Food Chemistry: X 13 (2022) 100199

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x

Analytical Methods

Polysaccharide prediction in Ganoderma lucidum fruiting body by hyperspectral imaging

Yu Liu [a,b], Yongbing Long [a,b], Houcheng Liu [c], Yubin Lan [a,b], Teng Long [a,b], Run Kuang [a,b], Yifan Wang [d], Jing Zhao [a,b,*]

[a] College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China
[b] National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology (NPAAC), Guangzhou, 510642, China
[c] College of Horticulture, South China Agricultural University, Guangzhou 510642, China
[d] Guangdong Yuewei Edible Fungi Technology Co., Guangzhou 510525, China

ARTICLE INFO

Keywords:
Hyperspectral imaging
Ganoderma lucidum
Polysaccharide
Nondestructive detection

ABSTRACT

Ganoderma lucidum is a traditional Chinese healthy food with many kinds of nutritious activities, and polysaccharide is one of its main active components. Ganoderma lucidum polysaccharide plays a vital role in improving human immunity and anti-oxidation. At present, the methods of detecting polysaccharide content of Ganoderma lucidum are destructive, and the steps are complicated and time-consuming. This study aims to explore the possibility of using hyperspectral imaging (HSI) to predict polysaccharide content in a nondestructive way during the growth of Ganoderma lucidum. The partial least square regression (PLSR) model shows good performance for Ganoderma lucidum ($R_p^2 = 0.924$, $RPD_p = 3.622$) with pretreatment method of Savitzky-Golay (SG) and standard normal variate (SNV), and feature selection method of successive projections algorithm (SPA). This study indicates that HSI can quickly and nondestructive detect the polysaccharide content of Ganoderma lucidum, provide guidance for the cultivation industry and improve the economic benefits of Ganoderma lucidum.

1. Introduction

Ganoderma is an edible mushroom, also known as Lingzhi in China, which has been used for centuries as nutraceuticals to improve health (Bishop et al., 2015; Huie & Di, 2004). Ganoderma belongs to Basidiomycota, Agaricomycetes, Polyporales, Ganodermataceae. There are many species of Ganoderma and Ganoderma lucidum is recognized as one of the most effective. At present, polysaccharides, triterpenes, steroids, nucleotides, amino acids, fatty acids, and lipids have been isolated from Ganoderma lucidum. Polysaccharides derived from plant foods are major components of the human diet, and cell wall polysaccharides are the major components of dietary fiber (Lovegrove et al., 2017). Numerous researches have demonstrated that polysaccharides are the most important components responsible for immunomodulatory (Ren et al., 2021). Due to its high nutritional value, Ganoderma is formulated in numerous health foods (functional foods or nutraceuticals, or dietary supplements) to promote longevity and improve human health. Traditional consumption methods include making soup, wine, and tea with the fruiting body of Ganoderma, the current study of Ganoderma food is the use of Ganoderma extract as raw materials, added to a variety of food and beverage, configured as a functional food with health effects (Li et al., 2019).

Although Ganoderma was considered to be a “fairy herb” in China, this precious food was not widely used due to lack of resources until Ganoderma was successfully cultivated by researchers from the Institute of Microbiology of the Chinese Academy of Sciences in the 1950 s (Lin, 2015). Since then, the artificial cultivation of Ganoderma fruiting body has become more and more popular in China, followed by Ganoderma-related food and industry. It is particularly important to detect the polysaccharide content of the Ganoderma lucidum fruiting body during its growth. The percentage content of polysaccharides in Ganoderma lucidum will change with growth, the highest in the mycelium and the lowest in the fruiting body, but the total content is still rising (Ren et al., 2021). To quantitatively analyze the polysaccharide content, the phenol sulfuric acid and anthrone sulfate methods are mainly used at present. The results of these methods are more accurate, but they also have many

* Corresponding author at: College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China.
E-mail address: edithzhao@scau.edu.cn (J. Zhao).

https://doi.org/10.1016/j.fochx.2021.100199
Received 18 September 2021; Received in revised form 20 December 2021; Accepted 28 December 2021
Available online 29 December 2021

disadvantages, such as complex operations, time-consuming, and destructive. To achieve non-destructive monitoring, real-time quantitative prediction of polysaccharide content, determine the optimal harvest time node of Ganoderma lucidum, and improve the quality, a method was proposed to predict the polysaccharide content of Ganoderma lucidum by using hyperspectral images in this paper. To our knowledge, the application of this method to the quantitative detection of polysaccharides in the cultivation process of Ganoderma lucidum has not been reported yet.

Hyperspectral imaging (HSI) is a technique that combines traditional imaging and spectroscopy to obtain spatial and spectral information of objects (Gowen et al., 2007). Although HSI was originally designed for remote sensing (Goetz et al., 1985), it has been increasingly used for nondestructive content detection, such as lipid prediction in cocoa beans (Caporaso et al., 2021), soluble solids content in kiwifruit (Ma et al., 2021) and Chlorophyll content of winter wheat (Cai et al., 2021). Yang et al. measured pectin polysaccharides using HSI in mulberry fruit showed that HSI is a promising alternative to the chemical method to rapidly and nondestructively measure the polysaccharides (Yang et al., 2021). Chen et al. collected the near-infrared spectrum of Ganoderma powder proving the feasibility of using HSI to measure polysaccharide content (Chen et al., 2012).

However, there are new challenges for the non-destructively polysaccharide detection of Ganoderma lucidum in the process of cultivation. Ganoderma lucidum fruiting body has low polysaccharide content, about 0.73% (Linghua et al., 2017), and the distribution is uneven, and the content of the cap surface is even lower. To improve the detection accuracy, more detailed research was presented in this paper.

A method was proposed to quickly and non-destructively detect the polysaccharide content of Ganoderma lucidum fruiting bodies using HSI technology. We collected both visible and near-infrared spectra of Ganoderma lucidum samples at the same time, extracted characteristic bands, and used machine learning algorithms to establish the regression relationship between the spectra of different tissue regions of the Ganoderma lucidum cap and polysaccharide content.

2. Materials and methods

2.1. Sample preparation

280 Ganoderma lucidum samples were provided by Guangdong Yuewei Edible Fungi Technology Co. These samples were cultured at a temperature of 26.8 °C±1.27 °C, the humidity of 81.78% ± 6.43%, the CO_2 concentrations of 2355.62 ppm ± 631.69 ppm and light intensity of 300 lx to 600 lx. Four collection points were designed in this experiment, the diameters of the Ganoderma lucidum cap were 3–5 cm (18–19 days after fruiting), 5–7 cm (21–22 days after fruiting), 7–10 cm (24–25 days after fruiting) respectively, and 20 days after fruiting body sprayed powder. The RGB images of four samples from the four periods taken by the mobile phone were shown in Fig. 1(a). There were 70, 150, 30, and 30 samples of Ganoderma lucidum from each period. When each batch of samples met the designed harvesting conditions, they were picked and transported to the laboratory to acquire the spectral image and numbered.

2.2. Hyperspectral imaging system and image acquisition

Hyperspectral image data of Ganoderma lucidum, also called hypercubes, were collected by GaiaField Pro-V10E (Sichuan Shuanglihepu Technology Co. Ltd, China), which consists of an imaging spectrograph (ImSpector V10E, Specim, Finland) with a spectral range of 402.6–1005.5 nm, the spectral resolution of 2.8 nm, a CCD with a maximum spatial resolution of 960 × 1040, and GaiaField Pro-N17E (Sichuan Shuanglihepu Technology Co. Ltd, China), which consists of an imaging spectrograph (ImSpector N17E, Specim, Finland) with a spectral range of 887–1703 nm, the spectral resolution of 5 nm, an InGaAs camera with a maximum spatial resolution of 320 × 400. The light source adopted a diffuse halogen light source (HSIA-LS-*T*-H, Sichuan Shuanglihepu Technology Co. Ltd, China), of which the spectral range was from 350 to 2500 nm. A software (SpecVIEW, Sichuan Shuanglihepu Technology Co. Ltd, China) was used to set acquisition parameters and collect hyperspectral images. The whole hyperspectral image acquisition system was shown in Fig. 1(b). The hyperspectral camera, halogen lamp, and sample were all located in an all-black environment.

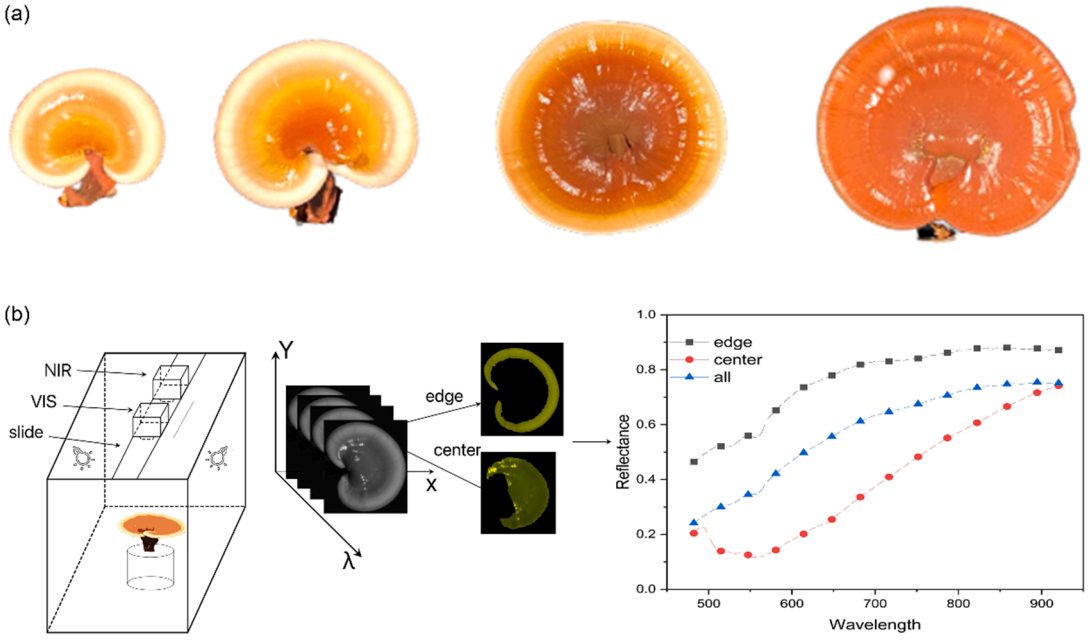

Fig. 1. Spectral collection environment and processing of samples. (a). RGB picture of samples from four periods, (b). Hyperspectral image acquisition system and ROI.

The angle between the halogen lamp and the horizontal plane was 45°, and the difference in optical power distribution was below 10%. The height between the hyperspectral camera lens and the sample was 400.0 mm. The acquisition parameters of GaiaField Pro-V10E (VIS) were set as spatial resolution 960 × 991, exposure time 3.7 ms. And the acquisition parameters of GaiaField Pro-N17E (NIR) were set as spatial resolution 320 × 333, exposure time 30 ms. Each spectral cube was a Ganoderma lucidum sample, which has 176 bands and 256 bands, respectively.

2.3. Measurement of polysaccharide

After the spectral image acquisition was completed, the polysaccharides in Ganoderma lucidum fruiting bodies were extracted following the modified phenol–sulfuric acid method (Nielsen, 2010; Xu et al., 2005; Yu et al., 2016). Samples from the first (P_1) were mixed and pulverized every 7 pieces, and the second period (P_2) were mixed and pulverized every 5 pieces. The samples in P_1 and P_2 can be seen from the Fig. 1(a) that the sample is small and needs to be mixed to perform chemical experiments. And there were 100 sets of data on polysaccharide content. Glucose was dissolved in distilled water at concentrations of 1.25, 2.5, 5, and 10 mg/ml. Add 20 μl of sample and 20 μl of phenol solution (5%) to a 96-well round bottom plate, and then add 100 μl of concentrated sulfuric acid to decompose polysaccharides. And record the absorbance at 490 nm. According to the standard curve, get the polysaccharide content of the sample.

2.4. Hyperspectral image processing

2.4.1. Spectrum correction

To reduce the dark noise of the detectors, the raw hyperspectral images were corrected with the white and dark reflectance with the Eq. (1):

$$R_{\lambda,n} = \frac{S_{\lambda,n} - D_{\lambda,n}}{W_{\lambda,n} - D_{\lambda,n}} \tag{1}$$

$R_{\lambda,n}$ and $S_{\lambda,n}$ were the relative reflectance and the brightness of the raw hyperspectral image at wavelength λ and pixel n respectively. D and W were the reflectance of the blackboard and the whiteboard, respectively. D was obtained by covering the lens with a black lens cap, while W was obtained by using the white reference board of the hyperspectral image system.

2.4.2. Region of interest

Firstly, the background noise and the reflective part on the surface of the Ganoderma lucidum cap were removed by threshold and we got the cap of Ganoderma lucidum. Since the samples of the first three periods have white growth areas, we have two ways to select the region of interest (ROI), as shown in Fig. 1(b). The first one is to select white (R_{edge}) and yellow areas (R_{center}) as ROI respectively and the second one is to select the entire area (R_{all}) as ROI. Compared with NIR, it is easier to extract ROI according to the first scheme in VIS. We used the threshold, morphological operation, and finally, extracted the connected components to get R_{center}, and performed logical XOR operation on R_{center} and R_{all} to get R_{edge}. The hyperspectral reflectance of each Ganoderma lucidum sample was obtained by averaging all the pixels in the ROI. For the samples of the P_1 and P_2, the reflectivity of 7 and 5 samples was averaged, respectively.

2.4.3. Spectral pretreatment

To eliminate the noises from scattering effects, different spectral pretreatment methods were used, including standard normal variate (SNV) and Savitzky-Golay filtering (SG).

The SNV processing formula was shown in equations (2) and (3).

$$X_{SNV} = \frac{X - \overline{X}}{\sqrt{\frac{\sum_{k=1}^{m}(X_k - \overline{X})^2}{m-1}}} \tag{2}$$

$$\overline{X} = \frac{\sum_{k=1}^{m} X_k}{m} \tag{3}$$

In the Eqs. (2) and (3), m was the total number of bands, k ∈ {1, 2, 3, ⋯, m}, X was the original spectrum, X_k was the k_{th} band of X.

SG was proposed by Savitzky and Golay (Savitzky & Golay, 1964) and has been widely used in data smoothing and denoising. The core idea of SG is the weighted filtering of the data in the window, and the weight is obtained by the least square fitting of the given high-order polynomial. Its advantage is that it can retain the information of the signal more effectively while being smoothing. Set the width of the window to $n = 2m + 1$, the order of the polynomial to k, and each point $x = (-m, \cdots, 0, \cdots, m)$ in the window is fitted with the order of $k-1$ polynomial, as shown in equation (4). Equation (5) was a matrix representation of equation (4).

$$\begin{pmatrix} y_{-m} \\ \vdots \\ y_0 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & -m & \cdots & (-m)^{k-1} \\ 1 & -m+1 & \cdots & (-m+1)^{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & m & \cdots & m^{k-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} e_{-m} \\ e_{-m+1} \\ \vdots \\ e_m \end{pmatrix} \tag{4}$$

$$Y_{(2m+1)\times 1} = X_{(2m+1)\times k} \cdot A_{k\times 1} + E_{(2m+1)\times 1} \tag{5}$$

Find the least-squares solution of A and get the smoothed result according to equation (6).

$$Y_{SG} = X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Y \tag{6}$$

Due to the low correlation, too large n is not appropriate, and k is generally less than 4. In this experiment, $k = 3$ and $n = 5$ were used for smoothing.

2.4.4. Feature selection

To reduce the dimension of features, in addition to standardizing and smoothing the spectral data, feature engineering was also carried out on the preprocessed results. The dimension of spectral features can be reduced by feature extraction and feature selection. The successive projections algorithm (SPA) was used for feature selection here. SPA is a forward selection method that uses simple operations in a vector space to minimize variable collinearity and is proposed as a novel variable selection strategy for multivariate calibration (Araújo et al., 2001). First, select a column randomly, calculate its projection to the remaining column vectors, and select the spectral wavelength of the maximum projection, and repeat these steps until the set number of features is reached. Then, for the extracted variables, multiple linear regression analysis (MLR) models are established to obtain the cross-validated root mean square error ($RMSE_{cv}$) of the modeling set, and the model with the smallest $RMSE_{cv}$ is selected. The corresponding variables are the optimal features (Soares et al., 2013).

2.5. Modeling and evaluation

Partial least squares regression (PLSR) was used to establish the model of the visible and near-infrared spectrum and polysaccharide content. PLSR is a robust and reliable algorithm to establish a model and it has been used in many spectral-content regression studies (Gabrielli et al., 2021; Zhu et al., 2021) because of its ability to reduce the impact of multicollinearity of the data. Cross-validation maximizes the amount of data used for training the model because each sample will be used for training as well as validation. In this study, 5-fold cross-validation (CV) was adopted. Each evaluation value of the model will generate 5 results

in the validation set respectively, and their average will be used as the evaluation value of CV. All models were evaluated using the coefficient of determination (R^2_{cal}, R^2_{val}, R^2_{cv}), root mean square error ($RMSE_{cal}$, $RMSE_{val}$, $RMSE_{cv}$) and residual prediction deviation (RPD_{cal}, RPD_{val}, RPD_{cv}). The relationship between the effect of the model and R^2 was very high when $R^2 > 0.85$ and good when $0.85 > R^2 > 0.6$ (Hu et al., 2021). Generally, the larger R^2 and RPD, the smaller $RMSE$ corresponds to a better model.

3. Results

3.1. Characteristics of samples

As can be seen from Fig. 1(a), the appearance of Ganoderma lucidum samples in four periods was very different. With the growth of Ganoderma lucidum, in addition to the gradual increase in the diameter and area of the cap, the color of the cap also changed significantly. The P_1 and the P_2 samples had a white rim and a yellow center, while the former had a lighter center. The center of the third period (P_3) samples were brown, with a yellow periphery and a very fine white margin. The caps of the samples in the fourth period (P_4) were all dark brown, without white or yellow edges. The white edge is the growth area of Ganoderma lucidum, so the weight and cap area of Ganoderma lucidum will stop changing or change very little at a certain moment between the P_3 and the P_4.

3.2. Polysaccharide content

The maximum, minimum, and mean polysaccharide content of samples and standard deviation in four periods were shown in Fig. 2(a). The mean values of polysaccharide content in samples of four periods were 1.14% ± 0.12%, 0.77% ± 0.09%, 0.61% ± 0.11%, and 0.52% ± 0.07% respectively, which was small and gradually decreased. The declining trend of polysaccharides between the four periods is different, the fastest period from the P_1 to the P_2, and the slowest period from the P_3 to the P_4. It can be seen from Fig. 2(b) that the polysaccharide content of different samples in different periods may be the same, which was related to the characteristics of the Ganoderma lucidum samples. For the growth cycle of a sample, this situation is also possible: the total weight of the sample is different but the polysaccharide content per unit weight is the same. Then we can find a moment to maximize the total polysaccharide content of the sample.

3.3. Spectral of the samples

The spectral reflectance of four periods (P_1, P_2, P_3, P_4) samples in the ROI_{center}, ROI_{edge} and ROI_{all} of VIS and NIR has been shown in Fig. 3. Wavelengths under 482 nm and over 920 nm, under 930 nm and 1640 nm were found unreliable because of noise. Therefore, the wavelengths of 482 nm to 920 nm and 930 nm to 1640 nm were selected, with a total of 128 bands of VIS and 222 bands of NIR. Since the samples of P_4 has no edge and center regions, the spectral reflectance of $R_{edge,vis}$ and $R_{center,vis}$ of P_4 were replaced by $R_{all,vis}$. Similarly, the spectral reflectance of $R_{edge,nir}$ and $R_{center,vis}$ of P_4 were replace by $R_{all,nir}$.

As can be seen from Fig. 3(a), (c), and (e), the original reflectivity curves of VIS in the first three periods differed greatly, especially the peaks and troughs between 520 nm and 580 nm. The original reflectance curves of NIR (Fig. 3(b), (d) and (f)) had a similar trend and the value in $R_{center,nir}$ was slightly larger. Since the reflectivity of $R_{all,nir}$ was obtained by the reflectivity value of pixels in $R_{center,nir}$ and $R_{edge,nir}$ on average, and the pixels in $R_{center,nir}$ were much larger than those in $R_{edge,nir}$, the reflectivity of $R_{all,nir}$ was more affected by $R_{center,nir}$, and the curve was closer to it. The absorptions at 950–1000 nm and 1400–1500 nm may be related to the first harmonic of O—H in the water, and the absorption at 1000–1100 nm may be related to the second harmonic of O—H in the polysaccharides and monosaccharides, while 1150–1300 nm may be related to the first harmonic of C—H combination (Xiaobo et al., 2010). Since there were more peaks and troughs in NIR, we expected that using NIR spectra to build polysaccharide regression models will have a better effect.

3.4. Effect of pretreatment

The spectral reflectance curve of $R_{all,nir}$ after three different spectral pretreatments were shown in Fig. 4(b)–(d). It can be seen from Fig. 4(b) that the SG pretreatment made the spectral curve smoother, such as smoother peaks and valleys. And it also made the difference between samples of the same period smaller, which can be seen from the figure that the curve of the sample of the same period was more compact. Through SG smoothing, the random noise in the spectral signal can be eliminated, and the signal-to-noise ratio of the sample can be improved. SNV was used to eliminate the influence of scattering on the spectrum due to uneven particle distribution and different particle sizes. The result of SNV preprocessing was a new numerical sequence with a mean value of 0 and a variance of 1.

3.5. Polysaccharide prediction

PLSR models for polysaccharide content of Ganoderma lucidum fruiting body were built from HSI scans of both VIS and NIR, after appropriate treatment of the information of the hyperspectral cubes and feature selection. The results of these prediction models were shown in

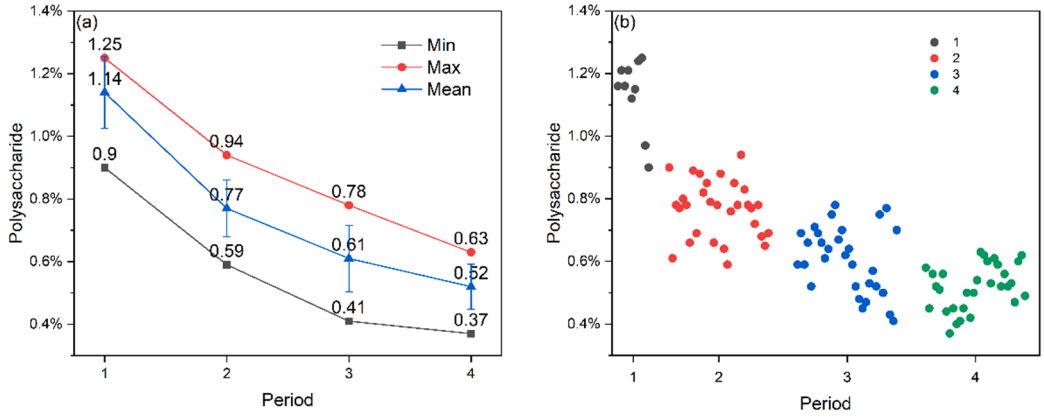

Fig. 2. The polysaccharide content of samples in four periods (a). The maximum, minimum, mean values, and standard deviation of Ganoderma lucidum polysaccharides in four periods, (b). The polysaccharide content of 100 samples in four periods.

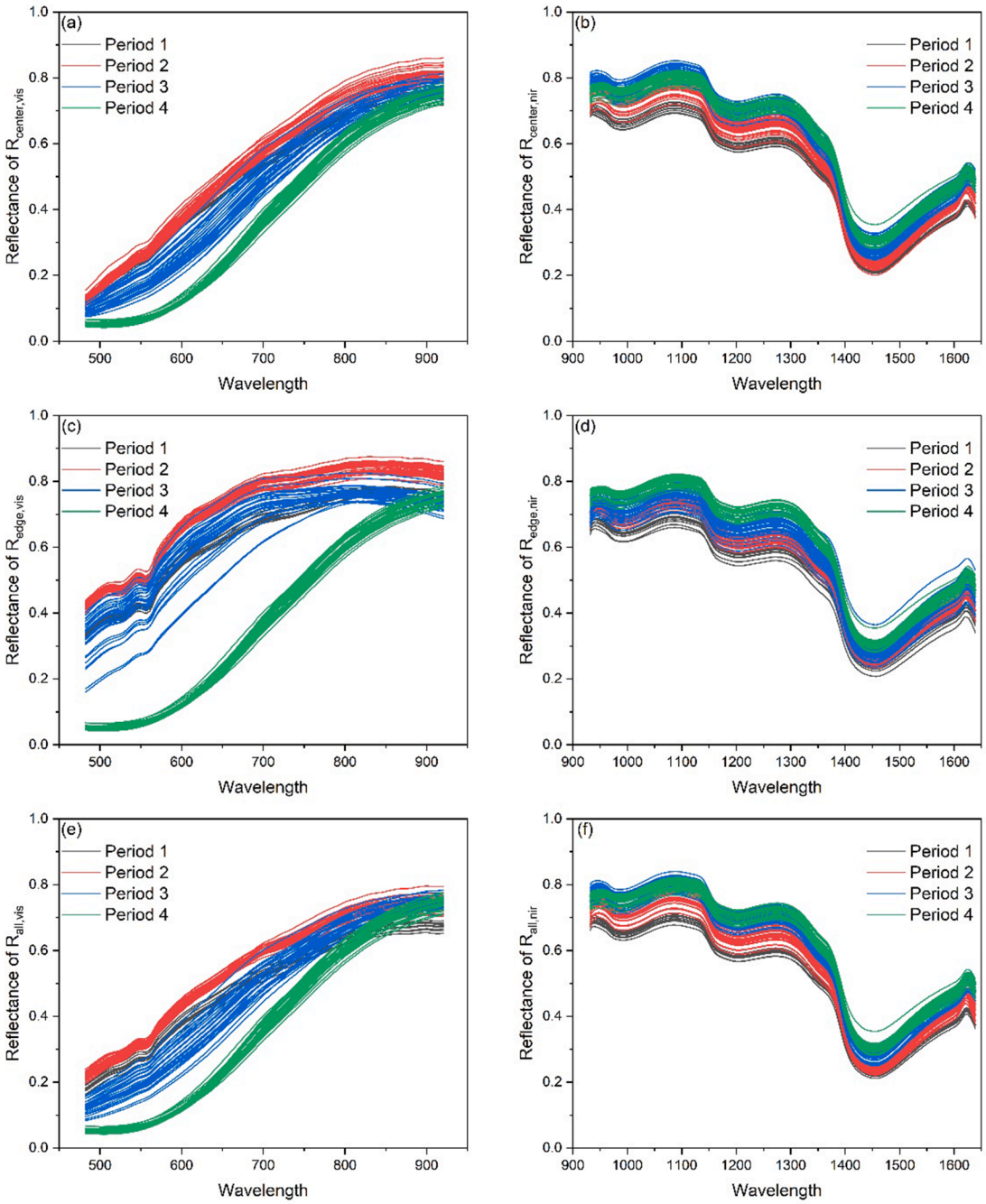

Fig. 3. Spectral reflectance in different ROI and spectral ranges. (a). $R_{center,vis}$, (b). $R_{center,nir}$, (c). $R_{edge,vis}$, (d). $R_{edge,nir}$, (e). $R_{all,vis}$, (f). $R_{all,nir}$.

Table 1. ROI_{center}, ROI_{edge}, and ROI_{all} represented the center, edge, and entire cap of Ganoderma lucidum as the ROI respectively. NONE in Preprocessing stands for no preprocessing, and SG + SNV means SG first, then SNV pretreatment. The subscripts cal, val, and cv of coefficient of determination (R^2), root mean square error (RMSE), and residual prediction deviation (RPD) represent calibration set, validation set, and cross-validation, respectively.

Since the results of cross-validation can better reflect the applicability of the preprocessing method and machine learning algorithm than the results of the calibration set and validation set, we give priority to the high value of R^2_{cv}, and consider R^2_{val} and R^2_{cal} comprehensively. All the results of the calibration and validation set were taken from the same fold from the 5-fold, so their data distribution was the same. All R^2_{cv} were greater than 0.74, indicating that both the original spectrum and pretreated spectra spectrum have good robustness to the polysaccharide content.

Modeling with the original spectra reflectance (no preprocessing, i.e., NONE) of ROI_{center}, the R^2_{val} of VIS and NIR are 0.86 and 0.786, respectively. After SG preprocessing, the effect of VIS is improved slightly, reaching 0.864, while the improvement of NIR is larger, reaching 0.866. The effect of SNV pretreatment on VIS is still very small, while NIR has increased from 0.786 to 0.828. We found that the original spectrum of VIS after the two pretreatments has a small increase or even a decrease, and it has an increase for NIR. After the two pretreatments are combined, as shown in Table 1, the effect of VIS has been reduced, and the effect of NIR has been greatly improved. The model established by $R_{center,nir}$ combined with SG and SNV preprocessing methods has the best effect, with R^2_{val} reaching 0.894.

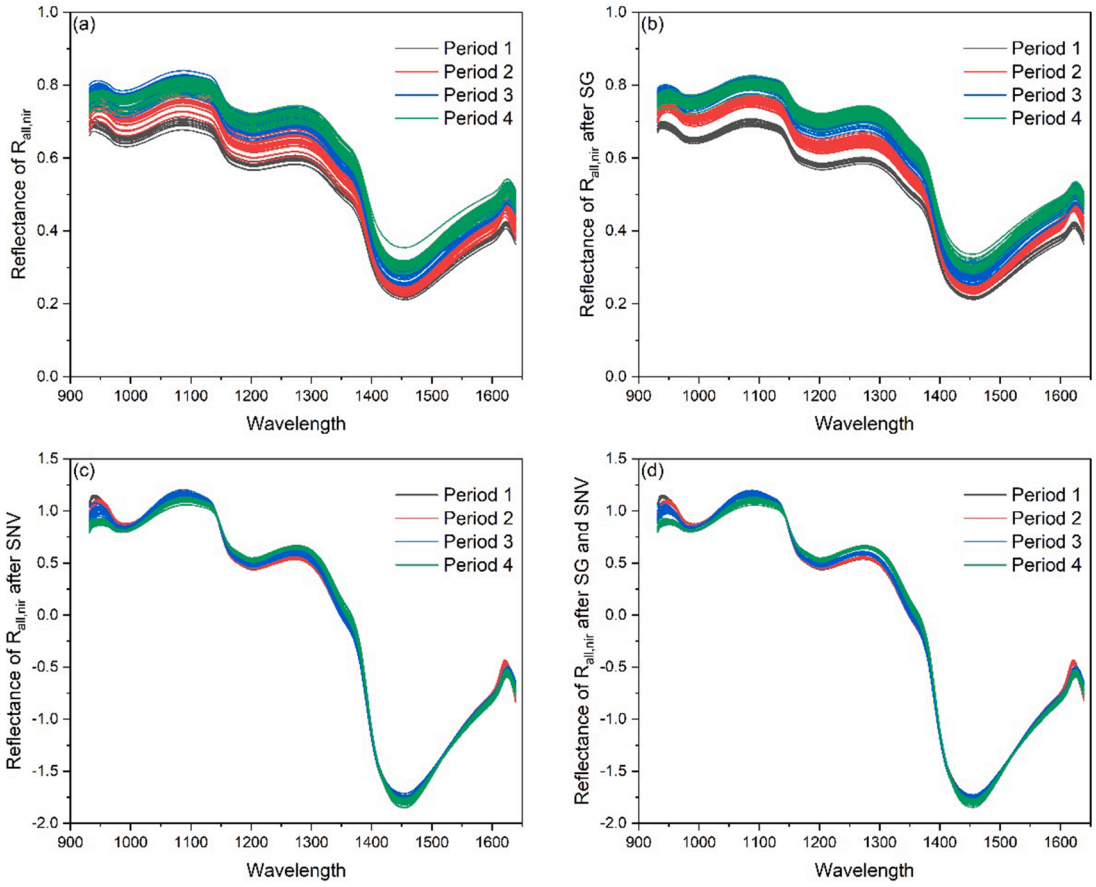

Fig. 4. Three spectral pretreatment results of $R_{all,nir}$. (a). Original spectrum, (b). SG, (c). SNV, (d). SG and SNV.

Table 1
Polysaccharide prediction result.

ROI	VIS/NIR	Preprocessing	R^2_{cal}	RMSE$_{cal}$	RPD$_{cal}$	R^2_{val}	RMSE$_{val}$	RPD$_{val}$	R^2cv	RMSE$_{cv}$	RPD$_{cv}$
ROI_{center}	VIS	NONE	0.815	8.66E−04	2.322	0.860	7.71E−04	2.674	0.782	9.27E−04	2.217
		SG	0.852	7.74E−04	2.600	0.864	7.61E−04	2.711	0.814	8.63E−04	2.351
		SNV	0.817	8.61E−04	2.335	0.836	8.35E−04	2.471	0.778	9.38E−04	2.175
		SG + SNV	0.839	8.07E−04	2.492	0.843	8.17E−04	2.524	0.825	8.40E−04	2.425
	NIR	NONE	0.819	8.56E−04	2.350	0.786	9.54E−04	2.162	0.769	9.69E−04	2.086
		SG	0.865	7.38E−04	2.724	0.866	7.55E−04	2.732	0.763	9.65E−04	2.150
		SNV	0.820	8.52E−04	2.359	0.828	8.56E−04	2.408	0.750	1.00E−03	2.039
		SG + SNV	0.887	6.77E−04	2.970	0.894	6.72E−04	3.070	0.818	8.51E−04	2.432
ROI_{out}	VIS	NONE	0.787	9.27E−04	2.168	0.842	8.19E−04	2.516	0.760	9.80E−04	2.081
		SG	0.819	8.55E−04	2.353	0.830	8.50E−04	2.427	0.784	9.32E−04	2.180
		SNV	0.801	8.97E−04	2.243	0.896	6.63E−04	3.108	0.808	8.75E−04	2.354
		SG + SNV	0.828	8.33E−04	2.413	0.867	7.52E−04	2.743	0.819	8.51E−04	2.382
	NIR	NONE	0.769	9.66E−04	2.082	0.855	7.86E−04	2.623	0.746	1.00E−03	2.048
		SG	0.825	8.41E−04	2.391	0.822	8.71E−04	2.367	0.780	9.40E−04	2.153
		SNV	0.799	9.01E−04	2.233	0.867	7.52E−04	2.740	0.756	9.80E−04	2.102
		SG + SNV	0.822	8.49E−04	2.369	0.831	8.49E−04	2.429	0.800	9.01E−04	2.244
ROI_{all}	VIS	NONE	0.785	9.33E−04	2.155	0.864	7.59E−04	2.716	0.768	9.53E−04	2.160
		SG	0.842	8.00E−04	2.513	0.874	7.32E−04	2.819	0.805	8.78E−04	2.330
		SNV	0.817	8.60E−04	2.338	0.893	6.74E−04	3.060	0.779	9.16E−04	2.290
		SG + SNV	0.841	8.02E−04	2.508	0.900	6.54E−04	3.155	0.819	8.37E−04	2.468
	NIR	NONE	0.796	9.07E−04	2.216	0.783	9.61E−04	2.146	0.762	9.73E−04	2.092
		SG	0.818	8.59E−04	2.341	0.815	8.88E−04	2.323	0.756	9.78E−04	2.100
		SNV	0.851	7.75E−04	2.595	0.863	7.64E−04	2.697	0.816	8.61E−04	2.358
		SG + SNV	0.886	6.80E−04	2.958	0.924	5.69E−04	3.622	0.823	8.35E−04	2.528

Modeled with original spectra reflectance of ROI$_{edge}$, the R^2_{val} of VIS and NIR was 0.842 and 0.855, respectively, and the effect was reduced and improved compared to the model of ROI$_{center}$, indicating that the original spectrum of ROI$_{edge}$ in NIR was more correlated with the polysaccharide content of the sample than the original spectrum of ROI$_{center}$. After preprocessing, the R^2_{val} of VIS and NIR dropped after SG and the R^2_{val} after SNV was improved. Combining the two preprocessing, the prediction effect also dropped. For this phenomenon, we assume that the

area of the ROI_{edge} occupies a relatively small area, resulting in a small number of pixels in ROI_{edge}, and the edge of ROI_{edge} is susceptible to light. When preprocessing the data, in this case, using SNV for standardization is better than SG smoothing. And for the ROI_{center}, the number of pixels is larger, and the edge of the area is not susceptible to light, the effect of SNV and SG preprocessing is related to the valid information retained by preprocessing. According to our assumption, since the spatial resolution of VIS is larger than that of NIR, the effect of the ROI_{edge} of VIS will be better than that of NIR. After SNV pretreatment, the best effect of VIS reached 0.896.

From the perspective of the two different regions of ROI_{center} and ROI_{edge}, the regression model established by corresponding preprocessing is very good, which shows that the spectrum of ROI_{center} and ROI_{edge} can reflect the polysaccharide content of the entire sample. Considering the difficulty of extracting two regions separately in actual application, to reduce the amount of calculation and improve the detection efficiency, it is the easiest to promote the application to use the entire cap of the Ganoderma lucidum sample to model and predict, and it is most reasonable to use the entire spectrum instead of part of the spectrum for modeling. Based on the original reflectance of the entire cap (ROI_{all}), the R^2_{val} of VIS and NIR reached 0.864 and 0.783, respectively, and multiple pretreatments have improved the effect. Among them, the combined effect of SG and SNV pretreatment reached the highest of VIS and NIR, R^2_{val} are 0.9 and 0.924, respectively. Fig. 5(a) and (b) have shown the prediction results of these two models.

3.6. Discussion

To accurately predict the polysaccharide content of the Ganoderma lucidum fruiting body using hyperspectral images, two sets of image acquisition equipment were used in this study, using visible and near-infrared spectra respectively. Moreover, we tried to predict the polysaccharide content by combing visible and near-infrared spectra, but the effect was not as good as using them alone, so we did not introduce it in detail in this manuscript.

For the ROI, three different schemes were carried out in this study: extracting the edge and center of the Ganoderma lucidum cap separately and extracting the whole Ganoderma lucidum cap. Meanwhile, a variety of spectral preprocessing methods and feature selection algorithms were used to find the best combination. The result showed that using the whole Ganoderma lucidum cap as ROI for prediction was the best. It indicates that the growth area on the periphery and the center of the Ganoderma ludicum jointly determine the polysaccharide content of Ganoderma lucidum, which is not directly related to the color it presents. For model building, We tried support vector machine (SVM), random forest (RF), and other algorithms, and finally, partial least square regression (PLSR) was the best. With the whole Ganoderma lucidum cap as ROI, after pretreatment by SG and SNV, and spectral feature extraction by SPA, PLSR modeling has the best effect, R^2_{cal} and R^2_{val} are 0.886 and 0.924, respectively.

Polysaccharide is the most important content of the Ganoderma lucidum. In addition, triterpene is also one of the main chemical components of Ganoderma lucidum. This experiment also tried to predict the triterpene content of the Ganoderma lucidum fruiting body by hyperspectral images. Supplementary data 1. (a) has shown the maximum, minimum, and mean triterpene content of samples and standard deviation in four periods. The mean values of triterpene content in samples of four periods were 2.46%±0.07%, 2.06%±0.37%, 2.46%±0.65%, and 1.8%±0.14% respectively. The two samples from the third period (P_3) were excluded and the spectra of the remaining 98 samples were used for modeling. Similar to the predicted polysaccharide, a variety of ROI selection schemes were adopted, different spectral preprocessing and feature extraction were used, and different regression algorithms were used, and the results of these prediction models were shown in supplementary data 2. Unlike polysaccharides, triterpenes were best predicted using the center of the Ganoderma lucidum cap as the ROI. Using the spectral of NIR, SG preprocessing first, and then SNV preprocessing, the model using the PLSR algorithm works best with a R^2_{cal} of 0.793 and a R^2_{val} of 0.705. This is the best result but it is not robust due to the R^2_{cv} of 0.588. The effect is poor, indicating that it is difficult to detect the content of triterpene through the Ganoderma lucidum fruiting body by HSI.

4. Conclusion

There are two main difficulties in this study. The first is that the fruiting body of Ganoderma lucidum usually has very low polysaccharide content. This poses a very big challenge to the accuracy of chemical detection and hyperspectral prediction. The second difficulty is the distribution of Ganoderma lucidum polysaccharides. Polysaccharide content is higher on the back of the Ganoderma lucidum cap because the Ganoderma lucidum spore powder is sprayed from there, which means the potential more accurate prediction. However, considering the actual application scenarios, it is difficult to collect the spectrum on the back of the cap due to the cover of the Ganoderma lucidum culture medium, so we finally used the surface of the cap to make predictions.

The polysaccharide in Ganoderma lucidum was predicted by combining machine learning and hyperspectral imaging. The best model used SG and SNV to preprocess the spectral data of NIR, SPA to extract features, and PLSR algorithm to model. The determination coefficient of calibration set (R^2_{cal}) and validation set (R^2_{val}) reached 0.886 and 0.924, respectively. The results of the VIS spectrum after the same preprocessing are 0.841 and 0.9. These two models have good results in the

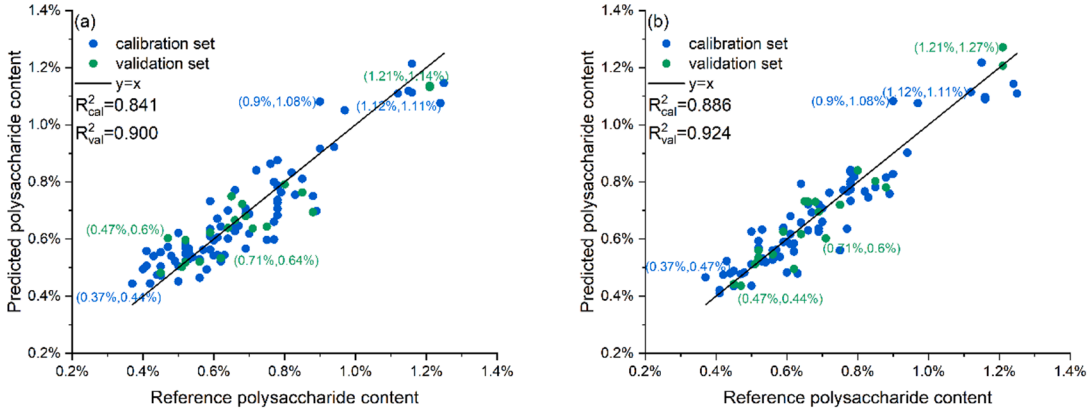

Fig. 5. Prediction results of the top two models. (a). PLSR on ROI_{all} in VIS with SG and SNV pretreatment, (b). PLSR on ROI_{all} in NIR with SG and SNV pretreatment.

prediction of Ganoderma lucidum fruiting body polysaccharides.

This study shows that it is feasible to use hyperspectral imaging equipment to collect the spectral cubes of Ganoderma lucidum from the top to predict its polysaccharide content, and highlights the great potential for rapid and nondestructive determination of polysaccharides in Ganoderma lucidum. VIS and NIR spectrum data can accurately predict the polysaccharide content of the Ganoderma lucidum fruiting body after being preprocessed by SG and SNV, and the effect of NIR is better. This study also tried modeling based on the spectra of different regions of Ganoderma lucidum, including the center, edge, and overall regions. The result showed that the model built with the overall works better.

We provided a feasible method for monitoring the polysaccharide content in the cultivation process of Ganoderma lucidum, obtaining growth information for growers, and determining the best harvesting time to improve its quality and economic value. It has the potential to be used in large-scale planting and high-throughput detection.

CRediT authorship contribution statement

Yu Liu: Conceptualization, Methodology, Software, Writing – original draft. **Yongbing Long:** Methodology. **Houcheng Liu:** Resources, Funding acquisition. **Yubin Lan:** Supervision. **Teng Long:** Conceptualization. **Run Kuang:** Conceptualization. **Yifan Wang:** Resources. **Jing Zhao:** Supervision, Conceptualization, Funding acquisition.

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.

Acknowledgments

This work was supported by a grant from the Key-Area Research and Development Program of Guangdong Province (2019B020214005-2).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2021.100199.

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*(2), 65–73. https://doi.org/10.1016/S0169-7439(01)00119-8

Bishop, K. S., Kao, C. H. J., Xu, Y., Glucina, M. P., Paterson, R. R. M., & Ferguson, L. R. (2015). From 2000years of Ganoderma lucidum to recent developments in nutraceuticals. *Phytochemistry, 114*, 56–65. https://doi.org/10.1016/j.phytochem.2015.02.015

Cai, Y., Miao, Y., Wu, H., & Wang, D. (2021). Hyperspectral estimation models of winter wheat chlorophyll content under elevated CO2. *Frontiers in Plant Science, 12*. https://doi.org/10.3389/fpls.2021.642917

Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2021). Total lipid prediction in single intact cocoa beans by hyperspectral chemical imaging. *Food Chemistry, 344*, 128663. https://doi.org/10.1016/j.foodchem.2020.128663

Chen, Y., Xie, M., Zhang, H., Wang, Y., Nie, S., & Li, C. (2012). Quantification of total polysaccharides and triterpenoids in Ganoderma lucidum and Ganoderma atrum by near infrared spectroscopy and chemometrics. *Food Chemistry, 135*(1), 268–275. https://doi.org/10.1016/j.foodchem.2012.04.089

Gabrielli, M., Lançon-Verdier, V., Picouet, P., & Maury, C. (2021). Hyperspectral imaging to characterize table grapes. *Chemosensors, 9*(4), 71. https://doi.org/10.3390/chemosensors9040071

Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for earth remote sensing. *Science, 228*(4704), 1147–1153. https://doi.org/10.1126/science.228.4704.1147

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral imaging – An emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology, 18*(12), 590–598. https://doi.org/10.1016/j.tifs.2007.06.001

Hu, N., Li, W., Du, C., Zhang, Z., Gao, Y., Sun, Z., … Wang, Z. (2021). Predicting micronutrients of wheat using hyperspectral imaging. *Food Chemistry, 343*, 128473. https://doi.org/10.1016/j.foodchem.2020.128473

Huie, C. W., & Di, X. (2004). Chromatographic and electrophoretic methods for Lingzhi pharmacologically active components. *Journal of Chromatography B, 812*(1), 241–257. https://doi.org/10.1016/j.jchromb.2004.08.038

Li, Z., Zhou, J., & Lin, Z. (2019). Development and Innovation of Ganoderma Industry and Products in China. In Z. Lin, & B. Yang (Eds.), *Ganoderma and Health: Biology, Chemistry and Industry* (Vol. 1181, pp. 187–204). Springer International Publishing Ag. https://doi.org/10.1007/978-981-13-9867-4_7.

Lin, Z. B. (2015). Chapter 1. Historical data of Chinese Ganoderma (Lingzhi) research. *Modern Research on Ganoderma*, 1–7.

Linghua, C., Zuxin, C., & Ming, X. (2017). Comparative Study on the Contents of Polysaccharides from 53 Different Ganoderma strains' Fruiting Bodies. *Natural Science Journal of Harbin Normal University*, 03.

Lovegrove, A., Edwards, C. H., De Noni, I., Patel, H., El, S. N., Grassby, T., … Shewry, P. R. (2017). Role of polysaccharides in food, digestion, and health. *Critical Reviews in Food Science and Nutrition, 57*(2), 237–253. https://doi.org/10.1080/10408398.2014.939263

Ma, T.e., Xia, Y.u., Inagaki, T., & Tsuchikawa, S. (2021). Non-destructive and fast method of mapping the distribution of the soluble solids content and pH in kiwifruit using object rotation near-infrared hyperspectral imaging approach. *Postharvest Biology and Technology, 174*, 111440. https://doi.org/10.1016/j.postharvbio.2020.111440

Nielsen, S. S. (2010). Phenol-sulfuric acid method for total carbohydrates. *Food Analysis Laboratory Manual, 47–53*. https://doi.org/10.1007/978-1-4419-1463-7_6

Ren, L.i., Zhang, J., & Zhang, T. (2021). Immunomodulatory activities of polysaccharides from Ganoderma on immune effector cells. *Food Chemistry, 340*, 127933. https://doi.org/10.1016/j.foodchem.2020.127933

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Soares, S. F. C., Gomes, A. A., Araujo, M. C. U., Filho, A. R. G., & Galvão, R. K. H. (2013). The successive projections algorithm. *TRAC Trends in Analytical Chemistry, 42*, 84–98. https://doi.org/10.1016/j.trac.2012.09.006

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta, 667*(1), 14–32. https://doi.org/10.1016/j.aca.2010.03.048

Xu, B., Dong, Y., Lin, L., & XU, Z. (2005). Determination of momordica charantia L. polysaccharide by improved phenol-sulfuric acid method. *Food Science and Technology, 7*, 79–82.

Yang, Liang, Gao, Huaqi, Meng, Liuwei, Fu, Xiaping, Du, Xiaoqiang, Wu, Di, & Huang, Lingxia (2021). Nondestructive measurement of pectin polysaccharides using hyperspectral imaging in mulberry fruit. *Food Chemistry, 334*, 127614. https://doi.org/10.1016/j.foodchem.2020.127614

Yu, H., Liu, Y., Zhou, S., Yan, M., Xue, L., Tang, Q., & Zhang, J. (2016). Comparison of the polysaccharides form fruiting bodies, mycelia and spore powder of Ganoderma lingzhi. *Mycosystema, 35*(02), 170–177.

Zhu, Y., He, H., Jiang, S., Ma, H., Chen, F., Xu, B., … Kang, Z. (2021). Mining hyperspectral data for non-destructive and rapid prediction of nitrite content in ham sausages. *International Journal of Agricultural and Biological Engineering, 14*(2), 182–187. https://doi.org/10.25165/ijabe.v14i2.5407.