ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Rapid and Accurate Authentication of Porcini Mushroom Species Using Fourier Transform Near-Infrared Spectra Combined with Machine Learning and Chemometrics

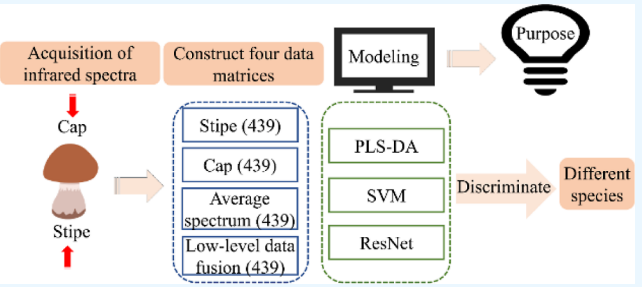
Hong Liu, Honggao Liu, Jieqing Li,* and Yuanzhong Wang*

Cite This: *ACS Omega* 2023, 8, 19663−19673

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Porcini mushrooms have high nutritional value and great potential, but different species are easily confused, so it is essential to identify them rapidly and precisely. The diversity of nutrients in stipe and cap will lead to differences in spectral information. In this research, Fourier transform near-infrared (FT-NIR) spectral information about imparity species of porcini mushroom stipe and cap was collected and combined into four data matrices. FT-NIR spectra of four data sets were combined with chemometric methods and machine learning for accurate evaluation and identification of different porcini mushroom species. From the results: (1) improved visualization level of t-distributed stochastic neighbor embedding (t-SNE) results after the second derivative preprocessing compared with raw spectra; (2) after using multiple pretreatment combinations to process the four data matrices, the model accuracies based on support vector machine and partial least-square discriminant analysis (PLS-DA) under the best preprocessing method were 98.73−99.04% and 98.73−99.68%, respectively; (3) by comparing the modeling results of FT-NIR spectra with different data matrices, it was found that the PLS-DA model based on low-level data fusion has the highest accuracy (99.68%), but residual neural network (ResNet) model based on the stipe, cap, and average spectral data matrix worked better (100% accuracy). The above results suggest that distinct models should be selected for dissimilar spectral data matrices of porcini mushrooms. Additionally, FT-NIR spectra have the advantages of being nondevastate and fast; this method is expected to be a promising analytical tool in food safety control.



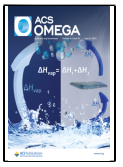
1. INTRODUCTION

Wild edible mushrooms are large fungal subentities rich in unique and appealing flavors and textures.[1] Because of their high nutritional and medicinal value, they are known as "delicacies from the mountains".[2] According to research, wild edible fungi are abundant in vitamins, proteins, fatty acids, minerals, etc.[3,4] They have many benefits such as antivirus, antibacterial, antioxidation, antitumor properties, etc.[5−7] The Food and Agriculture Organization (FAO) also brought up that "one meat dishes, one vegetable dishes, and one mushroom dishes" are the best diet structures for human beings.[8]

Among them, wild edible porcini mushrooms are among the most extensively consumed and valuable wild food mushrooms all over the world.[9] Based on the China Edible Fungi Association, in 2020, the output of porcini will be about 98,000 tons (https://www.cefa.org.cn/web/index.html?webId=1). The porcini mushroom industry plays a crucial role in the economic development of remote mountainous areas. Some of the known species of porcini mushrooms are toxic, such as *Lanmaoa asiatica*, which can cause hallucinogenic or even toxic phenomena, threatening the life and health safety of consumers if not processed properly or not cooked.[10] However, porcini mushrooms are similar in appearance, and it is difficult to identify their species quickly and accurately, even for experienced experts. In addition, at room temperature, the quality guarantee period of fresh porcini mushrooms is usually 1−3 days.[11,12] Therefore, porcini mushrooms are usually sold in dried flakes in the market, which also increases the difficulty of recognizing them. Commercial fraud is commonplace in the mushroom supply chain.[13] To prevent inferior and even toxic porcini mushrooms from entering the market, ensure that high-quality porcini mushrooms will not be confused; at the same time, to safeguard the health and rights of consumers, it is eager to seek a fast, economic, and accurate appraisal technology for porcini mushroom species.

To solve this problem, several methods are widely used. Conventionally, usual assessment by professionals uses easy

Received: February 23, 2023
Accepted: May 12, 2023
Published: May 23, 2023



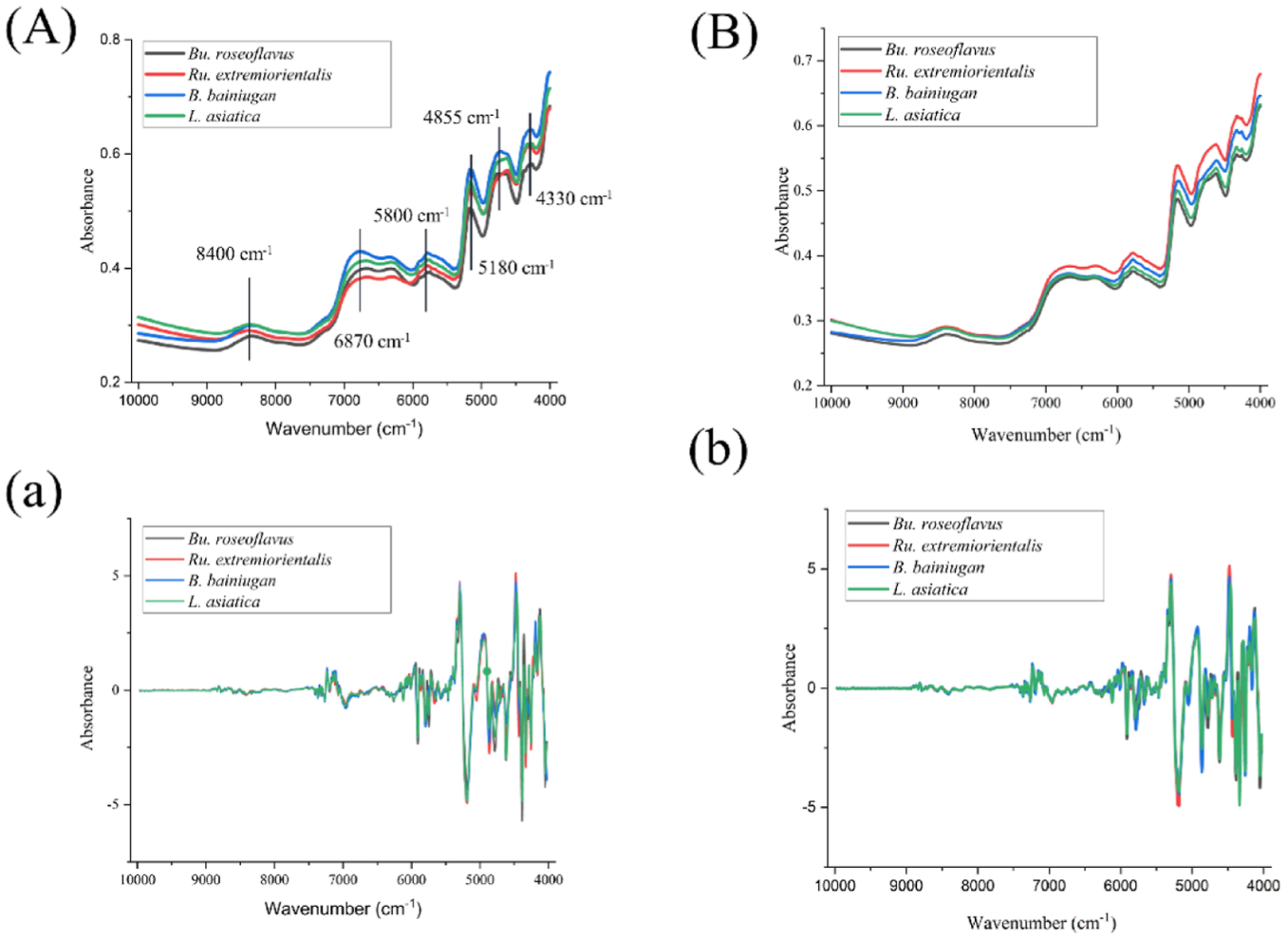


Figure 1. Average raw FT-NIR spectra of the stipe (A); average raw FT-NIR spectra of the cap (B); average FT-NIR spectra of stipe after SD + SNV pretreatment (a); and average FT-NIR spectra of the cap after SD + SNV pretreatment (b).

but low-confidence analysis of human senses.[14] This method is simple to use, but the accuracy is low and is vulnerable to subjective factors.[15] More reliable results were obtained by molecular identification techniques.[16] Nuclear magnetic resonance (NMR)[17] gas chromatography−mass spectrometry (GC−MS),[18] high-performance liquid chromatography (HPLC),[19] and other technologies are also extensively used to distinguish species. However, these methods are time consuming and consume large amounts of chemicals, require specialized operators, and cannot meet the market demand for porcini mushrooms.

In the present day, with the rapid growth of instrumentation feasibility and chemometrics, as an efficient and nondestructive technique, NIR spectroscopy has gone deep into food safety and other fields.[20] However, the subentities of porcini mushrooms are chemical composition complex and irregularly shaped, and the type and levels of nutrients in various parts are diverse.[3,4] In spectral measurements, spectral information is usually collected for two parts (stipe and cap). To our knowledge, there is little scientific literature on which part the spectrum of the porcini is more effective for species identification. In addition, porcini mushrooms as a commercially available product have long been popular with consumers and are not allowed to be wrongly classified. To compare the classification capabilities of different data sets, the data matrix is divided into four types, single data matrix of stipe and cap, respectively, low level data fusion matrix based on stipe and cap, and the average spectrum of (stipe + cap). The purpose of creating these data matrices is to evaluate whether changing the type of data sets will improve the classification performance of the model. Furthermore, when building discriminative models using NIR spectra, it is necessary to first remove noise and scattering information from the spectral data.[21] Also, some studies have pointed out that the results obtained by using only one preprocessing technique are usually not the best, and the data obtained by combining different preprocessing techniques have complementary information and can effectively improve the classification performance of the near-infrared model.[22,23]

Therefore, in this study, four common species of porcini mushrooms from Yunnan were used as experimental materials. First, the spectral information of the stipe and cap of different species of porcini mushrooms was collected, and the spectra were then combined into four data matrices [stipe, cap, the average spectrum of (stipe + cap), and low-level data fusion]. Spectral data from four data matrices are processed using single and multiple preprocessed combinations. Partial least-square discriminant analysis (PLS-DA), support vector machine (SVM), and residual neural network (ResNet) models were built for different data matrices, exploring the feasibility of near-infrared spectroscopy combining chemometrics and machine learning for nondestructive discrimination of porcini. This study not only compared the modeling

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

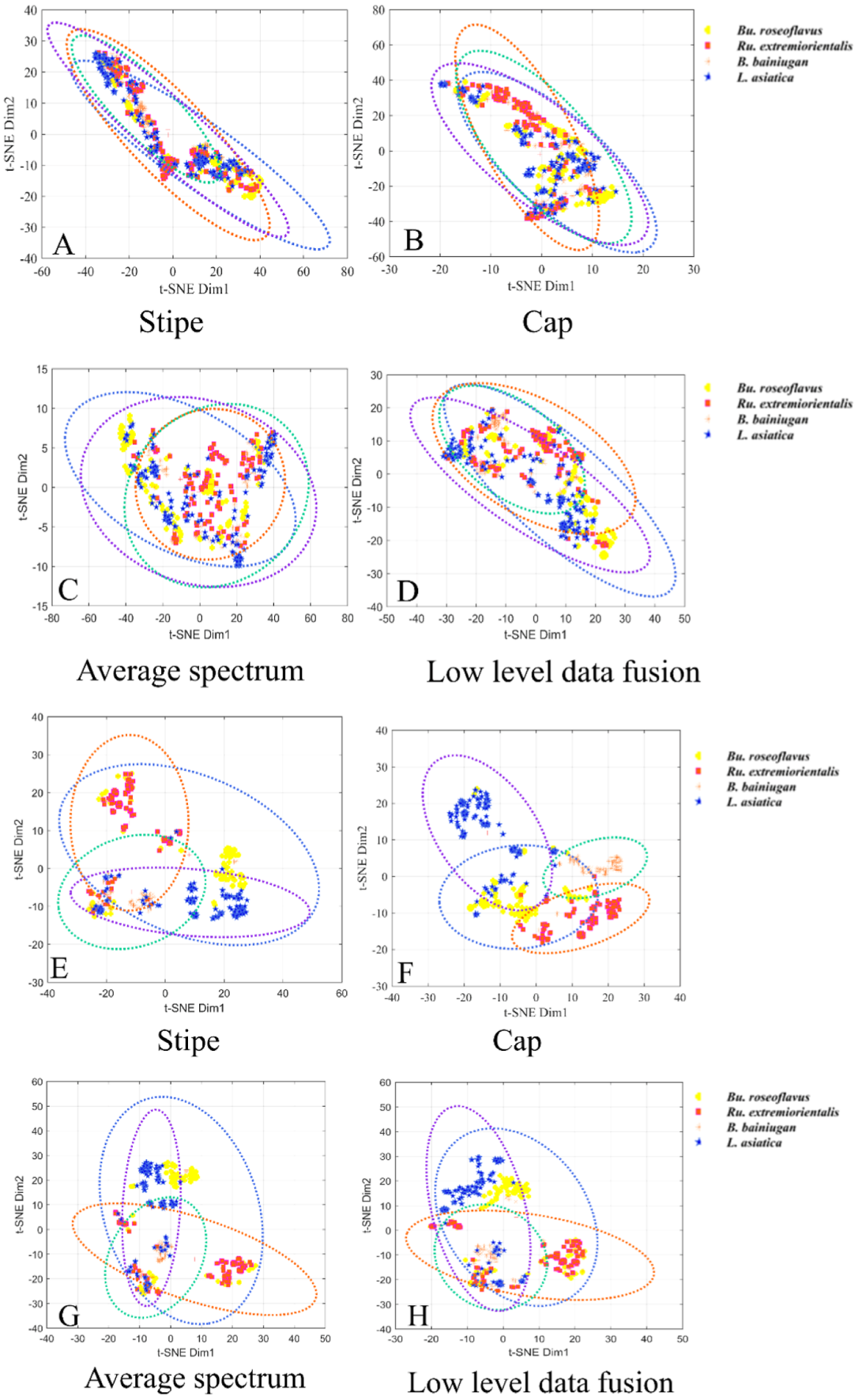


Figure 2. Cores scatter plots of four different species of porcini mushrooms using t-SNE [raw-averaged FT-NIR spectra (A−D) and SD and SNV preprocessing (E−H)].

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

Table 1. Modeling Results of PLS-DA Based on Four Data Matrices after Different Preprocessing Methods[a]

spectral pretreatment	LVs	R^2	Q^2	RMSEE	RMSECV	RMSEP	train acc	test acc
				Stipe				
none	23	0.861444	0.743689	0.16421	0.243926	0.143962	99.04%	100%
FD	18	0.881599	0.803119	0.151723	0.200615	0.130676	99.04%	100%
SD	11	0.892131	0.816067	0.143168	0.202378	0.187846	99.68%	96.8%
MSC	18	0.83478	0.730739	0.17773	0.249341	0.16463	97.77%	100%
SNV	18	0.835138	0.700228	0.177415	0.264556	0.161315	97.77%	100%
SG	22	0.850695	0.741302	0.16975	0.244333	0.148487	98.09%	100%
SD + SG	13	0.88927	0.793465	0.145312	0.216767	0.188461	99.36%	98.39%
SD + MSC	11	0.900641	0.79749	0.137933	0.212094	0.186846	99.68%	98.4%
SD + SNV	**13**	**0.921457**	**0.833303**	**0.122644**	**0.192983**	**0.165751**	**99.68%**	**99.2%**
				Cap				
none	22	0.871905	0.797843	0.156746	0.207288	0.139033	98.41%	98.4%
FD	12	0.914791	0.849767	0.128539	0.215361	0.229414	96.8%	91.72%
SD	11	0.898648	0.777056	0.137624	0.260586	0.161543	98.73%	97.6%
MSC	22	0.888078	0.800177	0.14631	0.20847	0.338099	98.41%	98.4%
SNV	23	0.890108	0.795028	0.144989	0.211812	0.14246	98.41%	98.4%
SG	22	0.869728	0.7969	0.158134	0.206839	0.139177	98.41%	98.4%
SD + SG	11	0.882162	0.781798	0.148446	0.25235	0.1679	98.09%	96.8%
SD + MSC	10	0.881688	0.803127	0.148983	0.208552	0.170227	98.41%	97.6%
SD + SNV	**11**	**0.902525**	**0.82970**	**0.13539**	**0.188601**	**0.156716**	**98.73%**	**97.6%**
				Average Spectrum				
none	18	0.844917	0.780295	0.170516	0.2151045	0.155472	96.18%	98.4%
FD	21	0.8650	0.7739	0.160243	0.232046	0.131503	98.09%	99.2%
SD	22	0.877462	0.777174	0.120677	0.176622	0.157634	99.04%	100%
MSC	17	0.835257	0.776169	0.173857	0.214565	0.153458	98.09%	99.2%
SNV	17	0.837927	0.837927	0.163322	0.19663	0.1546	98.09%	98.4%
SG	18	0.84266	0.770543	0.16373	0.230845	0.140366	98.09%	99.2%
SD + SG	14	0.913716	0.827337	0.1284191	0.188498	0.1639	99.04%	100%
SD + MSC	13	0.921597	0.849414	0.122221	0.184303	0.181535	99.36%	100%
SD + SNV	**12**	**0.920769**	**0.851822**	**0.116684**	**0.185881**	**0.149816**	**99.04%**	**100%**
				Low-Level Data Fusion				
none	21	0.848288	0.748594	0.151846	0.198558	0.141857	98.73%	100%
FD	17	0.886735	0.790753	0.143265	0.226442	0.115049	99.04%	100%
SD	13	0.941491	0.875818	0.0880801	0.165064	0.124546	100%	100%
MSC	24	0.884514	0.7701	0.1502181	0.237749	0.1192502	98.09%	99.2%
SNV	21	0.852555	0.71333	0.169289	0.259761	0.140642	98.09%	100%
SG	26	0.882446	0.766062	0.1515898	0.223054	0.122057	99.04%	100%
SD + SG	**13**	**0.949236**	**0.881138**	**0.0988546**	**0.154978**	**0.127564**	**99.68%**	**100%**
SD + MSC	13	0.937799	0.862661	0.10938	0.172403	0.133601	99.68%	100%
SD + SNV	13	0.937829	0.862756	0.1093532	0.17235	0.133552	99.68%	100%

[a]LVs: latent variables; R^2: correlation coefficient, indicates the fit of the model; Q^2: indicates the predictive power of the model; RMSEP: root-mean-square error of prediction; RMSEE: root-mean-square error of estimation; RMSECV: root-mean-square error of cross validation; Acc: accuracy; SG: Savitzky−Golay; SNV: standard normal variate; MSC: multiplicative signal correction; SD: second derivative: first derivative; and none: original spectrum.

effects between the original spectra and different preprocessed spectra, and by comparing the results of different models (PLS-DA, SVM, ResNet) built from different data matrices [Fourier transform near-infrared (FT-NIR) spectra collected from different parts of porcini or by combining spectra collected from different parts in multiple ways], construction of an accurate and efficient method for species distinguishes porcini mushrooms.

2. RESULTS AND DISCUSSION

2.1. FT-NIR Spectral Analysis. For facilitating the observation of variations between different porcini mushrooms species, the FT-NIR spectrum of the stipe and cap of porcini mushrooms was mapped. Figure 1 shows the obtained average FT-NIR spectra of four species of porcinis. The FT-NIR absorption peaks of the four porcini mushroom spectra were similar and the positions of the peaks were the same; however, there is a slight difference in peak height (absorbance). This phenomenon illustrates that the chemical composition species in the four porcini mushroom species were generally similar but accumulated in different amounts. When the spectra of the different parts were analyzed, greater differences in the intensity of absorption peaks of FT-NIR spectra were obtained from the porcini mushroom stipe (Figure 1A). In particular, absorption peaks in the 7000−5000 cm^{-1} band may imply a wide variation in the distribution and levels of the chemical compound in the stipe of porcini mushrooms. It may be easier to distinguish different species of porcini mushrooms using the spectral data set of the stipe. After second derivative (SD) preprocessing, more spectral information was displayed than

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

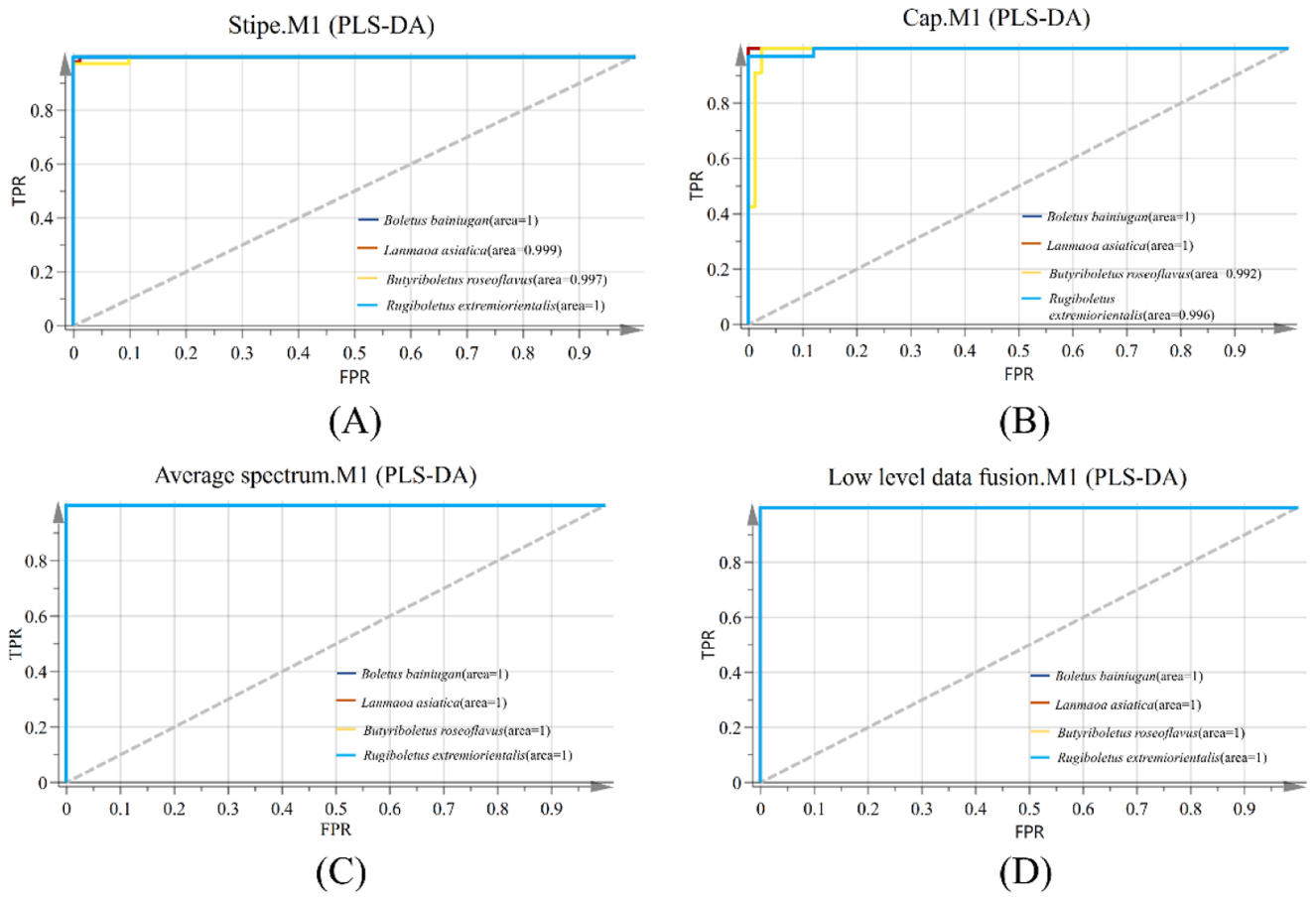


Figure 3. ROC) curves were used to evaluate the performance of the PLS-DA model in the test set: under the optimal preprocessing method (the stipe data matrix was preprocessed with SD + SNV) (A); the cap data matrix was preprocessed with SD + SNV (B); the average spectrum data matrix was preprocessed with SD + SNV (C); and the low-level data fusion data matrix was preprocessed with SD + SNV (D).

the original spectrum. As you can see from the graph, there are six common characteristic absorption peaks in four different porcini mushrooms, including three strong absorption peaks (6870, 5180, and 4600 cm^{-1}) and three weak absorption peaks (8400, 5800, and 4300 cm^{-1}). Among them, the first weak absorption peak is located near 8400 cm^{-1}. It corresponds with the SD overtone of C═C−H stretching; the main characterization is the unsaturated lipids in porcini mushrooms. Another lipid-related absorption band is located near 5800 cm^{-1}, which is made of the C−H stretch first overtone of methylene. There are two weaker absorption peaks near 5180 and 6870 cm^{-1}, mainly attributed to the first overtone of O−H stretching resp and a combo of O−H stretching and H−O−H deformation of water. In addition to this, some feature absorption zones were identified near 4855 cm^{-1}, mainly related to the amide group in the protein. Near 4330 cm^{-1} is the antisymmetric stretching vibration about CH_2 in polysaccharides.[40−42] It can be seen that porcini mushrooms contain many compounds. Unfortunately, porcini mushroom species cannot be identified by the subtle differences in the FT-NIR spectra In addition, the serious spectral overlap phenomenon makes it difficult to provide detailed information on different species. Therefore, it is necessary to use a chemometric approach to differentiate between the different species of porcini.

2.2. Analysis of t-SNE. Figure 2 offers a visualization result of the t-distributed stochastic neighbor embedding (t-SNE) between different porcini mushrooms based on four data matrices. The sample distribution for each species was represented by a 95% confidence ellipse. The shape of the ellipse meant the detailed orientation of the different categories of samples. The separation obtained using the original spectra (A−D) of the four data matrices was poor, with only a small fraction of *Ru. extremiorientalis* gathering was observed from low-level data fusion, and the remaining samples were scattered. Two-dimensional visualization results indicated that the FT-NIR info of various species of porcini mushroom samples was relatively similar and could not be easily distinguished. After SD + standard normal variate (SNV) pretreatment, the overlap between different samples is reduced. However, the results obtained from different data matrices were not significantly different; the distribution of *B. bainiugan* and *Ru. extremiorientalis* samples is more concentrated, denoting fewer individual differences. There is a small overlap between *L. asiatica* and *Bu. roseoflavus* samples, demonstrating that the chemical information of the two porcini mushroom species is similar and difficult to separate. These exploratory analyses reveal similarities in chemical information between different species of porcini; although the overlap between samples was reduced after pretreatment, there were still sample classification errors. The results illustrated that most samples could not be classified according to the predivision labels during the data visualization. Therefore, further analysis of distinguishing porcini mushroom species was carried out using pattern recognition.

2.3. Three Models Built Using FT-NIR. *2.3.1. Discriminant Results of PLS-DA Model.* The classification results of

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

the PLS-DA model based on four data matrices are implied in Table 1. To eliminate spectral interference, the influence of several various pretreatment approaches on model capabilities was compared. It is mentioned in Section 4.4.1 that the performance of a good model should include lower root-mean-square error of estimation (RMSEE), root-mean-square error of cross validation (RMSECV), and LV values and higher Q^2 and accuracy. The low-level data fusion matrix based on SD + Savitsky−Golay (SG) preprocessing had the best model performance, and the remaining three data matrices have the best model performance after SD + SNV preprocessing. Q^2 is in between 0.82970 and 0.881138, and root-mean-square error of prediction (RMSEP) is in between 0.127564 and 0.165751. This is mainly because the influence of solid particles and surface scattering on the spectrum can be eliminated after SD + SNV and SD + SG pretreatment, thus improving the performance of the model. However, not all preprocessing approaches can enhance the performance of the model. For example, after SNV pretreatment of the data set of stipes, compared with the model established by the original spectrum, Q^2 decreased by 5.8%, while RMSEP increased by 12%. The main reason may be that the SNV loses some important message that affects the predictive capability of the model while removing noise, which leads to the decline of the model performance. On the other hand, the performance of the discriminant model established with the four data types under the best preprocessing method is compared, and the confusion matrix of the PLS-DA model is shown in Table S1. The results of the classification of the four data types were as follows: classification on the ground of low-level data fusion > classification on the ground of the average spectrum of (stipe + cap) > classification on the ground of stipe > classification on the ground of cap. The PLS-DA model based on low-level data fusion matrix achieves optimal model results. The accuracy of the test set and training set reached 100 and 99.68%, respectively. Figure S1 shows the scatter plots of PLS-DA based on the best preprocessing method.

To better validate the classification capability of the PLS-DA model for various types of each category of porcini, Figure 3 computes and draws the area under curve (AUC) values for each class of porcini. This is a method that combines specificity and sensitivity to comprehensively assess model prediction accuracy. The AUC value is closer to 1, the higher the distinguishability of the two groups. The results imply that the AUC value on the ground of low-level data fusion is 1, exceeding AUC values based on stipe and cap. The results of the experiment further indicate that the model built on the ground of low-level data fusion is more effective. The main reason is that low-level data fusion is able to combine pieces of information from various parts of porcini and enriched spectral data to obtain more accurate and comprehensive results than using information from a single part. To test if the PLS-DA model is overfitting, this study conducted 200 iterations of persistence quizzes on the model built by low-level data fusion after SD + SG preprocessing. As shown in Figure S2, the left side is the simulated value and the right side is the true value. The standard to judge whether the model is overfitted is that the primitive points on the right are higher than all the blue Q^2 values on the left, and the regression line of the Q^2 point converges is lower than zero with the vertical axis (left).[43] The permutation values of the porcini mushroom species identification model (left) are lower than R^2 and Q^2 of the primitive points (right). The results revealed that after the SD + SG preprocessing, the PLS-DA model built using a low-level data fusion data matrix was not overfitting and could effectively identify different species of porcini.

2.3.2. SVM Model. Table 2 shows the outcomes of the SVM models constructed using disparate preprocessing approaches from the four data matrices. Figure S3 shows the classification of the test set in SVM based on the optimal preprocessing method. The accuracy of the SVM model training set and test set established by using the original spectral data of cap is 100%, but the *c*-value was too large (1024) and the model was easily overfitting. Compared with the original spectra, the risk of overfitting is reduced and the classification accuracy is increased by the four data matrices after the SD + SNV and SD

Table 2. Modeling Results of SVM Based on Four Data Matrices after Different Preprocessing Methods[a]

spectral pretreatment	*c*	*g*	train acc	test acc
		Stipe		
none	71,455.2002	$8.637 \times 10^{-0.5}$	96.82%	76.8%
FD	2048	1.7263×10^{-4}	99.04%	99.2%
SD	2	1.5625×10^{-2}	98.73%	100%
MSC	65,536	$6.1035 \times 10^{-0.5}$	96.50%	60.8%
SNV	23170.475	1.2207×10^{-4}	96.50%	99.2%
SG	741,455.2002	1.2207×10^{-4}	96.82%	99.2%
SD + SG	2896.3094	1.28×10^{2}	98.09%	99.2%
SD + MSC	2	1.1049×10^{-2}	98.73%	100%
SD + SNV	**1.4142**	$\mathbf{1.1049 \times 10^{-2}}$	98.97%	100%
		Cap		
none	1024	2.4414×10^{-4}	100%	100%
FD	128	5.5243×10^{-3}	98.09%	99.2%
SD	4	7.8125×10^{-3}	98.41%	99.2%
MSC	1,048,576	$2.1579 \times 10^{-0.5}$	98.41%	74.4%
SNV	741,455.2002	$2.1579 \times 10^{-0.5}$	98.1%	74.4%
SG	65,536	4.8828×10^{-4}	95.22%	98.4%
SD + SG	8	7.8125×10^{-3}	98.73%	97.2%
SD + MSC	5.6569	1.1049×10^{-2}	98.09%	97.6%
SD + SNV	**16**	$\mathbf{3.9063 \times 10^{-3}}$	98.30%	97.26%
		Average Spectrum		
none	370,727.6001	3.4527×10^{-4}	69.6%	76.8%
FD	1448.1547	2.4414×10^{-4}	99.04%	69.6%
SD	1.4142	2.2097×10^{-2}	99.04%	69.6%
MSC	2896.3094	9.7656×10^{-4}	99.04%	100%
SNV	4096	4.8828×10^{-4}	100%	99.2%
SG	524,288	3.4527×10^{-4}	96.82%	100%
SD + SG	1.4142	3.125×10^{2}	99.04%	98.4%
SD + MSC	2	2.2097×10^{-2}	98.73%	100%
SD + SNV	**2**	$\mathbf{3.125 \times 10^{-2}}$	98.99%	100%
		Low-Level Data Fusion		
none	8192	3.4527×10^{-4}	94.9%	99.2%
FD	362.0387	2.4414×10^{-4}	100%	69.6%
SD	1	2.2097×10^{-2}	99.04%	100%
MSC	5792.6188	2.4414×10^{-4}	94.59%	98.4%
SNV	524,288	$1.9073 \times 10^{-0.6}$	94.9%	66.4%
SG	11,585.2375	3.4527×10^{-4}	94.59%	99.2%
SD + SG	1	2.2097×10^{-2}	99.04%	100%
SD + MSC	**1.4142**	$\mathbf{7.8125 \times 10^{-3}}$	99.04%	100%
SD + SNV	1.4142	7.8125×10^{-3}	98.73%	100%

[a] *g*: kernel parameter; *c*: penalty factor; Acc: accuracy. SG: Savitzky−Golay; SNV: standard normal variate; MSC: multiplicative signal correction; SD: second derivative; FD: first derivative; none: original spectrum.

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

+ multiple scatter correction (MSC) preprocessing. Under the optimal preprocessing method, the SVM model built on low-level data fusion data matrix works best; the accuracy values of training and test sets are 99.04 and 100%, respectively. Furthermore, there is no risk of overfitting in the model. It implied that low-level data fusion has obtained a more comprehensive data set, which can better represent the overall information of porcini mushrooms than the data matrix of stipe and cap.

2.3.3. Discrimination Results of ResNet. Traditional one-dimensional NIR spectrum suffers from a peak overlap; this deficiency can be overcome by transforming 1D linear spectra to 2D images. To reduce the computational complexity and speed up computing efficiency, the characteristic absorption bands (7000−4000 cm^{-1}) of porcini mushrooms were intercepted to produce two-dimensional correlation spectroscopy (2DCOS) images, which laid the foundation for further deep-learning (DL) analysis.

As a new DL model, ResNet enables us to solve the problem of gradient bombing and gradient disappearance of the traditional DL model to a certain extent, speed up the model computing speed, and improve the accuracy rate.

The performances of 20 ResNet models built on four data matrices were compared. Table 3 shows the results of the four data sets at various epochs. The accuracy curves and loss value curves of the model at different epochs are shown in Figure S4. We can see that as epochs increase, the loss value function decreases. The training and test sets of the model tend to stabilize when the epoch reaches 5, and the accuracy of the models built by the three data matrices reached 100%, except for the low-level data fusion data set. In addition, confusion matrix by observation (Figure S5), when the epoch number is 40, the accuracy of external validation is 100% in all cases. However, when the iteration epochs increase to 50, the accuracy of the model decreases the runtime increases. Moreover, by comparing the results of the four data matrices in Table 3 when the epoch is 40 runs, only the low-level data fusion test set did not achieve 100% accuracy. The identification results were similar for the average spectrum of (cap + stipe) and the cap and slightly worse for the stipe. Interestingly, the constructed ResNet model gives the opposite results to PLS-DA and SVM. It took more time to run the model according to the low-level data fusion data matrix, but the worst outcomes are obtained, reducing the accuracy of the model. Results were obtained similar to those of Chen et al.[44] It showed that low-level fusion enriches the spectral data while superimposing invalid information on each other, reducing the correct classification rate. In all epochs, the model results based on low-level data fusion are not excelled to those based on the stipe and cap due to the influence of interference information.

Table 3. ResNet Model Results Based on the Original Spectra of the Four Data Matrices[a]

epochs	data types	train acc	test acc	external validation set (%)	loss value
10	stipe	0.99	1	100.00	0.07
	cap	1	1	100.00	0.056
	average spectrum	0.99	0.97	100.00	0.068
	low-level data fusion	0.90	0.83	100.00	0.261
20	stipe	1	1	100.00	0.037
	cap	1	1	100.00	0.023
	average spectrum	1	1	100.00	0.017
	low-level data fusion	0.96	0.87	100.00	0.17
30	stipe	1	1	95.00	0.024
	cap	1	1	100.00	0.014
	average spectrum	1	1	100.00	0.01
	low-level data fusion	0.95	0.81	90.00	0.128
40	**stipe**	**1**	**1**	**100.00**	**0.014**
	cap	**1**	**1**	**100.00**	**0.006**
	average spectrum	**1**	**1**	**100.00**	**0.006**
	low-level data fusion	**1**	**0.92**	**100.00**	**0.034**
50	stipe	1	0.96	95.00	0.012
	cap	1	1	100.00	0.005
	average spectrum	1	1	100.00	0.007
	low-level data fusion	0.99	0.85	97.00	0.039

[a]Acc: accuracy.

ResNet, PLS-DA, and SVM models based on four data matrices differed in their ability to identify porcini mushroom species, as well as in the comprehensive performance of the models. Sensitivity and specificity can comprehensively evaluate the accuracy of the model. As shown in Table 4, we can see the contrasting outcomes of the three models and the discrimination abilities of the three models based on the average spectrum of (stipe + cap) and low-level data fusion is similar, and sensitivity and specificity are both 1. However, from the results in Table 3, the model built by low-level data fusion has the worst performance; this may be because all information of the sample is displayed in the low-level data fusion and too many invalid data will reduce the accuracy of the model; the outcome in Table 4 also represents the poor capability of PLS-DA models based on stipe and cap data sets; this may be due to large variations in the chemical contents of various parts of porcini mushrooms, and specific parts have unique compounds that are not found in other parts. This further indicates that spectra based on individual parts cannot characterize the complete information of porcini mushroom subentities. Another aspect is that the model capabilities of SVM and PLS-DA need to be improved by various preprocessing methods. In contrast, the ResNet model built using synchronous 2DCOS images does not require any preprocessing and manual extraction of feature variables. Therefore, in addition to using low-level data fusion modeling, the ResNet model has an absolute advantage in using the other three data matrices.

2.3.4. Kruskal−Wallis Test. To compare whether there is a significant difference between the three models, we evaluate the model performance using the expert scoring method. The results of the scores scored by the experts are aggregated, and then the experts' scores are analyzed using the Kruskal−Wallis test. The results are shown in Table 5. $p < 0.05$ proved that there is a significant difference between the three models. Also, from the results of rank average, ResNet has the best model performance, the model performance of PLS-DA takes second place, and SVM has the worst performance. This further indicates that the ResNet model is expected to be an effective method for accurate and rapid identification of porcini species.

3. CONCLUSIONS

Our work proved that FT-NIR spectroscopy could be an effective method for assessing the authenticity of porcini.

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

Table 4. Different Model Parameters Are Evaluated[a]

		stipe			cap			average spectrum			low-level data fusion		
methods	class	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc
PLS-DA	*Boletus bainiugan*	1	1	1	1	1	1	1	1	1	1	1	1
	Lanmaoa asiatica	1	0.99	0.99	1	0.97	0.98	1	1	1	1	1	1
	Butyriboletus roseoflavus	0.97	1	0.99	0.92	0.99	0.97	1	1	1	1	1	1
	Rugiboletus extremiorientalis	1	1	1	0.97	1	0.99	1	1	1	1	1	1
SVM	*Boletus bainiugan*	1	1	1	1	1	1	1	1	1	1	1	1
	Lanmaoa asiatica	1	1	1	1	1	1	1	1	1	1	1	1
	Butyriboletus roseoflavus	1	1	1	0.97	1	1	1	1	1	1	1	1
	Rugiboletus extremiorientalis	1	1	1	1	0.99	1	1	1	1	1	1	1
ResNet	*Boletus bainiugan*	1	1	1	1	1	1	1	1	1	1	1	1
	Lanmaoa asiatica	1	1	1	1	1	1	1	1	1	1	1	1
	Butyriboletus roseoflavus	1	1	1	1	1	1	1	1	1	1	1	1
	Rugiboletus extremiorientalis	1	1	1	1	1	1	1	1	1	1	1	1

[a]Sen: sensitivity; Spe: specificity; and Acc: accuracy.

Table 5. Results of Kruskal–Wallis Test Are Based on the Expert Scoring Method[a]

designation	rank average	*p*
PLS-DA	2.20	
SVM	1.10	0.015
ResNet	2.70	

[a]$p < 0.05$ indicates significant difference.

Typically, classification results using the spectral data set of the cap and stipe alone were weaker, but the model also had a correct rate of over 90%, a demonstration of the potential of FT-NIR spectroscopy to identify porcini mushroom species. SVM and PLS-DA models built on low-level data fusion matrix achieved the best results. However, for the ResNet model, the accuracy rate of the model obtained built on the low-level data fusion matrix is below that obtained from the data set based on the cap and average spectra. This shows that the performance of the three models changes when the way the spectral data sets (data matrix) of the different parts of porcini are combined varies. In comparison, considering the simplicity of data processing and the accuracy of model operation, the results by the Kruskal–Wallis test both indicate the ResNet built using stipe and cap, and the average spectrum of (stipe + cap) may be the most effective way to identify porcini species.

In this study, only several common species of Yunnan Province were tested. If the range of factors such as sampling points and species type is expanded, a robust FT-NIR spectral database will be created while enhancing the adaptability of the model. This approach may hold promise for online spectroscopic monitoring in the porcini supply chain, a reliable and effective method for the identification of porcini mushrooms. In conclusion, the capability of three models based on various data matrices for the classification of porcini mushrooms species is discussed in this study from multiple perspectives. This study provides a reference for spectral acquisition and modeling analysis of other edible mushrooms and irregularly shaped samples (e.g., eggs).

4. MATERIALS AND METHODS

4.1. Information of Sample. We acquired 439 porcini subentities from Yunnan Province, including 111 subentities of *Butyriboletus roseoflavus*, 112 subentities of *Rugiboletus extremiorientalis*, 69 subentities of *Boletus bainiugan*, and 147 subentities of *L. asiatica*. The sample information is detailed in Table S2. Figure 4 shows the surrounding environment and morphological characteristics of the samples. Figure S6 shows the sample distribution of four species of porcini mushrooms. All subentities of porcini were appraised by Dr. Liu Honggao from Yunnan Agricultural University. After collecting *bolete*, the sediment was washed with deionized water and then sliced and dried at 60 °C to a constant weight. After that, the cap and stalk of the same porcini mushroom were separated, the stipe and cap of porcini were crushed, respectively, with a pulverizer (Huaxin Instrument Factory, Tianjin, China), and sieved

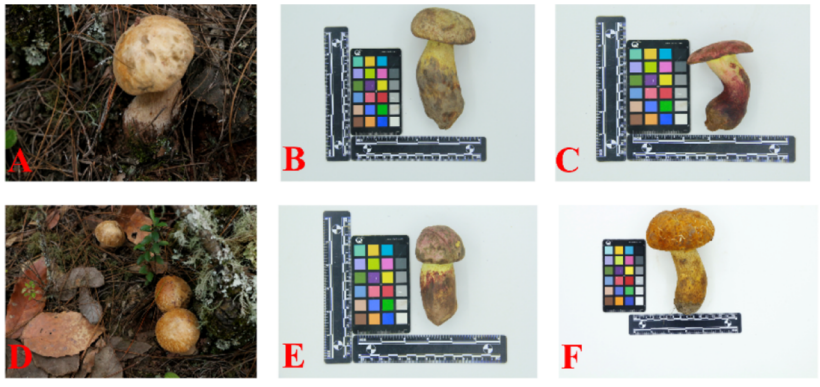


Figure 4. Sample information [(A,D): sample surroundings; (B): *Boletus bainiugan*; (C): *Lanmaoa asiatica*; (E): *Butyriboletus roseoflavus*; and (F): *Rugiboletus extremiorientalis*].

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

through 100 μm mesh sieves; finally, 439 samples of stipe and cap were obtained, respectively. The fine powder is collected and stored in a closed, self-sealing bag for subsequent research.

4.2. Spectral Acquisition. *4.2.1. FT-NIR Spectral Measurement and Preprocessing.* FT-NIR spectra of four porcinis were gathered by an FT-NIR instrument (Thermo Scientific Inc., Antaris II). The resolution was 4 cm^{-1}, the acquisition range was 10,000−4000 cm^{-1}, the signal was scanned 64 times, and the sample underwent repeated scan three times to obtain the mean value. For adequate representation, the chemical information of porcini and spectra of different parts were obtained: (1) stipe and (2) cap. Finally, a total of 878 spectral data were obtained (439 spectra for each part).

In order to improve the accuracy of the model, we compared different spectral pretreatment methods and their combinations during the development of the model. SG can reduce the noise of the spectrum,[24] SD and first derivative (FD) can remove baseline drift,[25] and SNV and MSC can eliminate the spectral differences caused by solid particles and different scattering levels.[26] Specifically, baseline drift is eliminated, spectral resolution is improved, and correlation between spectra and data is enhanced through various preprocessing methods,[27] making model results more reliable. The training and test sets were divided using the Kernnard Stone (K−S) algorithm. In this study, 293 porcini samples as the training set and 146 porcini samples as the test set were used.

t-SNE serves as one of the best nonlinear dimension reduction methods, and it can preserve the partial and total structure of the raw data.[28] This technique is characterized by the ability to make similar points close together in two dimensions. To observe the distribution of samples, we used t-SNE to visualize high-dimensional data sets. Differences in the distribution of samples in different data sets were observed by comparing the visualization results.

$$S(v) = \begin{bmatrix} s(v, t_1) \\ s(v, t_2) \\ \vdots \\ s(v, t_m) \end{bmatrix} \quad (1)$$

4.2.2. 2DCOS Image Acquisition. The traditional FT-NIR spectrum has the phenomenon of broadening and overlapping peaks.[29] 2DCOS can work around this problem and enhance the resolution.[30] In this study, 2DCOS is calculated by the discrete generalized 2DCOS algorithm. S denotes the dynamic spectral intensity at variable v, where t is the external perturbation and m is the spectrum measured at m steps with equal intervals of perturbation t.[30]

The 2DCOS images consist of integrative 2DCOS, asynchronous 2DCOS, and synchronous 2DCOS. Based on previous studies, the model accuracy of synchronous 2DCOS is significantly better than integrative and asynchronous 2DCOS.[23,29] Therefore, for the sake of avoiding repetitive work, this study focused on the exploration of synchronous 2DCOS. Synchronous 2DCOS can be obtained by eq 2

$$\Phi(v_1, v_2) = \frac{1}{m-1} s(v_1)^{\mathrm{T}} \cdot S(v_2) \quad (2)$$

According to the study of Yue et al.[31] different preprocessing methods will not affect the modeling effect of 2DCOS, and the preprocessing methods applicable to different data sets cannot be unified. Because of these factors, in this study, synchronous 2DCOS was obtained based on the original spectral data. Before generating 2DCOS, we selected 60% (263) of the samples from each species' data set as the training set, 30% (132) samples are selected as the test set, and external validation selected 10% (44) of the samples. The resulting images were then saved in JPEG format for subsequent deep-learning modeling analysis.

4.3. Low-Level Data Fusion Strategy. Low-level data fusion is to simply connect spectral data of stipe and cap from the same sample. It explains all the information in the data set, forming a more comprehensive data set, but requires a lot of calculation.[32]

4.4. FT-NIR Classification. *4.4.1. PLS-DA.* As an extension of PLS, PLS-DA is a common linear discriminant analysis approach,[33] which is characterized by transforming regression methods into classification tools.[34] R^2 and Q^2 are usually used to evaluate the properties of PLS-DA models. Under normal conditions, R^2 represents the degree to which the model fits the data, and R^2 close to 1 is a requisite for a positive model; when the value of this indicator is lower than 0.5, it means that it contains a lot of noise and extraneous pieces of information. Q^2 represents the ability to predict unknown digital, and a bigger Q^2 means better prediction ability. Other than that, lower estimated RMSEE, RMSECV, and RMSEP are important factors for better model performance.[35] In addition, specificity and sensitivity are also important indicators to assess the PLS-DA model, they are negatively correlated, and the relationship between the two is called the ROC curve. The accuracy of the model is assessed by the AUC. To inspect whether the model is over-fitting, this research carried out 200 iterations of permutation tests. In this study, the PLS-DA model was built in SIMCA 14.1 software using spectral matrix X and classification label Y.

4.4.2. SVM. SVM is a nonlinear method for solving training and classification problems with small samples. The purpose of SVM-based machine learning is to find the best hyper-flat. The two most significant factors that influence SVM categorization results are c (error penalty parameter) and g (kernel functions). Where, if c is too large, the model tends to overfit, and if c is too small, the model becomes underfitting. Not the same kernel functions have distinct category results, while various parameters of the identical kernel function have various categorization results.[36] Another aspect is that the higher the number of g, the less the support vector, and the lower the value of g, the greater the support vector, and the amount of support vectors influences the rate of model training and testing.[37] This study found the optimal combination of parameters using the grid search (GS) algorithm. First, establish a three-dimensional coordinate system and arrange the parameters to be searched C on the X-axis ($\log_2 C$) and g on the Y-axis ($\log_2 g$); the cross-validation accuracies of different parameter combinations are sequentially in the Z-axis. Finally, the combo (C, g) with the best cross-validation accuracy is chosen, where the extent of g is $[2^{-4}, 2^4]$ and the extent of C is $[2^{-2}, 2^4]$.[38] In this research, the SVM model was built using MATLAB 2017a.

4.4.3. Residual Neural Network. ResNet is a distinguished CNN, which simplifies the learning objectives and difficulty. First, a one-dimensional convolution calculation is executed on the import synchronous 2DCOS image. The Batchnorm layer was normalized and nonlinear activation was performed using the Relu function. The treated information is inputted into five

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673

identity blocks and four convolution blocks, respectively, for feature extraction. Key features were extracted and variables were reduced using global average pooling.[29] For more information about the model presentation, please refer to ref 29. We built a ResNet model with a learning rate of 0.01 and the weight attenuation coefficient λ was 0.001. First, synchronous 2DCOS images based on four data matrices were classified as external validation (44), prediction set (132), and training set (263). The model was trained by choosing different epochs (10, 20, 30, 40, and 50). The capabilities of the model were validated using a test set, and external verification was used to confirm the accuracy and generalization capability of the model. The optimal model was determined by comparing the running time and accuracy of the model.

In addition to the above parameters, three other parameters play an important role, which is specificity (SPE), sensitivity (SEN), and accuracy (ACC). In this research, the capabilities of the three models were assessed using these three parameters. These parameters can be obtained from the following 3 equations

$$\mathrm{SEN} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{3}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \tag{4}$$

$$\mathrm{ACC} = \frac{\mathrm{TN} + \mathrm{TP}}{(\mathrm{TN} + \mathrm{TP} + \mathrm{FP} + \mathrm{FN})} \tag{5}$$

In which, TN indicates true negative, TP denotes true positives, FN represents false negative, and FP is on behalf of false positive.[39]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c01229.

Confusion matrix for the PLS-DA model based on four data matrices; information about four porcini mushrooms from different geographical locations; scatter plot of PLS-DA based on the optimal preprocessing method; permutation test results of the PLS-DA models; SVM classification results of the optimal preprocessing method; results of ResNet models built on 2DCOS images of four data matrices at different epochs (accuracy of the training set and test set, loss value); external validation results of ResNet models built on 2DCOS images of four data matrices at different epochs; and geographical distribution of 439 samples of porcini mushroom subentities in Yunnan province, China (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Jieqing Li − *College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China*; Email: lijieqing2008@126.com

Yuanzhong Wang − *Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650200, China*; orcid.org/0000-0001-5376-757X; Email: Boletus@126.com

Authors

Hong Liu − *College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China; Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650200, China*

Honggao Liu − *Yunnan Key Laboratory of Gastrodia and Fungi Symbiotic Biology, Zhaotong University, Zhaotong 657000 Yunnan, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c01229

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

I am grateful for the valuable comments from editors and reviewers. This study was supported by the National Natural Science Foundation of China (grant number: 32160735); Special Program for the Major Science and Technology Projects of Yunnan Province (202002AA100007); and Special Program for the Major Science and Technology Projects of Yunnan Province (202102AE090051-1-01).

■ REFERENCES

(1) Xiao, Y.; Chen, L.; Fan, Y.; Yan, P.; Li, S.; Zhou, X. The effect of boletus polysaccharides on diabetic hepatopathy in rats. *Chem.-Biol. Interact.* **2019**, *308*, 61−69.

(2) Barros, L.; Baptista, P.; Correia, D.; Casal, S.; Oliveira, B.; Ferreira, I. Fatty acid and sugar compositions, and nutritional value of five wild edible mushrooms from Northeast Portugal. *Food Chem.* **2007**, *105*, 140−145.

(3) Li, T.; Wang, Y. Z.; Zhang, J.; Zhao, Y. L.; Liu, H. G. Trace element content of Boletus tomentipes mushroom collected from Yunnan, China. *Food Chem.* **2011**, *127*, 1828−1830.

(4) Tan, Y. Q.; Zeng, N.; Xu, B. J. Chemical profiles and health-promoting effects of porcini mushroom (Boletus edulis): A narrative review. *Food Chem.* **2022**, *390*, 133199.

(5) Salihović, M.; Šapčanin, A.; Špirtović-Halilović, S.; Mahmutović-Dizdarević, I.; Jerković-Mujkić, A.; Veljović, E.; Pehlić, E.; Gaši, F.; Zećiri, S. *Antimicrobial Activity of Selected Wild Mushrooms from Different Areas of Bosnia and Herzegovina*; Springer International Publishing: Cham, 2019.

(6) Zhang, L.; Hu, Y.; Duan, X. Y.; Tang, T. T.; Shen, Y. B.; Hu, B.; Liu, A. P.; Chen, H.; Li, C.; Liu, Y. T. Characterization and antioxidant activities of polysaccharides from thirteen boletus mushrooms. *Int. J. Biol. Macromol.* **2018**, *113*, 1−7.

(7) Guo, L.; Lan, N.; Li, H.; Xiang, P.; Kan, H. Effect of hot air drying temperature on the quality and antioxidant activity of Boletus edulis Bull.: Fr. *J. Food Process. Preserv.* **2021**, *45*, No. e15540.

(8) Liu, H.; Liu, H. G.; Li, J. Q.; Wang, Y. Z. Review of Recent Modern Analytical Technology Combined with Chemometrics Approach Researches on Mushroom Discrimination and Evaluation. *Crit. Rev. Anal. Chem.* **2022**, 1−24.

(9) Chen, J.; Liu, H. G.; Li, J. Q.; Wang, Y. Z. A rapid and effective method for species identification of edible boletes: FT-NIR spectroscopy combined with ResNet. *J. Food Compos. Anal.* **2022**, *112*, 104698.

(10) Gawlikowski, T.; Romek, M.; Satora, L. Edible mushroom-related poisoning. *Hum. Exp. Toxicol.* **2015**, *34*, 718−724.

(11) EISSA, H. A. A. Effect Of Chitosan Coating On Shelf Life And Quality Of Fresh-Cut Mushroom. *J. Food Qual.* **2007**, *30*, 623−645.

(12) Singh, P.; Langowski, H.; Wani, A. A.; Saengerlaub, S. Recent advances in extending the shelf life of fresh Agaricus mushrooms: a review. *J. Sci. Food Agric.* **2010**, *90*, 1393−1402.

(13) Wei, Y. M.; Li, L.; Liu, Y.; Xiang, S.; Zhang, H. Y.; Yi, L. Z.; Shang, Y.; Xu, W. T. Identification techniques and detection methods of edible fungi species. *Food Chem.* **2022**, *374*, 131803.
(14) Zhuang, J. D.; Xiao, Q.; Feng, T.; Huang, Q. R.; Ho, C.; Song, S. Q. Comparative flavor profile analysis of four different varieties of Boletus mushrooms by instrumental and sensory techniques. *Food Res. Int.* **2020**, *136*, 109485.
(15) Li, L. J.; Li, G. Y.; Xie, Q. T. Research progress on poisonous mushroom toxins classification and recognition. *Chinese J. Food Hyg.* **2013**, *25*, 383−387.
(16) Zhang, Y.; Mo, M. I.; Yang, L.; Mi, F.; Cao, Y.; Liu, C. L.; Tang, X. Z.; Wang, P. F.; Xu, J. P. Exploring the Species Diversity of Edible Mushrooms in Yunnan, Southwestern China, by DNA Barcoding. *J. Fungi* **2021**, *7*, 310.
(17) Alanne, A.; Issakainen, J.; Pihlaja, K.; Jokioja, J.; Sinkkonen, J. Metabolomic discrimination of the edible mushrooms Kuehneromyces mutabilis and Hypholoma capnoides (Strophariaceae, Agaricales) by NMR spectroscopy. *Z. Naturforsch., C: J. Biosci.* **2019**, *74*, 201−210.
(18) Zhou, J. J.; Feng, T.; Ye, R. Differentiation of Eight Commercial Mushrooms by Electronic Nose and Gas Chromatography-Mass Spectrometry. *J. Sens.* **2015**, *2015*, 1−14.
(19) Shi, X. M.; Zhang, J. S.; Tang, Q. J.; Yang, Y.; Hao, R. X.; Pan, Y. J. Fingerprint analysis of Lingzhi (Ganoderma) strains by high-performance liquid chromatography coupled with chemometric methods. *World J. Microbiol. Biotechnol.* **2008**, *24*, 2443−2450.
(20) Qu, J. H.; Liu, D.; Cheng, J. H.; Sun, D. W.; Ma, J.; Pu, H. B.; Zeng, X. A. Applications of near-infrared spectroscopy in food safety evaluation and control: a review of recent research advances. *Crit. Rev. Food Sci. Nutr.* **2015**, *55*, 1939−1954.
(21) Isaksson, T.; Næs, T. The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Appl. Spectrosc.* **1988**, *42*, 1273e−1284e.
(22) Puneet, M.; Roger, J. M.; Marin, F.; Biancolillo, A.; Rutledge, D. N. Parallel pre-processing through orthogonalization (PORTO) and its application to near-infrared spectroscopy. *Chemom. Intell. Lab. Syst.* **2021**, *212*, 104190.
(23) Yan, Z. Y.; Liu, H. G.; Li, T.; Li, J. I.; Wang, Y. Z. Two dimensional correlation spectroscopy combined with ResNet: Efficient method to identify bolete species compared to traditional machine learning. *LWT* **2022**, *162*, 113490.
(24) John, A.; Sadasivan, J.; Seelamantula, C. S. Adaptive Savitzky-Golay Filtering in Non-Gaussian Noise. *IEEE Trans. Signal Process.* **2021**, *69*, 5021−5036.
(25) Shi, H. T.; Yu, P. Q. Using Molecular Spectroscopic Techniques (NIR and ATR-FT/MIR) Coupling with Various Chemometrics to Test Possibility to Reveal Chemical and Molecular Response of Cool-Season Adapted Wheat Grain to Ergot Alkaloids. *Toxins* **2023**, *15*, 151.
(26) Mishra, P.; Lohumi, S. Improved prediction of protein content in wheat kernels with a fusion of scatter correction methods in NIR data modelling. *Biosyst. Eng.* **2021**, *203*, 93−97.
(27) Cortés, V.; Blasco, J.; Aleixos, N.; Cubero, S.; Talens, P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends Food Sci. Technol.* **2019**, *85*, 138−148.
(28) Luo, N.; Yang, X. T.; Sun, C. H.; Xing, B.; Han, J. W.; Zhao, C. J. Visualization of vibrational spectroscopy for agro-food samples using t-Distributed Stochastic Neighbor Embedding. *Food Control* **2021**, *126*, 107812.
(29) Dong, J.; Zuo, Z.; Zhang, J.; Wang, Y. Geographical discrimination of Boletus edulis using two dimensional correlation spectral or integrative two dimensional correlation spectral image with ResNet. *Food Control* **2021**, *129*, 108132.
(30) Noda, I. Two-Dimensional Infrared (2D IR) Spectroscopy: Theory and Applications. *Appl. Spectrosc.* **1990**, *44*, 550−561.
(31) Yue, J. Q.; Huang, H. Y.; Wang, Y. Z. A practical method superior to traditional spectral identification: Two-dimensional correlation spectroscopy combined with deep learning to identify Paris species. *Microchem. J.* **2021**, *160*, 105731.
(32) Sun, W. J.; Zhang, X.; Zhang, Z. Y.; Zhu, R. H. Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb. *Spectrochim. Acta, Part A* **2017**, *171*, 72−79.
(33) Fu, H.; Huan, S.; Xu, L.; Tang, L.; Jiang, J.; Wu, H.; Shen, G.; Yu, R. Moving Window Partial Least-Squares Discriminant Analysis for Identification of Different Kinds of Bezoar Samples by near Infrared Spectroscopy and Comparison of Different Pattern Recognition Methods. *J. Near Infrared Spectrosc.* **2007**, *15*, 291−297.
(34) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis − a marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10−23.
(35) Liu, W. B.; Zhang, B. Y.; Xin, Z. Q.; Ren, D. B.; Yi, L. Z. GC-MS Fingerprinting Combined with Chemometric Methods Reveals Key Bioactive Components in Acori Tatarinowii Rhizoma. *Int. J. Mol. Sci.* **2017**, *18*, 1342.
(36) Lee, L. H.; Wan, C. H.; Rajkumar, R.; Isa, D. An enhanced Support Vector Machine classification framework by using Euclidean distance function for text document categorization. *Appl. Intell.* **2012**, *37*, 80−99.
(37) Lei, C. K.; Deng, J.; Cao, K.; Xiao, Y.; Ma, L.; Wang, W. F.; Ma, T.; Shu, C. M. A comparison of random forest and support vector machine approaches to predict coal spontaneous combustion in gob. *Fuel* **2019**, *239*, 297−311.
(38) Liu, X. S.; Zhang, Z. B. Parameter optimization of Support Vector Machine based on improved grid search method. *J. Jiangxi Univ. Sci. Technol.* **2019**, *40*, 5−9.
(39) Gholami, R.; Raoof, N. *Support Vector Machine: Principles, Parameters, And Applications*; Elsevier Inc., 2017.
(40) Büning-Pfaue, H. Analysis of water in food by near infrared spectroscopy. *Food Chem.* **2003**, *82*, 107−115.
(41) Buijs, K.; Choppin, G. R. Near-Infrared Studies of the Structure of Water. I. Pure Water. *J. Chem. Phys.* **1963**, *39*, 2035−2041.
(42) Burns, D. A.; Ciurczak, E. W. *Handbook of Near-Infrared Analysis*, 3rd ed.; Taylor & Francis Group: 6000 Broken Sound Parkway N W, Suite 300, 2008.
(43) Wang, L.; Li, J. Q.; Li, T.; Liu, H. G.; Wang, Y. Z. Method Superior to Traditional Spectral Identification: FT-NIR Two-Dimensional Correlation Spectroscopy Combined with Deep Learning to Identify the Shelf Life of Fresh Phlebopus portentosus. *ACS Omega* **2021**, *6*, 19665−19674.
(44) Chen, X.; Li, J. Q.; Liu, H. G.; Wang, Y. Z. A fast multi-source information fusion strategy based on deep learning for species identification of boletes. *Spectrochim. Acta, Part A* **2022**, *274*, 121137.

https://doi.org/10.1021/acsomega.3c01229
ACS Omega 2023, 8, 19663−19673