

基于颜色气味信息融合的苦杏仁走油程度判别分析与质量预测模型建立

欧阳少琴¹, 陈慧荣², 拱健婷³, 王亚顺¹, 崔阳¹, 翟恩爱¹, 梁敬妮¹, 邹慧琴^{1*}, 闫永红^{1*}

1. 北京中医药大学中药学院, 北京 102488
2. 北京海金格医药科技股份有限公司, 北京 100053
3. 北京市中医药研究所, 北京 100005

摘要: **目的** 建立基于颜色气味数字化融合信息的苦杏仁 *Armeniaca Semen Amarum* 走油程度判别分析和内在质量预测模型。**方法** 基于颜色气味数字化的融合信息, 联合机器学习算法对其走油程度进行判别, 对比其正判率, 寻找识别效果最好的算法。利用 SPSS 分析平台, 开展苦杏仁颜色气味融合信息数字化与内在化学成分的相关性分析, 同时建立含量预测回归方程并且检验其拟合度。**结果** 基于粉末颜色和气味融合信息的模型建立中, Logistic、IBK、K-Star、LMT 和 Random Forest 算法正判率较高, 可完成对不同走油程度苦杏仁的分类鉴别; 基于剖面颜色和气味融合信息的模型建立中, Logistic 算法和 K-Star 算法可完成走油程度的判定。基于粉末颜色及气味融合信息建立苦杏仁质量预测模型, 预测方程: $Y_{\text{苦杏仁苷}} = 2.175 - 1.340 F_{1-2} + 0.529 F_{1-1}$, $R^2 = 0.732$; $Y_{\text{酸值}} = 2.113 + 1.7249 F_{1-2} - 0.667 F_{1-1}$, $R^2 = 0.719$; 基于剖面颜色及气味融合信息建立苦杏仁质量预测模型, 预测方程如下: $Y_{\text{苦杏仁苷}} = 2.153 + 1.242 F_{2-2} + 0.5 F_{2-1} - 0.689 F_{2-3}$, $R^2 = 0.775$, $Y_{\text{酸值}} = 2.226 - 1.946 F_{2-2} - 0.785 F_{2-1}$, $R^2 = 0.738$, 拟合度结果优良。**结论** 通过颜色气味数字化信息可快速推断苦杏仁化学成分的变化趋势, 颜色气味融合信息测量可发展为苦杏仁质量评价的新方法。

关键词: 苦杏仁; 走油; 气味数字化; 颜色数字化; 信息融合; 质量评价

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2023)12-3806-09

DOI: 10.7501/j.issn.0253-2670.2023.12.007

Study on discriminative analysis and quality prediction models of *Armeniaca Semen Amarum* with rancidness degrees based on information fusion of color and odor

OUYANG Shao-qin¹, CHEN Hui-rong², GONG Jian-ting³, WANG Ya-shun¹, CUI Yang¹, ZHAI En-ai¹, LIANG Jing-ni¹, ZOU Hui-qin¹, YAN Yong-hong¹

1. School of Traditional Chinese Medicine, Beijing University of Traditional Chinese Medicine, Beijing 102488, China
2. Beijing Haikinger Pharmaceutical Technology Co., Beijing 100053, China
3. Beijing Institute of Traditional Chinese Medicine, Beijing 100005, China

Abstract: Objective To establish discriminative analysis and quality prediction models of Kuxingren (*Armeniaca Semen Amarum*, ASA) with different rancidness degrees based on fusion information of color and odor digital values. **Methods** Firstly different models were established by various classifiers to discriminate different rancidness degrees of ASA samples with positive judgment rate as evaluation index. And then the optimal algorithm was screened out based on fusion information of color and odor digital values. Using SPSS analysis platform, the correlation analysis between the digitalized odor and color fusion information and the chemical composition of ASA samples was carried out. Afterwards the content prediction regression equation was established and its fit was checked. **Results** In the model building based on the fusion information of powder color and odor, four kinds of algorithms, namely Logistic, IBK, K-Star, LMT and Random Forest, possess higher positive judgment rate and could complete the classification and identification of ASA samples with different rancidness degrees; In the model building based on the fusion information of longitudinal

收稿日期: 2022-11-07

基金项目: 国家自然科学基金面上项目 (81573542); 国家自然科学基金青年科学基金项目 (81403054)

作者简介: 欧阳少琴 (1999—), 女, 硕士研究生, 研究方向为中药质量评价。E-mail: oysq18801399681@163.com

*通信作者: 邹慧琴, 副研究员, 硕士研究生导师, 主要从事中药战略情报分析与中药智能仿生鉴别研究。E-mail: zouhuiqin_bucm@sina.cn
闫永红, 教授, 博士研究生导师, 主要从事中药品质评价研究。E-mail: lxdyyh@yeah.net

section color and odor, Logistic algorithm and K-Star algorithm could complete the determination of rancidness degrees. The prediction models of ASA quality were established based on the fusion information of powder color and odor, and the prediction equations were as follows: $Y_{\text{amygdalin}} = 2.175 - 1.340 F_{1-2} + 0.529 F_{1-1}$, $R^2 = 0.732$; $Y_{\text{acid value}} = 2.113 + 1.724 F_{1-2} - 0.667 F_{1-1}$, $R^2 = 0.719$; based on the fusion information of longitudinal section color and odor ASA quality prediction models were established with the following prediction equations: $Y_{\text{amygdalin}} = 2.153 + 1.242 F_{2-2} + 0.5 F_{2-1} - 0.689 F_{2-3}$, $R^2 = 0.775$, $Y_{\text{acid value}} = 2.226 - 1.946 F_{2-2} - 0.785 F_{2-1}$, $R^2 = 0.738$, with excellent fit results. **Conclusions** The color-odor digital information was used to profile the trend of chemical composition of ASA samples, and the color-odor fusion information measurement could be used for the quality evaluation of ASA samples.

Key words: *Armeniaca Semen Amarum*; rancidness; odor digitization; color digitization; information fusion; quality evaluation

质优、稳定、可控的中药材及其中药饮片是中医临床疗效以及临床用药安全的重要保障,而中药质量与贮藏条件密切相关。中药材在贮藏过程中,受自然因素(湿度,温度)及中药本身所含成分性质影响发生变质现象。其中走油是中药饮片普遍存在的问题。“走油”,又称“泛油”,是指某些药材油脂溢出药材表面,或因药材受潮、变色、变质后表面溢出油样物质^[1]。药材走油后会导致疗效下降,甚至会引发毒性反应^[2]。

目前,判断药材走油程度的方法主要是传统感官鉴别法和理化测定法。传统感官鉴别法操作简单、方便、容易实现,但该方法存在主观性强、鉴定者需要积累丰富经验等局限性。理化测定法客观性强,结果准确可重复,但耗费时间长,检验流程复杂,检测费用高,且单一成分难以反映中药的整体性和多元性。基于此,以上2种方法均无法对走油药材的质量进行客观、全面、快速的评价。随着现代仿生技术的不断发展及其与中药鉴定学科的交叉结合,色度仪、电子鼻等仿生技术被迅速应用于中药品质评价研究。色度仪是基于三滤色片原理,利用仪器内部的标准光源照明样本,样本选择性吸收反射光线,光电探测器检测反射光并与标准光源作比较并计算出颜色坐标^[3]。智能感官仪器电子鼻,通过传感器阵列模拟人和动物的嗅觉细胞,分析、识别和检测复杂的气味和挥发性成分^[4]。电子感官技术能够对颜色气味量化、客观化,对有内涵、有经验的传统感官鉴别是一种有力的、科学的补充,同时也是可持续发展的重要研究课题。

苦杏仁 *Armeniaca Semen Amarum* 是常用大宗中药,苦杏仁中含丰富脂肪油(占总量的40%~56%)^[5],是易走油变质的典型代表。研究发现,苦杏仁走油后种皮颜色会由黄白变棕黄,子叶颜色由乳白色不断加深至黄棕色,伴有哈喇味^[6],药效成分苦杏仁苷含量下降,油脂发生氧化酸败,脂肪酸值与过氧化值均显著增高,脂肪酸种类与比例发

生变化,产生毒性成分^[7-8]。针对走油苦杏仁建立一种快速、精准的分析手段和质量评价方法是一项重要研究课题。本课题组在苦杏仁走油快速鉴别研究领域已经积累了研究基础,并且引入现代仿生技术电子鼻和色度仪使中药颜色、气味的表述不再是主观性、抽象化的定性描述,而是数字化、客观化的定量描述。但目前研究大多局限在一种或一类化合物的检测,不能代表苦杏仁的整体品质。而单一的“颜色-质量”或“气味-质量”研究来检测走油苦杏仁的质量存在局限性。

因此,本研究以苦杏仁颜色、气味指标值为参数,依据其与苦杏仁走油过程中化学成分的变化规律,探讨颜色与气味参数之间的信息融合体系,建立基于颜色气味信息融合的苦杏仁走油监测模型,为苦杏仁走油的在线准确监测提供新思路。

1 仪器与材料

1.1 仪器

α -FOX3000型电子鼻,法国Alpha M.O.S.公司;FW-135型中草药万能粉碎机,天津泰斯特仪器有限公司;BS-124S型电子分析天平,德国Sartorius公司;3010型紫外可见分光光度计,日本日立公司;ZN-02型小型粉碎机,北京兴时利和科技发展有限公司;CM-5型分光光度计,日本东京柯尼卡美能达控股有限公司;Waters 2695型高效液相色谱仪,美国Waters公司;Zorbax SB-C₁₈色谱柱(250 mm×4.6 mm, 5 μ m),美国安捷伦公司;药典筛,绍兴市上虞圣超仪器设备有限公司。

1.2 试剂与样品

苦杏仁苷对照品,批号ST02010120MG,质量分数93.4%,上海诗丹德标准技术服务有限公司;甲醇,分析纯/色谱纯,赛默飞世尔科技(中国)有限公司;去离子水,北京中医药大学中药学院。从内蒙古和河北采集了苦杏仁生品、焯品和炒品,经北京中医药大学中药资源与鉴定系闫永红教授按照《中国药典》2020年版一部^[9]有关项下进行鉴定为薺

薇科杏属植物山杏 *Prunus armeniaca* L. var. *ansu* Maxim. 的干燥成熟种子。

2 方法与结果

2.1 不同走油程度苦杏仁样品的准备

取净苦杏仁，置沸水中略烫，至外皮微胀时，捞出，用凉水稍浸，取出搓开种皮，晒干后簸去种皮，取仁制得焯苦杏仁样品。

取焯苦杏仁，置炒制容器内，用文火加热，炒至表面黄色，取出晾凉制得炒苦杏仁样品。

将样品混合均匀，分别采用自然留样法和加速试验法制备不同走油程度苦杏仁样品。自然留样法即将样品放置于北京室温 [温度 (25±2) °C]、通风环境 [相对湿度 (60±10)%] 中存储，每3个月采样1次；加速试验法即将样品放置于药物稳定性试验箱高温高湿环境 (温度 40 °C，相对湿度 75%) 的存储添加中，每10天采样1次。样品信息见表1。采用传统感官鉴别法对苦杏仁样品的走油程度划分等级^[10]。

2.2 不同走油程度样品图的划分结果

按“2.1”项下方法，制备不同走油程度苦杏仁样品，并通过传统感官鉴别划分4个等级。I~IV表示走油程度逐级加深，I为正常苦杏仁样品，IV为重度走油苦杏仁样品。结果见图1。

2.3 不同走油程度苦杏仁颜色测定

2.3.1 粉末颜色信息测量 将3种炮制品的苦杏仁样品去皮粉碎，过2号筛，混合均匀，取适量粉末



图1 不同走油程度 (I~IV) 样品图

Fig. 1 Sample chart of different rancid degree (I—IV)

装入石英测色皿中。检测条件：起止波长 380~780 nm，扫描速度 600 nm/min，狭缝宽度 1 nm，光源 D65，视场 10°。

2.3.2 剖面颜色信息测量 将苦杏仁沿种脊切成两半，用 CM-5 分光测色仪测定苦杏仁切片的颜色。测量条件：起止波长 360~740 nm，光源 D65，视场 10°，照明系统：SCI 反射，测试区域 3 mm。

2.4 不同走油程度苦杏仁气味测定

将样品粉碎，过2号筛，准确称定质量 0.30 g，

表1 样品信息

Table 1 Sample information

存储条件	放置时间	样品编号		
		生品	焯品	炒品
自然留样法	0月	S1	D1	C1
	3月	S2	D2	C2
	6月	S3	D3	C3
	9月	S4	D4	C4
	12月	S5	D5	C5
	24月	S6	D6	C6
	27月	S7	D7	C7
	30月	S8	D8	C8
	33月	S9	D9	C9
	36月	S10	D10	C10
加速试验法	10 d	S11	D11	C11
	20 d	S12	D12	C12
	30 d	S13	D13	C13
	40 d	S14	D14	C14
	50 d	S15	D15	C15
	60 d	S16	D16	C16
	70 d	S17	D17	C17
	80 d	S18	D18	C18
	90 d	S19	D19	C19
	100 d	S20	D20	C20
	110 d	S21	D21	C21
	120 d	S22	D22	C22

放入顶空进样瓶中，用压盖密封。测量条件：孵化时间 600 s，孵化温度 45 °C，振荡速度 250 r/s，进样量 2500 μL，载气为压缩空气，体积流量 150 mL/min，数据采集时间 120 s，采集周期 1 s，吹扫时间 600 s。

2.5 苦杏仁苷的测定

2.5.1 色谱条件 色谱柱为 Zorbax SB-C₁₈ 柱 (250 mm×4.6 mm, 5 μm)；流动相为甲醇-水 (20:80)；体积流量为 1 mL/min；进样量 10 μL；检测波长 207

nm。

2.5.2 对照品溶液的制备 取苦杏仁苷对照品适量,精密称定,置棕色瓶中,加甲醇定容,制成含苦杏仁苷 203.2 $\mu\text{g}/\text{mL}$ 的对照品溶液,并超声处理 5 min,过 0.22 μm 有机滤膜进液相色谱仪样瓶中,待测。

2.5.3 供试品溶液的制备 取苦杏仁粉未经 2 号筛过筛约 0.25 g,精密称定,置具塞锥形瓶中,精密加入甲醇 25 mL,密塞,称定质量,超声处理(功率 250 W、频率 50 kHz) 30 min,放冷,再称定质量,用甲醇补足减失的质量,摇匀,滤过,精密量取续滤液 5 mL,置 25 mL 量瓶中,加 50% 甲醇稀释至刻度,摇匀,滤过,取续滤液过 0.22 μm 有机滤膜进液相色谱仪样瓶中,待测。

2.5.4 线性关系考察 取苦杏仁苷对照品适量,精密称定,制得苦杏仁苷质量浓度分别为 1.02、5.08、20.32、30.48、50.80、71.12、101.60、121.92、132.08 $\mu\text{g}/\text{mL}$ 系列对照品溶液,用“2.5.1”项方法测得峰面积。以峰面积为纵坐标(Y),对照品质量浓度为横坐标(X),绘制标准曲线,得苦杏仁苷回归方程 $Y=10\ 549 X+6\ 522.7$, $R^2=0.999\ 9$,苦杏仁苷在 1.02~132.08 $\mu\text{g}/\text{mL}$ 呈良好线性关系。

2.5.5 精密度实验 取“2.5.2”项下方法制备的苦杏仁苷对照品溶液(43.28 $\mu\text{g}/\text{mL}$),按上述色谱条件进样 10 μL 测定,分别重复进样 6 次,测得相对峰面积的 RSD 为 0.354%,表明实验仪器精密性良好。

2.5.6 稳定性实验 取“2.5.3”项下方法制备的同一供试品溶液(S15)在 0、1、2、4、8、12、18、24、36、48 h 分别进样 10 μL 测定,测得苦杏仁苷的相对峰面积 RSD 为 0.66%,表明供试品溶液在 48 h 内稳定性良好。

2.5.7 重复性实验 取同一批苦杏仁样品(S15),按照“2.5.3”项下方法操作,平行制备 6 份供试品溶液,按照上述色谱条件进样 10 μL 测定,测得苦杏仁苷质量分数的 RSD 为 0.92%,表明该色谱条件下重复性良好。

2.5.8 加样回收实验 取已测定苦杏仁苷含量的苦杏仁样品(S15) 6 份,每份约 0.12 g,精密称定。分别置于锥形瓶中,各加入苦杏仁苷对照品溶液 2 mL(苦杏仁苷质量浓度为 827.9 $\mu\text{g}/\text{mL}$),重复以上步骤,计算加样回收率为 100.21%,RSD 为 1.07%,符合要求。

2.5.9 苦杏仁苷含量测定结果 按“2.5”项下方法,

苦杏仁苷测量结果见表 2,结果表明随着苦杏仁走油程度加深,苦杏仁苷含量降低。

表 2 不同走油程度样品中苦杏仁苷含量

Table 2 Range of amygdalin content in samples with different rancid degree

走油程度	苦杏仁苷/($\text{mg}\cdot\text{g}^{-1}$)		
	生品	煨品	炒品
I	3.73~5.07	1.22~3.56	2.27~3.41
II	1.53~5.09	2.44~3.22	1.89~3.23
III	0.13~0.58	0.95~1.74	0.57~1.25
IV	0.00~0.02	0.06~0.73	0.06~0.32

2.6 酸值与过氧化值的测定

酸值测定方法参照《中国药典》2020 年版通则 0713 的方法^[9];苦杏仁中过氧化值的测定方法参照《中国药典》2020 年版通则 2303 的方法^[9]。

结果表明,酸值和过氧化值测量结果显示,随着苦杏仁样品走油程度的加深,苦杏仁过氧化值和酸值整体趋势为逐渐升高。

2.7 数据处理

2.7.1 传感器阵列的优化 以 α -FOX3000 的 12 根原始传感器阵列为研究对象,基于 Kruskal-Wallis H 检验、累积相关系数分析、聚类分析结合典则指标筛选法、逐步判别分析法,摸索并确立了传感器阵列的优化方法^[10]。

2.7.2 苦杏仁颜色与气味的信息特征级融合

(1) Z-score 标准化:将原始数据标准化为 Z-score 标准化,即标准差标准化。为了不影响模型的精度,需要将采集到的数据进行 Z-score 标准化,以消除因为数据数量级相差大造成的变量对模型作用差异大^[11]。

(2) 颜色与气味信息融合:采用 SPSS 23.0 软件,将粉末颜色特征值和剖面颜色特征值分别与电子鼻传感器最大响应值进行标准化处理,然后用因子分析方法对复杂数据矩阵降维,得到粉末颜色与气味的融合信息^[12]。

2.7.3 基于融合信息对苦杏仁走油程度的判别 采用不同的机器学习算法根据样品粉末颜色与气味融合信息区分苦杏仁的走油程度,并采用十折交叉和外部测试集 2 种验证方法进行验证^[13],并与单一的粉末颜色和气味信息所建立的走油程度判别模型相比较,验证信息融合所建模型的科学性。

2.7.4 颜色气味融合信息的苦杏仁内在质量预测

利用 SPSS 软件分别对苦杏仁的化学成分、酸值、过氧化值与粉末/剖面颜色及气味融合信息进行相关性分析。并且以粉末/剖面颜色及气味融合信息为自变量，分别以苦杏仁苷含量、酸值和过氧化值为因变量，随机选取 80%的样本作为训练样本用逐步回归法建立回归方程，同时用另外 20%的样本检验方程的拟合程度。

2.8 传感器优化后阵列结果

按“2.7.1”项下数据处理，在检测不同走油程度苦杏仁时根据 Kruskal-Wallis *H* 检验、累积相关系数分析对 α -FOX3000 气味指纹分析仪进行了传感器的优化，优化后气味处理去除冗余信息只保留 7 根电子鼻，结果见表 3。

表 3 优化传感器阵列后 7 根金属氧化传感器及检测范围
Table 3 Optimized sensor array after seven metal oxide sensors and detection range

序号	传感器型号	检测范围
X1	LY2/LG	氧化气体
X2	LY2/G	氨气/有化胺类、一氧化碳
X3	LY2/AA	乙醇
X4	LY2/gCTL	硫化氢
X5	P10/1	烃类、甲烷
X6	P10/2	甲烷
X7	PA/2	乙醇、氨气/有机胺类

2.9 颜色与气味信息融合结果

2.9.1 粉末颜色与气味信息融合 按“2.7.2”项下数据处理，粉末颜色与气味信息经过标准化处理，主成分分析 (principal component analysis, PCA) 降维后，按照特征值大于 1，方差百分比大于 5，累积贡献率大于 85%原则^[14]，从表 4 可以看出前 2 个因子的特征值都大于 1，前 2 个因子总贡献率为 91.385%，故提取前 2 个因子 (F_{1-1} 、 F_{1-2}) 进行分析。

2.9.2 剖面颜色与气味信息融合 按“2.7.2”项下数据处理，剖面颜色与气味信息经过标准化处理，

表 4 粉末颜色与气味信息融合解释的总方差

Table 4 Total variance explained by the fusion of powder color and odor information

成分	初始特征值			提取平方和载入		
	合计	方差比/%	累积贡献率/%	合计	方差比/%	累积贡献率/%
1	11.187	74.581	74.581	11.187	74.581	74.581
2	2.521	16.804	91.385	2.521	16.804	91.385

PCA 降维后，选取原则同上，从表 5 可以看出 3 个因子总贡献率为 94.055%，故提取前 3 因子 (F_{2-1} 、 F_{2-2} 、 F_{2-3}) 进行分析。本实验进一步以主成分的分差贡献率为权重对各个主成分的线性组合的系数进行了加权平均的归一化，并采用 PCA 法进行权重计算^[14]。

表 5 剖面颜色与气味信息融合解释的总方差

Table 5 Total variance explained by fusion of profile color and odor information

成分	初始特征值			提取平方和载入		
	合计	方差比/%	累积贡献率/%	合计	方差比/%	累积贡献率/%
1	11.196	74.641	74.641	11.196	74.641	74.641
2	1.830	12.199	86.840	1.830	12.199	86.840
3	1.082	7.215	94.055	1.082	7.215	94.055

2.10 基于融合信息对苦杏仁走油程度的判别结果

2.10.1 基于粉末颜色气味融合信息对不同走油程度苦杏仁的判别 按“2.8.3”项下数据处理，不同的机器学习算法对苦杏仁走油程度进行判别，结果见表 6。结果显示，基于样品粉末颜色和气味融合信息的模型建立中，Logistic、IBK、K-Star、LMT 和 Random Forest 5 种算法所得正判率较高，为较优算法，可完成对不同走油程度苦杏仁的分类鉴别。

2.10.2 基于剖面颜色气味融合信息对不同走油程度苦杏仁的判别 按“2.8.3”项下数据处理，不同的机器学习算法对苦杏仁走油程度进行判别，结果见表 7。结果显示，基于样品剖面颜色和气味融合信息的模型建立中，Logistic 和 K-Star 这 2 种算法所得正判率较高，为较优算法，可完成对不同走油程度苦杏仁的分类鉴别。

2.11 颜色与气味融合信息的苦杏仁内在质量预测

2.11.1 苦杏仁走油的理化指标与信息融合相关性分析

(1) 苦杏仁走油理化指标与粉末颜色及气味信息融合相关性分析：按“2.7.4”项下数据处理，采用 SPSS 23.0 分析软件对粉末颜色及气味信息融合值与苦杏仁走油理化指标 (苦杏仁苷含量、酸值、过氧化值) 进行相关性分析^[15]，结果见表 8。

(2) 苦杏仁走油理化指标与剖面颜色及气味信息融合相关性分析 按“2.7.4”项下数据处理，同法采用 SPSS 23.0 分析软件对剖面颜色及气味信息融合值与苦杏仁走油理化指标进行相关性分析，结

表 6 基于粉末颜色、电子鼻传感器优化阵列、两者信息融合的不同走油程度苦杏仁机器学习结果

Table 6 Based on powder color, electronic nose sensor optimization array, both information fusion of different rancid degree *Armeniaca Semen Amarum* machine learning results

机器学习法	正判率/%					
	十折交叉验证法			外部测试集验证法		
	粉末颜色	气味信息	信息融合	粉末颜色	气味信息	信息融合
朴素贝叶斯 (Naive Bayes)	89.86	56.52	84.06	85.71	71.43	92.86
支持向量机 (SVM)	76.81	53.62	91.30	78.57	57.14	85.71
逻辑算法 (Logistic)	84.06	69.57	95.65	64.29	57.14	100.00
多层感知 (Multiple Layer Perception)	84.96	71.01	95.65	57.14	50.00	92.86
径向基函数网络 (RBF Network)	85.51	62.32	85.51	71.43	57.14	85.71
IBK*	85.51	73.91	95.65	57.14	64.29	100.00
K-Star*	85.51	71.01	94.02	64.29	57.14	100.00
过滤分类器 (Filtered Classifier)	85.51	56.52	81.16	57.14	42.86	85.71
随机委员会 (Random Committee*)	84.06	73.91	92.75	71.43	50.00	92.86
C4.5 决策树 (J48)	79.71	73.91	86.96	64.29	42.86	85.71
逻辑模型树 (LMT*)	85.51	63.77	94.02	64.29	50.00	100.00
随机森林 (Random Forest*)	86.96	66.67	94.02	71.43	42.86	100.00

*Logistic、IBK、K-Star、LMT 和 Random Forest 等为较优算法

*Logistic, IBK, K-Star, LMT and Random Forest are the superior algorithms

表 7 基于剖面颜色、电子鼻传感器优化阵列、两者信息融合的不同走油程度苦杏仁机器学习结果

Table 7 Based on profile color, electronic nose sensor optimization array, both information fusion different rancid degree *Armeniaca Semen Amarum* machine learning results

机器学习法	正判率/%					
	十折交叉验证法			外部测试集验证法		
	剖面颜色	气味信息	信息融合	剖面颜色	气味信息	信息融合
朴素贝叶斯 (Naive Bayes)	85.51	56.52	79.71	64.29	71.43	85.71
支持向量机 (SVM)	76.81	53.62	84.06	64.29	57.14	78.57
逻辑算法 (Logistic)	82.61	69.57	89.86	78.57	57.14	85.71
多层感知 (Multiple Layer Perception)	85.51	71.01	88.41	78.57	50.00	71.43
径向基函数网络 (RBF Network)	78.26	62.32	85.51	78.57	57.14	78.57
IBK*	84.06	73.91	85.51	64.29	64.29	85.71
K-Star*	81.16	71.01	82.61	71.43	57.14	100.00
过滤分类器 (Filtered Classifier)	82.61	56.52	78.26	71.43	42.86	92.86
随机委员会 (Random Committee*)	82.61	73.91	78.26	71.43	50.00	85.71
C4.5 决策树 (J48)	81.16	73.91	68.12	71.43	42.86	85.71
逻辑模型树 (LMT*)	88.41	63.77	88.41	64.29	50.00	78.57
随机森林 (Random Forest*)	86.96	66.67	78.26	71.43	42.86	71.43

*Logistic 和 K-Star 为较优算法

*Logistic and K-Star are the superior algorithms

果见表 9。

2.11.2 苦杏仁走油的理化指标与信息融合回归方程的建立与检验

(1) 理化指标与粉末颜色及气味信息融合回归方程的建立与检验：按“2.7.4”项下数据处理，自变量为粉末颜色及气味信息融合综合性指标，因变

表 8 苦杏仁走油理化指标与粉末颜色及气味信息融合的 Spearman 相关系数 (n = 66)

Table 8 Spearman's correlation coefficients for fusion of *Armeniaca Semen Amarum rancid physicochemical indexes with powder color and odor information (n = 66)*

理化指标	F_{1-1}	$F_{1-1} P$ 值	F_{1-2}	$F_{1-2} P$ 值
苦杏仁苷含量	0.239	0.054	-0.797**	0.000
酸值	-0.252*	0.041	0.788**	0.000
过氧化值	-0.162	0.194	0.630**	0.000

表 9 苦杏仁走油理化指标与剖面颜色及气味信息融合的 Spearman 相关系数 (n = 66)

Table 9 Spearman's correlation coefficients calculated for fusion of *Armeniaca Semen Amarum rancid physicochemical indicators with profile color and odor information (n = 66)*

理化指标	F_{2-1}	$F_{2-1} P$ 值	F_{2-2}	$F_{2-2} P$ 值	F_{2-3}	$F_{2-3} P$ 值
苦杏仁苷含量	0.207	0.095	0.610**	0.000	-0.452**	0.000
酸值	-0.225	0.070	-0.682**	0.000	0.236	0.057
过氧化值	-0.150	0.229	-0.588**	0.000	0.154	0.217

*为相关检验显著性在 0.05 水平 (双尾检验), **为相关检验显著性在 0.01 水平 (双尾检验)

* Correlation test significance at the 0.05 level (two-tailed test), ** correlation test significance at the 0.01 level (two-tailed test)

表 10 不同化学成分的回归模型

Table 10 Regression models for different chemical components

理化指标	预测方程	P 值	显著性	R^2	R^2 校正
苦杏仁苷含量	$Y=2.175-1.340 F_{1-2}+0.529 F_{1-1}$	0.000	显著	0.732	0.721
酸值	$Y=2.113+1.724 9 F_{1-2}-0.667 F_{1-1}$	0.000	显著	0.719	0.707
过氧化值	$Y=0.013+0.004 F_{1-2}$	0.000	显著	0.424	0.412

杏仁苷含量的预测方程和酸值的预测方程进行拟合误差检验,得到回归方程预测结果和真实值散点图。结果见图 2、3。

(2) 苦杏仁走油理化指标与剖面颜色及气味信息融合回归方程的建立与检验:方法同“2.11.2(1)”项,建立回归模型,结果见表 11。测试样本对苦杏仁苷含量的预测方程和酸值的预测方程进行拟合误差检验,得到回归方程预测结果和真实值散点图,

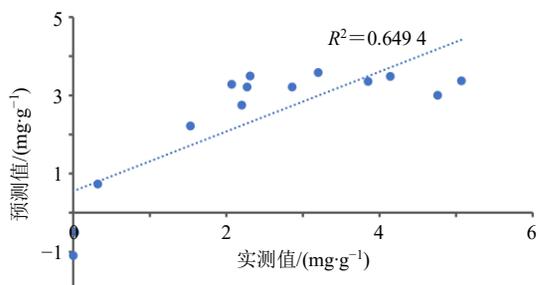


图 2 苦杏仁苷实测值与预测值对比

Fig. 2 Comparison of measured and predicted values of amygdalin

量分别为走油苦杏仁的理化指标 (苦杏仁苷含量、酸值、过氧化值)。

从样品中随机选取 53 个数据作为训练集用逐步回归法建立回归模型,余下的 13 个数据作为验证集验证回归模型^[16]。通过方程计算不同走油程度苦杏仁中的苦杏仁苷含量、酸值、过氧化值等指标的预测值与测量值之间的相关系数,综合检验方程的显著性系数 (P)、拟合优度 (R^2)、修正自由度后的拟合优度 (R^2 校正),结果见表 10。测试样本对苦

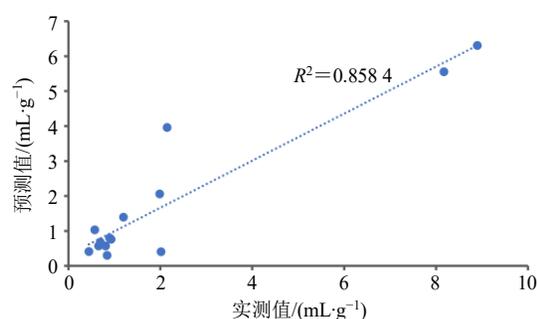


图 3 酸值实测值与预测值对比

Fig. 3 Comparison of measured and predicted acid values

结果见图 4、5。

3 分析与讨论

3.1 不同走油程度样品图的划分结果分析

经过研究证实,高温高湿是走油的主要影响因素,在相同的存储条件下存放时间越长,其走油程度越高。

3.2 不同走油程度苦杏仁机器学习结果判别分析

由表 6、7 可知,粉末颜色与气味融合信息来判

表 11 不同化学成分回归模型

Table 11 Regression models for different chemical components

理化指标	方程	P 值	显著性	R ²	R ² 校正
苦杏仁苷含量	$Y=2.153+1.242 F_{2-2}+0.500 F_{2-1}-0.689 F_{2-3}$	0.000	显著	0.775	0.765
酸值	$Y=2.226-1.946 F_{2-2}-0.785 F_{2-1}$	0.000	显著	0.738	0.728
过氧化值	$Y=0.013-0.004 F_{2-2}$	0.000	显著	0.397	0.385

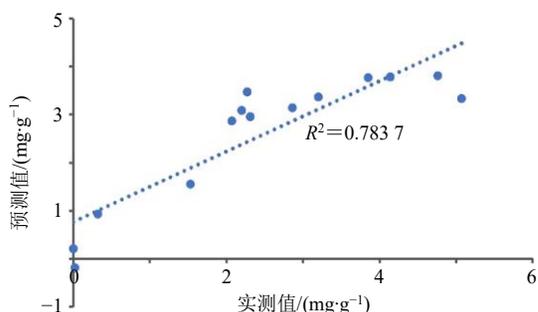


图 4 苦杏仁苷含量实测与预测值对比

Fig. 4 Comparison of measured and predicted values of amygdalin content

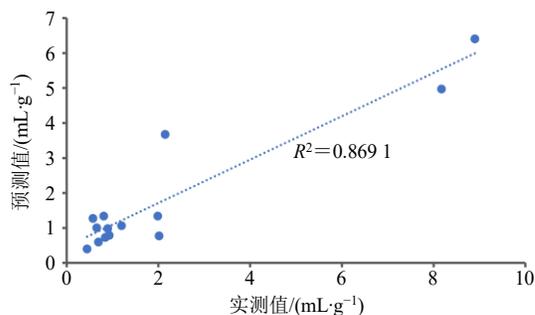


图 5 酸值含量实测与预测值对比

Fig. 5 Comparison of measured and predicted acid content

别苦杏仁走油程度的正判率明显高于单一的粉末颜色或气味信息的判别正确率。其中 Logistic、IBK、K-Star、LMT 和 Random Forest 算法正判率均高于 94%；而剖面颜色及气味融合信息来判别苦杏仁走油程度的正判率虽然低于粉末颜色及气味信息融合的正判率，但比单一的剖面颜色信息和气味信息的正判率高。其中，Logistics 算法和 K-Star 算法的正判率高于 82%，都可以很好的实现对各种走油苦杏仁的分类识别。

3.3 苦杏仁走油理化指标与颜色及气味信息融合值相关性分析

由表 8、9 可知，经显著性检验粉末颜色及气味信息融合值 F_{1-2} 与苦杏仁化学成分，走油指标具有相关性，其中，苦杏仁苷含量与 F_{1-2} 显著负相关；而酸值与 F_{1-1} 成负相关， F_{1-2} 显著正相关；过氧化值

与 F_{1-2} 显著正相关。而剖面颜色及气味信息融合值与苦杏仁走油理化指标具有相关性，其中，苦杏仁苷含量与 F_{2-2} 显著正相关，与 F_{2-3} 显著负相关；而酸值与 F_{2-2} 显著负相关，过氧化值与 F_{2-2} 显著负相关。相关性分析与后来的预测方程结果是保持一致。

3.4 回归方程的建立与检验结果分析

由表 10、11 可知，不同化学成分的回归模型中， $R^2=0.50$ 是可以接受结果的底线，此时随机因素的影响已经有 50%，但自变量和因变量的依存关系仍占主导地位^[17]。上述建好的 6 个方程中，过氧化值与粉末颜色及气味信息融合的预测方程 $R^2_{校正}=0.412<0.50$ ；与剖面颜色及气味信息融合的预测方程 $R^2_{校正}=0.3761<0.50$ ，方程不可取。而苦杏仁苷和酸值的回归模型中，方程通过。同时由图 2、3，图 4、5 可知，通过测试样本对苦杏仁苷含量的预测方程和酸值的预测方程进行拟合误差检验，其拟合度均高于 50%，方程可行。

3.5 讨论

本实验采用电子鼻和色度仪技术测量不同走油程度苦杏仁的气味和颜色信息，通过特征级融合方式来融合苦杏仁颜色与气味的信息并与苦杏仁走油过程中的内在化学成分的变化规律建立客观的联系，建立了苦杏仁不同走油程度的判别分析和内在质量预测模型，为苦杏仁走油的在线准确监测提供新思路。

本实验采用了 12 种不同的机器学习方法，对苦杏仁走油程度的判别进行了研究。在基于颜色、气味和融合信息的识别模型中，Naive Bayes、Logistic、Multiple Layer Perception 和 IBK 等方法具有较高的正确性。其他类似的研究也可以将这些算法作为首选。其中也有一些问题值得探讨，目前还没有得到过氧化值变化的预测模型。分析其原因，一方面是因为大多数样品的过氧化值偏低，而采用滴定方法测定的结果存在一定的误差。另外，过氧化物很容易分解为小分子挥发性物质，这也是原因之一。研究发现，反向传播 (back propagation, BP) 神经网络

络名为神经网络,构建药向量训练模型,在中药的药性与功效具有明显关联性的基础上,对中药药性的初始量化值进行修正,使中药药性的量化值更精确^[18]。鉴于此,课题组其它研究曾经尝试用BP神经网络建立的过氧化值预测模型,结果表现为粉末颜色和气味融合信息的过氧化值拟合度尚佳,可为过氧化值快速检测提供参考。

本实验仅从感官与化学成分角度对苦杏仁的走油进行了研究,建立监测模型,后续试验可增加药理相关研究,从感官、理化、药效3个角度全面地构建出一套完整的中药品质评估体系。由于本项目的建模和验证所使用的样品数量较少,因此在以后的试验中需要更多的样品来进一步完善。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 康廷国. 中药鉴定学 [M]. 第4版. 北京: 中国中医药出版社, 2016: 29.
- [2] 程云霞, 刘伟, 王凯, 等. 中药材贮藏过程中“走油”变质的影响因素品质评价及其机制研究进展 [J]. 中国现代中药, 2022, 24(6): 1134-1138.
- [3] 姚月保, 杨晓芸, 彭莲, 等. 基于仿生识别系统对金银花颜色与“气”的鉴别研究 [J]. 世界科学技术—中医药现代化, 2020, 22(4): 1270-1277.
- [4] Nešpor J, Karabín M, Štulíková K, *et al.* An HS-SPME-GC-MS method for profiling volatile compounds as related to technology used in cider production [J]. *Molecules*, 2019, 24(11): 2117.
- [5] 程云霞, 于金倩, 王凯, 等. 苦杏仁贮藏过程中走油变质机制分析 [J]. 中国实验方剂学杂志, 2021, 27(16): 124-130.
- [6] 郭星好, 曹光昭, 拱健婷, 等. 基于数学模型的苦杏仁走油变质过程拟合研究 [J]. 现代中药研究与实践, 2022, 36(4): 36-41.
- [7] 曹光昭, 陈慧荣, 刘玢清, 等. 不同走油程度苦杏仁“电子鼻”气味与止咳药效的相关性分析 [J]. 中成药, 2020, 42(5): 1249-1254.
- [8] 拱健婷, 赵丽莹, 徐东, 等. 电子鼻联合支持向量机建立苦杏仁不同炮制品及走油程度快速判别模型 [J]. 中国中药杂志, 2020, 45(10): 2389-2394.
- [9] 中国药典 [S]. 一部. 2020: 210.
- [10] 陈慧荣. 基于颜色气味数字化及信息融合的苦杏仁走油监测系统研究 [D]. 北京: 北京中医药大学, 2019.
- [11] 拱健婷, 李莉, 邹慧琴, 等. 基于灰色关联分析和功效组分的大黄药材质量评价 [J]. 世界科学技术—中医药现代化, 2018, 20(10): 1808-1815.
- [12] 李伟, 蒋沙沙, 欧美凤, 等. 基于数据挖掘探讨含钙中药药性规律 [J]. 中成药, 2021, 43(7): 1873-1877.
- [13] 李欢, 熊梦莹, 聂斌, 等. 融合因子分析的随机森林研究 [J]. 计算机工程与应用, 2019, 55(23): 125-130.
- [14] 李晓钟, 王欢. 互联网对我国经济发展影响的区域差异比较研究 [J]. 中国软科学, 2020(12): 22-32.
- [15] 洪伟峰, 赵丽莹, 拱健婷, 等. 基于电子鼻建立不同炮制规格苦杏仁“走油”预警模型 [J]. 中药材, 2021, 44(8): 1844-1847.
- [16] Liu C, Liu L C, Gao J L, *et al.* Identification of two long non-coding RNAs AC010082.1 and AC011443.1 as biomarkers of coronary heart disease based on logistic stepwise regression prediction model [J]. *Front Genet*, 2021, 12: 780431.
- [17] 冯艳春, 张琪, 胡昌勤. 药品近红外光谱通用性定量模型评价参数的选择 [J]. 光谱学与光谱分析, 2016, 36(8): 2447-2454.
- [18] 邓乐. 基于BP神经网络的中药药性量化研究 [D]. 长沙: 湖南中医药大学, 2020.

[责任编辑 郑礼胜]