ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

Identification of Cultivated and Wild Mushrooms Using Laser-Induced Breakdown Spectroscopy Technology Combined with Machine Learning Algorithms

Jiayuan Tang, Shaohua Tang

School of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua, 321000, China

Abstract: *This study aimed to identify the cultivated and wild growth modes of mushrooms using Laser-Induced Breakdown Spectroscopy (LIBS) combined with machine learning algorithms. A total of 230 mushroom slice samples were collected, and their LIBS spectral data were analyzed. Multiple machine learning algorithms, including Multilayer Perceptron (MLP), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), were employed for classification. The results showed that the integration of Principal Component Analysis (PCA) significantly reduced the modeling analysis time for MLP, SVM, and KNN. Among the six classification methods, PCA+KNN achieved the highest average accuracy of 94.7%. This indicates the feasibility of using LIBS combined with machine learning algorithms for distinguishing different types of mushroom samples. The innovation of this study lies in the utilization of PCA for dimensionality reduction and the combination of various machine learning algorithms for mushroom sample classification, leading to improved accuracy and modeling speed.*

Keywords: *Laser-induced breakdown spectroscopy, Mushroom, Machine learning algorithm, Principal component analysis*

1. Introduction

Cultivated mushrooms, such as the renowned shiitake mushroom, hold a prestigious status among consumers in China due to their tender texture and widespread availability. However, there exist differences in both economic and nutritional values between wild and cultivated mushrooms, with wild mushrooms possessing higher value in these aspects. In the pursuit of greater economic benefits, some traders in the market falsely represent cultivated mushrooms as wild mushrooms. Therefore, accurate identification of mushroom cultivation methods holds significant practical implications.

In recent years, researchers in China and abroad have conducted classification and identification studies on mushrooms using various techniques. For instance, Shi et al. from Anhui University [1] successfully assessed the quality of shiitake mushrooms based on texture features and the maximum between-class variance method. Xia et al. from Hubei Academy of Agricultural Sciences [2] employed near-infrared spectroscopy to differentiate shiitake mushrooms from different origins, providing a novel method for the identification of dried shiitake mushrooms and shiitake mushroom products. Wu Xueqian from the Biotechnology Research Institute of Jiangsu Academy of Forestry [3] conducted identification of different types of cultivated shiitake mushrooms using Sequence Characterized Amplified Region (SCAR) molecular analysis. Wang Dali from the Institute of Applied Mycology at Huazhong Agricultural University [4] utilized functional gene SNP markers for the classification and identification of shiitake mushrooms. Furthermore, Sugawara et al. [5] conducted research on wild mushrooms using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). However, most of these techniques require sample pre-processing, and some rely solely on visual features for identification, thereby limiting their accuracy and the feasibility of achieving online real-time detection.

Laser-Induced Breakdown Spectroscopy (LIBS) is a method that utilizes laser-generated high-energy plasma to analyze emitted spectra and obtain information regarding the composition of materials. Currently, LIBS technology has found extensive applications in fields such as metal detection [6-7], environmental monitoring [8-9], health management [10-11], and deep space exploration [12].

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

LIBS technology has made notable advancements in material classification and identification. For example, Lu et al. from Lanzhou University of Finance and Economics [13] successfully achieved rapid and accurate classification of plastics by combining LIBS technology with GA-BP neural networks and GA-SVM methods. Sun et al. from the School of Information and Electronic Engineering at Zhejiang University of Science and Technology [14] attained 99% accurate classification of soil lead content using LIBS technology and convolutional neural networks. Li et al. [15] employed LIBS technology in conjunction with the XGBoost method for aluminum alloy identification, successfully distinguishing aluminum alloys produced by different manufacturers. Moreover, Bilge et al. [16] utilized LIBS technology to study beef, pork, and chicken, and by incorporating the PCA algorithm, achieved a correct classification rate of 83.37% for these three meat types. However, there is currently no literature documenting the use of LIBS technology for the classification and identification of mushrooms.

This study aims to utilize Laser-Induced Breakdown Spectroscopy (LIBS) technology for the classification of mushrooms grown under different cultivation methods. Two types of mushrooms with different growth methods but from the same domestic origin were selected as the research subjects. The feasibility of combining LIBS technology with machine learning algorithms for differentiating mushroom cultivation methods will be explored. This research endeavors to provide new methods and technologies for the rapid differentiation and detection of mushrooms in the future.

2. Experimental Section

2.1 Sample Preparation

The dried shiitake mushrooms were manually sliced into thin slices measuring 20 mm × 20 mm × 2 mm. A total of 50 slices were obtained for wild shiitake mushrooms (Guizhou) and 50 slices for cultivated shiitake mushrooms (Guizhou).

2.2 Optical Setup and Experimental Instruments

The experimental setup is shown in Figure 1. The laser used was a Plite 200-II laser with a wavelength of 1064nm, pulse frequency of 1Hz, and pulse width of 8ns. The spectrometer employed was an AvaSpec-2048-USB2 fiber-optic spectrometer with a data acquisition range of 196-510nm, integration time of 2ms, and spectral resolution of 0.09-0.1nm. By optimizing the experimental parameters, the laser energy was set at 50mJ, and the delay between the laser pulse and spectrometer acquisition was set at 700ns. The mushroom slices were placed on a two-dimensional movable stage, and the sample laser collection position was adjusted in real time by controlling a stepper motor.

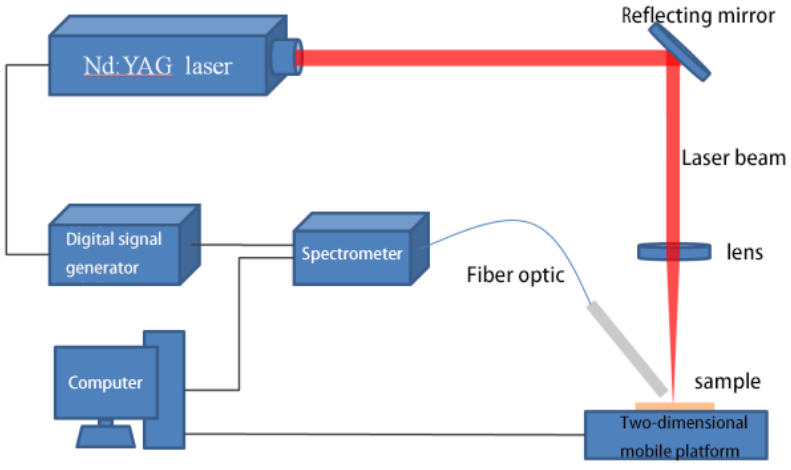


Figure 1: The diagram of LIBS experimental device.

2.3 Experimental Procedure

Under the same experimental conditions, the spectra of 50 slices of cultivated mushrooms and 50 slices of wild mushrooms were collected sequentially. For each mushroom sample, spectra were obtained from 50 points on both the front and back sides, resulting in a total of 5000 spectra for cultivated mushrooms and 5000 spectra for wild mushrooms.

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

3. Results and Discussion

3.1 Data Preprocessing

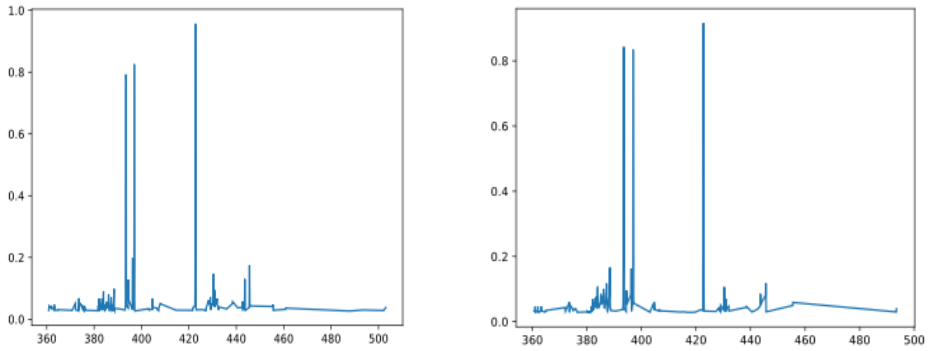
The normalization technique of Maximum-Minimum (MinMax) normalization was applied to preprocess the spectral data, which is a commonly used data processing technique. This method linearly maps the original spectral data to the range of [0, 1]. The normalization formula is as follows:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Where x represents the original spectral data, x_{max} is the maximum intensity value in the original spectral data, x_{min} is the minimum intensity value in the original spectral data, and x* is the normalized spectral data after the Maximum-Minimum normalization.

During the spectral acquisition process, we observed clear signals at 421.952 nm for both types of mushrooms. Therefore, based on the signal-to-noise ratio at this wavelength, we selected the top 200 spectra with the highest signal-to-noise ratio from each type of mushroom as the dataset for machine learning.

After the selection process, we averaged the 200 spectra obtained for each mushroom type to generate the spectral profiles shown in Figure 2. From the observation of the figure, it can be noted that the LIBS spectra of the two mushroom types exhibit similar features.



(a) Domestic Shiitake Mushroom Spectrum *(b) Wild Shiitake Spectrum*

Figure 2: Mushroom spectrum: (a) Domestic Shiitake Mushroom Spectrum and (b) Wild Shiitake Spectrum.

3.2 Selection of Training and Testing Sets

The division of the training and testing sets is as follows, with the specific ratios outlined in Table 1:

Training Set: Randomly select 70% of the spectral data for each mushroom type as the training set.

Testing Set: The remaining 30% of the spectral data is used as the testing set.

In machine learning, it is common practice to split the dataset into training and testing sets to evaluate the performance of a model. The testing set is used to assess the accuracy of the model's predictions. However, the random selection process of training and testing sets can yield different partition results, which may affect the accuracy estimation.

To address this issue, a commonly used approach is to perform multiple independent repetitions of the experiment, each time using different training and testing set partitions. By conducting multiple experiments and calculating the average accuracy, the influence of the random partitioning process can be reduced, providing a more stable and reliable evaluation metric. The purpose of this approach is to minimize the error caused by the randomness of the dataset partition and obtain more reliable performance evaluation results. Therefore, by conducting multiple repetitions of the experiment and calculating the average accuracy, we can better assess the accuracy and stability of the algorithm.

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

Table 1: The division of training set and test set.

Sample name	Number of training samples (instances)	Number of training samples (instances)
Cultivated shiitake mushrooms	71	29
Wild shiitake mushrooms	69	31
Total	140	60

3.3 Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is a feedforward artificial neural network model that is widely used in various fields for mapping multiple input datasets to a single output dataset. It has strong expressive and modeling capabilities. The basic principle of MLP is illustrated in Figure 3:

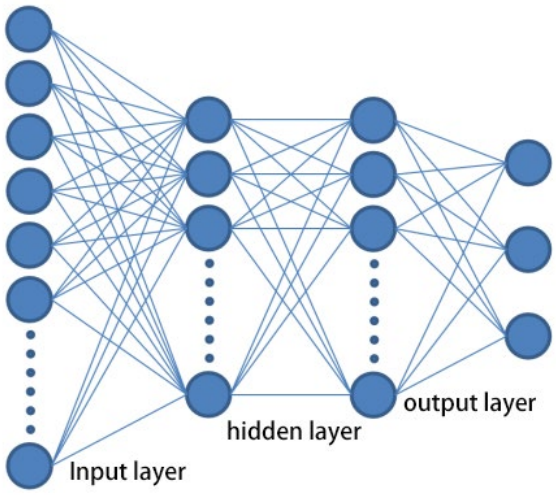


Figure 3: The Schematic diagram of MLP.

The study constructed a fully connected neural network model, consisting of an input layer, hidden layers, and an output layer. A learning rate of 0.0005 was chosen, and ReLU and Softmax activation functions were used to perform the multi-classification task.ReLU (Rectified Linear Unit) is a commonly used activation function that sets negative input values to 0 while leaving positive input values unchanged. Its simple computation and non-linear characteristics help improve the expressive power of the neural network.The Softmax activation function is commonly used for multi-classification tasks. It transforms the output of the neural network into a probability distribution, with output values ranging from 0 to 1 and summing up to 1. The Softmax function interprets the neural network's output as the probabilities of the different classes, facilitating classification decisions.After 150 training iterations, the neural network model achieved an accuracy of 87.4%. This indicates that your model has performed well and demonstrated the ability to classify mushroom samples accurately. The training process of the neural network is an iterative process that continuously adjusts the weights and biases through backpropagation, gradually optimizing the model and improving its accuracy. In this experiment, satisfactory results were obtained after 150 training iterations.

3.4 KNN

K-Nearest Neighbor (KNN) learning is a commonly used supervised learning method, and its working mechanism is quite simple. Given a test sample, it finds the k nearest training samples based on a certain distance measure. Then, it uses the information from these k neighbors to make predictions.The selection of the value of k is crucial in the KNN algorithm, as the optimal k value leads to the best classification prediction results. In this study, the 10-fold cross-validation method was used to train the KNN model on the training set. By comparing the prediction accuracy of the KNN model for each k value, the optimal k value was determined. Based on the optimal k value, classification predictions were made, and the results of the 10-fold validation are shown in Figure 4. It can be observed that the optimal k value is 5, with a training set accuracy of 92.5%.

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

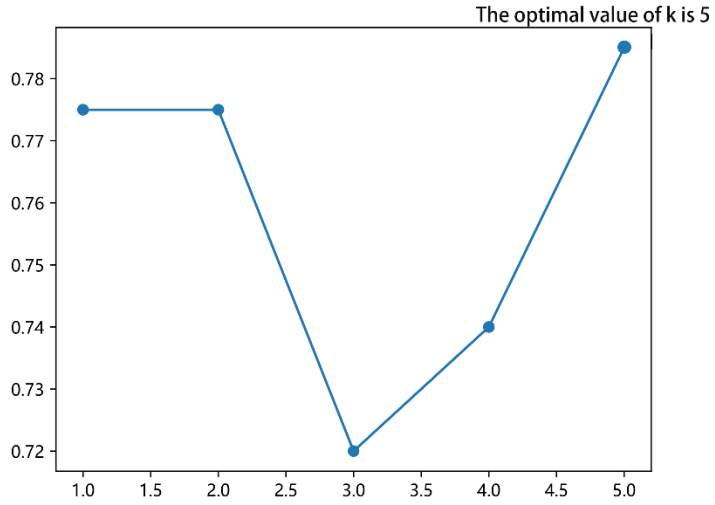


Figure 4: 10 schematic diagram of double cross validation results.

The KNN model with the optimal k value of 5 was tested on the test set, and the results were presented using a confusion matrix, as shown in Figure 5. In the confusion matrix, 0 represents cultivated mushrooms, and 1 represents wild mushrooms. The vertical axis represents the predicted labels, and the horizontal axis represents the true labels.Based on the test results, the KNN model achieved an overall accuracy of 91.7% on the validation set.

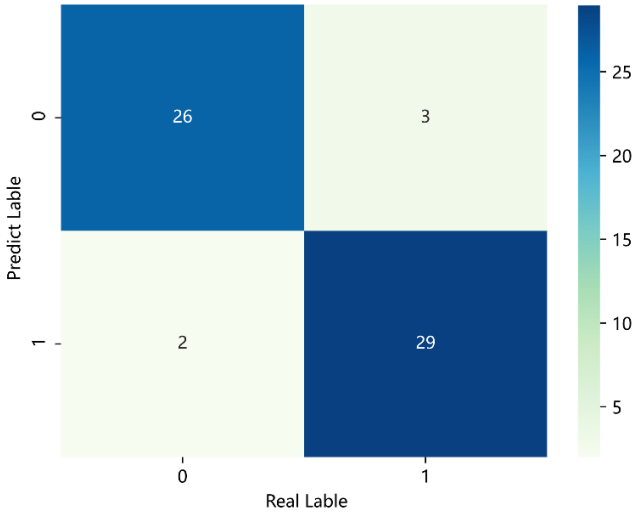


Figure 5: Confusion matrix results.

3.5 SVM

Support Vector Machine (SVM) is a machine learning algorithm proposed by Vapnik and others in the 1990s. SVM is a binary classification model that determines the decision boundary by solving for the maximum-margin hyperplane.

During the optimization process, SVM uses the hinge loss function to calculate the empirical risk. The hinge loss function applies a margin to correctly classified samples, and the loss is zero only when the samples are within the margin. This makes SVM more sensitive to misclassified samples and enables the generation of classifiers with better robustness.

Table 2 presents the average prediction accuracy of SVM models using different kernel functions on the mushroom test samples:

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

Table 2: The accuracy rate of different kernel functions of the SVM model.

(kernel function)	(Average prediction accuracy)
Linear kernel	88.53%
Radial basis function (RBF) kernel	84.51%
Polynomial kernel	78.40%
Non-linear kernel (Sigmoid)	68.42%

So, the SVM model used in this experiment ultimately employed a linear kernel, and the result of 15,000 independent experiments was 88.53%.

3.6 PCA

Principal Component Analysis (PCA), also known as principal component regression analysis or principal component regression, is an unsupervised data dimensionality reduction method. In spectral data analysis, the amount of data in spectra can be very large. By applying PCA, the data volume can be effectively reduced while preserving the key information of the spectral data, thus improving the speed of model building and analysis.

First, the ranking of principal components is computed. Here, the importance ratio of the top 8 principal components in the data is calculated. The x-axis represents the eight different principal components, and the y-axis represents the proportion of importance of each principal component, displayed as a bar chart. The red step chart represents the cumulative importance ratio obtained by adding up the top n principal components. The results are shown in Figure 6.

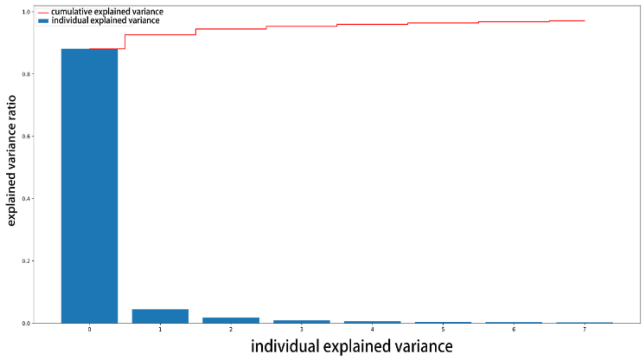


Figure 6: Schematic diagram of the importance ranking of principal components.

From the above figure, it can be observed that the cumulative importance of the top two principal components exceeds 93%. Therefore, a two-dimensional scatter plot of the test set is plotted using the top two principal components, resulting in the graph shown in Figure 7. Here, 0 represents cultivated mushrooms, and 1 represents wild mushrooms.

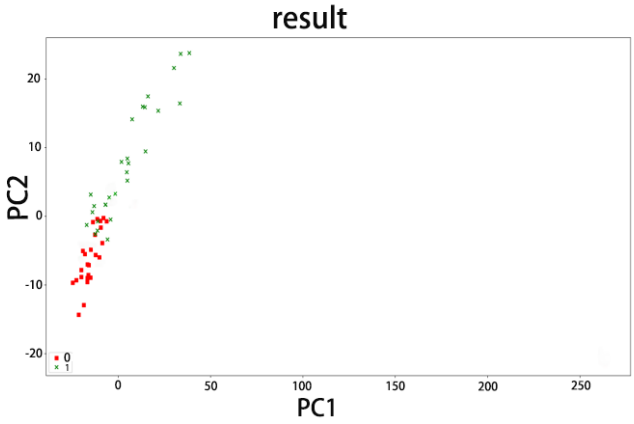


Figure 7: Two dimensional scatter plot of PCA of two kinds of Lentinus edodes.

Based on this, I further applied the KNN and SVM algorithms for classification prediction, repeating

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

the experiment independently for 15,000 times. The average prediction accuracy of PCA+KNN was 94.75%, while PCA+SVM achieved an average prediction accuracy of 89.25%. Additionally, when combined with MLP, PCA+MLP achieved an average prediction accuracy of 88.4%. The results of the six methods, MLP, PCA+MLP, SVM, PCA+SVM, KNN, and PCA+KNN, are summarized in Table 3.

Table 3: Summary of the results of the six methods.

Algorithm	Average prediction accuracy	Modeling time required
MLP	87.4%	27s (150 times)
PCA+MLP	88.4%	10s (150 times)
SVM	88.5%	30s (15000 times)
PCA+SVM	89.4%	8s (15000 times)
KNN	91.7%	25s (15000 times)
PCA+KNN	94.7%	12s (15000 times)

By using PCA dimensionality reduction technique in combination with different classifiers, we observed some performance improvements in our experiments. Specifically, when combining PCA with MLP, we achieved a 1% increase in accuracy compared to using MLP alone. The improvements seen with SVM combined with PCA and KNN combined with PCA can be attributed to several factors. Firstly, PCA reduces the complexity of the data, thereby decreasing the risk of overfitting and allowing the classifiers to generalize better to unseen samples. The reduced-dimensional data obtained through PCA also better reflects the similarities between samples, which aids in improving classification accuracy. In the case of KNN, the dimensionality reduction provided by PCA can reduce the computational cost of distance calculations, thereby enhancing the performance of the KNN algorithm. Hence, the improvements observed with SVM combined with PCA and KNN combined with PCA contribute to an overall increase in the average accuracy of the classification algorithms. PCA's role in dimensionality reduction reduces the complexity of the data, prevents overfitting, and speeds up the modeling and analysis process. These combined factors lead to an improvement in the performance of the classification algorithms.

4. Conclusion

In this study, laser-induced breakdown spectroscopy (LIBS) technique combined with six machine learning algorithms was used to differentiate the cultivation methods of mushrooms. The experimental results showed that when using MLP for classification, the average accuracy was 87.4%, while KNN achieved an average accuracy of 90%, and SVM achieved 88.5%. Among them, KNN algorithm performed better in terms of classification accuracy. To further improve the classification performance, principal component analysis (PCA) was applied to preprocess the spectral data, and the first two principal components were extracted as input variables for MLP, SVM, and KNN. By combining PCA with MLP, SVM, and KNN, new classification methods were obtained: PCA+MLP, PCA+SVM, and PCA+KNN. The average accuracies of these methods were 88.4%, 89.4%, and 94.7%, respectively. Compared to using MLP, SVM, and KNN directly, all the methods improved the classification accuracy and accelerated the modeling and analysis process. Therefore, the results of this study demonstrate that PCA provides a new technique for distinguishing mushroom cultivation methods. By using PCA for dimensionality reduction, the performance of the classification algorithms can be enhanced, resulting in higher classification accuracy and faster modeling and analysis speed. The findings of this research provide feasible methods for accurate identification of mushroom cultivation methods and have practical application value.

References

[1] Shi Yu, Sun Dong, Gao Qingwei, Lu Yixiang, Ji Mingwei. Mushroom Classification Method Based on Texture Features[J]. Foreign Electronic Measurement Technology, 2022, 41(01): 158-162. DOI: 10.19652/j.cnki.femt.2103035.

ISSN 2616-5775 Vol. 6, Issue 13: 161-168, DOI: 10.25236/AJCIS.2023.061323

[2] Xia Zhenzhen, Zheng Dan, Xia Hong, Yao Jingjing, Wang Shengpeng, Qiu Jianfei. Rapid Identification of Cultivated Mushrooms' Origin by Near-Infrared Spectroscopy Analysis[J]. Chinese Journal of Analysis and Testing, 2020, 39(11): 1371-1377.
[3] Wu Xueqian, Li Haibo, Wei Hailong, Fu Lizhong, Wu Qingqi, Peng Huazheng, Zhu Muyuan. Application of SCAR Molecular Marker Technology in Mushroom Strain Identification[J]. Mycosystema, 2005(02): 259-266. DOI: 10.13346/j.mycosystema.2005.02.016.
[4] Wang Dali, Bian Yinbing, Huang Chenyang, Zhang Jinxia. Application of Functional Gene SNP Markers in Discrimination of Mushroom Varieties[C]//. Abstracts of the 5th Membership Representative Conference of the Chinese Mycological Society and the 2011 Academic Annual Meeting. [Publisher not specified], 2011: 100-101.
[5] A, Ryota Sugawara, et al. "Rapid and reliable species identification of wild mushrooms by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)." Analytica Chimica Acta 934(2016):163-169.
[6] Liu Shiming. Research on Elemental Analysis Method in Metal Oxide Nanofilm Based on LIBS Technology [D]. Shandong University of Technology, 2021. DOI: 10.27276/d.cnki.gsdgc.2021.000303.
[7] Chen Boyang. Research on Analysis Method of Partial Metal Elements in Iron Ore Based on LIBS Technology [D]. Xi'an Shiyou University, 2020. DOI: 10.27400/d.cnki.gxasc.2020.000012.
[8] Zhang Ranran. LIBS Detection of Heavy Metal Elements in Aquatic Products [D]. Zhejiang Normal University, 2021. DOI: 10.27464/d.cnki.gzsfu.2021.000203.
[9] Sun Miao. Research on Industrial Heavy Metal Pollution Detection Model Based on Machine Learning and LIBS [D]. Zhejiang University of Science and Technology, 2020. DOI: 10.27840/d.cnki. gzjkj.2020.000097.
[10] Yang Sibo. Research on Elemental Imaging and Cluster Analysis of Breast Cancer Tissues Based on LIBS Technology [D]. Harbin Institute of Technology, 2019. DOI: 10.27061/d.cnki. ghgdu.2019. 002308.
[11] Liu Xianyun, Wang Zhenya, Hao Liqing, Zhao Wenwu, Huang Mingqiang, Long Bo, Zhang Weijun. Application of Laser-Induced Breakdown Spectroscopy in Biomedicine [J]. Laser Technology, 2008(02): 134-136.
[12] M. Darby Dyar et al. Strategies for Mars Remote Laser-Induced Breakdown Spectroscopy Analysis of Sulfur in Geological Samples [J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2010, 66(1): 39-56.
[13] Lu Yonghua. Research on Plastic Classification Based on LIBS Combined with GA-BP Neural Network and GA-SVM [J]. Laser & Infrared, 2022, 52(02): 273-279.
[14] Sun Miao, Guindo Mahamed Lamine, Zhuang Zhenhua, Zhong Haipeng, Zhao Yun. Rapid Classification of Soil Lead Content Levels Based on LIBS Technology and Convolutional Neural Network [J]. Journal of Zhejiang University of Science and Technology, 2019, 31(05): 373-380.
[15] Li Chenyang, Chen Xiongfei, Zhang Yong, Wang Yawen, Tian Zhongchao, Wang Shigong, Zhao Zhenyang, Liu Ying, Liu Pengyu. Aluminum Alloy LIBS Spectrum Classification and Recognition Method Based on XGBoost [J]. Spectroscopy and Spectral Analysis, 2021, 41(02): 624-628.
[16] Bilge G, et al. "Identification of meat species by using laser-induced breakdown spectroscopy." Meat Science 119.sep. (2016):118-122.