



OPEN ACCESS

EDITED BY

Anil Kumar Puniya,
National Dairy Research Institute (ICAR),
India

REVIEWED BY

Qiang Li,
Chengdu University,
China
Tao Li,
Yuxi Normal University,
China

*CORRESPONDENCE

Yuan-Zhong Wang
✉ boletus@126.com
Hong-Gao Liu
✉ honggaoliu@126.com

SPECIALTY SECTION

This article was submitted to
Food Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 04 September 2022

ACCEPTED 21 December 2022

PUBLISHED 12 January 2023

CITATION

Li J-Q, Wang Y-Z and Liu H-G (2023)
Application of spectral image processing
with different dimensions combined with
large-screen visualization in the
identification of boletes species.
Front. Microbiol. 13:1036527.
doi: 10.3389/fmicb.2022.1036527

COPYRIGHT

© 2023 Li, Wang and Liu. This is an open-
access article distributed under the terms
of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Application of spectral image processing with different dimensions combined with large-screen visualization in the identification of boletes species

Jie-Qing Li¹, Yuan-Zhong Wang^{2*} and Hong-Gao Liu^{1,3*}

¹College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China,

²Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, China,

³Zhaotong University, Zhaotong, China

Boletes are favored by consumers because of their unique flavor, rich nutrition and delicious taste. However, the different nutritional values of each species lead to obvious price differences, so shoddy products appear on the market, which affects food safety. The aim of this study was to find a rapid and effective method for boletes species identification. In this paper, 1,707 samples of eight boletes species were selected as the research objects. The original Mid-Infrared (MIR) spectroscopy data were adopted for support vector machine (SVM) modeling. The 11,949 spectral images belong to seven data sets such as two-dimensional correlation spectroscopy (2DCOS) and three-dimensional correlation spectroscopy (3DCOS) were used to carry out Alexnet and Residual network (Resnet) modeling, thus we established 15 models for the identification of boletes species. The results show that the SVM method needs to process complex feature data, the time cost is more than 11 times of other models, and the accuracy is not high enough, so it is not recommended to be used in data processing with large sample size. From the perspective of datasets, synchronous 2DCOS and synchronous 3DCOS have the best modeling results, while one-dimensional (1D) MIR Spectrum dataset has the worst modeling results. After comprehensive analysis, the modeling effect of Resnet on the synchronous 2DCOS dataset is the best. Moreover, we use large-screen visualization technology to visually display the sample information of this research and obtain their distribution rules in terms of species and geographical location. This research shows that deep learning combined with 2DCOS and 3DCOS spectral images can effectively and accurately identify boletes species, which provides a reference for the identification of other fields, such as food and Chinese herbal medicine.

KEYWORDS

boletes species, 2DCOS images, 3DCOS images, Alexnet, Resnet, large-screen visualization

1. Introduction

Wild edible mushrooms are rich in protein, vitamins and a variety of mineral elements. Regular consumption can enhance immunity and promote metabolism. In addition, it still has medicinal value, can effectively prevent and treat tumors, edema and other diseases (Li et al., 2008; Cheung, 2010; Yao et al., 2018). Recently, it has been reported in the literature that mushrooms, including *Ganoderma lucidum*, should be considered as functional foods. It also pointed out that *Ganoderma lucidum* can treat coronavirus disease of 2019 (COVID-19; El Sheikh, 2022). Another study suggests that mushrooms can be used as bioinsecticidal agents (El Sheikh, 2021). China is the world's largest producer of edible mushrooms and truffles (El Sheikh and Hu, 2018). Moreover, China is the main producer and exporter of edible mushrooms, accounting for more than 70% of the world's annual output, with a total output of approximately 36 million tons, and export trade is up to more than 130 countries (Zhang et al., 2020; Yan et al., 2021). Bolete is a world-famous high-quality edible fungus. China is rich in natural resources and has great potential for its development and utilization. Yunnan has a unique climate and diverse vegetation, which provides a good environment for the growth of wild boletes, so it is one of the largest boletes production areas in the world (Yang et al., 2016, 2017). 882 species of wild edible mushrooms are known to be distributed in Yunnan, accounting for 44.1% of the 2000 species in the world and 91.3% of the 966 species in China. In 2021, Yunnan province produced 280,400 tons of wild edible mushrooms, among which the boletes from Chuxiong are nationally famous¹. Nevertheless, the price of boletes varies greatly depending on species, and some species of boletes may cause poisoning. At present, there are illegal traders in the market who use fake as real and use shoddy as good, which damages the interests of consumers and affects food safety at the same time (El Sheikh, 2018; Yan et al., 2022a). Therefore, rapid and accurate identification of bolete species is urgently required.

Mass spectrometry, chromatography, spectroscopy and other methods are often used to evaluate the quality of wild edible mushrooms. Malheiro et al. used headspace solid phase microextraction (HS SPME) and gas chromatography/ion trap mass spectrometry (GC/IT MS) to analyze the volatilities of edible mushrooms, and used stoichiometric methods to identify six species of edible mushrooms. The results showed that volatile substances could be used as an important classification basis (Malheiro et al., 2013). Marekov et al. (2012) used gas chromatography-mass spectrometer (GC-MS) to determine 31 fatty acid components of 15 edible mushrooms belonging to 9 genera and 5 families in Bulgaria. Stoichiometric methods were used to analyze the differences in fatty acid components and classify the species. Mohač ek-gro šev et al. used Fourier transform infrared spectroscopy (FTIR) to make spectral

analysis of the spores and fruiting bodies of more than 70 species of wild edible mushrooms belonging to 37 genera. The results showed that vibration spectroscopy could characterize the information of the content of polysaccharides in fungi, and the vibration spectra of spores of different species of the same genus and some of the fruiting bodies were very similar. The chemical composition of cap and stipe of the same fruiting body was quite different (Mohacek-Grosev et al., 2001). Each of these reported identification methods has advantages and disadvantages. Among them, morphological taxonomy identification technology is easily affected by subjective factors, and the identification accuracy is low. Mass spectrometry and chromatography are expensive and require expertise and technology (Mohacek-Grosev et al., 2001; Li et al., 2011, 2013). FTIR is an efficient and nondestructive discrimination technique, but its accuracy is not high because of spectral overlap.

In recent years, 2DCOS technology has been used to identify adulterants in food and herbal medicines. Walkowiak et al. (2019) successfully detected ginkgo biloba adulterants in dietary supplements using 2DCOS, and initially established an effective evaluation method for screening ginkgo biloba adulterants. Chen et al. (2018) used i2DCOS to visually identify adulterated medicinal materials. The results showed that i2DCOS technology has potential development in the visual identification and quality control of traditional Chinese medicine and other complex mixtures. Ma et al. (2016) used 2DCOS to accurately identify seven species of bolete mushrooms of the same genus, and applied the 2DCOS method to the identification of edible mushrooms. Nevertheless, these reports all use traditional machine learning methods that require complex data processing, high time costs, and accuracy needs to be improved. Recently, Dong et al. (2021b) used deep learning combined with 2DCOS and i2DCOS to identify the origin of *Boletus Edulis*, Yan et al. (2022b) used Resnet combined with 2DCOS to identify bolete species, Yue et al. (2021) used the same method to identify the part and region of the medicinal plant. However, the deep learning methods in these reports are single, only Resnet is employed, and there is a lack of comparison with other deep learning models. Moreover, only 2DCOS data were obtained, so fewer datasets can be used.

In this paper, the generation method of the 3DCOS dataset is proposed for the first time, which is expanded from single 2DCOS images to 1D MIR spectral images, 2DCOS images and 3DCOS images. Among them, 2DCOS and 3DCOS datasets include synchronous spectrum, asynchronous spectrum and integrative spectrum, which significantly increases the types of datasets and enables full verification of the model on each dataset. At the same time, we used two deep learning algorithms, Alexnet and Resnet, to process the acquired data sets one by one and compared them with the traditional SVM algorithm. Finally, the optimal method for identifying species of boletes was obtained. In addition, in order to analyze the species information and

¹ <http://www.yncoop.com/sxjs/zsjs/08196173330559241148>

sample distribution of boletes more intuitively, we developed a large-screen of boletes data visualization using ECharts, HTML, CSS and JavaScript technology. We hope that the results of this study and the large-screen visualization technology can be extended to the field of food and traditional Chinese medicine identification.

2. Materials and methods

2.1. Sample collection and preparation

We collected 1,707 samples of boletes which belong to eight species from 46 sampling sites in 13 regions of Southwest China. All samples were identified by Professor Honggao Liu from Yunnan Agricultural University. The appearances of these 8 boletes species are shown in [Supplementary Figure S1](#). The geographical information and other details of the sample are shown in [Supplementary Table S1](#). In the laboratory, the sundries on the samples were removed, and they were cleaned with a SY3200-T type ultrasonic cleaning instrument. After that, the materials were dried to constant weight with a dryer at medium temperature. Finally, they were pulverized with a FW-100 high speed grinder, the powder was sieved by an 80 μm mesh and stored for later use.

2.2. FT-MIR spectra acquisition

Precisely weigh 1.5 ± 0.2 mg powder of each sample and 150 ± 20 mg KBr powder in the ratio of 1: 100, put them into agate mortar, mix and grind them into fine powder. Then pour the fine powder into a pressed grinding tool to produce thin slices with uniform thickness. The sample was measured after preheating the Fourier transform infrared spectrometer for 30 min, and the sample was repeated twice to get the average spectrum. The interference of CO_2 and H_2O was deducted from the blank sample before scanning.

2.3. The spectral images acquisition

In this paper, 11,949 spectral images were obtained from 1707 boletes samples, including 1D MIR spectrum, 2DCOS (synchronous, asynchronous, integrative) and 3DCOS (synchronous, asynchronous, integrative). Among them, the theoretical basis of 2DCOS is as follows.

As shown in Equation 1, t represents the interval of perturbation, m represents the number of steps measured by the spectrum, the column vector P represents the dynamic spectral intensity at variable v (Yang et al., 2020).

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

Then, the synchronous 2DCOS (Φ), asynchronous 2DCOS (Ψ) and integrative 2DCOS (I) can be calculated as (Cheung, 2010; Chen et al., 2018; Dong et al., 2021b, 2022):

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

$$\Psi(v_1, v_2) = \frac{1}{m-1} P(v_1)^T \cdot N \cdot P(v_2) \quad (3)$$

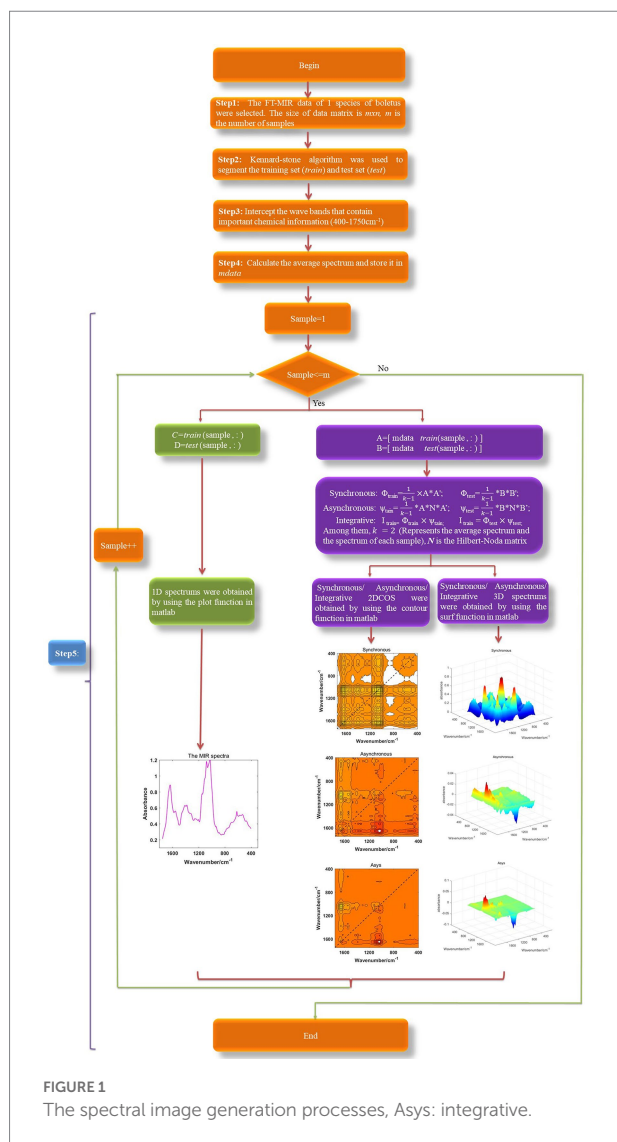
$$I(v_1, v_2) = \Phi(v_1, v_2) \cdot \Psi(v_1, v_2) \\ = \frac{1}{(m-1)^2} \left[P(v_1)^T \cdot P(v_2) \right] \cdot \left[P(v_1)^T \cdot N \cdot P(v_2) \right] \quad (4)$$

where N is the Hilbert-Noda matrix, which is defined as:

$$N_{jk} = \begin{cases} 0, & j = k \\ \frac{1}{\pi(k-j)}, & j \neq k \end{cases} \quad (5)$$

If there are m samples, then the 1D spectral data of all samples can be converted into an $m \times n$ data matrix, where n is the feature points contained in the spectrum. In this study, the matrix P ($m \times n$) contains two spectra data (that is, $m=2$), the average spectrum is the first and the i -th spectrum about each species of boletes is the second (Yang et al., 2013; Yang R. J. et al., 2014; Dong et al., 2021a).

The synchronous 3DCOS, asynchronous 3DCOS and integrative 3DCOS is generated in the same way as 2DCOS, but using 3D display. We intercept $1,750\text{--}400$ cm^{-1} fingerprint regions to generate various spectral images to reduce the computational burden. To improve efficiency and avoid generating spectral images one by one like Origin and OMNIC software, we wrote a program based on Matlab2017 to generate and automatically save these spectral images in bulk. All the spectral image generation processes are shown in [Figure 1](#). Firstly, all spectral data were obtained, and Kennard–Stone algorithm was used to divide the train set and test set. Then, various spectral images were obtained according to Equations 2–4 and different display methods. Finally, these images were separately saved to prepare for later model establishment.



2.4. Establishment of the identification model

2.4.1. SVM model

SVM is a classification method based on the realization of the minimization of constitutive risk. With strong generalization ability and good robustness, it is widely used in image recognition, text classification, biotechnology, face recognition and other fields (Hu et al., 2022). Based on the limited sample information, SVM seeks the best between the complexity of the model and the learning ability, in order to obtain the best generalization ability. It shows unique advantages in solving small samples, nonlinear and high-dimensional pattern recognition (Li et al., 2022). The SVM model has two very vital parameters c and g where c is the penalty coefficient. If c is too large or too small, the generalization ability of the model will deteriorate. If g increases, the number of support vectors increases which will affect the speed of training and prediction. Therefore, the SVM model has a good effect on

small samples, but is not suitable for large samples. In this paper, libsvm-3.21 toolbox was used to identify eight species of boletes.

2.4.2. Alexnet model

Alexnet model was proposed by Alex in 2012 and won the championship of the 2012 image recognition competition, which makes CNN become the core algorithm model in image classification.

As is shown in Supplementary Figure S2, the input data of the first layer is the original 64×64 size image, which is convolved by 96 11×11 convolution kernels, and each convolution of the original image generates a new pixel, batch normalization is then performed. These pixel layers are processed by the Maxpooling operation. The scale of the pooling operation is 3×3, and the step size of the operation is 2, so the size of the pooled image is $(64-3)/2 + 1 = 32$ and the image of 32×32×96 is obtained. The input data of the second layer is the pixel layer of 32×32×96 after the first pooling, which is convolved by 256 convolution kernels of 5×5 and then batch normalized. After the pooling operation with the scale of 3×3 and the operation step of 2, 16×16×256 images are obtained. Then, it passes through three convolution layers with a convolution kernel of 3×3 and stride of 1, and the number of convolution kernels is 384,384,256, respectively. The third Maxpooling operation is also with the scale of 3×3 and the operation step of 2. Therefore, the size of the input data in the sixth layer is 8×8×256. The operation results are output through 4,096 neurons, and then these 4,096 neurons in the seventh layer are fully connected to output 4,096 data. Lastly, the trained value is output after being fully connected to 8 neurons in the eighth layer.

In order to run Alexnet model, we used an 8-core 16G ECS cloud server, selected tensorflow deep learning framework developed by Google, installed anaconda3 deep learning environment, and used matplotlib library to visualize output data. All spectral images were uniformly clipped to 64×64 before input to the model.

2.4.3. Resnet model

ResNet was proposed by Microsoft Research's Kaiming He in 2015, and it won first place in the ImageNet competition Classification task. It solves the problems of overfitting, weight attenuation and gradient disappearance caused by the deepening of the layer of CNN, and has excellent performance.

According to the Resnet theory, we first construct the identity residual block and the convolution residual block. When the dimension of input data is consistent with the dimension of output data, the identity residual block (identity block) is employed, while the convolution residual block (conv block) is used when the dimension of input data is inconsistent with that of output data. Then a 12 layers Resnet network display in Supplementary Figure S3 is established based on these two residual blocks. 1D, 2DCOS, and 3DCOS images were fed into the model as input data, and convolution, batch normalization, Relu nonlinear activation were performed first. Then three identity blocks and two conv blocks are used for processing. After that, global average pooling (GAP)

and flatten were carried out. Finally, softmax was used to output the discrimination results.

Similarly, we used an 8-core 16G ECS cloud server, selected Amazon's Mxnet deep learning framework, installed anaconda3 deep learning environment, and used matplotlib library to visualize output data. All spectral images were uniformly clipped to 64×64 before input to the model.

2.4.4. Data set partition

In classical machine learning, the ratio of training set to test set is 7:3 or 8:2. In general, the larger the proportion of samples in the training set, the more data information it contains, and the better the model training results will be. In order to enhance the accuracy of SVM model, the ratio of 8:2 was selected to divide the training set and the test set. For deep learning based on image processing, the generalization ability of its model is essential. In order to evaluate the generalization error of the model and make a choice through the test set, the training set and test set of Resnet and Alexnet models are divided into 7:3. The specific division results are illustrated in [Supplementary Table S2](#).

2.5. The identification strategy of boletes species

The identification strategy of eight boletes species is shown in [Supplementary Figure S4](#). There are two modeling methods. One method is to directly use spectral data for SVM modeling, and the other method is to convert spectral data into images, then conduct deep learning modeling based on image processing. We convert the spectral dataset into seven image datasets including 1D, 2D (synchronous, asynchronous, integrative), and 3D (synchronous, asynchronous, integrative). Then, seven Resnet and seven Alexnet models were established respectively, and the final identification results were output. Therefore, a total of 15 recognition models based on traditional machine learning and deep learning are established in this paper.

2.6. The large-screen visualization

Data visualization large-screen is a means to transform boring, professional and unintuitive data content into interesting, simple and intuitive content with the help of visual language expression and convey it to large-screen viewers. With the support of the current new technology, data visualization is not only visible, but also communicative and interactive. The essence of this technique is the mapping of data space to graphics space ([Yao, 2021](#)).

In this paper, based on ECharts, HTML, CSS and JavaScript technology, a large-screen of boletes information visualization was developed to display the origin, sampling site, species and other information of boletes from various aspects. The layout and proportion of large-screen panels are shown in

[Supplementary Figure S5](#). The front-end editor uses Visual Studio Code. ECharts is a pure JavaScript diagram library, the bottom layer depends on the lightweight Canvas class library ZRender, based on BSD open source protocol, is a very excellent visual front-end framework. HTML is a markup language used to design web pages. The files written by the browser are interpreted and executed. On the HTML page, you can write program segments nested in scripting languages. CSS is a computer language used to represent the style of HTML files. It can not only statically decorate web pages, but also dynamically format the elements of web pages with various scripting languages. JavaScript is a scripting language embedded in HTML and used in the browser to add interactive behavior to the HTML page. It is directly embedded in the HTML page and interpreted by the browser to execute the code without precompilation.

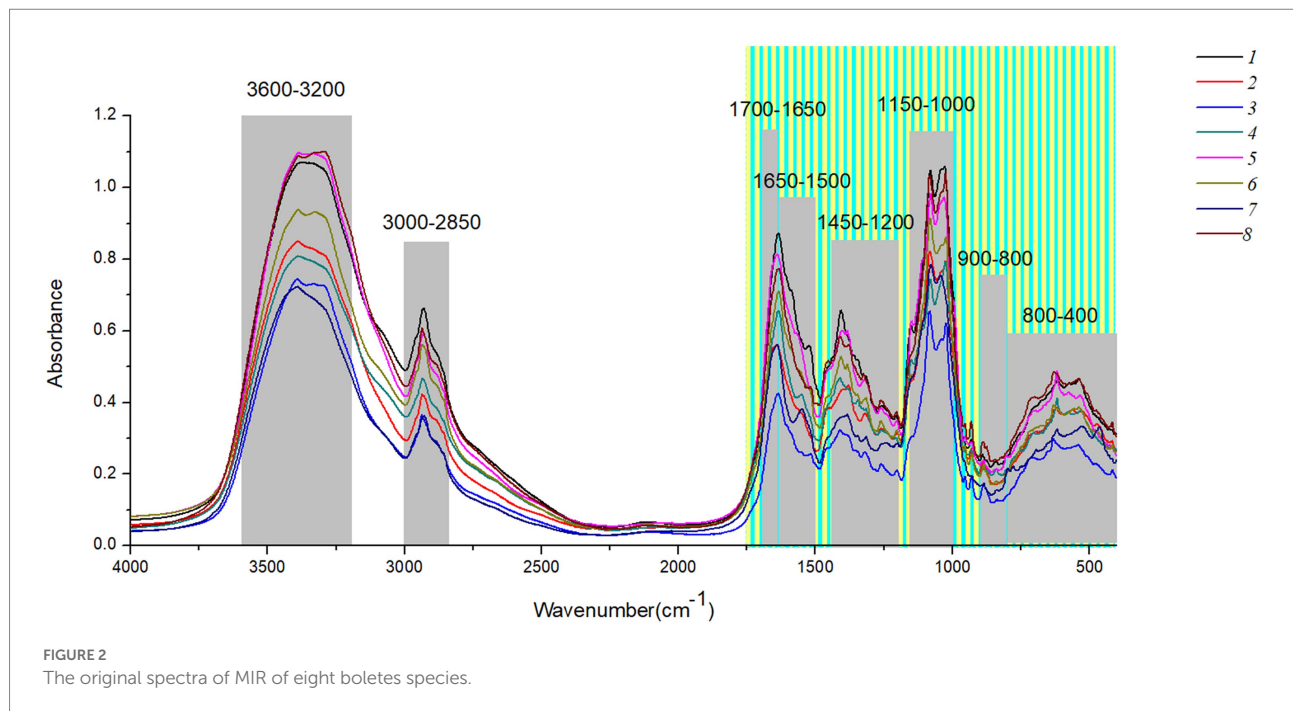
3. Results and discussion

3.1. MIR spectral analysis

[Figure 2](#) shows the original spectra of MIR of eight boletes species. The band 3,600–3,200 cm^{-1} is caused by O-H stretching, and the peak 3,342 cm^{-1} represents strong water interference ([Hirri et al., 2016](#); [Wang, 2020](#)). The 3,000–2,850 cm^{-1} band is caused by the stretching of lipid methylene group and the pyranose ring, among which 2,928 cm^{-1} and 2,855 cm^{-1} are the absorption peaks of fatty acid components ([Zhao et al., 2015](#); [He, 2019](#)). The 1,700–1,000 cm^{-1} band contains the organic material, C-C stretching, C-O-H, C-H and CH_2 bending ([Mohacek-Grosecv et al., 2001](#); [Nie et al., 2007](#); [Dong et al., 2021b](#)). 1,700–1,650 cm^{-1} is mostly protein, 1,650–1,500 cm^{-1} mainly caused by amide I and amide II. The main components of 1,450–1,200 cm^{-1} are proteins, fatty acids, and polysaccharides ([Chen et al., 2021](#); [Dong et al., 2021b](#)). In addition, the 1,000–1,200 cm^{-1} band is caused by carbohydrates, and the two peaks 1,032 cm^{-1} and 1,080 cm^{-1} mainly contain chitin ([Yang T. W. et al., 2014](#); [Zhang et al., 2018](#)). Finally, the interval of 900–400 cm^{-1} can be effectively used for chemometric analysis, among which the band of 900–800 cm^{-1} mainly contains glucan and mannan ([Qi et al., 2018](#); [Chen et al., 2021](#)). [Supplementary Table S3](#) summarizes the Peak assignments on the FT-MIR spectra of boletes in different characteristic bands.

3.2. Spectral image analysis of three dimensions

In this paper, Matlab2017 was used to generate 1D MIR spectral images in batches, and 2DCOS and 3DCOS images were generated according to relevant theories. The spectral images of three dimensions about eight boletes species are shown in [Figure 3](#). It can be seen that the 1D spectra have obvious absorption peaks near wave numbers 3,300, 2,900, 1,640, 1,420, 1,070, 1,040, and 650 cm^{-1} . However, the absorption peak is



difficult to show due to spectral overlap, and the spectra are too similar to distinguish the eight boletes species. 2DCOS can show more information because the peak is expanded by the correlation operation. The synchronous 2DCOS has strong auto-peaks at 1650, 1080, 550 cm^{-1} and obvious cross-peaks at 1650, 1500, 1,080, and 550 cm^{-1} . Asynchronous 2DCOS has no auto-peak, but it has strong cross-peaks at 1650, 1080 and 550 cm^{-1} . The integrated 2DCOS only has two cross-peaks at 1650 and 1,080 cm^{-1} , which contains relatively little information. 3DCOS is a 3D display of 2DCOS, and the size and number of peaks can be seen more intuitively from the 3DCOS. Among them, the synchronous 3DCOS has two large auto-peak at 1080 and 550 cm^{-1} , a relatively small auto-peak at 1650 cm^{-1} , and two large cross-peaks at 1080 and 1,500 cm^{-1} . In addition, there are several smaller cross-peaks. Both asynchronous 3DCOS and integrated 3DCOS have two large cross peaks, one of which is positive and the other is negative. Therefore, from the distribution of peak values, synchronous 3DCOS can better identify boletes species. However, there are exceptions, such as *Retiboletus griseus*, which has many relatively large auto-peaks and cross-peaks.

3.3. Discrimination results of SVM model

In order to establish the SVM model, we selected 1,367 samples (80%) as the training set, and the remaining 340 samples (20%) as the test set. The Kennard–Stone algorithm was used to partition the dataset. [Supplementary Figure S6A](#) displays the optimal SVM parameters obtained by grid search method, among which the best $c = 1.049 \times 10^6$, the best $g = 0.125$, and the accuracy of the training set is 92.1%. The classification

results of 340 test set samples by SVM algorithm are shown in [Supplementary Figure S6B](#), 331 samples are accurately classified, only 9 samples are incorrectly classified, and the classification accuracy of the test set is 97.4%. It can be clearly seen that one sample of the third species was misclassified to the sixth species, five samples of the sixth species were misclassified to the first species, two samples of the sixth species were misclassified to the seventh species, and one sample of the seventh species was misclassified to the fifth species. Consequently, the algorithm has a low recognition rate for the sixth species (*Boletus bicolor*). Therefore, we need to find better identification methods.

3.4. Discrimination results of Alexnet model

[Figure 4](#) is the discrimination results of Alexnet model, where X is 1D MIR spectrum, Y is 2DCOS spectrum, Z is 3DCOS spectrum, A is the accuracy curves, B is the cross-entropy cost function, C is the confusion matrix. The number of samples in the training set was 1,195 (70%), and the number of samples in the test set was 512 (30%). In the Alexnet model, the initial learning rate is 0.01, the decay rate is 0.99, and the decay step is 3. To improve the convergence speed of the model, the Nesterov gradient descent method was used, and the learning rate was 0.001. Underfitting may occur if the number of epochs is too small. Conversely, if there are too many epochs, overfitting is likely to occur. Therefore, we use early stopping to select the number of epochs and stop training when the model's performance on the test set does not increase. The parameter for early stopping is set

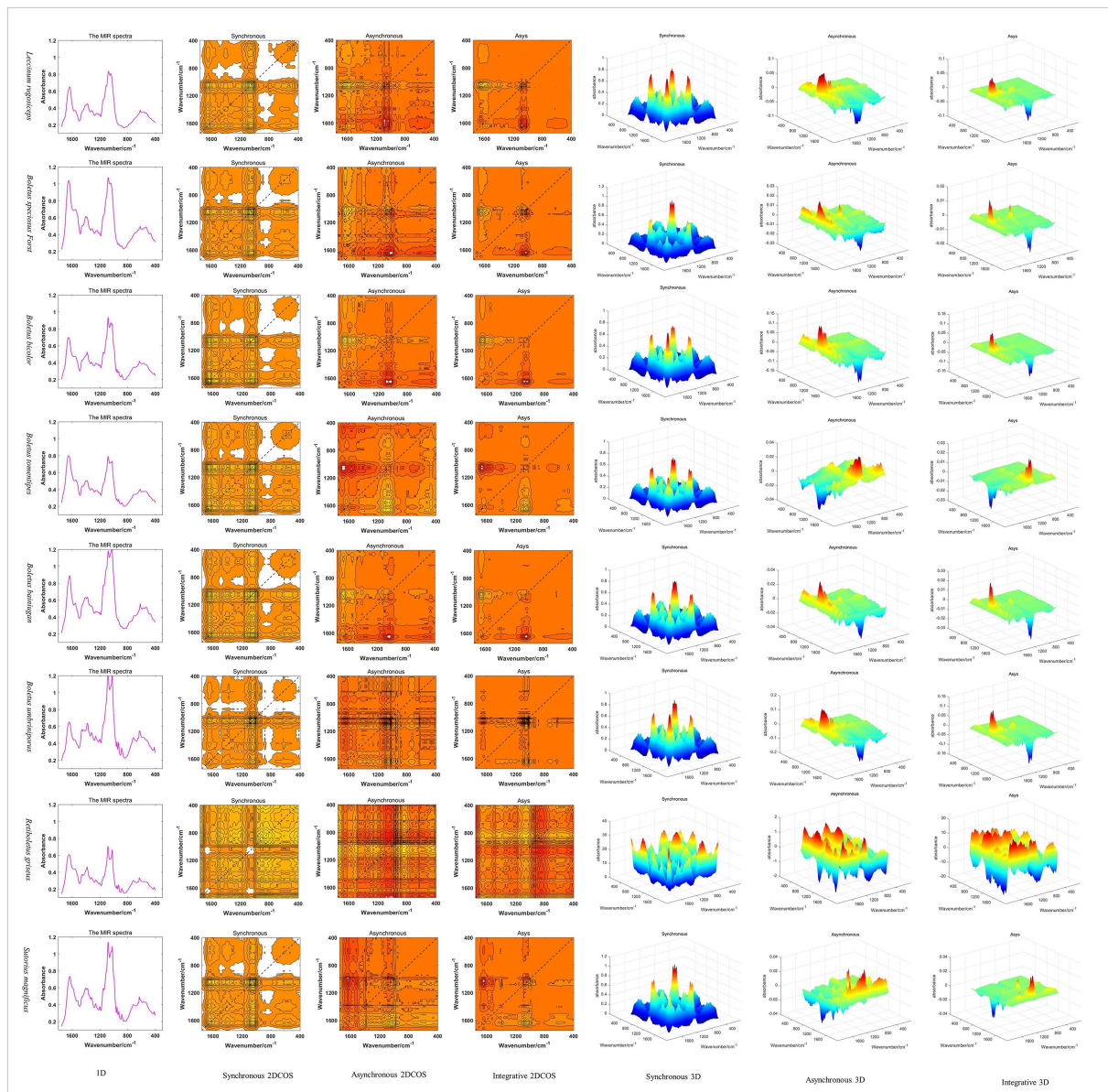


FIGURE 3 The spectral images of three dimensions about eight boletes species.

as follows: monitor is the accuracy rate of the test set, min_delta is 0.001, patience is 5.

It can be seen that early stopping occurs on the 1D MIR spectrum dataset when the epoch is 10. The highest accuracy rate of the training set is 49.8%, while that of the test set is only 24.4%. The cross-entropy cost function decreases at first and then increases sharply, and the minimum loss value is 1.441, which is comparatively large. The confusion matrix showed that 68 samples of the eighth species (*Leccinum rugosiceps*) were correctly identified, and 444 samples of the other seven species were incorrectly identified. As a result, the model's performance is poor in both accuracy and loss value. Early stopping occurs at the 38th epoch on the synchronous 2DCOS dataset. The accuracy rate is

98.9% on the training set and 100% on the test set. The minimum loss value is 0.038. All 512 samples in the confusion matrix were accurately classified, so the model performed well on this data set. For the asynchronous 2DCOS dataset, early stopping occurs at the 38th epoch. The accuracy rate is 37.6% on the training set and 43.2% on the test set. In addition, the minimum loss value is 1.618. The confusion matrix shows that 212 samples were correctly identified, and 300 samples were incorrectly identified. Therefore, the model of this data set is poor. The model stopped early at the 85th epoch on the Integrative 2DCOS dataset. The accuracy rate in the training set and test set is 69.6% and 79.3%, respectively. The minimum loss value is 0.817. In the confusion matrix, 100 samples were incorrectly identified and the rest were

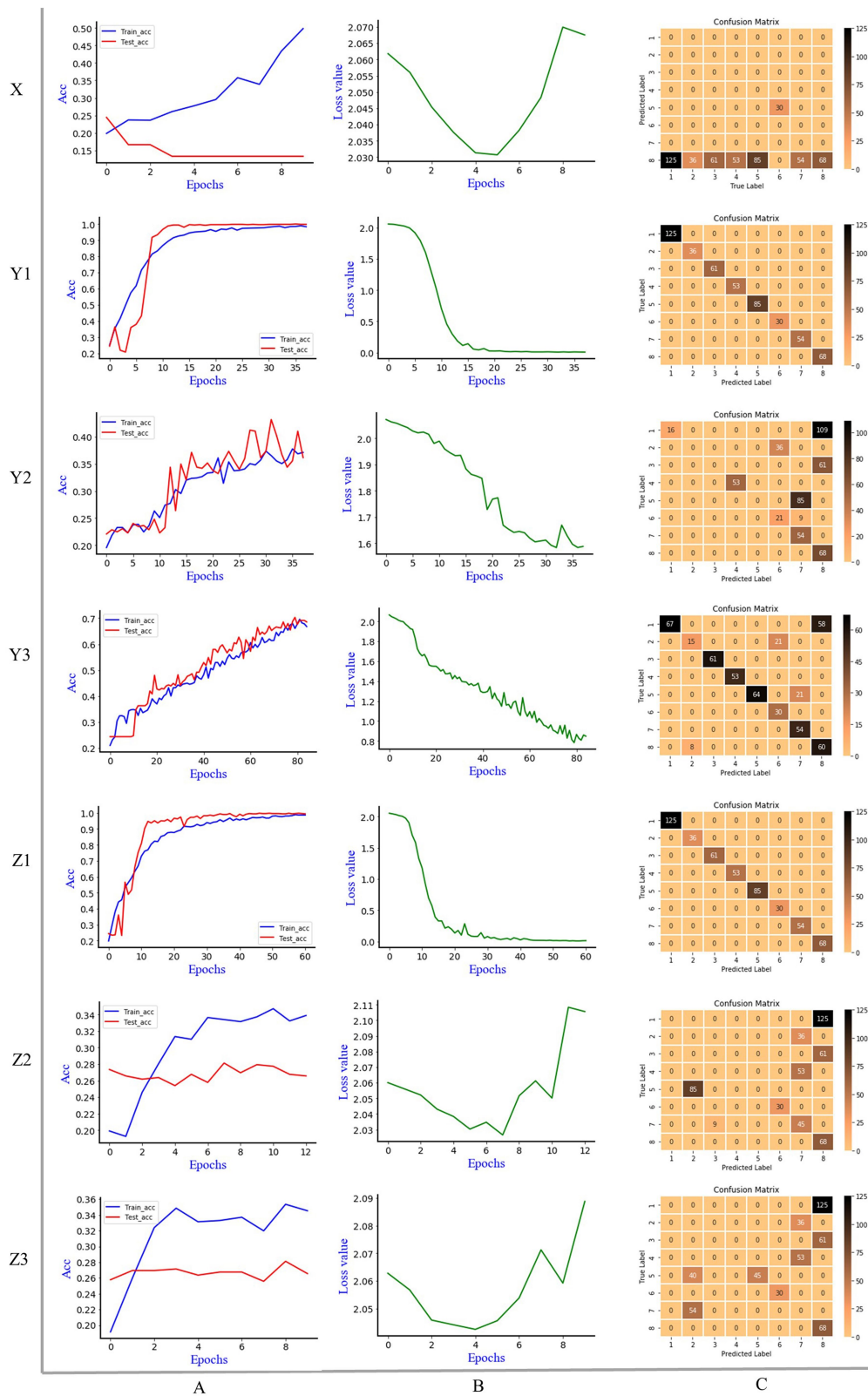


FIGURE 4
 The discrimination results of Alexnet model, X: 1D MIR spectrum; Y1: synchronous 2DCOS; Y2: asynchronous 2DCOS; Y3: integrative 2DCOS; Z1: synchronous 3DCOS; Z2: asynchronous 3DCOS; Z3: integrative 3DCOS; A: accuracy curves; B: the cross-entropy cost function; C: the confusion matrix.

correctly identified. The performance of the model on this dataset is better than that of asynchronous 2DCOS, but markedly worse than that of synchronous 2DCOS. Consequently, synchronous 2DCOS has the best effect in the three 2DCOS datasets and can be used for the identification of boletes species.

On the synchronous 3DCOS dataset, the early stop occurs at the 61st epoch, and the accuracy of the training and test sets is 98.9% and 99.8%, respectively. The distribution of 512 samples in the test set on the confusion matrix is exactly correct. With the increase of epochs, the loss value gradually decreases, and the minimum value reaches 0.039, which is close to zero, indicating that the error of the model is small. It shows that the model performs well on this dataset and can be used for the identification of boletes species. For the asynchronous 3DCOS dataset, the model stops early at the 13th epoch. At this time, the accuracy is 34.7% on the training set and 28.1% on the test set, and the minimum loss value is 1.702, implying that the error of the model is large. The confusion matrix showed that only 143 samples of the sixth (*Boletus bicolor*), seventh (*Boletus speciosus*) and eighth (*Leccinum rugosiceps*) species were correctly identified. The accuracy of the training set and test set on integrative 3DCOS are 35.4% and 28.1%. Early stopping occurred at the 10th epoch, the minimum loss value was 1.652, and only 143 samples were correctly identified in the confusion matrix. These results show that the performance of the model on integrative 3DCOS and asynchronous 3DCOS datasets is similar, with low accuracy, high loss value and poor discrimination effect. Therefore, among the three 3DCOS datasets, Alexnet had the best performance in the synchronous 3DCOS dataset, which was suitable for the identification of boletes species.

3.5. Discrimination results of Resnet model

Figure 5 is the discrimination results of Resnet model, where L is 1D MIR spectrum, M is 2DCOS spectrum, N is 3DCOS spectrum, D is the accuracy curves, E is the cross-entropy cost function, F is the confusion matrix. Same as the Alexnet model, the number of samples in the training set was 1,195 (70%), and the number of samples in the test set was 512 (30%). In the Resnet model, stochastic gradient descent (SGD) was used to modify the parameters of the optimization model to minimize the cross-entropy loss, L2 regularization was used to avoid overfitting, the weight decay factor was set to 0.0001, and the learning rate was 0.01.

When the epoch is 15, the accuracy of Resnet model on the training set is 100%, and the accuracy on the test set is 29%, and the minimum loss value is 0.014. Although the loss value is close to zero, the low accuracy of the test set leads to the poor effect of the model, which is not suitable for the identification of boletes species. This may be caused by overfitting due to the comparatively complex structure of the model due to the small amount of data. The confusion matrix demonstrated that only 18 samples of the

second species (*Sutorius magnificus*) and 68 samples of the eighth species (*Leccinum rugosiceps*) were correctly identified, while the rest were incorrectly identified. Figure 5 shows that the Resnet model performs better on the synchronous 2DCOS dataset. When the epoch is 15, the accuracy of the training set and test set is 100%, and 512 samples are correctly identified in the confusion matrix. At this time, the loss value is 0.052, which is close to zero. This indicates that the error of the model is small. Therefore, the synchronous 2DCOS model has high accuracy, small loss value, and strong generalization ability. The model can be used for the identification of boletes species. When the epoch is 39, the accuracy of the Resnet model on the training set and test set is 100% and 77%, respectively, on the asynchronous 2DCOS dataset, and the loss value is 0.041. The confusion matrix indicated that 313 samples were correctly identified and the remaining 189 samples were incorrectly identified. Although the loss value is small, the accuracy of the test set is low and the generalization ability is poor, which leads to the unavailability of the model. Similarly, when the epoch is 39, the accuracy of the Resnet model on the integrative 2DCOS dataset is 100% and 77% for the training and testing sets, and the loss value is 0.064. The confusion matrix revealed that 339 samples were correctly identified and the remaining 173 samples were incorrectly identified. Therefore, the generalization ability of this model is poor and it is not available. Therefore, in the three datasets of 2DCOS, Resnet performs perfectly on the synchronous 2DCOS dataset, while the test set accuracy on the other two datasets is relatively low.

Compared with synchronous 2DCOS, the accuracy of the Resnet model in the training and testing sets on the synchronous 3DCOS dataset is also 100%. Although the loss value is 0.1, which is higher than that of the synchronous 2DCOS dataset, its epoch is only 15, which saves the computational complexity and time cost. All test samples are correctly identified in the confusion matrix. As a result, the Resnet model shows different advantages on synchronous 2DCOS and synchronous 3DCOS datasets, which are both perfect models. When the epoch is 39, the loss value of the Resnet model on the asynchronous 3DCOS dataset is 0.07. The accuracy was 100% on the training set and 58% on the test set. In the confusion matrix, 297 samples were accurately identified and 215 samples were incorrectly identified. In the case of the same epoch value, the loss value of the Resnet model on the integrative 3DCOS dataset is 0.09. The accuracy on the training set was 100%, while the accuracy on the test set was 65%. In the confusion matrix, 308 samples were correctly identified and 204 samples were incorrectly identified. Therefore, the effect of Resnet model on integrative 3DCOS and asynchronous 3DCOS datasets is similar with relatively low accuracy, while the effect of Resnet model on synchronous 3DCOS datasets is the best, which can be applied to actual species discrimination.

In addition, by analyzing the confusion matrix, we found that the Alexnet and Resnet models performed best in identifying the eighth species (*Leccinum rugosiceps*) in all data sets, with almost no misjudgment. This may be related to the chemical characteristics, molecular structure, phenological characteristics

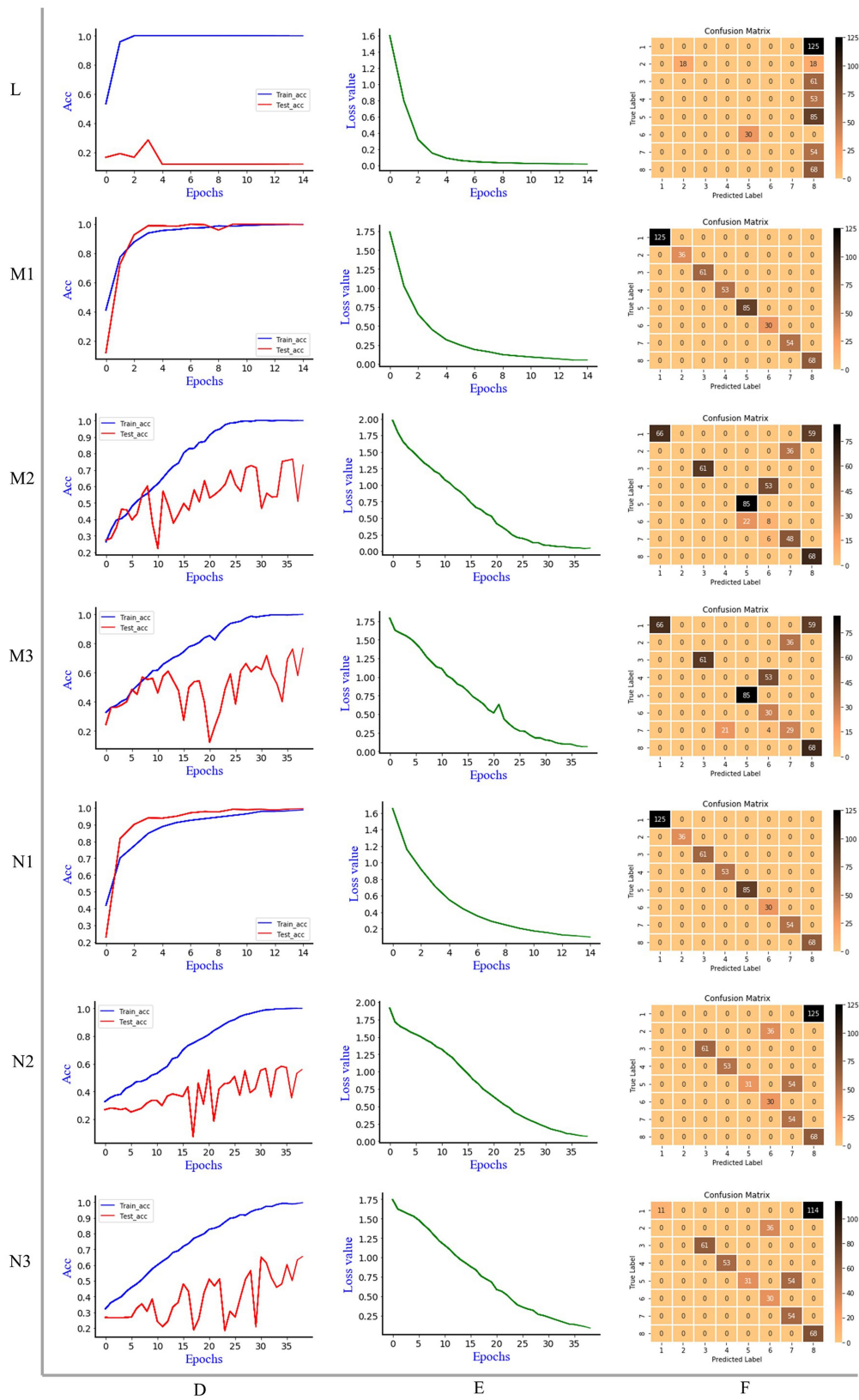


FIGURE 5

The discrimination results of Resnet model. L: 1D MIR spectrum; M1: synchronous 2DCOS; M2: asynchronous 2DCOS; M3: integrative 2DCOS; N1: synchronous 3DCOS; N2: asynchronous 3DCOS; N3: integrative 3DCOS; D: accuracy curves; E: the cross-entropy cost function; F: the confusion matrix.

TABLE 1 The comparison of model results.

Methods	Dimensions of images	Image types	Epoch	Loss value	Train set accuracy	Test set accuracy	Time costs (minutes)
SVM	/	/	/	/	92.10%	97.35%	1,092
Alexnet	1D	The MIR spectrum	10	2.032	49.80%	24.40%	21.2
		synchronous	38	0.038	98.90%	100.00%	66.3
		asynchronous	38	1.618	37.60%	43.20%	68.2
	2D	integrative	85	0.817	69.60%	79.30%	183.1
		synchronous	61	0.039	98.90%	99.80%	152.1
		asynchronous	13	1.702	34.70%	28.10%	23.2
	3D	integrative	10	1.652	35.40%	28.10%	20.3
		synchronous	15	0.014	100%	29%	43.1
		asynchronous	39	0.041	100%	77%	178.6
Resnet	1D	The MIR spectrum	15	0.014	100%	29%	43.1
		synchronous	15	0.052	100%	100%	41.5
		asynchronous	39	0.041	100%	77%	178.6
	2D	integrative	39	0.064	100%	77%	125.2
		synchronous	15	0.1	100%	100%	66.6
		asynchronous	39	0.07	100%	58%	157.3
	3D	integrative	39	0.09	100%	65%	152.8
		synchronous	15	0.1	100%	100%	66.6
		asynchronous	39	0.07	100%	58%	157.3
3D	integrative	39	0.09	100%	65%	152.8	
	synchronous	15	0.1	100%	100%	66.6	
	asynchronous	39	0.07	100%	58%	157.3	

and growing environment of this species, which lead to the obvious characteristics of its spectral image compared with other species and make it easier to identify.

3.6. The comparison of model results

Table 1 shows the results of species identification of boletes species using various models. It can be seen that the Resnet model has the best effect on the synchronous 2DCOS dataset. The accuracy of the training and test sets is 100% which shows that the model has high accuracy. The loss value is close to zero, which indicates that the error of the model is small. The epoch value is only 15 and the time cost is 41.5 min, which proves that the model works quickly. The external verification is all correct, which indicates that the model has high reliability and strong generalization ability. In addition, the deep learning model processes spectral images directly without complex data parsing, which is simple and convenient. The Resnet model also performs well on the synchronous 3DCOS dataset. The accuracy of the training and testing sets is 100%, and the epoch value is 15. Nevertheless, the loss value is 0.1 and the time cost is 66.6 min. It takes more time to achieve the same effect as the synchronous 2DCOS dataset, and the loss value also increases. The accuracy of the Alexnet model on the synchronous 2DCOS dataset is 98.9% for the training set and 100% for the test set, and the epoch value is 38. However, the loss value is 0.038, and the time cost is 66.3 min. The training effect of the model on the

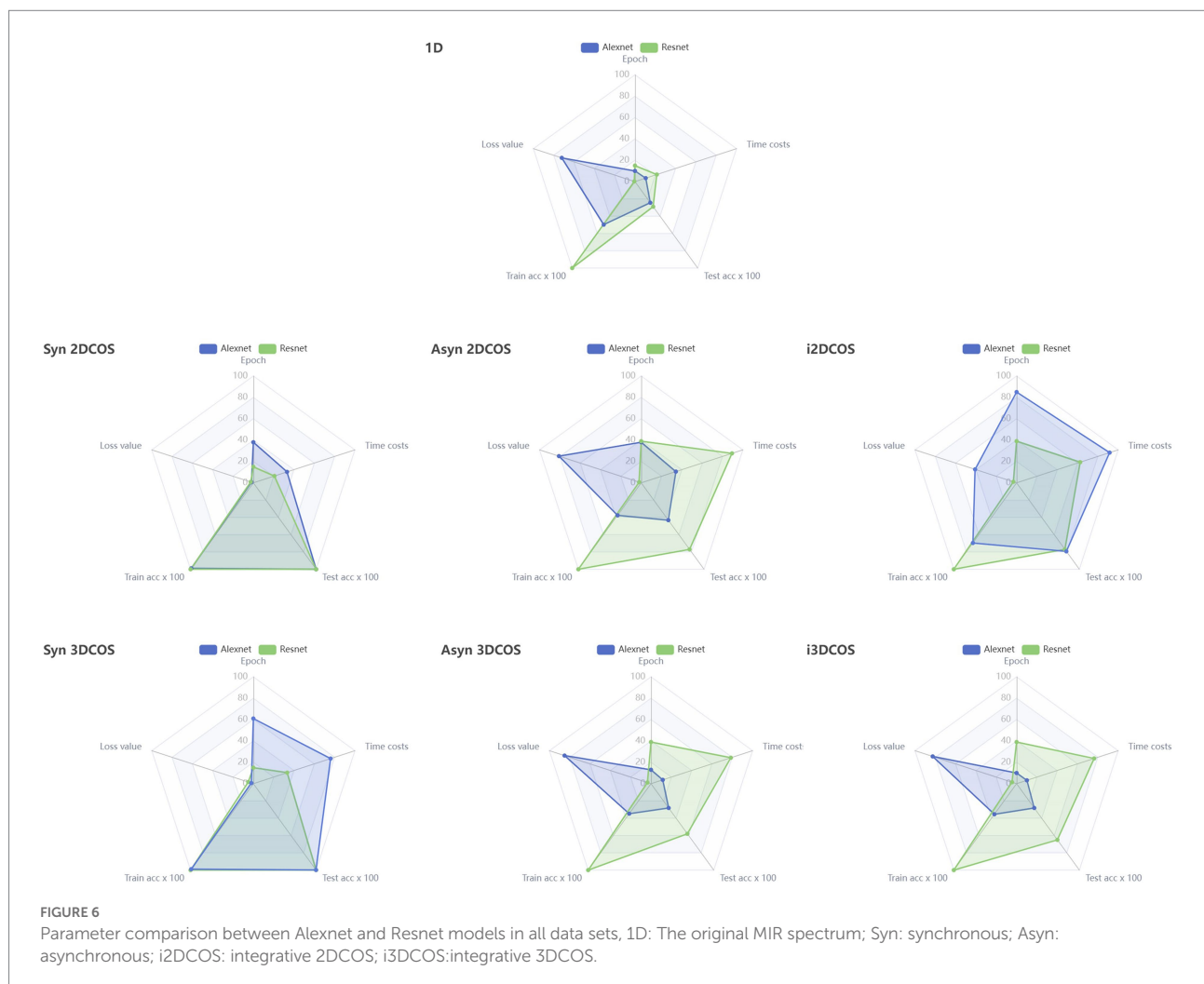
synchronous 3DCOS dataset is similar to that on the synchronous 2DCOS dataset. The accuracy of the training set is 98.9%, the accuracy of the test set is 99.8%, and the loss value is 0.039, but the epoch expands to 61. The time cost increases to 152.1 min, more than twice as much as on the Synchronous 2DCOS dataset. The above four models can be used for the identification of boletes species. Among them, the Resnet model has the highest accuracy, the best effect, the smallest error and the least time cost on the synchronous 2DCOS dataset. In this case, the sensitivity of the model is 100%, which indicates that the model is relatively stable. As shown in Supplementary Figure S1, four boletes species (*Sutorius magnificus*, *Boletus speciosus*, *Boletus bainiugan*, *Boletus bicolor*) have similar appearance, which are difficult to distinguish in morphology. However, the Resnet model on the synchronous 2DCOS dataset can be used to identify them accurately, and the external verification accuracy is 100%, which can be inferred that the sensitivity of the model is very high.

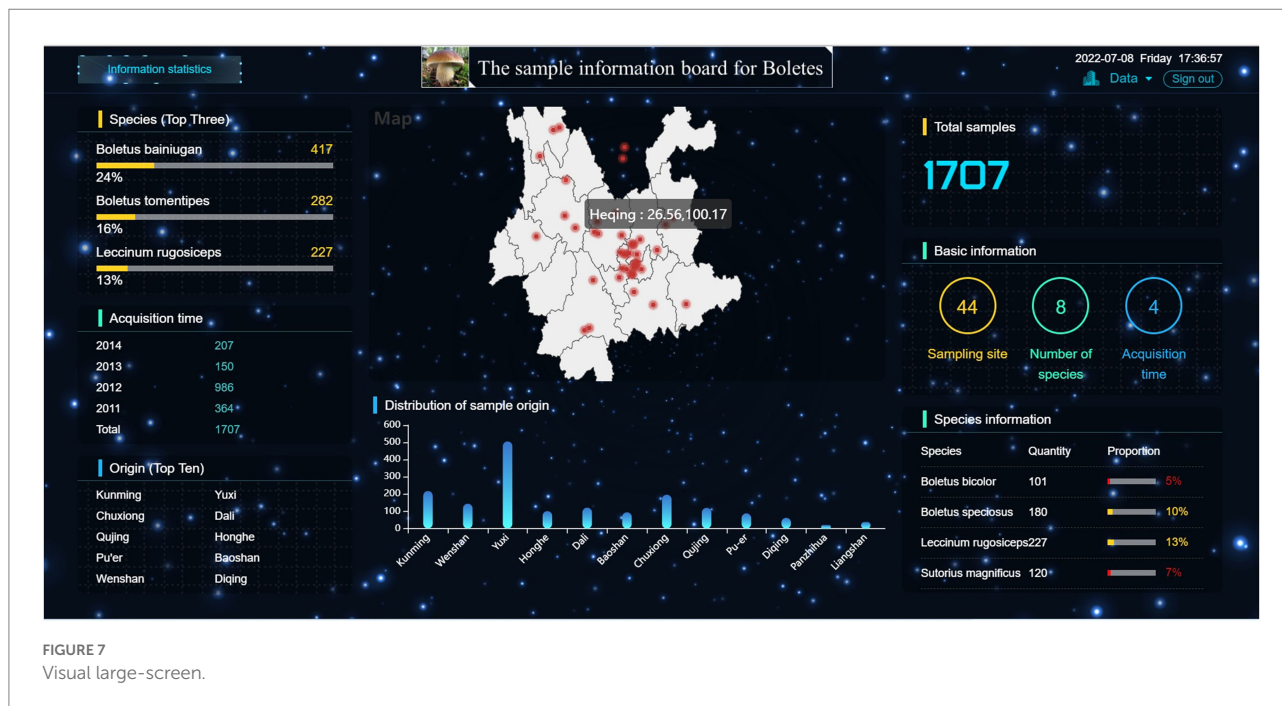
The SVM model does not process the spectral images, but trains the original spectral data directly. Although the accuracy of the training set is 92.1% and the test set is 97.35%, it is much higher than the accuracy of Resnet and Alexnet models on asynchronous 2DCOS, Integrative 2DCOS, asynchronous 3DCOS and Integrative 3DCOS datasets. However, under the same hardware conditions, the training time is as high as 1,092 min, so the time cost of this model is too high, and it is not recommended to be used. In general, Resnet model is better than Alexnet and SVM model.

Figure 6 compares the parameters of Resnet and Alexnet models on different data sets through radar charts. We can see that these two models perform best on synchronous spectral images (including 2DCOS and 3DCOS), where Resnet has smaller epochs and lower time complexity compared to Alexnet. The two models performed the worst on the original MIR spectrum, and their test set had the lowest accuracy. Although the model has a high accuracy in the training set on the other four data sets, the accuracy in the test set is not high, and Alexnet has a high loss value, and the time cost of Resnet is relatively high. In summary, synchronous 2DCOS and synchronous 3DCOS datasets perform better than other datasets. This may be because the synchronous spectral image contains more auto-peaks and cross-peaks, which leads to its obvious features and makes it easier to adopt the deep learning method for image recognition. However, the original MIR Spectrum image is a simple curve, which contains less feature information and its feature peaks overlap, so its recognition effect is the worst.

3.7. Visualization analysis

Figure 7 displays the detailed information of the samples in this research. The right side of the large-screen shows that the sample size of this study is 1707, including 44 sampling sites, eight boletes species and 4 sample collection years. The map module in the middle shows the sampling sites and their latitude and longitude. It can be seen that the sampling sites practically cover the whole Yunnan Province, with the most dense distribution in Central Yunnan, followed by the Northwest Yunnan, and fewer sampling sites in Northeast Yunnan. In addition, it can be seen that there are two sampling sites in southern Sichuan Province. The main reason is that boletes is abundant in Central Yunnan and Northwest Yunnan due to the influence of climate and environmental conditions. This is also proved by the Origin (Top 10) and distribution of sample origin modules. From the distribution of sample origin modules, it can be seen that Yuxi has the largest number of





samples, more than 500, while Liangshan, Panzhihua and Diqing have the smallest number of samples, among which Panzhihua has only 20 samples. In the Species (Top Three) module on the left, it can be seen that *Boletus Bainiugan* has the largest number, with 417 samples, accounting for 24% of the total sample size, followed by *Boletus Tomentipes* and *Leccinum Rugosiceps*. They accounted for 16 and 13% of the total sample size, which implied that Yunnan was rich in these types of boletus. From the perspective of acquisition time, the largest sample size was collected in 2012, which may be due to the abundant rainfall in Yunnan from July to September 2012, which promoted the high yield of boletes.

4. Conclusion

In this paper, based on deep learning, a method using 2DCOS and 3DCOS spectral image processing and recognition technology is proposed to identify the species of boletes. Among them, 3DCOS was proposed for the first time. Experimental results show that the method is rapid, accurate and effective. In our study, 15 models were established through three algorithms and eight datasets. Although the accuracy rate of SVM algorithm was higher than 90%, the time cost of is too high, and it needed to run for more than 18h on our hardware. The remaining 14 models were based on deep learning method, and the results show that the Resnet algorithm is the best model in the synchronous2DCOS dataset with 100% accuracy, small loss value and low time complexity. In addition, the detailed information of all samples is displayed visually on a large visual

screen, and the correlation analysis is carried out to obtain the relevant conclusion of the sample distribution. This research overcomes previous problems such as single algorithm, lack of data sets and less choice caused by fewer models. In summary, the method recommended in this paper is effective and reliable, and this method can be applied to other discrimination fields in future research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

J-QL: conceptualization. Y-ZW: ethodology and writing—original draft, formal analysis, data curation, and project administration. H-GL: resources and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supposed by the National Natural Science Foundation of China (Grant number: 32160735).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Chen, J., Li, J. Q., Li, T., Liu, H. G., and Wang, Y. Z. (2021). Application of UV-vis and infrared spectroscopy on wild edible bolete mushrooms discrimination and evaluation: a review. *Crit. Rev. Anal. Chem.*, 1–18. doi: 10.1080/10408347.2021.1984870 [Epub ahead of print]
- Chen, J. B., Wang, Y., Rong, L. X., and Wang, J. J. (2018). Integrative two-dimensional correlation spectroscopy (i2DCOS) for the intuitive identification of adulterated herbal materials. *J. Mol. Struct.* 1163, 327–335. doi: 10.1016/j.molstruc.2018.02.061
- Cheung, P. C. K. (2010). The nutritional and health benefits of mushrooms. *Appl. Physiol. Nutr. Metab.* 35, 292–299. doi: 10.1111/j.1467-3010.2010.01859.x
- Dong, J. E., Zhang, J., Li, T., and Wang, Y. Z. (2021a). The storage period discrimination of bolete mushrooms based on deep learning methods combined with 2DCOS and i2DCOS. *Front. Microbiol.* 12:771428. doi: 10.3389/fmicb.2021.771428
- Dong, J. E., Zhang, S., Li, T., and Wang, Y. Z. (2022). 2DCOS combined with CNN and blockchain to trace the species of boletes. *Microchem. J.* 177:107260. doi: 10.1016/j.microc.2022.107260
- Dong, J. E., Zuo, Z. T., Zhang, J., and Wang, Y. Z. (2021b). Geographical discrimination of boletus edulis using two dimensional correlation spectral or integrative two dimensional correlation spectral image with ResNet. *Food Control* 129:108132. doi: 10.1016/j.foodcont.2021.108132
- El Sheikha, A. F. (2018). *Molecular Techniques in Food Biology: Safety, Biotechnology, Authenticity & Traceability*, John Wiley & Sons Ltd., Chichester, UK, ISBN 978-1-1193-7460-2, pp. 472.
- El Sheikha, A. F. (2021). Can mushrooms and their derivatives be efficient bioalternatives to conventional synthetic insecticides? *Int. J. Med. Mushrooms* 23, 1–14. doi: 10.1615/IntJMedMushrooms.2021041319
- El Sheikha, A. F. (2022). Nutritional profile and health benefits of Ganoderma lucidum Lingzhi, Reishi, or Mannentake as functional foods: current scenario and future perspectives. *Foods* 11:1030. doi: 10.3390/foods11071030
- El Sheikha, A. F., and Hu, D. M. (2018). How to trace the geographic origin of mushrooms? *Trends Food Sci. Technol.* 78, 292–303. doi: 10.1016/j.tifs.2018.06.008
- He, Z. D. (2019). Contribution of China's edible fungi export trade to economic growth. *Edible Fungi of China* 38, 73–76. doi: 10.13629/j.cnki.53-1054.2019.10.022
- Hirri, A., Bassbasi, M., Platikanov, S., Tauler, R., and Oussama, A. (2016). FTIR spectroscopy and PLS-DA classification and prediction of four commercial grade virgin olive oils from Morocco. *Food Anal. Methods* 9, 974–981. doi: 10.1007/s12161-015-0255-y
- Hu, J., Wang, X. T., Mao, Y. M., and Liu, W. (2022). Parallel SVM algorithm based on Relief and bacterial foraging optimization algorithm. *Appl. Res. Comput.* 39, 447–455. doi: 10.19734/j.issn.1001-3695.2021.08.0314
- Li, J., Cheng, C., Dong, L. N., Wu, M. S., and Zhang, L. (2022). Differential diagnosis of early osteoarthritis baseS on in-line phase contrast imaging and support vector machine. *J. Capital. Med. Univ.* 43, 120–126. doi: 10.3969/j.issn.1006-7795.2022.01.020
- Li, T., Wang, Y. Z., and Li, R. C. (2008). Major nutritional components of *Boletus tomentipes* fruit bodies. *Journal of Edible Fungi* 15, 70–72.
- Li, C. Y., Wu, G., and Yang, Z. L. (2013). DNA barcoding of edible boletes (Boletaceae) from Yunnan, China. *Plant Divers.* 35, 725–732. doi: 10.7677/ynzwjy201313201
- Li, S. H., Zhao, Y. C., Yu, F. Q., Wang, X. H., Zhang, X., and Liu, P. G. (2011). Systematics of the easily confusing poisonous boletes from yunnan wild mushroom

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1036527/full#supplementary-material>

markets. *Ethnomycol. For. Manag.* 30, 34–36. doi: 10.13629/j.cnki.53-1054.2011.05.013

Ma, D. X., Liu, G., Ou, Q. H., Yu, H. C., Li, H. M., and Liu, Y. (2016). Discrimination of seven species of boletus with fourier transform infrared spectroscopy. *Guang pu xue yu guang pu fen xi = Guang pu* 36, 2479–2486. doi: 10.3964/j.issn.1000-0593(2016)08-2479-08

Malheiro, R., Guedes De Pinho, P., Soares, S., Da Silva, C., Ferreira, A., and Baptista, P. (2013). Volatile biomarkers for wild mushrooms species discrimination. *Food Res. Int.* 54, 186–194. doi: 10.1016/j.foodres.2013.06.010

Marekov, I., Momchilova, S., Grung, B., and Nikolova-Damyanova, B. (2012). Fatty acid composition of wild mushroom species of order Agaricales—examination by gas chromatography–mass spectrometry and chemometrics. *J. Chromatogr. B* 910, 54–60. doi: 10.1016/j.jchromb.2012.10.025

Mohacek-Grosov, V., Bozac, R., and Puppels, G. J. (2001). Vibrational spectroscopic characterization of wild growing mushrooms and toadstools. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 57, 2815–2829. doi: 10.1016/s1386-1425(01)00584-4

Nie, M., Luo, J., Xiao, M., Chen, J., Bao, K., Zhang, W., et al. (2007). Structural differences between fusarium strains investigated by FT-IR spectroscopy. *Biochemistry* 72, 61–67. doi: 10.1134/s0006297907010075

Qi, L. M., Liu, H. G., Li, J. Q., Li, T., and Wang, Y. Z. (2018). Feature fusion of ICP-AES, UV-vis and FT-MIR for origin traceability of boletus edulis mushrooms in combination with chemometrics. *Sensors* 18, 241–254. doi: 10.3390/s18010241

Walkowiak, A., Ledziński, A., Zapadka, M., and Kupciewicz, B. (2019). Detection of adulterants in dietary supplements with Ginkgo biloba extract by attenuated total reflectance Fourier transform infrared spectroscopy and multivariate methods PLS-DA and PCA. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 208, 222–228. doi: 10.1016/j.saa.2018.10.008

Wang, M. (2020). Development direction and industrial structure optimization of edible mushroom export in China. *Edible Fungi China* 39, 90–92. doi: 10.13629/j.cnki.53-1054.2020.01.023

Yan, Z. Y., Liu, H. G., Li, T., Li, J. Q., and Wang, Y. Z. (2022a). Two dimensional correlation spectroscopy combined with ResNet: efficient method to identify bolete species compared to traditional machine learning. *LWT* 162:113490. doi: 10.1016/j.lwt.2022.113490

Yan, Z. Y., Liu, H. G., Li, J. Q., and Wang, Y. Z. (2021). Application of identification and evaluation techniques for edible mushrooms: a review. *Crit. Rev. Anal. Chem.* 1–21. doi: 10.1080/10408347.2021.1969886 [Epub ahead of print]

Yan, Z. Y., Liu, H. G., Zhang, S., Li, J. Q., and Wang, Y. Z. (2022b). Superiority of two-dimensional correlation spectroscopy combined with ResNet in species identification of bolete. *Infrared Phys. Technol.* 125:104303. doi: 10.1016/j.infrared.2022.104303

Yang, T. W., Li, T., Zhang, J., Li, J. Q., Liu, H. G., and Wang, Y. Z. (2014). Rapid identification of bolete mushrooms by UV spectroscopy combined with euclidean distance and principal component analysis. *Food Sci.* 35, 105–109.

Yang, R. J., Liu, R., and Xu, K. X. (2013). Detection of adulterated milk using two-dimensional correlation spectroscopy combined with multi-way partial least squares. *Food Biosci.* 2, 61–67. doi: 10.1016/j.fbio.2013.04.005

Yang, R. J., Liu, C. Y., Yang, Y. R., Wu, H. Y., Jin, H., Shan, H. Y., et al. (2020). Two-trace two-dimensional (2T2D) correlation spectroscopy application in

- food safety: a review. *J. Mol. Struct.* 1214:128219. doi: 10.1016/j.molstruc.2020.128219
- Yang, R. J., Yang, Y. R., Dong, G. M., Zhang, W. Y., and Yu, Y. P. (2014). Multivariate methods for the identification of adulterated milk based on two-dimensional infrared correlation spectroscopy. *Anal. Methods* 6, 3436–3441. doi: 10.1039/c4ay00442f
- Yang, T. W., Zhang, J., Li, T., Wang, Y. Z., and Liu, H. G. (2016). Discrimination of *Boletus tomentipes* from different regions based on infrared spectrum combined with principal component analysis and cluster analysis. *Guang pu xue yu guang pu fen xi = Guang pu* 36, 1726–1730. doi: 10.3964/j.issn.1000-0593(2016)06-1726-05
- Yang, T. W., Zhang, J., Li, J. Q., Wang, Y. Z., and Liu, H. G. (2017). Study on the discrimination of species of bolete mushrooms and content prediction of cadmium by infrared spectroscopy. *Spectrosc. Spectr. Anal.* 37, 2730–2736. doi: 10.3946/j.issn.1000-0593(2017)09-2730-07
- Yao, Y. N. (2021). Research and analysis on large screen UI design for data visualization. *Process Automation Instrumentation* 42, 107–110. doi: 10.16086/j.cnki.issn1000-0380.2019120061
- Yao, S., Li, T., Liu, H. G., Li, J. Q., and Wang, Y. Z. (2018). Identification of geographical origin of *Boletus tomentipes* by multi-spectral data fusion. *Food Sci.* 39, 212–217. doi: 10.7506/spkx1002-6630-201808033
- Yue, J. Q., Li, W. Y., and Wang, Y. Z. (2021). Superiority verification of deep learning in the identification of medicinal plants: taking *Paris polyphylla* var. *Yunnanensis* as an example. *Front. Plant Sci.* 12:752863. doi: 10.3389/fpls.2021.752863
- Zhang, J., Baralkiewicz, D., Wang, Y. Z., Falandysz, J., and Cai, C. T. (2020). Arsenic and arsenic speciation in mushrooms from China: a review. *Chemosphere* 246:125685. doi: 10.1016/j.chemosphere.2019.125685
- Zhang, Y., Li, J. Q., Li, T., Liu, H. G., and Wang, Y. Z. (2018). Research on identification of *Boletus tomentipes* with different storage period by UV and FT-IR combined with data fusion. *Mod. Food Sci. Technol.* 34, 218–224. doi: 10.13982/j.mfst.1673-9078.2018.2.034
- Zhao, Y., Zhang, J., Jin, H., Zhang, J., Shen, T., and Wang, Y. (2015). Discrimination of *Gentiana rigescens* from different origins by Fourier Transform infrared spectroscopy combined with chemometric methods. *J. AOAC. Int.* 98, 22–26. doi: 10.5740/jaoacint.13-395