Nitrogen detection of *Dendrobium nobile* based on hyperspectral images

Teng Long[1,2], Yongbing Long[1,2], Houcheng Liu[3], Hailin Liu[4], Zaihua Wang[4*], Jing Zhao[1,2*]

(1. *School of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China*;
2. *Lingnan Modern Agricultural Science and Technology Guangdong Laboratory, Guangzhou 510642, China*;
3. *College of Horticulture, South China Agricultural University, Guangzhou 510642, China*;
4. *Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences/Guangdong Provincial Key Lab of Ornamental Plant Germplasm Innovation and Utilization, Guangzhou 510640, China*)

Abstract: *Dendrobium nobile* Lindl. (*D. nobile*), as an important traditional Chinese medicine and highly ornamental value plant, has attracted more and more people's attention. In order to meet the needs of tracking and testing the growth status of *D. nobile*, the visible-near infrared hyperspectral imaging technology was proposed for nitrogen nutrients detection in vivo in its different growth stages. Firstly, collecting the hyperspectral images of *D. nobile* in spectral range 400-1000 nm. Secondly, extracting the region of interesting (ROI). The 2G-R-B algorithm was used to segment the background and plants, and then the RGB threshold method was used to separate the leaf sheath and stems. Removing noise by two masks' or-operations, and then the ROI area was finally extracted by selecting the largest area. After that, the reflectance spectrum of the ROI area was extracted, and then two kinds of feature extraction methods and two kinds of optimizing band selection methods were researched for dimension reduction of hyperspectral images. Finally, Support vector machine (SVM) model was established to classify the nitrogen level of *D. nobile*. The results showed that the LDA combined with the SVM algorithm had the highest classification accuracy. The classification accuracy of training sets in the three growth stages were 97.47%, 95.03%, and 95.97%, respectively, and the classification accuracy of test set reached 97.00%, 88.8%, 92.67%. The visible-near infrared hyperspectral imaging technology combining LDA-SVM classification model could effectively distinguish *D. nobile* cultivated by gradient nitrogen in each growth stage. It is a potential technology applied in decision-making of precise nutrition supply.

Keywords: facility gardening, hyperspectral imaging technology, *Dendrobium nobile* Lindl., nitrogen detection, SVM
DOI: 10.33440/j.ijpaa.20200302.87

Citation: Long T, Long Y B, Liu H C, Liu H L, Wang Z H, Zhao J. Nitrogen detection of Dendrobium nobile based on hyperspectral images. Int J Precis Agric Aviat, 2020; 3(2): 73–82.

1 Introduction

Dendrobium nobile Lindl. (*D. nobile*), also known as Flat gold hairpin, Astragalus sibiricus, etc., is a *Dendrobium* plant in the family Orchidaceae, which has been listed as the primary source of medicinal *Dendrobium* in the *Pharmacopoeia of the People's Republic of China (2015)*, because of its functions of moistening lungs and cough, improving eyesight, strengthening body, and replenishing stomach. Its main medicinal parts are fresh or dried stems, while its main active ingredients include dendromine and polysaccharides[1,2]. At the same time, the flowers of *D. nobile* are elegant and colorful, known as one of the "four ornamental orchids"[3,4], also called as "appreciated food" plant, because it has both ornamental value and medicinal value. During the growth of *Dendrobium*, the supply of nitrogen fertilizer which is related to the yield and quality level of *Dendrobium*[5-10] is one of the most important aspects of *Dendrobium* cultivation. When the nitrogen content of cultivation liquid increases, the nitrogen content of *Dendrobium* leaves increases, chlorophyll synthesis accelerates, and the synthesis of active ingredients also accelerates (Shi Jun et al.[11]). And with the increase of nitrogen concentration, the leaves, plant height, leaf width, and leaf length of *Dendrobium* show a very significant upward trend (Li Yuan et al.[12]). Wang Zaihua et al.[13] found that nitrogen was one of the main factors affecting the growth of the main stem of *Dendrobium* and the absorption of mineral elements. A suitable nitrogen levels can accelerate the growth of the main stem and normal flowering, and the flowering time is obviously about 2 weeks in advance in their experiment. It is so important to accurately identify the nitrogen content level of *D. nobile* in vivo for guaranteeing the quality of plant growth[14-18]. At the same time, it is also the key to realize reasonable fertilization in the refined management of facility horticulture[19-21].

Nowadays, Chemical analysis were common method for detecting plant nutrients[22]: obtain the nitrogen content by extracting the sample solution and measure the absorbance of the solution complex. Such a detection method is simple in operation and accurate in result, but it will damage the plants, cannot achieve in-vivo detection or detect all plants, and is not suitable for a large number of samples. In order to achieve real-time, non-destructive detection of *Dendrobium*, the hyperspectral imaging technology was proposed in this paper.

Hyperspectral imaging is a kind of multi-narrow band images data technology, which combined imaging technology and

Received date: 2020-04-10 **Accepted date:** 2020-06-15
Biographies: Teng Long, master candidate, research interests: application of hyperspectral imaging technology in precision agriculture, Email:2997978135@qq.com; **Yongbing Long**, PHD, professor, research interests: machine vision, artificial intelligence in precision agriculture, Email:yongbinglong@126.com; **Houcheng Liu**, PhD, Professor, research interests: protected horticulture production, Email: liuhch@scau.edu.cn; **Hailin Liu**, Assistant research fellow, research interests: exploitation and utilization of edible medicinal plant resources and physiological research, Email:liuhailin2006@163.com.
***Corresponding author: Jing Zhao**, PHD, associate professor, research interests: machine vision, artificial intelligence in precision agriculture, Email: edithzj@126.com; **Zaihua Wang**, Research associate, director, research interests: facility agriculture and healthy flowers, Email: wangzaihua@163.com.

spectroscopy technology to detect two-dimensional geometric space and one-dimensional spectral information of the target, so it could obtain continuous, narrow-band images with high spectral resolution. Since hyperspectral imaging can simultaneously obtain the composition information and its spatial distribution information of plants, the application in the detection of nutrient elements in plants has received more and more attention[23-25]. For example, Shi Jiyong et al.[26] collected the hyperspectral images of cucumbers leaves whose nutrients supply were accurately controlled , and then combined with chemometric methods to detect and find out the distribution of chlorophyll of leaves lacking of *N* and *Mg*. Zhu Wenjing[27] found out the sensitive bands of reflection hyperspectral images for tomato nutrition detection. Sun Jun et al.[28] selected the ROI region of hyperspectral images of lettuce cultivated by different nitrogen concentration, and then constructed the classification model of different leaf nitrogen level of born vegetables. *Dendrobium* has more stringent requirements on nutritional supply[3,4], and the non-destructive detection technology for *Dendrobium* nutrient elements has not been reported as far as we know. The hyperspectral imaging technology was proposed to collect images of *D. nobile* in this paper, and ROI selection and dimensionality reduction methods were discussed. At last, the classification model based on support vector machine was established to realize the classification of *D. nobile* cultivated by gradient nitrogen level. The study showed the potential of hyperspectral imaging in intelligent monitoring for the precise nutrition supply of *D. nobile*.

2 Materials and method

2.1 Materials

The samples of *D. nobile* from different seedling stages were collected from Baiyun Base, Institute of Environmental Horticulture, Guangdong Academy of Agricultural Sciences. As for the sample gradient nutrition supply cultivation, we referred to the modified Hogland nutrient solution formula, and prepared 5 different nutrient solution fertilizer 1.5N, N, 1/2N, 1/3N, 0N nutrient solutions, corresponding to *D. nobile* T_1, T_2, T_3, T_4, T_5 (CK) treatment. The nutrient solution was irrigated every 7 days, and 100 mL/ time was irrigated in each pot. A total of 450 samples were taken from 30 samples at each of three growth stages and 5 treatments at the seedling stage, growth stage and mature stage. In this paper, the seedling stage (*S*), growth stage (*M*) and maturity stage (*L*) of *D. nobile*. in T_i (i=1,2,…,5) cultivating method were denoted as S_i, M_i, L_i (i=1,2,…,5).

2.2 Experimental device and image acquisition

A hyperspectral imaging device (Gaiafield Pro-V10 kit, Shuanglihepu, Sichuan, China) was adopted, which consisted of a transmission grating (ImSpector-V10, Specim, Finland), an electronically controlled scanning platform (Zolix-SC300, Zhuoli Hanguang Instrument, Beijing, China), and a scientific grade 16-bit CCD grayscale camera (Lt365R, Lumenera, Canada). The light source adopted diffuse halogen light source (HSIA-LS-TH, Shuanglihepu, Sichuan, China), of which the spectral range was from 350 to 2500 nm.

The constructed imaging system and the acquired hyperspectral images were shown in Figure 1. The entire sampling device was in a dark box, and the center of the two halogen lamps was kept 50 cm from the center of the object. The angle between the halogen lamp and the horizontal plane was 45°, and the difference in optical power distribution was below 10%, so that the shooting area was evenly illuminated. The acquisition parameters were set as spectral resolution 991×960, spectral range 400-1000 nm, spectral resolution 3.5 nm at 700 nm, exposure time 6.6 ms. Each spectral cube was for one *D. nobile* sample, which has 176 bands. So, there were 450 spectral cubes totally. The hyperspectral images of each *D. nobile.* was recorded as a 3D data block, which included image information in the *x* and *y* axes and spectral information in the *z* axis.

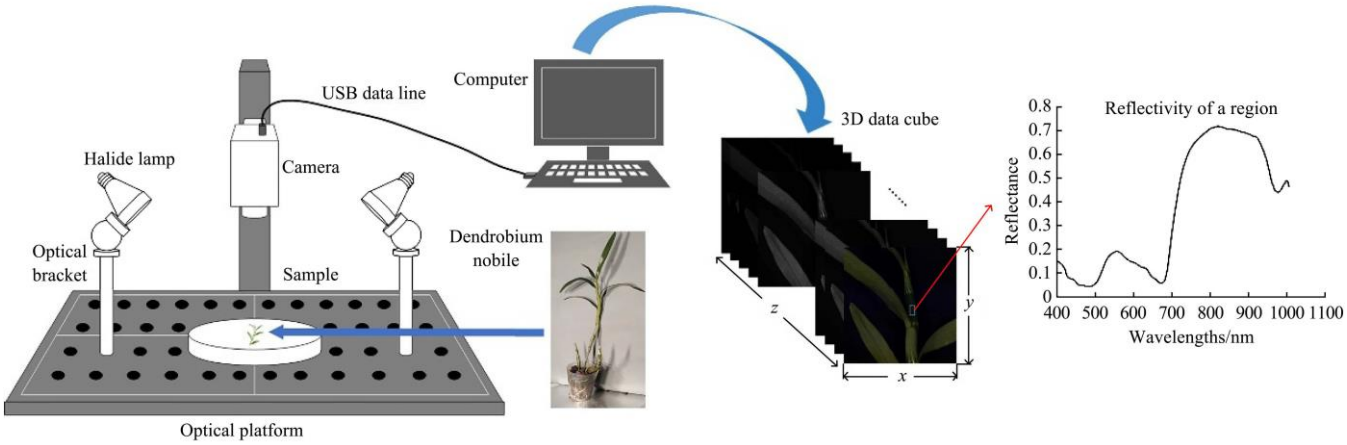


Figure 1 Hyperspectral images acquisition

2.3 Spectral image preprocessing

In order to eliminate the dark noise of the CCD and the influence of the light source, the reflectance of the spectral image was corrected. We obtained dark noise by covering the camera lens, and collected standard whiteboard reflection spectrum images. Finally, correcting the dots of the spectral image to obtain the plant reflectance spectrum, by using the following equations:

$$I(i,j)=\frac{R_{raw}(i,j)-R_{black}(i,j)}{R_{white}(i,j)-R_{black}(i,j)} \quad (1)$$

where, $I(i,j)$ was the (i,j) pixel reflectance in the image at a certain wavelength after correction; $R_{raw}(i,j)$ was the original gray value corresponding to (i,j), and R_{black} was the gray value of the dark noise of (i,j); $R_{white}(i,j)$ was the gray value of the standard whiteboard corresponding to (i,j).

2.4 ROI area selection

Compared with leaves and stems covered by leave sheath of *D. nobile*, bare stems are more sensitive to Nitrogen supply level. So bare stem region in images were ROI we wanted. As shown in Figure 2, a series of grey masks were invited for ROI selection here. Firstly, the 450 nm, 550 nm, and 650 nm images in the hyperspectral image cube were selected to compose a pseudo-color image *a*. Secondly, the green plant and the background were

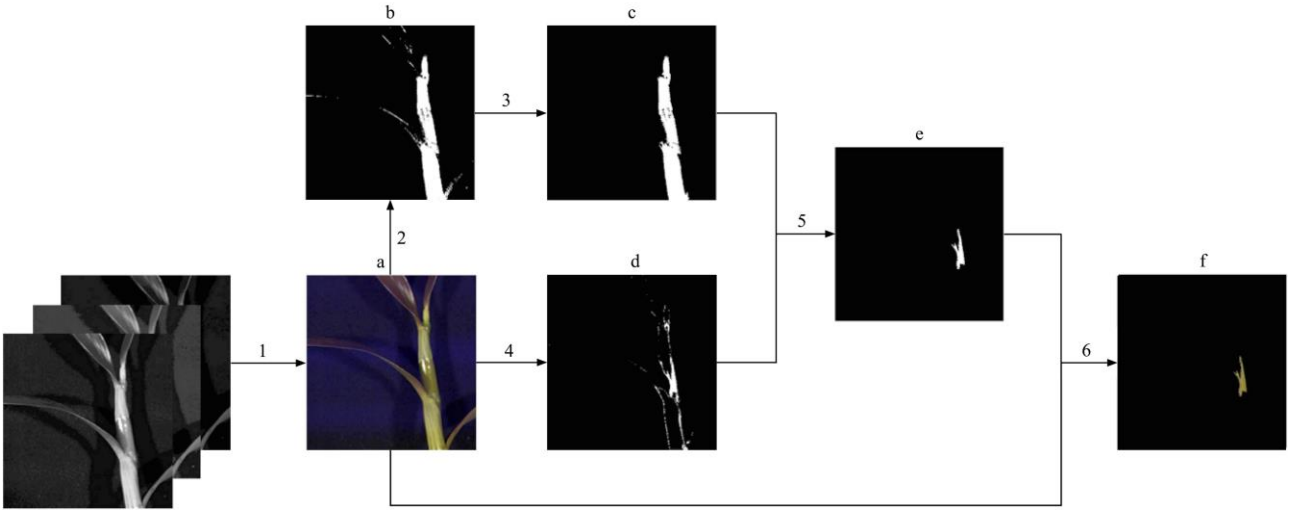
segmented by the 2G-R-B method to obtain the mask b of the green plant, which may include unremoved noise.

The divided pixel can be mathematically formulated as below:

$$P(i,j)=2G(i,j)-b(i,j)-R(i,j) \quad (2)$$

where, $P(i,j)$ was the gray value after processing; G, B, and R correspond to 550 nm, 450 nm, and 650 nm images, respectively.

The third step was to remove the noise by opening and closing operation to obtain the mask c of green plants. The fourth step, after completing a segmentation, separated the stem with high brightness and the bare stem with low brightness by RGB threshold method. The gray value with the most occurrences was used as the threshold to extract part of the ROI area here to guarantee the pureness of ROI. The threshold value of three channels was set from [25, 65, 70] to [40, 90, 90], respectively. If the R, G, B gray value of a pixel in the picture was not within these three thresholds, its R, G, B gray value was set to [0, 0, 0], and the Mask d was got, which included ROI area and some of unremoved noise. The fifth step was to perform an or-operation on masks c and d to remove the background and leaf sheath at the same time. Then various areas were sorted from large to small, since the largest area was bare stem which was the ROI area. In the sixth step, the mask e was added to the pseudo-color image to obtain the image f of the ROI.



1. Synthetic pseudo-color image 2. 2G-R-B segmentation 3. Open-close operation 4. RGB threshold segmentation 5. Mask or-operation 6. ROI region intercept

a. Pseudo-color image b. Mask for green plants c. Mask after noise removal d. Mask for removing leaf sheath e. ROI area mask f. ROI area

Figure 2 ROI area selection

2.5 Analysis Software

The environment in which the program runs in this article was Python 3.7.6, and the programming software was the jupyter notebook under anaconda3. The python third-party toolkit sklearn was used in this paper.

3 Spectral data analysis

In order to reduce the data noise caused by possible specular reflection in the ROI area and sample contamination in small areas, the ROI area was divided into three parts horizontally, which means 3 groups of data there were for each *D. nobile* plant. Since there were 30 samples of each growth stage combined with each cultivation method, a total of 1350 spectra were got for analysis.

3.1 Outliers elimination

Outliers have a great influence on the data modeling and often lead to incorrect results. The standard deviation between each data and the average value of 30 samples is usually used as the prediction error distribution, and those with a larger standard deviation from the average are considered as potential outliers. To remove outliers, we put the plant data of the same growth period and the same cultivation method into a group, took the average value of the spectral reflectance of this group I_{mean}, and then, removed the spectral data with a large standard deviation from the average value. Calculation formula (3) and (4) are used for removing outliers:

$$I_{mean}=\sum_{p=1}^{30}\sum_{q=1}^{3}I_{p,q} \quad (3)$$

$$E_{p,q}=\sum_{w=1}^{176}\sqrt{I_{p,q}^{2}(w)-I_{mean}^{2}(w)} \quad (4)$$

Among them, the formula (3) is the average reflectance of each part of the 3 parts of *D. nobile*. $I_{p,q}$ is the reflectance data of the q-th part of the p-th *D. nobile*, formula (4) is the standard deviation between the reflectance of the q-th part of the p-th *D. nobile* and the average reflectance of the group, w is the number of spectral data bands, and $I_{p,q}(w)$ is the *p-th* reflectance data for the w-th band of the q-th part of *D. nobile*.

By eliminating outliers, there were 5×70 groups of data for seedling stage *S*, 5×75 for growing stage *M*, and 5×80 data for mature stage *L*, respectively. 2/3 of them were randomly selected as the training set and the remaining 1/3 as the test set.

3.2 Spectral feature extraction

3.2.1 PCA dimensionality Reduction

PCA dimensionality reduction projects all dimensions into a low-dimensional space to maximize the variance D_1 of each data into this space, as shown in formulas (5) and (6):

$$Z(m*k)=X(m*n)\times W(n*k) \quad (5)$$

$$D_1=\sum_{m=1}^{450}\omega^{T}x_m \quad (6)$$

where, m is the number of samples, n is the number of bands; k is the dimension after dimensionality reduction; $X(m*n)$ is the training sample set, and $W(n*k)$ is the eigenvector of the first k covariance matrices selected after $X(m*n)$ is sorted by eigenvalue, $Z(m*k)$ is the projection of $X(m*n)$ on the new orthogonal subspace formed by $W(n*k)$, and also new matrix after dimensionality reduction; ω^T is the transpose of $W(n*k)$, and $\omega^T x_m$ is the value of each spectrum projected onto the low-dimensional space.

The dimension of PCA dimensionality reduction is generally determined by the contribution rate. We balanced both the contribution rate and the change rate of the cumulative contribution

rate to get the suitable dimension here. We set the contribution rate to be larger than 99%, and the change rate of the cumulative contribution rate to be less than 0.01% as the threshold to determine the principal component training model, which is shown in formula (7) and (8):

$$r_n = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{176} \lambda_i} \geq 99\% \quad (7)$$

$$\frac{dr_n}{dn} \leq 0.01\% \quad (8)$$

Formula (7) represented the proportion of the eigenvalue of the *n*-th principal component to the total eigenvalue. Formula (8) represented the rate of change of the cumulative contribution rate *r* of the *n*-th principal component.

3.2.2 LDA dimension reduction

LDA dimensionality reduction maximizes the variance of each data to low-dimensional space while projecting high-dimensional data to low-dimensional data. The main difference from PCA is that while the variance between classes is maximized, the variance within the classes is minimized. Formula (9) and (10) were used to represent the inter-class distance D_2 and the intra-group distance D_3, respectively.

$$D_2 = (\omega^T \mu_0 - \omega^T \mu_1)^2 \quad (9)$$

$$D_3 = \sum_{x \in X_0} (\omega^T x - \omega^T \mu_0)^2 + \sum_{x \in X_1} (\omega^T x - \omega^T \mu_1)^2 \quad (10)$$

Among them, the matrix was divided into X_0 and X_1 according to class labels 0 and 1, and their corresponding mean vectors were μ_0 and μ_1. $\omega^T\mu_0$ and$\omega^T\mu_1$ were the mean vectors of the points of class label 0 and 1, and these vectors were projected to the low value in space, where the distance between them was the class distance after being projected to the low-dimensional space. $\sum_{x \in X_0} (\omega^T x - \omega^T \mu_0)^2$ and $\sum_{x \in X_1} (\omega^T x - \omega^T \mu_1)^2$ were the variance of the points of the class labels 0 and 1 projected onto the low-dimensional space. The contribution rate and the rate of change of the cumulative contribution rate were also used to determine the dimension here.

3.2.3 GA genetic algorithm feature selection

The genetic algorithm reduces the dimension by optimizing the wave band. 20 bands were randomly selected from 176 bands in this paper, and the data of 20 bands of 100 groups were taken out separately. The distance between classes and the distance between groups were calculated from largest to smallest. The crossover rate was set to 0.5-0.9, where the first group and the second group were selected to exchange the band with a small contribution rate in proportion to the crossover rate. This process was repeated. At the same time, the mutation rate was set to 0.01-0.1, where one of the 176 bands were selected randomly to replace the random one of the original set of bands. This process was repeated. The number of iterations was determined by the mean square error. The last 20 bands selected had the largest distance between classes and the least distance within a group.

3.2.4 Decision tree feature selection

The decision tree uses entropy and information gain to determine the amount of information contained in each attribute. The larger the information gain, the more information in this attribute. Formula (11) and formula (12) showed the calculation method of entropy and information gain.

$$Ent(D) = -\sum_{i=1}^{176} p_i \log_2 p_i \quad (11)$$

$$Gain(D, x_i) = Ent(D) - \sum_{v=1}^{V} \frac{|D_v|}{D} Ent(D_v) \quad (12)$$

where, *D* was the sample set of two cultivation methods. For each band x_i in the band $X=\{x_1, x_2, \ldots, x_{176}\}$, its probability was described as $P(X = x_i) = p_i$ (i = 1,2,…,176). In formula (12), the data of band x_i had a number of values, which were divided into V = 10 equal parts from large to small, to obtain data sets D_1 to D_{10}. Each data set contained a different number of data. After that, we calculated the proportion of each set in the entire data set $\frac{|D_v|}{D}$, respectively, to find the information entropy and weighted average of these 10 data sets. The difference from the previous entropy was the information gain.

3.3 Establishment of classification model based on SVM

The SVM classification model was invited to classify the data after dimensionality reduction in the paper. In this research, the "one-to-one" method was used to build 10 support vector machines by combining five classes of training data in pairs. Each support vector machine trained two different classes of data. After that once the test set was imported into the model, its category was determined by voting.

Considering the kernel function of SVM, Gaussian kernel function was selected since it could map the sample space to higher dimensions for classification of complicated data, which solves the overfitting phenomenon caused by linear kernel.

The penalty coefficient was gradually increased from 0.001-100 in 10 times steps. With the accuracy of the training set and test set as the standard, the span of the penalty coefficient with the best effect was selected, and then the selected spans were divided into 10 equal parts. The penalty coefficient was determined by repeating the above process. At last, ten-fold cross-validation was used to train each dimension 10 times, and the average of training results was used as the prediction model.

4 Results

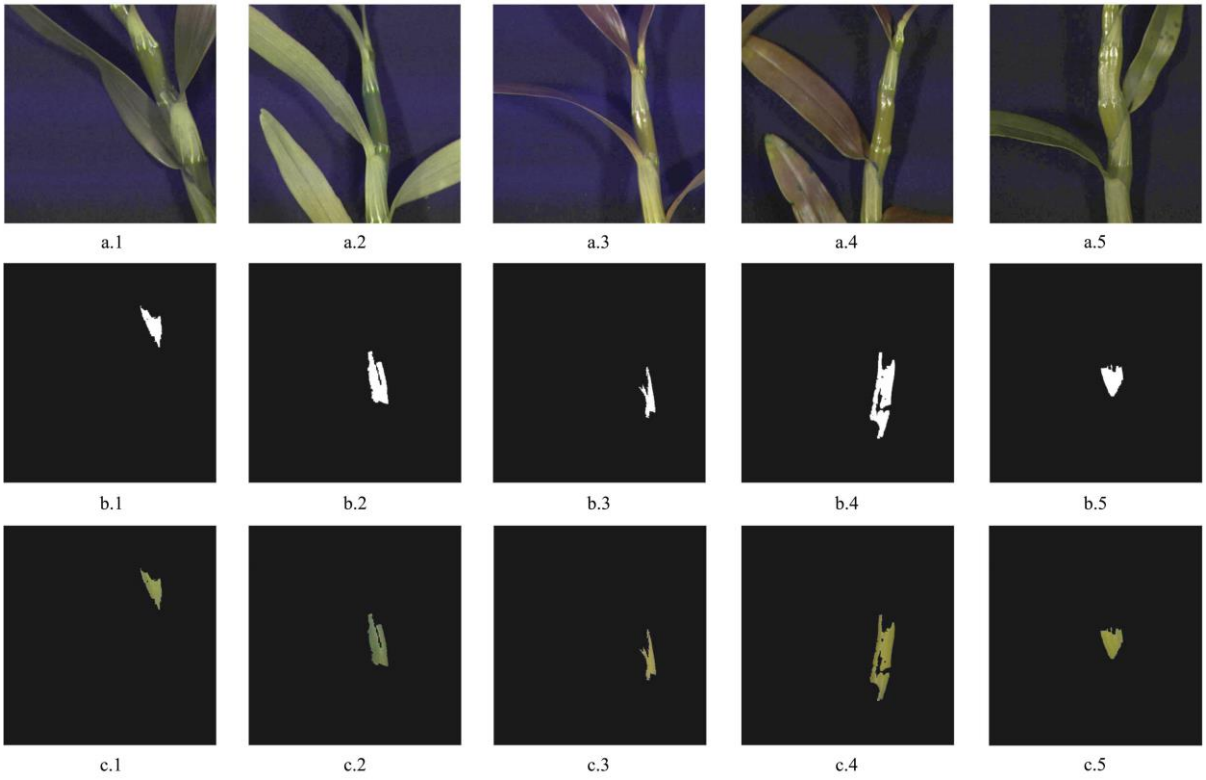
4.1 ROI area selection results

Spectral images of 460 nm, 560 nm, and 660 nm in the extracted spectral images composed the pseudo-color images. The images of *Dendrobium* in mature stage (*L*) as a sample were shown in Figure 3. Figures a.1-a.5 were five different cultivation methods L_1–L_5, *a* was a pseudo-color image, *b* was a mask obtained by secondary segmentation, and *c* was the final ROI area.

It could be seen from the Figure 3 that the method can effectively remove the background of stems and leaves wrapped by leaf sheath.

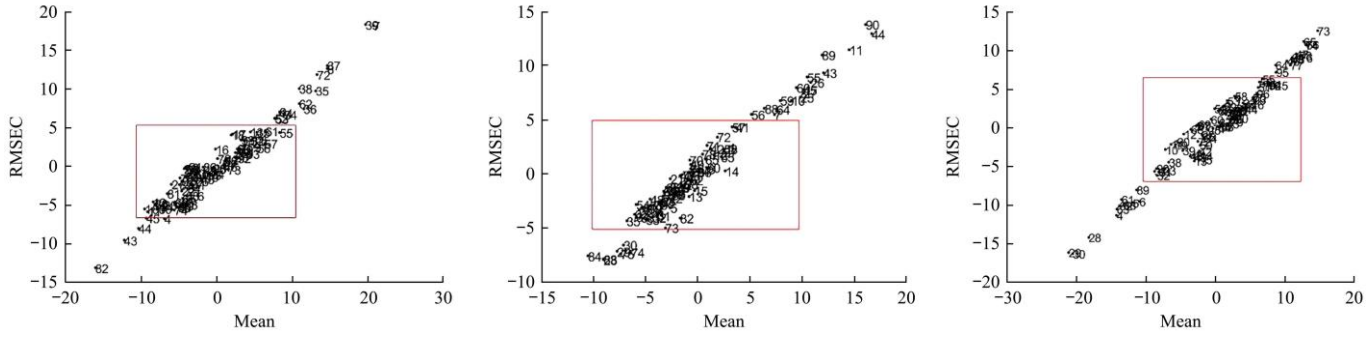
4.2 Outliers elimination results

The difference and standard deviation distribution of the full-band spectrum of T_5 cultivation mode was shown in Figure 4. The horizontal and vertical coordinates were difference and standard deviation of the full-band spectrum, respectively. The points whose absolute value of the standard deviation from the center was larger than 5 and absolute value of the average difference larger than 10 were considered to be outliers. The reflectance spectra of L_5 without removing outliers were shown in Figure 5a, and the reflectance spectra of L_5 after removing outliers were shown in Figure 5b. The average value of the removed and unremoved outliers and the dispersion of the reflectance of each band were shown in Figure 5c. Comparing the data dispersion before and after eliminating outliers, the band fluctuations were significantly reduced in the full band range. According to formula (9), the difference between the curve *x* and the mean μ was reduced, so the intra-class gap was smaller, which was helpful to improve the classification accuracy.



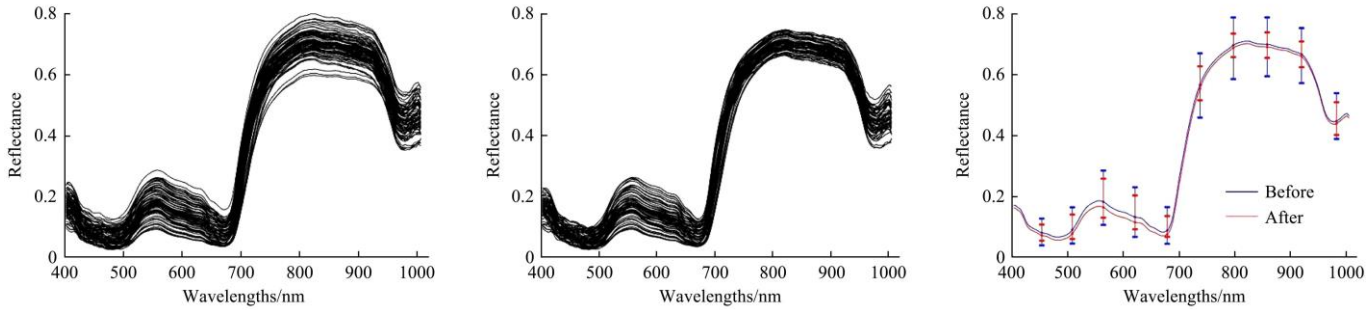
a.1-a.5: L_1–L_5 pseudo color image; b.1-b.5: mask after secondary segmentation; c.1-c.5 ROI area.

Figure 3 L_1–L_5 ROI region selection results



a. S_5 b. L_5 c. M_5

Figure 4 Means/Mean Squares of Remove Outliers



a. L_5 reflectance spectrum without removing outliers b. L_5 reflectance spectrum after outliers have been removed c. L_5 With / without removing the average value of outliers and the dispersion of reflectivity of each band

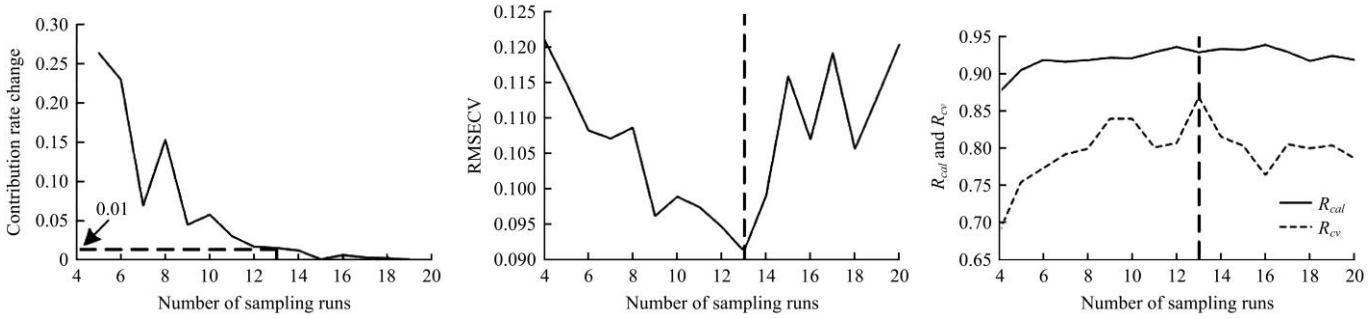
Figure 5 Spectral reflectance data of unremoved/ removed outliers

4.3 Data analysis results

4.3.1 PCA-SVM classification results

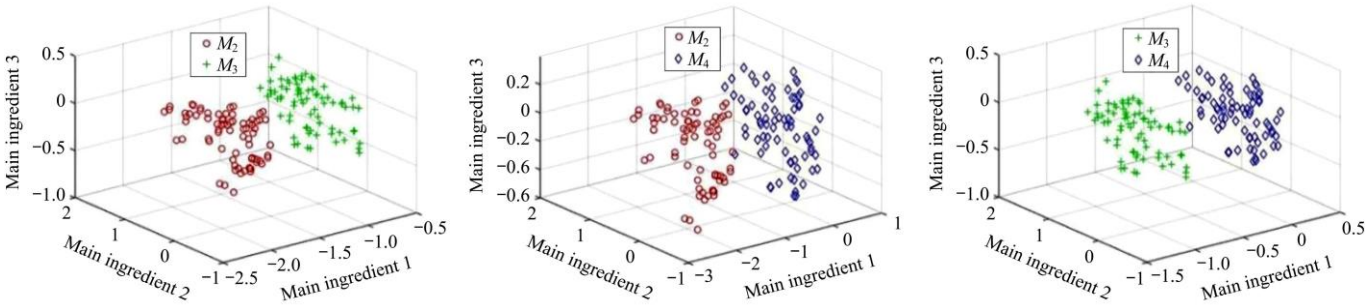
We used PCA combined with SVM algorithm to train the classification network on the reflectance data in this paper. Firstly, we should determine the optimal number of principal components. Figure 6.a is the change rate of the main component contribution rate when the growth stage is *M*. The change rate was larger before 4 dimensions. With the dimension increasing, the contribution rate increased, and the change rate tended to gradually decrease. The contribution rate was larger than 99% and the change rate was less than 0.01% at 13 dimensions. The lowest test set root mean square error RMSECV, the highest training set accuracy rate R_{cal}, and the highest test set accuracy rate R_{cv} were obtained together when first 13 dimensions were selected (shown in Figure 6). The same pattern existed in the other two growth stage samples. The data distribution of the first three principal

components between the two cultivation methods of M_2, M_3 and M_4 were shown in Figure 7. It showed that most of the data could be successfully separated even if only the first three principal components were used.



a. RMSECV b. R_{cal} c. R_{cv}

Figure 6 Principal component contribution rate change rate selection band



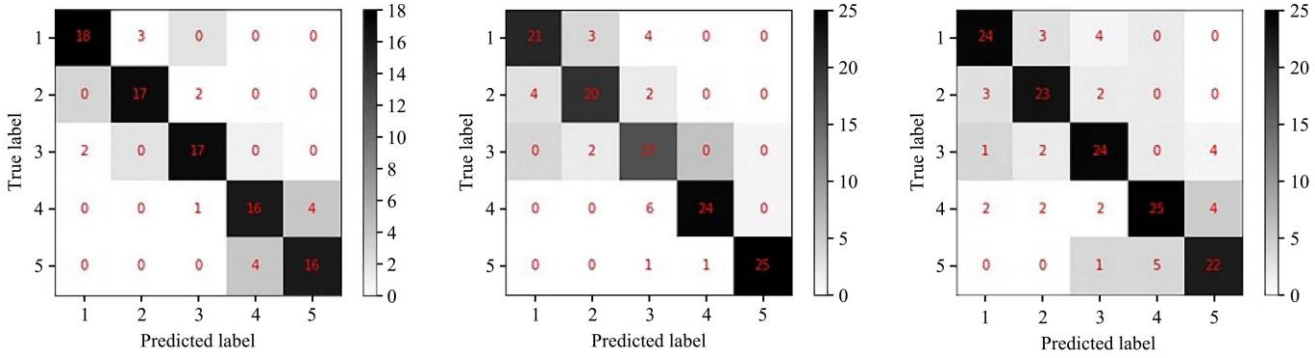
a. M_2, M_3 classification b. M_2, M_4 classification c. M_3, M_4 classification

Figure 7 M_2, M_3, M_4 one versus one classification

The classification results of the training set and the test set gotten by PCA-SVM model were shown in Table 1. The test set confusion matrix of the three growth stages were shown in Figures 8a to 8c. The higher the accuracy, the deeper the data color. It showed that growth stage of M has highest classification accuracy when PCA-SVM model was used.

Table 1 PCA-SVM Calibration set and test set results

Types	Livespan	LVs	R_{cal}	RMSEC	R_{cv}	RMSECV
PCA	*S*	11	90.17%	0.03242	84.00%	0.0348
	M	13	92.67%	0.09227	87.73%	0.09283
	L	10	87.83%	0.03918	78.67%	0.04592



a. Seedling stage *S* b. Growth stage *M* c. Mature stage *L*

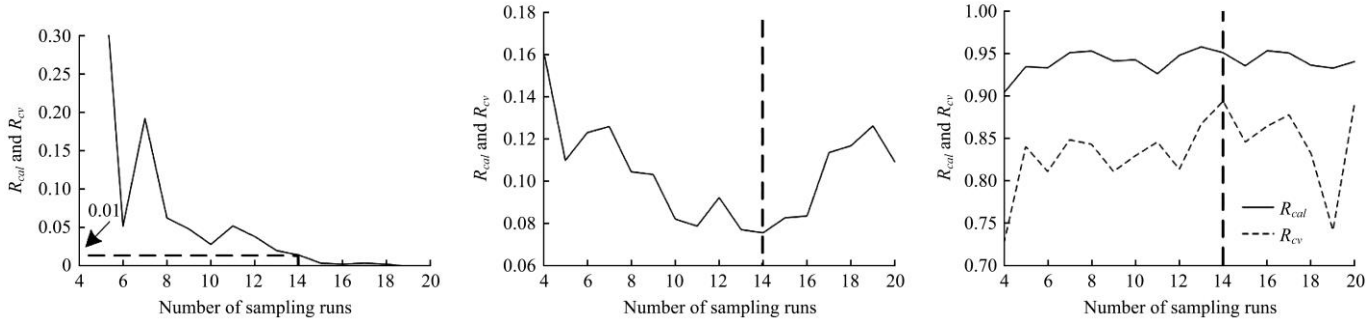
Figure 8 Confusion matrix of the three growth stages test sets of *S*, *M* and *L*

4.3.2 LDA-SVM Classification results

We used LDA combined with SVM algorithm to train a classification network on reflectance data. Figure 9a is the rate of change of the principal component contribution rate during the growth stage (*M*). The rate of change was higher before 4 dimensions, the contribution rate was larger than 99% and the change rate was less than 0.01% at 14 dimensions. The lowest test set RMSECV, the highest R_{cal}, and the highest R_{cv} were obtained together when first 14 dimensions were selected (shown in Figure 9). The same pattern existed in the other two growth stage samples. The data distribution of the first three principal components between the three cultivation methods of M_2, M_3 and M_4 were shown in Figure 10. It showed that most of the data could be successfully separated even if only the first three principal components were used. Compared with PCA, the data was more concentrated. It showed that LDA required more dimensions than PCA, but the classification results were more accurate.

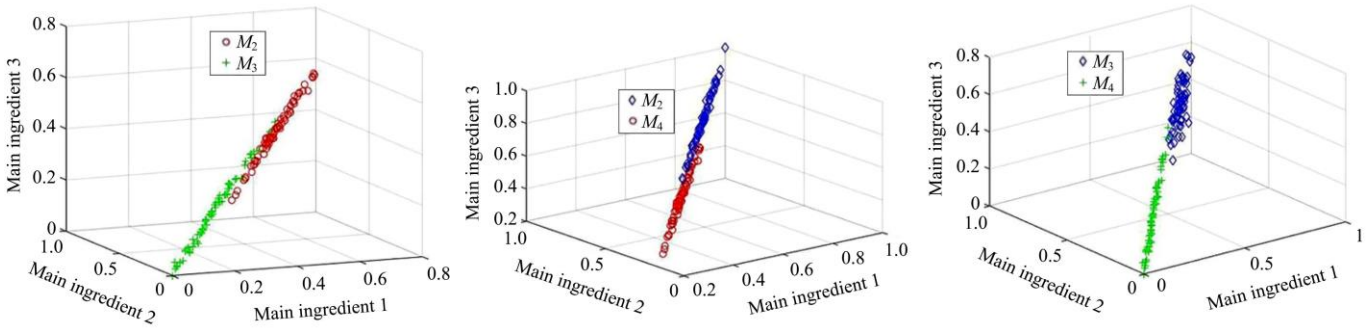
The classification results of the training set and the test set gotten by LDA-SVM model were shown in Table 2. The test set confusion matrix of the three growth stages were shown in Figures 11a to 11c. It showed that LDA can establish a model with good performance for all three growth stages, and meet the previous conjecture. The control group T_5 and the other four groups were misjudged as 0. Most of the misjudgments occurred between T_2, T_3, and T_4, which were probably caused by similar nitrogen cultivation methods. The accuracy rate of the growth period (*M*) was low, and the RMSECV was larger than that of growth stage (*M*)

and mature stage (*L*), which was the same law in four dimensionality reduction methods in this paper.



a. RMSECV b. R_{cal} c. R_{cv}

Figure 9 Principal component contribution rate change rate selection band

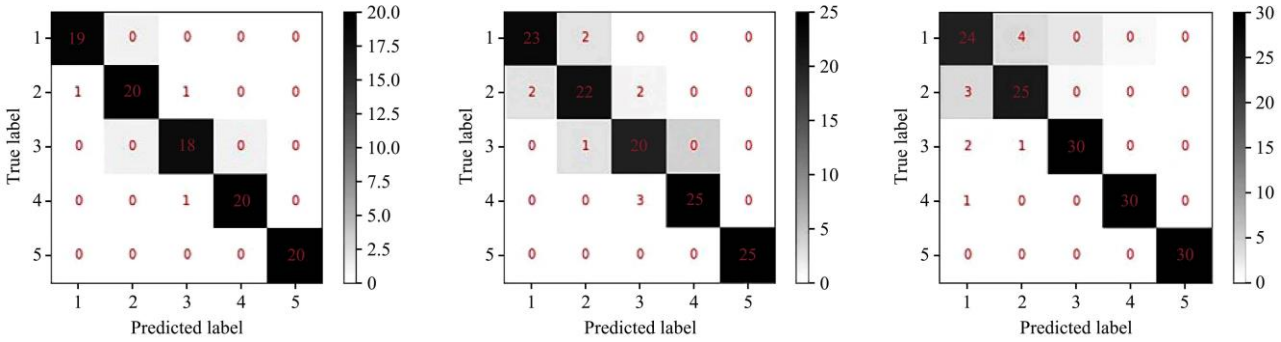


a. M_2, M_3 classification b. M_2, M_4 classification c. M_3, M_4 classification

Figure 10 M_2, M_3, M_4 one versus one classification

Table 2 LDA-SVM Calibration set and test set results

Types	Livespan	LVs	R_{cal}	RMSEC	R_{cv}	RMSECV
LDA	*S*	15	97.47%	0.02375	97.00%	0.02522
	M	14	95.03%	0.07735	88.80%	0.08376
	L	11	95.97%	0.03156	92.67%	0.03572



a. Seedling stage *S* b. Growth stage *M* c. Mature stage *L*

Figure 11 Confusion matrix of the three growth stage test sets of *S*, *M* and *L*

4.3.3 GA -SVM classification result

The population number, the crossover rate, the mutation rate, and the number of iterations were set to 200, 0.7, 0.02 and 200 respectively. Since the contribution rate cannot be observed, the model was trained from the first 4 characteristic bands. When the number of wavelengths selected by the GA genetic algorithm was equal to 7 or 8, the training set and the test set had the highest accuracy. The selected wavelengths (blue solid circles) corresponding to the original spectrum were shown in Figure 12, with the variance ratio as a contribution for the evaluation criteria of the rate. The larger the characteristic value, the wider the bar. The selected band was used as the input of the SVM to establish a model for separating 5 different nutrient cultivation methods. The prediction results were listed in Table 3. Figures 13a to 13c were the confusion matrix of the test sets of three growth stages, respectively.

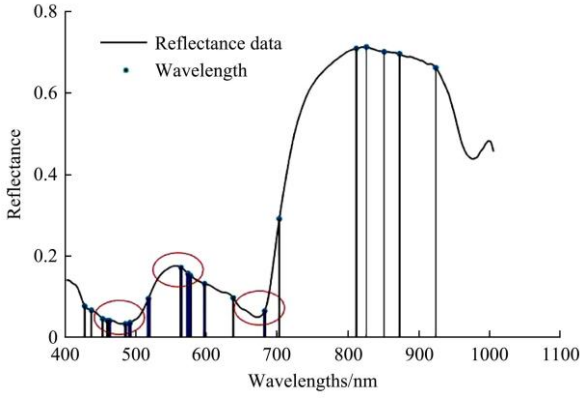
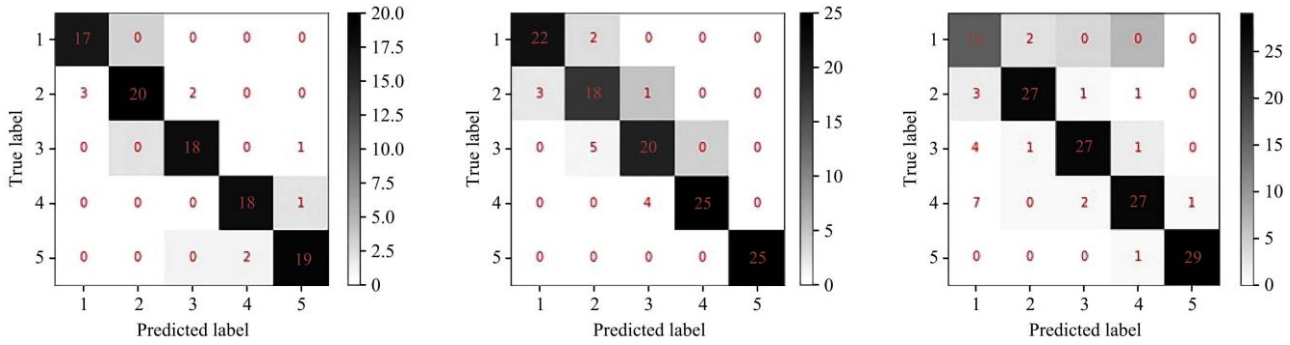


Figure 12 Feature band selection



a. Seedling stage *S* b. Growth stage *M* c. Mature stage *L*

Figure 13 Confusion matrix of the three growth stage test sets of *S*, *M* and *L*

Table 3 GA-SVM Calibration set and test set results

Types	Livespan	LVs	R_{cal}	RMSEC	R_{cv}	RMSECV
Genetic algorithm	*S*	8	92.80%	0.04709	92.00%	0.04862
	M	7	93.63%	0.08417	88.00%	0.08793
	L	7	92.20%	0.05512	84.00%	0.06081

It showed that band 492 nm had the largest contribution rate in Figure 12. The best results can be obtained when the bands 492 nm, 547.5 nm, 495.2 nm, 692.5 nm, 709.8 nm, 475.8 nm, 696 nm (red circles) were used for classification. The most misjudgments occurred between T_3 and T_4, and the least misjudged were between T_1 and T_5, similar with other models. So classification error rate of adjacent gradients (1/2N and 1/3N) was the highest.

4.3.4 Decision tree DT-SVM classification result

Information gain was used to select features. By slowly increasing the maximum depth of the decision tree, the difference of accuracy rate between the correction set and the test set within 10% was found, and the maximum depth was determined to be 3. Since each plant took three points, we cut off the branches with less than three data, which meant the minimum number of samples contained in the leaf node was equal to 3. The cutting step above reduced the misclassification of different types of points on the same plant. Since there must be at least two side branches on each node, the minimum sample node was set equal to 2.

After running the decision tree for feature selection, some wavelengths with small information gain were removed, while the key wavelengths were retained. The first 20 selected wavelengths (blue solid circles) corresponding to the original spectrum were shown in Figure 14. It showed that band 574nm had the largest contribution rate. The best results can be obtained when bands 574 nm, 547.5 nm, 564.1 nm, 678.8 nm, 577.3 nm, 661.7 nm, 492 nm, 485.5 nm, 665.1 nm, 668.5 nm, 894.6 nm, 495.2 nm, and 475.8 nm (red circles) were selected for classification. The SVM model was established with 4-20 bands after feature selection. The test set confusion matrix for *S*, *M*, *L* were in Figures 15a to 15c. It showed that T_5 had the lowest misjudgment from other cultivation methods, and T_3 and T_4 were produced more misjudgments in all three growth stages. The training results of this model were not as good as GA-SVM. The dimensions selected and the classification results for each growth stage were listed in Table 4.

The results of both feature band selection methods showed that the characteristic bands with biggest information gain were at the valley of 480 nm, 680 nm and the peak of 550 nm, which also were the characteristic bands of chlorophyll a and chlorophyll b. This was consistent with the correlation between nitrogen content and chlorophyll concentration, which proved the effectiveness of feature band selection methods.

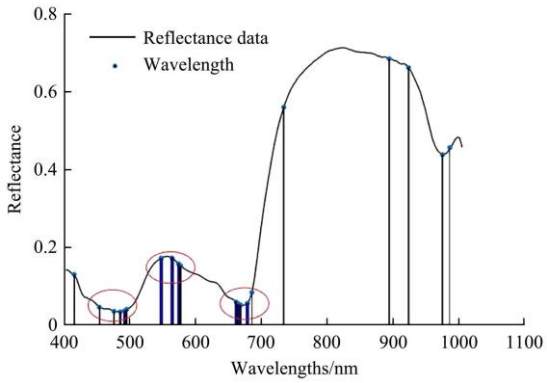
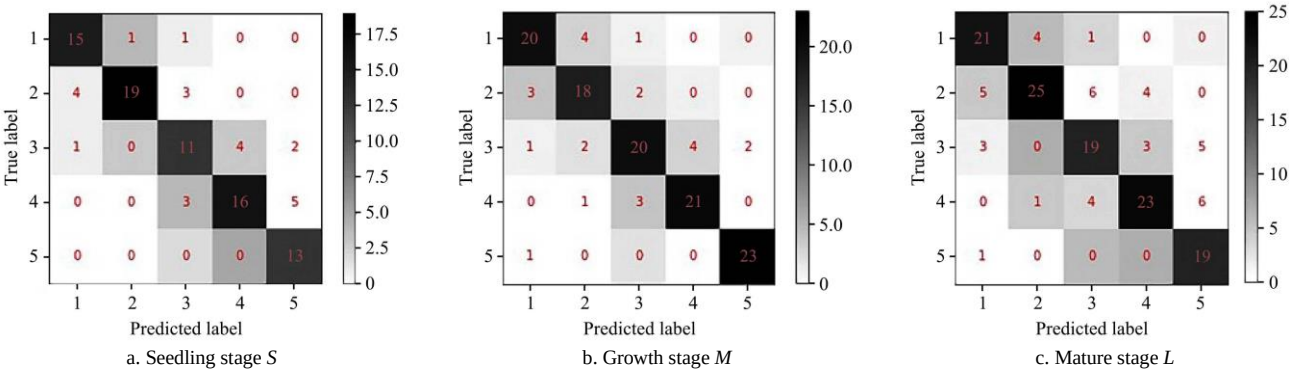


Figure 14 Characteristic bands

Table 4 DT-SVM Calibration set and test set results

Types	Livespan	LVs	R_{cal}	RMSEC	R_{cv}	RMSECV
Decision tree	*S*	11	82.11%	0.03470	76.00%	0.03745
	M	13	87.83%	0.08986	81.60%	0.09487
	L	12	77.54%	0.05426	71.33%	0.05677



a. Seedling stage *S* b. Growth stage *M* c. Mature stage *L*

Figure 15 Confusion matrix of the three growth stage test sets of *S*, *M* and *L*

5 Discussion

Aiming at the demand of *D. nobile* nutritional supply tracking and detection, the hyperspectral imaging method was proposed and the classification model was established to obtain the relationship between the hyperspectral image data and the gradient nutrition supply method of the plant.

An effective ROI selection method for *Dendrobium* plants was proposed. The mask of green plants and the mask of dark bare stems were provided to perform an or-operation for removing the background, the leaves and the leaf sheath, respectively, and then the ROI area were obtained.

The methods of reducing the dimensions, PCA and LDA, as well as the method of optimizing bands, GA genetic and the decision tree algorithm, were compared in the paper. These four spectral feature extraction methods combined with SVM multi-classification method were used for modeling. The classification results showed that the correct rate of the LDA -SVM model was higher than the other three methods for both the training set and the test set. The accuracy rates of the *S*, *M*, and *L* training sets reached up to 97.47%, 95.03%, and 95.97%, respectively. The accuracy rate of the test set reached 97.00%, 88.80%, 92.67%, which can meet the needs of *D.nobile* detection.

The bands selected by the method of optimizing bands are especially valuable for high-throughput detection in the field, such as multi-spectral remote sensing. The classification accuracy of GA-SVM model was higher than that of DT-SVM in the paper. In the GA genetic algorithm, the bands 492 nm and 547.5 nm were the preferred bands in the three growth stages, with the smallest root mean square error RMSEC, the highest correction set accuracy rate R_{cal}, and the highest test set accuracy rate R_{cv}. The GA-SVM model using the selected bands 475.8 nm, 495.2 nm, 492 nm, 547.5 nm, 692.5 nm, 696 nm, 709.8 nm could provide the accuracy rates more than 80%.

In brief, the classification of *D. nobile* cultivated by different nitrogen supply could be better achieved by hyper-spectral images acquisition, ROI extraction, spectral dimensionality reduction and SVM classification modeling. The results could be used directly for nitrogen detection of *D. nobile*, and the method in the paper provided a new idea for intelligent monitoring for precise nutrition supply of plant.

Acknowledgments

We deeply thank for the Key-Area Research and Development Program of Guangdong Province (No. 2019B020214005), Guangdong Provincial Rural strategic revitalization project in 2019 (YCN (2019) No. 73).

[References]

[1] Zhou W, Xia J, Sun W B, et al. Research status of the chemical constituents and pharmacological effects of Dendrobium nobile. Chinese Journal of New Drugs, 2017; 26(22): 2693–2700. (in Chinese)

[2] Xie M M, Xiao Liu, Yang Lei, et al. Isolation and purification of Dendrobium nobile polysaccharide and its anti-aging activity. Modern Chinese Medicine, 2018; 20(12): 1489–1493. doi: 10.13313/j.issn.1673-4890.20180622003. (in Chinese)

[3] Chen X Q, Ji Z H. China Orchid Complete Book. Beijing: Chinese Forestry Fruits and Vegetables, 2013; 11: 52–53. Publishing House, 1998. (in Chinese)

[4] Jiang Y, Huang H J. Research progress of Dendrobium nobile. Yunnan Journal of Traditional Chinese Medicine, 2015; 36(01): 77–78. doi: 10.16254/j.cnki.53-1120/r.2015.01.038. (in Chinese)

[5] Chen B L, Gong J Y, Wang H X, et al. Effects of different fertilization methods on the growth of Dendrobium officinale plants under forest. Heilongjiang Agricultural Sciences, 2019; (02): 98–103. doi: 10.11942/j.issn.1002-2767.2019.02.0098. (in Chinese)

[6] Hu X J, Zeng Y Y, Ao F X, et al. Effects of different cultivation media on growth and physiology of two dendrobium species. Journal of Southwest Normal University (Natural Science Edition), 2019; 44(11): 29-35. doi: 10.13718/j.cnki.xsxb.2019.11.005. (in Chinese)

[7] Cai L, Xu Y M, Zhao Z, et al. Study on the quality change of Dendrobium nobile Lindl. at different harvest time. Journal of Mountain Agriculture and Biology, 2019; 38(01): 19–23. doi: 10.15958 /j.cnki.sdnyswxb.2019.01.004. (in Chinese)

[8] Wang X M, Cheng H T, Kang Z M, et al. Effects of several fertilizers on the growth characteristics of Dendrobium candidum. Journal of Anhui Agricultural Sciences, 2017; 23(07): 122–124. doi: 10.16377/ j.cnki.issn1007-7731.2017.07.055. (in Chinese)

[9] Zhang Z P. Effects of different cultivation methods on the growth and effective components of Dendrobium candidum. Nanchang University, 2016. (in Chinese)

[10] Guo J X, Li Q M, Wu P, et al. Effects of tiller number and cultivation substrate on stem and leaf growth of Dendrobium nobile. Anhui Agricultural Sciences, 2016; 44(36): 156–158, 258. doi: 10.13989/j.cnki.0517-6611.2016.36.052. (in Chinese)

[11] Shi J, Yang H T, Xu C, et al. Effects of nitrogen nutrition on the growth and physiology of Dendrobium candidum. Jiangsu Agricultural Sciences, 2018; 46(23): 119–122. doi: 10.15889 /j.issn.1002-1302.2018.23.029. (in Chinese)

[12] Wang Z H. Effects of organic fertilization and chemical fertilization on the growth and quality of three dendrobium trees. Chinese Horticultural Society Ornamental Horticulture Professional Committee, National Flower Engineering Technology Research Center. Chinese Ornamental Horticulture Research Progress 2016. Chinese Horticultural Society Ornamental Horticulture Professional Committee, National Flower Engineering Technology Research Center: Chinese Horticultural Society, 2016; 407–411. (in Chinese)

[13] Li Y, Wang M J, Ji J. Effects of different nitrogen levels on the growth and total flavonoids production of Erigeron breviscapus under UV-B radiation. Yunnan Agricultural Science and Technology, 2012 Suppl; 182–188. (in Chinese)

[14] Kong D D, Shen H L, Zhong Y X, et al. Effects of different fertilization measures on the growth and quality of Dendrobium candidum. Agricultural Science and Technology Communication, 2015; (08): 108–111. (in Chinese)

[15] Xue M, Yang J Y, Dan S Y, et al. Effects of combined application of nitrogen, phosphorus and potassium on seedling growth of Dendrobium nobile. Guizhou Agricultural Sciences, 2014; 42(11): 166–169, 175. (in Chinese)

[16] Li S K. Effects of fertilization on tissue culture seedlings of Dendrobium candidum. Agricultural Technology Services, 2013; 30(05): 492+494. (in Chinese)

[17] Zhou H G. Preliminary report on the observation of stem growth of Dendrobium nobile. China Tropical Agriculture, 2012; (04): 37–39. (in Chinese)

[18] Tang L, Zhang L X, Wang Y Q, et al. Study on the quality standard of Dendrobium nobile seedlings grading. Traditional Chinese Medicine, 2012; 35(01): 12–15. doi: 10.13863/j.issn.1001-4454.2012.01.005. (in Chinese)

[19] Wang Z H, Zhu G F, Cao J X, et al. Effects of N, P, and K fertilization on vegetative growth and flowering of Dendrobium candidum. Chinese Agricultural Science Bulletin, 2011; 27(16): 248–254. (in Chinese)

[20] Wang Z H, Zhu G F, Cao J X, et al. Effects of different fertilization treatments on the growth characteristics and mineral content of Dendrobium candidum. Guangdong Agricultural Sciences, 2011; 38(05): 83–86. doi: 10.16768/j.issn.1004-874x.2011.05.056. (in Chinese)

[21] Li Y. Identification of different cultivation types of Dendrobium nobile Lindl. by infrared spectroscopy. China Commodity Society. Proceedings of the Second National Conference on Chinese Medicine Commodities. China Commodity Society: China Commodity Society, 2010; 191–197. (in Chinese)

[22] Li S H. Discussion on issues related to the determination of ammonia nitrogen in water. Northern Environmental, 2012; 24 (04): 176–177.

(in Chinese)

[23] Liu R T. Research and application of rapid identification of Dendrobium species based on near infrared spectroscopy. Zhejiang Agriculture and Forestry University, 2015. (in Chinese)

[24] Zhang Y T, Ge H L, Sun Zhen, et al. Hyperspectral inversion of Dendrobium officinale polysaccharide content. Spectroscopy and Spectral Analysis, 2014; 34(06): 1645–1648. doi: 10.3964/j.issn.1000-0593(2014)06-1645-04. (in Chinese)

[25] Shi J Y. Diagnosis of nutrient element deficiency of facility cultivation crops based on hyperspectral image technology. Jiangsu University, 2012. (in Chinese)

[26] Shi J Y, Li W T, Hu X T, et al. Rapid diagnosis of nitrogen and magnesium deficiency in cucumber based on chlorophyll foliage distribution characteristics. Transactions of the Chinese Society of Agricultural Engineering, 2019; 35(13): 170–176. doi: 10.11975/j.issn.1002-6819.2019.13.019. (in Chinese)

[27] Zhu W J. Research on the detection of tomato nitrogen, phosphorus and potassium and their interaction based on polarization-hyperspectral multi-dimensional light information. Jiangsu University, 2014. (in Chinese)

[28] Sun J, Wei A G, Mao H P, et al. Qualitative analysis of nitrogen levels in lettuce leaves based on hyperspectral images and ELM. Journal of Agricultural Machinery, 2014; 45(07): 272–277. doi: 10.6041/j.issn.1000-1298.2014.07.042. (in Chinese)