

基于高光谱信息特征选择的玉米霉变程度 Fisher 鉴别方法

Fisher discriminant analysis for moldy degrees of maize samples
based on the feature selection of hyperspectral data

戴松松 殷 勇

DAI Song-song YIN Yong

(河南科技大学食品与生物工程学院, 河南 洛阳 471023)

(College of Food and Bioengineering, Henan University of Science and Technology, Luoyang, Henan 471023, China)

摘要:为了提高高光谱鉴别玉米霉变程度的正确率,分别对全波长和特征波长下霉变玉米进行鉴别分析。利用高光谱图像采集系统获得250个霉变玉米样本的高光谱数据,并用标准正态变量变换(standard normal variate, SNV)和多元散射校正(Multiplicative scatter correction, MSC)2种方法对原始数据进行预处理,再对未预处理和预处理后的原始数据进行判别,优选出多元散射校正的预处理方法;运用偏最小二乘回归系数选择了9个特征波长;运用Fisher判别分析(Fisher discriminant analysis, FDA)分别对全波长和特征波长下的训练集进行判别分析,并用对应的测试集进行检验。FDA结果表明,全波长下判别模型的训练集和测试集的准确率分别为97.71%, 97.33%, 9个特征波长下训练集和测试集的准确率分别为100.00%, 98.67%。研究表明,利用特征光谱能够较好地表征玉米的霉变程度,有利于提高玉米霉变程度的鉴别正确率。

关键词:高光谱;霉变玉米;特征波长;Fisher判别分析

Abstract: In order to improve the identification accuracy of moldy degrees of maize samples using hyperspectral, the identification effects of moldy maize at the full wavelength and the characteristic wavelength were investigated in this study, respectively. Firstly, the hyperspectral data of 250 moldy maize samples were obtained by hyperspectral image acquisition system, and the standard normal variate (SNV) and multiplicative scatter correction (MSC) were employed to preprocess the original data; and then the MSC was adopted by comparing the results of the preprocessed and non-preprocessed data. Secondly, nine characteristic wavelengths were selected by using the

partial least squares regression coefficients. Finally, Fisher discriminant analysis (FDA) was used to analyze the training set at full wavelengths and characteristic wavelengths, and examined by the corresponding test set. The FDA results showed that the accuracy of the training set and test set were 97.71% and 97.33% for full wavelengths case, respectively, and were 100.00% and 98.67% for characteristic wavelengths case, respectively. The accuracy of the training set and test set at 9 characteristic wavelengths were 100.00% and 98.67% respectively. The research findings showed that the characteristic wavelengths could be used to represent the moldy degrees of maize samples, which was helpful to improve the identification correct rate of moldy maize. Moreover, the research findings might provide a reference for identifying the other agricultural products using hyperspectral technology.

Keywords: hyperspectral; moldy maize; characteristic wavelength; fisher discriminant analysis

玉米极易被霉菌感染,发生霉变,人畜食用发霉的玉米,毒素会在体内蓄积,产生严重危害。目前,检测霉变玉米的方法主要有感官评定^[1]、高效液相色谱法(HPLC)^[2]、聚合酶链式反应(PCR)^[3]、气质联用法(GC-MS)^[4]及DNA探针法^[5]、酶联免疫(ELISA)法^[6-7]等,这些方法虽然比较准确,但缺点是存在主观性强、结果可靠性差,而且破坏物料、操作复杂、成本高等,不能满足玉米贮藏和运输过程中快速检测霉变的需要。相比之下,高光谱技术具有高效、无损、低成本和可重复等优势^[8],已广泛应用于烟叶^[9]、干果^[10-11]、水果^[12-13]、油^[14]、肉^[15-16]等农副产品检测中,并且取得了较好的分析结果。近年来,高光谱技术在玉米霉变检测中也有报道。袁莹等^[17]利用傅里叶变换近红外光谱对不同霉变程度的玉米颗粒进行检测区分,以支持向量机作为判别模型,其分类模型得出的训练集和预测集的预测准确率分别为93.3%, 91.7%。张楠楠等^[18]利用高光谱图像技术对霉变玉

基金项目:河南省科技攻关项目(编号:172102210256)

作者简介:戴松松,女,河南科技大学在读硕士研究生。

通信作者:殷勇(1966—),男,河南科技大学教授,硕士生导师,博士。

E-mail: yinyong@haust.edu.cn

收稿日期:2017-10-11

米进行检测,其正确检出率为93.75%。

已有的研究成果表明,高光谱技术有能力对霉变玉米进行鉴别,但正确率不高,而且检测时均需往新鲜玉米中注射不等量的真菌毒素,来培育不同等级的霉变玉米,不具有随机性。本研究以5种发霉程度不同的玉米(霉变玉米是实验室模拟毒素最适宜生长时的温度和湿度来培育的霉变玉米,并由商丘市质量技术监督检验测试中心测出毒素值)为研究对象,利用高光谱图像采集系统采集这些霉变玉米的高光谱图像,提取各个样本的平均光谱反射值,用标准正态变量变换和多元散射校正2种预处理方法对原始平均光谱进行预处理,再对预处理后的数据和未预处理的原始数据进行FDA判别,并比较三者的FDA结果;为了降低光谱维度,采用偏最小二乘回归系数法选取特征波长,建立特征波长下的Fisher判别模型,并考察全波长与特征波长的判别效果,以期为高光谱成像技术应用于玉米霉变的在线分级提供参考。

1 材料与方法

1.1 试验材料

玉米样品:中单909,形状和大小基本一致,购于洛阳市中原农贸城。由于玉米本身自带芽孢和孢子,芽孢是细菌的休眠体,孢子由霉菌产生,它们在适宜的生长环境下可使玉米产生霉变^[1],在实际试验中,发现玉米样品的发霉程度随着培养时间的延长而相应发生变化(表1),因此采用不同的培养时间来表征玉米样品的不同发霉等级。方法为:将新鲜玉米放入相对湿度为90%,温度为30℃的恒温恒湿培养箱(LHS-HC-100型,上海资一仪器设备有限公司)中培育。采用不同培育天数的玉米来表征不同的霉变程度。试验选用培育0,2,4,6,8 d的霉变玉米,每个等级的样品各取50个样本(35个训练集,15个测试集),每个样本量为(60.0±0.5) g。这些霉变等级样本委托商丘市质量技术监督检验测试中心测定(每个霉变等级测试5个平行样本),结果见表1。

表1 发霉玉米中3种霉菌测量结果

Table 1 Moldy maize three kinds of mold measurement results

培育天数/d	玉米赤霉烯酮/ (μg·kg ⁻¹)	脱氧雪腐镰刀菌 烯醇/(μg·kg ⁻¹)	黄曲霉毒素 B ₁ / (μg·kg ⁻¹)
0	5.6	81.1	3.78
2	24.8	406.8	8.96
4	50.2	880.5	16.68
6	70.8	1 100.9	30.18
8	81.3	1 208.6	39.96

1.2 高光谱采集系统

高光谱图像采集系统,是由成像光谱仪(IST 50-3810型,德国Inno-spec公司)、计算机、4个500 W的光纤卤素灯(RK90000420108型,德国ESYLUX公司)和传送装置等组成。可采集到的反射光谱范围为371.05~1 023.82 nm,光谱分辨率2.8 nm,共得到1 288个不同的波段。高光谱扫描设定范围为Width760、Heigh550。得到的高光谱图像分辨率

为760×550。因此,最终得到的每个样本大小为760×550×1 288的高光谱图像。

1.3 高光谱数据采集与校正

采集样本高光谱数据时,将样本均匀地平铺在规格为8 cm×8 cm的培养皿中,然后将盛有样品的培养皿放在搭建的传送带上,由SICap-STVR V1.0.x软件平台驱动控制成像仪,采集样本光谱大小在371.05~1 023.82 nm的反射光谱。然后对采集的高光谱图像信息进行黑白校正,校正方法详见文献[19]。

针对所采集的高光谱图像,用ENVI 4.8提取每个样本的所有像素点在各波长下的平均反射值,以此得到250个样本的平均反射光谱数据,然后通过MATLAB R2014a软件对光谱数据进行处理。

1.4 光谱预处理

高光谱图像采集时,受到电路噪声以及随机误差等因素的影响,噪音是光谱信号探测时无法避免的现象^[20]。因此,需对高光谱数据进行预处理;采用标准正态变量变换和多元散射校正2种预处理方法对原始光谱进行预处理,并将不预处理的原始数据与2种预处理后的数据作为Fisher判别模型的输入,进行初步的判别。

1.4.1 标准正态变量变换处理 高光谱在采集图像过程中,由于霉变玉米的颗粒度、表面散射等原因,会在采集过程中产生噪声,因此,选用标准正态变量变换预处理方法来减弱这种噪声。SNV的方程式为:

$$X_{a,SNV} = \frac{X_a - \bar{X}_a}{\sqrt{\frac{1}{p-1} \sum_{i=1}^p (X_{ai} - \bar{X}_a)^2}}, \tag{1}$$

式中:

X_a ——第 a 个样本的光谱向量;

$\bar{X}_a$ ——光谱向量平均值;

p ——波长变量数。

1.4.2 多元散射校正处理 多元散射校正法可以增强与成分含量相关的光谱吸收信息,消除了光谱数据中散射带来的影响。使用这种方法的前提是要先建立一个样品的“理想光谱”,即样品中待测成分的含量与光谱变化满足直接线性关系,以该光谱为标准对其他样品光谱通过基线平移和偏移校正进行修正,首先计算样品高光谱的平均值,然后以平均值作为标准光谱,将样品光谱与标准光谱进行一元线性回归运算,最后进行多元散射校正^[21],公式如下:

$$\bar{A}_{i,j} = \frac{1}{n} \sum_{i=1}^n A_{i,j}, \tag{2}$$

$$A_i = m_i \bar{A} + b_i, \tag{3}$$

$$A_{i(MSC)} = \frac{A_i - b_i}{m_i}, \tag{4}$$

式中:

A —— $n \times p$ 维光谱数据矩阵;

n ——样品数;

p ——光谱采集所用的波长点数;

$\bar{A}$ ——所有样品的原始高光谱在各个波长点处求平均值所得到的平均光谱矢量;

A_i —— $1 \times p$ 维矩阵;

m_i, b_i ——分别表示各样品高光谱 A_i 与平均光谱 $\bar{A}$ 进行一元线性回归后得到的相对偏移系数和平移量。

1.5 光谱特征波长提取

由于采集到的高光谱信息是高达 1 288 个波段的数据,再加上光谱数据中相邻光谱值存在极高的相关性,含有大量的冗余信息,使得特征吸收峰不明显。所以特征波长的提取对减少计算量、提高检测精度至关重要。因此,通过偏最小二乘回归系数法对 1 288 个波段上样本的平均光谱反射值进行分析,借助于回归系数的大小来提取特征波长。

由于表达式中每个波段的回归系数表征了各波段的贡献比重,系数绝对值越大对回归模型的影响越大。因此,通过比较回归系数绝对值的大小可以选取出光谱的特征波长。

1.6 Fisher 判别分析

为了有效地鉴别发霉玉米的霉变程度,选用 Fisher 判别分析作为模式识别的工具,其基本的思想是投影,即将 k 组 m 元数据投影到某一方向,投影后组和组之间尽可能地区分开,借助方差分析思想来导出判别函数,这个判别函数可以是线性的,也可以是比较一般的函数。在实际应用中,因为线性判别函数是最为方便的,所以选用线性判别函数。为探究采用 FDA 分析全波长和特征波长下不同等级霉变玉米的有效性,采用本实验室在 MATLAB R2014a 平台上自行研制的 FDA 分析程序,分别对全波长和特征波长下的 5 种玉米训练集及其对应的测试集进行分析。其 FDA 的相关应用参见文献[1]。

2 结果与分析

2.1 霉变玉米的光谱特征

由图 1 可知,培育 0,2,4 d 的发霉玉米,随着培育时间的延长,平均光谱反射值逐渐降低;培育 6,8 d 时,平均光谱反射值没有降低反而变的更高了,可能是与霉变玉米中的某些霉变产物有关。总体来看,不同等级霉变玉米的平均光谱曲线总体变化规律具有相似性和可分性,这为建立 Fisher 判别模型提供了一定的基础。

2.2 原始光谱预处理

随机选取 175 个玉米样本为训练集,75 个样本为测试集,把原始光谱数据和不同预处理后的光谱数据,分别作为建立 Fisher 判别模型的输入,结果见表 2。由表 2 可知,经过 MSC 预处理的光谱 FDA 判别模型的正确识别率最高,其训练集和测试集的识别率分别为 0.977 1,0.973 3;经 MSC、SNV 预处理后的训练集和测试集识别率均高于原始光谱,因此,选用 MSC 为原始光谱预处理方法。

2.3 特征波长的选取

由图 2 可知,可选出 9 个特征波长(538.5,583,624.5,649.6,699.1,767.3,759.2,842.7,932.6 nm),这些特征波长能较好地体现出大部分原光谱信息。因此,在建立特征波长下的 Fisher 判别模型时,选择这 9 个波长作为模型的输入。

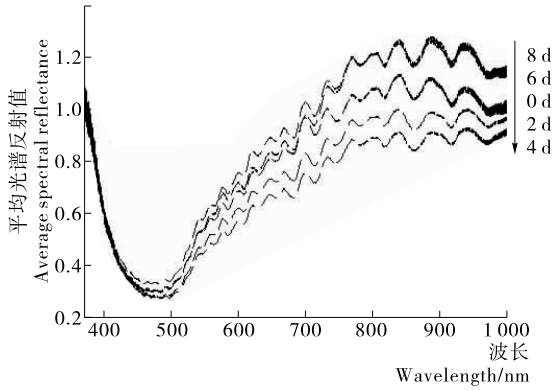


图 1 5 种玉米样本平均光谱反射值曲线
Figure 1 The average spectral reflectance curve of five kinds of maize samples

表 2 原始光谱与预处理光谱的 FDA 模型比较
Table 2 Comparison of FDA model between raw and pretreated spectra

类别	正确识别率	
	训练集	测试集
原始光谱	0.954 7	0.943 9
SNV	0.967 3	0.958 9
MSC	0.977 1	0.973 3

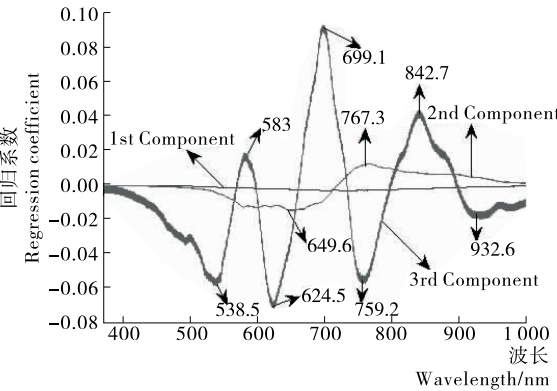


图 2 偏最小二乘回归系数图
Figure 2 Coefficient map of partial least squares regression

2.4 玉米霉变等级 FDA 鉴别分析

2.4.1 全波长下的 FDA 分析 随机选取 175 个玉米样本为训练集,75 个样本为测试集,以 MSC 预处理后的全波长光谱信息作为 Fisher 判别模型的输入,其 FDA 鉴别结果见图 3,FDA 的鉴别正确率见表 3。由图 3(a)可知,全波长下的不同程度霉变玉米大部分能够区分开来,但培育 2,4 d 与培育 6,8 d 的仍有小部分重叠,FDA 测试结果比较分散。由表 3 可知,培育 2 d 的 35 个玉米样本,有 1 个样本错分为培育 4 d,判别准确率为 97.14%;培育 6 d 的 35 个玉米样本,有 1 个样本错分为培育 8 d,判别准确率为 97.14%;而培育 8 d 有 2 个样本错分为培育 6 d,判别准确率为 94.29%。由图 3(b)可知,不同程度霉变玉米大部分也能够区分开来,培育 2,4 d 与培育 6,8 d 的也有小部分重叠。由表 3 可知,培育 2 d 的 15 个玉米样本,有 1 个样本错分为培育 4 d,判别准确率为 93.33%,培育 6 d 的 15 个玉米样本有 1 个样本错分

为培育8 d,判别准确率为93.33%。全波长下的FDA判别模型训练集和测试集的整体判别准确率分别为97.71%,97.33%,因此,在鉴别分析中运用全波长能较好地鉴别不同等级的霉变玉米。

2.4.2 特征波长下的FDA分析 为了减少计算量、提高检测精度,从1 288个全波长中提取了9个特征波长(538.5,583.624.5,649.6,699.1,767.3,759.2,842.7,932.6 nm)作为Fisher判别模型的输入,其FDA鉴别结果见图4,FDA的鉴别正确率见表4。由图4可知,特征波长下的不同程度霉变玉米都能够区分开来,FDA测试结果中每个等级也比较聚集。从表4中的训练集和测试集的总体判别结果分别为

100.00%,98.67%。可以得出,建立的特征波长下的判别模型能够有效地鉴别玉米的霉变程度,与全波长下的FDA鉴别结果相比,特征波长下的鉴别能力明显提高,而且特征波长判别模型的变量降到9个,远低于全波长下的1 288个变量,提高了计算速度。

3 结论

为了提高高光谱鉴别玉米霉变程度的正确率,在提取特征波长的基础上,分别研究了全波长和特征波长下的Fisher判别模型的判别效果。结果显示,在全波长和特征波长下均能有效地鉴别玉米霉变程度,而特征波长的鉴别效果优于全波长的,鉴别正确率可以达到98.67%。表明基于特征选择

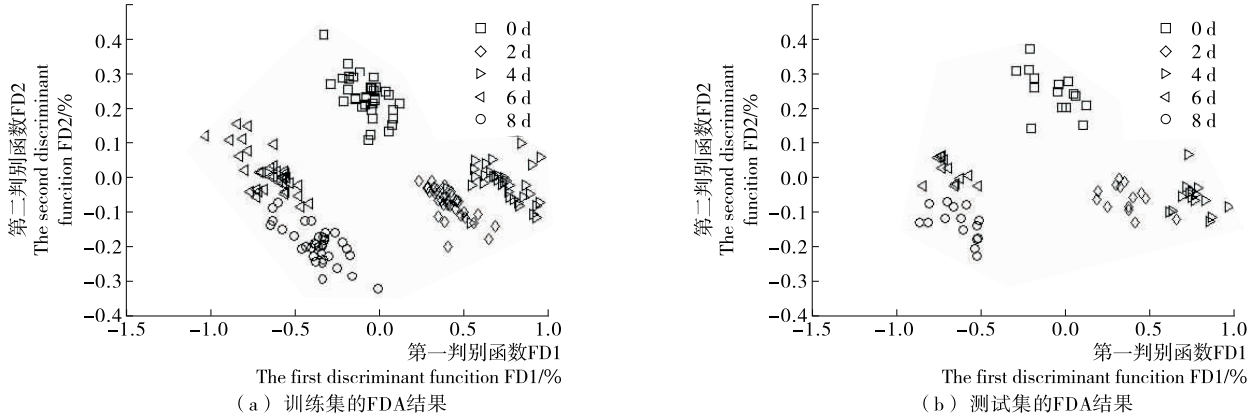


图3 基于全波长下的FDA结果

Figure 3 FDA results based on the full wavelengths

表3 全波长下的Fisher判别模型判别准确率

Table 3 accuracy of Fisher discriminant model under full wavelength

实际培育 天数/d	判别培育天数/d											
	训练集						测试集					
	0	2	4	6	8	准确率/%	0	2	4	6	8	准确率/%
0	35	0	0	0	0	100.00	15	0	0	0	0	100.00
2	0	34	1	0	0	97.14	0	14	1	0	0	93.33
4	0	0	35	0	0	100.00	0	0	15	0	0	100.00
6	0	0	0	34	1	97.14	0	0	0	14	1	93.33
8	0	0	0	2	33	94.29	0	0	0	0	15	100.00
总体准确率/%	97.71						97.33					

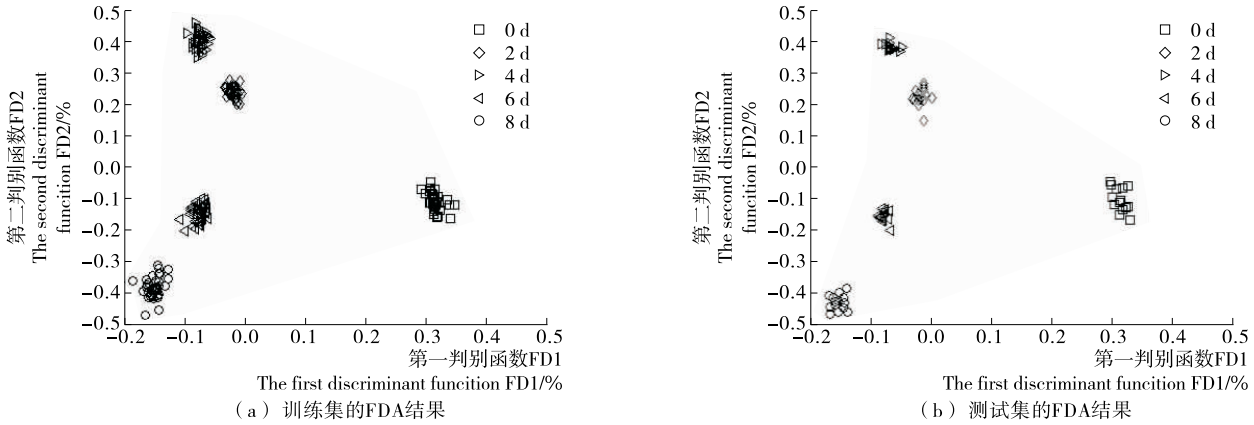


图4 基于特征波长下的FDA结果

Figure 4 FDA results based on the characteristic wavelengths

表 4 特征波长下的 Fisher 判别模型判别准确率

Table 4 The accuracy of Fisher discriminant model under characteristic wavelength

实际培育 天数/d	判别培育天数/d											
	训练集						测试集					
	0	2	4	6	8	准确率/%	0	2	4	6	8	准确率/%
0	35	0	0	0	0	100.00	15	0	0	0	0	100.00
2	0	35	0	0	0	100.00	0	15	0	0	0	100.00
4	0	0	35	0	0	100.00	0	0	15	0	0	100.00
6	0	0	0	35	0	100.00	0	0	0	14	1	93.33
8	0	0	0	0	35	100.00	0	0	0	0	15	100.00
总体准确率/%	100.00						98.67					

及 Fisher 判别的高光谱鉴别玉米霉变程度是可行、有效的;而且进行特征波长的提取也是必要的,可减少计算量、提高检测精度及减少信息的冗余现象。该研究也为高光谱鉴别其他物品提供重要的理论依据。

参考文献

[1] 殷勇,郝银凤,于慧春.基于多特征融合的电子鼻鉴别玉米霉变程度[J].农业工程学报,2016,32(12):254-260.

[2] 高秀芬,计融,李燕俊,等.高效液相色谱法测定玉米中的黄曲霉毒素[J].中国食品卫生杂志,2007,19(2):105-108.

[3] 黄海泉.实时荧光定量 PCR 技术在植物检疫中应用的研究进展[J].湖北农业学,2012,51(1):5-8.

[4] 梁颖,张春晖,刘邻渭.液相色谱-质谱测定小麦中 B 类单端孢霉烯族毒素[J].中华预防医学杂志,2006,40(4):293-295.

[5] JONES J L. Application of DNA probes in the food industry[J]. Trends Food Science Technology, 1991, 2(6): 28-32.

[6] 王俊双,许定花,孙秀兰.食品中脱氧雪腐镰刀菌烯醇三种检测方法的比较研究[J].食品工业科技,2008(2):287-290.

[7] 邓舜洲,何庆华,章英,等.竞争间接 ELISA 检测饲料中脱氧雪腐镰刀菌烯醇和玉米赤霉烯酮[J].江西农业大学学报,2006,28(2):289-292.

[8] 姜庆虎,童芳,余明珠,等.高光谱技术—生态学领域研究的新方法[J].植物科学学报,2015,33(5):633-640.

[9] YIN Yong, XIAO Yu-juan, YU Hui-Chun. An image selection method for tobacco leave grading based on image information[J]. Engineering in Agriculture, Environment and Food, 2015, 8: 148-154.

[10] 于慧春,王润博,殷勇,等.基于不同波段的枸杞多糖及总糖高光谱成像检测[J].食品科学,2017,38(8):191-197.

[11] 马菁,张学俭.枸杞 4 大病虫害的遥感近地高光谱特征[J].贵州农业科学,2016,44(3):65-68.

[12] NAKARIYAKUL S, CASASSENT D P. Classification of internally damaged almond nuts using hyperspectral imagery[J]. Journal of Food Engineering, 2011, 103(1): 62-67.

[13] LORENTE D, ALEIXOS N, GOMEZ-SANCHIS J, et al. Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks[J]. Food Bioprocess Technol, 2013, 6(2): 530-541.

[14] 褚璇,王伟,赵昕,等.近红外光谱和特征光谱的山茶油掺假鉴别方法研究[J].光谱学与光谱分析,2017,37(1):75-79.

[15] KAMRUZZAMAN M, MAKINO Y, OSHITA S. Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging[J]. Food Chemistry, 2016, 196: 1 084-1 091.

[16] 田卫新,何丹丹,杨东,等.一种基于高光谱图像的熟牛肉 TVB-N 含量预测方法[J].食品与机械,2016,32(12):70-74.

[17] 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.

[18] 张楠楠,刘伟,王伟,等.玉米霉变及黄曲霉毒素的图像处理检测方法[J].中国粮油学报,2014,29(2):82-88.

[19] 张若宇.番茄可溶性固形物和硬度的高光谱成像检测[D].杭州:浙江大学,2014:51-52.

[20] 季江,高鹏飞,贾南南,等.自适应多尺度窗口平均光谱平滑[J].光谱学与光谱分析,2015,35(5):1 445-1 449.

[21] 谢安国.冷冻冷藏过程中猪肉的光谱特性研究及其品质的快速检测[D].广州:华南理工大学,2016:77-78.

(上接第 67 页)

[15] 荣瑞芬,李鸿玉,叶磊,等.杏鲍菇营养及功能成分分析测定[J].亚太传统医药,2007(5):85-87.

[16] BARROS L, BAPTISTA P, CORREIA D M, et al. Fatty acid and sugar compositions, and nutritional value of five wild edible mushrooms from Northeast Portugal[J]. Food Chemistry, 2007, 105(1): 140-145.

[17] 李书倩,辛广,张博,等.红蘑、猴头菇、香菇三种食用菌中脂肪酸的气相色谱-质谱分析[J].食品工业科技,2012,33(8):56-58.

[18] GOYAL R, GREWAL R B, GOYAL R K. Fatty acid composition and dietary fibre constituents of mushrooms of North India[J].

Emirates Journal of Food and Agriculture, 2015, 27(12): 927-930.

[19] ERGÖNÜL P G, AKATA I, KALYONCU F, et al. Fatty acid compositions of six wild edible mushroom species[J]. The Scientific World Journal, 2013, DOI:10.1155/2013/163964.

[20] RIBEIRO B, DE PINHO P G, ANDRADE P B, et al. Fatty acid composition of wild edible mushrooms species: a comparative study[J]. Microchemical Journal, 2009, 93(1): 29-35.

[21] 许峰. 大球盖菇富硒液体培养条件优化及抗氧化能力初步研究[D].泰安:山东农业大学,2006:5-8.

[22] 闫培生,李桂舫,蒋家慧,等.大球盖菇菌丝生长的营养需求及环境条件[J].食用菌学报,2001,8(1):5-9.