

基于性状电子检测技术结合机器学习的不同产地薏苡仁的快速鉴别

张云飞, 张玖捌, 陈 鹏, 史亚博, 张科卫*, 毛春芹, 陆兔林*

南京中医药大学药学院, 江苏 南京 210023

摘要: 目的 基于性状电子检测技术, 结合多元分析算法, 建立不同产地薏苡仁的快速鉴别方法。方法 使用 CM-5 分光测色仪测定薏苡仁的色度值, 建立决策树 (decision tree, DT) 模型、k 最邻近算法 (k-nearest neighbor, KNN) 模型和贝叶斯 (Bayes) 判别模型。其次, 根据超快速气相电子鼻检测气味成分, 建立不同产地薏苡仁的判别因子分析 (discriminant factor analysis, DFA) 模型及热图, 探究不同产地薏苡仁的气味信息差异。同时, 采用偏最小二乘法判别分析 (partial least squares discrimination analysis, PLS-DA) 模型探究不同产地的差异标志物。最后, 将薏苡仁表面颜色与内在气味成分进行相关性分析。结果 2-丙烯酸、2-甲基丁醛、甲苯、丙醛可以作为区分不同产地薏苡仁的主要气味标志物, 丁二酮、2,3-乙酰基丙酮、3-己醇等多种气味成分与 L^* 、 a^* 、 b^* 值存在显著相关性。结论 电子眼联合超快速气相电子鼻技术能够快速、准确鉴别不同产地的薏苡仁, 该方法对于多产地中药的鉴别和质量控制具有借鉴意义。

关键词: 薏苡仁; 性状电子检测; 机器学习; 快速鉴别; 气味标志物; 2-丙烯酸; 2-甲基丁醛; 丁二酮; 3-己醇

中图分类号: R286.2 文献标志码: A 文章编号: 0253-2670(2024)12-4185-09

DOI: 10.7501/j.issn.0253-2670.2024.12.025

Rapid identification of coicis semen from different habitats based on electronic trait detection technology combined with machine learning

ZHANG Yunfei, ZHANG Jiuba, CHEN Peng, SHI Yabo, ZHANG Kewei, MAO Chunqin, LU Tulin

College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China

Abstract: Objective To establish a rapid identification method of coicis semen from different habitats based on electronic trait detection technology and multiple analysis algorithm. **Methods** Firstly, the chromaticity value of coicis semen was determined by CM-5 spectrophotometer, and the Decision Tree (DT) model, k-nearest neighbor (KNN) model and Bayes discriminant model were established. Secondly, the discriminant factor analysis (DFA) model and heat map of coicis semen of different habitats were established according to the detection of odor components by ultra-fast electronic nose analysis to explore the difference of odor information of coicis semen in different habitats. At the same time, the partial least squares discriminant analysis (PLS-DA) model was used to explore difference markers from different habitats. Finally, the correlation between the surface color of coicis semen and the internal odor components was analyzed. **Results** 2-Propenoate, 2-methylbutanal, toluene, propanal can be used as the main odor markers to distinguish coicis semen from different habitats, butan-2-one, 2,3-pentanedione, 3-hexanol and other odor components were significantly correlated with L^* , a^* , and b^* values. **Conclusion** Electronic eye combined with ultra-fast gas phase electronic nose technique can identify coicis semen from different habitats quickly and accurately. This method can be used as reference for the identification and quality control of multi-habitats Chinese medicines.

Key words: Coicis semen; electronic trait detection; machine learning; rapid identification; odor markers; 2-propenoate; 2-methylbutanal; 1,2-butanediol; 3-hexanol

薏苡仁为禾本科植物薏苡 *Coix lacryma-jobi* L. var. *ma-yuen* (Roman.) Stapf 的干燥成熟种仁。秋季果

实成熟时采割植株, 晒干, 打下果实, 再晒干, 除去外壳、黄褐色种皮和杂质, 收集种仁^[1]。薏苡仁性寒,

收稿日期: 2023-11-02

基金项目: 国家重点研发计划 (2018YFC1707000)

作者简介: 张云飞, 男, 硕士研究生, 主要从事中药炮制及中药饮片质量标准研究。E-mail: 1292868646@qq.com

*通信作者: 张科卫, 博士, 副教授, 主要从事中药炮制及中药饮片质量标准研究。E-mail: kewei@njucm.edu.cn

陆兔林, 教授, 博士生导师, 主要从事中药炮制及中药饮片质量标准研究。E-mail: ltl2021@njucm.edu.cn

入肺胃及脾经,有利脾健胃、渗湿利尿、清热排脓等功效,可用于治疗脾虚泄泻、小便不畅、痈疽不溃以及水肿等病症^[2]。薏苡仁中的主要活性成分包括脂肪酸类、多糖类^[3]、黄酮类、三萜类、甾醇类等^[4],其中脂肪酸类的主要成分是甘油三油酸酯,除此之外还含有薏苡素、生育酚、角鲨烯等多种活性成分。现代研究表明,薏苡仁具有调节血脂、降低血糖、抗肿瘤、抗炎、提高机体免疫力等药理作用^[5-7]。

我国是薏苡仁种植大国,栽培历史十分悠久,目前全国的薏苡仁种植面积已经超过了一百万亩(66 667 公顷)。除甘肃、宁夏等少数省份外,在其他各省份均有分布,其中以贵州、福建和云南等省为主产地^[8]。相关研究表明,地理位置、气候、温湿度等外部环境因素的不同导致了各地薏苡仁中甘油三油酸酯和薏苡素等有效成分的含量差异较大。薏苡仁药材品质的差异对其临床疗效有很大的影响,因此需要对不同产地的薏苡仁进行综合分析考察,以保证薏苡仁的质量和临床功效^[9-12]。

近年来,性状电子检测技术用于中药的全面分析和质量控制已成为一种趋势,其中以分光测色技术和超快速气相电子鼻技术尤为典型。分光测色技术通过将检测样品的光谱数据转换为颜色参数(L^* 、 a^* 、 b^*),实现颜色的客观量化,建立 $L^*a^*b^*$ 模型。该模型是一种三维空间的色彩模型, L^* 、 a^* 、 b^* 分别代表颜色的亮度、红绿色方向、黄蓝色方向。该技术操作简便,只需将样品打粉过筛,无需进行其他处理和使用化学试剂。超快速气相电子鼻是一种基于顶空气相基本原理的新型气味分析仪器,拥有氢离子火焰检测器和 2 根极性不同的色谱柱,分别为 MTX-5(弱极性)和 MTX-1701(中极性),能够

全面客观地识别中药整体气味信息,实现气味的量化。该仪器分析时间短,检测效率高,仅需 2~3 min 即可完成样品的检测。作为新兴智能性状电子检测技术,分光测色技术和超快速气相电子鼻技术都具有快速、准确、便捷的特点,能够准确地将传统的鉴别经验转化为客观数据,目前已应用在中药的产地鉴别、真伪鉴别、炮制品鉴别等多个领域^[13-19]。

因此,本实验采用电子眼联合气相电子鼻技术,对不同产地薏苡仁的进行快速鉴别,并成功筛选出不同产地薏苡仁的气味标志物。同时对薏苡仁的表面颜色与内在气味成分进行相关性分析,探索颜色与成分之间的相关性,为薏苡仁的“辨状论质”提供客观依据。

1 材料与试剂

1.1 材料

26 批薏苡仁药材分别收集于云南(Y1~Y5)、江苏(J1~J6)、福建(F1~F4)、山东(S1~S4)和贵州省(G1~G7),具体来源信息见表 1。所有样品经南京中医药大学药学院陈建伟教授鉴定,均为禾本科植物薏苡 *C. lacryma-jobi* L. var. *ma-yuen* (Roman.) Stapf 的干燥成熟种仁。分析前,样品均储存在室温、干燥、密封的条件下。CM-5 分光光度计(日本柯尼卡美能达公司);Heracles NEO 超快速气相色谱电子鼻(法国 Alpha MOS 公司)。

1.2 试剂

正构烷烃 $nC_6 \sim nC_{16}$ 混合对照品(批号 A0142930),购自美国 RESTEK 有限公司。

2 方法

2.1 颜色测定

使用 CM-5 分光光度计,样本观察脉冲下氙弧

表 1 26 批薏苡仁样品来源

Table 1 Sources of 26 batches of *Coicis Semen*

编号	产地	批号	编号	产地	批号
Y1	云南楚雄	221016	F3	福建浦城	220911
Y2	云南楚雄	221017	F4	福建三明	221207
Y3	云南楚雄	221115	S1	山东沂水	221202
Y4	云南西双版纳	220910	S2	山东沂水	221203
Y5	云南西双版纳	221125	S3	山东临沂	221128
J1	江苏宿迁	221209	S4	山东东营	221129
J2	江苏宿迁	221210	G1	贵州兴仁	221201
J3	江苏宿迁	221211	G2	贵州兴仁	221204
J4	江苏宿迁	221212	G3	贵州兴仁	221018
J5	江苏南京	221205	G4	贵州兴仁	220927
J6	江苏南京	221206	G5	贵州兴仁	221208
F1	福建浦城	221126	G6	贵州安顺	220918
F2	福建浦城	220907	G7	贵州安顺	220919

灯 (D65, $\phi 3\text{mm}$) 的视角设置 10° , 校准后, 选择 SCE 模式。将薏苡仁样品粉碎 (过 3 号筛), 放入检测皿中, 每个样品重复测定 2 次, 以平均值为最终测定结果。

2.2 超快速气相电子鼻分析

2.2.1 样品孵化 将薏苡仁样品粉末 (3 g) 放入 20 mL 顶空瓶中, 机械臂将顶空瓶送至孵化箱中, 在 80°C 条件下孵化 20 min, 孵化炉转速为 500 r/min, 部分气味挥发并聚集在小瓶顶部, 最终达到平衡的状态。

2.2.2 电子鼻分析方法 使用 Heracles NEO 超快速气相色谱电子鼻对薏苡仁样品进行检测分析。该仪器由自动采样装置、包含 2 根不同极性色谱柱的检测系统、2 个氢火焰离子化检测器组成。

进样体积为 5 000 μL , 进样口的温度为 200°C , 压力为 10 kPa, 体积流量为 125 $\mu\text{L/s}$ 。捕集阱初始温度为 40°C , 并以 1.0 mL/min 的速度持续 50 s。程序升温方式: 初始温度为 50°C , 先以 1°C/s 升至 80°C , 然后以 3°C/s 升至 250°C , 保持 21 s, 总采集时间为 110 s。氢火焰离子化检测器温度为 260°C , 使用体积流量为 1.0 mL/min 的氮气作为载气。每个样品重复测定 2 次。

2.2.3 气味成分的鉴定 采用正构烷烃标准溶液 ($n\text{C}_6\sim n\text{C}_{16}$) 进行校准, 将保留时间转化为保留指数, 然后通过 Kovats 保留指数定性, 在 Arochembase 数据库中对挥发性化合物进行分析鉴定。

2.3 数据分析

使用 Matlab 2021a 软件对色度值进行 DT 和 KNN 分析, 以真正率 (true positive rate, TPR) (公式 1) 和假负率 (false negative rate, FNR) (公式 2) 来评价模型的预测能力, 使用受试者工作特征曲线 (receiver operating characteristic curve, ROC 曲线) 评判模型的综合性能。

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}) \quad (1)$$

$$\text{FNR} = \text{FN} / (\text{TP} + \text{FN}) \quad (2)$$

TP 代表真实类别是正类, 预测结果也是正类; FN 代表真实类别是正类, 预测结果是负类

一般来说, TPR 的值越高, FNR 的值越低, ROC 曲线下的面积越大, 表明模型的预测能力越好。

使用 IBM SPSS Statistics 26.0 software (SPSS Inc., Chicago, IL, 美国) 建立不同产地薏苡仁的颜色参考范围和 Bayes 判别模型。使用 Simca-p 软

件 (Simca Imola S.C., Imola, Bologna, 意大利) 对电子鼻分析结果进行 PLS-DA 分析; 最后采用皮尔逊 (pearson) 检验对颜色值和挥发性成分进行关联分析。

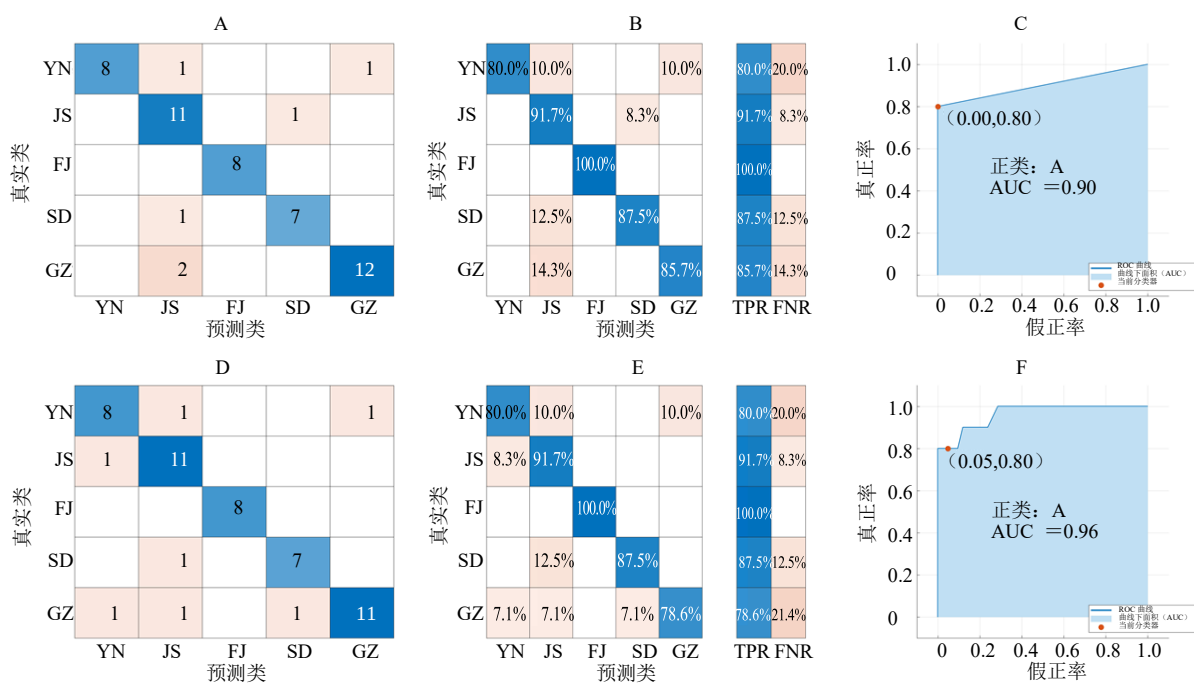
3 结果与分析

3.1 电子眼分析

3.1.1 颜色测定方法的验证 对颜色测定方法进行验证, 连续 6 次测定同一份样品, 颜色值 (L^* 、 a^* 、 b^*) 的相对标准偏差 (RSD) 均小于 1.60%, 表明该仪器有良好的精密度。分别在 0、2、4、8、12、24 h 测定同一份样品, 颜色值 (L^* 、 a^* 、 b^*) 的 RSD 均小于 2.90%, 表明该样品粉末在 24 h 内有良好的稳定性。取同一批样品, 平行制备 6 份粉末, 分别测定颜色值 (L^* 、 a^* 、 b^*), 6 组颜色值的 RSD 均小于 1.00%, 表明该方法具有良好的重复性。

3.1.2 基于机器学习算法的薏苡仁产地区分 机器学习算法 (machine learning, ML) 在图像处理和大数据分析中是一项至关重要的技术, 被广泛应用于医学、药学、经济学、生物学等多个科学领域中^[20-23]。其中最主要的分类方法是根据训练过程中得到监督的类型进行区分, 可将其分为监督学习、非监督学习、半监督学习等^[24]。决策树 (decision tree, DT) 模型是其中一种被普遍应用的监督学习方法, 原理是通过构建分类树或分类规则来对数据进行分类^[25]。谭有珍等^[26]的研究表明, 决策树模型可用于区分沉香及其混伪品, 该方法快速高效。 k 最近邻算法 (k -nearest neighbor, KNN) 模型是一种直观且容易实现的有监督分类算法, 其核心思想是通过寻找与待分类的样本在特征空间中距离最近的 k 个已标记样本, 将 k 个样本中占比最高的类别, 标记给待分类样本^[27]。

本实验基于薏苡仁外观色泽建立了 DT 和 KNN 模型, 用来区分不同产地的薏苡仁, 结果见图 1。使用混淆矩阵图将 2 种模型的区分结果进行可视化, 采用 k 折交叉验证法 ($k=10$) 进行测试, 其中 YN、JS、FJ、SD、GZ 分别代表云南、江苏、福建、山东、贵州。以 DT 模型为例, 从图 1-A 中可以看出, 云南产地的 10 个样本有 8 个被正确划分, 贵州产地 12 个样本有 11 个被正确划分, 福建产地 8 个样本均被正确划分, 山东产地 8 个样本有 7 个被正确划分, 贵州产地 14 个样本有 12 个被正确划分。由图 1-B 可知, DT 模型中各产地预测正确率 (TPR)



A-DT 模型混淆矩阵; B-DT 模型 TPR 和 FNR 值; C-DT 模型 ROC 曲线; D-KNN 模型混淆矩阵; E-KNN 模型 TPR 和 FNR 值; F-KNN 模型 ROC 曲线。

A-obfuscation matrix of DT model; B-TPR and FNR of DT model; C-ROC curve of DT model; D-obfuscation matrix of KNN model; E-TPR and FNR of KNN model; F-ROC curve of KNN model.

图 1 基于颜色参数的 DT 和 KNN 模型

Fig. 1 DT and KNN models based on color parameters

分别为 80.0%、91.7%、100%、87.5%、85.7%，均高于 80.0%，而 FNR 均小于 20.0%，表明 DT 具有较强的区分能力。图 1-D、E 显示，KNN 模型中各产地的 TPR 均高于 75.0%，而 FNR 均小于 25.0%，表明 KNN 同样具备较强的区分能力。从图 1-C 和 F 可以得知，2 种模型的 ROC 曲线下面积均大于 0.8，表明模型预测能力可靠。

3.1.3 颜色参考范围的建立 虽然 DT 和 KNN 模型能够较好区分薏苡仁的产地，但是无法客观量化不同产地的颜色差异。因此，在 DT 与 KNN 分析的基础上，使用 SPSS 26.0 软件，以 95% 为置信水平，选择均值建立不同产地薏苡仁的颜色参考范围，结果见表 2。

采用单因素方差分析验证所建立的颜色参考范围的合理性，结果见表 3。结果显示，不同产地 L^* 、 a^* 、 b^* 的 P 值均 < 0.01 ，表明 5 个产地薏苡仁的颜色参考范围具有显著性差异，所建立的颜色参考范围能够进行有效区分。

3.1.4 不同产地薏苡仁快速判别模型的建立 在客观量化不同产地薏苡仁的颜色差异后，将薏苡仁样品的色度值录入 SPSS 26.0 软件，采用判别分析中

表 2 不同产地薏苡仁颜色参考范围

Table 2 Reference range of color value of coicis semen from different habitats

产地	参考值			
	L^*	a^*	b^*	平均值
云南	82.10~82.91	1.21~1.57	11.18~12.46	82.51, 1.39, 11.82
江苏	83.44~83.76	1.12~1.23	11.24~11.57	83.60, 1.17, 11.41
福建	83.98~90.30	0.44~1.10	8.84~10.80	87.14, 0.77, 9.82
山东	82.83~83.21	1.12~1.30	11.53~12.02	83.02, 1.21, 11.77
贵州	83.02~83.17	1.20~1.24	11.54~11.78	83.09, 1.22, 11.66

表 3 不同产地薏苡仁颜色值方差分析结果

Table 3 Results of variance analysis of color value of coicis semen from different habitats

颜色值	组间			组内均方
	均方	自由度 (F)	显著性 (P)	
L^*	99.074	6.660	7.28×10^{-4}	100.411
a^*	2.367	13.739	3.00×10^{-6}	1.163
b^*	25.895	17.204	3.65×10^{-3}	10.160

的典型判别函数-Bayes 判别法建立数学判别模型，快速判别薏苡仁的产地来源。判别函数式如下：

$$Y_{\text{云南}} = 12.015 L^* - 386.693 a^* + 199.026 b^* - 2\ 776.941$$

$$Y_{\text{江苏}} = 45.588 L^* - 406.540 a^* + 203.829 b^* - 2\ 831.162$$

$$Y_{\text{福建}} = 46.635 L^* - 388.272 a^* + 194.427 b^* - 2\ 838.484$$

$Y_{\text{山东}}=45.578 L^*-410.911 a^*+206.105 b^*-2\ 851.388$

$Y_{\text{贵州}}=45.417 L^*-408.768 a^*+205.053 b^*-2\ 828.340$

将未知样本的颜色值代入上述判别函数计算 Y ，并以最大 Y 确定未知样本的处理程度：如果 $Y_{\text{云南}}$ 最大，则样本为云南产地的薏苡仁；如果 $Y_{\text{江苏}}$ 最大，则为江苏产地的薏苡仁；如果 $Y_{\text{福建}}$ 最大，则为福建

产地的薏苡仁；如果 $Y_{\text{山东}}$ 最大，则为山东产地的薏苡仁；如果 $Y_{\text{贵州}}$ 最大，则为贵州产地的薏苡仁。最后，采用初始和交叉验证的方法评估所建模型的预测能力，结果见表 4。在交叉验证实验中，各产地交叉验证的正确率均为 100%，表明该模型预测能力较强，可用于快速判别不同产地的薏苡仁。

表 4 不同产地薏苡仁交互验证

Table 4 Interactive verification results of coicis semen from different habitats

验证方式	产地	预测组成员计数					正确率/%
		云南	江苏	福建	山东	贵州	
初始	云南	5	0	0	0	0	100
	江苏	0	6	0	0	0	100
	福建	0	0	4	0	0	100
	山东	0	0	0	4	0	100
	贵州	0	0	0	0	7	100
交叉验证	云南	5	0	0	0	0	100
	江苏	0	6	0	0	0	100
	福建	0	0	4	0	0	100
	山东	0	0	0	4	0	100
	贵州	0	0	0	0	7	100

3.2 超快速气相电子鼻分析

3.2.1 超快速气相电子鼻检测条件的优化 为了得到良好的分析效果，通过单因素实验对气相电子鼻主要参数进行了优化：样品用量（1、2、3、4 g）、进样量（1 000、2 000、3 000、4 000、5 000 μL ）、孵化温度（60、65、70、75、80 $^{\circ}\text{C}$ ）、孵化时间（15、20、25、30 min）。结果表明，当样品用量为 3 g，进样量为 5 000 μL ，孵化温度为 80 $^{\circ}\text{C}$ ，孵化时间为 20 min 时，色谱图信息良好。

3.2.2 超快速气相电子鼻检测方法的验证 连续 6 次检测同一批次样品，5 个特征峰的保留时间 RSD 均 $<0.30\%$ ，峰面积 RSD 均 $<4.80\%$ ，表明该方法有

良好的精密度。取同一批样品，平行制备 6 份粉末，进样检测，5 个特征峰的保留时间 RSD 均 $<0.40\%$ ，峰面积 RSD 均 $<4.90\%$ ，表明该方法有良好的重复性。分别在 0、2、4、8、12、24 h 检测同一批次样品，5 个特征峰的保留时间 RSD 均 $<0.40\%$ ，峰面积 RSD 均 $<5.00\%$ ，表明该样品粉末在 24 h 内有良好的稳定性。

3.2.3 判别因子分析与热图分析 使用超快速气相电子鼻对不同产地薏苡仁的气味进行分析，结果见图 2。综合 MTX-5 和 MTX-1701，从整体气味信息来看，云南、贵州、福建产地的薏苡仁气味强度明显高于江苏和山东产地的薏苡仁。

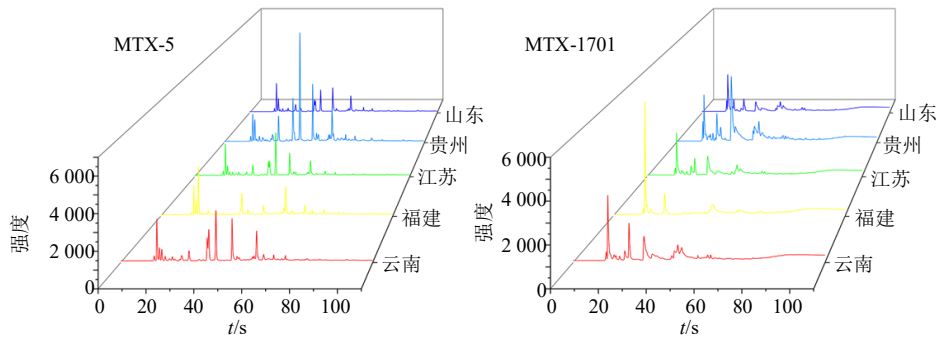


图 2 不同产地薏苡仁气味指纹图谱

Fig. 2. Odor fingerprints of coicis semen from different habitats

为了进一步探究不同产地薏苡仁的气味差异,根据电子鼻检测结果,进行判别因子分析,结果见图 3-A、图 3-B。提取贡献率最大的 3 个判别因子,贡献率分别为 87.034%、9.922%、1.800%,累积贡献率 98.756%,这说明 DFA 模型能充分体现出各个产地薏苡仁样品的气味信息和特征。由图 3-A、B 可知,各产地薏苡仁样本距离较远,表明不同产地的薏苡仁在气味信息方面存在较明显的差异。

在 DFA 分析基础上,将不同产地薏苡仁的气味强度差异进行可视化。根据超快速气相电子鼻的分析结果,将每个成分的最大峰面积设置为 100%,并对其余色谱峰进行归一化处理,进行热图分析,结果如图 3-C 所示。根据图 3-C 可知,不同产地的薏苡仁在气味强度方面存在明显的差异。其中福建产地的薏苡仁气味强度最高,其次为贵州和云南产地的薏苡仁,山东和江苏产地的薏苡仁气味强度较弱,这与气味指纹图谱的分析结果相一致。

3.2.4 气味成分鉴定与差异标志物确定 通过计算不同产地薏苡仁共有峰的 Kovats 保留指数,将其与 Arochembase 数据库进行比较,确定各个产地薏苡仁的挥发性成分,结果见表 5。

为了寻找不同产地薏苡仁的气味差异标志物,使用 simca-p 软件进行 PLS-DA 分析,结果如图 4 所示。由得分矩阵图可知,PLS-DA 模型中所有数据点均在 95%置信区间内,聚类效果良好,区分明显,说明不同产地的薏苡仁样品在气味化学成分上存在着一定的差异性。启动 loading 程序,得到荷载散点图,结果如图 4-B 所示。由图 4-B 可知,丙烯醛、丙醛、2-丙烯酸、甲苯、2-己醇等成分距离原点中心较远,表明这些成分对分组贡献度较大。PLS-DA 模型解释率参数 R^2_X 为 0.721,明显高于 0.5,说明模型解释率较高。设置分类 Y 矩阵变量随机排列 200 次做置换检验,结果如图 4-C 所示。 R^2 拟合直线在 Y 坐标轴的截距小于 0.3,说明该模型较为可靠, Q^2 拟合直线在 Y 坐标轴的截距小于 0.05,说明该模型不存在过度拟合的现象。

变量重要性投影值 (VIP) 可以进一步直观地反映具体变量对模型分类的整体贡献度,在 PLS-DA 模型中,变量的 VIP 值大于 1.0 则说明该变量对整体模型的贡献度高于平均水平,对样本分类的影响具有统计学意义,可以作为差异标志物。使用 simca-p 软件,启动 VIP 分析程序,结果如图 4-D 所示。

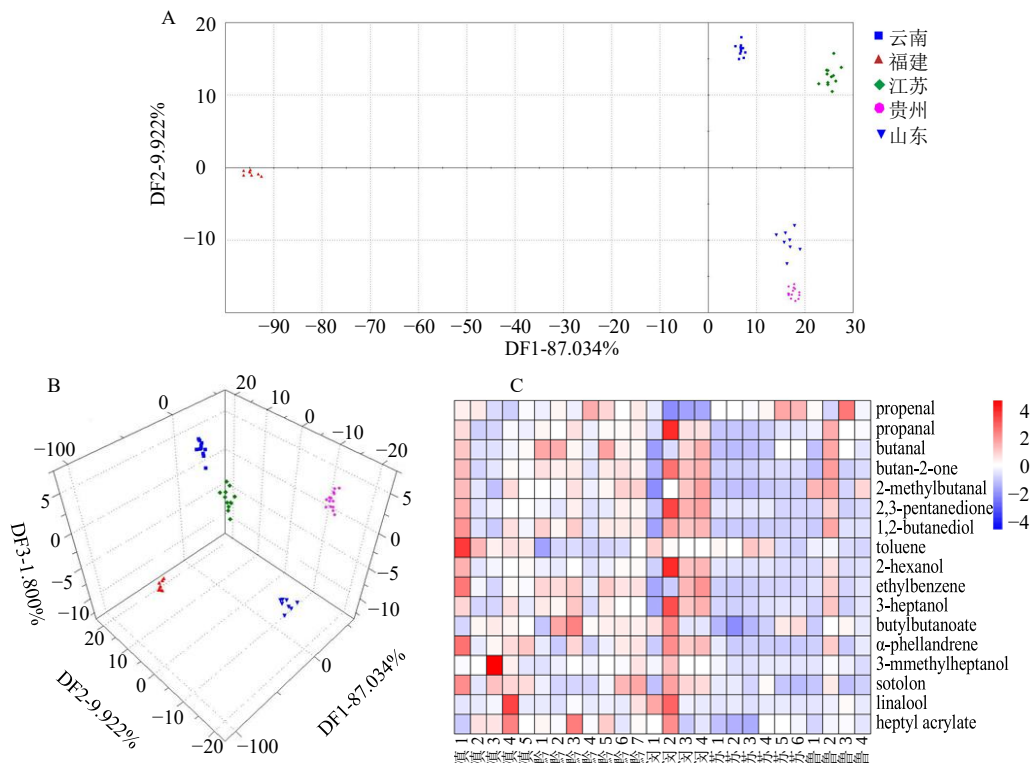
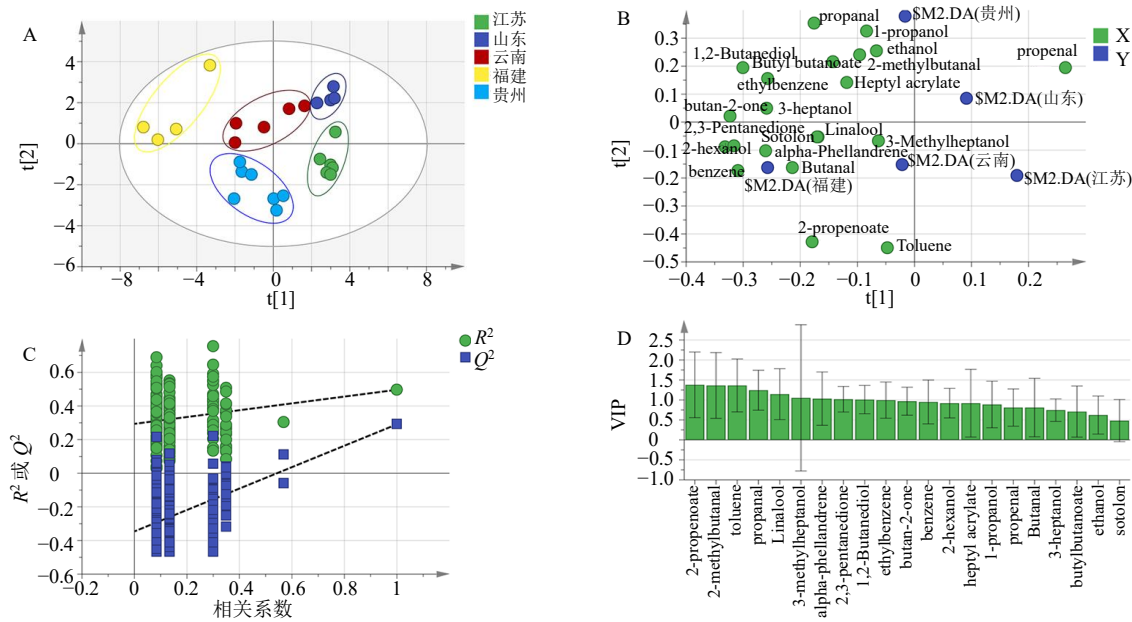


图 3 不同产地薏苡仁电子鼻 DFA 得分图 (A、B) 与热图 (C)

Fig. 3 DFA score plots (A, B) and heatmap (C) of coicis semen from different habitats

表 5 不同产地薏苡仁中挥发性化合物及感官描述信息

序号	分子式	RI		可能的化合物	相关指数	气味信息
		MXT-5	MXT-1701			
1	C ₂ H ₆ O	467	583	乙醇	64.83	酒精; 乙醇; 辛辣的; 甜的
2	C ₃ H ₄ O	467	584	丙烯醛	65.39	杏仁; 樱桃
3	C ₃ H ₆ O	493	563	丙醛	81.06	乙醛; 可可; 辛辣的
4	C ₄ H ₈ O	552	698	正丁醛	88.69	巧克力; 可可; 霉味; 辛辣的
5	C ₄ H ₈ O	600	698	丁二酮	68.07	丙酮; 奶酪; 巧克力; 芳香的; 辛辣的; 甜的
6	C ₅ H ₁₀ O	662	750	2-甲基丁醛	95.53	杏仁; 苹果; 可可; 咖啡; 麦芽; 霉味; 酸味
7	C ₅ H ₈ O ₂	700	802	2,3-乙酰基丙酮	92.20	杏仁; 苹果; 黄油; 奶油糖果; 焦糖; 奶酪; 辛辣的; 甜的
8	C ₄ H ₁₀ O ₂	769	947	1,2-丁二醇	98.45	酒精; 黄油
9	C ₇ H ₈	775	824	甲苯	93.45	焦糖; 辛辣的; 甜的
10	C ₆ H ₁₄ O	802	913	2-己醇	97.50	花椰菜; 水果; 葡萄酒
11	C ₈ H ₁₀	872	913	乙苯	89.47	芳族; 甜的
12	C ₇ H ₁₆ O	894	968	1-丙醛	60.88	酒精
13	C ₇ H ₁₆ O	903	997	3-己醇	89.14	绿色; 草药的
14	C ₃ H ₈ O ₃	975	1 281	2-丙烯酸	84.71	杏仁; 苦味; 淡味的; 无味的
15	C ₆ H ₆ O ₂	983	1 164	苯	90.29	酸性的; 杏仁; 焦糖; 辣
16	C ₈ H ₁₆ O ₂	996	1 078	丁酸丁酯	88.12	香蕉; 樱桃; 绿色; 菠萝; 甜的
17	C ₁₀ H ₁₆	1 015	1 031	α -水芹烯	52.99	桔子; 薄荷味; 松脂; 木头味
18	C ₈ H ₁₈ O	1 037	1 164	3-甲基庚醇	91.69	/
19	C ₁₀ H ₁₈ O	1 107	1 206	芳樟醇	86.14	茴香; 香柠檬; 桔子; 芳香的; 百合; 辣; 甜的;
20	C ₁₀ H ₁₈ O ₂	1 196	1 281	丙烯酸庚酯	86.69	/
21	C ₆ H ₈ O ₃	1 107	1 345	索托隆	86.82	焦糖; 咖啡; 棉花糖; 甜的



A-得分散点图; B-荷载散点图; C-置换检验图; D-VIP 值图。

A-score plot; B-loading scatter plot; C-permutation test plot; D-VIP plot.

图 4 不同产地薏苡仁的 PLS-DA 模型图

Fig. 4 PLS-DA model of coicis semen from different habitats.

由图 4-D 可知, VIP 值大于 1 的气味成分有 9 个, 分别为 2-丙烯酸、2-甲基丁醛、甲苯、丙醛、芳樟醇、3-甲基庚醇、 α -水芹烯、2,3-乙酰基丙酮、1,2-丁二醇。其中 2-丙烯酸、2-甲基丁醛、甲苯、丙醛 4 种成分的 VIP 值大于 1.2, 可视为主要的气味差异标志物。对不同产地薏苡仁的上述 4 种成分进行单因素方差分析, 结果表明 P 值均小于 0.05, 具有显著性差异。因此, 确定上述 4 种成分作为不同产地薏苡仁的气味差异标志物。

3.3 颜色-气味相关性分析

在 SPSS 26.0 软件中, 采用 Pearson 相关性分析, 将表面色度值 (L^* 、 a^* 、 b^*) 与内在气味成分进行相关性分析, 旨在为薏苡仁“辨状论质”提供科学依据, 结果见表 6 和图 5。在图 5 中, 红色代表正相关, 绿色代表负相关, 圆圈大小表示相关性强弱。一般认为, 相关系数的绝对值大于 0.5 可认为两者存在显著相关性。根据表 6 和图 5 可知, 丁二酮、2,3-乙酰基丙酮、3-己醇与 L^* 值呈正相关且与 a^* 值、 b^* 值呈负相关; 丙醛与 L^* 值呈负相关且与 b^* 值呈正相关; 2-丙烯酸与 2-己醇与 L^* 值呈正相关且与 b^* 值呈负相关; 1,2-丁二醇与 L^* 值呈正相关。上述成分的 P 值均小于 0.05, 具有统计学意义。

表 6 不同产地薏苡仁表面颜色与内在气味成分的相关性

Table 6 Correlation between surface color and internal odor components of coicis semen from different habitats

成分	L^*	a^*	b^*
乙醇	0.225	-0.239	-0.193
丙烯醛	-0.517**	0.425**	0.505**
丙醛	0.270	-0.176	-0.153
正丁醛	0.441*	-0.192	-0.399*
1-丙醛	-0.142	0.030	0.102
丁二酮	0.629**	-0.537**	-0.592**
2-丙烯酸	0.552**	-0.430*	-0.561**
苯	0.431*	-0.331	-0.465*
2-甲基丁醛	0.127	-0.01	-0.001
2,3-乙酰基丙酮	0.677**	-0.514**	-0.623**
1,2-丁二醇	0.578**	-0.409*	-0.428*
甲苯	0.024	0.053	-0.104
2-己醇	0.571**	-0.426*	-0.543**
乙苯	0.247	-0.235	-0.241
3-己醇	0.693**	-0.615**	-0.596**
丁酸丁酯	0.341	-0.215	-0.253
α -水芹烯	0.420*	-0.251	-0.358
3-甲基庚醇	-0.126	0.054	-0.020
芳樟醇	0.165	-0.003	-0.123
索托隆	0.457*	-0.138	-0.255
丙烯酸庚酯	0.237	-0.069	-0.069

* $P<0.05$, ** $P<0.01$ 。

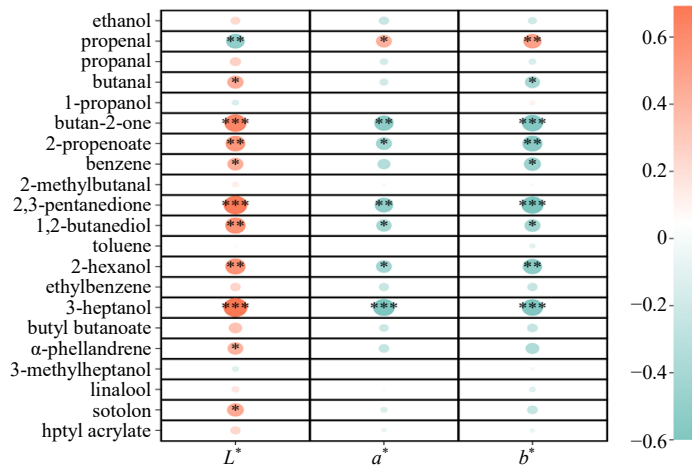


图 5 不同产地薏苡仁颜色与气味成分相关性热图

Fig. 5 Correlation heatmap between color and odor components of coicis semen from different habitats

4 讨论

颜色和气味是中药材质量识别的重要指标, 但传统的经验鉴别容易受到主观因素的影响。近年来, 由于性状电子检测技术具有快速、准确、便捷的特点, 能够将传统的外观性状进行客观量化, 在中药质量控制方面备受关注。本研究首次采用分光测色技术联用超快速气相电子鼻技术, 对不同产地的薏

苡仁药材进行快速鉴别研究, 并探究薏苡仁外在颜色和内在气味成分的关联性。在颜色测定方面, DT 和 KNN 模型和 Bayes 判别式均能够有效区分所有产地的薏苡仁。超快速气相电子鼻分析结果显示, 同产地的薏苡仁药材之间相似度较高, 整体气味信息类似, 而不同产地薏苡仁的气味成分则存在较大的差异。同时采用 PLS-DA 模型成功筛选出 4 种成

分作为鉴别不同产地薏苡仁的气味差异标志物。最后,对颜色和气味进行相关性分析,结果显示薏苡仁的色泽与气味成分之间存在显著关联性。因此,采用分光测色计、超快速气相电子鼻结合化学计量学快速鉴别不同产地薏苡仁是一种快速准确的手段,该方法在多产地中药的质量控制方面具有巨大的应用前景^[28]。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国药典 [S]. 一部. 2020: 393-394.
- [2] 马明慧, 刘秀峰, 余伯阳. 薏苡仁的本草考证 [J]. 中国民族民间医药, 2021, 30(19): 32-37.
- [3] Zhu F. *Coix*: Chemical composition and health effects [J]. *Trends Food Sci Technol*, 2017, 61: 160-175.
- [4] 王金丹. 薏苡仁中有效成分提取工艺的研究发展 [J]. 农业科技与装备, 2021(2): 43-45.
- [5] 毕天琛, 杨国宁, 马海春. 中药薏苡仁化学成分及药理活性研究进展 [J]. 海峡药学, 2019, 31(11): 52-56.
- [6] 罗云云, 杜伟锋, 应泽茜, 等. 响应面法协同正交试验设计优化薏苡仁多糖提取工艺研究 [J]. 中华中医药杂志, 2019, 34(10): 4847-4851.
- [7] 蒋星月, 张喻惠, 高嵩, 等. 薏苡仁多糖的药理作用及应用研究进展 [J]. 中药与临床, 2022, 13(5): 135-139.
- [8] 吕峰. 我国薏苡仁资源主要品质及薏苡仁活性多糖的研究 [D]. 福州: 福建农林大学, 2008.
- [9] 崔媛, 王小明, 杨勇, 等. 薏苡仁油融合指纹图谱研究 [J]. 中草药, 2014, 45(12): 1698-1701.
- [10] 李厚聪, 刘圆, 袁玮, 等. RP-HPLC 法测定薏苡中薏苡素的含量 [J]. 西南大学学报: 自然科学版, 2009, 31(11): 154-157.
- [11] 郑利, 陈丹, 曾令军, 等. UPLC-MS 测定不同产地薏苡仁中甘油三油酸酯的含量 [J]. 中国现代应用药学, 2014, 31(2): 200-204.
- [12] 任江剑, 俞旭平, 张斌, 等. 不同种源薏苡仁中甘油三油酸酯含量比较 [J]. 中国现代中药, 2011, 13(4): 15-17.
- [13] 张礼欣, 于小聪, 王泽琨, 等. 基于气味信息的荆芥穗产地鉴别 [J]. 中国药房, 2021, 32(18): 2203-2209.
- [14] 刘梦楚, 邹晓红, 蓝伦礼, 等. 基于电子鼻及顶空-气质联用技术结合化学计量学区分不同产地的砂仁 [J]. 中国实验方剂学杂志, 2017, 23(6): 35-42.
- [15] 卢一. 基于“气味”信息分析的中药饮片快速鉴别研究 [D]. 成都: 成都中医药大学, 2017.
- [16] Fei C H, Ren C C, Wang Y L, *et al*. Identification of the raw and processed *Crataegi Fructus* based on the electronic nose coupled with chemometric methods [J]. *Sci Rep*, 2021, 11(1): 1849.
- [17] 梅茜, 许金国, 苏联麟, 等. 基于 Heracles NEO 超快速气相电子鼻对枳壳麸炒前后气味差异标志物的快速识别研究 [J]. 中草药, 2023, 54(16): 5165-5171.
- [18] Deng C, Xue R, Wang J, *et al*. Discrimination between *Zingiberis Rhizoma Praeparatum* and carbonised ginger by colour measurement and fingerprint analysis [J]. *Phytochem Anal*, 2021, 32(6): 921-931.
- [19] 李亚飞, 汤璐璐, 赵明方, 等. 基于 Heracles NEO 超快速气相电子鼻麸炒白术炮制全过程气味变化识别研究 [J]. 中草药, 2023, 54(15): 4812-4822.
- [20] 韩丰, 龙明盛, 李月安, 等. 循环神经网络在雷达临近预报中的应用 [J]. 应用气象学报, 2019, 30(1): 61-69.
- [21] 陆虹, 翟盘茂, 覃卫坚, 等. 低温雨雪过程的粒子群-神经网络预报模型 [J]. 应用气象学报, 2015, 26(5): 513-524.
- [22] Cramer S, Kampouridis M, Freitas A A, *et al*. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives [J]. *Expert Syst Appl*, 2017, 85: 169-181.
- [23] 王在文, 郑祚芳, 陈敏, 等. 支持向量机非线性回归方法的气象要素预报 [J]. 应用气象学报, 2012, 23(5): 562-570.
- [24] Douglass M J J. Book Review: Hands-on machine learning with Scikit-Learn, Keras, and Tensorflow, 2nd edition by Aurélien Géron [J]. *Phys Eng Sci Med*, 2020, 43(3): 1135-1136.
- [25] 魏向祎, 孔令然, 肖磊. 基于决策树的京津冀地区 Landsat 图像分类 [J]. 现代信息科技, 2023, 7(5): 41-44.
- [26] 谭有珍, 陈碧莹, 吴霞, 等. 基于 ¹H-NMR 指纹图谱建立沉香快速识别模型的研究 [J]. 中药材, 2021, 44(4): 818-823.
- [27] Banerjee M B, Roy R B, Tudu B P, *et al*. Black tea classification employing feature fusion of E-Nose and E-Tongue responses [J]. *J Food Eng*, 2019, 244: 55-63.
- [28] Liu X, Jiang W, Su M, *et al*. Quality evaluation of traditional Chinese medicines based on fingerprinting [J]. *J Separat Sci*, 2020, 43(1): 6-17.

[责任编辑 时圣明]