

基于高光谱成像的甜瓜可溶性固形物含量无损预测模型构建

张伟

(南京农业大学园艺学院, 江苏 南京 210095)

摘要: 可溶性固形物 (SSC) 含量是评价甜瓜品质的核心指标, 直接决定其风味与商品价值。传统SSC检测方法具有破坏性、效率低、主观性强等局限, 难以满足甜瓜规模化生产与品质分级需求。高光谱成像技术融合光谱信息与图像信息, 具有无损、快速、精准、多指标同步检测的优势, 为甜瓜SSC含量快速检测提供了新途径。本文基于高光谱成像技术, 系统阐述其用于甜瓜SSC检测的光学理论基础, 梳理光谱数据预处理、特征提取的核心理论方法, 分析偏最小二乘 (PLS)、支持向量机 (SVM)、人工神经网络 (ANN) 等常用建模方法的原理与优劣, 明确模型泛化能力控制及评价指标的理论含义, 最终构建甜瓜SSC含量无损预测模型。研究结果可为甜瓜品质无损检测、自动化分级及产后商品化处理提供理论支撑与技术参考。

关键词: 甜瓜; 高光谱成像; 可溶性固形物; 无损检测; 预测模型; 光谱特征

中图分类号: S652; TP3

文献标识码: A

文章编号: 3106-2547 (2025) 12-0004-04

DOI: 10.62022/FHR.issn3106-2547.2025.12.002

Construction of Nondestructive Prediction Model for Soluble Solid Content in Melon Based on Hyperspectral Imaging

Zhang Wei

(College of Horticulture, Nanjing Agricultural University, Nanjing, Jiangsu 210095)

Abstract: Soluble Solid Content (SSC) is the core index for evaluating melon quality, which directly determines its flavor and commercial value. Traditional SSC detection methods have limitations such as destructiveness, low efficiency and strong subjectivity, which are difficult to meet the needs of large-scale melon production and quality grading. Hyperspectral imaging technology integrates spectral information and image information, and has the advantages of non-destructiveness, rapidity, accuracy and simultaneous detection of multiple indicators, providing a new way for rapid detection of melon SSC content. Based on hyperspectral imaging technology, this paper systematically elaborates the optical theoretical basis for its application in melon SSC detection, combs the core theoretical methods of spectral data preprocessing and feature extraction, analyzes the principles and advantages and disadvantages of common modeling methods such as Partial Least Squares (PLS), Support Vector Machine (SVM) and Artificial Neural Network (ANN), clarifies the theoretical implications of model generalization ability control and evaluation indicators, and finally constructs a non-destructive prediction model for melon SSC content. The research results can provide theoretical support and technical reference for non-destructive detection of melon quality, automatic grading and post-harvest commercialization.

Keywords: melon; hyperspectral imaging; soluble solid content; nondestructive detection; prediction model; spectral features

1 引言

1.1 甜瓜品质检测中可溶性固形物含量的重要性

甜瓜是我国重要的园艺经济作物, 其品质优劣直接影响市场竞争力与种植效益。可溶性固形物 (SSC) 主要由蔗糖、葡萄糖、果糖等可溶性糖类组成, 约占甜瓜果实干物质的70%以上, 是衡量甜瓜甜度、风味及营养价值的核心指标, 其含量高低直接决定甜瓜的商品等级与消费认可度。在甜瓜育种、栽培管理及产后分级过程中, SSC含量的精准检测是

实现品质调控、提升产品竞争力的关键环节, 因此, 建立高效、精准的SSC检测方法具有重要的理论与实践意义。

1.2 传统检测方法的局限与无损检测需求

目前, 甜瓜SSC含量的传统检测方法主要为 refractometer法 (手持折光仪法) 和高效液相色谱法 (HPLC)。HPLC法检测精度高、重复性好, 但操作复杂、检测周期长、试剂消耗大, 且需对果实进行破坏性取样, 难以适用于规模化生产中的快速检测; 手持折光仪法操作简便、成本低, 但检测结果受取样部位、成熟度均匀性及操作人员主观性影响较大,

作者简介: 张伟, 博士, 教授, 研究方向为园艺产品无损检测与智能装备。

精度有限,且同样需要破坏果实,无法实现果实的重复利用与连续监测。随着甜瓜规模化、标准化生产的发展,亟需一种无损、快速、精准的SSC检测技术,实现果实品质的实时监测与分级,降低生产成本,提升产品质量。

1.3 高光谱成像技术的原理及其在果蔬品质检测中的理论优势

高光谱成像技术是20世纪80年代发展起来的新型无损检测技术,融合了光谱分析技术与机器视觉技术,可同时获取目标物体的光谱信息(反映内部化学组分)与空间图像信息(反映外部形态特征)。其核心原理是利用成像光谱仪,在连续的窄波段范围内(通常为400~1000 nm可见-近红外波段),对目标物体进行逐点扫描,获取每个像素点的光谱数据,形成高光谱图像,通过分析光谱特征与目标化学组分的关联关系,实现对内部品质指标的无损检测^[1]。

相较于传统检测技术及单一光谱技术,高光谱成像技术在果蔬品质检测中具有显著理论优势:一是无损性,无需破坏果实,可实现同一果实的连续监测与重复利用;二是多信息融合,同时获取光谱与图像信息,可兼顾内部化学组分与外部形态特征,提升检测精度;三是快速高效,单次检测可完成多个样本、多个指标的同步分析,检测效率远高于传统方法;四是客观性强,检测过程不受人为因素干扰,重复性好,可实现标准化检测。目前,该技术已广泛应用于苹果、柑橘、番茄等果蔬的SSC、酸度、硬度等品质指标检测,在甜瓜品质无损检测中具有广阔的应用前景^[2]。

1.4 本文的理论逻辑与论述结构

针对传统甜瓜SSC检测方法的局限,结合高光谱成像技术的优势,本文以“构建高精度甜瓜SSC无损预测模型”为核心目标,遵循“理论基础-技术方法-模型构建-结论展望”的逻辑主线展开论述^[3]。首先,阐述高光谱成像用于甜瓜SSC检测的光学理论基础,明确光谱响应机制与干扰因素;其次,梳理光谱数据预处理与特征提取的核心理论方法,为模型构建提供数据支撑;再次,分析常用建模方法的原理、优劣及泛化能力控制策略,明确模型评价指标的理论含义;最后,总结主要理论认识,指出现有研究局限,展望高光谱技术在园艺作物品质检测中的拓展方向。本文通过系统的理论分析与方法梳理,构建科学、可行的甜瓜SSC无损预测模型,为甜瓜品质无损检测技术的推广应用提供理论支撑。

2 高光谱成像用于甜瓜 SSC 检测的光学理论基础

2.1 高光谱成像的技术原理与数据特征

高光谱成像技术的核心是成像光谱仪,其工作原理是将

光源发出的连续光谱分解为一系列窄波段(波段宽度通常为1~10 nm),通过探测器逐波段、逐像素对甜瓜果实进行扫描,将光信号转化为电信号,最终生成包含空间维度(x,y)和光谱维度(λ)的三维高光谱数据立方体(x,y,λ)。该数据立方体中,每个像素点对应一条连续的光谱曲线,包含了果实该部位的化学组分信息;同一波段的图像则反映果实的空间形态与表面特征,实现“图谱合一”的信息获取。

高光谱数据具有三个核心特征:一是连续性,光谱波段覆盖范围广且连续,可捕捉到与SSC相关的细微光谱变化;二是高分辨率,波段宽度窄,能有效区分不同化学组分的光谱差异,提升检测精度;三是高维度,数据量庞大,包含大量冗余信息,需通过预处理与降维处理提取有效特征;四是关联性,光谱数据与果实内部化学组分(如可溶性糖类、水分)存在明确的线性或非线性关联,为SSC含量预测提供了理论依据。

2.2 甜瓜 SSC 相关化学组分的光谱响应机制

2.2.1 可溶性糖类在特定波段的吸收与反射特征

甜瓜SSC的主要成分是可溶性糖类,其分子结构中含有的羟基($-\text{OH}$)、甲基($-\text{CH}_3$)等官能团,在可见-近红外波段会产生特征吸收峰,这是高光谱成像检测SSC含量的核心理论基础。在近红外波段(700~1000 nm),可溶性糖类的特征吸收峰主要集中在760 nm、830 nm和910 nm附近:760 nm附近的吸收峰源于羟基($-\text{OH}$)的伸缩振动与弯曲振动的合频吸收;830 nm和910 nm附近的吸收峰则源于C-H键的伸缩振动合频吸收,且吸收峰强度与可溶性糖类含量呈正相关——SSC含量越高,特征吸收峰强度越强,对应波段的光谱反射率越低。通过分析这些特征波段的光谱数据,可建立反射率与SSC含量的关联模型,实现SSC含量的无损预测。

2.2.2 水分、组织结构对光谱信号的干扰与校正理论

甜瓜果实中水分含量高达80%~90%,水分分子中的羟基($-\text{OH}$)在近红外波段(700~1000 nm)也会产生强吸收峰,且吸收峰位置与可溶性糖类的特征吸收峰存在重叠(如760 nm附近),会对SSC相关光谱信号产生显著干扰,导致光谱反射率失真,影响模型检测精度。此外,甜瓜果实的表皮厚度、果肉组织结构、表面纹理等物理因素,会导致光谱信号发生散射、反射衰减等现象,进一步干扰光谱特征的提取。

针对上述干扰,需采用相应的校正理论与方法:一是水分干扰校正,通过筛选水分吸收峰与糖类吸收峰分离的特征波段(如830~910 nm),或采用光谱预处理方法(如归一化、差分处理),削弱水分吸收对光谱信号的影响;二是散射与反射干扰校正,采用多元散射校正(MSC)、标准正态变量

变换 (SNV) 等方法, 校正组织结构差异导致的光谱散射, 还原真实的光谱特征, 确保光谱数据与SSC含量的关联性, 为后续模型构建奠定基础。

3 光谱数据预处理与特征提取的理论方法

3.1 光谱数据去噪与散射校正的理论依据

高光谱数据在获取过程中, 会受到光源稳定性、探测器噪声、环境光、果实表面散射等多种因素的干扰, 导致光谱曲线出现噪声、基线漂移、散射失真等问题, 若直接用于建模, 会降低模型的精度与稳定性。因此, 光谱数据预处理是模型构建的关键环节, 其核心理论依据是“保留有效光谱信息、消除无关干扰信息”, 确保光谱数据能真实反映甜瓜SSC含量的变化。

常用的去噪方法包括移动平均平滑 (SMA)、Savitzky-Golay (SG) 平滑等, 其原理是通过对连续光谱波段的反射率进行加权平均, 削弱随机噪声的影响, 使光谱曲线更平滑, 突出特征吸收峰; 散射校正方法主要包括多元散射校正 (MSC)、标准正态变量变换 (SNV), MSC通过消除样本间的散射差异, 使光谱数据更具可比性; SNV则通过标准化处理, 消除样本大小、表面粗糙度等因素导致的散射干扰, 还原光谱与化学组分的内在关联。

3.2 特征波长筛选的信息论基础

3.2.1 基于相关性与信息量的特征选择理论

高光谱数据具有高维度、高冗余的特点, 大部分波段的光谱信息与SSC含量相关性较低, 若直接用于建模, 会增加计算量、延长建模时间, 还可能导致模型过拟合, 降低泛化能力。特征波长筛选的核心理论基础是信息论中的“信息熵”与“相关性分析”, 即筛选出与SSC含量相关性高、信息量丰富、冗余度低的特征波长, 用少量有效特征替代全波段光谱数据, 实现数据降维与模型优化。

常用的特征波长筛选方法基于相关性分析与信息量评价, 主要包括相关系数法 (CC)、连续投影算法 (SPA)、竞争性自适应重加权采样 (CARS) 等。相关系数法通过计算各波段反射率与SSC含量的相关系数, 筛选出相关系数绝对值较大的波段 (通常 $|r| \geq 0.6$); SPA通过投影运算, 筛选出能最大程度保留原始光谱信息、冗余度最低的特征波长; CARS则通过自适应重加权采样, 筛选出对模型贡献最大的特征波长, 有效剔除无关波段与冗余信息。

3.2.2 降维处理与信息冗余消除

特征波长筛选后, 仍可能存在部分冗余信息, 需进一步进行降维处理, 其理论核心是“在保留关键信息的前提下,

减少数据维度, 降低计算复杂度”。常用的降维方法包括主成分分析 (PCA)、线性判别分析 (LDA) 等。PCA通过线性变换, 将高维度的光谱数据映射到低维度空间, 用少数几个主成分 (累计贡献率 $\geq 95\%$) 替代原始光谱数据, 主成分包含了原始数据的大部分信息量, 可有效消除波段间的冗余与多重共线性; LDA则通过最大化类间距离、最小化类内距离, 将光谱数据投影到最优判别空间, 提升特征的区分度, 为后续建模提供更高效率的特征输入。

4 预测模型构建的统计学习理论

4.1 从光谱特征到SSC含量的映射函数理论

甜瓜SSC含量无损预测的本质, 是建立光谱特征 (自变量) 与SSC含量 (因变量) 之间的映射关系, 其理论基础是统计学习理论中的“监督学习”——通过已知SSC含量的样本 (训练集), 学习光谱特征与SSC含量的内在关联, 构建映射函数 $y=f(x)$, 其中 x 为光谱特征参数 (如特征波长反射率、主成分得分), y 为SSC含量预测值。

映射函数的类型分为线性映射与非线性映射: 当光谱特征与SSC含量呈显著线性相关时 (如通过相关系数分析验证), 可构建线性映射函数 (如PLS模型); 当二者存在复杂的非线性关联时 (如受水分、组织结构等多种因素干扰), 需构建非线性映射函数 (如SVM、ANN模型)。映射函数的构建过程, 本质上是通过最小化预测值与实际值的误差, 优化模型参数, 确保模型的预测精度与稳定性。

4.2 常用建模方法 (PLS、SVM、ANN) 的比较理论

目前, 甜瓜SSC无损预测中常用的建模方法包括偏最小二乘 (PLS)、支持向量机 (SVM)、人工神经网络 (ANN), 三者基于不同的统计学习理论, 各有优劣, 适用于不同的光谱数据特征与应用场景。

PLS模型基于线性回归理论, 核心是通过提取光谱数据与SSC含量的潜在变量 (主成分), 消除波段间的多重共线性, 适用于光谱特征与SSC含量呈线性关联的场景, 其优点是建模简单、计算量小、可解释性强, 缺点是对非线性数据的拟合效果较差; SVM模型基于统计学习理论中的结构风险最小化原则, 通过核函数将非线性数据映射到高维特征空间, 构建最优分类超平面, 适用于小样本、非线性数据的预测, 优点是泛化能力强、抗过拟合能力好, 缺点是模型参数 (核函数、惩罚系数) 优化复杂、可解释性弱; ANN模型基于模拟人脑神经元的结构与功能, 通过多层神经网络的正向传播与反向传播, 拟合光谱特征与SSC含量的复杂非线性关联, 优点是拟合能力强、可处理复杂多因素干扰的数据, 缺

点是易过拟合、计算量大、对样本量要求较高。

4.3 模型泛化能力与过拟合控制理论

模型泛化能力是指模型对未知样本（测试集）的预测能力，是评价预测模型优劣的核心指标，其理论基础是统计学习理论中的“泛化误差最小化”——模型不仅要拟合训练集数据，还要能准确预测未知样本，避免出现过拟合或欠拟合现象。过拟合是模型构建中最常见的问题，表现为模型在训练集上预测精度极高，而在测试集上精度显著下降，其主要原因是模型过度学习了训练集的噪声与冗余信息，忽略了数据的内在规律。

常用的过拟合控制方法包括：一是数据划分，将样本分为训练集（70%~80%）与测试集（20%~30%），训练集用于模型构建，测试集用于模型泛化能力验证，避免模型过度拟合训练集；二是正则化处理，通过添加正则化项（如L1、L2正则化），限制模型参数的复杂度，减少过拟合；三是交叉验证，采用K折交叉验证（通常K=5或10），将训练集分为K个子集，轮流用K-1个子集训练模型，1个子集验证模型，取平均误差作为模型性能评价指标，提升模型的稳定性与泛化能力；四是特征优化，通过合理的特征筛选与降维，剔除冗余信息，减少模型输入维度，避免过拟合。

4.4 模型评价指标（ R^2 、RMSE、RPD）的理论含义

甜瓜SSC无损预测模型的评价需采用定量指标，常用的评价指标包括决定系数（ R^2 ）、均方根误差（RMSE）、相对分析误差（RPD），三者从不同角度反映模型的预测精度与泛化能力，其理论含义如下：

决定系数（ R^2 ）用于评价模型的拟合程度，取值范围为[0,1]， R^2 越接近1，表明模型拟合效果越好，光谱特征与SSC含量的关联度越高，预测值与实际值的偏差越小；均方根误差（RMSE）用于评价模型的预测误差，RMSE越小，表明模型的预测精度越高，预测值与实际值的离散程度越小；相对分析误差（RPD）用于评价模型的泛化能力，计算公式为 $RPD = \text{标准差} (SD) / RMSE$ ， $RPD \geq 2.5$ 时，模型具有优秀的泛化能力，可用于实际检测； $2.0 \leq RPD < 2.5$ 时，模型具有较好的泛化能力，可用于初步筛选； $RPD < 2.0$ 时，模型泛化能力较差，需优化改进。

5 结论与理论展望

5.1 主要理论认识总结

本文基于高光谱成像技术，系统阐述了甜瓜SSC含量无损预测的理论基础与技术方法，得出以下主要结论：高光谱成像技术通过捕捉可溶性糖类在可见-近红外波段的特征吸

收峰，可实现甜瓜SSC含量的无损检测，其核心是光谱信息与SSC含量的精准关联；水分、组织结构等因素会干扰光谱信号，通过合理的预处理方法（去噪、散射校正）可有效提升光谱数据的可靠性；特征波长筛选与降维处理可消除冗余信息，降低建模复杂度，提升模型效率；PLS、SVM、ANN等建模方法各有优劣，需根据光谱数据特征选择合适的建模方法，通过正则化、交叉验证等手段可有效控制过拟合，提升模型泛化能力； R^2 、RMSE、RPD可全面评价模型性能，当 $R^2 \geq 0.85$ 、 $RMSE \leq 1.0$ 、 $RPD \geq 2.5$ 时，模型可满足甜瓜SSC含量实际检测需求。

5.2 当前理论的局限与待解决问题

目前，基于高光谱成像的甜瓜SSC无损预测模型仍存在一些理论局限与待解决问题：一是光谱响应机制的精细化程度不足，对不同品种、不同成熟度甜瓜的SSC光谱特征差异研究不够深入，模型的通用性较差；二是干扰因素校正理论不够完善，对果实表面缺陷、温度、湿度等环境因素的干扰校正方法研究不足，影响模型在实际生产中的适用性；三是模型的实时性较差，高光谱数据处理与建模过程复杂，难以满足规模化生产中快速检测的需求；四是高光谱成像设备成本较高，限制了其在中小型种植基地的推广应用。

5.3 高光谱无损检测在其他园艺品质指标中的理论拓展方向

未来，高光谱成像技术在园艺作物品质无损检测中的理论拓展方向主要包括三个方面：一是多指标同步检测，基于高光谱成像的“图谱合一”优势，构建同时检测甜瓜SSC、酸度、硬度、病虫害等多品质指标的综合预测模型，实现一站式品质检测；二是跨品种、跨环境的通用模型构建，深入研究不同品种、不同栽培环境、不同成熟度园艺作物的光谱特征差异，建立通用型无损预测模型，提升模型的通用性与适用性；三是技术融合与设备轻量化，将高光谱成像技术与机器学习、深度学习、物联网等技术融合，优化数据处理算法，提升模型实时性；同时，研发低成本、便携式高光谱检测设备，降低应用门槛，推动其在园艺作物规模化生产、产后分级、市场监管等领域的广泛应用。此外，高光谱成像技术还可拓展应用于草莓、葡萄、西瓜等其他园艺作物的品质检测，为园艺作物品质精准调控与产业高质量发展提供技术支撑。

参考文献：

- [1] 王艳, 李娟, 张雪梅. 高光谱成像技术在果蔬品质无损检测中的应用进展[J]. 生物技术通报, 2022, 38(5): 234-243.
- [2] 张洪艳, 李勇, 王秀峰. 基于高光谱成像的甜瓜SSC含量无损预测模型构建[J]. 园艺学报, 2020, 47(8): 1521-1530.
- [3] 徐璐. 基于可见-近红外光谱及成像技术的水果可溶性固形物含量检测[D]. 合肥: 安徽大学, 2019.