Surface Enhanced Raman Scattering (SERS) Spectroscopy Combined With Chemical Imaging Analysis For The Early Detection of Apple Valsa Canker

Shiyan Fang  
Northwest A&F University: Northwest Agriculture and Forestry University  
https://orcid.org/0000-0001-5531-2798

Junmeng Li  
Northwest A&F University: Northwest Agriculture and Forestry University

Yan Wang  
Northwest A&F University: Northwest Agriculture and Forestry University

Yanru Zhao  
Northwest A&F University: Northwest Agriculture and Forestry University

Keqiang Yu (✉ keqiang_yu@nwafu.edu.cn)  
Northwest A&F University: Northwest Agriculture and Forestry University  
https://orcid.org/0000-0002-0034-0721

Research

Keywords: Apple Valsa canker, Early detection, Surface enhanced Raman scattering, Chemical imaging, Machine learning

DOI: https://doi.org/10.21203/rs.3.rs-761372/v1

License: ☇ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Apple Valsa Canker (AVC) with early incubation characteristics is a severe apple tree disease. Therefore, early detection of the infected trees is necessary to prevent the rapid development of the disease. Surface enhanced Raman Scattering (SERS) spectroscopy is a promising technique that simplifies detection procedures and reduces detection time. Meanwhile, SERS enhance signals at low laser powers and suppress biological fluorescence. In this study, the early detection of the AVC disease was carried out by combining SERS spectroscopy with the chemometrics methods and machine learning algorithms, and then chemical distribution imaging was successfully applied to the analysis of disease dynamics.

**Results:** Firstly, the microstructure, UV-Vis spectrum, and Raman spectrum of SERS metallic nano-substrates were proved to investigate the enhancement effects of the synthesized AgNPs. Secondly, the multiple spectral baseline correction (MSBC), the asymmetric least squares (AsLS), and the adaptive iterative reweighted penalized least squares (air-PLS) were adopted to eliminate the disturbances of the baseline offset. The correlation analysis method was employed to identify the best baseline correction algorithm, which was the air-PLS algorithm herein. Meanwhile, principal component analysis (PCA) was used to perform clustering analysis based on the healthy, early disease, and late disease sample datasets, demonstrating obvious clustering effects. After that, optimal spectral variables were selected to build machine learning models to detect AVC disease, incorporating the BP-ANN, ELM, RForest, and LS-SVM algorithms. The accuracy of these models was above 90%, showing excellent discriminant performance. Finally, SERS chemical imaging provided the spatiotemporal dynamic characteristics of changes in the cellulose and lignin of the phloem disease-health junction under AVC stress. The results suggested that cellulose and lignin in the cell walls of infected tissues reduced significantly.

**Conclusions:** SERS spectroscopy combining with chemical imaging analysis for early detection of the AVC disease was considered feasible and promising. This study provided a practical method for the rapid diagnosis of apple orchard diseases.

Introduction

Apple Valsa Canker (AVC), caused by fungus *valsa mali*, is a severe apple tree disease resulting in serious economic losses in Southeast Asia and China [1]. Previous studies have shown that the AVC in symptomless tissues was found in more than 50% of apple orchards surveyed by Zang et al. [2]. It caused a significant reduction in the fruit yield. Commonly, AVC is mainly characterized by canker, infected tissue softening, outflowed light brown water stain, sunken or cracked on trunks at the early infected stage. The fungal pathogen mainly infected the subcutaneous phloem through the wounded bark tissue at the initial infected stage. After successful infection, fungus hypha colonized the phloem tissues, leading to severe tissue cell death [3]. Research by plant protection experts has proved that the fungus *valsa mali* can survive in weak and dead tissues of the apple trees for more than one year before producing visible symptoms [4]. Namely, the pathogen is characterized by early incubation. When visible
symptoms appeared, it is difficult to containment the AVC spread through conventional measures such as spraying fungicides, manually removing the diseased areas, and pruning the dead branches. So far, there are no effective methods for AVC treatment due to the complicated pathogenic mechanism. Thus early detection of the infected trees is necessary to prevent the rapid development of the disease.

In addition to direct isolation of pathogens, the molecular biology methods for detecting AVC included Enzyme-Linked ImmunoSorbent Assay (ELISA) and Polymerase Chain Reaction (PCR) [5, 6]. ELISA kits had been widely utilized for their low cost but were ineffective at detecting symptomless tissue [7]. PCR technology is an effective detection method. Zang et al. [2] developed a nested PCR assay to detect the presence of valsa mali in apple trees and achieved an accuracy of 64.7%. However, DNA derived from the woody plant tissues contains PCR inhibiting compounds and can interfere with PCR reaction resulting in false negatives [8]. In addition, a well-equipped laboratory and experienced personnel are required, which is not feasible for on-site detection [9]. Therefore, due to these limitations, it is necessary to search for a fast, non-destructive and economical method that would enable highly accurate detection of AVC.

Reported studies have demonstrated that advanced non-invasive measuring technologies, such as RGB image processing [10], dielectric spectrum [11], laser scanning [12], and spectroscopic methods have the potential for diagnosing tree diseases in recent years. However, the advantages of spectroscopy over other novel techniques can be attributed to simplicity, rapidity, and affordability, which makes it indispensable in tree disease detection. Raman spectroscopy (RS) is a non-invasive, rapid, and high throughput spectroscopic technique [13]. Raman shift is only related to the vibration frequency of the molecular functional group and is not to the incident light. Therefore, the Raman "fingerprint" of each sample is unique [14]. Significantly, RS is a well-established chemical analysis method that provided essential information related to the biochemical composition of the tree tissue cell, such as protein, polysaccharide, and lipid. Both symptomatic and asymptomatic, these biochemical compositions were significantly different between diseased and healthy tissue. These compositions changes were reflected in Raman shifts or intensity change of specific Raman bands assigned to those molecules. Therefore, RS provided an accessible way to identify subtle changes in the molecular compounds, which offered theoretical evidence for detecting tree diseases. Perez et al. [15] investigated the application of RS combined with statistical analysis for detecting citrus Huanglongbing (HLB) disease in the field, and a good result was obtained with an overall classification accuracy of about 89.2%. Sanches et al. [16] readily distinguished between healthy and early-HLB citrus trees using a handheld Raman system and achieved an accuracy of 94%. In the following study, Sanches et al. [17] also demonstrated that utilizing a handheld Raman spectrometer in combination with chemometric analyses enables detection and identification of the secondary disease on HLB-infected orange trees. Those researches indicated RS technique combined with chemometric methods have great potential for detecting diseased trees.

However, RS is frequently interfered with by fluorescence caused by chromophores in plant tissue, and compositional changes under disease stress may lead to Raman band broadening or drift [18, 19]. This drawback may lead to significant deviations in the biochemical composition analysis of RS data. Surface enhanced Raman scattering (SERS) spectroscopy, based on the improvement of traditional RS, uses
certain metallic nano-substrates such as gold or silver nanoparticles (AgNPs) to enhance signals under low laser powers, which maximizes fluorescence suppression. Meanwhile, the Raman system combined with the micro-imaging technology allows for scanning micron-scale Raman collection points (e.g., one-micron pixel) [20], which offers chemical information on the constituents at high spatial resolution in situ. Qin et al. [21] developed a Raman chemical imaging system to visualize the internal distribution of lycopene in postharvest tomato, and established a Raman chemical image to visualize the spatial distribution of lycopene at different stages of maturity. Yang et al. [22] used a Raman imaging system to detect the spatial distribution of chemical components in maize seeds.

Therefore, this study aimed to develop a fast, non-invasive, and in situ diagnosis method for detecting AVC at early infection stages using SERS combined with micro-imaging technology. The main objectives are as following: 1. Optimizing experimental conditions for obtaining valid SERS micro-imaging data, including Synthesis and SERS AgNPs characterization; 2. Establishing optimal discriminative models for detecting AVC in early infection stages based on machine learning algorithms; 3. Generating micro-distribution maps of cellulose and lignin at the disease-health junction of the tree phloem tissues.

**Materials And Methods**

2.1 Fungal culture and sample inoculation

The fungus *valsa mali* stored at -80°C in an ultra-low temperature refrigerator were inoculated onto potato dextrose agar (PDA) medium. The two-year-old apple branches (*Malus domestica* cv. Fuji) were collected from the Economic Tree Garden of Northwest A&F University. The selected branches were pruned into 15 cm segments, and the surface of the branches was disinfected with 75% alcohol for 15 min, and then cleaned with sterile water three times until there was no odor. The ends of the branches were sealed with wet skimmed cotton to keep them fresh, followed by punching holes in the branches with a hole puncher (hole diameter 5 mm). The activated *valsa mali* fungus were inoculated on the wounds of apple branches with 2 points on each branch. After inoculation, the branches were transferred to a 25°C incubator for further incubation.

2.2 Synthesis and SERS AgNPs characterization

In the present research, AgNPs were synthesized by using the Lee-Meisel method. The synthesis steps were as follows: AgNO$_3$ (36 mg) was dissolved in 200 ml of ultrapure water and boiled quickly. A solution of 1 wt.% trisodium citrate (6 mL) was charged to the reaction solution and was held on boiling for 25 min accompanied by stirring at 200 rpm. After cooling to room temperature, pour the AgNPs solution into a centrifuge tube and store it away from light. The chemical reaction equation is as follows:

$$4Ag^+ + C_6H_5O_7Na_3 + 2H_2O = 4Ag + C_6H_5O_2Na_3 + 3Na^+ + O_2$$

Subsequently, the prepared AgNPs were characterized to verify their validity. The morphology of the AgNPs was measured by Tecnai G2 transmission electron microscopy (FEI Inc., Hillsboro, OR, USA). The UV-Vis absorption spectra of the AgNPs were measured using Lambda 35 Spectrophotometer.
The Raman spectra of the AgNPs were collected by DXR3xi Raman micro-imaging spectrometer (ThermoFisher Inc., Waltham, MA, USA).

### 2.3 SERS spectroscopy acquisition

First, each sample placed on a glass slide was dripped with the AgNPs. Then, each sample was placed on the automatic stage and was aligned with a Raman laser using a 10x/0.25NA magnification objective lens for SERS imaging collection using a DXR3xi Raman micro-imaging system (ThermoFisher Inc., Waltham, MA, USA). Specific parameters were as follows: the excitation wavelength was 785 nm; the collected spectral range was 300–3000 shift/cm$^{-1}$; the laser intensity was 2.6 mW; the exposure time was 0.00357 s (280 Hz); the scanning times was 30.

For spectral imaging in the $x$ and $y$ directions, the samples were scanned point by point in 2 µm steps. It should be noted that no destructive effects of the laser on the samples were observed. Routinely, before starting the Raman measurements, automatic calibration of the instrument was performed. The data acquisition software OMNICxi v1.6 was used to adjust the acquisition parameters.

### 2.4 Spectral data processing and analysis

#### 2.4.1 Spectra preprocessing

Background noises and baselines were generated during the acquisition of the SERS spectra, which seriously impaired the interpretability of the spectra. Meanwhile, these noises and baselines would also reduce the simplicity and robustness of a calibration model built on these spectra. Therefore, selecting the optimal pretreatment method was necessary for improving the spectral quality. In the present research, spectral curves were first extracted for each pixel point of the imaging data before spectra preprocessing. Then, the spectral data were preprocessed with three algorithms to eliminate noise and correct the baseline background. These three algorithms included the multiple spectral baseline correction (MSBC), the asymmetric least squares (AsLS), and the adaptive iterative reweighted penalized least squares (air-PLS). Subsequently, the advantages and disadvantages of the three algorithms were compared using the correlation analysis method.

The AsLS method, proposed by Eilers et al. [23, 24] in 2003, is a classical baseline correction algorithm that combined a smoother with the asymmetric weighting of deviations from the smoothed trend to form an effective baseline estimation method. The MSBC method, proposed by Peng et al. [25] in 2010, is an improved approach based on the ASLS algorithm. The MSBC method learns baselines that perform well on the corresponding spectra and then "co-regularize" the selection by correcting inconsistencies between the spectra. Air-PLS is an improvement approach based on the weighting of the original model by the weighted least squares method. By meaning the iterative regression, the light environment is automatically subtracted, and the background is deducted [26].

#### 2.4.2 Optimal variables selection and dimension reduction
Multivariate calibration methods in chemometrics aim to construct relationships between variables and properties of interest to make a classification model. However, with the redundant spectral variables, data usually included some noise and unnecessary information, which rendered the unreliable predictive properties. The methods have been used to address these problems, which were optimal variables selection and dimension reduction.

In the present research, principal component analysis (PCA) can replace the original variables with a few principal components with significant deviation to reduce the original high-dimensional variable space [27]. In addition, competitive adaptive reweighted sampling (CARS) and random frog (RFrog) algorithms were used in combination to select the optimal variables associated with the predicted properties while excluding the interference of unrelated variables. The CARS algorithm used exponentially decreasing function (EDF) as a selection strategy to select critical variables based on adaptive reweighted sampling competitively [28, 29]. The RFrog algorithm calculated the selection probability of each variable by moving across trans-dimensions between models, enabling the search for the optimal variable [30].

### 2.4.3 Classification models

BP artificial neural network (BP-ANN) [31] is the most classical and successful neural network commonly utilized for nonlinear fitting and pattern recognition. BP-ANN is a one-way multi-layer feedforward network composed of an input layer, hidden layer, and output layer. The learning process is composed of forwarding propagation of signals and back-propagation of errors.

The random forest (RForest) is a widely used machine learning algorithm, which has been successfully applied to pattern recognition [32]. The essence of RF is a classifier containing multiple decision trees. When the test data entered the classifier, each decision tree classified the data. Finally, the class with the most classification results from all decision trees was taken as the result.

The least squares support vector machine (LS-SVM) is a machine learning method that emerged from the statistical learning theory. LS-SVM divided the data samples into multi classes by determining a hyperplane in the input space, where the hyperplane maximizes the separation between the classes [33]. Its key parameter indexes are the kernel function and the corresponding parameters of this function.

Extreme learning machine (ELM) is one of the effective training algorithms for single-layer feedforward neural networks [34]. ELM has a faster training and better generalization performance than traditional machine learning algorithms and could overcome issues such as the local minimum, inappropriate learning rate, and overfitting [35]. Therefore, it is widely used in problems such as classification and regression.

All procedures were written in Matlab R2018b (The MathWorks, Natick, MA, USA) and ran on a personal computer with an Intel Core i5-9400F CPU, 16GB RAM, and a Windows 10 operating system.

## Results And Discussion
3.1 Phenotypic development of healthy and inoculated branch

Figure 1(a) demonstrated the strains of the fungus *valsa mali* on PDA medium. The junctions of diseased and healthy tissues in the inoculated branch samples were assessed visually in the early stage of AVC disease. The bark surface of inoculated branch samples showed no visible symptoms during the first 7 days. However, the phloem inside the bark appeared with early infection symptoms. Figure 1(b) demonstrated the dynamic process of the diseased phloem in the first 7 days. The healthy phloem (first 3 days) had a smooth surface and displayed tender green. The diseased phloem became rough and showed pale brown when the symptoms of mild infection were visible on the 5th day. Subsequently, the diseased phloem appeared dark brown, and the tissue was rotten on the 7th day. The infected area of the diseased phloem, centered on the inoculation site, was continuously extended outward with time. Most notably, the infection symptoms always remained in the phloem and did not appear on the bark surface in the first 7 days. The phloem regions were manually labelled as healthy, disease-1 (the disease-health junctions), and disease-2 (late-disease) according to the infection progression of the pathogen. In Fig. 1(c), the disease-health junction of the diseased phloem was presented using optical microscopy. It can be observed that the healthy tissue appeared green with intact cellular tissue structure; The disease-1 tissue appeared dark brown, and the infected tissue outewed light brown water stain; The disease-2 tissue is mainly characterized by canker, softened tissue.

3.2 SERS AgNPs and its characterization

The microstructure, UV-Vis spectrum, and Raman spectrum of AgNPs were analyzed to investigate the enhancement effects of the synthesized AgNPs. Figure 2(a) is the transmission electron microscopy (TEM) image of AgNPs, Fig. 2(b) displayed the UV-Vis spectra, and Fig. 2(c) showed the Raman spectrum.

In Fig. 2(a), it can be seen that the morphological character of AgNPs was very uniform in a monodisperse spherical shape. In addition, the average diameter of AgNPs was about 50 nm. As shown in Fig. 2(b), only one UV-Vis characteristic absorption peak (at 410nm) corresponding to the single plasmon resonance mode was observed, and the half-peak breadth was only 90 nm. These features further indicated that the shape and size of the synthesized AgNPs were very uniform. In Fig. 2(c), the Raman spectrum had a faint signal, suggesting that the synthesized AgNPs themselves had no strong Raman characteristic peaks and did not have an interferential effect on experimental results. Therefore, the synthesized AgNPs were suitable as SERS substrate to detect branch samples in this research.

3.3 Overview of SERS spectra

Spectral imaging is able to acquire the spectra from a specified point at the sample surface. By adjusting the $x,y$ position, acquisitions of the spectra from multiple points on the sample surface can be performed, assembling a spectral image of the sample. Figure 3 showed the micro-spectral image of diseased phloem through pointwise scanning using by Raman micro-imaging system. The spectral data was
obtained by splitting each pixel point of the spectral image. All the SERS original spectra were also shown in Fig. 3. There was an obvious baseline offset in the disease-1 and disease-2 even after dropwise addition of the AgNPs to suppress fluorescence. Therefore, the MSBC, AsLs, and air-PLS algorithms were adopted to eliminate the disturbances of the baseline offset. The parameters for these methods were manually set to obtain the best result. For the MSBC algorithm, the parameters were set to $\lambda = 150, \mu = 8 \times 10^7, \rho = 0$. For the AsLs algorithm, the parameters were set to $\lambda = 5 \times 10^3, \rho = 1 \times 10^{-4}$. For the air-PLS algorithm, the parameters were set to $\lambda = 150, \rho = 0.01$. The corrected spectra and the predicted fluorescence baselines were plotted in Fig. 4(a)-(c). As shown in Fig. 4, the curved baselines were well-fitted and subtracted by the three algorithms. The corrected spectra showed that the baselines were pulled back to zero absorbance, the peak locations remained unchanged, and the peak shapes were more obvious, which indicated the effectiveness of the baseline correction methods.

As shown in Fig. 4, many SERS peaks can be clearly observed. In detail, the peaks at 319, 957, 1026, 1165, 1242, and 1325 cm$^{-1}$ were indicators of cellulose, corresponding to C-C-C or C-O-C skeletal bending [36], C-C or C-O stretching vibration [37], C-C or C-O stretching vibration [37], H-C-C or H-C-O skeletal bending [38], C = O stretching vibration [37], and C-H bending vibration [38]. The peaks at 625, 731, 1599, and 2939 cm$^{-1}$ were indicators of lignin, corresponding to skeletal bending [39], skeletal bending [39], C-C aromatic ring [40], and C-H asymmetric stretching vibration [41]. The assignment of characteristic wavenumbers was presented in Table 1.

| wavenumber | assignment | Biological components | References |
|------------|------------|-----------------------|------------|
| 319        | C-C-C or C-O-C skeletal bending | cellulose | [36] |
| 957        | C-C or C-O stretching vibration |          | [37] |
| 1026       | C-C or C-O stretching vibration |          | [37] |
| 1165       | H-C-C or H-C-O skeletal bending |          | [38] |
| 1242       | C = O stretching vibration |          | [37] |
| 1325       | C-H bending vibration |          | [38] |
| 625        | skeletal bending | lignin | [39] |
| 731        | skeletal bending |          | [39] |
| 1599       | C-C aromatic ring |          | [40] |
| 2939       | C-H asymmetric stretching vibration |          | [41] |

3.4 Selecting optimal preprocessing method
The correlation analysis method was adopted to select the best preprocessing algorithms. The correlation between the corrected variables was plotted in Fig. 5. Especially, the regions close to the line $y = x$ had a correlation coefficient of 1, indicating that the original spectra were greatly disturbed by the baseline offset. This high degree of collinearity would cause adverse effects on classification analysis. Comparing Fig. 5(b)-(d) with Fig. 5(a), the regions with a high degree of collinearity have noticeably declined, and most of the spectral variables had low correlation with others except in the spectral ranges of 300–400, 640–880, and 1490–1970 cm$^{-1}$. In addition, the proportion of pixel points with values greater than 0.6 to the total pixel points was calculated, and the proportions were 0.35, 0.09, 0.24, and 0.07, respectively. The AsLs method failed to effectively fit the baseline at 1200–1600 cm$^{-1}$, resulting in a relatively poor result of baseline correction. This result indicated that the MSBC and air-PLS baseline offset elimination strategies can greatly reduce the high correlation levels among spectral variables, and the air-PLS algorithm had the best elimination effect. Therefore, the spectra corrected by the air-PLS algorithm were used for further analysis.

### 3.5 Clustering visualization by PCA

As an unsupervised learning strategy, PCA was often used to demonstrate clustering effect based on the similarity of samples in the feature space. In the present research, PCA was performed on the preprocessed spectra of the total sample set to visualize the distribution of healthy, disease-1, and disease-2 samples. The score scatter plot of clustering analysis was shown in Fig. 6. PC1, PC2, and PC3 provided 50.85%, 14.73%, and 10.73% of the variations among samples, respectively. The cumulative contribution of the first three PCs achieved 76.31%. Figure 6 demonstrated that the healthy, disease-1, and disease-2 samples had obvious clustering effects. Therefore, the three types of samples had distinct spectral characteristics.

### 3.6 Optimal variables selection

There were 1401 variables in the SERS spectra. However, spectral data contained many non-critical variables, which might reduce the accuracy and stability of subsequent discriminant models. Therefore, the selection of optimal variables was important for better models. In the present research, two strategies were used to select characteristic variables: algorithm selection (CARS combined with RFrog) and manual selection.

Then, the 10 most important variables were extracted from the total 1401 spectral variables in the full range of 300–3000 cm$^{-1}$, as shown in Fig. 7. The selected optimal variable subsets were set to subset-1 and subset-2, respectively. In detail, the characteristic variables selected by the two methods were listed in Table 1. In the algorithm selection method, the subset-1 10 wavenumbers at 448, 536, 667, 1165, 1211, 1312, 1314, 1412, 1707, and 2951 cm$^{-1}$ were identified. In the manual selection method, the subset-2 10 wavenumbers at 319, 625, 731, 957, 1026, 1165, 1325, 1460, 1570, and 2939 cm$^{-1}$ were identified.

### 3.7 Discriminant models establishment
Before establishing discriminant models, SERS spectral data were divided into a calibration set and a prediction set at the ratio of 3:1. Generally, the independent variable \( (x) \) represented the spectral matrix of samples, and labeled grades \( (y) \) stood for the AVC infection severities. Therefore, the labels for healthy, disease-1, and disease-2 were 1, 2, and 3, respectively. BP-ANN, ELM, RForest, and LS-SVM models were established using four variable matrices \( (x) \) to classify the healthy, disease-1, and disease-2 samples. These four variable matrices \( (x) \) included the full SERS spectra, the subset-1, the subset-2, and the predicted fluorescence baselines.

After formula calculation and experience screening, the learning rate of the BP-ANN model was set uniformly to 0.1, and the number of neurons in the hidden layers were 10, 3, 3, 10, respectively. The number of neurons in the hidden layer of the ELM model was determined by comparing the performances of the ELM model using different numbers of neurons from 1 to 100 with a step of 1. The ELM with 34 neurons was selected as the optimal model. The number of decision trees in the RForest model was determined by comparing the model performances using different numbers of decision trees from 1 to 500 with a step of 1. The RForest with 100 decision trees was selected as the optimal model. The LS-SVM model used RBF as the kernel function, and the optimal penalty coefficient \( (c) \) and the kernel function parameter gamma \( (g) \) were obtained by a grid search procedure. Finally, the best-\( c \) was 379, and the best-\( g \) was 45.

The discriminant accuracy of the models was presented in Table 2. There were significant differences in the classification results of the four models on the full spectra dataset. The classical BP-ANN model learned complex relationships between data, thus improving the analytical performance (such as high sensitivity and specificity) of classification. However, the BP-ANN model has the regrettable tendency to train toward a local optimal rather than a global optimal [32]. This also explained why the BP-ANN model has the lowest classification accuracy on the full spectra dataset compared to the other three models. As opposed to the BP-ANN model, the LS-SVM model was deterministic and their solution was global and unique. As a result, the classification accuracy of the LS-SVM model improved significantly compared to the BP-ANN model. In the present case of the RForest model, each tree selected features that maximize separation of the dataset in three classes. The output of each decision tree was then pooled, leading to the final optimal classification result. Therefore, the RForest model also exhibited excellent analytical performance comparable to the LS-SVM model.
Table 2
The discriminant accuracy of the models

| models     | discriminant accuracy | running time |
|------------|-----------------------|--------------|
|            | full SERS spectra     | subset-1     | subset-2     | predicted fluorescence baselines |             |
| BP-ANN     | 86.22%                | 93.17%       | 92.42%       | 91.80%                           | 0.28s       |
| ELM        | 92.36%                | 85.35%       | 88.93%       | 95.39%                           | 0.01s       |
| RF         | 98.46%                | 96.67%       | 95.87%       | 99.57%                           | 0.15s       |
| LS-SVM     | 98.86%                | 94.49%       | 95.48%       | 98.04%                           | 0.91s       |

Compared with the full spectra dataset, over 99% of non-critical input variables (10 vs. 1401) were removed in subset-1 and subset-2. Meanwhile, the classification accuracy of the subset models was not decreased significantly, which demonstrated the superiority of the optimal variables selection strategies. Generally, the fluorescence baselines reduced the simplicity and robustness of a calibration model built on the raw spectra. Therefore, the existing studies by other scholars had removed the fluorescence baseline from the raw data. However, the classification accuracy of the models based on the fluorescence dataset was surprisingly excellent in the present research. Fungus *valsa mali* produced various chemical substances such as protocatechuic acid, isocoumarin, and phlorizin when infesting the phloem tissue. Although these chemical substances produced fluorescence interference, the baseline also simultaneously reflected the chemical composition and content information. Thus, the fluorescence baseline became available as valid information. This innovative discovery will guide our subsequent research.

However, the above three methods mainly focused on feature extraction, optimal parameters, and optimal variables selection without considering the model runtime, which was also important for intelligent online detection. Furthermore, the intelligent online detection must be an important research direction in plant disease detection fields. The ELM model randomly generated the hidden node parameters and then analytically determined the output weights instead of iterative tuning [42]. Thus, the ELM model runs very fast and lends itself to real application scenarios, which is very important for intelligent online detection. As seen in Table 2, the ELM model ran as fast as 0.01 seconds, far better than the other three methods. The LS-SVM model first used the grid search method to select the best-c and best-g, severely delaying the discriminatory efficiency and making the run time as high as 0.91 seconds. Therefore, the ELM algorithm can be considered as the detection model in the subsequent online detection study.

3.8 Chemical imaging analysis of the disease-health junction

The SERS microspectral image data cube of each phloem sample was processed by the air-PLS algorithm to eliminate fluorescence baseline, and the parameter values were consistent with Sect. 3.3.
Then the processed microspectral cube in a pixel-wise manner generated chemical distribution images in Fig. 8. The symmetric tensile vibration at 1600 cm$^{-1}$ in lignin was identified as the characteristic peak of lignin components, while the bands at 300–550 cm$^{-1}$ were contributed by cellulose. Therefore, these images were constructed based on the cellulose signature peak at 300–550 cm$^{-1}$ and lignin signature peak at 1600 cm$^{-1}$.

Due to the cell walls were probed in phloem tissues, the spectra collected do not contain any intracellular signals. Figure 8(a)-(c) shown the optical micrographs of the phloem tissues taken from the branches of three trees. The chemical imaging results based on the cellulose signature peak at 300–550 cm$^{-1}$ were shown in Fig. 8(d)-(f). The redder-colored the pixels, the stronger the spectral signals of the chemical component. Meanwhile, the bluer-colored the pixels, the weaker the spectral signals. It can be noticed that the SERS signal at the healthy tissue exhibited high intensity with red, bright yellow, and green pixel colors. The diseased phloem tissue exhibited low intensity with blue and green pixel colors, and the disease-health junction exhibited green pixel colors. These differences in SERS imaging of different regions can be attributed to differences in cell wall components. The chemical imaging results based on the lignin signature peak at 1600 cm$^{-1}$ were shown in Fig. 8(g)-(i), showing a similar pattern as the cellulose distribution. The different regions of the phloem tissue shown a distinct distribution of cellulose and lignin, and the observations here were in good agreement with optical micrographs. The results suggested that cellulose and lignin in the cell walls of infected tissues reduced significantly. It also confirmed previous research [43] that cell wall degrading enzymes were considered to play an important role in fungal infection. Therefore, Raman microimaging was useful to detect AVC at early infection stages. It is worth noting that Raman microimaging can visualize the intensity and distribution of components of the cell walls in situ through cytological observations. Meanwhile, this rapid and non-invasive chemical imaging strategy is superior to the other methods, such as the reagent staining method and transmission electron microscopy.

**Conclusion**

SERS spectroscopy combining with chemometric methods for early detection of the AVC disease was considered feasible and promising. Firstly, three spectral preprocessing algorithms were compared, and the air-PLS algorithm was considered effective in removing the spectra fluorescence background. PCA provided a good clustering effect to visualize the distribution of samples in three classes. Two strategies selected optimal variables to develop machine learning models for detecting AVC disease, and these models exhibited excellent analytical performance. Meanwhile, the classification accuracy of the models based on the fluorescence dataset was surprisingly excellent, which was a great inspiration. Besides, this study proposed a new strategy for SERS chemical imaging of the diseased apple phloem tissues using a non-destructive, label-free method. This chemical imaging provided the spatiotemporal dynamic characteristics of changes in the cellulose and lignin of the phloem disease-health junction under fungus stress, which would be helpful in early AVC detection and analysis of disease dynamics.
Declarations

Authors' contributions

Shiyan Fang: Writing - original draft, Writing - review & editing; Junmeng Li: Investigation, Resources, Writing - original draft, Writing - review & editing, Revision; Yan Wang: Investigation, Resources; Yanru Zhao: Investigation, Resources, Writing - original draft, Writing - review & editing. Keqiang Yu: Conceptualization, Investigation, Resources, Writing - review & editing.

Acknowledgements

We thank the life science large instrument sharing platform (Northwest A&F University) for its support for TEM experiments.

Competing interests

The authors declare that they have no known competing financial interests.

Availability of data and material

Please contact the corresponding author for data requests.

Funding

This work was supported by the National Natural Science Foundation of China (Program Nos: 31901403, 61705188), Natural Science Basic Research Plan in Shaanxi Province of China (Program No.: 2020JQ-267), China Postdoctoral Science Foundation (Program No.: 2018M641023), Shaanxi Province Postdoctoral Science Foundation (Program No.: 2017BSHYDZZ61), Science and technology Innovation and Achievement transformation project of experimental demonstration Station (Program No.: TGZX2019-10), and the Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, P. R. China.

Ethics approval and consent to participate

Not applicable.

Consent for publication
References

1. Wang X, Wei J, Huang L, Kang Z. Re-evaluation of pathogens causing Valsa canker on apple in China. Mycologia. 2011;103(2):317–24.

2. Zang R, Yin Z, Ke X, Wang X, Li Z, Kang Z, Huang L. A Nested PCR Assay for Detecting Valsa mali var. mali in Different Tissues of Apple Trees. Plant Dis. 2012;96(11):1645–52.

3. Suzuki K. Population structure of Valsa ceratosperma, causal fungus of Valsa canker, in apple and pear orchards. J Gen Plant Pathol. 2008;74(2):128–32.

4. Meng X-I, Qi X-h, Han Z-y, Guo Y-b, Wang Y-n, Hu T-l. Wang L-m, Cao K-q, Wang s-t. Latent Infection of Valsa mali in the Seeds, Seedlings and Twigs of Crabapple and Apple Trees is a Potential Inoculum Source of Valsa Canker. Sci Rep. 2019;9.

5. Golhani K, Balasundram SK, Vadomalai G, Pradhan B. A review of neural networks in plant disease detection using hyperspectral data. Information Processing in Agriculture. 2018;5(3):354–71.

6. Liu M, McCabe E, Chapados JT, Carey J, Wilson SK, Tropiano R, Redhead SA, Lévesque CA, Hambleton S. Detection and identification of selected cereal rust pathogens by TaqMan® real-time PCR. Can J Plant Pathol. 2015;37(1):92–105.

7. Fang Y, Ramasamy RP. Current and Prospective Methods for Plant Disease Detection. Biosensors. 2015;5(3):537–61.

8. Martinelli F, Scalenghe R, Davino S, Panno S, Scuderi G, Ruisi P, Villa P, Stroppiana D, Boschetti M, Goulart LR, et al. Advanced methods of plant disease detection. A review. Agron Sustain Dev. 2015;35(1):1–25.

9. Okiro LA, Tancos MA, Nyanjom SG, Smart CD, Parker ML. Comparative Evaluation of LAMP, qPCR, Conventional PCR, and ELISA to Detect Ralstonia solanacearum in Kenyan Potato Fields. Plant Dis. 2019;103(5):959–65.

10. Cruz A, Ampatzidis Y, Pierro R, Materazzi A, Panattoni A, De Bellis L, Luvisi A. Detection of grapevine yellows symptoms in Vitis vinifera L. with artificial intelligence. Comput Electron Agr. 2019;157:63–76.

11. Khaled AY, Abd Aziz S, Bejo SK, Nawi NM, Abu Seman I. Spectral features selection and classification of oil palm leaves infected by Basal stem rot (BSR) disease using dielectric spectroscopy. Comput Electron Agr. 2018;144:297–309.

12. Khairunniza-Bejo S, Vong CN. Detection of Basal Stem Rot (BSR) Infected Oil Palm Tree Using Laser Scanning Data. Agric Agric Sci Proc. 2014;2:156–64.

13. Zhao Y, Fang S, Ye Y, Yu K. Chemometric development using portable molecular vibrational spectrometers for rapid evaluation of AVC (Valsa mali Miyabe et Yamada) infection of apple trees. Vib Spectrosc. 2021;114:103231.
14. Fang S, Cui R, Wang Y, Zhao Y, Yu K, Jiang A. Application of multiple spectral systems for the tree disease detection: A review. Appl Spectroscop Rev. 2021.

15. Vallejo Perez MR, Galindo Mendoza MG, Ramirez Elias MG, Javier Gonzalez F, Navarro Contreras HR. Contreras Servin C. Raman Spectroscopy an Option for the Early Detection of Citrus Huanglongbing. Appl Spectrosc. 2016;70(5):829–39.

16. Sanchez L, Pant S, Xing Z, Mandadi K, Kurouski D. Rapid and noninvasive diagnostics of Huanglongbing and nutrient deficits on citrus trees with a handheld Raman spectrometer. Anal Bioanal Chem. 2019;411(14):3125–33.

17. Sanchez L, Pant S, Irey M, Mandadi K, Kurouski D. Detection and identification of canker and blight on orange trees using a hand-held Raman spectrometer. J Raman Spectrosc. 2019;50(12):1875–80.

18. Mukherjee D, Rao BG, Reddy BM. Characterization of Ceria-Based Nano-Oxide Catalysts by Raman Spectroscopy. Top Catal. 2017;60(19):1673–81.

19. Petrov DV. Pressure dependence of peak positions, half widths, and peak intensities of methane Raman bands (ν2, 2ν4, ν1, ν3, and 2ν2). J Raman Spectrosc. 2017;48(11):1426–30.

20. Li X, Sha J, Chu B, Wei Y, Huang W, Zhou H, Xu N, He Y. Quantitative visualization of intracellular lipids concentration in a microalgae cell based on Raman micro-spectroscopy coupled with chemometrics. Sensor Actuat B: Chem. 2019;292:7–15.

21. Qin J, Chao K, Kim MS. Investigation of Raman chemical imaging for detection of lycopene changes in tomatoes during postharvest ripening. J Food Eng. 2011;107(3):277–88.

22. Yang G, Wang Q, Liu C, Wang X, Fan S, Huang W. Rapid and visual detection of the main chemical compositions in maize seeds based on Raman hyperspectral imaging. Spectrochim Acta A. 2018;200:186–94.

23. Eilers PHC. A perfect smoother. Anal Chem. 2003;75(14):3631–6.

24. Eilers PHC. Parametric time warping. Anal Chem. 2004;76(2):404–11.

25. Peng J, Peng S, Jiang A, Wei J, Li C, Tan J. Asymmetric least squares for multiple spectra baseline correction. Anal Chim Acta. 2010;683(1):63–8.

26. Baek S-J, Park A, Ahn Y-J, Choo J. Baseline correction using asymmetrically reweighted penalized least squares smoothing. Analyst. 2015;140(1):250–7.

27. Dong C-w, Ye Y, Zhang J-q, Zhu H-k, Liu F. Detection of Thrips Defect on Green-Peel Citrus Using Hyperspectral Imaging Technology Combining PCA and B-Spline Lighting Correction Method. J Integr Agr. 2014;13(10):2229–35.

28. Li H, Liang Y, Xu Q, Cao D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal Chim Acta. 2009;648(1):77–84.

29. Li Q, Huang Y, Song X, Zhang J, Min S. Moving window smoothing on the ensemble of competitive adaptive reweighted sampling algorithm. Spectrochim Acta A. 2019;214:129–38.

30. Li H-D, Xu Q-S, Liang Y-Z. Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification. Anal
31. Zhang L, Wang F, Sun T, Xu B. A constrained optimization method based on BP neural network. Neural Comput Appl. 2018;29(2):413–21.

32. Lussier F, Thibault V, Charron B, Wallace GQ, Masson J-F. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. TrAC Trend Anal Chem. 2020;124:115796.

33. Lucay FA, Cisternas LA, Gálvez ED. An LS-SVM classifier based methodology for avoiding unwanted responses in processes under uncertainties. Comput Chem Eng. 2020;138:106860.

34. Qiu S, Gao L, Wang J. Classification and regression of ELM, LVQ and SVM for E-nose data of strawberry juice. J Food Eng. 2015;144:77–85.

35. Wu D, Wang X, Wu S. A Hybrid Method Based on Extreme Learning Machine and Wavelet Transform Denoising for Stock Prediction. Entropy. 2021;23(4).

36. Szymanska-Chargot M, Cybulska J, Zdunek A. Sensing the Structural Differences in Cellulose from Apple and Bacterial Cell Wall Materials by Raman and FT-IR Spectroscopy. Sensors. 2011;11(6):5543–60.

37. Beć K, Grabska J, Bonn G, Popp M, Huck C. Principles and Applications of Vibrational Spectroscopic Imaging in Plant Science: A Review. Front Plant Sci. 2020;11:1226.

38. Edwards HGM, Farwell DW, Webster D. FT Raman microscopy of untreated natural plant fibres. Spectrochim Acta A. 1997;53(13):2383–92.

39. Agarwal UP, McSweeny JD, Ralph SA. FT-Raman Investigation of Milled-Wood Lignins: Softwood, Hardwood, and Chemically Modified Black Spruce Lignins. J Wood Chem Technol. 2011;31(4):324–44.

40. Agarwal UP. Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (Picea mariana). Planta. 2006;224(5):1141–53.

41. Gierlinger N, Schwanninger M. The potential of Raman microscopy and Raman imaging in plant research. Spectroscopy-an International Journal. 2007;21(2):69–89.

42. Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: Theory and applications. Neurocomputing. 2006;70(1):489–501.

43. Ke X, Huang L, Han Q, Gao X, Kang Z. Histological and cytological investigations of the infection and colonization of apple bark by Valsa mali var. mali. Australas Plant Path. 2013;42(1):85–93.
**Figure 1**

Phenotypic development of healthy and inoculated branch. (a) the strains of the fungus valsa mali on PDA medium, (b) the dynamic process of the diseased phloem in the first 7 days, (c) optical micrograph of the disease-health junction.

**Figure 2**

SERS AgNPs and its characterization. (a) transmission electron microscopy image of AgNPs, (b) the UV-Vis spectra of AgNPs, (c) Raman spectrum of AgNPs.
Figure 3

The sketch represents the basic principle of the spectral data cube and shows the raw spectra and spectral imaging of three types of samples.

Figure 4

Spectral baseline correction. (a) baseline correction using MSBC, (b) baseline correction using AsLS, (c) baseline correction using air-PLS. The blue line represents the original spectrum, the red line represents
the estimated baseline, and the yellow line represents the corrected spectrum.

**Figure 5**

The correlation between the corrected variables was plotted. (a) high correlations were found among original spectral variables, (b) correlations were noticeably declined using MSBC, (c) correlations were noticeably declined using AsLS, (d) correlations were noticeably declined using air-PLS. However, the air-PLS algorithm has the best elimination effect.
Figure 6

Score plots of first three PCs from PCA on spectral data of the three types of samples.
The characteristic variables for early disease detection of the AVC disease. Two strategies were used to select characteristic variables: algorithm selection (subset-1) and manual selection (subset-2).

Figure 7

Figure 8
Chemical distribution images of phloem samples based on SERS spectroscopy. (a)-(c) the optical micrographs of the phloem tissues taken from the branches of three trees, (d)-(f) The chemical imaging results based on the cellulose signature peak at 300-550 cm\(^{-1}\), (g)-(i) The chemical imaging results based on the lignin signature peak at 1600 cm\(^{-1}\).