

殷海, 李斌, 刘燕德, 等. 基于反射率、吸光度和 Kubelka-Munk 光谱数据的黄桃早期损伤程度检测[J]. 华中农业大学学报, 2023, 42(3): 220-229. DOI: 10.13300/j.cnki.hnlkxb.2023.03.026

基于反射率、吸光度和 Kubelka-Munk 光谱数据的 黄桃早期损伤程度检测

殷海, 李斌, 刘燕德, 张烽, 苏成涛, 欧阳爱国

华东交通大学机电与车辆工程学院/水果智能光电检测技术与设备国家与地方联合工程研究中心, 南昌 330013

摘要 为提高黄桃损伤程度无损检测识别的准确率, 采集健康和不同损伤程度黄桃(*Amygdalus persica*)的反射光谱(R)、吸收光谱(A)、Kubelka-Munk 光谱(K-M), 并基于反射光谱、吸收光谱、Kubelka-Munk 光谱等原始光谱和 RAW、BOC、DT、SG、SNV 等预处理方法后的光谱建立偏最小二乘判别分析(partial least squares discriminant analysis, PLS-DA)、极限梯度提升(extreme gradient boosting, XGBoost)和随机森林(random forest, RF)模型, 比较3种模型检测效果, 选出正确率较高模型并构建其特征波长下的模型, 并对结果再次进行比较。结果显示, 基于3种原始光谱和SG预处理后光谱的RF模型判别效果较优, 整体准确率均达到了90.00%以上。利用竞争性自适应重加权(CARS)和无信息变量消除(UVE)算法对3种原始光谱和SG预处理后的光谱进行波长筛选, 并再次建立RF模型。结果显示, A-RAW-CARS-RF模型和K-M-SG-CARS-RF模型相比于全光谱下的RF模型判别效果得到了改善, 并且在基于特征波长建立的RF模型中, A-RAW-CARS-RF模型的判别效果最好, 整体准确率达到97.12%, 对4个子类别的误判数分别为0、1、1、1。

关键词 黄桃; 损伤程度; 反射光谱; 吸收光谱; Kubelka-Munk 光谱; 无损检测; 高光谱检测; 表型检测

中图分类号 T255 **文献标识码** A **文章编号** 1000-2421(2023)03-0220-10

水果类果实在采摘、挑选、分级和包装等处理过程中会发生机械损伤, 进而引起果实生理代谢紊乱和微生物感染, 导致果实加速老化腐烂, 造成果实营养严重流失^[1]。黄桃(*Amygdalus persica*)富含多种微量元素, 具有预防贫血等疾病和美容养颜的功能^[2], 但黄桃质地较软, 不耐储存^[3]。对损伤程度的客观定量描述不仅是生产者和消费者评价黄桃品质的重要依据, 而且对改进黄桃采后处理和贮藏操作的研究具有重要意义^[4]。因此, 对黄桃的损伤程度进行识别和量化是非常必要的。

机器视觉、近红外光谱、热成像、结构光以及高光谱等技术已经被广泛应用于水果碰撞损伤方面检测^[5-7], 其中高光谱技术因其具有快速、无损等优势, 成为苹果、柑橘等水果缺陷检测识别的热点技术之一^[8-10]。Zhang等^[11]利用高光谱成像技术采集了324张受损苹果的高光谱图像, 在多元散射校正(MSC)下的自适应提升模型(AdaBoost)准确率达到

97.63%。Liu等^[12]利用高光谱技术得到健康、碰伤和受感染草莓的光谱数据, 基于全波长下的支持向量机(SVM)模型的校准集和预测集准确率分别达到96.91%和92.59%。Fu等^[13]基于支持向量机模型(SVM)在损伤后15 min可识别出不同损伤程度的梨, 准确率为93.33%。Tan等^[14]基于网格搜索参数优化的支持向量机(GS-SVM)与逐次投影算法(SPA)相结合的方法检测不同碰伤程度的苹果, 准确率达到95.00%。Li等^[15]基于图谱融合建立最小二乘支持向量机(LS-SVM)模型, 识别不同损伤程度枇杷, 其分类准确率为91.3%。以上研究都是基于反射光谱进行建模, 表明利用高光谱技术检测水果不同碰伤程度具有现实可行性, 但不意味着适用于对所有水果的检测, 存在一定的局限性。目前鲜见利用其他类型光谱区分黄桃碰伤程度的报道。

因此, 本研究将基于反射光谱、吸收光谱、Kubelka-Munk 光谱(K-M)以及5种预处理方法对3种原

收稿日期: 2022-09-04

基金项目: 国家自然科学基金项目(12103019); 国家科技奖后备项目培育计划(20192AEI91007)

殷海, E-mail: yh19970912@163.com

通信作者: 欧阳爱国, E-mail: ouyang1968711@163.com

始光谱预处理后的光谱建立判别分析模型,探究基于3种原始光谱数据集下对不同损伤程度的黄桃进行鉴定和分类的准确性,旨在为后续高光谱技术在黄桃碰撞损伤检测中的应用提供技术支持。

1 材料与方 法

1.1 样品制备

试验所用的黄桃(*Amygdalus persica*)来自于山东省临沂市。试验前筛选出420个新鲜无损且成熟度、硬度和大小尺寸相近的黄桃样本,并逐个编号后

分为对照组(104个)和试验组(316个)。试验前将编号的黄桃储存在室温25℃环境中12 h,使样品温度与室温保持一致。将直径30 mm、质量为100 g的铁球固定在落球冲击试验机上,分别从0.6、1.2、1.8 m等3个高度释放,冲击黄桃造成其损伤,并定义从0.6、1.2、1.8 m释放造成的损伤分别为Ⅰ级、Ⅱ级和Ⅲ级损伤^[16]。试验装置及碰伤样品如图1所示。黄桃样本受到冲击损伤后均储存在室温为24℃、湿度为42%的黑暗条件下,等待数据采集。

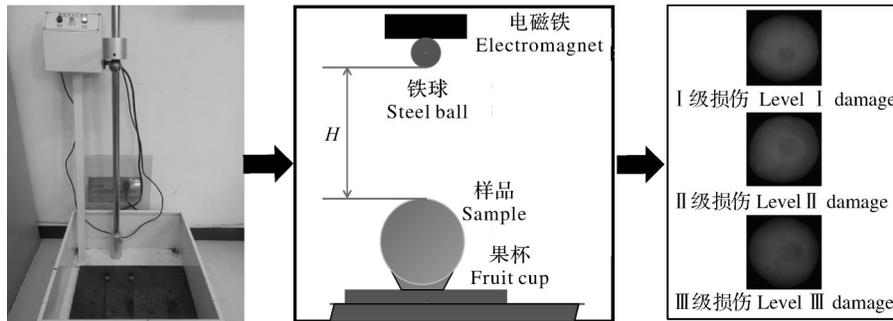


图1 落球冲击试验机及碰伤样品

Fig.1 Falling ball impact tester and bruised sample

1.2 高光谱图像采集

本研究试验数据是通过盖亚(GaiaSorter)高光谱分选仪采集的,其组成结构示意图和其采集的三

维数据立方体如图2所示。系统硬件的组成主要包括成像光谱仪、4盏卤素灯(20 W)、位移平台和步进电机。软件主要采用SpectralView。

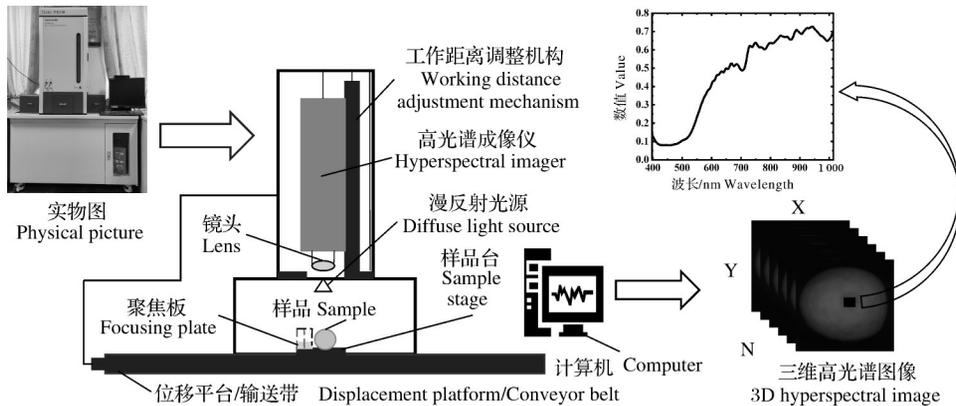


图2 高光谱系统示意图及三维数据立方体

Fig.2 Schematic diagram of the hyperspectral system and 3D data cube

在图像采集之前,为防止无关因素对试验结果造成干扰,需将高光谱系统打开预热0.5 h。设置的参数为:相机曝光时间6.0 ms;光谱分辨率3.5 nm,光谱范围397.5~1 014.0 nm;位移平台移动速度2.5 m/s。操作步骤:(1)将黄桃样品放在位移平台上;(2)打开位移平台,并点击保存按钮,采集高光谱图像;(3)通过计算机软件进行记录,最终形成1个包含了影像信息和光谱信息的三维数据立方体。

1.3 光谱提取

感兴趣区域为大小约为1 600个像素点的圆形区域,通过ENVI 4.5软件进行提取,碰伤黄桃样品选择正对镜头的碰伤区域,健康黄桃样品则选择位于赤道位置且正对着镜头的区域。由于CCD相机中存在暗电流,导致一些光强度较低的波段出现大量的噪声。因此,在进一步的数据处理和分析之前,高光谱图像需用白色和暗色的参照物进行校正^[17]。在本

研究中,高光谱成像系统的图像采用公式(1)进行校正。

$$R = \frac{R_0 - B}{W - B} \times 100 \quad (1)$$

式(1)中, R 、 R_0 、 B 和 W 分别为校正后的图像、原始图像、黑色参考和白色参考。

从高光谱图像中提取的光谱数据是反射光谱,通过公式(2)和(3)将反射光谱分别转换成吸收度光谱和K-M光谱。

$$A = -\lg R \quad (2)$$

$$K-M = \frac{(1-R)^2}{2R} \quad (3)$$

通过公式转换得到了3种不同类型的光谱数据集,即 R 、 A 和K-M,并用于判别黄桃的损伤程度。

1.4 数据处理

1)光谱数据预处理。高光谱成像系统包括来自各种波长的反射数据,所获信息的复杂性应在不同的分析阶段加以解决。同时,光谱数据包含多种无用信息,本研究采用基线偏移校正(BOC)、去趋势(DT)、卷积平滑(SG)、标准正态变量变换(SNV)和多元散射校正(MSC)对光谱数据进行处理,然后比较不同预处理下模型的判别效果。

2)特征波长的筛选。利用高光谱数据进行分类研究时,去除与样本不相关或冗余数据的特征具有

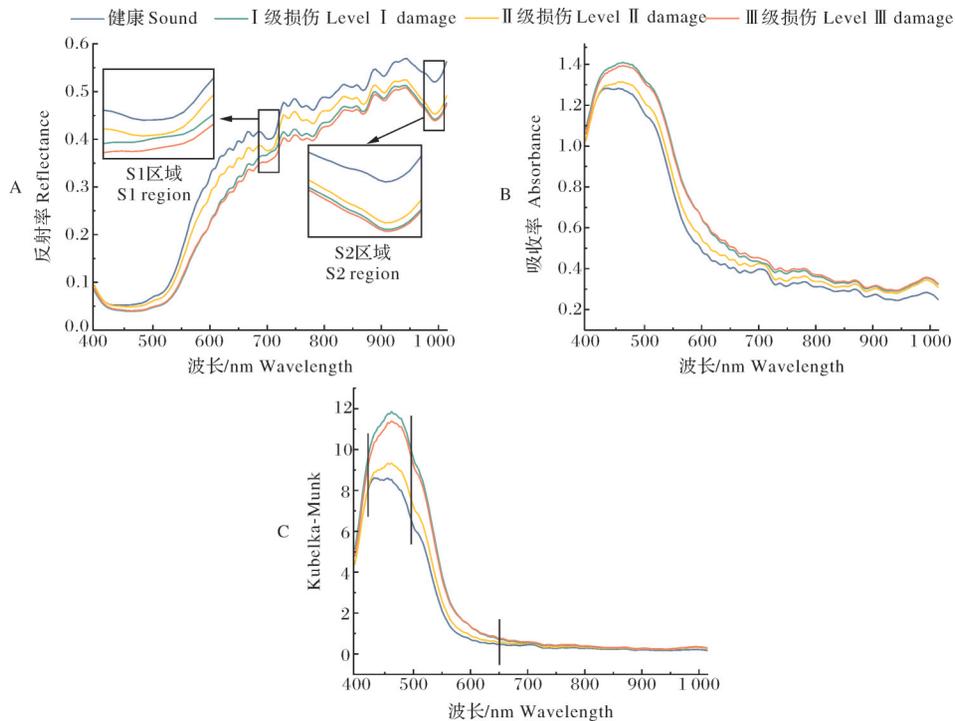
重要意义,本研究采用竞争性自适应重加权(CARS)算法和无信息变量消除(UVE)算法对光谱数据进行变量的挑选。然后比较在这2种波长选择算法下模型的判别效果。

3)模型的建立与评估。PLS-DA作为经典的判别分析方法,当不同类别之间差别较大或同类别之间差别较小时,PLS-DA可以很好对其区分^[18]。极限梯度提升(extreme gradient boosting, XGBoost)算法相比于梯度提升决策树(GBDT)算法, XGBoost在损失函数、正则化(控制模型的复杂度)、切分点查找和并行化设计等方面进行了改进^[19]。随机森林(random forest, RF)是一种有监督学习算法,通过组合多个弱分类器,最终结果通过投票得到,并且使得整个模型具有较高的精确度和泛化能力,同时也有很好的稳定性^[20]。本研究基于反射光谱(R)、吸收光谱(A)和Kubelka-Munk光谱(K-M)分别建立PLS-DA、XGBoost和RF模型,用来判别不同损伤程度的黄桃,模型的判别效果通过预测集的准确率进行评价。

2 结果与分析

2.1 样品的光谱分析

图3显示了健康黄桃和不同损伤程度黄桃的反射光谱、吸收光谱和Kubelka-Munk光谱曲线,从图



A:反射光谱曲线 Reflection spectral curve; B:吸收光谱曲线 Absorption spectral curve; C:K-M光谱曲线 Kubelka-Munk spectral curve.

图3 健康黄桃和不同损伤程度黄桃的平均光谱曲线

Fig.3 Average spectral curve of healthy yellow peach and yellow peach with different damage degrees

3A可知,健康黄桃与不同损伤程度黄桃的光谱曲线趋势相近,健康黄桃的光谱反射强度高于损伤黄桃,主要原因是细胞膜破裂,该结果与健康受损苹果、草莓和梨的光谱一致^[21]。在S1区域(685.0~720.0 nm)和S2区域(950.0~1 010.0 nm)处存在2个明显的吸收谷,其中S1区域的吸收谷随着损伤程度的增加逐渐消失,这可能与叶绿素a有关,损伤程度的增加,致使样品中叶绿素的降解^[12]。S2区域的吸收谷主要是水和糖类成分的O-H键振动引起的,随着损伤程度的增加,光谱反射率逐渐降低,这是因为冲击损伤对黄桃果肉细胞造成破坏,增加了果肉中的羟基(-OH)数量^[22]。从图3B、C可知,健康黄桃的吸收光谱和K-M光谱强度都低于损伤黄桃。对于平均吸收光谱曲线,Ⅱ级损伤与Ⅲ级损伤之间的光谱吸收强度较为接近,其他光谱曲线之间重叠较少。对于平均K-M光谱曲线,在397.5~425.0 nm和650.0~1 014.0 nm光谱相互重叠,在425.0~500.0

nm几乎没有光谱重叠。

2.2 模型的建立与分析

运用Kennard-stone(KS)算法分别将反射光谱、吸收光谱和K-M光谱3种光谱数据集均按照3:1的比例分成建模集和预测集。然后,分别基于3种原始光谱和各种预处理后的光谱建立PLS-DA、XGBoost和RF模型,各种模型的预测集整体准确率如表1所示。由表1可知,基于反射光谱、吸收光谱和K-M光谱以及各种预处理后光谱建立的XGBoost和RF模型的判别效果均优于PLS-DA模型,并且所有PLS-DA模型的总体准确率均未超过80.00%,表明PLS-DA模型不适合本研究中黄桃损伤程度的分类。对于XGBoost和RF模型,总体准确率在70.00%以上,并且RF模型的判别效果大多优于XGBoost模型,甚至部分RF模型的总体准确率超过了90.00%,表明RF模型适合于本研究中黄桃损伤程度的分类。

表1 基于反射光谱、吸收光谱和K-M光谱及各种预处理后光谱建立的模型预测集准确率

Table 1 Prediction results of models based on reflection, absorption and K-M spectra and various pretreated spectra

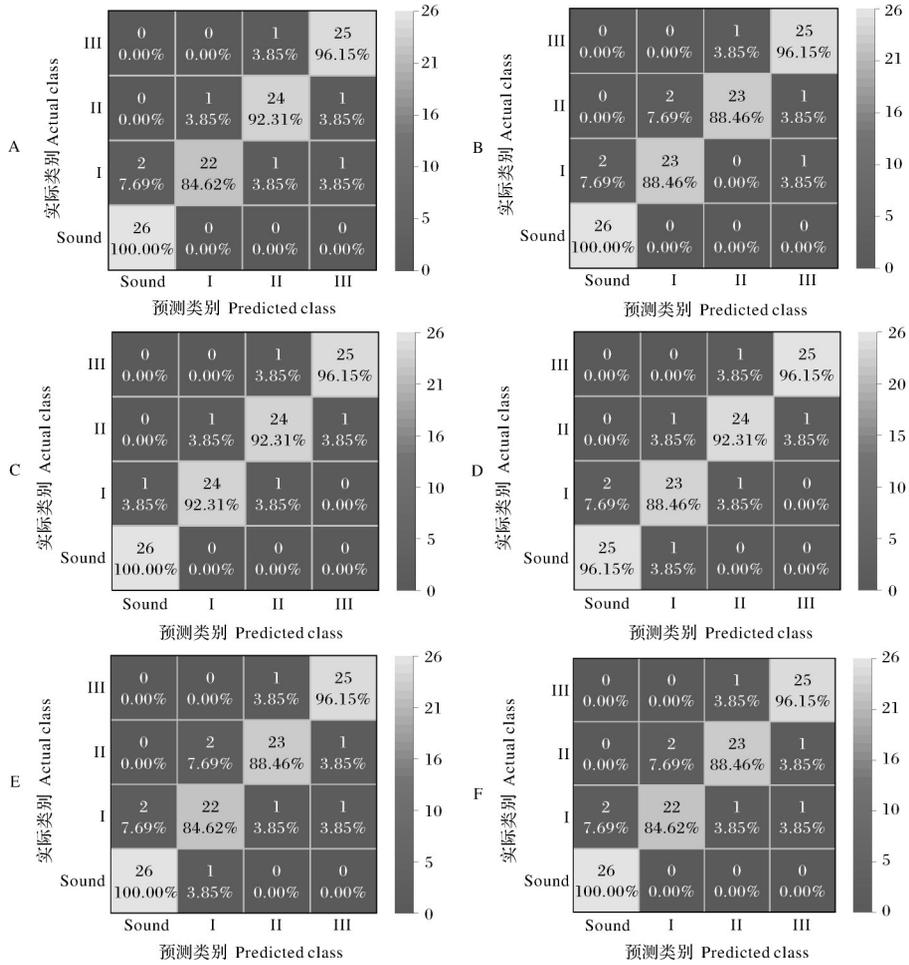
光谱类型 Spectral type	模型 Model	RAW	BOC	DT	SG	SNV	%
反射光谱 Reflection spectrum	PLS-DA	76.92	76.92	72.12	76.92	77.88	
	XGBoost	84.62	77.88	81.73	84.62	85.58	
	RF	93.27	84.62	81.73	93.27	84.62	
吸收光谱 Absorption spectrum	PLS-DA	68.27	71.15	66.35	69.23	74.04	
	XGBoost	84.62	76.92	81.73	86.54	77.88	
	RF	95.19	85.58	84.62	93.27	84.62	
Kubelka-Munk 光谱 Kubelka-Munk spectrum	PLS-DA	61.54	56.73	57.69	66.35	62.50	
	XGBoost	84.62	74.04	73.08	85.58	80.77	
	RF	92.31	76.92	78.85	92.31	84.62	

由表1可知,在基于反射光谱(R)所建立的模型中,R-RAW-RF和R-SG-RF模型判别效果最优,总体准确率均为93.27%。图4为RF模型(总体准确率 $\geq 90\%$)对每个子类别的判别结果。由图4A和图4B可知,R-RAW-RF和R-SG-RF模型对4个子类别的误判数分别为0、4、2、1和0、3、3、1,2个模型对健康黄桃的判别准确率均达到了100.00%。基于吸收光谱所建的模型中,A-RAW-RF模型是基于3种光谱下所建立模型中判别效果最优,其总体准确率为95.19%。A-SG-RF模型虽然判别效果略差于A-RAW-RF模型,但总体准确率也达到了93.27%。由图4C、D可知,A-RAW-RF和A-SG-RF模型对4个子类别的误判数分别为0、2、2、1和1、3、2、1。基于K-M光谱所建立的模型中,K-M-RAW-RF和K-M-

SG-RF模型判别效果最优,总体准确率均为92.31%。由图4E、F可知,K-M-RAW-RF和K-M-SG-RF模型对4个子类别的误判数均为0、4、3、1,并且2个模型对健康黄桃的判别准确率均达到了100.00%。

2.3 特征波长筛选

全光谱的使用通常会引入噪声,导致过度拟合、非线性以及效率或准确性的损失^[23],并且通过对基于全光谱建立的模型进行对比,发现基于反射光谱、吸收光谱和K-M光谱等原始光谱和SG预处理后光谱的RF模型总体准确率都达到了90.00%以上。本研究中CARS和UVE算法分别用于3种原始光谱和SG预处理后的光谱(397.5~1 014.0 nm,176个波段)。以原始吸收光谱(A-RAW)为例,介绍CARA



Sound:健康样品 Health sample; I、II和III: I、II和III级损伤 I, II and III level damage yellow peaches; A:反射光谱-原始光谱-随机森林 R-RAW-RF; B:反射光谱-卷积平滑-随机森林 R-SG-RF; C:吸收光谱-原始光谱-随机森林 A-RAW-RF; D:吸收光谱-卷积平滑-随机森林 A-SG-RF; E:Kubelka-Munk 光谱-原始光谱-随机森林 K-M-RAW-RF; F: Kubelka-Munk 光谱-卷积平滑-随机森林 K-M-SG-RF.

图4 不同模型预测结果的混淆矩阵

Fig. 4 Confusion matrix for the prediction results of different models

和 UVE 算法筛选变量的过程。图 5 展示了 CARS 算法筛选变量的过程,从图 5 可以看出,采样次数 0~100 过程中,变量的数量呈指数下降,同时交叉验证均方根误差 (RMSECV) 先降后升,这是因为 CARS 在步骤结束时消除了相关性高的变量。特征波长由采样过程中的最低 RMSECV 确定^[24]。如图 5 所示,最小的 RMSECV 对应采样次数发生在 23 次。

CARS 算法对 3 种原始光谱和 SG 预处理后光谱挑选变量的位置如图 6 所示。从图 6 可知,R-RAW 和 R-SG 光谱分别被挑选出 35 和 26 个变量,占总波长的 19.89% 和 14.77%,A-RAW 和 A-SG 光谱分别被挑选出 65 和 52 个变量,占总波长的 36.93% 和 29.55%,K-M-RAW 和 K-M-SG 光谱分别被挑选出 41 和 36 个变量,占总波长的 23.30% 和 20.45%。

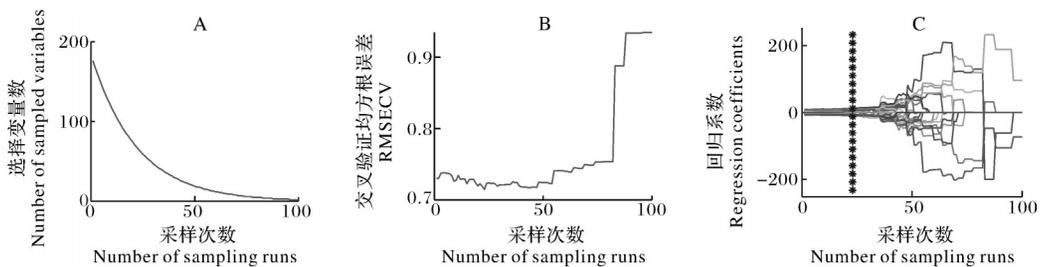
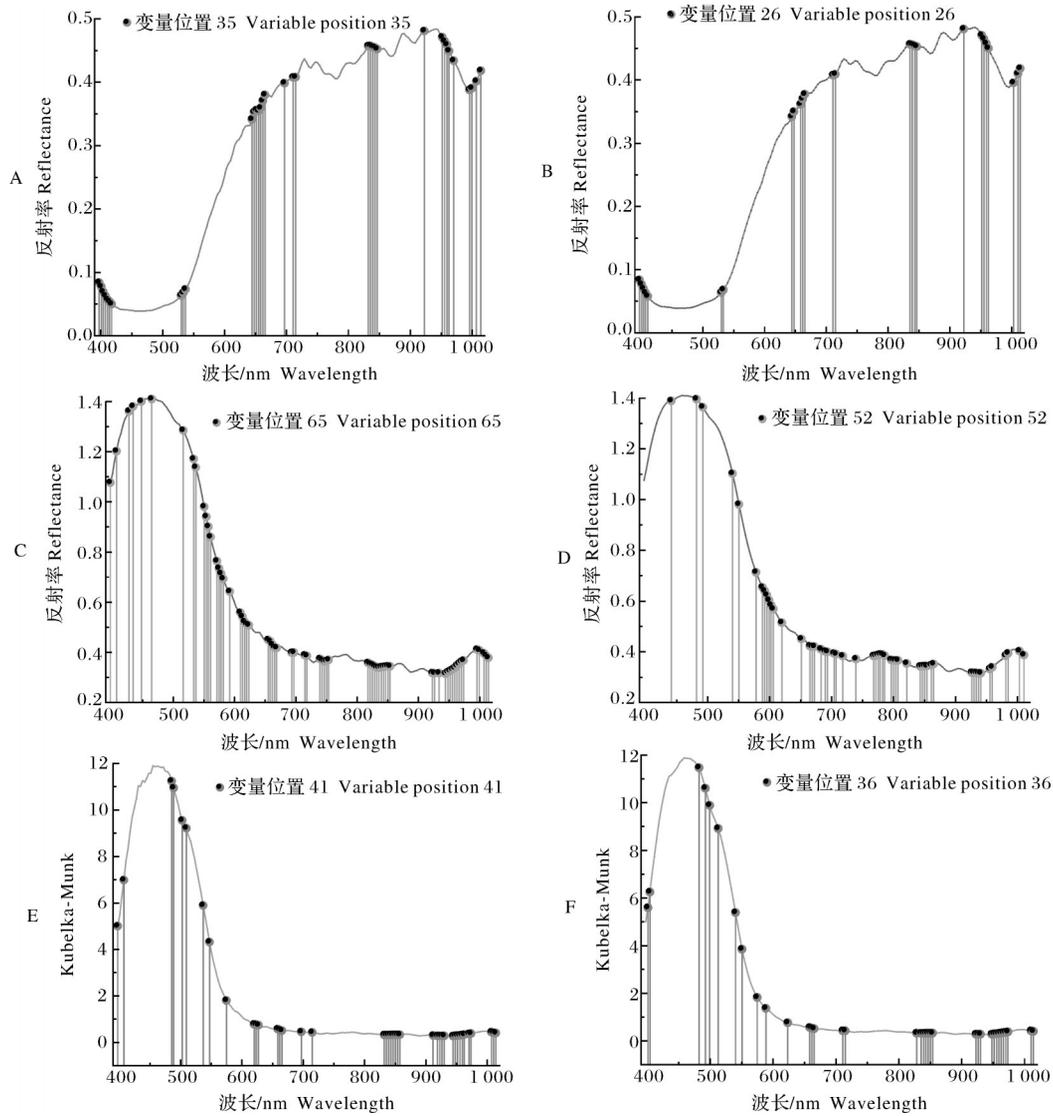


图5 CARS算法筛选变量过程

Fig.5 The process of screening the variables of CARS algorithm



A:反射光谱-原始光谱 R-RAW;B:反射光谱-卷积平滑 R-SG;C:吸收光谱-原始光谱 A-RAW;D:吸收光谱-卷积平滑 A-SG;E:Kubelka-Munk 光谱-原始光谱 K-M-RAW;F:Kubelka-Munk 光谱-卷积平滑 K-M-SG.

图 6 CARS 算法挑选波长结果

Fig.6 The results of CARS algorithm to select wavelength

图 7 展示了 UVE 算法筛选变量的过程,随机变量的个数设置为 176 个。从图 7 可以看出,2 条横向虚线对应的稳定性系数分别为 20 和 -20,即筛选变量的阈值,在 2 条横向虚线之间的变量是无效的,应被剔除,而 2 条横向虚线之外的变量则是挑选出来的特征变量,用于后续模型的建立与分析。 UVE 算法对 3 种原始光谱和 SG 预处理后光谱挑选变量的位置如图 8 所示。从图 8 可知,R-RAW 和 R-SG 光谱分别被挑选出 63 和 94 个变量,占总波长的 35.80% 和 53.41%,A-RAW 和 A-SG 光谱分别被挑选出 85 和 82 个变量,占总波长的 48.30% 和 46.59%,K-M-RAW 和 K-M-SG 光谱分别被挑选出 66 和 77 个变量,占总波长的 37.50% 和 43.75%。

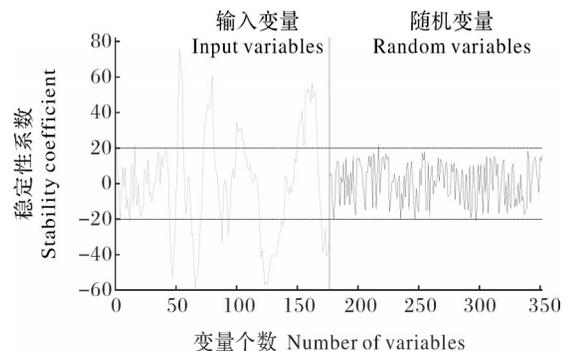
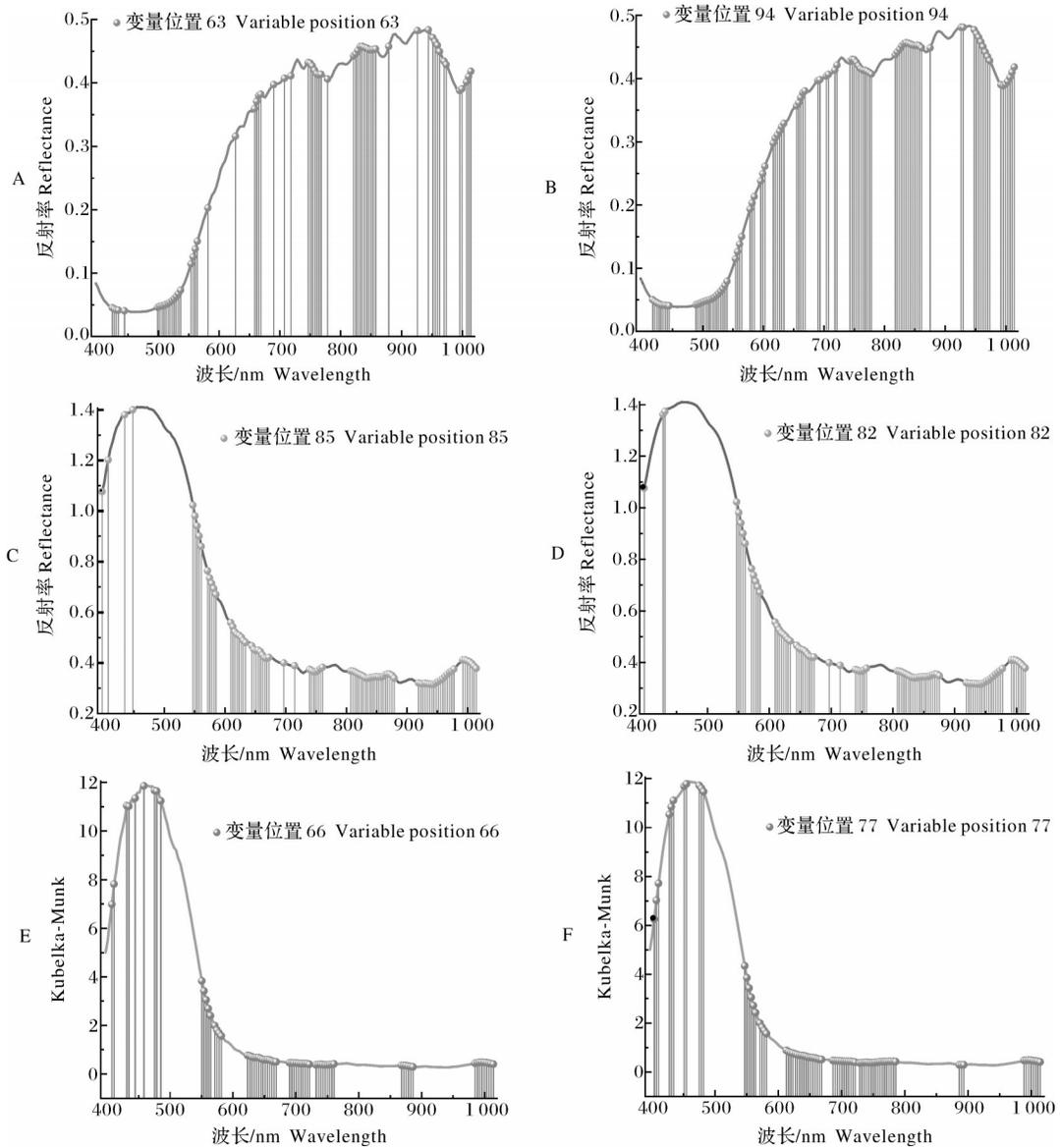


图 7 UVE 算法筛选变量过程

Fig.7 The process of screening variables of UVE algorithm

2.4 特征波长下模型的建立

利用 CARS 和 UVE 算法对 3 种原始光谱及 SG



A: 反射光谱-原始光谱 R-RAW; B: 反射光谱-卷积平滑 R-SG; C: 吸收光谱-原始光谱 A-RAW; D: 吸收光谱-卷积平滑 A-SG; E: Kubelka-Munk 光谱-原始光谱 K-M-RAW; F: Kubelka-Munk 光谱-卷积平滑 K-M-SG.

图8 UVE算法挑选波长结果

Fig.8 The results of UVE algorithm to select wavelength

预处理后光谱进行特征波长筛选,利用筛选后的波长再次建立RF模型。基于特征波长建立的RF模型预测集整体准确率如表2所示。从表1和表2可知,A-RAW-CARS-RF模型和K-M-SG-CARS-RF模型相比于全光谱下的RF模型判别效果得到改善,其整体准确率分别提高1.93和0.96个百分点,分别达到97.12%和93.27%,且在基于特征波长建立的RF模型中,A-RAW-CARS-RF模型整体准确率最高。其他模型相比于全光谱下的RF模型整体准确率保持持平或出现轻微下降。图9为A-RAW-CARS-

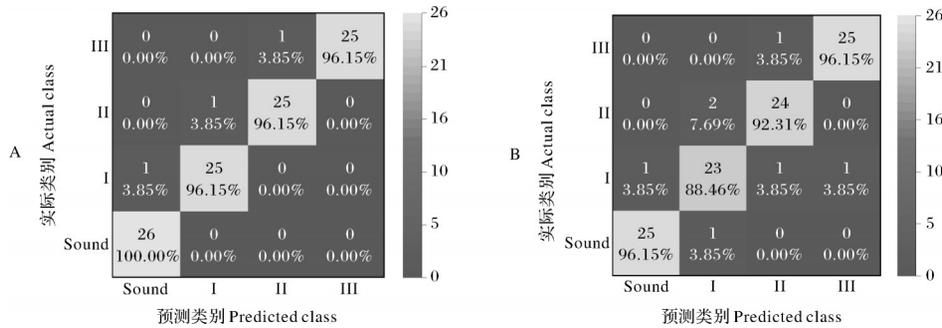
RF模型和K-M-SG-CARS-RF模型对每个子类别的判别结果,从图9可知,2个模型对4个子类别的误判数分别为0、1、1、1和1、3、2、1。结合图4可知,A-RAW-CARS-RF模型对I级损伤和II级损伤的判别相比于A-RAW-RF模型误判数均减少1个,其余子类别不变。K-M-SG-CARS-RF模型对I级碰撞和II级碰撞的判别相比于K-M-SG-RF模型误判数均减少1个,对III级碰撞的判别不变,但对健康黄桃的判别上误判数增加1个,这可能是由于波长筛选不仅去除了无用信息,也可能去除一部分有效信息。

表 2 基于特征波长建立的 RF 模型预测集准确率

Table 2 RF model prediction results based on the characteristic wavelength

%

波长筛选方法 Wavelength screening method	反射光谱 Reflection spectrum		吸收光谱 Absorption spectrum		Kubelka-Munk 光谱 Kubelka-Munk spectrum	
	RAW	SG	RAW	SG	RAW	SG
	CARS	89.42	89.42	97.12	88.46	91.35
UVE	91.35	93.27	91.35	93.27	89.42	92.31



Sound:健康样品 Health sample; I、II 和 III: I、II 和 III 级损伤 I, II and III level damage yellow peaches; A:吸收光谱-原始光谱-竞争性自适应重加权-随机森林 A-RAW-CARS-RF; B:Kubelka-Munk 光谱-卷积平滑-竞争性自适应重加权-随机森林 K-M-SG-CARS-RF.

图 9 不同模型预测结果的混淆矩阵

Fig.9 Confusion matrix for different model prediction results

3 讨 论

客观定量描述水果损伤程度不仅是评估其品质的重要参考,也对改善黄桃采后处理和贮存操作具有重要意义。前人检测水果损伤程度大部分都是基于反射光谱建立模型,文献[13]中对不同损伤程度梨的识别率为 93.33%,文献[14]中对不同擦伤程度苹果的识别率为 95.00%,虽然这些研究都取得了不错的效果,但是其模型可能存在普适性和可移植性不高等问题。

本研究观察并比较了高光谱成像的反射光谱、吸收光谱和 K-M 光谱对黄桃早期损伤程度的检测能力。基于原始光谱和各种预处理后的光谱建立 PLS-DA、XGBoost 和 RF 模型,发现基于 3 种原始光谱和 SG 预处理后光谱的 RF 模型判别效果较优,整体准确率均达到 90.00% 以上。然后,利用 CARS 和 UVE 算法对 3 种原始光谱和 SG 预处理后的光谱进行波段筛选,并再次建立 RF 模型,发现 A-RAW-CARS-RF 模型和 K-M-SG-CARS-RF 模型相比于全光谱下的 RF 模型判别效果得到改善,并且在基于特征波长建立的 RF 模型中,A-RAW-CARS-RF 模型的判别效果最好,整体准确率达到 97.12%,对 4 个子类别的误判数分别为 0、1、1、1。综上所述,研究表明,基于吸收光谱检测黄桃早期损伤程度具有可

行性,同时,利用 3 种光谱同时检测水果的损伤程度,可大大提高模型的普适性和可移植性。

参考文献 References

[1] TAN W Y, SUN L J, YANG F, et al. Study on bruising degree classification of apples using hyperspectral imaging and GS-SVM [J]. Optik, 2018, 154: 581-592.

[2] 黄余年, 张维, 张群, 等. 采前套袋与未套袋处理对黄桃采后贮藏品质的影响 [J]. 中国食品学报, 2021, 21 (6): 231-242. HUANG Y N, ZHANG W, ZHANG Q, et al. Effects of pre-harvest bagging and non-bagging treatment on postharvest storage quality of yellow-flesh peach [J]. Journal of Chinese Institute of Food Science and Technology, 2021, 21 (6): 231-242 (in Chinese with English abstract).

[3] 于笑颜, 吕健, 毕金峰, 等. 基于果胶特性改变的罐藏黄桃质构软化机制 [J]. 食品科学, 2020, 41 (19): 45-52. YU X Y, LÜ J, BI J F, et al. Mechanism for texture softening of canned yellow peaches based on modification of pectin characteristics [J]. Food science, 2020, 41 (19): 45-52 (in Chinese with English abstract).

[4] LI Z G, THOMAS C. Quantitative evaluation of mechanical damage to fresh fruits [J]. Trends in food science & technology, 2014, 35(2): 138-150.

[5] 安小松, 宋竹平, 梁千月, 等. 基于 CNN-Transformer 的视觉缺陷柑橘分选方法 [J]. 华中农业大学学报, 2022, 41(4): 158-169. AN X S, SONG Z P, LIANG Q Y, et al. A CNN-Transformer-based method for sorting citrus with visual defects [J]. Journal of

- Huazhong Agricultural University, 2022, 41(4):158-169(in Chinese with English abstract).
- [6] 刘燕德,黎丽莎,李斌,等.多品种苹果可溶性固形物近红外无损检测通用模型研究[J].华中农业大学学报,2022,41(2):237-244.LIU Y D, LI L S, LI B, et al. General near-infrared model of soluble solids content in multi-variety apples[J]. Journal of Huazhong Agricultural University, 2022, 41(2):237-244 (in Chinese with English abstract).
- [7] LU Y, SAEYS W, KIM M, et al. Hyperspectral imaging technology for quality and safety evaluation of horticultural products: a review and celebration of the past 20-year progress[J/OL]. Postharvest biology and technology, 2020, 170: 111318 [2022-09-04]. <http://doi.org/10.1016/j.postharvbio.2020.111318>.
- [8] EKRAMIRAD N, KHALED A Y, DOYLE L E, et al. Non-destructive detection of codling moth infestation in apples using pixel-based NIR hyperspectral imaging with machine learning and feature selection[J/OL]. Foods (Basel, Switzerland), 2021, 11(1):8[2022-09-04]. <https://doi.org/10.3390/foods11010008>.
- [9] TIAN X, ZHANG C, LI J B, et al. Detection of early decay on citrus using LW-NIR hyperspectral reflectance imaging coupled with two-band ratio and improved watershed segmentation algorithm[J/OL]. Food chemistry, 2021, 360: 130077 [2022-09-04]. <https://doi.org/10.1016/j.foodchem.2021.130077>.
- [10] TANG Y, GAO S J, ZHUANG J J, et al. Apple bruise grading using piecewise nonlinear curve fitting for hyperspectral imaging data[J]. IEEE access, 2020, 8:147494-147506.
- [11] ZHANG M, LI G H. Visual detection of apple bruises using AdaBoost algorithm and hyperspectral imaging[J]. International journal of food properties, 2018, 21(1):1598-1607.
- [12] LIU Q, SUN K, PENG J, et al. Identification of bruise and fungi contamination in strawberries using hyperspectral imaging technology and multivariate analysis[J]. Food analytical methods, 2018, 11(5):1518-1527.
- [13] FU X P, WANG M Y. Detection of early bruises on pears using fluorescence hyperspectral imaging technique[J]. Food analytical methods, 2022, 15(1):115-123.
- [14] TAN W Y, SUN L J, YANG F, et al. The feasibility of early detection and grading of apple bruises using hyperspectral imaging[J/OL]. Journal of chemometrics, 2018, 32(10):e3067 [2022-09-04]. <https://doi.org/10.1002/cem.3067>.
- [15] LI B, HAN Z Y, WANG Q, et al. Study on qualitative impact damage of loquats using hyperspectral technology coupled with texture features[J/OL]. Foods (Basel, Switzerland), 2022, 11(16):2444 [2022-09-04]. <https://doi.org/10.3390/foods11162444>.
- [16] LI B, YIN H, LIU Y D, et al. Study on qualitative impact damage of yellow peaches using the combined hyperspectral and physicochemical indicators method[J/OL]. Journal of molecular structure, 2022, 1265:133407 [2022-09-04]. <https://doi.org/10.1016/j.molstruc.2022.133407>.
- [17] LI B, YIN H, LIU Y D, et al. Detection storage time of mild bruise's yellow peaches using the combined hyperspectral imaging and machine learning method[J/OL]. Journal of analytical science and technology, 2022, 13(1):24 [2022-09-04]. <https://doi.org/10.1186/s40543-022-00334-5>.
- [18] 王晓宾,马泉,杨蕾,等.基于PLS-DA拉曼光谱特征提取的中性笔油墨MLP模式识别[J].激光与光电子学进展,2021,58(1):329-335.WANG X B, MA X, YANG L, et al. Multi-layer perceptron pattern recognition of handwriting ink based on PLS-DA Raman spectral feature extraction[J]. Laser & optoelectronics progress, 2021, 58(1):329-335 (in Chinese with English abstract).
- [19] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 13-17, 2016, San Francisco, California, USA. New York: ACM, 2016:785-794.
- [20] 李欣海. 随机森林模型在分类与回归分析中的应用[J]. 应用昆虫学报, 2013, 50(4):1190-1197. LI X H. Using "random forest" for classification and regression[J]. Chinese journal of applied entomology, 2013, 50(4):1190-1197 (in Chinese with English abstract).
- [21] LEE W H, KIM M S, LEE H, et al. Hyperspectral near-infrared imaging for the detection of physical damages of pear[J]. Journal of food engineering, 2014, 130:1-7.
- [22] PAN T T, CHYNGYZ E, SUN D W, et al. Pathogenetic process monitoring and early detection of pear black spot disease caused by *Alternaria alternata* using hyperspectral imaging[J]. Postharvest biology and technology, 2019, 154:96-104.
- [23] LI X P, JIANG H Z, JIANG X S, et al. Identification of geographical origin of Chinese chestnuts using hyperspectral imaging with 1D-CNN algorithm[J/OL]. Agriculture, 2021, 11(12):1274 [2022-09-04]. <https://doi.org/10.3390/agriculture11121274>.
- [24] SU W H, YANG C, DONG Y H, et al. Hyperspectral imaging and improved feature variable selection for automated determination of deoxynivalenol in various genetic lines of barley kernels for resistance screening[J/OL]. Food chemistry, 2021, 343:128507 [2022-09-04]. <https://doi.org/10.1016/j.foodchem.2020.128507>.

Detection of early damage level of yellow peaches based on reflectance, absorbance and Kubelka-Munk spectral data

YIN Hai, LI Bin, LIU Yande, ZHANG Feng, SU Chengtao, OUYANG Aiguo

*School of Mechatronics & Vehicle Engineering/National and Local Joint Engineering
Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment,
East China Jiaotong University, Nanchang 330013, China*

Abstract Yellow peaches are soft and prone to damage, and the different level of damage can directly affect the end use and sale price of yellow peaches. The reflection (R), absorption (A), and Kubelka-Munk spectra (K-M) of yellow peaches were obtained by using hyperspectral techniques and used to detect the early damage level of yellow peaches. Partial least squares discriminant analysis (PLS-DA), extreme gradient boosting (XGBoost) and random forest (RF) models based on three raw spectra and various pre-treated spectra were established. The results were compared to select the model with higher correctness. The model with its characteristic wavelength was constructed and compared again. The results showed that RF models based on the three raw spectra and SG pretreated spectra were superior in discriminating, with the overall accuracy rates all above 90.00%. The wavelength screening of the raw spectra and SG pretreated spectra was performed with the competitive adaptive reweighting (CARS) and uninformative variable elimination (UVE) algorithms, and the RF models were established again. The results showed that the A-RAW-CARS-RF model and the K-M-SG-CARS-RF model were improved in discriminating compared with the RF model at full spectrum. Among the RF models established based on the characteristic wavelengths, the A-RAW-CARS-RF model had the best discriminating effect with an overall accuracy of 97.12%. The number of misclassifications for the four subcategories were 0, 1, 1, and 1. It is indicated that the feasibility of detecting the early damage level of yellow peaches based on absorption spectroscopy (A). It will provide some theoretical basis for detecting fruit bruise with hyperspectral techniques in the future.

Keywords yellow peach; damage level; reflection spectroscopy; absorption spectroscopy; Kubelka-Munk spectroscopy; nondestructive testing; hyperspectral detection; phenotypic testing

(责任编辑:陆文昌)