

基于拉曼光谱和 CNN 算法的特级初榨 橄榄油的掺假量化

乌文泽¹, 何 凯², 吴东雷³

(1. 内蒙古交通职业技术学院, 内蒙古 赤峰 024000; 2. 太原理工大学 信息与计算机学院, 太原 030024;
3. 中国空间技术研究院 西安分院, 西安 710100)

摘要:旨在为特级初榨橄榄油掺假快速定量分析提供参考,以掺假菜籽油的特级初榨橄榄油为例,采用激光拉曼光谱实验系统获取油样的拉曼光谱数据,运用基于 Inception V2 结构的卷积神经网络(CNN)算法提取拉曼光谱特征并完成光谱特征与掺假量的非线性关系映射。结果表明:特级初榨橄榄油与菜籽油的拉曼光谱存在较大的差异,其中类胡萝卜素、碳碳双键、甲基和亚甲基产生的拉曼特征峰是引起差异的主要因素;所建立的 CNN 模型效果较好,训练集、验证集、测试集的决定系数均大于 0.99,均方根误差均小于 0.026;在低剂量掺假中,模型的预测结果仍具有一定的参考价值。综上,拉曼光谱结合基于 Inception V2 结构的 CNN 算法所建立的模型可以满足特级初榨橄榄油掺假量的快速检测。

关键词:特级初榨橄榄油;拉曼光谱;掺假量化;Inception V2 结构;卷积神经网络

中图分类号:TS225.1;O433 文献标识码:A 文章编号:1003-7969(2024)05-0133-05

Quantification of adulteration in extra virgin olive oil based on Raman spectroscopy and CNN algorithm

WU Wenze¹, HE Kai², WU Donglei³

(1. Inner Mongolia Vocational and Technical College of Communications, Chifeng 024000, Inner Mongolia, China; 2. College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China; 3. China Academy of Space Technology(Xi'an), Xi'an 710100, China)

Abstract: To provide a reference for the rapid quantitative analysis of adulteration in extra virgin olive oil (EVOO), taking EVOO adulterated with rapeseed oil as an example, the Raman spectral data of the oil samples were obtained using a laser Raman spectroscopy experimental system. The convolutional neural network(CNN) algorithm based on the Inception V2 structure was used to extract Raman spectral features and complete the nonlinear relationship mapping between spectral features and adulteration amount. The results showed that there was a significant difference between the Raman spectra of extra virgin olive oil and rapeseed oil, and the Raman characteristic peaks generated by carotenoids, C=C, methyl, and methylene groups were the main factors causing the differences. The established CNN model performed well, with determination coefficients greater than 0.99 for the training, validation, test sets and root mean square errors less than 0.026. In low-dose adulteration, the model's predictive performance still had a specific reference value. In summary, the model established by combining Raman spectroscopy with CNN algorithm based on the Inception V2 structure can meet the rapid detection of adulteration

amount in extra virgin olive oil.

Key words: extra virgin olive oil; Raman spectroscopy; quantification of adulteration; Inception V2 structure; convolutional neural network

收稿日期:2023-02-15;修回日期:2024-01-31

基金项目:国家自然科学基金(62103296)

作者简介:乌文泽(1987),女,讲师,主要从事机器学习与大数据应用的研究(E-mail)fashion102@yeah.net。

特级初榨橄榄油(EVOO)是采用机械压榨等物理方式直接从油橄榄鲜果中制取的无任何添加剂的、酸值(KOH)不超过 1.6 mg/g 的油品,其保留了大量天然营养成分和风味。目前,国内市场需求的特级初榨橄榄油仍依赖于大量进口,且其市场价格远高于普通植物油,这使得特级初榨橄榄油成为掺假概率较高的植物油。因此,如何快速筛选高价油品并完成掺假量化的检测是维护特级初榨橄榄油市场稳定,促进国内橄榄油产业健康发展的前提。

光谱检测技术是植物油掺假量化分析的重要手段之一,根据产生方式的差异,可将光谱技术分为原子光谱和分子光谱。激光诱导击穿光谱是一种已应用于植物油检测的典型原子光谱技术^[1],而应用于植物油检测的分子光谱技术包括拉曼光谱^[2]、布里渊散射光谱^[3]、红外光谱^[4]以及荧光光谱^[5-6]等。其中,拉曼光谱技术具有高灵敏度、无损伤、高效率、无样本前处理、可提供分子结构信息以及量化等优点,被广泛应用于材料科学、物理化学、生物医学、地质环境等领域。目前,已有基于拉曼光谱技术对植物油掺假的量化研究,如:周秀军等^[7]提出一种基于拉曼光谱的食用植物油快速鉴别方法,根据类中心最小距离分类法预测样本类别;张朝晖等^[8]研究认为便携式拉曼光谱仪可作为在线初步筛查橄榄油真伪的工具,但未提出具体的定量分析算法;余小敏等^[9]提出一种基于拉曼光谱现场快速鉴别火锅油与食用植物油,以及检测火锅油质量的方法。以上基于拉曼光谱结合传统化学计量学方法在植物油鉴别和量化中的研究表明,拉曼光谱技术具有推广至

一线质量监测的应用潜力。

卷积神经网络(Convolutional neural network, CNN)是一种由输入层、卷积层、激励层、池化层、全连接层和输出层组成的深度学习算法。其中:卷积层实现数据的特征提取,卷积层的卷积核是 CNN 算法的核心,其尺寸、步长决定了提取不同层次的数据特征;激活函数是激励层实现非线性映射的关键,作用于卷积层的输出;池化层实现特征压缩和主要特征的提取;全连接层是对前面逐层所有特征数据的整合。与常规的化学计量学算法相比,CNN 算法具有很强的自适应学习能力,在处理非线性映射问题中具有很好的适应性和泛化能力,此外 CNN 算法还具有层次化特征提取能力、高效的并行计算能力以及结构化信息处理能力等。在 CNN 算法结构优化方面,Ioffe 等^[10]提出的将 Inception V2 结构嵌入卷积神经网络中^[11],增强了 CNN 算法非线性映射的性能并减少了网络参数数量,提高了训练模型的效率。

本研究以掺假菜籽油(RO)的特级初榨橄榄油为例,采用便携式蓝绿激光共聚焦拉曼光谱实验系统采集其拉曼光谱,并采用基于 Inception V2 结构的 CNN 算法对数据进行处理,以期为特级初榨橄榄油掺假快速定量分析提供参考。

1 材料与方法

1.1 实验材料

特级初榨橄榄油,陇南市祥宇油橄榄开发有限责任公司;菜籽油,四川省粮油集团有限责任公司。

便携式蓝绿激光共聚焦拉曼光谱实验系统如图 1 所示。

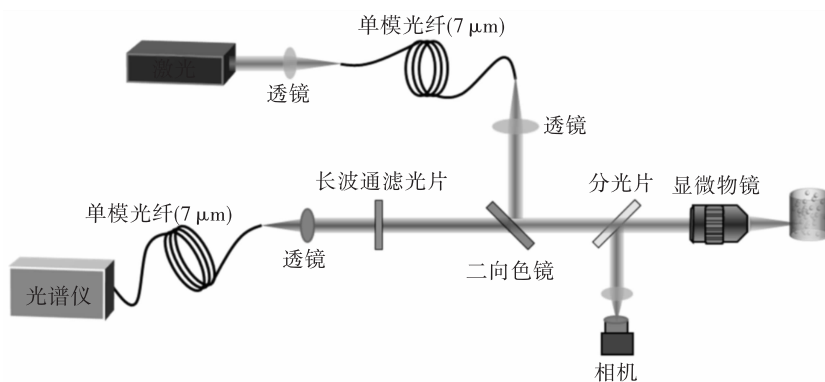


图 1 拉曼光谱实验系统光路示意图

如图 1 所示,该系统激励源为 532 nm 连续激光且功率可调(50 ~ 150 mW),作用于目标焦点的激光能量为 30 mW,通过单模光纤耦合的方式实现光学系统的准共焦。为提高系统光谱的信噪比,采用制冷低噪声的 XS7031 光纤光谱仪(上海如海光电科技有限公司)采集激光拉曼光谱,该光谱仪采集

范围 300 ~ 3 200 cm^{-1} (脂肪酸的拉曼特征峰在 3 100 cm^{-1} 以内^[12]),狭缝宽度 25 μm ,光学分辨率优于 0.8 nm。另外,该系统采用消色差 40 倍显微物镜用于激励光源的聚焦和信号收集。该系统集成了显微成像部件,用于统一激光在油样中的聚焦深度,避免了油样中激光焦点位置不同带来的光谱差

异。另外,由于采用单模光纤实现了准共焦,因此背景环境对拉曼原始光谱的采集影响较小。

1.2 实验方法

1.2.1 掺假油样的制备

向特级初榨橄榄油中依次添加不同体积的菜籽油并混合均匀,得到菜籽油体积分数分别为 1.0、0.60、0.55、0.45、0.30、0.20、0.10 以及 0.00 的掺假油样,每种样本制备 15 个。

1.2.2 光谱数据采集及预处理

采用便携式蓝绿激光共聚焦拉曼光谱实验系统采集油样的光谱数据,每条拉曼光谱的积分时间为 1 s,每个样本采集 60 条拉曼光谱。由于原始拉曼光谱数据常伴有荧光光谱、仪器暗电流产生的背景及仪器噪声,未经处理会严重影响定量分析结果,因此需要对原始拉曼光谱数据进行预处理以实现光谱的基线校正和噪声去除。本文采用文献[13]的方法对拉曼光谱数据进行预处理。

1.2.3 掺假识别模型的建立及评价

1.2.3.1 模型建立

采用基于 Inception V2 结构的 CNN 算法建模。Inception V2 结构是 Inception V1 结构的一种变形,其提高了网络解决非线性回归问题的能力。图 2 为基于 Inception V2 结构的 CNN 算法结构示意图。

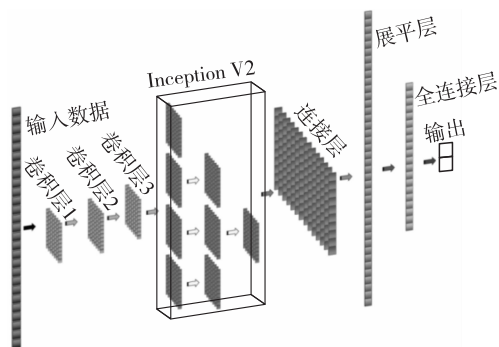
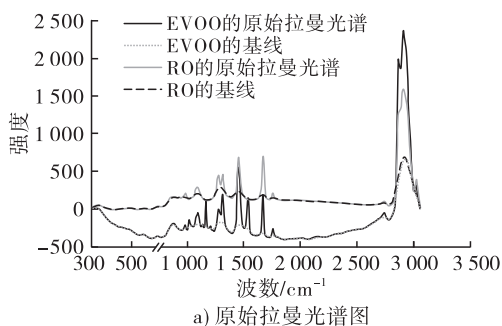


图2 基于 Inception V2 结构的 CNN 算法结构示意图

如图 2 所示,Inception V2 结构中 4 个分支 1×1 卷积核用于降低特征图数量,2、3 分支 1×3 卷积核



a) 原始拉曼光谱图

进一步提取特征,4 分支最大池化层降低特征维度并保持平移不变。Inception V2 结构是对 3 个串联卷积层提取的特征数据的进一步深入提取,全连接层是将 2 个及以上的特征按照维度进行拼接,展平层是卷积层到全连接层的过渡。

拉曼光谱为一维数据,因此需要对基于 Inception V2 结构的 CNN 定量分析模型的卷积层进行参数设置^[11],如表 1 所示。

表 1 CNN 定量分析模型卷积层参数设置

结构名称	卷积核大小	卷积核移动步长	卷积核
卷积层 1	1×3	1	8
卷积层 2	1×3	1	8
卷积层 3	1×3	1	4
Inception V2			
分支 1	1×1	1	4
分支 2	1×1	1	4
	1×3	1	4
分支 3	1×1	1	4
	1×3	1	4
	1×3	1	4
分支 4	1×1	1	4
	最大池化	1	-

将采集的拉曼光谱数据按照 4:1:1 的比例随机生成训练集、验证集和测试集样本,训练集样本用于建模,验证集样本用于模型验证,测试集样本用于模型测试。

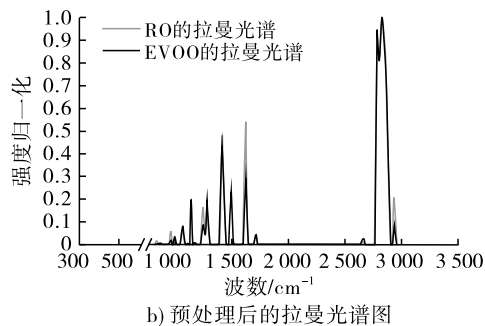
1.2.3.2 模型评价

通过决定系数 (R^2) 和均方根误差 (RMSE) 评价 CNN 模型的性能。其中: R^2 用于衡量回归模型预测效果, R^2 越趋近于 1,说明模型预测效果越好;RMSE 是评估模型预测值与实际值偏差的参数,RMSE 越小,说明量化模型的预测值与实际值之间的差异越小,即模型的预测准确性越高。

2 结果与分析

2.1 光谱数据预处理效果

特级初榨橄榄油和菜籽油的原始拉曼光谱图和预处理后的拉曼光谱图如图 3 所示。



b) 预处理后的拉曼光谱图

图 3 特级初榨橄榄油和菜籽油拉曼光谱预处理前后的光谱图

由图3可知,特级初榨橄榄油和菜籽油的原始拉曼光谱数据经处理后,消除了基线等的影响,说明本研究采用的光谱数据预处理方法有效。预处理后拉曼光谱表征了2种植物油所含成分差异,即拉曼光谱谱峰强度的差异反映了2种植物油所含脂肪酸等营养物质成分差异,而不同掺假比例则通过拉曼谱峰强度的增强或减弱得以体现。

2.2 不同掺假量油样的拉曼光谱图

图4为不同掺假量油样的拉曼光谱图(经预处理后的)。

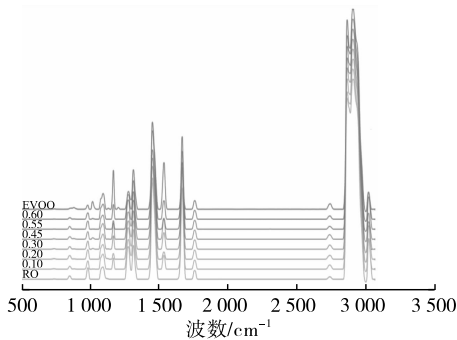


图4 不同掺假量油样的拉曼光谱图

由图4可知,某些特征谱峰的峰强与浓度存在正相关性,如类胡萝卜素产生的位于1008、1161、1528 cm^{-1} 的拉曼特征谱峰^[14]。特级初榨橄榄油保留了丰富的天然类胡萝卜素成分,其可以作为掺

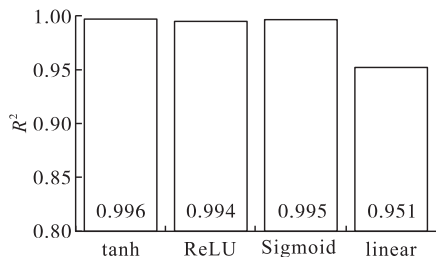


图5 不同激活函数对CNN模型的影响

由图5可见,在CNN模型训练集的性能表现中,tanh激活函数的 R^2 和RMSE均优于ReLU函数、Sigmoid函数以及linear函数,说明CNN模型在使用tanh激活函数时,模型预测效果的精度最优,因此选用tanh激活函数建模。

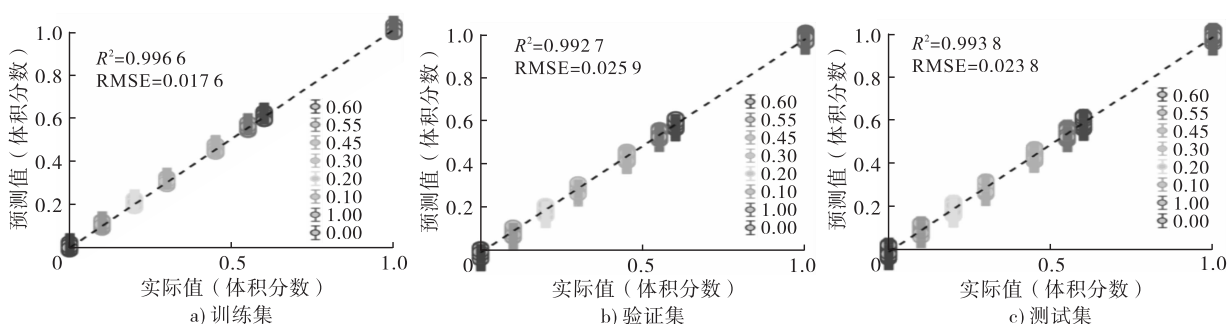


图6 基于Inception V2结构的CNN模型预测结果

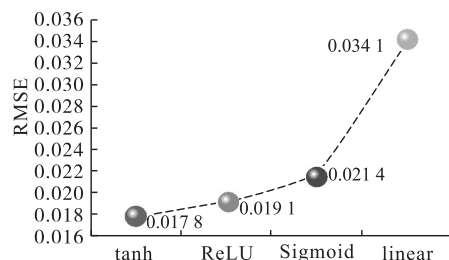
假定量分析的特征依据。此外,特级初榨橄榄油与菜籽油的脂肪酸种类及其含量存在显著差异,因此在碳碳双键、亚甲基、甲基振动产生的特征拉曼谱峰带存在明显差异^[12],这些特征谱峰和谱带的强度(1270、1305、1663 cm^{-1} 谱峰以及2800~3050 cm^{-1} 谱带)与掺假量呈现出非线性关系。

光谱仪共有2048个像元,因此采集的每条光谱共有2048个数据点,即每条光谱有2048个数据维度。图4显示在某些区域内存在大量的没有特征信息的冗余数据,因此在进行CNN模型训练时,对冗余信息进行人为干预,即删除无用数据,保留特征信息。本研究选取800~1800 cm^{-1} 以及2800~3050 cm^{-1} 内的870个数据点用于CNN模型训练、验证和测试。

2.3 基于Inception V2结构的CNN掺假定量分析模型

2.3.1 激活函数确定

8种不同掺假量油样共采集了7200条光谱数据,按4:1:1比例随机生成训练集、验证集和测试集。激活函数是影响CNN模型量化性能的重要指标之一,对比了tanh函数、ReLU函数、Sigmoid函数以及linear函数等4种常见激活函数对CNN定量分析模型的影响,结果如图5所示。



2.3.2 模型评价

在确定激活函数后,按1.2.3.1方法建模,并对CNN定量分析模型进行训练、验证和测试,结果如图6所示。

由图6可知:训练集 R^2 为0.9966, RMSE为0.0176;验证集 R^2 为0.9927, RMSE为0.0259;测试集 R^2 为0.9938, RMSE为0.0238。上述结果说明在特级初榨橄榄油掺假菜籽油的定量分析方面,采用拉曼光谱结合基于Inception V2结构的CNN算法可以快速精准地预测掺假量。

2.4 模型对低剂量掺假的分析评估

为进一步评估拉曼光谱结合基于Inception V2结构的CNN算法分析特级初榨橄榄油掺假的检出限问题,共制备了20个低掺假量的混合油品,其中菜籽油的体积分数为0.010~0.029,变化梯度为0.001,模型预测结果见图7。

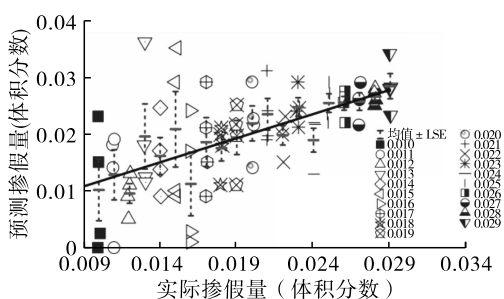


图7 低掺假量特级初榨橄榄油的基于Inception V2结构的CNN模型预测结果

由图7可知,随着掺假量的降低,CNN定量分析模型的预测效果表现较差。此外,一定范围内定量误差大于浓度变化梯度,这主要由两方面因素造成,一方面是制备样本混合油品时,操作人员引入的人为误差,另一方面是便携式拉曼光谱仪对低浓度混合油品的响应和光谱抖动造成的仪器误差,这些误差很难避免,但有优化空间。总之,这种误差虽会引起检出限和定量限降低的问题,但与拉曼光谱快速、无损定量分析的优势相比,这些缺陷在快速检测掺假应用中可暂时忽略。

3 结论

在特级初榨橄榄油掺假菜籽油的定量分析中,拉曼光谱技术结合基于Inception V2结构的CNN模型的训练集、验证集和测试集的 R^2 均大于0.99, RMSE均小于0.026,说明拉曼光谱技术结合基于Inception V2结构的CNN模型可以实现对特级初榨橄榄油掺假快速量化分析。

参考文献:

- [1] 吴宜青,刘津,莫欣欣,等. 共轴双脉冲激光诱导击穿光谱和最小二乘支持向量机法定量分析植物油中铬[J]. 分析化学, 2016, 44(12): 1919-1926.
- [2] 蒋雪松,莫欣欣,孙通,等. 食用植物油中反式脂肪酸含量的激光拉曼光谱检测[J]. 光谱学与光谱分析, 2019, 39(12): 3821-3825.
- [3] 曹晨鹏,郝仕国,罗宁宁,等. 基于受激布里渊散射及紫外-可见吸收光谱的橄榄油等级鉴别与分析[J]. 中国激光, 2018, 45(9): 165-170.
- [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.
- [5] 周孟然,王锦国,宋红萍,等. 核极限学习机和激光诱导荧光技术在食用油识别中的应用[J]. 激光与光电子学进展, 2020, 57(20): 387-393.
- [6] 周孟然,戴荣英,杨晨,等. 基于荧光光谱和堆栈自编码器的食用油快速无损检测[J]. 激光与光电子学进展, 2022, 59(8): 550-558.
- [7] 周秀军,戴连奎,李晟. 基于拉曼光谱的食用植物油快速鉴别[J]. 光谱学与光谱分析, 2012, 32(7): 1829-1833.
- [8] 张朝晖,严华,顾强,等. 便携式激光拉曼光谱法快速鉴别橄榄油掺假[J]. 食品安全质量检测学报, 2015, 6(11): 4324-4329.
- [9] 余小敏,尹蔚,罗文松. 拉曼光谱法快速鉴别劣质火锅油[J]. 中国油脂, 2021, 46(7): 99-103.
- [10] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[R]. Lille, France: Proceedings of the 32nd international conference on machine learning, 2015.
- [11] 曹学强. 基于卷积神经网络的类火星矿物成分定量分析[D]. 山东威海: 山东大学, 2021.
- [12] WANG H P, WAN X. Effect of chlorophyll fluorescence quenching on quantitative analysis of adulteration in extra virgin olive oil[J/OL]. Spectrochim Acta A, 2021, 248: 119183[2023-02-15]. <https://doi.org/10.1016/j.saa.2020.119183>
- [13] 焦青亮,刘明,于坤,等. 基于卷积神经网络的光谱预处理方法[J]. 光谱学与光谱分析, 2022, 42(1): 292-297.
- [14] 李晓丽,许凯雯,何勇. 基于拉曼光谱技术的茶叶叶片中类胡萝卜素含量的无损快速检测[J]. 光谱学与光谱分析, 2017, 37(11): 3465-3470.