

基于机器学习与近红外光谱的高粱淀粉定量分析

李珂¹ 欧阳润坤¹ 戴奕杰² 李杰¹ 侯爱香¹ 杨雨¹ 李宗军¹

(1. 湖南农业大学食品科学技术学院, 湖南 长沙 410128; 2. 贵阳学院生物与环境工程学院, 贵州 贵阳 550005)

摘要: [目的] 针对高粱淀粉快速定量建立一种基于近红外光谱技术与机器学习算法的模型。[方法] 以酿酒高粱为原料, 通过近红外光谱仪器采集近红外漫反射光谱, 利用孤立森林算法剔除异常样本, 优化数据集后, 运用多种建模软件对比多种预处理方法及回归模型, 分析并建立高粱淀粉定量模型。[结果] 经孤立森林预处理后, 偏最小二乘回归法 (PLSR) 模型性能最优, 其校正集决定系数 (R_c^2) 为 0.993, 预测均方根误差 (RMSECV) 为 3.401。[结论] 该模型简便、快速、精准, 适用于高粱淀粉含量的定量分析。

关键词: 近红外光谱; 机器学习; 淀粉; 定量分析; 高粱

Quantitative analysis of sorghum starch using machine learning and near-infrared spectroscopy

LI Ke¹ OUYANG Runkun¹ DAI Yijie² LI Jie¹ HOU Aixiang¹ YANG Yu¹ LI Zongjun¹

(1. College of Food Science and Technology, Hunan Agricultural University, Changsha, Hunan 410128, China;
2. College of Biology and Environmental Engineering, Guiyang University, Guiyang, Guizhou 550005, China)

Abstract: [Objective] To establish a model for the rapid quantification of sorghum starch based on near-infrared spectroscopy and machine learning algorithms. [Methods] Brewing sorghum was used as the raw material. Near-infrared diffuse reflectance spectra were collected using a near-infrared spectrometer. The Isolation Forest algorithm was employed to remove outlier samples and optimize the dataset. Subsequently, multiple preprocessing methods and regression models were compared using various modeling software to analyze and establish a quantitative model for sorghum starch. [Results] After preprocessing with the Isolation Forest algorithm, the partial least squares regression (PLSR) model exhibited the best performance, with a coefficient of determination for the calibration set (R_c^2) of 0.993 and a root mean square error of cross-validation (RMSECV) of 3.401. [Conclusion] The model is simple, rapid, and accurate, and is suitable for the quantitative analysis of starch content in sorghum.

Keywords: near-infrared spectroscopy; machine learning; starch; quantitative analysis; sorghum

在全球粮食生产格局中, 高粱凭借其广泛的种植面积, 被确立为世界第五大核心粮食作物, 其栽培规模紧随玉米、水稻、小麦及大麦之后^[1]。高粱的用途广泛, 可以用于酿酒、食用、饲料和生物能源生产等^[2]。其主要成分为淀粉、蛋白质、脂肪和单宁等。不同品种和场地的 高粱, 其成分组成、理化指标和酿造特性存在差异, 因此对白酒发酵过程也会产生不同影响^[3]。高粱作为主要的能量原料, 在快速检测分析中不仅可以把控原料质量, 还可以指

导原料的高效采购和贮藏, 因此高粱原料颗粒模型的建立非常关键^[4]。Bian 等^[5]通过监测澳大利亚高粱和东北高粱在发酵过程中的理化性质、风味成分及微生物群落结构变化, 为优化澳大利亚高粱在小曲酒生产中的利用提供了理论依据。在酿酒需求条件下, 高粱中的淀粉含量尤为重要。但在酿酒高粱快速检测分析中, 基于机器学习的高效检测领域仍缺乏大量研究。

根据美国材料与试验协会 (ASTM) 的界定, 近红外光

基金项目: 湖南省重点研究计划 (编号: 2022NK2040)

通信作者: 李宗军 (1967—), 男, 湖南农业大学教授, 博士。E-mail: hnlizongjun@163.com

收稿日期: 2025-09-17 改回日期: 2026-03-22

引用格式: 李珂, 欧阳润坤, 戴奕杰, 等. 基于机器学习与近红外光谱的高粱淀粉定量分析[J]. 食品与机械, 2026, 42(6): 60-66.

Citation: LI Ke, OUYANG Runkun, DAI Yijie, et al. Quantitative analysis of sorghum starch using machine learning and near-infrared spectroscopy[J]. Food & Machinery, 2026, 42(6): 60-66.

谱(NIRs)是指波长介于780~2 526 nm的电磁波,该区域也是非可见光区域^[6]。近红外光谱技术现已成为食品、医药和农业等领域不可或缺的核心分析技术之一,这主要得益于其高效、快速、无损检测与绿色环保的特质,这些特质共同促成了其在全球范围内的广泛应用^[7-10]。Ning等^[11]成功实现了茶园土壤有机质和总氮含量的快速无损评估,以及土壤肥力水平的分类。姚力等^[12]运用近红外光谱法实现了简便、快速、安全、实用、准确,适用于含高纤维素、木质素物料的水分含量的快速测定。赵玉霞等^[13]采用UV、NIR及紫外—近红外中级数据融合光谱结合反向传播神经网络法建立了一种快速、无损、高效的“互助”青稞酒判别分类模型。综上,运用近红外光谱可以进行快速准确的检测与判别。

机器学习在本质上是植根于数据驱动的交叉学科,其通过整合计算机科学的计算能力与统计学的推断理论,从构建数据中学习并做出预测或决策的模型,使得机器能够通过数据学习完成特定的任务^[14]。机器学习算法的选择多样化,涵盖支持向量机、随机森林、K最近邻、梯度回归、逻辑回归、线性回归、朴素贝叶斯模型等传统算法,以及深度学习算法^[15]。黄林峰等^[16]基于高光谱特征波段的监测模型拟合效果良好,实现了梨树果实膨大期和成熟期氮磷钾含量的准确监测。宋少忠等^[17]采用高光谱成像技术结合机器学习算法在高粱品种鉴别中的有效性,实现了快速精确的高粱品质检测。研究拟基于机器学习算法与近红外光谱技术及孤立森林算法处理,将化学计量学方法与样本预处理技术结合,系统探究预处理策略、光谱波段选择及建模算法对定量分析性能的影响,建立一种适用于高粱淀粉含量的高效无损检测模型,以期为高粱的品质评价及后续生产研究提供依据。

1 材料与方法

1.1 材料、试剂与仪器

1.1.1 材料与试剂

高粱样品:186份,网购或生产厂商赠送,具体信息见表1;

盐酸、氢氧化钠、乙酸铅、硫酸钠、浓硫酸、 α -萘酚、石油醚(60~90℃)等:分析纯,成都市科隆化学品有限公司。

表1 高粱样品信息

Table 1 Sample information of sorghum

样品产地	样品数量	样品产地	样品数量
山东济南	10	陕西榆林	10
山东聊城	10	陕西延安	10
辽宁朝阳	20	山西忻州	10
辽宁沈阳	15	贵州贵阳	10
辽宁锦州	15	云南楚雄	15
吉林长春	10	美国	11
贵州安顺	10	澳洲	10
贵州遵义	20		

1.1.2 主要仪器设备

傅立叶变换近红外光谱仪:NIRM54型,附带NIRareOperator光谱分析软件、NIRCal 5.4-化学计量学软件、Manangementconsole软件,瑞士BUCHI公司;

电子天平:ATY124型,日本岛津公司;

电陶炉:C22-CS03型,浙江苏泊尔股份有限公司;

恒温水浴锅:XMTD-7000型,永光明医疗仪器有限公司;

多功能粉碎机:2500C型,永康市红太阳机电有限公司。

1.2 方法

1.2.1 样品前处理 所有高粱均经过筛选,确保无发霉或破损,粉碎,过60目筛,样品经过近红外光谱扫描,采用酸水解法测定其淀粉含量。样品经粉碎、脱脂及乙醇溶液洗涤去除可溶性糖类后,在加热条件下与盐酸进行回流,使淀粉充分水解为葡萄糖。水解液经中和、除铅等净化处理后,以碱性酒石酸铜溶液为氧化剂,通过滴定法精确测定其还原糖含量,并同步进行空白试验以校正试剂干扰,最终折算得到样品中淀粉含量。

1.2.2 样品近红外光谱图采集条件 检测前将样品平铺摆匀至近红外仪样品杯中,输入待测样品信息,启动仪器,自动完成光谱采集,通过NIRareOperator光谱分析软件输出光谱采集结果。光谱采集范围为12 000~4 000 cm^{-1} ,仪器分辨率为8 cm^{-1} ,扫描次数为32次。分别采用内置背景与外部背景作为参比,所有样本均重复测定3次,计算平均光谱以用于后续建模。

1.2.3 光谱扫描及光谱数据处理 采用GB 5009.9—2023测定的淀粉含量为输入的手工值,建立定标模型赋值。去除异常样,最终得到186个样品光谱,样本集按148:38的比例划分为校正集与验证集。其中,校正集用于模型构建,验证集则用以综合评价模型的稳健性与预测精度。利用Visual Studio Code去趋势光谱(DT)、均值中心化(CT)、数据标准化(SS)、光谱经平滑处理(SG)、移动平均(MA)、最大归一化(MMA)、一阶导数(D1)、二阶导数(D2)、多元散射校正(MSC)等光谱预处理后选出最佳预处理方法^[18]。

为了系统比较和评估全波长模型与特征波长模型的性能,采用校准均方根误差(RMSEC)、预测均方根误差(RMSEP)、交叉验证均方根误差(RMSECV)、校准相关系数(R_c)、预测相关系数(R_p)和决定系数(R^2)多个统计指标进行评估,相关计算见式(1)~式(6)。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad (1)$$

$$R_p = \frac{\sum_{i=1}^m (y_i - \bar{y})(\dot{y}_i - \dot{\bar{y}})}{\sqrt{\sum_{i=1}^m (y_i - \bar{y})^2 \sum_{i=1}^m (\dot{y}_i - \dot{\bar{y}})^2}}, \quad (2)$$

$$R_c = \frac{\sum_{i=1}^n (y_i - \bar{y})(\dot{y}_i - \dot{\bar{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\dot{y}_i - \dot{\bar{y}})^2}}, \quad (3)$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}}, \quad (4)$$

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (5)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m-1}}, \quad (6)$$

式中:

y_i ——真实值;

$\hat{y}_i$ ——预测值;

m ——样本集的数量;

n ——校准样本集的数量。

RMSEC、RMSEP 和 RMSECV 均为越接近于 0 越好, RMSEC 反映模型对训练数据的拟合精度, RMSEP 反映模型对未知数据的预测能力。二者差值小意味着模型不仅在训练数据上表现良好, 还能准确预测新数据, 说明模型泛化能力强、稳定性高。 R^2 反映了因变量的总变异中, 由所建模型解释的部分所占的比例。 R^2 越接近 1, 说明相关程度越高。

1.3 数据处理

1.3.1 机器学习建模方法 偏最小二乘回归法(PLSR)是一种线性的、有监督的多元校正方法, 具有广泛的适用性和高效的模型建立能力。该方法在统计学和化学领域具有广泛的适用性, 其优势在于模型的简洁性和强大的预测能力, 尤其是在面对小样本量时仍能建立稳定可靠的模型^[19]。PLS 法因其在统计分析和多元数据处理中的独特优势, 已成为化学计量学和相关领域的标准分析工具, 是目前应用最为广泛的多元校正方法。对于定量分析, 为了比较和评估全波长模型以及特征波长模型, 建立了多种 PLSR 模型, 包括全波长 PLSR、遗传算法偏最小二乘回归(GA-PLSR)和变量组合种群分析算法偏最小二乘回归(VCPL-PLSR)模型。

随机森林(RF)是一种基于决策树的集成学习方法, 其核心思想是通过袋装法和随机子空间法, 使其生成多棵决策树, 并将这些树的预测结果进行集成, 以此提升模型的泛化能力和预测稳定性。在随机森林算法中, 每棵决策树的生成采用分类与回归树(CART)算法, 该算法通过最小化基尼指数来选择最优分裂特征。CART 算法不仅可以用于分类任务, 也可以应用于回归分析。所以随机森林算法在处理分类和回归任务时表现出较高的预测准确性和模型稳定性, 成为机器学习领域中被广泛应用的重要算法^[20]。

岭回归(RIDGE)是一种用于解决多重共线性问题的线性回归技术。其核心是通过引入 L2 正则化项来收缩模型参数, 减少过拟合以及提升模型泛化能力。多重共线性描述的是回归模型中自变量之间存在高度线性相关的现象, 使普通最小二乘法(OLS)的回归系数估计值误差增大导致稳定性下降, 从而影响统计推断的可靠性, 使得模

型的预测性能下降。岭回归的主要优势包括有效缓解多重共线性问题, 能够有效减小估计量的方差; 同时, 该方法能够在高维数据环境中保持良好的预测性能, 有效避免过拟合问题。

套索回归(LASSO)是一种有效的特征选择方法, 能够在全光谱区间数据中进行特征选择和维度降低。通过优化该目标函数, LASSO 算法能够自动将与响应变量相关性较低的变量的系数估计值压缩为 0, 从而实现特征的选择和剔除, 最终被保留的非零系数对应的变量即为所需的特征变量^[21]。该方法不仅能够有效降低数据维度, 还为模型的简化和解释提供了重要支持, 是统计学习和数据分析中一种重要的工具。

梯度提升决策树(GBDT)是机器学习领域中一种高效且强大的集成学习方法。它通过迭代地添加决策树以逐步降低预测误差, 从而在各种任务中, 尤其是回归和分类问题上表现出色。梯度提升是一种迭代的机器学习算法, 其核心思想是利用前一个模型的残差作为当前模型的学习目标, 通过不断添加决策树, 逐步降低训练数据的损失函数值, 直至达到预设的停止条件。

1.3.2 模型样本预处理方法 孤立森林(IF)是一种基于集成学习的异常检测算法, 特别适用于连续数据的异常检测。孤立森林异常检测的思路是利用异常数据在多个评价方法下的“孤立性”检测异常^[22]。孤立森林通过随机构建二叉树对数据空间进行递归划分。异常的样本因其特征疏离, 其被孤立至叶节点所需的平均路径长度更短, 该长度直接量化为异常得分。

1.3.3 数据处理软件 试验的主要算法和模型参数优化及回归模型构建等均在 Visual Studio Code 和 MATLAB R2021 软件平台中实现。

2 结果与分析

2.1 光谱数据分析

由图 1~图 2 和表 2~表 3 可知, CT 预处理过程优化了原始光谱的数据质量, 使光谱曲线更为光滑, 反映了样本组成的关键谱图特征, 从而放大了不同样本间的差异性。模型的最佳 R_c^2 值为 0.951, R_p^2 值为 0.888, RMSEC 值为

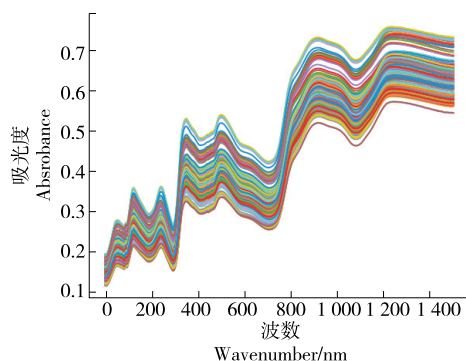


图 1 高粱样品的原始光谱图

Figure 1 NIR raw spectra of starch samples

表2 不同预处理方法对PLSR建模的影响

Table 2 Effect of different preprocessing methods on PLSR modeling

预处理	R_c^2	R_p^2	RMSEC	RMSEP	RMSECV
CT	0.951	0.888	2.104	3.569	3.676
D1	0.894	0.831	3.097	4.398	4.882
D2	0.965	0.766	1.776	5.167	4.600
DT	0.936	0.869	2.397	3.870	4.068
MA	0.959	0.870	1.925	3.852	3.914
MMS	0.959	0.870	1.925	3.852	3.914
SS	0.959	0.870	1.925	3.852	3.914
MSC	0.939	0.850	2.341	4.143	3.688
SG	0.952	0.855	2.074	4.065	3.811

2.104, RMSEP 值为 3.569, RMSECV 值为 3.676。说明利用近红外漫反射光谱进行淀粉含量检测,全波段均能观察到明显的特征峰。

2.2 基于孤立森林算法的模型样本优化

由图3可知,分布稀疏的样本相较于正常样本更易于在迭代划分过程中优先分离并被提早分配到叶节点中。相比之下,正常样本通常表现出较高的相似性和集群特

表3 多种预处理方法叠加对PLSR建模的影响

Table 3 Effect of multiple preprocessing methods superimposed on PLSR modeling

预处理	R_c^2	R_p^2	RMSEC	RMSEP	RMSECV
CT+MSC	0.939	0.850	2.341	4.144	3.688
CT+SG	0.949	0.864	2.142	3.933	3.911
MSC+SNV	0.955	0.874	2.008	3.789	3.613
MSC+SG	0.943	0.849	2.275	4.156	3.864

性,因此需要经过多次划分才能被成功孤立。由于异常样本更早被孤立,其所在叶节点到根节点的路径长度相对更短,而正常样本的路径长度则相对较长。通过对样本点在孤立树中的路径长度进行分析,可以有效地识别出异常样本。

运用 MATLAB R2021 软件将高粱光谱特征输入 Visual Studio Code 中预处理,在构建的孤立森林模型中进行异常分数计算,并结合异常分数阈值对预测高粱淀粉含量进行异常判定,光谱数量由 186 条变为 149 条,剔除后的光谱信息见表4,导出处理好后的光谱。

由表5可知,应用孤立森林算法完成异常光谱剔除

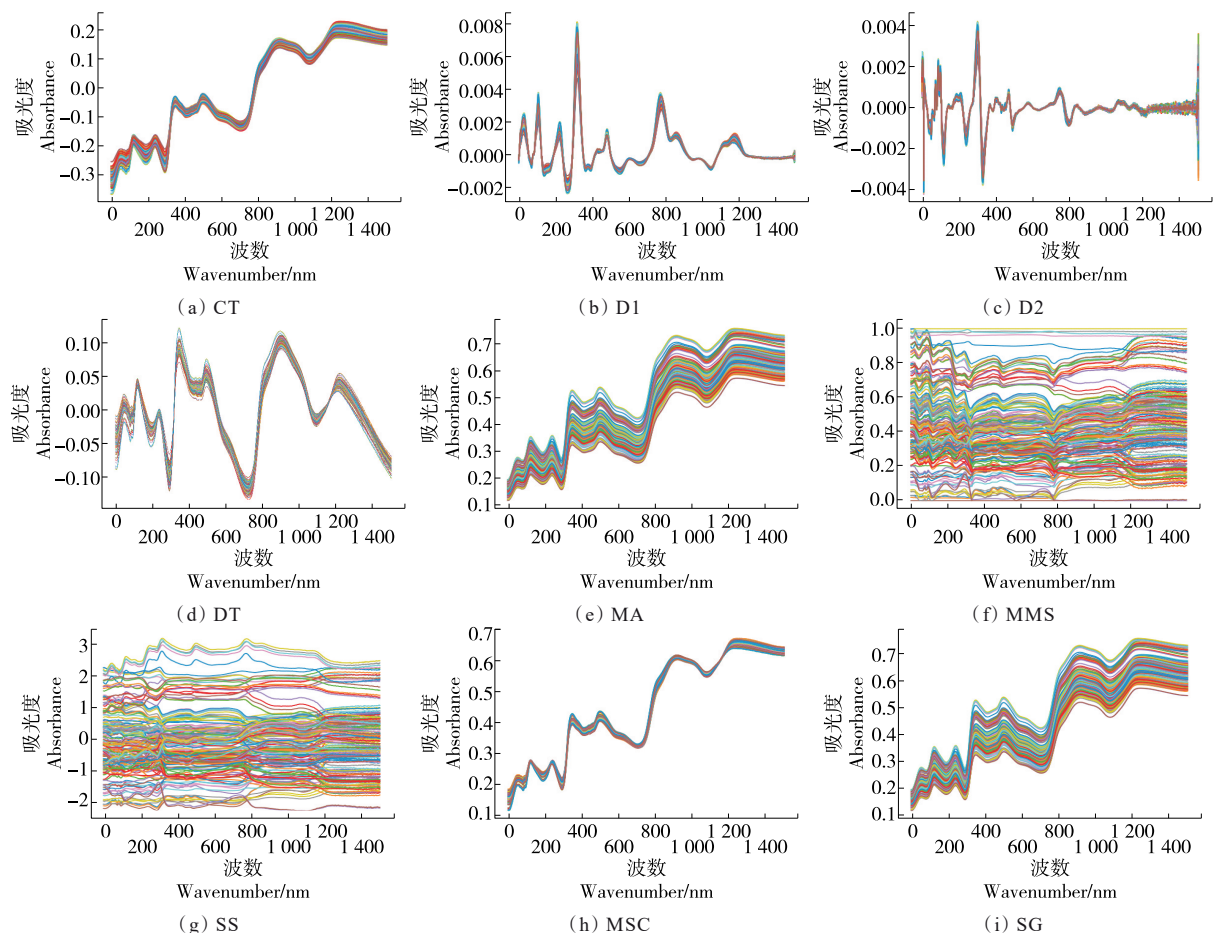


图2 高粱样品预处理后的光谱图
Figure 2 NIR of sorghum samples after pretreatment

后,依次采用PLSR、RF、RIDGE、LASSO以及GBDT构建预测模型。各模型在训练集上的决定系数(R^2)由原先的0.951,0.945,0.733,0.620,0.999分别变为0.993,0.916,0.768,0.715,0.999,测试集决定系数(R^2)由0.888,0.649,0.540,0.553,0.857变为0.903,0.728,0.656,0.564,0.329,由此可得在孤立森林算法处理后的模型相关程度更高。

2.3 基于机器学习回归模型建立及定量预测结果对比

2.3.1 预测结果对比 由图4和表6可知,PLSR和RF模型均能准确定性检测出高粱中的淀粉,且具有较好的定量预测能力;RIDEG和LASSO模型表现较差;PLSR模型表现最佳,其RMSEC为0.776,RMSEP为3.176, R_c^2 为0.993, R_p^2 为0.903。

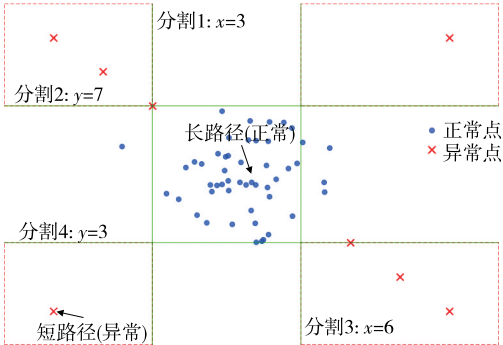


图3 孤立森林异常检测方法
Figure 3 Isolation Forest anomaly detection method

表4 孤立森林算法剔除光谱
Table 4 Isolated Forest algorithm for spectral thinning

品种	化学检测值/%	合理范围/%	绝对差值	相对误差/%	品种	化学检测值/%	合理范围/%	绝对差值	相对误差/%
山东济南	57.96	60~70	2.04	3.40	陕西延安	58.27	60~70	1.73	2.88
山东济南	57.88	60~70	2.12	3.53	贵州贵阳	86.71	70~80	6.71	8.39
山东聊城	81.16	70~80	1.16	1.45	山西忻州	60.81	60~70	0.81	1.35
山东聊城	80.91	70~80	0.91	1.14	山西忻州	60.51	60~70	0.51	0.85
辽宁朝阳	58.56	60~70	1.44	2.40	山西忻州	59.85	60~70	0.15	0.25
辽宁朝阳	59.13	60~70	0.87	1.45	山西忻州	59.12	60~70	0.88	1.47
辽宁沈阳	49.22	60~70	10.78	17.97	山西忻州	58.94	60~70	1.06	1.77
辽宁沈阳	49.19	60~70	10.81	18.02	云南楚雄	57.81	60~70	2.19	3.65
辽宁盘锦	60.51	60~70	0.51	0.85	云南楚雄	59.34	60~70	0.66	1.10
贵州安顺	55.92	60~70	4.08	6.80	云南楚雄	59.13	60~70	0.87	1.45
贵州遵义	82.97	70~80	2.97	3.71	云南楚雄	58.35	60~70	1.65	2.75
贵州遵义	81.51	70~80	1.51	1.89	美国	52.51	60~70	7.49	12.48
贵州遵义	81.77	70~80	1.77	2.21	美国	54.81	60~70	5.19	8.65
贵州遵义	82.85	70~80	2.85	3.56	美国	53.15	60~70	6.85	11.42
陕西榆林	61.33	60~70	2.04	2.22	美国	54.31	60~70	5.69	9.48
陕西榆林	60.03	60~70	0.03	0.05	美国	50.81	60~70	9.19	15.32
陕西延安	59.72	60~70	0.28	0.47	澳洲	52.41	60~70	7.59	12.65
陕西延安	59.61	60~70	0.39	0.65	澳洲	56.43	60~70	3.57	5.95
陕西延安	57.96	60~70	2.04	3.40	澳洲	57.56	60~70	2.44	4.07

2.3.2 检测结果的外部验证 采用验证样品集进行定标模型准确性验证,PLSR模型建模,然后进行外部验证。为验证高粱淀粉检测模型的准确性,随机取20份样本不参与建模,作为预测集样本。淀粉真实值参照GB 5009.9—

2023中的酸水解法测定,分析结果见表7。由表7可知,高粱淀粉含量的相对误差均值为1.46%,说明基于偏最小二乘回归法算法的高粱淀粉预测值与真实值的吻合度较高,模型效果较好,表明随机选择的外部验证样品具有代表性,且基于偏最小二乘回归法算法的高粱淀粉近红外

表5 孤立森林算法前后建模比较
Table 5 Comparison of modeling performance before and after applying isolated forest algorithm

模型	预处理前		预处理后	
	R_c^2	R_p^2	R_c^2	R_p^2
PLSR	0.951	0.888	0.993	0.903
RF	0.945	0.649	0.916	0.728
RIDGE	0.733	0.540	0.768	0.656
LASSO	0.620	0.553	0.715	0.564
GBDT	0.999	0.857	0.999	0.329

表6 机器学习算法验证与评价
Table 6 Machine learning algorithm validation and evaluation

模型	R_c^2	R_p^2	RMSEC	RMSEP	RMSECV
PLSR	0.993	0.903	0.776	3.176	3.401
RF	0.916	0.728	2.766	5.323	7.282
RIDEG	0.768	0.656	4.710	5.568	5.286
LASSO	0.715	0.564	4.689	3.820	6.529
GBDT	0.999	0.329	0.268	9.066	5.235

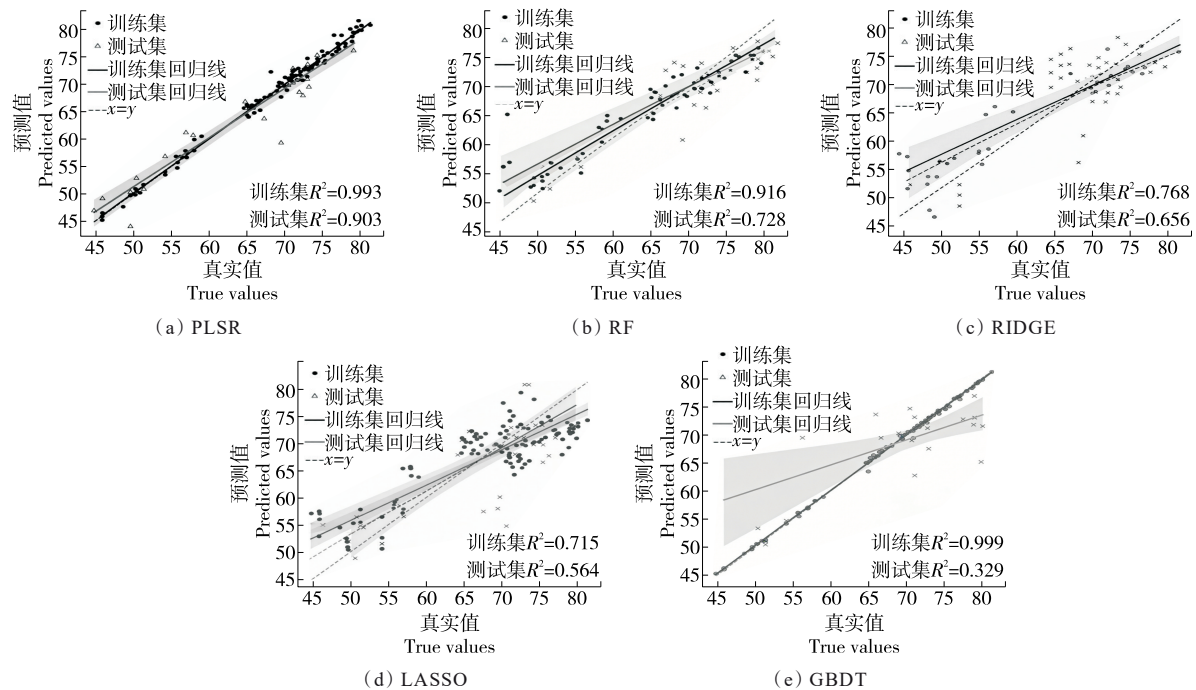


图4 基于5种机器学习算法模型的验证拟合图

Figure 4 Validated fitted plots based on four machine learning algorithm models

表7 外部验证数据

Table 7 External validation data

样品编号	国标检测数据/%	光谱检测数据/%	绝对差值	相对误差/%	样品编号	国标检测数据/%	光谱检测数据/%	绝对差值	相对误差/%
1	49.52	48.23	1.29	2.68	11	64.89	66.51	1.62	2.44
2	77.01	77.74	0.73	0.94	12	64.21	64.71	0.49	0.76
3	71.45	72.78	1.33	1.83	13	70.71	72.41	1.70	2.35
4	76.18	77.66	1.48	1.90	14	69.97	71.30	1.32	1.86
5	79.66	78.87	0.79	1.01	15	65.85	66.40	0.55	0.83
6	61.43	62.64	1.20	1.92	16	74.71	75.02	0.31	0.41
7	65.95	66.96	1.01	1.51	17	71.57	72.58	1.01	1.39
8	65.20	64.21	0.99	1.54	18	72.41	73.33	0.92	1.25
9	64.91	64.44	0.48	0.74	19	72.39	71.67	0.72	1.01
10	70.80	71.63	0.83	1.16	20	69.01	67.90	1.11	1.64

模型预测结果准确度较高。

3 结论

研究成功构建了一种基于近红外光谱与机器学习的高粱淀粉含量快速定量分析模型。结果表明,均值中心化预处理方法在保留有效光谱信息方面表现最佳,其偏最小二乘回归法模型预测性能显著优于其他复杂预处理方法。通过引入孤立森林算法进行数据清洗,有效剔除了19.9%的异常样本,使偏最小二乘回归法模型的校准相关系数从0.951提升至0.993,凸显了数据质量在光谱建模中的关键作用。在5种机器学习算法的系统比较中,偏最小二乘回归法模型展现出最优的预测能力与稳健性,为性能优越的模型。值得注意的是,梯度提升决策树模型

出现严重的过拟合现象,提示其在小样本光谱建模中需谨慎选择复杂算法。最终外部验证显示,模型预测值与标准方法测定值间的平均相对误差仅为1.46%,满足实际检测需求。未来研究将进一步扩大样本多样性,探索卷积神经网络与深度学习算法在复杂组分分析中的潜力。

参考文献

- [1] 赵珊珊, 郭志强, 朱立勋, 等. 高粱高亲和硝酸盐转运蛋白 *NRT2/3* 基因家族鉴定、表达与 DNA 变异分析[J]. 生物工程学报, 2023, 39(7): 2 743-2 761.
- ZHAO S S, GUO Z Q, ZHU L H, et al. Identification, expression and DNA variation analysis of high affinity nitrate transporter *NRT2/3* gene family in *Sorghum bicolor*[J]. Chinese Journal of Biotechnology, 2023, 39(7): 2 743-2 761.

- [2] XIN Z G, WANG M L, CUEVAS H E, et al. Sorghum genetic, genomic, and breeding resources[J]. *Planta*, 2021, 254(6): 114.
- [3] HASHEMI S S, ABBASI-RIYAKHUNI M, DENAYER J F M, et al. Efficient bioremediation of distillery and dairy wastewaters: a three-stage biorefinery for high-quality aquaculture feed and bioenergy generation[J]. *Process Safety and Environmental Protection*, 2023, 180: 566-574.
- [4] 向娜娜, 王晓琼, 孙美珍, 等. 基于近红外光谱技术快速测定高粱中淀粉及单宁含量的研究[J]. *中国饲料*, 2024(23): 345-349, 366.
- XIANG N N, WANG X Q, SUN M Z, et al. Research on rapid determination of starch and tannin content in sorghum based on near-infrared spectroscopy technique[J]. *China Feed*, 2024(23): 345-349, 366.
- [5] BIAN M H, FANG Y L, YANG K L, et al. Comparative analysis of microbial communities and flavor compounds in fermented grains from different sorghum varieties used in Sichuan Xiaoqu liquor[J]. *LWT-Food Science and Technology*, 2025, 222: 117640.
- [6] 陆婉珍. 现代近红外光谱分析技术[M]. 2版. 北京: 中国石化出版社, 2007: 1-11.
- LU W Z. Modern near infrared spectroscopy analytical technology [M]. 2nd ed. Beijing: China Petrochemical Press, 2007: 1-11.
- [7] 乔宁, 饶敏, 黄雪媛, 等. 基于便携式近红外光谱仪和随机森林方法快速鉴别蜂蜜品种[J]. *中国口岸科学技术*, 2024, 6(8): 75-80.
- QIAO N, RAO M, HUANG X Y, et al. Rapid identification of honey varieties based on portable near-infrared spectroscopy and random forest algorithm[J]. *China Port Science and Technology*, 2024, 6(8): 75-80.
- [8] 王雪, 张广月, 马铁民, 等. 基于近红外光谱的灌浆期玉米籽粒水分定量分析通用模型[J]. *农业工程学报*, 2025, 41(8): 291-300.
- WANG X, ZHANG G Y, MA T M, et al. Generalized model for the quantitative moisture analysis of maize grains during filling stage based on near-infrared spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2025, 41(8): 291-300.
- [9] 张云霞, 余佑, 李运通, 等. 基于近红外光谱法的馥郁香型白酒基酒中 4 种主要有机酸检测模型构建[J]. *食品与机械*, 2025, 41(4): 72-80.
- ZHANG Y X, YU J, LI Y T, et al. Detection modelling of four major organic acids in the base wine of Fuyuxiangxing crude Baijiu based on near-infrared spectroscopy[J]. *Food & Machinery*, 2025, 41(4): 72-80.
- [10] 李楠, 杨春杰. 基于近红外光谱技术的小米产地溯源研究[J]. *食品与机械*, 2020, 36(9): 97-101.
- LI N, YANG C J. Geographic origin determination of millet based on near infrared spectroscopy technique[J]. *Food & Machinery*, 2020, 36(9): 97-101.
- [11] NING J M, SHENG M G, YI X Y, et al. Rapid evaluation of soil fertility in tea plantation based on near-infrared spectroscopy[J]. *Spectroscopy Letters*, 2018, 51(9): 463-471.
- [12] 姚力, 李宗军, 朱门君, 等. 近红外漫反射光谱法快速检测高纤维素、木质素物料水分含量[J]. *食品与机械*, 2024, 40(2): 69-73.
- YAO L, LI Z J, ZHU M J, et al. Rapid determination of moisture content of high cellulose and lignin materials by near-infrared diffuse reflectance spectroscopy[J]. *Food & Machinery*, 2024, 40(2): 69-73.
- [13] 赵玉霞, 张明锦, 王茹, 等. 反向传播神经网络结合紫外-近红外融合光谱对“互助”青稞酒的判别研究[J]. *光谱学与光谱分析*, 2025, 45(5): 1 290-1 299.
- ZHAO Y X, ZHANG M J, WANG R, et al. Discriminative study on Huzhu Qingke liquor by back propagation neural network combined with ultraviolet-near infrared fusion spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2025, 45(5): 1 290-1 299.
- [14] BROWN T, JACOBS B. An introduction to computer science with emphasis on statistics and the social sciences[J]. *Computers & Education*, 1978, 2(4): 267-270.
- [15] 侯江霞, 姜金辉, 王琛鑫, 等. 基于机器学习的抑菌活性物质筛选研究进展[J]. *食品科学*, 2025, 46(14): 366-375.
- HOU J X, JIANG J H, WANG C X, et al. Research progress in the screening of antimicrobial substances based on machine learning[J]. *Food Science*, 2025, 46(14): 366-375.
- [16] 黄林峰, 蒋雪松, 贾志成, 等. 基于深度学习的梨树养分含量高光谱监测[J]. *光谱学与光谱分析*, 2024, 44(12): 3 543-3 552.
- HUANG L F, JIANG X F, JIA Z C, et al. Deep learning-based monitoring of nutrient content in pear trees[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(12): 3 543-3 552.
- [17] 宋少忠, 刘园园, 周紫阳, 等. 基于高光谱图像技术的高粱品种识别研究[J]. *光谱学与光谱分析*, 2024, 44(5): 1 392-1 397.
- SONG S Z, LIU Y Y, ZHOU Z Z, et al. Identification of sorghum breed by hyperspectral image technology[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(5): 1 392-1 397.
- [18] FEMENIAS A, BAINOTTI M B, GATIUS F, et al. Standardization of near infrared hyperspectral imaging for wheat single kernel sorting according to deoxynivalenol level [J]. *Food Research International*, 2021, 139: 109925.
- [19] ZHANG Y L, DAI A J, WU H Y, et al. Development of a benzothianone-based NIR fluorescent probe and its application of Pd0 detection in bioimaging and smartphone-assisted water samples[J]. *Journal of Photochemistry and Photobiology A: Chemistry*, 2025, 469: 116576.
- [20] LEE Y, CAPPELLATO M, DI CAMILLO B. Machine learning-based feature selection to search stable microbial biomarkers: application to inflammatory bowel disease[J]. *GigaScience*, 2022, 12: giad083.
- [21] 陈闻鹤, 程龙生, 常志朋, 等. 基于 Lasso 稳健马田系统的相对贫困识别方法[J]. *系统工程理论与实践*, 2022, 42(2): 527-544.
- CHEN W H, CHENG L S, CHANG Z P, et al. Identification of relative poverty based on Lasso-based robust Mahalanobis-Taguchi system[J]. *Systems Engineering-Theory & Practice*, 2022, 42(2): 527-544.
- [22] XIONG Z M, ZHU D F, LIU D F, et al. Anomaly detection of metallurgical energy data based on iForest-AE[J]. *Applied Sciences*, 2022, 12(19): 9 977.