

岑忠用,雷顺新,雷蕾,等. 近红外光谱法鉴别6种根茎类中药材[J]. 华中农业大学学报, 2021, 40(3): 271-277.
DOI:10.13300/j.cnki.hnlkxb.2021.03.031

近红外光谱法鉴别6种根茎类中药材

岑忠用¹, 雷顺新², 雷蕾², 严军³, 张晖英¹

1. 河池学院化学与生物工程学院, 宜州 546300; 2. 广西横县综合检验检测中心, 横县 530300;
3. 广西民族大学化学化工学院/广西高校食品安全与药物分析化学重点实验室, 南宁 530008

摘要 为建立山豆根、百两金、千斤拔、北豆根、滇豆根、云南豇豆等6种易混淆根茎类中药材的快速鉴别方法, 利用近红外漫反射光谱分别采集其在900~1700 nm内的光谱信息, 结合主成分分析、系统聚类分析、K近邻法、线性判别分析建立定性判别模型。结果显示, 6种药材在主成分分析及系统聚类分析中表现出了明显的分类聚集特征, K近邻法和线性判别分析对46个药材盲样的鉴别准确率分别达到了93.48%和95.65%。结果表明, 近红外漫反射光谱结合模式识别方法可用于根茎类中药材的定性鉴别。

关键词 近红外漫反射光谱; 中药材鉴别; 山豆根; 系统聚类; K近邻法; 线性判别分析; 模式识别

中图分类号 R 931.5; O 657.33 **文献标识码** A **文章编号** 1000-2421(2021)03-0271-07

近年来我国的自然生态环境逐渐恶化, 野生的中药材生长环境遭到破坏, 导致部分野生中药材供应短缺, 再加上中药材种类较多, 历代本草记载、地区用药名称和使用习惯不同, 类同品、代用品和民间用药不断出现, 中药材同名异物、同物异名、品种混乱现象比较普遍, 使得中药鉴别成为一个难题^[1]。很多中药材形态相似, 但其化学成分、性味、毒性、用量、药理作用和功能等方面均不完全相同, 因此可能因药材误用而产生严重后果。传统中药鉴别方法有性状鉴别、显微鉴别、理化鉴别等, 效率较为低下, 且对于亲缘关系近、生物形态相似度高的药材难以获得准确的鉴别结果^[2-4]。色谱以及DNA分子鉴别能获得极高的药材鉴别准确率, 但是往往需要复杂的前处理, 且分析时间长、成本高、操作繁琐, 难以进行原位在线的快速鉴别^[5-6]。近红外漫反射光谱具有分析速度快、成本低、无损、前处理简单等优点, 近年来在中药材质量控制领域得到了广泛应用^[7-9]。

根茎类中药指入药部分是根茎或带有少量根部或肉质鳞叶的地下茎类药材, 由于根茎类药材的植物形态相似性较高, 导致了中药材市场上真伪混淆难以辨别的问题, 并进一步引起严重的临床用药危险。近年来, 根茎类中药材如百两金、山豆根、千斤拔、滇豆根、北豆根等相似药材因地区名称和使用习

惯不同而被误用, 或者由于不法商人为了牟利而故意售卖易混淆药材, 研究者已针对这一问题建立了一系列鉴别方法^[10-11], 但所建立的方法多是依赖于性状鉴别和理化鉴别。本研究利用近红外漫反射光谱技术结合模式识别方法建立6种易混淆的根茎类中药材(山豆根、北豆根、滇豆根、百两金、千斤拔、云南豇豆)的定性鉴别模型, 以期对中药材的快速、准确鉴别和质量控制提供参考。

1 材料与方法

1.1 药材采集及鉴定

共采集6种易混淆中药材, 即百两金(BLJ)、山豆根(SDG)、千斤拔(QJB)、滇豆根(DDG)、云南豇豆(YNJD)、北豆根(BDG), 产地包括广西、云南、吉林等地。样品经由野外采集及中药材市场购买共2种方式获得, 所有中药材均由河池学院化学与生物工程学院鉴定。试验所用药材样品的信息如表1所示。

1.2 设备及参数

中药粉碎机(宝利, 中国江阴); 近红外光谱仪(必达泰克*i-Spec*型, 美国), 配置积分球采样附件及InGaAs检测器, 光谱采集参数: 扫描范围为900~1700 nm, 积分时间为4000 μ s, 扫描次数20次, 每个样本采集3次光谱, 并用平均光谱进行数据分析。

收稿日期: 2021-01-07

基金项目: 国家自然科学基金项目(21563010); 河池学院校级重点实验室项目(2016-91)

岑忠用, E-mail: dqz8@163.com

通信作者: 严军, E-mail: yanjun03@163.com

表 1 6 种根茎类中药材样品信息表

Table 1 Sample information table of six rhizome Chinese medicinal materials

中药材名称 Name of Chinese medicinal materials	产地 Origin	采集方式 Collection method	样本数 Samples number
山豆根 Subprostrate Sophora	广西 Guangxi	野外采集 Field collection	30
百两金 Ardisia crispa	广西 Guangxi	野外采集 Field collection	30
北豆根 Rhizoma Menispermii	吉林 Jilin Province	市场购买 Market purchases	34
千斤拔 Philippine Flemingia	广西 Guangxi	野外采集 Field collection	20
滇豆根 Yunnan bean	云南 Yunnan Province	市场购买 Market purchases	13
云南豇豆 Yunnan cowpea	云南 Yunnan Province	野外采集 Field collection	12

1.3 药材前处理

所有药材均通过日晒的方式进行干燥以除去水分,干燥后的药材经中药粉碎机粉碎后过孔径为 0.25 mm 的筛子,再次干燥所得粉末样品,并经减重法证明样品干燥前后质量无显著差别,即说明样品水分含量极低,不影响近红外光谱分析,随后取适量样品装入透明密封袋中,压实,粉末样品厚度约 7 mm,袋装粉末样品直接用于近红外漫反射光谱测定。

1.4 数据分析

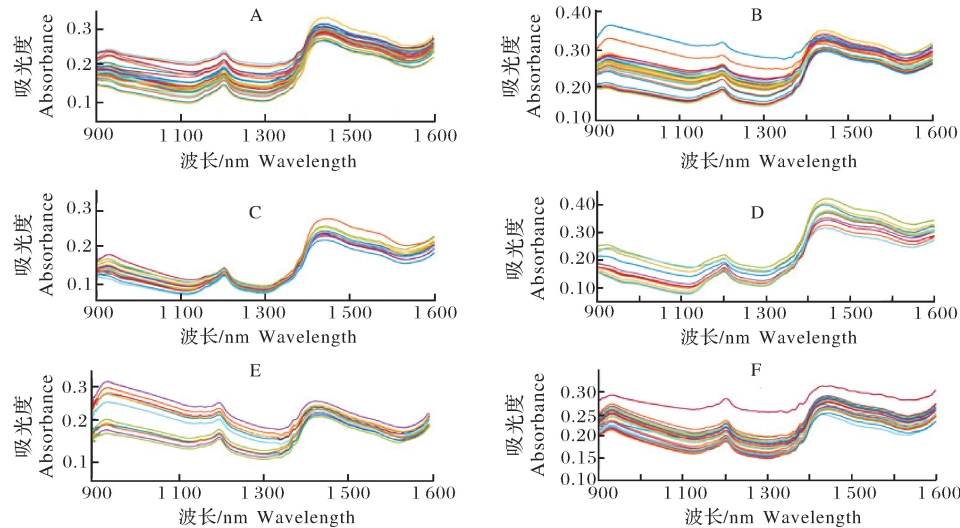
本研究先用无监督分析法(主成分分析、系统聚类分析)对药材样本进行分类,主成分分析和系统聚类分析基于不同的数学原理,且均可对样本分类实现可视化,便于呈现数据结构。然后用有监督分析法(*K* 近邻法、线性判别分析)构建分类模型,所选方法同样基于不同数学原理,以从不同角度证明近红外光谱用于中药材识别的可行性。在

有监督分析中,137 个药材样本通过 Kennard-Stone 算法分为训练集(91 个)和测试集(46 个)两部分,每种药材按 2 : 1 的质量比例均匀分布在训练集和测试集中,训练集用于构建分类模型,测试集用于评价模型的预测能力。本试验所用算法均通过软件 Matlab 2015a 编程,在计算机 Windows 7.0 系统下运行。

2 结果与分析

2.1 近红外漫反射光谱

6 种药材的近红外漫反射光谱(900~1 700 nm)如图 1 所示,其中 950 nm 为 O—H 二级倍频,1 200 nm 为 C—H 二级倍频,1 450 nm 为 O—H 一级倍频,1 360~1 390 nm 为—CH₃和—CH₂的合频,3 个主要的吸收谱带位于 900~950 nm、1 200~1 250 nm、1 400~1 500 nm。通过谱图比较可见,6 种药材的近红外漫反射光谱总体具有明显的相似性,尤其



A:百两金; B:山豆根; C:千斤拔; D:滇豆根; E:云南豇豆; F:北豆根。图 3 同。A: Ardisia crispa; B: Subprostrate Sophora; C: Philippine Flemingia; D: Yunnan bean; E: Yunnan cowpea; F: Rhizoma Menispermii. The same as Fig.3.

图 1 6 种药材的原始近红外漫反射

Fig.1 Original near-infrared diffuse spectra of six medicinal materials

在1200~1700 nm内。同时,6种药材的近红外光谱在不同波长下的吸收强度具有一定的差异,比如云南豇豆(YNJD)和北豆根(BDG)相对于其他4种药材在900~1200 nm波段具有更强的吸收。其中,个别样品表观颜色较深,导致其吸收度偏大,因此,所得近红外光谱与同类样品的近红外光谱在强度上呈现差异(例如北豆根(BDG)样品的最上一条光谱曲线)。由于近红外光谱反映的是化合物中分子振动的倍频和合频信息,信号弱、重叠度高、不具有特征峰,加之药材的化学组成复杂,因此,仅从近红外光谱图的表现图谱特征无法对不同的药材进行鉴别,必须借助于数学分析手段。

2.2 主成分分析

主成分分析是一种常用的无监督分析技术,可以通过将样本在第一主成分(PC1)、第二主成分(PC2)、第三主成分(PC3)上进行投影实现数据可视

化。本试验中,通过对由139个样本和511个波长点构成的近红外光谱数据矩阵 $X_{139 \times 511}$ 进行奇异值分解,主成分分析投影图如图2所示。方差分析结果表明,PC1能解释73.95%的信息量,PC2能解释25.04%的信息量,前2个主成分解释的累积信息量达到了98.99%。由图2可见,6种药材在PC1和PC2的投影分布总体上具有良好的分类聚集特征,滇豆根(DDG)、千斤拔(QJB)、云南豇豆(YNJD)3种药材均能和其他药材完全区分,而百两金(BLJ)、山豆根(SDG)、北豆根(BDG)3种药材的分布存在一定的重叠。在PC1、PC2、PC3上进一步做主成分投影分析,由于PC3只解释了0.49%的信息量,所以样本在前3个主成分上的分类效果相对于前2个主成分并没有显著提高。样本重叠的原因在于百两金、北豆根、山豆根3种药材的近红外光谱所携带的样本信息具有部分相似性。

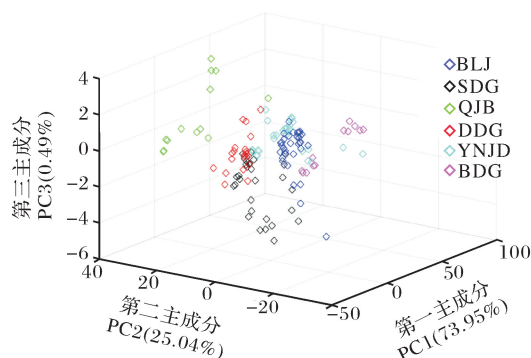
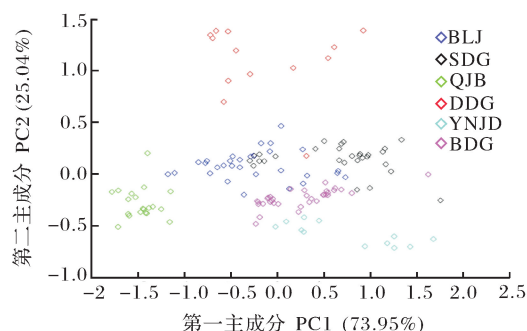


图2 6种药材的主成分分析投影图

Fig.2 Principal component analysis projection diagram of six medicinal materials

2.3 系统聚类分析

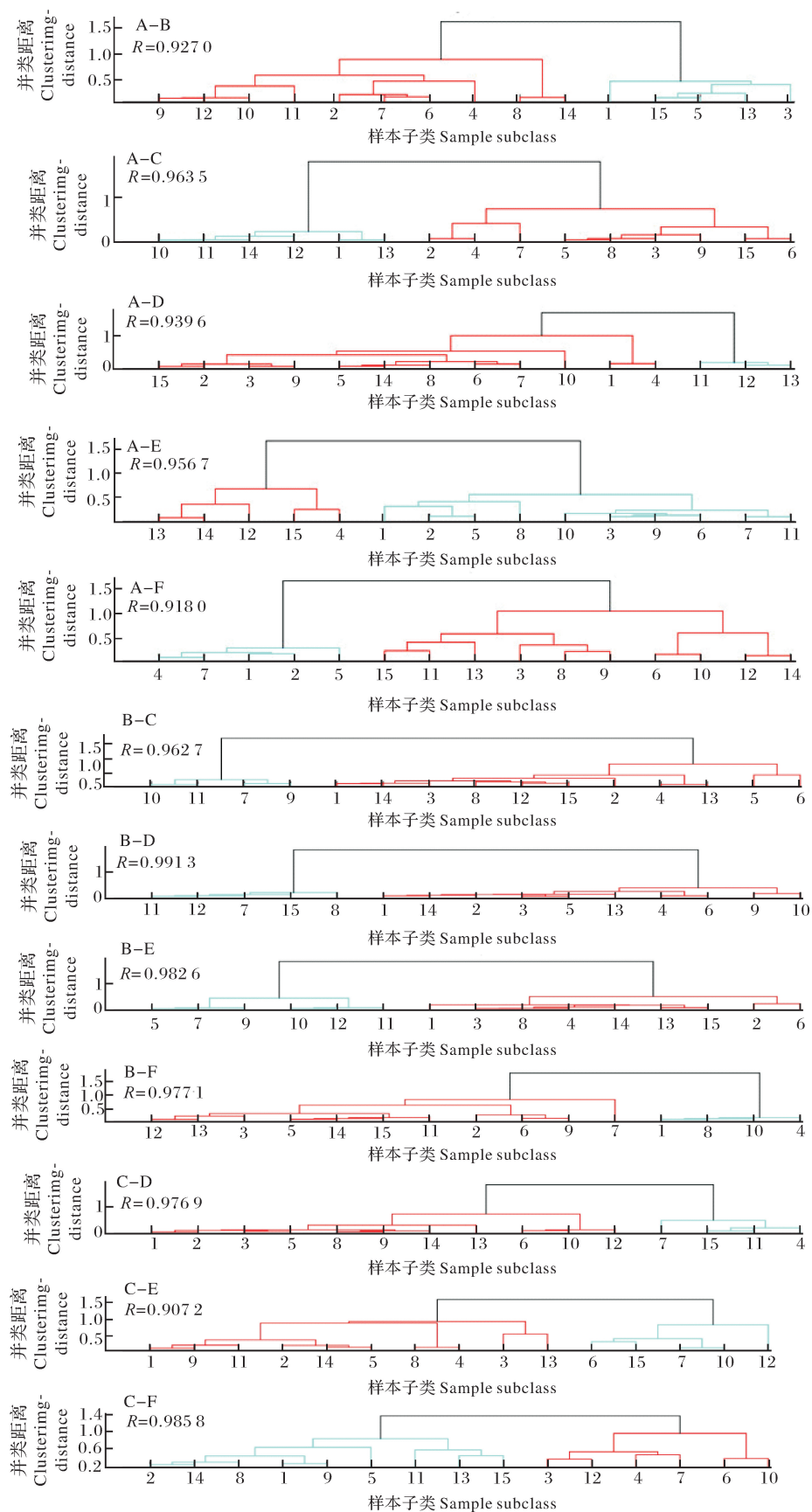
易混药材鉴定过程中面对最多的问题是2种相似药材之间的鉴别,因此,在本试验中,我们将6种药材进行两两配对,一共得到15种组合,并对每一个组合进行系统聚类分析,研究系统聚类分析在药材鉴别中的可行性。

本试验中分别考察了9种样本间距离(欧氏距离、标准化欧氏距离、绝对值距离、闵可夫斯基距离、夹角余弦、相关系数、斯皮尔曼距离、汉明距离、Jaccard系数)和7种类间距离(类平均法、重心法、最长距离法、中间距离法、最短距离法、离差平方和法、加权平均法)对系统聚类分析效果的影响,得到了最佳的参数组合:(1)“千斤拔-北豆根”组合采用夹角余

弦作为样本间距离,其余14种组合均采用斯皮尔曼距离;(2)所有组合的类间距离均通过类平均法定义。系统聚类分析谱系图(图3A、3B、3C)表明对于15种药材组合均能有良好的聚类效果,相关系数(R)在0.9063~0.9912,最大显示节点数为15。结果表明,通过近红外光谱结合系统聚类分析方法,可以对易混药材进行较为准确的分类识别。

2.4 K近邻法分析

K近邻法是一种常用的有监督模式识别方法,可根据K个样本的主要类别对分类样本进行归类。为进行模型验证,将原数据集分为训练集和测试集,具体信息见表2。



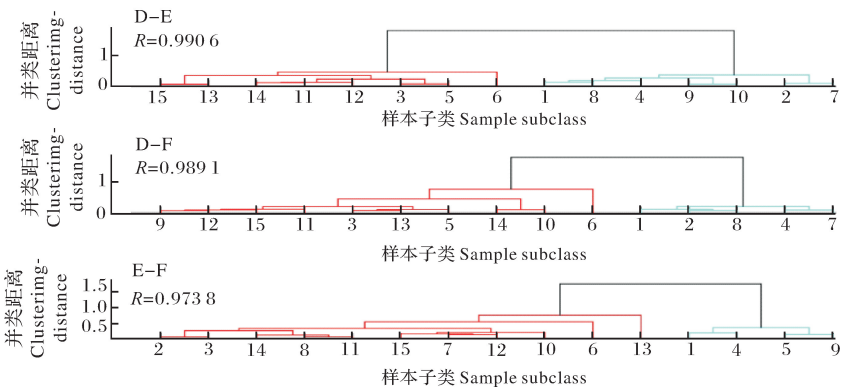


图 3 6 种易混药材的系统聚类谱系图

Fig.3 The hierarchical clustering pedigree of six miscible medicinal materials

表 2 K 近邻法的训练集及测试集个数
Table 2 The number of training set and test set
for k -nearest neighbor method

药材名称 Medicinal name	训练集($n=93$) Training set	测试集($n=46$) Test set
百两金 <i>Adisia crispa</i>	20	10
山豆根 <i>Subprostrate Sophora</i>	20	10
千斤拔 <i>Pilippine Flemingia</i>	13	7
滇豆根 <i>Yunnan bean</i>	8	5
云南豇豆 <i>Yunnan cowpea</i>	8	4
北豆根 <i>Rizoma Menisperm</i>	24	10

在本试验中,由于 PCA 能对 6 种中药材有较好的聚类效果,且前 5 个主成分几乎解释了 100% 的方差,因此,将主成分分析的前 5 个主成分作为 K 近邻法的输入数据,标准欧氏距离作为距离参数。此外, K 近邻法中的 K 值对分类效果有显著影响,选择合适的 K 值能有效地改善分类效果。为确定最佳的主成分个数(PCs)和 K 值,本试验采用留一法交互检验进行参数优化,如图 4 所示。交互检验

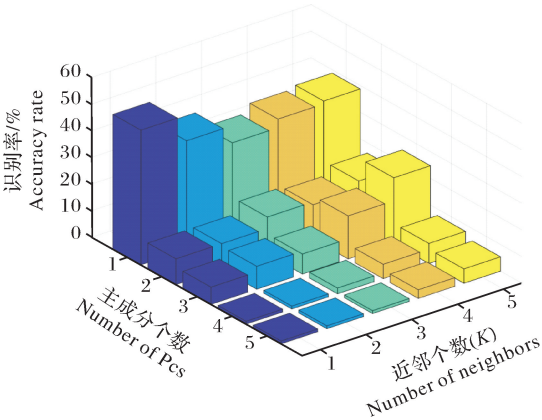


图 4 留一法交互检验参数(PCs 和 K)优化结果示意图

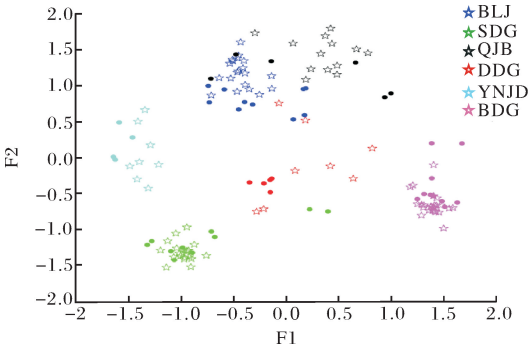
Fig.4 Schematic diagram of optimization results of leave-one-out interactive test parameters (PCs and K)

结果表明,当采用 5 个主成分和 $K=3$ 时分类效果最好,交互检验的分类准确率达到 98.92%。

为评价 K 近邻法对未知样本的分类效果,通过优化得到的 K 近邻法分类模型对 46 个未知样本进行分类。在 46 个样本中,只有 3 个样本被错误识别,其中 2 个山豆根被错判为北豆根,1 个千斤拔被错判为百两金,总体分类准确率达到 93.48%。

2.5 线性判别分析

本研究采用 Fisher 判别法,将近红外光谱数据的前 5 个主成分经标准化处理后作为输入数据,建立了线性判别分析模型。通过将训练集和测试集的样本投影到前 2 个判别函数(F_1 和 F_2),可以看到训练集样本具有完全的分类聚集分布特征,而测试集的绝大多数样本也能够落入训练集样本的投影区域内(图 5)。 F_1 和 F_2 解释的累积方差达到了总方差的 83.86%。通过对 46 个未知样本进行分类,结果表明线性判别分析具有很高的准确度,判别准确率达到 95.65%,除了 2 个山豆根被分别误判为云



☆训练集样本 Training set samples; ●测试集样本 Test set samples.

图 5 线性判别分析结果示意图

Fig.5 Schematic diagram of linear discriminant analysis results

南豇豆和北豆根之外,其余 44 个样本全部判别正确。

3 讨 论

中药材的准确鉴别是保证安全用药和中药复方开发的前提,也是中药材质量控制的重要组成部分。近红外光谱作为一种无损分析技术具有分析时间短、操作简单、分析成本低等优势,近年来在中药材鉴别工作中日益受到关注。梁华伦等^[12]利用近红外光谱建立了不同厂家小柴胡颗粒的快速鉴别方法,可有效鉴别出不同厂家的药品真伪;余梅等^[13]通过采集不同产地陈皮的内外侧近红外光谱,建立的分类模型准确率可达到 91.67%。但是,关于山豆根、千斤拔、百两金、北豆根等性状相似的根茎类药材,目前大多采用性状鉴别、显微鉴别等技术,近红外光谱在根茎类中药材的鉴别中尚未见报道。

系统聚类的基本思想是根据不同样本之间的“距离”进行分类,基于“相近者相似”的原理形成一个亲疏关系图谱。影响系统聚类分析效果的主要参数包括样本之间的“距离”定义和聚类过程中类与类之间的“距离”。本研究针对 6 种易混淆的根茎类中药材(百两金、千斤拔、山豆根、滇豆根、云南豇豆、北豆根)建立了一种基于近红外光谱和化学计量学的药材快速鉴别方法。研究结果显示,6 种中药材在主成分分析和系统聚类中均表现出明显的分类聚集趋势,说明 6 种中药材的近红外光谱信息具有足够的差异性。进一步通过 K 近邻法和线性判别分析建立的分类判别模型,具有较高的准确率,并通过外部测试集对模型的预测能力进行了评估。结果表明,将所建立的模型对包含 6 类药材的 46 种未知样本进行分类,准确率在 93.48%~95.65%,说明模型具有良好的预测能力。与目前已报道的方法相比,本试验方法可以有效缩短药材鉴别时间,降低成本,同时保证鉴别结果的可靠性。

然而,任何模型的应用范围都受到样本空间的限制,尽管本试验所建立的模型在交互检验和外部测试中均表现出良好的准确度和稳健性,但是要作为一种实用技术进行推广仍有许多工作需要完善。一方面,需要进一步扩大样本空间,收集不同产地、不同采集时间的样本,提高模型的普适性;另一方面,研究不同实验室或仪器之间的模型转移误差是否在可接受范围之内。

参考文献 References

- [1] 陈科力,黄林芳,刘义梅.中药鉴定方法学发展历程[J].中国中药杂志,2014,39(7):1203-1208.CHEN K L, HUANG L F, LIU Y M. Development history of methodology of Chinese medicines' authentication[J]. China journal of Chinese materia medica, 2014, 39(7):1203-1208(in Chinese with English abstract).
- [2] 宋希贵,李青.北豆根及其混淆品鉴别[J].时珍国医国药,2000,11(1):45-46.SONG X G, LI Q. Identification of Beidougen and its confusions[J]. Lishizhen medicine and materia medica research, 2000, 11(1):45-46 (in Chinese).
- [3] 鞠康,胡云飞,王德胜,等.微性状鉴别法在中药检验中应用研究进展[J].中成药,2019,41(2):411-415.JU K, HU Y F, WANG D S, et al. Research progress in the application of micro-character identification methods in traditional Chinese medicine testing[J]. Chinese traditional patent medicine, 2019, 41(2):411-415(in Chinese).
- [4] 陈文婷,童家赞,钟慧怡,等.朱砂根等四种易混淆中药的性状和显微比较鉴别[J].中药材,2020,43(1):52-54.CHEN W T, TONG J Y, ZHONG H Y, et al. The characteristics and microscopic comparison and identification of four confusing Chinese medicinal materials including cinnabar root[J]. Journal of Chinese medicinal materials, 2020, 43(1):52-54 (in Chinese).
- [5] 马欣,德白,潘一滨,等.复黄片全方中药鉴别方法的研究[J].中草药,2018,49(12):2878-2882. MA X, DE B, PAN Y B, et al. Identification of all herb medicines in Fuhuang tablet[J]. Chinese traditional and herbal drugs, 2018, 49(12):2878-2882 (in Chinese with English abstract).
- [6] 方强强,王燕,彭春,等.中药 DNA 条形码分子鉴定技术的应用与展望[J].中国实验方剂学杂志,2018,24(22):197-205. FANG Q Q, WANG Y, PENG C, et al. Application and prospect of DNA barcode molecular identification technology in traditional Chinese medicine[J]. Chinese journal of experimental traditional medical formulae, 2018, 24(22):197-205(in Chinese with English abstract).
- [7] RUKUNDO I R, DANA O M G C, WELLER C L, et al. Use of a handheld near infrared spectrometer and partial least squares regression to quantify metanil yellow adulteration in turmeric powder[J]. Journal of near infrared spectroscopy, 2020, 28(2): 81-92.
- [8] WANG C Y, XIANG B R, ZHANG W. Application of two-dimensional near-infrared (2D-NIR) correlation spectroscopy to the discrimination of three species of *Dendrobium*[J]. Journal of chemometrics, 2009, 23(9):463-470.
- [9] DONG F, LIN J T, YOU J H, et al. A chemometric modeling-free near infrared barcode strategy for smart authentication and geographical origin discrimination of Chinese ginseng[J/OL]. Spectrochimica acta part A: molecular and biomolecular spectroscopy, 2020, 226:117555[2021-01-07]. <https://doi.org/10.1016/j.saa.2019.117555>.
- [10] 华应顺.山豆根与其常见 7 种伪品的鉴别[J].西部中医药,

- 2015,28(4):21-24. HUA Y S. Identificaiton of Shandougen and its seven common conterfeits[J]. Western journal of traditional Chinese medicine, 2015, 28(4): 21-24 (in Chinese with English abstract).
- [11] 张云方,山豆根及其常见混淆品的鉴别[J].海峡药学,2011,23(12):50. ZHANG Y F. Identification of Shandougen and its common confounding products[J]. Strait pharmaceutical journal, 2011, 23(12):50 (in Chinese).
- [12] 梁华伦,谭昌成,江秀娟,等. 近红外光谱快速鉴别不同厂家小柴胡颗粒研究[J]. 中医药导报,2021,27(1):62-64. LIANG H L, TAN C C, JIANG X J, et al. Study on rapid identification of Xiaochaihu granules from different manufacturers by near infrared spectroscopy[J]. Journal of traditional Chinese medicine and pharmacology, 2021, 27(1): 62-64 (in Chinese with English abstract).
- [13] 余梅,李尚科,杨菲,等. 基于近红外光谱技术与优化光谱预处理的陈皮产地鉴别研究[J]. 分析测试学报,2021,40(1):65-71. YU M, LI S K, YANG F, et al. Identification on different origins of citri reticulatae pericarpium using near infrared spectroscopy combined with optimized spectral pretreatments[J]. Journal of instrumental analysis, 2021, 40(1): 65-71 (in Chinese with English abstract).

Identification of 6 kinds of rhizome Chinese medicine by near infrared spectroscopy

CEN Zhongyong¹, LEI Shunxin², LEI Lei², YAN Jun³, ZHANG Huiying¹

1. School of Chemistry and Biological Engineering, Hechi University, Yizhou 546300, China;

2. Hengxian Comprehensive Inspection and Testing Center, Hengxian 530300, China;

3. School of Chemistry and Chemical Engineering, Guangxi University for Nationalities/

Key Laboratory of Guangxi Colleges and University for Food Safety and

Pharmaceutical Chemistry, Nanning 530006, China;

Abstract Rhizome Chinese medicine refers to rhizomes or underground stem medicinal materials with a small amount of roots or fleshy scales. Due to the high similarity of plant morphology of rhizome medicinal materials, it has caused the problem of confusion and difficulty in distinguishing the authenticity of Chinese medicinal materials in the market, and further cause serious clinical medication risk. Traditional Chinese medicine identification methods are cumbersome to operate, long analysis time, high cost, and relatively low efficiency. In order to establish the identification method of six kinds of easily confused rhizome Chinese medicinal materials including Subprostrate Sophora, Ardisia crispa, Rhizoma Menisperm, Philippine Flemingia, Yunnan bean, and Yunnan cowpea, etc, which has rich chemical composition and high medicinal value. In this research we used near-infrared diffuse reflectance (NIR) spectroscopy to collect spectral information within 900-1 700 nm, and then combined with principal component analysis (PCA), systematic cluster analysis (SCA), k -nearest neighbor (k -NN), and linear discriminant analysis (LDA) to develop qualitative discriminant models. Results showed that the six kinds of medicinal materials showed obvious classification and aggregation characteristics in principal component analysis and systematic cluster analysis, the classification accuracy rate for 46 unknown samples up to 93.48% and 95.65% for k -NN and LDA, respectively. From the results of the experiment, we come to the following conclusion that NIR coupled with pattern recognition techniques, which is rapid, accurate, and low-cost, has potential to discriminate Rhizome Chinese medicine.

Keywords near infrared diffuse reflectance spectroscopy; identification of traditional Chinese medicinal materials; Subprostrate Sophora; systematical cluster; k -NN; LDA; pattern recognition

(责任编辑:赵琳琳)