

Detection of scab in wheat ears using in situ hyperspectral data and support vector machine optimized by genetic algorithm

Linsheng Huang¹, Hansu Zhang^{1,2}, Chao Ruan^{1,2}, Wenjiang Huang^{1,2*},
Tingguang Hu^{1,2}, Jinling Zhao¹

(1. National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei, China;
2. Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China)

Abstract: A new method was proposed to extract sensitive features and to construct a monitoring model for wheat scab based on in situ hyperspectral data of wheat ears to achieve effective prevention and control and provide theoretical support for its large-scale monitoring. Eight sensitive features were selected through correlation analysis and wavelet transform. These features were as follows: three original bands of 350–400 nm, 500–600 nm, and 720–1000 nm; three vegetation indices of modified simple ratio (MSR), normalized difference vegetation index, and structural independent pigment index; and two wavelet features of WF01 and WF02. By combining the selected sensitive features with support vector machine (SVM) and SVM optimized by genetic algorithm (GASVM), a total of 16 monitoring models were built, and the monitoring accuracies of the two types of models were compared. The ability of the monitoring models built by GASVM to identify scab was better than that of SVM algorithm under the same characteristic variables. Among the 16 models, MSR combined with GASVM had an overall accuracy of 75% and a Kappa coefficient of 0.47. GASVM can be used to monitor wheat scab and its application can improve the accuracy of disease monitoring.

Keywords: wheat scab, hyperspectral data, correlation analysis, genetic algorithm, wavelet transform, support vector machine

DOI: 10.25165/j.ijabe.20201302.5331

Citation: Huang L S, Zhang H S, Ruan C, Huang W J, Hu T G, Zhao J L. Detection of scab in wheat ears using in situ hyperspectral data and support vector machine optimized by genetic algorithm. *Int J Agric & Biol Eng*, 2020; 13(2): 182–188.

1 Introduction

Wheat is one of the three major grains in the world, and wheat diseases have become the focus of research worldwide. Scab is a worldwide epidemic disease caused by *Fusarium asiaticum* and *F. graminearum* that mainly occurs in warm and humid areas. It decreases wheat yield and produces the deoxynivalenol toxin. This toxin can cause poisoning and even death of humans and animals. If the disease is not monitored and controlled in time, it will cause a large-scale reduction in wheat yield and degradation of grain quality, thereby resulting in economic losses^[1]. Therefore, the problem of scab must be addressed.

Traditional monitoring methods, such as on-site sampling analysis, are mainly performed by plant pathologists or experts. Such methods are costly, time consuming, and labor-intensive; they are also unsuitable for large-area applications^[2]. Remote sensing has the characteristics of multiscale and multi-time resolution. In recent years, it has been widely used in crop growth and area

monitoring and production forecasting^[3]. Hyperspectral remote sensing technology is one of the greatest achievements in the development of remote sensing technology in the late 20th century. It is applied by an increasing number of scholars who are involved in crop disease research because of its rich band information. Hyperspectral technology can obtain continuous spectral curves with spectral resolutions up to the order of nanometers, and its ability to recognize features is strong^[4]. Many researchers use hyperspectral imaging technology to directly identify the disease of crop kernels. For example, Delwiche and Kim^[5,6] identified scab in wheat kernels by using a custom-made hyperspectral imaging system and near-infrared hyperspectral system (1000–1700 nm). Liang et al.^[7] used hyperspectral technology to identify infected kernels via spectral analysis and pattern processing. The scab identification model constructed by support vector machine (SVM) and back propagation neural network achieved excellent results and an accuracy of more than 90%. Ewa et al.^[8] constructed a classification model based on texture parameters of hyperspectral images to identify infected kernels, and ventral kernels were classified with 100% accuracy. These studies all achieved good results, and they all directly identified the infected or uninfected kernels through hyperspectral technology in the laboratory. However, they excluded the influence of various factors, such as field conditions, weather, and leaves. Thus, they achieved ideal research results. Studies^[9] have demonstrated that leaf area index, chlorophyll content, and aboveground biomass are the main indices affecting scab occurrence. Li et al.^[10] directly used non-imaging hyperspectral technology to conduct canopy-scale research from the practical perspective of wheat growth conditions. A remote sensing estimation model of scab was established based on spectral reflectance data, climatic factors, and growth parameters, and this

Received data: 2019-08-12 **Accepted data:** 2019-12-23

Biographies: Linsheng Huang, PhD, Professor, research interests: Remote sensing information processing, Email: linsheng0808@163.com; Hansu Zhang, Master candidate, research interests: Remote sensing monitoring of crop diseases and pests, Email: hansu_zhang@qq.com; Chao Ruan, PhD candidate, research interests: Pests and diseases of wheat, Email: 17611337552@163.com; Tingguang Hu, Master candidate, research interests: Remote sensing monitoring of crop diseases and pests, Email: hutingguang@qq.com; Jinling Zhao, PhD, Associate Professor, research interests: Research on remote sensing monitoring and forecasting, Email: zhaojl@ahu.edu.cn.

***Corresponding author:** Wenjiang Huang, PhD, Professor, research interests: Quantitative vegetation remote sensing. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. Tel: 86-10-82178169, Email: huangwj@aircas.ac.cn.

model was used to provide a reference for the information acquisition of disease prevention and monitoring for winter wheat production in Yangze and Huai river region. A single wheat ear in the field contains purer spectral information, which can provide a theoretical basis for canopy studies and even large-scale research. Therefore, some researchers directly studied scab on a single wheat ear. Huang et al.^[11] directly measured the spectrum of a single wheat ear by using non-imaging hyperspectral technology, established an effective disease severity recognition model by performing Fisher's linear discriminant analysis and SVM based on radial basis function (RBF), and identified scab on a single wheat plant in the front, side, and upright directions. Non-imaging remote sensing technology also has great potential in the monitoring of wheat scab. However, most previous studies focused on wheat powdery mildew, stripe rust, or aphids; few studies were conducted on scab^[11]. Therefore, the present study aimed to use the non-imaging hyperspectral technique to identify wheat scab on the ear scale.

SVM is a model construction method based on statistical theory. It is typically used in pattern recognition, classification, and regression analysis^[12]. It first constructs the optimal hyperplane to minimize the classification error. Then, it transforms the input space into a high-dimensional space by nonlinear transformation of the appropriate kernel function to find the optimal classification surface in the new space^[13]. However, effectively selecting the kernel function and determining the parameters when using this algorithm are still controversial. The traditional grid search algorithm is inefficient, computationally intensive, time consuming,

and produces an unsatisfactory effect. Genetic algorithm (GA) is good for solving global optimization problems; it has strong robustness and a simple process^[14]. GA has been widely used in various studies, such as face and text recognition.

SVM optimized by genetic algorithm (GASVM) has been applied to the monitoring and identification of wheat diseases^[14], but it has not been used for wheat scab. Hence, this study aimed to do the following: (1) analyze the spectral information of wheat ears measured by a non-imaging spectrometer and select the spectral features that are sensitive and significantly different from disease severity; and (2) establish effective models for identifying wheat scab by using SVM and GASVM algorithms and prove that the GASVM algorithm is more conducive to the information detection of scab.

2 Materials and methods

2.1 Study site

Scab is sensitive to humidity and temperature, and it often occurs in temperate regions where the climate is warm, moist, and rainy^[10]. Guohe Town (31°29'N, 117°13'E), Baihu Town (31°14'N, 117°27'E), and Shucheng County (31°32'N, 116°59'E) in Anhui Province were selected as the research sites (Figure 1), and the field spectrum acquisition time was May 2018. The average temperature is 14 °C-17 °C, and the annual rainfall is 770-1700 mm^[15]. This is the proper temperature and the sufficient amount of moisture suitable for scab occurrence. Hyperspectral data were measured during the grain filling stage of wheat, which is an important period for scab detection.

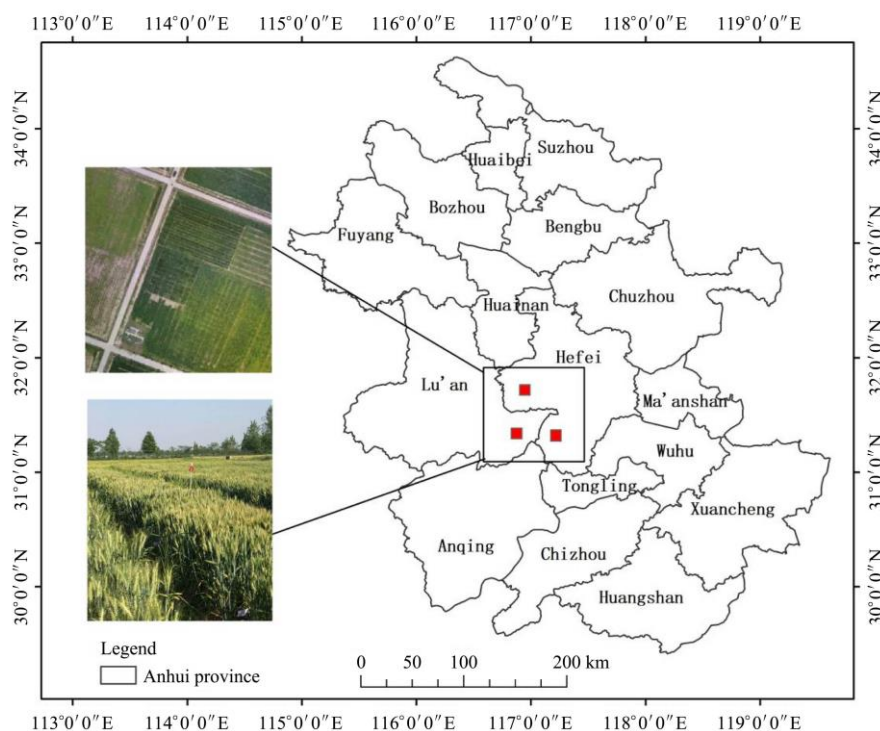


Figure 1 Location of the study area

2.2 Data acquisition

The statistical results of the wheat ear samples measured in the field are shown in Table 1. The analytical spectral device FieldSpec Pro full range spectrometer (350-2500 nm) was used to collect spectral information. Its spectral resolution was 3 nm within the 350-1000 nm range and 10 nm within the 1000-2500 nm range^[11]. All in situ hyperspectral data were measured in a windless, cloudless, and sunny environment from 10:00-14:00.

To eliminate the interference of other wheat ears, we cut a hole in the center of a 1 m×1 m black cloth, inserted the wheat ear vertically into the black cloth, and placed the probe of the spectrometer on top of the ear to measure the spectrum. Each ear was measured 10 times, and a 40 cm×40 cm BaSO₄ calibration panel was used for spectrum correction before each measurement. The average value of the 10 measurements was recorded. All spectral curves were resampled at 1 nm intervals before

pretreatment. The spectral reflectance could be obtained by calculating the ratio of radiation brightness between wheat ears and panel radiation. The calculation formula is as follows:

$$R_1 = \frac{DN_1}{DN_2} \times R_2 \tag{1}$$

where, R_1 is the target spectral reflectance; DN_1 is the gray value of the target spectrum; DN_2 is the gray value of the calibration panel, and R_2 is the calibration panel reflectance.

Table 1 Number of healthy and diseased samples collected

Number of samples collected		
Total number	Number of healthy samples	Number of diseased samples
72	24	48

2.3 Assessment of disease severity

According to the rules for monitoring and forecast of the wheat head blight (GB/T15796-2011), disease severity is the proportion of diseased spikelets (ear rot, decay, etc.) to the total number spikelets. The grading standards of wheat scab occurrence severity are shown in Table 2, and the samples of levels 0-4 were 8, 16, 11, 20, and 17, respectively. In order to reduce the difficulty of identifying the severity of the disease, the severity of scab was further reclassified into a healthy class and diseased class. Given that the infected wheat ears in levels 0 and 1 were difficult to distinguish, we classified the level 1 samples as healthy wheat ears, and the level 2-4 samples were divided into infected wheat ears.

Table 2 Severity grading of wheat scab occurrence

Level	The proportion of diseased spikelets to all spikelets
0	0
1	0-1/4
2	1/4-1/2
3	1/2-3/4
4	3/4-1

2.4 Data analysis

2.4.1 Selection of vegetation indices

Vegetation index-based analysis is a major approach for studying and practicing remote sensing of pests and diseases. According to the spectral characteristics of crops under stress, researchers have constructed a variety of vegetation indices for monitoring crop diseases and insect pests^[16]. To determine the sensitive features of physiological and biochemical changes induced by scab, we selected 10 vegetation indices (Table 3) that were combined and transformed by different wavebands as the primary feature sets of the monitoring model and discussed their applicability in assessing scab. These indices were as follows: modified simple ratio (MSR) and normalized difference vegetation index (NDVI) related to biophysical pigments; nitrogen reflectance index (NRI) related to water and nitrogen content; photochemical reflectance index (PRI) and physiological reflectance index (PhRI) related to photosynthetic activity; ratio vegetation stress index (RVSI) related to cell structure; and structural independent pigment index (SIPI), normalized pigment chlorophyll index (NPCI), anthocyanin reflectance index (ARI), and triangular vegetation index (TVI) related to pigment variation.

2.4.2 Wavelet transform

Wavelet transform can realize data filtering and de-noising. It has multiresolution characteristics. Each channel can obtain local detail features of the data by adopting multichannel filtering, which highlights the sensitive information of the data. Therefore, the utilization of spectral information is optimized to some

extent^[12].

Table 3 Characteristics of vegetation indices

Vegetation index	Formula	Reference
Modified Simple Ratio (MSR)	$(\rho_{800}/\rho_{670}-1)/(\rho_{800}/\rho_{670}+1)^{1/2}$	[17,18]
Normalized Difference Vegetation Index (NDVI)	$(\rho_{840}-\rho_{675})/(\rho_{840}+\rho_{675})$	[19]
Nitrogen Reflectance Index (NRI)	$(\rho_{570}-\rho_{670})/(\rho_{570}+\rho_{670})$	[20]
Photochemical Reflectance Index (PRI)	$(\rho_{570}-\rho_{531})/(\rho_{570}+\rho_{531})$	[21]
Structural Independent Pigment Index (SIPI)	$(\rho_{800}-\rho_{445})/(\rho_{800}-\rho_{680})$	[22]
Physiological Reflectance Index (PhRI)	$(\rho_{550}-\rho_{531})/(\rho_{550}+\rho_{531})$	[23]
Normalized Pigment Chlorophyll Index (NPCI)	$(\rho_{680}-\rho_{430})/(\rho_{680}+\rho_{430})$	[24]
Anthocyanin Reflectance Index (ARI)	$(\rho_{550})^{-1} - (\rho_{700})^{-1}$	[25]
Ratio Vegetation Stress Index (RVSI)	$[(\rho_{712}+\rho_{752})/2] - \rho_{732}$	[26]
Triangular Vegetation Index (TVI)	$60(R_{750}-R_{550}) - 100(R_{670}-R_{550})$	[27]

Gabor wavelet enables simultaneous local analysis of time and frequency, thereby analyzing stationary signals easily. Gabor wavelet transform solves the expansion coefficient of Gabor^[28]. In this study, the Gaussian function was adopted as the mother wavelet to construct the wavelet kernel function, the vegetation index was convoluted with the wavelet kernel function, and the amplitude after convolution was used as the modeling feature information. The formula is as follows^[29]:

$$\begin{cases} g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \\ h(x, y) = g(x, y) \exp(2\pi jWx) \\ H(u, v) = \exp\left(-\frac{1}{2}\left(\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)\right) \end{cases} \tag{2}$$

where, $g(x, y)$ represents a Gaussian modulation function; σ_x and σ_y represent the standard deviation on the X and Y axes; $h(x, y)$ represents a wavelet function; W is the frequency of the sine function on the X -axis; $H(u, v)$ is the Fourier transform; u is the frequency of independent variable; v is the amplitude value of the frequency signal, and σ_u and σ_v are the standard deviations on the U and V axes, respectively.

$$S(x, y) = \sqrt{(h_R * I)^2 + (h_I * I)^2} \tag{3}$$

where, $(h * I)$ is the convolution of filter h with data I ; h_R and h_I are the real and imaginary parts of filter h , and $S(x, y)$ is the characteristic obtained by Gabor filter. $h(x, y)$ as the mother wavelet can be scaled and rotated to obtain a set of self-similar filters:

$$h_m(x, y) = \alpha^{-m} h(x', y') \tag{4}$$

where, $x' = \alpha^{-m}(x \cos \theta + y \sin \theta)$, $y' = \alpha^{-m}(-x \sin \theta + y \cos \theta)$, $\alpha > 1$, $\theta = n\pi/K$, α^{-m} is scale factors, $m=0, 1, \dots, t-1, n=0, 1, \dots, K-1$, and t and K are the number of scales and directions, respectively. In this experiment, $t = 4$, $K = 8$, and $\alpha = \sqrt{2}$.

2.4.3 Classification method

The principle of SVM involves finding an optimal hyperplane that satisfies the classification requirements. The hyperplane maintains the classification accuracy and maximizes the interval between the two types of classification samples^[30]. It is widely used in remote sensing classification because of its simple structure, strong adaptability, and global optimal characteristics. The discriminant function of the model is as follows:

$$f(x) = \text{sgn}\left(\sum_{x_i \in S_v}^n k\alpha_i y_i(x_i, x) + b\right) \quad (5)$$

where, α_i is the Lagrange multiplier; S_v is the support vector; x_i and y_i are support vectors in two classes; b is the threshold, and $k(x_i, x)$ is the kernel function. In this study, RBF was selected as the kernel function of SVM.

How to set the penalty factor and kernel parameter is the problem that SVM faces in practical application. The traditional methods of parameter selection are mostly by trial and error. Cross-validation is widely used in modeling applications, but this method is inefficient and requires a heavy workload. GA was used to optimize the SVM. The steps are as follows^[31]:

- 1) The population was initialized.
- 2) Select training data and validation data. Seventy-two sample plots in the field were investigated during the study period. Forty-two samples were designated as the training samples. The remaining 24 samples were the verification samples.
- 3) Using SVM to train and test the data. The fitness function was selected, and the fitness value of each individual was calculated.
- 4) The maximum genetic algebra (100) of the initial setting was assessed if it was satisfied, and the optimal penalty factor and kernel parameter were obtained when the condition was met. Otherwise, the next step was performed.
- 5) Crossover operator and mutation operator were performed to form a new generation of individuals. Step 2 was performed again to continue the algorithm's optimization until the termination condition is met to exit the loop.
- 6) The parameter-optimized SVM model was used to detect wheat scab.

3 Results and discussion

3.1 Selection of sensitive features

3.1.1 Features of spectral reflectance

Spectral reflectance is the simplest and most direct feature. The spectral reflectance signals in visible and near-infrared regions reflect the changes in physical and biochemical components caused by vegetation stress. These signals have been widely used in remote sensing monitoring and early stress diagnosis of crop diseases and pests^[16]. Figure 2 shows the spectral reflectance curves of healthy and infected wheat ears. The morphological difference between the two spectra was not obvious. To perceive the changes in the spectrum more intuitively, we calculated the reflectance ratio between the diseased and healthy wheat ears (Figure 3). The reflectance ratio reached the maximum in the 500-600 nm band regions, thereby indicating that the spectral reflectance of wheat after stress in this band was greatly improved. This result was obtained, because when wheat is under stress, the content of chlorophyll decreases and the absorption ability of visible regions is weakened, thus, the reflectance at the "green peak" increases^[4]. Correlation analysis was performed to assess whether significant relationships exist between band reflectance and disease severity. In Figure 4, the reflectance of 350-400 nm in the visible regions exhibited the highest correlation with disease severity, and all correlation coefficients in the near-infrared regions (720-1000 nm) were greater than 0.7. Therefore, the bands of 500-600 nm with the greatest variation and 350-400 nm and 720-1000 nm with the largest correlation coefficients in the visible and near-infrared regions were selected as the preferred features of the original spectral feature set.

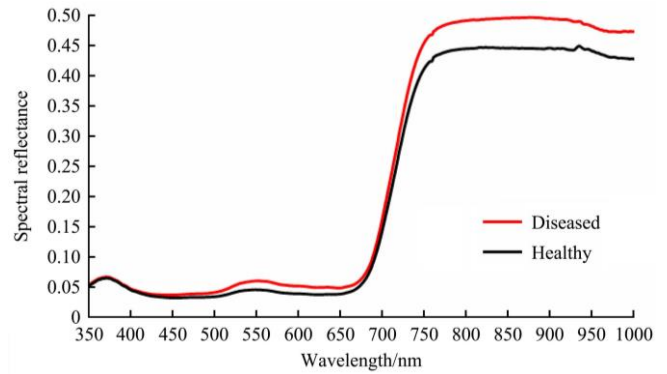


Figure 2 Comparison of hyperspectral curves between diseased and healthy ears

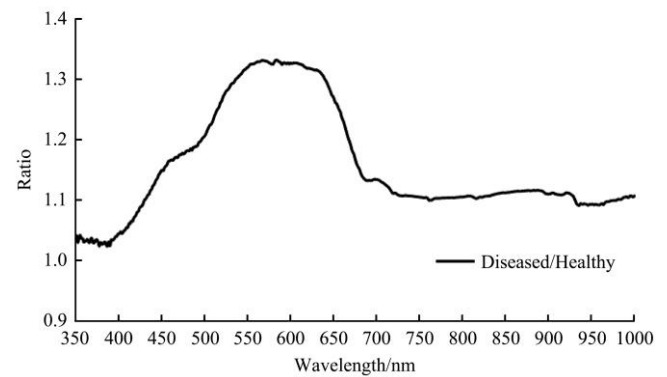


Figure 3 Curve of reflectance ratios between diseased and healthy wheat ears

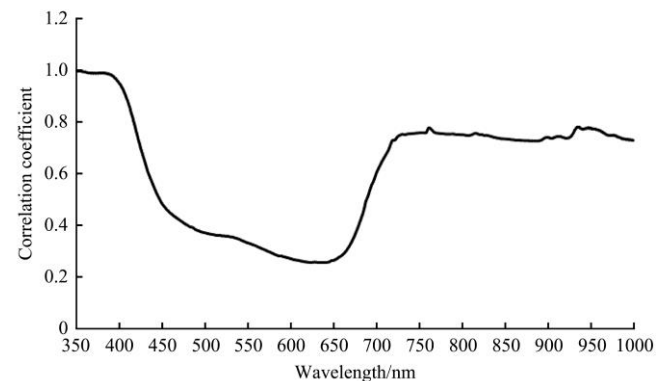


Figure 4 Correlation coefficient curve between the spectral reflectance of infected wheat and disease severity

3.1.2 Vegetation indices

Vegetation indices based on a certain physiological significance can enhance and highlight some spectral changes to obtain a more ideal result^[16]. Table 4 summarizes the responses of all vegetation indices to wheat scab. Three vegetation indices, namely, MSR, NDVI, and SIPI, were significantly correlated (p -value < 0.01) with disease severity; the correlation coefficients were 0.62, 0.58, and 0.55, respectively. Other indices did not show significant responses to disease severity. Thus, we chose MSR, NDVI, and SIPI as the sensitive vegetation indices.

3.1.3 Wavelet features

Three vegetation indices, namely, MSR, NDVI, and SIPI, were transformed by wavelet transform. A total of 32 wavelet kernel functions (four scales and eight directions) were constructed, and they increased the data dimension by 32 times. To determine the best wavelet features, we analyzed the correlation between wavelet features and disease severity and selected the features with significant differences (p -value < 0.01) as the sensitive features by

t-test. The scale, direction, and features' names are shown in Table 5.

Table 4 Correlation between vegetation indices and disease severity

Vegetation index	R
MSR*	0.62
NDVI*	0.58
NRI	0.04
PRI	0.54
SIPI*	0.55
PhRI	0.26
NPCI	0.54
ARI	0.38
RVSI	0.51
TVI	0.43

Note: * Indicates the correlation is significant at the 0.99 confidence level.

Table 5 Wavelet features

Vegetation index	Scale	Direction	Wavelet feature
MSR	4	2	WF01
NDVI	1	0	WF02

3.2 Construction of disease monitoring model

3.2.1 Construction of monitoring models based on original spectrum

A total of 72 samples were collected in this experiment, among which 42 random samples were used as the training samples, and the remaining 24 samples were used as the verification samples. According to the three bands regions of 350-400 nm, 500-600 nm, and 720-1000 nm selected in 3.1.1, the three bands with the largest correlation in the three regions were 354 nm, 525 nm, and 761 nm. These bands were the input variables, and SVM and GASVM were used for the model construction. The monitoring results, overall accuracy (OA), and Kappa coefficient of the six models are shown in Table 6.

Table 6 Accuracy analysis of the classification models based on original bands

Feature	Model	Healthy	Diseased	Sum	U/%	OA/%	Kappa	
λ354	SVM	Healthy	6	8	14	42.9	58.3	0.21
		Diseased	2	8	10	80		
		Sum	8	16	24			
		P/%	75	50				
	GASVM	Healthy	6	7	13	46.2	62.5	0.27
		Diseased	2	9	11	81.8		
		Sum	8	16	24			
		P/%	75	56.2				
λ525	SVM	Healthy	5	9	14	35.7	50	0.05
		Diseased	3	7	10	70		
		Sum	8	16	24			
		P/%	62.5	43.7				
	GASVM	Healthy	5	8	13	38.5	54.2	0.11
		Diseased	3	8	11	72.7		
		Sum	8	16	24			
		P/%	62.5	50				
λ761	SVM	Healthy	6	8	14	43	58.3	0.21
		Diseased	2	8	10	80		
		Sum	8	16	24			
		P/%	75	50				
	GASVM	Healthy	4	4	8	50	66.7	0.25
		Diseased	4	12	16	75		
		Sum	8	16	24			
		P/%	50	75				

The accuracy of the monitoring model established by GASVM was better than that by SVM algorithm based on the same characteristic variables. In the SVM model, the overall accuracy of the monitoring model at 761 and 354 nm, both of which were 58.3%, was slightly higher than that at 525 nm, whereas the monitoring model at the 525 nm band had a monitoring accuracy of only 50%. In the GASVM model, the overall accuracy at 756 nm was 4.2% and 12.5% higher than that at 354 nm and 525 nm, respectively, but it was only 66.7%, and the Kappa coefficient of the model was 0.25. In general, the overall accuracy of all models was not high.

3.2.2 Construction of monitoring models based on vegetation indices

Three vegetation indices, namely, MSR, NDVI, and SIPI, were used as the input variables to build the model. Table 7 shows the results of the models constructed by SVM algorithm and GASVM for scab detection. MSR had the highest identification ability for scab, followed by NDVI and SIPI. With the same input variables, the accuracy of the GASVM model was higher than that of the SVM model. The ability of the MSR-GASVM model to detect scab was better than that of the other models. The overall accuracy was 75%, and the Kappa coefficient was 0.47. In general, the vegetation index was better at detecting scab than the original waveband, because the vegetation index enhances the difference between healthy and infected samples by combining and transforming band reflectance^[32].

Table 7 Accuracy analysis of the classification models based on vegetation indices

Feature	Model	Healthy	Diseased	Sum	U/%	OA/%	Kappa	
NDVI	SVM	Healthy	5	7	12	41.7	58.3	0.17
		Diseased	3	9	12	75		
		Sum	8	16	24			
		P/%	62.5	56.3				
	GASVM	Healthy	5	5	10	50	66.7	0.29
		Diseased	3	11	14	78.6		
		Sum	8	16	24			
		P/%	62.5	68.8				
MSR	SVM	Healthy	5	5	10	50	66.7	0.29
		Diseased	3	11	14	78.6		
		Sum	8	16	24			
		P/)	62.5	68.8				
	GASVM	Healthy	6	4	10	60	75	0.47
		Diseased	2	12	14	85.7		
		Sum	8	16	24			
		P/%	75	75				
SIPI	SVM	Healthy	2	5	7	28.6	54.2	-0.065
		Diseased	6	11	17	64.7		
		Sum	8	16	24			
		P/%	25	68.8				
	GASVM	Healthy	6	7	13	46.2	62.5	0.27
		Diseased	2	9	11	81.8		
		Sum	8	16	24			
		P/%	75	56.3				

3.2.3 Construction of monitoring models based on wavelet features

As discussed in Section 3.1.3, WF01 and WF02 were sensitive to scab at ear scale. We designed four scab identification models using these two sensitive features. Table 8 shows the results of discriminating between healthy wheat and wheat infected by scab

by using WF0 and WF1. We can see that the model constructed by GASVM had better identification results than the SVM model, and the overall accuracy was 70.8%. For the SVM monitoring model, the highest overall accuracy of the verified samples was only 62.5%, and the Kappa coefficient was 0.23. The results show the same property as the monitoring model constructed by vegetation indices and original wavebands. The wavelet features were superior to waveband features in terms of the ability to identify scab, but they were inferior to MSR, thereby indicating that the MSR-GASVM model can ideally identify scab-infected ears.

Table 8 Accuracy analysis of the classification models based on wavelet features

Feature	Model	Healthy	Diseased	Sum	U/%	OA/%	Kappa	
WF01	SVM	Healthy	4	7	11	36.4	54.2	0.06
		Diseased	4	9	13	69.2		
		Sum	8	16	24			
		P/%	50	56.3				
	GASVM	Healthy	3	2	17	17.6	70.8	0.49
		Diseased	5	14	7	20		
		Sum	8	16	24			
		P/%	37.5	87.5				
WF02	SVM	Healthy	5	6	11	45.5	62.5	0.23
		Diseased	3	10	13	76.9		
		Sum	8	16	24			
		P/%	62.5	62.5				
	GASVM	Healthy	4	3	7	57.1	70.8	0.32
		Diseased	4	13	17	76.5		
		Sum	8	16	24			
		P/%	50	81.3				

When wheat is infected by scab, its appearance or internal structure changes and shows some differences in spectral reflectance and radiation characteristics. This phenomenon is also the fundamental theoretical basis for the identification of wheat scab by remote sensing spectroscopy^[9]. In addition to selecting the appropriate algorithm for feature extraction, choosing the suitable modeling method also has a great impact on the improvement of the monitoring level of crop diseases^[33]. According to the results obtained in the present study, the identifications of disease severity based on GASVM models were generally ideal, and the MSR-GASVM model achieved the best classification accuracy (75%) in terms of discriminating between healthy wheat and wheat infected by scab. GA has an advantage in terms of global optimization of solution space. SVM approximates the nonlinear arbitrarily and establishes an excellent nonlinear mapping model with small samples. It overcomes the defects of some traditional intelligent algorithms that easily fall into the local minimum and has high stability and robustness in prediction and control; however, it has difficulties in kernel function or parameter selection^[34]. Combining these two algorithms is beneficial because they complement each other and greatly improve the overall accuracy of the model. This combined model has a higher practical significance than the disease recognition model constructed by a single SVM algorithm.

4 Conclusions

In this study, we used the non-imaging hyperspectral technique to collect spectral information and analyzed the spectral changes in wheat ears under stress. Using the original spectral features,

vegetation indices, and wavelet features combined with SVM algorithm and GASVM to develop wheat scab monitoring models. Both model types identified wheat infected with scab, but the overall accuracy of the GASVM model was higher than that of the SVM model. The MSR-GASVM model had the best identification performance. The results had relevant implications for identifying wheat scab at the ear scale and provided a theoretical reference for the further study of the identification of scab at the canopy or regional scale.

Acknowledgments

This work was supported by National Natural Science Foundation of China (41571354, 41871339), Natural Science Research Project of Anhui Provincial Education Department (KJ2019A0030) and Anhui Provincial Science and Technology Project (16030701091, 201904f06020038), supported by Hainan Provincial Key R&D Program of China (ZDYF2018073), National special support program for high-level personnel recruitment (Wenjiang Huang).

References

- [1] Guan E Q, Cui G J, Bian K, Zheng Z H. SIMCA identification model establishment of gibberellic disease wheat grain based on near infrared spectrum characteristics. *Journal of the Chinese Cereals and Oils Association*, 2016; 31(11): 125–129. (in Chinese)
- [2] Zheng Q, Huang W J, Cui X M, Shi Y, Liu L Y. New spectral index for detecting wheat yellow rust using Sentinel-2 multispectral imagery. *Sensors*, 2018; 18(3): 868. doi: 10.3390/s18030868.
- [3] Huang Q, Li D D, Chen Z X, Liu J, Wang L M. Monitoring of planting area and growth condition of winter wheat in China based on MODIS data. *Transactions of the CSAM*, 2012; 43(7): 163–167. (in Chinese)
- [4] Fan Y B, Gu X H, Wang S T, Yang G J, Wang L, Wang L Z, et al. Analysis of canopy spectral characteristics of winter wheat powdery mildew and disease index inversion. *Journal of Triticeae Crops*, 2017; 37(1): 136–142. (in Chinese)
- [5] Delwiche S R, Kim M S. Hyperspectral imaging for detection of scab in wheat. *Biological Quality and Precision Agriculture II. International Society for Optics and Photonics*, 2000; 4203: 13–20.
- [6] Delwiche S R, Kim M S. Fusarium damage assessment in wheat kernels by Vis/NIR hyperspectral imaging. *Sensing and Instrumentation for Food Quality and Safety*, 2011; 5: 63–71.
- [7] Liang K, Du Y Y, Lu W, Wang C, Xu J H, Shen M X. Identification of fusarium head blight wheat based on hyperspectral imaging technology. *Transactions of the CSAM*, 2016; 47(2): 310–315. (in Chinese)
- [8] Ropelewska E, Zapotoczny P. Classification of Fusarium-infected and healthy wheat kernels based on features from hyperspectral images and flatbed scanner images: A comparative analysis. *European Food Research and Technology*, 2018; 244(8): 1453–1462.
- [9] Li W G, Chen H, Jin Z T, Zhang Z Z, Ge G X, Ji F J. Remote sensing monitoring of winter wheat scab based on suitable scale selection. *Journal of Triticeae Crops*, 2018; 38(11): 1374–1380. (in Chinese)
- [10] Li W G, Huang W J, Dong Y Y, Chen H, Wang J J, Shan J. Estimation on winter wheat scab based on combination of temperature, humidity and remote sensing vegetation index. *Transactions of the CSAE*, 2017; 33(23): 203–210. (in Chinese)
- [11] Huang L S, Wu Z C, Huang W J, Ma H Q, Zhao J L. Identification of fusarium head blight in winter wheat ears based on fisher's linear discriminant analysis and a support vector machine. *Applied Sciences*, 2019; 9(18): 3894. doi: 10.3390/app9183894.
- [12] Huang L S, Liu W J, Huang W J, Zhao J L, Song F R. Remote sensing monitoring of winter wheat powdery mildew based on wavelet analysis and support vector machine. *Transactions of the CSAE*, 2017; 33(14): 188–195. (in Chinese)
- [13] Chi D C, Zhang L F, Li X, Wang K, Wu X M, Zhang T N. Through prediction model based on genetic algorithm optimization support vector machine (SVM). *Journal of Shenyang Agricultural University*, 2013; 44(2): 190–194. (in Chinese)

- [14] Huang L S, Ruan C, Huang W J, Shi Y, Peng D L, Ding W J. Wheat powdery mildew monitoring based on GF-1 remote sensing image and relief-mRMR-GASVM model. *Transactions of the CSAE*, 2018; 34(15): 167–174. (in Chinese)
- [15] Zhao J L, Xu C, Xu J P, Huang L S, Zhang D Y, Liang D. Forecasting the wheat powdery mildew (*Blumeria graminis f. Sp. tritici*) using a remote sensing-based decision-tree classification at a provincial scale. *Australasian Plant Pathology*, 2018; 47: 53–61.
- [16] Huang W J, Zhang J C, Shi Y, Dong Y Y, Liu L Y. Progress in monitoring and forecasting of crop pests and diseases by remote sensing. *Journal of Nanjing University of Information Science and Technology (Nature Science Edition)*, 2018; 10(1): 30–43. (in Chinese)
- [17] Ma H Q, Huang W J, Jing Y S. Wheat powdery mildew forecasting in filling stage based on remote sensing and meteorological data. *Transactions of the CSAE*, 2016; 32(9): 165–172. (in Chinese)
- [18] Zhao J, Huang W J, Zhang Y H, Jing S Y. Inversion of leaf area index during different growth stages in winter wheat. *Spectroscopy and Spectral Analysis*, 2013; 33(9): 2546–2552. (in Chinese)
- [19] Schell J A. Monitoring vegetation systems in the great plains with ERTS. *NASA Special Publication*, 1973; 351: 309.
- [20] Filella I, Serrano L, Serra J, Penuelas J. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Science*, 1995; 35(5): 1400–1405.
- [21] Gamon J A, Penuelas J, Field C B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 1992; 41(1): 35–44.
- [22] Penuelas J, Baret F, Filella I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 1995; 31(2): 221–230.
- [23] Guan Q S, Huang W J, Zhao J L, Liu L Y, Liang D, Huang L S, et al. Quantitative identification of yellow rust, powdery mildew and fertilizer-water stress in winter wheat using in-situ hyperspectral data. *Sensor Letters*, 2014; 12: 876–882.
- [24] Penuelas J, Gamon J A, Fredeen A L, Merino J, Field A B. Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sensing of Environment*, 1994; 48(2): 135–146.
- [25] Gitelson A A, Merzlyak M N, Chivkunova O B. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 2001; 74(1): 38–45.
- [26] Merton R N, Huntington J F. Early simulation results of the ARIES-1 satellite sensor for multi-temporal vegetation research derived from AVIRIS. In: *Proceedings of the Eighth Annual JPL Airborne Earth Science Workshop*. Pasadena, CA: JPL publication, 1999; 299–307.
- [27] Broge N H, Leblanc E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 2000; 76(2): 156–172.
- [28] Ye C. The optimization algorithm of face recognition based on Gabor wavelets and SVM. Taiyuan: North University of China, 2014.
- [29] Lu J J, Huang W J, Jiang J B, Zhang J C. Comparison of wavelet features and conventional spectral features on estimating severity of stripe rust in winter wheat. *Journal of Triticeae Crops*, 2015; 35(10): 1456–1461. (in Chinese)
- [30] Tang C C, Huang W J, Luo J H, Liang D, Zhao J L, Huang L S. Forecasting wheat aphid with remote sensing based on relevance vector machine. *Transactions of the CSAE*, 2015; 31(6): 201–207. (in Chinese)
- [31] Chen L. Evaluation of the ecology and environment of Shanxi-Shaanxi-Gansu-Ningxia area based on PCA-GA-SVM. *Arid Land Geography*, 2015; 38(6): 1262–1269. (in Chinese)
- [32] Jiang J B, Li Y F, Guo H Q, Liu Y Q, Chen Y H. Spectral characteristics and identification research of soybean under different disease stressed. *Spectroscopy and Spectral Analysis*, 2012; 32(10): 2775–2779.
- [33] Ma H Q, Huang W J, Jing S Y, Dong Y Y, Zhang J C, Nie C W, et al. Remote sensing monitoring of wheat powdery mildew based on AdaBoost model combining mRMR algorithm. *Transactions of the CSAE*, 2017; 33(5): 162–169. (in Chinese)
- [34] Man C T, Liu B, Cao Y C. The application based on support vector machine optimized by particle swarm optimization and genetic algorithm. *Journal of Harbin University of Science and Technology*, 2019; 24(3): 88–92. (in Chinese)