

Abstract

This work was aimed at developing a prototype system based on multispectral digital photogrammetry to support tests required by international regulations for new Plant Protection Products (PPPs). In particular, the goal was to provide a system addressing the challenges of a new PPP evaluation with a higher degree of objectivity with respect to the current one, which relies on expert evaluations. The system uses Digital Photogrammetry, which is applied to multispectral acquisitions and Artificial Intelligence (AI). The goal of this paper is also to simplify the present screening process, moving it towards more objective and quantitative scores about phytotoxicity. The implementation of an opportunely trained AI model for phytotoxicity prediction aims to convert ordinary human visual observations, which are presently provided with a discrete scale (forbidding a variance analysis), into a continuous variable. The technical design addresses the need for a reduced dataset for training the AI model and relating discrete observations, as usually performed, to some proxy variables derived from the photogrammetric multispectral 3D model. To achieve this task, an appropriate photogrammetric multispectral system was designed. The system operates in multi-nadiral-view mode over a bench within a greenhouse exploiting an active system for lighting providing uniform and diffuse illumination. The whole system is intended to reduce the environmental variability of acquisitions tending to a standard situation. The methodology combines

advanced image processing, image radiometric calibration, and machine learning techniques to predict the General Phytotoxicity percentage index (PHYGEN), a crucial measure of phytotoxicity. Results show that the system can generate reliable estimates of PHYGEN, compliant with existing accuracy standards (even from previous PPPs symptom severity models), using limited training datasets. The proposed solution addressing this challenge is the adoption of the Logistic Function with LASSO model regularization that has been shown to overcome the limitations of a small sample size (typical of new PPP trials). Additionally, it provides the estimate of a numerical continuous index (a percentage), which makes it possible to tackle the objectivity problem related to human visual evaluation that is presently based on an ordinal discrete scale. In our opinion, the proposed prototype system could have significant potential in improving the screening process for new PPPs. In fact, it works specifically for new PPPs screening and, despite this, it has an accuracy consistent with the one ordinarily accepted for human visual approaches. Additionally, it provides a higher degree of objectivity and repeatability.

2.2.1 Introduction

Researchers in the field of Plant Protection Products (PPPs) need to bridge the gap between evaluations from traditional human-based approaches and those enabled by Artificial Intelligence (AI) [22]. Specifically, new PPPs undergo a rigorous safety screening before market entry. PPP developers must meticulously formulate and dose these PPPs to avoid harmful phytotoxic effects on crops, thus maintaining selectivity [21]. Traditionally, experimenters assess the severity of phytotoxicity through visual observations. The reliability of these assessments depends on low variability among experimenters' observations and proper rating scales [17]. In Europe, technicians are required to operate according to Good Experimental Practice (GEP), which is based on international laws [2]. GEP is a set of standards that ensures objectivity and precision in scientific experiments. The World Trade Organization Agreement on Sanitary and Phytosanitary Measures [9] designates the International Plant Protection Convention (IPPC) as the authority for plant health standards [45]. The European Union falls under the European and Mediterranean Plant Protection Organization (EPPO) within IPPC. EPPO is responsible for setting phytosanitary and PPP standards. EPPO standards address crop selectivity [5] by providing evaluation methods involving both discrete and continuous values. However, experimenters often prefer using quantitative ordinal discrete scales due to their practicality [18]. As observed by Chiang et al. [17], percentage scales with intervals of 10% can reduce rater uncertainty. That is because 10% is commonly accepted as inter-rater error. This can

potentially lead to inconsistencies with theoretical assumptions in variance analysis [48, 8]. Nevertheless, the selectivity of PPPs is inherently a continuous variable, assumed to be inversely proportional to the percentage of phytotoxicity symptoms and their intensity. According to EPPO, phytotoxicity symptoms include (i) modifications in the development cycle, (ii) thinning, (iii) modifications in color, (iv) necrosis, (v) deformation, and (vi) effects on quantity and quality of the yield [5]. General Phytotoxicity (PHYGEN) is an aggregate indicator that summarizes the above symptoms by defining the percentage of damage to a plant compared to a perfectly healthy reference plant [44].

Imaging sensors have already been demonstrated to improve precision and objectivity in the detection of pathological symptoms [18, 39]. Some spectral properties of plants, as recorded through multispectral sensors [38] are recognized as indicators of photosynthetic efficiency [25, 16]. Various methods, including multi-view approaches [46, 35, 53], can be used to create 3D models of plants [39]. Spectral and geometric features of plants can be used to virtually reproduce the plant appearance, as observed by an experimenter during assessment. When working with three-dimensional and multispectral data, a summary is necessary to obtain an accurate estimate of PHYGEN, like a direct human-based evaluation approach. Machine learning (ML) models from artificial intelligence (AI) can synthesize vast amounts of digital information in a robust and reasonable manner when guided by expert (low variation) experimenter annotations [38]. Open platforms offer large labeled training datasets, allowing users to customize ML algorithms to their re-

quirements [32, 29] Convolutional Neural Networks (CNNs) were found to be the most accurate method for symptom classification [50, 41] while working with image-based data. CNNs were shown to be capable of rating EPPO symptoms, specifically “modifications in color”, at both leaf and canopy levels [26]. Gómez-Zamanillo et al. [28] proposed a method for assessing PHYGEN by classifying the most common symptoms. Their study demonstrated the effectiveness of CNNs as feature extractors for predicting PHYGEN rates or similar measures. The study utilized CNN to identify and classify color-related phytotoxicity symptoms from RGB images. Severity estimates were determined by assigning arbitrary weights to the detected symptoms. Rather, they relied on expert experimenters to quantify weights without optimizing scores. Currently, no CNN-based model has been proposed to generate a reasonable estimate of PHYGEN based on a comprehensive analysis of all symptoms. Weight optimization is highly appreciated as it is expected to enhance the accuracy of estimates and provide insights into the significance of each symptom in the toxicological mechanism of PPPs. Further challenges associated with the deployment of CNNs for plant disease detection and scoring are reported in Barbedo et al. [13, 12]. In particular, these include (i) sensitivity of deductions to environmental and sensor-related issues, (ii) capability of generalization of the model, and (iii) training dataset quality. It is important to note that the quality of the training dataset is highly significant as it must be properly calibrated for the specific type of PPP being tested. Therefore, pre-trained networks relying on training datasets generated for different symptoms from different PPPs should not be used to test

new PPPs. It is worth noting that, in order for CNN training to be robust and accurate enough, it requires huge training datasets consisting of thousands of images. Table 1 shows some of the methods proposed in the literature for the estimation of PHYGEN, enhancing their suitability for new PPPs PHYGEN prediction.

Table 1: Related works.

Paper	Method	Accuracy ¹	Suitability ²
Human raters	Depending on the rater, the method recommended maximum error is 10% [17]	Traditional	-
Ali et al. [10]	Image processing no AI involved, and no monitorable stability	Not reported	Involving only biomass estimation,
Chu et al. [19]	Shallow CNN	80%	Destructive and only spectral signature involved
Ghosal et al. [26]	CNN	From 50% to 90% depending on rater	Not phytotoxicity-specific, destructive
Gómez-Zamanillo et al. [28]	CNN	93.26%	Not suitable for new PPPs because of the amount of training data required

¹ It indicates the accuracy of phytotoxicity severity with respect to human raters.

² For new PPPs PHYGEN screening.

Typical trials for new PPPs usually involve only a few hundred plants. This may not provide a sufficient dataset for robust training, testing, and deployment of a new CNN. It is noteworthy that CNNs maintain their efficacy when symptoms of phytotoxicity are well-documented and recognized within the training dataset. This specificity is a true challenge in ML optimization for the newer PPP-related trials since the explored symptomatology may not be cataloged. This work emphasizes that symptoms of phytotoxicity resulting from new PPPs can be unique due to their novelty, making them unpredictable. Therefore, screening trials are necessary. The proposed method involves a PHYGEN evaluation via a CV ML system for new PPPs operating in a greenhouse environment that overcome such limitations. The system is specifically designed to address three key challenges in adopting AI, and specifically CV ML for new PPPs screening: small amount of training data, stability, and accuracy. Moreover, the model prediction suitability for ANOVA testing is also discussed. The presented method requires only a small training sample with respect to CNN algorithms because it relies on a single linear regression and a logistic function. It takes a small training sample from the available study population, effectively addressing issues of under-representation of training datasets [13], which is typical when testing new PPP phytotoxicity.

The system was found to reduce the impact of environmental and sensor-related factors on plant symptom detection, increasing the stability of plant pictures and measures. This is achieved through proper platform calibration techniques and a multi-view image capture approach that allows for the monitoring of errors of the geo-

metrical and radiometric measures used to train and test the model. Model stability was tested using cross-validation. The results confirmed the robustness of the method regardless of the sample adopted. The accuracy of the model's prediction was compared to the precision of human raters as described in the literature (10%) [17] and to the state-of-the-art (SOTA) model for PHYGEN of non-new PPPs (6.74%) [28]. It was not possible to find a direct comparison of a model predicting PHYGEN for new PPPs by CV ML in the literature. Therefore, the accuracy must be considered satisfactory if it is higher than the precision of human raters, and it is expected to be lower than that of CNN models with a greater amount of training data. The methodology also addresses the challenge of adopting discrete quantitative scales in the ML training step. It has been shown to improve the prediction of PHYGEN as a continuous scale variable, starting from quantitative ordinal discrete values, such as those obtained from ordinary approaches. Furthermore, as the PHYGEN estimates are now on a continuous scale, the ANOVA test can be more appropriately utilized, resolving the cumbersome lack of adaptation to the statistical theory that is often observed in the field of PPP screening.

2.2.2 Materials and Methods

2.2.2.1 Hardware Platform

A platform was developed and integrated into a greenhouse structure for multispectral photogrammetric data acquisition. The integration was achieved using a framework consisting of two 10-m-long aluminum extruded profiles affixed to the roof and walls of the greenhouse. To enable the sensing system to move along the Y-axis, two parallel linear rail guides were mounted on these profiles. In addition, a 6-m-long aluminum support was installed perpendicular to the initial rails. This support incorporates a linear guide rail, which enables camera movement along the X-axis. Adjustments along the Z-axis were made possible by altering the brackets on the Y-axis rails. The proximity of the sensing system to the bench, where the pots were situated, was adjustable within a range of 1.1 to 1.5 m. The camera's position along the 6-m rail could be adjusted using fixing brackets, as shown in fig. 1.

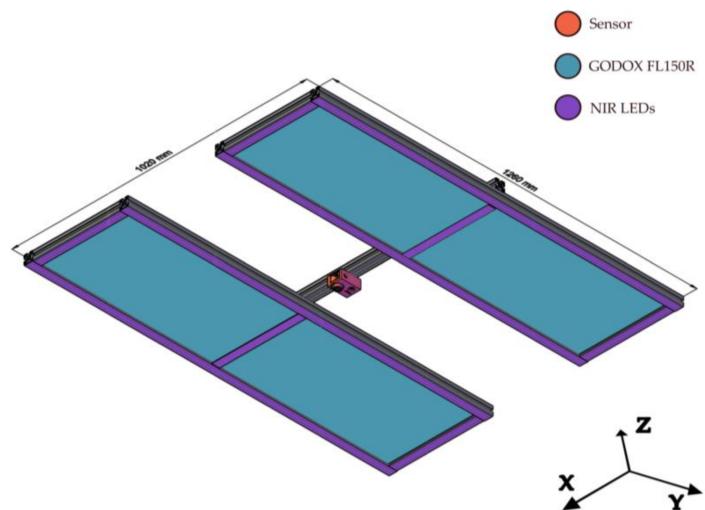
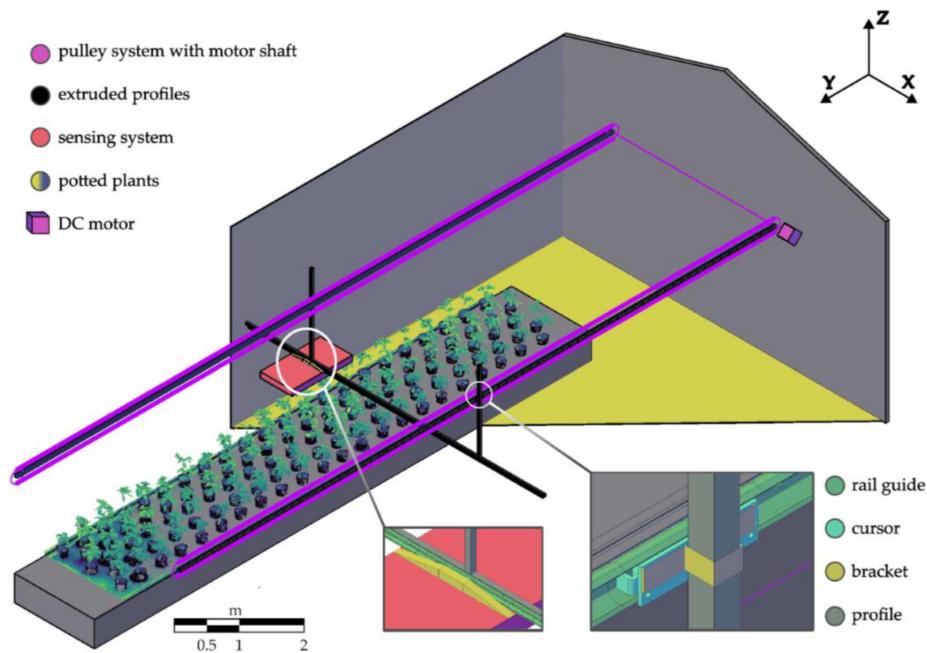


Figure 1: Platform and sensing system (top) and only the sensing system (bottom).

Camera movement along the Y-axis in the greenhouse was controlled by a DC motor that operates through a pulley system. This

system works similarly to a bridge crane that moves the imaging compound automatically with a speed of about 0.08 m/s along the Y-axis. The motion was manually activated and stopped. The whole moving platform was made of (i) one MAPIR Survey3W (PeauProductions, San Diego, CA, USA) camera (S3) multispectral camera, (ii) two Light-Emitting Diode (LED) panels (GODOX FL150R) (Godox, Shenzhen, Guangdong, China) each measuring 1.2×0.3 m, and (iii) a 6-m LED strip emitting with a peak at 850 nm that encircles the GODOX FL150R panels to ensure that adequate Near-Infrared (NIR) radiation reaches the plants. Panels (the entire imaging system) ran parallelly to the bench hosting the pots to be imaged to ensure uniform illumination. Furthermore, shading curtains were installed on the walls and ceiling of the greenhouse to reduce exterior light contribution during data collection. A preliminary test was conducted to ensure the consistency of the spectrum provided by LEDs through its comparison with the reflectance spectrum acquired by an RS-5400 Spectroradiometer (Spectral Evolution, Haverhill, MA, USA). The acquisition was performed using calibrated panels of the RS-5400 instrument fig. 2 in lighting conditions replicating the operational environment.

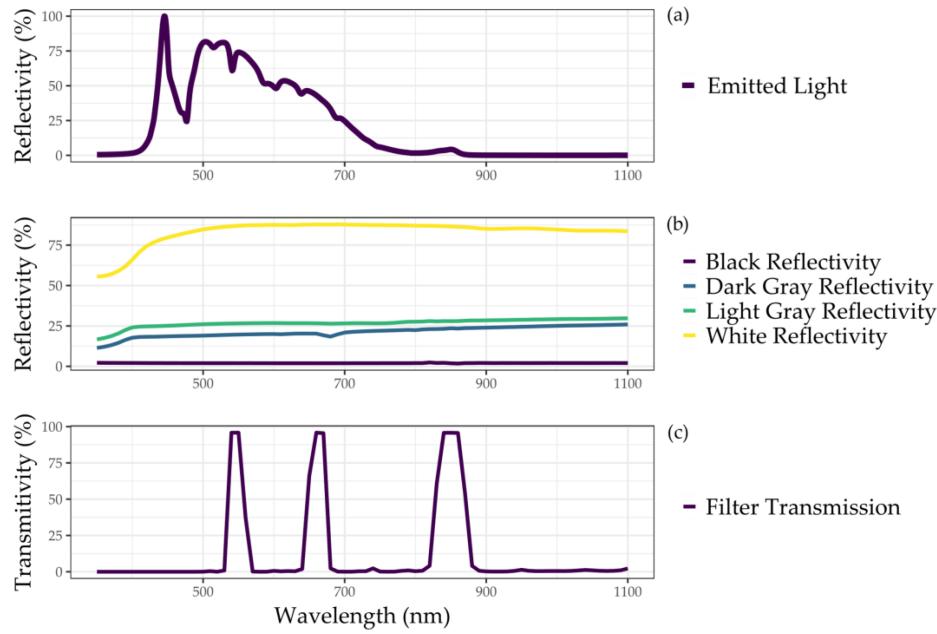


Figure 2: (a) Spectral signature of the reference panel, lighted with the tested LEDs and measured using the RS-5400 Spectroradiometer. (b) Reflectivity of the MAPIR calibration panels corresponding to the different grayscale levels (yellow, light-green, blue, and violet colors in the graph) provided by the factory. The dark green line shows the filter sensitivity of MAPIR for the different bands. (c) Transmissivity of the S3 camera filter.

The S3 camera was used for image capture, as detailed in table 2. A white balance setting was employed during acquisition to increase the intensity of the Red and NIR bands, resulting in a reduction of green band sensitivity.

Table 2: S3 and system integration specifics.

Parameter	Value
Focal Length	3.37 mm
Aperture	f/2.8 (fixed)
Lens Distortion	<1%
Focal Length	Fixed
Hyper-focal Distance	81.5 cm
Sensor Size	3000 × 4000 pixels
Pixel Physical Size	1.55 μm
Bands	Green, Red, and NIR (Figure 2c)
Camera Shift (Y-axis)	20.3 cm per shot
Frames per second	1/3
Horizontal Footprint ¹	202–276 cm
Vertical Footprint ¹	152–207 cm

1.1–1.5 m distance.

2.2.2.2 Experimental Design

An experiment was conducted to assess the reliability of the system and the processing workflow with respect to EPPO standards. The selectivity of a herbicide with an unknown mode of action was tested in a controlled environment greenhouse following EPPO standards [5, 4, 6, 7]. This allowed uniform growing conditions to be maintained throughout the greenhouse. Forty-four pots, each 40 × 30

cm, were sown with oilseed rape (OSR) and treated with the experimental product before emergence. The treatments were applied using an automatic spray chamber. To ensure a balanced set of PHYGEN, different concentrations of the herbicide, including a control group, were used to cover a range of phytotoxicity intensities. Visual and digital evaluations were carried out simultaneously. The PHYGEN assessment values were recorded as Day After Treatment (DAA) in table 3.

Table 3: PHYGEN observations.

DAA ¹	0%	13%	38%	63%	88%
3	11	9	8	7	9
7	5	4	15	10	10
14	15	14	9	6	0
TOT	31	27	32	23	19

¹ Days After Application.

Only five discrete PHYGEN values were retained for scoring: 0%, 13%, 38%, 63%, and 88%. This emphasizes the nature of the data generated by the visual assessment and the extreme use of the discrete quantitative scale. It is important to note that all five values were assigned during the three assessments, except on the last day, when the highest value (88%) was not observed. This resulted in an imperfectly balanced distribution of PHYGEN over time. The interval between consecutive discrete values was 25%, except for the interval between 0% and 13%. The 0% value may be unreliable for treated pots due to the inevitable effect of herbicides, even for re-

sistant crops. The true value in the range between 0% and 13% is difficult to detect, even for expert experimenters upon visual inspection, and is usually interpreted as having no effect on the harvest. Despite this, 0% values were always recorded, as assessed by the experimenters.

2.2.2.3 Data Processing

The workflow starts with planning the image acquisition of the experimental plants. Then, the images are used to retrieve the multispectral 3D reconstruction of the plants. The parameters of the observed plants are extracted by the 3D model. Finally, the ML model is trained on the extracted parameters and validated. The workflow is summarized in fig. 3.

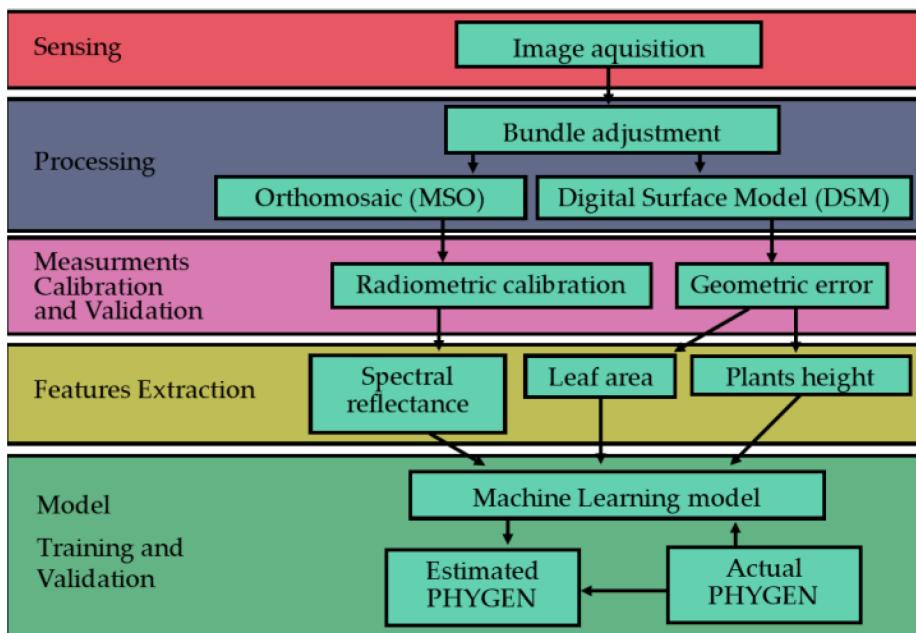


Figure 3: General workflow of the suggested method.

2.2.2.4 Planning the Acquisition

The camera movements were planned to capture stereoscopic images using a local Euclidean coordinate system, hereinafter called Coordinate Reference System (CRS), having the origin located at the lower left corner of the bench hosting plants. Image block bundle adjustment was intended to refine both position and attitude image Exterior Orientation (EO) parameters, using nominal coordinates of the focal point position and a nadiral orientation as an initial solution during the adjustment. Nominal values for image focal point position (X_0 , Y_0 , Z_0) were determined assuming (i) X_0 as the horizontal distance between adjacent strips; (ii) Y_0 was computed by considering the speed of the camera shifts along the bars, and (iii) Z_0 was set to a fixed value, which is discussed in the next paragraph. The camera was positioned with the longer side (4000 pixels) aligned across the track. The nominal Z_0 value was determined based on two conditions. First, the resulting image footprint must be consistent with the expected target size of plants. Second, targets should be visible at the smallest distance longer than the hyper-focal distance (0.815 m for S3). This condition ensures the maximum obtainable resolution, known as Ground Sampling Distance (GSD), which maximizes the efficiency and quality of tie point recognition. It is important to note that GSD is proportional to the physical pixel size according to eq. (1).

$$GSD = \delta \frac{H}{f} \quad (1)$$

where H is the camera-to-target distance, f is the camera focal length, and δ is the physical pixel size. As the height of the assessed plants can vary greatly during the same acquisition, H can range from 0.815 to 1.500, resulting in a ground sample distance (GSD) that varies between 0.37 and 0.69 mm·pixel⁻¹.

When planning an acquisition, it is important to ensure that the coarser GSD (which depends on H) is smaller than the smallest feature that needs to be recognized. Tie point recognition depends on both the forward and side overlap among images. The forward overlap is determined by the baseline (B), which is the distance between consecutive focal points along the same strip. On the other hand, the side overlap is determined by the distance between two adjacent strips. The platform is designed to operate with a strip distance equal to the baseline (0.2 m), resulting in 95ln digital photogrammetry, it is widely acknowledged that the Z coordinate of target points is the most critical to estimate accurately. Its precision can be evaluated using eq. (2) [20, 14, 34].

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

where H is the camera-target distance, σ_x is the precision of parallax measures in the image domain (assumed to be half the physical pixel size, i.e., 1.685 µm for S3), B is the baseline, f is the sensor focal length, and σ_z is the estimated precision of the Z coordinate of the target point. The graphs in fig. 4 relate the theoretical (expected) σ_z with the baseline B while varying the camera-to-target distance at three reference values. The B interval was considered to be within

the minimum and maximum overlap.

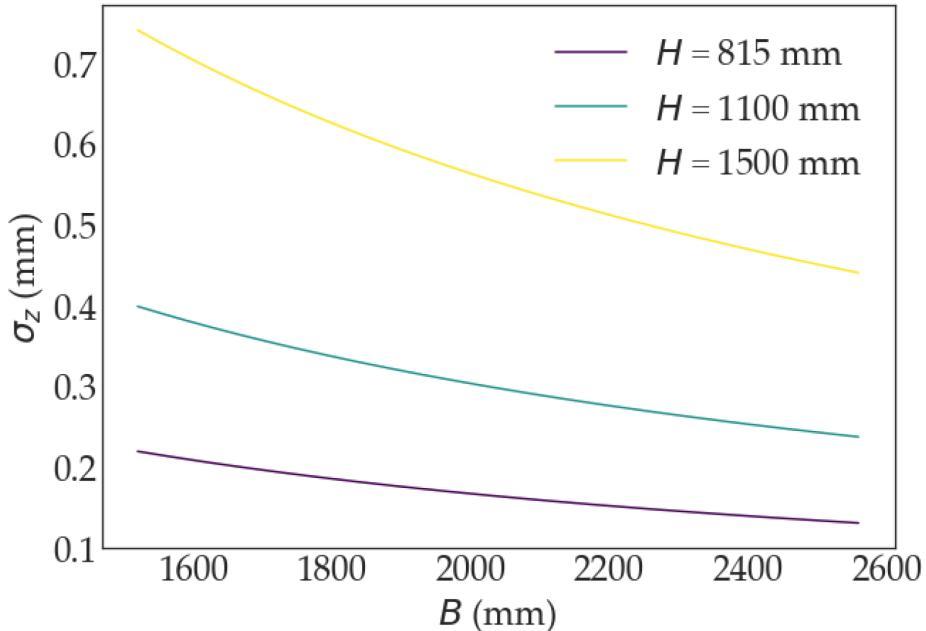


Figure 4: σ_z estimates computed by eq. (2). Colored curves refer to different D values.

eq. (2) was used to estimate the actual Z precision from the bundle adjustment solution and compare it to the expected (theoretical) precision (σ_z). To enhance the robustness of validation and test for geometrical errors, four metered tapes were placed over the bench (fig. 5), and at least nine GCPs were manually positioned throughout the scene for each acquisition date. The GCPs were positioned in a pattern to ensure a uniform distribution across the image block in both longitude and latitude. The GCPs were at three different heights: 0 m, 0.35 m, and 0.7 m.

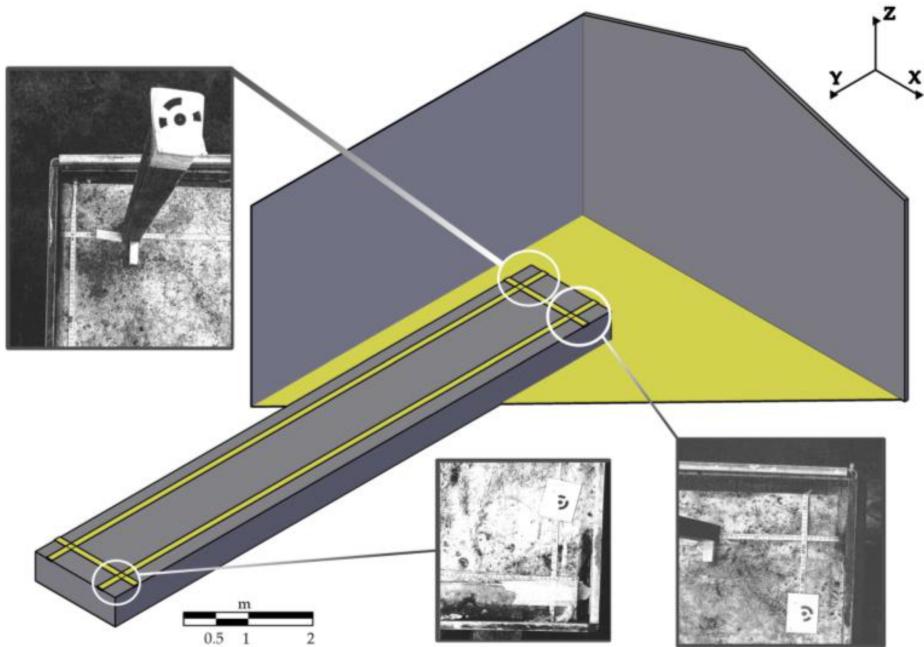


Figure 5: Metered tapes and GCPs on the bench.

In summary, the only adjustable parameters for planning the acquisition were (i) the Z position of the camera and (ii) the distance between strips (side overlap). The Z position was set at 1.5 m from the bench, and the distance between strips (on the X axis) was 20 cm in all three assessments.

2.2.2.5 Bundle Adjustment and Point Cloud Generation

Digital photogrammetry software utilize computer vision algorithms, such as the Scale-Invariant Feature Transform (SIFT), to automatically identify potential tie points in images [42, 36, 37]. Photogrammetric software may use various algorithms to match these points

across images, including Random Sample Consensus (RANSAC) or other methods, depending on computational efficiency and accuracy requirements [3]. After matching the points, software uses bundle adjustment to estimate the spatial locations of the points and the camera positions. This process takes into account the matched points and the camera's Exterior and Interior Orientation (EO/IO) parameters [34, 30, 27]. This study employed tie point identification, matching, and bundle adjustment using Agisoft Metashape version 2.1.0 (Agisoft LLC, St. Petersburg, Russia). To support image bundle adjustment, a portion of the GCPs and initial camera EO/IO parameters were provided [11, 51, 40]. As far as IO parameters are concerned, the initial values used to bootstrap adjustment were the following: (i) focal length as supplied by S3 and (ii) lens distortion parameters = 0, coordinates of the Principal Point of Autocollimation (PPA) equal to the physical center of the image (fiducial point). Sensor array and physical pixel size were set to their nominal values. The solution was spatially referenced using GCP coordinates, which are referred to as a local reference system (CRS). The resulting point cloud associates spectral values from bands to each point. These values were obtained as the mean value of the image pixels corresponding to the target points. Bundle adjustment provides estimated camera EO and IO parameters and their uncertainties, as well as all GCP coordinates estimated by the model and their corresponding errors. The GCPs involved in the bundle adjustment allow for the detection of outliers and refinement of the solution by running the bundle adjustment again after removing the outliers. To ensure accuracy, the solution was checked by three GCPs, which were not

involved in bundle adjustment. The adjustment solution was considered satisfactory if the difference between these three GCP values from the model and the reference values was less than or equal to the expected error.

2.2.2.6 Products

A digital surface model (DSM) with a GSD inherited from the previous steps was generated from the point cloud data. The DSM was then utilized to create the final multi-spectral orthomosaic (MSO) [27]. Both the DSM and MSO are projected in the CRS.

2.2.2.7 Radiometric Calibration of the Multi-Spectral Orthomosaic

MSO radiometric calibration was performed using an empirical line approach with reference reflectance values obtained from the S3 calibrated panel provided by the MAPIR company [1]. The average pixel value from each squared area of the panel having the same grey level was computed for all the bands of the non-calibrated orthomosaic. Reference reflectance values from the MAPIR calibration panel were compared with the averaged ones from the orthomosaic by scatterplot. An Ordinary Least Squares approach was used to calibrate a linear function modeling the relationship between MSO Digital Numbers and the correspondent “expected” reflectance values [15]. Calibration function definition was carried out separately for each band. The resulting functions were then applied to all the

pixels of MSO bands, resulting in a calibrated (reflectance) version of MSO fig. 6.

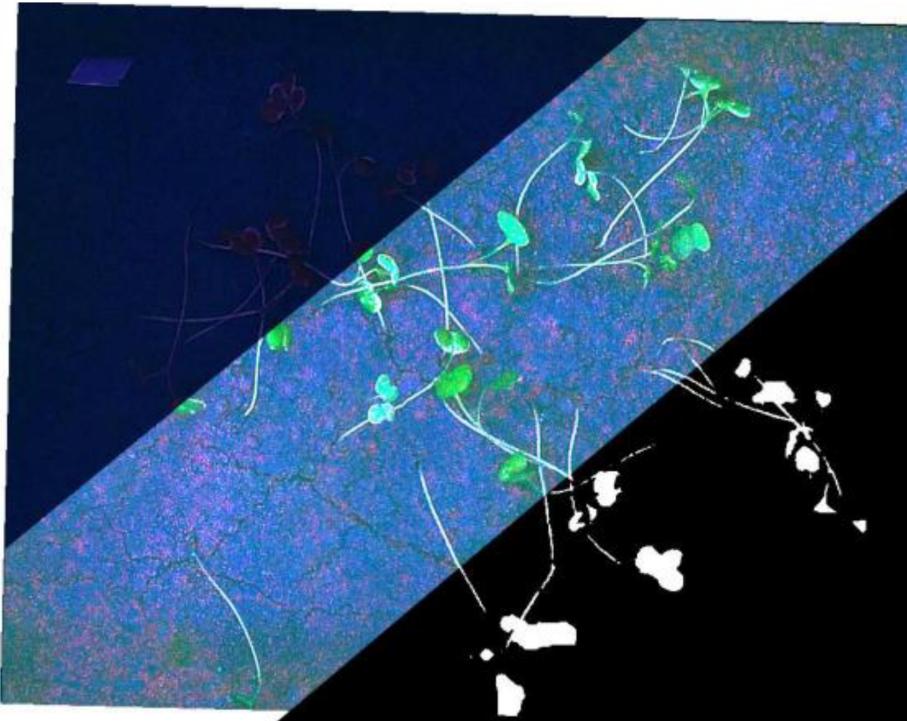


Figure 6: Non-calibrated (dark) and calibrated MSO are shown together with the last vegetation mask (white pixels) on an oil seed rape pot.

The radiometric calibration accuracy was computed as the Mean Absolute Error (MAE) between the panel ground truth values and the forecasted values [52] according to Equation

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (3)$$

where y_i is the expected reflectance value of the i-calibration panel square, x_i the estimated correspondent one, and n the number of

observations.

2.2.2.8 MSO Classification

A vector format file was generated to map the area of each potted plant. A local coordinate system (CRS) was adopted. The file contains two essential pieces of information: a unique identifier for each plant and the date of assessment. The second process involved manually isolating the plant from the soil in the pot using thresholding, focusing only on the plant pixels. The soil was identified and masked by applying a bimodal threshold [43]. to the green band. The mask was then refined using a semi-automatic technique [47]. This step produced the final vegetation mask (VM) fig. 6, effectively isolating the plants for analysis.

2.2.2.9 Predictors

It is important to note that a PHYGEN estimate, in terms of a continuous variable, is the main expected outcome of this work. To achieve this task, the VM-derived area was assumed as a proxy for the Leaf Area Index (LAI). Differently, the mean (μ) and standard deviation (σ) of the following bands/indices from the calibrated S3 orthomosaic were computed: (i) Red, Green, and NIR bands, (ii) Normalized Difference Vegetation Index (NDVI), and (iii) Soil Adjusted Vegetation Index (SAVI). Additionally, the mean (μ) and standard deviation (σ) of heights of pixels belonging to VM were obtained by differencing DSM values of pixels within VM and the average of DSM values

of soil pixels. Finally, the date of acquisition (defined as DAA) was also considered to calibrate the prediction model.

Table 4: Predictors and their meanings.

Predictors	Variables meaning
(4) $NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$	where ρ_{NIR} and ρ_{RED} are the calibrated reflectance values from MSO
(5) $SAVI = \frac{1.5 \cdot (\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED} + 0.5)}$	where ρ_{NIR} and ρ_{RED} are the calibrated reflectance values from MSO
(6) $H_P = H_V - \overline{H_S}$	where H_P is the computed pixel relative average height of the vegetation contained in a pot, H_V is the absolute height of vegetation pixel in a pot, and $\overline{H_S}$ is the average absolute height of soil level in a pot.

2.2.2.10 ML Model

The available dataset is made of 132 multivariate observations (n), each providing 14 different predictors (p). To simplify the model and reduce parameters, the least absolute shrinkage and selection operator (LASSO) model (7) was used [31]. The PHYGEN variable (y) originally expressed as a percentage, was transformed into a probability by dividing it by one hundred. As PHYGEN values range

between 0 and 1, a linear regression model is unsuitable. A logistic function was used to adjust the linear predictions from the LASSO model to the PHYGEN scale, which is relevant to human vision. Twelve variables from MSO and DAA were normalized and used as independent variables. The dataset was split into an 80% training set and a 20% testing set. A K-fold ($K = 10$) strategy was applied to train and cross-validate the model [33]. To ensure a balanced splitting of observations, a stratified method was used based on PHYGEN values and acquisition dates. The human visual PHYGEN was fitted using a multivariate regression model with a L_1 regularization term [23] and a least squares adjusting method. The hyperparameter λ of L_1 was determined through a cross-validation involving 5 subsets of the training data, each representing a different part of a logarithmic range developing approximately between 0.003 and 0.67. The trained model outputs were then used as inputs for a logistic function (LF) (8), which was fitted to the PHYGEN data. The function parameters were estimated using non-linear least-squares optimization [24, 49], with initial values inferred from the PHYGEN distribution. The optimization aimed to minimize two error functions in the model, thereby enhancing the accuracy of the PHYGEN prediction:

Table 5: Models, their loss functions, and model outputs.

Model	Loss function	Model output
LASSO	$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \hat{\beta}_j x_{ij} \right)^2$ (7) $L_1 = \min; L_1 = \lambda \sum_{j=1}^p \beta_j $	
LogisticFunction (LF)	$\sum_{i=1}^n \left(y_i - \frac{L}{1+e^{-k(\hat{y}_i-y_0)}} \right) \hat{L}, \hat{k}, \hat{y}_0$ (8) \min	

where y_i is the i-PHYGEN observed rate, x_{ij} (7) is the observed value of the j-th explaining variable, β_0 (7) is the intercept of the function, β_j (7) is the weight corresponding to j-th variable, and \hat{y}_i (8) is the predicted value of the PHYGEN rate computed using weights estimated values from LASSO ($\hat{\beta}_j, \hat{\beta}_0$). The logistic function (8) has three parameters: L , y_0 , and k . These correspond to the higher limit of the function, the inflection point of the sigmoid, and the rate of growth, respectively. The estimated values for L , k , and y_0 are, respectively, \hat{L} , \hat{k} , and \hat{y}_0 , the correspondent estimated values. Initial values of \hat{L} , \hat{k} , and \hat{y}_0 , needed to run the not-linear least squares were set to 100, 50, and a random value extracted in the range [0, 1], respectively.

2.2.3 Results and Discussion

2.2.3.1 Measurement Errors

The surveyed 3D coordinates of GCPs were compared to those obtained from the photogrammetric restitution of the adjusted image block to assess errors associated with geometric features. To ensure a reasonable level of robustness for the accuracy assessment despite the low number of surveyed points, a Leave One Out method was used. MAE was used as an error measure. Similarly, the accuracy of radiometric calibration was assessed using a Leave One Out (LOO) approach. An assessment was performed separately for the different dates, and the corresponding Mean Absolute Percentage Error (MAPE) values were computed. Finally, MAPE values from the different dates were averaged to define the final reference value for radiometric calibration accuracy.

2.2.3.2 Geometric Assessment Errors

Accuracy assessment concerning image block bundle adjustment was achieved at a single date level. MAE values (for each coordinate) are reported in table 6.

Table 6: XYZ errors from photogrammetric restitution in mm.

DAA	MAE_x (mm)	MAE_y (mm)	MAE_z (mm)
3	0.57	0.61	0.62
7	0.65	0.70	0.91
14	0.67	0.68	0.89

The retained solution was deemed suitable, assuming that the differences between the main geometric features of diseased and healthy plants are greater than the reported errors. A comparison between MAE_z with the theoretical accuracy expected for the Z coordinate measure through photogrammetry eq. (2) showed that they were consistent.

2.2.3.3 Radiometric Validation

The Mean Absolute Percentage Error (MAE) of the calibration function training sample table 7 was used to estimate the goodness of function fitting.

Table 7: Radiometric Mean Absolute Percentage Error (Rad-MAPE) and ratio with the expected values obtained for the different bands and grey levels averaged along the three dates.

Band	Black (%)	Dark Gray (%)	Light Gray (%)	White (%)
Red	76.7	14.3	19.2	4.1
Green	82.8	47.5	53.2	18.1
NIR	119.6	29.2	20.7	6.2

The higher Rad-MAPE value was found for the green band, which is expected given the white balancing strategy adopted during image pre-processing. MAPE for red and NIR bands was found to be high as well, suggesting further refinements in the future to improve radiometric calibration.

2.2.3.4 Stability

The stability of the LASSO and Logistic model coefficients was analyzed. A 10-fold strategy was performed to generate an estimate for the mean and the standard deviation of coefficient estimates. fig. 7 and table 8 show related statistics.

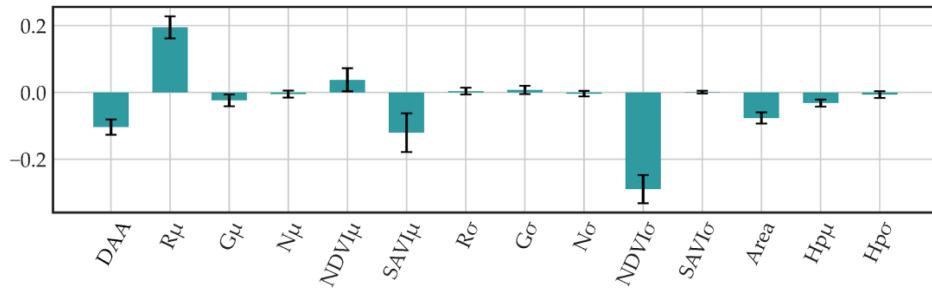


Figure 7: Mean values of LASSO β coefficients from the 10-fold approach, given for all the predictors. Whisker bars show 1-sigma LASSO β estimates.

Table 8: Mean, standard deviation, and coefficient of variation¹ values for the coefficients of the LASSO and logistic functions estimated using the 10-fold strategy.

Model	Parameter	Mean	Std. dev.	Coef.Var.¹
LASSO	β_{DAA}	-0.099	0.017	0.17
	$\beta_{\text{R}\mu}$	0.2	0.05	0.25
	$\beta_{\text{SAVI}\mu}$	-0.14	0.7	0.5
	$\beta_{\text{NDVI}\sigma}$	-0.28	0.04	0.14
	β_{Area}	-0.08	0.01	0.13
	λ	0.0028	0.0013	0.46
LF	L	94.53	1.22	<0.1
	k	0.06	0.001	<0.1
	y_0	47.15	0.69	<0.1

Coef.Var. is calculated with the absolute value of the mean.

nsights into the stability of the model can be gained by observing the coefficient of variation (Coef.Var.) of the most influencing parameters as estimated through the 10-fold strategy. Low values of Coef.Var. across all parameters proved that model stability is ensured. Bands and spectral indices showed the highest values of Coef.Var. This can be related to the significant uncertainty of calibrated reflectance, thus confirming the strict correspondence between measurement errors and the stability of the model (Barbedo et al. [13]).

2.2.3.5 Model Performances

Descriptive statistics of accuracy metrics were calculated with respect to the K-adjusted models used for predicting PHYGEN. MAE and the adjusted coefficient of determination $Adj R^2$ were calculated for the LASSO model, whereas the coefficient of determination R^2 was calculated for the Logistic function trained on LASSO predictions. The adjusted R^2 residual degrees of freedom were maintained equal to the number of the LASSO nonzero coefficients [54]. table 9 shows the results.

Table 9: Fit evaluation metric statistics.

Model	MAE (PHYGEN %)		R^2	$Adj R^2$
LASSO	Mean	11.77%	-	0.89
	Std	0.67%	-	0.03
LASSO + LF	Mean	10.66%	0.9	-
	Std	0.83%	0.03	-

The stacked model predictions ensure a mean absolute error, slightly overcoming the 11% and having a minimum coefficient of determination R_2 of about 0.9. Regarding the main goal of this work, it is worth noting that whatever the approach used to obtain an estimate of PHYGEN, its accuracy should be consistent with the one of human evaluation. According to the values reported above, the proposed method is able to provide PHYGEN scores similar to the one from experts. Our estimated accuracy (about 11%) is close to the reference threshold ordinarily accepted for PPP tests, which is 10%.

Moreover, it presents an R^2 value similar to the SOTA model that is trained with a huge amount of data from already tested PPPs due to its CNN architecture [28]. In contrast, MAE values for PHYGEN from our model were about double that obtained from SOTA, which can exploit a huge training set more effectively. Despite this, we believe that our method is promising and affordable when considering the actual operational conditions for the estimate of PHYGEN for new and untested PPPs, which escapes from the field of application of SOTA, basing deductions on a small training set.

2.2.3.6 Compliance with ANOVA Assumptions

As previously stated, ANOVA, t-tests, and Z-tests cannot be used with ordinal discrete scale dependent variables [48]. fig. 8 shows both the ordinal discrete data used to test the model and the continuous ones from the model. This is a great improvement in the ordinary screening procedures since it enables the possibility of testing group differences through an ANOVA-based approach that a discrete variable excludes.

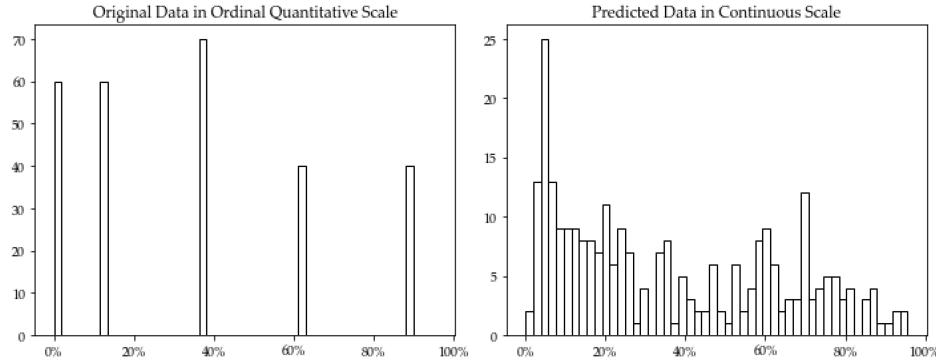


Figure 8: Discrete PHYGEN scores from the ordinary human vision-based approach (left). Continuous PHYGEN scores from the model proposed in this work (right).

2.2.4 Conclusions

The goal of this study was to test the operability and effectiveness of a controllable simple system based on multispectral digital photogrammetry and AI to support (and improve) current procedures for new PPP screening. This means that the system must be able to generate estimates of ordinarily recognized standard parameters (i.e., PHYGEN) and define the level of phytotoxicity of new PPPs before they enter the market. Basic requirements concern both compliance with accuracy standards and the robustness of the model output. The proposed method can be made operational if proper Geomatics and AI skills are properly integrated. Geomatic skills are related to proper management of the acquisition system that involves both geometric (image block bundle adjustment) and radiometric-related operations needed to prepare the data that the predictors of the PHYGEN have to be extracted from. Hardware solutions pro-

posed for the system exploit the abovementioned skills with the aim of reducing environmental and sensor-related issues. This makes acquired images more similar, partially overcoming one of the biggest problems recognized for the proper adoption of ML in phytopathometry: image features variability. A strong constraint introduced by this specific field of study is the lack of a huge training dataset that cannot be reasonably supplied for new PPPs to be screened. In such situations, this type of screening is required. The system operates in an effectively prepared greenhouse and requires significant infrastructure for the proper movement of the camera and lighting platform. In this work, we present a simple solution to these requirements. In particular, after suggesting how to pre-process the data from a photogrammetric and radiometric point of view, we found some predictors for the model to be trained that are able to exploit both the geometric and spectral content of acquired data. The predictors were analyzed and selected. They were used to train an ML algorithm integrating a LASSO and a logistic function to generate continuous estimates of PHYGEN. The robustness of the model was tested by conducting the training with a k-fold strategy and the correspondent statistics analyzed. The proposed method/system showed stability (robustness), proving to be independent of the training sample. The accuracy of PHYGEN prediction from our model is consistent with the ones from traditional methods. Compared to other AI-based approaches (i.e., SOTA), it showed slightly higher performances in terms of correlation with expert scores applied for new PPPs (our model: $R^2 = 0.9$, SOTA: $R^2 = 0.89$). In contrast, our model was not able to reach SOTA accuracy in PHYGEN scores

prediction (our model: MAE = 10.66%, SOTA: MAE = 6.74%). However, it must be noted that SOTA is not intended for predictions concerning new PPPs, and the reference values we reported refer to previously tested PPPs (i.e., providing a huge amount of training data). A surprising capability of the model was to overcome the discrete nature of expert-based scores for PHYGEN. In fact, it is able to generate continuous scores of PHYGEN, even if trained on discrete ones. Their continuous nature provides a high added value since it makes it possible to test differences among groups using ordinary ANOVA-based methods. However, some improvements are desirable, mostly in relation to a refinement of the hardware of the acquisition platform. A better-performing multispectral camera showing a higher spectral resolution and more rigorous calibration metadata is certainly a first step for future work. The active system providing controlled lighting can also be improved by using light sources that are able to generate a wider spectrum. Camera motion can be improved by using a stepper motor, allowing the possibility to stop the camera during image acquisition, thus avoiding blurring and reducing geometric deformations. Image processing could be also enhanced by strengthening automation in vegetation mask calculation from orthomosaic. The most significant improvement of the model would be to train a CNN with such a small amount of data. The final activation layer of this CNN should be set to the logistic function proposed in this work. Further studies must test data augmentation techniques and such activation layers with MAE loss to predict PHYGEN in similar setups. Regardless of the solution, we maintain that the explicability of the model, where the physical meaning of pre-

dictors and their relationships can be somehow recognized, is an added value for those applications where precise decision making is involved.

Bibliography

- [1] MAPIR_Survey3_Camera_Datasheet_English.pdf.
www.mapir.camera.
- [2] Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. page 50.
- [3] FAST APPROXIMATE NEAREST NEIGHBORS WITH AUTOMATIC ALGORITHM CONFIGURATION:. In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 331–340, Lisboa, Portugal, 2009. SciTePress - Science and Technology Publications.
- [4] Design and analysis of efficacy evaluation trials. *EPPO Bulletin*, 42(3):367–381, December 2012.
- [5] PP 1/135 (4) Phytotoxicity assessment. *EPPO Bulletin*, 44(3):265–273, December 2014.
- [6] PP 1/319 (1) General principles for efficacy evaluation of plant protection products with a mode of action as plant defence inducers. *EPPO Bulletin*, 51(1):5–9, April 2021.

- [7] PP 1/181 (5) Conduct and reporting of efficacy evaluation trials, including good experimental practice. *EPPO Bulletin*, 52(1):4–16, 2022.
- [8] Alan Agresti. Analysis of Ordinal Categorical Data.
- [9] R. Alcala, H. Vitikkala, and G. Ferlet. The World Trade Organization Agreement on the Application of Sanitary and Phytosanitary Measures and veterinary control procedures. *El Acuerdo sobre la Aplicación de Medidas Sanitarias y Fitosanitarias de la Organización Mundial del Comercio y los procedimientos de control veterinario.*, 39(1):253–261, January 2020.
- [10] Asif Ali, Jens C. Streibig, Joachim Duus, and Christian Andreasen. Use of Image Analysis to Assess Color Response on Plants Caused by Herbicide Application. *Weed Technology*, 27(3):604–611, 2013.
- [11] Keith B. Atkinson, editor. *Close Range Photogrammetry and Machine Vision*. Whittles, Caithness, reprinted edition, 1996.
- [12] Jayme G. A. Barbedo. Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, 172:84–91, August 2018.
- [13] Jayme G. A. Barbedo. Deep learning applied to plant pathology: The problem of data representativeness. *Tropical Plant Pathology*, 47(1):85–94, February 2022.
- [14] Enrico Borgogno Mondino. Multi-temporal image co-registration improvement for a better representation and quan-

tification of risky situations: The Belvedere Glacier case study. *Geomatics, Natural Hazards and Risk*, 6(5-7):362–378, July 2015.

- [15] MAPIR CAMERA. MAPIR Camera Reflectance Calibration Ground Target Package (V2). <https://www.mapir.camera/products/mapir-camera-reflectance-calibration-ground-target-package-v2>.
- [16] Gregory A. Carter and Alan K. Knapp. Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4):677–684, 2001.
- [17] K. S. Chiang, C. H. Bock, M. El Jarroudi, P. Delfosse, I. H. Lee, and H. I. Liu. Effects of rater bias and assessment method on disease severity estimation with regard to hypothesis testing. *Plant Pathology*, 65(4):523–535, 2016.
- [18] Kuo-Szu Chiang and Clive H. Bock. Understanding the ramifications of quantitative ordinal scales on accuracy of estimates of disease severity and data analysis in plant pathology. *Tropical Plant Pathology*, 47(1):58–73, February 2022.
- [19] Hangjian Chu, Chu Zhang, Mengcen Wang, Mostafa Gouda, Xinhua Wei, Yong He, and Yufei Liu. Hyperspectral imaging with shallow convolutional neural networks (SCNN) predicts the early herbicide stress in wheat cultivars. *Journal of Hazardous Materials*, 421:126706, January 2022.

- [20] Samuele De Petris, Filippo Sarvia, and Enrico Borgogno-Mondino. RPAS-based photogrammetry to support tree stability assessment: Longing for precision arboriculture. *Urban Forestry & Urban Greening*, 55:126862, November 2020.
- [21] EPPO. PP 1/135 (4) Phytotoxicity assessment. *EPPO Bulletin*, 44(3):265–273, 2014.
- [22] 2022-06-27/29 EPPO. Digital Technology and Efficacy Evaluation of Plant Protection Products. <https://www.eppo.int/MEETINGS/2022>.
- [23] Jerome H. Friedman, Trevor Hastie, and Rob Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33:1–22, February 2010.
- [24] Burton S. Garbow. MINPACK-1, Subroutine Library for Nonlinear Equation System. April 1984.
- [25] David M. Gates, Harry J. Keegan, John C. Schleter, and Victor R. Weidner. Spectral Properties of Plants. *Applied Optics*, 4(1):11–20, January 1965.
- [26] Sambuddha Ghosal, David Blystone, Asheesh K. Singh, Baskar Ganapathysubramanian, Arti Singh, and Soumik Sarkar. An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, 115(18):4613–4618, May 2018.

- [27] Mario A. Gomarasca. Elements of Photogrammetry. In Mario A. Gomarasca, editor, *Basics of Geomatics*, pages 79–121. Springer Netherlands, Dordrecht, 2009.
- [28] Laura Gómez-Zamanillo, Arantza Bereciartua-Pérez, Artzai Picón, Liliana Parra, Marian Oldenbuerger, Ramón Navarra-Mestre, Christian Klukas, Till Eggers, and Jone Echazarra. Damage assessment of soybean and redroot amaranth plants in greenhouse through biomass estimation and deep learning-based symptom classification. *Smart Agricultural Technology*, 5:100243, October 2023.
- [29] Mohd Asif Hajam, Tasleem Arif, Akib Mohi Ud Din Khanday, and Mehdi Neshat. An Effective Ensemble Convolutional Learning Model with Fine-Tuning for Medicinal Plant Leaf Identification. *Information*, 14(11):618, November 2023.
- [30] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2 edition, 2004.
- [31] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY, 2009.
- [32] David P Hughes and Marcel Salathé. An open access repository of images on plant health to enable the development of mobile disease diagnostics.
- [33] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An Introduction to Statistical Learn-*

ing: With Applications in Python. Springer Texts in Statistics.
Springer International Publishing, Cham, 2023.

- [34] Karl Kraus. *Photogrammetry: Geometry from Images and Laser Scans*. De Gruyter, October 2011.
- [35] Dawei Li, Lihong Xu, Xue-song Tang, Shaoyuan Sun, Xin Cai, and Peng Zhang. 3D Imaging of Greenhouse Plants with an Inexpensive Binocular Stereo Vision System. *Remote Sensing*, 9(5):508, May 2017.
- [36] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.
- [37] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pages 1150–1157 vol.2, Kerkyra, Greece, 1999. IEEE.
- [38] A.-K. Mahlein, M.T. Kuska, J. Behmann, G. Polder, and A. Walter. Hyperspectral Sensors and Imaging Technologies in Phytopathology: State of the Art. *Annual Review of Phytopathology*, 56(1):535–558, 2018.
- [39] Anne-Katrin Mahlein. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease*, 100(2):241–251, February 2016.

- [40] Pierre Moulon. Positionnement robuste et précis de réseaux d'images. page 193.
- [41] B. V. Nikith, N. K. S. Keerthan, M. S. Praneeth, and Dr. T Amrita. Leaf Disease Detection and Classification. *Procedia Computer Science*, 218:291–300, January 2023.
- [42] Ives Rey Otero and Mauricio Delbracio. Anatomy of the SIFT Method. *Image Processing On Line*, 4:370–396, December 2014.
- [43] Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, January 1979.
- [44] Michael D. Owen, Damian D. Franzenburg, Dean M. Grossnickle, and James F. Lux. Evaluation of Application Timings of Warrant Herbicide for Soybean Phytotoxicity. *Iowa State University Research and Demonstration Farms Progress Reports*, 2012(1), January 2013.
- [45] Françoise Petter, Anne Sophie Roy, and Ian Smith. International standards for the diagnosis of regulated pests. *European Journal of Plant Pathology*, 121(3):331–337, July 2008.
- [46] Riccardo Rossi, Claudio Leolini, Sergi Costafreda-Aumedes, Luisa Leolini, Marco Bindi, Alessandro Zaldei, and Marco Moriondo. Performances Evaluation of a Low-Cost Platform for High-Resolution Plant Phenotyping. *Sensors*, 20(11):3150, January 2020.

- [47] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "GrabCut": Interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 309–314, New York, NY, USA, August 2004. Association for Computing Machinery.
- [48] S. S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, June 1946.
- [49] Rainer Storn and Kenneth Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359, December 1997.
- [50] Lijuan Tan, Jinzhu Lu, and Huanyu Jiang. Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods. *AgriEngineering*, 3(3):542–558, September 2021.
- [51] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle Adjustment — A Modern Synthesis. In Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883, pages 298–372. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.
- [52] Clair Wyatt. *Radiometric Calibration: Theory and Methods*. Elsevier, December 2012.
- [53] Jing Zhou, Xiuqing Fu, Leon Schumacher, and Jianfeng Zhou. Evaluating Geometric Measurement Accuracy Based on 3D

Reconstruction of Automated Imagery in a Greenhouse. *Sensors*, 18(7):2270, July 2018.

- [54] Hui Zou, Trevor Hastie, and Robert Tibshirani. On the “degrees of freedom” of the lasso. *The Annals of Statistics*, 35(5):2173–2192, October 2007.