

基于主成分聚类 and PLS 回归分析的穿心莲质量等级评价研究

崔丹丹^{1,3}, 曾令杰^{1,2*}, 黄嘉玲^{1,2}, 冯晓云^{1,2}, 张晓元³, 冯 凯^{1,2}

1. 广东药科大学中药学院, 广东 广州 510006

2. 国家中医药管理局岭南中药材生产与开发重点研究室, 广东 广州 510006

3. 韶关市华工高新技术产业研究院, 广东 韶关 512027

摘要: **目的** 结合性状指标和内在质量指标对穿心莲进行质量等级评价, 建立等级评价模型, 为穿心莲及其他中药材等级标准的制定提供参考。 **方法** 采收不同批次的穿心莲药材样品, 测量和描述穿心莲药材的性状, 采用数量分类学的方法对其评价指标进行编码, 以 HPLC 法测定药材的 4 个二萜内酯类成分的含量, 将编码后的性状指标和 4 个内酯的总量及醇溶性浸出物含量做相关性分析, 初步筛选出 13 个评价指标, 应用主成分聚类分析法对这 13 个指标进行分析, 划分等级; 并以偏最小二乘-判别分析 (PLS-DA) 对所划分的等级进行判别分析; 最后, 通过偏最小二乘回归分析 (PLS) 建立穿心莲药材等级的预测模型。 **结果** 根据主成分聚类分析法可以将穿心莲药材划分为 3 个等级, PLS-DA 分析表明该等级划分合理; PLS 回归分析所建立的穿心莲质量等级的预测模型为药材等级 (Y) = $3.761 - 0.020 \times \text{叶含量} - 0.388 \times \text{穿心莲内酯含量} - 1.117 \times \text{新穿心莲内酯含量} - 0.274 \times \text{去氧穿心莲内酯含量} - 0.287 \times \text{脱水穿心莲内酯含量} - 0.302 \times 4 \text{ 个内酯总量} - 0.104 \times \text{醇溶性浸出物含量} - 0.015 \times \text{茎颜色} - 0.008 \times 4 \times \text{叶颜色} - 0.003 \times \text{茎基部直径} + 0.020 \times \text{分枝数} + 0.137 \times \text{茎上部直径} + 0.011 \times \text{株高}$, 若 Y 在 0.7~1.3, 则预测穿心莲药材为一等品; 若 Y 在 1.7~2.3, 则为二等品; 若 Y 在 2.7~3.3 则为合格品。 **结论** 主成分聚类分析法结合 PLS 回归建立的穿心莲等级快速评价模型评价效果较为理想, 可作为穿心莲药材质量等级的快速评价模型, 为穿心莲及其他中药材质量等级评价及其等级标准的制定提供新思路。

关键词: 穿心莲; 等级标准; 主成分聚类分析; PLS-DA 分析; PLS 回归分析

中图分类号: R282.6 **文献标志码:** A **文章编号:** 0253-2670(2019)13-3200-07

DOI: 10.7501/j.issn.0253-2670.2019.13.029

Study on quality grade evaluation of *Andrographis Herba* based on principal component clustering and PLS regression

CUI Dan-dan^{1,3}, ZENG Ling-jie^{1,2}, HUANG Jia-ling^{1,2}, FENG Xiao-yun^{1,2}, ZHANG Xiao-yuan³, FENG Kai^{1,2}

1. School of Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China

2. Key Laboratory of State Administration of Traditional Chinese Medicine for Production & Development of Cantonese Medicinal Materials, Guangzhou 510006, China

3. Research Institute of Shaoguan Huagong High-tech Industry, Shaoguan 512027, China

Abstract: Objective To combine macroscopical characteristic indices and chemical indices of *Andrographis Herba* to evaluate its quality grade. **Methods** Both macroscopical characteristic indices and chemical indices (the content of four active diterpenoids and the content of ethanol-soluble extractives) of different batches of *Andrographis Herba* were determined. The macroscopical characteristic indices were encoded using the method of numerical taxonomy, and the content of four active diterpenoids were determined by HPLC. To screen out the appropriate indices for classification, the correlational analyses were conducted between encoded macroscopical characteristic indices and chemical indices. The quality grade was made by principal component clustering analysis according these evaluation indices, and then was analyzed through partial least squares discriminant analysis (PLS-DA). Furthermore, a partial least squares (PLS) regression was constructed for the quality grade prediction of *Andrographis Herba*. **Results** It showed that the samples could be divided into three grades according to the principal component clustering analysis, and was reasonable evaluating by PLS-DA. The PLS regression model for quality grade of *Andrographis Herba* was constructed as follows:

收稿日期: 2019-02-01

基金项目: 国家新兴产业重大工程包中药标准化项目: 槟榔等 20 种中药饮片标准化建设 (ZYBZH-Y-GD-13); 广东省科技计划项目: 韶关高新区食品安全质量研究公共技术服务平台建设 (2017A020108006)

作者简介: 崔丹丹 (1993—), 女, 在读硕士研究生, 研究方向为生药的质量控制研究。E-mail: cdd1677322833@126.com

*通信作者 曾令杰 (1970—), 男, 教授, 博士, 研究方向为药用植物品质评价研究。E-mail: lingjiezheng@gdpu.edu.cn

grade $Y=3.761-0.020\times\text{the leaf content}-0.388\times\text{the content of andrographolide}-1.117\times\text{the content of neoandrographolide}-0.274\times\text{the content of deoxyandrographolide}-0.287\times\text{the content of 14-deoxy-11,12-didehydro-andrographolide}-0.302\times\text{the content of four active diterpenoids}-0.104\times\text{the content of ethanol-soluble extractives}-0.015\times\text{the color of stem}-0.0084\times\text{the color of leaf}-0.003\times\text{the diameter of base part of stem}+0.020\times\text{the number of branch}+0.137\times\text{the diameter of the upper stem}+0.011\times\text{plant height}$, if $Y=0.7-1.3$, the predicted quality was grade A, if $Y=1.7-2.3$, then B grade, and if $Y=2.7-3.3$, C grade or qualified product.

Conclusion The model of grade evaluation we constructed using principal component clustering analysis combining with PLS regression analysis performed well, which was applicable in evaluating the quality grade of *Andrographis Herba* and other traditional Chinese medicines. It also provided a new strategy for study on grade standards of traditional Chinese medicines.

Key words: *Andrographis Herba*; grade standard; principal component clustering analysis; PLS-DA analysis; PLS regression analysis

穿心莲药材 *Andrographis Herba* 为爵床科植物穿心莲 *Andrographis paniculata* (Burm. f.) Nees 的干燥地上部分, 主要含有穿心莲内酯、新穿心莲内酯、去氧穿心莲内酯、14-去氧-11,12-二去氢穿心莲内酯(又称脱水穿心莲内酯素)、穿心莲苷(又称穿心莲内酯苷)及新穿心莲内酯苷元等二萜内酯类成分, 具有抗菌、消炎、抗病毒、抗肿瘤等药理活性^[1-3], 主要分布于广西、广东、福建、湖南、四川等地, 由于各地的生态环境条件及药材栽培管理、采收加工方式的不同导致市场上穿心莲药材质量的差异较大^[4-5]。研究表明, 不同来源的穿心莲药材中穿心莲内酯类成分含量变化较大^[6-9], 《中国药典》2015 年版^[1]以穿心莲内酯和脱水穿心莲内酯的总量不低于 0.8% 作为穿心莲药材质量考察指标; 但仅用二者含量控制和评价穿心莲药材及其制剂的质量, 难以全面反映其药材质量等级及优劣。目前穿心莲药材尚没有统一的等级标准, 难于区分其质量优劣等级。因此, 开展穿心莲药材等级标准研究, 对穿心莲规范化生产和标准化具有十分重要的意义。但如何评级穿心莲药材的等级, 迄今为止, 文献并无具体方法的研究报道。传统意义上对穿心莲药材鉴别是以外观质量为依据的“辨状论质”, 而现代评价则多以药效成分穿心莲内酯和脱水穿心莲内酯含量为主, 如果仅以其中一种方法判断其质量优劣则有失偏颇, 只有全面地结合其外观质量性状和内在品质等多个方面的影响因素进行综合评价, 才能有效地判定其质量等级。本研究以不同产地的穿心莲药材为研究对象, 将传统的外观质量性状和内在活性成分等指标相结合, 采用统计学、化学计量学等方法, 开展了穿心莲药材的质量等级评价研究, 建立了穿心莲药材的质量等级快速评价模型, 为穿心莲药材等级评价及等级质量标准的制定提供了新思路, 为其他中药材的等级标准制定提供参考。

1 材料与仪器

1.1 样品来源

穿心莲药材采自广东(S1~S25)、广西(S26~S30、S36)、福建(S31~S35)共36批, 经广东药科大学中药学院曾令杰教授鉴定为穿心莲 *Andrographis paniculata* (Burm. f.) Nees。

1.2 试剂与仪器

对照品穿心莲内酯(批号 L23S6Y3682)、新穿心莲内酯(批号 Y27J8S40807)、去氧穿心莲内酯(批号 Y26D6S7349)购自上海源叶生物科技有限公司; 脱水穿心莲内酯(批号 J1020AS)购自大连美伦生物技术有限公司, 质量分数均大于 98%。乙腈为色谱纯, 其他试剂均为分析纯。岛津 LC-20AT HPLC 色谱仪, Agilent 5 TC-C₁₈ (2) 色谱柱(250 mm×4.6 mm, 5 μm), KQ5200DE 数控超声波清洗仪(昆山舒美仪器有限公司), BP211D 电子天平(德国赛多利斯公司)。

2 方法与结果

2.1 性状指标的测量及编码

根据传统的经验鉴别方法, 选取穿心莲药材的叶含量(叶质量占药材总质量的百分比)、分枝数、株高、茎直径(上、中、下3个部位)、茎颜色、叶颜色、质地等作为穿心莲外观性状指标进行量化测定。随机抽取穿心莲样品每批次20株, 分离出其叶、茎及杂质, 分别称定质量, 计算其叶质量占总质量的百分比, 作为叶含量; 测量其植株的高度, 即完整植株茎基部(茎的最下部1 cm处)到顶部(茎的最上部1 cm处)的距离, 作为株高, 精确至0.01 cm; 采用游标卡尺测量植株茎基部直径(茎的最下部1 cm处)、中部直径(株高的1/2处)、上部直径(茎的最上部1 cm处), 精确至0.01 mm; 记录完整植株的一级分枝的数目作为分枝数; 描述并记录其茎的颜色、叶的颜色、质地等性状。

对上述性状进行数量分类学的性状编码处理^[10]: 数值性状(株高、茎基部直径、茎中部直径、茎上

部直径、分枝数、叶含量) 无需编码, 直接取其测量值的平均值; 二元性状(质地脆与否)按“0”(否)、“1”(是)编码; 有序多态性状, 按“0、1、2、3…”编码(茎、叶颜色黄色赋值“0”, 黄绿“1”, 灰绿“2”, 绿“3”, 墨绿“4”)。

2.2 穿心莲内酯活性成分的含量测定

穿心莲醇溶性浸出物参照《中国药典》2015 年版四部通则项下的热浸法, 用乙醇作溶剂测定^[1]。

研究表明穿心莲地上部分的主要药效成分为二萜内酯类化合物, 其中以穿心莲内酯、新穿心莲内酯、去氧穿心莲内酯和脱水穿心莲内酯为其主要有效成分。故本实验以这 4 种二萜内酯为指标采用 HPLC 测定其含量。色谱条件: 流动相为乙腈-水(33:67)等度洗脱 40 min, 体积流量为 1 mL/min, 柱温为室温, 检测波长为 223 nm。参照《中国药典》2015 年版穿

心莲项下的规定制备供试品溶液和对照品溶液, 按上述条件进行方法学考察(线性关系、精密密度试验、重复性试验、稳定性试验、加样回收率试验), 计算相关系数 r 、平均加样回收率和 RSD, 结果表明此方法的线性关系($r > 0.999$)、回收率、重复性、稳定性和仪器精密密度(RSD 均小于 3%)均较好。按照上述条件测定穿心莲样品中 4 种内酯的含量。

2.3 数据分析

利用 SPSS 23.0 软件做统计描述和相关性分析; 用 SIMCA-P 14.0 软件进行主成分分析(PCA)、系统聚类分析(HCA)、偏最小二乘判别分析(PLS-DA)和偏最小二乘回归分析(PLS)等数据处理分析。不同批次的穿心莲样品的质地均一致, 因此, “质地”这个指标不作为分级参考指标。各批次穿心莲药材的性状指标和内在质量指标的结果见表 1。

表 1 各批次穿心莲药材各指标测定情况 ($\bar{x} \pm s, n = 3$)

Table 1 Determination results of different batches of *Andrographis Herba* ($\bar{x} \pm s, n = 3$)

编号	分枝数	茎基部直径/mm	茎中部直径/mm	茎上部直径/mm	株高/cm	叶含量/%	穿心莲内酯/%	新穿心莲内酯/%	去氧穿心莲内酯/%	脱水穿心莲内酯/%	4 个内酯总量/%	醇溶性浸出物/%	茎颜色	叶颜色
S1	36.00±0.60	4.91±0.75	2.60±0.27	0.89±0.20	54.90±0.22	25.45±1.16	0.55±0.14	0.29±0.82	0.57±0.36	0.97±0.14	2.38±0.27	10.02±0.11	0	1
S2	38.00±0.86	4.56±0.60	2.93±0.50	0.97±0.23	73.68±0.73	18.12±0.58	0.37±0.29	0.15±0.28	0.31±0.32	0.52±0.07	1.35±0.69	7.82±0.78	1	1
S3	41.00±0.86	5.25±1.25	2.67±0.68	0.58±0.13	62.96±1.04	29.46±0.43	0.97±0.41	0.26±0.10	0.76±0.02	0.74±0.04	2.73±0.25	11.00±0.07	0	2
S4	37.00±1.10	4.90±0.43	2.42±0.36	0.66±0.11	59.39±0.83	29.14±1.22	0.62±0.02	0.38±0.19	0.54±0.06	0.90±0.06	2.44±0.26	11.04±0.11	1	0
S5	36.00±0.77	4.93±0.85	2.71±0.38	0.93±0.37	54.86±1.32	21.18±0.64	0.59±0.05	0.31±0.92	0.38±0.26	0.75±0.05	2.02±0.18	9.20±0.08	0	0
S6	31.00±1.04	4.75±0.58	2.46±0.31	0.80±0.52	48.90±0.77	27.75±1.05	0.88±0.58	0.27±0.42	0.71±0.03	0.75±0.35	2.61±0.28	10.73±0.44	0	0
S7	41.00±0.91	4.41±0.64	2.27±0.38	0.62±0.30	66.50±1.20	19.58±0.82	0.37±0.26	0.13±0.01	0.38±0.37	0.57±0.34	1.45±0.10	7.60±0.53	1	4
S8	37.00±0.22	4.98±0.96	2.25±0.22	0.53±0.14	58.10±0.28	24.17±1.43	0.71±0.14	0.28±0.67	0.47±0.38	0.85±0.42	2.31±0.61	10.07±0.56	1	1
S9	40.00±0.82	5.17±0.56	2.15±0.27	0.31±0.08	57.40±0.50	23.98±0.64	0.77±0.05	0.52±0.30	0.43±0.14	0.87±0.06	2.39±0.58	10.97±0.21	0	1
S10	40.00±0.46	4.55±0.91	2.06±0.35	0.67±0.10	55.60±0.77	17.09±0.46	0.76±0.16	0.37±0.92	0.32±0.32	0.44±0.52	1.66±0.07	8.53±0.09	2	1
S11	35.00±0.53	4.31±0.66	2.83±0.58	0.67±0.28	77.25±0.93	25.16±0.86	0.21±0.01	0.37±0.16	0.66±0.37	0.42±0.06	1.51±0.63	8.51±0.04	3	3
S12	35.00±0.35	4.88±0.40	2.63±0.30	0.96±0.14	54.20±1.10	22.52±0.91	0.71±0.03	0.43±0.64	0.38±0.47	0.58±0.64	1.85±1.05	10.96±0.09	0	0
S13	39.00±0.67	3.40±0.58	2.19±0.26	0.60±0.26	73.20±0.92	21.04±1.39	0.71±0.36	0.36±0.17	0.52±0.50	0.57±0.63	1.94±0.48	8.71±0.26	0	0
S14	40.00±0.97	4.33±0.98	2.13±0.27	0.30±0.10	70.10±0.41	19.60±1.11	0.78±0.17	0.28±0.66	0.61±0.21	0.81±0.37	2.34±1.50	7.53±0.04	2	1
S15	34.00±1.04	4.84±1.00	2.26±0.36	0.86±0.25	59.80±0.55	18.38±1.25	0.46±0.01	0.47±0.39	0.41±0.07	0.67±0.17	1.69±0.69	9.06±0.43	2	1
S16	41.00±1.01	5.51±0.59	2.42±0.32	0.36±0.08	61.05±1.20	27.37±1.06	0.35±0.03	0.34±0.07	0.97±0.07	0.84±0.06	2.46±0.24	10.40±0.07	2	2
S17	39.00±0.33	5.32±1.25	2.38±0.37	0.79±0.24	63.60±0.85	22.03±1.45	0.24±0.22	0.41±0.14	0.63±0.12	0.47±0.17	1.53±0.06	8.53±0.08	2	3
S18	41.00±0.41	5.59±0.94	2.17±0.47	0.43±0.08	61.70±0.75	22.20±0.68	0.35±0.22	0.31±0.22	0.41±0.15	0.52±0.39	1.42±1.02	9.54±0.65	1	3
S19	44.00±1.15	4.46±0.56	2.20±0.29	0.43±0.13	62.00±0.81	24.42±1.25	0.28±0.22	0.55±0.39	0.43±0.18	0.37±0.25	1.24±1.10	9.92±0.06	0	1
S20	42.00±1.04	5.40±1.13	2.18±0.38	0.42±0.13	60.98±1.40	26.99±0.73	0.29±0.05	0.23±0.04	0.67±0.02	0.76±0.05	1.96±0.09	10.35±0.10	1	2
S21	40.00±0.42	4.39±1.45	2.30±0.32	0.43±0.13	65.19±0.58	19.79±1.22	0.64±0.10	0.36±0.18	0.39±0.15	0.53±0.02	1.67±0.54	8.75±0.10	1	1
S22	37.00±1.01	5.55±0.68	2.54±0.49	0.39±0.09	56.84±0.61	22.40±0.79	0.59±0.12	0.45±0.03	0.55±0.02	0.78±0.09	2.14±0.16	12.33±0.62	1	1
S23	43.00±1.01	4.67±1.20	2.09±0.64	0.44±0.10	61.25±1.09	21.30±0.86	0.35±0.42	0.26±0.19	0.56±0.06	0.58±0.42	1.68±1.05	9.36±0.34	3	3
S24	39.00±0.60	5.98±0.83	2.29±0.58	0.50±0.11	56.65±0.71	26.63±1.29	0.58±0.10	0.51±0.39	0.68±0.12	0.98±0.17	2.47±0.88	12.06±0.38	2	2
S25	41.00±0.56	5.47±1.37	2.61±0.71	0.50±0.12	60.59±1.14	18.94±0.32	0.58±0.00	0.64±0.64	0.46±0.20	0.65±0.13	1.87±1.01	8.45±0.09	1	1
S26	41.00±1.08	5.98±0.79	2.22±0.44	0.41±0.06	55.01±0.71	30.10±0.53	0.64±0.02	0.95±0.06	0.40±0.05	0.69±0.02	2.04±0.02	10.64±0.17	1	2
S27	43.00±0.68	5.32±0.56	2.37±0.24	0.62±0.20	66.80±0.55	28.02±1.06	0.37±0.32	0.46±0.36	0.53±0.24	0.47±0.37	1.55±0.36	10.02±0.60	1	3
S28	42.00±0.88	5.56±0.79	2.86±0.56	0.47±0.06	66.56±1.29	28.86±0.87	0.53±0.29	0.54±0.44	0.49±0.26	0.54±0.46	1.76±0.52	9.55±0.34	2	2
S29	39.00±0.59	5.42±0.61	2.47±0.72	0.44±0.10	60.22±0.28	22.39±1.02	0.64±0.28	0.50±0.27	0.43±0.18	0.69±0.31	2.06±0.15	9.83±0.40	2	3
S30	40.00±0.23	5.50±0.99	2.02±0.51	0.44±0.06	62.51±0.97	38.16±1.19	0.50±0.76	0.39±0.15	0.87±0.01	0.43±0.05	1.97±0.87	9.21±0.06	3	3
S31	33.00±1.01	3.66±0.41	2.68±0.39	0.44±0.15	66.50±0.41	25.44±0.72	0.60±0.25	0.43±0.21	0.27±0.04	0.41±0.04	1.49±0.30	8.54±0.47	3	4
S32	32.00±0.25	4.34±1.17	2.57±0.54	0.61±0.25	66.50±0.34	19.79±0.48	0.74±0.07	0.22±0.10	0.18±0.01	0.27±0.04	1.11±0.07	7.33±0.07	3	4
S33	37.00±1.00	3.42±0.37	2.74±0.49	0.49±0.07	85.90±0.57	24.84±1.47	0.54±0.00	0.26±0.02	0.18±0.01	0.30±0.05	1.43±0.10	8.68±0.83	3	3
S34	34.00±1.27	2.79±0.53	1.85±0.42	0.41±0.15	64.20±0.87	31.49±0.63	1.37±0.23	0.61±0.19	0.23±0.06	0.27±0.05	2.61±0.53	11.50±0.19	1	3
S35	32.00±0.86	4.11±1.20	2.59±0.39	0.44±0.06	68.10±0.84	34.44±0.58	0.82±0.48	0.59±0.04	0.48±0.34	0.47±0.35	2.36±0.25	11.04±0.06	3	3
S36	31.00±1.04	4.01±0.86	2.49±0.23	0.43±0.13	57.60±1.33	21.28±0.73	0.58±0.09	0.34±0.07	0.70±0.02	0.58±0.18	2.08±0.27	10.90±0.36	3	3

2.3.1 穿心莲外观性状指标和内在质量指标的相关性分析 利用 SPSS 23.0 软件对包括外观性状和内在指标在内的 14 个指标进行统计描述和相关性分析, 结果见表 2、3。由表 2 可知分枝数、茎中部直径的 RSD 为 9.44%、10.82% 较小, 说明各批次间差异较小。由表 3 可知茎中部直径除与茎上部直径相关外, 与其他指标无明显相关, 故茎中部直径不适合作为划分穿心莲等级的指标; 分枝数虽然 RSD 小, 但与其他外观性状(如茎基部直径)和内在指标(穿心莲内酯含量、新穿心莲内酯含量)相关性较高, 因此, 保留其作为评价指标。其他指标之间均存在着较为明显的相关性, 根据传统鉴别经验及现代研究, 其他指标均可作为评价穿心莲质量等级的重要指标, 应保留。

表 2 穿心莲各指标描述统计分析

Table 2 Describe statistical analysis of indices of *Andrographis Herba* samples

指标	最小值	最大值	平均值	标准差	RSD/%
分枝数	31.00	44.00	38.08	3.60	9.44
茎基部直径/mm	2.79	5.98	4.80	0.74	15.47
茎中部直径/mm	1.85	2.93	2.41	0.26	10.82
茎上部直径/mm	0.30	0.97	0.56	0.19	33.54
株高/cm	48.90	85.90	62.68	7.23	11.53
叶/%	17.09	38.16	24.43	4.78	19.56
穿心莲内酯/%	0.21	1.37	0.58	0.23	39.48
新穿心莲内酯/%	0.12	0.74	0.24	0.13	52.49
去氧穿心莲内酯/%	0.18	0.97	0.50	0.18	36.24
脱水穿心莲内酯/%	0.27	0.98	0.61	0.19	31.55
4 个内酯总量/%	1.11	2.73	1.93	0.43	22.50
醇溶性浸出物/%	7.33	12.33	9.69	1.29	13.32
茎颜色	0	3	1.44	1.08	75.07
叶颜色	0	4	1.89	1.24	65.45

表 3 穿心莲各指标相关性分析

Table 3 Correlational between indices of *Andrographis Herba*

指标	分枝数	茎基部直径	茎中部直径	茎上部直径	株高	叶含量	穿心莲内酯含量	新穿心莲内酯含量	去氧穿心莲内酯含量	脱水穿心莲内酯含量	4 个内酯总量	醇溶性浸出物	茎颜色	叶颜色
分枝数	1.000													
茎基部直径	0.470**	1.000												
茎中部直径	0.312	0.028	1.000											
茎上部直径	0.325	0.022	0.418*	1.000										
株高	0.063	-0.507**	0.274	0.131	1.000									
叶含量	0.028	0.129	0.086	0.279	0.033	1.000								
穿心莲内酯含量	-0.369*	-0.403*	0.174	0.136	0.123	0.249	1.000							
新穿心莲内酯含量	-0.389*	0.265	0.115	0.183	0.145	0.549**	0.605**	1.000						
去氧穿心莲内酯含量	0.187	0.452**	0.089	0.149	0.251	0.389*	0.242	0.015	1.000					
脱水穿心莲内酯含量	0.121	0.546**	0.024	0.032	-0.544**	0.003	0.012	0.047	0.477**	1.000				
4 个内酯总量	0.163	0.147	0.176	0.168	-0.448**	0.458**	0.612**	0.622**	0.486**	0.663**	1.000			
醇溶性浸出物	0.102	0.283	0.129	0.204	-0.523**	0.520**	0.352*	0.586**	0.327	0.432**	0.678**	1.000		
茎颜色	0.230	0.175	0.080	0.306	0.415*	0.091	0.206	0.004	0.026	-0.412*	0.293	0.283	1.000	
叶颜色	0.030	0.139	0.017	-0.366*	0.395*	0.208	0.197	0.082	0.087	-0.525**	-0.356*	0.213	0.679**	1.000

在* $P < 0.05$ 水平, 相关性显著; 在** $P < 0.01$ 水平, 相关性极显著

*indicated significant correlation at 0.05 level, and ** indicated significant correlation at 0.01 level

2.3.2 基于 PCA 和 PLS-DA 模式的穿心莲药材等级划分 PCA 和系统聚类分析是 2 种常用的非监督模式的分析方法, PCA 分析利用降维的思想, 在最大程度保留原始信息的基础上, 把多个指标数据转化为少数几个新的综合变量来描述原数据特征^[12]。而聚类分析(HCA)是一种常用的凝聚聚类分析方法, 其思想是先将样本各自看成一类, 然后规定样本之间的距离和类与类之间的距离, 最开始时样本间的距离与类间距离是等价的, 选择距离小的聚成一个新类, 重新计算新类与其他类之间的距离, 再

将距离小的 2 个合并成一类, 如此反复, 直到所有样本都成为一类^[13]。PCA 有利于反映原始数据的整体情况, 但是无法进行分类, 当组间差异较小, 而组内差异较大时无法得出正确的结论^[14], 所以利用 PCA 进行降维处理得到精简的具有代表性的新指标, 再根据各样本在各主成分的得分情况进行系统 HCA, 可以得到较好的分类结果。PLS-DA 为有监督分类方法, 是基于 PLS 回归的一种判别分析方法, 主要反映 X 变量与因变量之间的线性关系, 这一方法人为加入了分组变量, 有利于强化组间差

异, 能对不同等级进行判别^[15]。

本实验将筛选得到的 13 个指标进行标准化处理后利用 simca-p 14.0 进行 PCA, 提取指标数据的主要信息, 得到的 PCA 得分图、载荷因子图见图 1, 特征值及方差贡献率见表 4, 然后利用提取的主成分采用 Ward 聚类法进行 HCA, 得到穿心莲药材的分类结果, 见图 2。为验证分类结果的正确性, 将对各样本进行有监督的 PLS-DA, 并对模型进行 200 次的置换检验实验, 结果见图 3。

由表 4 可知, 前 4 个主成分 PC1、PC2、PC3、PC4 的特征值大于 1, 其累积贡献率为 77.56%, 可以较好地表征原始信息。从 PCA 载荷图(图 1-b)可以看出 4 个内酯总量、醇溶性浸出物、脱水穿心莲内酯含量、株高这 4 个指标在第 1 主成分上载荷较大, 新穿心莲内酯含量、穿心莲内酯含量、茎基部直径、分枝数在第 2 主成分上载荷较大, 可以看出第 1 主成分主要体现的是穿心莲药材的 4 个内酯总量及与其相关性较高的内在化学指标和性状指标指标, 第 2 主成分主要体现的是新穿心莲内酯含量、穿心莲内酯含量及其他的一些外在指标。从 PCA 得分图(图 1-a)看, 穿心莲样本有聚类的趋势, 大致分为 3 类, 但组 1 中 3 个样本较为分散, 组 2 中 S29、S30 分类不明显; 结合 HCA 结果可知, 穿心莲样品分为 3 类: I 类: S34~36; II 类: S1、S3~6、S8、S9、S12、S16、S20、S22、S24、S26、S29、S30; III 类: S2、S7、S10、S11、S13~

表 4 PCA 特征值及方差贡献率

Table 4 Characteristic value and variance contribution rate of PCA

成分	特征值	信息贡献率/%	累积贡献率/%
1	4.059	31.22	31.22
2	2.799	21.53	52.75
3	2.145	16.50	69.25
4	1.080	8.31	77.56
5	0.768	5.91	83.47
6	0.718	5.52	88.99
7	0.350	2.69	91.68
8	0.341	2.63	94.31
9	0.280	2.16	96.47
10	0.200	1.54	98.00
11	0.166	1.28	99.28
12	0.093	0.72	100.00
13	0.001	0.01	100.00

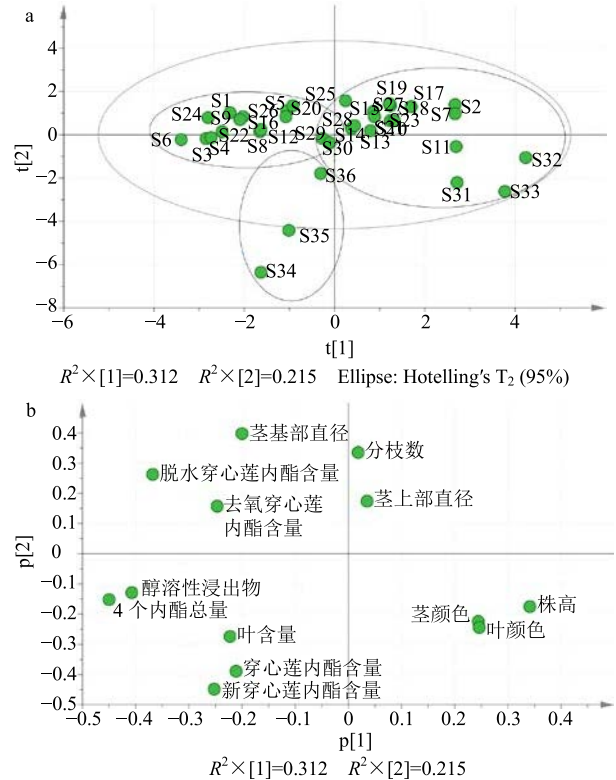


图 1 穿心莲药材 PCA 得分图 (a) 及载荷图 (b)

Fig. 1 PCA score scatter plot (a) and loading scatter plot (b) of *Andrographis Herba* samples

15、S17~19、S21、S23、S25、S27、S28、S31~33。从 PLS-DA 结果图(图 3-a)可以看出, 穿心莲样本可以明显的分为 3 类, 其结果与主成分聚类结果一致, 说明该分类比较合理, 由模型验证(图 3-b)可以看出, 位于左边的 R^2_Y 均低于最右侧的 R^2_Y , 且 R^2 的回归线截距为 0.0862 (< 0.3), Q^2_Y (模拟) 均低于右侧真实模型 Q^2_Y 值, 且 Q^2 的回归线截距为负, 说明该模型预测能力较好, 没有出现拟合现象。

2.3.3 穿心莲药材等级分类预测模型的建立 将 36 批穿心莲药材的 13 个评价指标作为自变量(X), 不同等级分类作因变量(Y), 并将其预设值定义如下: 1 为一等品, 2 为二等品, 3 为合格品, 进行 PLS 回归分析, 若输出值在 0.7~1.3, 则穿心莲样品判别为一等品, 若输出值在 1.7~2.3, 则穿心莲样品判别为二等品, 若输出值在 2.7~3.3, 则判别为合格品。采用软件 simca-p 14.0 建立 PLS 回归模型, 由表 5 可知, 提取第 1 个 PLS 成分对等级分类的交叉有效性是 0.721, 第 2 个成分的是 -0.026 9, 一般认为交叉有效性大于 0.097 5 时引入的新成分对模型的预测效果才有明显改善作用, 但本研究

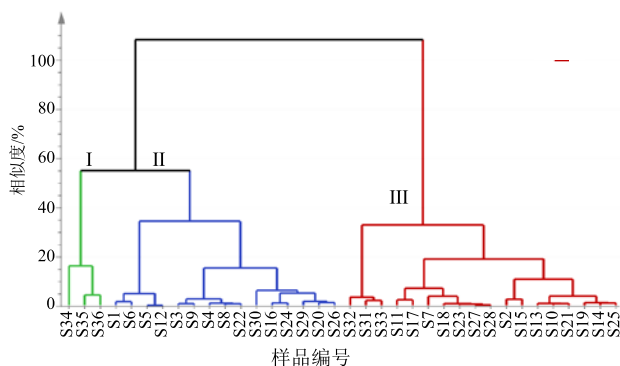


图 2 穿心莲药材等级划分 HCA 结果

Fig. 2 HCA results of quality grade for *Andrographis Herba* samples

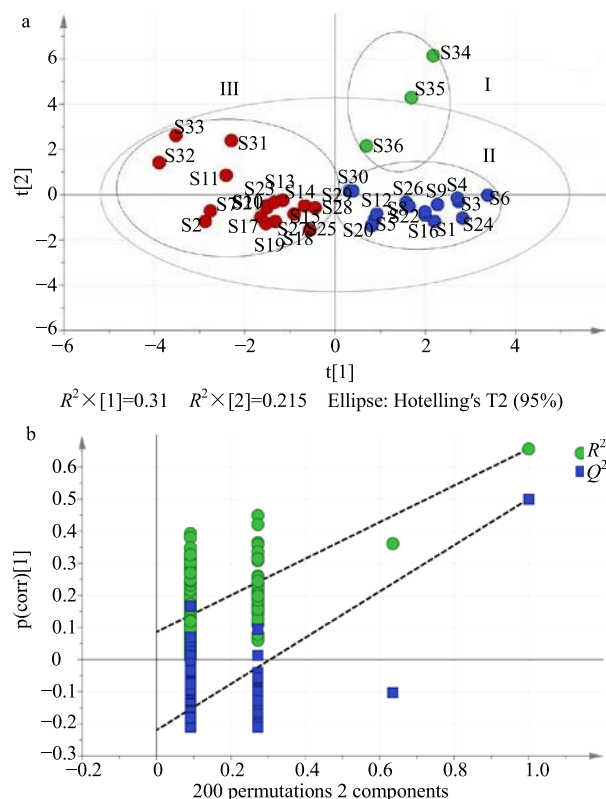


图 3 穿心莲药材等级划分 PLS-DA 得分图 (a) 及模型验证图 (b)

Fig. 3 PLS-DA score scatter plot (a) and permutation plot (b) of different quality grades of *Andrographis Herba* samples

引入第 2 个主成分时 R^2_X 为 0.211, R^2_Y 为 0.043 7, 表示增加一个主成分, 对 X 的解释能力增加 21.1%, 对 Y 的解释能力增加 4.37%, 为了更好地分析结果, 本研究拟提取 2 个主成分进行分析, 其 R^2_X (cum) 为 0.504 说明该模型对变量 X 的解释能力为 50.4%, R^2_Y (cum) 为 0.835, 说明提取的主成分能解释 83.5% 的因变量的变异, Q^2 (cum) 为 0.714 (大于 0.5) 表示该预测模型较为理想。建立穿心莲等级分类与 13 个自变量的偏最小二乘回归模型: $Y=3.761-0.020 X_1-0.388 X_2-1.117 X_3-0.274 X_4-0.287 X_5-0.302 X_6-0.104 X_7-0.015 X_8-0.0084 X_9-0.003 X_{10}+0.020 X_{11}+0.137 X_{12}+0.011 X_{13}$ ($X_1 \sim X_{13}$ 依次为叶含量、穿心莲内酯含量、新穿心莲内酯含量、去氧穿心莲内酯含量、脱水穿心莲内酯含量、4 个内酯总含量、醇溶性浸出物、茎颜色、叶颜色、茎基部直径、分枝数、茎上部直径、株高)。

为考察建立的回归模型的预测精确度, 将原始数据代入上述回归方程计算预测值, 并与实际观测值绘制散点图 (图 4) 进行比较, 可以看出实测值与预测值的样本点基本分布在对角线上, 说明模型拟合效果较好, 可用来预测穿心莲药材的等级。为分析各自变量对等级分类因变量 Y 的影响作用大小, 绘制变量投影重要性指标 (VIP) 得分图 (图 5) 来。从图 5 可以看出, VIP 的顺序为新穿心莲内酯的含量 > 醇溶性浸出物含量 > 4 个内酯总量 > 叶含量 > 穿心莲内酯含量 > 株高 > 分枝数 > 脱水穿心莲内酯含量 > 去氧穿心莲内酯含量 > 茎上部直径 > 茎颜色 > 叶颜色 > 茎基部直径, 其中前 5 个自变量的 VIP 值大于 1, 说明这 5 个指标对预测穿心莲等级分类能起到最重要的作用; 株高、分枝数这 2 个指标的 VIP 值大于 0.8, 表明这 2 个指标在预测等级分类中起较为重要的作用。

3 讨论

本研究探讨了穿心莲药材性状指标与内在化学品质的相关性, 通过主成分聚类分析初步将穿心莲药材分为 3 个等级, 通过 PLS 建立了穿心莲药材等级分类的预测模型: 药材等级 $Y=3.761-0.020 \times$ 叶含量 $-0.388 \times$ 穿心莲内酯含量 $-1.117 \times$ 新穿心莲内酯含量 $-0.274 \times$ 去氧穿心莲内酯含量 $-0.287 \times$ 脱水

表 5 穿心莲药材等级分类 PLS 回归的交叉有效性与精度分析

Table 5 Cross validity and precision analysis values of PLS regression of *Andrographis Herba* samples

组分	对 X 的解释能力	对 X 的累积解释能力	特征值	对 Y 的解释能力	对 Y 的累积解释能力	交叉有效性	置信水平	模型预测能力
1	0.293	0.293	3.81	0.791 0	0.791	0.721 0	0.05	0.721
2	0.211	0.504	2.74	0.043 7	0.835	-0.026 9	0.05	0.714

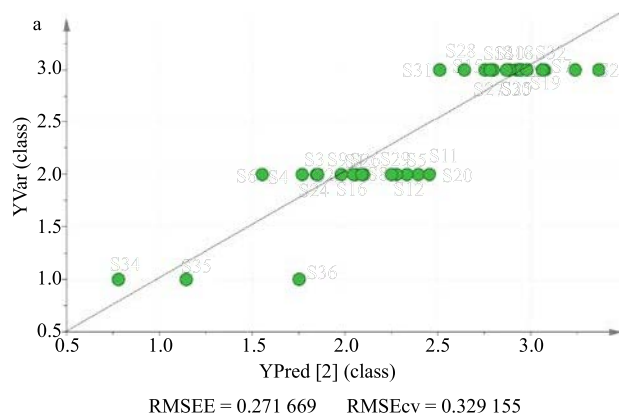


图 4 穿心莲药材等级分类实测值与预测值比较

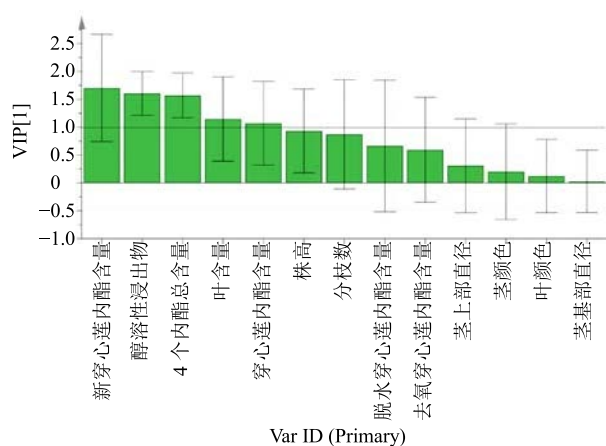
Fig. 4 Comparison between measured and predicted values of different grades of *Andrographis Herba* samples

图 5 穿心莲药材等级影响因素的 VIP 得分图

Fig. 5 Variable projection importance indicator (VIP) score plot of factors for quality grades of *Andrographis Herba* samples

穿心莲内酯含量 -0.302×4 个内酯总量 $-0.104 \times$ 醇溶性浸出物含量 $-0.015 \times$ 茎颜色 $-0.008 \times 4 \times$ 叶颜色 $-0.003 \times$ 茎基部直径 $+0.020 \times$ 分枝数 $+0.137 \times$ 茎上部直径 $+0.011 \times$ 株高, 若 Y 在 $0.7 \sim 1.3$, 则预测穿心莲药材为一等品, 若 Y 在 $1.7 \sim 2.3$, 则预测为二等品, 若 Y 在 $2.7 \sim 3.3$ 则预测为合格品。本研究表明穿心莲药材等级与新穿心莲内酯含量、醇溶性浸出物含量、4个内酯总量、叶含量、穿心莲内酯含量显著相关, 与株高和分枝数相关, 其中新穿心莲内酯含量、醇溶性浸出物含量、4个内酯总量、叶含量、穿心莲内酯含量这5个指标含量越高; 分枝数较少、株高较小穿心莲药材的等级越优。

本研究首次将 PCA 与 HCA 相结合应用于穿心莲药材质量等级的划分研究之中, 将 PLS 应用于穿心莲等级分类的预测, 同时评价了性状指标与内在

化学品质以及确定了各指标的权重, 筛选出了对等级评价具有较大影响的指标, 建立了穿心莲等级的快速评价模型, 为穿心莲药材的优质优价提供了有利的理论基础。这种分析模型可为其他药材的等级评价的研究提供参考。此外, 在今后研究中, 还可同时结合药材指纹图谱及药效进行谱效评价, 建立更加全面、完善的穿心莲药材等级标准。

参考文献

- [1] 中国药典 [S]. 一部. 2015.
- [2] Hossain M, Urbi Z, Sule A, et al. *Andrographis paniculata* (Burm. f.) Wall. ex Nees: A review of ethnobotany, phytochemistry, and pharmacology [J]. *Sci World J*, 2014, doi:10.1155/2014/274905.
- [3] Shailaja A, Bindu B B V, Srinath M, et al. In silico structural and functional analysis of copalyl diphosphate synthase enzyme in *Andrographis paniculata* (Burm. f.) Wall. ex Nees: A plant of immense pharmaceutical value [J]. *Ann Phytomed*, 2018, 7(1): 69-77.
- [4] Qing L I, Jian Z, Liu J H, et al. Morphological and chemical studies of artificial *Andrographis paniculata* polyploids [J]. *Chin J Nat Med*, 2018, 16(2): 81-89.
- [5] Pholphana N, Rangkadilok N, Saehun J, et al. Changes in the contents of four active diterpenoids at different growth stages in *Andrographis paniculata* (Burm. f.) Nees (Chuanxinlian) [J]. *Chin Med*, 2013, 8(1): 2-8.
- [6] Lata M P, Vijaya K. Variation in morphological parameters and andrographolide content in *Andrographis paniculata* (Burm. f.) Nees collected from different provenances of chhattisgarh [J]. *Res J Biotechnol*, 2012, 7(4): 168-174.
- [7] Devi H, Dhuri H, Tirkey A, et al. Genetic variability and heritability studies in *Andrographis paniculata* (Burm. f.) Wall. ex Nees [J]. *Int J Cur Microbiol Appl Sci*, 2017, 6(11): 1847-1855.
- [8] 林青, 匡艳辉, 黄琳, 等. 一测多评法测定穿心莲及其制剂中内酯类成分 [J]. *中草药*, 2012, 43(12): 2406-2411.
- [9] Kumar A, Semwal D P, Bhatt K C, et al. Characterization of indigenous germplasm of *Andrographis paniculata* (Burm. f.) Wall. ex Nees for variability analysis [J]. *Med Plants-Int J Phytomed Related Ind*, 2014, 6(4): 277-281.
- [10] 徐静静, 赵冰, 申惠娟, 等. 15个杜鹃花品种叶片解剖和表型数量分类研究 [J]. *西北林学院学报*, 2017, 32(1): 142-149.
- [11] 中国药典 [S]. 四部. 2015.
- [12] Jolliffe I T, Cadima J. Principal component analysis: A review and recent developments [J]. *Philosoph Transact Royal Soc*, 2016, doi:10.1098/rsta.2015.0202.
- [13] 梁逸曾, 吴海龙, 俞汝勤. 分析化学手册: 化学计量学 [M]. 北京: 化学工业出版社, 2016.
- [14] 阿基业, 何骏, 孙润彬. 代谢组学数据处理-主成分分析十个要点问题 [J]. *药学报*, 2018, 53(6): 929-937.
- [15] Hirri A, Bassbasi M, Platikanov S, et al. FTIR spectroscopy and PLS-DA classification and prediction of four commercial grade virgin olive oils from Morocco [J]. *Food Analyt Meth*, 2016, 9(4): 974-981.