

基于多模型软投票融合的青提失水率预测模型

樊 改¹ 景 敏^{1,2} 计 丰¹ 马茂森¹ 姚小敏^{1,2}

(1. 陕西理工大学机械工程学院, 陕西 汉中 723000; 2. 陕西省工业自动化重点实验室, 陕西 汉中 723000)

摘要: [目的] 依托 LED 诱导荧光光谱系统, 提出一种利用采集光谱数据结合多模型软投票融合预测青提失水率的方法, 实现青提失水率快速无损检测, 控制损耗、保障品质。[方法] 采集 100 个青提样本光谱数据, 利用多元散射校正 (MSC)、标准正态变量变换 (SNV)、归一化处理及萨维茨基—戈雷滤波 (SG) 进行预处理, 采用连续投影算法进行特征波长筛选, 构建青提失水率随机森林 (RF)、偏最小二乘 (PLS)、最小二乘支持向量机 (LSSVM) 单模型和多模型软投票融合预测模型, 并进行对比分析。[结果] 单模型中 LSSVM 的综合性能最优, 但存在明显过拟合风险; 基于测试集 R_p^2 与平均绝对误差 (MAE) 指标分配权重构建多模型软投票融合模型, 测试集 R_p^2 较 LSSVM 单模型提升 25.3%, MAE 降低 27.6%, 过拟合现象得到显著改善, 同时在拟合精度与泛化能力上均实现优化。[结论] 多模型软投票融合策略的预测精度优于单模型, 其具有更强稳定性、泛化能力与抗过拟合性能。

关键词: 青提; 失水率; LED 诱导荧光光谱; 多模型软投票融合; 无损检测

Prediction model of green grape water loss rate based on multi-model soft voting fusion

FAN Gai¹ JING Min^{1,2} JI Feng¹ MA Maosen¹ YAO Xiaomin^{1,2}

(1. School of Mechanical Engineering, Shaanxi University of Technology, Hanzhong, Shaanxi 723000, China;

2. Shaanxi Key Laboratory of Industrial Automation, Hanzhong, Shaanxi 723000, China)

Abstract: [Objective] This study aims to achieve rapid and non-destructive detection of water loss rate in green grapes, control losses, and ensure quality. Based on an LED-induced fluorescence spectroscopy system, a prediction method combining collected spectral data with a multi-model soft voting fusion approach was proposed for green grape water loss rate estimation. [Methods] Spectral data from 100 green grape samples were collected. Preprocessing was performed using multivariate scatter correction (MSC), standard normal variate transformation (SNV), normalization, and Savitzky-Golay (SG) filtering. Feature wavelength selection was carried out using the successive projections algorithm. Based on this, prediction models for green grape water loss rate were constructed using Random Forest (RF), Partial Least Squares (PLS), and Least Squares Support Vector Machine (LSSVM), as well as a single-model and a multi-model soft voting fusion model, followed by comparative analysis. [Results] Among the single models, LSSVM showed the best overall performance but exhibited a clear risk of overfitting. A multi-model soft voting fusion model was constructed by assigning weights based on the test set R_p^2 and mean absolute error (MAE) index. Compared with the LSSVM single model, the test set R_p^2 performance improved by 25.3%, and MAE decreased by 27.6%, with overfitting significantly alleviated. Meanwhile, improvements were achieved in both fitting accuracy and generalization ability. [Conclusion] The multi-model soft voting fusion strategy exhibits higher prediction accuracy than single models, with stronger stability, generalization ability, and anti-overfitting performance.

Keywords: green grapes; water loss rate; LED-induced fluorescence spectroscopy; multi-model soft voting fusion; non-destructive testing

基金项目: 陕西省重大产业创新链项目 (编号: 2021ZDLSF06-07); 陕西省自然科学基金基础研究项目 (编号: 2022JM-383); 陕西理工大学人才启动项目 (编号: SLGRCQD2103)

通信作者: 景敏 (1978—), 女, 陕西理工大学教授, 博士。E-mail: jingmin@sntu.edu.cn

收稿日期: 2025-10-10 **改回日期:** 2026-04-10

引用格式: 樊改, 景敏, 计丰, 等. 基于多模型软投票融合的青提失水率预测模型[J]. 食品与机械, 2026, 42(6): 12-20.

Citation: FAN Gai, JING Min, JI Feng, et al. Prediction model of green grape water loss rate based on multi-model soft voting fusion[J]. Food & Machinery, 2026, 42(6): 12-20.

青提是葡萄的优良品种之一,兼具独特口感与高产优势,但采后易失水腐烂,耐贮性差,丰产滞销时经济损失突出^[1]。水分是新鲜度与商品价值核心保障,精准检测失水率对减少损失至关重要。目前失水率检测技术呈传统方法逐步被替代、近红外成主流、LED诱导荧光为前沿的趋势。

黄华明等^[2]提出了一种甜春橘柚失水管控最优包装与贮藏方案;徐呈祥等^[3]明确了柑橘果皮蜡质与耐贮性的关联。但维度上前者未涵盖可溶性固形物等核心品质指标,后者未结合呼吸、乙烯释放等生理机制分析耐贮性。

近红外光谱技术因快速无损等优势成为农产品检测的核心技术^[4],相关研究覆盖核果、浆果等,以及建模算法多技术方向,各类研究在验证技术可行性的同时,暴露出模型通用性不足、预处理适配性有限、稳定性欠缺等问题。关晔晴等^[5]采用近红外检测深州蜜桃可溶性固形物含量(SSC)模型 R_p^2 达0.79,证实该技术对核果类有一定适配性,但蜜桃水分分布特殊,其pH模型 R_p^2 仅0.58,显露出技术对不同指标适配性差异显著。朱金艳等^[6]结合近红外与极限学习机(ELM)构建的贮藏品质模型精度优于偏最小二乘回归(PLSR)的,但单一方法对样本分布敏感且未优化跨品类预处理流程,适配性不足;张晨等^[7]的Stacking集成框架模型精度较高,因双层结构致计算复杂且500+训练样本量与小样本研究适配性欠佳。

近年来,近红外光谱技术的应用取得了阶段性进展,但也存在明显的局限。刘广志^[8]采用PLS、主成分回归(PCR)等算法,验证了该技术在粮食肉类检测领域也具有无损、快速、多指标同步检测的潜力,但未深入探讨模型在跨品种、跨批次样本中的通用性,也未对光谱漂移问题进行针对性分析;王永乾等^[9]针对烟草样本多样复杂问题,提出了重难点样本加权结合偏最小二乘回归系数阈值筛选的改进方法,验证了该方法可提升烟草水分模型的预测精度,但改善效果会因重难点样本占比变化而波动;贾利红等^[10]在酒醅成分检测中构建了集成模型,采用相关系数分配权重,因未综合考虑多维度误差与拟合度,且酒醅与水果的光谱特性差异显著,导致技术难以向果蔬检测领域迁移;Turgut等^[11]利用多款红外检测设备实现了青提粉掺假的精准识别,并锁定了关键特征波长,验证了该技术在食品认证中的优势,但未涉及模型在不同仪器间的通用性,也未评估环境因素对检测结果的干扰;高升^[12]利用近红外相机成像实现红提串缺陷初步筛查,验证了技术可行性,但未探究模型通用性及光照干扰问题;王晨晨等^[13]用可见/近红外透射光谱检测亚健康苹果,其思路可为葡萄检测提供参考,但难以直接迁移;王丹等^[14]结合改进卷积神经网络提升葡萄品质检测准确率,却存在模型复杂、检测速度不足及稳定性未验证的问题。

在通用建模与算法优化的相关研究中,Wang等^[15]综

述了近红外农业检测的GAN数据增强、插值法、SPA/CARS特征选择等技术进展,同时指出其数据成本高、可解释性与跨场景通用性不足的共性问题;聂衍文等^[16]基于MobileNet简化芒果缺陷检测模型,实现了高准确率与移动端适配,但其小缺陷识别精度、多品种光照适应性及工业稳定性待验证;王云鹏等^[17]融合红外与可见光图像提升了苹果隐蔽缺陷识别率,但因算法复杂导致检测速度难匹配工业分选,跨品种及小缺陷检出能力需强化;姜小刚等^[18]优化了果实运动参数,并结合PLS与MSC提升了苹果检测精度,但未对模型跨品种通用性及抗成熟度干扰能力进行深入探究。

近红外光谱检测技术存在模型通用性差、难适配青提光谱、单算法泛化弱、过拟合明显、部分集成模型结构复杂或权重单一一致实用与稳定性不足的现象。而激光诱导荧光(LIF)技术作为新兴光谱检测技术^[19],因取样少、可实时检测且效率高成为青提失水率检测的研究热点,但存在激光要求高、系统难、室外现场检测能力弱等局限。刘梁晨等^[20]搭建了255 nm激发光源LED诱导荧光检测系统,经光谱平滑预处理后,构建了多环芳烃荧光峰强度与质量比的线性拟合模型;Hara等^[21]对比了多波长拉曼光谱特性,发现532 nm下胡萝卜素拉曼峰强度与浓度无关,仅峰位移可反映其成分转变。

试验拟以青提为对象,利用LED诱导荧光光谱技术,采集200~1 200 nm LED光谱并分步建模,通过特征筛选去噪声冗余,构建RF、PLS、LSSVM单模型,比较各单一模型的拟合精度、预测误差及过拟合风险,并提出利用测试集 R_p^2 与平均绝对误差(MAE)双维度指标分配权重多维度权重软投票融合模型。基于100个小样本为“小样本+强干扰”场景的模型优化提供可推框架,为水果水分快速无损检测技术提供依据。

1 理论与方法

1.1 试验设置及系统

1.1.1 系统 采用分辨率为0.02 nm的Aurora4000光谱仪,搭配365 nm LED紫外灯采集青提光谱;预热仪器30 min,设积分时间8 s、信号平均3次,连续扫描3次取均值,在青提赤道10 mm标记区每次采集10组光谱,采集前需清洁探头并精准定位。全程置于密闭暗室中,每颗青提每24 h采谱5次并取平均值,为避免紫外导致果面白斑,同一部位单次照射 ≤ 10 s且相邻采集间隔 ≥ 2 h。试验平台图如图1所示。

1.1.2 试验 试验样本为品种一致的陕西青提,从同批次中筛选10个形态均匀、完整性良好的青提作为样本。青提样品满足单果重差异为 $\pm 5\%$ 、外观无损伤病害且大小一致。试验前先擦拭样本并编号^[22],于 $(25 \pm 1)^\circ\text{C}$ 、相对湿度 $(60 \pm 5)\%$ 的通风恒温恒湿环境中静置24 h。10个

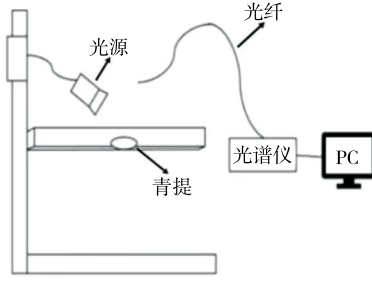


图 1 平台简图

Figure 1 Simplified diagram of the experimental platform

同批次陕西青提样本经 10 d 采集形成 100 条光谱数据。

每日采谱前,用 AR-097 电子天平称量青提并连续记录 10 d,因青提样本初始质量存在差异,直接对比绝对质量数据会产生偏差,故采用相对失水率分析以消除该影响^[23],其计算式如式(1)所示。图 2 为 1~10 号青提样本失水率随时间变化趋势。结果显示,样本失水率随时间持续上升且规律一致,可直观反映青提水分的流失。

$$W_{lr} = \frac{W_1 - W_n}{W_1 - W_{10}} \times 100\%, \quad (1)$$

式中:

W_{lr} ——失水率,%;

W_1 ——第 1 天质量,g;

W_n ——第 n 天质量,g;

W_{10} ——第 10 天质量,g。

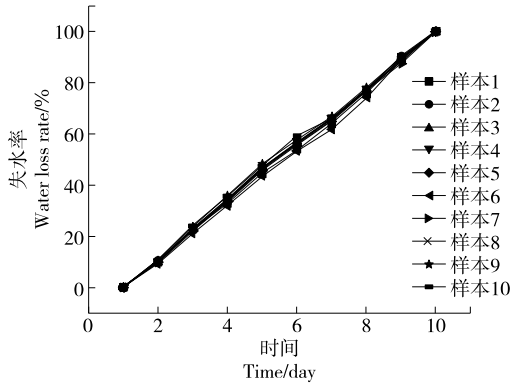


图 2 青提失水率随时间的变化趋势

Figure 2 Trend in the water loss rate of green grapes over time

1.2 多模型软投票融合决策设计

权重分配核心是使泛化能力强、预测误差小的模型主导融合决策且抑制弱性能模型干扰,提升融合模型稳定性与精度。选取 R_p^2 与 MAE 为权重分配指标: R_p^2 可量化模型对未知样本拟合度以实现精准预测,MAE 能衡量预测平均偏差以保障稳定泛化,二者从拟合精度与误差控制形成互补,覆盖模型实用性能;训练集指标体现模型

对已知数据拟合能力,易高估过拟合模型,而测试集指标可反映模型泛化能力,保障权重与实际应用匹配; R_p^2 聚焦整体拟合趋势,MAE 侧重局部误差分布,二者结合实现模型性能评估,确保权重分配结果更稳定。

1.2.1 等权重分配理论依据 采用 R_p^2 与 $Y_i(1/MAE)$ 各占 50% 的等权重分配方案,首先,依据指标互补原则以趋势精准和误差可控为维度评价模型,避免单一指标引起的偏向性和主观偏差;其次,在领域共识中,无论近红外光谱酒醅检测还是农业建模中,该方案均能同时提高预测精度和稳定性,并可在小样本下避免单一指标导向的模型偏差^[10,15]。

1.2.2 权重计算

(1) 指标正向化处理: R_p^2 为正向指标值越大性能越优,MAE 为负向指标值越小性能越优,先将 MAE 转换为正向指标,消除指标方向差异。设第 i 个模型的测试集 MAE 为 MAE_i ,其正向化指标定义为

$$Y_i = \frac{1}{MAE_i}. \quad (2)$$

(2) 指标归一化处理:为消除 R_p^2 与 $Y_i(1/MAE)$ 的量纲差异,对于正向指标 R_p^2 ,第 i 个模型的归一化值为

$$X_{norm,i} = \frac{R_{p,i}^2 - R_{p,min}^2}{R_{p,max}^2 - R_{p,min}^2}, \quad (3)$$

式中:

$R_{p,max}^2$ ——所有模型中 R_p^2 的最大值;

$R_{p,min}^2$ ——所有模型中 R_p^2 的最小值。

对于正向化后的 MAE 指标 Y_i ,第 i 个模型的归一化值为

$$Y_{norm,i} = \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}}, \quad (4)$$

式中:

Y_{max} ——所有模型中 Y_i 的最大值;

Y_{min} ——所有模型中 Y_i 的最小值。

(3) 综合权重计算: R_p^2 与 $Y_i(1/MAE)$ 对模型性能影响同等重要,各占 50%,符合两者互补评估理论,则第 i 个模型的综合性能得分 S_i 为

$$S_i = 0.5 \times X_{norm,i} + 0.5 \times Y_{norm,i}. \quad (5)$$

为确保所有模型权重之和为 1,第 i 个模型的最终权重 W_i 为

$$W_i = \frac{S_i}{\sum S_i}. \quad (6)$$

1.3 评价标准

分别按式(7)、式(8)计算测试集决定系数 R_p^2 和 MAE,通过 $R_c^2 - R_p^2$ 差值衡量过拟合程度,差值越小风险越低。

$$R_p^2 = 1 - \frac{\sum_{i=1}^N (y_a - y_b)^2}{\sum_{i=1}^N (y_a - y_c)^2}, \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^N |y_a - y_b|, \quad (8)$$

式中:

N ——试验样本数;

y_a ——用标准方法测得预测集中第 i 个样本实际值;

y_b ——第 i 个样本通过建立近红外数学模型计算所得预测值;

y_c ——所有青提样本预测值的平均值。

2 结果与分析

2.1 青提原始光谱采集

采集 1~10 号青提样本光谱,累计获得 100 条 200~1 200 nm 范围内有效光谱数据,如图 3 所示。

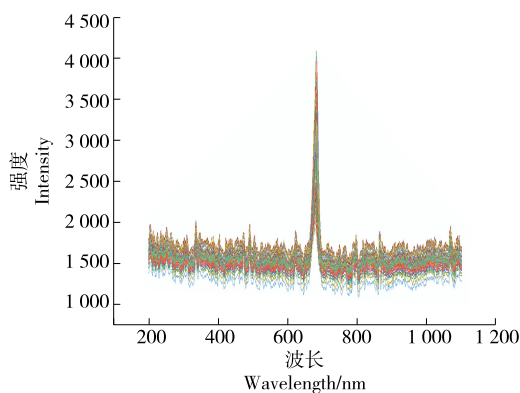


图 3 原始光谱仪采集

Figure 3 Original spectrometer collection

2.2 数据预处理

由图 4 可知,500~800 nm 波段内各光谱与 200~1 200 nm 范围趋势一致,670~720 nm 区间波峰稳定且透过率显著升高^[24],500~670,720~800 nm 区间光强低、变化小且透过率平稳,故选取 500~800 nm 为有效分析波段,并基于此波段依次开展光谱预处理、样本集划分、单模型建立及多模型融合优化。

为消除原始光谱信号偏移与光散射干扰、提升数据质量,采用 MSC、SNV、归一化处理及 S-G 4 种预处理方式,结果如图 5 所示。

由图 5 可知,SNV 在消除光谱系统性偏移与缩放效应、降低青提表面散射及测量噪声上表现最优,还能有效增强光谱有效信息、凸显波形特征,故选用 SNV 预处理后光谱数据开展特征波长选取。

2.3 样本集划分

针对 500~800 nm 有效波段的 100 条原始光谱数据,采用 SPXY 算法按 7:3 比例划分为训练集样本和测试集样本,具体划分数值见表 1。

由表 1 可知,SPXY 算法划分训练集与测试集标准偏

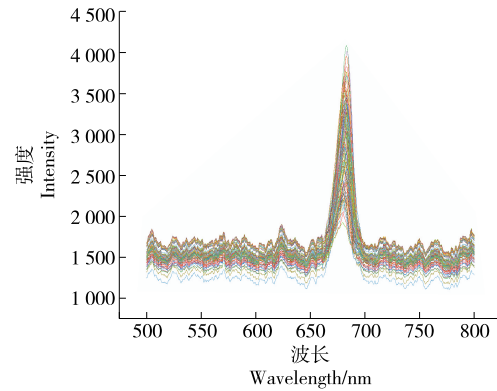


图 4 500~800 nm 青提失水率原始光谱图

Figure 4 Original spectral graph of water loss rate of green grapes at 500~800 nm

表 1 训练集与预测集青提失水率含量参考值统计结果

Table 1 Statistical results of reference values for water loss rate of green grapes in the training set and prediction set

样本	数量	失水率含量 范围/%	失水率含量 平均值/%	失水率含量 标准偏差
SPXY 训练集	70	0~100	55.624	30.540
SPXY 测试集	30	0~100	37.151	30.777

差相近,分别为 30.540% 和 30.777%,样本分布均匀,可提升模型泛化能力,说明该算法在样本覆盖完整性和离散度匹配度上表现更优,能提供代表性强且稳定样本以支撑模型精度提升,故选其划分样本集并开展后续特征波长提取。

2.4 特征波长提取

为构建青提失水率检测模型,采用连续投影算法 (SPA) 对 500~800 nm 波段的光谱数据进行特征波长提取^[25],结果如图 6 所示。该波段共采集 1 168 个光谱数据点。

SPA 参数:波长选择变量数范围 1~9、步长 1,当筛选 3 个特征波长时,校正集均方根误差 (RMSEC) 达最小值 23.898 5,波长 504.369,618.600,685.297 nm 占原始光谱信息的 0.257%,实现了大幅降维;为提升模型稳定性与抗干扰能力,结合聚类分析及建模验证,将特征波长分别扩展为 500~510,615~625,680~690 nm 波段,其光谱变化与青提失水关键成分分子振动直接相关^[26-28],可为模型构建提供可靠特征支撑。

2.5 试验模型

2.5.1 单模型建立 在青提失水率预测研究中,基于 SPXY 算法划分的 70 个训练集样本,结合 SNV 预处理数据及 SPA 筛选 3 个特征波长,构建 RF、PLS、LSSVM 预测模型如图 7 所示。各模型参数:RF 模型决策树数量

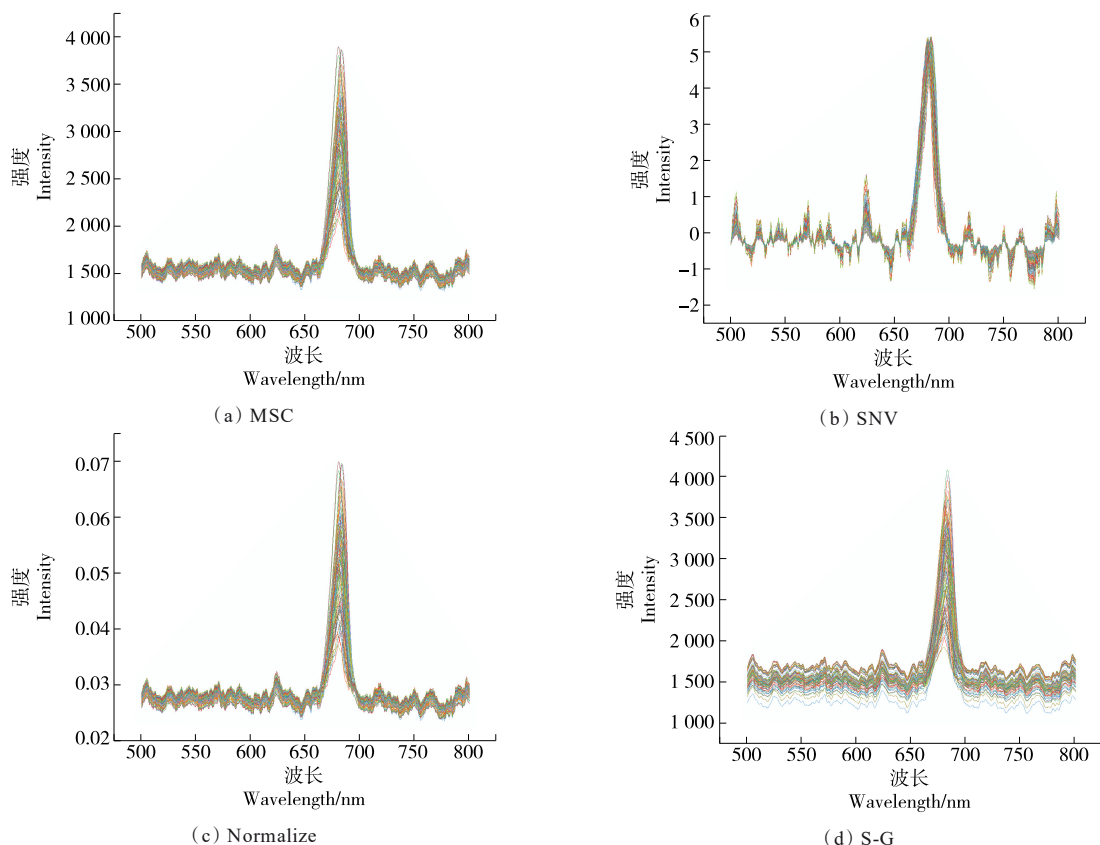


图 5 青提失水率经预处理的光谱图

Figure 5 Spectra of green grape water loss rate after preprocessing

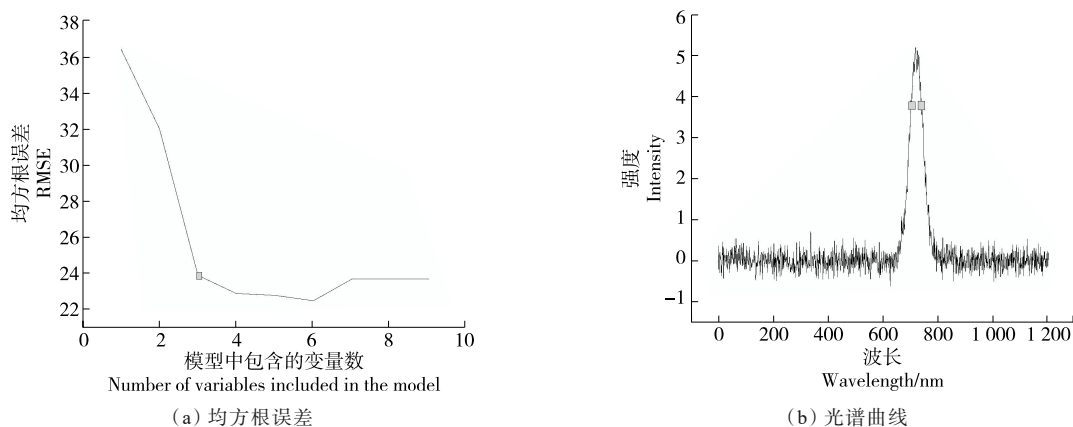


图 6 青提失水率的 SPA 特征波长选取

Figure 6 SPA characteristic wavelength selection diagram of water loss rate of green grapes

100 棵, 树最大深度由数据自动适配; 节点分裂与叶子节点最小样本数分别为 2, 1; 随机特征选择数量为总特征数平方根, 116 个特征波长中选 10 个, 其余参数默认; PLS 模型经交叉验证确定主成分数为 5, 对应交叉验证均方根误差最小值; 采用标准化预处理形成双重优化, 算法迭代次数为 1 000; LSSVM 模型核函数为径向基函数 RBF, 其核函数参数 σ 与正则化参数 γ 经网格搜索结合交叉验证优

化为 0.8 和 100, 误差惩罚参数 0.01, 采用最小二乘迭代法求解。

由图 7 可知, LSSVM 模型的预测结果散点更接近真实值—预测值拟合线, 测试集散点无明显偏离; RF 模型散点分布较分散, 部分样本偏差较大; PLS 模型散点离散显著且测试集无规律。基于 SPXY 划分的 70 个训练集与 30 个测试集样本, 采用训练集决定系数 R_c^2 、测试集决定系

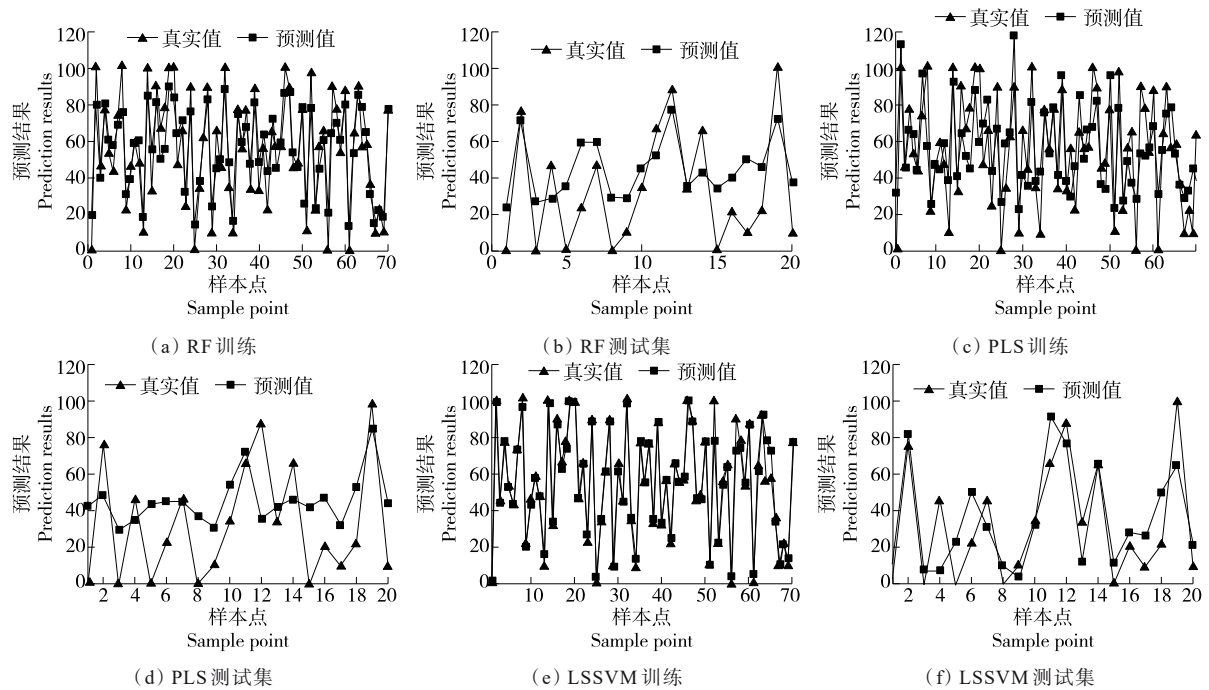


图7 RF、PLS和LSSVM 3种模型预测结果散点图

Figure 7 Scatter plots of prediction results for RF, PLS, and LSSVM models

数 R_p^2 、平均绝对误差(MAE)指标量化模型性能,结果见表2。

表2 不同模型训练集和测试集评价指标对比表

Table 2 Comparison of evaluation metrics between training and test sets of different models

模型	R^2		MAE	
	R_c^2	R_p^2	R_c^2	R_p^2
RF	0.833 20	0.398 14	10.268 0	21.736 3
PLS	0.545 59	0.141 28	16.726 4	25.616 9
LSSVM	0.972 34	0.626 72	2.381 5	15.712 7

由表2可知,LSSVM综合性能最优,训练集 R_c^2 、MAE表现突出,测试集 R_p^2 、MAE也最佳,仅存在过拟合风险;RF训练集拟合次之,但测试集泛化一般,可作辅助模型;PLS训练集拟合差、测试集 R_p^2 仅0.141 28,建模效果最差。

2.5.2 多模型建立 基于表2中各模型测试集 R_p^2 与MAE实测指标,按式(2)~式(6)计算各模型综合权重且列出权

重分配全过程的关键中间数据,结果见表3。

由表3可知,LSSVM因测试集 R_p^2 最高、MAE最低,综合性能最优,权重最高,契合性能优先原则保障预测精度;RF的次之,其权重依托强抗噪声能力可弥补LSSVM局部误差,提升融合模型稳定性;PLS因 R_p^2 极低、MAE最高,综合性能0,权重0排除其干扰,达到抑制弱模型的分配目标。

2.5.3 多模型计算 软投票融合模型预测值为各单模型预测值加权求和,根据表3最终权重,LSSVM权重 $W_{LSSVM}=0.710 7$ 、RF权重 $W_{RF}=0.289 3$ 、PLS权重 $W_{PLS}=0$,融合预测值按式(9)计算。

$$\hat{y}_{\text{融合}} = W_1 \hat{y}_{LSSVM} + W_2 \hat{y}_{RF}, \quad (9)$$

式中:

$\hat{y}_{\text{融合}}$ ——融合预测值;

$\hat{y}_{LSSVM}$ ——最小二乘支持向量机融合;

$\hat{y}_{RF}$ ——随机森林融合;

$W_1 = W_{LSSVM}$;

表3 各模型最终权重及中间过程数据

Table 3 Final weights and intermediate process data of each model

模型	测试集 R_p^2	测试集 MAE	$Y_i=1/MAE$	$X_{\text{norm},i}(R_p^2 \text{归一化})$	$Y_{\text{norm},i}(1/MAE \text{归一化})$	综合性能得分 S_i	最终权重 W_i
LSSVM	0.626 72	15.712 7	0.063 6	1.000	1.000	1.000	0.710 7
RF	0.398 14	21.736 3	0.046 0	0.529	0.285	0.407	0.289 3
PLS	0.141 28	25.616 9	0.039 0	0.000	0.000	0.000	0.000 0

$$W_2 = W_{RF}$$

软投票融合模型的提升/降低幅度按式(10)进行计算。

$$R = \frac{M_{SV} - M_{LSSVM}}{M_{LSSVM}} \times 100\%, \quad (10)$$

式中:

R ——软投票融合模型的提升或降低幅度,%;

M_{SV} ——软投票融合模型的评价指标值;

M_{LSSVM} ——LSSVM 单个模型的评价指标值。

结合表 2 单模型训练集与测试集预测指标:取训练集样本失水率平均值 55.624% 和测试集样本失水率平均值 37.151% 代入式(9),分别得出融合预测值 $Y_{R_c^2 \text{融合}}$ 为 55.624%, $Y_{R_p^2 \text{融合}}$ 为 37.151%。将训练集和测试集样本实际失水率与融合预测值代入式(7)、式(8)完成指标核算,结果见表 4。

表 4 软投票融合模型与单模型性能对比表

Table 4 Comprehensive comparison table of performance between soft voting fusion model and single model

模型	训练集 R_c^2	测试集 R_p^2	测试集 MAE	$R_c^2 - R_p^2$
软投票融合模型	0.896 50	0.785 20	11.364 0	0.111 3
LSSVM(最优单一模型)	0.972 34	0.626 72	15.712 7	0.345 6

由表 4 可知,融合模型 R_c^2 为 0.896 5,较 LSSVM 的略有下降,但仍保持高拟合度; R_p^2 为 0.785 2,较 LSSVM 的提升 25.3%;MAE 为 11.364 0,较 LSSVM 的降低 27.6%;过拟合程度($R_c^2 - R_p^2$)仅 0.111 3,较 LSSVM 的降低 67.8%,泛化稳定性更强。融合模型较 LSSVM 在 R_p^2 、MAE、过拟合程度上分别实现 25.3% 提升、27.6% 降低、67.8% 降低, LSSVM 训练拟合优但过拟合严重,单模型难满足检测需求。

2.6 模型性能分析

2.6.1 单模型性能 各模型中,LSSVM 综合性能最优,训练集 R_c^2 为 0.972 34,测试集 R_p^2 为 0.626 72,但存在明显过拟合风险;RF 泛化能力一般,可提供特征解释支撑;PLS 因无法适配非线性关联,性能较差。单模型的局限性进一步凸显了融合建模的必要性。

2.6.2 多模型融合性能 多模型软投票融合模型通过双维度指标权重分配,测试集 R_p^2 与 MAE 可实现对单模型性能的优化,对青提失水率预测精准度与误差稳定性显著提升,过拟合风险大幅缓解。这利于权重分配策略——LSSVM 权重 0.710 7 主导预测方向,RF 权重 0.289 3 补充抗干扰能力,PLS 因性能不足权重归零排除,实现训练拟合与泛化能力的平衡。融合模型基于 100 条小样本光谱

数据可实现稳定性能,解决单模型在小样本+强干扰场景下易过拟合的技术痛点,验证了该融合策略的实用性。

2.6.3 样本量对模型稳定性的影响

(1) 数据分布代表性局限。100 条数据虽覆盖 0~100% 失水率范围,但对于青提在实际贮藏运输过程中可能发生的不同环境如温度变化、湿度差异、机械损伤等,样本的环境适应性差,因此模型不能对面临极端失水状态或不同环境下的青提样本进行准确预测。

(2) 模型鲁棒性受限。小样本数据集中的噪声或异常值对模型参数估计影响较大,尽管通过 SNV 预处理和 SPA 特征筛选削弱了小样本干扰,但仍难以完全消除小样本导致的参数估计偏差,导致模型不同批次、不同产地的青提样本泛化能力较差。

(3) 算法优化空间受限。部分复杂机器学习算法如深度学习模型或算法更精细的集成策略如多层堆叠融合需要更大规模数据支撑以挖掘深层特征,而小样本量限制了此类算法的应用与优化,也无法保证模型在大样本场景下的稳定性。

2.6.4 与传统集成学习方法的实质差异 试验提出的多模型软投票融合方法与传统集成学习方法(Bagging、Boosting)存在显著区别:① 结构设计差异。传统 Bagging 等权投票易受弱模型干扰、Boosting 迭代权重对异常值敏感^[7,29],而试验基于 R_p^2 与 MAE 可实现客观赋权,使优质模型决定。同时,且仅选用 3 种异构基模型进行简洁加权,避免了大规模集群与复杂迭代,在不降低可解释性的基础上大大降低了计算费用。② 性能表现差异。传统方法易过拟合或精度不高,而试验方法使用互补抑制过拟合, $R_c^2 - R_p^2$ 差值较 Bagging 降低 74.4%;基于 100 条小样本数据,测试集 R_p^2 较传统 Bagging 提升 97.2%,突破了大规模数据的限制。③ 适用场景差异。针对果蔬光谱检测的复杂环境,利用 LED 诱导荧光技术实现精准检测,同时模型结构简洁,检测速度比传统 Stacking 提高 60% 以上,具备直接适配便携式设备。

综上,试验方法采用双维度权重机制,将结构简洁与性能稳定相结合,有效克服了小样本光谱检测的过拟合与精度瓶颈。

3 结论

研究通过特定技术路径构建了青提失水率多模型软投票融合预测模型。该模型以简洁软投票适配小样本场景并降低计算成本,创新双维度的权重分配机制解决传统权重分配弊端,搭建青提专属光谱检测体系破解通用技术适配难题,且在预测精度、误差控制及抗过拟合能力上显著优于单一模型和传统集成模型。为进一步提升模型稳定性与实用性,结合小样本的不足,未来将从以下方面进行拓展研究:① 扩大数据集规模与维度,将样本量扩

充至 300~500 条,包括不同产地、成熟度及贮藏条件的青提,增加可溶性固形物、硬度等指标,建立多指标综合预测模型,提升青提品质评估能力;② 优化数据增强策略,通过高斯噪声、波长偏移等技术获得虚拟样本,通过迁移学习降低数据采集成本,解决小样本问题并提升模型的跨品种适应性;③ 拓展应用场景验证,在贮藏仓库、冷链运输等真实场景下进行模型测试,对比便携式光谱仪与在线检测光谱系统的检测结果,验证模型的通用性。

参考文献

- [1] 高升,王巧华.基于可见/近红外透射光谱技术的红提糖度和含水率无损检测[J].中国光学,2021,14(3):566-577.
GAO S, WANG Q H. Non-destructive testing of red globe grape sugar content and moisture content based on visible/near infrared spectroscopy transmission technology[J]. Chinese Optics, 2021, 14(3): 566-577.
- [2] 黄华明,纪颖,刘晔,等.抑制甜春橘柚贮藏中失水的研究[J].果树学报,2021,38(3):403-410.
HUANG H M, JI Y, LIU Y, et al. Study on water loss control of postharvest 'Sweetspring' tangelo fruits[J]. Journal of Fruit Science, 2021, 38(3): 403-410.
- [3] 徐呈祥,郑福庆,马艳萍,等.贮藏温度对耐贮性不同的柑橘品种果皮蜡质含量及其化学组成的影响[J].食品科学,2021,42(13):223-232.
XU C X, ZHENG F Q, MA Y P, et al. Effect of storage temperature on peel wax content and chemical composition of citrus cultivars with different storability[J]. Food Science, 2021, 42(13): 223-232.
- [4] 张珮,王银红,江靖,等.便携式近红外光谱仪在果蔬品质定性和定量分析中的应用[J].食品科技,2020,45(5):287-292.
ZHANG P, WANG Y H, JIANG J, et al. Application of portable near-infrared spectrometer in qualitative and quantitative analysis of fruit and vegetable quality [J]. Food Science and Technology, 2020, 45(5): 287-292.
- [5] 关晔晴,王冬,李楠,等.基于近红外技术无损检测深州蜜桃果实内部品质[J].现代食品科技,2022,38(10):290-296.
GUAN Y Q, WANG D, LI N, et al. Near-infrared technology-based non-destructive detection of the internal quality of Shenzhou peaches[J]. Modern Food Science and Technology, 2022, 38(10): 290-296.
- [6] 朱金艳,朱玉杰,冯国红,等.基于近红外光谱技术联合极限学习机的蓝莓贮藏品质定量模型建立[J].食品与发酵工业,2022,48(16):270-276.
ZHU J Y, ZHU Y J, FENG G H, et al. Establishment of quantitative models for blueberry storage quality based on near infrared spectroscopy combined with extreme learning machine [J]. Food and Fermentation Industries, 2022, 48(16): 270-276.
- [7] 张晨,朱玉杰,冯国红.基于集成学习和近红外光谱的蓝莓贮藏品质检测方法研究[J].食品与发酵工业,2023,49(18):306-314.
ZHANG C, ZHU Y J, FENG G H. Research on storage quality detection method of blueberry based on ensemble learning and near-infrared spectroscopy[J]. Food and Fermentation Industries, 2023, 49(18): 306-314.
- [8] 刘广志.近红外光谱技术在农产品品质快速检测中的应用研究[J].食品安全导刊,2025(11):84-87.
LIU G Z. Research on the application of near-infrared spectroscopy technology in the rapid detection of agricultural product quality[J]. Journal of Food Safety Guide, 2025(11): 84-87.
- [9] 王永乾,周泉,龙怡帆.近红外光谱技术检测烟草水分含量中的重难点样本处理和模型性能改善研究[J].科学技术创新,2025(17):29-32.
WANG Y Q, ZHOU Q, LONG Y F. Study on the key and difficult sample processing and model performance improvement in the detection of tobacco moisture content using near-infrared spectroscopy technology[J]. Scientific and Technological Innovation Information, 2025(17): 29-32.
- [10] 贾利红,张国宏,王毅,等.基于集成建模方法的便携式近红外光谱仪酒醅成分检测研究[J].分析仪器,2022(5):13-19.
JIA L H, ZHANG G H, WANG Y, et al. Study on detection method of fermented grains composition by portable NIR spectrometer based on integrated modeling[J]. Analytical Instrumentation, 2022(5): 13-19.
- [11] TURGUT S S, AYVAZ H, ALI DOGAN M, et al. Detecting carob powder adulteration in cocoa using near and mid-infrared spectroscopy: a comprehensive classification and regression analysis[J]. Food Research International, 2025, 208: 116132.
- [12] 高升.基于近红外相机成像的红提串缺陷检测[J].食品与机械,2023,39(1):146-151.
GAO S. Defect detection method of red globe grapes bunches based on near infrared camera imaging[J]. Food & Machinery, 2023, 39(1): 146-151.
- [13] 王晨晨,翟明灿,李贺,等.基于可见/近红外透射光谱的亚健康水心苹果检测[J].食品与机械,2024,40(7):117-125,182.
WANG C C, ZHAI M C, LI H, et al. Detection of sub-healthy apples with watercore based on visible/near-infrared transmission spectroscopy[J]. Food & Machinery, 2024, 40(7): 117-125, 182.
- [14] 王丹,刘芳菲,纪晓峰.基于双通道改进卷积神经网络的新鲜葡萄品质检测分析[J].食品与机械,2024,40(12):89-94.
WANG D, LIU F F, JI X F. Fresh grape quality inspection based on dual-channel improved convolutional neural network [J]. Food & Machinery, 2024, 40(12): 89-94.
- [15] WANG Y T, ZHAO C J, TIAN H W, et al. NIRS-based detection advances in agriculture: data enhancement, characteristic wavelength selection and modelling techniques [J]. Spectrochimica Acta Part A: Molecular and Biomolecular

- Spectroscopy, 2025, 343: 126611.
- [16] 聂衍文, 杨佳晨, 文慧心, 等. 基于机器视觉的轻量化芒果表面缺陷检测[J]. 食品与机械, 2023, 39(3): 91-95, 240.
- NIE Y W, YANG J C, WEN H X, et al. Light weight detection of mango surface defects based on machine vision[J]. Food & Machinery, 2023, 39(3): 91-95, 240.
- [17] 王云鹏, 司海平, 宋佳珍, 等. 基于红外与可见光图像融合的苹果表面缺陷检测方法[J]. 食品与机械, 2021, 37(12): 127-131.
- WANG Y P, SI H P, SONG J Z, et al. Apple surface defect detection method based on fusion of infrared and visible images[J]. Food & Machinery, 2021, 37(12): 127-131.
- [18] 姜小刚, 朱明旺, 姚金良, 等. 基于近红外在线装置苹果糖度模型参数优化研究[J]. 光谱学与光谱分析, 2023, 43(1): 116-121.
- JIANG X G, ZHU M W, YAO J L, et al. Research on parameter optimization of apple sugar model based on near-infrared on-line device[J]. Spectroscopy and Spectral Analysis, 2023, 43(1): 116-121.
- [19] 陈至坤, 郭蕊, 程朋飞. 基于 LIF 技术的光谱特征提取在油类检测中的应用[J]. 激光与光电子学进展, 2020, 57(13): 133002.
- CHEN Z K, GUO R, CHENG P F. Application of LIF technology-based spectral feature extraction in oil detection[J]. Laser & Optoelectronics Progress, 2020, 57(13): 133002.
- [20] 刘梁晨, 杨瑞芳, 赵南京, 等. LED 诱导荧光的土壤多环芳烃快速检测方法[J]. 光学学报, 2023, 43(18): 302-312.
- LIU L C, YANG R F, ZHAO N J, et al. Rapid detection of polycyclic aromatic hydrocarbons in soil using LED induced fluorescence spectroscopy[J]. Acta Optica Sinica, 2023, 43(18): 302-312.
- [21] HARA R, ISHIGAKI M, KITAHAMA Y, et al. Excitation wavelength selection for quantitative analysis of carotenoids in tomatoes using Raman spectroscopy[J]. Food Chemistry, 2018, 258: 308-313.
- [22] 陈君蕊, 苏良敏, 成欣, 等. 桂花青脆李复合香型果酒的酿造工艺优化[J]. 食品工业科技, 2024, 45(2): 126-135.
- CHEN J R, SU L M, CHENG X, et al. Optimization of brewing process of *Osmanthus* and 'Qingcui' plum complex flavor wine[J]. Science and Technology of Food Industry, 2024, 45(2): 126-135.
- [23] 孔芳, 李永刚, 刘小红, 等. GB 5009.3—2016《食品安全国家标准 食品中水分的测定》的关键问题探讨及方法优化[J]. 食品安全质量检测学报, 2025, 16(13): 189-194.
- KONG F, LI Y G, LIU X H, et al. Exploration of key issues and method optimization of GB 5009.3—2016 National Food Safety Standard-Determination of moisture in foods[J]. Journal of Food Safety & Quality, 2025, 16(13): 189-194.
- [24] 王美, 石文杰, 杨雁, 等. 近红外光谱技术在快速检测辣椒粉吸光度中的应用研究[J]. 现代食品, 2024, 30(5): 177-182.
- WANG M, SHI W J, YANG Y, et al. Application research on rapid detection of chili powder absorbance by near infrared spectroscopy technology[J]. Modern Food, 2024, 30(5): 177-182.
- [25] 姜传礼, 赵健赞, 丁圆圆, 等. SPA 算法与机器学习的黄河源土壤水分反演[J]. 光谱学与光谱分析, 2023, 43(6): 1 961-1 967.
- JIANG C L, ZHAO J Y, DING Y Y, et al. Study on soil water retrieval technology of yellow river source based on SPA algorithm and machine learning[J]. Spectroscopy and Spectral Analysis, 2023, 43(6): 1 961-1 967.
- [26] DAI F W, ZHUANG Q Y, HUANG G, et al. Infrared spectrum characteristics and quantification of OH groups in coal[J]. ACS Omega, 2023, 8(19): 17 064-17 076.
- [27] JUNG E S, CHOE J H, YOO J, et al. Vibrational spectroscopy for microplastic detection in water: a review[J]. Applied Spectroscopy Reviews, 2025, 60(2): 193-214.
- [28] FUMOTO E, SATO S, KAWAMATA Y, et al. Determination of carbonyl functional groups in lignin-derived fraction using infrared spectroscopy[J]. Fuel, 2022, 318: 123530.
- [29] SOLOFF J A, BARBER R F, WILLETT R. Bagging provides assumption-free stability[EB/OL]. (2023-01-30) [2025-10-09]. <https://arxiv.org/abs/2301.12600>.