



OPEN ACCESS

EDITED BY

Leizi Jiao,
Beijing Academy of Agriculture and
Forestry Sciences, China

REVIEWED BY

Leiqing Pan,
Nanjing Agricultural University, China
Seyed Ahmad Mireei,
Isfahan University of Technology, Iran

*CORRESPONDENCE

Guohui Han
✉ hghui2007@126.com

SPECIALTY SECTION

This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

RECEIVED 04 January 2023

ACCEPTED 06 March 2023

PUBLISHED 27 March 2023

CITATION

Li X, Wei Z, Peng F, Liu J and Han G (2023)
Non-destructive prediction and
visualization of anthocyanin
content in mulberry fruits using
hyperspectral imaging.
Front. Plant Sci. 14:1137198.
doi: 10.3389/fpls.2023.1137198

COPYRIGHT

© 2023 Li, Wei, Peng, Liu and Han. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Non-destructive prediction and visualization of anthocyanin content in mulberry fruits using hyperspectral imaging

Xunlan Li, Zhaoxin Wei, Fangfang Peng, Jianfei Liu
and Guohui Han*

Research Institute of Pomology, Chongqing Academy of Agricultural Sciences, Chongqing, China

Being rich in anthocyanin is one of the most important physiological traits of mulberry fruits. Efficient and non-destructive detection of anthocyanin content and distribution in fruits is important for the breeding, cultivation, harvesting and selling of them. This study aims at building a fast, non-destructive, and high-precision method for detecting and visualizing anthocyanin content of mulberry fruit by using hyperspectral imaging. Visible near-infrared hyperspectral images of the fruits of two varieties at three maturity stages are collected. Successive projections algorithm (SPA), competitive adaptive reweighted sampling (CARS) and stacked auto-encoder (SAE) are used to reduce the dimension of high-dimensional hyperspectral data. The least squares-support vector machine and extreme learning machine (ELM) are used to build models for predicting the anthocyanin content of mulberry fruit. And genetic algorithm (GA) is used to optimize the major parameters of models. The results show that the higher the anthocyanin content is, the lower the spectral reflectance is. 15, 7 and 13 characteristic variables are extracted by applying CARS, SPA and SAE respectively. The model based on SAE-GA-ELM achieved the best performance with R^2 of 0.97 and the RMSE of 0.22 mg/g in both the training set and testing set, and it is applied to retrieve the distribution of anthocyanin content in mulberry fruits. By applying SAE-GA-ELM model to each pixel of the mulberry fruit images, distribution maps are created to visualize the changes in anthocyanin content of mulberry fruits at three maturity stages. The overall results indicate that hyperspectral imaging, in combination with SAE-GA-ELM, can help achieve rapid, non-destructive and high-precision detection and visualization of anthocyanin content in mulberry fruits.

KEYWORDS

hyperspectral imaging, mulberry fruit, anthocyanin content, SAE, ELM

1 Introduction

Mulberry (*Morus L.*) is widely planted around the world. Tender, juicy and delicious mulberry fruits have long been used as traditional medicine as well as edible fruits in countries such as China, India and Turkey (Jan et al., 2021). Modern researches show that black and red mulberry fruits are rich in anthocyanins, which, with the properties of antioxidant, anti-inflammatory and chemical protection, play a positive role in reducing the risk of cardiovascular diseases and cancers (Chen et al., 2006; Krishna et al., 2018). Anthocyanins are considered to be one of the most important indicators for mulberry fruits of good quality by researchers and consumers.

Anthocyanin contents are usually determined by adopting wet chemical methods, such as spectrophotometry (Jiang and Nie, 2015) and high-performance liquid chromatography (Zou et al., 2012). The samples need to be ground and extracted with the use of chemical reagents such as ethanol or acetone. These methods are destructive and will produce chemical residues. And only a small number of samples can be analyzed at a time. It is difficult to detect anthocyanin content in mulberry fruits on a large scale by applying the existing time-consuming and inefficient detecting methods. For efficient agricultural management and production, it is necessary to find a reliable, fast and non-destructive method for anthocyanin content detection.

Hyperspectral imaging (HSI) can obtain the spectral data of each pixel in the sample image simultaneously. This is of potential value in non-destructive detection of uneven distribution of quality indicators. There are reports about visualizing anthocyanin contents of purple sweet potato (Liu et al., 2017), lychee pericarp (Yang et al., 2015), and grape (Chen et al., 2015) by using HSI. The research by Huang et al. (2017) has shown that 400-1000nm and 900-1700nm HSI, in combination with least squares support vector machine (LS-SVM), has great potential in evaluating total anthocyanin content and antioxidant activity of mulberry fruits. This is the only study on determining anthocyanin of mulberry by using HSI. And further research endeavors to visualize anthocyanin content of mulberry fruit have not been reported yet.

The variable selection is an essential step for modeling. From previous researches, variable selection methods, such as interval partial least square, successive projections algorithm (SPA) and competitive adaptive reweighted sampling (CARS) are often used to reduce the number of input variables before modeling (Zhu et al., 2017; Silva and Melo-Pinto, 2021). When using these variable selection methods, the average spectrum of all pixels in the hyperspectral image is applied, while efficient big data analysis of each pixel spectrum is ignored. Depth feature extraction and dimension reduction can be conducted by using the stacked auto-encoder (SAE), a nonlinear unsupervised neural network, which is capable of effectively analyzing the spectral data of all pixels of the hyperspectral image and then select variables (Yu et al., 2018). In terms of modeling, the LS-SVM has been shown to be of good potential in non-destructive detection. Research reports show that the extreme learning machine (ELM), a single hidden layer

feedforward neural network model, is able to achieve similar or much better performance at a much faster learning speed than traditional LS-SVM (Huang et al., 2011; Zheng et al., 2014).

This study is meant for developing a rapid, non-destructive, high-precision method to detect and visualize the anthocyanin content of mulberry fruit. The main research objects are as follows: (1) analyzing the differences in anthocyanin content and corresponding spectral data between two mulberry varieties at different maturity stages; (2) reducing the dimension of high-dimensional spectral data by using SPA, CARS and SAE, and selecting the most effective feature variables; (3) using LS-SVM and ELM to build the models for predicting mulberry anthocyanin and selecting the best prediction model so as to achieve rapid, non-destructive and high-precision prediction of the anthocyanin content of mulberry fruit; (4) mapping distribution of anthocyanin content in mulberry fruit.

2 Materials and methods

2.1 Materials

The sampled varieties, Dashi (*Morus nigra L.*) and Siji (*Morus nigra L.*) were collected from the mulberry resource conservation nursery of Chongqing Academy of Agricultural Sciences on April 23, 2020. Disease-free fruits at three maturity stages (S1: red maturity, S2: red to purple maturity and S3: full maturity) were randomly picked, then stored in ice boxes. They were brought back to the laboratory for hyperspectral image collection and anthocyanin content detection (Figure 1). Six fruits at the same maturity stage were randomly selected as one sample for anthocyanin content detection. A total of 180 samples were obtained, and the numbers of Dashi and Siji were 90 respectively. The samples were randomly divided into the training set and the testing set at a ratio of 7:3.

2.2 Collection and calibration hyperspectral images

The hyperspectral imaging system was used to collect hyperspectral images of mulberry fruits (Figure 2). The hyperspectral imaging system consists of a spectrograph (ImSpector V10E, SPECIM, Finland), an EMCCD camera (DL-604E, Andor Technology plc., N. Ireland), two halogen light sources (150 W/21 V halogen lamp, Illuminator Technologies, Inc, USA), an electric mobile platform and controller (SC30021A, Zolix, China), and a laptop. The wavelength range of the spectrum collected was 305-1 090 nm. The two light sources were at an angle of 45° with the mobile platform respectively. The camera exposure time was 60 ms. The spectral resolution was 2.8 nm. The platform moving speed was 1.87 mm/s. The distance between the objective lens and the platform was 40 cm. After preheating for half an hour by the light source, the mulberry fruits were placed on the black cardboard for hyperspectral image collection.

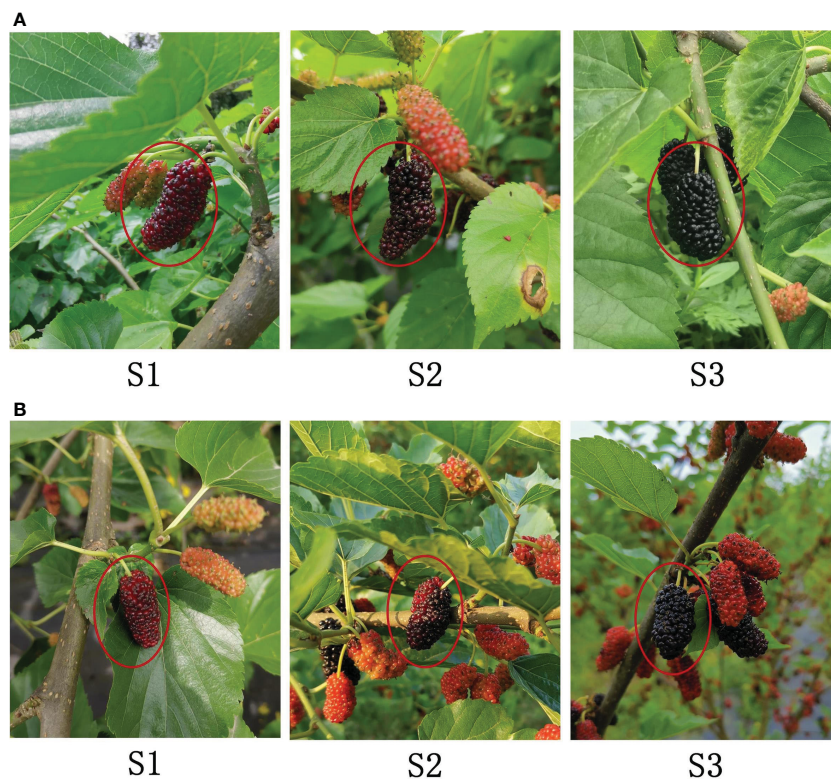


FIGURE 1
Fruit images of Dashi (A) and Siji (B) at three maturity stages: (S1) red maturity; (S2) red to purple maturity; (S3) full maturity.

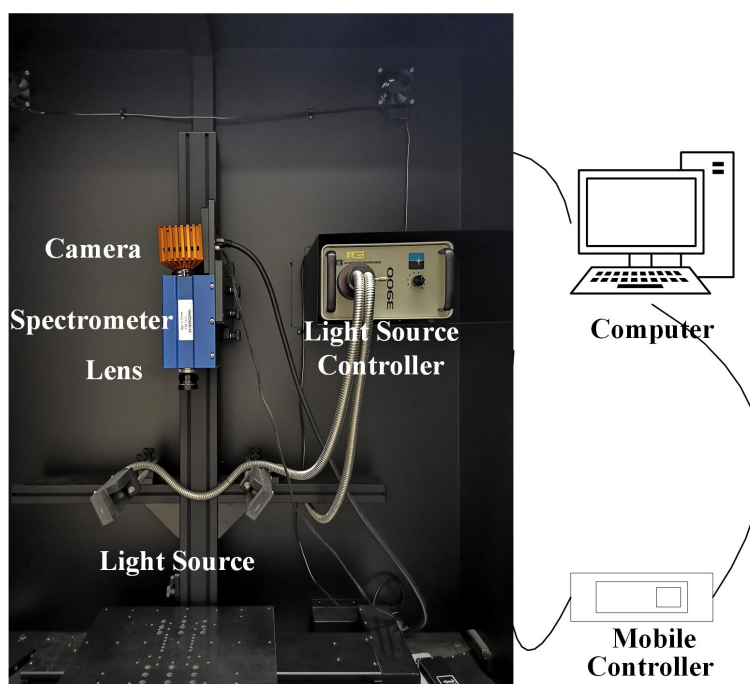


FIGURE 2
The hyperspectral imaging system.

The collected hyperspectral images need to be calibrated so as to avoid the effect caused by uneven light source intensity distribution and dark current during the image collecting process. Under the same conditions as the sample images were collected, the white reference image W was obtained by scanning the standard white reference panels. The dark reference image D was obtained by scanning with the lens covered. And the image calibration was completed on the basis of formula (1),

$$R_{\lambda} = \frac{I_{\lambda} - D_{\lambda}}{W_{\lambda} - D_{\lambda}} \quad (1)$$

where R_{λ} is the calibrated image, I_{λ} is the raw image, W_{λ} is the white reference image, and D_{λ} is the dark reference image.

2.3 Anthocyanin content extraction

Anthocyanin content was detected by pH-differential spectrophotometry (Lee et al., 2005). 0.5 ± 0.001 g of grinded fresh mulberry fruits was added to 10 ml of acidified ethanol (95% ethanol and 1% concentrated hydrochloric acid, the volume ratio of ethanol to hydrochloric acid was 60:40) for 1 h ultrasound extraction and 2 min centrifugation at 8000 r·min⁻¹. 1 ml of supernatant was taken and the volume was fixed to 25 ml by adding buffer solutions of pH 1.0 and pH 4.5 respectively. The absorbance was measured at 520 nm and 700 nm after letting it stand for 15 min with an ultra-violet-visible spectrophotometer (UV-6000PC ShanghaiMetash. Co. Ltd, China). The anthocyanin content was calculated by the formula (2) and (3).

$$A = (A_{520nm} - A_{700nm}) \text{ at PH } 1.0 - (A_{520nm} - A_{700nm}) \text{ at PH } 4.5 \quad (2)$$

$$\begin{aligned} & \text{Total anthocyanin content(mg/g)} \\ & = (A * MW * DF * V) / (\epsilon * 1 * M) \end{aligned} \quad (3)$$

Where is the absorbance, A_{520nm} and A_{700nm} are the absorbance at the 520 nm and 700 nm respectively. MW (molecular weight) = 449.2 g/mol for cyanidin-3-glucoside (cyd-3-glu). DF (Dilution factor) = 25. V is the original volume of 10 ml. The molar extinction coefficient $\epsilon=26900$. M is the weight of the sample.

2.4 Region of interest and spectral data extraction

In this study, the whole fruit with the fruit stalk removed was treated as the region of interest (ROI). The whole mulberry fruit and the collection background plate were segmented at 800 nm, with the reflectance of 0.2 as the minimum value. The petiole was removed from the whole fruit at 550 nm and 670 nm, with the difference value of 0.04 as the maximum value. Then the ROI was obtained by conducting mask processing. The average spectrum of ROI at each wavelength was calculated for subsequent SPA and CARS feature wavelength extraction. To create a data set for deep learning, 400

pixels (20 * 20) corresponding to spectral data were randomly selected from the ROI of each sample, totaling 72,000, for SAE training.

2.5 Spectral data processing

2.5.1 Spectral data pretreatment

Owing to the existence of strong noises in the beginning and ending bands of the raw spectral data, spectral data within the range of 450-1050 nm, a total of 379 variables were selected for subsequent analysis. In this study, standard normal variate transform (SNV) was used to preprocess the spectral data, to eliminate the scattering caused by uneven particle distribution and different particle sizes, and the influence of optical path change on the spectral data.

2.5.2 Feature extraction

Successive projections algorithm(SPA), Competitive adaptive reweighted sampling(CARS) and Stacked auto-encoder (SAE) were respectively used in this study to extract spectral data features for the purpose of reducing the number of input variables, improving model efficiency, eliminating redundant information of spectral data, and improving the prediction accuracy of the model.

Successive projections algorithm (SPA) is a forward variable selection algorithm. By this method, the cycle of forward is conducted with a wavelength initially selected and the projection value of the remaining wavelength calculated. Then the projection vector is combined with the wavelength corresponding to the maximum projection value until the cycle ends. The minimum variable group can be effectively obtained by calculating the band projection value, thus minimizing the collinearity between variables (Araújo et al., 2001).

Competitive adaptive reweighted sampling (CARS) is a method based on Monte Carlo sampling and the PLS regression coefficient. By this method, characteristic variables are primarily screened out by using the PLS regression coefficient in combination with the exponential decline function. Then the initially selected characteristic variables are competitively screened out by using adaptive reweighted sampling. And the final characteristic variables are screened out from the wavelength combinations according to the cross-validation root mean square error. The detailed algorithm of CARS can be found in reference (Li et al., 2009). In this study, the number of CARS samples was set to 50, and the ten-fold cross-validation method was used.

Stacked auto-encoder (SAE) is a deep neural network consisting of multilayer auto-encoders (AE), by which better feature extraction is obtained with the hidden layers added to the simple auto-encoders. AE consists of encoders and decoders. The input layers map the input data to the hidden layers through the activation function to obtain the encoding features, which is called encoding. Through the same steps, the encoding features are mapped to the output layers by using the activation function to obtain the decoding features, which is called decoding. In terms of SAE, the decoding features of the previous AE are used as the input of

the next hidden layer of AE, and code and decode the next layer of AE. By analogy, these hidden input layers are connected to form SAE (Xu et al., 2022).

2.5.3 Model construction and evaluation

Least squares support vector machine (LS-SVM) is a machine learning algorithm based on SVM, boasting good generalization ability and nonlinear regression processing ability (Suykens and Vandewalle, 1999). The fitting ability of LS-SVM mainly depends on the selection of kernel parameters (C and γ). Kernel parameter C affects the fitting accuracy and generalization ability of the model, and kernel parameter γ directly determines the calculation amount and efficiency of the model.

Extreme learning machine (ELM) is a feedforward neural network with a single hidden layer, which has a fast learning ability and strong nonlinear approximation ability (Huang et al., 2006). Compared with traditional neural network learning algorithms, such as back propagation neural network, ELM presents the advantages of strong generalization ability and fast calculation speed (Ye et al., 2022). Over-fitting is liable to occur, since the weight and offset of ELM are randomly determined.

Genetic algorithm (GA) is a search algorithm for obtaining the global optimal solution based on the biological evolution mechanism of “survival of the fittest” (Mirjalili, 2019). In this study, GA is used to optimize the important parameters of the RBF kernel function and the offset and weight of ELM. In this case, the value ranges of kernel parameters (C and γ) were set to 0.01-100, the population size was set to 20, and the number of maximum evolution times was set to 200. When GA was used to optimize ELM, the population size was set to 20, the maximum number of evolutions was set to 300, and the number of neurons in the hidden layer of ELM was set to 90.

The training set determination coefficient (R^2_c), testing set determination coefficient (R^2_p), training set root mean square error (RMSEC), and testing set root mean square error (RMSEP) were used as indicators to evaluate the performance of models. The closer to 1 the determination coefficient (R^2) is, the better the model

fitting effect is. And the smaller RMSEC and RMSEP are, the higher the precision of the model is.

The hyperspectral image calibration in this study was completed by the software of the hyperspectral image acquisition system. ROI segmentation, spectral data extraction and processing were completed by using MATLAB 2022a, with SPA, CARS, GA, SAE, and LS-SVM realized by using SPA_GUI, Lib PLS1.98, GATBX, Deep Learning toolbox, and LS-SVMlab v1.8 toolbox. The overall flow is shown in Figure 3.

3 Results and analysis

3.1 Anthocyanin content and spectral characteristics of mulberry fruits

The anthocyanin contents of two mulberry varieties at three maturity stages were analyzed and measured, and the mean anthocyanin content and corresponding spectral reflectance of two mulberry varieties at different maturity stages were calculated (Figure 4). It was shown in Figure 4A that the higher the maturity of mulberry fruits was, the higher the anthocyanin content was, which followed the description of the report of Saracoglu (Saracoglu, 2018). The anthocyanin content of Dashi was higher than that of Siji at the same maturity stage. Anthocyanins are the main reasons why mulberry has red and purple (Li et al., 2020). From the analysis of the spectrum reflection curve of mulberry fruit, it can be seen that the spectral reflection value in the range from 500 to 700 nm was very low. According to Qin and Lu (Qin and Lu, 2008), the maximum absorbance of anthocyanin pigments is about 535 nm. However, the difference between the mulberry fruits of different maturity was not obvious at 535 nm in Figure 4B. This may be because the black substances have strong absorption in the visible light area, and the reflectance value is not attributed to a single compound, the spectra are the sum of the major mulberry fruit composition spectra (Cozzolino et al., 2004). A small reflective valley could be seen near 680 nm in red maturity fruits, which is related to the existence of

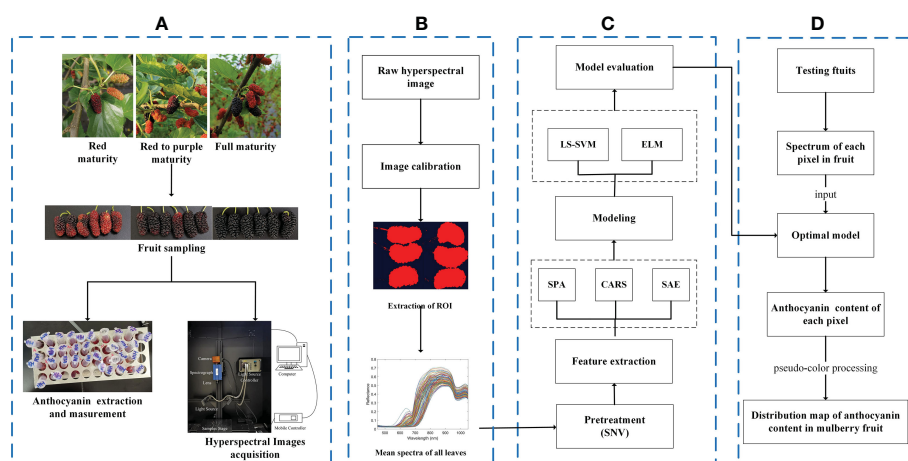


FIGURE 3

Overall flow chart. (A) Acquisition of data; (B) hyperspectral image processing; (C) analysis of spectral data; (D) visualization of anthocyanin content.

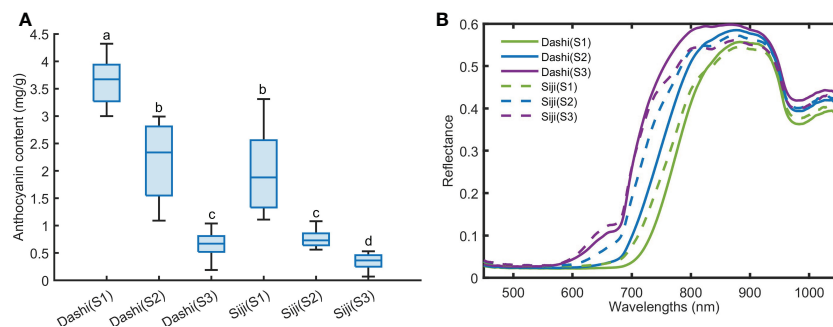


FIGURE 4

The anthocyanin content (A) and average spectra (B) of mulberry fruit at three maturity stages. Values with the same letter (i.e. a, b, c or d) are not significantly different ($p < 0.05$).

chlorophyll. The spectral reflectance was lower with the increase of maturity and anthocyanin content within the range of 590–800 nm. The two varieties showed obvious absorption peaks near 970 nm and 840 nm. This is speculated to be related to water and sugar absorption (ElMasry et al., 2008; Zheng et al., 2008). The differences in spectral characteristics of the mulberry fruits above show that hyperspectral imaging has the potential to distinguish the mulberry fruits of different anthocyanin contents.

3.2 The results of feature extraction

When hyperspectral imaging is used to detect the anthocyanin contents of mulberry fruits, the redundant information is often eliminated and the amount of calculation is compressed by screening out the characteristic wavelengths to improve the accuracy and robustness of the diagnostic models. In this paper, SPA, CARS and SAE were used to extract feature variables from the 379 variables.

SPA was used to screen characteristic wavelengths from spectral data of SNV pretreatment in 450–1,050 nm region, and the results were shown in Figure 5. It can be seen from Figure 5A that when the number of characteristic wavelengths increased from 1 to 7, the value of RMSE decreased in a ladder shape and then leveled off. And 7 characteristic wavelengths at 684.88, 703.98, 747.15, 798.58, 842.15, 923.11 and 962.05 nm were obtained.

The process of screening wavelengths by using CARS was shown in Figure 6. With the increase in sampling times, the number of selected wavelengths decreased gradually at the speed from fast to slow. This reflected the two stages, preliminary screening and fine screening, of using CARS to screen out key variables. With the increase in sampling times, the root mean squares error of cross-validation (RMSECV) value gradually decreased. And the RMSECV value obtained was the lowest when the 31st sampling was conducted. This is an indication that some irrelevant variables are removed during the sampling process. After the 31st sampling, the RMSECV value presented a stepwise progression, indicating the removal of some key information. Therefore, the wavelengths obtained at the 31st sampling were

the characteristic wavelengths. Fifteen characteristic wavelengths, 450.08, 451.59, 601.16, 703.98, 707.17, 708.77, 743.95, 795.36, 796.97, 798.58, 832.45, 963.67, 965.29, 1038.29 and 1046.39 nm, were screened out by using CARS.

Based on the analysis of the characteristic wavelengths, the positions and numbers of characteristic wavelengths screened out by using SPA and CARS were found to be different. And the wavelength positions are concentrated within the ranges of 703–835 nm and 963–1046 nm.

The feature variables of SAE screening are shown in Figure 7. When it comes to SAE, it is not necessarily the case that the more hidden layers are, the better the effect is. In this study, 379–300–150–h–150–300–379 was set to be the basic network. h denotes the number of neurons in the last coding layer, and it is also the number of feature variables extracted. Based on experience and many previous attempts, sigmoid was set as the activation function, iterate was set to 40 times, the batch size was set to 200, the initial learning rate was set to 0.001, and h was set to 13. From the results shown in Figure 7A, the reconstructed spectral reflectance curve is highly coincident with the original spectral curve, indicating that the original spectral data can be perfectly reconstructed by using SAE. The last coding layer was extracted as the spectral feature variables (Figure 7B). It can be seen that the corresponding values of the 13 feature variables of samples at different maturity stages are obviously different.

3.3 The results of modeling

All wavelengths and feature variables extracted by using SPA, CARS and SAE were used as the model inputs. Regression models of mulberry anthocyanin contents were built based on GA-LS-SVM and GA-ELM respectively. And the regression results were evaluated (Figure 8). Models were constructed by using the two non-linear regression methods that achieved good performance, R^2 values of the training sets and those of the testing sets of GA-LS-SVM and GA-ELM models built on the basis of full wavelengths and variables extracted by using SPA, CARS and SAE were greater than 0.90, RMSE was less than 0.38 mg/g. The models based on

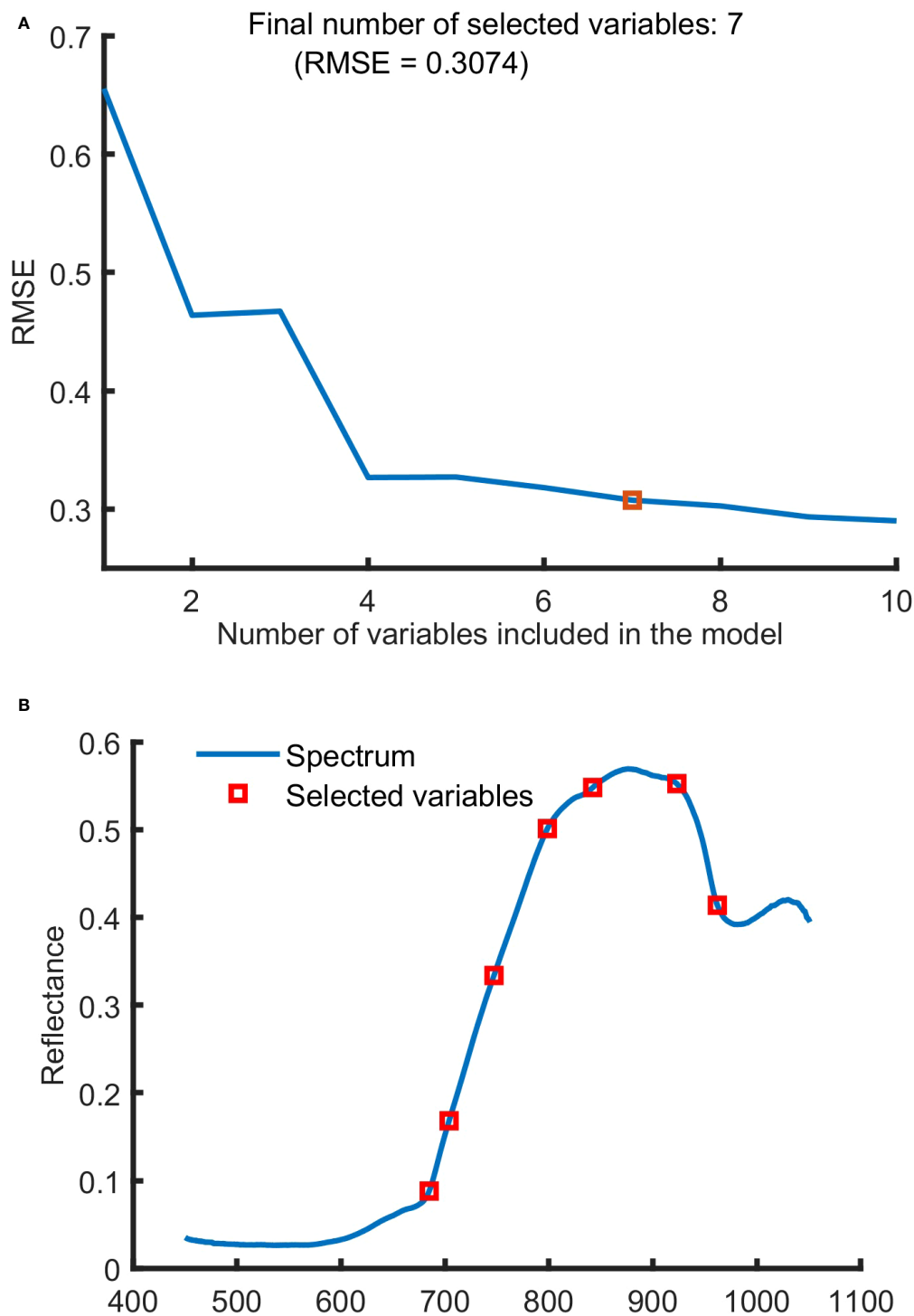


FIGURE 5

The characteristic wavelengths selected by SPA. (A) Variation of RMSE with the number of variables, (B) the selected wavelengths.

variables selected by SPA, CARS and SAE achieved better performances than those based on full-band spectral data, indicating that SPA, CARS and SAE can reduce the redundancy of model input variables and help improve the accuracy of the model. Many researches show that ELM has the advantages of fast

learning speed and good generalization ability (Wong et al., 2013; Huang et al., 2014). In this study, The SAE-GA-ELM models, requiring only 13 input variables, has achieved the best predictive performance, with the values of R^2_c and R^2_p reaching 0.97, and RMSEC and RMSEP being only 0.22 mg/g, obtained.

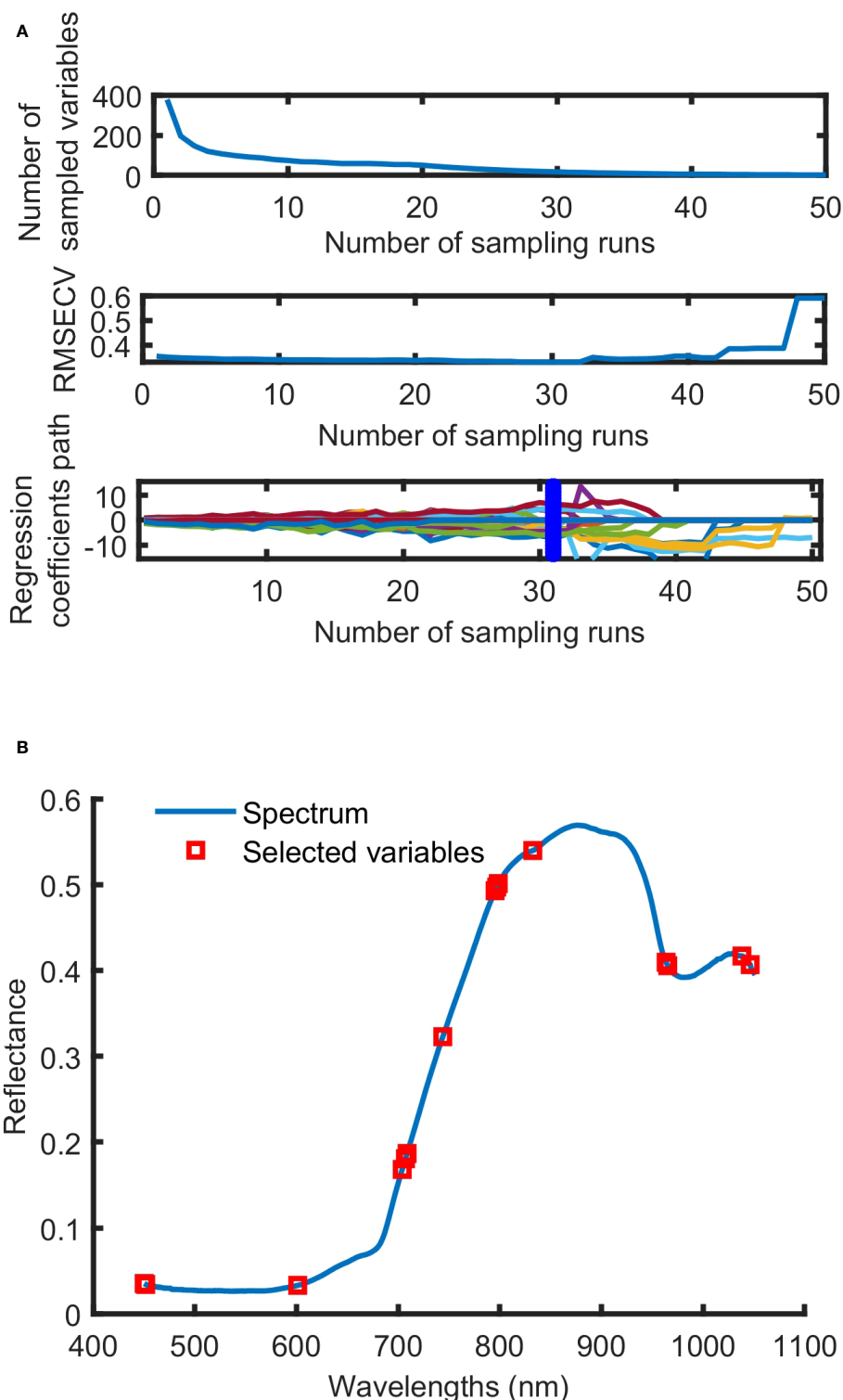


FIGURE 6 The process (A) and result (B) of characteristic wavelength selection by CARS.

3.4 Visualization of anthocyanin content

The visualization of anthocyanin content distribution in mulberry fruits is needed for more intuitively observing the changes in anthocyanin contents of mulberry fruits at different

maturity levels. One of the advantages of hyperspectral imaging is that spectral data of each pixel can be obtained by using hyperspectral imaging. This makes it possible for the prediction about each pixel to be made, thus helping create distribution prediction maps. The visualization can be achieved with the average spectra applied for

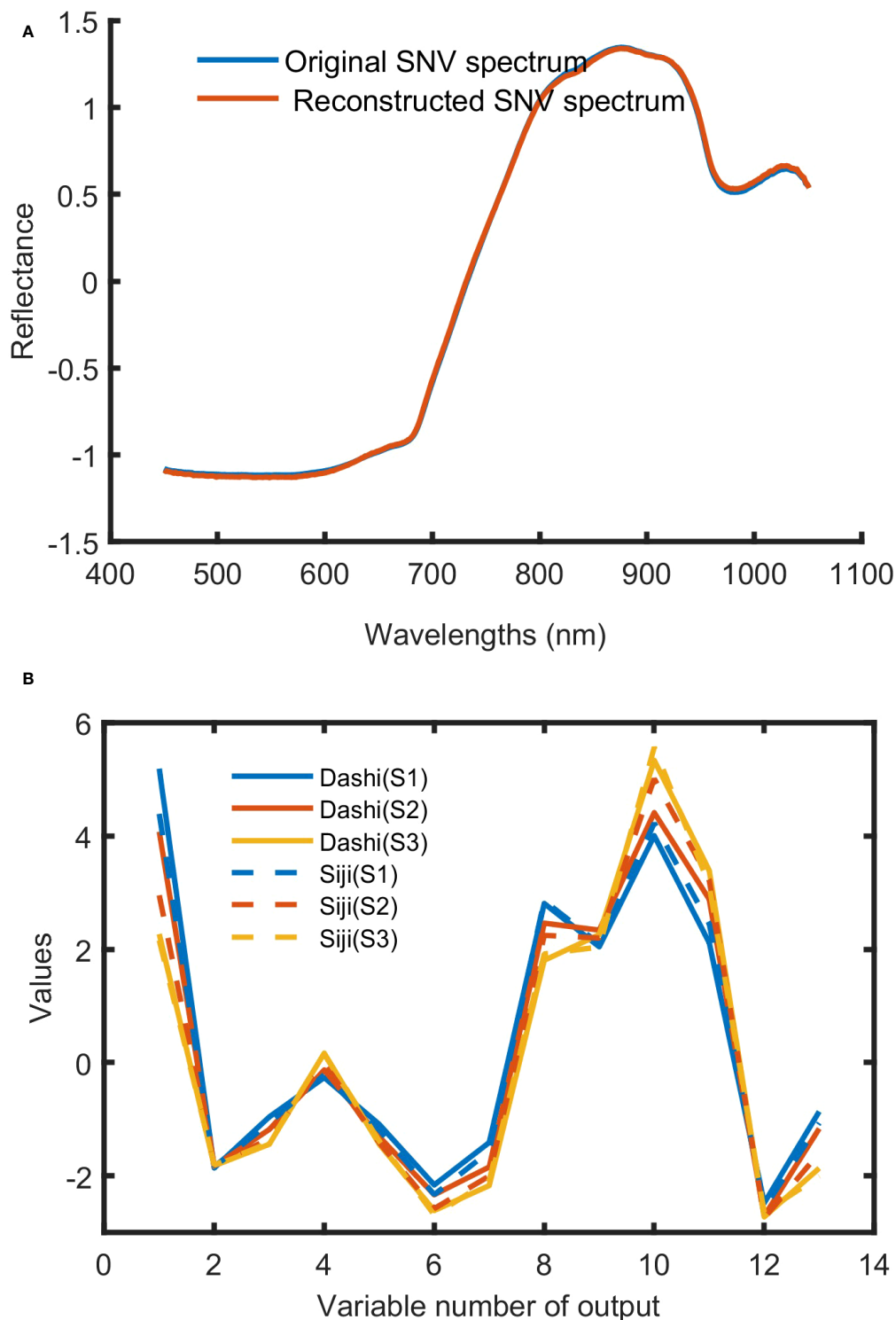


FIGURE 7

The training results of SAE. (A) Original SNV spectrum and reconstructed SNV spectrum; (B) deep spectral features of Dashi and Siji.

modeling and all of the single-pixel spectra in the hyperspectral image used for the best prediction model (Sun et al., 2019; Xiao et al., 2020). In this study, SAE-GA-ELM, the best model for anthocyanin content detection, was applied to visualize anthocyanin content distribution. All the single-pixel spectrum was processed by the same treatment used in the modeling. Figure 9 shows the visualization maps of

eighteen samples representing different maturity levels of two varieties. we can see from Figure 9 that the higher the maturity level of mulberry fruits is, the higher the anthocyanin content is, and that the anthocyanin content of Dashi is higher than that of Siji at the same maturity stage, which is consistent with the results shown in Figure 4A. It can be seen that the distribution of anthocyanin content

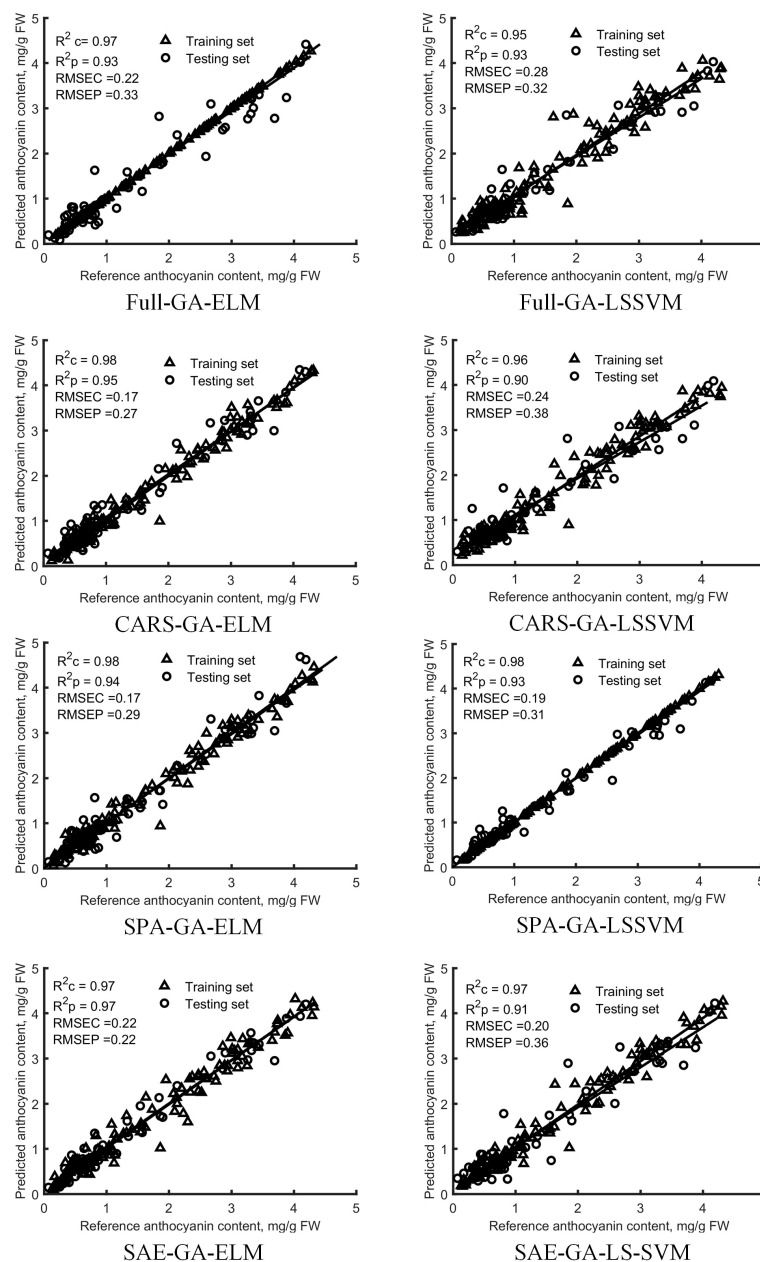


FIGURE 8

Diagnosis results of anthocyanin content in the training set and testing set by GA-ELM and GA-LS-SVM models based on all-band and feature variables.

of the mulberry fruits at the red maturity stage is not consistent with that of the content of the mulberry fruits at the red to purple maturity stage, which is speculated to be the result of the uneven distribution of such anthocyanin synthesis regulators as sugars and hormones in the fruits (Aramwit et al., 2010; Mo et al., 2022).

4 Conclusions

In this study, with Dashi and Siji mulberry varieties selected as research objects, and SPA, CARS and deep learning methods

SAE used to screen out feature variables, models for predicting anthocyanin content in mulberry fruits are built based on GA-LS-SVM and GA-ELM. The SAE-GA-ELM has achieved the best performance with R^2_c and R^2_p reaching the value of 0.97 under the condition of RMSEC and RMSEP being only 0.22 mg/g. By applying this best model to each pixel of the mulberry fruit images, distribution maps are created for visualizing the changes in anthocyanin content of mulberry fruits at three maturity stages. The results indicate that the hyperspectral imaging, in combination with SAE-GA-ELM could realize the fast, non-destructive, and high-precision detection of

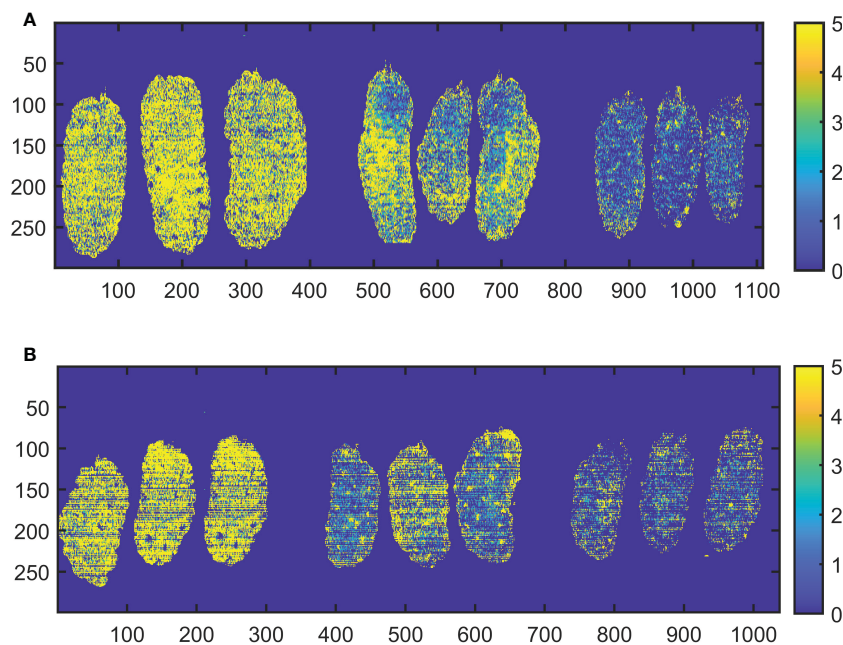


FIGURE 9
Visualization of anthocyanin content in mulberry fruits of Dashi (A) and Siji (B) at three maturity stages.

anthocyanin content of mulberry fruits, which means a new reference for rapid and nondestructive evaluation of physiological traits for the breeding, cultivation, harvesting and selling of the fruits.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XL completed data collection, model construction, and paper writing. ZW and FP helped to collect data and provided comments and suggestions to improve the manuscript. JL edited the manuscript. GH directed the paper revision and provided the main idea. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Youth Innovation Team Project of Chongqing Academy of Agricultural Sciences (Grant No. NKY-

2019QC08), Performance Incentive and Guidance Special Project of Chongqing Research Institute (Grant No. cqaas2021jxjl08) and Excellent Germplasm Innovation Project of Chongqing (Grant No. NKY-2021AB019).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aramwit, P., Bang, N., and Srichana, T. (2010). The properties and stability of anthocyanins in mulberry fruits. *Food Res Int* 43, 4, 1093–1097. doi: 10.1016/j.foodres.2010.01.022

Araújo, M. C. U., Saldanha, T. C. B., Galvao, R. K. H., Yoneyama, T., Chame, H. C., Visani, V. J. C., et al. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent

- analysis. *Chemom. Intell. Lab. Syst. 57* (2), 65–73. doi: 10.1016/S0169-7439(01)00119-8
- Chen, P.-N., Chu, S.-C., Chiou, H.-L., Kuo, W.-H., Chiang, C.-L., and Hsieh, Y.-S. (2006). Mulberry anthocyanins, cyanidin 3-rutinoside and cyanidin 3-glucoside, exhibited an inhibitory effect on the migration and invasion of a human lung cancer cell line. *Cancer Lett.* 235, 2, 248–259. doi: 10.1016/j.canlet.2005.04.033
- Chen, S., Zhang, F., Ning, J., Liu, X., Zhang, Z., and Yang, S. (2015). Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chem.* 172, 788–793. doi: 10.1016/j.foodchem.2014.09.119
- Cozzolino, D., Esler, M., Dambergs, R., Cynkar, W., Boehm, D., Francis, I., et al. (2004). Prediction of colour and pH in grapes using a diode array spectrophotometer (400–1100 nm). *J Near Infrared Spectrosc* 12, 2, 105–111. doi: 10.1255/jnirs.414
- ElMasry, G., Wang, N., Vigneault, C., Qiao, J., and ElSayed, A. (2008). Early detection of apple bruises on different background colors using hyperspectral imaging. *LWT - Food Sci. Technol* 41, 2, 337–345. doi: 10.1016/j.lwt.2007.02.022
- Huang, M., Wang, Q., Zhang, M., and Zhu, Q. (2014). Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology. *J. Food Eng.* 128, 24–30. doi: 10.1016/j.jfoodeng.2013.12.008
- Huang, G.-B., Zhou, H., Ding, X., and Zhang, R. (2011). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42, 2, 513–529. doi: 10.1109/TSMCB.2011.2168604
- Huang, L., Zhou, Y., Meng, L., Wu, D., and He, Y. (2017). Comparison of different CCD detectors and chemometrics for predicting total anthocyanin content and antioxidant activity of mulberry fruit using visible and near infrared hyperspectral imaging technique. *Food Chem.* 224, 1–10. doi: 10.1016/j.foodchem.2016.12.037
- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing* 70, 1–3, 489–501. doi: 10.1016/j.neucom.2005.12.126
- Jan, B., Parveen, R., Zahiruddin, S., Khan, M. U., Mohapatra, S., and Ahmad, S. (2021). Nutritional constituents of mulberry and their potential applications in food and pharmaceuticals: A review. *Saudi J. Biol. Sci.* 28, 7, 3909–3921. doi: 10.1016/j.sjbs.2021.03.056
- Jiang, Y., and Nie, W.-J. (2015). Chemical properties in fruits of mulberry species from the xinjiang province of China. *Food Chem.* 174, 460–466. doi: 10.1016/j.foodchem.2014.11.083
- Krishna, P. G. A., Sivakumar, T. R., Jin, C., Li, S.-H., Weng, Y.-J., Yin, J., et al. (2018). Antioxidant and hemolysis protective effects of polyphenol-rich extract from mulberry fruits. *Pharmacognosy Magazine* 14, 53, 103. doi: 10.4103/pm.pm_491_16
- Lee, J., Durst, R. W., Wrolstad, R. E., and Kupina, C. (2005). Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: Collaborative study. *J. AOAC Int.* 88, 5, 1269–1278. doi: 10.1093/jaoac/88.5.1269
- Li, H., Liang, Y., Xu, Q., and Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta* 648, 1, 77–84. doi: 10.1016/j.aca.2009.06.046
- Li, H., Yang, Z., Zeng, Q., Wang, S., Luo, Y., Huang, Y., et al. (2020). Abnormal expression of bHLH3 disrupts a flavonoid homeostasis network, causing differences in pigment composition among mulberry fruits. *Horticult. Res.* 7, 83. doi: 10.1038/s41438-020-0302-8
- Liu, Y., Sun, Y., Xie, A., Yu, H., Yin, Y., Li, X., et al. (2017). Potential of hyperspectral imaging for rapid prediction of anthocyanin content of purple-fleshed sweet potato slices during drying process. *Food Anal. Methods* 10, 3836–3846. doi: 10.1007/s12161-017-0950-y
- Mirjalili, S. (2019). “Genetic algorithm,” in *Evolutionary algorithms and neural networks* (Springer), 43–55.
- Mo, R., Han, G., Zhu, Z., Essemine, J., Dong, Z., Li, Y., et al. (2022). The Ethylene Response Factor ERF5 Regulates Anthocyanin Biosynthesis in ‘Zijin’ Mulberry Fruits by Interacting with MYBA and F3H Genes. *Int. J. Mol. Sci.* doi: 10.3390/ijms23147615
- Qin, J., and Lu, R. (2008). Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest Biol. Technol.* 49, 3, 355–365. doi: 10.1016/j.postharvbio.2008.03.010
- Saracoglu, O. (2018). Phytochemical accumulation of anthocyanin rich mulberry (*Morus laevigata*) during ripening. *Food Measure.* 12, 3, 2158–2163. doi: 10.1007/s11694-018-9831-3
- Silva, R., and Melo-Pinto, P. (2021). A review of different dimensionality reduction methods for the prediction of sugar content from hyperspectral images of wine grape berries. *Appl. Soft Comput.* 113, 107889. doi: 10.1016/j.asoc.2021.107889
- Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., Wang, P. J. C., et al. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. *Comput. Electron. Agric.* 160, 153–159. doi: 10.1016/j.compag.2019.03.004
- Suykens, J. A., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 3, 293–300. doi: 10.1023/A:1018628609742
- Wong, K. I., Wong, P. K., Cheung, C. S., and Vong, C. (2013). Modeling and optimization of biodiesel engine performance using advanced machine learning methods. *LWT - Food Sci. Technol.* 55, 519–528. doi: 10.1016/j.lwt.2007.02.022
- Xiao, Q., Bai, X., and He, Y. (2020). Rapid screen of the color and water content of fresh-cut potato tuber slices using hyperspectral imaging coupled with multivariate analysis. *Foods.* 9, 1, 94. doi: 10.3390/foods9010094
- Xu, M., Sun, J., Yao, K., Cai, Q., Shen, J., Tian, Y., et al. (2022). Developing deep learning based regression approaches for prediction of firmness and pH in kyoho grape using Vis/NIR hyperspectral imaging. *Infrared Phys. Technol.* 120, 104003. doi: 10.1016/j.infrared.2021.104003
- Yang, Y.-C., Sun, D.-W., Pu, H., Wang, N.-N., and Zhu, Z.-W. (2015). Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. *Postharvest Biol. Technol.* 103, 55–65. doi: 10.1016/j.postharvbio.2015.02.008
- Ye, A., Zhou, X., and Miao, F. (2022). Innovative hyperspectral image classification approach using optimized CNN and ELM. *Electronics.* 11, 5, 775. doi: 10.3390/electronics11050775
- Yu, X., Lu, H., and Wu, D. (2018). Development of deep learning method for predicting firmness and soluble solid content of postharvest korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. *Postharvest Biol. Technol.* 141, 39–49. doi: 10.1016/j.postharvbio.2018.02.013
- Zheng, W., Fu, X., and Ying, Y. (2014). Spectroscopy-based food classification with extreme learning machine. *Chemom. Intell. Lab. Syst.* 139, 42–47. doi: 10.1016/j.chemolab.2014.09.015
- Zheng, Y., Lai, X., Bruun, S. W., Ipsen, H., Larsen, J. N., Lföwenstein, H., et al. (2008). Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy. *J. Pharm. Biomed. Anal.* 46, 3, 592–596. doi: 10.1016/j.jpba.2007.11.011
- Zhu, H., Chu, B., Fan, Y., Tao, X., Yin, W., and He, Y. (2017). Hyperspectral imaging for predicting the internal quality of kiwifruits based on variable selection algorithms and chemometric models. *Sci. Rep.* 7, 1, 1–13. doi: 10.1038/s41598-017-08509-6
- Zou, T., Wang, D., Guo, H., Zhu, Y., Luo, X., Liu, F., et al. (2012). Optimization of microwave-assisted extraction of anthocyanins from mulberry and identification of anthocyanins in extract using HPLC-ESI-MS. *J. Food Sci.* 77, 1, C46–C50. doi: 10.1111/j.1750-3841.2011.02447