



A non-destructive determination of protein content in potato flour noodles using near-infrared hyperspectral imaging technology

Jing Zhang^a, Zhen Guo^a, Zhishang Ren^c, Sihua Wang^a, Xiang Yin^a, Dongliang Zhang^a, Chenjie Wang^a, Hui Zheng^b, Juan Du^{a,*}, Chengye Ma^{a,*}

^a School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, China

^b Gaoqing Inspection and Testing Center, Zibo 265400, China

^c Zibo Institute for Food and Drug Control, Zibo 255000, China



ARTICLE INFO

Keywords:

Hyperspectral imaging
Potato flour noodles
Protein content
Multivariate analysis
Visualization

ABSTRACT

Hyperspectral imaging (HSI) with the near-infrared (NIR) bands (900–2250 nm) was employed to investigate the non-destructive prediction of protein content in potato flour noodles. The protein content (8.906–10.515 %) of 120 potato flour noodles was studied to establish the prediction model. Partial least squares regression (PLSR) model was established based on the spectra of potato flour noodles to predict the protein content showing high performance. After optimization, orthogonal signal correction (OSC) was used to preprocess the original spectra, and the competitive adaptive reweighted sampling algorithm (CARS) was chosen to select characteristic wavelengths, therefore, OSC-CARS-PLSR was established. Next, 77 samples were selected as the calibration set, and the remaining 38 samples were used as the prediction set. The coefficient of determination (R^2) and the root mean square error (RMSE) were used to evaluate the performance of the model. OSC-CARS-PLSR showed a high performance with R^2 values of 0.9606 and 0.8925 and RMSE values of 0.070 % and 0.1385 % in the calibration set and prediction set, respectively. The visualization image was used to identify protein distribution in potato flour noodles. Overall, the results indicate that HSI technology could accurately predict the protein content in potato flour noodles providing a rapid and non-destructive method to detect protein and other compositions in grains and foods.

1. Introduction

Noodles made from wheat flour are popular worldwide [1]. There are many kinds of noodles around the world, such as pasta, udon noodles, cold noodles, and Chinese noodles. Chinese noodle is one of the staple foods with a large share in the category of convenience foods. Potato is rich in dietary fibre, protein, vitamin C, minerals, and potassium, etc. [2–4]. The noodles made with whole potato flour meet the requirements of potato staple food and have a huge market. Nonetheless, the development of the potato industry is affected by the detection technology of the main ingredients, including starch, protein and lipid. Protein is an essential indicator that could determine the texture, nutrition and cooking quality of the noodles. However, the detection of protein content is mostly done using chemical methods, which are destructive, hazardous and environmentally harmful at present [5]. Thus, developing a rapid, non-destructive, accurate and reliable protein detection technique for noodles is necessary.

Spectral technologies, such as infrared spectrum technology, hyperspectral imaging (HSI) technology, and Raman spectrum technology, have been widely used in the field of food quality and safety detection. In recent years, infrared spectral analysis technology has played a critical role in agricultural product and food quality detection, such as the detection of acidity index in peanuts [6], rapid recognition of Auricularia auricular varieties [7], and determination of the protein content in peanut [8], and moisture content in green tea [9], which has the advantages of rapid detection and non-destructive determination, and makes up for the limitation of destroying samples by the traditional detection methods. Meanwhile, Raman spectroscopy has been widely used in dairy and agricultural products in recent years, such as for rapid evaluation of milk acidity and identification of milk adulteration [10], detection of starch in banana fruit [11], fast identification of soybean varieties [12], and distinguishing rice producing areas [13]. However, infrared and Raman spectroscopy cannot reflect the spectral distribution of the whole field of view and spectral technology is susceptible to

* Corresponding authors.

E-mail addresses: dujuan0427@163.com (J. Du), mccycn2002@163.com (C. Ma).

background noise [14].

Compared with infrared spectroscopy and Raman spectroscopy, HSI technology can collect the spectral and spatial information of the detected objects at the same time; therefore, this technology can detect the external quality of the object as machine vision technology and the internal quality of the object as spectrum technology. Near-infrared (NIR) HSI provides information on vibrations of overtones and combination bands such as O—H (water, carbohydrates), N—H (proteins), and C—H (lipids) [15]. Therefore, NIR-HSI technology coupled with chemometrics could detect protein and other substances qualitatively and quantitatively. HSI is composed of hundreds of single-band images at a specific wavelength of the NIR spectrum for each pixel [5]. Each spectrum contains information on samples that could be extracted from the spectrum by the chemometrics regression algorithm, and quantitative determination could be performed.

The application of NIR-HSI technology in food quality and safety evaluation has been widely developed [16] to determine the chemical composition, such as meat [17], grain [18], vegetables, and fruits [19], etc. However, only a few studies have applied NIR-HSI detection on potato flour noodles. Traditional spectroscopic techniques have significant limitations in the determining thin and flat samples, such as noodles, as it is not possible to obtain accurate spectral information of a whole noodle to assess the overall level of noodle protein. Noodles are more commonly consumed in China, and with the stapling of potatoes, the process of mixing whole potato whole flour and wheat flour to make noodles is gaining increasing interest.

HSI technology based on chemometrics and composition prediction technology is usually combined with spectral preprocessing technology [20]. In this study, the NIR-HSI technology was employed to forecast the protein content in potato flour noodles. Data analysis plays an important role in the HSI of food component detection, and the accuracy could be improved by spectral preprocessing, characteristic wavelength extraction and various modeling methods [21]. The optimal algorithm was established by combining different algorithms at each step. This work followed these steps: the model was compared to select the best preprocessing algorithm; the optimal regression algorithm was confirmed

by comparing three regression methods of partial least squares regression (PLSR), principal component regression (PCR), and support vector regression (SVR); three algorithms of successive projections algorithm (SPA), uninformative variable elimination (UVE) and competitive adaptive reweighted sampling (CARS) were used to extract the characteristic wavelengths; finally, the distribution of protein content in the potato flour noodle samples was depicted by the developed image processing algorithms. Fig. 1 shows the main steps of this study.

This work verified the feasibility of NIR-HSI technology for rapid, non-destructive detection of protein content in potato flour noodles, visualized the protein content and distribution, and promoted the development of infrared technology in the field of food quality detection.

2. Materials and methods

2.1. Prepare samples

Potato whole flour does not contain gluten protein, the processing performance of noodles made from potato whole flour is poor, so it needs to be mixed with wheat flour to make noodles [22]. Wheat flour (Zhejiang Cereals, Oils and Foodstuffs Import and Export Co., Ltd.) was blended with a random proportion of potato whole flour (Gansu Zhenyang Modern Agriculture Service Co., Ltd.) between 0 and 35 % and mixed using the dough mixer (Zaoyang Juxin Machinery Co., Ltd.) with an addition of water (total moisture content 32 %) to make potato flour noodles samples [23].

2.2. Compositional analysis

After collecting the spectral data, the potato flour noodles were grinded and screened. The crude protein content was estimated based on AOAC 960.52 [24]. The reference protein content of samples were measured by the automatic Kjeldahl apparatus (K9860, Shandong Haineng Scientific Instrument Co. Ltd). A total of 0.5000 g grinded potato flour noodle sample, 3 g Kjeldahl catalyst and 10 mL 98 % sulfuric acid

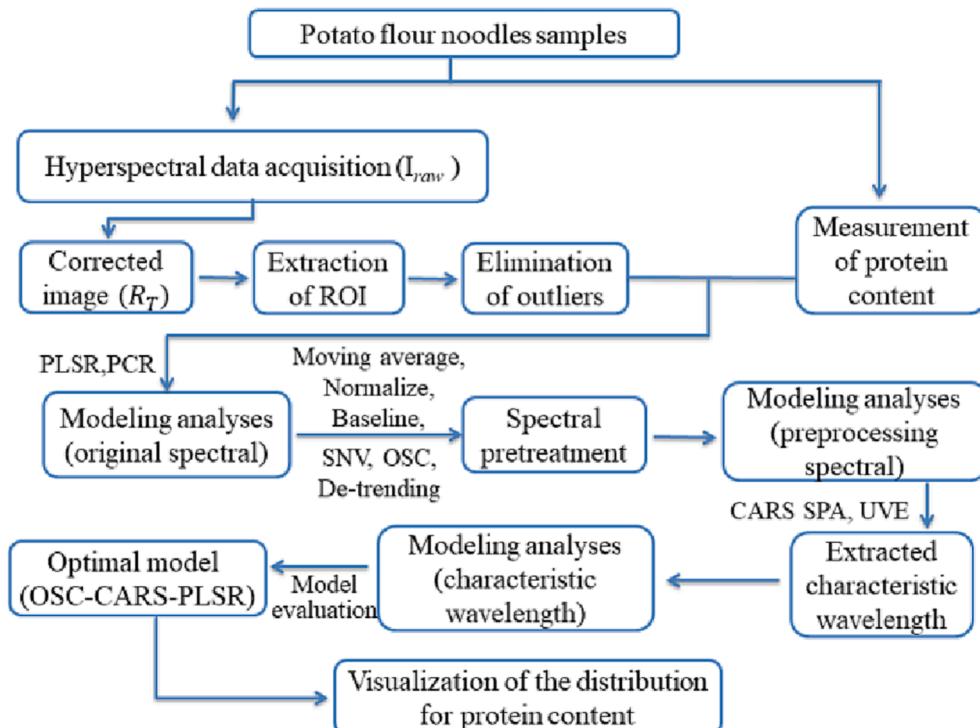


Fig. 1. Schematic diagram illustrating the workflow of data processing.

were added in a glass digestive tube, respectively. Then, the glass digestive tube was heated at 420 °C for 4 h until the liquid became transparent. Next, the glass digestive tube was cooled at room temperature. Afterward, put the glass digester into the automatic Kjeldahl nitrogen apparatus, set the parameters as 20 mL water and 40 mL 400 g/L sodium hydroxide solution were added into the glass digestive tube, and the glass digestive tube was heated again until boiling the liquid for 5 min, and the steam was received by 20 mL 20 g/L boric acid solution with 3 drops of methyl orange-bromocresol green mixture indicator. Finally, the boric acid solution that absorbed the steam was titrated by 0.1000 mol/L hydrochloric acid solution. The measured protein contents of samples ranged from 8.906 to 10.515 % with the mean value of 9.777 % and a standard deviation value of 0.397 %.

2.3. Spectral data acquisition and preprocessing

2.3.1. HSI data measurement

The HSI system was applied to obtained reflective hyperspectral images, ranging from 900 to 2500 nm (Isuzu Optics Corp., Taiwan, China). The schematic of the HSI equipment was shown in Fig. 2. The main hardware device for the HSI system includes a line scanning spectrometer, charged couple device (CCD) camera, two halogen lamps, a conveyor belt regulated by a stepper motor, a camera obscura and a computer. The specific models and manufacturers are shown in Table 1. The experimental data were collected by the software (Spectral Image software, Isuzu Optics Corp., Taiwan, China).

2.3.2. Acquire and calibrate image

The samples spectral information was collected by the HSI system with a band of 900–2500 nm. After that, the information was calibrated by using black reference and white reference.

To improve the accuracy of hyperspectral data and reduce the impact of changes in illumination intensity on hyperspectral data, the original gathered HSI information was corrected to obtain the spectral relative reflection information. The calculation formula was as follows (1) [25].

$$R_T = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}} \quad (1)$$

Where R_T , I_{raw} and I_{dark} , I_{white} represented the corrected and the raw hyperspectral image, and the blackboard and whiteboard calibration

Table 1
HSI system.

Component name	Model	Manufacture factory
Line scanning spectrometer	V10E	Specim, Spectral Imaging ltd., Oulu, Finland
CCD camera	DL-604 M	Andor, Ireland
Halogen lamp	IT 3900e	Illumination Technologies Inc., New York, USA
Camera lens	OLE23	Schneider, German
Conveyor belt	IRCP0076-1COMB	Isuzu Optics Corp., Taiwan, China

image, respectively.

Each collected HSI was a three-dimensional (3-D) data cube including a set of two-dimensional images arranged in sequence at different wavelengths [26]. HSI could be regarded as spectral $I(l)$ of each individual pixel (x, y) , or image $I(x, y)$ of each individual wavelength l [27]. The spectral data of reflectivity and wavelength can be obtained automatically by selecting the scanned image of each sample with the spectral analysis software of the Isuzu optical instrument.

2.3.3. ROI selection

The region of interest (ROI) was selected from the HSI after black calibration and white calibration, and the average spectrum of ROI was used to improve the reproducibility of spectral information of potato flour noodles [28]. The sample spectrum and the background can be easily distinguished because they have significantly different spectral information; the sample can be separated from the background by simple threshold segmentation [29]. Due to the obvious spectral contrast between potato flour noodles and the background, a single-band grayscale image at 1396 nm was applied to establish mask templates for NIR hyperspectral images. The OTSU method was applied for threshold segmentation of the acquired images [30]. A binary mask was built for each sample, assigning values 0 and 1 to the pixels corresponding to the background and potato flour noodles, respectively. The average spectrum of each image was obtained by averaging the index value of each potato flour noodles pixel and 120 spectral data points were produced. The spectral information of potato flour noodles was concatenated to acquire the final spectral matrix used to establish the model. In this matrix, the rows and columns represented the number of

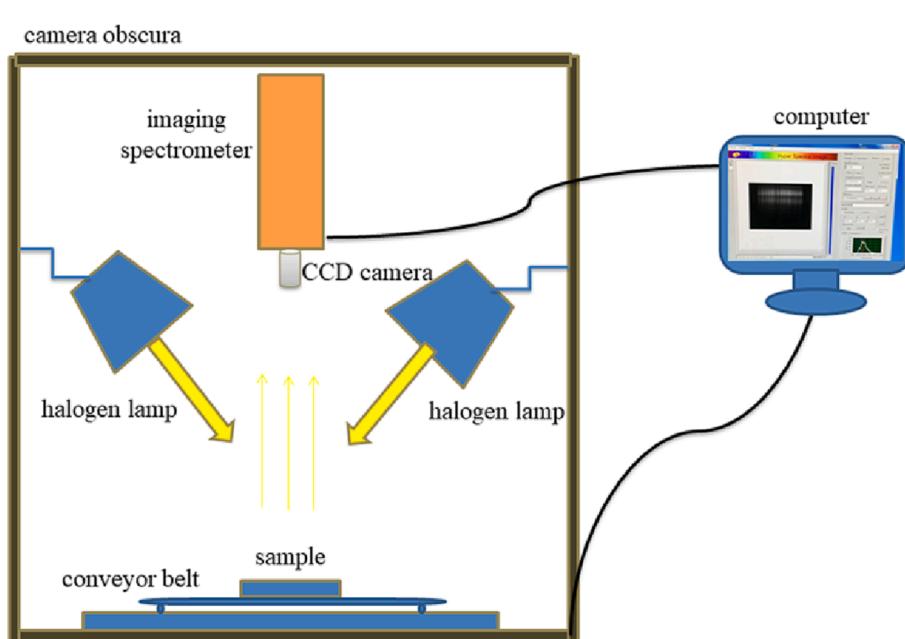


Fig. 2. Schematic of HSI equipment.

potato flour noodles images and the wavelengths, respectively [31]. In consequence, the 120×230 dimensional spectral matrix was obtained for protein content prediction. The specific operation processes are reported in Fig. 3. The spectral image extracted from the software HSI Analyzer (ITT Visual Information Solutions, Boulder, CO, USA).

2.4. Sample and spectral image multivariate analysis

2.4.1. Sample set division

The research samples could ensure the performance of the model by being reasonably partitioned into the calibration set and prediction set. The samples in the calibration set should be representative and include the extreme value. In general, the ratio between the calibration and prediction sets could be between 2:1 and 3:1. The sample set division methods commonly include random method (RS), Kennard-Stone method (KS), and sample set partitioning based on joint X-Y distance method (SPXY).

In the SPXY method, based on the KS method and first proposed by Galvao [32], when the spatial distance of the samples was calculated, not only spectral variables but also reference variables were considered. The advantage of SPXY was that the partition sample set synchronously considered the spectral and concentration matrix, improving the capability of the model [33]. The distance between samples was calculated using the following equations [34] (2)–(4):

$$d_x(p, q) = \sqrt{(x_p - x_q)^2} = |x_p - x_q|; p, q \in [1, n] \quad (2)$$

$$d_y(p, q) = \sqrt{(y_p - y_q)^2} = |y_p - y_q|; p, q \in [1, n] \quad (3)$$

$$d_{x,y}(p, q) = \frac{d_x(p, q)}{\max_{p,q \in [1,n]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p,q \in [1,n]} d_y(p, q)}; p, q \in [1, n] \quad (4)$$

where $d_x(p, q)$ was the distance of spectrum and $d_y(p, q)$ represents the distance of the chemical measurements.

2.4.2. Outliers elimination method

In the process of hyperspectral images acquisition, environment and instrument would cause data errors. Therefore, the outliers were eliminated in this study [35]. Monte Carlo cross-validation was used to eliminate outliers, which filtered the singular samples by taking spectral information and checking the physical and chemical reference values [36].

The specific steps were as follows: all the samples were used as calibration set to establish the PLSR model. The best performance principal component number was confirmed according to the lowest RMSECV. Commonly, outliers and normal samples in the two distributions had a significant difference. Therefore, a suitable tuning value (2–2.5 times the average value of the main body of the data set) was used to classify outliers and normal samples according to real circumstances [37]. Therefore, Monte Carlo cross-validation method was used to calculate the standard deviation (STD) and mean of the prediction error distribution for each sample. The 2.5 times values of mean and STD were taken as the thresholds, and a mean-STD scatter plot was drawn based on the calculated relevant parameters. Generally, the samples not in the threshold range were considered outliers.

2.4.3. Preprocessing algorithms

Various preprocessing algorithms were widely used in the models to increase the significant information including the spectra and to remove the inessential data from the interference of noise, background and the light scattering influence [38]. In this study, six preprocessing methods, including Baseline, Standard normal variate (SNV), De-trending, Orthogonal signal correction (OSC), Moving average, and Savitzky-Golay were used to preprocess the raw spectral. As a common

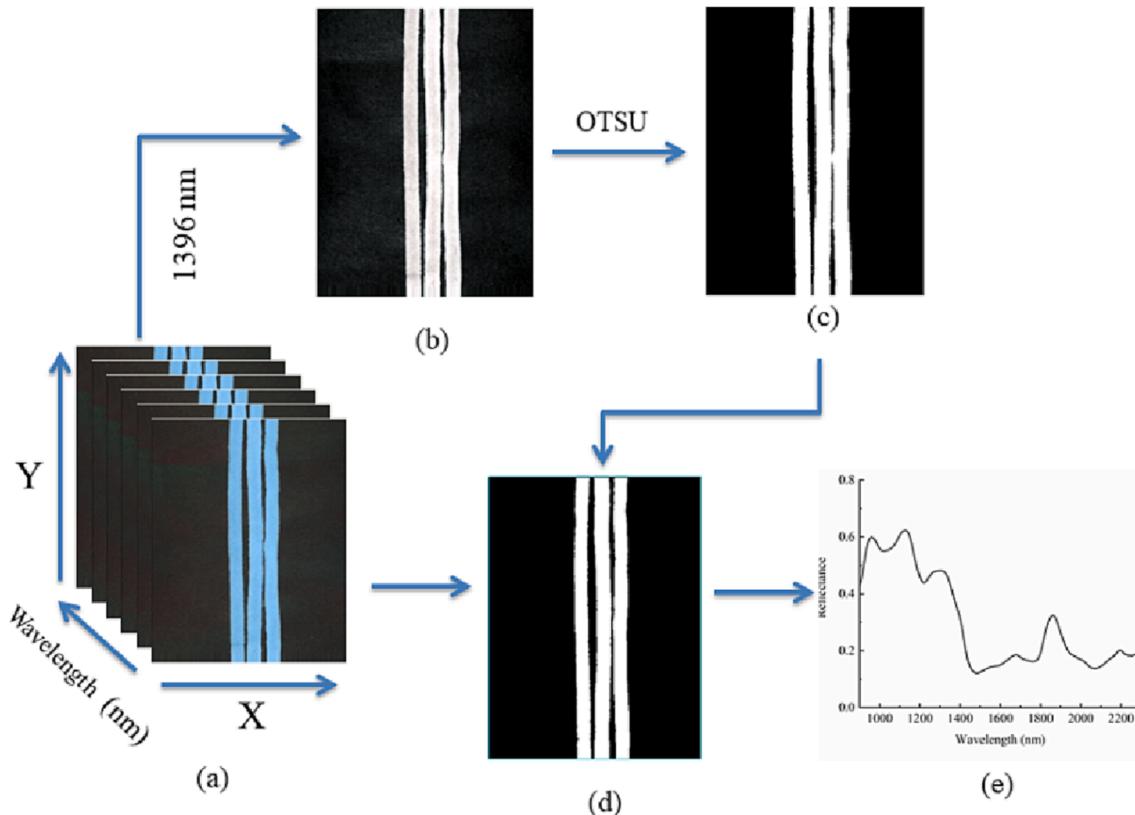


Fig. 3. The process of hyperspectral images segmentation, (a) raw hyperspectral image, (b) 1396 nm gray image, (c) mask image, (d) mask hyperspectral image after setting all background pixels to zero, (e) average spectral curve.

preprocessing algorithm, Baseline can effectively eliminate the baseline offsets and tilt caused by instrument background and sample surface inhomogeneity from spectral data. SNV was applied to eliminate spectral errors caused by optical path changes or sample dilution. To eliminate the baseline drift from the spectrum De-trending was mainly arithmetic, which was usually used for spectra after SNV processing, or could be used separately. De-trending was beneficial for removing nonlinear trends in spectroscopic data [39]. OSC was an effective algorithm that could eliminate the variation of a spectra matrix X that was quadrature to the response matrix Y [40]. And the OSC could correct spectra as a widespread algorithm by removing the unacceptable systematic irrelevant data. The moving average was one of the empirical tools for smoothing data [41]. Savitzky-Golay was calculated by derivative and polynomial calculation to reduce the original spectral noise [42].

2.4.4. Characteristic wavelengths selection

The samples spectral data were characterized by high dimensionality with redundancy, multi collinearity among contiguous wavelength bands, which would inevitably reduce the calculation speed of the model development and prediction performance [43]. Therefore, selecting the characteristic wavelengths with the exhibit of high sensitivity and relevant data could accelerate the calculation speed and simplify the modeling process of the prediction models [44]. However, the relevant characteristic wavelength algorithms for establishing models were proposed for different samples, and there was no unified optimal wavelength selection method. In this investigation, the characteristic wavelengths were selected by three algorithms including CARS, SPA and UVE. The CARS was a potential method to establish high-performance calibration model, which imitated the survival of the fittest in Darwinian evolution theory, selected and eliminated variables step by step through full spectra combined with PLSR [45]. SPA was used to solve the problem of minimum redundancy collinearity, which established a candidate variable subset based on a series of projection operations related to the instrument response matrix column [46]. UVE was a method that the independent variables with non-contribution to the calibration model were eliminated and then retained the useful information from the remaining wavelengths [47]. This investigation primarily established a full wavelength model, after that selected optimal wavelengths by three methods to establish the simplified models.

2.4.5. Model construction and evaluation

Establishing the model was a significant step in multivariate data analysis. Two chemometrics quantitative calibration models, namely, PLSR and PCR models were built to reflect the relation of spectra and reference value [48]. PLSR and PCR were robust and reliable multivariate methods that have been widely applied to establish fast and non-destructive prediction parameter models to determine the quality of agricultural products [49]. When there was a linear relationship between the independent variables and identified dependent variables, PLSR was especially propitious in this condition [50]. PCR was another multivariate collinearity algorithm to process the experimental images and information in chemistry and tends to model among the variables by fitting a linear equation to the observed data [49].

Besides, some characteristic indices were calculated to assess the predictive power of the PLSR and PCR models, including the determination coefficients in calibration (R_C^2), prediction (R_P^2), and the root mean squared errors by calibration and prediction (RMSEC and RMSEP, respectively). Normally, the well-performing model tends to have a high-value of R_C^2 and R_P^2 and a low-value of RMSEC and RMSEP, and they have a slight difference [51]. Each model was constructed several times to test the reproducibility of the model, and the results were found to be indistinguishable, indicating that the model approach was independent and reliable for the data. The R^2 and RMSE were defined using the following equations (5)–(6):

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (6)$$

where \hat{y} was the predicted value, y was the measured value, \bar{y} refers to the mean value of samples, and n represents the number of sample spectra.

2.5. Visualization of protein distribution

Compared with the traditional spectroscopic technique, the HSI technique could create images with different colors to present the spatial pattern of the component content, so that it has the individual advantage in efficient chemical imaging instruments. Currently, each pixel of the HSI could be transformed into visual images of concentration gradients with different colours by using the optimized models. And the visual images would predict the protein content allocation in potato flour noodles. The colours expressed content variation and distinguished the composition by relying on the colour bar showed on the right side of the visual images [52]. This corresponding pseudo-color map could obtain the distribution of the protein content in potato flour noodles.

3. Results and discussion

3.1. Features of spectral

Fig. 4 shows the raw average curves of the spectra reflectance obtained from 120 samples with different protein contents. In the wavelength range the waveforms revealed a similar spectral trend for different potato flour noodles. Nevertheless, when the wavelength exceeded 2250 nm, the trend of the spectrum curves was irregular, which could be attributed to the excessive energy at the scanning step and the decreased noise reduction effect. Therefore, in the present study, the range of hyperspectral wavelength was 900–2250 nm and the spectral reflectivity was 0.1–0.7, which were selected for the subsequent work. The essential absorption bands represented by the peaks in **Fig. 4** were the absorption covalent bonds such as O—H, C—H, N—H, and C—O bonds and several amides and carboxyl groups with their strong overtone effect and synergistic effect respectively [53].

The essential characteristic peaks of the protein appeared at 962,

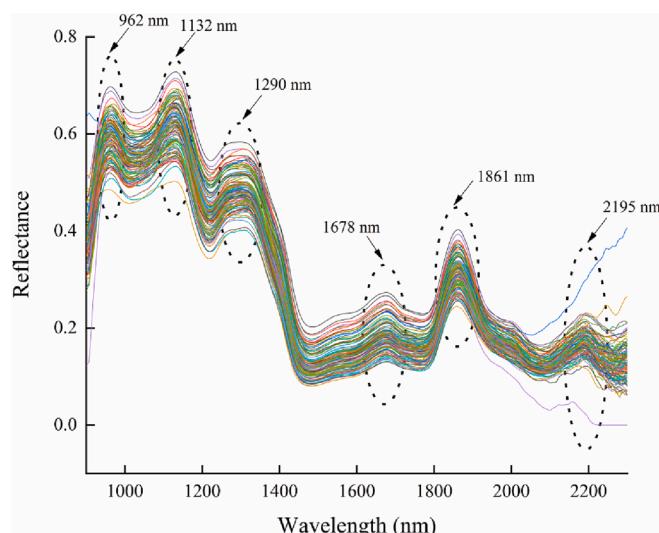


Fig. 4. Raw average curves of spectral reflectance.

1132, 1290, 1678, 1861 and 2195 nm. Most of these peaks indicated the presence of C—H, O—H and N—H covalent bonds, overtone or frequency-doubling absorption. The band near 962 nm was formed by the combined effect of O—H, which primarily represents carbohydrates. The combined frequency vibration of C—H caused the spectral feature peaks at 1132 nm. The feature peaks at 1290 nm was formed by the in-plane bending of C—H. The spectral reflectance values at 1678 nm appeared due to the effect of the amide groups [54–55]. The feature peaks of protein appeared obviously after 1600 nm, specifically in the range of 2050–2200 nm. The feature peak at 1861 nm was formed relevant to the absorption of water molecules. The spectral reflectance values at 2195 nm represented the connected absorption values of C—H and N—H.

3.2. Eliminating outliers

The accuracy and robustness of the models were prominently influenced by abnormal samples, therefore, it's necessary to exclude the outliers prior to establishing the model. Generally, based on the PLSR model to identify and remove the outliers, even a single outlier could have a remarkably greater and harmful effect, causing misleading statistics. However, PLSR mainly focuses on solving the spectral information from samples irrespective of the concentration information, this study used the Monte Carlo cross-validation method to exclude outliers (Fig. 5), which was more efficient. Monte Carlo-Partial Least Squares (MCPLS) [56], based on the Monte Carlo cross-validation methods, has the advantage of simultaneously detecting spectral outliers and physical and chemical reference outliers. The specific method was to take all the sample sets as the correction set and establish the model. The minimum RMSECV was the best principal component score. Moreover, the characteristic statistical parameters (mean and standard deviation) of the prediction error of each sample were determined. The threshold samples were screened by adopting and taking 2.5 times the average of the two data of MEAN and STD, and generally, the samples not in the threshold range were considered outliers.

The samples circled in red (23, 38, 41, 48, 76, 90, 108, 110 and 112) that had prominent mean or STD were regarded as potential outliers (Fig. 5). As shown in Table 2, the PLSR prediction model R^2_C and cross-validation RMSECV for the protein content in potato flour noodles were based on the 9 abnormal samples and all other samples. In Table 2, the outlier samples were singly eliminated, when sample numbers 41, 48, 110, 108, and 76 were eliminated, the R^2_C of the model improved. However, the R^2_C of the model decreased after the elimination of sample

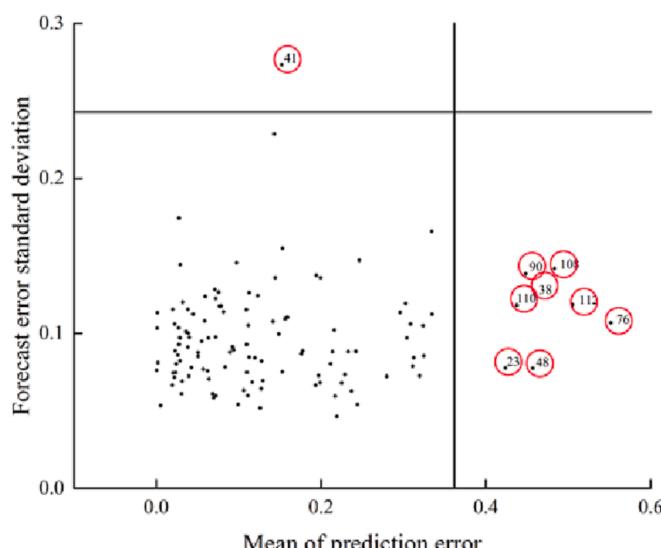


Fig. 5. Monte-Carlo cross validation method to eliminate protein outliers.

Table 2
PLSR model of eliminating outliers.

The number of outliers sample	No. Principal components	RMSECV/%	R^2_C	RMSEC/%
All samples	8	0.1412	0.9174	0.1151
41	8	0.1420	0.9180	0.1145
23	8	0.1425	0.9166	0.1161
48	8	0.1395	0.9199	0.1138
110	8	0.1393	0.9181	0.1137
38	8	0.1412	0.9171	0.1146
90	8	0.1396	0.9171	0.1141
108	8	0.1410	0.9182	0.1150
112	8	0.1403	0.9170	0.1140
76	8	0.1407	0.9183	0.1137

numbers 23, 38, 90 and 112. After eliminating the outliers, the PLSR model showed better performance in terms of prediction, the differences were smaller between RMSEC and RMSECV, and the overall function was obviously enhanced. Therefore, sample numbers 41, 48, 110, 108 and 76 were designed as outlier samples, and the rest 115 samples were used in the subsequent work.

3.3. Sample set partition

After excluding the outliers, according to the ratio of 2:1, the SPXY sample segmentation method selected 77 samples from the remaining 115 samples as the calibration set [57]. The rest 38 samples were used as the prediction set to evaluate the performance of the model. The range of the protein content in the calibration set (8.906–10.515 %) was covered in the prediction set (9.019–10.510 %) (Table 3), indicating that the division method was reasonable and acceptable.

3.4. Spectral pretreatment

PLSR and PCR models were established to compare the protein content in potato flour noodles according to the original spectral message. The effectiveness of the discrepancy pretreatment methods was verified by establishing the PLSR and PCR models, and the results were shown in Table 4. The model performance of protein content could be evaluated by different spectral pretreatment algorithms. Based on mathematical preprocessing, certain irrelevant variable noise interference, the scattering amplitude and baseline drift could be decreased.

It was evident that four pretreatment algorithms were beneficial in improving the PLSR prediction model compared with the original data without preprocessing under the same model (Table 4). The PLSR model based on the spectral data and corrected by the OSC emerged with perfect accuracy, wherein the R^2_C value was 0.9641 and RMSECV was 0.1158 %, and the number of principal components reduced from 8 to 6. As for the PCR model with different pretreatments, half of the methods increased the accuracy, especially spectra preprocessed by Moving Average filter produced models with a better capacity, and the R^2_C was 0.9014 and RMSECV was 0.1550 %. The availability of the other pretreatment methods decreased possibly, caused by the spectral correction losing useful details and information. When considering the full spectrum, the PLSR model possessed higher capability than that of PCR. In this study, the leave-one-out cross-validation (LOOCV) method was used

Table 3
Statistical results of protein content in potato flour noodles by SPXY method in the calibration set and prediction set.

Sample set	Number of samples	Protein content (%)			
		Min	Max	Mean	SD
Calibration	77	8.906	10.515	9.804	0.3849
Prediction	38	9.019	10.510	9.749	0.4284

SD: Standard Deviation.

Table 4

The protein content prediction results for PLSR and PCR model with different preprocessing methods.

model	Pretreatment methods	No. Principal components	RMSECV / %	calibration set	
				R _C ²	RMSEC / %
PLSR	None	8	0.1429	0.9364	0.0964
	Moving average	9	0.1475	0.9443	0.0902
	Normalize	8	0.1511	0.9284	0.1023
	Baseline	12	0.1351	0.9561	0.0801
	SNV	10	0.1397	0.9625	0.0741
	De-trending	8	0.1502	0.9314	0.1002
	OSC	6	0.1158	0.9641	0.0724
	None	16	0.1573	0.8978	0.1222
PCR	Moving Average	16	0.1550	0.9014	0.1201
	Normalize	17	0.1604	0.8973	0.1225
	Baseline	17	0.1596	0.8963	0.1231
	SNV	15	0.1550	0.8983	0.1219
	De-trending	16	0.1587	0.8936	0.1247
	OSC	9	0.1392	0.8982	0.1220

to establish the calibration set under complete cross-validation using the PLSR model. Based on the consequences of OSC preprocessing of the PLSR model, it was necessary to use spectrum pretreatment in the current study.

3.5. Extracted characteristic wavelength

In this study, the optimal wavelength selection for protein content in potato flour noodles was acquired by comparison of the following three types of algorithms.

3.5.1. Selecting characteristic wavelengths by CARS

To decrease the irrelevant spectral message, the CARS algorithm was used to select the characteristic wavelengths of potato flour noodles. The

samples with full spectral information and protein content in the calibration set were considered entering values. The amount of Monte Carlo sampling runs was installed as 200, and the wavelengths chosen were decided by RMSECV. Fig. 6 exhibit the variation trend with the number of wavelengths, RMSECV and the regression coefficient path with an increase in Monte Carlo sampling runs, respectively. The process of selecting characteristic wavelength by CARS had two steps roughly and precise selection (Fig. 6a). Particularly, when the number of sampling runs was smaller than 80 the number of sampled variables was sharply reduced and then gently decreased. In the inception phase, the chosen wavelength number rapidly declined on account of the employment of exponentially decreasing function and then gradually tended to stabilize [58]. RMSECV cross-verified value (Fig. 6b) declined with sampling runs 1–80 because the irrelevant wavelength was removed, which then changed slightly with sampling runs 80–94. When the sampling runs number was 94, the RMSECV reached the lowest value of 0.1118 %. Ultimately, the values rose fast when the sampling runs exceeded 94, due to some worthy and relevant wavelengths were excluded, thus, stopping the process when the sampling runs was 94. As shown in Fig. 6c, each line expressed the regression coefficient path of each variable at different sampling runs. At the beginning of the curve, the regression coefficient value of the whole wavelength was close to 0. The blue line represents the optimal wavelength subset with the minimum RMSECV. In particular, when the RMSECV value jumped to the new level at dotted line L1, the coefficient of wavelength (point P1) decreased to 0, thereby removing this wavelength. With the wavelength elimination, the precious of the model declined quickly. On the basis of these results, the consequence of the characteristic variable was concluded. Therefore, after employing the CARS algorithm, the number of characteristic wavelengths accounted for 10.9 % of the total wavelength, leading to 25 characteristic wavelengths (1007, 1057, 1126, 1290, 1408, 1618, 1654, 1666, 1714, 1720, 1744, 1762, 1779, 1803, 1815, 1884, 1986, 2025, 2041, 2085, 2164, 2174, 2180, 2190 and 2195 nm). Finally, a subset of characteristic wavelengths was determined to predict the protein content.

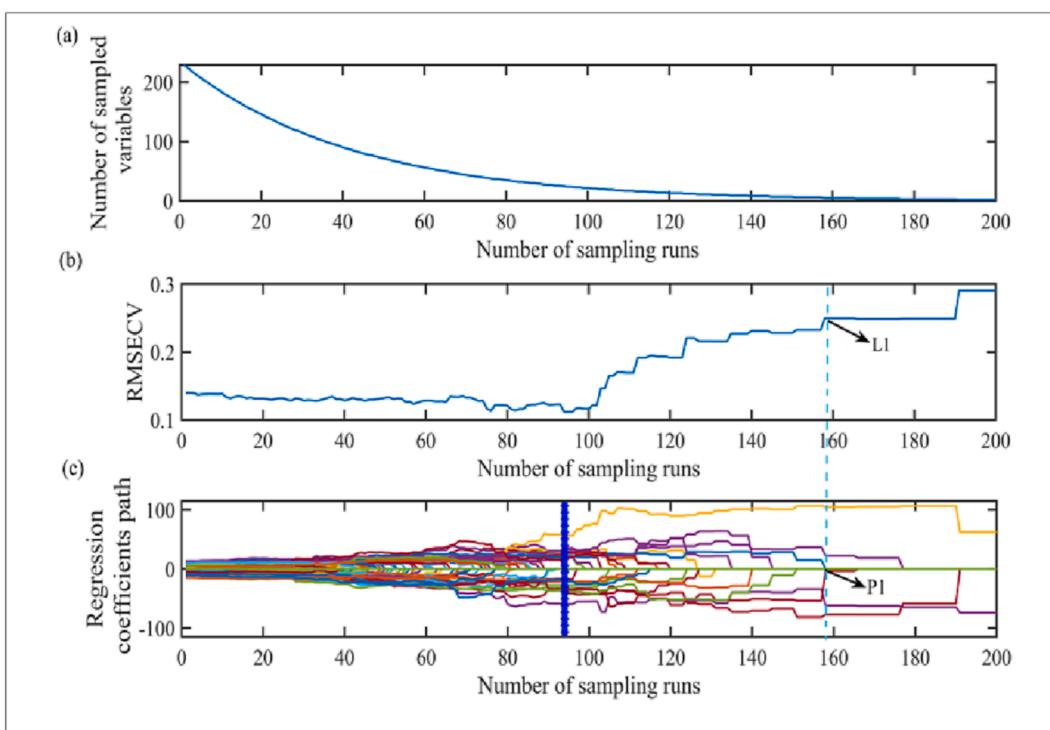


Fig. 6. The process of screening characteristic wavelength by CARS (a) the number of selected wavelengths in the change process diagram of CARS algorithm, (b) RMSECV value changes with the sampling runs, (c) the change process of regression coefficient of each variable with sampling times. The blue line represents the position with the lowest RMSECV. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5.2. Selection of the characteristic wavelengths with SPA

The SPA was extensively applied in characteristic wavelength choices for hyperspectral data [59]. SPA has the advantage of minimizing multicollinearity in vector space and excluding redundant messages of the primary spectral matrix. When the SPA operation counts the RMSE of the model, the optimal wavelengths were decided by the subset with the smallest RMSE [60]. As shown in Fig. 7a, a sharp decline occurred at the beginning part of the RMSE curve, with the number of selected useful wavelengths increasing from 1 to 13. When the number of selected optimal wavelengths position (red square) was 13, it corresponded to the minimum value (RMSE = 0.1361 %). The unique nature of the SPA selected for the 13 optimal wavelengths was expounded in Fig. 7b, wherein the abscissa was the serial number of wavelengths; the characteristic values were respectively 919, 938, 1007, 1057, 1145, 1340, 1408, 1684, 1913, 1975, 2014, 2090 and 2169 nm, accounting for 5.7 % of the full spectral band. The intensity of reflectance with different wavelengths in the HSI was related to the protein content in potato flour noodles. Thus, the protein content in potato flour noodles could be detected by the discrepancy in spectral reflectance.

3.5.3. Selected characteristic wavelengths for the UVE

UVE algorithm was a method of variable selection based on a stability analysis of PLS regression coefficients, used to analyze the stability of each wavelength with the corresponding regression coefficients of the PLS model, and then excluded the wavelength variables with low absolute correlation coefficients [61]. Fig. 8 confirmed the stability values of the acquired variables for protein content in potato flour noodles. The two horizontal dashed lines performed the UVE cut-off threshold value. The wavelength whose stability was interiorly the thresholds were excluded as independent variables. Meanwhile, the remained were considered as correlated variables and as the import to the regression model. This research, 34 characteristic variables (994, 1001, 1007, 1013, 1045, 1051, 1057, 1095, 1214, 1277, 1346, 1352, 1358, 1377, 1383, 1390, 1396, 1402, 1408, 1415, 1421, 1427, 1654, 1660, 1666, 1672, 1779, 1785, 1791, 1913, 1930, 2101, 2159, 2195 nm) were chosen from the primitive 229 wavelengths by UVE. The chosen variables were proportion of 14.8 % for the original bands, which desired the meeting choice potency for beneficial variables for protein content in potato flour noodles.

3.6. Comparison of the optimal modeling

On account of the full spectral data of protein content in potato flour noodles, PLSR and PCR models were established. Table 5 displays the results of the calibration set and prediction set for different PLSR and PCR models. In the case of full-spectrum PLSR and PCR model, the R^2

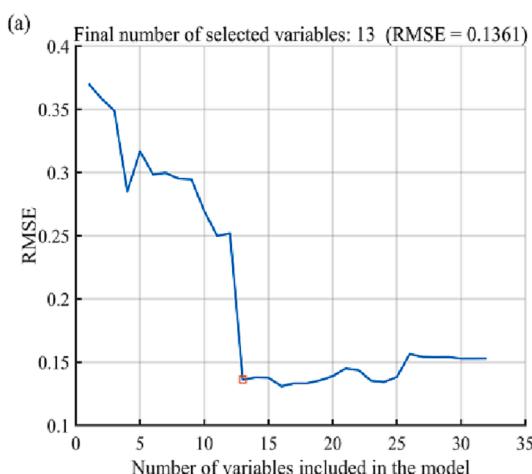


Fig. 7. Selected characteristic wavelength by SPA (a: RMSE growth pattern of variable, b: sequence number of selected characteristic wavelength).

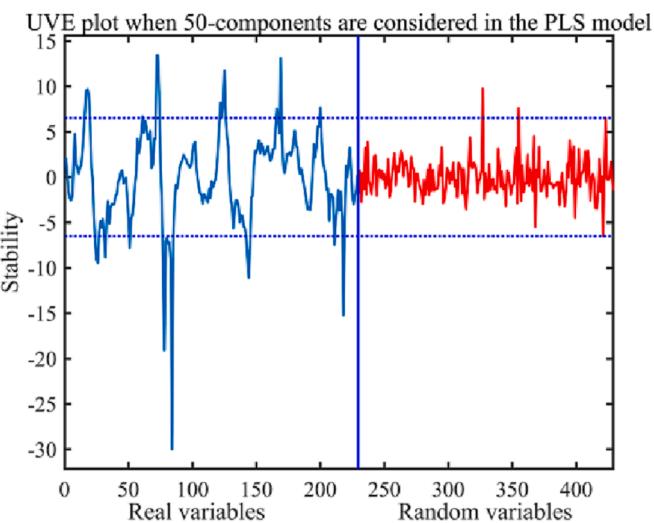


Fig. 8. Selection process of characteristic wavelength by UVE algorithm.

and R^2 were 0.9641, 0.9014 and 0.8506, 0.8005 respectively, RMSEC and RMSEP were 0.0724 %, 0.1201 % and 0.1478 %, 0.1888 % respectively. For the OSC-CARS-PLSR model, the obtained results ($R^2_C = 0.9606$, RMSEC = 0.0760 %, $R^2_P = 0.8925$, RMSEP = 0.1385 %) were outperformed by the OSC-SPA-PLSR, OSC-UVE-PLSR and FS-PLSR models. Meanwhile, for the Moving average-CARS-PCR models the obtained results ($R^2_C = 0.9331$, RMSEC = 0.0989 %, $R^2_P = 0.8388$, RMSEP = 0.1697 %) were better than those based on the other variable selection models. Therefore, using CARS algorithm to select the characteristic wavelengths was the optimal approach. Comparing the PLSR and PCR models, the former was slightly better than the latter. That suggests the information variables could be well correlated by the OSC-CARS-PLSR model, and that the model robustness also improved based on the wavelength selection method. Fig. 9 shows the model renderings of the protein calibration model and the prediction model, respectively.

Compared with other studies, Liu et al., [62] determined the protein content of wheat flour using Fourier transform infrared photoacoustic spectroscopy, and based on the whole wavenumber dataset, the RMSEP was 0.362 % and R^2_P was 0.84. However, the optimal 23 wavenumbers are selected using SPA, the RMSEP was 0.443 % and R^2_P was 0.75, and the accuracy of Fourier transform infrared photoacoustic spectroscopy needed to be improved. Chen et al., [63] determined foxtail millet protein content in the range of 950–1650 nm using NIR spectroscopy methods, the values of R^2_P , RMSEP and RPD were 0.94, 0.28 %, and 4.07 of the obtained optimum models. These findings were close to the results

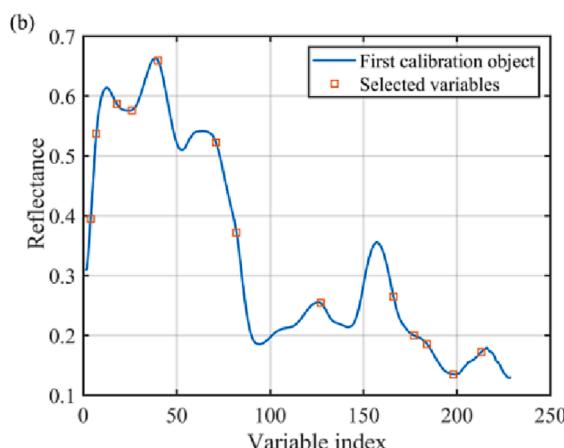
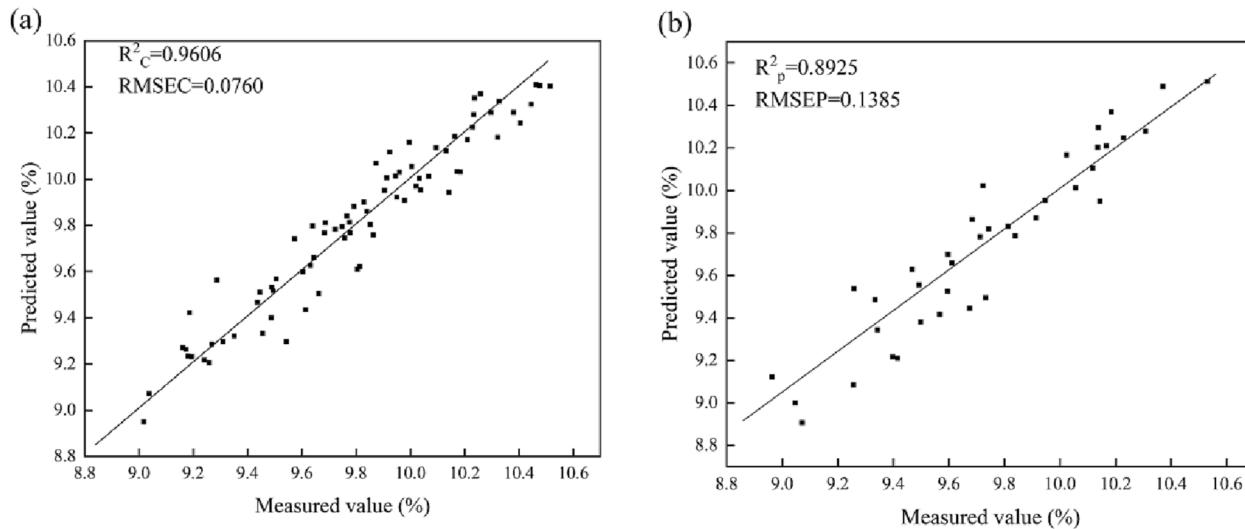


Table 5

Comparison of different characteristic wavelengths selection methods to predict the protein content.

Model	Pretreatment methods	Characteristic wavelength selection method	Wavelengths number	No. principal components	Calibration set		Prediction set	
					R^2_C	RMSEC/%	R^2_P	RMSEP/%
PLSR	OSC	FS	229	8	0.9641	0.0724	0.8506	0.1478
		CARS	25	13	0.9606	0.0760	0.8925	0.1385
		SPA	13	1	0.8795	0.1327	0.8338	0.1723
		UVE	34	1	0.9092	0.1152	0.8360	0.1712
PCR	Moving average	FS	229	16	0.9014	0.1201	0.8005	0.1888
		CARS	25	18	0.9331	0.0989	0.8388	0.1697
		SPA	13	13	0.8795	0.1327	0.7773	0.1995
		UVE	34	11	0.9017	0.1199	0.6904	0.2352

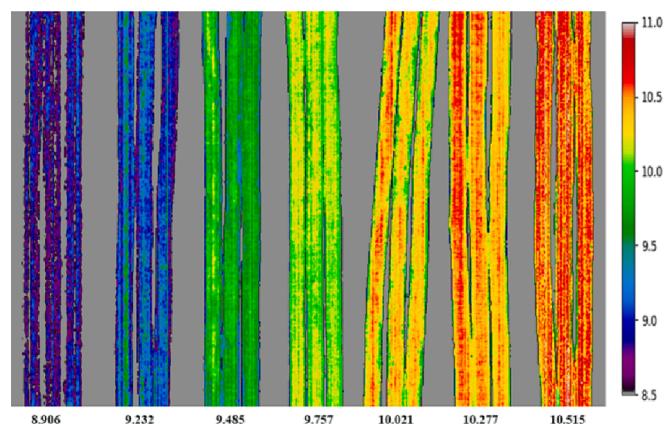
FS represented the full spectra.

**Fig. 9.** OSC-CARS-PLSR calibration model (a) and prediction model (b) for noodle protein content.

of the present study, but the spatial pattern distribution of the component content was unknown. Ma et al., [24] determined the protein content of rice (with husk) using HSI technology, which also showed the feasibility of HSI in the determination of rice (with husk). In this study, the spectral detection in the 938–2215 nm region showed that the main absorption wavebands of protein were concentrated at 2100–2200 nm, and 2190 nm was retained. In this study, 2195 nm was selected both as the reflectance peak and the characteristic wavelength using the CARS and UVE algorithms, indicating that 2195 nm is strongly associated with proteins. In conclusion, NIR-HSI was a suitable method to determine the protein content in potato flour noodles.

3.7. Visualization of the distribution of protein content

The outstanding advantage of HSI technology was that could create the spatial distribution map of samples chemical properties in a pixel-level model. Taking full used of the advantage, the visualization of protein content distribution was realized, which used a linear color bar to present different protein content values to reflect the protein content changes in potato flour noodles. Fig. 10 was the application for the best OSC-CARS-PLSR model to predict distribution of protein content. The image with the linear colour scale on the right intuitively showed the distribution of protein content in potato flour noodles. Variations of the ingredients were allocated with a linear colour scale. In the visual distribution map, the dark purple color of noodles indicated the lowest protein content which was 8.906 %, while the red color indicates the highest protein content which was 10.515 %. The color of the noodles changed from dark purple to blue, then to green, then to yellow, and finally to red, indicating that the protein content of the noodles is constantly increasing, ranging from 8.5 % to 11 %.

**Fig. 10.** Visualization map of protein content in potato flour noodles.

The protein content in the potato flour noodles was obviously differentiated from the color, and the distribution was uneven. It is concluded the change of protein content in potato flour noodles would be distinguished through color changes, and the visualization detection of protein content in potato flour noodles was realized. Therefore, the visualization images from NIR-HSI could be easily used for the detection of protein content in potato flour noodles. That was a simple and useful method used for a fast and accurate non-destructive visible system. As compared to single-

point detection technology, HSI technology could accomplish the visualization of protein content in any local area of potato flour noodles, intuitively reveal the potato flour noodles changes the protein contents, and contribute to the integrated assessment for the quality of potato flour noodles.

4. Conclusion

This study indicated that NIR (900–2500 nm) combined with data fusion analysis could accurately detect the protein content in potato flour noodles using the NIR-HSI technology. After using different pre-treatment algorithms to establish PLSR and PCR models, the spectral data discarded using OSC preprocessing method for the PLSR model were discovered to produce the optimal results ($R^2_C = 0.9641$, RMSECV = 0.1158 %). CARS, SPA and UVE algorithms were applied to extract the characteristic wavelengths. On the basis of these results, the CARS algorithm extracted 25 characteristic variables from 229 wavelengths, and the optimal model was OSC-CARS-PLSR, with the R^2_P and RMSEP value being 0.8925 and 0.1385 %, respectively. The spectral data were substituted into the established prediction model to predict the protein content separately and then combined with the established featured variable model. Finally, the visualization of the protein content in potato flour noodles was accomplished. Compared with other spectral or optical methods, HSI technology involves the coconstantaneous expression of spectral and spatial data along with non-destructive online detection, making the detection work more efficient and faster, and promoting the development of infrared technology in related fields. Therefore, the obtained model was successfully applied to the ingredient prediction of samples, avoiding the dreary preparation, and it was dependable and sufficient for feasible quality control procedures. This work exhibited that NIR-HSI technology could be used as an effective tool to quantify and evaluate the allocation of ingredients.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors appreciate the financial support provided by the National Natural Science Foundation of China (32171910), Shandong Natural Science Foundation (ZR2020MC085), Major Scientific and Technological Innovation Project of Shandong Province (2019JZZY010734).

References

- [1] M.A. Lambrecht, I. Rombouts, M.A. Nivelle, J.A. Delcour, The impact of protein characteristics on the protein network in and properties of fresh and cooked wheat-based noodles, *J. Cereal Sci.* 75 (2017) 234–242, <https://doi.org/10.1016/j.jcs.2017.04.014>.
- [2] Y. Cao, H. Zhang, Z. Yang, M. Zhang, P. Guo, H. Li, Influence of the fermentation time and potato pulp addition on the technological properties and volatile compounds of wheat dough and steamed bread, *LWT* 128 (2020) 109377, <https://doi.org/10.1016/j.lwt.2020.109377>.
- [3] R. Cui, F. Zhu, Effect of ultrasound on structural and physicochemical properties of sweetpotato and wheat flours, *Ultron. Sonochem.* 66 (2020) 105118, <https://doi.org/10.1016/j.ultronch.2020.105118>.
- [4] J. Valcarcel, K. Reilly, M. Gaffney, N. O'Brien, Total Carotenoids and L-Ascorbic Acid Content in 60 Varieties of Potato (*Solanum tuberosum* L.) Grown in Ireland, *Potato Res.* 58 (1) (2015) 29–41, <https://doi.org/10.1007/s11540-014-9270-4>.
- [5] D.F. Barbin, G. Elmasry, D.W. Sun, P. Allen, Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging, *Food Chem.* 138 (2–3) (2013) 1162–1171, <https://doi.org/10.1016/j.foodchem.2012.11.120>.
- [6] H. Jiang, L. Liu, Q. Chen, Rapid determination of acidity index of peanuts by near-infrared spectroscopy technology: Comparing the performance of different near-infrared spectral models, *Infrared Phys. Technol.* 125 (2022), 104308, <https://doi.org/10.1016/j.infrared.2022.104308>.
- [7] C. Yang, X. Ma, H. Guan, L. Li, B. Fan, A rapid recognition method of *Auricularia auricula* varieties based on near-infrared spectral characteristics, *Infrared Phys. Technol.* 125 (2022), 104239, <https://doi.org/10.1016/j.infrared.2022.104239>.
- [8] M. Zhou, L. Wang, H. Wu, Q. Li, M. Li, Z. Zhang, Y. Zhao, Z. Lu, Z. Zou, Machine learning modeling and prediction of peanut protein content based on spectral images and stoichiometry, *LWT* 169 (2022), 114015, <https://doi.org/10.1016/j.lwt.2022.114015>.
- [9] Z. Liu, R. Zhang, C. Yang, B. Hu, X. Luo, Y. Li, C. Dong, Research on moisture content detection method during green tea processing based on machine vision and near-infrared spectroscopy technology, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 271 (2022), 120921, <https://doi.org/10.1016/j.saa.2022.120921>.
- [10] W. Huang, D. Fan, W. Li, Y. Meng, T.-C.-Y. Liu, Rapid evaluation of milk acidity and identification of milk adulteration by Raman spectroscopy combined with chemometrics analysis, *Vib. Spectrosc.* 123 (2022), 103440, <https://doi.org/10.1016/j.vibspec.2022.103440>.
- [11] S. Nakajima, S. Kuroki, A. Ikehata, Selective detection of starch in banana fruit with Raman spectroscopy, *Food Chem.* 401 (2023), 134166, <https://doi.org/10.1016/j.foodchem.2022.134166>.
- [12] W. Li, F. Tan, J. Cui, B. Ma, Fast identification of soybean varieties using Raman spectroscopy, *Vibr. Spectrosc.* 123 (2022), <https://doi.org/10.1016/j.vibspec.2022.103447>.
- [13] F. Tian, F. Tan, H. Li, An rapid nondestructive testing method for distinguishing rice producing areas based on Raman spectroscopy and support vector machine, *Vib. Spectrosc.* 107 (2020), 103017, <https://doi.org/10.1016/j.vibspec.2019.103017>.
- [14] H. Sun, N. Liu, L. Wu, T. Zheng, M.Z. Li, J.Z. Wu, Visualization of Water Content Distribution in Potato Leaves Based on Hyperspectral Image, *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis* 39 (3) (2019) 910–916, [https://doi.org/10.3964/j.issn.1000-0593\(2019\)03-0910-07](https://doi.org/10.3964/j.issn.1000-0593(2019)03-0910-07).
- [15] D. Kimuli, W. Wang, W. Wang, H. Jiang, X. Zhao, X. Chu, Application of SWIR hyperspectral imaging and chemometrics for identification of aflatoxin B1 contaminated maize kernels, *Infrared Phys. Technol.* 89 (2018) 351–362, <https://doi.org/10.1016/j.infrared.2018.01.026>.
- [16] K. Kobayashi, Y. Matsui, Y. Maebuchi, T. Toyota, S. Nakauchi, Near infrared spectroscopy and hyperspectral imaging for prediction and visualisation of fat and fatty acid content in intact raw beef cuts, *J. Near Infrared Spectrosc.* 18 (2010) 301–315, <https://doi.org/10.1255/jnirs.896>.
- [17] Q. Yang, D.W. Sun, W. Cheng, Development of simplified models for nondestructive hyperspectral imaging monitoring of TVB-N contents in cured meat during drying process, *J. Food Eng.* 192 (JAN) (2017) 53–60, <https://doi.org/10.1016/j.jfooodeng.2016.07.015>.
- [18] C. Erkinbaev, K. Derkens, J. Paliwal, Single kernel wheat hardness estimation using near infrared hyperspectral imaging, *Infrared Phys. Technol.* 98 (2019) 250–255, <https://doi.org/10.1016/j.infrared.2019.03.033>.
- [19] W. Lan, B. Jaillais, C. Renard, A. Leca, S. Chen, C.L. Bourville, et al., A method using near infrared hyperspectral imaging to highlight the internal quality of apple fruit slices, *Postharvest Biol. Technol.* 175 (2021) 11497, <https://doi.org/10.1016/j.postharvbio.2021.111497>.
- [20] N. Caporaso, M.B. Whitworth, I.D. Fisk, Protein content prediction in single wheat kernels using hyperspectral imaging, *Food Chem.* 240 (2018) 32–42, <https://doi.org/10.1016/j.foodchem.2017.07.048>.
- [21] Z. Chu, E. Wwacd, Z. Lei, E. Hcad, E. Xyacd, H. Yong, Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging, *Food Chem.* 319 (2020) 12636, <https://doi.org/10.1016/j.foodchem.2020.126536>.
- [22] N. Malhotra, S. Sharma, P. Sahni, B. Singh, S.P. Sharma, Nutritional composition, techno-functionality, in-vitro starch digestibility, structural characteristics and storage stability of sweet potato flour and mash supplemented specialty pasta, *LWT* 168 (2022), 113886, <https://doi.org/10.1016/j.lwt.2022.113886>.
- [23] NY/T 3521-2019, Technical specification for potato noodles processing[S].
- [24] C. Ma, Z. Ren, Z. Zhang, J. Du, X. Yin, Development of Simplified Models for Nondestructive Testing of Rice (with Husk) Protein Content Using Hyperspectral Imaging Technology, *Vib. Spectrosc.* 114 (1) (2021), 103230, <https://doi.org/10.1016/j.vibspec.2021.103230>.
- [25] J. Ma, J.H. Cheng, D.W. Sun, D. Liu, Mapping changes in sarcoplasmatic and myofibrillar proteins in boiled pork using hyperspectral imaging with spectral processing methods, *LWT* 110 (2019) 338–345, <https://doi.org/10.1016/j.lwt.2019.04.095>.
- [26] X. Zhang, F. Liu, Y. He, X. Li, Application of Hyperspectral Imaging and Chemometric Calibrations for Variety Discrimination of Maize Seeds, *Sensors* (Basel, Switzerland) 12 (12) (2012) 17234–17246, <https://doi.org/10.3390/s121217234>.
- [27] F. Tao, Y.K. Peng, Y.Y. Li, Feature extraction method of hyperspectral scattering images for prediction of total viable count in pork meat, *Int. J. Agric. Biol. Eng.* 8 (4) (2015) 95–105, <https://doi.org/10.3965/j.ijabce.20150804.1559>.
- [28] A.A. Gowen, M. Taghizadeh, C.P. O'Donnell, Identification of mushrooms subjected to freeze damage using hyperspectral imaging, *J. Food Eng.* 93 (1) (2009) 7–12, <https://doi.org/10.1016/j.jfoodeng.2008.12.021>.

- [29] G. Elmasry, D.W. Sun, P. Allen, Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging, *Food Res. Int.* 44 (9) (2011) 2624–2633, <https://doi.org/10.1016/j.foodres.2011.05.001>.
- [30] G. Ma, X. Yue, An improved whale optimization algorithm based on multilevel threshold image segmentation using the Otsu method, *Eng. Appl. Artif. Intel.* 113 (2022), 104960, <https://doi.org/10.1016/j.engappai.2022.104960>.
- [31] G. Wan, G. Liu, J. He, R. Luo, L. Cheng, C. Ma, Feature wavelength selection and model development for rapid determination of myoglobin content in nitrite-cured mutton using hyperspectral imaging, *J. Food Eng.* 287 (2020), 110090, <https://doi.org/10.1016/j.jfoodeng.2020.110090>.
- [32] R.K.H. Galvão, M.C.U. Araújo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, A method for calibration and validation subset partitioning, *Talanta* 67 (4) (2005) 736–740, <https://doi.org/10.1016/j.talanta.2005.03.025>.
- [33] J.L. Xu, C. Riccioli, D.W. Sun, Development of an alternative technique for rapid and accurate determination of fish caloric density based on hyperspectral imaging, *J. Food Eng.* 190 (2016) 185–194, <https://doi.org/10.1016/j.jfoodeng.2016.06.007>.
- [34] L. Tao, W. Li, Y. Jin, Y. Yang, Y. Wu, X. Liu, Application of near-infrared spectroscopy combined with chemometrics for online monitoring of Moluodan extraction, *J. Chemom.* 32 (5) (2018) e2979, <https://doi.org/10.1002/cem.2979>.
- [35] H.J. He, D.W. Sun, Selection of Informative Spectral Wavelength for Evaluating and Visualising Enterobacteriaceae Contamination of Salmon Flesh, *Food Anal. Methods* 8 (10) (2015) 2427–2436, <https://doi.org/10.1007/s12161-015-0122-x>.
- [36] D. Wu, H. Shi, S. Wang, Y. He, Y. Bao, K. Liu, Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system, *Anal. Chim. Acta* 726 (2012) 57–66, <https://doi.org/10.1016/j.aca.2012.03.038>.
- [37] D.S. Cao, Y.Z. Liang, Q.S. Xu, H.D. Li, X. Chen, A new strategy of outlier detection for QSAR/QSPR, *J. Comput. Chem.* 31 (3) (2010) 592–602, <https://doi.org/10.1002/jcc.21351>.
- [38] D. Wu, D.W. Sun, Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part II: Applications, *Innov. Food Sci. Emerg. Technol.* 19 (2013) 15–28, <https://doi.org/10.1016/j.ifset.2013.04.016>.
- [39] S. Wajizah, A.A. Munawar, Fast and simultaneous prediction of animal feed nutritive values using near infrared reflectance spectroscopy, *IOP Conf. Ser. Earth Environm. Sci.* 122 (1) (2018), 012112, <https://doi.org/10.1088/1755-1315/122/1/012112>.
- [40] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, *Chemometr. Intell. Lab. Syst.* 44 (1) (1998) 175–185, [https://doi.org/10.1016/S0169-7439\(98\)00109-9](https://doi.org/10.1016/S0169-7439(98)00109-9).
- [41] S.A. Akrami, A. El-Shafie, M. Naseri, C.A.G. Santos, Rainfall data analyzing using moving average (MA) model and wavelet multi-resolution intelligent model for noise evaluation to improve the forecasting accuracy, *Neural Comput. Applic.* 25 (7–8) (2014) 1853–1861, https://doi.org/10.1007/s00521-014-1675-0_25.
- [42] H. Chen, T. Pan, J. Chen, Q. Lu, Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods, *Chemometr. Intell. Lab. Syst.* 107 (1) (2011) 139–146, <https://doi.org/10.1016/j.chemolab.2011.02.008>.
- [43] H. Jin, L. Li, J. Cheng, Rapid and Non-destructive Determination of Moisture Content of Peanut Kernels Using Hyperspectral Imaging Technique, *Food Anal. Methods* 8 (10) (2015) 2524–2532, <https://doi.org/10.1007/s12161-015-0147-1>.
- [44] J. Sun, G. Wang, H. Zhang, L. Xia, W. Zhao, Y. Guo, et al., Detection of fat content in peanut kernels based on chemometrics and hyperspectral imaging technology, *Infrared Phys. Technol.* 105 (2020), 103226, <https://doi.org/10.1016/j.infrared.2020.103226>.
- [45] H. Li, Y. Liang, Q. Xu, D. Cao, Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, *Anal. Chim. Acta* 648 (1) (2009) 77–84, <https://doi.org/10.1016/j.jaca.2009.06.046>.
- [46] J.H. Cheng, J.H. Qu, D.W. Sun, X.A. Zeng, Visible/near-infrared hyperspectral imaging prediction of textural firmness of grass carp (*Ctenopharyngodon idella*) as affected by frozen storage, *Food Res. Int.* 56 (2014) 190–198, <https://doi.org/10.1016/j.foodres.2013.12.009>.
- [47] T. Iwata, S. Yoshioka, S. Nakamura, Y. Mizutani, T. Yasui, Prediction of the Thickness of a Thin Paint Film by Applying a Modified Partial-Least-Squares-1
- [48] Method to Data Obtained in Terahertz Reflectometry, *J. Infrared, Millimeter, Terahertz Waves* 34 (10) (2013) 646–659, <https://doi.org/10.1007/s10762-013-0012-2>.
- [49] X. Peng, T. Shi, A. Song, Y. Chen, W. Gao, Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods, *Remote Sens. (Basel)* 6 (2014) 2699–2717, <https://doi.org/10.3390/rs6042699>.
- [50] S. Mahesh, D.S. Jayas, J. Paliwal, N.D.G. White, Comparison of Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) Methods for Protein and Hardness Predictions using the Near-Infrared (NIR) Hyperspectral Images of Bulk Samples of Canadian Wheat, *Food Bioproc. Tech.* 8 (1) (2015) 31–40, <https://doi.org/10.1007/s11947-014-1381-z>.
- [51] M. Taghizadeh, A. Gowen, C.P. O'Donnell, Prediction of white button mushroom (*Agaricus bisporus*) moisture content using hyperspectral imaging, *Sens. Instrumen. Food Qual.* 3 (4) (2009) 219–226, <https://doi.org/10.1007/s11694-009-9088-y>.
- [52] A. Xie, D.W. Sun, Z. Zhu, H. Pu, Nondestructive Measurements of Freezing Parameters of Frozen Porcine Meat by NIR Hyperspectral Imaging, *Food Bioproc. Tech.* 9 (9) (2016) 1444–1454, <https://doi.org/10.1007/s11947-016-1766-2>.
- [53] W.-H. Su, D.-W. Sun, Multispectral Imaging for Plant Food Quality Analysis and Visualization, *Compr. Rev. Food Sci. Food Saf.* 17 (2018) 220–239, <https://doi.org/10.1111/1541-4337.12317>.
- [54] V. Wiedemair, D. Mair, C. Held, C.W. Huck, Investigations into the use of handheld near-infrared spectrometer and novel semi-automated data analysis for the determination of protein content in different cultivars of *Panicum miliaceum* L, *Talanta* 205 (2019), 120115, <https://doi.org/10.1016/j.talanta.2019.120115>.
- [55] E. Lancelot, D. Bertrand, M. Hanafi, B. Jaillais, Near-infrared hyperspectral imaging for following imbibition of single wheat kernel sections, *Vib. Spectrosc* 92 (2017) 46–53, <https://doi.org/10.1016/j.vibspec.2017.05.001>.
- [56] L.C.R. Silva, G.S. Folli, L.P. Santos, I.H.A.S. Barros, B.G. Oliveira, F.T. Borghi, et al., Quantification of beef, pork, and chicken in ground meat using a portable NIR spectrometer, *Vib. Spectrosc* 111 (2020), 103158, <https://doi.org/10.1016/j.vibspec.2020.103158>.
- [57] W. Guo, Y. Du, Y. Zhou, S. Yang, J. Lu, H. Zhao, Y. Wang, L. Teng, At-line monitoring of key parameters of nisin fermentation by near infrared spectroscopy, chemometric modeling and model improvement, *World J. Microbiol. Biotechnol.* 28 (2012) 993–1002, <https://doi.org/10.1007/s11274-011-0897-x>.
- [58] R.K.G. Harrop, A method for calibration and validation subset partitioning, *Talanta* 4 (67) (2005) 736–740, <https://doi.org/10.1016/j.talanta.2005.03.025>.
- [59] Z. Wang, S. Fan, J. Wu, C. Zhang, F. Xu, X. Yang, et al., Application of long-wave near infrared hyperspectral imaging for determination of moisture content of single maize seed, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 254 (2021), 119666, <https://doi.org/10.1016/j.saa.2021.119666>.
- [60] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, *Chemom. Intel. Lab. Syst.* 57 (2) (2001) 65–73, [https://doi.org/10.1016/S0169-7439\(01\)00119-8](https://doi.org/10.1016/S0169-7439(01)00119-8).
- [61] K.-Q. Yu, Y.R. Zhao, Z.Y. Liu, X.L. Li, F. Liu, Y. He, Application of Visible and Near-Infrared Hyperspectral Imaging for Detection of Defective Features in Loquat, *Food Bioproc. Tech.* 7 (11) (2014) 3077–3087, <https://doi.org/10.1007/s11947-014-1357-z>.
- [62] Z. Wang, J. Chen, Y. Fan, Y. Cheng, X. Wu, J. Zhang, et al., Evaluating photosynthetic pigment contents of maize using UVE-PLS based on continuous wavelet transform, *Comput. Electron. Agric.* 169 (2020), 105160, <https://doi.org/10.1016/j.compag.2019.105160>.
- [63] Q. Liu, W. Zhang, B. Zhang, C. Du, N. Wei, D. Liang, K. Sun, K. Tu, J. Peng, L. Pan, Determination of total protein and wet gluten in wheat flour by Fourier transform infrared photoacoustic spectroscopy with multivariate analysis, *J. Food Compos. Anal.* 106 (2022), 104349, <https://doi.org/10.1016/j.jfca.2021.104349>.
- [64] J. Chen, X. Ren, Q. Zhang, X. Diao, Q. Shen, Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy, *J. Cereal Sci.* 58 (2) (2013) 241–247, <https://doi.org/10.1016/j.jcs.2013.07.002>.