

王新忠, 卢青, 张晓东, 等. 基于高光谱图像的黄瓜种子活力无损检测[J]. 江苏农业学报, 2019, 35(5): 1197-1202.
doi: 10.3969/j.issn.1000-4440.2019.05.028

基于高光谱图像的黄瓜种子活力无损检测

王新忠^{1,2}, 卢青^{1,2}, 张晓东^{1,2}, 吴又新^{1,2}, 承银辉^{1,2}

(1. 江苏大学农业装备工程学院, 江苏 镇江 212013; 2. 江苏大学现代农业装备与技术教育部重点实验室, 江苏 镇江 212013)

摘要: 为实现对黄瓜种子的快速、无损检测, 以人工老化 0 h、36 h、72 h 的 3 个不同活力梯度的黄瓜种子为研究对象, 利用波长 400~1 000 nm 的可见光光谱对黄瓜种子活力进行检测。对比了多元散射校正(MSC)、标准正态变换(SNV)、卷积平滑(S-G) 3 种预处理方法, 结果显示 SNV 预处理的效果最优。从特征提取和特性选择 2 个角度进行降维分析。分别使用主成分分析法和连续投影算法, 对比各个主成分数的正确分类率, 选取最佳的主成分数。通过连续投影算法(SPA) 选择 9、12、13 个特征波长, 通过对比分类正确率, 选出最佳波长数为 12 个。最后将提取出的最佳主成分和选择的最佳特征波长作为支持向量机的输入, 分别选择线性核函数和径向基核函数, 结合网格搜索方法, 确定模型的惩罚因子 c 和径向基核函数中的参数 γ , 建立判别分析模型。所有模型分类正确率均达到 97.3% 以上, 其中 SPA-SVM(基于 RBF 核函数) 效果最佳, 分类正确率达到 98.6%。可见, 利用高光谱图像技术结合 SPA-SVM 能有效地鉴别黄瓜种子的活力。

关键词: 高光谱图像; 种子活力; 主成分分析; 连续投影算法; 支持向量机

中图分类号: S625.5 **文献标识码:** A **文章编号:** 1000-4440(2019)05-1197-06

Non-destructive detection of cucumber seeds vigor based on hyperspectral imaging

WANG Xin-zhong^{1,2}, LU Qing^{1,2}, ZHANG Xiao-dong^{1,2}, WU You-xin^{1,2}, CHENG Yin-hui^{1,2}

(1. College of Agricultural Equipment Engineering, Jiangsu University, Zhenjiang 212013, China; 2. Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang 212013, China)

Abstract: To detect the cucumber seeds rapidly, precisely and nondestructively, the cucumber seeds were subjected to three different gradient aging treatments of 0 h, 36 h, and 72 h. The seeds were detected using visible light spectra with wavelengths ranging from 400 nm to 1 000 nm. The three pretreatment methods of multiplicative scatter correction (MSC), Savitzky-Golay (S-G) and standard normal variate (SNV) were compared. The result showed that the effect of standard normal variate method was optimal. It was dimensionality-analyzed from two aspects: feature extraction and feature selection. Principal component analysis and continuous projection algorithm were used to compare the correct classification rate under each principal component number, and the optimal principal component number was chosen. The nine, twelve and thirteen characteristic wavelengths were selected by successive projection algorithm (SPA), and the optimal wavelength was selected by comparing the correct classification rate to 12. Finally, the extracted optimal principal component and the selected optimal feature wavelength were used as the input of the support vector machine, and the linear kernel function and the radial basis kernel function were selected, respectively. Moreover, the grid search method was combined to determine the penalty factor of the model and the parameter in the radial basis kernel function, and the discriminant a-

收稿日期: 2019-04-11

基金项目: 国家自然科学基金项目(61771224); 江苏省重点研发计划项目(BE2016323)

作者简介: 王新忠(1969-), 男, 河北石家庄人, 博士, 教授, 主要从事农业设施装备研究。(E-mail) xzwang@Ujs.edu.cn

通讯作者: 卢青, (Tel) 18605243646

optimal feature wavelength were used as the input of the support vector machine, and the linear kernel function and the radial basis kernel function were selected, respectively. Moreover, the grid search method was combined to determine the penalty factor of the model and the parameter in the radial basis kernel function, and the discriminant a-

analysis model was established. The classification accuracy of all models was above 97.3%, the SPA-SVM (based on RBF kernel function) had the best effect, and the correct classification rate reached 98.6%. It can be seen that the use of hyperspectral image technology combined with SPA-SVM can effectively identify the vigor of cucumber seeds.

Key words: hyperspectral image; seed vigor; principal component analysis; successive projections algorithm; support vector machine

农业生产中种子是基础生产物资之一。种子活力反映了种子质量状况,种子活力检测对种子质量检测和储存具有重要意义。黄瓜为设施种植中重要蔬菜之一。种子活力的高低直接影响黄瓜的种子出苗率和产量。所以,对黄瓜种子进行活力检测具有重要意义^[1-2]。

常用种子活力检测方法有直接法和间接法。直接法通过发芽试验确定种子的活力状况,间接法通过酶活性、ATP含量等检测评定种子活力。这些活力检测方法可以直观地评定种子活力,但存在一定的不足,例如工作量大,效率低,成本高,对种子有不可逆的损害^[3-4]。

随着老化时间的增加,种子可溶性蛋白质含量和保护酶活性会明显下降。崔鸿文^[5]采用化学方法测得人工老化的黄瓜种子随着老化时间的增加,过氧化氢酶和脱氢酶的活性都递减,这些蛋白质大分子在光谱中均有对应的信号。变质的蛋白质大分子的可见光及近红外光吸收波长和强度会有差异,形成的光谱图像也有差异。即不同老化程度的种子对应的高光谱图像也存在差异^[6]。因此,本研究尝试将高光谱图像技术用于种子活力的分级检测。

Huang等^[7]运用高光谱技术结合偏最小二乘法,成功建立了分类模型,能无损高效地鉴别不同年份收获的玉米种子。Wakholi等^[8]运用高光谱技术对不同活力的玉米种子进行分类,对比分析了线性判别分析(LDA)、偏最小二乘法判别分析(PLS-DA)和支持向量机(SVM)3种分类模型的效果,试验证明SVM的效果最佳。2015年,Ambrose等^[9]运用高光谱技术成功鉴别区分有活力和无活力的玉米种子。2014年,Nansen^[10]等人运用高光谱图像技术结合LDA选为10的特征波段,对3种不同老化程度的澳大利亚树种进行分类。可见,光谱图像可以作为鉴别不同活力种子的依据。

大田撒播类作物(例如玉米、小麦、水稻)的种子活力对产量的影响较小,而在精准定量播种的设

施农业中,种子活力对产量的影响较大^[11]。本研究尝试将高光谱技术运用于黄瓜种子的活力检测及分类中,并进一步对比分析线性核函数和径向基核函数(非线性)对SVM模型正确分类的影响,结合参数寻优方法,进一步优化种子分类模型,提高模型的效率与准确性。

1 材料与方法

1.1 黄瓜种子的处理

采用新津春4号黄瓜种子。将种子分为3组,每组8g,分别进行不同程度老化处理。将干燥器底部装50ml水,将陶瓷片换成铁丝架,更利于水分的均匀扩散。提前12h将干燥器置于45℃的恒温箱中预热^[12]。将1~3组种子分别装入纱布袋中。在温度45℃、湿度100%下对1~3组种子进行0、36、72h的老化处理,得到3个不同活力梯度的种子。从外观看,经过人工老化的种子无明显差异。黄瓜种子在老化处理前、后都置于常温干燥处保存。老化处理的种子干燥后,每组选取100粒,送往实验室进行高光谱图像采集。

1.2 试验设备

高光谱图像采集系统主要结构包括高光谱图像摄影仪、150W光纤卤素灯、控制箱和电控位移台、计算机、暗箱等(图1)。在测量高光谱图像过程中,需关闭暗箱,避免阳光和灯光对试验的干扰。

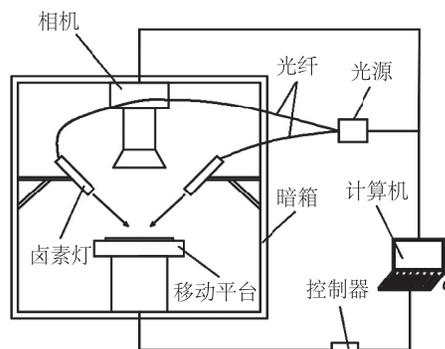


图1 高光谱成像系统

Fig.1 Hyperspectral imaging system

1.3 黄瓜种子高光谱图像采集与校正

在进行黄瓜种子高光谱图像采集试验前,需要对高光谱成像系统进行黑白标定试验^[13],获取校正过的图像。对样本进行图像信息采集时,将黄瓜种子平铺于白纸上,每粒种子间相距一段距离,再将白纸置于移动平台上,进行高光谱的采集。为了避免出现高光谱图像的失真情况,需要对高光谱仪设定一些参数。2个卤素灯与移动平台之间的夹角设置为45°,相机的曝光时间设置为20 ms,移动平台的移动速度设置为1 mm/s,样本与镜头之间距离约为28 cm。最后依次采集3组样本的高光谱图像。

1.4 黄瓜种子高光谱数据的提取与处理

在提取高光谱数据前,需要先确定黄瓜种子高光谱图像的ROI(感兴趣区)。ROI选取是影响之后建立的模型性能的重要因素之一,本试验统一选取单粒种子的全部区域作为ROI。采用ENVI5.1软件处理高光谱图像^[14]。然后通过计算每一粒种子ROI区域内像素的平均值,得到每粒种子的平均高光谱数据。图2为不同活力梯度的黄瓜种子平均高光谱曲线。从图2可看出,未老化的黄瓜种子光谱曲线平均反射率最高,随着老化程度增加,平均反射率逐渐降低。总体来看,不同老化梯度的黄瓜种子具有一定的可分类性。

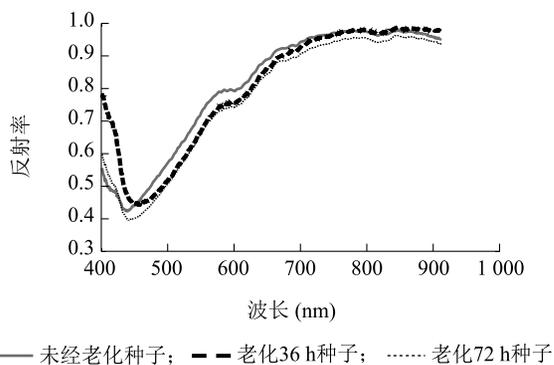


图2 不同活力梯度黄瓜种子的平均高光谱曲线

Fig.2 Mean hyperspectral curve of cucumber seeds with different vigor gradients

1.5 黄瓜种子活力验证试验

将采集过光谱的黄瓜种子,进行发芽试验,每个活力梯度各100粒种子。在发芽皿中铺一层滤纸,用少量水将滤纸湿润。将黄瓜种子按活力梯度分别置于不同发芽皿中,在28℃、4000 lx光照条件下进行发芽,每天观察并浇水,保持滤纸的湿润。第6 d

时测定各个活力梯度的黄瓜种子的发芽率和芽长。测得未经老化的种子发芽率为93%,老化36 h的种子发芽率为88%,老化72 h的种子发芽率为76%。发芽后测得各个梯度的幼苗高度(图3),从图3可看出未经老化的幼苗高度普遍大于老化36 h的幼苗,老化36 h的幼苗高度大于老化72 h的幼苗。综上所述,人工老化的种子存在活力的差异。

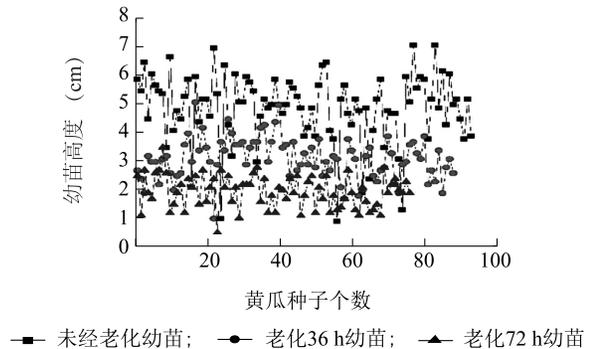


图3 不同活力梯度黄瓜种子发芽后的幼苗高度

Fig.3 Height of cucumber seedlings with different vigor gradients

1.6 数据处理

1.6.1 不同老化梯度种子的光谱特征 在黄瓜种子高光谱图像采集和提取过程中,通常会存在很多干扰因素^[15]。为了减少干扰因素,对采集的高光谱数据分别进行归一化、多元散射校正(Multiplicative scatter correction, MSC)、标准正态变换(Standard normal variate, SNV)、卷积平滑(Savitzky-Golay, S-G)处理,对比选择效果最优的预处理方法。将分别采用基于径向基核函数和线性核函数的SVM进行建模分析。运用网格搜索法^[16]确认惩罚因子 c 和径向基核函数中的 γ 值。建模时将所有300个样本按3:1分成训练集和测试集,训练集225个样本,测试集75个样本。从表1可以看出不同预处理方法得到的结果不同。归一化+原始数据、归一化+S-G、归一化+MSC 3种方法均没有达到较好的效果。归一化+SNV的效果最好,基于径向基核函数(Radial basis function, RBF)的SVM分类错误数量为0,正确分类率达到100.0%;基于线性核函数的SVM分类错误数量为1,正确分类率达98.6%。因此,选择归一化+SNV预处理方法。

1.6.2 高光谱的特征提取和特征选择 高光谱原始数据具有数据繁杂、波段多、误差信息多等特点。

如果直接对原始数据进行建模可能会导致建模效率低下,且模型的分辨性能差^[17]。因此,须在高光谱数据建模前进行降低数据维度的处理,去除多余的信息。特征提取和特征选择是目前2种主要的降维方法,特征提取即运用数学方法对样本波段进行压缩处理,将大部分有用信息压缩在某些波段区间内,然后剔除信息量较少的波段,本研究采用主成分分析法(Principal component analysis,PCA)^[18];特征选择即从获得的波段数据中提取出最有效的波段数据,本研究采用连续投影算法(Successive projection algorithm,SPA)^[19-20]。

表1 不同预处理方法正确分类率

Table 1 Accuracy of different pretreatment methods

预处理方法	径向基核函数的样本正确分类率(%)	线性核函数的样本正确分类率(%)
归一化+原始数据	69.3	68.0
归一化+S-G	66.7	69.3
归一化+MSC	82.6	69.3
归一化+SNV	100.0	98.6

1.6.2.1 基于PCA的特征提取 对3个不同活力梯度的黄瓜种子进行PCA分析。前7个主成分PC1~PC7的累计贡献率分别为87.0%、97.4%、98.1%、98.7%、99.0%、99.2%、99.3%。说明PC1、PC2能够表达绝大部分的光谱信息。图4为PC1、PC2的得分分布图。从图4可以看出,3个不同活力梯度的黄瓜种子分类趋势较明显,说明PC1、PC2对于3种不同活力梯度的种子有着较好的分类作用,能够鉴别不同活力的种子。但3种不同活力梯度的种子相邻部分交叉较明显,少数样本产生重叠现象,没有做到完全区分,所以需要对所有样本数据进行进一步处理。

1.6.2.2 基于SPA的特征选择 SPA是一种前向波长提取方法,不断循环计算一个波长在其余未选波长上的投影,从而找到包含冗余信息量最少的波长^[21-24]。这种方法能较好地降低输入数据组的共线性。运用matlab2014a软件对样本的光谱数据进行SPA特征选择,将401~1000 nm全波段光谱数据通过SPA进行降维,制定波长数 n 的范围为5~30,根据压缩后的验证均方根误差确定光谱的最佳波长数量。由图5可以看出,随着特征波长数量的增加,交叉验证均方根误差不断减小。当选取到9个波长

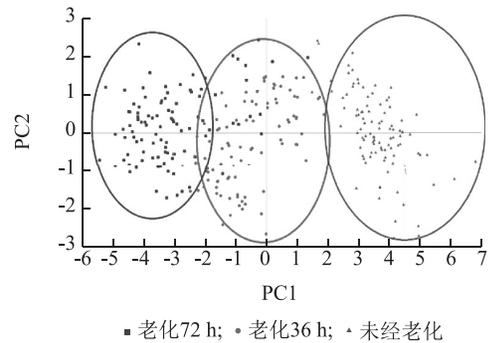


图4 PC1和PC2得分分布图

Fig.4 Score distribution map of PC1 and PC2

的时候,之前明显下降的曲线变成较为平缓,当波长数为12时,曲线又有一个较陡的下降,波长数到达13之后,曲线变得更为平缓,之后的曲线没有太大的起伏波动,即交叉验证均方根误差也趋于平缓。前13个特征波长分别为447 nm、448 nm、449 nm、539 nm、667 nm、715 nm、717 nm、721 nm、746 nm、751 nm、753 nm、889 nm、907 nm。

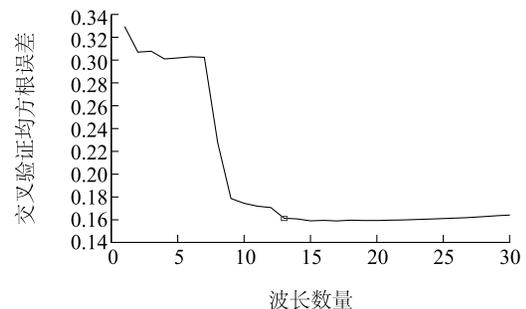


图5 验证均方根误差变化图

Fig.5 Variations of root mean square error

2 结果与分析

建模试验分为2个部分。一部分运用SVM对主成分分析法提取的前5个主成分数据进行建模,另一部分运用SVM对连续投影算法选出的13个特征波长及数据进行建模。试验中,SVM的核函数分别采用了线性核函数和径向基核函数。经过归一化和SNV处理后的高光谱数据作为输入量,对其进行主成分分析。将SVM作为分类器,在1~5之间不断改变主成分数量,根据最后的识别正确率确定最优主成分数。图6显示,主成分数为3时,当采用基于径向基核函数的SVM时,分类正确率达到98.6%,主成分数继续增加,最后分类正确率没有变

化,此时基于线性核函数的主成分-支持向量机模型(PCA-SVM)最佳主成分数为3。当采用线性核函数的SVM建模时,主成分数为4时,分类正确率达到97.3%。当主成分继续增加,分类正确率保持不变,此时基于RBF核函数的主成分-支持向量机模型(PCA-SVM)最佳主成分数为4。

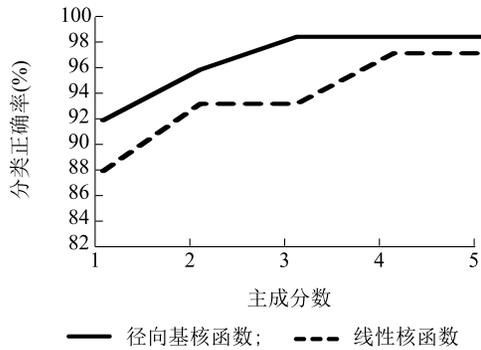


图6 不同主成分数的SVM模型分类正确率

Fig.6 Classification accuracy of SVM models with different principal component number

分别选择9个特征波长、12个特征波长、13个特征波长作为模型的输入量,使用分别基于径向基核函数和线性核函数的SVM模型,对比两者的正确分类率。由图7可以看出,基于RBF核函数的连续投影-支持向量机模型(SPA-SVM)的分类效果高于基于线性核函数的SVM模型。当选取9个特征波长作为输入量时,基于径向基核函数的SVM分类模型分类正确率达到97.3%,略高于基于线性核函数的SVM模型。当选取12个特征波长时,两个模型的正确分类率均有提高,基于径向基核函数的SVM分类模型正确分类率达到98.6%,基于线性核函数的SVM模型正确分类率达到97.3%。当选取13个特征波长时,两个模型的正确分类率都没有提升。因此特征波长的最佳数量为12,SPA-SVM(径向基核函数)模型分类效果最优。

3 结论

针对黄瓜种子活力检测的问题,本研究利用高光谱采集系统,采取了3组不同老化程度的黄瓜种子共300个高光谱图像。利用不同预处理方法对得到的高光谱数据进行处理,选择效果最好的SNV。采用主成分分析法(PCA)和连续投影算法(SPA)对高光谱数据进行特征降维。采用网格搜寻法,确定

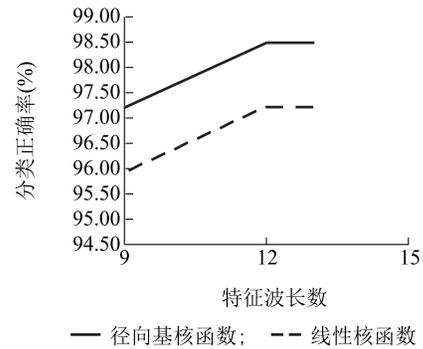


图7 不同特征波长的SVM分类正确率

Fig.7 Classification accuracy of SVM models with different characteristic wavelengths

模型的惩罚因子 $c=4$ 和径向基核函数中的参数 $g=21.1121$,进一步优化模型参数。分别讨论了PCA-SVM(线性核函数)、SPA-SVM(线性核函数)、PCA-SVM(径向基核函数)、SPA-SVM(径向基核函数)的4种分类模型。从测试集的正确率来看,这几种模型都具有较理想的分类效果,分类正确率均达到97.3%以上,其中SPA-SVM(径向基核函数)分类效果最为理想,测试集正确率达到98.6%。可见,运用高光谱技术可无损、高效地鉴别不同活力的黄瓜种子。

参考文献:

- [1] 雷雨,韩德俊,曾庆东,等.基于高光谱成像技术的小麦条锈病害程度分级方法[J].农业机械学报,2018,49(5):233-239.
- [2] 朱大洲,王坤,周光华,等.单粒大豆的近红外光谱特征及品种鉴别研究[J].光谱学与光谱分析,2010,30(12):3217-3221.
- [3] 姚侠妹,张瑞娥,偶春,等.人工老化处理对桔梗种子生理生化特性的影响[J].西北农林科技大学学报(自然科学版),2015,43(2):203-209.
- [4] 张瑛,滕斌,吴敬德,等.水稻种子高温高湿人工加速老化试验方法研究[J].中国粮油学报,2010,25(10):8-12.
- [5] 崔鸿文.黄瓜种子人工老化过程中某些生理生化规律研究[J].西北农林科技大学学报(自然科学版),1992,20(1):51-54.
- [6] 梁琨,杜莹莹,卢伟,等.基于高光谱成像技术的小麦籽粒赤霉病识别[J].农业机械学报,2016,47(2):309-315.
- [7] HUANG M, TANG J, YANG B, et al. Classification of maize seeds of different years based on hyperspectral imaging and model updating[J]. Computers & Electronics in Agriculture, 2016, 122:139-145.
- [8] WAKHOLI C, KANDPAL L M, LEE H, et al. Rapid assessment

- of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics[J]. *Sensors and Actuators B: Chemical*, 2017, 255:498-507.
- [9] AMBROSE A, KANDPAL L M, KIM M S, et al. High speed measurement of corn seed viability using hyperspectral imaging[J]. *Infrared Physics & Technology*, 2016, 75:173-179.
- [10] NANSEN C, ZHAO G, DAKIN N, et al. Using hyperspectral imaging to determine germination of native Australian plant seeds[J]. *J PhotochemPhotobiol B*, 2015, 145:19-24.
- [11] 姚侠妹,张瑞娥,偶春,等.人工老化处理对桔梗种子生理生化特性的影响[J].*西北农林科技大学学报(自然科学版)*, 2015, 43(2):203-209.
- [12] 张帅堂,王紫烟,邹修国,等.基于高光谱图像和遗传优化神经网络的茶叶病斑识别[J].*农业工程学报*, 2017, 33(22):200-207.
- [13] 马惠玲,蔡骋,王若琳,等.一种基于高光谱成像的苹果品种快速鉴别方法[J].*农业机械学报*, 2017, 48(4):305-312.
- [14] 赵欣欣,宋丽荣,权薇竹,等.高温高湿老化处理对不同豆类种子活力及生长能力的影响[J].*作物杂志*, 2017(5):168-172.
- [15] SHRESTHA S, MATEJ K, ZIBRAT U, et al. Single seed near-infrared hyperspectral imaging in determining tomato (*Solanum lycopersicum* L.) seed quality in association with multivariate data analysis[J]. *Sensors and Actuators B: Chemical*, 2016, 237:1027-1034.
- [16] KANDPAL L M, LOHUMI S, KIM M S, et al. Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds[J]. *Sensors & Actuators B Chemical*, 2016, 229:534-544.
- [17] 孙俊,路心资,张晓东,等.基于高光谱图像的红豆品种GA-PNN神经网络鉴别[J].*农业机械学报*, 2016, 47(6):215-221.
- [18] 孙俊,金夏明,毛罕平,等.高光谱图像技术在掺假大米检测中的应用[J].*农业工程学报*, 2014, 30(21):301-307.
- [19] 张初,刘飞,章海亮,等.近地高光谱成像技术对黑豆品种无损鉴别[J].*光谱学与光谱分析*, 2014, 34(3):746-750.
- [20] 袁莹,王伟,褚璇,等.光谱特征波长的SPA选取和基于SVM的玉米颗粒霉变程度定性判别[J].*光谱学与光谱分析*, 2016, 36(1):226-230.
- [21] 孙俊,张梅霞,毛罕平,等.基于高光谱图像的桑叶农药残留种类鉴别研究[J].*农业机械学报*, 2015, 46(6):251-256.
- [22] 李波,刘占宇,黄敬峰,等.基于PCA和PNN的水稻病虫害高光谱识别[J].*农业工程学报*, 2009, 25(9):143-147.
- [23] 刘飞,张帆,方慧,等.连续投影算法在油菜叶片氨基酸总量无损检测中的应用[J].*光谱学与光谱分析*, 2009, 29(11):3079-3093.
- [24] 章海亮,刘雪梅,何勇. SPA-LS-SVM检测土壤有机质和速效钾研究[J].*光谱学与光谱分析*, 2014, 34(5):1348-1351.

(责任编辑:张震林)