

# 基于多参考相关系数法和 BP-ANN 建立紫石英的近红外光谱定性模型

徐子杰<sup>1</sup>, 陈龙<sup>1,2</sup>, 刘义梅<sup>1</sup>, 陈科力<sup>1\*</sup>

(1. 湖北中医药大学 教育部中药资源和中药复方重点实验室, 武汉 430065;  
2. 南漳县人民医院, 湖北 襄阳 441500)

**[摘要]** **目的:** 基于多参考相关系数法和误差反向传递人工神经网络 (BP-ANN) 建立矿物药紫石英的近红外光谱定性模型, 用于紫石英的生品、煅制品、醋煅品和伪品的快速鉴别。**方法:** 采集紫石英、紫石英煅制品、紫石英醋煅品及紫石英伪品这 4 类不同紫石英样品的近红外光谱, 对光谱进行二阶导数和 9 点平滑预处理, 计算多项相关系数。在 Matlab 2014a 软件中将多项相关系数作为输入数据, 以 BP-ANN 建立 4 类不同紫石英样品的快速鉴别模型。**结果:** 建立了紫石英近红外光谱 BP-ANN 鉴别模型, 模型验证结果显示, 15 批验证样品中 14 批样品预测结果正确, 仅 1 批样品预测有误, 准确率达 93.33%。**结论:** 建立的紫石英近红外光谱 BP-ANN 鉴别模型能通过一次性整合的运算区分紫石英生品、煅制品、醋煅品及其伪品, 鉴别结果准确可靠。此外, 模型在近红外光谱相关系数法基础上, 以多个参考光谱为对照计算所得的多组相关系数作为网络特征输入数据, 实现了光谱数据的压缩。

**[关键词]** 紫石英; 饮片鉴定; 多参考相关系数法, 近红外光谱; 误差反向传递人工神经网络

**[中图分类号]** R282.5; R931.5; R284 **[文献标识码]** A **[文章编号]** 1005-9903(2017)22-0037-06

**[doi]** 10.13422/j.cnki.syfjx.2017220037

**[网络出版地址]** <http://kns.cnki.net/kcms/detail/11.3495.R.20170906.1114.006.html>

**[网络出版时间]** 2017-09-06 11:14

## Establishment of Near Infrared Spectral Qualitative Model of Fluoritum Based on Multi-reference Correlation Coefficient Method and BP-ANN

XU Zi-jie<sup>1</sup>, CHEN Long<sup>1,2</sup>, LIU Yi-mei<sup>1</sup>, CHEN Ke-li<sup>1\*</sup>

(1. Key Laboratory of Chinese Medicine Resource and Compound Prescription, Ministry of Education, Hubei University of Chinese Medicine, Wuhan 430065, China;  
2. Nanzhang People's Hospital, Xiangyang 441500, China)

**[Abstract]** **Objective:** To establish a near infrared spectral qualitative model of Fluoritum for rapidly identifying raw, calcined, vinegar calcined products and its counterfeits based on multi-reference correlation coefficient method (MRCC) and back-propagation artificial neural networks (BP-ANN). **Method:** The near infrared spectra of 4 kinds of Fluoritum samples were collected, including raw products, calcined products, vinegar calcined products and counterfeit products. The second derivative and 9 points smoothing were used to calculate multiple correlation coefficients. Multiple correlation coefficients were used as input data in Matlab 2014a software, and a rapid identification model for 4 different types of Fluoritum samples was established by BP-ANN. **Result:** BP-ANN identification model of near infrared spectroscopy of Fluoritum was established. The prediction results of 14 batches of samples were correct, only 1 batch of samples was predicted error, the accuracy rate was

**[收稿日期]** 20170428(011)

**[基金项目]** 武汉市 2012 年高新技术产业发展行动计划生物技术与新医药专项(20126053193)

**[第一作者]** 徐子杰, 在读硕士, 从事中药资源及其品质研究, Tel:15827342040, E-mail:405320756@qq.com

**[通讯作者]** \* 陈科力, 教授, 博士生导师, 从事中药鉴定学研究, Tel:027-68890106, E-mail:kelichen@126.com

93.33%。 **Conclusion:** The BP-ANN identification model of near infrared spectroscopy of Fluoritum can be used to distinguish among 4 different types of samples by a one-time integrated operation, and the result is accurate and reliable. In addition, based on the correlation coefficient method of near infrared spectroscopy, the multi-group correlation coefficient calculated by the multiple reference spectra is used as the input data of the network characteristics, the compression of spectral data is realized.

**[Key words]** Fluoritum; pieces identification; multi-reference correlation coefficient method; near infrared spectroscopy; back-propagation artificial neural networks

紫石英为氟化物类矿物萤石族萤石,具有温肾暖宫、镇静安神的作用<sup>[1]</sup>。其在临床中主要使用的炮制品为醋煅紫石英<sup>[2]</sup>。在中药材市场调查时发现,紫石英混淆品及炮制不合格品较多,亦见仅煅烧而未醋淬的煅紫石英药材作为炮制品销售<sup>[3]</sup>。为了鉴别紫石英及其炮制品真伪,急需一种明确的且简单快速的客观鉴别方法。近红外光谱技术(near infrared, NIR)具有快速无损、样品用量少、成本低、操作简单、准确性高等优点<sup>[4-5]</sup>。矿物药虽响应近红外光谱的基团不明显,但因其结构及晶形变化多,且物相结构稳定,已成功应用于许多矿物药的分析<sup>[6-7]</sup>。

相关系数法建模是较为简便实用的近红外光谱建模较常用的分析计量学方法,已成功应用于一些矿物药的定性鉴别中<sup>[8]</sup>。目前,应用单一的相关系数对某一些复杂矿物药样品进行鉴别的效果并不理想,而分步采用多个相关系数逐一建立模型(多项相关系数法模型)则可以解决上述问题,但其缺点在于需要建立多个相关系数模型,同时需要分步进行多次运算,应用较繁琐复杂。本课题组成功建立了多参考相关系数法(multi-reference correlation coefficient method, MRCC),其可以简化运算过程<sup>[5]</sup>。人工神经网络(artificial neural networks, ANN)是一类模拟人脑功能的全新的数据和知识等信息加工处理系统<sup>[9]</sup>。其中多层误差反向传递人工神经网络(back-propagation artificial neural networks, BP-ANN)应用尤广,具有很强的非线性建模能力,适合解决复杂的映射问题<sup>[10]</sup>。

本实验将基于近红外光谱获得的多项相关系数作为 BP-ANN 的输入数据,建立紫石英 BP-ANN 定性鉴别模型,可通过一次整合的运算同时区分 4 类紫石英样品,包括生品、煅制品、醋煅品及其伪品。

## 1 材料

MPA 型傅里叶变换近红外光谱仪(德国布鲁克光学仪器公司,配备固体积分球漫反射附件,OPUS 7.5 光谱采集和处理软件),Matlab 2014a 数据分析

软件(美国 MathWorks 公司)。

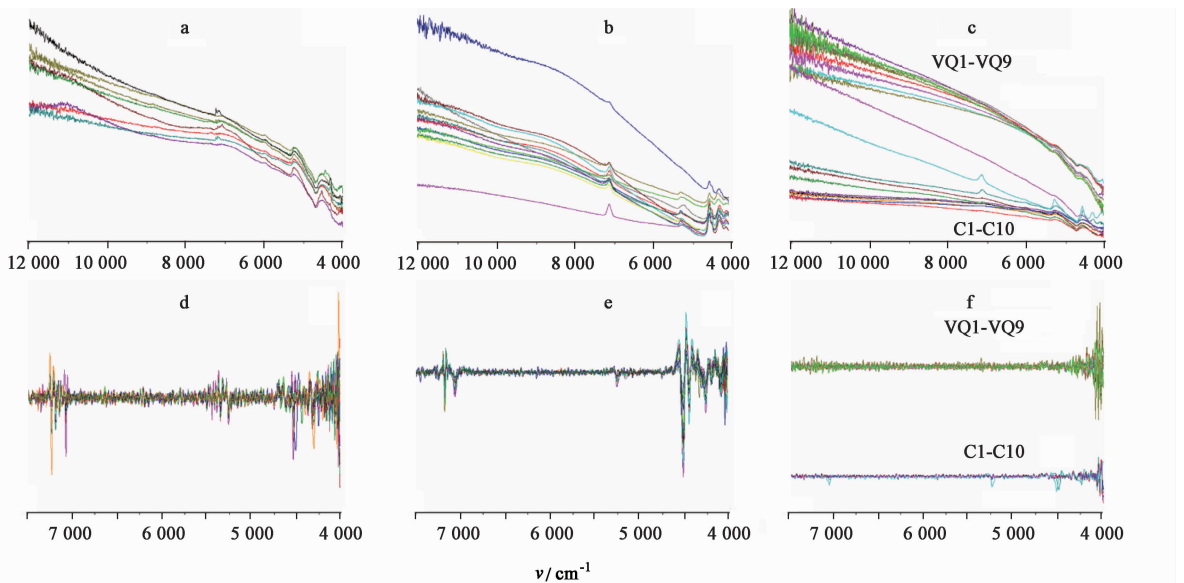
25 批紫石英样品分别购于安国、亳州、禹州等中药材专业市场,经湖北中医药大学药教研室陈科力教授鉴定,样品均利用粉末 X 射线衍射法确定其矿物药的物相组成和基原<sup>[11]</sup>,其中紫石英生品 13 批(F1 ~ F13),伪品 11 批(S1 ~ S11),煅制品(C1)1 批;随机取 9 批正品和 1 批伪品,共 10 批,每批取 2 份,每份 10 g,分别以明煅法和醋煅法<sup>[12]</sup>对上述 10 批样品进行炮制,得 9 批紫石英煅制品(C2 ~ C10),9 批醋煅品(VQ1 ~ VQ9)和 2 批煅制伪品(S12 ~ S13),2 批煅制伪品均作伪品处理。购买及自制样品共计 45 批,粉碎,过 60 目筛,并在玛瑙研钵中混合均匀,用于实验分析。

## 2 方法

**2.1 光谱采集** 所有样品分别取 5 g 置于样品杯中,采用积分球漫反射测试模式扫描近红外光谱。光谱扫描范围 4 000 ~ 12 000  $\text{cm}^{-1}$ ,扫描数 64 次,仪器分辨率 8  $\text{cm}^{-1}$ 。每个样品重复扫描 3 次,取平均光谱作为该样品的分析光谱。

**2.2 光谱预处理** 各类紫石英样品在近红外区域的主要特征峰在 7 500 ~ 4 000  $\text{cm}^{-1}$ ,结果见图 1。对比 4 类紫石英样品在该区域的近红外光谱图谱,结果发现紫石英生品与炮制品光谱较为一致,而生品与伪品图谱差异较大。由于生品和 2 种炮制品在水分含量上的不同,因此,在 7 000  $\text{cm}^{-1}$  附近和 5 000 ~ 5 400  $\text{cm}^{-1}$  等水分的近红外特征吸收谱带表现一定差异。各类样品的近红外光谱图谱都发生明显的基线漂移,且谱带重叠严重。因此,采用二阶导数(9 点平滑)对近红外光谱图谱进行预处理。预处理后的 7 500 ~ 4 000  $\text{cm}^{-1}$  谱段用于紫石英近红外光谱分析,见图 1。

**2.3 近红外光谱相关系数的计算** 近红外光谱相关系数是评价样品光谱与参考光谱相似程度的指标,公式为  $c = \text{Cov}(y_k, y_r) / (\sigma_k \cdot \sigma_r)$ ,式中  $\text{Cov}(y_k, y_r)$  为样品光谱  $y_k$  与参考光谱  $y_r$  的协方差,  $\sigma_k$  和  $\sigma_r$  分别为样品光谱  $y_k$  和参考光谱  $y_r$  的标准偏差<sup>[13]</sup>。



a, d. 生品; b, e. 伪品; c, f. 煨制品和醋煨品

图 1 紫石英样品近红外原始光谱 (a, b, c) 及二阶导数光谱 (d, e, f)

Fig. 1 NIR original spectra (a, b, c) and second derivative spectra (d, e, f) of Fluoritum samples

以 2:1 的比例将 4 类共 45 批样品随机分为训练集样品 (30 批) 和验证集样品 (15 批), 将训练集样品中紫石英生品近红外光谱的平均光谱作为紫石英的标准参考图谱; 同样获得紫石英伪品、煨制品、醋煨品的标准参考图谱; 各参考图谱分别记为参考光谱 1, 2, 3, 4。在 OPUS 7.5 软件中, 依次计算各样品与 4 个标准参考光谱经二阶导数 (9 点平滑) 处理后在  $7\ 500 \sim 4\ 000\ \text{cm}^{-1}$  谱段的相关系数, 分别记为  $c_1, c_2, c_3, c_4$ 。

**2.4 BP-ANN 模型建立** 利用 Matlab 2014a 软件的人工神经网络工具箱, 建立 3 层的 BP-ANN 网络, 即输入层、隐含层和输出层。用各样品与 4 个参考光谱的相关系数  $c_1, c_2, c_3, c_4$  作神经网络的输入, 输入层节点数为 4。为了更直观地区分 4 类样品, 故将伪品、生品、煨品和醋煨品分别设定为平面直角坐标系中的点 ( $X, Y$ ) ( $X$  为样品点的横坐标值,  $Y$  为样品点的纵坐标值), 以点 (0, 0) 代表伪品, 点 (1, 0) 代表生品, 点 (0, 1) 代表煨品, 点 (1, 1) 代表醋煨品, 并分别将各样品点在  $X$  轴和  $Y$  轴的坐标值作为网络训练的目标数据, 输出层节点数为 2。输入层到隐含层传递函数采用对数 S 型传递函数 (logsig), 隐含层到输出层采用线性传递函数 (purelin), 学习函数为 BP 学习规则 (learnbd), 训练函数为量化共轭梯度法 (traincg), 性能分析函数为均方差性能分析函数 (MSE)。网络建立后, 对网络权值进行初始化。将训练集中的 30 批样品随机排列, 所导入的训练集输入数据须经标准化处理, 并对网络参数进行优化,

确定隐含层节点数为 2, 神经网络学习速率 0.1, 动量因子为 0.9, 学习数 500 次。

确定网络结构和参数后, 对网络进行训练, 采用内部交叉验证和外部验证对神经网络模型的预测能力进行验证。内部交叉验证以内部交叉验证均方差 (RMSECV), 训练集样品预测值与目标值的决定系数 ( $R^2$ ) 为指标, 外部验证以训练集样品的预测均方差 (RMSEP) 和  $R^2$  为指标; RMSEP 和 RMSECV 越小,  $R^2$  越大, 模型预测效果越好。分别考察此训练后的网络对训练集和验证集样品进行预测所得的样品点  $X$  轴和  $Y$  轴坐标预测值的 RMSECV, RMSEP,  $R^2$ , 在网络达到较好效果时, 停止训练, 保存训练后的网络, 作为 BP-ANN 模型。模型建立后, 利用验证集样品对模型预测能力进行评价。

### 3 结果

**3.1 多参考的相关系数** 训练集中 4 类共计 30 批样品的近红外光谱与 4 个参考光谱的相关系数范围见表 1。在 NIR 定性分析建模过程中, 若建模所用训练集光谱与一个标准参考光谱进行比较时, 计算各类别样品光谱的相关系数不重叠, 可使各类别样品互相区分开, 则可取训练集中两相邻类别的样品光谱相关系数的临界值的平均值为阈值, 建立定性模型。

由表 1 可知, 所计算的 4 组相关系数  $c_1, c_2, c_3, c_4$  均不能达到单独使用就可将 4 类紫石英样品互相区分开。利用各样品  $c_1$  仅能将生品与其他样品区

表 1 紫石英样品的 NIR 相关系数

Table 1 NIR correlation coefficients of Fluoritum samples

样品类别	训练集样品数	参考光谱	$c_1$	$c_2$	$c_3$	$c_4$
生品	9	1	0.438 9 ~ 0.642 2	0 ~ 0.623 0	0 ~ 0.685 8	0 ~ 0.562 5
伪品	9	2	0.174 0 ~ 0.381 7	0.756 0 ~ 0.966 2	0 ~ 0.406 8	0 ~ 0.322 5
煨制品	6	3	0.288 1 ~ 0.428 6	0.037 8 ~ 0.151 6	0.942 2 ~ 0.982 3	0
醋煨品	6	4	0.051 8 ~ 0.342 4	0.000 9 ~ 0.058 2	0 ~ 0.291 1	0.305 8 ~ 0.672 1

分开,但不能将伪品、煨制品和醋煨品互相区分开;同样的,利用  $c_2$  和  $c_3$  分别能将伪品、煨制品区分出来;而 4 类样品  $c_4$  范围互相叠加,不能区分出任何一类样品。说明基于单一参考光谱的相关系数模型不可行。

依据上述分析,虽然基于单一参考光谱的各样品的相关系数难以将 4 类不同的紫石英样品区分开,但若应用由多个参考光谱组合而成的多项相关系数模型组,则可实现 4 类样品的互相区分。即首先在作为指标的相关系数模型中鉴别出紫石英生品,然后将光谱数据应用于  $c_2$  为指标的相关系数模型中鉴别出伪品,再用  $c_3$  为指标的相关系数模型鉴别出煨制品,最后未被鉴别的样品即为醋煨品。

实际上,该应用多项相关系数的模型组方法在鉴别应用上较为繁琐、复杂。样品以 4 个参考光谱为对照,所得的 4 个相关系数,反映了样品的近红外光谱分别与生品、伪品、煨制品和醋煨品的相似度。这 4 个相似度即为样品光谱的 4 个特征数,联合利用则可用于区分所有的 4 类紫石英样品。基于此,尝试将 4 组训练集样品所得相关系数作为紫石英近红外光谱定性模型的光谱特征输入值,并提出对建模的近红外光谱多参考的相关系数数据进行数据压缩的一种新方法。

**3.2 BP-ANN 模型** 在应用 BP-ANN 进行 NIR 建模运算时,NIR 往往具有大量的数据点,若直接将这些数据点输入一个网络,则网络规模会非常庞大,这些数据点也可能存在严重的共线性,不仅使网络运算速度减慢,更易陷入局部最优。因此有必要对光谱数据进行压缩,通常采用主成分分析或偏最小二乘法等降维方法,以得到的主因子得分作为神经网络的输入变量,或采用遗传算法等谱段优选方法,以优化所得特征谱段作为神经网络的输入变量。在 NIR 相关系数法基础上,将以多个参考光谱为对照计算所得的多组相关系数为网络特征输入数据,实践光谱数据的压缩。以训练集和验证集各样品光谱

的 4 项相关系数  $c_1, c_2, c_3, c_4$  为 BP-ANN 模型的输入数据,以内部交叉验证和外部验证进行评价,最终所建紫石英 NIR 光谱 BP-ANN 模型的效果见表 2。

表 2 紫石英近红外光谱 BP-ANN 模型的预测效果

Table 2 Predictive effect of BP-ANN model of near infrared spectroscopy of Fluoritum

验证方法	模型参数	X 轴坐标值	Y 轴坐标值
内部交叉验证	RMSECV	0.010 2	0.001 1
	$R^2$	0.999 6	0.100 0
外部验证	RMSEP	0.263 4	0.078 4
	$R^2$	0.758 2	0.979 3

根据训练集样品的预测结果,在坐标系中绘制各样品点,见图 2。利用图 2 判断 BP-ANN 模型所预测的各样品类别,其中  $X < 0.5$  且  $Y < 0.5$  为伪品; $X \geq 0.5$  且  $Y < 0.5$  为生品; $X < 0.5$  且  $Y \geq 0.5$  为煨制品; $X \geq 0.5$  且  $Y \geq 0.5$  为醋煨品。

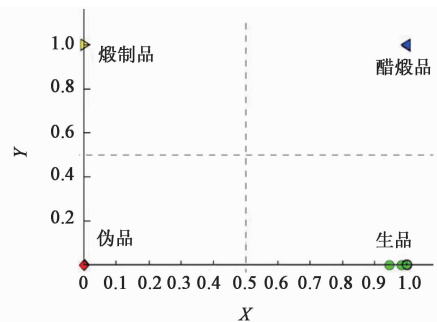


图 2 紫石英近红外光谱的训练集样品预测

Fig. 2 Prediction of training sets samples of near infrared spectroscopy of Fluoritum

**3.3 BP-ANN 模型的评价** 利用 BP-ANN 建立的鉴别模型对 15 个验证集样品进行预测,见表 3。将预测所得 X 和 Y 轴坐标值在坐标系中表示,见图 3。结果发现所建 BP-ANN 模型预测准确率较高,15 个验证集样品光谱经 BP-ANN 模拟计算,最终所得结果中,有 14 个样品预测结果正确,仅 1 批样品预测有误,准确率达 93.33%,说明该模型可用于紫石英

的快速鉴别,鉴别结果较准确。此外,对预测有误的 C1 样品进行考察,该样品为市场所购的煅制品,为进一步确定 C1 误判的原因,分别取醋煅品(VQ3),煅品(C4)和 C1 粉末 1 g,各加水 5 mL,振摇,滤过,依次测定滤液 pH。结果发现醋煅品 pH 6.628,

煅制品 pH 7.637, C1 样品 pH 7.175。C1 样品 pH 在煅制品和醋煅品之间,具有一定的酸性,其原本可能为醋煅品,但由于存储时间过久或保存不当,其颜色较浅,醋煅后的气味减弱,被作为煅制品销售和使用,导致模型判断失误。

表 3 BP-ANN 模型预测 15 个紫石英样品

Table 3 Prediction effect of 15 batches of Fluoritum samples based on BP-ANN model

编号	X			Y			预测结果
	真实值	预测值	偏差	真实值	预测值	偏差	
C1	0	0.999 2	0.999 2	1	0.989 0	-0.011 0	醋煅品(错误)
C4	0	0.003 6	0.003 6	1	1.000 5	0.000 5	煅品(正确)
C6	0	-0.001 1	-0.001 1	1	0.996 0	-0.004 0	煅品(正确)
C9	0	0.000 1	0.000 1	1	1.000 5	0.000 5	煅品(正确)
F2	1	1.000 1	0.000 1	0	-0.000 6	-0.000 6	生品(正确)
F5	1	1.000 5	0.000 5	0	0.263 9	0.263 9	生品(正确)
F8	1	1.000 6	0.000 6	0	0.149 7	0.149 7	生品(正确)
F11	1	1.000 7	0.000 7	0	-0.000 6	-0.000 6	生品(正确)
S2	0	-0.000 5	-0.000 5	0	0.001 0	0.001 0	伪品(正确)
S5	0	0.205 6	0.205 6	0	-0.000 5	-0.000 5	伪品(正确)
S9	0	-0.000 5	-0.000 5	0	-0.000 5	-0.000 5	伪品(正确)
S12	0	-0.000 5	-0.000 5	0	0.000 4	0.000 4	伪品(正确)
VQ3	1	1.000 1	0.000 1	1	1.000 4	0.000 4	醋煅品(正确)
VQ5	1	1.000 1	0.000 1	1	1.000 4	0.000 4	醋煅品(正确)
VQ9	1	1.000 1	0.000 1	1	1.000 4	0.000 4	醋煅品(正确)

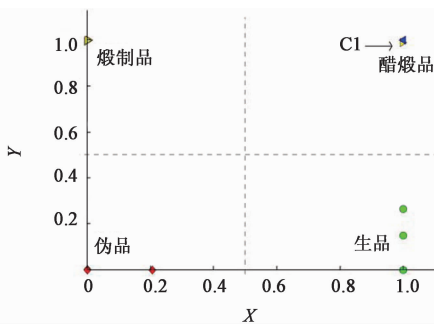


图 3 紫石英近红外光谱 BP-ANN 模型对训练集样品的预测  
Fig. 3 Prediction of training sets samples of near infrared spectroscopy of Fluoritum based on BP-ANN model

#### 4 讨论

本研究结果表明利用各样品分别与 4 个参考光谱的相关系数  $c_1, c_2, c_3, c_4$  为特征输入数据所建立的近红外光谱 BP-ANN 模型,可用于同时鉴别生品、伪品、煅制品和醋煅品,结果准确可靠。BP-ANN 能显著提高常规线性算法(主成分回归、偏最小二乘回归等)所建模型的性能,适用于矿物类中药的光

谱鉴别,并且此模型具有较强的预测能力,可用于其他矿物类中药的近红外光谱分析。

本研究提出了采用多参考相关系数法对定性建模的近红外光谱数据进行压缩,并证明了该方法具有可行性。这种多参数的相关系数法,相对于单一参数光谱的相关系数法,充分考察了各样品与所有样品类别之间的关系,对样本特性的表达更准确、更可靠、更全面。综合上述,多参数的相关系数法建立了 BP-ANN 智能分类模型,这也是智能算法引入矿物类中药分析中的一次创新,为解决近红外光谱法分析矿物类中药提供了新思路。

矿物药临床用量较小,使用频率较低,因而在中药药理学研究中一直缺乏重视,为现代中药药理学研究中的薄弱环节。但矿物类中药药性独特、临床效果好,是现代中医药学不可缺少的一部分。本文提出的一次性鉴别紫石英炮制品的方法,能够达到准确快速鉴别的目的。运用近红外光谱鉴别矿物药,能够进一步指导矿物药的真伪分辨。

[参考文献]

- [1] 国家药典委员会. 中华人民共和国药典. 一部[M]. 北京: 中国医药科技出版社, 2010: 316-317.
- [2] 朱传静, 常琳, 康琛, 等. 紫石英研究概况[J]. 中国实验方剂学杂志, 2011, 17(14): 306-311.
- [3] 陈龙, 袁明洋, 陈科力. 常见矿物药近红外漫反射光谱特征归纳与分析[J]. 中国中药杂志, 2016, 41(19): 3528-3536.
- [4] 李云, 毕宇安, 王振中, 等. 近红外光谱技术在热毒宁注射液栀子提取液浓缩过程中的应用[J]. 中国实验方剂学杂志, 2016, 22(12): 1-6.
- [5] SUN Y B, CHEN L, HUANG B S, et al. A rapid identification method for calamine using Near-Infrared spectroscopy based on multi-reference correlation coefficient method and back propagation artificial neural network[J]. Appl Spectrosc, 2016, 71(7): 1447-1456.
- [6] 李军华, 吴炜, 何艳, 等. 近红外光谱对天然岩石中矿物成分含量测定的研究[J]. 光谱学与光谱分析, 2013, 33(1): 85-88.
- [7] 闫蔚, 曾柏淋, 孟江, 等. 石膏红外图谱鉴定研究[J]. 光谱学与光谱分析, 2016, 36(7): 2098-2103.
- [8] 黄必胜, 袁明洋, 余驰, 等. 基于近红外特征谱段相关系数法鉴别真伪龙齿[J]. 中国药师, 2014, 17(4): 619-622.
- [9] 史永刚, 粟斌, 田高友, 等. 化学计量学方法及MATLAB实现[M]. 北京: 中国石化出版社, 2010: 284.
- [10] 刘志霄, 梁亮, 俞晓莹. 基于可见-近红外光谱技术与BP-ANN算法的污水类型鉴定[J]. 激光与红外, 2009, 39(11): 1153-1157.
- [11] 陈龙, 明晶, 曹艳, 等. 24份紫石英样品的XRD谱分析[J]. 中药材, 2016, 39(1): 42-47.
- [12] 朱传静, 常琳, 康琛, 等. 紫石英炮制的历史沿革[J]. 中国实验方剂学杂志, 2011, 17(6): 270-274.
- [13] 张学博, 马金金, 曹丽梅. 近红外光谱相关系数法用于快速检测药品的质量[J]. 光谱实验室, 2013, 3(4): 2010-2015.

[责任编辑 刘德文]