

Classification of pine wilt disease at different infection stages by diagnostic hyperspectral bands

Niwen Li^{a,b,c}, Langning Huo^c, Xiaoli Zhang^{a,b,*}

^a Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China

^b The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing 100083, China

^c Department of Forest Resource Management, Swedish University of Agriculture Sciences, SE-901 83 Umeå, Sweden

ARTICLE INFO

Keywords:

Hyper spectral
Pine wilt disease
Infection stages
Diagnostic spectral bands screening
Forest health monitoring

ABSTRACT

Pine wilt disease (PWD) is a very destructive forest disease that causes the mortality of pine. The infected trees usually die within three months, and the disease spreads fast with the long-horned beetle as the medium if the infected trees are not removed from the forest in time. Therefore, detecting the infected trees at different infection stage, especially the early infection, is crucial for preventing PWD spread. This study aims to exhibit the spectral differences of the pine needles between healthy pines and infected pines at different infection stages and reveal the diagnostic spectral bands for classifying the different infected stage trees. We collected needle samples from healthy, early-, middle-, late-stage infected trees in a Japanese pine (*Pinus densiflora*) forest and a Korean pine (*Pinus koraiensis*) forest in northern China to explore the spectral and biochemical properties differences of these four classes, and selected the sensitive bands combining competitive adaptive reweighted sampling (CARS) and successive projections algorithm (SPA). The selected bands were used for the four infection stages classification by linear discriminant analysis (LDA) algorithm. The results show that Chlorophyll *a*, chlorophyll *b*, carotenoids, and moisture content decreases with the aggravation of infection. The green (510–530 nm), red-edge (680–760 nm), and short-wave infrared (1400–1420 nm and 1925–1965 nm) bands are the sensitive bands, and the overall accuracy is 77 % and 78 % for the Japanese pine and Korean pine respectively when using these bands for classifying healthy, early-, middle-, late-stage infected trees. The results demonstrate that physiological parameters including Chlorophyll *a*, chlorophyll *b*, carotenoids, and moisture content can be used as the diagnostic parameters of PWD, and the selected sensitive spectral bands are feasible for detecting the stress symptoms of the Japanese pine and Korean pine.

1. Introduction

Pine wilt disease (PWD) is a devastating disease caused by pinewood nematode (*Bursaphelenchus xylophilus*) (Futai, 2013; Kim et al., 2019). The disease causes rapid mortality with a high mortality rate, and is difficult to control and prevent from spreading, causing severe forest and ecological damage and serious economic losses (Tóth, 2011; Vicente et al., 2012). PWD is native to North America (Ikegami and Jenkins, 2018), but is now widely distributed in Asia (Hyun et al., 2007; Mamiya, 2004; Ye, 2019) and Europe (Abelleira et al., 2011; Fonseca et al., 2012; Mantas et al., 2022; Mota et al., 1999; Robertson et al., 2011). There are currently 52 countries that classify pinewood nematodes as quarantine pests (Escuer et al., 2004; Jones et al., 2008; Wu et al., 2020b). The first discovery of PWD in China was in the subtropical zone, but it has

extended to the north of China and the damage has aggravated with global warming and the impact of human factors caused by global trade. According to the Chinese Forestry Administration's No. 5 and No. 14 National Pinewood Nematode Epidemic Area Announcement in 2021, the disease has spread to 19 provinces including the regions in the temperate zone (Cheng et al., 2015; He et al., 2012; Li and Zhang, 2018). Therefore, a feasible and accurate disease monitoring method must be developed to better control the spread of the disease.

The nematode cannot be transmitted outside the wood independently but is spread by the main insect vector, pine sawyer longhorn-beetles (*Monochamus* spp.) during feeding and oviposition (Kim et al., 2019). Currently, there is no way to eradicate PWD, and timely felling of the infected trees and incineration or fumigation in situ is the most effective method for controlling its spread (Kim et al., 2011; Shin, 2008),

* Corresponding author at: Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China.

E-mail address: zhang-xl@263.net (X. Zhang).

<https://doi.org/10.1016/j.ecolind.2022.109198>

Received 11 May 2022; Received in revised form 9 July 2022; Accepted 19 July 2022

Available online 26 July 2022

1470-160X/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

which requires detecting the infected trees in a timely and accurate manner. Previous studies have shown the potential of using remote sensing techniques for forest damage detection and monitoring. For example, Syifa et al. (Syifa et al., 2020) used drone RGB images to detect dead trees with an accuracy of 86 % to 94 %. Tao et al. (Tao et al., 2020) used the UAV RGB images to identify dead trees due to PWD in pure forest with an accuracy of 65 % to 80 %. Zhang et al. (Zhang et al., 2021a) used revisited high-resolution satellite images of PlanetScope multispectral images with 3 m resolution to separate dead trees from PWD and other objects with an accuracy of 81.2 %. Mantas et al. (Mantas et al., 2022) used Sentinel-2 data to detect decline trees with an overall accuracy of 95 %. Zarco-Tejada et al. (Zarco-Tejada et al., 2019) used Sentinel-2A and hyperspectral imagery, tree decline was detected by chlorophyll content estimation. However, the above studies only detected the dead or declined trees, while detecting infected trees at the early stage with high accuracy to reduce the spread is still a challenge.

In the early stage of disease, the spectral difference between healthy trees and infected trees is small, while hyperspectral technology can acquire spectral information in hundreds of narrow bands in the visible, near-infrared (NIR), red-edge, and short-wave infrared (SWIR) ranges (Ghamisi et al., 2017). These narrow bands are highly sensitive to the subtle changes in the vegetation damaged by pests and diseases (Zhang et al., 2020b) or other reasons, so hyperspectral technology shows its great potentials for early detection of vegetation pest and disease. It has been widely used in agriculture pest and disease monitoring. Liu et al. (Liu et al., 2018) used ASD FieldSpec data to discriminate asymptomatic and symptomatic rice leaves, and the overall accuracies were 91.3 % – 93.1 %. Susiç et al. (Susiç et al., 2018) used hyperspectral images to discriminate abiotic and biotic drought stress of tomatoes, and the accuracies were 90 % – 100 %. Couture et al. (Couture et al., 2018) used non-imaging hyperspectral data to detect the Potato virus Y infected leaves with the mean validation kappa of 0.73. Polder et al. (Polder et al., 2019) used hyperspectral images to detect Potato Virus Y with the accuracy of over 78 %.

In recent years, PWD detection based on hyperspectral technology has attracted wide attention. Iordache et al. used airborne multispectral and hyperspectral images to identify suspicious infected trees with 83 % accuracy, demonstrating the feasibility of early identification of PWD (Iordache et al., 2020). Yu et al. used ground-based hyperspectral data and UAV hyperspectral data for early detection, with an accuracy of 57.69 % – 73.91 % and 48.28 % – 57.69 %, respectively (Yu et al., 2021c). And in another study, they used UAV hyperspectral images combined with LiDAR data for early detection with an accuracy of 50 % – 64.1 % (Yu et al., 2021b). Compared with the results of early detection using UAV-based multispectral imagery and RGB images with the accuracy of 42.36 % – 48.88 % (Yu et al., 2021a) and 46.5 % – 50.8 % (Wu et al., 2021), hyperspectral detection has high accuracy.

However, when using hyperspectral data for PWD monitoring, most studies directly use vegetation index or spectral characteristic parameters for identification. There is a lack of analysis of the spectral changes after being stressed by pinewood nematode. In addition, according to the physiological and ecological mechanism of PWD, the physiological properties at the different infection stages will change (Huang, 2020), which could be the important characteristics for PWD early detection. Hyperspectral data contains rich spectral signals for disease damage detection, but it also has spectral redundancy, so spectral change analysis and diagnostic spectral bands screening are the basis for early identification. Therefore, the screening of the diagnostic spectrum and the joint analysis with the physiological parameters are of great significance for PWD infection stages classification. However, both spaceborne data and airborne data, are affected by environmental factors such as atmosphere and moisture when acquiring data, which leads to measurement errors. Even if these errors can be corrected by pre-processing, the data will still be affected by canopy architecture and bidirectional reflectance distribution factor. Therefore, we use ground-based leaf hyperspectral data and physiological parameters to reveal the

mechanism and process of leaf spectrum and biochemical property changes after infection, which is important for large-scale PWD monitoring and control based on airborne or spaceborne optical remote sensing technology.

The objectives of this study can be formulated as the following:

- (1) Analyze the changes in spectral and physiological parameters of the pine needles following infection.
- (2) Obtain the diagnostic spectral bands and physiological parameters through spectral dimensionality reduction and correlation analysis.
- (3) Construct the detection models to provide an effective method to identify different infection stages.

2. Materials and methods

2.1. Study area

The two study areas were located in Weihai, Shandong province (37°25' – 37°25' N, 121°58' – 122°17' E) and Fushun, Liaoning province (41°14'– 42°28'N, 123°39'– 125°28'E), northeast China (Fig. 1). Weihai is located in the warm temperate zone and has a temperate monsoon climate with abundant rain and a moderate annual temperature. The average annual precipitation is 730.2 mm, and the average annual temperature is about 11.9°C. Fushun is located in the temperate zone and has a continental monsoon climate with hot summer and more rain, and a cold and long winter with a large temperature difference. The annual average precipitation is 804.2 mm, and the annual average temperature is about 6.6°C. The Japanese pine (*Pinus densiflora*) forest in Weihai and the Korean pine (*Pinus koraiensis*) forest in Fushun were largely attacked by PWD and caused substantial financial loss.

2.2. Field data

The field inventory was conducted in September 2019 in Weihai and August 2020 in Fushun. We selected the plots that have been verified being infected with pine wood nematode disease with the help of the local forestry staff. Ten plots in Weihai and five plots in Fushun were set up with a 25 m × 25 m size, containing pines at different infection stages. We randomly selected 10 – 20 sampling trees in each plot, and obtained samples from 106 Japanese pines and 83 Korean pines in Weihai and Fushun, respectively. Sampling is concentrated in 1–2 days to ensure that the trees do not decline over time. The Japanese pine dataset was randomly divided into a 70-sample training set and a 36-sample validation set, and the Korean pine dataset was randomly divided into a 55-sample training set and a 28-sample validation set. The training set was used for model training, and the validation set was used to test the accuracy of the model.

We classified the pines as healthy or infected at three different stages according to the colors of the needles as:

- (1) Healthy, with all needles green.
- (2) Early-stage infection, with less than half of the sampling needles turning yellow, or <1/3 of the sampling needles turning red.
- (3) Middle-stage infection, with more than half of the needles yellow, or the proportion of red needles between 1/3 and 3/4.
- (4) Late-stage infection, with >3/4 needles red.

We collected branches that roughly represent the average spectrum of each tree. Those branches of each sampled tree were cut from four directions from the upper, middle, and lower layers (Zhang et al., 2018). The numbers of healthy, early-stage, middle-stage, and late-stage infected trees were 33, 33, 18, and 22 in Weihai (Japanese pine) and 33, 25, 10, and 15 in Fushun (Korean pine).

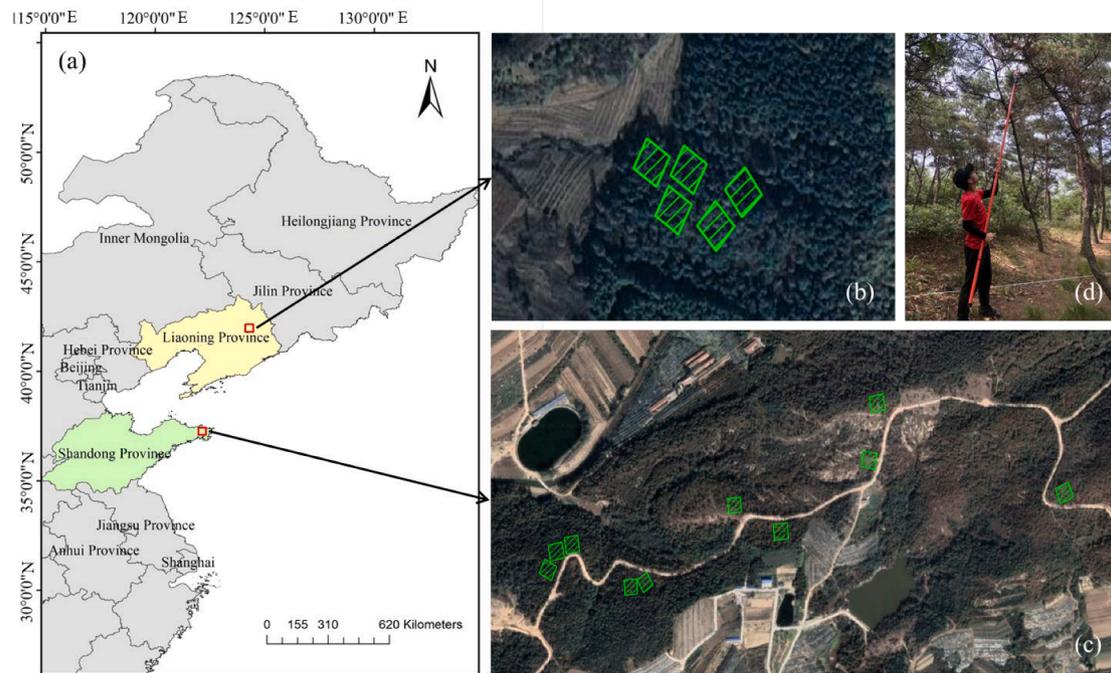


Fig. 1. Location of the study area (a), sample plots in Fushun City, Liaoning Province (b), sample plots in Weihai City, Shandong Province (c). The green squares represent the location of the sample plots. (d) is sampled using the high branch shear. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Physiological parameter measurements

The measured physiological parameters includes chlorophyll *a*, chlorophyll *b*, carotenoids, and moisture content. We cut 0.1 g of needles, soaked them in 95 % ethanol until colorless, and then diluted them to 50 ml. We measured the absorption values at 663 nm, 645 nm, and 470 nm using a spectrophotometer, and calculated the contents of chlorophyll *a*, chlorophyll *b*, and carotenoids by the Arnon formula (Duan, 1992). Each sample was measured three times, and the average value was used as the final result. In addition, to determine moisture content, the needles with a weight of m_1 were dried in an oven dryer at 105 °C for 30 min, then adjusted the temperature to 80 °C and dried to a constant weight. We measured the weight after drying, recorded it as m_2 , and used the following formula to calculate the moisture content:

$$\text{Moisture Content} = (m_1 - m_2) / m_2 * 100\% \tag{1}$$

2.4. Spectrum data acquisition

The spectral reflectance of the needles of sampling trees was measured indoors using a FiledSpec 4 HR NG spectrometer (Analytical Spectral Diveces, Inc. Boulder, Co, USA) ranging from 350 to 2500 nm (Table 1). The spectrum of each sampled tree is calculated from the average spectrum of collected branches. The needles were placed on a black cover and lighted only by the supporting light resource of the spectrometer. The spectrum values were averaged every ten scans as the spectrum of each sample to minimize the spectral signal noise. We calibrated the spectrometer using the white broad every ten minutes during the measurements. The measurements with spectral range < 400

nm and > 2400 nm were removed due to noise.

2.5. Sensitive bands selection

Hyperspectral data contains rich spectral information, but at the same time, the huge amount of data also causes redundancy in the data processing. Not all the bands are sensitive to PWD infection monitoring, so we need to select the sensitive bands without losing important information by suitable dimensionality reduction methods.

In this study, three algorithms are used for bands selection: competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), and the CARS-SPA algorithm. CARS is a variable selection method that combines Monte Carlo sampling with partial least squares (PLS) model regression coefficients over many sampling runs. In each sampling run, CARS performs four successive procedures, including Monte Carlo model sampling, exponentially decreasing function (EDF) forced variable removal, adaptive reweighted sampling (ARS) competing variable removal, and minimum root mean squared error of cross validation (RMSECV) calculation for each subset. Finally, the subset with the lowest RMSECV value is taken as the optimal wavelength subset (Li et al., 2009).

SPA is a forward selection method. It starts with one wavelength, and then incorporates a new wavelength at each iteration, until a specified number of wavelengths are reached. The principle of variable selection by SPA is that the selected new variable is one of all the remaining variables, which has the maximum projection value on the orthogonal subspace of the previous selected variables. Root mean square error (RMSE) is used as the evaluation criterion to determine the optimal bands. The purpose of SPA is to select the wavelength with the minimum

Table 1
Main parameters of FiledSpec 4 HR NG spectrometer.

Spectral Range (nm)	Spectral resolution (nm)	Spectral sampling (bandwidth) (nm)	Wavelength reproducibility (nm)	Wavelength accuracy (nm)	Field of view (°)
350 nm-2500 nm	3@700 6@1400/2100	1.4@350-1000 1.1@1001-2500	0.1	0.5	25

redundancy of information content in order to solve collinearity problems. The detailed description of SPA can be found in previous studies (Araújo et al., 2001; Soares et al., 2013; Xiaobo et al., 2010).

Combining two complementary band selection algorithms may achieve a superimposed effect (Su et al., 2018, 2019). The CARS algorithm runs fast, but the number of selected bands is large and there is redundancy. The SPA algorithm is time-consuming, but the selected bands are all valid. Therefore, we combine these two algorithms and call it “CARS-SPA”. This joint algorithm not only shortens the running time

of the algorithm, but also ensures that the selected bands are valid variables with minimum collinearity. This joint algorithm has been applied in crop growth monitoring and seed quality assessment (Wang et al., 2021; Zhang et al., 2021b). The results show that compared with using one algorithm alone, this joint algorithm can improve the band selection efficiency by 0.78 % – 20.23 %. Currently, this combined algorithm has not been applied to the spectral selection of forest pests and diseases. The algorithm was implemented using MATLAB 2018a.

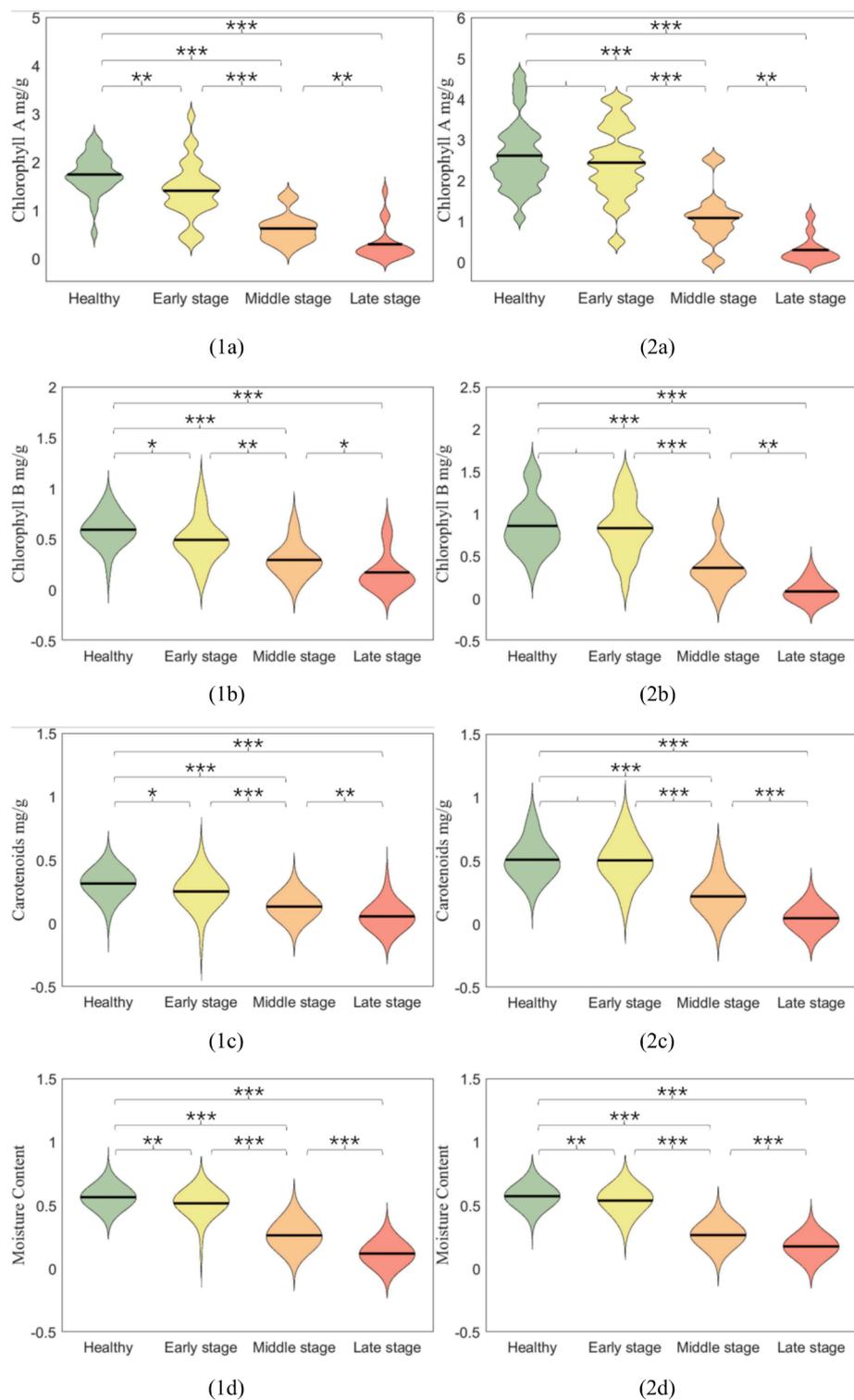


Fig. 2. Physiological parameter content at different infection stages of Japanese pine (1), Korean pine (2), and (a) Chlorophyll a, (b) Chlorophyll b, (c) Carotenoids, (d) Moisture content. The black line in the figure is the average value. Statistical significance differences with $p > 0.05$ were marked as null. Statistical significance differences with $0.01 \leq p < 0.05$ were marked as “*”. Statistical significance differences with $0.001 \leq p < 0.01$ were marked as “***”. Statistical significance differences with $p < 0.001$ were marked as “****”.

2.6. Classification and validation

In this study, two algorithms are used for classification: linear discriminant analysis (LDA) and support vector machine (SVM). The LDA algorithm is used to model after band screening, and the SVM algorithm is used to model in all bands.

The LDA is a linear projection technique proposed by Fisher (Fisher, 1936). Its principle is to project the training samples onto a straight line during training, which can make the projection points of the same type of sample as close as possible and the projection points of different types of samples as far away as possible. During prediction, the data to be predicted is projected onto the straight line learned above, and the category it belongs to is determined according to the position of the projection point (Tharwat et al., 2017). In order to eliminate the error caused by the division of the sample set as much as possible, when modeling with LDA, we conduct 10 experiments, and take the average accuracy as the final accuracy. In each experiment, the training set and the verification set were randomly assigned.

The SVM is a supervised nonparametric statistical learning technique, which was introduced within the framework of statistical learning theory developed by Vapnik (Cortes and Vapnik, 1995; Mountrakis et al., 2011). SVM is effective for high-dimensional feature classification, even when the feature dimension is larger than the number of samples. Therefore, we used SVM for full-band classification. The SVM algorithm is implemented by LIBSVM 3.25 software (Chang and Lin, 2011). The division of training set and validation set was written in section 2.2. In this study, the SVM type is C-Support Vector Classification (C-SVC). The kernel functions used include linear kernel function, polynomial kernel function and radial basis function (RBF) kernel function.

The accuracy of the model is evaluated using the correct identification rate (CIR), which is the percentage of the number of correctly identified samples in the total number of samples. The larger the value, the better the recognition effect of the model.

3. Results

3.1. Physiological parameters at different infection stages

With the development of infection, chlorophyll *a*, chlorophyll *b*, carotenoids, and moisture content tended to decrease (Fig. 2). All parameters showed statistically significant differences in early and middle,

middle and late stages, so the spectral absorption related to chlorophyll and water would show some differences in the spectral curves of different infection stages. The moisture content showed significant differences between healthy and early stage of the two tree species, so the parameters of spectral absorption related to water have the greatest potential for early identification. For the Japanese pine, chlorophyll *a* also showed good potential for early identification.

3.2. Reflectance spectrum and sensitive bands

The needles from healthy and early-stage infected trees showed a similar spectrum to the typical vegetation spectrum. With the development of infection, the spectral curves gradually lost the following features (Fig. 3):

- (1) The absorption spectrum around the green peak caused by Chlorophyll absorption at the green bands and red bands disappeared (Zone A and B).
- (2) The slope at the red edge was gentler than the healthy samples.
- (3) The absorption valleys in the NIR bands weakened due to change of cell structure. (Zone C and D).
- (4) The reflectance at the SWIR bands was higher than in healthy samples due to low water absorption. The absorption valleys at zones E, F, and G were less obvious than in healthy samples.

By comparing the two tree species, we found that the spectral reflectance values of Korean pine were higher than those of Japanese pine at all stages. In the NIR band, the average spectrum curve of Japanese pine in the middle stage was higher than that in the late stage, while the average spectrum curve of Korean pine in the middle stage is almost the same as that in the late stage. Although there are some differences in the spectra of the two species, the changes in each absorption valley and red edge at different susceptible stages are the same.

The mean reflectance in the red, NIR and SWIR bands showed big differences among different stages. However, the reflectance at the NIR bands also showed high standard deviation (Fig. 4), i.e., large variance within groups, making classifying different stages challenging. The reflectance at the red-edge bands showed low differences among different stages, but also low standard deviation within stages, thus it could potentially contribute to the classification (Fig. 5).

In terms of tree species, we found that the spectrum of Japanese pine at various susceptible stages and healthy trees has little difference in the

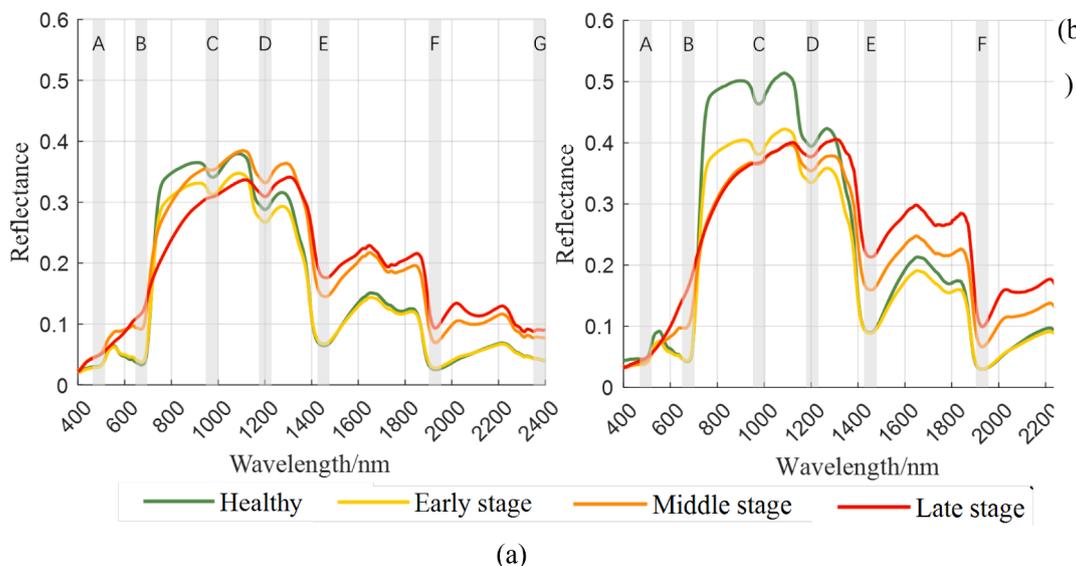


Fig. 3. The average reflectance spectrum of samples at different stages of infection: (a) Japanese pine, (b) Korean pine. Spectral absorption valleys are marked in grey (A-G).

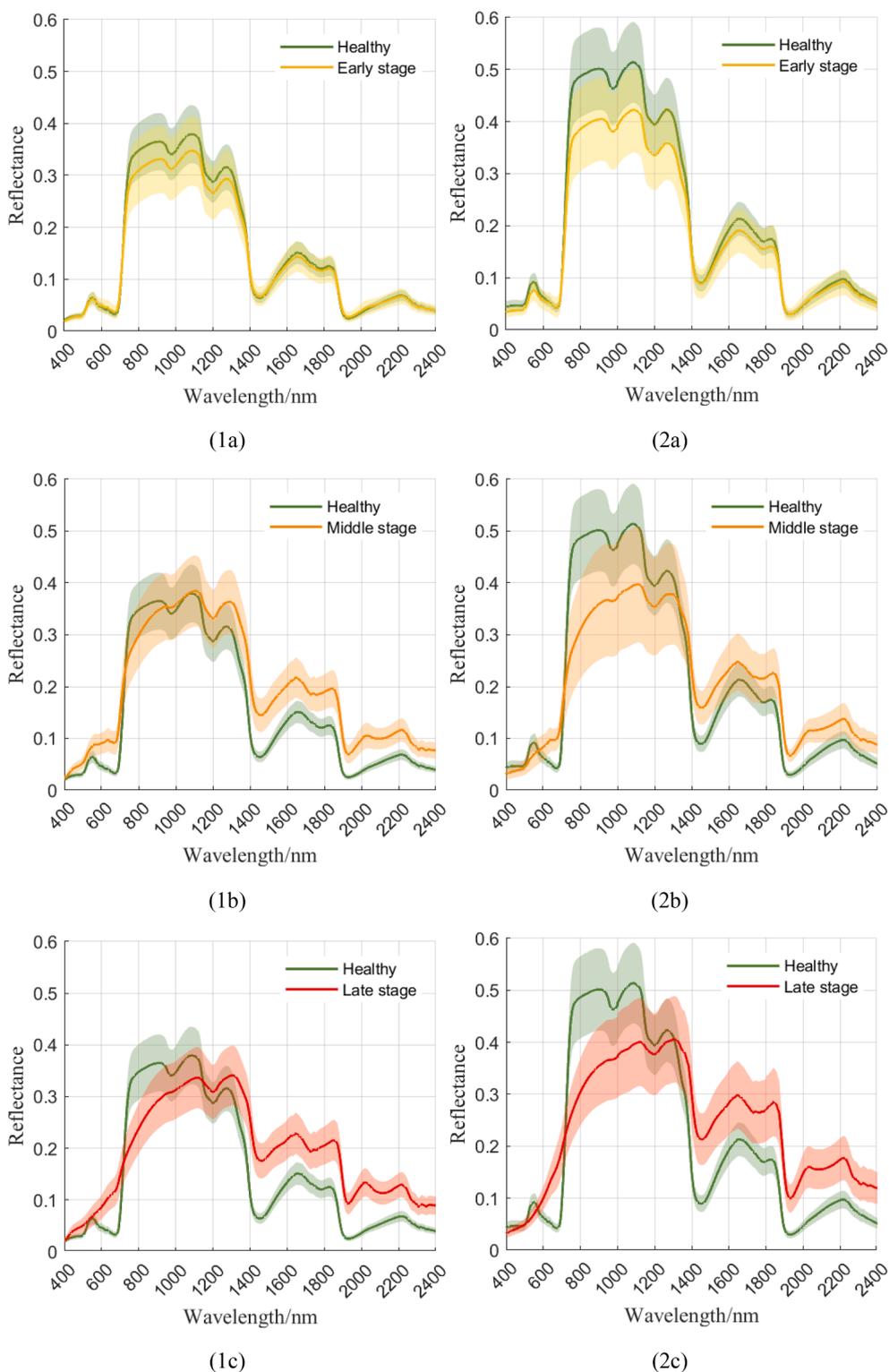


Fig. 4. The mean and standard deviation of the spectrum of Japanese pine (1) and Korean pine (2) needles from (a) healthy trees and early-stage infected trees, (b) healthy trees and middle-stage infected trees, and (c) healthy trees and late-stage infected trees.

NIR band, especially in the early stages, which may lead to a low early recognition rate. The spectrum of Korean pine at each susceptible stage was quite different from that of healthy trees.

When using SPA to select the sensitive bands, six and twelve bands were selected for Japanese pine and Korean pine, respectively. The sensitive bands of Japanese pine corresponded to the chlorophyll absorption bands around the green peak, the red-edge bands, and the water

absorption zones in the SWIR bands. In addition to the above bands, the sensitive band of Korean pine also has the absorption zones in the NIR bands. When using CARS to select the sensitive bands, fifteen and fourteen bands were selected for Japanese pine and Korean pine, respectively. The sensitive bands of Korean pine are the chlorophyll absorption bands, the red-edge bands, and the water absorption zones in the SWIR bands. In addition to the above bands, the sensitive bands of

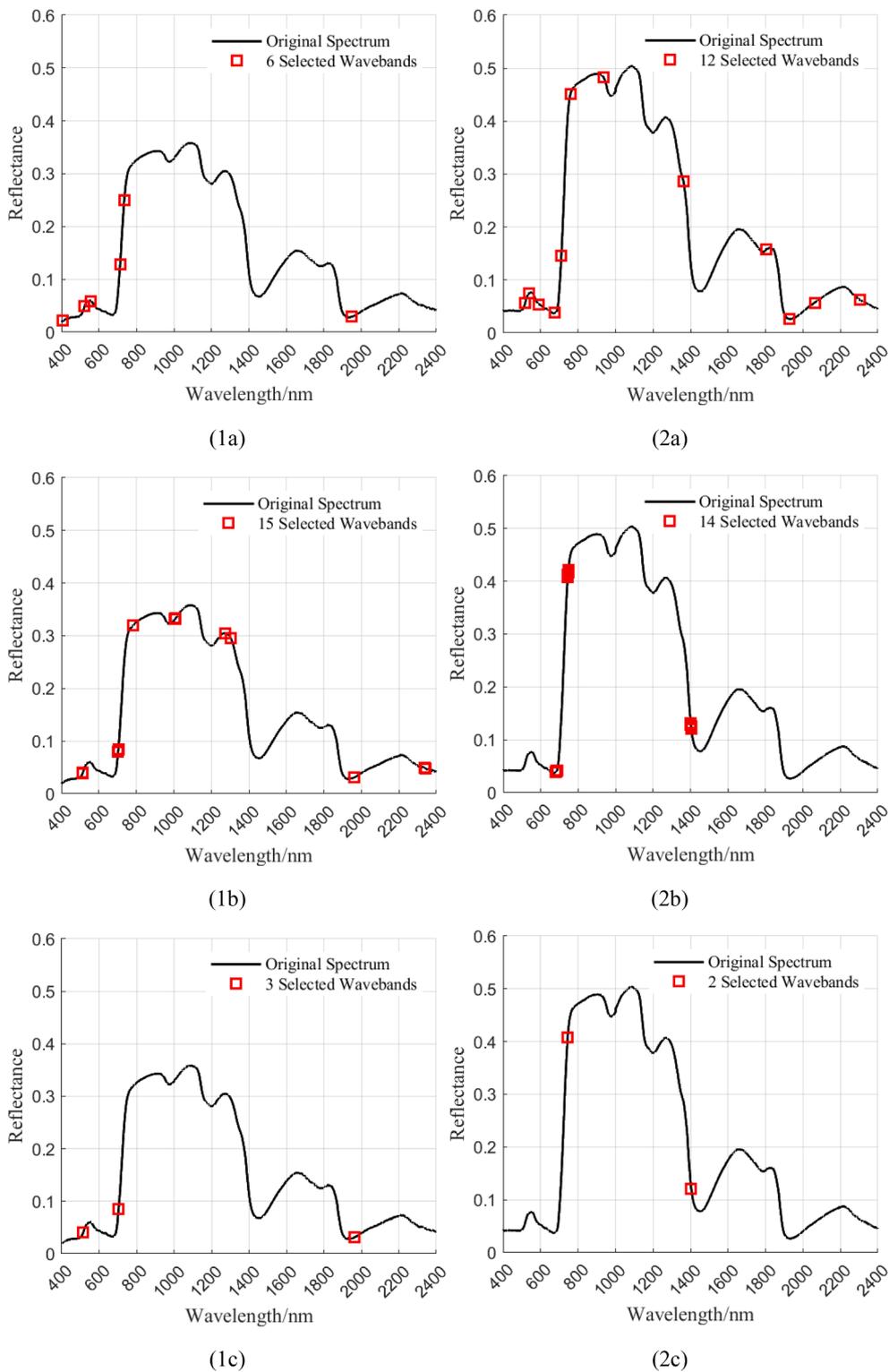


Fig. 5. The sensitive bands of Japanese pine (1), Korean pine (2) selected by (a) SPA, (b)CARS, (c) CARS-SPA.

Japanese pine also include the absorption bands in the NIR bands. In summary, all the above sensitive bands include the chlorophyll absorption bands around the green peak, the red-edge bands, the two absorption zones in the NIR bands, and the three water absorption zones in the SWIR bands. When using CARS-SPA for the band selection, three and two bands were sensitive to the Japanese pine and Korean pine, respectively. Compared to the SPA and CARS selections, the slopes of the spectrum curves caused by absorptions were judged to be more sensitive

than the peak or valley values.

3.3. Infection stage classification

All bands were used for classification by the SVM algorithm, and the selected bands from SPA, CARS, and CARS-SPA were used for classification by the LDA algorithm. The results showed that there was overfitting when using the SVM algorithm in all bands, and the linear

kernel function showed the best classification accuracy, which was 0.78 for Japanese pine and 0.79 for Korean pine. Among the three band selection algorithms, CARS-SPA used the least number of bands but had the highest classification accuracy, with the accuracy of Japanese pine and Korean pine being 0.77 and 0.78, respectively (Table 2). Classification using CARS-SPA selected bands (3 and 2 respectively for Japanese pine and Korean pine) were only 0.01 less accurate than classification using the full band, demonstrating the effectiveness of the selected bands. The CARS-SPA-LDA models separated middle- and late-stage infected samples with high accuracy, but yielded low accuracy in separating early-stage samples from the healthy ones.

4. Discussion

Hyperspectral data provides rich spectral information with a large number of narrow bands. It can capture the subtle changes in vegetation and is thus widely used for disease detection and monitoring in forestry and agriculture. Some of the bands respond earlier and more significantly during the decrease in vitality. Therefore, selecting the most responsive bands and applying them to disease detection is crucial to decreasing data redundancy and avoiding overfitting. In this study, the SPA and CARS algorithms were used to select sensitive bands responding to different infection stages of PWD. We also used a CARS-SPA algorithm for the band selection, resulting in the least number of bands chosen and higher classification accuracy compared to when using CARS and SPA individually. This algorithm was also used to determine the fruit soluble solids content and cultivated soil moisture content (Wu et al., 2020; Zhang et al., 2019; Zhang et al., 2020a).

This study illustrated that the red-edge bands were very sensitive to different infection stages of PWD, followed by the 510–530 nm band of green light, the 1 400–1 420 nm of SWIR bands and the 1 925–1 965 nm of SWIR bands. This result partially agrees with previous studies. For example, a study measured the spectra of trees inoculated with PWD using a GER-3700 spectrometer and observed the increased reflectance in the red and SWIR bands after 67 days of inoculation (Kim et al., 2018). Lee et al. (Lee et al., 2014) analyzed the spectral characteristics of different infection stages of PWD using a ground hyperspectral camera and the 688 nm band was found to be the optimal band, while in another study, the mid-infrared bands were shown to be the most sensitive bands for the early stage.

During the infection of PWD, the water transportation is blocked by the PWD, which leads to a decrease of water content in the crowns, and gradually slows down the photosynthesis, which thus leads to a decrease in chlorophyll content (Fukuda, 1997; Kim et al., 2018; Wu et al., 2020b). In this study, the same results were obtained by analyzing the differences in chlorophyll content and moisture content in different susceptible stages, that is, with the development of the susceptible stage, both chlorophyll and water content decreased. There is a significant difference in moisture content between healthy and early stages, so the

related spectral parameters have good potential for early identification. This result was consistent with the research of other stem borers (Abdullah et al., 2018; Liu et al., 2021). The absorption spectrum around the green peak and the SWIR bands reflected the decreased water and chlorophyll content (Jones and Vaughan, 2010; Roberts et al., 2012). The NIR bands were sensitive to the cell structure, which thus showed considerable variation within the group and less depending on the infection stages. On the contrary, the red-edge bands showed little variation within the group and were thus more sensitive to the infection stages. The classification accuracy was similar when using the red-edge bands with or without NIR bands. Therefore, we considered the red-edge bands to respond better to the infection than the NIR bands.

In this study, we classified healthy and infected trees at three stages with the selected bands. The classification performed well in distinguishing between healthy and infected trees at the middle and late stages, while separating healthy and early-stage infection was more challenging. Other studies have reached the same conclusion (Yu et al., 2021b; Yu et al., 2021c). An infected tree usually turns red and dies within three months, while a study has shown that the spectra of the needles did not change significantly until 67 days after the PWD inoculation (Kim et al., 2018), which exhibited the difficulties of early detection. We used the LDA algorithm for the classification, considering many selected bands from the CARS-SPA algorithm as the variables. The result showed that compared with SVM using full-band classification, the classification using fewer selected bands also has similar accuracy. We assume using vegetation indices by the bands selected from the CARS-SPA algorithm (green, red, red-edge or SWIR bands) could also achieve the early classification which needs further studies. Since the focus of this study is revealing the spectra differences and exploring the accuracy differences between the models using different numbers of bands, we did not use other models than the LDA to reduce the influence caused by the model selection.

One shortage in this study is that the infection was not verified by the presence of PWD in the lab for every tree, but we set up plots with infected trees verified in the past, and assumed the declined trees were infected by PWD based on the spreading characteristic of PWD. This might cause errors in the spectral analysis if other factors caused the discoloration, but we considered the probability very low. It has been shown that trees that decline for PWD and trees that decline for other reasons do not have unique traits in the spectrum (Mantas et al., 2022). Therefore, we hope this shortage is not so big to influence this study providing some insights on PWN.

The needle spectra presented in this study could be a reference for further studies using hyperspectral or multispectral remote sensing data, such as drone and satellite images. The spectra in remote sensing images, particularly satellite images, may be influenced by the atmosphere, branches, canopy gaps, and surrounding trees; thus, the spectral difference between healthy and infected trees would be less dependent on their vitality. Therefore, the hyperspectral measurement from this study

Table 2

Classification accuracy of Japanese pine and Korean pine samples using different selected bands and all bands. H, E, M, L represent the classes of healthy, early-, middle-, and late-stage infected samples.

Modeling method	Band select Algorithm	Modeling algorithm	Number of bands selected	Training set					Validation set				
				H	E	M	L	Total	H	E	M	L	Total
Japanese pine	All bands	SVM-Linear	2001	0.95	0.81	1.00	1.00	0.93	0.82	0.58	1.00	0.89	0.78
		SVM- Polynomial	2001	0.95	1.00	1.00	1.00	0.99	0.73	0.83	1.00	0.56	0.75
		SVM- RBF	2001	0.95	0.90	1.00	1.00	0.96	0.27	0.42	1.00	0.78	0.53
	SPA	LDA	6	0.84	0.68	0.90	0.92	0.82	0.73	0.63	0.85	0.83	0.75
	CARS	LDA	15	0.90	0.87	0.95	0.97	0.91	0.70	0.63	0.79	0.79	0.72
	CARS-SPA	LDA	3	0.80	0.62	0.87	0.93	0.78	0.85	0.58	0.83	0.90	0.77
Korean pine	All bands	SVM-Linear	2001	0.82	0.76	1.00	1.00	0.85	0.91	0.75	0.40	1.00	0.79
		SVM- Polynomial	2001	0.86	1.00	1.00	1.00	0.95	0.27	0.88	0.40	1.00	0.57
		SVM- RBF	2001	0.86	0.82	0.60	1.00	0.85	0.82	0.63	0.00	1.00	0.64
	SPA	LDA	12	0.87	0.81	0.96	0.88	0.87	0.85	0.60	0.73	0.71	0.74
	CARS	LDA	14	0.91	0.81	0.88	0.92	0.88	0.75	0.67	0.76	0.88	0.74
	CARS-SPA	LDA	2	0.84	0.73	0.91	0.79	0.81	0.86	0.65	0.82	0.87	0.78

could be a good reference to highlight the possible spectra changes due to infection rather than other factors. In the following research, we will compare the sensitive bands and the classification performance from the spectrometer and drone images, and explore which bands lost the sensitivity to infections on the remote sensing images. Achieving early detection with higher accuracy is another topic that needs further study. We consider the deviations of the spectra have the potential to perform better early detection based on the spectra we observed in this study.

5. Conclusion

This study analyzed the spectra and biochemical properties of needles from Japanese pines (*Pinus densiflora*) and Korean pines (*Pinus koraiensis*) with early, middle, and late stages of PWD infection. The results showed that chlorophyll *a*, chlorophyll *b*, carotenoids, and moisture content decreased during the infection, and water content had the most significant difference between healthy and early stage. The CARS-SPA spectral dimensionality reduction algorithm showed that the green (510–530 nm), red-edge (680–760 nm), and short-wave infrared (1400–1420 nm and 1925–1965 nm) bands were the diagnostic spectral bands. By using the LDA algorithm, the healthy, early-, middle-, and late-stage infected trees were classified with an overall accuracy of 77 % and 78 % for the Japanese pine and Korean pine respectively. This study provides an effective method to identify different infection stages, and the needle spectra presented in this study provide a reference for further studies using hyperspectral or multispectral remote sensing data, such as drone and satellite images.

Funding

This research was funded by the National Natural Science Foundation of China (31870534), and the ESA-MOST Dragon 5 Cooperation (59257).

CRediT authorship contribution statement

Niwen Li: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Visualization.
Langning Huo: Writing – original draft, Writing – review & editing.
Xiaoli Zhang: Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

References

- Abdullah, H., Darvishzadeh, R., Skidmore, A.K., Groen, T.A., Heurich, M., 2018. European spruce bark beetle (*Ips typographus*, L.) green attack affects foliar reflectance and biochemical properties. *International journal of applied earth observation and geoinformation* 64, 199–209.
- Abelleira, A., Picoaga, A., Mansilla, J., Aguin, O., 2011. Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain. *Plant disease* 95, 776–776.
- Araújo, M.C.U., Saldanha, T.C.B., Galvão, R.K.H., Yoneyama, T., Chame, H.C., Visani, V., 2001. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics Intell. Lab. Syst.* 57, 65–73.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2, 1–27.
- Cheng, G., Lü, Q., Feng, Y., Li, Y., Wang, Y., Zhang, X., 2015. Temporal and spatial dynamic pattern of pine wilt disease distribution in China predicted under climate change scenario. *Scientia Silvae Sinicae* 51, 119–126.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine learning* 20, 273–297.
- Couture, J., Singh, A., Charkowski, A., Groves, R., Gray, S., Bethke, P., Townsend, P., 2018. Integrating spectroscopy with potato disease management. *Plant disease* 102, 2233–2240.
- Duan, G., 1992. Derivation of Arnon formula in the determination of chlorophyll content. *Plant Physiology Communications* 28, 1.
- Escuer, M., Arias, M., Bello, A., 2004. Occurrence of the genus *Bursaphelenchus* Fuchs, 1937 (Nematoda: Aphelenchida) in Spanish conifer forests. *Nematology* 6, 155–156.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics* 7, 179–188.
- Fonseca, L., Cardoso, J., Lopes, A., Pestana, M., Abreu, F., Nunes, N., Mota, M., Abrantes, I., 2012. The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia* 49, 96–103.
- Fukuda, K., 1997. Physiological process of the symptom development and resistance mechanism in pine wilt disease. *Journal of forest research* 2, 171–181.
- Futai, K., 2013. Pine wood nematode, *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 51, 61–83.
- Ghamisi, P., Plaza, J., Chen, Y., Li, J., Plaza, A.J., 2017. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine* 5, 8–32.
- He, S., Wen, J., Luo, Y., Zong, S., Zhao, Y., Han, J., 2012. The predicted geographical distribution of *Bursaphelenchus xylophilus* in China under climate warming. *Chinese Bulletin of Entomology* 49, 236–243.
- Huang, B., 2020. Monitoring *Bursaphelenchus xylophilus* with multispectrum camera in UAV. *Guangxi Forest Sci* 49, 380–384.
- Hyun, M.W., Kim, J.H., Suh, D.Y., Lee, S.K., Kim, S.H., 2007. Fungi isolated from pine wood nematode, its vector Japanese pine sawyer, and the nematode-infected Japanese black pine wood in Korea. *Mycobiology* 35, 159–161.
- Ikegami, M., Jenkins, T.A., 2018. Estimate global risks of a forest disease under current and future climates using species distribution model and simple thermal model–Pine Wilt disease as a model case. *Forest ecology and management* 409, 343–352.
- Iordache, M.-D., Mantas, V., Baltazar, E., Pauly, K., Lewycky, N., 2020. A Machine Learning Approach to Detecting Pine Wilt Disease Using Airborne Spectral Imagery. *Remote Sensing* 12, 2280.
- Jones, J.T., Moens, M., Mota, M., Li, H., Kikuchi, T., 2008. *Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host–parasite interactions. *Molecular plant pathology* 9, 357–368.
- Jones, H.G., Vaughan, R.A., 2010. *Remote sensing of vegetation: principles, techniques, and applications*. Oxford University Press.
- Kim, N., Jeon, H.W., Manna, M., Jeong, S.I., Kim, J., Kim, J., Lee, C., Park, A., Kim, J.C., Seo, Y.S., 2019. Induction of resistance against pine wilt disease caused by *Bursaphelenchus xylophilus* using selected pine endophytic bacteria. *Plant Pathology* 68, 434–444.
- Kim, M.-I., Lee, W.-K., Kwon, T.-H., Kwak, D.-A., Kim, Y.-S., Lee, S.-H., 2011. Early detecting damaged trees by pine wilt disease using DI (Detection Index) from portable near infrared camera. *Journal of Korean Society of Forest Science* 100, 374–381.
- Kim, S.R., Lee, W.K., Lim, C.H., Kim, M., Kafatos, M.C., Lee, S.H., Lee, S.S., 2018. Hyperspectral Analysis of Pine Wilt Disease to Determine an Optimal Detection Index. *Forests* 9, 12.
- Lee, J.B., Kim, E.S., Lee, S.H., 2014. An analysis of spectral pattern for detecting pine wilt disease using ground-based hyperspectral camera. *Korean J. Remote Sensing* 30, 665–675.
- Li, H., Liang, Y., Xu, Q., Cao, D., 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta* 648, 77–84.
- Li, Y., Zhang, X., 2018. Analysis on the trend of invasion and expansion of *Bursaphelenchus xylophilus*. *For. Pest Dis* 1, 1–4.
- Liu, Z.-Y., Qi, J.-G., Wang, N.-N., Zhu, Z.-R., Luo, J., Liu, L.-J., Tang, J., Cheng, J.-A., 2018. Hyperspectral discrimination of foliar biotic damages in rice using principal component analysis and probabilistic neural network. *Precision Agriculture* 19, 973–991.
- Liu, Y.J., Zhan, Z.Y., Ren, L.L., Ze, S.Z., Yu, L.F., Jiang, Q., Luo, Y.Q., 2021. Hyperspectral evidence of early-stage pine shoot beetle attack in Yunnan pine. *Forest Ecology and Management* 497.
- Mamiya, Y., 2004. Pine wilt disease in Japan, The pinewood nematode. *Bursaphelenchus xylophilus*. *Brill* 9–20.
- Mantas, V., Fonseca, L., Baltazar, E., Canhoto, J., Abrantes, I., 2022. Detection of Tree Decline (*Pinus pinaster* Aiton) in European Forests Using Sentinel-2 Data. *Remote Sensing* 14, 2028.
- Mota, M.M., Braasch, H., Bravo, M.A., Penas, A.C., Burgermeister, W., Metge, K., Sousa, E., 1999. First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1, 727–734.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 247–259.
- Polder, G., Blok, P.M., de Villiers, H.A., van der Wolf, J.M., Kamp, J., 2019. Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. *Frontiers in plant science* 10, 209.
- Roberts, D., Roth, K., Perroy, R., 2012. *hyperspectral vegetation indices. Hyperspectral Remote Sensing of Vegetation*, 309. CRC Press/Taylor and Francis Group.
- Robertson, L., Arcos, S.C., Escuer, M., Merino, R.S., Esparrago, G., Abelleira, A., Navas, A., 2011. Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhrer, 1934 (Nickle, 1970) in Spain. *Nematology* 13, 755–757.
- Shin, S.-C., 2008. Pine wilt disease in Korea. *Pine wilt disease*. Springer 26–32.
- Soares, S.F.C., Gomes, A.A., Galvao, A.R., Araujo, M.C.U., Galvao, R.K.H., 2013. The successive projections algorithm. *Trac-Trends Anal. Chem.* 42, 84–98.

- Su, W.-H., Bakalis, S., Sun, D.-W., 2018. Fourier transform mid-infrared-attenuated total reflectance (FTMIR-ATR) microspectroscopy for determining textural property of microwave baked tuber. *Journal of Food Engineering* 218, 1–13.
- Su, W.-H., Bakalis, S., Sun, D.-W., 2019. Chemometrics in tandem with near infrared (NIR) hyperspectral imaging and Fourier transform mid infrared (FT-MIR) microspectroscopy for variety identification and cooking loss determination of sweet potato. *Biosystems Engineering* 180, 70–86.
- Susić, N., Žibrat, U., Širca, S., Strajnar, P., Razingger, J., Knapić, M., Vončina, A., Urek, G., Stare, B.G., 2018. Discrimination between abiotic and biotic drought stress in tomatoes using hyperspectral imaging. *Sensors and actuators B: Chemical* 273, 842–852.
- Syifa, M., Park, S.-J., Lee, C.-W., 2020. Detection of the pine wilt disease tree candidates for drone remote sensing using artificial intelligence techniques. *Engineering* 6, 919–926.
- Tao, H., Li, C., Zhao, D., Deng, S., Jing, W., 2020. Deep learning-based dead pine trees detection from unmanned aerial vehicle images. *International Journal of Remote Sensing*.
- Tharwat, A., Gaber, T., Ibrahim, A., Hassanien, A.E., 2017. Linear discriminant analysis: A detailed tutorial. *AI communications* 30, 169–190.
- Tóth, Á., 2011. *Bursaphelenchus xylophilus*, the pinewood nematode: its significance and a historical review. *Acta Biologica Szegediensis* 55, 213–217.
- Vicente, C., Espada, M., Vieira, P., Mota, M., 2012. Pine wilt disease: a threat to European forestry. *European journal of plant pathology* 133, 89–99.
- Wang, Z.L., Fan, S.X., Wu, J.Z., Zhang, C., Xu, F.Y., Yang, X.H., Li, J.B., 2021. Application of long-wave near infrared hyperspectral imaging for determination of moisture content of single maize seed. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* 254.
- Wu, W., Zhang, Z., Zheng, L., Han, C., Wang, X., Xu, J., Wang, X., 2020b. Research Progress on the Early Monitoring of Pine Wilt Disease Using Hyperspectral Techniques. *Sensors (Basel)* 20.
- Wu, B., Liang, A., Zhang, H., Zhu, T., Zou, Z., Yang, D., Tang, W., Li, J., Su, J., 2021. Application of conventional UAV-based high-throughput object detection to the early diagnosis of pine wilt disease by deep learning. *Forest Ecology and Management* 486, 118986.
- Wu, T., Yu, J., Lu, J., Zou, X., Zhang, W., 2020. Research on Inversion Model of Cultivated Soil Moisture Content Based on Hyperspectral Imaging Analysis. *Agriculture* 10, 292.
- Xiaobo, Z., Jiewen, Z., Povey, M.J., Holmes, M., Hanpin, M., 2010. Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta* 667, 14–32.
- Ye, J., 2019. Epidemic status of pine wilt disease in China and its prevention and control techniques and counter measures. *Scientia silvae sinicae* 55, 1–10.
- Yu, R., Luo, Y., Zhou, Q., Zhang, X., Wu, D., Ren, L., 2021a. Early detection of pine wilt disease using deep learning algorithms and UAV-based multispectral imagery. *Forest Ecology and Management* 497, 119493.
- Yu, R., Luo, Y., Zhou, Q., Zhang, X., Wu, D., Ren, L., 2021b. A machine learning algorithm to detect pine wilt disease using UAV-based hyperspectral imagery and LiDAR data at the tree level. *International Journal of Applied Earth Observation and Geoinformation* 101, 102363.
- Yu, R., Ren, L., Luo, Y., 2021c. Early detection of pine wilt disease in *Pinus tabulaeformis* in North China using a field portable spectrometer and UAV-based hyperspectral imagery. *Forest Ecosystems* 8, 44.
- Zarco-Tejada, P.J., Hornero, A., Beck, P., Kattenborn, T., Kempeneers, P., Hernández-Clemente, R., 2019. Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline. *Remote sensing of environment* 223, 320–335.
- Zhang, J.J., Cheng, T., Guo, W., Xu, X., Qiao, H.B., Xie, Y.M., Ma, X.M., 2021b. Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. *Plant Methods* 17.
- Zhang, D., Xu, Y., Huang, W., Tian, X., Xia, Y., Xu, L., Fan, S., 2019. Nondestructive measurement of soluble solids content in apple using near infrared hyperspectral imaging coupled with wavelength selection algorithm. *Infrared Physics & Technology* 98, 297–304.
- Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L., Zhao, C., 2020b. A Review of Advanced Technologies and Development for Hyperspectral-Based Plant Disease Detection in the Past Three Decades. *Remote Sensing* 12, 3188.
- Zhang, B., Ye, H., Lu, W., Huang, W., Wu, B., Hao, Z., Sun, H., 2021a. A Spatiotemporal Change Detection Method for Monitoring Pine Wilt Disease in a Complex Landscape Using High-Resolution Remote Sensing Imagery. *Remote Sensing* 13, 2083.
- Zhang, H., Zhan, B., Pan, F., Luo, W., 2020a. Determination of soluble solids content in oranges using visible and near infrared full transmittance hyperspectral imaging with comparative analysis of models. *Postharvest Biology and Technology* 163, 111148.
- Zhang, N., Zhang, X., Yang, G., Zhu, C., Huo, L., Feng, H., 2018. Assessment of defoliation during the *Dendrolimus tabulaeformis* Tsai et Liu disaster outbreak using UAV-based hyperspectral images. *Remote Sensing of Environment* 217, 323–339.