Smart Agricultural Technology 12 (2025) 101364

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Synergistic 3D, multispectral, and thermal image analysis via supervised machine learning for improved detection of root rot symptoms in hydroponically grown flat-leaf parsley

Avinash Agarwal [a,f,*], Filipe de Jesus Colwell [b], Julian Bello Rodriguez [a], Sarah Sommer [a], Monica Barman [c], Viviana Andrea Correa Galvis [b], Tom R. Hill [d], Neil Boonham [a,e], Ankush Prashar [a,f,*]

[a] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK
[b] Crop Science R&D Division, Infarm - Indoor Urban Farming B.V., the Netherlands
[c] Science Support Platform, Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Großbeeren, Germany
[d] Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[e] Fera Science Ltd., Sand Hutton, UK
[f] Institute for Bio- and Geosciences: Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, Jülich, Germany

ARTICLE INFO

Keywords:
RGB
Infrared thermography
3D-imaging
Machine learning
Disease detection
Root rot
Hydroponics

ABSTRACT

Root rot in hydroponically grown leafy vegetables is difficult to detect via conventional manual and machine vision-based approaches as symptoms of infection are not clearly visible on the canopy at earlier stages of infection. Hence, the present study investigates the potential of using machine learning for assessing canopy information obtained from multiple imaging platforms synergistically to improve root rot detection. Herein, flat-leaf parsley seedlings were grown in an experimental hydroponic vertical farm and inoculated with *Pythium irregulare* and *Phytophthora nicotianae*. Subsequently, the seedlings were imaged via 3D, multispectral, and thermal sensors at various stages of growth to obtain twenty-six image-based plant features. Following a preliminary screening of redundant features via regression analysis, data for seventeen image features associated with morphometric, spectral, and thermal attributes was co-analyzed using supervised machine learning by Support Vector Machines (SVM). Models using all eleven spectral features provided 98 % accuracy compared to 90 % for all five morphometric features and 94 % for canopy temperature alone. Inclusion of temporal data improved model performance by ca. 0.5 %, 1.5 %, and 8 % for spectral, thermal, and morphometric datasets, respectively. Exhaustive feature selection using different SVM kernels and maximum feature thresholds showed that combining features across the three imaging platforms along with temporal information enabled better identification of infected samples (>99 %) with as low as three features in comparison to using considerably more features from individual imaging systems. Hence, fusion of data from multiple imaging systems and using it with temporal information enabled better real-time high-throughput monitoring of root rot.

1. Introduction

Large-scale indoor crop production employing hydroponics has grown rapidly in the past decade. Beyond its traditional implementation within glasshouses, hydroponic cultivation systems are now being used in vertical farms as well. Such indoor farming operations provide greater control over plant growth conditions, have high water-use efficiency, and have the flexibility of being located closer to the point of consumption such as cities, reducing transportation-related carbon footprint [1–3]. Further, indoor farms produce high yield per unit area and facilitate year-round crop production, thereby reducing storage requirements and associated losses [3]. In addition to being highly resource-efficient, these operations also provide a high level of biosecurity by physically isolating the crops from pests and pathogens, thus minimizing the need for biocides.

Despite better biosecurity, complete exclusion of pathogenic

* Corresponding authors at: School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK; Institute for Bio- and Geosciences: Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, Jülich, Germany.
E-mail addresses: avinash.agarwal.1612@gmail.com, a.agarwal@fz-juelich.de (A. Agarwal), ankush.prashar@newcastle.ac.uk (A. Prashar).

https://doi.org/10.1016/j.atech.2025.101364
Received 14 May 2025; Received in revised form 22 August 2025; Accepted 23 August 2025
Available online 23 August 2025

microorganisms is practically impossible in large-scale indoor cultivation setups. Contamination due to bacterial and fungal pathogens may still occur due to inadequate phytosanitary conditions, limited treatment of seeds and substrates, or entry of air-borne spores through the ventilation system [4,5]. High humidity, monoculture, and favorable ambient temperature within indoor farms create an ideal environment for the proliferation of these microorganisms. Moreover, high planting density and circulatory irrigation via hydroponics increases the likelihood of rapid pathogen spread, thus increasing the risk of extensive crop loss [4–6]. The situation is exacerbated due to increased likelihood of mycotoxins entering the food-supply chain [7]. Since hydroponics-based indoor crop production aims at minimizing biocide use, real-time high-throughput monitoring of crops becomes imperative for timely detection of infected plants in such cultivation systems.

Crop monitoring and plant disease detection via machine vision has grown in popularity over the past decade as it overcomes limitations such as subjectiveness, low throughput, and poor reproducibility associated with the conventional practice of manual plant health assessment [8]. Imaging technologies such as 3D, multispectral, hyperspectral, and thermal sensors have been successfully applied for assessing plant health status and detecting plant stress, with each of these sensors providing information about distinct aspects of plant health [9,10]. For instance, 3D scanners and stereovision approaches can provide information pertaining to plant morphology in terms of height as well as canopy cover and structure [11–14], thus enabling reliable assessment of plant growth. In contrast, spectral features reflect the physiological status of a plant and can be used to monitor foliar symptoms appearing as changes in pigmentation patterns [15–18], whereas thermal sensors provide more in-depth real-time plant sensing by enabling users to monitor rapid variations in plant temperature in response to stress [19–24].

Despite their versatility, each of these sensing technologies has its own limitations. For example, inferences from 3D scans specifically rely on alterations in physical structure or canopy volume, and hence only prolonged or severe stress resulting in distinctive canopy damage would be clearly perceptible via this method. Conversely, slow physiological responses to stressors and limited sensing capacity for stresses that cause minimal changes in plant spectral profile are constraints associated with multispectral imaging. In contrast, diurnal variations and high sensitivity to environmental conditions are some limitations of thermal imaging [10,25,26]. However, as each of these imaging techniques focuses on distinct plant attributes, improvement in real-time plant stress detection in indoor cultivation systems could be achieved by computationally fusing the information coming from such sensors via machine learning (ML).

Currently, a large variety of ML methods are being implemented for modeling complicated data patterns across diverse agricultural sectors, including yield prediction [27,28], forecasting the price of agricultural products [29,30], and assessing soil quality [31,32]. In this context, such computational resources have also been pivotal in improving sensor data analysis for crop health assessment [33]. Unsupervised data processing tools such as principal component analysis and *k*-means clustering have been used to distinguish between healthy and stressed plants based on multiple spectral features [15,34–37], whereas supervised ML algorithms such as Support Vector Machines (SVM), *k*-Nearest Neighbors, and Random Forest have been employed for crop monitoring by identifying underlying trends in image datasets [34,36,38–40]. Further, computationally intensive deep-learning tools such as neural networks (NN) have also been implemented for disease detection by recognizing patterns on leaves such as lesions and chlorotic patches [41–43]. Although the feasibility of plant image analysis via ML for high-throughput crop monitoring has been extensively explored for the different imaging platforms deployed individually [9,44,45], reports on ML-based amalgamation of information from multiple imaging resources for plant stress and disease detection, especially in hydroponic cultivation systems, are very limited.

Amongst various plant stressors, root rot has emerged as a serious concern for hydroponics-based indoor crop production, with zoosporic oomycetes belonging to the *Pythium* and *Phytophthora* genera frequently identified as the causal organisms [4,6,46–48]. Although direct root imaging offers high accuracy in detecting root rot [49–51], such an approach may not be practical for real-time monitoring owing to practical limitations within large-scale production systems. Since these water-borne necrotrophs infiltrate and damage roots, timely detection of symptoms via standard machine vision focusing on canopy traits is challenging as the damage is hidden underneath the surface. However, earlier studies have demonstrated the possibility of monitoring root rot by canopy image analysis of plants such as lentils, ginseng, pepper, and avocado [50,52–56], with some even exploring data fusion from two sensor types for enhanced detection [37,57].

Considering the limited research on multi-sensor data fusion for root rot detection using canopy imaging, this study advances the field by employing ML to integrate morphometric, spectral, and thermal features for high-precision crop monitoring. For this, flat-leaf parsley, a popular culinary herb, was grown in a customized hydroponic vertical farming unit, and root rot was induced by inoculation with mycelial fragments of specimen belonging to the *Pythium* and *Phytophthora* genera. Plants were imaged at regular intervals via 3D, multispectral, and thermal sensors to assess the efficacy of stress detection through integration of canopy features from these sensors. To enhance our understanding of how individual plant attributes from different sensors contribute to disease detection, we employed exhaustive feature selection (EFS) to filter the dataset and subsequently assessed the accuracy of the ML models based on this refined data. Hence, our findings highlight the potential of integrating three complementary imaging techniques with ML to optimize feature selection and develop predictive models for detecting "hidden" diseases such as root rot more precisely.

2. Material and methods

2.1. Plant trials

Seedlings of flat-leaf parsley (*Petroselinum crispum* var. *neapolitanum*) were dark-germinated in coco-peat plugs in a nursery (Aralab-InFarm UK Ltd., London, UK) at a density of ~25 seedlings/plug, in line with commercial production standards. Once the seedlings were ca. 2 cm in height, they were transplanted to an experimental setup with six "deep water culture" hydroponic units (Fig. 1) in a growth chamber having regulated environment (Newcastle University, Newcastle upon Tyne, UK). Each hydroponic unit comprised of a dark-grey polypropylene reservoir (inner dimensions: $L \times W \times H$ 56×36×11 cm) filled with 18 L commercial hydroponics solution, an opaque-white tray-lid with a 7×4 array of circular empty slots for seedling plugs (Fig. 1). Each unit was equipped with a submersible water pump for root aeration as well as overhead broad-spectrum LED lighting (L28-NS12, Valoya Ltd., Finland; 300–350 μmol.m^{-2}s^{-1} PPFD), and was connected to a circulating water bath to maintain the water temperature (Fig. 1). A total of 26 seedling plugs were placed within each hydroponic unit, as depicted. Plants were allowed to grow for 20 days under controlled conditions: 22±1 °C temperature, 75±5 % relative humidity, and 16/8 h day-night cycle. Four independent trials were conducted by varying inoculation stages and pathogen concentration to obtain a wide range of plant responses for ML analysis (Table 1). In each trial, two units each were inoculated with *Pythium irregulare* and *Phytophthora nicotianae* (described in Section 2.2) on specific days post transplantation (DPT), while the two remaining units acted as the control (Fig. 1).

2.2. Pathogen isolation and inoculation

Py. irregulare and *Ph. nicotianae* isolates were obtained from diseased plants found in a commercial hydroponic vertical farm. *Py. irregulare* was cultured on PARP+B semi-selective medium (corn meal agar amended with pimaricin, 5 mg/L; ampicillin, 250 mg/L; rifampicin, 10

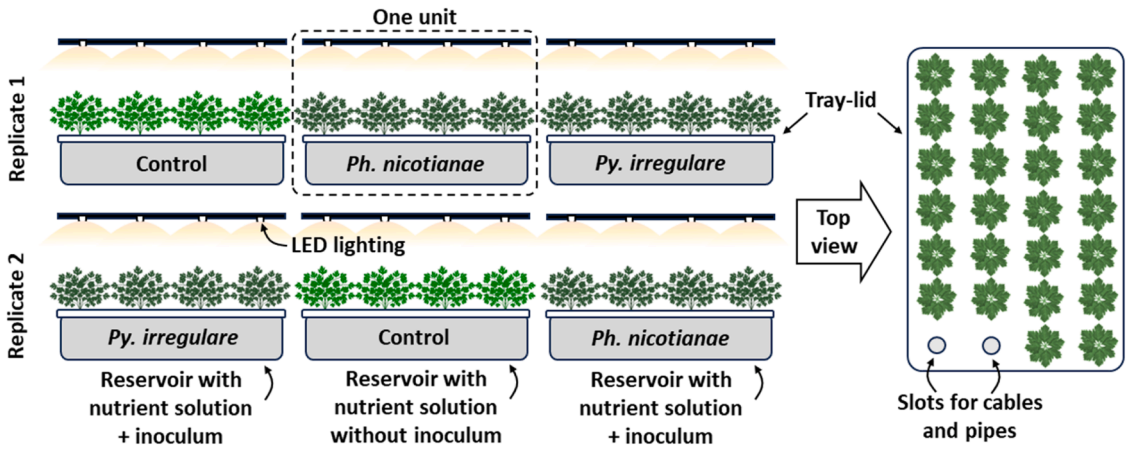

Fig. 1. Schematic layout of the experimental setup consisting of six hydroponic units. Each unit (represented using dotted lines) comprised of a reservoir for nutrient solution, a tray-lid, and overhead LED lighting. As shown in the top view, each tray-lid had 28 circular slots; 26 slots with plant samples, and the two empty slots being used for cables and pipes connected to a submersible air pump for root aeration and a circulating water bath to maintain water temperature. The six units were grouped into two replicates of three trays each. One unit from each replicate was inoculated with either *Pythium irregulare* (*Py. irregulare*) or *Phytophthora nicotianae* (*Ph. nicotianae*) as indicated, while the remaining unit acted as the control.

Table 1
Overview of the experimental design for inoculation of parsley with root rot-causing oomycetes and acquisition of thermal, 3D, and multispectral images.

Trial No.	Days post transplantation (DPT)		Inoculum strength (mycelia/mL)	No. of samples*		
	Inoculation	Imaging		Control	*Py. irr.*	*Ph. nic.*
1	4	6, 8, 11, 13, 15	1×10^4	52	52	52
2	4	7, 11, 14, 18	1×10^4	52	52	52
3	5	13, 16, 20	1×10^4	52	52	52
4	1	4, 6, 8, 11, 13, 15	2×10^4	52	52	52

DPT, no. of days after transferring the seedlings from the nursery to the experimental hydroponics setup; *Py. irr., Pythium irregulare; Ph. nic., Phytophthora nicotianae.*
*Values indicate the number of seedling plugs, where each plug contained ~25 seedlings, considered as a single sample.

mg/L; pentachloronitrobenzene, 50 mg/L; and benomyl 10 mg/L) [58], whereas *Ph. nicotianae* was cultured on PARP+H medium (corn meal agar amended with pimaricin, 10 mg/L; ampicillin, 100 mg/L; rifampicin, 10 mg/L; pentachloronitrobenzene, 50 mg/L; hymexazol, 50 mg/L) [59]. Identity of the isolates was confirmed via PCR amplification of specific genes using primers reported earlier [60–62](Supplementary Table S1), followed by sequencing and alignment. The pathogens were grown in bulk using clarified-V8 broth as described by McGehee et al. [63] with minor modifications [37]. Briefly, a 4-mm plug of the inoculated PARP medium was transferred to a sterile Petri dish and 20 mL of V8 broth was added, followed by incubation in darkness for 5 days at 25 °C. Subsequently, the mycelial mats were liquefied in ddH_2O for 2 min. The resulting slurry was used for inoculating seedling plugs for the designated hydroponic units (Fig. 1) at different growth stages and concentrations in each of the four trials (Table 1).

2.3. Plant imaging

Non-invasive data collection was performed via thermal, multispectral, and 3D imaging at different stages of plant growth by briefly transferring individual sample trays to a customized imaging setup. Stage of plant growth at each imaging interval, indicated as DPT for imaging (Table 1), served as the temporal marker for the imaging dataset.

Thermal images were captured using a T1030sc thermal camera (Teledyne FLIR LLC, USA; spectral range 7.5–14 μm; focal plane array uncooled microbolometer with HD detector; spatial resolution 1024×768 pixels) that acquired thermal and Red-Green-Blue (RGB) digital images concurrently via adjacent lenses, as described previously [24]. Briefly, canopy thermal images were acquired at a room temperature of 25±1 °C under a neutral-white LED light source. An adjustable camera stand was used to position the camera vertically above the canopy while maintaining a fixed distance (2 m) between the camera objective and the hydroponics tray-lid. Camera parameters such as reflected, atmospheric, and optics temperatures were fixed [64]. The trays were imaged within 3–5 min of being taken out of the experimental setup to limit shifts in sample temperature during imaging. RGB images captured by the thermal camera were used for extracting thermal data (described in Section 2.4) and for plant scoring (described in Section 2.6), but were excluded during ML-based data analyses to avoid redundancy with the spectral information generated from the 3D-multispectral scanner.

Subsequently, each tray was scanned using a PlantEye F500 3D-multispectral scanner (Phenospex, The Netherlands, www.phenospex.com) to simultaneously record the morphometric and spectral features of the samples [65]. Briefly, the scanning setup comprised of a fixed platform for placing trays with plant samples, along with an overhead scanning unit equipped with both 3D and multispectral sensors [37]. The scanner moved horizontally (Y-axis) along a conveyor at a speed of 50 mm/s, at a fixed distance of 100 cm (Z-axis) from the tray-lid (X-Y axis). The setup provided an approximate resolution of 0.7 mm X-axis, 1 mm Y-axis, and 0.2 mm Z-axis. Spectral features were recorded in the blue (B; λ = 460–485 nm), green (G; λ = 530–540 nm), red (R; λ = 620–645 nm), and near-infrared (NIR; λ = 720–750 nm) ranges by illuminating the samples using in-built LED lights with corresponding wavebands, whereas a laser scanner (λ = 940 nm) was used for recording morphometric features in 3D. The scans were processed using HortControl software (Phenospex), which superimposed the spectral and 3D information based on internal calibrations to create point-cloud (.ply) data files containing the spatial (X, Y, Z) and spectral (R, G, B, NIR) values of each pixel. Uniformity in lighting provided by the in-built R-G-B-NIR LEDs nullified the need for further spectral calibrations.

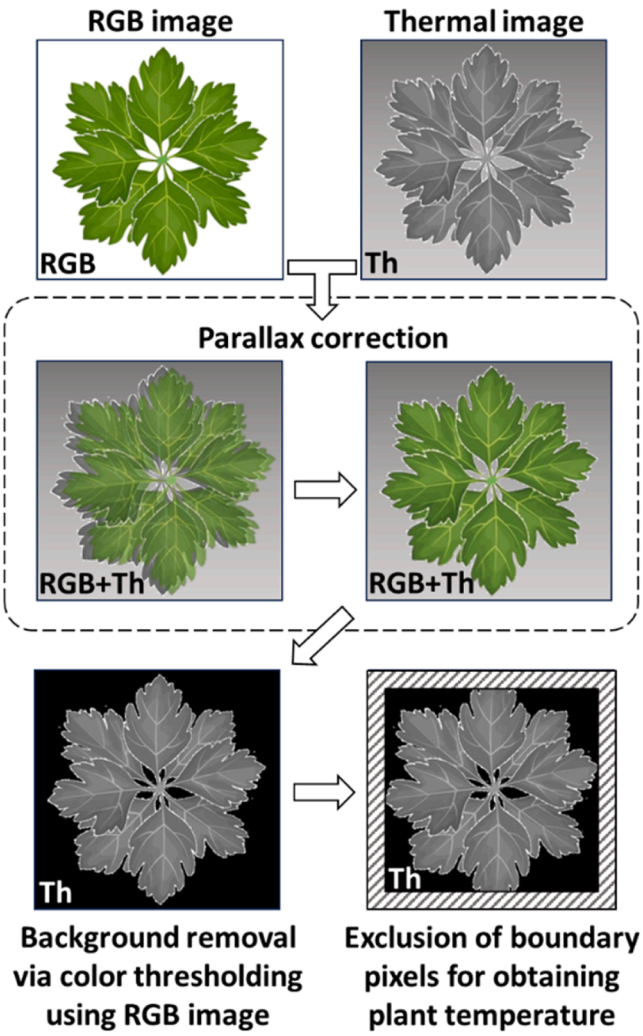

Fig. 2. Overview of the thermal image processing pipeline for extracting canopy temperature of each sample. First, the parallax error between each RGB and thermal image was corrected to ensure adequate superimposition. Subsequently, threshold RGB values were used to create a mask for background removal, which was applied to the thermal image. This resulted in the removal of background (tray) pixels within the thermal image. Next, a border of 10 pixels was omitted on all four sides from the masked thermal image to minimize the interference from overlapping neighboring samples. Finally, the remaining pixels occupied by the leaves were used for extracting average canopy temperature for each sample using at least 1000 pixels. RGB, color image consisting of Red, Green, and Blue channels; Th, thermal image; RGB+Th, superimposed RGB and thermal images.

2.4. Extracting plant temperature from thermal images

Canopy temperature for individual plants was extracted using a customized thermal image processing pipeline (Fig. 2) created using Python programming (www.python.org) as reported earlier [24], with minor modifications. Briefly, the process involved the following steps: 1) correction of parallax error between the thermal and RGB images arising due to non-coaxial thermal and RGB sensors; 2) RGB color thresholding to create a mask for removing background (tray) pixels in the thermal image; 3) isolation of individual plants using a 7×4 grid as per tray design to create regions of interest (ROIs, 95×105 pixels) corresponding to each sample; 4) excluding a border region of 10 pixels on all four sides of each ROI to minimize the effect of overlapping leaves from adjacent ROIs; and 5) computation of average canopy temperature for each ROI using at least 1000 plant pixels. As an effective environmental correction for absolute errors, the difference between observed plant temperature (T_{obs}) and the tray surface temperature (T_{tray}) was used along with a constant (25 °C) to obtain the normalized canopy temperature (T_c) as follows:

$$T_c\ (^{\circ}\mathrm{C}) = T_{obs} + (25 - T_{tray}) \quad (1)$$

2.5. Morphometric and spectral feature extraction

The 3D-multispectral scans were processed using HortControl software (Phenospex). Briefly, the morphometric and spectral data for individual plants was obtained by dividing the scanning area into a 7×4 array of identical sectors, with each sector containing one sample (seedling plug). The 3D point-cloud data was processed using the software to extract nine morphometric parameters (Table 2): mean plant height, maximum plant height (Ht_max), total leaf area (TLA), digital biomass (DB), leaf area index (LAI), projected leaf area (PLA), leaf angle, leaf inclination (LInc), and light penetration depth (LPD). Similarly, multispectral data was processed by the software to calculate five spectral indices as follows (Table 2): Green Leaf Index (GLI), Hue angle, Normalized Difference Vegetation Index (NDVI), Normalized Pigment Chlorophyll ratio Index (NPCI), and Plant Senescence Reflectance Index (PSRI). Raw R, G, B, and NIR reflectance data were extracted from the 3D point-cloud files using a customized pipeline created using Python programming. This data was used to calculate other spectral indices, viz., R+G+B, R+G-B, R+G, R/G, G/R, G-minus-R (GMR) [15,66,67] and Augmented Green-Red Index (AGRI, [GMR]×[G/R]) [37].

2.6. Data pre-processing

Since both pathogens resulted in realistically similar foliar symptoms, the analysis followed a generalized ML-based disease detection approach by collating the samples with both types of inoculations within one class of "infected" samples, which was compared against the healthy "control" class. The full dataset comprised of twenty-seven features, i.e., nine morphometric and sixteen spectral attributes (Table 2), plant temperature (T_c), and the temporal data corresponding to each imaging interval in terms of DPT (Table 1). Preliminary sample screening was carried out manually to minimize erroneous ML trends by excluding perfectly healthy samples from the infected dataset and clearly stressed/damaged samples from the control dataset. This was done by visual scoring of the RGB image of each sample six times following two rounds of blinded labelling by three members of the team (healthy = 3; intermediate = 2; stressed = 1). Samples from the control cohort that appeared perfectly healthy (average score > 2.5) as well as samples from the inoculated trays that showed clear signs of stress such as aberrant growth and/or poor pigmentation (average score < 1.5) were used for subsequent analyses; the remaining samples were ignored. All twenty-seven features for the selected samples were subsequently subjected to Pearson's correlation analysis using the Data Analysis ToolPak within Microsoft Excel 365 (Microsoft Corp., USA). Features exhibiting strong identical linear trends ($r^2 > 0.95$) were selectively excluded to reduce redundancy. The features that were retained, henceforth referred to as "shortlisted features", were used for generating ML models.

2.7. ML for disease detection

Considering the possibility of processing high-dimensional data and the scope of testing different kernels using SVM [68], a Python-based ML pipeline using the Scikit-learn module (https://scikit-learn.org/stable/) for SVM was implemented for sample classification. Briefly, ML tests with the shortlisted features were carried out using three kernels, viz., *linear* (*lin*), *polynomial* (*poly*), and *radial basis function* (*rbf*). Herein, the *lin* kernel employs linear mathematical expressions to generate classification boundaries or hyperplanes, whereas the *poly* and *rbf* kernels create polynomial and Gaussian expressions, respectively, which are capable of classifying samples based on non-linear spatial distributions

Table 2
Definitions of morphometric (M) and spectral (S) indices measured by the 3D multispectral scanner.

Type	Parameter	Definition
M	Mean plant height	Average height of the top 10 % points within the canopy
	Maximum plant height (Ht_max)	Height of highest point within the canopy
	Total leaf area (TLA)	Sum of all triangulated 3D surfaces on the canopy
	Leaf area index (LAI)	TLA/sector area
	Projected leaf area (PLA)	Two-dimensional projection of TLA
	Digital biomass (DB)	Mean plant height × TLA
	Leaf angle	Weighted average of all angles for every face in the triangulated plant mesh
	Leaf inclination (LInc)	TLA/PLA
	Light penetration depth (LPD)	Lowest point in the canopy detectable by the laser
S	Green Leaf Index (GLI)	[(2× G)-R-B]/[(2× G)+R+B])
	Hue angle	Numeric representation of a tone on the color wheel within the HSV color space
	Normalized Difference Vegetation Index (NDVI)	(NIR-R)/(NIR+R)
	Normalized Pigment Chlorophyll ratio Index (NPCI)	(R-B)/(R+B)
	Plant Senescence Reflectance Index (PSRI)	(R-G)/NIR

Multispectral channels: B, Blue; G, Green; NIR, Near-infrared; R, Red. Sector area = total scanned area/28, as per the 7× 4 array for each tray (Fig. 1).

of data.

In preliminary ML tests (ML-1), efficacy of disease detection using data from individual sensors was assessed by analyzing the shortlisted features obtained from each type of sensor, viz., spectral, morphometric, and thermal, with and without temporal data. All three kernels were deployed and optimal values of hyperparameters C and γ were selected following a limited grid search with C = [0.001, 0.01, 0.1, 1, 10, 100] and γ = [0.0001, 0.001, 0.01, 0.1, 1, 10] to maximize model accuracy. Here, C represents penalty weight of deviations, and γ (*poly*- and *rbf*-specific hyperparameter) defines the range of influence for a single training instance. Stratified 80:20 train–test split and five-fold cross-validation was implemented for all tests to account for changes in model performance with variations in training datasets.

In a series of subsequent ML tests (ML-2), EFS was performed by analyzing all shortlisted features simultaneously to evaluate the likelihood of each feature being selected automatically by the ML algorithm, as well as to understand how model accuracy changed upon increasing the total number of features. For this, the *mlxtend* library (https://rasbt.github.io/mlxtend/) was utilized along with the Scikit-learn module. The threshold for maximum features (*max_feats*) was increased stepwise from 1 to 10 for all three kernels. Each threshold was tested using ten different random states, i.e., arbitrary seed values for data shuffling prior to the 80:20 train–test split. This resulted in a total of 300 EFS models as follows: 3 kernels × 10 *max_feats* thresholds × 10 random states. In each iteration, the EFS algorithm attempted to identify the smallest feature subset that could give the highest accuracy as dictated by the unique "kernel + *max_feats* + random state" criterion by exhaustively testing all possible combinations of features within the threshold limit following five-fold cross-validation. Frequency of feature selection was recorded for each *max_feats* value, and feature score was calculated for each kernel as follows:

$$\text{Feature score} = \sum_{i=1}^{n} F_i \tag{2}$$

Here, F_i indicates how frequently each feature was selected across ten trials with different random states at the i^{th} *max_feats* threshold, and n indicates the total number of *max_feats* thresholds, i.e., from 1 to 10, creating a score range of 0 to 100 for each feature within each kernel. A higher feature score indicated higher feature selection frequency by the respective kernel, and vice versa. The scores were used to assign ranks to each feature, which indicated its overall performance in ML modelling.

Since the present study focused on evaluating feature significance and the role of sensor data fusion in enhancing disease detection rather than optimizing model performance through extensive hyperparameter tuning, additional validation tests were not undertaken for the sake of simplicity. Instead, five-fold cross validation and varying random states were employed to account for overfitting and assessing model stability across different training datasets.

2.8. Statistical analysis

Datasets for all shortlisted features collated from all intervals and trials were subjected to two-sample Kolmogorov–Smirnov (KS) non-parametric test using R programming (www.r-project.org) to assess the statistical difference of data distribution between control and infected samples. Here, KS = 0 implies identical distribution, whereas KS = 1 indicates high degree of dissimilarity.

Further, overlap between the numerical ranges of control and infected datasets was assessed by calculating the Jaccard index (JI) and the Szymkiewicz–Simpson overlap coefficient (SS) using basic mathematical functions in Microsoft Excel 365 as follows:

$$\text{JI} = \frac{|R_C \cap R_I|}{|R_C \cup R_I|} \tag{3}$$

$$\text{SS} = \frac{|R_C \cap R_I|}{min\,(|R_C|,\ |R_I|)} \tag{4}$$

Here, R_C and R_I indicate the range of values for the control and infected datasets, respectively. The $|R|$ function finds the size of the specified range, *min* function finds the smallest amongst the specified ranges, whereas ∩ and ∪ operators find the intersection and union of two ranges, respectively. JI represents the proportion of overlapping data across the entire data distribution, with JI = 0 indicating no overlap and JI = 1 indicating complete overlap between both datasets. In contrast, SS represents the proportion of the dataset with the smaller range overlapping with the dataset with greater range, where SS = 0 indicates no overlap and SS = 1 indicates that the low-range dataset lies entirely within the high-range dataset. Outliers were excluded using the interquartile range while calculating JI and SS to improve the conciseness of results.

3. Results

3.1. Feature shortlisting, data distribution, and overlap analysis

Following the correlation analysis ($r^2 < 0.95$) of twenty-seven features using the data for 1804 samples from the healthy (n = 954) and infected (n = 850) classes, the following eighteen features were shortlisted: morphometric features– DB, Ht_max, LAI, LInc, and LPD; spectral features– Hue, GLI, NDVI, NPCI, PSRI, R, G, NIR, G/R, GMR, and AGRI; thermal data (T_c); and temporal data for imaging intervals in terms of DPT (Supplementary file S1).

Collating data across different intervals from all four trials revealed diverse trends in the numerical ranges of the digitally recorded features for both control and infected samples (Figs. 3, 4). Values of morphometric features such as DB, Ht_max, and LAI were generally lower for infected samples (Fig. 3a–c), which indicates suppressed growth. In

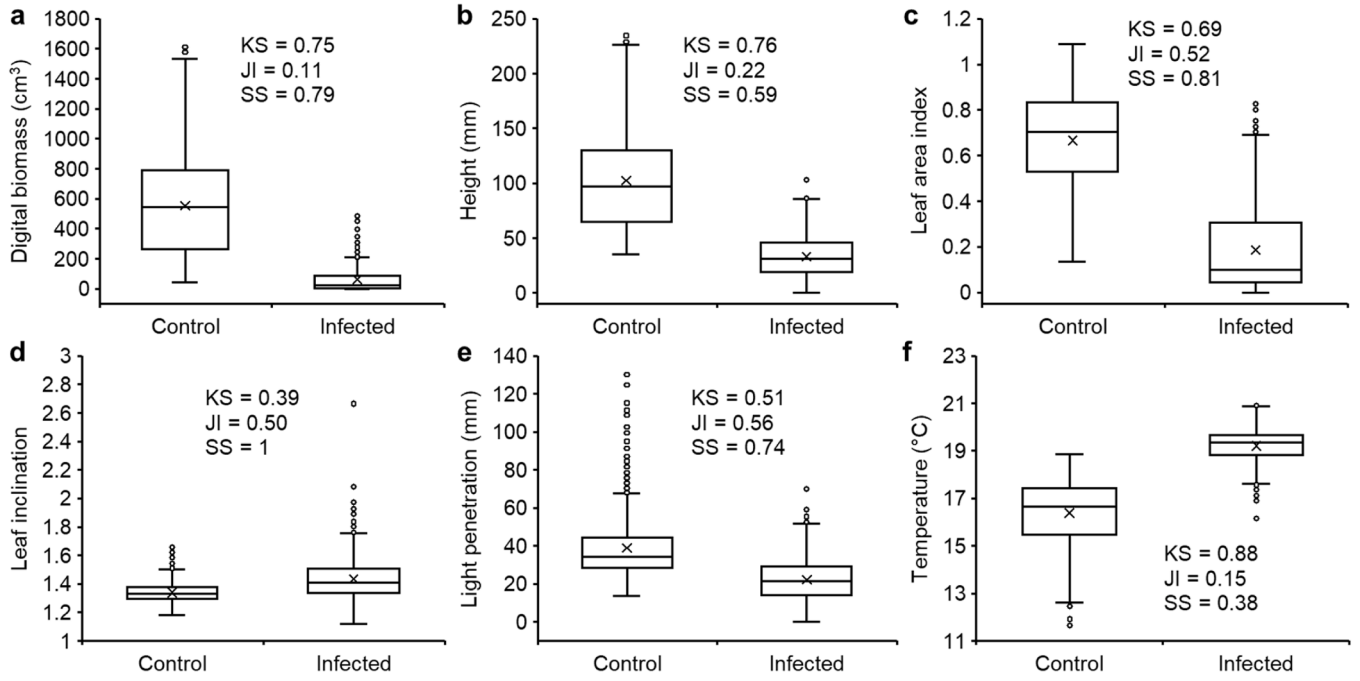

Fig. 3. Morphometric features (a–e) and canopy temperature (f) of healthy and infected flat-leaf parsley (control: $n = 954$; infected: $n = 850$) recorded throughout the period of observation. Morphometric features: digital biomass (a), maximum plant height (b), leaf area index (c), leaf inclination (d), light penetration depth (e). Data represents the summarized range of values observed across all imaging intervals from all four experimental trials (Table 1). Box-and-Whisker plots include the mean (×), median (horizontal line), interquartile range (box), whiskers representing 5 and 95 % percentile, and the outliers (o). Data distributions for all parameters were significantly different ($p < 0.05$) as per two-sample Kolmogorov–Smirnov (KS) test. JI, Jaccard Index of similarity; SS, Szymkiewicz–Simpson overlap coefficient.

contrast, T_c values exhibited a reversed trend, i.e., the infected plants were generally warmer than control samples (Fig. 3f). LPD was relatively lower in infected samples, whereas LInc did not exhibit any clear trend (Fig. 3d, e). Various spectral features, including GLI, NDVI, G/R, GMR, and AGRI, had higher average values for the control as compared to the infected samples (Fig. 4b, c, i–k), whereas the average values for R and G reflectance as well as PSRI were markedly higher in the infected samples (Fig. 4e–g). In contrast, NPCI and NIR exhibited only minor differences in mean values between both classes (Fig. 4d, h). Notably, while the numerical range of Hue for control samples was very narrow, the infected samples showed high variability (Fig. 4a).

While features such as T_c, Hue, GLI, G/R, and AGRI showed a strong distinction in data distribution (KS $>$ 0.8) between the control and infected samples (Figs. 3f, 4a, b, i, k), relatively higher similarity in data distribution (KS $\leq$ 0.55) was observed for LInc, LPD, NPCI, and NIR (Figs. 3d, e, 4d, h). Concomitantly, JI $\geq$ 0.5 along with SS $\geq$ 0.7 for LAI, LInc, LPD, NPCI, and NIR indicated considerable overlap amongst the control and infected samples (Figs. 3c–e, 4d, h). In contrast, SS $\geq$ 0.7 with JI $<$ 0.5 for features such as DB, Hue, NDVI, PSRI, R, G, GMR, and AGRI suggested that the treatment showing the smaller range of values was considerably subsumed within the broader range of values from the other treatment, despite minimal overlap across the entire range of the latter (Figs. 3a, 4a, c, e–g, j, k).

3.2. ML with individual feature categories (ML-1)

In the ML-1 tests, grouping of features based on sensor type revealed that model training with the eleven shortlisted spectral features had the highest prediction accuracy of ca. 98 %, whereas models with the five shortlisted morphometric features yielded ca. 90 % accuracy (Table 3). Interestingly, the model trained with thermal data only, i.e., T_c, had an accuracy of ~94 %. Inclusion of temporal data (DPT) improved the accuracy of spectral and thermal models by <1 % and ~2 %, respectively, whereas accuracy for the morphometric model was increased by ~8 %.

3.3. ML with exhaustive feature selection (ML-2)

EFS tests (ML-2) revealed a sharp increase in model accuracy from 93.4–95.7 % to 99.0–99.3 % between *max_feats* = 1 to 3 (Fig. 5a–c). Overall accuracy peaked around *max_feats* = 5 at 99.4–99.7 %, with no realistic improvement upon increasing the number of features further. In general, classification accuracy of 93–96 % could be achieved with only one feature, whereas an accuracy of >99 % could be achieved for *max_feats* $\geq$ 3 (Fig. 5a–c), i.e., when selection of three or more features was permitted. In contrast to the similar trends in accuracy observed across the three kernels, total number of features selected at different thresholds varied considerably for *max_feats* $>$ 5 (Fig. 5d–f). For example, amongst the different models at *max_feats* = 10 the total number of selected features ranged between 6–10, 9–10, and 6–9 for the *lin*, *poly*, and *rbf* kernels, respectively.

Feature co-occurrence matrices for the three kernels (Supplementary file S2) provided an insight into feature selection and pairing preferences at different *max_feats* thresholds. Herein, AGRI was the only feature that was selected for all 30 ML models at *max_feats* = 1, i.e., for creating classification models using just one feature, but was rarely selected for higher *max_feats* values. Subsequently, all 30 models with *max_feats* = 2 selected DPT and paired it with one spectral feature, viz., GLI ($n = 11$), NDVI ($n = 10$), or G/R ($n = 9$). Further, 29 out of 30 models with *max_feats* = 3 used DPT along with T_c and one spectral feature, whereas one model used DPT along with two spectral features. Herein, GLI was selected most frequently (27/30 models), followed by NDVI (3/30 models); T_c was replaced by PSRI in the model where two spectral features were paired with DPT. For *max_feats* = 4, all ten models employing the *lin* kernel had DPT and T_c along with one spectral and one morphometric feature, whereas 4/10 *rbf*-based models had such a combination of features. In contrast, use of all four feature categories was not observed in any of the models implementing the *poly* kernel at

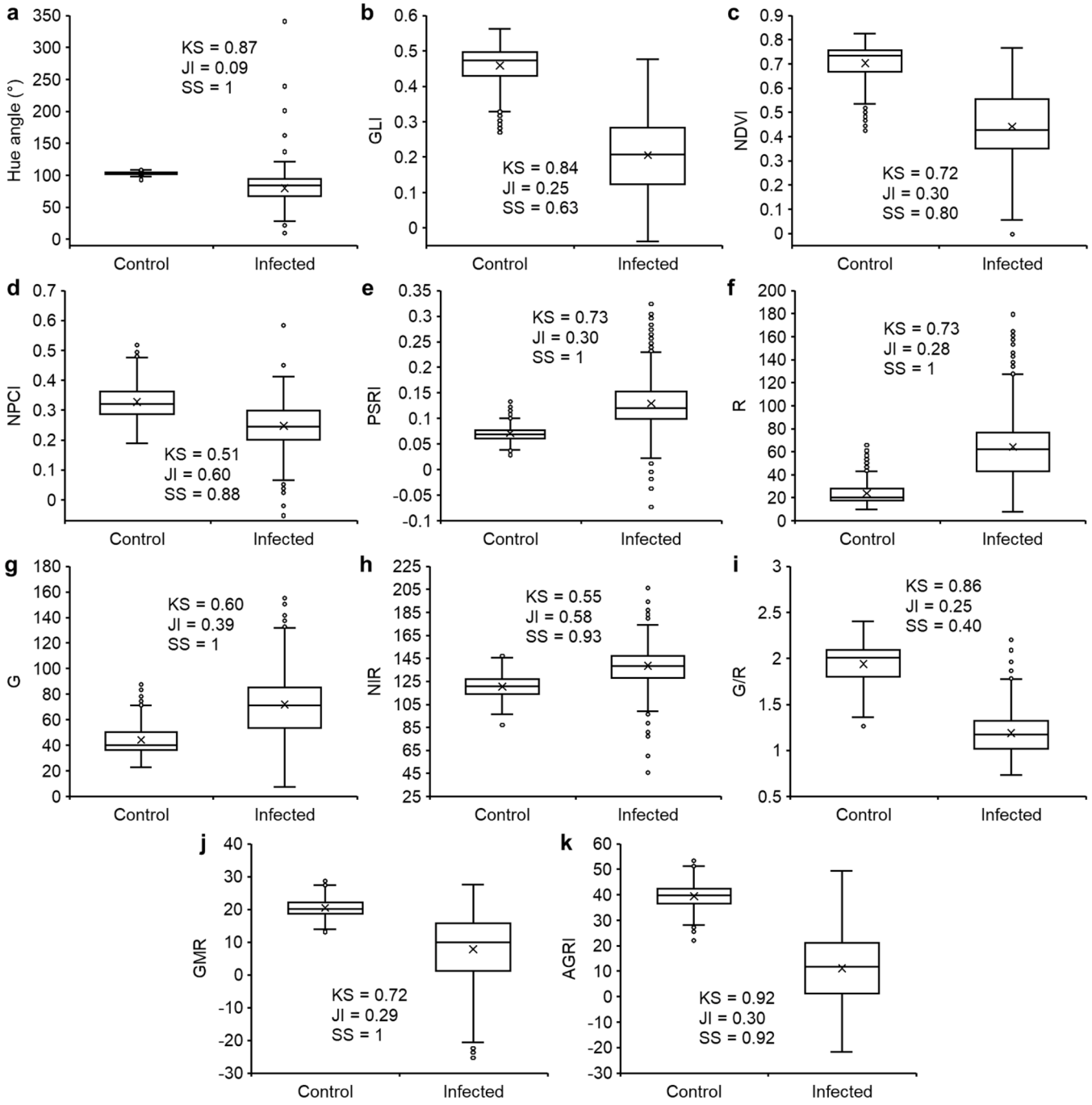

Fig. 4. Spectral features of healthy and infected flat-leaf parsley (control: $n = 954$; infected: $n = 850$) recorded throughout the period of observation. Spectral features: Hue (a), Green Leaf Index (GLI; b), Normalized Difference Vegetation Index (NDVI; c), Normalized Pigment Chlorophyll ratio Index (NPCI; d), Plant Senescence Reflectance Index (PSRI; e), Red reflectance (R; f), Green reflectance (G; g), Near-Infrared reflectance (NIR; h), Green-Red reflectance ratio (G/R; i), Green-minus-Red reflectance (GMR; j), and Augmented Green-Red Index (AGRI; k). Data represents the summarized range of values observed across all imaging intervals from all four experimental trials (Table 1). Box-and-Whisker plots include the mean (×), median (horizontal line), interquartile range (box), whiskers representing 5 and 95 % percentile, and the outliers (o). Data distributions for all parameters were significantly different ($p < 0.05$) as per two-sample Kolmogorov–Smirnov (KS) test. JI, Jaccard Index of similarity; SS, Szymkiewicz–Simpson overlap coefficient.

max_feats = 4. Instead, 9/10 *poly* models selected DPT and T_c along with two spectral features, whereas one model selected DPT, one morphometric feature, and two spectral features. Upon increasing the *max_feats* threshold to 5, all 20 models based on *lin* and *poly* kernels as well as 9/10 *rbf* models used DPT and T_c along with at least one spectral and one morphometric feature. Every ML model with *max_feats* $\geq$ 6 utilized features from all categories irrespective of the kernel.

Feature rankings based on the frequency of occurrence during EFS indicated that features such as DPT, T_c, GLI, and Ht_max were selected most often by the ML algorithm for all kernels (feature score > 55/100; Table 4). In general, DPT received the highest rank with a feature score of 90/100 for all three kernels, and was selected by all 270 models with *max_feats* $\geq$ 2. Further, T_c was ranked second, being selected in 238 out of 240 models with *max_feats* $\geq$ 3 and attaining a score of 79–80/100 for each kernel. GLI and Ht_max were ranked third and fourth, respectively, for both *poly* and *rbf* kernels, although the ranks were reversed for the *lin*

Table 3
Accuracy of machine learning-based sample classification using different types of sensor data without feature selection (ML-1).

Feature type	Total no. of features	Accuracy (%)*	Kernel#	No. of training samples		No. of test samples	
				Healthy	Infected	Healthy	Infected
S	11	98.34±0.65	*rbf*	763	680	191	170
S+D	12	98.95±0.6	*rbf*	763	680	191	170
M	5	89.92±1.1	*rbf*	763	680	191	170
M+D	6	98.61±0.68	*rbf*	763	680	191	170
T	1	94.18±1.06	*lin*	763	680	191	170
T+D	2	95.62±1.71	*rbf*	763	680	191	170

Feature types: D, temporal data; M, morphometric features; S, spectral features; T, temperature. *Accuracy values indicate mean±SD of five-fold cross-validation. #Kernel resulting in the highest accuracy amongst *linear* (*lin*), *polynomial*, and *radial basis function* (*rbf*).

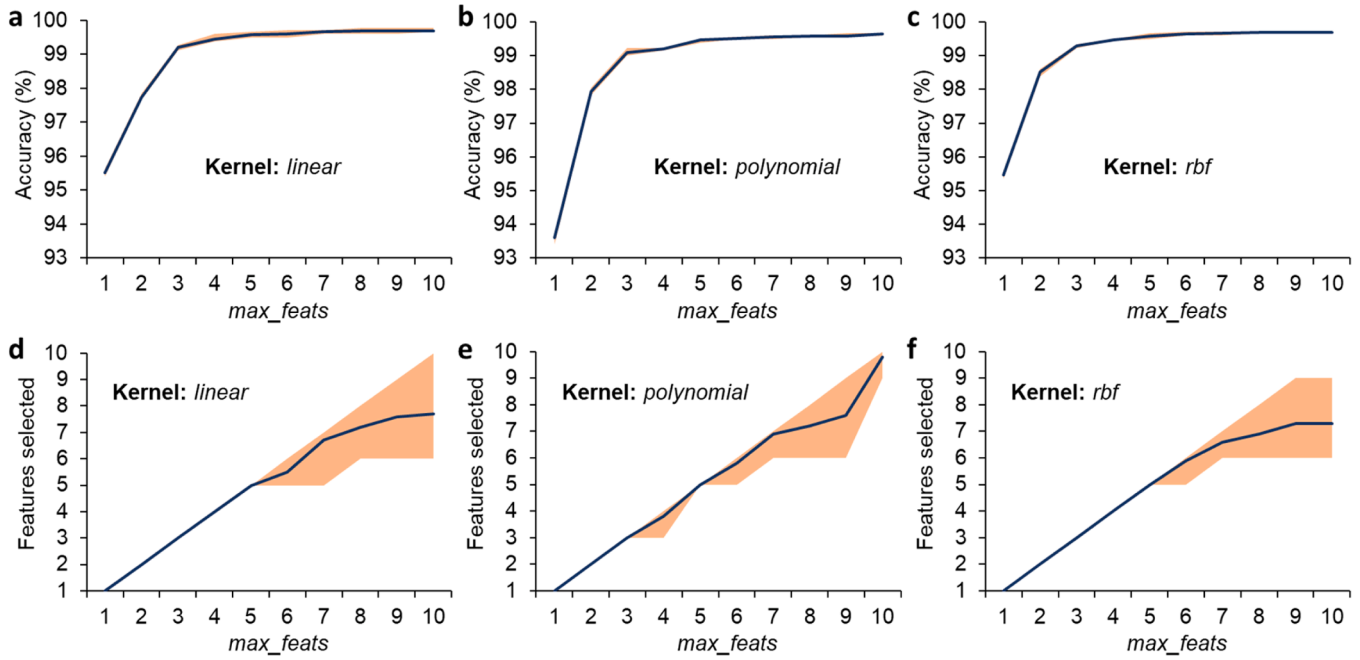

Fig. 5. Prediction accuracy (a–c) and the corresponding feature selection frequency (d–f) of machine learning models at different maximum feature thresholds (*max_feats*; abscissa) for *linear, polynomial*, and *radial basis function* (*rbf*) kernels in tests following exhaustive feature selection (ML-2). Lines represent the average output for $n = 10$ model instances generated using different random states and the shaded region indicates the range of output values for prediction accuracy (a–c) and number of features selected (d–f) at each *max_feats* threshold.

kernel. NDVI was ranked fifth for the *lin* and *poly* kernels, whereas LAI received the fifth rank for the *rbf* kernel. The features having relatively lower ranks varied considerably across the three kernels, with features such as DB, LPD, G/R, and NIR having the lowest overall ranks (Table 4).

4. Discussion

4.1. Trends in data distribution and preliminary ML analyses

Different inoculation strengths across different growth stages (Table 1) were used to mimic real-world pathogen uncertainty, and to obtain a more varied trend in data distribution between the healthy and infected plants. In practice, factors such as level of pathogen exposure, age of plants at infection, and degree of plant damage may result in diverse symptomatic changes, complicating the distinction of stressed plants from healthy ones using fixed thresholds for individual parameters. For instance, parameters such as T_c (Fig. 3f), Ht_max (Fig. 3b), and GLI (Fig. 4b) showed very low overlap between the control and infected samples. In contrast, parameters such as LAI, LInc, LPD (Fig. 3c–e), NPCI, and NIR (Fig. 4d, h) exhibited significant overlap (JI $\geq$ 0.5 and SS $\geq$ 0.7) between the two classes. Hence, using ML-based disease detection systems capable of distinguishing between healthy and infected plants by assessing simultaneous trends amongst multiple plant attributes at early stages of infection may help bypass this issue and enhance the precision of disease diagnosis.

In this context, as a preliminary co-assessment of multiple plant attributes, ML-1 tests revealed that prediction accuracy was dependent on the type of feature rather than the total number of features within each model (Table 3). Specifically, models trained using 11 spectral features as well as models with only one thermal feature had accuracies of 98.34 % and 94.18 %, respectively, whereas the model with only morphometric data, i.e., 5 features, had an accuracy of 89.92 %. This indicates that using a high number of features may not necessarily improve model performance. Instead, it may increase computational load leading to slowed data processing [69,70]. Conversely, using fewer but more reliable or "informative" features would be more beneficial for optimizing model performance. In-depth analyses by EFS with different kernels and *max_feats* thresholds in ML-2 tests allowed us to explore these trends with more granularity.

4.2. EFS-based minimal feature subsets for multi-sensor data fusion

Feature subset refinement via EFS in ML-2 analyses was helpful in reducing "superfluous data", i.e., data that increases computational load without making any significant contribution towards model improvement [69]. Herein, models with *max_feats* = 1 achieved accuracies of

Table 4
Scores and ranks calculated using the frequency of occurrence during tests with exhaustive feature selection (ML-2) for image-based canopy attributes of flat-leaf parsley samples.

Overall rank#*	Features		*lin*		*poly*		*rbf*		Total score
	Name	Type	Score	Rank#	Score	Rank#	Score	Rank#	
1	DPT	D	90	1	90	1	90	1	270
2	T_c	T	80	2	79	2	79	2	238
3	GLI	S	62	4	73	3	69	3	204
4	Ht_max	M	68	3	57	4	56	4	181
5	NDVI	S	50	5	50	5	24	7	124
6	PSRI	S	46	6	41	6	16	8	103
7	LAI	M	0	17	13	11	49	5	62
8	R	S	3	15	9	13	28	6	40
9	NPCI	S	4	13	20	10	14	10	38
9	G	S	32	7	1	16	5	15	38
9	GMR	S	14	8	24	7	0	18	38
12	LInc	M	4	13	21	9	9	13	34
13	AGRI	S	11	11	10	12	10	11	31
14	Hue	S	0	17	23	8	3	17	26
15	G/R	S	14	8	0	18	10	11	24
15	NIR	S	5	12	4	15	15	9	24
17	LPD	M	12	10	1	16	9	13	22
18	DB	M	2	16	5	14	4	16	11
	Maximum Score		***100***	–	***100***	–	***100***	–	***300***

Features: AGRI, Augmented Green-Red Index; DB, Digital biomass; DPT, days post transplantation; G, Green reflectance; G/R, Green-Red reflectance ratio; GLI, Green Leaf Index; GMR, Green-minus-Red reflectance; Ht_max, maximum plant height; Hue, hue angle; LAI, leaf area index; LInc, leaf inclination; LPD, light penetration depth; NDVI, Normalized Difference Vegetation Index; NIR, Near-Infrared reflectance; NPCI, Normalized Pigment Chlorophyll ratio Index; PSRI, Plant Senescence Reflectance Index; R, Red reflectance; T_c, canopy temperature. **Feature types:** D, temporal; M, morphometric; S, spectral; T, thermal. **Kernels:** *lin, linear; poly, polynomial; rbf, radial basis function.* Score, the number of times each feature was selected out of 100 trials for each kernel. *Maximum score*, maximum possible score for 10 random states × 10 maximum feature thresholds. #Features with the same score were assigned the same rank. Total score, sum of scores across the three kernels. *Overall rank was calculated based on total score. Details of feature co-occurrence are provided in Supplementary file S2.

93.4–95.7 % (n = 30) (Fig. 5a–c). Strikingly, all thirty models relied exclusively on AGRI (Supplementary file S2), and performed marginally better than the models generated using T_c alone in ML-1 (accuracy 92.8–95.6 % across five cross-validation trials; Table 3). However, AGRI was rarely selected for higher *max_feats* values (Supplementary file S2), wherein the algorithm could combine features to capture stronger interactions and relative trends amongst the selected features. Thus, while AGRI was highly effective as a stand-alone ML attribute without temporal data, incorporating additional features provided greater flexibility and yielded more accurate models when larger feature sets and temporal information were available (Fig. 5, Supplementary file S2). Exploring this aspect via gradual increments in EFS *max_feats* values highlighted the role of feature diversity and interactions for ML.

In general, prediction accuracy improved by increasing the *max_feats* threshold from 1 to 5, with a tendency of performance saturation around *max_feats* = 3–5 (Fig. 5a–c). In particular, model accuracy increased steadily from *max_feats* = 1 to *max_feats* = 3 by ca. 3–6 %, following which the accuracy improved only by <0.7 % up to *max_feats* = 5, with negligible further increase (<0.1 %) up to *max_feats* = 10. Further, all three kernels were able to deliver >99 % accuracy at *max_feats* = 3. This suggests that the present method could yield very reliable results with as low as three to five input parameters irrespective of the modelling approach, i.e., linear or non-linear. Notably, such high-accuracy, low-feature-threshold (*max_feats* = 3–5) models frequently combined features belonging to multiple categories, i.e., spectral, morphometric, thermal, and temporal. This clearly indicates that amalgamation of data from different sources created the most efficient plant stress detection models due to the complementarity of features with reduced redundancy in available information.

Interestingly, use of one feature each from three different data categories (Supplementary file S2) yielded higher prediction accuracy than using considerably more features belonging to the same category (Table 3). For example, use of DPT, GLI, and T_c, i.e., three features from three different categories, gave an average accuracy of 99.3 % for seven *rbf* models (Supplementary file S2) as compared to the accuracy of 98.61 % or 98.95 % obtained by using DPT with all five morphometric or all eleven spectral features, respectively, as observed in ML-1 (Table 3). This trend of better model prediction using fewer plant attributes obtained from diverse imaging platforms further corroborates our hypothesis that amalgamation of information from multiple plant sensors may potentially improve plant stress detection via ML.

In the current scenario, where numerous features or variables can be obtained from plant images with relative ease, identifying the most useful image attributes is crucial for optimizing the modeling pipeline to minimize data processing time and overfitting by eliminating redundant and/or repetitive information by dimensionality reduction. To achieve this, a variety of approaches, ranging from statistical filters such as correlation analysis, Lilliefors test for normality, analysis of variance, and multiple comparison tests [37,71], as well as advanced techniques such as recursive feature elimination, adaptive rain optimization [72], adaptive red fox algorithm [73], and black widow optimization [74], have been explored. Further, some of these studies have even employed multiple feature selection techniques for consensus-based selection or stepwise feature elimination.

In the present study, we followed stepwise feature selection, starting with regression analysis to identify and eliminate highly repetitive features. However, unlike many other studies, we implemented EFS, which is considered a "brute-force" algorithm as it tests all possible feature combinations to identify the best possible subset of features. Although EFS is highly computationally expensive and impractical for very large number of features [75], it was adopted here considering the relatively low number of features being tested. Moreover, while many studies focusing on plant disease detection have utilized feature selection for identifying texture and color related features using RGB images [76] or spectral indices and wavebands for hyperspectral images [8,55], this study investigates the selection of features across three types imaging platforms simultaneously, which remains hitherto unexplored for plant disease detection.

4.3. Impact of kernel on feature selection

Assessment of feature rankings based on EFS scores allowed the

identification of attributes that would be more suitable for detecting stressed plants via ML with specific kernels. Herein, rankings of sensor-derived attributes (Table 4) indicated that T_c, GLI, and Ht_max were generally favored by all three kernels, whereas features such as DB, LPD, G/R and NIR were selected least frequently. Notably, feature preference differed considerably for each kernel. For example, while LAI ranked fifth with the *rbf* kernel (score 49/100), it ranked eleventh for the *poly* kernel (score 13/100), and was never selected by the *lin* kernel (score 0/100). Similarly, G was ranked seventh (score 32/100) with the *lin* kernel, but was ranked fifteenth (score 5/100) and sixteenth (score 1/100) for the *rbf* and *poly* kernels, respectively. This highlights the diversity in feature preference across the different kernels depending on how multiple features are co-interpreted for generating the classification model by each kernel.

Despite variations in individual feature preferences by the different kernels, simultaneous selection of attributes from multiple feature categories remained consistent. Although this trend could be expected for very high *max_feats* thresholds, it was notable even at lower threshold values (Supplementary file S2). For instance, 29/30 models with *max_feats* = 3 selected one feature each from three different feature categories. Further, 10/10 *lin* models and 4/10 *rbf* models at *max_feats* = 4 chose one feature of each type, i.e., temporal, thermal, spectral, and morphometric, whereas 10/10 *poly* models and the remaining 6/10 *rbf* models chose features belonging to three different categories. At *max_feats* = 5, selection of features from all four categories was observed in 29/30 models. Thus, it may be inferred that the modelling algorithms for all kernels inherently attempted to select a more diverse dataset in terms of feature type to maximize model accuracy while using as few features as possible.

4.4. Significance of feature informativeness

Frequent selection of plant attributes such as Ht_max, T_c, and GLI by all three kernels (Table 4) may be attributed to the feasibility of distinguishing between healthy and infected samples based on the numerical ranges, as also supported by the three statistical metrics for data distribution and overlap (Figs. 3b, f, 4b). In contrast, parameters with lower feature scores such as LPD and NIR (Table 4) were chosen less frequently during EFS, likely because of greater similarity between the datasets of control and infected samples (Figs. 3e, 4h). Likewise, high overlap between both classes in DB values (SS = 0.79; Fig. 3a) explains its infrequent selection as well. Hence, it may be inferred that features which allowed better separability between the control and infected samples due to limited overlap in numerical ranges were selected more frequently by the algorithm as they were deemed more "informative" for classification.

Notably, in addition to the overall feature quality in terms of discernibility between control vs. infected samples, non-redundancy of data also played a role in determining feature selection by the algorithm. For instance, G/R was selected infrequently by the *lin* and *rbf* kernels, and was never selected by the *poly* kernel (Table 4) despite having very low similarity between the control and infected sample datasets (Fig. 4i). Notably, only 15/240 models with *max_feats* $\geq$ 3 selected both G/R and GLI simultaneously, whereas the *lin* models with *max_feats* = 2 always chose either of the two (Supplementary file S2). A possible explanation for this might be the high correlation between GLI and G/R datasets (r^2 = 0.92; Supplementary file S1), which might have led the algorithm to consider the G/R data redundant in the presence of GLI, with the latter being selected more frequently due to potentially better interaction with other features and stronger compatibility with the different kernels. Thus, in addition to feature informativeness, its uniqueness also played an important part during EFS. While a stringent threshold of $r^2 > 0.95$ for preliminary feature shortlisting addressed this issue up to some extent (Supplementary file S1), a more flexible threshold of $r^2 > 0.9$ would enable more redundant features to be screened via correlation analysis, and could be tested in the future.

4.5. Role of temporal information

Comparing model accuracies with and without DPT in the ML-1 tests revealed that model performance improved for all types of sensor datasets upon including temporal information (Table 3). Although the extent of improvement varied between the morphometric, spectral, and thermal datasets, the findings concurrently suggest that temporal information helped generate more robust ML-based disease detection models. Interestingly, despite being numerically identical for both control and infected datasets, DPT had the highest feature score in ML-2 tests as it was selected most frequently amongst all features (Table 4), and was the only feature that was selected for all models with *max_feats* $\geq$ 2 (Supplementary file S2). It enabled the creation of models capable of providing highly reliable outcomes (97.5–99 % accuracy) even with as low as one sensor-based attribute, as observed in the ML-2 tests with *max_feats* = 2 (Supplementary file S2). Moreover, additional EFS trials for *max_feats* 1 to 5 excluding DPT (Supplementary Fig. S1) indicated that model performance was ca. 1 % less accurate for all kernels as compared to the corresponding models which included DPT.

Since multitemporal imaging allows the tracking of disease symptom progression, it adds a key dimension to image dataset: time. This creates the scope for more accurate predictions by creating a timeline within the dataset. While symptoms of plant disease become more obvious with time, the advantage of computer vision and ML lies in the potential for early detection via even the slightest symptoms based on the time-series models. Deep-learning approaches involving NNs have been used frequently to generate such temporal data maps for crop monitoring [77]. However, since the present study was conducted with a relatively low sample size ($n < 1000$) for each class, an SVM-based model was adopted instead of NNs to avoid overfitting. While NN-based models can store observations as a function of time, here DPT was included within the dataset as an independent attribute parallel to the image-derived features for simplifying modeling and analysis. Nonetheless, the positive impact of including this temporal attribute was clearly visible for all types of models tested (Tables 3, 4).

As demonstrated in our earlier study [37], morphological and spectral differences between healthy and infected plants became more prominent over time as the effects of infection intensified. Thus, it would be logical to assume that the changes in plant attributes could be interpreted by the ML algorithm as temporal functions. However, this can only be possible in the presence of temporal indicators such as DPT. Hence, it can be inferred that DPT values, despite being identical for both classes, provided crucial supporting information during sample classification by adding an extra dimension of temporal resolution for better mapping of sensor-based traits. This phenomenon further reiterates the importance of feature interaction, wherein the ML algorithm co-evaluates information from multiple features and classifies samples based on mutual trends. Furthermore, based on the experimental design and high accuracy of predictions (Table 3, Fig. 5), it may be assumed that such models could potentially detect strongly affected plants as early as 2–3 days after being infected, highlighting the scope for early disease detection.

4.6. Limitations, current perspectives, and future scope

Simultaneously utilizing multiple imaging sensors for crop monitoring is on the rise [78,79], with ML playing a crucial role in streamlining image processing [45]. Earlier studies focusing on multi-sensor data fusion by ML have explored diverse aspects of crop monitoring by combining spectral and 3D data, including estimation of leaf nitrogen content in rice via SVM [80], wheat biomass prediction by implementing three ML approaches, viz., Support Vector Regression, Random Forest, and Extreme Learning Machine [81], as well as assessment of biomass and nitrogen-fixation in legume-grass mixtures using Random Forest [82]. Additionally, a recent study by our group demonstrated the potential of improving root-rot detection by combining 3D and spectral

data via Principal Component Analysis, an unsupervised ML technique [37]. Notably, all these studies adopted different "shallow learning" methods, which are generally more resource-efficient than deep learning methods in terms of training dataset requirements, computational load, and processing time, while still being able to provide realistically reliable outcomes. This was also indicated in the study by Pérez-Bueno et al. [57], wherein various ML methods were compared while co-analyzing thermal and spectral data for detecting root rot in avocado trees.

The present study delved deeper into this topic to ascertain the potential of synergistically utilizing 3D, multispectral, and thermal data for detecting root rot accurately. Although a limited experimental investigation was carried out using one hydroponically-grown crop and only one type of ML algorithm, our study highlighted that co-analysis of data from all three types of sensing platforms along with temporal information improved plant stress detection remarkably. Future investigations with diverse crops and a range of diseases within both indoor and field production systems will further expand the knowledge base for monitoring plant health status via multi-sensor approaches. Additionally, while the current study puts equal emphasis on data collected between 2 to 18 DPT, future experimental trials focusing more on the earlier intervals could help develop more robust ML models with early detection capabilities. For this, experiments incorporating a range of shallow and deep ML algorithms, including Decision Trees and Convolutional NN, could further enhance the ML pipeline. Integrating different feature selection and dimensionality reduction tools, such as Principal Component Analysis, Linear Discriminant Analysis, and Autoencoders, would further optimize data processing by minimizing data redundancy and boosting computational efficiency, hence paving the way for higher prediction accuracies during multi-sensor crop monitoring.

5. Conclusion

Considering the ever-growing need for improving machine vision-based protocols for crop monitoring, our study highlights the potential for identifying diseased plants with high accuracy by fusing data from 3D, spectral, and thermal sensors via ML. Our findings indicate that combining the information from all three sensors provides more accurate results than using the sensors independently. Further, pairing sensor-based data with temporal information improves model performance considerably. Assessing data distribution between the control and infected samples along with feature scores and co-occurrence trends indicated that feature informativeness in terms of its discerning capacity for the two classes as well as its uniqueness play an important role during automated feature selection for generating the classification model. Future investigations with diverse plant–pathogen combinations, growth conditions, monitoring intervals, and ML approaches implementing NN would help further optimize such sensor fusion-based disease detection models and enable us to implement them with higher fidelity in large scale cultivation systems.

CRediT authorship contribution statement

Avinash Agarwal: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Filipe de Jesus Colwell:** Writing – review & editing, Resources, Methodology. **Julian Bello Rodriguez:** Writing – review & editing, Methodology. **Sarah Sommer:** Writing – review & editing, Methodology. **Monica Barman:** Writing – review & editing, Formal analysis. **Viviana Andrea Correa Galvis:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Tom R. Hill:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Neil Boonham:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Ankush Prashar:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We thank the InFarm UK team for supplying seedlings and providing technical support, along with InFarm Crop Science team (Germany) for their support. We acknowledge all the partners (RoboScientific, Marks and Spencer, and InFarm) for their feedback and support in the project. We also thank the staff at Newcastle University for their technical, administrative and logistical support. AA thanks Forschungszentrum Jülich GmbH for support during manuscript preparation.

Funding

This work was funded by Innovate UK (Technology Strategy Board – CR&D) [grant number: TS/V002880/1].

Ethics statement

Not applicable: This manuscript does not include human or animal research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2025.101364.

Data availability

Data will be made available on request.

References

[1] C. Eigenbrod, N. Gruda, Urban vegetable for food security in cities. a review, Agron. Sustain. Dev. 35 (2015) 483–498, https://doi.org/10.1007/s13593-014-0273-y.
[2] B. Goldstein, M. Hauschild, J. Fernández, M. Birkved, Urban versus conventional agriculture, taxonomy of resource profiles: a review, Agron. Sustain. Dev. 36 (2016) 9, https://doi.org/10.1007/s13593-015-0348-4.
[3] K. Benke, B. Tomkins, Future food-production systems: vertical farming and controlled-environment agriculture, Sustain.: Sci. Pract. Policy 13 (2018) 13–26, https://doi.org/10.1080/15487733.2017.1394054.
[4] J. Vallance, F. Déniel, Floch G Le, L. Guérin-Dubrana, D. Blancard, P. Rey, Pathogenic and beneficial microorganisms in soilless cultures, in: E. Lichtfouse, M. Hamelin, M.D.P. Navarrete (Eds.), Sustainable Agriculture, Sustainable Agriculture, 2, Springer, Netherlands, 2011, pp. 711–726, https://doi.org/10.1007/978-94-007-0394-0_31.
[5] J.M. Roberts, T.J.A. Bruce, J.M. Monaghan, T.W. Pope, S.R. Leather, A. M. Beacham, Vertical farming systems bring new considerations for pest and disease management, Ann. Appl. Biol. 176 (2020) 226–232, https://doi.org/10.1111/aab.12587.
[6] T.C. Paulitz, Biological control of root pathogens in soilless and hydroponic systems, HortScience 32 (1997) 193–196, https://doi.org/10.21273/HORTSCI.32.2.193.
[7] G.C.S. Laevens, W.C. Dolson, M.M. Drapeau, S. Telhig, S.E. Ruffell, D.M. Rose, B. R. Glick, A.A. Stegelmeier, The good, the bad, and the fungus: insights into the relationship between plants, fungi, and oomycetes in hydroponics, Biol. (Basel) 13 (2024) 1014, https://doi.org/10.3390/biology13121014.
[8] C.H. Bock, G.H. Poole, P.E. Parker, T.R. Gottwald, Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging, CRC. Crit. Rev. Plant Sci. 29 (2010) 59–107, https://doi.org/10.1080/07352681003617285.
[9] J. Zhang, Y. Huang, R. Pu, P. Gonzalez-Moreno, L. Yuan, K. Wu, W. Huang, Monitoring plant diseases and pests through remote sensing technology: a review, Comput. Electron. Agric. 165 (2019) 104943, https://doi.org/10.1016/j.compag.2019.104943.

[10] P. Waiphara, C. Bourgenot, L.J. Compton, A. Prashar, Optical imaging resources for crop phenotyping and stress detection, in: P. Duque, D. Szakonyi (Eds.), Methods in Molecular Biology, Methods in Molecular Biology, 2494, Humana, New York, 2022, pp. 255–265, https://doi.org/10.1007/978-1-0716-2297-1_18.
[11] M. Bietresato, G. Carabin, R. Vidoni, A. Gasparetto, F. Mazzetto, Evaluation of a LiDAR-based 3D-stereoscopic vision system for crop-monitoring applications, Comput. Electron. Agric. 124 (2016) 1–13, https://doi.org/10.1016/j.compag.2016.03.017.
[12] J. Sandhu, F. Zhu, P. Paul, T. Gao, B.K. Dhatt, Y. Ge, P. Staswick, H. Yu, H. Walia, PI-plat: a high-resolution image-based 3D reconstruction method to estimate growth dynamics of rice inflorescence traits, Plant Methods 15 (2019) 162, https://doi.org/10.1186/s13007-019-0545-2.
[13] L. Xiang, Y. Bao, L. Tang, D. Ortiz, M.G. Salas-Fernandez, Automated morphological traits extraction for sorghum plants via 3D point cloud data analysis, Comput. Electron. Agric. 162 (2019) 951–961, https://doi.org/10.1016/j.compag.2019.05.043.
[14] Jesus de, F. Colwell, J. Souter, G.J. Bryan, L.J. Compton, N. Boonham, A. Prashar, Development and validation of methodology for estimating potato canopy structure for field crop phenotyping and improved breeding, Front. Plant Sci. 12 (2021) 612843, https://doi.org/10.3389/fpls.2021.612843.
[15] A. Agarwal, S. Dutta Gupta, Assessment of spinach seedling health status and chlorophyll content by multivariate data analysis and multiple linear regression of leaf image features, Comput. Electron. Agric. 152 (2018) 281–289, https://doi.org/10.1016/j.compag.2018.06.048.
[16] A. Agarwal, P.K. Dongre, S. Dutta Gupta, Smartphone-assisted real-time estimation of chlorophyll and carotenoid concentrations and ratio using the inverse of red and green digital color features, Theor. Exp. Plant Physiol. 33 (2021) 293–302, https://doi.org/10.1007/s40626-021-00210-4.
[17] F. Zhu, M. Saluja, J.S. Dharni, P. Paul, S.E. Sattler, P. Staswick, H. Walia, H. Yu, PhenoImage: an open-source graphical user interface for plant image analysis, Plant Phenome J. 4 (2021) e20015, https://doi.org/10.1002/ppj2.20015.
[18] S. Osuna-Caballero, T. Olivoto, M.A. Jiménez-Vaquero, D. Rubiales, N. Rispail, RGB image-based method for phenotyping rust disease progress in pea leaves using R, Plant Methods 19 (2023) 86, https://doi.org/10.1186/s13007-023-01069-z.
[19] A. Prashar, H.G. Jones, Infra-red thermography as a high-throughput tool for field phenotyping, Agronomy 4 (2014) 397–417, https://doi.org/10.3390/agronomy4030397.
[20] M.A. Jiménez-Bello, C. Ballester, J.R. Castel, D.S. Intrigliolo, Development and validation of an automatic thermal imaging process for assessing plant water status, Agric. Water. Manage 98 (2011) 1497–1504, https://doi.org/10.1016/j.agwat.2011.05.002.
[21] M.F. Buitrago, T.A. Groen, C.A. Hecker, A.K. Skidmore, Changes in thermal infrared spectra of plants caused by temperature and water stress, ISPRS J. Photogramm. Remote Sens. 111 (2016) 22–31, https://doi.org/10.1016/j.isprsjprs.2015.11.003.
[22] M. Pineda, M. Barón, Pérez-Bueno ML, Thermal imaging for plant stress detection and phenotyping, Remote Sens. 13 (2021) 68, https://doi.org/10.3390/rs13010068.
[23] M. Smigaj, A. Agarwal, H. Bartholomeus, M. Decuyper, A. Elsherif, A. de Jonge, L. Kooistra, Thermal infrared remote sensing of stress responses in forest environments: a review of developments, challenges, and opportunities, Curr. For. Rep. 10 (2024) 56–76, https://doi.org/10.1007/s40725-023-00207-z.
[24] A. Agarwal, F. de Jesus Colwell, R. Dinnis, V.A. Correa Galvis, T.R. Hill, N. Boonham, A. Prashar, Infrared thermography in plant factories: solving spatiotemporal variations via Machine Learning, Mod. Agric. 3 (2025) e70012, https://doi.org/10.1002/moda.70012.
[25] A. Prashar, J. Yildiz, J.W. McNicol, G.J. Bryan, H.G. Jones, Infra-red thermography for high throughput field phenotyping in Solanum tuberosum, PLoS. One 8 (2013) e65816, https://doi.org/10.1371/journal.pone.0065816.
[26] A.V. Zubler, J.Y. Yoon, Proximal methods for plant stress detection using optical sensors and machine learning, Biosens. (Basel) 10 (2020) 193, https://doi.org/10.3390/BIOS10120193.
[27] S.M. Shawon, F.B. Ema, A.K. Mahi, F.L. Niha, H.T. Zubair, Crop yield prediction using machine learning: an extensive and systematic literature review, Smart Agric. Technol. 10 (2025) 100718, https://doi.org/10.1016/j.atech.2024.100718.
[28] D.S. Dhaliwal, M.M. Williams, Sweet corn yield prediction using machine learning models and field-level data, Precis. Agric. 25 (2024) 51–64, https://doi.org/10.1007/s11119-023-10057-1.
[29] B. Jin, X. Xu, Machine learning coffee price predictions, J. Uncertain Syst. 17 (2024) 2450023, https://doi.org/10.1142/S1752890924500235.
[30] B. Jin, X. Xu, Y. Zhang, Peanut oil price change forecasts through the neural network, Foresight. 27 (2025) 595–612, https://doi.org/10.1108/FS-01-2023-0016.
[31] N. Serda Kaya, O. Dengiz, Assessment of the neutrosophic fuzzy-AHP and predictive power of some machine learning approaches for maize silage soil quality, Comput. Electron. Agric. 226 (2024) 109446, https://doi.org/10.1016/j.compag.2024.109446.
[32] K. Rangzan, Z. Zaheri Abdehvand, S.R. Mousavi, D. Karimi, Spatial analysis of soil quality in agricultural land using machine learning and environmental covariates: a case study of Khuzestan Province, Soil. Tillage Res. 252 (2025) 106591, https://doi.org/10.1016/j.still.2025.106591.
[33] L.C. Ngugi, M. Abelwahab, M. Abo-Zahhad, Recent advances in image processing techniques for automated leaf pest and disease recognition – A review, Inf. Process. Agric. 8 (2021) 27–51, https://doi.org/10.1016/j.inpa.2020.04.004.
[34] H. Lin, H. Sheng, G. Sun, Y. Li, M. Xiao, X. Wang, Identification of pumpkin powdery mildew based on image processing PCA and machine learning, Multimed. Tools. Appl. 80 (2021) 21085–21099, https://doi.org/10.1007/s11042-020-10419-1.
[35] Y. Song, G. Teng, Y. Yuan, T. Liu, Z. Sun, Assessment of wheat chlorophyll content by the multiple linear regression of leaf image features, Inf. Process. Agric. 8 (2021) 232–243, https://doi.org/10.1016/j.inpa.2020.05.002.
[36] S.M. Javidan, A. Banakar, K.A. Vakilian, Y. Ampatzidis, Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning, Smart Agric. Technol. 3 (2023) 100081, https://doi.org/10.1016/j.atech.2022.100081.
[37] A. Agarwal, F. de Jesus Colwell, J. Bello Rodriguez, S. Sommer, V.A. Correa Galvis, T. Hill, N. Boonham, A. Prashar, Monitoring root rot in flat-leaf parsley via machine vision by unsupervised multivariate analysis of morphometric and spectral parameters, Eur. J. Plant Pathol. 169 (2024) 359–377, https://doi.org/10.1007/s10658-024-02834-z.
[38] M.R. Larijani, E.A. Asli-Ardeh, E. Kozegar, R. Loni, Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means, Food Sci. Nutr. 7 (2019) 3922–3930, https://doi.org/10.1002/fsn3.1251.
[39] A. Wójtowicz, J. Piekarczyk, B. Czernecki, H. Ratajkiewicz, A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale, J. Photochem. Photobiol. B 223 (2021) 112278, https://doi.org/10.1016/j.jphotobiol.2021.112278.
[40] A. Agarwal, F. de Jesus Colwell, V.A. Correa Galvis, T.R. Hill, N. Boonham, A. Prashar, Assessing nutritional pigment content of green and red leafy vegetables by image analysis: catching the “red herring” of plant digital color processing via machine learning, Biol. Methods Protoc. (2025), https://doi.org/10.1093/biomethods/bpaf027.
[41] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016) 1419, https://doi.org/10.3389/fpls.2016.01419.
[42] K.P. Ferentinos, Deep learning models for plant disease detection and diagnosis, Comput. Electron. Agric. 145 (2018) 311–318, https://doi.org/10.1016/j.compag.2018.01.009.
[43] Y. Jiang, C. Li, Convolutional neural networks for image-based high-throughput plant phenotyping: a review, Plant Phenomics. (2020) 4152816, https://doi.org/10.34133/2020/4152816, 2020.
[44] A.M. Mutka, R.S. Bart, Image-based phenotyping of plant disease symptoms, Front. Plant Sci. 5 (2015) 734, https://doi.org/10.3389/fpls.2014.00734.
[45] A. Singh, S. Jones, B. Ganapathysubramanian, S. Sarkar, D. Mueller, K. Sandhu, K. Nagasubramanian, Challenges and opportunities in machine-augmented plant stress phenotyping, Trends. Plant Sci. 26 (2021) 53–69, https://doi.org/10.1016/j.tplants.2020.07.010.
[46] C.X. Hong, G.W. Moorman, Plant pathogens in irrigation water: challenges and opportunities, CRC. Crit. Rev. Plant Sci. 24 (2005) 189–208, https://doi.org/10.1080/07352680591005838.
[47] E. Minchinton, J. Petkowski, D. deBoer, F. Thomson, L. Trapnell, L. Tesoriero, L. Forsyth, J. Parker, H. Pung, A. McKay, Identification of IPM strategies for Pythium induced root rots in Apiacae vegetable crops, Sydney (2013).
[48] G.P. Suárez-Cáceres, L. Pérez-Urrestarazu, M. Avilés, C. Borrero, J.R. Lobillo Eguíbar, V.M Fernández-Cabanás, Susceptibility to water-borne plant diseases of hydroponic vs. aquaponics systems, Aquaculture 544 (2021) 737093, https://doi.org/10.1016/j.aquaculture.2021.737093.
[49] E.G. Kokko, R.L. Conner, B. Lee, A.D. Kuzyk, G.C. Kozu, Quantification of common root rot symptoms in resistant and susceptible barley by image analysis, Can. J. Plant Pathol. 22 (2000) 38–43, https://doi.org/10.1080/07060660009501159.
[50] A. Marzougui, Y. Ma, C. Zhang, R.J. McGee, C.J. Coyne, D. Main, S. Sankaran, Advanced imaging for quantitative evaluation of aphanomyces root rot resistance in lentil, Front. Plant Sci. 10 (2019) 383, https://doi.org/10.3389/fpls.2019.00383.
[51] L.D. Pierz, D.R. Heslinga, C.R. Buell, M.J. Haus, An image-based technique for automated root disease severity assessment using PlantCV, Appl. Plant Sci. 11 (2023) e11507, https://doi.org/10.1002/aps3.11507.
[52] E. Granum, M.L. Pérez-Bueno, C.E. Calderón, C. Ramos, A. de Vicente, F. M. Cazorla, M. Barón, Metabolic responses of avocado plants to stress induced by *Rosellinia necatrix* analysed by fluorescence and thermal imaging, Eur. J. Plant Pathol. 142 (2015) 625–632, https://doi.org/10.1007/s10658-015-0640-9.
[53] P.K. Jayapal, E. Park, M.A. Faqeerzada, Y.-S. Kim, H. Kim, I. Baek, M.S. Kim, D. Sandanam, B.-K. Cho, Analysis of RGB plant images to identify root rot disease in Korean ginseng plants using deep learning, Appl. Sci. 12 (2022) 2489, https://doi.org/10.3390/app12052489.
[54] A.S.A. Salgadoe, A.J. Robson, D.W. Lamb, E.K. Dann, C. Searle, Quantifying the severity of phytophthora root rot disease in avocado trees using image analysis, Remote Sens. 10 (2018) 226, https://doi.org/10.3390/rs10020226.
[55] E. Park, Y.S. Kim, M.A. Faqeerzada, M.S. Kim, I. Baek, B.K. Cho, Hyperspectral reflectance imaging for nondestructive evaluation of root rot in Korean ginseng (*Panax ginseng* Meyer), Front. Plant Sci. 14 (2023) 1109060, https://doi.org/10.3389/fpls.2023.1109060.
[56] S. Weksler, O. Rozenstein, N. Haish, M. Moshelion, R. Wallach, E. Ben-dor, Pepper plants leaf spectral reflectance changes as a result of root rot damage, Remote Sens 13 (2021) 980, https://doi.org/10.3390/rs13050980.
[57] M.L. Pérez-Bueno, M. Pineda, C. Vida, D. Fernández-Ortuño, J.A. Torés, A. de Vicente, F.M. Cazorla, M. Barón, Detection of white root rot in avocado trees by remote sensing, Plant Dis. 103 (2019) 1119–1125, https://doi.org/10.1094/PDIS-10-18-1778-RE.
[58] R.L. Matthiesen, A.A. Ahmad, A.E. Robertson, Temperature affects aggressiveness and fungicide sensitivity of four pythium spp. That cause soybean and corn

damping off in Iowa, Plant Dis. 100 (2016) 583–591, https://doi.org/10.1094/PDIS-04-15-0487-RE.
[59] P.V Oudemans, Phytophthora species associated with cranberry root rot and surface irrigation water in New Jersey, Plant Dis. 83 (1999) 251–258, https://doi.org/10.1094/PDIS.1999.83.3.251.
[60] T.J. White, T. Bruns, S. Lee, J.W Taylor, Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: M.A. Innis, D.H. Gelfand, J. J. Sninsky, T.J. White (Eds.), PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., New York, 1990, pp. 315–322.
[61] F.N. Martin, Phylogenetic relationships among some Pythium species inferred from sequence analysis of the mitochondrially encoded cytochrome oxidase II gene, Mycologia 92 (2000) 711–727, https://doi.org/10.1080/00275514.2000.12061211.
[62] G.J. Bilodeau, F.N. Martin, M.D. Coffey, C.L. Blomquist, Development of a multiplex assay for genus- and species-specific detection of phytophthora based on differences in mitochondrial gene order, Phytopathology 104 (2014) 733–748, https://doi.org/10.1094/PHYTO-09-13-0263-R.
[63] C.S. McGehee, R.E. Raudales, W.H. Elmer, R.J. McAvoy, Efficacy of biofungicides against root rot and damping-off of microgreens caused by pythium spp, Crop Prot. 121 (2019) 96–102, https://doi.org/10.1016/j.cropro.2018.12.007.
[64] A. Prashar, H.G. Jones, Assessing Drought Responses Using Thermal Infrared imaging. Methods in Molecular Biology, 1398, Humana Press Inc., Duque, Paula, New York, 2016, pp. 209–219, https://doi.org/10.1007/978-1-4939-3356-3_17.
[65] V. Zieschank, R.R. Junker, Digital whole-community phenotyping: tracking morphological and physiological responses of plant communities to environmental changes in the field, Front. Plant Sci. 14 (2023) 1141554, https://doi.org/10.3389/fpls.2023.1141554.
[66] S. Kawashima, M. Nakatani, An algorithm for estimating chlorophyll content in leaves using a video camera, Ann. Bot. 81 (1998) 49–54, https://doi.org/10.1006/anbo.1997.0544.
[67] S.P. Yadav, Y. Ibaraki, S. Dutta Gupta, Estimation of the chlorophyll content of micropropagated potato plants using RGB based image analysis, Plant Cell Tissue Organ. Cult. 100 (2010) 183–188, https://doi.org/10.1007/s11240-009-9635-6.
[68] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297, https://doi.org/10.1007/BF00994018.
[69] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (2014) 16–28, https://doi.org/10.1016/j.compeleceng.2013.11.024.
[70] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: a data perspective, ACM Comput. Surv. 50 (2017) 94, https://doi.org/10.1145/3136625.
[71] C.A. Rivera-Romero, E.R. Palacios-Hernández, J.U. Muñoz-Minjares, O. Vite-Chávez, R. Olivera-Reyna, I.A Reyes-Portillo, Recognition in the early stage of powdery mildew damage for cucurbits plants using spectral signatures, Biosyst. Eng. 252 (2025) 144–158, https://doi.org/10.1016/j.biosystemseng.2025.03.001.
[72] R.K. Dubey, D.K. Choubey, An efficient adaptive feature selection with deep learning model-based paddy plant leaf disease classification, Multimed. Tools. Appl. 83 (2023) 22639–22661, https://doi.org/10.1007/s11042-023-16247-3.
[73] R.K. Dubey, D.K. Choubey, Adaptive feature selection with deep learning MBi-LSTM model based paddy plant leaf disease classification, Multimed. Tools. Appl. 83 (2023) 25543–25571, https://doi.org/10.1007/s11042-023-16475-7.
[74] A. Jain, R.K. Dwivedi, Hybrid segmentation-based agricultural leaf disease detection (Hy-SALDD) using black widow optimization for feature selection, and Bayesian-optimized SVM classification, Int. J. Inf. Technol. 17 (2025) 3197–3211, https://doi.org/10.1007/s41870-025-02478-0.
[75] M. Nayab, A. Waris, M. Jawad Khan, D. AlQahtani, A. Imran, S.O. Gilani, U. H Shah, Gaussian process latent variable models-ANN based method for automatic features selection and dimensionality reduction for control of EMG-driven systems, Front. Artif. Intell. 8 (2025) 1506042, https://doi.org/10.3389/frai.2025.1506042.
[76] A. Singla, A. Nehra, K. Joshi, A. Kumar, N. Tuteja, R.K. Varshney, S.S. Gill, R. Gill, Exploration of machine learning approaches for automated crop disease detection, Curr. Plant Biol. 40 (2024) 100382, https://doi.org/10.1016/j.cpb.2024.100382.
[77] Y.E. García-Vera, A. Polochè-Arango, C.A. Mendivelso-Fajardo, Gutiérrez-Bernal FJ, Hyperspectral image analysis and machine learning techniques for crop disease detection and identification: a review, Sustainability 16 (2024) 6064, https://doi.org/10.3390/su16146064.
[78] K. Berger, M. Machwitz, M. Kycko, S.C. Kefauver, S. Van Wittenberghe, M. Gerhards, J. Verrelst, C. Atzberger, C. van der Tol, A. Damm, U. Rascher, I. Herrmann, V.S. Paz, S. Fahrner, R. Pieruschka, E. Prikaziuk, M.L. Buchaillot, A. Halabuk, M. Celesti, G. Koren, E.T. Gormus, M. Rossini, M. Foerster, B. Siegmann, A. Abdelbaki, G. Tagliabue, T. Hank, R. Darvishzadeh, H. Aasen, M. Garcia, I. Pôças, S. Bandopadhyay, M. Sulis, E. Tomelleri, O. Rozenstein, L. Filchev, G. Stancile, M. Schlerf, Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: a review, Remote Sens Env. 280 (2022) 113198, https://doi.org/10.1016/j.rse.2022.113198.
[79] U. Ahmad, A. Nasirahmadi, O. Hensel, S. Marino, Technology and data fusion methods to enhance site-specific crop monitoring, Agronomy 12 (2022) 555, https://doi.org/10.3390/agronomy12030555.
[80] J. Sun, S. Shi, W. Gong, J. Yang, L. Du, S. Song, B. Chen, Z. Zhang, Evaluation of hyperspectral LiDAR for monitoring rice leaf nitrogen by comparison with multispectral LiDAR and passive spectrometer, Sci. Rep. 7 (2017) 40362, https://doi.org/10.1038/srep40362.
[81] N. Lu, J. Zhou, Z. Han, D. Li, Q. Cao, X. Yao, Y. Tian, Y. Zhu, W. Cao, T. Cheng, Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system, Plant Methods 15 (2019) 17, https://doi.org/10.1186/s13007-019-0402-3.
[82] E. Grüner, T. Astor, M. Wachendorf, Prediction of biomass and N fixation of legume–grass mixtures using sensor fusion, Front. Plant Sci. 11 (2021) 603921, https://doi.org/10.3389/fpls.2020.603921.