

高光谱成像的红提总酸与硬度的预测及其分布可视化

高升¹, 徐建华²

(1.青岛理工大学信息与控制工程学院, 山东 青岛 266520; 2.中国民航大学空中交通管理学院, 天津 300300)

摘要: 利用高光谱成像技术实现对红提总酸和硬度无损检测和分布可视化。首先, 利用高光谱采集生长期360个红提样本在波段450~1 000 nm的高光谱图像信息后用化学方法测定对应样本的总酸, 用质构仪测定硬度。采用KS (Kennard-Stone) 算法将总样本按照3:1的比例划分为训练集 (270个样本) 和测试集 (90个样本)。对红提原始光谱数据分别利用标准正态变量变换 (standard normal variate transformation, SNV)、卷积平滑 (Savitzky-Golay, SG) 处理法、多元散射校正 (multivariate scatter correction, MSC)、归一化等光谱预处理方法处理, 确定最优光谱预处理方法。然后, 分别采用一次降维 (竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS)、连续投影算法 (successive projections algorithm, SPA)、遗传算法 (genetic algorithm, GA)、无信息变量消除法 (uninformative variable elimination, UVE)) 算法和组合降维算法 (CARS-SPA、UVE-SPA) 6种降维方法对光谱信息进行特征变量提取; 分别建立红提总酸和硬度的偏最小二乘回归 (partial least square regression, PLSR) 最优预测模型。最后, 根据所建最优预测模型计算红提图像每个像素点的总酸和硬度, 得到灰度图像并对该灰度图像进行伪彩色变换, 实现红提总酸和硬度的分布可视化。结果表明根据提取到的特征波长对生长期内的红提总酸和硬度进行建模分析得到: 总酸的最优检测模型为MSC-CARS-SPA-PLSR, 其预测集相关系数 R_p 和均方根误差分别为0.985 1、1.348 2, 残差预测偏差 (residual predictive deviation, RPD) 为5.664 3; 硬度的最优检测模型为SG-CARS-PLSR, 其预测集相关系数 R_p 和均方根误差分别为0.929 1、7.935 4, RPD为2.510 8。综上利用高光谱成像技术可以实现红提总酸和硬度的检测与可视化分布, 为生长期红提总酸和硬度的检测及可视化找到一种新方法。

关键词: 红提; 总酸; 硬度; 高光谱成像; 无损检测; 可视化

Hyperspectral Imaging for Prediction and Distribution Visualization of Total Acidity and Hardness of Red Globe Grapes

GAO Sheng¹, XU Jianhua²

(1. School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China;

2. College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In this paper, hyperspectral imaging technology was used for nondestructive detection and distribution visualization of total acidity and firmness of red globe grapes. The hyperspectral information of 360 samples of growing red globe grapes in the wavelength range from 450 to 1 000 nm was collected using a hyperspectral instrument, and the total acidity and firmness of these samples were determined by titration and a texture analyzer, respectively. The Kennard-Stone (KS) algorithm was used to divide the total samples into a training set (270 samples) and a test set (90 samples) in a 3:1 ratio. The collected raw spectral data were preprocessed using various methods such as standard normal variate (SNV), Savitzky-Golay (SG), multivariate scatter correction (MSC), and normalization to determine the best spectral preprocessing method. Then, the feature variables were extracted from the spectral information using six dimensionality reduction algorithms: competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), genetic algorithm (GA), uninformative variable elimination (UVE), CARS-SPA, and UVE-SPA. Using partial least squares regression (PLSR), optimal prediction models for total acidity and firmness were developed separately. Finally, the total acidity and hardness for each pixel of the hyperspectral image were calculated according to the proposed optimal prediction models, and a gray

收稿日期: 2022-03-06

基金项目: 国家自然科学基金面上项目 (31871863; 32072302); 湖北省自然科学基金项目 (2012FKB02910);

湖北省研究与开发计划项目 (2011BHB016)

第一作者简介: 高升 (1988—) (ORCID: 0000-0003-1320-1242), 男, 讲师, 博士, 研究方向为农产品无损智能检测技术及装备。E-mail: gaosheng@qut.edu.cn

scale image was obtained and pseudo-color transformed to visualize the distribution of total acidity and firmness of red globe grapes. The results showed that the optimal prediction model for total acidity was MSC-CARS-SPA-PLSR, with correlation coefficient for the prediction set (R_p), root mean square errors of prediction (RMSEP) and residual predictive deviation (RPD) of 0.985 1, 1.348 2 and 5.664 3, respectively. The optimal prediction model for firmness was SG-CARS-PLSR, with R_p , RMSEP and RPD of 0.929 1, 7.935 4 and 2.510 8, respectively. In summary, hyperspectral imaging provides a new method for the detection and visualization of total acidity and firmness of growing red globe grapes.

Keywords: red globe grapes; total acidity; firmness; hyperspectral imaging; nondestructive detection; visualization

DOI:10.7506/spkx1002-6630-20220306-078

中图分类号: O433.4; O439; S371

文献标志码: A

文章编号: 1002-6630 (2023) 02-0327-10

引文格式:

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. DOI:10.7506/spkx1002-6630-20220306-078. <http://www.spkx.net.cn>

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. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-20220306-078. <http://www.spkx.net.cn>

葡萄是世界四大水果之一, 红提葡萄又名红地球葡萄, 原产于美国, 是美国加利福尼亚州立大学研究人员通过杂交实验培育而成的一个葡萄品种, 自1986年引入我国以来, 在我国广泛种植^[1]。红提因含有较丰富的维生素、矿物质、氨基酸等, 成为我国鲜食葡萄的主要品种之一^[2]。红提果实饱满、鲜食可口、营养价值高, 深受人们的喜爱^[3]。随着消费者购买能力的提升, 红提的内部品质成为决定销售的决定性因素, 直接影响商品的价值。总酸和硬度一直是反映水果成熟度的重要指标^[4-5]。总酸是水果的重要品质之一, 决定了果实的风味, 是消费者进行购买时的重要参考, 也是反映水果成熟状态的重要指标^[5-6]。总酸含量是葡萄品质和风味形成的关键物质, 同时也影响葡萄酒的结构性和清爽性, 此外, 葡萄酒中的果香也是由葡萄内部的酸在发酵过程中参与酯化反应而产生^[7]。水果的硬度是判断其成熟状态和腐烂程度的一个重要指标, 它决定了水果的耐贮性和成熟度^[8], 采后红提果实质地不断发生变化, 内部组织变软, 风味变差, 特别是在长距离运输过程中, 挤压易造成果粒机械损伤, 严重影响红提的感官性能、品质和商品价值^[9-10]。因此检测红提的总酸和硬度具有重要意义。传统检测红提总酸和硬度方法主要是进行破坏性生化实验, 且存在检测时间长、步骤繁琐、只能进行抽样检测等弊端^[11-12], 因此, 利用高光谱成像技术实现对红提硬度和总酸的快速、无损检测, 同时利用无损检测技术, 准确掌握生长期水果总酸和硬度指标及分布规律, 将对果园合理管理、果实成熟预测、果实最佳采摘期、产量预测及成熟后果实的质量检测具有重要的意义^[8]。

高光谱成像技术结合了机器视觉技术和光谱检测分析的优点, 能够同时获取样本的图像信息和光谱信息^[13-15]。具有操作步骤简单、检测速度快、检测准确性高、无损环保的优点, 已成功应用于水果和蔬菜等品质的分析与检测^[16-18]。Baiano等^[19]利用高光谱成像技

术采集Italia、Baresana、Pizzutello、Red Globe、Michele Palieri、Crimson Seedless和Thompson Seedless七个葡萄品种的数据, 分析并建立了葡萄的pH值、可溶性固形物和总酸的偏最小二乘回归 (partial least square regression, PLSR) 模型, 证明光谱信息可以较好地表征每个理化指标的可行性。Nogales-Bueno等^[20]利用近红外高光谱成像技术建立红葡萄和白葡萄的总酚含量、糖度、总酸和pH值的PLSR模型, 实现了对总酚含量、糖度、总酸和pH值的无损检测。许锋等^[21]利用光谱仪 (400~1 000 nm) 运用竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS) 对预处理过的光谱进行特征波长筛选, 最终构建随机森林预测模型, 糖度和总酸的预测集相关系数 (R_p) 分别为0.956 8和0.940 5; 预测均方根误差 (root mean square error of prediction, RMSEP) 分别为0.318 5和0.311 2, 实现了对红提内部的糖度和总酸含量进行定量预测。Dong Jinlei等^[11]利用高光谱成像技术结合连续投影法和无信息变量消除法, 对苹果的可溶性固形物、硬度、水分和pH值进行预测研究。Leiva-Valenzuela等^[22]使用高光谱成像技术采集在500~1 000 nm波段范围内蓝莓高光谱图像, 建立了硬度和可溶性固形物的偏最小二乘预测模型。Fan Shuxiang等^[23]利用高光谱成像技术采集梨的反射图像信息并提取感兴趣区域 (region of interest, ROI) 的平均光谱, 对比分析利用连续投影算法 (successive projections algorithm, SPA)、CARS等不同特征波长提取方法的建模效果, 最终建立梨的可溶性固形物和硬度的最优PLSR预测模型。孙红等^[24]利用高光谱成像实现了对马铃薯叶片含水率进行检测和可视化研究。陈远哲等^[25]采用高光谱成像技术实现了对出缸期皮蛋凝胶品质的含水率和弹性进行可视化检测与不同品质预测。郑涛等^[26]利用高光谱成像技术实现了马铃薯叶片含量无损检测以及叶绿素分布可视化表达。目前已有研究多集中于红提采后贮藏期间内部品质的检

测, 而对生长期红提总酸和硬度的研究相对较少, 且利用建立好的总酸和硬度模型实现生长期红提的伪彩色图像的可视化鲜见报道。

本实验以生长期红提为研究对象, 利用高光谱结合红提硬度和总酸进行研究, 采集生长期红提的高光谱信息, 找出能够表征红提总酸和硬度最优波长组合, 并建立光谱信息与红提硬度和总酸的最优无损预测模型, 进而实现生长期红提硬度和总酸的可视化分析。

1 材料与方 法

1.1 材 料

红提果粒样本均处于生长期。在武汉小宛葡萄园选取10棵红提植株并编号用来研究红提硬度和总酸的变化。由红提开花时间确定实验周期为从红提开花后生长发育的第61天持续到第116天, 每隔5d进行1次采样, 每次采集实验样本为10串红提, 共采样12次, 在每串的上部、中部、下部, 各挑选2粒(每串共6粒红提)大小相近的完整红提果粒共720个, 各选取360个红提果粒作为红提总酸和硬度的实验样本。为了保证实验效果, 样本于实验当天采摘, 采摘后将红提样本进行编号, 需将编号后的样本立即放入温度设置为(22±1)℃, 相对湿度为65%的恒温恒湿箱中保存12h。

1.2 仪 器 与 设 备

TMS-PRO型高精度专业食物物性分析仪 美国FTC公司; HPX-25085H-III恒温恒湿箱 上海新苗医疗器械制造有限公司。

Zolix Hyper SIS-VNIR-CL高光谱成像系统为美国海洋光学公司, 该系统主要由高光谱成像光谱仪(芬兰Spectral Imaging Ltd.公司)、CCD相机(日本Hamamatsu公司)、4个50W的卤素灯(北京卓立汉光仪器有限公司)、1台丝杆式位移控制平台(北京卓立汉光仪器有限公司)等主要部件组成(图1)。该系统采集的光谱波长范围为391~1043nm(含有520个波长), 分辨率为2.8nm, 整个采集系统置于暗箱内。

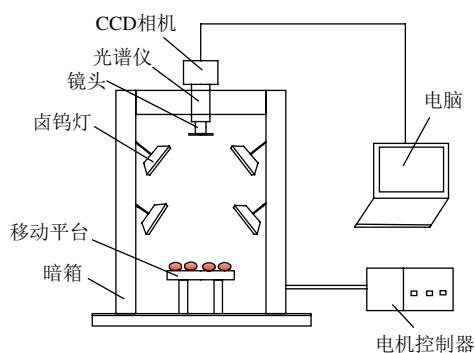


图1 高光谱图像采集系统

Fig. 1 Schematic diagram of hyperspectral image acquisition system

1.3 方 法

1.3.1 高光谱图像采集与校正

高光谱成像系统的参数设置: 相机曝光时间为0.15s, 平台移动速率为1.7mm/s, 移动范围0~245mm, 样本平台与镜头的距离420mm。高光谱成像系统在实验前应进行预热, 本实验设定预热时间为30min。受暗电流及CCD相机芯片不稳定的影响, 高光谱系统采集到的图像会带有一定噪声, 为保证实验精度还需提取对高光谱图像进行黑白校正。黑白校正过程如下: 1) 将标准白板(聚四氟乙烯长方形白板)放在图像信息采集平台上, 获得白板的图像信息 I_w ; 2) 盖上相机盖, 获得全黑图像的图像信息 I_D ; 3) 将红提样本放在采集平台上获得原始高光谱漫反射图像信息 I_R 。

根据式(1)得到校正后样本的图像信息 $R^{[27]}$:

$$R/\% = \frac{I_R - I_D}{I_w - I_D} \times 100 \quad (1)$$

1.3.2 红提硬度和总酸的测定

总酸测定: 将通过挤压装置得到的果汁, 参照GB/T 12456—2008《食品中总酸的测定》方法, 按式(2)计算:

$$\text{总酸含量}/(\text{g}/\text{kg}) = \frac{c \times (V - V_0) \times K}{m} \times 1000 \quad (2)$$

式中: c 为NaOH标准溶液浓度/(mol/L); V 为样品滴定时消耗NaOH溶液体积/mL; V_0 为空白实验消耗的NaOH溶液体积/mL; K 为换算系数(红提中主要为酒石酸, K 取0.075); m 为样品质量/g。

硬度测定: 将进行光谱检测后的红提样品横向放置在质构仪的实验台上, 采用P100/R探头, 设置测前速率为2.0mm/s, 测试速率为1.0mm/s, 测后速率为2.0mm/s, 起始力设置为0.05N, 压缩距离为8.0mm, 对红提果粒进行横向压缩, 压缩部位为红提的中部位置。在进行压缩时, 随着位移的增大, 果粒受到的压力显著增大, 当压缩到一定的位置时, 果实较大压力下会出现脆断, 压力迅速减小, 实验选取在整个压缩过程中果粒所能承受的最大压力作为评判果实硬度的指标。

1.4 数 据 处 理

1.4.1 红提单个果粒光谱数据提取

高光谱的数据采集在自制的带孔载物平板上进行, 因自制的载物平板上可以同时放置15粒红提果粒, 需要分别对单粒图像作为ROI进行分割, 然后提取单粒红提的光谱信息进行研究。在Matlab 2017a软件上对高光谱图像进行处理, 选择图像中整粒红提的图像作为ROI, 每粒红提分3种不同的放置模式(果柄侧朝上、果柄侧朝下, 横放)分别采集单粒红提的3次高光谱信息^[12], 平均光谱为上述3种放置模式获得的数据进行平均, 因之前研究已经表明平均光谱的效果明显好于其他的放置模式, 本研究

直接提取感兴趣区域的平均光谱作为原始光谱进行分析研究。分析时剔除因噪声、外界光照等因素的影响,本实验选取450~1 000 nm(含有439个波长点)的波长点进行建模。在进行感兴趣区域提取时,由前期的研究结果可知^[12],背景与红提果粒区域的反射率在726.6 nm处差值较大,因此选取726.6 nm处的灰度图像进行果粒区域的提取,首先采用Otsu阈值分割方法获得二值图像,然后利用中值滤波和腐蚀运算,得到二值图像并将该二值图像作为掩膜模板对灰度图像进行掩膜操作,最后依次将每个果粒分割出来,依次提取高光谱图像中单个红提果粒的平均光谱信息。

1.4.2 光谱数据预处理

因外界环境、仪器噪声的影响,使采集的原始光谱包含较多的无用信息和噪声,影响后期模型的建立和稳定性。选择合理的光谱预处理方法可有效消除冗余信息,提高模型的稳定性和精度,因此,在进行建模前进行光谱预处理。本研究建模前先进行光谱预处理,能有效消除由于仪器噪声、暗电流等因素的影响,通常为了保证模型的准确性,本研究采用光谱预处理方法包括:标准正态变量变换(standard normal variable transformation, SNV)、卷积平滑处理法(savitzky-golay, SG)等,并针对不同的研究指标确定一种适合的光谱预处理方法。

1.4.3 特征波长提取

SPA^[28]可以挑选出最少冗余信息的变量组,有效解决信息重叠和不共线的问题。

CARS^[29]是一种基于蒙特卡罗采样法对模型进行变量选择的方法,利用PLS建模并去掉权重较小的变量,将交互验证均方根误差(root mean square error of cross validation, RMSECV)最小的变量集作为最终选定的最优变量子集。

无信息变量消除法(uninformative variable elimination, UVE)^[30]的原理为在光谱中加入一个随机矩阵,建立PLSR交互验证模型,将回归系数均值与标准差的商作为衡量指标,挑选出需要的光谱变量。

1.4.4 样本集的划分

KS(Kennard-Stone)算法,原理为把所有的样本都看作训练集候选样本,依次从中挑选样本进训练集。首先选择欧氏距离最远的2个样本进入训练集,其后通过计算剩下的每一个样品到训练集内每一个已知样品的欧式距离,找到拥有最大最小距离的待选样本放入训练集,以此类推,直到达到所要求的样本数目。该方法的优点是能保证训练集中的样本按照空间距离分布均匀。本实验对样本的总酸和硬度按照3:1比例利用KS算法划分为270个校正集样本,90个预测集样本。计算分析校正集和预测集样本的最小值、最大值、平均值、标准差。

1.4.5 模型的建立与评价

高光谱图像数据上每一个像素点都包含一条全波长的光谱反射率曲线,将每个像素点上的光谱数据代入到上述建立的PLSR模型中计算出相应像素点的总酸和硬度,并得到红提果粒的灰度图像;再利用Matlab 2017a软件中伪彩色图像处理的方法将不同总酸和硬度用不同的颜色表示,绘制红提果粒的总酸和硬度的可视化分布,具体步骤如下:

1)采集红提果粒的高光谱图像信息,然后依次分割出单粒红提果粒高光谱图像信息并进行平均,获得单粒果粒的光谱数据;2)利用光谱预处理、特征波长提取等方法分别建立红提果粒的PLSR预测模型;3)提取相应特征波长图像中每个像素的光谱反射率值;4)将高光谱图像数据上每一个像素点所包含一条全波长的光谱反射率数据代入建立的PLSR模型,计算红提果粒每个像素点的总酸和硬度值,得到红提果粒的灰度图像;5)对灰度图像进行伪彩色处理,得到红提果粒总酸和硬度可视化彩色分布。

2 结果与分析

2.1 生长期红提总酸和硬度的变化及光谱特征

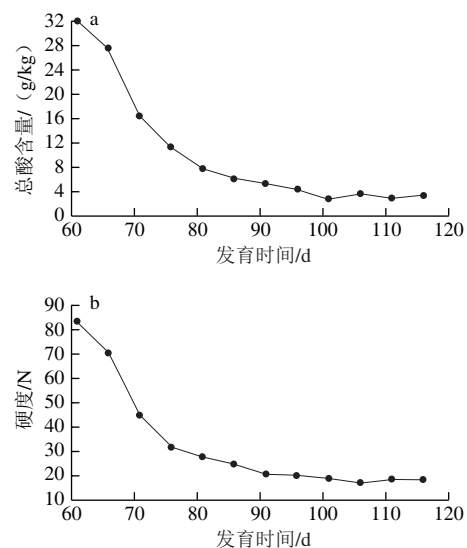


图2 不同生长期红提总酸(a)和硬度(b)的变化

Fig. 2 Changes in total acidity (a) and firmness (b) of red globe grapes at different growth stages

进行12次实验,记录每次实验测定的硬度、总酸数据,取平均值作为最终实验数据,分别绘制样本的内部品质参数随时间的变化图,如图2所示。在红提的整个生长期,总酸指标随着红提成熟而减小。总酸变化分为2个阶段,第1个阶段持续25 d(从第61天至第81天),总酸

极剧下降,第2阶段总酸基本固定,从91 d一直持续到实验结束变化较小。硬度的变化与总酸类似,第1阶段下降明显,第2阶段基本保持在一个固定值。

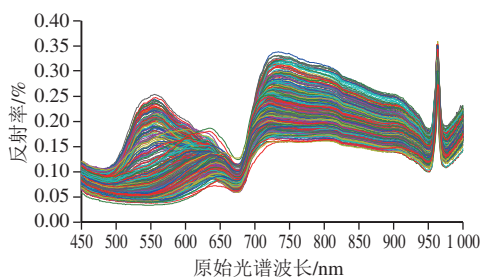


图3 原始光谱

Fig. 3 Original spectra of red globe grapes

从图3可以看出,所有红提样品的光谱都呈现出相同的变化趋势,在450~550 nm区间内曲线平滑,吸光度变化较小,550 nm之后反射强度快速升高,725~820 nm波长光强度较高但变化较小。

表1 采用不同预处理方法的全波长PLSR预测模型

Table 1 Performance of full-wavelength PLSR prediction models using different preprocessing methods

指标	预处理	主因子数	校正集		预测集	
			R_c	RMSEC	R_p	RMSEP
总酸	原始	20	0.981 0	1.994 8	0.979 7	1.329 0
	SNV	20	0.981 2	1.980 1	0.972 7	1.491 6
	SG	14	0.977 7	2.182 2	0.977 8	1.586 3
	MSC	13	0.981 7	1.970 5	0.980 4	1.313 0
	归一化	20	0.980 8	1.994 2	0.975 7	1.483 4
硬度	RAW	10	0.920 5	9.196 8	0.885 6	8.508 4
	SNV	10	0.911 4	9.937 9	0.906 3	8.299 0
	SG	12	0.928 7	8.233 9	0.918 8	7.452 1
	MSC	10	0.919 4	9.535 3	0.918 3	8.051 2
	归一化	10	0.921 1	9.120 2	0.866 5	8.512 1

注:校正均方根误差(root mean square error of calibration, RMSEC)。

由表1可知,对于总酸指标,原始光谱经过多元散射校正(multivariate scatter correction, MSC)预处理后所建PLSR模型的校正集相关系数(R_c)和预测集相关系数(R_p)分别为0.981 7、0.980 4,都高于其他预处理方法后所建模型, RMSEC和RMSEP较小,表明其他的预处理方法与原始光谱经过MSC预处理后所建模型相比并未提高模型预测精度,且原始光谱与其他预处理方法所建模型的预测集相关系数明显低于经过MSC预处理所建模型。因此,对于总酸,选取经过MSC预处理后的光谱进行特征波长提取和建模分析。

对于硬度指标,原始光谱进行SG预处理后所建PLSR模型建立模型的 R_c 和 R_p 分别为0.928 7、0.918 8,都高于其他预处理方法后所建模型, RMSEC和RMSEP较小,分别为8.233 9和7.452 1。表明原始光谱进行SG预处理后的

光谱数据建立的PLSR模型取得了最佳效果。对于硬度,选取经过SG预处理后的光谱进行特征波长提取和建模分析。

2.2 样本集的划分

表2 生长期红提样品利用KS算法划分样品集的数据统计

Table 2 Statistics of total acidity and firmness of sample sets divided by KS algorithm

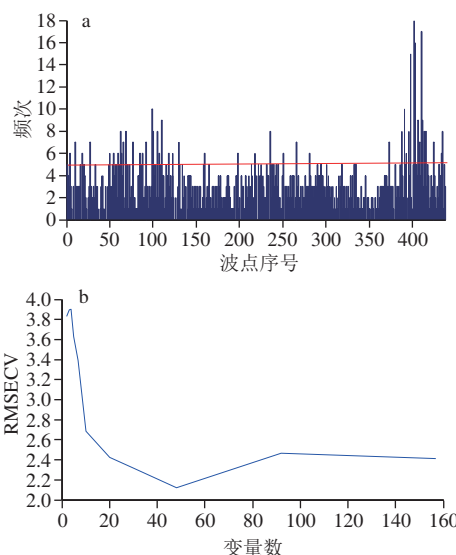
品质参数	校正集				预测集			
	数量	最小值	最大值	平均值	数量	最小值	最大值	平均值
总酸含量/(g/kg)	270	2.320	37.663	11.283	90	2.254	35.415	6.575
硬度/N	270	9.414	121.305	34.061	90	13.447	106.964	30.651

实验中总酸和硬度各采集了360个样本,按照3:1比例利用KS算法划分为270个校正集样品,90个预测集样品。由表2可以看出,总酸含量分布范围为2.254~37.663 g/kg,校正集和预测集的平均值分别为11.283、6.575 g/kg;硬度分布范围为9.414~121.305 N,校正集和预测集的平均值分别为34.061、30.651 N。通过KS法所划分指标的校正集的分布范围比预测集的分布范围更广,证明所挑选出来的校正集样品具有代表性。

2.3 特征波长提取

2.3.1 GA提取特征波长

KS划分后的经过预处理的原始光谱通过GA优选出波长序列分别建立红提总酸和硬度的PLSR模型。以利用GA提取特征波长建立红提总酸PLSR模型为例,在GA运算过程中,设定初始群体为30,交叉率为50%,变异率为1%,迭代次数为100,以最小的RMSECV为标准,筛选出波长点在迭代过程中出现频次较多的波长点为最优波长点;经过20次随机搜索后,最终选定特征波长点为48个,如图4所示。



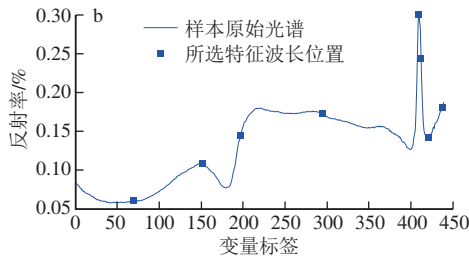
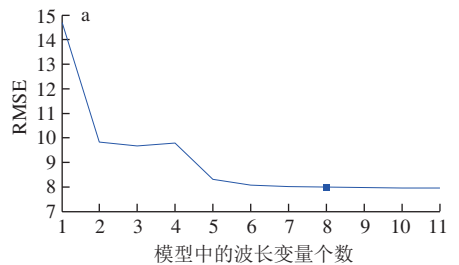
a. GA筛选图; b. RMSECV变化图。

图4 基于GA算法的红提总酸特征波长选取图

Fig. 4 Characteristic wavelength selection for total acidity of red globe grapes based on GA algorithm

2.3.2 SPA提取特征波长

KS划分后的经过预处理的原始光谱通过SPA优选出波长序列分别建立红提总酸和硬度的PLSR模型。以利用SPA提取特征波长建立红提硬度PLSR模型为例,根据SPA的原理,利用均方根误差(root mean square error, RMSE)决定所建模型的优劣, RMSE越小模型的稳定性越好、精度越高。RMSE值随着波长个数的变化,会不断的波动,如图5a所示,当变量个数为8时, RMSE最小为7.99;在原始光谱中所选特征波长位置如图5b所示。



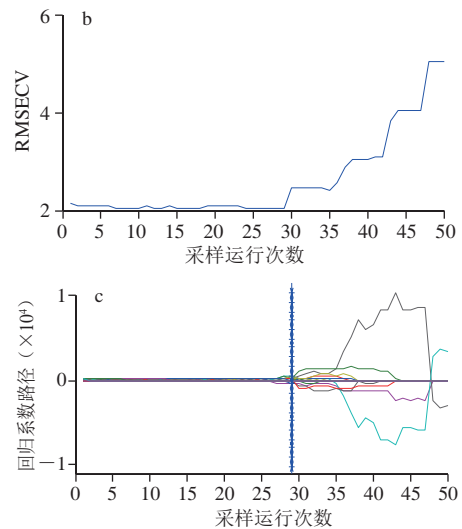
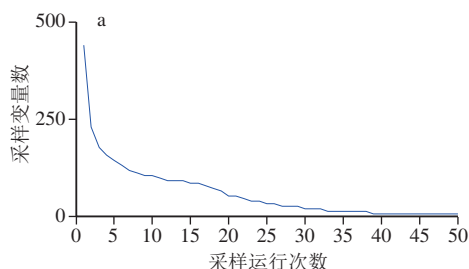
a. RMSE变化图; b. SPA选取波长变量。

图5 基于SPA算法的红提硬度特征波长选取图

Fig. 5 Characteristic wavelength selection for firmness of red globe grapes based on SPA algorithm

2.3.3 CARS提取特征波长

KS划分后的经过预处理的原始光谱通过CARS优选出波长序列分别建立红提总酸和硬度的PLSR模型。以利用CARS提取特征波长建立红提总酸PLSR模型为例,本研究设定蒙特卡罗采样为50次,采用5折交叉验证法。由图6可知,取采样50次所建立的PLSR模型中所对应的最小MSECV作为最优结果,由图6b可知,当RMSECV达到最小值时,各变量的回归系数位于图6c中竖直线位置,采样运行28次。



a.采样变量数; b. RMSECV; c.回归系数路径。

图6 基于CARS算法的红提总酸特征波长选取图

Fig. 6 Characteristic wavelength selection for total acidity of red globe grapes based on CARS algorithm

2.3.4 UVE提取特征波长

KS划分后的经过预处理的原始光谱通过UVE优选出波长序列分别建立红提总酸和硬度的PLSR模型。以利用UVE提取特征波长建立红提硬度PLSR模型为例,应用UVE对光谱数据进行有效信息提取,设定噪声矩阵处最大稳定性绝对值的99%作为剔除阈值,图像黄色曲线代表光谱变量的稳定性值,红色曲线代表噪声变量的稳定性值,两水平虚线为变量的选择阈值(± 46.41),虚线内部的为无用信息被消除,外部的为有用信息,选取结果如图7所示,所对应的波长为选择的特征波长。通过UVE共选择177个特征波长。

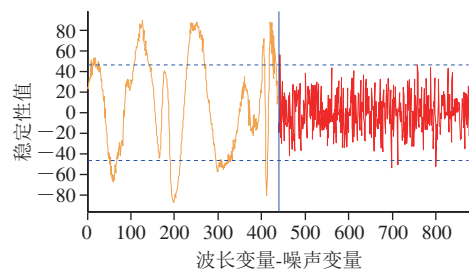


图7 基于UVE算法的红提硬度特征波长选取图

Fig. 7 Characteristic wavelength selection for firmness of red globe grapes based on UVE algorithm

由表3可知,原始光谱经过MSC预处理建立的总酸PLSR模型的校正集相关系数与预测集相关系数相差不大,但建模所用的特征波长数量较多,需要进行特征波长的提取进一步提高模型的稳定性和预测性能。采用不同的特征提取方法,根据最优的波段组合分别建立PLSR模型,UVE算法提取的波段与经过CARS、GA和SPA算法提取的特征波长相比,所提取的波长个数较

多, UVE提取特征波长的数量为298, 占原始光谱数量的67.882%, 预测集的相关系数为0.979 4。通过CARS和SPA一次波段提取算法, 所建模型的校正集和预测集的相关系数都小于原始光谱所建模型的相关系数, 模型虽然大大简化, 但模型的预测性能有所降低。经过UVE-SPA二次波段提取后, 模型的波段数量进一步减小, 模型进一步简化, 但模型的稳定性和预测效果有所减小, 不适合总酸模型的建立。经过CARS-SPA二次特征波段提取后, 模型的校正集相关系数和RMSEC分别为0.987 1、1.662 3, 预测集的相关系数和RMSEP为0.985 1、1.348 2, 残差预测偏差(residual predictive deviation, RPD)为5.664 3, 模型的预测性能显著提高, 提取的特征波长数量为10, 占原始光谱数量的2.278%, 与通过其他算法处理相比, 得到的特征波段最少, 大大简化了模型。

表3 基于不同特征波长提取方法建立的PLSR模型效果
Table 3 Performance of PLSR models based on different feature wavelength extraction methods

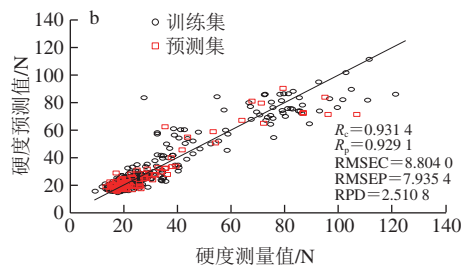
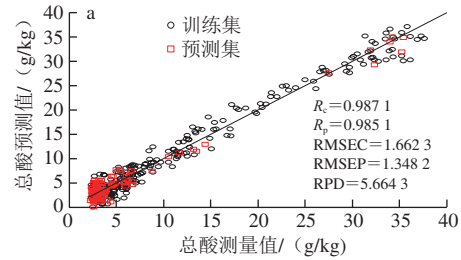
指标	特征波段提取方法	因子数	特征波段个数	校正集		预测集		RPD
				R_c	RMSEC	R_p	RMSEP	
总酸	MSC	13	439	0.981 7	1.970 5	0.980 4	1.313 0	5.223 5
	MSC-CARS	9	20	0.980 4	2.042 6	0.976 7	1.787 1	4.328 9
	MSC-SPA	13	13	0.972 0	2.468 0	0.971 0	1.391 4	4.544 2
	MSC-GA	20	48	0.985 7	1.767 2	0.977 2	1.410 2	4.733 1
	MSC-UVE	18	298	0.986 6	1.713 0	0.979 4	1.355 9	4.953 7
	MSC-CARS-SPA	10	10	0.987 1	1.662 3	0.985 1	1.348 2	5.664 3
硬度	MSC-UVE-SPA	11	23	0.980 6	2.059 4	0.971 1	1.584 1	4.201 1
	SG	12	439	0.928 7	8.233 9	0.918 8	7.452 1	2.297 8
	SG-CARS	9	35	0.931 4	8.804 0	0.929 1	7.935 4	2.510 8
	SG-SPA	8	8	0.914 3	9.589 7	0.911 2	8.292 0	2.292 2
	SG-GA	8	26	0.917 1	9.437 7	0.914 3	8.233 0	2.299 3
	SG-UVE	7	177	0.916 0	9.499 6	0.905 9	8.702 4	2.171 0
	SG-CARS-SPA	7	7	0.908 6	9.887 6	0.906 1	8.701 3	2.175 7
	SG-UVE-SPA	7	7	0.914 3	9.590 9	0.910 0	8.250 4	2.281 3

原始光谱经过SG预处理建立的硬度指标的PLSR模型的校正集相关系数与预测集相关系数较小, 需要进行特征波长的提取进一步提高模型的稳定性和预测性能。采用不同的特征提取方法, 根据最优的波段组合分别建立PLSR模型, UVE算法提取的波段与经过CARS、GA和SPA算法提取的特征波长相比, 所提取的波长个数较多, UVE提取特征波长的数量为177, 占原始光谱数量的40.319%, 预测集的相关系数为0.905 9, 低于通过CARS算法提取特征后所建立的模型。通过SPA、GA和UVE一次波段提取算法, 所建模型的校正集和预测集的相关系数都小于原始光谱经过SG预处理所建模型的相关系数, 模型虽然大大简化, 但模型的预测性能有所降低。经过CARS-SPA二次波段提取后, 模型的波段数量进一步减小, 模型进一步简化, 但模型的稳定性和预测效果没有明显改善, 不适合硬度指标模型的建立。经过CARS特征波段提取后, 模型的校正集相关系数和RMSEC分别为0.931 4、8.804 0, 预测集相关系数和RMSEP为0.929 1、7.935 4, RPD为2.510 8, 模型的预测性能显著提高, 提

取的特征波长数量为35, 占原始光谱数量的7.973%, 大大简化了模型。

2.4 模型的建立与结果比较

分别利用最优特征波长组合建立红提总酸和硬度的最佳PLSR模型, 训练集和预测集样本的预测值和理化测量值之间的散点图如图8所示。



a. 基于MSC-CARS-SPA-PLSR红提总酸最优模型; b. 基于SG-CARS-PLSR红提硬度最优模型。

图8 基于最优特征波长组合建立的红提总酸和硬度的PLSR模型
Fig. 8 PLSR models for total acidity and firmness based on optimal combination of characteristic wavelengths of red globe grape

由图8可知, 所建立的红提总酸和硬度的最优PLSR模型的校正集相关系数 R_c 分别为0.987 1、0.931 4, RMSEC为1.662 3、8.804 0; 预测集相关系数 R_p 分别为0.985 1、0.929 1, RMSEP为1.348 2、7.935 4, 模型的RPD分别为5.664 3、2.510 8, 均大于2.5, 表明预测效果较好, 两者的预测性能都满足精度的要求。

进行特征波长提取后, 所建红提果粒总酸和硬度指标的最优模型最优特征波长如表4所示。

表4 红提总酸和硬度的PLSR预测模型的最优特征波点列表
Table 4 List of optimal characteristic wavelengths for the establishment of PLSR prediction models for total acidity and hardness of red globe grapes

指标	建模方法	波长/nm
总酸 (10个)	MSC-CARS-SPA-PLSR	499.23、531.89、727.85、752.97、908.74、951.44、969.03、971.54、975.31
		450.24、451.50、464.06、504.26、510.54、518.08、519.33、538.17、541.94、543.20、549.48、553.25、572.09、573.35、575.86、578.37、
硬度 (35个)	SG-CARS-PLSR	612.29、681.37、810.76、857.23、938.88、943.91、955.21、957.73、964.01、969.03、975.31、977.82、980.34、982.85、986.62、991.64、995.41、997.92、1 000.43

基于光谱信息建立的红提总酸和硬度的PLSR预测模型的最优特征波点所建方程:

$$Y_{\text{总酸}} = -97.666\lambda_{499.23} + 170.126\lambda_{531.89} - 529.021\lambda_{727.85} + 787.844\lambda_{752.97} - 641.820\lambda_{908.74} + 184.022\lambda_{951.44} - 112.858\lambda_{962.75} - 460.869\lambda_{969.03} + 601.607\lambda_{971.54} + 438.774\lambda_{975.31} + 11.283$$

$$Y_{\text{硬度}} = 240.290\lambda_{450.24} + 411.825\lambda_{451.50} + 144.668\lambda_{464.06} - 294.831\lambda_{504.26} - 293.54\lambda_{510.54} - 302.601\lambda_{518.08} - 256.819\lambda_{519.33} + 333.855\lambda_{538.17} + 267.675\lambda_{541.94} + 310.327\lambda_{543.20} + 223.103\lambda_{549.48} + 187.807\lambda_{553.25} - 277.903\lambda_{572.09} - 277.428\lambda_{573.35} - 302.071\lambda_{575.86} - 227.419\lambda_{578.37} + 494.527\lambda_{612.29} - 433.806\lambda_{681.37} - 263.993\lambda_{810.76} - 345.177\lambda_{857.23} + 508.772\lambda_{938.88} + 499.032\lambda_{943.91} + 550.131\lambda_{955.21} + 294.009\lambda_{957.73} - 261.226\lambda_{964.01} - 427.250\lambda_{969.03} + 375.825\lambda_{975.31} + 268.700\lambda_{977.82} + 234.524\lambda_{980.34} + 346.980\lambda_{982.85} + 408.840\lambda_{986.62} + 197.043\lambda_{991.64} - 401.206\lambda_{995.41} - 702.009\lambda_{997.92} - 390.396\lambda_{1000.43} + 34.061$$

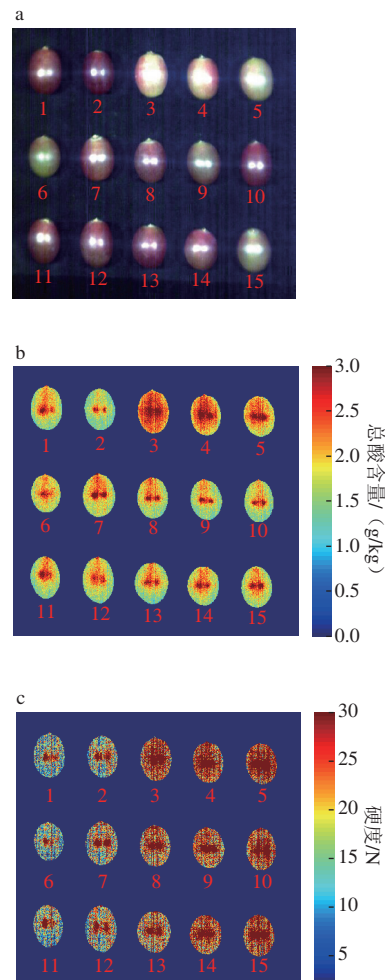
式中： $Y_{\text{总酸}}$ 为总酸的预测值； $Y_{\text{硬度}}$ 为硬度的预测值； λ_i 为特征波长处的反射率（下标*i*表示波长/nm）。

2.5 红提果粒总酸与硬度的分布

利用高光谱图像采集系统，分别采集生长期红提果粒的高光谱图像信息，采集的高光谱原始图像如图9a所示，该图像中包含15粒红提果粒。选取其中的一次高光谱采集图像，首先分割出每粒红提的图像，利用Matlab 2017a编程提取高光谱图像中每一个像素点的光谱信息，根据MSC-CARS-SPA-PLSR模型计算出各个像素点的红提果粒总酸分布图，如图9b所示。图中蓝色代表总酸的最低值0 g/kg，红色代表总酸的最高值3 g/kg，从蓝色到红色代表总酸逐渐上升。通过可视化可以清晰的明确红提总酸的分布规律，由图9b可知，1、2、10、11、12、13号红提果粒为深红色，果实相对成熟，通过总酸的伪彩色图像可视化分布可知红色像素点相对较少，表明总酸的含量较低，3、4、5、6号红提果粒颜色较绿，果实相对不成熟，通过总酸的伪彩色图像可视化分布可知红色像素点相对较多，表明总酸的含量较高，总酸含量较高，表明随着果粒的逐渐成熟，总酸的含量降低，与实际的化学实验结果相一致^[10]，从而证明了红提果粒总酸可视化的正确性。对单个的果粒进行分析，红提中间位置的总酸含量相对较高，可能与内部的红提果核有关，表皮附近的总酸含量相对较少。

利用高光谱图像采集系统，分别采集生长期红提果粒的高光谱图像信息，选取其中的一次高光谱采集图像，首先分割出每粒红提的图像，利用Matlab 2017a编程提取高光谱图像中每一个像素点的光谱信息，根据SG-CARS-PLSR模型计算出各个像素点的红提果粒硬度分布图，如图9c所示。图中蓝色代表硬度的最低值

5 N，红色代表硬度的最高值为30 N，从蓝色到红色代表硬度逐渐上升。通过可视化可以清晰的明确红提硬度的分布规律，由图9c可知，1、2、11、12号红提果粒为深红色，果实相对成熟，通过硬度的伪彩色图像可视化分布可知红色像素点相对较少，表明硬度的含量较低，3、4、5、9、15号红提果粒颜色较绿，果实相对不成熟，通过硬度的伪彩色图像可视化分布可知红色像素点相对较多，表明硬度的含量较高，硬度较高，6号红提果粒虽然颜色较轻，但果实较软，因此硬度值较低；表明随着果粒的逐渐成熟，硬度降低，与实际实验结果一致^[10]，从而证明了红提果粒硬度可视化的正确性。对每个红提果粒进行分析可知，在6、7、11、12号红提果粒中红提果柄侧及中间位置的红色区域面积相对果柄另外一侧较多，表明果柄侧及中间位置硬度含量相对较高，也与实际情况中果柄侧颜色相对较青一致。



a.高光谱原始图像；b.总酸分布可视化图；c.硬度分布可视化图。

图9 红提果粒总酸和硬度的分布可视化图

Fig. 9 Visualization of distribution of total acidity and firmness of red globe grapes

3 讨论与结论

本研究为尽量消除中心反光的问题,采用了2种方法:第1种方法:选用4个环绕卤素灯光源进行照射,并不断优化4个卤素灯的位置和照射角度,使尽量避免中心反光的问题;第2种方法:在进行建模时,选择应用SG等光谱预处理在一定程度上可消除中间位置反光等干扰信息的影响。红提果粒成球形,每个果粒都会在果粒中心位置因球形的果粒外形产生反光区域,中间反光的影响依然存在,但中心反光区域的面积较小,整体对果粒的影响有限,光谱曲线虽然会应反光使中心位置的光谱强度增强,但是光谱曲线吸收等位置的特性与其中内部氢键有关,在该反光区域实现了反射率的扩大使中间反射光的部位的确有更高的反射率,导致中间位置在高光谱可视化时候造成一定的影响。使在可视化时中间反光区域位置的总酸和硬度预测值较高,后期将不断对所采集原始光谱的光照反射率进行调整和优化,后期在本研究基础上将根据反射光强度变化程度,对采集的反射光谱数据进行相应的缩减,以进一步减小反光的影响。

以可高光谱成像技术为主要技术手段,以生长期的红提果粒为研究对象,采集红提果粒高光谱信息构建了红提果粒总酸和硬度检测模型,通过比较分析,得到红提果粒总酸和硬度的最优预测模型,并通过伪彩色图像实现红提果粒总酸和硬度的可视化分布。主要结论如下:

红提果粒总酸的最佳模型为通过二次波段提取的方法,红提果粒硬度的最佳模型为通过一次波段提取的方法。对生长期内的红提果粒进行建模分析,深入分析不同检测方法对各指标的预测优劣,得到最佳预测模型:总酸的最优检测模型为MSC-CARS-SPA-PLSR,其预测集相关系数 R_p 和RMSEP分别为0.985 1、1.348 2,RPD为5.664 3;硬度的最优检测模型为SG-CARS-PLSR模型,其预测集相关系数 R_p 和RMSEP分别为0.929 1、7.935 4,RPD为2.510 8。分别利用MSC-CARS-SPA-PLSR、SG-CARS-PLSR预测模型计算出高光谱图像中每个像素点的红提果粒的总酸和硬度值,结合伪彩色图像处理技术得到可视化分布图,结果表明采用高光谱成像技术可以实现红提果粒总酸和硬度的分布可视化,为直观监测红提果粒生长状态及成熟度判别提供可靠的理论依据。

参考文献:

- [1] 黄静. 鲜食葡萄酒特性及葡萄酒品质的研究[D]. 石河子: 石河子大学, 2014.
- [2] 赵胜建. 红地球葡萄国内外栽培现状及特点[J]. 北方园艺, 1998(5): 34-35.
- [3] 陈辰, 鲁晓翔, 张鹏, 等. 基于可见-近红外漫反射光谱技术的葡萄贮藏期间可溶性固形物定量预测[J]. 食品科学, 2015, 36(20): 109-114. DOI:10.7506/spkx1002-6630-201520020.
- [4] GUO W C, ZHAO F, DONG J L. Nondestructive measurement of soluble solids content of kiwifruits using near-infrared hyperspectral imaging[J]. Food Analytical Methods, 2016, 9: 38-47. DOI:10.1007/s12161-015-0165-z.
- [5] ZHANG D Y, XU L, WANG Q Y, et al. The optimal local model selection for robust and fast evaluation of soluble solid content in melon with thick peel and large size by vis-NIR spectroscopy[J]. Food Analytical Methods, 2019, 12: 136-147. DOI:10.1007/s12161-018-1346-3.
- [6] 高升, 王巧华, 李庆旭, 等. 基于近红外光谱的红提维生素C含量、糖度及总酸含量无损检测方法[J]. 分析化学, 2019, 47(6): 941-949. DOI:10.19756/j.issn.0253-3820.191124.
- [7] CONDE C, SILVA P, FONTES N, et al. Biochemical changes throughout grape berry development and fruit and wine quality[J]. Food, 2006, 1(1): 1-22.
- [8] 邓云, 吴颖, 李云飞. 温度和相对湿度对采后葡萄浆果硬度的影响[J]. 食品科学, 2007, 28(2): 46-49. DOI:10.3321/j.issn:1002-6630.2007.03.006.
- [9] 田海龙, 张平, 农绍庄, 等. 基于TPA测试法对1-MCP处理后葡萄果实实质构性能的分析[J]. 食品与机械, 2011, 27(3): 104-107. DOI:10.3969/j.issn.1003-5788.2011.03.033
- [10] 贾楠, 尹勇刚, 李敏敏, 等. ‘玫瑰香’葡萄果实成熟过程中质地与解剖结构的变化[J]. 中国农业大学学报, 2022, 27(2): 78-86. DOI:10.11841/j.issn.1007-4333.2022.02.08.
- [11] DONG J L, GUO W. Nondestructive determination of apple internal qualities using near-infrared hyperspectral reflectance imaging[J]. Food Analytical Methods, 2015, 8(10): 2635-2646. DOI:10.1007/s12161-015-0169-8.
- [12] 高升, 王巧华, 付丹丹, 等. 红提糖度和硬度的高光谱成像无损检测[J]. 光学学报, 2019, 39(10): 355-364. DOI:10.3788/AOS201939.1030004.
- [13] 黄蒂云, 李景彬, 尤佳, 等. 基于高光谱技术融合图像信息的脱绒棉品种分类检测研究[J]. 光谱学与光谱分析, 2018, 38(7): 2227-2232. DOI:10.3964/j.issn.1000-0593(2018)07-2227-06.
- [14] 蔡庆空, 李二俊, 蒋金豹, 等. 联合光谱-空间信息的短波红外高光谱图像茶叶识别模型[J]. 光谱学与光谱分析, 2019, 39(8): 2522-2527. DOI:10.3964/j.issn.1000-0593(2019)08-2522-06.
- [15] 苗荣慧, 黄锋华, 杨华, 等. 基于光谱和Gabor纹理信息融合的油桃品种识别[J]. 江苏农业科学, 2019, 47(6): 174-178. DOI:10.15889/j.issn.1002-1302.2019.06.040.
- [16] 孙世鹏, 彭俊, 李瑞, 等. 基于近红外高光谱图像的冬枣损伤早期检测[J]. 食品科学, 2017, 38(2): 301-305. DOI:10.7506/spkx1002-6630-201702047.
- [17] 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. DOI:10.7506/spkx1002-6630-201516018.
- [18] 郭文川, 董金磊. 高光谱成像结合人工神经网络无损检测桃的硬度[J]. 光学精密工程, 2015, 23(6): 1530-1537. DOI:10.3788/OPE.20152306.1530.
- [19] BAIANO A, TERRACONE C, PERI G, et al. Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes[J]. Computers & Electronics in Agriculture, 2012, 87: 142-151. DOI:10.1016/j.compag.2012.06.002.
- [20] NOGALES-BUENO J, HERNÁNDEZ H, JOSÉ M, et al. Determination of technological maturity of grapes and total

- phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: a preliminary approach[J]. *Food Chemistry*, 2014, 152: 586-591. DOI:10.1016/j.foodchem.2013.12.030.
- [21] 许锋, 付丹丹, 王巧华, 等. 基于MCCV-CARS-RF建立红提糖度和酸度的可见-近红外光谱无损检测方法[J]. *食品科学*, 2018, 39(8): 149-154. DOI:10.7506/spkx1002-6630-201808024.
- [22] LEIVA-VALENZUELA G A, LU R, AGUILERA J M. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging[J]. *Journal of Food Engineering*, 2013, 115(1): 91-98. DOI:10.1016/j.jfoodeng.2012.10.001.
- [23] FAN S X, HUANG W Q, GUO Z M, et al. Prediction of soluble solids content and firmness of pears using hyperspectral reflectance imaging[J]. *Food Analytical Methods*, 2015, 8(8): 1936-1946. DOI:10.1007/s12161-014-0079-1.
- [24] 孙红, 刘宁, 吴莉, 等. 高光谱成像的马铃薯叶片含水率分布可视化[J]. *光谱学与光谱分析*, 2019, 39(3): 910-916. DOI:10.3964/j.issn.1000-0593(2019)03-0910-07.
- [25] 陈远哲, 王巧华, 高升, 等. 皮蛋凝胶品质含水率和弹性的高光谱预测及其可视化[J]. *食品科学*, 2022, 43(2): 324-331. DOI:10.7506/spkx1002-6630-20201215-173.
- [26] 郑涛, 刘宁, 孙红, 等. 基于高光谱成像的马铃薯叶片叶绿素分布可视化研究[J]. *农业机械学报*, 2017, 48(增刊1): 153-159; 340. DOI:CNKI:SUN:NYJX.0.2017-S1-025.
- [27] 付丹丹, 王巧华. 鸡蛋新鲜度、pH值及黏度的高光谱检测模型[J]. *食品科学*, 2016, 37(22): 173-179. DOI:10.7506/spkx1002-6630-20160411-126.
- [28] YE S F, WANG D, MIN S G. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection[J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, 91(2): 194-199. DOI:10.1016/j.chemolab.2007.11.005.
- [29] 李江波, 彭彦昆, 陈立平, 等. 近红外高光谱图像结合CARS算法对鸭梨SSC含量定量测定[J]. *光谱学与光谱分析*, 2014, 35(5): 1264-1269. DOI:10.3964/j.issn.1000-0593(2014)05-1264-06.
- [30] 刘燕德, 肖怀春, 孙旭东, 等. 柑桔叶片黄龙病光谱特征选择及检测模型[J]. *农业工程学报*, 2018, 34(3): 180-187. DOI:10.11975/j.issn.1002-6819.2018.03.024.