

基于改进 ABR 模型的小麦籽粒类胡萝卜素含量 近红外检测方法

罗斌^{1,2} 刘洁琼^{1,2} 陈泉² 周亚男² 潘大宇²

(1. 新疆农业大学机电工程学院, 乌鲁木齐 830052; 2. 北京市农林科学院智能装备技术研究中心, 北京 100097)

摘要: 小麦籽粒类胡萝卜素含量是衡量小麦营养价值和育种品质的关键指标。为实现小麦籽粒类胡萝卜素含量快速无损检测,设计了一种小麦籽粒近红外光谱快速采集装置,可实现待测样品等量分样、快速多次采集近红外光谱数据,提高光谱采集效率。以180个小麦籽粒作为研究对象,获取900~1700 nm范围内的近红外光谱数据,采用 Savitzky-Golay 平滑(Savitzky-Golay, SG)、标准正态变量变换(Standard normal variate transformation, SNV)、多元散射校正(Multiplicative scatter correction, MSC)、趋势校正(Trend correction, TC)、Savitzky-Golay 平滑+趋势校正(SG+TC)、Savitzky-Golay 平滑+多元散射校正(SG+MSC)、Savitzky-Golay 平滑+标准正态变量变换(SG+SNV)、Savitzky-Golay 平滑+一阶导数(SG+1D)、Savitzky-Golay 平滑+二阶导数(SG+2D)、Savitzky-Golay 平滑+三阶导数(SG+3D)10种预处理方法,Relief 算法、遗传算法(Genetic algorithm, GA)、方差阈值算法(Variance threshold, VT)、连续投影算法(Successive projections algorithm, SPA)4种特征选择算法建立偏最小二乘回归(Partial least squares regression, PLSR)、支持向量机回归(Support vector regression, SVR)、自适应提升回归(Adaptive boosting regression, ABR)3种机器学习模型,对小麦籽粒类胡萝卜素含量进行预测。为进一步提高模型预测精度,引入纵横交叉(CSO)和自适应收敛因子(ACF)2种改进策略对 ABR 模型进行优化。结果表明,使用 SNV 预处理、Relief 特征选择算法和 CSO-ACF 改进策略建立 ABR 模型预测效果最佳,校正集决定系数 R_c^2 为 0.90,均方根误差(RMSEC)为 0.29 $\mu\text{g/g}$,预测集决定系数为 R_p^2 为 0.90,均方根误差(RMSEP)为 0.32 $\mu\text{g/g}$,RPD 为 3.16。因此,该装置与模型算法相结合可实现小麦类胡萝卜素含量快速、无损预测。

关键词: 小麦籽粒;类胡萝卜素;无损检测;近红外光谱;自适应提升回归算法

中图分类号: S512 文献标识码: A 文章编号: 1000-1298(2026)03-0324-08

OSID:



Near Infrared Detection of Carotenoids in Wheat Grain Based on Improved ABR Model

LUO Bin^{1,2} LIU Jieqiong^{1,2} CHEN Quan² ZHOU Ya'nan² PAN Dayu²

(1. College of Mechanical and Electrical Engineering, Xinjiang Agricultural University, Urumqi 830052, China

2. Research Center of Intelligent Equipment, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China)

Abstract: Carotenoid content in wheat grain is the key index to measure the nutritional value and breeding quality of wheat. In order to realize the rapid and nondestructive detection of carotenoid content in wheat grain, a near-infrared spectrum rapid acquisition device for wheat grain was designed, which can realize the equal sample size of the sample to be measured, and quickly collect near-infrared spectrum data for many times, so as to improve the spectral acquisition efficiency. Taking 180 wheat grains as the research object, the near-infrared spectral data in the range of 900~1700 nm were obtained. Savitzky-Golay (SG), standard normal variate transformation (SNV), multiplicative scatter correction (MSC), trend correction (TC), SG+TC, SG+MSC, SG+SNV, Savitzky-Golay+first derivative (SG+1D), Savitzky-Golay+second derivative (SG+2D), Savitzky-Golay+third derivative (SG+3D) ten pretreatment methods, four feature selection algorithms, relief algorithm (Relief), genetic algorithm (GA), variance threshold (VT) and successive projections algorithm (SPA), were used to establish three mathematical models of partial least squares regression (PLSR),

收稿日期: 2025-03-03 修回日期: 2025-08-20

基金项目: 国家重点研发计划项目(2023YFD2000405)

作者简介: 罗斌(1981—),男,副研究员,博士,主要从事作物表型信息检测和农用传感器技术装备研究, E-mail: luob@nrcita.org.cn

support vector regression (SVR) and adaptive boosting regression (ABR) to predict carotenoid content in wheat seeds. The results showed that the prediction effect of ABR model based on SNV preprocessing, relief feature selection algorithm and CSO - ACF improved strategy was the best. The determination coefficient R_c^2 of correction set was 0.90, the root mean square error (RMSEC) was 0.29 $\mu\text{g/g}$, the determination coefficient R_p^2 of prediction set was 0.90, the root mean square error (RMSEP) was 0.32 $\mu\text{g/g}$, and the RPD was 3.16. Therefore, the device combined with the model algorithm can achieve rapid and nondestructive prediction of carotenoid content in wheat.

Key words: wheat grain; carotenoids; non-destructive testing; near infrared spectroscopy; adaptive boosting regression

0 引言

小麦籽粒中含有淀粉、蛋白质、维生素、矿物质、类胡萝卜素等多种营养物质,可以为人体生长发育提供基础能量^[1]。其中,类胡萝卜素可以通过抑制脂肪氧化酶活性保障小麦胚乳营养品质,延长小麦储藏时间^[2]。类胡萝卜素含量对小麦食品色泽的呈现也十分重要,会直接影响小麦制品的颜色外观^[3]。此外,类胡萝卜素在维持人体健康方面也发挥着重要作用,可以调节人体免疫系统,降低癌症风险^[4]。因此,开展小麦籽粒类胡萝卜素含量的研究具有重要意义,可以为小麦营养价值和品质评估提供参考^[5]。目前,测定小麦籽粒类胡萝卜素含量的方法主要有紫外分光光度法、薄层色谱法和高效液相色谱法等化学方法,化学方法检测精度高,但具有一定的局限性,难以满足快速无损检测需求^[6]。因此,利用高效、快速、无损检测技术测定小麦籽粒类胡萝卜素含量对小麦品质提升和育种改良有重要意义。

近红外光谱技术在农作物品质含量预测方面具有快速、无损、成本较低等优势^[7]。文献[8]利用平均光谱数据建立了径向基函数网络预测模型,实现了面粉中偶氮二甲酰胺含量的检测。文献[9]使用近红外光谱数据建立预测模型对小麦中谷蛋白、醇溶蛋白和麦谷蛋白等物质含量进行预测。文献[10]通过多目标优化算法对特征波长进行筛选,实现小麦脂肪酸的高精度分析。文献[11]提出了一种改进粒子群优化-反向传播神经网络算法对小麦储藏品质进行预测。文献[12]基于近红外光谱技术对燕麦籽粒 β -葡聚糖含量进行预测研究。文献[13]利用近红外光谱技术对西葫芦果皮和果肉中总类胡萝卜素、叶黄素等物质含量进行检测。文献[14-15]分别建立了玉米籽粒和马铃薯中类胡萝卜素含量近红外预测模型。现有研究证明近红外光谱技术可用于农作物的品质检测,但有关小麦籽粒类胡萝卜素含量近红外检测的方法鲜有研究。利用近红外光谱数据建立模型预测农作物品质参数时,使用具有代表性的光谱数据建立模型可以有效

提升模型预测精度与稳定性。但是获取具有代表性的光谱数据,需多次扫描样品并计算平均光谱。目前近红外光谱采集装置大多仅具备单次采集光谱数据的功能,在获取平均光谱时,需要人为进行多次重复操作,耗时费力。

针对小麦近红外光谱采集装置和小麦籽粒类胡萝卜素含量测定方法的不足,本文设计小麦籽粒近红外光谱快速采集装置,并采用 SG、SNV、MSC、TC、SG + TC、SG + MSC、SG + SNV、SG + 1D、SG + 2D、SG + 3D 共 10 种预处理方法,Relief、GA、VT、SPA 共 4 种特征选择算法建立 PLSR、SVR、ABR 共 3 种机器学习模型对小麦籽粒类胡萝卜素含量进行预测。为提高模型预测精度,使用纵横交叉(CSO)和自适应收敛因子(ACF)两种改进策略对 ABR 模型进行优化。设计近红外光谱快速采集装置和改进模型算法,以期对小麦籽粒类胡萝卜素含量快速无损预测提供方法。

1 检测装置设计

该装置可快速多次采集小麦籽粒近红外光谱信息,提供更具代表性的平均光谱,用于建立小麦籽粒类胡萝卜素含量预测模型。为确保近红外光谱数据质量和采集效率,设计了一种侧面接触式光谱采集装置,检测装置原理图如图 1 所示。首先将待测样品进行等量分样,然后进行样品光谱信息采集,采集结束后显示光谱曲线。小麦籽粒等量分样及回收部分通过步进电机驱动滚珠丝杆滑台模组,带动移动挡板做直线往复循环运动,实现小麦籽粒等量分样及回收功能。近红外光谱采集部分主要负责采集小

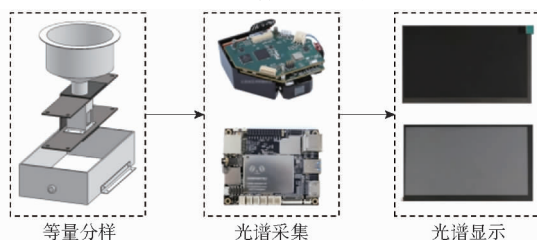


图 1 检测装置原理图

Fig. 1 Schematic of detection device

麦籽粒样品的近红外光谱信息,用于建立小麦籽粒类胡萝卜素含量检测模型。光谱显示部分则负责将结果进行可视化展示。

近红外快速采集装置装配图如图2所示,整体装置由样品杯、微型光谱仪、固定板、X86开发板、7 inch显示屏、显示屏支撑架、支撑板、待测样品盒、样品收集盒、滚珠丝杆滑台、连接板、电机、移动挡板、机架组成。

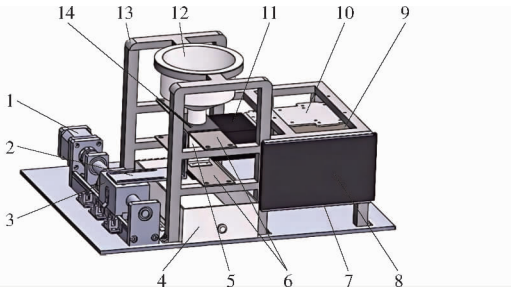


图2 近红外光谱快速采集装置装配图

Fig.2 Assembly drawing of near infrared spectrum rapid acquisition device

1. 电机 2. 连接板 3. 滚珠丝杆滑台 4. 样品收集盒 5. 待测样品盒 6. 支撑板 7. 显示屏支撑架 8. 显示屏 9. X86 开发板 10. 固定板 11. 微型光谱仪 12. 样品杯 13. 机架 14. 移动挡板

装置实物图如图3所示,整体装置外形尺寸为420 mm × 325 mm × 196 mm。其中,微型光谱仪型号为DLP 2010 NIR,由DLP 4500NIR DMD型光谱仪光学引擎、单元件扩展型InGaAs检测器、DLPC 350 DMD控制器、AM 3358处理器、ADS 1255 ADC驱动器电子元件以及卤素灯6部分组成。微型光谱仪波长范围700 ~ 2 500 nm,像素间距5.4 μm。使用装置前应对微型光谱仪进行白板校正,消除系统误差,提高测量的一致性和重复性。进行白板校正的具体步骤为:首先准备白板样品,选择一个经过验证、反射率高、反射均匀的白色标准样品;将标准白板放置在近红外光谱采集装置的测量区域内,确保白板放置平稳且无遮挡;选择“开始校正”选项,等待近红外光谱采集装置进行自动校正;校正完成后显示校正结果。如果校正成功,则继续进行后续工作;如果校正失败,则需要重新进行上述操作。

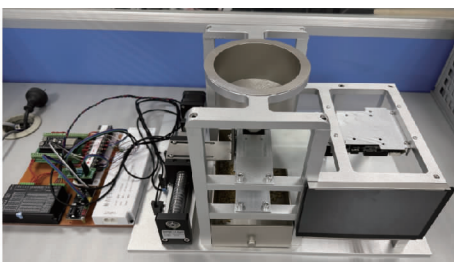


图3 近红外快速采集装置

Fig.3 Near infrared fast acquisition device

小麦籽粒类胡萝卜素含量近红外快速采集装置可以将一批小麦籽粒进行等量分样操作,并将分样后的小麦籽粒依次下落输送至待测样品盒中,采集每次分样后样品的近红外光谱数据,从而获取这一批小麦籽粒样品的平均光谱曲线。小麦籽粒近红外快速采集装置开始工作时,小麦籽粒等量分样与回收部件由步进电机驱动滚珠丝杆滑台模组带动移动挡板向左做直线运动,进料口挡块打开,小麦籽粒落入待测样品盒中,当移动挡板到达规定位置时,向右做直线运动,关闭进料口。接着近红外光谱采集部件开始工作,采集待测样品盒中小麦籽粒样品的近红外光谱数据。当小麦籽粒近红外光谱数据采集结束后,移动挡板继续向右做直线运动,打开待测样品盒下方出口,小麦籽粒落入样品收集盒。样品收集结束后,电机带动移动挡板回到初始位置,往复循环上述运动过程,多次采集小麦籽粒的近红外光谱数据,最后获取样品平均近红外光谱数据。

2 近红外光谱采集

以济麦22和济麦23两个品种的小麦籽粒作为实验对象,每个品种选取90组样本,共计样本180组。设置波长测定范围为900 ~ 1 700 nm,设置测量参数为吸光度,共228个波长点,扫描时间预计10 s,每份样品扫描10次,共计获得光谱数据1 800条,计算每个样本的平均光谱曲线作为对应样本的分析光谱进行后续数据分析^[16]。

3 小麦类胡萝卜素含量的测定

将小麦样品使用Tissue Lyser II型高效样品破碎机进行粉碎,得到全麦粉备用,按照样本编号密封避光保存,利用紫外分光光度法测定小麦类胡萝卜素含量,采用UV-3600型紫外可见分光光度计测定吸光度,每份样品设置3个重复,类胡萝卜素含量(质量比)的计算公式为^[17]

$$X = C \frac{V}{m}$$

式中 X ——类胡萝卜素含量, μg/g

C ——吸光度对应质量浓度, μg/mL

V ——类胡萝卜素溶液体积, mL

m ——全麦粉质量, g

4 数据分析与模型构建

4.1 光谱数据预处理

由于近红外光谱数据在采集过程中容易受到外界因素的干扰,因此采用SG^[18]、SNV^[19]、MSC^[20]、TC^[21]、SG + TC^[22]、SG + MSC^[23]、SG + SNV^[24]、

SG + 1D^[25]、SG + 2D^[26]、SG + 3D^[27]共 10 种预处理方法对原始光谱数据进行处理,去除干扰信息,提升光谱数据精度^[28]。通过比较不同预处理方法对模型精度的影响,确定适合该样本集的最佳预处理方法。

4.2 光谱特征选择

原始光谱数据中包含大量的无关特征变量,通过特征选择,可以有效提取关键光谱特征,采用 Relief^[29]、GA^[30]、VT^[31]、SPA^[32]共 4 种特征选择算法对原始数据进行特征筛选,通过比较全波段光谱数据与使用特征选择算法筛选得到的光谱数据对模型精度的影响,确定最佳特征选择算法。

4.3 模型的建立

建立 PLSR^[33]、SVR^[34]、ABR^[35]模型以及使用 CSO - ACF^[36-37]改进后的 ABR 模型,比较不同模型对小麦籽粒类胡萝卜素含量的预测精度,得到最优模型。

4.4 模型评价指标

采用校正集决定系数 R_c^2 和均方根误差 (RMSEC) 来评价校正集模型精度,采用预测集决定系数 R_p^2 和均方根误差 (RMSEP) 来评价预测集模型精度,使用剩余预测偏差 (RPD) 评价模型的预测能力^[38]。

5 结果与分析

5.1 全波段下预处理方法的比较

利用 SG、SNV、MSC、TC、SG + TC、SG + MSC、SG + SNV、SG + 1D、SG + 2D、SG + 3D 共 10 种预处理方法对原始光谱数据 (Raw) 进行处理,建立 PLSR 预测模型,比较原始光谱数据与预处理光谱数据对模型精度的影响,选择最佳预处理方法,结果见表 1。原始光谱数据建立预测模型后,校正集 R_c^2 为 0.88, RMSEC 为 0.36 $\mu\text{g/g}$, 预测集 R_p^2 为 0.68, RMSEP 为 0.55 $\mu\text{g/g}$ 。使用 SG、SNV、TC、SG + TC、SG + MSC、SG + SNV、SG + 1D、SG + 2D、SG + 3D 共 9 种预处理光谱数据建模效果均优于原始光谱数据。采用 SNV 预处理光谱数据建模后,校正集 R_c^2 为 0.91, RMSEC 为 0.29 $\mu\text{g/g}$, 预测集 R_p^2 为 0.77, RMSEP 为 0.51 $\mu\text{g/g}$ 。采用 TC 预处理光谱数据建模后,校正集 R_c^2 为 0.85, RMSEC 为 0.39 $\mu\text{g/g}$, 预测集 R_p^2 为 0.77, RMSEP 为 0.49 $\mu\text{g/g}$ 。对比 TC 预处理光谱数据的模型精度,使用 SNV 预处理数据建模后,校正集 R_c^2 提高了 0.06。SNV 预处理光谱数据与原始光谱数据模型精度相比,校正集 R_c^2 提高 0.03, RMSEC 降低 0.07 $\mu\text{g/g}$, 预测集 R_p^2 提高 0.09, RMSEP 降低 0.04 $\mu\text{g/g}$ 。因此,选择 SNV 预处理光

表 1 PLSR 模型下预处理方法的比较

Tab.1 Comparison of preprocessing methods under PLSR model

| 预处理方法 | 校正集 | | 预测集 | |
|----------|---------|---|---------|---|
| | R_c^2 | RMSEC/ ($\mu\text{g}\cdot\text{g}^{-1}$) | R_p^2 | RMSEP/ ($\mu\text{g}\cdot\text{g}^{-1}$) |
| Raw | 0.88 | 0.36 | 0.68 | 0.55 |
| SG | 0.86 | 0.36 | 0.70 | 0.56 |
| SNV | 0.91 | 0.29 | 0.77 | 0.51 |
| MSC | 0.91 | 0.29 | 0.67 | 0.61 |
| TC | 0.85 | 0.39 | 0.77 | 0.49 |
| SG + TC | 0.87 | 0.36 | 0.75 | 0.53 |
| SG + MSC | 0.87 | 0.35 | 0.76 | 0.52 |
| SG + SNV | 0.79 | 0.46 | 0.75 | 0.53 |
| SG + 1D | 0.87 | 0.35 | 0.73 | 0.52 |
| SG + 2D | 0.94 | 0.25 | 0.73 | 0.53 |
| SG + 3D | 0.81 | 0.44 | 0.71 | 0.56 |

谱数据进行后续数据分析。

5.2 PLSR 模型下不同特征选择算法的比较

采用 Relief、GA、VT、SPA 共 4 种特征选择算法对最佳预处理光谱数据进行特征筛选,建立 PLSR 预测模型,比较全波段光谱数据 (Full-spectrum, FS) 与使用特征选择算法筛选后的光谱数据对 PLSR 模型预测精度的影响,获得适合该模型的最佳特征选择算法,结果见表 2。使用全波段光谱数据建立预测模型后,校正集 R_c^2 为 0.91, RMSEC 为 0.29 $\mu\text{g/g}$, 预测集 R_p^2 为 0.77, RMSEP 为 0.51 $\mu\text{g/g}$, RPD 为 2.11。使用 VT 特征选择算法建立预测模型时,模型预测精度优于 Relief、GA、SPA 这 3 种特征选择算法。其中,校正集 R_c^2 为 0.84, RMSEC 为 0.40 $\mu\text{g/g}$, 预测集 R_p^2 为 0.69, RMSEP 为 0.58 $\mu\text{g/g}$, RPD 为 1.79。相比于全波段数据模型预测精度,使用特征选择算法建立预测模型,预测精度并未得到提升。导致这种结果的原因可能是在特征选择过程中,忽略了特征间的相互作用,造成了关键特征遗漏或无关特征保留。因此,基于 PLSR 模型,使用全波段数据建立模型后预测精

表 2 PLSR 模型下特征变量选取算法比较

Tab.2 Comparison of feature variable selection algorithm under PLSR model

| 特征选择算法 | 校正集 | | 预测集 | | RPD |
|--------|---------|---|---------|---|------|
| | R_c^2 | RMSEC/ ($\mu\text{g}\cdot\text{g}^{-1}$) | R_p^2 | RMSEP/ ($\mu\text{g}\cdot\text{g}^{-1}$) | |
| FS | 0.91 | 0.29 | 0.77 | 0.51 | 2.11 |
| Relief | 0.54 | 0.71 | 0.52 | 0.61 | 1.45 |
| GA | 0.70 | 0.58 | 0.61 | 0.55 | 1.61 |
| VT | 0.84 | 0.40 | 0.69 | 0.58 | 1.79 |
| SPA | 0.63 | 0.63 | 0.59 | 0.61 | 1.57 |

度优于使用 Relief、GA、VT、SPA 这 4 种特征选择算法的建模精度。

5.3 SVR 模型下不同特征选择算法的比较

将最佳预处理光谱数据分别采用 Relief、GA、VT、SPA 这 4 种特征选择算法进行特征筛选,建立 SVR 预测模型,通过比较全波段光谱数据与特征筛选后光谱数据在 SVR 模型中的预测精度,得到适合该模型的最佳特征选择算法,结果见表 3。在 SVR 预测模型中,相比全波段光谱数据,使用特征选择的光谱数据模型预测精度都得到了提升。使用 VT 特征选择算法建立模型的预测效果最佳,校正集 R_c^2 为 0.85, RMSEC 为 $0.40 \mu\text{g}/\text{g}$, 预测集 R_p^2 为 0.83, RMSEP 为 $0.42 \mu\text{g}/\text{g}$, RPD 为 2.40。使用 Relief 特征选择算法后模型预测效果次之, SPA 特征选择算法模型预测精度低于其他 3 种。使用 VT 特征选择算法与全波段数据建模效果相较,预测集 R_p^2 提高了 0.09, RPD 提高了 0.45。相比于利用 Relief、GA、SPA 这 3 种特征选择算法,使用 VT 特征选择算法建立模型后,预测集 R_p^2 分别提高了 0.03、0.05、0.06, RPD 分别提高了 0.16、0.28、0.29。因此,基于 SVR 预测模型,VT 特征选择对模型预测精度有显著提升,SVR 模型中,核函数 Kernel 选择径向基函数,核系数设置为默认值。

表 3 SVR 模型下特征变量选取算法比较

Tab. 3 Comparison of feature variable selection algorithm under SVR model

| 特征选择算法 | 校正集 | | 预测集 | | RPD |
|--------|---------|---|---------|---|------|
| | R_c^2 | RMSEC/ ($\mu\text{g}\cdot\text{g}^{-1}$) | R_p^2 | RMSEP/ ($\mu\text{g}\cdot\text{g}^{-1}$) | |
| FS | 0.98 | 0.13 | 0.74 | 0.58 | 1.95 |
| Relief | 0.81 | 0.43 | 0.80 | 0.50 | 2.24 |
| GA | 0.91 | 0.31 | 0.78 | 0.44 | 2.12 |
| VT | 0.85 | 0.40 | 0.83 | 0.42 | 2.40 |
| SPA | 0.81 | 0.45 | 0.77 | 0.44 | 2.11 |

5.4 ABR 模型下不同特征选择算法的比较

将最佳预处理光谱数据分别采用 Relief、GA、VT、SPA 这 4 种特征选择算法进行特征筛选并建立 ABR 预测模型,比较全波段光谱数据与特征筛选后光谱数据在 ABR 模型中的预测精度,获得适合该模型的最佳特征选择算法,结果见表 4。使用 Relief 特征选择算法后得到光谱数据在模型中的表现最佳,校正集 R_c^2 为 0.87, RMSEC 为 $0.37 \mu\text{g}/\text{g}$, 预测集 R_p^2 为 0.86, RMSEP 为 $0.37 \mu\text{g}/\text{g}$, RPD 为 2.70。相比于全波段数据模型预测精度,预测集 R_p^2 提高了 0.12, RPD 提高了 0.74。相比于使用 GA、VT、SPA 特征选择后的光谱数据,预测集 R_p^2 分别提高了

表 4 ABR 模型下特征变量选取算法比较

Tab. 4 Comparison of feature variable selection algorithm under ABR model

| 特征选择算法 | 校正集 | | 预测集 | | RPD |
|--------|---------|---|---------|---|------|
| | R_c^2 | RMSEC/ ($\mu\text{g}\cdot\text{g}^{-1}$) | R_p^2 | RMSEP/ ($\mu\text{g}\cdot\text{g}^{-1}$) | |
| FS | 0.90 | 0.31 | 0.74 | 0.58 | 1.96 |
| Relief | 0.87 | 0.37 | 0.86 | 0.37 | 2.70 |
| GA | 0.84 | 0.40 | 0.79 | 0.47 | 2.19 |
| VT | 0.89 | 0.33 | 0.84 | 0.39 | 2.55 |
| SPA | 0.88 | 0.36 | 0.79 | 0.42 | 2.20 |

0.07、0.02、0.07, RPD 分别提高了 0.51、0.15、0.50。因此,基于 ABR 预测模型,使用 Relief 特征选择后得到的光谱数据进行建模,得到的结果最佳,在 ABR 模型中,弱评估器数量设置为 10,学习率设置为 0.9。

5.5 ABR 模型下不同优化算法的比较

由表 2~4 结果可知,在 ABR 模型中,使用 Relief 特征选择算法建立模型后,模型预测效果优于其他特征选择方法。为进一步提升模型预测精度,对 ABR 预测模型进行改进。利用 CSO-ACF、PSO、GA 对 ABR 模型进行改进,建立预测模型,比较不同改进算法对模型预测精度的影响,结果见表 5。使用 CSO-ACF 优化算法改进后的 ABR 模型预测效果优于 PSO 和 GA 优化算法,校正集 R_c^2 为 0.90, RMSEC 为 $0.29 \mu\text{g}/\text{g}$, 预测集 R_p^2 为 0.90, RMSEP 为 $0.32 \mu\text{g}/\text{g}$, RPD 为 3.16。相比于 PSO-ABR 模型预测精度,预测集 R_p^2 提高了 0.11, RPD 提高了 0.98。相比于使用 GA 优化算法改进后的 ABR,预测集 R_p^2 提高了 0.16, RPD 提高了 1.17。因此,基于 CSO-ACF 改进算法优化后的 ABR 模型预测效果最佳。并将小麦籽粒类胡萝卜素含量实测值与预测值进行统计显著性检验,偏差 (Bias) 为 0.025 8, $P = 0.603 9 > 0.05$,即在 95% 置信区间内,实测值和预测值无显著性差异。在 CSO-ACF-ABR 模型中,设置测试数据集大小为 0.2,保留特征数设置为 10,ABR 回归模型中弱评估器数量设置为 44。

表 5 ABR 模型下不同优化算法的比较

Tab. 5 Comparison of different optimization algorithm under ABR model

| 改进算法 | 校正集 | | 预测集 | | RPD |
|---------|---------|---|---------|---|------|
| | R_c^2 | RMSEC/ ($\mu\text{g}\cdot\text{g}^{-1}$) | R_p^2 | RMSEP/ ($\mu\text{g}\cdot\text{g}^{-1}$) | |
| CSO-ACF | 0.90 | 0.29 | 0.90 | 0.32 | 3.16 |
| PSO | 0.89 | 0.33 | 0.79 | 0.48 | 2.18 |
| GA | 0.78 | 0.47 | 0.74 | 0.51 | 1.99 |

5.6 不同建模方法的比较

由表 2~5 可知,在 PLSR 模型中,使用全波段数据建模效果最好。在 SVR 模型中,使用 VT 特征选择算法后,预测效果优于其他特征选择算法。在 ABR 模型中,使用 Relief 特征选择算法后,模型预测效果优于其他特征选择算法。在使用 CSO-ACF、PSO、GA 优化后的 ABR 模型中,使用 CSO-ACF 改进算法后建模精度优于其他优化算法。与使用全波段数据建立的 PLSR 预测模型相比,预测集 R_p^2 提高 0.13,RPD 提高 1.05。与 VT-SVR 模型相比,校正集 R_c^2 提高 0.05, RMSEC 降低 0.11 $\mu\text{g/g}$, 预测集 R_p^2 提高 0.07, RMSEP 降低 0.10 $\mu\text{g/g}$, RPD 提高 0.76。与 Relief-ABR 预测模型相比,校正集 R_c^2 提高 0.03, RMSEC 降低 0.08 $\mu\text{g/g}$, 预测集 R_p^2 提高 0.04, RMSEP 降低 0.05 $\mu\text{g/g}$, RPD 提高 0.46。因此,SNV 预处理后的光谱数据,使用 Relief 特征选择算法,建立 CSO-ACF-ABR 模型预测精度最好,散点图如图 4 所示。

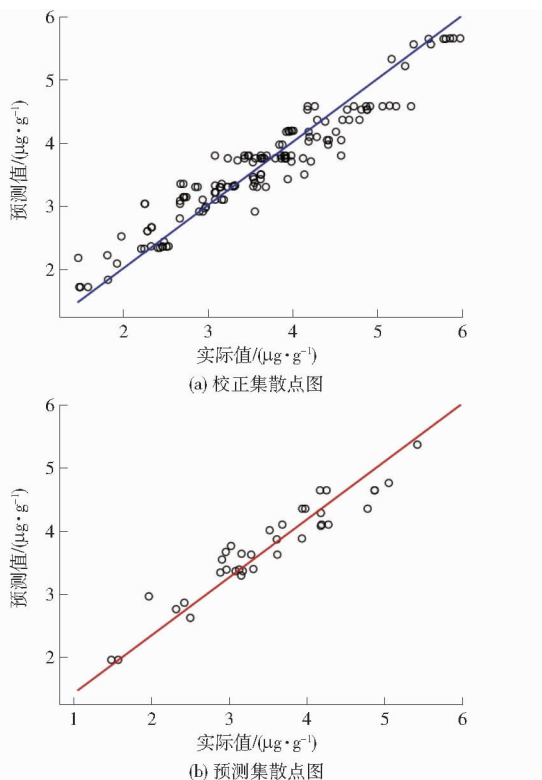


图 4 基于 SNV-Relief-CSO-ACF-ABR 的小麦类胡萝卜素含量近红外预测模型散点图

Fig. 4 Scatter plots of near-infrared prediction model for carotenoid content in wheat based on SNV-Relief-CSO-ACF-ABR

5.7 模型检测精度验证

选择不在校正集和预测集内的小麦籽粒样本用于模型检测精度验证。使用所设计的装置采集样本近红外光谱数据,每个样品采集 10 次光谱,以平均

光谱数据作为分析光谱,并使用化学方法测定小麦籽粒类胡萝卜素含量,验证模型的准确性和稳定性,试验结果如图 5 所示。结果表明,小麦籽粒类胡萝卜素含量验证决定系数 R^2 为 0.77,相比于预测集 R_p^2 有所下降,在后续研究中,应增大样本量,进一步优化模型,提高模型泛化能力和准确性。

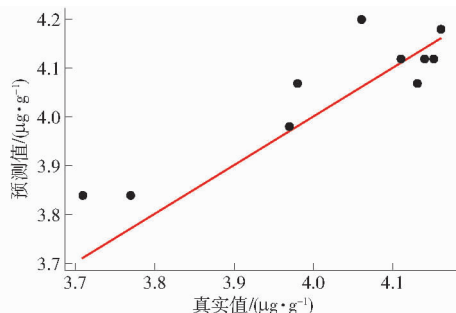


图 5 模型检测精度验证

Fig. 5 Verification of model detection accuracy

6 讨论

利用近红外光谱技术对小麦籽粒品质指标进行预测已广泛应用于农业领域。传统方法测定类胡萝卜素含量具有局限性,因此,本文设计了一种小麦籽粒近红外光谱快速采集装置并建立数学模型,对小麦籽粒类胡萝卜素含量进行预测。首先利用 SG、SNV、MSC、TC、SG+TC、SG+MSC、SG+SNV、SG+1D、SG+2D、SG+3D 共 10 种预处理方法对原始光谱数据进行处理,建立预测模型,通过比较不同预处理光谱数据模型的预测精度,得到最佳预处理方法为 SNV。

使用 Relief、GA、VT、SPA 共 4 种特征选择算法,建立 PLSR、SVR、ABR 预测模型。在 ABR 预测模型中,使用 Relief 特征选择算法得到的预测效果最佳,导致该结果的原因可能是相比于其他特征选择算法,Relief 特征选择算法可以通过评估特征的重要性去除无关特征,提升模型精度。在近红外光谱 960~980 nm、1 400~1 420 nm 内含有 O—H 基团,1 150~1 220 nm、1 360~1 390 nm 含有 C—H 基团,而类胡萝卜素分子中的 C—H 等基团在近红外光区域有特定的吸收^[13-15],Relief 特征选择后得到的 1 177、1 184、1 360、1 364、1 370 nm 附近光谱特征含有 C—H、O—H 等基团,对类胡萝卜素含量预测模型具有更为关键的意义。

ABR 模型的预测精度优于其他两种模型主要原因在于,ABR 模型可以有效处理复杂非线性关系并具有较强的适应性。但是 ABR 模型容易陷入局部最优解,存在全局搜索能力受限、收敛速度缓慢等问题。因此选择 CSO-ACF、PSO、GA 这 3 种优化算法对

ABR 模型进行改进。CSO - ACF 优化后的 ABR 模型对小麦籽粒类胡萝卜素含量的预测精度有进一步提升,CSO 算法在种群中通过交叉操作来交换个体之间的信息,从而产生新的候选解。ACF 算法则可以根据算法运行的状态,自适应地调整交叉操作的相关因子。CSO 通过交叉操作产生新个体,而 ACF 可以根据当前种群的多样性、目标函数的变化等因素,引导 CSO 的交叉操作向着更有利于收敛的方向进行。因此,CSO 和 ACF 具有跳出局部最优解、扩展搜索范围、增强全局搜索能力、加速算法收敛的能力,可以有效避免 ABR 模型存在的问题。

7 结 论

(1)根据小麦籽粒的外观形状和尺寸,设计了一种小麦籽粒近红外光谱快速采集装置,该装置外形尺寸为 420 mm × 325 mm × 196 mm,可通过调节移

动挡板位置实现小麦籽粒等量分样,快速高效采集样品的多条近红外光谱数据,获得具有代表性的平均光谱数据,提高光谱采集效率,有助于今后利用近红外光谱技术开展小麦籽粒内部品质分析工作时提高效率。

(2)采集小麦样品 900 ~ 1 700 nm 近红外光谱数据,通过比较 SG、SNV、MSC、TC、SG + TC、SG + MSC、SG + SNV、SG + 1D、SG + 2D、SG + 3D 共 10 种预处理方法和 Relief、GA、VT、SPA 共 4 种特征选择算法对 PLSR、SVR、ABR 这 3 种模型预测精度的影响,得到较好的预测模型。并使用 CSO - ACF 改进策略对 ABR 模型进行优化,改进后的 CSO - ACF - ABR 模型,校正集 R_c^2 为 0.90, RMSEC 为 0.29 $\mu\text{g/g}$, 预测集 R_p^2 为 0.90, RMSEP 为 0.32 $\mu\text{g/g}$, RPD 为 3.16。因此,改进的机器学习算法可以有效提高检测精度,为小麦类胡萝卜素含量无损预测提供理论依据。

参 考 文 献

- [1] DU Chengsi, SUN Laijun, BAI Hongyi, et al. Quantitative detection of azodicarbonamide in wheat flour by near-infrared spectroscopy based on two-step feature selection[J]. *Chemometrics and Intelligent Laboratory Systems*, 2021, 219: 104445.
- [2] NISHINO H, MURAKOSHI M, TOKUDAY H, et al. Cancer prevention by carotenoids[J]. *Archives of Biochemistry and Biophysics*, 2009, 483: 165 - 168.
- [3] 王君婵, 吴旭江, 吴迪, 等. 小麦面粉及鲜面片色泽的影响因素[J]. *麦类作物学报*, 2020, 40(5): 560 - 567.
WANG Junchan, WU Xujiang, WU Di, et al. Influencing factors of wheat flour and fresh dough sheet color[J]. *Journal of Triticeae Crops*, 2020, 40(5): 560 - 567. (in Chinese)
- [4] ROWLES J L, ERDMAN J W. Carotenoids and their role in cancer prevention[J]. *BBA-Molecular and Cell Biology of Lipids*, 2020, 1865(11): 158613.
- [5] 任得强, 吴媛媛, 周健, 等. 小麦品种(系)籽粒类胡萝卜素含量及其与其他品质性状的相关性[J]. *麦类作物学报*, 2014, 34(6): 868 - 873.
REN Deqiang, WU Yuanyuan, ZHOU Jian, et al. Analysis on carotenoids content and some other quality traits of wheat varieties (lines)[J]. *Journal of Triticeae Crops*, 2014, 34(6): 868 - 873. (in Chinese)
- [6] 孙延芳, 王成社, 杨进荣, 等. 硬粒小麦类胡萝卜素含量的定量分析[J]. *西北农林科技大学学报(自然科学版)*, 2007, 35(5): 102 - 106.
SUN Yanfang, WANG Chengshe, YANG Jinrong, et al. Study on quantitative analysis of total carotenoids content in triticum durum[J]. *Journal of Northwest A&F University (Natural Science Edition)*, 2007, 35(5): 102 - 106. (in Chinese)
- [7] QUINTELAS C, RODRIGUES C, SOUSA C, et al. Cookie composition analysis by Fourier transform near infrared spectroscopy coupled to chemometric analysis[J]. *Food Chemistry*, 2024, 435: 137607.
- [8] GAO Shang, SUN Laijun, HUI Guangyan, et al. Prediction of azodicarbonamide in flour using near-infrared spectroscopy technique[J]. *Food Analytical Methods*, 2016, 9(9): 2642 - 2648.
- [9] SCHUSTER C, HUEN J, SCHERF K A. Prediction of wheat gluten composition via near-infrared spectroscopy[J]. *Current Research in Food Science*, 2023, 6: 100471.
- [10] CHEN Run. Determination of fatty acid of wheat by near-infrared spectroscopy with combined feature selection based on CARS and NSGA-III[J]. *Infrared Physics and Technology*, 2023, 129: 104572.
- [11] 蒋华伟, 郭陶, 杨震. 基于改进粒子群优化-反向传播神经网络算法的小麦储藏品质预测模型[J]. *科学技术与工程*, 2021, 21(21): 8951 - 8956.
JIANG Huawei, GUO Tao, YANG Zhen. Prediction model of wheat storage quality based on IPSO - BPNN algorithm[J]. *Science Technology and Engineering*, 2021, 21(21): 8951 - 8956. (in Chinese)
- [12] YANG Zhongyu, CHENG Ze, SU Pengyan, et al. A model for the detection of β -glucan content in oat grain based on near infrared spectroscopy[J]. *Journal of Food Composition and Analysis*, 2024, 129: 106105.
- [13] MARTÍNEZ-VALDIVIESO D, FONT R, BLANCO-DÍAZ M T, et al. Application of near-infrared reflectance spectroscopy for predicting carotenoid content in summer squash fruit[J]. *Computers and Electronics in Agriculture*, 2014, 108: 71 - 79.
- [14] BERARDO N, BRENNA O, AMATO A, et al. Carotenoids concentration among maize genotypes measured by near infrared

- reflectance spectroscopy (NIRS)[J]. *Innovative Food Science and Emerging Technologies*, 2004, 5(3): 393–398.
- [15] BONIERBALE M, GRÜNEBERG W, AMOROS W, et al. Total and individual carotenoid profiles in *Solanum phureja* cultivated potatoes: II. Development and application of near-infrared reflectance spectroscopy (NIRS) calibrations for germplasm characterization[J]. *Journal of Food Composition and Analysis*, 2008, 22(6): 509–516.
- [16] 康明月, 罗斌, 周亚男, 等. 基于近红外光谱技术结合改进的 CS-BPNN 樱桃番茄 SSC 和 Vc 含量检测[J]. *现代食品科技*, 2023, 39(8): 287–295.
- KANG Mingyue, LUO Bin, ZHOU Ya'nian, et al. The detection of SSC and Vc content in cherry tomatoes based on near infrared spectroscopy combined with improved CS-BPNN[J]. *Modern Food Science and Technology*, 2023, 39(8): 287–295. (in Chinese)
- [17] AACC. Approved methods of the American Association of Cereal Chemists (9th edition)[M]. St Paul: American Association of Cereal Chemists Incorporation, 1995.
- [18] ZIMMERMANN B, ACHIM K. Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy[J]. *Applied Spectroscopy*, 2013, 67(8): 892–902.
- [19] GUO Peng, LI Ting, GAO Han, et al. Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using vis-nir spectroscopy[J]. *Remote Sensing*, 2021, 13(19): 4000.
- [20] MANLEY M. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials[J]. *Chemical Society Reviews*, 2014, 43(24): 8200–8214.
- [21] ZHOU Ying, ZHANG Han, SHENG Xiaohui, et al. Evaluation of saffron quality using rapid quantitative inspection technology with near-infrared spectroscopy[J]. *Molecules*, 2024, 29(17): 3983.
- [22] 方向, 王文才, 金秀, 等. 土壤速效磷可见-近红外光谱检测方法[J]. *江苏农业学报*, 2019, 35(5): 1112–1118.
- FANG Xiang, WANG Wencai, JIN Xiu, et al. Study on visible-near infrared spectroscopy for detection of available phosphorus in soil[J]. *Jiangsu Journal of Agricultural Science*, 2019, 35(5): 1112–1118. (in Chinese)
- [23] 陈华舟, 潘涛, 陈洁梅. 多元散射校正与 Savitzky-Golay 平滑模式的组合优选应用于土壤有机质的近红外光谱分析[J]. *计算机与应用化学*, 2011, 28(5): 518–522.
- CHEN Huazhou, PAN Tao, CHEN Jiemei. Combination optimization of multiple scatter correction and Savitzky-Golay smoothing modes applied to the near infrared spectroscopy analysis of soil organic matter[J]. *Computers and Applied Chemistry*, 2011, 28(5): 518–522. (in Chinese)
- [24] SADEGHI V P, HITZMANN B. Comparison of various classification techniques for supervision of milk processing[J]. *Engineering in Life Sciences*, 2021, 22(3–4): 279–287.
- [25] 韦紫玉, 斯中发, 王月. 基于近红外漫反射光谱技术的小麦蛋白质含量检测[J]. *轻工科技*, 2018, 34(5): 41–42, 57.
- WEI Ziyu, SI Zhongfa, WANG Yue. Detection of protein content in wheat based on near-infrared diffuse reflection spectroscopy technology[J]. *Light Industry Science and Technology*, 2018, 34(5): 41–42, 57. (in Chinese)
- [26] 刘红芸, 吴雪梅, 李德仑, 等. 基于高光谱技术的采摘期烟叶水分含量研究[J]. *中国农机化学报*, 2021, 42(9): 157–163.
- LIU Hongyun, WU Xuemei, LI Delun, et al. Study on the moisture content of tobacco leaves during the picking period based on hyperspectral technology[J]. *Journal of Chinese Agricultural Mechanization*, 2021, 42(9): 157–163. (in Chinese)
- [27] 汪紫阳, 尹世遼, 李颖, 等. 基于可见/近红外光谱识别东北地区常见木材[J]. *浙江农林大学学报*, 2019, 36(1): 162–169.
- WANG Ziyang, YIN Shikui, LI Ying, et al. Identification of common wood species in northeast China using Vis/NIR spectroscopy[J]. *Journal of Zhejiang A & F University*, 2019, 36(1): 162–169. (in Chinese)
- [28] 刘文政, 周雪健, 平凤娇, 等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测[J]. *农业机械学报*, 2024, 55(2): 372–383.
- LIU Wenzheng, ZHOU Xuejian, PING Fengjiao, et al. Detection of key indicators of ripening quality in table grapes based on visible-near-infrared spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(2): 372–383. (in Chinese)
- [29] URBANOWICZ J R, MEEKER M, CAVA L W, et al. Relief-based feature selection: introduction and review[J]. *Journal of Biomedical Informatics*, 2018, 85: 189–203.
- [30] SYAFRUDIN M, ALFIAN G, FITRIYANI N L, et al. A self-care prediction model for children with disability based on genetic algorithm and extreme gradient boosting[J]. *Mathematics*, 2020, 8(9): 1590.
- [31] WANG Weihao, LU Lixin, WEI Wang. A novel supervised filter feature selection method based on Gaussian probability density for fault diagnosis of permanent magnet DC motors[J]. *Sensors*, 2022, 22(19): 7121.
- [32] 高常军, 蒋侠朋, 甄佳宁, 等. 耦合 WorldView-2 和珠海一号影像的红树林物种分布[J]. *遥感学报*, 2022, 26(6): 1155–1168.
- GAO Changjun, JIANG Xiapeng, ZHEN Jianing, et al. Mangrove species classification with combination of Worldview-2 and Zhuhai-1 satellite images[J]. *National Remote Sensing Bulletin*, 2022, 26(6): 1155–1168. (in Chinese)

- LIANG Kun, SONG Jinpeng, ZHANG Chi, et al. Classification of common bunt of wheat kernels based on series fusion of scatter correction techniques[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(5): 263 – 272. (in Chinese)
- [26] MOMENY M, JAHANBAKHSI A, NESHAT A A, et al. Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks[J]. Ecological Informatics, 2022, 71: 101829.
- [27] ZHONG N, LI Y P, LI X Z, et al. Accurate prediction of salmon storage time using improved Raman spectroscopy[J]. Journal of Food Engineering, 2021, 293: 110378.
- [28] WANG X, ZHANG M, GUO Q, et al. Estimation of soil organic matter by in situ Vis – NIR spectroscopy using an automatically optimized hybrid model of convolutional neural network and long short-term memory network[J]. Computers and Electronics in Agriculture, 2023, 214: 108350.
- [29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.
- [30] LEE Y, PARK J. Centermask: real-time anchor-free instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13906 – 13915.
- [31] WOO S, PARK J, LEE J, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.
- [32] WANG Q, WU B, ZHU P, et al. ECA – Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534 – 11542.
-

(上接第 331 页)

- [33] 赵娟, 沈懋生, 浦育歌, 等. 基于近红外光谱与多品质指标的苹果出库评价模型研究[J]. 农业机械学报, 2023, 54(2): 386 – 395.
ZHAO Juan, SHEN Maosheng, PU Yuge, et al. Out-of-warehouse evaluation and prediction model of apple based on near-infrared spectroscopy combined with multiple quality indexes [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 386 – 395. (in Chinese)
- [34] YU Shui, HUAN Kewei, LIU Aoxi. Application of quantitative non-destructive determination of protein in wheat based on pretreatment combined with parallel convolutional neural network[J]. Infrared Physics & Technology, 2023, 135: 104958.
- [35] KODURI S B D, GUNISETTI L, CH R R, et al. Prediction of crop production using Adaboost regression method[J]. Journal of Physics: Conference Series, 2019, 1228 (1): 012005.
- [36] 梁昔明, 张洋, 龙文. 含有纵横交叉策略的蜘蛛猴优化算法[J]. 数学的实践与认识, 2022, 52(12): 144 – 158.
LIANG Ximing, ZHANG Yang, LONG Wen. Spider monkey optimization algorithm with crisscross optimization [J]. Mathematics in Practice and Theory, 2022, 52(12): 144 – 158. (in Chinese)
- [37] 李新宇. 改进自适应收敛因子的灰狼优化算法研究[J]. 信息与电脑(理论版), 2021, 33(24): 91 – 94.
LI Xinyu. Research on grey wolf optimization algorithm with improved adaptive convergence factor[J]. China Computer & Communication, 2021, 33(24): 91 – 94. (in Chinese)
- [38] 相志勇, 苗玉彬. 基于 CNN 和近红外光谱的蜜柑 SSC 预测模型研究[J]. 中国农机化学报, 2024, 45(11): 139 – 144.
XIANG Zhiyong, MIAO Yubin. Study on prediction model of citrus SSC based on CNN and near infrared spectroscopy[J]. Journal of Chinese Agricultural Mechanization, 2024, 45(11): 139 – 144. (in Chinese)