https://doi.org/10.17221/129/2022-CJFS

Rapid non-destructive identification of selenium-enriched millet based on hyperspectral imaging technology

Fu Zhang[1,2,3], Xiahua Cui[1], Chaochen Zhang[1], Weihua Cao[1], Xinyue Wang[1], Sanling Fu[4], Shuai Teng[1]*

[1]*College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, China*
[2]*Key Laboratory of Modern Agricultural Equipment and Technology, Jiangsu University, Zhenjiang, China*
[3]*Collaborative Innovation Center of Machinery Equipment Advanced Manufacturing of Henan Province, Luoyang, China*
[4]*College of Physical Engineering, Henan University of Science and Technology, Luoyang, China*

**Corresponding author: fusanling@126.com*

Citation: Zhang F., Cui X.H., Zhang C.C., Cao W.H., Wang X.Y., Fu S.L., Teng S. (2022): Rapid non-destructive identification of selenium-enriched millet based on hyperspectral imaging technology. Czech J. Food Sci., 40: 445–455.

Abstract: To meet rapid and non-destructive identification of selenium-enriched agricultural products selenium-enriched millet and ordinary millet were taken as objects. Image regions of interest (ROI) were selected to extract the spectral average value based on hyperspectral imaging technology. Reducing noise by the Savitzky-Golay (SG) smoothing algorithm, variables were used as inputs that were screened by successive projections algorithm (SPA), competitive adaptive reweighted sampling (CARS), uninformative variable elimination (UVE), CARS-SPA, UVE-SPA, and UVE-CARS, while sample variables were used as outputs to build support vector machine (SVM) models. The results showed that the accuracy of CARS-SPA-SVM was 100% in the training set and 99.58% in the test set equivalent to that of CARS-SVM and UVE-CARS-SVM, which was higher than that of SPA-SVM, UVE-SPA-SVM, and UVE-SVM. Therefore, the method of CARS-SPA had superiority, and CARS-SPA-SVM was suitable to identify selenium-enriched millet. Finally, 454.57 nm, 484.98 nm, 885.34 nm, and 937.1 nm, which were obtained by wavelength extraction algorithms, were considered as the sensitive wavelengths of selenium information. This study provided a reference for the identification of selenium-enriched agricultural products.

Keywords: millet; selenium element; spectroscopy; secondary variable screening; support vector machine

Selenium plays an important role that cannot be neglected in antioxidant and disease prevention, but the body cannot synthesize the required selenium. So dietary selenium supplementation is the most important way of selenium supplement (Mojadadi et al., 2021). Therefore, the identification of selenium-enriched foodstuffs has a high research significance. Millet is a better carrier of selenium (Mu et al. 2018; Liang et al. 2020). However, the quality of millet on the market is not easy to judge. The correct identification

Supported by the National Key Research and Development Program of China (Project No. 2017YFD0301106), the Scientific and Technological Project of Henan Province (Project No. 212102110029), the High-tech Key Laboratory of Agricultural Equipment and Intelligence of Jiangsu Province (Project No. JNZ201901).

https://doi.org/10.17221/129/2022-CJFS

of selenium-enriched millet provides the possibility to improve the human body's selenium deficiency state (Yu et al. 2018; Rayman et al. 2000).

Nowadays, the detection methods of selenium in foodstuffs include fluorescence spectrophotometry, hydride generation atomic fluorescence spectrometry, inductively coupled plasma mass spectrometry (Zhao et al. 2021). However, both fluorescence spectrophotometry and hydride generation fluorescence spectrometry need chemical treatment of the samples, resulting in waste. The mass spectrum detector required by an inductively coupled plasma mass spectrometer is difficult to popularise.

Hyperspectral imaging technology has the advantages of multi-wavelengths, high resolution, and map integration, without sample pre-treatment and pollution. Compared with the area sampling method of near-infrared spectroscopy, selecting the region of interest (ROI) area in the hyperspectral images can better reflect the overall situation and make the data more random. So, it has already been widely used in agricultural engineering (Khan et al., 2021; Wang et al. 2021). Millet samples from different areas were identified based on hyperspectral imaging technology and machine learning, and it was found that various algorithm models had different effects on millet identification (Ji et al. 2019). A partial least squares (PLS) model of total selenium in millet was established which contents of ninety-three samples. The results showed that the ratio of the standard error of the test set and that of the training set was 1.073, indicating that the model prediction accuracy and robustness were high. It could realize rapid detection of total selenium content (Wang et al. 2021). It was judged whether high-quality rice was mixed with inferior rice using hyperspectral imaging technology. The results showed that the cross-validation accuracy of SVM model based on characteristic wavelength was 95% and the prediction set was 96%, while the cross-validation accuracy of SVM model based on the optimal principal component number was 94% and the prediction set was 98%. It provided a fast and non-destructive detection method for the problem of rice adulteration (Sun et al. 2014). Hyperspectral imaging technology was used to analyse the spectral image features from three types of rice. The results showed that rice had prominent characteristic peaks in the band of 480–550 nm, and various kinds of rice had different light and dark ratios in binary images (Si et al. 2020). Based on visible/near-infrared hyperspectral image analysis technology (NIRS^{TM} DS2500; FOSS, Denmark), a total of 480 samples of eight millet varieties were analysed, and spectral and image features were extracted, then the SVM models and the attentional convolutional neural network models were established. The results showed that the attention-radial basis function neural network models could greatly improve the recognition accuracy of varieties (Wang et al. 2020).

Hyperspectral imaging technology has attracted much attention due to its non-destructive detection characteristics (Garcia-Martin et al. 2020; Liu et al. 2021). Selenium-enriched millet and ordinary millet were distinguished based on the hyperspectral imaging technology in the study, which will provide a reference for the efficient identification of selenium-enriched millets.

MATERIAL AND METHODS

Material

The purchased experimental samples were bagged selenium-enriched fragrant Duojiao millet and small grain Shennong millet. The labelled selenium content of selenium-enriched fragrant Duojiao millet was 0.14 mg kg^{-1}, and small grain Shennong millet was ordinary millet without selenium (detected by Agricultural Products Quality Supervision and Testing Center of Ministry of Agriculture in China). Forty sets of samples were taken from each kind of millet samples, a total of 80 sets of samples were taken, and each of them had a mass of 50 g.

Hyperspectral image acquisition

The hyperspectral image acquisition system used in the experiment mainly included a complementary metal oxide semiconductor (CMOS) camera (Hurricane-40-U2; Photonfocus, Switzerland), a line-scan spectrograph (ImSpector V10E; Specim, Finland), two 250 W optical fibre halogen tungsten lamps (ALPHA-1501, 21 V / 250 W), a set of conveying devices and other components. The spectral range was 400–1 000 nm, and the spectral resolution was 1 nm.

During the collection, the samples of each group were laid flat in a petri dish and placed on a mobile platform for image collection. The exposure time of the camera and the speed of the platform movement were determined in advance to ensure the image was clear. Finally, a 300 × 400 × 350 hyperspectral image was obtained. Then the original image $I_{original}$ was corrected to get the image I following the calculation formula:

$$I = \frac{I_{original} - I_{\text{black}}}{I_{\text{white}} - I_{\text{black}}} \quad (1)$$

In the formula, I was the corrected image, which was used for subsequent processing.

https://doi.org/10.17221/129/2022-CJFS

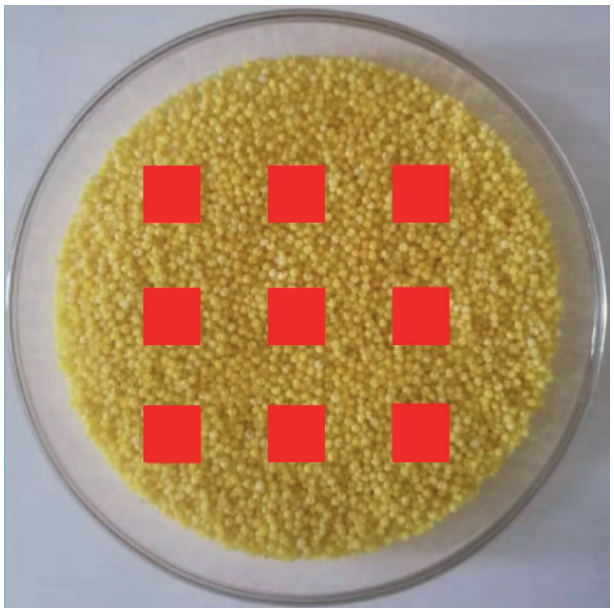

Figure 1. Schematic diagram of millet sample ROI selection

ROI – region of interest

Region of interest extraction

To make full use of millet samples and to increase the number of model training samples, nine ROIs were selected for each hyperspectral image. The schematic diagram of ROI selection is shown in Figure 1.

The number of pixels for each ROI was 400 and the average spectral value in the ROI was used as a spectral record of the sample. Finally, a total of 720 spectral curves were collected. Then the spectral curves of the two kinds of millet were averaged. Because of the noise at both ends of the extracted spectral curves, a total of 478 wavelengths were selected in the middle region of 450.53–950.09 nm. The average original spectral curves are shown in Figure 2.

For more useful information, the Savitzky-Golay (SG) smoothing algorithm was applied to reduce the noise. The smoothing point was set to three. The spectral curves after noise reduction are shown in Figure 3.

Characteristic wavelength selection

There were lots of redundancy and multicollinearity problems in the 478 wavelengths. Selecting representative characteristic wavelengths was one of the effective ways to solve the problems.

Successive projections algorithm (SPA). The SPA is an approach to forwarding the characteristic variable selection. SPA selects the least collinearity variables combination that contains the least redundant information (Araújo et al. 2001; Cao et al. 2021). The brief steps are as follows:

Assuming the starting iteration vector $x_{k(0)}$, the number of variables N needed to be extracted and the spectral matrix of column J are given.

Step 1. Select one optional column of the spectral matrix (j column), assign the j column of the modelling set to x_j, and, and note it as $x_{k(0)}$.

Step 2. Let the set of unselected wavelengths be S, that is:

$$S = \left\{ j,\ 1 \le j \le J,\ j \notin \left[k(0), \cdots k\left(n-1\right) \right] \right\} \tag{2}$$

Step 3. Calculate the projection of x_j to the other column vectors separately:

$$P_{x_j} = x_j - \left(x_j^T x_{k(n-1)} \right) x_{k(n-1)} \left(x_{k(n-1)}^T x_{k(n-1)} \right)^{-1}, \quad j \in s \tag{3}$$

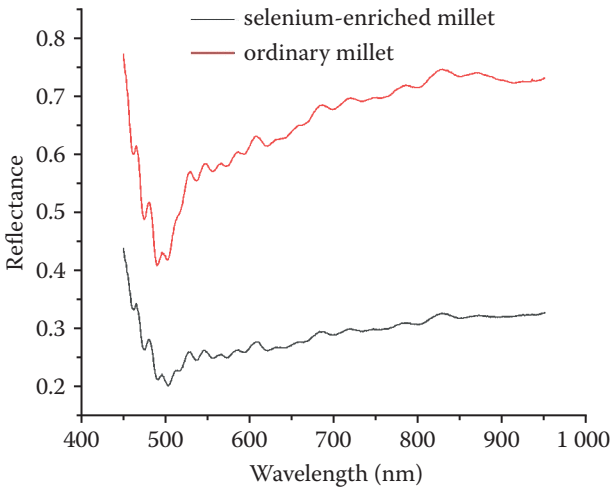

Figure 2. Original spectral curves of selenium-enriched millet and ordinary millet

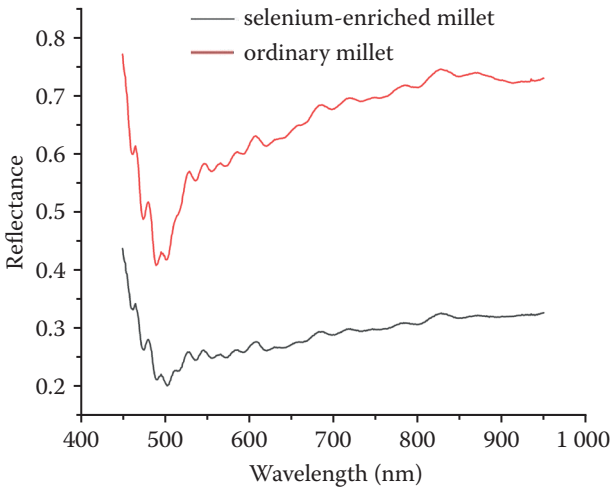

Figure 3. Savitzky-Golay (SG) corrected spectral curves of selenium-enriched millet and ordinary millet

https://doi.org/10.17221/129/2022-CJFS

Step 4. Extract the spectral wavelength of the largest projection vector:

$$k(n) = arg\left[max\left(P_{(x_j)} \right),\ j \in s \right] \quad (4)$$

Step 5. Let $x_j = p_x,\ j \in s$.

Step 6. Let $n = n + 1$, if $n < N$, then resume Step 1 to start loop computing.

Finally, the extracted wavelengths are

$$\left(x_{k(n)} = 0,\ \cdots,\ N-1 \right) \quad (5)$$

Competitive adaptive reweighted sampling (CARS). The CARS firstly uses adaptive reweighted sampling technology and exponential decay function to select the wavelengths with the largest absolute value of regression coefficient in the PLS model, then it removes the wavelengths with small weight. Finally, the set is chosen for the optimal variable's combination with the lowest root mean square error of cross validation (RMSECV) of prediction by cross-validation (Li et al. 2009; Guo et al. 2020). The steps of the CARS algorithm are as follows:

Step 1. Draw a fixed proportion of samples randomly from the data set by Monte Carlo sampling (MCS) and establish the PLS model to obtain the absolute value of the regression coefficient.

Step 2. Select the corresponding wavelengths of the regression coefficient with a large absolute value by exponentially decreasing function (EDF).

Step 3. Process the data corresponding to the selected wavelength based on adaptive reweighted sampling (ARS) and establish the PLS model to calculate its RMSECV.

Step 4. Repeat the above steps to substitute the new subset of the remaining wavelengths into the calculation.

Finally, select the most optimal subset of wavelengths with the smallest RMSECV.

Uninformative variable elimination (UVE). UVE gets rid of variables that are not as important as random variables in the model after adding artificial random variables to the data as a reference (Zou et al. 2010; Guo et al. 2016). The steps of the UVE algorithm are as follows:

Step 1. Establish the PLS model of the spectral matrix $X_{n\times m}$ and the target component content matrix $Y_{n\times 1}$, then select the optimal factor f through the leave-one-out cross-validation, where n is the number of samples, and m is the number of wavelength variables, and the following is the same.

Step 2. Generate a noise matrix $R_{n\times m}$ artificially, and it consists of random numbers with a data interval of 0.0-1.0 multiplied by a tiny constant to ensure that the impact on the model is negligible. Combine spectral matrix $X_{n\times m}$ and noise matrix $R_{n\times m}$ to form dilation matrix $X^R_{n\times 2m}$.

Step 3. Establish the PLS model of the extended matrix $X^R_{n\times 2m}$ and the target component content matrix $Y_{n\times 1}$ to obtain the regression coefficient matrix $B_{n\times 2m}$ by the leave-one-out cross-validation.

Step 4. Calculate the standard deviation $S(b)$ and the mean value $mean(b)$ of the regression coefficient matrix $B_{n\times 2m}$ in columns, then calculate $C_i = mean(b_i) / S(b_i)$, $i = 1, 2, \ldots, 2m$.

Step 5. Take the maximum absolute value in the stability index C of random variables as the variable selection cut-off value on $(m + 1, 2m)$. It is:

$$C_{max} = max\left[abs(C) \right] \quad (6)$$

Step 6. Define the variables of spectral matrix $X_{n\times m}$ corresponding to $C_i < C_{max}$ as uninformative variables and remove them, then form the remaining variables into a new matrix X_{new}.

Secondary variable selection. To further simplify the model by reducing the number of variables, UVE and CARS are selected as primary variable selection methods, and then SPA is used for the secondary variable selection.

Discriminant Analysis

To identify the selenium-enriched situation of millet effectively, SVM models were established to screen out the best performance prediction model. The models were based on SPA, CARS, UVE, UVE-SPA, CARS-SPA, and UVE-CARS characteristic wavelength extraction methods. The higher the correct recognition accuracy, the better the identification performance of the model. The number of input variables was the number of characteristic wavelengths required to maintain the corresponding identification accuracy. It was used as a standard to measure the practicality of the model.

RESULTS AND DISCUSSION

According to the statistical results of the samples, the spectral values of the data set were divided into two parts at the ratio of 2 : 1, as shown in Table 1. In the analysis, the categorical variable of selenium-enriched millet was set to 1, and the categorical variable of ordinary millet was set to 2.

Table 1. Division results of set samples number

Data set	Selenium-enriched millet	Ordinary millet
Training set	240	240
Test set	120	120
Total	360	360

https://doi.org/10.17221/129/2022-CJFS

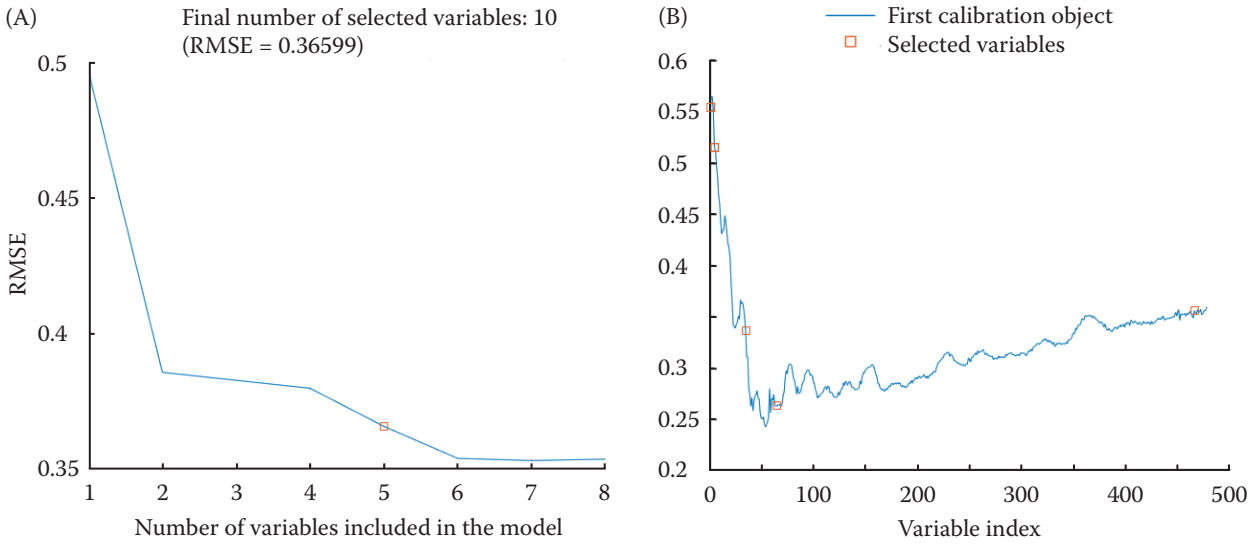

Figure 4. Key variables selection by SPA: (A) variation of RMSE, (B) key variables extracted

RMSE – root mean square error; SPA – successive projections algorithm

Successive projections algorithm (SPA)

The process of extracting feature parameters by SPA is shown in Figure 4. According to the principle of minimum root mean square error (RMSE), when the number of variables was five, the change of RMSE was no longer significant, and the RMSE was 0.36599, as shown in Figure 4A. Finally, five wavelength variables were selected. They were 450.53 nm, 454.57 nm, 484.98 nm, 515.53 nm, and 937.1 nm, as shown in Figure 4A.

Competitive adaptive reweighted sampling (CARS)

The process of extracting feature parameters by CARS is shown in Figure 5, Figure 5A shows the changes in the number of wavelength variables, Figure 5B

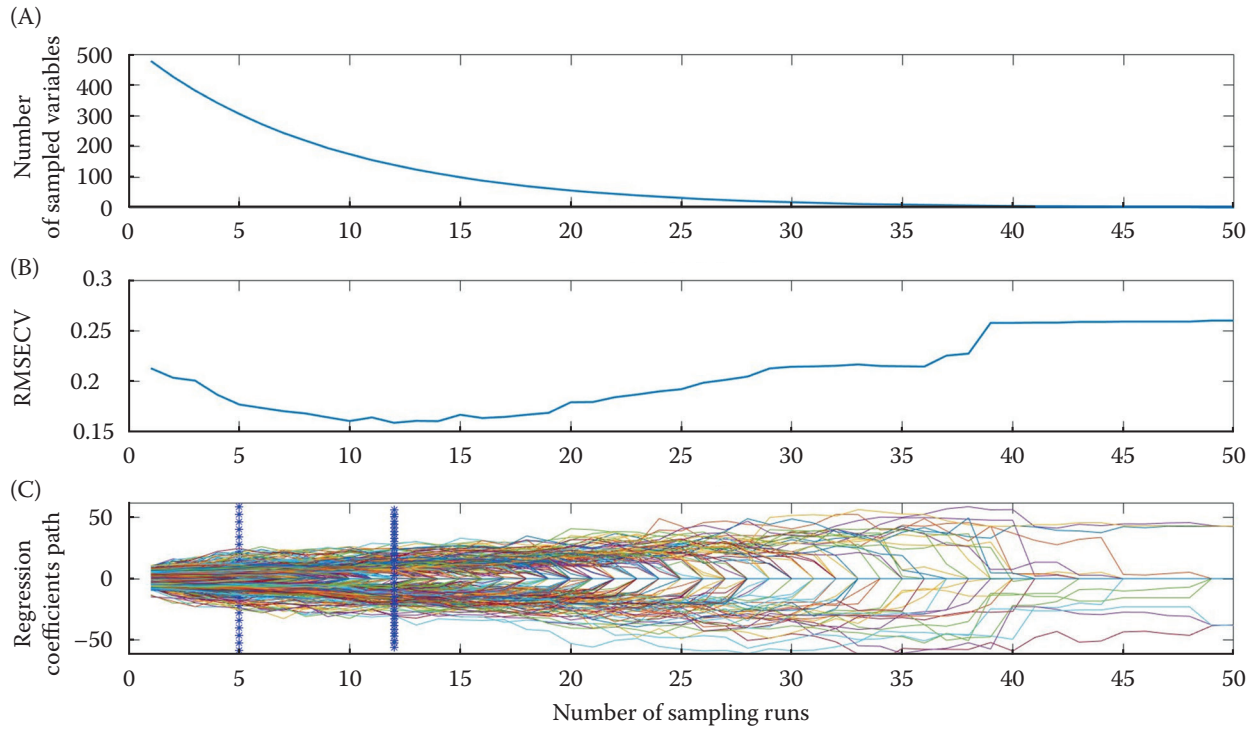

Figure 5. Key variables selection by CARS: (A) changes in the number of wavelength variables, (B) variation of RMSECV, (C) path of variables regression coefficients

CARS – competitive adaptive reweighted sampling; RMSECV – root mean square error of cross validation

https://doi.org/10.17221/129/2022-CJFS

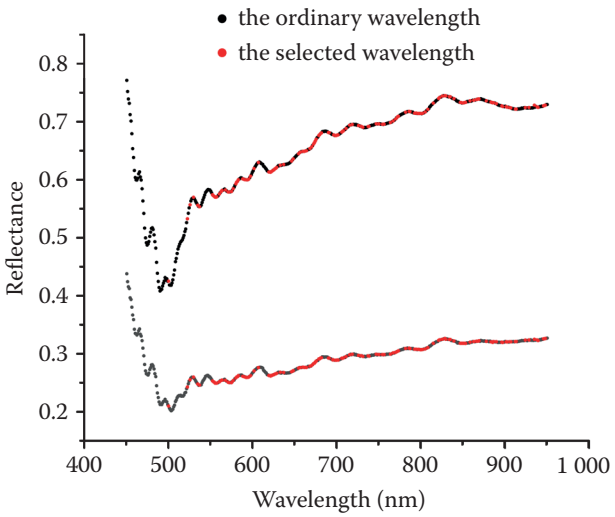

Figure 6. Original spectral curves of selenium-enriched millet and ordinary millet

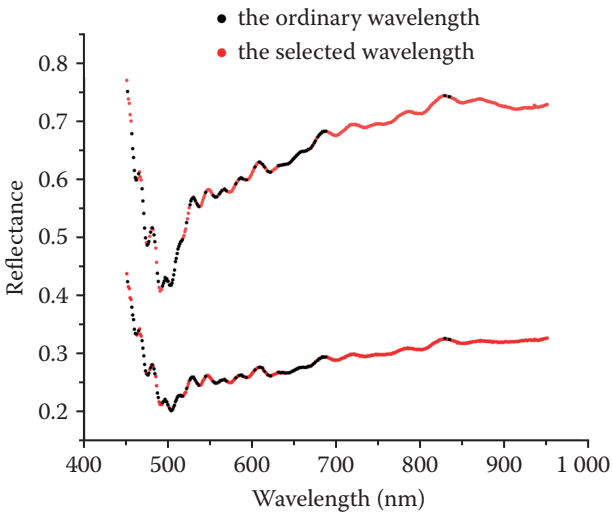

Figure 8. Key variables selection results by UVE

UVE – uninformative variable elimination

shows the variation of RMSECV, and Figure 5C shows the path of variables regression coefficients. It was easy to see after 12 samplings by the CARS algorithm, a total of 140 variables between 500.24 nm and 947.92 nm was selected, as shown in Figure 6, where the red dots are the selected variables.

Uninformative variable elimination (UVE)

The stability distribution map was optimised by UVE variables, as shown in Figure 7, and the mean coefficient/standard deviation was used to evaluate the stability value. The red part in the figure is the corresponding stable value of the artificially added random variable. It determined the threshold value of spectral variables through the fluctuation of the stable value of the random variable, i.e. the corresponding variable was discarded when the stability value was within the threshold and variables outside the threshold were retained for model building. Finally, the number of variables was reduced from 478 to 356 after UVE screening, while the variables were between 450.53 nm and 950.09 nm as shown in Figure 8, where the red dots are the selected variables.

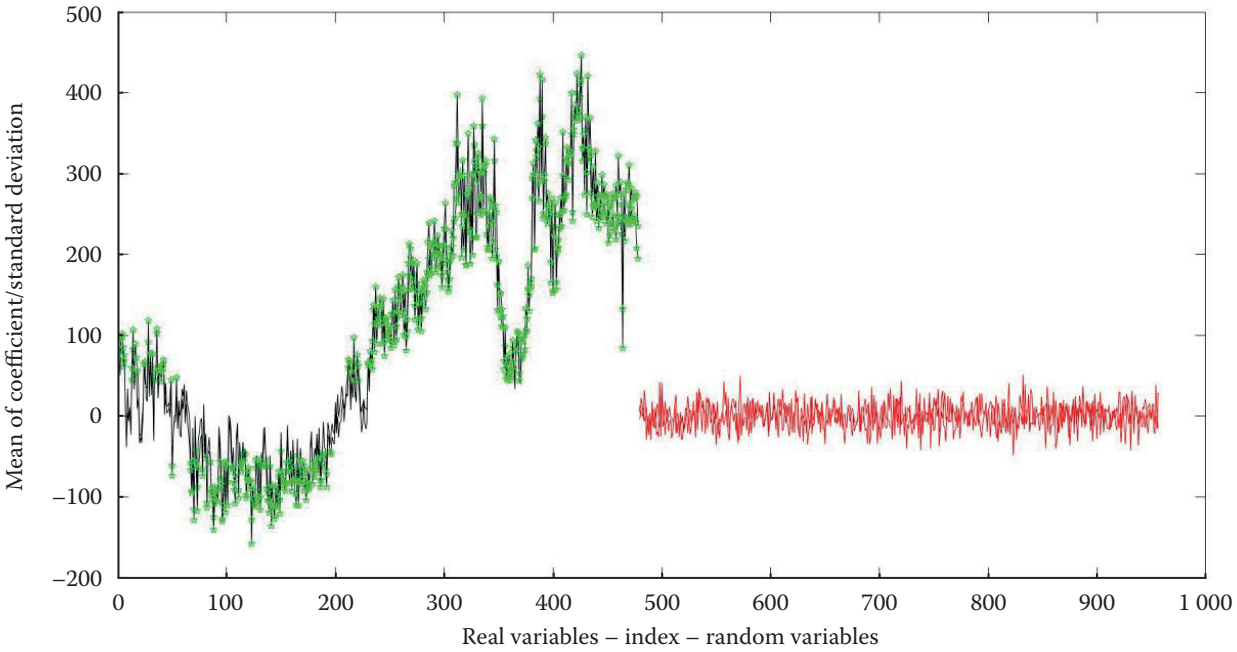

Figure 7. Key variables selection by UVE

UVE – uninformative variable elimination

https://doi.org/10.17221/129/2022-CJFS

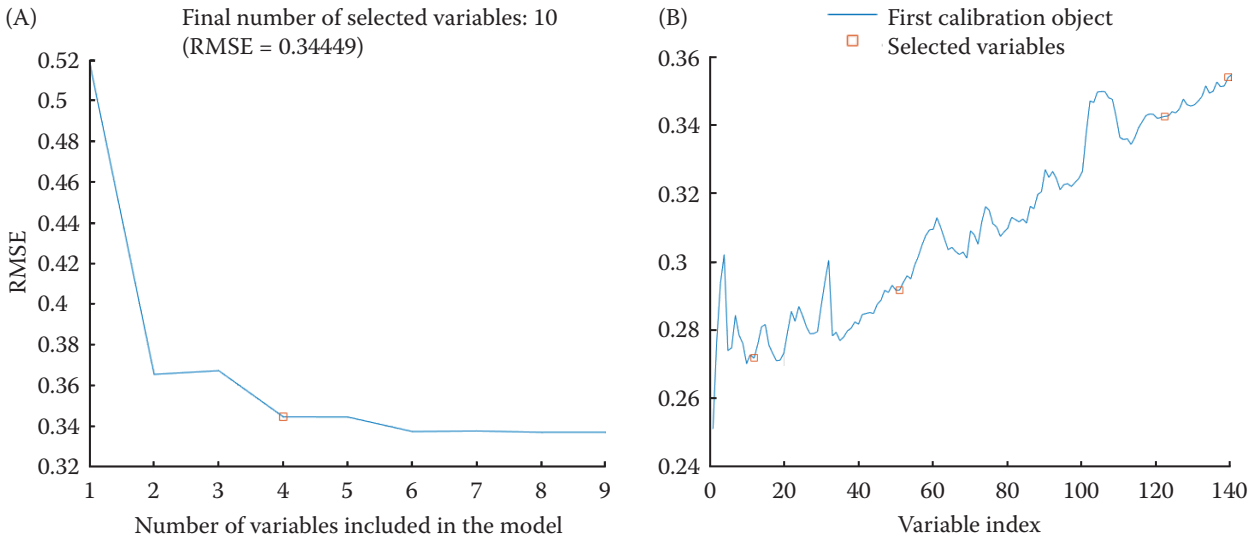

Figure 9. Key variables selection by CARS-SPA: (A) variation of RMSE, (B) key variables extracted

CARS – competitive adaptive reweighted sampling; SPA – successive projections algorithm; RMSE – root mean square error

Secondary variable selection

The number of variables affected the computation amount and complexity of the model. Therefore, based on UVE and CARS algorithms, SPA was taken as the secondary variable selection method to screen variables again.

CARS-SPA. As shown in Figure 9, the four wavelength variables were selected finally: 557.51 nm, 664.09 nm, 885.34 nm, and 946.84 nm. After the second selection, the number of variables selected finally was reduced from 140 to 4.

UVE-SPA. The process of extracting feature parameters by UVE-SPA is shown in Figure 10. Finally, the four wavelength variables were selected: 454.57 nm, 484.98 nm, 542.12 nm, and 937.1 nm. After the second selection, the number of variables selected finally was reduced from 356 to 4.

UVE-CARS. Since the number of variables screened by UVE was greater than the number of variables screened by CARS, indicating that CARS was better than UVE in screening selenium-containing information on millet samples, the CARS algorithm was

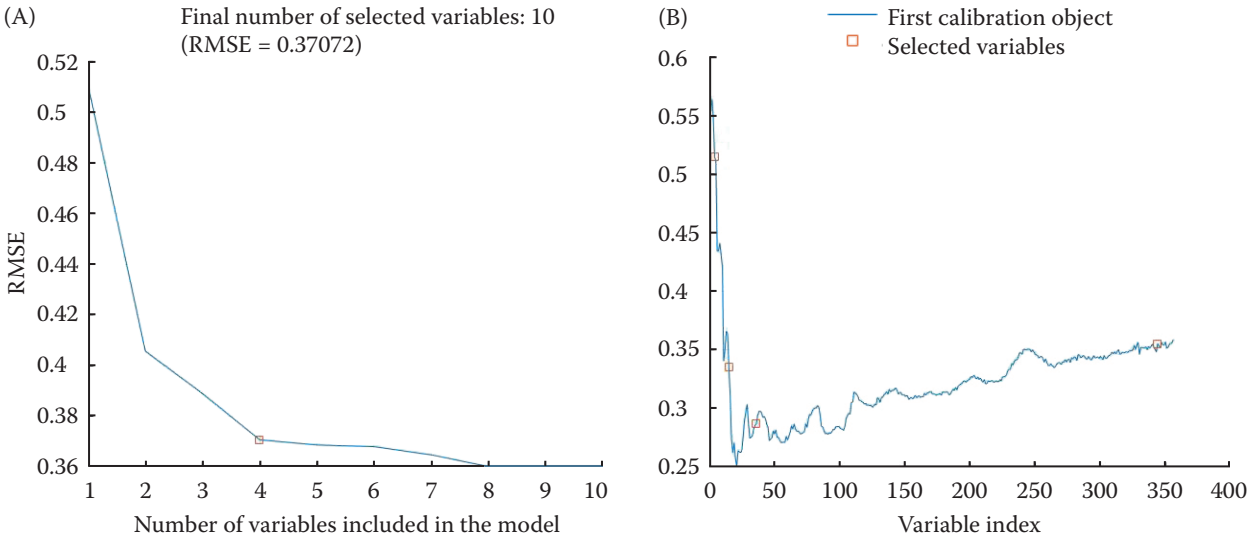

Figure 10. Key variables selection by UVE-SPA: (A) variation of RMSE, (B) key variables extracted

SPA – successive projections algorithm; UVE – uninformative variable elimination; RMSE – root mean square error

https://doi.org/10.17221/129/2022-CJFS

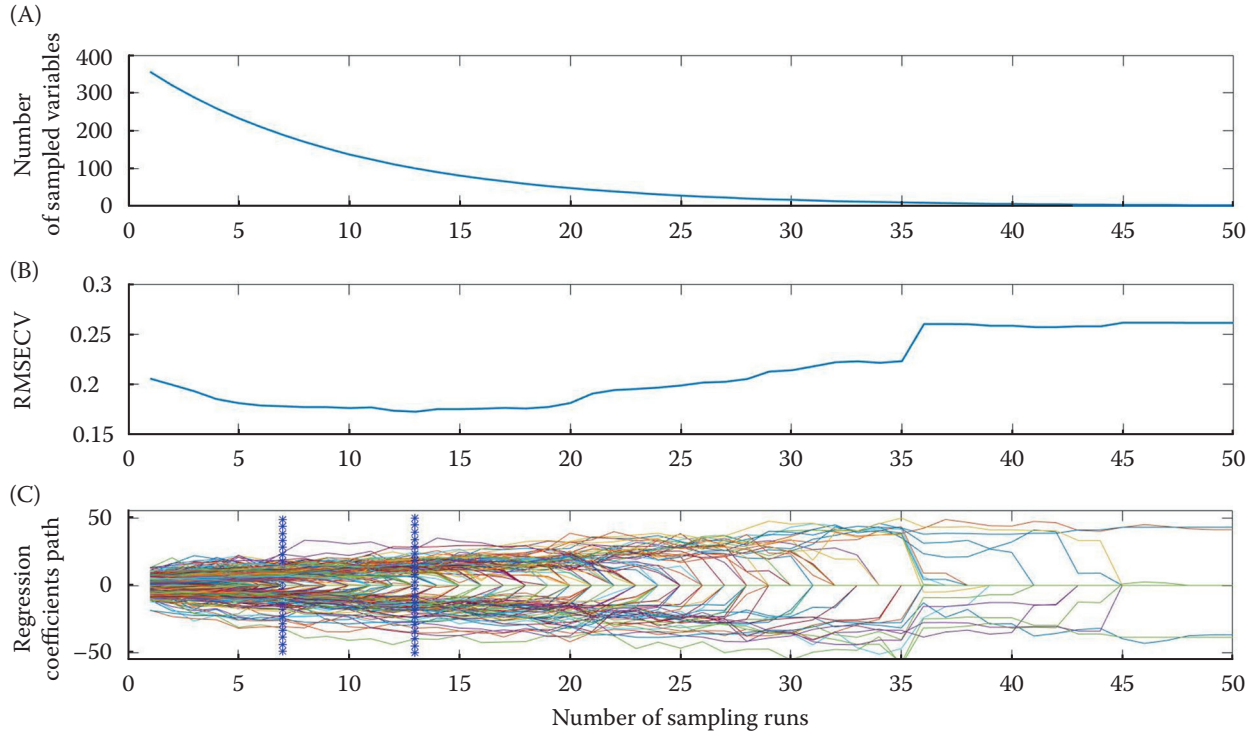

Figure 11. Key variables selection by UVE-CARS: (A) changes in the number of wavelength variables, (B) variation of RMSECV, (C) path of variables regression coefficients

CARS – competitive adaptive reweighted sampling; RMSECV – root mean square error of cross validation; UVE – uninformative variable elimination

used to screen the variables after UVE, and the key variables after secondary selection are shown in Figure 11. Finally, 100 variables between 522.68 nm and 945.76 nm were selected after 13 samplings by the CARS algorithm.

The support vector machines classification model was established. The radial basis function (RBF) was selected, and the samples were randomly divided into ten groups for cross-validation. The effects of *c* and *g* parameters on the accuracy are compared in Table 2. It was found that the accuracy rate gradually increased with the increase of parameters. When *c* was 100 and *g* was 10, the accuracy reached the maximum value of 100%, and then the accuracy began to decrease. Therefore, the best parameter combination was selected that *c* was 100 and *g* was 10.

Table 2. The accuracy of different combinations of c and g (%)

g	*c*				
	0.01	0.1	1	10	100
0.01	72.50	72.50	74.31	86.67	88.47
0.1	72.78	74.86	86.81	87.92	97.08
1	76.67	87.50	88.47	97.36	99.44
10	87.92	87.78	96.25	99.31	100
100	89.31	91.94	99.58	99.86	99.86

g – kernel function parameter; *c* – penalty coefficient

Based on SPA, CARS, UVE, CARS-SPA, UVE-SPA, and UVE-CARS characteristic wavelength extraction methods, SVM predictive analysis models were established respectively, while the parameter *c* was 100 and *g* was 10. The accuracy of the training set and test set was used as the evaluation index to evaluate whether the model was suitable for the identification of selenium-enriched millet. The results are shown in Table 3.

After variables optimization, the models were greatly simplified. After SPA, CARS, UVE, CARS-SPA, UVE-SPA and UVE-CARS, the number of SVM model variables was 5, 140, 356, 4, 4 and 100, respectively, and the number of spectral variables was reduced 98.95, 70.71, 25.52, 99.16, 99.16 and 79.08% compared with that before treatments. Therefore, CARS-SPA and UVE-SPA had the strongest simplification ability for the models. Compared with CARS and UVE, SPA had the strongest ability to simplify the number of characteristic wavelengths. The SPA algorithm for feature variables screening could be used to significantly reduce the number

https://doi.org/10.17221/129/2022-CJFS

Table 3. The results of different models

Models	Number of wavelengths	Accuracy (%)	
		training set	test set
SPA-SVM	5	99.79	98.33
CARS-SVM	140	100	99.58
UVE-SVM	356	100	99.17
CARS-SPA-SVM	4	100	99.58
UVE-SPA-SVM	4	100	98.75
UVE-CARSSVM	100	100	99.58

SPA – successive projections algorithm; CARS – competitive adaptive reweighted sampling; UVE – uninformative variable elimination; SVM – support vector machine

of variables, providing a reference for the model simplification.

After the optimization of variables, the accuracy of the model was improved, and the classification ability was enhanced. Firstly, it was easy to see from Table 3 that the accuracy of SPA-SVM was the lowest, the accuracy of the training set was 99.79%, and the accuracy of the test set was only 98.33%. While the accuracy of CARS-SPA-SVM and UVE-SPA-SVM models was higher than that of SPA-SVM in the training set, and the accuracy was increased by 1.25% and 0.42% in the test set, respectively. It showed that the secondary selection using SPA could effectively enhance the accuracy of models that used SPA in primary variable selection. Secondly, the accuracy of models was 100% in the training set, and the accuracy was 99.58% in the test set that used CARS for variables selected, including primary selection and secondary selection. It could be seen that the models using CARS had the best accuracy. It showed that the CARS algorithm retained the characteristic wavelengths which were strongly correlated with the selenium element information on millet in the variables. Finally, although the number of variables screened by CARS was 140, the number of variables screened by UVE was 356, the identification accuracy of CARS-SVM was better than that of UVE-SVM. It could be inferred that the accuracy in the test set decreased because too many useless variables interfered with the calculation of the models. Because redundant variables would increase nonlinear correlation factors, which had a certain restrictive effect on the accuracy and stability of the prediction model.

In summary, using the CARS algorithm for the primary screening and the SPA algorithm for the secondary selection for variables not only retained the characteristic wavelengths that were strongly related to the selenium element in millet, but also greatly reduced the number of spectral variables. It could be inferred that the combination of the SPA algorithm and the CARS algorithm had obvious advantages in feature variables selection. So, CARS-SPA-SVM was considered the best model for discriminating the selenium-enriched information on millet. The identification results are shown in Figure 12.

Finally, the sensitive wavelengths with the highest contribution to the model were selected after variables optimization. Because the number of variables screened by UVE algorithm was the largest, CARS was

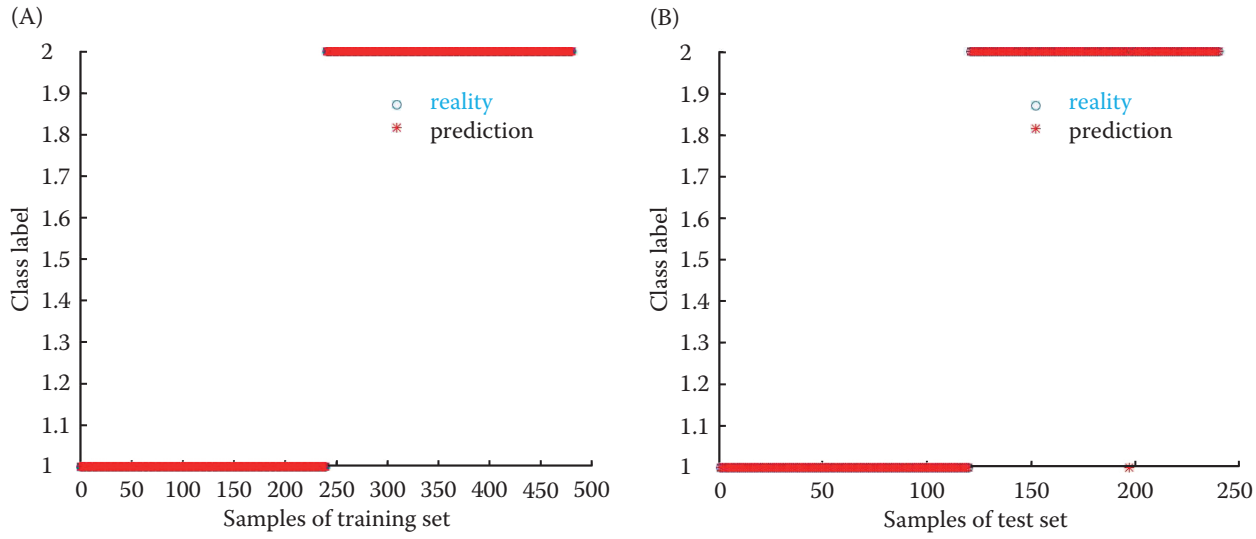

Figure 12. Identification results of CARS-SPA-SVM model: (A) identification result of the training set, (B) identification result of the test set

CARS – competitive adaptive reweighted sampling; SPA – successive projections algorithm; SVM – support vector machine

https://doi.org/10.17221/129/2022-CJFS

the second, and SPA was the least. Therefore, the variables screened by SPA and CARS were compared with the variables screened by UVE respectively to obtain wavelength variables sensitive to the information on selenium enrichment. After UVE processing, the number of the remaining variables without irrelevant information was 356 between 450.53 nm and 950.09 nm. After SPA processing, the number of the remaining variables without irrelevant information was 5, and a total of three variables, which were 454.57 nm, 484.98 nm, and 937.1 nm, respectively, coincided with selection variables by UVE. After CARS processing, the number of the remaining variables without irrelevant information was 140 between 500.24 nm and 947.92 nm, and one of them coincided with selection variables by UVE, which was 885.34 nm. Therefore, it was considered that 454.57 nm, 484.98 nm, 885.34 nm, and 937.1 nm were the sensitive wavelengths of selenium information in millet samples. The results indicated that these sensitive wavelengths had a more significant effect on the molecular structure of Se-enriched millet.

CONCLUSION

In this study, image data of selenium-enriched millet and ordinary millet were collected based on the hyperspectral image acquisition system, and then the ROIs were extracted. The average spectral value in the ROIs was taken as a spectral record of the samples, and a total of 720 spectral curves were extracted finally. The SG smoothing algorithm was performed on the original spectra, then SPA, CARS, UVE, CARS-SPA, UVE-SPA, and UVE-CARS were used to extract characteristic wavelengths for the denoised spectral. The number of spectral variables was reduced 98.95, 70.71, 25.52, 99.16, 99.16, and 79.08% compared with that before treatment. For CARS and UVE, SPA had the strongest ability to simplify the model. To identify selenium-enriched millet, SVM models were established with the proposed characteristic wavelengths. The discrimination accuracy was used as the evaluation criterion. The discriminant accuracy of the training set was 99.79% in SPA-SVM model, while it was 100% in CARS-SVM, UVE-SVM, CARS-SPA-SVM, UVE-SPA-SVM model and UVE-CARS-SVM model. The discriminant accuracy of the test set was 98.33% in the SPA-SVM model, while it was 99.17% in UVE-SVM model, 98.75% in UVE-SPA-SVM model, 99.58% in CARS-SVM, CARS-SPA-SVM and UVE-CARS-SVM model. The results showed that the models using CARS to select variables including primary selection and secondary selection had the best accuracy. And the combination of the SPA algorithm and the CARS algorithm had obvious advantages in feature variables selection. Therefore, the CARS-SPA-SVM model was the best model for identifying selenium-enriched millet.

REFERENCES

Araújo M.C.U., Saldanha T.C.B., Galvão R.K.H., Yoneyama T., Chame H.C., Visani V. (2001): The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems, 57: 65–73.

Cao C.L., Wang T.L., Gao M.F., Li Y., Li D.D., Zhang H.J. (2021): Hyperspectral inversion of nitrogen content in maize leaves based on different dimensionality reduction algorithms. Computers and Electronics in Agriculture, 190: 106461.

Garcia-Martin J.F., Badaró A.T., Barbin D.F., Álvarez-Mateos P. (2020): Identification of copper in stems and roots of *jatropha curcas* L. by hyperspectral imaging. Processes, 8: 823.

Guo W.C., Gu J.S., Liu D.Y., Shang L. (2016): Peach variety identification using near-infrared diffuse reflectance spectroscopy. Computers and Electronics in Agriculture, 123: 297–303.

Guo Z.M., Wang M.M., Agyekum A.A., Wu J.Z., Chen Q.S., Zuo M., El-Seedi H.R., Tao F.F., Shi J.Y., Ouyang Q., Zou X.B. (2020): Quantitative detection of apple watercore and soluble solids content by near infrared transmittance spectroscopy. Journal of Food Engineering, 279: 109955.

Ji H.Y., Ren Z.Q., Rao Z.H. (2019): Discriminant analysis of millet from different origins based on hyperspectral imaging technology. Spectroscopy and Spectral Analysis, 39: 2271–2277.

Khan I.H., Liu H.Y., Li W., Cao A.Z., Wang X., Liu H.Y., Cheng T., Tian Y.C., Zhu Y., Cao W.X., Yao X. (2021): Early detection of powdery mildew disease and accurate quantification of its severity using hyperspectral images in wheat. Remote Sensing, 13: 23612.

Li H.D., Liang Y.Z., Xu Q.S., Cao D.S. (2009): Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Analytica Chimica Acta, 648: 77–84.

Liu Y., Qiao F., Wang S. W., Wang R. T., Xu L. L. (2021): Application of hyperspectral imaging technology for rapid identification of ruditapes philippinarum contaminated by heavy metals. RSC Advances, 11: 33939–33951.

Liang K.H., Lu L.G., Zhu D.Z., Zhu H. (2020): Research progress of selenium content and selenium enrichment in millet of China. Journal of Chinese Institute of Food Science and Technology, 20: 337–343.

https://doi.org/10.17221/129/2022-CJFS

Mojadadi A., Au A., Salah W., Witting P., Ahmad G. (2021): Role for selenium in metabolic homeostasis and human reproduction. Nutrients, 13: 3256.

Mu T.T., Zhang F.Y., Li Z.H., Liu, Z., Tian G. (2018): Effects of spraying selenium on selenium content, conversion rate of organic selenium and quality of foxtail millet in different stages. Acta Agriculturae Boreali-Sinica, 33: 193–198.

Rayman M.P. (2000): The importance of selenium to human health. The Lancet, 356: 233–241.

Si G.Z., Yue X., Lyu Z., Yang H., Wang S.N., Li F.J., Song S.Z., Wen C.L., Tan Y. (2020): Rapid classification of northeast rice varieties based on hyperspectral imagery. Journal of Terahertz Science and Electronic Information Technology, 18: 687–691.

Sun J., Jin X.M., Mao H.P., Wu X.H., Yang N. (2014): Application of hyperspectral imaging technology for detecting adulterate rice. Transactions of the Chinese Society of Agricultural Engineering, 30: 301–307.

Wang H., Yu G. H., Hou Y., Hou S. Y., Han Y. H., Li H. Y., Xing G. F. (2021): Establishment and evaluation of near-infrared prediction model for selenium content in millet. Journal of China Agricultural University, 26: 157–163.

Wang J., Cheng J.J., Liu Y., Chang J.L., Wang Z.H. (2021): Identification of rice variety based on hyperspectral imaging technology. Journal of Agricultural Science and Technology, 23: 121–128.

Wang X., Li Z., Zheng D., Wang W. (2020): Nondestructive identification of millet varieties using hyperspectral imaging technology. Journal of Applied Spectroscopy, 87: 54–61.

Zhao C.X., Gao J., Fu Z.B., Chang Q.F. (2021): Research progress on the detection methods of total selenium in food. Journal of Food Safety and Quality, 12: 1653–1661.

Zou X.B., Zhao J.W., Malcolm J.W.P., Mel H., Mao H.P. (2010): Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta, 667: 14–32.

Yu Q., Zhang X., Si X.Z., Suo Y.Y., Li L., Mao J.W. (2018): Research progress of selenium in crops. Journal of Shanxi Agricultural Sciences, 46: 2122–2126.

Received: July 18, 2022
Accepted: November 15, 2022
Published online: December 19, 2022