

Rapid determination of protein, starch and moisture content in wheat flour by near-infrared hyperspectral imaging

Jing Zhang^a, Zhen Guo^a, Zhishang Ren^b, Sihua Wang^a, Minghui Yue^a, Shanshan Zhang^a, Xiang Yin^a, Kuijie Gong^c, Chengye Ma^{a,*}

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

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

^c Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan 250100, China



ARTICLE INFO

Keywords:

Hyperspectral imaging
Wheat flour
Near-infrared
Chemical composition
Wavelength selection

ABSTRACT

In this study, near-infrared (NIR) hyperspectral imaging (HSI) (969–2174 nm) technology was investigated for the determination of protein, starch and moisture content in 77 different varieties of wheat flour. The NIR-HSI system combined with effective wavelengths (EWs) algorithms was applied to obtain spectral information with wheat flour. Then, four regression models based on the original spectral information and the EWs were established to determine the relationship between the spectrum and detection index. Five EWs algorithms were applied to select EWs to optimise the models. The coefficient of determination and root mean square error for prediction of the obtained optimum models were 0.9859 and 1.1580 g/100 g for protein, 0.9243 and 0.2068 g/100 g for starch, and 0.8646 and 2.1669 g/100 g for moisture, respectively. The visualisation of the protein, starch and moisture content was achieved using the optimal models. The result indicated that the NIR-HSI technology is a promising approach for predicting the protein, starch and moisture content of wheat flour effectively and accurately. It also provided a reference for non-destructive and rapid prediction of other chemical compositions in foxtail millet.

1. Introduction

As one of the most cultivated and staple commodity food crops worldwide, wheat is an important cash crop used in the human diet and animal feed (Li et al., 2021). Wheat flour is an important ingredient in multiple food products; amongst its quality parameters, protein, starch, moisture, gluten, ash and fat content are considered important indices for measuring its quality (Mæhre et al., 2018; Tian et al., 2021). The processing characteristics of wheat flour could be evaluated and appropriate wheat flour could be selected for different food formulas by understanding these indices to be detected. The determination of these parameters is closely related to the processing characteristics and nutritional quality of wheat flour. As mentioned above, protein, starch and moisture content play a decisive role in the final cost and the processing suitability of wheat flour, as well as the stability during storage (Caporaso et al., 2018). Traditional chemical detection methods are accurate and reliable. However, these methods are disruptive and they consume manpower, chemicals and time, especially when performing

large numbers of samples. Thus, a highly efficient, low-cost, chemical-free and accurate quality evaluation system for food-industry detection methods must be developed.

In recent years, spectral analysis technology has been used for testing seed quality, such as the determination of protein, wet gluten, moisture, ash, and sedimentation in wheat flour (Chen et al., 2021), protein content, moisture content, and flour color *b** values in grain, flour and dough (Dowell et al., 2006), protein content and farinograph stability in wheat flour (Barbon Junior et al., 2020) and protein content in wheat (Ye et al., 2018). Spectral analysis technology has the advantages of rapid and non-destructive and it was established to make up for the deficiency of traditional detection methods in destroying samples. Fatty acid value in wheat flour was detected by a self-built portable near-infrared spectrometer (899.22–1724 nm) during the storage, the extreme learning machine model achieved a good effect ($R^2 > 0.96$) (Jiang et al., 2020a). Three handheld NIR spectrometers were used to detect the flour quality parameters including protein, wet gluten, moisture, ash, and sedimentation, the wavelength ranges of those

* Corresponding author.

E-mail addresses: zj2021zj@126.com (J. Zhang), 806559062@qq.com (Z. Guo), 1525169572@qq.com (Z. Ren), 1932611183@qq.com (S. Wang), 1778322325@qq.com (M. Yue), 1974742249@qq.com (S. Zhang), xiangyinxz@163.com (X. Yin), 1820787057@qq.com (K. Gong), mcycn2002@163.com (C. Ma).

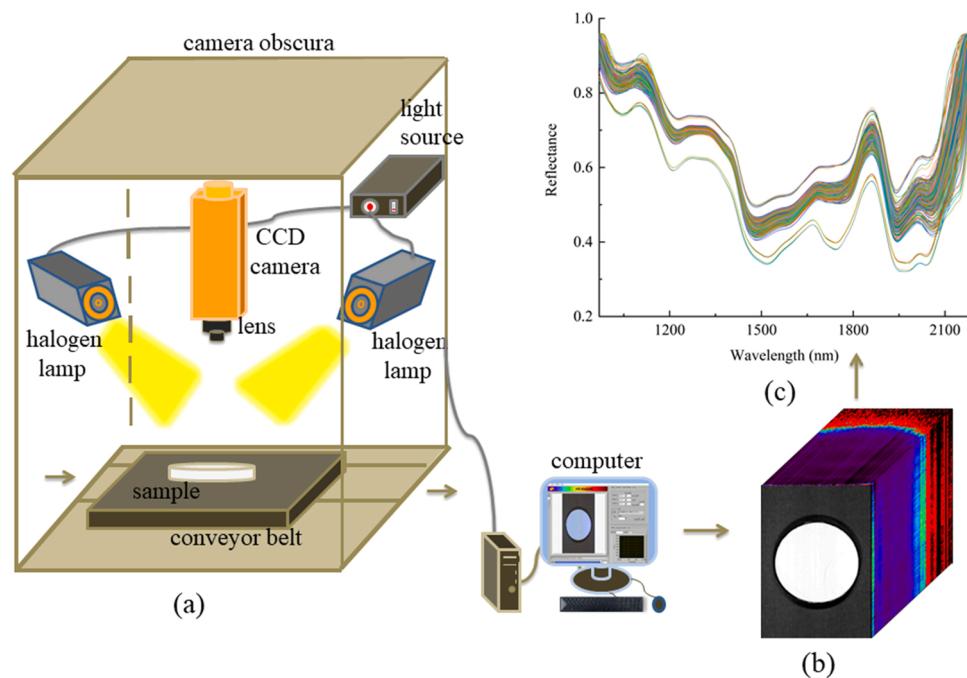


Fig. 1. (a) Schematic diagram of the hyperspectral imaging system, (b) hyperspectral image 3D data matrix, (c) reflectance curves of spectra.

spectrometers were 1550–1950 nm, 908–1676 nm and 1298–2606 nm (Chen et al., 2021).

However, NIR spectroscopy is a spot area dimension technique; in many cases, the spectral data acquired were inadequate or unrepresentative of the whole samples on the account of the uneven surface of the sample (Mahesh et al., 2015). Meanwhile, it is necessary to keep a relatively fixed distance between the camera lens and the sample to focus the light source, which is limited to expand the detection range of the sample surface using NIR spectroscopy. Therefore, spectral analysis technology is not suitable for the detection of the average information of large quantities of wheat flour samples. For the past few years, hyperspectral imaging (HSI) technology has appeared as a promising burgeoning technology for non-invasive pollution-free quality assessment of food products (Srivastava and Mishra, 2022). HSI technology combines the advantages of spectral technology and image technology to obtain physical and chemical information about wheat flour.

Depending on the chemical properties of food, HSI could operate over a wide band range (780–2500 nm), and the specific wavelength range of 900–2300 nm is referred to as near-infrared HSI (NIR-HSI). NIR spectroscopy operates in accordance with the interaction of light radiation with the sample, especially the absorption of molecular overtone combination vibrations to establish the relationship between the spectrum and the detected index (Huan et al., 2021). NIR-HSI obtains spatial and spectral information from the captured images and affords a characteristic chemical fingerprint for each pixel associated with that information. NIR-HSI technology overcomes the boundedness of spectroscopy by obtaining a 3D data matrix containing thousands of consecutive images, narrowband spectra and 2D images of spectral information. Accordingly, this study aimed to calibrate the regression model to enhance the performance of predicting protein, starch and moisture content in wheat flour. NIR-HSI technology has the advantage of collecting abundant sample data, and it showed unique superiority for qualitative and quantitative analyses in fruits and vegetables (Ji et al., 2019; Xia et al., 2019), agricultural products (Long et al., 2022), dairy products (Iymen et al., 2020), meat (Cheng et al., 2018) and other fields.

Accordingly, the primary goals of this research are as follows: (1) determine protein, starch and moisture content in wheat flour by using traditional chemical methods and obtain NIR-HSI images information; (2) establish and analyse four calibration models on the basis of raw

spectra data as input, namely, partial least square regression (PLSR), principal component regression (PCR), support vector machine regression (SVMR) and multiple linear regression (MLR); (3) select the effective wavelengths (EWs) by iteratively retains informative variables (IRIV), variable combination population analysis (VCPA), interval variable iterative space shrinkage approach (IVISSA), model adaptive space shrinkage (MASS) and interval random frog (IRF) to verify the stability and rapid quantification for multispectral regression model and (4) visualise the chemical composition distribution of all pixels in wheat flour by applying the optimal model combined with image processing algorithm.

2. Materials and methods

2.1. Wheat flour samples

A total of 77 wheat flour samples were applied in this investigation. Amongst them, 12 varieties were purchased from the location supermarkets and 65 varieties were obtained from the Shandong Grain Group Co. Ltd (Table S1). In this study, wheat flour can be divided into low-gluten flour and high-gluten flour according to the gluten content.

The same variety of wheat flour was divided into six parallel samples. Six samples weighing 300 g were taken from each variety and placed in six clean petri dishes respectively. After obtaining the hyperspectral images, each parallel sample was divided into three equal parts by weight to measure the content of protein, starch and moisture respectively. Each index was measured three times and the average value was calculated. Therefore, a total of 462 wheat flour samples were obtained and the hyperspectral images were collected. The image libraries of 462 wheat flour images were established for the protein, starch and moisture content regression models.

2.2. Determination of chemical composition

The protein content was determined by the Kjeldahl method according to the AOAC 960.52 (Ma et al., 2021). The total wheat flour nitrogen compound is multiplied by the conventional factor 5.7 to obtain the protein content.

The starch content was determined by the amyloglucosidase/

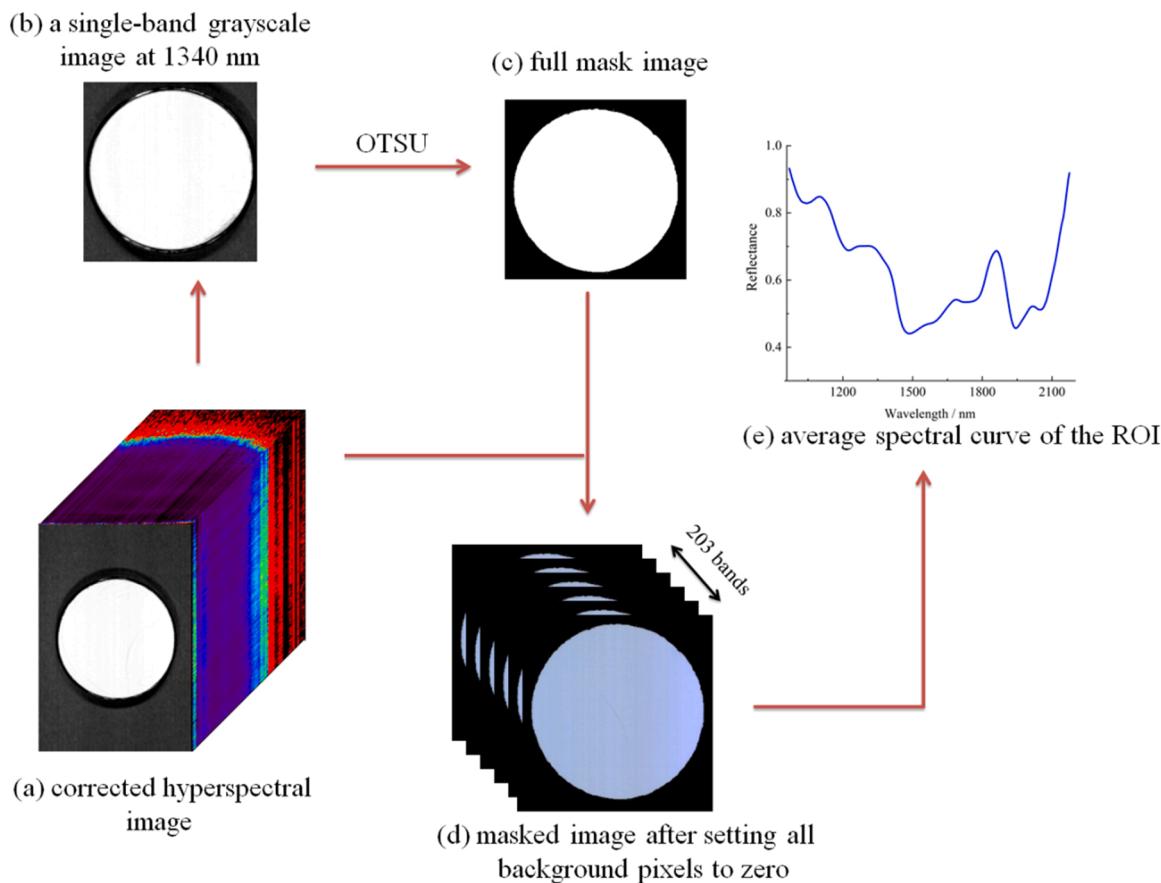


Fig. 2. Main steps of image segmentation and hyperspectral data extraction.

α -amylase method (Megazyme International Ireland Ltd., Wicklow, Ireland) according to the AOAC 996.11 (Zhang et al., 2019). Briefly, samples were incubated with thermostable α -amylase in sodium acetate buffer containing calcium chloride in a boiling water bath (at ~ 100 °C) for 15 min with periodic mixing to gelatinize and hydrolyze the starch to maltodextrins. After cooling to 50 °C over 5 min, amyloglucosidases was added which catalyses complete hydrolysis of the maltodextrins to glucose within 30 min. The mixture was clarified, and appropriately diluted, and glucose was determined using a glucose oxidase/peroxidase reagent that is both sensitive and specific for glucose. The moisture content of the wheat flour was measured was determined by oven drying method according to the AOAC 930.15 (Panda et al., 2022).

2.3. Acquisition and calibration of hyperspectral imaging

In this study, all of the wheat flour was scanned by a push-broom NIR-HSI system (Isuzu Optics Corp., Taiwan, China) (reflectance-mode) at the range of 969–2174 nm with 6.2 ± 0.2 nm intervals and a total of 203 bands were collected. Fig. 1(a) exhibited that the NIR-HSI system consists of a line-scan spectrograph (Specim V10E, Spectral Imaging Ltd., Oulu, Finland); a high-performance 2560 × 2160 CCD camera (Zyla 5.5sCMOS, Oxford Instruments, Oxford, England); a camera lens (OLE23, Spectral Imaging Ltd., Oulu, Finland); an illumination unit, including two halogen lamps (3900-ER, Illumination Technologies Inc., New York, USA); a mobile platform controlled by a stepper motor (IRCP0076-1COMB, Isuzu Optics Corp., Taiwan, China); data acquisition software (Spectral Image software, Isuzu Optics Corp., Taiwan, China) and an image processor (DELL, USA) with 4 G memory and a 1 TB hard drive. For elimination of the image size and spatial resolution distortion, the moving speed of the displacement platform was set as 14.97 mm/s, the object distance between the imaging

spectrometer and conveyor belt was set to about 400 mm and the exposure time was set to 3.1 ms. Scanned images of the typical hypercube of dimension with the size of 300 pixels × 300 pixels × 462 were obtained from 462 samples.

Before image acquisition was conducted, the instrument should be preheated for 1 h to avoid baseline drift caused by the spectrometer. The change of room temperature did not exceed 1 °C controlled by air conditioning in the process of collecting hyperspectral images. And the instrument was integrated in a dark box which was not affected by ambient light. The original hyperspectral image must be corrected with a black reference and white reference image by using Eq. (1) to improve the accuracy of spectral information and reduce the effects of light sources and dark noise (Zhang et al., 2017).

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

where R_T and I_{raw} represent the corrected and original hyperspectral images, respectively, I_{white} and I_{dark} represent the white reference image and dark reference image, respectively.

2.4. Image segmentation and spectral data extraction

During the extraction of spectral information from wheat flour samples, the mask was required to distinguish wheat flour from the image background. Due to the obvious spectral contrast between wheat flour and the background, a single-band grayscale image at 1340 nm was applied to establish mask templates for NIR hyperspectral images. The OTSU method was applied for threshold segmentation of the acquired images. A binary mask was then built for each sample, assigning values 0 and 1 to the pixels corresponding to the background and wheat flour, respectively. The average spectrum of each image was obtained by

Table 1

Statistics reference measurement results of protein, starch and moisture content in wheat flour of the calibration set and prediction set.

Subsets	No. of samples	Min	Max	Mean	SD	SEM	CV
Protein content (g/100 g)							
Total	462	8.5817	16.7890	10.3497	1.7267	0.1994	0.1668
Calibration set	347	8.5817	16.7890	10.2515	1.6649	0.0902	0.1624
Prediction set	115	8.7704	16.6065	10.6570	1.8453	0.1768	0.1732
Starch content (g/100 g)							
Total	462	51.5749	70.4446	65.4236	4.3961	0.2072	0.0672
Calibration set	347	51.5749	70.4446	65.6572	4.2354	0.2294	0.0645
Prediction set	115	52.4238	69.8398	64.6929	4.8125	0.4610	0.0744
Moisture content (g/100 g)							
Total	462	9.1652	13.1145	12.2144	0.6437	0.0424	0.0527
Calibration set	347	9.1652	13.1145	12.3144	0.5087	0.0394	0.0413
Prediction set	115	9.4093	13.0189	12.1006	0.6152	0.0815	0.0508

SD: standard deviation; SEM: stand error of mean; CV: Co-efficient of variation

averaging the index value of each wheat flour pixel and 462 spectral data points were produced. The spectral information of whole wheat flour was concatenated to acquire the final spectral matrix used to establish the model. In this matrix, the rows and columns represented the number of wheat flour images and the wavelengths, respectively (Wan et al., 2020). In consequence, a 462×203 dimensional spectral matrix was obtained for protein, starch and moisture content prediction. The specific operation processes are reported in Fig. 2.

2.5. EWs selection

EWs selection is an essential segment of spectral data analysis in this study, 203 high-dimensionality and multi-collinearity wavelengths bands were obtained from the hyperspectral images. For the purpose of improving the calculation velocity and prediction precision, the frequently used method was dimensionality reduction, which selected certain sensitive or feature wavelengths from the whole spectral matrix to condense the hyperspectral information (Dou et al., 2021). The useful information contained in the EWs could accurately predict the protein, starch and moisture content of wheat flour samples and improve the calculation speed and the robustness of the model. In the present study, several methods, such as IRIV, VCPA, IVISSA, MASS and IRF algorithms, were used to select EWs (An et al., 2022; Ren et al., 2020; Zhang et al., 2022). The effects of different algorithms on the models performance were summarised and analysed. R software (version 3.6.3, Lucent Technologies, Murray Hill, NJ, USA) was used to select the EWs and establish the model.

2.6. Model construction and evaluation

The optimization calibration model was obtained on the basis of the four algorithms for multivariate calculation, such as PLSR, PCR, SVMR and MLR (Feng et al., 2019; Li et al., 2019).

Evaluating the model is one of the paramount steps in multivariate data analysis. The coefficient of determination (R^2) for calibration (R_C^2), cross-validation (R_{CV}^2) and prediction (R_P^2), the root mean square errors (RMSE) of calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP) and residual prediction deviation (RPD) of prediction were applied to assess the performance of the models. Generally, the high-performance model ought to have high RPD and R^2 (near to 1) and low RMSE (close to 0), with little variation amongst the models (Sun et al., 2019).

2.7. Image visualisation and prediction map

The chemical composition value of each pixel must be obtained to assess the protein, starch and moisture content difference in different varieties of wheat flour or even the same wheat flour (Yu et al., 2021). HSI technology could predict the distribution map of chemical composition change on the basis of the spatial distribution of each pixel. The

scalar product between the spectrum of each image pixel and the regression coefficient of the preferred performance model was calculated in accordance with the acquired images with different EWs. The optimal model was applied to the prediction of physical and chemical parameters of each pixel. Therefore, pseudo colour and synthetic colour images could be used to map and visualise the spatial distribution of each physical and chemical index of wheat flour (Jiang et al., 2020b).

3. Results and discussions

3.1. Reflectance spectra

The original spectra acquired by the HSI system were in the band of 969–2174 nm. The original spectra of different wheat flour varieties had similar variation trends but also certain discrepancies [Fig. 1(c)]. These discrepancies may be due to differences in the internal chemical composition and surface information of wheat flour varieties. The prominent feature peaks and valleys of the wavelengths range 969–1310, 1470–1860, 1935–2025 and 2040–2170 nm were related to the first and second overtone stretching and combined band and bending vibrations of C-H, O-H, N-H from the protein, starch and moisture content in wheat flour (Silva et al., 2020; Wang et al., 2018; Wiedemair et al., 2019). Therefore, using NIR-HSI technology to predict the protein, starch and moisture content in wheat flour was feasible.

3.2. Samples set division

The whole wheat flour samples were firstly divided into the calibration set and prediction set as this was closely related to the accuracy of the obtained model, followed by multivariate data analysis. The calibration set samples were applied to optimize the EWs and construct the quantitative model. The prediction set samples were used to prove the majorization result of the EWs and the repeatability capacity of the established quantitative models (Achata et al., 2021). Kennard-Stone (KS) algorithm was used to select the sample set, with an objective to uniformly cover the multi-dimensional space by maximizing Euclidean distance amongst the system responses. KS method was performed to divide the whole samples into the calibration set and prediction set at a ratio of 3:1. Therefore, 347 samples were selected as the calibration set and the remaining 115 samples were used as the prediction set. The distributions of protein, starch and moisture values in wheat flour samples from the calibration set and prediction set are shown in Table 1. The calibration set samples contained the variation range of the prediction set samples. These data indicated that the results of the samples set partitioning method were reasonable and the selected samples for constructing the model were highly representative.

3.3. Modelling analysis based on full wavelengths

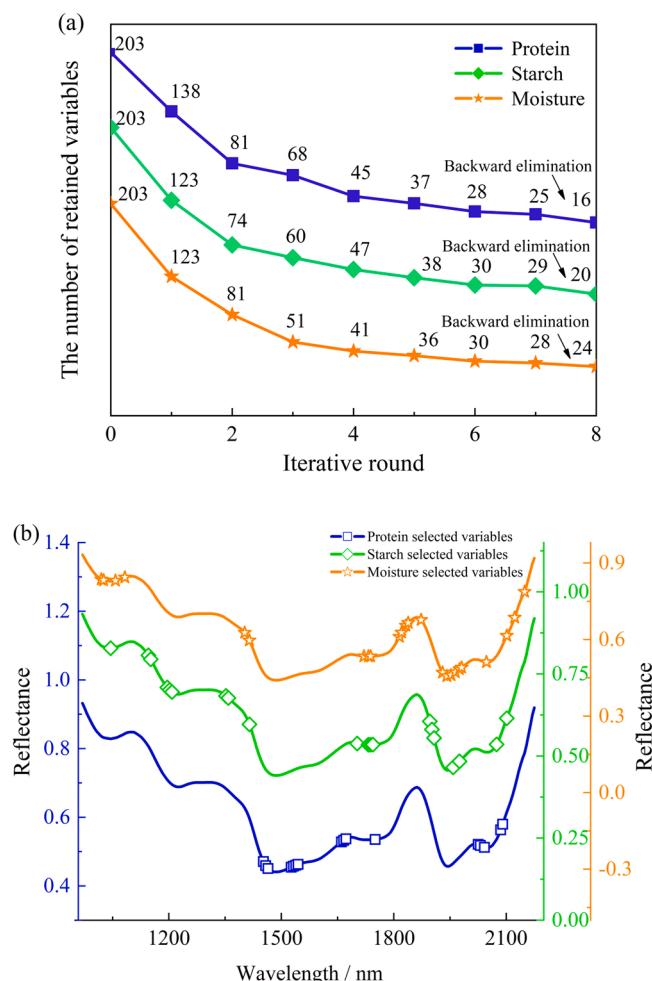
The correction models of the full spectral band were established in

Table 2

Model performances for predicting protein, starch and moisture content in wheat flour by using hyperspectral imaging technology.

	Model	Calibration set		Cross-validation set			Prediction set		RPD
		R _C ²	RMSEC /g/100 g	R _{CV} ²	RMSECV /g/100 g	R _P ²	RMSEP /g/100 g		
Protein content (g/100 g)	PLSR	0.9884	0.4666	0.9872	0.4934	0.9553	2.0031	0.9212	
	PCR	0.9861	0.5201	0.9843	0.5548	0.9580	1.8223	1.0126	
	SVM	0.9098	1.6813	0.9017	1.7471	0.8774	1.7667	1.0445	
	MLR	0.9783	0.2834	0.9876	0.4933	0.9546	0.7192	2.5658	
Starch content (g/100 g)	PLSR	0.9047	1.9318	0.8867	2.1040	0.8898	4.4953	1.0706	
	PCR	0.8932	2.0452	0.8755	2.2127	0.8442	4.5278	1.0629	
	SVM	0.892	2.3239	0.8802	2.4689	0.8327	4.7953	1.0036	
	MLR	0.9171	1.1087	0.8908	2.0704	0.8954	3.8357	1.2547	
Moisture content (g/100 g)	PLSR	0.8797	0.1752	0.7721	0.2409	0.8264	0.2540	2.4220	
	PCR	0.7008	0.3191	0.5287	0.3507	0.6622	0.3544	1.7359	
	SVM	0.8882	0.1715	0.6868	0.2849	0.6707	0.5553	1.1079	
	MLR	0.7486	0.2635	0.7068	0.2743	0.5623	0.4471	1.3760	

PLSR: partial least square regression; PCR: principal component regression; SVMR: support vector machine regression; MLR: multiple linear regression

**Fig. 3.** Results of selection by IRIV (a: remained variables growth pattern with iterations rounds, b: selected effective wavelengths).

accordance with the whole spectral data extracted from the hyperspectral images of wheat flour and their corresponding protein, starch and moisture reference values, and five algorithms were used to predict the chemical composition contents. Table 2 presents the results of R_C², RMSEC, R_P², RMSEP, R_{CV}², RMSECV and RPD of the corresponding optimised calibration models for PLSR, PCR, SVMR and MLR. The optimal full-wavelength prediction models for wheat flour protein, starch and moisture content were PCR, MLR and PLSR, respectively. The corresponding R_C² and R_P² were 0.9861, 0.9171, 0.8797 and 0.9580, 0.8954,

0.8264, respectively, and the associated RMSEC and RMSEP were 0.5201 g/100 g, 1.1087 g/100 g, 0.1750 g/100 g and 1.8223 g/100 g, 3.8357 g/100 g, 0.2540 g/100 g, respectively. And the RPD were 1.0126, 1.2547 and 2.4220. The spectra in the range of 969–2174 nm contained an abundance of redundant information, which was not conducive to improving the robustness and prediction speed of the model. Therefore, EWs were selected from the original spectra to further optimize the models.

3.4. Selection of EWs

3.4.1. Selection of EWs by IRIV

IRIV is a novel variable selection method that is based on the binary matrix shuffling filter (BMSF) (Yao et al., 2022). IRIV algorithm eliminates non-information and weak information variables through multiple iterations and effective information variables as EWs are retained. Fig. 3(a) illustrates the process of the changes in the number of remaining wavelengths with the increase in iteration rounds. IRIV algorithm was used to conduct a total of eight iterations round for protein, starch and moisture in the wavelength range of 962–2174 nm. During the first four iterations round, the number of wavelengths decreased sharply from 203 to 45, 47 and 41 because many irrelevant information wavelengths were eliminated and then decreased slowly in the subsequent round. This result was stable at the seventh iteration round, followed by backward elimination of 9, 9 and 4 variables. Fig. 3(b) shows 16, 20 and 24 effective variables for protein, starch and moisture selected from the original wavelengths, accounting for 7.88%, 9.85% and 11.82% of the total wavelengths, respectively.

3.4.2. Selection of EWs by VCPA

VCPA iteratively selected the best-performance EWs subset on the basis of exponentially decreasing function (EDF) and binary matrix sampling (BMS) (Huan et al., 2021). The parameters of VCPA were set as follows: the EDF runs was 50 times, the BMS runs was 1000 times, the selected wavelengths were determined by 5-fold cross validation and the ratio of the optimal subset was 0.1. Fig. 4(a–c) shows the variation trend of RMSECV during EDF operation. The characteristic space shrunk and RMSECV presented a downward trend as a whole with the repeated operation of EDF. The final step was to select the variable subset with the smallest RMSECV and 8 (protein), 9 (starch) and 12 (moisture) variables were extracted (Fig. 4(d)), accounting for 3.94%, 4.43% and 5.91% of the total wavelengths, respectively.

3.4.3. Selection of EWs by IVISSA

IVISSA combines global search and local search to intelligently optimize the position, width and combination of spectral intervals in an iterative manner (Ma et al., 2019). Fig. 5(a) shows the variation trend of RMSECV during iterations. In the present study, after 27, 30 and 26

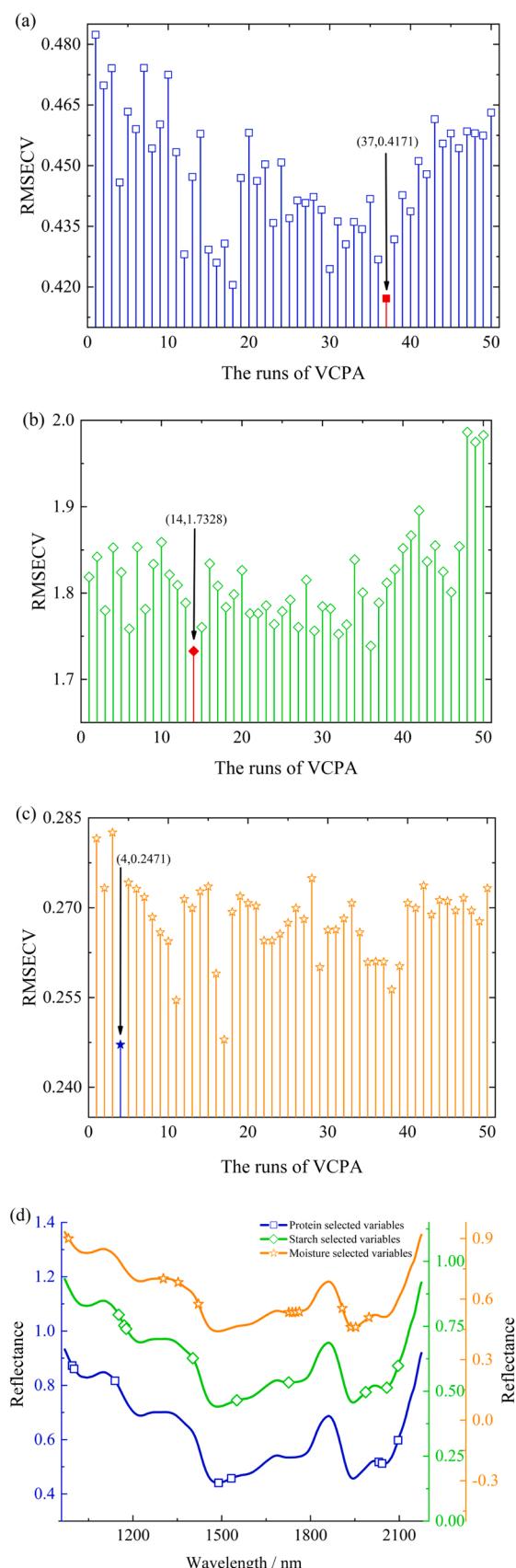


Fig. 4. Results of selection by VCPA (a–c): EWs were determined according to the minimum RMSECV for protein, starch and moisture content, respectively; (d): EWs distribution selected by IRIV.

iterations, the IVISSA algorithm selected 82 EWs for protein, 95 EWs for starch and 102 EWs for moisture. The algorithm extracted abundant EWs and the spacing between the bands was minor. Generally, approximate or similar wavelengths possessed the same or semblable information. Therefore, further extracting the EWs for the reduction in hyperspectral image dimensionality was necessary to decrease the non-effective information between adjacent bands to effectively improve the operation speed of the models. On the basis of IVISSA, IRIV was used to further select the EWs and a combination of IVISSA–IRIV was proposed to extract the optimal characteristic variables for 11, 28 and 36 for protein, starch and moisture (Fig. 5(b)), respectively, accounting for 5.42%, 13.79% and 17.73% of the total wavelengths, respectively.

3.4.4. Selection of EWs by MASS

The MASS algorithm mainly applies continuous model space shrinkage and weighted iteration strategy to obtain high-performance models in the model space. In this process, MASS applies a random sample program to combine outlier masking and variable combination effects to obtain the optimal model (Wen et al., 2016). Fig. 5(c) shows the variation trend of RMSECV during iterations, which was a declining trend. After 42, 39 and 41 iterations of the MASS algorithm, the minimum values of RMSECV dropped to 0.2587 g/100 g, 1.2138 g/100 g and 0.1762 g/100 g, respectively. A total of 64, 56 and 52 EWs for protein, starch and moisture were retained, respectively. In this work, MASS algorithm was combined with IRIV algorithm to establish a hybrid variable selection method to solve the problem of feature variables. The final variables subset obtained by the MASS-IRIV algorithm is shown in Fig. 5(d). A total of 13, 13 and 17 EWs for protein, starch and moisture were extracted from the variable sets of 42, 39 and 41, accounting for 6.40%, 6.40% and 8.37% of the total wavelengths, respectively.

3.4.5. Selection of EWs by IRF

IRF is a novel wavelength selection method that is based on random frog PLS framework (Yang et al., 2021). During the process, IRF calculates the selection probability of each wavelength in the 300 variable quantum sets generated after 300 iterations and arranges them in descending order. Each group of wavelengths is cross-verified to obtain RMSECV separately. The wavelength in the minimum group of RMSECV is the selected wavelength. As shown in Fig. 5(e), the first 81 variable subsets were selected as the protein EWs. IRF finally selected 105 EWs, including 976–1019, 1101–1164, 1415–1489, 1507–1550, 1581–1593, 1618–1678, 1690–1773, 1785–1797, 1873–1907 and 1964–2101 nm. The operation process of starch and moisture characteristic variables selection was the same as that of protein. The numbers of EWs selected were 111 and 94 (Fig. 5(f)). IRF retained many wavelength variables. The IRF results were further selected in combination with IRIV to further improve the robustness and speed of the model. The numbers of wavelengths of protein, starch and moisture were reduced from 105, 111 and 94–20, 15 and 20, respectively, in which the effectively reduced spectral dimension are shown in Fig. 5(g), accounting for 9.85%, 7.39% and 9.85% of the full wavelengths, respectively.

3.5. Modelling effect comparison

The performance of PLSR, PCR, SVMR and MLR models established based on the full NIR-HSI spectra were firstly assessed. A suitable EWs extraction algorithm was advantageous to acquire superior robustness and accuracy, whilst unsuitable algorithms may cause damage of EWs and decline the prediction accuracy (Jiang et al., 2020c). The whole wavelengths and extracted EWs were regarded as input data to the PLSR, PCR, SVMR and MLR models to assess the effect of characteristic extraction for the prediction models. The prediction results of protein, starch and moisture content by different models are indicated in Fig. 6. Comparing the performance of the established prediction models, except for the SVMR models that had a certain degree of overfitting, the models

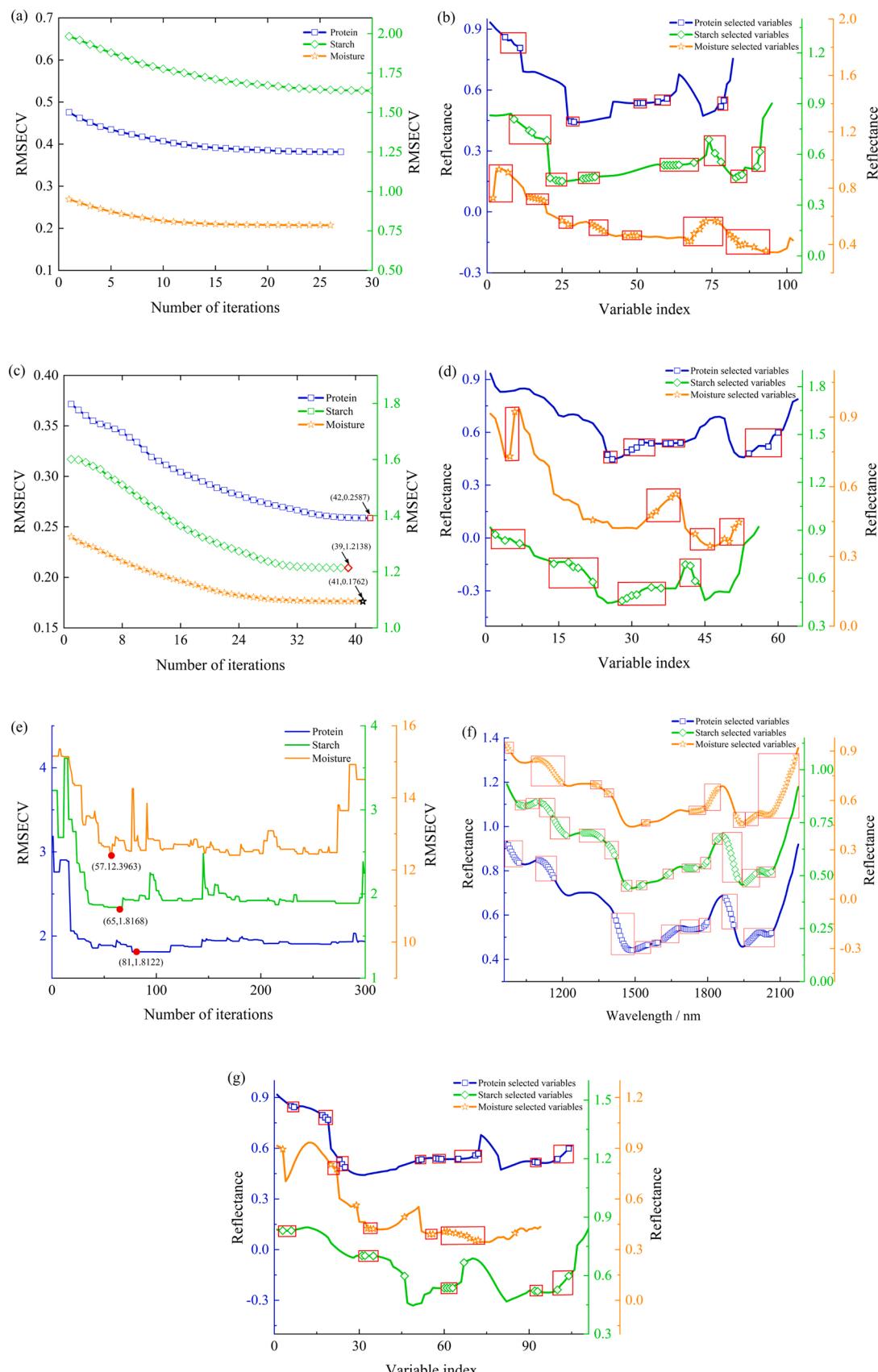


Fig. 5. IVISSA, IVISSA-IRIV, MASS, MASS-IRIV, IRF and IRF-IRIV algorithms to select EWs (a, c and e: variation trend of RMSECV during iterations; b, d, f and g: sequence number of selected EWs).

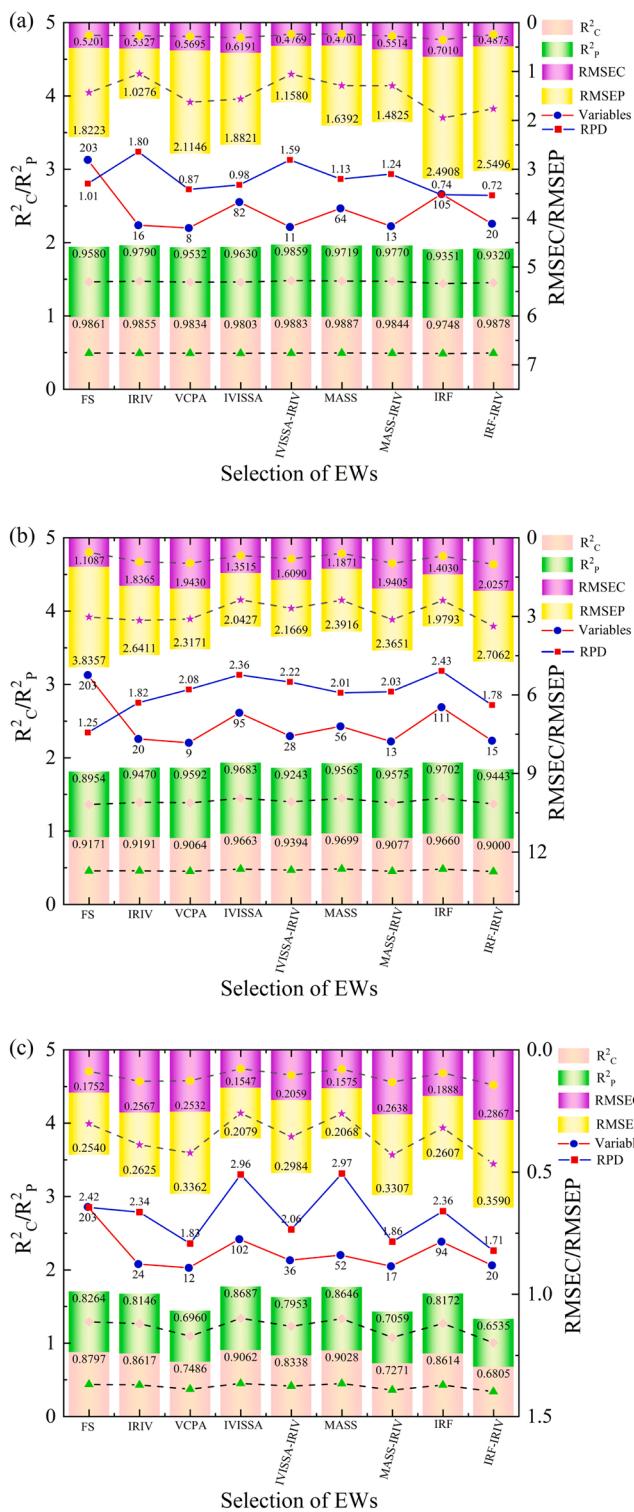


Fig. 6. Prediction results by the different models using full spectra and selected spectra (a: PCR-protein, b: MLR-starch and c: PLSR-moisture).

based on EWs could acquire high effectively prediction performance for protein, starch and moisture content of wheat flour. Although the IVISSA, MASS and IRF algorithms selected the major EWs to improve the performance of the models, the simplified performance of the models was inconspicuous. Therefore, it should be further combined with the high accuracy of IRIV algorithm to extract the EWs. In the quantitative analysis models of protein, starch and moisture, the FS-PCR, FS-MLR and FS-PLSR models had better performance than the other three models, the

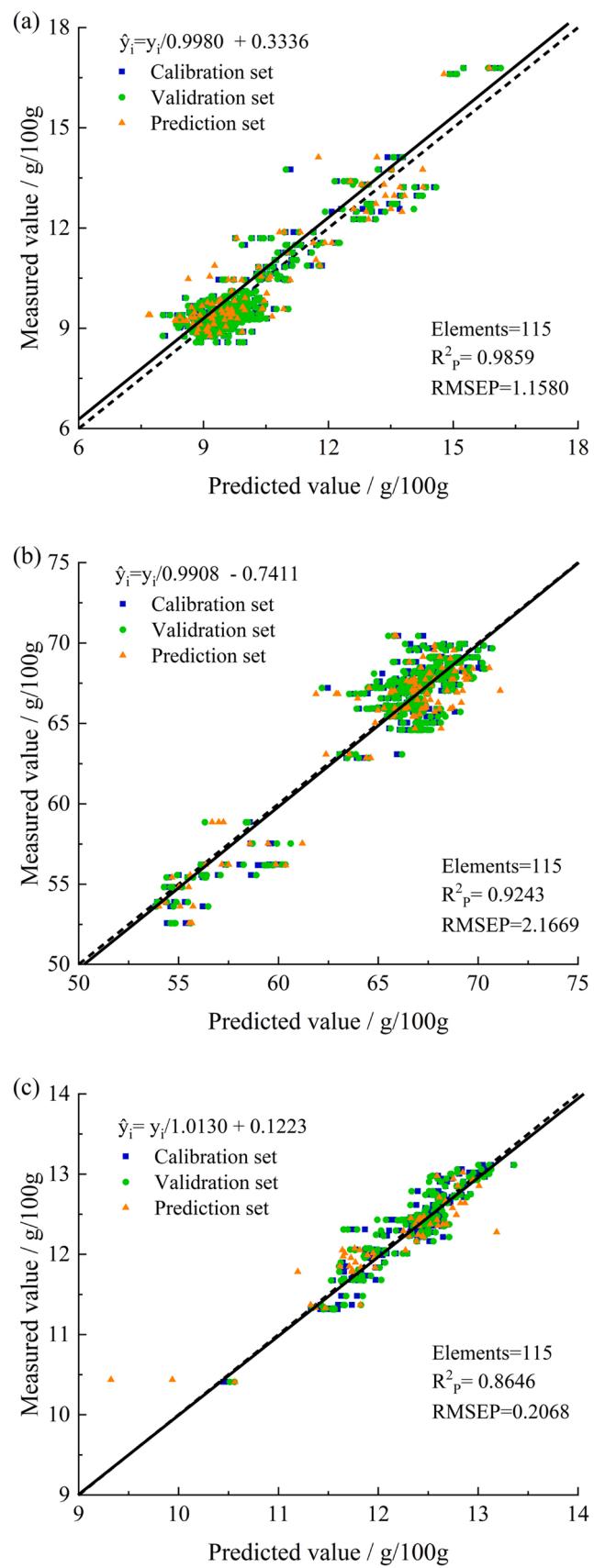


Fig. 7. Scatter plots of predicted protein, starch and moisture content obtained by (a) IVISSA-IRIV-PCR, (b) IVISSA-IRIV-MLR and (c) MASS-PLSR.

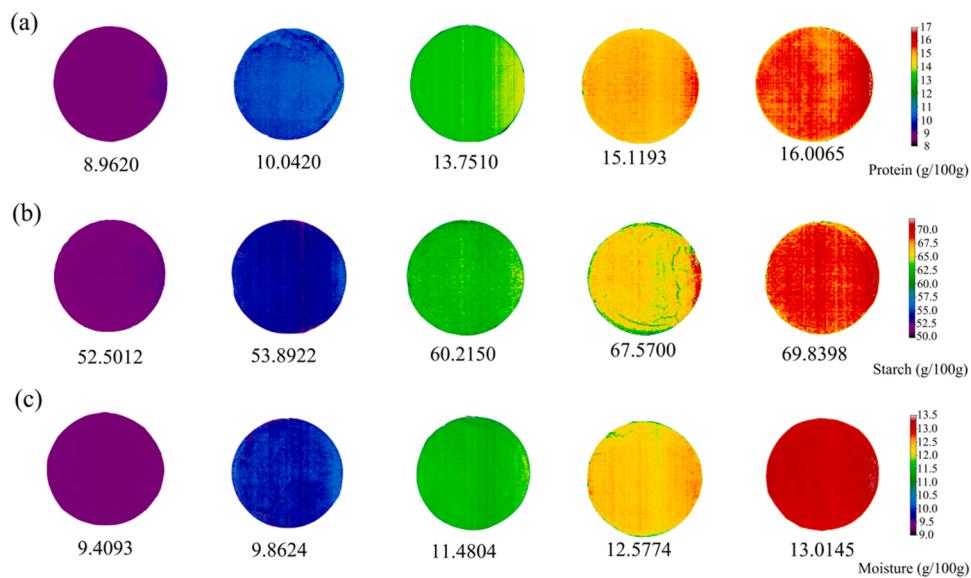


Fig. 8. Visualization map of protein, starch and moisture content in wheat flour.

prediction effect were introduced in the above analysis ($R_p^2 = 0.9580$, 0.8954 and 0.8264 and RMSEP = 1.8223 g/100 g, 3.8357 g/100 g and 0.2540 g/100 g, and RPD= 1.0126, 1.2547 and 2.4220, respectively). In the prediction models of protein, starch and moisture, the selected EWs algorithms were evaluated and compared to verify the accuracy and effectiveness of the EWs and the rapid determination. The optimal models were IVISSA-IRIV-PCR in predicting protein content ($R_c^2 = 0.9883$, $R_p^2 = 0.9859$, RMSEC = 0.4769 g/100 g, RMSEP = 1.1580 g/100 g, RPD = 1.5935), IVISA-IRIV-MLR in predicting starch content ($R_c^2 = 0.9394$, $R_p^2 = 0.9243$, RMSEC = 0.1575 g/100 g, RMSEP = 2.1669 g/100 g, RPD = 2.2209) and MASS-PLSR in predicting moisture content ($R_c^2 = 0.9028$, $R_p^2 = 0.8646$, RMSEC = 1.6090 g/100 g, RMSEP = 0.2068 g/100 g, RPD = 2.9749), which extracted 11, 28 and 52 EWs, respectively. Amongst them, IVISSA, MASS and IRF algorithms were used to identify the EWs in the moisture prediction model. The model still contained abundant bands, which were disadvantageous in terms of improving the calculation speed. Although the combination of these three algorithms and IRIV greatly reduced the number of wavelengths in the calibration set, the accuracy of the prediction model decreased, and the ideal prediction effect was not achieved, probably because some useful information related to moisture content was lost in the process of extracting important bands, which reduced the robustness of the model.

Fig. 7 was a scatter plot of the predicted and actual values of protein, starch and moisture contents in wheat flour based on the IVISSA-IRIV-PCR, IVISA-IRIV-MLR and MASS-PLSR models. The dotted line represented the regression line for the ideal correlation between the actual and predicted values of protein, starch and moisture. The sample points were closely distributed near the regression line, demonstrating that the prediction performance of the model was better. In this study, wheat flour can be divided into low-gluten flour and high-gluten flour according to the gluten content. High-gluten flour has higher protein content but lower starch content, while low-gluten flour has higher starch content but lower protein content. Therefore, it is reasonable that the protein and starch contents of the sample are distributed in two clusters. The purpose of choosing high-gluten flour and low-gluten flour as samples in this work is to expand the detection range of the model and lay a foundation for the application of the model in the future.

In conclusion, the selection of EWs could decrease the high dimensionality of hyperspectral data and complexity and improve the accuracy and calculation speed of the HSI system. The results showed that the NIR-HSI imaging technique could accurately achieve the detection of

protein content, starch content and moisture content in wheat flour. In conclusion, NIR-HSI technology is a suitable method for detecting wheat flour chemical composition.

3.6. Results of visualisation

NIR-HSI technology could simultaneously provide spectral and image information of the samples compared with conventional spectroscopic techniques (Yu et al., 2021). The pseudo colour map of protein, starch and moisture content models generated by the EWs in pixel-wise disposing visualisation on the account of linear colour scale, where the red area represents high content and the purple area represents low content, is shown in Fig. 8. A simplified model was established to predict the protein content of each pixel on the hyperspectral image by using the 11 EWs extracted by IVISSA-IRIV. Finally, a visualisation map of the detection index was constructed (Fig. 8(a)). Accordingly, the simplified models established by the EWs (28 and 52) extracted by IVISSA-MLR and MASS-PLSR were used to draw a visualisation map of the starch and moisture content of the samples (Fig. 8(b and c)). The colour changes exhibited in the predicted samples were automatically concentrated on a linear colour bar, where the different colours correspond to the different values of protein, starch and moisture content in wheat flour. The visualisation maps could directly reflect the spatial variation of protein, starch and moisture content in different variety samples or even in the same variety of samples, which beneficial to grasp the relative distribution of the substance content. Pseudo-colour map could display the distribution of basic chemical components in different wheat flours, intuitively allowing food processing industries and acquirers to select the desired wheat flour.

4. Conclusions

In this study, a rapid and reliable method for determining the changes in protein, starch and moisture content in wheat flours were proposed using NIR-HSI technique (with the spectral range of 969–2174 nm) combined with chemometrics. IRIV, VCPA, IVISSA, MASS and IRF and their combinatorial algorithms were applied as data dimensionality reduction methods to discover the EWs, which could reflect the variation of wheat flour's effective information in the spectral curves with spectral noise and overlap. A series of predictive fusion models of internal components (protein, starch and moisture) on the account of PLSR, PCR, SVMR and MLR were established with the FS and

EWs subset as input variables. The optimal models for protein, starch and moisture contents were IVISSA-IRIV-PCR ($R^2_C = 0.9883$, $R^2_P = 0.9859$), IVISA-IRIV-MLR ($R^2_C = 0.9394$, $R^2_P = 0.9243$) and MASS-PLSR ($R^2_C = 0.9028$, $R^2_P = 0.8646$), which extracted 11, 28 and 52 EWs, respectively. The protein, starch and moisture content in wheat flour could be estimated by the visualization map. The result indicated that the NIR-HSI technique is a highly efficient, rapid, non-destructive and feasible prediction method for evaluating protein, starch and moisture content in wheat flour. The method developed in this study provides a reference for the determination of basic substances in other agricultural products.

CRediT authorship contribution statement

Jing Zhang: performed the experiments, data curation and writing the original manuscript; Zhen Guo and Sihua Wang: contributed to the software, conception and methodology of this study; Minghui Yue and Shanshan Zhang: investigation, resources; Xiang Yin, Juan Du, Kuijie Gong and Chengye Ma: funding acquisition, supervision, project administration and writing-review and editing.

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), Key R&D Program of Shandong Province (Major Science and Technology Projects) (2020CXGC0108053).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.jfca.2023.105134](https://doi.org/10.1016/j.jfca.2023.105134).

References

- Achata, E.M., Esquerre, C., Ojha, K.S., Tiwari, B.K., O'Donnell, C.P., 2021. Development of NIR-HSI and chemometrics process analytical technology for drying of beef jerky. *Innov. Food Sci. Emerg. Technol.* 69, 102611 <https://doi.org/10.1016/j.ifset.2021.102611>.
- An, T., Yu, S., Huang, W., Li, G., Tian, X., Fan, S., Dong, C., Zhao, C., 2022. Robustness and accuracy evaluation of moisture prediction model for black tea withering process using hyperspectral imaging. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 269, 120791 <https://doi.org/10.1016/j.saa.2021.120791>.
- Barbon Junior, S., Mastelini, S.M., Barbon, A.P.A.C., Barbin, D.F., Calvini, R., Lopes, J.F., Ulrici, A., 2020. Multi-target prediction of wheat flour quality parameters with near infrared spectroscopy. *Inf. Process. Agric.* 7 (2), 342–354. <https://doi.org/10.1016/j.inpa.2019.07.001>.
- Caporaso, N., Whitworth, M.B., Fisk, I.D., 2018. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chem.* 240, 32–42. <https://doi.org/10.1016/j.foodchem.2017.07.048>.
- Chen, X., Siebler, H.W., Yan, H., 2021. Rapid analysis of wheat flour by different handheld near-infrared spectrometers: a discussion of calibration model maintenance and performance comparison. *Spectrochim. Acta - Part A: Mol. Biomol. Spectrosc.* 252, 119504 <https://doi.org/10.1016/j.saa.2021.119504>.
- Cheng, W., Sun, D.-W., Pu, H., Wei, Q., 2018. Characterization of myofibrils cold structural deformation degrees of frozen pork using hyperspectral imaging coupled with spectral angle mapping algorithm. *Food Chem.* 239, 1001–1008. <https://doi.org/10.1016/j.foodchem.2017.07.011>.
- Dou, Z., Gao, K., Zhang, X., Wang, H., Han, L., 2021. Band selection of hyperspectral images using attention-based autoencoders. *IEEE Geosci. Remote Sens. Lett.* 18 (1), 147–151. <https://doi.org/10.1109/LGRS.2020.2967815>.
- Dowell, F.E., Maghirang, E.B., Xie, F., Lookhart, G.L., Pierce, R.O., Seabourn, B.W., Bean, S.R., Wilson, J.D., Chung, O.K., 2006. Predicting wheat quality characteristics and functionality using near-infrared spectroscopy. *Cereal Chem.* 83 (5), 529–536. <https://doi.org/10.1094/CC-83-0529>.
- Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C., 2019. Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods* 15 (1), 91. <https://doi.org/10.1186/s13007-019-0476-y>.
- Huan, K., Chen, X., Song, X., Dong, W., 2021. Variable selection in near-infrared spectra: application to quantitative non-destructive determination of protein content in wheat. *Infrared Phys. Technol.* 119, 103937 <https://doi.org/10.1016/j.infrared.2021.103937>.
- Iymen, G., Tanriver, G., Hayirlioglu, Y.Z., Ergen, O., 2020. Artificial intelligence-based identification of butter variations as a model study for detecting food adulteration. *Innov. Food Sci. Emerg. Technol.* 66, 102527 <https://doi.org/10.1016/j.ifset.2020.102527>.
- Ji, Y., Sun, L., Li, Y., Li, J., Liu, S., Xie, X., Xu, Y., 2019. Non-destructive classification of defective potatoes based on hyperspectral imaging and support vector machine. *Infrared Phys. Technol.* 99, 71–79. <https://doi.org/10.1016/j.infrared.2019.04.007>.
- Li, L., Xie, S., Ning, J., Chen, Q., Zhang, Z., 2019. Evaluating green tea quality based on multisensor data fusion combining hyperspectral imaging and olfactory visualization systems. *J. Sci. Food Agric.* 99 (4), 1787–1794. <https://doi.org/10.1002/jsfa.9371>.
- Jiang, H., Liu, T., Chen, Q., 2020a. Quantitative detection of fatty acid value during storage of wheat flour based on a portable near-infrared (NIR) spectroscopy system. *Infrared Phys. Technol.* 109, 103423 <https://doi.org/10.1016/j.infrared.2020.103423>.
- Jiang, H., Jiang, X., Ru, Y., Wang, J., Xu, L., Zhou, H., 2020b. Application of hyperspectral imaging for detecting and visualizing leaf lard adulteration in minced pork. *Infrared Phys. Technol.* 110, 103467 <https://doi.org/10.1016/j.infrared.2020.103467>.
- Jiang, H., Liu, T., He, P., Chen, Q., 2020c. Quantitative analysis of fatty acid value during rice storage based on olfactory visualization sensor technology. *Sens. Actuators B: Chem.* 309, 127816 <https://doi.org/10.1016/j.snb.2020.127816>.
- Li, S., Luo, J., Zhou, X., Li, X., Wang, F., Liu, Y., 2021. Identification of characteristic proteins of wheat varieties used to commercially produce dried noodles by electrophoresis and proteomics analysis. *J. Food Compos. Anal.* 96, 103685 <https://doi.org/10.1016/j.jfca.2020.103685>.
- Long, Y., Huang, W., Wang, Q., Fan, S., Tian, X., 2022. Integration of textural and spectral features of Raman hyperspectral imaging for quantitative determination of a single maize kernel mildew coupled with chemometrics. *Food Chem.* 372, 131246 <https://doi.org/10.1016/j.foodchem.2021.131246>.
- Ma, C., Ren, Z., Zhang, Z., Du, J., Jin, C., Yin, X., 2021. Development of simplified models for nondestructive testing of rice (with husk) protein content using hyperspectral imaging technology. *Vib. Spectrosc.* 114, 103230 <https://doi.org/10.1016/j.vibspec.2021.103230>.
- Ma, J., Cheng, J., Sun, D., Liu, D., 2019. Mapping changes in sarcoplasmatic and myofibrillar proteins in boiled pork using hyperspectral imaging with spectral processing methods. *LWT* 110, 338–345. <https://doi.org/10.1016/j.lwt.2019.04.095>.
- Mæhre, H.K., Dalheim, L., Edvinsen, G.K., Ellevoll, E.O., Jensen, I.J., 2018. Protein determination—method matters. *Foods* 7 (1). <https://doi.org/10.3390/foods7010005>.
- Mahesh, S., Jayas, D.S., Paliwal, J., White, N.D.G., 2015. 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 Bioprocess Technol.* 8 (1), 31–40. <https://doi.org/10.1007/s11947-014-1381-z>.
- Panda, B.K., Mishra, G., Ramirez, W.A., Jung, H., Singh, C.B., Lee, S.H., Lee, I., 2022. Rancidity and moisture estimation in shelled almond kernels using NIR hyperspectral imaging and chemometric analysis. *J. Food Eng.* 318, 110889 <https://doi.org/10.1016/j.jfoodeng.2021.110889>.
- Ren, G., Ning, J., Zhang, Z., 2020. Intelligent assessment of tea quality employing visible-near infrared spectra combined with a hybrid variable selection strategy. *Microchem. J.* 157, 105085 <https://doi.org/10.1016/j.microc.2020.105085>.
- Silva, L.C.R., Folli, G.S., Santos, L.P., Barros, I.H.A.S., Oliveira, B.G., Borghi, F.T., Santos, F.D., Filgueiras, P.R., Romão, W., 2020. Quantification of beef, pork, and chicken in ground meat using a portable NIR spectrometer. *Vib. Spectrosc.* 111, 103158 <https://doi.org/10.1016/j.vibspec.2020.103158>.
- Srivastava, S., Mishra, H.N., 2022. Detection of insect damaged rice grains using visible and near infrared hyperspectral imaging technique. *Chemom. Intell. Lab. Syst.* 221, 104489 <https://doi.org/10.1016/j.chemolab.2021.104489>.
- Sun, J., Shi, X., Zhang, H., Xia, L., Guo, Y., Sun, X., 2019. Detection of moisture content in peanut kernels using hyperspectral imaging technology coupled with chemometrics. *J. Food Process Eng.* 42 (7), e13263 <https://doi.org/10.1111/jfpe.13263>.
- Tian, W., Chen, G., Zhang, G., Wang, D., Tilley, M., Li, Y., 2021. Rapid determination of total phenolic content of whole wheat flour using near-infrared spectroscopy and chemometrics. *Food Chem.* 344, 128633 <https://doi.org/10.1016/j.foodchem.2020.128633>.
- Wan, G., Liu, G., He, J., Luo, R., Cheng, L., Ma, C., 2020. Feature wavelength selection and model development for rapid determination of myoglobin content in nitrite-cured mutton using hyperspectral imaging. *J. Food Eng.* 287, 110090 <https://doi.org/10.1016/j.jfoodeng.2020.110090>.
- Wang, Y., Lv, Y., Liu, H., Wei, Y., Zhang, J., An, D., Wu, J., 2018. Identification of maize haploid kernels based on hyperspectral imaging technology. *Comput. Electron. Agric.* 153, 188–195. <https://doi.org/10.1016/j.compag.2018.08.012>.

- Wen, M., Deng, B., Cao, D., Yun, Y., Yang, R., Lu, H., Liang, Y., 2016. The model adaptive space shrinkage (MASS) approach: a new method for simultaneous variable selection and outlier detection based on model population analysis. *Analyst* 141 (19), 5586–5597. <https://doi.org/10.1039/C6AN00764C>.
- Wiedemair, V., Mair, D., Held, C., Huck, C.W., 2019. 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 miliaecum L.* *Talanta* 205, 120115. <https://doi.org/10.1016/j.talanta.2019.120115>.
- Xia, Y., Huang, W., Fan, S., Li, J., Chen, L., 2019. Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Phys. Technol.* 97, 467–477. <https://doi.org/10.1016/j.infrared.2019.01.012>.
- Yang, X., Liu, G., He, J., Kang, N., Yuan, R., Fan, N., 2021. Determination of sugar content in Lingwu jujube by NIR-hyperspectral imaging. *J. Food Sci.* 86 (4), 1201–1214. <https://doi.org/10.1111/1750-3841.15674>.
- Yao, K., Sun, J., Chen, C., Xu, M., Zhou, X., Cao, Y., Tian, Y., 2022. Non-destructive detection of egg qualities based on hyperspectral imaging. *J. Food Eng.* 325, 111024 <https://doi.org/10.1016/j.jfoodeng.2022.111024>.
- Ye, D., Sun, L., Zou, B., Zhang, Q., Tan, W., Che, W., 2018. Non-destructive prediction of protein content in wheat using NIRS. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 189, 463–472. <https://doi.org/10.1016/j.saa.2017.08.055>.
- Yu, H., Qing, L., Yan, D., Xia, G., Zhang, C., Yun, Y., Zhang, W., 2021. Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness. *Food Chem.* 348, 129129 <https://doi.org/10.1016/j.foodchem.2021.129129>.
- Zhang, C., Jiang, H., Liu, F., He, Y., 2017. Application of near-infrared hyperspectral imaging with variable selection methods to determine and visualize caffeine content of coffee beans. *Food Bioprocess Technol.* 10 (1), 213–221. <https://doi.org/10.1007/s11947-016-1809-8>.
- Zhang, J., Ma, Y., Liu, G., Fan, N., Li, Y., Sun, Y., 2022. Rapid evaluation of texture parameters of Tan mutton using hyperspectral imaging with optimization algorithms. *Food Control* 135, 108815. <https://doi.org/10.1016/j.foodcont.2022.108815>.
- Zhang, Z., Yin, X., Ma, C., 2019. Development of simplified models for the nondestructive testing of rice with husk starch content using hyperspectral imaging technology. *Anal. Methods* 11 (46), 5910–5918. <https://doi.org/10.1039/C9AY01926J>.