IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 10, 2019, accepted August 31, 2019, date of publication September 5, 2019, date of current version September 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939579

Comparative Analysis of Non-Destructive Prediction Model of Soluble Solids Content for *Malus micromalus Makino* Based on Near-Infrared Spectroscopy

QIANG GAO[1,4], MEILI WANG[2,4], YANGYANG GUO[1,4,5], XIAOQIANG ZHAO[3], AND DONGJIAN HE[1,4,5]

[1]College of Mechanical and Electronic Engineering, Northwest A&F University, Xianyang 712100, China
[2]College of Information Engineering, Northwest A&F University, Xianyang 712100, China
[3]School of Communication and Information, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[4]Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture, Xianyang 712100, China
[5]Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Xianyang 712100, China

Corresponding authors: Xiaoqiang Zhao (Zxq7703@126.com) and Dongjian He (hdj168@nwsuaf.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 31301242, in part by the Science and Technology Innovation Team for Talent Promotion Plan of Shaanxi Province under Grant 2019TD-028, in part by the Key Research and Development Project of Shaanxi Province under Grant 2019NY-167, and in part by the Special Planned Project for Serving Local Areas of Education Department of Shaanxi Provincial Government under Grant 18JC029.

ABSTRACT To investigate the feasibility of using near-infrared (NIR) spectral technology to detect the soluble solids content (SSC) of *Malus micromalus Makino*, rapid and non-destructive prediction models of SSC were studied using least-square support vector regression (LS-SVR), partial least squares regression (PLSR), and the error back propagation artificial neural network (BP-ANN). First, 110 samples of NIR diffuse reflectance spectra in the wavelength range of 400.41–1083.89 nm were obtained, and then were divided into the calibration set and prediction set by sample set partitioning based on the joint x–y distance (SPXY) algorithm. Second, we compared the prediction performance of the PLSR model after preprocessing by nine spectral preprocessing methods, and applied data dimension reduction methods (random frog, the successive projections algorithm (SPA), and principal component analysis) for variable selection. Finally, the effect of applying full spectrum and characteristic spectrum modeling on SSC prediction accuracy was compared and analyzed. The comparison studies confirmed that the optimal fusion model of SPA-LS-SVR had the best performance ($R_C = 0.9629$, $R_P = 0.9029$, RMSEC = 0.199, RMSEP = 0.271). The experimental results could provide a reference for future development of the internal component analysis system for *Malus micromalus Makino* based on NIR spectroscopy and its classification system using SSC as the classification standard.

INDEX TERMS Least-square support vector regression, *Malus micromalus Makino*, near-infrared spectroscopy, soluble solids content, successive projection algorithm.

I. INTRODUCTION

Malus micromalus Makino, Rosaceae apple, a rare fruit found only in China, is an important contributor to food production and nutrition [1]. The trees are distributed in the area of Fugu county in the city of Yulin, in the areas of Hequ, Pianguan, and Baode in Shanxi province, and in Jungar Banner of inner Mongolia [2]. Compared with other fruits, it has a sweet and sour taste, crisp and juicy meat, and is becoming more and more popular with the local people. It is a good raw material for fruit processing and is rich in soluble solids content (SSC), acids, and calcium. It has been processed into brandy, beverages, pastries, and other products. In recent years, due to its high nutritional and economic value, it has been widely promoted by the local government as a special agricultural product of Fugu Valley to be shipped to all parts of the country and across the world.

Sugar content is one of the most important internal qualities of *Malus micromalus Makino*. About 85% of the components in the sugar content are the SSC, and therefore SSC

The associate editor coordinating the review of this manuscript and approving it for publication was Geng-Ming Jiang.

is often used to reflect the sugar content [3]. A previous study [4] showed that the SSC not only indirectly reflects the fruit's growth environment, cultivation conditions, processing, and storage and transportation conditions, but it also has a significant impact on commodity value. In Fugu, about 40,000 tons of *Malus micromalus Makino* are processed and graded to enter the market every year. Surprisingly, because of the scarcity of *Malus micromalus Makino* varieties and the strong regionality of planting, there are few studies focusing on a rapid measurement method of SSC for *Malus micromalus Makino*. Thus, it is of great significance to study a fast prediction method for determining SSC of *Malus micromalus Makino*; a classification standard can be developed, and production enterprises will benefit economically.

At present, the refractometer is widely used to measure the SSC of fruits at home and abroad. However, it has a fatal disadvantage in that it destroys the integrity of fruits. Fortunately, it was demonstrated in a number of studies that near-infrared (NIR) spectral technology is a non-destructive and easy-to-use detection technology that can be used in food [5], chemical [6], and medical [7] industries. From the perspective of principle analysis, NIR spectroscopy is a type of high-energy vibrational spectroscopy and the NIR electromagnetic radiation is invisible to the naked eye [8]. Furthermore, near-infrared spectroscopy is produced by the transition of molecular vibration from a ground state to a high energy level. In the case of influence by external energy, when the electrons outside the nucleus transition from the ground state to the excited state, or return to the ground state from the excited state, the light energy of a certain characteristic wavelength is absorbed or released [9]–[11]. Under certain conditions, the intensity of the emission and absorption spectra of the same substance is proportional to the content of the substance. Thus, we can qualitatively and quantitatively analyze a material by assessing its spectral intensity.

In recent years, NIR spectroscopy has been widely used as an efficient and simple detection technology in the detection of fruit pest defects [12], natural damage [13], recessive damage [14], and pesticide residues [15]. Guo *et al.* [16] used NIR diffuse reflectance spectroscopy combined with an extreme learning machine to establish a damage discrimination analysis model for kiwifruit during storage. The effects of the number of characteristic wavelengths on simplified operation were compared for uninformative variable elimination (UVE), the successive projections algorithm (SPA), and the UVE-SPA algorithm. The results showed that NIR diffuse reflectance spectroscopy was feasible for damage detection of kiwifruit. Li *et al.* [17] used NIR spectroscopy analysis technology and machine vision technology to analyze the spectral characteristics of the reflectivity of apples with and without bruises. A miniature on-line quality detector for apples was developed. The accuracy of the system for non-invasive detection of apples was 94%. Liu *et al.* [18] established the partial least squares discriminant (PLSD) model and peak area discriminant model for healthy pears and black heart pears by NIR diffuse transmission spectroscopy. The PLSD had the better recognition accuracy, which proves that the visible/near-infrared diffuse transmission spectrum is feasible for the detection of black heart disease in 'Yali' pear products. Jamshidi *et al.* [19] developed an optical system based on NIR spectroscopy and graphical software for determination of pesticide residues of cucumbers. It was concluded that the partial least squares regression (PLSR) model and the PLSD model have good performance for detection of diazinon residue.

In addition, NIR spectroscopy has also been applied in the analysis of vitamin C [20], phenols [21], SSC [22], and other components [23] of kiwifruit, Fuji apples, and pears. A non-destructive methodology based on NIR spectroscopy was developed by Fernández-Novales *et al.* [23] to estimate the grape amino acid content along maturation on intact berries. Although the prediction model can obtain good precision, only the prediction model based on PLSR is established and discussed. Therefore, the potential and performance of the model (except PLSR) still need further comparative analysis. Urraca *et al.* [24] developed a hand-held NIR spectrometer for in-field assessment of total soluble solids in grape berries based on near-infrared–reflectance spectroscopy. However, it may difficult to use this spectrometer in different field conditions because it is composed of different elements. The results of [25] also showed that apple SSC prediction based on short-wave NIR and long-wave NIR spectroscopy is feasible. However, SSC prediction of apples requires a combination of multiple spectral bands instead of a single spectral band or a unique characteristic absorption band.

The above research shows that near-infrared spectroscopy can be used to predict the SSC of apples, pears, kiwifruit, and other fruits. However, there is a lack of research on SSC detection and prediction methods for *Malus micromalus Makino*. Therefore, we try to use near-infrared spectroscopy and chemometrics to establish a non-destructive prediction model of SSC for *Malus micromalus Makino*. Figure 1 shows the main procedures of SSC models using NIR for *Malus micromalus Makino*. After data preparation, the full spectrum (FS) and characteristic spectrum extracted by principal component analysis (PCA), SPA, and random frog (RF) algorithms were used as input for the least-square support vector regression (LS-SVR), PLSR, and error back propagation artificial neural network (BP-ANN) models, respectively. We then compared the prediction accuracy of various feasible fusion models: FS-PLSR, RF-PLSR, SPA-PLSR, PCA-PLSR, FS-LS-SVR, RF-LS-SVR, SPA-LS-SVR, PCA-LS-SVR, FS-BP-ANN, RF-BP-ANN, SPA-BP-ANN, and PCA-BP-ANN.

II. MATERIALS AND METHODS

A. SAMPLE PREPARATION

A total of 110 samples were randomly collected from multiple fruit trees planted in Fugu County Economic Forest Demonstration Base, Yulin City, Shaanxi Province on September 25, 2018. According to the existing enterprise

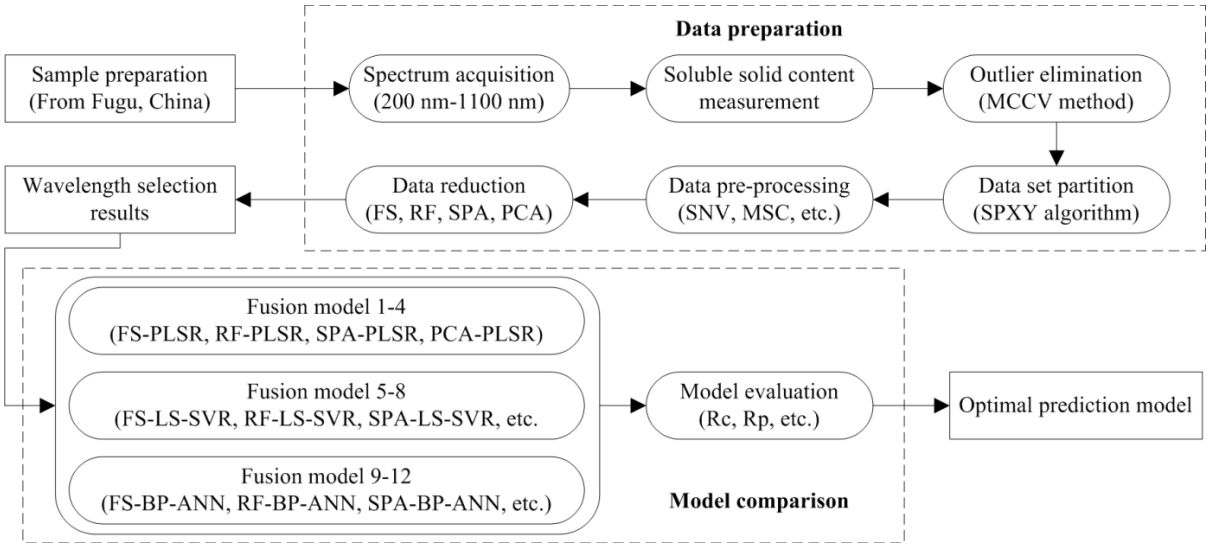


FIGURE 1. A flowchart of the experimental procedure.

grading standard of *Malus micromalus Makino*, selected samples were all with plump fruit, glossy appearance, regular shape, no pests, and no damage. In addition, the equatorial diameter of every sample ranged from 33 mm to 37 mm. Before the spectral acquisition experiment, all samples were cleaned, dried naturally, and numbered. Then they were stored in the laboratory (temperature: 20°C ± 1°C, relative humidity: 60%) for 24 hours so that the samples could be brought to room temperature in order to lower the effect of *Malus micromalus Makino* temperature on the veracity of the forecast results [26].

B. SPECTRUM ACQUISITION

The NIR spectral data acquisition system for samples is shown in Fig. 2. It consists of a visible light source (a halogen tungsten lamp (HL-2000, Ocean Optics, USA)), a spectrometer collimation lens, an optical fiber and spectrometer collimation lens (OFS-1100, Ocean Optics, USA), and a personal computer (Intel (R) Core (TM) i7-6700HQ CPU @ 2.60GHz 2.39GHz). In order to reduce the influence of ambient light on the spectrum [27], we designed a sealed light-shielding cabin with a door on the outside and painted it black on the inside to block the light. The inside of the block was also black. The effective scanning range of the spectrometer was 200–1100 nm with a resolution of 0.43 nm and a total of 2093 wavelength points. The spectrometer, collimation lens, and computer were connected by an optical fiber and USB. The instrument was preheated for 30 minutes before spectral measurement to prevent the detection results from deviating from the true value due to wavelength drift [28]. Under room temperature (20°C–22°C) acquisition conditions, the diffuse reflectance information was obtained by direct contact between the optical fiber probe and the pericarp with reference to the built-in background of the instrument. In the equatorial direction of the sample, the average spectra of three points were taken as the spectral values of the

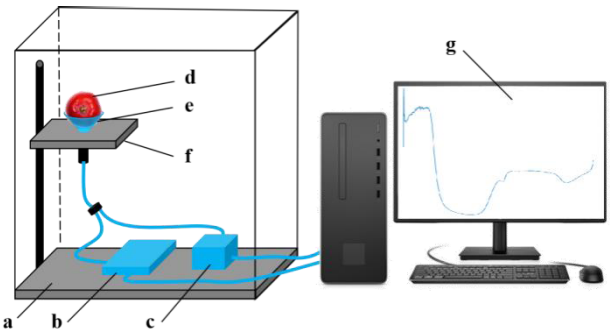


FIGURE 2. Schematic diagram of the near-infrared spectroscopy data acquisition system. (a) Black box for data collection, (b) Optical fiber and geographic spectrometer, (c) Visible light source (halogen tungsten lamp), (d) Sample of *Malus micromalus Makino*, (e) The shading utensil, (f) The height-adjustable stage, and (g) The personal computer.

sample after each rotation of 120 degrees. Spectral data were collected and converted by *Spectra-Suite* and *Unscrambler X* 10.4 (CAMO, Nedre Vollgate, Oslo, Norway). Data processing and modeling analysis were conducted with MATLAB 2016a (The Math Works, USA) and a neural network training tool.

C. SSC MEASUREMENT

After the spectra acquisition, the SSC of *Malus micromalus Makino* was measured using traditional destructive tests for reference. *Malus micromalus Makino* juice was squeezed using a manual fruit squeezer from the same position where spectra were acquired. The fruit sugar refractometer (PAL-BX/ACID5, ATAGO, Japan) was calibrated with distilled water before measurement. Then, at the same points of the spectrum measurement, we removed the peel, and took about 2.5 g of pulp and pressed it into juice. After filtering, we dropped 0.1 ml of the juice onto the fruit sugar refractometer. The average values of six measurements were taken as the SSC values of the samples.

TABLE 1. Sample dividing results and SSC statistical indicators.

Subset	Number	SSC /%					
		Min.	Max.	Mean	Median	S.D.	CV/%
Calibration Set	80	14.50	18.10	16.24	16.20	0.728	0.045
Prediction Set	26	15.00	17.30	16.01	16.05	0.613	0.038

D. OUTLIER ELIMINATION AND DATASET PARTITIONING

Outliers were eliminated by the Monte Carlo cross validation method [29]. This method has the advantages of simultaneously detecting spectral outliers and physicochemical reference outliers [30]. At first, the sample set (we randomly chose 80% of samples as the calibration set and used the remaining 20% of samples as the prediction set) was used to establish 1,000 partial least squares models with random test cross validation. When the root mean square error of cross validation (RMSECV) reached the minimum, the optimal principal component number of the model was obtained. Then, the statistical characteristic parameters of each model after sorting and the cumulative value of the prediction residual error sum of squares of each sample were determined. At last, outliers were recognized with the accumulative value [31]. In this paper, samples 4, 59, 62, and 110 were removed by the above method, and the regression coefficient (R_C) of the model increased from 0.6311 to 0.8511 after eliminating the abnormal samples.

In this study, the sample set partitioning based on the joint x–y distance (SPXY) algorithm proposed by Galvão *et al.* [32] was applied to divide the *Malus micromalus Makino* samples. Based on the Kennard–Stone algorithm, SPXY took both x and y variables into account when calculating the distance between samples. In this way, the multi-dimensional vector space could be effectively covered, leading to a more representative calibration set. Therefore, the prediction performance of the model was greatly improved. Experiments demonstrated that the SPXY algorithm could be effectively utilized in the establishment of a near-infrared quantitative model [33]. Based on [32] and [33], the present paper adopted the SPXY method to study the 106 valid samples remaining after eliminating the abnormal samples. The valid samples were divided into a calibration set and a prediction set (approximately 3:1; 80 were in the calibration set and 26 were in the prediction set). The results of sample partitioning and SSC statistical indicators are shown in Table 1. The minimum value of SSC in the calibration set was less than the minimum value in the prediction set; meanwhile, the maximum value in the calibration set was larger than the maximum value in the prediction set. This indicated that the division was reasonable and representative.

III. EXPERIMENT AND RESULTS

A. SPECTRAL PREPROCESSING

In order to effectively eliminate the noise information contained in the first and last parts of the spectrum, the spectrum ranging from 400.41 nm to 1083.89 nm was captured as a valid reference. In the FS, 1560 wavelengths were reserved for SSC prediction, as shown in Fig. 3. In addition, for the sake of the model's stability [34], nine pretreatment methods were adopted to process the retained spectral data: Savitzky–Golay smoothing (SG), multiplicative scattering correction (MSC), moving average, median filtering, normalization, standard normal variable transformation (SNV), baseline, de-trending, and direct differential first-order derivatives (DDFOD). Thus, we got rid of the high-frequency random noise, baseline drift, and light scattering, as well as eliminated the component differences of SSC between samples. Besides, the SSC prediction models using different pretreatment methods based on PLSR were established (details below). Table 2 shows that the RMSECV pretreated by the SNV algorithm was 0.357; this was the minimum value compared with the other pretreatment methods. The maximum R_C was 0.9179. This illustrated that the model utilizing the SNV algorithm for pretreatment had better stability and cross validation effect than the other models, and therefore it was selected for further analysis.

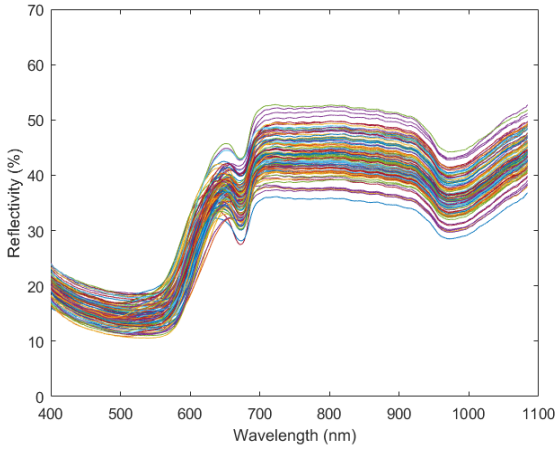


FIGURE 3. The original diffuse reflectance spectroscopy in the range of 400.41–1083.89 nm after noise is removed.

B. VARIABLE SELECTION METHOD

1) RANDOM FROG ALGORITHM

RF [35] is a new algorithm for feature band selection. Modeled by a small number of variable iterations, it can output the probability of every variable selection before selecting the variables. This is an effective method for selecting high-dimensional data variables [36]. In this paper, 1560 near-infrared spectral wavelengths were taken as the potential candidate set. After an effective pretreatment of spectral

TABLE 2. Performance of the SSC prediction model (PLSR) for *Malus micromalus Makino* based on different pretreatment methods. PCs: the scores of the principal compounds.

Method	R_C	RMSEC	RMSECV	PCs
Without Pretreatment	0.8511	0.451	0.472	7
SG	0.8798	0.351	0.368	5
MSC	0.8954	0.460	0.463	8
SNV	0.9179	0.284	0.357	6
Moving Average	0.8591	0.311	0.398	7
Median Filter	0.8752	0.438	0.409	6
Normalization	0.8815	0.312	0.381	7
Baseline	0.8732	0.364	0.367	5
De-trending	0.8616	0.409	0.434	6
DDFOD	0.8933	0.413	0.422	5

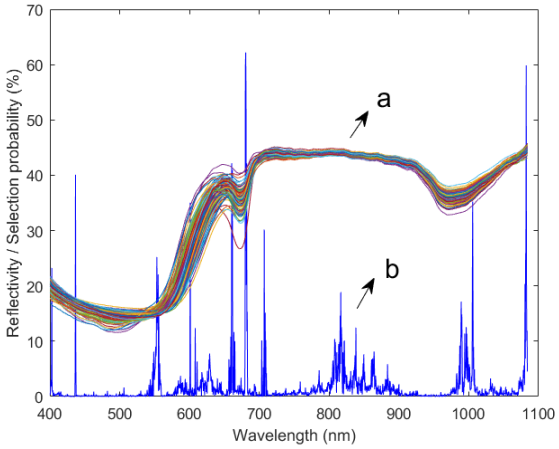


FIGURE 4. The result of feature variable extraction using the RF algorithm. (a) Spectrogram after SNV pretreatment. (b) Variable selection probability after executing the RF algorithm.

data, spectral bands were screened by the open source RF toolbox (http://code.google.com/p/randomfrog/). In order to ensure the convergence of results, the iteration parameter N=11,000 of the algorithm was set, and the probability of each band being selected as the characteristic wavelength was calculated. Then, all the wavelength variables were sorted. Figure 4 displays the results of feature wavelength extraction obtained by the RF algorithm. In order to contain the vast majority of effective information in the original spectrum, the selection threshold was set to 10%, and 52 valid characteristic band data were extracted for subsequent establishment of the SSC prediction model for *Malus micromalus Makino*.

2) SUCCESSIVE PROJECTIONS ALGORITHM

SPA adopted forward selection, searching for minimum redundant information and minimum collinear sets of variables [37]. This algorithm evaluated the candidate variable subset based on RMSEC and eliminated redundant variables through variable elimination. It is one of the most important methods for extracting characteristic wavelength variables, and has been widely applied in the optimization of spectral characteristic wavelengths. In this study, RMSEC varied with the increase in effective wavelength, and the number of

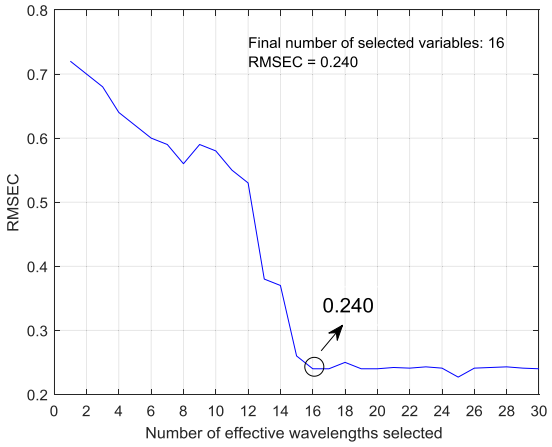


FIGURE 5. The variation of RMSEC with the number of SPA characteristic wavelengths.

wavelengths was set to 1–50. Figure 5 demonstrates that the RMSE reached the minimum value of 0.240 when 16 characteristic wavelengths were selected. Therefore, 16 characteristic wavelengths were used as the input of the prediction model. The distribution of selected characteristic wavelengths in the whole spectrum is charted in Fig. 6; wavelengths near 550 nm, 680 nm, and 970 nm were selected to build the model. This result indicated that the low reflectance in the original spectrum meant high absorptivity, and that some effective information was contained in the absorption peak.

3) PRINCIPAL COMPONENT ANALYSIS ALGORITHM

PCA [38], a classical descending dimension method, eigen-decomposed the covariance matrix in order to gain eigenvectors and their eigenvalues. Therefore, PCA can expose the complex structure of high-dimensional data through a few comprehensive indicators. It can better explain the data variables, simplify an algorithm, and improve computational speed.

When spectral data of SSC from *Malus micromalus Makino* were processed by PCA, the contribution rate of the first 12 principal components and their cumulative contribution rates were as shown in Table 3. The contribution rates of the first principal component (PC1), second principal component (PC2), and third principal component (PC3) were 45.62%, 25.92%, and 8.46%, respectively. The cumulative contribution rates of the three components reached 84.00%. In order to effectively prevent under-fitting of the prediction model caused by too few components, and to avoid the over-fitting situation caused by redundant component information, the first 11 principal components were finally selected as the input of the SSC prediction model after numerous single-factor modeling experiments.

C. PREDICTION MODEL ESTABLISHMENT

1) PARTIAL LEAST SQUARES REGRESSION

PLSR is an important kind of multivariate statistical and data analysis method [39]. By projecting independent variable X

TABLE 3. Principal component contribution rates and the accumulated contribution rates.

Parameter	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Contribution (%)	49.62	25.92	8.46	7.43	1.87	1.36	0.84	0.76	0.55	0.43	0.25	0.20
Total contribution (%)	49.62	75.54	84.00	91.43	93.30	94.66	95.50	96.26	96.81	97.24	97.49	97.69

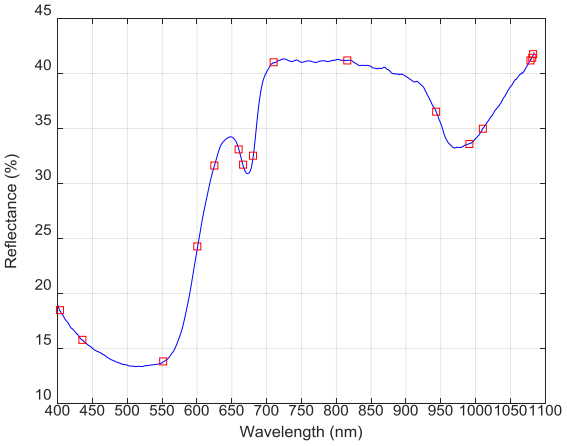


FIGURE 6. Visual distribution of effective bands selected by SPA.

and dependent variable Y's high-dimensional data space into the corresponding low-dimensional space, mutual orthogonal eigenvectors of X and Y are obtained. Then the linear regression relationship between X and Y eigenvectors can be established, thus effectively resolving the multicollinearity. PLSR emphasizes the interpretation and prediction of X to Y while selecting eigenvectors, which can eliminate the influence of regressive and useless noises, thus leading to a minimum number of variables contained in the model.

It is crucial to select principal factors when applying PLSR for regression analysis. Selecting an improper number of factors may result in over-fitting or under-fitting, thus reducing the prediction accuracy of the model. Therefore, PLSR was used for spectral data SSC modeling and analysis; the wavelength range was 400.41 nm–1083.89 nm. The ten-fold cross validation method was applied to optimize the selection, and the number of principal components of FS, RF, SPA, and PCA were confirmed as 103, 52, 16, and 11, respectively.

2) LEAST-SQUARE SUPPORT VECTOR REGRESSION

LS-SVR, proposed by Chauchard *et al.* [40], took the loss function learned by a machine and applied two norms into the objective function of its optimization. Then, equality constraints replaced inequality constraints in the standard SVM algorithm, which simplified the problem of constrained convex quadratic programming to a set of linear equations obtained by Kuhn–Tucker. This greatly reduced the computational complexity and improved the model's speed.

The steps for establishing the regressive model for *Malus micromalus Makino* are described as follows:

In this study, the PC scores of FS, RF, SPA, and PCA were also considered as input vectors for the fusion

I. The SSC regression problem for *Malus micromalus Makino* can be described as in formula (1), given a calibration set $S = \{\mathbf{x}_k, y_k\}_{k=1}^{n}$, where $f(\mathbf{x})$ is the regression equation, $\varphi(\mathbf{x})$ is the non-linear mapping from data ω^{T} is the weighted vector, b is the threshold, and ε is the random error.

$$\begin{cases} f(\mathbf{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}) + b, \quad \mathbf{x} \in \mathbf{R}^m,\ y \in \mathrm{R} \\ y = f(\mathbf{x}) + \varepsilon \end{cases} \tag{1}$$

II. Add the constraint condition and establish optimal objectives as in formula (2), where γ is the penalty factor, and e_k the relaxation variable.

$$\begin{cases} \min\limits_{\omega,b} \quad J(\boldsymbol{\omega}, e_k) = \dfrac{1}{2}\|\omega\|^2 + \dfrac{\gamma}{2}\sum\limits_{k=1}^{n} e_k^2 \\ \text{s.t.} \quad y_k = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_k) + \mathbf{b} + e_k,\ k = 1, \ldots, n \end{cases} \tag{2}$$

III. The corresponding Lagrange function is shown in formula (3), where α_k is the Lagrange multiplier.

$$L = \frac{1}{2}\|\omega\|^2 + \frac{\gamma}{2}\sum\nolimits_{k=1}^{n} e_k^2 - \sum\nolimits_{k=1}^{n} \times \alpha_k \left(\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_k) + b\right) + e_k - y_k\right) \tag{3}$$

IV. Then the empirical equation (formula (4)) can be obtained, where $k(\mathbf{x}, \mathbf{x}_k)$ is the kernel function.

$$\hat{f}(\mathbf{x}) = \sum\nolimits_{k=1}^{n} \hat{\alpha}_k \mathbf{K}(\mathbf{x}, \mathbf{x}_k) + \hat{b} \tag{4}$$

models of FS-LS-SVR, RF-LS-SVR, PCA-LS-SVR, and SPA-LS-SVR. The regularization parameter γ and the RBF kernel parameter σ^2 were optimized by rough selection and selection of K-fold cross validation. γ determines the trade-off between the training error and model simplicity. σ^2 is the bandwidth, and implicitly defines the non-linear mapping from input space to some high-dimensional feature space. After using the open source library of toolbox (https://www.esat.kuleuven.be/sista/lssvmlab/) to implement the algorithm, the optimal combinations of (γ, σ^2) were found at the values of (13.124, 41.031), (9.362, 52.112), (1.355, 59.367), and (25.771, 100.804), respectively.

3) ERROR BACK PROPAGATION ARTIFICIAL NEURAL NETWORK

BP-ANN [41]–[43], a multi-layer feed forward neural network trained by the error back propagation algorithm, is the most widely used regressive neural network at present [44].

It features great non-linear mapping, self-organization, and fault tolerance.

A BP neural network with a single hidden layer structure was modeled in this paper. ''tansig'' and ''logsig'' were applied in the hidden layer and output layer, respectively, as transfer functions. ''trainlm'' was selected in the training function, and the maximum number of iterations was 2000. The target error was set to 0.0001. According to empirical formulas and the trial-and-error method, the optimal number of hidden layer nodes for the fusion models FS-BP, RF-BP, PCA-BP, and SPA-BP were confirmed as 5, 15, 13, and 11, respectively.

D. MODEL EVALUATION

The effect of applying different feature variable modeling on SSC prediction accuracy was analyzed. The performance evaluation indices of the SSC prediction model were the same as those in [45], [46]. The correlation coefficient of calibration (R_C), the correlation coefficient of prediction (R_P), the root mean square error of calibration (RMSEC), and the root mean square error of prediction (RMSEP) were used to describe the performance of the model in this paper. Furthermore, we used the $R_C + R_P$ value and |RMSEC-RMSEP| to evaluate the overall detection accuracy and robustness of the model, respectively. The definitions of R_C, R_P, RMSEC, and RMSEP are as follows:

$$R_c = \sum_{i=1}^{n_c} \left(y_{mi} - \bar{y}_m\right)\left(y_{pi} - \bar{y}_p\right) \Bigg/ \sqrt{\sum_{i=1}^{n_c} \left(y_{mi} - \bar{y}_m\right)^2} \times \sqrt{\sum_{i=1}^{n_c} \left(y_{pi} - \bar{y}_p\right)^2}; \quad (5)$$

$$R_p = \sum_{i=1}^{n_p} \left(y_{mi} - \bar{y}_m\right)\left(y_{pi} - \bar{y}_p\right) \Bigg/ \sqrt{\sum_{i=1}^{n_p} \left(y_{mi} - \bar{y}_m\right)^2} \times \sqrt{\sum_{i=1}^{n_p} \left(y_{pi} - \bar{y}_p\right)^2}; \quad (6)$$

$$RMSEC = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} \left(y_{pi} - y_{mi}\right)^2}; \quad (7)$$

$$RMSEP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} \left(y_{pi} - y_{mi}\right)^2}. \quad (8)$$

Here, n_c and n_p are the number of *Malus micromalus Makino* samples in the calibration set and prediction set, respectively, y_{pi} is the predicted value of SSC in sample number i, y_{mi} is the measured value of SSC in sample number i, and $\overline{y_m}$ and $\overline{y_p}$ are the mean SSC values in the calibration set and prediction set, respectively. As mentioned in [47], in general, a model with good performance should have higher R_C and R_P values and lower RMSEC and RMSEP values.

IV. DISCUSSION

In the above experiment, we first used the SPXY method to divide 106 samples of *Malus micromalus Makino* into the calibration set and prediction set (at a ratio of about 3:1). Then, after SNV pretreatment and analysis of the original spectral information, FS, the characteristic wavelength extracted by RF, SPA, and the principal components extracted by PCA were used as input variables for PLSR, LS-SVR, and BP-ANN models, respectively. Thus, the SSC prediction models were established. Finally, 12 trained fusion models were used to predict the samples. The results show that the minimum and maximum values of SSC are 14.50% and 18.10% after measuring all samples by traditional destructive tests. Moreover, the prediction results of fusion models indicated that the values of SSC of calibration set samples and prediction set samples are between 14% and 19%. Thus, the reference values and predicted values range from 14%–19% were selected in Fig. 7. Figure 7 shows the results of predicted values and reference values of the fusion models.

When NIR diffuse reflectance spectroscopy is used for quantitative analysis of fruits, high-dimensional spectral data is involved, which dramatically increases the complexity of component analysis. In addition, the stacking problem of the spectral data of the sample will lead to a large amount of redundant information. Therefore, the computation of the model can easily fall into the local minimum and thus reduce prediction performance. However, in this paper, we avoid this problem. As can be seen from Fig. 7, the fusion model of SSC based on RF, SPA, and PCA can achieve the same prediction performance as that based on FS by using only a few characteristic spectra. Besides, no over-fitting or under-fitting problems were observed in any models, and the evaluation indices of some fusion models based on RF and SPA were better than those based on FS. This indicates that it is necessary to reduce the dimension of high-dimensional spectral data in SSC prediction using near-infrared spectroscopy analysis. This can greatly reduce the complexity of the model, improve the operational efficiency, and guarantee or even improve the performance of the model.

In this paper, a total of 16 feature wavelengths were selected by SPA, i.e., 680.46, 1082.24, 994.2, 660.46, 436.71, 1005.55, 1081.82, 553.25, 402.71, 660.9, 816.64, 708.37, 627.88, 655.56, and 945.56 nm. Furthermore, a total of 52 feature wavelengths were selected by RF, i.e., 680.46, 1082.24, 680.02, 680.91, 660.46, 681.8, 661.35, 436.71, 1005.55, 600.99, 707.05, 1081.82, 1081.41, 553.25, 402.71, 1005.13, 555.06, 1082.65, 660.9, 1080.58, 681.35, 1005.97, 816.64, 554.15, 989.14, 552.8, 682.68, 708.37, 555.96, 989.98, 815.77, 553.7, 996.3, 817.51, 838.33, 608.17, 554.61, 1083.48, 682.24, 663.57, 994.2, 997.14, 988.72, 659.12, 989.56, 817.95, 807.94, 709.26, 679.58, 703.51, 821.85, and 1083.06 nm. Among them, the feature wavelengths are arranged in descending order according to the selected probability. As can be seen by analyzing the characteristic wavelengths extracted by the SPA and RF methods, some of

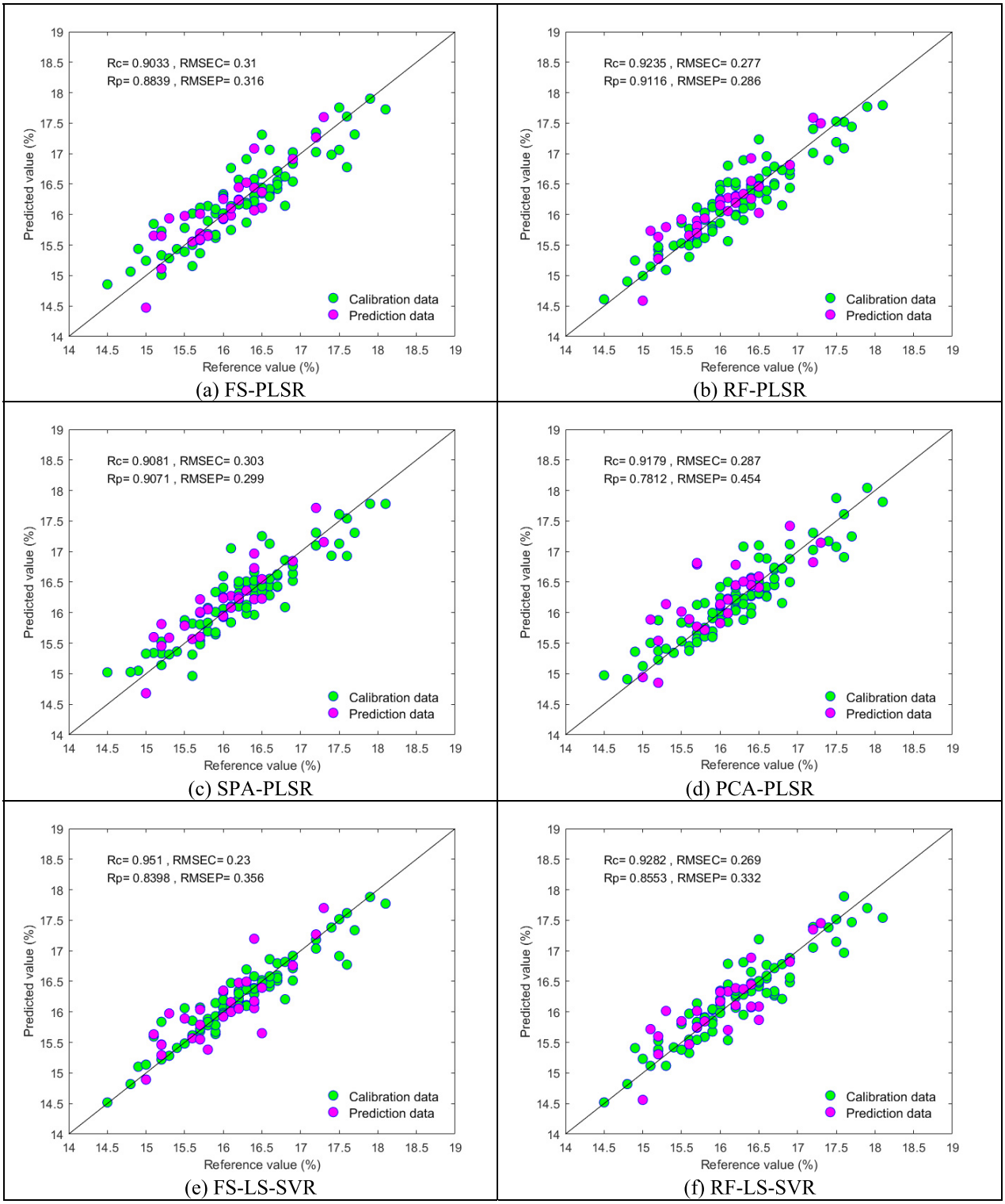


FIGURE 7. Correlations between predicted and reference values of fusion models. (a)–(d) Results of fusion models based on PLSR; (e)–(h) Results of fusion models based on LS-SVR; and (i)–(l) Results of fusion models based on BP-ANN.

the selected bands were the same, such as 680.46, 1082.24, 994.2, and 553.25 nm, and the wavelengths of the remaining features were very close. In both the spectra (Figs. 3–6) and the experimental results, the main absorption peaks of SSC of *Malus micromalus Makino* were located close to 550, 680, and 960 nm. The characteristic absorption band close to 960 nm was related to the second-order frequency doubling peak of the O-H bond, and to the triple frequency doubling peak of the C-H bond of SSC of *Malus micromalus Makino*.

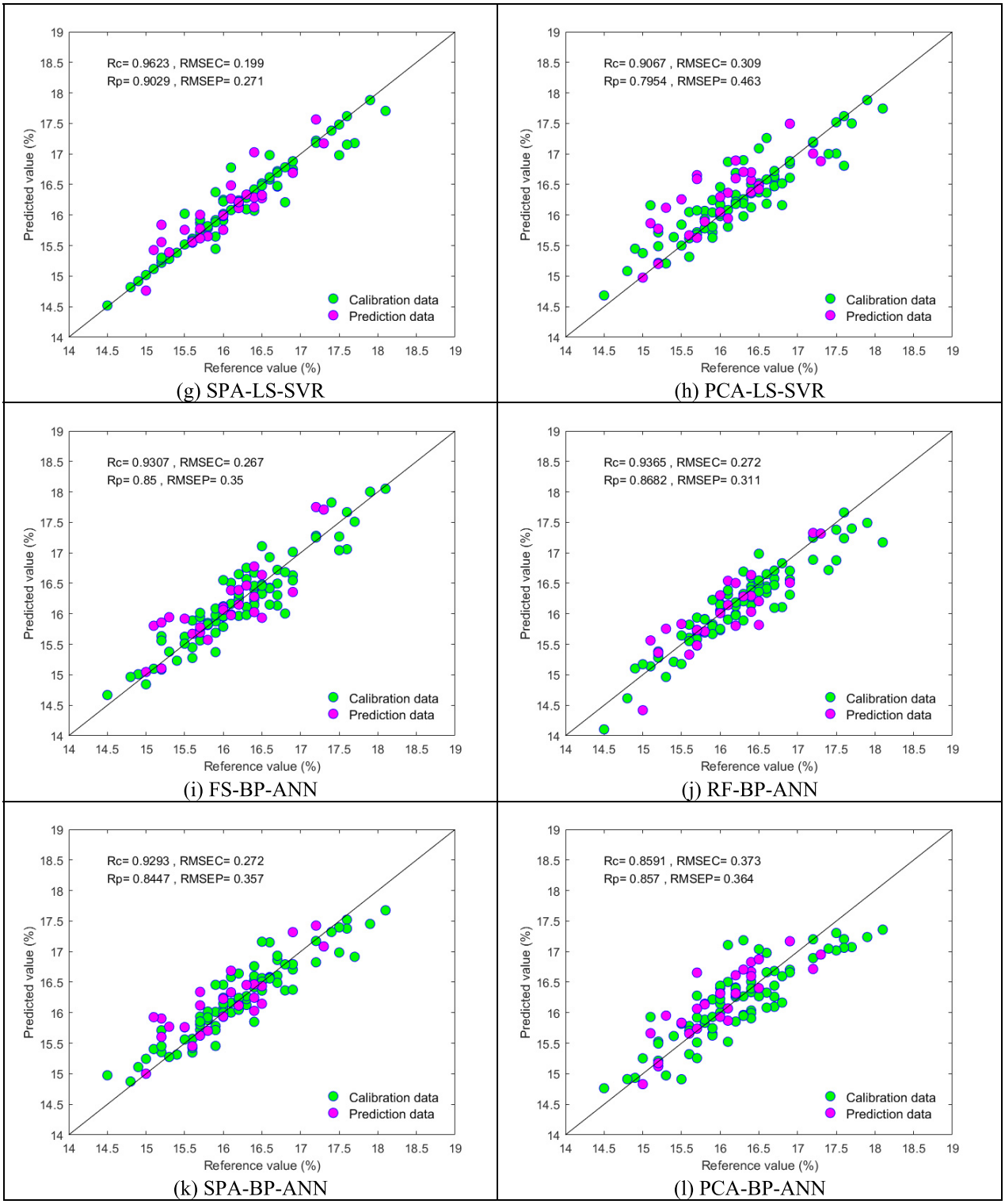


FIGURE 7. *(Continued.)* **Correlations between predicted and reference values of fusion models. (a)–(d) Results of fusion models based on PLSR; (e)–(h) Results of fusion models based on LS-SVR; and (i)–(l) Results of fusion models based on BP-ANN.**

Comparisons of model evaluation indicators are shown in Fig. 8 and Table 4. Of the PLSR-based fusion models, the RF-PLSR model achieves the best performance, with $R_C + R_P = 1.8260$, |RMSEC - RMSEP| = 0.012%, and an 8.5 s runtime. On the contrary, the PCA-PLSR model has the lowest overall detection accuracy and the worst robustness, with $R_C + R_P = 1.6991$ and |RMSEC - RMSEP| = 0.167%. Similarly, of the LS-SVR-based fusion models, we can easily

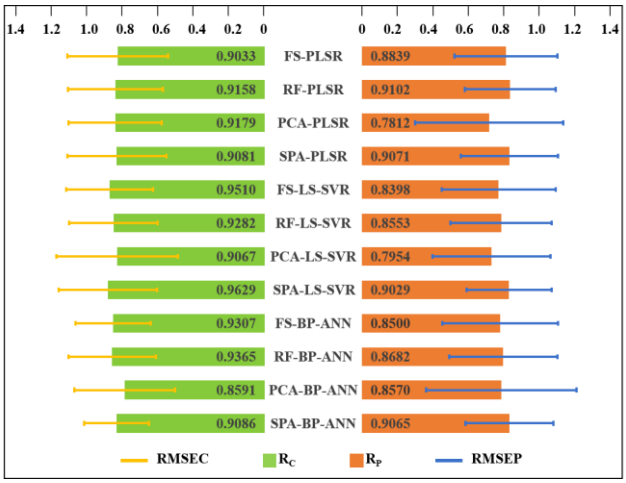


FIGURE 8. Relevance and error result analysis of predicted and reference values of SSC in different fusion models.

TABLE 4. Comparison of different fusion models' runtime, overall detection accuracy, and robustness.

Fusion Model	Model evaluation		
	Runtime (s)	R_C+R_P	|RMSEC-RMSEP|
FS-PLSR	51.2	1.7872	0.006
RF-PLSR	8.5	1.8260	0.012
PCA-PLSR	16.5	1.6991	0.167
SPA-PLSR	29.8	1.8152	0.004
FS-LS-SVR	2101.5	1.7908	0.126
RF-LS-SVR	11.2	1.7835	0.063
PCA-LS-SVR	15.6	1.7021	0.154
SPA-LS-SVR	33.1	1.8658	0.072
FS-BP-ANN	151.2	1.7807	0.083
RF-BP-ANN	9.3	1.8047	0.039
PCA-BP-ANN	13.4	1.7161	0.009
SPA-BP-ANN	30.6	1.8151	0.042

find that the PCA-LS-SVR model is the worst predictor for SSC. It is worth pointing out that the SPA-LS-SVR model achieves the best prediction performance on both the calibration set and prediction set, with $R_C = 0.9629$, $R_P = 0.9029$, RMSEC = 0.199%, and RMSEP = 0.271%. Furthermore, it takes 33.1 s to complete the SSC prediction task. Of the BP-ANN-based fusion models, the predictive performances are ranked from good to bad as follows: SPA-BP-ANN, RF-BP-ANN, FS-BP-ANN, and PCA-BP-ANN. Moreover, the overall prediction accuracies of the SPA-BP-ANN model and the RF-BP-ANN model are greater than 1.8000, but the RF-BP-ANN model took less time to complete the task. The PCA-BP-ANN model also has poor performance, with $R_C = 0.8591$, $R_P = 0.8570$, RMSEC = 0.373%, and RMSEP = 0.364%. This indicated that some useful information in the spectrum is lost when using the PCA algorithm to reduce the dimension of the data.

It is noteworthy that the overall detection accuracy and robustness of the RF-PLSR model are almost the same as those of the SPA-LS-SVR model. Furthermore, the former takes only 8.5 s to perform the prediction task, which saves about 25 s compared with the SPA-LS-SVR model. Thus, under the model evaluation rules proposed in this paper, the SPA-LS-SVR model is the best; however, from the perspective of engineering applications, the RF-PLSR model is more valuable for commercial promotion.

V. CONCLUSION

In this paper, we reported on a novel technology based on NIR spectroscopy analysis to predict the SSC of *Malus micromalus Makino* in a rapid and non-destructive way. Based on the experimental results, the following conclusions were obtained:

(1) The method for predicting SSC in *Malus micromalus Makino* by using spectra in the wavelength range of 400.41–1083.89 nm is feasible.

(2) Compared with FS, the RF, PCA, and SPA methods can effectively reduce variable selection and the complexity of the model. Moreover, the SPA algorithm has a better capacity to select variables for modeling.

(3) Among all prediction models, the fusion model of SPA-LS-SVR achieved the best performance, with $R_C + R_P = 1.8658$ and |RSMEC – EMSRP| = 0.072%. This provides a technical basis for developing a portable SSC detection instrument suitable for *Malus micromalus Makino*.

REFERENCES

[1] Y. Hu, Z. Ma, M. Dang, X. Feng, Y. Sun, X. Yuan, X. Ouyang, and K. E. Woeste, "The complete plastid genome of the endangered species midget crabapple (Malus micromalus)," *Conservation Genet. Resour.*, vol. 10, no. 3, pp. 531–533, Sep. 2018.

[2] H. Yang, J. L. Hua, and C. Wang, "Anti-oxidation and anti-aging activity of polysaccharide from *Malus micromalus Makino* fruit wine," *Int. J. Biol. Macromolecules*, vol. 121, pp. 1203–1212, Jan. 2019.

[3] J. L. Dong and W. C. Guo, "Nondestructive detection of soluble solid content of postharvest kiwifruits based on hyperspectral imaging technology," *Food Sci.*, vol. 36, no. 16, pp. 101–106, Mar. 2015.

[4] A. Moghimi, M. H. Aghkhani, A. Sazgarnia, and M. Sarmad, "Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit," *Biosyst. Eng.*, vol. 106, no. 3, pp. 295–302, Jul. 2010.

[5] C. Mees, F. Souard, C. Delporte, E. Deconinck, P. Stoffelen, C. Stévigny, J.-M. Kauffmann, and K. D. Braekeleer, "Identification of coffee leaves using FT-NIR spectroscopy and SIMCA," *Talanta*, vol. 177, no. 4, pp. 4–11, Jan. 2017.

[6] A. D. de Oliveira, A. F. de Sá, M. F. Pimentel, J. G. A. Pacheco, C. F. Pereira, and M. S. Larrechi, "Comprehensive near infrared study of jatropha oil esterification with ethanol for biodiesel production," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 170, pp. 56–64, Jan. 2017.

[7] A. Sakudo, "Near-infrared spectroscopy for medical applications: Current status and future perspectives," *Clinica Chim. Acta*, vol. 455, pp. 181–188, Apr. 2016.

[8] C. Pasquini, "Near infrared spectroscopy: A mature analytical technique with new perspectives— A review," *Analytica Chim. Acta*, vol. 1026, pp. 8–36, Oct. 2018.

[9] C. Pasquini, "Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications," *J. Brazilian Chem. Soc.*, vol. 14, no. 2, pp. 198–219, Apr. 2003.

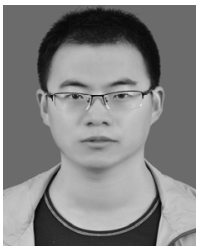
[10] W. F. McClure, "204 years of near infrared technology: 1800–2003," *J. Near Infr. Spectrosc.*, vol. 11, no. 6, pp. 487–518, 2003.

[11] Y. Ozaki, "Near-infrared spectroscopy—Its versatility in analytical chemistry," *Anal. Sci.*, vol. 28, no. 6, pp. 545–563, Jun. 2012.

[12] Y.-D. Liu, H.-C. Xiao, X.-D. Sun, R.-B. Han, L.-Y. Ye, L. Huang, Y.-S. Xiao, and X.-H. Liao, "Study on the quick non-destructive detection of citrus huanglongbing based on the spectrometry of VIS and NIR," *Spectrosc. Spectral Anal.*, vol. 38, no. 2, pp. 528–534, Feb. 2018.
[13] T. Wu, P. R. Armstrong, and E. B. Maghirang, "Vis- and NIR-based instruments for detection of black-tip damaged wheat kernels: A comparative study," *Trans. ASABE*, vol. 61, no. 2, pp. 461–467, Jan. 2018.
[14] Q. Chi, Z. Wang, and T. Yang, "Recognition of early hidden bruises on kiwifruits based on near-infrared hyperspectral imaging technology," *Trans. Chin. Soc. Agricult. Mach.*, vol. 46, no. 3, pp. 235–241, May 2015.
[15] S. Jiang, J. Sun, Z. Xin, H. Mao, X. Wu, and Q. Li, "Visualizing distribution of pesticide residues in mulberry leaves using NIR hyperspectral imaging," *J. Food Process Eng.*, vol. 40, no. 4, Aug. 2017, Art. no. e12510.
[16] G. U. Wenchuan, W. A. Minghai, and G. U. Jingsi, "Identification of bruised kiwifruits during storage by near infrared spectroscopy and extreme learning machine," *Opt. Precis. Eng.*, vol. 21, no. 10, pp. 2720–2727, Oct. 2013.
[17] L. Li, Y. Peng, and Y. Li, "Design and experiment on grading system for online non-destructive detection of internal and external quality of apple," *Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 9, pp. 267–275, May 2018.
[18] Y.-D. Liu, Y.-F. Li, Z.-Y. Gong, and X.-D. Sun, "The discrimination of blackheart pears research based on visible/near-infrared diffuse transmission spectrum on-line detector," *Spectrosc. Spectral Anal.*, vol. 37, no. 12, pp. 3714–3718, Dec. 2017.
[19] B. Jamshidi, E. Mohajerani, and J. Jamshidi, "Developing a Vis/NIR spectroscopic system for fast and non-destructive pesticide residue monitoring in agricultural product," *Measurement*, vol. 89, pp. 1–6 Jul. 2016.
[20] J. He, C. Qiao, D, Li, H. Zhang, H. Dong, Q. Shan, K. Gao, and R. Ma, "Non-destructive detection of vitamin C content in 'Lingwu Changzao' jujubes (Zizyphus jujuba Mill. cv. Lingwu Changzao) using visible near infrared hyperspectral imaging," *Food Sci.*, vol. 39, no. 6, pp. 194–199, Mar. 2018.
[21] H. Xiao, K. Sun, K. Tu, and L. Pan, "Development and application of a specialized portable visible and near-infrared instrument for grape quality detection," *Food Sci.*, vol. 40, no. 8, pp. 300–305, Apr. 2019.
[22] X. Xu, H. Xu, L. Xie, and Y. Ying, "Effect of measurement position on prediction of apple soluble solids content (SSC) by an on-line near-infrared (NIR) system," *J. Food Meas. Characterization*, vol. 13, no. 1, pp. 506–512, Mar. 2019.
[23] J. Fernandez-Novales, T. Garde-Cerdán, J. Tardáguila, G. Gutiérrez-Gamboa, E. PilarPérez-Álvarez, and M. P. Diago, "Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening," *Talanta*, vol. 199, pp. 244–253, Jul. 2019.
[24] R. Urraca, A. Sanz-Garcia, J. Tardaguila, and M. P. Diago, "Estimation of total soluble solids in grape berries using a hand-held NIR spectrometer under field conditions," *J. Sci. Food Agricult.*, vol. 96, no. 6, pp. 3007–3016, Jul. 2016.
[25] Z. M. Guo, W. Huang, Y. Peng, Q. Chen, Q. Quyang, and J. Zhao, "Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of 'Fuji' apple," *Postharvest Biol. Technol.*, vol. 115, pp. 81–90, May 2016.
[26] J. Li, W. Huang, L. Chen, S. Fan, B. Zhang, Z. Guo, and C. Zhao, "Variable selection in visible and near-infrared spectral analysis for noninvasive determination of soluble solids content of 'Ya' pear," *Food Anal. Methods*, vol. 7, no. 9, pp. 1891–1902, Feb. 2014.
[27] Y. Lei, D. He, Z. Zhou, H. Zhang, and D. Su, "Detection of moldy core of apples based on visible/near infrared transmission energy spectroscopy," *Trans. Chin. Soc. Agricult. Mach.*, vol. 47, no. 4, pp. 193–200, Apr. 2016.
[28] S. Travers, M. G. Bertelsen, K. K. Petersen, and S. V. Kucheryavskiy, "Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy," *LWT-Food Sci. Technol.*, vol. 59, no. 2, pp. 1107–1113, Dec. 2014.
[29] S. Gourvenec, J. A. F. Pierna, D. L. Massart, and D. N. Rutledge, "An evaluation of the PoLiSh smoothed regression and the Monte Carlo cross-validation for the determination of the complexity of a PLS model," *Chemometrics Intell. Lab. Syst.*, vol. 68, no. 1, pp. 41–51, Oct. 2003.
[30] Y. Xiao-Yu, L. Gui-Shan, D. Jia-Xing, C. Ya-Bin, F. Meng-Meng, M. Chao, and H. Jian-Guo, "A rapid evaluation of vc content on lingwu long jujube using hyperspectral technique," *Spectrosc. Spectral Anal.*, vol. 39, no. 1, pp. 230–234, Jan. 2019.
[31] Z. Liu, W. Cai, and X. Shao, "Outlier detection in near-infrared spectroscopic analysis by using Monte Carlo cross-validation," *Sci. China B-Chem.*, vol. 51, no. 8, pp. 751–759, Aug. 2008.
[32] R. K. H. Galvão, M. C. U. Araujo, G. E. José, M. J. C. Pontes, E. C. Silva, T. Cristina, and B. Saldanha, "A method for calibration and validation subset partitioning," *Talanta*, vol. 67, no. 4, pp. 736–740, Oct. 2005.
[33] S.-F. Wang, P. Han, G.-L. Cui, D. Wang, S.-S. Liu, and Y. Zhao, "The NIR detection research of soluble solid content in watermelon based on SPXY algorithm," *Spectrosc. Spectral Anal.*, vol. 39, no. 3, pp. 738–742, Mar. 2019.
[34] S. Abasi, S. Minaei, B. Jamshidi, D. Fathi, and M. H. Khoshtaghaza, "Rapid measurement of apple quality parameters using wavelet denoising transform with Vis/NIR analysis," *Scientia Horticulturae*, vol. 252, pp. 7–13, Jun. 2019.
[35] H.-D. Li, Q.-S. Xu, and Y.-Z. Liang, "Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification," *Analytica Chim. Acta*, vol. 740, pp. 20–26, Aug. 2012.
[36] J. Zhao, P. Quan, M. Ma, L. Li, D. He, and H. Zhang, "Comparative analysis of harvest maturity model for Fuji apple based on visible/near spectral nondestructive detection," *Trans. Chin. Soc. Agricult. Mach.*, vol. 49, no. 12, pp. 347–354, Dec. 2018.
[37] W. Luo, Y. Z. Du, and H. L. Zhang, "Discrimination of varieties of cabbage with near infrared spectra based on principal component analysis and successive projections algorithm," *Spectrosc. Spectral Anal.*, vol. 36, no. 11, pp. 3536–3541, Mar. 2016.
[38] Z. F. Yap, Y. Lei, and D. J. He, "Identification of Powdery Mildew and Stripe Rust in Wheat Using Hyperspectral Imaging," *Spectrosc. Spectral Anal.*, vol. 39, no. 3, pp. 969–976, Mar. 2019.
[39] Q. Chen, J. Zhao, M. Li, J. Cai, and J. Liu, "Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms," *J. Pharmaceutical Biomed. Anal.*, vol. 46, no. 3, pp. 568–573, Feb. 2008.
[40] F. Chauchard, R. Cogdill, S. Roussel, J. M. Roger, and V. Bellon-Maurel, "Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust and portable sensor for acidity prediction in grapes," *Chemometrics Intell. Lab. Syst.*, vol. 71, no. 2, pp. 141–150, May 2004.
[41] X. Zhang, L. Chen, Y. Sun, Y. Bai, B. Huang, and K. Chen, "Determination of zinc oxide content of mineral medicine calamine using near-infrared spectroscopy based on MIV and BP-ANN algorithm," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 193, pp. 133–140, Mar. 2018.
[42] S. Khoshnoudi-Nia and M. Moosavi-Nasab, "Comparison of various chemometric analysis for rapid prediction of thiobarbituric acid reactive substances in rainbow trout fillets by hyperspectral imaging technique," *Food Sci. Nutrition*, vol. 7, no. 5, pp. 1875–1883, May 2019.
[43] Z. Hong, M. Yan, L. Yong, Z. Ruili, Z. Xuejun, and Z. Rui, "Rupture energy prediction model for walnut shell breaking based on genetic BP neural network," *Trans. Chin. Soc. Agricult. Eng.*, vol. 30, no. 18, pp. 78–84, Sep. 2014.
[44] C. Wang, X.-H. Wu, L.-Q. Li, Y.-S. Wang, and Z.-W. Li, "Convolutional neural network application in prediction of soil moisture content," *Spectrosc. Spectral Anal.*, vol. 38, no. 1, pp. 36–41, Jan. 2018.
[45] X. Tian, Q. Wang, J. Li, F. Peng, and W. Huang, "Non-destructive prediction of soluble solids content of pear based on fruit surface feature classification and multivariate regression analysis," *Infr. Phys. Technol.*, vol. 92, pp. 336–344, Aug. 2018.
[46] A. C. Dotto, R. S. D. Dalmolin, A. T. Caten, and S. Grunwald, "A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra," *Geoderma*, vol. 314, pp. 262–274, Mar. 2018.
[47] S. Fan, J. Li, Y. Xia, X. Tian, Z. Guo, and W. Huang, "Long-term evaluation of soluble solids content of apples with biological variability by using near-infrared spectroscopy and calibration transfer method," *Postharvest Biol. Technol.*, vol. 151, pp. 79–87, May 2019.

QIANG GAO received the M.S. degree from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2017. He is currently pursuing the Ph.D. degree with the College of Mechanical and Electronic Engineering, Northwest A&F University. His research interests include intelligent detection and control technology, digital signal processing, and automatic control.

MEILI WANG received the Ph.D. degree from Bournemouth University, U.K., in 2011. She is currently an Associate Professor with the College of Information Engineering, Northwest A&F University, Xianyang, China. Her research interests include digital signal processing, computer modeling, and image processing.



YANGYANG GUO is currently pursuing the master's and Ph.D. degrees with Northwest A&F University, Xianyang, China. His research interests include machine vision, computer vision, and image processing.

XIAOQIANG ZHAO received the Ph.D. degree from the Xi'an University of Technology, Xi'an, China, in 2015. He is currently a Full Professor with the Xi'an University of Posts and Telecommunications, Xi'an. His current research interests include agricultural Internet of Things technology, intelligent detection and control technology, and environmental engineering.

DONGJIAN HE received the Ph.D. degree from Northwest A&F University, Xianyang, China, in 1998, where he is currently a Full Professor with the College of Mechanical and Electronic Engineering. His current research interests include intelligent detection and control technology, agricultural information acquisition technology, and equipment and computer vision.