

## LIBS 中药材石斛等级识别研究

郑培超<sup>1</sup>, 郑爽<sup>1</sup>, 王金梅<sup>1\*</sup>, 廖香玉<sup>1</sup>, 李晓娟<sup>1</sup>, 彭锐<sup>2</sup>

1. 重庆邮电大学光电工程学院, 光电信息感测与传输技术重庆重点实验室, 重庆 400065
2. 重庆市中药研究院, 重庆 400065

**摘要** 石斛是一种常用的中药材,经常使用新鲜的或干燥的茎条入药,有益胃生津、滋阴清热的效果。近年来,药理学研究探索出石斛具有抗白内障、抗氧化、抗肿瘤、提高免疫力的作用,其在许多病例中疗效显著,引起了国内外学者的关注,然而不同时间采集的石斛中氨基酸、微量元素等含量各不同,其对应药用价值,价格也不同,因此石斛价格等级分辨的研究具有重要意义。为快速鉴别不同价格、不同药效的石斛,研究了随机森林分类模型结合激光诱导击穿光谱技术(LIBS)对石斛价格等级进行分析建模。选取5个等级的石斛样品进行建模,为了对样品进行精确稳定分析,所有石斛样品均通过粉碎压片减小实验误差。采用1064 nm波长的Nd:YAG脉冲激光器作为激发光源,设置激光脉冲能量50 mJ,探测延时1  $\mu$ s,采集五个等级石斛样本的光谱数据,每个等级的样本采集40组光谱,共200组数据,并采用归一化处理,使所有的光谱数据转换到-1~1之间。采用归一化处理后的光谱数据进行主成分分析,通过主成分分析获得前7个主成分的得分矩阵,其累计解释95.24%的光谱信息。将选取的7个主成分作为输入,建立波段为220~880 nm的随机森林鉴别模型。并将石斛样本编号打乱,任意选取50%的光谱数据作为训练集,剩下50%的光谱数据作为测试集,默认决策树个数 $n_{tree}$ 为500,分裂属性集中属性个数 $m_{try}$ 为5,建立模型对不同等级的石斛进行分类。等级一、二、三、四、五的识别率分别为95.45%,100%,78.26%,94.12%和85%,平均识别率为90.57%。为提高识别率,研究了不同的 $n_{tree}$ 和 $m_{try}$ 对分类模型的影响,利用袋外数据误差率估计对随机森林的两个参数进行了优化。选择 $n_{tree}$ 为300, $m_{try}$ 为1,等级一、二、三、四、五的识别率分别为100%,100%,92.31%,100%和90%,平均识别率为96.46%,识别率提高了5.89%。综上所述,采用LIBS技术结合优化后的随机森林模型鉴别石斛等级具有一定的可行性,为未来快速鉴定不同价格的石斛等级分类提供了可行性的判别系统。

**关键词** 中药材石斛; LIBS; 随机森林; 等级识别

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)03-0941-04

### 引言

石斛一年四季均可采摘,但不同产地、不同采收时间的石斛,功效具有一定的差异,在适合的时间段采摘才可获得药用价值比较高的石斛。石斛在春天萌发,夏天生长,但是到了十月份左右便不再继续生长,冬天是石斛营养储蓄的阶段。这个阶段石斛内部的光合作用不是很强,含有的氨基酸、多糖和微量元素都处在一年中的峰值;其药用成分累积在茎条中,这时采摘的石斛比较适用于药用<sup>[1]</sup>。然而石斛中部分未知成分的存在,在分析和质量控制方面存在很大的挑

战,严重限制石斛的广泛应用,因此,石斛的鉴别对其使用和加工具有重要意义。常用的中药材检测技术,如原子薄层色谱法<sup>[2]</sup>、气相色谱分析<sup>[3]</sup>和电感耦合等离子体发射光谱法<sup>[4]</sup>等。虽然具有较高的检测精度、稳定度等优点,然而需要对实验样品进行繁琐的化学预处理,检测周期长同时容易产生二次污染,所以,迫切需要发展一种方便、迅速的中药材检测技术。

激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)是近几年以来迅速发展的一种多元素分析技术,具有快速响应、远程检测能力、高灵敏度和样品制备简单等固有优势,被广泛应用于地质勘探<sup>[5]</sup>、能源分析<sup>[6-7]</sup>和食品

收稿日期: 2019-03-01, 修订日期: 2019-07-14

基金项目: 重庆市留学归国人员创新创业项目(cx2017126, cx2018127)资助

作者简介: 郑培超, 1980年生, 重庆邮电大学光电工程学院教授 e-mail: zhengpc@cqupt.edu.cn

\* 通讯联系人 e-mail: wangjm@cqupt.edu.cn

安全<sup>[8-9]</sup>等。近年来, LIBS 技术逐渐应用于中药材的检测, 刘晓娜<sup>[10]</sup>等采用 LIBS 对树脂类药材的元素进行快速分析, 通过主成分分析和偏最小二乘判别方法对 3 类树脂类药材进行鉴别。Rai 等<sup>[11]</sup>使用 LIBS 对草药制剂中存在的元素种类进行分析。我们课题组<sup>[12-13]</sup>采用 LIBS 对中药川芎中铅(Pb)和铜(Cu)进行了定量分析, 结合主成分分析和人工神经网络对不同产地和部位的当归、党参、白芷根等进行了分析鉴定。然而, 少有关于利用 LIBS 技术对不同等级石斛的识别和定量分析的报道。

不同时间采集的石斛中氨基酸、多糖、微量元素含量各不相同, 价格也不同, 在中药材市场, 通常会以出售价格和品相区分不同等级规格的石斛, 但是当人工无法辨别时, 可能引起相似药材的以假充次, 导致治疗效果低甚至损害人的身体。因此, 石斛的等级识别对药材的使用是有帮助的, 为此, 基于中药材石斛样品的 LIBS 光谱图建立了随机森林分类(random forests, RF)模型, 实现了石斛等级的快速鉴别。

## 1 实验部分

### 1.1 随机森林分类模型原理

RF 分类模型由  $i$  个决策树模型  $\{H(x, \Theta_i), i=1, 2, \dots, i\}$  组合而成, 其中  $\{\Theta_i, i=1, 2, \dots, i\}$  为独立同分布的随机向量, 通过  $i$  次训练, 获得一个分类模型序列组  $\{H_1(x), H_2(x), \dots, H_i(x)\}$ <sup>[14]</sup>, 用其构建一个多分类的系统, 最优分类的结果为获得票数最多的, 最终的分类决策

$$h(x) = \arg \max_y \sum_{i=1}^k M(H_i(x) = y)$$

其中,  $h(x)$  为组合分类模型,  $H_i$  为单个决策树分类模型,  $y$  表示输出变量,  $M(\cdot)$  为示性函数<sup>[14]</sup>。每次构建决策树时, 通过重复抽样得到一个数据用于训练决策树, 还有大约  $1/e$  ( $e$  为自然常数) 的数据未被利用, 该部分数据可以用于对决策树的性能进行评估, 计算模型的预测错误率, 称为袋外数据误差(out-of-bag error, OOB error)。

### 1.2 样本制备

所用的石斛样品均采购于重庆中药材市场, 石斛样品均产自于同一产地(浙江), 由于不同采收时间的石斛中微量元素含量各不相同, 价格有一定的差异, 根据价格对石斛设定五个等级, 等级一、二、三、四、五对应的市场价分别为 90 元/50 g、70 元/50 g、65 元/50 g、60 元/50 g、30 元/50 g。为了对样品进行精确稳定分析, 所有石斛样品均通过粉碎压片的方式减小实验误差。

### 1.3 LIBS 装置

实验装置如图 1 所示, 激发光源为 1 064 nm 波长 Nd:YAG 脉冲激光器(Quantel Ultra 100, 法国 Quantel), 重复频率 20 Hz 可调, 输出的激光能量 1~100 mJ 可调。样品直接放置于三维移动平台上, 激光器发出的高能激光经反射镜反射之后, 由焦距为 100 mm 的石英平凸透镜聚焦到位于平台上的石斛样品表面。样品被激发产生高温等离子体, 发出的光由光纤耦合传输到多通道光谱仪(MX2500+, 海洋光学), 光谱范围达 198~877 nm, 最高分辨率优于 0.1 nm, 最

后由电脑记录采集的光谱。

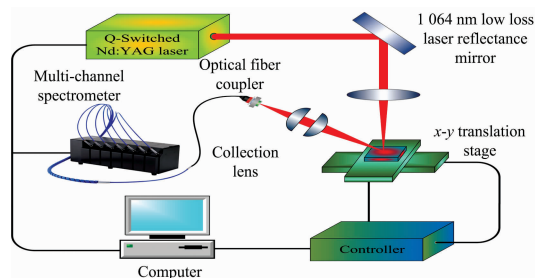


图 1 LIBS 装置示意图

Fig. 1 LIBS system

根据课题组之前的研究经验<sup>[12-13]</sup>选取激光脉冲能量为 50 mJ, 探测延时为  $1 \mu\text{s}$ , 利用 MaxLIBS(版本 1.0) 软件采集光谱数据, 在样品表面不同采样点进行采集, 光谱数据是 5 次采集光谱的平均值, 每个等级的石斛采集 40 组光谱, 共 200 组数据用于后续分析。

## 2 结果与讨论

### 2.1 光谱数据的获得

采集石斛样品 200~880 nm 范围的光谱图, 由图 2 可以看出, 在石斛等离子体光谱中出现了 Ca, Na, Al, K, Fe 和 Mg 等金属谱线, 同时也出现了 C, H 和 O 等元素谱线以及 CN 分子的辐射, 另外, 一些微量元素如 Cr 和 Cu 在光谱中也能看到。而在其他文献中所提到的石斛中的 Hg, As, Pb 和 Cd 等元素在本实验中未被发现<sup>[1]</sup>, 可能是这些元素在石斛中含量很少并且难以激发。

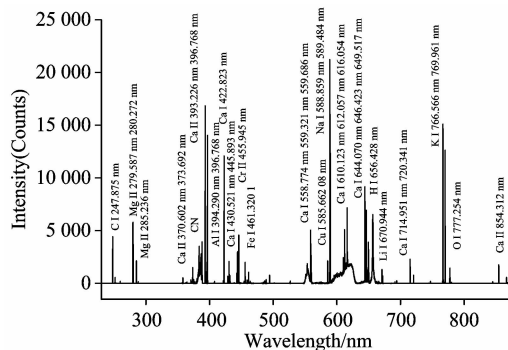


图 2 等级 1 的石斛样品等离子体光谱

Fig. 2 Plasma spectra of grade 1 dendrobium

### 2.2 光谱数据处理及结果

在光谱采集过程中, 样品环境变化的影响, 导致每次采集的光谱有轻微的差异, 为了补偿由环境变化引起的光谱变化, 所有的光谱数据均进行归一化处理, 使所有的数据转换到  $-1 \sim 1$  之间。归一化处理后的光谱数据采用 PCA 进行降维, 其中前 7 个主成分累积解释了全谱数据信息的 95.24%, 因此选取前七个主成分做进一步分析。

从图 3 可以看出, 等级 1 和 4 的样品能比较明显被识别, 显示出良好的聚类效果, 但其他三个样品大部分重叠,

为了提高光谱准确识别率,引入随机森林模型进行分析。以 7 个主成分作为输入,所有样本编号打乱,随机选取 50% 的光谱数据作为训练集,剩下 50% 的光谱数据作为测试集,默认决策树个数  $n_{tree}$  为 500 和分裂属性集中的属性个数  $m_{try}$  为 5,建立模型进行训练。最终的结果为同一数据集 30 次重复测试平均结果,每次运行结果均一致,分类模型具有良好的稳定性,分类结果如表 1。

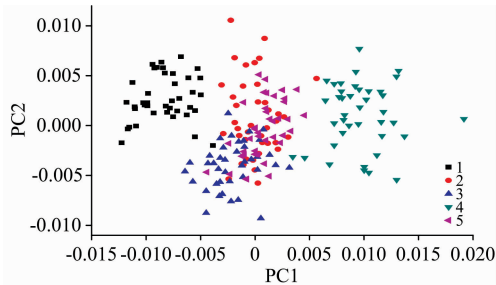


图 3 石斛主成分二维得分图

Fig. 3 Two-dimensional score of main components for dendrobium

表 1 随机森林分类模型结合 LIBS 对石斛的等级预测结果

Table 1 The prediction results of Dendrobium by RF classification model combined with LIBS

Grade	Rate of identification ( $n_{tree}=500, m_{try}=5$ )	Rate of identification ( $n_{tree}=300, m_{try}=1$ )
Grade 1	95.45%	100%
Grade 2	100%	100%
Grade 3	78.26%	92.31%
Grade 4	94.12%	100%
Grade 5	85%	90%

通过袋外误差估计研究了不同的  $n_{tree}$  和  $m_{try}$  对分类模型的影响,如图 4 所示,当  $n_{tree}$  达到一定值时,分类器的泛化误差趋于有限上界,即如果  $n_{tree}$  增加到最优值以上,运行速度一般会变慢,但结果并没有显著改善。从袋外错误率来看,

选择参数  $n_{tree}$  为 300,  $m_{try}$  为 1 为最优参数对五种不同等级的石斛进行分类,结果如表 1 所示。

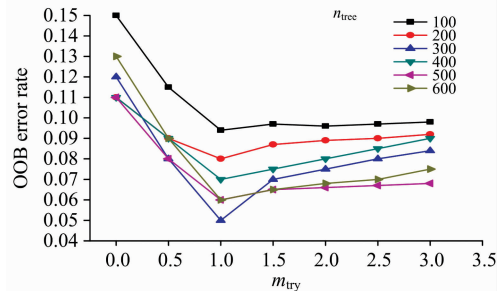


图 4 决策树个数和属性个数优化趋势图

Fig. 4 The trend diagram of the numbers of decision trees and attributes for optimization

可以看出优化后的随机森林模型对石斛的等级鉴别平均识别率为 96.46%,其中等级 1 的石斛是最优质的,其品质较其他等级有很大的区别,等级 4 的石斛为人工加工的耳环石斛,由于加工过程中破坏了样品的角质层,所以石斛形态及所含胶质有较大区别,易于鉴别出来。等级 3 与等级 5 分辨率较差,可能在主成分分析中,提取出的前七个主成分虽然阐述了石斛原始输入数据的大部分信息,消除了局部相互重叠的信息,但是没有考虑主成分对输出变量的解释能力,可能导致部分等级分辨率效果较差。

### 3 结 论

采用随机森林对五个不同价格等级的石斛进行鉴定,随机森林模型采用的 7 个主成分作为输入,默认决策树的颗数  $n_{tree}$  为 500,分裂属性集中的属性个数  $m_{try}$  为 5,平均识别率为 90.57%,通过优化,选择参数  $n_{tree}$  为 300,  $m_{try}$  为 1,平均识别率提高了 5.89%。这表明 LIBS 技术结合随机森林模型对不同等级的中药材进行鉴别是可行的,拓宽了中药材鉴别的方法,为未来快速鉴别中药材等级鉴别提供了新的可能性。

### References

- [1] JIANG Xue-yan, ZHANG Ling-ju, LI Ya, et al(蒋雪嫣,张玲菊,李 娅,等). Chinese Archives of Traditional Chinese Medicine(中华中医药学刊), 2018, 36: 1.
- [2] Pozzi F, Shibayama N, Leona M, et al. Journal of Raman Spectroscopy, 2013, 44(1): 102.
- [3] Cai Hao, Cao Gang, Zhang Hongyan, et al. Chinese Journal of Integrative Medicine, 2015, 23(4): 261.
- [4] Qasim M, Anwar-ul-haq M, Sher Afgan M, et al. Analytical Letters, 2016, 49(13): 2108.
- [5] Xia H, Dong F Z, Wu B, et al. Chinese Physics B, 2015, 24(3): 180.
- [6] Zhu Z, Li J, Guo Y, et al. Journal of Analytical Atomic Spectrometry, 2018, 33: 205.
- [7] Guo Y, Deng L, Yang X, et al. Journal of Analytical Atomic Spectrometry, 2017, 32: 2401.
- [8] Zheng P C, Liu H D, Wang J M, et al. Journal of Analytical Atomic Spectrometry, 2015, 30(4): 867.
- [9] Wang J M, Zheng P C, Liu H D, et al. Analytical Methods, 2016, 8(15): 3204.
- [10] LIU Xiao-na, ZHANG Qiao, SHI Xin-yuan, et al(刘晓娜,张 乔,史新元,等). China Journal of Traditional Chinese Medicine and Pharmacy(中华中医药杂志), 2015, 5(1): 1610.
- [11] Rai P K, Pathak A K, Ghatak S, et al. Journal of Food Measurement and Characterization, 2013, 7(3): 114.
- [12] Wang J M, Xue S W, Zheng P C, et al. Analytical Letters, 2017, 50(12): 1.
- [13] Wang Jinmei, Liao Xiangyu, Zheng Peichao, et al, Analytical Letters, 2018, 51: 4575.

- [14] FANG Kuang-nan, WU Jian-bin, ZHU Jian-ping, et al(方匡南, 吴见彬, 朱建平, 等). Statistics & Information Forum(统计与信息论坛), 2011, 26(3): 32.

## Study on Grade Identification of Dendrobium by LIBS

ZHENG Pei-chao<sup>1</sup>, ZHENG Shuang<sup>1</sup>, WANG Jin-mei<sup>1\*</sup>, LIAO Xiang-yu<sup>1</sup>, LI Xiao-juan<sup>1</sup>, PENG Rui<sup>2</sup>

1. Chongqing Municipal Level Key Laboratory of Photoelectronic Information Sensing and Transmitting Technology, College of Optoelectronic Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. Chongqing Academy of Chinese Medicine, Chongqing 400065, China

**Abstract** Dendrobium is a commonly used Chinese herbal medicine, often using fresh or dry stems into the medicine, beneficial to the stomach, nourishing yin and clearing heat. In recent years, pharmacological studies have found that Dendrobium has the functions of anti-cataract, anti-oxidation, anti-tumor and improving immunity. It has remarkable effects in many cases, which has attracted the attention of scholars at domestic and abroad. However, the contents of amino acids and trace elements in Dendrobium collected at different times are different, and their medicinal value and price are different. So the study of price grade discrimination of Dendrobium is of great significance. In order to quickly identify Dendrobium with different price and efficacy, the random forest classification model combined with laser induced breakdown spectroscopy (Laser-induced Breakdown Spectroscopy, LIBS) was developed to model the price grade of Dendrobium. In this paper, five samples of Dendrobium were selected for modeling. In order to analyze the samples accurately and stably, all Dendrobium samples were pressed to reduce the experimental error. The Nd : YAG pulse laser with 1 064 nm wavelength was used as the excitation light source, the detection delay of 50 mJ, laser pulse energy was set to 1  $\mu$ s, the spectral data of five grades of Dendrobium were collected, 40 sets of spectra were collected from each grade of samples, and a total of 200 sets of data were collected. Normalized processing was used to convert all spectral data from  $-1$  to  $1$ . The principal component analysis (PCA) was used to analyze the normalized spectral data. The score matrix of the first seven principal components was obtained by principal component analysis, and the cumulative interpretation of the total spectral information was 95.24%. So seven principal components were selected as input, and a random forest identification model with 220~880 nm was established. The number of Dendrobium samples was disrupted, and 50% spectral data were randomly selected as training sets, and 50% spectral data were left as test sets. The default number of decision trees ( $n_{tree}$ ) was 500, and the number of attributes in the split attribute set ( $m_{try}$ ) was 5. The model was established to classify Dendrobium in different grades. And the recognition rates of grades one, two, three, four and five were 95.45%, 100%, 78.26%, respectively. 94.12%, 85%, with an average recognition rate of 90.57%. In order to improve the recognition rate, the influence of different  $n_{tree}$  and  $m_{try}$  on the classification model was studied, and the two parameters of the random forest were optimized by using the out-of-bag data error rate estimation. The  $n_{tree}$  was 300, the  $m_{try}$  was 1, the recognition rates of grade one, two, three, four and five were 100%, 100%, 92.31%, 100%, 90%, the average recognition rate was 96.46%, and the recognition rate was increased by 5.89%. In conclusion, it is feasible to identify the Dendrobium classification by LIBS technology combined with the optimized random forest model, which provides a feasible discrimination system for the rapid identification of Dendrobium classification with different prices in the future.

**Keywords** Dendrobium; Libs; Random forests; Level to identify

(Received Mar. 1, 2019; accepted Jul. 14, 2019)

\* Corresponding author