

DOI: 10.16210/j.cnki.1007-7561.2024.06.020

张凯萍, 杨青波. 基于近红外拉曼光谱结合 LSTM-CNN 模型提高牡丹籽油掺假浓度的预测精度[J]. 粮油食品科技, 2024, 32(6): 170-179.

ZHANG K P, YANG Q B. Improving the prediction accuracy of peony seed oil adulteration concentration based on near-infrared Raman spectroscopy combined with LSTM-CNN model[J]. Science and Technology of Cereals, Oils and Foods, 2024, 32(6): 170-179.

基于近红外拉曼光谱结合 LSTM-CNN 模型提高牡丹籽油掺假浓度的预测精度

张凯萍, 杨青波

(许昌电气职业学院 信息工程学院, 河南 许昌 461002)

摘要: 为建立一种适用于市场在线监测玉米油掺假牡丹籽油的方法, 研究了便携式 785 拉曼光谱技术应用于一线市场质检的潜力, 制备了 525 个玉米油掺假牡丹籽油的油品, 评估了每种混合油品便携式近红外拉曼光谱数据集的稳定性和 21 种混合油品拉曼光谱谱峰的差异性, 分析了拉曼光谱谱峰的振动归属。随机选择 16 种混合油品的拉曼光谱数据集用于训练定量模型, 将剩余的 5 种混合油品的拉曼光谱数据集用于测试定量模型。牡丹籽油与玉米油的近红外拉曼光谱存在较大差异, 根据特征峰位的归属判定, 拉曼光谱谱峰差异很好地反映了植物油间的脂肪酸含量差异。提出一种将长短期记忆网络结合卷积神经网络 (LSTM-CNN) 的模型应用于玉米油掺假牡丹籽油的量化预测, 对比了偏最小二乘回归算法的预测效果。结果表明便携式 785 拉曼光谱的稳定性良好, 具备实现对牡丹籽油掺假量的在线、快速量化的性能, 且结合 LSTM-CNN 模型可以实现对玉米油掺假牡丹籽油含量的预测, 其模型评价参数决定系数 (R^2) 为 0.990 8, 均方根误差 (RMSE) 为 0.029 9。便携式近红外拉曼光谱结合 LSTM-CNN 模型是一种快速、高效、可行的鉴别玉米油掺假牡丹籽油的有效方法。

关键词: 牡丹籽油; 玉米油; 长短期记忆网络; 卷积神经网络; 近红外拉曼光谱; 便携式拉曼光谱仪

中图分类号: TS201.6; O433 文献标识码: A 文章编号: 1007-7561(2024)06-0170-10

网络首发时间: 2024-11-07 11:46:47

网络首发地址: <https://link.cnki.net/urlid/11.3863.TS.20241107.0903.004>

Improving the Prediction Accuracy of Peony Seed Oil Adulteration Concentration Based on Near-infrared Raman Spectroscopy Combined with LSTM-CNN Model

ZHANG Kai-ping, YANG Qing-bo

(School of Information Engineering, Xuchang Electrical Vocational College, Xuchang, Henan 461002, China)

Abstract: To establish a method applicable for online monitoring of adulteration of corn oil with peony

收稿日期: 2024-04-25

基金项目: 河南省重点研发与推广专项 (232400410357); 河南省高等教育学会高等教育研究项目 (2021SXHLX190)

Supported by: Key Research and Development Project of Henan Province (No. 232400410357); Higher Education Research Project of Henan Institute (No. 2021SXHLX190)

第一作者: 张凯萍, 女, 1983 年出生, 硕士, 副教授, 研究方向为机器学习与大数据, E-mail: kp_z@foxmail.com

seed oil in the market, this paper studied the potential of portable 785 Raman spectroscopy technology for quality inspection in the first-line market and prepared 525 corn oil adulterated peony seed oil products. The stability of the portable near-infrared Raman spectroscopy data set of each kind of mixed oil and the differences of Raman spectrum peaks of 21 types of mixed oil were evaluated. The vibration attribution of Raman spectrum peaks was analyzed. The Raman spectral data sets of 16 kinds of oil mixtures were randomly selected to train the quantitative model, and the Raman spectral data sets of the remaining 5 kinds of oil mixtures were used to test the quantitative model. Significant difference existed between the near-infrared Raman spectra of peony seed oil and corn oil. According to the attribution of characteristic peaks, the difference in Raman spectra peaks could well reflect the difference in fatty acid content between vegetable oils. A model combining long short-term memory network and convolutional neural networks (LSTM-CNN) was proposed for quantitative prediction of corn oil adulterated with peony seed oil, and its prediction performance was compared with the partial least squares regression algorithm. The results show that the portable 785 nm Raman spectrometer had good stability and could achieve online and rapid quantification of peony seed oil adulteration levels. Combined with the LSTM-CNN model, it could predict the content of corn oil adulterated peony seed oil. The model evaluation parameter's coefficient of determination (R^2) was 0.990 8, and the root mean square error (RMSE) was 0.029 9. Portable near-infrared Raman spectroscopy combined with the LSTM-CNN model is a fast, efficient, and feasible method to identify the adulteration of corn oil with peony seed oil.

Key words: peony seed oil; corn oil; long short-term memory; convolutional neural networks; near infrared raman spectroscopy; portable raman spectrometer

牡丹籽油是由牡丹籽仁提取的本本坚果植物油, 其含有丰富的维生素和矿物质, 且饱和脂肪酸的含量在 90%以上, 其中 α -亚麻酸 (α -Linolenic acid, ALA) 含量高达 40%以上, 这表明牡丹籽油不仅具有很高的营养价值, 而且具有一定的保健功效^[1]。牡丹籽油是一种新兴油品, 其行业标准和国家标准都晚于其他普通植物油, 且大众对牡丹籽油的认知度普遍较低, 不法商贩借此趁机制假贩假。因此, 牡丹籽油的快速检测技术成为推动市场健康发展的“助推器”。目前国内外开展牡丹籽油掺假鉴别的研究工作较少, 牡丹籽油作为一种本土的优质油品, 加快牡丹籽油鉴别检测技术的基础研究具有现实意义。

激光光谱技术是一种极具潜力的鉴别植物油掺假的技术手段, 因其无损、高效、无样本前处理的优势, 被广泛应用于高值植物油的真伪鉴别和掺假定量的研究^[2-4]。拉曼光谱是激光光谱的一个典型代表, 在食品检测^[5]、生物医学^[6]、材料科学^[7]及环境科学^[8]中具有广泛的应用。针对不同的应用场景和需求, 拉曼光谱技术可结合显微镜形

成精密的台式仪器^[9], 也可结合集成度较高的便携探头形成便携式仪器^[10]。

在食用油拉曼光谱检测方面的研究工作中, 便携式 785 nm 拉曼光谱技术已成功应用于食用油品质鉴别和定量分析的领域。其中, 针对高品质植物油的鉴别研究问题, 黄秀丽等^[11]采用 785 nm 便携式拉曼光谱仪开展了灵芝孢子油掺假的鉴别研究工作, 利用特征峰强比值和峰面积比值参数鉴别了灵芝孢子油与菜籽油、大豆油、玉米油、葵花油及花生油, 发现该方法可作为一种初步鉴别灵芝孢子油的筛查技术。张凤娟等^[12]利用 785 nm 便携式拉曼光谱仪开展了牛油果油鉴别的研究工作, 比较了牛油果油、棕榈油、玉米油、芥花籽油、核桃油及亚麻籽油的拉曼光谱特征峰差异, 发现了牛油果油中疑似类胡萝卜素特征拉曼谱峰的存在, 并依次开展了疑似掺假的研究工作。针对进口散装橄榄油品质判定的问题, 马金鸽等^[13]开展了 785 nm 便携式拉曼光谱技术结合传统化学计量学方法鉴别进口散装橄榄油品质现场鉴别的研究, 发现了便携式近红外拉曼光谱结

合角度度量法的预测效果最优。张凤娟等^[14]针对性开展了便携式拉曼光谱技术结合主成分分析和偏最小二乘回归法实现掺伪量的检测研究,验证了便携式拉曼光谱结合化学计量学方法的应用潜力。

此外,针对检测对象的光谱特性,研究人员还开展了不同实验参数拉曼光谱技术的研究工作。郭宗昱等^[15]采用紫外激光器开展了转基因大豆油的紫外拉曼光谱的研究工作,实现了对转基因油品的快速识别。由于荧光对拉曼光谱信号的干扰问题不可忽略。赵迎等^[16]提出一种基于荧光差分拉曼光谱技术开展食用油复热时长的预测,提高了食用油复热时长定量模型的稳定性和精确度。通过文献查新发现国内外还没有开展基于便携式近红外拉曼光谱技术在线评估牡丹籽油掺假的研究工作。WANG Hongpeng 等^[17]自研了一台共聚焦拉曼光谱仪并探测了不同掺伪浓度的牡丹籽油的拉曼光谱,理论上基于密度泛函理论计算并解译了主要脂肪酸分子的振动光谱和振动模式,实验上分析了偏最小二乘回归法、主成分分析回归法和多元线性回归法在预测牡丹籽油掺假浓度的效果,首次实现了针对牡丹籽油掺假定量分析问题的拉曼光谱理论和实验相结合的综合性研究工作,为牡丹籽油掺假定量分析提供了全面性、可供参考的指导性成果,该研究面向大豆油掺假牡丹籽油的制假行为,缺少其他廉价植物油掺假定量分析的比较,也没有对比人工神经网络的预测效果。总之,国内外开展牡丹籽油掺假预测的研究工作较少,拉曼光谱技术量化牡丹籽油掺假的研究报道则更加欠缺。

本研究采用便携式近红外 785 拉曼光谱仪开展了玉米油掺假牡丹籽油的浓度预测研究,在开展牡丹籽油掺伪的研究工作中,将玉米油作为一种潜在的廉价掺伪植物油,主要考虑因素是玉米油中不饱和脂肪酸含量高达 80%~85%,其不饱和脂肪酸含量接近于牡丹籽油。本文分析了拉曼光谱谱峰的主要成分信息,对比了相同掺伪浓度内拉曼光谱的稳定性和不同掺伪浓度间拉曼光谱的差异性,建立了基于长短期记忆网络结合卷积神经网络(Long short-term memory-convolutional neural networks, LSTM-CNN)算法定量反演牡丹

籽油掺假的模型,本项研究对于牡丹籽油掺假的量化分析具有重要的借鉴价值,为相关部门提供了一种可供现场原位检测的方法。

1 材料与方法

1.1 材料与试剂

实验选用的牡丹籽油产地为山东菏泽,原料为丹凤牡丹籽,其品牌为凤祥牡丹,生产过程中执行的产品标准号为 LS/T3242;玉米油产地为广东广州,原料为非转基因玉米油,其品牌为金龙鱼,生产过程中执行的产品标准号为 Q/BBAH0025S。

1.2 仪器与设备

便携式近红外拉曼光谱仪为长春新产业光电技术有限公司研发的便携式拉曼光谱仪 RamanSys-785(如图 1 所示),设备主要由激光器、光谱仪及拉曼探头组成,其中激光器为单模激光器,波长为 (785 ± 0.3) nm,线宽小于 3.2 cm^{-1} ,最大输出功率为 500 mW,光谱仪探测范围为 785~1 100 nm,光谱分辨率优于 8 cm^{-1} ,信噪比为 1 000:1,积分时间可调范围为 17 ms~10 s。785 拉曼探头焦距为 7.5 mm,激励光光纤芯径为 100 μm ,信号光光纤芯径为 200 μm ,两根光纤的 N.A.均为 0.22。

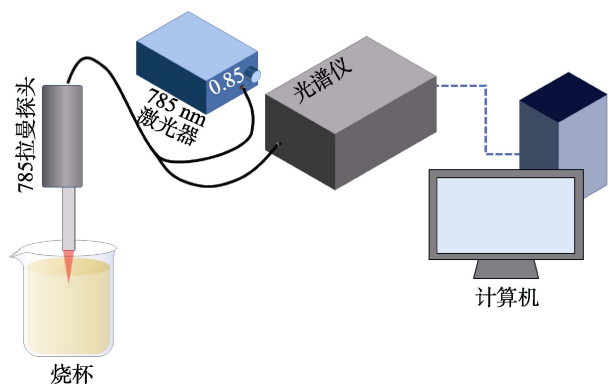


图 1 便携式近红外拉曼光谱仪实验系统示意图

Fig.1 Schematic diagram of portable near infrared Raman spectrometer experimental system

玉米油掺假牡丹籽油的油品均存放在 10 mL 的烧杯中,为避免玻璃容器对油品信号的干扰,拉曼探头从上往下探测油品,其中激光焦点浸入油品液面约 1.5 mm。单条光谱的积分时间为 850 ms,每个油品各随机探测 15 条光谱,每种掺伪浓度可获得 375 条光谱数据。

1.3 实验方法

1.3.1 样本制备方法

考虑到定量模型的鲁棒性并防止模型欠拟合问题的出现,在样品制备环节制备了由低到高尽可能多的掺伪体积比。采用不同体积比的制备方式制备玉米油掺假牡丹籽油的油品,其中牡丹籽油的体积比分别为 0.00、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95 及 1.00。每种混合油品均制备 25 个独立样本,因此共采集了 7 875 条光谱数据。

1.3.2 预处理过程

便携式拉曼光谱仪的稳定性和对外界背景光的抗干扰能力均不及台式系统,因此便携式拉曼光谱仪的原始数据中包含了大量的外界噪声信息和激光能量波动引起的光谱抖动。预处理过程主要包含了噪声去除和光谱稳定性修正两部分,其中采用小波算法去除环境背景光、基体荧光及仪器暗电流等噪声的影响^[17],采用最大值归一化修正激光能量波动引起的光谱抖动。

1.3.3 LSTM-CNN 算法模型

长短期记忆网络 (Long short-term memory, LSTM) 是一种特殊的循环神经网络 (Recurrent neural network, RNN),它是针对传统 RNN 的一种优化结构^[18]。LSTM 具有记忆和遗忘单元,可以对长序列数据保留重要信息并丢弃无意义信息,其擅长处理一维时间序列的分析和预测^[19],而光谱数据也可视为一种特殊的一维“时间序列”数据。LSTM 改善了 RNN 在处理长序列数据的梯

度消失问题,但当光谱数据维度特征较多时,LSTM 也可能无法“记忆”长序列数据的依赖性和相关性。因此,在 LSTM 中引入注意力机制 (Attention mechanism, AM) 保留所有源隐藏状态并将注意力集中在长序列数据的最相关部分^[20]。卷积神经网络 (Convolutional neural network, CNN) 是一种广泛应用于计算机视觉领域的深度学习算法^[21]。CNN 的设计灵感源自生物视觉系统,其主要目标是通过模仿人类视觉处理方式来解决计算机视觉中的问题。相较于传统的机器学习算法,CNN 在处理图像和视频等高维数据时表现出色。CNN 通过多层卷积层和池化层来提取图像的特征。卷积层通过滑动一个小的卷积核在图像上提取局部特征,然后通过多个卷积核来提取不同的特征。池化层则用于降低特征图的维度,减少计算复杂度,并保留主要的特征信息。在训练过程中,CNN 通过反向传播算法来自动调整网络中的权重和偏置,以最小化预测输出与真实标签之间的差距。这使得 CNN 能够从大量的训练数据中学习到有效的特征,从而提高对未知数据的泛化能力。

图 2 为 LSTM-CNN 模型流程示意图,原始光谱数据中包含大量的无用信息或非拉曼光谱信息,在将数据送入模型前需要进行 1.3.2 小节的预处理;预处理后的光谱数据直接输入 LSTM 模型中,并得到新的一维“序列”数据;将新的序列数据送入 AM 增强后直接送入 CNN 模型中,数据在经过多次卷积和池化后,通过全连接层的矩阵相乘输入到隐藏层,最终得到预测结果。

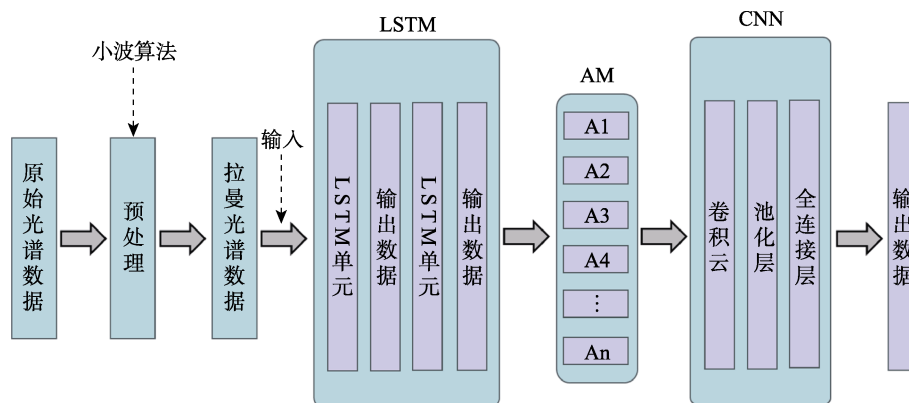


图 2 LSTM-CNN 模型流程示意图

Fig.2 LSTM-CNN model process diagram

LSTM-CNN 模型预测效果与传统化学计量学方法—偏最小二乘回归 (Partial least squares regression, PLSR) 算法进行比较, PLSR 是一种经典的化学计量学算法, 可解决多元线性回归中自变量间高度相关的问题^[22]。PLSR 具有主成分分析、典型相关分析和线性回归分析的多种优点, 不仅可以解决回归问题, 而且可以实现主成分分析的功能, 因此其可提供更多有用的信息。主成分是影响 PLSR 模型定量的关键性参数, 因此在建立 PLSR 定量分析模型时, 需要针对性优化主成分, 以实现定量分析模型的最优化^[22]。

定量模型的评价参数通常采用决定系数 (Coefficient of determination, R^2) 和均方根误差 (Root mean square error, RMSE), 其中 R^2 取值范围 0~1, 当 R^2 越趋近于 1, 说明定量模型的预测效果越好; 当 RMSE 越小说明定量模型预测值与真实值的偏差越小, 其可反映数据中的异常值。

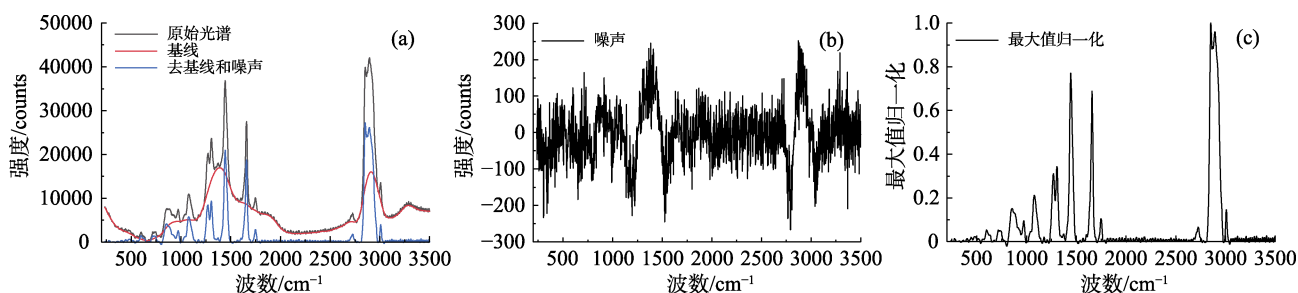
2 结果与分析

2.1 信号预处理过程

图 3 为玉米油拉曼光谱的信号预处理过程。图 3a 为原始光谱、基线和拉曼光谱的信号, 小波算法实现了基线与拉曼光谱的信号分离, 提取了油品拉曼光谱特征峰信息。图 3b 为原始光谱中小波算法去除的高频噪声信号, 图 3c 为对去基线和降噪后拉曼光谱的最大值归一化结果。

2.2 光谱稳定性

光谱的稳定性是实现掺伪定量检测的基础。影响光谱稳定性的因素主要包含两部分: 人为配比的浓度误差; 光谱在表征同一种油样的一致性表现。为说明这两个因素对光谱抖动的影响, 图 4 展示了体积比为 0.5 的玉米油掺假牡丹籽油的所有拉曼光谱, 图 4 表明人为干预和仪器抖动对光谱稳定性的影响较少, 信号预处理后的光谱表现出较好的稳定性。



注: (a) 基线校正, (b) 无效噪声, (c) 归一化。

Note: (a) baseline correction, (b) invalid noise, (c) normalization.

图 3 玉米油拉曼光谱的信号预处理过程

Fig.3 Signal preprocessing process of Raman spectrum of corn oil

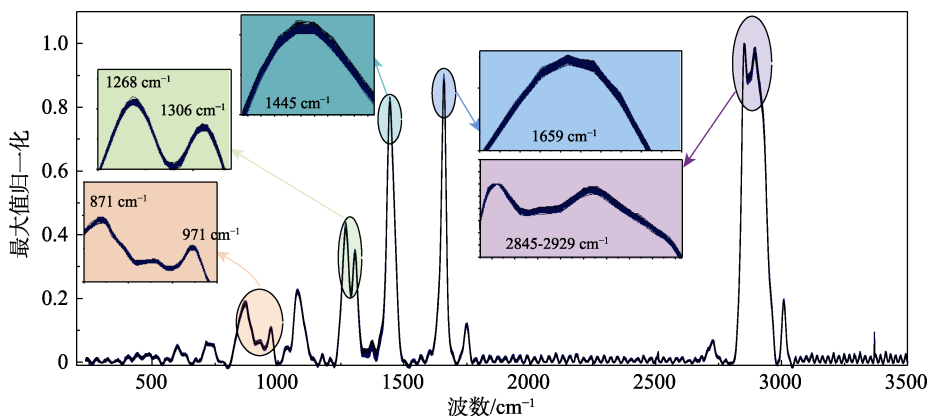


图 4 体积比为 0.5 的玉米油掺假牡丹籽油的拉曼光谱

Fig.4 Raman spectra of peony seed oil adulterated with corn oil with volume ratio of 0.5

图 5 为 11 种不同掺伪浓度混合油品拉曼光谱的抖动评价, 结果表明相同浓度内拉曼光谱的抖动可忽略不计, 即同一掺伪浓度玉米油掺假牡丹籽油的所有拉曼光谱的稳定性较好, 可用于定量分析模型的建模和测试。

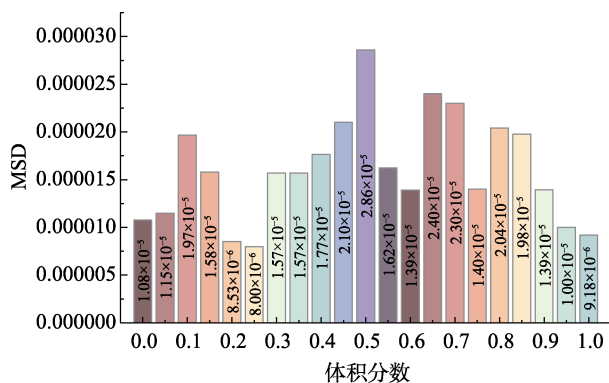


图 5 不同掺伪浓度混合油品拉曼光谱的抖动评价

Fig. 5 Jitter evaluation of Raman spectra of mixed oils with different adulteration concentrations

2.3 拉曼光谱差异性分析及特征谱峰的归属判定

图 5 表明相同掺伪浓度混合油品的拉曼光谱稳定性较好, 为了评估不同掺伪浓度混合油品的光谱差异性, 本研究对比了不同体积比牡丹籽油的平均拉曼光谱, 如图 6 所示。图 6 表明牡丹籽油和玉米油拉曼光谱的特征峰位相似度较高, 但在部分特征峰强上存在一定的差异性, 主要表现在 871、971、1 268、1 306、1 445 以及 2 845~2 929 cm^{-1} 的特征峰或区域。这些光谱差异是构建定量分析模型的主要依据, 即可理解为通过算法

模型挖掘拉曼光谱差异, 并实现与掺伪浓度的非线性映射。

图 6 中不同体积比玉米油掺假牡丹籽油的平均拉曼光谱主要为脂肪酸官能团的拉曼光谱特征信息, 为进一步深入研究造成光谱差异的主要因素, 有必要对拉曼光谱特征峰进行归属判定。

图 7 展示了不同体积比玉米油掺假牡丹籽油拉曼光谱的主要特征谱峰, 可以发现某些特征峰强 (例如 D 峰) 与掺伪浓度间存在正相关性, 即谱峰峰强反映了某种脂肪酸含量的变化。不同脂肪酸的分子结构存在差异, 可通过拉曼光谱的特征信息加以区分, 这为不同体积比玉米油掺假牡丹籽油光谱表征提供了可行性, 可理解为拉曼光谱谱峰强度反映了多种脂肪酸的不同含量信息。图 7 表明经过信号预处理后便携式近红外拉曼光谱仪收集的拉曼光谱信号具有较好的信噪比和稳定性, 满足分析该类问题的能力。

为了解拉曼光谱谱峰代表的分子信息, 本研究对图 7 中主要拉曼光谱谱峰进行了归属判定, 如表 1 所示。谱峰 A 和 C 的振动归属一致, 均为脂肪酸碳长链上亚甲基的振动产生, 谱峰 B、D 及 G 的振动源自不饱和碳双键的振动, 谱峰 E、F 及谱带 H 由 C—H 间的振动产生, 谱峰 I 则产生于 C—H 分子键的振动。因此, 差异性最大的谱峰 D 反映了不同体积比玉米油掺假牡丹籽油中不饱和碳双键的信息, 可理解为对植物油中饱和脂肪酸和不饱和脂肪酸含量的表征。综上所述, 近红

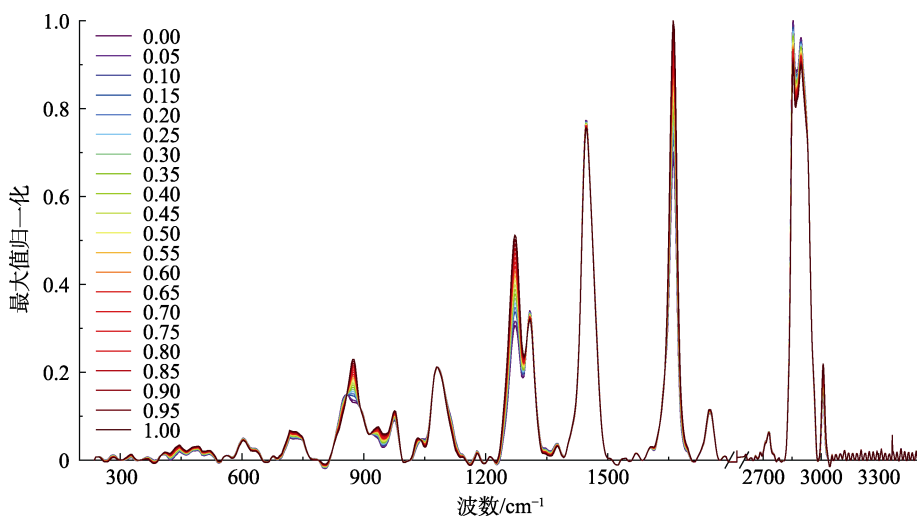
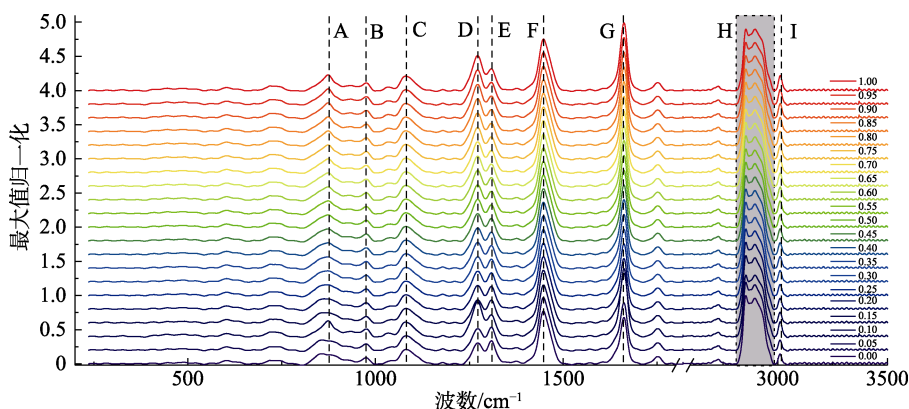


图 6 不同体积比玉米油掺假牡丹籽油的平均拉曼光谱

Fig.6 Average Raman spectra of peony seed oil adulterated with different volume ratio of corn oil



注：A~I 分别代表拉曼光谱特征峰或特征谱带。

Note: A~I represent the characteristic peak or band of Raman spectrum respectively.

图 7 不同体积比玉米油掺假牡丹籽油平均拉曼光谱及其特征峰

Fig.7 Average Raman spectra and characteristic peaks of peony seed oil adulterated with different volume ratio of corn oil

外拉曼光谱很好的表征了植物油的物质成分信息，拉曼光谱谱峰不仅反映了脂肪酸碳骨架伸缩振动信息，也反映了 C—H 基团的振动信息。

酸的种类和含量的差异引起了拉曼光谱特征峰位和强度的区别，这是进行植物油定性定量分析的最主要依据。

表 1 拉曼光谱谱峰归属判定

Table 1 Determination of Raman spectrum peak attribution

编号	峰位/cm ⁻¹	归属	参考文献
A	871	—(CH ₂) _n —	[17]
B	971	RHC=CHR	[23]
C	1 075	—(CH ₂) _n —	[17]
D	1 268	(RHC=CHR)	[23]
E	1 306	C—H	[23]
F	1 445	C—H	[24]
G	1 659	C=C	[23]
H	2 845~2 929	C—H	[17]

牡丹籽油中总不饱和脂肪酸含量高达 92%，且 ALA 含量在 40%以上，这种天然优势是其他植物油 (<10%) 无法比拟的^[1]。ALA 是一种具有 3 个双键结构的多元不饱和脂肪酸，对人体健康十分有益，长期食用牡丹籽油对人体摄取并吸收 ALA 极为有利^[1]，因此牡丹籽油对人体健康改善方面可发挥重要作用。牡丹籽油中高含量且具有 ALA 使其不饱和碳双键结构有别于其他植物油，从而引起拉曼光谱 C=C 谱峰峰强的强化，例如图 7 中的 D 峰 (1 268 cm⁻¹)。由于玉米油和牡丹籽油的不饱和脂肪酸的含量均接近 90%，而 ALA 分子中 3 个不饱和碳双键增加了牡丹籽油中不饱和碳双键的总量，因此，牡丹籽油在该位置表现出更高的信号强度。总之，不同植物油所含脂肪

2.4 LSTM-CNN 定量模型

在进行定量模型的构建中，采用信号预处理后的光谱强度作为定量分析模型的输入。21 种不同体积比的样本，将体积比为 0.0、0.05、0.15、0.20、0.25、0.30、0.40、0.45、0.55、0.60、0.70、0.75、0.80、0.85、0.95 及 1.00 的玉米油和牡丹籽油的混合油产生的拉曼光谱作为定量模型的训练集，并采用留一法进行模型的交叉验证，将体积比 0.10、0.35、0.50、0.65 及 0.90 的油样产生的拉曼光谱作为定量模型的测试集。其中，训练集共有 6 000 条拉曼光谱，测试集共有 1 875 条拉曼光谱。

PLSR 模型：由于主成分数直接影响了 PLSR 模型的定量效果，因此需要对其进行优化，基于训练集数据开展对 PLSR 主成分数的优化，其中采用 RMSE 作为优化评价参数，如图 8 所示。图 8 表明当主成分数为 8 时，PLSR 模型评价参数 RMSE 趋于稳定，因此，本研究 PLSR 模型的核心参数——主成分数设置为 8。

图 9 为基于 PLSR 算法玉米油掺假牡丹籽油的定量效果，图 9(a)为定量模型训练集的预测结果，图 9(b)为定量模型测试集的预测结果，图 9(c)为定量模型训练集和测试集的预测结果。通过对

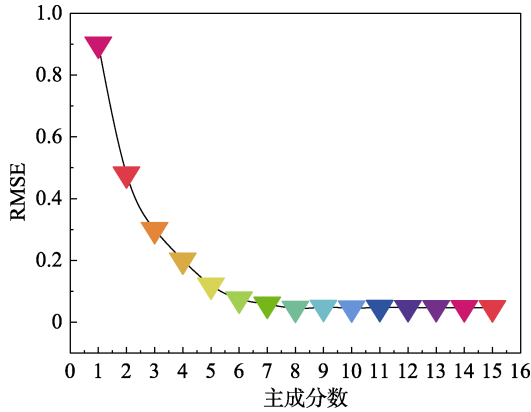


图 8 PLSR 模型主成分数优化

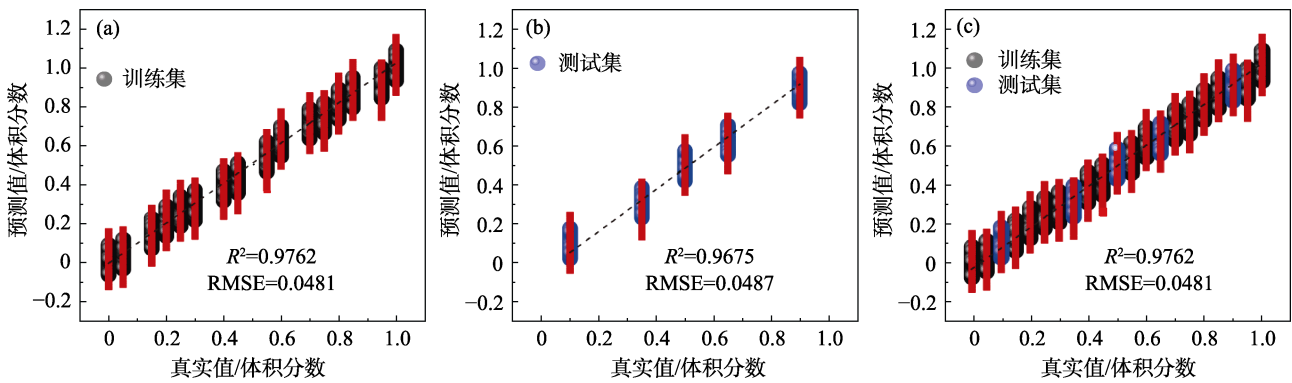
Fig.8 Optimization of principal components of PLSR model

比 R^2 和 RMSE 可以发现 PLSR 模型表现出较好的预测效果, R^2 大于 0.976 2, RMSE 小于 0.048 1。

LSTM-CNN 模型:该模型由 LSTM 和 1DCNN 网络组成, LSTM 有输入门、输出门和遗忘门组

成, 通过 3 个控制门左右历史信息的走向, 达到长时间数据的历史记忆功能。1DCNN 利用局部连接和权值共享的方式, 通过一维卷积核滤波器进行卷积、池化和特征。本模型通过引入 Dropout 层防止模型出现过拟合。LSTM-CNN 模型的失真较小, 这主要得益于 AM 对序列数据不同程度权重的分配, 使得 LSTM-CNN 模型可以获取前向和后向的特征, 更加全面的提取特征信息, 实现精准预测。图 10 为 LSTM-CNN 模型对牡丹籽油掺假对的定量预测效果。

图 9 和 10 表明在相同实验参数的条件下, LSTM-CNN 算法比 PLSR 预测效果更优, 其中模型整体 R^2 为 0.990 8, RMSE 为 0.029 9。相比于 PLSR 预测效果, LSTM-CNN 模型对牡丹籽油掺假浓度预测的 R^2 提高了 1.87%。

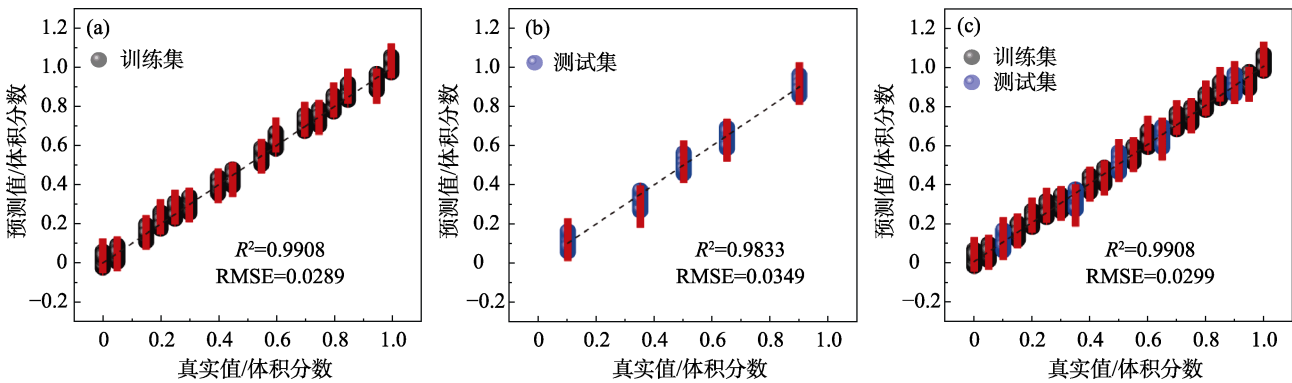


注: a 为训练集 PLSR 的预测效果; b 为测试集 PLSR 的预测效果; c 为所有光谱数据 PLSR 的预测效果。

Note: (a) represents the prediction performance of the training set PLSR; (b) represents the prediction performance of the test set PLSR; (c) represents the predictive performance of PLSR for all spectra.

图 9 基于 PLSR 算法玉米油掺假牡丹籽油的定量效果

Fig.9 Quantitative effect of peony seed oil adulterated with corn oil based on PLSR algorithm



注: a 为训练集 LSTM-CNN 的预测效果; b 为测试集 LSTM-CNN 的预测效果; c 为所有光谱数据 LSTM-CNN 的预测效果。

Note: (a) represents the prediction performance of the training set LSTM-CNN; (b) represents the prediction performance of the test set LSTM-CNN; (c) represents the predictive performance of LSTM-CNN for all spectra.

图 10 基于 LSTM-CNN 算法玉米油掺假牡丹籽油的定量效果

Fig.10 Quantitative effect of peony seed oil adulterated with corn oil based on LSTM-CNN algorithm

2.5 灵敏度分析

基于 LSTM-CNN 算法结合便携式近红外拉曼光谱实现了玉米油以不同掺伪浓度混合牡丹籽油的定量预测, 通常情况下不法商贩在造假的过程中永远以利益最大化作为衡量造假的标准, 因此, 一旦对纯牡丹籽油掺入其他廉价植物油, 会考量造假“成本”与利益量的平衡, 即掺入的廉价植物油一定不是一种低剂量。但为了衡量该探测方法预测玉米油掺伪牡丹籽油的检测极限, 本文针对性开展了低剂量掺伪的预测研究, 在掺伪体积浓度 0~0.1 范围内共制备了 4 个浓度梯度(每种浓度样本单独制备 5 份), 分别为 0.02、0.04、0.06 及 0.08。通过便携式近红外拉曼光谱装置探测 4 种低剂量掺伪浓度混合油品的拉曼光谱数据(每个样本随机收集 2 条光谱), 然后将预处理后的光谱数据送入 LSTM-CNN 模型中进行预测, 结果如图 11 所示。

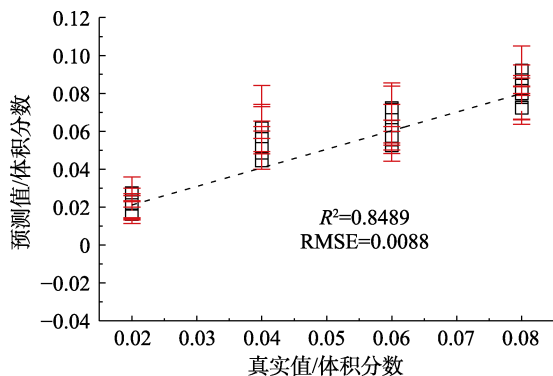


图 11 基于 LSTM-CNN 算法玉米油掺伪牡丹籽油的低浓度预测效果

Fig.11 Prediction effect of low concentration peony seed oil adulterated with corn oil based on LSTM-CNN algorithm

在低浓度掺伪定量预测中, 浓度梯度变化为 0.02 时, LSTM-CNN 模型的预测效果一般, 这主要取决于两方面因素, 一是定量模型的复杂度, 二是掺伪浓度配比的人为误差。但无论哪一种因素, 就目前定量模型的预测效果而言完全满足对实际市场上玉米油掺伪牡丹籽油检测的应用需求。


3 结论

针对牡丹籽油掺假的行为, 采用便携式近红外拉曼光谱技术实现了对不同体积比玉米油掺假牡丹籽油光谱表征, 分析了典型拉曼光谱谱峰代

表的分子信息, 评估了相同掺伪浓度油品的光谱稳定性, 以及不同掺伪浓度油品的光谱特征谱峰的差异性。结合 LSTM-CNN 算法实现了不同体积比玉米油掺假牡丹籽油的定量分析, 其中模型评价参数 R^2 为 0.990 8, RMSE 为 0.029 9。结果表明, 便携式近红外拉曼光谱结合 LSTM-CNN 算法具有面向现场原位检测牡丹籽油掺伪量的潜力, 虽然便携式仪器的稳定性和可靠性不及台式系统, 但通过光谱预处理和 LSTM-CNN 算法可以实现拉曼光谱特征信息与牡丹籽油掺伪量的映射。然而, 实际应用需要考虑商品容器对拉曼光谱的影响, 因此为深入探究该技术的应用潜力, 后续工作有必要针对性开展商品容器对拉曼光谱定量分析结果的影响。本研究提出的方法对快速、无损检测牡丹籽油的掺伪量具有一定的借鉴价值, 且针对便携式近红外拉曼光谱技术定量反演牡丹籽油掺伪的工作未见发表。

参考文献:

- [1] 郑雅琪, 武艺, 袁玮琼, 等. 多产地油用牡丹籽油脂肪酸构成及基本理化指标评价[J]. 食品工业科技, 2023, 44(15): 312-319. ZHENG Y Q, WU Y, YUAN W Q, et al. Study on the fatty acid composition and basic physical and chemical indicators of peony seed oil in different origins[J]. Science and Technology of Food Industry, 2023, 44(15): 312-319.
- [2] 邓晓军, 马金鸽, 杨巧玲, 等. 基于拉曼-紫外可见融合光谱技术的进口橄榄油质量等级可视化快速鉴别方法研究[J]. 光谱学与光谱分析, 2023, 43(4): 1117-1125. DENG X J, MA J G, YANG Q L, et al. Visualized fast identification method of imported olive oil quality grade based on Raman-UV-visible fusion spectroscopy technology[J]. Spectroscopy and Spectral Analysis, 2023, 43(4): 1117-1125.
- [3] 卢先勇, 何文绚, 陈昊聪, 等. 基于植物油三维荧光光谱的茶油鉴定模型[J]. 光谱学与光谱分析, 2020, 40(12): 3854-3859. LU X Y, HE W X, CHEN H C, et al. Identification model of camellia oil based on 3D fluorescence spectra of vegetable oils[J]. Spectroscopy and Spectral Analysis, 2020, 40(12): 3854-3859.
- [4] 王哲, 李晨曦, 钱蕊, 等. 二维相关近红外光谱的植物油鉴别方法[J]. 光谱学与光谱分析, 2020, 40(10): 3230-3234. WANG Z, LI C X, QIAN R, et al. Research on vegetable oils classification based on two-dimensional correlation near-infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(10): 3230-3234.
- [5] 唐雪梅, 王朗, 李备, 等. 拉曼光谱在食品安全检测中的应用[J]. 食品安全质量检测学报, 2023, 14(23): 51-58.

- TANG X M, WANG L, LI B, et al. Application of Raman spectroscopy in food safety detection[J]. *Journal of Food Safety & Quality*, 2023, 14(23): 51-58.
- [6] 何传箴, 许露云, 林敏琪, 等. 表面增强拉曼光谱液体活检技术研究进展[J]. *光散射学报*, 2023, 35(2): 160-173.
- HE C Z, XU L Y, LIN M Q, et al. Research progress on surface-enhanced Raman spectroscopy liquid biopsy technology [J]. *Chinese Journal of Light Scattering*. 2023, 35(2): 160-173.
- [7] 吴娟霞, 谢黎明. 二维材料的拉曼光谱研究进展[J]. *科学通报*, 2018, 35:3727-3746.
- WU J X, XIE L M. Progress on Raman spectroscopy of two-dimensional materials (in Chinese)[J]. *Chin Sci Bull*, 2018, 63: 3727-3746.
- [8] 邱梦情, 徐青山, 郑守国, 等. 农药残留检测中表面增强拉曼光谱的研究进展[J]. *光谱学与光谱分析*, 2021, 41(11): 3339-3346.
- QIU M Q, XU Q S, ZHENG S G, et al. Research progress of surface-enhanced Raman spectroscopy in pesticide residue detection[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(11): 3339-3346.
- [9] 孔维刚, 王倩楠, 常先磊, 等. 安纳咖毒品的激光共聚焦显微拉曼光谱快速检验法[J]. *中国药物依赖性杂志*, 2023, 32(4): 329-335.
- KONG W G, WANG Q N, CHANG X L, et al. Rapid detection of sodium caffeine by laser confocal micro Raman spectroscopy [J]. *Chinese Journal of Drug Dependence*, 2023, 32(4): 329-335.
- [10] 张凯萍, 李国霞. 基于反向传播神经网络结合便携式拉曼光谱特级初榨橄榄油掺假定量分析[J]. *光散射学报*. 2023, 35(1): 64-70.
- ZHANG K P, LI G X. Quantitative analysis of adulteration in extra virgin olive oil based on back propagation neural network and portable raman spectroscopy[J]. *Chinese Journal of Light Scattering*. 2023, 35(1): 64-70.
- [11] 黄秀丽, 詹云丽, 李菁, 等. 便携式激光拉曼光谱仪快速鉴别灵芝孢子油掺伪[J]. *食品工业科技*, 2016, 37(20): 59-6267.
- HUANG X L, ZHAN Y L, LI J, et al. Rapid detection of spore oil from *Ganoderma lucidum* by portable laser Raman spectroscopy [J]. *Science and Technology of Food Industry*, 2016, (20): 59-62.
- [12] 张凤娟, 黄敏, 刘振方. 基于激光拉曼光谱快速无损检测牛油果油的研究[J]. *光散射学报*, 2020, 32(3): 288-294.
- ZHANG F J, HUANG M, LIU Z F. Study on rapid non-destructive testing of avocado oil based on laser Raman spectroscopy[J]. *Chinese Journal of Light Scattering*. 2020, 32(3): 288-294.
- [13] 马金鸽, 杨巧玲, 邓晓军, 等. 基于便携式拉曼光谱的进口散装橄榄油品质现场快速无损鉴别方法[J]. *光谱学与光谱分析*, 2021, 41(9): 2789-2794.
- MA J G, YANG Q L, DENG X J, et al. On-site rapid and non-destructive identification method for imported bulk olive oil quality based on portable Raman spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(9): 2789-2794.
- [14] 张凤娟, 黄敏. 基于激光拉曼光谱的橄榄油掺伪定量检测[J]. *应用激光*, 2023, 43(10): 164-169.
- ZHANG F J, HUANG M. Quantitative detection of adulteration in olive oil based on Laser Raman spectroscopy[J]. *Applied Laser*, 2023, 43(10): 164-169.
- [15] 郭宗昱, 郭一新, 金伟其, 等. 基于紫外拉曼光谱的转基因大豆油快速识别方法研究[J]. *光谱学与光谱分析*, 2022, 42(12): 3830-3835.
- GUO Z Y, GUO Y X, JIN W Q, et al. Rapid identification of transgenic soybean oil based on ultraviolet Raman spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2022, 42(12): 3830-3835.
- [16] 赵迎, 沈学静, 李小佳. 基于荧光差分拉曼光谱技术预测食用油复热时长[J]. *核农学报*, 2023, 37(10): 2034-2041.
- ZHAO Y, SHEN X J, LI X J. Prediction of reheating time of edible oil based on shifted excitation Raman difference spectroscopy [J]. *Journal of Nuclear Agricultural Sciences*. 2023, 37(10): 2034-2041.
- [17] WANG H P, XIN Y J, MA H Z, et al. Rapid detection of chinese-specific peony seed oil by using confocal Raman spectroscopy and chemometrics[J]. *Food Chem*, 2021, 362: 130041.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [19] 郭蕴颖, 丁云峰. 基于 CNN 和 LSTM 联合预测并修正的电量缺失数据预测[J]. *计算机系统应用*, 2020, 29(8): 192-198.
- GUO Y Y, DING Y F. Prediction and correction of power loss data based on CNN and LSTM[J]. *Computer Systems and Applications*, 2020, 29(8): 192-198.
- [20] 宋晓宇, 朱林, 冯小蓓, 等. 基于改进的 LSTM-CNN 高血压组合预测模型[J]. *软件工程与应用*, 2021, 10(3): 337-347.
- SONG X Y, ZHU L, FENG X B, et al. Combined forecast model of hypertension based on improved LSTM-CNN[J]. *Software Engineering and Applications*, 2021, 10(3): 337-347.
- [21] 马瑶, 智敏, 殷雁君, 等. CNN 和 Transformer 在细粒度图像识别中的应用综述[J]. *计算机工程与应用*, 2022, 58(19): 53-63.
- MA Y, ZHI M, YIN Y J, et al. Review of applications of CNN and transformer in fine-grained image recognition[J]. *Computer Engineering and Applications*, 2022, 58(19):53-63.
- [22] 付洪波, 吴边, 王华东, 等. 激光诱导击穿光谱定量分析锂矿石中锂元素[J]. *光谱学与光谱分析*, 2022, 42(11): 3489-3493.
- FU H B, WU B, WANG H D, et al. Quantitative analysis of Li in lithium ores based on laser-induced breakdown spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(11): 3489-3493.
- [23] FAN D S, HUANG W, LIU T C, et al. Quantitative analysis of blended oils by confocal Raman spectroscopy and chemometrics in situ[J]. *Food Control*, 2022, 142: 109244.
- [24] WU X J, ZHANG X, DU Z R, et al. Raman spectroscopy combined with multiple one-dimensional deep learning models for simultaneous quantification of multiple components in blended olive oil[J]. *Food Chem*, 2024, 431: 137109. 
- 备注: 本文的彩色图表可从本刊官网 (<http://lyspkj.ijournal.cn>)、中国知网、万方、维普、超星等数据库下载获取。