ATR-FTIR 和 UV-Vis 结合数据融合策略鉴别滇黄精产地

张 娇[1,2]，王元忠[1]，杨维泽[1]，张金渝[1*]

1. 云南省农业科学院药用植物研究所，云南 昆明 650200
2. 云南中医药大学中药学院，云南 昆明 650500

摘 要 黄精药材品质优劣与基原植物产地环境因子密切相关，建立简单、快速且能够准确鉴别药材产地的方法对保证其质量可控及用药安全具有重要的理论意义和应用前景。研究中以云南、四川和广西9个产地的133份滇黄精 *Polygonatum kingianum* coll. et Hemsl 根茎为试验材料，采集衰减全反射-傅里叶变换红外光谱(ATR-FTIR)和紫外-可见光光谱(UV-Vis)数据预处理后分别建立单一光谱随机森林(Random forest，RF)模型；将 ATR-FTIR 与 UV-Vis 数据直接串联完成低级融合，提取两种光谱的主成分数(PCs)和潜在变量(LVs)以实现中级(中级融合$_{PCs}$和中级融合$_{LVs}$)和高级数据融合(高级融合$_{PCs}$和高级融合$_{LVs}$)，基于不同数据融合策略分别建立 RF 模型；比较不同模型的正确率(ACC)、灵敏度(SEN)和特异性(SPE)，筛选产地鉴别最佳模型。结果显示，不同产地滇黄精 ATR-FTIR 和 UV-Vis 峰型相似，吸光度略有差异，ATR-FTIR 显示14个共有峰，与糖类、甾体皂苷、黄酮类和生物碱类物质有关，其 UV-Vis 共有峰主要位于272及327 nm处，与黄酮类物质有关；ATR-FTIR、UV-Vis 和低级融合的 RF 模型，训练集和预测集 ACC 分别为(76.34%，95.00%)，(80.65%，95.00%)和(83.87%，100.00%)，但 SEN 和 SPE 值较低，故不宜采用；中级融合$_{PCs}$和中级融合$_{LVs}$的 RF 模型的 SEN 和 SPE 分别为大于0.91和0.98，训练集 ACC 分别为91.40%和97.85%，预测集 ACC 均为97.50%；高级融合$_{PCs}$和高级融合$_{LVs}$的 RF 训练集 ACC 分别为77.42%和97.85%，预测集 ACC 均为95.00%，高级融合$_{PCs}$的 RF 模型鉴别效果较差，高级融合$_{LVs}$的 RF 模型存在过拟合现象；模型鉴别能力为中级融合$_{LVs}$＞中级融合$_{PCs}$＞低级融合＞ UV-Vis＞ATR-FTIR＞高级融合$_{PCs}$；提取 LVs 对产地鉴别的方法优于 PCs；中级融合$_{LVs}$建立的 RF 模型鉴别 ACC 最高，SEN 和 SPE 大于0.98，模型性能最佳。该方法可为黄精药用资源的科学评价提供理论依据和技术支撑。

关键词 滇黄精；产地鉴别；数据融合；数据衰减全反射-傅里叶变换红外光谱；紫外-可见吸收光谱(UV-Vis)
中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2021)05-1410-07

引 言

滇黄精 *Polygonatum kingianum* coll. et Hemsl 是百合科(Liliaceae)黄精属药食同源植物，主要分布于我国云南、贵州、四川等西南地区和越南、缅甸、日本等国家[1]。其干燥根茎具有补气养阴、健脾、润肺和益肾之功效，是中药黄精的主要来源之一[2]。滇黄精主要药效成分是多糖和甾体皂苷，此外还含有黄酮、生物碱、氨基酸等成分，现代药理学研究表明其具有降血糖、抗衰老、抗肿瘤等作用[3]。调查发现，滇黄精栽培范围逐年扩大，不同产地的气候、土壤条件等均影响其药材质量。李婧等[4]以4种黄酮类成分含量为指标筛选影响其含量的环境因子，结果表明降水量、年平均温度、黏土量等环境因子对黄酮类成分影响最大。研究发现不同产地黄精中的多糖、薯蓣皂苷元[5]和挥发性成分[6]等均存在显著差异。为保证黄精质量的有效性和一致性，产地鉴别研究是其中的关键环节和重要前提条件。目前，可进行准确定量的色谱技术、液(气)质联用技术及电化学指纹图谱技术广泛应用于其产地溯源研究[7]，但这些技术存在操作复杂、成本高和耗时长等缺点，因此找到一种快速、简便且可靠的方法显得尤为重要。

衰减全反射-傅里叶变换红外-光谱(attenuated total re-

收稿日期：2020-05-03，**修订日期**：2020-08-12
基金项目：云南省重大科技专项(2018ZF010)，云南省科技计划项目(2017RA001)，中医药公共卫生服务补助专项国家重大项目“云南中药资源普查”第5批项目资助
作者简介：张 娇，女，1994年生，云南中医药大学中药学院硕士研究生 e-mail：jzhang2019@126.com
*通讯作者 e-mail：jyzhang2008@126.com

flection-Fourier transform infrared spectra，ATR-FTIR)和紫外-可见光光谱(ultraviolet-visible spectra，UV-Vis)技术具有方便、快速、无损等特点，已被广泛应用于食品与中药的产地鉴别。Zhao 等[8]利用 FTIR 技术和化学计量学方法鉴别滇龙胆产地，结果显示正确率达到 97.22%，为滇黄精产地鉴别提供参考。但单一指纹图谱通常不能全面反映样品化学信息，采用数据融合策略能够弥补此方面的不足。Yao 等[9]采用 FTIR 和 UV-Vis 技术对 7 个产地牛肝菌进行鉴别，训练集和预测集正确率(accuracy，ACC)分别为 80.18% 和 94.14%，使用中级数据融合策略后达到 99%。Wu 等[10]采集 ATR-FTIR 和 UV-Vis 信息结合高级数据融合策略鉴别 6 个产地野生滇重楼，分类正确率为 98.88%。由以上研究结果可知，数据融合策略可有效提高产地鉴别正确率，能够实现中药产地的快速、方便和无损鉴别。

本研究拟采集 9 个产地共 133 份滇黄精根茎样品的 ATR-FTIR 和 UV-Vis 光谱信息，经预处理及特征变量筛选后，建立单一(ATR-FTIR，UV-Vis)和数据融合(低级、中级和高级)随机森林(random forest，RF)模型，通过比较其灵敏度、特异性和分类正确率参数，最终确定快速鉴别滇黄精产地的最佳模型和方法，为其药用资源评价提供理论依据。

1　实验部分

1.1　材料

采集于云南、四川和广西 9 个产地的 133 份样品，由云南省农业科学院药用植物研究所张金渝研究员鉴定为滇黄精 *Polygonatum kingianum* coll. et Hemsl 的根茎[图 1(a，b，c)和表 1]。样品去除须根，用去离子水将附着的杂质和泥土清洗干净，切片，于 55 ℃烘箱中干燥至恒重。粉碎过筛(100 目)后保存于自封袋中备用。

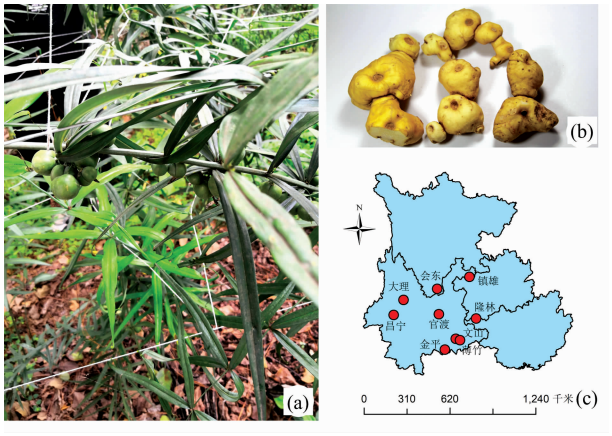

图 1　滇黄精样品和采集地图片

(a)：全株植物；(b)：根茎；(c)：样品采集地

Fig. 1　Collectionorigins of *P. kingianum* samples

(a)：The whole plant；(b)：Rhizome；(c)：Samples collection area

1.2　光谱采集

1.2.1　ATR-FTIR 采集

ATR-FTIR 光谱通过配备 ZnSe 衰减全反射附件及氘化硫酸三甘氨酸(DTGS)检测器的 FTIR 光谱仪(frontier perkin elmer，USA)采集。扫描范围为 4 000～550 cm^{-1}，扫描信号累加 16 次，分辨率为 4 cm^{-1}。每个样品重复 3 次，取平均光谱。

表 1　滇黄精样品信息

Table 1　Information of *P. kingianum* samples

产地	海拔/m	样本量	经纬度
云南省保山市昌宁县	1 698	14	N20°14′，E100°12′
云南省大理市挖色镇	1 968	12	N25°25′，E100°27′
云南省昆明市官渡区	1 901	13	N24°54′，E103°03′
四川省凉山州会东县	1 683	14	N26°12′，E103°03′
云南省红河州金平县	1 272	13	N23°26′，E101°47′
广西壮族自治区百色市隆林县	617	14	N24°22′，E105°41′
云南省文山州文山市	1 255	20	N23°40′，E103°51′
云南省文山州文山市薄竹镇	1 514	12	N23°06′，E103°27′
云南省昭通市镇雄县	1 705	21	N27°17′，E105°19′

1.2.2　UV-Vis 采集

UV-Vis 光谱通过配有积分球检测器的 UV2700 紫外-可见分光光度计(Shimadzu，Japan)采集。使用石英容器压片，压制成 1 mm 薄片进行光谱采集。样品测试前使用 $BaSO_4$ 进行背景扫描。扫描范围 220～850 nm，采样间隔为 1，狭缝宽度为 5.0 nm。每个样品重复 3 次，取平均光谱。

1.3　随机森林算法

RF 是以决策树为基础的有监督学习算法[11]。建模前使用 Kennard-Stone 算法[12]将每个产地样品的 2/3 划分为为训练集，1/3 划分为预测集。采用训练集样品建立模型，预测集样品用来验证模型的性能。建模时从原始训练集中使用自助法随机且有放回地取出 m 个样品，共进行 n 次取样，得到 n 个训练集并对每一个训练集训练，根据袋外-误差率(out-of-bag error，OOB)最小来选择最优的 ntree 棵决策树(classification and regression tree，CART)。在 CART 分类过程中没有进行剪枝处理。每个样品有 M 个变量，随机变量数(mtry)决定每棵树的分类性能，在建模过程中使用±10 来寻找最优 mtry。最后根据找到的最优参数 ntree 和 mtry 建立最终鉴别模型。通过集成多个 CART 的分类结果进行投票获得最后的分类结果，即使数据分布不平衡或有多个缺失值，也能提供稳定、准确度高的分类模型[13]。采用灵敏度(sensitivity，SEN)、特异性(specificity，SPE)和正确率 ACC 来衡量模型是否稳定。ACC，SEN 和 SPE 值越接近于 1，模型的性能越好。计算公式见式(1)和式(2)

$$\text{灵敏度} = \frac{\text{真阳性}}{\text{真阳性} + \text{假阴性}} \tag{1}$$

$$\text{特异性} = \frac{\text{真阴性}}{\text{真阴性} + \text{假阳性}} \tag{2}$$

1.4　数据融合

数据融合属于化学计量学方法之一，是将不同来源数据有效结合后再建立分类模型的一种策略[13]，通常分为低级、中级和高级融合。低级融合是指将不同来源的数据合并成一个新数据矩阵再建立分类模型。中级融合是分别提取单一光谱、色谱或者波谱的特征变量串联形成一个新数据矩阵来建

立分类模型。主成分数(principal components，PCs)、潜在变量(latent variables，LVs)、变量投影重要性等，是数据融合中常用的特征变量提取方法。高级融合是在中级融合的基础上，用特征变量分别建立单一模型，融合单一模型结果，根据模糊集合论的最大值(maximum，Max)、最小值(minimum，Min)、乘积(product，Pro)和平均值(average，Ave)进行投票得到最终结果。为了使数据处理方便在数据融合前进行归一化处理。

1.5 数据处理

通过 OMNIC 9 软件将 ATR-FTIR 透光率转化为吸光度。使用 SIMCA 14.1 软件对光谱数据进行预处理。用 ORIGIN 9.1 软件作图。R studio 软件用于建立 RF 模型。光谱易受噪音和样品性质的影响，对其进行适当的预处理是必要的。采用一阶导数(first derivative，FD)、二阶导数(second derivative，SD)和标准正态变量(standard normal variable，SNV)对光谱进行预处理，根据决定系数(determination coefficient，R^2)、交叉验证均方根误差(root mean square error of cross validation，RMSEcv)和校正均方根误差(root mean square error of estimation，RMSEE)及 ACC 来选择最佳预处理方式。UV-Vis 波长在 700～850 nm 范围内受噪音影响较大，在建立 RF 模型时去除这一波段。ATR-FTIR 在建模分析时去除 4 000～3 700 cm^{-1}的基线区、682～653 cm^{-1}的 CO_2 光谱区和 2 500～1 799 cm^{-1}的 ZnSe 晶体光谱区[14]。当 Q^2(模型预测能力)第一次达到最大值时，提取相应特征变量[15]。从 ATR-FTIR 的 133×1 775个变量中分别提取 133×17 个 PCs 和 133×13 个 LVs，从 UV-Vis 的 133×467 个变量中分别提取 133×9 个 PCs 和 133×5 个 LVs 用于建立模型。

2 结果与讨论

2.1 光谱分析

2.1.1 ATR-FTIR 分析

图 2(a)是 9 个产地滇黄精原始 ATR-FTIR 平均图。不同产地的 ATR-FTIR 有 14 个共有峰，波数分别为 3 338，2 924，2 845，1 735，1 625，1 416，1 362，1 316，1 262，1 115，1 024，928，873 和 824 cm^{-1}。共有峰的峰形和峰位相似，其差异较大的是吸光度值。云南镇雄县和四川会东县两个产地样品的吸光度明显高于其他 7 个产地。3 338 cm^{-1}表征—OH或者—NH—的伸缩振动；2 924 和 2 825 cm^{-1}由吡喃糖环中 C—H 的伸缩振动引起；1 735 cm^{-1}表征 C═O 的伸缩振动；1 625 cm^{-1}表征 C═C 和 C═N 的伸缩振动，与黄酮类、甾体皂苷和生物碱有关；1 416，1 362 和 1 316 cm^{-1}表征 C—H 或者—OH 的弯曲振动；1 262，1 115 和 1 024 cm^{-1}表征 C—O 的伸缩振动；928，873 和 824 cm^{-1}表征 ═CH 的弯曲振动[16-17]。

2.1.2 UV-Vis 分析

图 2(b)为 9 个产地滇黄精原始 UV-Vis 平均图。其特征峰波长为 272 和 327 nm，推测与滇黄精中黄酮类物质有关[18]。部分样品在 668 nm 处的可见光区存在吸收峰。整体而言，UV-Vis 吸收峰较少，主要反映芳香族和含有共轭体系的黄酮类物质信息。此外，产自大理的滇黄精 UV-Vis 吸光度次之，与 ATR-FTIR 显示产自镇雄的滇黄精吸光度结果不一致，因此使用不同性质和原理的指纹图谱来评价或鉴别滇黄精产地是必要的。

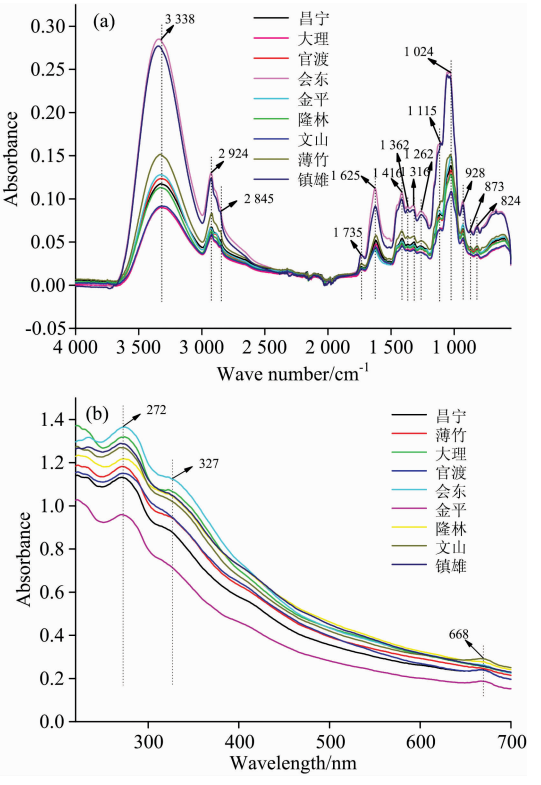

图 2 9 个产地滇黄精的平均光谱图

(a)：ATR-FTIR；(b)：UV-Vis

Fig. 2 Average spectra of 9 origins in *P. kingianum*

(a)：ATR-FTIR；(b)：UV-Vis

2.2 单一光谱鉴别分析

两种光谱经 FD，SD，SNV 预处理(表 2)，UV-Vis 在预处理(除 SD)后建模不成功，SD 为两种光谱的最佳预处理方式，鉴别能力较差，选择非线性的 RF 算法对滇黄精产地进行鉴别分析。ATR-FTIR 的 RF 模型最优 ntree 为 1 140，mtry 为 35。UV-Vis 的 RF 模型最优 ntree 为 1 041，mtry 为 42。结果如表 3 所示，ATR-FTIR 光谱结果显示训练集 ACC＝76.34%，预测集 ACC＝95.00%，训练集的 SEN 为 0.77(＜0.8)，训练模型时对样品识别能力较差，模型存在不稳健现象；UV-Vis 的 RF 模型 SEN 和 SPC 值分别为0.8和0.98，训练集ACC＝80.65%，预测集ACC＝95.00%，对产

表 2 单一光谱预处理结果

Table 2 Single spectral pretreatment results

光谱	预处理方法	R^2	RMSEE	RMSEcv	训练集 ACC/%
ATR-FTIR	FD	0.329	0.277	0.281	63.83
	SD	0.667	0.233	0.272	95.24
	SNV	0.452	0.266	0.274	75.53
UV-Vis	SD	0.307	0.253	0.274	72.34

地鉴别效果较差。采用数据融合策略建立 RF 模型对这 9 个产地的滇黄精进行鉴别。

2.3　数据融合

2.3.1　低级融合

将 ATR-FTIR 的 133×1775 个变量和 UV-Vis 的 133×467 个变量串联起来形成一个新的数据矩阵建立 RF 模型，其最优 ntree 和 mtry 值如图 3(a)所示。模型的灵敏度和特异性大于 0.83，训练集和预测集的正确率(表 3)分别为 83.87%和 100.00%，训练模型时对样品的识别能力较弱，表明光谱含有一些对产地鉴别冗余的波段，需要挖掘对产地鉴别有用的信息。

表 3　数据融合的结果

Table 3　The results of data fusion

数据融合	特征变量	训练集			预测集		
		SEN	SPC	ACC/%	SEN	SPC	ACC/%
ATR-FTIR	/	0.77	0.97	76.34	0.94	0.99	95.00
UV-Vis	/	0.80	0.98	80.65	0.95	0.99	95.00
低级	/	0.83	0.94	83.87	1.00	1.00	100.00
中级	PCs	0.91	0.99	91.40	0.98	1.00	97.50
	LVs	0.98	1.00	97.85	0.98	1.00	97.50
高级	PCs	0.80	0.97	77.42	0.95	0.89	95.00
	LVs	0.96	1.00	97.85	0.93	0.99	95.00

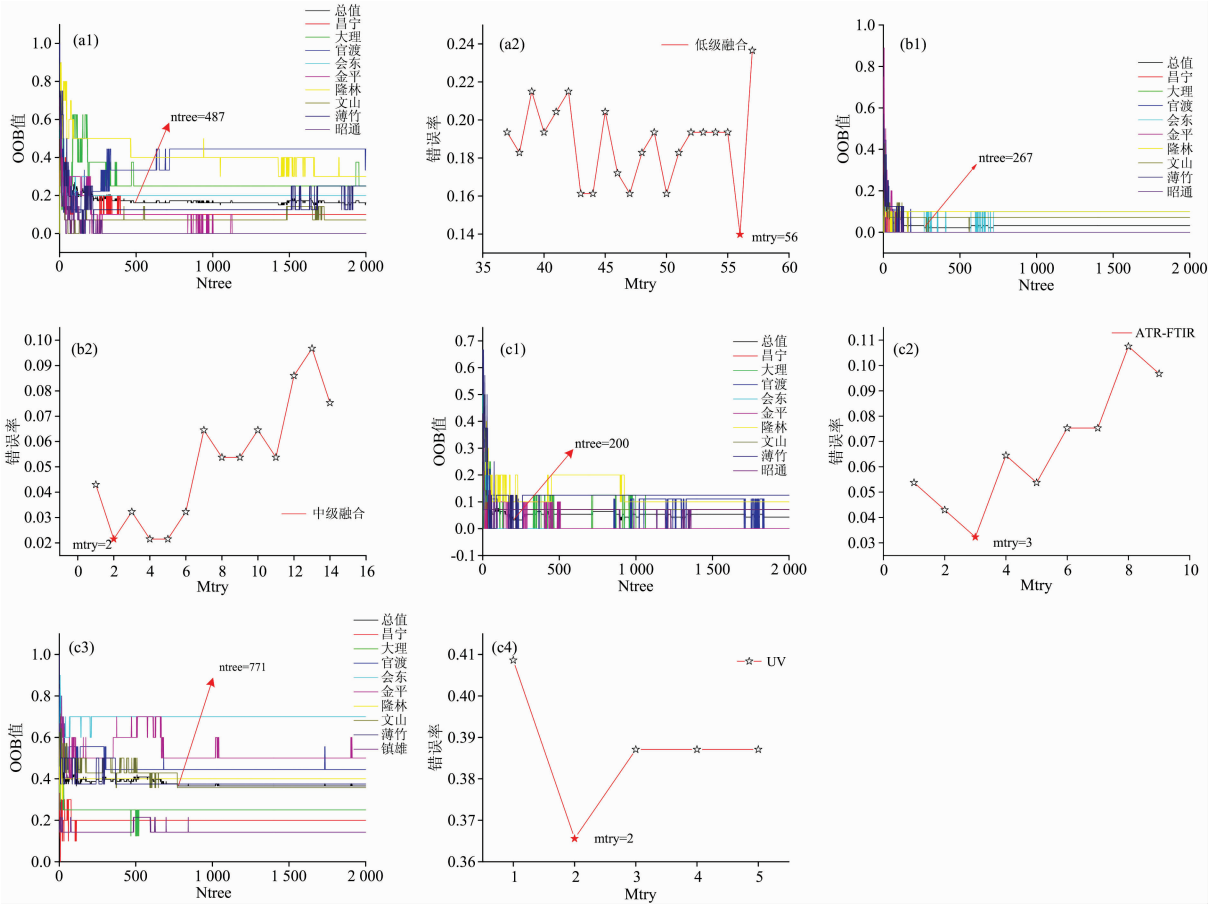

图 3　数据融合的最佳 ntree 和 mtry 值

(a)：低级融合；(b)：中级融合；(c)：高级融合

(a1)：ntree 值，(a2)：mtry 值；(b1)：ntree 值，(b2)：mtry 值；(c1)：ATR-FTIR 的 ntree 值，(c2)：ATR-FTIR 的 mtry 值，(c3)：UV-Vis 的 ntree 值，(c4)：UV-Vis 的 mtry 值

Fig. 3　Optimal ntree and mtry values for data fusion

(a)：Low-level data fusion；(b)：Mid-level data fusion；(c)：High-level data fusion

(a1)：ntree，(a2)：mtry；(b1)：ntree，(b2)：mtry；(c1)：the ntree values of ATR-FTIR；(c2)：the mtry values of ATR-FTIR；(c3)：the ntree values of UV-Vis；(c4)：the mtry values of UV-Vis

2.3.2　中级融合

在中级融合中，使用 PCs 和 LVs 来建立 RF 产地鉴别模型，比较两种特征变量融合对产地鉴别的能力。提取特征变量结果如图 4 所示，ATR-FTIR 的 133×17 个 PCs 和 UV-Vis 的 133×9 个 PCs 被提取建立 RF 模型，ATR-FTIR 的 133×13 个 LVs 和 UV-Vis 的 133×5 个 LVs 提取建立 RF 模型。结果如表 3 所示，LVs 建立中级融合(中级融合$_{LVs}$)的 RF 模型中 3 个样品被分类错误；PCs 建立中级融合(中级融合$_{PCs}$)的 RF 模型中 8 个样品被分类错误。基于 LVs 建立的中级融合 RF 模型训练集和预测集的灵敏度、特异性和 ACC

均高于 PCs 结果，其分类正确率均大于 97.85%，因此在中级融合中 LVs 选择为产地鉴别的特征变量。从原始的 133×2 242个降到 133×17 个变量，明显缩短模型拟合时间，提高了产地鉴别能力。LVs 建立的 RF 模型的参数优化如图 3(b1,b2)所示，最优的 ntree 为 267，mtry 为 2。

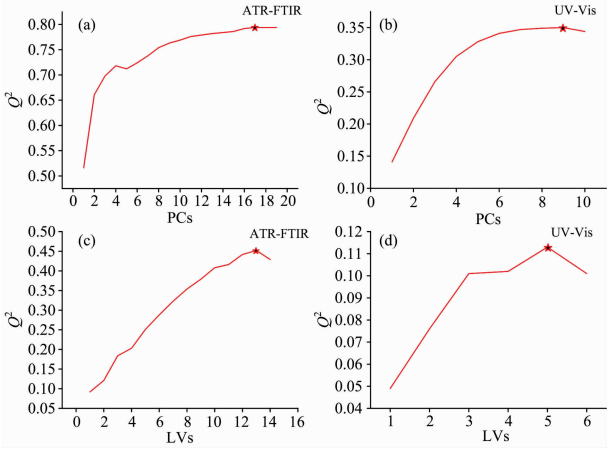

图 4 特征变量提取结果

(a)：ATR-FTIR 的主成分数；(b)：UV-Vis 的主成分数；
(c)：ATR-FTIR 的潜在变量数；(d)：UV-Vis 的潜在变量数

Fig. 4 Feature variable selection results

(a)：PCs of ATR-FTIR；(b)：PCs of UV-Vis；
(c)：LVs of ATR-FTIR；(d)：LVs of UV-Vis

2.3.3 高级融合

高级融合的结果如表 4 所示，PCs 的高级融合 RF 模型(高级融合$_{PCs}$)训练集和预测集的 SEN，SPC 和 ACC 分别为 0.80，0.97，77.42%和 0.95，0.89 和 95.00%，其鉴别能力较差。LVs 的高级融合(高级融合$_{LVs}$)RF 模型对 9 个产地鉴别进行鉴别，其 RF 模型最优 ntree 和 mtry 结果如图 3(c)所示，图 3(c1)和(c2)是 ATR-FTIR 的 RF 模型最优 ntree 和 mtry 结果，图 3(c3)和(c4)是 UV-Vis 的 RF 模型最优 ntree 和 mtry 结果。在高级融合中的 133 个样品中有 42 个样品需根据 CART 进行投票。42 个样品中有 37 个样品经投票后分类正确，有 1 个样品分类出现分歧(No. 74，文山薄竹)，其余 4 个样品分类错误。分类错误及分歧样品的投票结果如表 4 所示。No. 74 样品被 UV-Vis 分类到 Class4，被 ATR-FT-IR 分类到 Class8，最后获得相同票数。4 个被误分的样品中有 2 个(No. 21、106)是 ATR-FTIR 投票结果正确，而 UV-Vis 投票错误导致分类错误，1 个(No. 51)是 ATR-FTIR 投票错误，UV-Vis 投票正确，最终被分类错误，剩余 1 个(No. 122)样品的 ATR-FTIR 和 UV-Vis 投票结果均错误。高级融合$_{LVs}$和中级融合$_{LVs}$鉴别能力较好，训练集和预测集的 SEN 和 SPC 均高于 0.93，其鉴别能力比低级融合和单光谱的鉴别能力增强，但高级融合$_{LVs}$模型存在过拟合现象。中级和高级融合结果表明：中级融合$_{LVs}$建立不同产地滇黄精鉴别模型，其训练集 ACC 为 97.85%，预测集 ACC 为 97.50%，鉴别能力最好。

表 4 高级数据融合分类错误的样品投票结果

Table 4 Voting results of misclassified samples in high-level data fusion

样品编号		Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	结果
	ATR-FTIR	0.057	0.092	**0.241**	0.011	0.195	0.115	0.092	0.069	0.126	
	UV-Vis	0.003	0.000	0.091	0.080	0.038	0.045	0.359	0.010	**0.373**	
No. 21	Max	0.057	0.092	0.241	0.080	0.195	0.115	0.359	0.069	**0.373**	
Class3	Min	0.003	0.000	0.091	0.011	0.038	0.045	0.092	0.010	**0.126**	
	Pro	0.000	0.000	0.022	0.001	0.007	0.005	0.033	0.001	**0.047**	
	Ave	0.030	0.046	0.166	0.046	0.117	0.080	0.225	0.040	**0.250**	Class9
	ATR-FTIR	0.019	0.092	0.091	0.060	**0.537**	0.119	0.045	0.119	0.045	
	UV-Vis	0.015	0.000	0.042	0.130	0.123	**0.487**	0.042	0.146	0.003	
No. 51	Max	0.019	0.092	0.042	0.130	**0.537**	0.487	0.045	0.146	0.045	
Class6	Min	0.015	0.000	0.091	0.060	**0.123**	0.119	0.042	0.119	0.003	
	Pro	0.000	0.000	0.001	0.008	**0.066**	0.058	0.002	0.017	0.000	
	Ave	0.017	0.026	0.029	0.095	**0.330**	0.303	0.043	0.133	0.024	Class5
	ATR-FTIR	0.245	0.196	0.045	0.020	0.057	0.049	0.024	**0.286**	0.078	
	UV	0.014	0.000	0.003	**0.734**	0.021	0.000	0.021	0.204	0.003	
No. 74	Max	0.245	0.196	0.045	**0.734**	0.057	0.049	0.024	0.286	0.078	
Class8	Min	0.014	0.000	0.003	0.020	0.021	0.000	0.021	**0.204**	0.030	
	Pro	0.003	0.000	0.000	0.015	0.001	0.000	0.001	**0.058**	0.000	
	Ave	0.129	0.098	0.024	**0.377**	0.039	0.024	0.023	0.245	0.041	Class4，8
	ATR-FTIR	0.020	0.035	0.010	**0.675**	0.000	0.050	0.090	0.000	0.120	
	UV-Vis	0.034	0.000	0.122	0.027	0.030	**0.752**	0.025	0.038	0.000	
No. 106	Max	0.034	0.035	0.122	0.675	0.003	**0.752**	0.090	0.038	0.120	
Class4	Min	0.020	0.000	0.010	0.027	0.000	**0.050**	0.025	0.000	0.000	
	Pro	0.001	0.000	0.001	0.018	0.000	**0.038**	0.002	0.000	0.000	
	Ave	0.027	0.018	0.066	0.351	0.001	**0.401**	0.057	0.019	0.060	Class6

续表 4

	ATR-FTIR	0.125	**0.360**	0.185	0.069	0.025	0.015	0.120	0.147	0.070	
	UV-Vis	**0.328**	0.071	0.083	0.250	0.069	0.009	0.043	0.075	0.000	
No. 122	Max	**0.328**	0.360	0.185	0.250	0.069	0.015	0.120	0.147	0.070	
Class7	Min	**0.125**	0.071	0.083	0.069	0.025	0.009	0.043	0.075	0.000	
	Pro	**0.041**	0.026	0.015	0.006	0.002	0.000	0.005	0.011	0.000	
	Ave	**0.227**	0.216	0.134	0.138	0.047	0.012	0.081	0.111	0.035	Class1

滇黄精的 UV-Vis 光谱在紫外-可见光区吸收峰是芳香族和含有共轭体系黄酮类成分的化学信息，其 ATR-FTIR 光谱的吸收峰显示的是官能团和化学键信息。两种指纹图谱反映不同的化学成分信息，融合两种光谱的化学信息可以更加全面的反映其化学信息，可对滇黄精实现更加全面的质量评价。

3　结　论

探讨了 ATR-FTIR 和 UV-Vis 及数据融合策略结合 RF 算法对 9 个产地滇黄精鉴别的可行性。通过两种光谱对滇黄精产地鉴别分析表明，单一光谱对产地评价不够全面，可以利用数据融合策略来弥补不足，提取光谱的两种特征值结合 RF 方法提高了对产地的鉴别效果。采用 SEN 和 SPE 和模型分类正确率筛选出最佳模型，其鉴别能力为中级融合$_{LVs}$＞中级融合$_{PCs}$＞低级融合＞ UV-Vis ＞ ATR-FTIR＞高级融合$_{PCs}$；提取 LVs 对产地鉴别的方法优于 PCs；中级融合$_{LVs}$建立的 RF 模型分类正确率最高，SEN 和 SPE 大于 0.98，模型性能最佳，为黄精药用资源的科学评价提供理论依据和技术支撑，同时为其它中药材鉴别新方法的建立有借鉴作用。

References

[1] Zhao P, Zhao C, Li X, et al. Journal of Ethnopharmacology, 2018, 214: 274.

[2] Chinese Pharmacopoeia Commission(国家药典委员会). Pharmacopoeia of the People's Republic of China(中华人民共和国药典), Part One(第一部). Beijing: China Medical Science Press(北京：中国医药科学出版社), 2015. 306.

[3] ZHANG Jiao, WANG Yuan-zhong, YANG Wei-ze, et al(张　娇，王元忠，杨维泽，等). China Journal of Chinese Materia Medica(中国中药杂志), 2019, 44(10): 1989.

[4] LI Jing, WANG Ying-zhe, LIU Yu-cui, et al(李　婧，王英哲，刘玉翠，等). China Journal of Chinese Materia Medica(中国中药杂志), 2019, 44(24): 5368.

[5] JIAO Ji, CHEN Li-ming, SUN Rui-ze, et al(焦　劼，陈黎明，孙瑞泽，等). Journal of Chinese Medicinal Materials(中药材), 2016, 39(3): 519.

[6] CHEN Long-sheng, DU Li-ji, CHEN Shi-jin, et al(陈龙胜，杜李继，陈世金，等). Journal of Chinese Medicinal Materials(中药材), 2018, 41(4): 894.

[7] Zhou Y H, Zuo Z T, Xu F R, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 226: 117619.

[8] Zhao Y L, Yuan T J, Zhang J, et al. Journal of Chemometrics, 2019, 33(4): e3115.

[9] Yao S, Li T, Li J Q, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 198: 257.

[10] Wu X M, Zhang Q Z, Wang Y Z. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 205: 479.

[11] Qiu S, Wang J. Food Chemistry, 2017, 230: 208.

[12] Wang Y, Zuo Z T, Huang H Y, et al. Royal Society Open Science, 2019, 6(5): 190399.

[13] Hou L, Liu Y, Wei A. Industrial Crops and Products, 2019, 134: 146.

[14] Rodríguez S D, Rolandelli G, Buera M P. Food Chemistry, 2019, 274: 392.

[15] Chen H, Tan C, Lin Z, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 189: 183.

[16] SUN Yu-qing, HU Yu-zhu, DU Ying-xiang, et al(孙毓庆，胡育筑，杜迎翔，等). Analytical Chemistry(分析化学). 3rd edition(第 3 版). Beijing: Science Press(北京：科学出版社), 2011. 136.

[17] Pei Y F, Wu L H, Zhang Q Z, et al. Analytical Methods, 2019, 11(1): 113.

[18] Yang Y G, Zhao Y L, Zuo Z T, et al. Journal of AOAC International, 2019, 102(2): 457.

Data Fusion of ATR-FTIR and UV-Vis Spectra to Identify the Origin of *Polygonatum Kingianum*

ZHANG Jiao[1,2], WANG Yuan-zhong[1], YANG Wei-ze[1], ZHANG Jin-yu[1*]

1. Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650200, China

2. College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming 650500, China

Abstract The quality of Polygonati Rhizoma medicinal materials is closely related to the original plants' origin environment. It is necessary to ensure their quality control and drug safety by establishing a simple, rapid and accurate origin identification method for the medicinal materials. In this study, the attenuated total Reflection-Fourier transform infrared (ATR-FTIR) spectra and ultraviolet visible (UV-Vis) spectra of 133 *Polygonatum kingianum* rhizomes from 9 geographic origins in Yunnan, Sichuan and Guangxi Provinces were collected to establish random forest (RF) modelafter data pretreatment, respectively. ATR-FTIR and UV-Vis spectra data were directly connected in series to complete the RF model of low-level data fusion. Principal components (PCs) and latent variables (LVs) of the two spectra were extracted to achieve RF model ofmid-level (mid-PCs and mid-LVs) and high-level (high-PCs and high-LVs) data fusion. The accuracy (ACC), sensitivity (SEN) and specificity (SPE) of different models were compared to select the best model for origin identification. The results showed that the peaks of ATR-FTIR and UV-Visspectrain *P. kingianum* were similar, and their absorbance were different. There were 14 common peaks in ATR-FTIR spectra of *P. kingianum*, which were related to carbohydrate, steroidal saponins, flavonoids and alkaloids. The common peaks of UV-Visspectra in *P. kingianum* were mainly at 272 and 327 nm, which were related to flavonoids. For the RF models of ATR-FTIR, UV-Vis and low-level fusion, the ACC of the training set and prediction set were respectively (76.34%, 95.00%), (80.65%, 95.00%) and (83.87%, 100.00%), however, the SEN and SPE values were so low that they were not suitable to use. The SEN and SPE of mid-PCs and mid-LVs RF models were greater than 0.91 and 0.98, respectively. The ACC of the training set was 91.40% and 97.85%, respectively, and that of the prediction set both were 97.50%. The ACC of RF training set with high-PCs and high-LVs was 77.42% and 97.85%, respectively, and the prediction set ACC both were 95.00%. The RF model with high-PCs has poor identification effect, and the RF model with high-LVs was over-fitted. In summary, the identification of model from high to low was: mid-LVs > mid-PCs > low fusion > UV-Vis > ATR-FTIR > high-PCs. LVs extraction method is better than PCs for origin identification. RF model of mid-LVs established has the highest ACC with the best model performance, and the SEN and SPE greater than 0.98, and, which can provide a theoretical basis for the scientific evaluation of medicinal resources of Polygonati Rhizoma.

Keywords *Polygonatum kingianum*; Origin identification; Data fusion; ATR-FTIR; UV-Vis

(Received May 3, 2020; accepted Aug. 12, 2020)

* Corresponding author