

基于近红外光谱的便携式油菜籽品质检测装置研制

彭望^{1,2}, 王前¹, 廖庆喜^{1,2*}, 汪昊¹, 徐义³

(1. 华中农业大学工学院, 武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室, 武汉 430070;
3. 武汉联农种业科技有限责任公司, 武汉 430070)

摘要:为解决油菜籽收获、运输、存储、销售等过程中对含油量、蛋白质、硫苷等内部品质的检测需求, 实现油菜籽多指标、便携式、快速无损的检测目标, 该研究设计了一种基于近红外光谱的便携式油菜籽内部品质检测装置, 集成微型光谱仪、LCD触摸屏、树莓派处理器、箱体、样品杯和电源, 装置尺寸为246 mm×128 mm×127 mm, 可在户外环境稳定工作6 h以上。以不同产区的65个油菜籽品种为研究对象, 采集900~1 633 nm的近红外漫反射光谱, 使用竞争性自适应重加权、最小角回归、无信息变量消除等降维算法与偏最小二乘、极限学习机、支持向量等回归算法, 建立了含油量、蛋白质、硫苷、含水率和芥酸高精度预测模型, 决定系数 R^2 分别为0.949、0.861、0.730、0.976、0.862, 均方根误差RMSE分别为1.39%、1.46%、20.7 $\mu\text{mol/g}$ 、0.36%、3.63%。使用QT Creator作为集成开发环境, PyTorch作为框架, 实现了模型的嵌入式部署与应用, 实现了油菜籽的多品质参数一键式无损检测。使用Flask、MQTT等技术, 开发了APP端、网页端和小程序端软件, 实现了预测数据的多端同步和实时监控。经检验测试, 含油量、蛋白质、硫苷、含水率和芥酸的预测决定系数 R^2 分别为0.932、0.855、0.734、0.968、0.761, 均方根误差RMSE分别为1.35%、1.67%、19.6 $\mu\text{mol/g}$ 、0.34%、2.96%, 检测过程为13 s, 在相对湿度40%~70%、温度15~35 $^{\circ}\text{C}$ 的环境下, 装置对含杂量不大于2%、质量3~7 g的成熟期油菜籽具有数据采集稳定性; 该仪器可用于对油菜籽的快速无损检测。

关键词:无损检测; 近红外; 便携式; 多指标; 油菜籽

doi: 10.11975/j.issn.1002-6819.202402106

中图分类号: S565.4

文献标志码: A

文章编号: 1002-6819(2024)-18-0013-10

彭望, 王前, 廖庆喜, 等. 基于近红外光谱的便携式油菜籽品质检测装置研制[J]. 农业工程学报, 2024, 40(18): 13-22.
doi: 10.11975/j.issn.1002-6819.202402106 <http://www.tcsae.org>
PENG Wang, WANG Qian, LIAO Qingxi, et al. Development of the portable device for rapeseed quality detection based on near-infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 13-22. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202402106 <http://www.tcsae.org>

0 引言

据国家统计局信息显示, 中国2022年油料作物产量3 654.2万t, 其中油菜籽产量1 553.1万t, 占国产油料作物总产量的42.5%^[1], 油菜籽作为国内主要的油料作物, 其品质检测在整个生产过程有重要作用。油菜籽内部品质评价指标主要包括含油量(oil content, OC)、蛋白质含量(protein content, PC)、硫苷(glucosinolates, GL)、芥酸(erucic acid, EA)和含水率(moisture content, MC)等^[2], 含油量与蛋白质含量是油菜籽重要的理化特性, 决定了油菜籽的营养价值^[3], 硫苷与芥酸是普通油菜籽与双低油菜籽的评价指标^[2], 含水率对贮存、运输、干燥等环节有重要影响。

目前, 油菜籽内部品质评价指标的常规检测方法是使用化学分析法进行破坏性抽样检测, 该方法精准、可

靠, 但成本高、人员技术要求高、检测时间长^[4], 且油菜籽多种理化值的化学测定方法各不相同, 如蛋白质测定需要使用凯氏定氮仪, 硫苷与芥酸使用气相色谱法检测^[5]。因此, 需要一种可进行多指标同步检测、高效、无损、便携式的油菜籽品质检测方法及设备。基于近红外的便携式检测装置已应用于籽粒作物的内部品质检测, LIN等^[6]使用便携式稻粒水分仪检测了水分含量为13%~30%的米粒样品, 决定系数为0.936; LIU等^[7]使用4组特征波长收集小麦漫反射敏感光谱, 测定小麦蛋白质和水分含量预测值的均方根误差分别为0.486 6和0.216 1%; CHENG等^[8]使用紫外-可见-近红外光谱系统进行单粒玉米籽黄曲霉毒素检测, 准确率为95%, 灵敏度为86%; AYKAS等^[9]使用手持式大豆品质筛选仪实现了决定系数为0.92、均方根误差为3.07%的含油量检测。现有研究证实了可见/近红外光谱技术可用于籽粒作物的品质检测^[10-11], 但针对油菜籽多参数同时检测的研究尚未见报道, 且缺少针对油菜籽的便携式内部品质检测仪器研究。

本文基于近红外光谱分析技术, 设计油菜籽快速无损便携式检测装置; 利用数据分类、预处理、特征提取、模型筛选等多种分析方法, 探究各理化值最优模型; 依托校正模型开发控制与分析软件, 拟实现嵌入式设备一键操作, 实时显示含油量、蛋白质含量、硫苷、芥酸和

收稿日期: 2024-02-29 修订日期: 2024-05-17

基金项目: 湖北省自然科学基金资助项目(2023AFB860)

作者简介: 彭望, 博士, 副教授, 研究方向为便携式检测装置、微纳米感知器件。Email: pengwang@mail.hzau.edu.cn

*通信作者: 廖庆喜, 博士, 教授, 博士生导师, 研究方向为油菜全程机械化生产技术、现代农业装备与计算机测控。

Email: liaoxq@mail.hzau.edu.cn

含水率的预测值；并开发小程序、APP 与网页，将预测结果显示在手机和用户网站上；通过试验检测装置性能，以实现油菜籽的实时无损检测。

1 检测装置集成设计

1.1 检测原理

该装置能够采集油菜籽的近红外漫反射光谱信息，用于检测收获后达到安全储存期的油菜籽含油量、蛋白质含量、硫苷、芥酸和含水率等理化指标。为确保检测距离固定，设计了底部接触式光谱测量法，使用蓝宝石玻璃底的样品杯采集光谱，检测原理见图 1。将 3~7 g 油菜籽放置到样品杯中，进行预试验达到采样条件，采集参考白板形成的暗室与待测样品的光强，根据嵌入式处理器内部吸光度公式，得到样品光谱数据，由内部模型分析光谱数据得到预测值，最终在嵌入式设备、手机、网页上显示预测结果。

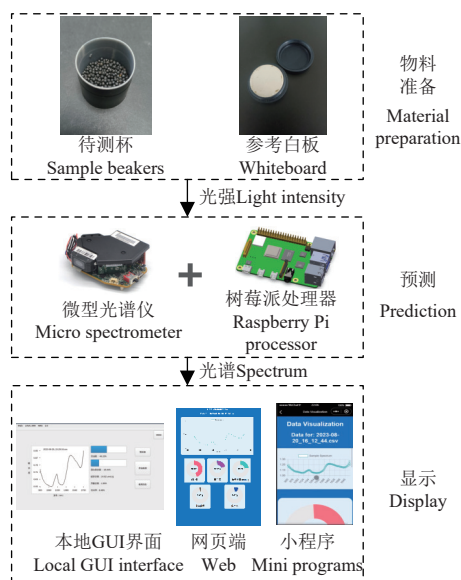


图 1 检测装置工作原理图

Fig.1 Working principle diagram of detection instrument

1.2 硬件设计

油菜籽受到近红外光 (400~2 500 nm) 光照射后，C-H、O-H、N-H 等含氢基团分子被激发，形成反映特征的近红外光谱，其中 1 210 nm 附近是含油量与含糖量相关的亚甲基 C-H 键二级倍频吸收峰，1 500~2 000 nm 附近存在蛋白质 N-H 的一级倍频与组合频率的吸收峰，1 034 nm 附近存在 N-H 键二级倍频吸收峰，1 440 nm 存在 O-H 键一级倍频吸收峰^[12-13]，可基于近红外光谱预测油菜籽相关内部品质（测量区间需涵盖以上主要吸收峰）。依据检测需求与检测原理，对关键部件进行优选，常见单片机为 STM32、FPGA、树莓派等，本装置涉及 GUI 设计、模型写入与预测、云端数据传输等操作，需较大的 SRAM 与丰富的扩展协议，因此，选择树莓派 4B 作为嵌入式处理器；常见可见/近红外微型光谱仪有日本滨松的 C10082CA，美国海洋光学的 NIR-M-R2，联合光科的 80311 等，本装置为便携式检测，对集成度有

较高要求，需使用漫反射或透射原理的光谱仪、波长区间需涵盖 1 210 与 1 440 nm，符合需求的为 TIDA-00554 与 NIR-M-R2，综合经济性，选择 NIR-M-R2 作为微型光谱仪。

油菜籽内部品质检测装置核心组件如图 2 所示，由装置箱体、锂离子电池、微型光谱仪、样品杯、7 英寸 LCD 触摸屏、箱体上板和树莓派处理器组成，其外形尺寸为 246 mm×129 mm×130 mm，装置总功耗约为 12 W（LCD 屏幕+光谱仪+树莓派），综合使用时间为 6 h，锂离子电池容量取 20 000 mA；样本杯直径需覆盖光源，取 2 cm，高度取 4 cm；参考白板使用温湿度不敏感的 PTFE 材料，直径较样品孔略小，取 3.6 cm，可覆盖光源孔形成暗室，用于环境校正。

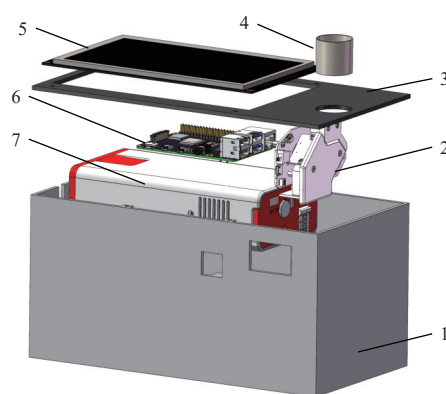


图 2 检测装置核心组件示意图
Fig.2 Schematic diagram of the core components of the detector

装置主要参数如表 1 所示。该装置采用接触反射方式采集 900~1 700 nm 波段的近红外光谱信息，整机质量 1.9 kg，持续作业时间大于 6 h，可通过采集 3~7 g 的油菜籽预测含油量、蛋白质含量、硫苷含量、芥酸含量与含水率，一次采集流程为 13 s，采用 PTFE 做参考板，工作温度为 15~35 ℃，各理化值决定系数大于 0.73。

表 1 装置参数表

Table 1 Device parameters

项目 Item	参数 Parameter
尺寸 Size/mm	246×128×127
质量 Weight / kg	1.9
取样范围 Sampling range/g	3~7
决定系数 Coefficient of determination	$R^2 > 0.73$
稳定性 Stability	15~35 ℃、40%~70% 相对湿度
波段 Wavelength range/nm	900~1 700
预测时间 Prediction time/s	13
信噪比 Signal-to-noise ratio	5 000 : 1

2 油菜籽近红外数据采集及处理

2.1 试验材料及试验方法

试验材料为华中、华南、西南等油菜主产区共 41 个油菜种植品种与各地收获油菜籽 24 个品种（共 65 种样品），理化值参考 GB/T 14488.1-2008、GB/T 14489.1-2008，

由国家油菜工程技术研究中心（湖北 武汉）使用凯氏定氮仪与液相色谱法测定^[5]。经测定，硫苷（8.3~270 $\mu\text{mol/g}$ ）、芥酸（0.7%~40%）呈现低态分布（63%的样品硫苷含量低于 35 $\mu\text{mol/g}$ 、69% 的样品芥酸含量低于 3%），油菜籽双低品种样品普适性较高；含油量（32.7%~54.9%）、蛋白质（12.1%~28.1%）与含水率（4.66%~9.96%）大致呈正态分布，符合常态取样规律。

为确保试验数据的可重复性、模型稳定性，对仪器进行光程筛选，每 10 min 执行预试验后采集参考白板的吸光度，重复 20 次。结果如图 3 所示，在 900~1 633 nm 波段内，吸光度基本稳定，有较高的信号稳定性。但在灰色区域（波长大于 1 636 nm）中，仪器吸光度曲线波幅增加，考虑试验的稳定与准确性，舍弃 1 636 nm 以上的波段，选择 900~1 633 nm 的光谱区间。

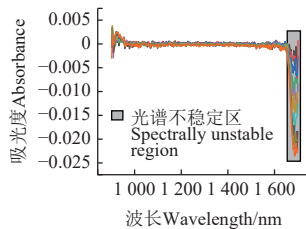


图 3 光程筛选

Fig.3 Pathlength screening

光谱采集的物料主要由参考白板、待测样品、样品杯组成，将样品放入杯中后，由微型光谱仪先采集参考白板的光谱，根据提示采集样品杯中的待测油菜籽光谱，根据式（1）得到样品吸光度，利用嵌入式处理器内校正模型与实际光谱数据对比，得到多种理化值的预测值。

$$A = -\lg \frac{I_s}{I_a} \quad (1)$$

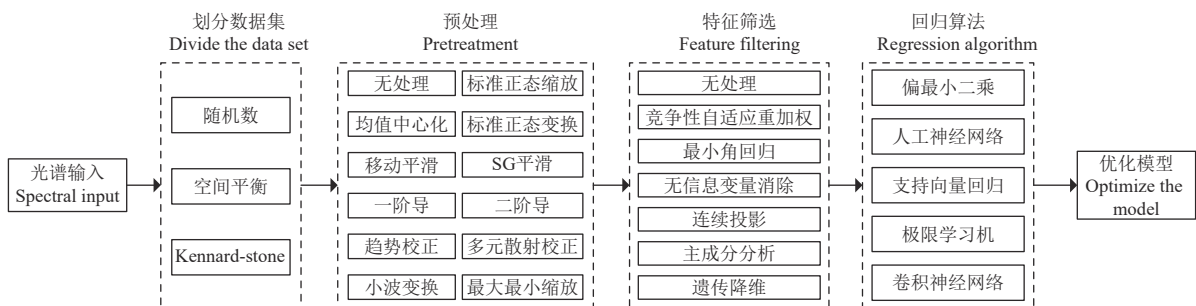


图 5 模型训练过程

Fig.5 Model training process

2.2.1 划分数据集

为支持机器学习模型的训练、验证和测试，以及在正式使用时能泛化到未知数据，本试验采用随机数（Random, None）、KS（Kennard-Stone, KS）与空间平衡随机（sample set partitioning based on joint X-Y distances, SPXY）^[16]3 种方法划分 65 种油菜籽的近红外光谱数据集，训练集与测试集比例为 8:2，使用训练集构建回归模型，使用测试集分析模型预测效果。

构建偏最小二乘（partial least squares, PLS）模型，

式中 A 为吸光度； I_a 为参考光谱； I_s 为样品光谱。

使用装置采集试验材料的光谱，每种油菜籽在室温（25 $^{\circ}\text{C}$ ）下采集 3 次（曝光时间为 0.635 ms，使用无平滑的 6 次平均得到初始光谱，增益 32 倍后得到实际光谱）。共获得 65 组样品的 195 条光谱，组内光谱平均后如图 4 所示，图中一定数量的光谱存在较大基线偏移是表面种衣与油菜品种差异双重因素导致的。例如，表面有暗淡色泽种衣的陕油 1309 平均吸光度为 1.413，表面有明亮色泽种衣的庆油 3 号平均吸光度为 0.637，两者吸光度相差 0.793；不含种衣的福油 518 与阳光 50 平均吸光度相差 0.364。

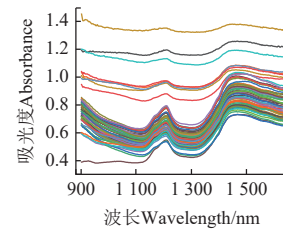


图 4 原始训练光谱

Fig.4 Raw training spectra

2.2 预测模型建立

为构建高精度、低误差的优化模型，以图 4 所示的光谱为原始数据，以测定的含油量、蛋白质、硫苷、芥酸、含水率为标签，将划分数据集（以均方根误差、决定系数为指标）、预处理（以光谱平滑度为指标）、数据降维（以降维与回归精度为指标）、回归算法（以均方根误差、决定系数为指标）模块化组合^[14-15]，如图 5 所示，分别确定各模块最优方案以实现油菜籽各理化值预测。

以测试集的均方根误差（root mean square error, RMSE）、决定系数（coefficient of determination, R^2 ）为主要评价指标，以均方误差（mean square error, MAE）为次要指标，3 种划分方法的 PLS 结果如表 2，含油量和蛋白质 KS 划分后的决定系数 R^2 为 0.901、0.770。因为 KS 方法通过考虑数据的特征值，寻找并保留主要的数据变化方向，对于常量元素具有更好的划分效果^[17]；针对含水率，随机数划分后的决定系数 R^2 为 0.890；硫苷与芥酸的样品含量范围跨度大且数据分布复杂，特定的划分方法捕

捉特性困难^[18]。基于任意数的随机划分具有最佳的预测值, 分别为 $R^2=0.635$ 和 $R^2=0.637$ 。

表 2 不同数据集划分方法下的偏最小二乘模型指标

Table 2 Partial least squares (PLS) model indicators for different dividing methods of dataset

划分方法 Dividing method	理化值 Physicochemical values	均方根误差 RMSE	决定系数 R^2	均方误差 MAE
随机数 Random	含油量/%	1.403	0.848	1.228
	蛋白质/%	2.078	0.593	1.493
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	23.018	0.635	17.640
	含水率/%	0.459	0.890	0.343
	芥酸/%	7.017	0.637	5.260
KS(Kennard-Stone)	含油量/%	1.374	0.901	1.153
	蛋白质/%	1.713	0.770	1.443
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	29.603	-0.316	24.530
	含水率/%	0.633	0.656	0.472
	芥酸/%	7.693	-0.229	5.522
空间平衡随机 (Sample set partitioning based on joint X-Y distances)	含油量/%	1.706	0.812	1.559
	蛋白质/%	1.935	0.749	1.694
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	18.350	0.300	16.845
	含水率/%	0.399	0.794	0.315
	芥酸/%	5.057	-0.102	4.140

注: 表中决定系数 R^2 为负数代表严重的过拟合, 表示划分方法与对应理化值不匹配。

Note: A negative coefficient of determination R^2 in the table indicates severe overfitting, meaning the dividing method does not match the corresponding physical and chemical values.

2.2.2 预处理

为去除噪点、降低环境影响、提高模型的稳健性与预测能力, 采用 SG 平滑滤波 (savitzky-golay, SG)、移动平均平滑 (moving average, MA)、标准正态缩放 (standardization and scaling, SS) 等预处理方法^[16], 以光谱曲线平滑度为评价指标计算预处理结果。

光谱曲线平滑度计算如式 (2) 所示, 基于平方二阶导数积分计算每条曲线平滑度, 将 65 条曲线的平滑度累加得一次预处理的光谱平滑度, 将 12 种预处理的光谱平滑度均一化处理, 计算得 SG 数据平滑度最高, 为 1.0, SNV (均标准正态变化) 次之, 为 0.981, D1 (一阶导数) 与 D2 (二阶导数) 最小, 为 0.486、0.032。本研究选择 SG 平滑 (5 窗口, 3 阶平滑) 方式。

$$S_i = \int (f_i''(x))^2 dx \quad (2)$$

$$S_{\text{总}} = \sum_{i=1}^{65} S_i \quad (3)$$

式中 $f_i''(x)$ 为光谱的二阶导; $S_{\text{总}}$ 为预处理的平滑度。

2.2.3 特征筛选

为判断原始数据是否满足单波长预测, 使用皮尔逊相关系数衡量理化值与吸光度单波长相关性^[19], 皮尔逊系数如式 (3) 所示, 分子代表吸光度与理化值的协方差估计值, 分母代表吸光度与理化值的标准差的乘积。

$$r = \frac{\sum_{i=1}^{65} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{65} (X_i - \bar{X})^2 \sum_{i=1}^{65} (Y_i - \bar{Y})^2}} \quad (4)$$

式中 r 为皮尔逊系数; X_i 为吸光度; Y_i 为理化值。 $\bar{X}$ 、 $\bar{Y}$ 为吸光度与理化值的均值。

由图 6 可知, 芥酸与吸光度相关性差, 1519 nm 处为最大值 0.041; 硫苷、含水率分别与 1086、1091 nm 处的吸光度存在 0.556、0.513 的最高相关度; 含油量、蛋白质分别与 974、1447 nm 的吸光度存在 -0.434、-0.449 的最高相关度。各理化值单波长预测精度低, 需进一步分析。

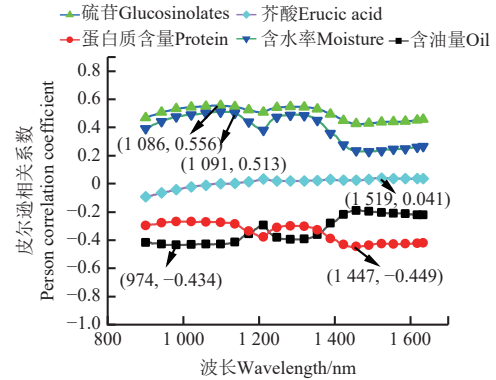


图 6 皮尔逊分析

Fig.6 Pearson analysis

为进一步从大量的光谱数据中筛选出具有特殊结构和功能的光谱峰、降低数据维度、提高运算速率, 采用主成分分析 (principal component analysis, Pca)、竞争性自适应重加权算法 (competitive adaptive reweighted sampling, Cars)、无信息变量消除 (uninformative variables elimination, Uve) 等算法执行特征筛选, Pca 算法采用累计贡献率大于 85% 的特征方向代替原始光谱, 实现降维; Cars 算法可消除无信息变量, 减少共线性数据的影响; Uve 算法可有效去除噪声数据, 实现数据降维^[20-21]。

1) Pca 分析

Pca 是一种非监督的降维算法, 通过正交变换将隐相关的变量转换为一组线性的主成分, 标准化赋权各特征、计算协方差矩阵线性关系并特征分解协方差矩阵, 筛选出新的特征空间 (主成分, PC)^[20], 实现数据降维。主成分分析结果如图 7 所示, 两组主成分解释了 99.6% 的方差差异性, 第一主成分代表了数据中最为显著的趋势, Pca 算法对数据压缩和简化有效。

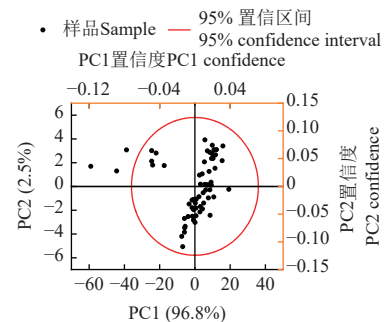


图 7 主成分分析

Fig.7 Principal component analysis (Pca)

2) Cars 分析

Cars 降维是结合蒙特卡洛采样与 PLS 模型回归系数的特征变量选择方法, 本研究设定 100 次蒙特卡洛采样,

五折交叉验证法^[22]。设定波长迭代方向为强制迭代, 固定剔除。波长迭代变化与含油量、蛋白质的均方根误差波动如图 8a 所示。在 Cars 算法的初期迭代 (0~30 轮) 中, 不重要的波长被移除, 初步优化了模型, RMSE 值波动中下降。后期迭代 (40~50 轮) 更多的波长被剔除, 导致模型丢失部分关键信息, RMSE 值波动中上升。硫苷、芥酸与含水率误差趋势如图 8b 所示, Cars 分析仅减少了波长, 对 RMSE 影响较小。含油量、蛋白质、硫苷、含水率和芥酸其理化值的 RMSE 分别为 1.58%、1.35%、17.913 $\mu\text{mol/g}$ 、0.525%、4.227%, 筛选后波长数量分别为 56、73、24、18、41。

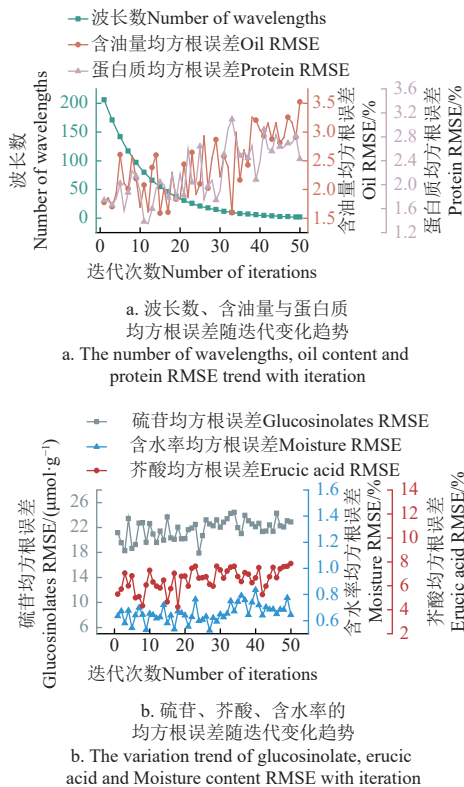


图 8 竞争性自适应重加权分析

Fig.8 Competitive adaptive reweighted sampling analysis (Cars)

3) Lars 分析

最小角回归 (Lars) 降维是一种用于高维数据的特征变量选择方法, 适合于变量个数大于观测数的情况^[23], 算法回归目标向量依次分解为若干组特征向量的线性组合, 最终使得与所有特征均线性无关的残差向量最小。Lars 分析如图 9 所示, 含油量与硫苷的光谱矩阵特征向量相关性高, 区分残差向量快; 蛋白质与芥酸为 Lars 不敏感矩阵, 在特征波长数超过 100 后才趋向稳定; 含水率模型最低误差为 1.4%, 远大于原始分析的 0.7% 与 Cars 分析的 0.525%, 因为含水率中 O-H 键 (氧-氢键) 的振动模式与偶极性对近红外 (NIR) 光谱区域的电磁波特别敏感^[24-25], Lars 分析对高敏感、多噪声变量易误判^[26], 因此含水率的特征降维误差较大。经过 Lars 降维后, 含油量、蛋白质、硫苷、含水率和芥酸其理化值的 RMSE 值为 4.173%、3.691%、22.765 $\mu\text{mol/g}$ 、1.428%、10.50%, 特征波长数量为 3、2、2、32、105。

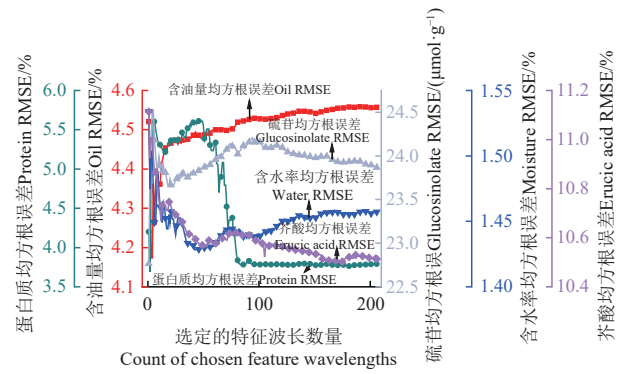


图 9 最小角回归分析

Fig.9 Least angle regression analysis (Lars)

4) Uve 分析:

Uve 是用噪声的无关变量信息统计去选择光谱自身的特征变量, 根据含噪自变量矩阵对目标矩阵的回归系数进行统计分布与变量判断的特征变量选择方法, 可有效减小过拟合^[27]。在添加噪声前, 需确定判定核的最优主成分, 以理化值的 RMSE 为指标, 执行 1~10 主成分数分析, 含油量、蛋白质、含水率与芥酸最佳主成分数为 7、7、7、8; 硫苷回归响应不积极, 在主成分为 2 时达到最佳。

人为产生一噪声矩阵 $R(n \times m)$, 将光谱矩阵 $X(n \times m)$ 与噪声矩阵 $R(n \times m)$ 组合形成矩阵 $XR(n \times 2m)$, 含噪组合矩阵如图 10a 所示, 深色区域为原始光谱标准差 15% 强度的高斯噪声区^[28], 噪声中心值为原始光谱标准差均值。图 10b 展示了含油量的特征矩阵及其筛选过程 (其余理化值类似), 图中水平线表示噪声矩阵的最大值 $C_{\max}$, 特征系数小于 $C_{\max}$ 的波长舍弃, 反之保留。经过无信息变量剔除后的含油量、蛋白质、硫苷、含水率和芥酸其理化值的 RMSE 值为 8.7%、13.4%、77.3 $\mu\text{mol/g}$ 、0.66%、6.5%, 特征波长数量为 69、32、23、87、9。

5) GA 分析:

遗传算法 (GA) 在降维中通过迭代过程中的选择、交叉和变异模拟自然选择过程, 保留对模型性能贡献最大的特征, 剔除冗余或不重要的特征^[29]。将光谱矩阵随机分层、采用五折交叉验证, 初始种群个数 50, 最大进化代数 200, 交叉组合率 0.7, 变异率 0.2。满足适应度误差小于固定值或进化代数时结束计算, 特征波长数与迭代精度于 160 轮最低。GA 分析后含油量、蛋白质、硫苷、含水率和芥酸其特征波长数为 81、75、80、80、78。

6) Spa 分析

连续投影算法 (Spa) 是一种有效消除变量共线性的降维算法^[30]。经过预处理后的原始矩阵使用 Spa 优选出波长序列, 分别建立含油量、蛋白质、硫苷、含水率、芥酸的 PLS 模型, 模型内波长序列对预测误差的影响如图 11 所示, 各理化值模型误差趋势均为先降低, 后稳定。以图中蛋白质模型为例, RMSE 随特征波长数先降低, 后稳定, 综合考虑波长序列与预测精度, 选择特征波长数为 12 个, 所选序列占原始信息的 5.8%

(含油量、硫苷、芥酸、含水率特征波长数为 18、16、12、19)。

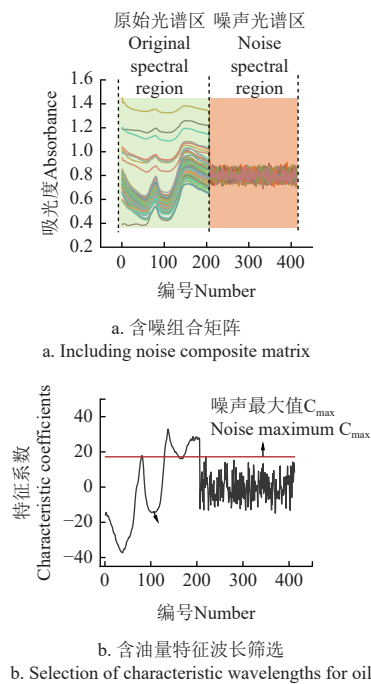


图 10 无信息变量消除分析

Fig.10 Uninformative variables elimination analysis (Uve)

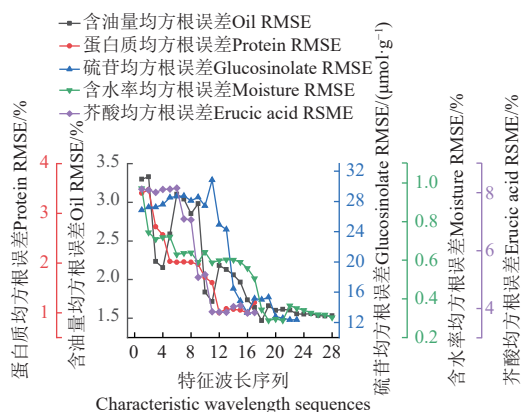


Fig.11 Successive projection algorithm analysis (Spa)

不同理化值的特征提取算法和相应的 PLS 模型参数如表 3 所示。特征降维前后, 含水率与硫苷模型的决定系数 R^2 分别损失 15.8%、9.35%, 对精度影响较大, 不选择特征降维; 含油量特征峰较为集中, 利用 Cars 分析得到关键组分后, 数据维度降低 72.82%, 预测的决定系数损失 0.9%; 蛋白质成分复杂 (菜籽蛋白 3 类、储藏蛋白 2 种、酶蛋白多种), 具有广泛而连续的吸收模式, 基于 Pca 的全局特征提取效果更好, 数据维度降低 99.03%, 决定系数提高 13.2%; 芥酸的 C=O 与 N-H 主吸收峰存在于 1 600~1 700 nm, 该波段仪器误差较大, 800~1 100 nm、1 150~1 300 nm 的弱吸收峰波段较长, 使用 Pca 可组合多组原始的微弱变量, 增强特征, 预测模型的决定系数提高 31.49%、均方根误差 RMSE 降低 40.78%。

2.2.4 光谱模型建立与评价

为探究不同回归算法对预测能力的影响, 获取最佳预测模型, 使用卷积神经网络 (convolutional neural network, CNN)、人工神经网络 (artificial neural network, ANN)、极限学习机 (extreme learning machine, ELM)、偏最小二乘 (partial least squares, PLS)、支持向量 (support vector regression, SVR)^[28, 31] 等回归算法对经过划分数据集、预处理、特征降维的油菜籽近红外光谱建立品质预测模型。

品质预测模型结果见表 4, 基于 PLS 模型整体误差与精度均较好, 因为 PLS 模型在 30 个样品以上即可有效划分反应型与形成型, 且结构方程模型是 Bootstrap 计算标准误差, 可有效用于非正态分布的特殊数据; 基于 CNN 与 ANN 的模型存在负数模型, 效果极差, 因为本研究将同种油菜籽取平均后视为一条光谱, 降低了数据丰度; 基于 SVR 的模型整体精度较低, 因为支持向量机对回归问题的学习策略为间隔最大化, 本质上是将分类问题向高维投影, 线性核、多项式核、高斯核等映射函数对数据仍具损失性; ELM 模型将单隐层前馈神经网络 (SLFN) 的后向传播改为随机输入权值, 使得光谱最敏感的水分 (正向或逆向) 较好的被区分出来, 决定系数 R^2 达 0.975。

表 3 不同特征筛选方法下测试集的预测精度

Table 3 Prediction accuracy under different feature selection methods

理化值 Physicochemical values	提取算法 Extraction algorithms	数据维度 Data dimensionality	RMSE	R^2	MAE
含油量 Oil/%	Pca	2	1.831	0.824	1.598
	Cars	56	1.579	0.892	1.071
	Lars	3	4.173	0.663	3.527
	Uve	69	8.732	0.563	1.690
	Spa	18	1.948	0.811	1.624
	GA	81	1.995	0.754	1.680
蛋白质 Protein/%	Pca	2	1.412	0.866	1.195
	Cars	73	1.351	0.787	1.617
	Lars	2	3.691	0.591	3.454
	Uve	32	13.414	0.488	1.725
	Spa	12	1.441	0.835	1.207
	GA	75	1.623	0.808	1.316
硫苷 Glucosinolates/ ($\mu\text{mol}\cdot\text{g}^{-1}$)	Pca	2	21.876	0.490	16.308
	Cars	24	17.913	0.444	15.484
	Lars	2	22.765	0.489	21.234
	Uve	23	77.327	0.258	16.824
	Spa	16	20.890	0.535	17.513
	GA	80	21.818	0.493	18.281
含水率 Moisture/%	Pca	2	0.504	0.671	0.414
	Cars	18	0.525	0.712	0.355
	Lars	32	1.428	0.513	1.422
	Uve	87	0.664	0.636	0.315
	Spa	19	0.512	0.694	0.433
	GA	80	0.399	0.807	0.314
芥酸 Erucic acid/%	Pca	2	4.155	0.838	3.626
	Cars	41	4.227	0.552	5.304
	Lars	105	10.50	0.280	10.679
	Uve	9	6.537	0.346	9.811
	Spa	12	7.040	0.534	5.190
	GA	78	6.885	0.554	5.351

综合各流程处理结果, 得各理化值最优模型参数, 含油量的校正模型采用 KS+SG+Cars+PLS, 蛋白质使用 Spxy+SG+Pca+PLS, 硫苷使用 Random+SG+PLS, 含水

率使用 Random+SG+ELM，芥酸使用 Random+SG+Pca+PLS。各模型指标如下：含油量的 RMSE, R^2 , MAE 分别为 1.40%、0.95、1.16%；蛋白质为 1.46%、0.86、1.24%；硫苷为 20.70 $\mu\text{mol/g}$ 、0.73、15.73 $\mu\text{mol/g}$ ；芥酸为 3.63%、0.86、3.28%；含水率为 0.36%、0.98、0.24%。

表 4 不同回归算法下的模型预测精度

Table 4 Model prediction accuracy under different regression algorithms

回归算法 Regression algorithm	理化值 Physicochemical values	RMSE	R^2	MAE
偏最小二乘 PLS	含油量/%	1.400	0.949	1.169
	蛋白质/%	1.456	0.862	1.240
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	20.708	0.730	15.732
	含水率/%	0.371	0.885	0.313
	芥酸/%	3.627	0.862	3.280
人工神经网络 ANN	含油量/%	8.056	0.591	4.979
	蛋白质/%	5.112	-0.465	4.531
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	26.066	0.276	24.460
	含水率/%	1.678	0.463	1.463
	芥酸/%	10.241	0.414	6.993
支持向量 SVR	含油量/%	3.170	0.226	2.516
	蛋白质/%	2.681	0.622	2.333
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	27.147	0.215	22.236
	含水率/%	0.800	0.668	0.584
	芥酸/%	11.591	0.564	5.449
极限学习机 ELM	含油量/%	2.358	0.572	1.871
	蛋白质/%	2.497	0.412	1.511
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	23.615	0.406	18.813
	含水率/%	0.363	0.976	0.236
	芥酸/%	9.221	0.200	5.728
卷积神经网络 CNN	含油量/%	2.611	0.475	2.135
	蛋白质/%	2.694	0.316	2.350
	硫苷/($\mu\text{mol}\cdot\text{g}^{-1}$)	30.986	-0.023	27.445
	含水率/%	0.739	0.716	0.572
	芥酸/%	10.288	0.460	6.801

3 检测仪 GUI 及云端开发

装置的本地端 GUI (graphical user interface) 如图 12 所示，由光谱可视化区、预测显示区、用户交互区组成。光谱可视化区利用 QGraphicsScene 与 QLabel 结合光谱采集标志位，将光谱数据可视化；预测显示区显示含油量、蛋白质、硫苷、芥酸、含水率的预测结果与油菜籽分类结果；用户交互区使用 QT 与自定义类，实现装置预热、光谱采集进程控制与中断、检测历史调用等常用功能。

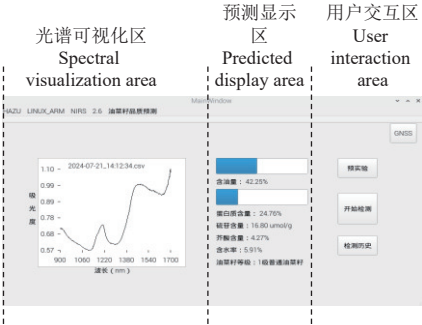


图 12 本地端展示界面
Fig.12 Local display interface

小程序与 APP 界面为基于 Flask 的数据展示界面，集成用户登录管理、MQTT 消息处理、以及数据可视化

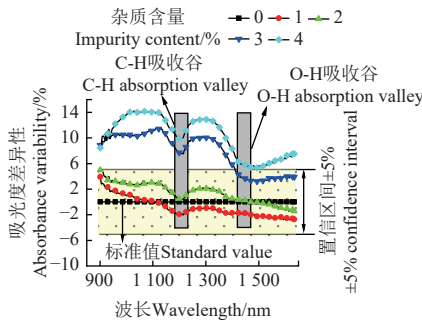
功能。后端使用 Flask-WTF 实现用户认证流程，利用 MQTT 协议质询服务器，实时接收和解析数据，使云端与本地端数据保持一致；前端定义组织用户登录、认证、注销过程，使用 Chart 以折线图和环形图展示后端数据，通过 Ajax 轮询实现数据的实时更新。

4 装置验证

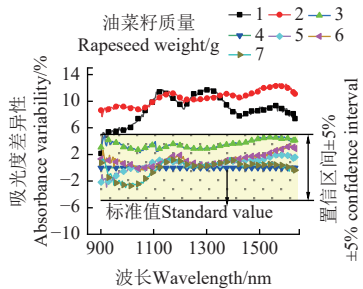
4.1 装置稳定性验证

4.1.1 含杂量与样品质量稳定性试验

依据小样杂质的掺混原则，混合质量分数 1%、2%、3%、4% 的杂质到未测试样品（5 g）中，吸光度差异性结果如图 13a 所示，吸光度差异性曲线左高右低，含杂量越高偏移越大（存在 1 210、1 440 nm 两处明显波谷）的趋势。因杂质通常为破碎的茎秆、叶片，土壤颗粒等，含油量与含糖量等成分与油菜籽重合度较低，在 1 210 nm 的与含油量与含糖量相关的亚甲基 C-H 键二级倍频吸收峰处出现吸光度差异的局部波谷。又因杂质中多为有机物，1 440 nm 的 O-H 键一级倍频吸收峰与油菜籽接近，因此在 1 440 nm 处出现差异性曲线的降低。以 $\pm 5\%$ 为置信区间，1%、2% 含杂量的全波段差异均在置信区间内，装置对含杂量不大于 2% 的样品具有良好预测能力。



a. 不同杂质百分比下的吸光度差异性
a. Difference in absorbance at different impurity percentages



b. 不同样品质量下的吸光度差异性
b. Difference in absorbance at different sample weight

图 13 装置含杂量与样品质量稳定性验证

Fig.13 Verification of device impurity content and sample weight stability

使用 1~7 g 的同一品种油菜籽进行试验，吸光度差异如图 13b 所示，取 1 g 油菜籽时，样品杯周向（圆周方向，即杯底平面）未被完全覆盖，有大范围光源透过，光谱结果整体吸光度误差约为 10%；取 3~7 g 油菜籽时，

样品杯周向与轴向均填充油菜籽, 无光斑出现, 吸光度误差约为 1.5% (系统误差)。综上, 装置对 3~7 g 的油菜籽具有良好的泛用性。

4.1.2 温湿度稳定性试验

为进一步测试装置稳定性与泛用性, 执行装置预试验后, 开展温湿度稳定性试验。以 15~35 °C 为区间开展装置温度稳定性试验。结果如图 14a 所示, 以 $\pm 5\%$ 的吸光度差异为置信区间, 温度稳定区间为 15~35 °C; 基于常见相对湿度 (40%~80%) 开展装置环境湿度稳定性试验, 由图 14b 可知, 当相对湿度为 70%、80% 时, 油菜籽样品的光强与平均湿度下差异较大, 这是因为装置光源孔温度较高 (30~35 °C)、样品杯为常温 (20~25 °C), 接触后发生热传递, 样本杯底 (蓝宝石结构) 发生过饱和, 水分凝结为小水滴, 影响了油菜籽样品的漫反射检测结果, 手动擦拭样品杯底的水雾或将样品杯调整至与装置同温 (将样本与样本杯放置于装置内部一段时间) 后, 装置的相对湿度稳定区间为 40%~70% (不擦拭或调整, 湿度上限为 60%)。

4.2 装置检测精度验证

检测精度验证选择为不在训练集内的收获与种植油菜籽, 理化值测定同试验材料。使用所设计的便携式油菜籽内部品质检测装置进行试验, 每个样品采集三次光谱取平均, 将预测值与真实值结果做线性分析, 试验结果如图 15 所示。由图 15a、图 15b、图 15d、图 15e 可知, 含油量为 52%、蛋白质含量为 30% 等特殊油菜品种的理化值预测仍具有较高的预测精度。因为试验材料综合了特育品种与种植品种, 提高了模型对超范围样品的拓展预测能力, 降低了模型对非常规理化值的预测偏差。图 15c、图 15e 中硫苷、芥酸预测误差较大, 因硫苷含量小且极值差异大, 线性模型的预测精度较低。综上,

便携式油菜籽内部品质检测装置含油量、蛋白质含量、硫苷、含水率、芥酸预测集与真实集的决定系数 R^2 分别为 0.932、0.855、0.734、0.968、0.761。RMSE 分别为 1.35%、1.67%、19.6 $\mu\text{mol/g}$ 、0.34%、2.96%。仪器有良好的预测能力, 外部泛用性较好, 可适用于未知样品的预测, 预测精度达到设计需求, 可用于油菜籽内部品质检测。

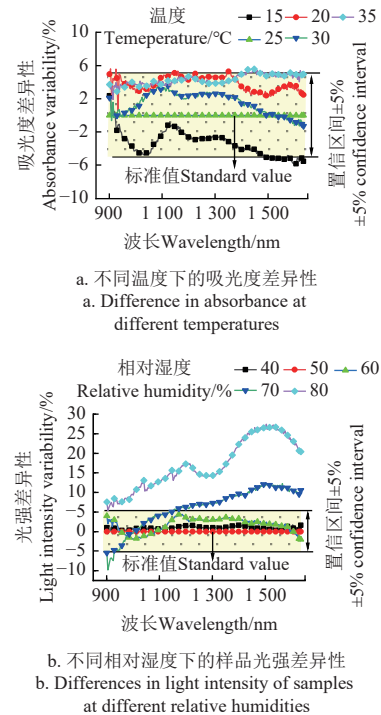


图 14 装置温湿度稳定性验证

Fig.14 Verification of temperature and humidity stability of the device

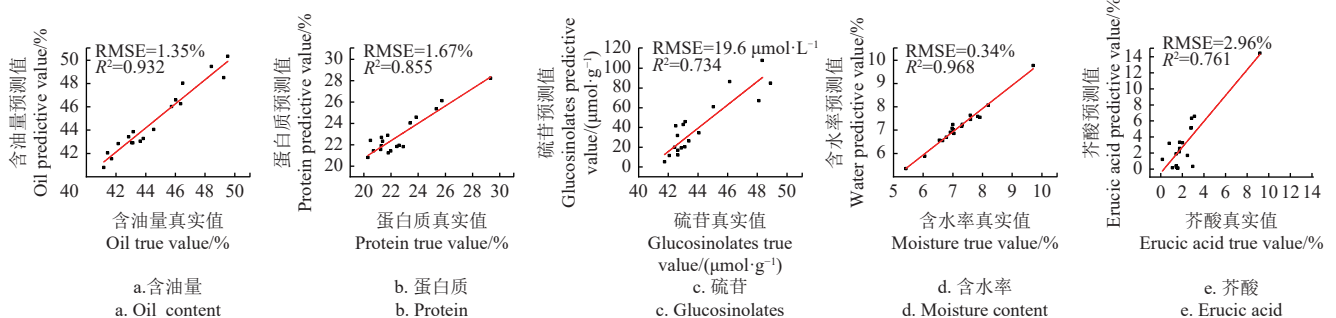


图 15 便携式油菜籽品质检测装置预测精度验证

Fig.15 Portable rapeseed quality testing device prediction accuracy verification

5 结论

1) 根据成熟期油菜籽的光学、物理特性, 设计基于近红外光谱的便携式油菜籽内部品质检测装置, 该仪器主要由装置箱体、锂离子电池、微型光谱仪、7 英寸 LCD 触摸屏、箱体上板和树莓派处理器组成, 其外形尺寸为 246 mm×129 mm×130 mm, 该装置充电后可不间断连续稳定测试 6 h, 不受环境光影响, 符合户外作业条件。

2) 采集油菜籽的 900~1 633 nm 近红外接触反射光谱信息, 对原始光谱进行划分数据集、预处理、特征筛选、回归模型等分析, 建立了油菜籽含油量、蛋白质、硫苷、含水率、芥酸的最优预测模型, 决定系数 R^2 分别为 0.949、0.861、0.730、0.976、0.862。RMSE 分别为 1.39%、1.46%、20.7 $\mu\text{mol/g}$ 、0.36%、3.63%。

3) 使用 Python 部署预测模型, 使用 QT creator 编写嵌入式油菜籽品质预测软件, 实现一键无损检测油菜

籽 5 种理化值。

4) 验证了便携式油菜籽内部品质检测装置的检测精度与稳定性。在相对湿度 40%~70%、温度 15~35 ℃ 的环境下, 装置对含杂量不大于 2%、质量 3~7 g 的油菜籽具有良好的数据采集稳定性, 油菜籽含油量、蛋白质、硫苷、含水率、芥酸的预测值与真实值的决定系数 R^2 分别为 0.932、0.855、0.734、0.968、0.761。均方根误差 RMSE 分别为 1.35%、1.67%、19.6 $\mu\text{mol/g}$ 、0.34%、2.96%。结果表明, 该装置可完成油菜籽达到安全储存期后的含油量、蛋白质、硫苷、含水率、芥酸的实时无损检测。

【参 考 文 献】

- [1] 国家统计局. 2023 年中国统计年鉴 [M]. 2023 年中国统计年鉴. 北京: 中国统计出版社, 2023.
- [2] 湖北省粮油食品质量监测站: GB/T 11762-2006 油菜籽[S]. 北京: 中国标准出版社, 2006.
- [3] HU Q, HUA W, YIN Y, et al. Rapeseed research and production in China[J]. *The Crop Journal*, 2017, 5(2): 127-135.
- [4] BARTHET V J, PETRYK M W, SIEMENS B. Rapid nondestructive analysis of intact canola seeds using a handheld near - infrared spectrometer[J]. *Journal of the American Oil Chemists' Society*, 2020, 97(6): 577-589.
- [5] CHMIELEWSKA A, KOZŁOWSKA M, RACHWAŁ D, et al. Canola/rapeseed protein—nutritional value, functionality and food application: a review[J]. *Critical Reviews in Food Science and Nutrition*, 2021, 61(22): 3836-3856.
- [6] LIN L, HE Y, XIAO Z, et al. Rapid-detection sensor for rice grain moisture based on NIR spectroscopy[J]. *Applied Sciences*, 2019, 9(8): 1654.
- [7] LIU Y, LI D, LI H, et al. Design of a phenotypic sensor about protein and moisture in wheat grain[J]. *Frontiers in Plant Science*, 2022, 13: 881560.
- [8] CHENG X, VELLA A, STASIEWICZ M J. Classification of aflatoxin contaminated single corn kernels by ultraviolet to near infrared spectroscopy[J]. *Food Control*, 2019, 98: 253-261.
- [9] AYKAS D P, BALL C, SIA A, et al. In-situ screening of soybean quality with a novel handheld near-infrared sensor[J]. *Sensors*, 2020, 20(21): 6283.
- [10] 金文玲, 曹乃亮, 朱明东, 等. 基于近红外超连续激光光谱的水稻种子活力无损分级检测研究[J]. *中国光学*, 2020, 13(5): 1032-1043.
JIN Wenling, CAO Nailiang, ZHU Mingdong, et al. Nondestructive grading test of rice seed activity using near infrared super-continuum laser spectrum[J]. *Chinese Optics*, 2020, 13(5): 1032-1043. (in Chinese with English abstract)
- [11] BEĆ K B, GRABSKA J, HUCK C W. Near-infrared spectroscopy in bio-applications[J]. *Molecules*, 2020, 25(12): 2948.
- [12] BEC K B, HUCK C W. Breakthrough potential in near-infrared spectroscopy: Spectra simulation. A review of recent developments[J]. *Frontiers in Chemistry*, 2019, 7: 48.
- [13] OZAKI Y. Infrared spectroscopy—Mid-infrared, near-infrared, and far-infrared/terahertz spectroscopy[J]. *Analytical Sciences*, 2021, 37(9): 1193-1212.
- [14] 李永玉, 吴继峰, 王威, 等. 便携式双档位柑橘多品质无损检测装置设计与试验[J]. *农业工程学报*, 2023, 39(23): 259-267.
- LI Yongyu, WU Jifeng, WANG Wei, et al. A portable dual-gear device for non-destructive testing on multi quality of citrus[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(23): 259-267. (in Chinese with English abstract)
- [15] 郭俊先, 张振振, 韩景, 等. 基于傅里叶变换近红外光谱的籽棉含水率无损检测[J]. *农业工程学报*, 2023, 39(21): 152-160.
GUO Junxian, ZHANG Zhenzhen, HAN Jing, et al. Non-destructive detection of seed cotton moisture content based on Fourier transform near-infrared spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(21): 152-160. (in Chinese with English abstract)
- [16] JIAO Y, LI Z, CHEN X, et al. Preprocessing methods for near-infrared spectrum calibration[J]. *Journal of Chemometrics*, 2020, 34(11): e3306.
- [17] FERREIRA R D A, TEIXEIRA G, PETERNELLI L A. Kennard-Stone method outperforms the Random Sampling in the selection of calibration samples in SNPs and NIR data[J]. *Ciência Rural*, 2021, 52(5): e20201072.
- [18] CHEN W, CHEN H, FENG Q, et al. A hybrid optimization method for sample partitioning in near-infrared analysis[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021, 248: 119182.
- [19] 张柏, 凌彩金, 李露青, 等. 基于数据融合策略的红茶发酵程度判别[J]. *农业工程学报*, 2022, 38(15): 339-347.
ZHANG Bai, LING Caijin, LI Luqing, et al. Discrimination of black tea fermentation degrees based on data fusion strategy[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(15): 339-347. (in Chinese with English abstract)
- [20] ANOWAR F, SADAUI S, SELIM B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)[J]. *Computer Science Review*, 2021, 40: 100378.
- [21] 刘莉, 陶红燕, 方静, 等. 基于近红外高光谱的梨叶片炭疽病与黑斑病识别[J]. *农业机械学报*, 2022, 53(2): 221-230.
LIU Li, TAO Hongyan, FANG Jing, et al. Identifying anthracnose and black spot of pear leaves on near-infrared hyperspectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(2): 221-230. (in Chinese with English abstract)
- [22] 彭彦昆, 孙晨, 赵苗. 苹果品质动态无损感知及分级机器人系统[J]. *农业工程学报*, 2022, 38(16): 293-303.
PENG Yankun, SUN Chen, ZHAO Miao. Dynamic nondestructive sensing and grading manipulator system for apple quality[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(16): 293-303. (in Chinese with English abstract)
- [23] LIU L, LI Y, HAN J, et al. A least angle regression assessment algorithm based on joint dictionary for visible and near-infrared spectrum denoising[J]. *Optik*, 2021, 242: 167093.
- [24] CHEN R C, DEWI C, HUANG S W, et al. Selecting critical features for data classification based on machine learning methods[J]. *Journal of Big Data*, 2020, 7(1): 52.
- [25] KNADEL M, CASTALDI F, BARBETTI R, et al. Mathematical techniques to remove moisture effects from visible–near-infrared–shortwave-infrared soil spectra[J]. *Applied Spectroscopy Reviews*, 2023, 58(9): 629-662.

- [26] CAO Y, BAO N, LIU S, et al. Reducing moisture effects on soil organic carbon content prediction in visible and near-infrared spectra with an external parameter orthogonalization algorithm[J]. *Canadian Journal of Soil Science*, 2020, 100(3): 253-262.
- [27] FILZMOSER P, NORDHAUSEN K. Robust linear regression for high - dimensional data: An overview[J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2021, 13(4): e1524.
- [28] 郭志明, 王郡艺, 宋焯, 等. 手持式可见近红外苹果品质无损检测系统设计与试验[J]. *农业工程学报*, 2021, 37(22): 271-277.
- GUO Zhiming, WANG Junyi, SONG Ye, et al. Design and experiment of the handheld visible-near infrared nondestructive detecting system for apple quality[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(22): 271-277. (in Chinese with English abstract)
- [29] JIANG W, LU C, ZHANG Y, et al. Moving-window-improved monte carlo uninformative variable elimination combining successive projections algorithm for near-infrared spectroscopy (NIRS)[J]. *Journal of Spectroscopy*, 2020, 2020(1): 3590301.
- [30] 张楠楠, 张晓, 王城坤, 等. 基于高光谱和连续投影算法的棉花叶面积指数估测[J]. *农业机械学报*, 2022, 53(S1): 257-262.
- ZHANG Nannan, ZHANG Xiao, WANG Chengkun, et al. Cotton LAI estimation based on hyperspectral and successive projection algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(S1): 257-262. (in Chinese with English abstract)
- [31] TIAN X, LI J, YI S, et al. Nondestructive determining the soluble solids content of citrus using near infrared transmittance technology combined with the variable selection algorithm[J]. *Artificial Intelligence in Agriculture*, 2020, 4: 48-57.

Development of the portable device for rapeseed quality detection based on near-infrared spectroscopy

PENG Wang^{1,2}, WANG Qian¹, LIAO Qingxi^{1,2*}, WANG Hao¹, XU Yi³

(1. College of Engineering, Huazhong Agriculture University, Wuhan 430070, China; 2. Key Laboratory of Agricultural Equipment in Mid-lower Yangtze River, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China; 3. Wuhan Liannong Seed Science and Technology Co., Ltd., Wuhan 430070, China)

Abstract: Oil content, protein, glucosinolates, and internal qualities are required to be detected during harvesting, transportation, storage, and sale of rapeseed. In this study, a portable device was designed to detect the rapeseed internal quality using near-infrared spectroscopy, in order to realize the multi-index, portable, and rapid non-destructive testing. The hardware was integrated with the dimensions of 246 mm × 128 mm × 127 mm, such as a mini spectrometer, an LCD touchscreen, and a Raspberry Pi processor. The device was operated stably outdoors for 6 hours. The 65 varieties of rapeseed were sampled from different production areas. Diffuse reflectance spectra were then collected from 900 to 1700 nm. The wavelength stability tests were carried out to determine a stable spectral region from 900 to 1 633 nm. Data was divided using KS (Kennard-Stone), SPXY (sample set partitioning based on joint X-Y distances), and Random number. The dataset division was obtained with R^2 as the index, where the KS was used for the oil content and protein, while the random numbers were for the glucosinolates, erucic acid, and moisture content. Various methods of data smoothing were evaluated with data smoothness as the index, such as SNV, SG, MSC, D1, and D2. SG smoothing (5-window, and 3rd order) was determined as the best preprocessing. Cars, Pca, GA, Lars, Uve, and Spa were used as data reduction to explore the best dimension reduction for each physicochemical index. Cars reduced the dimensions of the oil content model by 73%. Pca reduced the dimensions of the protein and erucic acid models by 98%. Thus the predictive accuracy of erucic acid was improved by 31.49%, whereas, the predictive error was reduced by 40.78%. PLS, ANN, CNN, SVR, and ELM models were used with RMSE, R^2 , and MAE as the indices. The calibration model was determined for the oil content using KS+SG+Cars+PLS, for the protein using SPXY+SG+Pca+PLS, for the glucosinolates using Random+SG+PLS, for the moisture content using Random+SG+ELM, and the for erucic acid using Random+SG+Pca+PLS. Model indices were as follows: RMSE, R^2 , and MAE for the oil content were 1.40%, 0.95, 1.16%; for the protein 1.46%, 0.86, 1.24%; for the glucosinolates 20.70 $\mu\text{mol/g}$, 0.73, 15.73 $\mu\text{mol/g}$; for the erucic acid 3.63%, 0.86, 3.28%; for the moisture content 0.36%, 0.98, 0.24%, respectively. The device was used to collect the spectra from 1 to 7 g of rapeseed seven times for the light transmission. Results showed that the absorption error for 3 to 7 g was within $\pm 5\%$. Electronic scales were then removed for direct sample testing during field tests. Five stability tests were conducted at temperatures from 15 to 35 $^{\circ}\text{C}$, all of which were within reliable ranges; Stability tests were set as the relative humidity from 40% to 80% and found condensation at 70% relative humidity, which was corrected to maintain stability from 40% to 70% relative humidity. Accuracy tests on the device showed the correlation coefficients (R^2) for oil content, protein content, glucosinolates, moisture content, and erucic acid between predicted and real sets at 0.932, 0.855, 0.734, 0.968, and 0.761, respectively; RMSE values were 1.35, 1.67, 19.6, 0.34, and 2.96, respectively. Therefore, the device can be expected to perform real-time non-destructive testing of oil content, protein, glucosinolates, moisture content, and erucic acid in rapeseed after safe storage.

Keywords: non-destructive testing; near-infrared; portable; multi-index; rapeseed