Cucumber powdery mildew detection using hyperspectral data1

Authors: Fernández, Claudio I., Leblon, Brigitte, Wang, Jinfei, Haddadi, Ata, and Wang, Keri

ARTICLE

Cucumber powdery mildew detection using hyperspectral data[1]

Claudio I. Fernández, Brigitte Leblon, Jinfei Wang, Ata Haddadi, and Keri Wang

Abstract: This study aimed to understand the spectral changes induced by *Podosphaera xanthii*, the causal agent of powdery mildew, in cucumber leaves from the moment of inoculation until visible symptoms are apparent. A principal component analysis (PCA) was applied to the spectra to assess the spectral separability between healthy and infected leaves. A spectral ratio between infected and healthy leaf spectra was used to determine the best wavelengths for detecting the disease. Additionally, the spectra were used to compute two spectral variables [i.e., the red-well point (RWP) and the red-edge inflexion point (REP)]. A linear support vector machine (SVM) classifier was applied to certain spectral features to assess how well these features can separate the infected leaves from the healthy ones. The PCA showed that a good separability could be achieved from 4 days post-inoculation (DPI). The best model to fit the RWP and REP wavelengths corresponded to a linear model. The linear model had a higher adjusted R^2 for the infected leaves than for the healthy leaves. The SVM trained with five first principal components scores achieved an overall accuracy of 95% at 4 DPI (i.e., two days before the visible symptoms). With the RWP and REP parameters, the SVM accuracy increased as a function of the day of inoculation, reaching 89% and 86%, respectively, when symptoms were visible at 6 DPI. Further research must consider a higher number of samples and more temporal repetitions of the experiment.

Key words: crop disease detection, red-edge point, support vector machines, red-well point, leaf spectroscopy.

Résumé : Les auteurs voulaient savoir comment *Podosphaera xanthii*, agent responsable du blanc, modifie le spectre dans les feuilles de concombre entre l'inoculation et l'apparition des symptômes de la maladie. Ils ont donc appliqué une analyse en composantes principales (ACP) au spectre pour voir s'il était possible de différencier les feuilles saines des feuilles malades d'après leurs caractéristiques spectrales. Les longueurs d'onde permettant de déceler le mieux la maladie ont été établies en fonction du rapport spectral entre les feuilles infectées et les feuilles bien portantes. Le spectre a aussi permis de calculer deux variables : le point de vitalité dans le rouge (PVR) et le point d'inflexion en bordure du rouge (PIR). Une machine à vecteurs de support (MVS) servant à la classification linéaire a été appliquée à certaines propriétés spectrales pour déterminer si ces dernières permettent de séparer les feuilles malades des feuilles saines. L'ACP indique qu'on obtiendrait une bonne séparation à partir de la quatrième journée suivant l'inoculation. Le modèle linéaire est celui qui s'ajuste le mieux à la longueur d'onde du PVR et du PIR. Après ajustement, le facteur R^2 du modèle linéaire est plus élevé pour les feuilles infectées que pour les feuilles saines. Une fois formée avec la valeur des cinq composantes principales, la MVS parvient à un degré d'exactitude général de 95 %, quatre jours après l'inoculation, soit deux jours avant l'apparition des symptômes. Avec les paramètres PVR et PIR, l'exactitude de la MVS s'accroît en fonction de la journée d'inoculation pour atteindre respectivement 89 % et 86 % à l'apparition des symptômes, six jours après l'infection. Les recherches ultérieures devront recourir à un plus grand nombre d'échantillons et à de plus nombreuses répétitions de l'expérience dans le temps. [Traduit par la Rédaction]

Mots-clés : dépistage des maladies des plantes, point d'inflexion en bordure du rouge, machine à vecteurs de support, point de vitalité dans le rouge, spectroscopie des feuilles.

C.I. Fernández and B. Leblon. Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, NB E3B 5A3, Canada.

J. Wang. Department of Geography and Environment, University of Western Ontario, London, N6G 2V4, Canada.

A. Haddadi and K. Wang. A&L Canada Laboratories, London, N5V 3P5, Canada.

Corresponding author: Claudio I. Fernández (email: ci.fe@unb.ca).

[1]This article is part of a Special Issue entitled "Advancements in Canadian horticulture, in celebration of the 100th year of horticulture research at the University of Saskatchewan".

Introduction

The Canadian greenhouse crops are the most significant and fastest-growing segment of Canadian horticulture. The surface has increased 21% over the last five years and 48% over the last decade [Agriculture and Agri-Food Canada (AAFC) 2020]. The province of Ontario has the highest greenhouse crop production, with 69% of the total production in Canada (AAFC 2020). Such as with other crops, fungal diseases can affect greenhouse crops and be a significant limiting production factor (Khater et al. 2017). Powdery mildew due to *Podosphaera xanthii* is a fungal disease affecting greenhouse cucumber (*Cucumis sativus* L.) production. The pathogen most likely survives from season to season in the asexual state on living cucurbits, spreading by wind-blown spores. Conidia may also survive in the greenhouse for short periods, infecting new cucumber crops, particularly when the new crop overlaps or follows too soon after removing the old crop (Pérez-García et al. 2009). The pathogen could produce between 30% and 50% yield losses in cucumber greenhouse production (Hafez et al. 2020). *P. xanthii* grows haustorium to interact with the cell walls of the leaves, petioles, and stems, establishing a close connection beneath the host cells (Nishizawa et al. 2016; Eskandari and Sharifnabi 2019). The biotrophic pathogen does not kill the host cells to obtain nutrients (Spanu and Panstruga 2017); however, it induces leaf chlorophyll degradation and internal structural damage of colonized cells. Chlorophyll variations can be detected in the visible spectral domain (400–700 nm) (Mutanga and Skidmore 2007), while the internal leaf structure changes can be detected in the near-infrared spectral domain (700–1300 nm) (Knipling 1970; Delalieux et al. 2007).

Another interesting spectral domain is the red-edge region, a transition zone located between 660 and 780 nm (Horler et al. 1983). The transition zone is between the maximal chlorophyll absorbance in the red wavelength and the strong near-infrared reflectance due to leaf mesophyll scattering. This region has two variables, the red-well point (RWP) and the red-edge inflection point (REP) (Pu et al. 2003). RWP is the wavelength in the red region, between 660 and 680 nm, corresponding to the minimum reflectance due to maximum chlorophyll absorption. REP is the wavelength corresponding to the inflexion point of the spectral curve between the red and near-infrared spectral domains. At REP, the reflectance slope in the red-edge region is maximal (Darvishzadeh et al. 2009). Shifts of the REP to longer or shorter wavelengths have already been related to chemical and morphological plant status (Peng et al. 2011). Therefore, changes in the reflectance of the red-edge region and its associated parameters, RWP and REP, can be related to powdery mildew in cucumber leaves.

Previous studies on the detection of cucumber powdery mildew inside greenhouses were based on RGB images when the symptoms were already visible (Wspanialy and Moussa 2016; Zhang et al. 2017, Zhang et al. 2019). Other studies used hyperspectral data; hyperspectral imagery combined with chlorophyll fluorescence and thermograms was used by Berdugo et al. (2014) to study cucumber leaves infected with powdery mildew. Atanassova et al. (2019) examined cucumber powdery mildew at the leaf level using vegetation indices computed with data acquired with a spectrometer having a spectral range between 450 and 1100 nm. Other cucumber diseases were also detected using remotely sensed data. Cucumber downy mildew (*Pseudoperonospora cubensis*) was detected over hyperspectral imagery in the range between 400 and 1100 nm by Tian and Zhang (2012). Remote sensing was also used to detect other significant damages in cucumber. For example, Cen et al. (2016) applied hyperspectral imagery acquired between 400 and 675 nm in reflectance mode and from 675 to 1000 nm in transmittance mode to study chilling damage in cucumber fruits. Diseases affecting other cucurbit species were also studied using remote sensing. Kalischuk et al. (2019) applied unmanned aerial vehicle (UAV) multispectral imagery for mapping and scouting gummy stem blight (*Stagonosporopsis cucurbitacearum*) in watermelon (*Citrullus lanatus*) fields, and hyperspectral UAV imagery was used by Abdulridha et al. (2020) to detect powdery mildew in squash (*Cucurbita pepo*).

This paper is the first part of a research project aiming to detect powdery mildew on cucumber leaves. This paper has the specific objective to understand the changes induced in hyperspectral data by powdery mildew in cucumber leaves from the moment of inoculation until visible symptoms are apparent. A principal component analysis (PCA) was applied to the spectra to assess the spectral separability between healthy and infected leaves. A spectral ratio between infected and healthy leaf spectra was used to determine the best wavelengths for detecting the disease. Additionally, the spectra were used to compute two spectral variables (i.e., RWP and REP). A linear support vector machine (SVM) classifier was applied to the principal component scores, RWP, or REP, to assess how well these features can separate the infected leaves from the healthy ones. The second part of the study (Fernández et al. 2021) will test the use of simulated multispectral band reflectance of the Micasense RedEdge camera and associated vegetation indices to detect powdery mildew on cucumber leaves because the main objective of the study is to test the use of a Micasense RedEdge camera for detecting cucumber powdery mildew.

Materials and Methods

Experiment

The study used data acquired during a controlled experiment in December 2019 in a walk-in growth chamber located at the Biotron facilities of the University of Western Ontario (London, Ontario). The chamber had a

constant relative humidity of 70%, a photoperiod of 12 h, with an air temperature of 23 °C, followed by a dark period of 12 h with an air temperature of 20 °C. Cucumber seeds cultivar Straight Eight (William Dan Seeds Ltd) were planted individually at 3 mm depth in 0.5 L pots filled with a volume of 0.475 L of the multipurpose Pro-Mix LP15 substrate. The plants were regularly well-watered. A dose of 10 mg per plant of MiracleGro All Purpose (NPK 24-8-16, with micronutrients) commercial soluble fertilizer was applied at 100% plant emergence and then every seven days after plant emergence until the end of the experiment. Ten plants were used and were split into two groups: healthy ($n = 5$) and infected ($n = 5$). The healthy plants were placed in a separate chamber to avoid cross-contamination.

Before inoculation, the leaves of each plant were marked on the petiole base with a small white tape, which was numbered from the bottom to the top leaf. Due to the large surface of the cucumber leaves, a region of interest (ROI) of 12 cm^2 was drawn on each leaf. These ROIs allow collecting spectral data precisely on the same inoculated area over time. The infected plants were inoculated as follows. Cucumber leaves infected with *P. xanthii* were collected one day before inoculation from Great Lakes Greenhouses Inc., located in Leamington, Ontario, Canada. Samples were wrapped in a paper towel to absorb humidity, bagged into a Ziploc® bag, and transported to A&L Laboratories Inc., where they were kept in a cold chamber at 4 °C until the inoculation. They were cut into small pieces, and the infected sections were used to contact each ROI marked on the leaves of the plants from the infected group five times. The ROI of the control group was also physically contaminated five times using non-infected sections of cucumber leaves.

Spectral data

Radiances between 325 and 1075 nm at 1.5 nm sampling intervals were measured at the leaf level four hours before inoculation (0 DPI), until seven days post-inoculation (DPI), with an ASD FieldSpec® HandHeld 2 spectroradiometer (ASD Inc., Boulder, Colorado, USA) having a 25° field of view (FOV) bare fibre optic cable. No data was collected at one-day post-inoculation (1 DPI) to properly establish the pathogen and reduce air dispersion risk of non-germinated spores and contamination of the healthy plants. The spectra were acquired and calibrated using the RS^3 version 6.4 software from ASD (ASD Inc., Boulder, Colorado, USA). Leaf reflectance spectra were acquired over a total of 183 healthy and 201 infected ROI, respectively. They were acquired on the centre of each marked ROI, including the midrib region when present, with the optic fibre coupled to an ASD high-intensity probe equipped with a light source. A leaf clip, having a white Goretex (99% reflectance) and a black reference, was attached to the probe. The calibration of the probe was performed every 10 min and set to

obtain one mean reflectance value from 10 ASD spectral measurements per reading.

Spectral data processing

All the data was processed using MATLAB R2020b (MathWorks, Inc., Natick, Massachusetts, USA) following the flowchart of Fig. 1. In the analysis, we did not consider the spectra obtained from the following cases: (*i*) ROIs where there were scars due to the friction during inoculation, and (*ii*) yellow bottom leaves due to natural senescence. The resulting data set contained 71 spectra from healthy leaves and 57 spectra from infected leaves. All the spectra were then cropped to the 400–900 nm range and converted to reflectance spectra using the ViewSpec Pro version 6.2.0 software (ASD Inc., Boulder, Colorado, USA). These reflectance spectra were then subjected to a Savitzky–Golay filtering method to reduce instrumental noise (Savitzky and Golay 1964). It was followed by a multiplicative scatter correction (MSC) to remove additive and multiplicative scattering effects (Isaksson and Næs 1988).

To determine the DPI when a spectral difference between healthy and infected leaves can be observed, the mean reflectance spectra were plotted as a function of the DPI. Such as in Fernández et al. (2020*a*), they were then used to compute the following spectral ratio (SR) at each wavelength between the spectra acquired over healthy and infected leaves or plants, which was then plotted as a function of DPI.

$$(1) \quad SR_\lambda = \frac{Ri_\lambda}{Rh_\lambda}$$

where SR_λ is the spectral ratio at wavelength n, Ri_λ is the mean reflectance at wavelength n for the spectra acquired from the infected leaves, and Rh_λ is the mean reflectance at wavelength n for the spectra acquired from the healthy leaves.

The SR plots have peaks corresponding to wavelengths with the highest spectral difference between healthy and infected leaves or plants.

Given that spectral data have high correlation and multicollinearity, following several studies on crop disease detection (Wang et al. 2012; Tian and Zhang 2012; Gulhane and Kolekar 2014), the spectra were subjected to a PCA. PCA is a multivariate statistical technique for transforming data based on eigenvalue analysis. A PCA on hyperspectral data transforms the data to a new set of uncorrelated variables that are linear combinations of the original dataset; and reduces the dimensionality of the data set while preserving the variance (Kumar et al. 2014; Martinez and Cho 2015). We applied PCA over the truncated reflectance dataset from 400 to 900 nm. Then, the first 2 PCA were plotted as a function of the day post-inoculation to observe the separability of the healthy and infected spectra as an effect of the disease development (Everis et al. 2001; Sankaran et al. 2010).

↵ Published by Canadian Science Publishing

Fig. 1. General workflow, from data collection to the computation of each evaluated parameter [adapted from Fernández et al. (2020*a*, 2020*b*)].
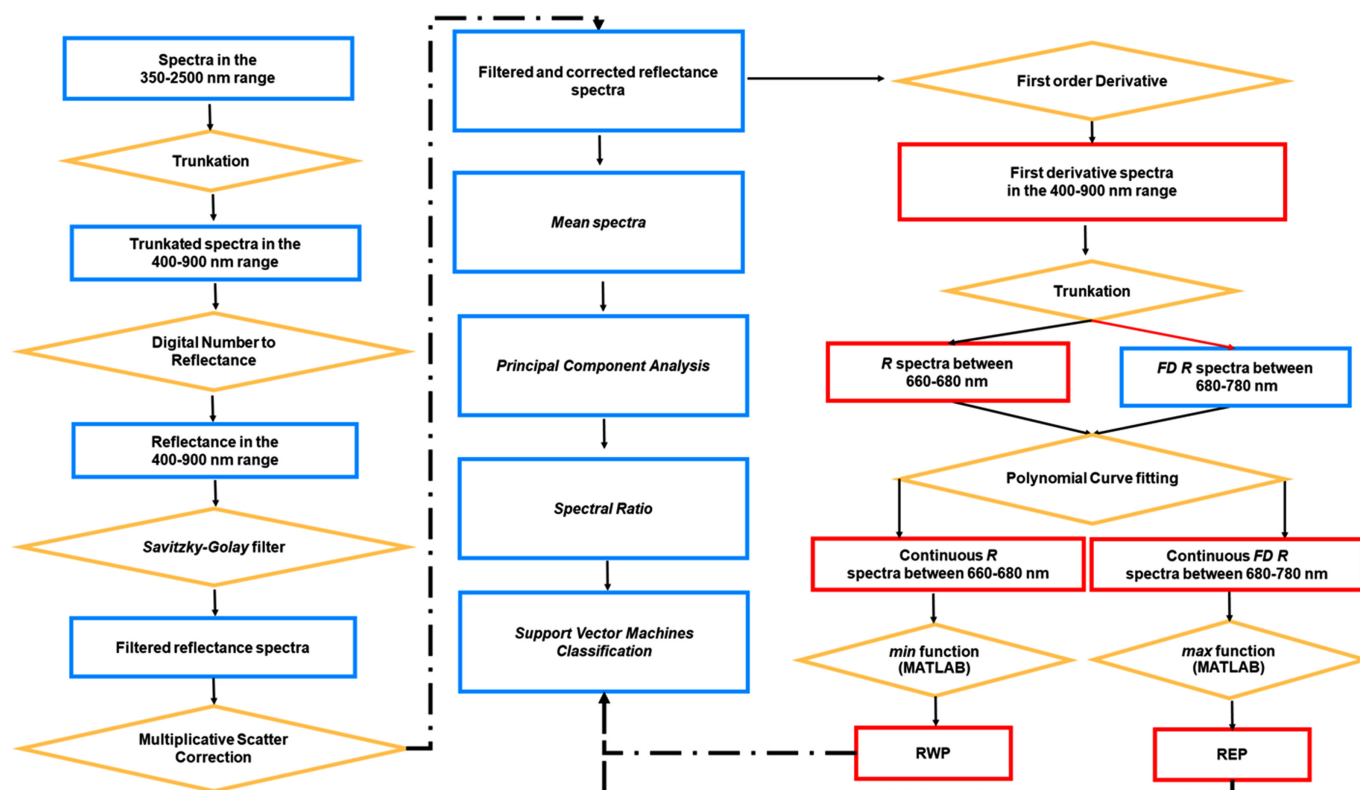

Table 1. Typical confusion matrix and related parameters.

Confusion matrix	True (infected)	False (healthy)
Predicted values	Number of true positive (TP) Number of false negative (FN)	Number of false positive (FP) Number of true negative (TN)
Sensitivity	$\dfrac{\text{TP}}{\text{TP} + \text{FN}}$	(2)
Precision	$\dfrac{\text{TP}}{\text{FP} + \text{TP}}$	(3)
F1 Score	$2 \times \left(\dfrac{\text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \right)$	(4)
Overall accuracy	$\dfrac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$	(5)

Such as in Fernández et al. (2020*b*), to compute RWP and REP, the spectra were subjected to first derivations. Both the reflectance and the first-order derivative spectra were cropped to the 660–780 nm spectral range. For each red-edge parameter (RWP and REP wavelengths), we created a unique vector containing all RWP or REP during all days of evaluation. The vectors containing the RWP or REP wavelengths of the healthy (infected) cases were 497 × 1 (399 × 1). Linear regression models were fitted to explain the RWP or REP trend as a function of the DPI using the fitlm function of MATLAB R2020b over each vector. Such modelling over each vector allows obtaining the residuals to perform normality assumption analysis properly. The one-sample Kolmogorov–Smirnov test was applied over the residuals of each fitted model using the kstest function of MATLAB R2020b. The normality assumption was followed by analyzing the variance homogeneity of residuals using the F-test (vartest2 function of MATLAB R2020b). Finally, to assess the trend of each red-edge parameter, we plotted the mean value of healthy(infected) RWP or REP as a function of the DPI.

↳ Published by Canadian Science Publishing

Fig. 2. Comparison between mean reflectance (with one standard deviation) acquired from healthy and infected leaves as a function of the number of days post-inoculation.
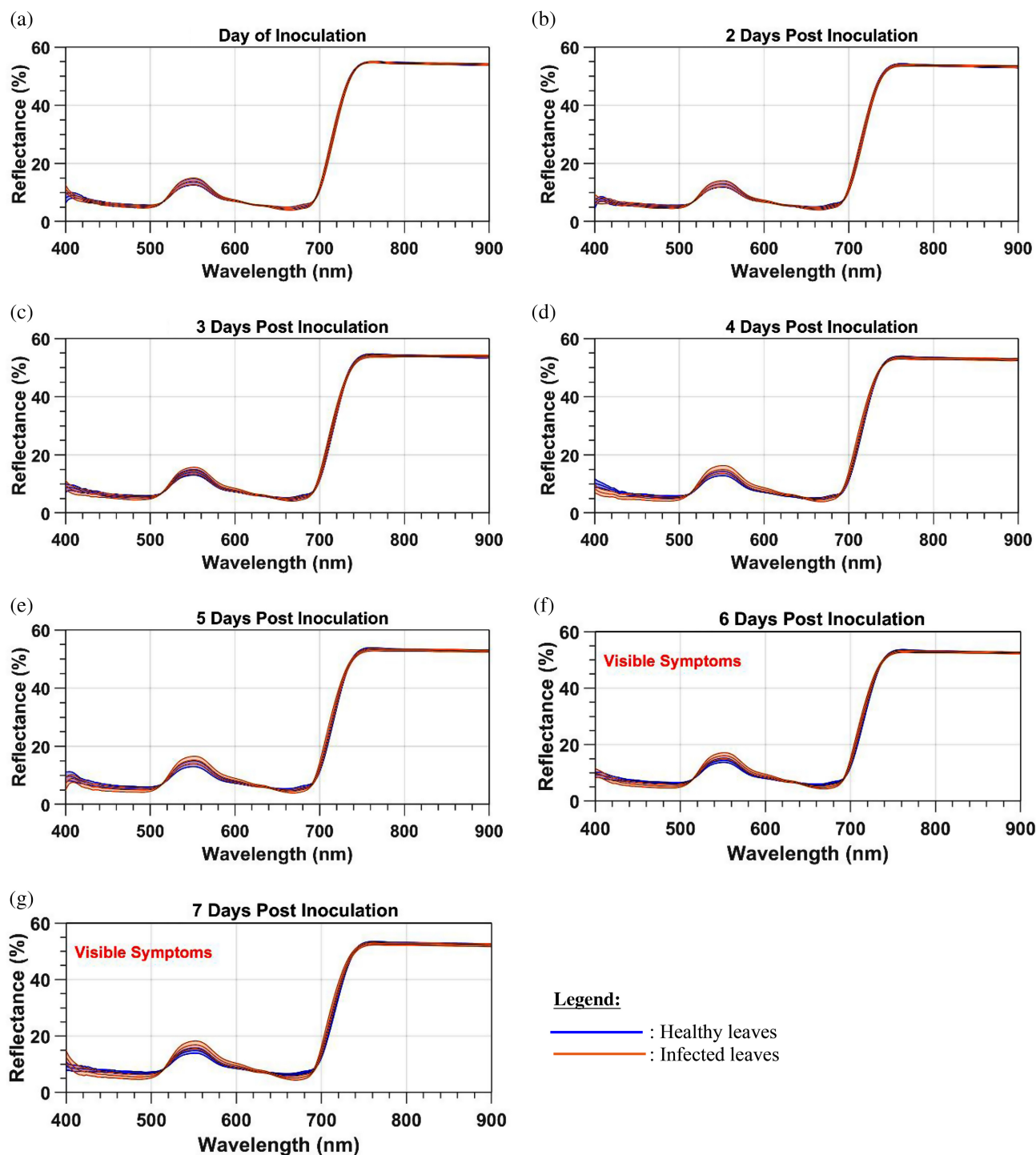
In order to test whether spectral feature allows sorting healthy and infected leaf ROI, we applied an SVM algorithm embedded in the MATLAB Statistics and Machine Learning Toolbox™. The SVM was applied to (*i*) the first five principal components, (*ii*) the RWP, (*iii*) and REP wavelengths. The SVM algorithm was calibrated using cross-validation and a 10 K-fold division to avoid overfitting. For each spectral feature, the SVM outputs a confusion matrix as a function of the DPI that was used to compute the SVM parameters sensitivity, precision, F1 score, and overall accuracy (Table 1).

↰ Published by Canadian Science Publishing

Table 2. Variance explained by first and second principal components (PC1 and PC2, respectively) as a function of the number of days post-inoculation at the leaf level.

Number of days post inoculation	PC1	PC2	PC3	PC4	PC5	Total
0	84.53	6.42	4.09	1.48	1.01	97.53
2	83.63	6.44	4.05	1.81	1.07	97.00
3	83.51	6.24	5.06	2.13	0.98	97.92
4	86.91	7.58	1.70	1.42	0.87	98.48
5	86.81	5.71	2.85	1.69	1.17	98.23
6	88.31	6.53	2.02	1.06	0.83	98.75
7	83.38	10.16	2.33	1.78	0.79	98.44

Results

Mean leaf reflectance and PCA

The mean leaf reflectance spectra for the healthy and infected leaves are presented in Fig. 2. The results indicate that across the whole spectra, between 400 and 900 nm, spectral changes between healthy and infected leaves are not observed until 3 DPI (Figs. 2a–2c). At 4 DPI, the reflectance of the infected leaves is lower than those of the healthy leaves in the 400–450 nm spectral region. Between 520 and 520 nm, the reflectance of the infected leaves is higher than those of the healthy leaves (Fig. 2d). These results are also reflected at 5 DPI (Fig. 2e). However, during the whole experiment period, 7 d, the magnitude of changes across the whole spectra (400–900 nm) is not large enough between healthy and infected leaves, especially in the near infrared (NIR) region (780–900 nm), where no visible differences are apparent (Figs. 2a–2g).

The total variance explained by the five first PCs changes with the DPI because of the powdery mildew development, but it is always near or above 97% of the total variance of the leaf reflectance spectra (Table 2). The first two PCs showed high variances and improved the visualization to discriminate between healthy and infected observations. For PC1, the explained variance dropped from 84.53% at 0 DPI until 83.51% at 3 DPI, then increased to 86.91% at 4 DPI, then had a slight reduction to 86.81% at 5 DPI, then increased and reached its maximum value of the experiment at 6 DPI (88.31%) and then dropped to 83.38% at 7 DPI (Table 2). For PC2, the explained variance showed a minor reduction from 6.42% at 0 DPI to 6.24% at 3 DPI (Table 2). Then, the variance increased to 7.58% at 4 DPI and decreased to 5.71% at 5 DPI (Table 2). It increased again to reach its maximum value (10.16%) at 7 DPI (Table 2).

Figure 3 presents the variation of the scores for the first and second principal components (PC1 and PC2, respectively) as a function of the DPI. It allows assessing the separability of the spectra of healthy and infected leaves. Both spectra are similar, and no separability can be seen on the day of inoculation (Fig. 3a) and at 2 DPI (Fig. 3b). From 3 DPI (Fig. 3c), some infected leaves have

spectra that begin to separate from the centre of the ellipse of their distribution. It is possible to observe that at 4 DPI (Fig. 3d), when spectra of most healthy leaves are closely related, the spectra of the infected leaves become distant from the spectra of healthy leaves. However, it is the opposite at 5 DPI (Fig. 3e). At 6 and 7 DPI, the spectra show greater separation (Figs. 3f and 3g). When the total explained variance of PC2 is over 6.5% at 4 DPI, the distribution of the spectra of the infected leaves is distant from the ones of the spectra of the healthy leaves.
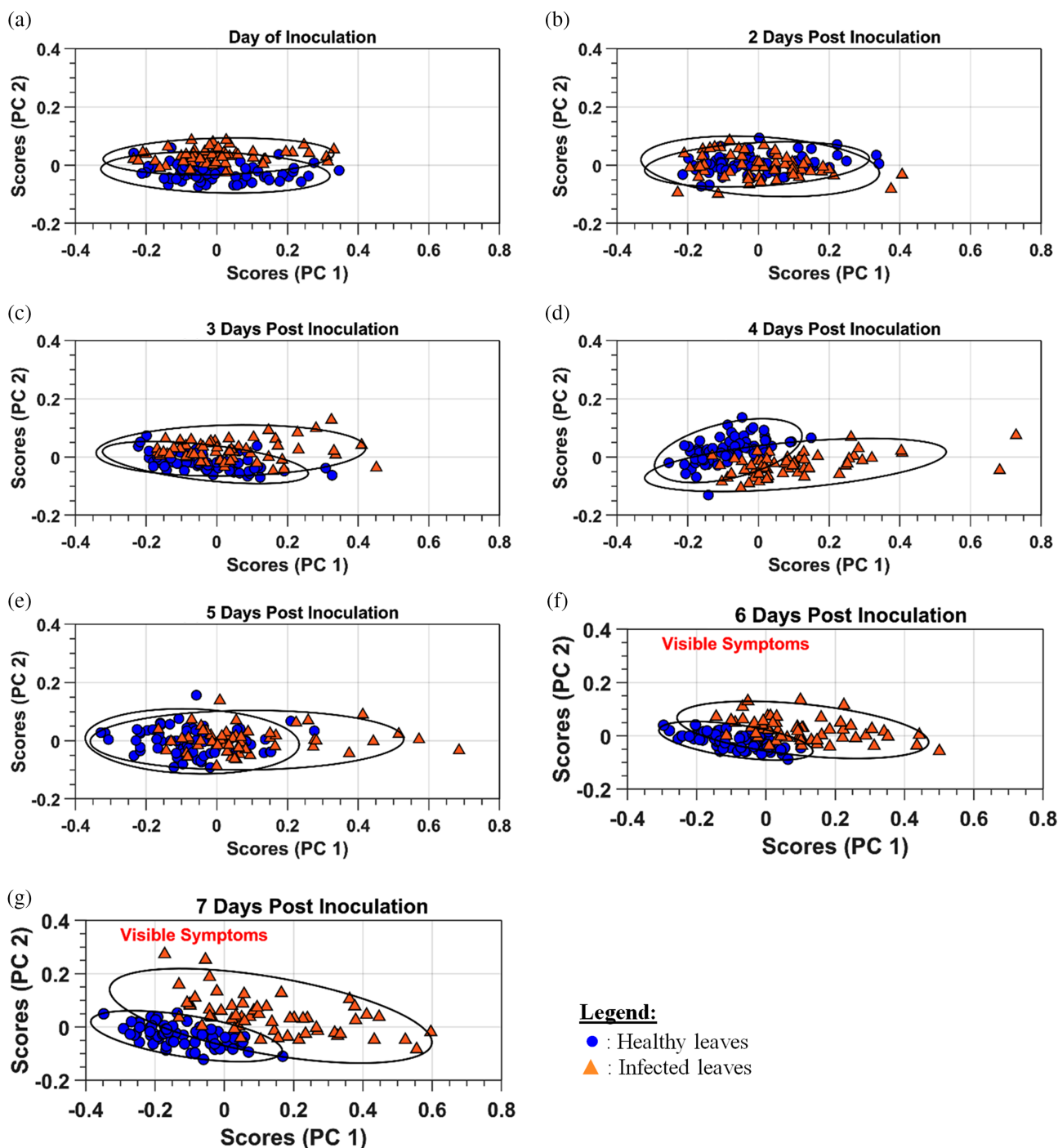
Spectral ratio

The SR between healthy and infected cases for each DPI is presented in Fig. 4. The curve peaks of these plots correspond to the wavelengths where the spectral difference between healthy and infected leaves is the largest. As expected, the spectra of the infected leaves are very similar to the spectra of the healthy leaves on the day of inoculation, leading to a spectral ratio closed to 1. The changes in the spectral ratio were observed from 2 DPI (Figs. 4b–4g). The peaks detected by the ratios are more pronounced for a longer time after inoculation and are presented mainly between 450 and 700 nm. We did not consider the spectral region between 400 and 430 nm during this evaluation because of apparent instrumental noise (Figs. 4a–4g). When symptoms were visible for the first time, at 6 DPI, peaks can be easily located at 444, 570, 667, and 703 nm (Fig. 4f and Table 3). Between 6 and 7 DPI, there was no significant variation in the peak positions at 445, 570, 665, and 704 nm (Table 3). The NIR region (800–900 nm) presented almost no variations on the computed spectral ratio.

RWP and REP

A linear regression model obtained the best fit for the variation of the mean RWP and REP wavelength positions as a function of the number of days post-inoculation (Fig. 5). In both cases, the adjusted R^2 of the linear model was not significant at $P > 0.05$ for the healthy leaves ($R^2 = 0.49$ for RWP and $R^2 = 0.18$ for REP). For the infected leaves, it was significant at $P \leq 0.01$ for RWP ($R^2 = 0.88$) and at $P \leq 0.001$ for REP ($R^2 = 0.91$).

↳ Published by Canadian Science Publishing

Fig. 3. Variation of the scores for the first and second principal components (PC1 and PC2, respectively) of healthy and infected leaves as a function of the number of days post-inoculation.

(a) Day of Inoculation
(b) 2 Days Post Inoculation
(c) 3 Days Post Inoculation
(d) 4 Days Post Inoculation
(e) 5 Days Post Inoculation
(f) 6 Days Post Inoculation — Visible Symptoms
(g) 7 Days Post Inoculation — Visible Symptoms

Legend:
● : Healthy leaves
▲ : Infected leaves

SVM

We applied a linear SVM classifier to either the first five PCs that explained more than 97% of the total variance, RWP, or REP. We obtained an inconsistent trend for the spectra classification. At 0 DPI, the overall accuracy was around 65%, then increased to 86% at 2 DPI to drop drastically to 77% at 3 DPI (Table 4). A high increment was observed from 3 to 4 DPI, reaching an overall accuracy of 95% (Table 4). Then again, the accuracy dropped by 10% to reach a value of 85% at 5 DPI (Table 4). When signs of the disease were visible at 6 DPI, the accuracy increased to 97%, with a slight reduction of 1% at 7 DPI (Table 4).

The best overall classification accuracy was obtained using RWP across all DPIs (Table 4). RWP constantly increased with the DPI and showed higher accuracy than

↘ Published by Canadian Science Publishing

Fig. 4. Variation as a function of the number of days post-inoculation of the spectral ratio computed using the mean spectrum, acquired from healthy and infected leaves, concerning the healthy leaf spectra according to eq. 1.
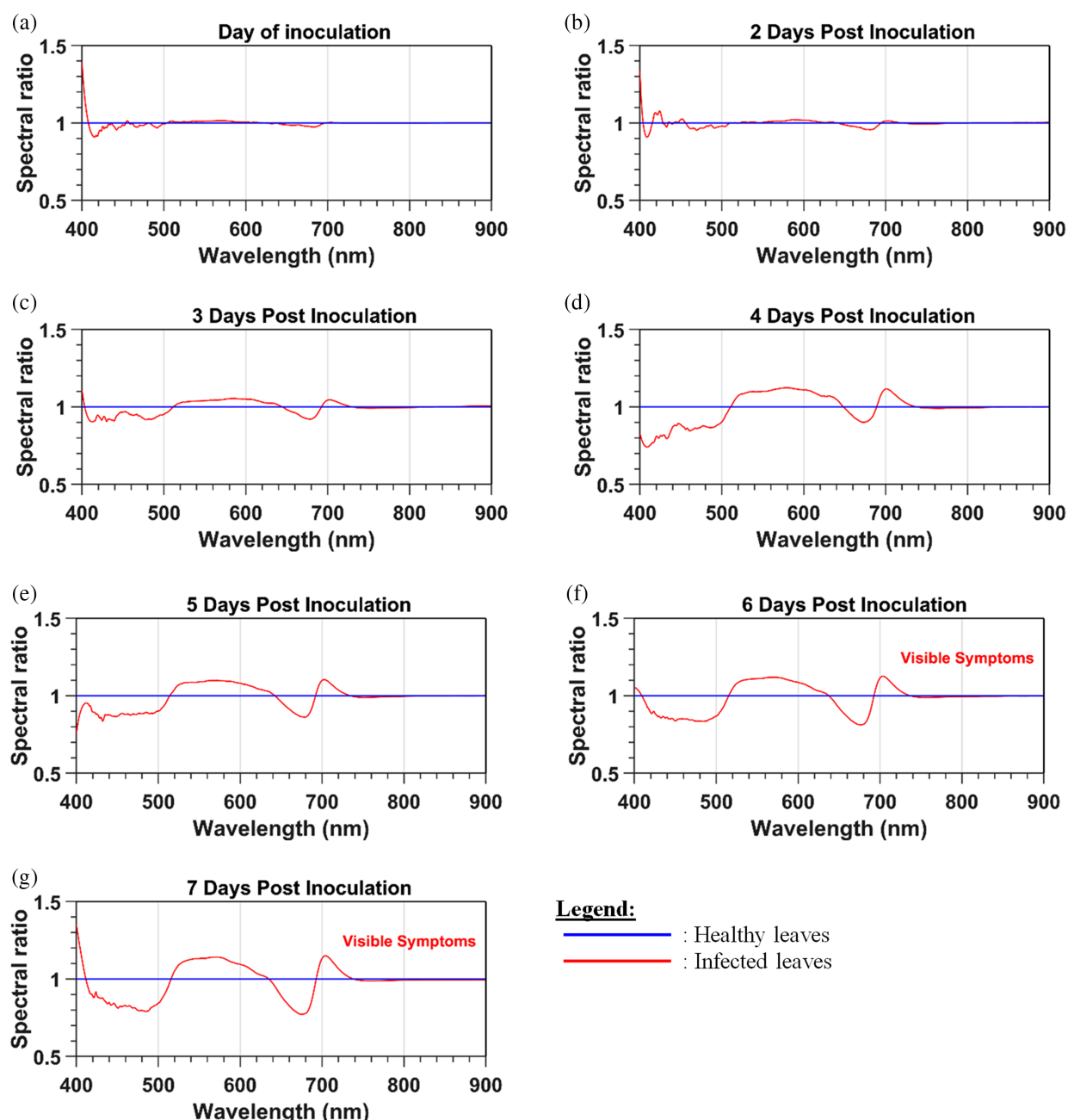
REP. The corresponding confusion matrix for the RWP classification is presented in Table 5. At 0 DPI, the overall classification using RWP was 55% (Table 4). The overall accuracy increased to 57.81% at 2 DPI, showing a higher increment to 72% at 3 DPI (Table 4). At 2 DPI, the corresponding confusion matrix showed that the number of correctly classified healthy and infected leaves was 60 (84%) and 14 (24%), respectively (Table 5). At 3 DPI, the number of adequately classified samples dropped to 59 (83%) in the case of healthy leaves and increased to 34 (59%) in the case of infected leaves

(Table 5). At 4 DPI and 5 DPI, it increased to near 78% (Table 4). At 4 DPI, the number of adequately classified healthy and infected observations was 58 (81%) and 42 (73%), respectively (Table 5). This number showed slight variations at 5 DPI to 55 (77%) and 45 (78%) for the healthy and infected leaves, respectively (Table 5). At 6 DPI, when symptoms were visible, the overall accuracy was 89%. It reaches its maximum value at 7 DPI with 94% (Table 4). The related confusion matrix at 6 DPI showed that 90% (64) of the healthy and 88% (50) of the infected leaves were classified correctly (Table 5). These numbers

Table 3. Position of the peak wavelength in the blue (431–499 nm), green (500–599 nm), red (600–699 nm), and the Red-Edge (700–799 nm) spectral regions.

DPI	Blue	Green	Red	RedEdge
0	432	565	683	704
2	430	586	681	703
3	430	586	678	702
4	441	579	672	701
5	436	571	678	703
6*	444	570	677	703
7*	445	570	675	704

Note: Data determined from the spectral ratio computed using the mean spectrum acquired over healthy and infected leaves concerning the healthy leaf spectra.

*Visible symptoms.

increased to 67 (94%) and 54 (95%) for the healthy and infected leaves at 7 DPI (Table 5).

With REP, the SVM classification presented an overall classification accuracy of around 55% at 0 DPI and 2 DPI (Table 4). At 3 DPI, the overall classification accuracy was 61% (Table 4). From 4 DPI, REP increased continuously from 71% to 86% at 6 DPI (Table 4). At 7 DPI, the REP's overall classification accuracy dropped slightly to 83% (Table 4).

Discussion

This study analyzed the changes in reflectance spectra in the visible-NIR regions induced by *P. xanthii* on cucumber leaves. In particular, we applied PCA to the spectra and computed two spectral parameters, RWP and REP. Only a few studies assessed these features for detecting biotrophic diseases such as cucumber powdery mildew with RWP, REP, or PCA score; hence we compared our results with studies on necrotic diseases, such as late blight (*Phytophthora infestans*) on potatoes.

In our experiment, the signs and symptoms of the disease were visible at 6 DPI, while Berdugo et al. (2014) observed signs at 4 DPI. Another difference with Berdugo et al. (2014) is that the evaluations were done every four days after inoculation. Our study considered daily evaluations (excepting 1 DPI to allow the pathogen establishment and reduce cross-contamination risk).

The spectral changes in the 400–900 nm were assessed using several data analysis methods. A graphical comparison of the mean reflectance spectra of the healthy and infected leaves as a function of the day of inoculation showed that the green reflectance of infected leaves has higher values than the healthy leaves. Cucumber leaves infected with *P. xanthii* showing higher reflectance values between 550 and 680 nm concerning healthy leaves were also reported by Berdugo et al. (2014). Atanassova et al. (2019) reported that the most considerable spectral differences between healthy leaves and those infected with *P. xanthii* were found between 540 and 680 nm, corresponding to the wavelengths between the green and red spectral regions. Mahlein et al. (2013) also reported higher reflectance values in the case of sugar beet (*Beta vulgaris* var. saccharifera) leaves infected with *Erysiphe betae* (powdery mildew) in the green and red-edge spectral domains. In the red-edge reflectance, we observed a slight blue spectral shift. Gitelson et al. (1996) attributed the blue shift of the red-edge reflectance to the chlorophyll loss and plant stress.

Such as in Fernández et al. (2020a) and Xie et al. (2017), we applied a PC analysis to reduce the high number of reflectance data to a new uncorrelated dataset containing 5 PC explaining near or above 97% of the total variance of the data set. The total variance explained by these 5 PCs was similar to the one reported by Fernández et al. (2020a) (96%), who studied late blight on potato leaves. The explained variance explained by the first 5 PCs was also in the same magnitude as the total variance of 98.88% observed at 5 DPI with the first 3 PC by Xie et al. (2017), who studied grey mould (*Botrytis cinerea*) on tomato leaves. We reached an overall accuracy of 95% at 4 DPI (i.e., two days before visible symptoms). Our accuracies are higher than those reported by Abdulridha et al. (2020) (82%), who applied a radial basis function to hyperspectral imagery to discriminate healthy squash leaves from those infected with *P. xanthii* before the symptoms were visible. When the symptoms were apparent (at 6 DPI), our accuracy increased to 97%. This accuracy was higher than the one of Tian and Zhang (2012) (90%). They classified a limited number of hyperspectral images (ten on healthy leaves and ten on infected leaves) acquired over healthy leaves and leaves infected with *P. cubensis* (cucumber downy mildew). However, our accuracy was lower than that of Abdulridha et al. (2020) (99%), who discriminated with a radial function healthy squash leaves from those infected with *P. xanthii* at a late development disease stage. Our accuracies were also lower than those reported by Cen et al. (2016) (100%), who applied an SVM algorithm to discriminate chilling injured and non-damaged cucumber fruits. We hypothesize that these changes in the classification accuracy when using PCA might be related to some noise in the 400–440 nm region. We can observe that the spectral ratio dropped drastically from 3 to 4 DPI (Figs. 4c and d) in this region and then increased at 5 DPI (Fig. 4e). Those changes could be attributable to instrumental noise. Therefore, the computation of a PCA over the whole visible spectra (400–900 nm) might be affected by instrumental noise in some spectral regions, i.e., below 450 nm in our case. Another possible explanation could be changing chlorophyll-a concentration due to the pathogen infestation, affecting the chlorophyll a/b ratio (Tanaka and Tanaka 2011). Finally, the low number of observations might significantly influence the overall accuracies even when few samples are misclassified.

↵ Published by Canadian Science Publishing

Fig. 5. Linear regression model fitted with the mean red-well point (RWP) and mean red-edge point (REP) wavelength (in nm) at the leaf level as a function of the number of days post-inoculation. **Visible symptoms.
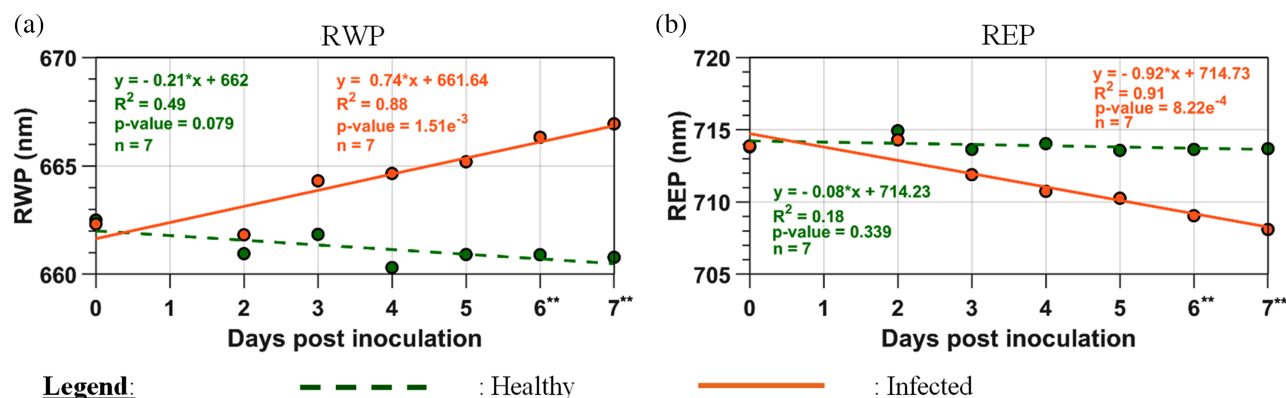

(a)

RWP

$y = -0.21*x + 662$
$R^2 = 0.49$
p-value = 0.079
n = 7

$y = 0.74*x + 661.64$
$R^2 = 0.88$
p-value = $1.51e^{-3}$
n = 7

(b)

REP

$y = -0.92*x + 714.73$
$R^2 = 0.91$
p-value = $8.22e^{-4}$
n = 7

$y = -0.08*x + 714.23$
$R^2 = 0.18$
p-value = 0.339
n = 7

Legend: – – – : Healthy   ——— : Infected

Table 4. Comparison between spectral features* of the linear support vector machine (SVM) classifier and overall accuracy as a function of the day post-inoculation (DPI).

DPI	PC	RWP	REP
0	65.35	55.47	55.47
2	86.61	57.81	54.69
3	77.17	72.66	61.72
4	95.28	78.13	71.88
5	85.04	78.13	69.53
6**	97.64	89.06	86.72
7**	96.85	94.07	83.59

Note: *PC, 5 first principal components; RWP, red-well point; REP, red-edge point. **Visible symptoms.

In order to determine the best wavelengths to detect powdery mildew, we computed spectral ratios between infected and healthy spectra, such as in Fernández et al. (2020a). When powdery mildew signs and symptoms were visible for the first time (at 6 DPI), the wavelengths we defined as being the best to detect the disease were the blue (444 nm) and green (570 nm) ones. They are closely located to those reported by Xie et al. (2015) for late blight disease detection (i.e., visible symptoms) on tomato leaves (442 and 573 nm, respectively). However, Xie et al. (2015) also reported 508, 696, and 715 nm as suitable wavelengths that are not closely related to our wavelengths. None of our wavelengths were close to those used by Xie et al. (2017) to detect *Botrytis cinerea* infecting tomato leaves. They were also different from those reported by Fernández et al. (2020a) (488, 556, 681, and 709 nm) for late blight disease detection in potato leaves when symptoms were visible for the first time at 3 DPI. At 5 and 6 DPI, the 703 nm wavelength matches the wavelength reported by Xie and He (2016), who tested spectral and image textural features to detect

Alternaria solani (early blight) on eggplant (*Solanum melongena*) leaves. The authors also reported wavelengths located at 408, 535, and 624 nm. These discrepancies might be explained by anatomical differences between plants species (Fernández et al. 2020a). Also, it must be noted that Fernández et al. (2020a), Xie et al. (2017), Xie et al. (2015), and Xie and He (2016) studied a necrotic pathogen, while cucumber *P. xanthii* is a biotrophic pathogen.

A linear SVM was applied to the first 5 PC scores, the RWP or REP wavelengths, to sort the leaf spectra according to the leaf status (healthy or infected) as a function of the DPI. Using five PCs, the lowest overall classification accuracy for healthy and infected leaves was around 65% at 0 DPI. The highest overall accuracy was around 97% at 6 DPI when the symptoms were visible. Our overall classification accuracy was higher than those of Fernández et al. (2020a). They reported an overall accuracy of around 89% at 3 DPI when late blight symptoms were visible for the first time on potato leaves, using a partial least square discriminant analysis (PLS-DA) applied leaf reflectance from 400 to 900 nm. However, our accuracies were lower than those (100%) reported at 4 DPI when disease signs were observed in Berdugo et al. (2014), who applied a stepwise discriminant analysis to discriminate healthy cucumber leaves from those infected with *P. xanthii* using the following features: effective quantum yield, SPAD values, maximum temperature difference (MTD), NDVI and the anthocyanin reflectance index (ARI).

The leaf spectra we measured were used to compute two main parameters of the red and red-edge regions, the RWP and REP wavelengths. For both parameters, the best fit was achieved using a linear model. For the REP, our results are not following Fernández et al. (2020b). They reported an exponential model to fit best the REP changes in the case of potato late blight, which is due to a hemibiotrophic pathogen that necrotizes infected tissue. This study studied variations induced by powdery mildew due to a biotrophic pathogen that

↴ Published by Canadian Science Publishing

Table 5. Confusion matrix and related statistics when a linear support vector machine (SVM) classifier is applied to the red-well point (RWP) wavelength to classify 71 healthy and 57 infected spectral samples as a function of the number of days post-inoculation (DPI).

DPI	Confusion matrix			Specificity	Precision	F1 Score	Overall accuracy (%)
0	—	I	H	Ind	0.00	Ind	55.47
	I	0	57				
	H	0	71				
2	—	I	H	0.56	0.25	0.34	57.81
	I	14	43				
	H	11	60				
3	—	I	H	0.74	0.60	0.66	72.66
	I	34	23				
	H	12	59				
4	—	I	H	0.76	0.74	0.75	78.13
	I	42	15				
	H	13	58				
5	—	I	H	0.74	0.79	0.76	78.13
	I	45	12				
	H	16	55				
6	—	I	H	0.88	0.88	0.88	89.06
	I	50	7				
	H	7	64				
7	—	I	H	0.93	0.95	0.94	94.07
	I	54	3				
	H	4	67				

Note: Ind, indeterminate.

interacts with the host without producing necrosis. The slope of the linear model was higher for the REP model than for the RWP model; hence the REP parameter is more sensitive to the disease than the RWP. For the infected leaves, the R^2 values were 0.88 with RWP and 0.91 with REP. The one for RWP was higher than the one reported by Fernández et al. (2020*b*) (0.83), but the one for REP was lower than the one reported by Fernández et al. (2020*b*) (0.99). When the symptoms were visible for the first time, our overall classification accuracies with the RWP and REP wavelengths position (89% and 86%, respectively) were higher than those (65% with RWP and 59% with REP) reported by Fernández et al. (2020*b*) who applied a linear SVM classifier to classify healthy potato leaves and those infected with *P. infestans.*

Early disease detection is a challenging topic and might not find a definitive solution yet. One main limitation for early disease detection can be related to the conditions where the experiment was performed. Such as the ones of the current study, most of the literature results on disease detection have been obtained under controlled conditions with optimal light illumination and plants having excellent growing conditions. There

is, therefore, the need for further work to up-scale the laboratory measurements to actual crop conditions. Gold et al. (2020) already reported that using leaf-level spectroscopy for disease scouting has been already reported as unfeasible. Khaled et al. 2018 suggested that it is necessary to consider the effect of environmental light for measuring spectral data in the field. Another limitation when studying plant-pathogen interactions is the spectral range of the available sensors (Khaled et al. 2018). Usually, the spectral range for image acquisition is focused in the visible-NIR range (400–1000 nm) because the majority of the commercial cameras operate in this range. Indeed, short-wave infrared sensors are up to five times more expensive than visible and NIR sensors (Mishra et al. 2020). Finally, a universal method will be difficult to develop as the specific wavelengths required to detect specific diseases depend on the nature of the pathogen. Three different types of pathogens can be found in nature: biotrophic, hemibiotrophic, or necrotrophic. In the case of biotrophic pathogens such as *P. xanthii* (this study) and hemibiotrophic pathogens such as *P. infestans,* the useful wavelengths for disease detection change with the disease progress (Fernández

et al. 2020*b*, Gold et al. 2020). For a necrotrophic pathogen such as *Alternaria solani* on potato leaves, the best wavelength was determined to be in the short-wave infrared region (Gold et al. 2020).

Conclusions

In this study, we determined which spectral variables could be used to detect powdery mildew on cucumber leaves, as well as the timing of onset following disease inoculation. Using principal component analysis over the whole visible spectra range, between 400 and 900 nm, allowed us to reduce the number of variables from 501 wavelengths to 5 components explaining above 97% of the total variance. However, the classification based on principal components is affected by instrumental noise observed in the 400–440 nm region. Regarding the two red and red-edge parameters (RWP and REP) we studied, REP was more sensitive to the disease than RWP. However, when a linear SVM is applied to REP and RWP to classify healthy and infected leaves, higher overall classification accuracy was achieved earlier with RWP than with REP.

Our results were obtained on Straight Eight cultivar cucumber plants. Further work is needed to test the method on other cucumber cultivars. While the results of this study are quite promising, they were acquired on a limited number of plants; therefore, there is a need to test the method over a large number of plants. Our results were obtained in a walk-in chamber with a controlled environment using point measurements. Future studies should address testing the method in actual greenhouse conditions, and on hyperspectral imagery acquired over those greenhouse conditions.

Acknowledgments

References

Abdulridha, J., Ampatzidis, Y., Roberts, P., and Kakarla, S.C. 2020. Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. Biosyst. Eng. **197**: 135–148. doi:10.1016/j.biosystemseng.2020.07.001.

Agriculture and Agri-Food Canada. 2020. [Online]. Available from https://www.agr.gc.ca/eng/canadas-agriculture-sectors/horticulture/horticulture-sector-reports/statistical-overview-of-the-canadian-greenhouse-vegetable-industry-2019/?id=1609252451699 [10 May 2021].

Atanassova, S., Nikolov, P., Valchev, N., Masheva, S., and Yorgov, D. 2019. Early detection of powdery mildew (*Podosphaera xanthii*) on cucumber leaves based on visible and near-infrared spectroscopy. Proc. 10th Jubilee International Conference of the Balkan Physical Union, Sofia, Bulgaria, 26–30 Aug. 2019, Proc. 2075, p. 160014. doi:10.1063/1.5091341.

Berdugo, C.A., Zito, R., Paulus, S., and Mahlein, A.-K. 2014. Fusion of sensor data for the detection and differentiation of plant diseases in cucumber. Plant Pathol. **63**: 1344–1356. doi:10.1111/ppa.12219.

Cen, H., Lu, R., Zhu, Q., and Mendoza, F. 2016. Non-destructive detection of chilling injury in cucumber fruit using hyperspectral imagery with feature selection and supervised classification. Postharvest Bio. Technol. **111**: 352–361. doi:10.1016/j.postharvbio.2015.09.027.

Darvishzadeh, R., Atzberger, C., Skidmore, A.K., and Abkar, A.A. 2009. Leaf area index deviation from hyperspectral vegetation indices and the red edge position. Int. J. Remote Sens. **30**: 6199–6218. doi:10.1080/01431160902842342.

Delalieux, S., Van Aardt, J., Keulemans, W., Schrevens, E., and Coppin, P. 2007. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications. Eur. J. Agron. **27**: 130–143. doi:10.1016/j.eja.2007.02.005.

Eskandari, S., and Sharifnabi, B. 2019. The modifications of cell wall composition and water status of cucumber leaves induced by powdery mildew and manganese nutrition. Plant Physiol. Biochem. **145**: 132–141. doi:10.1016/j.plaphy.2019.10.037.

Everis, B., and Dunn, G. 2001. Principal component analysis. Pages 48–73 in Everis, B., Dunn, G., eds. Applied multivariate data analysis. 2nd ed. John Wiley & Sons, Ltd, Chichester, England.

Fernández, C.I., Leblon, B., Haddadi, A., Wang, J., and Wang, K. 2020*a*. Potato late blight detection at the leaf and canopy level using hyperspectral data. Can. J. Remote Sens. **46**: 390–413. doi:10.1080/07038992.2020.1769471.

Fernández, C.I., Leblon, B., Haddadi, A., Wang, K., and Wang, J. 2020*b*. Potato late blight detection at the leaf and canopy levels based on red and red-edge spectral regions. Remote Sens. **12**: 1292. doi:10.3390/rs12081292.

Fernández, C.I., Leblon, B., Wang, J., Haddadi, A., and Wang, K. 2021. Detecting infected cucumber plants with close-range multispectral imagery. Remote Sens. **13**: 2948. doi:10.3390/rs13152948.

Gitelson, A.A., Merzlyak, M.N., and Lichtenthaler, H.K. 1996. Detection of red-edge position and chlorophyll content by reflectance measurements near 700 nm. J. Plant Physiol. **148**: 501–508. doi:10.1016/S0176-1617(96)80285-9.

Gold, K.M., Townsend, P.A., Chlus, A., Herrmann, I., Couture, J.J., Larson, E.R., and Gevens, A.J. 2020. Hyperspectral measurements enable pre-symptomatic detection and differentiation of contrasting physiological effects of late blight and early blight in potato. Remote Sens. **12**: 12020286. doi:10.3390/rs12020286.

Gulhane, V.A., and Kolekar, M.H. 2014. Diagnosis of diseases on cotton leaves using principal component analysis classifier. Proc. Annual IEEE India Conference (INDICON), Pune, India, 11–13 December 2014. pp. 1–5. doi:10.1109/INDICON.2014.7030442.

Hafez, Y.M., Attia, K.A., Kamel, S., Alamery, S.F., El-Gendy, S., Al-Doss, A.A., et al. 2020. Bacillus subtilis as a bio-agent combined with nano molecules can control powdery mildew

disease through histochemical and physiobiochemical changes in cucumber plants. Physiol. Mol. Plant Pathol. 111: 101489. doi:10.1016/j.pmpp.2020.101489.

Horler, D.N.H., Dockray, M., and Barber, J. 1983. The red edge of plant leaf reflectance. Int. J. Remote Sens. 4: 273–288. doi:10.1080/01431168308948546.

Isaksson, T., and Næs, T. 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. Appl. Spectrosc. 42: 1273–1284. doi:10.1366/0003702884429869.

Kalischuk, M., Paret, M.L., Freeman, J.H., Raj, D., Da Silva, S., Eubanks, S., et al. 2019. An improved crop scouting technique incorporating unmanned aerial vehicle-assisted multispectral crop imaging into conventional scouting practice for gummy stem blight in watermelon. Plant Dis. 103: 1642–1650. doi:10.1094/PDIS-08-18-1373-RE.

Khaled, A.Y., Aziz, S.A., Bejo, S.K., Nawi, N.M., Semana, I.A., and Onwude, D.I. 2018. Early detection of diseases in plant tissue using spectroscopy – applications and limitations. Appl. Spectrosc. Rev. 53: 36–64. doi:10.1080/05704928.2017.1352510.

Khater, M., de la Escosura-Muñiz, A., and Merkoçi, A. 2017. Biosensors for plant pathogen detection. Biosens. Bioelectron. 93: 72–86. doi:10.1016/j.bios.2016.09.091.

Knipling, E.B. 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sens. Environ. 1: 155–159. doi:10.1016/S0034-4257(70)80021-9.

Kumar, N., Bansal, A., Sarma, G.S., and Rawal, R.K. 2014. Review Chemometrics tools used in analytical chemistry: an overview. Talanta, 123: 186–199. doi:10.1016/j.talanta.2014.02.003.

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plumer, L., Steiner, U., and Oerke, E.-C. 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sens. Environ. 128: 21–30. doi:10.1016/j.rse.2012.09.019.

Martinez, W.L., and Cho, M. 2015. Statistics in Matlab® A Primer. CRC Press, Boca Raton, Florida, USA, pp. 182–222.

Mishra, P., Polder, G., and Vilfan, N. 2020. Close range spectral imaging for disease detection in plants using autonomous platforms: a review on recent studies. Curr. Robot. Rep. 1: 43–48. doi:10.1007/s43154-020-00004-7.

Mutanga, O., and Skidmore, A.K. 2007. Red edge shift and biochemical content in grass canopies. ISPRS J. Photogramm. Remote Sens. 62: 34–42. doi:10.1016/j.isprsjprs.2007.02.001.

Nishizawa, Y., Mochizuki, S., Yokotani, N., Nishimura, T., and Minami, E. 2016. Molecular and cellular analysis of the biotrophic interaction between rice and Magnaporthe oryzae – exploring situations in which the blast fungus controls the infection. Physiol. Mol. Plant Pathol. 95: 70–76. doi:10.1016/j.pmpp.2016.02.001.

Peng, Y., Gitelson, A.A., Keydan, G., Rundquist, D.C., and Moses, W. 2011. Remote estimation of gross primary production in maize and support for a new paradigm based on total crop chlorophyll content. Remote Sens. Environ. 115: 978–989. doi:10.1016/j.rse.2010.12.001.

Pérez-García, A., Romero, D., Fernández-Ortuño, D., López-Ruiz, F., de Vicente, A., and Torés, J.A. 2009. The powdery mildew fungus *Podosphaera fusca* (synonym *Podosphaera xanthii*), a constant threat to cucurbits. Mol. Plant Pathol. 10: 153–160. doi:10.1111/j.1364-3703.2008.00527.x.

Pu, R., Gong, P., Biging, G.S., and Larrieu, M.R. 2003. Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. IEEE Trans. Geosci. Remote Sens. 41: 916–921. doi:10.1109/TGRS.2003.813555.

Sankaran, S., Mishra, A., Ehsani, R., and Davis, C. 2010. A review of advanced techniques for detecting plant diseases. Comp. Electron. Agric. 72: 1–13.

Savitzky, A., and Golay, M.J.E. 1964. Smoothing and differentiation of data by simplified least-squares procedures. Anal. Chem. 36: 1627–1639. doi:10.1021/ac60214a047.

Spanu, P.D., and Panstruga, R. 2017. Biotrophic plant-microbe interactions. Front. Plant Sci. 8: 192. doi:10.3389/fpls.2017.00192.

Tanaka, R., and Tanaka, A. 2011. Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. Biochim. Biophys. Acta Bioenerg. 1807: 968–976. doi:10.1016/j.bbabio.2011.01.002.

Tian, Y., and Zhang, L. 2012. Study on the methods of detecting cucumber downy mildew using hyperspectral imaging technology. Phys. Procedia. 33: 732–750. doi:10.1016/j.phpro.2012.05.130.

Wang, H., Li, G., Ma, Z., and Li, X. Image recognition of plant disease based on principal component analysis and neural networks. In Proceedings of the 8th International Conference on Natural Computation (ICNC 2012), Chongqing, China, 29–31 May 2012. pp. 246–251. doi:10.1109/ICNC.2012.6234701.

Wspanialy, P., and Moussa, M. 2016. Early powdery mildew detection system for application in greenhouse automation. Comp. Electron. Agric. 127: 487–494. doi:10.1016/j.compag.2016.06.027

Xie, C., and He, Y. 2016. Spectrum and image texture feature analysis for early blight disease detection on eggplant leaves. Sensors, 16: 676. doi:10.3390/s16050676.

Xie, C., Shao, Y., Li, X., and He, Y. 2015. Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. Sci. Rep. 5: 16564. doi:10.1038/srep16564.

Xie, C., Yang, C., and He, Y. 2017. Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities. Comp. Electron. Agric. 135: 154–162. doi:10.1016/j.compag.2016.12.015.

Zhang, S., Wu, X., You, Z., and Zhang, L. 2017. Leaf image based cucumber disease recognition using sparse representation classification. Comp. Electron. Agric. 134: 135–141. doi:10.1016/j.compag.2017.01.014.

Zhang, S., Zhang, S., Zhang, C., Wang, X., and Shi, Y. 2019. Cucumber leaf disease identification with global pooling dilated convolutional neural network. Comp. Electron. Agric. 162: 422–430. doi:10.1016/j.compag.2019.03.012.