Title

Clustering symptomatic pixels in broomrape-infected carrots facilitates targeted evaluations of alterations in host primary plant traits

Authors

Atsmon, Guy

Pourreza, Alireza

Kamiya, Yuto

et al.

Publication Date

2024-05-01

DOI

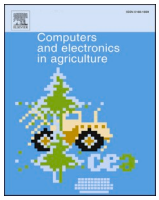
10.1016/j.compag.2024.108893

Computers and Electronics in Agriculture 220 (2024) 108893

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Clustering symptomatic pixels in broomrape-infected carrots facilitates targeted evaluations of alterations in host primary plant traits

Guy Atsmon [a,b,*], Alireza Pourreza [c,*], Yuto Kamiya [c], Mohsen B. Mesgaran [d], Fadi Kizel [e], Hanan Eizenberg [a,b], Ran Nisim Lati [a]

[a] Department of Plant Pathology and Weed Research, Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay 30095, Israel
[b] The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot 7610001, Israel
[c] College of Biological and Agricultural Engineering, University of California, Davis, United States
[d] Department of Plant Sciences, University of California, Davis, United States
[e] Department of Mapping and Geo-Information Engineering, Technion–Israel Institute of Technology, Haifa, Israel

ARTICLE INFO

Keywords:
Hyperspectral
Spectral clustering
Radiative transfer modeling
Broomrape
Stress detection
Site specific weed management

ABSTRACT

In this study, we explore spectral heterogeneity within plant canopies, a characteristic often observed in stressed plants where certain leaves or intra-leaf regions exhibit stress symptoms while others remain unaffected. Considering this variability in spectral signatures holds promise for enhancing remote sensing methodologies aimed at plant stress detection. Typically, remote sensing techniques analyze the plant as a whole, potentially overlooking stress-related spectral signatures due to the inclusion of unaffected pixels. We used a clustering-based technique, which incorporates semi-supervised learning elements for tuning hyper-parameters, to differentiate spectral patterns associated with and unique to pixels from broomrape-infected (*Orobanche* spp. and *Phelipanche* spp.) carrots from unrelated patterns. Ground-based hyperspectral (400–1000 nm) images of broomrape-infected and non-infected carrot canopies were used in an agglomerative clustering procedure followed by spectral angle mapper (SAM) analysis to identify a spectral endmember indicative of broomrape infection symptoms. Pixels from this cluster constituted an average of 8.5–11.5 % from the canopies of infected plants. Subsequently, we: (a) examined the relationship between carrot leaf mineral content and the percentage of symptomatic pixels to explore stress-induced alterations creating the unique spectral signatures of infected plants; and (b) utilized the inverse mode of PROSPECT, a radiative transfer model (RTM), to derive primary plant traits from the distinct spectral data of each cluster. We found that deficits in two macro elements, phosphorous and potassium, along with two pigments, chlorophyll and carotenoid, were correlated with the symptomatic cluster in infected plants. The methodology presented in this study paves the way for further research into broomrape detection in various crop species, as well as other plant stressors.

1. Introduction

Broomrapes (*Orobanche* spp. and *Phelipanche* spp.) are root parasitic weeds significantly hindering the cultivation of numerous vegetable and field crops worldwide. These chlorophyll-lacking holoparasitic plants attach to the host plant by forming haustoria (vascular connections) with the host's roots. Then, the parasite becomes a sink and depletes water, nutrients, and minerals from the host for its own development, including tubercle production, apex formation, and the emergence of inflorescences above the soil surface (Parker, 2012). Carrot (*Daucus carota, Apiaceae*) is a high-income cash crop widely grown throughout Europe and the Mediterranean area. Two species of broomrape, *Orobanche crenata* and *Phelipanche aegyptiaca* parasitize the carrot roots and in highly infested fields, can cause total yield loss (Bernhard et al., 1998; Cochavi et al., 2016b; Parker, 2012).

Herbicide application is the most effective method to control broomrape in carrots and other crop species. The primary limitation of existing chemical control protocols is the uniform application of

Abbreviations: SAM, spectral angle mapper; SSWM, site-specific weed management; ML, machine learning; RTM, radiative transfer models.

* Corresponding authors at: Department of Plant Pathology and Weed Research, Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay 30095, Israel (G. Atsmon); College of Biological and Agricultural Engineering, University of California, Davis, United States (A. Pourreza).
E-mail addresses: guy.atsmon@mail.huji.ac.il (G. Atsmon), apourreza@ucdavis.edu (A. Pourreza).

https://doi.org/10.1016/j.compag.2024.108893
Received 6 November 2023; Received in revised form 17 March 2024; Accepted 23 March 2024
Available online 9 April 2024

herbicides, which ignores the spatial distribution of the broomrape infestations within the fields (Cohen et al., 2017; Eizenberg and Goldwasser, 2018). Adopting site-specific weed management (SSWM) approaches becomes imperative. Nonetheless, as most of the parasite's life cycle occurs below the soil surface, detecting parasitized plants in the field through remote sensing presents a formidable challenge.

Recently, several studies investigated the potential to detect broomrape-infected plants by remote sensing of the host plants. These preliminary studies, conducted on sunflowers, demonstrated that plants parasitized with sunflower broomrape (*Orobanche cumana*) exhibited differences in spectral reflectance (Atsmon et al., 2022; Cochavi et al., 2017; Ortiz-Bustos et al., 2016) and morphological attributes (Lati et al., 2019) compared to non-infected plants. Atsmon et al., (2022) also demonstrated the feasibility to utilize these differences for classification between infected and non-infected plants. An alternative approach for broomrape detection involves employing physically-based models accounting for the mechanistic link between measured data and plant traits. Until now, models demonstrated for broomrape detection have been empirical, relying on ML algorithms or statistical models focusing exclusively on improving classification accuracy. Empirical methods often overlook variables influencing remote sensing data accuracy, including image acquisition conditions, sun-view angle geometry, and the crop's architecture and phenological stage (Jafarbiglu and Pourreza, 2023), rendering the transferability of such classification models uncertain (Wang et al., 2023). In contrast, physically-based approaches, leveraging the inversion of radiative transfer models (RTMs), have been suggested to achieve robust remote sensing of crops (Féret et al., 2019; Li et al., 2018; Shiklomanov et al., 2016). PROSPECT, for instance, is an optical model capable of simulating reflectance and transmittance at the leaf level across the electromagnetic spectrum from 400 nm to 2500 nm (Féret et al., 2021). When operated in the inverse mode, PROSPECT can retrieve primary plant traits from spectral data. Combining RTMs and ML algorithms, simulated data generated by the RTM can be used to train a statistical model, which is eventually applied to real spectral observations (Peanusaha et al., 2024).

An important insight gained by previous broomrape detection studies revealed a non-uniform spectral response within canopies of infected plants. Ortiz-Bustos et al., (2016) showed significant differences in the levels of red and far-red fluorescence reflected by infected and non-infected plants depended on the leaf pairs (e.g., 1st and 2nd vs. 3rd and 4th). Similar trend was observed by Atsmon et al., (2022) that showed the impact of the spatial location of the sampled area within the crop canopy (leaf center vs. edges) on the classification accuracy between infected and non-infected plants. Considering this variation within the canopy response to the broomrape-related stress has potential to benefit future research and practice of remote sensing of broomrape infected plants. These attempts to address spectral heterogeneity within broomrape-infected canopies required manual labeling of suspected symptomatic regions within the plant (Atsmon et al., 2022; Ortiz-Bustos et al., 2016). However, recent advancements in sensors, computing capabilities, and image processing systems have ushered in new data analysis methodologies enhancing the automation, precision, and robustness in identifying symptomatic regions in stressed plants (Jin et al., 2021; Singh et al., 2016). In a recent study, Omidi et al. (2022) suggested a semi-supervised approach to cluster symptomatic and asymptomatic leaves in root lesion nematode-infected walnut trees. An agglomerative style of hierarchical clustering based on leaf-level hyperspectral measurements was applied to identify a symptomatic cluster unique to infected plants. This technique could cluster asymptomatic leaves of infected trees with healthy leaves from non-infected trees, based on their spectral similarity, thereby isolating the symptomatic leaves to better represent the stress' spectral signature. The development of such methodologies for identification of stress symptomatic regions at the intra-plant level holds substantial promise for various aspects related to remote sensing of plant stress, including broomrape-infected plants.

In this study, agglomerative style of hierarchical clustering was employed. Symptomatic and asymptomatic pixels were clustered from hyperspectral images of broomrape-infected and non-infected carrot plants at early, pre-symptomatic stages. Subsequently, a new methodology was developed, leveraging the clustering capability for broomrape symptomatic pixels to improve the specificity of two approaches for investigating broomrape-related alterations in plant traits. In the first approach, the percentage of symptomatic pixels within the plants was correlated to measured leaf concentrations of nitrogen (N), magnesium (Mg), potassium (K) and phosphorous (P). The second approach relied on the inversion of PROSPECT to retrieve primary plant traits from spectral data derived uniquely from pixels symptomatic to the broomrape infection. Applying the model exclusively to broomrape-related spectra facilitated a more accurate linkage between foliar traits and leaf spectra.

2. Materials and methods

2.1. Plant material and experimental design

On January 1st 2021, 147 seeds of the carrot cultivar 'Nairobi' (Bejo Seeds, Oceano, CA, USA) were seeded into 3 L pots filled with infested or non-infested soil (Newe Ya'ar soil, Chromic Haploxererts, fine clay, montmorillonitic, thermic, 55 % clay, 25 % silt and 20 % sand, 2 % organic matter at pH 7.2). *Phelipance aegyptiaca* and *O. crenata* seeds were collected in 2014 from Ramat David and 2018 from Gazit, Israel, respectively. The seeds were passed through a 300-mesh sieve and stored in the dark at 4∘C. A germination test was performed under standard conditions at 25∘C, with germination stimulated by the synthetic stimulant GR-24, commonly used for broomrape germination tests. GR-24 was applied at 10 mg kg^{-1} after 12 days of preconditioning, resulting in a germination rate of >90 %. Parasite seeds were mixed into the soil using a 50 L cement mixer, creating an infestation level of 15 mg kg^{-1}. Throughout the experiment, soil temperature was monitored using temperature data-loggers (UA-001-08 data logger, Onset, Co., Bourne, MA, United States) buried in the soil at a depth of 10 cm. Below-ground parasitism dynamics were estimated using the thermal model developed by Cochavi et al., (2016a) and facilitated to perform the imaging at the specific parasitic developmental stage that is most susceptible to treatment. Plants were grown in a net-house (Newe Ya'ar Research Center, Israel) and were randomly placed to ensure similar conditions between the treatment groups (Fig. 1a). Plants were irrigated for field capacity plus 10 % in order to remove salts from the upper soil level. Fertilizer (27:10:17N-P-K) was applied at 1 g fertilizer/500 L water every 14 days starting January 18th.

Treatment groups were as follows: 32 plants of non-infested control, 56 *P. aegyptiaca* infected plants, and 59 *O. crenata* infected plants. Measurements were conducted 77 days after planting (DAP), equivalent to 974 growing degree days (GDD), according to thermal time model (Cochavi et al., 2016a). At that stage, most parasite tubercles had attached to the host roots, but only a few inflorescences emerged above the soil (Fig. 1b). At the end of the experiment, the soil was gently removed from the carrot plants to produce ground-truth infection levels by counting the attached parasite tubercles and inflorescences, and carrot leaves were cut to experimentally determine leaf mineral content (Fig. 1c).

2.2. Determination of leaf mineral concentrations

At the end of the experiment, following the spectral measurements, the whole canopy of all carrot plants was oven-dried for 48 h at 65∘C. Nitrogen levels were measured using a FLASH 2000 CHNS/O Analyzer (Thermo Fisher Scientific, Waltham, MA, United States). Phosphorous, potassium, and magnesium measurements were taken using Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES, Varian 720 ES, Agilent Technologies, Santa Clara, CA, United States). Dry plant

Fig. 1. (a) Experimental and hyperspectral imaging set up, (b) Emergence of *P. aegyptiaca* inflorescences, (c) At the end of the experiment, the soil was washed from the roots and the number of broomrape tubercles (red) and shoots (blue) were counted, (d) Masking out background pixels with NDVI and EGI indices.

material was extracted by overnight ash in a 470∘C oven, followed by nitric acid extraction. Samples were diluted with double-distilled water before measurements were taken.

2.3. Ground truth measurements of broomrape infection levels

Broomrapes that had already flowered are likely to have a more significant cumulated effect on the host than tubercles that are yet to flower. Therefore, the final values of infection levels (Fig. 2) were calculated by multiplying the number of tubercles at tubercle stages of S4 (Eizenberg et al., 2004) by three and adding the number of tubercles at tubercle stages of S1-S3, as follows:

$$A = 3B + C$$

Where A is the calculated infection level, B is the number of broomrape inflorescences, and C is the number of broomrape tubercles.

2.4. Hyperspectral data

Plants were imaged using a ground-based hyperspectral camera (Specim IQ, Specim Ltd., Oulu, Finland) that relies on a push-broom line scanner mounted on an internal rotating stage. The camera provides images of 512 × 512 pixels with 204 spectral bands between 400 and 1000 nm. The spectral resolution is 7 nm at full width half maximum (FWHM). Each image included five plants, in a random order, allowing a short image acquisition timeframe (11:00 A.M – 12:00 P.M), thereby reducing environmental variables during data collection. In addition, a white reference panel was placed in each frame to calibrate the raw data and account for changes in lighting conditions. The camera's built-in calibration procedure produced calibrated reflectance values during the image acquisition steps (Behmann et al., 2018). The camera was mounted on a tripod, and the distance between the camera lens and the top end of the plant canopy was 0.5–1 m (Fig. 1a). In this range, the spatial resolution of one pixel is about 1 mm. Imaging was performed outside with the sun as the light source for illumination, and the scanning integration time was set to a minimum of 1 ms per line. 147 hyperspectral images were taken of the whole canopy of the carrot plant. Normalized difference vegetation index (NDVI) and excess green index (EGI) were used to mask out background pixels (Fig. 1d).

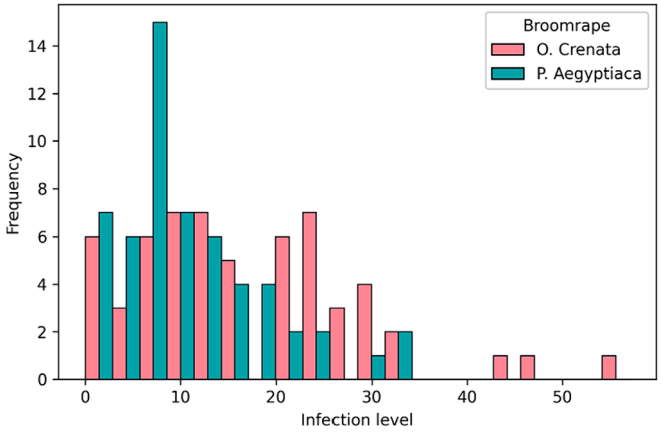


Fig. 2. Histogram of broomrape infection levels.

2.5. Sample clustering and SAM analysis

In our approach, we utilized the agglomerative style of hierarchical clustering as suggested by Omidi et al., (2022) for this type of dataset and objectives. This method of clustering is superior in handling complex and high-dimensional hyperspectral data, supports any pairwise distance metric and performs well on common clustering test cases (El-Hamdouchi and Willet, 1989). While the core algorithm, agglomerative hierarchical clustering, operates unsupervised, semi-supervised learning elements were incorporated in the hyper-parameter optimization phase. This was facilitated primarily through fine-tuning the distance threshold, which was informed by prior knowledge about the characteristic cluster of infected samples.

Complete linkage, which uses the maximum distances between all observations of two sets, was chosen to prevent intermediate samples between dense clusters from fusing those dissimilar dense clusters. Testing with other linkage methods, such as single linkage, produced poor clustering relative to complete linkage and were not used further. Cosine, which measures the angular similarity between two vectors, was set as the affinity type. Cosine was chosen as it performs well and commonly used with high-dimensional data, such as hyperspectral matrices.

Employing agglomerative clustering on all pixels from 147 hyperspectral images suffers from intrinsic computational complexity as the algorithm relies on a distance metric that grows quadraticaly in size as the number of samples increases. Therefore, a two-step process was developed for making the analyses feasible. Clustering was initially performed per each plant individually, resulting with a significant decrease in representative spectra. A fixed distance threshold of 0.002 % was used as it was found to reduce the number of samples in each image to between 20 and 250, thus still preserving the spectral heterogeneity of the image (Fig. 3a and Fig. 3b). In the second step, agglomerative clustering was performed on all retrieved spectra from the first step (n = 23,519). The prior knowledge about which plants belonged to the control and infected treatments facilitated to enhance the specificity of the model to the unique aspects of our research. A candidate cluster for representing broomrape stress contains symptomatic pixels which come exclusively from infected plants and not from control plants. Ideally, the candidate cluster would include samples from several different infected plants, i.e., a spectral pattern common among different broomrape infected plants. The distance threshold for clustering was tuned to

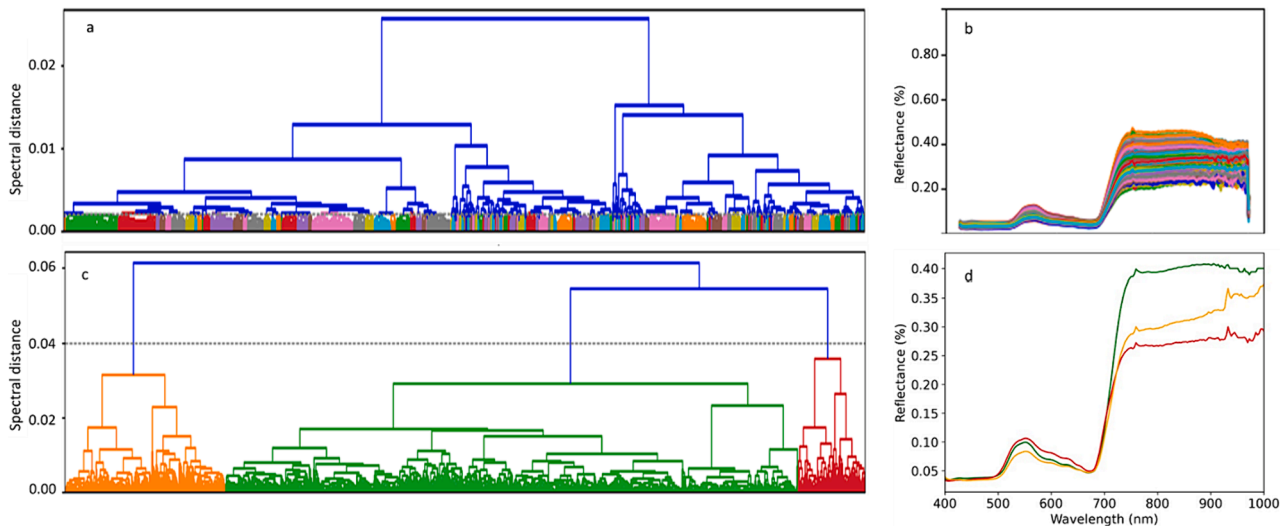


Fig. 3. (a) Agglomerative hierarchical clustering was performed per plant - Example of a broomrape infected plant, originally containing 10,830 pixel spectra, resulting in 220 sub-clusters when a distance threshold of 0.002% was applied. (b) Heterogeneity in a carrot canopy – average reflectance of the obtained sub-clusters from one broomrape infected plant. (c) Agglomerative hierarchical clustering performed on 23,519 representative sub-clusters obtained from all plants resulted in three clusters when a distance threshold of 0.04% was applied. (d) Reflectance patterns of the obtained clusters.

produce the largest candidate cluster of broomrape-infected samples. Increasing the complete linkage distance threshold parameter resulted in fewer clusters. It was finally set to 0.04 %, resulting in three spectral clusters, amongst them the broomrape symptomatic cluster which was largest of all candidates (Fig. 3c and Fig. 3d).

Finally, SAM was used to quantify the spectral similarity between each pixel spectrum and the reference spectra of the three obtained clusters. Pixels were assigned to clusters based on the highest cluster similarities (Fig. 4). SAM was chosen as the distance metric since it is sensitive to the shape of the spectral reflectance curve regardless of its magnitude, and it proved suitable in other hyperspectral studies (Kizel et al., 2017; Omidi et al., 2022; Zhang and Li, 2014).

2.6. Cluster occupancies

The percentage of pixels from each of the obtained three clusters within each plant's canopy was determined and referred to as the cluster occupancy.

2.7. Radiative transfer modeling (RTM)

RTMs are used to investigate and retrieve various vegetation characteristics (Féret et al., 2021). Here, a two-step process using the PROSPECT model was utilized. Initially, PROSPECT, a common RTM that simulates the reflectance and transmittance of different plants at the leaf level, was run in forward mode to generate a comprehensive look-up table (LUT), simulating leaf reflectance and transmittance across a range of plant traits (such as chlorophyll, carotenoids, and anthocyanin). This

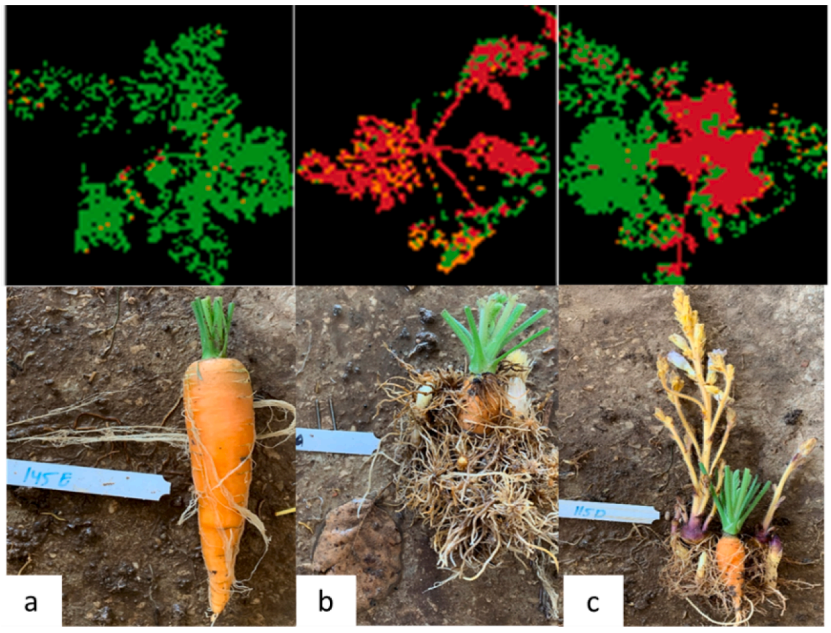


Fig. 4. Examples of soil-washed (a) control, (b) *O. crenata* infected, and (c) *P. aegyptiaca* infected carrots with their respective canopies, represented by the different clusters (top).

LUT served as a basis for the inverse mode operation, where we matched the observed spectral data from our study against this table. The mathematical essence of this process involves minimizing the difference between the observed reflectance and the simulated reflectance in the LUT to infer the corresponding plant trait values. N-structure (a leaf structure parameter) was not retrieved because our data is not based on pure proximity measurements. The input data for the model were the spectra of all sub-clusters (n = 23,519) obtained in the first clustering step which was performed on each hyperspectral image separately (as described in section 2.5.). Plotting the retrieved trait levels according to the final three clusters obtained after the second step of the clustering was facilitated by linking each sub-cluster to the final cluster in which it consisted. PROSPECT was analyzed using the package 'pracma' in R 3.2.

3. Results

3.1. Clusters identification

A candidate cluster representing broomrape stress should predominantly consist of symptomatic pixels from infected plants and not control plants. This prior knowledge regarding which plants were in the control and infected groups incorporates semi-supervised elements into the clustering technique. Ideally, the candidate cluster should include samples from several different infected plants. The first step towards identifying the candidate broomrape cluster entailed examining the canopies of all plants, represented by the obtained clusters corresponding to each pixel following SAM (Fig. 4). Upon analysis, it was observed that the red cluster was predominantly associated with *O. crenata* and *P. aegyptiaca* infected plants, in contrast to the control plants which were mainly represented by the 'green' cluster. This observation led to the conclusion that the red cluster represents pixels that are spectrally symptomatic and exclusive to the broomrape infection, while the green cluster represents 'healthy' pixels unrelated to infection. Consequently, the green cluster pixels in infected plants were identified as asymptomatic pixels, demonstrating the spectral variation of carrot leaves in response to infection.

3.2. Pixel cluster occupancy

The second step entailed analyzing the pixel occupancies of the three clusters for each plant's canopy concerning infection level (Fig. 5). This analysis clearly illustrated that the canopies of all control plants were primarily composed of pixels from the green, asymptomatic cluster, with an average occupancy of 94.6 % (Table. 1). However, it was also noted that the green cluster significantly represented *P. aegyptiaca* and *O. crenata* infected plants as well, with occupancies of 83.3 % and 78.5 %, respectively, underscoring the relevance of acknowledging the asymptomatic pixels within infected plants.

In contrast, the red pixels, indicative of symptomatic regions, had an average occupancy 0.7 % (maximum 4.1 %) within control plants, compared to 8.5 % (maximum 53.6 %) and 11.5 % (maximum 84.2 %) in *P. aegyptiaca* and *O. crenata* infected plants, respectively. This result further validates the observation that the red cluster, which was labeled symptomatic, is associated with infected plants. In general, the occupancy of the symptomatic cluster was found to correspond with infection level. Higher infection levels in both *P. aegyptiaca* and *O. crenata* infected plants, were associated with an increase in symptomatic pixels and a decrease in asymptomatic pixel occupancy. Interestingly, an increase in the symptomatic pixels only begins at a certain level of infection. For instance, red cluster occupancy above 5 % (Fig. 5 – dashed line) was observed only in plants with infection levels of five or more for both broomrape species.

The third cluster, colored in orange, accounted for an average of 4.5 % in control plants and 8.1 % and 9.9 % in *P. aegyptiaca* and *O. crenata* infected plants, respectively. However, given its lack of correlation with infection levels (Table. 1), it is hypothesized that the orange cluster represents mixed pixels, encapsulating spectral patterns of both symptomatic and asymptomatic pixels.

The average occupancies of all clusters was significantly different between infected plants (*P. aegyptiaca* and *O. crenata*) and control plants as validated by Tukey HSD at the 0.05 level.

3.3. Cluster reflectance spectra

Analyzing the reflectance patterns revealed distinct trends between the symptomatic and asymptomatic clusters. Specifically, the mean leaf reflectance from the symptomatic cluster exhibited higher green and red reflectance (~500–600 nm) and lower far-red and near infra-red reflectance (NIR ~ 750–100 nm) reflectance compared to the mean leaf reflectance of the asymptomatic cluster (Fig. 3d). High reflectance within the visible region (~500–600 nm) was previously associated with reduced foliar pigments, such as chlorophyll, carotenoids, and anthocyanin (Hennessy et al., 2020). A similar chlorophyll-reflectance relationship in broomrape-infected sunflower plants was also previously demonstrated (Ortiz-Bustos et al., 2016). Furthermore, lower NIR reflectance is often associated with stressed plants (Gill et al., 2022) and has also been observed in broomrape-infected sunflower plants, where it was correlated to an increase in the amount of air spaces within the leaf mesophyll of infected plants (Cochavi et al., 2017). Thus, the spectral patterns observed within the broomrape symptomatic and

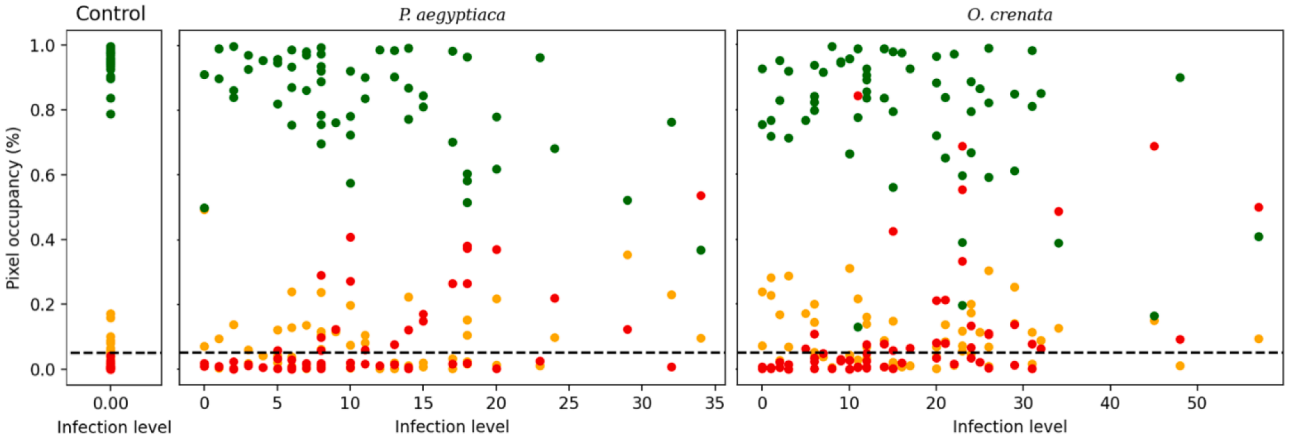


Fig. 5. Pixel cluster occupancies – each plant is represented by three dots, one from each cluster, which sum to a total occupancy of 1 (100%). (a) Canopies of control plants consisted mainly of pixels from the green cluster, (b) Increased infection levels of *P. aegyptiaca* and (c) *O. crenata* infected plants resulted in an increase in red cluster pixels and a decrease in green cluster pixels. Dot colors correspond to cluster colors. Dashed line at 0.05 occupancy.

Table 1
Cluster occupancies within the carrot canopies under different treatments.

	Symptomaic pixels			Asymptomatic pixels			Unknown pixels		
	Control	*P. aegyptiaca*	*O. crenata*	Control	*P. aegyptiaca*	*O. crenata*	Control	*P. aegyptiaca*	*O. crenata*
Average occupancy	0.007	0.085	0.115	0.946	0.833	0.785	0.045	0.081	0.099
Minimum occupancy	0	0	0	0.787	0.367	0.13	0.003	0	0.001
Maximum occupancy	0.041	0.536	0.842	0.995	0.998	0.994	0.17	0.49	0.31

asymptomatic clusters align with findings documented in the literature, corroborating the spectral indicators of plant stress and infection.

3.4. A stress-related decrease in the levels of macro elements leads to higher occupancies of symptomatic pixels

Identifying the symptomatic cluster unique to infected plants enabled an investigation into whether changes in mineral levels in the leaves of infected plants are associated with the representation of this cluster in infected plants. Generally, increased infection levels resulted in decreased mineral levels within the leaves of infected plants (Fig. 6a-d). However, plotting the symptomatic pixel occupancies in relation to the levels of these minerals in the carrot leaves revealed that only decreases in phosphorous and potassium concentrations corresponded with higher pixel occupancies of the symptomatic cluster in both broomrape species (Fig. 6f-g). The increase in symptomatic pixels (orange and blue dots) began in plants when mineral levels fell below 0.25 % phosphorous and 3 % potassium, which also represent the lower range of these minerals levels in control plants (green dots). These results indicate that deficiencies in phosphorous (R = -0.58, R^2 = 0.33) and potassium (R = -0.43, R^2 = 0.19) in infected plants are responsible for at least part of the distinctive spectral patterns of the symptomatic cluster that characterize canopies of broomrape-infected plants. In contrast, existing but less significant decreases in nitrogen (R = −0.23, R^2 = 0.05) and magnesium (R = −0.14, R^2 = 0.02) did not appear to influence the occupancy of the symptomatic cluster (Fig. 6e and Fig. 6h).

3.5. Utilizing PROSPECT for prediction of chlorophyll, carotenoid, and anthocyanin levels in the broomrape symptomatic pixels

The identification of symptomatic and asymptomatic clusters, coupled with the application of the inverse mode of PROSPECT, provided a platform to estimate the levels of chlorophyll, carotenoid and anthocyanin specific to each cluster. Fig. 7 illustrates the levels of these plant traits across all sub-clusters, grouped by the final three clusters. The broomrape symptomatic cluster exhibited the lowest levels of chlorophyll and carotenoids, while anthocyanin levels were comparable across all clusters. The average level of chlorophyll in the symptomatic cluster was 31.6 (ug/cm2), significantly lower than the 71.1(ug/cm2) observed in the asymptomatic cluster (44 % of the asymptomatic level). Similarly, the average carotenoid levels were 10.4 and 19.6 (ug/cm2) for symptomatic and asymptomatic clusters, respectively, with the symptomatic cluster having only 53 % of the carotenoid level seen in the asymptomatic cluster.

4. Discussion

4.1. Clustering

In this study, a clustering methodology which incorporates semi-supervised learning elements, was utilized for the first time on broomrape infected crops. This methodology harnesses prior knowledge about the characteristic cluster of infected samples, for fine-tuning the hyper-parameter optimization of the agglomerative style of hierarchical clustering algorithm. Our results demonstrated the benefits of utilizing the

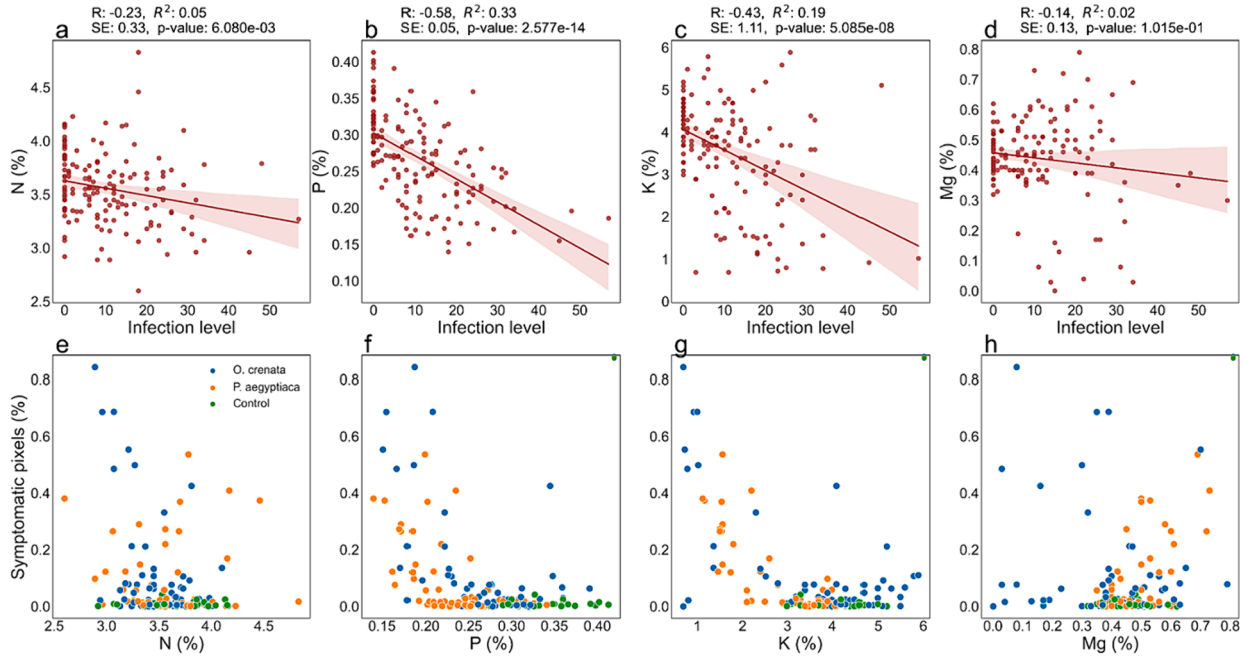


Fig. 6. (a-d) Mineral levels in relation to infection level, and (e-f) Occupancy of symptomatic pixels in relation to mineral levels – broomrape-related decreases in the concentrations of phosphorous and potassium led to higher pixel occupancies of the symptomatic cluster in the canopies of *O. crenata* and *P. aegyptiaca* infected carrots.

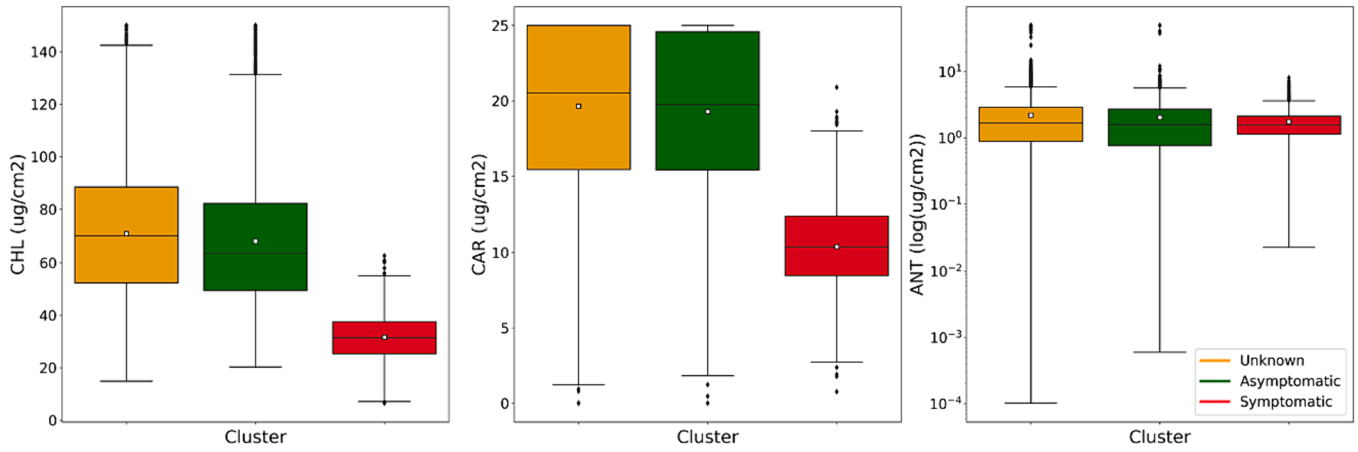


Fig. 7. Inverse mode of PROSPECT – lower levels of chlorophyll and carotenoids were observed in the broomrape symptomatic cluster compared to other clusters. (N = 1688, 17,429 and 4405 sub-clusters for the symptomatic, asymptomatic and 'unknown' clusters, respectively).

clustering methodology on image data, in contrast to previous use (Omidi et al., 2022) which was based on proximal leaf scans resulting in leaf-level resolution. The technique proved effective in discerning spectral variabilities within the canopies of broomrape-infected carrots. Clustering was successfully executed on images captured during the early, pre-symptomatic developmental stage of the parasite when the broomrape resided beneath the soil surface. The utilization of image data allowed for pixel-wise analysis and for exploration of possible matches between cluster spectra and broomrape infection while utilizing the number of pixels per cluster in each canopy as a feature vector.

A spectral cluster, distinct to pixels of infected plants, was identified. The pixels within this cluster were determined symptomatic of broomrape infection and were shown to encompass only 8.5 % and 11.5 % of *P. aegyptiaca* and *O. crenata* infected plants, respectively. From an agronomic standpoint, interestingly, the symptomatic pixels in both *P. aegyptiaca* and *O. crenata* infected plants exhibited identical spectral patterns, clustering together into one symptomatic cluster. This suggests the robustness of this methodology for future broomrape detection tasks, regardless of the broomrape species involved. Another important finding was the increasing occupancies of symptomatic pixels at infection levels of five attachments per plant. This implies that in practice, plants with low infection levels might fall outside the detection threshold.

4.2. Utilizing the symptomatic cluster for evaluation of broomrape-related alterations within host plant traits

In this study, we demonstrate two downstream approaches that leverage the clustering of symptomatic pixels for targeted investigation of broomrape-related alterations in host plant traits.

4.2.1. Correlating occupancies of symptomatic pixels to broomrape to leaf mineral concentrations

This study unveiled a decrease in the levels of measured N, P, K, and Mg as the broomrape infection escalated. However, only P and K deficiencies correlated with increased occupancies of the symptomatic cluster. By analyzing cluster occupancies concerning trait levels, the clustering method assisted in pinpointing which host plant traits were not only affected by the infection but also contributed to the unique spectral fingerprint of infected plants. Thus, compared to common approaches for correlating plants spectra and traits, this approach is less prone to obscured results which might occur due the inclusion of asymptomatic pixels together with the symptomatic pixels. The observed deficiencies align with the findings of Cochavi et al., (2017), who demonstrated that sunflower broomrape parasitism in sunflower resulted in decreased levels of several macro elements, including P and K, which are crucial for plant growth and metabolism.

4.2.2. Cluster level RTM

RTMs harbor substantial promise for enhancing the robustness and scalability of remote sensing of crops through prediction of the end-member's spectral signature (Féret et al., 2021). To the best of our knowledge, this study introduces for the first time a methodology that leverages the identification of distinct spectral clusters for a targeted deployment of an RTM specific to each cluster. The inverse mode of PROSPECT was performed on all sub-spectra linked to the three obtained clusters, focusing the estimation of plant traits solely on the symptomatic pixels exclusive to broomrape-infected plants. This trait evaluation zeroes in on the stress induced by broomrape, yielding more precise insights into the desired symptoms while minimizing interference from the predominantly asymptomatic pixels in the canopy of infected plants.

The results revealed markedly lower levels of total chlorophyll and carotenoids per area (ug/cm^2) in the symptomatic leaf regions of broomrape-infected carrots compared to asymptomatic regions. This reduction in carotenoid and chlorophyll levels aligns with previous studies. A study in *P. aegyptiaca* infected carrots found a significant 24 % decrease in carotenoid levels and linked the decrease to hampered carrot development, given the crucial role of carotenoids in plant growth and bioactive compound biosynthesis. The authors suggested that *P. aegyptiaca* may alter the carotenoids obtained from the host and utilize them for synthesizing additional carotenoid compounds (Emran et al., 2020). Other studies highlighted chlorophyll level reduction in broomrape-infected plants, attributing the damage to *P. ramosa* infected tomatoes (Mauromicale et al., 2008) and *O. crenata* infected faba beans (El-okkiah et al., 2015) to the parasite-induced decrease in leaf chlorophyll concentration, impacting the host's photosynthetic machinery.

4.3. Future perspectives and limitations

RTMs have been suggested to achieve robust remote sensing of crops by simulating reflectance and transmittance of plants under varying conditions and stressors (Féret et al., 2019; Li et al., 2018; Shiklomanov et al., 2016). Our approach, which integrates clustering prior to inverse RTM application, unveils significant potential to benefit more accurate RTM estimations of plant traits, specific to broomrape infection. In a future study, the simulated data from RTM can be utilized in aiding the training of machine learning models for real spectral observation analysis for broomrape detection (Peanusaha et al., 2024). It is worth exploring how such workflow could improve the preformance of existing broomrape detection methods of (Atsmon et al., 2022; Cochavi et al., 2017; Ortiz-Bustos et al., 2016). Recently, Atsmon et al., (2022) utilized

ground-based hyperspectral images of infected and non-infected sunflower plants to train a logistic regression classifier and achieved 88 % classification accuracies. However, the robustness and transferability of such empirical methods, relying on small, high-resolution datasets, to varying environmental conditions and lower-resolution data, is uncertain. The prospect of the hybrid retrieval workflows (RTM coupled with machine learning models) have been implemented in several operational processing chains, mainly in combination with neural networks and is very promising for vegetation property mapping, carrying advantages that include robustness to variation in environments, cultivars and image acquisition conditions (Berger et al., 2020; Verger et al., 2011).Further, RTM-based detection requires significantly less spectral while in a recent study it was also shown to retain an acceptable and consistent accuracy compared to data-driven approaches (Peanusaha et al., 2024).

In addition to precise weed management, the method used in this study to isolate the symptomatic regions within plant canopies under diverse stressors or diseases can significantly contribute to a broad spectrum of precision agriculture fields. For instance, in high-throughput phenotyping for resistant genotypes, a workflow akin to that proposed by (Omidi et al., 2022) can be adopted. Clustering applied to a dataset consisting of images of different cultivars of a given crop (e.g., carrot), once a symptomatic cluster is identified, allows for the assessment of symptomatic pixel occupancies in each cultivar to rank its resistance/tolerance to broomrape infection compared to a manual phenotyping procedure of the germplasm. The methodology could also enhance deep-learning-based disease/infection-severity evaluations, where it can be used to isolate symptomatic regions prior to analysis, as these regions encapsulate the most pertinent information (Barbedo, 2018). Further, it may facilitate the detection of subtle plant health and stress level changes that might otherwise be obscured when examining the entire plant, thereby augmenting sensitivity to early stress indicators.

The methodology presented in this study is challenged by an inherent computational limitation of the clustering method and by an agronomic limitation, which is specific to our case-study. Applying the clustering approach on image data suffers from a computational complexity. To overcome it, a two-step clustering protocol was suggested. Further, in the case of large datasets, the clustering can be based on a proportion of the sample population, while the downstream, less complex analysis, (i.e., SAM) is applied to all samples. From an agronomic standpoint, this study found that plants with low infection levels consisted low levels of symptomatic pixels, possibly bellow detection threshold of future broomrape detection models. While such potential misdetection should be considered, it is unlikely to significantly impact the final yield, as damage at such infection levels is typically minimal (Barker et al., 1996). Further, this study performed clustering on images captured during early parasitism stages, a crucial period for treatment intervention. Extending this methodology to later stages, where the parasitic impact on host plants is more pronounced, might unveil symptoms even in plants with low infection levels. While herbicidal treatment at these later stages may not be efficient for the current season's yield, it can eradicate the broomrape seed production and contribute to lower infestation levels in the subsequent seasons. Moreover, late detection still furnishes invaluable infestation maps, aiding farmers in making informed decisions in the years to follow.

5. Conclusion

The results of this study demonstrate the variability of pixel spectral patterns within broomrape infected carrot canopies. This variability should be considered in future precision agriculture practices and studies that aim to detect broomrape infected plants in the field. Agglomerative clustering followed by SAM successfully highlighted spectral symptoms exclusive to broomrape-infected plants, albeit without differentiating between the two broomrape species. A pioneering approach, melding clustering methodology with radiative transfer models, was developed to extract specific parasite-induced physiological alterations in host plants, which culminate in the distinct spectral fingerprint of infected plants.

The identification of symptomatic pixels data holds great potential for advancing site-specific management of broomrape-infected crops and streamlining high-throughput phenotyping of crop cultivars for tolerance/resistance to broomrape infection. The demonstrated suitability of this method to image data is relevant for upscaling purposes where aerial-scale imaging sensor data collection is anticipated, facilitating the collection of larger data sets. Nevertheless, it is necessary to investigate the effect of lower spectral and spatial resolution data, i.e., multispectral cameras mounted on unmanned airborne vehicles, on the clustering performance. The methodology presented in this study bears the potential for broader application across diverse broomrape and crop species as well as in other plant stress and disease studies.

CRediT authorship contribution statement

Guy Atsmon: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alireza Pourreza:** Writing – review & editing, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Yuto Kamiya:** Software, Methodology, Investigation, Formal analysis. **Mohsen B. Mesgaran:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Fadi Kizel:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Hanan Eizenberg:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Ran Nisim Lati:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Atsmon, G., Nehurai, O., Kizel, F., Eizenberg, H., Lati, R.N., 2022. Hyperspectral imaging facilitates early detection of Orobanche cumana below-ground parasitism on sunflower under field conditions. Comput. Electron. Agric. 196, 106881 https://doi.org/10.1016/j.compag.2022.106881.

Barbedo, J.G.A., 2018. Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng. 172, 84–91. https://doi.org/10.1016/j.biosystemseng.2018.05.013.

Barker, E.R., Press, M.C., Scholes, J.D., Quick, W.P., 1996. Interactions between the parasitic angiosperm Orobanche aegyptiaca and its tomato host: Growth and biomass allocation. New Phytol. 133, 637–642. https://doi.org/10.1111/j.1469-8137.1996.tb01932.x.

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., Bohnenkamp, D., Kuska, M., Jussila, J., Salo, H., Mahlein, A., Rascher, U., 2018. Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. Sensors (Switzerland) 18, 201–221. https://doi.org/10.3390/s18020441.

Berger, K., Verrelst, J., Féret, J.B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., Hank, T., 2020. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sens. Environ. 242, 111758 https://doi.org/10.1016/j.rse.2020.111758.

Bernhard, R.H., Jensen, J.E., Andreasen, C., 1998. Prediction of yield loss caused by Orobanche spp. in carrot and pea crops based on the soil seedbank. Weed Res. 38, 191–197. https://doi.org/10.1046/j.1365-3180.1998.00089.x.

Cochavi, A., Rubin, B., Achdari, G., Eizenberg, H., 2016a. Thermal time model for egyptian broomrape (Phelipanche aegyptiaca) parasitism dynamics in carrot (daucus carota L.): Field validation. Front. Plant Sci. 7, 1–11. https://doi.org/10.3389/fpls.2016.01807.

Cochavi, A., Rubin, B., Smirnov, E., Achdari, G., Eizenberg, H., 2016b. Factors Affecting Egyptian Broomrape (Orobanche aegyptiaca) Control in Carrot. Weed Sci. 64, 321–330. https://doi.org/10.1614/ws-d-15-00123.1.

Cochavi, A., Rapaport, T., Gendler, T., Karnieli, A., Eizenberg, H., Rachmilevitch, S., Ephrath, J.E., 2017. Recognition of orobanche cumana below-ground parasitism through physiological and hyper spectral measurements in sunflower (Helianthus annuus L.). Front. Plant Sci. 8, 1–12. https://doi.org/10.3389/fpls.2017.00909.

Cohen, Y., Roei, I., Blank, L., Goldshtein, E., Eizenberg, H., 2017. Spatial spread of the root parasitic weed phelipanche aegyptiaca in processing tomatoes by using ecoinformatics and spatial analysis. Front. Plant Sci. 8, 1–13. https://doi.org/10.3389/fpls.2017.00973.

Eizenberg, H., Colquhoun, J., Mallory-Smith, C.A., 2004. The relationship between temperature and small broomrape (Orobanche minor) parasitism in red clover (Trifolium pratense). Weed Sci. 52, 735–741. https://doi.org/10.1614/ws-03-157r.

Eizenberg, H., Goldwasser, Y., 2018. Control of egyptian broomrape in processing tomato: A summary of 20 years of research and successful implementation. Plant Dis. 102, 1477–1488. https://doi.org/10.1094/PDIS-01-18-0020-FE.

El- Hamdouchi, A., Willet, P., 1989. Comparison of Hierarchic Agglomerative Clustering Methods for Document Retrieval. Comput. J. 32, 220–227.

El-okkiah, S.A.F., Emara, H., Engineering, G., Abdellatif, K.F., 2015. Morphological, physiological, histological and biochemical characteristics of faba bean (Vicia faba L.) Infected by broomrape (Orobanche crenata). J. Agric. Res. 41, 1073–1093.

Emran, S., Nawade, B., Yahyaa, M., Abu Nassar, J., Tholl, D., Eizenberg, H., Ibdah, M., 2020. Broomrape infestation in carrot (Daucus carota): Changes in carotenoid gene expression and carotenoid accumulation in the parasitic weed Phelipanche aegyptiaca and its host. Sci. Rep. 10, 2–11. https://doi.org/10.1038/s41598-019-57298-7.

Féret, J.B., le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., Cheraiet, A., Oliveira, J.C., Ponzoni, F.J., Solanki, T., de Boissieu, F., Chave, J., Nouvellon, Y., Porcar-Castell, A., Proisy, C., Soudani, K., Gastellu-Etchegorry, J.P., Lefèvre-Fonollosa, M.J., 2019. Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. Remote Sens. Environ. 231, 110959 https://doi.org/10.1016/j.rse.2018.11.002.

Féret, J.B., Berger, K., de Boissieu, F., Malenovský, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sens. Environ. 252 https://doi.org/10.1016/j.rse.2020.112173.

Gill, T., Gill, S.K., Saini, D.K., Chopra, Y., de Koff, J.P., Sandhu, K.S., 2022. A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. Phenomics 2, 156–183. https://doi.org/10.1007/s43657-022-00048-z.

Hennessy, A., Clarke, K., Lewis, M., 2020. Hyperspectral classification of plants : a review of waveband selection generalisability. Remote Sens. 12 https://doi.org/10.3390/rs12010113.

Jafarbiglu, H., Pourreza, A., 2023. Impact of sun-view geometry on canopy spectral reflectance variability. ISPRS J. Photogramm. Remote Sens. 196, 270–286. https://doi.org/10.1016/j.isprsjprs.2022.12.002.

Jin, X., Zarco-Tejada, P.J., Schmidhalter, U., Reynolds, M.P., Hawkesford, M.J., Varshney, R.K., Yang, T., Nie, C., Li, Z., Ming, B., Xiao, Y., Xie, Y., Li, S., 2021. High-throughput estimation of crop traits: a review of ground and aerial phenotyping platforms. IEEE Geosci. Remote Sens. Mag. 9, 200–231. https://doi.org/10.1109/MGRS.2020.2998816.

Kizel, F., Shoshany, M., Netanyahu, N.S., Even-Tzur, G., Benediktsson, J.A., 2017. A stepwise analytical projected gradient descent search for hyperspectral unmixing and its code vectorization. IEEE Trans. Geosci. Remote Sens. 55, 4925–4943. https://doi.org/10.1109/TGRS.2017.2692999.

Lati, R.N., Filin, S., Elnashef, B., Eizenberg, H., 2019. 3-D image-driven morphological crop analysis : A novel method ford etection of sunflower. Sensors 19. https://doi.org/10.3390/s19071569.

Li, D., Cheng, T., Jia, M., Zhou, K., Lu, N., Yao, X., Tian, Y., Zhu, Y., Cao, W., 2018. PROCWT: Coupling PROSPECT with continuous wavelet transform to improve the retrieval of foliar chemistry from leaf bidirectional reflectance spectra. Remote Sens. Environ. 206, 1–14. https://doi.org/10.1016/j.rse.2017.12.013.

Mauromicale, G., Monaco, A.L., Longo, A.M.G., 2008. Effect of branched broomrape (Orobanche ramosa) infection on the growth and photosynthesis of tomato. Weed Sci. 56, 574–581. https://doi.org/10.1614/ws-07-147.1.

Omidi, R., Pourreza, A., Moghimi, A., Zuniga-Ramirez, G., Jafarbiglu, H., Maung, Z.T.Z., Westphal, A., 2022. A Semi-supervised approach to cluster symptomatic and asymptomatic leaves in root lesion nematode infected walnut trees. Comput. Electron. Agric. 194, 106761 https://doi.org/10.1016/j.compag.2022.106761.

Ortiz-Bustos, C.M., Pérez-Bueno, M.L., Barón, M., Molinero-Ruiz, L., 2016. Fluorescence imaging in the red and far-red region during growth of sunflower plantlets. Diagnosis of the early infection by the parasite orobanche cumana. Front. Plant Sci. 7, 1–10. https://doi.org/10.3389/fpls.2016.00884.

Parker, C., 2012. Parasitic weeds: A world challenge. Weed Sci. 60, 269–276. https://doi.org/10.1614/ws-d-11-00068.1.

Peanusaha, S., Pourreza, A., Kamiya, Y., Fidelibus, M.W., Chakraborty, M., 2024. Nitrogen retrieval in grapevine (Vitis vinifera L.) leaves by hyperspectral sensing. Remote Sens. Environ. 302, 113966 https://doi.org/10.1016/j.rse.2023.113966.

Shiklomanov, A.N., Dietze, M.C., Viskari, T., Townsend, P.A., Serbin, S.P., 2016. Quantifying the influences of spectral resolution on uncertainty in leaf trait estimates through a Bayesian approach to RTM inversion. Remote Sens. Environ. 183, 226–238. https://doi.org/10.1016/j.rse.2016.05.023.

Singh, A., Ganapathysubramanian, B., Singh, A.K., Sarkar, S., 2016. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci. 21, 110–124. https://doi.org/10.1016/j.tplants.2015.10.015.

Verger, A., Baret, F., Camacho, F., 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens. Environ. 115, 415–426. https://doi.org/10.1016/j.rse.2010.09.012.

Wang, Z., Féret, J.B., Liu, N., Sun, Z., Yang, L., Geng, S., Zhang, H., Chlus, A., Kruger, E. L., Townsend, P.A., 2023. Generality of leaf spectroscopic models for predicting key foliar functional traits across continents: A comparison between physically- and empirically-based approaches. Remote Sens. Environ. 293 https://doi.org/10.1016/j.rse.2023.113614.

Zhang, X., Li, P., 2014. Lithological mapping from hyperspectral data by improved use of spectral angle mapper. Int. J. Appl. Earth Obs. Geoinf. 31, 95–109. https://doi.org/10.1016/j.jag.2014.03.007.