



OPEN

Quantitative measurement of internal quality of carrots using hyperspectral imaging and multivariate analysis

Arcel Mutombo Mulowayi^{1,4}, Zhen Hui Shen^{1,2,4}, Witness Joseph Nyimbo³, Zhi Feng Di^{1,4}, Nyumah Fallah³ & Shu He Zheng^{1,4}✉

The study aimed to measure the carotenoid (Car) and pH contents of carrots using hyperspectral imaging. A total of 300 images were collected using a hyperspectral imaging system, covering 472 wavebands from 400 to 1000 nm. Regions of interest (ROIs) were defined to extract average spectra from the hyperspectral images (HSI). We developed two models: least squares support vector machine (LS-SVM) and partial least squares regression (PLSR) to establish a quantitative analysis between the pigment amounts and spectra. The spectra and pigment contents were predicted and correlated using these models. The selection of EWs for modeling was done using the Successive Projections Algorithm (SPA), regression coefficients (RC) from PLSR models, and LS-SVM. The results demonstrated that hyperspectral imaging could effectively evaluate the internal attributes of carrot cortex and xylem. Moreover, these models accurately predicted the Car and pH contents of the carrot parts. This study provides a valuable approach for variable selection and modeling in hyperspectral imaging studies of carrots.

Keywords Hyperspectral imaging, Internal attribute evaluation, Carrot, Variable selection, Quantitative analysis model

Carrot (*Daucus carota* L.) is a widely consumed root vegetable crop known for its high nutritional value, including essential micronutrients such as vitamins A and C^{1,2}. Carrot production is rising worldwide, with China leading the way as the top producer³. Although carrots are typically orange, they also exhibit a range of other colors including purple, red, and yellow, thereby enriching the diversity within the spectrum⁴. Moreover, these crops provide significant amounts of antioxidants, provitamin A, and carotenoids, which have been linked to various health benefits, including a lower risk of prostate cancer and improved heart and liver health⁵⁻⁸.

With its unique pH value, carrot juice is susceptible to spoilage and pathogenic organisms⁹. Key quality indicators for carrots include factors like color, absence of bruises, provitamin A content, vitamin C levels, and firmness, all of which impact shelf life, market value, and consumer satisfaction¹⁰. Carrots' shelf life, selling price, and customer satisfaction depend on their quality. Enhancing carrot quality inspection and developing rapid quality control technologies that give precise and detailed information about nutritional content is crucial, given rising consumption and the effects of climate change^{11,12}. This information can be utilized to ascertain the most suitable time for harvesting, refine storage parameters, and enhance the nutritional quality of processed carrot derivatives.

The simultaneous collection of spectral and image data from the tested sample using hyperspectral imaging (HSI) merges conventional spectroscopy and digital imaging technology into a system¹³⁻¹⁵. HSI technology is used in various industries, including agriculture, food¹⁶, environmental management, and urban planning. It can provide substantial information in spectral and spatial domains¹⁷. In recent years, HSI technology has played a pivotal role in detecting the internal quality of agricultural products, ranging from moisture and starch detection contents¹⁸ to protein and fat analysis. Furthermore, HSI has also been leveraged to investigate crop diseases¹⁹,

¹College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China. ²Engineering College, Fujian Jiangxia University, Fuzhou 350108, China. ³Fujian Provincial Key Laboratory of Agro-Ecological Processing and Safety Monitoring, College of Life Sciences, Agriculture and Forestry University, Fuzhou 350002, China. ⁴Fujian University Engineering Research Center for Modern Agricultural Equipment, Fuzhou 350002, China. ✉email: zsh@fafu.edu.cn

nutrient deficiency²⁰, and estimating biochemical and biophysical characteristics essential for understanding vegetable physiological status and predicting crop yields. Moreover, this tool can investigate soil properties, including moisture content, organic matter, and carbon content^{21,22}, total capsaicinoids²³, and pH²⁴. Munera et al.^{25,26} mentioned that the evaluation of fruit quality is a recently developed application. For instance, studies on the quality detection of bakery goods, meat, and fresh vegetables have already been published²⁷.

Research has shown that visible/near-infrared HSI technology has been extensively employed in the non-destructive assessment of interior fruit attributes, including soluble solids content (SSC) and firmness. Nevertheless, the current research on predicting Car and pH content in various regions of fruits, such as the cortex and xylem, is limited from a scientific standpoint. To comprehensively evaluate the internal quality attributes of carrots, this study aimed to investigate the potential of hyperspectral reflectance imaging for predicting the Car and pH content of carrots. We sought to investigate these parameters in two central regions of carrots (cortex and xylem) using visible and near-infrared (Vis/NIR) HSI. The specific objectives of this study were to:

1. Acquire hyperspectral images of carrot samples and extract spectral data from them.
2. Build partial least squares regression (PLSR) and least squares support vector machine (LS-SVM) models using the entire spectrum.
3. Choose representative wavelengths using successful projection algorithms (SPA) and regression coefficients (RC) from PLSR.
4. Develop simplified LS-SVM and PLSR models.
5. Use the best model to predict the quality attributes of each sample pixel and compare its performance to Fig. 1.

Methods

Sample preparation

A stratified sampling approach was applied to select carrot samples for analysis. A comprehensive collection of 300 carrot samples, exhibiting comparable shape and size, was procured from Putian (Pt) and Fuzhou (Fz) City, located within the geographical boundaries of Fujian Province, China. Each sample weighted between 55 to 65 g.

Following an exhaustive washing process, carrots that exhibited cracks, rust, dysmorphia, or dark discoloration were excluded from the sample set. As a result, 240 samples remained, all meeting the predefined quality criteria. Among the samples selected for investigation, 120 carrots were sourced from the Fz, while the remaining 120 originated from the Pt. The carrots were stored in a sealed plastic bag at 3 °C for 2 days. Later, each carrot was divided into two halves to investigate Car and pH contents. The plant material used in this work complies with relevant institutional, national, and international guidelines and legislation.

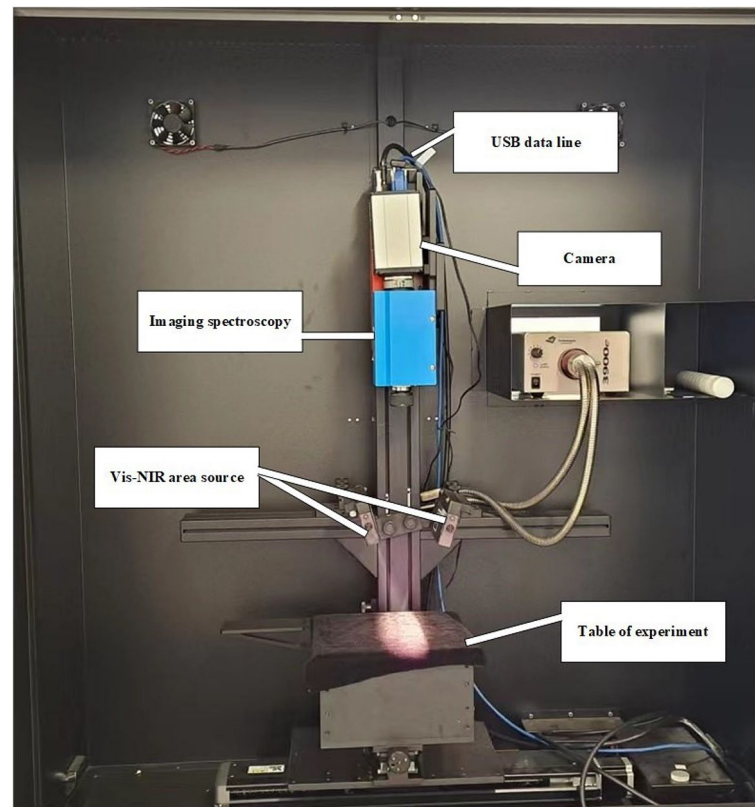


Figure 1. Hyperspectral imaging system.

HSI system and image acquisition

Experimental HSI was conducted using a high-performance CCD digital camera (Sencicam QE Taiwan) and a hyperspectral camera (HIS-V10E-sCMOS) that covered the wavelength range of 400–1000 nm with a spectral resolution of 2–8 nm. The system was equipped with Oriol Instruments USA halogen tungsten light bulbs, a spatial resolution point radius of 9 m, a light source supply system with a feedback controller, and a computer. The camera was operated using Camera Control Kit V219. We used this camera to capture hyperspectral images. The camera consisted of a focal length of 170mm, a scanning line distance of 2mm, and a light source beam's optical center located 2mm from the scanning line. We integrated data from four evenly spaced places over the equator using a 22-binning technique to provide a full spectral image to conduct the HSI of the carrots. See Fig. 1 for a graphic depiction of the HSI equipment.

Image processing

One of the most important steps in pre-processing the hyperspectral images was calibrating the raw data to exclude dark current effects from the CCD camera. After calibration, an area of interest (ROI) was found in the calibrated images, and spectrum data was then taken out of these ROIs, as Fig. 2 shows. To reduce differences caused by illumination, detector sensitivity, camera specs, and subtleties in the physical setup, raw hyperspectral photos were corrected by comparing them to black-and-white reference images²⁸.

The camera lens was covered with its opaque cap, and the light source was turned off to provide a black reference image. Alternatively, a spectral image of a uniformly white tile with approximately 99.9% reflectance was captured to create a white reference image²⁹. The following equation was used to adjust the uncorrected hyperspectral pictures:

$$R = \frac{I - I_d}{I_w - I_d} \quad (1)$$

Here, R represents the corrected hyperspectral image, while I represents the sample's initial spectral image, I_d denotes the dark reference image and I_w stands for the white reference image. We leveraged image acquisition software to correct the image.

Spectral pre-treatment

The hyperspectral data were extracted from the acquired ROIs for spectral processing. Undesired variations were compensated (negative effects from random and systematic noise), and unnecessary or noisy wavelengths were removed to improve prediction accuracy. Pre-treatment was applied to the spectral data in the form of operations, including smoothing, derivatives, multiplicative scatter correction (MSC)^{14,15}, standard normal variate (SNV), and Savitzky–Golay (SG)³⁰. The SG smoothing method with a window width of three points was used to reduce high-frequency noise, baseline excursion, and dispersion to stabilize the baseline and reduce noise.

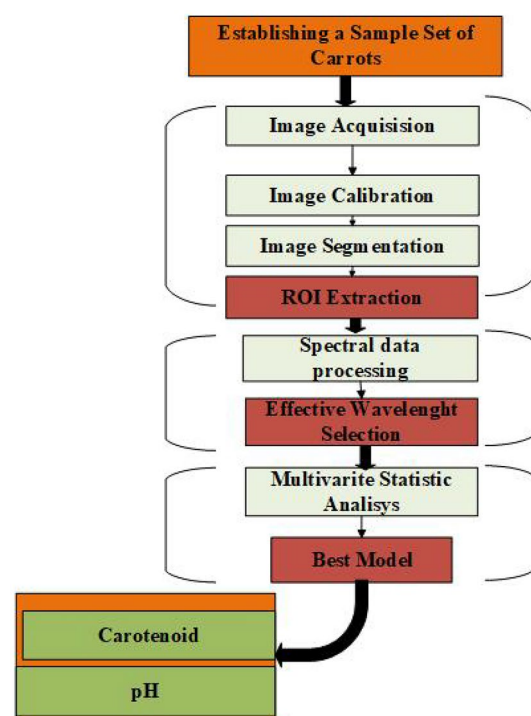


Figure 2. Flow diagram of the experimental steps.

Moreover, MSC was applied to adjust for additive and multiplicative scatter effects, which improved and corrected the obtained hyperspectral data.

The SG filter³¹ is a widely used technique for smoothing data based on approximating the raw data using polynomials in a defined data frame. The SG filter has two degrees of freedom, including polynomial order and window length. The first parameter enables the smoothed data to follow the raw data as closely as possible. This process demonstrates the importance of preserving the edges of the data. However, it also entails the drawback of tracking noise fluctuations. The window length neutralizes the high-frequency noise contribution for the second degree of freedom by smoothing its fluctuations through polynomial fitting^{32,33}. The SG filter searches for the optimal $n + 1$ polynomial coefficients for a given n -degree polynomial to best suit the raw data and assesses the outcome in the window center^{34,35}. The polynomial function was applied to the signal point by point. The measured value of the window's midpoint was replaced with the polynomial function's estimated value. The degree of smoothing was altered by changing the window's width and polynomial order. In addition to SG smoothing, other spectral pre-treatment methods, such as MSC and SNV, are commonly used to compensate for undesired variations and remove unnecessary or noisy wavelengths^{14,15,30}.

Effective wavelength selection methods

The spectrum data set may comprise thousands of variables/wavelengths and hundreds or thousands of samples^{36,37} due to the high resolution of modern spectroscopic instruments. Such large-scale data can make hyperspectral image inspection techniques more time-consuming. Moreover, variable selection (wavelength selection) is crucial in identifying the relevant variables and eliminating highly correlated ones to reduce computational complexity, increase detection effectiveness, and meet the industry-required inspection speed^{38,39}. While no definitive method has been established for selecting optimal wavelengths, various approaches have been recommended⁴⁰. For instance, SPA, RC, uninformative variable elimination (UVE), simulated annealing (SA), K-nearest neighbors regression (K-NNR), and genetic algorithm (GA) are a few multivariate algorithms that have been suggested for developing quantitative models.

In this study, the wavelength selection techniques utilized included RC, K-NNR, and SPA. SPA identified wavelengths with the least redundant information. SPA has been described as a method for identifying relevant features in a forward direction by comparing projection vectors resulting from projecting wavelengths onto other wavelengths. It chooses the most significant projection vector wavelength and incorporates it into the candidate subset of characteristic wavelengths. Studies have described SPA as a method that identifies relevant features in a forward direction by comparing projection vectors resulting from projecting wavelengths onto other wavelengths. It selects the most significant projection vector wavelength and includes it in the candidate subset of characteristic wavelengths⁴¹. Here, the performance of different subsets was evaluated using a regression model. SPA aims to identify a combination of variables that contains the least redundant information and the least covariance, thereby reducing model complexity and improving accuracy. Overall, SPA is a useful tool for feature selection in various applications, such as regression, classification, and data mining.

It has been established that RC plays a decisive role in creating a predictive model for specific data collection. Weighted RC, also known as b-coefficients, which are equivalent to the model with full spectra, are used to calculate RC. The best wavelengths are determined by selecting those with the highest absolute b-coefficient values. This approach enables the identification of the most crucial wavelengths for forecasting the response variable, leading to a more accurate and effective model⁴². The use of fewer wavelengths in spectral analysis has the potential to improve model performance^{24,24}. The method representing a small number of wavelengths, RC, and SPA, was chosen for modeling following the selection of EWs.

Modeling methods and model evaluation

Model validation is an important step in multivariate data analysis. The prediction model for this study was constructed utilizing PLSR and LS-SVM, which are linear multivariate algorithms. This is because of its efficacy when a linear relationship exists between spectra and object properties^{43–45}. PLSR is widely employed in chemometrics to analyze the correlation between spectral data and reference quality indicators. A set of statistically uncorrelated latent variables was utilized by the PLSR model to forecast Car and pH levels. Through decomposition, this method generates principal factors from the independent and dependent variables as they are projected into a new multidimensional space. Seven PLSR factors were chosen for this investigation according to the correlation strength of the principal factors. Notably, PLSR and LS-SVM were applied to the prediction model. However, PLSR was solely utilized to model the full spectra⁴⁴.

Based on concepts from statistical learning theory, SVM can be used for classification and nonlinear regression. LS-SVM is an enhancement of traditional SVM. It uses least-squares linear systems as the loss function rather than traditional convex quadratic programming⁴⁶. LS-SVM is more than SVM because of its low computational complexity and efficiency.

$$y(x) = \sum_{k=1}^N a_k K(x, x_k) + b \quad (2)$$

In the given context, $K(x, x_k)$ represents the kernel function, x_k indicates the input vectors, k denotes the support values, and b indicates the bias factor. The computation of similarity between the input vectors is the responsibility of the kernel function, and the kernel function selection influences the efficacy of the model.

Furthermore, a correlation analysis was conducted to assess the RC of the simplified models, investigating the association between the EWs and the quality features.

Model evaluation

The prediction capacities of the models were assessed by calculating statistical metrics, including the coefficient of determination of calibration (R^2_{cal}), coefficient of determination of prediction (R^2_{pre}), root mean square error (RMSEC, RMSEP), and RPD can be described as follows:

$$R^2 = \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y} - y_i)^2} \quad (4)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}} \quad (5)$$

where m represents the number of samples, \hat{y}_i represents the predicted value, y_i represents the actual value and \bar{y} represents the mean value of the actual value. SD is the standard deviation of the validation sample.

When the RPD value is greater than 2.5, it indicates a high capacity for prediction⁴⁷. Spectral data extraction was conducted on ENVI 4.8 (ITT, Visual Information Solutions, Boulder, USA). All computations and multivariate data analyses were performed with chemometric software Unscrambler[®] 9.7 (CAMO AS, Oslo, Norway) and MATLAB R 2009b (The Math Works, Natick, USA).

Biochemical analyses

After acquiring hyperspectral images, the samples were immediately sliced and weighed for subsequent chemical analysis. Each measurement was performed three times⁴⁸. We used 0.1 g of fresh-weight material immersed in a 20 ml solution containing 80% acetone and 100% ethanol (1:1 ratio) for 24 h in darkness to extract the pigments. The pH composite electrode was mixed in pure water and then shaken dry after being thoroughly washed. The pH meter was placed into the 4.00 pH calibration solution to calibrate it. Once the calibration was finished, the meter was rinsed with distilled water and dried. The pH meter was calibrated using a standard buffer solution with a pH of 7.00^{49,50}. The meter was cleaned with pure water, dried, and calibrated using a pH 9.18 solution. The three-point calibration has been accomplished at this stage. A sufficient amount of pulp was extracted from each sample, squeezed to obtain juice, and then the electrode was immersed in the juice to measure the pH value. Next, the electrode was immersed in the juice, and the pH value was measured. Each sample underwent three measurements following the described procedure. The average of the three readings was considered for the pH value.

We measured the Car levels with a 752UV/Vis spectrophotometer and determined based on fresh weight using standard techniques^{51,52}.

Results

Hyperspectral reflectance spectra

Figure 3a shows how to identify the carrot region of interest. Figure 3b and c show the 400–1000 nm xylem and cortex spectral of the Fz carrot cultivar, respectively. These spectra were taken from the hyperspectral image of calibration set samples. It is evident that the spectra from all sides follow the same pattern across the whole wavelength range, but there were some notable deviations. The spectral curves exhibited distinct absorption and reflection peaks, as can be seen in Fig. 4. The reflectivity of the Fz-xylem and Pt-xylem side is slightly higher than that of the Fz-cortex and Pt-cortex side within the visible light range of 420 to 680 nm, which was based on the spectral images obtained from the Fz-xylem and Pt-xylem, as well as the Fz-cortex and Pt-cortex side. However, the reflectivity of the Fz-xylem and Pt-xylem side significantly increases compared to the Fz-cortex and Pt-cortex side within the near-infrared range of 780 to 1000 nm.

A typical Car absorption band at 680nm corresponds to the first discernible absorption peak. Around 750 nm is the peak of the second absorption center, and a relatively wide absorption band is connected to the band C–H's fourth overtone. The second overtone of band O–H may be related to the tiny absorption band at 950 nm⁵³.

In addition to the typical absorption characteristics, the spectral intensities of different samples were different, indicating differences in chemical components, which was conducive to constructing the Car and pH quantitative analysis model.

PLSR models based on the full spectra

We leveraged PLSR to establish regression models with the xylem and cortex datasets. The regression results are shown in Tables 1 and 2. PLSR models, the samples were taken in the same order on the carrot xylem and cortex side. The calibration and prediction sets for both regions of the carrot were also the same.

As indicated by the R^2_{pre} , RMSEP, and RDP values, the results displayed in Table 1 illustrate the ability to predict carotenoid quality and pH in the xylem and bark regions of the Fz-Pt cultivar. The RMSEP values for Fz and Pt in the xylem region were 0.026 and 0.027, respectively, while the R^2_{pre} values for predicting carotenoid quality ranged from 0.903 to 0.915 for Fz and from 0.885 to 0.876 for Pt. Furthermore, the region where Fz and Pt had RDP values of 2.19 and 2.21, respectively, demonstrated a greater ability to predict carotenoid quality. In contrast, the R^2_{pre} values obtained to infer pH in the xylem and cortex regions were comparatively lower, ranging from 0.666 to 0.702. Furthermore, RMSEP values ranged from 0.022 to 0.035. All RDP values were less than 2, indicating a satisfactory level of predictive accuracy despite the lower R^2_{pre} values. It was also observed that the

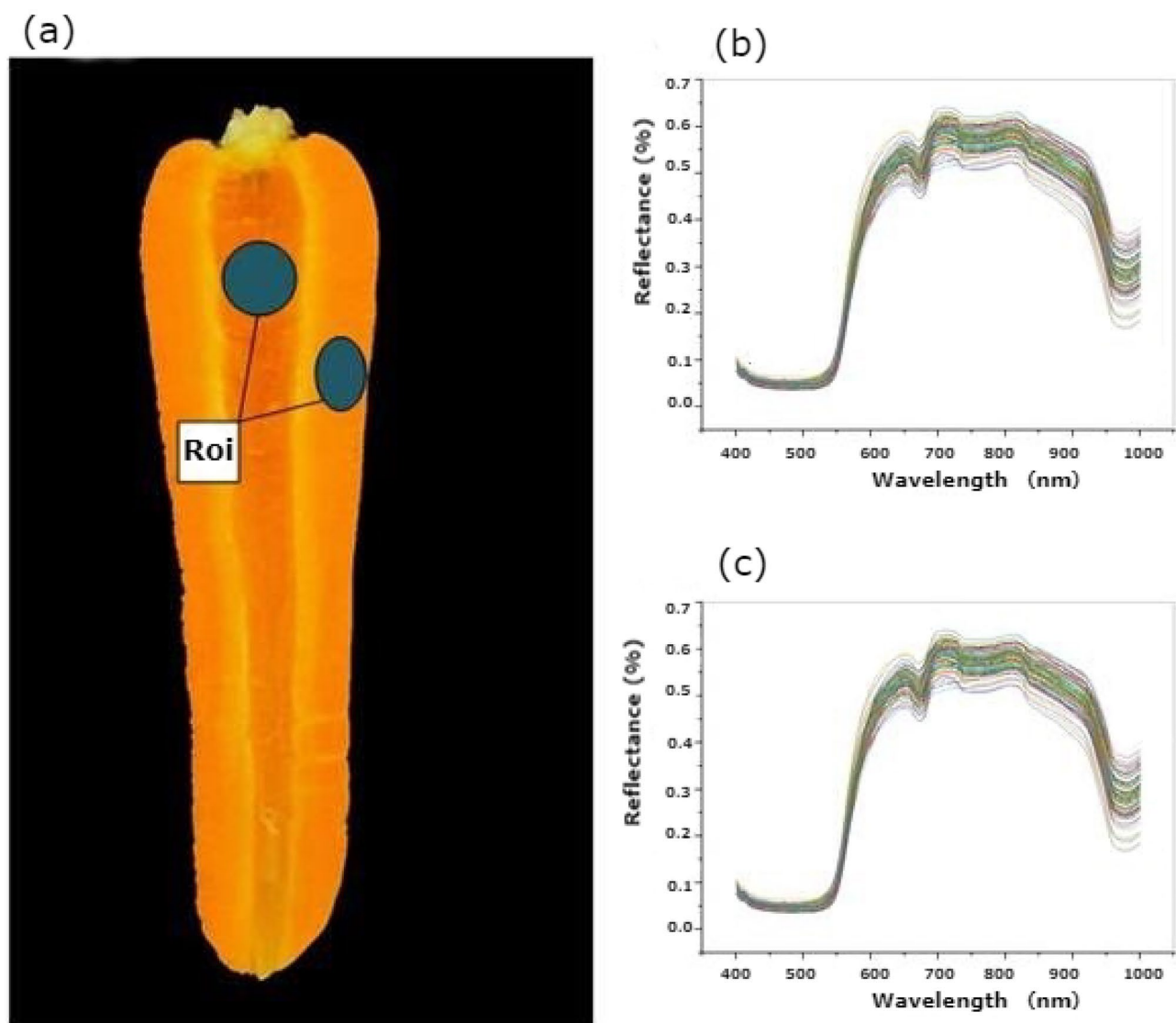


Figure 3. Main steps for image and spectra processing of carrot: (a) identification of the Region of Interest (ROI), (b) Fz-cortex raw mean reflectance spectrum, and (c) Fz-xylem raw mean reflectance spectrum.

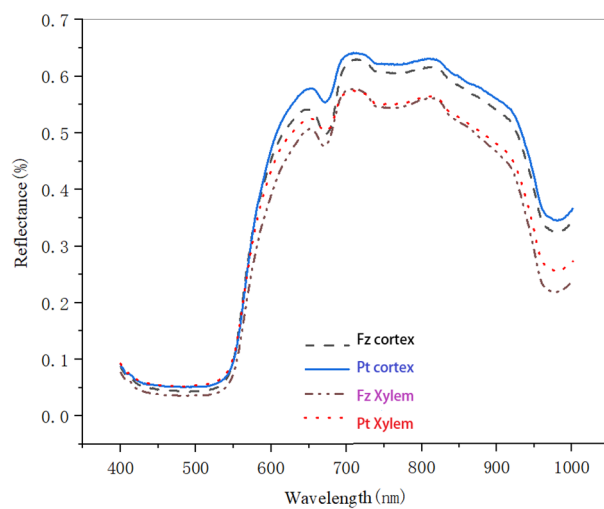


Figure 4. Average spectral curves of cortex and xylem of the Fz and Pt carrot cultivar.

Cultivar	Region	Calibration set		Prediction set		
		R ² _{Pre}	RMSEC	R ² _{Pre}	RMSEP	RPD
Fz	Xylem	0.924	0.021	0.903	0.026	2.19
Fz	Cortex	0.897	0.023	0.885	0.025	1.61
Pt	Xylem	0.930	0.022	0.915	0.027	2.21
Pt	Cortex	0.906	0.025	0.876	0.030	1.72

Table 1. Results and parameters of the calibration and prediction sets of Car by partial least squares regression (PLSR) models.

Cultivar	Region	Calibration set		Prediction set		
		R ² _{Pre}	RMSEC	R ² _{Pre}	RMSEP	RPD
Fz	Xylem	0.799	0.024	0.666	0.022	1.54
Fz	Cortex	0.841	0.024	0.682	0.025	1.72
Pt	Xylem	0.854	0.023	0.674	0.027	1.66
Pt	Cortex	0.871	0.024	0.702	0.035	1.80

Table 2. Results and parameters of the calibration and prediction sets of pH by PLSR models.

RDP values in the cortex region were slightly higher than those in the xylem region. This discrepancy implies that pH prediction performance was significantly improved in the cortex region.

Selection of effective wavelengths

Choosing a configuration with fewer wavebands is recommended to enhance the stability and integrability of a multispectral imaging system from a scientific standpoint (ElMasry et al. 2019⁵⁴). SPA was used to identify the EWs carrying crucial information for determining scaling rates and reducing data dimensionality. These EWs remove unnecessary information by including the whole spectral data range (400–1000 nm), representing the most important data among the EWs. Table 3 demonstrates that only the important wavelengths are required to estimate Car and pH. The pH decreased the number of wavelengths from 5 to 9, contrasting the pH range (8 to 14) detected in the xylem of both cultivars, which showed a wavelength range from (Table 3).

Prediction of pH

Table 3 illustrated that the xylem spectra had less spectral than the cortex spectra. However, their predictive capability was much better.

Besides, Table 4 showed that the prediction under the RC-PLSR model had an R²_{Pre} of 0.672 and an RMSEP of 0.030, while the counterpart had an R²_{Pre} of 0.752 and an RMSEP of 0.029. However, the RC-LS-SVM model

Cultivar	Parameter	Selected range (nm)	Methods	Region	Number	Selected EWs (nm)
Fz	Car	400–1000	RC	Xylem	10	910, 643, 441, 561, 628, 880, 664, 731, 450, 945
				Cortex	7	410, 580, 671, 900, 950, 660, 730
Pt	Car	400–1000	RC	Xylem	14	994, 444, 561, 594, 628, 880, 577, 634, 658, 660, 730, 739, 740, 765
				Cortex	10	775, 410, 580, 670, 952, 950, 662, 685, 704, 736
Fz	Car	400–1000	SPA	Xylem	11	902, 440, 560, 681, 594, 625, 880, 958, 956, 828, 691
				Cortex	8	489, 522, 555, 584, 582, 671, 911, 953
Pt	Car	400–1000	SPA	Xylem	14	994, 574, 444, 561, 580, 628, 881, 577, 634, 659, 662, 730, 739, 765
				Cortex	13	543, 595, 674, 775, 410, 673, 952, 952, 662, 685, 699, 704, 736
Fz	pH	400–1000	RC	Xylem	8	441, 561, 594, 628, 880, 994, 674
				Cortex	7	410, 580, 671, 900, 950, 704, 735
Pt	pH	400–1000	SPA	Xylem	8	441, 561, 594, 628, 880, 994
				Cortex	5	410, 580, 671, 900, 950
Pt	pH	400–1000	SPA	Xylem	8	443, 556, 592, 628, 880, 992, 674, 710
				Cortex	9	410, 714, 721, 580, 673, 900, 950, 702, 734

Table 3. The selected EWs for Car and pH using RC and SPA.

Cultivar	Models	Region	Calibration set		Prediction set		
			R ² _{Pre}	RMSEC	R ² _{Pre}	RMSEP	RPD
Fz	RC-PLSR	Xylem	0.901	0.025	0.672	0.030	1.67
Fz	RC-PLSR	Cortex	0.961	0.021	0.752	0.029	1.72
Pt	RC-PLSR	Xylem	0.892	0.026	0.653	0.029	1.55
Pt	RC-PLSR	Cortex	0.924	0.022	0.711	0.030	1.70
Fz	RC-LS-SVM	Xylem	0.876	0.024	0.701	0.032	1.56
Fz	RC-LS-SVM	Cortex	0.957	0.030	0.802	0.026	1.92
Pt	RC-LS-SVM	Xylem	0.861	0.023	0.696	0.030	1.57
Pt	RC-LS-SVM	Cortex	0.946	0.030	0.812	0.025	2.00
Fz	SPA-PLSR	Xylem	0.923	0.027	0.678	0.031	1.61
Fz	SPA-PLSR	Cortex	0.964	0.024	0.841	0.024	2.08
Pt	SPA-PLSR	Xylem	0.933	0.025	0.681	0.031	1.62
Pt	SPA-PLSR	Cortex	0.953	0.024	0.791	0.025	2.11
Fz	SPA-LS-SVM	Xylem	0.826	0.025	0.731	0.030	1.67
Fz	SPA-LS-SVM	Cortex	0.954	0.023	0.816	0.028	1.79
Pt	SPA-LS-SVM	Xylem	0.841	0.024	0.742	0.029	1.81
Pt	SPA-LS-SVM	Cortex	0.970	0.024	0.820	0.026	2.40

Table 4. LS-SVM and PLSR models calibration and prediction of pH using EWs (RC and SPA).

had an R²_{Pre} of 0.701 and an RMSEP of 0.032, while the counterpart had an R²_{Pre} of 0.802, an RMSEP of 0.026, an R²_{Pre} of 0.757, and an RMSEP of 0.024.

The SPA-PLSR had an R²_{Pre} of 0.678 and an RMSEP of 0.031, while the SPA-LS-SVM had an R²_{Pre} of 0.731 and an RMSEP of 0.030, and the counterpart had an R²_{Pre} of 0.816 and an RMSEP of 0.028. The identical outcome observed in the Fz sample was found to be applicable in the Pt samples, as evidenced by the data presented in Table 4. Our findings indicate that the spectral characteristics of the xylem were more effective in predicting Car than those of the cortex.

Prediction of carotenoid

Here, we used EWs to build Car and pH-predicting models in two carrot cultivars, namely Fz and Pt. The obtained results of the car prediction are shown for the Fz sample in Table 5.

Table 5 also showed that the Fz cultivar RC-PLSR model had an R²_{Pre} value of 0.892 and an RMSEP value of 0.023 in the xylem, slightly lower than its counterpart (using the cortex), with R²_{Pre} = 0.854 and RMSEP = 0.030. The RC-LS-SVM results showed an R²_{Pre} value of 0.933, an RMSEP value of 0.022, and an RPD of 2.27, while its counterpart had an R²_{Pre} value of 0.883 and an RMSEP value of 0.026, and an RPD of 1.8.

Cultivar	Models	Region	Calibration set		Prediction set		
			R ² _{Pre}	RMSEC	R ² _{Pre}	RMSEP	RPD
Fz	RC-PLSR	Xylem	0.937	0.022	0.892	0.023	2.17
Fz	RC-PLSR	Cortex	0.906	0.025	0.854	0.030	1.67
Pt	RC-PLSR	Xylem	0.941	0.021	0.890	0.024	2.16
Pt	RC-PLSR	Cortex	0.961	0.024	0.872	0.029	2.00
Fz	RC-LS-SVM	Xylem	0.961	0.023	0.933	0.022	2.27
Fz	RC-LS-SVM	Cortex	0.914	0.024	0.883	0.026	1.82
Pt	RC-LS-SVM	Xylem	0.957	0.024	0.878	0.022	2.13
Pt	RC-LS-SVM	Cortex	0.930	0.023	0.842	0.026	1.57
Fz	SPA-PLSR	Xylem	0.933	0.022	0.896	0.023	2.08
Fz	SPA-PLSR	Cortex	0.913	0.024	0.815	0.024	2.08
Pt	SPA-PLSR	Xylem	0.923	0.022	0.944	0.024	2.27
Pt	SPA-PLSR	Cortex	0.933	0.023	0.811	0.022	1.76
Fz	SPA-LS-SVM	Xylem	0.967	0.025	0.934	0.022	2.30
Fz	SPA-LS-SVM	Cortex	0.922	0.022	0.893	0.024	2.17
Pt	SPA-LS-SVM	Xylem	0.964	0.024	0.932	0.022	2.26
Pt	SPA-LS-SVM	Cortex	0.924	0.021	0.831	0.023	1.80

Table 5. LS-SVM and PLSR models calibration and prediction of pH using EWs (RC and SPA).

For the SPA-PLSR model, the R^2_{pre} value was 0.896, the RMSEP value was 0.023, and an RDP of 2.17, while its counterpart had an R^2_{pre} value of 0.815 and an RMSEP value of 0.024 and an RPD of 2.08. On the other hand, the LS-SVM model had an R^2_{pre} value of 0.934 and an RMSEP value of 0.022, with an RPD of 2.27, and its counterpart had an R^2_{pre} value of 0.893 and an RMSEP value of 0.024 and 2.08 for the RPD. Regarding prediction accuracy for Car content in both the calibration and prediction sets, the outcomes demonstrate that the LS-SVM models exhibited superior performance overall than the PLSR models. In contrast to the PLSR models, the LS-SVM models demonstrated superior RMSEC and RMSEP values.

The LS-SVM model exhibited notably robust outcomes for the Xylem region among the cultivars, whereas Fz maintained a consistently high performance across all models and regions. Conversely, the cultivar Pt obtained superior performance from the PLSR model in the Cortex region. As listed in Table 5, the Pt samples exhibited the identical pattern identified in the Fz sample. Superior suitability for Car prediction was observed in the xylem spectra compared with the cortex.

Figures 5 and 6 illustrate the optimal prediction outcomes according to the selected-range spectra.

Discussion

Analysis of characteristic wavelengths

Here, the spectral window ranged from 400 to 1000 nm. The outcomes of the wavelength selection are shown in Table 1. The EWs for Car were determined in the xylem and cortex, ranging from 410 and 956 nm, whereas the EWs for pH contents were between 500 and 900 nm. In related research, Car pigments were observed at wavelengths between 400 and 500 nm⁵⁵, 450 nm, and 580 nm^{25,26}, as well as 400–600 nm⁵⁶. Additional peaks were observed at the xylem area at 820 and 980 nm and in the cortex region at 814 and 970 nm. Additionally, acids were found to be present at 800 nm⁵⁶, and sugars were detected at 835 nm²⁴ and 840 nm⁵⁷. Therefore, the peak at 820 nm in the carrot xylem and 814 nm in the cortex could be related to acids and sugars in both cases. Water was detected at 960 nm⁵⁸, 970 nm⁵⁶, and 976 nm²⁴. Hence, the peaks reported at 980 nm and 970 nm may be attributed to water and sugars. Conversely, water and sugars have been observed at wavelengths of 970 nm^{59,60}, 960–980 nm⁶¹, and 970–980 nm⁶², respectively. In particular, some minor differences in wavelength reflectance were observed in the xylem and cortex.

Selection of effective wavelengths

The research emphasized the criticality of EW selection, improving the stability and integrability of multispectral imaging systems. The study employed SPA and RC to forecast quality metrics associated with Car and pH, as suggested in ElMasry et al. (2019⁵⁴) work for a system with a diminished quantity of wavebands. By the selection procedure, informative EWs containing crucial information for detecting scaling rates were discerned, resulting in the compilation of a more efficient dataset comprising the most valuable spectral data. The 400–1000 nm wavelengths were considered essential for predicting Car and pH characteristics. Notably, the chosen EWs

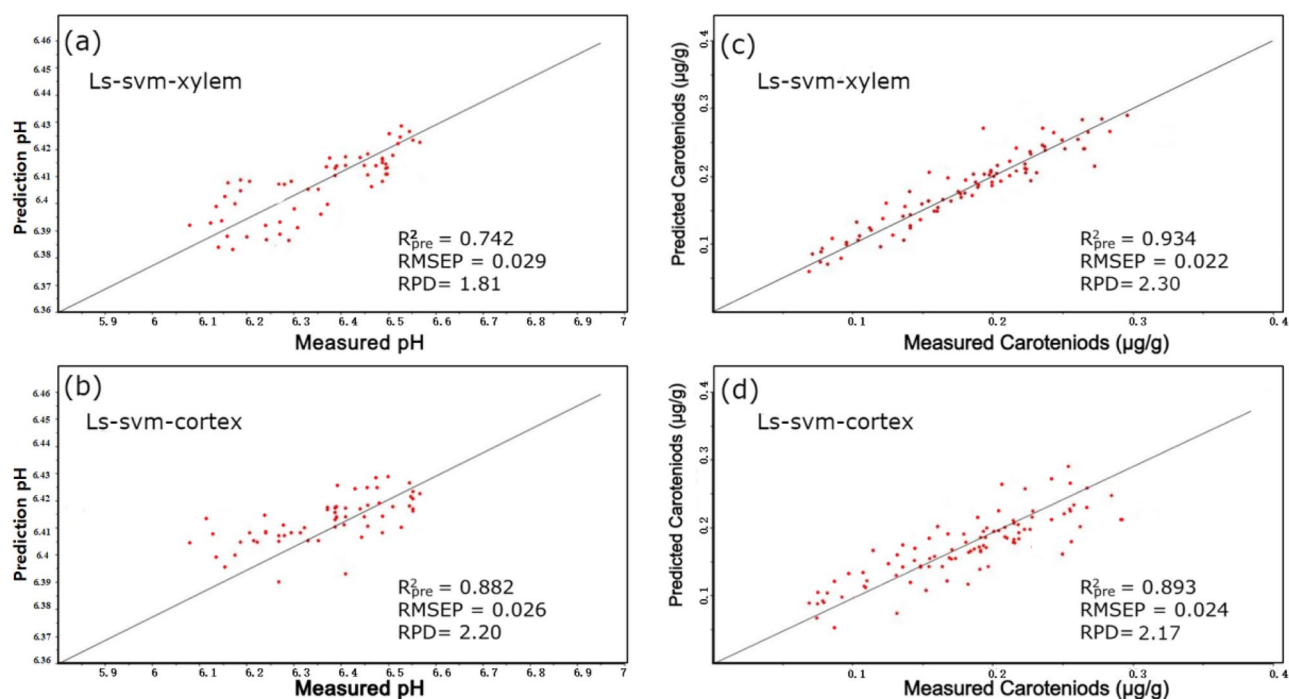


Figure 5. The performances of the top prediction models to detect the quality of Fuzhou (Fz) carrots using EWs are as follows: the pH and carotenoid (Car) prediction is achieved by employing the combination of successive projections algorithm, and least squares support vector machine (SPA-LS-SVM) model.

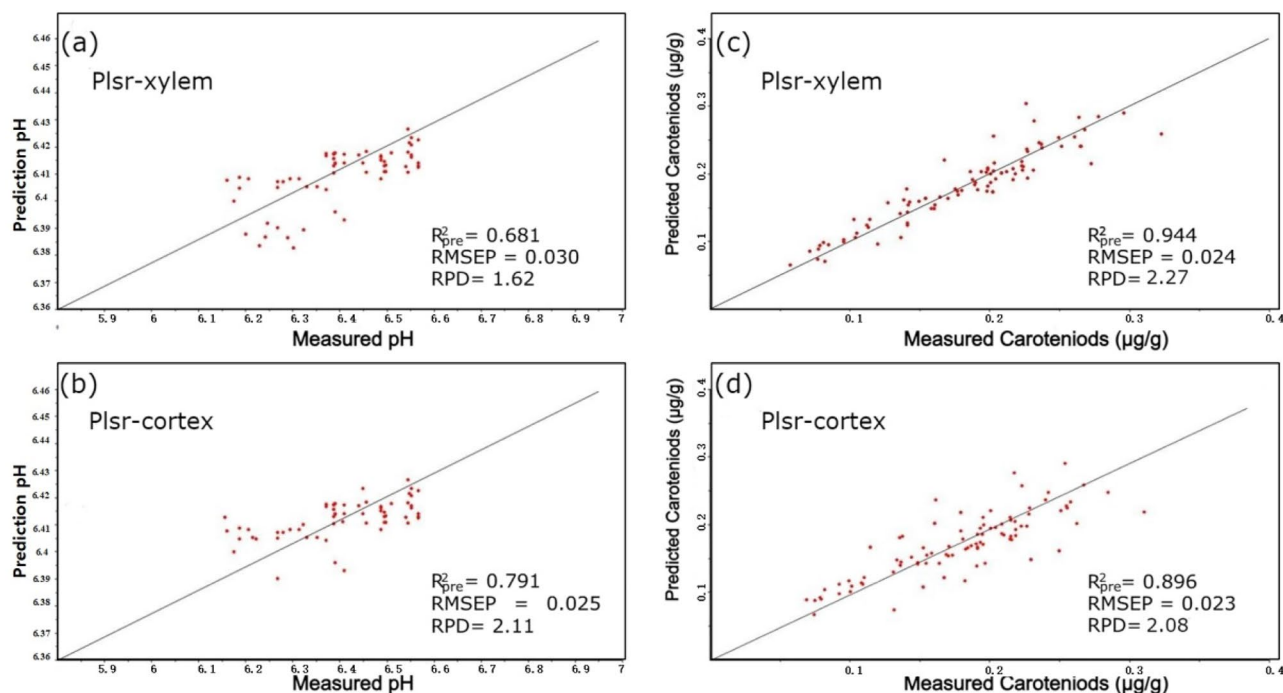


Figure 6. The performances of the top prediction models on Fuzhou (Fz) for detecting quality attributes based on EWs are as follows: the pH and carotenoid (Car) prediction is achieved by employing the combination of successive projections algorithm and partial least squares regression (SPA-PLSR) model.

exhibited variations between the xylem and cortex regions of the cultivars, suggesting that the estimation of Cars requires distinct spectral information needs.

In addition, the analysis unveiled particular wavelength requirements for pH estimation, suggesting the presence of more concentrated and accurate spectral data that is essential for precise pH forecasting. The application of SPA in ascertaining the chosen EWs enhanced the accuracy of the prediction models by revealing the crucial wavelengths associated with each quality attribute and emphasizing the combined benefits of RC and SPA methods. Consistent with previous investigations into internal fruit properties^{63,64}, the results of this study provide additional evidence for the importance of vibrational energy changes induced by NIR light on chemical bonds, including C–H, N–H, O–H, and C–O, which influence Near-Infrared Reflectance Spectroscopy^{65,66}.

The inconsistency between the NIR method and the number of wavelengths chosen for pH estimation may be attributable to internal pH components that are not perfectly aligned; this highlights the significance of selecting wavelengths tailored to particular attributes⁶⁷.

In general, the predictive capabilities of the models for Car and pH attributes have been improved through the careful selection and analysis of EWs utilizing RC and SPA methods. This highlights the significance of custom wavelength selection in multispectral imaging applications that require accurate and efficient quality attribute predictions in the agricultural and scientific sectors.

Modeling based on full wavelengths

Tables 1 and 2 show PLSR model carotenoid and pH predictions. Our findings showed that xylem side spectra models outperformed cortex region models. The higher concentration of carotenoids in the xylem area, which transports water and nutrients, may explain this performance differential. Therefore, HSI of xylem side spectra may improve carrot carotenoid predictions.

In contrast, pH prediction performed poorly. Fz-Pt xylem regions had prediction quality of 0.666 to 0.674, while cortical regions had 0.674 to 0.702. Chemical composition and physiological mechanisms may explain the pH prediction performance differential between the xylem and cortex. Plant xylem, which transports water, may have a more stable pH than the cortex, which stores and performs other functions^{68,69}. This pH stability variance may explain the decreased xylem forecast accuracy. Even if all RDP values are below 2, the pH prediction models need more development and optimization.

The PLSR models made essentially identical predictions, proving their reliability and consistency.

Modeling based on effective wavelengths

This study's reference indices for model evaluation were the root mean square error, R^2_{pre} , and RPD, as shown in Table 3. The regression equations for RC-PLSR-xylem and SPA-LS-SVM-Xylem exhibited high R^2_{pre} and RPD values and low RMSE values. This suggests that these two differential orders provide superior predictive performance compared to other examples. In summary, the performance of the SPA-PLSR model was found to be slightly worse compared with the other models. However, the SPA-LS-SVM models exhibited exceptional

performance. In general, the early warning signals (EWs) identified by the signal processing algorithm (SPA) exhibited more efficacy compared to those discovered by the rule-based classifier (RC).

We noticed that the SPA-LS-SVM method for pH prediction was more accurate than RC, despite RC having more variables. The findings verified that the chosen EWs and spectral morphological parameters were appropriate for spectral dimension reduction and feature extraction. However, they employed different routines for spectral analysis. The former emphasized the morphological differences and spectra variations in samples with diverse Car and pH, while the latter mainly reflected the reflectance or absorbance values. Based on the Car and pH prediction samples, the absorbance spectra calibrated by SG-MSC based on the EWs proved the best in constructing PLSR and LS-SVM predictive models for Car and pH in carrot samples.

On the other hand, Car estimates outperformed pH predictions due to the low organic acid concentration in fruits. Given that the reference pH values did not span a broad spectrum, it is reasonable to infer that they were suitable for establishing a dependable and precise calibration model. The results of the top prediction models for identifying quality attributes using the various variable selection techniques are shown in Figs. 5 and 6.

Prediction of pH

The prediction of pH using the full wavelength range was not satisfactory because it did not yield accurate results. However, EWs exhibited an improved prediction⁷⁰. Despite this improvement, the RPD of the pH prediction could not match or exceed the Car prediction. This discrepancy can be attributed to the relatively low concentration of organic acids in the slip fruits⁷¹. Here, Car prediction relied on a more distinct spectral signature, which made it easier to detect and quantify. The pH values in the cortex tissue were higher than those in the xylem tissue.

Regarding the levels of Car and pH in the different tissues, it is likely that there could be differences in the distribution of these compounds. It has been pointed out that Car are synthesized and stored in plastids in different concentrations in different tissue types⁷². It is also possible that the metabolism of Car and pH in the different tissues is regulated differently, leading to different Car and pH levels^{38,39}. Therefore, it is possible to have higher levels of Car in the xylem and lower levels in the cortex, and viceversa for pH levels. However, more research is needed to confirm this phenomenon.

We noticed that the LS-SVM model effectively evaluated the internal qualities of carrots, specifically the pH level, as shown in Table 4. On the other hand, the PLSR presents a substantial discrepancy between the correction set and the prediction generated. Furthermore, the degree of dispersion in the predictions for the data points is considerable, suggesting that the PLSR model exhibits inadequate accuracy in fitting predictions and stability concerning pH quality, as can be seen in Figs. 5a,b and 6a,b. This model can be applied to detect and assess the pH of carrots, providing valuable theoretical support and serving as a foundation for developing online carrot detection equipment. Future research can focus on exploring additional spectral regions and refining the models to improve pH prediction accuracy and further enhance the overall performance of the models.

Moreover, the implementation of the LS-SVM model for assessing carrot quality holds promise for enhancing the agricultural industry. Accurately assessing the pH level of carrots enables farmers and producers to make informed decisions regarding harvesting, storage, and distribution. This process ultimately enhances the market value of the crop and overall quality. Furthermore, successfully applying the LS-SVM model in carrot quality evaluation highlights its potential for use in other fruits and vegetables. This result aligns with the Shao et al.⁷³ findings. These authors demonstrated that LS-SVM models exhibit strong predictive capabilities for internal fruit attributes^{74,75}. This discovery opens doors for further research and development in agricultural technology, potentially enhancing quality control processes and overall productivity within the industry.

Prediction of carotenoid

Research has extensively studied Car in carrot root systems, including tomatoes and peppers^{76,77}. For example, Perrin et al. found various Car in carrots with varying root colors. A related work revealed that the orange plant phloem had more Car than its xylem and was in the carrot roots. Another study also documented that the red genotype phloem and xylem Car were similar^{78,79}. The overall results of the Car and pH prediction of the two sides showed significant differences using different regression methods. Further analysis of the characteristic wavelengths for Car and pH showed significant similarities between the two regions. Here, Car was mainly distributed on the xylem side, with higher concentrations in the pericarp region of the carrot slice. This difference may be due to differences in Car synthesis and transport between the two tissues. Perrin et al.⁷⁸ found that the varying expression patterns of the genes involved in Car production in various tissues could account for the variations in the accumulation in various root tissues. This aligns with our finding, implying that the cortex side had much lower Car concentrations compared with the xylem tissue. The LS-SVM model predicted significant set of R^2_{pre} values, smaller RMSE values, greater prediction accuracy, and an RPD of 2.30 for the Car of Fz compared to PLSR prediction models, as shown in Table 2 and Figs. 5c,d, and 6c,d. With the least dispersion, the measured and predicted values of the Car of Fz in the LS-SVM model's calibration and prediction sets are situated on opposite sides of the 45° line. This signifies that the model possesses optimal fitting accuracy and stability.

Conclusion

This study investigated HSI to determine carrots' internal quality, Car, and pH, focusing on the effects of sampling regions of two cultivars. The results indicated that the cortex or xylem region can accurately predict Car and pH, with a significant difference in prediction performance between the two regions. The characteristic wavelengths for Car and pH prediction using different sampling regions were not identical. This suggests that assessing the cortex and xylem regions could be used to predict these attributes more precisely and cost-effectively. This finding is significant as it provides flexibility in the sampling process and allows a more straight forward implementation

of HSI in determining internal quality attributes in carrots. This technique can also help farmers better understand their crop's condition and monitor different attributes depending on their supply chains. Overall, our study provides a reference for implementing multispectral technologies for the internal quality assessment of carrots. Adopting HSI can provide accurate and non-destructive testing of internal quality attributes in carrots, benefiting the food industry and, ultimately, the consumers. Further research is needed to fully understand the potential of these models for predicting pH levels in different fruit tissues and under different environmental conditions. Studies are also required to explore different spectral regions to improve the prediction of internal quality attributes in carrots.

Data availability

The data used in this study are available upon request. Please contact Shu he zheng at zsh@fafu.edu.cn for access to the data.

Received: 12 December 2023; Accepted: 8 April 2024

Published online: 12 April 2024

References

1. Arscott, S. A. & Tanumihardjo, S. A. Carrots of many colors provide basic nutrition and bioavailable phytochemicals acting as a functional food. *Compr. Rev. Food Sci. Food Saf.* **9**, 223–239. <https://doi.org/10.1111/j.1541-4337.2009.00103.x> (2010).
2. Nicolle, C., Simon, G., Rock, E., Amouroux, P. & Révész, C. genetic variability influences carotenoid, vitamin, phenolic, and mineral content in white, yellow, purple, orange, and dark-orange carrot cultivars. *J. Am. Soc. Hortic. Sci.* **129**, 523–529. <https://doi.org/10.21273/JASHS.129.4.0523> (2004).
3. Zhao, D. *et al.* First report of black rot of carrot caused by *Alternaria carotiincultae* in China. *Plant Dis.* **108**, 223 (2023).
4. Becaro, A. *et al.* Postharvest quality of fresh-cut carrots packaged in plastic films containing silver nanoparticles. *Food Bioprocess Technol.* <https://doi.org/10.1007/s11947-015-1656-z> (2016).
5. Bashir, R. *et al.* Foliar application of γ -aminobutyric acid (GABA) improves vegetative growth, and the physiological and antioxidative potential of *Daucus carota* L. Under water deficit conditions (2019)..
6. Deding, U., Baatrup, G., Christensen, L. P. & Kobaek-Larsen, M. Carrot intake and risk of colorectal cancer: A prospective cohort study of 57,053 Danes. *Nutrients* **12**, 332 (2020).
7. Donaldson, M. S. Nutrition and cancer: A review of the evidence for an anti-cancer diet. *Nutr. J.* **3**, 1–21 (2004).
8. Xu, X. *et al.* Dietary carrot consumption and the risk of prostate cancer. *Eur. J. Nutr.* **53**, 1615–1623 (2014).
9. Harris, L. J. *et al.* Outbreaks associated with fresh produce: Incidence, growth, and survival of pathogens in fresh and fresh-cut produce. *Compr. Rev. Food Sci. Food Saf.* **2**, 78–141 (2003).
10. Djoufack, M. M. T. *et al.* Sensory quality and nutritional composition of carrot (*Daucus carota* L.) genotypes as affected by fertilization in production system in Cameroon. *CABI Agric. Biosci.* **4**, 22 (2023).
11. Turner, N. C., Molyneux, N., Yang, S., Xiong, Y. & Siddique, K. H. Climate change in south-west Australia and north-west China: Challenges and opportunities for crop production. *Crop Pasture Sci.* **62**, 445–456 (2011).
12. White, J. W., Hoogenboom, G., Kimball, B. A. & Wall, G. W. Methodologies for simulating impacts of climate change on crop production. *Field Crops Res.* **124**, 357–368 (2011).
13. Gao, Z. *et al.* Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artif. Intell. Agric.* **4**, 31–38. <https://doi.org/10.1016/j.iaia.2020.04.003> (2020).
14. Li, B. *et al.* Detection of waxed chestnuts using visible and near-infrared hyper-spectral imaging. *Food Sci. Technol. Res.* **22**, 267–277. <https://doi.org/10.3136/fstr.22.267> (2016).
15. Li, J., Tian, X., Huang, W., Zhang, B. & Fan, S. Application of long-wave near infrared hyperspectral imaging for measurement of soluble solid content (SSC) in pear. *Food Anal. Methods* **9**, 3087–3098. <https://doi.org/10.1007/s12161-016-0498-2> (2016).
16. Taghinezhad, E., Szumny, A. & Figiel, A. The application of hyperspectral imaging technologies for the prediction and measurement of the moisture content of various agricultural crops during the drying process. *Molecules* **28**, 2930 (2023).
17. Barreto, A., Cruz-Tirado, J. P., Siche, R. & Quevedo, R. Determination of starch content in adulterated fresh cheese using hyperspectral imaging. *Food Biosci.* **21**, 14–19 (2018).
18. Zhang, Z., Yin, X. & Ma, C. Development of simplified models for nondestructive testing of rice with husk starch content using hyperspectral imaging technology. *Anal. Methods* <https://doi.org/10.1039/C9AY01926J> (2019).
19. Apan, A., Held, A., Phinn, S. & Markley, J. Detecting sugarcane 'orange rust'disease using EO-1 Hyperion hyperspectral imagery. *Int. J. Remote Sens.* **25**, 489–498 (2004).
20. Goel, P. K. *et al.* Potential of airborne hyperspectral remote sensing to detect nitrogen deficiency and weed infestation in corn. *Comput. Electron. Agric.* **38**, 99–124 (2003).
21. Lu, B., Dao, P. D., Liu, J., He, Y. & Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* **12**, 2659 (2020).
22. Mulla, D. J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **114**, 358–371 (2013).
23. Mo, C. *et al.* Development of a non-destructive on-line pungency measurement system for red-pepper powder (2013)..
24. Zhu, H. *et al.* Hyperspectral imaging for predicting the internal quality of kiwifruits based on variable selection algorithms and chemometric models. *Sci. Rep.* **7**, 7845. <https://doi.org/10.1038/s41598-017-08509-6> (2017).
25. Munera, S. *et al.* Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. *J. Food Eng.* **214**, 29–39 (2017).
26. Munera, S. *et al.* Astringency assessment of persimmon by hyperspectral imaging. *Postharvest Biol. Technol.* **125**, 35–41. <https://doi.org/10.1016/j.postharvbio.2016.11.006> (2017).
27. Kamruzzaman, M., Makino, Y. & Oshita, S. Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging. *Food Chem.* **196**, 1084–1091 (2016).
28. Kamruzzaman, M., Makino, Y., Oshita, S. & Liu, S. Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef. *Food Bioprocess Technol.* **8**, 1054–1062. <https://doi.org/10.1007/s11947-015-1470-7> (2015).
29. Yuan, R. R. *et al.* Quantitative damage identification of lingwu long jujube based on visible Near-Infrared hyperspectral imaging. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectrosc. Spectr. Anal.* **41**, 1182–1187. [https://doi.org/10.3964/j.issn.1000-0593\(2021\)04-1182-06](https://doi.org/10.3964/j.issn.1000-0593(2021)04-1182-06) (2021).
30. Luypaert, J., Heuerding, S., Heyden, Y. & Massart, D. The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams. *J. Pharm. Biomed. Anal.* **36**, 495–503. <https://doi.org/10.1016/j.jpba.2004.06.023> (2004).

31. Jongeneel, M. & Saccon, A. *Geometric Savitzky-Golay filtering of noisy rotations on SO(3) with simultaneous angular velocity and acceleration estimation.* (2022).
32. Cerro, G., Angrisani, L., Capriglione, D., Ferrigno, L. & Miele, G. *On employing a Savitzky-Golay filtering stage to improve performance of spectrum sensing in CR applications concerning VDSA approach* (Meas. Syst. Metrol, 2016). <https://doi.org/10.1515/mms-2016-0019>.
33. Kagawade, V. C. & Angadi, S. A. Savitzky-Golay filter energy features-based approach to face recognition using symbolic modeling. *Pattern Anal. Appl.* **24**, 1451–1473. <https://doi.org/10.1007/s10044-021-00991-z> (2021).
34. Angrisani, L., Capriglione, D., Cerro, G., Ferrigno, L. & Miele, G. *The Effect of Savitzky-Golay Smoothing Filter on the Performance of a Vehicular Dynamic Spectrum Access Method* (2014).
35. Maleki, M. *et al.* Estimation of gross primary productivity (GPP) phenology of a short-rotation plantation using remotely sensed indices derived from sentinel-2 images. *Remote Sens.* **12**, 2104 (2020).
36. Liu, H., Guo, W. & Yue, R. Non-destructive detection of kiwifruit firmness based on near-infrared diffused spectroscopy. *Nongye Jixie Xuebao Trans. Chin. Soc. Agric. Mach.* **42**, 145–149 (2011).
37. Sun, T., Lin, H., Xu, H. & Ying, Y. Effect of fruit moving speed on predicting soluble solids content of ‘Cuiguan’ pears (*Pomaceae pyriformis* Nakai cv. Cuiguan) using PLS and LS-SVM regression. *Postharvest Biol. Technol.* **51**, 86–90 (2009).
38. Liu, D., Sun, D. & Zeng, X. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food Bioprocess Technol.* **7**, 307–323 (2014).
39. Liu, L., Zhiyong, S., Zhang, M. & Wang, Q. Regulation of carotenoid metabolism in tomato. *Mol. Plant* <https://doi.org/10.1093/mp/ssu121> (2014).
40. Iqbal, J. & Zhang, M. Key wavelengths selection from near infrared spectra using Monte Carlo sampling–recursive partial least squares. *Chemom. Intell. Lab. Syst.* **128**, 17–24. <https://doi.org/10.1016/j.chemolab.2013.07.009> (2013).
41. Yu, B., Yan, C., Yuan, J., Ding, N. & Chen, Z. Prediction of soil properties based on characteristic wavelengths with optimal spectral resolution by using Vis-NIR spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **293**, 122452. <https://doi.org/10.1016/j.saa.2023.122452> (2023).
42. Yamashita, G., Anzanello, M., Soares, F., Rocha, M. & Fogliatto, F. Selecting relevant wavelength intervals for PLS calibration based on absorbance interquartile ranges. *Chemom. Intell. Lab. Syst.* **231**, 104689. <https://doi.org/10.1016/j.chemolab.2022.104689> (2022).
43. Andersson, M. A comparison of nine PLS1 algorithms. *J. Chemom.* **23**, 518–529. <https://doi.org/10.1002/cem.1248> (2009).
44. Dangel, S. R. S., Sanderman, J., Wills, S. & Ramirez-Lopez, L. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. *Soil Syst.* **3**, 11 (2019).
45. Sarkar, S., Basak, J. K., Moon, B. E. & Kim, H. T. A comparative study of PLSR and SVM-R with various preprocessing techniques for the quantitative determination of soluble solids content of hardy kiwi fruit by a portable Vis/NIR spectrometer. *Foods* **9**, 1078 (2020).
46. Pei, H., Wang, K. & Zhong, P. Semi-supervised matrixized least squares support vector machine. *Appl. Soft Comput.* **61**, 72–87. <https://doi.org/10.1016/j.asoc.2017.07.040> (2017).
47. Fu, C., Xiong, H. & Tian, A. Fractional modeling for quantitative inversion of Soil-Available phosphorus content. *Mathematics* **6**, 330 (2018).
48. Britton, G. Structure and properties of carotenoids in relation to function. *FASEB J.* **9**, 1551–1558. <https://doi.org/10.1096/fasebj.9.15.8529834> (1995).
49. Leghari, M., Sheikh, S., Memon, N., Soomro, A. H. & Khooharo, A. Quality attributes of immature fruit of different mango varieties. *J. Basic Appl. Sci.* **9**, 52–56. <https://doi.org/10.6000/1927-5129.2013.09.09> (2013).
50. Wang, X. *et al.* Non-Destructive detection of pH value of kiwifruit based on hyperspectral fluorescence imaging technology. *Agriculture* **12**, 208 (2022).
51. Arnon, D. I. Copper enzymes in isolated chloroplasts polyphenoloxidase in *Beta vulgaris*. *Plant Physiol.* **24**, 1–15. <https://doi.org/10.1104/pp.24.1.1> (1949).
52. Yang, C. *et al.* Comparative effects of salt-stress and alkali-stress on the growth, photosynthesis, solute accumulation, and ion balance of barley plants. *Photosynthetica* **47**, 79–86. <https://doi.org/10.1007/s11099-009-0013-8> (2009).
53. Yang, X. *et al.* Determination of the soluble solids content in korla fragrant pears based on visible and near-infrared spectroscopy combined with model analysis and variable selection. *Front. Plant Sci.* **13**, 938162. <https://doi.org/10.3389/fpls.2022.938162> (2022).
54. ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., and Rousseau, D. Recent applications of multispectral imaging in seed phenotyping and quality monitoring—an overview. *Sensors* **19**, 1090 (2019).
55. Falcioni, R., Antunes, W. C., Dematté, J. A. M. & Nanni, M. R. A novel method for estimating chlorophyll and carotenoid concentrations in leaves: A two hyperspectral sensor approach. *Sensors* **23**, 3843 (2023).
56. Munera, S. *et al.* Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine. *Food Control* **86**, 1–10 (2018).
57. Wei, X., He, J., Ye, D. & Jie, D. Navel orange maturity classification by multispectral indexes based on hyperspectral diffuse transmittance imaging. *J. Food Qual.* <https://doi.org/10.1155/2017/1023498> (2017).
58. ElMasry, G., Wang, N., ElSayed, A. & Ngadi, M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *J. Food Eng.* **81**, 98–107 (2007).
59. Ana, M. C., Dário, P., Rosa, M. P., Maria, D. A. & Rui, G. Nondestructive assessment of citrus fruit quality and ripening by visible–near infrared reflectance spectroscopy. In (eds. Muhammad, S. K. & Iqbal, A. K.) 13 (IntechOpen, 2021).
60. 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).
61. Dong, J., Guo, W., Wang, Z., Liu, D. & Zhao, F. Nondestructive determination of soluble solids content of ‘Fuji’ apples produced in different areas and bagged with different materials during ripening. *Food Anal. Methods* **9**, 1087–1095 (2016).
62. Xuan, G., Gao, C. & Shao, Y. Spectral and image analysis of hyperspectral data for internal and external quality assessment of peach fruit. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **272**, 121016. <https://doi.org/10.1016/j.saa.2022.121016> (2022).
63. Escribano, S., Biasi, W. V., Lerud, R., Slaughter, D. C. & Mitcham, E. J. Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries. *Postharvest Biol. Technol.* **128**, 112–120 (2017).
64. Goke, A., Serra, S. & Musacchi, S. Postharvest dry matter and soluble solids content prediction in d’Anjou and Bartlett pear using near-infrared spectroscopy. *HortScience* **53**, 669–680 (2018).
65. Murray, I. Chemical principles of near-infrared technology. *Near-Infrared Technology in the Agricultural and Food Industries*, 17–34 (1987).
66. Su, W. H. & Sun, D. W. Multispectral imaging for plant food quality analysis and visualization. *Compr. Rev. Food Sci. Food Saf.* **17**, 220–239 (2018).
67. Liu, F. & He, Y. Application of successive projections algorithm for variable selection to determine organic acids of plum vinegar. *Food Chem.* **115**, 1430–1436 (2009).
68. Gomes, M. *et al.* Does abscisic acid and xylem sap pH regulate stomatal responses in papaya plants submitted to partial root-zone drying?. *Theor. Exp. Plant Physiol.* **1**, 1. <https://doi.org/10.1007/s40626-023-00275-3> (2023).
69. Hahn, K. *Not Hydraulic But an Adsorption Water Transport Occurs in the Xylem of Land Plants.* (2023).

70. Andersen, P. V., Afseth, N. K., Gjerlaug-Enger, E. & Wold, J. P. Prediction of water holding capacity and pH in porcine longissimus lumborum using Raman spectroscopy. *Meat Sci.* **172**, 108357. <https://doi.org/10.1016/j.meatsci.2020.108357> (2021).
71. Jiang, B. *et al.* Exogenous salicylic acid regulates organic acids metabolism in postharvest blueberry fruit. *Front. Plant Sci.* **13**, 1024909. <https://doi.org/10.3389/fpls.2022.1024909> (2022).
72. Howitt, C. A. & Pogson, B. J. Carotenoid accumulation and function in seeds and non-green tissues. *Plant Cell Environ.* **29**, 435–445. <https://doi.org/10.1111/j.1365-3040.2005.01492.x> (2006).
73. Shao, Y., Bao, Y. & He, Y. Visible/near-infrared spectra for linear and nonlinear calibrations: A case to predict soluble solids contents and pH value in peach. *Food Bioprocess Technol.* **4**, 1376–1383 (2011).
74. Liu, Y., Wang, Q., Gao, X. & Xie, A. Total phenolic content prediction in Flos Lonicerae using hyperspectral imaging combined with wavelengths selection methods. *J. Food Process Eng.* **42**, e13224 (2019).
75. Zhang, H., Zhan, B., Pan, F. & Luo, W. Determination of soluble solids content in oranges using visible and near infrared full transmittance hyperspectral imaging with comparative analysis of models. *Postharvest Biol. Technol.* **163**, 111148 (2020).
76. Gómez-García, M. D. R. & Ochoa-Alejo, N. Biochemistry and molecular biology of carotenoid biosynthesis in chili peppers (*Capsicum* spp). *Int. J. Mol. Sci.* **14**, 19025–19053 (2013).
77. Nisar, N., Li, L., Lu, S., Khin, N. C. & Pogson, B. J. Carotenoid metabolism in plants. *Mol. Plant* **8**, 68–82. <https://doi.org/10.1016/j.molp.2014.12.007> (2015).
78. Perrin, F. *et al.* Carotenoid gene expression explains the difference of carotenoid accumulation in carrot root tissues. *Planta* <https://doi.org/10.1007/s00425-016-2637-9> (2017).
79. Perrin, F. *et al.* Carotenoid gene expression explains the difference of carotenoid accumulation in carrot root tissues. *Planta* **245**, 737–747. <https://doi.org/10.1007/s00425-016-2637-9> (2017).

Author contributions

Mulowayi Mutombo Arcel and Zhen Hui Shen conceived of the presented idea. Mulowayi Mutombo Arcel, Zhen Hui Shen and Shu He Zheng developed the theory and performed the computations. Witness Joseph Nyimbo and Di Zhi Feng verified the analytical methods. Shu He Zheng supervised the findings of this work. Zhen Hui Shen and Nyumah Fallah prepared Table and figures. All authors discussed the results and contributed to the final manuscript.

Funding

The authors gratefully appreciate the financial support provided by the Natural Science Foundation of Fujian Province (Grant 427 number 2023J011108), the Pilot Project of the Fujian Provincial Department of Science and Technology (Grant number 2022N0009), 428, and the Science and Technology Project of Fujian Agriculture and Forestry University (Grant number 000-71202103B).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.H.Z.

Reprints and permissions information is available at 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) 2024