Computers and Electronics in Agriculture 193 (2022) 106717

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis

Xiaohu Zhao [a], Jingcheng Zhang [a,*], Yanbo Huang [b], Yangyang Tian [a], Lin Yuan [c]

[a] College of Artificial Intelligence, Hangzhou Dianzi University, Hangzhou 310018, China
[b] Crop Production Systems Research Unit, United States Department of Agriculture, Agricultural Research Service, PO Box 350, Stoneville, MS 38776, USA
[c] School of Information Engineering and Art and Design, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China

ARTICLE INFO

Keywords:
Tea green leafhopper
Anthracnose
Sunburn
Continuous wavelet analysis
Discrimination

ABSTRACT

Compared with the traditional visual detection method, hyperspectral imaging enables efficient and non-destructive plant monitoring. Besides, it has great potential in plant phenotyping in response to disease and insect infections. However, most previous studies on hyperspectral imaging have focused on detecting a single disease, which can rarely discriminate between multiple co-occurring diseases and insects. In this study, three tea plant stresses with similar symptoms, including the tea green leafhopper (*Empoasca (Matsumurasca) onukii* Matsuda), anthracnose (*Gloeosporium theae-sinesis* Miyake), and sunburn (disease-like stress), were evaluated. A multi-step approach was proposed based on hyperspectral imaging and continuous wavelet analysis (CWA) to discriminate the plant stresses. The process entailed: (1) Feature extraction for detection and discrimination of tea plant stresses based on CWA; (2) Detecting abnormal areas on tea leaves via the k-means clustering and support vector machine algorithms; (3) Construction of a model for identification and discrimination of the three tea plant stresses via the random forest algorithm. The results showed that CWA could effectively identify spectral features for distinguishing the three stresses. The overall accuracy (OA) of the proposed approach reached 90.26%-90.69%, with anthracnose having the highest OA (94.12%-94.28%), followed by tea green leafhopper (93.99%-94.20%), while sunburn damage was the least (82.50%-83.91%). Therefore, hyperspectral imaging is effective for plant phenotyping after diseases and insect infections.

1. Introduction

Tea is an important economic crop with an increasing global demand (Xia et al., 2017; Fang et al., 2019; Bora et al., 2019). However, diseases and insect infestations pose severe threats to the yield and quality of tea leaves, thus limiting crop productivity. At present, there are nearly 1,000 types of tea plant pests and more than 380 tea plant diseases globally, which cause yield losses as high as 43% (Roy et al., 2019; Gnanamangai and Ponmurugan, 2012; Sun et al., 2019). Effective monitoring and identification of major tea plant diseases and insects could guide precise disease prevention and insect control. This could help reduce environmental pollution caused by the excessive application of insecticides and pesticides (Sankaran et al., 2010; Hu et al., 2019). A high-throughput phenotypic analysis is also applied during tea plant breeding to detect diseases and insect infestations. Presently, manual inspection and sampling methods, including field scouting of symptoms, disease incidence rate, and severity, or taking samples of diseased plants, and detecting the pathogens through immune or biochemical procedures are commonly used for monitoring of diseases and insects in tea gardens (Zhang et al., 2019a). Although these conventional methods can achieve relatively accurate and reliable results, they can be subjective and inefficient, limiting rapid investigation over large areas. Therefore, non-destructive pest detection methods are needed (Mahlein et al., 2019).

Hyperspectral imaging (HSI) has recently emerged as a non-destructive sensing technology with great potential to detect crop diseases and insect infestations. HSI can simultaneously obtain both spatial and spectral information. Therefore, it captures the spectral leaf changes caused by the symptoms of diseases and insect infestations. This makes it valuable for the diagnosis and determination of the degree of disease and insect infestations. Various studies have combined spectral analysis with pattern recognition algorithms to effectively identify plant diseases. For instance, Xie et al. (2017) used 655, 746, 759, 760, and 761 nm spectral bands as the feature set. These were combined with the KNN model to build a classification model of tomato gray mold, and obtained

* Corresponding author.
E-mail address: zhangjcrs@hdu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.compag.2022.106717
Received 22 October 2020; Received in revised form 11 August 2021; Accepted 13 January 2022
Available online 21 January 2022

Nomenclature table

AH	Anthracnose
CWA	Continuous wavelet analysis
FLDA	Fisher linear discriminant analysis
GL	Tea green leafhopper
HSI	Hyperspectral imaging
NM	Normal areas
OA	Overall accuracy
OFs-AD	Optimal features for abnormality detection
OFs-SD	Optimal features for stresses discrimination
RF	Random forest
RMV	Maximum differential value in red edge
ROI	Regions of interest
RPMV	Position of the maximum differential value in red edge
RSV	Sum of differential values in red edge
SB	Leaf sunburn
SFs-AD	Spectral features for abnormality detection
SFs-SD	Spectral features for stresses discrimination
SVM	Support vector machine
WFs-AD	Wavelet features for abnormality detection
WFs-SD	Wavelet features for stress discrimination
YMV	Maximum differential value in yellow edge
YPMV	Position of the maximum differential value in yellow edge
YSV	Sum of differential values in yellow edge

an OA of 97.22%. Sinha et al. (2019) also selected 16 Visible and Near-infrared (VIS-NIR) bands to detect grapevine leafroll disease. Similarly, quadratic discriminant analysis (QDA) has been used to develop a disease recognition model with an OA greater than 75%. The vegetation indices can detect some important physical and chemical plant traits by applying a mathematical combination of some original wavebands. Abdulridha et al. (2019) tested the ability of 31 vegetation indices to detect citrus canker. The results showed that the improved chlorophyll absorption reflectance index (ARI and TCARI1) combined with the radial basis function could effectively detect the disease, with OA of 94%. Yuan et al. (2019) also constructed two disease indices based on HSI data (tea anthracnose ratio index (TARI) and tea anthracnose normalization index (TANI)) which were then combined with the ISO-DATA image clustering algorithm for the adaptive detection of tea anthracnose.

However, recent studies that have used spectral analysis to detect plant diseases and insect infestations mainly focus on the leaf scale. Besides, corresponding techniques have been primarily designed to identify single stress. However, various disease and insect stresses concurrently occur in the actual farmland environment. For instance, tea green leafhopper (GL) and anthracnose (AH) co-exist in the green tea planting area. In addition, tea plantations severely damaged by diseases and insect infestations, such as the tea green leafhopper, are also more vulnerable to drought and heat damage, resulting in leaf sunburn (SB) (Han et al., 2013). GL, AH, and SB have different methods for their prevention and control. For instance, GL and AH should be picked multiple times in batches, pruned reasonably, and fertilized properly. However, SB usually needs timely irrigation without picking, pruning, and fertilizing. Therefore, effective technology for diagnosing and distinguishing between the various tea plant stresses is needed since inappropriate procedures can severely affect many cases. However, spectral similarity may exist between different diseases and insect infestations or other stresses. Therefore, spectral analysis methods should be precise, especially to identify subtle spectral differences (Mahlein et al., 2013; Yuan et al., 2014; Baranowski et al., 2015). Among the available spectral analysis methods, continuous wavelet analysis (CWA) is superior to traditional approaches in spectral feature extraction of biochemical parameters, such as vegetation moisture (Cheng et al., 2011; Cheng et al., 2014). In recent years, CWA has shown encouraging results in identifying and differentiating plant diseases and insects, such as yellow rust, powdery mildew, and aphids in winter wheat (Zhang et al., 2014; Zhang et al., 2017). Therefore, the ability of CWA in distinguishing tea plant diseases and insect infestations should be examined.

Tea plants in Southern China are prone to GL, AH, and SB attacks. Although the above stresses exhibit similar symptoms, different procedures are required to prevent or alleviate their occurrence and damages. Herein, hyperspectral imaging technology and CWA were used to: (1) assess the spectral characteristics of the three similar tea plant stresses and identify optimized feature sets for detection and

Fig. 1. Symptoms of GL, AH, and SB at canopy and leaf level.

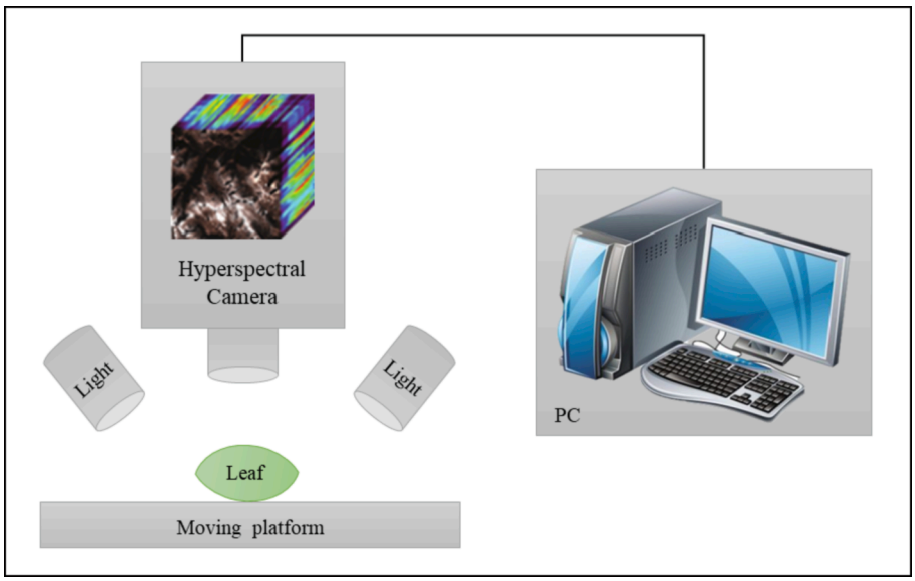

Fig. 2. Schematic diagram of a hyperspectral imaging system.

discrimination of tea plant stresses; (2) integrate image clustering analysis and multiple classification methods (support vector machine (SVM), Fisher linear discriminant analysis (FLDA), and random forest (RF)) to develop a protocol for identifying and discriminating the stresses.

2. Materials and methods

2.1. Experimental materials

A plant pathologist identified and collected the tea leaf samples of GL, AH, and SB from the experimental tea gardens of the Chinese Academy of Agricultural Sciences, Hangzhou, Zhejiang, China. AH (cultivar: Longjing 43) samples were collected in September 2017, GL (cultivar: Longjing 43) and SB (cultivar: Baiye No. 1) samples were collected in June 2019 at the growing stage of the summer shoot of tea plants. The Longjing 43 and Baiye No. 1 are widely used in Zhejiang Province to produce green tea. Both cultivars are susceptible to AH and GL and are vulnerable to sunburn. GL is characterized by leaf yellowing and edge curling, and premature leaf abscission (Jin et al., 2012; Mei et al., 2017). AH infection is characterized by gray-brown spots on tea leaves (Chen et al., 2012). SB, a disease-like stress condition, is characterized by dehydration, tarnishing, reddening, scorching, and abscission (Wahid et al., 2007; Han and Xiao, 2013) on the crown surface of tea plants (Fig. 1). The leaf samples (118 GL, 72 AH, and 119 SB) were transported to the laboratory in a small incubator. An indoor imaging hyperspectral testing system was used to obtain the hyperspectral imaging data.

2.2. Measuring hyperspectral imaging

A hyperspectral imager (UHD185, Cubert GmbH, Ulm, Germany) in a dark box under two 50 W halogen lamps (ASD Pro Lamp, ASD Inc., Boulder, Colorado, USA) was used for hyperspectral imaging. A schematic diagram of the HSI system is shown in Fig. 2. The image had a spatial resolution of 1000 * 1000 pixels and a spectral resolution of 4 nm with 125 bands at 450–950 nm. The hyperspectral image had a spatial resolution of 0.1 mm/pixel, exposure time of 10 ms. The distance between leaf and camera lens is 500 mm, and the lens f = 25 mm. Image radiometric correction was conducted using the reference PTFE whiteboard and blackboard (https://sphereoptics.de/) to derive the reflectance. The spectral signals could saturate because the wax layer of the tea tree was relatively thick. Therefore, such samples were removed to ensure the quality of the data for subsequent analysis. Finally, 100 GL, 60 AH, and 100 SB leaves were retained. Half of the samples (130 images) were randomly selected from each stress category to form the data set for model training, and the remaining half was used for model validation.

2.3. ROI selection of hyperspectral images

It is necessary to select regions of interest (ROI) on the leaves for spectral information extraction when studying the spectral characteristics of the three foliar damages. Herein, two ROIs were selected for each leaf image in the training data set, located in the center of normal (NM) and abnormal areas. The selected ROIs were kept symmetrical with the leaf vein, and the areas with large spectral changes, such as leaf veins, were avoided. In each leaf sample, all the pixels of two ROIs for normal and lesion areas were averaged to obtain the spectra representing the ROI. A total of 130 lesion spectra and 130 normal spectra were obtained.

2.4. Feature extraction for abnormality detection and stress discrimination

The conventional spectral differential analysis, continuum removal analysis, and continuous wavelet analysis (Demetriades-Shah et al., 1990; Pu and Gong, 2004; Cheng et al., 2010) were used for abnormality detection and stress discrimination. They are all classic methods used to analyze the spectral shape and capture the spectral reflecting and absorbing features. Further, they mitigate the background influence of radiation variations. Therefore, the spectral shape identified based on the above methods has achieved excellent performance in several studies (Pu, 2017). Herein, “detection” refers to identifying abnormal areas (all the three stresses were defined as abnormal) in tea leaves. Each pixel is judged as healthy or stressed. “Discrimination” aims at determining the stress type (GL, AH, or SB) of the identified abnormal pixels.

2.4.1. Extraction of continuous wavelet features

The CWA technique decomposes the reflectance spectra with continuous scale and wavelength based on the wavelet transformation, then extracts the wavelet features from the obtained wavelet coefficient spectrum. Compared with the commonly used spectral indices, wavelet features can effectively capture the spectral shape information, which can detect both the microscopic and macroscopic characteristics in the

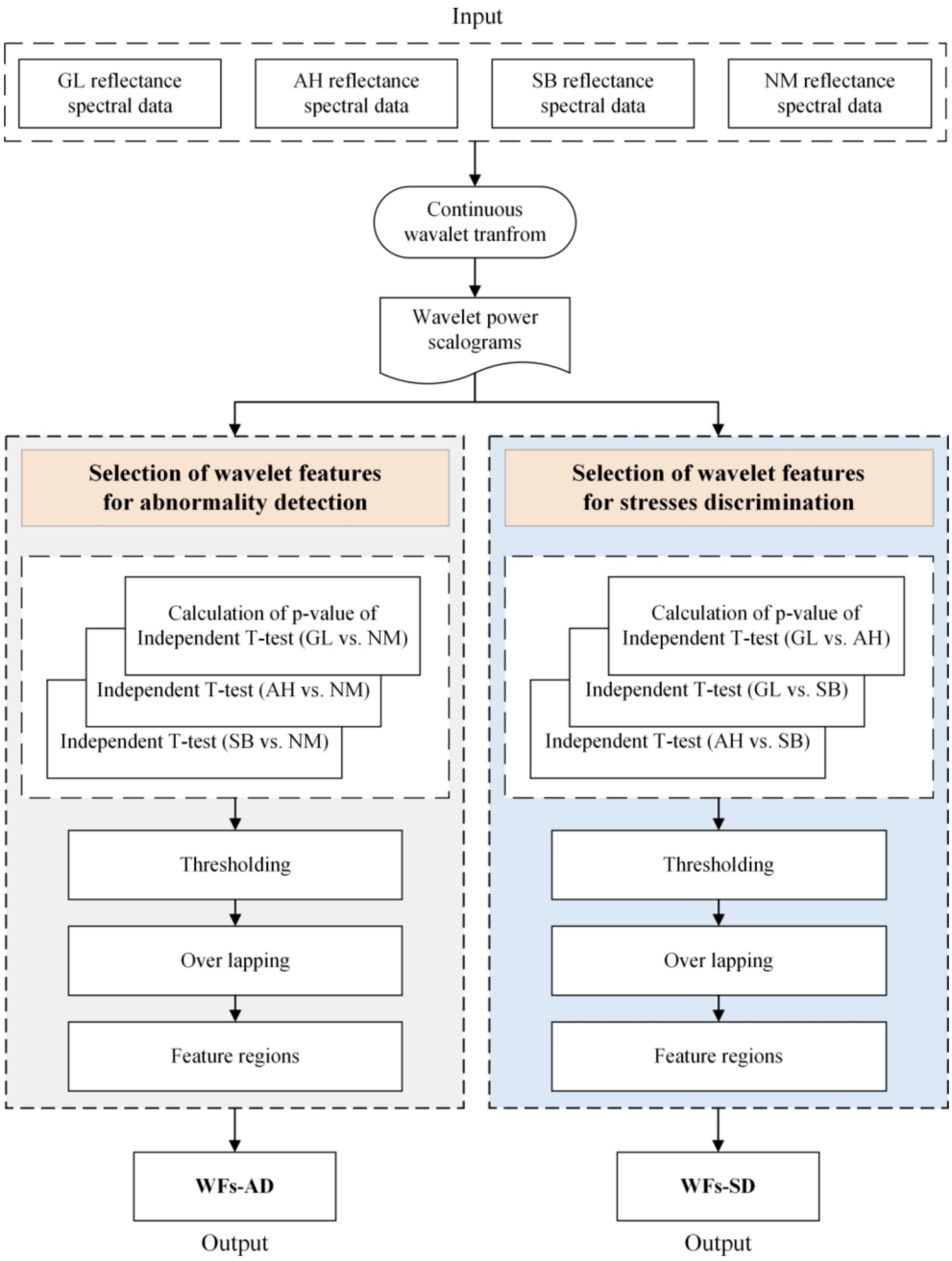

Fig. 3. Flow chart of feature selection for abnormality detection and stresses discrimination based on continuous wavelet analysis. (GL, AH, SB, NM indicate tea green leafhopper, anthracnose, sunburn, and normal leaves, respectively).

spectra (Cheng et al., 2010). Herein, Mexican Hat (Quasi-Gaussian function) was used as the wavelet basis (Torrence and Compo, 1998). The decomposition scale was restricted to 2^i (i = 0, 1, 2, to 10) to reduce the complexity of the analysis (Cheng et al., 2010; Cheng et al., 2011).

After completing CWA, each spectral curve was used to generate corresponding wavelet coefficients at various decomposition scales. A sensitivity analysis should be conducted to identify the most appropriate features for detecting plant diseases and pests (Zhang et al., 2019b). Herein, an independent *t*-test was conducted among each pair of stresses (GL vs. AH, GL vs. SB, and AH vs. SB) and normal against each stress type (GL vs. NM, AH vs. NM, and SB vs. NM). Six *p*-value scalograms were obtained, indicating significant differences across all wavelengths and scales. The retained wavelet elements from the GL vs. NM, AH vs. NM, and SB vs. NM *p*-value scalograms were overlapped to obtain the wavelet feature regions for abnormality detection ($p < 0.001$). Meanwhile, the overlapped wavelet feature regions from the GL vs. AH, GL vs. SB, and AH vs. SB *p*-value scalograms were identified for stress discrimination (Fig. 3). For the extraction and calculation of the wavelet features within each identified wavelet feature region, the wavelength and scale corresponding to the minimum value in the region was determined as the optimal wavelet feature. The wavelet features corresponding to the t-value peaks in local sub-regions within the sensitive region were selected as the optimal wavelet features where the feature area was a continuous large region (L > 100 nm) to avoid missing crucial information. The optimal wavelet features formed the wavelet features for abnormality detection (WFs-AD) and stress discrimination (WFs-SD).

2.4.2. *Extraction of differential and continuum features*

Besides CWA, the conventional spectral differential analysis and continuum removal analysis were used to capture significant variations

Table 1
Conventional shape based on spectral features.

Feature Type	Position	Band Range (nm)	Feature	Reference
First-order differential features	Yellow edge	540–620	Maximum differential value (YMV)	Gong et al. (2002)
			Position of the maximum differential value (YPMV)	
			Sum of differential values (YSV)	
	Red edge	660–780	Maximum differential value (RMV)	
			Position of the maximum differential value (RPMV)	
			Sum of differential values (RSV)	
Continuum removal features	VIS-NIR	530–770	Depth	Pu et al. (2003)
			Width	
			Area	

in spectral shape with certain spectral transformations (Gong et al., 1997; Pu and Gong, 2004). Nine spectral characteristics from the first-order differential and continuum removal features were then selected to form a candidate spectral feature set (Table 1).

A similar spectral sensitivity analysis (Section 2.4.1) was performed on the above spectral differential and continuum removal features via the independent *t*-test to obtain the spectral features for abnormality detection (SFs-AD) and stress discrimination (SFs-SD). To avoid information redundancy among the identified WFs and SFs, a cross-correlation analysis was performed on each pair of features from the WFs and SFs for abnormality detection and stress discrimination, respectively. A criterion of $R^2 > 0.8$ was used to remove features with a high correlation, after which attributes with stronger sensitivity (marked as the smaller *p*-value) were retained. All feature pairs within the spectral features were traversed to form the optimal features for abnormality detection (OFs-AD) and stress discrimination (OFs-SD).

2.5. Construction of a step-by-step abnormality detection and stress discrimination model

A step-by-step abnormality detection and stress discrimination model was established based on images of optimal features (OFs) extracted from the hyperspectral imaging data, combined with various machine learning approaches (Fig. 4). The K-means algorithm was used for image clustering. The K-means algorithm can divide the whole data set into k disjoint clusters, process large data sets, and quickly converge them to local optimum (Zahra et al., 2015). Meanwhile, three machine learning approaches, including SVM, FLDA, and RF, were used, and their stress detection and discrimination abilities were compared. FLDA is a standard technology used for dimension reduction. It can find the projection direction between very distant classes and very close classes to achieve a good classification effect (Duda and Hart, 1973). SVM can effectively analyze high-dimensional data when the size of training samples is small, and the difference between classification samples is not apparent (Tarabalka et al., 2010; Jiménez-Carvelo et al., 2019). RF is a combination of decision trees based on the idea of bagging. The RF classifier averages many noisy but approximately unbiased models, thus reducing variance (Pal, 2005). The FLDA does not include parameters for tuning. For SVM, through preliminary analysis that based on the training data and the criteria of the highest overall accuracy, the C-SVC

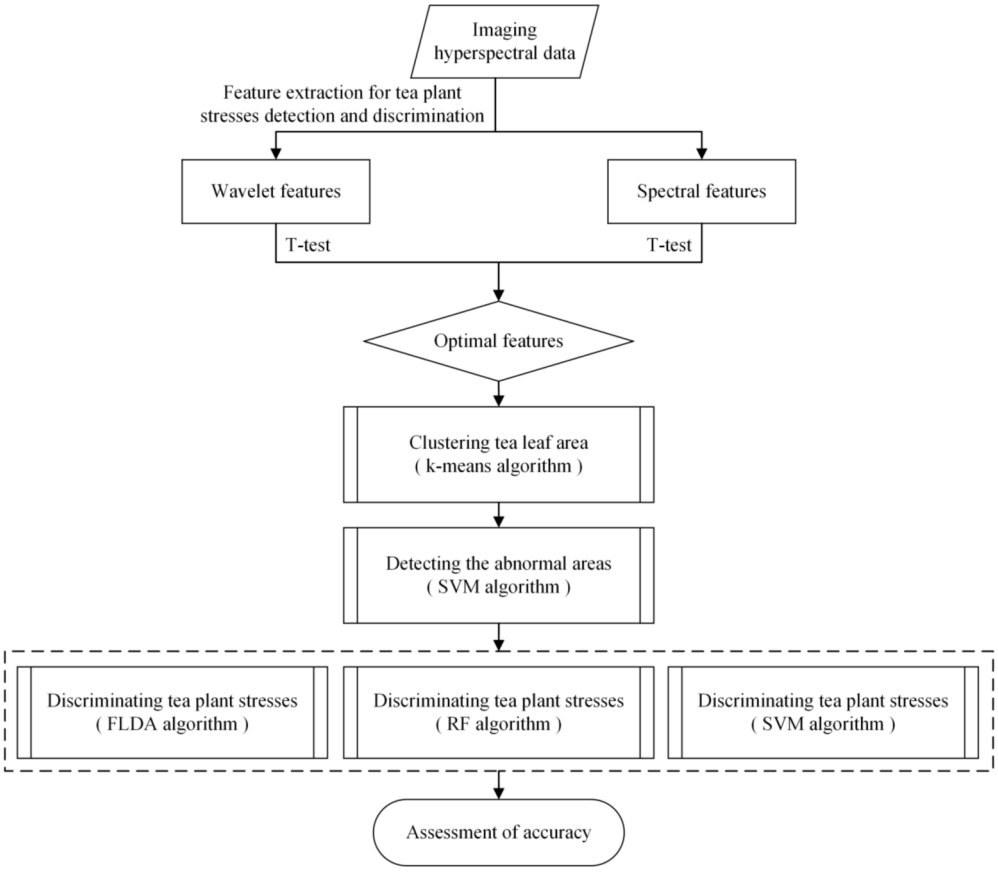

Fig. 4. Technical workflow of tea plant stresses detection and discrimination based on hyperspectral imaging data.

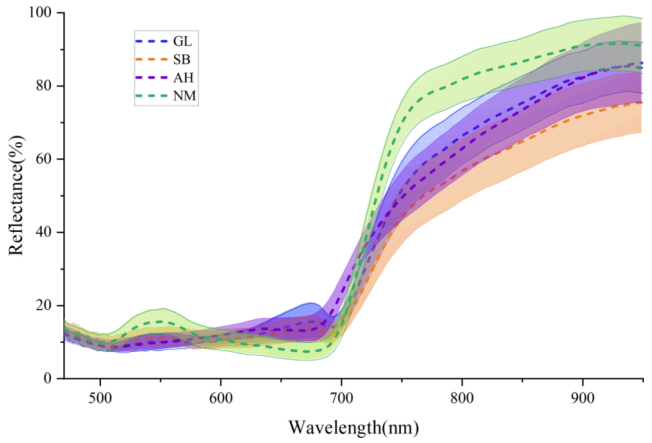

Fig. 5. Average reflectance and standard deviations for the three tea plant stresses. (GL, AH, SB, NM indicate tea green leafhopper, anthracnose, sunburn, and normal leaves, respectively).

algorithm was determined as the core algorithm. The RBF kernel function was selected from Linear, Polynomial, RBF, and Sigmoid, and the parameter C was set to 1 from a range of [1–10] with a step of 1. The same process was used to determine the number of the tree for RF. The number was determined as 1000 from a range of [100–2000] with a step of 100. Notably, both SVM and RF can self-adapt under given parameters and data.

The step-by-step strategy for detection and discrimination of plant stresses was constructed as follows:

(1) The k-means algorithm was used to cluster the tea leaf area. To avoid irregular leaf lesion edges and "salt pepper" fragments in classification, the k-means algorithm was used on OFs-AD images for image clustering before subsequent analysis. The obtained clusters were used as analysis units during stress detection and discrimination (Yuan et al., 2019). The clustering number n was set to 20 through trial and error.
(2) The SVM algorithm was used to detect the abnormal areas on tea leaves based on the leaf image clusters. Each cluster was classified into the normal or abnormal area using the training data and SVM algorithm. The abnormal regions were further processed using the subsequent stress discrimination to simplify the problem and improve the overall discrimination accuracy.
(3) Based on the identified abnormal areas, FLDA, RF, and SVM algorithms were used on OFs-SD feature images to compare their performance on discrimination of tea stresses.

2.6. Assessment of accuracy

For sample validation, the hyperspectral leaf images were visually interpreted, and the ROI of abnormal areas was extracted manually to act as a reference. The proposed methods were then compared with the reference ROIs to determine the numbers of correctly classified and misclassified pixels, thus generating a confusion matrix for accuracy evaluation. The accuracy evaluation was done based on overall accuracy (OA) = (TN + TP)/(TN + TP + FN + FN), precision = TP/(TP + FP) and recall = TP/(TP + FN). MATLAB software (MathWorks Inc., Natick, Massachusetts, USA) was used for all statistical analyses and modeling.

3. Results and discussion

3.1. Detection and discrimination features of tea plant stresses

The spectral curves of the three tea plant stresses (GL, AH, and SB) and corresponding normal leaves are shown in Fig. 5. Compared with the normal samples, the three stresses induced similar spectral changes. The main changes included a decreased reflectance around the green peak (near 550 nm) and near-infrared platform (near 770 nm) and increased reflectance around the red valley (near 680 nm). The spectrum curves of the three stresses revealed a convergence at the visible region. However, the spectral curve in the near-infrared region was significantly lower in SB than in GL and AH. By viewing the spectra of samples from different stresses (Fig. 5), it can be observed that the average spectra of different classes approach to each other. Therefore, some in-depth spectral analysis is necessary for effectively classifying these stresses.

The sensitivity analysis of spectral features used for detecting abnormal areas was performed on wavelet and conventional features to screen the optimal feature sets. The *t*-test results of wavelet scalograms of the three tea plant stresses versus their normal samples are shown in Fig. 6. Overlapping of the sensitive regions on each scalogram identified five WFs-AD (Table 2) located at the green (543 nm), red (623 nm), red-edge (700 nm), and near red (721–745 nm) spectral regions. For the large-area contiguous sensitive region, four wavelet features were

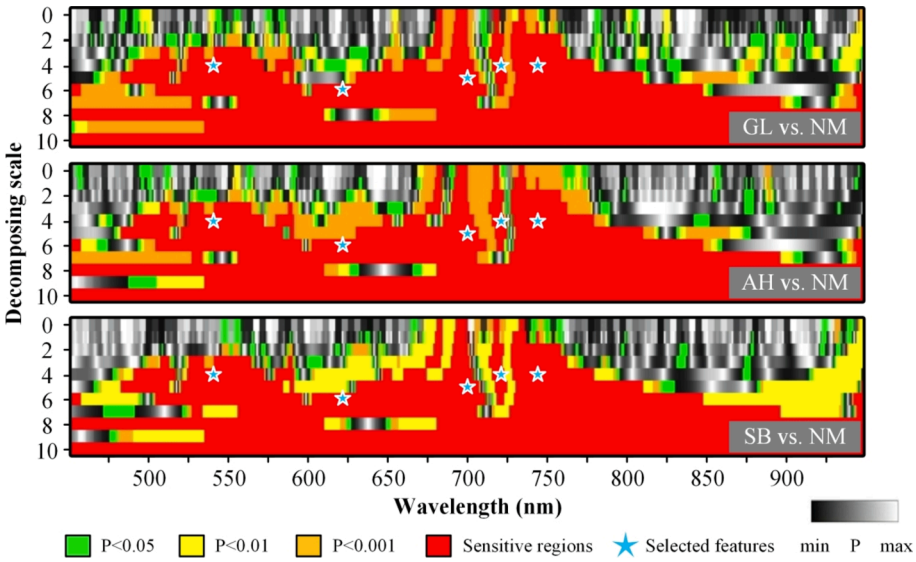

Fig. 6. Wavelet feature selection based on *t*-test scalograms for abnormality detection. (GL, AH, SB, NM indicate tea green leafhopper, anthracnose, sunburn, and normal leaves, respectively).

Table 2
Parameters of selected wavelet features for abnormality detection.

	Wavelet Features	Scale	Wavelength Region (nm)	Wavelength (nm) at minimal *p*-value
WFs-AD	WF_{A01}	4	538–564	543
	WF_{A02}	6	618–633	623
	WF_{A03}	5	695–701	700
	WF_{A04}	4	717–724	721
	WF_{A05}	4	738–760	745

Table 3
Sensitivity of conventional spectral features on abnormality detection.

Spectral Features	Significance of independent T-test		
	GL vs. NM	AH vs. NM	SB vs. NM
YMV			
YPMV		+	+
YSV*	+	+	+
RMV*	+	+	+
RPMV		+	+
RSV*	+	+	+
Depth	+	+	
Width*	+	+	+
Area	+		+

Note: "+" indicates that the significance of the *t*-test satisfies $p < 0.001$, and * is the selected spectral features for abnormality detection. GL, AH, SB, and NM represent tea green leafhopper, anthracnose, sunburn, and normal leaves, respectively.

selected at four local extreme value regions to ensure the diversity of the wavelength and scale distributions of wavelet features. The sensitivity analysis results of the conventional SFs obtained four SFs-AD (YSV, RMV, RSV, and Width) (Table 3). These features were mainly located around the chlorophyll absorption regions (450 nm and 650 nm), red-edge (680–750 nm), and near-infrared areas (700–750 nm).

Cross-correlation analysis of the identified WFs-AD and SFs-AD detected and removed considerable redundancy (using a criterion of $R^2 < 0.8$). Subsequently, OFs-AD constructed using fewer features (WF_{A04}, WF_{A05}, YSV, RMV, and Width) were obtained as the final inputs for detecting the abnormal leaf areas.

A similar procedure was used on the *t*-test scalograms (Fig. 7) to screen the features of discriminating tea plants stresses. Comparing the sensitivity patterns shown in the scalograms for stress detection (Fig. 6), the sensitive regions significantly shrunk, possibly due to the relatively similar spectral variations between GL and AH. However, the superiority of continuous wavelet analysis was used to assess all the wavelength and scales, identifying seven sensitive areas that could capture spectral differences (Table 4). The WFs-SD were mainly located around the blue, green, and red-edge spectral regions. Besides, none of the nine conventional features passed the sensitivity analysis for all the three stress pairs (Table 5). Conventional SFs can capture the spectral shape variations around specific spectral regions. In this study, the SFs did not capture some subtle spectral changes between stresses with similar symptoms. For instance, most SFs except for YSV could not differentiate between GL and AH. Meanwhile, wavelet features had superior adaptive

Table 4
Parameters of selected wavelet features for stress discrimination.

	Wavelet Features	Scale	Wavelength Region(nm)	Wavelength(nm) at Minimal *p*-value
WFs-SD	WF_{S01}	3	451–459	451
	WF_{S02}	4	475–486	484
	WF_{S03}	4	499–506	503
	WF_{S04}	5	513–524	521
	WF_{S05}	5	622–639	633
	WF_{S06}	4	647–652	650
	WF_{S07}	6	711–723	719

Table 5
Sensitivity of conventional spectral features to stress discrimination.

Spectral Features	Significance of independent T-test		
	GL vs. AH	GL vs. SB	SB vs. AH
YMV			
YPMV			
YSV	+		+
RMV		+	+
RPMV		+	
RSV		+	+
Depth		+	+
Width			+
Area		+	+

Note: "+" indicates that the significance of the *t*-test satisfies $p < 0.001$. GL, AH, and SB represent tea green leafhopper, anthracnose, and sunburn, respectively.

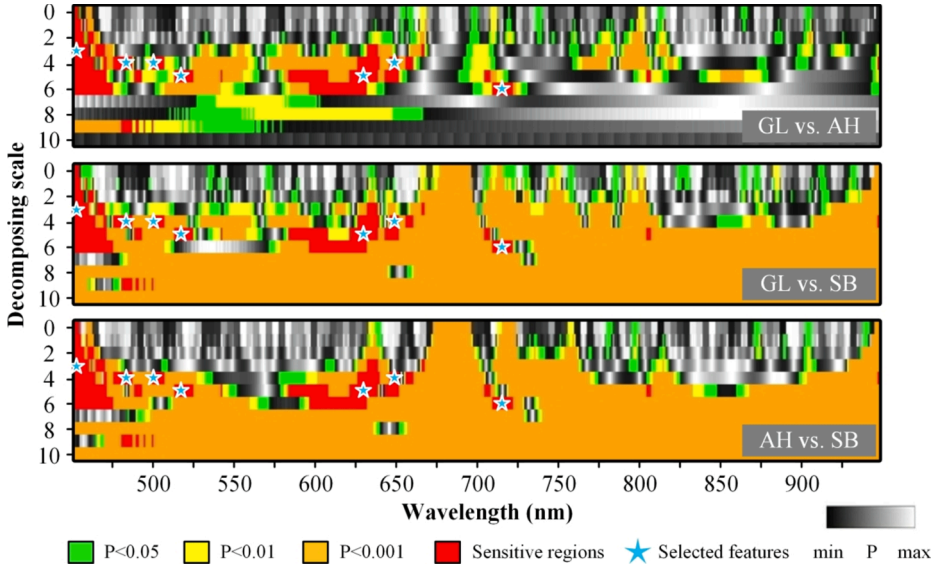

Fig. 7. Wavelet feature selection based on *t*-test scalograms for stress discrimination. (GL, AH, SB indicate tea green leafhopper, anthracnose, and sunburn, respectively).

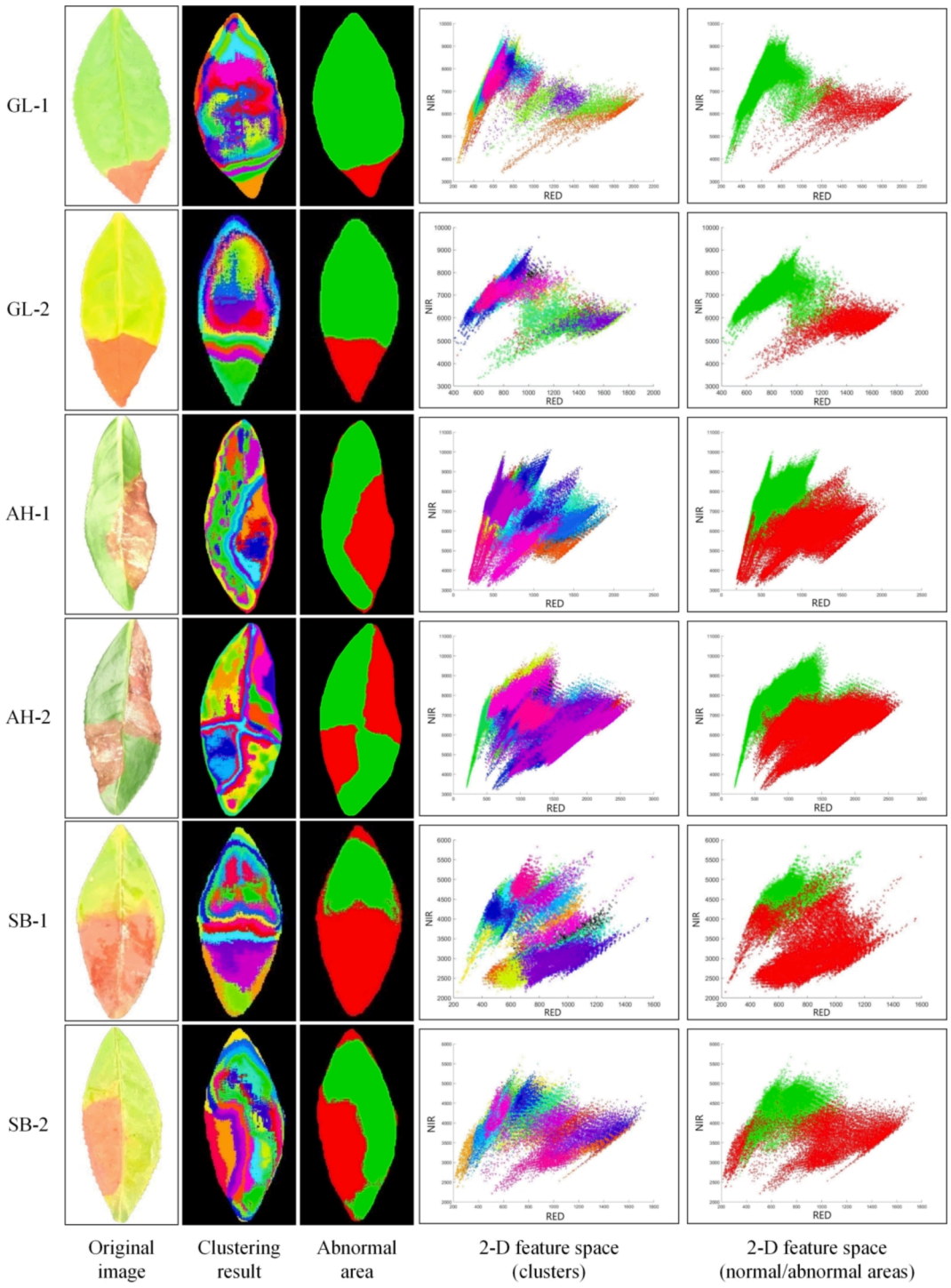

Fig. 8. Demonstration of the leaf abnormal area detecting process. Note: GL, AH, SB indicate tea green leafhopper, anthracnose, and sunburn, respectively. The clustering results were generated based on the optimal spectral features identified for leaf abnormal area. The clusters are shown in different colours in both the leaf image and their distributions in the RED-NIR 2-D feature space.

spectral feature extraction capability that could capture the subtle variations in the spectral shapes corresponding to different stresses. Also, all the seven features were included in the OFs-AD because all WFs-AD passed the cross-correlation check.

Tea plant stresses, such as diseases, insects, and sunburns, destroy the leaf chloroplast structure, weakening the absorption of corresponding bands and increasing the reflectance (Belward, 1991; Wang et al., 2014). Meanwhile, decreased chlorophyll concentration also causes a shift of the red edge towards the shortwave direction (Jago et al., 1999). Besides, the leaf damages cause substantial cell destruction in the leaf sponge tissue and decrease water content in some severe cases. These damages can reduce the spectral reflectance in the near-infrared region (Wang et al., 2014; Abdulridha et al., 2019). SB usually occurs when tea plants exposed to intense sunlight and severe dehydration develop a lower spectral reflectance in the near-infrared band than GL and AH (Han and Xiao, 2013). Notably, in the Wavelet features, $WF_{A02\text{-}03}$, $WF_{S01\text{-}02}$, and $WF_{S05\text{-}06}$ located around the chlorophyll absorption bands indicate the damage of the chloroplast. $WF_{A04\text{-}05}$

Table 6
Accuracy of tea plant stress detection and discrimination.

Tea plant stresses	Accuracy of abnormality detection	Accuracy of stress discrimination		
	SVM	FLDA	RF	SVM
GL	94.29%	93.99%	94.03%	94.20%
AH	94.32%	94.28%	94.12%	94.17%
SB	83.97%	82.50%	83.91%	83.63%
Average	90.86%	90.26%	90.69%	90.67%

Note: GL, AH, and SB indicate tea green leafhopper, anthracnose and sunburn, respectively.

and WF_{S07} located around the Near-infrared platform indicate the spectral change induced by cellular structure changes.

3.2. Step-by-step detection and discrimination model of tea plant stress

3.2.1. Abnormality detection model of tea plants

A tea plant model for detecting abnormal areas was constructed based on OFs-AD, k-means clustering, and SVM algorithms. "Red-NIR" (Fig. 8) shows the distribution of all leaf pixels from different clusters that were obtained via k-means (shown in different colours) in the Red-NIR two-dimensional feature space. The results showed that many pixels diffused outward from the centre and were distributed in a fan-shaped two-dimensional scattering pattern. The abnormal area pixels are often located at the lower right corner of the fan-shaped area near the x-axis, indicating that the red band reflectance of the pixels in this region is higher. In contrast, the near-infrared reflectance is lower, consistent with the spectral characteristics shown in Fig. 5. The OA of the model was 90.86% (Table 6), indicating that the model can effectively distinguish normal and stressed leaves. However, the detection accuracy of SB (83.97%) was significantly lower than that of GL (94.29%) and AH (94.32%). This finding could be because SB symptoms are not similar to GL and AH symptoms. The edge of the SB spot is relatively blurred with a gradual change in pattern (Han and Xiao, 2013). Besides, a certain degree of deviation may occur through methods, such as clustering, challenging its identification. Notably, when performing abnormality detection on the test set, errors always occur at the petiole and leaf edge (misclassification of the normal area as abnormal), as shown in Fig. 8 with SB-1, SB-2, and AH-1. This finding may be attributed to the fact that the petiole contains low pigment and high lignin and cellulose concentrations, resulting in different spectral responses (Taiz et al., 2018).

3.2.2. Stress discrimination model of tea plants

The tea plant stress discrimination models were constructed using three classification algorithms (FLDA, RF, and SVM) based on the OFs-SD images of the detected abnormal leaf areas. The results suggested that the three classifiers can effectively distinguish different tea plant stresses (Table 6). The overall classification accuracies were 90.26%, 90.69%, and 90.67% for FLDA, RF, and SVM, respectively. The relatively high classification accuracy could be related to the preliminary procedure for detecting abnormal leaf areas, which reduces the complexity of the classification. A confusion matrix was also generated, which only evaluated the stress discrimination (GL, SB, and AH. The normal pixels determined in the previous step were not included). The stress classification results of the three algorithms in the abnormal area are shown in Fig. 9. OA of SVM (98.07%), RF (96.93%), and FLDA (94.93%) all reached a relatively high level with slight difference. Besides, the precision and recall of the three classifiers for all the stresses were more than 91%. Although the OA of RF and SVM was slightly higher than that of FLDA, the simple and explainable structure of FLDA makes it superior in analyzing efficiency. Therefore, FLDA is recommended for distinguishing tea plant stresses.

The step-by-step leaf abnormality detection and stresses discrimination procedure was established based on the advanced spectral features screening technique and machine learning methods. The performance of the stress discrimination model is demonstrated in Fig. 10. The abnormality detection conducted before stress discrimination simplified the spectral classification of stresses. Although most damaged areas were identified and classified, some imperfect recognized regions occurred around the petiole and leaf edge of the leaves. The misclassified regions could be due to the less explicit spectral characteristics at these positions. The success of the model could be due to: (1) The wavelet features generated from the CWA can provide adaptive capturing of subtle changes of different stresses; (2) The clustering generated by the k-means algorithm provides an excellent segmentation, producing naturally-looking edges between the normal and abnormal areas on the leaf; (3) The stepwise procedure can effectively reduce the complexity of stress discrimination, which may improve performance. Hyperspectral imaging can monitor plant diseases and pests (Yuan et al., 2019; Xie et al., 2017; Sinha et al., 2019). The differentiation of leaf diseases and pests is based on the spectral response caused by leaf injury symptoms. Therefore, it is necessary to analyze the spectral response and establish the detecting models at the leaf level. The leaf level findings can then be used to upscale the detecting features and models to the canopy level, which could be more complicated due to the influence from the canopy structural factors (plants morphology and leaf angle distributions). Therefore, the applicability of the spectral features and models at the canopy level should be assessed in the future for effective monitoring of tea garden stresses using real airborne and satellite images. Notably, the development of hyperspectral imaging technology will reduce the cost of the sensors. Besides, the cost of the sensors in some well-established application scenarios may further decrease by customizing some parameters, such as the spectral range,

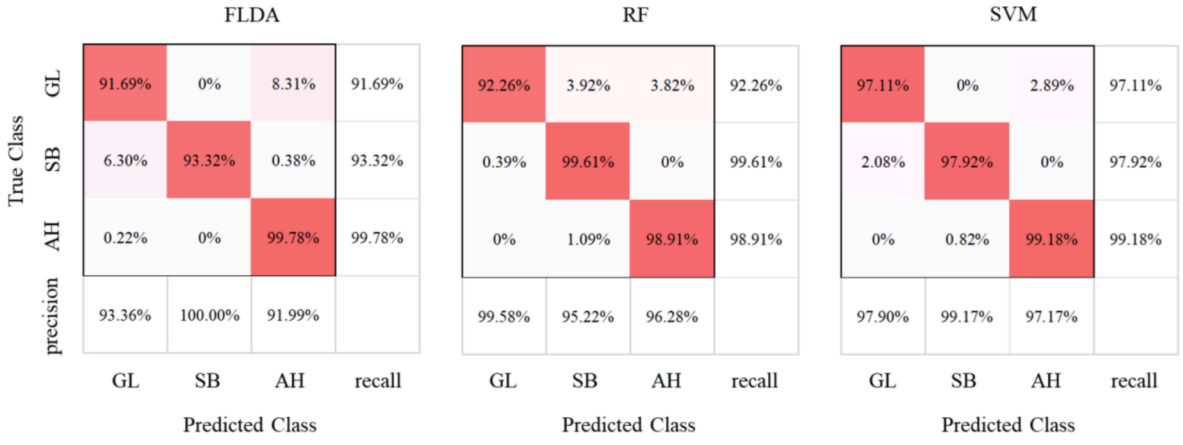

Fig. 9. Confusion matrix of tea plant discrimination.

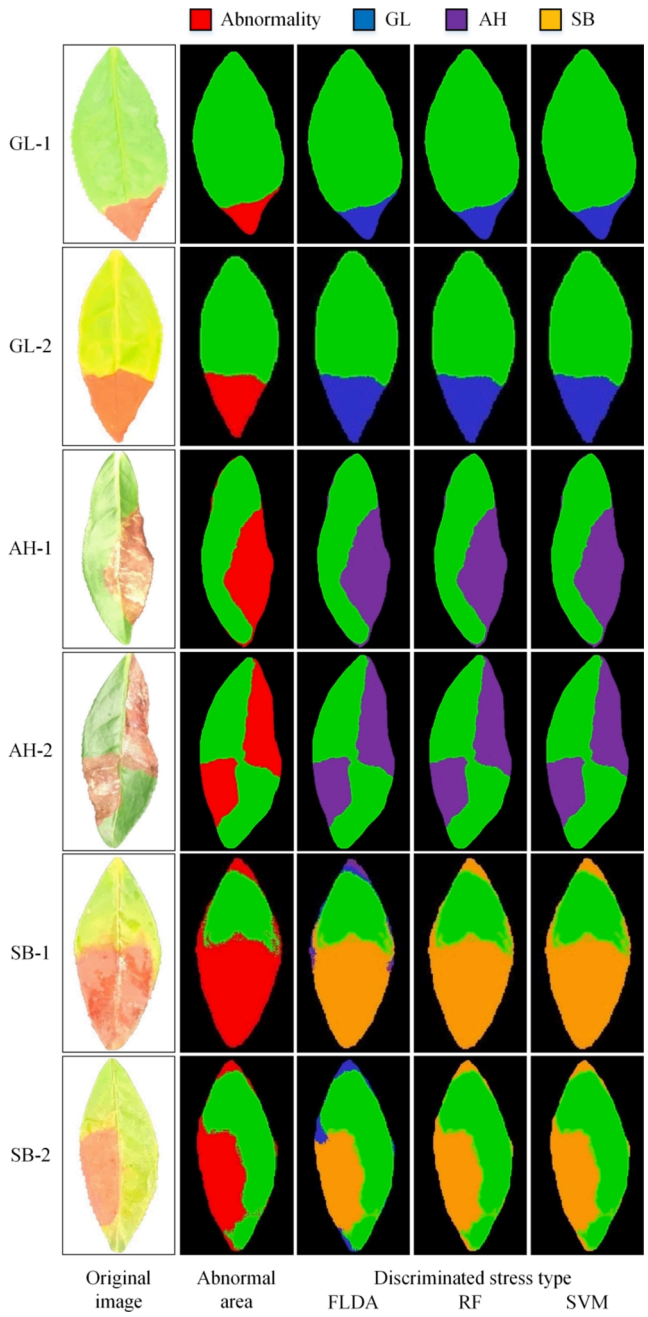

Fig. 10. Demonstration of stress discrimination on tea leaves. Note: GL, AH, and SB indicate tea green leafhopper, anthracnose, and sunburn, respectively. The right 1–3 columns show the discrimination results of the stress type (indicated by different colours) with different algorithms.

spectral resolution and lens parameters. Therefore, hyperspectral monitoring systems can promote the monitoring of diseases and pests over a large area in the future, essential for high-level breeding and cultivation practices of tea plants (Zhang et al., 2019b).

4. Conclusion

This study aimed at detecting and differentiating three tea plant stresses with similar symptoms, including the tea green leafhopper, anthracnose, and sunburn. Subsequently, a step-by-step detection and discrimination method of tea plant stresses was proposed and verified. The main conclusions of the study are as follows: (1) The abnormal area of tea leaves showed noticeable spectral changes under tea green leafhopper, anthracnose, and sunburn stress. However, the spectral characteristics between stresses were relatively similar. (2) The optimal spectral feature sets for stress detection (WF_{A04}, WF_{A05}, YSV, RMV, and Width) and discrimination ($WF_{S01\text{-}07}$) were obtained via continuous wavelet analysis and a rigorous feature sensitivity analysis. (3) A step-by-step procedure combining the identified optimal spectral features, image clustering, and machine learning algorithms was established, achieving an overall discrimination accuracy of more than 90%. Besides, the capability of the spectral features and models in detecting and discriminating plant stresses at leaf and canopy levels still need be assessed in the future.

CRediT authorship contribution statement

Xiaohu Zhao: Methodology, Data curation, Writing – original draft. **Jingcheng Zhang:** Conceptualization, Project administration, Supervision, Investigation, Funding acquisition. **Yanbo Huang:** Visualization, Writing – review & editing. **Yangyang Tian:** Investigation, Data curation. **Lin Yuan:** Conceptualization, Funding acquisition, Software, Validation.

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.

Acknowledgments

This work was supported by National Key R&D Program of China (2019YFE0125300), National Natural Science Foundation of China (42071420), Zhejiang Agricultural Cooperative and Extensive Project of Key Technology (2020XTTGCY04-02), and the Graduate Scientific Research Foundation of Hangzhou Dianzi University (CXJJ2020057).

References

Abdulridha, J., Batuman, O., Ampatzidis, Y., 2019. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. Remote Sensing 11 (11), 1373. https://doi.org/10.3390/rs11111373.

Baranowski, P., Jedryczka, M., Mazurek, W., Babula-Skowronska, D., Siedliska, A., Kaczmarek, J., 2015. Hyperspectral and Thermal Imaging of Oilseed Rape (Brassica napus) Response to Fungal Species of the Genus Alternaria. PLoS One 10, e122913.

Belward, A., 1991. Spectral Characteristics of Vegetation, Soil and Water in the Visible, Near-Infrared and Middle-Infrared Wavelengths, pp. 31–53.

Bora, K., Sarkar, D., Konwar, K., Payeng, B., Sood, K., Paul, R., Datta, R., Das, S., Khare, P., Karak, T., 2019. Disentanglement of the secrets of aluminium in acidophilic tea plant (Camellia sinensis L.) influenced by organic and inorganic amendments. Food Res. Int. 120, 851–864.

Chen, Z.M., Sun, X.L., Dong, W.X., 2012. Genetics and Chemistry of the Resistance of Tea Plant to Pests, pp. 343–360.

Cheng, T., Rivard, B., Sánchez-Azofeifa, A.G., Féret, J.-B., Jacquemoud, S., Ustin, S.L., 2014. Deriving leaf mass per area (LMA) from foliar reflectance across a variety of plant species using continuous wavelet analysis. ISPRS J. Photogram. Remote Sensing 87, 28–38.

Cheng, T., Rivard, B., Sánchez-Azofeifa, G.A., Feng, J., Calvo-Polanco, M., 2010. Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation. Remote Sensing Environ. 114 (4), 899–910.

Cheng, T., Rivard, B., Sánchez-Azofeifa, A., 2011. Spectroscopic determination of leaf water content using continuous wavelet analysis. Remote Sensing Environ. 115 (2), 659–670.

Demetriades-Shah, T.H., Steven, M.D., Clark, J.A., 1990. High resolution derivative spectra in remote sensing. Remote Sensing Environ. 33 (1), 55–64.

Duda, R., Hart, P., 1973. Pattern Classification and Scene Analysis. Chapter 7.5 Template Matching 95.

Fang, J., Sureda, A., Silva, A.S., Khan, F., Xu, S., Nabavi, S.M., 2019. Trends of tea in cardiovascular health and disease: A critical review. Trends Food Sci. Technol. 88, 385–396.

Gnanamangai, B.M., Ponmurugan, P., 2012. Evaluation of various fungicides and microbial based biocontrol agents against bird's eye spot disease of tea plants. Crop Protect. 32, 111–118.

Gong, P., Pu, R., Heald, R.C., 2002. Analysis of in situ hyperspectral data for nutrient estimation of giant sequoia. Int. J. Remote Sensing 23 (9), 1827–1850.

Gong, P., Pu, R.L., Yu, B., 1997. Conifer species recognition: an exploratory analysis of in situ hyperspectral data. Remote Sensing Environ. 62, 189–200.

Han, W.Y., Xiao, Q., 2013. Causes and control suggestions of drought and heat damage in tea garden in summer of 2013 (in Chinese). China Tea 35 (9), 18–19.
Hu, G., Wu, H., Zhang, Y., Wan, M., 2019. A low shot learning method for tea leaf's disease identification. Comput. Electron. Agric. 163, 104852. https://doi.org/10.1016/j.compag.2019.104852.
Jago, R.A., Cutler, M.E.J., Curran, P.J., 1999. Estimating canopy chlorophyll concentration from field and airborne spectra. Remote Sensing Environ. 68 (3), 217–224.
Jiménez-Carvelo, A.M., González-Casado, A., Bagur-González, M.G., Cuadros-Rodríguez, L., 2019. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity - A review. Food Res. Int. 122, 25–39.
Jin, S., Chen, Z.M., Backus, E.A., Sun, X.L., Xiao, B., 2012. Characterization of EPG waveforms for the tea green leafhopper, Empoasca vitis Göthe (Hemiptera: Cicadellidae), on tea plants and their correlation with stylet activities. J. Insect Physiol. 58 (9), 1235–1244.
Mahlein, A.-K., Kuska, M.T., Thomas, S., Wahabzada, M., Behmann, J., Rascher, U., Kersting, K., 2019. Quantitative and qualitative phenotyping of disease resistance of crops by hyperspectral sensors: seamless interlocking of phytopathology, sensors, and machine learning is needed! Curr. Opin. Plant Biol. 50, 156–162.
Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sensing Environ. 128, 21–30.
Mei, X., Liu, X., Zhou, Y., Wang, X., Zeng, L., Fu, X., Li, J., Tang, J., Dong, F., Yang, Z., 2017. Formation and emission of linalool in tea (*Camellia sinensis*) leaves infested by tea green leafhopper (*Empoasca (Matsumurasca) onukii* Matsuda). Food Chem. 237, 356–363.
Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sensing 26 (1), 217–222.
Pu, R.L., 2017. Hyperspectral Remote Sensing: Fundamentals and Practices. CRC Press.
Pu, R.L., Ge, S., Kelly, N.M., Gong, P., 2003. Spectral absorption features as indicators of water status in Quercus Agrifolia leaves. Int. J. Remote Sensing 24, 1799–1810.
Pu, R.L., Gong, P., 2004. Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. Remote Sensing Environ. 91, 212–224.
Roy, S., Barooah, A.K., Ahmed, K.Z., Baruah, R.D., Prasad, A.K., Mukhopadhyay, A., 2019. Impact of climate change on tea pest status in northeast India and effective plans for mitigation. Acta Ecol. Sin. 40 (6), 432–442.
Sankaran, S., Mishra, A., Ehsani, R., Davis, C., 2010. A review of advanced techniques for detecting plant diseases. Comput. Electron. Agric. 72 (1), 1–13.
Sinha, R., Khot, L.R., Rathnayake, A.P., Gao, Z., Naidu, R.A., 2019. Visible-near infrared spectroradiometry-based detection of grapevine leafroll-associated virus 3 in a red-fruited wine grape cultivar. Comput. Electron. Agric. 162, 165–173.
Sun, Y., Jiang, Z., Zhang, L., Dong, W., Rao, Y., 2019. SLIC_SVM based leaf diseases saliency map extraction of tea plant. Comput. Electron. Agric. 157, 102–109.
Taiz, L., Zeiger, E., Moller, I.M., Murphy, A., 2018. Fundamentals of Plant Physiology. Oxford University Press, Oxford, UK.
Tarabalka, Y., Fauvel, M., Chanussot, J., Benediktsson, J.A., 2010. SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images. IEEE Geosci. Remote Sensing Lett. 7 (4), 736–740.
Torrence, C., Compo, G.P., 1998. A Practical Guide to Wavelet Analysis. Bull. Am. Meteorol. Soc. 79 (1), 61–78.
Wahid, A., Gelani, S., Ashraf, M., Foolad, M., 2007. Heat tolerance in plants: an overview. Environ. Exp. Botany 61 (3), 199–223.
Wang, X.Q., Ran, L., Peng, P., Cui, Z.L., 2014. Analysis of the hyperspectral characteristics of tea leaves under anthracnose disease stress. Plant Protect. 40 (06), 13–17.
Xia, E.-H., Zhang, H.-B., Sheng, J., Li, K., Zhang, Q.-J., Kim, C., Zhang, Y., Liu, Y., Zhu, T., Li, W., Huang, H., Tong, Y., Nan, H., Shi, C., Shi, C., Jiang, J.-J., Mao, S.-Y., Jiao, J.-Y., Zhang, D., Zhao, Y., Zhao, Y.-J., Zhang, L.-P., Liu, Y.-L., Liu, B.-Y., Yu, Y., Shao, S.-F., Ni, D.-J., Eichler, E.E., Gao, L.-Z., 2017. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol. Plant 10 (6), 866–877.
Xie, C., Yang, C.e., He, Y., 2017. Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities. Comput. Electron. Agric. 135, 154–162.
Yuan, L., Huang, Y., Loraamm, R.W., Nie, C., Wang, J., Zhang, J., 2014. Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. Field Crops Res. 156, 199–207.
Yuan, L., Yan, P., Han, W., Huang, Y., Wang, B., Zhang, J., Zhang, H., Bao, Z., 2019. Detection of anthracnose in tea plants based on hyperspectral imaging. Comput. Electron. Agric. 167, 105039. https://doi.org/10.1016/j.compag.2019.105039.
Zahra, S., Ghazanfar, M.A., Khalid, A., Azam, M.A., Naeem, U., Prugel-Bennett, A., 2015. Novel centroid selection approaches for KMeans-clustering based recommender systems. Inform. Sci. 320, 156–189.
Zhang, H., Zhang, M.T., Wang, F.K., Ren, M.X., Qing, Y., Kang, X.H., 2019a. Effects of four intercropping crops on the occurrence of major leaf diseases in tea plantations in summer and autumn. J. Tea Sci. 39, 318–324.
Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., Huang, W., 2019b. Monitoring plant diseases and pests through remote sensing technology: a review. Comput. Electron. Agric. 165, 104943. https://doi.org/10.1016/j.compag.2019.104943.
Zhang, J., Wang, N., Yuan, L., Chen, F., Wu, K., 2017. Discrimination of winter wheat disease and insect stresses using continuous wavelet features extracted from foliar spectral measurements. Biosyst. Eng. 162, 20–29.
Zhang, J., yuan, L., Pu, R., Loraamm, R.W., Yang, G., Wang, J., 2014. Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat. Comput. Electron. Agric. 100, 79–87.