



李艳坤,张伟,刘彦伶. 数据融合策略在食用油真实性鉴别中的研究与应用进展[J]. 轻工学报,2024,39(5):50-59.
LI Y K,ZHANG W,LIU Y L. Research and application progress of data fusion strategy in authenticity identification of edible oil[J]. Journal of Light Industry,2024,39(5):50-59. DOI:10.12187/2024.05.006

数据融合策略在食用油真实性鉴别中的研究与应用进展

李艳坤^{1,2},张伟¹,刘彦伶¹

- 华北电力大学 环境科学与工程系,河北 保定 071003;
- 华北电力大学 河北省燃煤电站烟气多污染物协同控制重点实验室,河北 保定 071003

摘要: 对基于光谱、质谱、色谱等检测技术的数据融合策略及其在食用油真实性鉴别中的研究及应用现状进行综述,指出:目前,广泛应用于食用油真实性鉴别的检测技术包括光谱、色谱、质谱、电子传感器等。然而,单一检测技术往往只关注某一特定的数据或指标,当食用油所含成分较复杂时,无法充分消除叠加效应、基线漂移、噪声等问题。数据融合策略分为数据层融合、特征层融合和决策层融合三类,结合化学计量学方法可以综合不同检测技术获取的数据,提取更丰富的数据特征信息,从而提高食用油真实性的鉴别效果。不同的新型检测技术之间,或将其与传统光谱、质谱、色谱等检测技术之间进行数据融合,可以快速、准确地实现食用油掺伪鉴别、品种分类和产地溯源,未来可就改进现有分析方法、结合深度学习算法开发新型融合算法、引入云计算提高食用油鉴别实时性等方面进行深入研究,以推动数据融合策略在食用油真实性鉴别领域的发展与创新。

关键词: 数据融合;食用油;真实性鉴别;检测技术;化学计量学

中图分类号: O657.3;TS227 **文献标识码:** A **文章编号:** 2096-1553(2024)05-0050-10

0 引言

食用油富含不饱和脂肪酸等营养成分,是日常烹饪中必不可少的食材之一^[1]。当前,食用油市场中掺假、以次充好等现象屡禁不止,食用油产地、品质造假问题也日渐引起关注^[2-3]。食用油真实性鉴别的目的是确定其成分、纯度和品质特征,以确保消费者能够选购到高品质、健康和安全的食用油^[4]。

目前,用于食用油真实性检测的技术主要包括色谱、质谱、光谱、电子传感器等^[5],将这些检测技

术与模式(分类)识别方法结合使用,能够实现食用油真实性的快速准确鉴别。根据有无训练样本,模式识别方法可以分为有监督识别和无监督识别。常用的有监督识别方法包括偏最小二乘判别分析(Partial Least Squares Discriminant Analysis, PLS-DA)、线性判别分析(Linear Discriminant Analysis, LDA)、支持向量机、神经网络等;无监督识别方法包括主成分分析(Principal Component Analysis, PCA)、距离判别分析、聚类分析等。李艳坤等^[6]基于中红外光谱结合不同模式识别方法对不同质量浓度的掺

收稿日期:2023-11-15;修回日期:2024-01-18;出版日期:2024-10-15

基金项目:中央高校基本科研业务费项目(2017MS135);国创计划资助项目(S202310079127)

作者简介:李艳坤(1977—),女,河北省邯郸市人,华北电力大学副教授,博士,主要研究方向为化学计量学、环境分析化学。
E-mail:liyankun_ncepu@foxmail.com

伪食用植物油进行研究,发现有监督识别方法的鉴别精确度均较高。王九玲等^[7]采集不同掺假浓度橄榄油样品的拉曼光谱和荧光光谱,并结合反向传播神经网络模型进行定量分析,发现该模型具有较高的预测效率和精确度。

通常情况下,单一检测技术难以全面地分析、评价物质组成或含量,适当融合多个检测技术的数据,可以克服单一检测技术的不足,获得更准确的检测结果。数据融合也称信息融合,是将多个不同来源或类别的量测数据进行融合、关联的过程^[8]。数据融合策略可以对收集到的数据进行加工处理,削弱甚至消除冗余、干扰信息,并就处理后的数据建立预测分析模型,实现对研究对象的定性、定量分析,提高分析模型的预测性能^[9-10]。本文基于 2018—2023 年期间的文献调研,综述了数据融合策略在食用油真实性鉴别中的应用,并对该领域未来的发展前景进行展望,以期数据融合策略在食用油真实性鉴别、掺伪分析、产地溯源等领域中的研究与应用提供参考。

1 食用油真实性鉴别技术及其应用

1.1 光谱技术

近红外光谱(Near Infrared Spectroscopy, NIR)和中红外光谱(Mid-Infrared Spectroscopy, MIR)因检测速度快、操作简便等特点被广泛应用于食用油真实性鉴别中。Q. W. Du 等^[11]通过 NIR 结合 PLS-DA 的方法研究了山茶油中掺杂玉米油、菜籽油和葵花籽油的情况,发现该方法区分纯山茶油和掺假山茶油的准确率为 100%。Z. Yuan 等^[12]提出了一种利用 NIR 对亚麻籽油多次掺假进行目标检测的方法,该方法通过正交偏最小二乘判别分析(Orthogonal Partial Least Squares Discriminant Analysis, OPLS-DA)选择重要变量,并采用高斯径向基核函数消除共线性,再结合单类偏最小二乘法构建检测模型,结果发现该模型能够有效检测单掺杂物、双掺杂物和多掺杂物,准确率为 95.8%。R. Jamwal 等^[13]采用衰减全反射-傅里叶变换红外光谱(Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy, ATR-FTIR)分别结合 PCA 和 LDA 检测芥子油中甾

甾粟油的掺假情况,发现结合 PCA 能够成功区分芥子油和掺假芥子油,而结合 LDA 还能够检测到芥子油中甾粟油的掺假占比。

拉曼光谱(Raman Spectroscopy, RS)被用于测定食用油中游离脂肪酸的含量和不饱和度^[14]。H. F. Zhao 等^[15]采集了 10 种不同品牌食用油的拉曼光谱数据集,并使用卷积神经网络、多项式逻辑回归、线性回归等 9 种监督机器学习算法进行分析,再与 PCA 模型进行比较,发现监督机器学习算法模型的分类速度更快且精确度更高。R. C. Castro 等^[16]利用 NIR 和 RS 结合多变量曲线分辨率-交替最小二乘模型对纯花生油中掺杂的玉米油和植物油进行鉴别,发现该模型能够准确提取复杂混合物中玉米油和植物油的光谱图。

诱导荧光光谱(Induced Fluorescence Spectroscopy, IFS)法是采用具有一定波长和强度的光波来诱导特定的荧光信号进行定性或定量分析的方法,已应用于特级初榨橄榄油(Extra Virgin Olive Oil, EVOO)掺假、茶叶分类等领域。T. Zhang 等^[17]利用发光二极管(Light Emitting Diodes, LED)诱导 IFS 测定 EVOO 的荧光光谱,并通过 PCA 结合多元线性回归和 PLS 建模,对 EVOO 中大豆油和花生油的掺假比例进行分析,发现该方法能有效识别并准确检测 EVOO 中二者的掺杂情况。Z. Y. Huan 等^[18]采用同步荧光光谱结合 LDA 模型对掺假花生油进行分类,发现该模型的识别率达 95.0%,可以有效区分不同类型的掺假花生油。

核磁共振(Nuclear Magnetic Resonance, NMR)因具有快速、无损分析等特点被广泛用于食品质量、检测食品掺假、测量食品性能等指标的评估。S. H. Wang 等^[19]利用低场核磁共振结合 OPLS-DA 模型鉴定掺杂不同比例大豆油或玉米油的橄榄油,发现该模型能够准确区分纯橄榄油和二元掺假橄榄油。

1.2 色谱、质谱技术

色谱技术可以分离食用油中的脂肪酸及其氧化产物等成分,并对其进行定性或定量分析;质谱(Mass Spectrum, MS)技术可以鉴别食用油中的甘油三酯、微量元素等成分。因此,将色谱与 MS 联用可以实现食用油中复杂成分和微量组分的分析与

鉴定。

X. R. Wang 等^[20]利用气相色谱(Gas Chromatography, GC)从不同地区种植的青海亚麻籽中获取了植物油的主要脂肪酸谱,并结合判别分析(Discriminant Analysis, DA)和PCA鉴定了亚麻籽油与其他品种植物油的掺杂程度,发现DA和PCA均能实现掺杂程度的鉴别,且DA还能鉴别掺入的植物油品种。A. R. Mansur 等^[21]通过GC结合PCA分析掺假芝麻油的脂肪酸、三酰基甘油和挥发性有机化合物含量,发现通过分析脂肪酸含量区分纯芝麻油和掺假芝麻油的精确度最高。侯颖焯等^[22]对元素分析-稳定同位素质谱结合PCA和OPLS-DA鉴定橄榄油掺假情况进行对比分析,发现这两种方法构建的鉴别模型均准确可靠。G. Ozcan-Sinir^[23]通过离子流动管-质谱结合化学计量学测定EVOO中挥发性成分的含量,成功区分了EVOO与掺假产品。

1.3 其他检测技术

目前,电子鼻(E-nose)、电子舌(E-tongue)、计算机视觉系统(Computer Vision System, CVS)、离子迁移谱(Ion Mobility Spectrometry, IMS)、DNA检测等技术也可以对食用油中脂肪酸、过氧化物、硫化物等成分进行鉴别。其中,E-nose技术通过模拟生物嗅觉,利用多个传感器对复杂食用油进行挥发性成分分析,然后采用模式识别技术分析收集的数据以确定食用油中的掺假类型和掺假比例。R. Calvini 等^[24]利用E-nose和CVS的组合系统结合聚类分析鉴定EVOO是否掺假,发现该组合系统可以准确、快速和无损地鉴定EVOO的掺假情况。E-tongue技术是一个多传感器系统,可以结合化学计量学方法获取更准确的分析结果^[25]。

IMS技术具有较高的选择性和稳定性,但其获得的多维数据较复杂,需要与合适的化学计量学方法相结合进行定量或定性分析。O. Clemente 等^[26]通过对比分析顶空(Headspace, HS)-GC-IMS与NIR的测试结果,发现HS-GC-IMS对三元混合组分和二元混合组分的分类准确率分别高于85%和90%,优于NIR的测试结果。

DNA检测技术通过识别油类样本的基因组序列来分析食用油品质,具有稳定性、可靠性和特异性

均较高的优势。S. C. Sheu 等^[27]通过利用环介导等温扩增技术和聚合酶链式反应检测技术扩增样品的特征DNA来鉴定橄榄油的掺假情况,发现两种技术都能够鉴别橄榄油样品是否掺假,且前者的鉴别效率和特异性更高。

然而,单一检测技术只关注某一特定的数据或指标,可能无法提供较全面和准确的分析结果。当食用油样品所含成分较复杂时,单独使用色谱或质谱技术可能会出现漏检或误判部分成分的问题,从而降低检测结果的精确度;单一光谱技术无法充分解决样品中存在的叠加效应、基线漂移、噪声等问题;单一传感器技术只能针对特定的物理或化学参数进行测量,分析范围存在较大的局限性。而数据融合策略可以实现不同检测技术之间的优势互补与特征信息的提取,通过结合化学计量学方法将互补的数据信息进行融合,可获得更全面和准确的预测结果。

2 数据融合策略的特点与方法

根据数据融合结构不同,可将数据融合策略分为数据层融合、特征层融合和决策层融合三类^[28],其流程如图1所示。

2.1 数据层融合

数据层融合(低级融合)作为最基本的融合方法,是将多个研究对象的初始数据直接进行融合的方法^[29]。该方法不但将各个来源的数据信息全部保留,更加全面地收集能够反映样本信息的数据,充分体现数据的完整性和真实性,而且操作相对简单,对技术的要求也较低。但串联如此庞大的数据信息可能会出现更多的噪音、冗余等干扰信息,进而对后续所建立模型的精确度产生影响;同时,数据信息量的增加也会加重计算负担,增加处理数据信息所消耗的时间,从而降低信息处理的时效性。J. Moros^[30]等

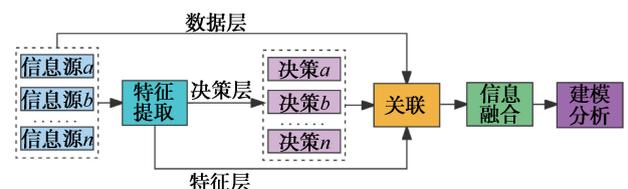


图1 数据融合策略流程图

Fig. 1 Data fusion strategy scheme

提出了4种数据融合方法,即串联融合(Equal Rights Fusion, ERF)、累加融合(Coaddition Fusion, CF)、外和融合(Outer Sum Fusion, OSF)和外积融合(Outer Product Fusion, OPF),其中串联融合的应用范围最为广泛。表1为4种数据层融合方法的特点与示例。

2.2 特征层融合

特征层融合(中级融合)是对不同来源的数据信息进行特征提取,再将提取的具有不同分类或回归贡献度的特征信息进行融合处理的方法。该方法是对提取的多个初始数据源的特征信息进行处理,数据信息的数量和处理时间相对数据层融合有所减少;此外,还可以剔除数据中的无关变量,降低建模的复杂程度,同时提高光谱的信噪比^[33]。在光谱融合建模过程中,数据的多重共线性问题与自变量高度相关,这可能导致模型的回归系数不稳定、预测能力差等问题,一般选用变量(波长)选择和特征提取两种方法,通过光谱降维的形式来解决这一问题。表2为7种常用的变量选择和特征提取方法及其特点^[34]。

2.3 决策层融合

决策层融合(高级融合)是在数据层融合和特征层融合的基础上,对不同来源数据信息分别进行独立处理和建模,然后将所有模型的预测结果进行关联整合,以获得最优预测结果^[42]。决策层信息融合具有较强的灵活性、容错性和抗干扰能力,但决

策层融合的处理成本较高,且损失了很多与数据相关的细节信息,可能会导致后续建模精确度的下降。在决策层融合中,针对体系多元分类和回归预测的不同性质和特点,发展了多种融合策略^[43]。常用的决策层融合策略有基于概率量化的贝叶斯法则、模糊集理论、核融合、多重投票等,其方法与特点见表3。

综上所述,数据层融合较简单,使用单一模型便可提取不同变量块之间的相关性;特征层融合通过特征提取可以显著降低数据的维度,且每个变量块可以独立对数据进行处理,有助于消除噪声等冗余信息,但通常需要结合精确的数据预处理方法;决策层融合的优点是每个信号矩阵都是单独处理,避免了其他融合级别下可能出现的模型整体性能恶化的情况,但未考虑不同来源数据源之间响应的相关性,可能会造成数据丢失。因此,不同数据融合策略各有优劣,需根据实际需求和数据特点进行选择。

3 数据融合策略在食用油真实性鉴别中的应用

通过数据融合可以实现不同检测技术间的优势互补,但由于数据融合策略的多样性,在实际应用中,还需结合合适的数据预处理和化学计量学方法以提高预测的准确性和效率。表4为近5年数据融合策略在食用油真实性鉴别中的典型应用。

表1 4种数据层融合方法的特点与示例

Table 1 Characteristics and examples of four data layer fusion methods

融合方式	特点	示例(以 \mathbf{R} 和 \mathbf{L} 向量融合为例)
串联融合 ^[31]	简单串联,可保留数据原始形貌信息和相关性,但融合后的数据未必能找到适用于不同来源数据的最佳模型	$\begin{bmatrix} r_1 & r_2 & r_3 & \cdots & r_n \end{bmatrix} + \begin{bmatrix} l_1 & l_2 & l_3 & \cdots & l_n \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & \cdots & r_n & l_1 & l_2 & l_3 & \cdots & l_n \end{bmatrix}$
累加融合 ^[31]	可以加大信息丰富、信噪比高的波段的权重,增强融合后数据的非线性,但可能会导致一些数据丢失	$\begin{bmatrix} r_1 & r_2 & r_3 & \cdots & r_n \end{bmatrix} + \begin{bmatrix} l_1 & l_2 & l_3 & \cdots & l_n \end{bmatrix} = \begin{bmatrix} r_1 + l_1 & r_2 + l_2 & r_3 + l_3 & \cdots & r_n + l_n \end{bmatrix}$
外和融合 ^[32]	对各仪器变量的复加性产生贡献,从而突出显示样本的信息特征,且融合后为二阶数据	$\begin{bmatrix} r_1 \\ r_2 \\ \cdots \\ r_3 \end{bmatrix} \oplus \begin{bmatrix} l_1 & l_2 & \cdots & l_n \end{bmatrix} = \begin{bmatrix} r_1 + l_1 & r_2 + l_1 & \cdots & r_n + l_1 \\ r_1 + l_2 & r_2 + l_2 & \cdots & r_n + l_2 \\ \cdots & \cdots & \cdots & \cdots \\ r_1 + l_n & r_2 + l_n & \cdots & r_n + l_n \end{bmatrix}$
外积融合 ^[32]	增强数据信号中样本特征相关信息的共同演化,融合后为二阶数据,但建模计算所消耗的时间有所增加	$\begin{bmatrix} r_1 \\ r_2 \\ \cdots \\ r_3 \end{bmatrix} \otimes \begin{bmatrix} l_1 & l_2 & \cdots & l_n \end{bmatrix} = \begin{bmatrix} r_1 l_1 & r_2 l_1 & \cdots & r_n l_1 \\ r_1 l_2 & r_2 l_2 & \cdots & r_n l_2 \\ \cdots & \cdots & \cdots & \cdots \\ r_1 l_n & r_2 l_n & \cdots & r_n l_n \end{bmatrix}$

3.1 不同光谱数据融合

MIR 和 NIR 技术可以根据官能团在吸收光谱中的出峰位置、吸收强度等信息对食用油中脂肪酸等特征组分进行定量分析^[55-56],两种光谱数据融合后可以更加全面、准确地鉴别食用油的真实性。Y. Li 等^[44]采用 NIR 和 MIR 融合的方法对橄榄油和掺入菜籽油的橄榄油进行定量分析,发现相较于单一光谱预测模型,低级融合包含了 NIR 和 MIR 的全部信息,提高了模型的预测精确度;中级融合减少了变量的数量,但并未提高模型的预测精确度;高级融合大幅提高了模型的预测精确度,可作为定量检测橄榄油掺假的可靠工具。张婧等^[48]采集了纯芝麻油和掺入玉米油的芝麻油的同步和异步二维近红外相关谱和中红外相关谱数据,结合 PLS-DA 构建了基于多维主成分分析的中级数据融合模型,判别准确率达 100%。

在 RS 技术中,由烯烃分子振动而产生的拉曼散射可用于食用油的品质分析,但 RS 技术对掺假食用油的定性、定量识别精确度不高,因此需要与其他光谱数据融合后使用^[57]。H. Liu 等^[51]利用 FTIR

和 RS 预测常见食用油在不同热氧化条件下的过氧化值和脂肪酸含量,对光谱全波长预处理后,结合 PLS 方法构建基于 SPA、UVE 的中级数据融合模型,发现 SPA-PLS 模型的精确度更高。邱薇纶等^[52]借助 ATR-FTIR 和表面增强拉曼光谱获取常见的植物油样本光谱数据,结合 Bayes 判别分析、多层感知器神经网络分别构建了低级串联融合和基于 PCA、广义最小平方、最大似然、主轴因式分解特征提取的中级融合模型,发现这些融合模型均能实现植物油的准确分类,且基于 PCA 的 Bayes 判别融合模型的识别准确率最高。

O. Hu 等^[49]将 FS 与 NIR 融合,并结合单类偏最小二乘算法建立分类模型分析山茶油的掺假问题,发现融合光谱的灵敏度为 0.954,特异性为 0.91,均优于单一光谱模型的预测结果。A. Fort 等^[58]通过低级融合和高级融合 FS 和 FT-RS 的数据来鉴别 EVOO 的产地,发现数据融合可以互补两组光谱信息,从而提高模型的灵敏度。H. Y. Gu 等^[50]采用低场核磁共振、FS 和基于探针的 FS 数据

表 2 7 种常用的变量选择和特征提取方法及其特点^[34]

Table 2 Seven variable selection and feature extraction methods and their characteristics

方法	首次提出者及时间	特点
PCA (Principal Component Analysis)	K. Pearson, 1901 ^[35]	可以有效降低数据维度,但对数据线性相关性的假设较强,可能导致特征信息丢失
UVE (Uninformative Variable Elimination)	V. Centner, 1996 ^[36]	可以有效消除无信息变量的影响,但随机噪声变量可能使结果不稳定,进而导致计算效率较低
iPLS (Interval PLS)	L. Nørgaard, 2000 ^[37]	选择预测均方根误差最低的波段区间,但可能会丢失一些具有样本信息的波段
CARS (Competitive Adaptive Reweighted Sampling)	H. D. Li, 2009 ^[38]	提取的特征变量较少,且光谱稳定性较低
SPA (Successive Projection Algorithm)	M. C. U. Araújo, 2001 ^[39]	可最小化多重共线性变量对模型的影响,但计算量大,不适合小样本数据
SA (Simulated Annealing)	N. Metropolis, 1953 ^[40]	保留了变量组合的优点,但待优化的变量组合较多,容易导致局部最优
ULDA (Uncorrelated Linear Discriminant Analysis)	Z. Jin, 2001 ^[41]	将变量特征提取与模型分类融为一体,保留了特征变量与样本类别之间的相关性

表 3 4 种常用决策层融后方法及其特点

Table 3 Four common decision methods and their characteristics

方法	原理	特点
贝叶斯法则	利用概率论整合不同来源的数据信息,更新先验概率,使决策更加动态和灵活	在数据量大或模型复杂时计算难度大,且先验概率对模型预测精确度影响较大
模糊集理论	通过找出隶属度最接近 1(相似性最强)的子集,使决策更接近实际情况	能够有效融合不同来源的数据信息,但计算复杂性高,且需要选择适当的隶属函数
核融合	通过在高维空间中构建核函数融合数据,提高分类或回归的准确性	能够处理复杂的非线性关系和高维数据,但模型参数和核函数的选择对模型预测精确度影响较大
多重投票	将多个模型的预测结果进行投票作为最终决策结果	能够有效降低单一模型的误差,需要合理设计投票机制,过多模型的集成会导致计算复杂程度增大

表4 近5年数据融合策略在食用油真实性鉴别中的典型应用

Table 4 Typical application of data fusion in authenticity identification of edible oil in five years

鉴别类型	食用油品种	检测技术	化学计量学方法	融合层级	文献
掺伪鉴别	橄榄油	NIR, MIR	SPA, PLS, BLR	低,中,高	[44]
			二进制线性回归(Binary Linear Regression, BLR)		
		FT-TR, HS-GC-IMS	PCA-LDA, PLS-DA	低	[45]
		E-nose, E-tongue, CVS	k 近邻算法(k-Nearest Neighbours, k-NN)	中	[46]
	芝麻油	LC-MS, GC-IMS, FGC-Enose	PLS-DA	低,中	[47]
	山茶油	NIR, ATR-IR	PCA	低	[48]
			标准正态变量变换(Standard Normal Variate, SNV), 单类偏最小二乘法(One-Class Partial Least Squares, OCPLS)	低	[49]
棕榈油	NMR, FS 等	遗传算法-支持向量回归机(Genetic Algorithm-Support Vector Regression, GA-SVR), 平行因子分析(Parallel Factor Analysis, PARAFAC)等	低,中,高	[50]	
品种分类		FTIR, RS	PLS, SPA, UVE	中	[51]
		ATR-FTIR, RS	PCA, 贝叶斯判别分析(Bayes Discriminant Analysis, Bayes 判别)等	低,中	[52]
产地溯源	棕榈油	配置不同检测器的 HPLC	PCA, iPLS, PLS-DA 等	中,高	[53]
	橄榄油	正相、反相 HPLC, GC 等	PLS-DA, PCA, 类独立建模分析(Soft In Dependent Modelling of Class Analogies, SIMCA)	中,高	[54]

进行低、中、高级数据融合后对煎炸油(棕榈油)的品质进行评估,发现单一光谱模型与融合数据建立的模型均能准确评估煎炸油的品质,但低、中、高级数据融合模型预测结果的精确度均高于单一光谱模型。

3.2 光谱与质谱、色谱数据融合

不同种类的光谱数据还可以与质谱或色谱数据融合,获得更全面和多元化的数据信息。高冰等^[59]结合 PLS-DA 构建了基于 GC 和 NIR 数据串联融合的食用油品质判别模型,发现融合模型的灵敏度和特异性均为 1,分类误差为 0。A. Maléchaux 等^[60]将 3 种 EVOO 样品的 GC 和 MIR 数据串联融合,并对加权、非加权多元偏最小二乘判别分析和 PLS-DA 模型的预测性能进行对比分析,发现相较于单一模型,融合模型的预测结果更佳、预测精确度更高,且非加权多元偏最小二乘判别分析模型的灵敏度更高。S. Schwolow 等^[45]利用 FT-IR 和 HS-GC-IMS

分析了 3 个不同地区的 53 份 EVOO 样品,将所获数据进行串联融合后,对比分析了 PCA-LDA 和 PLS-DA 模型分类效率和性能,发现结合 PLS-DA 的融合模型具有更好的预测精确度和稳定性。N. Jurado-campos 等^[61]将紫外检测-毛细管电泳与 GC-IMS 进行高级数据融合后,对橄榄油样本进行分类分析,证明了数据融合技术的潜力和可靠性。

3.3 质谱与色谱数据融合

目前,质谱与色谱数据融合在食用油真实性鉴别领域的研究相对较少,但已有研究^[53-54]表明,色谱与质谱数据具有协同性,将二者融合用于食用油真实性鉴别具有可行性。K. A. Obisesan 等^[53]使用配备不同检测器的 HPLC 鉴别棕榈油的产地,并结合 PLS-DA 分别建立基于 PCA、iPLS 特征信息提取的中级数据融合模型和基于模糊集理论的高级数据融合模型,发现与使用单一检测技术相比,3 种数据融合模型均能准确鉴定棕榈油的产地,且基于 iPLS

特征信息提取的中级数据融合模型分类效果最好。D. N. Vera 等^[54]运用反相和正相液相色谱与高温气相色谱获得橄榄油样品的色谱指纹图谱,并结合 PLS-DA、软独立类比建模法分别构建了基于 PCA 特征信息提取的中级融合模型和基于模糊集理论的高级数据融合模型,发现两种数据融合模型均能实现橄榄油产地的准确鉴定,且基于 PCA 特征提取的中级数据融合模型分类效果最佳,这也证明了不同色谱数据协同融合的可行性。

3.4 其他检测技术数据融合

近年来,针对 E-nose、E-tongue、CVS 等不同传感器检测技术之间进行数据融合的研究,以及传感器与光谱、色谱之间进行数据融合的研究逐年增多。M. R. Zarezadeh 等^[62]分别采用超声系统和 E-nose 对 EVOO 进行分类,发现超声系统比 E-nose 的预测精确度高;将两个传感器的数据进行融合后,发现融合模型比单一传感器模型的分类精确度更高,准确率最高达 95.51%。L. Blandon-Naranjo 等^[63]将 E-tongue 中每个工作电极获得的橄榄油电化学数据结合人工神经网络进行低级串联融合,并利用离散小波变换特征提取的数据进行中级融合,发现中级融合模型的相对误差为 8%~10%,分析准确率高于低级融合模型(相对误差为 10%~18%)。S. Buratti 等^[46]采用 E-nose、E-tongue 和 CVS 获得橄榄油的理化数据,结合 K 近邻算法建立中级数据融合模型,根据新鲜度对橄榄油样品进行分类,准确率达 94%。A. Tata 等^[47]分别采用串联两种 LC-MS 的低级数据融合策略,串联 GC-IMS 和 FGC-Enose 的低级数据融合策略,串联多模式 LC-MS、FGC-Enose 和 GC-IMS 的低级数据融合策略及基于 PLS-DA 特征提取的中级数据融合策略,对 EVOO 掺假问题进行研究,发现 4 种数据融合策略均可鉴别 EVOO 掺假情况,且串联 GC-IMS 和 FGC-Enose 的低级数据融合模型分类性能最优。X. Zhou 等^[64]将热重分析和 GC-MS 数据结合,基于 PCA 的中级数据融合研究橄榄油和大豆油掺假,发现数据融合能成功鉴别橄榄油和大豆油掺假,且在橄榄油样品掺假中,预测准确率达 100%。综上,不同新型检测技术之间或新型检测技术与传统光谱、质谱、色谱等技术之间的

数据融合均可以快速、准确地鉴别食用油的真实性。

4 总结与展望

本文综述了光谱、色谱、质谱、E-nose、CVS 等检测技术及基于上述技术的数据融合策略在食用油真实性鉴别,包括食用油掺伪鉴别、品种分类和产地溯源领域中的应用现状和研究进展,发现单一检测技术在分析成分复杂的食用油样品时,存在漏检或误判,无法充分消除叠加效应、基线漂移、噪声等问题。数据融合策略结合化学计量学方法可以实现不同检测技术之间的优势互补,并提取更丰富的数据特征信息。根据应用场景分析需求、数据信息特征等不同的实际情况,选用适宜的数据融合策略结合化学计量学方法,可以高效地获得精确的分析结果。然而,由于不同层次融合方法的多样性,在选择融合策略时,需要采用适宜的数据预处理方法和化学计量学方法进行分析。因此,在不同层次的数据融合策略与方法上仍需深入展开以下研究:一是除了发展新型融合算法,还可以将深度学习与先进的数据融合算法相结合以提高预测模型的稳定性和准确性,更好地解决建模复杂的问题;二是针对特定应用场景,研究如何选择适宜的融合策略和化学计量学方法,以实现不同检测技术间的最佳协同效应;三是探索使用云计算,寻求更高效、更可靠的数据管理和分析解决方案,实现对食用油的即时监测和鉴别,对于食用油真实性鉴别技术的发展与创新具有重要意义。

参考文献:

- [1] ZHANG T, LIU Y Y, DAI Z P, et al. Quantitative detection of extra virgin olive oil adulteration, as opposed to peanut and soybean oil, employing LED-induced fluorescence spectroscopy[J]. *Sensors*, 2022, 22(3): 1227.
- [2] 孔令琦, 宋佳琪, 陈林林, 等. 食用油掺伪鉴别技术及模型建立的研究进展[J]. *食品安全质量检测学报*, 2022, 13(19): 6132-6139.
- [3] SALAH W A, NOFAL M. Review of some adulteration detection techniques of edible oils[J]. *Journal of the Science of Food and Agriculture*, 2021, 101(3): 811-819.
- [4] 程慧, 刘顺, 关洪宣. 两种滴定法测定食用油中过氧化值和酸价的不确定度评价[J]. *食品与机械*, 2022, 38

- (1):73-77,99.
- [5] 邓焯文,陈喆,付家顺,等.数据融合策略在食品产地溯源中的应用进展[J].分析化学,2023,51(1):11-21.
- [6] 李艳坤,许东情.基于中红外光谱模型对食用植物油掺伪的判别[J].河北大学学报(自然科学版),2022,42(6):605-610.
- [7] 王九玲,罗文,李文凯.反向传播神经网络算法结合拉曼荧光光谱法定量检测特级初榨橄榄油掺假[J].食品安全质量检测学报,2023,14(22):126-133.
- [8] 戴嘉伟,王海朋,陈瀑,等.多光谱数据融合分析技术的研究和应用进展[J].分析化学,2022,50(6):839-849.
- [9] 杨巧玲,邓晓军,孙晓东,等.光谱数据融合技术在食品检测中的应用研究进展[J].食品工业科技,2020,41(18):324-329.
- [10] 孙婷婷,钟瑾璟,刘剑波,等.茶油掺伪定性鉴别模型的对比分析[J].中国粮油学报,2022,37(11):245-252.
- [11] DU Q W, ZHU M T, SHI T, et al. Adulteration detection of corn oil, rapeseed oil and sunflower oil in camellia oil by *in situ* diffuse reflectance near-infrared spectroscopy and chemometrics[J]. Food Control, 2021, 121: 107577.
- [12] YUAN Z, ZHANG L X, WANG D, et al. Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis[J]. LWT-Food Science and Technology, 2020, 125: 109247.
- [13] JAMWAL R, AMIT, KUMARI S, et al. Attenuated total Reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy coupled with chemometrics for rapid detection of argemone oil adulteration in mustard oil[J]. LWT-Food Science and Technology, 2020, 120: 108945.
- [14] ARSLAN F N, AKIN G, KARUK ELMAS N, et al. Rapid detection of authenticity and adulteration of cold pressed black cumin seed oil: A comparative study of ATR-FTIR spectroscopy and synchronous fluorescence with multivariate data analysis[J]. Food Control, 2019, 98: 323-332.
- [15] ZHAO H F, ZHAN Y L, XU Z, et al. The application of machine-learning and Raman spectroscopy for the rapid detection of edible oils type and adulteration[J]. Food Chemistry, 2022, 373: 131471.
- [16] CASTRO R C, RIBEIRO D S M, SANTOS J L M, et al. Comparison of near infrared spectroscopy and Raman spectroscopy for the identification and quantification through MCR-ALS and PLS of peanut oil adulterants[J]. Talanta, 2021, 230: 122373.
- [17] ZHANG T, LIU Y Y, DAI Z P, et al. Quantitative detection of extra virgin olive oil adulteration, as opposed to peanut and soybean oil, employing LED-induced fluorescence spectroscopy[J]. Sensors, 2022, 22(3): 1227.
- [18] HUYAN Z Y, DING S X, LIU X L, et al. Authentication and adulteration detection of peanut oils of three flavor types using synchronous fluorescence spectroscopy[J]. Analytical Methods, 2018, 10(26): 3207-3214.
- [19] WANG S H, LAI G Y, LIN J Z, et al. Rapid detection of adulteration in extra virgin olive oil by low-field nuclear magnetic resonance combined with pattern recognition[J]. Food Analytical Methods, 2021, 14(7): 1322-1335.
- [20] WANG X R, HAN Y Z, LI Y X, et al. Detection of Qinghai-Tibet Plateau flaxseed oil adulteration based on fatty acid profiles and chemometrics[J]. Food Control, 2021, 130: 108332.
- [21] MANSUR A R, JEONG H R, LEE B H, et al. Comparative evaluation of triacylglycerols, fatty acids, and volatile organic compounds as markers for authenticating sesame oil[J]. International Journal of Food Properties, 2018, 21(1): 2509-2516.
- [22] 侯颖焯,王志元,谢建军,等.元素分析-稳定同位素质谱法结合化学计量学鉴别橄榄油掺假[J].中国油脂,2023,48(6):73-78.
- [23] OZCAN-SINIR G. Detection of adulteration in extra virgin olive oil by selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics[J]. Food Control, 2020, 118: 107433.
- [24] CALVINI R, PIGANI L. Toward the development of combined artificial sensing systems for food quality evaluation: A review on the application of data fusion of electronic noses, electronic tongues and electronic eyes[J]. Sensors, 2022, 22(2): 577.
- [25] MIRHOSEINI-MOGHADDAM S M, YAMAGHANI M R, BAKHSHIPOUR A. Application of electronic nose and eye systems for detection of adulteration in olive oil based on chemometrics and optimization approaches[J]. Journal of Universal Computer Science, 2023, 29(4): 300-325.
- [26] CLEMENTE O, ROCÍO R, L. D G, et al. Comparing the potential of IR-spectroscopic techniques to gas chromatography coupled to ion mobility spectrometry for classifying virgin olive oil categories[J]. Food Chemistry(X), 2023, 19: 100738.
- [27] SHEU S C, WANG Y J, HUANG P C, et al. Authentication of olive oil in commercial products using specific, sensitive, and rapid loop-mediated isothermal amplification[J]. Journal of Food Science and Technology, 2023, 60(6): 1834-1840.
- [28] AZCARATE S M, RÍOS-REINA R, AMIGO J M, et al. Data handling in data fusion: Methodologies and applications[J]. TrAC Trends in Analytical Chemistry, 2021, 143: 116355.
- [29] SUN Z B, DAVIS J, GAO W. Estimating error covariance and correlation region in UV irradiance data fusion by combining TOMS-OMI and UVMRP ground observations[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(1): 355-370.
- [30] MOROS J, JAVIER LASERNA J. Unveiling the identity of

- distant targets through advanced Raman-laser-induced breakdown spectroscopy data fusion strategies [J]. *Talanta*, 2015, 134: 627-639.
- [31] 徐伟杰, 武中臣, 朱香平, 等. 基于光谱融合的火星表面相关矿物分类方法研究[J]. *光谱学与光谱分析*, 2018, 38(6): 1926.
- [32] 王清亚, 李福生, 江晓宇, 等. 基于XRF和Vis-NIR光谱数据融合的土壤镉含量定量分析法[J]. *分析测试学报*, 2020, 39(11): 1327-1333.
- [33] CHANDRA S, CHAPMAN J, POWER A, et al. Origin and regionality of wines: The role of molecular spectroscopy [J]. *Food Analytical Methods*, 2017, 10(12): 3947-3955.
- [34] 李艳坤, 董汝南, 张进, 等. 光谱数据解析中的变量筛选方法[J]. *光谱学与光谱分析*, 2021, 41(11): 3331-3338.
- [35] PEARSON K. On lines and planes of closest fit to systems of points in space [J]. *Philosophical Magazine*, 1901, 2(11): 559-572.
- [36] CENTNER V, MASSART D L, DE NOORD O E, et al. Elimination of uninformative variables for multivariate calibration [J]. *Analytical Chemistry*, 1996, 68(21): 3851-3858.
- [37] NØRGAARD L, SAUDLAND A, WAGNER J, et al. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy [J]. *Applied Spectroscopy*, 2000, 54(3): 413-419.
- [38] LI H D, LIANG Y Z, XU Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. *Analytica Chimica Acta*, 2009, 648(1): 77-84.
- [39] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 57(2): 65-73.
- [40] METROPOLIS N, ROSENBLUTH A W, ROSENBLUTH M N, et al. Equation of state calculations by fast computing machines [J]. *The Journal of Chemical Physics*, 1953, 21(6): 1087-1092.
- [41] JIN Z, YANG J Y, HU Z S, et al. Face recognition based on the uncorrelated discriminant transformation [J]. *Pattern Recognition*, 2001, 34(7): 1405-1416.
- [42] GENG X J, LIANG Y, JIAO L M. Multi-frame decision fusion based on evidential association rule mining for target identification [J]. *Applied Soft Computing*, 2020, 94: 106460.
- [43] DANKOWSKA A. Data fusion of fluorescence and UV spectroscopies improves the detection of cocoa butter adulteration [J]. *European Journal of Lipid Science and Technology*, 2017, 119(8): 1156-1172.
- [44] LI Y, XIONG Y M, MIN S G. Data fusion strategy in quantitative analysis of spectroscopy relevant to olive oil adulteration [J]. *Vibrational Spectroscopy*, 2019, 101: 20-27.
- [45] SCHWOLOW S, GERHARDT N, ROHN S, et al. Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys-is it worth to go the extra Mile? [J]. *Analytical and Bioanalytical Chemistry*, 2019, 411(23): 6005-6019.
- [46] BURATTI S, MALEGORI C, BENEDETTI S, et al. E-nose, e-tongue and e-eye for edible olive oil characterization and shelf life assessment: A powerful data fusion approach [J]. *Talanta*, 2018, 182: 131-141.
- [47] TATA A, MASSARO A, DAMIANI T, et al. Detection of soft-refined oils in extra virgin olive oil using data fusion approaches for LC-MS, GC-IMS and FGC-Enose techniques: The winning synergy of GC-IMS and FGC-Enose [J]. *Food Control*, 2022, 133: 108645.
- [48] 张婧, 单慧勇, 杨仁杰, 等. 基于近-中红外相关谱融合判定掺伪芝麻油 [J]. *光子学报*, 2019, 48(6): 62-68.
- [49] HU O, CHEN J, GAO P F, et al. Fusion of near-infrared and fluorescence spectroscopy for untargeted fraud detection of Chinese tea seed oil using chemometric methods [J]. *Journal of the Science of Food and Agriculture*, 2019, 99(5): 2285-2291.
- [50] GU H Y, HUANG X Y, SUN Y H, et al. Intelligent evaluation of total polar compounds (TPC) content of frying oil based on fluorescence spectroscopy and low-field NMR [J]. *Food Chemistry*, 2021, 342: 128242.
- [51] LIU H, CHEN Y, SHI C, et al. FT-IR and Raman spectroscopy data fusion with chemometrics for simultaneous determination of chemical quality indices of edible oils during thermal oxidation [J]. *LWT-Food Science and Technology*, 2020, 119: 108906.
- [52] 邱薇纶, 周燕舞, 石孟良. 基于数据融合策略植物油光谱模式的识别 [J]. *中国油脂*, 2023, 48(7): 62-66, 89.
- [53] OBISESAN K A, JIMÉNEZ-CARVELO A M, CUADROS-RODRIGUEZ L, et al. HPLC-UV and HPLC-CAD chromatographic data fusion for the authentication of the geographical origin of palm oil [J]. *Talanta*, 2017, 170: 413-418.
- [54] VERA D N, JIMÉNEZ-CARVELO A M, CUADROS-RODRÍGUEZ L, et al. Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics [J]. *Talanta*, 2019, 203: 194-202.
- [55] SRINATH K, KIRANMAYEE A H, BHANOT S, et al. Detection of palm oil adulteration in sunflower oil using ATR-MIR spectroscopy coupled with chemometric algorithms [J]. *MAPAN*, 2022, 37(3): 483-493.
- [56] UNCU O, OZEN B. A comparative study of mid-infrared, UV-visible and fluorescence spectroscopy in combination

- with chemometrics for the detection of adulteration of fresh olive oils with old olive oils [J]. *Food Control*, 2019, 105: 209–218.
- [57] ARSLAN F N, AKIN G, KARUK ELMAS N, et al. Rapid detection of authenticity and adulteration of cold pressed black cumin seed oil: A comparative study of ATR-FTIR spectroscopy and synchronous fluorescence with multivariate data analysis [J]. *Food Control*, 2019, 98: 323–332.
- [58] FORT A, RUISÁNCHEZ I, CALLAO M P. Chemometric strategies for authenticating extra virgin olive oils from two geographically adjacent Catalan protected designations of origin [J]. *Microchemical Journal*, 2021, 169: 106611.
- [59] 高冰, 吴鹏飞, 许晓栋, 等. 基于色谱和光谱数据融合的不同植物源食用油判别方法与模型 [J]. *分析测试学报*, 2020, 39(11): 1398–1403.
- [60] MALÉCHAUX A, LAROUCSI-MEZGHANI S, LE DRÉAU Y, et al. Multiblock chemometrics for the discrimination of three extra virgin olive oil varieties [J]. *Food Chemistry*, 2020, 309: 125588.
- [61] JURADO-CAMPOS N, ARROYO-MANZANARES N, VIÑAS P, et al. Quality authentication of virgin olive oils using orthogonal techniques and chemometrics based on individual and high-level data fusion information [J]. *Talanta*, 2020, 219: 121260.
- [62] ZAREZADEH M R, ABOONAJMI M, GHASEMI-VARNAMKHAHI M. The effect of data fusion on improving the accuracy of olive oil quality measurement [J]. *Food Chemistry (X)*, 2023, 18: 100622.
- [63] BLANDON-NARANJO L, ALANIZ R D, ZON M A, et al. Development of a voltammetric electronic tongue for the simultaneous determination of synthetic antioxidants in edible olive oils [J]. *Talanta*, 2023, 261: 124123.
- [64] ZHOU X, LI X Q, ZHAO B, et al. Discriminant analysis of vegetable oils by thermogravimetric-gas chromatography/mass spectrometry combined with data fusion and chemometrics without sample pretreatment [J]. *LWT-Food Science and Technology*, 2022, 161: 113403.

Research and application progress of data fusion strategy in authenticity identification of edible oil

LI Yankun^{1,2}, ZHANG Wei¹, LIU Yanling¹

1. *Department of Environmental Science and Engineering, North China Electric Power University, Baoding 071003, China;*

2. *Hebei Key Laboratory of Power Plant Flue Gas Multi-Pollutants Control, North China Electric Power University, Baoding 071003, China*

Abstract: An overview of data fusion strategies based on spectroscopy, mass spectrometry, chromatography and other detection technologies and their current research and application in authenticity identification of edible oils was presented, pointing out that: at present, detection technologies widely used for authenticity identification of edible oils including spectroscopy, chromatography, mass spectrometry and electronic sensors. However, a single detection technique often focused only on a specific data or index, which could not fully eliminate the superposition effect, baseline drift and noise when the ingredients contained in edible oils were more complex. Data fusion strategies were categorized into three types: data layer fusion, feature layer fusion and decision layer fusion. Combined with chemometrics methods, the data obtained by different detection technologies could be integrated to obtain and extract richer data feature information, thus improving the authenticity identification of edible oils. Data fusion between various novel detection technologies, or between new and traditional spectroscopy, mass spectrometry, chromatography and other detection technologies, which could quickly and accurately achieved the identification of adulteration of edible oils, variety classification and origin traceability. In the future, in-depth research could be carried out on the improvement of the existing analytical methods, the development of new fusion algorithms combined with deep learning algorithms, and the introduction of cloud computing to improve real-time edible oil identification, so as to promote the development of data fusion strategy in the field of edible oil authenticity identification.

Key words: data fusion; edible oil; authenticity identification; testing technology; chemometrics

[责任编辑: 王晓波]