Original Article
Food Science and Technology

ISSN 0101-2061 (Print)
ISSN 1678-457X (Online)

DOI: https://doi.org/10.1590/fst.38922

Characteristic wavelengths selection of rice spectrum based on adaptive sliding window permutation entropy

Sen YANG[1], Houqing ZHANG[1]*, Wenmin FAN[1]

Abstract

Due to the redundancy of rice spectral wavelengths and the strong correlation between adjacent wavelengths, the modeling classification accuracy based on traditional characteristic wavelengths selection methods is insufficient. Thus, a rice spectral characteristic wavelengths selection method based on adaptive sliding window permutation entropy (ASW-PE) was proposed in this paper. Firstly, the ASW-PE algorithm is constructed by combining the adaptive sliding window (ASW) method and permutation entropy (PE) method. Then, for the spectral data of rice varieties WC, XS, YS and YG, based on ASW-PE, sliding window permutation entropy (SW-PE), analysis of variance (ANOVA), competitive adaptive reweighted sampling (CARS) and successive projections algorithm (SPA) to carry out the characteristic wavelengths selection experiment, and evaluated the computational efficiency of the five algorithms from the perspective of time complexity. Finally, a partial least squares (PLS) rice varieties classification model was established based on the spectral characteristic wavelengths selected by the above algorithms, and the characteristic selection performance of the five algorithms was evaluated with the classification accuracy. Experimental results show that the ASW-PE algorithm has a speed advantage in selecting characteristic wavelengths for large sample spectral data. Compared with SW-PE, ANOVA, CARS and SPA algorithms, the accuracy of modeling classification based on ASW-PE method is improved by 5.6%, 22.6%, 8.6% and 15.2%, respectively.

Keywords: rice varieties classification; infrared spectrum; characteristic wavelengths selection; adaptive sliding window permutation entropy.

Practical Application: Characteristic wavelengths selection is a key step in rice varieties classification or origin traceability using spectroscopy, which can simplify the model and improve the classification accuracy of the model.

1 Introduction

The geographical symbol rice is favored by the market because of its excellent quality. Due to its limited output and high product value, adulteration is rampant in the market, which has triggered extensive research and attention by researchers from all walks of life on the classification and identification of geographical symbol rice varieties (Lee et al., 2020; Mongkontanawat et al., 2022). In the classification research of rice varieties, near infrared spectroscopy has been widely used due to its advantages of rapidity, pollution-free, and simple operation (Murtaza et al., 2022; Munarko et al., 2022).

Spectral characteristic wavelengths selection is a key step in spectral analysis. Selecting characteristic wavelengths for original spectral data can not only simplify the model, but also improve the model performance (Fan et al., 2019). Whether the selected characteristic wavelengths can best characterize the spectral differences of different rice varieties directly determines the classification accuracy of rice varieties. Currently, the commonly used methods for selecting spectral characteristic wavelengths mainly include ANOVA algorithm (Qian et al., 2018; Song et al., 2017), CARS algorithm (Diallo et al., 2019; Weng et al., 2020) and SPA algorithm (Weng et al., 2020; Tian et al., 2020). When Qian et al. (2018) and Song et al. (2017) used near infrared spectroscopy to study rice origin traceability, they determined characteristic spectral wavelengths based on spectral absorption peak and ANOVA algorithm, which improved the accuracy of the model to a certain extent. Diallo et al. (2019) used near infrared spectroscopy to study the prediction model of lignocellulose and organic elements in rice straw. Compared with PLS algorithm, modeling based on CARS-PLS algorithm has better prediction effect, which verified the effectiveness of CARS algorithm in selecting characteristic wavelengths. Weng et al. (2020) used hyperspectral image technology to study the nondestructive identification method of famous rice, and established a rice varieties classification model based on the spectral characteristic wavelengths selected by SPA algorithm and CARS algorithm. Compared with full spectrum modeling, the accuracy of rice varieties identification was further improved. Tian et al. (2020) used Raman spectroscopy to study the classification of rice production areas, they established a classification model of rice production areas based on the effective wavelengths in the rice Raman spectrum selected by the SPA algorithm, and realized the accurate classification of rice in adjacent production areas. Compared with the original spectral modeling (Jin et al., 2022; Mishra et al., 2021; Onmankhong et al., 2022; Wang & Tan, 2021), the above algorithms can simplify the model input and improve its classification accuracy. However, due to the common problems of wavelength variable redundancy and strong correlation between adjacent wavelengths in the near infrared spectrum, the classification accuracy of the classification

Received 10 Mar., 2022
Accepted 02 May, 2022
[1] *College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, China*
*Corresponding author: zhq1695641968@163.com

model established based on the above traditional characteristic wavelengths selection methods still need to be improved.

To solve the above problems, an ASW-PE algorithm was proposed to select the spectral characteristic wavelengths of different rice varieties. First, the rice near infrared spectral data was segmented using adaptive sliding window, and the permutation entropy algorithm was used to detect the mutation of the spectral data within the sliding window, so as to maximize the retention of spectral characteristic information. Then, based on the ASW-PE, SW-PE, ANOVA, CARS and SPA algorithms, the characteristic wavelengths selection experiments were carried out for the spectra of four rice varieties respectively, and the computational efficiency of the five algorithms was evaluated from the perspective of time complexity. Finally, rice varieties classification model was established based on the characteristic wavelengths selected by the five algorithms, and the effectiveness of different algorithms in rice varieties classification was evaluated from the perspective of classification accuracy.

2 Materials and methods

2.1 Spectral data collection

In this experiment, Tianjin ENERGY spectrum iCAN9 Fourier transform infrared spectrometer and PIKE Technology PN044-60XX diffuse reflectance accessory was used for spectral data collection. The spectral measurement range of the instrument is 4000 cm^{-1} ~ 7800 cm^{-1}, the resolution is 4 cm^{-1}, and the scanning frequency is 32 times per minute. To ensure the regional authenticity of the samples, four specific varieties were collected in different geographical indication areas, including WC rice, XS rice, YS rice and YG rice. The obtained rice varieties were evenly ground into rice flour, which were put into different grinding-mouth bottles according to the varieties and stored at room temperature for one week to ensure the stability and repeatability of spectral data collection. During spectrum collection, 10 g of a single sample were collected for three times and the average value was taken. The number of spectral wavelength points of the sample was 1972. The training set and the validation set were divided according to the ratio of 1 : 1, with a total of 320 samples. The original spectral curves of the four rice varieties are shown in Figure 1.

2.2 Spectral characteristic wavelengths selection method based on ASW-PE

The PE analysis of time series can show the similarity and difference between sequences (Araujo et al., 2019; Feng et al., 2021), and then accurately detect the randomness and dynamic variation behavior of time series. In order to make the detection accuracy of the PE algorithm more accurate, this paper integrates the adaptive sliding window segmentation method on the basis of the PE algorithm to construct the ASW-PE method, selects the time series through the sliding window, and establishes the window forgetting factor (μ) based on the similarity of the old and new window data information. The window forgetting factor is used to adaptively update the subsequence window size, and then the PE value is calculated for each subsequence. The sliding window model is a dynamic data block matrix in which the observation samples are arranged sequentially, and the *KTH* data window is expressed as (Equation 1):

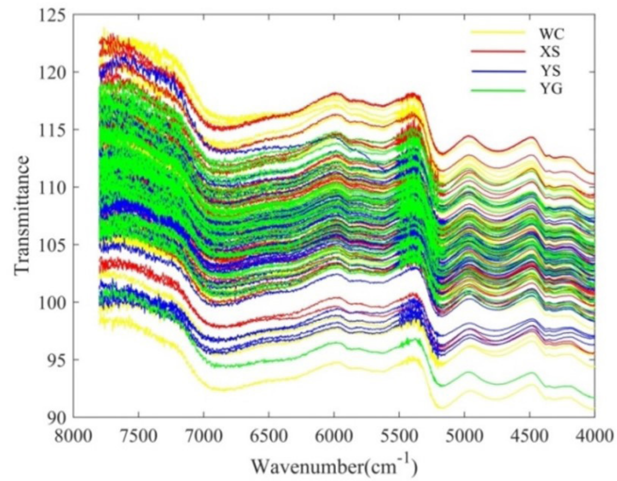

Figure 1. Original spectral curves of four rice varieties.

$$X_k = \left[x_{k+1}^{T}, x_{k+2}^{T}, \cdots x_{k+L}^{T}\right]^{T} \tag{1}$$

Where L is the window width, and the statistical characteristics of the window are described by standard deviation matrix Σ_k, mean value m_k and Gram matrix G_k (Equations 2-3):

$$\sum\nolimits_k = Diag\left(\sigma_k^1, \sigma_k^2, \cdots, \sigma_k^L\right), \sigma_k^i = \sqrt{\left(x_{ij} - m_j\right)^2 / L} \tag{2}$$

$$G_k = X_k^T X_k \tag{3}$$

Where X_k is the normalized data block matrix.

The Gram matrix of the $k + 1$ new data window is defined as G_H, and the mixed Gram matrix of the old and new data window is established (Equation 4):

$$G_\Omega = \left[G_k \, G_H\right]^{T} \tag{4}$$

The eigenvalue decomposition of G_Ω is carried out, and its eigenvalue diagonal matrix and eigenvector are Λ_Ω and P_Ω, respectively, and the transformation matrix is defined (Equation 5):

$$P = P_\Omega \Lambda_\Omega^{-\frac{1}{2}} \tag{5}$$

G_k and G_H are transformed by transformation matrix P, $G_k' = P^T G_k P$ and $G_H' = P^T G_H P$ are obtained, from which it can be inferred that the transformed matrix meets the following conditions (Equation 6):

$$P^T G_\Omega P = G_k' + G_H' = I \tag{6}$$

Υ_i^K and Υ_i^H are obtained by eigenvalue decomposition of G_k' and G_H', and the value of $\Upsilon_i^K + \Upsilon_i^H = 1$ is satisfied. Therefore, the closer the value of Υ_i^K and Υ_i^H is to 0.5, the higher the similarity of G_k and G_H is. Forgetting factor μ was used to distinguish the similarity of two process data (Equation 7):

$$\mu = \frac{m - 4\sum_{i=1}^{m}\left(\Upsilon_i^k - 0.5\right)}{m} \tag{7}$$

By sliding the window and setting the forgotten number $l = \mu L$ to carry out adaptive discarding of the old window data, the $k + 1$ data window can be obtained as (Equation 8):

$$X_{k+1} = \left[x_{k+l+1}^{T}, x_{k+l+2}^{T}, \cdots, x_{k+l+H}^{T}\right]^{T} \tag{8}$$

Where H is the sliding step length, l is the number of self-adaptive discarded samples.

Phase space reconstruction was carried out on the time series after adaptive sliding window segmentation, and the new time series was obtained as (Equation 9):

$$\begin{bmatrix} X(1) = \{x(1), x(1+\tau), \cdots, x(1+(m-1)\tau)\} \\ \vdots \\ X(i) = \{x(i), x(i+\tau), \cdots, x(i+(m-1)\tau)\} \\ \vdots \\ X(N) = \{x(N), x(N+\tau), \cdots, x(N+(m-1)\tau)\} \end{bmatrix} \tag{9}$$

Where m is the embedding dimension and τ is the delay time.

By arranging the reconstructed components of time series in ascending order, we can obtain (Equation 10):

$$\left[x(1+(j_1-1)\tau) \le x(1+(j_2-1)\tau) \le \cdots x(1+(j_m-1)\tau)\right] \tag{10}$$

$j_1, j_2, \cdots, j_m$ represent the index positions of each element.

If equal sequences are observed in the reconstructed component, they are sorted according to the index positions, $x(1 + (j_a - 1)\tau) \le x(1 + (j_b - 1)\tau)$, (a < b). The arrangement of m vectors is m!, such as $P = \{p_j, j = 1, 2, \cdots, m!\}$, and use this to estimate Shannon entropy (Equation 11):

$$S_P = -\sum_{j=1}^{m!} P_j \ln P_j \tag{11}$$

The permutation entropy indicates the randomness and predictability of the signal. H_p is obtained through the following data transformation (Equation 12):

$$H_P = \frac{S_P}{S_{max}} = \frac{S_P}{\ln(m!)} \tag{12}$$

Where S_{max} is the value obtained from an equal probability sequence mode, and the value range of H_p is $0 \le H_p \le 1$, the change of H_p reflects and magnifies the fine details of time series.

The smaller the value is, the more regular the time series is; otherwise, the more random the time series is. The spectral characteristic wavelengths selection process of ASW-PE algorithm is shown in Figure 2.

Based on ASW-PE method of rice varieties classification spectral characteristic wavelengths selection procedure is as follows:

1. First, initialize the parameters of the sliding window, and then perform adaptive sliding segmentation on the time series based on the window forgetting factor to obtain subsequences. The time series of this experiment are the preprocessed data of four kinds of rice spectra.
2. Perform phase space reconstruction on the subsequence, and the reconstructed components are arranged in ascending order.
3. Calculate the relative frequency ($m!$) of the subscript order of each component as the probability of the component.
4. The sum of the information entropy of the spectral subsequence components of the sample is the PE value of the rice sample, select the optimal window and locate the characteristic wavelength according to the PE ratio (PE ratio > ± 1.5). It should be noted that the PE ratio is the value of the spectrum of a single rice variety divided by the average of the spectra of four rice varieties.

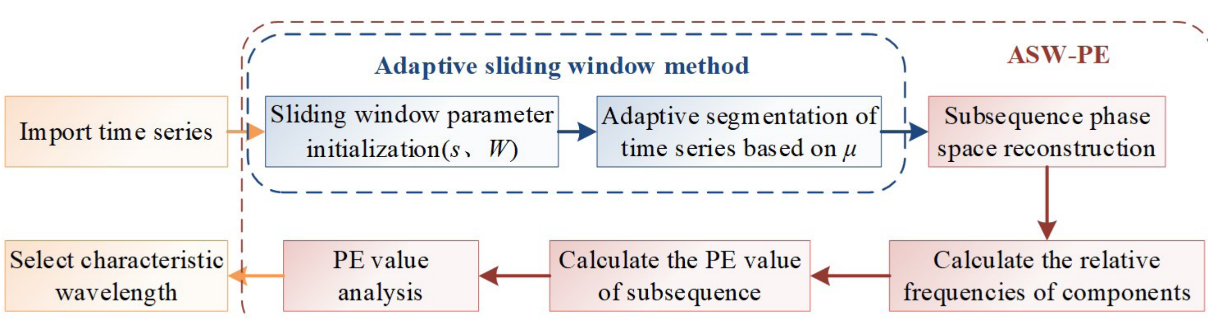

Figure 2. Spectral characteristic wavelength selection process of the ASW-PE algorithm.

3 Results and discussion

In this chapter, the full spectral data of the four kinds of rice in Figure 1 were preprocessed by standard normal variable transformation, and the ASW-PE, SW-PE, ANOVA, CARS and SPA algorithms were used to select the spectral characteristic wavelengths, and the time complexity as an index to compare the computational efficiency of different algorithms. Based on the spectral characteristic wavelengths selected by the five algorithms, a PLS rice varieties classification model was established, and the characteristic wavelengths selection performance of different algorithms was evaluated from the perspective of model classification accuracy.

3.1 Results of rice spectral characteristic wavelengths selection based on different methods

Characteristic wavelengths selection based on ASW-PE algorithm

In the ASW-PE algorithm, the initialization of sliding window parameters (window width W and sliding step size s) and the setting of arrangement entropy parameters need to be considered. Among them, the selection rule of sliding window width is $W > 5\ m!$. This paper sets the sliding window parameters to $W = 130$ and $s = 1$, the PE algorithm parameters to $m = 4$ and $\tau = 1$, and uses the PE ratio to analyze the differences between spectral sequences. The PE values and PE ratios of four kinds of rice spectra based on ASW-PE and S-PE algorithms are shown in Figure 3.

It can be seen from Figure 3a and 3c that the PE values of the near-infrared spectra of the four rice varieties still have a lot of overlap. However, the PE ratio in Figure 3b and 3d can more intuitively identify and analyze the PE values of rice varieties. The more obvious the peak of PE ratio, the greater the difference of PE value between different varieties. According to the principle of "least information and maximum features", the window sequence with the absolute value of the PE ratio peak value greater than 1.5 was selected. Finally, the ASW-PE algorithm selects 3 window sequences through adaptive sliding window segmentation, with a total of 300 characteristic wavelength points; the SW-PE algorithm selects 2 window sequences through sliding segmentation, with a total of 260 characteristic wavelength points. The corresponding spectral characteristic wavelengths are shown in Table 1.

Characteristic wavelengths selection based on other methods

The ANOVA algorithm can reflect the degree of influence of each factor on the index. For the selection of the characteristic wavelengths of rice near-infrared spectrum, since the rice varieties do not constitute an influencing factor, only the difference between the data is analyzed from the spectral wavelengths, so the one-factor variance analysis is selected. Through the variance multiple comparison analysis of the near-infrared spectral data of four kinds of rice, 103 characteristic wavelength points

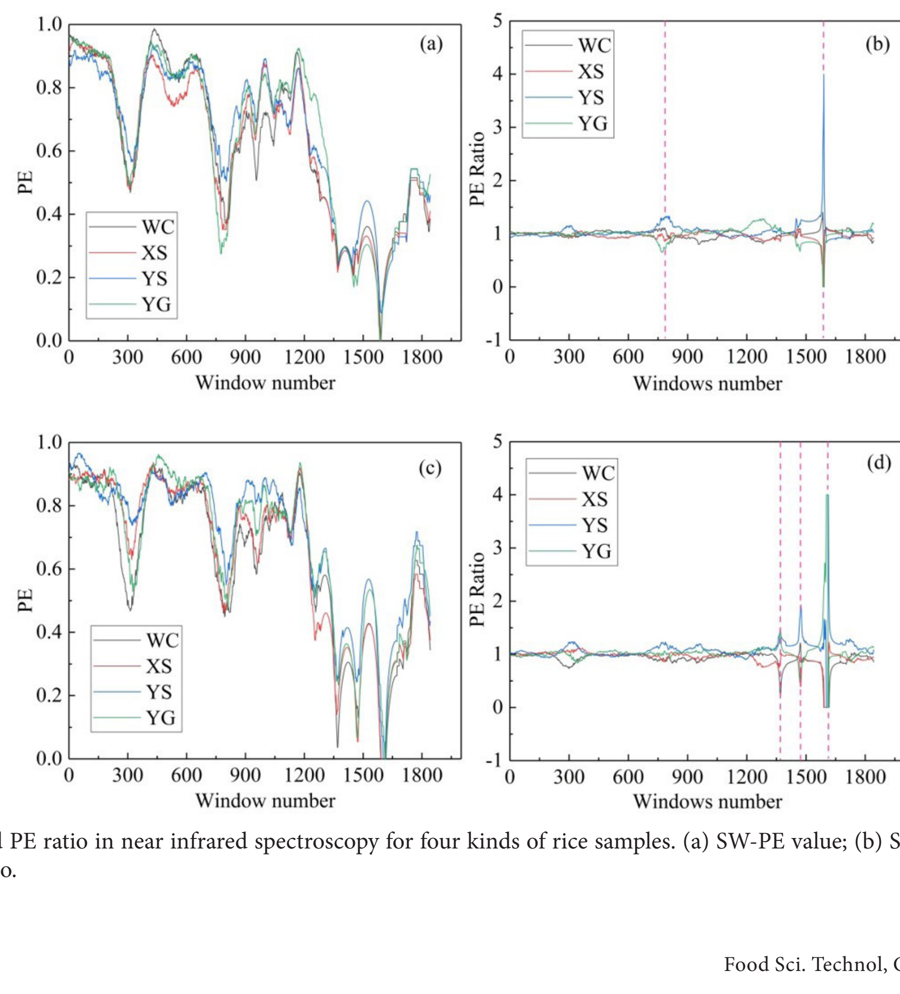

Figure 3. PE value and PE ratio in near infrared spectroscopy for four kinds of rice samples. (a) SW-PE value; (b) SW-PE ratio; (c) ASW-PE value; (d) ASW-PE ratio.

were finally screened out, mainly distributed in the range of 4904 cm^{-1} ~ 5102 cm^{-1}.

In the CARS algorithm, the ten-fold cross-validation method is used, the number of Monte Carlo sampling is set to 150 times, and the minimum mean square error (RMSECV) value of the PLS model is used to determine the spectral characteristic wavelength points. For the near-infrared spectral data of four rice varieties, the characteristic wavelengths selection process based on the CARS algorithm is shown in Figure 4. It can be seen from Figure 4a that the number of characteristic wavelengths screened gradually decreases with the increase of sampling times, and approaches 0 when the sampling times is 90. In addition, as can be seen from Figure 4b that the RMSECV value decreases slowly in the first 54 sampling intervals, indicating that the effect of eliminating useless information is better at this stage. Combining with Figure 4c, it can be seen that the minimum value of RMSECV is 0.194 when the number of sampling is 54. After 54 samplings, the value of RMSECV increases and the performance of the algorithm decreases. Finally, 157 characteristic wavelength points were selected by sampling, and the corresponding wavelengths were concentrated in the range of 4535 cm^{-1} ~ 5139 cm^{-1}.

The SPA algorithm uses a continuous projection strategy to screen the optimal variables, which can effectively eliminate the collinearity problem between variables. In the SPA algorithm, it is first necessary to determine the optimal Maximum, traverse the range of 20 to 120 at intervals of 20, and screen out the optimal Maximum value according to the root mean square error of prediction (RMSEP) of the PLS model. The parameter optimization results are shown in Table 2.

It can be seen from Table 2 that after the Maximum reaches 80, the RMSEP and the number of characteristic wavelengths tend to be stable. At this time, the RMSEP has a minimum value of 0.3528, and the corresponding number of characteristic wavelengths is 51. Therefore, the optimal value of Maximum is set to 80. The distribution of spectral characteristic wavelengths based on the SPA algorithm is shown in Figure 5. As can be seen from Figure 5, the 51 characteristic wavelength points selected by the SPA algorithm are mainly concentrated in the three wavelength ranges of 5352 cm^{-1} ~ 5548 cm^{-1}, 5033 cm^{-1} ~ 5158 cm^{-1} and 4277 cm^{-1} ~ 4372 cm^{-1}.

Table 1. Spectral characteristic wavelengths selected based on ASW-PE and SW-PE algorithms.

Method	Window number	Wavenumber (cm^{-1})
ASW-PE	1360, 1480, 1615	4967 ~ 5179, 4755 ~ 4948, 4495 ~ 4688
S-PE	765, 1608	6071 ~ 6327, 4451 ~ 4701

SW-PE: sliding window permutation entropy.

Time complexity analysis of five algorithms

Time complexity is an important index for evaluating algorithms, which can directly reflect the computational efficiency of data processing. The running times of the five algorithms are compared using the spectral data of different sample numbers as shown in Figure 6.

It can be seen from Figure 6 that at the starting point when the number of samples is 40, the running times from small to large are ANOVA, CARS, SW-PE, ASW-PE, and SPA. With

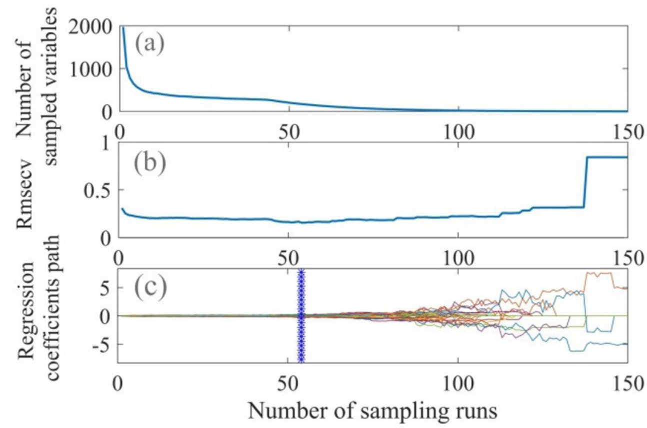

Figure 4. Selection process of characteristic wavelengths based on the CARS algorithm. (a) number of sampled variables; (b) RMSECV value; (c) regression coefficients path.

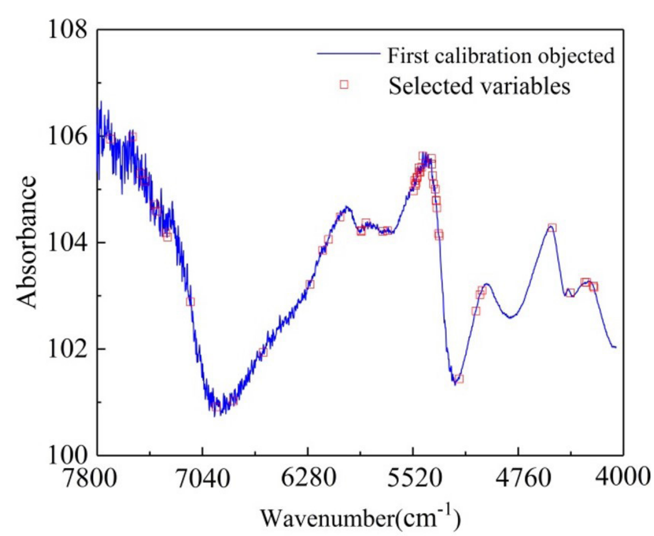

Figure 5. Spectral characteristic wavelengths distribution based on the SPA algorithm.

Table 2. Parameter optimization results of the SPA algorithm.

Maximum	20	40	60	80	100	120
RMSEP	0.4531	0.4021	0.3814	0.3528	0.3528	0.3528
Number of wavelengths	10	25	40	51	51	51

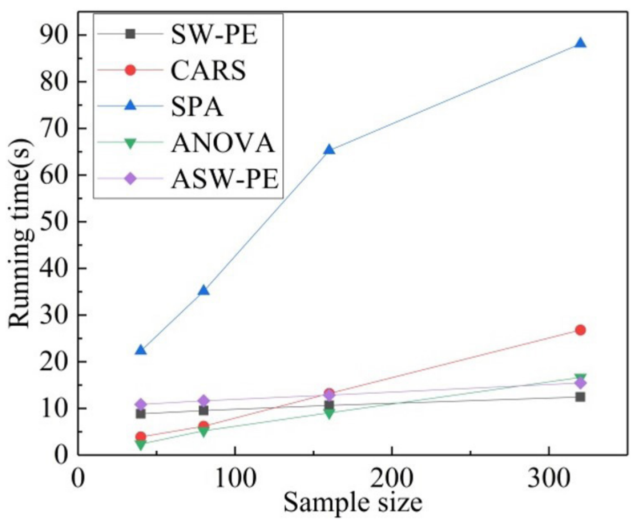

Figure 6. The running time of the five algorithms.

the increase of the number of samples, the running time of the five algorithms is also prolonged. The largest increase is SPA, followed by CARS, and then ANOVA. It is worth noting that ASW-PE and SW-PE are basically stable. When the sample size exceeds 280, the running time of ASW-PE and SW-PE is the shortest, and from the trend of algorithm running time, with the continuous increase of the sample size, the advantages of these two algorithms in terms of operation speed are more obvious. It can be concluded that the SW-PE and ASW-PE algorithms have more advantages in the selection of characteristic wavelengths of large sample near infrared spectral data.

3.2 Analysis of results of rice varieties classification experiment

In this paper, a rice varieties classification model was established based on the PLS method, and the lowest value of the classification accuracy of the four rice varieties was selected as the classification accuracy of the model. The modeling process is as follows:

1. Divide the experimental data into the training set and the validation set in a ratio of 1 : 1.
2. The samples in the training set were assigned the dummy variable Y_i as the reference value of varieties by using the fixed assignment method, in which the four types of rice WC, XS, YS and YG were set as 1, 2, 3 and 4, respectively.
3. Use the training set data to train PLS to obtain a classification model, and delineate the classification threshold range.
4. The classification model was tested using the validation set data to evaluate the classification accuracy of the rice samples.

The full spectrum data and the spectral characteristic wavelengths data obtained based on five characteristic wavelengths selection algorithms were used as modeling data to establish a rice varieties classification model. Then, the characteristic wavelengths selection performance of the five algorithms is characterized based on the model classification accuracy. The higher the classification accuracy, the better the characteristic wavelengths selection performance of the algorithm. The classification accuracy of rice varieties based on full spectrum data and five characteristic wavelengths selection algorithms is shown in Table 3. It can be seen from Table 3 that compared with the full spectrum modeling, the use of five characteristic wavelengths selection algorithms can not only simplify the input data of the classification model, but also improve the classification accuracy to different degrees. Among them, the classification accuracy obtained by the ASW-PE algorithm proposed in this paper is optimal. Compared with the modeling classification accuracy based on full spectrum data, the classification accuracy based on ASW-PE, SW-PE, CARS, SPA and ANOVA algorithms is improved by 65.2%, 56.5%, 52.2%, 43.5% and 34.8%, respectively. Comparing the five algorithms, the classification accuracy based on the ASW-PE algorithm is improved by 5.6%, 22.6%, 8.6% and 15.2% compared with the SW-PE, ANOVA, CARS and SPA algorithms, respectively. The superiority of the ASW-PE algorithm in the selection of characteristic wavelengths of near infrared spectra for rice variety classification was verified.

Table 3. Rice varieties classification accuracy based on full spectral data and five characteristic wavelengths selection algorithms.

Method	Number of wavelengths	Accuracy rate (%)
Full spectra data	1972	57.5
ASW-PE	300	95
SW-PE	260	90
ANOVA	103	77.5
CARS	157	87.5
SPA	51	82.5

4 Conclusions

To improve the classification accuracy of rice varieties from the perspective of spectral characteristic wavelengths selection, this paper proposed an ASW-PE algorithm to be applied to the selection of near infrared spectral characteristic wavelengths for rice varieties classification. Firstly, the adaptive sliding window segmentation method and the permutation entropy algorithm were combined to construct a spectral characteristic wavelengths selection algorithm based on ASW-PE; then the characteristic wavelengths of rice varieties were selected based on ASW-PE, SW-PE, ANOVA, CARS and SPA algorithms respectively, and analyze the time complexity of five algorithms. Finally, the rice varieties classification model was constructed based on the PLS algorithm, and the performance of the five algorithms in the selection of near infrared spectral characteristic wavelengths was compared. The experimental results showed that the classification accuracy of rice varieties using the ASW-PE algorithm was improved by 5.6%, 22.6%, 8.6% and 15.2% compared with the SW-PE, ANOVA, CARS and SPA algorithms respectively. In addition, it shows the advantage of computational efficiency when processing large sample data, which provides an effective means for improving the classification accuracy of rice varieties using near infrared spectroscopy.

References

Araujo, F. H. A., Bejan, L., Rosso, O. A., & Stosic, T. (2019). Permutation entropy and statistical complexity analysis of Brazilian agricultural commodities. *Entropy*, 21(12), 1220. http://dx.doi.org/10.3390/e21121220.

Diallo, A. A., Yang, Z. L., Shen, G. H., Ge, J. Y., Li, Z. C., & Han, L. J. (2019). Comparison and rapid prediction of lignocellulose and organic elements of a wide variety of rice straw based on near infrared spectroscopy. *International Journal of Agricultural and Biological Engineering*, 12(2), 166-172. http://dx.doi.org/10.25165/j.ijabe.20191202.4374.

Fan, L., Zhao, J., Xu, X., Liang, D., Yang, G., Feng, H., Yang, H., Wang, Y., Chen, G., & Wei, P. (2019). Hyperspectral-based estimation of leaf nitrogen content in corn using optimal selection of multiple spectral variables. *Sensors*, 19(13), 2898. http://dx.doi.org/10.3390/s19132898. PMid:31262053.

Feng, A., Li, H. X., Liu, Z. X., Luo, Y. J., Pu, H. B., Lin, B., & Liu, T. (2021). Research on a rice counting algorithm based on an improved MCNN and a density map. *Entropy*, 23(6), 721. http://dx.doi.org/10.3390/e23060721. PMid:34198797.

Jin, B., Zhang, C., Jia, L., Tang, Q., Gao, L., Zhao, G., & Qi, H. (2022). Identification of rice seed varieties based on near-infrared hyperspectral imaging technology combined with deep learning. *ACS Omega*, 7(6), 4735-4749. http://dx.doi.org/10.1021/acsomega.1c04102. PMid:35187294.

Lee, J. Y., Pavasopon, N., Napasintuwong, O., & Nayga, R. M. Jr. (2020). Consumers' valuation of geographical indication-labeled food: the case of hom mali rice in Bangkok*. *Asian Economic Journal*, 34(1), 79-96. http://dx.doi.org/10.1111/asej.12196.

Mishra, P., Angileri, M., & Woltering, E. (2021). Identifying the best rice physical form for non-destructive prediction of protein content utilising near-infrared spectroscopy to support digital phenotyping. *Infrared Physics & Technology*, 116, 103757. http://dx.doi.org/10.1016/j.infrared.2021.103757.

Mongkontanawat, N., Ueda, Y., & Yasuda, S. (2022). Increased total polyphenol content, antioxidant capacity and γ-aminobutyric acid content of roasted germinated native Thai black rice and its microstructure. *Food Science and Technology*, 42, e34521. http://dx.doi.org/10.1590/fst.34521.

Munarko, H., Sitanggang, A. B., Kusnandar, F., & Budijanto, S. (2022). Germination of five Indonesian brown rice: evaluation of antioxidant, bioactive compounds, fatty acids and pasting properties. *Food Science and Technology*, 42, e19721. http://dx.doi.org/10.1590/fst.19721.

Murtaza, G., Huma, N., Sharif, M. K., & Zia, M. A. (2022). Probing a best suited brown rice cultivar for the development of extrudates with special reference to physico-chemical, microstructure and sensory evaluation. *Food Science and Technology*, 42, e103521. http://dx.doi.org/10.1590/fst.103521.

Onmankhong, J., Ma, T., Inagaki, T., Sirisomboon, P., & Tsuchikawa, S. (2022). Cognitive spectroscopy for the classification of rice varieties: a comparison of machine learning and deep learning approaches in analysing long-wave near-infrared hyperspectral images of brown and milled samples. *Infrared Physics & Technology*, 123, 104100. http://dx.doi.org/10.1016/j.infrared.2022.104100.

Qian, L. L., Song, X. J., Zhang, D. J., Zhang, L. Y., Ruan, C. Q., & Lu, B. X. (2018). Tracing the geographical origin of Sanjiang and Wuchang rice grown in different years by near infrared spectroscopy. *Shipin Kexue*, 39(16), 321-327.

Song, X. J., Qian, L. L., Zhang, D. J., Wang, X. H., Yu, G., Yu, J. C., & Zhou, Y. (2017). Tracing the geographical origin of rice grown in different crop years based on diffuse reflectance Fourier transform near infrared spectroscopy. *Shipin Kexue*, 38(18), 286-291.

Tian, F. M., Tan, F., & Li, H. (2020). An rapid nondestructive testing method for distinguishing rice producing areas based on Raman spectroscopy and support vector machine. *Vibrational Spectroscopy*, 107, 103017. http://dx.doi.org/10.1016/j.vibspec.2019.103017.

Wang, Y., & Tan, F. (2021). Extraction and classification of origin characteristic peaks from rice Raman spectra by principal component analysis. *Vibrational Spectroscopy*, 114, 103249. http://dx.doi.org/10.1016/j.vibspec.2021.103249.

Weng, S. Z., Tang, P. P., Zhang, X. Y., Xu, C., Zheng, L., Huang, L. S., & Zhao, J. L. (2020). Non-destructive identification method of famous rice based on image and spectral features of hyperspectral imaging with convolutional neural network. *Guangpuxue Yu Guangpu Fenxi*, 40(9), 2826-2833.