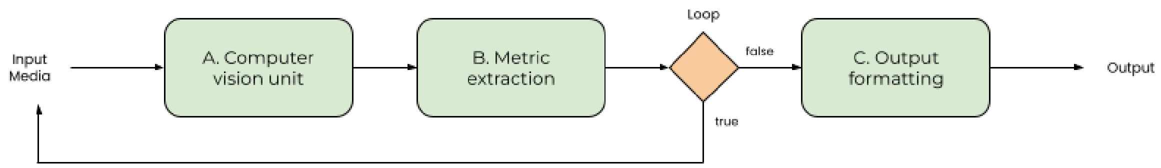

Figure 1.5. Synthetic view of the software architecture of Task 3.1.

The development method is based on software modularity which ensure code reusability, allowing to leverage existing modules across different parts of the software or in multiple applications and facilitate incremental evolution of the solution

- Module A will be dedicated to image analysis. In this module, Computer Vision and Deep Learning algorithms will be used. It aims to be as generic as possible to be reused for many applications and consequently many outputs.
- Module B will be dedicated to metric extraction. This module aims to link generic approaches from Module A to the specific application of FOLOU regarding food loss. Module A and B can then be treated individually (image by image), or globally (at the field level, using an orthomosaic). By considering individual input, it will provide precise feedback on one specific part of the field. On the other hand, working globally might provide information at the scale of an entire field or geographical area.
- Finally, computed metrics can be formatted in Module C to suit then end-user needs as good as possible.

1.6 Expected outcomes

Running these modules on a dataset taken on one single moment results in an estimate of the number of vegetables (in this case cauliflower) present at that time in the field. By taking this measurement before and after harvest and comparing the results, we can calculate the yield and the combination of pre-harvest, harvest and post-harvest losses. We are still exploring the optimal timing (minimizing measurement costs, and maximizing information) as well as the need to differentiate the cauliflower in several classes (i.e., suited for food consumption; not suited for food consumption)

1.7 Overview of the progress

Table 1.3 shows the current stage and progress of this task.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

	Functionality	State
Module A: Computer Vision unit	Detection neural network	Under development
	Anomaly neural network	Not yet started
Module B: Metric extraction	Homogeneity metric	Done
	Additional metrics	Not yet started
Module C: Output format	Single output representation	Done
	Cartography representation	In progress

Table 1.3. Overview of the progress of Task 3.1.

1.8 Key successes and challenges

To conclude, the key successes and challenges of the activities performed until June 2024 are presented below.

Key successes:

Data is collected

Annotation started

Prototyping is ongoing

Challenges:

Working with tractor, the reason why it was replaced by the drone.

Find the right crops

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

2. UAV-based high-resolution RGB for food loss assessment in apple (UGent-T3.2)

2.1 Overview of the challenge

Commodity: Fruit (currently tested on Jonagold apples)

Food loss category: Pre-harvested losses | Harvest losses

Technology: Deep learning on high-resolution RGB data acquired with UAVs

Current stage: Under development: First datasets collected; data annotation performed; deep learning network under development.

Desired outcome: Methodology to apply UAVs to estimate pre-harvest and harvest losses in fruit orchards.

Apples are crucial for a healthy diet, rich in vitamins, fiber, and antioxidants. In Europe, major apple producers include Poland, Italy, France, and Belgium, significantly boosting the agricultural economy. For instance, in 2020, Poland was the leading producer with about 3.4 million tons of apples, followed by Italy, France, and Belgium. Despite its productivity, European apple farming faces challenges like pest infestations, diseases, and climate changes. To address these issues, the integration of remote sensing and artificial intelligence has become essential. These technologies enable precise monitoring and management of orchard health and estimate the food loss. Hence this challenge is focusing on the application of deep learning (AI) technology to measure pre-harvest and harvest losses using high resolution imagery acquired from UAV-based RGB cameras of fruit crops. The method under development must be able detect and count damaged fruit to directly estimate the yield, pre-harvest and harvest losses. Apart from the deep learning methodology, we also need to explore the most optimal viewing angle and measurement condition.

2.2 Data acquisition

UAV flights were performed to collect very high resolution RGB imagery from five Jonagold orchards in Flanders and northern France. The Jonagold variety was selected because it is of high economic importance and because it is sensitive to scab disease which will lead to potential pre-harvest losses.

2.2.1 Material & equipment

UAV data: A DJI M350 UAV equipped with a high-resolution camera (DJI Zenmuse P1 with 50mm lens, 45MP camera). Images were collected over the five different orchards in September 2023, close to the time of harvest. In each orchard, the flight height was 12 m, and images were collected from 5 different viewing angles (see further). At this altitude, the imagery has a Ground Sampling Distance (GSD) of 0.12 cm.

Ground truth reference data: In each orchard, 30 trees were randomly sampled, collecting the number of good (marketable) apples, apples fallen on the ground, and damaged apples

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

(non-marketable) on the tree. Also, the exact location of the tree was recorded with a high precision RTK GNSS system.

Figure 2.1. Equipment for high resolution UAV acquisition.

Figure 2.2. Tree sampling.

2.2.2 Protocol

Standard UAV acquisitions position the camera to look straight down (nadir viewing, 90° angle relative to the horizon). However, a relatively small part of the canopy is then visible, with the largest part being occluded from the view. An initial test revealed that the maximum viewing angle still allowing to view the full tree and the ground was 50°. Therefore, we collected dataset having a nadir view, 70° and 50° viewing angle. For the 50° and 70° viewing angles, we collected data from both sides of the tree line, so we can see the

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

full canopy of the tree. Further image processing will reveal which viewing angle or combination of angles is most suited.

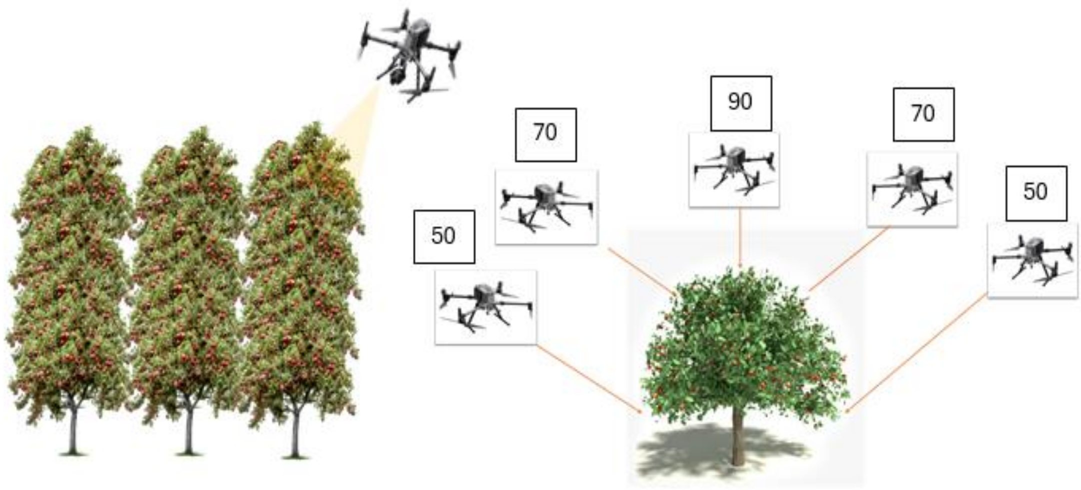

Figure 2.3. Different viewing angles for high resolution RGB acquisitions.

Figure 2.4. Samples of images from different angle views.

2.2.3 Outcomes of data acquisition

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

A total number of 8204 images were collected over the five orchards; an overview per orchard is given in Table 2.1.

High-Resolution RGB datasets	Localisation	Resolution	Number of images
Dataset 1	Michelbeke (Belgium)	8192x 5460	2456
Dataset 2	Sint Gillis Waas (Belgium)	8192x 5460	2317
Dataset 3	Zottegem (Belgium)	8192x 5460	1757
Dataset 4	Pitgam (France)	8192x 5460	829
Dataset 5		8192x 5460	845

Table 2.1. High resolution RGB dataset statistics

2.2.4 Data preparation

Data preparation before annotation

To prepare our datasets for annotation, relevant subregions were extracted from the full image. Each subregion has a size of 3500x4000 pixels (14 MP). Subregions were selected to contain areas of the orchard canopy that were likely to contain damaged apples, since damaged apples are sparse and not uniformly distributed throughout the orchard. This approach allows us to avoid annotating unnecessary parts and saves time and resources, especially given the size of the images, potential overlap, and the overall size of our datasets. Python scripts were created in order to cut the relevant region from the raw image, save the chunk and the coordinates of the relevant region in a text file.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

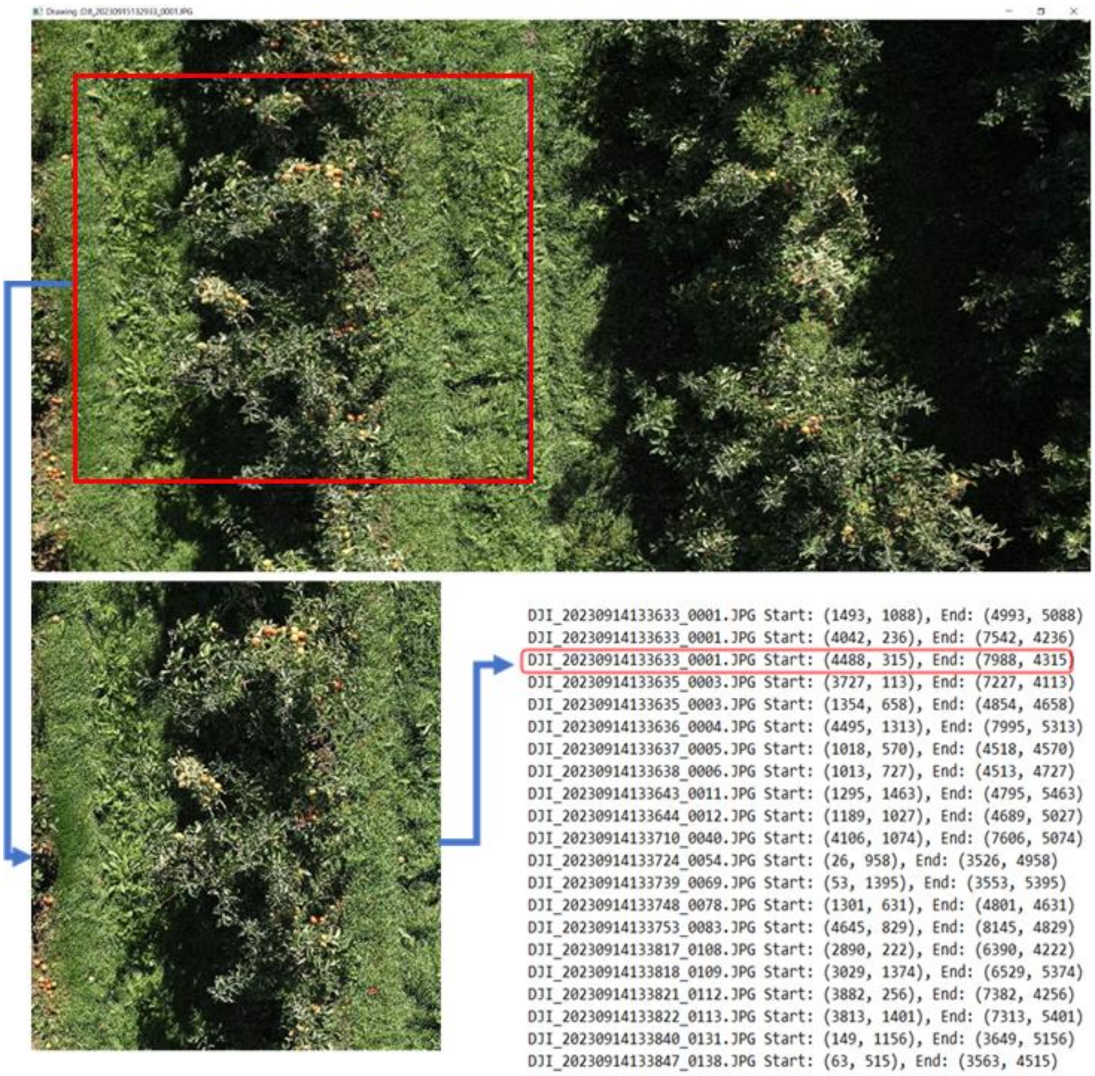

Figure 2.5. Selection of relevant region and recording the coordinates in text file. (Top = complete image; cut-out left = selected region for annotation; text right = reference format in text file format)

In the Figure 2.5. shows extracting relevant subregions from the high-resolution RGB images captured by UAVs, focusing specifically on areas likely to contain damaged apples. This targeted approach, facilitated by custom Python scripts, ensures efficient processing and maximizes the relevance of the data for annotation. The information of the coordinates of the relevant region and the name of the raw image are stored in a text file.

<u>Data preparation after annotation</u>

To expand the size and diversity of our datasets, data augmentation techniques will be implemented. This step is indispensable for the development of robust deep learning models, as they rely on large datasets to achieve optimal performance. Gaussian filter, mirror and flipping are among the potential annotation techniques that will be employed.

2.3 Data annotation

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Annotation in high-resolution UAV images is a time-consuming and labor-intensive process. To manage this, we outsource this task to specialized annotation service providers. An annotation guide document was created to standardize and streamline this process. The annotation is done manually which affects the efficiency and time required for annotation. Concerning the annotation, polygon technic is preferred over bounding boxes, since this is more accurate and provides grain fine details of the annotated object especially in the case of complex shapes. However, marking polygons is also more time-consuming, which is why we will later include a comparison of the model performance of the polygons vs the bounding box techniques.

The annotation uses 3 main classes (Fig. 2.6):

- Class 1: Good apples

  Healthy apples (marketable) that look good and do not have any spots, lesions, or reduced size. The level of ripening of apples does not matter.

- Class 2: Fallen apples

  Apples on the ground, in whatever stage

- Class 3: Damaged apples

  Damaged (unmarketable) apples include apples with spots, lesions or growth problems. In this class, an attribute concerning the level of visibility is also added. This attribute will allow us to improve the angle for more visibility and study the behavior of the models regarding the visibility of the apples.

- Class 4: Apple (uncertain): We cannot decide if it is an apple or not.
- Class 5: Fallen apples (uncertain): The scenario where we are uncertain whether it is a fallen apple or something else.
- Class 6: Damage (uncertain): Not clear if the apple is damaged or not.

Figure 2.6. Sample images of the different classes.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Based on a review on supervised deep learning research for fruit detection, a target of 40000 annotations was set. The annotation process is currently in process, an example of a first batch of this annotation is given in Fig. 2.7.

Figure 2.7. Sample of image annotation with the classes and visibility attribute.

2.4 The proposed method

Deep learning architectures, such as YOLO (You Only Look Once) and Faster R-CNN (Region-based Convolutional Neural Network) will be explored for the detection of damaged apples in high-resolution RGB UAV images.

The method uses the following workflow:

- Stage 1: Development of model that is capable of detecting healthy and damaged apples with high accuracy.
- Stage 2: Comparison of datasets under different viewing angles, and recommendations (optimization) of viewing angles and image settings
- Stage 3: Upscaling of method to field and regional scale (variations in canopy, density, Covering other variety of apples and diseases)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Each stage in the workflow above involves iterative refinement, algorithmic adjustments, fine-tuning, and continuous experimentation which are integral parts of this ongoing work.

The models will be trained and tested on different (independent) datasets to verify model robustness.

2.5 Expected outcomes

The expected outcome is the development of an automated method based on high resolution UAV RGB images and deep learning to directly measure the pre-harvest and harvest loss in apple orchards. The solution can be also generalized to other type of orchards with some adjustments.

2.6 Development states

Main Tasks	Subtasks	State
Remote sensing / Field campaign	High resolution RGB data acquisitions	Done (first year)
	Data acquisition	Done (first year)
	Data field analysis	Done (first year)
Data processing	Data preprocessing	Done
	Data annotation	In progress
	Data augmentation	In progress
Development of object detection algorithms using nadiral view.	Prototyping	Under development
	Model development	Not yet started
	Model evaluation	Not yet started
Development of object detection algorithms using multiple views	Prototyping	Not yet started
	Developing the model	Not yet started
	Testing the model	Not yet started

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Development of food loss estimation algorithm	Development of food loss estimation algorithm	Not yet started

Table 2.2. Overview of the progress of Task 3.2.

2.7 Key successes and challenges

In conclusion, the following summarizes the key successes and challenges of the activities conducted up to June 2024:

<u>Key successes:</u>

Data is collected from various orchards(diverse and rich data)

Preprocessing is done

Annotation is ongoing

Prototyping is ongoing

<u>Challenges:</u>

Difficulty to find collaborative farmers especially in the period of pre-harvesting and harvesting.

Find damage crops.

Preprocessing our huge dataset

Annotation of the images: we should deal with :

Big images

Complex background

Occlusion

Size and the shape of the apples & the damage

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

3. Repeated multispectral and satellite data for production loss assessment in potato (UGent, T3.2 & T3.3)

3.1 Overview of the challenge

Commodity: Potato

Food loss category: Production losses

Technology: Crop growth model, fed with data from multispectral and multitemporal datasets

Current stage: Ongoing; Experimental trial was carried out in 2023 and will be repeated in 2024 and possibly 2025. Similarly, large-scale data collections are planned for 2024 for satellite data input.

Desired outcome: Automated method estimating production losses for potato based on satellite or UAV data.

Belgium is a notable potato producer within Europe, contributing significantly to the region's agricultural output. In 2020, Belgium produced approximately 4.5 million tons of potatoes. These potatoes are essential for both the local diet and culinary uses, valued for their carbohydrate and potassium content. Despite its success, potato farming in Belgium faces several challenges, including diseases, pests, and variable climate conditions. To address these difficulties, advancements in remote sensing and artificial intelligence are increasingly being utilized. These technologies enhance the monitoring and management of crop health and environmental factors, leading to improved management practices, better yields, and reduced losses.

3.2 Data acquisition

An experiment was set up in the experimental farm of Bottelare (Flanders, Belgium) in the 2023 growing season. This trial explored differences between potato due to biostimulants aiming to improve root growth, drought resistance and yield (see further).

Two types of data were collected from this trial:

- Remote sensing data: multitemporal multispectral and thermal imagery
- Field data measurements: Including pre-harvest and harvest sampling, fresh and dry weigh of leaves and stem, roots and tubers, size and number of the tubers.

3.2.1 Material & equipment

Field trial: The experimental design chosen for the trials is a randomized complete block design (RCBD), consisting of 6 treatments(Humifirst, seaweed, alphasol, delphan+, quantum, and irrigation)and one untreated treatment (control) arranged in three blocks, with each block containing 14 plots. The trial comprises a total of 42 elementary plots.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Each plot measures 12m in length and 3m in width, resulting in a total trial area of 36m x 42m = 1512m².

Trial Map Treatment Description		
Trt	Code	Description
1	CHK	control
2		Humifirst 50 L/ha
3		Seaweed 2 L/ha;Seaweed 2 L/ha;Seaweed 2 L/ha
4		Alphasol 2 L/ha;Alphasol 2 L/ha;Alphasol 2 L/ha
5		Delfan plus 2 L/ha;Delfan plus 2 L/ha;Delfan plus 2 L/ha
6		Quantum 2 L/ha;Quantum 2 L/ha;Quantum 2 L/ha
7		irrigation 2 L/ha

501 5	502 6	503 2	504 3	505 4	506 1	507 7	601 2	602 3	603 1	604 6	605 4	606 5	607 7
301 3	302 5	303 4	304 6	305 1	306 2	307 7	401 4	402 6	403 2	404 5	405 1	406 3	407 7
101 1	102 2	103 3	104 4	105 5	106 6	107 7	201 5	202 1	203 3	204 4	205 2	206 6	207 7

Figure 3.1. Description of trial Map treatments.

Pests and pathogens were monitored and treated according to good agricultural practices and the overview of the treatments will follow later.
During the preharvest period, the average of 10 leaves per plot was measured by SPAD for chlorophyll content measurement.

During midharvest(24th of august 2023), four samples were taken from each plot. These samples were then separated into leaves and stems, and roots and tubers. The weight of all parts was measured directly for each treatment. The size and number of tubers were measured using the flat sizing template. The dry weight of each sample was determined, after steaming at 65°C for one week.

The harvest was carried out manually due to adverse weather conditions, as the soil was wet from the rain. Consequently, 50% of our tubers remained underground. After the harvest, the tubers were sorted by size, number, fresh weight, and dry weight. Additionally, we recorded the under-water weight (UWW) and dry matter (DM) yield. We aimed to identify the most effective biostimulant to produce high-quality potatoes within the optimal market size of 35-50 mm, considering dry matter content and treatment efficacy to enhance marketability and productivity while avoiding losses of potatoes outside this size range.

Remote sensing protocol : Aerial imagery were captured with a DJI Mavic 3 MS and with a DJI Matrice 600 Pro UAV equipped with a multispectral (MicaSense RedEdge Dual MX Camera system) as well as a thermal camera (Teax ThermalCapture): Eleven flights were conducted over the field between June 12, 2023, and September 15, 2023.

Drones	Equipment	Dates
DJI Matrice 600 Pro UAV	**Multispectral:** MicaSense RedEdge Dual MX Camera system)	18/07/2023 26/07/2023

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

	Thermal: Teax ThermalCapture	09/08/2023 14/08/2023 21/08/2023 31/08/2023 08/09/2023 15/09/2023
Mavic 3 MS	RGB camera 20 MP	12/06/2023 26/06/2023 04/07/2023 18/07/2023 14/08/2023 21/08/2023 08/09/2023
	Multispectral camera	

Table 3.1. Data Acquisition Equipment & Material

Permanent ground control points (GCP) were positioned in the experimental area. Before each flight, a reference grey panel was measured with the UAV cameras. Additionally, 6 gray referent targets were placed near the target area. A plate covered with aluminum foil was positioned in the view of the flight path to measure incoming longwave radiation. Flights were performed under sunny conditions (if possible) and at a flight altitude of 50m, with 75% horizontal and vertical overlap (for the thermal images).

Figure 3.2. Left: white reference panel captured in band 5 (642 nm–658 nm) of the MicaSense dual camera system. Right: Aerial image of the 6 RRTs captured in band 2 (459 nm–491 nm).

3.2.2 Data preparation

Remote sensing data preparation: The UAV images are processed with Agisoft Metashape to generate orthomosaic images. These orthomosaic images are then imported into Matlab R2024, where we performed calibration and necessary corrections using empirical line methods. This calibration ensured that the images accurately represented the field conditions.

Finally, various vegetation indices, such as NDVI (Normalized Difference Vegetation Index), SAVI, GNDVI.... are calculated to assess crop health, vigor, and stress levels. These detailed analyses provide valuable insights into the overall condition of the crops throughout the growing season.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

3.3 The proposed method

This section outlines the integration of remote sensing (RS) data with the crop growth model (APSIM) through the use of a radiative transfer model (PROSAIL). The APSIM model requires inputs such as soil properties, climate data, and crop management practices. Inverse modeling uses RS data, including vegetation indices and soil moisture estimates, and SPAD measurements for chlorophyll content to calibrate APSIM. PROSAIL, a radiative transfer model, simulates canopy reflectance to estimate leaf and canopy parameters, which refine APSIM inputs. Actual yield simulations reflect current conditions, while attainable yield simulations assume optimal conditions. This integration minimizes non-RS data input, enhances yield prediction accuracy, and improves crop management and productivity.

In 2024 and 2025, we will sample regular growing fields of potato, in collaboration with project partner Warnez. In these fields, a similar approach will be used to estimate attained and attainable yield, but then based on satellite data of vegetation development from Sentinel-2, complemented with other datasets concerning the soil type (publicly available datasets), weather conditions and possible soil moisture content (from Sentinel-1). The field sampling during the growing season includes the reference sampling of chlorophyll content and leaf area to verify whether the PROSAIL-estimation of these key variables using the Seninel-2 data is correct. At the end of the growing season, samples of potato yield and potato yield losses will be collected here.

3.4 Expected outcomes

We aim to integrate UAV and satellite remote sensing data in a radiative transfer model and a crop growth model to predict yield and production losses and identify areas for improvement. This integrated approach allows us to make data-driven decisions to optimize biostimulant use, irrigation, and other agricultural practices. Ultimately, our objective is to correctly predict production losses, but also to be able improve the quality and quantity of potato harvest, ensuring that we meet market demands and reducing the incidence of suboptimal or wasted produce.

3.5 Current stage

Main Tasks	Subtasks	State
Remote sensing / Field campaign	Multispectral data acquisition	Done (first year)
	Data field analysis	Done (first year)
	Data preprocessing	Done (first year)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Data processing	Data annotation	In progress
	Data augmentation	In progress
	Prototyping	Under development
Development of crop modelling	Development of crop modelling	Not yet started
Development of production loss estimation algorithm	Development of production loss estimation algorithm	Not yet started

Table 3.2. Overview of the progress of the method development

3.6 Key successes and challenges

In conclusion, the following summarizes the key successes and challenges of the activities conducted up to June 2024:

Key successes:

Integration of advanced remote sensing technologies.

Field data collection and managements.

Innovative data analysis and modelling (predictive crop growth modelling and vegetation indices calculations).

Potential for yield improvement and loss reduction.

Challenges:

Difficulty to find collaborative farmers especially in the.

Adverse weather conditions (impact on harvest and yield).

Data collection and processing complexity (High volume and variety of Data, calibration and accuracy).

Technological and methodological constraints (Model development and validation, combining field data and RS data together in order to estimate the food losses).

Storage and Post Harvest management (preventing Post harvest losses and integration of new parameters).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

4. Repeated multispectral and satellite data for production loss assessment in maize, corn, faba bean and sunflower (UNIVPM-T3.2 & T3.3)

4.1 Overview of the challenge

Commodity: Maize, corn, faba bean, sunflower

Food loss category: Production losses

Technology: Repeated multispectral imagery (UAV or satellite-based) to estimate production losses

Current stage: Under development; field experimental trial is ongoing, data have been collected for one season and will continue to be collected in the upcoming seasons; method for production losses under development

Desired outcome: Automated method that estimates production losses of selected crops using repeated UAV or satellite imagery.

Maize, corn, faba beans, and sunflowers are crucial crops in Europe, essential for human nutrition and animal feed. Leading producers include France, Romania, Hungary, the UK, Germany, Ukraine, Russia, and Bulgaria. Italy is also a significant player, particularly in maize and sunflower production, with over 6 million tons of maize and 2 million tons of sunflower seeds harvested annually. These crops face challenges like pest infestations and climate variations. The integration of remote sensing, crop medeling, and AI, will allow improving crop health monitoring, optimizing resource use, and increasing yields. Moreover an accurate production loss can be estimated.The research team of Università Politecnica delle Marche (UNIVPM) is focusing on the agronomic perspective, considering production losses (PL) as the 'gap between actual crop yield (attained yield) and potential crop yield (attainable yield)' (Pérez-Méndez et al., 2021). Crop yield loss is due to the action and interaction between biotic (e.g., pests, insects) and abiotic stresses such as drought and heat waves (Ramegowda and Senthil-Kumar, 2015). Both biotic and abiotic stresses can be limited by various agronomic management practices, that include irrigation, pests and insects control and soil tillage regimes. With the aim to detect and quantify the gap between actual crop yield and potential crop yield, an open field trial was set up at the 'Pasquale Rosati' experimental farm of UNIVPM located in Agugliano, Province of Ancona, Italy (**Error! Reference source not found.** 4.1 (left)).

4.2 Data acquisition

4.2.1 Material & equipment

Three different agronomic management types subjected to two soil tillage regimes are tested:

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

a) 'Business-as-usual' (BAU) is the locally common management approach representative in terms of irrigation, nutrition, and pest control. In this management, no irrigation is applied, standard nutrient management is conducted, and pest control followed the Integrated Pest Management principles of the Marche region.

b) The 'zero-stress' (ZST) management is designed to entirely prevent biotic and abiotic stresses throughout the entire growth cycle. This is achieved by implementing a fertigation system to deliver the optimal amount of water and nutrients (which can also be applied for canopy cooling), along with a pest management control schedule. The result can be considered close to the 'attainable' yield.

c) The 'enhanced conventional' (ECV) management includes supplemental irrigation, standard nutrient management, and pest control according to the Integrated Pest Management principles of the Marche region.

While the comparison between BAU and ZST managements would allow to assess the gap between actual crop yield and potential crop yield, ECV will address crop adaptation to climate change and has the potential to be the most readily adopted by farmers. The two different tillage regimes are as follows:

a) Ploughing, the most widespread tillage at national level and within the Mediterranean area (Alcántara et al., 2016).

b) Minimum tillage, a more conservative and sustainable option compared to ploughing (Xiao et al., 2023).

The three management options and the two tillage regimes will be tested on four of the most representative crops of Mediterranean cropping systems (Cramer et al., 2018) in a 4-year rotation scheme: wheat (*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.), sunflower (*Helianthus annuus* L.), faba bean (*Vicia faba* L var. minor) and maize (*Zea mays* L.).

The three management systems (i.e., BAU, ZST, and ECV) have been assigned to a single nearby field, homogeneous in terms of previous management and soil characteristics, each measuring 120 × 30 m. Within each field, two tillage regimes (main plot, 15 × 120 m) and four crops (sub-plots, 4 × 8 m) were arranged according to a split plot experimental design, replicated thrice. Therefore, the total number of subplots is 72 (3 fields × 2 tillage regimes × 4 crops × 3 replicates).

To upscale (e.g., basin, territorial, regional scale) the results of ground sampling and drone monitoring, a validation of the satellite data is necessary. The validation will be carried out by planning simultaneous flights over the "calibration" and the "plot field" (Figure 4.1 (right)).

The Calibration trial (43°32'44.1" N 13°22'12.3" E) is managed according to BAU management and is shown in Figure 4.1. The choice of 1 ha area was driven by the need to acquire an adequate number of satellite images for validation.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

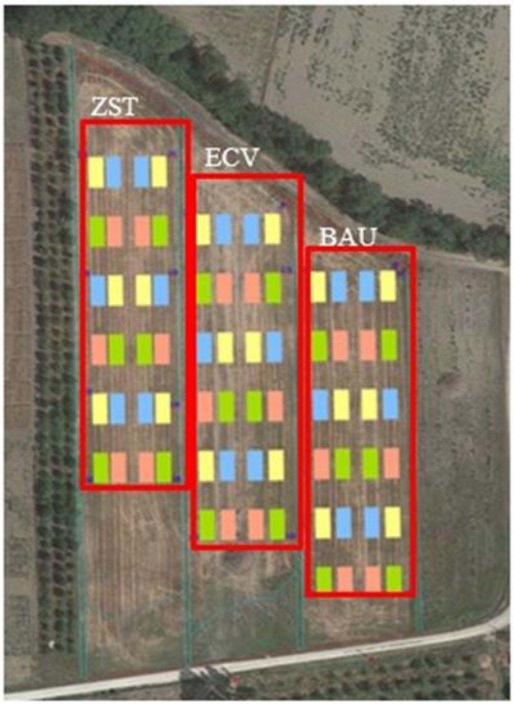

Figure 4.1. A map of the experimental design of the macro-plot trial (43°32'39.8"N 13°21'39.5"E) (left) and the Calibration trial (43°32'44.1" N 13°22'12.3" E) (right). The macro-plot trial shown in the figure is an example of rotation in a phase where all four crops are simultaneously in the field. BAU: Business-As-Usual, ZST: Zero Stress

The trial was set up in the central block plots of the open field trial described in the previous subsection. Soil moisture was measured with LoRaWAN technology and irrigation was scheduled.

4.2.2 Remote sensing data collection

The data collection campaign relies on different kinds of sensors:

a) UAV-based multispectral imagery
   1. Micasense Altum
   2. MAIA S2 and MAIA WV2
b) Satellite-based multispectral imagery
   1. Sentinel-2
   2. Planet

Micasense Altum

The Altum sensor integrates a radiometric thermal camera with five narrow bands, producing thermal, multispectral, and medium-resolution imagery in a single flight. Altum is equipped with Direct Light Sensor (DLS). This sensor is installed on a DJI M200 UAV for all the flights.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Figure 4.2 shows an example of spectral index over the test and calibration field.

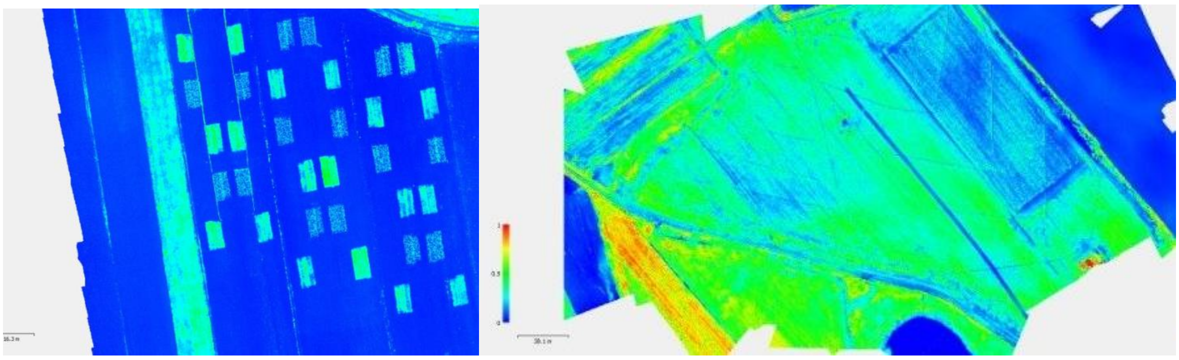

Figure 4.2. Left: Spectral Index (NDVI) over wheat and faba bean on the test field, Right: Spectral Index (NDVI) over wheat and faba bean on calibration field (Feb 2024).

Maia S2 and MAIA WV

MAIA is a multispectral camera designed and developed by SAL Engineering and Eoptis to be used on board UAVs or RPAS, entirely made in Italy. MAIA is composed of 9 1.2 MP sensors (9 monochromatic sensors with related bandpass filters in the MAIA S2 filter-set, and 8 monochromatic + 1 RGB in the MAIA WV filter-set) to acquire images in the VIS -NIR spectrum. MAIA WV has the same wavelength ranges as DigitalGlobe's WorldView-2™ satellite, from 395 nm to 950 nm. MAIA S2 has the same wavelength ranges as ESA's Sentinel-2™ satellite, from 433 to 899.5 nm. The systems are equipped with an Incident light sensor (ILS) to measure and correct for ambient radiation. The MAIA is also installed on the DJI M600 Pro UAV.

Figure 4.3. Left: MAIA Sensor. Right: Incident Light Sensor (ILS).

PlanetScope

PlanetScope (by Planet) is a constellation of approximately 130 satellites, able to image the entire land surface of the Earth every day (a daily collection capacity of 200 million km²/day). PlanetScope images are approximately 3 meters per pixel resolution. The PlanetScope satellite constellation consists of multiple Dove satellites.

Figure 4.4 is an example of the false color composite of the Agugliano area (test field and calibration field). Red colors indicate green vegetation, with denser and healthier vegetation showing a more bright red color.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

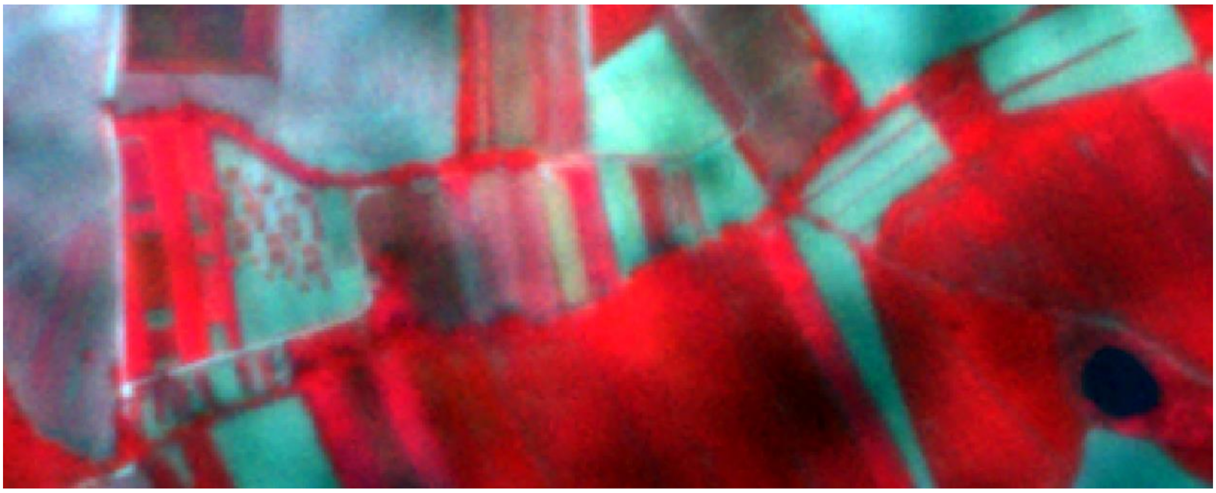

Figure 4.4. Example of acquisition using NRG representation of the Agugliano site. Data: Planet – 8 band (instrument: PSB.SD). Acquisition date: 2024-05-14T10:04:09

Sentinel-2

The Copernicus SENTINEL-2 mission includes two polar-orbiting satellites, S2-A and S2-B, which are arranged in the same sun-synchronous orbit, positioned 180° apart. This mission is designed to observe changes in land surface conditions. It features a broad swath width of 290 kilometers, and high revisit time (10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes) to support monitoring of Earth's surface changes.

4.2.3 Protocol

Remote monitoring

Combined with surveys carried out on physiological status and at harvest, flights by DJI Matrice 600 Pro drone with two multispectral maya S2 and WV2 cameras and by DJI Matrice 200 drone with the Altum sensor are performed. These flights last all the growing season with a frequency of 1 flight/week.
Permanent ground control points (GCPs) have been deployed over the areas of interest and reflectance panels are also used to generate geometrically and radiometrically corrected data.
The type of data used for validation of the drone-collected data will be derived from Sentinel-2 and PlanetScope remote sensing images.

Soil analysis

Before the main soil tillage different profiles (three for each management) are excavated and classified. These profiles will assess soil organic matter content at baseline (T0) and monitor its dynamic over time. Chemical and physical soil status will be evaluated through composite and core sampling at fixed depths after pedological characterization. This sampling establishes the baseline for future soil monitoring campaigns. Both organic and

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

mineral soil layers will be sampled and analyzed regularly (e.g., annually) to detect potential chemical changes. The following variables will be monitored at varying frequencies, calibrated for each:

a. Soil water content: A fertigation system and a network of soil moisture sensors were installed. A nearby weather station measures main agro-meteorological parameters daily.

b. Crop Phenology: Ten plants/plot, showing regular growth since emergence, will be selected. They'll be geo-referenced, individually tagged, and observed weekly (from May for summer crops, and November/December for fall/winter crops) using BBCH scale for sunflower and faba bean, Ritchie's scale for maize, and Zadoks' scale for wheat.

c. Crop physiological status: Monthly surveys will monitor crop status, becoming weekly at critical stages like flowering. Gas exchanges will be analyzed using an infrared gas exchange analyzer during maximum photosynthetic activity (12:00 h - 14:00 h, 1000–1800 μ mol m−2 s−1). Chlorophyll content and leaf area index will be estimated using SPAD-502 and a ceptometer, respectively.

d. Yield and yield components: Yield components will be determined on the ten plants used for phenological surveys:

1. For maize: plant density, ears per plant (and per square meter), rows per ear, grains per ear, and 100-grain weight.
2. For sunflower: plant density, heads per plant (and per square meter), achenes per head, and 100-grain weight. At flowering, flower numbers will be counted destructively on five heads per plot.
3. For wheat: tiller density, spikes per plant (and per square meter), spikelets per spike, kernels per spikelet, and 100-kernel weight.
4. For faba bean: plant density, pods per plant (and per square meter), seeds per pod, and 100-seed weight.

4.2.4 Outcomes of data acquisition

Starting from the acquired data it will be possible to have:

- Soil analysis
- Time-series of Soil volumetric water content, temperature and salinity using the drill&drop probes
- Time-series of Phenology measurements (set of regular reports)
- Time series of Crop physiological status (set of regular reports)
- Yield and yield components
- High resolution remotely sensed time-series from UAVs (multispectral orthophotos)
- Medium/Low resolution remotely sensed time-series from satellites (multispectral images)

4.2.5 Data preparation

Proximal acquired data have been archived and are arranged in spreadsheets. Remote data are being processed through different pipeline according to the payload.

Data from Micasense Altum are directly processed through state of the art software (i.e. Pix4D, Metashape,....)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Data from MAIA S2 and WV2 require a pre-proccesing. Here is the list of main pre-processing tasks:

- Geometric correction and generation of 'undistorted' images through the calibration parameters that are included with each sensor (enabled as a tick; otherwise disabled).
- Coregistration (or Stitching) based on a reference image for image stitching of each band, with pixel-by-pixel convergence (enabled as a tick; otherwise disabled).
- Radial radiometric correction: This is active as a tick, and it corrects the border effects of the image (usually darker pixels) that can arise due to lens curvature.
- Radiometric correction: This is based on the following selectable options in the field 'Radiometric'

Data of Sentinel-2 (L2A) are fetched using a custom python script that invokes third party APIs.

4.3 The proposed method

The attainable and attained yield will be evaluated using two main approaches. One model will be data-driven considering weather data and vegetation index-based time-series. We will extend our previous work "Time Series from Sentinel-2 for Organic Durum Wheat Yield Prediction Using Functional Data Analysis and Deep Learning" (Mancini et al., 2024) to establish a correlation between observed variable from remotely sensed data (UAVs and satellites) with ground measurement (at the end of the season). The use of different field with similar crop and management observed at different scale is a key point to establish a way to map areas with different performances in terms of nitrogen content and the potential yield. The fields are monitored not only using remote sensing devices, but also probes installed in the soil that can support the understanding of water availability for roots. Starting from the analysis of time series it is also possible to evaluate the different way to reach the final maturation state that is another key point. Different kind of images will be used ranging from high-resolution imagery collected using UAVs to satellite platforms with pixel size that ranges from 3 to 20m.

Another approach will consider crop simulation models which use biophysical parameters to simulate crop growth and yield. The model chosen is the Decision Support System for Agrotechnology Transfer (DSSAT). DSSAT integrates modules on crops, climate and soil models to simulate the crop growth cycle and predict the impact of different environmental and agronomic management factors on production trends (Jones et al., 2003). The objective is to combine remotely sensed data from UAVs and satellite with the crop model to improve model calibration and the accuracy of estimates through a data assimilation process.

The integration of a radiative transfer model with a crop model, using a canopy structure variable such as the Leaf Area Index (LAI), enables the simulation of variables like reflectance in the VIS and NIR range across the entire spatial domain at any date when remotely sensed reflectance data are available. By comparing these simulated reflectance values with actual measurements, it is possible to re-estimate certain model parameters or initial conditions. This process enhances the accuracy of simulating state variables related to reflectance, particularly LAI. Since crop yield is highly dependent on intercepted radiation, which in turn is closely linked to LAI, refining the LAI simulation leads to more precise yield predictions. This optimization process, known as "assimilation", allows for localized adjustments in the crop model, thereby improving its overall performance (Launay and Guerif, 2005).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

4.4 Expected outcomes

The objective is to enhance the outcomes of ground surveys by incorporating satellite and drone monitoring images. Tailored analyses will be conducted, with a particular emphasis on estimating production loss (i.e., the difference between the attainable and attained yield) and subsequent yield variability based on the crops under evaluation. This entails measuring both a) pre-harvest loss, such as natural seed drop, and b) harvest loss, which includes mechanical losses incurred during harvesting processes.). This initiative builds upon recent advancements in the field and aims to refine and extend existing methodologies to estimate food loss.

4.5 Development states

Main Tasks	Subtasks	State
Remote sensing / Field campaign	High resolution RGB data Acquisition	Done
	Multispectral data Acquisition	Done
	Data field analysis	Done
Data processing	Data preprocessing	Done
	Data Annotation	N/A
	Data Augmentation	N/A
Development of crop modelling	Prototyping	Under development
	Testing the model	Not yet started
Development of production loss estimation algorithm	Prototyping	Not yet started
	Testing the model	Not yet started

Table 4.1. Overview of the progress of the tasks T3.2 & T3.3.

4.6 Key Successes and Challenges

Key Successes:

- Data collected from different fields (test and calibration fields; different and large data with ground and remote sensing data)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

- Preprocessing: ongoing
- Annotation ongoing
- Prototyping in progress

<u>Challenges:</u>

- Complex management of trials (different management and different crops)
- Difficulty in finding collaborative farmers, particularly during pre-harvesting and harvesting periods over different management
- Preprocessing and annotating our large dataset (ground data, remotely sensed data).

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

5. Automated fish egg sorting using multispectral camera technology in trout aquaculture (UNIVPM-T3.4)

<u>Commodity</u>: Aquaculture, case study in trout

<u>Food loss category</u>: Production

<u>Technology</u>: Machine learning to detect infected fish eggs using multispectral camera; automated sorting out of infected fish eggs

<u>Current stage:</u> Proof of concept of multispectral system; Prototype of fish

<u>Desired outcome</u>: Automated fish egg sorting system using multispectral imagery, machine learning and applied robotics.

The fisheries and aquaculture sectors have been increasingly recognized for their essential contribution to global food security and nutrition in the twenty-first century. Further expansion of this contribution requires the acceleration of transformative changes in policy, management, innovation and investment to achieve sustainable and equitable global fisheries and aquaculture.

In 2019, the world trout production was 939878 tons (FAO) and has been increasing since 2015 (+21 % in volume between 2015 and 2019). Main species farmed is the rainbow trout (Oncorhynchus mykiss) which accounted for 97 % of the total volume in 2019. The EU is the second largest producer in the world ($183 \cdot 10^3$ tons in 2019: 20 % of world production). The rainbow trout farming industry has been developing for several hundred years, and many aspects are highly efficient, using well-established systems. However, current research and development is continually attempting to increase production efficiency and sales by increasing rearing densities, improving recirculation technology, developing genetically superior strains of fish for improved growth, controlling maturation and gender, improving diets, reducing phosphorous concentrations of effluents, and developing better marketing.

5.1 Overview of the challenge

The production cycle of rainbow trout can be divided into different phases, each characterized by important critical aspects that have to be considered to ensure high productivity rates and high-quality standards. As reported in the literature, the knowledge on the majority of these aspects has been deeply consolidated and this allows farmers to minimize losses in the supply chain. **The most critical phase is represented by the early embryo development** from fertilization to the achievement of the eyed egg stage, subjected to the absence of standard procedures or technologies designed to estimate losses. Fish egg loss due to fungal and bacterial infections **represents the main issue in the entire productive cycle.** Mortality during the incubation of eggs from a contamination by water molds (*Saprolegnia* spp.) and common aquatic pathogens ubiquitous in water results in annual production losses in the hatchery production of rainbow trout. The management of fungal and bacterial infections has historically relied on the use of chemical treatments which may have negative effects on human and environmental health. Currently, the early embryo development is characterized by the use of upwelling incubators subjected to visual check, manual removal of dead embryos (white), and use of chemicals for disinfection of the whole batch. **This is time-consuming and not sufficiently accurate to prevent spreading of bacteria and fungi** (Figure 5.1). High rates of bacterial and fungal infection can result in the discard of whole egg batches (including healthy embryos), representing the main gap within the supply chain. Since oomycetes species like *Saprolegnia* sp. can colonize surface of the dead eggs to then suffocate the surrounding living ones, a constant and quick identification and removal of infected eggs is necessary. In fact, this step is characterized by the higher percentage of loss compared to the other steps (Δ ~ 20%)

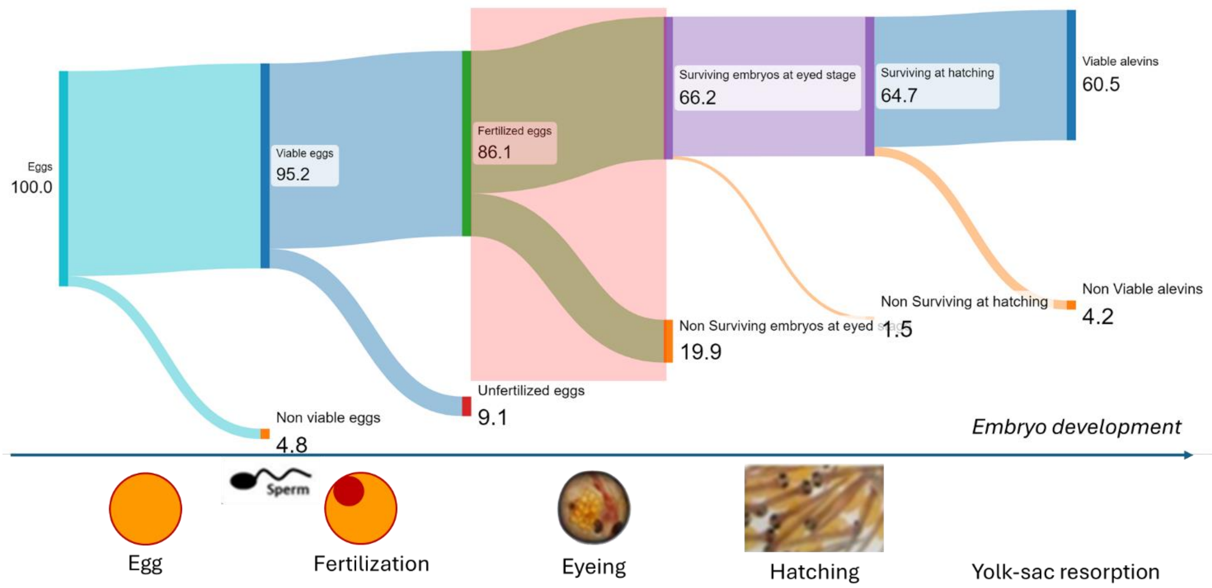

Figure 5.1. Representation of embryo development connected to food loss. Values are in %. Data from from Cardona et al 2021. Fish Physiol and Biochem 47, 671-679. https://doi.org/10.1007/s10695-020-00844-2 . In light red the phase of interest for our task.[1]

Fast and automatic identification and removal of whitish dead eggs in early developing trout embryos is needed to reduce microbial spread and avoid early food losses deriving from the whole batch discard.

In this context, the T3.4 of FOLOU project aims to develop and validate an automatic system based on multi-spectral cameras that will be used to precociously identify, remove, and count dead embryos. In a preliminary stage, the collected images are labelled by domain experts and then used to train a deep learning model. In order to have a real evaluation of the dead eggs during the incubation period an optic system able to precociously recognize and move the dead whitish eggs will be developed.

This device could improve up to 5% (but reaching 20% considering data from some publications) the number of eggs that pass the step 3 and achieve the eyed stage (less delicate), thus improving the productivity of the farming. This would lead to an increase, for example, in Italian rainbow trout production of 1723 tons (19%) and of 9190 tons at European level.

[1] Image modified from Cardona et al 2021. Fish Physiol and Biochem 47, 671-679. https://doi.org/10.1007/s10695-020-00844-2.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

5.2 Data acquisition

5.2.1 Material & equipment

- Good and bad just-fertilized eggs were obtained from an Italian supplier; eggs were immediately fixed in Formol and stored at 4 °C until use.
- Specific cameras (VIS and NIR) for images acquisition manufactured by XIMEA (VIS4x4 – NIR5x5)

- Plastic material to develop prototypes of vertical trays using a specific cochlea to gently lift the embryos
- Plastic beads of same weight and size of trout embryos to perform the preliminary tests.

5.2.2 Protocol

VIS-NIR acquisitions

For preliminary VIS and NIR acquisitions, good and bad eggs were obtained from an Italian fish farm and placed on a net in a single layer with or without water (Figure 5.2).

Figure 5.2. Good (orange) and bad (whitish) trout eggs.

Through specific cameras the acquisition based on 4x4 and 5x5 VIS-NIR mosaic sensors (**Error! Reference source not found.**) was performed and data were verified by human experts.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

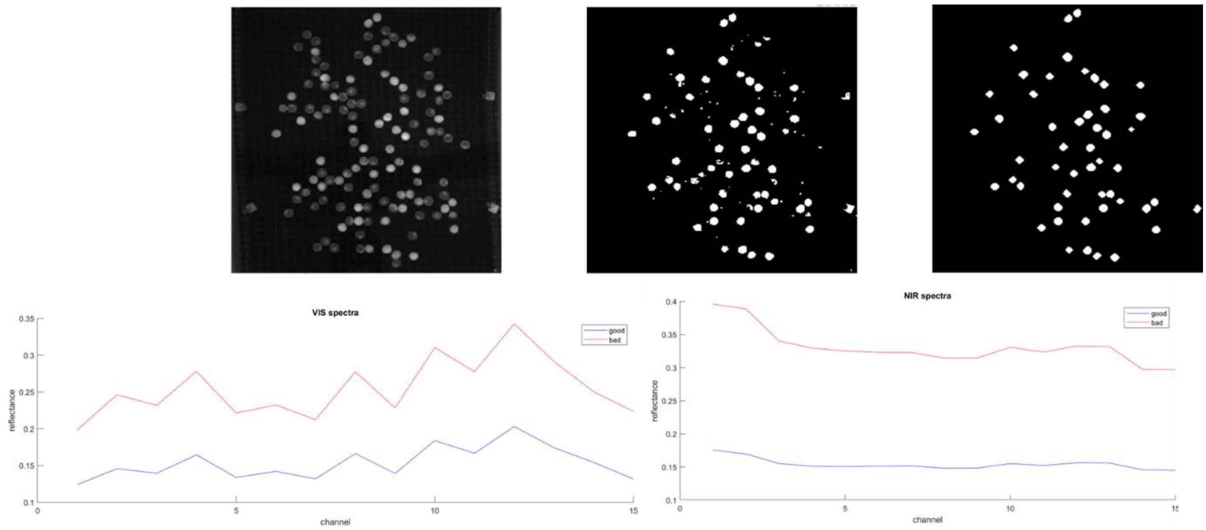

Figure 5.3. Example of VIS and NIR good and bad egg spectra. The top gives the image for three band waves, the bottom two figures the average reflectance over the different bands (left: visual camera; right: near-infrared camera) of good and bad eggs.

.

Incubation system

A first prototype of the incubation system was developed and tested and is represented in Figure .

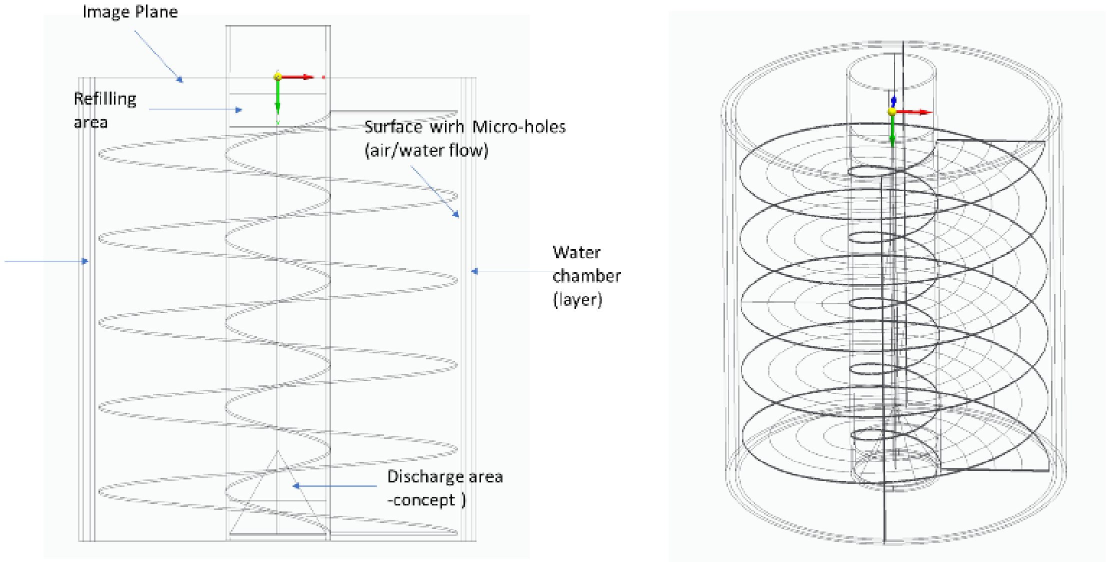

Figure 5.4. First prototype of hatching system under development.

However, this system was not particularly efficient in lifting the embryos and thus a second prototype was designed considering this problem. Specifically, it was evident from the first prototype that to efficiently lift the embryos a cochlea set up with a 45° angle was necessary.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Figure 5.5 shows the last prototype which was efficient in lifting the plastic beads and which is now under construction.

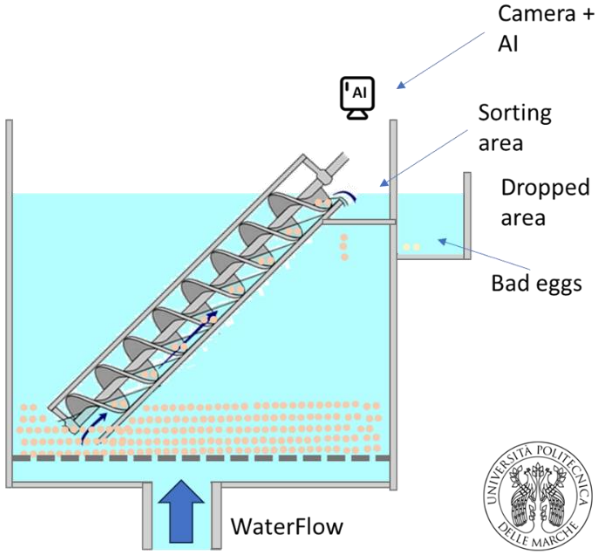

Figure 5.5. Second prototype under development

5.2.3 Outcomes of data acquisition

VIS-NIR Acquisition

Images are in tiff and png format 16 bit with multiple channels (15 VIS, 24 NIR). Results obtained from VIS/NIR acquisition were verified by trained human personnel and showed a proper sensitivity. Dataset of multi-spectral images of eggs were labelled over different classes (good, bad, uncertain).

At present, a more extensive set of images is not yet ready since the second prototype is under development.

5.2.4 Data preparation

Originally, each camera produces a single image, with the different bandwidths measured in different pixels. This needs to converted in multispectral image tiles, in which each image represents a single bandwidth. This pre-processing step is called de-bayering and requires a complex algorithm that tries to generate multi-spectral image form snapshot sensors applying calibration coefficients provided by the manufacturer of color filter array (IMEC, Belgium). We apply also white and black calibration to have reflectance values instead of raw digital number.

Images have been augmented using different data augmentation algorithms as: rotation, flip (vertical / horizontal), blurring.

The main reason to use VIS-NIR camera was to evaluate the performance in the detection using VIS-NIR vs classical RGB or RGB with a modified set of filters.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

5.3 Data annotation

A preliminary set of images have been labelled from experts to train an object detection algorithm to detect eggs also providing a class (bad, good, uncertain).

We used the labelbox platform. Figure 5.6 shows an example of labelling on acquired images.

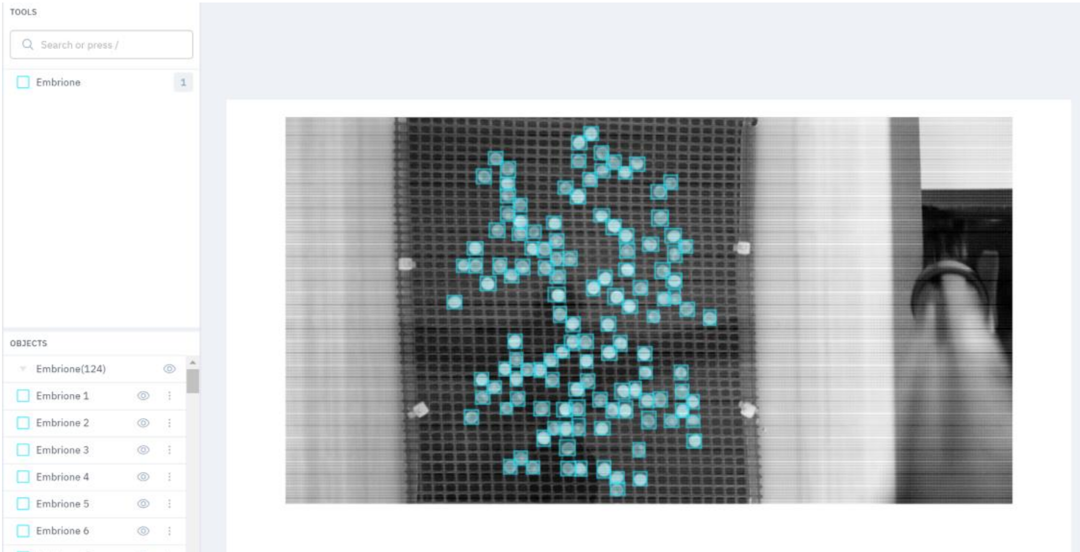

Figure 5.6. Data labelling through a web-based tool (Labelbox); the labelling was performed by experts.

5.4 The proposed method

The expected outcome is the development of a new system able to gently lift trout early developing embryos equipped with a set of cameras and specific tools able to recognize and fastly remove non-developing embryos.

The deep learning approach relies on detection algorithms. We started by training a model based on You only look once (YOLO) algorithms; we used v5 and v8; augmentation was also used to increase the number of samples (rotation and flip). Models have been training using pre-trained models on State-of-the-Art dataset using transfer learning techniques.

The overall welfare of egg batches incubated in the new incubation system will be compared to those incubated in standard systems.

The models will be enriched using a wider set of images. Bad eggs will be removed using a mechanical system that will route good and bad eggs to dedicated collection areas. This step will be evaluated in the second prototype.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

5.5 Expected outcomes

The expected outcome is the development of a new system powered by artificial intelligence approaches able to gently lift trout early developing embryos equipped with a set of cameras and specific tools able to recognize and quickly remove non-developing embryos. Non-developing embryos are also a potential site for bacterial infections, from which also healthy eggs will be eventually infected. Hence, it is an important step in preventing production losses.

The overall welfare of egg batched incubated in the new incubation system will be compared to those incubated in standard systems.

5.6 Development state

Main Tasks	Subtasks	State
Data acquisition	Good and bad egg samples collection	Done
	Multispectral data Acquisition of good and bad eggs	Done
Data processing	Data preprocessing	Done
	Data Annotation	In progress
	Data Augmentation	In progress
Development of object detection algorithms	Prototyping	1st prototype developed, 2nd one under development
	Testing the model	Not yet started
Development of food loss estimation algorithm	Prototyping	Not yet started

Table 5.1. Overview of the progress of Task 3.4.

5.7 Key Successes and Challenges

Key Successes:

- Data collected with different cameras (RGB + multi-spectral) completed 1st run
- Preprocessing: completed (1st run)
- Annotation ongoing (done 1st run)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

- Prototyping in progress

Challenges:

- Different design to properly manage the embryos avoiding mechanical stresses.
- Preprocessing and annotating our large dataset (required expert)
- Robustness to environmental conditions of embryos classification algorithm (e.g. light condition)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

6. Blockchain technology implementation in mussel aquaculture

Commodity: mussels

Food loss category: Production losses | Pre-harvest losses | Losses during Harvest

Technology: Blockchain methodology for tracking food losses along the production chain

Current stage: Under development: mussel datasets collected; blockchain methodology under development

Desired outcome: Provide a decentralized digital platform based on blockchain technology capable of ensuring the non-repudiation and non-alterability of data regarding food loss.

Aquaculture farming in the EU yielded an estimated 1.1 million tons of aquatic organisms in 2021, corresponding to one quarter of the output of European fisheries as a whole. Particularly, mussels (Mytilus sp.) aquaculture reached alone about 40% of the total production in EU with Italy being the second European producer (Figure 6.1).

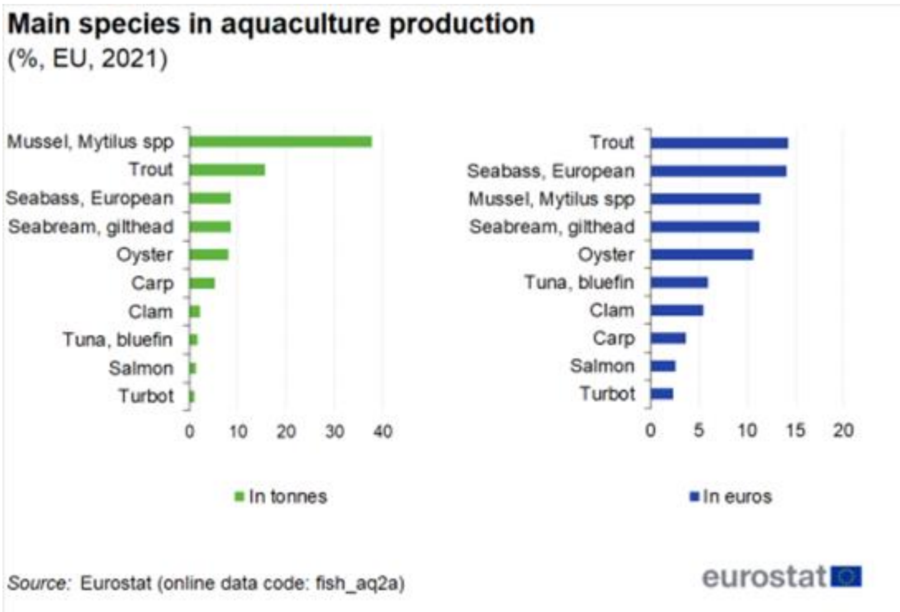

Figure 6.1. Main species in EU aquaculture production (%)[2]

[2] Source: EUROSTATS. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Aquaculture_statistics#EU_Aquaculture

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

6.1 Overview of the challenge

To carry out an effective analysis and reduction of food loss phenomena along any food chain, it is essential to have detailed and reliable data. Of crucial importance is the non-repudiation and non-alterability over time of the collected data , which require the use of appropriate technological tools. Blockchain technology provides a particularly suitable tool for this purpose, providing an unmodifiable decentralized digital platform that can be used as a pillar for the design and implementation of technological methods and solutions capable of ensuring the non-repudiation and non-alterability of data regarding food loss.

The reduction of food losses in the first stages of the food production chain can be supported by the application of the cold chain, where the reference legislation for food products of animal origin is the Regulation (EC) No 853/2004. The regulation applies to the production and distribution of food of animal origin intended for human consumption.

Blockchain technology, and in particular public blockchain infrastructures such as Ethereum, offer an unchangeable and persistent digital infrastructure to which the function of ensuring the integrity and immutability over time of collected data can be delegated. To take advantage of these features, however, it is necessary to set up a technological infrastructure to support the collection of data, its organization and, most importantly, its blockchain-based certification and the corresponding verification of certified data.

To this end, a preliminary analysis was carried out to identify the most suitable technological solutions for application in the context of interest and for the intended use cases. The analysis performed has led to the definition of an architecture of the type schematically described in the Figure 6.2.

The productive cycle of mussels can be affected by several factors that can represent potential sources of food loss. The acquisition of data from different variables during mussel farming can represent a suitable solution to provide data to monitor the productive cycle and possibly prevent and minimize food loss. However, this process is often endangered by the integrity of data themselves. Indeed, it can happen that data are either modified or purposely deleted. Or, analogously, it can happen that important pieces of information get lost because of malfunctioning of the used technology. For this reason, there is a crucial need for an efficient and secure way to store and certify data, aiming for data integrity and availability. Blockchain technology (or more in general, distributed ledgers technology) represents the perfect technological means to face this challenge. Blockchain technology, indeed, is a distributed ledger of records. The distributed nature of the blockchain technology guarantees data integrity since data, once inserted, cannot be modified or eliminated from it; blockchain, in fact, is defined as immutable. Moreover, the absence of single points of failure guarantees that data are always available.

To validate this, the aim is to propose a blockchain-based platform to certify the data obtained during the analysis of a case study (mussel farm located in Senigallia, Marche, IT).

Mussel (*Mytilus galloprovincialis*) aquaculture in the Adriatic Sea is environmental-friendly and does not significantly cause alterations in the marine ecosystem, both as functioning and trophic state. It represents a sustainable and relatively cost-effective way to produce seafood. Particularly, the rearing of Mytilus galloprovincialis is always extensive. The young mussels are collected from the sea and can be cultured on suspended ropes; these ropes, which are covered with mussel seeds kept in place by nylon nets, are suspended from

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

longlines of floating plastic buoys. Here, mussels increase in size until they reach the marketable size, feeding on the fine particulate organic matter (POM), including phytoplankton, bacteria, and organic detritus, from suspension in the water column around them. This form of rearing represents a great advantage in terms of cost for the farmers compared to finfish aquaculture but can be characterized by different drawbacks. Particularly, mussels are reared in the sea and can, thus, be exposed to: (i) eventual alterations of the water quality; (ii) adverse meteorological and marine conditions; (iii) presence of parasites or predators; (iv) they are sold alive and must be properly stored. In addition, during the productive cycle, mussels are not regularly inspected, but are periodically subjected to declumping and thinning procedures in relation to their growth rate, an activity that can cause damage to the shell.

In addition, from the cold chain point of view, the temperature of storage and therefore of transport during the production phase are very relevant and it should be kept in the desired temperature range. The TOR (Time Out of Refrigeration), the maximum time that the product can spend out of the range of temperatures, could be a relevant parameter to be certified in the blockchain framework. The regulation (853/2004/CE) states: "Fishery products kept alive must be kept at a temperature and under conditions which do not affect food safety or their viability." To be sold alive, during transport, mussels need to be stored at a temperature that prevents their death, between +4 and +6°C. For this reason, the mussels must be exposed and stored in refrigerators other than those of fish, which must be stored at a temperature around 0°C, from which it must be kept separated also to avoid bacterial contamination (cross contamination)3. Figure 6.3 shows the cold chain framework for shellfish.

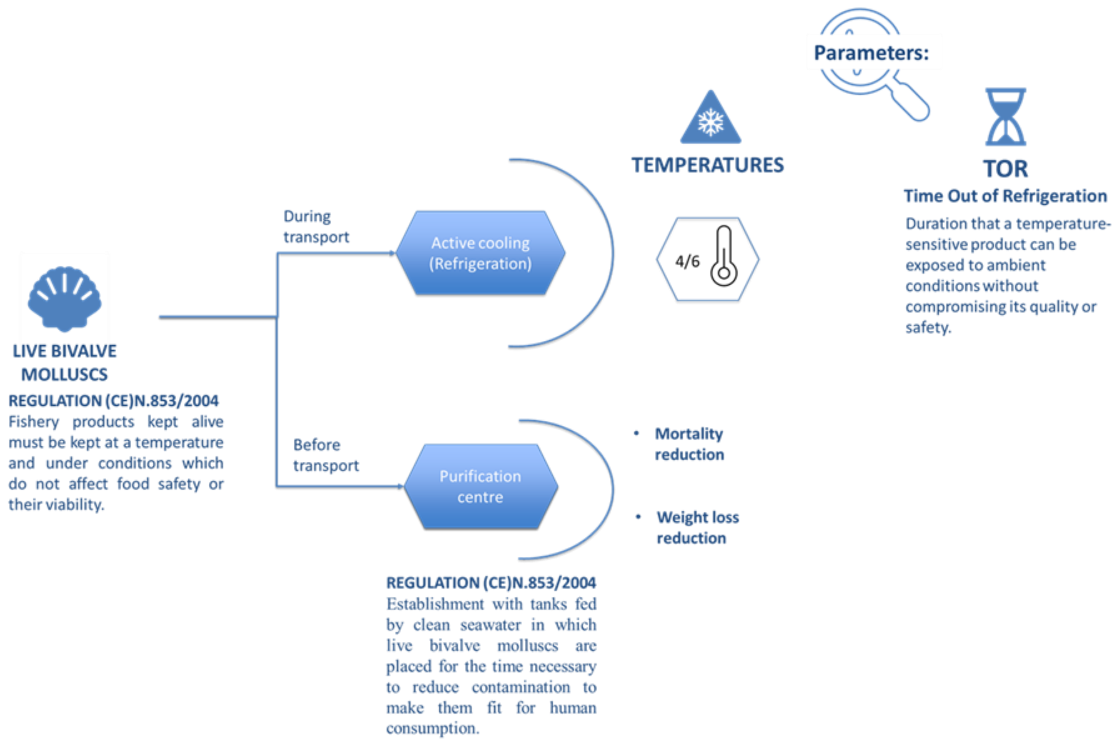

Figure 6.3. Scheme of the methodology for control of the cold chain for mussels.

[3] Chemical, Gli obblighi dell'operatore per i Molluschi Bivalvi Vivi. (Available online: Gli obblighi dell'operatore per i Molluschi Bivalvi Vivi – Chemichal), access on: 15 December 2023

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

6.2 Data acquisition

6.2.1 Material & equipment

- Agreement with company (Sena Gallica Soc. Cooperativa, Senigallia, IT)
- Identification of 3 sampling sites within the mussel farm that were sampled in 4 key-times during the productive cycle, with emphasis on the market size.
- Mussels' samples for laboratory analysis including total weight, single animal weight and length, edible part %, presence of parasites, dead specimens vs alive ones
- Prototype: an IoT device (currently under development) to measure the growing rate

A blockchain-based platform used to certify data. Special focus on the certification of the different processes involved in the use-case related to mussel aquaculture.

Sampling sites within the farm were named Small, Medium, and Big according to the size of the seeds at the beginning of the productive cycle (Figure 6.4). The Small site was the closest to the coastline.

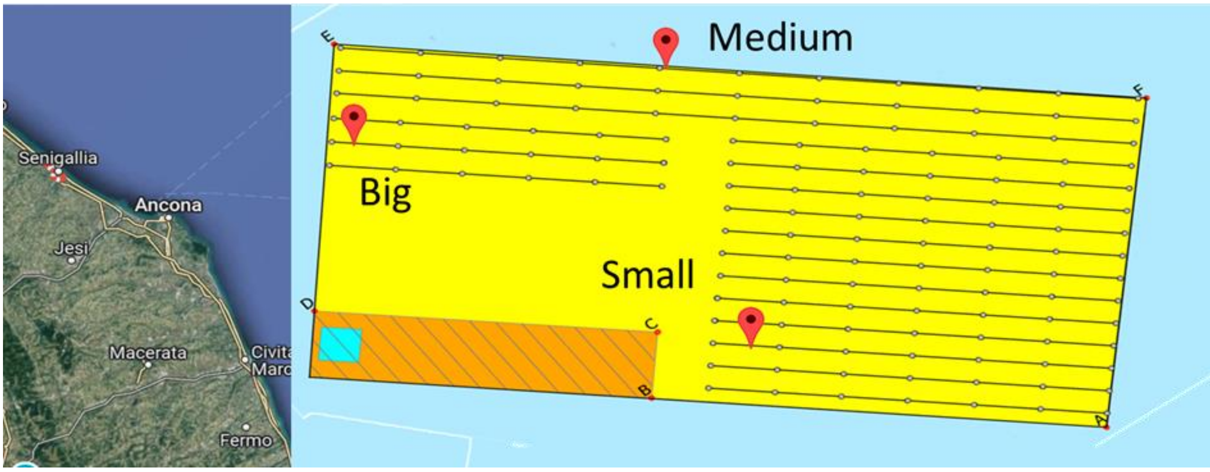

Figure 6.4. Sampling sites within the farm named Small, Medium, and Big according to the size of the seeds at the beginning of the productive cycle. Cyan and orange areas are out of the scope of sampling protocol due to a different management). Yellow color represents the useful area for our task.These sites were sampled in different times of the productive cycle (from pre-harvest to harvest) to determine the mortality and the growth parameters of mussels. Particularly, 3 one-meter samples from each site were obtained at (i) Settlement (July 2023; pre-harvest); (ii) Intermediate sampling (November 2023; pre-harvest); (iii) Market Size (February 2024; Harvest 1 – only Medium and Big sites); Product for summer season (April 2024; Harvest 2 – only Small site).

Harvest 1 - From each one-meter sample (Figure 6.5) collected in the Small, Medium, and Big sites, alive and dead specimens were counted and weighted in light to measure the production in terms of alive biomass per sample and to obtain the mortality rate (%). In addition, on 20 specimens per one-meter sample (60 specimens per site at each sampling time), individual parameters were measured: length, total wet weight, wet weight of the edible part, dry weight of the edible part, shell thickness.

In addition, the presence of mussel predators (flatworms of the class Turbellaria; Figure 6.6) was assessed and the number of alive specimens per one-meter sample was recorded.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

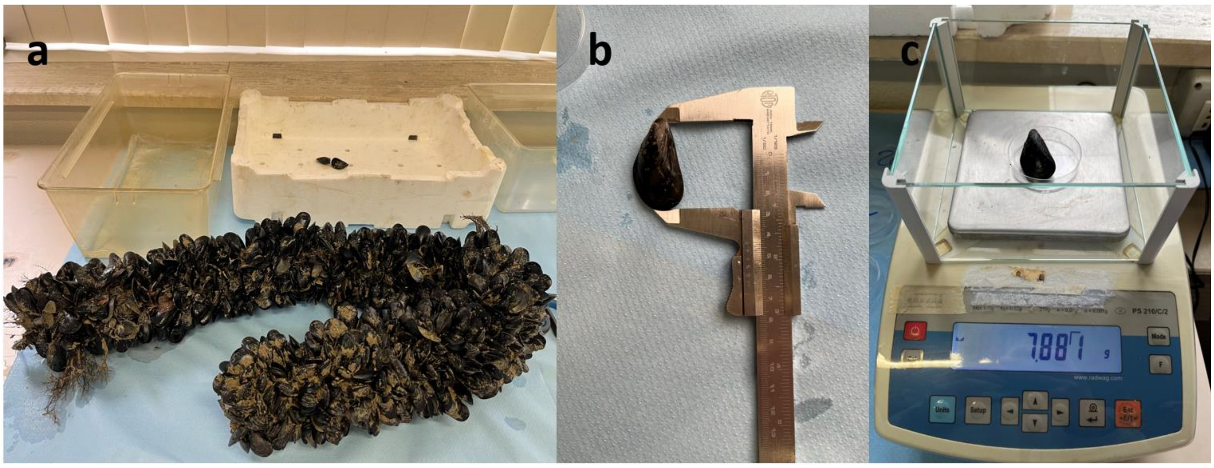

Figure 6.5. Details of measurements on mussels. (a) example of a one-meter sample; (b) individual length; (c) individual total wet weight

Figure 6.6. Example of flatworms found alive within mussels' sample

Harvest 2. Differently from mussels from Harvest 1 that are sold as they are since the shell does not have a proper thickness for the cleaning procedures, those from Harvest 2, can be cleaned due to proper shell thickness, a procedure that can possibly add another source of food loss. For that reason, the whole socks were weighted as they are and then were cleaned resulting in the separation of: (i) cleaned sellable product at market size; (ii) dead mussels at market size; (iii) dead smaller mussels and fouling; (iv) mussels' seed (recovered by the farmers as a stock for a new production); (v) sand; (vi) plastic nets. The cleaned sellable product was analysed following the same protocol reported for Harvest 1. In addition, the number of mussels broken by the cleaning procedure (Figure 6.7) was measured.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Figure 6.7. Examples of damaged mussels after the cleaning procedure at Harvest 2

Data obtained from the samplings will be coupled with data obtained from the: (i) normal operations of farm maintenance; (ii) harvesting of the final product; (iii) transport conditions (i.e. cold chain, packaging); (iv) meteorological and marine data obtained by the station located near the farm; (v) the IoT device for measuring the growing rate; (vi) installation of a tracking device to monitor the vessel position and temperature.

6.2.2 Outcome of data acquisition

- Weight and number of alive and dead mussels at each sampling site during the productive cycle to meet the project goals; emphasis was given to the food losses at harvest (commercial size; Harvest 1 and Harvest 2).
- Number of mussels broken by the cleaning procedure to compare the two harvesting procedures (Harvest 1 and Harvest 2). In terms of food loss.
- Weather data collected by data buoy managed by National Council of Research (CNR)
- Blockchain to guarantee integrity and certification of acquired data.

6.2.3 Data preparation

Data of measurements conducted on mussels' samples were analyzed to remove eventual outliers. Results were then expressed as mean ± standard deviation.

6.3 Data annotation

- Data on total productivity of the farm: excel file provided by the farmers

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

- Data of measurements conducted on mussels' samples: excel file provided by the UNIVPM team
- Blockchain data certification and verification

6.4The proposed method

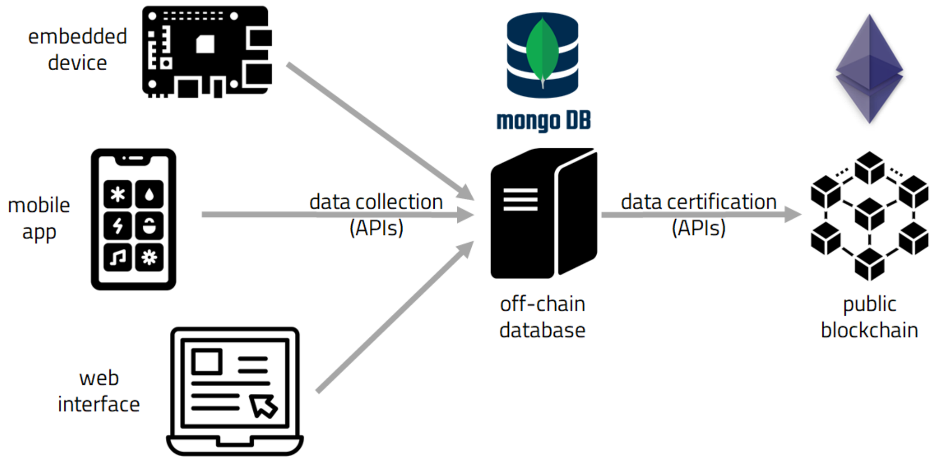

Figure 6.2. Scheme of the proposed blockchain-based architecture for data certification.

The designed architecture is based on a central component, consisting of an off-chain server equipped with storage capacity for the data to be collected and certified, as well as capacity to execute the cryptographic functions required for their certification. Such a central server is expected to perform the following core functions:

- Expose API (Application Program Interface) functions to enable the collection of data from several possible sources. It is expected that, for example, such sources may be represented by web interfaces (usable from computers or via smartphone apps) for manual entry, or cyber-physical devices for automated collection.
- Perform cryptographic functions to compute certification data derived from the data itself, which enable certification and integrity of collected data to be ensured, without including the data itself.
- Interface with a public blockchain (such as Ethereum) to perform writing of certification data to ensure its persistence and immutability, without disclosing the collected data.

Once the collection and blockchain-based certification of data has been carried out, anyone who comes into possession of the certified data and the related certification information can verify its correctness directly. This can be done by querying the central server again, or in an entirely decentralized manner, via functions executed locally.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Data are also collected through a data buoy managed and maintained by the National Council of Research (CNR) – IRBIM4.

The data collected during the mussel aquaculture case study will be used to test the blockchain-based platform for certifying data. The aim, in fact, is to address the problem of certification and data traceability by identifying an architecture that can exploit the inherent advantages of distributed ledger technologies.

In the proposed architecture, data to be certified can come from different sources, such as embedded devices, mobile apps, or web interfaces, as previously shown in Figure 6.2. These data of different types (i.e., strings, files, pictures, etc.) can be uploaded by users on a dedicated database through some ad-hoc application. Then, some processing needs to be done between the database, containing the information to be certified, and the public blockchain (i.e., Ethereum), to immutably notarize data.

To this aim, an innovative solution based on the use of Merkle trees for organizing pieces of information is proposed. The primary goal of this approach is to limit the number of blockchain transactions to be generated, making the certification process more efficient and cost-effective. The idea is to independently consider data coming from different sources and organize them into Merkle trees, whose numerosity (i.e., the number of leaves) is a design choice that considers various factors aimed at minimizing costs. Merkle trees are data compressing structures (see Figure 6.8).

Exploiting this structure, the data to be certified are compressed using a hashing algorithm (i.e., SHA256) and then placed in a leaf. The tree is then built from the leaves to the root by applying the SHA256 compression to the concatenation of each pair of successive leaves. Then the upper nodes, obtained from this compression step, are again compressed in the same way. This process continues until the root (i.e., merkle root) of the tree is reached. The resulting hash digest is the compressed representation of all the leaves and will therefore be the only data stored on the blockchain.

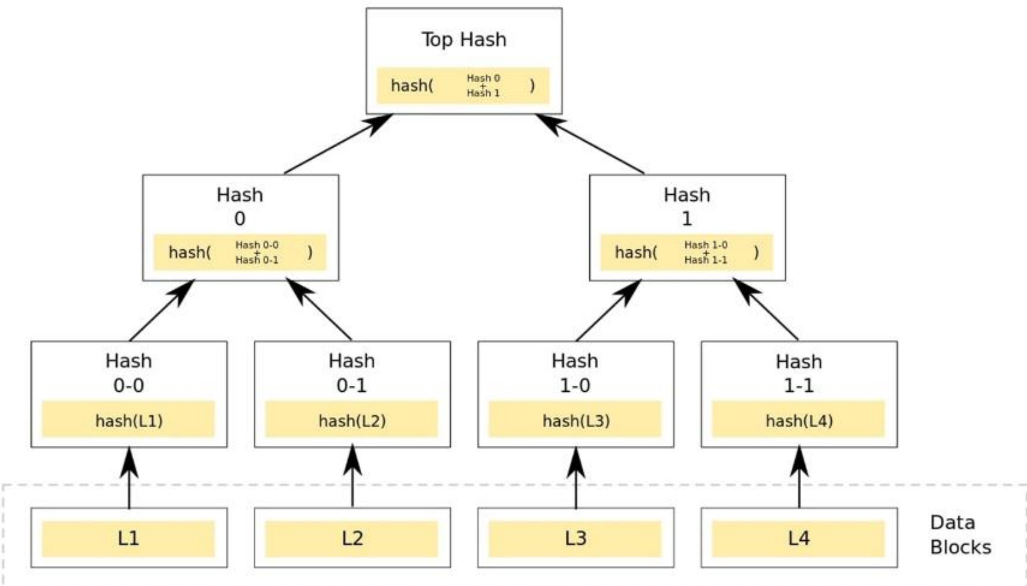

Figure 6.8. Merkle tree structure.

[4] http://rmm.an.irbim.cnr.it/index.php/meda-senigallia (last access March 2024)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Since compression is not invertible, to verify that one piece of information is not modified after the certification, a set of node values needs to be extracted from the tree to test the integrity of the document. Through the so called Merkle Proof, indeed, it is possible to demonstrate that data in a leaf have not been modified.

The certification process is summarized in Figure 6.9. Initially, data that need to be certified are grouped together in a waiting queue and, after a pre-defined amount of time, the certification process begins. Basically, these grouped data are used as leaves to build the Merkle Tree and the Merkle root is computed. While the tree is stored in an off-chain database, the Merkle root is sent to the blockchain through a transaction. Transaction information is stored in the off-chain database as well, together with all the information that allows the verification process (e.g., the Merkle proof).

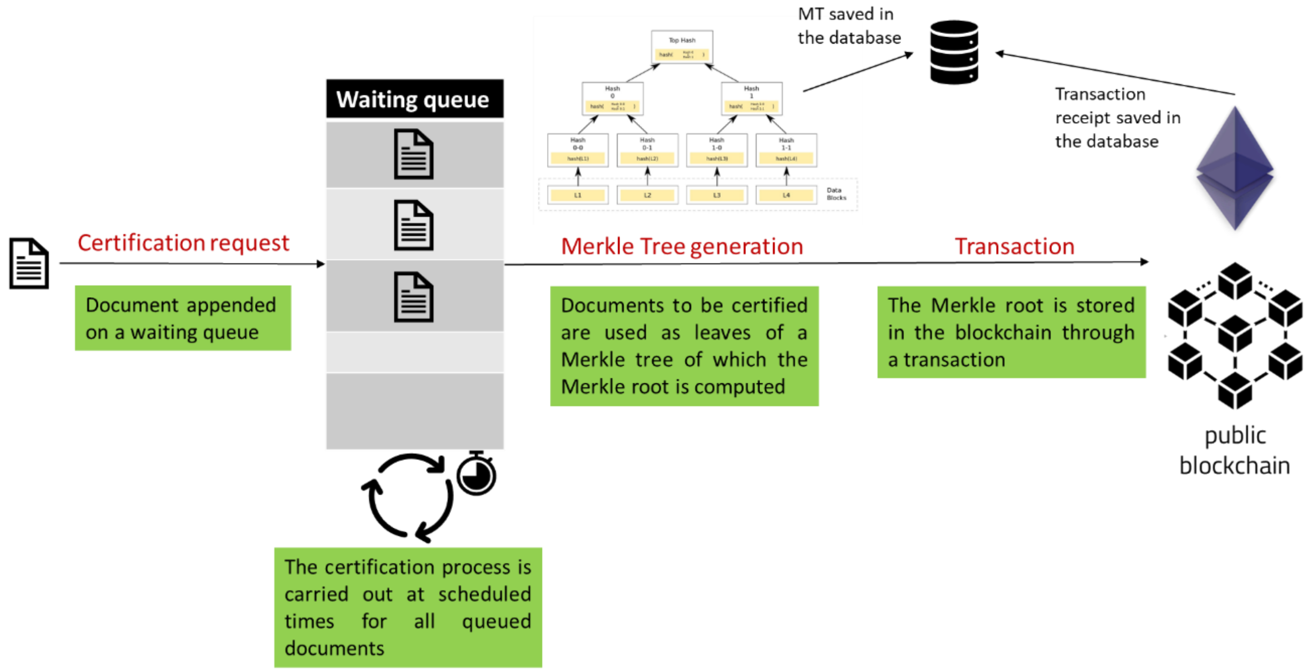

Figure 6.9. Scheme of the certification process.

When a user needs to verify a document, the proof is extracted from the off-chain database. Then, the transaction that contains the in-chain root needs to be located in order to start the actual verification function, which includes the following operations:

1. calculate the hash of the file,
2. extract the root from the blockchain,
3. locally calculate the root through the proof and the hash found in step 1),
4. compare the values obtained in steps 2) and 3); in case of equality, the verification is successful.

The user will have the possibility to upload data and to verify them through an easy interface that will show, for example, if data are in queue or certified, or if the considered data has maintained their integrity or not.

The proposed blockchain-based architecture gives some useful properties:

- Certification via blockchain:
    - data are certified via blockchain transactions
    - Data integrity depends on the security of the employed blockchain
    - Data can be easily and securely verified (thanks to fast cryptographic functions)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

o When an IoT Device is used, human actions are not required

- Off-chain database (e.g., MongoDB):
    - o Heterogeneous data (several formats and sources)
    - o Database is responsible for data maintenance
    - o Different solutions for data certification (single vs multiple data)
- User-friendly interface:
    - o Updating data can be done in a few steps
    - o Verifying data integrity can be done is a few steps

The proposed blockchain-based architecture will be validated using the mussels aquaculture case study data. An example of a simple prototype for this application is shown in Figure 6.10.

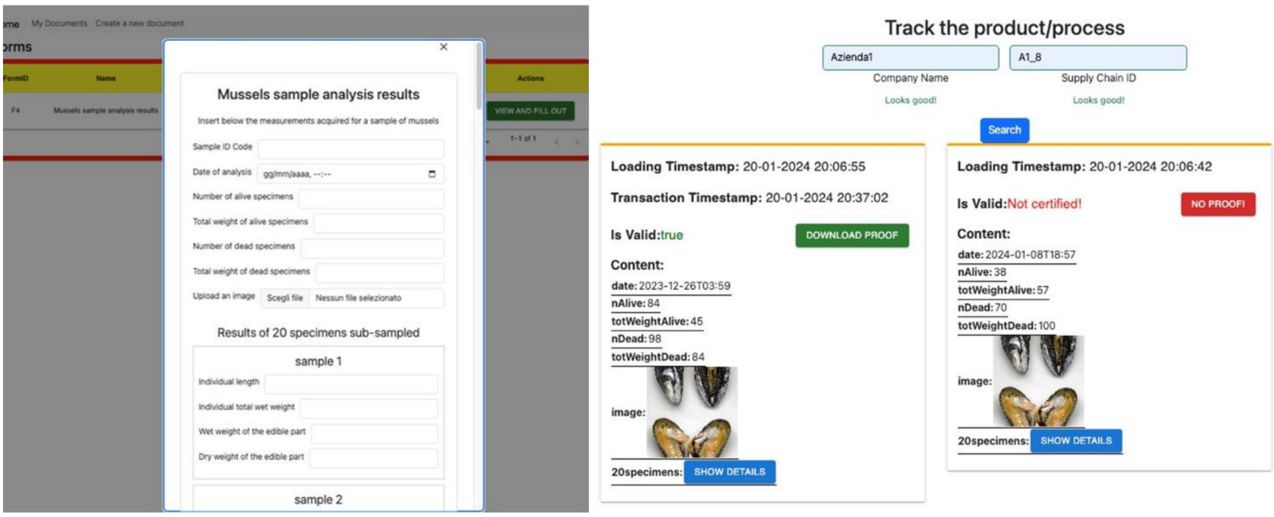

Figure 6.10. Left: Example of data input interface. Right: Example of data certification process.

6.5 Expected outcomes

The expected outcome is the development of a blockchain-based platform to certify data and to guarantee their integrity. This application can be used to efficiently store data related to food processing and, eventually, food losses. The platform will be tailored to the case study on mussel aquaculture. However, the method and blockchain architecture can be potentially applied to any commodity.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

6.6 Development state

Main Tasks	Subtasks	State
Field campaign	Data Acquisition on mussels production during a whole productive cycle	Done
	Data Acquisition on mussels pre harvest, harvest food losses	Done
	Data field analysis	Done
Data processing	Data preprocessing on mussels	Done
	Cryptographic certification algorithm design	Done
	Data Annotation on mussels	In progress
	Data Augmentation	In progress
Development of the blockchain-based platform for data certification	Prototyping	Under development
	Testing	Not yet started
Development of food loss estimation algorithm	...	.....

Table 6.1. Overview of the progress of Task 3.5.

6.7 Key Successes and Challenges

Key Successes:

- Data collected over 1 aquaculture field over 1 season
- Preprocessing: completed
- Data Analysis: completed
- Blockchain platform: in progress

Challenges:

- Acquisition of samples and their processing takes time.
- Reduction of Blockchain and (big) data certification complexity
- Increase of temporal resolution

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

7. Market demand tools from social networks (T3.6, CIRCE)

7.1 Overview of the challenge

Commodity: : Grain and pulses (pasta), fruits and vegetables (tomato), root tubers and oil crops (olive oil), meat and animal derived products (pork), and fish(hake).

Food loss category: Surplus: rough forecasts of food demand limit the efficiency of the supply chain.

Technology: Natural language processing for messages from social networks, and machine learning to model actual consumption based on text features.

Current stage: Under development: data acquired for food consumption, data annotation and modeling in progress.

Desired outcome: Models to predict food consumption based on messages from social networks.

Predicting customer behavior is crucial for reducing food loss in production, retail and hospitality. Advanced analytics, including natural language processing (NLP) and machine learning (AI), can estimate food consumption based on social media messages. This approach helps adapt the production, optimize inventory, reducing overstocking and minimizing waste. The development analyzes five Spanish commodities: grains and pulses, fruits and vegetables, root tubers and oil crops, meat and animal products, and fish. Different input features and algorithms are studied to improve the accuracy of consumption models. By understanding consumption patterns, businesses can better align production and supply with demand, enhancing product availability, reducing spoilage, and improving operational efficiency. Moreover, the demand forecasts can be compared to supply information to identify potential food surplus and quantify food loss.

7.2 Data acquisition

Information is collected for the case of Spain between 2018 and 2023, from web data sources with open access via web. Food products (pasta, tomato, olive oil, pork and hake) were selected due to its importance in Spanish consumption.

7.2.1 Material & equipment

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Social messages: Messages from the social network X (formerly Twitter) are collected from the Internet Archive. This digital library provides a sample of all the messages published worldwide between 2011 and 2023. Text messages and their creation timestamp are grouped in a file structure of months, days, hours and minutes.

Queries to search engine: Apart from social networks, popularity is studied based on queries to the search engine Google. Google Trends computes a quantitative variable of search interest in time series format, between 2004 and nowadays.

Food consumption: Domestic food consumption in Spain was acquired from the Spanish Ministry of Agriculture, Fisheries and Food (MAPA). This government department provides yearly and monthly information from 1990 to 2023, characterized by food, region, socio-demographic profile and sales channel. Besides MAPA, complementary data was acquired from other free official entities: Spanish National Statistics Institute (INE), European Food Safety Authority (EFSA) and Food and Agriculture Organization of the United Nations (FAO).

7.2.2 Protocol

All the data sources are open access via web, and data for the time period between 2018 and 2023 is manually downloaded on local servers for further processing. Social messages are stored in up to 10^5 compressed JSON files, which are automatically decompressed by a Python script. Search interest was exported in a single CSV file per food, while data about food consumption was saved in a XLSX file per year.

7.2.3 Outcomes of data acquisition

The data acquired from each data source is summarized in

Table 7.1.

Source	Data type	Size of collected dataset
X	Social messages	17 TB of decompressed files (3500 million of messages)
Google Trends	Search interest	72 samples per food
MAPA	Food consumption	72 samples per food

Table 7.1. Summary of the data collected from each source.

7.2.4 Data preparation

The dataset of social messages includes texts for additional topics other than food, written in multiple languages. Thus, relevant messages are filtered using Spanish hashtags

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

for the considered foods. Data about food consumption was manually selected to handle inconsistencies in food names and categories.

The consumption datasets of the four food entities (MAPA, INE, EFSA and FAO) were organized creating a relational database, promoting its normalization, elimination of redundancy and prevention of cyclic dependencies. Variables were encoded, date formats were parsed, measurement units were normalized, and additional features and implicit information were extracted.

7.3 Data annotation

Social messages, search interest and food consumption are labelled with their corresponding timestamp, enabling their comparison. For example, Figure 7.1 shows the consumption and search interest for tomato in Spain, between 2021 and 2023. A model of sentiment analysis evaluated for X (XLM-T) is applied, providing three emotion scores (positive, neutral and negative) for each user text.

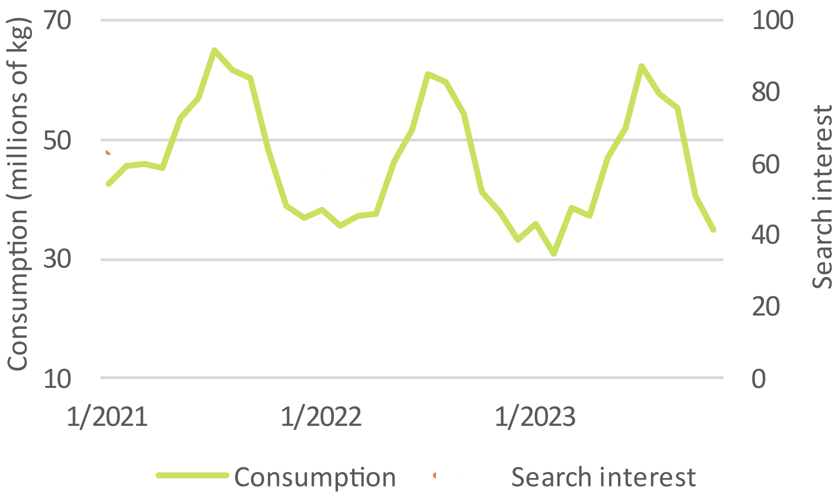

Figure 7.1. Consumption and search interest for tomato in Spain, between 2021 and 2023.

7.4The proposed method

To match the granularity of search interest and food consumption, monthly features will be extracted from the social messages, including text number and average emotion. Official food news will be analyzed to study their relationship with social messages and search interest. Food consumption will be studied with time series methods, such as moving average and autoregressive integrated moving average. Machine learning algorithms will be used to predict food consumption based on social messages and search interest. The accuracy of the models will be evaluated for training and test subsets, analyzing different

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

social features and learning algorithms, as support vector machines and multilayer perceptron.

7.5 Expected outcomes

This development will provide models to estimate consumption based on social networks for five foods, using Spanish data between 2018 and 2023. With that purpose, techniques of natural language processing and machine learning are applied.

Conventional techniques to monitor food consumption provide measures with delay, which can reach several months in some cases. By employing dynamic data generated in social networks, the developed models can be used to provide an earlier estimation of food consumption and demand. Moreover, the demand forecasts can be compared to supply information to identify food surplus and quantify food loss.

7.6 Development stages

Main Tasks	Subtasks	State
Data acquisition	Social messages	In progress
	Search interest	Done
	Food consumption	Done
Data annotation	Sentiment analysis of messages	In progress
Data modelling	Monthly characterization of messages	Not started
	Time series analysis of food consumption	Not started
	Estimation of food consumption based on social data	In progress

Table 7.2. Overview of the progress of Task 3.6.

7.7Key successes and challenges

To conclude, the key successes and challenges of the activities performed until June 2024 are presented below.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Key successes:

- End of data acquisition for search interest and food consumption
- Start of data acquisition and processing for social messages
- Start of consumption modeling based on social data

Challenges:

- Policy changes of X and the limitation of alternatives to collect social messages
- High volume of social messages to access and filter
- Lack of standardization of consumption data

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Conclusion

The current report offers a technical overview of the diverse methodologies and approaches currently being implemented and developed in Work Package 3 of the FOLOU project. This work package has significantly advanced the handling of critical food loss challenges within various agricultural commodities. Innovations include the deployment of high-resolution UAVs and multispectral imaging coupled with AI for precise food loss estimation, crop modelling for production loss, the application of blockchain technology for reliable tracking of food loss throughout the supply chain, and the analysis of consumer food trends on social media to better predict food production needs. all methods are presently under development and have demonstrated promising progress by successfully confirming the first key milestones. Next step for most the tasks will be finishing the annotation, analyzing the data, and successfully implement the first protypes of the models.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.