

基于高光谱成像技术的油茶果硬度测定及其分布可视化

袁伟东, 姜洪喆, 张聪, 周禹, 蒋雪松, 周宏平*

南京林业大学机械电子工程学院

摘要:【目的】油茶果的硬度是影响其加工性能以及储存和运输过程中品质的关键因素。本研究旨在结合机器学习算法, 利用可见-近红外高光谱成像技术 (VNIR-HIS, 374.98~1038.79 nm) 实现对油茶果硬度的快速、无损、高通量测定。【方法】获取并校准油茶果的高光谱图像后, 从明确的感兴趣区域 (ROI) 中提取全光谱数据, 利用质构仪测定硬度的理化值。将数据集按 3:1 比例随机划分为校准集和预测集, 并对原始光谱数据进行多元散射校正 (MSC)、标准正态变量变换 (SNV)、卷积平滑 (SG)、卷积平滑一阶导数 (SG1) 和二阶导数 (SG2) 等预处理。基于这些预处理方法构建了全光谱偏最小二乘回归 (PLSR) 模型, 并确定 SG1 为最佳预处理方法。为了降低数据维度, 采用连续投影算法 (SPA)、竞争自适应重加权算法 (CARS)、蒙特卡罗-无信息变量消除法 (MC-UVE)、随机蛙跳算法 (RF) 和迭代保留信息变量算法 (IRIV) 选择特征变量, 并基于这些特征变量分别构建 PLSR、粒子群支持向量回归 (PSO-SVR) 和粒子群极限学习机 (PSO-ELM) 模型。【结果】经过波段筛选后, 特征变量占全光谱的 15.7% 至 22.6%。研究表明, CARS-PSO-ELM 模型的预测效果最佳, 其 R_c^2 和 R_p^2 分别为 0.908 和 0.907, RMSEC 和 RMSEP 分别为 1.038 和 0.831。最终, 利用最优预测模型 (CARS-PSO-ELM) 计算油茶果图像中每个像素点的硬度值, 并通过伪彩色映射实现了每颗油茶果硬度分布的可视化。【结论】总体而言, VNIR-HSI 可用于无损快速测量油茶果的硬度检测及可视化。基于 CARS-PSO-ELM 模型建立的硬度预测方法具备稳健性和准确性。这一研究为油茶果硬度检测提供了理论支撑, 也为其他农产品硬度的评估提供了有益参考。

关键词: 高光谱成像技术; 油茶果; 机器学习; 硬度; 可视化

Determination of *Camellia oleifera* Fruit Hardness and Visualization of its Distribution Based on Hyperspectral Imaging Technology

Weidong Yuan, Hongzhe Jiang, Cong Zhang, Yu Zhou, Xuesong Jiang, Hongping Zhou*

College of Mechanical and Electronic Engineering, Nanjing Forestry University

Abstract: 【Objective】The hardness of *Camellia oleifera* fruit is a critical factor influencing its processing performance, as well as its quality during storage and transportation. The aim of this study was to achieve rapid, non-destructive, and high-throughput assessment of *Camellia oleifera* fruit hardness by integrating machine learning algorithms with visible-near infrared hyperspectral imaging (VNIR-HSI, 374.98~1038.79 nm). 【Method】After acquiring and calibrating the hyperspectral images of *Camellia oleifera* fruit, the full spectral data were extracted from designated regions of interest (ROI). The physicochemical values of hardness were determined using a texture analyzer. The dataset was randomly split into a calibration set and a prediction set in a 3:1 ratio. Several preprocessing methods were applied to the raw spectral data, including multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky-Golay smoothing (SG), and its first- and second-order derivatives (SG1 and SG2, respectively). Full spectral partial least squares regression (PLSR) models were constructed based on these preprocessing methods, with SG1 identified as the optimal preprocessing technique. To reduce data dimensionality,

feature variables were selected using methods such as successive projections algorithm (SPA), competitive adaptive reweighted sampling (CARS), Monte Carlo uninformative variable elimination (MC-UVE), random frog (RF), and iteratively retained informative variables (IRIV). PLSR, particle swarm optimization-support vector regression (PSO-SVR), and particle swarm optimization-extreme learning machine (PSO-ELM) models were then constructed based on these feature variables. **【Result】** After band selection, the feature variables accounted for 15.7% to 22.6% of the full spectra. The study revealed that the CARS-PSO-ELM model exhibited the best prediction performance, with R_c^2 and R_p^2 values of 0.908 and 0.907, and RMSEC and RMSEP of 1.038 and 0.831, respectively. Finally, the optimal prediction model (CARS-PSO-ELM) was used to calculate the hardness value of each pixel in the hyperspectral images of *Camellia oleifera* fruit, and pseudo-color mapping was applied to visualize the hardness distribution for each fruit. **【Conclusion】** Overall, VNIR-HSI can be utilized for non-destructive, rapid measurement and visualization of *Camellia oleifera* fruit hardness. The hardness prediction method based on the CARS-PSO-ELM model demonstrates robustness and accuracy, providing theoretical support for *Camellia oleifera* hardness assessment and offering valuable insights for evaluating the hardness of other agricultural products.

Key words: hyperspectral imaging; *Camellia oleifera* fruit; machine learning; hardness; visualization.