

基于高光谱技术的蓝莓SSC无损检测通用模型构建与分析

陈国良, 杨冕清, 代景源, 汪国政, 刘大洋*

(东北林业大学计算机与控制工程学院, 黑龙江哈尔滨 150006)

摘要: 为了开发适应生产线随机摆放场景的蓝莓 SSC 通用检测模型, 该研究采集了 300 颗蓝莓样本的 SSC 值及其 4 个面的高光谱数据, 分别使用偏最小二乘回归、支持向量回归和长短期记忆网络 (Long Short-Term Memory, LSTM) 构建局部模型和基于多面光谱数据的通用模型。结果表明, 基于单一面光谱数据的局部模型在跨面预测时表现较差, 而基于多面光谱数据构建的通用模型显著提高了泛化能力。特别是, LSTM 通用模型在外部验证集上对 4 个面的预测相关系数分别为 0.963、0.954、0.964 和 0.962; 预测均方根误差分别为 0.513%、0.521%、0.472% 和 0.483%; 剩余预测偏差分别为 3.32、3.27、3.61 和 3.52。该研究验证了基于多面光谱数据构建 SSC 通用检测模型的可行性, 并证明 LSTM 模型在蓝莓 SSC 检测中的潜在应用价值。研究构建的 LSTM 通用检测模型为生产线上蓝莓品质无损检测提供了新型技术方案, 其多面光谱融合方法可为其他小型水果品质检测模型开发提供理论参考。

关键词: 蓝莓; 高光谱成像; 可溶性固形物含量; 通用预测模型

文章编号: 1673-9078(2026)3-333-340

DOI: 10.13982/j.mfst.1673-9078.2026.3.1915

Construction and Analysis of A Universal Model for the Non-destructive Detection of Blueberry SSC Based on Hyperspectral Technology

CHEN Guoliang, YANG Mianqing, DAI Jingyuan, WANG Guozheng, LIU Dayang*

(College of Computer and Control Engineering, Northeast Forestry University, Harbin 150006, China)

Abstract: To develop a universal blueberry soluble solids content (SSC) detection model adaptable to randomly arranged production line scenarios, this study collected the SSC values of 300 blueberry samples and their hyperspectral data on four surfaces. Partial least-squares regression, support vector regression, and Long Short-Term Memory (LSTM) were applied to construct both local and universal models based on multi-surface spectral data. The results indicated that the local models constructed using single-surface spectral data performed poorly in cross-surface prediction. In contrast, the universal models, constructed based on multi-surface spectral data, significantly improve generalizability. Specifically, prediction correlation coefficients of the universal LSTM model on the external validation set for the four surfaces were 0.963, 0.954, 0.964, and 0.962, respectively. The root-mean-square errors of the predictions were 0.513%, 0.521%, 0.472%, and 0.483%, respectively, and the residual prediction biases were 3.32, 3.27, 3.61, and 3.52, respectively. This study validates the feasibility of constructing a universal SSC detection model based on multi-

引文格式:

陈国良, 杨冕清, 代景源, 等. 基于高光谱技术的蓝莓 SSC 无损检测通用模型构建与分析[J]. 现代食品科技, 2026, 42(3): 333-340.

CHEN Guoliang, YANG Mianqing, DAI Jingyuan, et al. Construction and analysis of a universal model for the non-destructive detection of blueberry SSC based on hyperspectral technology [J]. Modern Food Science and Technology, 2026, 42(3): 333-340.

收稿日期: 2024-12-24; 修回日期: 2025-03-14; 接受日期: 2025-03-17

基金项目: 国家自然科学基金项目 (32202147); 中国博士后基金项目 (2021M690573); 中央高校基本科研业务费专项资金项目 (2572020BF05)

作者简介: 陈国良 (1999-), 男, 硕士研究生, 研究方向: 智能化检测与技术, E-mail: chen_guoliang222@163.com

通讯作者: 刘大洋 (1990-), 男, 博士, 副教授, 研究方向: 农产品智能化检测与技术, E-mail: ldy333ldy@163.com

surface spectral data and demonstrates the potential application of the LSTM model in blueberry SSC detection. The LSTM universal detection model constructed in this study provides a novel technical solution for the non-destructive detection of blueberry quality on the production line. The multi-surface spectral fusion method can provide a theoretical reference for the development of quality detection models for other small-sized fruits.

Key words: blueberry; hyperspectral imaging; soluble solids content; universal prediction model

蓝莓以其丰富的香气物质、适宜的糖酸比和独特风味广受喜爱，并因富含花青素、维生素 C 等抗氧化成分而备受消费者青睐^[1,2]。近年来，随着蓝莓种植面积和消费量的不断增加，市场对蓝莓品质的要求也日益提高^[3]。蓝莓的品质检测与分级在产业链中起到关键作用，不仅保障产品标准化销售，还增强消费者对蓝莓品质的认可^[4]。其中，可溶性固形物含量（Soluble Solids Content, SSC）作为衡量蓝莓品质的重要指标，直接影响果实的风味、成熟度和市场接受度^[5,6]。然而，现有的 SSC 检测方法多为破坏性测试，难以满足现代生产对高效、快速无损检测的需求。因此，开发适应生产需求的无损检测技术，不仅能提高蓝莓品质分级效率，还能对市场应用提供支持，具有重要意义。

高光谱成像（Hyperspectral Imaging, HSI）技术具有无损、高精度、多参数同时检测的特点，已被广泛应用于梨、苹果等水果的内部品质检测^[7-10]。该技术可以提供样品的空间和光谱特征信息，正确识别和应用光谱特征可以对蓝莓 SSC 进行更准确的评估。Hu 等^[11]采集 3 个品种蓝莓的光谱数据，建立单品种及多品种 SSC 预测模型，并验证多品种模型在应对品种间生物变异时的鲁棒性最优。Zheng 等^[12]采集 3 个品种跨季节蓝莓的光谱数据，建立自适应 SSC 预测模型，通过五种校正方法应对生物变异影响，验证了基于预处理的模型效果最佳。Leiva 等^[13]利用高光谱反射和透射模式获得的全光谱对蓝莓 SSC 进行了预测。结果显示，使用反射光谱预测效果更好，校正集相关系数为 0.93，剩余预测偏差为 2.6。尽管上述研究在解决品种间生物变异和光谱模式选择方面取得了一定成果，但这些工作多集中于光谱模式及校正方法的优化，而针对生产线场景中随机摆放方式对光谱数据和 SSC 预测的具体影响仍缺乏系统研究。在实际蓝莓生产线上，蓝莓的摆放方式通常是随机的，而在高光谱镜头下，不同摆放方式会导致蓝莓表面光谱信号的采集区域差异，不同的采集区域对应着蓝莓的不同面（如果柄面、花萼面或赤道面），这些差异可能引起光谱信息的变化，从而影响 SSC 预测的准确性。目前的研究表明，不同采集区域对苹果等大型水果的模型预测有显著

影响^[14,15]，而对蓝莓等小型水果的相关研究相对较少^[16-21]。蓝莓作为小型水果，其生物异质性和随机摆放特性增加了光谱信号采集和模型构建的复杂性，这对开发适应不同摆放方式的 SSC 预测通用模型提出了挑战。

本研究基于蓝莓果柄面、花萼面及两种赤道面的高光谱数据和 SSC 数据，开发了一种适用于不同摆放方式的蓝莓 SSC 预测全局通用模型，为生产线上蓝莓品质的快速无损分级提供技术支持。

1 材料与方法

1.1 实验样本

本研究所使用的样本于 2024 年 4 月采集自辽宁省丹东市的一处果园。选取 300 颗外形完好无明显缺陷的蓝莓作为实验样本，样本保存于温度为 4 °C 的冷藏室内^[22]。实验分 3 d 进行，每天使用 100 颗蓝莓进行实验。每天实验开始前 12 h，从冷藏室取出 100 颗样本，使其达到室温（23±2 °C）以满足实验条件^[23]。

1.2 高光谱数据采集与校正

本研究使用光谱范围为 935~1 720 nm 的高光谱成像系统进行高光谱数据采集。该成像系统由一台光谱分辨率为 8 nm、光谱波段数为 224 的高光谱相机（Specim FX17, Specim, Finland）、六盏 20 W 卤素灯（DECOSTAR 51 ALU, Specim, Finland）、电动扫描平台（LabScanner 40×20, Specim, Finland）、数据采集软件（Lumo Scanner, Specim, Finland）和一台计算机（OMEN by HP Laptop, Intel® Core (TM) i7-7700HQ CPU @ 2.80 GHz, RAM 16.0 GB）组成。高光谱成像系统示意图如图 1 所示。

每次实验前，高光谱成像系统需预热 30 min，以消除因随机热波动引起的暗流影响，从而确保采集到高质量的数据。为了获得高质量的高光谱图像，实验调整后确定了最佳参数设置如下：样品与相机之间的距离、曝光时间、扫描平台的速度以及扫描台上放置的蓝莓数量分别设置为 32 cm、3.8 ms、8 mm/s 和 50。此外，还需对采集到的原始图像进行黑白板校正^[24]，其校正方法如公式（1）所示。

$$R = \frac{R_i - R_d}{R_w - R_d}$$

(1)

式中：
 R ——校正后的高光谱图像；
 R_i ——原始高光谱图像；
 R_w ——标准白板图像；
 R_d ——全黑标定图像。

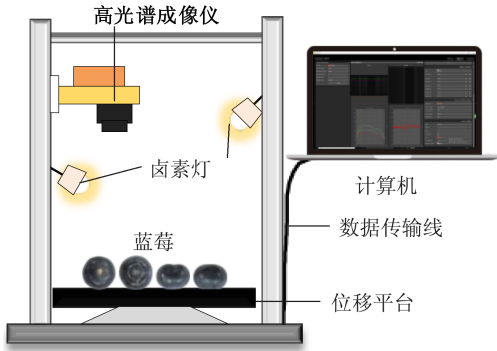


图 1 高光谱成像系统
Fig.1 Hyperspectral imaging system

1.3 高光谱数据处理

本研究使用 ENVI 软件及其工具箱，从原始高光谱数据中提取样本的平均光谱，具体过程如下：首先，对原始高光谱数据进行黑白校正，以完成初步校准（参见第 1.2 节）。随后，使用 ENVI 的“感兴趣区域（Region of Interest, ROI）”工具，通过在 1 109.1 nm 波长下的原始高光谱图像中设置阈值范围（2 025~10 000），生成确保选中视野内所有蓝莓的 ROI。在此波长下，蓝莓与背景（黑色扫描平台）的对比度最大，便于实现目标的精确分割。其次，利用 ENVI 的“构建掩膜（Build Mask）”插件生成二元掩膜，并通过“应用掩膜（Apply Mask）”插件将其应用于高光谱图像。这一步骤有效消除了背景干扰，仅保留蓝莓的高光谱数据。此过程确保每颗蓝莓在处理后的光谱图像中独立存在，便于后续单颗蓝莓平均光谱的提取。最后，在处理后的光谱图像中，通过 ENVI 的“生长（Grow）”模块标记整颗蓝莓的 ROI，以精确定义其边界和区域。对 ROI 区域内每个像素点的光谱值进行求和并取平均，计算出代表该蓝莓光谱特征的平均光谱数据。

1.4 可溶性固形物（SSC）含量测定

在样本光谱数据采集完成后，使用破坏性方法测定蓝莓 SSC。操作步骤如下：将每颗蓝莓沿赤道方向切割成均等的两半。一半蓝莓用无纺布包裹后挤压榨汁，将获得的汁液滴入全自动便携式数字折射仪（PAL-3, ATAGO Co. Ltd., Tokyo, Japan）的测量窗口进行测定。每半颗蓝莓重复测量 3 次，取平均值作为该

半颗蓝莓的 SSC 参考值^[25]。另一半蓝莓采用相同方法测定。两半蓝莓的 SSC 参考值平均后，作为该蓝莓的 SSC 参考值。每次测量后，使用蒸馏水清洁折射仪测量窗口，保证测量准确性^[26]。

1.5 建模方法

本研究基于经过黑白校正和平均光谱提取的质量较高的全波段光谱数据，无需进行光谱数据预处理，分别采用偏最小二乘法、支持向量回归以及长短期记忆网络构建通用模型。

1.5.1 PLSR

偏最小二乘回归（Partial Least Squares Regression, PLSR）是一种常用的线性回归建模方法，旨在通过将预测变量和响应变量同时投影到一个新空间，建立能够解释二者最大协方差的线性回归模型^[27]。该方法通过提取潜在因子对预测矩阵和响应矩阵的关系进行建模，特别适合处理变量众多、存在多重共线性的高维数据，同时能够有效降低数据维度并保留关键信息。在本研究中，潜在变量（Latent Variables, LVs）的个数通过五折交叉验证（5-Fold Cross-Validation）确定。操作上，该方法将数据集随机分为五个等分，每次选择其中一个分组作为验证集，其余四个分组作为训练集，循环计算验证误差，从而优化 PLSR 模型的复杂度，确保模型参数选择的稳定性和泛化能力，有效避免过拟合或欠拟合的风险。

1.5.2 SVR

支持向量回归（Support Vector Regression, SVR）是一种基于支持向量机理论的回归方法，适用于线性或非线性、高维数据的建模，通过优化损失函数平衡预测误差与模型泛化能力^[28]。该方法通过引入核函数实现非线性映射，用于在高维空间中处理复杂的数据关系，常见的核函数包括径向基核函数、线性核函数和多项式核函数等。其中，线性核函数因其计算速度快、可解释性强，适合处理线性可分数据，因此在本研究中被选为核函数。同时，为了避免过拟合并提高模型的预测性能，SVR 采用带容忍区间的损失函数控制预测误差范围，并通过五折交叉验证优化惩罚因子和核参数。SVR 能有效降低数据复杂性，同时提取关键特征信息，用于构建高效的预测模型。

1.5.3 LSTM

长短期记忆网络（Long Short-Term Memory, LSTM）是一种特殊的递归神经网络（Recurrent Neural Network, RNN），旨在解决传统 RNN 中因梯度消失或爆炸导致的长时依赖问题^[29]。LSTM 通过引入遗忘门、

输入门和输出门的门控机制，有效地控制信息的记忆与遗忘过程，从而在时间序列任务中捕获长期依赖关系，广泛应用于自然语言处理、时间序列预测等领域^[30,31]。为了将 LSTM 应用于高光谱数据的回归任务，本研究对传统 LSTM 模型进行了定制化调整，调整体现在以下 3 方面：

(1) 手动初始化隐藏状态，优化独立序列建模。传统 LSTM 通过在时间序列中传递隐藏状态来捕捉长期依赖关系，而本研究中则对其进行了调整，在每个 batch 中将隐藏状态初始化为零，使模型能够独立学习每个输入序列的特征。这一调整充分利用了光谱数据中样本相互独立的特性，避免了跨样本间的依赖干扰，从而更专注于建模每个样本内部的特征关系。该策略不仅提高了模型的稳定性，还显著降低了解决长期依赖问题的计算复杂度。

(2) Dropout 正则化与早停机制，提升模型泛化能力。在 LSTM 层中加入 Dropout 机制（概率设定为 50%），以减少模型过拟合的风险，同时引入早停机制，当验证集损失在 500 个 epoch 内未显著改善时终止训练。Dropout 通过随机屏蔽部分神经元输出，有效防止模型对训练集的过度拟合，而早停机制则确保训练在性能达到最优时适时停止，既节约计算资源，又进一步降低过拟合风险。

(3) 特殊结构优化与标准化处理，简化计算并加速收敛。本研究在模型结构和数据处理方面进行了优化与标准化，以简化计算并加速收敛。为此，仅取 LSTM 最后一个时间步的输出作为预测值，减少了参数量并使特征学习更为集中；对光谱数据和 SSC 值进行标准化处理，确保了特征尺度的一致性；采用 3 层 LSTM 结构以提升模型对复杂光谱特征的表达能力。这种结构优化避免了逐时间步输出的冗余计算，更契合回归任务的目标；标准化处理有效提升了训练的稳定性和收敛速度；多层 LSTM 逐步提取光谱数据中的复杂特征，增强了对非线性关系的捕捉能力，从而显著提高了预测精度。

1.6 样本集划分及用途

本研究对样本集进行了合理的划分，旨在验证单一面光谱数据与多面数据混合构建通用模型的可行性，并设计了相应的实验流程。建模所使用的数据来源于 300 颗蓝莓的光谱数据和 SSC 数据。每颗蓝莓的 4 个面均采集了高光谱数据，并对应一个统一的 SSC 值标签。因此，每颗蓝莓的数据包含 4 组光谱数据和 1 个 SSC 值标签。本文采用随机样本划分，将所有蓝莓样本按照 2:1 的比例分为训练与内部验证集和外部验证

集，其对应的 SSC 统计结果见表 1。训练与内部验证集的主要用途是用于构建和评估局部模型与通用模型的训练性能，而外部验证集则完全独立于模型训练，专用于评估模型的泛化能力，以验证其在未知数据上的适用性和鲁棒性。

表 1 300颗蓝莓SSC参考值

Table 1 SSC reference values of 300 blueberries

指标	样本	范围/%	平均值/%	标准差/%
	所有样本（300）	7.90~17.05	11.68	1.75
SSC	训练与内部验证集（200）	7.90~17.05	11.52	1.75
	外部验证集（100）	8.95~17.00	12.00	1.71

为实现研究目标，本实验设计了两个关键步骤。首先，在验证单一面数据建模能力的过程中，分别使用训练与内部验证集中 4 个面的数据（果柄面、花萼面、赤道面-1 和赤道面-2），并按照 4:1 的比例划分为训练子集和内部验证子集，用于构建局部模型；随后，利用外部验证集对局部模型进行评估，以分析其在不同面数据上的预测性能。在此基础上，为实现通用模型的构建，将训练与内部验证集合并，形成一个包含 800 个数据的综合数据集，综合数据集同样按照 4:1 的比例划分为训练子集和内部验证子集，用于尝试构建通用模型并评估其性能。在这一过程中，通过学习综合数据集中的跨面特征，探索是否可以建立一个能够适应任意面光谱数据的预测模型；最后，利用外部验证集对模型进行独立验证，以分析其在不同面上的适用性和预测能力。

1.7 模型评价指标

为了评估模型的预测效果与精度，采用相关系数（ r ）、均方根误差（RMSE）和剩余预测偏差（RPD）作为验证指标。相关系数 r 越接近 1，代表模型精度越高；RMSE 越小，说明模型具有更好的稳健性；RPD 则用来评估模型的预测能力。RPD 值小于 1.5 表明模型预测效果较差，介于 1.5 至 2.0 之间表明模型能够区分低高值，2.0 至 2.5 表明模型具有基本的定量预测能力，而 RPD 大于 2.5 则说明模型具备出色的预测性能^[32]。

1.8 数据分析

本研究的数据分析采用 ENVI 5.3（Exelis Visual Information Solutions, America）、MATLAB R2022a（MathWorks, America）和 Python 3.9（Python Software Foundation, America）完成。具体分析过程为：首先利用 ENVI 软件对采集到的高光谱数据进行黑白校正、感兴趣区域提取和平均光谱数据计算；其次使用

MATLAB 进行主成分分析, 验证不同面光谱数据之间的差异性; 最后采用 Python 及相关开源库构建和评估 PLSR、SVR 及 LSTM 模型的预测性能, 并根据相关系数、均方根误差和剩余预测偏差等指标评估模型的精度和泛化能力。

2 结果与讨论

2.1 光谱分析

图 2 为 200 颗建模数据集蓝莓在 935~1 720 nm 波长区域的平均光谱反射率。可以看出, 由于蓝莓颜色较深, 其各个面的光谱反射率普遍较低, 在该波长范围内的反射率均低于 0.65。从整体上看, 四条曲线在波长范围 935~1 720 nm 内呈现出相似的变化趋势, 表明蓝莓不同面的光谱分布具有较高的一致性。尽管总体趋势一致, 但蓝莓不同面间的光谱强度仍存在一定的差异。花萼面的光谱强度整体高于其他面, 尤其是在 1 150~1 300 nm 范围内, 其光谱曲线与其他三个面明显分离。这与蓝莓花萼面的生物特征密切相关: 花萼面表面粗糙且轻微凹陷 (残留萼片组织), 其表皮较厚、蜡质层和色素 (如花青素) 含量较高, 导致入射光多次散射和反射; 此外, 花萼面的水分和代谢物分布差异也可能增强其光谱响应^[33]。相比之下, 果柄面、赤道面 -1 和赤道面 -2 三个方向的光谱曲线较为接近, 但在 1 200~1 400 nm 区间仍存在微小差异, 反映出不同面细微的光谱分布差异。

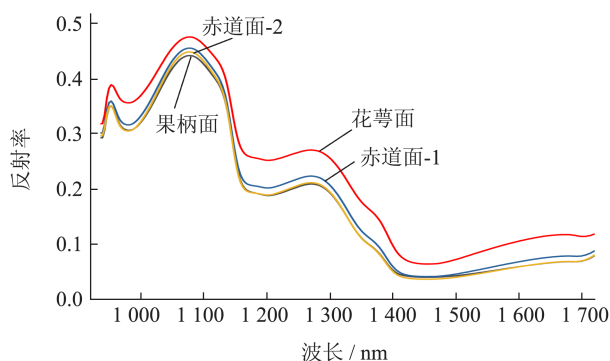


图 2 蓝莓四个面的平均光谱曲线

Fig.2 Average spectral curves of the four surfaces of blueberries

2.2 PCA 得分图

为了验证蓝莓不同面光谱数据之间的差异, 尤其是花萼面与其他面的光谱特征是否具有显著差异, 本

研究对训练与内部验证集的光谱数据进行了主成分分析 (Principal Component Analysis, PCA)。PCA 是一种数据降维方法, 能够将高维数据映射到低维空间, 通过提取主成分来最大程度地保留数据的方差信息, 从而揭示样本间的分布特征和潜在规律^[34]。

训练与内部验证集光谱数据的 PCA 得分图如图 3 所示, 其中不同的符号和颜色分别代表蓝莓的不同面, 包括果柄面 (绿色)、花萼面 (橙色)、赤道面 -1 (蓝色) 和赤道面 -2 (红色)。图中的椭圆表示不同面样本对应的 95% 置信区间, 横轴 (PC1) 和纵轴 (PC2) 分别代表第一和第二主成分。这两个主成分共解释了 99.2% 的总方差, 其中 PC1 解释了 91.0% 的方差, PC2 解释了 8.2% 的方差。

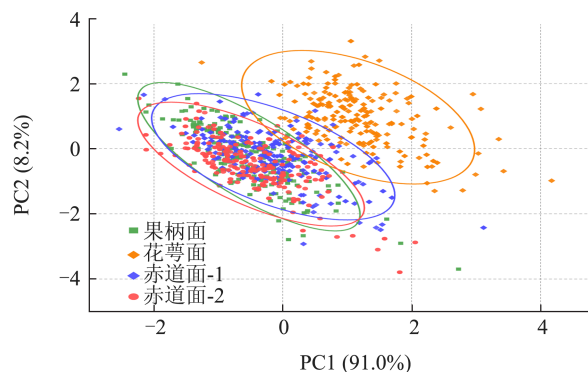


图 3 蓝莓不同面光谱数据的 PCA 得分图

Fig.3 PCA score plot of spectral data from different surfaces of blueberries

从 PCA 得分图可以看出, 蓝莓不同面样品在第一主成分与第二主成分空间中表现出明显的聚类趋势。在图中, 赤道-1 面和赤道-2 面的样品分布较为接近, 其 95% 置信椭圆高度重叠, 表明这两个面的光谱特征高度相似。与此同时, 果柄面 (绿色) 与赤道面 -2 (红色) 样品的置信椭圆也存在较多重叠, 说明果柄面与赤道面之间的光谱特征较为接近。然而, 花萼面 (橙色) 的置信椭圆相对独立, 分布位置明显偏离其他三个面, 形成了较为清晰的分离趋势。该结果与已有研究中关于苹果不同位置 (如果柄、赤道) 光谱差异的发现一致, 表明果实不同部位的表皮结构和化学成分差异会显著影响其光谱特征分布^[35]。类似地, 这也表明花萼面在光谱反射特征上具有显著的独特性, 与其他面存在明显差异。这一结果与 2.1 节光谱曲线分析的结论一致。

2.3 局部模型建模分析

本小节旨在通过蓝莓单一面数据构建局部模型,评估其预测所有面 SSC 值的效果。表 2 和表 3 分别为基于 PLSR 模型和 SVR 模型,利用训练与内部验证集的单面数据所构建的 4 个局部模型在外部验证集上的表现。结果显示,在多数情形下,基于单一面数据构建的模型在预测其他面的 SSC 时表现出较差的精度,具体表现为 RMSEP 值较高,而 r_p 和 RPD 值较低。特别是,使用花萼面数据构建的 PLSR 和 SVR 模型在外部验证集上预测当前面 SSC 时表现良好,但在预测其他面 SSC 时,RPD 值均未超过 1.8,且 RMSEP 值均高于 0.90。这可能是由于蓝莓花萼面具有与其他面不同的光谱特征,导致花萼面局部模型在建模过程中过度捕捉这些特定特征,限制了模型的泛化能力。这表明,基于单一面数据建立的局部模型无法准确预测其他面的 SSC,说明仅依赖单个面数据构建的模型在跨面预测时存在显著局限性。因此,为了提高对蓝莓所有面(包括果柄面、花萼面、赤道面-1 和赤道面-2)SSC 的预测能力,分别使用 PLSR、SVR 和 LSTM 模型,在包含 4 个面数据的综合数据集上进行建模,尝试构建通用模型,并筛选出性能最优的通用模型。

2.4 通用模型建模分析

表 4 为基于综合数据集构建的模型在外部验证集上的表现,所使用的模型分别为 PLSR、SVR 和 LSTM。可以看出,3 个模型对 4 个面的 SSC 均有较好的预测能力,预测所有面的 RPD 值都在 2.5 以上,且 RMSEP 值均低于 0.65。与局部模型相比,使用综合数据集建模的模型在预测蓝莓其他面 SSC 时表现出更好的效果。基于 PLSR、SVR 和 LSTM 构建的模型在果柄、花萼、赤道-1 和赤道-2 这四个面上预测的 RPD 分别为:PLSR 模型为 2.79、3.20、3.16 和 3.21;SVR 模型为 2.67、3.25、3.04 和 3.22;LSTM 模型为 3.32、3.27、3.61 和 3.52。这表明 PLSR 和 SVR 模型的效果相近,而 LSTM 模型表现最佳,构建的通用模型性能最优。在外部验证集上,LSTM 模型在四个面的 r_p 分别为 0.963、0.954、0.964 和 0.962;RMSEP 分别为 0.513%、0.521%、0.472% 和 0.483%;RPD 分别为 3.32、3.27、3.61 和 3.52。使用 LSTM 模型构建的通用模型在外部验证集上的预测结果散点图如图 4 所示。从图中可以看出,通用模型在四个面上均表现出优异的预测性能,拟合线与散点分布紧密,表明模型具有较高的预测可靠性。

表 2 局部模型 (PLSR) 在外部验证集上的回归结果

Table 2 Regression results of the local model (PLSR) on the external validation set												
局部模型 (PLSR)	果柄面			花萼面			赤道面-1			赤道面-2		
	r_p	RMSEP	RPD	r_p	RMSEP	RPD	r_p	RMSEP	RPD	r_p	RMSEP	RPD
果柄面模型	0.946	0.652	2.61	0.885	0.983	1.73	0.928	0.870	1.95	0.947	0.655	2.59
花萼面模型	0.857	0.998	1.70	0.970	0.431	3.94	0.921	1.296	1.31	0.888	1.509	1.12
赤道面-1 模型	0.947	0.642	2.65	0.932	0.870	1.95	0.961	0.530	3.21	0.958	0.649	2.62
赤道面-2 模型	0.941	0.664	2.56	0.925	1.806	0.94	0.947	0.550	3.09	0.959	0.519	3.28

表 3 局部模型 (SVR) 在外部验证集上的回归结果

Table 3 Regression results of the local model (SVR) on the external validation set												
局部模型 (SVR)	果柄面			花萼面			赤道面-1			赤道面-2		
	r_p	RMSEP	RPD	r_p	RMSEP	RPD	r_p	RMSEP	RPD	r_p	RMSEP	RPD
果柄面模型	0.944	0.682	2.49	0.930	0.871	1.95	0.939	1.009	1.68	0.957	0.732	2.32
花萼面模型	0.879	0.946	1.80	0.975	0.420	4.04	0.936	1.117	1.52	0.901	1.345	1.26
赤道面-1 模型	0.957	0.569	2.99	0.946	0.872	1.95	0.961	0.533	3.19	0.960	0.603	2.82
赤道面-2 模型	0.950	0.529	3.21	0.916	0.859	1.98	0.952	0.529	3.22	0.969	0.431	3.94

表 4 通用模型（PLSR、SVR和LSTM）在外部验证集上的回归结果

Table 4 Regression results of the universal model (PLSR, SVR, and LSTM) on the external validation set

模型名称	果柄面			花萼面			赤道面-1			赤道面-2		
	r_p	RMSEP	RPD	r_p	RMSEP	RPD	r_p	RMSEP	RPD	r_p	RMSEP	RPD
PLSR 模型	0.959	0.608	2.79	0.954	0.531	3.20	0.965	0.539	3.16	0.963	0.530	3.21
SVR 模型	0.962	0.637	2.67	0.956	0.524	3.25	0.970	0.560	3.04	0.972	0.527	3.22
LSTM 模型	0.963	0.513	3.32	0.954	0.521	3.27	0.964	0.472	3.61	0.962	0.483	3.52

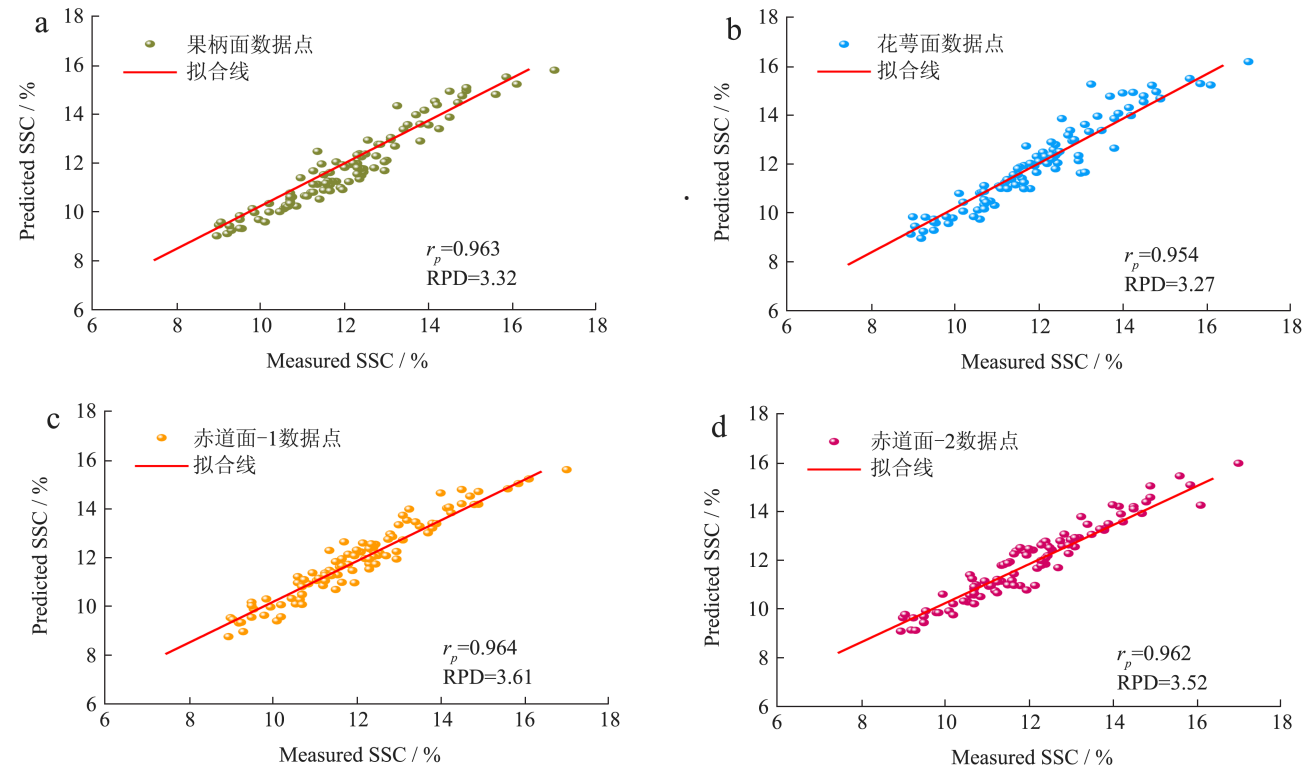


图 4 通用模型（LSTM）在外部验证集上对蓝莓不同面预测值与实测值的回归分析图

Fig.4 Regression analysis plot of predicted versus measured values of the universal model (LSTM) on the external validation set for the different surfaces of blueburry

注：(a) 果柄；(b) 花萼；(c) 赤道-1；(d) 赤道-2。

3 结论

本研究基于高光谱成像技术，结合蓝莓的 SSC 数据，对蓝莓不同面（果柄面、花萼面、赤道面-1 和赤道面-2）的光谱特征进行了分析，并采用 PLSR、SVR 和 LSTM 三种方法构建局部模型和通用模型。结果表明，不同面光谱反射特征在 935~1 720 nm 范围内整体一致，但花萼面在 1 150~1 300 nm 区域表现出更高的光谱强度，PCA 分析进一步证实了花萼面的光谱分布与其他面存在显著差异，显示其在跨面数据分析中的独特性。基于单一面数据构建的局部模型在外部验证集上的泛化性能较差，难以实现 SSC 的跨面精准预测；而基于综合数据集构建的通用模型表现出显著优势，通用模型的 RPD 值均超过 2.5，显示了较强的预测能

力，其中 LSTM 模型预测性能最佳，平均 RPD 达到 3.43，RMSEP 最低，具备优异的准确性与泛化能力。本研究验证了高光谱成像技术结合多面光谱数据在蓝莓 SSC 预测中的可行性，为后续跨面光谱分析和模型构建提供了重要参考。

参考文献

[1] YANG J N, YIN J T, WANG K, et al. Advanced technology in fruit preservation: Effects of nanoscale charged water particles on storage quality and reactive oxygen species in blueberries [J]. Food Research International, 2024, 198: 115331.

[2] G M T, R R S, SHRUTI S, et al. Genotypic variability in nutritional and functional attributes of blueberry varieties grown in northern-western Himalayas [J]. Journal of Food Science and Technology, 2020, 57(6): 2251-2258.

- [3] QIAO S C, TIAN Y W, GU W J, et al. Research on simultaneous detection of SSC and FI of blueberry based on hyperspectral imaging combined MS-SPA [J]. *Engineering in Agriculture, Environment and Food*, 2019, 12(4): 540-547.
- [4] ZHANG M Y, JIANG Y, LI C Y, et al. Fully convolutional networks for blueberry bruising and calyx segmentation using hyperspectral transmittance imaging [J]. *Biosystems Engineering*, 2020, 192(C): 159-175.
- [5] BAI Y H, FANG Y L, ZHANG B H, et al. Model robustness in estimation of blueberry SSC using NIRS [J]. *Computers and Electronics in Agriculture*, 2022, 198: 107073.
- [6] 刘小路,薛璐,鲁晓翔,等.蓝莓可溶性固形物、总酸近红外无损检测模型的建立[J].*食品与生物技术学报*,2016,35(7):752-756.
- [7] LI Y T, YOU S C, WU S S, et al. Exploring the limit of detection on early implicit bruised ‘Korla’ fragrant pears using hyperspectral imaging features and spectral variables [J]. *Postharvest Biology and Technology*, 2024, 208: 112668.
- [8] ZHANG D Y, XU Y F, HUANG W Q, et al. Nondestructive measurement of soluble solids content in apple using near infrared hyperspectral imaging coupled with wavelength selection algorithm [J]. *Infrared Physics and Technology*, 2019, 98: 297-304.
- [9] XUAN G T, GAO C, SHAO Y Y. Spectral and image analysis of hyperspectral data for internal and external quality assessment of peach fruit [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2022, 272: 121016.
- [10] PAN T, MENG Q H, WU Z F, et al. Detection of mango soluble solid content using hyperspectral imaging technology [J]. *Infrared Physics and Technology*, 2023, 129: 104576.
- [11] HU M H, DONG Q L, LIU B L. Modelling postharvest quality of blueberry affected by biological variability using image and spectral data [J]. *Journal of the Science of Food and Agriculture*, 2016, 96(10): 3365-3373.
- [12] ZHENG W, BAI Y H, LUO H, et al. Self-adaptive models for predicting soluble solid content of blueberries with biological variability by using near-infrared spectroscopy and chemometrics [J]. *Postharvest Biology and Technology*, 2020, 169: 111286.
- [13] LEIVA-V A G, LU R F, AGUILERA M J. Assessment of internal quality of blueberries using hyperspectral transmittance and reflectance images with whole spectra or selected wavelengths [J]. *Innovative Food Science and Emerging Technologies*, 2014, 24: 2-13.
- [14] FAN S, ZHANG B, LI J, et al. Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple [J]. *Biosystems Engineering*, 2016, 143: 9-19.
- [15] 田有文,程怡,王小奇,等.基于高光谱成像的苹果虫伤缺陷与果梗/花萼识别方法[J].*农业工程学报*,2015,31(4):325-331.
- [16] ZHENG Y J, CAO Y C, YANG J, et al. Enhancing model robustness through different optimization methods and 1-D CNN to eliminate the variations in size and detection position for apple SSC determination [J]. *Postharvest Biology and Technology*, 2023, 205: 112513.
- [17] WANG Z J, WU S S, ZUO C Z, et al. Exploring the variability and heterogeneity of apple firmness using visible and near-infrared hyperspectral imaging [J]. *LWT*, 2024, 192: 115704.
- [18] BOBELYN E, SERBAN A-S, NICU M, et al. Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance [J]. *Postharvest Biology and Technology*, 2009, 55(3): 133-143.
- [19] HUANG Y P, XIONG J, JIANG X S, et al. Assessment of firmness and soluble solids content of peaches by spatially resolved spectroscopy with a spectral difference technique [J]. *Computers and Electronics in Agriculture*, 2022, 200: 107212.
- [20] YU Y, YAO M B. Is this pear sweeter than this apple? A universal SSC model for fruits with similar physicochemical properties [J]. *Biosystems Engineering*, 2023, 226: 116-131.
- [21] CHANGYEUN M, MOON S K, GIYOUNG K, et al. Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging [J]. *Biosystems Engineering*, 2017, 159: 10-21.
- [22] 刘鸿阳,孔德国,罗华平,等.基于多光谱图像角度融合测定库尔勒香梨理化指标[J].*光谱学与光谱分析*,2024,44(3):649-655.
- [23] WU D, WAN G L, JING Y X, et al. Hyperspectral imaging combined with deep learning for discrimination of Lingwu long jujube in terms of the time after bruising [J]. *Microchemical Journal*, 2023, 194: 109238.
- [24] LI S Y, SONG Q M, LIU Y J, et al. Hyperspectral imaging-based detection of soluble solids content of loquat from a small sample [J]. *Postharvest Biology and Technology*, 2023, 204: 112454.
- [25] LEIVA-V A G, LU R F, AGUILERA M J. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging [J]. *Journal of Food Engineering*, 2013, 115(1): 91-98.
- [26] 黄华,刘亚,马毅航,等.基于可见/近红外光谱和函数型线性回归模型的成熟期苹果可溶性固形物含量预测[J].*光谱学与光谱分析*,2024,44(7):1905-1912.
- [27] FU P, KATHERINE M, GUAN K Y, et al. Estimating photosynthetic traits from reflectance spectra: A synthesis of spectral indices, numerical inversion, and partial least square regression [J]. *Plant, Cell & Environment*, 2020, 43(5): 1241-1258.
- [28] ZHANG P P, SHEN B B, JI H W, et al. Nondestructive Prediction of Mechanical Parameters to Apple Using Hyperspectral Imaging by Support Vector Machine [J]. *Food Analytical Methods*, 2022, 15(5): 1397-1406.
- [29] XU Y, YANG G, LI B, et al. Flow inferential measurement for proportional control valves by combining wavelet denoising and a dual-attention-based LSTM network [J]. *Flow Measurement and Instrumentation*, 2024, 100: 102713.
- [30] LI M H, ZHOU Q Y, HAN X, et al. Prediction of reference crop evapotranspiration based on improved convolutional neural network (CNN) and long short-term memory network (LSTM) models in Northeast China [J]. *Journal of Hydrology*, 2024, 645(PA): 132223.
- [31] YU S, FAN J C, LU X J, et al. Deep learning models based on hyperspectral data and time-series phenotypes for predicting quality attributes in lettuces under water stress [J]. *Computers and Electronics in Agriculture*, 2023, 211: 108034.
- [32] SUN X D, DU Y J, NAWAZ A M, et al. Apple SSC estimation using hand-held NIRS instrument for outdoor measurement with ambient light correction [J]. *Postharvest Biology and Technology*, 2024, 217: 113101.
- [33] 李瑞,傅隆生.基于高光谱图像的蓝莓糖度和硬度无损测量[J].*农业工程学报*,2017,33(S1):362-366.
- [34] ABDI H, WILLIAMS J L. Principal component analysis [J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, 2(4): 433-459.
- [35] XU X, XU H, XIE L, et al. Effect of measurement position on prediction of apple soluble solids content (SSC) by an on-line near-infrared (NIR) system [J]. *Journal of Food Measurement and Characterization*, 2019, 13: 506-512.