Geomatic Techniques to Support Phytosanitary Products Tests whithin the EPPO Standard Framework

Samuele Bumbaca

University of Turin

August 28, 2025

The Traditional Approach to Agricultural Trials

Block 3	R	C	T
Block 2	T	R	C
Block 1	C	T	R

C Control
T Tested Product
R Reference Product

ANOVA Model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Where:

- y_{ij} = response
- μ = overall mean
- α_i = treatment effect
- β_j = block effect
- ε_{ij} = random error

Note:

This is the additive model. Modern approaches may include interaction terms: $\alpha_i \times \beta_j$

Key Assumptions of Traditional ANOVA

Statistical Assumptions:

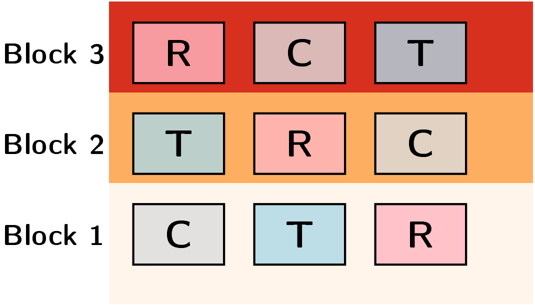
- **Randomization**: Treatments randomly assigned within blocks
- **Replication**: Each treatment appears in each block
- **Independence**: Observations are independent given the design
- **Homoscedasticity** : Equal variances across treatments
- **Normality**: Residuals follow normal distribution

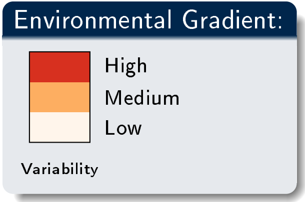
Consequences of Assumption Violations:

- **Invalid conclusions of parametric tests**: Need for non-parametric tests leading to reduced statistical power

Based on R. A. Fisher, Statistical Methods for Research Workers, in S. Kotz & N. L. Johnson (eds.), Breakthroughs in Statistics: Methodology and Distribution, pp. 66–70, Springer, New York, 1992.

The Right Blocking: Capturing Environmental Variability



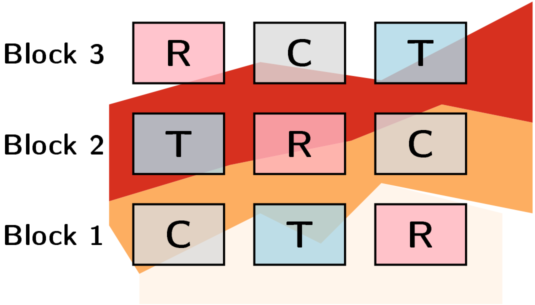


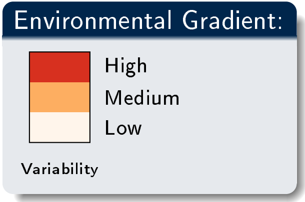
C Control
T Tested Product
R Reference Product

Success of Blocking Strategy:

- **Within-block homogeneity**: Treatments compared under similar conditions
- **Between-block heterogeneity**: Environmental gradient captured by block effects

The Wrong Blocking: Assumption Violation





C Control
T Tested Product
R Reference Product

Heteroscedasticity Assumption Violation Problem:

- **Blocks fail to capture environmental variability**: Treatments compared under different conditions
- **Invalid parametric test**: Residual variance differs across treatments

Current Limitations in Statistics for Agricultural Trials

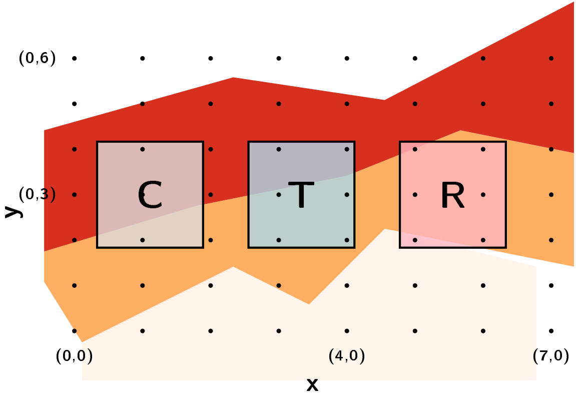
Traditional Approach Issues:

- **Human-dependent blocking**: Environmental variability assessment relies on experimenter experience
- **A priori identification**: Must identify variance sources BEFORE data collection

The Challenge:

How can we capture environmental variability mathematically rather than through human judgment?

Geostatistical Approach: Spatial Linear Mixed Models



C/T/R	Control/Tested/Reference
•	Georeferenced observations

Spatial LMM:

$$y(s_i) = \mu + \alpha_j + f(s_i) + \varepsilon_i$$

Where:

- $y(s_i)$ = response at s_i
- μ = overall mean
- α_j = treatment effect
- $f(s_i)$ = spatial random field
- ε_i = error
- $s_i = (x_i, y_i)$ = coordinates

Benefits:

- **No blocking**: Spatial correlation captures variability
- **Post-hoc**: No a priori variance identification
- **Homoscedasticity**: Assumption satisfied in more cases in respect blocking

Statistical Methods Comparison: Introduction

Comparison Objective:

Evaluate the performance of **traditional RCBD** versus **spatial geostatistical methods** (SpATS) in capturing environmental variability and estimating treatment effects.

Synthetic Dataset:

- **54 observations**(6×9 grid)
- **3 treatments**: Control (0 t/ha), Reference (0.5 t/ha), Test (1.0 t/ha)
- **3 blocks**(18 plots each)
- **Environmental zones**: Low (-1.5 t/ha), Medium (0 t/ha), High (+1.5 t/ha)

Tested Models:

1. **RCBD Model**: Linear Mixed Model with random block effects

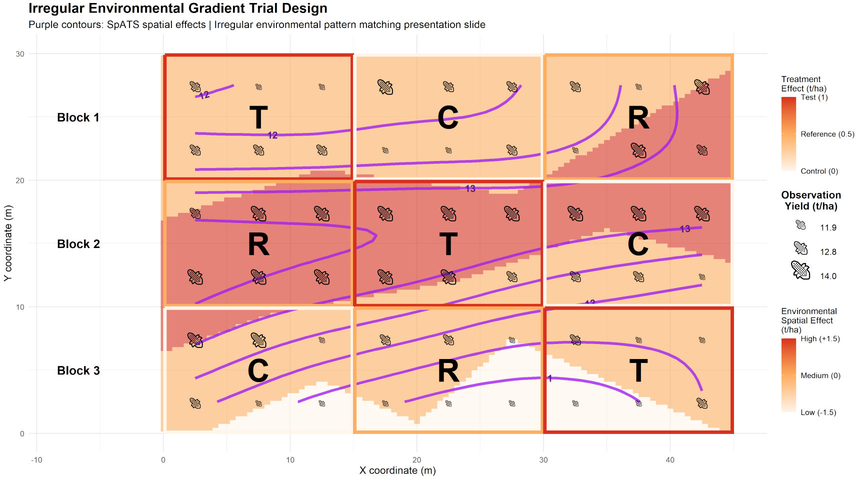
$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

2. **SpATS Model**: Spatial model with PSANOVA splines

$$y(s) = \mu + \alpha_i + f(s) + \varepsilon(s)$$

Where: α_i = treatment effects, β_j = block effects, $f(s)$ = spatial smooth

Statistical Methods Comparison: The Field Trial Design



Statistical Methods Comparison: Results

Model Performance (Mean Absolute Errors tonn/ha):

Model	Treat. Error	Env. Error
RCBD Model	0.13	0.62
SpATS Spatial	0.03	0.45

Treatment Effect Estimation (tonn/ha):

Treatment	True	RCBD	SpATS
Control	0.00	0.00	0.00
Reference	0.50	0.40	0.45
Test	1.0	0.69	0.94

Key Findings:

- **Both models satisfied assumptions**
- **SpATS outperformed RCBD**:
  - 3.8× better treatment effect estimation
  - 1.4× better environmental effect estimation
- **RCBD underestimated** by 20-31%
- **SpATS <6% error**

Implications:

Even when traditional RCBD meets statistical assumptions, **spatial modeling provides superior accuracy** in treatment effect estimation by properly accounting for environmental spatial variability.

The Missing Link: Spatial Coordinates

Geostatistical Methods Advantages:

- ✓ **Mathematical modeling** of environmental variability
- ✓ **Post-hoc analysis** - no need for prior knowledge of the environment variables and of their distribution
- ✓ **Superior performance** in handling spatial heterogeneity
- ✓ **EPPO recognized** approach (PP1/152(4) - Design and analysis of efficacy evaluation trials)

Current Barrier:

- × **Requires spatially referenced observations**
- × **Traditional manual assessments lack coordinates**
- × **Implementation gap** in practical field trials

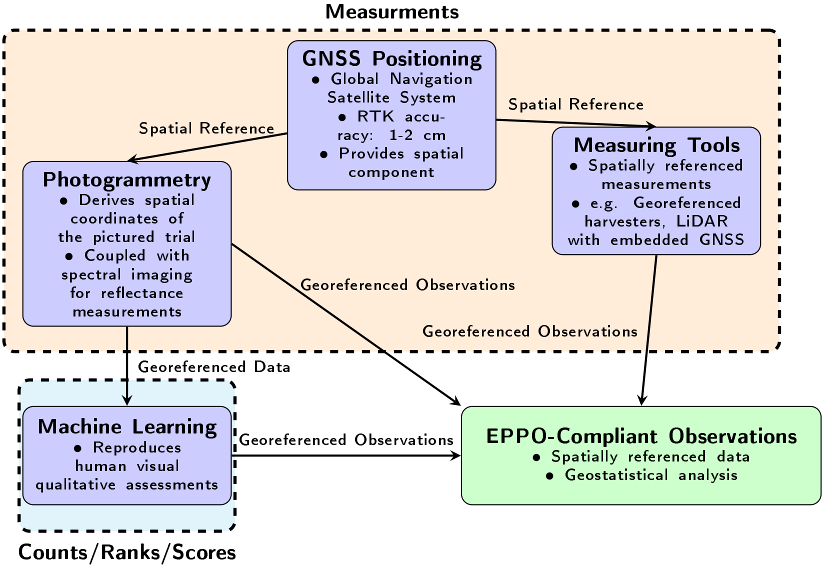
Central Research Question

Can geomatics technologies provide spatially referenced observations that enable geostatistical analysis within EPPO-compliant Plant Protection Product trials?

Specific Objectives:

1. Establish which geomatics technologies can be used to collect spatially referenced observations
2. Demonstrate the feasibility of collect spatially referenced observations in compliant with EPPO standards
3. Validate performance against traditional methods
4. Provide practical implementation guidelines

Geomatic Technologies: Workflow for Spatially Referenced Observations



Georeferencing EPPO Standard Assessments

Table: EPPO's types of variables

Type of Variable	Measurement	Ranking	Scoring
Continuous not limited	X		
Continuous limited	X		
Discrete	X		
Ordinal		X	X
Nominal			X
Binary			X

Summary from EPPO PP 1/152: Design and analysis of efficacy evaluation trials

Current State of Georeferencing in Agricultural Trials:

Tool-based measurements (e.g., yield harvesters) can be easily georeferenced by integrating GNSS receivers on the tool. For visual assessments as counting, scoring or ranking, a method to transform georeferenced data to georeferenced observations is needed.

Machine Learning Inference on Georeferenced Data

Machine Learning for Visual Assessments:
Machine Learning (ML) can reproduce human visual assessments, allowing for spatially referenced observations.

Machine Learning Limitations:
ML needs training data to reproduce human assessments. Understand data requirment to fit EPPO standard is critical.

Application Case Studies:
This thesis provides a case study to prove the possibility to get georeferenced observations for each variable type that need for a ML step.

EPPO ML integration

EPPO PP 1/333(1): Digital Technologies in PPP Trials

ML integrated assessments must meet the same quality standards as manual assessments and require validation through comparison with manual assessments (golden sample).

Validation Benchmarks[1]

[1] Based on EPPO PP 1/333(1): Use of digital technologies in efficacy and selectivity trials

- **Continuous/Discrete**: $R^2 > 0.85$ (1:1 relationship)
- **Ordinal/Nominal**: Cohen's $\kappa > 0.7$
- **Binary**: Accuracy > 0.85

Georeferencing Gap in EPPO Standard Assessments

	Type of Variable	Measurement	Ranking	Scoring
	Continuous not limited	X		
	Continuous limited	X		
→	Discrete	X		
→	Ordinal		X	X
→	Nominal			X
→	Binary			X

Case Studies:

This thesis aim to prove the reliability of georeferencing every EPPO standard assessmentEach case study addresses a specific variable type as defined in the EPPO standards

- **Discrete (Counts)** : Plant counting
- **Ordinal** : Phytotoxicity scoring
- **Nominal** and **Binary** : Disease detection

Georeferencing Counts (Discrete Variable)

	Type of Variable	Measurement	Ranking	Scoring
	Continuous not limited	X		
	Continuous limited	X		
→	Discrete	X		
	Ordinal		X	X
	Nominal			X
	Binary			X

Georeferencing Counts:

- **Counts** are discrete variables required for measuring density of individuals (e.g. plant density in PP1/46 (3) - Wireworms).
- the **Case Study**: Counting plants from georeferenced photogrammetric orthomosaics by ML Object Detection.
- this study is discussed in the scientific article **Bumbaca, S.; Borgogno-Mondino, E.C. On the Minimum Dataset Requirements for Fine-Tunining an Object Detector for Arable Crop Plant Counting: A Case Study on Maize Seedlings. Remote Sens. 2025, 17, 2190. DOI: 10.3390/rs1713219061**

Arable Crop Plant Counting by Object Detection

The Critical Need after EPPO Assessments:

- Plant counting is **fundamental** also in precision agriculture and plant breeding
- Traditional manual counting is **time-consuming** and bring **human error** risks
- **Computer vision** offers a solution, but requires **dataset size and quality** characterization to prove the reliability for this task.

EPPO Benchmark Standards:

Coefficient of determination (R^2) ≥ 0.85 of ML method w.r.t. manual counting (no bias nor slope linear first order relation)

Research Gap:

What are the minimum dataset requirements to achieve this benchmark across different inference datasets?

Photogrammetric Orthomosaics for Plant Counting

Advantages over other kind of data:

- **Geographical coordinates**: Suitable for spatial analysis
- **Fixed scale and orientation images**: Eliminate perspective inconsistencies
- **Achievable High-resolution**: From low altitude nadiral overlapping images[1]

[1] Kraus, K. Photogrammetry: Geometry from Images and Laser Scans. De Gruyter: Berlin, Germany, 2011

Limitations:

- **Occlusions**: Overlapping vegetation canopy issues [1] -> Target crop and phenological stage selection
- **Georeferencing errors**: Due to low-quality/insufficient GNSS embedded systems or Ground Control Points (GCPs) [2] -> Hardware requirements
- **Computational demand**: Processing time constraints for large-area orthomosaics -> Not real-time suitable

[1] Habib et al. Automated Ortho-Rectification of UAV-Based Hyperspectral Data over an Agricultural Field Using Frame RGB Imagery. *Remote Sens.*, 8:796, 2016

[2] Pugh et al. Comparison of image georeferencing strategies for agricultural applications of small unoccupied aircraft systems. *Plant Phenome J.*, 4:e20026, 2021

Plants Occlusion Solution: Maize Seedlings at BBCH 13-15 Stage

- **Optimal detection conditions**: Regular planting pattern, minimal plant overlapping at BBCH 13-15 stage [1]
- **Data availability**: Most represented plant in scientific [2] [3] and public datasets
- **Economic importance**: World's highest-production crop [4]
- **Rappresentative crop**: Findings applicable to other row crops [5] (e.g. Sunflower, Sugar beet)

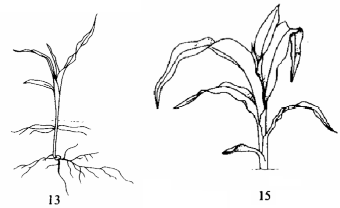
[1] Meier et al. The BBCH system to coding the phenological growth stages of plants. *J. Für Kult.*, 61:41–52, 2009

[2] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[3] Liu et al. IntegrateNet: A deep learning network for maize stand counting from UAV imagery. *IEEE Geosci. Remote Sens. Lett.*, 19:6512605, 2022

[4] FAO. *Agricultural Production Statistics 2010–2023*. FAOSTAT, Rome, Italy, 2024

[5] Torres-Sánchez et al. Early Detection of Broad-Leaved and Grass Weeds in Wide Row Crops Using Artificial Neural Networks and UAV Imagery. *Agronomy*, 11:749, 2021



Suitable Hardware and Photogrammetric Picturing

- **UAV Platform**: Phantom 4 Pro v2.0 (DJI, Shenzhen, China)
- **Camera**: Default series RGB camera
- **Flight Altitude**: 10 m above ground level
- **Original GSD**: 2.7 mm/pixel
- **GNSS Mode**: VRS-NRTK for GCP surveying
- **Bundle Adjustment Error**: 38 mm
- **Final Orthomosaic GSD**: 5 mm/pixel
- **Reference System**: WGS84/UTM 32 N

Key Processing Steps:

1. Nadiral image capture with 70%-80% overlapping patterns
2. Ground Control Points (GCPs) surveyed with high-precision GNSS
3. Photogrammetric bundle adjustment and orthomosaic generation
4. Georeferenced orthomosaic output ready for spatial analysis

Plant Counting - Object Detection Methods

Machine (Deep) Learning Approaches:

Convolutional Neural Networks [1] Grid-based convolutions

- **Faster R-CNN** [2]
- **YOLO variants** for faster inference

[1] LeCun et al. Deep learning. *Nature*, 521:436–444, 2015

[2] Faster R-CNN: Towards real-time object detection with region proposal networks

Transformer Architectures [3] Image patches processing (attention-based)

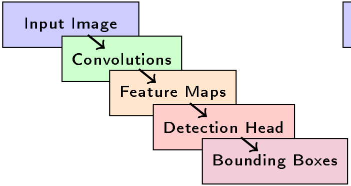
- **DETR** [4]
- **Hybrid approaches** with convolutions and attention

Superior with scarce data [5]

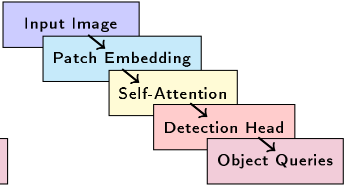
[3] Vaswani et al. Attention is all you need. *NIPS*, 2017

[4] Carion et al. End-to-end object detection with transformers. *arXiv:2005.12872*, 2020

[5] Rekavandi et al. Transformers in small object detection: A benchmark and survey. *arXiv:2309.04902*, 2023



CNN-based



Vision Transformer

Plant Counting - Research Gap

Critical Research Gaps:

- **Minimum dataset requirements** None of the studies taken into account systematically tested[1] the minimum dataset requirements for robust plant detectors (EPPO benchmark).
- **In-domain vs. out-of-distribution data** Despite some author already studied the impact[2] [3] none did it in a quantitative and systematic way.

---

[1] Sun et al. Revisiting unreasonable effectiveness of data in deep learning. *arXiv:1707.02968*, 2017

[2] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[3] Andvaag et al. Counting canola: Toward generalizable aerial plant detection. *Plant Phenomics*, 6:0268, 2024

Study Aim

Primary Objective:

Establish the **minimum dataset requirements** for accurate maize seedling detection (EPPO benchmark) in georeferenced orthomosaics across different object detection paradigms

Key Definitions:

- **Dataset size**: Amount of annotated images in training set
- **Dataset quality**: Accuracy of annotations (percentage of correct annotations relative to ground truth)

Specific Research Questions:

1. How does training data source (in-domain vs. out-of-distribution) affect required dataset size and quality?
2. Untill which extent different architectures affect training dataset requirements?

Plant Counting - Material and Methods - Research Methodology

- **Objective**: Investigate minimum dataset size and quality for robust object detection
- **Classic Object Detectors requirements**:
  - with out-of-distribution (OOD) training datasets.
  - with in-domain (ID) training datasets.
- **Empirical Modeling Approach**:
  - Analyze the relationship between dataset size/quality and model performance
  - Fit empirical functions to characterize this relationship
  - Use fitted functions to predict performance with varying dataset size/quality

Plant Counting - Material and Methods - Dataset

Dataset Classification:

- **Out-of-Distribution (OOD)**: Training datasets from different sources than inference target
- **In-Domain (ID)**: Training datasets from same source/distribution as testing dataset

OOD Scientific Datasets:

Source: Scientific literature

OOD Internet Datasets:

Source: Internet repositories

ID Datasets:

Source: Collected by the author

Key Processing Parameters:

All dataset preprocessed to get standard **tile size**: 224$\times$224 pixels (1.12$\times$1.12 m field coverage for georeferenced)

Plant Counting - Material and Methods - Datasets

Dataset	Phenological Stage	Train Size	Test Size
OOD Scientific			
DavidEtAl.2021 [1]	V3	182 tiles	N/A*
LiuEtAl.2022 [2]	V3	596 tiles	N/A*
OOD Internet			
OOD_int_1 [3]	V3	216 tiles	N/A*
OOD_int_2 [4]	V5	174 tiles	N/A*
ID [5]			
ID_1	V3	150 tiles	20 tiles
ID_2	V3	150 tiles	20 tiles
ID_3	V5	150 tiles	20 tiles

* N/A indicates that these datasets were used only for training purposes and do not have separate test sets in this study.

[5] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[5] Liu et al. IntegrateNet: A deep learning network for maize stand counting from UAV imagery. *IEEE Geosci. Remote Sens. Lett.*, 19:6512605, 2022

[5] Maize_seeding dataset. `https://universe.roboflow.com/objectdetection-hytat/maize_seeding`

[5] Maize-seedling-detection dataset. `https://universe.roboflow.com/fyxdds-icloud-com/maize-seedling-detection`

[5] Bumbaca, Samuele. 'The Original Dataset for the Paper "on the Minimum Data Set Requirements for Fine-tuning an Object Detector for Arable Crop Plant Counting: A Case Study on Maize Seedlings"'. Zenodo, 17 April 2025. https://doi.org/10.5281/zenodo.15235602.

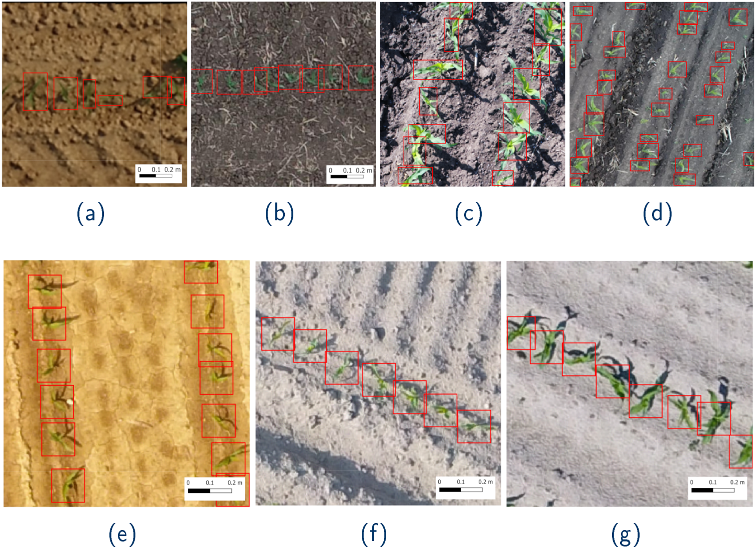


Figure: Image examples taken from each dataset, ground truth bounding boxes are shown in red. (**a**) DavidEtAl.2021, (**b**) LiuEtAl.2022, (**c**) Internet Maize stage V3, (**d**) Internet Maize stage V5, (**e**) ID_1, (**f**) ID_2, (**g**) ID_3.

Primary Evaluation Metrics:

Performance assessed using R^2 and mAP for counting and detection respectively

Counting Metric:

Coefficient of Determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Detection (Spatial) Metric:

Mean Average Precision:

$$mAP = \frac{1}{|IoU|} \sum_{t \in IoU} AP_t$$

Metric Interpretation:

R^2: 1 = perfect, 0 = mean prediction, negative = worse than mean | mAP: IoU threshold 0.5

Plant Counting - Material and Methods - Training Protocols

Training Dataset Configuration:

90% training / 10% validation split

Dataset Size Evaluation:

10 to 150 images (15 steps of 10)

Quality Assessment:

- **Annotation reduction**: 100% to 10% (10 steps)
- **Constant dataset size**: During quality evaluation
- **OOD vs ID influence**: Same experimental protocol

Plant Counting - Material and Methods - Predictive Modeling

Empirical Function Testing:

Three mathematical functions tested to model dataset size/quality vs performance relationships

Logarithmic:

$$f(x) = a\ln(x) + b$$

Behavior: Diminishing returns pattern **Theory**: Asymptotic performance approach

Arctangent:

$$f(x) = a\arctan(bx) + c$$

Behavior: Saturating performance **Theory**: Bounded metrics plateau

Algebraic Root:

$$f(x) = ax^{1/b} + c$$

Behavior: Power-law relationships **Theory**: Flexible scaling dynamics

Model Selection Criteria:

- **Goodness-of-fit**: $GoF = R^2_{fit}$ for function selection
- **Best predictor**: Highest fit determines model-metric combination
- **Practical guidance**: Annotation planning through interpolation/extrapolation

Plant Counting - Materials and Methods - Classic Detectors Architecture

YOLOv5 - Baseline CNN Architecture:

- **Backbone**: CSP (Cross Stage Partial) with PANet neck
- **Agricultural dominance**: Most widely adopted in crop monitoring [1]
- **Reference point**: Well-established baseline for dataset requirements

[1] Badgujar et al. Agricultural object detection with YOLO algorithm. *Comput. Electron. Agric.*, 223:109090, 2024

YOLOv8 - Improved CNN Architecture:

- **Backbone improvement**: C2f blocks for enhanced efficiency
- **Detection head**: Anchor-free design with decoupled heads
- **Performance**: Superior accuracy-speed trade-offs [2]

[2] Terven et al. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.*, 2023

CNN Architecture Benefits:

Computational efficiency, **proven agricultural performance**[3] [4], and **established baseline** for dataset requirement comparison

[3] Kitano et al. Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geosci. Remote Sens. Lett.*, 16:1–5, 2019

[4] Barreto et al. Automatic UAV-based counting of seedlings. *Comput. Electron. Agric.*, 191:106493, 2021

Plant Counting - Materials and Methods - Classic Detectors Architecture

YOLO11 - Transformer-mixed:

- **Key innovation**: Multi-scale deformable attention mechanisms (for small object detection)
- **Hybrid approach**: YOLO backbone + Transformer attention

RT-DETR - CNN+Transformer Hybrid:

- **Architecture**: CNN backbone + Transformer decoder
- **Attention mechanism**: Deformable attention for adaptive feature sampling
- **Global relationships**: Models object interactions across entire image
- **Real-time performance**: Parallel prediction heads
- **Agricultural proven**: Superior inference performances in respect pure-CNN YOLOs [1]

[1] Zhao et al. DETRs beat YOLOs on real-time object detection. *arXiv:2304.08069*, 2024

Research Question:

Do Transformer-mixed improvements affect minimum dataset requirements for small object detection compared to pure CNN approaches?

Plant Counting - Materials and Methods - Detectors Architecture

Unified Training Configuration and Implementation:

- **Library**: Ultralytics open-source implementation [1]
- **Consistency**: Same framework enables fair architectural comparison
- **Hardware**: Intel Xeon E5-2670 v3, 64GB RAM, NVIDIA RTX A5000 (24GB VRAM)

[1] Jocher, Glenn; Qiu, Jiarui; Chaurasia, Anil. GitHub Ultralytics YOLO. 2023. Available online: https://github.com/ultralytics/ultralytics

Training Hyperparameters:

- Batch size: 16
- Max epochs: 200
- Early stopping: 15 epochs without improvement

Data Augmentation Protocol:

- **Geometric**: Random scaling, Translation
- **Photometric**: HSV augmentation
- **Composition**: Mosaic augmentation, Horizontal flip

Excluded Alternatives:

Faster R-CNN: Computational overhead[2] | **Pure DETR**: Prohibitive training requirements for small datasets [3]

[2] Velumani et al. Estimates of Maize Plant Density from UAV RGB Images Using Faster-RCNN Detection Model. *Plant Phenomics*, 2021:9824843, 2021

[3] Carion et al. End-to-end object detection with transformers. *arXiv:2005.12872*, 2020

Plant Counting - Materials and Methods - Architecture Summary

Table: Summary of tested architectures and model sizes (millions of parameters)

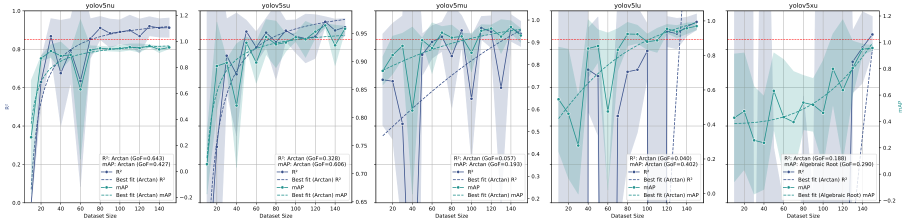
Architecture	**Shots**	**n**	**s**	**m**	**l**	**x**
YOLOv5	many	1.9	7.2	21.2	46.5	86.7
YOLOv8	many	3.2	11.2	25.9	43.7	68.2
YOLO11	many	4.0	12.5	28.0	50.0	75.0
RT-DETR	many	-	-	-	60.0	80.0

n: nano, **s**: small, **m**: medium, **l**: large, **x**: extra-large

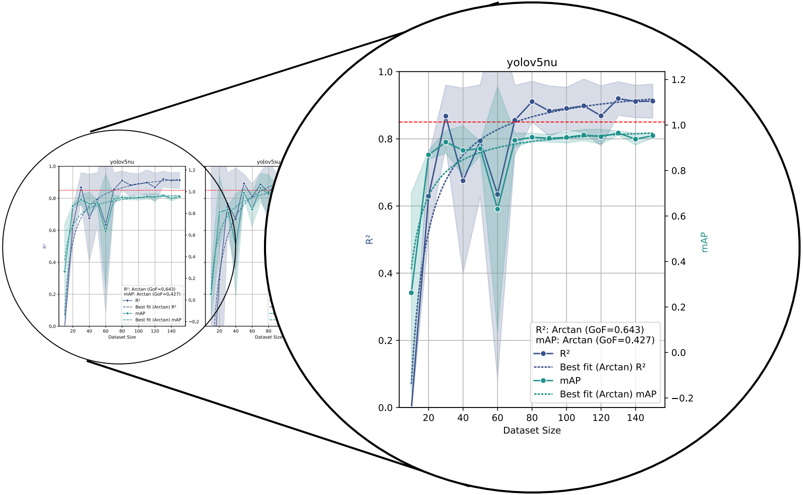
Architecture Selection Strategy:

Parameter count affects dataset requirements: larger models may need more data for training but offer better feature extraction capabilities for complex tasks

Plant Counting - Results - YOLOv5 dataset size

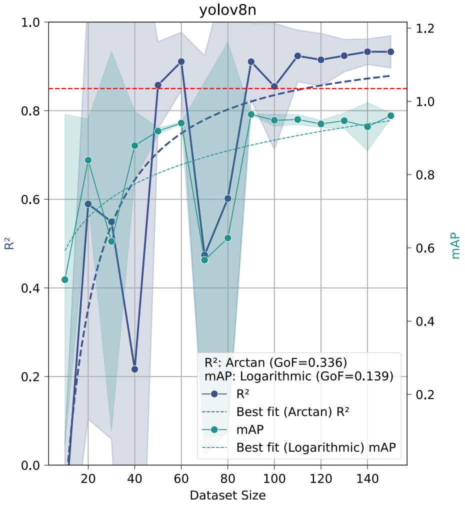


Plant Counting - Results - YOLOv5 dataset size



YOLOv5 demonstrates that traditional CNN architectures with low parameter amount are sufficient even with only 130 samples

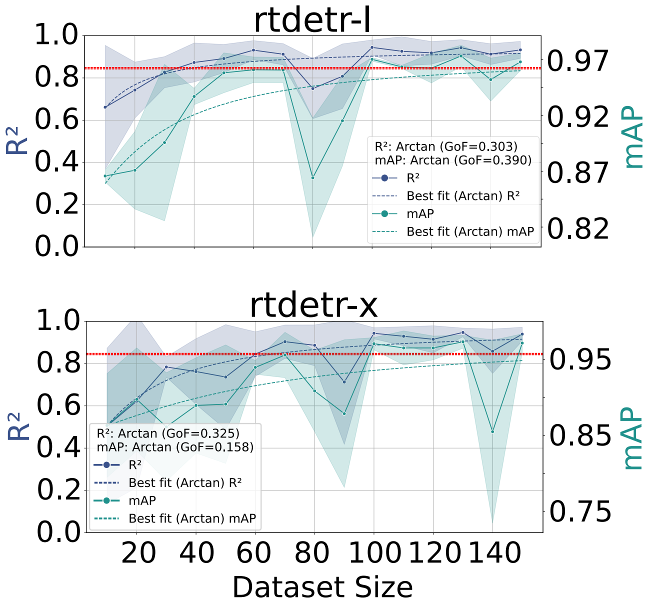
Plant Counting - Results - YOLOv8 dataset size



Evolution Impact:

- Reductions in annotation burden (110 images)
- Only low amount of parameters succeded like in YOLOv5

Plant Counting - Results - RT-DETR dataset size



Plant Counting - Results - Dataset Size

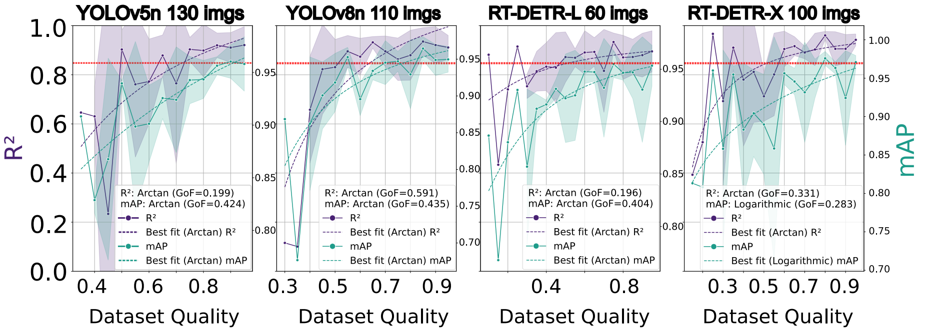
Architecture	Parameters	Dataset Size
YOLOv5	1.9 (n)	130
YOLOv8	3.2 (n)	110
RT-DETR	60 (l)	60
RT-DETR	80 (x)	100

Transformer-Mixed Superiority at a higher parameters price:

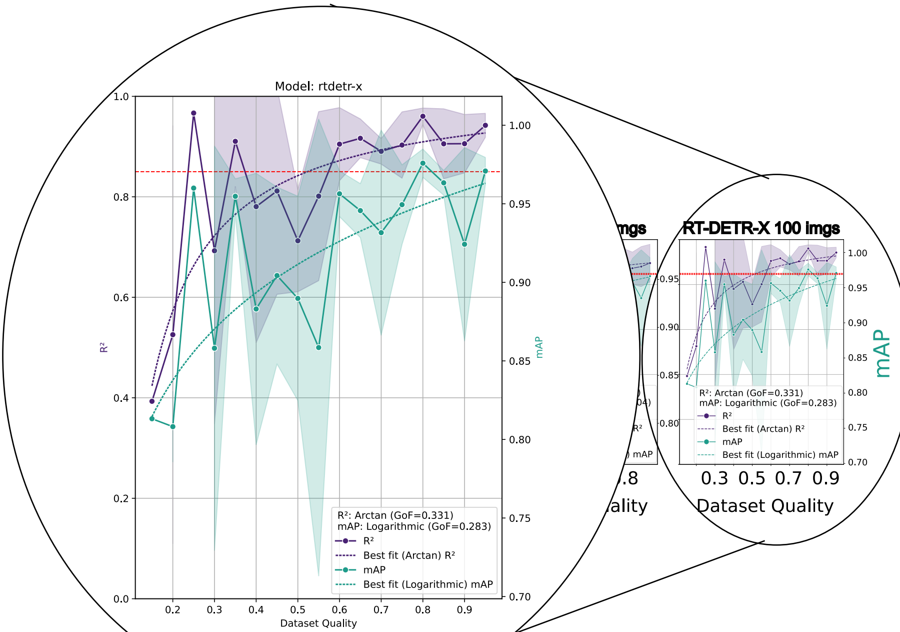
RT-DETR demonstrates reduced dataset requirements in respect pure-CNN counter parts.
YOLO11 did not succeded to reach the benchmark with any dataset and parameter size.

RT-DETR L predictions trained on 60 images

Plant Counting - Results - Dataset Quality Requirements



Plant Counting - Results - Dataset Quality Requirements



RT-DETR X predictions with 35% quality reduction

Quality vs Quantity Trade-offs:

Strategic insight: RT-DETR architectures offer flexibility - achieve benchmark with either minimal size and high-quality data (60 images, 100% quality) or more abundant medium-quality data (100 images, 65% quality)

Plant Counting - Discussion: Key Findings

Critical Findings - Dataset Source Impact:

- **In-domain data mandatory**: No OOD-trained model achieved $R^2 = 0.85$ benchmark
- **GoF limitation**: OOD models showed GoF < 0.2, indicating poor predictability with limited datasets (1168 images)

Architecture vs Dataset Requirements:

- **CNN complexity trade-off**: Higher parameters = higher dataset needs
- **Transformer superiority**: RT-DETR L achieves benchmark with only 60 samples
- **Predictable scaling**: GoF > 0.3 enables practical annotation planning

Quality vs Quantity Trade-offs:

- **Quality tolerance**: Models maintain benchmark with 65-90% annotation quality
- **Strategic flexibility**: RT-DETR offers choice between minimal high-quality (60 images, 100%) or abundant low-quality data (100 images, 65%)

Plant Counting - Conclusions and Future Directions

Practical Implementation Guidelines:

- **Minimum requirements**: Focus on minimum viable dataset of 60 images with high annotation quality or 100 images with 65% quality
- **Architecture selection**: Use CNNs models with few parameters or RT-DETR when allowed by computational resources

Future Research Directions:

Domain-specific pre-training: Agricultural aerial orthomosaic imagery backbones may further reduce dataset requirements

Core Conclusion:

EPPO compliance: The minimum dataset size of 60-100 images (about 75 to 125 m^2) is achievable for any efficacy or selectivity trial (about 500-1000 m^2), enabling practical implementation of geostatistic in EPPO standard framework

Georeferencing Rankings (Ordinal Variable)

	Type of Variable	Measurement	Ranking	Scoring
	Continuous not limited	X		
	Continuous limited	X		
	Discrete	X		
$\rightarrow$	Ordinal		X	X
	Nominal			X
	Binary			X

Georeferencing Rankings:

- **Rankings** are ordinal variables required for estimating phytotoxicity sympthoms (e.g. PP 1/135 (4) Phytotoxicity assessment).
- the **Case Study**: Ranking phytotoxicity sympthoms from georeferenced photogrammetric orthomosaics by ML regression.
- this study is discussed in the scientific article **Bumbaca, S.; Borgogno-Mondino, E. Supporting Screening of New Plant Protection Products through a Multispectral Photogrammetric Approach Integrated with AI. Agronomy 2024, 14, 306. DOI: 10.3390/agronomy14020306**

PHYGEN: General Phytotoxicity Index

- **Definition**: Aggregate indicator summarizing phytotoxicity symptoms as percentage of damage compared to healthy reference plant
- **EPPO Symptoms**: (i) development cycle modifications, (ii) thinning, (iii) color modifications, (iv) necrosis, (v) deformation, (vi) yield effects[1]
- **Selectivity Assessment**: Critical for Plant Protection Product (PPP) market approval

[1] EPPO. PP 1/135(2) - Design and analysis of efficacy evaluation trials. *EPPO Bull.*, 42:367–381, 2014

Statistical Measurement Theory Concerns

Stevens Scale Theory Problem[2]:

- **Current practice**: Ordinal discrete scales (0%, 13%, 38%, 63%, 88%)
- **Statistical limitation**: Ordinal data violates ANOVA assumptions[3]
- **Rater variability**: 10% inter-rater error commonly accepted[4]

[2] Stevens, S.S. On the Theory of Scales of Measurement. *Science*, 103:677–680, 1946

State of the Art in Automated Phytotoxicity Assessment

Current Approaches Limitations

- **Human assessment**: Subjective, 10% maximum error[1]
- **CNN methods**: Require thousands of images for training[2]
- **Deep learning**: Not suitable for new PPPs with limited training data

[1] Chiang, K.-S.; et al. Effects of rater bias and assessment method on disease severity estimation. *Plant Dis.*, 100:2530–2538, 2016

[2] Gómez-Zamanillo, A.; et al. Damage assessment in soybean and redroot pigweed plants exposed to herbicides. *Agronomy*, 13:2523, 2023

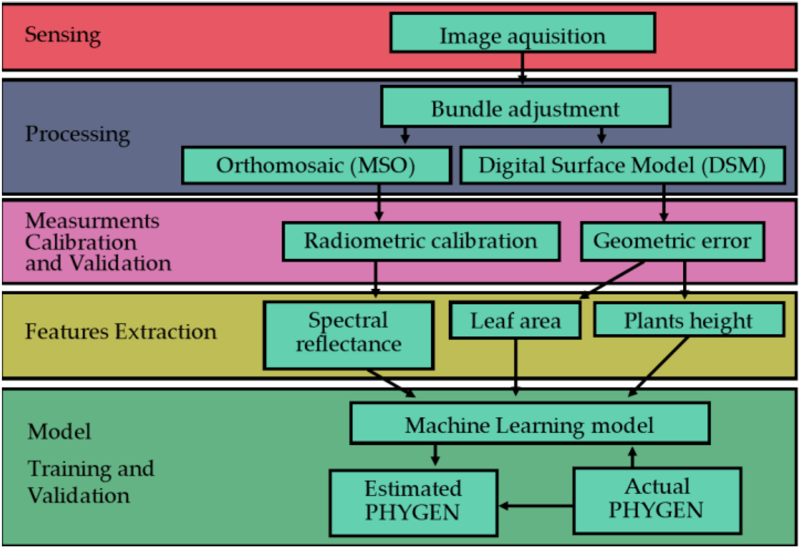
Table: Related works comparison

Method	Accuracy	New PPP Su
Human raters	90% (±10%)	Traditional
CNN (Ghosal et al.)	50-90%	Destructive
CNN (Gómez-Z. et al.)	93.26%	Not suitable
This work	**89.34%**	**Suitable**

New PPP Challenge

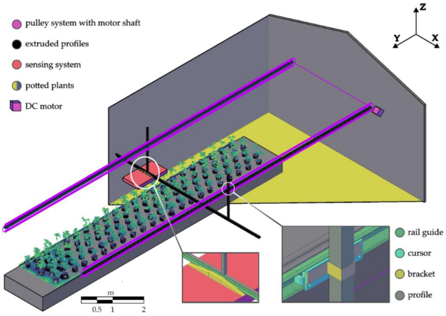
- **Unique symptoms**: Unpredictable for new products
- **Small datasets**: Few hundred plants typical
- **No pre-training**: Symptoms not cataloged

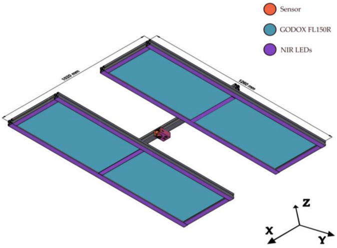
Methodology: Workflow Overview



Workflow Steps

Hardware Platform: Controlled Greenhouse System





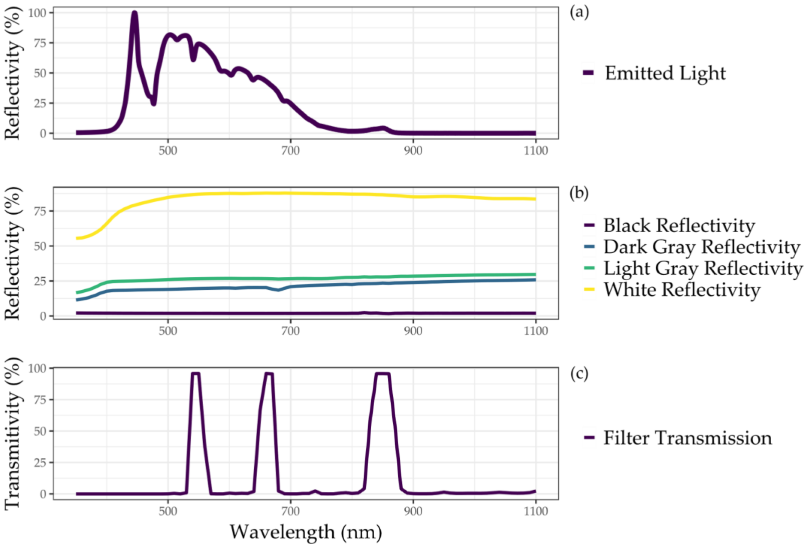
Platform Components

- **MAPIR Survey3W**: Multispectral camera (Green, Red, NIR)
- **LED lighting**: GODOX FL150R panels + 850nm NIR strip
- **Motion system**: DC motor pulley system (0.08 m/s)
- **Positioning**: 3-axis adjustable (1.1-1.5m height)

System Specifications

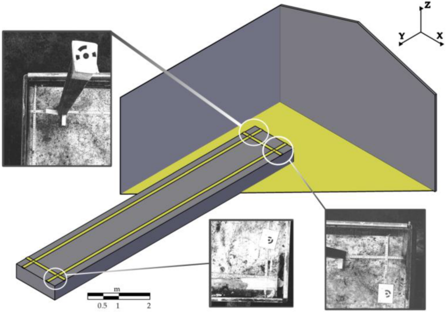
- **Sensor**: 3000$\times$4000 pixels, 1.55 μm pixel size
- **Footprint**: 202-276 cm horizontal
- **GSD**: 0.37-0.69 mm/pixel
- **Overlap**: 95% forward, strip distance = 20cm

Radiometric Calibration and Spectral Validation



Calibration Process

Geometric Validation and Bundle Adjustment



Geometric Control

- **Ground Control Points**: $\geq$9 GCPs per acquisition
- **Height levels**: 0m, 0.35m, 0.7m distribution
- **Metered tapes**: 4 reference tapes for validation
- **Bundle adjustment**: Agisoft Metashape 2.1.0

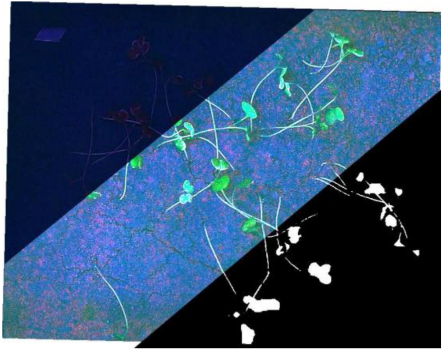
Precision Estimation

Z-coordinate precision:

$$\sigma_z = \frac{H^2}{Bf}\sigma_x$$

- H: camera-target distance
- B: baseline (0.2m)
- f: focal length (3.37mm)

Products Generation and Plant Segmentation



Generated Products

- **DSM**: Digital Surface Model from point cloud
- **MSO**: Multi-Spectral Orthomosaic (calibrated)
- **Vegetation Mask**: Isolated plant pixels

Segmentation Process

- **Bimodal thresholding**: Otsu method on Green band
- **Refinement**: Semi-automatic GrabCut technique
- **Output**: Binary mask separating plants from soil
- **Quality control**: Manual validation and correction

Experimental Design and Observations

Experimental Setup

- **Crop**: Oilseed rape (OSR) in greenhouse
- **Sample size**: 44 pots (40×30 cm each)
- **Treatment**: Herbicide with unknown mode of action
- **Design**: Multiple concentrations + control group
- **Assessment**: 3 time points (3, 7, 14 DAA)

Table: PHYGEN observations distribution

DAA	0%	13%	38%	63%	88%
3	11	9	8	7	9
7	5	4	15	10	10
14	15	14	9	6	0
Total	**31**	**27**	**32**	**23**	**19**

Data Characteristics

- **Discrete values**: Only 5 PHYGEN levels used
- **Uneven intervals**: 25% between most levels, 13% for first
- **Temporal variation**: Imbalanced distribution over time
- **Total dataset**: 132 multivariate observations

Machine Learning Model: LASSO + Logistic Function

Extracted Predictors (14 variables)

- **Spectral bands**: Red, Green, NIR (μ, σ)
- **Vegetation indices**: NDVI, SAVI (μ, σ)
- **Geometric**: Plant area (LAI proxy), height (μ, σ)
- **Temporal**: Days After Application (DAA)

Model Architecture

Two-stage approach:

- **Stage 1**: LASSO regression with L_1 regularization
- **Stage 2**: Logistic function on LASSO output
- **Cross-validation**: K-fold (K=10) for hyperparameter tuning
- **Split**: 80% training, 20% testing (stratified)

Table: Model equations and objectives

Model	Description
LASSO	RSS + L_1 penalty regularization
Logistic	Nonlinear least squares with sigmoid curve

Advantages for New PPPs

- **Small dataset suitable**: Only 132 observations needed
- **Regularization**: Prevents overfitting with limited data
- **Interpretability**: Physical meaning of predictors maintained

Results: Measurement Errors and Model Stability

Geometric Assessment

- **Precision**: Sub-millimeter accuracy achieved
- **X-axis MAE**: 0.57-0.67 mm across all DAA
- **Y-axis MAE**: 0.61-0.70 mm across all DAA
- **Z-axis MAE**: 0.62-0.91 mm across all DAA

Model Stability (10-fold CV)

- **LASSO coefficients**: All CV < 0.25
- **DAA parameter**: CV $= 0.17$ (very stable)
- **NDVI parameter**: CV $= 0.14$ (excellent)
- **Area parameter**: CV $= 0.13$ (excellent)

Radiometric Calibration

- **Best performance**: White reference (4.1-18.1% MAPE)
- **Challenging**: Black reference (76.7-119.6% MAPE)
- **NIR band**: Highest variability across targets
- **Overall**: Acceptable for relative measurements

Key Findings

- **Geometric accuracy**: Sub-millimeter precision achieved
- **Model stability**: Low coefficient of variation (<0.25)
- **Logistic parameters**: Very stable (CV $<$ 0.1)

Results: Model Performance and Accuracy

Performance Metrics

Model	MAE (%)	R^2
LASSO	11.77 ± 0.67	-
LASSO + LF	$\mathbf{10.66 \pm 0.83}$	$\mathbf{0.9 \pm 0.03}$

Benchmark Comparison

- **Human raters**: 10% accepted error threshold[1]
- **Our model**: 10.66% MAE ($\approx$ human performance)
- **SOTA CNN**: 6.74% MAE (but requires huge datasets)
- **Correlation**: $R^2 = 0.9$ (excellent agreement)

[1] Chiang, K.-S.; et al. Effects of rater bias and assessment method on disease severity estimation. *Plant Dis.*, 100:2530–2538, 2016

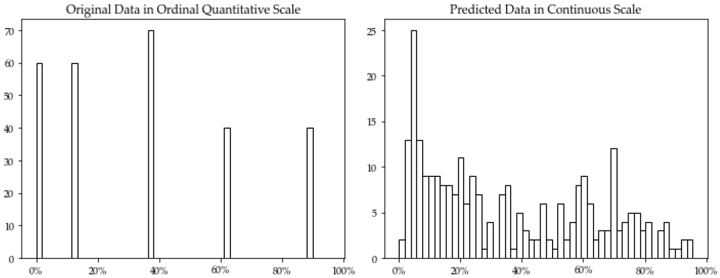
Performance Analysis

- **Accuracy achievement**: Meets EPPO standards for new PPPs
- **Small dataset advantage**: Only 132 observations vs. thousands for CNNs
- **Stability proven**: Robust across different training samples
- **Practical applicability**: Suitable for operational new PPP screening

Operational Context

- **Target application**: New PPP screening (limited data)
- **Infrastructure**: Greenhouse platform required
- **Skills needed**: Photogrammetry + AI expertise
- **Processing time**: Automated after setup

Results: ANOVA Compliance and Statistical Advancement



Statistical Advancement

- **Input**: Discrete ordinal PHYGEN scores (left)
- **Output**: Continuous PHYGEN estimates (right)
- **Compliance**: Now suitable for parametric ANOVA testing
- **Benefit**: Enables proper statistical

Statistical Theory Resolution

Stevens Scale Problem Solved[1]:

- **Before**: Ordinal data → Non-parametric tests only
- **After**: Continuous data → ANOVA, t-tests enabled
- **Implication**: Proper variance analysis for PPP trials
- **Geostatistics**: Compatible with LMM spatial analysis

Conclusions: Automated PHYGEN Assessment for New PPPs

Key Achievements

- **Accuracy**: 10.66% MAE, $R^2 = 0.9$ (comparable to human raters)
- **Small dataset capability**: Only 132 observations vs. thousands for CNNs
- **Stability**: Robust model performance across different training samples
- **Statistical compliance**: Converts ordinal to continuous data for ANOVA

Methodological Contributions

- **Controlled environment**: Reduced sensor and environmental variability
- **Multispectral photogrammetry**: Geometric + spectral feature integration
- **LASSO + Logistic model**: Optimized for limited training data scenarios
- **Validation framework**: Comprehensive error assessment and stability testing

Impact for EPPO Framework

Enables geostatistical analysis: Continuous PHYGEN scores are now

Georeferencing Rankings (Ordinal Variable)

	Type of Variable	Measurement	Ranking	Scoring
	Continuous not limited	X		
	Continuous limited	X		
	Discrete	X		
	Ordinal		X	X
→	Nominal			X
→	Binary			X

Georeferencing Classifications:

- **Nominal and Binary** variables are fundamental for report the presence of pathogens sympthoms (e.g. PP 1/5 (3) Efficacy evaluation of fungicides against Venturia inaequalis and V. pirina).
- the **Case Study**: Anomaly detection and clusterization on encoded images of healthy/sympthomatic plant organs.

Anomaly Detection Approaches

Neural Network Encoders for Plant Disease Detection Across Laboratory and Field Conditions

[Anomaly Detection Overview Figure]

Introduction: The Anomaly Detection Paradigm

The Challenge

- Plant diseases cause **20-40% global crop losses**[1]
- Traditional detection: labor-intensive, subjective[2]
- Supervised ML: requires extensive labeled data[3]
- Performance gap: lab vs. field conditions[4]

[1] Savary et al. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.*, 3:430–439, 2019

[2] Barbedo, J.G.A. Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.*, 172:84–91, 2018

[3] Vallabhajosyula et al. Novel hierarchical framework for plant disease classification. *Appl. Sci.*, 14:3721, 2024

[4] Toda & Okura. How convolutional neural networks diagnose plant disease. *Plant Methods*, 15:9, 2019

Anomaly Detection Advantages

- **Minimal labeling**: Only healthy samples needed
- **Novel disease detection**: Identifies unseen pathogen symptoms[6]
- **Agricultural alignment**: Healthy plants = majority class
- **Practical deployment**: Resource-efficient approach

[6] Bumbaca et al. Supporting screening of new plant protection products. *Agronomy*, 14:306, 2024

Study Workflow and Methodology

[Study Workflow Diagram]

Feature Extraction $\rightarrow$ *Dimensionality Reduction* $\rightarrow$ *Anomaly Detection* $\rightarrow$ *Evaluation*

Evaluation Strategies

1. **Anomaly Detection**: Healthy vs. All diseases
2. **Clustering Classification**: Disease-specific grouping

Analysis Pipeline

1. Feature extraction (56 encoders)
2. Dimensionality reduction (4 methods)
3. Anomaly detection (3 algorithms)
4. Performance evaluation

Datasets: Laboratory vs. Field Conditions

Table: Apple leaf disease datasets

Dataset	Samples	Size	Environment
Plant Village[1]			
Healthy	516	256×256	Laboratory
Cedar rust	275	256×256	Controlled
Apple scab	583	256×256	Uniform bg
Plant Pathology[2]			
Healthy	516	Variable	Field
Cedar rust	275	Variable	Natural
Apple scab	583	Variable	Complex bg

[Dataset Comparison Placeholder]

Dataset	Plant Village	Plant
Environment	Laboratory	
Conditions	Controlled	N
Quality	High	V

Key Differences

[1] Hughes & Salathé. An open access repository of images on plant health. *arXiv:1511.08060*, 2015

[2] Thapa et al. The Plant Pathology 2020 challenge dataset. *arXiv:2004.11958*, 2020

- **Laboratory**: Controlled lighting,

Dataset Characteristics

- **Balanced classes**: Identical sample counts
- **Disease spectrum**: Healthy, fungal infections
- **Real-world challenge**: 5-10% performance gap expected

Convolutional Neural Network Encoders

Traditional CNN Families

- **ResNet family**: Deep residual learning[1]
  - ResNet18, 34, 50, 101, 152
  - Skip connections for gradient flow
- **DenseNet family**: Dense connectivity
  - DenseNet121, 161, 169, 201
  - Feature reuse and efficiency
- **VGG family**: Sequential architecture
  - VGG11, 13, 16, 19
  - Foundation CNN design

Efficient CNN Families

- **EfficientNet family**: Compound scaling
  - EfficientNet-B0 to B7
  - Optimal width/depth/resolution
- **MobileNet family**: Depthwise separable
  - MobileNet v2, v3
  - Mobile deployment optimized
- **ShuffleNet family**: Channel shuffling
  - ShuffleNet v2 variants
  - **Best performer**: 2.3M parameters

Performance Insight

[1] Faster R-CNN: Towards real-time object detection

Vision Transformer Encoders

Vision Transformer Family

- **ViT (Vision Transformer)**[1]
  - ViT-Base, Large variants
  - Self-attention mechanism
  - 86M+ parameters
- **Swin Transformer**: Hierarchical structure
  - Shifted window attention
  - Multi-scale feature extraction
- **ConvNeXt**: CNN-Transformer hybrid
  - Modernized ConvNet design
  - Transformer-inspired blocks

DINOv2 Family

- **DINOv2**: Self-supervised learning
  - DINOv2-Small to Giant
  - Up to 300M parameters
  - Rich semantic representations
- **Key advantages**:
  - No fine-tuning required
  - Strong feature extraction
  - General-purpose encoders

Surprising Finding

Large transformers (300M params) underperformed lightweight CNNs (2.3M) in field conditions

[1] Vaswani et al. Attention is all you need. *NIPS*, 2017.

Feature Extraction Methodology

[Feature Extraction Methodology]

Encoder → *Feature Maps* → *Global Average Pooling* → *Feature Vector*

Extraction Process

1. **Input**: RGB images (224×224)
2. **Encoder**: Pre-trained ImageNet weights
3. **Output**: High-dimensional features
4. **No fine-tuning**: Off-the-shelf performance

Key Advantages

- **No training required**: Immediate deployment
- **Computational efficiency**: Single forward pass
- **Scalability**: Easy architecture comparison
- **Robustness**: Pre-trained stability

Feature Characteristics

- **Dimensionality**: 512-2048 features

Evaluation Strategy

Evaluation Strategies

Strategy 1: Anomaly Detection

- **Training**: Only healthy samples
- **Testing**: Healthy vs. All diseases
- **Goal**: Identify any pathological condition
- **Metrics**: Precision, Recall, F1-score, AUC

Practical application: Early disease screening in agricultural monitoring

Anomaly Detection Pipeline

Strategy 2: Clustering Classification

- **Training**: Unsupervised clustering
- **Testing**: Disease-specific grouping
- **Goal**: Distinguish between disease types
- **Metrics**: Adjusted Rand Index, Silhouette

Practical application: Disease type identification for targeted treatment

Clustering Pipeline

1. Feature extraction

Dimensionality Reduction Techniques

Linear Methods

- **Principal Component Analysis (PCA)**
  - Linear dimensionality reduction
  - Variance maximization
  - Interpretable components
- **Linear Discriminant Analysis (LDA)**
  - Supervised dimensionality reduction
  - Class separation optimization
  - Limited to n-1 dimensions

Non-linear Methods

- **t-SNE (t-Distributed Stochastic Neighbor Embedding)**
  - Non-linear manifold learning
  - Local structure preservation
  - **Best performer** across datasets
- **UMAP (Uniform Manifold Approximation)**
  - Topology preservation
  - Global structure maintenance
  - Faster than t-SNE

Advantages

Key Finding

t-SNE consistently outperformed linear methods for both anomaly

Anomaly Detection Algorithms

Local Outlier Factor (LOF)

- **Principle**: Density-based anomaly detection[1]
- **Method**: Local density comparison
- **Advantages**:
  - Handles varying densities
  - Local anomaly assessment
  - Robust to cluster shapes
- **Best performance**: Most stable across datasets

[1] Breunig et al. LOF: identifying density-based local outliers. *SIGMOD*, 2000

LOF Score Interpretation

One-Class SVM

- **Principle**: Support vector boundary[2]
- **Method**: Maximum margin separation
- **Advantages**:
  - Kernel flexibility
  - Theoretical foundation
  - Memory efficient

[2] Schölkopf et al. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13:1443–1471, 2001

Isolation Forest

- **Principle**: Path-based isolation[3]
- **Method**: Random tree partitioning

Experimental Setup and Evaluation Metrics

Experimental Design

- **Cross-validation**: 5-fold stratified CV
- **Training**: 80% healthy samples only
- **Testing**: All samples (healthy + diseases)
- **Repetitions**: 10 independent runs
- **Statistical testing**: Wilcoxon signed-rank

Evaluation Metrics

Anomaly Detection:

- Precision, Recall, F1-score
- Area Under ROC Curve (AUC)
- Average Precision (AP)

Clustering:

- Adjusted Rand Index (ARI)
- Silhouette Score
- Normalized Mutual Information

Hyperparameter Optimization

- **Grid search**: Algorithm-specific parameters

Performance Analysis

- **Lab vs. Field**: Direct comparison
- **Architecture ranking**:

Results: Anomaly Detection Performance

Table: Top anomaly detection performance

Encoder	Lab F1	Field F1	Gap
ShuffleNet_v2	0.892	0.834	-6.5%
MobileNet_v3	0.885	0.821	-7.2%
EfficientNet-B0	0.878	0.815	-7.2%
ResNet34	0.874	0.808	-7.6%
DenseNet121	0.871	0.805	-7.6%
Large Models			
DINOv2-Giant	0.856	0.782	-8.6%
ViT-Large	0.851	0.776	-8.8%

Key Findings

- **5-10% performance drop** from lab to field

[Anomaly Detection Performance]

Best F1-scores: ShuffleNet (0.892), MobileNet (0.885), EfficientNet (0.878)

Performance Analysis

- **Consistent ranking**: ShuffleNet_v2 top performer
- **Efficiency advantage**: 2.3M vs. 300M parameters
- **Field robustness**: Smaller models less sensitive
- **t-SNE reduction**: Best dimensionality technique

Results: Clustering-based Disease Classification

[Clustering Visualization]

t-SNE projection showing disease group separation

Table: Clustering performance (ARI scores)

Method	Lab ARI	Field ARI	Algorithm
ShuffleNet+DBSCAN	0.745	0.682	
MobileNet+K-means	0.738	0.675	
EfficientNet+DBSCAN	0.731	0.669	
ResNet34+DBSCAN	0.728	0.664	
Large Models			
DINOv2+K-means	0.712	0.645	
ViT+DBSCAN	0.708	0.641	

Disease Group Characteristics

- **Healthy clusters**: Tight, well-defined boundaries
- **Cedar rust**: Distinct spectral signatures
- **Apple scab**: Variable manifestations
- **t-SNE projection**: Optimal class separation

Clustering Insights

- **DBSCAN superiority**: Density-based clustering excels
- **Disease separation**: Clear

Validation Metrics

Discussion: Implications and Challenges

Key Findings

- **Lightweight superiority**: Small models outperform large ones
- **Performance gap**: 5-10% reduction lab-to-field
- **Method stability**: LOF most robust anomaly detector
- **Feature quality**: t-SNE essential for performance

Practical Advantages

- **Minimal labeling**: Only healthy samples required

Challenges Addressed

- **Environmental variability**: Field condition robustness
- **Computational constraints**: Mobile deployment feasibility
- **Labeling costs**: Reduced annotation requirements
- **Scalability**: Cross-crop applicability potential

Agricultural Integration

- **Early warning systems**: Automated disease screening
- **Precision agriculture**: Targeted intervention

Conclusions: Anomaly Detection for Agricultural Monitoring

Major Contributions

1. **Comprehensive evaluation**: 56 neural network encoders
2. **Practical insights**: Lightweight models excel in field conditions
3. **Methodological framework**: Robust anomaly detection pipeline
4. **Performance benchmarks**: Lab-to-field transition analysis

EPPO Integration Potential

- **Pathogen symptom localization**: Spatial anomaly mapping
- **Geostatistical integration**: Disease distribution analysis
- **EPPO workflow enhancement**: Digital technology adoption[1]
- **Regulatory compliance**: Standardized assessment protocols

[1] EPPO. PP 1/333(1) - Adoption of Digital Technology for Data Generation for the Efficacy Evaluation of Plant Protection Products. *EPPO Bull.*, 55:14–19, 2025

Agricultural Impact