

# 基于高光谱的茄子外部缺陷检测

池江涛, 张淑娟\*, 任锐, 廉孟茹, 孙双双, 穆炳宇

(山西农业大学农业工程学院, 山西太谷 030801)

**摘要:** 利用高光谱(900~1700 nm)对完好、木栓化和烂果茄子进行识别研究。共采摘了 252 个茄子样本, 包含完好茄子 170 个, 木栓化茄子 60 个和烂果茄子 22 个, 利用高光谱成像系统采集完好、木栓化和烂果 3 种区域一共 252 个样本的高光谱图像, 然后提取合理的感兴趣区域(ROI)获得样本光谱数据。采用多种预处理方法进行光谱预处理, 建立偏最小二乘(partial least squares method, PLS)判别分析模型, 结果表明, 经 normalize 预处理后模型的预测效果最好, 因此选择 normalize 作为预处理方法。基于预处理后的光谱数据, 采用连续投影法(SPA)、回归系数法(RC)和竞争性自适应重加权算法(CARS)提取特征波长, 并分别建立偏最小二乘(PLS)和多元线性回归(MLR)判别模型进行研究。结果表明: CARS-MLR 模型对 3 种类型样本鉴别效果最佳, 其校正集决定系数  $R_c^2$  为 0.94, 预测集决定系数  $R_p^2$  为 0.90, RMSEC 和 RMSEP 分别为 0.19 和 0.21, 预测集判别准确率达到 96.82%。本研究采用高光谱可以对完好、木栓化和烂果茄子进行有效鉴别, 为茄子的缺陷无损检测提供了理论参考。

**关键词:** 茄子; 缺陷; 高光谱; 分类检测

文章编号: 1673-9078(2021)09-279-284

DOI: 10.13982/j.mfst.1673-9078.2021.9.0034

## Detection of Eggplant External Defects Using Hyperspectral Technology

CHI Jiang-tao, ZHANG Shu-juan\*, REN Rui, LIAN Meng-ru, SUN Shuang-shuang, MU Bing-yu

(College of Agricultural Engineering, Shanxi Agricultural University, Taigu 030801, China)

**Abstract:** Hyperspectral technology (in the range of 900~1700 nm) was employed to distinguish sound, suberized and rotten eggplants. A total of 252 eggplant samples were collected, including 170 sound eggplants, 60 suberized eggplants and 22 rotten eggplants. The hyperspectral imaging system was used to acquire hyperspectral images of 252 eggplant samples in the three types of areas (sound, suberized and rotten fruit regions), and a reasonable region of interest (ROI) was extracted to obtain the spectral data. A variety of preprocessing methods were used for spectral pretreatment and a partial least squares (PLS) discriminant analysis model was established. The results show that the prediction was the best after the model was subjected to normalization pretreatment. Therefore, normalization was selected as the pretreatment method, based on the pre-treated spectral data, the characteristic wavelengths were extracted by successive projections algorithm (SPA), regression coefficient (RC) and competitive adaptive reweighted sampling (CRAS) methods, and partial least squares (PLS) and multiple linear regression (MLR) discriminant models were established for analysis. The results showed that the CRAS-MLR model discriminate most effectively the 3 types of samples, with the coefficient of determination for the calibration set ( $R_c^2$ ) as 0.944, the coefficient of determination for prediction set ( $R_p^2$ ) as 0.901, and the RMSEC and RMSEP as 0.199 and 0.213, respectively. The discriminant accuracy of the prediction set was 96.8%. In this study, hyperspectral technology can be used to distinguish effectively the sound, suberized and rotten eggplants, which provides a theoretical reference for the nondestructive detection of eggplant defects.

**Key words:** eggplant; defects; hyperspectral; classification detection

引文格式:

池江涛,张淑娟,任锐,等.基于高光谱的茄子外部缺陷检测[J].现代食品科技,2021,37(9):279-284,+178

CHI Jiang-tao, ZHANG Shu-juan, REN Rui, et al. Detection of eggplant external defects using hyperspectral technology [J]. Modern Food Science and Technology, 2021, 37(9): 279-284, +178

收稿日期: 2021-01-10

基金项目: 国家自然科学基金青年基金项目(31801632); 山西省高等学校科技创新项目(2019L0396)

作者简介: 池江涛(1996-), 男, 硕士研究生, 研究方向: 农产品无损检测技术及装备, E-mail: 741737059@qq.com

通讯作者: 张淑娟(1963-), 女, 教授, 研究方向: 农产品无损检测技术及装备, E-mail: zsujuan1@163.com

茄子 (*Solanum melongena* L.) 又称昆仑瓜、矮瓜、落苏和酪酥等, 属于非呼吸跃变型果实, 起源于亚洲东南热带地区, 西汉时传入我国<sup>[1]</sup>。茄子富含膳食纤维、维生素、多酚、蛋白质和矿物质等多种营养物质, 具有降血脂、防治高血压和糖尿病、保肝以及抗氧化等保健功效<sup>[2]</sup>。茄子的生产和开发利用市场前景广阔, 然而茄子存在木栓化和烂果等缺陷问题, 严重影响了茄子的产量和品质, 进而降低了其商品性。木栓化可能是高温或者气候变化异常导致茄子钙硼缺失引起; 烂果则可能由虫害、菌害、雨水和光照等因素造成。在实际生产过程中, 将茄子木栓化和烂果样本剔除仍然靠人工来完成, 不仅耗时耗力、效率低下, 还易造成漏选, 因此实现一种能够快速、准确识别木栓化和烂果茄子的方法则尤为重要。

高光谱成像技术将光谱分析技术和数字成像技术相结合, 可以同时获得样本大量波段的空间图像信息和每一像素点的光谱信息, 具有灵敏度高、测量速度快和抗干扰能力强等优点, 广泛应用于农产品无损检测、病害检测等领域中<sup>[3]</sup>。刘立新等<sup>[4]</sup>利用高光谱技术结合机器学习算法, 基于 CARS 提取的特征波段建立 LDA、KNN 和 SVM 模型来鉴别红枣品种, 总准确率分别达到 85.53%、98.68% 和 98.25%; A. Del Fiore 等<sup>[5]</sup>利用高光谱成像技术, 在光谱范围为 400~1000 nm 内对玉米产毒真菌进行研究, 实现了玉米早期病害的检测; Folch-Fortuny 等<sup>[6]</sup>利用高光谱成像技术, 采用排列测试法提取特征波长并建立多路 PLS 判别模型, 实现了柑橘腐烂病的检测; Sun 等<sup>[7]</sup>利用高光谱成像技术对无损、轻度腐烂、中度腐烂和重度腐烂的桃子进行病害检测, 进一步提高了其病害识别率; 李宏强等<sup>[8]</sup>采用高光谱分析技术结合模式识别, 通多分层、分步骤建立了 8 种马铃薯微型种薯的分类模型, 其平均正确识别率达到 89.75%, 表明高光谱分析技术可用于马铃薯微型种薯的分类检测; YEH 等<sup>[9]</sup>基于高光谱成像技术, 对草莓叶状炭疽病进行研究, 利用光谱角度映射器实现了对病害的 3 个不同感染阶段的有效检测; LAN 等<sup>[10]</sup>运用多种机器学习算法, 分别对健康和黄龙病感染的柑橘多光谱样本建模, 实现了很强的分类效果。上述研究均为利用高光谱成像技术对各种农产品进行识别分类, 就目前研究状况来看, 尚未有学者利用高光谱对茄子外部缺陷类型进行鉴别研究。

本研究采用高光谱技术对茄子完好、木栓化和烂果进行检测, 通过多种预处理方法对原始光谱数据进行预处理, 并建立 PLS 判别模型比较分析, 选择最佳预处理方法进行后续研究。采用 SPA、RC 和 CARS 对预处理后的光谱数据提取特征波长, 基于特征波长

分别建立 PLS 和 MLR 判别模型进行比较分析, 以实现茄子缺陷的定性判别, 为进一步开发茄子在线分选装备提供了理论依据。

## 1 材料与方法

### 1.1 试验材料

所有用于试验的茄子样本均采摘于山西省晋中市太谷区明星乡武家堡村 (茄子于 2020 年 5 月 23 日移栽, 7 月 20 日采摘), 品种为“紫光”茄子。为保证研究的可靠性, 采摘时选择大小 (单果质量 450 g~680 g) 均匀, 形状为近圆球形以及缺陷类型 (完好、木栓化和烂果) 齐全得茄子作为试验样本, 图 1 所示为 3 类样本图。运回实验室后, 对其表面泥土进行清理以避免误判。共挑选 252 个样本, 包含完好样本 170 个, 木栓化样本 60 个和烂果样本 22 个。采集各样本的高光谱图像, 然后从中提取 252 个光谱数据, 运用 Kennard-Stone 算法将 3 类样本按近似 3:1 的比例随机划分为校正集样本 189 个和预测集样本 63 个。

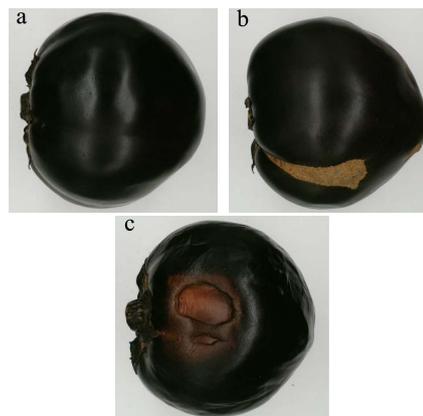


图 1 采集的茄子样本

Fig.1 Sample of eggplant collected

注: a: 完好; b: 木栓化; c: 烂果。

### 1.2 高光谱信息采集

本试验采集样本信息的仪器为北京卓立公司开发的“盖亚”高光谱分选仪, 仪器部件包括: 计算机、Image-λ-N17E 光谱相机、4 个溴钨灯、暗箱以及电控移动平台, 采集到的光谱波长范围为 900~1700 nm。采集的信息过度饱和会出现失真现象, 因此在信息采集前需调试曝光时间和平台的移动速度。根据实验经验, 设置曝光时间为 20 ms, 平台移动速度为 2 cm/s, 样本与镜头的距离为 22 cm 时, 采集效果最佳。

为消除光强的变化和镜头中暗流对成像产生影响, 以及计算扫描物体的相对反射光谱值, 因此要在光谱采集前进行黑白板校正<sup>[11]</sup>。计算公式如式 (1):

$$R = \frac{I - B}{W - B} \quad (1)$$

式中, R 是校正后的高光谱图像, I 是原始高光谱图像, B 是黑板校正后的高光谱图像, W 是白板校正后的高光谱图像。

### 1.3 光谱分析和数据处理软件

研究中的数据处理和分析基于 ENVI 4.7、Matlab R2012a、The Unscrambler X 10.1、Origin 8.5 和 Microsoft Excel 2010 等软件完成。

## 2 基于高光谱数据的茄子外观品质定性判别分析

### 2.1 茄子完好、木栓化和烂果 3 种区域的平均光谱曲线图

使用 ENVI 4.7 软件的提取感兴趣区域 (ROI) 函数分别提取茄子的木栓化、烂果和完好 3 种区域的光谱数据, 然后计算并求取各类样本的平均光谱, 如图 2 所示。“紫瓜”茄子呈类球状, 且表皮光滑发亮, 导致采集到的高光谱图像中间区域漫反射强度大、信噪比高, 影响建模精度以及试验可靠性。因此, 在利用 ENVI4.7 提取感兴趣区域时应避开中间反光区域。

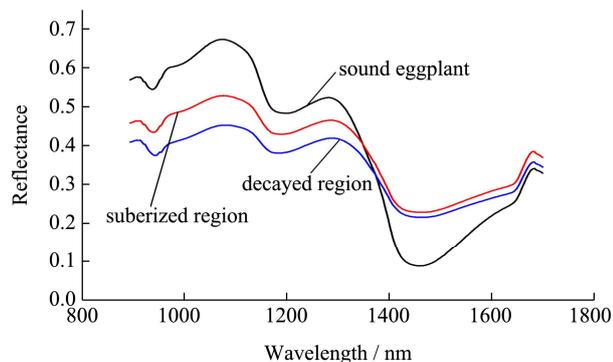


图 2 完好、木栓化和烂果区域平均光谱图

Fig.2 Average spectral image of sound, suberized and decayed regions

由图 2 可知, 完好茄子、木栓化区域和烂果区域的平均光谱曲线具有很大差别, 在 900~1300 nm 范围内, 完好区域的反射率最高, 原因可能是完好茄子表皮光滑, 对光的反射最强; 在 1200 nm 附近的 3 种曲线均为波谷, 这是由于茄子表皮叶绿素的 C-H 基团二级倍频吸收作用<sup>[12]</sup>; 大于 1350 nm 的波段范围, 完好区域的反射率低于木栓化和烂果区域的反射率。

### 2.2 光谱数据预处理

对采集到的原始光谱数据预处理, 可以有效降低或消除因采集背景、噪声干扰和暗电流等产生的大量与样本固有性质无关的冗余信息, 起到提高模型精度和预测能力的作用<sup>[13]</sup>。本研究采用标准归一化 (standard normalized variate, SNV)、多元散射校正 (multiplicative scatter correction, MSC)、中值滤波 (median filter, MF)、卷积平滑 (savitzkygolay, SG) 和归一化 (normalize) 等方法对原始光谱数据进行预处理, 通过建立 PLS 模型, 比较模型参数决定系数 ( $R_c^2$ 、 $R_p^2$ ) 和标准偏差 (RMSEC、RMSEP) (决定系数越大, 标准偏差越小, 其建模精度越高) 以确定最佳光谱预处理方法。

表 1 不同预处理方法对 3 类茄子 PLS 分类模型精度的影响

Table 1 Effects of different pretreatment methods on the accuracy of PLS classification model of eggplant

预处理方法	校正集		预测集	
	$R_c^2$	RMSEC	$R_p^2$	RMSEP
Raw	0.70	0.35	0.82	0.29
SNV	0.69	0.39	0.79	0.32
MSC	0.62	0.39	0.68	0.38
MF	0.71	0.35	0.82	0.28
SG	0.70	0.34	0.82	0.29
normalize	0.74	0.33	0.85	0.26

由表 1 可以看出, 经 Normalize 预处理后所建立的 PLS 模型精度最高, 其校正集决定系数  $R_c^2$  最大, 为 0.74; 均方根误差 RMSEC 最小, 为 0.33。其预测集决定系数  $R_p^2$  为 0.85, 均方根误差 RMSEP 为 0.26, 同样均为最优。因此, 最终选择 Normalize 作为预处理方法进行后续研究。

### 2.3 特征波段提取

特征波段来源于全光谱波段, 携带其最重要的光谱判别信息。其作用主要有: 消除原始数据的线性相关性、奇异性和不稳定性; 降低数据维数, 减少变量数, 排除多余的干扰信息等。特征波段的提取直接影响模型建立的效率以及建模后预测结果的可靠性和准确性。

#### 2.3.1 连续投影法 (SPA)

连续投影算法 (successive projection algorithm, SPA) 是一种使矢量空间共线性最小化的前向变量选择算法。作为一种新兴的特征波长筛选方法, 它能够有效消除波长变量之间共线性的影响, 进而有效提取出特征波长变量<sup>[14]</sup>。对 normalize 预处理后的样本光谱数据进行 SPA 特征波长提取, 如图 3 所示。当特征波长数为 14 时, RMSE 值为 0.3274, 且值达到最小;

所提取的特征波长分别为：931.02、924.64、1399.29、1093.68、950.17、902.3、1380.21、1147.86、895.91、1345.23、1265.68、1332.5、1173.34、982.08 nm，其重要程度依次递减。

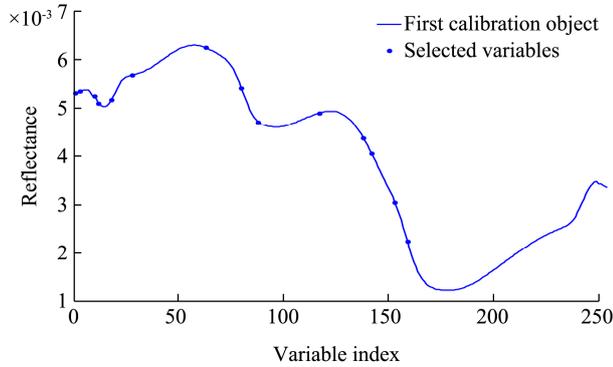


图3 特征波长分布图

Fig.3 Characteristic wavelength distribution map

### 2.3.2 回归系数法 (RC)

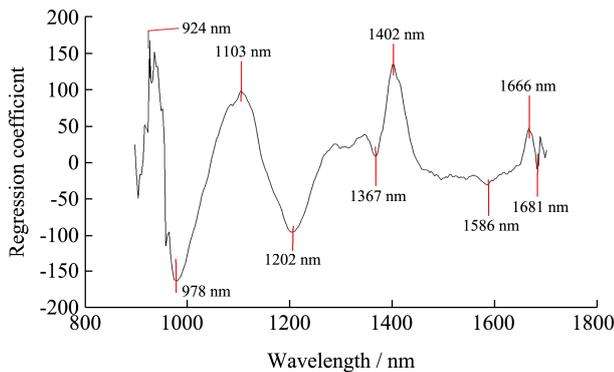


图4 RC提取的特征波长

Fig.4 Key variables selection results of RC

回归系数法<sup>[15]</sup> (RC)：通过对预处理后的样本光谱数据建立 PLS 判别模型，并从模型中提取回归系数。本研究选取了9个特征波长值，分别为924、978、1103、1202、1367、1402、1586、1666、1681 nm，所依据的原则为：将局部极值作为特征波长值，如图5所示。

### 2.3.3 竞争性自适应重加权算法 (CARS)

竞争性自适应重加权算法基于达尔文生物进化论“适者生存”的法则，将蒙特卡罗采样与偏最小二乘回归系数相结合，以实现变量优选。在变量优选过程中，指数衰减函数决定变量剔除数量。变量剔除后，将保留的变量数据利用自适应重加权采样建立 PLS 模型，比较模型的交叉验证均方根误差 RMSECV 值，算出 RMSECV 值最小时对应的 PLS 模型，其对应的子集变量即 CARS 方法优选出的特征变量<sup>[16]</sup>，如图5所示。由图5可以看出，第32次采样所得 RESECV 最小，因此第32次采样保留的变量为所提取的特征波长，分

别为：18、20、65、67、78、79、81、157、158、159、163、230 nm。

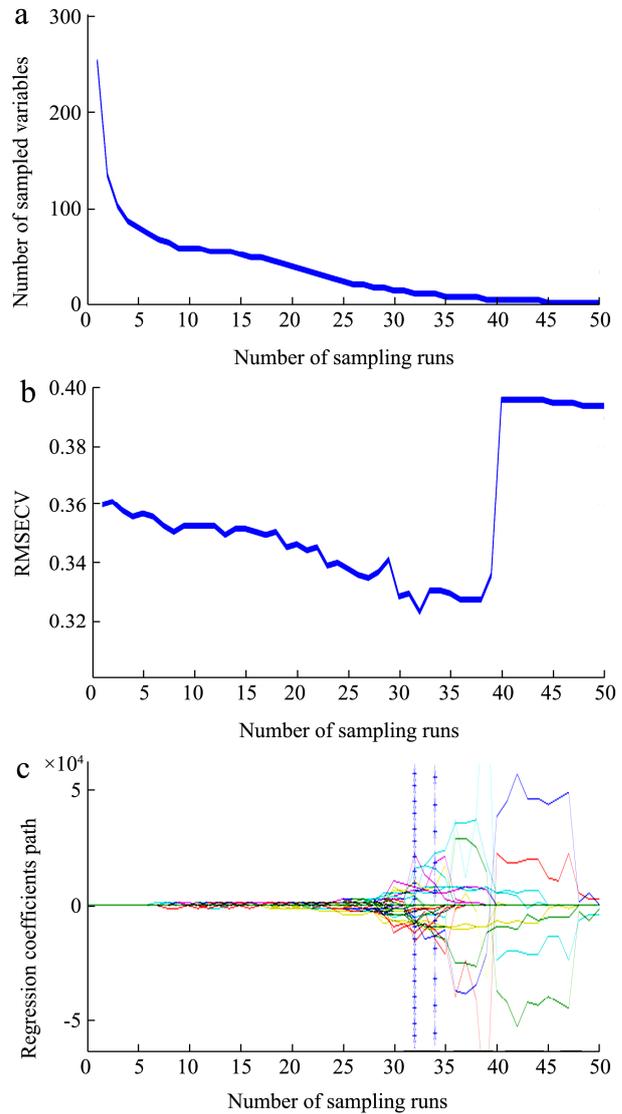
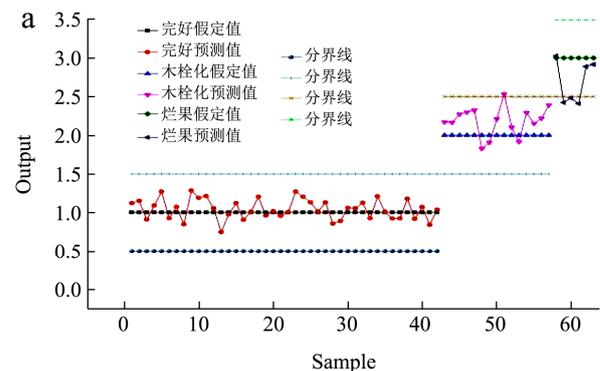


图5 CARS 关键变量选择

Fig.5 Key variables selection results of CARS

### 2.4 基于特征波长的判别模型建立及性能比较

较



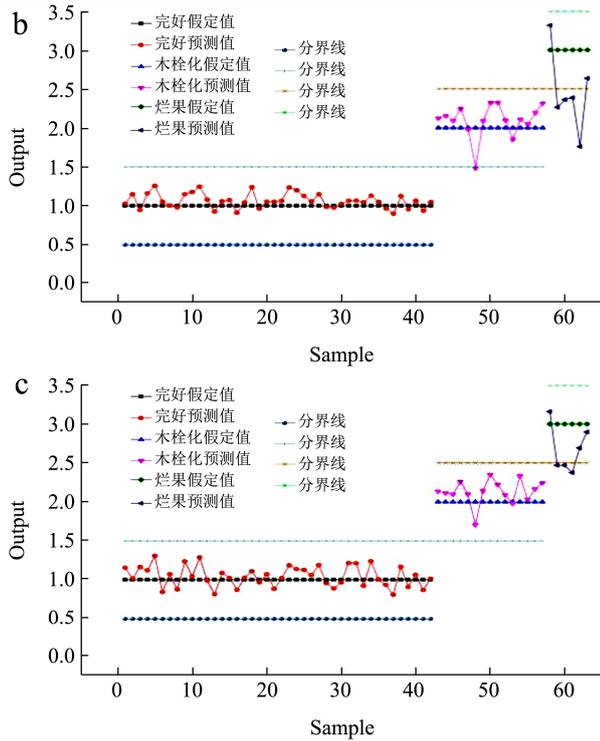


图6 PLS模型预测集鉴别结果

Fig.6 Prediction set of discrimination results of PLS model

注: a: SPA-PLS; b: RC-PLS; c: CARS-PLS.

对3类样本各假定一个值作为判别缺陷类型的依据,将完好样本赋值为1,木栓化样本赋值为2,烂果样本赋值为3。基于上述3种特征波长优选方法,分别以SPA算法提取的14个特征波长、RC提取的9个特征波长以及CARS提取的12个最优变量和类别值作为输入,建立偏最小二乘法(PLS)和多元线性回归(MLR)模型对预测集进行识别分类,如图6和图7。由图可以知道,模型预测结果数值非整数,因此将最大偏离值设定为0.5,即预测值与假定值之差的绝对值在0~0.5范围内时,则将其认定为此类样本,否则

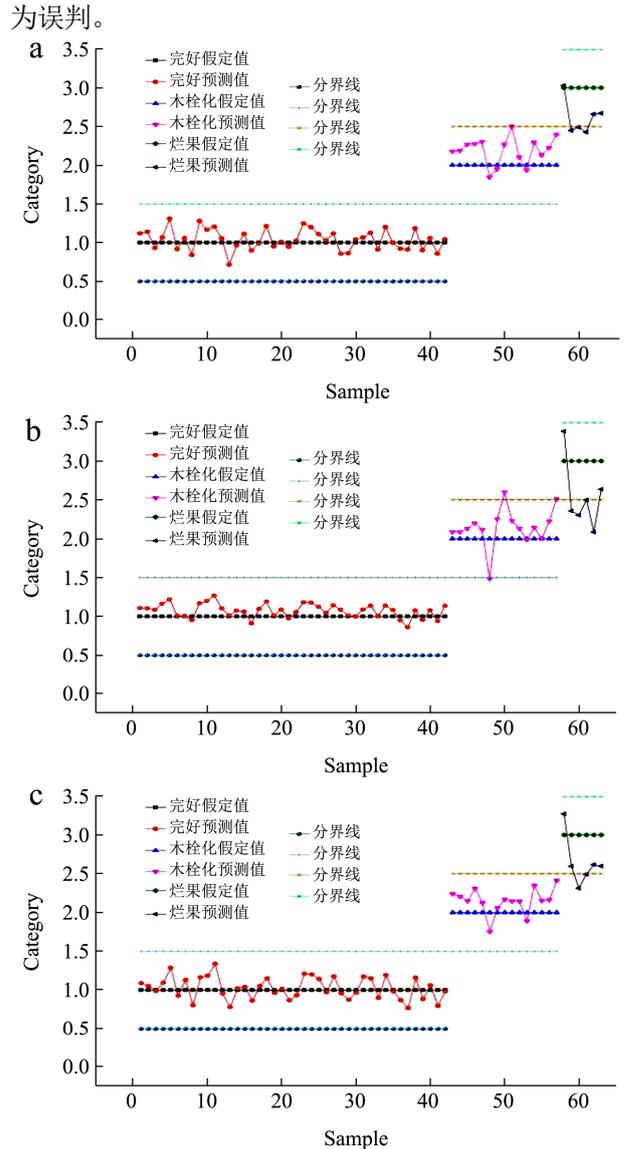


图7 MLR模型预测集鉴别结果

Fig.7 Prediction set of discrimination results of MLR model

注: a: SPA-MLR; b: RC-MLR; c: CARS-MLR.

表2 PLS预测模型与MLR预测模型性能比较

Table 2 Performance comparison of PLS and MLR models

建模方法	校正集		预测集		预测集样本判别准确率		
	$R_c^2$	RMSEC	$R_p^2$	RMSEP	总数	误判数	正确率/%
SPA-PLS	0.91	0.20	0.89	0.22	63	4	93.65
RC-PLS	0.87	0.23	0.85	0.26	63	5	92.06
CARS-PLS	0.92	0.18	0.89	0.21	63	3	95.24
SPA-MLR	0.92	0.20	0.90	0.21	63	4	93.65
RC-MLR	0.90	0.23	0.87	0.24	63	7	88.89
CARS-MLR	0.94	0.19	0.90	0.21	63	2	96.82

表2列出了以3种特征波长提取方法所优选的变量和类别值作为输入时,PLS模型和MLR模型对茄子完好、木栓化和烂果的判别结果。由表2可得,CARS提取的12个最优变量能更好地代替原始光谱信息,原

因是SPA算法和RC在提取特征变量时,虽然降低了原始波长的冗余度,但是同时也剔除了部分有用的信息。因此,基于CARS提取的特征波长所建立的预测模型效果最好。比较模型对样本的预测性能及参数可

得, CARS-MLR 模型的预测精度最高, 其  $R_c^2$  值为 0.94,  $R_p^2$  值为 0.90, 相比于其他模型,  $R_c^2$  和  $R_p^2$  值较优, 同时 RMSEC 和 RMSEP 值都相对较小, 分别为 0.19 和 0.21; 比较模型对预测集样本判别的准确率, CARS-MLR 模型的准确率最高, 为 96.82%, 其次为 CARS-PLS (95.24%), RC-MLR 模型判别准确率最低, 为 88.89%。综合判别模型的决定系数  $R^2$ 、均方根误差 RMSE 以及预测集判别准确率三大指标, 认为 CARS-MLR 模型对茄子外部缺陷的鉴别分类效果最优。

### 3 结论

3.1 基于高光谱技术采集茄子样本的高光谱数据。比较原始光谱数据和经过多种预处理方法预处理后建立 PLS 模型, 结果表明, 经 Normalize 预处理后的 PLS 判别模型效果最佳, 其校正集决定系数  $R_c^2$  为 0.74, 均方根误差 RMSEC 为 0.33; 其预测集决定系数  $R_p^2$  为 0.85, 均方根误差 RMSEP 为 0.26。

3.2 采用 SPA、RC 和 CARS 分别对 Normalize 预处理后的光谱数据提取特征波长, 基于特征波长分别建立 PLS 和 MLR 模型。比较多种模型可知, CARS-MLR 模型效果最优, 其校正集决定系数  $R_c^2$  为 0.94, 预测集决定系数  $R_p^2$  为 0.90, RMSEC 和 RMSEP 分别为 0.19 和 0.21, 预测集判别准确率达到 96.82%, 较好地实现了茄子外部缺陷的检测。

### 参考文献

- [1] 范林林,王清,左进华,等.外源 NO 处理对茄子贮藏品质的影响[J].中国食品学报,2017,17(1):186-192  
FAN Lin-lin, WANG Qing, ZUO Jin-hua, et al. The effect of exogenous NO treatment on eggplant quality during storage [J]. Journal of Chinese Institute of Food Science and Technology, 2017, 17(1): 186-192
- [2] 张婷婷,姚文思,朱惠文,等.冷激处理减轻茄子冷害与活性氧代谢的关系[J].食品科学,2018,39(23):205-211  
ZHANG Ting-ting, YAO Wen-si, ZHU Hui-wen, et al. Effect of cold shock treatment on reducing chilling injury in eggplant and its relationship with reactive oxygen species metabolism [J]. Food Science, 2018, 39(23): 205-211
- [3] 刘爽,谭鑫,刘成玉,等.高光谱数据处理算法的小麦赤霉病籽粒识别[J].光谱学与光谱分析,2019,39(11):3540-3546  
LIU Shuang, TAN Xin, LIU Cheng-yu, et al. Recognition of fusariumhead blight wheat grain based on hyperspectral data processing algorithm [J]. Spectroscopy and Spectral Analysis, 2019, 39(11): 3540-3546
- [4] 刘立新,何迪,李梦珠,等.基于高光谱技术与机器学习的新疆红枣品种鉴别[J/OL].中国激光:1-13[2020-11-30].  
http://kns.cnki.net/kcms/detail/31.1339.TN.20200713.1747.016.html  
LIU Li-xin, HE Di, LI Meng-zhu, et al. Identification of Xinjiang jujube varieties based on hyperspectral technique and machine learning [J/OL]. Chinese Journal of Lasers: 1-13[2020-11-30].  
http://kns.cnki.net/kcms/detail/31.1339.TN.20200713.1747.016.html
- [5] A Del Fiore, M Reverberi, A Ricelli, et al. Early detection of toxigenic fungi on maize by hyperspectral imaging analysis [J]. International Journal of Food Microbiology, 2010, 144(1): 64-71
- [6] A Folch-Fortuny, J M Prats-Montalbán, S Cubero, et al. VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits [J]. International Journal of Food Microbiology, 2016, 156: 241-248
- [7] Ye Sun, Hui Xiao, Sicong Tu, et al. Detecting decayed peach using a rotating hyperspectral imaging testbed [J]. LWT-Food Science and Technology, 2018, 87: 326-332
- [8] 李鸿强,孙红,李民赞.基于高光谱的马铃薯微型种薯分类检测[J].分析测试学报,2020,39(11):1421-1426  
LI Hong-qiang, SUN Hong, LI Min-zan. Classification detection of potato micro seed potato based on hyperspectral [J]. Journal of Instrumental Analysis, 2020, 39(11): 1421-1426
- [9] Yu-Hui F Yeh, Wei-Chang Chung, Jui-Yu Liao, et al. A comparison of machine learning methods on hyperspectral plant disease assessments [J]. IFAC Proceedings Volumes, 2013, 46(4): 361-365
- [10] Yubin Lan, Zixiao Huang, Xiaoling Deng, et al. Comparison of machine learning methods for citrus greening detection on UAV multispectral images [J]. Computers and Electronics in Agriculture, 2020, 171
- [11] Piotr Baranowski, Wojciech Mazurek, Joanna Wozniak, et al. Detection of early bruises in apples using hyperspectral data and thermal imaging [J]. Journal of Food Engineering, 2012, 110(3): 345-355

(下转第 178 页)