ISSN: 1813-4890

Detection of soluble solids content of cherry tomatoes based on hyperspectral imaging technology

Man Li [1], Li Deng [2], Qian Zhang [3] and Hongyan Wang [4]

[1]Chongqing University of International Business and Economics Chongqing Super City Digital Governance College, Hechuan 401520, China;

[2]Chongqing University of International Business and Economics Chongqing Super City Digital Governance College, Hechuan 401520, China;

[3]Chongqing University of International Business and Economics Chongqing Super City Digital Governance College, Hechuan 401520, China;

[4]Chongqing University of International Business and Economics Chongqing Super City Digital Governance College, Hechuan 401520, China.

Abstract

Using hyperspectral imaging technology to accurately and non-destructively detect the quality of cherry tomatoes can provide a basis for fruit storage. Based on the reflectance spectrum of soluble solids content (SSC) of cherry tomatoes, the characteristic wavelengths were screened by spectral preprocessing method, uninformative variable elimination (UVE), competitive adaptive weight algorithm and successive projections algorithm (SPA). In addition, SPA-UVE method was used to further optimize the screening process. Based on the selected characteristic wavelengths, the SSC detection model based on multiple scattering correction spectra was constructed by using partial least squares regression, principal component regression and BP neural network. The results show that 12 characteristic wavelengths are selected, and the model detection accuracy based on multiple scattering correction and SPA-UVE method is better than other models. The model can accurately detect the SSC of cherry tomatoes on both the calibration set and the prediction set (RC=0.950, RMSEC=0.262; RP=0.928, RMSEP=0.255; RPD=2.063). The method of detecting the quality of cherry tomato based on hyperspectral imaging technology provides new technical support for the sustainable development of cherry tomato production.

Keywords

Hyperspectral ; saint fruit ; soluble solids content ; BP neural network.

1. Introduction

The cherry tomato is known as 'little golden fruit 'and' love fruit ' because of its lovely appearance and unique flavor. It is one of the most popular ' four major fruits ' in the world [1]. It is rich in vitamin C, which can not only effectively prevent diseases such as vascular sclerosis, hypertension and coronary heart disease, but also enhance immunity and promote human metabolism [2]. Soluble solid content (SSC) is an important index to evaluate the quality of cherry tomato, which directly reflects the maturity and nutrient content of fruit, and has a significant effect on flavor, taste and shelf life [3]. However, although the traditional chemical detection methods are effective, they are difficult to meet the needs of large-scale detection due to the problems of time-consuming, complex operation and destructiveness. Therefore, it is of great significance to develop a rapid, efficient and non-destructive detection technology for the quality detection of cherry tomatoes, which is helpful to improve the efficiency of fruit grading and sorting and promote industrial development.

Hyperspectral imaging technology emerged in the 1980 s. It combines image processing, optics, computer and other disciplines to obtain high-resolution image information reflecting the chemical composition and internal structure of samples [4]. This technology has the characteristics of fast, non-destructive, low cost and ' combination of map and spectrum ', and has been widely used in the field of fruit quality detection [5]. For example, Huang et al. [6] used hyperspectral technology to detect the soluble solids content of apples, used continuous projection algorithm (SPA) to extract characteristic wavelengths and establish prediction models, and achieved high prediction accuracy. Li et al. [7] and Jiali Xu et al. [8] also used hyperspectral imaging technology to non-destructively detect the sugar content of pears and kiwifruits, respectively, and verified the effectiveness of the technology. However, the research on the soluble solids content of cherry tomatoes is still limited.

Therefore, this study takes cherry tomatoes as the test object, proposes an optimization method for feature wavelength extraction through hyperspectral imaging technology combined with stoichiometric methods, and compares and analyzes the effectiveness of different modeling methods, in order to provide new technical means and theoretical basis for non-destructive detection of soluble solids content in cherry tomatoes.

2. Materials and methods

2.1. Testing material

The experimental research object was the cherry tomatoes purchased from the fruit market in Wanzhou District.A total of 209 samples were selected, and the single fruit weight ranged from 3.73 g to 17.46 g. The fruit volume was similar, the peel color was uniform and the surface was not damaged. Before the experiment, all the samples were numbered in order, and the surface was wiped clean, and then placed in a room temperature environment for 12 h to ensure that the samples were basically consistent with the room temperature.

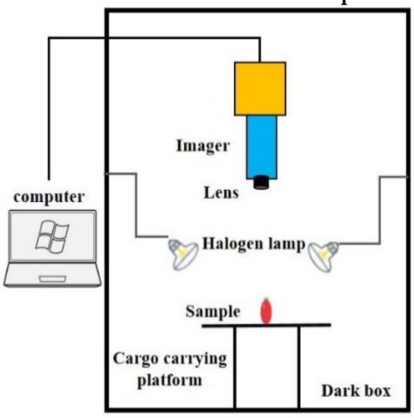

Fig. 1 Hyperspectral imaging system structure diagram

2.2. Hyperspectral image acquisition and correction

In this study, a five-bell optical hyperspectral imaging system was used, see Fig. 1. The system consists of two 150 W halogen lamps, an imaging spectrometer (ImSpectorV10E, Finland), a CCD camera (RaptorEM285CL, USA), a displacement electronic control platform and a computer. The spectral acquisition range is 400 ~ 1000 nm, and the spectral resolution is 2.8 nm. Before each hyperspectral data acquisition, the system needs to be opened and preheated for 0.5 h to ensure the stability of the light source and the acquisition system. The cherry tomato sample was placed on the displacement electronic control platform, and the sample was scanned at a speed of 2.6 mm/s through a distance of 450 mm from the CCD camera. In order to avoid image distortion or distortion, the CCD camera exposure time is set to 11 ms.

Due to the dark current and noise in the CCD camera, it may interfere with the accuracy of the spectral information. Therefore, the original image after acquisition needs to be immediately corrected in black and white [9]. Subsequently, ENVI5.3 software was used to select three 25×25 pixel regions around the equatorial region of cherry tomatoes as the region of interest (ROI), and the average spectral data of all pixels in the ROI were used as the original spectral data of cherry tomatoes.

2.3. Hyperspectral image acquisition and correction

After the hyperspectral image acquisition was completed, the cherry tomato samples were transversely cut along the equator to determine the soluble solids content (SSC). According to the national agricultural standard 'NY/T 26367-2014 Determination of soluble solids content in fruits and vegetables- Refractometer method '. The SSC of the samples was measured by SW-32A digital refractometer produced by Guangzhou Fast Electronic Technology Co., Ltd. The measurement range of the instrument is 0 ~ 32° Brix, the resolution is 0.1° Brix, and the error is ± 2%. In order to reduce the measurement error, each sample was measured three times, and the average value was taken as the final value of SSC of cherry tomatoes.

2.4. Extracting characteristic wavelengths

The spectral data obtained by hyperspectral image acquisition and correction has high dimension and information content, which not only brings great difficulty to data storage, transmission and processing, but also is easily disturbed by uninformative variables in modeling. The feature wavelength extraction method can select the wavelengths that are strongly correlated with the target variables (such as SSC) from the high-dimensional spectral data, thereby eliminating redundant information and improving the accuracy and generalization ability of the model. In this experiment, four feature wavelength extraction methods were used to extract the original spectral data by Matlab R2022 a software : uninformative variable elimination algorithm (UVE) [10], continuous projection algorithm (SPA) [11], competitive adaptive reweighted algorithm (CARS) [12] and SPA-UVE combination algorithm. By comparing and analyzing the effects of these four methods on the SSC detection of cherry tomatoes, the characteristic wavelengths that can characterize the SSC-related spectral information of cherry tomatoes were selected.

2.5. Modeling Method and Evaluation

The spectral data obtained by hyperspectral image acquisition and correction has high dimension and information content, which not only brings great difficulty to data storage, transmission and processing, but also is easily disturbed by uninformative variables in modeling. The feature wavelength extraction method can select the wavelengths that are strongly correlated with the target variables (such as SSC) from the high-dimensional spectral data, thereby eliminating redundant information and improving the accuracy and generalization ability of the model. In this experiment, four feature wavelength extraction methods were used to extract the original spectral data by Matlab R2022 a software : uninformative variable elimination algorithm (UVE) [10], continuous projection algorithm (SPA) [11], competitive adaptive reweighted algorithm (CARS) [12] and SPA-UVE combination algorithm. By comparing and analyzing the effects of these four methods on the SSC detection of cherry tomatoes, the characteristic wavelengths that can characterize the SSC-related spectral information of cherry tomatoes were selected.

3. Results and discussion

3.1. Abnormal sample elimination and sample division

The sample collection process of cherry tomatoes may be affected by factors such as hyperspectral noise interference or human operation, resulting in some abnormal samples, thereby reducing the accuracy and stability of the model. In order to improve the accuracy and stability of the model, the Monte Carlo cross validation method [16] was used to eliminate the abnormal samples. The number of cycles is set to 1000 times to ensure the stability of the results, and the preprocessing method is 'center' to eliminate the data deviation. The mean value (Mean) and standard deviation (STD) of the sample prediction residuals are calculated, and a scatter plot with Mean as the abscissa and STD as the ordinate is drawn, see Fig. 2. The 2.5 times of the mean value of Mean and STD are regarded as outliers [17], and the abnormal sample data with Mean greater than 1.245 and STD greater than 0.259 are eliminated in turn. A total of 7 samples are regarded as abnormal samples. After eliminating the four samples of No.74, No.89, No.123 and No.209, the RC value decreased from 0.823 to 0.815, indicating that the four samples were not abnormal data. Finally, the remaining 206 samples of cherry tomatoes were subjected to subsequent modeling analysis.

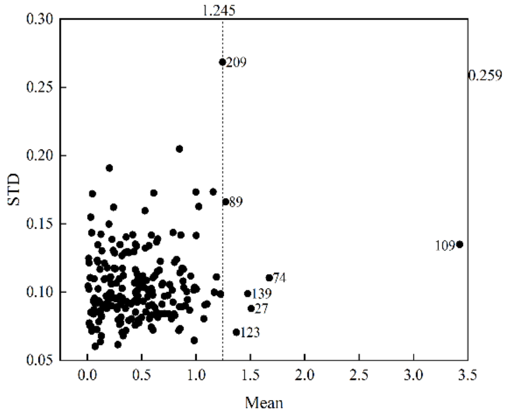

Fig. 2 Monte Carlo cross validation method to eliminate the results

After eliminating the abnormal spectral data, the remaining 206 samples of cherry tomatoes were modeled. The samples were divided into 154 calibration set samples and 52 prediction set samples according to the ratio of 3 : 1 by using the spectral-physicochemical value co-occurrence distance method [18]. The statistical results of soluble solids content of cherry tomatoes, see Table 1. The calibration set ranges from 3.9° Brix to 8.0° Brix, and the prediction set ranges from 4.3 ° Brix to 6.7° Brix. The calibration set covers the prediction set, and the sample distribution is uniform, indicating that the sample division is representative and can effectively support subsequent modeling analysis.

Table 1 The results of sample division of cherry tomatoes

Sample set	Sample size	Minimum	Maximum
Calibration set	154	3.9	8.0
Prediction set	52	4.3	6.7
Total sample	206	3.9	8.0

3.2. The original spectrum and pretreatment of cherry tomatoes

Due to the problems of baseline drift, light scattering, random noise and other irrelevant information in the original spectral data, the spectral data are biased, thus reducing the accuracy of the model. In order to improve the resolution of the spectrum and the stability of the model, it is necessary to preprocess the original spectral data to eliminate the influence of noise and irrelevant information. Multiplicative scatter correction (MSC), standard normal variable (SNV), savitzky-golay (SG) and other preprocessing methods were used. These methods can effectively eliminate background interference and improve the correlation between spectra and data [19]. Different preprocessing methods based on Partial Least Squares Regression (PLSR) model have different effects, see Table 2. Compared with the original spectral data, the RP after MSC pretreatment was 0.739, RMSEP was 0.281, and RPD was 1.873. Therefore, MSC is finally selected for subsequent modeling processing.

Table 2 Comparison of PLSR models based on different pretreatments

pretreatment method	RC	RMSEC	RP	RMSEP	RPD
Original	0.855	0.284	0.727	0.304	1.736
MSC	0.859	0.281	0.739	0.281	1.873
SNV	0.875	0.264	0.721	0.292	1.807
SG	0.852	0.287	0.728	0.303	1.740

3.3. Extracting characteristic wavelengths

Despite MSC preprocessing, there is still redundant information in the spectral data. In order to further improve the computational efficiency, Matlab R2022 a software is used to extract the characteristic wavelengths of the preprocessed spectral data to reduce the data dimension and improve the model performance. Competitive adaptive reweighting algorithm (CARS), uninformative variable elimination algorithm (UVE), successive projections algorithm (SPA) and SPA-UVE combination algorithm were used to extract characteristic wavelengths. The specific results are as follows : CARS algorithm extracted 24 characteristic wavelength variables, accounting for 6.2% of the full spectrum wavelength; the UVE algorithm extracted 142 characteristic wavelength variables, accounting for 36.8% of the full spectrum wavelength. The SPA algorithm extracted eight characteristic wavelength variables, accounting for 2.1% of the full spectrum wavelength; the SPA-UVE algorithm extracted 12 characteristic wavelength variables, accounting for 3.1% of the full spectrum wavelength.

In order to compare and analyze the optimal characteristic wavelengths, PLSR model was established based on the extracted characteristic wavelengths to evaluate the performance of different methods, see Table 3. Compared with single methods such as CARS, UVE and SPA, the model established by SPA-UVE combination method has higher prediction set correlation coefficient and residual prediction deviation, and lower root mean square error. In addition, the number of characteristic wavelengths extracted by SPA-UVE is only half of that of CARS. In summary, the SPA-UVE method performs well in screening characteristic wavelengths and optimizing model performance, which can effectively eliminate redundant information and improve computational efficiency and model accuracy, see Fig. 3. The results showed that the 12 characteristic wavelengths extracted based on SPA-UVE method could be used to establish the SSC quantitative detection model of cherry tomatoes.

Table 3 Characteristic wavelength results based on PLSR model

Extraction method	Number of wavelengths	RC	RMSEC	RP	RMSEP	RPD
CARS	24	0.822	0.304	0.739	0.284	1.854
SPA	8	0.755	0.369	0.655	0.322	1.635
UVE	142	0.896	0.241	0.721	0.288	1.827
SPA-UVE	12	0.865	0.275	0.750	0.272	1.935

3.4. Comparative Analysis of Different Models

The partial least squares regression (PLSR), principal component regression (PCR) and BP neural network were used to model and analyze the characteristic wavelengths extracted by SPA-UVE after the spectral data of cherry tomatoes were subjected to outlier removal, preprocessing and characteristic wavelength extraction. By comparing the performance of different models, the best prediction model is selected, see Table 4. The SSC prediction results of cherry tomatoes based on the PCR model showed that the correlation coefficient of the calibration set, the correlation coefficient of the prediction set and the residual prediction deviation were lower than the other three models, while the root mean square error of the calibration set and the root mean square error of the prediction set were higher than the other three models, indicating that the prediction effect of the PCR model was the worst. The prediction effect of PLSR model is better, but the prediction effect of MLR model is slightly better than that of PLSR model. In contrast, the prediction effect of BP neural network model is significantly better than that of other linear regression models due to its strong nonlinear fitting ability. In summary, the BP neural network model based on feature wavelength extraction can significantly improve the prediction ability and operation efficiency, and its model effect is better than other models, see Fig. 3.

Table 4 Comparative analysis of different detection models

modeling method	RC	RMSEC	RP	RMSEP	RPD
PLSR	0.865	0.275	0.750	0.272	1.935
PCR	0.801	0.333	0.726	0.291	1.808
BP	0.950	0.262	0.928	0.255	2.063

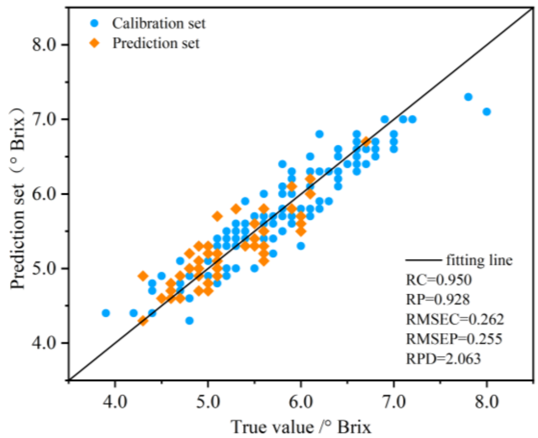

Fig. 3 Scatter plot of SPA-UVE-BP model

4. Conclusion

Hyperspectral imaging technology combined with spectral outlier elimination, preprocessing method and characteristic wavelength extraction method were used to analyze the quantitative model of soluble solids content of cherry tomatoes. The main conclusions are as follows :

(1) Hyperspectral imaging technology was used to obtain hyperspectral images, and the ROI region and original spectral data were manually obtained by software ENVI5.3. After eliminating abnormal samples, the original spectral data were compared with the results of PLSR model combined with different preprocessing methods such as MSC, SNV and SG. It was found that the spectral data modeling effect of MSC preprocessing method was the best, and the signal-to-noise ratio was significantly improved.

(2) After preprocessing, there is still redundant information in the spectral data, so the characteristic wavelengths are extracted and the prediction model is established. Through comparative analysis, it is found that the SPA-UVE method is used to extract 12 characteristic wavelengths, accounting for 3.1 % of the full wavelength, which significantly simplifies the calculation amount. At the same time, the correlation coefficient RC value and RP value of the model reached 0.865 and 0.750 respectively, the root mean square error RMSEC value and RMSEP value were 0.275 and 0.272 respectively, and the residual prediction deviation RPD value was 1.935, indicating that the method improved the stability of the model to a certain extent.

(3) The prediction model is established based on the characteristic wavelength. Among the various prediction models, the BP neural network model shows the best performance. The correlation coefficient RC value and RP value were 0.950 and 0.928 respectively, the root mean square error RMSEC value and RMSEP value were 0.262 and 0.255 respectively, and the residual prediction deviation RPD value was 2.603. The MCS-SPA-UVE-BP method can effectively improve the accuracy of the model and provide a theoretical basis for SSC online quality detection of cherry tomatoes.

Acknowledgements

Natural Science Project of Chongqing University of International Business and Economics (KYZK2024043).

References

[1] W.T. Zhang, W.Q. Zhao, X.X. Lu, et al. Effects of four logistics storage temperatures on the quality of cherry tomatoes, Food Industry Technology, Vol. 36 (2015) No.5, p.329-333.

[2] W.B. Li, H.Y. Sun, J.Y. Shen, et al. Study on the preparation process of pitaya-sacred cherry compound beverage, Food Industry, Vol. 36 (2015) No.5, p.109-112.

[3] Y.J. Pan, Y.X. Sun, Y.M. Li, et al. Response of soluble solids content of tomato fruit to phenotypic traits based on different varieties and parks, Shandong Agricultural Sciences, Vol. 54 (2022) No.10, p.30-36.

[4] Q.X. Tong, B. Zhang, L.F. Zhang. Frontier progress of hyperspectral remote sensing in China, Journal of Remote Sensing, Vol. 20 (2016) No.5, p.689-707.

[5] M.Y.He, Q.H. Meng, L. Tang, et al. Research progress of hyperspectral imaging technology in fruit quality detection, Journal of Fruit Science, Vol. 38 (2021) No.9, p.1590-1599.

[6] W.Q Huang, J.B. Li, L.P. Chen. Hyperspectral data were used to effectively predict the soluble solids content of apple, Spectroscopy and Spectral Analysis, Vol. 33 (2013) No.10, p.2843-2846.

[7] J.B. Li, X. Tian; W.Q. Huang, et al. Application of Long-Wave Near Infrared Hyperspectral Imaging for Measurement of Soluble Solid Content (SSC) in Pear, Food Analytical Methods, Vol. 9 (2016) No.11, p.3087-3098.

[8] L.J. Xu, M. Chen, Y.C. Wang, et al. Non-destructive determination of sugar content in kiwifruit by hyperspectral imaging, Spectroscopy and Spectral Analysis, Vol. 41 (2021) No.7, p.2188-2195.

[9] T. Li, M.L. Qu, Y.Q. Zhong. Hyperspectral imaging technology to identify Rhodiola crenulata and Rhodiola tetrapoda, West China Journal of Pharmacy, Vol. 36 (2021) No.6, p.644-648.

[10] S. Liu, H.Y.Yu, Z.J. Pu, et al. Research on the method of extracting characteristic wavelengths based on hyperspectral technology to invert soybean physiological information, Spectroscopy and Spectral Analysis, Vol. 40 (2020) No.11, p.3542-3548.

[11] L.J. Cheng, G.S. Liu, J.G. He, et al. Hyperspectral non-destructive detection of sucrose content in Lingwu long jujube, Food Science, Vol. 40 (2019) No.10, p.285-291.

[12] H.T. Tang, X.T. Meng, X.X. Su, et al. Hyperspectral prediction of different types of soil organic matter based on CARS algorithm, Journal of Agricultural Engineering, Vol. 37 (2021) No.2, p.105-113.

ISSN: 1813-4890

[13] R.C. Jiang, M.S. Gu, Q.H. Zhao, et al. Detection of pesticide residues in Chinese cabbage by hyperspectral and convolutional neural network, Spectroscopy and Spectral Analysis, Vol. 42 (2022) No.5, p.1385-1392.

[14] Q.L. Meng, J. Shang, X. Yang, et al. Hyperspectral nondestructive testing of apple hardness based on BP network, Packaging Engineering, Vol. 41 (2020) No.15, p.14-18.

[15] C. Wang, H. Yu, W.R. Yao, et al. Near-infrared hyperspectral imaging technology for screening the illegal addition of anti-inflammatory drug diclofenac sodium in health food , Spectroscopy and Spectral Analysis, Vol. 41 (2021) No.2, p.592-598.

[16] T.L. Liu, X.C. Zhu, X.Y. Bai, et al. Construction and accuracy comparison of hyperspectral estimation models for soil organic matter content, Wisdom Agriculture (Chinese and English), Vol. 2 (2020) No.3, p.129-138.

[17] M.W. Jiang, C.H. Wang, Q.H. Zhang. Study on wheat hardness determination based on CARS variable selection method, Journal of HeNan University of Technology (Natural Science Edition), Vol. 41 (2020) No.6, p.91-95+105.

[18] J. Sun, X. Zhou, X. Wu, et al. Identification of moisture content in tobacco plant leaves using outlier sample eliminating algorithms and hyperspectral data, Biochemical and Biophysical Research Communications, Vol. 471 (2016) No.1, p.226-232.

[19] X.Y. Yang, G.S. Liu, J.X. Ding, et al. Hyperspectral rapid detection of VC content in Lingwu long jujube, Spectroscopy and Spectral Analysis, Vol. 39 (2019) No.1, p.230-234.