

doi: 10.12452/j.fxcxb.240816316

基于近红外光谱技术的文冠果油人工掺杂快速分析研究

陈启文¹, 向超群¹, 李欣怡¹, 余丹华¹, 王乐琪^{1,2}, 肖雪^{1,2*}

(1. 广东药科大学 广东省代谢病中西医结合研究中心(中医药研究所), 广东 广州 510006; 2. 济源市广东药科大学神经健康产业研究院, 河南 济源 454600)

摘要: 该研究以文冠果油为例, 以5%为梯度, 人工制备文冠果油含量为0%~100%的文冠果油掺杂样品并采集其近红外光谱图。针对文冠果油掺杂油的鉴别问题, 采用偏最小二乘法-判别分析(PLS-DA)构建不同品种掺杂油的识别方法, 结合不同预处理方法比较优选并优化了模型性能。采用偏最小二乘回归法(PLSR)构建掺杂油的掺杂比例预测模型, 结合不同光谱预处理方法和不同波长选择方法优选并优化了模型性能。结果表明, 通过卷积一阶导数预处理(1st Dec), PLS-DA模型鉴别不同品种掺杂油达到了98%的准确率; 结合1st Dec和竞争自适应重加权采样变量选择(CARS), 文冠果油中不同品种掺杂油的近红外光谱模型的相对分析误差(RPD)分别为5.90、40.00、5.20和4.90, 显示出良好的预测性能。该研究实现了文冠果油样品在不同掺杂品种和比例下的定性定量分析, 为文冠果油掺杂的质量快检开展了探索性研究。

关键词: 文冠果油; 掺杂; 近红外光谱; 定性分析; 定量分析

中图分类号: O657.3; R284.1 **文献标识码:** A **文章编号:** 1004-4957(2025)02-0253-06

Research on Rapid Analysis of Artificially Mixed *Xanthoceras Sorbifolia* Oil Based on Near-infrared Spectroscopy Technology

CHEN Qi-wen¹, XIANG Chao-qun¹, LI Xin-yi¹, YU Dan-hua¹, WANG Le-qi^{1,2}, XIAO Xue^{1,2*}

(1. Guangdong Metabolic Diseases Research Center of Integrated Chinese and Western Medicine (Institute of Traditional Chinese Medicine), Guangdong Pharmaceutical University, Guangzhou 510006, China; 2. Jiyuan Neurohealth Industry Research Institute of Guangdong Pharmaceutical University, Jiyuan 454600, China)

Abstract: The issue of adulteration in the high-end edible oil market is becoming increasingly severe, with a notable lack of effective rapid detection methods. In this study, *Xanthoceras sorbifolia* oil was used as a case study. Artificially adulterated samples with *Xanthoceras sorbifolia* oil content ranging from 0% to 100% at 5% increments were prepared, and their near-infrared (NIR) spectra were collected. Partial least squares discriminant analysis (PLS-DA) was employed to develop a method for identifying adulterated *Xanthoceras sorbifolia* oil. Several preprocessing techniques were evaluated and optimized to enhance the performance of the model. To address the issue of quantifying the adulteration ratio in *Xanthoceras sorbifolia* oil, a model was constructed using partial least squares regression (PLSR). Model performance was further optimized by integrating different spectral preprocessing techniques and wavelength selection methods. The results indicate that the PLS-DA model, using convolutional first derivative preprocessing, achieved a 98% accuracy rate in identifying different types of adulterated oils. Furthermore, combining first derivative deconvolution preprocessing with competitive adaptive reweighted sampling (CARS) variable selection yielded NIR model RPD values of 5.90, 40.00, 5.20 and 4.90 for the adulterated oils in *Xanthoceras sorbifolia* oil, demonstrating excellent predictive performance. This study successfully conducted both qualitative and quantitative analyses of *Xanthoceras sorbifolia* oil samples adulterated with various types and proportions of other oils, offering an exploratory approach for rapid quality assessment of *Xanthoceras sorbifolia* oil adulteration.

Key words: *Xanthoceras sorbifolia* oil; adulteration; near-infrared spectroscopy; qualitative anal-

收稿日期: 2024-08-16; 修回日期: 2024-10-03

基金项目: 广东省中医药局中医药科研项目(科研平台专项)(20224046); 中国仪器仪表学会科学仪器托举计划项目(CISTJ2024); 国家药品监督管理局药品快速检验技术重点实验室开放课题(KF2022006)

* 通讯作者: 肖雪, 博士, 副研究员, 研究方向: 中药分析与质量评价, E-mail: erxiaohappy@163.com

ysis; quantitative analysis

随着高端药食两用植物油脂产量和消费量的持续增长,此类产品面临着日益严重的掺杂问题。油脂掺杂不仅侵犯了消费者权益,还对消费者的健康构成潜在威胁,对高端油脂产业的健康发展造成了不利影响^[1-3]。文冠果油是我国一种特种木本粮油,在食品、医药和保健领域应用前景广阔^[4]。然而,由于文冠果油的产出效率低下和供求错位,市场上的文冠果油产品容易受到掺杂问题的影响,对文冠果产业的可持续发展构成了挑战。

近年来已有多种检测技术被应用于油脂的掺假检测,包括基于理化参数的传统检测方法、电子鼻技术、液相色谱-串联质谱法和多光谱成像技术等^[5-10]。但上述方法或因成本高、操作复杂,或因检测效率低、维护成本高,不适用于即时检测。近红外光谱(NIRS)技术具有快速、无损、绿色和低成本等优点,在多个行业中得到了广泛应用^[11-13]。NIRS通过分析物质内部氢键振动的倍频和合频吸收,收集物质内部信息,实现对样品成分的快速定性和定量分析。尽管国内外研究人员已使用NIRS对多种油类样品进行了定性和定量鉴别研究^[14-16],但针对文冠果油的掺杂定性及定量鉴别研究较少。本研究采用有监督识别算法,结合不同的光谱预处理技术和变量选择方法,构建文冠果油人工掺杂油的定性识别和定量分析模型,旨在建立一种快速、准确的文冠果油掺杂产品检测方法,为实现其质量快检提供有价值的参考。

1 实验部分

1.1 仪器与样品

SupNIR-2600型近红外光谱仪(聚光科技(杭州)股份有限公司)、YP50001B型电子天平(上海市力辰仪器科技有限公司)、超声波清洗机(德国Elma公司);Vortex-2型涡旋混匀机(上海沪析实业有限公司)、RZSCF220-65-02L型超临界萃取设备(南通睿智超临界科技发展有限公司)。文冠果油(深圳极致本草生物医药有限公司提供),金龙鱼玉米油(益海(广州)粮油有限公司)、蓓琳娜特级初榨橄榄油(西班牙EXIOM FOOD S.L.)、桃仁油和茶油为实验室自制。

1.2 文冠果油掺假样品制备

样品组别:文冠果油、茶油、橄榄油、桃仁油、玉米油、茶油-文冠果油混合油、橄榄油-文冠果油混合油、桃仁油-文冠果油混合油、玉米油-文冠果油混合油。除各种油品外,其余4组,分别以5%的梯度递增(0%~100%)配制文冠果油比例,制得适量文冠果油混合油样品。

1.3 光谱采集、模型建立和软件

实验前,将近红外光谱仪预热30 min进行仪器自检,以保证仪器运行稳定。基于透射方式进行样品分析,波长范围为1 000~1 800 nm,分辨率为1 nm,温度为25 ℃,湿度为60%,软件为RIMP Client。以空白背景作为参比。样品置于2 mm石英比色皿的2/3,每个样品平行扫描采集5张光谱作为该样品的光谱。

采用偏最小二乘法对所采集的数据建立定性、定量模型。分别采用原始光谱(None)、标准正态变量变换(SNV)、一阶解卷积导数(1st Dec)、二阶解卷积导数(2nd Dec)、矢量归一化(VN)对数据进行预处理,比较不同预处理方法对定量模型的影响,选取最优的模型。在定性分类模型中,若模型具有较高的准确率,则认为结果较佳。在定量模型中,若模型具有较高的决定系数(R^2)、相对分析误差(RPD),以及较低的交叉验证均方根误差(RMSECV)、预测均方根误差(RMSEP),则认为结果较佳。

1.4 数据处理

采用The Unscrambler X 10.4(挪威Camo analytics公司, Demo)软件对光谱数据进行预处理和偏最小二乘法(PLS)模型的建立; Matlab软件(R2024a, MathWorks, Natick, 美国)对光谱数据进行竞争自适应重加权采样变量选择(CARS)算法、最小绝对收缩与选择算子(LASSO)算法处理,实现异常样本的筛选和变量选择。CARS算法采用个体逐步淘汰的选择机制,通过指数衰减函数控制变量的保留率,引入了模型集群分析(MPA)的概念,对模型子集所选定的参数进行有效统计分析^[17]。LASSO算法则是一种基于线性回归的特征筛选方法,通过在模型中引入范式惩罚项,将回归系数的绝对值限制在预设

的阈值内,同时最小化模型的残差平方和。该方法通过产生稀疏矩阵来识别重要的特征,其中非零系数对应的特征被选为有效特征^[18]。LASSO算法在机器学习领域被广泛用于解决过拟合问题。

2 结果与讨论

2.1 光谱分析和主成分分析

图1展示了各类不同比例掺杂的油样品经不同预处理方法处理后的近红外光谱。图1A中,原始光谱之间高度重叠,使得目标成分的光谱信息难以区分。在采集近红外光谱数据时,光谱常受到散射、背景噪声等因素的影响。因此,在建立模型之前,必须应用化学计量学方法对光谱数据进行适当的预处理。本研究采用SNV、1st Dec、2nd Dec和VN等预处理技术,以改善光谱特征的可区分性,图1展示了预处理后的光谱结果。

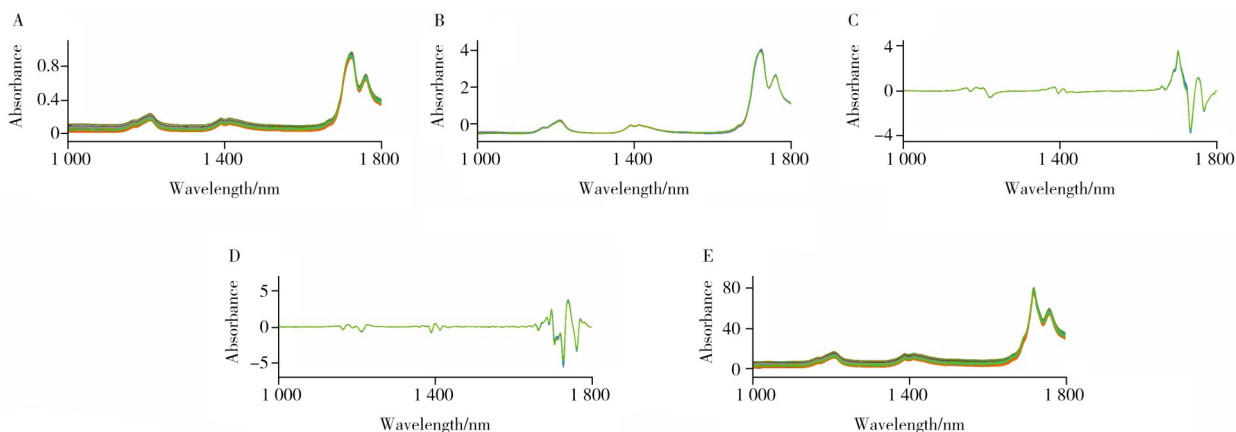


图1 各类不同比例掺杂油样品经预处理后的近红外光谱图

Fig. 1 NIRS of oil samples with various adulteration ratios following different pretreatments

A: none; B: SNV; C: 1st Dec; D: 2nd Dec; E: VN

为了初步评估4种文冠果油掺杂样品的分类情况,本研究对415批样品进行了主成分分析(PCA),结果如图2所示。PCA是一种常用的降维技术,用于提取数据集中的关键信息,并可作为无监督模式识别的工具。图2展示了基于前3个主成分中第一主成分(PC1)、第二主成分(PC2)和第三主成分(PC3)的散点图,这些成分共同解释了99.04%的数据方差。具体来说,PC1、PC2和PC3大约分别解释了方差的95%、3%和2%。从图2的散点图分析中可以看出,4种文冠果油掺杂样品之间的区分度并不理想,这可能是由基线漂移、噪声和光散射效应等因素所致。结果表明,单纯依赖PCA分析可能不足以有效识别和区分4种文冠果油掺杂样品。

2.2 分类模型的优化和建立

基于对PCA结果的深入分析,进一步采用偏最小二乘法-判别分析(PLS-DA)技术构建了分类模型,以实现更优的分类效果。在模型建立过程中,从415个样本中通过Kennard-Stone(KS)算法随机选取样本,按照大约3:1的比例划分为校正集和预测集,其中校正集包含315个样本,预测集包含100个样本。为保证各类别样本的代表性,本研究采用分层抽样方法。此外,为评估模型的稳定性并防止过拟合,本研究应用10倍交叉验证方法,并比较了不同预处理方法对模型性能的影响,具体结果见表1。

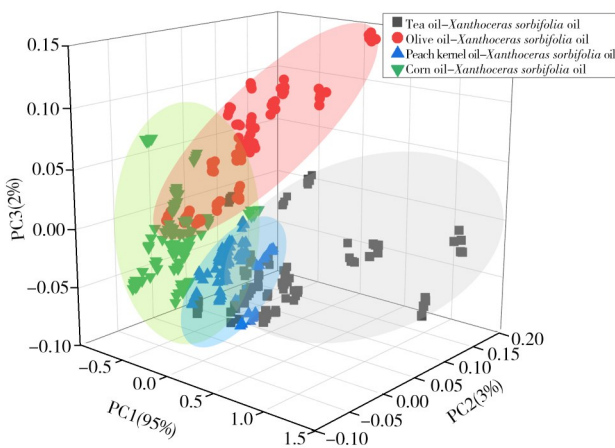


图2 4种调和样品的PCA因子三维图

Fig. 2 PCA factor three-dimensional graph of four blended samples

表1 PLS-DA 建模预处理优化结果

Table 1 Optimization results of PLS-DA model by different pretreatments

Method	LVs	Calibration set accuracy/%	Prediction set accuracy/%
None	8	96	95
SNV	7	96	96
1 st Dec	5	98	96
2 nd Dec	5	97	96
VN	9	98	98

优化后的模型性能分析显示, 经过预处理的光谱数据在分类能力上有所提升。特别是在 1st Dec 预处理条件下, PLS-DA 模型展现出较好的鉴别能力, 这表现在最优主成分的数量减少和模型性能的提升上。图 3 进一步展示了预处理对模型性能的积极影响, 其中 1st Dec 预处理后的模型在定性鉴别分析中表现最为突出。

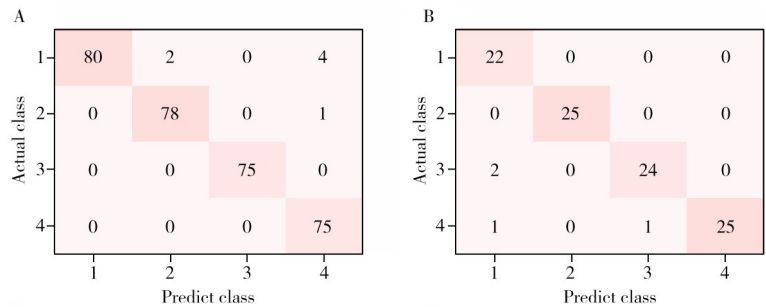


图3 最优预处理后模型的混淆矩阵

Fig. 3 Confusion matrix of the optimal pretreatment models

A: cross-validation set; B: prediction set

2.3 定量模型的优化和建立

为分析掺杂样品中掺杂品种的含量, 建立并优化了定量模型。在建立定量模型之前, 基于光谱-理化值共生距离 (SPXY) 算法, 按照大约 3 : 1 的比例选取校正集和预测集, 如表 2 所示。由于 SPXY 算法同时考虑了光谱和样品浓度, 因此校正集的浓度范围涵盖了验证集的浓度范围。建立了 PLSR 定量检测模型, 并比较了不同预处理方法和不同波长选择方法对文冠果油掺杂定量模型预测结果的影响。

表2 定量建模数据集划分

Table 2 Quantitative modeling dataset division

Sample	Total sample	Calibration set sample	Prediction set sample
茶油-文冠果油	110	75	35
橄榄油-文冠果油	105	75	30
桃仁油-文冠果油	110	75	35
玉米油-文冠果油	110	75	35

2.3.1 预处理比较 通过选取前 12 个潜在变量数对样品的光谱数据进行降维, 结合 10 折交叉验证对文冠果油掺杂样品集建立了 4 个偏最小二乘回归 (PLSR) 定量检测模型, 通过对比 None、VN、SNV、1st Dec、2nd Dec 对光谱的预处理效果, 比较不同预处理对建模结果的影响。结果如表 3 所示, 4 种油类掺杂文冠果油的检测中, 经上述方法预处理后, 模型校正集、交叉验证集的预测精度较原始光谱均得到提升。其中 1st Dec 预处理后的文冠果油掺杂茶油、橄榄油和桃仁油模型结果最佳, 而 SNV 预处理得到的文冠果油掺杂玉米油模型最稳定, 预测能力最好。经 1st Dec 预处理后预测集的精度略有提升, 同时校正集的过拟合现象有所缓解, 模型稳定性得到增强。

表3 定量模型预处理比较

Table 3 Comparison of quantitative model preprocessing

Sample	Method	LVs	Calibration set			Prediction set		
			R^2_c	RMSECV	RMSE	R^2_p	RMSEP	RPD
茶油-文冠果油	None	9	0.83	14.01	10.20	0.88	13.28	2.91
	2 nd Dec	8	0.83	13.78	13.50	0.83	14.23	2.43
	1 st Dec	8	0.85	12.86	10.70	0.88	12.04	2.97

(续表3)

Sample	Method	LVs	Calibration set			Prediction set		
			R^2_c	RMSECV	RMSE	R^2_p	RMSEP	RPD
橄榄油-文冠果油	SNV	6	0.79	15.43	17.00	0.59	23.10	1.56
	VN	11	0.91	9.92	8.20	0.92	10.79	3.69
	None	9	0.95	1.89	1.40	0.99	1.55	23.50
	2 nd Dec	2	0.98	3.41	2.20	0.99	2.19	14.46
	1 st Dec	3	0.99	2.43	1.80	0.99	1.76	18.04
桃仁油-文冠果油	SNV	8	0.99	1.52	1.20	0.99	1.32	26.14
	VN	9	0.99	1.80	1.20	0.99	1.39	26.00
	None	6	0.59	20.02	16.30	0.64	19.16	1.68
	2 nd Dec	2	0.37	24.86	15.20	0.41	23.36	1.30
	1 st Dec	3	0.59	20.60	13.90	0.57	20.07	1.53
玉米油-文冠果油	SNV	7	0.38	17.64	13.30	0.76	14.87	2.05
	VN	6	0.59	20.06	16.40	0.64	19.25	1.67
	None	9	0.95	6.39	4.95	0.97	5.02	6.59
	2 nd Dec	4	0.93	7.70	7.17	0.94	8.26	4.27
	1 st Dec	4	0.94	6.90	5.69	0.97	5.57	5.90
	SNV	8	0.97	4.73	4.12	0.98	4.22	8.00
	VN	9	0.96	6.08	4.81	0.97	4.84	6.83

2.3.2 波长选择比较 为增强定量模型的适用性，本研究对几种常见的波长选择方法进行了比较分析，主要包括基于蒙特卡洛采样的CARS算法和LASSO的特征选择算法。由表4可知，CARS或LASSO选择的波长模型在 R^2 方面较全波长模型表现更佳，且整体上CARS-PLS模型的性能更优。

表4 定量模型波长选择比较
Table 4 Comparison of wavelength selection for quantitative models

Sample	Method	Wavenumber	LVs	Calibration set			Prediction set		
				R^2_c	RMSECV	RMSE	R^2_p	RMSEP	RPD
茶油-文冠果油	CARS	144	6	0.93	9.18	5.62	0.97	5.85	5.90
	LASSO	60	10	0.93	7.33	5.31	0.97	6.05	6.40
	None	800	8	0.85	12.86	10.70	0.88	12.04	2.90
橄榄油-文冠果油	CARS	144	6	0.99	1.03	0.81	0.99	0.91	40.00
	LASSO	41	3	0.99	2.12	1.67	0.99	1.68	19.80
	None	800	3	0.99	2.43	1.80	0.99	1.76	18.00
桃仁油-文冠果油	CARS	163	7	0.94	7.30	5.42	0.96	6.15	5.20
	LASSO	66	4	0.89	10.07	7.43	0.93	8.35	3.80
	None	800	3	0.59	20.60	13.90	0.57	20.07	1.50
玉米油-文冠果油	CARS	128	5	0.92	8.53	6.14	0.95	7.14	4.90
	LASSO	86	5	0.88	10.53	10.09	0.87	12.32	2.70
	None	800	8	0.97	4.73	4.12	0.98	4.22	8.00

2.3.3 PLSR模型 为了更直观地展示表4中优选的模型结果，绘制了散点图(图4)。如图4所示，所有数据点均紧密聚集在对角线周围。4个CARS-PLSR定量模型均得到了 $R^2_p>0.90$ 和 $RPD>3.0$ 的结果，表现出良好的稳健性和预测准确性。样本中的RMSECV和RMSEP没有显著差异，表明模型训练未出现明显的过拟合，同时校正集的浓度范围覆盖了验证集的浓度范围。

3 结 论

本研究采用近红外光谱法对4种文冠果油掺杂制品进行定性和定量分析。在明确PLS-DA对4种文冠果油掺杂制品具有识别能力的基础上，进一步比较并筛选多种光谱预处理和波长选择方法，得到的最优模型在校正集和预测集上的 $R^2>0.90$ ， $RPD>3.0$ ，表明其具有非常好的预测能力。同样，PLS-DA对文冠果油掺杂制品具有优秀的识别能力，建立的CARS-PLS定量模型能准确预测掺杂比例成分。本研究表明近红外光谱结合化学计量学技术可作为文冠果油掺杂制品质量控制和快速评价的参考方法。

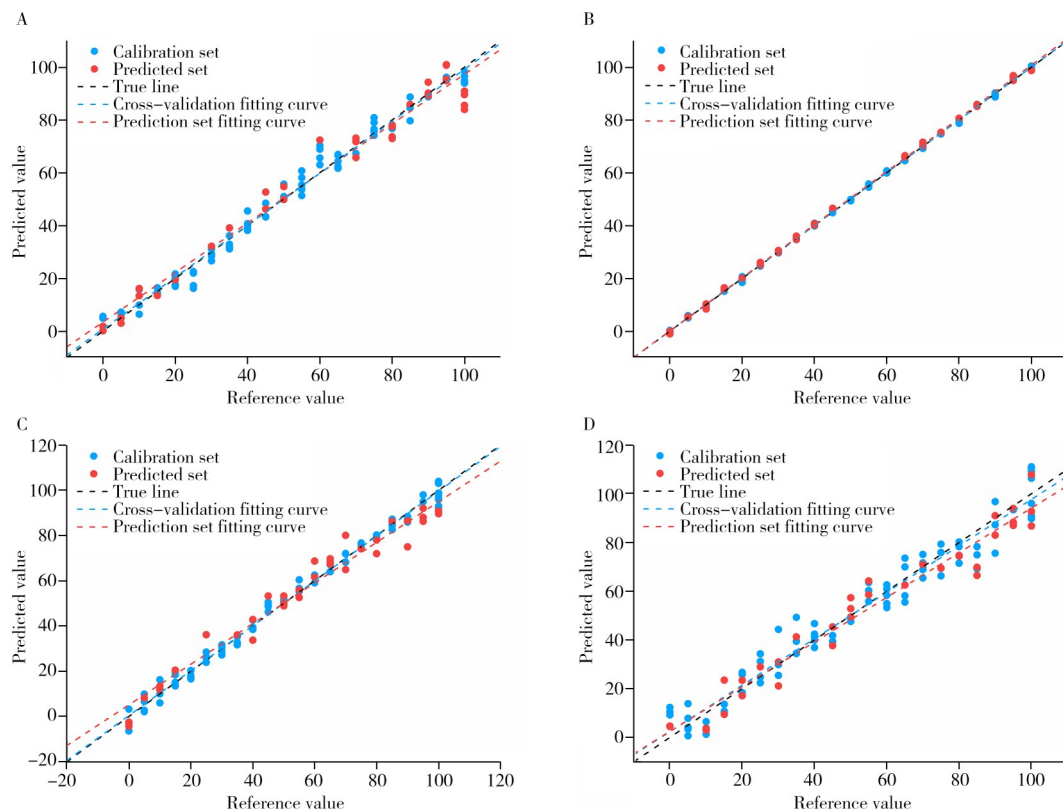


图4 PLSR内部验证结果

Fig. 4 Internal validation results of PLSR

A: tea oil-*Xanthoceras sorbifolia* oil; B: olive oil-*Xanthoceras sorbifolia* oil; C: peach kernel oil-*Xanthoceras sorbifolia* oil;
D: corn oil-*Xanthoceras sorbifolia* oil

参考文献:

- [1] Zhang J Y, Xu B B, Zheng J X. *China Oils and Fats* (张婧好, 许本波, 郑家喜. 中国油脂), **2022**, 47(3): 5-10.
- [2] Zhai J, Wu Z H, Chen Y, Wang J L. *Fine Chem. Intermed.* (翟潜, 伍振煌, 陈逸, 王俊龙. 精细化工中间体), **2023**, 53(2): 12-15.
- [3] Rifna E J, Pandiselvam R, Kothakota A, Subba R K V, Dwivedi M, Kumar M, Thirumdas R, Ramesh S V. *Food Chem.*, **2022**, 369(1): 1-16.
- [4] Xiao J, Sun L, Pan Y N, Bai X L, Chen G, Zhang X L, Chen X X, Li N. *Forests*, **2023**, 14(1): 1-9.
- [5] Jiang H, He Y C, Xu W D, Chen Q S. *Food Anal. Methods*, **2021**, 14: 1826-1835.
- [6] Drira M, Guclu G, Portolés T, Jabeur H, kelebek H, Selli S, Bouaziz M. *Food Anal. Methods*, **2021**, 14: 2121-2135.
- [7] Heidari M, Talebpour Z, Abdollahpour Z, Adib N, Ghanavi Z, Aboul-Enein H Y. *J. Food Sci. Technol.*, **2020**, 57(9): 3415-3425.
- [8] Dou X J, Mao J, Zhang L X, Xie H L, Chen L, Yu L, Ma F, Wang X P, Zhang Q, Li P W. *Molecules*, **2018**, 23(2): 1-10.
- [9] Puspita I, Irawati N, Madurani K A, Kurniawan F, Koentjoro S, Hatta A M. *Photonic. Sens.*, **2022**, 12(4): 1-10.
- [10] Libish T M, Bobby M C, Linesh J, Mathew S, Pradeep C, Nampoore V P N, Biswas P, Bandyopadhyay S, Dasgupta K, Radhakrishnan P. *Laser Phys.*, **2013**, 23(4): 1-6.
- [11] Srinath K, Kiranmayee A H, Bhanot S, Panchariya P C. *Mapan*, **2022**, 37(3): 483-493.
- [12] Sánchez-López E, Sánchez-Rodríguez M. I, Marinas A, Marinas J M, Urbano F J, Caridad J M, Moalem M. *Talanta*, **2016**, 156(15): 180-190.
- [13] Uncu O, Ozen B. *Food Control*, **2019**, 105: 209-218.
- [14] Ye Q, Meng X H. *Food Chem.*, **2022**, 385(15): 1-8.
- [15] Da Costa G B, Fernandes D D S, Gomes A A, de Almeida V E, Veras G. *Food Chem.*, **2016**, 196(1): 539-543.
- [16] Xu Y, Hassan M M, Kutsanedzie F Y H, Li H H, Chen Q S. *Qual. Assur. Saf. Crops Foods*, **2018**, 10(4): 411-421.
- [17] Wang K Y, Yang S, Guo C Y, Bian X H. *J. Instrum. Anal.* (王恺怡, 杨盛, 郭彩云, 卞希慧. 分析测试学报), **2022**, 41(3): 398-402.
- [18] Pfeiffer P, Ronai B, Vorlauffer G, Dörr N, Filzmoser P. *Chemom. Intell. Lab. Syst.*, **2022**, 228(15): 1-10.

(责任编辑: 丁 岩)