

安 婧, 韩俊英. 基于高光谱成像的苹果品质指标的无损检测方法[J]. 江苏农业学报, 2026, 42(5): 973-981.
doi: 10.3969/j.issn.1000-4440.2026.05.011

基于高光谱成像的苹果品质指标的无损检测方法

安 婧, 韩俊英

(甘肃农业大学信息科学技术学院, 甘肃 兰州 730070)

摘要: 为实现苹果内部品质的快速无损检测,本研究基于380~1 018 nm 高光谱成像技术构建了融合多种预处理、特征选择与建模算法的预测框架。采用 Savitzky-Golay (SG) 平滑、离散小波变换(DWT)、多元散射校正(MSC)、标准正态变量变换(SNV)、自适应加权平滑(AWS)、鲁棒性主成分分析(RPCA)和基线校正(BC)预处理光谱数据,结合连续投影算法(SPA)、竞争性自适应重加权采样法(CARS)和基于 Bootstrap 的优化光谱选择方法(BOSS)提取特征波段,构建偏最小二乘回归模型、支持向量回归模型与岭回归模型。结果表明,不同指标的最优建模路径呈现显著差异。对于维生素 C 含量,SG-SPA-SVR 组合建模预测效果最佳。对于可滴定酸含量,BC-原始光谱-Ridge 组合建模效果最佳。对于淀粉含量,BC-原始光谱-SVR 组合建模预测效果最佳。对于可溶性固形物含量,MSC-原始光谱-Ridge 组合建模预测效果最佳。本研究构建的多策略建模体系可为苹果品质无损检测及便携式高光谱系统开发提供理论支持与技术参考。

关键词: 高光谱成像; 苹果品质检测; 机器学习; 图像处理

中图分类号: TS255.7;O657.3 **文献标识码:** A **文章编号:** 1000-4440(2026)05-0973-09

A non-destructive detection method for apple quality indices based on hyperspectral imaging

AN Jing, HAN Junying

(College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China)

Abstract: To achieve rapid and non-destructive detection of internal quality in apples, this study established a prediction framework integrating multiple preprocessing, feature selection, and modeling algorithms based on hyperspectral imaging technology in the range of 380–1 018 nm. Spectral data were preprocessed using Savitzky-Golay (SG) smoothing, discrete wavelet transform (DWT), multiplicative scatter correction (MSC), standard normal variate transformation (SNV), adaptive weighted smoothing (AWS), robust principal component analysis (RPCA), and baseline correction (BC). Characteristic wavelengths were selected using the successive projections algorithm (SPA), competitive adaptive re-weighted sampling (CARS), and Bootstrap-based spectral selection (BOSS). Partial least squares regression, support vector regression, and ridge regression models were constructed. The results showed that the optimal modeling paths for different quality indices differed significantly. For vitamin C content, the combination of SG-SPA-SVR achieved the best prediction performance. For titratable acid content, the BC-raw spectrum-Ridge combination performed best. For starch content, the BC-raw spectrum-SVR combination showed the optimal prediction effect. For soluble solids content, the MSC-raw spec-

收稿日期:2025-07-19

基金项目:国家自然科学基金项目(32360437);甘肃省自然科学基金项目(25YFNA040)

作者简介:安 婧(2000-),女,甘肃天水人,硕士研究生,研究方向为农作物品质检测。(E-mail)2750214516@qq.com

通讯作者:韩俊英,(E-mail)hanjy@gsau.edu.cn

trum-Ridge combination yielded the best prediction results. The multi-strategy modeling system established in this study can provide theoretical support and technical reference for non-destructive detection of apple quality and the development of portable hyperspectral systems.

Key words: hyperspectral imaging; apple quality

detection; machine learning; image processing

苹果是全球范围内广泛种植与消费的重要水果之一,苹果品质评价指标不仅包括颜色、大小和形状等外观指标,还包括维生素 C 含量、可滴定酸含量、可溶性固形物含量和淀粉含量等内部理化指标^[1-2]。然而,传统的化学滴定、重量法和酶分析等方法操作复杂、检测周期长且具有破坏性,难以满足现代果品产业对高效、快速和无损检测的实际需求^[3]。近年来,高光谱成像技术(HSI)凭借其融合光谱分析与机器视觉的独特优势,在果蔬品质无损检测领域得到广泛应用^[4-5]。该技术可同步获取样品的的光谱信息与空间分布信息,显著提升检测结果的鲁棒性^[6]。已有研究将 HSI 与深度学习算法相结合,用于检测富士苹果可溶性固形物含量以及预测苹果总糖含量和硬度^[7-9]。

然而,高光谱成像技术在实际应用中仍面临若干挑战^[10]。首先,高光谱数据具有高维度和信息冗余特点,增加了数据处理与建模的难度^[11-12]。其次,光谱数据易受光照不均、杂散光和基线漂移等噪声干扰,这些干扰影响了建模准确性^[13]。为克服上述问题,研究者发展了多种数据预处理与特征提取方法,包括 Savitzky-Golay 平滑(SG)、移动平均平滑(MA)、归一化(NM)、多元散射校正(MSC)、基线校正、标准正态变换(SNV)、去趋势(DT)、一阶导数(1st Der)和二阶导数(2nd Der)等^[14-17]。此外,遗传算法(GA)、连续投影算法(SPA)和竞争性自适应重加权采样算法(CARS)等也被用于优选与目标变量高度相关的特征波长^[18]。

值得注意的是,目前高光谱成像技术多应用于单一指标的检测,忽略了不同营养成分之间的内在关联,难以全面衡量果实整体品质^[19-20]。本研究拟以苹果为研究对象,采集其高光谱反射率信息,系统比较 7 种光谱预处理方法、3 种特征波段选择算法及 3 种回归建模方法,筛选出预测维生素 C 含量、可滴定酸含量、可溶性固形物含量和淀粉含量最优组合模型,以期为苹果品质快速无损检测提供理论依据。

1 材料与方 法

1.1 苹果光谱数据采集

试验样品选取 5 个苹果品种,分别为成纪 1 号、瑞雪、瑞阳、烟富 3 号和静宁 1 号,于 2023 年 10 月中

旬采摘自中国甘肃省静宁县果树研究所。每个品种选取 12 个果实,共计 60 个苹果,要求果实成熟度基本一致、大小均匀,且表面无明显机械损伤、无病虫害及畸形。采摘后使用塑料泡沫进行保护,装箱后运送至实验室。对每个苹果采集 4 张高光谱图像,共获得 240 张高光谱图像。图像按照 7:3 划分为训练集和测试集,其中 168 张图像的光谱数据用于模型训练,72 张图像的光谱数据用于模型验证。

GaiaField 便携式高光谱系统(四川双利光谱成像技术有限公司产品)如图 1 所示,该系统主要由高光谱相机、4 个无阴影光源、可调节载物台及数据采集软件 SpecView 构成。其中,高光谱相机的光谱波长为 380~1 018 nm,光谱分辨率为 2.8 nm,4 个光源的照射角度均调整为 30°。图像采集前,先将高光谱相机预热 30 min。将相机镜头与载物台之间的距离调整为 40 cm,设置曝光时间为 49 ms,增益模式设为 2,实际扫描速度设为 18 mm/s,以确保图像不失真。整个采集过程在暗箱内进行,以避免杂乱光线的干扰。对每个苹果采集 4 次高光谱图像。每完成第一次采集后,将苹果围绕叶柄-花萼连线(即果实纵轴)旋转 90°,保持叶柄-花萼连线与载物台基本平行,并确保苹果中心与相机对齐。利用该系统采集了 380~1 018 nm 波长的 320 个光谱波段的苹果反射率数据。

为消除暗电流及光源不均匀性对成像结果的影响,需对原始高光谱图像(I_{raw})进行黑白校正。在与样品采集相同的条件下,分别采集标准白板图像(I_{white})和全黑图像(I_{dark})。校正后的图像(I_{cor})计算公式如下:

$$I_{\text{cor}} = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}} \quad (1)$$

利用 ENVI 5.3 软件从校正后的高光谱图像中提取光谱数据,在每个苹果图像上手动选取一个感兴趣区域(ROI),提取该区域内所有像素点的平均反射率,作为原始光谱数据。共获得 240 条光谱数据,每条光谱包含 320 个波段。

1.2 理化指标的测定方法

1.2.1 维生素 C 含量的测定 采用 2,6-二氯酚靛定法测定苹果中的维生素 C 含量。取 10 mL 抗坏血酸标准溶液,加入 30 mL 2% 草酸溶液作为提取

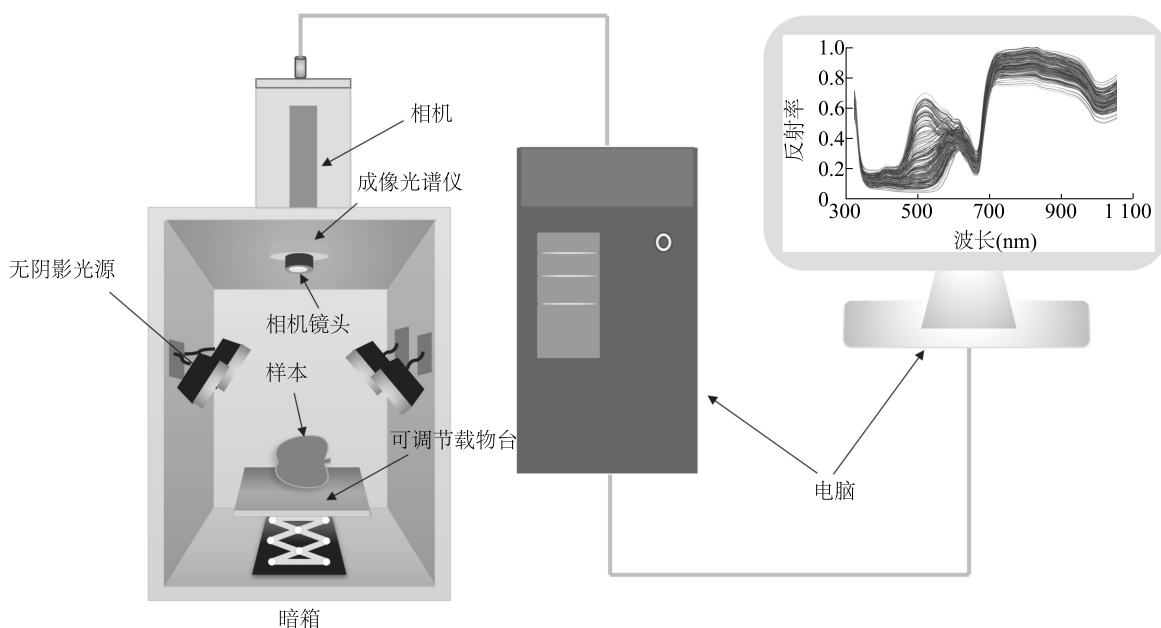


图1 高光谱系统组成

Fig.1 Components of hyperspectral system

剂。然后用2,6-二氯酚钠溶液进行滴定,至溶液呈粉红色并持续30 s不褪色即为滴定终点。记录消耗的染料体积以确定滴定度。苹果样品采用相同方法进行滴定,维生素C含量计算见公式(2):

$$\text{维生素 C 含量}(\text{mg/g}) = \frac{(v_3 - v_2) \times T \times v_1}{m \times v_s} \quad (2)$$

式中, v_2 和 v_3 分别为空白对照和样品滴定消耗的染料体积(mL); T 为滴定度(mg/mL); m 为样品重量(g); v_1 为提取液体积(mL); v_s 为滴定时样品溶液体积(mL)。

1.2.2 可滴定酸含量的测定方法 采用酸碱滴定法测定苹果中可滴定酸含量。称取250 g苹果可食用部分,加入等量水后置于高速组织捣碎机中匀浆2 min。取50 mL匀浆液,用水冲洗并转移至250 mL容量瓶中,于75~80℃水浴加热30 min,冷却后加水定容至刻度,摇匀后过滤。取50 mL滤液,加入酚酞指示剂,用氢氧化钠标准溶液进行滴定,至溶液呈微红且维持30 s不褪色为滴定终点。另外用蒸馏水进行滴定,所消耗的氢氧化钠溶液体积为空白对照。可滴定酸含量计算公式如下:

$$\text{可滴定酸含量}(\text{g/kg}) = \frac{c \times (v_1 - v_0) \times k \times F}{m} \times 1000 \quad (3)$$

式中, c 为氢氧化钠标准溶液浓度(mol/L); v_1 和 v_0 分别为样品和空白对照滴定消耗的氢氧化钠溶

液体积(mL); k 为酸的换算系数; F 为稀释倍数, m 为样品重量(g)。

1.2.3 可溶性固形物含量的测定方法 采用折光仪法测定苹果样品中可溶性固形物含量。称取250 g苹果可食用部分,切碎并充分混合后,置于高速组织捣碎机中匀浆。匀浆液经双层纱布过滤,收集滤液。取适量清液滴加于折光仪测量棱镜表面,待温度稳定后读取折光仪数值,即为SSC含量。

1.2.4 淀粉含量的测定方法 采用3,5-二硝基水杨酸(DNS)比色法测定苹果中的淀粉含量。在试管中加入葡萄糖溶液,再加入DNS试剂,摇匀后置于沸水中保温5 min,取出后自然冷却至室温。以葡萄糖含量为横坐标、吸光度为纵坐标绘制标准溶液的吸光度曲线,并计算标准线性回归方程。测定苹果溶液吸光度,依据标准曲线计算糖含量,淀粉含量计算公式如下:

$$\text{淀粉含量}(\text{mg/g}) = \frac{c \times (v_1/v) \times d}{m \times 0.9} \quad (4)$$

式中, c 为样品中糖含量(mg/mL); v_1 为提取液总体积(mL); v 为测定溶液体积(mL); d 为稀释倍数; m 为样品重量(g);0.9为糖转化为淀粉的换算系数。

1.3 光谱数据预处理方法

为提升模型的准确性与鲁棒性,在建模前需对原始光谱进行预处理,以消除系统误差、噪声及散射

效应等干扰。针对高频噪声,采用 Savitzky-Golay (SG)平滑预处理和离散小波变换(DWT)预处理。SG平滑通过局部多项式拟合对光谱数据进行平滑处理,在保留光谱局部特征的同时滤除噪声;DWT预处理基于多尺度分解原理,将光谱分解为不同频段,可有效保留低频主信息并抑制高频噪声,尤其适用于非平稳信号处理。针对由样品粒径及光程变化引起的散射效应,采用多元散射校正(MSC)预处理和标准正态变量变换(SNV)预处理。MSC预处理通过将每条光谱与基准光谱进行线性回归,校正基线平移和乘性散射效应,使各光谱在散射水平上趋于一致;SNV预处理则对每条光谱独立进行中心化和标准化处理,以消除光程差异和颗粒散射带来的干扰,增强光谱间的可比性。为进一步提升信号质量与模型鲁棒性,引入自适应加权平滑(AWS)预处理和鲁棒性主成分分析(RPCA)预处理。AWS预处理依据光谱梯度变化自适应调整平滑强度,在降噪的同时保留关键信息;RPCA预处理将光谱数据分解为低秩结构和稀疏噪声,可有效抑制离群点和局部异常。针对仪器漂移或背景光变化引起的基线偏移,采用基线校正(BC)预处理方法修正光谱整体漂移,使光谱变化更准确地反映样本特征。

1.4 特征波段选择方法

光谱数据的波段中既蕴含与目标变量密切相关的有效信息^[21-25],也混杂着大量冗余数据,可能导致建模效率降低和精度下降。特征波段提取可以降低数据维度,剔除无关干扰,提升模型的性能与稳定性。本研究采用3种特征提取方法:连续投影算法(SPA)、竞争性自适应重加权采样法(CARS)和Bootstrap优化光谱选择算法(BOSS)。

SPA基于向量空间投影原理,通过迭代选取冗余度最小的波段,有效缓解光谱变量间的多重共线性问题。该算法通过保证所选变量间的低相关性,提升了模型的稳定性与可解释性,且结构简单、计算高效,适用于快速筛选任务。CARS利用偏最小二乘回归系数的绝对值评估变量重要性,并引入蒙特卡洛抽样机制,在多次迭代中动态调整波段权重,逐步压缩变量空间。该方法能有效识别对模型预测能力贡献显著的波段,尤其适用于目标属性与光谱变量间关系复杂的建模场景。BOSS结合Bootstrap重抽样与软阈值筛选策略,在多次建模与评估过程中保留预测贡献显著的波段。该方法适用于处理峰型复杂、噪声干扰较

强的数据集,能在确保模型稳健性的同时实现特征维度压缩,有助于提升整体建模性能与简洁性。

1.5 回归模型

本研究采用3种回归建模算法:偏最小二乘回归(PLSR)、支持向量回归(SVR)和岭回归(Ridge),分别与预处理及特征提取方法组合,构建苹果品质指标的预测模型。

偏最小二乘回归是一种适用于高维小样本数据的线性回归方法^[26]。该算法通过将原始自变量投影至潜变量空间,在最大化自变量与因变量间协方差的同时,有效缓解多重共线性问题。偏最小二乘回归计算效率高、模型结构清晰,尤其适用于变量远多于样本数量的线性建模场景。支持向量回归是一种典型的非线性建模算法,基于结构风险最小化原则构建。该算法通过引入核函数,将原始数据映射到高维特征空间,从而将非线性问题转化为高维空间中的线性回归问题。支持向量回归特别适用于处理非线性关系强、数据分布复杂的问题,具有较强的泛化能力与抗干扰性能。岭回归在最小二乘目标函数中引入L2正则化项,通过对回归系数施加惩罚约束,有效抑制高维共线性引发的系数波动。该方法适用于线性关系下的多变量回归任务,能够提升模型的稳定性与预测能力。

1.6 模型性能评估指标

为科学评估模型的拟合能力与泛化性能,本研究将原始光谱数据按照7:3划分为训练集与测试集。通过决定系数(R^2)、均方根误差(RMSE)、预测偏差比(RPD)综合评价各组合模型对苹果品质参数的预测效果,计算公式如下:

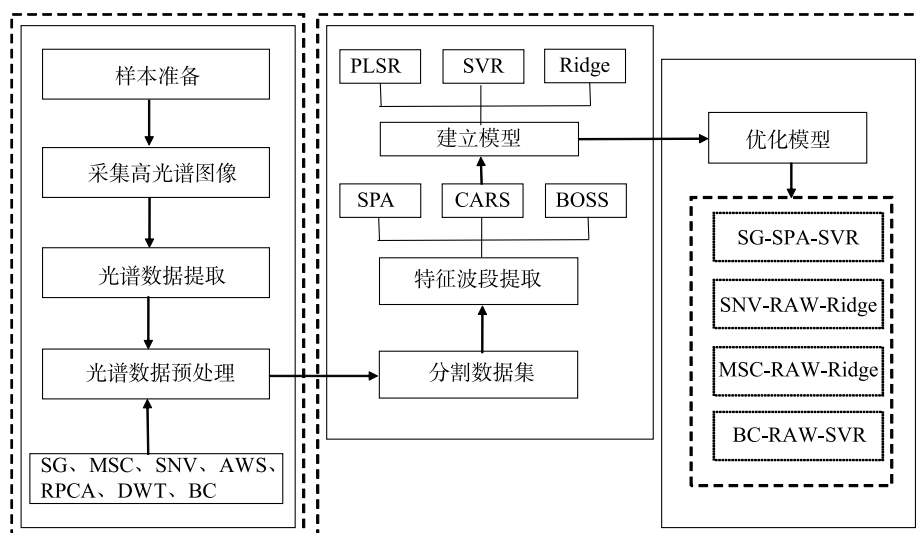
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$RPD = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}} \quad (7)$$

式中, y_i 为第*i*个样本的实际值, \hat{y}_i 为预测值, \bar{y} 为所有实际值的均值, n 为样本总数。

数据处理与建模流程如图2所示。



SG: Savitzky-Golay 平滑; MSC: 多元散射校正; SNV: 标准正态变量变换; AWS: 自适应加权平滑; RPCA: 鲁棒性主成分分析; DWT: 离散小波变换; BC: 基线校正; SPA: 连续投影算法; CARS: 竞争性自适应重加权采样法; BOSS: 优化光谱选择算法; PLSR: 偏最小二乘回归; SVR: 支持向量回归; Ridge: 岭回归。

图2 数据处理与建模流程

Fig.2 Data processing and modeling workflow

2 结果与分析

2.1 苹果光谱曲线特征分析

本研究共采集 240 份苹果样本的高光谱图像, 获取其在 380~1 018 nm 波段共计 320 个波段的光谱反射率数据。如图 3 所示, 不同苹果品种的平均光谱曲线整体呈现相似的变化趋势。在 530~630 nm 波段, 各苹果品种的光谱曲线表现出显著差异, 表明不同品种苹果的内部成分含量存在明显区别。在 500 nm 和 680 nm 波长附近的吸收特征与类胡萝卜素和叶绿素等色素成分有关。在近红外区, 975 nm 附近的吸收特征主要与果实内部水分含量有关, 而 840 nm 附近的光谱变化则通常与糖分、可溶性固形物含量有关。

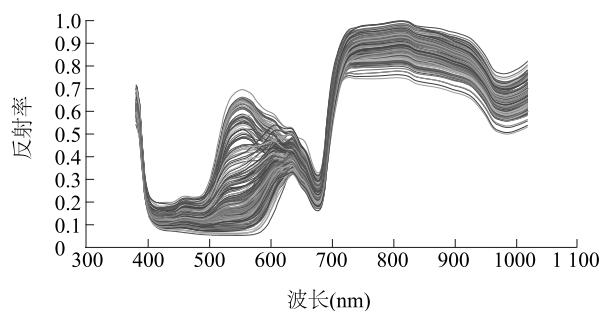


图3 不同品种苹果的原光谱曲线

Fig.3 Original spectral curves of different apple varieties

2.2 苹果理化指标

苹果维生素 C 含量、可滴定酸含量、可溶性固形物含量及淀粉含量如表 1 所示。

表 1 苹果理化指标

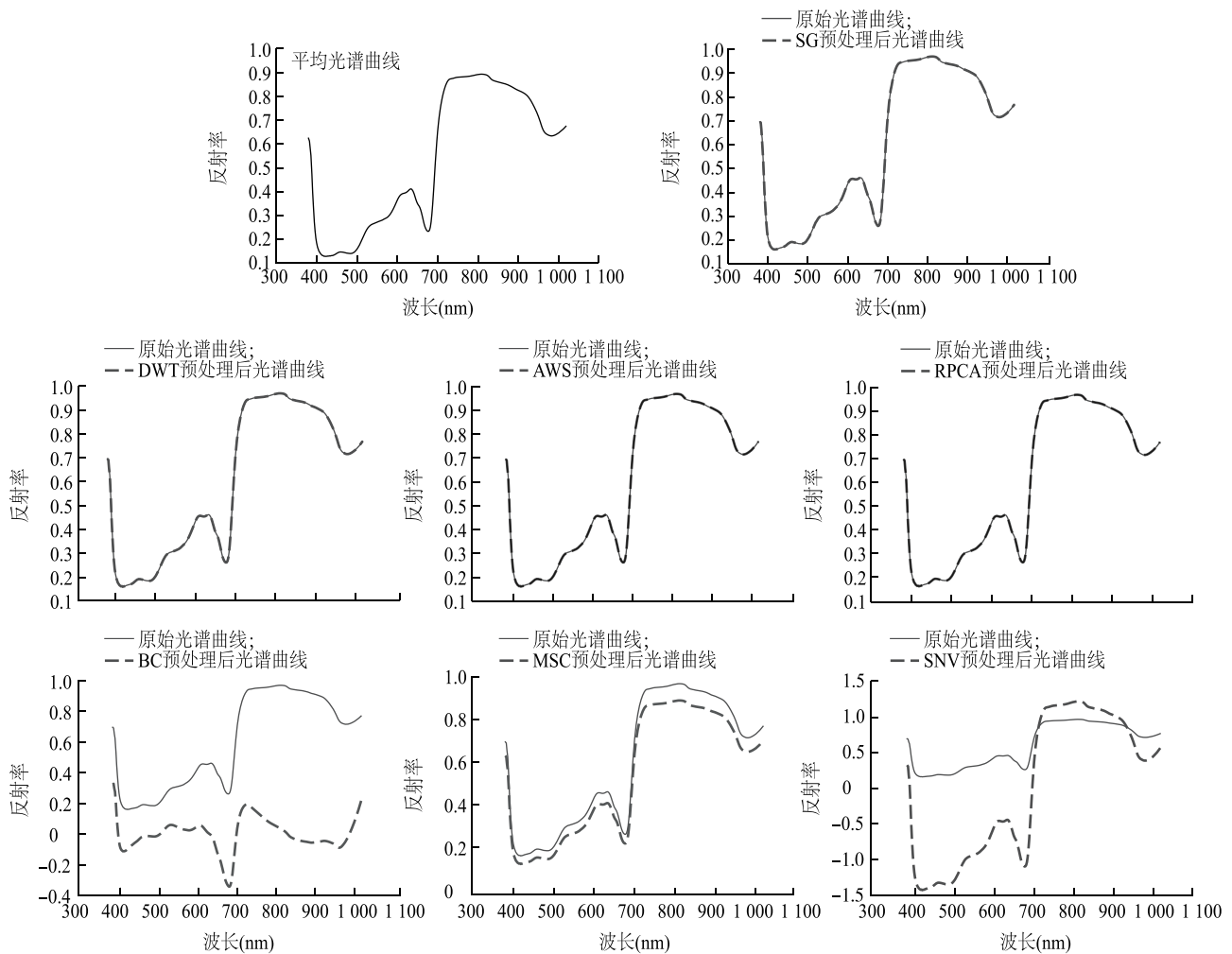
Table 1 Apple physicochemical indicators

理化指标	平均数	最大值	最小值
维生素 C 含量 (mg/g)	0.019 0±0.004 2	0.022 9	0.011 0
可滴定酸含量 (g/kg)	3.41±0.95	5.13	2.46
可溶性固形物含量 (%)	13.27±1.55	16.00	11.40
淀粉含量 (mg/g)	7.44±2.59	10.82	4.32

2.3 预处理方法对模型性能的影响

对原始光谱数据分别进行 SG 平滑预处理、MSC 预处理、SNV 预处理、AWS 预处理、RPCA 预处理、DWT 预处理和 BC 预处理, 并将预处理后的光谱数据分别用于建模。如图 4 所示, SG 平滑预处理降低了光谱噪声, BC 预处理消除了光谱基线偏移, SNV 预处理和 MSC 预处理校正了光谱散射差异, AWS 预处理和 DWT 预处理不仅降低了光谱噪声, 还进一步增强了光谱特征。

采用 PLSR 模型对预处理后的光谱数据进行建模。如表 2 所示, 对于维生素 C 含量, 基于 SG 预处



SG: Savitzky-Golay 平滑; DWT: 离散小波变换; AWS: 自适应加权平滑; RPCA: 鲁棒性主成分分析; BC: 基线校正; MSC: 多元散射校正; SNV: 标准正态变量变换。

图 4 原始光谱与预处理后光谱对比

Fig.4 Comparison between original spectra and preprocessed spectra

理的光谱数据建立的 PLSR 模型预测效果最佳。而对于可滴定酸含量和淀粉含量, 基于 BC 预处理的光谱数据建立的 PLSR 模型预测效果最佳。表 2 仅展示模型性能排名前 3 的预处理方法。

2.4 组合建模预测性能分析

分别采用连续投影算法 (SPA)、竞争性自适应重加权采样法 (CARS) 和基于 Bootstrap 的优化光谱选择算法 (BOSS) 3 种特征选择方法, 从预处理后的光谱数据中筛选出 20 个特征波段。针对每个检测指标, 结合前期确定的 3 种最优光谱预处理方法、上述 3 种特征选择方法与 3 种回归模型进行组合建模, 并从所有组合中筛选出性能排名前 3 的配置。如表 3 所示, 对于维生素 C 含量, SG-SPA-SVR 组合

建模预测效果最佳, 决定系数 (R^2) 为 0.928 9, 均方根误差 (RMSE) 为 0.115 7。对于可滴定酸含量, BC-原始光谱-Ridge 组合建模效果最佳, 模型 R^2 为 0.955 5, RMSE 为 0.189 3。对于淀粉含量, BC-原始光谱-SVR 组合建模预测效果最佳, 决定系数 (R^2) 为 0.926 5, 均方根误差 (RMSE) 为 0.698 8。对于可溶性固形物含量, MSC-原始光谱-Ridge 组合建模预测效果最佳, 决定系数 (R^2) 为 0.855 8, 均方根误差 (RMSE) 为 0.613 9。

SVR 在非线性成分建模中表现最佳, 尤其适用于维生素 C 含量和淀粉含量的预测; PLSR 则更适合处理光谱变量间相关性较强的成分, 如淀粉含量; 而 Ridge 模型在变量维度较高且存在线性关系的建

模任务中更具稳定性,例如可溶性固形物含量的预测。

表 2 光谱数据预处理对建模性能的影响

Table 2 Effects of spectral data preprocessing on modeling performance

指标	预处理	训练集			测试集		
		R^2	$RMSE$	RPD	R^2	$RMSE$	RPD
维生素 C 含量	SG	0.890 0	0.135 0	3.024 3	0.879 2	0.150 8	2.897 5
	RAW	0.891 8	0.133 9	3.049 2	0.874 5	0.153 7	2.842 8
	DWT	0.891 8	0.133 9	3.049 2	0.874 5	0.153 7	2.842 8
可滴定酸含量	BC	0.969 7	0.169 8	5.763 4	0.935 9	0.244 2	3.700 0
	MSC	0.959 4	0.196 4	4.980 6	0.932 6	0.232 9	3.878 5
	SNV	0.957 9	0.200 3	4.885 7	0.917 9	0.257 0	3.515 4
可溶性固形物含量	MSC	0.895 6	0.487 5	3.103 5	0.816 3	0.692 9	2.349 5
	SG	0.909 1	0.454 7	3.327 3	0.804 9	0.714 0	2.280 0
	SNV	0.899 7	0.477 8	3.166 4	0.814 1	0.697 1	2.335 5
淀粉含量	BC	0.939 7	0.638 8	4.085 8	0.886 3	0.869 0	2.986 0
	DWT	0.937 7	0.649 5	4.018 3	0.875 8	0.908 0	2.857 6
	SG	0.930 2	0.687 6	3.795 9	0.882 7	0.882 3	2.940 9

SG; Savitzky-Golay 平滑; MSC; 多元散射校正; SNV; 标准正态变量变换; DWT; 离散小波变换; BC; 基线校正; RAW; 原始光谱; R^2 ; 决定系数; $RMSE$; 均方根误差; RPD ; 相对预测偏差。

表 3 不同组合建模策略对苹果品质指标的预测性能

Table 3 Prediction performance of different combined modeling strategies for apple quality indices

指标	预处理	特征提取	模型	训练集			测试集		
				R^2	$RMSE$	RPD	R^2	$RMSE$	RPD
维生素 C 含量	SG	SPA	SVR	0.892 9	0.133 2	3.064 2	0.928 9	0.115 7	3.776 3
				0.886 4	0.137 2	2.975 9	0.907 7	0.131 8	3.315 5
				0.886 4	0.137 2	2.975 9	0.907 7	0.131 8	3.315 5
可滴定酸含量	BC	RAW	Ridge	0.980 8	0.135 1	7.244 8	0.955 5	0.189 3	4.773 5
				0.977 1	0.147 7	6.626 6	0.954 5	0.191 3	4.722 3
				0.957 5	0.201 1	4.865 7	0.951 1	0.198 5	4.552 2
可溶性固形物含量	MSC	RAW	Ridge	0.919 4	0.428 3	3.532 7	0.855 8	0.613 9	2.651 7
				0.873 8	0.535 9	2.823 4	0.839 8	0.647 1	2.515 8
				0.869 2	0.545 5	2.773 6	0.838 3	0.650 1	2.504 1
淀粉含量	BC	RAW	SVR	0.925 1	0.712 4	3.663 7	0.926 5	0.698 8	3.713 3
				0.927 1	0.702 5	3.715 1	0.912 8	0.761 1	3.409 4
				0.923 2	0.721 3	3.618 5	0.912 2	0.763 5	3.398 6

SG; Savitzky-Golay 平滑; MSC; 多元散射校正; SNV; 标准正态变量变换; DWT; 离散小波变换; BC; 基线校正; RAW; 原始光谱; R^2 ; 决定系数; $RMSE$; 均方根误差; RPD ; 相对预测偏差; SPA; 连续投影算法; CARS; 竞争性自适应重加权采样法; BOSS; 优化光谱选择算法; PLSR; 偏最小二乘回归; SVR; 支持向量回归; Ridge; 岭回归。

2.5 敏感波段的筛选与优化

为降低数据冗余维度并提升模型稳定性,本研究对特征波段进行筛选与优化。首先,对380~1 018 nm 波段的光谱数据与维生素 C 含量、可滴

定酸含量、可溶性固形物含量、淀粉含量进行皮尔逊相关性分析。如图 5 所示,不同品质指标在各波段上的相关响应特征存在差异,其中可滴定酸含量和淀粉含量在 806~912 nm 波段范围内具有

较高相关性。表明不同品质指标具有各自敏感的特征波段。

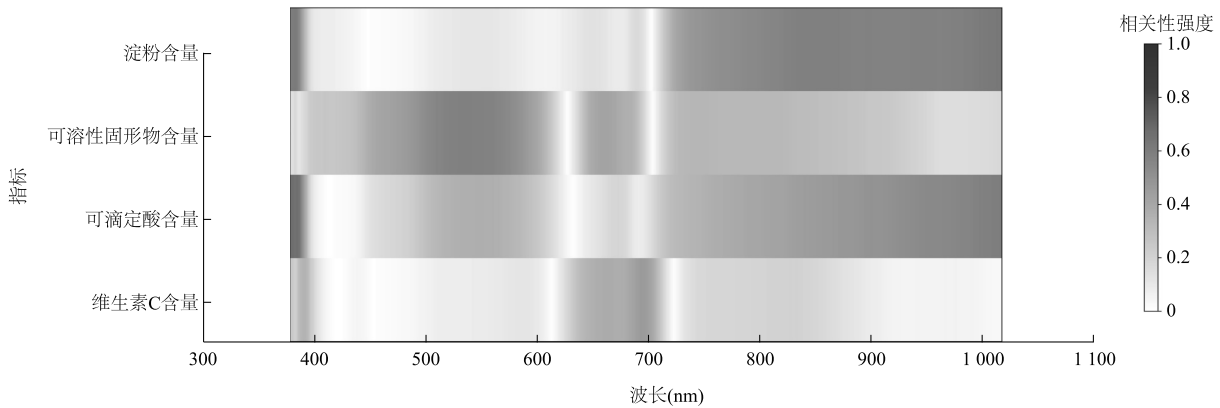
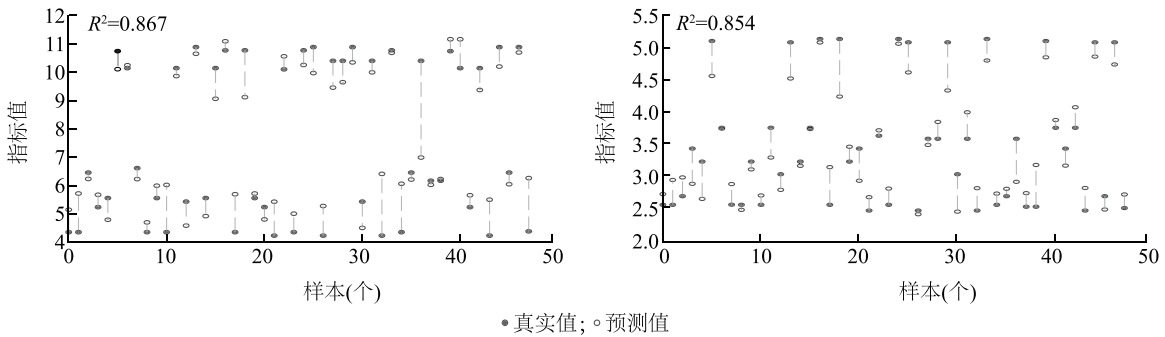


图 5 光谱波段与品质指标的相关性热力图

Fig.5 Heatmap of correlation between spectral bands and quality indices

结合随机森林回归模型的重要性排序,对筛选得到的敏感波段进行精简。以淀粉含量和可滴定酸含量为目标变量,分别构建基于精简波段的岭回归(Ridge)模型。采用随机划分方式按 8 : 2 将样本集

划分为训练集和测试集。如图 6 所示,淀粉含量和可滴定酸含量的预测值与实测值均呈现较好的一致性。



R^2 : 决定系数。

图 6 相关性预测结果回归散点图

Fig.6 Regression scatter plot of correlation prediction results

3 讨论与结论

本研究构建了一套融合光谱预处理、特征波段选择与回归建模的多策略组合方法,以实现苹果维生素 C 含量、可滴定酸含量、可溶性固形物含量及淀粉含量的准确预测。在光谱预处理环节,对比分析了 Savitzky-Golay 卷积平滑(SG)、标准正态变量变换(SNV)及基线校正(BC)等方法,他们分别在抑制噪声、校正散射影响及消除基线漂移方面表现出良好的适用性。在特征波段选择上,采用连续投影算法(SPA)、竞争性自适应重加权算法(CARS)和优化光谱选择算法(BOSS)对高维光谱数据进行

降维与冗余消除。结果表明,不同算法对各品质指标的适应性存在差异:SPA 在维生素 C 含量的预测中效果最佳。在回归模型构建阶段,比较了支持向量回归(SVR)、偏最小二乘回归(PLSR)及岭回归(Ridge)的预测性能。其中,SVR 模型对维生素 C 含量和淀粉含量等非线性变量的建模效果最优,而 PLSR 和 Ridge 则更适用于光谱冗余度高或变量存在线性关系的预测目标。

各品质指标的最优建模路径呈现显著差异性:对于维生素 C 含量,SG-SPA-SVR 组合建模预测效果最佳。对于可滴定酸含量,BC-原始光谱-Ridge 组合建模效果最佳。对于淀粉含量,BC-原始光谱-

SVR 组合建模预测效果最佳。对于可溶性固形物含量, MSC-原始光谱-Ridge 组合建模预测效果最佳。本研究构建的多策略建模体系可为果品无损分选、品质追溯及在线检测系统提供技术支持。

尽管本研究构建的多策略建模体系取得了较好的预测效果,但仍面临样本区域单一、模型迁移性不足等问题,未来研究可进一步拓展数据样本多样性,结合便携设备部署、高光谱图像融合算法与深度学习技术,提升其在多品种、多场景下的适用性与智能化水平。

参考文献:

- [1] 高 冲. 基于高光谱与深度学习的苹果品质检测方法及装置研究[D]. 泰安: 山东农业大学, 2022.
- [2] BAI S H, TOOTOONCHY M, KÄMPER W, et al. Predicting carbohydrate concentrations in avocado and macadamia leaves using hyperspectral imaging with partial least squares regressions and artificial neural networks[J]. *Remote Sensing*, 2024, 16(18): 3389.
- [3] WANG X J, HAN J Y, LIU C Z, et al. Non-destructive assessment of apple internal quality using rotational hyperspectral imaging[J]. *Frontiers in Plant Science*, 2024, 15: 1432120.
- [4] ZHU D Y, HAN J Y, LIU C Z, et al. Modeling of flaxseed protein, oil content, linoleic acid, and lignan content prediction based on hyperspectral imaging[J]. *Frontiers in Plant Science*, 2024, 15: 1344143.
- [5] 曾柯宜, 刘禹彤, 张 倩, 等. 基于多模态融合的玉米种子成熟度的无损检测[J]. *食品安全质量检测学报*, 2025, 16(2): 171-177.
- [6] GAO W J, CHENG X, LIU X H, et al. Apple firmness detection method based on hyperspectral technology[J]. *Food Control*, 2024, 166: 110690.
- [7] WANG F, ZHAO C J, YANG H, et al. Non-destructive and in-site estimation of apple quality and maturity by hyperspectral imaging[J]. *Computers and Electronics in Agriculture*, 2022, 195: 106843.
- [8] QIAO M M, CUI T, XIA G Y, et al. Integration of spectral and image features of hyperspectral imaging for quantitative determination of protein and starch contents in maize kernels[J]. *Computers and Electronics in Agriculture*, 2024, 218: 108718.
- [9] 肖宏辉, 邓 浩, 李春霖, 等. 基于高光谱成像技术的橙子可溶性固形物和可滴定酸含量无损预测研究[J]. *核农学报*, 2025, 39(5): 1012-1020.
- [10] DONG J L, GUO W C. Nondestructive determination of apple internal qualities using near-infrared hyperspectral reflectance imaging[J]. *Food Analytical Methods*, 2015, 8: 2635-2646.
- [11] 王宏恩, 冯国红, 徐华东, 等. 基于可见-近红外光谱和深度森林的蓝莓成熟度判别[J]. *光谱学与光谱分析*, 2024, 44(11): 3280-3286.
- [12] MO C, KIM M S, KIM G, et al. Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging[J]. *Biosystems Engineering*, 2017, 159: 10-21.
- [13] ZHANG P P, WANG H W, JI H W, et al. Hyperspectral imaging-based early damage degree representation of apple: a method of correlation coefficient[J]. *Postharvest Biology and Technology*, 2023, 199: 112309.
- [14] XU D H, WANG H W, JI H W, et al. Quantitative evaluation of impact damage to apple by hyperspectral imaging and mechanical parameters[J]. *Food Analytical Methods*, 2019, 12: 371-380.
- [15] JIANG D Y, CHANG Q R, ZHANG Z J, et al. Monitoring the degree of mosaic disease in apple leaves using hyperspectral images[J]. *Remote Sensing*, 2023, 15(10): 2504.
- [16] ZHANG P, SHEN B B, JI H W, et al. Nondestructive prediction of mechanical parameters to apple using hyperspectral imaging by support vector machine[J]. *Food Analytical Methods*, 2022, 15(5): 1397-1406.
- [17] 李泽祺, 杨 铮, 周壮飞, 等. 高光谱与宽度学习检测草莓叶片含水率[J]. *光谱学与光谱分析*, 2025, 45(6): 1534-1542.
- [18] LIU Y F, ZHANG Y, JIANG D Y, et al. Quantitative assessment of apple mosaic disease severity based on hyperspectral images and chlorophyll content[J]. *Remote Sensing*, 2023, 15(8): 2202.
- [19] WANG Z J, WU S S, ZUO C Z, et al. Exploring the variability and heterogeneity of apple firmness using visible and near-infrared hyperspectral imaging[J]. *LWT-Food Science and Technology*, 2024, 192: 115704.
- [20] ZHAN C Y, MAO H Y, FAN R S, et al. Detection of apple sucrose concentration based on fluorescence hyperspectral image system and machine learning[J]. *Foods*, 2024, 13(22): 3547.
- [21] 张永亮, 汪 泓, 肖玖军, 等. 基于高光谱的山区耕地土壤有机质含量估测[J]. *江苏农业学报*, 2024, 40(1): 112-120.
- [22] 李贝贝, 王伟宁, 赵 森, 等. 云烟 87 叶片生理指标对镁浓度的响应特征及高光谱反演[J]. *南方农业学报*, 2025, 56(11): 3341-3351.
- [23] 余德焰, 江晓东, 杨莹颖, 等. 基于 RGB 与高光谱图像的小麦低温胁迫评估模型[J]. *南方农业学报*, 2025, 56(1): 97-110.
- [24] 李 伟, 宋 睿, 刘明江, 等. 光学遥感在区域灌溉监测中的应用的研究进展[J]. *排灌机械工程学报*, 2024, 42(11): 1157-1165.
- [25] 刘江凡, 赵泽艺, 李朝阳, 等. 基于无人机多光谱遥感的苹果树冠层 SPAD 反演[J]. *排灌机械工程学报*, 2024, 42(5): 525-531.
- [26] 王 宇, 汪 泓, 肖玖军, 等. 基于 MCC-GAPLS-PLSR 的辣椒叶绿素含量高光谱定量反演[J]. *江苏农业学报*, 2024, 40(5): 865-873.

(责任编辑:成纾寒)