Data Challenge

Kais Kharrat
Adam Saoud
Guillaume Vindry
David Sergio Arrustico Villanueva

Executive Summary

A **5-day data challenge** utilizing **satellite imagery (Sentinel-2)**, **NDVI**, and **UNet models** to **predict crop types,** offering insurers a **SaaS solution for fraud prevention, faster claims, and better risk assessment**.

- **5-Day Data Challenge:** Analyzed time-series satellite images to predict crop types and parcel

- **Data Cleaning using NDVI:** Cleaned and normalized satellite data for consistent analysis using the NDVI vegetation index.

- **UNet Model :** Used UNet for precise segmentation and classification of agricultural parcels.

- **Business Solution :** SaaS tool automating fraud detection, speeding claims, improving risk assessment, and providing ongoing client support.

Data quality issues

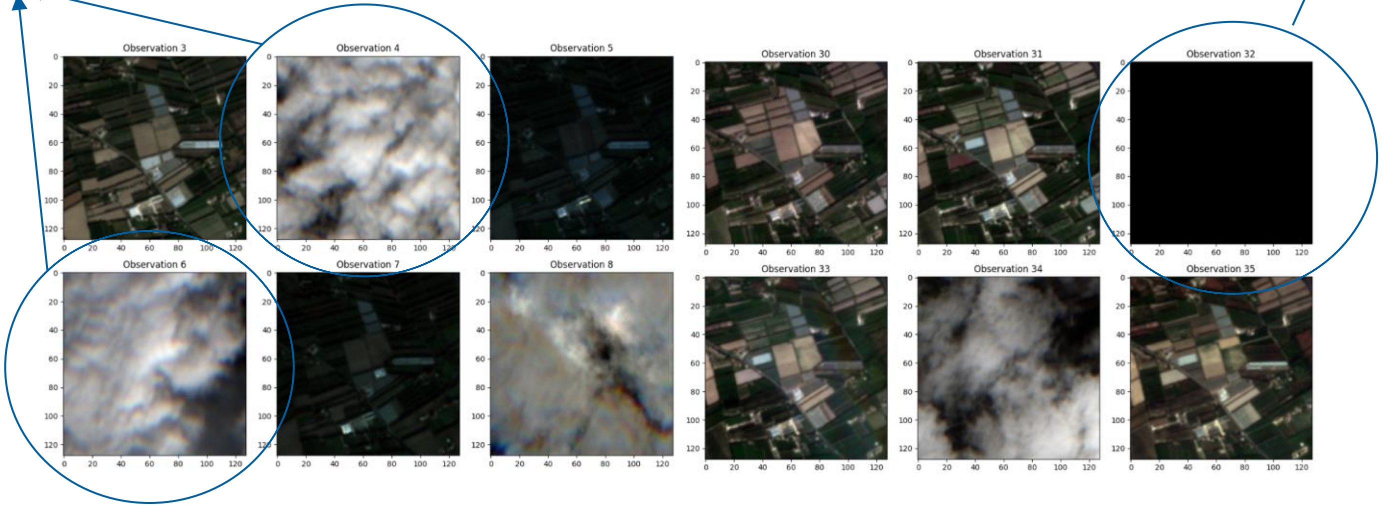
2 main types of problems

Cloud images

- ➔ Disturbance of visual image quality
- ➔ Make image segmentation difficult

Completely black images

- ➔ No relevant information



NDVI (Normalized Difference Vegetation Index)

An index widely used to measure vegetation health from satellite images

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}}$$

-Higher Values indicate healthier and denser vegetation

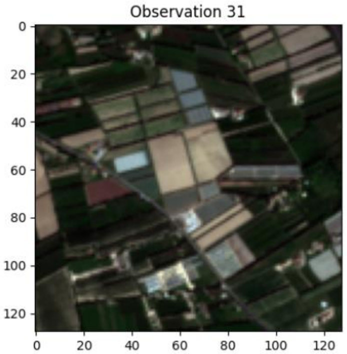
-Lower values indicate less dense vegetation or bare ground

NDVI values range from -1 to 1

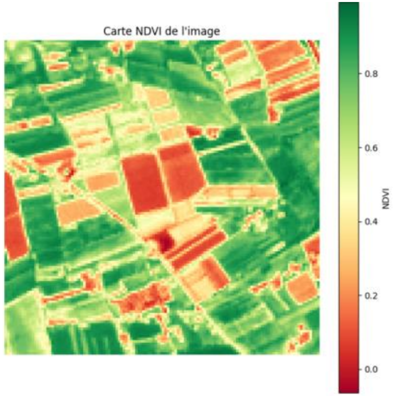
⇒ Delete images whose mean NDVI is below a threshold

NDVI (Normalized Difference Vegetation Index)

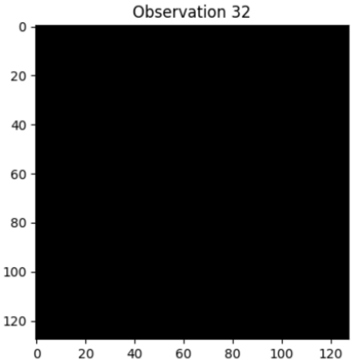
Image with clearly visible vegetation



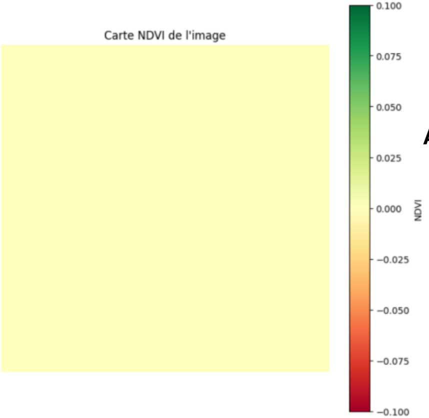
Average NDVI = 0.6



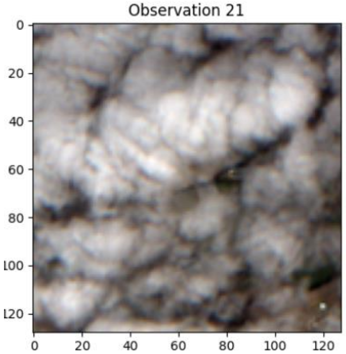
Black image



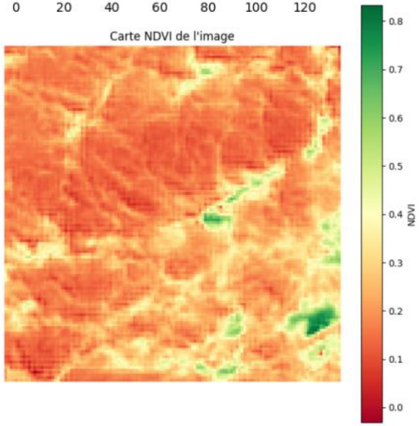
Average NDVI = 0



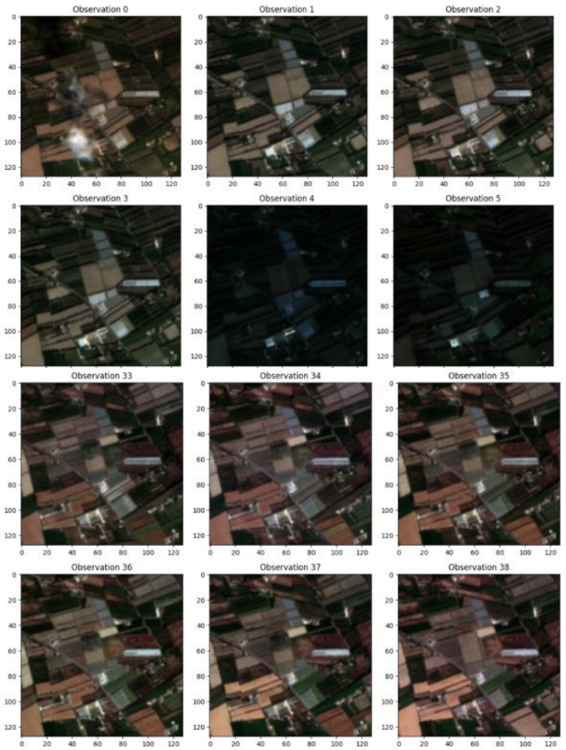
Cloud image

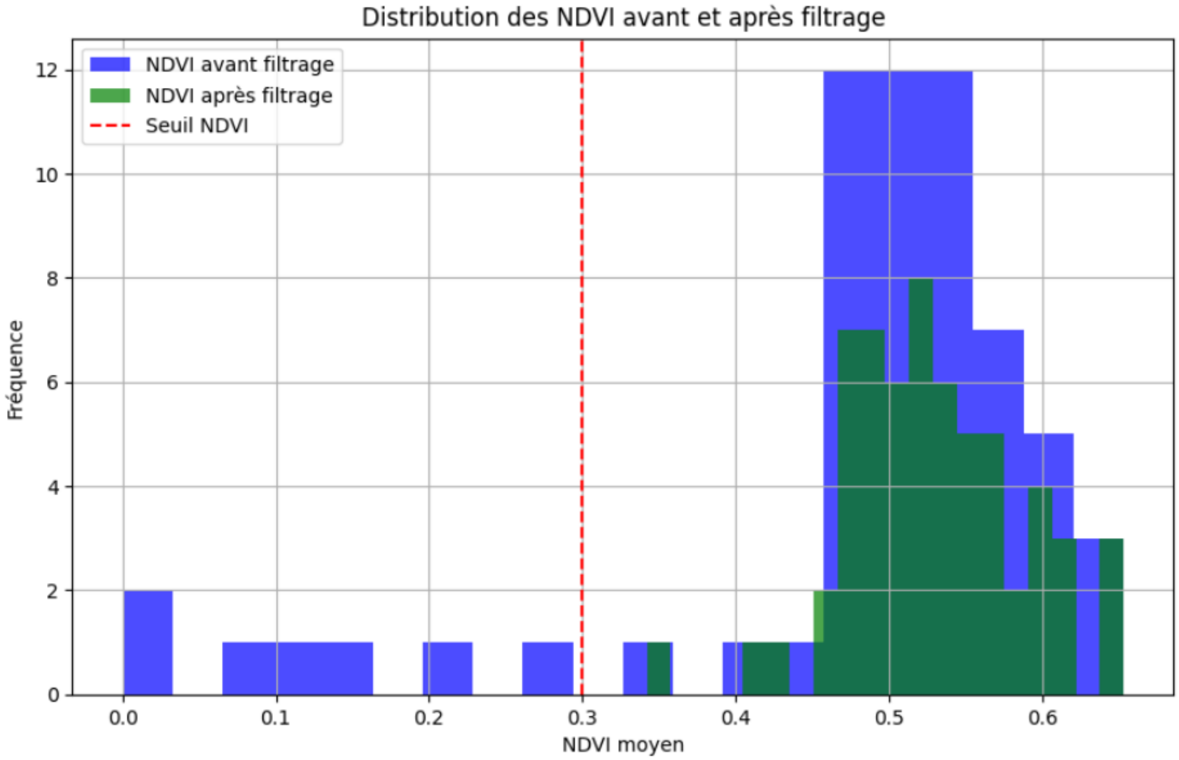


Average NDVI = 0.226



NDVI Cleaning Impact: Enhanced Data Quality





MINES PARIS | PSL

Capgemini

Data Modelling : pretrained models

We need a good semantic segmentation model

Our metric : IoU (measures the overlap between prediction and real surfaces).

First idea : existing pretrained models :

- Problem : <u>most pretrained models are RGB</u>, thus the shape must be changed, we lose a lot of the benefit of pretraining.

- <u>YOLO-seg</u> : good segmentation model, but needs to be fine-tuned with a special masking format : we would need to reprocess the entire dataset and store it somewhere.

- Other datasets pretrained on satellite images: danger of having been trained on PASTIS thus being biased.

Data Modelling

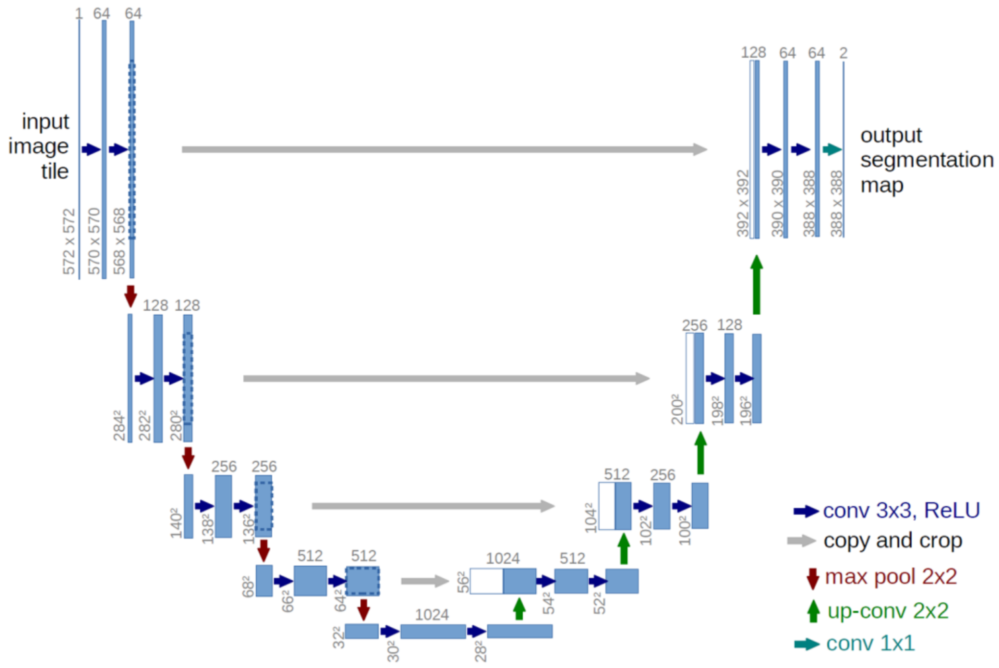
We train a new model from scratch. 2 architectures are considered:

- UNet (CNN) : Classic approach, good enough for many semantic segmentation tasks.
- Vision transformers : More powerful in theory, but need more data and more training time.
    - We thought transformers might be useful to capture evolutions in the time dimension.

Data preprocessing:

- Augmentation : Rotating the images.
- Preprocessing : removing the cloudy images with a threshold on the NDVI.

Data Modelling



UNet : Encoder - decoder structure with skip connections.
Reduces the dimension of the image by convolution to extract the crop categories.

Implementation from : https://github.com/milesial/Pytorch-UNet

The priority was to have a good enough spatial segmentation, with the time evolution as an additional information.

Thus we reduce the time series by taking their median. Empirically, it works better than taking the average.

Training : around 30 epochs gave the best results, then it starts to overfit, even with augmentation.

Further Work

Ideas for improvement:

- Training a 4D CNN on the time series to take into account evolution of the crops, which may give us insight into the type of plants, which may grow at different speeds and times of the year
- Using NDVI as an additional channel.
- Other data augmentation techniques like changing the contrast, adding noise or removing some areas.

Current Inefficiencies in Agricultural Insurance

The **High Cost of Inefficiency** in Agricultural Insurance: **Fraud, Delays, and Mispricing**

THE PAIN

2023

- **Fraudulent Claims :** Up to **10% of all claims** in some regions are fraudulent, costing **millions annually**
- **Slow Execution: Manual inspections**, causing long delays. On average, it takes **4-6 weeks** to process a claim
- **Inaccurate Risk Assessment :** Insurance premiums are often calculated based on **outdated or generalized data**, resulting in mispricing

Why?

- **Lack of Real-Time Data :** Insurers rely on manual inspections or periodic data snapshots
- **Reliance on Manual Inspections :** Processes require physical visits to farms for damage assessment
- **Inaccurate Risk Assessment :** Static and Generalized Risk Models based on historical data leading to mispricing premium

Why?

- **Inefficient Data Providers**
- **Lack of Automation**
- **Limited Use of Advanced Predictive Tools**

OPPORTUNITY

THE SOLUTION

Satellite Insights for Agricultural Insurance : Enhancing Accuracy and Efficiency

Capgemini invent | MINES PARIS | PSL

- **Automated, Continuous Monitoring**
- **Automated Parametric Insurance**
- **Dynamic Risk Assessment & Pricing**

Current Approach

- **Manual Verification**
- **Slow, Manual Processes**
- **Static Risk Models & Limited Predictive Analytics:**

Sources : (cognitivemarketresearch.com)

PRODUCT | SAAS and client support

An advanced satellite imagery model that **continuously monitors agricultural parcels.** This core technology **feeds into a series of AI-driven processes**, providing insurers with innovative tools to combat fraudulent claims, streamline execution, and enhance risk assessment accuracy , complemented by **dedicated client support** for seamless integration and ongoing assistance.

Parcel Detection and Classification

- Detects, monitors, and classifies agricultural parcels using satellite imagery.
- Forms the foundation of our solution with real-time insights.
- Provides accurate data on the current state of agricultural lands.

Automated Continuous Monitoring (Fraud Prevention)

- 7/7 automated surveillance of crops
- Insurers can verify the validity of claims.
- Reduces the chances of fraudulent claims by offering a comprehensive record of crop conditions over time

Automated Parametric Insurance (Faster Execution):

- Integration of real-time satellite data with parametric models
- The model automates policy adjustments and claim payouts based on predefined triggers like weather changes or crop damage
- Reduces human intervention, accelerates execution, and ensures transparent, criteria-based payouts.

Dynamic Risk Assessment & Pricing (Accurate Risk Evaluation):

- Leverages satellite model outputs alongside traditional insurer data.
- Utilizes advanced AI models for dynamic risk assessment.
- Continuously adjusts risk profiles and pricing in real-time.

Client Support

- **Onboarding & Training**: Insurers can seamlessly integrate and make the most of the platform's features.
- **Real-Time Analytics & Reporting**: On-demand insights and reports to make informed decisions quickly and efficiently.
- **Continuous Support & Updates**: Technical and business support, with continuous feature updates
- **Customizable Dashboards**: Tailored dashboards enable clients to monitor key performance indicators and track claims, risks, and payouts in real-time.

The Market

A mature market **with deep traditions and high entry barriers,** we will leverage our **trusted relationships in the insurance sector** to propose our innovative solution and ongoing support.

Market Opportunity

Agricultural Insurance Market Size: Global market size was estimated **at 39 514.2 M$** with 30% in Europe (11854.26 M$) in 2024. The market will grow at a compound annual growth rate (CAGR) of 4.5% from 2024 to 2031.
In France, the market size was estimated at **1090.59 M$** in 2024 with CAGR of 3,7% during the same perode

Growth Drivers: Rising need for automated, data-driven risk management solutions in agricultural insurance to address fraud, slow claim processing, and inaccurate risk assessments.

Competitive Landscape

Our Edge: We integrates satellite imagery, AI, and parametric models in one solution, offering insurers a more comprehensive, automated approach.

Go-to-Market Strategy

- **Target Customers:**
  - ✓ **Primary** : Insurance companies looking to modernize claims processing and improve risk management in agriculture.
  - ✓ **Secondary** : Insurers with existing in-house solutions who seek to benchmark and compare their models with our satellite-driven AI system for enhanced accuracy and efficiency.
- **Partnerships**: Collaboration with satellite data providers to ensure accurate and real-time data streams.
- **Revenue Model**: **SaaS subscription-based** with additional usage fees for data processing and tiered **client support packages**..

Financial Overview and Risk Assessment (2025-2030)

A total investment **of $1.5 million** with the acquisition of **20 clients**, projecting revenues of **$5.68 million** over six years, resulting in an impressive **ROI of approximately 418%** during the period of 2025-2030 with manageable risks.

Allocation of Funds(1,5M$) :

- **Product Development**: 50%
- **Marketing & Sales**: 10%
- **Operational Costs**: 20%
- **Client Support**: 10%

Expected Outcomes :

- Launch product by Q4 2025.
- Acquire 2 clients in 2025, expanding to 20 clients by 2030.
- Linear revenue forecast during the 2025-2030 period

Total Annual Revenue :
5 685 000 $

ROI = 418,5 %

- **Market Risk** ★★
  Increased competition in the agricultural insurance sector could limit market share.
- **Technological Risk** ★★★
  Dependence on advanced technologies (e.g., AI modeling, satellite image processing) requires constant updates and may lead to unforeseen costs.
- **Client Risk** ★
  Difficulty in acquiring and retaining clients due to hesitancy in adopting new technological solutions.
- **Regulatory Risk** ★★
  Changes in insurance regulations or agricultural practices could impact product viability.

→ Manageable risks

2026-2030: Expansion and Revenue Growth
2025: Product Launch & Initial Client Acquisition

MINES PARIS | PSL

Data Challenge

Kais Kharrat
Adam Saoud
Guillaume Vindry
David Sergio Arrustico Villanueva

Annex :Financial Overview (2025-2030)

Fees per Client:

- Subscription Fees (Year)($) =5000$
- Usage fees (Monthly)($) = 1000+500*year
- Consulting Fees (Monthly)($)= 2000+500*year

Year	Clients	Subscription Fees (Year)($)	Usage Fees (Monthly)($)	Consulting Fees (Monthly)($)	Total Monthly Revenue($)	Total Annual Revenue($)
2025	2	10000	2000	4000	6833,33333	82000
2026	5	25000	7500	12500	22083,3333	265000
2027	10	50000	20000	30000	54166,6667	650000
2028	15	75000	37500	52500	96250	1155000
2029	17	85000	51000	68000	126083,333	1513000
2030	20	100000	70000	90000	168333,333	2020000

Papers

[1] 4D U-Nets for Multi-Temporal Remote Sensing Data Classification (Giannopoulos et al. 2022)