

· 资源与质量评价 ·

红外光谱结合化学计量学鉴别獐牙菜属植物

于叶霞^{1,2}, 李鹂^{1*}, 王元忠^{2*}

(1. 吉首大学 植物资源保护与利用湖南省高校重点实验室, 湖南 吉首 416000;
2. 云南省农业科学院 药用植物研究所, 昆明 650200)

[摘要] **目的:** 采用傅里叶变换红外光谱(FTIR)结合化学计量学实现獐牙菜属植物快速、准确鉴别。**方法:** 采集川东獐牙菜、青叶胆、紫红獐牙菜、狭叶獐牙菜和西南獐牙菜不同部位(根、茎、叶)543份样品红外光谱信息,原始数据经标准正态变量(SNV),多元散射校正(MSC),平滑(SG),一阶导数(1D),二阶导数(2D),三阶导数(3D)等处理后删减4 000~3 700,2 799~1 800 cm⁻¹和682~653 cm⁻¹波段,建立偏最小二乘判别分析(PLS-DA)和支持向量机(SVM)模型。**结果:** 5种獐牙菜属植物相同部位平均光谱较为相似,无法区分,同一物种不同部位光谱特征峰有差异,复杂程度为叶>茎>根;根、茎和叶3个部位PLS-DA和SVM模型均能准确鉴别5种獐牙菜属植物,且MSC+SG+2D预处理效果最佳。PLS-DA模型R²Y>0.8, RMSEP<RMSECV,所建模型稳定,效果好,Q²超过0.6,预测集正确率达到100%,所有预测集样品分类正确,模型预测能力强。根、茎和叶SVM模型最优惩罚参数c分别为22.627 4,2和1.414 2,均在正常范围内,预测集正确率均为100%,分类准确率高。**结论:** FTIR结合PLS-DA和SVM模型能准确区分不同獐牙菜属植物,模型预测效果好,为其他植物鉴别提供一定的参考依据。

[关键词] 獐牙菜属; 川东獐牙菜; 青叶胆; 紫红獐牙菜; 狭叶獐牙菜; 西南獐牙菜

[中图分类号] R289;R284.1;R22;R2-031 **[文献标识码]** A **[文章编号]** 1005-9903(2019)20-0114-07

[doi] 10.13422/j.cnki.syfjx.20192014

[网络出版地址] <http://kns.cnki.net/kcms/detail/11.3495.r.20190711.0847.001.html>

[网络出版时间] 2019-07-15 14:24

Discrimination of Different Species in *Swertia* Using FTIR Combined with Chemometrics

YU Ye-xia^{1,2}, LI Li^{1*}, WANG Yuan-zhong^{2*}

(1. Key Laboratory of Plant Resources Conservation and Utilization of Hunan Province, Jishou University, Jishou 416000, China;

2. Institute of Medicinal Plants, Yunnan Academy of Agricultural Sciences, Kunming 650200, China)

[Abstract] **Objective:** To realize the rapid and accurate discrimination of *Swertia* plants by Fourier transform infrared spectroscopy (FTIR) and chemometrics. **Method:** The original infrared spectra data from different parts (roots, stems, leaves) of all of the 543 samples of *S. davidii*, *S. mileensis*, *S. punicea*, *S. angustifolia* and *S. cincta* were collected and preprocessed by multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky-Golay filter (SG), first derivative (1D), second derivative (2D), third derivative (3D). Then, the spectral ranges of 4 000-3 700, 2 799-1 800 cm⁻¹ and 682-653 cm⁻¹ were deleted before PLS-DA and SVM analysis. **Result:** The samples of the five species could not be distinguished with similar averaged infrared spectra in the same part. The characteristic peaks of different parts in the same species were different, and the sequence of complexity was leaves > stems > roots. The five species of *Swertia* could accurately be

[收稿日期] 20190509(020)

[基金项目] 国家自然科学基金项目(81760695,31260102)

[第一作者] 于叶霞,在读硕士,从事植物生态学研究,E-mail:yyx921583014@163.com

[通信作者] *李鹂,博士,教授,从事药用植物学研究,E-mail:lilyjsu@126.com;

*王元忠,博士,副研究员,从事药用植物资源学研究,Tel: 0871-65033575,E-mail:boletus@126.com

identified by PLS-DA and SVM models established by spectra data in roots, stems and leaves. MSC + SG + 2D showed the best preprocessing effect, and the prediction accuracies of all models were 100%. The values of $R^2 Y$ in PLS-DA of all of the parts were more than 0.8, and the RMSEP was less than RMSECV, indicating that the model was stable and more effective. Furthermore, the value of Q^2 exceeded 0.6, and the accuracy of prediction set reached 100%, indicating a high classification accuracy. It showed that PLS-DA models had a strong prediction ability. The c values in SVM model of roots, stems and leaves were 22.627 4, 2 and 1.414 2, respectively, which were all within the normal ranges. The accuracy of prediction set was 100%, suggesting a high accuracy.

Conclusion: FTIR combined with PLS-DA and SVM could accurately distinguish different species from *Swertia*, and the model has a good prediction effect and provides certain reference for the identification of other plants.

[Key words] *Swertia*; *S. davidii*; *S. mileensis*; *S. punicea*; *S. angustifolia*; *S. cincta*

中药材在我国使用历史悠久,不仅可以用于预防、治疗疾病,部分药材还可美容保健。中药材有效成分是其发挥功效的物质基础,易受种类、产地、采收期等多种因素影响,致使中药材质量不稳定,真假难辨,妨碍其进一步开发和利用^[1-2]。因此,对中药材进行定性鉴别是实现中医药现代化和国际化发展的基础。

龙胆科 (Gentianaceae) 獐牙菜属 (*Swertia*) 为一年生或多年生草本植物,常以干燥全草入药,世界上已报道约 170 种,主要分布在亚洲、非洲和北美洲;我国有 79 种,主要分布在西南山岳地区^[3]。现代植物化学与药理学研究表明,獐牙菜属植物主要包括环烯醚萜类、三萜类、吡啶类、黄酮类、生物碱等,具有保肝、降血糖、抗氧化、抗菌、抗癌等作用^[4]。川东獐牙菜 (*S. davidii*) 性寒、味极苦,具有清热解毒、利胆健胃的功效。在湘西土家族地区,当地医生常用其治疗黄疸、恶疮疥癣、急性肠炎、菌痢等病。1975 年收入《中国药典》的青叶胆 (*S. mileensis*) 对急性黄疸型肝炎有显著效果,是“青叶胆片”“肝复康”等保肝药主要原材料。“布什都补”为紫红獐牙菜 (*S. punicea*) 彝语音译药名,常用于治疗肝炎、胆囊炎、内外出血、百日咳等^[5]。狭叶獐牙菜 (*S. angustifolia*) 在少数民族地区常用于治疗肝炎、胆囊炎。西南獐牙菜 (*S. cincta*) 主要用于黄疸型肝炎、风热感冒、咽喉肿痛、烧烫伤治疗^[6]。

随着对獐牙菜属植物市场需求的不断增加,药材混杂、以假充数现象时有发生。由于不同物种化学成分种类与含量不同,混淆用药可能会导致药效下降甚至一定程度上损害健康。因此,物种鉴别是保证正确用药的前提。传统鉴别方法主要通过植物形态和性状鉴定^[7],易受主观因素干扰,加之同属植物表型上具有一定重叠性,制成干品后外观更是相似,难以分辨。现代鉴别方法主要包括分子鉴

别^[8]、色谱鉴别^[9]与光谱鉴别。其中,分子鉴别价格昂贵,仪器操作复杂,专业性强、费力耗财,不利于推广。目前色谱鉴别广泛应用于中药材种类鉴别研究,如狄准等^[10]采用高效液相色谱技术鉴别川东獐牙菜、狭叶獐牙菜和紫红獐牙菜。LI 等^[11]利用超高效液相色谱结合紫外光谱对青叶胆及同属植物进行研究,结果显示,狭叶獐牙菜、川东獐牙菜和紫红獐牙菜化学组成与青叶胆差异较大,不能用作青叶胆替代品。色谱技术需对中药材进行预处理,所需试剂价格昂贵、污染环境,分析时间长,效率低。

傅里叶变换红外光谱 (Fourier transform infrared spectroscopy, FTIR) 反映分子转动和某些官能团振动信息,用于分析样品间物质组分差异,具有操作简单、用样量少、样品无损、能全面反映植物内部分子结构信息等特点^[12],广泛用于鉴别植物种类、产地、部位、采收期等。YANG 等^[13]通过 FTIR 结合偏最小二乘回归和支持向量机回归分析准确鉴别 5 种重楼属植物。WANG 等^[14]采集了 3 个省份滇龙胆不同部位 (根、茎、叶、花) 红外光谱图,经自动基线校正、二阶导数、平滑等预处理后建立支持向量机和偏最小二乘判别模型,结果表明此方法可有效鉴别滇龙胆来源。吴喆等^[15]通过 FTIR 结合偏最小二乘判别和偏最小二乘回归对不同采收期滇重楼进行定性鉴别与定量分析,结果显示偏最小二乘判别模型可准确区分不同采收期滇重楼样品。目前,利用红外光谱鉴别獐牙菜属不同物种研究未见系统报道。

本研究采集了青叶胆、川东獐牙菜、紫红獐牙菜、狭叶獐牙菜与西南獐牙菜不同部位 (根、茎、叶) 543 份野生样品红外光谱信息,通过标准正态变量 (standard normal variate, SNV), 多元散射校正 (multiplicative signal correction, MSC), 平滑 (Savitzky-Golay filter, SG), 一阶导数 (first derivative, 1D), 二阶导数 (second derivative, 2D), 三

阶导数 (third derivative, 3D) 等方法进行对原始光谱预处理, 采用偏最小二乘判别分析 (partial least squares discriminant analysis, PLS-DA) 和支持向量机 (support vector machine, SVM) 等化学计量学方法, 建立不同种类獐牙菜属植物快速鉴别方法, 以期对獐牙菜属植物质量评价提供理论基础。

1 材料

Frontier 型傅里叶变换红外光谱仪 (配备 DTGS 检测器和 ATR 附件, ZnSe 晶体材料, 美国珀金埃尔默); CH936B 型除湿机 (森井电气公司); DFT-50A 型高速粉碎机 (温岭市林大机械有限公司); 100 目标准筛盘 (浙江上虞市道墟五四仪器厂)。

实验采集 5 种獐牙菜属植物不同部位 (根、茎、叶) 野生样品共 543 份, 经吉首大学生物资源与环境科学学院李鹏教授鉴定为龙胆科狭叶獐牙菜 *S. angustifolia*, 西南獐牙菜 *S. cincta*, 川东獐牙菜 *S. davidii*, 青叶胆 *S. mileensis*, 紫红獐牙菜 *S. punicea*, 样品信息见表 1。

表 1 獐牙菜属植物样品信息

Table 1 Information of plant samples *Swertia*

物种	编号	地点	部位/份		
			根	茎	叶
狭叶獐牙菜	Sa	贵州省	11	11	11
西南獐牙菜	Sc	云南省	12	12	12
川东獐牙菜	Sd	重庆市、湖南省、湖北省、四川省	57	60	60
青叶胆	Sm	云南省	61	86	87
紫红獐牙菜	Sp	贵州省	20	22	21

2 方法

2.1 样品制备 新鲜植株材料洗净根茎部, 分成根、茎和叶 3 个部位, 45 °C 烘干至恒重, 粉碎, 过 100 目筛, 于自封袋避光保存, 备用。

2.2 数据采集 室内空气干燥的情况下 (室温 25 °C), 取样品粉末于洁净且干燥的 ZnSe 晶体材料上, 采集红外光谱图。光谱采集条件, 扫描范围 4 000 ~ 550 cm^{-1} , 分辨率 4 cm^{-1} , 扫描次数 16 次, 压力范围 129 ~ 132 Pa。

2.3 数据处理 中红外 4 000 ~ 3 700 cm^{-1} 波段为基线, 2 799 ~ 1 800 cm^{-1} 波段为晶体光谱区, 682 ~ 653 cm^{-1} 波段为二氧化碳光谱区。这 3 个波段未提供样品相关光谱信息, 建模时删除^[16]。

FTIR 采集光谱过程中易受样品颗粒不均、光散射和仪器噪音等无关变量干扰。为减少或消除以上

影响, 常用多元散射校正、标准正态变量、平滑、求导等预处理对原始光谱进行优化。多元散射校正用于消除样品研磨不均匀或颗粒分布不均产生的散射干扰和基线漂移影响; 标准正态变量计算光谱总变量标准偏差, 通过该标准偏差值将所有光谱标准化, 从而用于消除斜率变化和校正散射影响; 平滑能减少高频噪音干扰和取样量不同的影响; 光谱求导能减少基线漂移影响, 放大细微差异以有效区分重叠峰和弱峰, 提高分辨率^[17-19]。

ORIGIN 2017 软件做平均光谱图。SIMCA 13.0 软件用于对獐牙菜属植物原始光谱数据进行 MSC, SNV, SG, 1D, 2D, 3D 等预处理和建立 PLS-DA 模型。MATLAB 2017a 软件进行 Kennard-Stone (KS) 算法, 建立 SVM 模型。SIMCA 13.0 软件建立 PLS-DA 模型。KS 算法用于训练集与预测集 (2:1) 的划分, 359 份样品数据作为训练集用于建立 SVM 和 PLS-DA 模型, 进行训练集内部交叉验证, 184 份预测集样品对模型预测能力进行外部验证, 见表 2。

表 2 训练集与预测集分类信息

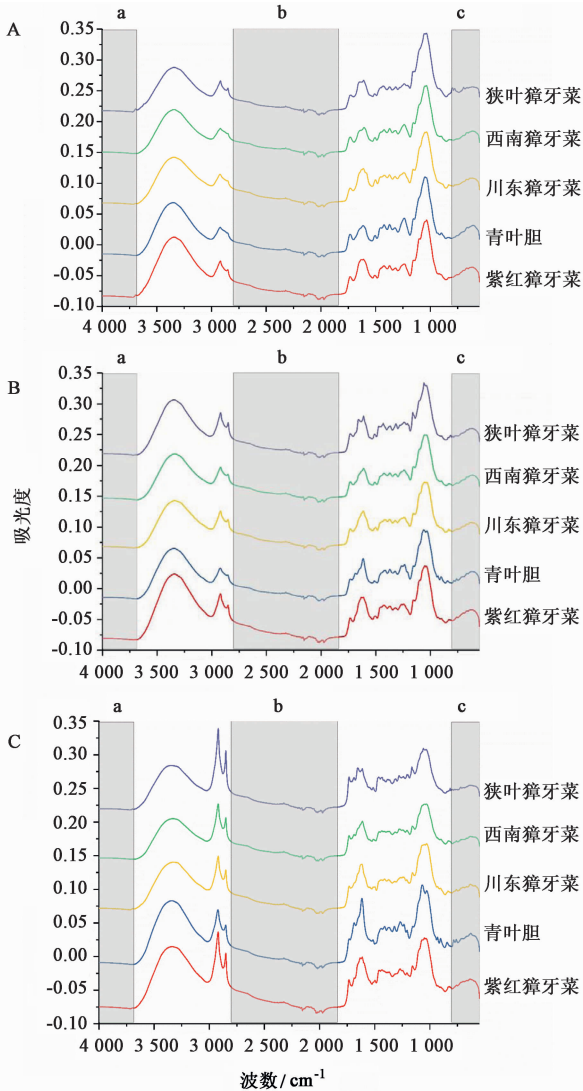
Table 2 Information of category for calibration set and prediction set

编号	根		茎		叶	
	训练集	预测集	训练集	预测集	训练集	预测集
Sa	7	4	7	4	7	4
Sc	8	4	8	4	8	4
Sd	38	19	40	20	40	20
Sm	40	21	57	29	58	29
Sp	13	7	14	8	14	7

3 结果与讨论

3.1 5 种獐牙菜属植物红外光谱分析 图 1 为 5 种獐牙菜属植物根、茎和叶的平均光谱图, 阴影部分未提供样品相关光谱信息, 1 700 ~ 700 cm^{-1} 波段为獐牙菜属特征吸收峰指纹区, 吸收峰较密集。主要吸收峰在 3 340, 2 920, 2 850, 1 735, 1 651, 1 614, 1 506, 1 234, 1 160, 1 027, 899 cm^{-1} 左右。3 340 cm^{-1} 附近主要为羟基 O-H 伸缩振动吸收峰, 2 920 cm^{-1} 附近主要为亚甲基 C-H 反对称伸缩振动吸收峰, 2 850 cm^{-1} 附近主要为亚甲基 C-H 对称伸缩振动吸收峰, 1 735 cm^{-1} 附近为酯类羰基 C=O 伸缩振动吸收峰^[20], 1 651 cm^{-1} 附近为酰胺 I 带 C=O 伸缩振动吸收峰, 1 614 cm^{-1} 附近为苯环 C=C 骨架振动吸收峰, 为龙胆苦苷主要吸收峰^[21], 1 506 cm^{-1} 附近为单核芳烃 C=C 伸缩振动吸收峰,

1 234 cm^{-1} 附近为 C-C 伸缩振动吸收峰或 C-H 面内弯曲振动吸收峰, 1 200 ~ 950 cm^{-1} 主要为皂苷等糖类物质振动吸收区^[22], 950 ~ 700 cm^{-1} 主要为糖类异构体的特征峰。光谱特征综合反映獐牙菜属植物化学成分复杂多样, 包含糖类、萜类、酯类等物质。



A. 根; B. 茎; C. 叶; a. 基线波段 (4 000 ~ 3 700 cm^{-1}); b. 晶体光谱区 (2 799 ~ 1 800 cm^{-1}); c. 二氧化碳光谱区 (682 ~ 653 cm^{-1})

图 1 5 种獐牙菜属植物不同部位样品原始平均 FTIR 光谱

Fig. 1 Original FTIR averaged infrared spectra of samples from five *Swertia* plants with different parts

比较 5 种獐牙菜属植物平均光谱图, 发现茎、叶光谱吸收峰数目多于根, 且部分吸收峰在峰形、峰强与峰位上有明显差异。在 2 920 cm^{-1} 和 2 850 cm^{-1} 附近与 1 700 ~ 1 200 cm^{-1} , 茎与叶吸收峰峰数与峰强均高于根, 1 200 ~ 700 cm^{-1} 吸收峰强度则表现为根 > 茎 > 叶, 推测与獐牙菜属植物自身遗传因子调

控有关。川东獐牙菜、青叶胆、狭叶獐牙菜、紫红獐牙菜和西南獐牙菜红外光谱较为相似, 但吸收峰强度存在一定差异。狭叶獐牙菜的根、茎和叶在 2 920 cm^{-1} 和 2 850 cm^{-1} 附近吸收峰强度最高, 紫红獐牙菜次之。狭叶獐牙菜和青叶胆茎与叶的红外光谱在 1 700 ~ 1 533 cm^{-1} 波段与川东獐牙菜、西南獐牙菜和紫红獐牙菜差异较明显, 表明狭叶獐牙菜与青叶胆的化学成分与其他 3 种獐牙菜属植物有所区别, 狭叶獐牙菜和青叶胆酯类含量可能更高。川东獐牙菜、紫红獐牙菜和西南獐牙菜的红外光谱吸收峰强度无明显差异, 3 种獐牙菜在化学组成上较为相似。综上所述, 仅比较狭叶獐牙菜、西南獐牙菜、川东獐牙菜、青叶胆及紫红獐牙菜红外光谱图, 难以区分獐牙菜属不同种, 因此将通过化学计量等方法对其进行物种鉴别。

3.2 PLS-DA PLS-DA 通过将数据转换成一组线性潜在变量用于分类变量预测, 该方法以自变量矩阵 (如光谱范围) 为 X 变量, 训练集样品分类类别 (如不同种类獐牙菜) 为 Y 变量, 通过已知类别作为训练集建立分类模型, 对未知样品进行预测, 是一种有监督的判别分析方法, 能解决维度高、噪声大的问题^[23-24]。PLS-DA 中, 参数 R^2Y 代表对潜在变量数的解释能力, 表明模型与数据的匹配程度, 其数值越接近 1 说明包含样品相关信息越多, 模型效果越好; Q^2 是经交叉验证所得的一项拟合参数, 代表模型预测未知数据的能力, Q^2 超过 0.5 时模型预测能力强; 校正均方根误差 (RMSEE), 内部交叉验证均方根误差 (RMSECV), 预测均方根误差 (RMSEP), 3 个参数值越小, 且 $\text{RMSECV} \geq \text{RMSEP}$, 说明所建判别模型越稳定, 效果越好。若 $\text{RMSECV} < \text{RMSEP}$, 模型可能过拟合。 A_T 为训练集分类正确率, A_P 为预测集分类正确率, 数值越大说明正确率越高^[25]。参数计算方法如下:

$$R^2Y = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2} \quad (1)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^{N_p} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_p} (\hat{y}_i - \bar{y})^2} \quad (2)$$

$$\text{RMSEE} = \sqrt{\frac{\sum_{i=1}^{N_T} (\hat{y}_i - y_i)^2}{N_T - 1}} \quad (3)$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{N_T} (\hat{y}_{\text{cv}i} - y_i)^2}{N_T - 1}} \quad (4)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N_p} (\hat{y}_i - y_i)^2}{N_p}} \quad (5)$$

其中, y_i 为样品分类真实值, \hat{y}_i 为第 i 个样品分

类预测值, \hat{y}_{cvi} 为第 i 个样品交叉验证估值, \bar{y} 为平均值, N 为样本数, N_T 为训练集样本数, N_P 为预测集样本数^[26-27]。

选用样品 3 699 ~ 2 802, 1 797 ~ 685 cm^{-1} 和 652 ~ 550 cm^{-1} 波段红外光谱数据建立川东獐牙菜、青叶胆、狭叶獐牙菜、紫红獐牙菜和西南獐牙菜不同部位(根、茎、叶) PLS-DA 模型。根和叶光谱数据通过 MSC + SG + 2D 处理后, R^2Y 和 Q^2 值最大, RMSEE 和 RMSECV 较小, A_T 为 100%, 故选择 MSC + SG + 2D 为最佳预处理方式。经过 MSC + SG + 2D 处理后预测集样品全部分类正确, A_P 为

100%, 且 $\text{RMSEP} < \text{RMSECV}$, 说明所建模型稳定、可靠, 预测效果好。茎原始光谱数据经 SNV + SG + 2D 和 MSC + SG + 2D 预处理后建立 PLS-DA 效果佳, $R^2Y > 0.88$, RMSEE 和 RMSECV 均 < 0.16 , 且 $\text{RMSEP} < \text{RMSECV}$, 表明模型稳定、可靠, $Q^2 > 0.75$, $\text{RMSEP} < 0.13$, 说明模型预测能力强, A_T 和 A_P 均为 100%, 全部样品分类正确。对比根、茎和叶最佳预处理 PLS-DA 模型, A_P 均为 100%, 表明 3 个部位构建 PLS-DA 模型均能准确鉴别川东獐牙菜、青叶胆、狭叶獐牙菜、紫红獐牙菜和西南獐牙菜, 模型预测能力好。见表 3。

表 3 不同预处理后 PLS-DA 模型参数

Table 3 Parameters of PLS-DA models with different pretreatment

部位	预处理	R^2Y	Q^2	RMSEE	RMSECV	$A_T/\%$	RMSEP	$A_P/\%$
根	RAW	0.773	0.608	0.172 5	0.215 8	97.17	0.175 3	100
	SNV + SG + 1D	0.713	0.628	0.184 3	0.202 0	95.28	0.157 8	100
	SNV + SG + 2D	0.813	0.624	0.150 9	0.205 6	99.06	0.154 0	96.36
	SNV + SG + 3D	0.825	0.654	0.144 4	0.188 8	99.06	0.178 1	94.55
	MSC + SG + 1D	0.712	0.613	0.184 2	0.205 4	95.28	0.157 4	100
	MSC + SG + 2D	0.855	0.601	0.134 1	0.206 5	100	0.151 2	100
	MSC + SG + 3D	0.822	0.649	0.145 6	0.188 5	99.06	0.175 7	94.55
茎	RAW	0.762	0.636	0.156 9	0.200 8	98.41	0.135 7	100
	SNV + SG + 1D	0.857	0.759	0.119 1	0.161 9	100	0.102 6	100
	SNV + SG + 2D	0.890	0.764	0.111 5	0.157 5	100	0.127 2	100
	SNV + SG + 3D	0.832	0.735	0.135 5	0.163 0	98.41	0.144 6	96.92
	MSC + SG + 1D	0.876	0.771	0.113 6	0.151 0	100	0.099 2	100
	MSC + SG + 2D	0.887	0.759	0.112 5	0.156 7	100	0.124 0	100
	MSC + SG + 3D	0.830	0.729	0.136 5	0.163 8	98.41	0.145 7	95.38
叶	RAW	0.850	0.731	0.128 0	0.191 0	99.21	0.112 2	100
	SNV + SG + 1D	0.882	0.817	0.108 1	0.146 0	99.21	0.091 4	100
	SNV + SG + 2D	0.899	0.820	0.103 3	0.137 2	99.21	0.098 3	100
	SNV + SG + 3D	0.903	0.799	0.103 2	0.141 7	99.21	0.103 2	100
	MSC + SG + 1D	0.881	0.808	0.108 7	0.144 8	99.21	0.093 7	100
	MSC + SG + 2D	0.902	0.822	0.102 7	0.137 4	99.21	0.100 7	100
	MSC + SG + 3D	0.901	0.770	0.104 3	0.149 6	99.21	0.100 0	100

3.3 SVM 分析 SVM 是由 Vapnik 等^[27]提出的一种有监督机器学习方法, 通过数据映射到高维空间并构造一个最优分类超平面为决策曲面, 使得不同类别间隔离边缘最大化, 从而可区分不同类型数据, 因此广泛用于药材鉴别研究^[28]。不同预处理光谱数据建立 SVM 判别模型, 对川东獐牙菜、青叶胆、狭叶獐牙菜、紫红獐牙菜和西南獐牙菜进行鉴别。543 份獐牙菜属植物样品采用 KS 算法按 2:1 划分训练集和预测集。建模前, 将 5 种獐牙菜训练集与预测集数据归一化到 $[0, 1]$ 区间, 采用网格搜索法(Grid Search, GS)优化惩罚参数 c 与核函数参数 g 。

先将 $\log_2 c$ 与 $\log_2 g$ 在 $[-10, 10]$ 区间内粗略搜索, 筛选出最佳 c 与 g 值后根据粗选结果进行精细选择, $\log_2 c$ 选择范围在 $[-5, 20]$ 区间, $\log_2 g$ 选择范围在 $[-20, 0]$ 区间, 得到惩罚参数 c 与核函数 g 的最佳值以及判别模型 A_T , 最佳惩罚参数 c 与核函数 g 代入 SVM 模型, 对训练好的模型准确性进行外部验证, 得到 A_P 。

表 4 为 5 种獐牙菜属植物根、茎和叶光谱数据预处理后建立 SVM 模型结果, Best c , Best g 分别代表精细筛选后最优惩罚参数与核函数参数。由表 4 可知, 根、茎原始数据建立 SVM 模型 c 值过大, 分别

为 5 792.618 8 和 8 192, 有过拟合风险^[29]。MSC + SG + 2D 为根最佳预处理方法, 最佳惩罚参数 $c = 22.627 4$, 核函数参数 $g = 0.001 7$, A_T 为 100%, 将 55 个预测数据集代入 SVM 模型, A_P 为 100%; 茎与叶光谱数据经 SNV + SG + 2D, SNV + SG + 3D, MSC + SG + 2D 与 MSC + SG + 3D 预处理构建 SVM 模型, Best c 与 Best g 均在正常范围, 且 A_T 与 A_P 均达到 100%。通过分析 21 种 SVM 模型, 发现预处理后 SVM 模型参数优于原始数据; SNV 与 MSC 预处理对叶和根的 SVM 模型效果基本一致, 无较大差别; 2D 与 3D 处理后 SVM 模型判别效果优于 1D, 由于导数阶数增加可能导致光谱信号失真、噪声放大以及样品化学成分信号衰减, 故预处理选用 2D; 根、茎、叶 3 个部位红外光谱结合 SVM 均能准确鉴别 5 种獐牙菜属植物, 且分类效果好, 结果见图 2。

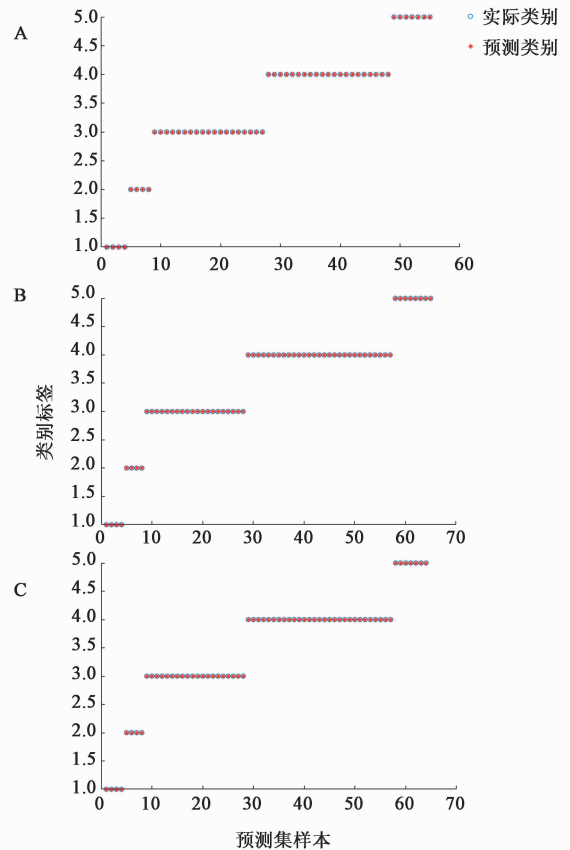
表 4 不同预处理后 SVM 模型参数

Table 4 Parameters of SVM models with different pretreatment

部位	预处理	Best c	Best g	$A_T/\%$	$A_P/\%$
根	RAW	5 792.618 8	0.000 1	93.40	96.36
	SNV + SG + 1D	4	0.022 10	95.28	100
	SNV + SG + 2D	16	0.002 8	95.28	100
	SNV + SG + 3D	2.828 4	0.015 6	94.34	94.55
	MSC + SG + 1D	2.828 4	0.015 6	95.28	100
	MSC + SG + 2D	22.627 4	0.001 7	96.23	100
	MSC + SG + 3D	4	0.011 0	95.28	96.36
茎	RAW	8 192	0.000 1	95.24	100
	SNV + SG + 1D	1	0.022 1	99.21	100
	SNV + SG + 2D	2	0.022 1	100	100
	SNV + SG + 3D	5.656 9	0.015 6	100	100
	MSC + SG + 1D	1	0.031 3	99.21	100
	MSC + SG + 2D	2	0.022 1	100	100
	MSC + SG + 3D	5.656 9	0.015 6	100	100
叶	RAW	64	0.002 0	95.28	100
	SNV + SG + 1D	1.414 2	0.031 3	99.21	100
	SNV + SG + 2D	2	0.015 6	100	100
	SNV + SG + 3D	1.414 2	0.022 1	100	100
	MSC + SG + 1D	1.414 2	0.031 3	99.21	100
	MSC + SG + 2D	1.414 2	0.015 6	100	100
	MSC + SG + 3D	2	0.015 6	100	100

4 结论

本研究采集 5 种獐牙菜属植物根、茎和叶的红外光谱信息, 选择 MSC, SNV, SG, 1D, 2D 和 3D 等方法对原始光谱数据进行预处理, 建立 PLS-DA 和 SVM 模型对不同种类獐牙菜属植物进行鉴别。结果发现, 5 种獐牙菜属植物的共有峰在 3 340, 2 920, 2 850, 1 735, 1 651, 1 614, 1 506, 1 234, 1 160, 1 027,



A. 根; B. 茎; C. 叶

图 2 SVM 对 5 种獐牙菜属植物不同部位测试集的实际分类与预测分类

Fig. 2 Actual and SVM predicted categories of test set samples for roots from five *Swertia* plants in different parts

899 cm^{-1} 处, 狭叶獐牙菜和紫红獐牙菜的峰强度高 于其他 3 种獐牙菜。同一物种的茎与叶吸收峰强度 明显高于根。根、茎和叶 3 个部位 PLS-DA 和 SVM 模型均可准确鉴别不同种獐牙菜属植物, 最佳预处 理方法为 MSC + SG + 2D。FTIR 技术无需对样品进 行前处理, 操作方便、快捷, 这为獐牙菜属植物鉴别 提供了一种高效、准确的鉴别方法。由于实验材料 限制, 仅采集了 5 种獐牙菜属植物样品, 如果要 将模型进行实际应用, 还需增加更多獐牙菜属植物, 不断 修正模型, 使其适用性更强。

[参考文献]

[1] HE S X, LIU X H, ZHANG W Y, et al. Discrimination of the *Coptis chinensis* geographic origins with surface enhanced Raman scattering spectroscopy [J]. Chemometr Intell Lab Syst, 2015, 146:472-477.
[2] PAN Z, FAN G, YANG R P, et al. Discriminating *Lamiophlomis rotata* according to geographical origin by ¹H-NMR spectroscopy and multivariate analysis [J].

- Phytochem Anal, 2015, 26(4): 247-252.
- [3] 中国科学院《中国植物志》编辑委员会. 中国植物志. 第62卷[M]. 北京: 科学出版社, 1988: 385.
- [4] LI J, ZHAO Y L, HUANG H Y, et al. Phytochemistry and pharmacological activities of the genus *Swertia* (Gentianaceae): a review[J]. Am J Chin Med, 2017, 45(4): 667-736.
- [5] 贾敏如, 李星玮. 中国民族药志要[M]. 北京: 中国医药科技出版社, 2005: 593.
- [6] 王国强. 全国中草药汇编(卷一、三)[M]. 北京: 人民卫生出版社, 2014: 667, 583.
- [7] Amirabadizadeh H, Jafari A, Mahmoodzadeh H. Comparative morphology, anatomy and palynological studies of perennial species of *Onobrychis* (Fabaceae) in northeast Iran [J]. Nordic J Bot, 2015, 33(2): 159-169.
- [8] 杨晶凡, 蒋超, 袁媛, 等. 快速PCR方法在山药真伪鉴别中的应用[J]. 中国实验方剂学杂志, 2018, 24(22): 45-49.
- [9] 施崇精, 王姗姗, 程中琴, 等. HPLC特征指纹图谱结合化学计量学比较川牛膝及其混淆品、掺混品化学成分差异[J]. 中国中药杂志, 2018, 43(11): 2313-2320.
- [10] 狄准, 张霁, 赵艳丽, 等. 3种獐牙菜属植物民族药UPLC指纹图谱研究[J]. 中草药, 2017, 48(9): 1860-1865.
- [11] LI J, ZHANG J, JIN H, et al. Using UHPLC and UV-vis fingerprint method to evaluate substitutes for *Swertia mileensis*: an endangered medicinal plant [J]. Pharmacogn Mag, 2017, 13(49): 13-20.
- [12] Karoui R, Downey G, Blecker C. Mid-Infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships-a review [J]. Chem Rev, 2010, 110(10): 6144-6168.
- [13] YANG Y, JIN H, ZHANG J, et al. Determination of total steroid saponins in different species of *Paris* using FTIR combined with chemometrics[J]. J AOAC Int, 2018, 101(3): 732-738.
- [14] WANG Y Z, SHEN T, ZHANG J, et al. Geographical authentication of *Gentiana rigescens* by high-performance liquid chromatography and infrared spectroscopy [J]. Anal Lett, 2018, 51(14): 2173-2191.
- [15] 吴喆, 张霁, 金航, 等. 红外光谱结合化学计量学对不同采收期滇重楼的定性定量分析[J]. 光谱学与光谱分析, 2017, 37(6): 1754-1758.
- [16] Horn B, Esslinger S, Pfister M, et al. Non-targeted detection of paprika adulteration using mid-infrared spectroscopy and one-class classification-is it data preprocessing that makes the performance [J]. Food Chem, 2018, 257: 112-119.
- [17] DeJong S A, O'Brien W L, LU Z, et al. Optimization of gap derivatives for measuring blood concentration of fabric using vibrational spectroscopy [J]. Appl Spectrosc, 2015, 69(6): 733-748.
- [18] Mezzettiab A, Kisha E, Roberta B, et al. Assignment of IR bands of isolated and protein-bound Peridinin in its fundamental and triplet state by static FTIR, time-resolved step-scan FTIR and DFT calculations [J]. J Mol Structure, 2015, 1090: 58-64.
- [19] Rinnan A, Berg F V D, Engelsen S B. Review of the most common pre-processing techniques for near-infrared spectra [J]. TrAC-Trends Anal Chem, 2009, 28(10): 1201-1222.
- [20] Cebi N, Durak M Z, Toker O S, et al. An evaluation of fourier transforms infrared spectroscopy method for the classification and discrimination of bovine, porcine and fish gelatins [J]. Food Chem, 2016, 190: 1109-1115.
- [21] 杨红霞, 马芳, 杜玉枝, 等. 藏药川西獐牙菜及其不同提取物的红外光谱分析[J]. 光谱学与光谱分析, 2014, 34(11): 2973-2977.
- [22] CHEN J B, SUN S Q, MA F, et al. Vibrational microspectroscopic identification of powdered traditional medicines: chemical micromorphology of *Poria* observed by infrared and Raman microspectroscopy [J]. Spectrochim Acta A Mol Biomol Spectrosc, 2014, 128: 629-637.
- [23] YAO S, LI T, LIU H, et al. Traceability of Boletaceae mushrooms using data fusion of UV-visible and FTIR combined with chemometrics methods [J]. J Sci Food Agr, 2018, 98(6): 2215-2222.
- [24] WANG Q Q, HUANG H Y, WANG Y Z. Geographical authentication of *Macrohyporia cocos* by a data fusion method combining ultra-fast liquid chromatography and Fourier transform infrared spectroscopy [J]. Molecules, 2019, 24(7): 1320.
- [25] 吴喆, 张霁, 张金渝, 等. 红外光谱法快速预测不同种类重楼中重楼皂苷含量[J]. 中国中药杂志, 2017, 42(17): 3403-3410.
- [26] Sarraguca M C, Lopes J A. Quality control of pharmaceuticals with NIR: from lab to process line [J]. Vib Spectrosc, 2009, 49(2): 204-210.
- [27] Vapnik V, Cortes C. Support vector networks [J]. Mach Learn, 1995, 20(3): 273-297.
- [28] 李运, 徐福荣, 张金渝, 等. FTIR结合化学计量学对三七产地鉴别及皂苷含量预测研究[J]. 光谱学与光谱分析, 2017, 37(8): 2418-2423.
- [29] Sobek S, Tranvik L J, Prairie Y T, et al. Patterns and regulation of dissolved organic carbon: an analysis of 7, 500 widely distributed lakes [J]. Limnol Oceanogra, 2007, 52(3): 1208-1219.

[责任编辑 顾雪竹]