Analysis of Aflatoxin B1 in Peanut Oil by near Infrared Spectroscopy

Zhui Hu, Zihao Cai, Xingxing Yang, Qiang Xu, Xiao Zheng*, Jianping Ye and Junsheng Yang

School of Mechanical Engineering, Wuhan Polytechnic University

*Corresponding Author Email: zhengxiao@whpu.edu.cn

Abstract: Aflatoxin is a naturally highly toxic substance. Aflatoxin B1 is the most common one, and is usually detected in moldy peanut, corn and other foods. In this paper, for the detection of aflatoxin, near-infrared spectroscopy technology was applied, and three preprocessing methods, two characteristic wavelength extraction algorithms and three parameter optimization methods were adopted to establish a variety of models for the quantitative prediction of aflatoxin B1 content in peanut oil. The influences of different spectral preprocessing methods, characteristic wavelength extraction methods, and parameter optimization algorithms on the prediction performances of these models were studied. Additionally, the ways to find an optimal quantitative prediction model and qualitative identification model were explored.

1. Introduction
The annual peanut oil consumption in China accounts for about 50% of that in the world. Currently, peanut oil has become the first choice of most consumers. However, peanuts are prone to mildew. Aflatoxin B1 content is an important index to the quality of peanut oil, because it is a highly toxic and carcinogenic natural secondary metabolite produced by Aspergillus flavus and Aspergillus parasitica [1]. Acute poisoning will occur when a large amount of Aflatoxin B1 is digested. And, Aflatoxin B1 is one of the strongest carcinogens ever known [2]. The content of aflatoxin B1 is directly related to the edible safety of peanut oil. It is very important to develop a simple, rapid and effective oil detection technology. With the development of technology, more and more advanced technologies have been applied to oil detection, and certain goals have been achieved. The primary detection methods for aflatoxin B1 include thin layer chromatography (TLC), high performance liquid chromatography (HPLC) and enzyme-linked immunosorbent assay (ELISA) [3]. Compared to the above methods, NIR spectroscopy is faster and simpler. In the determination by near infrared spectroscopy, the samples are irradiated, and the absorption region of the near-infrared spectral region is consistent with that of the frequency doubling and combination band of molecular vibration of hydrogen containing groups in the oil samples. NIR can reflect the information of specific components in the samples, so NIR can be used to detect the characteristic components in the oil samples [4]. Near infrared spectroscopy (NIR) is a simple, rapid, effective and non-destructive method for analysis and detection. It has many advantages such as low cost and time saving, and has been widely used in many fields. In this paper, aflatoxin B1 in peanut oil was quantitatively detected and qualitatively identified by NIR.

2. Sample Preparation and Spectra Acquisition

2.1. Preparation of Samples
Preparation of aflatoxin-containing peanut oil was necessary in the experiment. Generally, the
formation of aflatoxin in peanut oil is caused by the mildewing of peanut raw materials. The moldy peanut is directly used to prepare peanut oil, so aflatoxin will remain in peanut oil. Therefore, moldy peanut was cultured in our laboratory. For convenience, peanut materials with high moisture contents were purchased in the market and placed in humid environment in the laboratory. And, water was regularly sprayed on the raw materials to keep the environment moist to facilitate mildew. After a certain period of cultivation, peanut was moldy, and then a part of the peanut was sampled at intervals. The moldy peanut was pressed with a small-sized expeller to derive crude peanut oil containing aflatoxin. Then, a total of 40 oil samples were prepared through filtration and centrifugation. The real contents of aflatoxin B1 in the samples were determined by chromatography.

2.2. Acquisition of Spectra of Samples
A custom-made laser near-infrared (NIR) vegetable-oil-quality detector was used in the experiments. The host was an Axsun XL410-type laser near-infrared spectrometer (The United States), with the scanning range of 1350–1800 nm and spectral resolution of 3.5 cm⁻¹. The scanning was performed 32 times. The temperature could be adjusted within the range of 20–100 °C. In this experiment, the spectrum of each sample was acquired 60 °C three times, and 120 original spectra were obtained, as shown in Figure 4.1.

![Figure 1. Original spectra of the samples tested](image)

3. Models for Quantitative Prediction
SNV-DT [5] - [6], MAS [7] and MSC [8] preprocessing methods were adopted. CARS-PLS [9] and SPA [10] algorithms were used for the extraction of feature wavelengths. The NONE algorithm [11] without the extraction of feature wavelengths was also adopted for comparison. Three parameter optimization methods including GS [12], GA [13], and PSO [14] - [15] were used to optimize the parameters. Finally, SVM [16] was used for the modeling.

The spectral data of 120 samples were divided into calibration and prediction sets with a ratio of 3:1. A total of 27 SVR quantitative prediction models were established by combining different algorithms. The parameters of each model and prediction results for the content of aflatoxin B1 in peanut oil are shown in Table 1.
Table 1. Quantitative prediction results of the contents of aflatoxin B1 in peanut oil

| Data processing method | Parameters | Calibration set | Prediction set |
|-----------------------|------------|-----------------|----------------|
| SNV-DT-SPA-PSO         | 100        | 931.0176        | 66.3183        | 31.3032        |
| SNV-DT-SPA-GA          | 27.5993    | 999.9762        | 36.4247        | 13.4023        |
| SNV-DT-SPA-GS          | 16         | 1024            | 32.6197        | 13.9565        |
| SNV-DT-CARS-PLS-PSO    | 1000       | 851.3131        | 28.8926        | 35.0238        |
| SNV-DT-CARS-PLS-GA     | 998.8165   | 995.5406        | 29.4428        | 35.5622        |
| SNV-DT-CARS-PLS-GS     | 1024       | 1024            | 29.6924        | 35.6046        |
| SNV-DT-NONE-PSO        | 810.1101   | 22.1427         | 100            | 97.3874        |
| SNV-DT-NONE-GA         | 706.6231   | 21.6437         | 100            | 97.3615        |
| SNV-DT-NONE-GS         | 1024       | 16              | 99.9989        | 97.1926        |
| MAS-SPA-PSO            | 1000       | 322.2271        | 100            | 99.8759        |
| MAS-SPA-GA             | 999.0063   | 11.4718         | 93.4676        | 93.1775        |
| MAS-SPA-GS             | 1024       | 256             | 100            | 99.8614        |
| MAS-CARS-PLS-PSO       | 386.062    | 307.045         | 99.9967        | 99.8522        |
| MAS-CARS-PLS-GA        | 999.6281   | 8.0566          | 91.9606        | 94.9018        |
| MAS-CARS-PLS-GS        | 1024       | 256             | 100            | 99.8614        |
| MAS-NONE-PSO           | 1000       | 54.5672         | 100            | 99.8509        |
| MAS-NONE-GA            | 974.3509   | 1.9503          | 93.7958        | 93.8927        |
| MAS-NONE-GS            | 256        | 64              | 100            | 99.8468        |
| MSC-SPA-PSO            | 1024       | 214.9902        | 91.347         | 89.0692        |
| MSC-SPA-GA             | 999.9485   | 260.669         | 91.7137        | 89.4407        |
| MSC-SPA-GS             | 1024       | 256             | 91.6972        | 89.4251        |
| MSC-CARS-PLS-PSO       | 1000       | 486.0703        | 93.0087        | 91.2801        |
| MSC-CARS-PLS-GA        | 999.4478   | 344.0374        | 92.4275        | 91.1531        |
| MSC-CARS-PLS-GS        | 1024       | 256             | 92.0091        | 90.8308        |
| MSC-NONE-PSO           | 1000       | 31.455          | 91.4674        | 89.1096        |
| MSC-NONE-GA            | 998.9729   | 68.4796         | 94.427         | 91.4388        |
| MSC-NONE-GS            | 1024       | 16              | 90.6971        | 88.6759        |

By analyzing and comparing the prediction results and parameters of various models in Table 1, it was found that the combination of MAS algorithm and PSO and GS parameter optimization algorithms showed the best performance. The correlation coefficient R of the calibration set was close to 1, and the correlation coefficient R of the prediction set was greater than 0.99. The model parameters (C, g) were in appropriate ranges. This model had a strong learning ability, and can be applied in practice. Moreover, the correlation was great, and the prediction error was small, showing a promising application prospect. The model with the combination of near infrared spectroscopy and support vector machine regression can be applied to fast and effectively quantitatively predict the content of aflatoxin B1 in peanut oil. The prediction results of the optimal MAS-SPA-GS model are shown in Figs. 2 and 3.
4. Models for Qualitative Identification

On the basis of the MAS-SPA-GS model, the qualitative model for the identification of the presence of aflatoxin B1 in peanut oil was established. According to the content of aflatoxin B1, the detection results were divided into three levels. Considering that aflatoxin B1 contents lower than 1 μg/kg were difficult to detect by current technologies, peanut oil samples with aflatoxin B1 contents lower than 1 μg/kg were classified as level I, denoted by 0. Considering the fact that 20 μg/kg was slightly higher than the China national standard of 20 μg/kg, the range of 1–16 μg/kg was classified as level II, denoted by 1, indicating a trace content. Considering the detection error, the aflatoxin B1 contents higher than 16 μg/kg were classified as level III, denoted by 2, indicating a high content. The output values of the functions of qualitative models included 0, 1 and 2. Accordingly, the qualitative model for the identification of aflatoxin B1 in peanut oil were established by the support vector machine classification method. The parameters and prediction results of these models are shown in Table 2.

Table 2. Qualitative prediction results for the presence of aflatoxin B1 in peanut oil

| Modeling method | Parameters (C, g) | Calibration set | Prediction set |
|-----------------|------------------|-----------------|---------------|
|                 | C                | Accuracy (%)    | Number of correct prediction cases | Accuracy (%) | Number of correct prediction cases |
| PSO-SVC         | 3.2263           | 97.7778         | 88/90         | 96.6667       | 29/30 |
| GS-SVC          | 256              | 100             | 90/90         | 100           | 30/30 |
| GA-SVC          | 10.2043          | 830.6149        | 100           | 100           | 30/30 |

Table 2 shows that the accuracy rates of both the calibration and prediction sets were greater than 96%, and the accuracy rates of the calibration and prediction sets of GS-SVC and GA-SVC models were as high as 100%, showing that the three optimization methods are very reliable and can effectively identify aflatoxin B1. In addition, the errors of the collected data were small and the data were accurate, so the errors had negligible impact on the establishment of quantitative models.

5. Conclusions

The applications of near infrared spectroscopy in the qualitative identification of aflatoxin B1 and
quantitative prediction of the aflatoxin B1 content in peanut oil were successfully conducted. The support vector machine regression model for the quantitative prediction of aflatoxin B1 content in peanut oil and support vector machine classification model for the classification and qualitative identification of the presence of aflatoxin B1 in peanut oil can realize the prediction of the content of aflatoxin B1 in peanut oil and classification of the content of highly toxic aflatoxin B1. Three different preprocessing methods were combined with two different extraction methods for feature wavelengths, and the combinations were analyzed and compared. The modeling performances of three parameter optimization methods were compared. The results showed that the MAS preprocessing method was superior for the modeling. The correlation coefficient R values of most models were greater than 0.95, and the average relative errors were almost smaller than 5%. MSC exhibited the best performance among all the preprocessing methods. Although the final modeling performance was not as good as those of the models preprocessed by MAS, the models were relatively stable without evident fluctuations. In contrast, the models established after SNV-DT preprocessing had poor performances, and the extraction of characteristic wavelengths negatively affected the prediction performances of these models. Compared with the parameter optimization methods of the models established after MAS preprocessing, it was found that the grid search method and particle swarm optimization algorithm were significantly better than the genetic algorithm, and the parameters (C, g) obtained were better. Finally, the correlation coefficient R value of the calibration set of the optimal MAS-SPA-GS model was as high as 1. Based on this MAS-SPA-GS model, three SVC models for the qualitative identification of aflatoxin B1 were built, and the accuracy rates of the calibration and prediction sets of GA-SVC and GS-SVC models were as high as 100%. This study provides a new way for the detection of the quality of edible oils. Compared to the complexity of high performance liquid chromatography, the near-infrared spectroscopy approach is simpler and faster.

6. References
[1] Shi Chunyue. Research progress on detection technology for aflatoxin in food [J]. Agricultural Products Processing, 2019 (14): 86-89
[2] Xu Weihao, Liu Yanming, Liu Jianlong. Technical approaches to reduce aflatoxin contents in agricultural products [J/OL]. China Fruit & Vegetable: 1-7 (2019-10-19).1
[3] Determination of aflatoxins B and G in foods (GB 5009.22-2016)
[4] Hu Weiping, Liu Yingchun, Meng Wei. Near infrared spectroscopy analysis of the viscosity of crude oil in Northern Xinjiang [J]. Chemical Management, 2016 (29): 1291
[5] Ying Guo, Ni Yongnian, Serge K. Evaluation of chemical components and properties of the jujube fruit infrared spectroscopy and chemometrica [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2016, 153: 79-861
[6] Wu Shuang. Study on the spectral quantitative detection of iodine value and saponification value of edible oils and identification of oil types by characteristic values [D]. Wuhan University of Light Industry, 2017
[7] Luo Qingsong. Spectral quantitative detection of acid value and aflatoxin content of edible oils and identification of oil types by characteristic values [D]. Wuhan University of Light Industry, 2019
[8] Lu Yongjun, Qu Yanling, Song Min. Study on the multivariate scattering correction processing of near-infrared-related spectra [J]. Spectroscopy and Spectral Analysis, 2007 (05): 877-880
[9] Wu Jingzhu, Xu Yun. Optimization of near infrared quantitative CARS-PLS-based models for the analysis of fatty acids in edible oils [J]. Transactions of The Chinese Society of Agricultural Machinery, 2011, 42 (10): 162-166
[10] Wu Di, Wu Hongxi, Cai Jingbo, Huang Zhenhua, He Yong. Study on the classification methods of white shrimp species by visible and near infrared spectroscopy based on uninformative variable elimination method and continuous projection algorithm [J]. Acta Infrared And Millimeter Wave, 2009, 28 (06): 423-427
[11] Tu bin. Study on the rapid detection method of edible oil quality and type based on the fusion of near infrared and Raman spectra [D]. Wuhan University of Light Industry, 2016
[12] Sun Ling. Research and application of a support vector machine method for the selection of
parameters [D]. Hangzhou University of Electronic Science and Technology, 2017

[13] Ma Yongjie, Yun Wenxia. Research progress of genetic algorithm [J]. Computer Application Research, 2012, 29 (04): 1201-12

[14] Particle Swarm Optimization [M]. New York: IEEE, 1995

[15] Liu Yanmin. Research and application of particle swarm optimization algorithm [D]. Shandong Normal University, 2011

[16] Du Zhe. Research on several variants of support vector machine algorithm [D]. Xi’an University of Electronic Science and Technology, 2009