

基于挥发性成分定性判别风味油茶籽油掺伪浸出油茶籽油 PCA 模型和逻辑回归模型的对比分析

孙婷婷^{1,2}, 陈志清^{1,2}, 刘剑波³, 任佳丽^{1,2}, 钟海雁^{1,2}, 周波^{1,2}

(1. 林产可食资源安全与加工利用湖南省重点实验室, 长沙 410004; 2. 中南林业科技大学 食品科学与工程学院, 长沙 410004; 3. 岳阳市检验检测中心食品药品检验所, 湖南 岳阳 414000)

摘要:为了解决风味(原香和烤香)油茶籽油掺伪浸出油茶籽油的定性判别问题,设计高、低两个掺伪梯度,基于挥发性成分构建并对比分析了定性判别风味油茶籽油掺伪浸出油茶籽油的主成分分析(PCA)模型和逻辑回归模型。结果表明:逻辑回归模型定性判别风味油茶籽油掺伪浸出油茶籽油的能力较强,优于PCA模型;高掺伪梯度下定性判别原香和烤香油茶籽油掺伪浸出油茶籽油,PCA模型的最低检出限分别为20%和60%,而逻辑回归模型的最低检出限均为10%;低掺伪梯度下定性判别原香和烤香油茶籽油掺伪浸出油茶籽油,PCA模型的判别不准确,而逻辑回归模型的最低检出限均为4%。逻辑回归模型能很好地定性判别风味油茶籽油掺伪浸出油茶籽油。

关键词:风味油茶籽油;浸出油茶籽油;挥发性成分;主成分分析模型;逻辑回归模型;定性判别
中图分类号:TS227;TQ646.4 文献标识码:A 文章编号:1003-7969(2023)03-0056-08

Comparative analysis between PCA model and logistic regression model for qualitative identification of flavor oil – tea camellia seed oil adulteration with leaching oil – tea camellia seed oil based on volatile components

SUN Tingting^{1,2}, CHEN Zhiqing^{1,2}, LIU Jianbo³, REN Jiali^{1,2},
ZHONG Haiyan^{1,2}, ZHOU Bo^{1,2}

(1. Hunan Key Laboratory of Forestry Edible Sources Safety and Processing, Changsha 410004, China; 2. School of Food Science and Engineering, Central South University of Forestry and Technology, Changsha 410004, China; 3. Food and Drug Inspection Institute of Yueyang City Inspection and Testing Center, Yueyang 414000, Hunan, China)

Abstract: Based on volatile components, principal component analysis (PCA) and logistic regression (LR) models were constructed and compared to solve the problem of qualitative identification of flavor (original/roasted) oil – tea camellia seed oil adulterated with leaching oil – tea camellia seed oil under high and low adulteration gradients. The results showed that LR model had a good ability to qualitatively identify the flavor oil – tea camellia seed oil adulterated with leaching oil – tea camellia seed oil, which was better than that of PCA model. Under the high adulteration gradient, the detection limit of the PCA model for original/roasted oil – tea camellia seed oil adulterated with leaching oil – tea camellia seed oil

was 20%/60%, and the detection limit of the LR model was 10%/10%, respectively. Under the low adulteration gradient, the detection limit of LR model for original/roasted oil – tea camellia seed oil adulterated with leaching oil – tea camellia seed oil was 4%/4%, but the discriminative ability of the PCA model was not accurate. The LR model can qualitatively identify

收稿日期:2021-12-30;修回日期:2022-10-26

基金项目:湖南省林业科技创新基金项目(XLK202101-02);湖南省市场监督管理局科技计划项目(2020KJJH55);中央引导地方科技发展专项资金区域创新体系建设专项(2020ZYQ036)

作者简介:孙婷婷(1994),女,硕士,研究方向为粮食、油脂与植物蛋白工程(E-mail)suntting0616@163.com。

通信作者:周波,教授(E-mail)zhbo78@126.com。

flavor oil – tea camellia seed oil adulterated with leaching oil – tea camellia seed oil.

Key words: flavor oil – tea camellia seed oil; leaching oil – tea camellia seed oil; volatile component; principal component analysis model; logistic regression model; qualitative identification

油茶 (*Camellia oleifera* Abel.) 属山茶科 (Theaceae) 山茶属 (*Camellia*) 的常绿小乔木或灌木, 与油棕、橄榄油和椰子齐名为世界四大木本食用油料树种^[1]。油茶籽油来源于油茶种子, 又名山茶油、山茶籽油或茶油, 是一种营养丰富的高品质食用油^[2-4]。油茶籽油营养价值和经济价值高, 有以次充好、掺杂使假、以假顶真的市场现象^[5-6], 因此建立科学、快速、有效的油茶籽油掺伪检测方法是维护油茶籽油正常市场秩序, 保护合法生产经营者和消费者权益的必然要求。

目前, 关于油茶籽油掺伪鉴别的方法主要是采用相关实验仪器和检测技术, 基于获得的大量复杂的结构化测量数据, 运用相关的数学统计分析方法, 挖掘和提取出掺伪鉴别所需的特征信息^[7]。油茶籽油结构化测量数据主要包括脂肪酸、挥发性成分、活性伴随物以及其他理化指标。油茶籽油掺伪鉴别研究大多集中在检测手段和分析方法上, 如常规理化检测法^[8-9]、色谱法^[10-11]、核磁共振法^[12]、近红外光谱法^[13]、拉曼光谱法^[14]、荧光光谱法^[15]、电子鼻技术^[16]及时间分辨荧光技术^[17]等, 这些技术和方法都能很好地实现油茶籽油掺伪的定性判别和较好的定量分析。采用红外光谱技术结合不同统计分析方法, 如红外光谱透射模式与透反射模式结合判别分析 (DA)^[18]和主成分分析 (PCA)^[19], 傅里叶变换红外光谱结合偏最小二乘法 (PLS)、支持向量机 (SVM) 和 BP 人工神经网络 (BPANN)^[20]等, 可实现压榨油茶籽油与浸出油茶籽油的定性判别。但基于我国风味油茶籽油产品分类分级体系缺乏、感官品质评价技术不完善、油茶籽油风味品质调控不精准和风味油茶籽油相关的结构化测量数据难以确定, 风味油茶籽油掺伪浸出油茶籽油的相关研究至今鲜有报道。

风味油茶籽油和浸出油茶籽油的主要区别在于其挥发性成分的组成及含量不同, 所以定性和定量判别风味油茶籽油掺伪浸出油茶籽油的关键在于采用何种检测技术和数据统计分析方法来挖掘和提取掺伪鉴别所需的信息。本研究基于通过 HS-SPME-GC-MS 获得的油茶籽油挥发性成分组成及含量, 运用 Python 语言构建并对比了定性判别风味 (原香和烤香) 油茶籽油掺伪浸出油茶籽油的 PCA 模型和逻辑回归模型, 以期能为风味油茶籽油掺伪浸出油茶

籽油的鉴别研究提供参考。

1 数据采集

原香和烤香油茶籽油掺伪浸出油茶籽油设置两个掺伪梯度, 自定义为高掺伪梯度 (0、5%、10%、15%、20%、40%、60%、80%) 和低掺伪梯度 (0.2%、4%、6%、8%、10%)。其中原香油茶籽油每个掺伪梯度均为 4 个样品, 烤香油茶籽油每个掺伪梯度均为 10 个样品。原香油茶籽油、烤香油茶籽油及其掺伪浸出油茶籽油挥发性成分组成及含量信息参考朱晓阳^[21]的报道, 原香油茶籽油掺伪浸出油茶籽油样品共检测出 45 种挥发性物质, 烤香油茶籽油掺伪浸出油茶籽油样品共检测出 56 种挥发性物质。

2 模型的建立及验证

2.1 数据处理

所有模型和算法均基于 Python 3.7 编程语言在 PyCharm 2018 IDE 平台 [JetBrains (Prague), Czech Republic] 进行程序编写, 分别采用 PCA 模型和逻辑回归模型定性判别风味油茶籽油掺伪浸出油茶籽油。逻辑回归模型原香和烤香油茶籽油掺伪样本标签都设置为“1”, 纯原香和烤香油茶籽油样本标签都设置为“0”。

2.2 逻辑回归模型评价指标

采用常见的分类模型评价指标如精确率、准确率、召回率、F1-分数等对模型的精度进行评价, 其中: 精确率代表预测为正的样本中有多少是实际的正样本; 准确率代表预测正确的样本占所有样本的比例; 召回率代表样本中的正例有多少被预测正确; F1-分数是兼顾了精确率和召回率的评估指标。按公式(1)~公式(4)计算精确率 (P)、准确率 (A)、召回率 (R)、F1-分数 (F_1)。

$$P = P_T / (P_T + P_F) \quad (1)$$

$$A = (P_T + N_T) / (P_T + N_T + P_F + N_F) \quad (2)$$

$$R = P_T / (P_T + N_F) \quad (3)$$

$$F_1 = 2PR / (P + R) = 2P_T / (2P_T + P_F + N_F) \quad (4)$$

式中: P_T 为被预测正确的正样本的数量; N_T 为被预测正确的负样本的数量; P_F 为被错判为正的负样本数量; N_F 为被错判为负的正样本数量。本文中, 预测正确是指预测标签与实际标签一致, 预测错误是指预测标签与实际标签不一致。

3 模型鉴别分析

3.1 风味油茶籽油掺伪浸出油茶籽油的 PCA 模型

3.1.1 原香油茶籽油高掺伪浸出油茶籽油的 PCA

利用 Sklearn.decomposition 库中的 PCA 函数对原香油茶籽油掺伪浸出油茶籽油挥发性成分含量开展 PCA, 设置最小累积贡献率阈值为 90%, 其余参数为缺省值, 得到特征值和方差贡献率, 见表 1。

由表 1 可知, 基于高掺伪梯度下原香油茶籽油的第 1 主成分的贡献率为 52.2%, 第 2 主成分的贡献率为 30.9%, 第 3 主成分的贡献率为 14.1%, 前 3 个主成分的累积贡献率达到了 97.2%, 满足预设的 90% 的最小累积贡献率阈值, 故 PCA 函数自动选择保留前 3 个主成分, 这 3 个主成分可反映样本的主

要信息。根据主成分载荷矩阵和特征向量矩阵(见表 2)和高掺伪梯度下不同掺伪量浸出油茶籽油的原香油茶籽油共 32 个样品的 45 个挥发性成分指标, 可得第 1 主成分、第 2 主成分、第 3 主成分的得分图, 结果见图 1。

表 1 原香油茶籽油掺伪浸出油茶籽油的 PCA 特征值和方差贡献率

梯度	主成分	特征值	贡献率/%	累积贡献率/%
高掺伪	1	128.354	52.2	52.2
	2	75.966	30.9	83.1
	3	34.744	14.1	97.2
低掺伪	1	200.476	72.7	72.7
	2	63.048	22.9	95.6

表 2 原香油茶籽油掺伪浸出油茶籽油的成分载荷矩阵和特征向量矩阵

挥发性成分	高掺伪梯度						低掺伪梯度			
	主成分载荷矩阵			主成分特征向量矩阵			主成分载荷矩阵		主成分特征向量矩阵	
	1	2	3	1	2	3	1	2	1	2
3-甲基丁醛	-0.600	-0.872	-1.067	-0.053	-0.100	-0.181	-0.637	1.660	-0.045	0.209
(E)-2-戊烯醛	-0.068	0.017	0.083	-0.006	0.002	0.014	-0.085	-0.111	-0.006	-0.014
己醛	-5.891	-3.225	-0.065	-0.520	-0.370	-0.011	-6.726	1.683	-0.475	0.212
庚醛	-1.020	0.723	0.666	-0.090	0.083	0.113	-1.373	-1.064	-0.097	-0.134
苯甲醛	0.759	-1.028	-0.908	0.067	-0.118	-0.154	1.090	1.485	0.077	0.187
辛醛	-3.263	1.456	2.770	-0.288	0.167	0.470	-4.205	-3.803	-0.297	-0.479
苯乙醛	1.960	-0.488	0.760	0.173	-0.056	0.129	2.549	-0.723	0.180	-0.091
(E,E)-2,4-庚二烯醛	-0.838	0.235	1.108	-0.074	0.027	0.188	-1.048	-1.397	-0.074	-0.176
(E)-2-辛烯醛	-1.326	-0.253	0.648	-0.117	-0.029	0.110	-1.586	-0.603	-0.112	-0.076
壬醛	-4.588	0.776	1.043	-0.405	0.089	0.177	-5.805	-1.421	-0.410	-0.179
(E)-2-壬烯醛	-0.612	0.183	0.047	-0.054	0.021	0.008	-0.793	-0.127	-0.056	-0.016
癸醛	-0.374	0.070	0.371	-0.033	0.008	0.063	-0.453	-0.453	-0.032	-0.057
(E,E)-2,4-壬二烯醛	-0.068	0.017	0.094	-0.006	0.002	0.016	-0.085	-0.119	-0.006	-0.015
(E)-2-癸烯醛	-0.555	4.863	-0.908	-0.049	0.558	-0.154	-1.614	-1.104	-0.114	-0.139
(E,E)-2,4-癸二烯醛	-1.110	1.020	0.307	-0.098	0.117	0.052	-1.557	-0.786	-0.110	-0.099
2-十一烯醛	-0.215	3.913	-1.332	-0.019	0.449	-0.226	-1.034	-0.183	-0.073	-0.023
椰子醛	0.748	-0.183	0.289	0.066	-0.021	0.049	0.963	-0.278	0.068	-0.035
2-甲基丁醇	-0.283	-1.159	-1.869	-0.025	-0.133	-0.317	-0.212	2.700	-0.015	0.340
3-甲基丁醇	-0.657	-0.540	-0.430	-0.058	-0.062	-0.073	-0.736	0.778	-0.052	0.098
1-己醇	-0.091	0.026	0.124	-0.008	0.003	0.021	-0.113	-0.159	-0.008	-0.020
1-庚醇	-0.091	-0.235	-1.155	-0.008	-0.027	-0.196	-0.113	1.445	-0.008	0.182
(E)-3-壬烯-1-醇	-0.804	0.227	1.055	-0.071	0.026	0.179	-0.991	-1.326	-0.070	-0.167
1-辛醇	-0.408	-0.122	-1.279	-0.036	-0.014	-0.217	-0.538	1.572	-0.038	0.198
苯乙醇	3.501	-1.177	0.931	0.309	-0.135	0.158	4.602	-0.667	0.325	-0.084
苯甲醇	0.453	-0.113	0.177	0.040	-0.013	0.030	0.595	-0.167	0.042	-0.021
苯乙烯	-1.020	-1.499	-1.863	-0.090	-0.172	-0.316	-1.062	2.898	-0.075	0.365
3-乙基-2-甲基-1,3-己二烯	-0.079	0.026	0.106	-0.007	0.003	0.018	-0.099	-0.127	-0.007	-0.016
惕格酸	-0.159	-0.619	-1.008	-0.014	-0.071	-0.171	-0.113	1.453	-0.008	0.183
(Z)-2-甲基-2-甲基丁酯-2-丁烯酸	-1.224	-0.941	-0.678	-0.108	-0.108	-0.115	-1.373	1.263	-0.097	0.159

续表 2

挥发性成分	高掺伪梯度						低掺伪梯度			
	主成分载荷矩阵			主成分特征向量矩阵			主成分载荷矩阵		主成分特征向量矩阵	
	1	2	3	1	2	3	1	2	1	2
辛酸	4.056	-0.967	0.424	0.358	-0.111	0.072	5.239	-0.167	0.370	-0.021
(E)-2-甲基-3-甲基丁酯-2-丁烯酸	-0.759	-0.880	-0.937	-0.067	-0.101	-0.159	-0.821	1.532	-0.058	0.193
壬酸	2.889	-0.471	0.819	0.255	-0.054	0.139	3.710	-0.834	0.262	-0.105
己酸	2.379	-0.017	0.678	0.210	-0.002	0.115	2.988	-0.850	0.211	-0.107
庚酸	1.020	-0.227	0.383	0.090	-0.026	0.065	1.331	-0.373	0.094	-0.047
乙酸乙酯	-0.453	-0.636	-0.772	-0.04	-0.073	-0.131	-0.481	1.207	-0.034	0.152
3-甲基-2-丁烯酸庚酯	-0.011	-0.061	-0.094	-0.001	-0.007	-0.016	-0.014	0.135	-0.001	0.017
丙位辛内酯	-0.023	-0.078	-0.136	-0.002	-0.009	-0.023	-0.014	0.191	-0.001	0.024
糠醛	-0.544	-0.331	-0.147	-0.048	-0.038	-0.025	-0.623	0.341	-0.044	0.043
(Z)-3,4-二甲基-3-己烯-2-酮	-0.363	-0.113	0.088	-0.032	-0.013	0.015	-0.425	-0.048	-0.030	-0.006
2-甲基萘	-0.045	0.017	0.065	-0.004	0.002	0.011	-0.057	-0.079	-0.004	-0.010
5-丁基-二氢-2(3H)-呋喃酮	0.351	-0.087	0.136	0.031	-0.010	0.023	0.453	-0.127	0.032	-0.016
甲基丁香酚	0.249	-0.061	0.094	0.022	-0.007	0.016	0.326	-0.095	0.023	-0.012
(Z)-2-庚烯醛	1.178	-0.296	0.454	0.104	-0.034	0.077	1.529	-0.437	0.108	-0.055
2,4-二甲基己烷	0.510	3.181	-1.214	0.045	0.365	-0.206	0.000	-0.016	0.000	-0.002
1-十五烯	0.136	0.819	-0.312	0.012	0.094	-0.053	0.000	-0.008	0.000	-0.001

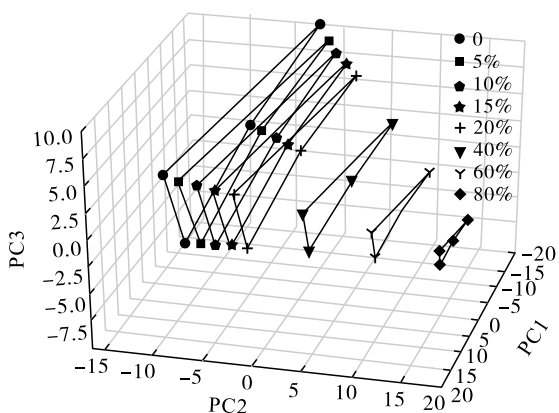


图1 高掺伪梯度下原香油茶籽油掺伪浸出油茶籽油的主成分得分图

由图1可知,当掺伪量在15%及以下时,PC1-PC2-PC3三维得分图中,掺伪浸出油茶籽油的样本与原香油茶籽油样本所处区域均存在部分重叠,利用PCA模型无法对其进行准确鉴别。而掺伪量在20%及以上时,通过PCA模型能较准确地识别掺伪样本,且掺伪量越高,掺伪样本与原香油茶籽油样本所处区域的偏离越明显。

3.1.2 原香油茶籽油低掺伪浸出油茶籽油的PCA

由表1可知,基于低掺伪梯度下原香油茶籽油的第1主成分的贡献率为72.7%,第2主成分的贡献率为22.9%,前2个主成分的累积贡献率达到了95.6%,满足预设的90%的最小累积贡献率阈值,故PCA函数自动选择保留前2个主成分,这2个主成

分可反映样本的主要信息。根据原香油茶籽油主成分载荷矩阵和特征向量矩阵(见表2)和低掺伪梯度下不同掺伪量浸出油茶籽油的原香油茶籽油共24个样品的45个挥发性成分指标,可得第1主成分、第2主成分的得分图,结果见图2。

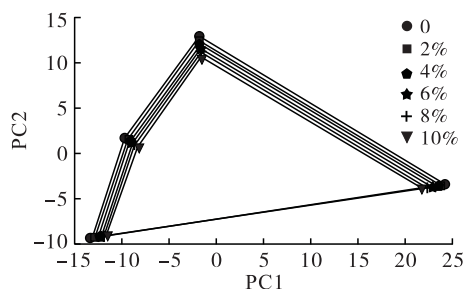


图2 低掺伪梯度下原香油茶籽油掺伪浸出油茶籽油的主成分得分图

由图2可知,在低掺伪梯度下,在PC1-PC2二维得分图中,各掺伪量的掺伪浸出油茶籽油的样本与原香油茶籽油样本所处区域均存在大面积重叠,利用PCA模型无法对其进行准确的鉴别。

3.1.3 烤香油茶籽油高掺伪浸出油茶籽油的PCA

利用Sklearn.decomposition库中的PCA函数对烤香油茶籽油掺伪浸出油茶籽油挥发性成分含量开展PCA,设置最小累积贡献率阈值为90%,其余参数为缺省值,得到特征值和方差贡献率,见表3。

表3 烤香油茶籽油掺伪浸出油茶籽油的 PCA 特征值和方差贡献率

梯度	主成分	特征值	贡献率/%	累积贡献率/%
高掺伪	1	279.029	52.7	52.7
	2	99.836	18.9	71.6
	3	82.542	15.6	87.2
	4	40.745	7.7	94.9
低掺伪	1	362.100	57.4	57.4
	2	134.771	21.3	78.7
	3	83.904	13.3	92.0

由表3可知,基于高掺伪梯度下烤香油茶籽油的第1主成分的贡献率为52.7%,第2主成分的贡献率为18.9%,第3主成分的贡献率为15.6%,第4主成分的贡献率为7.7%,前4个主成分的累积贡献率达到了94.9%,满足预设的90%的最小累积贡献率阈值,故PCA函数自动选择保留前4个主成分,这4个主成分可反映样本的主要信息。根据主成分载荷矩阵和特征向量矩阵(见表4)和高掺伪梯度下不同掺伪量浸出油茶籽油的烤香油茶籽油共80个样品的56个挥发性成分指标,可得第1主成分、第2主成分、第3主成分的得分图,结果见图3。

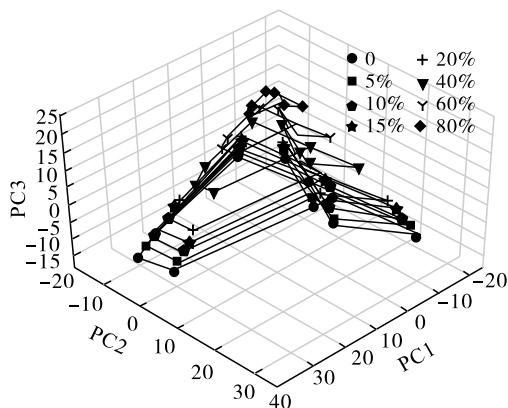


图3 高掺伪梯度下烤香油茶籽油掺伪浸出油茶籽油的 主成分得分图

由图3可知,当掺伪量在40%及以下时,在PC1-PC2-PC3三维得分图中,掺伪浸出油茶籽油的样本与烤香油茶籽油样本所处区域均存在部分重叠,利用PCA模型无法对其进行准确鉴别。掺伪量在60%及以上时,PCA模型可准确识别掺伪样本,且掺伪量越高,掺伪样本与烤香油茶籽油样本所处区域的偏离越明显。

3.1.4 烤香油茶籽油低掺伪浸出油茶籽油的 PCA

由表3可知,基于低掺伪梯度下烤香油茶籽油的第1主成分的贡献率为57.4%,第2主成分的贡献率为21.3%,第3主成分的贡献率为13.3%,前3个主成分的累积贡献率达到了92.0%,满足预设的90%的最小累积贡献率阈值,故PCA函数自动选择

保留前3个主成分,这3个主成分可反映样本的主要信息。根据烤香油茶籽油主成分载荷矩阵和特征向量矩阵(见表4)和低掺伪梯度下不同掺伪量浸出油茶籽油的烤香油茶籽油共60个样品的56个挥发性成分指标,可得第1主成分、第2主成分、第3主成分的得分图,结果见图4。

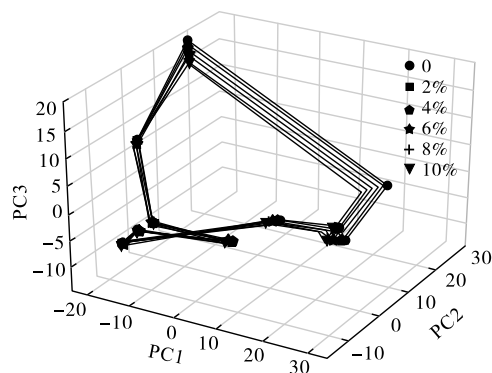


图4 低掺伪梯度下烤香油茶籽油掺伪浸出油茶籽油的 主成分得分图

由图4可知,在低掺伪梯度下,在PC1-PC2-PC3三维得分图中,各掺伪量的掺伪浸出油茶籽油的样本与烤香油茶籽油样本所处区域均存在大面积重叠,利用PCA模型无法对其进行准确的鉴别。

3.2 风味油茶籽油掺伪浸出油茶籽油的逻辑回归模型

基于样品的挥发性成分含量,利用Sklearn.linear_model库中的Logistic regression函数对数据构建逻辑回归模型。利用留一法训练和评估逻辑回归模型,计算得到高掺伪梯度下逻辑回归模型判别结果:对纯原香油茶籽油和浸出油茶籽油掺伪量大于或等于10%的原香油茶籽油的鉴别结果全部正确,但对掺伪量5%的4个样本鉴别错误,得出高掺伪梯度下逻辑回归模型鉴别原香油茶籽油掺伪浸出油茶籽油的检出限为10%;对纯烤香油茶籽油和浸出油茶籽油掺伪量大于或等于10%的烤香油茶籽油的鉴别结果全部正确,但掺伪量5%的10个样本中有5个样本鉴别错误,得出高掺伪梯度下逻辑回归模型鉴别烤香油茶籽油掺伪浸出油茶籽油的检出限为10%。低掺伪梯度下逻辑回归模型判别结果:对纯原香油茶籽油和浸出油茶籽油掺伪量大于或等于4%的原香油茶籽油的鉴别结果全部正确,但对掺伪量2%的4个样本鉴别错误,得出低掺伪梯度下逻辑回归模型鉴别原香油茶籽油掺伪浸出油茶籽油的检出限为4%;对纯烤香油茶籽油和浸出油茶籽油掺伪量大于或等于4%的烤香油茶籽油的鉴别结果全部正确,但对掺伪量2%的10个样本中有5个样本鉴别错误,得出低掺伪梯度下逻辑回归模型鉴别烤香油茶籽油掺伪浸出油茶籽油的检出限为4%。

表4 烤香油茶籽油掺伪浸出油茶籽油主成分载荷矩阵和特征向量矩阵

挥发性成分	高掺伪梯度												低掺伪梯度															
	主成分载荷矩阵						主成分特征向量矩阵						主成分载荷矩阵			主成分特征向量矩阵												
	1	2	3	4	1	2	1	2	3	4	1	2	1	2	3	1	2	3										
庚醛	-0.434	-0.500	0.200	-0.223	-0.026	-0.050	0.022	-0.035	-0.076	-0.058	-0.073	-0.004	-0.005	-0.008	-0.200	0.190	0.454	0.970	-0.012	0.019	0.050	0.152	-0.457	0.488	-1.182	-0.024	0.042	-0.129
苯甲醛	-3.291	-2.078	0.581	0.243	-0.197	-0.208	0.064	0.038	-2.283	-0.662	-1.530	-0.120	-0.057	-0.167	-0.017	0.010	-0.018	0.045	-0.001	0.001	-0.002	0.007	-0.038	-0.023	-0.027	-0.002	-0.002	-0.003
桃醛	-3.474	-3.387	1.544	-1.609	-0.208	-0.339	0.170	-0.252	-1.199	-0.174	-0.614	-0.063	-0.015	-0.067	-2.806	-2.738	0.827	-1.615	-0.168	-0.274	0.091	-0.253	-1.066	-0.604	-0.137	-0.056	-0.052	-0.015
(E)-2-癸烯醛	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2-十一烯醛	1.403	-0.310	-1.517	4.130	0.084	-0.031	-0.167	0.647	0.609	-2.194	-3.563	0.032	-0.189	-0.389	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2-甲基吡嗪	0.033	-0.020	-0.227	0.472	0.002	-0.002	-0.025	0.074	-0.095	-0.313	-0.403	-0.005	-0.027	-0.044	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2-乙基-6-甲基吡嗪	0.668	-0.250	-0.645	0.798	0.040	-0.025	-0.071	0.125	0.590	-0.906	-0.586	0.031	-0.078	-0.064	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2-乙基-5-甲基吡嗪	-0.551	0.590	-0.036	1.002	-0.033	0.059	-0.004	0.157	-1.104	0.186	-0.760	-0.058	0.016	-0.083	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2-乙基-3-甲基吡嗪	0.401	0.030	-0.136	0.581	0.024	0.003	-0.015	0.091	0.266	-0.174	-0.513	0.014	-0.015	-0.056	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
3-乙基-2,5-二甲基吡嗪	-0.067	0.030	-0.091	-0.006	-0.004	0.003	-0.010	-0.001	-0.133	-0.081	0.027	-0.007	-0.007	0.003	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
3,5-二乙基-2-甲基吡嗪	-1.954	5.026	-2.916	-1.787	-0.117	0.503	-0.321	-0.280	-4.967	-0.128	5.414	-0.261	-0.011	0.591	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
糠醛	-0.184	3.187	-1.454	-1.360	-0.011	0.319	-0.160	-0.213	-1.694	0.464	3.902	-0.089	0.040	0.426	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
糠醇	-0.969	2.708	-0.945	-1.334	-0.058	0.271	-0.104	-0.209	-2.303	0.685	3.188	-0.121	0.059	0.348	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
5-甲基糠醛	-0.535	0.500	-1.544	-1.155	-0.032	0.050	-0.170	-0.181	-1.199	-1.393	1.484	-0.063	-0.120	0.162	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
麦芽酚	0.902	-0.330	-1.227	0.626	0.054	-0.033	-0.135	0.098	0.818	-1.556	-0.156	0.043	-0.134	-0.017	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
甲基环戊烯醇酮	0.184	0.100	-0.127	-0.051	0.011	0.010	-0.014	-0.008	0.152	-0.070	0.202	0.008	-0.006	0.022	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2-乙酰基吡咯	-0.167	0.050	-0.164	0.364	-0.010	0.005	-0.018	0.057	-0.304	-0.197	-0.266	-0.016	-0.017	-0.029	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
2,3-二氢-3,5-二羟基-6-甲基-4H-吡喃-4-酮	14.249	-2.728	-2.053	-1.698	0.853	-0.273	-0.226	-0.266	16.822	-3.216	2.382	0.884	-0.277	0.260	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
己醛	-0.017	0.010	0.082	0.115	-0.001	0.001	0.009	0.018	-0.038	0.081	-0.165	-0.002	0.007	-0.018	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
苯乙醛	5.345	3.847	7.159	-0.013	0.320	0.385	0.788	-0.002	5.385	10.297	-0.330	0.283	0.887	-0.036	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
葵醛	-0.134	-0.180	0.100	-0.083	-0.008	-0.018	0.011	-0.013	0.000	0.012	-0.037	0.000	0.001	-0.004	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
α -亚乙基-苯乙醛	0.284	-0.100	0.027	0.070	0.017	-0.010	0.003	0.011	0.343	-0.035	-0.119	0.018	-0.003	-0.013	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
椰子醛	-0.017	0.010	0.073	0.096	-0.001	0.001	0.008	0.015	-0.038	0.070	-0.147	-0.002	0.006	-0.016	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
己酸	-0.334	-0.380	0.372	0.006	-0.020	-0.038	0.041	0.001	-0.076	0.174	-0.366	-0.004	0.015	-0.040	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
辛酸	-0.785	-0.629	0.436	0.134	-0.047	-0.063	0.048	0.021	-0.552	0.035	-0.843	-0.029	0.003	-0.092	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007
壬酸	-0.568	-0.250	0.109	0.319	-0.034	-0.025	0.012	0.050	-0.666	-0.128	-0.788	-0.035	-0.011	-0.086	0.000	0.560	-0.581	0.683	0.000	0.056	-0.064	0.107	-0.495	-0.348	-0.064	-0.026	-0.030	-0.007

续表 4

挥发性成分	高渗仍梯度				低渗仍梯度											
	主成分载荷矩阵				主成分特征向量矩阵				主成分载荷矩阵				主成分特征向量矩阵			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
庚酸	-0.017	-0.020	0.009	-0.013	-0.001	-0.002	0.001	-0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
月桂醇	-0.017	0.000	-0.009	0.006	-0.001	0.000	-0.001	0.001	-0.019	-0.012	-0.009	-0.001	-0.001	-0.001	-0.001	-0.001
(E)-2-十二烯醛(Z)-2,3-二甲基丙烯醛	-0.017	0.010	-0.018	-0.019	-0.001	0.001	-0.002	-0.003	-0.019	-0.012	0.018	-0.001	-0.001	-0.001	0.002	0.002
2-甲基四氢呋喃-3-酮	-0.017	0.010	-0.018	-0.019	-0.001	0.001	-0.002	-0.003	-0.019	-0.012	0.027	-0.001	-0.001	-0.001	0.003	0.003
2-N-辛基呋喃	-0.017	0.020	-0.036	-0.045	-0.001	0.002	-0.004	-0.007	-0.038	-0.023	0.055	-0.002	-0.002	-0.002	0.006	0.006
5-乙酰氧基甲基-2-呋喃醛	-0.033	0.030	-0.055	-0.070	-0.002	0.003	-0.006	-0.011	-0.057	-0.035	0.082	-0.003	-0.003	-0.003	0.009	0.009
2-乙基-2-己烯醛	-0.100	0.100	-0.173	-0.217	-0.006	0.010	-0.019	-0.034	-0.190	-0.128	0.256	-0.010	-0.010	-0.011	0.028	0.028
2,5-二氢-2,5-二甲基呋喃	-0.017	0.100	0.000	-0.057	-0.001	0.010	0.000	-0.009	-0.057	0.070	0.119	-0.003	0.006	0.006	0.013	0.013
呋喃酮	0.000	0.020	0.000	-0.013	0.000	0.002	0.000	-0.002	-0.019	0.012	0.027	-0.001	0.001	0.001	0.003	0.003
5-羟基尿嘧啶	-0.184	0.779	0.009	-0.453	-0.011	0.078	0.001	-0.071	-0.457	0.534	0.953	-0.024	0.046	0.046	0.104	0.104
惕格酸	0.451	0.929	1.172	-0.243	0.027	0.093	0.129	-0.038	0.343	1.881	0.366	0.018	0.162	0.162	0.040	0.040
乙酯	0.017	0.010	0.036	0.006	0.001	0.001	0.004	0.001	0.019	0.046	-0.009	0.001	0.004	-0.001	0.001	0.001
2-己烯酸乙酯	0.033	0.020	0.073	0.006	0.002	0.002	0.008	0.001	0.038	0.093	-0.018	0.002	0.008	-0.002	0.008	-0.002
2-甲基丁醇	0.401	0.290	0.963	0.077	0.024	0.029	0.106	0.012	0.476	1.231	-0.293	0.025	0.106	-0.032	0.032	0.032
(1-乙基-1-丙烯基)-苯	0.050	0.040	0.136	0.013	0.003	0.004	0.015	0.002	0.057	0.174	-0.037	0.003	0.015	-0.004	0.004	0.004
3-甲基丁醛	1.320	-0.480	-0.554	-0.204	0.079	-0.048	-0.061	-0.032	1.636	-0.824	0.357	0.086	-0.071	0.039	0.039	0.039
DL-泛酰内酯	0.418	-0.150	-0.191	-0.070	0.025	-0.015	-0.021	-0.011	0.514	-0.267	0.128	0.027	-0.023	0.014	0.014	0.014
2,4-二羟基-2,5-二甲基-3(2H)-呋喃-3-酮	0.184	-0.070	-0.100	-0.038	0.011	-0.007	-0.011	-0.006	0.228	-0.139	0.064	0.012	-0.012	0.007	0.007	0.007
2-吡咯烷酮	0.084	-0.030	-0.091	-0.026	0.005	-0.003	-0.010	-0.004	0.095	-0.116	0.064	0.005	-0.010	0.007	0.007	0.007
(Z)-2-庚烯醛	-0.501	-0.669	0.318	-0.396	-0.030	-0.067	0.035	-0.062	0.000	0.000	-0.009	0.000	0.000	-0.001	-0.001	-0.001
辛醛	-0.651	-0.869	0.418	-0.517	-0.039	-0.087	0.046	-0.081	0.000	0.000	-0.009	0.000	0.000	-0.001	-0.001	-0.001
(E)-2-壬烯醛	-0.200	-0.270	0.127	-0.160	-0.012	-0.027	0.014	-0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(E,E)-2,4-癸二烯醛	-0.635	-0.849	0.400	-0.498	-0.038	-0.085	0.044	-0.078	0.000	0.000	-0.009	0.000	0.000	-0.001	-0.001	-0.001
1-庚醇	-0.200	-0.270	0.127	-0.160	-0.012	-0.027	0.014	-0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1-辛醇	-0.367	-0.480	0.227	-0.287	-0.022	-0.048	0.025	-0.045	0.000	0.000	-0.009	0.000	0.000	-0.001	-0.001	-0.001
2,4-二甲基己烷	-1.687	-2.248	1.072	-1.334	-0.101	-0.225	0.118	-0.209	-0.019	0.000	-0.027	-0.001	0.000	-0.003	-0.003	-0.003
1-十五烯	-0.434	-0.580	0.273	-0.345	-0.026	-0.058	0.030	-0.054	0.000	0.000	-0.009	0.000	0.000	-0.001	-0.001	-0.001

不同掺伪梯度下逻辑回归模型的掺伪鉴别精度指标见表5。

表5 不同掺伪梯度下逻辑回归模型的掺伪鉴别精度指标

油茶籽油	掺伪梯度	精确率/ %	准确率/ %	召回率/ %	F1 - 分数/ %	Kappa 系数
原香油	高	100	87.5	85.7	92.3	0.60
茶籽油	低	100	83.3	80.0	88.9	0.57
烤香油	高	100	93.8	92.9	96.3	0.76
茶籽油	低	100	91.7	90.0	94.7	0.75

由表5可知,逻辑回归模型对原香和烤香油茶籽油高掺伪浸出油茶籽油的鉴别精确率(预测为掺伪的样本中真实掺伪样本的比例)都为100%,准确率(预测正确的样本占有所有样本的比例)分别为87.5%和93.8%,召回率(真实掺伪样本中被预测正确的比例)分别为85.7%和92.9%,F1 - 分数分别为92.3%和96.3%,Kappa系数分别为0.60和0.76。逻辑回归模型对原香和烤香油茶籽油低掺伪浸出油茶籽油的鉴别精确率都为100%,准确率分别为83.3%和91.7%,召回率分别为80.0%和90.0%,F1 - 分数分别为88.9%和94.7%,Kappa系数分别为0.57和0.75。

4 结论

本文构建的PCA模型和逻辑回归模型能够对风味油茶籽油是否掺伪浸出油茶籽油进行有效的定性判别。其中,对于原香油茶籽油,PCA模型对高掺伪梯度下油茶籽油样本是否掺伪浸出油茶籽油的检出限为20%,但对低掺伪梯度下样本是否掺伪不具有准确的鉴别能力。对于烤香油茶籽油,PCA模型对高掺伪梯度下样本是否掺伪浸出油茶籽油的检出限为60%,但对低掺伪梯度下样本是否掺伪同样不具有准确的鉴别能力。相比之下,逻辑回归模型具有更强的定性判别能力。对于原香及烤香油茶籽油,逻辑回归模型对高掺伪梯度下样本是否掺伪浸出油茶籽油的检出限均为10%,对低掺伪梯度下样本是否掺伪浸出油茶籽油的检出限均为4%。烤香油茶籽油挥发性成分多,运用机器学习算法可训练的数据集较多,建立的模型精准度较高。在后续研究中,本研究团队将基于本文研究成果,继续优化油茶籽油掺伪鉴别模型,以降低掺伪检出限。

参考文献:

[1] 姚小华,王开良,任华东,等. 油茶资源与科学利用研究[M]. 北京:科学出版社,2012.

[2] ROBARDS K,PRENZLE P,RYAN D, et al. Camellia oil and tea oil[M]// MOREAU R A, KAMAL - ELDIN A. Gourmet and health - promoting specialty oils. Urbana Illinois: AOCS Press, 2009: 313 - 343.

[3] 邓小莲,谢光盛,黄树根. 保健茶油的研制及其调节血脂的作用[J]. 中国油脂,2002,27(5):96 - 98.

[4] 钟小荣. 山茶油营养价值和发展研究[J]. 中国食品工业,2021(22):72 - 75, 128.

[5] 赵淑娟,郭平,万建春,等. 山茶油掺伪鉴别技术研究进展[J]. 食品工业,2018,39(5):306 - 309.

[6] SHI T, WU G C, JIN Q Z, et al. Camellia oil adulteration detection using fatty acid ratios and tocopherol compositions with chemometrics [J/OL]. Food Control,2022, 133:108565 [2022 - 10 - 20]. <https://doi.org/10.1016/j.foodcont.2021.108565>.

[7] MUNCK L, NØRGAARD L, ENGELSENS B, et al. Chemometrics in food science: a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance [J]. Chemometr Intell Lab, 1998, 44: 31 - 60.

[8] 庞妍妍,陈敏,王蓓,等. 基于吸收与散射特性的掺伪山茶油检测研究[J]. 粮食科技与经济,2020,45(11):90 - 94,130.

[9] 张岑. 折射率法测定茶油掺伪[J]. 中外交流,2021,28(5):514.

[10] 余佳荣,晁燕,谭利娟. 基于气相色谱法的湖南油茶籽油掺杂判定[J]. 中南林业科技大学学报,2014,34(5):88 - 91.

[11] 陈沙,陈春燕,李晴,等. 基于GC和GC - MS技术检测山茶油掺伪[J]. 食品安全导刊,2021(28):59 - 62.

[12] 石婷. 核磁共振技术用于茶油品质快速检测方法的建立及其应用[D]. 南昌:南昌大学,2018.

[13] LI S F, ZHU X R, ZHANG J H, et al. Authentication of pure camellia oil by using near infrared spectroscopy and pattern recognition techniques [J]. J Food Sci,2012, 77(4): C374 - C380.

[14] 邓平建,梁裕,杨冬燕,等. 基于拉曼光谱 - 聚类分析快速鉴别掺伪茶籽油[J]. 中国粮油学报,2016,31(4):72 - 75, 80.

[15] 孙艳辉,吴霖生,翁长晟,等. 应用同步荧光光谱和支持向量机快速鉴别油茶籽油真伪[J]. 食品工业科技,2012,33(4):52 - 55.

[16] 鲁小利,王俊. 仿生电子鼻在芝麻油掺伪检测中的应用研究[J]. 粮食与油脂,2016,29(6):75 - 77.

[17] 郭丽. 时间分辨荧光技术在食用植物油品质检测中的应用研究[D]. 江苏 镇江:江苏大学,2017.

[18] 张菊华,朱向荣,苏东林,等. 茶油品质鉴别的透射和透反射模式分析比较[J]. 食品与机械,2012,28(1):101 - 104, 218.

[19] 张菊华,朱向荣,李高阳,等. 近红外光谱法结合化学计量学方法用于茶油真伪鉴别分析[J]. 分析化学,2011,39(5):748 - 752.

[20] 王泽富,吴雪辉. 基于红外光谱快速鉴别压榨油茶籽油与浸出油茶籽油的研究[J]. 中国油脂,2018,43(11):63 - 68.

[21] 朱晓阳. 茶油主要感官品质的分析及其在掺伪鉴别中的应用[D]. 长沙:中南林业科技大学,2019.