

基于高光谱成像技术和机器学习的猕猴桃果实可溶性固形物含量预测

刘子涵¹, 李明^{3#}, 赵峙尧¹, 陈谦², 李佳利², 于家斌^{1*}, 钱建平^{2*}

(¹北京工商大学计算机与人工智能学院, 北京 100048; ²中国农业科学院农业资源与农业区划研究所/北方干旱半干旱耕地高效利用全国重点实验室, 北京 100081; ³中国农业科学院郑州果树研究所/中国农业科学院中原研究中心/果蔬园艺作物种质创新与利用全国重点实验室, 郑州 450009, 河南新乡 453500)

摘要:【目的】可溶性固形物含量 (SSC) 是评价猕猴桃果实品质的关键指标。旨在利用高光谱技术构建猕猴桃果实 SSC 预测方案, 实现无损、准确评估果实内部品质。【方法】以米良一号猕猴桃果实为研究对象, 对高光谱图像进行白板校正、感兴趣区域提取; 光谱数据预处理采用 MSC、SG 平滑、SG-MSC 和 SG-SNV 方法以消除噪声影响, 并通过 PLSR 模型确定最优方法; 结合 CARS、SPA 和 RF 算法分别提取与果实 SSC 相关特征波段; 建立 PLSR、SVR、RFR、BPNN 模型, 比较特征波段与 SSC 实测值之间的耦合关系, 选出最优模型, 并利用 PSO 算法优化其预测精度, 以实现果实内部品质的泛化预测。【结果】MSC 方法在全波段回归中表现最佳; CARS 算法有效简化模型并提取关键特征波段; SVR 模型预测精度最高, 经 PSO 优化后训练集和测试集决定系数分别为 $R_c^2=0.949$, $R_p^2=0.913$; 均方根误差分别为 RMSEC=0.341 2, RMSEP=0.364 9。【结论】相比于单一环节的算法优化, MSC+CARS+PSO-SVR 的组合模型在猕猴桃果实可溶性固形物含量预测表现更优, 为果品品质监测和分级分选提供技术支持。

关键词: 高光谱成像技术; 猕猴桃; 可溶性固形物含量; 机器学习; 品质预测

中图分类号: S663.4 文献标志码: A 文章编号: 1009-9980(2024)12-0001-08

Prediction of Soluble Solids Content in Kiwifruit Based on Hyperspectral Imaging Technology and Machine Learning

LIU Zihan¹, LI Ming^{3#}, ZHAO Zhiyao¹, CHEN Qian², LI Jiali², YU Jiabin^{1*}, QIAN Jianping^{2*}

(¹School of Computer and Artificial Intelligence, Beijing Technology And Business University, Beijing 100048, China; ²Institute of Agricultural Resources and Regional Planning Chinese Academy of Agricultural Sciences, State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Beijing 100081, Chinese Academy of Agricultural

收稿日期: 2024-08-14 接受日期: 2024-10-21

基金项目: 国家重点研发项目 (2022YFD1600703; 2022YFF1101103; 2022YFD1600700); 黑龙江省“揭榜挂帅”科技攻关课题“规模化农场天空地一体化智慧协同感知技术研究” (2021ZXJ05A0501); 中国农科院创新工程 (CAAS-ASTIP-2023-ZFRI-03); 北京市自然科学基金项目 (4222042, 6242004); 河南省重点研发专项 (221111111800); 国家现代农业产业技术体系 (CARS-26); 河南省现代农业产业技术体系 (HARS-22-09-S)。

作者简介: 刘子涵, 女, 在读硕士研究生, 主要从事高光谱无损检测研究, E-mail: 17610409060@163.com; #为共同第一作者。

*通信作者 Author for correspondence. 于家斌, 男, 副教授, 博士, 主要从事食品风险评估、预警与决策方法研究。Tel: 13811268611, E-mail: yujiabin@th.btbu.edu.cn; 钱建平, 男, 博士, 研究员, 主要从事农产品绿色供应链与智能化追溯。Tel: 010-82108770, Email: qianjianping@caas.cn。

Sciences, Beijing 100081, China ; ³Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences/Zhongyuan Research Center, Chinese Academy of Agricultural Sciences/National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, Zhengzhou 450009, Xinxiang 453500, Henan)

Abstract: 【Objective】In the context of predicting soluble solids content (SSC) for ‘Miliang No. 1’ kiwifruit, SSC is a key quality indicator representing the concentration of soluble sugars, which are important for determining the sweetness and maturity of the fruit. Accurate and timely SSC assessment is crucial for both consumer satisfaction and market pricing. Traditional methods such as refractometry and liquid chromatography, while accurate, are time-consuming, costly, and destructive, making them unsuitable for large-scale or real-time monitoring. To address these challenges, this study aims to develop a non-destructive SSC prediction model using hyperspectral imaging technology, integrating multiple preprocessing methods, feature extraction algorithms, and machine learning models. The goal is to enhance the robustness and generalization of SSC predictions by optimizing the entire prediction process, rather than focusing on individual steps like preprocessing or feature extraction, which has been the primary focus of many previous studies.**【Methods】**This study was conducted using 150 ‘Miliang No. 1’ kiwifruit samples, which were randomly divided into a training set of 120 samples and a test set of 30 samples. Hyperspectral images were captured using a Rikola portable hyperspectral imager, covering the spectral range of 500 nm to 900 nm with a wavelength interval of 2 nm, resulting in 194 spectral bands. The imaging was conducted in a controlled dark-box laboratory environment to ensure data consistency and minimize external interference. After the hyperspectral images were captured, SSC measurements were performed using an ATAGO PAL-BX/ACID 8 refractometer. Three SSC measurements were taken for each sample, and the arithmetic mean of the three values was used as the actual SSC value. To improve the quality of the spectral data, various preprocessing methods were applied. Four specific methods were employed to enhance data consistency and eliminate noise: multiplicative scatter correction (MSC), Savitzky-Golay smoothing (SG), SG combined with MSC (SG-MSC), and SG combined with standard normal variate (SG-SNV). The optimal preprocessing method was determined based on the performance of the partial least squares regression (PLSR) model, with MSC identified as the most effective method for reducing noise and correcting baseline drift. On this basis, feature extraction was performed using competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), and random frog (RF) to identify key spectral bands most relevant to SSC. These extracted spectral bands were then used as inputs for four machine learning models: partial least squares regression (PLSR), support vector regression (SVR), random forest regression (RFR), and backpropagation neural network (BPNN). The coupling relationships between the spectral data and the actual SSC measurements were evaluated, and their predictive performances were compared. Based on the best-performing model, particle swarm optimization (PSO) was further introduced to fine-tune the model parameters, aiming to enhance both prediction accuracy and generalization ability.**【Results】**After

applying the four preprocessing methods to the spectral data, the MSC method was found to be the most effective at eliminating noise and baseline drift, leading to a significant overlap in the spectral curves. The MSC-CARS-PLSR, MSC-SPA-PLSR, and MSC-RF-PLSR models demonstrated improved performance compared to the full-band PLSR model. Specifically, the R^2 value for these models increased by 0.01 to 0.092, while the RMSEC decreased by 0.0383 to 0.1341. The three feature extraction methods were particularly successful in reducing interfering variables and improving the predictive power of the models. It was noted that the majority of the spectral feature bands identified through the feature extraction process were concentrated within the 750 nm to 900 nm range, indicating that this range is the most sensitive interval for predicting the SSC of kiwifruit. Following feature extraction, the performance of the four machine learning models was evaluated, and the MSC-CARS-SVR model was found to exhibit the best predictive performance. After PSO parameter optimization, and the comparison revealed that MSC-CARS-PSO-SVR model had the best prediction effect, with the coefficient of determination $R_c^2=0.949$, $R_p^2=0.913$, the root mean square error RMSEC=0.3412, and the RMSEP=0.3649. These results indicate that the SVR model, especially when optimized using PSO, was highly effective at handling complex, high-dimensional, and small-sample data, making it particularly well-suited for predicting SSC in kiwifruit and other quality metrics. However, the worst prediction effect was achieved by utilizing the BPNN model, in which the CARS-BPNN test set $R_p^2=0.633$, RMSEP=1.2308. It indicates that the characteristics of the dataset used in this experiment may not be applicable to neural network prediction models, as its complexity or size may not be sufficient to effectively avoid model overfitting, which in turn may lead to limited prediction performance and affect the accuracy of the results. 【 Conclusion 】 The results of this study demonstrate that the MSC-CARS-PSO-SVR model is highly effective at predicting the internal quality indicators of kiwifruit, particularly SSC. This model provides a scientific basis for non-destructive quality inspection of agricultural products. By combining data preprocessing, feature extraction, and machine learning techniques with hyperspectral imaging, the study presents a rapid, non-destructive method for SSC detection in kiwifruit. The findings offer valuable technical support for intelligent fruit quality monitoring, grading, and sorting systems, and have the potential to be applied across a wide range of fruit and agricultural products in related industries.

Key words: Hyperspectral imaging technology; Kiwifruit; Soluble solids content; Machine learning; Quality prediction

猕猴桃因富含维生素 C、膳食纤维和多种矿物质而受到消费者的喜爱^[1]。可溶性固形物含量 (soluble solids content, SSC) 主要指可溶性糖类含量, 是衡量猕猴桃果实口感甜度和成熟度的关键品质指标之一^[2-3], 直接影响消费者购买意愿和水果市场价格^[4]。因此, 实现猕

猴桃果实 SSC 的精准、快速定量检测，对于监测果实内部品质、优化种植管理以及提升市场流通竞争力具有重要意义。

目前，果品内部品质指标分析测定主要依靠高准确性的有损检测技术，包括折射仪、液相色谱法等，但在实际应用中存在着成本高、破坏性强等局限性^[5]。为满足快速、无损的果品品质监测需求，核磁共振、光谱分析和电子鼻等技术^[6]，已被广泛研究应用于检测水果的内部品质。其中，高光谱成像技术作为新兴光学检测方法，可同时获取待测对象的内外品质信息，即二维空间和一维光谱信息。二维空间信息用于直接提取待测对象的外部品质特征（如大小、形状）；将一维光谱信息与对象特定成分及含量等特征进行耦合分析，可以实现果品内部品质预测评估^[7]。因此，国内外学者已逐渐将高光谱技术用于高效、无损的果品品质监测与分级研究中。Shao 等^[8]利用可见光和近红外（Vis-NIR）高光谱成像技术监测不同成熟期的冬枣果实 SSC，以分析确定其货架期。林娇娇等^[9]利用近红外高光谱（NIR-HSI）成像技术实现了不同品种芒果 SSC 的分析预测。高光谱技术相较于传统近红外技术，具有更广的光谱覆盖范围和更高的光谱分辨率^[10]，能够捕捉果实内外部可溶性固形物含量的精细空间分布，提供更深入的皮下成分信息，从而在非破坏性检测中实现更接近于传统破坏性测量的精确性，同时具备分析更复杂成分的潜力。其在果品内部品质检测中提供了更加深入的数据基础，但仍需通过预处理如 Savitzky-Golay（SG）平滑、多元散射校正（multiplicative scatter correction, MSC）等方法来被用于预先校准光谱数据，去除环境噪声，从而提升原始光谱数据可靠性^[11]。进一步，为保留关键有效信息并实现数据降维，特征波段提取如主成分分析（principal component analysis, PCA）、遗传算法（genetic algorithm, GA）等^[12]方法被采用，以平衡模型准确性和预测效率。最终，针对不同果品及其内部品质指标的光谱响应独特性，还需明确品质指标与其敏感特征波段之间的耦合关系，以建立可泛化的果品品质预测模型。

针对光谱技术在果品品质预测中的应用，机器学习模型主要可分为线性和非线性模型两类。其中，线性模型如偏最小二乘回归（partial least squares regression, PLSR）模型、岭回归（ridge regression）模型等，已被广泛应用于光谱数据分析领域，具有可解释性强、计算效率高的优势^[13]。但面对复杂高维的光谱数据时泛化能力较弱，难以动态适应跨域监测场景。目前，非线性机器学习在处理光谱数据方面展现出良好的适应性，该模型能够自主学习特征波段与内部品质指标间的潜在耦合关系并不断优化提升预测效果，适用于大规模数据集的分析任务^[14-15]。因此，将光谱成像技术与非线性机器学习算法结合，可以充分发挥两者优势，以实现不同批次、实时、准确的果品品质无损检测。Li 等^[16]利用高光谱成像技术结合偏最小二乘回归（PLSR）、支持向量机回归（support vector machine regression, SVR）、反向传播神经网络（back propagation neural network, BPNN）以及卷积神经网络（convolutional neural network, CNN）等机器学习方法对枇杷 SSC 进行无损检测。研究结果表明，在小样本情况下非线性机器学习（SVR）预测模型的精度高于其他模型，能够实现快速、精准预测

果品内部品质指标。这些研究强调了非线性模型在光谱数据分析中展现出优越的预测性能，尤其在处理复杂数据时具有更强的适应性和泛化能力。因此，将非线性机器学习算法应用于果品品质的无损检测，不仅能够提升预测精度，还为大规模、实时果品品质检测提供了可行的解决方案。

尽管已有大量研究^[17-18]在光谱数据处理的某个环节（数据预处理、特征波段提取等）取得了显著进展，但对于整个预测流程的系统协调和优化仍存在不足，限制了模型的稳健性和普适性。因此，本研究针对米良一号猕猴桃果实的 SSC 预测，提出了一种系统优化策略，深入分析了其光谱特性，综合考虑数据预处理、特征波段提取及模型构建等多个环节，旨在提高模型的预测精度与泛化能力。首先，研究对多种光谱数据预处理方法进行了比较分析，筛选出最优的预处理方法以提升数据质量。接着，采用了多种特征波段提取方法，分别提取与猕猴桃 SSC 相关的关键光谱波段。在此基础上，构建了 PLSR、SVR、RFR 和 BPNN 模型，评估各模型与 SSC 实测值的耦合关系，并对其预测性能进行比较。基于性能最优的模型，本研究进一步引入粒子群优化算法（particle swarm optimization, PSO），对模型参数进行优化，提升预测精度和泛化能力。通过对不同环节、不同方法的组合预测效果分析，本研究构建了一套基于高光谱成像技术的猕猴桃果实 SSC 预测最适组合方案，为实现猕猴桃果实品质监测和分级分选的产业化、智能化提供理论依据。

1 材料和方法

1.1 试验材料

试验材料为米良一号猕猴桃果实，种植于河南省郑州市，试验区为中国农业科学院郑州果树研究所猕猴桃试验园。随机选择 15 棵长势一致的植株，按照 NY/T 1392—2015《猕猴桃采收与贮运技术规范》中的适宜采收期指标统一采集果实，每株采集发育良好、大小均匀一致的果实 10 个，共选取 150 个样本，装入采样箱，并运至实验室进行后续相关数据采集工作。

1.2 猕猴桃果实高光谱图像数据采集

本试验搭建了一套专门的光谱信息采集系统用于采集猕猴桃果实高光谱图像数据，包括环境模拟和数据采集模块，如图 1 所示。其中，在位移平台上放置的暗箱斜上方安装固定两枚钨卤素灯（200 W），以模拟稳定均匀的自然光照环境。将 Rikola 便携式高光谱成像仪（500~900 nm，北京德中天地）镜头朝下，架设于暗箱正上方 50 cm 处，并通过数据线连接外置装有高光谱相控软件（rikola hyper spectral imager, Rikola HSI）的计算机，实现猕猴桃果实的高光谱图像实时采集。试验开展于 2023 年 10 月 12 日，处于猕猴桃成熟期，采集过程中将猕猴桃果实标记编号，与标准白板（JY-WS1，广州景颐光电）一起均匀放置在位移平台上进行拍摄，波长间隔设置为 2 nm，共采集了 194 个波段的猕猴桃果实高光谱图像。

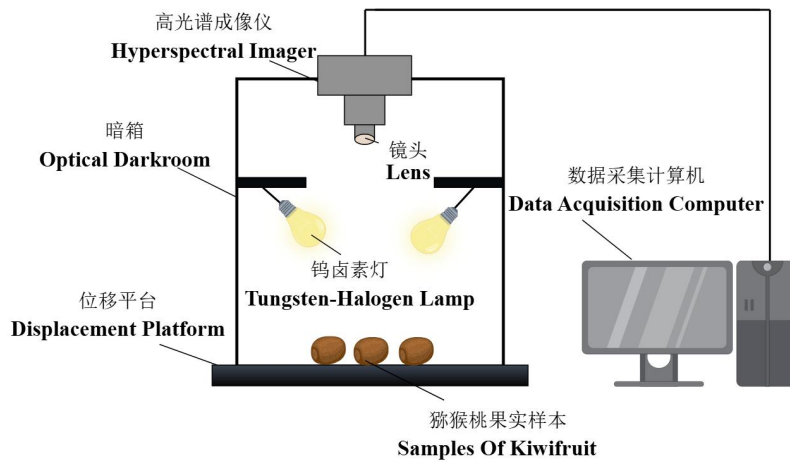


图 1 猕猴桃果实高光谱信息采集系统示意图

Fig. 1 Schematic diagram of kiwifruit fruit hyperspectral information acquisition system

1.3 高光谱数据处理

为提高采集的光谱数据质量，降低数据冗余，保障模型预测准确高效，本试验分别对原始高光谱图像进行以下数据处理：（1）白板校正，确定全白基准，保证光谱反射率数据准确性；（2）ROI（Region of Interest, ROI）区域提取，以明确有效图像数据范围；（3）光谱预处理，进一步消除噪声干扰，减小数据误差；（4）特征波段处理，实现数据降维，提取关键特征信息。相关具体操作如下：

1.3.1 白板校正

高光谱成像仪在采集光谱时，暗电流和光源亮度分布不均匀等现象会导致光谱采集对象的反射率波动，进而影响高光谱图像的整体质量。因此，需针对原始高光谱图像反射率进行白板校正^[19]，在 ENVI（Environment for Visualizing Images software, Research Systems Inc., Boulder, Co, USA）软件中选取标准白板区域的多个像素点，将其平均辐亮度作为标准值与相应波段下的反射率匹配，以校正图像中其他区域像素点的各波段反射率。校正公式如下：

$$R = \frac{R_w}{L_w} \times L \quad (1)$$

其中， R 为待计算像素点的反射率， R_w 为白板反射率， L_w 为白板像素点辐亮度值， L 为待计算像素点辐亮度值。

1.3.2 ROI 区域数据提取

在高光谱图像校正的基础上，利用 ENVI 软件选取猕猴桃果实部分作为感兴趣区域（ROI），并计算不同波段下 ROI 内所有像素点的光谱反射率平均值，作为对应样本果实的反射率。在提取并保存猕猴桃果实高光谱图像中的有效信息后，利用 MATLAB R2022b 软件执行光谱数据的后续预处理及分析与预测等相关工作。

1.3.3 光谱预处理

为进一步降低因光线、噪音、基线漂移等不确定因素造成的光谱数据误差，提高数据质量，需要对有效数据进行预处理^[20]。本研究分别采用了四种常见预处理方法对光谱数据进行优化，包括多元散射校正（MSC）、Savitzky-Golay 平滑（SG）、Savitzky-Golay 平滑结合多元散射校正（SG-MSC）和 Savitzky-Golay 平滑结合标准正态变量变换（SG-SNV）。以上四种方法各具特点，其中，MSC 可以有效降低样本散射对光谱信息的影响，从而增强光谱吸收信息与成分含量的相关性，提高吸收光谱的信噪比^[21]；SG 平滑通过拟合局部光谱趋势并去除不符合趋势的噪声成分，使光谱数据更加平滑准确^[22]；将 SG 平滑分别和 MSC、SNV 结合，前者能校正散射效应引起的光谱失真，后者可以补偿由颗粒大小和表面散射引起的偏差^[23]。为获得更清晰准确的光谱数据，本研究系统对比四种常见光谱预处理方法在猕猴桃 SSC 预测中的应用效果，结合偏最小二乘回归（PLSR）模型量化各方法的优劣，旨在筛选出最佳预处理策略，以优化数据质量，为后续建模分析与预测提供高效、准确的数据支持，从而增强整体分析流程的准确性和效率。

1.3.4 特征波段提取

由于高光谱数据具有复杂高维的特征信息，为有效消除原始数据中的线性相关性和不稳定性，需提取特征波段信息以解决维数过多等问题^[24]。本研究分别采用了竞争性自适应重加权采样（Competitive Adaptive Reweighted Sampling, CARS）、连续投影（Successive Projections Algorithm, SPA）、随机蛙跳（Random Frog, RF）三种算法提取与猕猴桃果实 SSC 变化相关的有效光谱特征波段。其中，CARS 通过模拟“生物进化”过程，自适应地对光谱波段进行重加权和选择，逐步淘汰冗余和不重要的波段，从而提高模型预测性能^[25]；SPA 算法则利用最少冗余且最具代表性的波长组合来选择光谱信息，以解决共线性问题^[26-27]；RF 算法通过序贯策略生成不同的模型，确定各波段变量的选择概率，以评估不同波段的重要性^[28]，进而实现对高维数据的有效变量筛选。这里，所提取特征光谱数据将作为猕猴桃果实 SSC 预测模型的输入训练样本。

1.4 猕猴桃果实 SSC 实测数据采集

为获取猕猴桃果实 SSC 实测数据作为其品质预测模型的预测效果验证，本试验采用 PAL-BX/ACID 8 数字折射仪（0.0~90.0 °Brix; 0.1 °Brix; ATAGO）测量样本猕猴桃果实 SSC 实测值；SSC 测定参考 NY/T 2637—2014《水果、蔬菜制品可溶性固形物含量的测定—折射仪法》，具体操作如下：

（1）对猕猴桃果实样本进行去皮处理，以消除果皮对果肉内部可溶性固形物含量测定的影响；

（2）取适量处理后的果肉进行压榨，提取出清澈的果汁样本，利用 ATAGO 数字折射仪进行测定。同时，为确保数据的准确性，对每份样本进行 3 次独立测量，计算平均值作为样本果实的 SSC 实测值。

原始数据集包含 150 个米良一号样本果实，按照 3 : 1 的比例随机划分为训练集和测试集。样本集的 SSC 含量参考值如表 1 所示。表 1 展示了训练集和测试集的最小值、最大值、平均值及标准差 (s)。

表 1 猕猴桃 SSC 含量参考值分布

Table 1 Distribution of results for SSC content in kiwifruit

样本	样本数	最小值	最大值	平均值	标准差
Sample	Number	Minimum value	Maximum value	Average value	Standard deviation
训练集	120	8.55	16.7	12.22	1.664
Training set					
测试集	30	9.24	15.9	12.42	1.641
Testing set					

1.5 果品品质预测模型建立

1.5.1 基于机器学习的猕猴桃果实 SSC 预测模型

通过对猕猴桃果实样本的 SSC 特征光谱数据及其实测值的耦合关系进行建模，可泛化实现猕猴桃果实 SSC 的无损预测。本试验利用随机数排序方法，将 150 个猕猴桃果实样本划分为训练集 (120 个) 和测试集 (30 个)，并分别采用偏最小二乘回归 (PLSR)、支持向量机回归 (SVR)、随机森林回归 (random forest regression, RFR)、反向传播神经网络 (BPNN) 四种机器学习模型来建立所需耦合模型。

为了验证不同机器学习方法在猕猴桃 SSC 预测中的效果，本研究基于各算法处理高维光谱数据的能力和适用性，采用对比分析方法评估其实际预测性能，并确定最优预测方案。常用的 PLSR、SVR 和 RFR 模型分别擅长于解决多重共线性问题、非线性回归和复杂数据的泛化问题。PLSR 模型基于主成分分析和最小二乘回归，通过探索输入变量 (样本光谱数据) 与输出变量 (SSC 预测值) 之间的最大协方差，在高度相关的样本变量下对光谱数据进行建模分析^[29]；SVR 模型在高维光谱特征空间中，通过最优超平面并最小化光谱数据点与其距离来稳定实现回归预测^[30]，对噪声和异常值具有鲁棒性，在捕捉光谱数据与目标变量关系上表现出较高准确性；RFR 模型通过综合多个决策树的预测结果，有效捕捉光谱数据与猕猴桃果实 SSC 值中的潜在非线性规律，擅长分析光谱数据与目标变量之间的复杂关联关系，有效降低过拟合风险，提高预测稳定性和泛化能力^[31]。这些算法为光谱数据建模提供了可靠的基础。

为了优化模型性能，本研究引入了反向传播神经网络 (BPNN) 模型，其利用反向传播算法优化神经网络的权重和偏置^[32]，以非线性方式学习猕猴桃果实高光谱数据与 SSC 之间的复杂映射关系，具有强适应性和精确预测特性，能够捕捉光谱数据中的细微变化。本研究

旨在通过比较四种模型在猕猴桃可溶性固形物含量（SSC）预测中的表现，系统评估其预测效果。为提升模型精度，选取表现最优的模型，借助粒子群优化算法（PSO）对模型参数进行全局搜索与优化，解决了高维光谱数据在传统算法下难以突破的精度瓶颈。通过组合常用算法和新兴算法，并引入 PSO 进行优化，本研究有效提升了猕猴桃果实 SSC 预测的准确性和模型的鲁棒性，为猕猴桃 SSC 的无损检测提供了一套更为高效可靠的最优组合方案。

粒子群优化算法（particle swarm optimization, PSO）是一种基于群体智能的优化算法，由 Kennedy 和 Eberhart 于 1995 年提出。该算法模拟鸟群觅食行为，通过一群粒子在解空间中的协同搜索找到最优解^[33]。每个粒子代表一个候选解，具有速度和位置两个属性。粒子根据自身的历史最优解（*pbset*）和全局最优解（*gbset*）来更新位置和速度，不断逼近最优解。在高光谱预测模型中，PSO 用于优化模型的参数配置。每个粒子代表一个候选的模型参数组合，并根据个体最优解和全局最优解的反馈，不断调整参数，直到找到最优参数配置。这种机制可以帮助机器学习预测模型，快速优化其参数，提高模型在猕猴桃 SSC 预测中的精度和泛化能力。

粒子速度的更新公式为：

$$v_i(t+1) = wv_i(t) + c_1r_1(pbset_i - x_i(t)) + c_2r_2(gbest - x_i(t)) \quad (2)$$

位置更新公式为：

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (3)$$

其中， v_i 表示粒子速度， x_i 表示粒子的参数值， w 为惯性权重， c_1 和 c_2 为加速因子， r_1 和 r_2 为随机数。

1.5.2 模型精度评价

这里，采用决定系数（Coefficient of determination, R^2 ）和均方根误差（Root Mean Squares Error, RMSE）两种误差指标评估模型性能。当 R^2 越接近 1，RMSE 越小（训练集和测试集的决定系数分别为 R_c^2 和 R_p^2 ，均方根误差分别为 RMSEC 和 RMSEP），表明模型的预测能力越强。其计算公式如下：

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

其中， N 是样本猕猴桃果实数量， y_i 是猕猴桃果实 SSC 实际观测值， \hat{y}_i 是猕猴桃果实 SSC 模型预测值。

本试验主要分为三部分，包括原始数据采集、光谱数据处理以及建立最适 SSC 预测模型，总体研究流程图如图 2 所示。

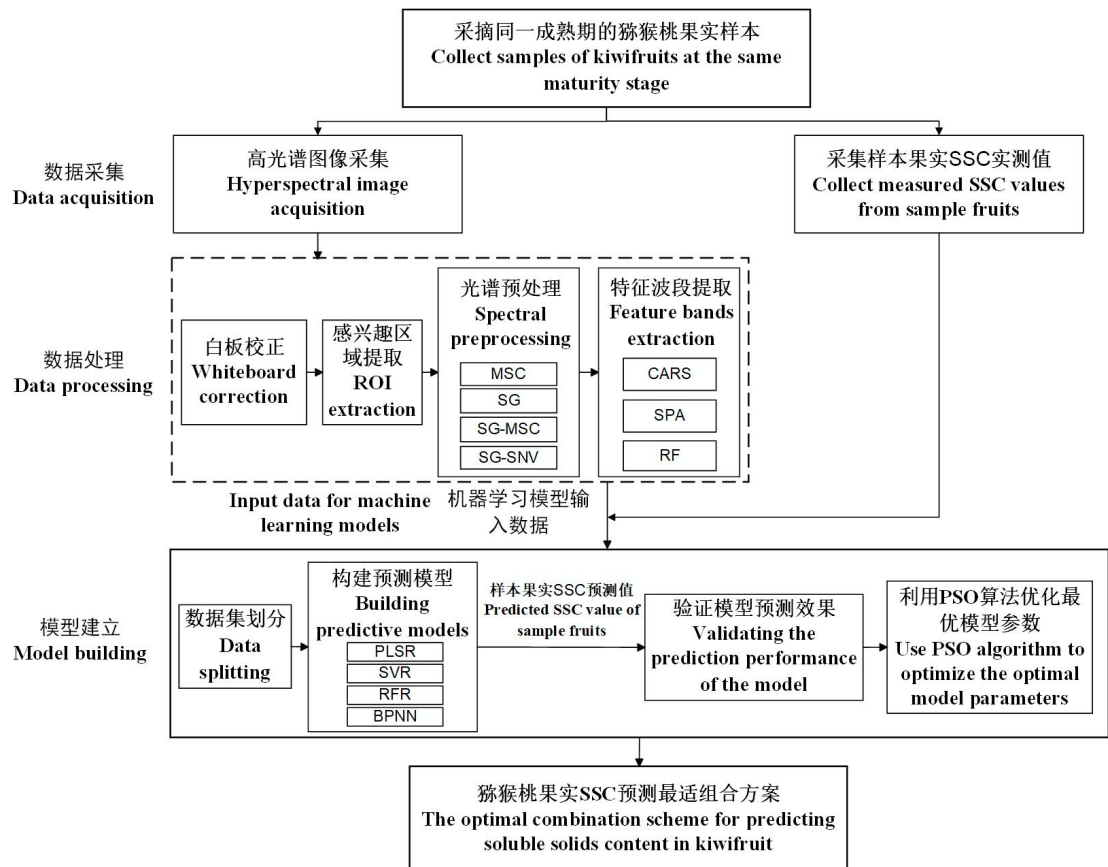


图2 总体研究流程图

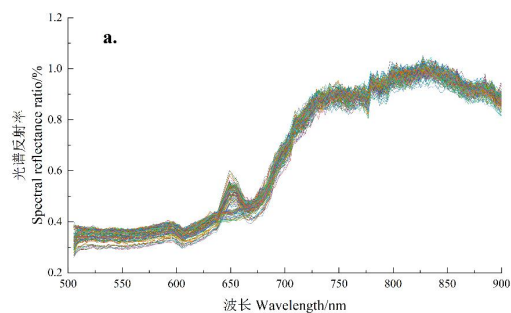
Fig. 2 The Overall Research Flowchart

2 结果与分析

2.1 基于不同预处理方式的全波段建模分析

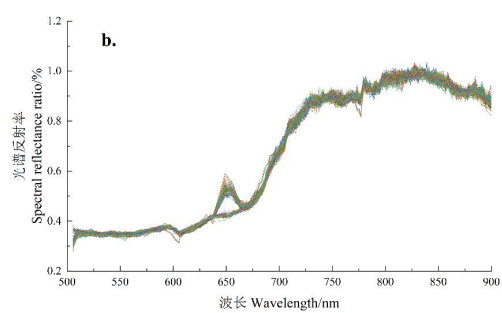
2.1.1 光谱曲线分析

本试验选取 500 nm~900 nm 范围内的光谱进行分析。首先，对 150 个米良一号猕猴桃果实样本的光谱反射率进行平均计算，并绘制平均反射率曲线，如图 3-a 所示。在 500 nm~630 nm 波段，果实的光谱反射率相对较低，这一结果与叶绿素吸收带内的情况相关，受到叶绿素中 C-H 光谱敏感基团对光的吸收影响；在 500 nm~610 nm 范围内，光谱变化缓慢；610 nm~630 nm 波段内，光谱反射率下降，继而稳定；630 nm~750 nm 波段内，光谱反射率显著增加，这表明果实表面色素的吸收性质发生了变化。在 750 nm~900 nm 范围内，光谱反射率持续高且略微波动，对应水的吸收峰，O-H 光谱敏感基团吸收率下降^[34-36]。同时原始光谱数据可能存在仪器产生的电噪声等干扰信息，为确保数据的准确性与可靠性，对原始光谱数据进行预处理^[37]，并对预处理后的光谱数据绘图，结果见图 3-b~e。



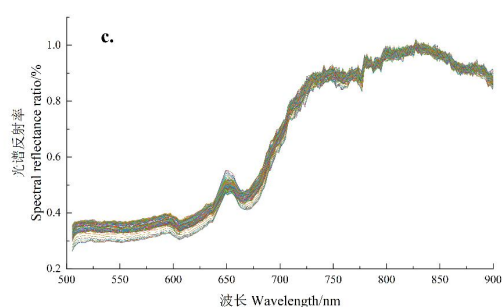
(a) 原始光谱曲线

(a) Original spectral curves



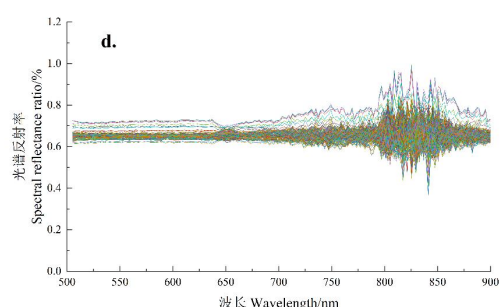
(b) MSC 处理后光谱曲线

(b) Spectral curves after MSC processing



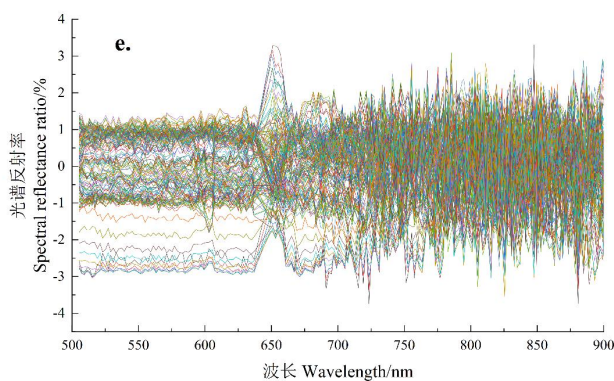
(c) SG 处理后光谱曲线

(c) Spectral curves after SG processing



(d) SG-MSC 处理后光谱曲线

(d) Spectral curves after SG-MSC processing



(e) SG-SNV 处理后光谱曲线

(e) Spectral curves after SG-SNV processing

图 3 样本果实不同预处理后的光谱曲线

Fig. 3 Spectral curves of sample fruits after different pre-treatments

在图 3-a 中, 651.43 nm 附近的光谱曲线受噪声和基线漂移的影响, 导致部分样本在该波长处未能形成明显清晰的峰值。由图 3-b 可知, 经过 MSC 预处理后, 光谱曲线高度重合, 基线漂移问题明显缓解, 但仍存在噪声影响, 部分样本在 651.43 nm 处的峰值仍不明显; SG 平滑处理能进一步提升光谱曲线的平滑度, 专注于消除局部噪声, 保留了光谱的细微反射率

差异。然而，单独使用 SG 处理可能无法全面解决问题。当将 SG 与 MSC 或 SNV 结合使用时，效果更为显著，不仅增强了光谱曲线的平滑度和特征峰的清晰度，还成功消除了 651.43 nm 处的无峰现象。虽然 SG-MSC 和 SG-SNV 处理会改变光谱的轮廓形状，但能够有效揭示潜在的光谱峰。此外，SG-SNV 虽能实现光谱标准正态化，但也伴随着噪声放大的问题。

2.1.2 全波段建模分析

分别对 MSC、SG、SG-MSC、SG-SNV 四种预处理后的光谱数据建立 PLSR 模型，以测试集的决定系数 R_p^2 和均方根误差 RMSEP 来确定最佳预处理方法，建模结果如表 2 所示。通过对比和分析发现，MSC 对样本果实光谱预处理的效果最优，其测试集的决定系数 R_p^2 大于 0.7，表明 MSC 处理后的光谱反射率与样本果实 SSC 的相关性得到显著增强，选择 MSC 处理后的光谱为研究光谱。

表 2 不同预处理方法建模结果对比

Table 2 Comparison of modelling results of different preprocessing methods

方法 Methods	训练集 Training set		测试集 Testing set	
	R_c^2	RMSEC	R_p^2	RMSEC
MSC	0.695	0.925 4	0.713	0.895 4
SG	0.641	0.945 6	0.688	0.917 8
SG-MSC	0.638	0.971 2	0.675	0.962 1
SG-SNV	0.635	0.978 8	0.592	1.084 7

2.2 特征波段提取结果分析

本试验采集的高光谱图像包含 194 个波段，使用全波段进行建模时，分析时间长且容易造成信息冗余。为此，采用 CARS、SPA、RF 三种算法对经过 MSC 预处理后的光谱数据分别进行特征波段提取，提取结果分布如图 4 所示。

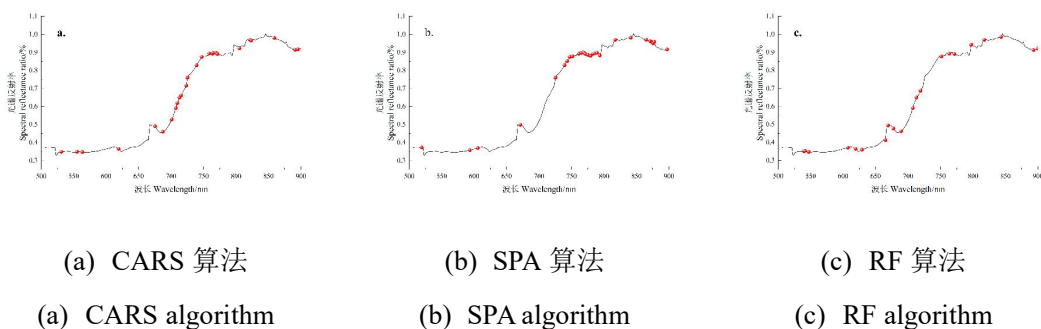


图 4 不同特征波段算法提取结果

Fig. 4 Extraction Results of Different Characteristic Wavelength Algorithms

表 3 为三种方法提取特征波段的具体结果，经 CARS 提取的 28 个特征点主要集中在 700 nm~850 nm，但分布较为不均，部分特征点位于 525 nm~700 nm；经 SPA 提取的 30 个

特征点主要集中在 725 nm~878 nm；经 RF 提取的特征点主要集中在 600 nm~700 nm 和 725 nm~825 nm。三种方法所选的特征波段主要集中在 750 nm~900 nm，主要因为在该波段范围内猕猴桃果实的光谱数据差异明显，包含更多与 SSC 值有关的光谱信息。

表 3 特征波段提取结果

Table 3 Feature bands extraction results

波段选择算法 Spectral band selection algorithm	特征波段数 Number of characteristic wavelength bands	特征波段 Characteristic wavelength bands/nm
CARS	28	531.02、554.96、563.22、619.21、675.27、687.19、701.14、707.46、709.48、710.98、715.95、723.39、725.24、739.14、749.15、761.13、763.38、765.05、767.42、769.19、771.32、805.14、821.36、823.38、825.39、858.85、891.05、893.15
SPA	30	519.11、593.31、605.5、671.14、725.24、739.14、743.28、747.25、751.34、761.13、765.05、767.42、769.19、771.32、775.33、779.28、782.79、785.33、787.30、789.22、793.39、817.32、841.21、864.95、871.19、873.31、875.36、878.56、897.22、901.34
RF	21	539.32、541.32、547.06、607.41、619.21、629.03、665.09、669.09、677.31、689.13、707.47、712.99、719.14、751.34、763.39、771.32、796.98、817.32、843.29、893.15、901.34

2.3 不同特征波段提取与机器学习方法组合的预测结果

为确定猕猴桃果实 SSC 预测最适的组合方案，本试验基于 CARS、SPA 和 RF 三种方法筛选后的特征波段光谱数据作为输入，以 SSC 预测值作为输出，分别建立了四种预测模型，包括 PLSR、SVR、RFR 以及 BPNN。表 4 为不同模型组合的预测结果。

表 4 不同特征提取方法机器学习模型结果

Table 4 Machine learning model results for different feature extraction methods

建模方法 Modeling method	输入变量 Input variable	训练集 Training set		测试集 Testing set	
		R_c^2	RMSEC	R_p^2	RMSEP
PLSR	CARS	0.78	0.791 3	0.81	0.799 7
		7		3	
	SPA	0.70	0.887 1	0.80	0.869 5
		5		7	
	RF	0.76	0.824 1	0.74	0.855 1

		2		4	
		0.93		0.88	
	CARS	0	0.388 7	2	0.526 0
		0.92		0.85	
SVR	SPA	5	0.392 2	0	0.541 2
		0.91		0.84	
	RF	4	0.426 0	1	0.591 4
		0.92		0.85	
	CARS	2	0.396 7	5	0.532 2
		0.90		0.83	
RFR	SPA	6	0.464 1	6	0.601 5
		0.89		0.82	
	RF	2	0.510 6	1	0.662 9
		0.68		0.63	
	CARS	5	1.185 6	3	1.230 8
		0.66		0.61	
BPNN	SPA	5	1.246 9	6	1.289 1
		0.64		0.62	
	RF	1	1.296 2	3	1.213 0

经过 CARS、SPA、RF 特征提取后，得到的特征点输入 PLSR 预测模型与全光谱 PLSR 模型相比，三种提取方法均提升了模型的预测效果。相较于四种预处理后的全光谱 PLSR 模型， R_c^2 提升 0.01~0.092，RMSEC 降低 0.038 3~0.134 1； R_p^2 提升 0.031~0.1，RMSEP 降低 0.025 9~0.095 7，这说明，以上三种方法有效地去除了光谱中的冗余信息，降低了数据的维度，提高了模型的精度，同时保留了用于反演 SSC 值的重要光谱信息。同时根据表 3 说明，不论采用哪种建模方法，CARS 特征波段的选取均能在不同程度上优化模型，使其在解决变量数多的同时更多保留有效信息。图 5 为基于 CARS 提取的特征波段作为输入的四种预测模型中，训练集和测试集中的样本预测值与实测值之间的关系散点图。

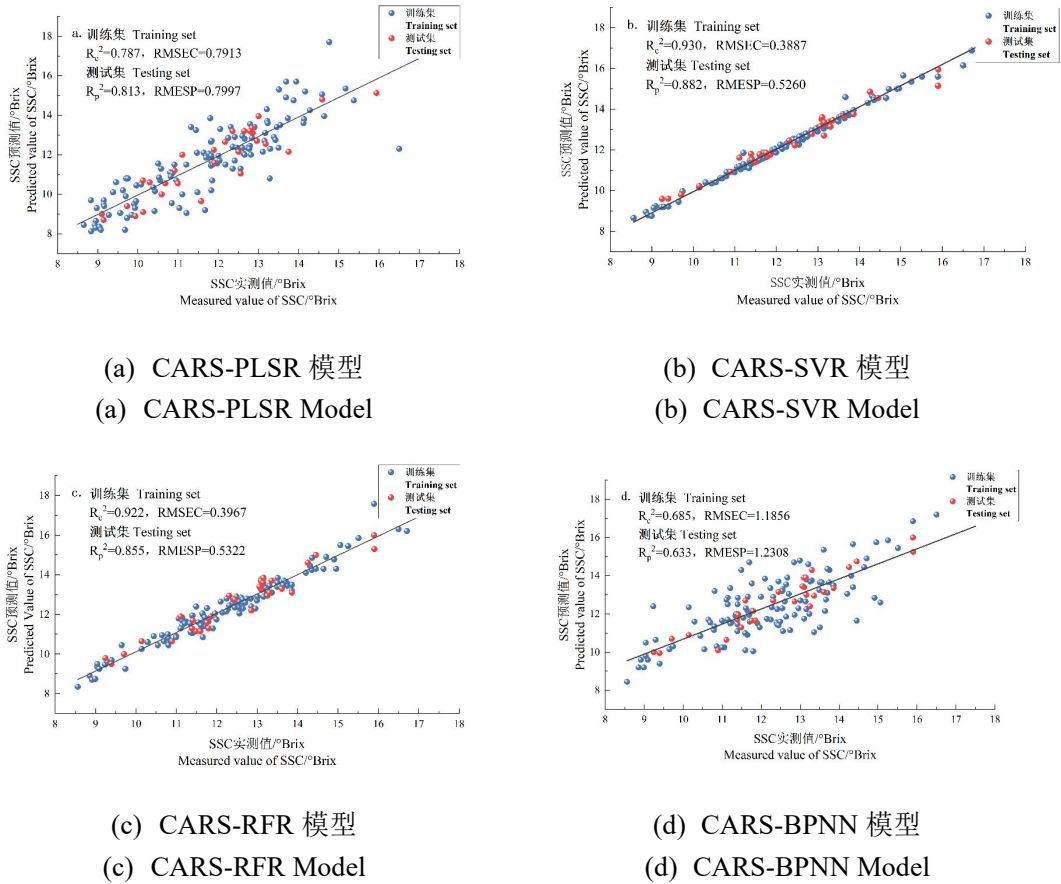


图 5 CARS 特征提取算法的不同模型预测结果

Fig. 5 Prediction results of different models of CARS feature extraction algorithm

SVR 模型大多数预测点集中在拟合曲线上,表现出较好的预测性能和稳定性。最佳模型为 CARS-SVR,训练集和测试集的决定系数分别为 $R_c^2=0.930$, $R_p^2=0.882$, RMSEC 为 0.3887, RMSEP 为 0.526 0, 其他 SVR 模型的测试集决定系数均高于 0.80, 这主要因为 SVR 能够有效处理高维数据,通过核函数将输入空间的样本数据映射到更高维的特征空间中,完成非线性变换。因此,SVR 在处理复杂、高维的小样本数据时具有显著优势。

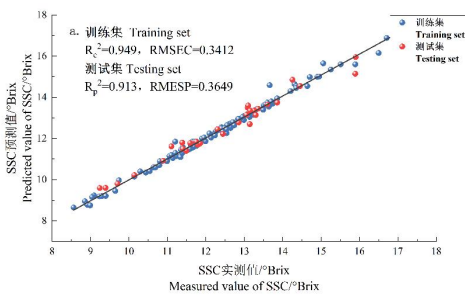
相比之下,RFR 模型对不同特征变量预测模型的均方根误差 RMSE 范围为 0.396~0.662,训练集和测试集的决定系数 R_c^2 范围为 0.892~0.922、 R_p^2 范围为 0.821~0.855。其中,CARS-RFR 模型在训练集中 ($R_c^2=0.922$ 和 $RMSEC=0.3967$) 和测试集中 ($R_p^2=0.855$, $RMSEP=0.5322$) 表现相对较好,但与 SVR 模型相比,其预测精度存在一定差距,这主要因为本试验样本数量较少,RFR 模型训练过程中可能过度适应训练集数据中的噪声,而这些噪声在测试集上并不存在泛化能力。BPNN 模型由于计算量大、训练速度慢,且易陷入局部最小值,导致其训练过程易出现过拟合或欠拟合问题^[38],因此性能不如 SVR 和 RFR 模型。然而,CARS-BPNN 模型的表现仍优于其他 BPNN 模型,三种 BPNN 模型 R^2 均在 0.616~0.685 之间, RMSE 处于 1.1~1.3 内。尽管 BPNN 预测效果相对较差,但具备一定的预测能力,后续可通过算法优化或数据增强进一步提升其性能。

表 5 经 PSO 优化后 SVR 模型结果

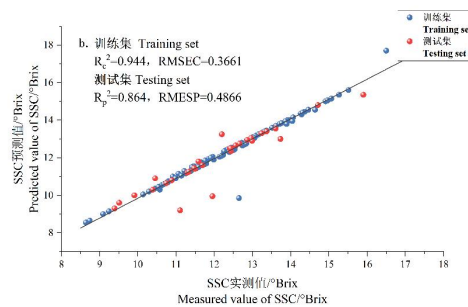
Table 5 The results of the SVR model after PSO optimization.

建模方法	输入变量	训练集 Training set		测试集 Testing set	
Modeling	Input				
method	variable	R_c^2	RMSEC	R_p^2	RMSEP
	CARS	0.94	0.341 2	0.91	0.364 9
		9		3	
PSO-SVR	SPA	0.94	0.366 1	0.86	0.486 6
		4		4	
RF	RF	0.93	0.406 0	0.85	0.645 7
		6		2	

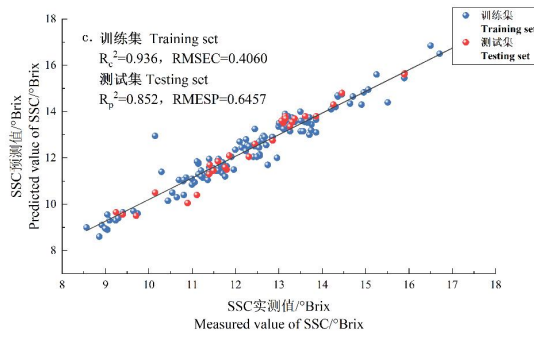
图 6 展示了经 PSO 优化后的猕猴桃果实可溶性固形物含量的最优预测模型。粒子群优化算法 (PSO) 主要用于优化 SVR 模型中的惩罚系数 c 和核函数参数 γ ，其中， c 控制模型的误差容忍度，影响模型的过拟合和欠拟合；而 γ 决定核函数的非线性映射能力。通过全局搜索，PSO 动态调整这两个参数，提升了模型的预测精度，有效避免陷入局部最优解。与基础 SVR 模型相比，PSO-SVR 模型的预测性能显著提升。在训练集和测试集上的决定系数 R_c^2 和 R_p^2 分别提升 0.019 和 0.031，RMSEC 和 RMSEP 分别降低 0.047 5 和 0.161 1。此结果与 Lin 等^[39]关于 PSO 在支持向量机模型的研究一致，表明 PSO 通过全局搜索有效优化了参数，避免了传统方法的局部最优问题。Houssein 等^[40]进一步指出，PSO 在处理复杂、高维数据时显著提升了 SVR 模型的泛化能力，尤其在小样本数据集上表现出更好的稳定性和预测精度。因此，PSO-SVR 模型在应对非线性和高维数据时表现尤为优异，为猕猴桃 SSC 预测提供了更为可靠的解决方案。



(a) CARS-PSO-SVR 模型
(a) CARS-PSO-SVR Model



(b) SPA-PSO-SVR 模型
(b) SPA-PSO-SVR Model



(c) RF-PSO-SVR 模型

(c) RF-PSO-SVR Model

图 6 不同特征提取算法的 PSO 优化 SVR 模型预测结果

Fig. 6 The prediction results of the PSO-optimized SVR model using different feature extraction algorithms.

3 讨论

本试验深入研究了猕猴桃果实可溶性固形物含量 (SSC) 与光谱反射率之间的对应关系。在 500 nm~900 nm 波段范围内, 针对高光谱数据, 采用多种预处理和特征提取方法, 并结合四种机器学习模型。为进一步提升模型的预测精度和泛化能力, 利用粒子群优化 (PSO) 算法对最优基础预测模型进行参数优化, 最终建立了猕猴桃 SSC 预测的最优方案。研究结果如下:

(1) 本研究在严格控制的实验室暗箱环境中, 使用高光谱成像技术对猕猴桃果实进行图像拍摄, 并对原始光谱图像进行白板校正, 利用 ENVI 软件中的 ROI 提取功能提取有效光谱数据。随后, 对有效光谱数据分别进行 MSC、SG、SG-MSC、SG-SNV 四种预处理。通过比较不同光谱预处理效果后, 发现 MSC 处理后的光谱曲线较原始光谱曲线更为紧凑, 能有效去除光谱噪声, 增强光谱特征之间的相关性。进一步利用 PLSR 模型对比分析四种预处理方法, 发现 MSC 处理在建模效果上优于其他三种方法, 其 $R_c^2=0.695$, $RMSEC=0.9254$; $R_p^2=0.713$, $RMSEP=0.8954$ 。这说明经过 MSC 处理后, 光谱数据的信噪比明显提升, 即信号 (果实 SSC 相关光谱特征) 相对于噪声 (如散射和仪器噪声引起的随机波动) 的比例增加。与刘美辰等^[41]的研究结果一致, MSC 预处理能够有效去除噪声, 并增强数据的一致性, 使不同样本之间的光谱数据在形状和趋势上更为接近, 有助于更容易地识别和提取与果实品质相关的特征信息, 为后续的特征提取和模型构建提供可靠的数据基础。

(2) 在建模过程中, 由于共线性问题较严重且高光谱数据冗余信息较多, 可能导致建模效果不佳, 因此需进行特征波长提取。经过 MSC 光谱预处理后, 采用 CARS、SPA 和 RF 三种特征波段提取方法, 分别获取了 28 个、30 个和 21 个特征点。波段位置数据显示, 猕猴桃果实可溶性固形物含量的光谱敏感区间主要位于 750 nm~900 nm 之间, 集中在 740 nm~800 nm 范围内。这与李浩等^[42]的研究结果相符, 经过 CARS 筛选波长后, 输入特征数量明显减少, 特征点间的相关性也得到增强。通过对比特征提取前后的 PLSR 模型, R^2 表

现出显著提升, RMSE 明显降低。这三种特征波段提取方法均有效地简化了模型结构, 显著提高了预测模型的准确性和效率。

(3) 构建了四个用于预测猕猴桃果实可溶性固形物含量的机器学习模型, 包括 PLSR、SVR、RFR、BPNN。模型输入为光谱特征波段, 输出为 SSC 预测值。通过比较这四个模型的 R^2 和 RMSE 发现, SVR 预测性能最佳, 具有更好的拟合效果, 预测结果的偏离程度和误差更小。此外, 利用 CARS 方法提取的特征波段建模效果最优, 不仅提高了模型的预测速度和精度, 同时提取的波段数量较少。最终确定 MSC-CARS-SVR 为最优模型, 其训练集和测试集的决定系数分别为 $R_c^2=0.930$, $R_p^2=0.882$; 均方根误差分别为 RMSEC=0.387 7, RMSEP=0.526 0。进一步利用粒子群优化算法 (PSO) 对 SVR 参数进行优化后, 模型性能显著提升, 在训练集和测试集上的决定系数 R_c^2 和 R_p^2 分别提升 0.019 和 0.031, RMSEC 和 RMSEP 分别降低 0.047 5 和 0.161 1。PSO 通过全局搜索和信息交换机制有效避免陷入局部最优解, 显著提高了 SVR 模型在小样本和高维光谱数据中的预测精度和泛化能力, 展现出在小样本条件下的优越性能。与董金磊等^[43]利用高光谱技术结合 SPA 算法和 BPNN 模型预测猕猴桃果实 SSC 的结果相比, 本研究的训练集相关系数 R_c^2 提高了 0.019, 均方根误差 RMSEC 减少了 0.554 8; 说明采用各环节的最优策略组合模型能够更好地预测猕猴桃 SSC。刘文政等^[44]的研究也表明, SVR 模型在预测葡萄中的总酚和单宁含量时, 预测性能优于 CNN 和 PLSR 模型, 进一步证实了 SVR 在处理非线性关系和噪声数据时的优势。

相比之下, BPNN 模型虽然具有较强的非线性解释能力, 但在本研究中 BPNN 模型预测效果最差, 其中 MSC-CARS-BPNN 的测试集 $R_p^2=0.633$, RMSEP=1.230 8。这可能是由于 BPNN 容易在训练过程中过度学习数据细节, 导致过拟合, 从而在新数据上表现出较差的泛化能力^[45]。尽管 BPNN 能够在一定程度上缓解低估和高估现象, 但其表现受于模型结构、测定时期和数据特性。此外罗浪琴等^[46]利用 BP 神经和 SVR 结合近红外光谱技术预测核桃仁可溶性蛋白质含量的研究结果与本研究不一致。其原因可能在于数据预处理方法和模型选择的差异, 说明特征提取和模型优化在不同研究中的预测性能具有关键影响。

目前, 本研究在建模数据方面主要关注猕猴桃单一品种和单一品质指标的预测。随着实际农业生产对不同成熟期、不同品种以及不同果园预测需求的增加, 本研究方法的适用性仍需进一步验证和优化。未来研究将扩展实验和建模分析范围, 重点关注不同成熟度果实以及多品质指标, 以期实现更全面、准确的水果内部品质指标预测, 有助于推动水果品质检测与分级技术的持续改进。

4 结论

本研究通过针对猕猴桃果实 SSC 的快速无损检测, 建立了一套基于高光谱技术的数据预处理、特征波段提取及机器学习预测的最优组合方案。结果表明, MSC-CARS-SVR 模型表现最佳。通过粒子群优化 (PSO) 算法优化 SVR 模型参数, 其测试集的决定系数 R_p^2 为 0.913, 均方根误差 RMSEP 为 0.364 9, 表明优化后的 SVR 模型能够显著提升预测精度, 有效地预

测猕猴桃果实的内部品质指标。该研究为农产品无损品质检测提供了科学依据，并为猕猴桃果实品质分级的精细化、智能化管理提供了便捷高效的技术手段。

参考文献 References:

- [1] 刘笑宏, 赵玲玲, 牟红梅, 唐美玲, 慈志娟, 肖慧琳, 苏佳明. 猕猴桃采后保鲜技术研究进展[J]. 保鲜与加工, 2021, 21(11): 121-128.
LIU Xiaohong, ZHAO Lingling, MU Hongmei, TANG Meiling, CI Zhijuan, XIAO Huilin, SU Jiaming. Research progress on preservation technology for postharvest kiwifruit[J]. Storage and Process, 2021, 21(11): 121-128.
- [2] MA T, XIA Y, INAGAKI T, TSUCHIKAWA S. Non-destructive and fast method of mapping the distribution of the soluble solids content and pH in kiwifruit using object rotation near-infrared hyperspectral imaging approach[J]. Postharvest Biology and Technology, 2021, 174: 111440.
- [3] ESCRIBANO S, BIASI W V, LERUD R, SLAUGHTER D C, MITCHAM E J. Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries[J]. Postharvest Biology and Technology, 2017, 128: 112-120.
- [4] LI L, PENG Y K, YANG C, LI Y Y. Optical sensing system for detection of the internal and external quality attributes of apples[J]. Postharvest Biology and Technology, 2020, 162: 111101.
- [5] LI J L, SUN D W, CHENG J H. Recent advances in nondestructive analytical techniques for determining the total soluble solids in fruits: a review[J]. Comprehensive Reviews in Food Science and Food Safety, 2016, 15(5): 897-911.
- [6] 孙静涛, 罗一甲, 史学伟, 马本学, 王文霞, 董娟. 葡萄品质无损检测技术的研究进展[J]. 光谱学与光谱分析, 2020, 40(9): 2713-2720.
SUN Jingtao, LUO Yijia, SHI Xuwei, MA Benxue, WANG Wenxia, DONG Juan. Research progress on non-destructive detection technology for grape quality[J]. Spectroscopy and Spectral Analysis, 2020, 40(9): 2713-2720.
- [7] TIAN P, MENG Q H, WU Z F, LIN J J, HUANG X, ZHU H, ZHOU X L, QIU Z Q, HUANG Y Q, LI Y. Detection of mango soluble solid content using hyperspectral imaging technology[J]. Infrared Physics & Technology, 2023, 129: 104576.
- [8] SHAO Y Y, JI S H, XUAN G T, WANG K L, XU L Q, SHAO J. Soluble solids content monitoring and shelf life analysis of winter jujube at different maturity stages by Vis-NIR hyperspectral imaging[J]. Postharvest Biology and Technology, 2024, 210: 112773.
- [9] 林娇娇, 蒙庆华, 吴哲锋, 常洪娟, 倪淳宇, 邱邹全, 李华荣, 黄玉清. 基于近红外高光谱技术的杧果可溶性固形物含量无损检测[J]. 果树学报, 2024, 41(1): 122-132.
LIN Jiaojiao, MENG Qinghua, WU Zhefeng, CHANG Hongjuan, NI Chunyu, QIU Zouquan, LI Huarong, HUANG Yuqing. Fruit soluble solids content non-destructive detection based on visible/near infrared hyperspectral imaging in mango[J]. Journal of Fruit Science, 2024, 41(1): 122-132.

- [10] BHARGAVA A, SACHDEVA A, SHARMA K, ALSHARIF M H, UTHANSAKUL P, UTHANSAKUL M. Hyperspectral imaging and its applications: a review[J]. *Heliyon*, 2024, 10(12): e33208.
- [11] 孙嘉豪, 张伟, 施鉴芬, 李艳坤. 光谱数据预处理策略选择及应用[J]. *计量学报*, 2023, 44(8): 1284-1292.
- SUN Jiahao, ZHANG Wei, SHI Jianqin, LI Yankun. Selection and application of spectral data preprocessing strategy[J]. *Acta Metrologica Sinica*, 2023, 44(8): 1284-1292.
- [12] 鲍浩, 张艳. 基于改进 Harris 鹰优化算法的光谱特征波段选择模型研究[J]. *光谱学与光谱分析*, 2024, 44(1): 148-157.
- BAO Hao, ZHANG Yan. Research on spectral feature band selection model based on improved Harris hawk optimization algorithm[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(1): 148-157.
- [13] 浦育歌. 基于可见/近红外光谱的苹果霉心病与可溶性固形物在线检测方法研究[D]. 杨凌: 西北农林科技大学, 2023.
- PU Yuge. Research on online detection methods for apple modly core disease and soluble solids based on visible/near-infrared spectroscopy[D]. Yangling: Northwest A & F University, 2023.
- [14] PRAJAPATI A, DEHAL A, KUMAR A R. Microplastics in soils and sediments: A review of characterization, quantitation, and ecological risk assessment[J]. *Water, Air, & Soil Pollution*, 2024, 235(3): 189.
- [15] ÖZGENÇ E. Advanced analytical techniques for assessing and detecting microplastic pollution in water and wastewater systems[J]. *Environmental Quality Management*, 2024, 34(1): 22217.
- [16] LI S Y, SONG Q M, LIU Y J, ZENG T H, LIU S Y, JIE D F, WEI X. Hyperspectral imaging-based detection of soluble solids content of loquat from a small sample[J]. *Postharvest Biology and Technology*, 2023, 204: 112454.
- [17] 高升, 徐建华. 高光谱成像的红提总酸与硬度的预测及其分布可视化[J]. *食品科学*, 2023, 44(2): 327-336.
- GAO Sheng, XU Jianhua. Hyperspectral imaging for prediction and distribution visualization of total acidity and hardness of red globe grapes[J]. *Food Science*, 2023, 44(2): 327-336.
- [18] 霍迎秋, 张晨, 李宇豪, 智文涛, 张炯, 刘景玲. 高光谱图像结合机器学习方法无损检测猕猴桃[J]. *中国农机化学报*, 2019, 40(4): 71-77.
- HUO Yingqiu, ZHANG Chen, LI Yuhao, ZHI Wentao, ZHANG Jiong, LIU Jingling. Nondestructive detection for kiwifruit based on the hyperspectral technology and machine learning[J]. *Journal of Chinese Agricultural Mechanization*, 2019, 40(4): 71-77.
- [19] 杨涵, 陈谦, 王宝刚, 李文生, 李文志, 王炳策, 钱建平. 利用高光谱技术预测采前猕猴桃干物质含量的可行性试验[J]. *农业工程学报*, 2022, 38(13): 133-140.
- YANG Han, CHEN Qian, WANG Baogang, LI Wensheng, LI Wenzhi, WANG Bingce, QIAN Jianping. Feasibility of estimating the dry matter content of kiwifruits before being harvested using hyperspectral technology[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(13): 133-140.

- [20] 郑丽娜. 基于高光谱技术的猕猴桃内部品质检测研究[D]. 雅安: 四川农业大学, 2019.
ZHENG Lina. Study on internal quality detection of kiwifruit based on hyperspectral technology[D]. Ya'an: Sichuan Agricultural University, 2019.
- [21] 沈兵兵, 姚星伟, 王怀文. 基于高光谱技术的花椰菜农药残留检测[J]. 包装工程, 2022, 43(19): 173-179.
SHEN Bingbing, YAO Xingwei, WANG Huaiwen. Detection of pesticide residues in cauliflower based on hyperspectral technology[J]. Packaging Engineering, 2022, 43(19): 173-179.
- [22] 王迪, 冯伟华, 郭军伟, 王锐, 刘惠民, 宗国浩, 刘绍锋, 王永胜, 赵乐. 基于 Savitzky-Golay 平滑插值的烟草近红外光谱模型转移[J]. 烟草科技, 2022, 55(8): 41-48.
WANG Di, FENG Weihua, GUO Junwei, WANG Rui, LIU Huimin, ZONG Guohao, LIU Shaofeng, WANG Yongsheng, ZHAO Le. Tobacco near infrared spectral model transfer based on Savitzky-Golay smooth interpolation[J]. Tobacco Science & Technology, 2022, 55(8): 41-48.
- [23] 刘昊灵, 张仲雄, 陈昂, 浦育歌, 赵娟, 胡瑾. 融合光谱形态特征的苹果霉心病检测方法[J]. 农业工程学报, 2023, 39(1): 162-170.
LIU Haoling, ZHANG Zhongxiong, CHEN Ang, PU Yuge, ZHAO Juan, HU Jin. Detection method for apple moldy cores based on spectral shape features[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(1): 162-170.
- [24] 李庆旭, 王巧华, 马美湖, 肖仕杰, 施行. 基于可见/近红外光谱和深度学习的早期鸭胚雌雄信息无损检测[J]. 光谱学与光谱分析, 2021, 41(6): 1800-1805.
LI Qingxu, WANG Qiaohua, MA Meihu, XIAO Shijie, SHI Hang. Non-destructive detection of male and female information of early duck embryos based on visible/near infrared spectroscopy and deep learning[J]. Spectroscopy and Spectral Analysis, 2021, 41(6): 1800-1805.
- [25] LIANG L, WEI L L, FANG G G, XU F, DENG Y J, SHEN K Z, TIAN Q W, WU T, ZHU B P. Prediction of holocellulose and lignin content of pulp wood feedstock using near infrared spectroscopy and variable selection[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 225: 117515.
- [26] YUAN R R, LIU G S, HE J G, WAN G L, FAN N Y, LI Y, SUN Y R. Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis[J]. Computers and Electronics in Agriculture, 2021, 182: 106043.
- [27] ZHU G Z, TIAN C N. Determining sugar content and firmness of 'Fuji' apples by using portable near-infrared spectrometer and diffuse transmittance spectroscopy[J]. Journal of Food Process Engineering, 2018, 41(6): e12810.
- [28] 张楷鑫. 基于光谱成像技术的煤岩特征信息检测与识别方法研究[D]. 西安: 西安科技大学, 2021.
ZHANG Kaixin. Research on detection and recognition of coal and rock feature information based on spectral imaging technology[D]. Xi'an: Xi'an University of Science and Technology, 2021.
- [29] OLAREWAJU O O, BERTLING I, MAGWAZA L S. Non-destructive evaluation of avocado fruit

maturity using near infrared spectroscopy and PLS regression models[J]. *Scientia Horticulturae*, 2016, 199: 229-236.

[30] 汪晓慧. 基于高光谱荧光成像技术的水蜜桃品质参数无损检测研究[D]. 雅安: 四川农业大学, 2023.

WANG Xiaohui. Study on non-destructive detection of peach quality parameters based on hyperspectral fluorescence imaging technology[D]. Ya'an: Sichuan Agricultural University, 2023.

[31] CHEN D S, ZHANG F, TAN M L, CHAN N W, SHI J C, LIU C J, WANG W W. Improved Na+ estimation from hyperspectral data of saline vegetation by machine learning[J]. *Computers and Electronics in Agriculture*, 2022, 196: 106862.

[32] YE F. Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data[J]. *PLoS One*, 2017, 12(12): e0188746.

[33] GAD A G. Particle swarm optimization algorithm and its applications: A systematic review[J]. *Archives of Computational Methods in Engineering*, 2022, 29(5): 2531-2561.

[34] WEI X, HE J C, YE D P, JIE D F. Navel orange maturity classification by multispectral indexes based on hyperspectral diffuse transmittance imaging[J]. *Journal of Food Quality*, 2017, 2017: 1023498.

[35] LI X L, WEI Y Z, XU J, FENG X P, WU F Y, ZHOU R Q, JIN J J, XU K W, YU X J, HE Y. SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology[J]. *Postharvest Biology and Technology*, 2018, 143: 112-118.

[36] 郑艺蕾. 基于高光谱和太赫兹光谱的甘薯品质检测方法研究[D]. 南昌: 华东交通大学, 2020.
ZHENG Yilei. Research on sweet potato quality detection method based on hyperspectral and terahertz spectroscopy[D]. Nanchang: East China Jiaotong University, 2020.

[37] 宋相中, 熊艳梅, 张录达, 闵顺耕. 分子光谱波长选择值得注意的几个问题[J]. *光谱学与光谱分析*, 2016, 36 (增刊 1): 181-182.

SONG Xiangzhong, XIONG Yanmei, ZHANG Luda, MIN Shungeng. Several notable problems of wavelength selection in molecular spectroscopy area[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(Suppl. 1): 181-182.

[38] ZHU N, WANG K, ZHANG S L, ZHAO B, YANG J N, WANG S W. Application of artificial neural networks to predict multiple quality of dry-cured ham based on protein degradation[J]. *Food Chemistry*, 2021, 344: 128586.

[39] LIN S W, YING K C, CHEN S C, LEE Z J. Particle swarm optimization for parameter determination and feature selection of support vector machines[J]. *Expert Systems with Applications*, 2008, 35(4): 1817-1824.

[40] HOUSSEIN E H, GAD A G, HUSSAIN K, SUGANTHAN P N. Major advances in particle swarm optimization: Theory, analysis, and application[J]. *Swarm and Evolutionary Computation*, 2021, 63: 100868.

[41] 刘美辰, 薛河儒, 刘江平, 代荣荣, 胡鹏伟, 黄清, 姜新华. 牛奶蛋白质含量的 SSA-SVM 高光谱预测模型[J]. 光谱学与光谱分析, 2022, 42(5): 1601-1606.

LIU Meichen, XUE Heru, LIU Jiangping, DAI Rongrong, HU Pengwei, HUANG Qing, JIANG Xinhua. Hyperspectral analysis of milk protein content using SVM optimized by sparrow search algorithm[J]. Spectroscopy and Spectral Analysis, 2022, 42(5): 1601-1606.

[42] 李浩, 于滢, 曹永研, 郝子源, 杨玮, 李民赞. 利用 CARS-CNN 模型的土壤有机质含量高光谱预测[J]. 光谱学与光谱分析, 2024, 44(8): 2303-2309.

LI Hao, YU Hao, CAO Yongyan, HAO Ziyuan, YANG Wei, LI Minzan. Hyperspectral prediction of soil organic matter content using CARS-CNN modelling[J]. Spectroscopy and Spectral Analysis, 2024, 44(8): 2303-2309.

[43] 董金磊, 郭文川. 采后猕猴桃可溶性固形物含量的高光谱无损检测[J]. 食品科学, 2015, 36(16): 101-106.

DONG Jinlei, GUO Wenchuan. Nondestructive detection of soluble solid content of postharvest kiwifruits based on hyperspectral imaging technology[J]. Food Science, 2015, 36(16): 101-106.

[44] 刘文政, 周雪健, 平凤娇, 苏媛, 鞠延仑, 房玉林, 杨继红. 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, SU Yuan, JU Yanlun, FANG Yulin, YANG Jihong. 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.

[45] 王丽爱, 马昌, 周旭东, 警妍, 朱新开, 郭文善. 基于随机森林回归算法的小麦叶片 SPAD 值遥感估算[J]. 农业机械学报, 2015, 46(1): 259-265.

WANG Liai, MA Chang, ZHOU Xudong, ZI Yan, ZHU Xinkai, GUO Wenshan. Estimation of wheat leaf SPAD value using RF algorithmic model and remote sensing data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 259-265.

[46] 罗浪琴, 王涛, 刘国庆, 赵文革, 张锐, 于军, 陆斌, 陈天财. 基于近红外光谱法建立核桃仁可溶性蛋白质含量检测模型[J]. 果树学报, 2023, 40(8): 1750-1761.

LUO Langqin, WANG Tao, LIU Guoqing, ZHAO Wenge, ZHANG Rui, YU Jun, LU Bin, CHEN Tiancai. A model for soluble protein content detection of walnuts based on near infrared spectroscopy[J]. Journal of Fruit Science, 2023, 40(8): 1750-1761.