

中图分类号: R932; R284.1; R286.0

文献标志码: A

文章编号: 1006-4931(2024)21-0062-05

doi:10.3969/j.issn.1006-4931.2024.21.015



红外光谱法结合化学计量学快速鉴别滑石和软滑石*

揭定华, 黄以财, 赵贵琴, 冷崇姣[△]

(重庆市永川食品药品检验所, 重庆 402160)

摘要:目的 建立快速鉴别滑石和软滑石的红外光谱法结合化学计量学方法。方法 采集25批滑石和15批软滑石样品的红外光谱, 选取 $1\,500\sim 400\text{ cm}^{-1}$ 和 $4\,000\sim 3\,000\text{ cm}^{-1}$ 范围内的红外光谱数据, 对数据进行相似度评价及标准正态变量变换(SNV)、多元散射校正(MSC)、SG平滑算法、二阶导数(2ND)、归一化等预处理, 对处理过的数据进行主成分分析(PCA)、层级聚类分析(HCA)、K最近邻算法(KNN)、偏最小二乘判别分析(PLS-DA)、支持向量机判别分析(SVM-DA)和随机森林(RF)分析, 比较混淆矩阵、受试者特征工作(ROC)曲线和曲线下面积(AUC)评价模型的预测性能。结果 PCA和HCA结果显示, 样品的红外光谱数据具有一定的特征性, 但聚类情况较差。SVM-DA的预测结果准确率较低, KNN、PLS-DA、RF分析的预测结果准确率较高, 其中RF分析的准确率可达100.0%, RF模型的预测性能均为最优。结论 所建立的以SNV-SG-2ND为光谱预处理方法结合RF判别模型能实现对滑石和软滑石的准确鉴别。

关键词: 红外光谱法; 化学计量学; 滑石; 软滑石; 快速鉴别

Rapid Identification of Talcum and Soft Talcum Based on Infrared Spectroscopy with Chemometrics

JIE Dinghua, HUANG Yicai, ZHAO Guiqin, LENG Chongjiao

(Yongchuan Institute for Food and Drug Control, Chongqing, China 402160)

Abstract: Objective To establish an infrared spectroscopy with the chemometric method for rapid identification of talcum and soft talcum. **Methods** Infrared spectra of 25 batches of talcum samples and 15 batches of soft talcum samples were collected. Infrared spectrum data within the ranges of $1\,500\sim 400\text{ cm}^{-1}$ and $4\,000\sim 3\,000\text{ cm}^{-1}$ were selected for similarity evaluation, and then preprocessed by standard normal variable (SNV), multivariate scattering correction (MSC), savitzky-golay smoothing (SG), second derivative (2ND) and normalization (NM), etc. The processed data were then put through the principal component analysis (PCA), hierarchical clustering analysis (HCA), K-nearest neighbor (KNN), partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), and random forest (RF) analysis. The confusion matrix, receiver operating characteristic (ROC) curve and area under the curve (AUC) were used to evaluate the predictive performance of the model. **Results** PCA and HCA showed that the infrared spectral data of the sample had certain characteristics, but the cluster of samples was poor. The accuracy of SVM-DA prediction result was relatively low, while that of KNN, PLS-DA, and RF analysis was higher. Among them, the accuracy of RF analysis could reach 100.0%, and the prediction performance of the RF model was the best. **Conclusion** The established spectral preprocessing SNV-SG-2ND method combined with RF discrimination model can achieve accurate identification of talcum and soft talcum.

Key words: infrared spectroscopy; stoichiometry; talcum; soft talcum; rapid identification

滑石作为传统利尿通淋、清热解暑中药, 在中医临床应用广泛, 在中成药制剂、化妆品中也大量使用^[1]。软滑石一般指高岭石, 曾在临床同滑石一样作为利尿通淋、清热解暑中药使用, 但二者主要成分完全不同, 且2020年版《中国药典(一部)》中仅收录滑石1项^[2]。滑石主含含水硅酸镁 $[\text{Mg}_3\text{Si}_4\text{O}_{10}(\text{OH})_2]$, 软滑石主含含水硅铝 $[\text{AlSi}_4\text{O}_{10}(\text{OH})_8]$ 。滑石与软滑石虽主要成分不同, 但性状极其相近, 单从性状特征难以区分, 且临床均为打粉入药, 打粉后更难区分^[3]。两者使用的鉴别方法均为化学鉴别法, 但实际运用中存在干扰大、反

应不充分、操作较烦琐等问题, 故需研究滑石和软滑石的其他鉴别方法, 优化质量检测手段, 以便鉴别区分^[4]。傅里叶红外光谱技术具有无损、快捷、重复性好、专属性强等特点, 能在化学定性方面提供较丰富的分子结构特征性数据信息, 已广泛应用于传统中药材的真伪鉴别及产地溯源^[5-10]。针对市场上滑石和软滑石混用的情况, 本研究中收集了不同地区、不同性状的滑石和软滑石, 采用红外光谱法结合化学计量学建立判别分析模型, 以快速鉴别滑石和软滑石, 为滑石的质量控制提供参考。现报道如下。

*基金项目: 重庆市永川区自然科学基金[2022yc-lhzz10002]。

第一作者: 揭定华, 男, 大学本科, 中级工程师, 研究方向为中药质量控制, (电子信箱)364965699@qq.com。

[△]通信作者: 冷崇姣, 女, 大学本科, 高级工程师, 研究方向为药物质量控制, (电子信箱)604033643@qq.com。

1 仪器与试药

1.1 仪器

Spectrum 100型红外光谱仪(美国PerkinElmer公司);JSP-24型粉末压片机(上海精胜公司);ME204E型电子天平(梅特勒托利多<上海>公司,精度为0.1 mg);SS-1022型多功能粉碎机(武义海纳电器公司);Omnics 8.0光谱软件(美国Thermo Fisher Scientific公司);Origin 2023b软件(美国Math works公司);Python 3.8软件(美国Python Software Foundation公司);Spectrum软件(美国Spectrum公司)。

1.2 试药

溴化钾(光谱纯,上海国药医药零售有限公司);滑石和软滑石样品来源于四川成都荷花池中药材批发市场、重庆众景中药饮片有限责任公司、网购及四川青川实地购买,共40批,其中编号为H1-H25的样品为滑石,编号为R1-R15的样品为软滑石,均经重庆市食品药品检验检测研究院副主任药师唐祖泽鉴定为正品。

2 方法与结果

2.1 试验条件

红外光谱扫描范围:4 000~400 cm^{-1} ;扫描信号时间:45 s;分辨率:4 cm^{-1} ;扫描时实时扣除水和二氧化碳的干扰。

2.2 样品处理

将滑石和软滑石样品粉碎,过200目筛,干燥,转移至密封袋,密封,保存,备用。

2.3 方法学考察

精密度试验:取2.2项下滑石样品粉末适量,精密称定,按2.1项下试验条件连续进样测定5次,用Spectrum软件计算光谱间的相关系数。结果的RSD为0.18%($n=5$),表明仪器精密度良好。

重复性试验:取同一滑石样品(编号为H1)粉末5份,按2.1项下试验条件连续进样测定,用Spectrum软件计算光谱间的相关系数。结果的RSD为0.34%($n=5$),表明方法重复性良好。

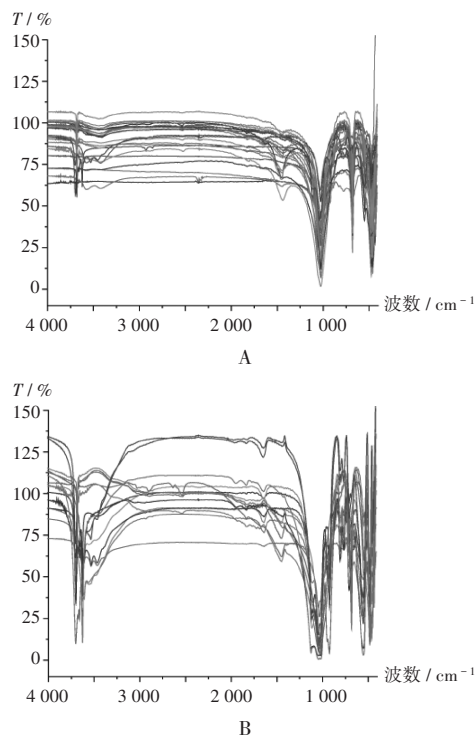
稳定性试验:取同一滑石样品(编号为H1)粉末,分别于0,2,4,8,16,24 h时按2.1项下试验条件进样测定,用Spectrum软件计算光谱间的相关系数。结果的RSD为0.22%($n=6$),表明样品在24 h内稳定性良好。

2.4 红外光谱采集与光谱波段范围选择

将溴化钾与样品粉末研磨均匀,按1:100(m/m)的比例混匀,用手动压片机制成厚度约为1 mm的透明錠片,按2.1项下试验条件进样测定。结果见图1。选择1 500~400 cm^{-1} 和4 000~3 000 cm^{-1} 2个范围内的红外光谱数据作为研究对象。

2.5 对照红外图谱建立与分析

将滑石和软滑石样品的红外光谱数据的平均值作



A. 滑石 B. 软滑石

图1 红外光谱图

A. Talcum B. Soft talcum

Fig. 1 Infrared spectra

为对照图谱的吸收强度,分析红外图谱共有特征峰,滑石和软滑石分别提取7个和11个特征峰。表征如下:滑石中,3 677 cm^{-1} 处为O—H伸缩振动;1 018 cm^{-1} 处为平行层的Si—O伸缩振动;670 cm^{-1} 处为O—H摇摆振动;535 cm^{-1} 处与垂直的Si—O—Mg伸缩振动有关;466,451 cm^{-1} 处为Si—O弯曲振动,一般滑石在460 cm^{-1} 附近出现强吸收带,同时显示弱分裂,分裂为466,451 cm^{-1} ,这是由其Si—O四面体对称特性导致的。软滑石中,3 695,3 677,3 620 cm^{-1} 处为O—H伸缩振动;1 102,1 018 cm^{-1} 处为Si—O伸缩振动,峰形与高岭石类矿物相近;912 cm^{-1} 处为外羟基摇摆振动;535 cm^{-1} 处与垂直的Si—O—Mg伸缩振动有关;670 cm^{-1} 处为O—H摇摆振动;797,756 cm^{-1} 处可能为Si—O—Si振动和Si—O—Al振动,也可能是特征的垂直振动;538 cm^{-1} 归属于Si—O—Al振动;470,435 cm^{-1} 处可能为Si—O弯曲振动、Al—O伸缩振动和O—H平动的耦合振动。详见图2。

2.6 相似度分析

对样品的红外光谱图进行相似度分析,以对照图谱为标准图谱计算相关系数。结果出现了对应品种的相关系数低于非本品类相关系数的现象,如编号为H2, H4, H10的样品本身为滑石,但与滑石对照图谱的相关系数低于软滑石对照图谱的相关系数;编号为R7, R8的样品本身为软滑石,但与软滑石对照图谱的相关系

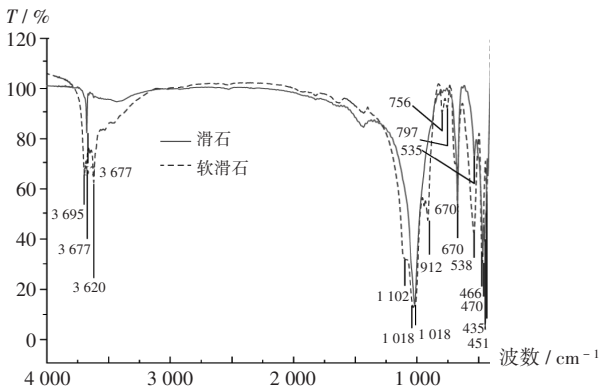


图2 滑石和软滑石的对照红外光谱图

Fig. 2 Comparative infrared spectra of talcum and soft talcum

表1 滑石和软滑石相似度分析结果

Tab. 1 Results of similarity analysis of talcum and soft talcum

编号	相关系数		编号	相关系数	
	滑石	软滑石		滑石	软滑石
H1	0.991 41	0.864 51	H21	0.903 86	0.758 26
H2	0.755 68	0.901 79	H22	0.962 07	0.850 81
H3	0.885 49	0.727 83	H23	0.995 04	0.855 14
H4	0.606 09	0.613 75	H24	0.985 45	0.844 95
H5	0.989 55	0.835 80	H25	0.757 46	0.921 58
H6	0.993 14	0.859 01	R1	0.748 69	0.885 16
H7	0.980 91	0.846 29	R2	0.660 90	0.936 24
H8	0.956 83	0.836 35	R3	0.677 44	0.860 85
H9	0.948 59	0.848 67	R4	0.585 07	0.857 16
H10	0.558 13	0.576 62	R5	0.733 77	0.830 40
H11	0.993 61	0.854 72	R6	0.674 99	0.924 41
H12	0.992 29	0.860 65	R7	0.875 61	0.868 05
H13	0.995 80	0.873 33	R8	0.953 11	0.738 07
H14	0.985 78	0.861 15	R9	0.933 74	0.828 55
H15	0.969 56	0.802 98	R10	0.973 15	0.812 45
H16	0.991 66	0.838 63	R11	0.705 13	0.822 80
H17	0.996 20	0.856 80	R12	0.652 79	0.925 13
H18	0.997 13	0.857 35	R13	0.680 17	0.914 45
H19	0.958 72	0.782 47	R14	0.660 84	0.908 41
H20	0.992 96	0.858 29	R15	0.951 55	0.909 85

数低于与滑石的对照图谱相关系数。故不能通过相似度分析实现对滑石和软滑石的区分,需进一步借助数学统计方法对数据进行处理和分析。详见表1。

2.7 红外光谱鉴别模型建立与验证

光谱预处理方法^[11]:本研究中采用了多元散射校正(MSC)法、标准正态变量变换(SNV)法、二阶导数(2ND)法、SG平滑算法、归一化(NM)法及其相互组合等多种方式降低样品及测量带来的误差。

无监督分析:分别将不同方式预处理样品的红外光谱数据组成待分析矩阵,并对矩阵进行主成分分析(PCA),PCA得分图见图3。可见,不同数据预处理方式

对样品在三维得分图中的分布有明显影响;图3 A中的原始红外数据并不能将滑石和软滑石分离开来,而图3 C中采用SNV-SG-2ND处理方式能实现二者的完全离散,故后续将SNV-SG-2ND预处理后的数据作为研究对象,层级聚类分析(HCA)结果见图4。可见,在PCA中即使能实现滑石和软滑石的完全分离,但其聚类效果依然不佳,虽然大部分样品能聚类成功,但少部分样品的聚类结果出现交叉。为得到更直观、准确的结果,借助监督分析的计量学模型进一步分析数据。

监督分析:K最近邻算(KNN)法是一种基本分类和回归方法,通过判断与样品最邻近的K个样本类别来决定待分样本所属类别^[12]。偏最小二乘判别分析(PLS-DA)法通过投影分别将预测变量和观测变量投影到一个新空间,寻找一个线性回归模型^[13]。支持向量机(SVM)是基于统计学理论的机器学习方法,在分类误差最小的情况下寻找最佳分割超平面,达到在低样本量情况下获得良好分类准确度的目的^[14]。随机森林(RF)是Breiman等在决策树基础上提出的一种机器学习算法,它是一个包含多个决策树的分类器,将多棵决策树在集成学习思维下作为分类器^[15]。将SNV-SG-2ND预处理的数据采用PLS-DA法、KNN法、SVM-DA法、RF法对样品进行建模,将样品划分为70%的训练集和30%的验证集,鉴别结果见表2。可见,RF法对验证集的鉴别准确率为100.0%,证明RF法可实现对滑石和软滑石红外光谱的准确区分。

判别分析模型评价:通过混淆矩阵直观反映分类模型的性能,评价各种模型的预测结果,即通过矩阵的形式表示样品的真实类别和模型的预测结果。结果显示,KNN模型和PLS-DA模型均有1批软滑石样品判定为滑石,SVM-DA模型中均将所有软滑石判定为滑石,对软滑石的误判率为100.0%,RF模型对滑石与软滑石鉴别的准确率均为100.0%。详见图5。但当模型样本分布发生变化时,混淆矩阵判别结果可能会随之发生变化。受试者特征工作(ROC)曲线评价模型时受样本分布影响较小,即样本分布发生变化时,分离效果基本不变^[16]。ROC曲线主要由于其曲线下面积(AUC)评价模型的泛化能力,而AUC值与分类器性能呈正相关。AUC值越接近1,模型分类性能越好。不同判别模型的ROC曲线见图6,4种模型的AUC值分别为0.943,0.929,0.943,1.000,表明RF模型在滑石和软滑石的鉴别中具有更好的性能和泛化能力。

3 讨论

3.1 光谱波段范围选择

红外光谱数据分布在4 000~400 cm⁻¹范围内,样品所得在该范围内的数据矩阵为40×1 869。在前期的滑石和软滑石红外光谱特征性分析中发现,在3 000~



A. SNV B. MSC C. SNV - SG - 2ND D. MSC - SG - 2ND E. SNV - 2ND - NM F. MSC - 2ND - NM

距离

1.2

1.0

0.8

0.6

0.4

0.2

0

1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 0 1 0 1 0 0 0 0 0 0 0 0

样本种类

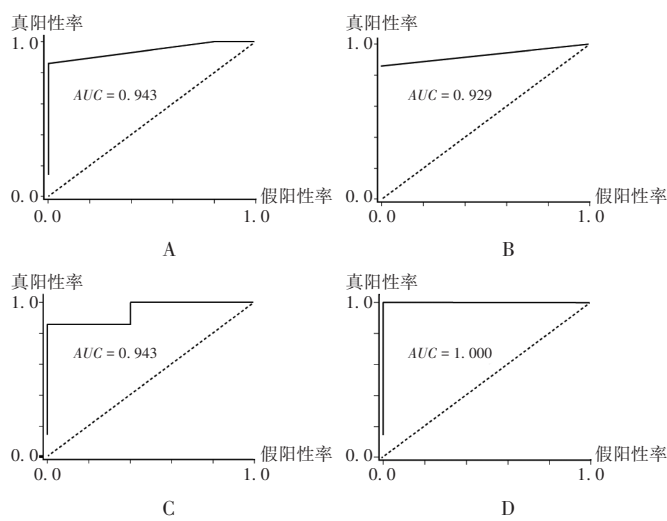
注:1 指滑石, 0 指软滑石。

Note:1 refers to the talcum,and 0 refers to the soft talcum.

Tab.2 Identification results of different discriminant models

A. KNN B. PLS - DA C. SVM - DA D. RF

1 500 cm^{-1} 范围内并无明显特征吸收峰,其特征吸收峰主要位于1 500~400 cm^{-1} 和4 000~3 000 cm^{-1} 2个范



A. KNN B. PLS - DA C. SVM - DA D. RF

图6 不同判别模型的ROC曲线

A. KNN B. PLS - DA C. SVM - DA D. RF

Fig. 6 ROC curves of different discriminant models

围内。为提高计算速度,减少冗余数据的影响,选择这2个范围的光谱数据进行分析。选定分析范围后的数据矩阵缩小为 $40 \times 1\,089$,在有效信息得到保留的基础上,数据量得到极大缩减。

3.2 光谱预处理方法比较

数据预处理可有效降低环境、压片差异、仪器噪声等干扰因素对光谱的影响。本研究中比较了6种具有代表性的预处理方式,分别为SNV, MSC, SNV - SG - 2ND, MSC - SG - 2ND, SNV - 2ND - NM, MSC - 2ND - NM。根据这6种预处理方式在PCA中得到了样品离散分布图,最终选择SNV - SG - 2ND作为数据预处理方式。

3.3 判别分析模型比较

红外光谱能全面反映中药材的信息,结合化学计量学能有效实现对中药材的真伪鉴别。本研究中借助常用的2种无监督和4种监督方法对滑石和软滑石的红外光谱数据进行分析。PCA和HCA结果显示,滑石和软滑石的红外光谱原始数据无明显特征性,判别分析的效果不佳,借助适当的数据预处理,能在PCA中实现良好分离,但离散程度不高。HCA结果中,虽然多数滑石和软滑石能聚类分开,但效果不佳,难以满足鉴别准确性的要求,需采用更先进的识别模式和算法来提高红外光谱技术的识别能力。与无监督分析不同,监督分析是基于分类特征值已知的情况下带入模型中计算,通过验证集的判别准确率、混淆矩阵、ROC曲线来衡量模型是否合适。本研究中将经过SNV - SG - 2ND预处理后的数据导入Python 3.8软件,分别基于KNN, PLS - DA, SVM - DA, RF 4种模型对滑石与软滑石进行鉴别,模型验证集的准确率分别为91.6%, 91.6%, 58.3%, 100.0%。除SVM模型外,其余模型的预测效果都较理想,其中RF可达到100.0%,实现对目标的准确区分。且

混淆矩阵、ROC曲线等多重指标显示,RF模型在滑石和软滑石的鉴别中具有较好的性能和泛化能力。故选择RF作为最终的判别分析模型,实现对滑石和软滑石的鉴别。

3.4 方法评价

本研究中以SNV - SG - 2ND为光谱预处理方法,基于RF算法建立红外光谱分析模型,能快速、准确地识别滑石和软滑石的红外光谱特征差异,有效区分滑石和软滑石,为滑石和软滑石提供了准确、快速的鉴别方法。且该方法具有无损、成本低、可在线检测等优点,在药品监管部门、医院、药店有广泛的应用前景。

参考文献

- [1] 任 蕾,赵 婧,吕晓武,等. 滑石不同炮制品的物质基础与抗菌活性研究[J]. 西北药学杂志,2023,38(3):35 - 38.
- [2] 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,2020:364.
- [3] 高 明,黄文青. 滑石浅谈[J]. 中药材,1987,47(6):28 - 30.
- [4] 张英华. 滑石质量评定的基础研究[J]. 国外医学:中医中药分册,2003,25(1):46.
- [5] 李秀梅,徐芳芳,张 欣,等. 基于近红外光谱和中红外光谱技术的金振口服液中间体含量预测模型研究[J]. 中草药,2023,54(24):8007 - 8017.
- [6] 孙琳琳,张 磊,孙辛未. 采用傅立叶红外光谱法鉴别牛皮和聚氨酯革的研究[J]. 合成材料老化与应用,2023,52(4):34 - 35.
- [7] 张志勇,韩 宁,周 涛,等. 基于二维相关近红外光谱的老陈醋醋龄鉴别[J]. 中国食品学报,2023,23(8):389 - 395.
- [8] 周飞翔,姜 红,胡晓光,等. 基于化学计量学的快递塑料包裹袋红外光谱特征分析[J]. 塑料工业,2023,51(6):110 - 114.
- [9] 田春婷,赵 宁,秦建伟,等. 基于机器学习的红外光谱数据鉴别中药材性能方法[J]. 兰州理工大学学报,2023,49(3):55 - 59.
- [10] 李艳萍,刘 冰,刘 楠,等. 不同种质宁夏枸杞红外光谱特征及其与表型性状的相关性分析[J]. 食品研究与开发,2023,44(9):178 - 182.
- [11] 孙嘉豪,张 伟,施鉴芬,等. 光谱数据预处理策略选择及应用[J]. 计量学报,2023,44(8):1284 - 1292.
- [12] 毋雪雁,王水花,张煜东. K最近邻算法理论与应用综述[J]. 计算机工程与应用,2017,53(21):1 - 7.
- [13] 张 琳,张黎明,李 燕,等. 偏最小二乘法在傅里叶变换红外光谱中的应用及进展[J]. 光谱学与光谱分析,2005,25(10):1610 - 1613.
- [14] 林香亮,袁 瑞,孙玉秋,等. 支持向量机的基本理论和研究进展[J]. 长江大学学报(自科版),2018,15(17):48 - 53.
- [15] 张 乐,吴静珠,李江波,等. 基于随机森林的单粒玉米种子水分近红外快速定量检测[J]. 中国粮油学报,2021,36(12):114 - 119.
- [16] 陈泽鹏. ROC分析在多分类问题中的应用[D]. 广州:广东工业大学,2022.

(收稿日期:2023 - 12 - 15;修回日期:2024 - 05 - 15)