Food Sci Biotechnol (2021) 30(6):783–791
https://doi.org/10.1007/s10068-021-00921-z

RESEARCH ARTICLE

Prediction of moisture content in steamed and dried purple sweet potato using hyperspectral imaging analysis

Suhyeon Heo[1] · Ji-Young Choi[1] · Jiyoon Kim[1] · Kwang-Deog Moon[1,2]

Received: 15 June 2020 / Revised: 27 March 2021 / Accepted: 18 May 2021 / Published online: 21 June 2021

Abstract Partial least squares regression (PLSR) modeling was performed to predict the moisture content in steamed, dried purple sweet potato based on spectral data obtained from hyperspectral imaging analysis. The PLSR model with a combination of multiplicative scatter correction, Savitzky–Golay, and first derivative exhibited the highest accuracy (R_P^2 = 0.9754). The wavelengths found that strongly affected the PLSR model were 961.12, 1065.50, 1083.93, 1173.23, and 1233.89 nm. These wavelengths were associated with the O–H second overtone and the second overtone of C–H, C–H_2, and C–H_3. When PLSR modeling was performed using these selected wavelengths, the prediction accuracy of the PLSR model exhibited high accuracy (R_P^2 = 0.9521). Therefore, the moisture content could be predicted with high accuracy using only five wavelengths rather than the full spectrum.

Keywords Hyperspectral imaging analysis · Partial least squares regression modeling · Selected wavelengths · Moisture content · Purple sweet potato

Abbreviations

AC	Anthocyanin contents
MC	Moisture content
MSC	Multiplicative scatter correction
PLS	Partial least squares
PLSR	Partial least squares regression
RMSE	Root mean square error
RMSECV	RMSE cross-validation
ROI	Region of interest
SNV	Standard normal variate
TPC	Total phenolic compounds
TSS	Total soluble solids

Introduction

Sweet potato (*Ipomoea batatas* L.) is a perennial herb, and it is known to have originated between the Yucatan Peninsula region of Central America and the Orinoco River estuary in South America (Kim et al., 2017). Danzami (*Ipomoea batatas* (L.) Lam.) was bred by crossing purple sweet potato (Yeonzami) and pumpkin sweet potato (Yeonhwangmi) in 2015 to improve the sweetness and taste. Although the total anthocyanin content of Danzami is lower than that of the previous purple sweet potato (Shinzami), their total polyphenol content and antioxidant activities are the same. Danzami possesses high sweetness and excellent flavor, making it a favorable candidate for steamed, dried sweet potatoes.

Sweet potato is consumed primarily in food, brewing, and starch production. Sweet potatoes contain a lot of

✉ Kwang-Deog Moon
kdmoon@knu.ac.kr

Suhyeon Heo
sewt1024@naver.com

Ji-Young Choi
chjiyeng91@naver.com

Jiyoon Kim
fuocoso95@naver.com

[1] School of Food Science and Biotechnology, Kyungpook National University, 80 Daehak-ro, Daegu 41566, South Korea

[2] Food and Bio-Industry Research Institute, Kyungpook National University, 80 Daehak-ro, Daegu 41566, South Korea

moisture, are weak against cold, and generate a lot of carbon dioxide gas. Therefore, there are difficulties involved in storage and transportation; the main issue is that they must be consumed within a short period following production. Sweet potatoes are processed in fried or dried form to increase industrial availability. Among the processed products, steamed, dried sweet potatoes are a dry product; and their quality is significantly affected by their MC. The excessive decrease in MC induced by drying causes the texture to becomes harder and less palatable (Shin and Lee, 2011). Furthermore, due to this rapid loss of water, the quality deteriorates, and it is subject to shrinking (Hong and Lee, 2004). Conversely, if the MC is too high, it becomes a product that easily breaks during storage and is difficult to store. Therefore, determining the optimal MC for steamed, dried sweet potatoes is an important step in the drying process. Moreover, measuring the MC during drying involves an atmospheric-pressure drying method, which is both destructive and time-consuming. In addition, the potential for error with this approach is vast as only a few representative samples are measured, as this method does not involve checking the MC for all of the samples. Therefore, overcoming this problem requires a new method to measure MC nondestructively; one such method is hyperspectral imaging analysis.

Hyperspectral image analysis is based on near-infrared (NIR) spectroscopy combined with imaging technology (Costa et al., 2011; Sun, 2004). Hyperspectral imaging analysis is similar in principle to NIR spectroscopy (Ye et al., 2014), where the spectrum generated by the absorption of infrared energy in response to the oscillation of covalent bonds (O–H, C–H, N–H, and S–H) is analyzed. Hyperspectral image analysis differs from NIR spectroscopy in that spectral information is obtained from an entire or partial image of the sample used to collect three-dimensional spectral data, and it can extract the spectral data corresponding to minute spaces in the sample separately. Therefore, the hyperspectral image analysis is suitable for monitoring a large amount of material as a scan. Numerous examples of hyperspectral imaging analysis have been reported, including the determination of banana quality and maturation stage (Rajkumar et al., 2012), organic residues (Feng and Sun, 2012), and the physical, chemical, and biological contamination of foods. Furthermore, it is also utilized prediction of corn hardness (Williams et al., 2009), beef color, pH, and shear force (Masry et al., 2012). Hyperspectral image analysis has previously been used to predict MC and exhibits high prediction accuracy (Huanga et al., 2014; Pu and Sun, 2016; Wu et al., 2012).

To our knowledge, hyperspectral image analysis in the NIR range (900–1700 nm) has not yet been used to predict the MC of steamed dried sweet potatoes. This study applies hyperspectral image analysis to the quality analysis of processed, steamed, dried sweet potatoes to quickly monitor the quality changes during processing rather than the raw materials, in order to filter out defective products. We also developed a PLSR model that predicts the most important MC to ensure quality in steamed, dried sweet potatoes.

Materials and methods

Sample preparation

The sweet potatoes used in this experiment were purple sweet potatoes (Danzami) produced in 2018 in Haenam-gun, Jeollanam-do, Korea. The sample weights were 157 ± 40 g. The sweet potatoes were washed by only water until they were clean, cooked for 25 min at 100 °C, cooled, and cut into a cylindrical shape (diameter 2.7 cm, thickness 1 cm). Each sample was dried at 55 °C for 0, 2, 4, 6, 8, or 10 h using a dryer (BL950903, Gumbok Stoke Co., Ltd., Seoul, Korea). Seventy samples per group were prepared, a total of 420 samples and the degree of drying was determined through PLS discriminant analysis (PLS-DA). Two hundred and seventy of these samples were used to develop a MC prediction model with PLSR.

Hyperspectral image acquisition and extraction

Hyperspectral images of the steamed, dried purple sweet potatoes were obtained in the Vis–NIR region 900–1700 nm using an ImSpector N17E (Specim, Spectral Imaging Ltd., Oulu, Finland). The N17E has a spectral range of 900–1700 nm, a variance of 110 nm/mm, and a spectral resolution of 5 nm, resulting in the spectral data being divided into 256 bands in the same spectral range. The light source consisted of two halogen lamps (1400 nm long-pass filter) fixed to the frame with a camera to illuminate the sample. The fixed camera could only obtain images directly under the lens, thus, requiring that the sample be moved in one direction to obtain a continuous hyperspectral image. The sample-conveying component was configured under the frame. Data were extracted using the ROI function of the ENVI software (version 5.4, Exelis Visual Information Solutions, Boulder, Colorado, USA) to obtain data only for the ROI from the acquired hyperspectral images.

Quality characteristics analysis

The moisture content (MC)

MC of the samples was measured using the atmospheric pressure drying method. Each sample was dried at 103 °C for 42 h using a drying oven (KMC-1202D3, Vision Scientific Co., Ltd., Daejeon, Korea). The MC was calculated using the weight differences found before and after drying. Total soluble solids (TSS) and reducing sugar content.

TSS content and the reducing sugar content were measured via the following method, 45 mL of distilled water was added to 5 g of each sample, followed by grinding with a hand blender (HR1604, Royal Philips Electronics N.V., Amsterdam, Netherlands) for 1 min. Following ultrasonic extraction for 2 h, centrifugation was performed at 10,000×*g* for 10 min. The supernatant of the centrifuged sample was filtered with a 0.45 μm filter, and the filtrate was used as a sample. The TSS was measured using a refractometer (N-1α, ATAGO Co., Ltd., Tokyo, Japan), and the results were expressed in ° Brix. Reducing sugar was added to 1 mL of the DNS reagent in the sample dilution and allowed to stand at 60 °C For 10 min, then the absorbance was measured at 540 nm using a UV–visible spectrophotometer (Evolution 201, Thermo Fisher Scientific Inc., Madison, WI, USA). Glucose was used as a standard.

Total phenolic compounds (TPC) and anthocyanin contents (AC)

TPC and AC were analyzed using the following method for extraction, 45 mL of 80% methanol was added to 5 g of the sliced sample, and extracted at 30 °C and 120 rpm in a shaking incubator (JSSI-300C, JS Research Inc., Gongju-city, Korea) for 12 h. The extract was then filtered with Whatman No. 2 and used as a sample. TPC was analyzed by applying the Folin-Denis method (Kim et al., 2016). Briefly, Folin-Denis reagent (JUNSEI Chemical Co. Ltd., Tokyo, Japan) was added to the diluted sample extract. This mixture was allowed to stand in the dark for 10 min, then 2 mL of a 10% Na_2CO_3 solution was added. Following 1 h in the dark, the absorbance was measured at 750 nm using a UV–visible spectrophotometer (Evolution 201, Thermo Fisher Scientific Inc., Madison, WI, USA) to calculate the gallic acid (mg / L) for use as a standard. Measurement of the AC involved the following, sample extracts were added to 0.025 M potassium chloride buffer (pH 1.0) and 0.4 M sodium acetate buffer (pH 4.5), respectively, and then allowed to stand for 15 min. Absorbance was measured at 515 nm and 700 nm using a UV–visible spectrophotometer (Evolution 201, Thermo Fisher Scientific Inc., Madison, WI, USA), and the anthocyanin content was calculated via the following equation:

$$\text{Anthocyanincontents(mg/L)} = \frac{A \times MW \times DF \times 1000}{\varepsilon \times 1}$$

$$A = (A_{515nm} - A_{700nm})_{pH1.0} - (A_{515nm} - A_{700nm})_{pH4.5}$$

A515nm = absorbance at 515 nm, A700nm = absorbance at 700 nm, MW = molecular weight of cyanidin-3-glucoside or 449.2, DF = dilution factor, ε = molar extinction coefficient of 26.900.

*Color value (CIE L*a*b*) and browning index (BI)*

Colors of the samples were represented by the L* (lightness), a* (redness), b* (yellowness) values, and a colorimeter (CR-400, Minolta Co., Osaka, Japan) calibrated with a standard white plate (L* = 97.79, a* = − 0.38, and b* = 2.05). The average value was measured 15 times for each group. Using the measured (L*, a*, b*) values, the BI was calculated using the following equation. It was calculated as X_n = 91.97, Y_n = 93.8, and Z_n = 107.98 used in the following equation (Ruiz et al., 2012).

$$X = X_n\left(\frac{a^*}{500} + \frac{(L^* + 16)}{116}\right)^3$$

$$Y = Y_n\left(\frac{(L^* + 16)}{116}\right)^3$$

$$Z = Z_n\left(\frac{-b^*}{200} + \frac{(L^* + 16)}{116}\right)^3$$

$$X = \frac{X}{(X + Y + Z)}$$

Duncan's multiple range test and Pearson's correlation analysis were performed using SPSS version 14.0 (SPSS, Inc., Chicago, IL, USA) to identify significant differences between the results ($p < 0.05$).

Partial least squares analysis

Partial least squares discriminant analysis (PLS-DA)

A regression model was used for PLS-DA that identifies the groups to be discriminated by setting them as virtual variables. In this study, the PLS-DA model was used to determine the dryness of purple sweet potatoes. PLS-DA model was performed using the Unscrambler (version 10.5, CAMO, Trondheim, Norway) program. As noted earlier, the dryness of the sweet potatoes was classified into six levels, and these six levels were confirmed by PLS-DA. A total of 294 dried sweet potatoes were used for developing the calibration model, and 126 were applied to the developed model to confirm the discrimination accuracy. The β

coefficient of the PLS-DA model could be used for meaningful wavelength selection (Cheng and Sun, 2015). The meaningful wavelength region greatly influenced the determination of the drying degree for the dried sweet potatoes and assisted with confirmation of the quality properties related to this region.

PLSR MC prediction model

The PLSR technique generalizes and combines the functions of principal component analysis and multiple regression analysis and aims to predict or analyze the dependent variables in a series of independent or predictor variables (Abdi, 2003). During spectrum acquisition, the spectral data of the sample were collected together with noise. Noise sources were generally from the scattering of light on the sample surface, the surrounding environment, or changes caused by the condition of the equipment. In order to obtain accurate spectral data, these noises must be removed prior to data analysis. Various pre-processing of the spectral data could improve the scattering effects, facilitate noise reduction, and provide more accurate spectra (Bae et al., 2016). In the present study, various pretreatments were used individually, and in combination including smoothing, normalization, SNV, MSC, Savitzky–Golay first derivative (SG D1), and Savitzky–Golay second derivative (SG D2). Smoothing replaced a certain range of values to be pre-processed with the mean value, in order to represent a spectrum with noise as a smooth curve. SNV and MSC pre-processing removed the influence of light scattering due to their irregular shape and size (Vidal and Amigo, 2012). The SG first-order (SG D1) and SG second-order (SG D2) differentiation methods eliminated the movement from the baseline of the spectrum and emphasized the spectral characteristics of the minute components (Rinnan et al., 2009). Since the spectral data were pre-processed using the most accurate pre-processing methods, the accuracy and efficiency of the PLSR prediction model of the full-wavelength and five selected wavelengths were compared. In this experiment, the calibration model was created using 192 samples based on the Kennard–Stone sampling algorithm, and 78 of these samples were used to estimate the MC using the calibration model. Therefore, a total of 270 sweet potatoes were used in the PLSR MC prediction model, and each of the samples was measured for MC. In order to evaluate the model accuracy, the following factors were considered the RMSE and R^2 were used to evaluate the performance of the PLS model for each pretreatment; RMSEC (calibration model), RMSECV, RMSEP (RMSE prediction), R_C^2 (R^2 calibration model), R_{CV}^2 (R^2 cross-validation), and R_P^2 (R^2 prediction) (Giovenzana et al., 2018).

PLSR prediction using a selected key wavelength

A prediction model calculation using the full wavelength is relatively time-consuming. In addition, the data is so vast that it may be inconvenient to process the data. Therefore, in order to speed up the development of the predictive model and to facilitate the calculation, the MC prediction model was calculated with a spectrum of several key wavelengths rather than the full-wavelength. Thus, by checking the regression coefficients, several wavelength regions with high absolute values were selected. The PLSR prediction model was developed using the MC and spectrum of the selected wavelengths. The accuracy and efficiency of the full-wavelength prediction model and the key selected wavelength prediction model were then compared.

Results and discussion

Quality characteristics

Various quality characteristics of the steamed, dried sweet potato samples according to drying time are shown in Table 1. The MC decreased significantly as the drying time was increased to 8 h ($p < 0.05$); however, at drying times longer than 8 h, the MC showed no significant differences. When sweet potatoes are heated, most of the starch is hydrolyzed to produce maltose (Suh et al., 1998). The sugar produced at this time affects the texture and preference of the sweet potatoes. TSS and reducing sugar increased significantly as the drying time increased. This was because the soluble solid content per unit weight had increased due to decreases in water content. A study by Shin and Lee (2011) showed similar results as they found that the soluble solid content increased as the water content of the agricultural products had decreased. Phenolic compounds are mostly flavonoids, primarily composed of flavones, flavanones, flavanols, anthocyanidins, and cathechin. Therefore, TPC and AC exhibited the same tendency as the drying time increased. Yang et al. (2010) reported the formation of phenolic compounds and increased antioxidant activity when steamed dried, which was similar to the results of this study. L* and a* had gradually decreased with drying time; however, no significant changes were found following 6 h of drying time. However, the b* and BI values increased as the drying time increased. Moreover, during the drying process, the dried sweet potato was darkened by the heat, producing browning substances via the polymerization of the sugar and amino acids. This result likely occurred because the anthocyanin content of the purple sweet potato had increased.

Table 1 Various quality characteristics of the steamed and dried purple sweet potato at different times during drying at 55 °C

Samples	Quality characteristics[1]								
	MC (%)	TSS (°Brix)	Reducing sugar content (glucose g/L)	TPC (gallic acid mg/L)	Monomeric anthocyanin content (mg/L)	L*	a*	b*	BI
Time (h)									
0	66.62 ± 2.17^{e2}	27.00 ± 1.00^{a}	8.12 ± 0.48^{a}	636.77 ± 172.49^{a}	60.78 ± 13.68^{a}	27.55 ± 2.39^{e}	18.61 ± 3.26^{d}	− 8.29 ± 1.21^{a}	14.90 ± 4.54^{a}
2	47.98 ± 4.53^{d}	38.67 ± 2.31^{b}	12.55 ± 1.39^{b}	841.59 ± 118.32^{ab}	78.95 ± 12.06^{ab}	23.75 ± 1.50^{b}	13.70 ± 1.06^{c}	− 5.22 ± 0.87^{b}	15.45 ± 3.01^{a}
4	35.51 ± 1.15^{c}	49.33 ± 4.16^{c}	15.56 ± 0.71^{c}	1076.44 ± 190.34^{bc}	95.07 ± 12.00^{bc}	21.83 ± 1.52^{a}	10.83 ± 1.66^{b}	− 2.34 ± 0.44^{c}	19.05 ± 3.44^{b}
6	25.94 ± 0.45^{b}	59.33 ± 3.79^{d}	19.19 ± 0.75^{d}	1117.22 ± 108.92^{bc}	101.60 ± 9.37^{bc}	20.10 ± 2.14^{a}	8.65 ± 2.03^{a}	− 0.65 ± 0.56^{d}	20.55 ± 4.93^{b}
8	19.99 ± 1.00^{a}	66.00 ± 2.65^{e}	19.41 ± 1.01^{d}	1256.61 ± 91.18^{bc}	96.90 ± 9.69^{bc}	21.40 ± 2.28^{a}	8.49 ± 2.95^{a}	− 0.35 ± 0.56^{de}	20.51 ± 6.14^{b}
10	18.07 ± 1.04^{a}	72.33 ± 0.58^{f}	21.81 ± 1.00^{e}	1188.03 ± 215.65^{c}	104.07 ± 13.93^{c}	20.15 ± 3.47^{a}	7.60 ± 1.82^{a}	0.11 ± 0.49^{e}	20.81 ± 6.15^{b}

[1]MC, moisture content; TSS, total soluble solids; TPC, total phenolic compounds; L*, lightness; a*, redness; b*, yellowness; BI, browning index

[2]The mean ± standard deviation (n = 3) with different letters (a–e) means significantly different by Duncan's multiple range test ($p < 0.05$)

PLS-DA

The accuracy of the calibration, cross-validation, and prediction discrimination of the PLS-DA modeling utilized raw data that was not pre-processed, as shown in Fig. 1(A). The discrimination accuracy of calibration, cross-validation, and the prediction was 82.99%, 80.61%, and 80.95%, respectively, and the overall accuracy was more than 80%. These results showed the potential of PLS-DA for use in the accurate discrimination of the dryness degree of the purple sweet potato. The influence of each wavelength region on the discrimination found in this PLS-DA model is depicted in Fig. 1(B). The 5 key wavelengths with remarkably high regression coefficients were 937, 1034, 1141, 1197, and 1339 nm, and were the highest-ranked in the order of 1339, 937, 1197, 1141, and 1034. The TSS content represented the organic molecules, which contained C–H, O–H, C–O, and C–C bonds (Liu et al., 2010). Glucose has an informative region of 840–1062 nm, where bands of glucose exist due to the first overtones of the C–H stretching modes (Workman and Weyer, 2007). Moreover, fructose exhibited informative regions of 816–1050 nm. These regions corresponded to band regions generated by C–H stretching combinations (Osborne et al., 1993). Sugar absorption bands (900 ~ 1000 nm), generated from OH stretching, and the third and fourth overtones of CH stretching, occurred near the strong water absorption regions (Amodio et al., 2017). While, absorption bands of 1100 ~ 1300 nm were related to the second overtone of CH, CH_2, CH_3 (Zou et al., 2010). Table 1 shows that the quality characteristics of the purple sweet potato are significantly different depending on the drying time, particularly the MC, soluble solids content, and reducing sugar content. Therefore, it seems that the sugar content of the purple sweet potatoes primarily influenced the discrimination of the samples in the PLS-DA model.

Table 2 shows the Pearson's correlation between the quality change values shown in Table 1. The MC of dried purple sweet potatoes are closely related to the TSS and the reducing sugar changes; thus, the correlation coefficients were − 0.983 and 0.989, respectively, and the absolute value was very high. TSS and reducing sugar showed the highest correlation coefficient with 0.992. Although the peaks related to moisture were not noticeable in the results of the regression coefficients, it was confirmed that the MC and the degree of drying could be predicted from the sugar content of dried sweet potatoes when looking at the Pearson correlation coefficient.

PLSR prediction

Moisture content prediction accuracy by pretreatment

The predicted accuracy of the water content by various pretreatments is shown in Table 3. The prediction accuracy of the raw data without pre-processing was RMSEP = 3.4026 and R_P^2 = 0.9616. Among the various individual pretreatment methods, the moisture prediction accuracy of the PLSR model had increased, with the exception of the smoothing and SG D2 method. Among the single pretreatment methods, the SNV treatment demonstrated the highest moisture prediction accuracy (RMSEC = 2.6048, RMSECV = 2.7374, RMSEP = 2.8787, R_C^2 = 0.9776, R_{CV}^2 = 0.9753, R_P^2 = 0.9729).

Combinations of the MSC, normalization, SNV, and SG D1 methods were used to improve accuracy. When a pretreatment combination is used, the accuracy of each of the two sequences differs. Therefore, the pretreatment combinations shown in Table 2 are listed in order of higher discrimination accuracy. The two combination pretreatment methods exhibited a higher discrimination accuracy than the raw data. Specifically, the SG D1 + MSC combination exhibited the highest accuracy (RMSEC = 2.2816, RMSECV = 2.4965, RMSEP = 2.6977, R_C^2 = 0.9828, R_{CV}^2 = 0.9794, R_P^2 = 0.9754). Figure 2 shows the results of the PLSR moisture prediction model with SG D1 and MSC pretreatment at full-wavelengths. In the calibration model and the cross-validation model, R^2 was very high at 0.9828 and 0.9794, respectively, and the predicted model also showed high accuracy at R_p^2 of 0.9754.

Key wavelengths selection and prediction model development

The influence on the PLSR model for each wavelength was obtained using the combination of the SG D1 + MSC methods, as depicted in Fig. 3. The greater the absolute value of the influence at zero, the greater the influence of the wavelength on the PLSR model. The most influential wavelengths were 561.12, 1065.50, 1083.93, 1173.23, and 1233.89 nm. Wavelengths from 960 to 990 nm were known to be related to the second overtone of O–H bonds in water molecules (Sivakumar et al., 2006; Williams and Norris, 2001). Wavelengths from 1100 to 1300 nm were associated with the second overtones of C–H, $C–H_2$, and

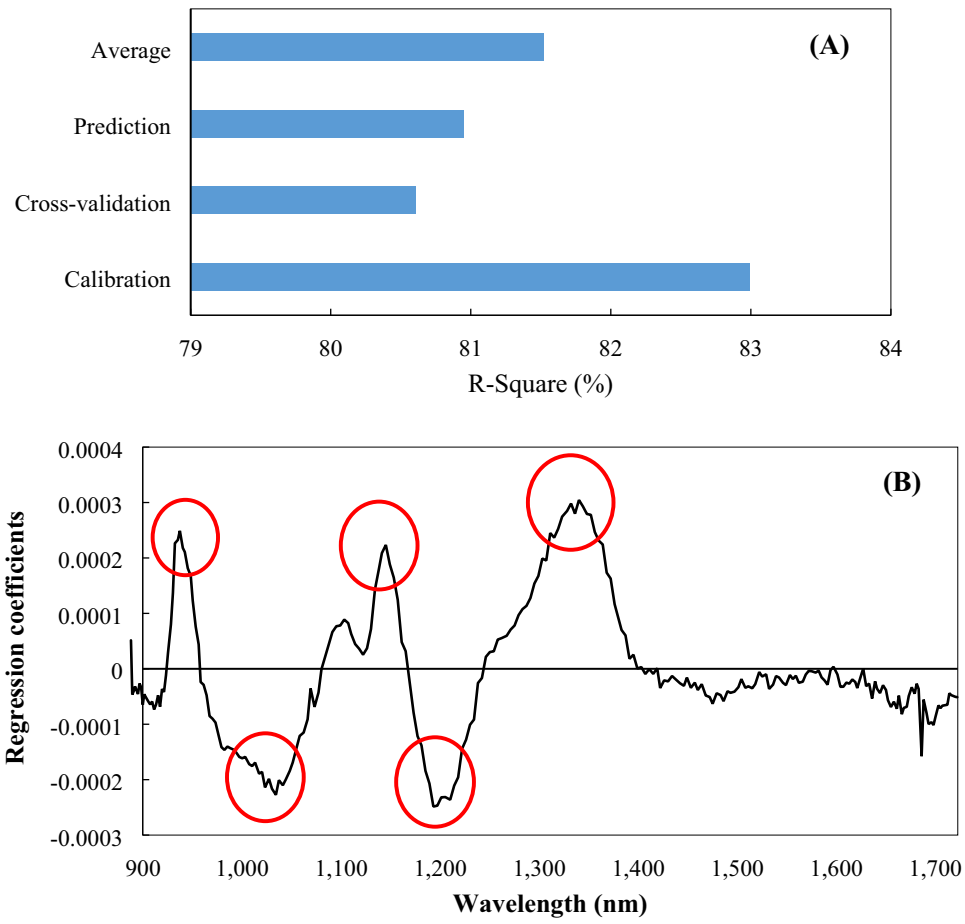


Fig. 1 Results for PLS-DA of the steamed and dried purple sweet potato using full-wavelengths. (**A**) Accuracy of classification; (**B**) Regression coefficients

Table 2 Pearson's correlation between the physicochemical properties

Quality properties	MC[1]	TSS	Reducing sugar content	TPC	Monomeric anthocyanin content	L*	a*	b*	BI
MC	1	− 0.983**[2]	− 0.989**	− 0.984**	− 0.970**	0.959**	0.994**	− 0.994**	− 0.952**
TSS	–	1	0.992**	0.956**	0.934**	− 0.914*	− 0.965**	0.972**	0.951**
Reducing sugar content	–	–	1	0.954**	0.966**	− 0.956**	− 0.985**	0.986**	0.956**
TPC	–	–	–	1	0.953**	− 0.929**	− 0.975**	0.984**	0.956**
Monomeric Anthocyanin content	–	–	–	–	1	− 0.993**	− 0.989**	0.986**	0.949**
L*	–	–	–	–	–	1	0.984**	− 0.973**	− 0.918**
a*	–	–	–	–	–	–	1	− 0.996**	− 0.949**
b*	–	–	–	–	–	–	–	1	0.972**
BI	–	–	–	–	–	–	–	–	1

[1]MC, moisture content; TSS, total soluble solids; TPC, total phenolic compounds; L*, lightness; a*, redness; b*, yellowness; BI, browning index

[2]**The correlation is significant at the 0.01 level; *THE correlation is significant at the 0.05 level

Table 3 Moisture prediction of the PLSR models with different preprocessing methods

	RMSEC[2] (%)	RMSECV (%)	RMSEP (%)	R_c^2	R_{cv}^2	R_P^2
Preprocessing						
Raw data	3.3034	3.4154	3.4026	0.9640	0.9615	0.9616
Smoothing	3.3147	3.4268	3.4153	0.9637	0.9612	0.9609
MSC[1]	2.6648	2.7952	2.9829	0.9765	0.9742	0.9707
Normalization	2.7425	2.8699	3.1181	0.9752	0.9728	0.9679
SNV	2.6048	2.7374	2.8787	0.9776	0.9753	0.9729
SG D1	2.6793	2.8973	2.9237	0.9763	0.9723	0.9715
SG D2	3.5474	4.4155	4.8889	0.9584	0.9657	0.9329
SG D1 + MSC	2.2816	2.4965	2.6977	0.9828	0.9794	0.9754
MSC + SNV	2.6048	2.7374	2.8787	0.9776	0.9752	0.9726
Normalization + MSC	2.6644	2.7947	2.9821	0.9766	0.9742	0.9707
Normalization + SNV	2.6048	2.7374	2.8787	0.9776	0.9753	0.9726
Normalization + SG D1	2.6247	2.8833	2.9140	0.9772	0.9726	0.9715
SG D1 + SNV	2.332	2.5505	2.7010	0.9820	0.9785	0.9753

[1]SNV, standard normal variate; MSC, multiple scatter correction; SG D1, Savitzky–Golay; SG D2, Savizky–Golay

[2]RMSEC, root mean square error of calibration; RMSECV, root mean square error of cross-validation

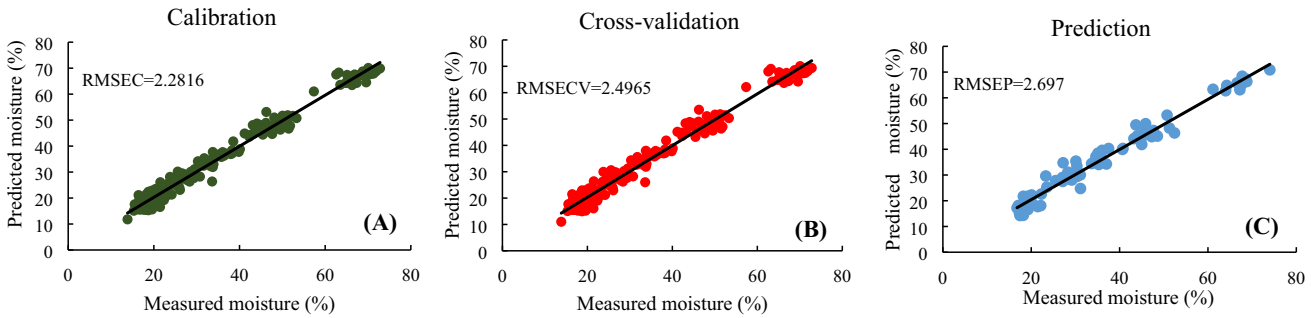
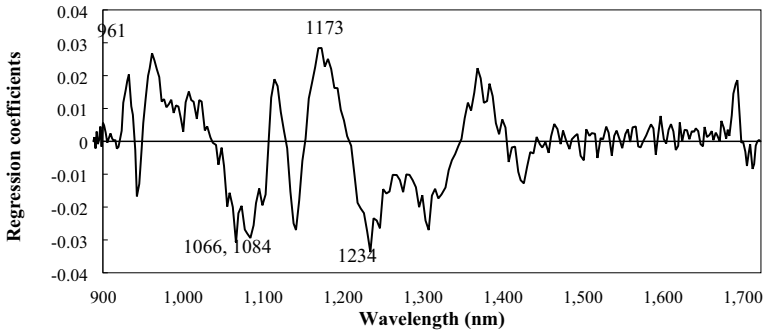


Fig. 2 PLSR model using full-wavelengths with pre-processing (Savitzky-Golay & 1st Derivative + MSC). (**A**) Results for the calibration model; (**B**) results of the full cross-validation model; (**C**) results of the prediction model



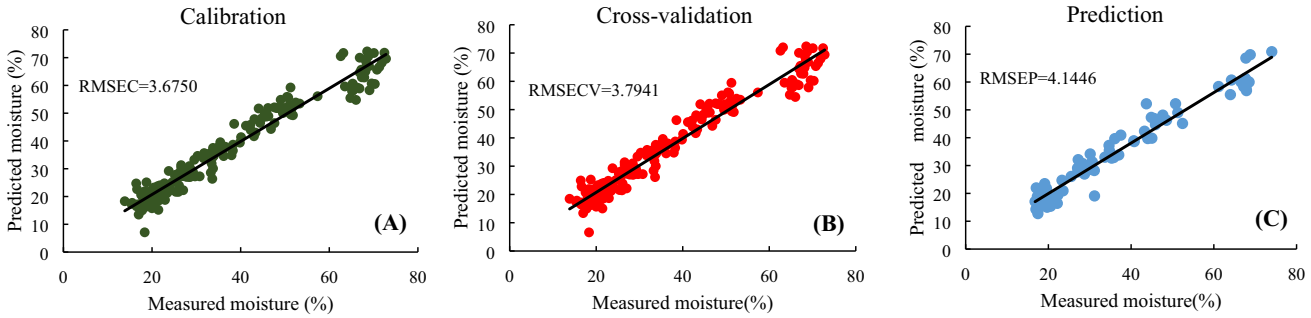


Fig. 3 Regression coefficients of the PLSR model with preprocessing (SG D1 + MSC) and results of the PLSR model using selected wavelengths with preprocessing (SG D1 + MSC). (**A**) Results of the calibration model; (**B**) Results for the full cross-validation of the PLSR model; (**C**) Results for prediction in the PLSR model

C–H_3 (Hourant et al., 2000; Zou et al., 2010). Selecting a few of the critical wavelengths in the spectral data, rather than all of them, for modeling is cost- and time-efficient when used in industry. Therefore, we performed PLSR modeling by selecting 5 influential wavelengths 961.12, 1065.50, 1083.93, 1173.23, and 1233.89 nm. The calibration, cross-validation, and prediction accuracy of the model are depicted in Fig. 3. The prediction accuracy of the PLSR model with the selected wavelength was RMSEC = 3.6750, RMSECV = 3.7941, and RMSEP = 4.1446. The RMSE value was 1–2% higher than the PLSR model using the data for the entire spectrum. R^2, which is another prediction accuracy indicator, was R_C^2 = 0.9554, R_{CV}^2 = 0.9525, and R_P^2 = 0.9521. Therefore, with this new PLSR model, the MC was predictable with an accuracy of R^2 = 0.95 or better using only five wavelengths.

Acknowledgements This manuscript do not receive any fund.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdi H. Partial least squares (PLS) regression. Encyclopedia of social sciences research methods, Sage, CA, USA. pp. 792-795 (2003)

Amodio ML, Capotorto I, Arif MM, Colelli, C.G. The use of hyperspectral imaging to predict the distribution of internal constituents and to classify edible fennel heads based on the harvest time. Computers and Electronics in Agriculture, 134: 1-10 (2017)

Bae H, Seo YW, Kim DY, Lohumi S, Park E, Cho BK. Development of Non-Destructive sorting technique for viability of watermelon seed by using hyperspectral image processing. Journal of the Korean Society for Nondestructive Testing, 36: 35-44 (2016)

Cheng JH, Sun DW. Rapid quantification analysis and visualization of escherichia coli loads in grass crap fish flesh by hyperspectral imaging method. Food Bioprocess Technology, 8: 951-959 (2015a)

Costa C, Antonucci F, Pallottino F, Aguzzi J, Sun DW, Menesatti P. Shape Analysis of agricultural products: A review of recent research advances and potential application to computer vision. Food Bioprocess Technology, 4: 673-692 (2011)

Feng YZ, Sun DW. Application of hyperspectral imaging in food safety inspection and control: A review. Critical Reviews in Food Science and Nutrition, 52: 1039-1058 (2012)

Giovenzana V, Beghi R, Romaniello R, Tamborrino A, Guidetti R, Leone A. Use of visible and near infrared spectroscopy with a view to on-line evaluation of oil content during olive processing. Biosystems Engineering, 172: 102-109 (2018)

Hong JH, Lee WY. Quality of characteristics of osmotic dehydrated sweet pumpkin by different drying methods. Journal of the Korean Society of Food Science and Nutrition. 33: 1573-1579 (2004)

Hourant P, Baeten V, Morales MT, Meurens M, Aparicio R. Oil and fat classification by selected bands of Near-Infrared spectroscopy. Applied Spectroscopy, 54: 1168-1174 (2000)

Huanga M, Wang QG, Zhang M, Zhu QB. Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology. Journal of Food Engineering, 128: 24-30 (2014)

Kim DH, Cho JS, Park JH, Kim JH, Moon KD. Quality characteristics of steamed rice cake with schisandra chinensis powder or extract added prior to storage. Korean Journal of Food Preservation, 23: 923-930 (2016)

Kim SY, Seo DW, Park JS, Kim SN, Choi YM, Nam JS, Lee JH, Kim SC, Yang MO, Hwang JB. Food composition of raw, boiled, and roasted sweet potatoes. The Korean Journal of Community Living Science, 28: 59-68 (2017)

Liu Y, Sun X, Ouyang A. Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN. LWT-Food Science and Technology, 43: 602-607 (2010)

Masry GE, Sun DW, Allen P. Near-Infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. Journal of Food Engineering, 110: 127-140 (2012)

Osborne BG, Fearn T, Hindle PH. Practical NIR spectroscopy: with applications in food and beverage analysis. Longman Scientific and Technical, Harlow, UK. p. 227 (1993)

Pu YY, Sun DW. Prediction of moisture content uniformity of microwave-vacuum dried mangoes as affected by different shapes using NIR hyperspectral imaging. Innovative Food Science and Emerging Technologies, 33: 348-356 (2016)

Rajkumar P, Wang N, Elmasry G, Raghavan SGV, Gariepy Y. Studies on banana fruit quality and maturity stages using hyperspectral imaging. Journal of Food Engineering, 108: 194-200 (2012)

Rinnan A, Berg FVD, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry, 28: 1201-1222. (2009)

Ruiz NAQ, Demarchi SM, Massolo JF, Rodoni LM, Giner SA. Evaluation of quality during storage of apple leather. LWT-Food Science and Technology, 47: 485-492 (2012)

Shin MY, Lee WY. Physical properties and preference of a steamed sweet potato slab after mild hot air drying. Korean Journal of Food and Cookery Science, 27: 73-81 (2011)

Sivakumar SS, Qiao J, Wang N, Gariépy Y, Raghavan GSV, McGill J. Detecting maturity parameters of mango using hyperspectral imaging technique. American Society of Agricultural and Biiological Engineers Meeting Presentation, 066183 (2006)

Suh HJ, Chung SH, Choi YM, Bae SH, Kim YS. Changes in sugar content of sweet potato by different cooking methods. Korean Society of Food Science and Technology, 14: 182-187 (1998)

Sun DW. Computer vision g maturity parameters of mangocontact quality evaluation tool for the food industry. Journal of Food Engineering, 61: 1-2 (2004)

Vidal M, Amigo JM. Pre-processing of hyperspectral images. Essential steps before image analysis. Chemometrics and Intelligent Laboratory Systems, 117: 138-148 (2012)

Williams P, Geladi P, Fox G, Manley M. Maize kernel hardness classification by near Infrared (NIR) hyperspectral imaging and multivariate data analysis. Analytica Chimica Acta, 653: 121-130 (2009)

Williams P, Norris KH. Variable affecting near infrared spectroscopic analysis. In: Williams, P., Norris, K.H. (Eds.), Nearinfrared Technology in the Agriculture and Food Industries, second ed. The American Association of Cereal Chemists, St. Paul, MN, USA. pp. 171–185 (2001)

Workman J, Weyer L. Practical guide to interpretative near-infrared spectroscopy. CRC Press, Inc., Boca Raton, FL, USA (2007)

Wu D, Shi H, Wang SJ, He Y, Bao YD, Liu KS. Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. Analytica Chimica Acta, 726: 57-66 (2012)

Yang J, Chen JF, Zhao YY, Mao LC. Effects of drying processes on the antioxidant properties in sweet potatoes. Agricultural Sciences in China, 9: 1522-1529 (2010)

Ye MQ, Yue TL, Yuan YH, Li Z. Application of FT-NIR spectroscopy to apple wine for rapid simultaneous determination of soluble solids content, pH, total acidity, and total ester content. Food and Bioprocess Technology, 7: 3055-3062 (2014)

Zou XB, Zhao JW, Povey MJW, Holmes M, Mao HP. Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta, 667: 14-32 (2010)