NON-DESTRUCTIVE PREDICTION OF SOLUBLE SOLID CONTENT IN KIWIFRUIT BASED ON VIS/NIR HYPERSPECTRAL IMAGING
/
基于可见/近红外高光谱成像无损预测猕猴桃可溶性固形物含量

Shibang MA*[1)], Ailing GUO [2)] [1]

[1)] School of Mechanical and Electrical Engineering, Nanyang Normal University, Nanyang / China;
[2)] Library, Nanyang Normal University, Nanyang / China
Tel: +86 13503774059; E-mail: mshibang@126.com
DOI: https://doi.org/10.35633/inmateh-70-42

Keywords: *nondestructive and rapid detection, kiwifruit, soluble solid content, Vis/NIR hyperspectral imaging, genetic algorithm*

ABSTRACT
Soluble solid content (SSC) is a major quality index of kiwifruits. Visible near-infrared (Vis/NIR) hyperspectral imaging with the genetic algorithm (GA) was adopted in this study to realize the non-destructive prediction of kiwifruit SSC. A laboratory Vis/NIR hyperspectral imaging system was established to collect the hyperspectral imaging of 120 kiwifruit samples at a range of 400–1100 nm. The average reflectance spectral data of the region of interest of the kiwifruit hyperspectral imaging were obtained after different preprocessing method, namely, Savitzky–Golay smoothing (SG), multiplicative scatter correction (MSC), and their combination method. The prediction models of partial least squares regression, multiple linear regression, and least squares support vector machine (LS-SVM) were built for determining kiwifruit SSC by using the average reflectance spectral data and effective feature wavelength variables selected by GA, respectively. The results show that SG+MSC is the best preprocessing method. The precisions of the prediction models built using the effective feature wavelength variables selected by GA are higher than that established using full average reflectance spectral data. The GA-LS-SVM prediction model has a best performance with correlation coefficient for prediction (R=0.932) and standard error of prediction (SEP=0.536° Bx) for predicting kiwifruit SSC. The prediction accuracy has been improved by 5.6% compared with that of the prediction models established by using the full-band reflectance spectral data. This study provides an effective method for non-destructive detection of kiwifruit SSC.

摘要
可溶性固形物含量是猕猴桃的主要品质指标。本研究采用可见/近红外高光谱成像结合遗传算法实现猕猴桃可溶性固形物含量的无损检测。构建了可见/近红外高光谱成像系统，采集了 120 个猕猴桃样品 400-1100 nm 的高光谱图像。获取猕猴桃高光谱图像感兴趣区域平均反射光谱，采用 Savitzky-Golay 平滑、散射校正及其组合预处理方法对其进行预处理。分别利用全波段反射光谱和遗传算法选取的有效特征波长光谱建立猕猴桃可溶性固形物含量的偏最小二乘回归、多元线性回归和最小二乘支持向量机预测模型。结果表明，Savitzky-Golay 组合散射校正是最好的预处理方法。利用遗传算法选择有效特征波长建立的预测模型精度高于全波段反射光谱建立的预测模型。GA-LS-SVM 预测模型预测猕猴桃可溶性固形物含量的相关系数(R=0.932)和标准误差(SEP=0.536°Bx)最好，相对于全波段反射光谱建立的预测模型，预测精度提高了 5.6%。本研究为无损检测猕猴桃可溶性固形物含量提供了有效的方法。

INTRODUCTION
A fruit with rich nutritional value is one of the most popular and favourite agricultural products worldwide. Kiwifruit, which contains a lot of vitamins, minerals, sugar, and other nutrients, has become a global agricultural commodity *(Berardinelli et al., 2019)*. It is loved by many consumers and called as the “fruit king”. Nowadays, consumers pay more attention to the internal quality of fruits *(Eunhee et al., 2018)*. Soluble solid content (SSC) is a major internal quality attribute index of a fruit. It is an important factor in the timing of fruit ripening and harvesting *(Umer et al., 2020)*. It also determines the taste or quality of the kiwifruit and affects the price and consumers’ purchasing intention *(Xu et al., 2020)*. Moreover, SSC affects the commercial value and market competitiveness of fruits *(Jiang et al., 2019)*.

[1] *Shibang Ma*, Associate Prof. Ph.D.; Ailing Guo, Assistant Librarian*

Predicting fruit SSC has an important reference value for the evaluation of a fruit harvest time, quality, and grading sale. The conventional measurements using a refractometer are reliable and objective. However, they are laborious, destructive, time-consuming and require experience. They cannot meet the demands for non-destructive and online real-time detection of fruits. Moreover, the requirement of consumers and the market such as non-destructive, rapid and easy operating cannot be met. Consequently, a rapid, nondestructive, and easy-to-operate evaluation method is required for evaluating fruit SSC. Hyperspectral imaging is a relatively emerging tool that can provide spectral data information and spatial information from the examined object *(Rosalba et al., 2017)*. Its detection principle is based on the interaction of the light with the examined object. The reaction strength of various objects to light at different wavelengths variable varies. The spectrum variation in each object is determined by its microstructure and chemical composition; thus, the SSC can be reflected by the intensity change in the relevant wavelength of examined object using chemometric methods *(Zhang et al., 2020)*. An examined object image obtained from a hyperspectral imaging system can be taken as a three-dimensional spectral data matrix *(Santosh et al., 2016)*. In recent years, hyperspectral imaging has been broadly recognized as an emerging technology for rapidly and non-destructively assessing the attributes, quality, and safety of fruits *(Ebrahiema et al., 2018)*. *Li et al. (2018)* used short-wave and long-wave near-infrared hyperspectral imaging technologies coupled with an improved watershed segmentation algorithm to detect early bruises on peaches rapidly and nondestructively. The results showed that short-wave NIR hyperspectral imaging model was highly suitable for detecting early bruises on peaches. Its detection results indicated that 96.5% of the bad peaches and 97.5% of the good peaches were accurately detected from all the test peaches. *Reddy R. et al. (2021)* used hyperspectral imaging and multivariate statistics to assess the SSC and flesh firmness of fresh cherry fruits nondestructively by developing a partial least squares regression (PLSR) model and Gaussian process regression (GPR) model. The test dataset results highlighted that GPR model can be used to detect SSC and firmness of cherry with correlation coefficients (R) of 0.88 and 0.60, respectively, and root mean standard errors (RMSE) of prediction of 0.43% and 0.38, respectively. Thus, hyperspectral imaging can be considered as a reliable and robust technology for estimating cherry fruit quality attributes. However, the GPR models showed low certainty with a prediction interval coverage probability of 0.90–0.97. *Ma et al. (2021)* built a NIR hyperspectral imaging detection system combined with a sample rotation holder to scan the whole kiwifruit surfaces and acquire the hyperspectral imaging with key wavelengths within the 1002–2300 nm range. Then, they constructed SSC calibration models via PLSR analysis. The results showed that the coefficient of determination (R_{cv}^2) and RMSE of the cross-validation set were 0.74 and 0.7%, respectively. However, *Ma et al. (2021)* only established a PLSR model. Moreover, the robustness and stability of the model is unknown.

These aforementioned studies showed that Vis-NIR hyperspectral imaging technology can predict the SSC and other qualities of different fruits, such as peach, plum, cherry, and kiwifruit. Those studies established prediction models mainly using the full wavelength spectrum data. The studies on the influence of different prediction model, different preprocessing methods and variable selection method on the prediction accuracy are few. Moreover, studies on the non-destructive prediction of kiwifruit SSC via Vis/NIR hyperspectral imaging with variable selection method are also fewer. The stability and accuracy of a SSC prediction model greatly affect its practical application and cost. Therefore, the stability and accuracy of the prediction model must be improved.

In the previous research process, the authors have explored the feasibility of using visible near infrared spectroscopy to detect the SSC of kiwifruit, and achieved good results. In this study, the authors further studied the feasibility of using hyperspectral imaging technology for non-destructive detection of kiwifruit SSC on the basis of the previous studies. Some experimental methods referred to the previous research process *(Ma, 2021)*.

The present study aimed to establish the approach of using Vis/NIR hyperspectral imaging combined with GA for the non-destructive prediction of kiwifruit SSC. Moreover, the influences of different spectral preprocessing methods, full wavelength variables, effective feature wavelength variables selected by GA, and the accuracy of different prediction models were studied. The best spectral preprocessing method and best prediction model for kiwifruit SSC were also determined. This study provides an effective method for non-destructive detection of fruit quality.

MATERIALS AND METHODS

Vis/NIR hyperspectral imaging system

A Vis/NIR hyperspectral imaging system in the wavelength range of 400–1100 nm was built and used to collect the hyperspectral imaging of the kiwifruit samples in the process of experiment. The schematic diagram of the Vis/NIR hyperspectral imaging system is shown in Fig. 1. The Vis/NIR hyperspectral imaging system comprised a charge coupled device camera, hyperspectral spectrograph, halogen lamp, a sample holder, DC power, power cable, a computer, data cable, and frame.

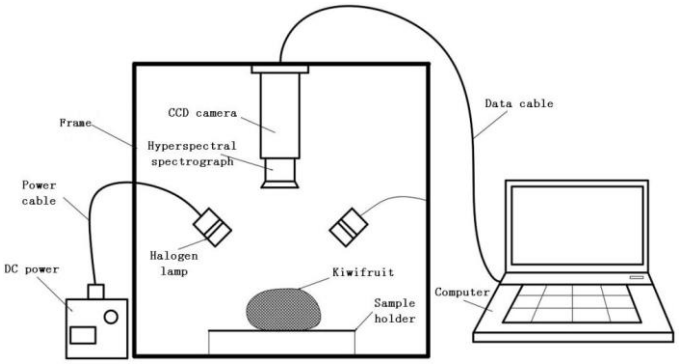


Fig. 1 - Schematic diagram of the Vis/NIR hyperspectral imaging system

Kiwifruit samples

One hundred and twenty fresh kiwifruit samples at the same maturity stage were collected from a kiwifruit orchard. All the samples were transferred to the optical testing laboratory room in a freshness preservation box, and stored under cold conditions at $0-3^{\circ}C$. The experiment was completed in four days. Every time before the experiment, the cold-stored samples were removed to room temperature conditions ($22\pm2^{\circ}C$) for approximately 8 h to reduce collection data variations caused by different temperature of kiwifruits. First, the hyperspectral imaging of each fruit sample was acquired by the Vis/NIR hyperspectral imaging system, then, the SSC values of each fruit sample were measured by the digital refractometer.

Vis/NIR hyperspectral imaging acquisition

Figure 2 shows a scene of the Vis/NIR hyperspectral system. The Vis/NIR hyperspectral imaging of each kiwifruit sample was acquired using the Vis/NIR hyperspectral imaging system. The experimental kiwifruit samples were placed on the sample tray for scanning and the hyperspectral imaging was collected. Each sample was scanned two times from the upper surface and bottom surface, and two Vin/NIR hyperspectral images were collected for each sample. The mean of two images was taken as the final imaging data. Before scanning, the parameters of exposure time, scanning speed and lens focusing distance were set. The raw Vis/NIR hyperspectral imaging was corrected by using the white reference image and dark reference image before each measurement to reduce the influences of external interference and dark current noise. The dark reference image was acquired by recording the Vis/NIR hyperspectral image after the lens was covered with black cap, whereas, the white reference image was acquired by collecting the Vis/NIR hyperspectral image of a white background plate.

The final calibrated Vis/NIR hyperspectral image of sample was calculated based on the following formula:

$$l_c = \frac{l_0 - l_d}{l_w - l_d} \tag{1}$$

where: l_c means the calibrated Vis/NIR hyperspectral image, l_0 indicates the raw Vis/NIR hyperspectral image, l_w represents the white reference images, and l_d denotes the dark reference images.

Spectral profile extraction

The calibrated Vis/NIR hyperspectral images of the samples were further analysed using the ENVI 5.3 software. The scanned region of kiwifruit was selected as the region of interest (ROI). The spectral profile was extracted from the ROI of each sample Vis/NIR hyperspectral image. The spectral profile extraction process is shown in Fig. 3.

The mean intensity of all pixels of each kiwifruit sample in the ROI was calculated and regarded as the full average reflectance spectrum for further constructing prediction model and wavelength variable selection.

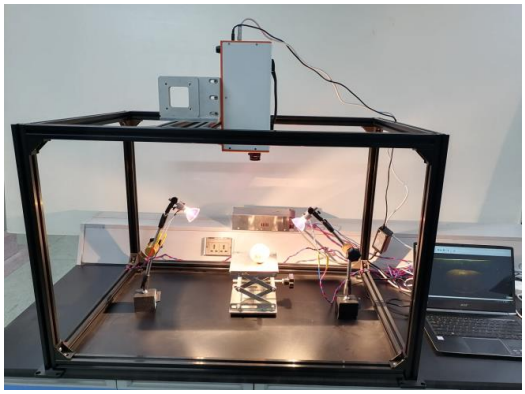


Fig. 2 - The scene of the Vis/NIR hyperspectral imaging system.

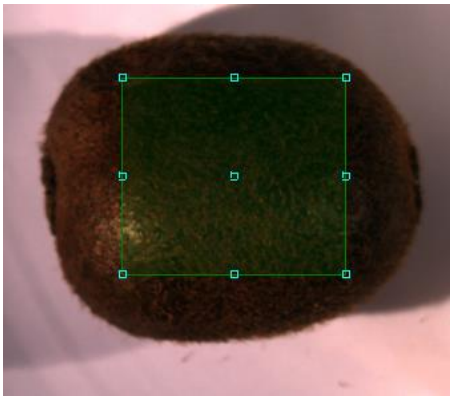


Fig. 3 - ROI selection and spectral profile extraction

Kiwifruit SSC real value

The SSC real value of each kiwifruit sample was measured after the Vis/NIR hyperspectral image collection. First, the kiwifruit sample was peeled and cut into flesh slices in its upper surface and bottom surface, respectively. Second, the flesh slices were squeezed into juice using a fruit juicer (HY-160, Joyoung, China). Then, the SSC of each kiwifruit sample was tested using a digital Brix refractometer (PR-101; ATAGO Co., Ltd., Japan). The average of the SSC real values of the upper and bottom scanned surface was regarded as the real value of kiwifruit SSC.

Data analysis and model establishment

Data preprocessing and model establishment methods

On the one hand, the kiwifruit is similar sphere. Its surface curvature affects the absorption and reflection of light. On the other hand, there are a lot of fluffs on the surface of kiwifruit. Scattering occurs when light hits the surface of kiwifruit. These affect the quality of the spectral data. The raw spectra of samples should be pretreated to improve their qualities. There are many kinds of spectral preprocessing methods. The spectra of samples were pretreated by Savitzky–Golay smoothing (SG) and multiple scattering correction (MSC) to eliminate noise interference and reduce analysis error. The detailed explanation and instruction of the SG algorithm can be found in literature *(Chen et al., 2011)*, whereas that of the MSC algorithm can be found in literature *(Tom et al., 2009)*. The prediction models of the SSC were developed by PLSR, MLR, and the least squares support vector machine (LS-SVM). Prediction model was optimized based on the optimal method to improve predictive ability by eliminating the useless information and redundant variables in the raw spectral data by using an effective variable selection algorithm. GA was used to decide optimized feature wavelength variables. Finally, the best pretreatment method and the best prediction model were determined.

Around 1975, the PLSR was proposed by Herman Wold to build complicated model. It is a widely used to solve the problem of multivariate linear regression. PLSR can be thought of relating two data matrices, X (predictors; i.e., spectrum of kiwifruits in this study) and Y (response; i.e., SSC in this study), through a linear multivariate model. In the present study, PLSR analysis method was used to build a prediction model for detecting kiwifruit SSC. The latent variables (LVs) were considered as an important parameter in establishing a PLSR model. They were considered potential eigenvectors of the raw spectra. LVs can compress the raw spectral data and reduce the dimensionality. The LVs were taken as the input variables of PLSR and simplified the PLSR model. The reasonable number of LVs helps avoid underfitting or overfitting the model. In general, full cross-validation is used to determine the reasonable number of LVs by using the root-mean-square error of cross-validation in establishing a PLSR model *(Li et al., 2019; Nturambirwe et al., 2019)*.

MLR is a well-known statistical analysis method *(Hamid et al., 2012)* that attempts to develop the relationship between an output variable with two or more input variables by fitting a linear regression equation.

The equation of MLR model is:

$$y_i = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \ldots + a_j x_{i,j} + e_i \quad (2)$$

where:

y_i is the output variable, a_0 is a constant, is an input variable, a_j is the regression coefficient vector, and e_i is a random measurement error.

Support vector machine (SVM) is an effective analysis tool for classification problem. However, the training problem of SVM model is a complex and thorny problem. Suykens and Vandewalle proposed a modified and simplified SVM algorithm called LS-SVM. LS-SVM can simplify the training process and improve the convergence rate of regression equation. Therefore, LS-SVM has been widely used to solve complex regression and classification problem and achieved acceptable results *(Xu et al., 2015)*. Establishing prediction model with LS-SVM belongs to a regression problem. In order to solve the regression problems with LS-SVM, suppose $T=(x_i, y_j)$ is the training set, and m is the sample number, where $i=1,2,3,\ldots,m$; x_i is the independent variable, and y_j is the dependent variable. The prediction regression problem of LS-SVM can be described as follows:

$$\begin{cases} \min J(w,a,\xi)=\frac{1}{2}w^T w+\frac{1}{2}\gamma\sum_{i=1}^{m}\xi_i^2 \\ s.t y_i[w^T\varphi(x_i)+b]=1-\xi_i,\ i=1,\ldots,m \end{cases} \quad (3)$$

where:

w is the weight vector, a is a partial vector, y is the penalty parameter used in adjusting the proportion of empirical risk and the confidence interval of LS-SVM ξ_i is the slack variable for x_i.

The LS-SVM parameters and proper kernel function are two important factors that influence the performance of LS-SVM. Therefore, the parameters and proper kernel function should be determined before using LS-SVM. The frequently used kernel functions include radial basis function (RBF), polynomial kernel, and linear kernel in LS-SVM. RBF is a nonlinear function that can simplify the training process and deal with the nonlinear relationships between input variables and output variables. RBF was selected as the kernel function of LS-SVM in present study. The parameter σ^2 and the penalty parameter γ of the RBF kernel function are two key parameters that should be determined. They affect prediction ability and prediction accuracy of LS-SVM. In this study, the γ and σ^2 were determined by using K-fold cross-validation method.

GA (genetic algorithm)

GA is one of the most popular methods for selecting effective features. GA can be regarded as computer simulations of evolutionary process from biological systems. It is able to find an optimal solution in dealing with the search problems and suitable for solving nonlinear complex optimization problems. GA starts with an initial population. Then, a new population is generated by crossover and mutation operations with the selected chromosomes of current population. The population and chromosomes are generated or chosen in term of their fitness at each stage. The selection of initial population, fitness function, chromosome, crossover rate and mutation rate are important steps in GA. The number of initial population is determined according to the training set and the number of extracted features. A new population is generated from the current population by chromosome selection, crossover, and mutation. A detailed description and computer program explanation of the GA can be found in Literature *(Leardi et al., 1998)*. GA has been widely used to select the feature variable and develop predictions. In present study, the GA was used to select the effective feature wavelength variables and develop prediction models.

Model performance criteria

The performance of prediction model was assessed based on calculating the differences between the real measured value and predicted value of kiwifruit SSC by the correlation coefficient and the standard error. The correlation coefficient of calibration set (R_C), the standard error of calibration (SEC), the correlation coefficient of prediction set (R_P), and the standard error of prediction (SEP) were defined as shown in Equations (4)–(7). The four indices were used to assess the performance of the prediction models.

$$Rc = \sqrt{\sum_{i=1}^{n_c}(\hat{y}_i - y_i)^2} \Big/ \sqrt{\sum_{i=1}^{n_c}(\hat{y}_i - y_m)^2} \quad (4)$$

$$Rp = \sqrt{\sum_{i=1}^{n_p}(\hat{y}_i - y_i)^2} \Big/ \sqrt{\sum_{i=1}^{n_p}(\hat{y}_i - y_m)^2} \quad (5)$$

$$SEC = \sqrt{\frac{1}{n_c - 1}\sum_{i=1}^{n_c}(\hat{y}_i - y_m)^2} \quad (6)$$

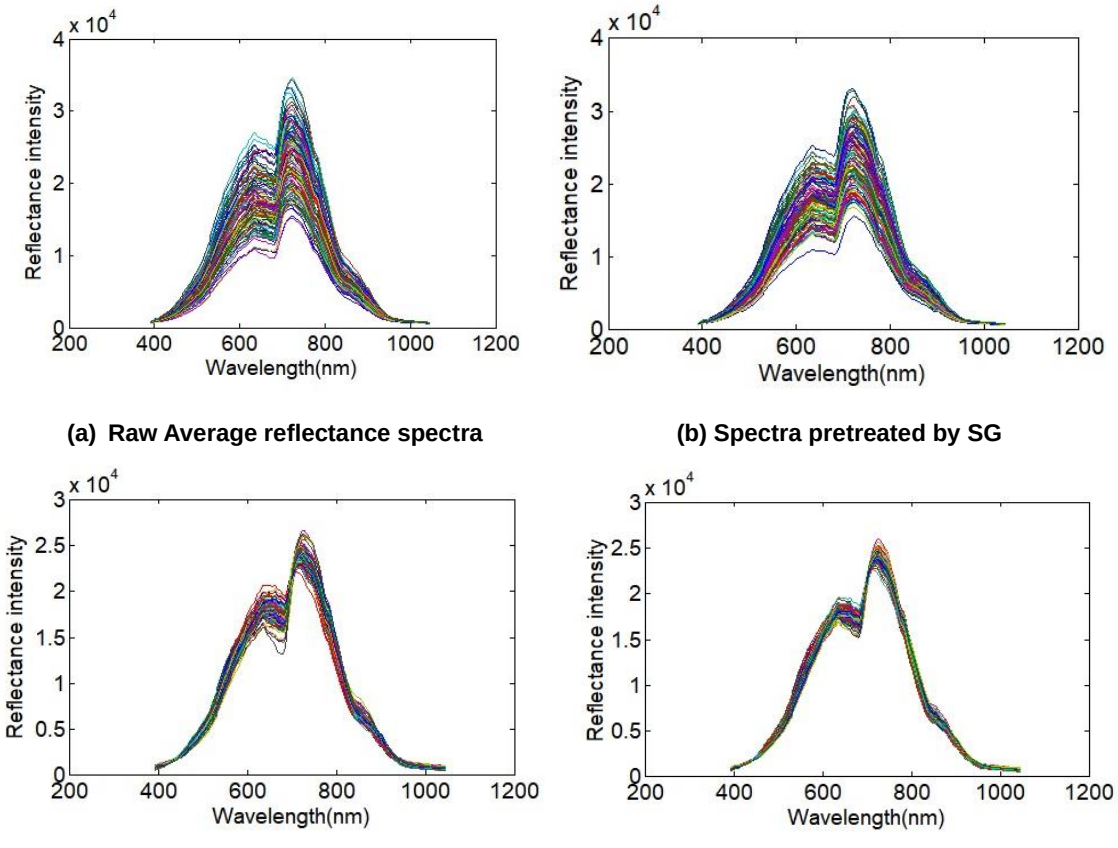
$$SEP = \sqrt{\frac{1}{n_p - 1}\sum_{i=1}^{n_p}(\hat{y}_i - y_m)^2} \quad (7)$$

where: $\hat{y}_i$ is the predicted value of the i^{th} kiwifruit sample, y_i is the real value of the i^{th} kiwifruit sample, y_m is the average value of the calibration or prediction set, n_C is the number of samples in the calibration set, and is the number of samples in the prediction set. In general, a good prediction model should have high values n_P of R_C and R_P, low values of SEC and SEP. In present study, the processing and analysis of all data were done by using MATLAB R2013 software (The Mathworks, Inc., USA).

RESULTS AND DISCUSSION

Spectral analysis of sample and SSC real values

The outliers of the samples' spectra and SSC measured values affect the accuracy and robustness of the prediction model. In order to improve the stability and precision of the prediction model, Mahalanobis distance outlier detection algorithm was used to eliminate the outliers of the samples' spectra and SSC measured values *(Christophe et al., 2018)*. Six outlier SSC value samples were determined and removed from 120 samples. The average reflectance spectra in the ROI of the hyperspectral image of the 114 kiwifruit samples are shown in Fig. 4 (a).



(a) Raw Average reflectance spectra **(b) Spectra pretreated by SG**

(c) Spectra pretreated by MSC **(d) Spectra pretreated by SG+MSC**

Fig. 4 – Raw average reflectance spectra and pretreated spectra of all kiwifruit samples

The reflection spectral curves of the samples had high similarity; the overall trends of the spectral curves were basically the same. Obvious absorption peaks existed at 713, 746, and 870 nm, which were principally due to C–H stretch third overtone and fourth overtone; they were mainly associated with the carbohydrate absorption bands in the kiwifruit *(Hu et al., 2017)*. The average reflection spectra exhibited obvious baseline drift and scattering. Therefore, pretreatment methods of SG, MSC, and SG+MSC were used to pretreat the average reflection spectra. The pretreatment spectra shown in Figs. 4(b) – 4(d) had no baseline drift and scattering.

One hundred and fourteen samples were randomly divided into two sets, namely, the calibration set and the prediction set. The calibration set was used to construct the prediction model, and the prediction set was used to verify the accuracy and robustness of established prediction model. All kiwifruit samples were randomly divided into the calibration set and prediction set according to the ratio of 3:1 by using Kenard-Stone algorithm. The calibration set has 86 kiwifruit samples and the prediction set has 28 kiwifruit samples. Table 1 summarizes the SSC measured values of kiwifruits in the calibration set and the prediction set. Measures of SSC ranged from 8.3 to 13.8 °Bx in the calibration set, and from 7.9 to 13.5 °Bx in the prediction set. The mean values of SSC in the calibration set and the prediction set were 10.5 °Bx and 10.4 °Bx, respectively. The SDs of SSC in the calibration and prediction sets were 1.23, 1.21, respectively. This shows the sample division was reasonable.

Prediction models built using full average reflectance spectrum

Before establishing prediction model, the full average reflectance spectra of samples in calibration and prediction sets were pretreated using the SG, MSC, and SG+MSC methods. Then, they were taken as the input variables to establish the prediction models of kiwifruit SSC. The kiwifruit SSC prediction models of PLSR, MLR, and LS-SVM were built based on the full average reflectance spectra, respectively. Table 2 shows the results of the prediction. The performance of the three prediction models pretreated with SG+MSC was better than that pretreated with SG or MSC only. Therefore, SG+MSC was taken as the pretreatment method for establishing the prediction models. The accuracy of the LS-SVM prediction model was the best in the three prediction models, with R_C = 0.893, SEC = 0.621, R_P = 0.882, and SEP = 0.752. Moreover, the performance of PLSR model was better than that of the MLR model. However, on the whole, the performances of the three prediction models were relatively low. The key reason may be that the full average reflectance spectra contain a lot of redundant, useless wavelength information, thereby reducing the prediction accuracy of the model. Thus, some measures should be taken to eliminate useless wavelength information. Moreover, effective feature wavelength variables should be picked out from the full average reflectance spectra to establish the prediction models and improve their prediction precision.

Prediction models established using effective wavelength variables

In order to optimize and simplify the prediction model and improve the prediction precision, the GA method was used to select the effective feature wavelength variables from the full average reflectance spectra of kiwifruit samples. The selected effective feature wavelength variables by GA were used as input variables to build the PLSR, MLR, and LS-SVM prediction model of SSC. GA parameters were set in term of the number of the spectral domains. The size of population and chromosome of GA was set to 30. Moreover, the probabilities of cross-over and mutation were set to 50% and 1%. The number of run times of GA was set to 100 for every model. At every run time, the initial chromosome population was generated randomly. The GA was randomly run five times to select the effective feature wavelength variables for every prediction model. Fig. 5 shows one of the five times variable selection processes. The wavelengths above the dotted line were selected feature variables by GA.

Overall, the GA-selected variables were mainly concentrated at the wavelengths of 420, 520, 600, 715, 750, and 870 nm. Moreover, the useless information and the number of variables for establishing the prediction model were greatly reduced. Fig. 6 illustrates the distribution of the effective feature wavelength variables picked out by GA at 400–1100 nm. The selected variables were used to establish the GA–PLSR, GA–MLR, and GA–LS-SVM prediction models of SSC. The results of the three different kind prediction models of SSC are shown in table 3. The results prove that the prediction precisions of the models established using the effective feature wavelength variables were better than those of the models established with the full-average reflectance spectra (Table 2).

The prediction accuracies of GA–LS-SVM are the best with R_C = 0.948, SEC = 0.403, R_P = 0.931, and SEP = 0.536 (Fig. 7).

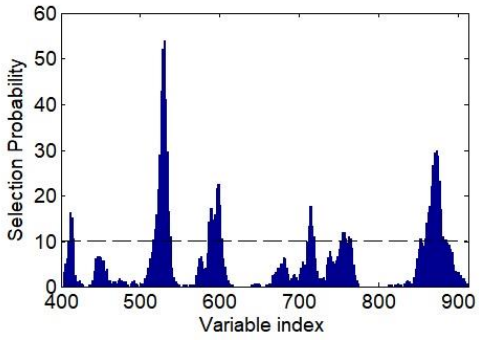


Fig. 5 - Frequencies of the variables selected by GA

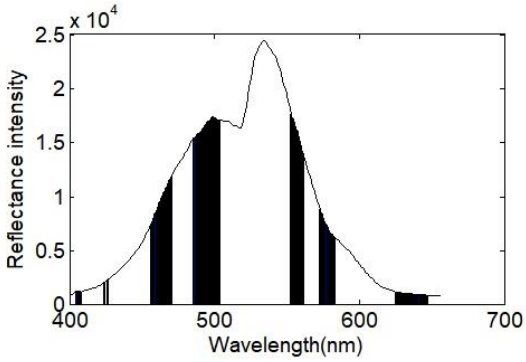


Fig. 6 - Distribution of the GA-selected effective wavelength variables

The prediction accuracy has been improved by 5.6% compared with the LS-SVM model established by using the full-band reflectance spectra. These results also proved that GA is an effective method for extracting effective feature wavelength variables and improving the prediction accuracy of the model.

Table 1

Statistics of the SSC real values

Set	Amount of sample	Maximum	Minimum	Mean	SD
Calibration set	86	13.8	8.3	10.5	1.23
Prediction set	28	13.5	7.9	10.4	1.21

Table 2

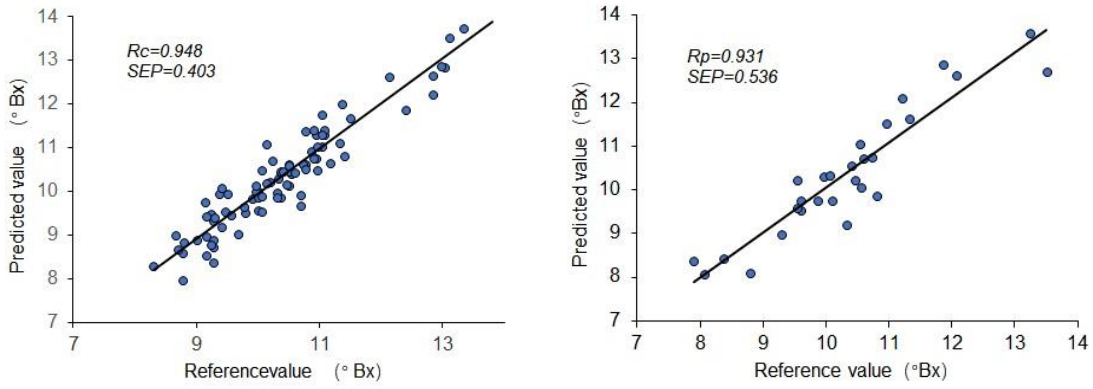
Prediction SSC values of different preprocessing methods and models

Preprocessing method	PLSR				MLR				LS-SVM			
	Rc	SEC	Rp	SEP	Rc	SEC	Rp	SEP	Rc	SEC	Rp	SEP
SG	0.846	0.702	0.844	0.783	0.837	0.715	0.834	0.872	0.855	0.662	0.835	0.876
MSC	0.865	0.633	0.822	0.818	0.866	0.675	0.845	0.783	0.877	0.648	0.799	0.920
SG +MSC	0.886	0.602	0.855	0.764	0.872	0.670	0.869	0.736	0.893	0.621	0.882	0.752

Table 3

Prediction results of GA–PLSR, GA–MLR, and GA–LS-SVM models

Run Times	GA–PLSR				GA–MLR				GA–LS-SVM			
	R_c	SEC	R_p	SEP	R_c	SEC	R_p	SEP	R_c	SEC	R_p	SEP
1	0.913	0.488	0.911	0.606	0.905	0.539	0.895	0.641	0.913	0.495	0.890	0.633
2	0.912	0.499	0.912	0.596	0.917	0.498	0.902	0.738	0.919	0.504	0.917	0.566
3	0.925	0.473	0.915	0.565	0.912	0.521	0.892	0.662	0.920	0.515	0.918	0.635
4	0.916	0.485	0.909	0.703	0.906	0.570	0.886	0.715	0.927	0.439	0.911	0.547
5	0.916	0.498	0.914	0.681	0.920	0.505	0.894	0.674	0.948	0.403	0.931	0.536



(a) Calibration set **(b) Prediction set**

Fig. 7 - Prediction SSC using GA–LS-SVM model by Vis/NIR hyperspectral imaging

CONCLUSIONS

GA-LS-VM prediction model based on Vis/NIR hyperspectral imaging was successfully utilized for non-destructive detection SSC of kiwifruit. The different preprocessing methods and prediction models were studied. GA was applied to select the effective feature wavelength variables to improve the precision of the prediction model. The conclusions are as follows:

(1) The GA is an effective method for selecting effective feature wavelength variables. The Vis/NIR hyperspectral imaging with GA can non-destructively determine the SSC of kiwifruits.

(2) The SG+MSC is the best pretreatment method among SG, MSC and SG+MSC in improving spectral quality.

(3) The precisions of the prediction models established with the effective feature wavelength variables selected are higher than that established with full wavelength variables. GA-LS-SVM is the best prediction model of kiwifruit SSC among PLSR, MLR, and LS-SVM. It shows a better accuracy with R_C = 0.948, SEC = 0.403 °Bx, R_P = 0.931, and SEP = 0.536 °Bx.

This study offers a non-destructive and rapid method of predicting kiwifruit SSC using Vis/NIR hyperspectral imaging with GA. However, this detection technique also has certain limitations, for example, it may need to construct different prediction models or modify the models for different varieties of fruit. When fruit farmers apply this technology, the prediction results may have errors due to the influence of temperature environment, and the prediction model needs to be constantly revised. In future studies, prediction models for different varieties and different internal quality indicators of kiwifruit should be further established, and detection methods should be further promoted and applied through cooperation with farms and factories.

ACKNOWLEDGEMENT

The authors grateful thank the financial support provided by the Science and Technology Research Project of Henan Province (Project No.222102110262)

REFERENCES

[1] Berardinelli A., Benelli A., Tartagni M., Ragni L. (2019). Kiwifruit flesh firmness determination by a NIR sensitive device and image multivariate data analyses. *Sensors and Actuators A*, 296, 265–271. https://doi.org/10.1016/j.sna.2019.07.027

[2] Chen H. ZH, Pan T., Chen J. M., Lu Q. P. (2011). Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods. *Chemometrics and Intelligent Laboratory Systems*, 107, 139-146. http://dx.doi.org /10.1016/j.chemolab.2011.02.008

[3] Christophe L., Olivier K., Yves D., Christophe L. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*, 74, 150-156. http://dx.doi.org/10.1016/j.jesp.2017.09.011

[4] Ebrahiema A., Olaniyi A. F., Lembe S. M., Umezuruike L. O. (2018). Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review. *Journal of Food Engineering*, 217, 11-23. http://dx.doi.org/10.1016/j.jfoodeng.2017.08.009

[5] Eunhee P., Yaguang L., Sasha C. M., Kathryne L. E., Shirley A. M., Samantha B., John S. (2018). Consumer preference and physicochemical evaluation of organically grown melons. *Postharvest Biology and Technology*, 141, 77–85. https://doi.org/10.1016/j.postharvbio.2018.03.001

[6] Hamid G., Golasa M., Hakimeh F. (2012). Spectroscopic studies on Solvatochromism of mixed-chelate copper (II) complexes using MLR technique. *Spectrochimica Acta Part A*, 85, 25-30. http://dx.doi.org /10.1016/j.saa.2011.08.042

[7] Hu W. H., Sun D. W., Jose B. (2017). Rapid monitoring 1-MCP-induced modulation of sugars accumulation in ripening 'Hayward' kiwifruit by Vis/NIR hyperspectral imaging. *Postharvest Biology and Technology*, 125, 168-180. http://dx.doi.org/10.1016/j.postharvbio.2016.11.001

[8] Jiang B., He J. R., Yang SH. Q., Fu H. F., Li T., Song H. B., & He D. J. (2019). Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues. *Artificial Intelligence in Agriculture*, 1, 1-8. https://doi.org/10.1016/j.aiia.2019.02.001

[9] Leardi R., & Amparo L. G. (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometrics and Intelligent Laboratory Systems*, 41,195-207.

[10] Li J.B., Chen L.P., & Huang W. Q. (2018). Detection of early bruises on peaches (*Amygdalus persica L.*) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biology and JeanTechnology*, 135, 104–113. http://dx.doi.org/10.1016/j.postharvbio.2017.09.007

[11] Li J. B., Zhang H. L., Zhan B. SH., Wang ZH. L, & Jiang Y. L. (2019). Determination of SSC in pears by establishing the multi-cultivar models based on visible-NIR spectroscopy. *Infrared Physics and Technology*, 102, 103066. https://doi.org/10.1016/j.infrared.2019.103066

[12] Ma SH. B. (2021). Nondestructive Determination of Kiwifruit SSC using Visible/Near-Infrared Spectroscopy with Genetic Algorithm. *Journal of Engineering Science and Technology Review*, 14 (1), 100-106. https://doi.10.25103/jestr.141.11

[13] Ma T., Xia Y., Tetsuya I., & Satoru T. (2021). Non-destructive and fast method of mapping the distribution of the soluble solids content and pH in kiwifruit using object rotation near-infrared hyperspectral imaging approach. *Postharvest Biology and Technology,* 174, 111440. https://doi.org/10.1016/j.postharvbio.2020.111440

[14] Nturambirwe J.F.I., Nieuwoudt H.H., Perold W.J.P., Opara U.L. (2019). Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization. *Scientific African*, 3, e00051. https://doi.org/10.1016/j.sciaf.2019.e00051

[15] Reddy R. Pullanagari, S. R. O., Mo L. & Research O. (2021). Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. *Journal of Food Engineering,* 289, 110177. https://doi.org/10.1016/j.jfoodeng.2020.110177

[16] Rosalba C., Jose M. A., & Alessandro U. (2017). Transferring results from NIR-hyperspectral to NIR-multispectral imaging systems: A filter-based simulation applied to the classification of Arabica and Robusta green coffee. *Analytica Chimica Acta*, 967, 33-41. http://dx.doi.org/10.1016/j.aca.2017.03.011

[17] Santosh S., Matej K., Uros Z., Lise C. D., & René G. (2016). Single seed near-infrared hyperspectral imaging in determining tomato (Solanum lycopersicum L.) seed quality in association with multivariate data analysis. *Sensors and Actuators B: chemical*, 237, 1027–1034. http://dx.doi.org/10.1016/j.snb.2016.08.170

[18] Tom F., Cecilia R., Ana G., & José E. G. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, 96, 22-26. http://dx.doi.org /10.1016/j.chemolab.2008.11.006

[19] Umer M. J., Lei G., Haileslassie G., Luqman B. S., Pingli Y., Zhao SH. J., Lu X. Q., He N., Zhu H. J., & Liu W. G. (2020). Expression pattern of sugars and organic acids regulatory genes during watermelon fruit development. *Scientia Horticulturae,* 265, 109102. https://doi.org/10.1016/j.scienta.2019.109102

[20] Xu J. C., Ren Q. W., & Shen ZH. ZH. (2015). Prediction of the strength of concrete radiation shielding based on LS-SVM. *Annals of Nuclear Energy*, 85, 296-300. http://dx.doi.org/10.1016/j.anucene.2015.05.030

[21] Xu L. J., Zhang L. N., Huang P., Chen H., & Kang ZH. L. (2020). Detection of kiwifruit dry matter content based on hyperspectral technology using uninformed variable elimination coupled with successive projection algorithm. *DYNA-Ingenieria e Industria,* 95(6), 654-660. https://doi.org/10.6036/9837

[22] Zheng W., Bai Y. H., Luo H., Li Y. H., Yang X., & Zhang B. H. (2020). Self-adaptive models for predicting soluble solid content of blueberries with biological variability by using near-infrared spectroscopy and chemometrics. *Postharvest Biology and Technology*, 169, 111286. https://doi.org/10.1016/j.postharvbio.2020.111286