

# 低温冷冻和机械损伤条件下马铃薯高光谱图像特征响应特性研究

邹志勇<sup>1</sup>, 吴向伟<sup>1</sup>, 陈永明<sup>2</sup>, 别云波<sup>1</sup>, 王 粒<sup>1</sup>, 林 萍<sup>2\*</sup>

1. 四川农业大学机电学院, 四川 雅安 625014

2. 盐城工学院电气工程学院, 江苏 盐城 224051

**摘 要** 开展了低温冷冻和机械损伤条件下马铃薯高光谱图像特征响应特性研究。采用卓立汉光公司 Image- $\lambda$ “谱像”系列高光谱相机获取完好的、低温冷冻和机械损伤条件下的光谱波段范围为 387~1 035 nm 的马铃薯高光谱图像; 截取校正后的像素尺寸大小为 60×60 的马铃薯高光谱中部完好的图像并计算该区域平均反射率值; 冻伤的马铃薯样本的反射光谱曲线在 440, 560 和 680 nm 附近有明显吸收峰; 机械损伤样本在 560 和 680 nm 附近有明显吸收峰, 在 680 nm 附近吸收峰谷值明显低于冻伤样本; 完好的马铃薯样本反射光谱曲线相对较为平滑, 在 560 和 680 nm 附近未见明显吸收峰; 撞伤样本在 440, 560 和 680 nm 附近存在吸收峰, 而在 410 nm 附近有一个明显的反射峰。四类马铃薯样本的反射光谱曲线特征峰值表现出一定的指纹特性, 因而可以被用于后续品质特征检测分析使用。由于仪器或检测环境、光照强弱等因素影响, 光谱数据中掺杂噪声, 因此采用化学计量学预处理方法消除噪声的影响; 随机选取 70% 的马铃薯四类样本的反射光谱作为训练数据, 剩余的 30% 作为测试集; 接着, 利用极端梯度提升算法、类型提升算法和轻量梯度提升机算法来获取马铃薯高光谱图像的有效特征波谱, 减少高维海量高光谱数据对后续品质分类模型的影响; 最后, 将提取到的有效特征波长构建马铃薯品质判别模型。在建立的分类模型中, 使用的轻量梯度提升机+逻辑斯蒂回归达到最高的判别精度 98.86%。该研究为将来高光谱图像成像技术在现代农业生产加工过程中马铃薯品质有效监测与控制提供理论基础和技术支撑。

**关键词** 马铃薯; 品质; 高光谱; 低温冷冻; 机械损伤

**中图分类号:** S24 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3571-08

## 引 言

马铃薯为仅次于小麦、水稻和玉米的全球第四大经济作物, 2016 年世界马铃薯产量为 3.77 亿吨, 中国马铃薯产量占世界总产量的 26% 以上。马铃薯内部约含有 79% 的水份、17% 的碳水化合物(其中 88% 为淀粉物质)、2% 的蛋白质。马铃薯中含有大量的膳食纤维, 约占马铃薯的 50%, 是日均膳食纤维的良好来源<sup>[1]</sup>。品质优良的马铃薯不但可以给人们提供丰富的营养元素, 并且在市场上可以占据较高的销售价格。因此现代农业企业生产加工过程迫切需要解决马铃薯品质快速、无损、准确的定量分析<sup>[2]</sup>, 高光谱成像技术针对农产品品质好坏的检测, 具有快速、无损、准确等优点。高光谱成像将波谱信息和图像信息结合起来, 综合利用具有连续性、多层次、多尺度的光谱分布特征信息来定量和定性分析

农产品品质特性。近年来, 高光谱被广泛运用于对马铃薯及其相关副产品品质特性研究, 如切片马铃薯<sup>[3]</sup>等相关品质检测。段丁丁等<sup>[4]</sup>采集马铃薯块茎形成期、块茎膨大期和淀粉积累期的冠层反射光谱曲线, 对曲线进行 Savitzky-Golay 滤波平滑和一阶微分处理, 用高光谱位置参数、振幅参数、面积参数、宽度参数和反射率参数以区分不同马铃薯品种的优劣, 各指标综合能力为: 面积参数 > 振幅参数 > 反射率参数 > 宽度参数 > 位置参数。金瑞等<sup>[5]</sup>利用高光谱结合极限学习机(DM-ELM)模型对多种马铃薯(发芽、绿皮、黑心和合格)同时进行识别, 其单一识别率分别为 97.30%, 93.55%, 94.44% 和 100%, 混合识别率达到 96.58%, 时间为 0.11 s。李小昱等<sup>[6]</sup>采用 V 型平面镜的高光谱并结合果蝇优化算法(FOA)检测马铃薯轻微碰伤, 对训练集和预测集的识别准确率均达到 100%。黄涛等提取了马铃薯可见光与近红外波段 390~1 040 nm 的光谱曲线来识别马铃薯内外部缺陷, 采用

收稿日期: 2019-07-22, 修订日期: 2019-09-30

基金项目: 国家自然科学基金项目(31601227, 31501221, 61803325), 江苏省自然科学基金项目(BK20161310, BK20181049)资助

作者简介: 邹志勇, 1984 年生, 四川农业大学机电学院讲师 e-mail: scauzyou007@126.com

\* 通讯联系人 e-mail: binglvcha007@126.com

有监督局部线性嵌入与基于纠错输出编码的最小二乘支持向量机多分类模型相结合, 最终获得马铃薯合格、发芽、绿皮和空心的样本识别率分别为 96.83%, 86.96%, 86.96% 和 95%, 其中混合识别率达到 93.02%, 实验表明高光谱技术对马铃薯缺陷检测有重要应用价值。

本研究的总体目标是验证利用马铃薯的高光谱图像特征进行低温冷冻和机械损伤条件下马铃薯品质检测的可行性。采用 17 种数据预处理方法处理获取到的马铃薯的高光谱曲线, 通过建模精度分析后选择出最优预处理方法, 最后将预处理后的样本波谱特征曲线运用机器学习的方法进行建模分析。

## 1 实验部分

### 1.1 材料

马铃薯样本由云南农业大学农产品加工工程技术研究中心提供, 品种为合作 88 号小马铃薯, 产地为中国云南省。为防止马铃薯种皮上的泥土和沙石对图像采集造成干扰, 对马铃薯种皮表面进行清洗预处理, 此过程严格控制清洗力度, 避免造成马铃薯表皮破损干扰最终的实验结果。清洗后的马铃薯样本放置在阴凉的环境下通风干燥 1 d, 主要是为了晾干马铃薯种皮表面的水分。共选取了 291 个(包括 88 个完好的、69 个机械损伤的、71 个冻伤的和 63 个撞伤的)清洗并干燥后的马铃薯样本(如图 1 所示), 其外观为均匀的椭圆形[长轴尺寸大小约为  $(730 \pm 63)$  mm, 短轴尺寸大小约为  $(440 \pm 52)$  mm], 每个样本重量范围为 53~112 g。

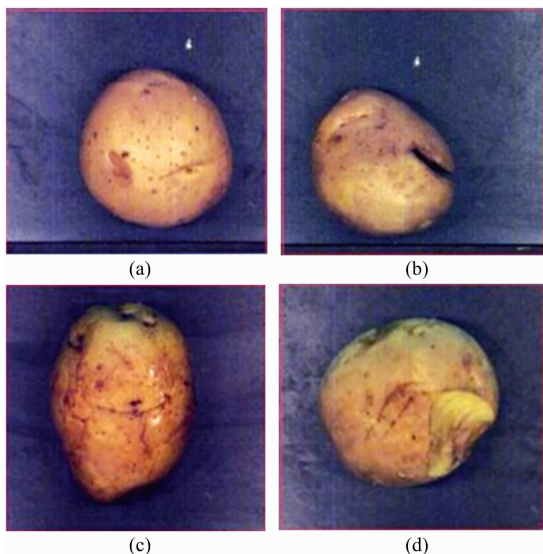


图 1 完好的(a)、机械损伤的(b)、冻伤的(c)和撞伤的(d)合作 88 号马铃薯样本

Fig. 1 Normal (a), mechanically damaged (b), frozen (c) and bruised (d) cooperative No. 88 potato samples

### 1.2 低温冷冻损伤试验

为了形成有效马铃薯样本冻伤状态, 将 72 个完好的马铃薯样本放入  $-18$  °C 低温冷冻实验装置中(型号: 捷盛 DW-

60W208) 连续存放 48 h。马铃薯出现冻伤后种皮的颜色发生改变, 出现黑褐色的斑点, 相比正常马铃薯颜色整体变的暗黑。冻伤的马铃薯样本内部出现了失水状态, 相比正常的样本的种皮形成明显的褶皱, 部分区域发生明显的凹陷, 部分区域伴有黑褐色的液体流出, 冻伤后的马铃薯样本的质量减小, 肉质变软。

### 1.3 机械冲击撞伤试验

马铃薯机械撞伤冲击损伤试验装置原理如图 2 所示, 固定硬塑料管道倾斜角度为  $35^\circ$ , 最上端垂直于地面高度为 1 147 mm, 每次把重 500 g 的实心铁球静止放置在管道的最上方, 让实心铁球自由滑落去撞击管道下端的马铃薯样本, 以保证每一次的撞击的力度均匀一致。取 66 个完好的马铃薯样本逐个放置在撞击管道下端进行冲击撞伤试验, 然后采集这些撞伤的马铃薯样本的高光谱图像。

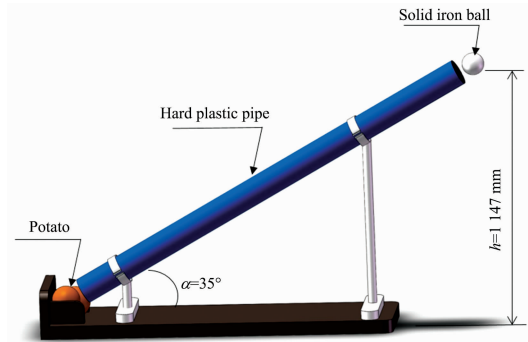


图 2 马铃薯机械撞伤冲击损伤试验装置原理图  
Fig. 2 Schematic diagram of potato mechanically damaged device

### 1.4 高光谱成像平台

马铃薯高光谱图像实验采用卓立汉光公司 Image~ $\lambda$ “谱像”系列高光谱仪, 其光谱有效波段范围为 387~1 035 nm, 波段分辨率为 2.8 nm, 共有 256 个波段, 成像的像素尺寸为  $1\ 344 \times 1\ 024$ , 每个样品测定速度设定为 60 s; 高光谱仪采用 12 V 直流稳压开关电源供电。高光谱相机物镜到马铃薯样本载物平台的物距设置为 170 mm, 电子输送移动载物平台装置的移动速度设定为  $4.6\text{ mm} \cdot \text{s}^{-1}$ , 高光谱相机曝光时间设定为 9 ms, 扫描宽带区域设定为 100 mm。采用 ENVI 4.8 (EXELIS Visual Information Solutions Inc., USA) 软件截取校正后的马铃薯高光谱图像中  $60 \times 60$  像素点感兴趣区域, 然后计算所有 387~1 035 nm 各个波段对应图像感兴趣区域上像素的平均值作为完好的、机械损伤的、低温冷冻的和撞伤的马铃薯特征反射光谱曲线。

## 2 理论和方法

### 2.1 高光谱数据预处理

为了消除高光谱波谱信息中的非品质因数信息的影响, 分别采用了 17 种光谱预处理方法<sup>[7-12]</sup>: 包括一阶导数(first derivative, FD)、二阶导数(second derivative, SD)、高斯窗平滑(Gaussian window smoothing, GWS), 盒平滑(box

smoothing, BS)、L2 范数归一化(L2 Norm Normalization, L2NN)、对数变换标准化(logarithmic transformation normalization, LTN)、Min-Max 标准化(min-max standardization, MMS)、移动平均法(moving average method, MAM)、多元散射校正(multiplicative scatter correction, MSC)、反余切校正(anti-cotangent normalization, CAN)、小波阈值去噪(wavelet threshold denoising, WTD)、指数平滑(exponential smoothing, ES)、中值滤波(median filtering, MF)、Z 分数标准化(Z-score standardization, ZSS)、局部回归加权线性最小二乘+一阶多项式模型(local regression-weighted linear least squares+first order polynomial model, LR1)、局部回归加权线性最小二乘+二阶多项式模型(local regression-weighted linear least squares+second order polynomial model, LR2)和 Savitzky-Golay 滤波(savitzky-golay filtering, SG)来消除原始光谱(Raw)数据中存在的噪声。

## 2.2 高光谱特征选择和建模

高光谱高维的变量经常包含大量无关信息, 冗余的变量会影响最终模型的分精度。采用 CatBoost, GBDT, LightGBM, 逻辑斯蒂回归(logistic regression, LR)、随机森林(random forest, RF)和 XGBoost 算法和堆叠集成学习算法(stacking)来有效选择高光谱特征变量选择和利用这些特征变量进行建模分析<sup>[13]</sup>。

## 2.3 模型性能评估

随机选取 70% 的合作 88 号马铃薯四类样本的反射光谱曲线数据作为训练集, 剩余 30% 的合作 88 号马铃薯四类样本的反射光谱曲线数据作为测试集, 利用机器学习算法构建判别模型验证四类完好、机械损伤、冻伤和撞伤条件下马铃

薯高光谱图像特征响应逻辑属性。准确率(Accuracy), F-Score, 对数损失度量(Log Loss)和海明损失度量(Hamming Loss)用于评估模型训练预测的效果<sup>[14]</sup>。

## 3 结果和讨论

### 3.1 马铃薯特征波谱分析

四类(完好的、机械损伤的、冻伤的和撞伤的)合作 88 号马铃薯样本的反射光谱曲线如图 3 所示。冻伤的合作 88 号马铃薯样本的反射光谱曲线在 440, 560 和 680 nm 附近有明显吸收峰; 机械损伤样本在 560 和 680 nm 附近有明显吸收峰, 在 680 nm 附近吸收峰谷值明显低于冻伤样本; 完好的马铃薯样本反射光谱曲线相对较为平滑, 在 560 和 680 nm 附近没有明显吸收峰; 撞伤样本在 440, 560 和 680 nm 附近存在吸收峰, 区别于其他三类样本的是它在 410 nm 附近形成一个明显的反射峰。在可见光波段 420~650 nm 所有样本有波谷, 在近红外波段 820~850 nm 所有样本有波峰, 机械损伤的波峰大于冻伤的波峰值, 因此四类马铃薯样本的反射光谱曲线特征峰值表现出一定的指纹特性。另外, 由于光谱曲线两端的噪声比较大, 因此选择 387~1 035 nm(合计 256 个波段)进行后续品质特征检测分析使用。完好的样本具有最高的反射率, 在可见光波段 420~650 nm 的冻伤样本有最低反射率, 在可见光波段 651~700 nm 的机械损伤样本有最低反射率, 并且可以观测到曲线突变; 在近红外波段 701~950 nm 三种损伤的光谱基本一样; 在 951~1 050 nm 的冻伤样本有最低反射率。因此, 正常样本与损伤样本的光谱反射特征局部特性不一致, 正常样本的反射率高于损伤的样本;

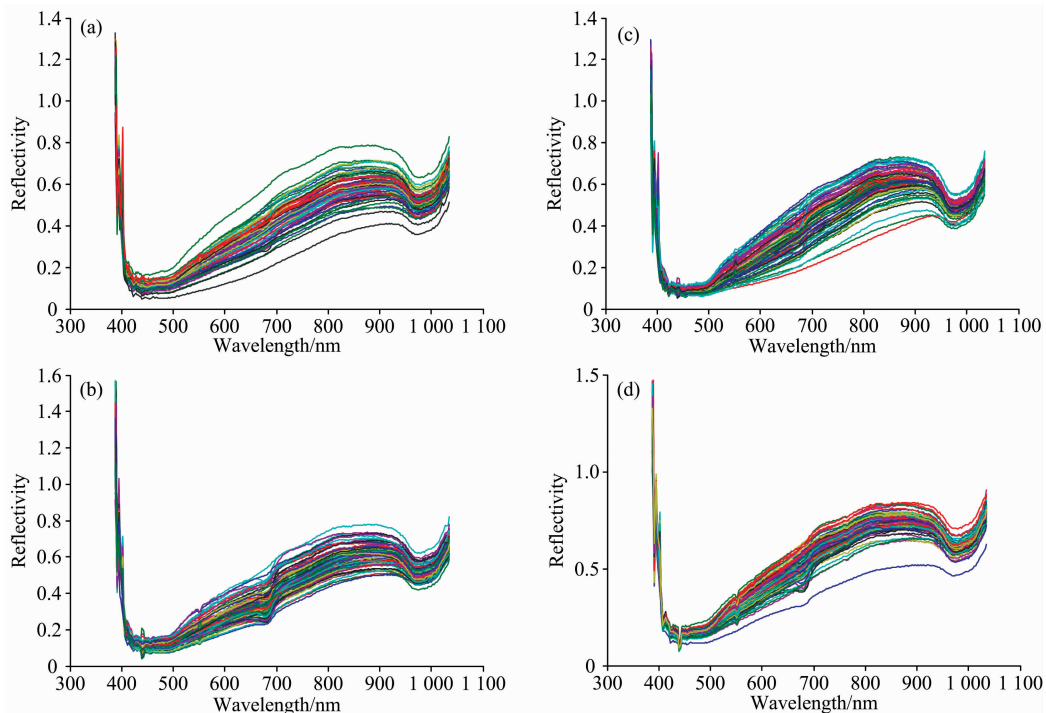


图 3 完好的(a)、机械损伤的(b)、冻伤的(c)和撞伤的(d)合作 88 号马铃薯样本的反射光谱曲线

Fig. 3 Reflectance spectra of normal (a), mechanically injured (b), frosted (c) and bruised (d) cooperative No. 88 potato samples

机械损伤、冻伤和撞伤的光谱反射趋势基本一致。这些光谱曲线之间的差异性为使用高光谱成像识别马铃薯损伤特性提供了有效判别依据。

### 3.2 马铃薯高光谱数据预处理

获取马铃薯光谱数据时,由于仪器或检测环境、光照强弱等因素影响,光谱数据中掺杂有噪声,这些掺杂的噪声会直接影响到后续模型的判别性能。对马铃薯原始光谱数据进行预处理的结果如表 1 所示。由表 1 可知,LTN, MAM, ZSS, LR1 和 LR2 算法处理后的预测精度与使用原始数据进行预测的精度无明显差异,都为 89.77%; MSC 和 BS 预测预测的精度分别为: 88.64%和 87.50%,比直接原始数据进行预测的效果较差; LTN 算法处理结果和使用原始数据预测结果一致,精度为 89.77%; MMS 相对预测能力较弱,获

得的精度值为 80.68%。基于导数的预处理方法, FD 和 SD 算法预测精度分别为 81.82%和 84.09%,比直接使用原始数据的总体预测效果差; GWS, ES 和 WTD 算法比直接利用原始数据进行预测的效果较好,预测精度为 90.91%; L2NN 算法比直接利用原始数据进行预测的效果较好,预测精度为 93.18%; MF 算法获得 96.59%的最高的预测精度。MSC 算法预处理用时最长为 20.71 s; MF 算法用时最短,为 14.72 s。MF 算法获得的 Log Loss 值最小,为 497.62; MMS 获得的 Hamming Loss 值最大,为 981.16; MF 算法获得的 Hamming Loss 值最小,为 0.03; MF 算法获得的 Jaccard Similarity 值最大,为 0.97; MMS 算法获得的 Jaccard Similarity 值最小,为 0.81。17 种预处理算法预测效果对比结果,最终选择 MF 算法预处理原始马铃薯高光谱特征曲线数据。

表 1 马铃薯高光谱特征曲线预处理结果

Table 1 Prediction results of potato hyperspectral curves preprocessed by different methods

Methods	Accuracy /%	F-Score Frostbite /%	F-Score Mechanical Injury/%	F-Score Normal /%	F-Score Bruise /%	Log Loss	Hamming Loss	Jaccard Similarity	Fitting time/s
Raw	89.77	90.48	82.05	95.38	86.67	704.85	0.10	0.90	18.65
FD	81.82	78.26	78.05	93.55	66.67	946.62	0.18	0.82	18.68
SD	84.09	68.42	90.00	98.41	68.57	877.54	0.16	0.84	18.37
GWS	90.91	90.48	84.21	96.97	86.67	670.31	0.09	0.91	18.75
BS	87.50	90.48	78.05	95.38	78.57	773.93	0.13	0.88	19.14
L2NN	93.18	95.24	90.48	93.75	92.86	601.23	0.07	0.93	19.95
LTN	89.77	90.48	82.05	95.38	86.67	704.85	0.10	0.90	17.92
MMS	80.68	71.11	80.00	98.46	58.06	981.16	0.19	0.81	18.70
MAM	89.77	93.02	80.95	96.97	80.00	704.85	0.10	0.90	19.54
MSC	88.64	90.48	76.92	95.38	86.67	739.39	0.11	0.89	20.71
ACN	88.64	90.48	80.00	95.38	82.76	739.39	0.11	0.89	18.74
WTD	90.91	93.02	82.05	95.38	89.66	670.31	0.09	0.91	18.36
ES	90.91	90.48	84.21	96.97	86.67	670.31	0.09	0.91	18.94
MF	96.59	92.68	95.00	100.00	96.77	497.62	0.03	0.97	14.72
ZSS	89.77	90.48	82.05	96.97	82.76	704.85	0.10	0.90	18.07
LR1	89.77	90.48	82.05	96.97	82.76	704.85	0.10	0.90	19.36
LR2	89.77	90.48	82.05	96.97	82.76	704.85	0.10	0.90	19.10
SG	84.09	87.80	70.00	92.06	81.25	877.54	0.16	0.84	18.57

### 3.3 马铃薯高光谱曲线特征波长选择

原始的马铃薯高光谱曲线包含 256 个特征变量,采用 CatBoost, GBDT, LightGBM, 逻辑斯蒂回归(logistic regression, LR)、随机森林(random forest, RF)和 XGBoost 算法来提取马铃薯高光谱曲线特征波长,选出最具贡献率的前 30 个的特征变量,最终获得的权重特征波长如图 4 所示。特征提取算法 GBDT, LightGBM 和 XGBoost 都把 439.9 nm 波长作为了首要的特征变量;在 CatBoost 检测模型中,该波长权重贡献率排名第五;在 RF 检测模型中,该波长权重贡献率排名第五;在 LR 检测模型中,该波长权重贡献率排名第四。实验结果表明这些特征提取算法在特征波长选择上具有一定的共性但也存在差别,这些获取的特征变量将作为后续判别算法的输入变量。

### 3.4 马铃薯品质高光谱响应结果分析

将 XGBoost, CatBoost 和 LightGBM 模型选择的特征波长训练建模,并且分别将此 3 个模型的第一层与 LR 和 RF 模型的第二层去构建一个新的 Stacking 训练模型,预测结果如表 2 所示。采用 6 种不同层结构建模后 XGBoost 和 CatBoost 算法获得相同的预测精度: 97.73%。CatBoost 选择的特征波长预测效果较差,预测精度值也为 97.73%,但与全波长相比,模型的复杂度和消耗时间都降低了,提高了效率;LightGBM 预测模型中使用 LR 算法处理光谱数据后的选择的特征波长获得最优的预测效果,精度达到 98.86%;同样,Stacking 预测模型中使用 LR 算法处理光谱数据后的选择的特征波长获得最优的预测效果,精度达到 98.86%;但是,LightGBM+LR 相对 Stacking+LR 建立模型耗时较

少, 为 0.23 s。综合表 2 中所列数据指标, 最后选择 LightG-  
BM+LR 组合模型作为马铃薯品质判别分析应用。

用混淆矩阵图观测完好的、机械损伤的、冻伤的和撞伤

的四类合作 88 号马铃薯样本分类结果。如图 5 所示, 机械损伤  
的马铃薯样本被误识别为冻伤的马铃薯样本; 冻伤的马铃  
薯样本被误识别为撞伤马铃薯样本, 因此对马铃薯不同损伤

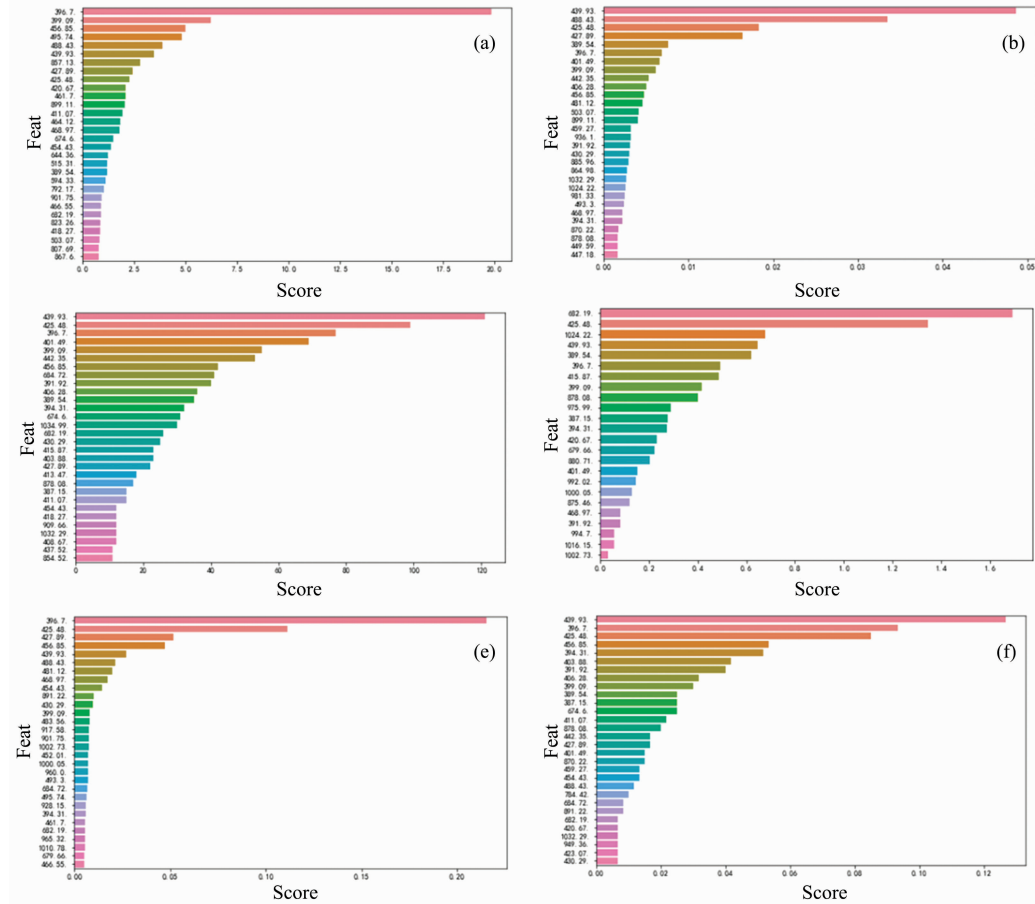


图 4 使用 CatBoost (a), GBDT (b), LightGBM (c), Logistic Regression (d),  
Random Forest (e) 和 XGBoost (f) 算法选择马铃薯高光谱曲线特征波长

Fig. 4 Using the CatBoost (a), GBDT (b), LightGBM (c), Logistic Regression (d), Random Forest (e)  
and XGBoost (f) algorithms to select the characteristic wavelengths of potato hyperspectral curves

表 2 不同算法提取到的特征波长的建立的预测模型性能

Table 2 Predictive performance of models using the characteristic wavelengths extracted by different algorithms

Methods	Layers	Accuracy /%	F-Score Frostbite /%	F-Score Mechanical Injury/%	F-Score Normal /%	F-Score Bruise /%	Log Loss	Hamming Loss	Jaccard Similarity	Fitting Time /s
XGBoost	CatBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.22
	GBDT	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.22
	LightGBM	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.22
	LR	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.22
	RF	97.73	97.56	95.00	100.00	96.77	463.08	0.02	0.98	0.22
	XGBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.22
LightGBM	CatBoost	96.59	95.24	92.31	100.00	96.77	497.62	0.03	0.97	0.23
	GBDT	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.23
	LightGBM	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.23
	LR	98.86	97.56	97.56	100.00	100.00	428.54	0.01	0.99	0.23
	RF	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.23
	XGBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	0.23

续表 2

CatBoost	CatBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	14.54
	GBDT	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	14.54
	LightGBM	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	14.54
	LR	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	14.54
	RF	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	14.54
	XGBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	14.54
Stacking	CatBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	2.87
	GBDT	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	3.07
	LightGBM	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	3.29
	LR	98.86	97.56	97.56	100.00	100.00	428.54	0.01	0.99	2.70
	RF	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	2.87
	XGBoost	97.73	95.24	95.00	100.00	100.00	463.08	0.02	0.98	2.97

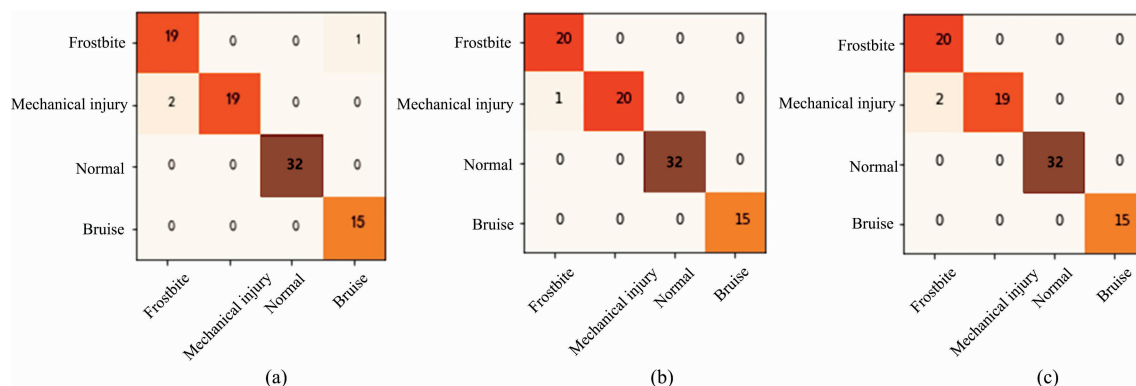


图 5 使用 CatBoost+XGBoost (a), LightGBM+LR (b) 和 XGBoost+XGBoost 预测结果混淆矩阵图  
 Fig. 5 Confusion matrix diagrams of models using the CatBoost+XGBoost (a), LightGBM+LR (b),  
 and XGBoost+XGBoost algorithms for prediction

性能的识别能力还有待进一步提升。三类损伤类别都未被误分类为完好的样本，同时完好的样本也未被误识别为损伤样本，即正确识别率达 100%，这表明建立的预处理模型对损伤与正常的识别性能较好，可以应用于现代农业自动化流水线生产加工过程中马铃薯品质有效监测。

## 4 结 论

分别采用了 17 种高光谱数据预处理方法消除原始马铃薯高光谱数据中的噪音影响，预处理结果显示不同的数据预处理方法预测结果差异性显著，并不是对原始高光谱数据实施的预处理方法就会带来模型预测性能的提升，采用 MMS 数据预处理是后的检测精度最低：80.68%，采用 MF 预处理算法获得的最高的检测精度：96.59%，因此需要合理选择的高光谱数据预处理算法来提升后续马铃薯品质检测的精度。特征提取算法 GBDT, LightGBM 和 XGBoost 都把 439.9

nm 波长作为了首要的特征变量；在 CatBoost 检测模型中，该波长权重贡献率排名第二；在 RF 检测模型中，该波长权重贡献率排名第五；在 LR 检测模型中，该波长权重贡献率排名第四。LightGBM+LR 多层次模型获得最高的判别精度：98.86%。结果表明这些特征提取算法在特征波长选择上具有一定的共性但也存在差别，这些最终获取的特征变量决定模型的最终预测效果。最终的品质分类结果显示：机械损伤的马铃薯样本被误识别为冻伤的马铃薯样本；冻伤的马铃薯样本被误识别为撞伤马铃薯样本，因此对马铃薯不同损伤性能的识别能力还有待进一步提升。三类损伤类别都未被误分类为完好的样本，同时完好的样本也未被误识别为损伤样本，即正确识别率达 100%，这表明建立的模型可以有效区分损伤与正常样本。本研究为将来高光谱图像成像技术在现代农业自动化流水线生产加工过程中马铃薯品质有效监测与控制提供理论基础和技术支撑。

## References

- [ 1 ] LI Yan, ZHANG Hai-fang, HAN Yu-mei, et al(李 艳, 张海芳, 韩育梅, 等). Food Science and Thchnology(食品科技), 2018, 43(12): 206.
- [ 2 ] Griffel L M, Delparte D, Edwards J. Computers and Electronics in Agriculture, 2018, 153: 318.
- [ 3 ] López A, Arazuri S, García I, et al. Journal of Agricultural and Food Chemistry, 2013, 61(23): 5413.
- [ 4 ] DUAN Ding-ding, HE Ying-bin, YU Jin-kuan, et al(段丁丁, 何英彬, 于金宽, 等). Journal of Anhui Agricultural University(安徽农业大学学报), 2019, 46(3): 521.
- [ 5 ] JIN Rui, LI Xiao-yu, YAN Yi-yun, et al(金 瑞, 李小昱, 颜伊芸, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(16): 258.
- [ 6 ] LI Xiao-yu, XU Sen-miao, FENG Yao-ze, et al(李小昱, 徐森淼, 冯耀泽, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2016, 47(1): 221.
- [ 7 ] Ma Xiaojun, Sha Jinglan, Wang Dehua, et al. Electronic Commerce Research and Applications, 2018, 31: 24.
- [ 8 ] WANG Fang-jie, WANG Fu-jian, WANG Yu-chen, et al(王芳杰, 王福建, 王雨晨, 等). Journal of Transportation Systems Engineering and Information Technology(交通运输系统工程与信息), 2019, 19(2): 19.
- [ 9 ] LI Hong-xiang, HUANG Hao, ZHENG Zi-xuan(李鸿祥, 黄 浩, 郑子旋). Finance(金融), 2019, 9(3): 137.
- [10] Ma Xiaojun, Sha Jinglan, Wang Dehua, et al. Electronic Commerce Research and Applications, 2018, 31: 24.
- [11] SHI Jia-qi, ZHANG Jian-hua(史佳琪, 张建华). Proceedings of the CSEE(中国电机工程学报), 2019, 5: 23.
- [12] Aishwarya Kakade, Baby Kumari, Pankaj Singh Dholaniya. Journal of Theoretical Biology, 2018, 457: 14.
- [13] XU Shao-cheng, LI Dong-xi(徐少成, 李东喜). Statistics and Decision(统计与决策), 2018, 34(18): 25.
- [14] Trajdos P, Kurzynski M. International Journal of Neural Systems, 2017, 28(9): 1750062.

## Investigation of Hyperspectral Imaging Technology for Detecting Frozen and Mechanical Damaged Potatoes

ZOU Zhi-yong<sup>1</sup>, WU Xiang-wei<sup>1</sup>, CHEN Yong-ming<sup>2</sup>, BIE Yun-bo<sup>1</sup>, WANG Li<sup>1</sup>, LIN Ping<sup>2\*</sup>

1. College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Ya'an 625014, China

2. School of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, China

**Abstract** The hyperspectral imaging technology was used to detect the frozen and mechanical damaged potatoes. The Zolix's Image~λ "spectrum" series of hyperspectral imaging device was employed to obtain the intact, frozen and mechanical damaged potato hyperspectral data within band range of 387~1 035 nm; Secondly, the 60×60 pixel sizes of region of interest in the intact, frozen and mechanically damaged potato hyperspectral image was cropped to calculate the average reflectance values; The reflectance spectral curves of frozen potato samples had the obvious absorption peaks near the visible wavelengths of 440, 560 and 680 nm; The reflectance spectral curves of mechanical damaged potato samples had the obvious absorption peaks near the visible wavelengths of 560 and 680 nm, and the absorption peaks and valleys near the visible wavelength of 680 nm were significantly lower than the frozen potato samples; The reflectance spectral curves of intact potato samples were relatively smooth, and there were no obvious absorption peaks appearing near the visible wavelengths of 560 and 680 nm; There were three absorption peaks near the visible wavelengths of 440, 560 and 680 nm in the bruised samples, and there was a significant reflectance peak near the visible wavelength of 410 nm. Four categories of potato samples demonstrated the different fingerprint characteristics in the reflectance spectral curves, which could be further used for the aim of potato quality discrimination. The instrument, detection environment, illumination intensity and other factors would add the noise variables to the obtained raw spectral data, so thirdly, the chemometric pretreatment methods were employed to eliminate the influence of noise in the raw spectral curves. There were 70 percent of the four kinds of potato samples randomly selected as the training dataset and the remaining 30 percent as test dataset; Fourthly, the method of local outlier factor (LOF) was used to identify the neighborhood point density of the spatial region of the collected potato spectral curves in order to find the abnormal non-nearest neighbor sample distribution to eliminate the abnormal samples; Fifthly, three types of boosting algorithms of extreme gradient boosting (XGBoost), categorical boost (CatBoost) and light gradient boosting machine (LightGBM) were used to extract the effective characteristic spectral bands from the potato hyperspectral curves, so that the dimensions of massive hyperspectral data for the subsequent classification mod-

eling were reduced; Finally, the characteristic wavelengths of extracted effective spectral data were used to construct the discriminant model of potato quality. The established classification model by using the LightGBM+Logistic regression reached the highest discriminant accuracy of 98.86%. Our study provided the theoretical basis and technical support for effectively monitoring potato quality in the process of modern agricultural production.

**Keywords** Potato; Quality; Hyperspectral; Frozen; Mechanically damaged

(Received Jul. 22, 2019; accepted Sep. 30, 2019)

\* Corresponding author