



## OPEN Nondestructive quality assessment and maturity classification of loquats based on hyperspectral imaging

Shunan Feng<sup>1</sup>, Jing Shang<sup>1,2✉</sup>, Tao Tan<sup>1</sup>, Qingchun Wen<sup>1</sup> & Qinglong Meng<sup>1,2</sup>

The traditional method for assessing the quality and maturity of loquats has disadvantages such as destructive sampling and being time-consuming. In this study, hyperspectral imaging technology was used to nondestructively predict and visualise the colour, firmness, and soluble solids content (SSC) of loquats and discriminate maturity. On comparison of the performance of different feature variables selection methods and the calibration models, the results indicated that the multiple linear regression (MLR) models combined with the competitive adaptive reweighting algorithm (CARS) yielded the best prediction performance for loquat quality. Particularly, CARS-MLR models with optimal prediction performance were obtained for the colour ( $R^2_p = 0.96$ , RMSEP = 0.45, RPD = 5.38), firmness ( $R^2_p = 0.87$ , RMSEP = 0.23, RPD = 2.81), and SSC ( $R^2_p = 0.84$ , RMSEP = 0.51, RPD = 2.54). Subsequently, distribution maps of the colour, firmness, and SSC of loquats were obtained based on the optimal CARS-MLR models combined with pseudo-colour technology. Finally, on comparison of different classification models for loquat maturity, the partial least square discrimination analysis model demonstrated the best performance, with classification accuracies of 98.19% and 97.99% for calibration and prediction sets, respectively. This study demonstrated that the hyperspectral imaging technique is promising for loquat quality assessment and maturity classification.

Loquat (*Eriobotrya japonica* Lindl.) is an evergreen fruit tree of the Rosaceae family, and its fruit is used as a dual-purpose medicine and food that has been cultivated in China for more than 2000 years<sup>1</sup>. It is used for clearing the pharynx, moistening the lungs, alleviating cough, and lowering phlegm<sup>2</sup>. The ripening pattern of loquats is similar to that of climacteric fruits. If harvested very early, it will have hard flesh and a bland flavour. As loquats have an active postharvest physiological metabolism, they are susceptible to water and nutrient loss and rot if harvested late<sup>3,4</sup>. Fruit quality has a direct impact on its commercial value. Colour, firmness, and soluble solid content (SSC) are important characteristics of loquats and are key parameters for evaluating their taste and maturity<sup>5</sup>. Therefore, the detection of postharvest loquats is crucial.

However, traditional determination methods have the disadvantage of destructive sampling and are not suitable for online detection. In recent years, hyperspectral imaging (HSI) techniques, which combine two-dimensional image information with one-dimensional spectral information, have been widely used to evaluate fruit quality and maturity. HSI has been used to determine multiple indicators (SSC, firmness, etc.) of fruits, including plums<sup>6</sup>, sweet cherries<sup>7</sup>, pears<sup>8</sup>, peaches<sup>9</sup>, and melons<sup>10</sup>. Extensive studies have been conducted to predict quality and ripeness of fruits. Wei et al.<sup>11</sup> used HSI to classify ripeness and predict the firmness of persimmons. Munera et al.<sup>12</sup> used the index of internal quality and maturity to assess the internal physicochemical attributes and sensory perception of 'Big Top' and 'Magique' nectarines. The ratio of total soluble solids (TSS) to titratable acidity (TA) was used as a pineapple ripeness index to analyse the effects of transmittance short-wavelength near-infrared spectroscopy and reflectance near-infrared hyperspectral imaging on the prediction of pineapple ripeness using the same procedure and model, respectively<sup>13</sup>. Benelli et al.<sup>14</sup> investigated the potential of using HSI directly in the field through proximal measurements under natural light conditions to predict the harvest time of 'Sangiovese' red grape. They split grape samples into two classes based on the reference value of SSC and established models to predict SSC and recognise the maturity stages, respectively. Zhang et al.<sup>15</sup> combined HSI

<sup>1</sup>Food and Pharmaceutical Engineering Institute, Guiyang University, Guiyang 550005, China. <sup>2</sup>Research Center of Nondestructive Testing for Agricultural Products of Guizhou Province, Guiyang 550005, China. ✉email: shji0124@163.com

with support vector machine (SVM) to evaluate strawberry ripeness. The results indicated that the SVM model performed the best, with classification accuracy of over 85%.

Furthermore, considerable attention has been given to visualise quality of fruits. Teerachaichayut et al.<sup>16</sup> applied HSI to perform nondestructive detection and visual analysis of TSS and TA and calculated TSS/TA as a measure of the maturity index in intact limes. The predictive distribution maps of TSS, TA and TSS/TA were generated by inputting the feature bands of each pixel into optimal models. Li et al.<sup>17</sup> realised the visualization of SSC and pH based on a colour scale in cherry fruits. Chu et al.<sup>18</sup> created the visualization maps for banana quality parameters using machine learning algorithm. The results indicated that the hyperspectral imaging is a useful tool to assess the quality of bananas. Additionally, due to the complexities involved in processing hyperspectral data and the inherent limitations of computer hardware capabilities, it is essential to select feature wavelengths instead of using full wavelengths to achieve similar precision in the operation. Zhang et al.<sup>19</sup> established partial least squares regression (PLSR) model for predicting caffeine content of coffee beans based on full wavelengths and feature wavelengths using HSI, respectively. The overall results indicated that, similar to PLSR models built on full wavelengths, all PLSR models based on feature wavelengths demonstrated robust performance. Li et al.<sup>20</sup> developed rapid and non-destructive models for detecting anthocyanin content in mulberry fruit using HSI, based on both full bands and feature variables, respectively. The results indicated that the models based on feature variables demonstrated superior performance compared to those using full bands. Sharma et al.<sup>21</sup> applied HSI to classify the ripening stages and predict the dry matter content of durian pulp. A comparison was conducted between the models using full wavelengths and feature wavelengths. The results indicated that the model based on full wavelengths showed comparable performance to the model based on feature wavelengths in maturity classification, while the model based on feature wavelengths achieved better results in predicting dry matter. Most of the above studies have confirmed the feasibility of fruit quality prediction and maturity classification using hyperspectral imaging, and it is crucial to choose feature variables for modelling during data processing. Nevertheless, little research has reported the utility of HSI technology to predict and visualise the colour, firmness, and SSC of loquats and discriminate maturity.

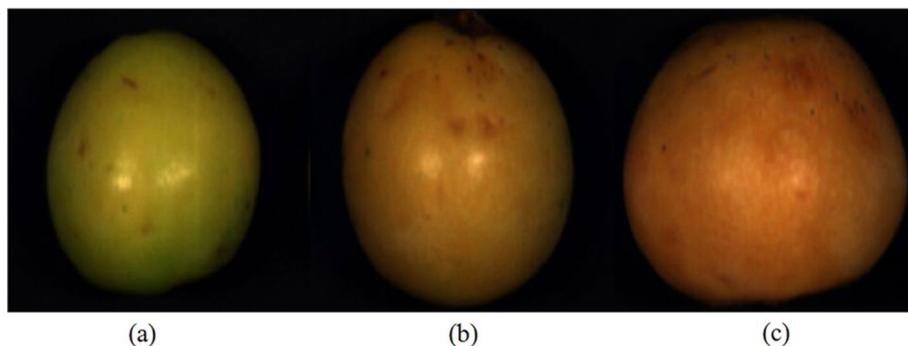
This study aimed to explore the feasibility of determining and visualising the colour, firmness, and SSC of loquats and discriminating maturity based on HSI. The specific objectives of this study were to (1) compare the performance of different feature variables selection methods including competitive adaptive reweighting algorithm (CARS), genetic algorithms (GA), and successive projections algorithm (SPA); (2) establish and compare calibration models for predicting quality including PLSR, principal components regression (PCR), multiple linear regression (MLR), extreme learning machine (ELM), and back-propagation neural network (BP); (3) visualise the spatial distribution of these quality parameters in loquats; and (4) develop recognition models for discriminating maturity including partial least square discrimination analysis (PLS-DA), simplified K-nearest neighbour (SKNN), and SVM models.

## Methods

**Sample preparation.** A total of 649 loquats (transverse diameter: 35–55 mm) without bruises were harvested from the commercial orchards (Loquat Green Planting Demonstration Garden of Kaiyang County) located in Guizhou Province, China, on 7 June 2022. The collectors took the permit, which was required at the time, and obtained the owner's permission. The selection of loquats was guided by experienced local growers based on visual observation of the external colour, ranging from dark green to dark orange. The samples were transported to the laboratory on the same day as the sampling, at a temperature of  $23 \pm 2$  °C. Before the experiment, the loquat surfaces were wiped and numbered. All methods were performed in accordance with the relevant guidelines and legislation.

Deng et al.<sup>22</sup> found a significant or highly significant correlation between the colour  $a^*$  value and loquat quality. On this basis, the 649 samples were divided into three maturity stages (stage I: 177, stage II: 331, and stage III: 141) based on the colour  $a^*$  value. Stage I represented colour  $a^*$  values less than 8.33, stage II covered colour  $a^*$  values between 8.33 and 15.41, and stage III encompassed colour  $a^*$  values greater than 15.41. The images of the three maturity stages are shown in Fig. 1.

To generate adequate variability and broaden the predictive range of colour, firmness and SSC, the samples were divided into four groups for experimentation. Among these samples, 140 were used for predicting loquat



**Figure 1.** Images of loquat maturity stage I (a), maturity stage II (b), and maturity stage III (c).

colour (stage I: 47, stage II: 63, and stage III: 30), another set of 140 for predicting loquat firmness (stage I: 45, stage II: 53, and stage III: 42), and 120 for predicting loquat SSC (stage I: 25, stage II: 65, and stage III: 30). The remaining 249 samples were used to classify loquat maturity (stage I: 60, stage II: 150, and stage III: 39).

**Hyperspectral image acquisition and correction.** Hyperspectral images of loquat samples were captured using a hyperspectral imaging system (GaiaFieldF-V10, Jiangsu Dualix Spectral Imaging Technology Co., Ltd). A schematic of the system is shown in Fig. 2. It primarily included a hyperspectral imaging spectrograph (Inspector V10, Spectral Imaging Ltd., Oulu, Finland), CCD camera (Imperx IPX-2 M30, Pixels: 696 × 1313), zoom lens (HSIA-OL23, Focal length: 23 mm), four 200 W halogen light sources (HSIA-LS-T-200 W), transportation plate, dark room (HSIA-T400-IMS), and computer with image acquisition software. The distance from the sample to the lens was 400 mm, and the exposure time of the spectral camera was 12.6 ms. The spectral resolution was 3.5 nm, and the spatial resolution was 0.2 mm/pixels. The spectrograph obtained spectral images covering a wavelength range from 390 to 1030 nm with 256 spectral bands.

When acquiring hyperspectral images each time, four loquats were placed regularly on the sample stage above the displacement platform according to their number<sup>23</sup>. To eliminate the effects of noise and dark current in the CCD camera, the acquired original images were used to correct the black and white images. The correction was performed based on Eq. (1). After the hyperspectral images were corrected, the spectral data from the entire sample area of loquat were extracted by using ENVI 5.4 (ITT Visual Information Solutions, Boulder, CO).

$$I = \frac{I_0 - B}{W - B} \quad (1)$$

where,  $I$  is the calibrated image,  $I_0$  is the original image,  $B$  is the dark reference image, and  $W$  is the white reference image.

**Reference values for measurement of quality parameters.** Following hyperspectral image acquisition, conventional destructive methods were used to measure the reference values for the colour, firmness, and SSC of the loquats. For the determination of colour, a spectrophotometer (Ci7800) was used to measure the colour parameters ( $L^*$ ,  $a^*$ , and  $b^*$  values), which were evaluated using colour  $e$  value calculated based on Eq. (2)<sup>24</sup>. The formula emphasizes the colour contrast in the  $a^*$  and  $b^*$  directions, enabling a more effective comparison of colour characteristics among different loquats.

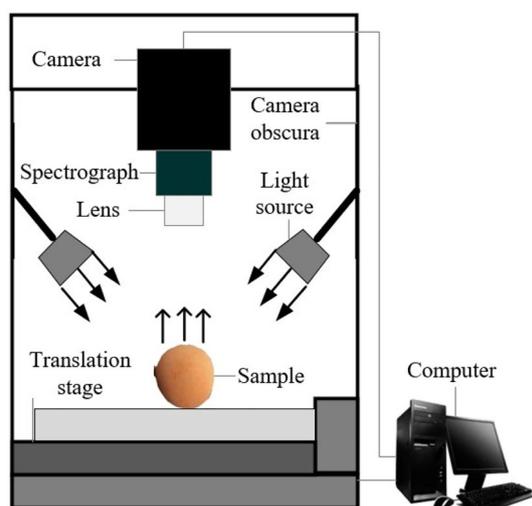
$$e = \frac{1000a^*}{(L^* \times b^*)} \quad (2)$$

Firmness was measured using a texture analyser (TA.XT.plus) with a cylindrical puncture probe of 2 mm at a test speed of 3 mm/s. The measurement required the peeling of the loquat around the equator.

The measurements of the SSC were carried out using a digital refractometer (PAL- $\alpha$ ) in the range 0–85%.

**Data preprocessing and feature variables selection.** To improve the accuracy and stability of the model, spectral pre-processing aims to eliminate instrument noise, scattering, and baseline shifts. Standard normal variation (SNV) was used to preprocess the original spectra; it can reduce the effects of surface scattering and light path alterations on diffuse reflection<sup>25</sup>.

Additionally, the hyperspectral data were characterised by redundancy and multicollinearity. To reduce the number of modelling calculations and improve the operational efficiency of the model, the CARS, GA, and SPA



**Figure 2.** Schematic diagram of the hyperspectral imaging acquisition system.

were applied to select the feature variables. Variable points with large absolute values of the regression coefficients in the PLSR model established by CARS are selected as the new correction set, and the subset with the smallest root mean square error was obtained after several cycles<sup>26</sup>. The GA simulates the mechanisms of natural selection and genetics and iteratively performs operations to generate a subset of variables<sup>27</sup>. Unlike GA, SPA is a forward feature variables selection method that minimises the collinearity between feature vectors<sup>28</sup>.

**Model building and evaluation.** Two commonly used tools for multivariate data analysis, PLSR and PCR models, were developed by combining chemical concentration and preprocessed data, respectively<sup>29</sup>. Subsequently, three feature variables models, namely, MLR, BP, and ELM models, were established based on the selected feature variables. MLR is used to characterise the relationship between spectral data and mass parameters using a linear fitting equation<sup>30</sup>. BP, which is one of the most typical multilayer forward network, is a local optimisation method based on gradient descent<sup>31</sup>. ELM is a high-efficiency single hidden layer feed-forward neural network that can map nonlinear relationships between input and output values<sup>32</sup>.

To evaluate the performances of the prediction models, the determination coefficient of the calibration set ( $R^2_C$ ), root mean square error of the calibration set (RMSEC), the determination coefficient of the prediction set ( $R^2_P$ ), root mean square error of the prediction set (RMSEP), and residual predictive deviation (RPD) were calculated. Generally, a model that performs well has higher values of  $R^2_C$ ,  $R^2_P$ , and RPD and lower values of RMSEC and RMSEP. The model performs poorly when the RPD is lower than 1.5, whereas an RPD between 1.5 and 1.99 indicates that the model performs moderately well. An RPD between 2 and 2.5 indicates that the model performs well, and the model performs excellently when the RPD is higher than 2.5<sup>33</sup>.

$$R^2_C = 1 - \frac{\sum_{i=1}^{n_c} [y_{act}(i) - y_{cal}(i)]^2}{\sum_{i=1}^{n_c} [y_{act}(i) - y_{mean}(i)]^2} \quad (3)$$

$$R^2_P = 1 - \frac{\sum_{i=1}^{n_p} [y_{act}(i) - y_{pre}(i)]^2}{\sum_{i=1}^{n_p} [y_{act}(i) - y_{mean}(i)]^2} \quad (4)$$

$$RMSEC = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} [y_{act}(i) - y_{cal}(i)]^2} \quad (5)$$

$$RMSEP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} [y_{act}(i) - y_{pre}(i)]^2} \quad (6)$$

$$RPD = \frac{SD}{RMSEP} \quad (7)$$

where  $n_c$  and  $n_p$  denote the number of samples in the calibration and prediction sets;  $y_{act}$  and  $y_{mean}$  denote the measured and mean values;  $y_{cal}$  and  $y_{pre}$  denote the predicted values in the calibration and prediction sets, respectively; and SD denotes the standard deviation of the measured values in the prediction set.

## Results and discussion

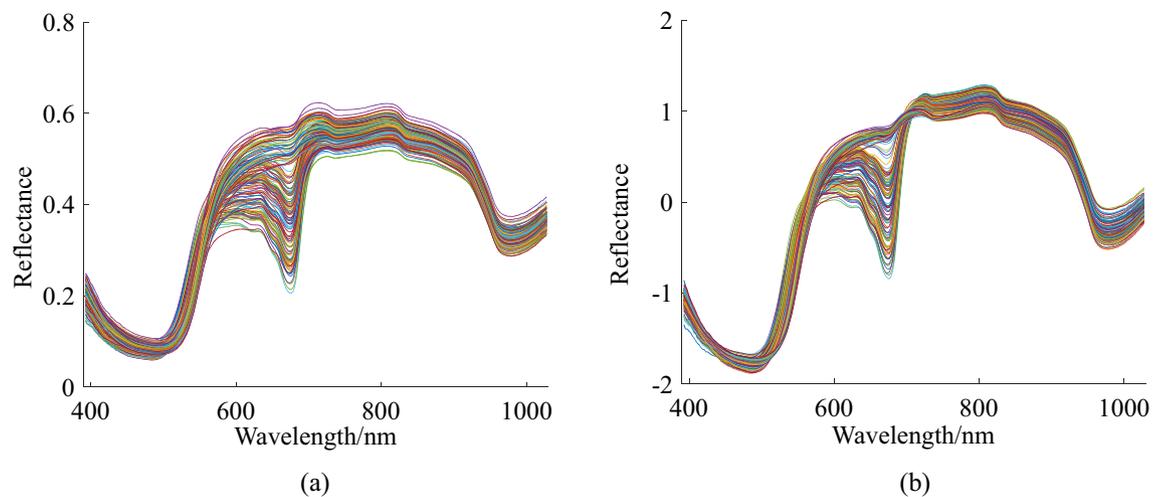
**Spectral characteristics.** The original and preprocessed (SNV) spectral curves are shown in Fig. 3. The spectra of the loquat samples showed the same tendency but with different reflection intensities. The preprocessed curves (Fig. 3b) were generally smoother than the original spectral curves (Fig. 3a), indicating a significant pretreatment effect. A clear absorption peak near 675 nm occurred, which correlated with the absorption of chlorophyll<sup>34</sup>. The more obvious absorption peak at approximately 980 nm may be attributed to the O–H chemical bond, which is related to water<sup>35</sup>.

**Statistical analysis of chemical concentration values.** Figure 4 shows colour  $e$  value, firmness, and SSC of loquat samples at three maturity stages; the data are shown as mean  $\pm$  SD. There is an increasing trend for colour  $e$  value and SSC of loquats and a downward trend for firmness with maturity stages.

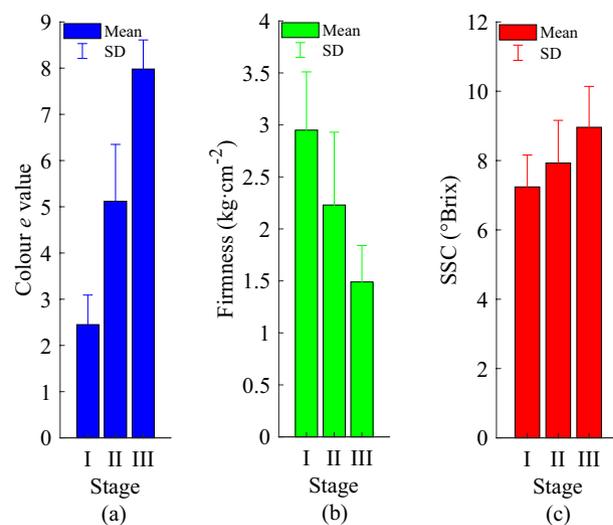
The SPXY algorithm<sup>36</sup> was used to divide all the samples into calibration and prediction sets. The ratio of the calibration set to the prediction set was 3:1. Table 1 presents the calibration and prediction sets statistics for colour  $e$  value, firmness, and SSC. The range of values of the calibration set was wider than that of the prediction set, which indicated that the results for the calibration and prediction sets were reasonable and the selected modelling samples were highly representative.

**Modelling based on full spectra.** PLSR and PCR models were built up to assess the parameters of loquat quality using spectra preprocessed with SNV. The prediction results for the PLSR and PCR models are listed in Table 2.

The prediction performances of the PLSR models for colour  $e$  value ( $R^2_P = 0.96$ , RMSEP = 0.49, RPD = 4.97), firmness ( $R^2_P = 0.82$ , RMSEP = 0.27, RPD = 2.39), and SSC ( $R^2_P = 0.72$ , RMSEP = 0.67, RPD = 1.92) were better



**Figure 3.** (a) Reflectance curves of raw spectra; (b) pretreated spectra of loquat samples.



**Figure 4.** (a) Colour *e* value of loquat samples at different maturity stages; (b) firmness of loquat samples at different maturity stages; (c) SSC of loquat samples at different maturity stages.

Quality parameter	Calibration set				Prediction set			
	Num	Range	Mean	SD	Num	Range	Mean	SD
Color <i>e</i> value	105	1.17–9.08	4.60	2.13	35	1.27–8.88	5.55	2.43
Firmness/(kg/cm <sup>2</sup> )	105	0.85–4.06	2.39	0.81	35	0.93–3.26	1.80	0.63
SSC/(°Brix)	90	5.50–10.50	7.79	1.21	30	6.60–10.50	8.81	1.29

**Table 1.** Statistics of colour *e* value, firmness and SSC of loquats.

than those of the PCR models. This may be because the PLSR method has the advantage of considering both matrices,  $x$  (spectral matrix) and  $y$  (concentration matrix).

**Feature variables selection.** *Feature variables selected by CARS.* When extracting the feature variables using CARS, the number of Monte Carlo sampling runs was set to 50, and the cross-validation of the group amount was set to five. The optimal feature variables was selected based on the minimal RMSECV, which corresponded to the sampling runs at 27, 23, and 28 for colour *e* value, firmness, and SSC, respectively. The selected variables were 20, 29, and 18 for colour *e* value, firmness, and SSC of loquats, respectively. Table 3 presents the detailed variables selected by CARS.

Model	Quality parameter	IVs	Calibration set		Prediction set		
			$R^2_C$	RMSEC	$R^2_P$	RMSEP	RPD
PLSR	Colour $e$ value	12	0.96	0.42	0.96	0.49	4.97
	Firmness	12	0.86	0.30	0.82	0.27	2.39
	SSC	17	0.87	0.43	0.72	0.67	1.92
PCR	Colour $e$ value	22	0.96	0.43	0.95	0.51	4.76
	Firmness	16	0.81	0.35	0.80	0.28	2.29
	SSC	27	0.81	0.53	0.65	0.75	1.72

**Table 2.** Performance of PLSR and PCR models for colour  $e$  value, firmness, and SSC.

Index	Methods	Number	Variables/nm
Colour $e$ value	CARS	20	397, 401, 404, 413, 418, 422, 432, 443, 498, 526, 541, 553, 621, 623, 641, 717, 750, 972, 974, 993
	GA	29	394, 397, 399, 401, 411, 413, 415, 420, 422, 425, 505, 507, 515, 517, 519, 522, 524, 526, 529, 536, 538, 541, 543, 546, 548, 551, 553, 555, 558
	SPA	3	551, 392, 577
Firmness	CARS	29	394, 399, 406, 408, 411, 413, 415, 555, 616, 619, 626, 641, 643, 678, 690, 693, 705, 707, 727, 730, 733, 743, 783, 785, 796, 974, 987, 1022, 1024
	GA	22	546, 548, 551, 553, 555, 558, 560, 675, 678, 680, 685, 688, 690, 693, 698, 700, 703, 801, 1006, 1019, 1022, 1024
	SPA	27	399, 406, 413, 415, 418, 422, 425, 429, 436, 439, 441, 448, 458, 493, 510, 526, 546, 577, 648, 678, 695, 745, 875, 914, 945, 1019, 1022
SSC	CARS	18	418, 425, 439, 488, 505, 695, 705, 720, 824, 883, 885, 888, 914, 919, 940, 961, 990, 1019
	GA	23	418, 420, 462, 465, 467, 474, 476, 479, 484, 507, 522, 524, 901, 904, 906, 917, 919, 982, 985, 987, 1016, 1019, 1022
	SPA	16	406, 408, 418, 425, 432, 443, 446, 450, 453, 481, 775, 867, 914, 937, 979, 1027

**Table 3.** Optimal variables for colour  $e$  value, firmness, and SSC selected by CARS, GA, and SPA.

*Feature variables selected by GA.* The GA has a strong global optimisation ability. When extracting the feature variables using the GA, the population size, crossover probability, mutation probability, and the number of iterations were set to 30, 0.5, 0.01, and 100, respectively. The optimal combination of variables with the minimal RMSECV was viewed as the key variable to determine the parameters in the loquat. The number of corresponding feature variables set with the minimal RMSECV was 29, 22, and 23 for colour  $e$  value, firmness, and SSC in loquats, respectively. Table 3 lists the variables selected by the GA.

*Feature variables selected by the SPA.* For SPA, the number of variables was selected based on the minimum root mean square error (RMSE). Firstly, the RMSE decreases rapidly owing to the elimination of unimportant redundant variables. When the redundant information variable set of spectral information was minimal, the number of corresponding feature variables sets was 3, 27, and 16 for colour  $e$  value, firmness, and SSC in the loquat, respectively. Table 3 presents the detailed variables selected by the SPA.

**Modelling based on feature variables.** The MLR, ELM, and BP models for predicting loquat quality were established based on these feature variables. The performances of the models are listed in Table 4.

As presented in Table 4, for colour  $e$  value, CARS was superior to the GA in setting the proper parameters. The models built based on the feature variables extracted by SPA exhibited the worst performance, with  $R^2_C$  lower than  $R^2_P$ , which might be caused by under-fitting. The number of feature variables selected using CARS was 20, which represented 7.81% of the full spectrum. Compared with other models built based on feature variables selected by CARS, the MLR model built based on the feature variables extracted by CARS obtained a higher RPD and lower RMSEC and RMSEP. Compared with the models based on full wavelengths shown in Table 2, the prediction accuracy of MLR, ELM, and BP models based on feature variables selected by CARS and GA was enhanced. Especially, the CARS-MLR model achieved the best performance ( $R^2_C = 0.97$ , RMSEC = 0.39,  $R^2_P = 0.96$ , RMSEP = 0.45, and RPD = 5.38) in predicting colour  $e$  value.

For firmness, the CARS appeared to be superior to the SPA and GA regarding setting appropriate parameters. The number of feature variables selected by CARS was 29, which was 11.33% of the full spectrum. Compared with other models built based on the feature variables selected by CARS, the MLR model built based on the feature variables extracted by CARS obtained higher  $R^2_C$ ,  $R^2_P$ , and RPD and lower RMSEC and RMSEP. Compared with the models based on full wavelengths shown in Table 2, the prediction accuracy of MLR, ELM, and BP models based on the feature variables selected by CARS and SPA was improved. Especially, the CARS-MLR model achieved the best performance ( $R^2_C = 0.90$ , RMSEC = 0.26,  $R^2_P = 0.87$ , RMSEP = 0.23, and RPD = 2.81) in predicting firmness.

For SSC, CARS appeared to be superior to the GA through the set of proper parameters. The accuracies of the SPA-ELM and SPA-BP models were lower than those of the CARS-ELM and CARS-BP models. The SPA-MLR model indicated the worst performance of  $R^2_C$  lower than  $R^2_P$ , which might be caused by under-fitting. The number of feature variables selected by CARS was 18, which was 7.03% of the full spectrum. Compared with other

Quality parameter	Model	Method	Band Num	Calibration set		Prediction set		
				$R^2_C$	RMSEC	$R^2_P$	RMSEP	RPD
Colour <i>e</i> value	MLR	CARS	20	0.97	0.39	0.96	0.45	5.38
		GA	29	0.97	0.40	0.96	0.49	5.01
		SPA	3	0.92	0.62	0.98	0.37	6.60
	ELM	CARS	20	0.97	0.39	0.96	0.46	5.31
		GA	29	0.96	0.41	0.96	0.50	4.87
		SPA	3	0.93	0.57	0.97	0.40	6.08
	BP	CARS	20	0.96	0.42	0.96	0.51	4.78
		GA	29	0.96	0.44	0.94	0.58	4.20
		SPA	3	0.91	0.63	0.97	0.44	5.56
Firmness	MLR	CARS	29	0.90	0.26	0.87	0.23	2.81
		GA	22	0.85	0.31	0.79	0.29	2.20
		SPA	27	0.88	0.28	0.84	0.25	2.51
	ELM	CARS	29	0.85	0.31	0.83	0.26	2.43
		GA	22	0.83	0.34	0.75	0.31	2.04
		SPA	27	0.81	0.35	0.79	0.29	2.21
	BP	CARS	29	0.89	0.27	0.83	0.26	2.46
		GA	22	0.83	0.34	0.77	0.30	2.12
		SPA	27	0.84	0.32	0.82	0.27	2.39
SSC	MLR	CARS	18	0.88	0.41	0.84	0.51	2.54
		GA	23	0.76	0.59	0.59	0.82	1.58
		SPA	16	0.74	0.61	0.77	0.61	2.13
	ELM	CARS	18	0.86	0.45	0.76	0.63	2.06
		GA	23	0.74	0.61	0.62	0.78	1.65
		SPA	16	0.75	0.60	0.69	0.70	1.84
	BP	CARS	18	0.87	0.43	0.77	0.61	2.11
		GA	23	0.60	0.77	0.59	0.81	1.59
		SPA	16	0.72	0.64	0.57	0.84	1.54

**Table 4.** Prediction results of the MLR, ELM, and BP models.

models built based on the feature variables selected by CARS, the MLR model established based on the feature variables extracted by CARS obtained higher  $R^2_C$ ,  $R^2_P$ , and RPD and lower RMSEC and RMSEP. Compared with the models based on full wavelengths shown in Table 2, the prediction accuracy of MLR, ELM, and BP models based on feature variables selected by CARS was improved. Especially, the CARS-MLR model achieved the best performance ( $R^2_C=0.88$ , RMSEC=0.41,  $R^2_P=0.84$ , RMSEP=0.51, and RPD=2.54) in predicting SSC.

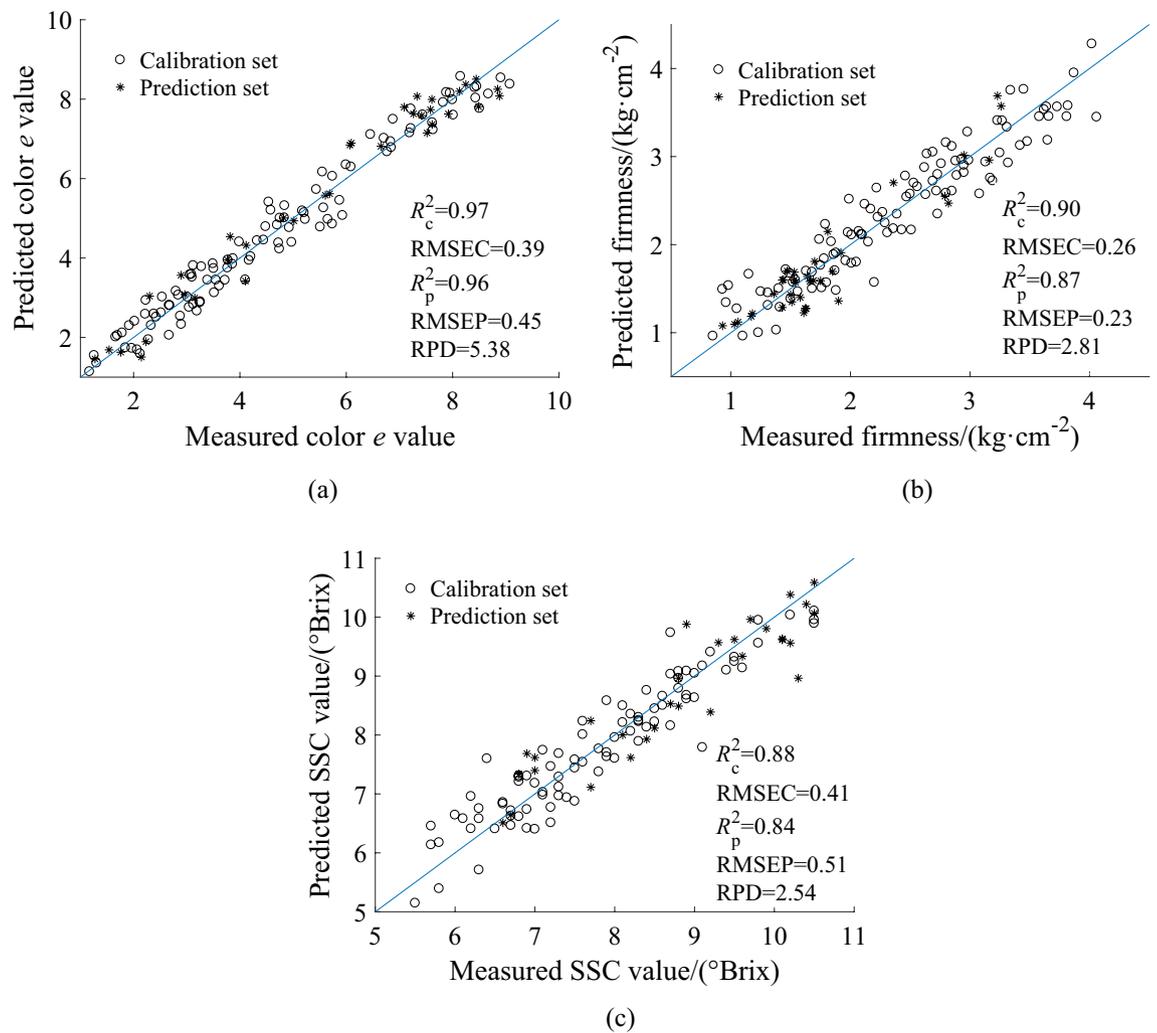
*Modelling based on the optimal combinations of variables.* MLR models using optimal feature variables selected by CARS were established to predict the quality of the loquats regarding colour *e* value, firmness, and SSC. The scatter plots of the actual measured and predicted values are shown in Fig. 5.

Figure 5 shows that the prediction errors of the three quality parameters were all small, and most of the data points were distributed near the fitting line, which indicates that the CARS-MLR model can predict loquat quality (colour *e* value, firmness, and SSC) very well.

The optimal CARS-MLR prediction model formulae for colour *e* value, firmness, and SSC of loquats are as follows:

$$\begin{aligned}
 Y_{\text{colour } e \text{ value}} = & 22.89 - 8.82\lambda_{397} + 12.37\lambda_{401} - 22.02\lambda_{404} + 28.37\lambda_{413} + 2.09\lambda_{418} \\
 & + 20.83\lambda_{422} - 11.90\lambda_{432} - 24.50\lambda_{443} + 23.75\lambda_{498} - 17.75\lambda_{526} + 20.45\lambda_{541} \\
 & - 16.34\lambda_{553} - 0.04\lambda_{621} + 21.10\lambda_{623} - 8.06\lambda_{641} - 15.07\lambda_{717} + 14.64\lambda_{750} \\
 & + 23.98\lambda_{972} - 0.96\lambda_{974} - 18.62\lambda_{993}
 \end{aligned} \quad (8)$$

$$\begin{aligned}
 Y_{\text{Firmness}} = & 13.36 - 11.57\lambda_{394} + 19.69\lambda_{399} - 14.56\lambda_{406} - 21.74\lambda_{408} + 11.40\lambda_{411} \\
 & + 21.95\lambda_{413} - 1.14\lambda_{415} + 5.08\lambda_{555} - 72.32\lambda_{616} + 98.67\lambda_{619} - 48.77\lambda_{626} \\
 & + 8.86\lambda_{641} + 5.32\lambda_{643} - 1.30\lambda_{678} - 31.62\lambda_{690} + 53.08\lambda_{693} + 22.26\lambda_{705} \\
 & - 61.38\lambda_{707} + 100.76\lambda_{727} - 95.06\lambda_{730} + 48.17\lambda_{733} - 11.88\lambda_{743} - 9.22\lambda_{783} \\
 & - 23.11\lambda_{785} + 10.39\lambda_{796} - 15.85\lambda_{974} + 1.51\lambda_{987} + 12.89\lambda_{1022} + 7.36\lambda_{1024}
 \end{aligned} \quad (9)$$



**Figure 5.** Scatter plots of the modelling results of the CARS-MLR model: (a) prediction results of colour  $e$  value; (b) prediction results of firmness; (c) prediction results of SSC.

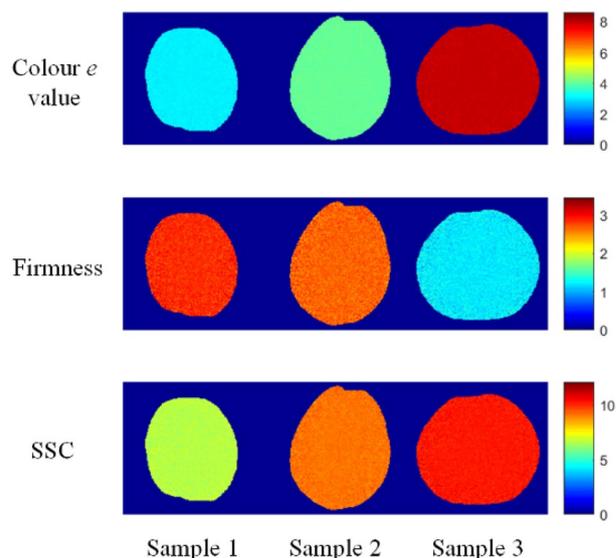
$$\begin{aligned}
 Y_{SSC} = & 36.33 - 76.42\lambda_{418} + 84.97\lambda_{425} + 26.85\lambda_{439} - 48.79\lambda_{488} + 30.30\lambda_{505} \\
 & - 21.61\lambda_{695} + 77.22\lambda_{705} - 67.04\lambda_{720} - 29.14\lambda_{824} - 31.90\lambda_{883} + 222.32\lambda_{885} \\
 & + 61.35\lambda_{888} - 176.46\lambda_{914} - 122.60\lambda_{919} + 72.51\lambda_{940} + 68.45\lambda_{961} - 134.16\lambda_{990} \\
 & + 81.12\lambda_{1019}
 \end{aligned} \quad (10)$$

where  $Y_{colour\ e\ value}$ ,  $Y_{Firmness}$ , and  $Y_{SSC}$  represent the predicted values for colour  $e$  value, firmness, and SSC, respectively.  $\lambda_i$  denotes the reflectance at the feature wavelength, where the subscript  $i$  indicates the wavelength (nm).

**Visualised distribution of quality parameters.** A feature of the HSI technique is that information can be gathered from each pixel of the test sample<sup>37</sup>. The information extracted from the hyperspectral images was used to generate visualisation distribution maps of the reference values (colour  $e$  value, firmness, and SSC), which enabled visualisation of the differences in the reference values between the samples<sup>38</sup>. Due to the approximately spherical shape of loquat fruit, the spectra of different pixels within the same fruit region may exhibit significant differences, potentially leading to poor imaging results. One specific application of loquat fruit detection is to evaluate the overall fruit quality, with secondary emphasis on expressing local characteristics. Building upon this fact, the deviation between the pixel values and the mean spectrum is compressed, and the sum of the compressed deviation and the mean spectrum is employed as the input variable<sup>17</sup>. In this study, the optimal CARS-MLR models were used to predict the quality parameter content of each pixel in loquat<sup>39</sup>. Figure 6 shows the intuitive distribution of colour  $e$  value, firmness, and SSC for samples 1, 2, and 3, respectively. The samples 1, 2, and 3 correspond to maturity stages I, II, and III, respectively.

As shown in Fig. 6, colour  $e$  value and SSC gradually increased with the different maturity stages, while firmness gradually decreased with the different maturity stages. And there were significant visual differences between the different samples. Therefore, the distribution map is useful for online monitoring of loquat quality.





**Figure 6.** Prediction maps for colour *e* value, firmness, and SSC in different loquat samples.

**Maturity stage classification.** A total of 249 samples were used for classifying loquat maturity, with 60 samples in stage I, 150 in stage II, and 39 in stage III. The Kennard–Stone algorithm was applied to partition the samples from each stage into calibration and prediction sets at a ratio of 2:1, resulting in 166 and 83 samples in the calibration and prediction sets, respectively. The PLS-DA, simplified K nearest neighbor (SKNN), and SVM models were applied to discriminate the maturity stages of loquats. The discrimination results are listed in Table 5.

As presented in Table 5, the PLS-DA model had a higher discrimination accuracy in the calibration set than the SKNN and SVM models. The three models had the same discrimination accuracy (97.59%) for the prediction set. Figure 7 shows the confusion matrix of the prediction set, in which two samples from Stage I were incorrectly

Model	Calibration set			Prediction set			Total accuracy/%
	Num	Error	Accuracy/%	Num	Error	Accuracy/%	
PLS-DA	166	3	98.19	83	2	97.59	97.99
SKNN	166	14	91.57	83	2	97.59	93.57
SVM	166	9	94.58	83	2	97.59	95.58

**Table 5.** Prediction results of maturity stages of loquat by PLS-DA, SKNN, and SVM models.

	I	II	III
I	18	2	0
II	0	50	0
III	0	0	13

**Figure 7.** Confusion matrix of prediction set.

identified as Stage II in each of the PLS-DA, SKNN, and SVM models. The results illustrated that the PLS-DA model had the best performance in discriminating loquat maturity.

## Conclusions

In this study, hyperspectral imaging technology was used to detect and visualise loquat quality and discriminate maturity. The main findings of this study are as follows.

1. Hyperspectral imaging coupled with chemometric algorithms is a feasible method for assessing loquat quality. Comparing full spectra models (PLSR and PCR) with simplified models (MLR, ELM, and BP network) based on feature variables selected by three effective variables selection algorithms (CARS, GA, and SPA), the CARS-MLR models with the optimal prediction performance were obtained for colour  $e$  value ( $R^2_p=0.96$ , RMSEP=0.45, RPD=5.38), firmness ( $R^2_p=0.87$ , RMSEP=0.23, RPD=2.81), and SSC ( $R^2_p=0.84$ , RMSEP=0.51, RPD=2.54), respectively.
2. The optimal prediction model combined with pseudo-colour technology could visualise the quality parameter distribution of loquats. The maps show that the distribution of the quality parameters essentially corresponded to the actual situation, and the content of the same quality parameters was significantly different between the loquat samples.
3. Hyperspectral imaging combined with pattern recognition can be used to evaluate loquat maturity. On comparison of the three maturity classification models (PLS-DA, SKNN, and SVM models), the PLS-DA model showed the best performance, with classification accuracies of 98.19% and 97.99% for calibration and prediction sets, respectively.

This study indicates that hyperspectral imaging technology can be used to non-destructively and rapidly determine loquat quality and maturity, providing a theoretical basis for the development of instruments in the future.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 18 June 2023; Accepted: 12 August 2023

Published online: 14 August 2023

## References

1. Fu, X. *et al.* Determination of soluble solid content and acidity of loquats based on FT-NIR spectroscopy. *J. Zhejiang Univ. Sci. B.* **10**(2), 120–125 (2009).
2. Huang, X. *et al.* Prediction of loquat soluble solids and titratable acid content using fruit mineral elements by artificial neural network and multiple linear regression. *Sci. Hortic.* **278**, 109873 (2021).
3. Pinillos, V., Hueso, J. J., Marcon Filho, J. L. & Cuevas, J. Changes in fruit maturity indices along the harvest season in 'Algerie' loquat. *Sci. Hortic.* **129**(4), 769–776 (2011).
4. Besada, C. *et al.* Physiological characterization of 'algerie' loquat maturity: External colour as harvest maturity index. *III Int. Symp. Loquat.* **887**, 351–356 (2010).
5. Cañete, M. L., Hueso, J. J., Pinillos, V. & Cuevas, J. Ripening degree at harvest affects bruising susceptibility and fruit sensorial traits of loquat (*Eriobotrya japonica* Lindl.). *Sci. Hortic.* **187**, 102–107 (2015).
6. Li, B. *et al.* Application of hyperspectral imaging for nondestructive measurement of plum quality attributes. *Postharvest Biol. Technol.* **141**, 8–15 (2018).
7. Pullanagari, R. R. & Li, M. Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. *J. Food Eng.* **289**, 110177 (2021).
8. Fan, S., Huang, W., Guo, Z., Zhang, B. & Zhao, C. Prediction of soluble solids content and firmness of pears using hyperspectral reflectance imaging. *Food Anal. Methods.* **8**(8), 1936–1946 (2015).
9. Jang, K. *et al.* Field Application of a Vis/NIR hyperspectral imaging system for nondestructive evaluation of physicochemical properties in 'Madoka' peaches. *Plants* **11**(17), 2327 (2022).
10. Sun, M., Zhang, D., Liu, L. & Wang, Z. How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method. *Food Chem.* **218**, 413–421 (2017).
11. Wei, X., Liu, F., Qiu, Z., Shao, Y. & He, Y. Ripeness classification of astringent persimmon using hyperspectral imaging technique. *Food Bioproc. Tech.* **7**(5), 1371–1380 (2014).
12. Munera, S. *et al.* Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. *J. Food Eng.* **214**, 29–39 (2017).
13. Tantantrakun, A., Sukwanit, S., Thompson, A. K. & Teerachaichayut, S. Nondestructive evaluation of SW-NIRS and NIR-HSI for predicting the maturity index of intact pineapples. *Postharvest Biol. Technol.* **195**, 112141 (2023).
14. Benelli, A., Cevoli, C., Ragni, L. & Fabbri, A. In-field and non-destructive monitoring of grapes maturity by hyperspectral imaging. *Biosyst. Eng.* **207**, 59–67 (2021).
15. Zhang, C. *et al.* Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *J. Food Eng.* **179**, 11–18 (2016).
16. Teerachaichayut, S. & Ho, H. T. Non-destructive prediction of total soluble solids, titratable acidity and maturity index of limes by near infrared hyperspectral imaging. *Postharvest Biol. Technol.* **133**, 20–25 (2017).
17. Li, X. *et al.* SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology. *Postharvest Biol. Technol.* **143**, 112–118 (2018).
18. Chu, X. *et al.* Green Banana maturity classification and quality evaluation using hyperspectral imaging. *Agriculture* **12**(4), 530 (2022).
19. Zhang, C., Jiang, H., Liu, F. & He, Y. Application of near-infrared hyperspectral imaging with variable selection methods to determine and visualize caffeine content of coffee beans. *Food Bioproc. Tech.* **10**, 213–221 (2017).
20. Li, X., Wei, Z., Peng, F., Liu, J. & Han, G. Non-destructive prediction and visualization of anthocyanin content in mulberry fruits using hyperspectral imaging. *Front. Plant Sci.* **14**, 1137198 (2023).

21. Sharma, S., Sumesh, K. C. & Sirisomboon, P. Rapid ripening stage classification and dry matter prediction of durian pulp using a pushbroom near infrared hyperspectral imaging system. *Measurement* **189**, 110464 (2022).
22. Deng, C. J. *et al.* Relationship between colour and the contents of sugar and acid in different maturity of loquat cultivar guifei. *Chin. J. Trop. Crops* **37**(09), 1747–1751 (2016).
23. Xie, C., Shao, Y., Li, X. & He, Y. Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. *Sci. Rep.* **5**(1), 1–11 (2015).
24. Olmo, M., Nadas, A. & García, J. M. Nondestructive methods to evaluate maturity level of oranges. *J. Food Sci.* **65**(2), 365–369 (2000).
25. Dong, J. & Guo, W. Nondestructive determination of apple internal qualities using near-infrared hyperspectral reflectance imaging. *Food Anal. Methods* **8**(10), 2635–2646 (2015).
26. Zhou, Y. *et al.* Early warning and diagnostic visualization of Sclerotinia infected tomato based on hyperspectral imaging. *Sci. Rep.* **12**(1), 1–13 (2022).
27. Su, W. H. & Sun, D. W. Comparative assessment of feature-wavelength eligibility for measurement of water binding capacity and specific gravity of tuber using diverse spectral indices stemmed from hyperspectral images. *Comput. Electron. Agric.* **130**, 69–82 (2016).
28. Li, X. L., Sun, C. J., Luo, L. B. & He, Y. Nondestructive detection of lead chrome green in tea by Raman spectroscopy. *Sci. Rep.* **5**(1), 1–9 (2015).
29. Asante, E. A., Du, Z., Lu, Y. & Hu, Y. Detection and assessment of nitrogen effect on cold tolerance for tea by hyperspectral reflectance with PLSR, PCR, and LM models. *Inf. Process. Agric.* **8**(1), 96–104 (2021).
30. Wu, D. & Sun, D. W. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals. *Innov. Food Sci. Emerg. Technol.* **19**, 1–14 (2013).
31. Yang, Y. C., Sun, D. W. & Wang, N. N. Rapid detection of browning levels of lychee pericarp as affected by moisture contents using hyperspectral imaging. *Comput. Electron. Agric.* **113**, 203–212 (2015).
32. Ding, S., Zhao, H., Zhang, Y., Xu, X. & Nie, R. Extreme learning machine: Algorithm, theory and applications. *Artif. Intell. Rev.* **44**(1), 103–115 (2015).
33. Askari, M. S., Cui, J., O'Rourke, S. M. & Holden, N. M. Evaluation of soil structural quality using VIS–NIR spectra. *Soil Tillage Res.* **146**, 108–117 (2015).
34. Munera, S. *et al.* Discrimination of common defects in loquat fruit cv'Algerie' using hyperspectral imaging and machine learning techniques. *Postharvest Biol. Technol.* **171**, 111356 (2021).
35. Camps, C. & Christen, D. Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy. *LWT.* **42**(6), 1125–1131 (2009).
36. Zhao, Y. R., Li, X., Yu, K. Q., Cheng, F. & He, Y. Hyperspectral imaging for determining pigment contents in cucumber leaves in response to angular leaf spot disease. *Sci. Rep.* **6**(1), 1–9 (2016).
37. Kong, W., Liu, F., Zhang, C., Zhang, J. & Feng, H. Non-destructive determination of Malondialdehyde (MDA) distribution in oilseed rape leaves by laboratory scale NIR hyperspectral imaging. *Sci. Rep.* **6**(1), 1–8 (2016).
38. Wang, B., He, J., Zhang, S. & Li, L. Nondestructive prediction and visualization of total flavonoids content in *Cerasus Humilis* fruit during storage periods based on hyperspectral imaging technique. *J. Food Process Eng.* **44**(10), e13807 (2021).
39. Wang, F., Wang, C., Song, S., Xie, S. & Kang, F. Study on starch content detection and visualization of potato based on hyperspectral imaging. *Food Sci. Nutr.* **9**(8), 4420–4430 (2021).

## Acknowledgements

This study was supported by the Fund Project of the Central Government Guide Local Science and Technology Department (QKZYD[2022]4050), the Fund Project of Guiyang Science and Technology Bureau (ZKHT[2021]43-15), and Special Funding of Guiyang Science and Technology Bureau and Guiyang University (GYU-KY-[2023]).

## Author contributions

S.F. collected data and wrote the main manuscript text. J.S. guided the experiments, checked the results, and approved the final version. T.T. and Q.W. investigated the background and processed data processing. Q.M. designed the experiments and made guidance for the writing of the manuscript. All authors revised the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023