DISCRIMINATION OF CERASUS HUMILIS FRUIT MATURITY BASED ON HYPERSPECTRAL IMAGING TECHNOLOGY
/
基于高光谱成像技术的欧李果成熟度判别

Bin WANG, Hua YANG*, Lili LI [1]
College of Information Science and Engineering, Shanxi Agricultural University, Taigu/China
Tel: +86-0354-6288165; E-mail: yanghuaxky@126.com
DOI: https://doi.org/10.35633/inmateh-70-10

Keywords: *Cerasus humilis fruit, maturity, hyperspectral imaging, nondestructive examination*

ABSTRACT

In order to realize the rapid and accurate identification of different maturity of Cerasus humilis fruit, this study explored the nondestructive testing method of Cerasus Humilis fruit maturity based on hyperspectral imaging technology. The hyperspectral data of 320 samples of Cerasus humilis fruit were collected by using a hyperspectral imaging system in the range of 895~1700 nm. By comparing the prediction accuracy of the partial least squares (PLS) model established by four preprocessing methods, the competitive adaptive reweighted algorithm (CARS), successive projection algorithm (SPA), and random frog (RF) were used to extract characteristic wavelengths, and partial least squares-discriminant analysis (PLS-DA) and least squares-support vector machine (LS-SVM) discriminant models were established. The results showed that the SPA-LS-SVM model had the highest discrimination accuracy for the four types of maturity samples, and the discrimination accuracy of the correction set and prediction set were 85.00% and 87.50%, respectively. This study provides a theoretical reference for the rapid and nondestructive testing of the maturity of Cerasus Humilis fruit by hyperspectral imaging technology.

摘要

为了实现对不同成熟度欧李果进行快速、准确识别，本研究探讨基于高光谱成像技术对欧李果成熟度进行无损检测研究的方法。利用 895~1700 nm 范围内的高光谱成像系统采集不同成熟时期（转色期、着色期、成熟期、完熟期）的欧李果共 320 个样本的高光谱数据。通过对比 4 种预处理方法建立的 PLS 模型预测精度，应用 CARS、SPA、RF 提取特征波长，并分别建立 PLS-DA 和 LS-SVM 判别模型。结果表明，SPA-LS-SVM 模型对 4 类成熟度样本的判别准确率最高，其校正集和预测集的判别准确率分别为 85.00% 和 87.50%。该研究为高光谱成像技术在欧李果成熟度的快速、无损检测提供了理论参考。

INTRODUCTION

As a unique fruit in China, the Cerasus Humilis fruit is also called “calcium fruit” because it is rich in active calcium and easy to be absorbed by the human body. The fruit is bright in color, unique in flavor and rich in nutrition. It can be eaten as fresh fruit or processed into fruit juice, wine, vinegar and other products. Maturity is an important factor that determines the shelf life, edible quality and postharvest storage of fruits. With the general improvement of social economic level and consumers' purchasing power, people pay more and more attention to the quality and safety of fruits. Determining the optimal maturity is the key to ensure the quality and storage of fruits. At present, the method of distinguishing the mature stage of the Cerasus humilis fruit mainly depends on people's experience and intuition. However, this method is inefficient and subjective, and cannot cope with large-scale production and modern fruit processing. Therefore, it is necessary to explore a fast and nondestructive method for determining the ripeness of Cerasus Humilis fruit to improve the market value and consumer satisfaction.

In recent years, hyperspectral imaging technology, a fusion technology integrating digital image and spectral technology, can provide spatial and spectral information of the target at the same time, and has been widely used in the detection of agricultural product maturity *(Yuan et al., 2021)*. *Shao et al. (2020)* used hyperspectral imaging technology (400~1000 nm) to analyze the maturity of Feicheng peach (green ripening stage, color changing stage).

[1] *Bin Wang, As Lec. Ph.D. Eng.; Hua Yang*, Prof. Ph.D. Eng.; Lili Li, As Lec. Ph.D. Eng.*

The sequential forward selection (SFS) algorithm was adopted to extract three characteristic wavelengths to establish an artificial neural network (ANN) prediction model. The total prediction accuracy of the model was 98.3%. *Li Lili et al. (2019)* used hyperspectral imaging technology (420~1000 nm) to identify Cerasus Humilis fruits with different maturity (immature, semi-mature, mature and over-mature). And the results showed that the established SPA-PLS model had the highest discrimination accuracy, and the accuracy reached 91.25%. *Pu et al. (2019)* used the hyperspectral imaging system to collect hyperspectral images of bananas at different maturity stages (the 2, 4 and 6 stages), and adopted regression coefficient method to choose characteristic wavelengths to establish three classification models (K-nearest neighbor algorithm, SIMCA, PLS-DA). Among them, the accuracy of PLS-DA model was the best (93.3%). *Zhang et al. (2020)* used the 300~1100 nm hyperspectral imaging system to divide bagged red Fuji apples into three maturity levels (immature, harvest maturity, edible maturity) with starch index as the maturity index, and established five discrimination models. The research showed that the RF-SPA-LS-SVM model had the highest classification accuracy (89.05%) for prediction sets. *Zou et al. (2019)* used hyperspectral imaging technology to classify peanuts with different maturity (immature and mature), and established multiple discrimination models (PLS-DA, LS-SVM). The results showed that LS-SVM model had the highest discrimination accuracy for immature and mature peanuts, which were 92.36% and 99.43% respectively. The above studies show that it is feasible to use hyperspectral imaging technology to identify fruit maturity. However, no researchers have used hyperspectral imaging to identify the maturity of Cerasus Humilis fruit in relevant studies.

In this study, the hyperspectral imaging system is used to obtain the hyperspectral image of the “Nongda No. 6” Cerasus humilis fruit samples in the range of 900~1700 nm, and extract the average spectral data of the region of interest (ROI). The effects of five spectral pretreatment methods on the performance of PLS model were analyzed and the optimal pretreatment method was determined. The competitive adaptive reweighted algorithm (CARS), successive projection algorithm (SPA), and random frog (RF) were used to extract the characteristic wavelength, and the partial least squares-discriminant analysis (PLS-DA) and least squares-support vector machine (LS-SVM) discriminant models were established based on the characteristic wavelength, so as to provide a theoretical basis for further development of online sorting equipment of Cerasus Humilis fruits in different mature periods.

MATERIALS AND METHODS

Sample Collection

In this study, fresh edible Cerasus Humilis fruit (Nongda No. 6) was collected from a Cerasus Humilis planting demonstration base in the agricultural high-tech industry demonstration zone (112°29'E, 37°23'N) of Jinzhong, China, July 30, 2020. After harvest, the sample is placed in a cryogenic crisper, transported to the laboratory on the same day as the sampling, to avoid the effects of individual differences on the testing results. A total of 320 samples (include 66 color turning stages, 90 coloring stages, 84 maturity stages, and 80 full ripe stages) without defects, bruises, scar, and relatively uniform shape were selected. Finally, the selected samples were numbered, rinsed, wiped. Prior to hyperspectral image acquisition, the samples were taken out and placed in a laboratory (25°C, 40% relative humidity) for 2h to avoid the effect of temperature on the spectrum and the fruit quality. Samples of Cerasus Humilis fruit at four maturity levels (immaturity, white maturity, early-red maturity, half-red maturity, and full maturity) are shown in Fig.1.

The SPXY algorithm *(Galvao et al., 2005)* was used to divide the Cerasus Humilis fruit dataset into a calibration set (240 samples) and a prediction set (80 samples) with a ratio of 3:1.

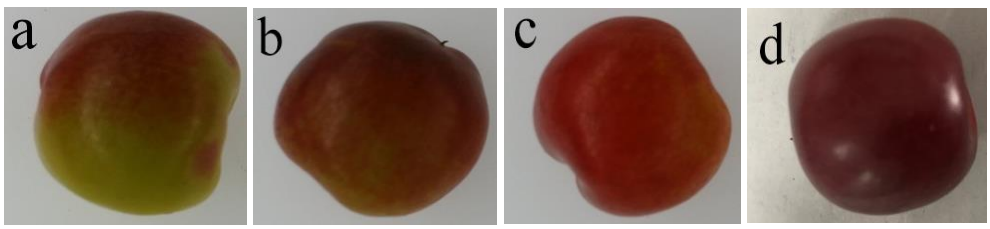


Fig. 1 - Samples of Cerasus Humilis fruit at different maturity stages
a) Color turning stage; b) Coloring stages; c) Maturity stage; d) Full ripe stage

Spectrum Acquisition

In this study, a hyperspectral imaging system (GaiaSorter, Zolix Instruments Co. Ltd., China) was used to collect spectral and image information. The spectral range is 895~1700 nm and the spectral resolution is 5 nm, and the number of pixels is 320×256. The entire system consists of a high-spectrum spectrometer, CCD camera, four 250 W halogen lamps, a precision mobile platform controlled by stepper motors, and a computer, as shown in Fig.2. Among them, four 250 W halogen lamps were placed as a lighting device at a 45°angle.

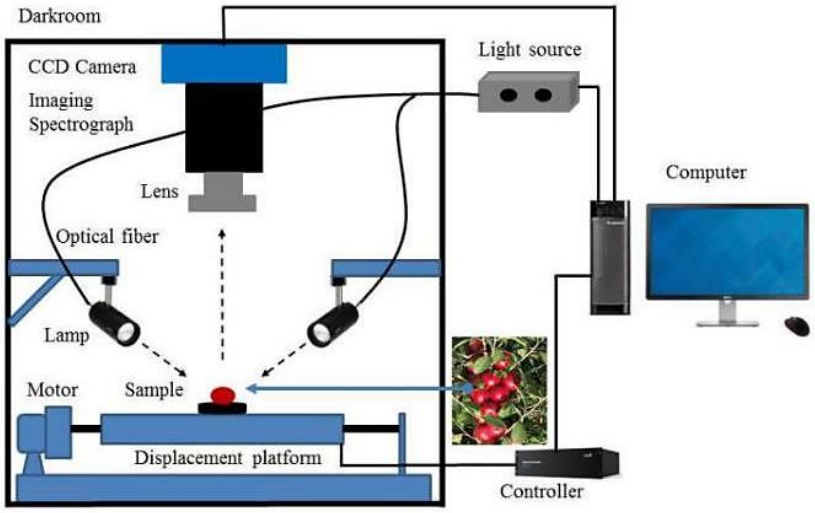


Fig. 2 - The schematic diagram of hyperspectral imaging system

To ensure clear images, avoid information oversaturation and imaging distortion. Based on the system configuration, the equatorial region of Cerasus humilis fruit was facing the camera, and the camera was fixed at a distance of 280 mm from the lens to the samples' surface. Each Cerasus humilis fruit was placed on the sample table to be scanned at an 8 mm/s constant speed line by line using 150 ms exposure time to create a hyperspectral image. Due to the uneven distribution of light intensity and the presence of dark currents in the sensor, the resulting image had high noise. Therefore, the raw images were corrected according to the following equation:

$$I = \frac{R - B}{W - B} \times 100\% \quad (1)$$

where:

I is the corrected image;
R is the original image;
B is the image of the blackboard correction;
W is the image of the whiteboard correction.

Discriminant Model

Partial least squares-discriminant analysis (PLS-DA) is a multivariate statistical analysis method combining partial least square method and linear discriminant analysis method. The optimal number of principal components is obtained by cross validation, and then linear discriminant analysis is carried out to solve the multicollinearity problem of independent variables in regression analysis. In this study, principal components were selected according to the interaction test, the maximum number of principal components was set as 10, and 10-fold interaction test was performed.

Least squares-support vector machine (LS-SVM) is a multivariate statistical method. Its algorithm is the least squares method. Its principle is structural risk minimization, which can effectively implement data classification and processing. It also reduces training time and simplifies computational complexity. In this study, the optimal characteristic wavelength is used as the input, and four maturity categories are used as the output parameters. After optimization, the optimal regular parameter (γ) and square bandwidth (σ^2) were obtained.

Software Tools

In this study, ENVI5.0 (Version 5.0, ITT Visual Information Solutions, Boulder, USA) software was used to analyze hyperspectral images. The Unscrambler X10.1'' (CAMO PROCESS AS, Oslo, Norway) and MATLAB R2009a software (The Mathworks, Inc., Natick, MA, USA) were used to process and analyze data. In addition, Origin Pro 8.0 SR0 (Origin Lab Corporation, Northampton, MA, USA) software was used to design graphs.

RESULTS AND DISCUSSION

Spectral Characteristics Analyses

The ROI with 40×60 pixels was manually extracted from hyperspectral image of equatorial region of each sample. The spectra of each pixel within the ROI were extracted, and then the mean spectrum was calculated, as shown in Fig. 3a. Meanwhile, the average spectra of various samples were obtained, as shown in Fig. 3b. Because of the low signal-to-noise ratios and large noise at 895~945 nm, 1675~1700 nm, the wavelength range of 945~1675 nm (230 bands) were selected for further analysis in this study.

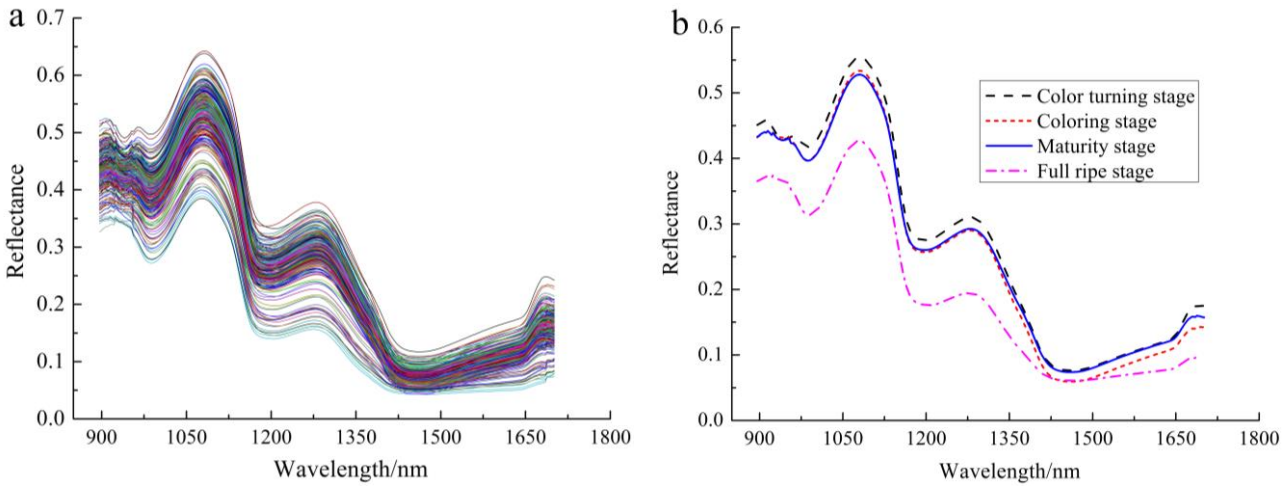


Fig. 3 - Raw reflectance spectra of Cerasus Humilis fruit samples at different maturity

As shown in Fig. 3a, spectra of samples at different maturity stages have some crossover and overlap, but the trend of spectral curves is very similar. It can be seen from Fig. 3a that the spectra have obvious absorption peaks at 980, 1195 and 1460 nm. The absorption peak at around 980 nm and 1195 nm were related to the second overtone of O-H stretching *(Shinzawa et al., 2011)* and the second overtone of C-H stretching, respectively *(Liu et al., 2010)*. There is an obvious absorption peak near 1460 nm, which is related to the first order frequency doubling of the stretching vibration of the O-H bond *(Osborne et al., 2006)*. It can be seen from Fig. 3b that the spectral curves at different maturity stages (color turning stage, coloring stage, maturity stage, and full ripe stage) have similar trends, but the spectral reflectance will vary greatly with the extension of the sample maturity stage, and the spectral reflectance of the sample at maturity stage is the lowest. In the range of 1180~1300 nm, the fruit reflectance varied greatly, and tended to decrease gradually, which may be related to the changes of the contents of chemical components in the fruits of Cerasus Humilis at different maturity stages.

Spectral Pretreatment

Selecting an appropriate spectral preprocessing method can eliminate or weaken the influence of non-target factors on the original spectral information, and improve the signal-to-noise ratio and the detection accuracy and stability of the model *(Chi et al., 2021)*. In the present work, four spectral pre-processing methods, including standard normal variate (SNV), multiplicative scatter correction (MSC), baseline correction (BC), and de-trending (De-T) were used to preprocess the raw spectral data. The PLS prediction models were established respectively, and the optimal pretreatment methods were compared and selected. The prediction results are shown in Table 1.

Table 1

Prediction results of PLS models built by different preprocessing methods

Pretreatment methods	Calibration set		Prediction set	
	Rc	RMSEC	Rp	RMSEP
Original spectra	0.87	0.51	0.88	0.50
SNV	0.76	0.68	0.85	0.54
MSC	0.85	0.58	0.85	0.57
BC	0.88	0.49	0.89	0.48
De-T	0.86	0.52	0.88	0.51

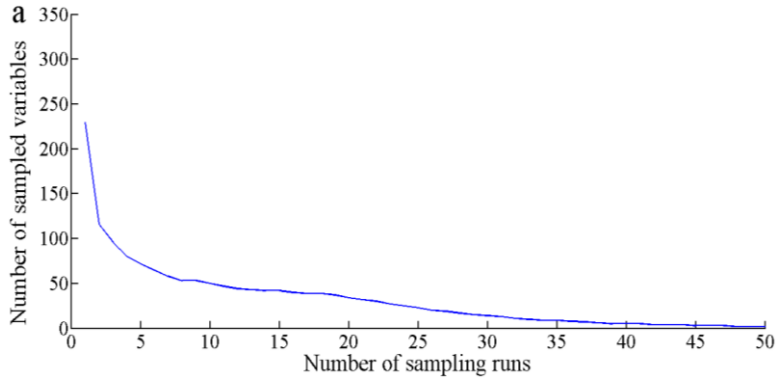
It can be seen from Table 1 that compared with the predicted results of the PLS model established by the original spectrum without pretreatment, the predicted results of the PLS model established by the three pretreatment methods (SNV, MSC, De-T) were poor. The PLS model established after BC method preprocessing has the highest prediction accuracy, with the correlation coefficient (Rc) of the correction set and the correlation coefficient (Rp) of the prediction set being 0.88 and 0.89, respectively. The root mean square error (RMSEC) of the correction set and the mean square error (RMSEP) of the prediction set being 0.49 and 0.48, respectively, both of which are relatively low and close. Therefore, the spectral data processed by BC were used for further analyses in this study.

Effective Wavelength Selection

Competitive Adaptive Reweighted Sampling (CARS)

The CARS algorithm is a feature information filtering method that has been widely used in recent years. It is based on Darwin's theory of evolution and follows the “survival of the fittest” principle. In the CARS procedure, the absolute coefficients of variables in the PLS model are set as an index for evaluating the importance of each variable and the wavelengths with large absolute coefficients were regarded as optimal wavelengths *(Wang et al., 2021)*.

In the process of CARS, the number of Monte Carlo sampling runs was set to 50 and the final variable number to be selected was determined by 5-fold cross validation. Fig. 4a, b and c show the trends of the number of sampled variables, RMSECV values, and regression coefficient paths for each variable as the number of Monte Carlo samples increases for each CARS run, respectively. From Fig. 4a, it can be seen that the number of wavelengths gradually decreases and finally plateaus with the gradual increase of the number of sampling runs, which verifies the fast selection phase and refined selection phase in the wavelength screening process. Fig. 4b showed the changing trends of RMSECV with sampling runs from each sampling. When the number of sampling runs gradually increases to 36, the cross-validation RMSECV gradually decreases and then shows a trend of increasing. When the RMSECV gradually becomes smaller, it means that the useless information in the spectral information was eliminated, when the RMSECV increases, it means that the useful information of the spectral information was eliminated. In Fig. 4c, each curve represents the regression coefficients at different sampling runs for each variable, the position of the vertical line marked with "*" in the Fig. 4c indicates that the RMSECV reaches a minimum value of 0.51 when the number of sampling runs was 36. Finally, the number of selected wavelengths was 14 from 230 wavelengths, these variables were 960, 1027, 1091, 1151, 1161, 1164, 1167, 1173, 1355, 1371, 1374, 1380, 1625, and 1650 nm, respectively.



a)

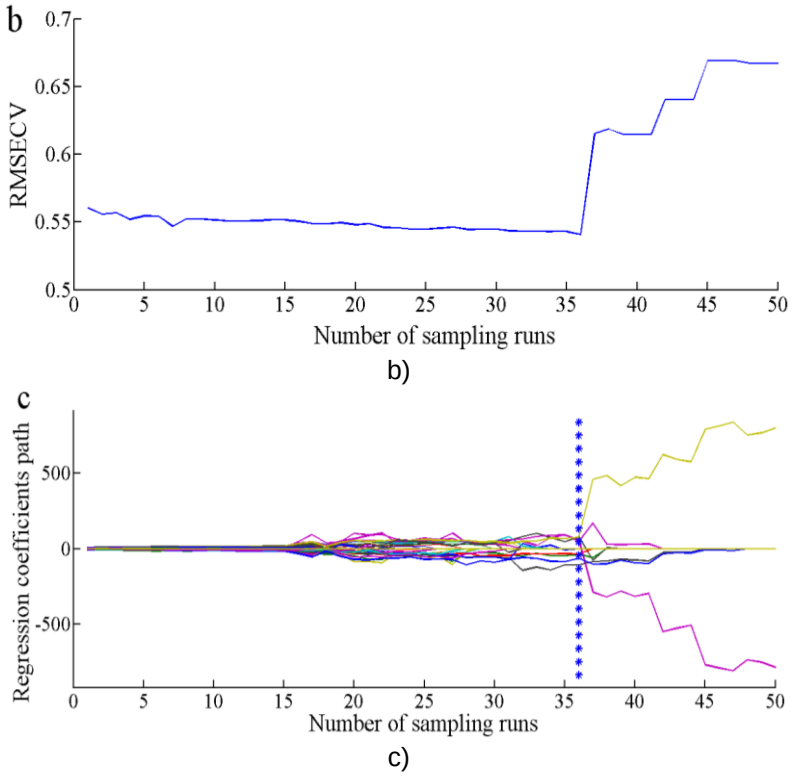


Fig. 4 - Key variables selection results of CARS

Successive Projection Algorithm (SPA)

SPA is a forward selection method proposed by Araújo *(Araújo et al., 2001)*. This algorithm is to avoid the overlap of effective information in spectral variables and eliminate the collinearity between them through multiple variable information iterations, find a group of characteristic variables with low redundant information, minimum collinearity and representativeness, which can represent most spectral information of tested samples, avoid the overlap of information to the greatest extent, and improve the modeling speed.

The SPA algorithm is used to extract the characteristic wavelengths from the pre-processed full spectrum data. In this study, the parameters of the minimum and maximum numbers of variables selected in the SPA procedure were 1 and 30, respectively. Fig. 5 shows the distribution of root mean square errors (RMSE) for the different number of variables chosen by the SPA algorithm. When the 19 variables were selected (marked as open blue square), the RMSEP reached its optimal value (with RMSE=0.46). Fig. 6 shows the selected 19 variables, the nineteen characteristic wave-lengths were 1609, 1377, 1113, 1580, 1068, 1275, 1355, 1212, 1641, 1672, 1141, 953, 1323, 957, 1399, 1488, 1669, 1676, and 1460 nm, respectively. Accounting for 3.91% of the total wavelength, the importance of the wavelength decreases in turn.

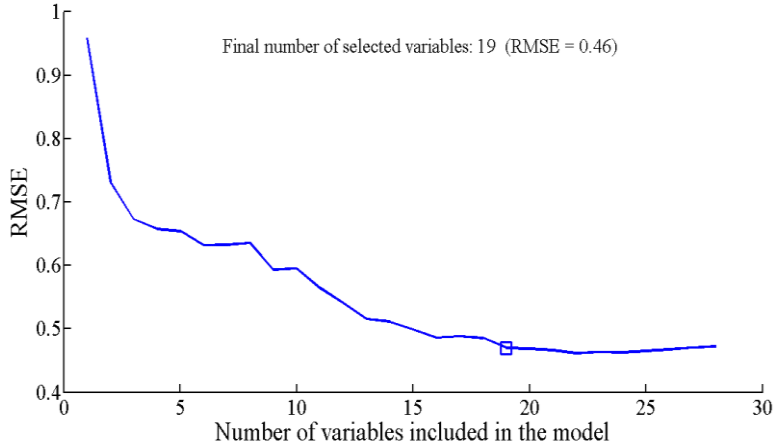


Fig. 5 - The relationship between wave number and root mean square error

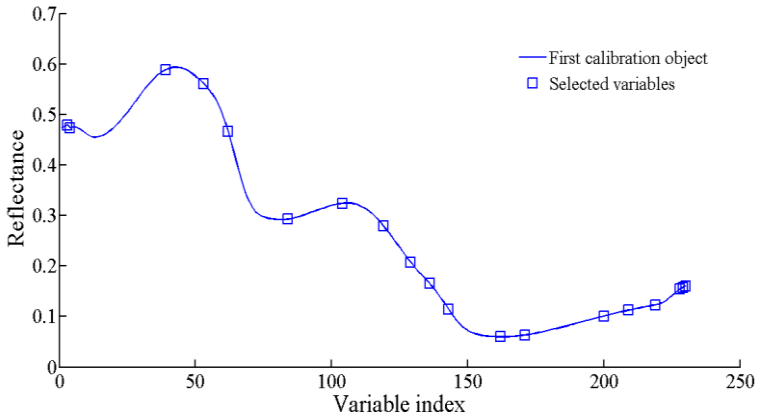


Fig. 6 - Feature wavelength extracted by SPA

Random Frog (RF)

Random Frog (RF) is a mathematically simple method with high computational efficiency. It is similar to reversible jump Markov chain Monte Carlo algorithm, which can iterate multidimensional data variables and calculate the weight value of each variable. The higher the value, the greater the probability of being selected, and the more important the corresponding wavelength *(Yan et al., 2020)*.

In order to reduce the influence of random factors, the algorithm needs to run several times and calculate the results. In this study, the frequency of RF algorithm was set as 2000, and when RMSECV reached the minimum value of 0.48, the threshold was selected as 0.31. The process of using RF algorithm to select characteristic wavelength was shown in Fig. 7. It can be seen from Fig. 7 that the band selection probability ranges from 0.0 to 0.8. Only a few variables have prominent selection probability peaks, which can be identified as characteristic wavelength. The first 10 data coordinate points with large probability values are (133,0.7525), (134,0.4245), (121,0.4135), (151,0.3670), (97,0.3495), (182,0.3460), (52,0.3290), (62,0.3175), (8,0.3135), (53,0.3075), respectively, arranged in descending order. The corresponding characteristic wavelengths are 1367, 1371, 1329,1425, 1253, 1523, 1110, 1142, 969, and 1113 nm respectively, accounting for 4.35% of the original wavelength.

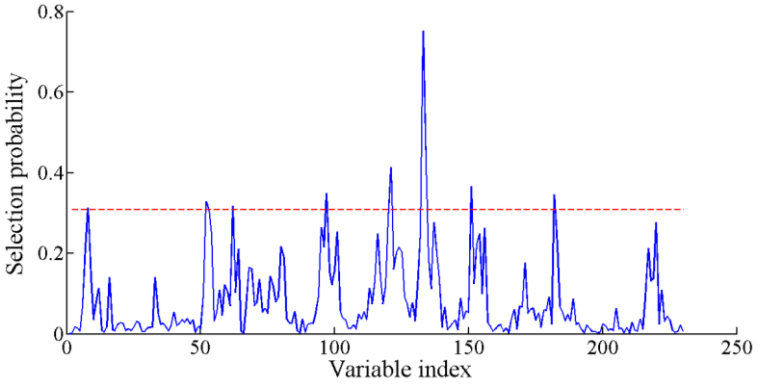


Fig. 7 - Feature wavelength extracted by RF

Establishment of a Maturity Model Using the Feature Wavelengths

The PLS-DA and LS-SVM discrimination models were established by assigning the samples of different maturity periods (color turning stage, coloring stage, maturity stage, and full ripe stage) with values of 1, 2, 3, and 4, respectively.

When the established discriminant model is used to discriminate samples at different mature stages, if the correct prediction value of color turning stage is within the threshold value [0.5, 1.5], it is determined that the prediction category is consistent with the assumed category, that is, the color turning stage. By analogy, the threshold range of the coloring stage is (1.5, 2.5], the maturity stage is (2.5, 3.5], and the full ripe stage is (3.5, 4.5]. Based on the discriminant results of different models of characteristic wavelength analysis, the most effective method of extracting characteristic wavelength and modeling method are determined. Table 2 shows the discrimination results of PLS-DA and LS-SVM models based on the characteristic wavelengths of Cerasus Humilis in different maturity periods.

Table 2

Discriminant results of models by the characteristic wavelength of Cerasus Humilis fruit maturity

Model	Variable selection methods (No. of variables)	Calibration set		Prediction set	
		Number of misjudgments	Discriminant accuracy (%)	Number of misjudgments	Discriminant accuracy (%)
PLS-DA	Original spectra（230）	76	68.33	27	66.25
	SPA（19）	68	71.67	20	75.00
	CARS（14）	70	70.83	25	68.75
	RF（10）	93	61.25	36	55.00
LS-SVM	Original spectra（230）	37	84.58	12	85.00
	SPA（19）	36	85.00	10	87.50
	CARS（14）	41	82.92	14	82.50
	RF（10）	68	71.67	27	66.25

As can be seen from Table 2, among the PLS-DA discriminant models, the RF-PLS-DA model has the worst discriminant results for Cerasus Humilis fruits at different maturity stages, and the accuracy of correction set and prediction set are 61.25% and 55.00%, respectively. Compared with the original spectra data, the discriminant results of the model based on the characteristic wavelengths extracted by SPA and CARS are slightly improved by 8.75% and 2.50% respectively. Among them, SPA-PLS-DA model has the best discrimination result for Cerasus Humilis fruits at different maturity stages. The number of selected characteristic wavelengths accounts for 8.26% of the total spectrum data, and the discrimination accuracy of correction set and prediction set are 71.67% and 75.00%, respectively.

In the LS-SVM discriminant model, compared with the discriminant results of the model established by the original spectrum, the model established by using the characteristic wavelength extracted by CARS and RF methods has significantly reduced the discriminant results of the prediction set, while the prediction set of RF-LS-SVM model has the lowest discriminant accuracy, with a value of 66.25%. It shows that the effective information in the original spectral data is eliminated in the process of extracting characteristic wavelength by RF algorithm. By comprehensive comparison, when $\gamma=6.17\times10^3$, $\sigma^2=71.21$, SPA-LS-SVM model has the best discriminative results for plum fruit at different maturity stages. The number of selected characteristic wavelengths only accounts for 8.26% of the whole spectrum data, and the discriminative accuracy of correction set and prediction set are 85.00% and 87.50%, respectively.

By comparing the discriminant accuracy of the optimal linear model and the optimal nonlinear model, it can be seen that the SPA-LS-SVM model has the highest discriminant accuracy. *Singh et al. (2021)* research on barley seed varieties recognition based on hyperspectral imaging technology combined with convolutional neural network, and discrimination accuracy of the nonlinear convolutional neural network model is superior to that of the linear model, which is consistent with the conclusion of this study. In this study, the accuracy of the optimal discriminant model for the prediction set (87.50%) was slightly lower than that of *Li et al., (2019),* (91.25%) and *Liu et al., (2022),* (90.12%). This may be due to the different research objects, which leads to the deviation of the hyperspectral image information collected. In addition, the 19 characteristic wavelengths extracted by SPA algorithm in this study are consistent with previous research reports *(Riza et al., 2017)*. That is 967~1095 nm is related to the overtone of C-H stretching, 1165~1390 nm is associated with the first overtone of bond O-H, and 1620~1800 nm is associated with the first overtone of bond C-H *(Wang et al., 2021)*. It is further indicated that the characteristic wavelength extracted by SPA can be used as the key variable to reflect the different ripening stages of Cerasus Humilis fruit, which can simplify the model and improve the accuracy of discriminant.

Fig. 8 shows the discriminant results of the prediction set of 4 types of Cerasus Humilis fruits based on the optimal SPA-LS-SVM model. Among them, 3 Cerasus Humilis fruits in the color turning stage were misjudged as the coloring stage, 4 Cerasus Humilis fruits in the coloring stage were misjudged as the color turning stage, and 3 Cerasus Humilis fruits in the maturity stage were misjudged as the coloring stage.

This may be due to the fact that a few Cerasus Humilis fruits in the color turning stage, the coloring stage and the maturity stage had no obvious color changes and the absorption spectrum was consistent, which was easy to cause confusion.

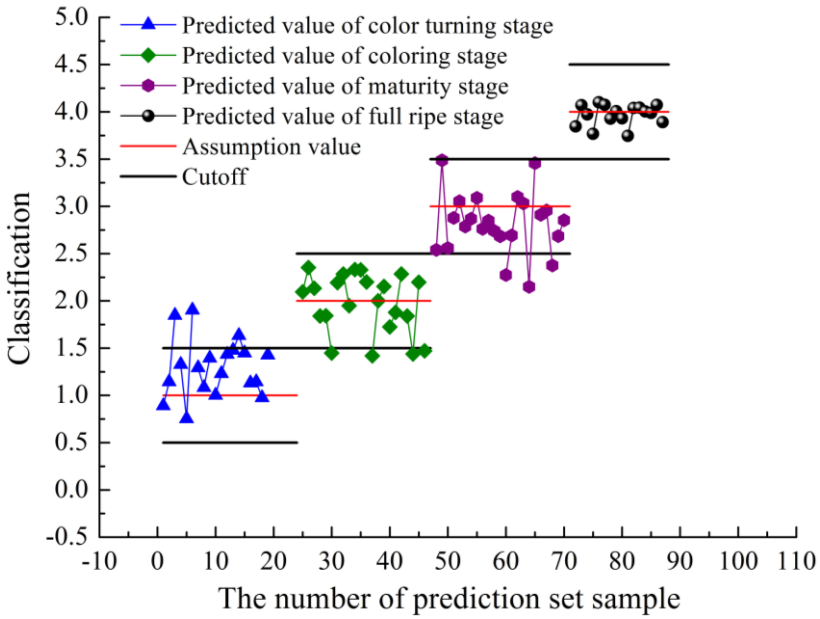


Fig. 8 - The discriminant result of SPA-LS-SVM model

CONCLUSIONS

In this study, hyperspectral imaging technology was used to nondestructive discriminant research of Cerasus Humilis fruit at different maturity stages (color turning stage, coloring stage, maturity stage, and full ripe stage), the effects of different spectral preprocessing methods and characteristic wavelength extraction methods on the accuracy of the model were compared. The main research conclusions are as follows:

(1) Based on the spectral information in the range of 945~1675 nm, the PLS prediction models established by four pretreatment methods (SNV, MSC, BC, De-T) were compared and analyzed. The results showed that the PLS model pretreated by BC method was the best, with Rc and RMSEC of 0.88 and 0.49, Rp and RMSEP of 0.89 and 0.48, respectively.

(2) Based on the original spectrum pretreated by BC method, the characteristic wavelength was extracted by CARS, SPA and RF algorithms, and the PLS-DA and LS-SVM discriminative models were established respectively. Comparative analysis showed that the SPA-LS-SVM model had the best discriminant results for different maturity stages of Cerasus Humilis fruit ($\gamma=6.17\times10^3$ and $\sigma^2=71.21$). The number of selected characteristic wavelengths only accounted for 8.26% of the full spectrum, and the discriminant accuracy of correction set and prediction set were 85.00% and 87.50%, respectively. This study provides a theoretical basis for rapid and non-destructive detection of the ripeness of Cerasus Humilis fruits.

ACKNOWLEDGEMENT

This research was supported by the Province Basic Research Program Project (Free Exploration) (Grant No.20220302123641), the Doctor Scientific Research Foundation of Shanxi Agricultural University (2023BQ63).

REFERENCES

[1] Araújo, M.C.U., Saldanha, T.C.B., Galvao, R.K.H., Yoneyama, T., Chame, H.C., & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and intelligent laboratory systems*, *57*(2), 65-73. https://doi.org/10.1016/S0169-7439(01)00119-8

[2] Chi, J.T., Zhang, S.J., Ren, R., Lian, M.R., & Mu, B.Y. (2021). Detection of eggplant external defects using hyperspectral technology. *Modern Food Science and Technology*, *37*(9), 279-284+178. https://doi.org/10.13982/j.mfst.1673-9078.2021.9.0034

[3] Galvao, R.K.H., Araujo, M.C.U., José, G.E., Pontes, M.J.C., & Saldanha, T.C.B. (2005). A method for calibration and validation subset partitioning. *Talanta,* 67, 736–740. https://doi.org/10.1016/j.talanta.2005.03.025

[4] Li, L.L., Wang, B., Zhang, X.H., & Zhang, S.J. (2019). Discrimination of plum fruit maturity based on hyperspectral imaging technology. *Modern Food Science and Technology*, *35*(6): 258-263. https://doi.org/10.13982/j.mfst.1673-9078.2019.6.034

[5] Liu, J. X., He, X. W., Luo, H. P., Xu, J. Y., & Shen, L. L. (2022). Maturity discrimination model of little white apricot based on hyperspectral imaging technology. *Food Research and Development*, *43*(15): 158-165. https://doi.org/10.12161/j.issn.1005-6521.2022.15.022

[6] Liu, Y., Sun, X., Zhang, H., & Aiguo, O. (2010). Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy. *Computers and Electronics in Agriculture*, *71*, S10-S14. https://doi.org/10.1016/j.compag.2009.09.005

[7] Osborne, B. G. (2006). Near-infrared spectroscopy in food analysis. *Encyclopedia of analytical chemistry: applications, theory and instrumentation*. https://doi.org/10.1002/9780470027318.a1018

[8] Pu, Y. Y., Sun, D. W., Buccheri, M., Grassi, M., Cattaneo, T. M., & Gowen, A. (2019). Ripeness classification of bananito fruit (Musa acuminata, AA): a comparison study of visible spectroscopy and hyperspectral imaging. *Food Analytical Methods*, *12*(8), 1693-1704. https://doi.org/10.1007/s12161-019-01506-7

[9] Al Riza, D. F., Suzuki, T., Ogawa, Y., & Kondo, N. (2017). Diffuse reflectance characteristic of potato surface for external defects discrimination. *Postharvest Biology and Technology*, *133*, 12-19. https://doi.org/10.1016/j.postharvbio.2017.07.006

[10] Shao, Y., Wang, Y., Xuan, G., Gao, C., Wang, K., & Gao, Z. (2020). Visual detection of SSC and firmness and maturity prediction for feicheng peach by using hyperspectral imaging. *Transactions of the Chinese Society for Agricultural Machinery*, *51*(8), 344-350. https://doi.org/10.6041/j.issn.1000-1298.2020.08.038

[11] Shinzawa, H., Ritthiruangdej, P., & Ozaki, Y. (2011). Kernel analysis of partial least squares (PLS) regression models. *Applied spectroscopy*, *65*(5), 549-556. https://doi.org/10.1366/10-06187

[12] Singh, T., Garg, N. M., & Iyengar, S. R. (2021). Nondestructive identification of barley seeds variety using near infrared hyperspectral imaging coupled with convolutional neural network. *Journal of Food Process Engineering*, *44*(10), e13821. https://doi.org/10.1111/jfpe.13821

[13] Wang, B., He, J., Zhang, S. J., & Li, L. L. (2021). Nondestructive prediction and visualization of total flavonoids content in Cerasus Humilis fruit during storage periods based on hyperspectral imaging technique. *Journal of Food Process Engineering*, *44*(10), e13807. https://doi.org/10.1111/jfpe.13807

[14] Wang, B., He, J. L., & Li, L.L. (2021). On-line detection of cerasus humilis fruit based on VIS/NIR spectroscopy combined with variable selection methods and GA-BP model. *INMATEH-Agricultural Engineering*, *63*(1), 199-210. https://doi.org/10.35633/inmateh-63-20

[15] Yan, L., Pang, L., Wang, H., & Xiao, J. (2020). Recognition of different Longjing fresh tea varieties using hyperspectral imaging technology and chemometrics. *Journal of Food Process Engineering*, *43*(4), e13378. https://doi.org/10.1111/jfpe.13378

[16] Yuan, R., Liu, G. S., He, J. G., Kang, N. B., & Ma, L. M. (2021). Quantitative damage identification of Lingwu long jujube based on visible near-infrared hyperspectral imaging. *Spectroscopy and Spectral Analysis*, *41*(4), 1182-1187. https://doi.org/10.3964/j.issn.1000-0593-1182-06

[17] Zhang, M., Zhang, B., Li, H., Tian, S., Zhang, H., & Zhao, J. (2020). Determination of bagged 'Fuji' apple maturity by visible and near-infrared spectroscopy combined with a machine learning algorithm. *Infrared Physics & Technology*, *111*, 103529. https://doi.org/10.1016/j.infrared.2020.103529

[18] Zou, S., Tseng, Y. C., Zare, A., Rowland, D. L., Tillman, B. L., & Yoon, S. C. (2019). Peanut maturity classification using hyperspectral imagery. *Biosystems Engineering*, *188*, 165-177. https://doi.org/10.1016/j.biosystemseng.2019.10.019