DOI: 10.19812/j.cnki.jfsq11-5956/ts.20251203002
引用格式: 陈若欣, 宁巍, 李明泽, 等. 深度学习与可解释人工智能驱动可见-近红外光谱无损检测牛肉硫代巴比妥酸含量[J]. 食品安全质量检测学报, 2026, 17(8): 76–84.
CHEN RX, NING W, LI MZ, *et al*. Non-destructive detection of the content of thiobarbituric acid in beef by visible-near infrared spectroscopy driven by deep learning and explainable artificial intelligence [J]. Journal of Food Safety & Quality, 2026, 17(8): 76–84. (in Chinese with English abstract).

深度学习与可解释人工智能驱动可见-近红外光谱无损检测牛肉硫代巴比妥酸含量

陈若欣, 宁 巍, 李明泽, 毕景然, 张公亮, 侯红漫*
(大连工业大学食品学院, 大连 116034)

摘 要: 目的 利用深度学习与可解释人工智能驱动的可见-近红外(visible near infrared, Vis-NIR)光谱技术实现牛肉中硫代巴比妥酸反应物(thiobarbituric acid reactive substances, TBARS)快速无损检测。**方法** 通过对 420 份牛肉样品进行 Vis-NIR 光谱采集和 TBARS 测定, 并比较不同的预处理方法、特征选择方法和建模方法, 实现牛肉 TBARS 的无损快速检测。**结果** 标准正态变换-逐步投影算法-长短期记忆网络模型表现最优[测试集决定系数(R-square of prediction set, R^2_P)=0.8378, 测试均方根误差(root mean square error of prediction set, RMSEP)=0.0679 mg/kg, 相对预测偏差(residual prediction deviation, RPD)=2.4833], 优于偏最小二乘回归(partial least squares regression, PLSR)和极端梯度提升(extreme gradient boosting, XGBoost)。进一步通过沙普利可加性解释(Shapley additive explanations, SHAP)可解释性分析, 揭示了 1080 nm、956 nm 等关键波长对 TBARS 预测的贡献机制。**结论** 本研究提出的方法为牛肉 TBARS 的快速、无损、精准且可解释检测提供了创新策略。

关键词: 硫代巴比妥酸反应物; 深度学习; 可解释人工智能; 长短期记忆网络; 无损检测; 可见-近红外光谱

Non-destructive detection of the content of thiobarbituric acid in beef by visible-near infrared spectroscopy driven by deep learning and explainable artificial intelligence

CHEN Ruo-Xin, NING Wei, LI Ming-Ze, BI Jing-Ran,
ZHANG Gong-Liang, HOU Hong-Man*
(*School of Food Science and Technology*, *Dalian Polytechnic University*, *Dalian* 116034, *China*)

ABSTRACT: Objective To achieve rapid and non-destructive detection of thiobarbituric acid reactive substances (TBARS) content in beef by leveraging deep learning and explainable artificial intelligence-driven visible-near

收稿日期: 2025-12-03
基金项目: 国家重点研发计划项目(2022YFD2100504)
第一作者: 陈若欣(2001—), 女, 硕士研究生, 主要研究方向为食品安全。E-mail: 3254079975@qq.com
***通信作者:** 侯红漫(1964—), 女, 博士, 教授, 主要研究方向为食品安全。E-mail: houhongman@dlpu.edu.cn

infrared (Vis-NIR) spectroscopy. **Methods** Vis-NIR spectra were collected from 420 beef samples, and the TBARS values were determined experimentally. Different spectral pretreatment methods, feature selection methods and modeling algorithms were compared to realize the non-destructive and rapid detection of TBARS in beef. **Results** It demonstrated that the standard normal variate preprocessing-successive projections algorithm for feature selection-long short-term memory network model achieved the best performance [R-square of prediction set (R^2_P)=0.8378, root mean square error of prediction set (RMSEP)=0.0679 mg/kg, residual prediction deviation (RPD)=2.4833], outperforming partial least squares regression (PLSR) and extreme gradient boosting (XGBoost). Furthermore, through the interpretability analysis using Shapley additive explanations (SHAP) revealed the contribution mechanisms of critical wavelengths, such as 1080 nm and 956 nm, to TBARS prediction. **Conclusion** The proposed method offers an innovative strategy for rapid, non-destructive, accurate and explainable detection of TBARS in beef.

KEY WORDS: thiobarbituric acid reactive substances; deep learning; explainable artificial intelligence; long short-term memory; non-destructive detection; visible-near infrared spectroscopy

0　引　言

随着经济水平的提升与健康饮食理念的普及, 作为优质蛋白质来源的牛肉需求持续增长, 但其易腐败特性对贮藏、运输及销售环节提出了严峻挑战。在牛肉贮藏过程中发生的一些化学反应中, 由多不饱和脂肪自氧化引发的脂质氧化是导致牛肉风味变差、营养价值降低以及货架期缩短的关键因素之一[1]。脂质氧化是肉类行业非常关注的问题, 因为它不仅会导致不良异味和酸败的产生, 还会伴随过渡金属离子(如 Fe^{2+}、Fe^{3+})的释放以及脂质氢过氧化物的形成, 这些物质可作为氧化反应的前体或催化剂, 促进活性氧的生成, 导致肉类进一步的有害变化。脂质氧化不仅会生成丙二醛(malondialdehyde, MDA)和胆固醇氧化产物[2]等具有毒性和潜在健康风险的物质, 还会产生大量挥发性羰基、醇和有机酸类化合物。这些挥发性氧化产物是肉品风味的重要组成部分, 但其过度积累可能导致异味和品质劣变。其中, MDA 含量通常通过硫代巴比妥酸反应物(thiobarbituric acid reactive substances, TBARS)值来衡量[3], 被广泛认为是评估牛肉脂质氧化程度和品质的重要指标。

然而, 传统的 TBARS 化学测定方法虽然准确可靠, 但其需要对样品进行前处理, 操作过程复杂、耗时, 且具有破坏性[4], 不适用于大规模的品质监控场景。可见近红外(visible-near infrared, Vis-NIR)光谱技术[5]以其快速、无损和可集成化等显著优势, 在农产品和食品品质检测中展现出巨大潜力[6]。通过分析样品在可见和近红外波段的吸收以及反射特征, 能够获取与样品内部化学成分(如水分、脂肪、蛋白质、肌红蛋白等)密切相关的信息[7], 从而间接预测其品质指标。高维光谱数据往往存在大量冗余信息、变量间强共线性以及环境噪声和仪器漂移等干扰[8]。这些问题不仅增加了模型训练的复杂性和计算负担, 也可能导致模型过拟合和预测性能下降。因此, 对光谱数据进行有效的数据预处理以消除无关噪声和散射效应, 以及在高维数据中及逆行特征选择以筛选出最优的特征波长[9], 对于构建稳定且精确的预测模型至关重要。

近年来, 人工智能技术发展迅猛。在此背景下的深度学习模型在光谱分析领域正受到越来越多的关注。ZHAO 等[10]利用 Vis-NIR 光谱技术, 将机器学习与深度学习相结合建立了卷积神经网络-支持向量回归(convolutional neural network-support vector regression, CNN-SVR)的融合深度学习模型, 实现了猪肉贮藏过程中总挥发性盐基氮含量(total volatile base nitrogen, TVB-N)和 pH 的高水准预测。在这些深度学习模型中, 长短期记忆(long short-term memory, LSTM)网络[11], 在光谱分析领域正受到越来越多的关注。LSTM 在处理序列数据和挖掘复杂非线性关系方面具有卓越的能力, 能够有效应对复杂数据的非线性特征, 通过多层网络结构实现对数据模式的深层次学习, 提升预测模型的精度、泛化能力和训练效率, 正逐步成为光谱分析领域的研究热点。

然而, 深度学习模型固有的“黑箱”特性, 使得其预测结果的依据和决策逻辑难以被人类理解, 这在对模型透明度要求较高的食品安全和品质控制领域构成了一大障碍。在此背景下, 可解释人工智能(explainable artificial intelligence, XAI)技术应运而生, 其中, 以沙普利可加性解释(Shapley additive explanations, SHAP)[12]为代表的工具, 为理解复杂深度学习模型如何做出决策提供了有效的途径[13], 有助于深入挖掘光谱特征与品质指标之间的内在联系[14], 从而增加模型的透明度和可信度, 加速其在实际应用中的落地。然而, 现有研究在利用深度学习模型进行光谱分析时, 仍主要侧重于预测精度的提升, 而对模型内部决策机制及其物理或化学意义的解释相对不足。尤其是

在肉类脂质氧化及 TBARS 含量预测领域，关于关键光谱特征与脂质氧化反应之间内在联系的系统分析仍较为有限。此外，不同光谱预处理、特征波长选择方法以及深度学习模型与传统机器学习模型在同一数据体系下的综合比较研究仍然匮乏，这在一定程度上限制了模型结果的可解释性和实际应用价值。

基于上述背景，本研究旨在探索深度学习与可解释人工智能(rtificial intelligence, AI)驱动 Vis-NIR 光谱无损预测牛肉 TBARS 的可行性。通过系统比较不同光谱数据预处理、特征波长选择和深度学习如 LSTM、传统机器学习如偏最小二乘回归(partial least squares regression, PLSR)和极端梯度提升(extreme gradient boosting, XGBoost)建模策略，筛选出最优的预测模型。同时，本研究将运用 SHAP 方法，深入揭示关键光谱特征对 TBARS 预测模型的影响机制，以期为牛肉品质的快速、无损、精准且可解释的监控提供理论依据和技术支持。

1 材料与方法

1.1 材料与试剂

牛通脊，市购。

三氯乙酸(分析纯，国药集团化学试剂有限公司)；硫代巴比妥酸(分析纯，上海阿拉丁生化科技股份有限公司)；乙二胺四乙酸(分析纯，上海麦克林生化科技有限公司)。

1.2 仪器与设备

SpectraMax M2 多功能酶标仪[美谷分子仪器(上海)有限公司]；Lambda 950 光谱仪(美国珀金埃尔默公司)；AX504 电子分析天平(精度 0.1 mg，上海天平仪器厂)；ST 16R 高速台式冷冻离心机(上海浦东物理光学仪器厂)。

1.3 方 法

1.3.1 样品准备

实验所用的样品是从当地屠宰场购买的当天宰杀的牛通脊肉，并在 30 min 内放入冰盒内被迅速运送到实验室。在实验室中，将牛通脊肉在无菌台上切成均匀的 6 cm×6 cm×2 cm 片，共取 420 份样品，并去除表面脂肪和筋膜，以保证样品表面均匀一致，并且采用自封聚乙烯袋密封包装。用于预测牛肉 TBARS 含量的数据集包含了牛肉样品在不同贮藏条件的变化。240 个样品的贮藏条件为 4 ℃避光贮藏，分别在贮藏 0、1、3、5、7、8、9、10 d 取样，每次检测 30 个样品；180 个样品的贮藏条件为室温 25 ℃避光贮藏，分别在贮藏 0、1、2、3 d 取样，每次检测 45 个样品。每次取样扫描 Vis-NIR 光谱，测量 TBARS 含量。

1.3.2 光谱采集

牛肉样品的光谱数据采用 Lambda 950 光谱仪在反射模式下采集。仪器的光源为氘和钨卤素光源，光谱范围为 400~1800 nm，分辨率为 4 nm，共包含 351 个变量。牛肉样品从环境培养箱中取出后，在 25 ℃左右的实验室中缓和 30 min。在牛肉表面随机选取 3 个点进行扫描，得到每个样品的 Vis-NIR 光谱，取平均光谱进行进一步分析。

1.3.3 TBARS 含量测定

TBARS 含量通过分光光度法测定。2 g 牛肉糜与 10 mL 含 0.1% (m/V)乙二胺四乙酸的 7.5% (m/V)三氯乙酸溶液混合均质化后，离心(8000 g, 10 min)分离上清液。取 5 mL 上清液与 5 mL 0.02 mol/L 硫代巴比妥酸溶液反应，于沸水浴中加热 40 min 并冷却。反应产物的吸光度在 532 nm 处测定，TBARS 随后依据公式(1)计算，以单位质量的 MDA 当量毫克数(mg/kg)表示。每个样品 3 个重复，取平均值。

$$\text{TBARS/(mg/kg)}=\frac{A_{532\ \text{nm}}}{m}\times 9.48 \tag{1}$$

式中：$A_{532\ \text{nm}}$ 为样品在 532 nm 测得的吸光度；m 为被测样品的质量，g；9.48 为实验方法综合换算常数，是由摩尔吸光系数、MDA 摩尔质量、提取液定容体积、比色皿光程及单位换算系数等共同推导得出的固定系数，用于将吸光度值直接换算为样品中丙二醛的含量(mg/kg)，该常数仅适用于本实验所采用的 TBARS 测定方法体系。

1.3.4 样本集划分

本研究对 420 份牛肉样品的光谱特征数据进行建模。采用基于联合 x–y 距离的样品集划分算法(sample set partitioning based on joint X–Y distances, SPXY)将样本集按 3:1 的比例划分为训练集和测试集，分别为 315 个样本和 105 个样本。

1.3.5 数据预处理

由于环境变化、电子噪声和样品散射的影响，原始光谱信号会受到干扰。因此，为了还原光谱的真实性和准确性，对光谱进行预处理是非常必要的。本研究采用乘法散射校正(multiplicative scattering correction, MSC)[15–16]、标准正态变换(standard normal variation, SNV)[17]、卷积平滑(savitzky-golay convolutional smoothing, SG)、卷积平滑一阶导数(savitzky-golay first-order derivative, SG+1D)和卷积平滑二阶导数(savitzky-golay second-order derivative, SG+2D)的预处理方法，对原始光谱进行优化，提高后续建模的性能。其中，SG、SG+1D 以及 SG+2D 预处理过程中窗口宽度均为 5，多项式阶数均为 3。通过合理选择和实施这些预处理方法，可以更准确地从光谱中提取相关信息。

1.3.6 特征选择

Vis-NIR 光谱数据与样品成分或性质之间通常存在复杂的非线性关系。若采用线性建模方法处理这些非线性信息，可能导致模型预测能力下降。此外，光谱波长变量间存在的高度线性相关性，不仅造成光谱信息冗余，增加了模型的计算复杂度，还可能因共线性问题而降低预测精度和模型稳定性。鉴于此，进行特征波长筛选显得尤为必要。

本研究利用竞争性自适应加权抽样法(competitive adaptive reweighted sampling, CARS)[18]、无信息变量消除(uninformative variable elimination, UVE)[19]和连续投影算法(successive projections algorithm, SPA)[20]这 3 种特征波长筛选算法建立模型。

1.3.7　模型建立

为实现牛肉 TBARS 含量的定量预测, 分别采用 PLSR、XGBoost 以及 LSTM 对处理后的 Vis-NIR 光谱数据进行建模。PLSR 作为经典的传统机器学习线性多变量回归方法[21], 用于建立光谱数据与 TBARS 之间的基线预测模型; XGBoost[22]和 LSTM[23−24]分别作为代表性的机器学习模型和深度学习模型, 用于捕捉光谱数据与 TBARS 之间的非线性关系。其中, PLSR 模型使用五折交叉验证确定最佳潜变量数; XGBoost 模型中决策树数量为 500, 最大树深为 3, 学习率为 0.01; LSTM 采用 2 层结构, 隐藏单元数为 32, 模型训练轮数为 1000, 损失函数为均方误差(mean squared error, MSE), 优化器为自适应矩估计(adaptive moment estimation, Adam), 学习率为 0.0005。各模型的关键超参数通过经验调优确定, 未特别说明的参数均采用默认设置。所有模型均基于相同的数据集和特征输入进行训练与测试, 以保证不同建模方法之间的可比性。

1.3.8　性能评估

在本研究中, 采用训练集的决定系数(R-square of correction set, R^2_C)、测试集的决定系数(R-square of prediction set, R^2_P)、训练均方根误差(root mean square error of correction set, RMSEC)和测试均方根误差(root mean square error of prediction set, RMSEP)以及相对预测偏差(residual prediction deviation, RPD)来评价模型的性能。一般来说, R^2 越接近于 1, RMSE 越小, RPD 越大, 说明模型预测效果越好。

1.3.9　模型的可解释分析

SHAP 值根植于合作博弈论, 通过评估所有特征相互作用并计算每个特征对预测的边际贡献, 为量化预测模型中的特征重要性提供了一个严格的框架[25]。SHAP 通过特征重要性图系统地对输入变量进行排序——给有影响力的特征分配更高的权重, 同时淡化不那么重要的特征。与传统指标不同, SHAP 的特点是结合了特征交互效应和方向影响, 从而能够精确解释深度神经网络等复杂模型[26]。SHAP 的理论基础是沙普利(Shapley)值, 每个特征对预测模型输出值的贡献度可通过对应的 Shapley 值来体现, Shapley 值可通过公式(2)来计算:

$$\Phi_j(f,x)=\sum \frac{|S|!(|F|-|S|-1)!}{|F|!}[f_x(S\cup\{j\})-f_x(S)] \quad (2)$$

式中: $\Phi_j(f,x)$为特征 j 对于给定样本 x 的预测 $f(x)$的贡献值; f 为被解释的机器学习模型; F 为包含了模型中所有输入特征的集合; S 为特征 j 的所有子集; !为阶乘运算符号。

在本研究中, 使用基于 Python 的 SHAP 库(版本 0.50.0)中的 Deep Explainer 工具计算每个特征的 Shapley 值, 并通过蜂巢图和条形图进行可视化, 以显示在使用选定的特征波长预测模型结果时每个特征的重要性。

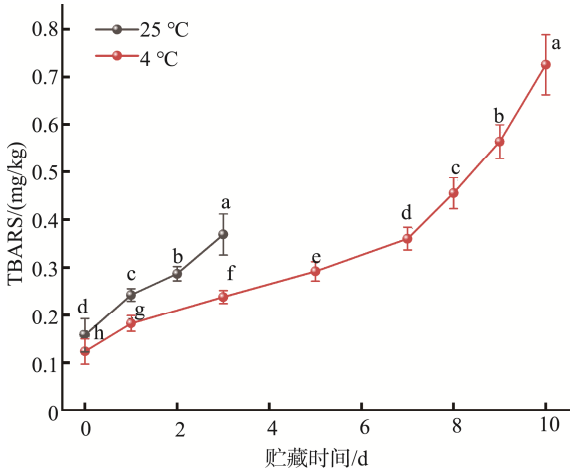
1.4　数据处理

所有代码都是在 PyCharm 集成开发环境中使用 Python 编写和测试的。SPSS 26 软件进行方差分析, $P<0.05$ 表示差异显著。使用 Origin 2024 软件绘图。TBARS 测量每个样品 3 个重复, 取平均值。

2　结果与分析

2.1　贮藏期间牛肉样品 TBARS 含量变化分析

由图 1 可知, 在 4 ℃和 25 ℃贮藏条件下, 牛肉样品的 TBARS 含量都随贮藏时间的延长而上升, 并且 25 ℃贮藏的样品比 4 ℃样品的 TBARS 增加速率更快。这是由于在贮藏过程中, 牛肉样品中的脂质被氧化, 发生酸败, 产生醛、酮、呋喃和 MDA 等化合物。并且温度越高, 脂质氧化速率越快。可以看出, 在 4 ℃冷藏条件下贮藏第 10 d 时牛肉样本的 TBARS 含量超过 0.7 mg/kg, 即将接近限值 1 mg/kg。结果表明肉质的氧化程度明显增加, 标志着牛肉的品质进一步下降。牛肉样本的TBARS含量的测量值按照SPXY方法3:1的比例划分为训练集和测试集的具体数据如表1所示, 其中训练集的范围大概覆盖了测试集中参考测量值的范围, 这意味着训练集和测试集中测量样本的参考数据的分布几乎相等, 避免了两组分布的偏差, 数据分布合理。



注: 同组数据标不同字母表示差异显著($P<0.05$)。

图 1　不同贮藏时间下牛肉样品 TBARS 含量的变化

Fig.1　Changes in TBARS content of beef samples during different storage periods

2.2　光谱分析

图 2A 展示了样品的原始 Vis-NIR 光谱。可以看出, 所有样品的光谱变化趋势基本一致, 尽管在反射强度上存在

一定差异。这些强度差异归因于牛肉在贮藏过程中化学成分的变化，这会影响样品的反射率。这一观测结果进一步证明，光谱包含了样品的大量信息。415 nm 和 470 nm 处的吸收波长主要与肌红蛋白相关联[27]。在 580 nm 附近的可见光区域存在一个明显的吸收带，主要对应于氧合肌红蛋白、肌红蛋白以及脱氧肌红蛋白的吸收特性[28]。在近红外区域，750 nm 附近的吸收被认为与 NH_3^+基团的 3 重态跃迁和 O-H 的第 3 泛音有关；而约 970 nm 处的吸收带则主要归因于水分子的吸收特性，对水分含量变化具有较高的敏感性。这些结果表明，脂质氧化和相关的生化变化显著影响 Vis-NIR 光谱特征。

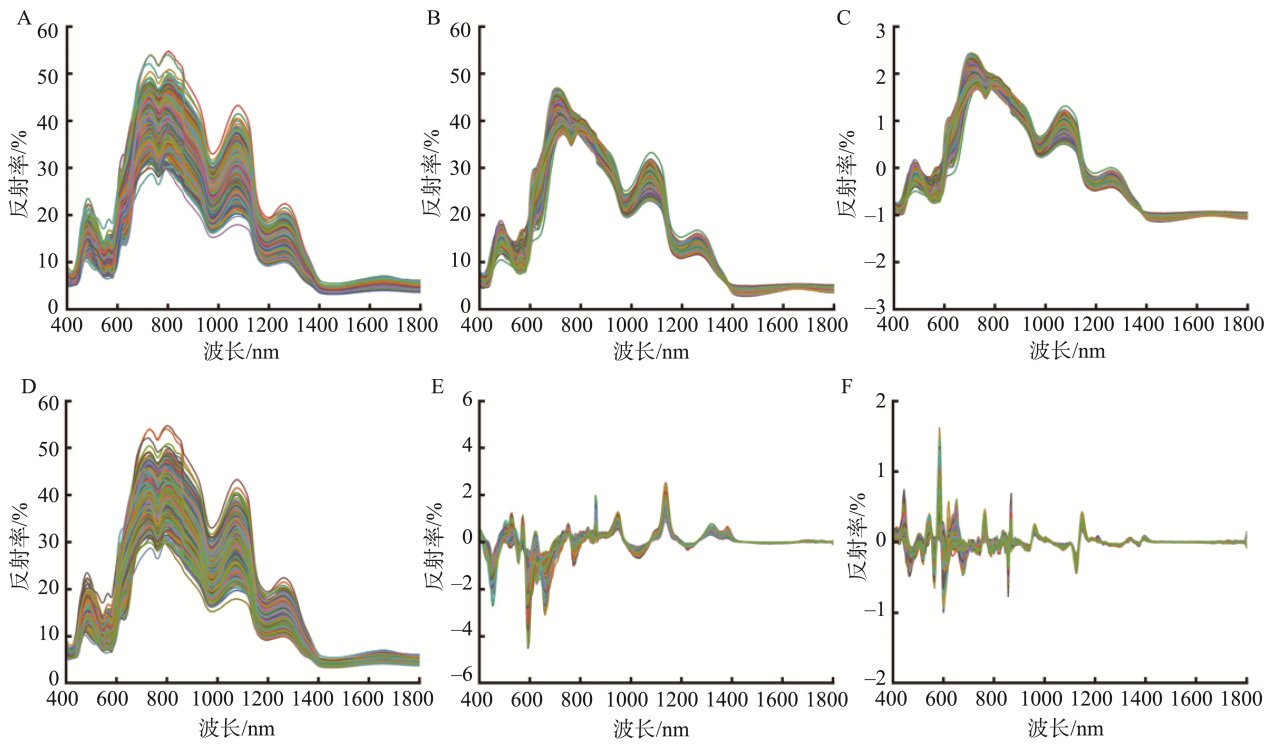
2.3 预处理方法的建模评价与分析

为了增强光谱数据的信息量，同时减轻噪声和误差对建模精度的不利影响[29]，使用几种初步方法对样品的光谱数据进行预处理。它可以减少实验误差和无关信息，从而提高光谱数据建模过程的准确性和稳定性。不同方法处理后的光谱数据曲线如图 2B~F 所示。利用 PLSR、XGBoost 和 LSTM 3 种模型在整个波长范围内进行建模。根据模型结果确定最优预处理方法，如图 3 所示。R^2_C 和 R^2_P 越高表明模型拟合越好，RMSE 值越低意味着预测值与实际值差异越小。采用 PLSR 和 LSTM 建模算法，得到最优预处理方法为 SNV。一般来说，SNV 改进了原始光谱的预测，可能是因为 SNV 消除了散射对不同样品粒度和不均匀分布光谱的影响。利用 XGBoost 建模算法得到的最优预处理方法为 SG。SG 平滑有效地消除了频谱中的随机噪声(如仪器振动和环境干扰)，同时保留了频谱特征峰的形状和位置，避免了过度平滑导致关键信息的丢失，从而达到最佳性能。

表 1 牛肉样品中 TBARS 含量的统计

Table 1 TBARS content statistics in beef samples

样本集	样本量	最大值/(mg/kg)	最小值/(mg/kg)	平均值/(mg/kg)	标准差/(mg/kg)
总数据集	420	0.920	0.066	0.323	0.164
训练集	315	0.920	0.066	0.327	0.167
测试集	105	0.782	0.100	0.311	0.153



注：A. 原始光谱；B. MSC；C. SNV；D. SG；E. SG+1D；F. SG+2D。

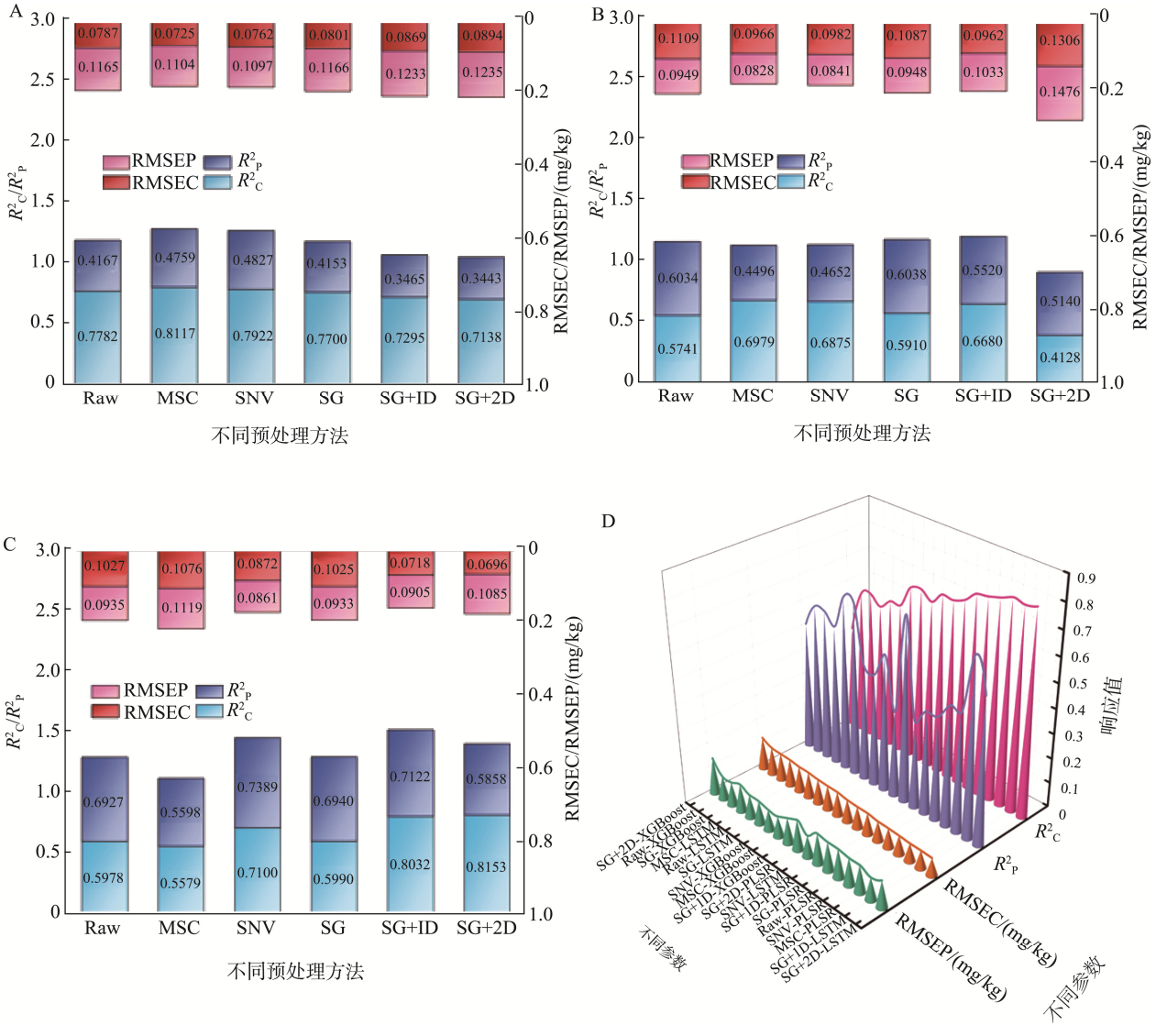
图 2 牛肉样品的原始 Vis-NIR 光谱和经预处理的 Vis-NIR 光谱

Fig.2 Raw and preprocessed Vis-NIR spectra of beef samples

在图 3A 中，SNV-PLSR 模型的 R^2_C、R^2_P、RMSEC 和 RMSEP 分别为 0.7922、0.4827、0.0762 mg/kg 和 0.1097 mg/kg。在图 3B 中，SNV-XGBoost 模型的 R^2_C、R^2_P、RMSEC 和 RMSEP 分别为 0.6875、0.4652、0.0982 和 0.0841 mg/kg。图 3C 中，SNV 预处理后的光谱 LSTM 模型的 R^2_C、R^2_P、RMSEC 和 RMSEP 分别为 0.7100、0.7389、0.0872 和 0.0861 mg/kg。因此，选择 SNV 作为后续分析的最佳预处理方法。

2.4　特征波长选择

Vis-NIR 光谱数据的高维性和强相邻变量相关性[30]给模型训练带来了重大挑战，包括复杂且耗时的过程。因此，进行有效的波长选择对光谱数据分析和模型简化至关重要。预处理后，采用 3 种不同的算法进行特征波长的选择，提高计算效率。图 4 显示了不同特征选择方法的最优波长分布。结果表明，CARS、UVE 和 SPA 分别提取了 49 个、49 个和 50 个特征波长。在构建精确的回归模型的过程中，特征波长的选择起着至关重要的作用。合理选择最有效的特征波长不仅可以减少数据维数，还可以最大限度地减少噪声干扰。此外，它有助于提高模型的可解释性，有助于发现潜在的相关性和模式，最终提高预测精度，同时降低过拟合的风险。



注：A. PLSR；B. XGBoost；C. LSTM；D. 基于不同预处理方法的建模评价。

图 3　不同预处理方法建模比较

Fig.3　Modeling comparison of different preprocessing methods

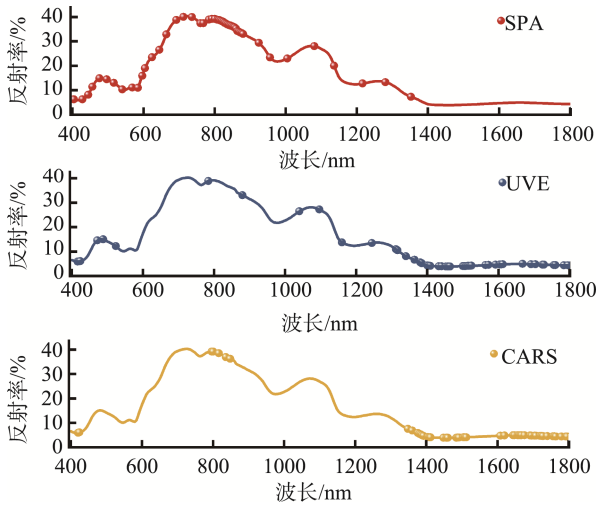


图 4 分布在整个平均光谱曲线上的特征波长提取结果

Fig.4 Results of feature wavelengths extraction distributed across the entire mean spectral curve

2.5 基于特征波长的模型分析

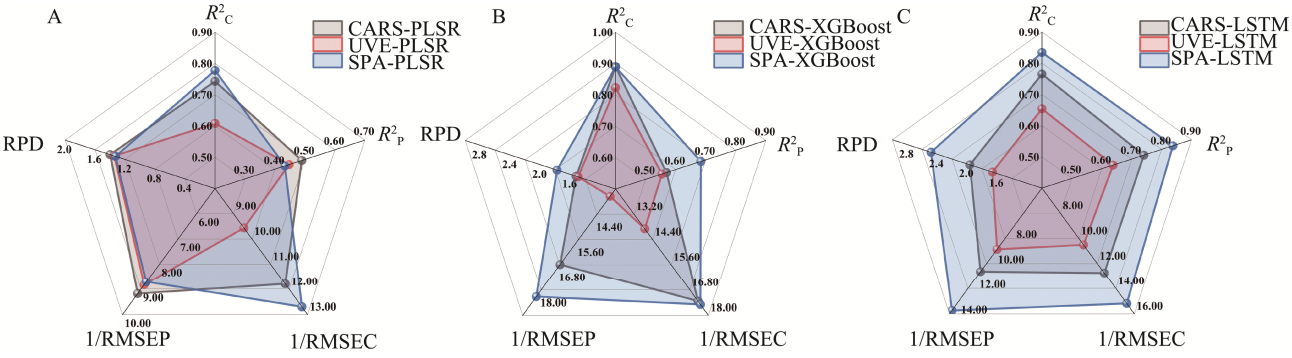
图 5 为基于特征波长的 PLSR(图 5A)、XGBoost(图 5B)和 LSTM(图 5C)定量预测模型的性能参数(R^2_C、R^2_P、1/RMSEC、1/RMSEP 和 RPD)的雷达图。对于传统机器学习模型，最优 PLSR 模型为 SNV-CARS-PLSR (R^2_C 为 0.7426, RMSEC 为 0.0848 mg/kg, R^2_P 为 0.4897, RMSEP 为 0.1090 mg/kg, RPD 为 1.3998)。最优 XGBoost 模型为 SNV-SPA-XGBoost (R^2_C为 0.8891, RMSEC 为 0.0572 mg/kg, R^2_P 为 0.6839, RMSEP 为 0.0552 mg/kg, RPD 为 1.7785)。对于深度学习模型，最优 LSTM 模型为 SNV-SPA-LSTM (R^2_C 为 0.8344, RMSEC 为 0.0659 mg/kg, R^2_P 为 0.8378, RMSEP 为 0.0679 mg/kg, RPD 为 2.4833)，表现出最优异的性能。对于 LSTM 深度学习算法，原始数据 SNV 预处理和 SPA 特征选择算法是最优算法。深度学习通过深度神经网络实现多层特征选择和端到端优化，具有较强的泛化能力和多模态数据处理优势，在复杂任务中得到广泛应用，优势显著。对于传统的机器学习，尽管 CARS 和 SPA 均分别对 PLSR 和 XGBoost 建模结果有一定提升，但它对手动特征选择和线性假设的依赖限制了它的性能上限。相比之下，深度学习模型在自动特征选择、复杂模式建模和小样本适应性方面明显优于传统机器学习。与 PLSR、XGBoost 模型相比，SNV-SPA-LSTM 模型的 R^2_P 准确率分别提高了 34.81%和 15.39%。SNV-SPA-LSTM 模型是 3 种建模方法中最好的回归模型，在牛肉的 TBARS 含量定量检测中表现优异。

2.6 最优模型的可解释 AI 分析

尽管 SPA 已成功从高维原始光谱中筛选出最具信息量的关键波长子集，但 LSTM 深度学习模型内部复杂的非线性映射关系仍需进一步阐明。为了深入理解最优 SNV-SPA-LSTM 模型对牛肉 TBARS 含量的预测机制，本研究进一步利用可解释 AI 中的 SHAP 框架对模型进行解释。SHAP 分析能够量化这些经优选后的特征波长对模型最终预测结果的贡献度，从而揭示模型如何利用这些精炼信息做出决策，并验证关键波长选择的有效性。

如图 6 所示，基于 SNV-SPA-LSTM 最优定量模型的特征波长重要性分析的 SHAP 值通过蜂窝图和柱状图两种可视化方法呈现了前 20 个特征波长的显著性分布。在条形图中，每个条形的长度代表了特征波长的重要性，量化依据是平均绝对 SHAP 值。蜂窝图将特征按重要性垂直排列，同时显示特征分布的密度和趋势特征。值得注意的是，SHAP 值的数值范围与特征与模型输出的相关性呈正相关。SHAP 值越大，该特征对模型输出的影响越显著。

1080、956、848、864、736 nm 处的特征对 TBARS 含量的预测影响最大。1080 nm 和 956 nm 这两个特征波长靠近 980 nm 和 1180 nm 这两个吸收峰，与脱氧肌红蛋白和氧合肌红蛋白密切相关。956 nm 处的 SHAP 值表明，该波



注: A. PLSR; B. XGBoost; C. LSTM。

图 5 基于特征波长预测牛肉 TBARS 含量的回归模型性能

Fig.5 Performance of regression models for predicting TBARS content in beef based on feature wavelengths

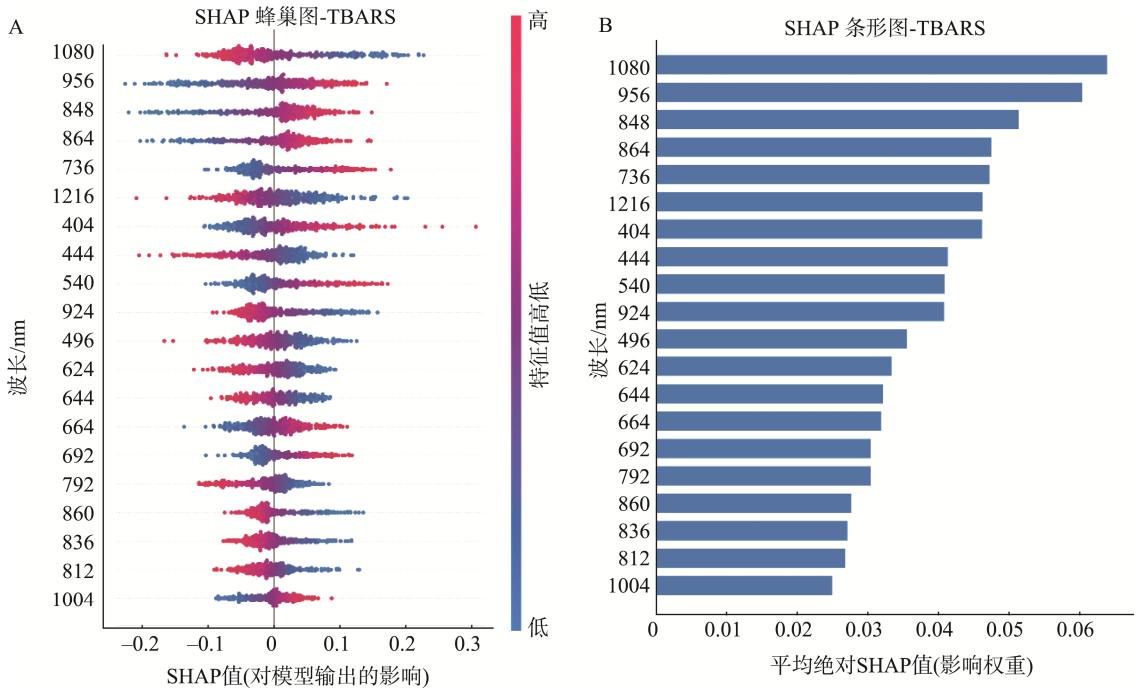


图 6　SNV-SPA-LSTM 模型中特征波长重要性 SHAP 分析

Fig.6　SHAP analysis of feature wavelength importance in the SNV-SPA-LSTM model

长对模型的预测呈现正向作用，较高的反射率值与较高的 TBARS 水平的预测有关，并且主要影响 TBARS 含量的是该波长所处的 970 nm 附近的吸收带，该吸收带主要与水分子中的 O-H 键有关。O-H 键可以与脂质自由基或脂质氧化过程中产生的活性氧相互作用，影响整个氧化过程。这些波长间的关联模式存在差异，对 TBARS 含量预测的影响方式也各不相同，这一点可由 SHAP 分析结果证实。848 nm 和 864 nm 位于 830~900 nm 之间，对应脂肪中 CH_2/CH_3(甘油三酯、磷脂等)，也包含蛋白侧链的 C–H。这些发现充分突出了预测模型中所选变量的重要性。

3　结　论

本研究利用 Vis-NIR 光谱系统在 400~1800 nm 波长范围内对牛肉贮藏过程中的脂质氧化程度进行评价。与 PLSR 和 XGBoost 模型相比，LSTM 模型对牛肉中 TBARS 含量的预测精度更高。为了降低原始光谱矩阵的维数，采用 CARS、UVE 和 SPA 算法分别从 351 个光谱变量中提取出 49、49 和 50 个关键变量。SPA 算法确定了能提供最高预测精度的最佳波长。简化后的 SNV-SPA-LSTM 模型对牛肉中 TBARS 含量的评估结果良好(R^2_C 为 0.8344，RMSEC 为 0.0659 mg/Kg，R^2_P 为 0.8378，RMSEP 为 0.0679 mg/Kg，RPD 为 2.4833)。尤为重要的是，本研究进一步引入 SHAP 可解释人工智能框架，对最优 SNV-SPA-LSTM 模型进行了深入解读，揭示了 1080、956、848、864 和 736 nm 等关键波长对 TBARS 预测的贡献机制，从而增强了模型的透明度和可信度。综上所述，本研究提出了一种深度学习与可解释 AI 驱动的可见近红外光谱无损预测牛肉 TBARS 的高效策略，为牛肉品质的智能检测、货架期评估和食品安全保障提供了一定的理论支撑和技术借鉴。

参考文献

[1] 李泽坤，肖宇，焦阳，等. 脂质和蛋白质氧化对肉品质的影响[J]. 中国食品学报, 2024, 24(7): 438–449.
LI ZK, XIAO Y, JIAO Y, *et al*. Effects of lipid and protein oxidation on meat quality [J]. Journal of Chinese Institute of Food Science and Technology, 2024, 24(7): 438–449.

[2] 陈紫婷，孙智达. 光氧化与氧气对牛肉保鲜过程中品质以及蛋白理化性质的影响[J]. 肉类研究, 2023, 37(10): 30–36.
CHEN ZT, SUN ZD. Effects of photooxidation and oxygen on beef quality and physicochemical properties of proteinduring storage [J]. Meat Research, 2023, 37(10): 30–36.

[3] GUO M, LIN H, FENG H, *et al*. Deep learning algorithm-assisted non-destructive detection of TBARS values of salmon flesh using multi-modal molecular spectra fusion [J]. Food Chemistry, 2025, 492: 145649.

[4] YI W, ZHAO X, YUN X, *et al*. Automated and explainable machine learning for monitoring lipid and protein oxidative damage in mutton using hyperspectral imaging [J]. Food Research International, 2025, 203. DOI: 10.1016/j.foodres.2025.115905

[5] 侯智轩，张朵朵，姚平波，等. 肉及肉制品新鲜度检测技术研究进展[J]. 肉类研究, 2025, 39(1): 64–71.
HOU ZX, ZHANG DD, YAO PB, *et al*. A review of technologies for freshness detection of meat and meatproducts [J]. Meat Research, 2025, 39(1): 64–71.

[6] 张旭，谢卓君，覃子权，等. 小型化可见/近红外光谱仪在叶片和果实

无损检测中的研究进展[J]. 光谱学与光谱分析, 2025, 45(10): 2720–2729.
ZHANG X, XIE ZJ, QIN ZQ, *et al*. Research progress of mini aturized Vis/NIR spectrometers in leaf and fruit nondestructive detection [J]. Spectroscopy and Spectral Analysis, 2025, 45(10): 2720–2729.

[7] 任智磊, 赵霄霄, 冯景, 等. 基于近红外光谱结合网格搜索-随机森林-自适应提升算法无损检测牛肉新鲜度[J]. 肉类研究, 2025, 39(11): 1–8.
REN ZL, ZHAO XX, FENG J, *et al*. Non-destructive detection of beef freshness using near infrared spectroscopy combined with grid search-random forest-adaptive boosting algorithm [J]. Meat Research, 2025, 39(11): 1–8.

[8] CHEN S, LIANG J, HUANG M, *et al*. Rapid and non-destructive prediction of pear firmness based on visible-near-infrared spectroscopy with a new method of spectra generating and wavelength selecting [J]. Infrared Physics & Technology, 2025, 149. DOI: 10.1016/j.infrared.2025.105903

[9] 罗毅智, 唐书奇, 金青婷, 等. 高光谱成像与混合波长选择的宽度学习系统预测猪肉 TVB-N 含量和 pH 值[J]. 食品科学, 2025, 46(16): 345–352.
LUO YZ, TANG SQ, JIN QT, *et al*. Prediction of pork TVB-N content and pH using broad learning system based onhyperspectral imaging with hybrid wavelength selection [J]. Food Science, 2025, 46(16): 345–352.

[10] ZHAO X, NING W, CHEN R, *et al*. Rapid non-destructive detection of pork freshness using visible-near infrared spectroscopy based on convolutional neural network hybrid models [J]. Journal of Food Composition and Analysis, 2025, 140. DOI: 10.1016/j.jfca.2025.107199

[11] TIRADO-KULIEVA VA, MECHATO J, SEMINARIO SRS, *et al*. Quantification of cumin powder adulteration with rice by-products using NIR spectroscopy combined with chemometric and machine learning models [J]. Journal of Food Composition and Analysis, 2025, 148(3): 108471.

[12] BACHMANN S. Efficient XAI: A low-cost data reduction approach to SHAP interpretability [J]. Journal of Artificial Intelligence Research, 2025, 83. DOI: 10.1613/jair.1.18325

[13] ANGELOV P, SOARES E, JIANG R, *et al*. Explainable artificial intelligence: An analytical review [J]. Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery, 2021, 11(5). DOI: 10.1002/widm.1424

[14] 颜玉洁, 俞玥, 孔天宇, 等. 深度学习驱动的机器视觉技术用于果蔬品质的智能感知: 进展、挑战、展望[J]. 食品科学, 2025, 46(22): 23–39.
YAN YJ, YU Y, KONG TY, *et al*. Deep learning-based machine vision for intelligent perception of fruit andvegetable quality: Progress, challenges, and prospects [J]. Food Science, 2025, 46(22): 23–39.

[15] CAI L, LI J, ZHANG H, *et al.* Determination of the SSC in oranges using Vis-NIR full transmittance hyperspectral imaging and spectral visual coding: A practical solution to the scattering problem of inhomogeneous mixtures [J]. Food Chemistry, 2025, 474: 143239.

[16] 李改英, 李品, 黄英杰, 等. 近红外光谱技术快速鉴别河南省地方品种羊肉[J]. 食品安全质量检测学报, 2025, 16(14): 89–96.
LI GY, LI P, HUANG YJ, *et al*. Rapid identification of local variety lamb in Henan Province by near infrared spectroscopy [J]. Journal of Food Safety & Quality, 2025, 16(14): 89–96.

[17] 赵世茹, 黄浦杰, 李宪秀, 等. 基于中红外光谱技术的安吉金花茶发酵过程中关键成分含量快速检测[J]. 食品科学, 2026, 47(1): 177–184.
ZHAO SR, HUANG PJ, LI XX, *et al*. Rapid determination of different components in *Eurotium cristatum* fermentation of AnjiGolden flower tea based on mid-infrared spectroscopy [J]. Food Science, 2026, 47(1): 177–184.

[18] GAO S, JI C, CUI J, *et al.* Competitive adaptive reweighted sampling algorithm identifies HIF-1α-regulated protein markers governing early energy metabolism in post-slaughter Tan sheep meat [J]. Food Quality and Safety, 2025, 9: fyaf039.

[19] WANG Z, CHEN J, ZHANG J, *et al*. Assessing canopy nitrogen and carbon content in maize by canopy spectral reflectance and uninformative variable elimination [J]. Crop Journal, 2022, 10(5): 1224–1238.

[20] CANOVA L, VALLESE F, PISTONESI M, *et al*. An improved successive projections algorithm version to variable selection in multiple linear regression [J]. Analytica Chimica Acta, 2023, 1274: 341560.

[21] LOMARAT P, PHECHKRAJANG C, SUNGHAD P, *et al*. Raman spectroscopy coupled with the PLSR model: A rapid method for analyzing gamma-oryzanol content in rice bran oil [J]. Food Chemistry-X, 2024, 24. DOI: 10.1016/j.fochx.2024.101923

[22] 张志从, 崔东, 赵阳, 等. 基于无人机载高光谱成像结合机器学习算法的新疆野苹果冠层叶绿素含量估算[J]. 果树学报, 2026, 43(3): 673–685.
ZHANG ZC, CUI D, ZHAO Y, *et al*. Estimation of chlorophyll content in the canopy of *Malus siersii* (Ledeb.) M. Roem based on unmanned aerial vehicle hyperspectral imaging combined with machine learning algorithms [J]. Journal of Fruit Science, 2026, 43(3): 673–685.

[23] NUNEKPEKU X, ZHANG W, GAO J, *et al*. Gel strength prediction in ultrasonicated chicken mince: Fusing near-infrared and Raman spectroscopy coupled with deep learning LSTM algorithm [J]. Food Control, 2025, 168. DOI: 10.1016/j.foodcont.2024.110916.

[24] WU X, DU Z, MA R, *et al*. Qualitative and quantitative studies of phthalates in extra virgin olive oil (EVOO) by surface-enhanced Raman spectroscopy (SERS) combined with long short term memory (LSTM) neural network [J]. Food Chemistry, 2024, 433. DOI: 10.1016/j.foodchem.2023.137300

[25] FU R, XIAO D, BUCHAN A, *et al*. A parametric non-linear non-intrusive reduce-order model using deep transfer learning [J]. Computer Methods in Applied Mechanics and Engineering, 2025, 438. DOI: 10.1016/j.cma.2025.117807

[26] ZHANG C, WANG D, WANG L, *et al*. Cause-aware failure detection using an interpretable XGBoost for optical networks [J]. Optics Express, 2021, 29(20): 31974–31992.

[27] WAGA M, NAKADE K. Light wavelengths that induce oxidation of oxymyoglobin in meat [J]. Meat Science, 2025, 219. DOI: 10.1016/j.meatsci.2024.109664

[28] ZHENG X, CHEN L, LI X, *et al.* Non-destructive detection of meat quality based on multiple spectral dimension reduction methods by near-infrared spectroscopy [J]. Foods, 2023, 12(2): 300.

[29] LEE G, LEE K. Feature selection using distributions of orthogonal PLS regression vectors in spectral data [J]. BioData Mining, 2021, 14(1): 7.

[30] ZHANG S, ZHANG J, WANG X, *et al*. ELS2T: Efficient lightweight spectral-spatial transformer for hyperspectral image classification [J]. Ieee Transactions on Geoscience and Remote Sensing, 2023, 61: 1–16.

(责任编辑: 韩晓红 蔡世佳)