Article

Identifying Optimal Wavelengths as Disease Signatures Using Hyperspectral Sensor and Machine Learning

Xing Wei [1,2], Marcela A. Johnson [1,3], David B. Langston, Jr. [1,2], Hillary L. Mehl [1,2,4] and Song Li [1,*]

1 School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, USA; xingwei@vt.edu (X.W.); maguileraf23@vt.edu (M.A.J.); dblangston@vt.edu (D.B.L.J.); hillary.mehl@usda.gov (H.L.M.)
2 Virginia Tech Tidewater Agricultural Research and Extension Center, Suffolk, VA 23437, USA
3 Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA 24061, USA
4 United States Department of Agriculture, Agricultural Research Service, Arid-Land Agricultural Research Center, Tucson, AZ 85701, USA
* Correspondence: songli@vt.edu

Abstract: Hyperspectral sensors combined with machine learning are increasingly utilized in agricultural crop systems for diverse applications, including plant disease detection. This study was designed to identify the most important wavelengths to discriminate between healthy and diseased peanut (*Arachis hypogaea* L.) plants infected with *Athelia rolfsii*, the causal agent of peanut stem rot, using in-situ spectroscopy and machine learning. In greenhouse experiments, daily measurements were conducted to inspect disease symptoms visually and to collect spectral reflectance of peanut leaves on lateral stems of plants mock-inoculated and inoculated with *A. rolfsii*. Spectrum files were categorized into five classes based on foliar wilting symptoms. Five feature selection methods were compared to select the top 10 ranked wavelengths with and without a custom minimum distance of 20 nm. Recursive feature elimination methods outperformed the chi-square and SelectFromModel methods. Adding the minimum distance of 20 nm into the top selected wavelengths improved classification performance. Wavelengths of 501–505, 690–694, 763 and 884 nm were repeatedly selected by two or more feature selection methods. These selected wavelengths can be applied in designing optical sensors for automated stem rot detection in peanut fields. The machine-learning-based methodology can be adapted to identify spectral signatures of disease in other plant-pathogen systems.

Keywords: soilborne diseases; peanut stem rot; *Athelia rolfsii*; *Sclerotium rolfsii*; spectroscopy; random forest; support vector machine; recursive feature elimination; feature selection; hyperspectral band selection

1. Introduction

Peanut (*Arachis hypogaea* L.) is an important oilseed crop cultivated in tropical and subtropical regions throughout the world, mainly for its seeds, which contain high-quality protein and oil contents [1,2]. The peanut plant is unusual because even though it flowers above ground, the development of the pods that contain the edible seed occurs below ground [3], which makes this crop prone to soilborne diseases. Stem rot of peanut, caused by *Athelia rolfsii* (Curzi) C.C. Tu & Kimbrough (anamorph: *Sclerotium rolfsii* Sacc.), is one of the most economically important soilborne diseases in peanut production [4]. The infection of *A. rolfsii* usually occurs first on plant tissues near the soil surface mid-to-late season following canopy closure [4]. The dense plant canopy provides a humid microclimate that is conducive for pathogen infection and disease development when warm temperatures (~30 °C) occur [5,6]. However, the dense plant canopy not only prevents foliar-applied fungicides from reaching below the canopy where infection of *A. rolfsii* initially occurs but also blocks visual inspection of signs and symptoms of disease.

Accurate and efficient diagnosis of plant diseases and their causal pathogens is the critical first step to develop effective management strategies [7–9]. Currently, disease

assessments for peanut stem rot are based on visual inspection of signs and symptoms of this disease. The fungus, *A. rolfsii*, is characterized by the presence of white mycelia and brown sclerotia on the soil surface and infected plant tissues [4,10]. The initial symptom in plants infected with *A. rolfsii* is water-soaked lesions on infected tissues [11], while the first obvious foliar symptoms are wilting of a lateral branch, the main stem, or the whole plant [4]. To scout for the disease in commercial fields, walking in a random manner and selecting many sites per field (~1.25 sites/hectare and ≥5 sites per field) are recommended to make a precise assessment of the whole field situation [12]. Peanut fields should be scouted once a week after plant pegging [13] as earlier detection before the disease is widespread is required to make timely crop management decisions. The present scouting method based on visual assessment is labor-intensive and time-consuming for large commercial fields, and there is a high likelihood of overlooking disease hotspots. Soilborne diseases including peanut stem rot typically have a patchy distribution, considered as disease hotspots, in the field. Spread of soilborne diseases is generally attributed to the expansion of these disease hotspots. If hotspots can be detected relatively early, preventative management practices such as fungicide application can be applied to the rest of the field. Thus, there is a need to develop a new method that can detect this disease accurately and efficiently.

Recent advancements of digital technologies including hyperspectral systems boost their applications in agriculture including plant phenotyping for disease detection [14–17]. Hyperspectral sensors have proven their potential for early detection of plant diseases in various plant-pathogen systems [8,9,18–26]. Visual disease-rating methods are based only on the perception of red, blue, and green colors, whereas hyperspectral systems can measure changes in reflectance with a spectral range typically from 350 to 2500 nm under sufficient light conditions [15,16,27]. Plant pathogens and plant diseases can alter the optical properties of leaves by changing the leaf structure and the chemical composition of infected leaves or by the appearance of pathogen structures on the foliage [14,15]. Specifically, the reflectance of leaves in the VIS region (400 nm to 700 nm), the near-infrared (NIR) region (700 to 1000 nm), and the shortwave infrared (SWIR) region (1000 to 2500 nm) is typically related to leaf pigment contents, leaf cell structure, and leaf chemical and water contents, respectively [14,28,29]. A typical hyperspectral scan can generate reflectance data for hundreds of bands. This large volume of high-dimensional data poses a challenge for data analysis to identify informative wavelengths that are directly related to plant health status [14,30].

Feature extraction and feature/band selection methods are commonly applied to hyperspectral data for dimensional reduction and spectral redundancy removal [31,32]. For example, some feature extraction methods project the original high-dimensional data into a different feature space with a lower dimension [33,34], and one of the most commonly used feature extraction methods is principal component transformation [35]. Band selection methods reduce the dimensionality of hyperspectral data by selecting a subset of wavelengths [32,36]. A variety of band selection algorithms has been used for plant disease detection, such as instance-based Relief-F algorithm [37], genetic algorithms [24], partial least square [8,20], and random forest [38]. In contrast to feature extraction methods that may alter the physical meaning of the original hyperspectral data during transformation, band selection methods preserve the spectral meaning of selected wavelengths [32,36,39].

In the past years, applications of machine learning (ML) methods in crop production systems have been increasing rapidly, especially for plant disease detection [30,40–42]. Machine learning refers to computational algorithms that can learn from data and perform classification or clustering tasks [43], which are suitable to identify the patterns and trends of large amounts of data such as hyperspectral data [15,30,41]. The scikit-learn library provides a number of functions for different machine learning approaches, dimensional reduction techniques, and feature selection methods [44]. Results of one of our previous studies demonstrated that hyperspectral radiometric data were able to distinguish between mock-inoculated and *A. rolfsii*-inoculated peanut plants based on visual inspection of leaf spectra after the onset of visual symptoms [45]. In order to more precisely identify

important spectral wavelengths that are foliar signatures of peanut infection with *A. rolfsii*, a machine learning approach was used in this study to further analyze the spectral reflectance data collected from healthy and diseased peanut plant infected with *A. rolfsii*. The specific objectives were (1) to compare the performance of different machine learning methods for the classification of healthy peanut plants and plants infected with *A. rolfsii* at different stages of disease development, (2) to compare the top wavelengths selected by different feature selection methods, and (3) to develop a method to select top features with a customized minimum wavelength distance.

2. Materials and Methods

2.1. Plant Materials

The variety 'Sullivan', a high-oleic Virginia-type peanut, was grown in a greenhouse located at the Virginia Tech Tidewater Agricultural Research and Extension Center (AREC) in Suffolk, Virginia. Three seeds of peanut were planted in a 9.8-L plastic pot filled with a 3:1 mixture of pasteurized field soil and potting mix. Only one vigorous plant per pot was kept for data collection, while extra plants in each pot were uprooted three to four weeks after planting. Peanut seedlings were treated preventatively for thrips control with imidacloprid (Admire Pro, Bayer Crop Science, Research Triangle Park, NC, USA). Plants were irrigated for 3-min per day automatically using a drip method. A Watchdog sensor (Model: 1000 series, Spectrum Technologies, Inc., Aurora, IL, USA) was mounted inside the greenhouse to monitor environmental conditions including temperature, relative humidity, and solar radiation. The temperature ranged from 22 °C to 30 °C during the course of experiments. The daily relative humidity ranged between 14% and 84% inside the greenhouse as reported previously [45].

2.2. Pathogen and Inoculation

The isolate of *A. rolfsii* used in this study was originally collected from a peanut field at the Tidewater AREC research farm in 2017. A clothespin technique, adapted from [46], was used for the pathogen inoculation. Briefly, clothespins were boiled twice in distilled water (30 min each time) to remove the tannins that may inhibit pathogen growth. Clothespins were then autoclaved for 20 min after being submerged in full-strength potato dextrose broth (PDB; 39 g/L, Becton, Dickinson and Company, Sparks, MD, USA). The PDB was poured off the clothespins after cooling. Half of the clothespins were inoculated with a three-day-old actively growing culture of *A. rolfsii* on potato dextrose agar (PDA), and the other half were left non-inoculated to be used for mock-inoculation treatment. All clothespins were maintained at room temperature (21 to 23 °C) for about a week until inoculated clothespins were colonized with mycelia of *A. rolfsii*.

Inoculation treatment was applied to peanut plants 68 to 82 days after planting. The two primary symmetrical lateral stems of each plant were either clamped with a non-colonized clothespin as mock-inoculation treatment or clamped with a clothespin colonized with *A. rolfsii* as inoculation treatment. All plants were maintained in a moisture chamber set up on a bench inside the greenhouse to facilitate pathogen infection and disease development. Two cool mist humidifiers (Model: Vicks® 4.5-L, Kaz USA, Inc., Marlborough, MA, USA) with maximum settings were placed inside the moisture chamber to provide ≥90% relative humidity. Another Watchdog sensor was mounted inside the moisture chamber to record the temperature and relative humidity every 15 min.

2.3. Experimental Setup

Two treatments were included: lateral stems of peanut mock-inoculated or inoculated with *A. rolfsii*. Visual rating and spectral data for this study were collected from 126 lateral stems on 74 peanut plants over four experiments where peanut plants were planted on 19 October and 21 December in 2018 and 8 January and 22 February in 2019 [45]. Each plant was inspected once per day, for 14 to 16 days after inoculation. For daily assessment, plants were moved outside the moisture chamber and placed on other benches inside the

greenhouse to allow the foliage to air dry before data collection. Plants were then moved back into the moisture chamber after measurements. Plants were kept inside the moisture chamber through the experimental course for the first three experiments. In the fourth experiment, plants were moved permanently outside the moisture chamber seven days after inoculation because they were affected by heat stress.

### 2.4. Stem Rot Severity Rating and Categorization

Lateral stems of each plant were inspected each day visually for signs and symptoms of disease starting from two days after inoculation. The lateral stems of plants were categorized based on visual symptom ratings into five classes, where 'Healthy' = mock-inoculated lateral stems without any disease symptoms, 'Presymptomatic' = inoculated lateral stems without any disease symptoms, 'Lesion only' = necrotic lesions on inoculated lateral stems, 'Mild' = drooping of terminal leaves to wilting of <50% upper leaves on inoculated lateral stems, and 'Severe' = wilting of ≥50% leaves on inoculated lateral stems. In this study, we did not categorize plants based on days after inoculation; instead, we categorized them based on distinct disease symptoms including the inoculated pre-symptomatic ones. Because disease development was uneven in the four independent greenhouse experiments, it made more sense to categorize this way rather than doing a time-course analysis.

2.5. Spectral Reflectance Measurement

The second youngest mature leaf on each treated lateral stem was labeled with a twist tie immediately before the inoculation treatment. Spectral reflectance of one leaflet on this designated leaf was measured from the leaf adaxial side once per day starting two days after inoculation using a handheld Jaz spectrometer and an attached SpectroClip probe (Ocean Optics, Dunedin, FL, USA; Figure 1A). The active illuminated area of the SpectroClip was 5 mm in diameter. The spectral range and optical resolution of the Jaz spectrometer were 200 to 1100 nm and ~0.46 nm, respectively. The spectrometer was set to boxcar number 5 and average 10 to reduce machinery noise. A Pulsed Xenon light embedded in the Jaz spectrometer was used as the light source for the hyperspectral system. The spectral range of the Pulsed Xenon light source was 190 to 1100 nm. The WS-1 diffuse reflectance standard, with reflectivity >98%, was used as the white reference (Ocean Optics, Dunedin, FL, USA).
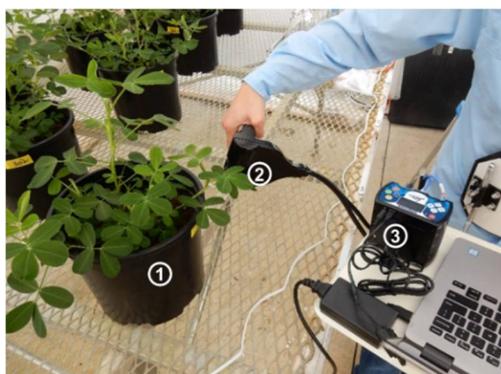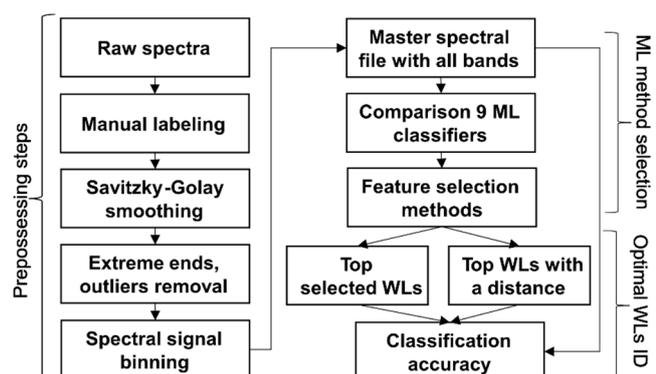
(A) (B)

Figure 1. (**A**) Spectral reflectance measurement of peanut leaves with the Jaz spectrometer system: (1) individual potted peanut plant, (2) SpectroClip probe, (3) Jaz spectrometer. (**B**) Data analysis pipeline for the wavelength selection to classify healthy peanut plants and plants infected with *Athelia rolfsii* at different stages of disease development. ML = machine learning; WL = wavelengths; ID = Identification.

2.6. Data Analysis Pipeline

The data analysis was conducted using the Advanced Research Computing (ARC) system at Virginia Tech (https://arc.vt.edu/, accessed on 4 January 2021). One node of a central processing unit (CPU) (Model: E5-2683v4, 2.1 GHz, Intel, Santa Clara, CA, USA) on the ARC system was used.

2.6.1. Data Preparation

The data analysis pipeline is illustrated in Figure 1B. The spectrum file collected by the Jaz spectrometer was in a Jaz data file format (.JAZ). Each spectrum file contains information of spectrometer settings and five columns of spectral data: W = wavelength, D = dark spectrum, R = reference spectrum, S = sample spectrum, and P = processed spectrum in percentage. Each spectrum file's name was labeled manually according to the experiment number, plant and its lateral stem ID, and date collected. Then spectrum files were grouped into five classes based on the visual ratings of disease symptoms for each individual assessment: 'Healthy', 'Presymptomatic', 'Lesion only', 'Mild', and 'Severe'.

2.6.2. Preprocessing of Raw Spectrum Files

The preprocessing steps of raw spectrum files were adapted from [38] using the R statistical platform (version 3.6.1). The processed spectral data and file name were first extracted from each Jaz spectral data file and saved into a Microsoft Excel comma-separated values file (.CSV). A master spectral file was created consisting of processed spectral data and file names of spectra from all five categories. Spectral data were then smoothed with a Savitzky-Golay filter using the second-order polynomial and a window size of 33 data points [47]. Wavelengths at two extreme ends of the spectrum (<240 nm and >900 nm) were deleted due to large spectral noises. Outlying spectrums were removed using depth measures in the Functional Data Analysis and Utilities for Statistical Computing (fda.usc) package [48]. The final dataset after outlier removal consisted of 399 observations including 82 for 'Healthy', 79 'Presymptomatic', 72 'Lesion only', 73 'Mild', and 93 'Severe'. The spectral signals were further processed using the prospectr package to reduce multicollinearity between predictor variables [38,49]. The bin size of 10 was used in this study, meaning the spectral reflectance for each wavelength was the average spectral reflectance values of 10 adjacent wavelengths before signal binning. The spectral resolution was reduced from 0.46 nm (1569 predictor variables/wavelengths) to 4.6 nm (157 predictor variables/wavelengths).

2.6.3. Comparison of Machine Learning Methods for Classification

The spectral data after preprocessing steps were analyzed using the Python programming language (version 3.7.10). Jupyter notebooks for the analysis in this work are provided in the following github repository (https://github.com/LiLabAtVT/HyperspecFeatureSelection, accessed on 26 May 2021). The performance of eight common machine learning methods was compared for the classification of mock-inoculated, healthy peanut plants and plants inoculated with *A. rolfsii* at different stages of disease development. The eight machine learning algorithms tested in this study included Gaussian Naïve Bayes (NB), K-nearest neighbors (KNN), linear discriminant analysis (LDA), multi-layer perceptron neural network (MLPNN), random forests (RF), support vector machine with the linear kernel (SVML), gradient boosting (GBoost), and extreme gradient boosting (XGBoost).

All eight algorithms were supervised machine learning methods, but they were compared in this work because they represent categories of different learning models [40]. The Gaussian Naïve Bayes (NB) algorithm belongs to the probabilistic graphical Bayesian model. The KNN uses instance-based learning models. The LDA is a commonly used dimensionality reduction technique. The MLPNN is one of the traditional artificial neural networks. The RF and the two gradient-boosting algorithms use ensemble-learning models by combining multiple decision-tree-based classifiers [40]. Support vector machine (SVM) uses separating hyperplanes to classify data from different classes with the goal

of maximizing the margins between each class [40,50]. The XGBoost classifier was from the library of XGBoost (version 1.3.3) [51], while all the others were from the library of scikit-learn (version 0.24.1) [44]. The default hyperparameters for each classifier were used in this study.

A commonly used chemometric method for classification, partial least square discriminant analysis (PLSDA), was also tested on the preprocessed spectral data of this study. The PLSDA method was implemented using the R packages caret (version 6.0-84) [52] and pls (version 2.7-2) [53]. The metric of classification accuracy was used to tune the model parameter, number of components/latent variables. The softmax function was used as the prediction method.

2.6.4. Comparison of Feature Selection Methods

Five feature selection methods from the scikit-learn library were evaluated in this study. They included one univariate feature selection method (SelectKBest using a chi-square test), two feature selection methods using SelectFromModel (SFM), and two recursive feature elimination (RFE) methods. Both SFM and RFE needed an external estimator, which could assign weights to features. SFM selects the features based on the provided threshold parameter. RFE selects a desirable number of features by recursively considering smaller and smaller sets of features [44]. In addition to the feature selection methods, principal component analysis (PCA), a dimensionality reduction technique, was also included in the comparison. The top 10 wavelengths or components selected from the feature-selection methods were then compared to classify the healthy and diseased peanut plants at different stages.

Unlike hyperspectral sensors, multispectral sensors using a typical RGB or monochrome camera with customized, selected band filters are cheaper and are more widely used in plant phenotyping research. However, the band filters typically have a broader wavelength resolution than hyperspectral sensors. To test whether our feature selection method can be used to guide the design of spectral filters, a method was developed to enforce a minimum bandwidth distance between the selected features. In detail, a python function was first defined using the function "filter" and Lambda operator to filter out elements in a list that were within the minimum distance of a wavelength. Second, the original list of features/wavelengths was sorted based on their importance scores in descending order. Third, a for loop was set up to filter out elements based on the order of the original feature list. All the features left after the filtering process were sorted in a new list. The order of the new list was reversed to obtain the final list of filtered features in descending order.

### 2.6.5. Statistical Tests for Model Comparisons

The classification accuracy was used as the metric in statistical tests to compare different models. Stratified 10-fold cross-validation repeated three times (3×10) was used for each classification model. The 3×10 cross-validation was followed by a Friedman test with its corresponding post hoc Nemenyi test to compare different machine-learning classifiers and different feature-selection methods, respectively. Friedman and Nemenyi tests were the non-parametric versions of ANOVA and Tukey's tests, respectively. The Friedman test was recommended because the two assumptions of the ANOVA, normal distribution and sphericity, were usually violated when comparing multiple learning algorithms [54]. If Friedman tests were significant, all pairwise comparisons of different classifiers were conducted using the nonparametric Nemenyi test with an α level of 0.05.

The best performing classifier and feature selection method was used for further analysis. The spectral data with all bands and with only the distanced top 10 selected wavelengths for all five classes were split into 80% training and 20% testing subsets (random state = 42). The models were trained using the training datasets with all bands and with the distanced top 10 selected wavelengths. Normalized confusion matrixes were plotted using the trained models on the testing datasets.

3. Results

3.1. Spectral Reflectance Curves

The average spectral reflectance values of peanut leaves from different categories were first compared using a one-way ANOVA test by each spectral region including ultraviolet (UV, 240 to 400 nm), visible (VIS, 400 to 700 nm), and near-infrared (NIR, 700 to 900 nm) regions. In the UV region, 'Mild' had the greatest reflectance, followed by 'Severe' and 'Lesion only', while 'Healthy' and 'Presymptomatic' had the lowest reflectance ($p < 0.0001$; Figure 2B). In the VIS region, with the exception of 'Presymptomatic' overlapping with 'Mild', there was a gradient increase of reflectance from 'Healthy' to 'Severe' ($p < 0.0001$; Figure 2C). In the NIR region, 'Severe' had the greatest reflectance, followed by 'Mild' and 'Presymptomatic', while the 'Lesion only' and 'Healthy' had the lowest reflectance ($p < 0.0001$; Figure 2D). The spectral reflectance curves in the NIR regions were noisier compared with the ones in the UV and VIS regions (Figure 2).
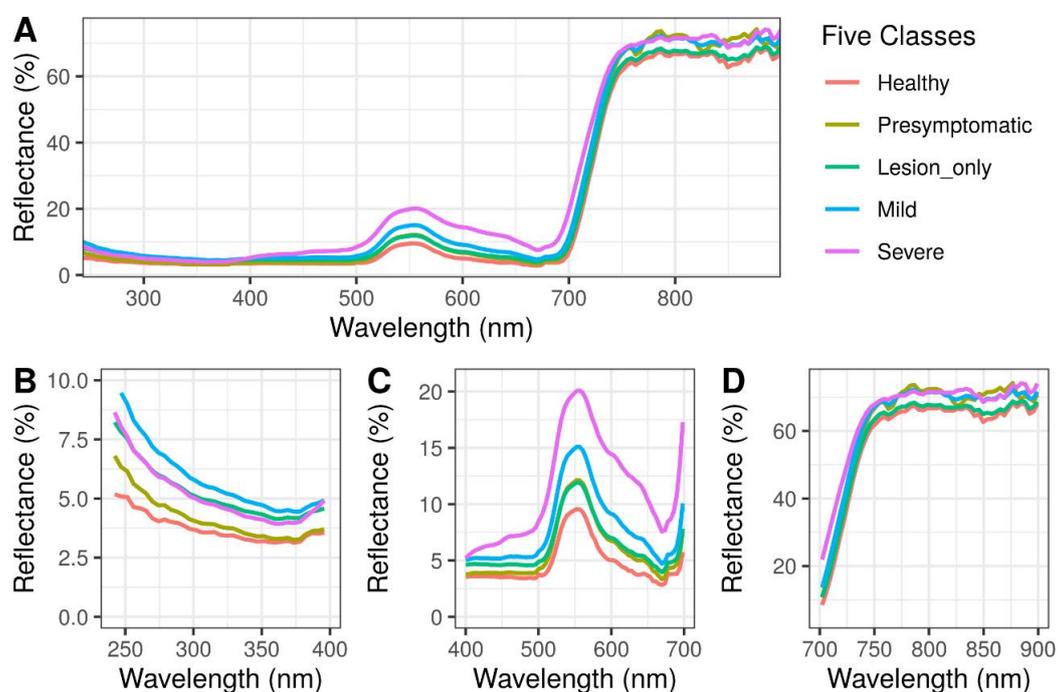

Figure 2. Spectral profiles of healthy and peanut plants infected with *Athelia rolfsii* at different stages of disease development. (**A**) Entire spectral region (240 to 900 nm); (**B**) Ultraviolet region (240 to 400 nm); (**C**) Visible region (400 to 700 nm); (**D**) Near-infrared region (700 to 900 nm).

3.2. Classification Analysis

To determine the best classifiers for downstream analysis, we compared the performance of nine machine learning methods for the classification of models with different input classes. The classification models were divided into three different classification problems: two classes ('Healthy' and 'Severe'), three classes ('Healthy', 'Mild' and 'Severe'), and five classes ('Healthy', 'Presymptomatic', 'Lesion only', 'Mild', and 'Severe'). Overall, there was a decrease in accuracy when the number of input classes increased (Figure 3). Specifically, all machine learning methods had >90% overall accuracy except the LDA classifier for the two-class classification ($p < 0.0001$). For three-class classification, KNN, RF, SVML, PLSDA, GBoost, and XGBoost had greater accuracy compared with NB, LDA, and MLPNN ($p < 0.0001$), and the average accuracy for all methods was approximately 80%. For the five-class classification, RF, SVML, GBoost, and XGBoost performed better than the rest of the classifiers ($p < 0.0001$) (Figure 3), and the accuracy for the best classifier was around 60%. The computational times of RF and SVML were also faster than

GBoost and XGBoost when performing the five-class classification. Based on these results, RF and SVML were selected for further analysis.
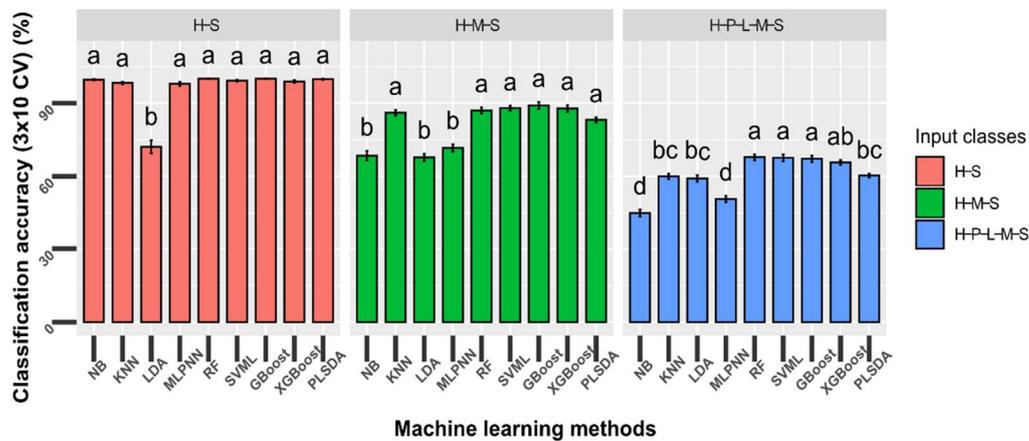
Figure 3. Comparison of the performance of nine machine learning methods to classify mock-inoculated healthy peanut plants and plants inoculated with *Athelia rolfsii* at different stages of disease development. Peanut plants from the greenhouse study were categorized based on visual symptomology: H = 'Healthy', mock-inoculated control with no symptoms; P = 'Presymptomatic', inoculated with no symptoms; L = 'Lesion only', inoculated with necrotic lesions on stems only; M = 'Mild', inoculated with mild foliar wilting symptoms (≤50% leaves symptomatic); S = 'Severe', inoculated with severe foliar wilting symptoms (>50% leaves symptomatic). Machine learning methods tested: NB = Gaussian Naïve Bayes; KNN = K-nearest neighbors; LDA = linear discriminant analysis; MLPNN = multi-layer perceptron neural network; RF = random forests; SVML = Support vector machine with linear kernel; GBoost = gradient boosting; XGBoost = extreme gradient boosting; PLSDA = partial least square discriminant analysis. Bars with different letters were statistically different using nonparametric Friedman and Nemenyi tests with an α level of 0.05. Error bars indicate standard deviation of accuracy using stratified 10-fold cross-validation (CV) repeated three times.

3.3. Feature Weights Calculated by Different Methods

Different machine learning methods determine the important features using different types of algorithms to decide the "weights" of each feature. For example, feature scores were calculated for each wavelength using the chi-square test method. The feature scores had similar shapes for models with different input classes (two classes, three classes, and five classes; Figure 4A). In general, higher scores were found in the VIS region compared with the UV and NIR regions using the chi-square test. Features from the regions of 590 to 640 nm all had similar scores. In contrast, the random forest method calculated feature "weights" using feature importance values. Similar to the chi-square test, greater importance values were assigned to the VIS region than the UV and NIR regions for the RF method (Figure 4B). Unlike the chi-square test, the peaks of importance values mostly occurred in two regions in the VIS region: 480 to 540 nm and 570 to 700 nm. In addition, two peaks were found on two wavelengths in the NIR regions—830 and 884 nm—for the three-class and five-class models, but not for the two-class model (Figure 4B). For peaks in the RF, feature importance scores were not continuous as was the case in the chi-square curve. Instead, there were discrete peaks in the feature importance scores, with two high peaks usually separated by wavelengths with lower importance scores. Finally, for the SVML method, the weights of each feature were calculated by averaging the absolute values of coefficients of multiple classes. Generally, peaks indicating important features occurred in all three regions: UV, VIS, and NIR (Figure 4C). In contrast to the chi-square and RF methods, greater weights were assigned to wavelengths in the NIR regions for the SVML method (Figure 4). Interestingly, for the two-class classification, the weights were much lower than those from the three- and five-class classifications. We also noticed that the absolute values of these "features weights" were not comparable between different approaches, which was expected.
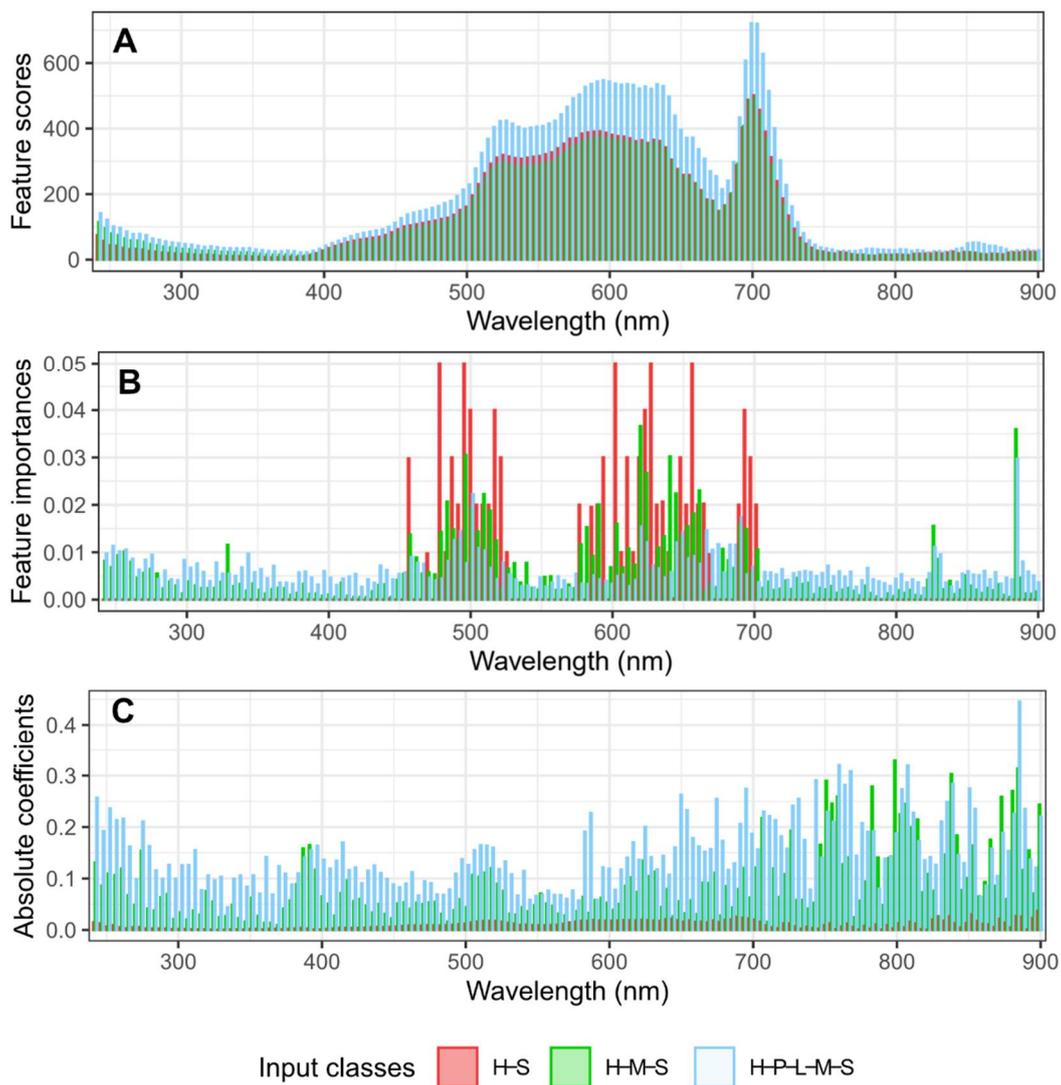
Figure 4. The weights of each feature/wavelength calculated by three different machine learning algorithms for the classification of peanut stem rot models. (**A**) The chi-square method; (**B**) Random forest; (**C**) Support vector machine with a linear kernel. H = 'Healthy', mock-inoculated control with no symptoms; P = 'Presymptomatic', inoculated with no symptoms; L = 'Lesion only', inoculated with necrotic lesions on stems only; M = 'Mild', inoculated with mild foliar wilting symptoms ($\leq$50% leaves symptomatic); S = 'Severe', inoculated with severe foliar wilting symptoms (>50% leaves symptomatic).

3.4. Dimension Reduction and Feature Selection Analysis

To understand the performance of different machine learning methods and the differences in the feature selection process, we performed PCA analysis for the dimension reduction of the spectral data from all five classes. In the PCA plot of the first two components, 'Healthy' (red) and 'Severe' (purple) samples could be easily separated by a straight line (Figure S1). 'Healthy' and 'Presymptomatic' were overlapping but some 'Presymptomatic' samples appeared in a region (PCA2 > 60) that had few samples from other classes. The 'Lesion-only' category overlapped with all other categories with a higher degree of overlapping with healthy and mild samples. In comparison, 'Mild' samples overlapped with all other categories, but had a high degree of overlap with 'Severe' samples compared to other categories (Figure S1). The first component explained nearly 60% of the variance, while the second component accounted for >20% of the variance of the data. Overall, the top three components explained >90% of the variance, and the top 10 components explained >99% of the variance (Figure S2).

To check whether unsupervised dimension reduction can be directly used as a way to select features and perform classifications, we tested the classifiers using the top 10 principal components or wavelengths for the five-class model (H-P-L-M-S). Feature selections were performed using the chi-square test, selection from model for random forest and SVM, and recursive feature selection for random forest and SVM. The top 10 principal components were also used as input features. Two classifiers including RF and SVML were used as classification methods for these selected features and the resulting accuracy was compared to all features without any selection. Regardless of classifiers, the top 10 components from the PCA analysis and top 10 wavelengths selected by the RFE-RF and RFE-SVML methods performed as well as using all bands for the five-class classification in terms of testing accuracy ($p > 0.05$), suggesting that using a few features does not reduce the model performance. RFE methods with either RF or SVML as classifiers performed better than the univariate feature selection (SelectKBest using the chi-square test) and the two SFM methods ($p < 0.0001$) (Figure 5A). Interestingly, features selected using RFE performed similarly, regardless of the classification method used. For example, REF-RF features had similar accuracy when tested using the SVM classifier, and vice versa.
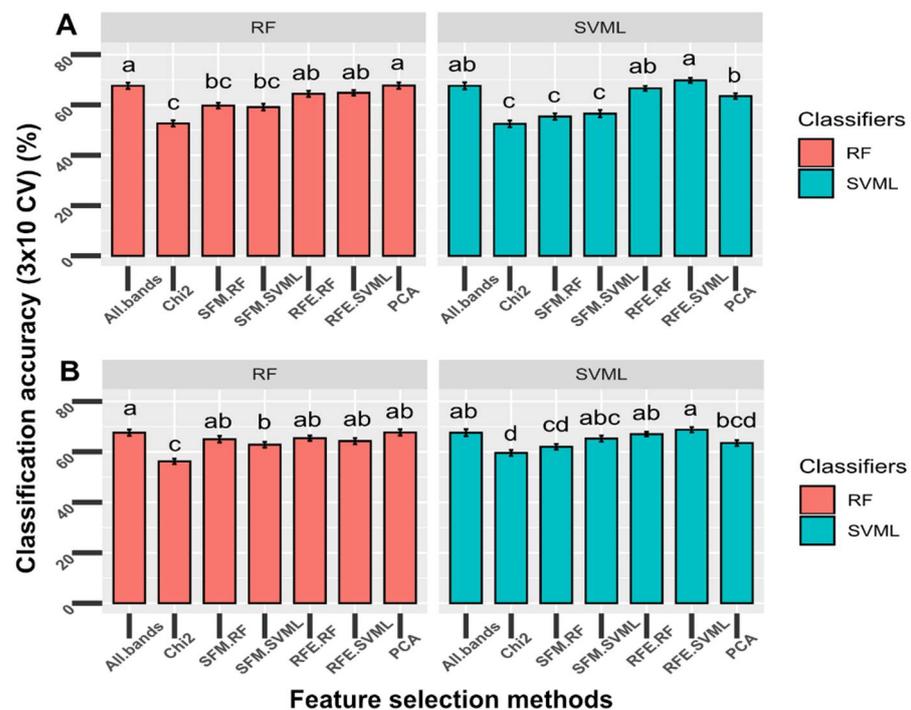
Figure 5. Comparison of the performance of the original top 10 selected wavelengths (**A**) and top 10 features with a minimum of 20 nm distance (**B**) by different feature-selection methods to classify the mock-inoculated healthy peanut plants and plants inoculated with *Athelia rolfsii* at different stages of disease development (input data = all five classes). Feature selection methods tested: Chi2 = SelectKBest (estimator = chi-square); SFM-RF = SelectFromModel (estimator = random forest); SFM-SVML = SelectFromModel (estimator = support vector machine with linear kernel); RFE-RF = Recursive feature elimination (estimator = random forest); RFE-SVML = Recursive feature elimination (estimator = support vector machine with linear kernel). The top 10 components were used as input data for the principal component analysis (PCA) method. The two classifiers tested: RF = random forest; SVML = support vector machine with the linear kernel. Bars with different letters were statistically different using nonparametric Friedman and Nemenyi tests with an α level of 0.05. Error bars indicate standard deviation of accuracy using stratified 10-fold cross-validation repeated three times.

3.5. Feature Selection with a Custom Minimum Distance

The spectral reflectance of wavelengths close to each other was highly correlated. It held true, especially for wavelengths within each region for the UV, VIS, and NIR regions. Wavelengths from different regions were less correlated with each other (Figure S3). The top wavelengths selected by some methods were neighboring or close to each other (Table 1A). Thus, the newly developed method described in Section 2.6.4 was implemented to enforce a minimum distance into the top selected features. The performance of the accuracies for the five-class classification was compared between top features with and without a customized minimum distance. The top 10 features with and without a minimum distance of 20 nm from each of the five feature selection methods were used as input data for the five-class classification (Table 1). Two classifiers, including RF and SVML, were tested. Regardless of classifiers, the top 10 features with a minimum distance of 20 nm performed better than the original top 10 features selected by chi-square and two SFM methods ($p < 0.05$; Figure 5B). The top 10 features with a minimum distance of 20 nm performed better than the original top 10 features in five out of ten methods compared ($p < 0.05$; Figure 6). In the other five methods, the top 10 features with the minimum distance performed the same as the original top 10 features, suggesting that using minimum distance filtering does not decrease the model performance (see discussion).

Table 1. The original top 10 features for the classification of peanut stem rot (**A**) and the top 10 features with a custom minimum distance of 20 nm (**B**) selected by different feature selection methods from the scikit-learn machine-learning library in Python.

Rank/Methods	Selected Wavelengths (nm)				
	Chi-Square	**SFM_RF**	**SFM_SVML**	**RFE_RF**	**RFE_SVML**
(**A**) Original top 10 selected features					
1	698	496	884	501	505
2	702	884	759	884	396
3	706	665	807	505	302
4	694	501	767	274	391
5	595	690	743	620	261
6	590	686	838	735	653
7	599	826	763	247	514
8	603	505	850	686	884
9	586	628	694	645	763
10	611	492	803	690	830
(**B**) Top 10 selected features with a custom minimum distance					
1	698	496	884	501	505
2	595	884	759	884	396
3	632	665	807	274	302
4	573	690	838	620	261
5	527	826	694	735	653
6	552	628	649	247	884
7	657	242	242	686	763
8	719	518	731	645	830
9	505	607	674	779	431
10	678	274	586	826	624

Notes: Chi-square = SelectKBest (estimator = chi-square); SFM_RF = SelectFromModel (estimator = random forest); SFM_SVML = SelectFromModel (estimator = support vector machine with the linear kernel); RFE_RF = Recursive feature elimination (estimator = random forest); RFE_SVML = Recursive feature elimination (estimator = support vector machine with the linear kernel). Wavelengths selected repeatedly by different methods were highlighted in different colors (Purple = Ultraviolet region; Green = Green region; Red = Red or red-edge region; Grey = Near-infrared region).

**Top 10 wavelengths with and without custom distance**
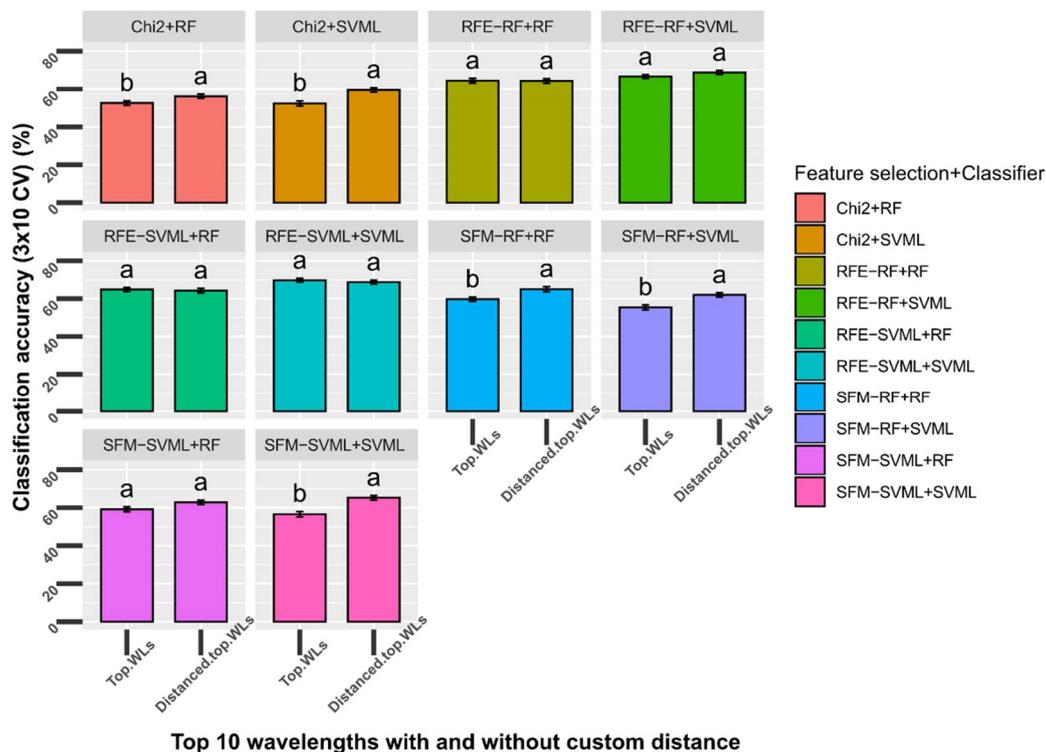
Figure 6. Comparison of top 10 selected wavelengths (WLs) with top 10 WLs with a minimum 20 nm distance between each feature for the classification of peanut stem rot. Five feature selection methods were tested: Chi2 = SelectKBest (estimator = chi-square); SFM-RF = SelectFromModel (estimator = random forest); SFM-SVML = SelectFromModel (estimator = support vector machine with linear kernel); RFE-RF = Recursive feature elimination (estimator = random forest); RFE-SVML = Recursive feature elimination (estimator = support vector machine with linear kernel). Each feature selection was tested using two classifiers: RF = random forest and SVML = support vector machine with the linear kernel. Bars with different letters were statistically different using a nonparametric Wilcoxon test with an α level of 0.05. Error bars indicate standard deviation of accuracy using stratified 10-fold cross-validation repeated three times.

3.6. Selected Wavelengths and Classification Accuracy for 5 Classes

The top 10 wavelengths selected by chi-square methods were limited to two narrow spectral regions (694 to 706 nm and 586 to 611 nm; Table 1). The top-four selected wavelengths were all within the narrow 694 to 706 nm range. Additional 20 nm limits helped to increase the diversity of the spectral wavelengths selected, such as adding 632 nm, 573 nm, and 527 nm in the top five selected features. However, no wavelengths were selected below 500 nm or above 720 nm with the chi-square method. Without 20 nm limits and using random forest, five out of the top-10 features were identical between the two feature selection methods. In contrast, using SVML, only three out of the top-10 features were identical between the two feature selection methods. Regardless of the machine learning method, one wavelength, 884 nm, was selected as an important feature in four method combinations (Table 1A). Among the rest of the features, 505 nm was selected by three methods, and 501, 690, 686, 694, and 763 nm were selected by two methods. There were more common wavelengths selected by different machine learning methods (six wavelengths) than between machine learning and chi-square methods (one wavelength).

By including a 20 nm minimum distance, three machine learning methods had the same top-three wavelengths and one machine learning method had the same top-two features, whereas the chi-square method only had one top feature (698 nm) remain the same because the original top-four features were all within a narrow bandwidth. Regardless of the type of machine learning method used, 884 nm was consistently an important feature. Another important feature was 496 to 505 nm, which was selected by three methods (Table 1B). If additional wavelengths can be included, other candidate ranges include

bands of 242 to 300 nm, 620 to 690 nm, 731 to 779 nm, or 807 to 826 nm. Regardless of whether using all bands or the top-10 selected wavelengths by RFE-SVML, the classification accuracy was low to medium for 'Presymptomatic' and 'Mild' (47 to 64%), medium to high for 'Lesion only' (64 to 91%), medium for 'Healthy' (59 to 77%), and high for 'Severe' (95%; Figure 7). The overall classification accuracy was 69.4% using all bands and 70.6% with the distanced top-10 selected wavelengths.

(**A**) (**B**)
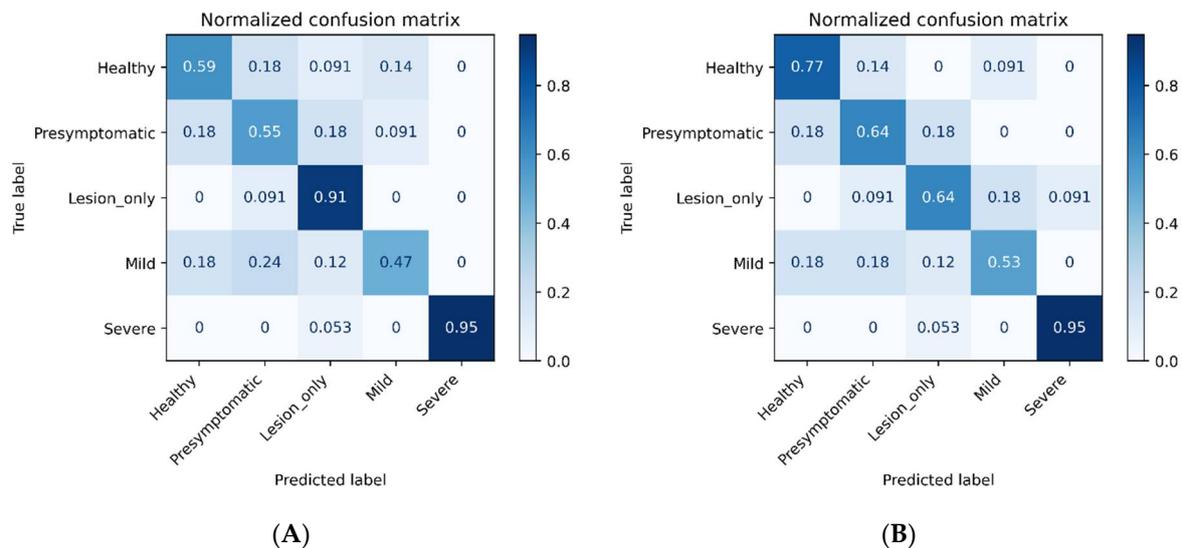
Figure 7. Confusion matrixes using all wavelengths (**A**) and the selected top-10 wavelengths (**B**) to classify mock-inoculated healthy peanut plants and plants inoculated with *Athelia rolfsii* at different stages of disease development. Feature selection method = recursive feature elimination with an estimator of the support vector machine, the linear kernel (RFE-SVML); Classifier = SVML.

4. Discussion

In this study, we compared nine methods to classify plants with different disease severities using reflectance wavelengths as input data. For two-class classification, all methods except for LDA performed well and similarly. One possible reason is that the LDA method is based on a linear combination of features and the classes cannot be separated using linear boundaries. For three-class classification, NB and MLPNN did not perform as well as other methods. NB assumes that different input features are independent of each other, which is not true for hyperspectral reflectance data. MLPNN is a neural network-based approach, which typically requires more samples for training due to high model complexity. For five-class classification, RF, SVM, GBoost, and XGBoost all perform similarly and slightly better than PLSDA. PLSDA has been widely used in spectral data analysis because these methods project the original data into latent structures that can help to remove the effect of multi-collinearity in the feature matrix. It is interesting that RF, SVM, and GBoost worked as well as PLSDA with our dataset; despite that, RF, SVM, and GBoost do not explicitly handle the dependency structure of the input features. These different methods also showed interesting differences in the important features that were selected to make the predictions, as discussed in the following paragraphs.

In this study, we also compared one feature-extraction method (PCA) and five feature-selection methods in the scikit-learn library to identify the most important wavelengths for the discrimination of healthy and diseased peanut plants infected with *A. rolfsii* under controlled conditions. A new method was also developed to select the top-ranked wavelengths with a custom distance. The distanced wavelengths method can utilize most of the wide spectral range covered by hyperspectral sensors and facilitate the design of disease-specific spectral indices and multispectral camera filters. Ultimately, the designed disease-specific spectral indices or filters can be used as a tool for detecting the beginning of

disease outbreaks in the field and guiding the implementation of preventative management tactics such as site-specific fungicide application.

Multiple feature-selection methods or an ensemble of multiple methods are recommended for application to hyperspectral data because the wavelengths selected are affected by the feature selection methods used [55,56]. Interestingly, SFM and RFE selected similar wavelengths with slightly different rankings using the RF estimator, but selected different wavelengths using the SVML estimator. Overall, the chi-square method tended to select adjacent wavelengths from relatively narrow spectral areas, while RFE methods with either an SVML or RF estimator selected wavelengths spread over different spectra regions. These machine learning-based methods are preferred because of the multicollinearity in hyperspectral data, especially among wavelengths in narrow spectral regions.

Several wavelengths were repeatedly selected as top features by different feature selection methods, including ones in the VIS region (501 and 505 nm), red-edge region (686, 690, and 694 nm), and NIR region (884 nm). Reflectance at 500 nm is dependent on combined absorption of chlorophyll a, chlorophyll b, and carotenoids, while reflectance around 680 nm is solely controlled by chlorophyll a [57,58]. Reflectance at 885 nm is reported to be sensitive to total chlorophyll, biomass, leaf area index, and protein [59]. The concentration of chlorophyll and carotenoid contents in the leaves may be altered during the disease progression of peanut stem rot, as indicated by discoloration and wilting of the foliage. The changes of leaf pigment contents may be caused by the damage to the stem tissues and accumulation of toxic acids in peanut plants during infection with *A. rolfsii*.

Wilting of the foliage is a common symptom of peanut plants infected with *A. rolfsii* [4] and plants under low soil moisture/drought stress [60–62]. However, compared to the wilting of the entire plant caused by drought, the typical foliar symptom of peanut plants during the early stage of infection with *A. rolfsii* is the wilting of an individual lateral branch or main stem [4]. In addition, the wilting of plants infected with *A. rolfsii* is believed to be caused by the oxalic acid produced during pathogenesis [63,64]. Our study identified two wavelengths including 505 and 884 nm. These wavelengths were used in a previous report on wheat under drought stress by Moshou et al. [65]. RGB color indices were reported recently to estimate peanut leaf wilting under drought stress [62]. To our knowledge, several other wavelengths identified in our study, including 242, 274, and 690 to 694 nm, have not been reported in other plant disease systems. Further studies should examine the capability of both the common and unique wavelengths identified in this study and previous reports to distinguish between peanut plants under drought stress and plants infected with soilborne pathogens including *A. rolfsii*.

Pathogen biology and disease physiology should be taken into consideration for plant disease detection and differentiation using remote sensing [8,37], which holds true also for hyperspectral band selection in the phenotyping of plant disease symptoms. Compared to the infection of foliar pathogens that have direct interaction with plant leaves, infection of soilborne pathogens will typically first damage the root or vascular systems of plants before inducing any foliar symptoms. This may explain why the wavelengths identified in this study for stem rot detection were different from ones reported previously for the detection of foliar diseases in various crops or trees [8,20,21,26,37,38,66]. Previously, charcoal rot disease in soybean was detected using hyperspectral imaging, and the authors selected 475, 548, 652, 516, 720, and 915 nm using a genetic algorithm and SVM [24]. Pathogen infection on stem tissues was measured via a destructive sampling method in the previous report [24], while reflectance of leaves on peanut stems infected with *A. rolfsii* was measured using a handheld hyperspectral sensor in a nondestructive manner in this study.

Regardless of whether using all bands or the distanced top-10 ranked wavelengths, the classification accuracy was low-to-medium for diseased peanut plants from the 'Presymptomatic' and 'Mild' classes. The five classes were categorized solely based on the visual inspection of the disease symptoms. It was not known if the infection occurred or not for inoculated plants in the 'Presymptomatic' class. The drooping of terminal leaves, the first foliar symptom observed, was found to be reversible and a potential response to heat

stress [45]. Plants with this reversible drooping symptom were included in the 'Mild' class, which may explain in part the low accuracy for this class. Future studies aiming to detect plant diseases during early infection may incorporate some qualitative or quantitative assessments of disease development using microscopy or quantitative polymerase chain reaction (qPCR) combined with hyperspectral measurements.

5. Conclusions

This study demonstrates the identification of optimal wavelengths using multiple feature-selection methods in the scikit-learn machine learning library to detect peanut plants infected with *A. rolfsii* at various stages of disease development. The wavelengths identified in this study may have applications in developing a sensor-based method for stem rot detection in peanut fields. The methodology presented here can be adapted to identify spectral signatures of soilborne diseases in different plant systems. This study also highlights the potential of hyperspectral sensors combined with machine learning in detecting soilborne plant diseases, serving as an exciting tool in developing integrated pest-management practices.

Author Contributions: Conceptualization, X.W., S.L., D.B.L.J., and H.L.M.; methodology, X.W., S.L., D.B.L.J., and H.L.M.; software, X.W., M.A.J., and S.L.; validation, X.W. and M.A.J.; formal analysis, X.W.; investigation, X.W.; resources, S.L., D.B.L.J., and H.L.M.; data curation, X.W. and M.A.J.; writing—original draft preparation, X.W.; writing—review and editing, X.W., M.A.J., S.L., D.B.L.J., and H.L.M.; visualization, X.W. and S.L.; supervision, D.B.L.J., H.L.M., and S.L.; project administration, X.W.; funding acquisition, D.B.L.J. and H.L.M. All authors have read and agreed to the published version of the manuscript.

References

1. Stalker, H. Peanut (*Arachis hypogaea* L.). *Field Crops Res.* **1997**, *53*, 205–217. [CrossRef]
2. Venkatachalam, M.; Sathe, S.K. Chemical composition of selected edible nut seeds. *J. Agric. Food Chem.* **2006**, *54*, 4705–4714. [CrossRef] [PubMed]
3. Porter, D.M. The peanut plant. In *Compendium of Peanut Diseases*, 2nd ed.; Kokalis-Burelle, N., Porter, D.M., Rodríguez-Kábana, R., Smith, D.H., Subrahmanyam, P., Eds.; The American Phytopathological Society: St. Paul, MN, USA, 1997; p. 1.

4.   Backman, P.A.; Brenneman, T.B. Stem rot. In *Compendium of Peanut Diseases*, 2nd ed.; Kokalis-Burelle, N., Porter, D.M., Rodríguez-Kábana, R., Smith, D.H., Subrahmanyam, P., Eds.; The American Phytopathological Society: St. Paul, MN, USA, 1997; pp. 36–37.

5.   Mullen, J. Southern blight, southern stem blight, white mold. *Plant Health Instr.* **2001**. [CrossRef]

6.   Punja, Z.K. The biology, ecology, and control of *Sclerotium rolfsii. Annu. Rev. Phytopathol.* **1985**, *23*, 97–127. [CrossRef]

7.   Agrios, G. *Encyclopedia of Microbiology*; Schaechter, M., Ed.; Elsevier Inc.: Amsterdam, The Netherlands, 2009; pp. 613–646.

8.   Gold, K.M.; Townsend, P.A.; Chlus, A.; Herrmann, I.; Couture, J.J.; Larson, E.R.; Gevens, A.J. Hyperspectral measurements enable pre-symptomatic detection and differentiation of contrasting physiological effects of late blight and early blight in potato. *Remote Sens.* **2020**, *12*, 286. [CrossRef]

9.   Zarco-Tejada, P.; Camino, C.; Beck, P.; Calderon, R.; Hornero, A.; Hernández-Clemente, R.; Kattenborn, T.; Montes-Borrego, M.; Susca, L.; Morelli, M. Previsual symptoms of *Xylella fastidiosa* infection revealed in spectral plant-trait alterations. *Nat. Plants* **2018**, *4*, 432–439. [CrossRef]

10.   Augusto, J.; Brenneman, T.; Culbreath, A.; Sumner, P. Night spraying peanut fungicides I. Extended fungicide residual and integrated disease management. *Plant Dis.* **2010**, *94*, 676–682. [CrossRef]

11.   Punja, Z.; Huang, J.-S.; Jenkins, S. Relationship of mycelial growth and production of oxalic acid and cell wall degrading enzymes to virulence in *Sclerotium rolfsii. Can. J. Plant Pathol.* **1985**, *7*, 109–117. [CrossRef]

12.   Weeks, J.; Hartzog, D.; Hagan, A.; French, J.; Everest, J.; Balch, T. *Peanut Pest Management Scout Manual*; ANR-598; Alabama Cooperative Extension Service: Auburn, AL, USA, 1991; pp. 1–16.

13.   Mehl, H.L. Peanut diseases. In *Virginia Peanut Production Guide*; Balota, M., Ed.; SPES-177NP; Virginia Cooperative Extension: Blacksburg, VA, USA, 2020; pp. 91–106.

14.   Mahlein, A.-K. Plant Disease Detection by Imaging Sensors—Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Dis.* **2016**, *100*, 241–251. [CrossRef]

15.   Mahlein, A.-K.; Kuska, M.T.; Behmann, J.; Polder, G.; Walter, A. Hyperspectral sensors and imaging technologies in phytopathology: State of the art. *Annu. Rev. Phytopathol.* **2018**, *56*, 535–558. [CrossRef]

16.   Bock, C.H.; Barbedo, J.G.; Del Ponte, E.M.; Bohnenkamp, D.; Mahlein, A.-K. From visual estimates to fully automated sensor-based measurements of plant disease severity: Status and challenges for improving accuracy. *Phytopathol. Res.* **2020**, *2*, 1–30. [CrossRef]

17.   Silva, G.; Tomlinson, J.; Onkokesung, N.; Sommer, S.; Mrisho, L.; Legg, J.; Adams, I.P.; Gutierrez-Vazquez, Y.; Howard, T.P.; Laverick, A. Plant pest surveillance: From satellites to molecules. *Emerg. Top. Life Sci.* **2021**. [CrossRef]

18.   Barreto, A.; Paulus, S.; Varrelmann, M.; Mahlein, A.-K. Hyperspectral imaging of symptoms induced by *Rhizoctonia solani* in sugar beet: Comparison of input data and different machine learning algorithms. *J. Plant Dis. Prot.* **2020**, *127*, 441–451. [CrossRef]

19.   Calderón, R.; Navas-Cortés, J.; Zarco-Tejada, P. Early detection and quantification of Verticillium wilt in olive using hyperspectral and thermal imagery over large areas. *Remote Sens.* **2015**, *7*, 5584–5610. [CrossRef]

20.   Gold, K.M.; Townsend, P.A.; Herrmann, I.; Gevens, A.J. Investigating potato late blight physiological differences across potato cultivars with spectroscopy and machine learning. *Plant Sci.* **2020**, *295*, 110316. [CrossRef] [PubMed]

21.   Gold, K.M.; Townsend, P.A.; Larson, E.R.; Herrmann, I.; Gevens, A.J. Contact reflectance spectroscopy for rapid, accurate, and nondestructive *Phytophthora infestans* clonal lineage discrimination. *Phytopathology* **2020**, *110*, 851–862. [CrossRef] [PubMed]

22.   Hillnhütter, C.; Mahlein, A.-K.; Sikora, R.; Oerke, E.-C. Remote sensing to detect plant stress induced by *Heterodera schachtii* and *Rhizoctonia solani* in sugar beet fields. *Field Crops Res.* **2011**, *122*, 70–77. [CrossRef]

23.   Hillnhütter, C.; Mahlein, A.-K.; Sikora, R.A.; Oerke, E.-C. Use of imaging spectroscopy to discriminate symptoms caused by *Heterodera schachtii* and *Rhizoctonia solani* on sugar beet. *Precis. Agric.* **2012**, *13*, 17–32. [CrossRef]

24.   Nagasubramanian, K.; Jones, S.; Sarkar, S.; Singh, A.K.; Singh, A.; Ganapathysubramanian, B. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods* **2018**, *14*, 1–13. [CrossRef] [PubMed]

25.   Rumpf, T.; Mahlein, A.-K.; Steiner, U.; Oerke, E.-C.; Dehne, H.-W.; Plümer, L. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput. Electron. Agric.* **2010**, *74*, 91–99. [CrossRef]

26.   Zhu, H.; Chu, B.; Zhang, C.; Liu, F.; Jiang, L.; He, Y. Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef] [PubMed]

27.   Thomas, S.; Kuska, M.T.; Bohnenkamp, D.; Brugger, A.; Alisaac, E.; Wahabzada, M.; Behmann, J.; Mahlein, A.-K. Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective. *J. Plant Dis. Prot.* **2018**, *125*, 5–20. [CrossRef]

28.   Carter, G.A.; Knapp, A.K. Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot.* **2001**, *88*, 677–684. [CrossRef] [PubMed]

29.   Jacquemoud, S.; Ustin, S.L. Leaf optical properties: A state of the art. In Proceedings of the 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, Aussois, France, 8–12 January 2001; pp. 223–332.

30.   Behmann, J.; Mahlein, A.-K.; Rumpf, T.; Römer, C.; Plümer, L. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precis. Agric.* **2015**, *16*, 239–260. [CrossRef]

31.   Miao, X.; Gong, P.; Swope, S.; Pu, R.; Carruthers, R.; Anderson, G.L. Detection of yellow starthistle through band selection and feature extraction from hyperspectral imagery. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1005–1015.

32.   Sun, W.; Du, Q. Hyperspectral band selection: A review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [CrossRef]

33. Dópido, I.; Villa, A.; Plaza, A.; Gamba, P. A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 421–435. [CrossRef]

34. Sun, W.; Halevy, A.; Benedetto, J.J.; Czaja, W.; Liu, C.; Wu, H.; Shi, B.; Li, W. UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 25–36. [CrossRef]

35. Hsu, P.-H. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 78–92. [CrossRef]

36. Yang, C.; Lee, W.S.; Gader, P. Hyperspectral band selection for detecting different blueberry fruit maturity stages. *Comput. Electron. Agric.* **2014**, *109*, 23–31. [CrossRef]

37. Mahlein, A.K.; Rumpf, T.; Welke, P.; Dehne, H.W.; Plümer, L.; Steiner, U.; Oerke, E.C. Development of spectral indices for detecting and identifying plant diseases. *Remote Sens. Environ.* **2013**, *128*, 21–30. [CrossRef]

38. Heim, R.; Wright, I.; Chang, H.C.; Carnegie, A.; Pegg, G.; Lancaster, E.; Falster, D.; Oldeland, J. Detecting myrtle rust (*Austropuccinia psidii*) on lemon myrtle trees using spectral signatures and machine learning. *Plant Pathol.* **2018**, *67*, 1114–1121. [CrossRef]

39. Yang, H.; Du, Q.; Chen, G. Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 544–554. [CrossRef]

40. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]

41. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [CrossRef]

42. Singh, A.; Jones, S.; Ganapathysubramanian, B.; Sarkar, S.; Mueller, D.; Sandhu, K.; Nagasubramanian, K. Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* **2020**. [CrossRef]

43. Samuel, A.L. Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* **1959**, *3*, 210–229. [CrossRef]

44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

45. Wei, X.; Langston, D.; Mehl, H.L. Spectral and thermal responses of peanut to infection and colonization with *Athelia rolfsii*. *PhytoFrontiers* **2021**. [CrossRef]

46. Shokes, F.; Różalski, K.; Gorbet, D.; Brenneman, T.; Berger, D. Techniques for inoculation of peanut with *Sclerotium rolfsii* in the greenhouse and field. *Peanut Sci.* **1996**, *23*, 124–128. [CrossRef]

47. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

48. Febrero Bande, M.; Oviedo de la Fuente, M. Statistical computing in functional data analysis: The R package fda. usc. *J. Stat. Softw.* **2012**, *51*, 1–28. [CrossRef]

49. Stevens, A.; Ramirez-Lopez, L. *An Introduction to the Prospectr Package*, R package Vignette R package version 0.2.1; 2020. Available online: http://bioconductor.statistik.tu-dortmund.de/cran/web/packages/prospectr/prospectr.pdf (accessed on 19 July 2021).

50. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

51. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

52. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; Team, R.C. *Caret: Classification and Regression Training*, R package version 6.0.-84; Astrophysics Source Code Library: Houghton, MI, USA, 2020.

53. Mevik, B.-H.; Wehrens, R.; Liland, K.H.; Mevik, M.B.-H.; Suggests, M. *pls: Partial Least Squares and Principal Component Regression*, R package version 2.7-2; 2020. Available online: https://cran.r-project.org/web/packages/pls/pls.pdf (accessed on 19 July 2021).

54. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

55. Chan, J.C.-W.; Paelinckx, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011. [CrossRef]

56. Hennessy, A.; Clarke, K.; Lewis, M. Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sens.* **2020**, *12*, 113. [CrossRef]

57. Gitelson, A.A.; Merzlyak, M.N. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *J. Plant Physiol.* **1996**, *148*, 494–500. [CrossRef]

58. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]

59. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; Van Der Meer, B. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* **2004**, *91*, 354–376. [CrossRef]

60. Balota, M.; Oakes, J. UAV remote sensing for phenotyping drought tolerance in peanuts. In Proceedings of the SPIE Commercial + Scientific Sensing and Imaging, Anaheim, CA, USA, 16 May 2017.

61. Luis, J.; Ozias-Akins, P.; Holbrook, C.; Kemerait, R.; Snider, J.; Liakos, V. Phenotyping peanut genotypes for drought tolerance. *Peanut Sci.* **2016**, *43*, 36–48. [CrossRef]

62. Sarkar, S.; Ramsey, A.F.; Cazenave, A.-B.; Balota, M. Peanut leaf wilting estimation from RGB color indices and logistic models. *Front. Plant Sci.* **2021**, *12*, 713. [CrossRef] [PubMed]

63. Higgins, B.B. Physiology and parasitism of *Sclerotium rolfsii* Sacc. *Phytopathology* **1927**, *17*, 417–447.
64. Bateman, D.; Beer, S. Simultaneous production and synergistic action of oxalic acid and polygalacturonase during pathogenesis by *Sclerotium rolfsii*. *Phytopathology* **1965**, *55*, 204–211. [PubMed]
65. Moshou, D.; Pantazi, X.-E.; Kateris, D.; Gravalos, I. Water stress detection based on optical multisensor fusion with a least squares support vector machine classifier. *Biosyst. Eng.* **2014**, *117*, 15–22. [CrossRef]
66. Conrad, A.O.; Li, W.; Lee, D.-Y.; Wang, G.-L.; Rodriguez-Saona, L.; Bonello, P. Machine Learning-Based Presymptomatic Detection of Rice Sheath Blight Using Spectral Profiles. *Plant Phenomics* **2020**, *2020*, 954085. [CrossRef]