

# 反射光谱结合光谱基二维卷积回归网络快速检测食用油中饱和脂肪酸

翁士状\*, 储昭结, 王满琴, 王 年

安徽大学, 农业生态大数据分析与应用技术国家地方联合工程研究中心, 安徽 合肥 230601

**摘要** 人们日常膳食中常见的食用油含有丰富的饱和脂肪酸, 饱和脂肪酸能为人体提供能量和必须营养物质, 但过量摄入会导致多种心血管疾病。结合反射率光谱和深度学习方法发展一种食用油中饱和脂肪酸含量的分析方法。首先, 测量了菜籽油、大豆油、葵花籽油、玉米油、橄榄油、芝麻油及花生油等7种食用植物油350~2500 nm范围的反射光谱, 并通过气相色谱-质谱分析法获得其软脂酸、花生酸及山萘酸等饱和脂肪酸的含量。使用中心化、多元散射校正、标准正态变量变换及标准化等算法做光谱预处理消除光谱噪声。然后, 构建了一种新型的二维光谱卷积回归网络(S2DCRN)用于脂肪酸分析, 而全卷积网络(FCN)、偏最小二乘回归(PLSR)、支持向量回归(SVR)及随机森林(RF)用于与S2DCRN模型相对比。最后, 采用序列前向选择(SFS)、随机蛙跳(RFrog)及遗传算法选取光谱特征的重要波长, 进而构建更为简单稳健的分析模型。实验结果表明, 对食用油的全光谱预处理后, S2DCRN模型性能最优, 其模型对预测集的决定系数( $R_p^2$ )达到0.9879, 均方根误差(RMSE<sub>p</sub>)为0.5100。基于重要波长的S2DCRN模型, RFrog-SFS为S2DCRN提供了最佳的预测结果 $R_p^2=0.9679$ , RMSE<sub>p}=0.4627。虽然变量选择后所取得的分析效果略差, 但光谱波长数目不足全光谱的1%, 节省了光谱数据采集工作并大幅降低了模型复杂度, 有助于后续便携式简化检测装置的研发。为进一步探究S2DCRN模型的通用性能, S2DCRN被用于食用油中花生酸和山萘酸含量分析。其中, S2DCRN对花生酸的预测结果较好 $R_p^2=0.9501$ , RMSE<sub>p}=0.1529。所提出的S2DCRN可实现反射率光谱对食用油中多种脂肪酸的准确快速分析。</sub></sub>

**关键词** 食用油; 饱和脂肪酸; 反射光谱; 卷积神经网络

**中图分类号:** O657.39 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)05-1490-07

## 引 言

食用植物油富含脂肪酸及维生素E、矿物质、类胡萝卜素等多种营养物质, 是人体能量和营养需要的重要来源<sup>[1]</sup>。当食用植物油中的饱和脂肪酸摄入过多时, 人体就会增加患胆固醇高和心血管疾病的风险。通过检测不同食用植物油中不同饱和脂肪酸含量的参考值, 不仅可以用于对食用植物油的分类, 还可以确定其变质程度和评定其氧化程度。因此, 测定出食用油中饱和脂肪酸含量对保障人们身体健康和食品安全具有重要意义<sup>[2]</sup>。

目前, 理化分析与图谱检测等方法常应用于食用油掺假、分类、新鲜度、品质及成分分析等应用中<sup>[3-4]</sup>。在理化分析方面, 气相色谱-质谱分析法(GC-MS)被用于测定食用油中6种甘油单酸酯标记物的含量, 用于区分用过的食用油和新鲜的食用油<sup>[1]</sup>。荧光光谱用于鉴定掺假植物食用油, 可在

2 min内检测出大豆油中0.4%精制掺假油<sup>[5]</sup>。理化分析方法准确度高, 但其操作过程中检测指标较多, 不适用于快速分析。相较于理化分析方法, 红外光谱、拉曼光谱和反射率光谱等图谱类方法可以做到快速且不破坏样品, 便于进行无损检测。中红外光谱被用于测定亚麻籽油与菜籽油、芝麻油和葵花籽油掺假, 准确度最高可达到99.0%, 预测误差为1.3%~2.5%<sup>[6]</sup>。红外光谱分析样品时, 样品前处理需要专业人员, 不好操控。Kwofie, Lavine等使用拉曼光谱技术对15种不同食用油分类准确率为95%<sup>[7]</sup>。拉曼光谱技术可提供快速、简单、可重复且无损的定性定量分析, 但存在灵敏度不高、需要特定波长的激光光源和精度低等问题。反射率光谱主要是通过测量样本的光谱特性计算其物理结构或化学属性来实现定量检测或定性识别。反射光谱具有波长范围宽、操作简单、扫描时间短、无需独立光源和仪器便携式等特点。反射率光谱分析带壳花生中总油脂和脂肪酸的含量, 对总油脂和脂肪酸预测决定系数 $R^2$ 值达到0.99<sup>[8]</sup>。在预测

收稿日期: 2021-01-06, 修订日期: 2021-06-29

基金项目: 国家自然科学基金项目(32001421), 安徽省重点研究与开发计划项目(202004a06020032)资助

作者简介: 翁士状, 1989年生, 安徽大学电子与信息工程学院副教授 e-mail: weng\_1989@126.com \* 通讯作者

两种山茶籽中油脂和水分含量时, 其中对油脂的预测模型  $R^2$  分别为 0.98 和 0.95, 对水分的预测模型  $R^2$  分别为 0.92 和 0.89<sup>[9]</sup>。反射率光谱分析样品时其设备方便小型便捷, 样本处理简单, 因此采用反射光谱技术对食用油脂肪酸进行分析。

利用机器学习方法实现反射率光谱的智能化、自动化分析, 可以使对食用油中脂肪酸的检测更快速、便捷。判别分析和主成分分析结合偏最小二乘回归(PLSR)对油品的 GC-MS 谱进行分析, 实现对芝麻油品掺假的判别准确率为 97.27%~100%<sup>[10]</sup>。应用 PLSR 对不同品牌不同种类食用油的中红外光谱数据进行掺假辨别, 其准确度最高可以达到 99.0%, 误差在 1.3%~2.5%之间<sup>[5]</sup>。在对油品监测分析方面, 传统机器学习快速简单, 对于小数据量、简单的关系很有效, 但在对大数据训练样本时难以很好地表达高度复杂的数据信息。区别于传统机器学习, 深度学习网络可以从大数据中自动提取特征, 从而提高模型的性能。在基于卷积神经网络(CNN)的光谱分析中, 将猪肉的中红外光谱数据集和大肠杆菌的拉曼光谱数据集可视化且基于类激活映射卷积神经网络, 平均准确率可达 98.83%和 100%<sup>[11]</sup>。Laarhoven 等利用一维 CNN 对啤酒、药片、葡萄酒、咖啡、橄榄油以及果汁的红外和拉曼光谱数据进行分类预测, 其平均准确率为 96%<sup>[12]</sup>。同样, 在将时间序列心电图数据输入 CNN 时, 平均 F1-score 为 78.2%。一维 CNN 的处理分析往往难以挖掘光谱数据的深度特征信息。为获得更精确的分析结果, 设计一个合理的能够充分挖掘光谱中信息的卷积回归网络是非常必要的。

本研究利用面向反射率光谱的光谱基二维卷积回归网络实现了食用油中饱和脂肪酸的测定。本研究的目的是: (1) 探索使用反射率光谱检测油品中饱和脂肪酸的可行性; (2) 评价中心化(CEN)、多元散射校正(MSC)、标准正态变量变换(SNV)及标准化(STA)等算法对光谱数据的去噪性能; (3) 构建一种新型的基于光谱的二维卷积回归网络(S2DCRN)用于脂肪酸分析, 并与全卷积网络(FCN)、偏最小二乘回归(PLSR)、支持向量回归(SVR)及随机森林(RF)比较; (4) 利用序列前向选择(SFS)、随机蛙跳(RFrog)及遗传算法(GA)选取重要波长, 构建简易、稳健的反射率光谱分析模型; (5) 探索 S2DCRN 模型在多种脂肪酸分析中的通用性。

## 1 实验部分

### 1.1 样本

13 种不同品牌的玉米油、13 种不同品牌的菜籽油、13 种不同品牌的芝麻油、13 种不同品牌的橄榄油、15 种不同品牌的花生油、15 种不同品牌的大豆油及 11 种不同品牌的葵花籽油均购于合肥大润发超市。每个油品取 4 份样分别置于玻璃烧杯中, 每份 25 mL, 共获得 372 个样本。

### 1.2 光谱数据采集

光谱范围为 350~2 500 nm 的反射率光谱用 PSR-3500 便携式地物光谱仪(Spectral Evolutions, Lawrence, MA, USA)测量, 其光谱分辨率为 1 nm, 测量环境如图 1 所示。

数据采集时, 每个样本测量 5 个反射光谱数据, 每测量一次移动一次位置。首先, 将光纤探头垂直对准标准白板并保持距离 2 cm 进行仪器校准, 再依次将样本置于采集台, 保持油样上表面与光纤探头垂直距离 2 cm。由于反射光谱受光照和系统物理结构差异等因素的影响, 用白板校准消除或最小化这些影响。校准公式如式(1)

$$\rho(\lambda) = \frac{V(\lambda)}{V_s(\lambda)} \rho_s(\lambda) \quad (1)$$

式(1)中,  $\rho(\lambda)$  为被测物体的反射率;  $\rho_s(\lambda)$  为标准白板的反射率;  $V(\lambda)$  和  $V_s(\lambda)$  分别为测量物体和标准白板的仪器测量值。

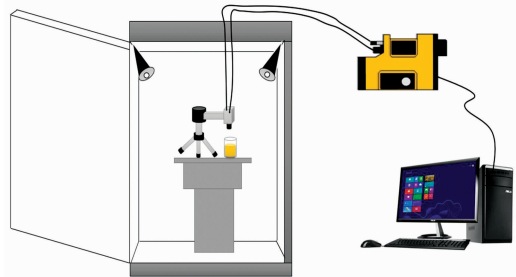


图 1 食用油的反射率光谱采集平台

Fig. 1 Image of acquisition system of the reflectance spectroscopy on edible oils

### 1.3 饱和脂肪酸参考含量检测

#### 1.3.1 仪器与环境

气相色谱质谱联用仪(GCMS-QP2010, Shimadzu, Japan)使用 DB-5MS 型毛细管柱(30 m×0.25 mm×0.25 μm), 氦气(99.99%)为载气, 设置恒定流速为 1 mL·min<sup>-1</sup>。柱箱温度从 60 °C 以 15 °C·min<sup>-1</sup> 速度升至 215 °C, 再以 10 °C·min<sup>-1</sup> 的速度升至 250 °C, 然后以 2 °C·min<sup>-1</sup> 的速度升至 260 °C, 最后以 5 °C·min<sup>-1</sup> 的速度升至 280 °C 并保持 2 min。进样口温度为 250 °C, 以 40:1 的分流比进样, 每次进样量为 1 μL。仪器的离子化方式选择电子轰击式离子源, 接口温度和离子源温度分别设置为 270 和 250 °C。将溶剂延迟时间设置为 10 min, 电子能量设置为 70 eV, 检测器电压设置为 1 500 V, 采集速率为每秒 20 个光谱。

#### 1.3.2 实验过程

由于食用油中各种脂肪酸的沸点比较高, 且含碳量相似的脂肪酸沸点相近, 不容易实现气化及分离, 因此使用 GC-MS 测定食用油中的脂肪酸含量之前需要对食用油进行甲酯化预处理。甲酯化预处理流程如下: 首先, 称量 100 mg 油样, 再加入 2 mL 正己烷并将混合物置于超声波清洗器中 20 min 使其完全溶解; 再加入 2 mL 氢氧化钾与甲醇溶液(28 mg·mL<sup>-1</sup> 氢氧化钾), 再将混合物置于超声波清洗器中 5 min, 使油同时皂化和甲酯化; 最后, 将 2 mL 盐酸加入溶液后置于超声波清洗器中 5 min 以中和过量的氢氧化钾; 静置 10 min 后取顶部有机溶液并过滤至进样瓶中, 最后进行 GC-MS 分析获取油样中软脂酸、花生酸以及山萘酸的含量。

#### 1.3.3 光谱预处理及变量选择方法

反射光谱数据除了含有油样本身的信息外, 还包含了其

他无关信息和噪声, 因此对于消除光谱数据中的无关信息及噪声的预处理举足轻重。对光谱数据的预处理主要有 CEN, MSC, SNV 和 STA 方法。CEN 可增加样品光谱之间的差异, 从而提高模型的稳健性和预测能力; MSC 可用来消除样本间的基线平移和漂移现象, 增强光谱特异性; SNV 主要用来消除固体颗粒大小、表面散射以及光程变化对漫反射的影响; STA 给光谱中所有变量相同的权重, 可以避免异常值和极端值的影响。

变量选择的方法主要包括 SFS, RFrog 及 GA。SFS 是指特征子集  $X$  从空集开始, 每次选择一个特征  $x$  加入特征子集  $X$ , 使得特征函数  $J(X)$  最优。即每次都选择一个使得评价函数的取值达到最优的特征加入, 其实就是一种简单寻找最优解的贪心算法。RFrog 利用少量的变量迭代, 输出每个变量选择可能性, 从而进行变量的选择, 是一种非常有效的高维数据变量选择方法。GA 通过数学的方式, 利用计算机仿真运算, 将问题的求解过程转换成类似生物进化中的染色体基因的交叉、变异等过程。

### 1.4 模型构建方法

#### 1.4.1 传统机器学习

在对食用油中饱和脂肪酸进行检测时, 利用数理统计方法中回归分析来确定光谱和脂肪酸变量间相互依赖的定量关系。PLSR 与主成分回归具有相似性, PLSR 没有寻找响应变量和预测变量之间最大方差的超平面, 而是分别将预测变量和响应变量投影到一个新空间, 寻找一个线性回归模型。RF

是一种由多棵决策树组成的非线性集成方法, 每棵树都依赖于独立采样的随机向量的值, 并且对森林中所有树具有相同的分布, 森林的泛化误差随着森林中树木数量的增加而收敛到一个极限。在回归分析中, 采用最小方差原则来最小化每棵树的方差。SVR 是一种非线性回归模型, 其主要思想是找到一个回归平面, 让一个集合内所有数据到该平面的距离最近。在解决非线性、小样本及高维数据实际问题时表现优异。

#### 1.4.2 FCN 网络

随着利用带标签数据量的快速增长和图形处理器的巨大进步, CNN 网络研究在各种任务上取得了较好的成果。经典的 CNN 架构除了输入层和输出层外, 还包括卷积层、池化层和完全连接层。卷积神经网络通过权值共享降低复杂度和参数, 提高神经网络的泛化能力, 通过池化操作减少神经元, 增强鲁棒性。卷积层的功能是对输入数据进行特征提取, 其内部包含多个卷积核, 组成卷积核的每个元素都对应一个权重系数和一个偏差量, 类似于一个前馈神经网络的神经元。卷积层在计算不同的特征图时, 首先可以通过将输入与学习的内核进行卷积, 然后应用基于元素的非线性来获得新的特征图, 由几个卷积核组成激活函数上的卷积结果。在卷积层进行特征提取后, 输出的特征图会被传递至池化层进行特征选择和信息过滤。池化层的每个特征图都连接到先前卷积层的相应特征图。全连接层的作用则是对提取的特征进行非线性组合从而得到输出。

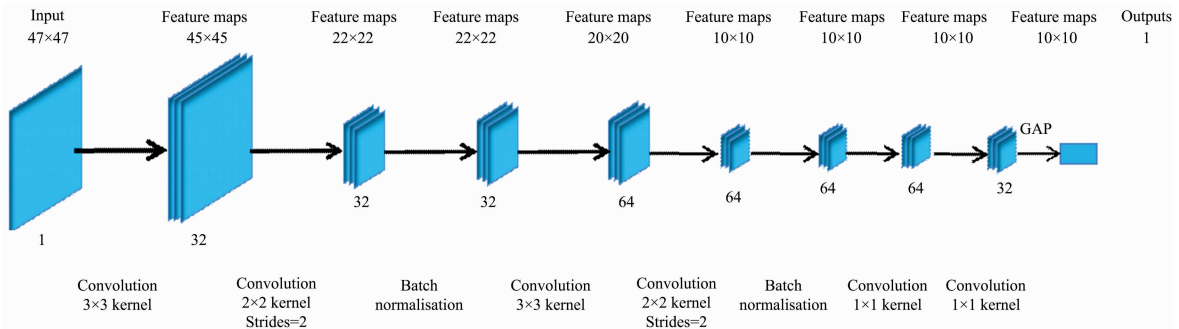


图 2 全卷积网络结构

Fig. 2 Structure of full convolution network

通常 CNN 在卷积之后会接上若干个全连接层, 将卷积层产生的特征图映射成为一个固定长度的特征向量。FCN 将 CNN 中的全连接层转化成一个个的卷积层。FCN 的架构 (图 2) 来源于 CNN, FCN 将池化层和全连接层替换为卷积层, 每一层的 sigmoid 函数都增加了网络的非线性度。在所有卷积层之后, 采用全局平均池化进行正则化, 加强特征图与标签之间的对应关系, 对输入空间变换具有更强的鲁棒性。由于 CNN 对二维矩阵形式的数据输入表现力较好, 故将一维向量形式的光谱数据转换为二维矩阵形式输入网络模型 (如图 3)。若  $n_1 \times n_2$  大于  $n$ , 则在最后补零使其构成  $n_1 \times n_2$  的矩阵。

#### 1.4.3 S2DCRN 网络

对食用油脂酸回归分析设计了光谱基二维卷积回归网

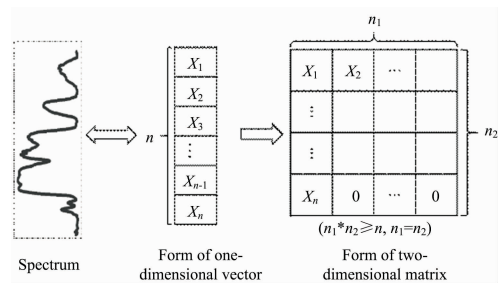


图 3 二维矩阵形式的光谱输入

Fig. 3 Spectral input form of two-dimensional matrix

络 (图 4), 为增加网络的非线性建模能力, 采用 sigmoid 函数与卷积层和池化层相结合的方法用于提取主要特征, 提高学

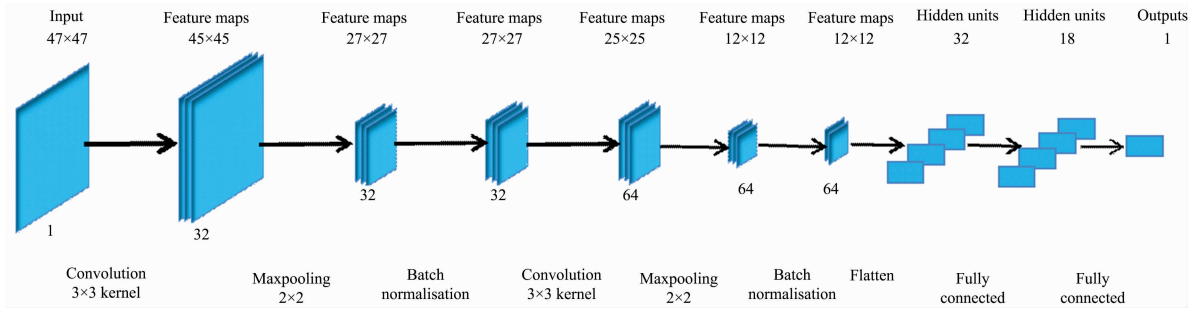


图 4 面向光谱的二维卷积回归网络结构图

Fig. 4 Structure of spectroscopy-based two Dimension convolution regression network

习效率。全连接层的核心操作就是矩阵向量乘积，相当于一个特征空间线性变换到另一个特征空间，可以把有用的信息提取整合。采用交叉熵损失作为损失函数，同时增加 1 个全连接层和 1 个未激活函数隐藏单元的输出层，该网络光谱数据同样以二维矩阵形式输入(图 3)。

1.4.4 模型评估

为评估机器学习与深度学习的效果，将 1 860 个食用油的反射光谱数据按 3 : 1 的比例分为训练集和测试集。利用 1 395 个光谱作为训练集和 465 个光谱作为预测集训练模型、调整模型参数和评级模型性能。用均方根误差(RMSE)和决定系数( $R^2$ )评价定量分析的质量。分别采用  $R^2_t$ ,  $RMSE_t$  和  $R^2_p$ ,  $RMSE_p$  来评价模型在训练集和测试集中的性能。RMSE 和  $R^2$  的计算公式如式(2)和式(3)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

其中  $y_i$  和  $\hat{y}_i$  分别表示真实值和预测值， $\bar{y}$  是真实值的平均值， $n$  为样本数。

PLSR, RF, SVR, FCN 和 S2DCRN 在 Python 和 Keras 的环境下实现，且所有方法均在 NVidia GeForce RTX

2080Ti 和 Intel Core i7-8700 CPU 的电脑设备进行。

2 结果与讨论

2.1 食用油反射率光谱特性

玉米油、菜籽油、芝麻油、橄榄油、花生油、大豆油和葵花籽油的原始光谱如图 5(a)所示。

每类油样品的平均反射光谱如图 5(b)所示。从图中可以看出，在反射光谱的 350~700 nm 波段范围内，7 类食用油的反射光谱之间存在较大的差别，其中花生油的反射率光谱变化随着波长增大而快速增加；在反射光谱的 700~2 500 nm 处，7 类食用油光谱的总体趋势是相似的。反射率变化与化学基团振动产生的倍频信息和合频信息有关，差异表现在不同种类食用油的反射率大小不同，这些差异是由于被测物在可见光谱范围内的颜色特征不同所导致的，具体到食用油的反射光谱检测中，350~700 nm 之间的反射光谱差异可能是不同种类食用油中叶绿素和胡萝卜素的含量不同引起的<sup>[8]</sup>。在反射光谱的 700~1 700 nm 处出现了几个波峰和波谷，其中位于 856 和 1 098 nm 处的波峰归属于 C—H 键的三级倍频振动，而 1 586 nm 处的波峰归属于 N—H 键的二级倍频振动。在反射光谱 1 320 nm 处的波峰与 C—H 键的弯曲振动有关，而 980 nm 处的波峰与 O—H 键的弯曲振动有关。

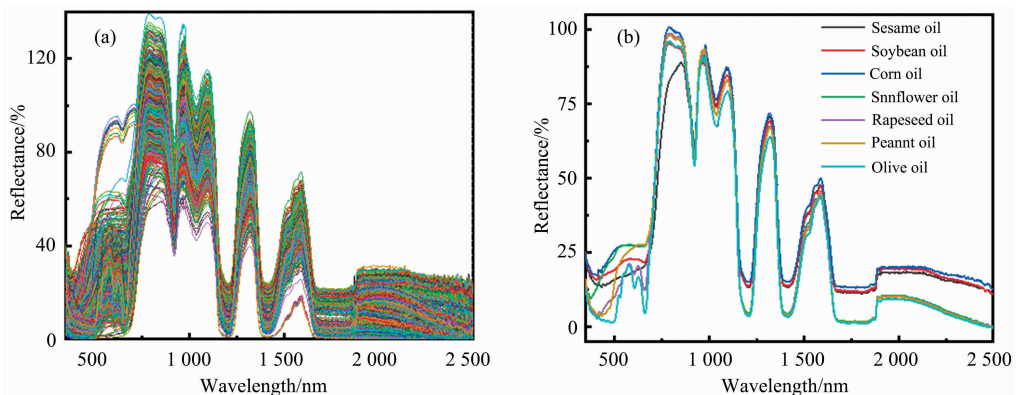


图 5 所有油样的原始反射率光谱 (a) 以及七种食用油的平均反射率光谱 (b)

Fig. 5 Reflectance spectra for all kinds of oil samples (a) and mean reflectance spectra of sesame oil, soybean oil, corn oil, sunflower oil, rapeseed oil, peanut oil and olive oil (b)



## 2.2 食用油中软脂酸含量的预测

采用 CEN, MSC, SNV 及 STA 等对食用油光谱去除噪声后, 采用 PLSR, RF, SVR, FCN, S2DCRN 等建立的回归模型对食用油中饱和脂肪酸含量进行预测分析。为选取适宜且最佳的预处理算法, 利用 PLSR 进行比较。经对比分析, SNV 预处理后的性能最优(表 1), 其  $R^2_{\text{T}}$  为 0.967 4,  $\text{RMSE}_{\text{T}}$  为 0.741 2,  $R^2_{\text{P}}$  为 0.887 8,  $\text{RMSE}_{\text{P}}$  为 1.322 1; MSC 较之略差,  $R^2_{\text{T}}$  为 0.958 2,  $\text{RMSE}_{\text{T}}$  为 0.769 3,  $R^2_{\text{P}}$  为 0.872 5,  $\text{RMSE}_{\text{P}}$  为 1.525 3; CEN 的  $R^2_{\text{T}}$  为 0.944 8,  $\text{RMSE}_{\text{T}}$  为 0.824 1,  $R^2_{\text{P}}$  为 0.866 2,  $\text{RMSE}_{\text{P}}$  为 1.598 5; STA 结果最差, 其  $R^2_{\text{T}}$  为 0.932 6,  $\text{RMSE}_{\text{T}}$  为 0.847 6,  $R^2_{\text{P}}$  为 0.853 5,  $\text{RMSE}_{\text{P}}$  为 1.664 7。

表 1 不同预处理方法后 PLSR 对食用油中软脂酸含量预测结果

Table 1 Prediction results of palmitic acid content in edible oils by PLSR after different pretreatment methods

Methods	$R^2_{\text{T}}$	$\text{RMSE}_{\text{T}}$	$R^2_{\text{P}}$	$\text{RMSE}_{\text{P}}$
CEN	0.944 8	0.824 1	0.866 2	1.598 5
MSC	0.958 2	0.769 3	0.872 5	1.525 3
SNV	0.967 4	0.741 2	0.887 8	1.322 1
STA	0.932 6	0.847 6	0.853 5	1.664 7

选定预处理方法 SNV 后, 分别采用 PLSR, RF, SVR, FCN 和 S2DCRN 对食用油中软脂酸含量进行分析。从表 2 可以看出, S2DCRN 的性能优于 PLSR, RF, SVR 和 FCN。结果显示, S2DCRN 的  $R^2_{\text{T}}$  为 0.991 1,  $\text{RMSE}_{\text{T}}$  为 0.185 1,  $R^2_{\text{P}}$  为 0.987 9,  $\text{RMSE}_{\text{P}}$  为 0.510 0, 预测结果最佳; FCN 结果次之,  $R^2_{\text{T}}$  为 0.939 3,  $\text{RMSE}_{\text{T}}$  为 0.925 1,  $R^2_{\text{P}}$  为 0.919 3,

表 2 采用 SNV 去噪后对食用油中软脂酸含量预测结果

Table 2 Prediction results of palmitic acid content in edible oils after SNV denoising

Methods	$R^2_{\text{T}}$	$\text{RMSE}_{\text{T}}$	$R^2_{\text{P}}$	$\text{RMSE}_{\text{P}}$
PLSR	0.967 4	0.741 2	0.887 8	1.322 1
RF	0.989 2	0.569 4	0.899 5	1.485 2
SVR	0.926 0	1.011 6	0.877 7	1.262 1
FCN	0.939 3	0.925 1	0.919 3	0.960 7
S2DCRN	0.991 1	0.185 1	0.987 9	0.510 0

$\text{RMSE}_{\text{P}}$  为 0.960 7; SVR 预测结果最不理想, 其  $R^2_{\text{T}}$  为 0.926 0,  $\text{RMSE}_{\text{T}}$  为 1.011 6,  $R^2_{\text{P}}$  为 0.877 7,  $\text{RMSE}_{\text{P}}$  为 1.262 1。分析结果表明, 深度学习网络的表现较优于传统机器学习, S2DCRN 性能最佳。因此, 利用反射率光谱技术与光谱基二维卷积回归网络相结合可以快速检测食用油中饱和脂肪酸含量, 此方法真实可靠。

## 2.3 重要波长的选择与验证

2 151 维的全光谱会降低模型的训练速度和分析效率, 无法满足快速检测的需求。特征提取或者特征选择能够减少光谱的维度, 筛选出重要的特征, 有效提高模型的鲁棒性。采用 SFS, RFrog 及 GA 等方法选取光谱特征的重要波长。在 RFrog 处理后, 食用油的光谱维数从 2 151 降至 64。但对于食用油中脂肪酸的快速检测, 光谱维数仍然太多。为进一步简化模型, 分别再利用 GA 和 SFS 对于选定的 64 个波长进一步筛选。通过 RFrog-GA 和 RFrog-SFS 选取, 分别从 2 151 个全光谱波长中选取了 16 和 14 个重要波长, 所选重要波长的重要性从大到小排序(见表 3)。由图 5(b)可知, 7 类食用植物油的反射率光谱在 350~700 nm 范围内差异较大, 在 700~2 500 nm 范围内光谱强度和趋势均相似。RFrog-SFS 选择的特征波长均位于在 350~700 nm 光谱范围内。虽通过 RFrog-GA 选择的特征波长数目与 RFrog-SFS 相当, 但仍有部分重要波长位于 1 200~2 200 nm 范围内。结果表明, 通过 RFrog-SFS 选择的重要波长数目最少且重要波长的分布范围最小。

PLSR, RF, SVR, FCN 和 S2DCRN 模型被应用于食用油中软脂酸含量分析, 其中 S2DCRN 模型的分析结果最佳。为评估重要波长的合理性, 采用 S2DCRN 对重要波长进行评估。利用重要波长, 重新构建 S2DCRN 模型并对软脂酸含量进行分析。从表 4 可以看出, 在 RFrog 处理后, S2DCRN 模型的  $R^2_{\text{P}}$  为 0.962 4,  $\text{RMSE}_{\text{P}}$  为 0.322 1; 基于 RFrog-GA 选择的重要波长, S2DCRN 在测试集中表现为  $R^2_{\text{P}}=0.947 6$ ,  $\text{RMSE}_{\text{P}}=0.376 4$ ; 而在 RFrog-SFS 处理后, S2DCRN 对于测试集的  $R^2_{\text{P}}$  为 0.967 9,  $\text{RMSE}_{\text{P}}$  为 0.462 7。结果表明, 重要波长的选取对于食用油中的脂肪酸预测是有效的, 但 RFrog-SFS 选择的重要波长数目最少具有更强的特异性。虽然有效波长选择的预测结果较全光谱而言略差, 但重要波长不仅可以消除光谱数据中的冗余信息并提高数据处理速度, 也为开发便携、低成本的仪器奠定了理论基础。

表 3 不同方法选取的重要波长

Table 3 Important wavelengths selected by different methods

算法	重要波长数	重要波长/nm
RFrog	64	646, 464, 414, 1647, 532, 546, 612, 580, 1 905, 554, 2 005, 551, 622, 2 006, 1 649, 535, 692, 669, 458, 2 158, 547, 534, 2 399, 460, 1 212, 1 253, 481, 629, 455, 645, 543, 732, 1 482, 1 900, 457, 548, 1 655, 632, 465, 388, 696, 439, 401, 530, 1 507, 584, 1 197, 1 638, 1 653, 2 351, 2 219, 1 964, 618, 2 156, 1 635, 545, 625, 456, 482, 1 576, 2 186, 2 217, 560, 1 645
RFrog-GA	16	532, 534, 692, 464, 629, 551, 646, 612, 1 905, 2 005, 1 253, 2 158, 554, 439, 458, 622
RFrog-SFS	14	388, 401, 414, 439, 455, 456, 457, 458, 460, 464, 465, 481, 482, 530

表 4 重要波长选择后 S2DCRN 对食用油中软脂酸含量预测结果

Table 4 Prediction results of S2DCRN on palmitic acid content in edible oil after important wavelength selection

变量选择方法	维数	$R_T^2$	RMSE <sub>T</sub>	$R_P^2$	RMSE <sub>P</sub>
RAW	2 151	0.991 1	0.185 1	0.987 9	0.510 0
RFrog	64	0.993 2	0.043 7	0.962 4	0.322 1
RFrog-GA	16	0.989 6	0.067 4	0.947 6	0.376 4
RFrog-SFS	14	0.987 3	0.090 4	0.967 9	0.462 7

#### 2.4 利用 S2DCRN 模型对食用油中花生酸和山嵛酸含量分析

将上述 S2DCRN 模型分别应用于对食用油中花生酸和山嵛酸的分析(表 5)。从表 5 可知, S2DCRN 结合全光谱对花生酸的预测结果较好, 与软脂酸分析结果相当, 其中  $R_T^2=0.996 7$ ,  $RMSE_T=0.185 1$ ,  $R_P^2=0.983 0$  和  $RMSE_P=0.090 2$ 。经过 RFrog-SFS 变量选择之后, 光谱维度降到 14,

表 5 重要波长选择后 S2DCRN 对食用油中花生酸和山嵛酸的含量预测结果

Table 5 Prediction results based on S2DCRN for arachidic acid and behenic acid content in edible oils after important wavelength selection

饱和脂肪酸种类	变量选择方法	维数	$R_T^2$	RMSE <sub>T</sub>	$R_P^2$	RMSE <sub>P</sub>
花生酸	—	2 151	0.996 7	0.185 1	0.983 0	0.090 2
	RFrog-SFS	14	0.973 5	0.102 9	0.950 1	0.152 9
山嵛酸	—	2 151	0.983 2	0.035 2	0.966 2	0.051 8
	RFrog-SFS	14	0.957 6	0.056 9	0.948 6	0.067 1

## References

- [1] Cao G D, Ding C, Ruan D L, et al. Food Control, 2019, 96: 494.
- [2] Nosratpour M, Farhoosh R, Sharif A. European Journal of Lipid Science and Technology, 2017, 119(12): 1700203.
- [3] Li X, Zhang L X, Zhang Y, et al. Trends in Food Science & Technology, 2020, 101: 172.
- [4] Rohman A. International Journal of Food Properties, 2017, 20(7): 1447.
- [5] Xu L Z, Xu X H, Xiong H, et al. Sensors and Actuators B: Chemical, 2014, 203: 697.
- [6] Jovic O. Food Chemistry, 2016, 213: 791.
- [7] Kwofie F, Lavine B K, Ottaway J, et al. Journal of Chemometrics, 2020, 34(7): 3173.
- [8] Sundaram J, Kandala C V, Holser R A, et al. Journal of the American Oil Chemists Society, 2010, 87(10): 1103.
- [9] Zhang Y Z, Zhang L B, Wang J, et al. Molecules, 2018, 23(9): 2332.
- [10] Xing C R, Yuan X Y, Wu X Y, et al. LWT-Food Science and Technology, 2019, 108: 437.
- [11] Zhang X L, Xu J F, Yang J, et al. Analytica Chimica Acta, 2020, 1119: 41.
- [12] Acquarelli J, van Laarhoven T, Gerretzen J, et al. Analytica Chimica Acta, 2017, 954: 22.

预测结果略有下降,  $R_T^2=0.973 5$ ,  $RMSE_T=0.102 9$ ,  $R_P^2=0.950 1$  和  $RMSE_P=0.152 9$ 。S2DCRN 在山嵛酸的分析中, 也取得相似的结果, 其中全光谱对应  $R_T^2=0.983 2$ ,  $RMSE_T=0.035 2$ ,  $R_P^2=0.966 2$  和  $RMSE_P=0.051 8$ ; 特征光谱对应  $R_T^2=0.957 6$ ,  $RMSE_T=0.056 9$ ,  $R_P^2=0.948 6$  和  $RMSE_P=0.067 1$ 。结果表明, 所提出的 S2DCRN 可实现反射率光谱对食用油中多种脂肪酸的准确快速分析, 也表明所提出的 S2DCRN 模型具有很强的适用性。同时, 结合 RFrog-SFS 可以大幅降低模型复杂度和所需的光谱变量, 为后续便携式简化检测装置的研发提供理论及新的思路。

## 3 结 论

使用反射光谱技术和多元分析方法快速检测食用植物油中饱和脂肪酸含量。利用 CEN, MSC, SNV 和 STA 对光谱进行预处理并进行比较, SNV 对油品的反射率光谱预处理效果最佳。PLSR, RF, SVR, FCN 和 S2DCRN 结合全波长反射光谱分析食用油中饱和脂肪酸含量, 其中 S2DCRN 模型取得了  $R_P^2=0.987 9$ ,  $RMSE_P=0.510 0$  的最佳分析结果。对于光谱重要波长, S2DCRN 取得的最优结果  $R_P^2=0.967 9$ ,  $RMSE_P=0.462 7$ 。虽然基于重要波长的结果较全波长分析结果略差, 但经 RFrog-SFS 筛选后, 光谱维数从 2 151 降至 14。在花生酸和山嵛酸含量分析中, 其结果与软脂酸的分析结果相当。S2DCRN 依然在全光谱中表现出最佳结果, 选择重要波长的结果略差。总之, 反射率光谱结合 S2DCRN 可以准确地分析食用植物油中软脂酸、花生酸以及山嵛酸的含量。虽然选择重要波长的分析结果略差, 但光谱维度不足全波长光谱的 1%, 大幅降低了模型复杂度, 有助于后续便携式简化检测装置的研发。所提出的反射率光谱结合 S2DCRN 可实现对食用油中的饱和脂肪酸含量快速、准确检测, 有望应用到其他食品的品质快速分析。

# Reflectance Spectroscopy for Accurate and Fast Analysis of Saturated Fatty Acid of Edible Oil Using Spectroscopy-Based 2D Convolution Regression Network

WENG Shi-zhuang\*, CHU Zhao-jie, WANG Man-qin, WANG Nian

National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei 230601, China

**Abstract** The edible oil in the humandaily diet is rich in saturated fatty acids, which can provide energy and other healthy nutrients for the human body, but excessive intake of saturated fatty acids can lead to a variety of cardiovascular diseases. In this study, a method for analyzing the content of saturated fatty acids in edible oils was developed by combining reflectance spectroscopy and machine learning. Firstly, the reflectance spectra of 7 edible vegetable oils, such as rapeseed oil, soybean oil, sunflower seed oil, corn oil, olive oil, sesame oil and peanut oil, were measured in the range of 350~2 500 nm, as well as the contents of palmitic acid, arachidonic acid and behenic acid were obtained by GC-MS. Spectral preprocessing algorithms were employed to eliminate the noise in spectra, including centralization, multiple scattering correction, standard normal variable transformation and standardization. Then, a novel two-dimensional spectral convolution regression network (S2DCRN) was constructed for fatty acids analysis, and a full convolutional network (FCN), partial least squares regression (PLSR), support vector regression (SVR) and random forest (RF) were compared with S2DCRN. Finally, sequential forward selection (SFS), random frog (RFrog) and genetic algorithm were used to select important wavelength spectra to re-build more simple and robust analysis models. The results showed that the S2DCRN obtained optimal performance after pretreatment of edible oil spectra with the determination coefficient of prediction set ( $R_p^2$ ) of 0.987 9 and the root mean square error of prediction set ( $RMSE_p$ ) of 0.510 0. Based on important wavelengths selected by combination RFrog and SFS, the S2DCRN exhibited excellent performance with  $R_p^2=0.967 9$  and  $RMSE_p=0.462 7$ . Although the results based on important wavelengths obtained is slightly worse, the number of wavelengths is less than 1% of the full spectra. It is convenient for the measurement of spectra and remarkably reduces the complexity of the model, which is helpful for the further development of portable and simplified detection devices. In addition, to further explore the generalization and applicability of S2DCRN, S2DCRN was used to analyze the content of arachidonic acid and behenic acid and gain a prediction result for arachidonic acid  $R_p^2=0.950 1$ ,  $RMSE_p=0.152 9$ . Therefore, the proposed method accurately and rapidly analysed various fatty acids in edible vegetable oils by reflectance spectroscopy.

**Keywords** Edible oil; Saturated fatty acids; Reflectance spectroscopy; Convolutional neural network

(Received Jan. 6, 2021; accepted Jun. 29, 2021)

\* Corresponding author