

DOI: 10.19812/j.cnki.jfsq11-5956/ts.20241110003

引用格式: 林伟琦. 基于傅里叶变换衰减全反射红外光谱鉴别山茶油掺假[J]. 食品安全质量检测学报, 2025, 16(2): 215-223.

LIN WQ. Identification of *Camellia* oil adulteration by attenuated total reflectance-Fourier transform infrared spectroscopy [J]. Journal of Food Safety & Quality, 2025, 16(2): 215-223. (in Chinese with English abstract).

基于傅里叶变换衰减全反射红外光谱鉴别 山茶油掺假

林伟琦*

(厦门市产品质量监督检验院, 厦门 361004)

摘要: 目的 基于傅里叶变换衰减全反射红外光谱(attenuated total reflectance-Fourier transform infrared spectroscopy, ATR-FTIR)结合聚类判别分析等化学计量学方法, 建立山茶油、大豆油、玉米油、葵花籽油和花生油 5 种植物油的快速鉴别模型, 及山茶油掺假模型。**方法** 采集山茶油、大豆油、玉米油、葵花籽油和花生油 5 种植物油共 99 份样品, 并按照不同质量百分比(掺伪 5%~95%)将大豆油、葵花籽油、玉米油、1:1 玉米大豆油、花生油、棕榈油掺入到山茶油中, 获得掺假山茶油样品 196 份, 采集 600~4000 cm^{-1} 波段的红外光谱信息, 建立偏最小二乘判别分析(partial least squares-discriminant analysis, PLS-DA)、主成分分析-判别分析(principal component analysis-linear discriminant analysis, PCA-LDA)、K 最近邻分类算法(K-nearest neighbor, KNN)以及数据驱动型簇类独立软模式分类(data driven soft independent modelling of class analogy, DD-SIMCA)模型, 并比较各方法建模效果, 确定最优识别模型。**结果** 各样品组红外吸收光谱非常类似, 具有相似的特征峰数、峰位置和峰形。DD-SIMCA 建立的鉴别模型能将山茶油和其他类别植物油样本完全分开; PLS-DA、PCA-LDA 和 KNN 模型判别经分析比较, 发现利用 PLS-DA 和 PCA-LDA 模型在 5 种植物油的分类中校正集和预测集中的各样品的预测值与实际值很接近, 除了花生油以外其余种类植物油的校正集和预测集样品的识别率和预测正确率均为 100.0%; ATR-FTIR 结合 PLS 的计量学方法能够准确进行山茶油掺假定量分析, 可用于掺杂大豆油、玉米油、葵花籽油等的定性定量分析, 结果可靠, 最低检出限可达 5%。**结论** ATR-FTIR 结合聚类判别分析等化学计量学方法实现对山茶油掺假的高效识别。

关键词: 傅里叶变换衰减全反射红外光谱; 山茶油; 掺假; 判别分析

Identification of *Camellia* oil adulteration by attenuated total reflectance-Fourier transform infrared spectroscopy

LIN Wei-Qi*

(Xiamen Products Quality Supervision & Inspection Institute, Xiamen 361004, China)

ABSTRACT: Objective To establish a rapid identification model for 5 different types of vegetable oils (*Camellia* oil, soybean oil, corn oil, sunflower seed oil and peanut oil) and adulterated *Camellia* oil, using attenuated total

收稿日期: 2024-11-10

基金项目: 厦门市自然科学基金项目(3502Z202374105)

第一作者/*通信作者: 林伟琦(1979—), 女, 高级工程师, 主要研究方向为食品安全检测。E-mail: linweiqi@xmzjy.org

reflectance-Fourier transform infrared spectroscopy (ATR-FTIR) and chemometrics methods such as cluster discriminant analysis. **Methods** The 99 samples of 5 different types of vegetable oils, including *Camellia* oil, soybean oil, corn oil, sunflower seed oil, and peanut oil were collected. According to the mass percentage of 5%–95%, soybean oil, sunflower seed oil, corn oil, 1:1 corn soybean oil, and palm oil was mixed into the *Camellia* oil, and 196 samples of the adulterated *Camellia* oil were obtained. Their infrared spectrum were collected in 600–4000 cm^{-1} region. The models for partial least squares discriminant analysis (PLS-DA), principal component analysis discriminant analysis (PCA-LDA), K-nearest neighbor (KNN), and data driven soft independent modeling of class analogy (DD-SIMCA) were established and compared to determine the optimal recognition model. **Results** The infrared spectra of each sample group had similar characteristic peaks, peak positions, and peak shapes were with slight differences. The discriminant model established by DD-SIMCA could completely separate *Camellia* oil samples from those of other types of vegetable oil. By comparison of PLS-DA, PCA-LDA, and KNN models, it was found that the predicted values of each sample in the training and testing sets of the classification of 5 types of edible vegetable oils samples using PLS-DA and PCA-LDA models were accurate and reliable. Except for peanut oil, the recognition and prediction accuracy of the training and testing sets of other edible vegetable oils were both 100.0%. The quantitative analysis of *Camellia* oil adulteration using ATR-FTIR combined with PLS could be accurately carried out, which could be used for qualitative and quantitative analysis of adulterated soybean oil, corn oil, sunflower seed oil, etc. The results were reliable, and the lowest limit of detection could reach 5%. **Conclusion** Adulterated *Camellia* oil can be determined accurately and efficiently based on ATR-FTIR combined with chemometric methods.

KEY WORDS: attenuated total reflectance-Fourier transform infrared spectroscopy; *Camellia* oil; adulteration; discriminant analysis

0 引言

我国是山茶油的原产地,也是世界上山茶油最主要的产地,优质山茶油在国际国内市场上倍受消费者青睐。2023 年国家林业和草原局、国家发展和改革委员会、财政部联合发布《加快油茶产业发展三年行动方案(2023—2025 年)》中明确目标为:3 年新增油茶种植 1917 万亩、改造低产林 1275.9 万亩,确保到 2025 年,全国油茶种植面积达到 9000 万亩以上、茶油产能达到 200 万 t。山茶油营养丰富,不饱和脂肪酸高达 90%以上,油酸达到 80%~83%、亚油酸达到 7%~13%,同时含有橄榄油所没有的特定活性物质山茶甙和茶多酚^[1-2]。油茶树生长结果需要 6~10 年,且多采取机械压榨法提取山茶油,生产加工工艺复杂,出油率不高,所以市场上的山茶油价格比其他食用植物油高 3~5 倍^[3-4]。然而市场上各种山茶油品质良莠不齐,甚至出现以低价的其他种类植物油掺入山茶油中以次充好的现象,因此开展快速无损、准确度高的山茶油掺假鉴别和品质检测方法的研究具有重要意义。

通常实验室进行理化检测酸价、过氧化值和碘值等指标来评价油品的品质等级与新鲜程度^[5]。酸价和过氧化值越低,油品氧化程度越低,就越新鲜。理化检测手段虽能较好的反应油脂的品质,但是操作复杂耗时较长,同时耗费试剂与人力,并且不能鉴别油脂品种,只能作为油脂品质检测的辅助手段。另一种较为常见的食用植物油掺假检测方法例如气相色谱法(gas chromatography, GC)或气相色

谱-质谱联用法(gas chromatography-mass spectrometry, GC-MS)^[6-7]、超高效液相色谱法(ultra performance liquid chromatography, UPLC)^[8]等。GC 可以通过检测油品的脂肪酸组成和含量,将脂肪酸甘油酯转化为脂肪酸甲酯后测定,用归一法确定各脂肪酸的组成比例^[9-10],再与国家标准中山茶油脂肪酸组成和含量比较,判别其质量。各类植物油的 40 余种脂肪酸组成成分类似,含量不同,仅用检测油品的脂肪酸组成和含量的方法来判别油类掺假情况工作量大而且较难实现^[11-12],且样品前处理时间长,设备操作复杂且维护成本高,不利于进行快速和实时的油品分析。因此,亟需建立一种准确、高效、无损的掺假山茶油鉴别技术。

红外光谱作为一种重要的化学分析技术,具有快速检测、特异性高、非破坏性等特点,同时能在几十秒甚至几秒内获得高质量的光谱数据,同时测定物质的多种成分数据^[13-15]。BAETEN 等^[16]利用傅里叶变换-拉曼光谱法分别鉴别出掺有 1%大豆油、5%玉米油和 10%初榨橄榄油的特级初榨橄榄油。YANG 等^[17]采用了近红外光谱法、傅里叶变换红外光谱法(Fourier transform infrared spectroscopy, FTIR)、傅里叶变换拉曼光谱法 3 种方法对掺入低级初榨橄榄油的特级初榨橄榄油进行鉴别和比较。匡俊豪等^[18]以玉米油、稻米油和大豆油作为掺假样品掺入山茶油中,得到山茶油的二元掺假样品和三元掺假样品,并对所有样品进行近红外拉曼光谱检测。可见红外光谱技术在食用油中掺假鉴定应用广泛。

食用植物油是一种复杂的混合物,含有甾醇、磷脂、

胡萝卜素、维生素 E 等多种物质,可能对红外/拉曼光谱产生干扰,给解析谱图增加了难度;甘油三酯的立体结构的差异也会给红外或拉曼光谱的分析带来一定的难度。而且,不同种类的食用油的 FTIR 具有相似的特征峰数、峰位置和峰形,很难单纯根据其红外光谱进行准确的分类^[19-25]。本研究拟建立基于傅里叶变换衰减全反射红外光谱(attenuated total reflectance-Fourier transform infrared spectroscopy, ATR-FTIR)数据解析的化学计量学算法用于山茶油的聚类、判别分析,建立识别假冒和掺假山茶油的方法,旨在获得一种能够高效鉴别假冒、掺假山茶油的新方法,为规范山茶油产品质量提供技术支持。

1 材料与方 法

1.1 材料与试剂

本研究收集了 5 种食用油的 99 个样本,分别是 17 个山茶油样本、22 个玉米油样本、40 个花生油样本、12 个葵花籽油样本和 8 个大豆油样本,见表 1。根据购买价格,可以看到山茶油是价格相对较高的植物油,其次为花生油。

无水乙醇(分析纯,国药集团化学试剂有限公司);正己烷(色谱纯,美国 Sigma 公司)

1.2 仪器与设备

NICOLET 6700 FTIR 傅里叶变换中红外光谱仪[红外附件: ZnSe 晶体材料水平衰减全反射(attenuated total reflectance, ATR)附件,美国赛默飞世尔科技公司]。

1.3 方 法

1.3.1 检测条件

扫描次数: 32; 光谱分辨率: 4 cm^{-1} ; 检测器: 氘化硫酸三甘肽晶体(deuterated triglycine sulfate, DTGS); 分束器: 溴化钾基片镀膜; 光源: 高强度空气冷却红外光源; 数据采集系统: OMNIC; 实验室温度: 25 $^{\circ}\text{C}$ 。

1.3.2 实验步骤

采集仪器本底(空气)光谱。保证 ATR 本底能量在 6.3 以上; 将样本混匀,滴入水平 ATR 附件中,覆盖 ATR 晶体表面,测量其红外光谱; 依次用正己烷、无水乙醇清洗

ATR 晶体后继续采集本底光谱; 测定下一个样本的光谱。选 4000~650 cm^{-1} 区间做光谱分析,所有的光谱数据都归一化后再进行处理。

1.4 化学计量学建模

设有 n 个油样,每个样品观测红外光谱 p 个波数吸光度,将收集的光谱原始数据以矩阵形式 X 表达,见式(1),其中 x_{np} 为第 n 个样品的第 p 个指标,并对其进行处理。

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

1.4.1 主成分分析

主成分分析(principal component analysis, PCA)是一种常用的无监督学习方法,通过正交变换将原始数据即量测矩阵 X 转换为一组新的正交变量,这些新变量称为主成分(principal component, PC)。先求出量测矩阵 X 的协方差阵,采用正交变换的方法来求出它的特征值和特征向量。

这些 PC 是原始数据的线性组合,且彼此之间线性无关,能够最大限度地保留原始数据的信息,达到了数据降维的效果。

1.4.2 主成分分析-判别分析法

判别分析法(linear discriminant analysis, LDA)是在样品标签已知的情况下,根据某一研究对象的各种特征值判别其类型归属问题,为有监督的分类模式。使用 LDA 法时样本数应大于自变量个数,所建的判别函数的拟合结果才好。将 LDA 应用在鉴别山茶油掺假时,拟将样品的光谱数据作为输入量,但由于特征变量也就是光谱数据的个数过大,超过了样本数,所以使用 PCA 来对光谱数据转换为一组新的正交变量,能保证光谱中特征信息不丢失,又降低了光谱中特征变量的维度。

1.4.3 偏最小二乘判别分析法

偏最小二乘法(partial least squares, PLS)主要是研究多个自变量与因变量之间的相关关系,即在自变量空间进行线性变化,同时因变量也进行线性变化,然后再将两种变换建立一定量的关系,并使两组变量之间的协方差最大,

表 1 实验中使用的植物油样本统计

Table 1 Statistics of plant oil samples used in the experiments

种类	样本数	等级	工艺	品牌	价格(500 mL)/元
山茶油	17	一级	压榨	金龙鱼、25 度、千岁好、山茶山、秋味坊、金浩、福临门、德尔乐、千岛源、迎福、广西金茶王(厂家)	52.0~112.0
玉米油	22	一级	压榨	西王、金龙鱼、长寿花、福临门、knife、傲鹏、鲁花、多力	7.5~20.0
大豆油	8	一级	压榨、浸出	元宝、盛洲、金龙鱼、傲鹏、嘉龙、御榨坊、福临门、惠宜	6.0~32.0
花生油	40	一级	压榨	金龙鱼、鲁花、knife、多里、福临门、盛洲、胡姬花、乐当家	9.3~46.0
葵花籽油	12	一级	压榨	金龙鱼、福临门、多力、鲁花、刀唛	8.8~15.0

这样各组分的贡献率最大,从而更能代表原始数据的信息。偏最小二乘法判别分析(partial least squares discriminant analysis, PLS-DA)是一种用于判别分析的多变量统计分析方法,主要用于处理分类和判别问题。其与 PCA 类似,不同的是 PCA 是无监督的,PLS-DA 是有监督的。PLS-DA 通过最大化组间的差异,使得组间差异最大化,组内差异最小化,从而更好地区分不同的样本组。其原理是对不同处理样本的特性分别进行训练,产生校正集,并检验校正集的可信度。可以 PLS-DA 建立特征光谱与样品类别之间的关系模型,来实现对样品类别的预测。

1.4.4 K 最近邻分类算法

K 最近邻分类算法(K-nearest neighbors, KNN)是一种简单且直观的监督学习算法,常用于分类和回归任务。其基本思想是:给定一个待分类的样本,找出与其距离最近的 K 个训练样本(邻居),然后通过这 K 个邻居的类别来决定待分类样本的类别。

KNN 分类算法的计算过程:计算待分类点与已知类别的点之间的距离;按照距离递增次序排序;选取与待分类点距离最小的 K 个点;确定前 K 个点所在类别的出现次数;返回前 K 个点出现次数最高的类别作为待分类点的预测分类。

作为 KNN 算法中唯一的一位超参数, K 值的选择对最终算法的预测结果会产生直观重要的影响。 K 值过大可能会导致模型泛化能力下降, K 值过小则可能导致模型过拟合。在应用中,通常采用交叉验证法来选择最优 K 值。

1.4.5 数据驱动的簇类独立软模式算法

簇类独立软模式分类(soft independent modelling of class analogy, SIMCA)是广泛用于化学计量学中的单分类器之一。它通过对每种类别进行无监督 PCA 建模分析,可以将样本关联到多个类或不关联任何类^[26]。MAGDAS 等^[27]的研究表明, SIMCA 分类与机器学习的预测模型在拉曼光谱判别蜂蜜产地和品种结果存在相关性,且两种算法分类结果正确率基本相同。有研究提出了基于 SIMCA 边界构建修改的数据驱动型簇类独立软模式分类(data driven soft independent modelling of class analogy, DD-SIMCA), DD-SIMCA 通过计算每个对象的得分距离和正交距离,从而建立两个公差阈值:给定显著性水平的接受区域和异常值区域,非常适合用于检测样品的掺假^[28-30]。

1.5 数据处理

数据采用 Omnic v8.2 光谱采集软件,实验结果使用 Origin 2020、MATLAB R2018a 数据处理软件进行处理,在个人计算机上运行,将植物油种类与光谱信息分别进行建模分析,并绘制聚类分析结果图、样本实际值和预测值相关图。

2 结果与分析

2.1 5 种食用油的红外光谱数据集

5 种食用油合计 99 个样本的 ATR-FTIR 图如图 1 所

示。从图 1 中可以看出各种食用油的 ATR-FTIR 图非常相似,很难只通过谱图区分各类食用油样本。5 种食用油的红外吸收光谱具有相似的特征峰数、峰位置和峰形,表明各种不同种类食用油的主要成分基本上是相同的。但不同种类食用油的特征峰相对强度有一定的差别,说明其不饱和脂肪酸、饱和脂肪酸和酯类等成分含量是不同的。

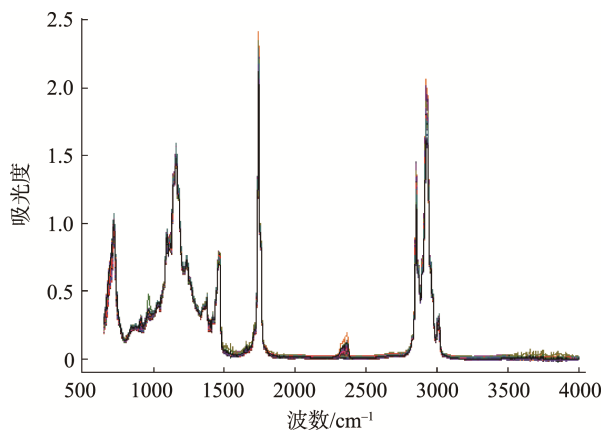


图 1 99 个食用油样本的 ATR-FTIR 图
Fig.1 ATR-FTIR spectra of 99 edible oil samples

从图 1 可以看出,主要的红外吸收峰分布在 $700\sim 1800\text{ cm}^{-1}$ 和 $2800\sim 3000\text{ cm}^{-1}$ 。由于样品池内没有充氮气净化,在波数范围 $2250\sim 2400\text{ cm}^{-1}$ 内是空气中的二氧化碳吸收峰。在光谱范围 $2000\sim 2250$ 、 $2400\sim 2750$ 和 $3100\sim 4000\text{ cm}^{-1}$ 内没有产生吸收,因此在数据分析过程中将这些波段去掉。

随机选取 99 个食用油样本其中 67 个样本作为校正集,剩下的 32 个作为预测集。然后使用化学计量学进行数据分析,包括 PCA、PLS、KNN 和 DD-SIMCA 算法,并比较各方法建模效果,确定最优识别模型。

2.2 基于 DD-SIMCA 的一类判别分析

DD-SIMCA 属于单类分类器,该分类器用于将一个特定的目标类别对象从所有其他对象类别中区分出来。将 2/3 的植物油样本数据集随机划分为校正集,剩下的样本数据作为预测集。为了防止过拟合,模型中最佳 PC 数分别采用杠杆率校正和留一法交叉验证确定。

将随机选取的校正集和预测集样本按类别分为两类,即山茶油和其他油。将各类植物油的红外光谱数据分别 DD-SIMCA 建模,设参数为:校正集数据采用中心化预处理,PC 数 3,接受区域类型选择卡方分布,并使用经典的估计方法($\alpha=0.01$; $\gamma=0.01$)。

在验证阶段,使用新数据(校正集)对模型进行评估。为了测试模型对目标类或者含有外部样本的新数据集的接受和拒绝能力,分别用预测集中非目标类、目标类和所有类数据进行验证,结果见图 2a,绿色的点代表山茶油样本,红色的点代表其他类别油样本,模型对目标类样本和非目标类样本的能完成区分开。图 2b 显示所有的校正集数据都

在公差范围内(以垂直线表示), 表明基于 DD-SIMCA 模型的校正集不包含异常值。基于 DD-SIMCA 建立的鉴定模型能将目标类(山茶油)和其他类别样本分开, 校正集和预测集的正确率均达到 100%。

2.3 基于 PLS-DA、PCA-LDA 和 KNN 的植物油分类判别

将植物油样本数据集分别按 7:3 随机分为校正集和预测集。校正集用于训练和优化模型, 预测集用来预测模型的泛化能力(即预测性能)评价。利用 MATLAB R2018a 软件, 通过 PLS-DA、PCA-LDA 和 KNN 算法, 将植物油种类与光谱信息分别进行建模分析。对 99 个植物油样本的光谱数据开展全光谱(650~4000 cm^{-1})聚类分析, 去除 2000~2250、2400~2750 和 3100~4000 cm^{-1} 范围内光谱数据, 以尽可能消除首位噪声, 避免因数据量大而造成冗余信息。

5 种食用油的特征光谱数据经聚类分析, 获得 PC 选择最佳 PC 数为 9。其中 PC1 贡献 69.82%, PC2 贡献率 17.20%, 9 个 PC 累计贡献率达 99.05%, 利用第 1、2 PC 得

分作散点图(图 3a)。作为一种常用的有监督的降维、判别分析方法, PLS-DA 能同时将响应矩阵和浓度矩阵进行分解提取主因子具有很强的提取信息的能力。本研究采用偏最小二乘法对样本的 8 个变量组成数据矩阵进行处理(见图 3b)。校正集数据经过 PCA-LDA 降维后训练得到 KNN 分类模型(K 值为 4), 最终使用预测集数据对得到的分类模型进行测试, 随机分组及分类识别重复进行 50 次后对分类识别正确率取平均值。采用 PLS-DA、PCA-LDA 和 KNN 算法建立的校正模型对样本的分类识别准确率如表 2~4 所示。

结果显示, 相比 KNN 模型, 利用 PLS-DA 和 PCA-LDA 模型在 5 种植物油的分类中校正集和预测集中各样本的预测值与实际值很接近, 只有一个花生样本在 PCA-LDA 模型中被错判到大豆油组中, 其余种类植物油的校正集和预测集样本的识别率和预报正确率均为 100.0%。因此, 可以认为采用 PLS-DA 和 PCA-LDA 模型所选的分析变量比较合理, 该法用于这 5 种油的类别鉴定结果准确。

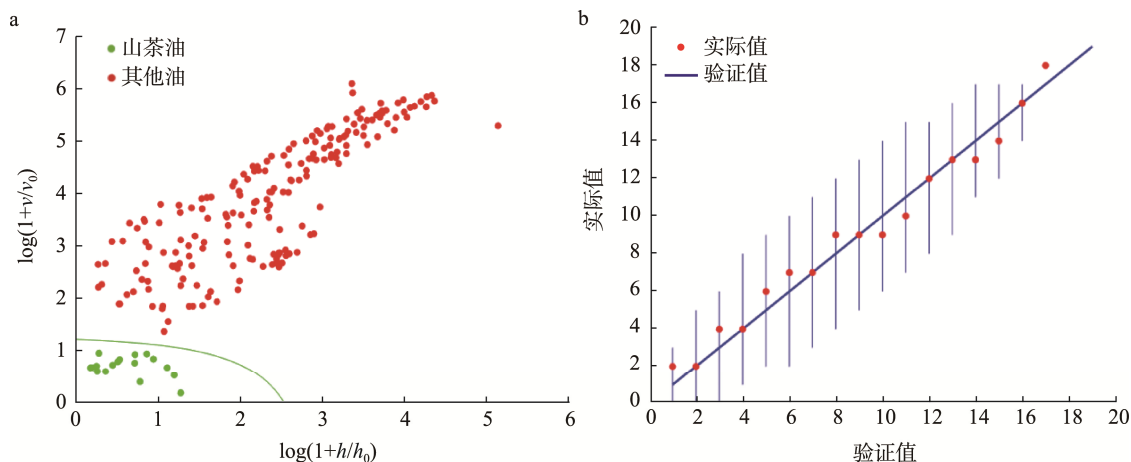


图 2 模型对预测集的接受度图(a)和校正集异常值验证图(b)

Fig.2 Acceptance chart of the model for the prediction set (a) and verification chart for outliers in the correction set (b)

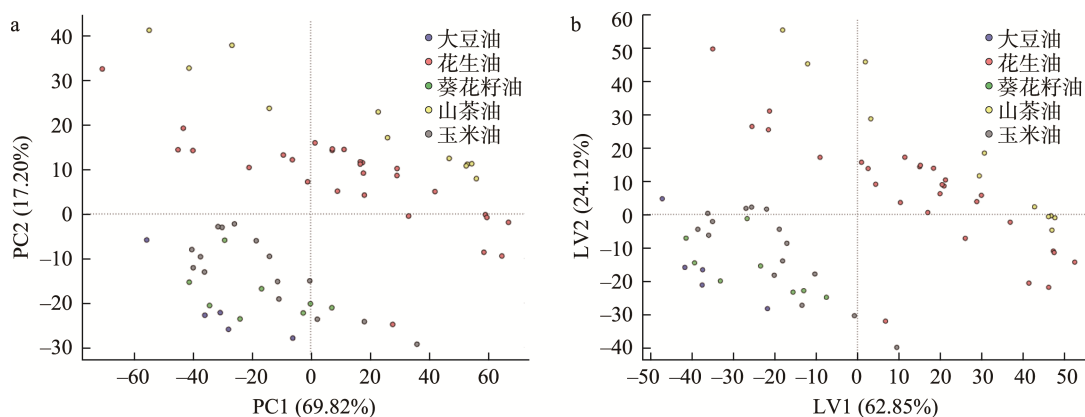


图 3 5 种植物油样本的 PCA-LDA 图(a)和 PLS-DA 图(b)

Fig.3 PCA-LDA (a) and PLS-DA (b) of 5 kinds of plant oil samples

表 2 KNN 模型对样本的分类识别准确率
Table 2 Classification recognition accuracy of samples by KNN model

混淆矩阵		真实值					准确率/%	
		大豆油	花生油	葵花籽油	山茶油	玉米油		
校正集	预测值	大豆油	3	0	0	0	0	100.0
		花生油	0	28	0	0	1	96.6
		葵花籽油	1	0	7	0	2	70.0
		山茶油	0	0	0	11	0	100.0
		玉米油	1	0	1	0	12	85.7
	灵敏度/%	60.0	100.0	87.5	100.0	80.0		
准确度/%		91.0						
预测集	预测值	大豆油	0	0	0	0	0	0.0
		花生油	2	12	1	1	1	70.6
		葵花籽油	1	0	2	0	1	50.0
		山茶油	0	0	0	5	0	100.0
		玉米油	0	0	1	0	5	83.3
	灵敏度/%	0	100.0	50.0	83.3	71.4		
准确度/%		75.0						

表 3 PLS-DA 模型对样本的分类识别准确率
Table 3 Classification recognition accuracy of samples by PLS-DA model

混淆矩阵		真实值					准确率/%	
		大豆油	花生油	葵花籽油	山茶油	玉米油		
校正集	预测值	大豆油	5	0	0	0	0	100.0
		花生油	0	28	0	0	0	100.0
		葵花籽油	0	0	8	0	0	100.0
		山茶油	0	0	0	11	0	100.0
		玉米油	0	0	0	0	15	100.0
	灵敏度/%	100.0	100.0	100.0	100.0	100.0		
准确度/%		100.0						
预测集	预测值	大豆油	3	0	0	0	0	100.0
		花生油	0	12	0	0	0	100.0
		葵花籽油	0	0	4	0	0	100.0
		山茶油	0	0	0	6	0	100.0
		玉米油	0	0	0	0	7	100.0
	灵敏度/%	100.0	100.0	100.0	100.0	100.0		
准确度/%		100.0						

表 4 PCA-LDA 模型对样本的分类识别准确率
Table 4 Classification recognition accuracy of samples by PCA-LDA model

混淆矩阵		真实值					准确率/%	
		大豆油	花生油	葵花籽油	山茶油	玉米油		
校正集	预测值	大豆油	5	1	0	0	0	83.3
		花生油	0	27	0	0	0	100.0
		葵花籽油	0	0	8	0	0	100.0
		山茶油	0	0	0	11	0	100.0
		玉米油	0	0	0	0	15	100.0
	灵敏度/%	100.0	96.4	100.0	100.0	100.0		
准确度/%		98.5						
预测集	预测值	大豆油	3	0	0	0	0	100.0
		花生油	0	12	0	0	0	100.0
		葵花籽油	0	0	4	0	0	100.0
		山茶油	0	0	0	6	0	100.0
		玉米油	0	0	0	0	7	100.0
	灵敏度/%	100.0	100.0	100.0	100.0	100.0		
准确度/%		100.0						

2.4 基于 PLS 的山茶油掺假定量分析

选取 2 个山茶油样本分别掺入不同比例(5%、10%、15%、20%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%)的大豆油、葵花籽油、玉米油以及 1:1 的玉米大豆油。由于花生油可以凭颜色和气味辨别出来, 因此山茶油掺花生油的最高比例为 60%; 而棕榈油由于熔点较低, 一般常温下为固态, 掺杂量较高时也可以辨别出来, 因此山茶油掺棕榈油的最高比例为 50%。植物油样本如表 5 和表 6 所示。

样本集里含 8 个纯的植物油样本和 196 个掺杂样本, 分别采集它们的 FTIR 信息, 将植物油样本数据集分别按 3:1 随机分为校正集和预测集。校正集用于训练和优化模型, 预测集用来预测模型的泛化能力(即预测性能)评价。利用 MATLAB R2018a 软件, 通过 PLS, 将不同掺杂量的植物油与光谱信息分别进行建模分析。山茶油掺假的真实值和预测值相关图见图 4 所示, 利用 PLS 对掺杂了不同种类、不同比例的山茶油样本进行定量分析, 具有良好的表现。

表 5 纯植物油样本种类和编号

Table 5 Types and numbers of pure vegetable oil samples

编号	1	2	3	4	5	6	7	8
品种(品牌)	山茶油 1 (厂家提供)	山茶油 2 (厂家提供)	玉米油 (金龙鱼)	大豆油 (元宝)	花生油 (鲁花)	玉米油+大豆油 (1:1)	葵花籽油 (福临门)	棕榈油 (天益佳)

表 6 山茶油掺杂样本

Table 6 Camellia oil adulterated samples

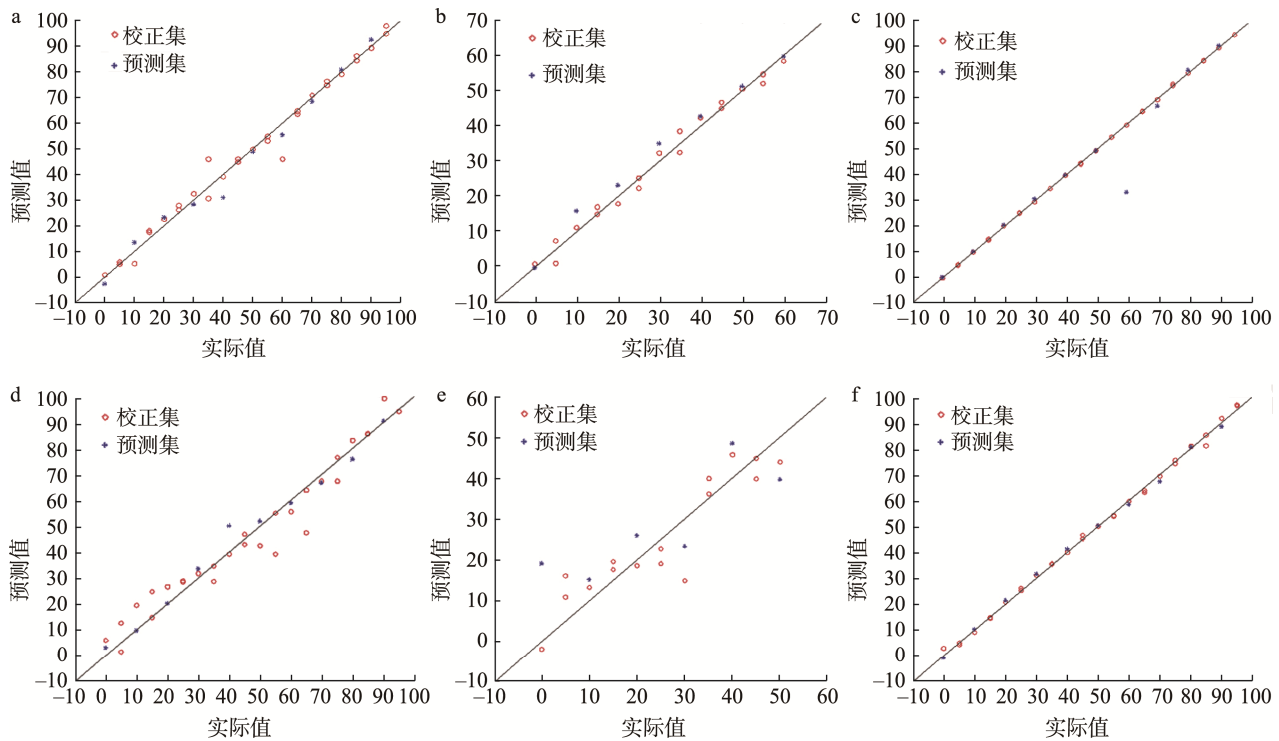
茶油掺玉米油	茶油掺大豆油	茶油掺花生油	茶油掺玉米大豆油	茶油掺葵花籽油	茶油掺棕榈油
1-3-5%	1-4-5%	1-5-5%	1-6-5%	1-7-5%	1-8-5%
1-3-10%	1-4-10%	1-5-10%	1-6-10%	1-7-10%	1-8-10%
1-3-15%	1-4-15%	1-5-15%	1-6-15%	1-7-15%	1-8-15%
1-3-20%	1-4-20%	1-5-20%	1-6-20%	1-7-20%	1-8-20%
1-3-25%	1-4-25%	1-5-25%	1-6-25%	1-7-25%	1-8-25%
1-3-30%	1-4-30%	1-5-30%	1-6-30%	1-7-30%	1-8-30%
1-3-35%	1-4-35%	1-5-35%	1-6-35%	1-7-35%	1-8-35%
1-3-40%	1-4-40%	1-5-40%	1-6-40%	1-7-40%	1-8-40%
1-3-45%	1-4-45%	1-5-45%	1-6-45%	1-7-45%	1-8-45%
1-3-50%	1-4-50%	1-5-50%	1-6-50%	1-7-50%	1-8-50%
1-3-55%	1-4-55%	1-5-55%	1-6-55%	1-7-55%	2-8-5%
1-3-60%	1-4-60%	1-5-60%	1-6-60%	1-7-60%	2-8-10%
1-3-65%	1-4-65%	2-5-5%	1-6-65%	1-7-65%	2-8-15%
1-3-70%	1-4-70%	2-5-10%	1-6-70%	1-7-70%	2-8-20%
1-3-75%	1-4-75%	2-5-15%	1-6-75%	1-7-75%	2-8-25%
1-3-80%	1-4-80%	2-5-20%	1-6-80%	1-7-80%	2-8-30%
1-3-85%	1-4-85%	2-5-25%	1-6-85%	1-7-85%	2-8-35%
1-3-90%	1-4-90%	2-5-30%	1-6-90%	1-7-90%	2-8-40%
1-3-95%	1-4-95%	2-5-35%	1-6-95%	1-7-95%	2-8-45%
2-3-5%	2-4-5%	2-5-40%	2-6-5%	2-7-5%	2-8-50%
2-3-10%	2-4-10%	2-5-45%	2-6-10%	2-7-10%	
2-3-15%	2-4-15%	2-5-50%	2-6-15%	2-7-15%	
2-3-20%	2-4-20%	2-5-55%	2-6-20%	2-7-20%	
2-3-25%	2-4-25%	2-5-60%	2-6-25%	2-7-25%	
2-3-30%	2-4-30%		2-6-30%	2-7-30%	
2-3-35%	2-4-35%		2-6-35%	2-7-35%	
2-3-40%	2-4-40%		2-6-40%	2-7-40%	
2-3-45%	2-4-45%		2-6-45%	2-7-45%	
2-3-50%	2-4-50%		2-6-50%	2-7-50%	
2-3-55%	2-4-55%		2-6-55%	2-7-55%	
2-3-60%	2-4-60%		2-6-60%	2-7-60%	
2-3-65%	2-4-65%		2-6-65%	2-7-65%	
2-3-70%	2-4-70%		2-6-70%	2-7-70%	
2-3-75%	2-4-75%		2-6-75%	2-7-75%	
2-3-80%	2-4-80%		2-6-80%	2-7-80%	
2-3-85%	2-4-85%		2-6-85%	2-7-85%	
2-3-90%	2-4-90%		2-6-90%	2-7-90%	
2-3-95%	2-4-95%		2-6-95%	2-7-95%	

注: 1-3-5%代表 1 号山茶油掺入 5% 3 号玉米油, 以下类推。

一般模型的好坏主要由校正均方根误差(root mean square error of cross validation, RMSECV)、预测均方根误差(root mean square error of prediction, RMSEP)以及决定系数(coefficient of determination, R^2)等综合评判。一般情况下, RMSECV 和 RMSEP 数值越低, R^2 数值越接近 1, 模型越好。

由表 7 可知, 山茶油掺杂其他种类植物油的校正 RMSECV 在 0.2413~6.5127 之间, 验证 RMSEP 在

1.4246~10.3908, R^2 在 0.8413~0.9999 之间, 尤其是掺杂葵花籽油和掺杂玉米大豆油的模型较稳健。说明利用 PLS 对掺杂不同种类、不同比例的山茶油样本进行定量分析, 结果准确可靠。因此本研究所构建的 PLS 模型能够满足山茶油中掺杂其他种类植物油的定量判别要求, 最低检测限可达 5%, 为山茶油品质鉴定与掺杂量预测提供了一种快速、简单、准确的方法, 在实际应用中具有一定的参考价值。



注: a. 山茶油掺大豆油; b. 山茶油掺花生油; c. 山茶油掺葵花籽油; d. 山茶油掺玉米油; e. 山茶油掺棕榈油; f. 山茶油掺玉米大豆油。

图 4 PLS 计算山茶油掺假的真实值和预测值相关图

Fig.4 Correlation between the true and predicted values of *Camellia* oil adulteration calculated by PLS

表 7 基于 PLS 的山茶油掺假定量分析结果比较

Table 7 Comparison of quantitative analysis results of *Camellia* oil adulteration based on PLS method

	掺大豆油	掺花生油	掺葵花籽油	掺玉米油	掺棕榈油	掺玉米大豆油
R^2	0.9832	0.9868	0.9999	0.9489	0.8413	0.9980
R	0.9915	0.9934	1.0000	0.9741	0.9172	0.9990
RMSECV	3.7404	2.0953	0.2413	6.5127	6.0984	1.2874
RMSEP	3.8705	3.3625	8.4564	3.9204	10.3908	1.4246

3 结论

本研究采用 ATR-FTIR 结合化学计量学的方法, 建立不同种类植物油及山茶油掺假的快速识别模型。DD-SIMCA 建立的鉴定模型能将山茶油和其他类别植物油完全分开; 基于 5 种植物油样本的 PLS-DA、PCA-LDA 和 KNN 模型判别经分析比较, PLS-DA 和

PCA-LDA 所选取的分析变量比较合理, 可用于 5 种植物油的类别鉴定; ATR-FTIR 结合 PLS 的计量学方法能够准确进行山茶油掺假定量分析; 可用于掺杂大豆油、玉米油、葵花籽油等的定性定量分析, 最低检测限可达 5%。研究表明, ATR-FTIR 结合化学计量学方法可以实现对山茶油掺假鉴别的高效识别, 为规范山茶油产品市场提供技术支持。

参考文献

- [1] TANG C, SHEN Q, SHAO S, *et al.* Quality authentication of *Camellia* (*Camellia oleifera* Abel.) oil based on fluorescence spectroscopy [J]. *Journal of Food Composition and Analysis*, 2024, 136: 106690.
- [2] SHI T, WU GC, JIN QZ, *et al.* Camellia oil adulteration detection using fatty acid ratios and tocopherol compositions with chemometrics [J]. *Food Control*, 2022, 133: 108565–108567.
- [3] 刘燕德, 谢庆华. 山茶油品质检测方法现状研究[J]. *食品工业*, 2016, 37(6): 253–257.
LIU YD, XIE QH. Research on the current status of quality testing methods for *Camellia* oil [J]. *The Food Industry*, 2016, 37(6): 253–257.
- [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.
ZHONG XR. Research on the nutritional value and development of *Camellia* oil [J]. *China Food Industry*, 2021(22): 72–75.
- [5] 殷莺倩, 王少敏, 刘翠玲, 等. 基于近红外光谱技术的食用油酸值和过氧化值定量分析研究[J]. *食品安全质量检测学报*, 2023, 14(5): 68–76.
YIN YQ, WANG SM, LIU CL, *et al.* Quantitative analysis of acid value and peroxide value of edible oil based on near-infrared spectroscopy [J]. *Journal of Food Safety & Quality*, 2023, 14(5): 68–76.
- [6] FLORES G, CASTILLO MLRD, PLANCH G, *et al.* Detection of the adulteration of olive oils by solid phase microextraction and multidimensional gas chromatography [J]. *Food Chemistry*, 2006, 97: 336–342.
- [7] CERCACI L, ESTRADA MTR, LERCKER G. Solid-phase extraction-thin-layer chromatography-gas chromatography method for the detection of hazelnut oil in olive oils by determination of esterified sterols [J]. *Journal of Chromatography A*, 2003, 985: 211–220.
- [8] 卢永欣, 李耀辉, 崔周平. 高效液相色谱法测定食用植物油中的桐油[J]. *预防医学文献信息*, 2002, 2(8): 48–49.
LU YX, LI YH, CUI ZP. Determination of tung oil in edible vegetable oil by high performance liquid chromatography [J]. *Liferatue and Information on Preventine Medicine*, 2002, 2(8): 48–49.
- [9] 于雪莲, 杨培艺, 郭娟, 等. 云南 6 种木本植物油脂肪酸组成分析及调和优化[J]. *中国油脂*, 2024, 49(12): 52–57.
YU XL, YANG PY, GUO J, *et al.* Fatty acid composition analysis and blending optimization of six woody palnt oils in Yunnan Province [J]. *China Oils and Fats*, 2024, 49(12): 52–57.
- [10] 章颖强. 食用植物油快速检测方法研究[D]. 杭州: 浙江大学, 2012.
ZHANG YQ. Research on rapid detection method for edible vegetable oil [D]. Hangzhou: Zhejiang University, 2012.
- [11] 孙玉萍, 杨亚, 黄国栋, 等. 橄榄油市场行情及其他植物油的脂肪酸组成和微量营养成分比较[J]. *中国油脂*, 2024, 49(3): 27–31.
SUN YP, YANG Y, HUANG GD, *et al.* Comparison of olive oil market trends and fatty acid composition and trace nutrients of other vegetable oils [J]. *China Oils and Fats*, 2024, 49(3): 27–31
- [12] RUIZ MV, DOBARGANES MC. Combination of chromatographic techniques for the analysis of complex deodoriser distillates from an edible oil refining process [J]. *Food Chemistry*, 2007, 103: 1502–1507.
- [13] ZHANG YL, CHEN JB, LEI Y, *et al.* Evaluation of different grades of ginseng using Fourier-transform infrared and two-dimensional infrared correlation spectroscopy [J]. *Journal of Molecular Structure*, 2010, 974: 94–102.
- [14] LIU HX, SUN SQ, LV GH, *et al.* Study on *Angelica* and its different extracts by Fourier transform infrared spectroscopy and two-dimensional correlation IR spectroscopy [J]. *Spectrochimica Acta Part A*, 2006, 64: 321–326.
- [15] 刘娅, 赵国华, 陈宗道, 等. 近红外光谱技术在油脂检测中的应用[J]. *粮油加工与食品机械*, 2000, 2: 37–39.
LIU Y, ZHAO GH, CHEN ZD, *et al.* Application of near infrared spectroscopy technology in oil and fat detection [J]. *Cereals and Oils Processing*, 2000, 2: 37–39.
- [16] BAETEN V, MEURENS M, MORALES MT, *et al.* Detection of virgin olive oil adulteration by Fourier transform raman spectroscopy [J]. *Journal of Agricultural and Food Chemistry*, 1996, 44(8): 2225–2230.
- [17] YANG H, IRUDAYARAJ J. Comparison of near-infrared, Fourier transform-infrared, and Fourier transform-Raman methods for determining olive pomace oil adulteration in extra virgin olive oil [J]. *Journal of the American Oil Chemists Society*, 2001, 78(9): 889–895.
- [18] 匡俊豪, 罗宁宁, 郝中骐, 等. 基于近红外拉曼光谱的山茶油掺假检测技术研究[J]. *南昌航空大学学报(自然科学版)*, 2022, 36(3): 104–110.
KUANG JH, LUO NN, HAO ZQ, *et al.* Research on the detection technology of *Camellia* oil adulteration based on near infrared Raman spectroscopy [J]. *Journal of Nanchang University of Aeronautics and Astronautics (Natural Science Edition)*, 2022, 36(3): 104–110.
- [19] HE YX, LIANG QF. Quality analysis of cooking oils by Fourier transform infrared spectroscopy [J]. *Technology & Development of Chemical Industry*, 2011, 40(6): 32–34.
- [20] WU JZ, LIU CL, LI H, *et al.* Application of NIR technology on identifying types and determining main fatty acid content of edible vegetable oil [J]. *Journal of Beijing Technology and Business University (Natural Science Edition)*, 2010, 28(5): 56–59.
- [21] WANG L, HU JH. Advance in discrimination method of edible on adulterated waste cooking oil [J]. *China Oils and Fats*, 2007, 32(9): 75–77.
- [22] BRENDEL R, SCHWOLOW S, GERHARDT N, *et al.* MIR spectroscopy versus MALDI-ToF-MS for authenticity control of honeys from different botanical origins based on soft independent modelling by class analogy (SIMCA)-A clash of techniques [J]. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy*, 2021, 263: 120225.
- [23] AKIN, GONOL KE, SUKRIYE NA, *et al.* Chemometric classification and quantification of cold pressed grape seed oil in blends with refined soybean oils using attenuated total reflectance-mid infrared (ATR-MIR) spectroscopy [J]. *Food Science and Technology*, 2019, 100: 126–137.
- [24] ALI H, MUHAMMAD S, ANSER MR, *et al.* Validation of fluorescence spectroscopy to detect adulteration of edible oil in extra virgin olive oil by applying chemometrics [J]. *Applied Spectroscopy*, 2018, 72(9): 1371–1379.
- [25] KIEFER J, LAMPE AI, NICOLI SF, *et al.* Identification of passion fruit oil adulteration by chemometric analysis of FTIR spectra [J]. *Molecules*, 2019, 24(18): 3219.
- [26] DUCA D, MANCINI M, ROSSINI G, *et al.* Soft Independent Modelling of class analogy applied to infrared spectroscopy for rapid discrimination between hardwood and softwood [J]. *Energy*, 2016, 117: 251–258.
- [27] MAGDAS DA, GUYON F, BERGHIAN GC, *et al.* Challenges and a step forward in honey classification based on Raman spectroscopy [J]. *Food Control*, 2021, 123: 107769.
- [28] GERÔNIMO DM, OLIVEIRA SCD, SOARES FLF, *et al.* Determination of main raw material source in bar soaps using mid-infrared spectroscopy combined with classification tools [J]. *Microchemical Journal*, 2021, 164: 106029.
- [29] PINTO FG, MAHMUD I, RUBIO VY, *et al.* Data-driven soft independent modeling of class analogy in paper spray ionization mass spectrometry-based metabolomics for rapid detection of prostate cancer [J]. *Analytical Chemistry*, 2022, 94(4): 1925–1931.
- [30] SALIH MAM. 牛奶掺假的近红外光谱分析技术研究[D]. 北京: 中国农业科学院, 2018.
SALIH MAM. Research on near infrared spectroscopy analysis technology for milk adulteration [D]. Beijing: Chinese Academy of Agricultural Sciences, 2018.

(责任编辑: 安香玉 韩晓红)