



# Research on peanut variety classification based on hyperspectral image

Zhiyong ZOU<sup>1</sup>, Li WANG<sup>1</sup> , Jie CHEN<sup>1</sup>, Tao LONG<sup>1</sup>, Qingsong WU<sup>1</sup>, Man ZHOU<sup>2\*</sup>

## Abstract

The classification algorithms of different peanut varieties were studied based on hyperspectral imaging technology. Firstly, the spectral images of five peanut species were collected by hyperspectral instrument produced by Zhuolihanguang Co., LTD. Then SpacVIEW was used to correct the spectral images in black and white, and ENVI5.1 was used to extract the interest in the spectral image of each peanut and calculate the mean spectral reflection value of the region. The spectral characteristic curves of the five peanut samples all showed certain differences, which lays a foundation for the feasibility of modeling in the next step. In order to eliminate the influence of non-quality factor information in hyperspectral spectral data, a variety of data preprocessing methods were used to eliminate noise in the original spectral data, and XGBoost, LightGBM, CatBoost and GBDT algorithms were used to extract feature bands. XGBoost and LightGBM were then used for classification modeling of extracted feature bands. In the classification model, both XGBoost and LightGBM can reach 99.33%, while other performance evaluation indexes cannot distinguish these two models well. Therefore, Optuna algorithm was selected to optimize the two algorithms respectively. After optimization, both LightGBM and XGBoost have improved to varying degrees, but LightGBM is relatively obvious, especially in fit\_time, which is 11 times faster than XGBoost and 16 times faster than before optimization. Therefore, the best classification algorithm selected in this study is MF-LightGBM-LightGBM-Optuna-LightGBM. The research of peanut classification method provides a strong theoretical basis and technical support for the revitalization of rural industry, the integration of peanut agriculture and industry and the acceleration of the modernization of agricultural industry system, production system and management system.

**Keywords:** peanut classification; hyperspectral classification method; modeling; LightGBM algorithm; optuna.

**Practical Application:** Non-destructive classification of peanut.

## 1 Introduction

China is one of the top three peanut producing countries in the world. There are more than 200 kinds of peanuts available now, and more than 30 kinds of fine varieties (Liu, 2011). Different varieties of peanuts are difficult to distinguish with the naked eye, and even the same varieties of peanuts also have different grain types. Even though peanut has a wide range of adaptations, its productivity is greatly influenced by soil physical conditions and environmental factors, especially temperature, water availability and radiation (Barbosa et al., 2014). Now peanuts are widely used in food industry processing in addition to oil and edible two main uses (Qin et al., 2015). Because peanut oil yield rate is high, generally in 40%-50%, so promote the development of peanut industry, can improve the supply capacity of edible oil (Tang et al., 2010), Peanuts are also rich in protein, and there are high food shortages of protein products today, mainly due to low consumption of animal-based foods at prices that are generally not available to the majority of the population (Camargo et al., 2011). China has always been a big producer and exporter of peanuts. In China, the annual planting area of peanuts is about

5000 Kkm<sup>2</sup>, its total output is about 15000 Kt, which is about 50% of the total output of other oil crops in China. The output of peanut oil is second only to that of rapeseed oil, and it is the second largest source of vegetable oil (National Bureau of Statistics of the People's Republic of China, 2019). Although the demand for edible oil has been rising in recent years, the planting area of oil crops has been declining due to various factors. Compared with other oil crops in China, peanut has the advantages of large production scale, high planting efficiency, high oil production efficiency, good oil quality, strong international competitiveness, huge demand space and so on. It has significant advantages and development potential in ensuring the supply of edible vegetable oil in China. This means that rapid and nondestructive identification of peanut varieties is of great significance to food processing (Liu et al., 2020). Hyperspectral technology provides a convenient, nondestructive and pollution-free detection method, which provides theoretical support for agricultural detection and embodies great technical advantages, and plays a role in promoting agricultural modernization (Burns & Ciurczak, 2007).

Received 07 Feb., 2022

Accepted: 03 Apr., 2022

<sup>1</sup> College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Ya 'an, China

<sup>2</sup> College of Food Science, Sichuan Agricultural University, Ya 'an, China

\*Corresponding author: 577937644@qq.com

Although the research of peanut quality nondestructive testing by hyperspectral image technology is not thorough, it develops fast and has made some achievements. Deshuai Yuan et al. studied the identification of moldy peanuts by a small number of critical band ensemble classifiers (EC) based on hyperspectral images, Support Vector Machine (SVM) based EC, partial least squares discriminant analysis (PLS-DA) and cluster independent soft pattern classifier (SIMCA) were used to select key wavelengths to identify healthy and moldy peanuts. The overall pixel classification accuracy of EC, SVM, PLS-DA and SIMCA were 97.66%, 97.53%, 95.31% and 97.36%, respectively (Yuan et al., 2020). Liu Cuiling et al. quickly classified peanuts of different varieties by using hyperspectral imaging technology and using network model as Google-Net model with 22-layer deep network (Liu et al., 2020), compared with the traditional spectral processing method of PLS-DA, it was found that Google-Net classification results were significantly better than those of PLS-DA in six feature bands. Zhenbo Li et al. used aspect ratio convolutional neural network, HOG convolutional neural network and Hu moment invariant convolutional neural network to extract features of peanut images respectively, and then used support vector machine (SVM) to classify peanut images. The accuracy of aspect ratio SVM algorithm, HOG + SVM algorithm and Hu moment invariant + SVM algorithm are 96.72%, 81.97% and 81.97%, respectively (Li et al., 2018). Qiao Xiaojun et al. obtained the first derivative of the asynchronous length after continuum removal, and calculated the  $Area_{500-600}$  index in the spectral region with better separability. Secondly, the shape and position of the spectrum were extracted by continuous wavelet transform, and the moldy anemone samples were identified by Indexcwt index. The results showed that the J-M distance of  $Area_{500-600}$  index was 1.95, and the J-M distance of Indexcwt model was 1.99, indicating that the characteristic space separability of  $Area_{500-600}$  index model and Indexcwt model was better, and the spectral characteristics of moldy peanuts were analyzed (Qiao et al., 2018). Yu Hongwei et al. used hyperspectral imaging technology combined with stoichiometry and the 2nd-RC-PLS algorithm to transform peanut hyperspectral image into protein content distribution map to study the protein content and distribution in peanut (Yu et al., 2016).

In the past 30 years, hyperspectral remote sensing technology has developed rapidly since it was proposed, and new breakthroughs have been made in hyperspectral image processing and information extraction. The research of hyperspectral image processing and information extraction mainly includes data dimension reduction, image classification, mixed pixel decomposition and target detection. Although hyperspectral imaging technology has provided a key technology for nondestructive testing of agricultural products, the stability and universality of agricultural product index testing model based on spectral technology still need to be improved. Due to natural factors and inevitable errors in operation, it is difficult for a certain peanut spectral model to be suitable for all regions and all varieties. Therefore, the future research objectives will be to establish the model database, make it applicable to different varieties and different regions of peanut, select the appropriate stoichiometry method and vegetation index, and improve the stability and universality of peanut quality detection model (Quiroga et al., 2020).

The overall goal of this paper is to develop an excellent model for nondestructive classification of peanut varieties based on hyperspectral image technology. A variety of data pretreatment methods are used to preprocess spectral data, XGBoost, LightGBM, CatBoost and GBDT algorithms are used to extract feature bands, and then XGBoost and LightGBM are used to classify and model the extracted feature bands. Finally, Optuna algorithm is used to optimize the two models, and the optimal model parameters are obtained, and then the optimal model of lossless classification is obtained.

## 2 Experimental part

### 2.1 Material

Main peanut varieties (Dabaisha, Huayu 16, Xiaobaisha, Haihua and Luhua) in the main planting provinces (Shandong, Henan, Jiangsu, etc.) in China in recent years were taken as research objects, as shown in Figure 1 below. We selected 400 peanuts from each peanut we purchased as samples for the experiment.

### 2.2 Hyperspectral image acquisition

In the experiment of peanut hyperspectral Image acquisition, image- $\lambda$  "spectral Image" series high spectrometers of Zhuolihanguang Company and SpacVIEW software were used, as shown in Figure 2 below. In order to improve the collection efficiency, 20 peanuts were collected each time. The effective spectral band range was 387-1035 nm, and the band resolution was 2.8 nm. There were 256 bands in total, and the pixel size of imaging was 1344\*1024. The determination time of each sample was set as 10 s, the object distance between the hyperspectral camera objective lens and the peanut sample carrier platform was set as 170 mm, the moving speed of the electronic transport mobile carrier platform device was set as 4.6 mm·s<sup>-1</sup>, the exposure time of the hyperspectral camera was set as 9 ms, and the scanning broadband area was set as 120 mm (Ma et al., 2021). In the process of collection, due to the influence of noise caused by ambient factors and dark current of the instrument, it is necessary to collect black and white frames respectively before sample collection, and conduct black and white correction in SpacVIEW according to the following Formula 1 after sample collection (Zou et al., 2019). And extract the regions of interest of 20\*20 pixels in the peanut hyperspectral images corrected by black and white using ENVI5.1 (Exelis Visual Information Solutions Inc., USA) software. Then the average value of pixels in the region of interest of all the corresponding images in each band of 387-1035 nm was calculated as the characteristic reflection spectral curve of peanut of different varieties.

$$R_1 = \frac{R_0 - R_B}{R_W - R_B} \quad (1)$$

Where,  $R_0$  is the original hyperspectral image,  $R_1$  is the corrected image,  $R_B$  is the black frame image,  $R_W$  is the white frame image.

### 2.3 Spectral data preprocessing

Because each ROI region of hyperspectral image extracted has 256 bands, corresponding to 256 data, the amount of data is



Figure 1. Five peanut sample.

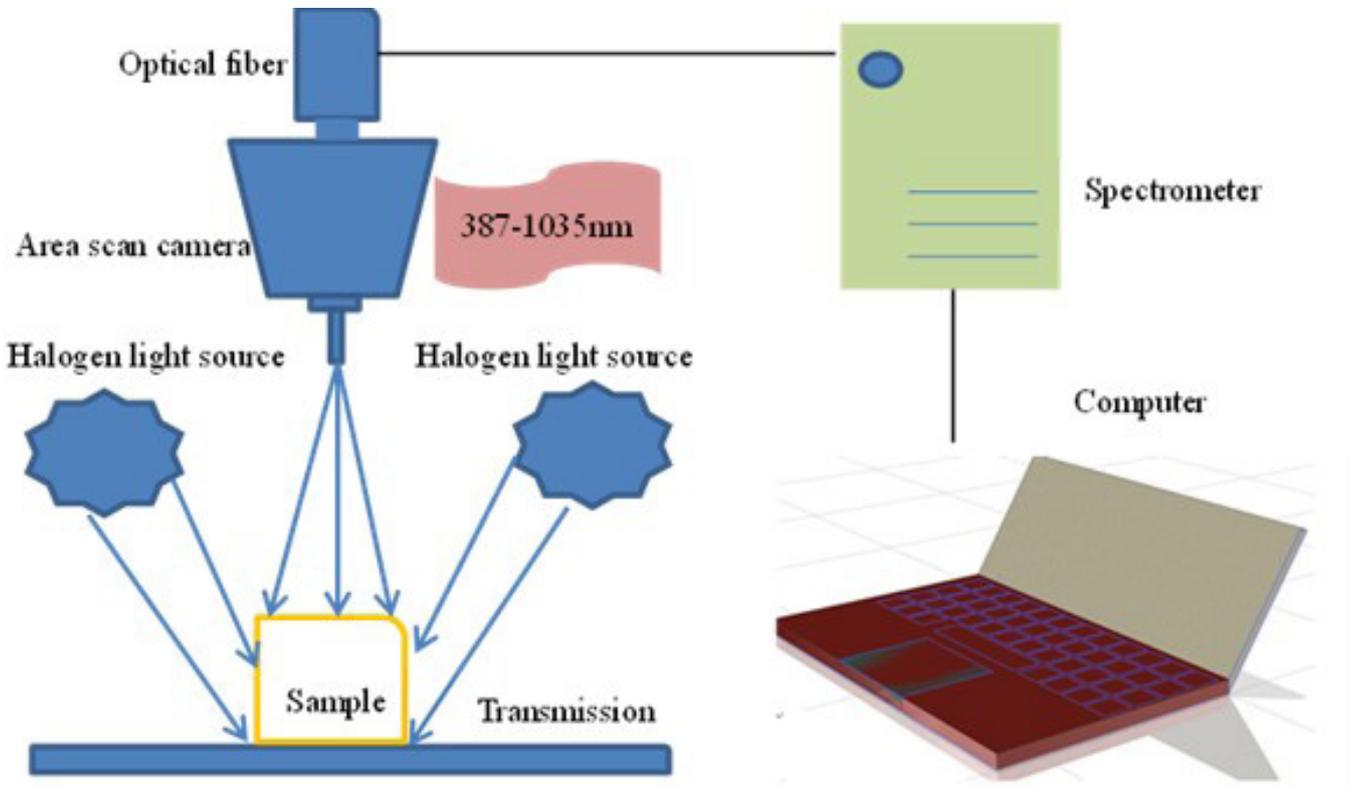


Figure 2. Image-λ “spectral Image” series high spectrometer of Zhuolihanguang Company.

large; And in hyperspectral data, weather, light, and the differences or artificial operation inevitable error, need to noise reduction of collected data, according to the strong correlations between the properties and spectral data band according to the amount of total high redundancy feature, so this research USES a variety of data preprocessing algorithm, and compared their results, get the optimal modeling spectrum data, including Derivative, Smoothing, Data Enhancement (Yang et al., 2021), Light Scattering Correction (Preys et al., 2008), Fourier Transform, Wavelet Transform (Xin et al., 2020), Local regression-weighted linear least squares + First Order Polynomial model, Local regression-weighted Linear least squares + Second Order Polynomial model (Zhang et al., 2017).

**2.4 Feature extraction of spectral data**

As the band selection of hyperspectral image is a very complex band combination optimization problem, it requires the selected band combination to have good performance, that is, to select the band combination with large information content, small correlation and good category separability. CatBoost, GBDT, XGBoost and LightGBM algorithms were used to extract important bands.

**2.5 Model selection**

The unsupervised learning algorithm was used to randomly select 70% spectral data of five peanut samples of Dabaisha, Huayu 16, Xiaobaisha, Haihua and Luhua as training set, and the rest data as test set. XGBoost and LightGBM deep learning algorithms were used to build models based on the extracted feature bands to classify the five peanut varieties (Thongsuwan et al., 2021; Żelasko, 2020; Ma et al., 2018; Aboozar & Georgina, 2020; Ma et al., 2020), the confusion matrix is obtained; Accuracy, F-Score, Log Loss and Hamming Loss are used to evaluate the effect of model training prediction.

*XGBoost*

XGBoost is a classification and regression algorithm based on gradient lifting decision tree (GBDT). XGBoost first builds a certain number of weak learners, most of which are classified regression trees, to train weak learners. It also performs weighted summation after training to obtain the final regression model. During construction, new learners are always added based on the residual error obtained from the last weak learner iteration.

*LightGBM*

LightGBM has three major optimizations on XGBoost:

1. Histogram algorithm: Histogram;
2. GOSS algorithm: gradient unilateral sampling algorithm;
3. EFB algorithm: mutually exclusive feature binding algorithm.

The relationship between LightGBM and XGBoost can be shown using the following Formula 2:

$$LightGBM = XGBoost + Histogram + GOSS + EFB \tag{2}$$

**2.6 Model optimization**

LightGBM and XGBoost were optimized using the Optuna algorithm. Optuna is a software framework that can automatically optimize hyperparameters. It can find the best and optimal value of hyperparameters through trial and error. It determines the value of the next hyperparameter to be tested based on the historical data of the run. Based on these data, it can estimate the region where hyperparameter comparison is likely to occur and make a hyperparameter search attempt in this region. As it gets new results, it updates the area and continues to search. In the process of continuous repeated search, evaluation and update to obtain better performance of the hyperparameter, its operation framework (Srinivas & Katarya, 2022) as shown in Figure 3 below.

**3 Results and discussion**

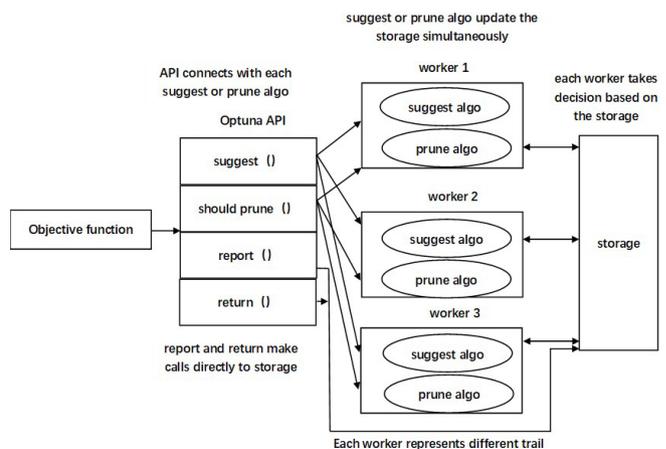
**3.1 Peanut spectral data analysis and pretreatment**

*Initial characteristic spectral analysis of peanut*

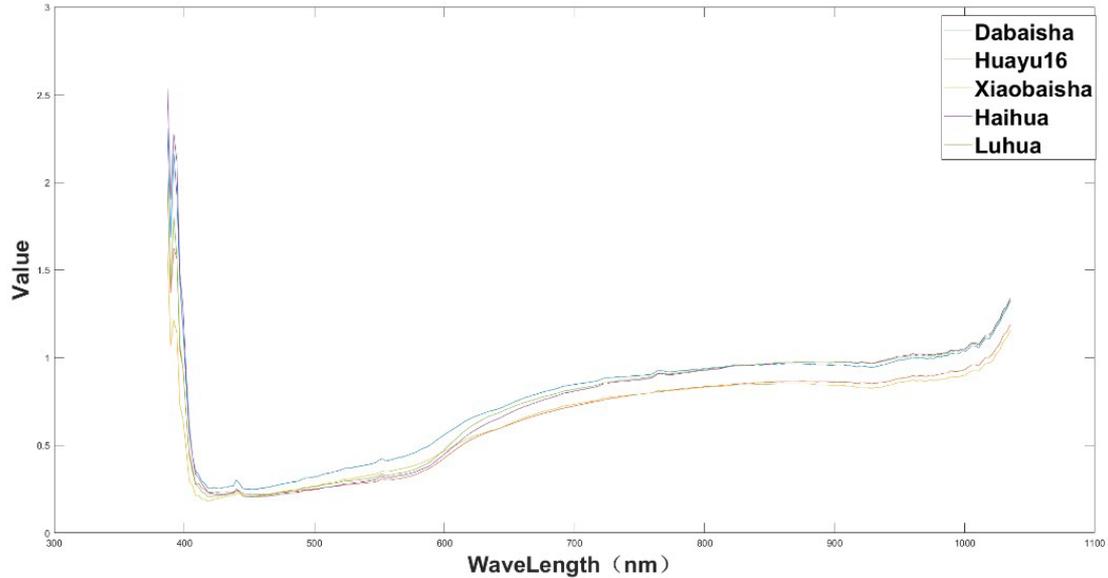
After SpacVIEW black and white correction treatment, ENVI5.1 was used to intercept each peanut region of interest and calculate its average reflection value. The spectral reflection value curve of the obtained data was drawn, as shown in Figure 4 below. It can be seen that from the initial band 387 nm, there was an absorption peak state, and the first wave peak appeared in the vicinity of 392 nm. The peak value of Haihua was the highest, and that of Xiaobaisha was the lowest, near 440 nm again when a wave, the Dabaisha crest value of the highest, Xiaobaisha reached the lowest, at 550 nm, a wave again, the Dabaisha value still is highest, Huayu16 got the lowest, from now, five kinds of peanut spectral reflection value began to rise, at 770 nm, the last wave peak appeared at this time, the wave peak of Dabaisha was still the highest, and huayu16 and Xiaobaisha showed a cross trend. According to the above analysis, it is feasible to classify these five peanut species based on hyperspectral image technology.

*Pretreatment of peanut spectral data*

After extracting the spectral data in ENVI5.1, in order to eliminate various noises during the collection, the spectral data



**Figure 3.** Optuna tuning framework.



**Figure 4.** Spectral images of average reflection values of five peanut species.

is preprocessed by using the method in 1.3, and the results are shown in Table 1:

As can be seen from Table 1, the recognition accuracy of LTN (logarithmic transformation standardization), SD (second derivative), GWS (Gaussian window smoothing), SG (Savitzky-Golay filtering), CAN (inverse cotangent correction), ZSS (Z score standardization) and WTD (wavelet threshold denoising) after pretreatment is 93.17% as that of the original data. LR1 (local regression-weighted linear least squares + first-order polynomial model), MMS (Min-max standardization), LR2 (local regression-weighted linear least squares + second-order polynomial model), MSC (multiple scattering correction) and ES (exponential smoothing) are not as accurate as the original data. FD (first derivative), L2NN (L2 norm normalization) and MF (median filtering) are more accurate than the original data, reaching 93.33%, 93.33% and 98.83% respectively. The hamming\_loss of MF algorithm is 0.01, so the misclassification of samples on a single label is better. The highest Jaccard\_similarity is 0.99, proving that the sample similarity is the highest. The log\_loss of MF is also the smallest among all the algorithms, that is, the regression loss is the smallest, and the adaptation time of MF is the shortest among the three algorithms, which is 0.96s. In summary, among the 17 pretreatment methods, MF algorithm stands out, so its preprocessed data are used for modeling and classification.

### 3.2 Extraction and modeling of peanut characteristic bands

XGBoost, LightGBM, GBDT and CatBoost were used to extract important feature bands, as shown in Figure 5:

As can be seen from Figure 5, both LightGBM and GBDT algorithms choose the band of 401.48999 nm as the best classification band, and this band is also important in other algorithms, indicating that the algorithms have some commonality.

XGBoost and LightGBM algorithms are used to model the feature bands and the classification results are shown as follows:

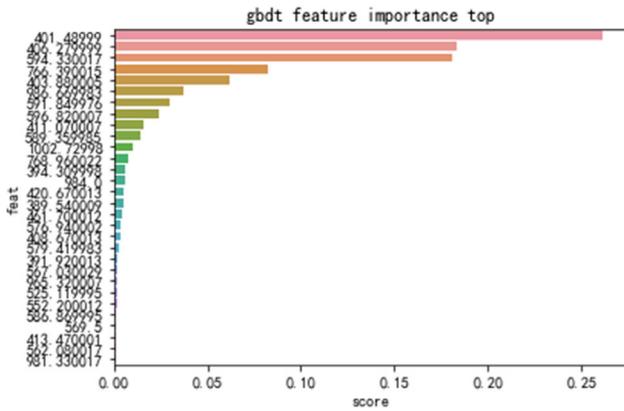
As can be seen from Table 2. The running results of the two algorithms are relatively good, and the modeling accuracy of the feature bands extracted by XGBoost and LightGBM reaches 0.993, showing a degree of differentiation from the other two algorithms. However, compared with the two algorithms, it is difficult to distinguish which one is superior and which one is inferior in other performance indicators. Therefore, The Optuna algorithm was used to optimize the parameters (max\_depth, N\_estimator, num\_leaves, subsample) of XGBoost and LightGBM, and the optimization times were set to 300 times in the Optuna algorithm. The optimization process is shown in the figure below:

It can be seen from the contour map that in the 300 times of optimization, the algorithm used multi-fusion mode to select the three parameters that are most suitable for one parameter. The four parameters are optimized in turn for 12 times to obtain the optimal parameters. The parameters obtained by the optimization algorithm when Objective Value = 1 are substituted into the XGBoost and LightGBM models, and the results are shown as follows (Figures 6 and 7).

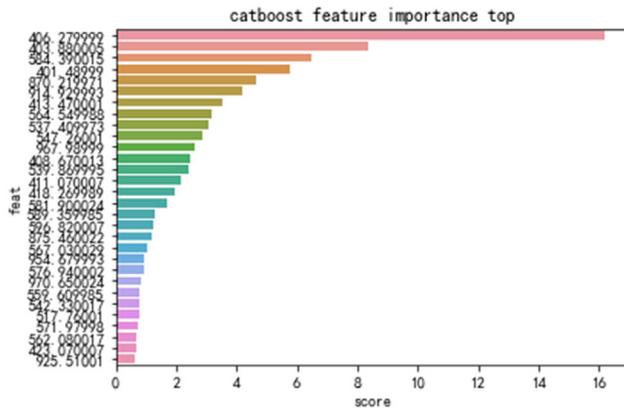
From the optimized performance indicators we can see in Table 3, it can be seen that each indicator of the two algorithms has been improved, which proves that the optimization is successful. In terms of refinement, the improvement of LightGBM is huge, and the LightGBM algorithm stands out in the feature band modeling extracted by it, especially in the performance of fit\_time. The unoptimized 1.625 s need only 0.098 s, which is 16 times of improvement. Therefore, the optimal peanut classification model is selected based on the performance indexes of all links, namely: MF-LightGBM-LightGBM-Optuna-LightGBM visualized confusion matrix for the optimal model is shown below. It can be seen that in the model, one of the

**Table 1.** Pretreatment results of spectral data.

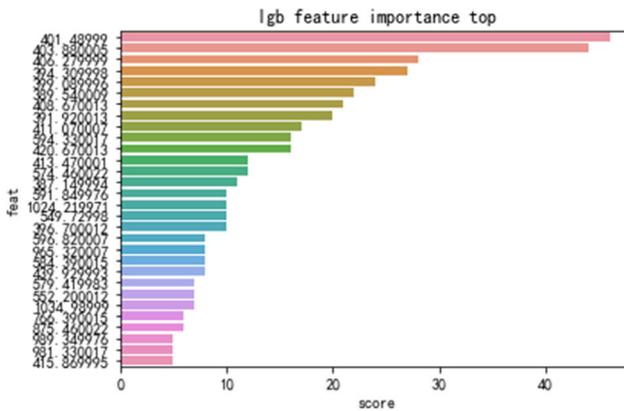
Method	ACC	Fscore_ dabaisha	Fscore_ huayu16	Fscore_ xiaobaisha	Fscore_ luhua	Fscore_ haihua	Log_Loss	Hamming_ Loss	Jaccard_ Similarity	fit_time
ORG	93.17%	94.31%	90.52%	94.98%	94.74%	91.14%	5254.25	0.07	0.93	0.88
LTN	93.17%	94.31%	90.52%	94.98%	94.74%	91.14%	5254.25	0.07	0.93	1.20
FD	93.33%	92.24%	96.46%	95.50%	91.39%	91.67%	5219.71	0.07	0.93	1.29
SD	93.17%	94.74%	90.43%	94.98%	94.74%	90.76%	5254.25	0.07	0.93	1.22
GWS	93.17%	94.74%	90.43%	94.98%	94.74%	90.76%	5254.25	0.07	0.93	1.22
BS	93.33%	94.74%	90.13%	94.98%	94.98%	91.74%	5219.71	0.07	0.93	1.18
LR1	92.67%	93.50%	89.36%	95.85%	94.30%	90.38%	5357.86	0.07	0.93	1.16
LR2	93.00%	93.88%	90.52%	95.45%	94.34%	90.76%	5288.79	0.07	0.93	1.15
L2NN	93.33%	95.16%	90.09%	92.79%	96.97%	90.98%	5219.71	0.07	0.93	1.30
MMS	83.50%	90.16%	80.17%	83.98%	90.30%	71.11%	7257.50	0.17	0.84	1.01
MAM	92.33%	93.39%	88.51%	94.55%	95.42%	89.63%	5426.94	0.08	0.92	1.16
MSC	91.67%	94.78%	85.84%	92.86%	96.53%	87.60%	5565.10	0.08	0.92	1.29
SG	93.17%	94.31%	90.52%	94.98%	94.74%	91.14%	5254.25	0.07	0.93	1.18
CAN	93.17%	94.31%	90.52%	94.98%	94.74%	91.14%	5254.25	0.07	0.93	1.17
ES	91.83%	92.18%	89.27%	94.55%	94.34%	88.70%	5530.56	0.08	0.92	1.17
MF	98.83%	98.80%	98.69%	99.55%	98.84%	98.33%	4079.93	0.01	0.99	0.96
ZSS	93.17%	94.31%	90.52%	94.98%	94.74%	91.14%	5254.25	0.07	0.93	1.16
WTD	93.17%	94.31%	91.23%	94.98%	94.74%	90.46%	5254.25	0.07	0.93	0.74



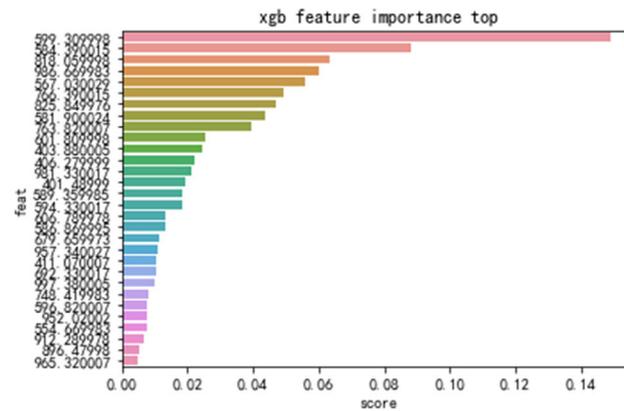
(a)



(b)



(c)



(d)

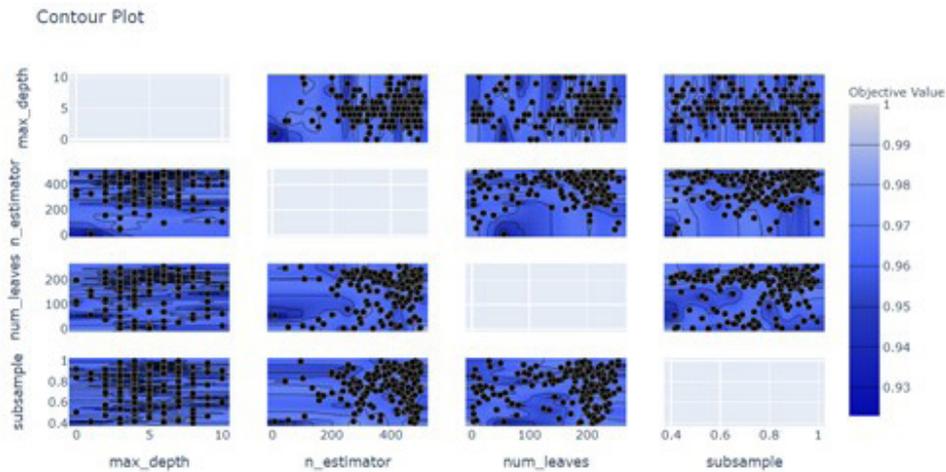
**Figure 5.** Extraction of important feature bands using GBDT(a),LightGBM(b) CatBoost, (c) and XGBoost(d).

**Table 2.** XGBoost and LightGBM classification performance indicators.

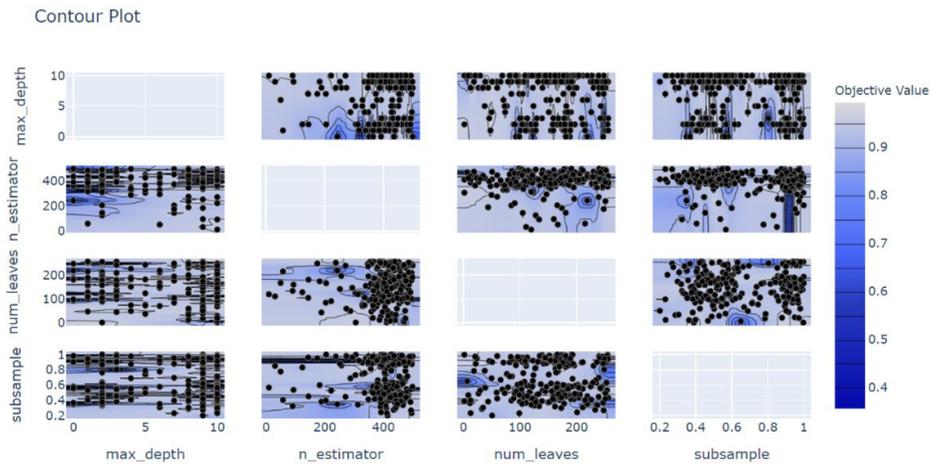
Model	Characters	ACC	Fscore_ Dabaisha	Fscore_ Huayu	Fscore_ Haihua	Fscore_ Luhua	Fscore_ Xiaobaisha	Log_Loss	Hamming_Loss	Jaccard_Similarity	fit_time
<b>XGB</b>	<b>xgb</b>	0.993	0.988	1.000	1.000	0.996	0.983	3976.313	0.007	0.993	1.384
	<b>lgb</b>	0.993	0.984	1.000	1.000	0.996	0.987	3976.313	0.007	0.993	1.555
	<b>cat</b>	0.985	0.980	0.987	0.995	0.989	0.975	4149.007	0.015	0.985	13.686
	<b>gbdt</b>	0.992	0.988	0.996	0.995	0.992	0.988	4010.852	0.008	0.992	6.648
<b>LGB</b>	<b>xgb</b>	0.993	0.988	1.000	1.000	0.996	0.983	3976.313	0.007	0.993	1.318
	<b>lgb</b>	0.993	0.984	1.000	1.000	0.996	0.987	3976.313	0.007	0.993	1.625
	<b>cat</b>	0.985	0.980	0.987	0.995	0.989	0.975	4149.007	0.015	0.985	13.490
	<b>gbdt</b>	0.992	0.988	0.996	0.995	0.992	0.988	4010.852	0.008	0.992	6.936

**Table 3.** Optimized classification performance indicators of XGBoost and LightGBM.

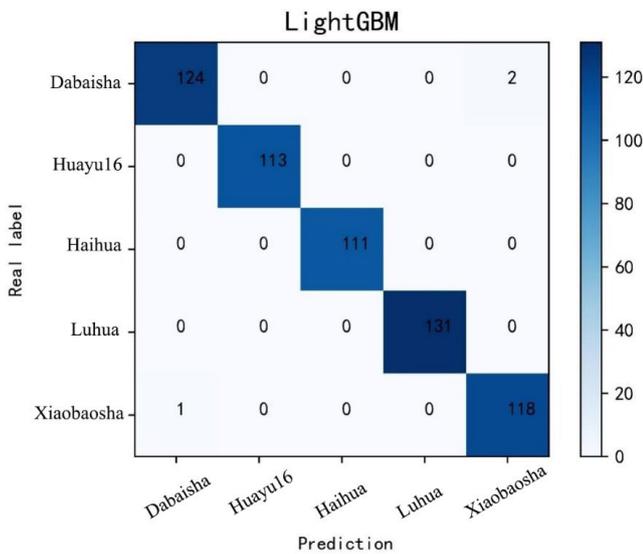
Model	Characters	ACC	Fscore_ Dabaisha	Fscore_ Huayu	Fscore_ Haihua	Fscore_ Luhua	Fscore_ Xiaobaisha	Log_Loss	Hamming_Loss	Jaccard_Similarity	fit_time
<b>XGB</b>	<b>xgb</b>	0.995	0.992	0.996	1.000	0.996	0.992	3941.774	0.005	0.995	1.087
	<b>lgb</b>	0.995	0.988	1.000	1.000	1.000	0.987	3941.774	0.005	0.995	1.073
<b>LGB</b>	<b>xgb</b>	0.995	0.992	0.996	1.000	0.996	0.992	3941.774	0.005	0.995	1.078
	<b>lgb</b>	0.995	0.988	1.000	1.000	1.000	0.987	3941.774	0.005	0.995	0.098



**Figure 6.** LightGBM optimization process.



**Figure 7.** XGBoost optimization process.



**Figure 8.** Confusion matrix after LightGBM optimization.

125 Dabaisha is mistakenly identified as Xiaobaisha, two of the 120 Xiaobaisha are mistakenly identified as Dabaisha, and the rest are not incorrectly identified. The classification effect is quite excellent (Figure 8).

## 4 Conclusion

Firstly, SpacVIEW was used to correct the black and white of peanut spectral images collected by Zhuolianguang hyperspectral instrument, and then ENVI was used to extract spectral data. Various pre-processing algorithms were used to de-noise the peanut spectral images, among which median Filtering (MF) had the best effect. The recognition rate reached 98.83%. XGBoost, LightGBM, GBDT and CatBoost algorithms were used to extract important feature bands in the spectral data preprocessed by MF. It can be seen that the optimal recognition variable is around 401nm. XGBoost and LightGBM algorithms are used for classification and modeling of feature bands. It can be seen from the model running results that LightGBM and XGBoost are not equal, and their recognition rate can reach 99.33%. Therefore, Optuna algorithm is used for parameter tuning of the two models. LightGBM has made a huge difference to XGBoost, especially in fit\_time, which is 11 times faster than XGBoost and 16 times faster than before optimization. Therefore, the optimal peanut classification model is selected based on the comprehensive analysis of all units, namely: MF-LightGBM-LightGBM-Optuna-LightGBM, the study of this method provides a strong theoretical basis and technical support for the revitalization of rural industry, the integration of agro-industry and the acceleration of the modernization of agricultural industrial system, production system and management system.

## References

Aboozar, T., & Georgina, C. T. M. M. (2020). AdaBoost-CNN: an adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning. *Neurocomputing*, 404, 351-366.

- Barbosa, R. M., Vieira, B. G. T. L., Martins, C. C., & Vieira, R. D. (2014). Physiological and health quality of peanut seeds during the production process. *Pesquisa Agropecuária Brasileira*, 49(12), 977-985. <http://dx.doi.org/10.1590/S0100-204X2014001200008>.
- Burns, D. A., & Ciurczak, E. W. (2007). *Handbook of near-infrared analysis*. 3rd ed. Boca Raton: CRC Press. <http://dx.doi.org/10.1201/9781420007374>.
- Camargo, A. C. D., Canniatti-Brazaca, S. G., Mansi, D. N., Domingues, M. A. C., & Arthur, V. (2011). Gamma radiation effects at color, antioxidant capacity and fatty acid profile in peanut (*Arachis hypogaea* L.). *Food Science and Technology*, 31(1), 11-15. <http://dx.doi.org/10.1590/S0101-20612011000100002>.
- Li, Z., Niu, B., Peng, F., Li, G., Yang, Z., & Wu, J. (2018). Classification of peanut images based on multi-features and SVM. *IFAC-PapersOnLine*, 51(17), 726-731. <http://dx.doi.org/10.1016/j.ifacol.2018.08.110>.
- Liu, B. (2011). On the advantages and strategies of peanut industry development in China. *Grain Science, Technology and Economy*, 36(1), 9-11.
- Liu, C. L., Lin, L., Yu, C. L., & Wu, J. Z. (2020). Study on peanut hyperspectral image classification method based on deep learning. *Computer Simulation*, 37(3), 189-192.
- Ma, W. Q., Zhang, M., Li, Y., Li, M. Z., Yang, L. L., Zhu, Z. J., & Cui, K. B. (2021). Quality detection and classification method of walnut kernel based on Hyperspectral imaging. *Chinese Journal of Analytical Chemistry*, 1-12.
- Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGBoost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, 31(8), 24-39. <http://dx.doi.org/10.1016/j.elerap.2018.08.002>.
- Ma, Z. F., Tian, H. P., Liu, Z. C., & Zhang, Z. W. (2020). A new incomplete pattern belief classification method with multiple estimations based on KNN. *Applied Soft Computing*, 90, 106175. <http://dx.doi.org/10.1016/j.asoc.2020.106175>.
- National Bureau of Statistics of the People's Republic of China. (2019). *China statistical yearbook*. China: China Statistics Press.
- Preys, S., Roger, J. M., & Boulet, J. C. (2008). Robust calibration using orthogonal projection and experimental design. Application to the correction of the light scattering effect on turbid NIR spectra. *Chemometrics and Intelligent Laboratory Systems*, 91(1), 28-33. <http://dx.doi.org/10.1016/j.chemolab.2007.10.007>.
- Qiao, X. J., Jiang, J. B., Li, H., Qi, X. T., & Yuan, D. S. (2018). Spectral Analysis and Index Models to Identify Moldy Peanuts Using Hyperspectral Images. *Guangpuxue Yu Guangpu Fenxi*, 38(02), 535-539. [In Chinese]
- Qin L., Han S., & Liu H. (2015). Research status of edible peanut in China. *Jiangsu agricultural sciences*, 43(11), 4-7.
- Quiroga, K., Bacca, J., & Arguello, H. (2020). Classification of cocoa beans based on their level of fermentation using spectral information. *Tecnológicas*, 24(50), e1654.
- Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control*, 73, 103456. <http://dx.doi.org/10.1016/j.bspc.2021.103456>.
- Tang, S., Yu, S., Liao, B., Zhang, X., & Sun, H. (2010). Industry status, existing problems and development strategy of peanut in China. *Journal of Peanut Science*, 39(3), 35-38.
- Thongsuwan, S., Jaiyen, S., Padcharoen, A., & Agarwal, P. (2021). ConvXGB: a new deep learning model for classification problems based on CNN and XGBoost. *Nuclear Engineering and Technology*, 53(2), 522-531. <http://dx.doi.org/10.1016/j.net.2020.04.008>.

- Xin, L., Liu, Z., Dou, J., Yang, Z., Zhang, X., & Liu, Z. (2020). A robust white-light interference signal leakage sampling correction method based on wavelet. *Optics and Lasers in Engineering*, 133, 106156. <http://dx.doi.org/10.1016/j.optlaseng.2020.106156>.
- Yang, Q., Wu, Y., Cao, D., Luo, M., & Wei, T. (2021). A lowlight image enhancement method learning from both paired and unpaired data by adversarial training. *Neurocomputing*, 433, 83-95. <http://dx.doi.org/10.1016/j.neucom.2020.12.057>.
- Yu, H. W., Wang, Q., Liu, L., Shi, A. M., Hu, H., & Liu, H. Z. (2016). Research progress of hyperspectral image detection technology for grain and oil quality safety. *Guangpuxue Yu Guangpu Fenxi*, 36(11), 3643-3650. PMID:30199206.
- Yuan, D. S., Jiang, J. B., Qi, X. T., Xie, Z., & Zhang, G. (2020). Selecting key wavelengths of hyperspectral imagine for nondestructive classification of moldy peanuts using ensemble classifier. *Infrared Physics & Technology*, 111, 103518. <http://dx.doi.org/10.1016/j.infrared.2020.103518>.
- Żelasko, D. (2020). Transmission quality classification in Pay&Require multi-agent managed network Means of Machine Learning techniques. *Simulation Modelling Practice and Theory*, 103, 102106. <http://dx.doi.org/10.1016/j.simpat.2020.102106>.
- Zhang, X., Kano, M., & Li, Y. (2017). Locally weighted kernel partial least squares regression based on sparse nonlinear features for virtual sensing Nonlinear time-varying processes. *Computers & Chemical Engineering*, 104, 164-171. <http://dx.doi.org/10.1016/j.compchemeng.2017.04.014>.
- Zou, Z. Y., Wu, X. W., Chen, Y. M., Bie, Y. B., Wang, L., & Lin, P. (2019). Study on hyperspectral image characteristic Response characteristics of potato under cryogenic freezing and mechanical damage. *Guangpuxue Yu Guangpu Fenxi*, 39(11), 3571-3578.