

贮藏大米中霉菌菌落总数的高光谱检测研究

丛孙丽, 刘晨, 鹿存莉, 高伟 (无锡太湖学院, 江苏省物联网应用技术重点建设实验室, 江苏无锡 214064)

摘要 [目的]为了快速、精确检测贮藏大米中的霉菌菌落总数,拟用高光谱图像技术实现无损检测。[方法]采用SG-SNV detrending的方式对原始光谱数据预处理,去除基线散射,平滑光谱曲线;然后分别利用SPA算法和CARS算法选取反映大米霉菌菌落总数特性的特征波长组合,最后采用SVR方法分别在全光谱数据和2种特征光谱数据的基础上建立预测模型,对比分析各SVR模型的预测效果。[结果]基于CARS特征选择的模型(CARS-SVR)预测效果优于基于全光谱数据的SVR模型和基于SPA特征选择的模型(SPA-SVR),其预测集决定系数(R_p^2)为0.8759、均方根误差(RMSEP)为0.0835。由于CARS-SVR模型的预测效果尚未达到农产品检测的精度要求,故引入GWO算法对SVR模型中的参数(c 和 g)寻优,优化后模型(CARS-GWO-SVR)的训练集和测试集决定系数(R_c^2 和 R_p^2)分别达0.9621和0.9511。[结论]利用高光谱技术对贮藏大米中的霉菌菌落总数实施无损检测具有可行性,可为其他农产品的霉菌检测提供参考依据。

关键词 高光谱;贮藏大米;霉菌菌落数;灰狼优化;支持向量回归

中图分类号 S379 文献标识码 A

文章编号 0517-6611(2020)19-0211-04

doi: 10.3969/j.issn.0517-6611.2020.19.055



开放科学(资源服务)标识码(OSID):

Hyperspectral Detection of the Total Number of Mold Colonies in Stored Rice

CONG Sun-li, LIU Chen, LU Cun-li et al (Jiangsu Key Construction Laboratory of IoT Application Technology, Taihu University of Wuxi, Wuxi, Jiangsu 214064)

Abstract [Objective] In order to quickly and accurately detect the total number of mold colonies in stored rice, hyperspectral image technology was proposed to achieve non-destructive detection. [Method] The original spectral data was preprocessed through the method of SG-SNV detrending for purpose of removing the baseline scattering and smoothing spectral curves. Then the combination of feature wavelengths reflecting the mold characteristics was selected by SPA and CARS, respectively. On the basis of feature data and full spectral data, the prediction models of mold colonies in stored rice samples were established through SVR method, and the prediction results of models would be compared and analyzed. [Result] The effect of CARS-SVR model was better than that of SVR model and SPA-SVR model. The R_p^2 and RMSEP of prediction set was 0.8759 and 0.0835. In order to improve the accuracy of CARS-SVR model, GWO algorithm was adopted to optimize c and g parameters in model, after that, the R_c^2 and R_p^2 were improved to 0.9621 and 0.9511. [Conclusion] Hyperspectral imaging technology is feasible for non-destructive detection of the total number of mold colonies in stored rice, providing a promising tool for mold detection of other agricultural products.

Key words Hyperspectral; Stored rice; Mold colonies; GWO; SVR

大米作为人们重要的主食之一,营养较为丰富,若贮藏不当易发生霉变。大米发生霉变开始产生霉菌,严重程度下甚至进而产生真菌毒素,霉菌和真菌毒素直接影响大米的品质、人们的食用安全,造成健康隐患^[1]。霉菌的多少直接反映了大米等食品的劣变程度,因此,寻求一种贮藏大米霉菌含量的快速、准确检测方法显得尤为必要。

常规情况下,大米霉菌计数可采用国标法(GB 4789.15—2016《食品微生物学检验 霉菌和酵母计数》^[2])进行测定。这种方法检测精度可靠性高,但前期样品准备工作复杂,且破坏了检测对象。随着光谱检测技术的快速发展,近红外光谱技术势头迅猛,作为无损检测方法之一在稻谷类霉菌或真菌毒素检测中应用典型^[3-4]。这些研究表明,近红外光谱用于农副产品霉菌或真菌毒素检测是可行的。然而,这一技术的缺点在于主要用于检测样本的平均光谱,并不能提供样本图像中每一像素包含的光谱信息。因此,利用近红外光谱定量检测霉菌的精度较低。高光谱图像技术将光谱和图像技术融于一体,同时提供检测目标的光谱和图像信息,在农产品无损检测中应用较为广泛。一些学者已经将高

光谱图像技术用于作物或农产品的霉菌或真菌毒素定性检测^[5-7],但相关定量研究仍较少。Tekle等^[8]获取了镰刀菌损伤的燕麦的高光谱图像,并建立偏最小二乘支持向量回归模型估算其中脱氧雪腐烯醇(DON)的含量,最终预测值与实测值之间的相关系数为0.82,预测精度较低。霉菌不仅会改变大米的颜色和其他表面特性,也会积累更多脂肪酸代谢物,从而提高大米中的脂肪酸含量。因此,脂肪酸含量的改变可以较好地表征大米霉变的状况^[9]。研究表明,可见光光谱可在样品的表面形态(包括纹理、形状等)的表征上达到较好的效果,但可见光的单波段信息不足以提高检测精度;而大米中脂肪酸含量的变化又会促使近红外光谱特性的变化^[10]。高光谱图像技术主要涵盖了可见光光谱和近红外光谱的波段范围,被视为大米霉菌含量检测最有效的方法。

笔者采用高光谱技术结合支持向量回归(support vector regression, SVR)模型实现大米样品中霉菌菌落总数的检测,为了寻找模型中的最佳参数(c 和 g)采用了新型群体智能算法——灰狼优化(gray wolf optimization, GWO)。通过对大米霉菌的高精度定量检测,可为其他农产品的霉菌检测提供新型方法,也为研制便携霉菌检测仪奠定理论基础。

1 材料与方法

1.1 人工诱导大米样品 试验选用大米为优质金龙鱼大米,购于镇江欧尚超市,确保大米完整、表面光滑无霉变。试

基金项目 国家自然科学基金项目(31471413);无锡市软科学课题(201913571004Z)。

作者简介 丛孙丽(1993—),女,江苏南通人,助教,硕士,从事计算机技术在农业工程方面的应用研究。

收稿日期 2020-04-08

验前须先测定所购大米的初始含水率,根据国标法(GB 5009.3—2016《食品中水分的测定》^[11])实际检测得出结果为11.4%,在大米安全贮藏标准范围内。首先,将大米分成100个样品盘(每盘100g),单独贴上标签。为去除大米中细菌带来的影响,所有大米均需经过消毒,将大米样品置于干净的工作台上,打开紫外线长达30min。灭菌完成后,在无菌状态下进一步诱导大米发生霉变。当储存过程中霉菌含量低于 10^4 CFU/g时,说明大米存放安全;一旦达到 10^5 CFU/g,大米便开始发生霉变;超过 10^6 CFU/g后,霉变则非常严重^[12]。因此,为了模拟贮藏大米霉变的过程,试验采取恒温恒湿箱培育大米样品,其中恒温恒湿箱的温度设定为30℃,相对湿度设定为90%(参照霉菌易于繁殖的温湿度环境)。基于大米贮藏过程中理化特性与感光指标随时间的变异情况,将培育阶段分成10个周期(每个周期15d),得到不同霉变时期内的样品各10份(共100份)。

1.2 高光谱扫描 试验中,采用光谱为871.607~1766.322nm的高光谱成像系统^[13],其中CCD相机与光谱仪共同构成高光谱图像摄影仪(ImSpector V10E, spectra imaging Ltd., Finland),2只150W的卤素灯构成系统的光源,一个电控位移平台及计算机。此系统采集的图像分辨率达到672像素×512像素,光谱分辨率达5nm。试验开始前,预先对高光谱系统进行黑白参考校正,然后将培育的各份大米样品均匀平铺于一宽口的PP试剂瓶瓶盖上,并缓慢放置于电控位移平台正中央的白纸中心,以待对样品进行扫描采集图像。扫描过程中摄像机曝光时间为20ms,平移台输送速度为1.25mm/s。所有样品扫描完成后利用感兴趣区域法将每个样品图像中的光谱数据提取出来,考虑到光谱范围的初始和结束部分均受到了外部噪声的影响,故将前后共40个波段的数据予以剔除,最终波长931~1666nm用于分析。

1.3 霉菌菌落总数检测 高光谱扫描后立即对各份大米样品进行霉菌菌落总数检测,检测方法参照文献[2]。检测过程中观察霉菌的滋生状况以CFU/g为单位进行平板计数。最终,将记录的数值取对数处理,降低量纲的影响,以便建模分析。

1.4 灰狼算法优化的支持向量回归机 支持向量机是1995年提出用于建立模型的方法,该研究中其回归功能SVR用于预测大米霉菌菌落数,一般将径向基核函数(radial basis function, RBF)用于模型。

GWO算法于2012年新提出,提出的依据是灰狼群体的

觅食行为,如捕食期间产生的跟踪、包围、追击等。该算法的机制简单、参数少、收敛速度快,且相对于其他智能算法(DE、PSO、GA等)优化效果更好^[14],也因此风速预测、水量分配、函数优化等领域应用广泛。该研究将GWO算法用于对SVR参数 c 和 g 寻优的具体步骤如下:①划分大米霉菌数据样本集。初始化SVR的参数 c 和 g 的取值区间,并设置GWO算法的相关参数(狼群数量SearchAgents_no以及最大迭代次数Max_iteration)。其中, c 和 g 的取值均为0.01~100.00,SearchAgents_no和Max_iteration取值均为20。②随机化灰狼的群体,各群体位置向量对应于一组参数(c 和 g)。③GWO算法根据适应度值对狼群中每个个体的位置进行更新,最终保留适应度值最优的位置。④判断更新次数是否超过Max_iteration值,若未超过,继续更新,否则更新结束。此时的参数 c 和 g 即最优参数,可利用其建立SVR最优模型进行菌落数预测分析。

2 结果与分析

2.1 大米霉菌菌落数参考值 采用平板计数法测得的霉菌菌落数参考值如表1所示。所有大米样品的霉菌菌落数分布在2.66~5.88lg(CFU/g)。根据大米贮藏的霉菌标准,该试验中霉菌菌落总数分布既有安全阶段,又有大米贮藏的不同霉变阶段,总体上霉菌菌落值符合正态分布。从表1可看出校正集和测试集样品的霉菌统计值。

表1 大米霉菌菌落总数统计值

Table 1 Statistical values of mold colonies number in rice samples

Sample set	Count	Mean	Minimum	Maximum
校正集 Correction set	50	4.82	2.66	5.88
测试集 Test set	50	4.95	3.34	5.85

2.2 数据预处理 依据1:1的比例对高光谱数据进行样本集划分,分成校正集和测试集两部分。多项式平滑(savitzky-golay, SG)、标准变量变换(standard normalized variable, SNV)及去趋势(detrending)均是可用于光谱预处理的方法^[15-17]。其中SG可有效消除基线漂移,SNV可校正光谱散射, detrending一般与SNV结合用于去趋势。考虑到每种预处理方法的差异性,以上3种方法结合(SG-SNV detrending)用于处理并保留光谱数据中的有效信息。预处理前后的光谱曲线如图1所示。

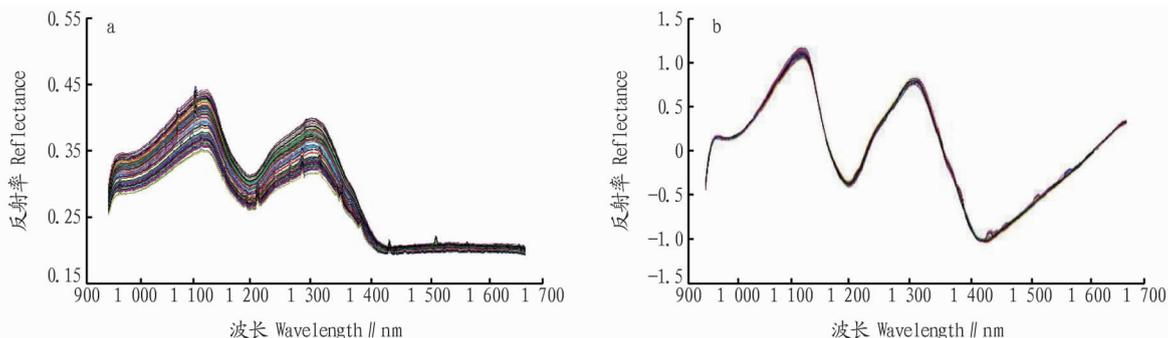


图1 预处理前(a)和预处理后(b)光谱曲线

Fig.1 Spectral curves before (a) and after pretreatment (b)

2.3 特征波长选择

2.3.1 采用 SPA 选择特征波长。连续投影算法 (successive projections algorithm, SPA) 因简单、快速的特点在特征选择上得到了广泛的应用^[18]。该研究将 SPA 波长最大值设为 30, 算法运行过程中不同的波长数会产生不同的均方根误差 (root mean square error, RMSE), 当 RMSE 值变化达到最小值时所产生的波长即最佳。该研究的 RMSE 最小为 0.602, 此时产生了 12 个特征波长。选取的波长如图 2 所示。

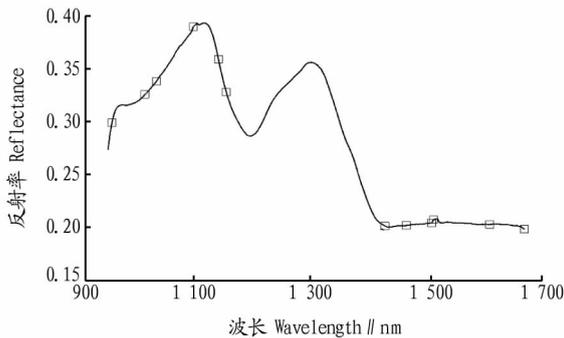


图 2 SPA 选择的特征波长分布

Fig. 2 Characteristic wavelengths distribution of SPA selection

2.3.2 采用 CARS 选择特征波长。竞争性自适应重加权采样 (competitive adaptive reweighted sampling, CARS) 作为特征

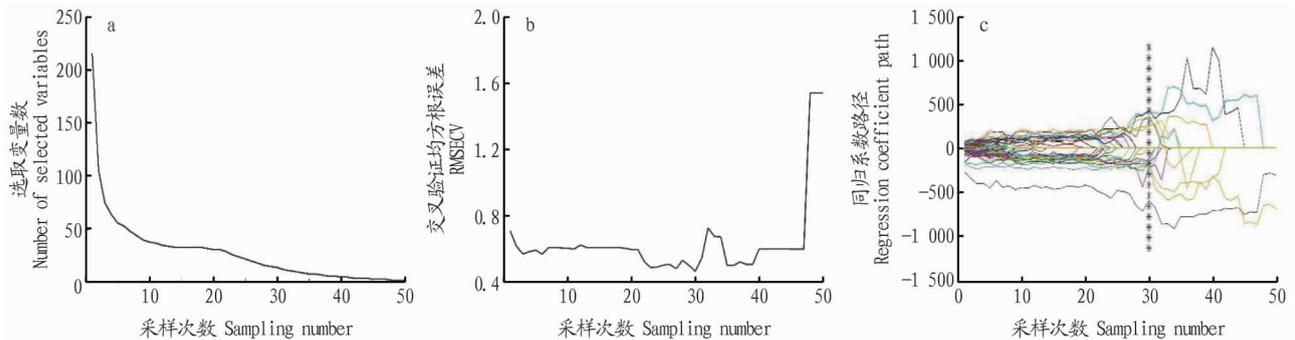


图 3 CARS 选择特征波长的过程

Fig. 3 The process of selecting characteristic wavelengths in CARS

2.4 SVR 模型的建立与分析 将特征光谱和全光谱数据分别用于 SVR 回归模型的输入变量, 大米霉菌菌落总数作为输出变量, 通过校正集样本和测试集样本的预测来评估最终模型的性能。从 SVR 各模型的评价指标对比 (表 2) 可以看出, 与 SVR 模型相比, SPA-SVR 模型虽然是通过 SPA 极大地减少了输入变量, 但性能却有所降低。原因可能是 SPA 算法在运行过程中剔除了部分与霉菌数量相关性较大的关键波长。相比之下, CARS-SVR 模型虽然采用的特征波长数较

表 2 SVR 各模型指标对比

Table 2 Index comparison of SVR models

模型 Mode	波长数 Number of wavelengths	R_c^2	RMSEC	R_p^2	RMSEP
SVR	216	0.978 9	0.010 1	0.691 4	0.149 9
SPA-SVR	12	0.700 7	0.123 8	0.669 0	0.151 3
CARS-SVR	14	0.879 5	0.046 8	0.875 9	0.083 5

选择法之一, 模拟“适者生存”准则, 通过指数衰减和自适应重加权采样 (ARS) 技术选出最终交叉验证均方根误差 (RMSECV) 最小下的特征集合^[19]。该研究设定 CARS 采样次数为 50 及五折交叉验证, 过程如图 3 所示。由图 3a 看出选择的特征波长数随着采样次数增加在逐渐减少; 图 3c 中星号垂线对应的点即为 b 图中 RMSECV 的最小值点 (0.462), 此时的采样次数为 30。此后, RMSECV 值有增加的趋势。因此, 选取采样次数为 30 时的特征波长共 14 个, 分别为 935.96、944.25、1 012.17、1 031.34、1 129.52、1 136.48、1 157.13、1 160.54、1 319.84、1 341.85、1 344.99、1 426.69、1 474.49、1 513.47 nm。

大米中最多的脂肪酸是油酸 (OFA) 和亚油酸 (LFA)^[20]。Mzimiri 等^[21]研究了这 2 种脂肪酸在 900 ~ 1 700 nm 处的高光谱特征, 其中 OFA 的特征波长理论上为 1 146.6、1 313.3 和 1 480.3 nm, 该研究 SPA 选择的波长中 1 146.85 nm 与 OFA 第 1 个理论波长较为接近, 而 CARS 选择的波长中 1 319.84、1 474.49 nm 与第 2、第 3 个较为接近; LFA 的特征波长理论上为 1 036.9 和 1 517.1 nm, 该研究 SPA 选择的波长中 1 031.34 nm 与 LFA 第 1 个理论波长较为接近, 而 CARS 选择的波长中 1 031.34、1 513.47 nm 与 2 个理论波长均较接近。因此, 相比之下 CARS 算法保留了反映大米样品霉菌菌落总数的更多特征信息, 大大降低建模复杂度。

多, 但模型的预测性能和鲁棒性明显更好, 也进一步印证了前面特征波长的分析。因此, 3 种模型中最佳的为 CARS-SVR, 达到了最高预测精度, R_p^2 为 0.875 9 且 RMSEP 为 0.083 5。

2.5 GWO-SVR 模型的建立 为提高 CARS-SVR 模型的检测精度, 在经过 CARS 特征光谱数据的基础上, 引入 GWO 算法优化 SVR 模型中的参数 c 和 g 。然后利用最优参数重新建立特征光谱下的 SVR 模型。优化后模型 (CARS-GWO-SVR) 的各评价指标分别为波长数量 14, $R_c^2 = 0.962 1$, RMSEC = 0.012 6, $R_p^2 = 0.951 1$, RMSEP = 0.036 0, 模型对测试集样本的预测结果如图 4 所示。从图 4 可以看出, CARS-GWO-SVR 模型预测效果得到了明显的提升, P_p^2 达到了 0.95 以上。同时, 与 CARS-SVR 相比, CARS-GWO-SVR 的 RMSEC 和 RMSEP 差异更小, 表明模型鲁棒性和预测能力更强。与 Tekle 等^[8]关于利用高光谱图像估算燕麦中 DON 含量的相比, 该研究大米霉菌含量的预测精度更高, 表明 GWO 算法确实优化了模型的性能。

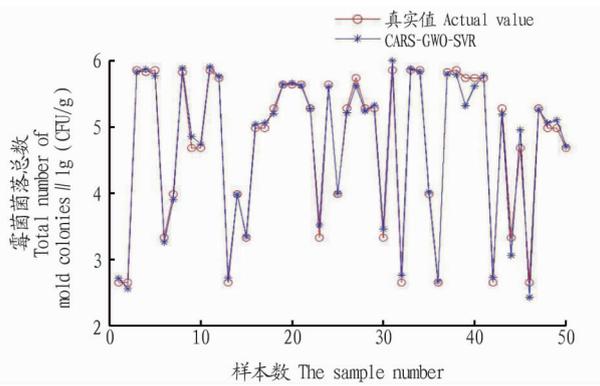


图4 CARS-GWO-SVR模型预测结果

Fig. 4 Predicted results of CARS-GWO-SVR model

3 结论

该研究培育了不同霉变程度的大米样品并采集了高光谱图像,通过感兴趣区域选取方式获取了32像素×32像素的光谱信息。在预处理之后的光谱基础上,比较了基于全光谱数据和不同特征光谱数据的SVR模型建模效果。结果表明:CARS-SVR模型获得了比全光谱模型和SPA-SVR模型更好的预测效果,其训练集 R_c^2 和RMSEC分别为0.8795、0.0468;测试集 R_p^2 和RMSEP分别为0.8759、0.0835。利用GWO算法对SVR模型中的参数 c 和 g 进行寻优,重新建立基于CARS特征光谱数据的SVR预测模型。最终,CARS-GWO-SVR模型的 R_c^2 和 R_p^2 分别提升为0.9621和0.9511,满足农产品的检测要求。

综上所述,该研究将高光谱图像技术用于大米霉菌菌落总数检测具有可行性,为其他农产品的霉菌检测提供新型方法,也为研制便携霉菌检测仪奠定理论基础。不足之处在于用于研究的样品数量有待提高以验证是否适用于更多种类的大米霉菌检测,有待进一步研究。

参考文献

[1] 周显青, 伦利芳, 张玉荣, 等. 大米储藏与包装的技术研究进展[J]. 粮油食品科技, 2013, 21(2): 71-75.
 [2] 中华人民共和国国家卫生和计划生育委员会. 食品微生物学检验 霉菌和酵母计数: GB 4789.15—2016[S]. 北京: 中国标准出版社, 2017.
 [3] 张强, 刘成海, 孙井坤, 等. 基于支持向量机的稻谷黄曲霉毒素 B_1 近红外无损检测[J]. 东北农业大学学报, 2015, 46(5): 84-88.

[4] 金昌福, 张强, 郑先哲. 贮藏稻谷霉菌菌落总数近红外光谱预测模型[J]. 农机化研究, 2016(10): 160-164.
 [5] FENG L, ZHANG D R, CHEN S S, et al. Early detection of gray mold on eggplant leaves using hyperspectral imaging technique[J]. Journal of Zhejiang University (Agriculture and life sciences), 2012, 38(3): 311-317.
 [6] DEL FIORE A, REVERBERI M, RICELLI A, et al. Early detection of toxigenic fungi on maize by hyperspectral imaging analysis[J]. International journal of food microbiology, 2010, 144(1): 64-71.
 [7] XIAO H, WANG Z J, SUN Y, et al. Growth simulation and discrimination of five fungi from rice storage using hyperspectral reflectance imaging technique[J]. Science and technology of food industry, 2016, 37(13): 276-281.
 [8] TEKLE S, MAGE I, SEGTNAN V H, et al. Near-infrared hyperspectral imaging of Fusarium-damaged oats (*Avena sativa* L.)[J]. Cereal chemistry, 2015, 92(1): 73-80.
 [9] 文韬, 洪添胜, 李立君, 等. 基于高光谱技术的霉变稻谷脂肪酸含量无损检测[J]. 农业工程学报, 2015, 31(18): 233-239.
 [10] NING J M, WAN X C, ZHANG Z Z, et al. Discriminating fermentation degree of Pu'er tea based on NIR spectroscopy and artificial neural network[J]. Transactions of the Chinese society of agricultural engineering, 2013, 29(11): 255-260.
 [11] 中华人民共和国国家卫生和计划生育委员会. 食品中水分的测定: GB 5009.3—2016[S]. 北京: 中国标准出版社, 2017.
 [12] 周建新, 鞠兴荣, 孙肖东, 等. 不同储藏条件下稻谷霉菌区系演替的研究[J]. 中国粮油学报, 2008, 23(5): 133-136.
 [13] 张晓东, 毛罕平, 周莹, 等. 基于高光谱成像技术的生菜叶片水分检测研究[J]. 安徽农业科学, 2011, 39(33): 20329-20331, 20714.
 [14] MIRJALILI S, HASHIM S Z M. A new hybrid PSO-GSA algorithm for function optimization[C]//International conference on computer and information application. Tianjin, China: IEEE, 2012: 374-377.
 [15] TAHIR H E, ZOU X B, LI Z H, et al. Rapid prediction of phenolic compounds and antioxidant activity of Sudanese honey using Raman and Fourier transform infrared (FT-IR) spectroscopy[J]. Food chemistry, 2017, 226: 202-211.
 [16] CHO J S, BAE H J, CHO B K, et al. Qualitative properties of roasting defect beans and development of its classification methods by hyperspectral imaging technology[J]. Food chemistry, 2017, 220: 505-509.
 [17] POLLOCK D S G. Trend estimation and de-trending via rational square-wave filters[J]. J. Econometrics, 2000, 99(2): 317-334.
 [18] 董金磊, 郭文川. 采后猕猴桃可溶性固形物含量的高光谱无损检测[J]. 食品科学, 2015, 36(16): 101-106.
 [19] 刘燕德, 施宇, 蔡丽君, 等. 基于CARS算法的脐橙可溶性固形物近红外在线检测[J]. 农业机械学报, 2013, 44(9): 138-144.
 [20] 王立峰, 陈超, 杨精华, 等. 大米小包装储藏过程中脂类及微观结构的变化[J]. 中国农业科学, 2015, 48(9): 1807-1817.
 [21] MZIMBIRI R, SHI A M, LIU H Z, et al. A review: Peanut fatty acids determination using hyper spectroscopy imaging and its significance on food quality and safety[J]. Food science & quality management, 2014, 28: 90-98.