

高光谱数据对损伤长枣的检测判别

袁瑞瑞¹, 王 兵², 刘贵珊^{1*}, 何建国¹, 万国玲¹, 樊奈昀¹, 李 月¹, 孙有瑞¹

1. 宁夏大学食品与葡萄酒学院, 宁夏 银川 750021

2. 宁夏大学物理与电子电气工程学院, 宁夏 银川 750021

摘 要 灵武长枣作为宁夏优势特色枣果, 具有重要的经济社会价值和科学研究意义。利用可见近红外(Vis/NIR)高光谱成像系统采集 60 颗完整长枣光谱图像, 然后利用损伤装置对 60 颗完整长枣进行损伤实验, 最终得到 60 颗损伤(内部瘀伤)长枣, 高光谱成像系统采集损伤后五个时间段(损伤后 2, 4, 8, 12 和 24 h)长枣的光谱图像。对采集的长枣光谱图像用 ENVI 软件提取感兴趣(ROI)区域, 并计算完整长枣和每个时间段长枣的平均光谱值。原始光谱利用 Savitzky-Golay 平滑的一阶导数(SG-1)和二阶导数(SG-2)、标准正态变换(SNV)和去趋势(Detrending)、以及 SNV-SG-1、SNV-SG-2、Detrending-SG-1、Detrending-SG-2 算法进行预处理, 原始光谱和预处理光谱建立偏最小二乘判别分析(PLS-DA)分类模型。选择最优的预处理光谱数据, 利用连续投影算法(SPA)、间隔随机蛙跳(IRF)、无信息消除变量(UVE)、变量组合集群分析法(VC-PA)、区间变量迭代空间收缩法(IVISSA)和 IRF-SPA、UVE-SPA、IVISSA-SPA 等算法进行特征变量选择, 对选择的特征变量建立 PLS-DA、线性判别分析(LDA)和支持向量机(SVM)分类判别模型。结果表明, 在原始光谱建立的 PLS-DA 模型中, 模型校正集和预测集准确率分别为 82.96% 和 90%。光谱经过预处理后得到 SNV-SG-2-PLS-DA 为最优分类判别模型, 模型校正集和预测集准确率分别为 91.11% 和 96.67%。在特征变量建立的分类模型中, SNV-SG-2-UVE-PLS-DA 模型校正集和预测集准确率分别为 86.3% 和 94.44%; SNV-SG-2-SPA-LDA 模型校正集和预测集准确率分别为 86.3% 和 83.33%; SNV-SG-2-UVE-SVM 模型校正集和预测集准确率分别为 77.78% 和 71.11%。对于分类模型来说线性分类模型(PLS-DA、LDA)分类结果优于非线性分类模型(SVM)分类结果, 在线性分类模型结果中 PLS-DA 优于 LDA 分类结果, PLS-DA 可以更好的提供分类效果。研究表明, 利用高光谱结合偏最小二乘判别分析分类模型, 可以有效的实现灵武长枣损伤后随时间变化的快速检测, 为灵武长枣在线检测提供理论依据。

关键词 灵武长枣; 高光谱; 偏最小二乘判别分析; 线性判别分析; 支持向量机

中图分类号: TS255 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)09-2879-07

引 言

枣(*Zizyphus jujuba Miller*)在中国已有 4 000 多年的历史, 它主要分布在亚洲的亚热带和热带地区, 已有三千多年的耕种历史。“灵武长枣”(Lingwu long jujube)是宁夏重要的经济林木之一, 并且由于其巨大的生态、社会和经济效益, 也是宁夏农业的主要组成部分^[1-2], 但是灵武长枣在采收、运输等过程中容易受到损伤, 并且损伤后不容易被观察到降低商品价值严重影响经济收益^[3]。因此, 亟需一种无损检测

技术快速有效地检测灵武长枣的内部损伤。

近年来, 高光谱成像系统作为一种快速无损、准确度高且具有高灵敏度的检测系统, 被用于苹果^[4]、梨^[5]、猕猴桃^[6]、草莓^[7]、蓝莓^[8]、桃子^[9]等的瘀伤检测。Zhang^[10]等利用高光谱成像系统结合 AdaBoost 算法对完整苹果和损伤后 5 个时间段(损伤后 1 min、1 天、2 天、3 天、4 天)的苹果进行了分类, 结果表明, 经 MSC(multiplicative scatter correction)和 CFS(correlation-based feature selection)预处理后, 所选波长建立的模型平均精度为 97.63%。Fan^[11]等利用最佳波长结合近红外高光谱反射成像系统对蓝莓内部瘀伤随时间

收稿日期: 2020-09-13, 修订日期: 2021-01-22

基金项目: 宁夏特色果蔬冷链关键技术装备研发与示范项目(2018ZDKJ0182), 国家自然科学基金项目(31560481)资助

作者简介: 袁瑞瑞, 1994 年生, 宁夏大学食品与葡萄酒学院硕士研究生 e-mail: ruiruiyuan123@163.com

王 兵, 1997 年生, 宁夏大学物理与电子电气工程学院硕士研究生 e-mail: wb731618660@163.com

袁瑞瑞, 王 兵: 并列第一作者 * 通讯作者 e-mail: liugs@nxu.edu.cn

的变化进行了检测研究,结果表明,蓝莓在撞击后 30 min、2 h、6 h 和 12 h 的波段比值图像建模分类精度分别为 77.5%, 83.8%, 92.5% 和 95.0%, 以及 CARS-LS-SVM(competitive adaptive reweighted sampling-least squares-support vector machine)模型的验证集中健康和瘀伤蓝莓准确率分别为 93.3% 和 95.9%。Lee^[5]等利用高光谱图像对梨的物理损伤进行了检测研究,结果表明,利用最佳阈值波段比检测结果的准确率为 92%。灵武长枣外部缺陷检测已有相关研究,但是对于灵武长枣内部损伤检测鲜有报道。

故以灵武长枣为研究对象,对完整长枣和损伤后五个时间段(损伤后 2, 4, 8, 12 和 24 h)长枣进行分类判别。利用高光谱成像系统获得高光谱图像,利用 ENVI 软件提取感兴趣(region of interest, ROI)区域,并计算平均光谱值。对原始光谱利用 Savitzky-Golay 平滑的一阶导数(first derivatives, SG-1)和二阶导数(second derivatives, SG-2)、标准正态变换(standard normal variate, SNV)和去趋势(Detrending)、以及 SNV-SG-1、SNV-SG-2、Detrending-SG-1、Detrending-SG-2 组合预处理,并建立 PLS-DA 分类模型;优选最优预处理算法得到的光谱数据,利用连续投影算法(successive projection algorithm, SPA)、间隔随机蛙跳(interval random frog, IRF)、无信息消除变量(uninformative variable elimination, UVE)、变量组合集群分析法(variable combination population analysis, VCPA)、区间变量迭代空间收缩法(interval variable iterative space shrinkage approach, IVISSA)五种算法和 IRF-SPA、UVE-SPA、IVISSA-SPA 三种组合算法进行特征变量选择,特征变量建立偏最小二乘判别分析(partial least squares-discriminant analysis, PLS-DA)、线性判别分析(linear discriminant analysis, LDA)和支持向量机(support vector machine, SVM)分类判别模型。为损伤灵武长枣在线检测提供理论依据。

1 实验部分

1.1 样品采集

从宁夏灵武某果园手工采摘大小颜色一致、完好无损伤的灵武长枣,放入保鲜袋中当天运回实验室,选取 60 个灵武长枣贮藏在(0±2)℃的冰箱备用。

1.2 损伤实验

采用如图 1 所示的损伤装置获得损伤灵武长枣,该装置由试验台、固定支架、活动摆臂、曲率半径为 8 mm 且重量为 22 g 的实心铁半球组成。实验过程中摆臂与固定支架夹角为 57°,每次铁半球冲击长枣赤道位置。一共 60 颗灵武长枣,每颗枣冲击一次,共得到 60 颗损伤枣。

1.3 仪器

Vis/NIR 高光谱成像系统,波长为 400~1 000 nm,该系统主要由高光谱成像仪,CCD 相机,4 个 150 W 的光纤卤素灯,电控移动平台,计算机和光谱数据处理等系统构成。由于高光谱成像系统中噪音的影响,需要对获得的光谱图像按式(1)进行校正^[12]。

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

式(1)中: R 是黑白校正后的灵武长枣图像; I_R 是灵武长枣原始光谱图像; I_D 是黑板图像; I_W 是白板图像。

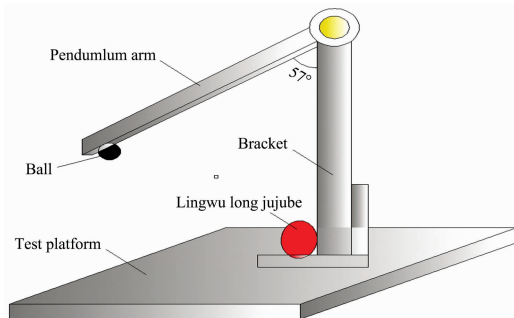


图 1 灵武长枣损伤装置

Fig. 1 Damage experimental device of Lingwu long jujube

1.4 高光谱数据获取

高光谱成像系统在采集样品图像前需要开机预热 30 min^[13]。首先采集 60 颗未损伤长枣的图像,接着利用损伤装置获得损伤长枣,采集损伤后 2, 4, 8, 12 和 24 h 长枣的光谱图像,最后一共得到 360 幅长枣的高光谱图像。利用 ENVI 软件获得感兴趣区域,计算得到完整枣和损伤后不同时间段长枣的平均光谱值。

1.5 数据处理及分析软件

原始光谱曲线有噪音和无用信息的干扰,所以利用 SG-1 和 SG-2, SNV 和 Detrending 以及 SNV-SG-1, SNV-SG-2, Detrending-SG-1, Detrending-SG-2 等预处理算法对原始光谱进行预处理,提高建模效果。为了减少全波段光谱数据的冗余,降低维数,去除无关信息,提取有效信息,建立一种低维的数据模型,所以利用 SPA, IRF, UVE, VCPA 和 IVISSA 五种算法和 IRF-SPA, UVE-SPA 和 IVISSA-SPA 三种组合算法进行特征变量的选择。PLS-DA 分类模型是基于 PLS 回归计算,全面考虑了每个变量的信息,从而能够高效准确的鉴别。LDA 是一种有监督的基于子空间的模式识别方法,该算法能够使原始样本经过线性变换后的信息更有利于分类。SVM 是一种可以分类、模式识别、拟合的监督学习模型。对于原始光谱、预处理光谱和选择的特征变量建立 PLS-DA, LDA 和 SVM 分类模型。光谱预处理、LDA 和 SVM 在 Unscramble X 10.4 程序中进行,PLS-DA 和特征变量选择在 MATLAB R 2014a 中进行,利用 Origin 2017 软件做图。

2 结果与讨论

2.1 光谱分析

图 2(a)为全部样本的曲线图,图 2(b)为完整长枣和损伤后不同时间段长枣的平均光谱曲线;由图中可以看出,6 条曲线具有相同的变化趋势,波峰波谷主要分布在 500, 645, 675, 900 和 970 nm 附近。675 nm 波段附近的吸收峰是由于样本中 C-H 伸缩振动引起^[14]。900~1 000 nm 之间的

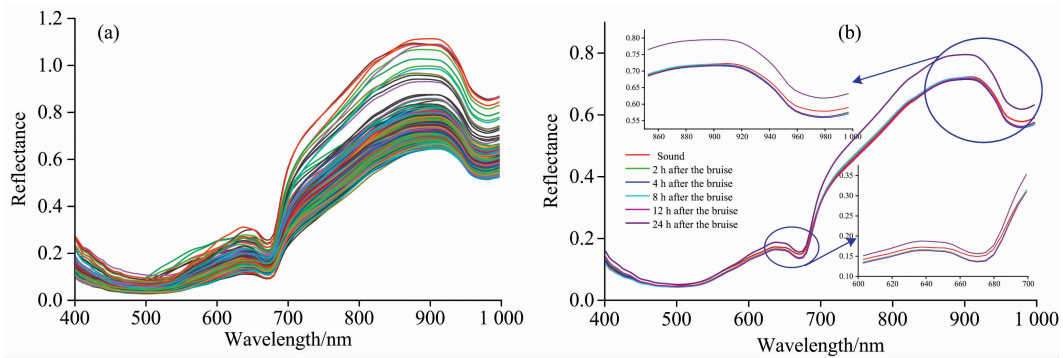


图 2 灵武长枣光谱曲线

(a): 全部样本曲线; (b): 平均光谱曲线

Fig. 2 Spectra of Lingwu long jujubes

(a): Original spectra of all samples; (b): Average spectral curves

吸收峰主要是由于样品内部水分的吸收引起, 该波段内存在水的 O—H 基团的二倍频特征吸收峰^[15]。

2.2 原始光谱与预处理光谱分类结果分析

原始光谱利用 SG-1, SG-2, SNV 和 Detrending 等算法以及不同预处理算法之间相互结合进行预处理, 建立 PLS-

DA 分类模型, 结果如表 1 所示, 所有模型的校正集和预测集准确率分别在 82.96%~91.11% 和 90%~96.67% 之间。利用不同预处理算法对原始光谱进行预处理都能提高模型分类效果, 分析得到 SNV-SG-2-PLS-DA 为最优分类模型, 该模型校正集和预测集分类准确率分别为 91.11% 和 96.67%。

表 1 原始光谱和预处理光谱的 PLS-DA 分类结果

Table 1 Classification results of PLS-DA of the original and pre-treated spectra

Pretreatment methods	Principal components	Calibration set (n=270)		Prediction set (n=90)	
		Correct	Accuracy/%	Correct	Accuracy /%
None	16	224	82.96	81	90
SG-1	19	235	87.04	84	93.33
SG-2	17	232	85.93	87	96.67
SNV	18	228	84.44	83	92.22
SNV-SG-1	17	230	85.19	83	92.22
SNV-SG-2	19	246	91.11	87	96.67
Detrending	18	231	85.56	84	93.33
Detrending-SG-1	17	236	87.41	83	92.22
Detrending-SG-2	17	233	86.30	87	96.67

Note: Correct; number of samples accurately identified, the same below

2.3 特征变量选择

利用 SPA, IRF, UVE, VCPA, IVISSA, IRF-SPA, UVE-SPA 和 IVISSA-SPA 等特征波长选择算法选择 SNV-SG-2 光谱数据的特征变量, 8 种算法选择的特征变量如表 2 所示, 特征变量位置如图 3 所示。SPA 算法共得到 23 个特征变量, 占总波长的 18.4%。IRF 算法选择特征变量, 得到 121 个间隔中排名前 10 的间隔如表 3 所示, 通过计算排名组合间隔 RMSECV 值最小, 得到 RMSECV 最小为第 68 个间隔, 通过计算共得到 108 个波长。图 4 为 UVE 算法选择特征变量过程, 虚线内的为无用的变量被剔除, 虚线以外对应的波长被选择, 共得到 68 个波长, 占总波长的 54.4%。VCPA 算法选择特征变量, 共得到 13 个波长, 占总波长的 10.4%。IVISSA 算法选择特征变量, 共得到 65 个波长, 占总波长的 52%。IRF-SPA, UVE-SPA 和 IVISSA-SPA 算法选择特征变量, 分别得到 17, 19 和 15 个波长, 分别占总波

长的 13.6%, 15.2% 和 12%。

2.4 基于特征变量的模型建立

特征变量建立的分类判别模型结果如表 4 所示。在 PLS-DA 模型分类结果中, 8 种特征变量选择算法选择的特征变量建立的模型校正集和预测集准确率分别在 72.96%~86.30% 和 74.44%~94.44% 之间。在 LDA 模型中需要变量数少于每个等级的样本数才能用于建模, 所以利用 SPA, VCPA, IRF-SPA, UVE-SPA 和 IVISSA-SPA 等 5 种算法选择的特征变量建立了 LDA 分类模型, 模型校正集和预测集准确率分别在 71.85%~86.3% 和 64.44%~83.33% 之间。在 SVM 模型分类结果中, 8 种算法选择的特征变量建立的模型校正集和预测集准确率分别在 41.49%~77.78% 和 34.44%~72.22% 之间。在建立的 PLS-DA 模型中, SNV-SG-2-UVE-PLS-DA 模型效果最好, 变量数为 68 个, 占总变量的 54.4%, 模型校正集和预测集准确率分别为 86.3%

表 2 不同算法选择的特征波长

Table 2 Characteristic wavelengths selected by different algorithms

Wavelength selection algorithms	Number of wavelengths	Wavelengths/nm
SPA	23	406, 411, 416, 421, 425, 435, 440, 445, 449, 454, 459, 469, 473, 478, 483, 488, 493, 497, 502, 507, 512, 521, 526
IRF	108	406~469 (14), 478~526 (11), 536~641 (23), 651~675 (6), 689~862 (37), 901~944 (10), 963~992 (7)
UVE	68	406, 411, 421, 425, 430, 440, 445, 449, 459, 469, 473, 483, 493, 497, 502, 507, 512, 517, 521, 531, 541, 545, 565, 569, 574, 589, 593, 598, 603, 608, 613, 617, 622, 632, 641, 670, 680, 694, 704, 709, 713, 723, 733, 737, 757, 761, 771, 795, 805, 814, 819, 829, 833, 838, 843, 862, 872, 877, 891, 906, 920, 934, 939, 949, 958, 973, 978, 987
VCPA	13	421, 469, 512, 517, 545, 579, 704, 709, 733, 771, 920, 939, 978
IVISSA	65	421, 425, 430, 435, 440, 445, 449, 454, 483, 488, 493, 497, 502, 507, 512, 517, 521, 526, 531, 560, 565, 569, 574, 579, 584, 589, 593, 598, 603, 608, 613, 617, 622, 627, 632, 637, 670, 675, 704, 709, 713, 733, 757, 761, 766, 771, 848, 853, 857, 862, 901, 906, 910, 915, 920, 925, 930, 934, 939, 968, 973, 978, 982, 987, 992
IRF-SPA	17	445, 517, 541, 617, 689, 733, 737, 757, 776, 781, 829, 838, 862, 915, 925, 939, 982
UVE-SPA	19	473, 502, 517, 541, 603, 608, 622, 641, 733, 737, 795, 805, 843, 862, 891, 920, 949, 973, 987
IVISSA-SPA	15	445, 483, 517, 569, 584, 608, 613, 709, 733, 757, 771, 920, 939, 973, 982

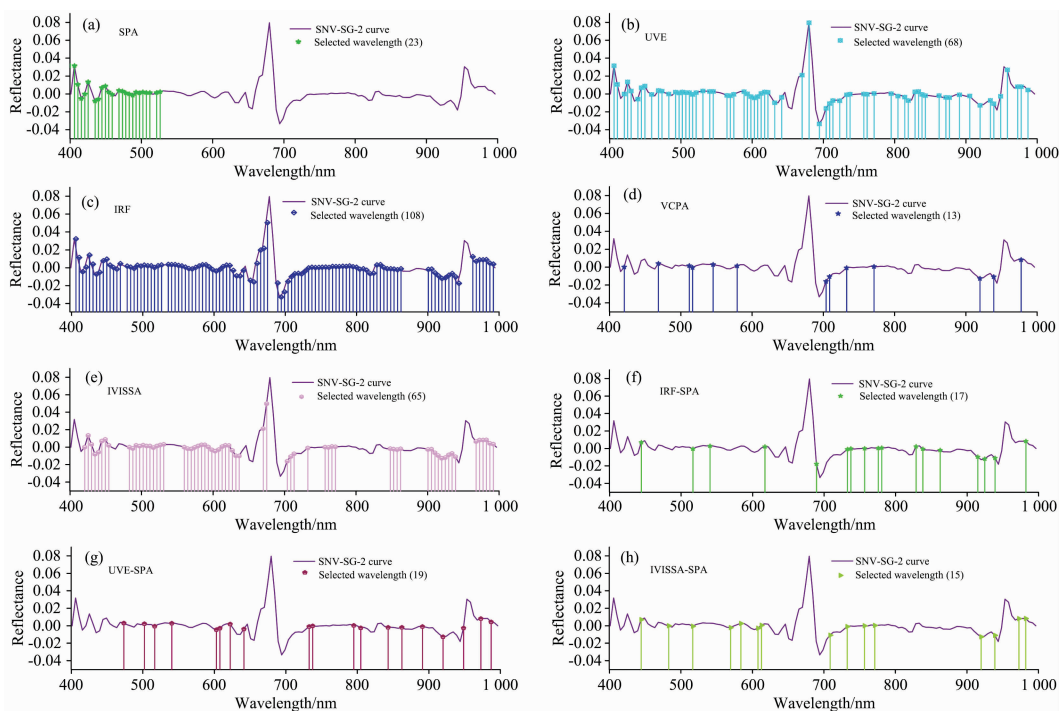


图 3 不同特征波长选择算法选择的波长

Fig. 3 Wavelengths selected by different feature wavelength selection algorithms

和 94.44%。在建立的 LDA 模型中, SNV-SG-2-SPA-LDA 模型效果最好, SPA 选择了 23 个特征变量, 占总变量的 18.4%, 模型校正集和预测集准确率分别为 86.3% 和 83.33%。在 SNV-SG-2-UVE-SVM 模型中, UVE 选择的变

量为 68 个, 模型校正集和预测集准确率分别为 77.78% 和 71.11%。通过分析, 线性判别模型(PLS-DA、LDA)的结果优于非线性判别模型(SVM), 在线性判别模型中 PLS-DA 模型分类结果优于 LDA 模型分类结果。

表 3 IRF 算法选择特征变量排名前 10 的波长间隔

Table 3 The top 10 intervals of feature variables selected by IRF

Ranking	Intervals	Ranking	Intervals
1	6~10	6	74~78
2	7~11	7	80~84
3	17~21	8	5~9
4	19~23	9	8~12
5	20~24	10	10~14

3 结 论

高光谱成像作为一种快速无损的检测方法被广泛应用。利用高光谱成像系统获得完整长枣和损伤后不同时间段(损伤后 2, 4, 8, 12 和 24 h)长枣的光谱图像, 提取感兴趣区域, 计算平均光谱值, 建立原始光谱和预处理光谱数据的 PLS-DA 分类模型, 选择 SNV-SG-2 光谱数据的特征变量建立线性(PLS-DA, LDA)和非线性(SVM)分类判别模型, 并对模型进行比较。在原始光谱数据建模中, 模型校正集和预测集准确率分别为 82.96% 和 90%。光谱经过预处理后得到 SNV-SG-2-PLS-DA 为最优分类判别模型, 模型校正集和预测集准确率分别为 91.11% 和 96.67%, 预处理可以有效提

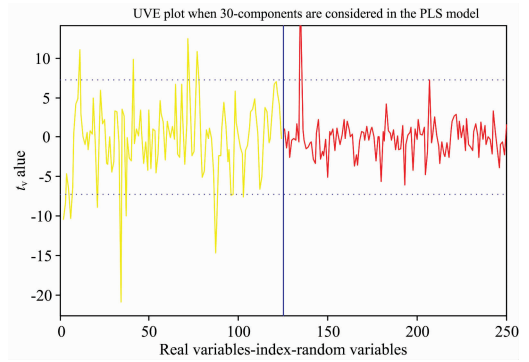


图 4 UVE 算法选择特征变量稳定性分布曲线

Fig. 4 Stability distribution curve of characteristic variables selected by UVE algorithm

高模型的分类准确率。在特征变量建立的分类模型中, SNV-SG-2-UVE-PLS-DA 模型校正集和预测集准确率分别为 86.3% 和 94.44%; SNV-SG-2-SPA-LDA 模型校正集和预测集准确率分别为 86.3% 和 83.33%; SNV-SG-2-UVE-SVM 模型校正集和预测集准确率分别为 77.78% 和 71.11%。对于特征变量选择算法来说, 有的可以提高建模准确率, 有的虽然减少了变量数, 但是使得建模效果降低, 不利于判别分

表 4 基于特征变量的分类结果

Table 4 The classification results based on characteristic wavelength

Model	Characteristic wavelength selection method	Number of characteristic wavelengths	PCs	Calibration set (n=270)		Prediction set (n=90)	
				Correct	Accuracy /%	Correct	Accuracy /%
PLS-DA	SPA	23	19	224	82.96	77	85.56
	IRF	108	17	232	85.93	85	94.44
	UVE	68	16	233	86.30	85	94.44
	VCPA	13	12	204	75.56	67	74.44
	IVISSA	65	17	229	84.81	82	91.11
	IRF-SPA	17	12	204	75.56	67	74.44
	UVE-SPA	19	14	208	77.04	77	85.56
	IVISSA-SPA	15	12	197	72.96	70	77.78
LDA	SPA	23	—	233	86.30	75	83.33
	IRF	108	—	—	—	—	—
	UVE	68	—	—	—	—	—
	VCPA	13	—	211	78.15	60	66.67
	IVISSA	65	—	—	—	—	—
	IRF-SPA	17	—	208	77.04	63	70
	UVE-SPA	19	—	220	81.48	63	70
	IVISSA-SPA	15	—	194	71.85	58	64.44
SVM	SPA	23	—	200	74.07	65	72.22
	IRF	108	—	198	73.33	65	72.22
	UVE	68	—	210	77.78	64	71.11
	VCPA	13	—	166	61.48	47	52.22
	IVISSA	65	—	208	77.04	64	71.11
	IRF-SPA	17	—	171	63.33	45	50
	UVE-SPA	19	—	148	54.81	45	50
	IVISSA-SPA	15	—	112	41.48	31	34.44

PCs: Principal components

类。对于建立的分类模型来说, 线性分类模型 (PLS-DA, LDA) 分类结果优于非线性分类模型 (SVM) 分类结果, 在线性分类模型分类结果中 PLS-DA 模型分类结果优于 LDA 模

型分类结果, 因此, PLS-DA 分类模型可以更好的为损伤灵武长枣在线检测提供分类效果。

References

- [1] Jiang W Q, Chen L H, Han Y R, et al. *Scientia Horticulturae*, 2020, 274: 109667.
- [2] Song L H, Cao B. *Acta Horticulturae*, 2016, 1116: 89.
- [3] Wang Y T, Dai Y P, Xue J R, et al. *EURASIP Journal on Image and Video Processing*, 2017, 2017: 34.
- [4] Keresztes J C, Goodarzi M, Saeyns W. *Food Control*, 2016, 66: 215.
- [5] Lee W, Kim M S, Lee H, et al. *Journal of Food Engineering*, 2014, 130: 1.
- [6] CHI Qian, WANG Zhuan-wei, YANG Ting-ting, et al(迟茜, 王转卫, 杨婷婷, 等). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2015, 46(3): 235.
- [7] Siedliska A, Baranowski P, Zubik M, et al. *Postharvest Biology and Technology*, 2018, 139: 115.
- [8] Hu M H, Dong Q L, Liu B L. *Computers and Electronics in Agriculture*, 2016, 122: 19.
- [9] Ye D, Sun L, Tan W, et al. *Chemometrics and Intelligent Laboratory Systems*, 2018, 177: 129.
- [10] Zhang M, Li G H. *International Journal of Food Properties*, 2018, 21(1): 1598.
- [11] Fan S X, Li C Y, Huang W Q, et al. *Postharvest Biology and Technology*, 2017, 134: 55.
- [12] CHENG Li-juan, LIU Gui-shan, HE Jian-guo, et al(程丽娟, 刘贵珊, 何建国, 等). *Food Science(食品科学)*, 2019, 40(10): 285.
- [13] GUO Wen-chuan, DONG Jin-lei(郭文川, 董金磊). *Optics and Precision Engineering(光学精密工程)*, 2015, 23(6): 1530.
- [14] Guo Z M, Wang M M, Agyekum A A, et al. *Journal of Food Engineering*, 2020, 279: 109955.
- [15] Siedliska A, Baranowski P, Zubik M, et al. *Postharvest Biology & Technology*, 2018, 139: 115.

Study on the Detection and Discrimination of Damaged Jujube Based on Hyperspectral Data

YUAN Rui-rui¹, WANG Bing², LIU Gui-shan^{1*}, HE Jian-guo¹, WAN Guo-ling¹, FAN Nai-yun¹, LI Yue¹, SUN You-rui¹

1. School of Food & Wine, Ningxia University, Yinchuan 750021, China

2. School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan 750021, China

Abstract Lingwu long jujube as Ningxia dominant characteristic jujube fruit. It has important economic and social value and scientific research significance. This paper has been lingwu long jujube as the research object. First, 60 intact jujubes images were collected Visible/near-infrared (Vis/NIR) using the hyperspectral imaging system. Damage tests were performed on 60 intact jujubes using the damaged device, and 60 damaged (internal bruising) jujube were obtained. The hyperspectral imaging system was used to collect the five time periods after damage (2, 4, 8, 12 and 24 h after damage) jujube spectral image. Region of interest (ROI) was extracted with ENVI software for the collected hyperspectral images of long jujube, and the average spectral value of intact long jujube and each time period long jujube were calculated. Then, the raw spectral data used Savitzky-golay smooth first derivatives (SG-1) and second derivatives (SG-2), standard normal variate (SNV) and de-trending, and the combined algorithms of SNV-SG-1, SNV-SG-2, de-trending-SG-1 and de-trending-SG-2 were pre-processed. The partial least squares-discriminant analysis (PLS-DA) classification model was established for the original spectrum and the pretreated spectrum. Finally, the optimal pre-processing spectral data were selected, and successive projection algorithm (SPA), interval random frog (IRF), uninformative variable elimination (UVE), variable combination population analysis (VCPA), interval variable iterative space shrinkage approach (IVISSA), IRF-SPA, UVE-SPA and IVISSA-SPA were used to select characteristic variables. The PLS-DA, linear discriminant analysis (LDA) and support vector machine (SVM) classification discriminant models were established for the selected feature variables. The results show that in the PLS-DA model based on the original spectral data, the accuracy of model calibration set and prediction set was 82.96% and 90%, respectively. After spectrum pretreatment, the SNV-SG-2-PLS-DA was obtained as the optimal classification discriminant model, and the accuracy of model calibration set and prediction set was 91.11% and 96.67%, respectively. In the classification model established by feature variables, the accuracy of the SNV-SG-2-UVE-PLS-DA model calibration set and prediction set were 86.3% and 94.44%, respectively. The accuracy of the SNV-SG-2-SPA-LDA model calibration set and prediction set were 86.3% and 83.33%,

respectively. The accuracy of the SNV-SG-2-UVE-SVM model calibration set and prediction set were 77.78 and 71.11%, respectively. For the classification model, the classification results of the linear classification model (PLS-DA, LDA) were superior to those of the nonlinear classification model (SVM). The results of the linear classification model, PLS-DA was superior to LDA classification results, and PLS-DA could provide a better classification effect. The results show that the hyperspectral combined with the partial least squares-discriminant analysis model could effectively realize the rapid detection of the damage of lingwu long jujube of the change of time, providing a theoretical basis for the online detection of lingwu long jujube.

Keywords Lingwu long jujube; Hyperspectral; Partial least squares-discriminant analysis; Linear discriminant analysis; Support vector machine

(Received Sep. 13, 2020; accepted Jan. 22, 2021)

YUAN Rui-rui and WANG Bing: joint first authors

* Corresponding author

敬告读者——《光谱学与光谱分析》已全文上网

从 2008 年第 7 期开始在《光谱学与光谱分析》网站(www.gpxygpx.com)“在线期刊”栏内发布《光谱学与光谱分析》期刊全文,读者可方便地免费下载摘要和 PDF 全文,欢迎浏览、检索本刊当期的全部内容;并陆续刊出自 2004 年以后出版的各期摘要和 PDF 全文内容。2009 年起《光谱学与光谱分析》每期出版日期改为每月 1 日。

《光谱学与光谱分析》期刊社