

# 基于 FTIR、RS 光谱结合化学计量学鉴别白矾真伪及其不同炮制品

杨欣欣, 黄梓骏, 李世昌, 康爽, 徐天艺, 杨波, 于澎\*  
(长春中医药大学, 吉林 长春 130117)

**摘要:** 目的 鉴别白矾真伪及其不同炮制品。方法 傅里叶变换红外光谱 (FTIR) 和拉曼光谱 (RS) 分别对正伪品及枯矾、蜡矾进行测定, 建立指纹图谱, 主成分分析 (PCA) 对数据进行处理, 单一光谱和数据融合策略建立偏最小二乘判别分析 (PLS-DA) 模型和支持向量机 (SVM) 模型, 对样品进行判别分析。结果 在 PCA 中, 各组间样品难以完全区分; 在 PLS-DA 和 SVM 模型中, RS 数据展现出较高的稳定性和 100% 的判别准确率, 并且优于经 FTIR、RS 光谱数据融合后的模型。结论 光谱学结合化学计量学在白矾鉴别、质量研究中具有可行性, 拉曼光谱在该药材真伪、不同炮制品鉴别中具有良好的适用性和发展潜力。

**关键词:** 白矾; 真伪; 炮制品; 傅里叶变换红外光谱 (FTIR); 拉曼光谱 (RS); 主成分分析 (PCA); 偏最小二乘判别分析 (PLS-DA); 支持向量机 (SVM)

中图分类号: R284.1

文献标志码: B

文章编号: 1001-1528(2025)10-3510-08

doi: 10.3969/j.issn.1001-1528.2025.10.051

白矾作为中医临床常用矿物药, 首载于《神农本草经》, 主要成分为含水硫酸铝钾  $[KAl(SO_4)_2 \cdot 12H_2O]$ , 具有消炎、止血、止泻等药理作用<sup>[1]</sup>, 临床应用中需根据病症选择生品或炮制品 (枯矾、蜡矾), 但市场流通的白矾约 88% 为铵明矾  $[NH_4Al(SO_4)_2 \cdot 12H_2O]$  伪品<sup>[2]</sup>, 而且生品与炮制品存在外观同质化问题。尽管 2020 年版《中国药典》<sup>[3]</sup> 中新增加了铵盐检查项以控制白矾及其炮制品质量, 但操作复杂, 缺乏快速无损的鉴别、质量评价方法。另外, 药典对矿物药进行鉴别及含量测定时, 是以相应金属离子的测定为主, 存在方法相同、专属性较弱等问题, 导致了伪品盛行, 生品和炮制品混用、滥用现象发生, 直接影响了其临床用药的安全性和有效性<sup>[4]</sup>。

由于矿物药主要成分明确, 故光谱技术能有效揭示其组成特征。目前, 已有研究采用 X 射线衍射<sup>[5-6]</sup>、红外光谱<sup>[7-9]</sup>、拉曼光谱<sup>[10-12]</sup> 等技术对同类矿物进行鉴别, 其中傅里叶变换红外光谱 (FTIR) 和拉曼光谱技术 (RS) 因其高效性和高灵敏度, 在化学分析、材料科学领域获得了广泛应用, 并且近年来两者联合化学计量学在多个研究方面的数据处理中<sup>[13-15]</sup>, 可用于优化复杂体系, 通过算法模型来简化流程并提升区分效果<sup>[16]</sup>, 整合互补信息来更全面地解析样品结构。因此, 本实验将 FTIR、RS 与化学计量学相结合, 针对白矾正伪品及 2020 年版《中国药典》收录的枯矾、古法炮制品蜂巢煅制白矾<sup>[17]</sup> (简称蜡矾), 通过光谱特征解析与数据融合策略构建机器学习模型, 旨在建立该药材快速无损的鉴别方法。

## 1 材料

白矾共 26 批, 产自山东、浙江等地, 按照 2020 年版《中国药典》白矾项下质量要求进行检查, 发现 16 批为正品 (编号 N1~N16), 另外 10 批铵盐检查不合格而为伪品 (编号 N17~N26), 具体见表 1。

表 1 白矾信息

编号	产地	批号	编号	产地	批号
N1	浙江	HX16C01	N14	浙江	2209001
N2	山东	20191127	N15	山东	8220526301
N3	山东	20190628	N16	浙江	1191101068
N4	山东	20190812	N17	—	—
N5	山东	20200515	N18	浙江	191201
N6	山东	20200613	N19	山东	20191006
N7	山东	200701	N20	山东	20191029
N8	山东	200801	N21	山东	20191108
N9	山东	20080101	N22	山东	20190602
N10	浙江	2004050	N23	山东	20190826
N11	山东	200801	N24	河北	20130601
N12	浙江	2207001	N25	浙江	—
N13	浙江	2207002	N26	浙江	—

将 16 批正品按课题组前期炮制工艺进行处理, 其中枯矾 (DN1~DN16) 制备工艺为第一阶段, 煅制温度 250 °C, 煅制时间 1.5 h; 第二阶段, 煅制温度 160 °C, 煅制时间 1 h, 即得<sup>[18]</sup>, 而蜡矾 (SN1~SN16) 制备工艺为在 587 °C 下煅制 60 min 后, 迅速加入白矾 0.6 倍量的蜂巢, 保持上述温度继续煅烧至 180 min, 研磨, 包裹在纸中, 埋入潮湿土壤里过夜, 次日取出, 即得<sup>[19]</sup>。再将上述所有样品磨

收稿日期: 2025-04-10

基金项目: 吉林省自然科学基金资助项目 (YDZJ202501ZYTS271)

作者简介: 杨欣欣 (1980—), 女, 博士, 副教授, 从事中药炮制研究。Tel: 17704314195, E-mail: yangxx@ccucm.edu.cn

\* 通信作者: 于澎 (1978—), 男, 硕士, 教授, 从事中药炮制研究。Tel: 13180800022, E-mail: yupeng@ccucm.edu.cn

碎, 过 100 目筛, 制成光谱分析所需的微细粉末。

## 2 方法

2.1 FTIR 数据采集 样品粉末、KBr 以 1 : 100 比例在玛瑙研钵中混合研磨, 在 10 MPa 下压制 2 min, 制成试样片, 设定分辨率 0.482  $1\text{ cm}^{-1}$ , 波数 4 000~400  $\text{cm}^{-1}$ , 环境温度 25  $^{\circ}\text{C}$  以下, 相对湿度 40% 以下, 进行 50 次累积扫描, 平行取样 3 次后各扫描 2 次, 取平均值。

2.2 RS 采集 取样品 0.2 g, 均匀铺设于载玻片上, 先通过 5 倍物镜调整至清晰视野, 再切换至 20 倍物镜继续调整以确保图像清晰, 设定激发光源波长 785 nm, 光谱测量范围 2 000~100  $\text{cm}^{-1}$ , 激光功率 50 mW, 光纤探头 1.5 m, 激光强度 100%, 扫描时间 10 s, 平行取样 3 次后各扫描 2 次, 取平均值。

2.3 数据处理 流程图见图 1。

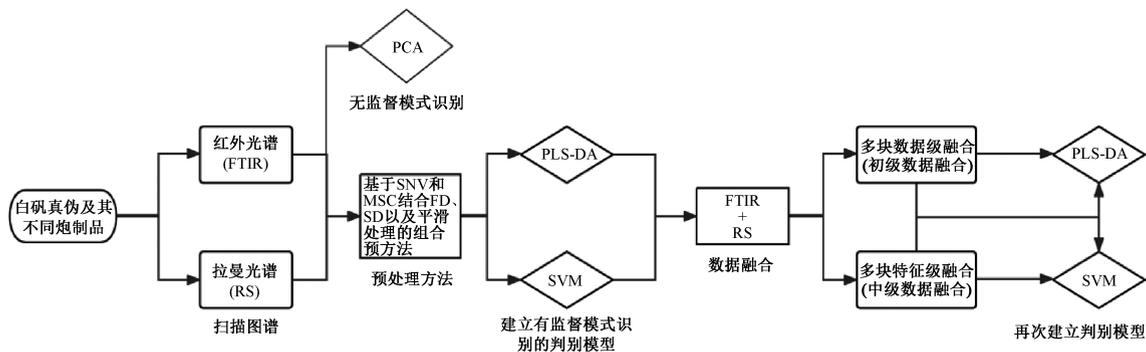


图 1 数据处理流程图

2.3.1 主成分分析 (PCA) 参照文献 [20] 报道, 提取特征峰后采用 SPSS 19.0 软件计算 58 组样本, SIMCA 14.1 软件运行模型得到散点图, Hotelling  $T^2$  检验筛选异常值。结果, 未发现异常样本, 具有较高的分类模型说服力。

2.3.2 数据预处理 光谱在采集过程中易受环境影响, 产生光散射等干扰, 预处理方法主要包括基线校正、散射校正和平滑。其中, 导数运算中的一阶导数 (FD) 和二阶导数 (SD) 可有效消除基线和其他背景干扰, 区分重叠峰; 散射校正中的标准正态变量变换 (SNV) 和多元散射校正 (MSC) 可消除颗粒分布和大小引起的散射影响<sup>[21-22]</sup>; Savitzky-Golay 平滑 (SG 平滑) 主要用于噪声的消除。本实验基于 SNV 和 MSC 结合 FD 和 SD, 以及平滑处理组合的预处理方法对光谱数据进行处理, 筛选最优方式。

### 2.3.3 机器学习算法

2.3.3.1 偏最小二乘判别分析 (PLS-DA) PLS-DA 是在 PCA 分析基础上, 同时对光谱矩阵和样品浓度矩阵进行分析的一种有监督的模式识别方法, 常用于解决样品数少于变量维度的问题。结合探索性分析结果发现, 本实验所用数据集与模型的契合度更高, 故采用 PLS-DA 进行分类判别。

2.3.3.2 支持向量机 (SVM) 本实验采用 MATLAB 2021 软件, 通过构造一个最优的线性分类来形成具有最大间隔的超平面, 同时使经验分类误差最小化, 几何间隔最大化。另外, 在不是线性可分的数据集的情况下, 将原始数据映射到由内核函数给出的高维特征空间进行分类。

2.3.3.3 数据融合 在处理源自不同途径 (如传感器类型或处理方式) 获得的数据时, 可利用多块信息集成策略来合并这些数据, 进一步支持全面的数据集成<sup>[23]</sup>, 主要包括多块数据级融合、多块特性级融合 (中级数据融合)、多块决策级融合 (高级数据融合)。本实验为了更深入地挖掘结果, 获取更详尽的结论, 选择多块数据级融合和特性级融

合策略<sup>[24]</sup>开展研究。

2.4 模型性能评价 本实验采用 PLS-DA 和 SVM, 对白矾正伪品及其不同炮制品的光谱特性进行分析。在 PLS-DA 模型中, 以测试集正确率 (ACC-text) 为首要评价指标, 结合  $R^2Y$ 、 $Q^2$ 、RMSEcv 和 RMSEP 对结果进行综合评估; 在 SVM 模型中, ACC-text 被视为主要评价指标, 结合惩罚因子  $c$  和核函数参数  $g$  进行综合评估<sup>[25]</sup>。另外, 混淆矩阵可反映模型对正确样本分类的准确性, 可用  $F_1$  值作为查准率、查全率的平均值及 2 个模型的综合度量指标<sup>[26]</sup>。

## 3 结果

### 3.1 光谱分析

3.1.1 FTIR 在前期研究<sup>[19]</sup>基础上, 本实验发现正品共有峰有 9 个, 波数分别为 618、918、988、1 113、1 402、1 618、1 670、2 480、3 441  $\text{cm}^{-1}$ ; 桔矾共有峰有 8 个, 波数分别为 477、618、688、1 112、1 620、1 645、2 442、3 458  $\text{cm}^{-1}$ ; 蜡矾共有峰有 7 个, 波数分别为 472、617、688、1 121、1 242、1 640、3 446  $\text{cm}^{-1}$ ; 伪品共有峰有 9 个, 波数分别为 618、916、988、1 115、1 403、1 618、1 670、2 468、3 455  $\text{cm}^{-1}$ 。其中, 伪品特征峰的峰位与正品的十分相近, 两者区别主要体现在其峰形的红外峰面积上, 各样品红外特征峰指征见表 2, FTIR 图谱见图 2。

3.1.2 RS 由图 3 可知, 正品共有峰有 8 个, 波数分别为 151、190、449、617、973、990、1 133、1 367  $\text{cm}^{-1}$ ; 伪品共有峰有 7 个, 波数分别为 156、193、445、610、973、990、1 130  $\text{cm}^{-1}$ ; 桔矾共有峰有 9 个, 波数分别为 194、220、319、484、610、655、985、1 075、1 302  $\text{cm}^{-1}$ ; 蜡矾共有峰有 11 个, 波数分别为 196、225、319、451、484、612、655、983、1 047、1 078、1 302  $\text{cm}^{-1}$ 。

其中, 正品在 990  $\text{cm}^{-1}$  处展现出最强的光谱峰, 并伴有 973  $\text{cm}^{-1}$  处的次强肩峰, 而伪品主峰位置虽与正品一致,

表2 各样品红外特征峰指征

样品	波数/cm <sup>-1</sup>	主要官能团	推测物质
白矾正品	3 441	$\nu$ O-H	H <sub>2</sub> O
	1 670	$\delta$ O-H	
	1 113	$\nu$ SO <sub>4</sub> <sup>2-</sup>	硫酸盐
白矾伪品	918, 618	$\delta$ SO <sub>4</sub> <sup>2-</sup>	
	3 455	$\nu$ O-H	H <sub>2</sub> O
	1 670	$\delta$ O-H	
桔矾	1 115	$\nu$ SO <sub>4</sub> <sup>2-</sup>	硫酸盐
	916, 618	$\delta$ SO <sub>4</sub> <sup>2-</sup>	
	3 458	$\nu$ O-H	H <sub>2</sub> O
蜡矾	1 645	$\delta$ O-H	
	1 112	$\nu$ SO <sub>4</sub> <sup>2-</sup>	硫酸盐
	477, 618, 688	$\delta$ SO <sub>4</sub> <sup>2-</sup>	
	3 446	$\nu$ O-H	H <sub>2</sub> O
蜡矾	1 640	$\delta$ O-H	
	1 121	$\nu$ SO <sub>4</sub> <sup>2-</sup>	硫酸盐
	472, 617, 688	$\delta$ SO <sub>4</sub> <sup>2-</sup>	
	1 242	$\delta$ C-O-H <sup>[27]</sup>	糖类物质 <sup>[28-29]</sup>

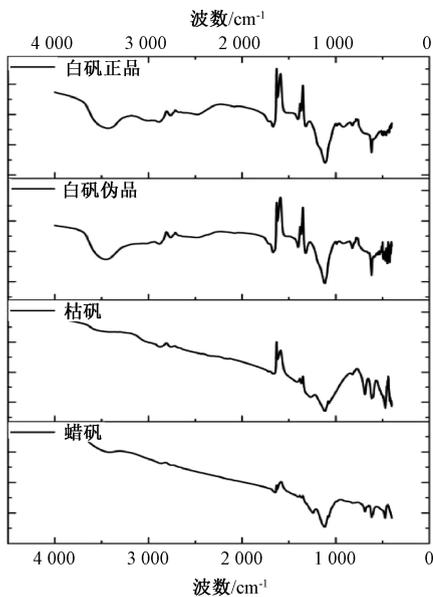


图2 各样品 FTIR 图谱

但 973 cm<sup>-1</sup>处的肩峰显著减弱, 而且缺失正品特有的 1 367 cm<sup>-1</sup>峰, 表明两者结构存在差异, 文献 [30] 报道, 伪品含水硫酸铝铵的最强峰不分裂, 与本实验结果一致, 进一步验证 973 cm<sup>-1</sup>处的肩峰可作为辅助鉴别特征; 桔矾最强峰在 1 075 cm<sup>-1</sup>, 而蜡矾峰值在 1 078 cm<sup>-1</sup>处, 并且在 1 047 cm<sup>-1</sup>处还存在肩峰, 均为其主成分硫酸铝钾中 SO<sub>4</sub><sup>2-</sup>的伸缩振动<sup>[31]</sup>, 同时因炮制过程中的高温煅烧而导致晶体结构发生变化, 脱去结晶水, 呈现不同最强峰位置<sup>[31-32]</sup>。

与生品比较, 桔矾与蜡矾在低波数范围内峰形尖锐, 表明晶体内物质纯净度提升, 与其所含的硫酸铝钾含量增加吻合, 并且后者在 983、1 302 cm<sup>-1</sup>处的峰表现出更高的强度和更尖锐的峰形, 即其含量与纯度高于前者。由于蜡矾炮制温度高于桔矾, 可能促使部分硫酸铝钾分解生成硫酸钾, 并且 983 cm<sup>-1</sup>处的峰与硫酸钾的已知 RS 特征符合,

可能为硫酸钾中 SO<sub>4</sub><sup>2-</sup>伸缩振动的表征。另外, 蜡矾在 451 cm<sup>-1</sup>处的峰属于 CCC 环变形振动<sup>[33]</sup>, 在 1 500、1 700 cm<sup>-1</sup>附近展现宽泛光谱峰, 可能源于炮制过程中加入的辅料蜂巢, 并且它含有黄酮类、苯丙素类、萜类、甾体、糖类成分, 在高温作用下发生碳化反应, 进而形成 2 个相关宽峰, 可能与碳基物质中 C-C 键的伸缩振动及无序碳材料中的缺陷结构有关。

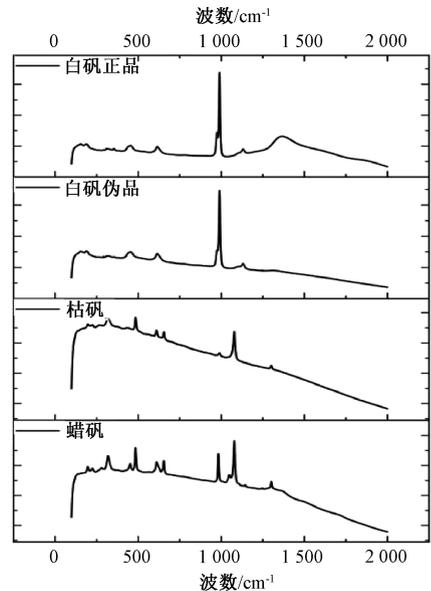


图3 各样品 RS 光谱

### 3.2 无监督模式识别分析

3.2.1 FTIR 光谱 以 477、618、688、918、1 113、1 670、3 450 cm<sup>-1</sup>处的峰为自变量, 其透射率为因变量, 采用 SPSS 19.0 软件对 58 组样本进行分析, 结果见表 3、图 4, 可知 3 个主成分涵盖了大部分信息。再采用 SIMCA 14.1 软件建立模型, 见图 5, 可知正品主要集中在偏左下方区域; 伪品主要位于下方区域, 点位分布松散; 桔矾分布分散; 蜡矾主要分布在上方区域, 虽然正品和蜡矾各自有较集中的聚类趋势, 但整体来看 4 组样品部分点位均出现交叉现象, 表明难以直接通过 PCA 分析来对其进行有效区分。

表3 FTIR 光谱主成分方差贡献率

主成分	参数		
	特征值	方差贡献率/%	累积方差贡献率/%
1	41.253	71.125	71.125
2	13.115	22.613	93.738
3	2.591	4.467	98.205

3.2.2 RS 光谱 以 451、484、610、655、973、985、990、1 075、1 130、1 302、1 367 cm<sup>-1</sup>处的峰为自变量, 其峰强为因变量, 采用 SPSS 19.0 软件对 58 组样本进行分析, 结果见表 4、图 6, 可知 4 个主成分涵盖了大部分信息。再采用 SIMCA 14.1 软件建立模型, 见图 7, 可知正品与伪品存在一定聚集现象, 主要分布在中上方区域; 桔矾与蜡矾聚集紧密, 主要分布在下方区域, 表现出聚类趋势, 表明两者差异较小, 但难以将白矾及其炮制品进行区分。

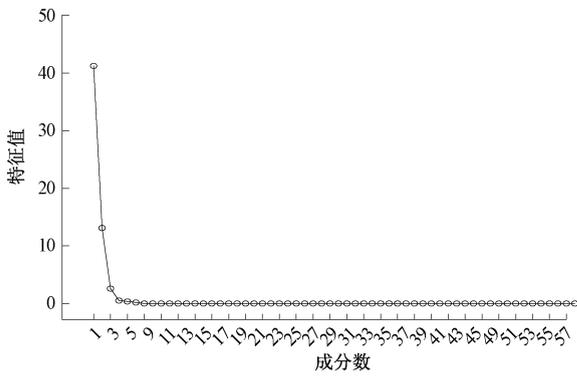


图4 FTIR光谱PCA分析碎石图

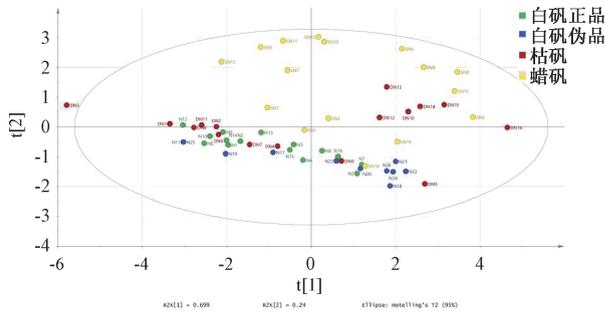


图5 FTIR光谱PCA分析得分图

表4 RS光谱主成分方差贡献率

主成分	参数		
	特征值	方差贡献率/%	累积方差贡献率/%
1	31.637	54.547	54.547
2	21.564	37.179	91.726
3	2.354	4.059	95.785
4	1.362	2.348	98.132

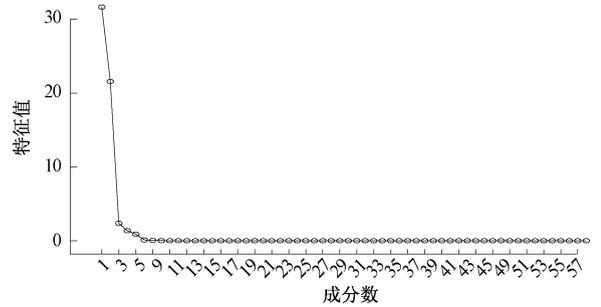


图6 RS光谱PCA分析碎石图

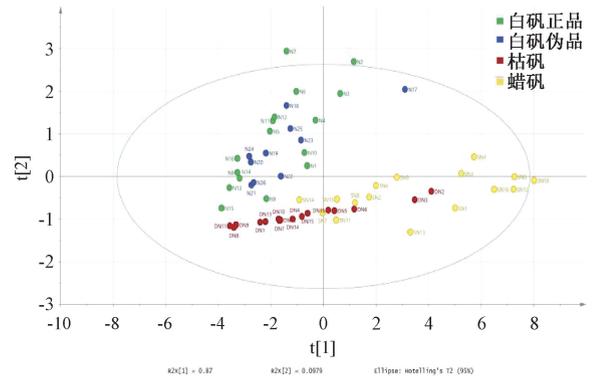


图7 RS光谱PCA分析得分图

3.3 有监督模式识别分析 由于生品与炮制品在颗粒粒径方面存在较大差异,在光谱中可能引发强烈的散射现象,故本实验采用散射校正作为主要预处理方法,并结合基线

校正和平滑处理,最终确定 SNV、MSC、FD+SNV、SD+SNV、FD+MSC、SD+MSC、FD+SNV+SG、SD+MSC+SG 用于 FTIR 光谱,SNV、FD+SNV、SD+SNV、FD+MSC、FD+SNV+SG、SD+MSC+SG 用于 RS 光谱,再建立 PLS-DA 模型和 SVM 模型,筛选最优预处理策略,结果见表 5。由此可知,对于 FTIR 数据块,建立 PLS-DA 模型时无预处理方法最优,而建立 SVM 模型时 FD+SNV 最优;对于 RS 数据块,建立 PLS-DA 模型时 SNV 预处理方法表现最优,而建立 SVM 模型时 FD+SNV 最优,故本实验采用 FTIR+RS 的多块

表5 不同数据块预处理模型参数 (%)

数据块	编号	预处理方法	PLS-DA				SVM			
			Trainingset-ACC	Testset-ACC	RMSEcv	RMSEP	Trainingset-ACC	Testset-ACC	Best c	Best g
FTIR	1	0	85.00	94.44	0.332 878	0.208 340	95.00	83.33	2.56×10 <sup>2</sup>	6.10×10 <sup>-5</sup>
	2	SNV	95.00	83.33	0.242 245	—	97.50	83.33	5.66	6.91×10 <sup>-4</sup>
	3	MSC	95.00	77.78	0.347 118	0.347 859	97.50	83.33	7.07×10 <sup>-1</sup>	5.52×10 <sup>-3</sup>
	4	FD+SNV	82.50	83.33	0.356 751	0.218 209	97.50	88.89	2.83	1.95×10 <sup>-3</sup>
	5	SD+SNV	80.00	83.33	0.379 910	0.326 498	97.50	72.22	4.00	2.76×10 <sup>-3</sup>
	6	FD+MSC	97.50	88.89	0.328 924	0.277 887	97.50	77.78	1.00	3.91×10 <sup>-3</sup>
	7	SD+MSC	82.50	83.33	0.363 350	0.318 819	97.50	83.33	1.28×10 <sup>2</sup>	4.32×10 <sup>-5</sup>
	8	FD+SNV+SG	82.50	83.33	0.356 515	0.219 236	97.50	88.89	2.83	1.95×10 <sup>-3</sup>
	9	SD+MSC+SG	82.50	83.33	0.362 961	0.320 929	97.50	83.33	1.60×10 <sup>1</sup>	3.45×10 <sup>-4</sup>
RS	1	0	87.50	94.44	0.304 999	0.317 153	97.50	100.00	3.62×10 <sup>2</sup>	6.91×10 <sup>-4</sup>
	2	SNV	100.00	100.00	0.130 196	0.200 102	100.00	100.00	1.13×10 <sup>1</sup>	2.21×10 <sup>-2</sup>
	3	FD+SNV	97.50	94.44	0.230 622	0.240 795	100.00	100.00	5.66	3.91×10 <sup>-3</sup>

续表 5

数据块	编号	预处理方法	PLS-DA				SVM			
			Trainingset-ACC	Testset-ACC	RMSEcv	RMSEP	Trainingset-ACC	Testset-ACC	Best <i>c</i>	Best <i>g</i>
	4	SD+SNV	97.50	83.33	0.244 629	0.300 762	100.00	100.00	2.26×10 <sup>1</sup>	7.81×10 <sup>-3</sup>
	5	FD+MSC	97.50	94.44	0.237 914	0.250 967	100.00	100.00	8.00	2.76×10 <sup>-3</sup>
	6	FD+SNV+SG	97.50	94.44	0.243 101	0.248 873	100.00	100.00	5.66	3.91×10 <sup>-3</sup>
	7	SD+MSC+SG	97.50	61.11	0.340 059	0.500 551	100.00	88.89	8.00	7.81×10 <sup>-3</sup>

信息融合以提升建模效率与判别准确性。

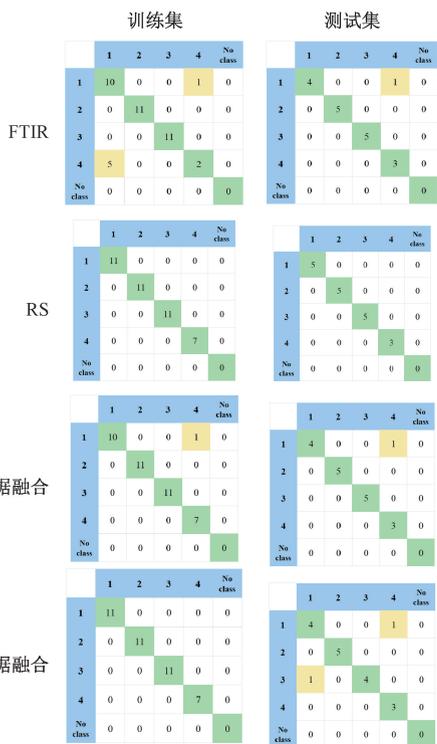
### 3.4 数据融合结果

3.4.1 PLS-DA 建立 PLS-DA 模型的关键在于准确确定潜在变量数量，并严格评估模型，应使 RMSEcv 最低、*Q*<sup>2</sup> 值最大以避免过拟合，并达到最优预测性能，同时模型评价需考量分类精度等因素，*Q*<sup>2</sup>、*R*<sup>2</sup>*Y* 等可反映模型性能<sup>[34-36]</sup>。本实验对 SNV+SG 处理后的 FTIR 数据集，以及 SD+SNV 处

理后的 RS 数据集分别施行初级数据融合（数据级融合）和中级数据融合（特征级融合），旨在获取解释能力更强的变量，再将 2 种融合策略所得变量进一步融合以建立 PLS-DA 模型，结果见表 6、图 8，再进行 200 次置换检验，结果见图 9，可知各类型 *Q*<sup>2</sup> 回归线的轴距为负数，表明不存在过拟合的情况，模型验证有效。

表 6 不同融合类型 PLS-DA 模型参数

类型	潜变量数	<i>R</i> <sup>2</sup> <i>Y</i>	<i>Q</i> <sup>2</sup>	RMSEcv	RMSEP	<i>F</i> <sub>1</sub>
FTIR 光谱	4	0.607	0.534	0.332 9	0.208 3	0.972 2
RS 光谱	13	0.978	0.879	0.130 2	0.200 1	1.000 0
初级数据融合	6	0.839	0.721	0.199 3	0.269 0	0.972 2
中级数据融合	3	0.844	0.574	0.304 8	0.361 3	0.922 2



注：1 为白矾正品，2 为白矾伪品，3 为枯矾，4 为蜡矾。

图 8 PLS-DA 分类模型混淆矩阵

另外，不同类型 PLS-DA 模型均展现出较高的判断准确率，揭示了模型之间的差异性，其中以 RS 光谱数据为基础的模型判断准确率达 100%；错误判断主要集中在正伪品之间，故更适合采用无监督模式识别分析；数据融合后，*R*<sup>2</sup>*Y*、*Q*<sup>2</sup> 等参数的表现均不如基于 RS 光谱数据的模型，表明红外数据可能会带来信息冗余，影

响分析结果。

3.4.2 SVM 在构建基于高斯核函数的 SVM 分类模型过程中，模型展现出优秀的分类、逼近性能，其性能优化主要依赖于惩罚因子 *c*、核函数参数 *g* 的细致调节。本实验建立 SVM 判别模型时，其策略与 PLS-DA 模型相仿，即在每种降维方案中均需对两部分数据块进行整合处理，并且通过网络搜索策略来探寻 *c*、*g* 最优组合，进而对相应样本数据执行网络训练，结果见表 7、图 10~13。

由此可知，SVM 模型呈现优越的辨识效能，展现了正伪品及其炮制品之间的显著区别，尤其是基于 RS 光谱时判别准确率达到 100%；数据融合后，*c*、*g* 等参数在效能上并未超越基于 RS 光谱的模型，可能是因为红外数据引入了额外的信息冗余，从而对结果产生不利影响。

## 4 讨论

虽然文献 [37] 采用 XRD 衍射峰和 FTIR 的 N-H 振动峰，并结合聚类分析及正交偏最小二乘法-判别分析来区分白矾与铵明矾，取得了较好的鉴别成果，但无监督模式识别方法面对种类复杂的样本时往往难以实现精准区分。本实验将多光谱技术与具有监督模式识别方法相结合，应用于矿物药白矾正伪品及枯矾、蜡矾鉴别中，获得了理想的结果，可为相关研究提供了新的思路和方法。

结果显示，RS 单一光谱模型优于数据融合模型，可能与前者对硫酸盐类矿物的 SO<sub>4</sub><sup>2-</sup> 振动峰响应更敏感，而且煅烧过程中结晶水脱除引起的峰位迁移在 RS 光谱中更清晰有关<sup>[31]</sup>。其中，SVM 模型通过参数优化实现了全类别正确区分，验证了其在小样本高维数据中的优势<sup>[25]</sup>，可为其他矿物药多光谱鉴别提供了方法学参考，也提示数据融合需结合具体样本特性选择策略。

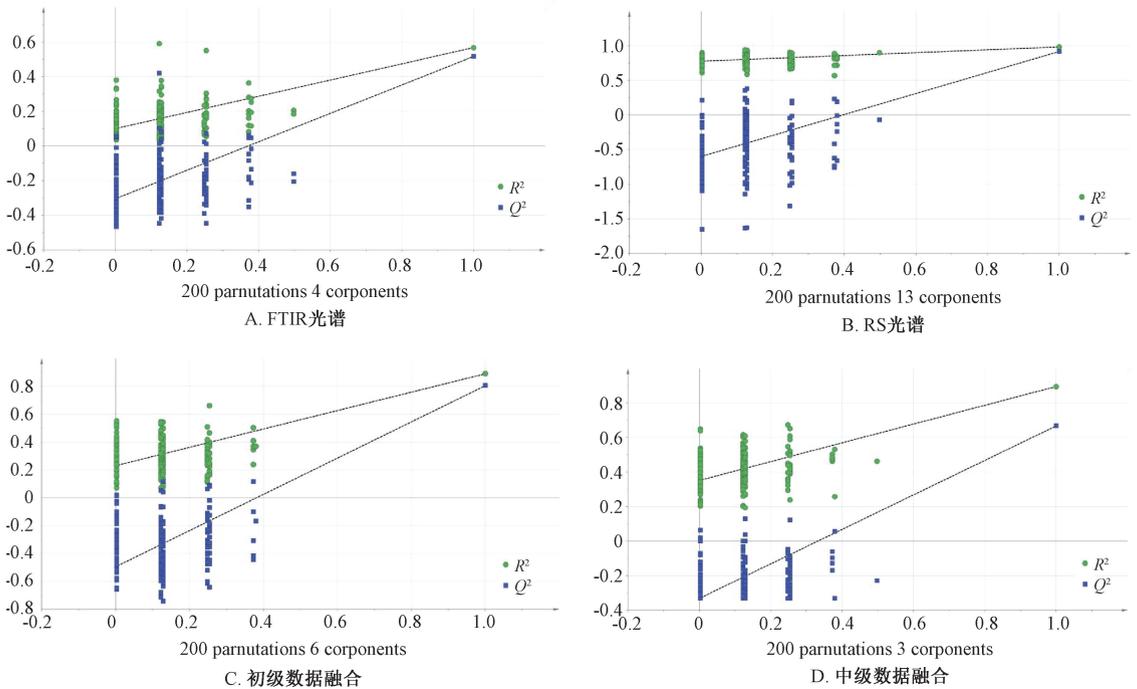


图9 置换检验图

表7 不同融合类型SVM模型指标

类型	训练集正确率/%	测试集正确率/%	Best $c$	Best $g$	$F_1$
FTIR 光谱	97.50	88.89	2.83	$1.95 \times 10^{-3}$	0.8295
RS 光谱	100.00	100.00	5.66	$3.91 \times 10^{-3}$	1.0000
初级数据融合	95.00	88.89	22.6	$1.22 \times 10^{-4}$	0.8860
中级数据融合	100.00	94.44	8.00	$6.25 \times 10^{-2}$	0.9365

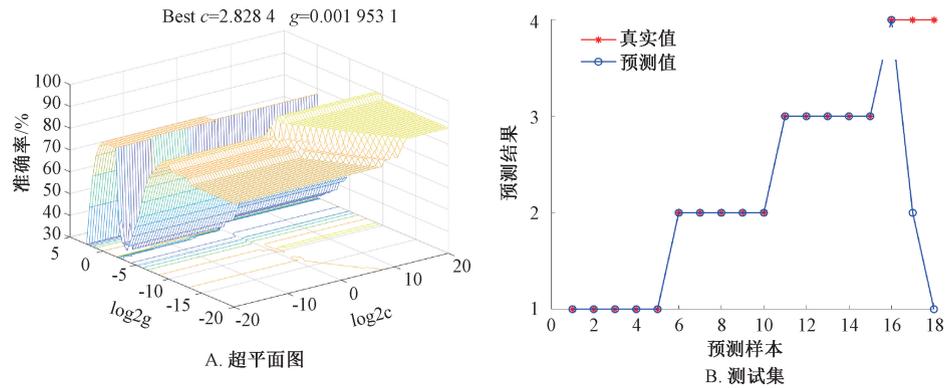


图10 基于FTIR光谱的SVM模型

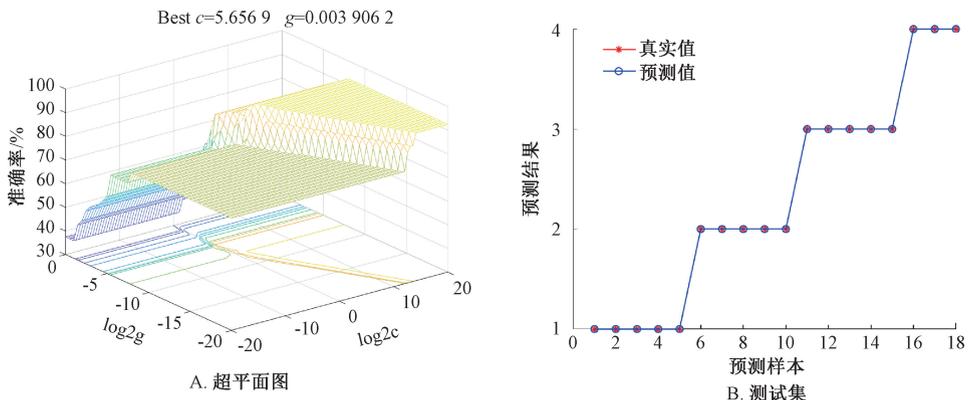


图11 基于RS光谱的SVM模型

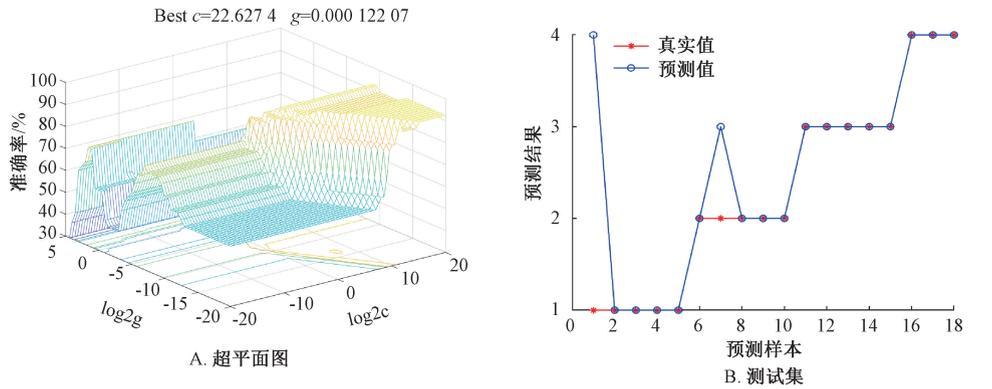


图 12 基于初级数据融合的 SVM 模型

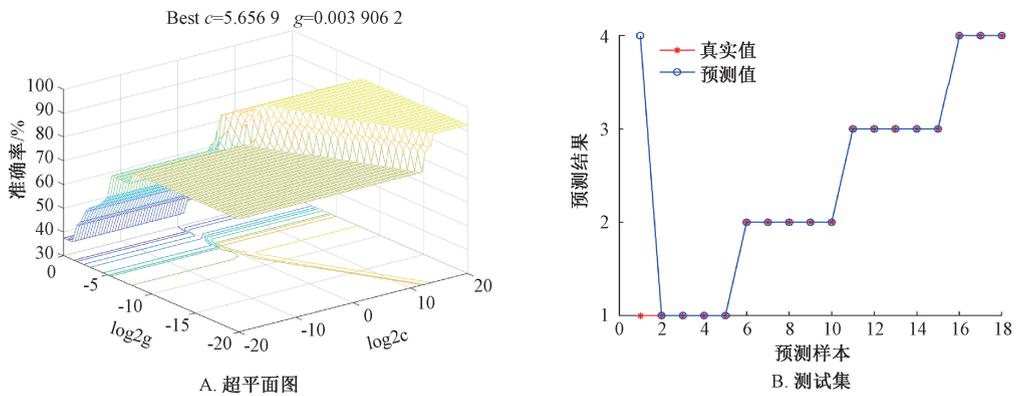


图 13 基于中级数据融合的 SVM 模型

随着机器学习技术的进一步发展及广泛应用，它在光谱数据分析领域的作用日益显著，特别是在处理包含复杂化学特性的光谱数据方面所展现出独特的模式优势，其中有监督学习相较于无监督学习，不仅能有效辨识不同化学特性，还可融合多来源数据，提高分析精度与可靠性，在提高鉴别精度和增强模型判别能力方面具有显著优势。然而，尽管本实验中数据融合表现突出，但在处理含有复杂化学特性的光谱数据时效果不理想，可能导致冗余信息增加，从而降低分析精度和准确性。因此，当光谱学结合模式识别技术对矿物药进行鉴别时，应选择最适合的方法，而不是盲目追求数据融合。

参考文献:

[ 1 ] 韩进庭. 白矾的药理作用及临床应用研究进展[J]. 现代医药卫生, 2006(24): 3763-3764.

[ 2 ] 林瑞超. 矿物药检测技术与质量控制[M]. 北京: 科学出版社, 2013: 685-708.

[ 3 ] 国家药典委员会. 中华人民共和国药典: 2020年版一部[S]. 北京: 中国医药科技出版社, 2020: 111-112.

[ 4 ] 王开心, 刘曼婷, 彭胡麟玥, 等. 《中华人民共和国药典》2020年版一部收载矿物药及含矿物药中成药的质量标准分析与探讨[J]. 中华中医药学刊, 2023, 41(1): 92-98; 280.

[ 5 ] 李 焱, 陈朝方, 林国生, 等. X 射线粉末衍射技术在矿物鉴别中的应用进展[J]. 冶金分析, 2023, 43(7): 11-20.

[ 6 ] 刘圣金, 乔婷婷, 马瑜璐, 等. 矿物药白矾、桔矾及其伪品的 SEM, XRD 鉴别分析[J]. 中国实验方剂学杂志,

2019, 25(5): 8-13.

[ 7 ] 张丽倩, 刘养杰. 4 种含铁矿物药矿物学鉴定及谱学特征对比研究[J]. 光谱学与光谱分析, 2022, 42(9): 2884-2889.

[ 8 ] 雷 咪, 陈 龙, 黄必胜, 等. 7 种硫酸盐类矿物药的近红外光谱鉴别方法研究[J]. 世界科学技术(中医药现代化), 2014, 16(11): 2385-2389.

[ 9 ] 闫 蔚. 硫酸盐类矿物药红外指纹图谱研究[D]. 广州: 广东药科大学, 2016.

[ 10 ] 雷 咪, 陈 龙, 黄必胜, 等. 6 种含硫酸盐的矿物类中药及其部分炮制品的拉曼光谱鉴别研究[J]. 中华中医药杂志, 2016, 31(7): 2811-2814.

[ 11 ] 韩斯琴高娃, 阿木古楞, 王士燕, 等. 基于拉曼光谱技术结合 PCA-SVM 算法对含汞矿物药的分类识别研究[J]. 中华中医药杂志, 2023, 38(4): 1815-1818.

[ 12 ] 韩斯琴高娃. 基于拉曼光谱技术的矿物药和成瘾性药物快速检测方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2022.

[ 13 ] 杨承恩, 武海巍, 杨 宇, 等. 中红外光谱和支持向量机的梅花鹿角帽粉假冒与掺假识别模型[J]. 光谱学与光谱分析, 2022, 42(8): 2359-2365.

[ 14 ] 江丽琴, 崔文波, 韩明澄, 等. 不同炮制方法黄精的红外光谱鉴别研究[J]. 光谱学与光谱分析, 2023, 43(S1): 53-54.

[ 15 ] 孙佳慧, 冯 源, 王 珊, 等. 冬虫夏草与蛹虫草的表面增强拉曼光谱结合化学计量学鉴定研究[J]. 药物分析杂志, 2022, 42(4): 727-733.

[ 16 ] 戴嘉伟, 王海朋, 陈 瀑, 等. 多光谱数据融合分析技术的

- 研究和应用进展[J]. 分析化学, 2022, 50(6): 839-849.
- [17] 雷 敷. 雷公炮炙论[M]. 张 骥, 补辑, 施仲安, 校注. 南京: 江苏科学技术出版社, 1985: 24.
- [18] 杨欣欣, 钱 圳, 许天阳, 等. 白矾煅制工艺优化[J]. 中成药, 2018, 40(6): 1351-1354.
- [19] 杨欣欣, 黄梓骏, 洪禹昕, 等. 基于《雷公炮炙论》的白矾蜂巢制炮制工艺研究[J]. 中草药, 2024, 55(12): 4008-4019.
- [20] 史静超, 张俊杰, 高晓霞, 等. 龟龄集胶囊无机元素指纹图谱的建立及主成分分析[J]. 中草药, 2018, 49(15): 3619-3624.
- [21] Rabatel G, Marini F, Walczak B, *et al.* VSN: Variable sorting for normalization[J]. *J Chemometr*, 2020, 34(2): e3164.
- [22] 王 靖. 近红外光谱和紫外-可见漫反射光谱结合化学计量学对三七及其相似品的鉴别及掺伪定量研究[D]. 天津: 天津工业大学, 2020.
- [23] Zhou J R, Hong X, Jin P Q. Information fusion for multi-source material data: Progress and challenges[J]. *Applied Sci*, 2019, 9(17): 3473.
- [24] 张 娇, 王元忠, 杨维泽, 等. ATR-FTIR 和 UV-Vis 结合数据融合策略鉴别滇黄精产地[J]. 光谱学与光谱分析, 2021, 41(5): 1410-1416.
- [25] Chen Y F, Chen Y, Feng X P, *et al.* Variety identification of orchids using fourier transform infrared spectroscopy combined with stacked sparse auto-encoder[J]. *Molecules*, 2019, 24(13): 2506.
- [26] Han M, Wang Y. Validating the feasibility of the method of ensemble learning combined with FT-MIR for the discrimination of wild *Paris polyphylla* var. *Yunnanensis* from different geographical sources[J]. *Microchemical J*, 2024, 202: 110824.
- [27] 曾 薇, 谭成森, 苏元翠, 等. 基于中红外光谱的含钩吻蜂蜜鉴别研究[J]. 光通信技术, 2023, 47(6): 77-82.
- [28] Santos M I, Araujo-Andrade C, Tymczynszyn E E, *et al.* Determination of amorphous/rubbery states in freeze-dried prebiotic sugars using a combined approach of near-infrared spectroscopy and multivariate analysis[J]. *Food Res Int*, 2014, 64: 514-519.
- [29] 王雨宵, 常相伟, 陈博文, 等. 蜂胶化学成分、药理作用及临床应用研究进展[J]. 中国现代中药, 2023, 25(3): 665-685.
- [30] 明 晶, 陈 龙, 姚 琪, 等. 基于 XRD, NIR 及拉曼光谱技术快速鉴别白矾、枯矾真伪[J]. 中国实验方剂学杂志, 2017, 23(19): 71-76.
- [31] 明 晶. 几种矿物药的 X 射线衍射、拉曼光谱及近红外光谱法鉴别研究[D]. 武汉: 湖北中医药大学, 2018.
- [32] 耿爱辉, 马艳梅, 李 敏, 等. 天青石的等温状态方程和拉曼光谱研究[J]. 原子与分子物理学报, 2009, 26(4): 753-756.
- [33] 马艳平, 刘英明, 张月芳, 等. 高温下石膏的拉曼光谱法研究[J]. 分析仪器, 2009(6): 45-47.
- [34] 张书迪, 陈玲玲, 陈芳芳, 等. 白蜂蜡和黄蜂蜡的拉曼光谱研究[J]. 香料香精化妆品, 2021(5): 34-39.
- [35] 孙 飞, 陈 雨, 王凯洋, 等. 基于红外光谱数据融合的姜半夏鉴别方法研究[J]. 北京中医药大学学报, 2019, 42(10): 862-868.
- [36] Lee L C, Liong C, Jemain A A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps[J]. *Analyst*, 2018, 143(15): 3526-3539.
- [37] 王 彬, 周靖惟, 张皇生, 等. 基于 XRD、FTIR、TG-DTA 结合化学计量学鉴别白矾与铵明矾[J]. 中国实验方剂学杂志, 2025, 31(13): 178-186.