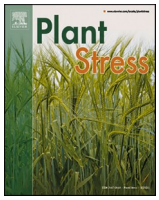
Plant Stress 15 (2025) 100732

Contents lists available at ScienceDirect

Plant Stress

journal homepage: www.sciencedirect.com/journal/plant-stress



Non-invasive and early detection of *tomato spotted wilt virus* infection in tomato plants using a hand-held Raman spectrometer and machine learning modelling

Ciro Orecchio [a,1], Camilla Sacco Botto [b,1], Eugenio Alladio [a], Chiara D'Errico [b,*], Marco Vincenti [a,**], Emanuela Noris [b]

[a] *Department of Chemistry, University of Turin, Turin 10125, Italy*
[b] *Institute for Sustainable Plant Protection, National Research Council of Italy, Turin 10135, Italy*

ARTICLE INFO

Keywords:
Plant viruses
Tomato disease
Raman spectroscopy
Partial least squares discriminant analysis
Machine Learning

ABSTRACT

Tomato spotted wilt virus (TSWV) is a polyphagous thrips-transmitted pathogen inducing significant economic losses in agriculture, particularly on tomato plants. The leading management and containment strategies to fight TSWV infection rely on growing resistant cultivars and spraying insecticides for thrips control. Therefore, its early detection is fundamental in sustainable crop management. Aim of the present work is to reveal TSWV infection using a hand-held Raman instrument and Machine Learning (ML) approaches. Artificially inoculated tomato plants were scored for symptom development for one month, while Raman spectra were collected 3 and 7 days after virus inoculation. After preliminary spectral pre-processing, a filter method based on Partial Least Squares Discriminant Analysis (PLS-DA) coefficients was applied to remove redundant and irrelevant variables. The resulting condensed dataset was checked with multivariate exploratory methods and exploited to build multiple PLS-DA models, using different random splitting of the samples between training and test sets. By interpreting the classification metrics, Raman spectroscopy coupled with ML techniques allowed us to discriminate infected from healthy tomato plants within the first 3–7 days after inoculation, with average accuracy of 90–95 % in validation. The model was also validated on two different sets of susceptible and resistant plants, achieving average accuracy higher than 85 %. Early detection of TSWV infection well before visual symptom occurrence represents an important advantage in a sustainable agricultural system. Notably, the use of a portable Raman spectrometer, much less expensive and cumbersome than benchtop instruments, allows the direct in-field execution of these diagnostic measurements.

Abbreviations

TSWV	Tomato spotted wilt virus
PCR	Polymerase chain reaction
RS	Raman spectroscopy
ML	Machine Learning
dpi	Days post inoculation
arPLS	Asymmetrically reweighted penalized least squares
SNV	Standard Normal Variate
PLS-DA	Partial Least Squares Discriminant Analysis
LV	Latent variable
PCA	Principal Components Analysis

1. Introduction

Tomato spotted wilt virus (TSWV) (family *Tospoviridae,* genus *Orthotospovirus,* species *Orthotospovirus tomatomaculae*) infects >1000 different plant species from 90 botanical families, including ornamental, fruit, horticultural, and agronomic crops (Ruark-Seward et al., 2020). TSWV is one of the 10 most destructive viruses infecting horticultural

* Corresponding author at: Chiara D'Errico, Institute for Sustainable Plant Protection, National Research Council of Italy, Strada delle Cacce 73 10135, Torino, Italy.
** Corresponding author at: Marco Vincenti, Department of Chemistry, University of Turin, Via Pietro Giuria 7 10125, Torino, Italy.
E-mail addresses: chiara.derrico@ipsp.cnr.it (C. D'Errico), marco.vincenti@unito.it (M. Vincenti).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.stress.2024.100732
Received 22 August 2024; Received in revised form 9 December 2024; Accepted 29 December 2024
Available online 1 January 2025

crops in the world (Scholthof et al., 2011), causing more than a billion dollars damage each year, therefore being one of the major threats to both specialty and staple crops around the world. This virus is transmitted by several thrips species, mainly the western flower thrips (*Frankliniella occidentalis*), but also the onion thrips (*Thrips tabaci*), and the chili thrips (*Scirtothrips dorsalis*) (Ullman et al., 2002). The thrips vector too is infected by this virus. On tomato plants, TSWV induces different symptoms, including leaf bronzing, small brown flecks and chlorotic spots, stunting, inward cupping and dropping of leaves, unilateral plant growth, dieback of growing tips, and ultimately death.

The major management and containment strategies of this disease rely on the use of resistant cultivars and insecticides for thrips control. However, to reduce the spread of the disease, limit crop damage, calibrate the use of insecticides, and support the breeding procedures to identify new sources of genetic resistance, early detection of this plant pathogen is fundamental in a sustainable crop management context. ELISA assays and polymerase chain reaction (PCR) are the most frequently utilized diagnostic techniques for TSWV recognition (Chinnaiah et al., 2022; Gao and Wu, 2022; Iturralde Martinez and Rosa, 2023; Roberts et al., 2000). However, these tests are laborious, invasive, costly, and time-consuming, therefore inappropriate for a broad range screening. Recently, innovative techniques for non-invasive disease detection have been developed, based on Raman spectroscopy (RS) (Payne and Kurouski, 2021; Saletnik et al., 2024), volatile compound detection (Li et al., 2019), and hyperspectral imaging (Nguyen et al., 2021).

RS exploits the interaction between photons emitted by a laser hitting the sample and its molecular components. The inelastic scattering of photons results in an energy shift of these photons, which is related to the molecular structure of the sample components and their vibrational modes, ultimately providing information about the chemical composition of the sample. RS has been applied for the detection of abiotic (Altangerel et al., 2017; Sanchez et al., 2020a) and biotic (Baratto et al., 2022; Egging et al., 2018; Farber and Kurouski, 2018; Kong et al., 2024; Mandrile et al., 2019, 2022; Sanchez et al., 2019a, 2019b) stress responses, fruit quality (Nekvapil et al., 2018), and chemical contamination (Mandrile et al., 2018). Moreover, RS is useful for quick and precise plant phenotyping, as well as for evaluating the nutritional content of grains (Farber et al., 2020; Krimmer et al., 2019).

We have previously described the RS capability of detecting the TSWV infection in tomato plants at an early stage, i.e. when symptoms are not yet visible (Mandrile et al., 2019). Moreover, we showed that RS allows to discriminate TSWV infection from the attack of other endemic pathogens affecting this crop (Mandrile et al., 2019). Recently, a RS-based approach was applied to confirm the infection by different TSWV strains in resistant and susceptible tomato varieties, but in this case RS analysis was performed only on symptomatic leaves collected at late time points after the artificial inoculation (Juárez et al., 2024).

Our previous experiments were conducted with a highly sensitive benchtop Raman spectrometer, which is nevertheless expensive, cumbersome, and unsuitable for non-destructive field applications. In this study, we show that the presence of different isolates of TSWV can be detected in both susceptible and resistant tomato plants at a very early stage of infection with a hand-held Raman device, ideal for field usage. To optimize the diagnostic performance of RS technique, advanced Machine Learning (ML) elaboration of the spectroscopic data has been implemented, allowing to achieve high accuracy levels. The model obtained was externally validated on two independent sets of susceptible and resistant plants. This study further supports the suitability of hand-held RS for point-of-care applications in plant pathology and plant phenotyping.

2. Materials and methods

2.1. Plants and virus isolates

Tomato plants (cv. Marmande) were grown in soil in a greenhouse compartment at an average temperature of 23 °C (day) and 19 °C (night). Plants ($n = 4$) were mechanically inoculated at the 4-leaf stage with TSWV (P105 isolate), as previously described (Mandrile et al., 2019). In brief, a homogenate obtained by grinding 1 g of tobacco leaves systemically infected by TSWV in 10 ml inoculation buffer (20 mM Na_2SO_3, 10 mM Na-diethyldithiocarbamate, 5 mM EDTA) was applied to the upper leaf surface of the fourth leaf starting from the apex (Fig. 1a), by gentle rubbing with carborundum. Mock-inoculated plants ($n = 4$) received the same treatment, using a homogenate made with healthy leaves. . A second inoculation experiment, using 3 plants (cv. Marmande) per condition, was conducted and the Raman spectra acquired at the same time points were used to validate the model. To validate the model protocol on different susceptibility rates, a third experiment was conducted using resistant tomato plants (F1 hybrid York). Plants ($n = 4$) were inoculated with the resistance-breaking TSWV T992 isolate (Ciuffo et al., 2005; https://www.european-virus-archive.com/) or the natural isolate I244, while mock-inoculated plants served as control. All TSWV isolates belong to the PLAVIT collection (IPSP-CNR Torino, Italy, World Data Center for Microorganism (WDCM) no. 1057).

2.2. Raman spectroscopy

Raman spectra of tomato leaves were collected using a hand-held Bruker BRAVO spectrometer (Bruker Optik GMBH, Ettlingen, Germany) equipped with two laser sources (785 and 853 nm). To suppress fluorescence BRAVO uses a sequentially shifted excitation (SSE) technology, as detailed by Jehlička et al. (2017). The spectrometer recorded spectra over a range of 300–3200 cm^{-1} with a sampling resolution of 10–12 cm^{-1}. The OPUS software (Bruker, version 8.2) was used for data transfer, while a Spectragryph (version 1.2.16.1) optical spectroscopy software was used for data conversion, in particular from 0.0 files to .csv files (Menges, 2001). For each plant, Raman spectra were acquired from the apical leaflet of the second and third leaves from the apex (Fig. 1a). Three spectra for each leaf were taken in different spots, at both 3 and 7 days post inoculation (dpi). Overall, 48 spectra, i.e. 24 spectra for each time point, were acquired for both mock-inoculated ('Healthy') and virus-inoculated ('TSWV') plants. In the second experiment, aimed at validating the ML model, additional 28 spectra for virus-inoculated ('TSWV') and 37 for mock-inoculated ('Healthy') plants were collected, using the Bruker BRAVO spectrometer. In the third experiment on York plants, Raman spectra were collected with the same procedure, using a hand-held Zolix Finder Edge spectrometer (Zolix Instruments Co., LTD, Beijing, China) equipped with a laser source at 1064 nm, recording spectra over a range of 200–2000 cm^{-1} with a sampling resolution of 14 cm^{-1}. The Spectragryph (version 1.2.16.1) optical spectroscopy software was used for data acquisition and data conversion, in particular from .txt to .csv files. In this case, 46 spectra for virus-inoculated ('TSWV') and 24 for mock-inoculated ('Healthy') were collected at 7 dpi on non-symptomatic leaves.

2.3. Chemometric analysis

2.3.1. Data matrix and Raman spectra pre-processing

The .csv files were converted to .xlsx files and analyzed with Python (Van Rossum, 1995) (version 3.11.8) after conversion in a matrix form (96 samples × 601 wavelengths), obtained by selecting the wavelengths range 600–1800 cm^{-1} as previously described (Mandrile et al. 2019; Vallejo-Pérez et al., 2021). Afterwards, three different pre-processing algorithms were applied sequentially to the spectra, namely (i) asymmetrically reweighted penalized least squares (arPLS) (Baek et al., 2015;

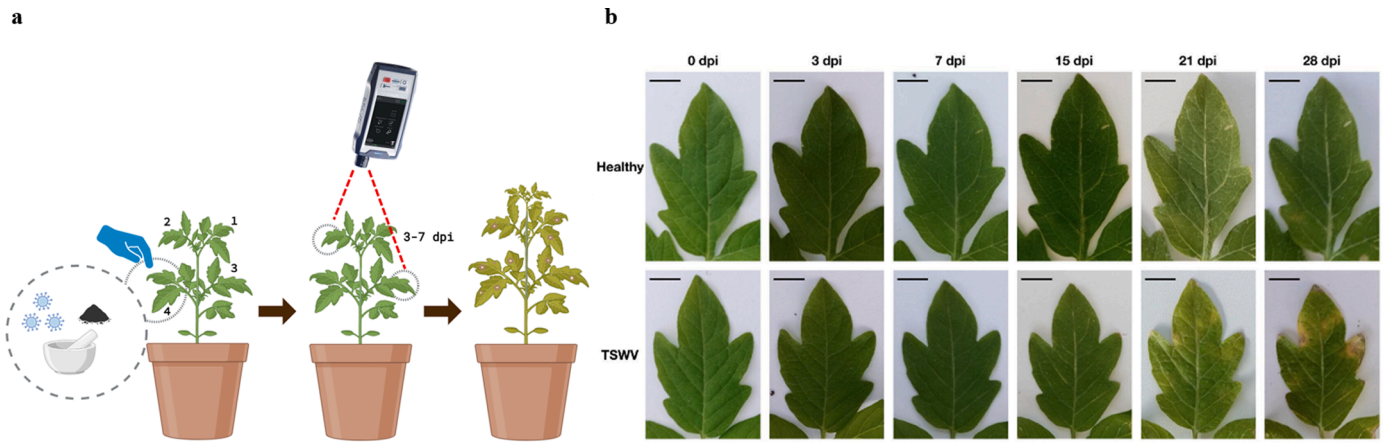


Fig. 1. Scheme of the experimental procedure (**a**) and images of tomato (cv. Marmande) leaflets (**b**) collected at the beginning of the experiment (0 days post inoculation, dpi) and at different times (3 to 28 dpi) after inoculation with TSWV; 'Healthy' leaflets represent mock-inoculated samples. Leaf numbering is indicated in the plant scheme on the left; TSWV inoculum was delivered on the fourth leaves from the apex, while Raman spectra were acquired on the apical leaflets of the second and third leaves from the apex; scale bar, 1 cm. Created in BioRender.com (2025).

Górski and Jakubowska, 2023) with smoothing for baseline correction (λ) set to 1000 and speed of algorithm convergence (*r*) set to $1{\cdot}10^{-6}$; (ii) Savitzky-Golay filter (Savitzky and Golay, 1964) for smoothing and differentiation of the signals with window length = 7 and polynomial order = 2; (iii) Standard Normal Variate (SNV), a powerful spectral data normalization algorithm that effectively removes the constant offset and the multiplicative effects between spectra (Martyna et al., 2020).

2.3.2. Features selection algorithm

Before applying the ML techniques described below, a variable selection strategy for classification purposes was performed. After the initial data-cropping, the full resulting matrix contained 601 variables. To remove the redundant or irrelevant variables (Yun et al., 2019; Li et al., 2021) and improve the model qualitative prediction, an appropriate features extraction algorithm was adopted, by implementing a Partial Least Squares Discriminant Analysis (PLS-DA) filter method based on regression coefficients, as described below (Chong and Jun, 2005; Florián-Huamán et al., 2022; Frenicha et al., 1995; Mehmood et al., 2012). First, a PLS-DA model was optimized on the full spectrum of 601 variables. Secondly, the model coefficients were extracted, in order to express the relationship between each feature (wavelength) and the qualitative response (class: Healthy vs. TSWV). Then, the dataset was subjected to an iterative approach which, at each cycle, discarded the wavelength with the lowest absolute value of the associated coefficient and rebuilt the classification model. By choosing the accuracy score in cross-validation as the evaluation parameter, the procedure was iterated in a loop which tuned the number n of latent variables at each iteration ($1 < n < 15$) until a maximum accuracy level was reached. Lastly, a total of 150 relevant wavelengths were extracted (Supplementary Fig. 1) and used as variables in the subsequent elaborations.

2.4. Machine learning techniques and classification metrics

The dataset resulting from the feature-selection process (96 samples × 150 wavelengths) was used for (i) an exploratory analysis through Principal Component Analysis (PCA) and (ii) to build an optimized PLS-DA classification model. For the development and validation of this classification model, a random data-splitting process was applied to the 96 spectra, considering 72 and 24 spectra as calibration set and test set, respectively. Since a unique random training/test split may affect the results (either by including or excluding outliers in the test set), this data-splitting procedure was repeated ten times so as to stress the model's robustness. Moreover, a stratified sampling strategy was adopted to preserve the class distribution within the categorical variable. Each training model was then optimized with a k-fold cross-validation ($k = 5$) strategy, again based on maximizing the accuracy score in setting the proper number of LVs. These LVs represent the hyper-spatial directions along which the maximum covariance between X (data matrix) and Y (categorical response) is captured (Ballabio and Consonni, 2013). In detail, a fixed number of five LVs was maintained, representing a good trade-off between prediction accuracy and model complexity while avoiding overfitting. The optimal, ten final trained PLS-DA models were validated by evaluating their performance on the corresponding test sets and the resulting classification reports and confusion matrices. In order to test this ML procedure on totally independent samples and data, further validation was achieved by building a final PLS-DA model using all 96 samples and applying it on an external dataset arising from 65 spectra, 37 of which belong to infected plants, inoculated 3–7 days before the spectral collection. Furthermore, the same ML protocol was applied to an additional dataset relative to the inoculation on the resistant tomato plants, consisting of 70 spectra, 46 of which were from infected plants, inoculated 7 days before spectra acquisition.

3. Results and discussion

3.1. Time course of virus infection

Systemic symptoms of TSWV commonly appear on tomato plants as typical chlorotic spots after more than one week, depending on plant cultivar and environmental conditions. In the present experiments, no visible systemic symptoms could be observed in any portion of the plant at 3 and 7 dpi, at the same time points of Raman spectra acquisitions on the apical leaflets. Chlorotic spots started to emerge at 15 dpi on some systemically infected leaves, while the apical leaflets used for Raman acquisition developed visible symptoms at 21 dpi (Fig. 1b). Mock-inoculated plants ('Healthy') remained symptomless across the whole experiment.

3.2. Preliminary analysis of tomato leaf spectra

The average leaf spectra of both Healthy and TSWV-infected tomato plants (acquired at 3 and 7 dpi) showed vibrational bands originating from the major chemical components of plants, *i.e.* cellulose, lignin, carotenoids, and chlorophyll (Fig. 2, Table 1). In detail, the most evident peaks were those relative to carotenoids (1004, 1156, 1186, 1526 cm^{-1}), chlorophyll (1224, 1326 - 1328 cm^{-1}), pectin (peaks between 740 and 746 cm^{-1}) and cellulose (peaks at 916 and 1094 cm^{-1}

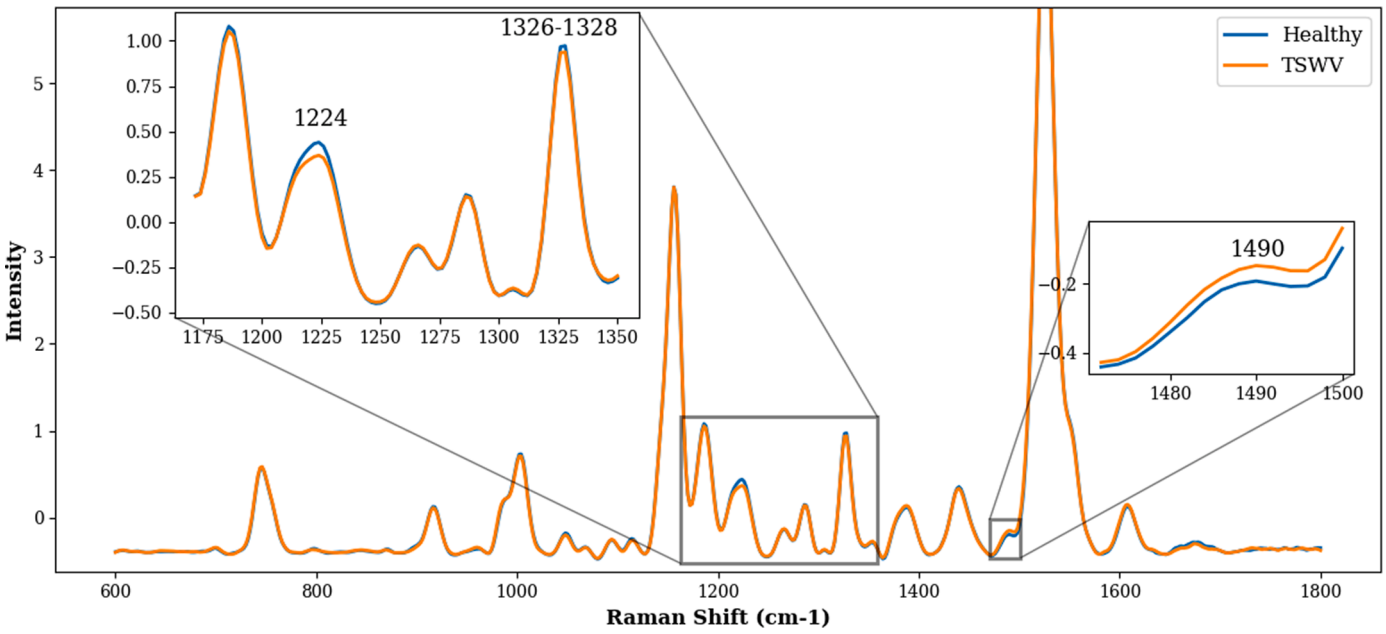


Fig. 2. Average Raman spectra for Healthy (mock-inoculated, blue line) and TSWV-infected (isolate P105, orange line) leaves of tomato (cv. Marmande), collected in the range 600–1800 cm^{-1}. The insets show the details of the peaks in the range 1175–1350 cm^{-1} and 1470–1500 cm^{-1} of the spectra.

Table 1
Main vibrational bands and their assignments for the Raman spectra collected from tomato plants.

Band (cm^{-1})	Vibrational mode	Assignment	References
744	γ(C–O–H) of COOH	Pectin	(Sanchez et al., 2020b)
916, 920	ν(C–O–C) in plane, symmetric	Cellulose, lignin	(Sanchez et al., 2020b; Vallejo-Pérez et al., 2021)
1004	ν3(C–CH3 stretching)	Carotenoids	(Mandrile et al., 2019; Sanchez et al., 2020b; Vallejo-Pérez et al., 2021)
1094	ν(CO) of secondary alcohol	Cellulose	(Sanchez et al., 2020b; Vallejo-Pérez et al., 2021)
1156	ν2(C‖C stretching)	Carotenoids	(Mandrile et al., 2019; Vallejo-Pérez et al., 2021)
1186	ν(C–O–H) next to aromatic ring + δ(CH)	Carotenoids	(Juárez et al., 2024; Saletnik et al., 2024)
1224	δ(CH). δ(CH2)	Chlorophyll	(Mandrile et al. 2019)
1326, 1328	δ(CH). ν(CN)	Chlorophyll	(Mandrile et al., 2019; Vallejo-Pérez et al., 2021)
1440	ν(phenyl ring)	Phenolics	(Mandrile et al., 2019)
1490	δ(CH2) and δ(CH3), bending	Aliphatic	(Mandrile et al. 2019)
1526	ν1(C − C)	Carotenoids	(Sanchez et al., 2020b; Vallejo-Pérez et al., 2021)
1610	ν(C–C) aromatic ring + σ(CH)	Lignin	(Mandrile et al., 2019; Sanchez et al., 2020b)
1650–1690	Amide I	Proteins	(Sanchez et al., 2020b; Vallejo-Pérez et al., 2021)

associated to ν(C–O–C) and ν(CO) modes of cellulose), while the peaks at 920 and 1610 cm^{-1} are related to lignin. Carotenoids generated peaks at 1004, 1156, and 1526 cm^{-1}, attributed to in-plane CH_3 rocking modes, and C = C and C—C stretching, respectively. Lastly, the peaks between 1650 and 1690 cm^{-1} can be assigned to proteins.

By averaging the 48 spectra collected from 'Healthy' plants and the 48 from 'TSWV', and comparing the averaged spectra, some minor differences could be evidenced (Fig. 2). In particular, it was observed that the signals relative to chlorophylls at 1224 and 1326–1328 cm^{-1}, carotenoids at 1156 and 1185 cm^{-1}, polyphenols at 1440 cm^{-1}, and those due to proteins at 1650–1690 cm^{-1} appeared altered in TSWV-infected samples, showing a slight reduction of peak intensity. Also, the differences observed in the shoulder peak at 1490 cm^{-1} and attributed to CH bending are likely to be exploited to discriminate 'Healthy' from 'TSWV' plants. Except for these slight differences, the average spectra of 'Healthy' and 'TSWV' plants were almost completely superimposable. Therefore, the adoption of Machine Learning techniques appears to be mandatory to enhance the information provided by the Raman spectra toward the detection of the plant infection.

3.3. Multivariate data analysis and spectra classification

The resulting dataset from feature selection consisting of 96 samples and 150 variables was analyzed with both unsupervised and supervised approaches. Preliminary exploratory analysis was conducted by PCA, while PLS-DA was adopted for the classification purpose, due to its proven effectiveness in elaborating spectral data for two-classes discrimination problems (Bevilacqua et al., 2017).

3.3.1. PCA

Multivariate unsupervised methods are generally used to verify whether any clustering exists in a dataset without exploiting any class information during the computation (Granato et al., 2018) and PCA represents the most common first step of any multivariate analysis (Bro and Smilde, 2014). In the present context, the first 8 PCs allowed to explain about 80 % of variance, but neither clusters nor trends could be highlighted among the samples, even following a features selection and a dimensionality reduction. This possibly results from the limited data visualization allowed in 2–3 dimensions only. The adoption of an efficient classification model could be functional also to a better data visualization (Barker and Rayens 2003). Moreover, when we applied PCA to the separate datasets of samples at 3 and 7 dpi, homogeneous distribution of data-points along the first and second PC and no clustering were observed for the two classes, as evident in Supplementary Fig. 2.

3.3.2. PLS-DA

Multivariate classification techniques are ML algorithms addressed to elaborate mathematical models able to predict the attribution of each sample to the correct class, on the basis of the values of a set of features/

measurements (Ballabio and Consonni, 2013). The PLS-DA algorithm relies on PLS regression, which captures the maximum covariance between X (data matrix) and Y (categorical response), where two distinct categories are identified with *a-priori* Y values of 0 and 1, respectively.

To discriminate 'Healthy' from 'TSWV' samples, a binary classification model was built. After selecting the appropriate strategy of repeated data-splitting, the final PLS-DA classification models were performed with five LVs, extracted as described in Section 2.4.

In a preliminary analysis, all the averaged spectra acquired from 'Healthy' and 'TSWV' plants were compared, separately considering the datasets with the samples at 3 and 7 dpi. At both timepoints, the model succeeded in discriminating 'Healthy' from 'TSWV' samples with accuracies of 89±13 % and 82±11 % at 3 and 7 dpi, respectively, as reported in Supplementary Fig. 3 . Since the accuracies at the two acquisition

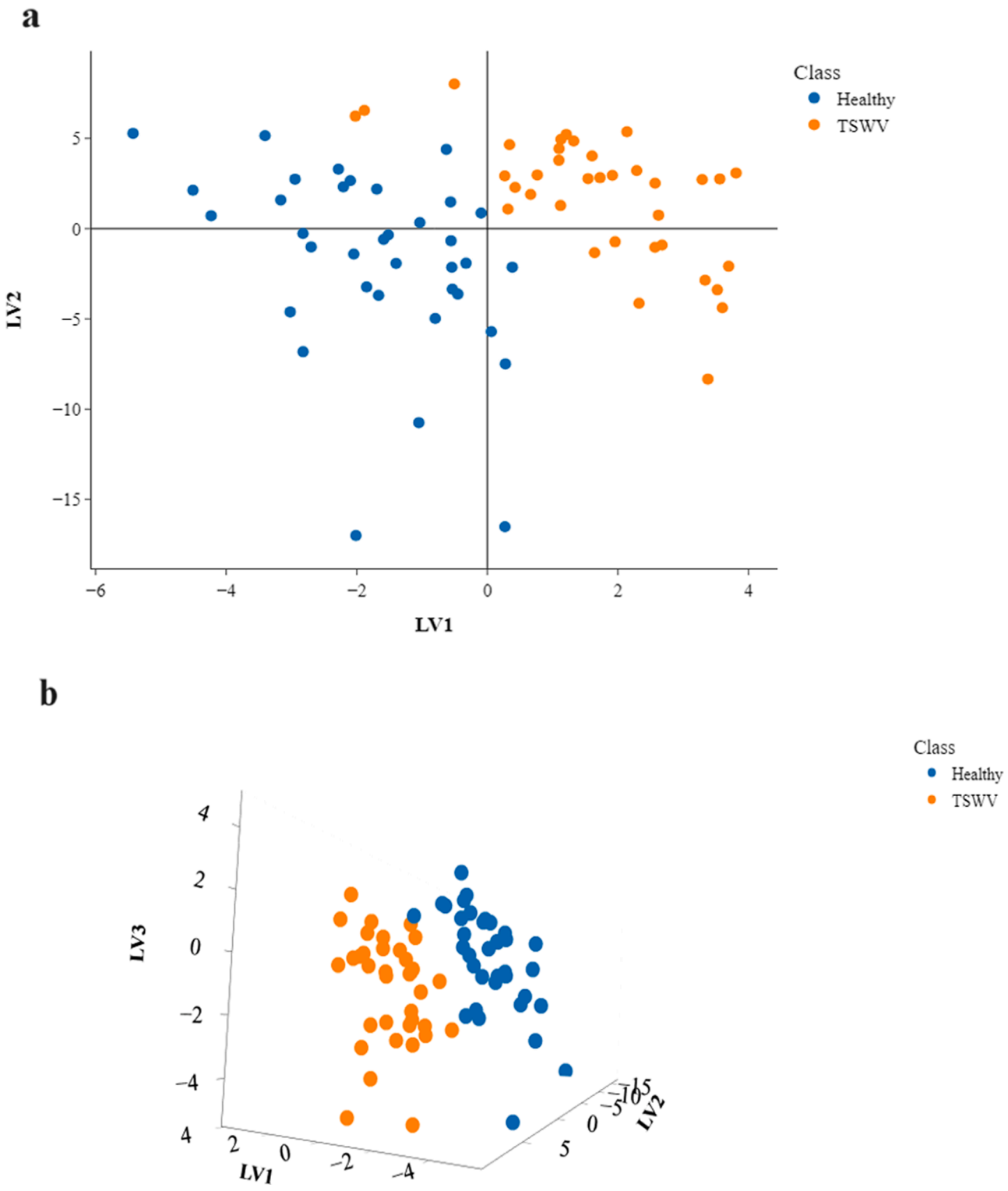


Fig. 3. Graphical outcomes of Partial Least Squares Discriminant Analysis (PLS-DA) related to Model 1: (a) Training set scores on the first two latent variables (LVs); (b) Projection on a selected plane of a three-dimensional representation of training set scores using LV1, LV2, and LV3.

times showed no statistically significant differences (p-value = 0.203), the final model was built by mixing the spectra acquired at the two timepoints, so as to improve the statistical significance of the 'TSWV' group. Fig. 3a shows the diagram reporting the training scores (coordinates of samples in the LVs projection hyperspace) on LV1 and LV2, relative to one of ten models computed ("Model 1″), used as benchmark. Notably, a comprehensive visual representation of all models is prevented by the fact that each single spectrum is sampled in several different models. Using PLS-DA, 'Healthy' samples could be discriminated from 'TSWV' ones, in particular along the bisector traced between the second and fourth quadrants (Fig. 3a). An even better visualization of the two class separation is detectable in the projection of the three-dimensional representation of the LVs hyperplane (Fig. 3b). This figure provides a partial intuitive overview of the initial dataset structure that the PLS-DA algorithm captured.

To evaluate the PLS-DA model performances, each classification model was applied to the corresponding external dataset (test set), and the true response (true class) was compared with the predicted one. The outcome of this procedure, relative to the single Model 1 and reaching an accuracy of 96 %, is summarized in the classification chart and confusion matrix reported in Table 2. Further validation of the present protocol was obtained by applying a comprehensive (i.e., built on all 96 samples) PLS-DA model on an external dataset of 65 new samples. The validation results, expressed by the figures-of-merit and confusion matrix, are also reported in Table 2. The general performance of PLS-DA models on different test dataset remains relatively high with accuracies above 85 %.

The PLS-DA classification criterion for the specific model considered in Table 2 is graphically represented in Fig. 4, where the predicted Y values are plotted against the 24 samples of the test set; the dashed green line represents the discriminant threshold, arbitrarily set to 0.5, even if this value is an adjustable parameter useful to improve the model performance. As mentioned in Section 2.4, to ensure the method robustness, ten different PLS-DA models were developed from random splitting of the samples between validation and test sets. A comprehensive representation of the most significant statistical parameters for all ten PLS-DA models (accuracy, sensitivity, and specificity) is reported in Fig. 5, showing the mean values of the three classification metrics (dashed lines, i.e. accuracy 93 %; sensitivity 94 %; specificity 92 %). Also the figure-of-metrics obtained from a totally independent validation dataset confirmed the model's reliability with accuracy 86 %, sensitivity 90 %, and specificity 82 % (Table 2). These values support the performances of the overall method proposed in this study. The accuracy obtained is greater than that reported with a hand-held Raman spectrometer on symptomatic leaves at a late stage after inoculation (Juárez et al., 2024), but also greater than that achieved with a benchtop instrument on asymptomatic leaves (Mandrile et al., 2019). These results represent a significant advance in the very early detection of TSWV infection, where the accuracy was <75 % before 8 dpi and reached 89 % later (Mandrile et al., 2019). Such improved performance likely relies on the different strategies adopted in both data treatment and machine learning processing. Moreover, having achieved such high performances with a portable instrument makes the technique particularly effective.

To better visualize the statistical distribution of the performance parameters for the ten different classification models, box-and-whisker plots were constructed (Fig. 6), showing consistency and comparability of all ten PLS-DA models.

From the analysis of classification metrics and diagnostic plots here reported, we can assert that RS coupled with adequate spectral preprocessing and PLS-DA interpretation allows a very early detection of TSWV infection on tomato plants, with average levels of accuracy, sensitivity, and specificity ranging between 90 % and 95 %. Therefore, the outcomes of this work are particularly useful in the agri-food field, considering that viral symptoms could be visible by eye two weeks later than by Raman detection (Fig. 1b).

3.4. Validation of the model on resistant tomato F1 hybrid plants

To further validate the protocol and explore the effect of a different susceptibility rate, another virus inoculation experiment was conducted using the commercial tomato F1 hybrid York, carrying the *Sw-5* gene providing resistance to TSWV. Moreover, in this case, two different TSWV isolates were used, the wild type I244 strain, unable to systemically spread in the York hybrid, and the T992 isolate (Ciuffo et al., 2005), overcoming the effect of the *Sw-5* gene. As shown in Supplementary Fig. 4, plants infected by both TSWV isolates showed a decrease in the intensity of peaks in the range 1175–1350 cm^{-1} and 1140–1170 cm^{-1}, related to both chlorophyll and carotenoids, confirming the results

Table 2
(A) Performance of the PLS-DA classification, expressed as figures of merit (precision, recall, specificity, accuracy) and calculated on the test set relative to Model 1 and the cross-validated (k = 4) external validation set, respectively. (B) Confusion matrix calculated on the test set relative to Model 1 and the cross-validated external validation set, respectively.

Test Set relative to Model 1							
		(A) Figures of Merits				(B) Confusion Matrix	
True class	N.	Precision	Recall	Specificity	Accuracy (N. 24)	Predicted Class	
						TSWV	Healthy
TSWV	12	1.00	0.92	1.00	0.96	11	1
Healthy	12					0	12

External Validation with an independent test set							
		(A) Figures of Merits				(B) Confusion Matrix	
True class	N.	Precision	Recall	Specificity	Accuracy (N. 65)	Predicted Class	
						TSWV	Healthy
TSWV	37	0.87	0.90	0.82	0.86	33	4
Healthy	28					5	23

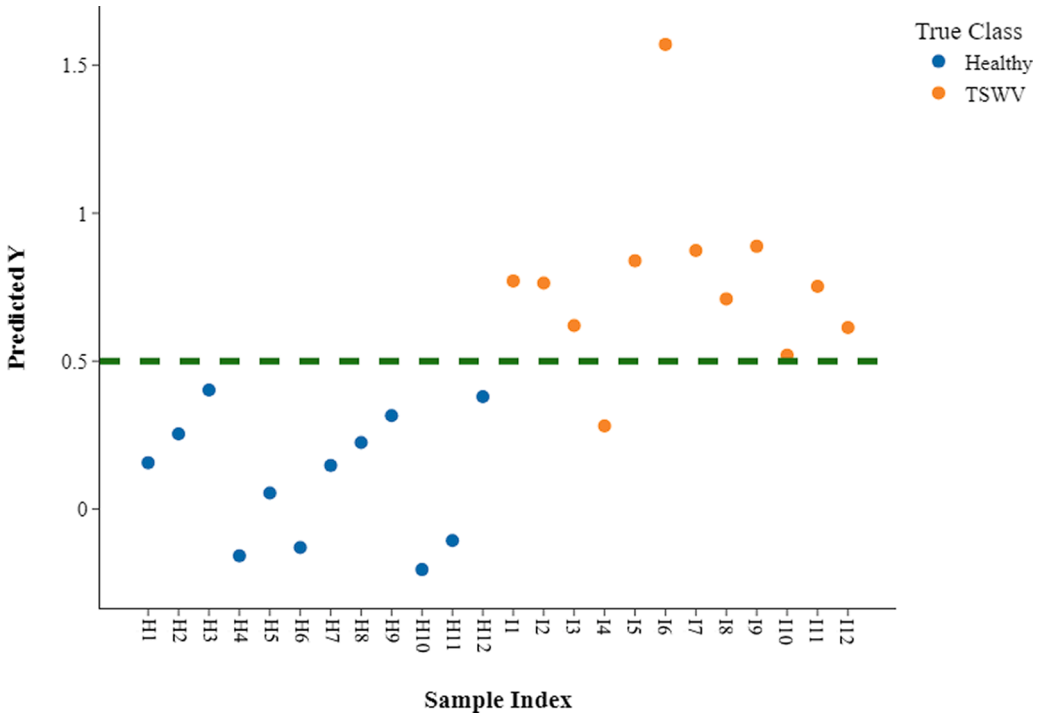


Fig. 4. Partial Least Squares Discriminant Analysis (PLS-DA) analysis in external validation: predicted Y vs. test samples for Model 1. The green dashed line represents the class-threshold. On the X axis, H or I represent Healthy or Infected samples, respectively.

obtained with the susceptible cv. Marmande (Fig. 2 and Supplementary Fig. 5). A decrease in the intensity of peaks at 1224 and 1326 cm^{-1}, associated to chlorophyll (Mandrile et al., 2019) and peaks at 1156 and 1185 cm^{-1}, ascribed to carotenoids, was also reported for TSWV-infected plants at later stages after inoculation (Mandrile et al., 2019; Juárez et al., 2024).

Interestingly, these analyses indicate that the impact of TSWV infection on these two leaf constituents (carotenoids and chlorophyll) occurs since the beginning of the infection process, even in the absence of systemic spread, as is the case for the T992 isolate. Such changes are in agreement with transcriptional data performed on TSWV-infected plants showing a wide downregulation of genes related to photosynthesis and pigment metabolism (Catoni et al., 2009). Since a decrease in carotenoid content also occurred in grapevine plants infected by *Grapevine fanleaf virus* (Mandrile et al., 2022), this can be considered an indication of a common virus-induced perturbation of hormonal responses, in particular abscisic acid and strigolactones, which originate from carotenoid precursors.

Moreover, the considerable signal reduction encountered in TSWV-infected plants for peaks around 1650–1690 cm^{-1}, possibly linked to a reduced protein content (Fig. 2 and Supplementary Fig. 4 and 5) could be attributed to a decrease of the major plant protein, the RuBisCO large subunit, forming over 50 % of leaf proteins in healthy plants (Kawashima and Wildman, 1970). Indeed, chlorosis is one of the first visible symptoms of TSWV, in line with the RuBisCO degradation recently reported in a proteomic analysis of TSWV-infected tomato plants (Gupta et al. 2020).

In order to validate the protocol through a different genotype, the PLS-DA model was applied on the new dataset consisting of 70 spectra (24 spectra for mock-inoculated and 46 spectra for virus-inoculated). The internal validation results, obtained by a repeated cross-validation procedure and expressed by the figures-of-merit and confusion matrix, are reported in Table 3, showing that the final accuracy is equal to 92 %, resulting from the average across the ten repetitions of the cross-validation procedure (chapter 2.4).

4. Conclusions

In the ever-increasing need of triggering more sustainable policies of crop production, the early detection of plant infections by affordable and easy-to-use instruments and methods represents a fundamental target. Unlike benchtop instruments, the less sophisticated portable devices, considerably cheaper and more maneuverable even though less performing, are likely to make practically feasible control strategies that are precluded to most agricultural enterprises.

The present study demonstrates that a simple portable Raman spectrometer is perfectly capable of recognizing asymptomatic TSWV-infected tomato plants two weeks before the infection effects become visible, provided that appropriate Machine Learning algorithms are targeted to the enhancement and interpretation of the minor alteration that the pathogen induces in the Raman spectra. Needless to say, this approach could be adopted to reduce or avoid the use of nucleic acid or protein-based diagnostic assays, allowing to save time and reduce costs and labor in horticultural practices.

The method's performing parameters were measured on independent test samples and were externally validated on two completely separated sets of plants, one of which performed on plants with a different genotype and with different virus isolates. Accuracies outscored 96 %, 86 %, and 92 % respectively, even on plants tested 3 days after inoculation. These repeated impartial test confirmations ensure the method's reliability and prevent overfitting that may arise from data-dependent ML modeling. Lastly, both spectral pre-processing and ML data elaborations were managed rapidly on a standard personal computer and could be applicable in routine controls.

From a broader perspective, the possibility to detect the presence of a pathogen when symptoms are visually undetectable represents an important advantage in the agricultural diagnostic sector, particularly considering that it can be achieved using a portable Raman device, allowing to perform field measurements in real time, with an instrument much less expensive than a corresponding benchtop spectrometer. Moreover, this allows the acquisition of a large number of spectra, as the procedure is not destructive and does not require storage and transfer of

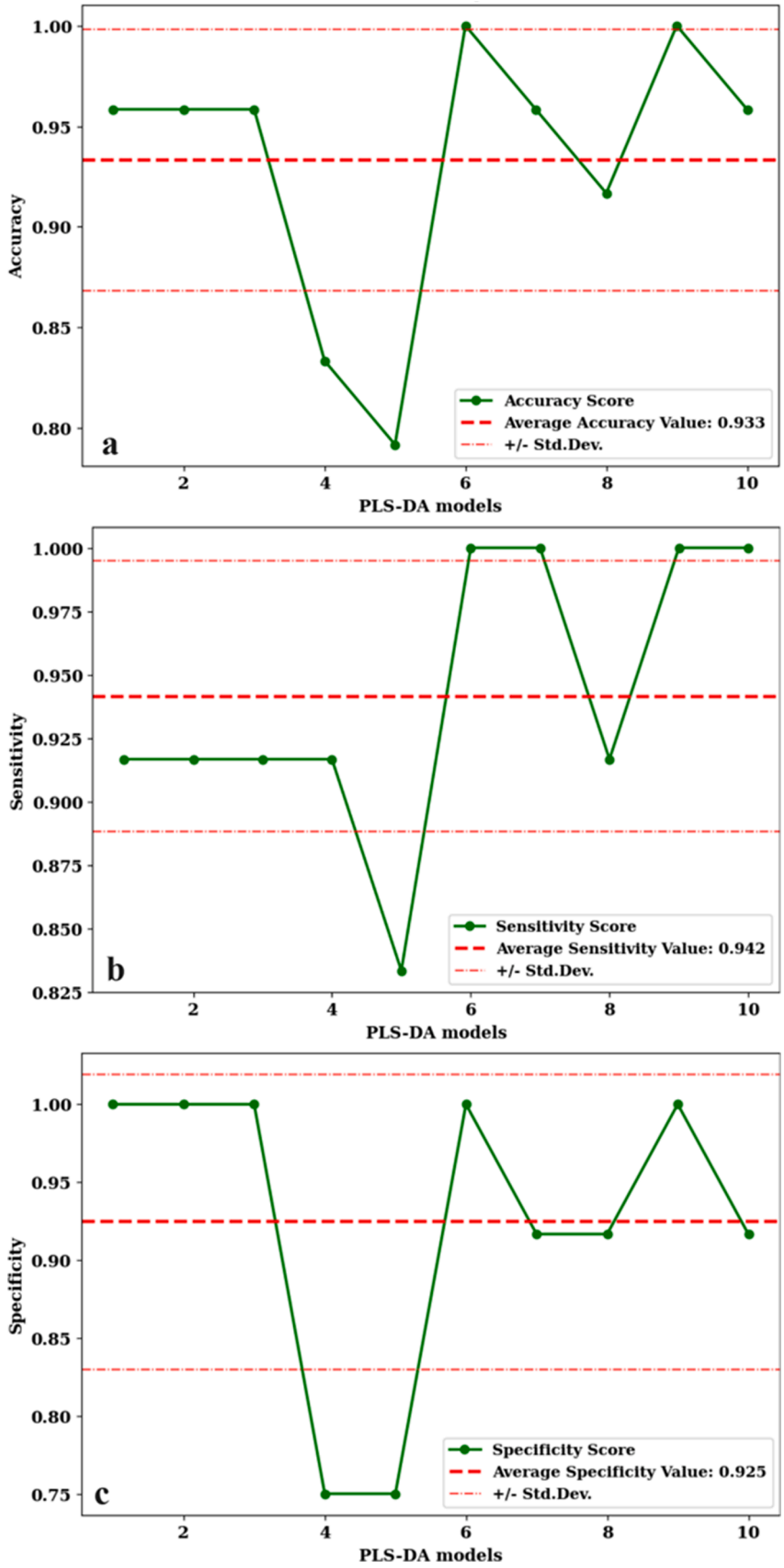


Fig. 5. Figures-of-merit obtained on each of the ten test sets randomly created by the 72:24 data-splitting between a training set (used to build each of the ten models) and the test set (used to validate the corresponding model). (a) Accuracy – fraction of samples (24) assigned to the correct class; (b) Sensitivity = Recall – fraction of TSWV samples (12) assigned to the correct class; (c) Specificity – fraction of healthy samples (12) assigned to the correct class. The dotted orange line represents the mean value obtained from the test sets relative to the ten models.

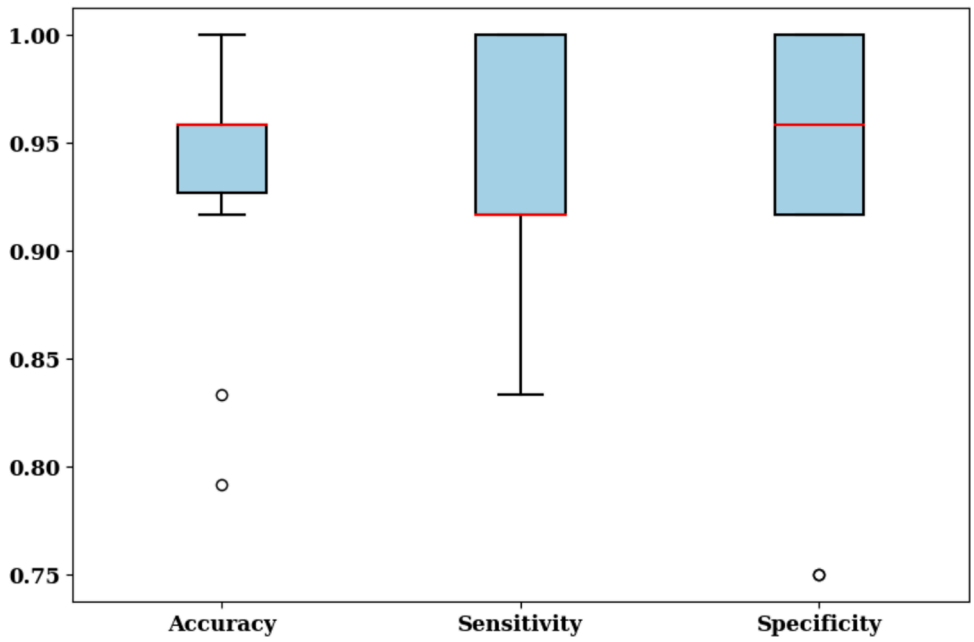


Fig. 6. Analysis of the classification of performance parameters calculated from the ten PLS-DA models randomly generated, represented by box-and-whisker plots. The red line represents the median value for each parameter and the black circles outside the box plots represent anomalous values (outliers). The y-axis represents the domain of the principal figures of merit, in a scale ranging between 0 and 1.

Table 3
(A) Performance of the PLS-DA classification, expressed as average figures of merit (precision, recall, specificity, accuracy) and calculated on the first cross-validated split. (B) Confusion matrix calculated on the test set relative to the first cross-validated split.

Repeated Cross Validation ($n = 10$, $\kappa = 4$) – *External Validation (f1 hybrid york)*							
		(A) Figures of Merits				(B) Confusion Matrix	
True classes	N.	Precision	Recall	Specificity	Accuracy N. 70	Predicted Class TSWV	Healthy
TSWV	46	0.94	0.94	0.88	0.91	43	3
Healthy	24					3	21

material, thus enabling to increase the accuracy and reliability of the technique.

However, for a more general application of such proximal sensor-based diagnostic techniques, deeper elucidations of the biological pathways altered during plant pathogen infection are sought, investigating and possibly verifying a commonality of responses induced by different isolates of a pathogen on different cultivars.

In a possible scenario of practical implementation of RS for diagnostic purposes of plant viruses, the most interesting example consists in breeding for resistance, which is currently the prevalent defense strategy against plant viruses. For tomatoes, due to the long domestication process of this crop (Ferrero et al., 2020), the majority of resistance genes against viruses are identified in wild germplasm and introduced into cultivated genotypes through hybrid breeding. Specifically, the most relevant disease management strategy for TSWV relies on the *Sw-5* locus, providing durable resistance against different tospovirus species and even against strains from diverse geographic locations (Turina et al., 2016). However, the frequent onset of resistance-breaking isolates (Aramburu et al., 2010; Latham and Jones, 1998) requires a continuous search of new sources of resistance, pushing towards the use of rapid diagnostic tools during breeding.

Here, we approached this issue testing the performances of a portable RS instrument on an ancient commercial tomato cultivar (cv. Marmande) susceptible to the majority of tomato viruses, including TSWV (Peiró et al., 2014); we expanded our analysis using a commercial tomato hybrid line carrying the *Sw-5* locus, testing two different TSWV isolates, a wild type one (I244) unable to systemically spread in this hybrid and the T992 isolate, obtaining rather high accuracy values. Nonetheless, a limitation of this diagnostic tool could result from its application on plants with different genetic background originating from crosses with wild relatives of domestic tomatoes to introgress new resistance traits (Qi et al., 2021), ultimately leading to uneven perturbations of the host basal metabolome and of altered metabolic responses following virus attack.

Indeed, substantial gaps in the knowledge of the biological pathways tackled during virus infection in cultivated vegetables with different genetic backgrounds and how such genetic variability modifies the response to pathogens have to be filled. A combinatorial approach integrating -omics studies with proximal sensor techniques could help to comprehensively elucidate the complex array of responses occurring during plant-pathogen interactions, for example combining transcriptomics and metabolomics studies of plants infected by TSWV (Lv et al., 2023; Liang et al., 2024) with RS techniques.

Funding

This study was supported by the following projects: (1) “Agritech National Research Center’’ funded by the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022), (2) “ON Foods - Research and innovation network on food and nutrition Sustainability, Safety and Security – Working ON Foods” (code PE00000003) within the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment 1.3 - Call for proposals

No 341 of 15 March 2022 of Italian MUR, funded by the European Union – NextGenerationEU, and (3) NUTRAGE, CNR project FOE-2021 DBA. AD005.225. (4) CSB was funded by the Fondazione Cassa di Risparmio di Torino, Project N. 2022.AI590.709. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version.

CRediT authorship contribution statement

Ciro Orecchio: Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Camilla Sacco Botto:** Methodology, Investigation. **Eugenio Alladio:** Supervision, Data curation. **Chiara D'Errico:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Marco Vincenti:** Writing – review & editing, Supervision, Data curation, Conceptualization. **Emanuela Noris:** Writing – review & editing, Writing – original draft, Supervision, 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.

Acknowledgements

IPSP-CNR authors are grateful to Luca Bordone and Elena Zocca for technical support in greenhouse experiments; Danila Fiaschi, Concetta Mottura, and José Saporita for administrative support, and Daniele Marian for purchasing procedures. All Authors give thanks to the Neutravel project and Cooperativa Alice for enabling them to complete the Raman spectra acquisition.

University of Turin authors acknowledge general support from Project CH4.0 under the MUR program "Dipartimenti di Eccellenza 2023–2027″ (CUP: D13C22003520001).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.stress.2024.100732.

Data availability

The data will be made available upon reasonable request to the corresponding authors.

References

Altangerel, N., Ariunbold, G.O., Gorman, C., et al., 2017. In vivo diagnostics of early abiotic plant stress response via Raman spectroscopy. Proc. Natl. Acad. Sci. U.S.A, 114, 3393–3396. https://doi.org/10.1073/pnas.1701328114.

Aramburu, J., Galipienso, L., Soler, S., Lopez, C., 2010. Characterization of ' Tomato spotted wilt virus isolates that overcome the Sw-5 resistance gene in tomato and fitness assays. Phytopathol. Mediterr. 49, 342–351.

Baek, S.J., Park, A., Ahn, Y.J., Choo, J., 2015. Baseline correction using asymmetrically reweighted penalized least squares smoothing. Analyst 140, 250–257. https://doi.org/10.1039/c4an01061b.

Ballabio, D., Consonni, V., 2013. Classification tools in chemistry. Part 1: linear models. PLS-DA. Anal. Methods 5, 3790–3798. https://doi.org/10.1039/c3ay40582f.

Baratto, C., Ambrosio, G., Faglia, G., Turina, M., 2022. Early detection of esca disease in asymptomatic vines by raman spectroscopy. IEEE Sens. J. 22 (23), 23286–23292. https://doi.org/10.1109/JSEN.2022.3211616.

Barker, M., Rayens, W., 2003. Partial least squares for discrimination. J. Chemom. 17, 166–173. https://doi.org/10.1002/cem.785.

Bevilacqua, M., Bro, R., Marini, F., et al., 2017. Recent chemometrics advances for foodomics. TrAC - Trends in Anal. Chem. 96, 42–51. https://doi.org/10.1016/j.trac.2017.08.011.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6, 2812–2831. https://doi.org/10.1039/c3ay41907j.

Catoni, M., Miozzi, L., Fiorilli, V., Lanfranco, L., Accotto, G.P., 2009. Comparative analysis of expression profiles in shoots and roots of tomato systemically infected by *Tomato spotted wilt virus* reveals organ-specific transcriptional responses. Mol. Plant-Microbe Interact. 22, 1504–1513. https://doi.org/10.1094/MPMI-22-12-1504.

Chinnaiah, S., Varagur Ganesan, M., Sevugapperumal, N., et al., 2022. A sequel study on the occurrence of Tomato spotted wilt virus (TSWV) in cut-chrysanthemum by DAS-ELISA using recombinant nucleocapsid protein to produce polyclonal antiserum. J. Virol. Methods 300, 114410. https://doi.org/10.1016/j.jviromet.2021.114410.

Chong, I.G., Jun, C.H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. 78, 103–112.. https://doi.org/10.1016/j.chemolab.2004.12.011.

Ciuffo, M., Finetti-Sialer, M.M., Gallitelli, D., Turina, M., 2005. First report in Italy of a resistance-breaking strain of *Tomato spotted wilt virus* infecting tomato cultivars carrying the *Sw5* resistance gene. Plant Pathol. 54, 564. https://doi.org/10.1111/j.1365-3059.2005.01203.x.

Egging, V., Nguyen, J., Kurouski, D., 2018. Detection and identification of fungal infections in intact wheat and sorghum grain using a hand-held raman spectrometer. Anal. Chem. 90, 8616–8621. https://doi.org/10.1021/acs.analchem.8b01863.

Farber, C., Kurouski, D., 2018. Detection and identification of plant pathogens on maize kernels with a hand-held raman spectrometer. Anal. Chem. 90, 3009–3012. https://doi.org/10.1021/acs.analchem.8b00222.

Farber, C., Sanchez, L., Rizevsky, S., et al., 2020. Raman spectroscopy enables non-invasive identification of peanut genotypes and value-added traits. Sci. Rep. 10, 7730. https://doi.org/10.1038/s41598-020-64730-w.

Ferrero, V., Baeten, L., Blanco-Sánchez, L., Planelló, R., Díaz-Pendón, J.A., Rodríguez-Echeverría, S., Haegeman, A., de la Peña, E., 2020. Complex patterns in tolerance and resistance to pests and diseases underpin the domestication of tomato. New. Phytol. 226, 254–266. https://doi.org/10.1111/nph.16353.

Florián-Huamán, J., Cruz-Tirado, J.P., Fernandes Barbin, D., Siche, R., 2022. Detection of nutshells in cumin powder using NIR hyperspectral imaging and chemometrics tools. J. Food Compos. Anal. 108. https://doi.org/10.1016/j.jfca.2022.104407.

Frenicha, A.G., Jouan-Rimbaudb, D., Massart, D.L., et al., 1995. Wavelength selection method for multicomponent spectrophotometric determinations using partial least squares. Analyst 120, 2787–2792. https://doi.org/10.1039/AN9952002787.

Gao, S., Wu, J., 2022. Detection of tomato spotted wilt virus (TSWV) infection in plants using DAS-ELISA and Dot-ELISA. In: Wang, A., Li, Y. (Eds.), Plant Virology : Methods in Molecular Biology. Springer US, New York, NY, pp. 253–261. https://doi.org/10.1007/978-1-0716-1835-6_24.

Górski, Ł., Jakubowska, M., 2023. A graphical user interface for arPLS baseline correction. Chemometr. Intelligent Labor. Syst. 238, 104848. https://doi.org/10.1016/j.chemolab.2023.104848.

Granato, D., Santos, J.S., Escher, G.B., et al., 2018. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective. Trends Food Sci. Technol. 72, 83–90. https://doi.org/10.1016/J.TIFS.2017.12.006.

Gupta, R., Min, C.W., Kim, S.W., et al., 2020. A TMT-based quantitative proteome analysis to elucidate the TSWV induced signaling cascade in susceptible and resistant cultivars of Solanum lycopersicum. Plants 9, 290. https://doi.org/10.3390/plants9030290.

Iturralde Martinez, J.F., Rosa, C., 2023. Reverse transcriptase recombinase polymerase amplification for detection of tomato spotted wilt orthotospovirus from crude plant extracts. Sci. Rep. 13, 9024. https://doi.org/10.1038/s41598-023-35343-w.

Jehlička, J., Culka, A., Košek, F., 2017. Obtaining Raman spectra of minerals and carbonaceous matter using a portable sequentially shifted excitation Raman spectrometer – a few examples. J. Raman Spectrosc. 48, 1583–1589. https://doi.org/10.1002/jrs.5105.

Juárez, I.D., Steczkowski, M.X., Chinnaiah, S., et al., 2024. Using Raman spectroscopy for early detection of resistance-breaking strains of tomato spotted wilt orthotospovirus in tomatoes. Front. Plant Sci. 14, 1283399. https://doi.org/10.3389/fpls.2023.1283399.

Kawashima, N., Wildman, S.G., 1970. A model of the subunit structure of fraction I protein. Biochem. Biophys. Res. Comm 41, 1463–1468. https://doi.org/10.1016/0006-291X(70)90551-6.

Kong, L., Liu, T., Qiu1, Honglin, Yu, X., Wang, X., Huang, Z., Huang, M., 2024. Early diagnosis of citrus Huanglongbing by Raman spectroscopy and machine learning. Laser Phys. Lett. 21, 015701. https://doi.org/10.1088/1612-202X/ad1097.

Krimmer, M., Farber, C., Kurouski, D., 2019. Rapid and Noninvasive Typing and Assessment of Nutrient Content of Maize Kernels Using a Handheld Raman Spectrometer. ACS. Omega 4, 16330–16335. https://doi.org/10.1021/acsomega.9b01661.

Latham, L.J., Jones, R.A.C., 1998. Selection of resistance breaking strains of Tomato spotted wilt tospovirus. Ann. Appl. Biol. 133, 385–402. https://doi.org/10.1111/j.1744-7348.1998.tb05838.x.

Li, P., Ma, J., Zhong, N., 2021. Raman spectroscopy combined with support vector regression and variable selection method for accurately predicting salmon fillets storage time. Optik. (Stuttg) 247, 167879. https://doi.org/10.1016/j.ijleo.2021.167879.

Li, Z., Paul, R., Ba Tis, T., et al., 2019. Non-invasive plant disease diagnostics enabled by smartphone-based fingerprinting of leaf volatiles. Nat. Plants 5, 856–866. https://doi.org/10.1038/s41477-019-0476-y.

Liang, Y., Wan, J., Zhang, X., Li, K., Su, J., Gui, M., Li, Y., Liu, Y., 2024. Comprehensive phytohormone metabolomic and transcriptomic analysis of tobacco (*Nicotiana tabacum*) infected by tomato spotted wilt virus (TSWV). Virus Res. 342, 199334. https://doi.org/10.1016/j.virusres.2024.199334.

Lv, J., Deng, M., Li, Z., Zhu, H., Wang, Z., Yue, Y., Yang, Z., Xu, J., Jiang, S., Zhao, W., Li, J., Zhao, K., 2023. Integrative analysis of the transcriptome and metabolome reveals the response mechanism to tomato spotted wilt virus. Hort. Plant J. 9 (5), 958–970. https://doi.org/10.1016/j.hpj.2022.12.008.

Mandrile, L., Giovannozzi, A.M., Durbiano, F., et al., 2018. Rapid and sensitive detection of pyrimethanil residues on pome fruits by Surface Enhanced Raman Scattering. Food Chem. 244, 16–24. https://doi.org/10.1016/j.foodchem.2017.10.003.

Mandrile, L., Rotunno, S., Miozzi, L., et al., 2019. Nondestructive raman spectroscopy as a tool for early detection and discrimination of the infection of tomato plants by two economically important viruses. Anal. Chem. 91, 9025–9031. https://doi.org/10.1021/acs.analchem.9b01323.

Mandrile, L., D'Errico, C., Nuzzo, F., et al., 2022. Raman Spectroscopy applications in grapevine: metabolic analysis of plants infected by two different viruses. Front. Plant Sci. 13, 917226. https://doi.org/10.3389/fpls.2022.917226.

Martyna, A., Menżyk, A., Damin, A., et al., 2020. Improving discrimination of Raman spectra by optimising preprocessing strategies on the basis of the ability to refine the relationship between variance components. Chemometr. Intell. Lab. Syst. 202, 104029. https://doi.org/10.1016/j.chemolab.2020.104029.

Menges F. (2001) Spectragryph - optical spectroscopy software - version 1.2.16.1, 2024. http://www.effemm2.de/spectragryph/. Accessed 19 Mar 2024.

Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in partial least squares regression. Chemometr. Intell. Lab. Syst. 118, 62–69. https://doi.org/10.1016/j.chemolab.2012.07.010.

Nekvapil, F., Brezestean, I., Barchewitz, D., et al., 2018. Citrus fruits freshness assessment using Raman spectroscopy. Food Chem. 242, 560–567. https://doi.org/10.1016/j.foodchem.2017.09.105.

Nguyen, C., Sagan, V., Maimaitiyiming, M., et al., 2021. Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning. Sensors 21 (3), 742. https://doi.org/10.3390/s21030742.

Payne, W.Z., Kurouski, D., 2021. Raman-Based diagnostics of biotic and abiotic stresses in plants. A Review. Front. Plant Sci. 11, 616672. https://doi.org/10.3389/fpls.2020.616672.

Peiró, A., Cañizares, M.C., Rubio, L., López, C., Moriones, E., Aramburu, J., Sánchez-Navarro, J., 2014. The movement protein (NSm) of Tomato spotted wilt virus is the avirulence determinant in the tomato *Sw-5* gene-based resistance. Mol. Plant Pathol. 15, 802–813. https://doi.org/10.1111/mpp.12142.

Qi, S., Zhang, S., Islam, M.M., El-Sappah, A.H., Zhang, F., Liang, Y., 2021. Natural Resources Resistance to *Tomato Spotted Wilt Virus* (TSWV) in Tomato (*Solanum lycopersicum*). Int. J. Mol. Sci. 22, 10978. https://doi.org/10.3390/ijms222010978.

Roberts, C.A., Dietzgen, R.G., Heelan, L.A., Maclean, D.J., 2000. Real-time RT-PCR fluorescent detection of tomato spotted wilt virus. J. Virol. Methods 88, 1–8. https://doi.org/10.1016/s0166-0934(00)00156-7.

Ruark-Seward, C.L., Bonville, B., Kennedy, G., Rasmussen, D.A., 2020. Evolutionary dynamics of Tomato spotted wilt virus within and between alternate plant hosts and thrips. Sci. Rep. 10, 15797. https://doi.org/10.1038/s41598-020-72691-3.

Saletnik, A., Saletnik, B., Zagula, G., Puchalski, C., 2024. Raman spectroscopy for plant disease detection in next-generation agriculture. Sustainability 16, 5474. https://doi.org/10.3390/su16135474.

Sanchez, L., Ermolenkov, A., Biswas, S., et al., 2020a. Raman spectroscopy enables non-invasive and confirmatory diagnostics of salinity stresses, nitrogen, phosphorus, and potassium deficiencies in rice. Front. Plant Sci. 11, 573321. https://doi.org/10.3389/fpls.2020.573321.

Sanchez, L., Ermolenkov, A., Tang, X.T., et al., 2020b. Non-invasive diagnostics of Liberibacter disease on tomatoes using a hand-held Raman spectrometer. Planta 251 (3), 64. https://doi.org/10.1007/s00425-020-03359-5.

Sanchez, L., Farber, C., Lei, J., et al., 2019a. Noninvasive and nondestructive detection of cowpea bruchid within cowpea seeds with a hand-held raman spectrometer. Anal. Chem. 91, 1733–1737. https://doi.org/10.1021/acs.analchem.8b05555.

Sanchez, L., Pant, S., Xing, Z., et al., 2019b. Rapid and noninvasive diagnostics of Huanglongbing and nutrient deficits on citrus trees with a handheld Raman spectrometer. Anal. Bioanal. Chem. 411, 3125–3133. https://doi.org/10.1007/s00216-019-01776-4.

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

Scholthof, K.B.G., Adkins, S., Czosnek, H., et al., 2011. Top 10 plant viruses in molecular plant pathology. Mol. Plant Pathol. 12, 938–954. https://doi.org/10.1111/j.1364-3703.2011.00752.x.

Turina, M., Kormelink, R., RO, Resende, 2016. Resistance to tospoviruses in vegetable crops: epidemiological and molecular aspects. Annu. Rev. Phytopathol. 54, 347–371. https://doi.org/10.1146/annurev-phyto-080615-095843.

Ullman, D.E., Meideros, R., Campbell, L.R., et al., 2002. Thrips as vectors of tospoviruses. Adv. Bot. Res. 36, 113–140. https://doi.org/10.1016/S0065-2296(02)36061-0.

Vallejo-Pérez, M.R., Sosa-Herrera, J.A., Navarro-Contreras, H.R., et al., 2021. Raman spectroscopy and machine-learning for early detection of bacterial canker of tomato: the asymptomatic disease condition. Plants 10, 1542. https://doi.org/10.3390/plants10081542.

Van Rossum G. (1995) Centrum voor Wiskunde en Informatica Python reference manual.

Yun, Y.H., Li, H.D., Deng, B.C., Cao, D.S., 2019. An overview of variable selection methods in multivariate analysis of near-infrared spectra. TrAC - Trends Analyt. Chem. 113, 102–115. https://doi.org/10.1016/j.trac.2019.01.018.