

## 基于高光谱成像技术的不同产地小米判别分析

吉海彦<sup>1,2</sup>, 任占奇<sup>1,2</sup>, 饶震红<sup>3</sup>

1. 中国农业大学, 现代精细农业系统集成研究教育部重点实验室, 北京 100083
2. 中国农业大学, 农业部农业信息获取技术重点实验室, 北京 100083
3. 中国农业大学理学院, 北京 100083

**摘要** 高光谱成像技术被广泛应用于农产品的检测。基于高光谱成像技术结合机器学习算法无损鉴别不同地区的小米样本。将来源7个省份共计23份样品的小米样本根据地理区域划分为东北地区、河北、陕西、山东和山西共5大类, 其中东北地区共6份样品, 山西地区5份样品, 河北、陕西和山东各4份样品。将每份样品均分为10等份并利用高光谱成像仪采集900~1700 nm波段内小米的高光谱数据。为了减少光照不均匀和暗电流对实验的影响, 对采集到的高光谱数据进行黑白校正。利用ENVI软件选取小米高光谱图像的感兴趣区域(ROI), 每份小米样品选取9个ROI。计算ROI内的平均光谱值, 以此平均值作为该样本的一条光谱记录, 最后共收集到2070条光谱曲线, 其中东北类540条, 山西类450条, 其他河北类、山东类、陕西类各360条。为了减少样品表面的不平整性引起的散射现象, 进而影响小米的真实光谱信息, 对收集到的原始光谱进行多元散射校正预处理(MSC)。采用随机划分法对校正后的光谱数据划分训练集和测试集, 测试集占的比例为0.3。利用线性判别分析(LDA)对不同产地小米的光谱数据进行可视化分析, 将测试集代入训练好的LDA模型, 做出预测结果的混淆矩阵(Confusion Matrix), 结果表明LDA对于陕西和山西类的预测准确率为0.84和0.99, 对于东北、河北和山东的预测准确率仅为0.68, 0.68和0.40。进而采用递归特征消除(RFE)对小米的光谱信息进行特征选择, 去除冗余的信息, 提高模型的预测准确率。将RFE分别与支撑向量机(SVM)和逻辑回归(LR)结合, 对不同产地小米的判别进行对比分析。将小米光谱数据的训练集分别代入SVM-RFE和LR-RFE模型并结合3折交叉验证技术, 以模型F值的微平均(Micro-averaging)最优选择出相应的特征子集。结果表明, LR-RFE选择的波长数为74个, 其模型的Micro\_F为0.59; SVM-RFE选择的波长数为220, 其模型的Micro\_F为0.66。将选择后的特征子集应用到测试集并将测试集分别代入SVM和LR模型, 采用模型预测结果的混淆矩阵和模型的受试者工作特征曲线(ROC)作为评价方法。结果表明SVM-RFE对东北地区、河北、陕西、山东和山西的预测准确率分别为1, 0.37, 0.72, 0和1, 其ROC曲线下面积(AUC)分别为0.82, 0.92, 0.93, 0.70和0.99。LR-RFE的预测准确率分别为0.92, 0, 0.97, 0和0.80, 其AUC分别为0.72, 0.74, 0.94, 0.66和0.88。从预测结果可以看出SVM-RFE模型的综合分类性能优于LR-RFE, 而对陕西类的判别LR-RFE要优于SVM-RFE, 对于河北类和山东类两个模型都不能有效判别。这两个模型的预测准确率相比LDA有了一定的提升。

**关键词** 高光谱成像; 小米; 判别分析; 递归特征消除

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)07-2271-07

### 引言

小米又称粟米, 含有丰富的营养成分, 具有保健和药用特性, 深受人们的喜爱, 是我国重要的粮食作物之一<sup>[1]</sup>。由于温度、水分等气候条件的不同, 不同地区的小米具有不同

的品质特性, 如小米的外观形态、所含营养物质等均具有一定的差异<sup>[2-5]</sup>, 因此不同地区的小米也具有价格差异。一些不法商家为了经济利益, 将不同种类的小米以次充好, 严重损害了消费者的利益。因此对小米的品质、加工及产地等方面的检测提出了更高的要求<sup>[6]</sup>。普通根据小米形态上不同进行鉴别的方法具有主观性强等缺点, 化学法鉴定则耗时耗

收稿日期: 2018-05-17, 修订日期: 2018-09-22

基金项目: 国家“十三五”重点研发计划课题(2016YFD0200602)资助

作者简介: 吉海彦, 1964年生, 中国农业大学信息与电气工程学院教授 e-mail: yuntian@cau.edu.cn

力、对样品具有破坏性、需要化学试剂等不足。高光谱图像技术具有波段多、光谱分辨率高、图谱合一等优点,检测过程无需对样品预处理、无污染,被广泛应用到现代农业检测分析<sup>[7-10]</sup>。

王庆国等基于高光谱图像技术结合偏最小二乘判别分析对不同产地和年份的玉米种子进行鉴别<sup>[11]</sup>;楚秉泉等基于近红外高光谱技术和特征波谱分析方法对不同品种的竹叶进行判别,最后采用受试者工作特征曲线(receiver operating characteristic curve, ROC)对模型进行验证,结果表明识别率能到达 0.98<sup>[12]</sup>;王婉娇等采用近红外高光谱成像技术结合线性判别分析对 3 类不同产地的圆枣进行判别分析,模型准确率达到 0.99<sup>[13]</sup>;Barbin 等基于高光谱图像技术结合主成分分析技术对猪肉的等级进行判别<sup>[14]</sup>;Liu 等利用高光谱图像系统采集了 122 份包含 3 类品种的荔枝样品,研究出了主成分分析(principal component analysis, PCA)结合支持向量机(support vector machine, SVM)为鉴别不同荔枝品种的最佳模型<sup>[15]</sup>;Rivera 等基于高光谱图像技术、机器学习算法和多种特征选择算法研究了对芒果早期机械损伤的鉴别<sup>[16]</sup>。国内外的研究表明高光谱图像技术被广泛应用于农产品的检测,但将高光谱图像技术应用在不同地区的小米判别分析中比较少见。

本研究以来源不同地区的小米样品为研究对象,利用高光谱成像技术、光谱预处理技术、递归特征消除技术和分类机器学习算法对不同产地的小米进行判别分析,将不同模型预测结果的混淆矩阵和 ROC 曲线作为评价方法,对比了不同模型的分类结果,为实现在线快速无损识别不同产地的小米提供参考。

## 1 实验部分

### 1.1 材料

实验用的小米来源 7 个省份共计 23 份样品,如表 1 所示。将每份样品均分为 10 等份,每等份如图 3 所示。根据产地不同,将样品按产地分为 5 类,其中辽宁、黑龙江和吉林省归为东北类,其他山西、陕西、河北、山东各归为一类。

表 1 小米样品的产地  
Table 1 The origin of millet samples

辽宁省	黑龙江省	吉林省	河北省	山东省	陕西省	山西省
朝阳县	哈尔滨	乾安县	南和县	临沂市	定边县	忻州市
	肇源县	昌邑县	邯郸市	济南市	府谷县	泽州县
		四平市	沧州市	安丘县	延川县	垣曲县
			张家口	金乡县	榆林市	兴县
						陵川县

### 1.2 仪器

实验用的仪器为高光谱分析仪,为北京卓立汉光公司产品,主要由高光谱成像仪、电控移动平台、均匀光源、计算机及控制软件等部分组成,如图 1 所示。光谱相机为卓立汉光公司的 Image 谱像系列,该相机能获取的波长范围为 900

~1 700 nm,共 256 个波长点。均匀光源由四个 200 W 的溴钨灯构成。该仪器的工作原理是通过光源照射放置于电控移动平台上的待测样品,样品的反射光被光谱相机捕获,随着电控移动平台的移动,最终获得一个同时包含图像信息和光谱信息的三维数据立方块,电控移动平台的速度由软件控制。

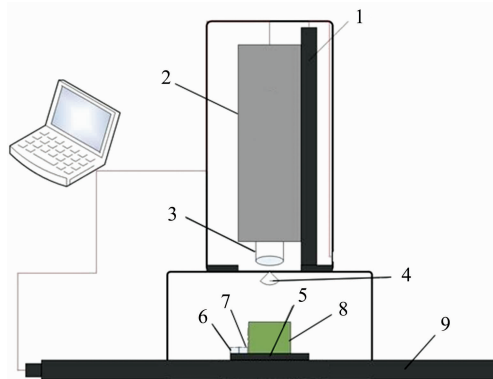


图 1 高光谱成像系统示意图

1: 工作距离调整台; 2: 高光谱成像仪; 3: 成像镜头; 4: 光源; 5: 样品台; 6: 白板; 7: 调焦板; 8: 样品; 9: 一维电移台

### Fig. 1 Schematic diagram of hyperspectral imaging system

1: Working distance adjustment plate; 2: Hyperspectral imager; 3: Imaging lens; 4: Light source; 5: Sample plate; 6: White board; 7: Focusing plate; 8: Sample; 9: One-dimensional electric station

### 1.3 高光谱数据的采集与校正

在采集高光谱数据前先将机器预热 30 min,为了得到清晰的图像需要多次的相机调焦和移动平台的速度测试,最终确定平台的移动速度为  $0.35 \text{ cm} \cdot \text{s}^{-1}$ ,相机曝光时间为 0.09 s。为了减少光照不均匀和暗电流对实验的影响,需要对采集到的高光谱数据进行黑白校正<sup>[17]</sup>,校正公式为

$$R = \frac{I - D}{W - D} \quad (1)$$

式(1)中,  $I$  为采集到的原始高光谱数据;  $D$  为盖上相机镜头采集到的数据(反射率接近 0);  $W$  为对准白板采集到的数据(反射率接近 1);  $R$  为校正后的高光谱数据。校正的工具为系统自带的 SpecVIEW 软件,后续用到的软件包括 Matlab 2016b, Envi5.1 和 python3.6。

### 1.4 递归特征消除

特征选择的目的是使得原始高光谱数据降维,且降维后的数据能够提高分类性能或者保持性能不变。特征选择可以去冗余的特征,减少模型运行的时间,其选取的方式分为筛选器、封装器和混合器三种<sup>[18]</sup>。递归特征消除(recursive feature elimination, RFE)是封装器的代表,它是一个框架需要结合特定的分类器使用,其流程图如图 2 所示。

## 2 结果与讨论

### 2.1 样品的原始光谱曲线和多元散射校正预处理

利用 Envi5.1 软件选取小米高光谱图像感兴趣区域

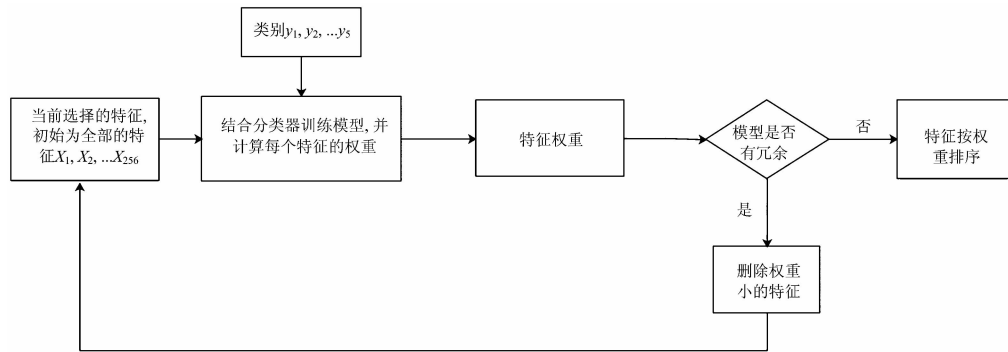


图 2 递归特征消除流程图  
Fig. 2 Recursive feature elimination flow chart

(region of interest, ROI), 为了充分利用小米样本同时增加模型训练样本数, 每个高光谱图像选取 9 个 ROI, ROI 的选取如图 3 所示, 每个 ROI 选取的像素点数为 50 个。计算 ROI 内的平均光谱值, 以此平均值作为该样本的一条光谱记录, 最后共收集 2 070 条光谱曲线, 其中东北类 540 条, 山西类 450 条, 河北类、山东类、陕西类各 360 条, 对每类样品的光谱曲线求平均, 平均光谱曲线如图 4 所示。在小米的高光谱数据采集过程中, 所测样品表面的不平整性可能会引起散射现象发生, 进而影响小米的真实光谱信息。多元散射校正算法(multiplicative scatter correction, MSC)可以校正漫反射光谱的基线漂移, 经过散射校正后能有效抑制样品不均匀性

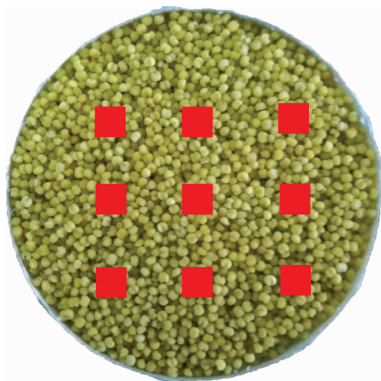


图 3 小米样本 ROI 选取示意图

Fig. 3 Schematic diagram of millet sample ROI selection

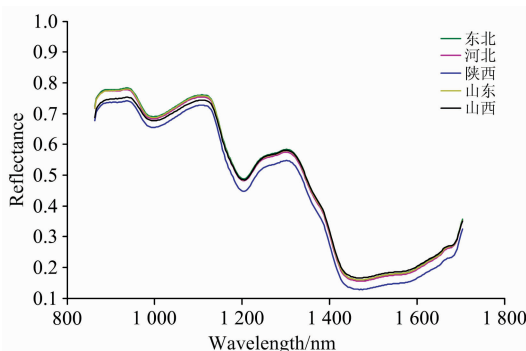


图 4 不同产地小米 ROI 平均光谱曲线

Fig. 4 Millet ROI average spectral curve from different origins

造成的噪声, 校正前后的光谱曲线如图 5 所示。

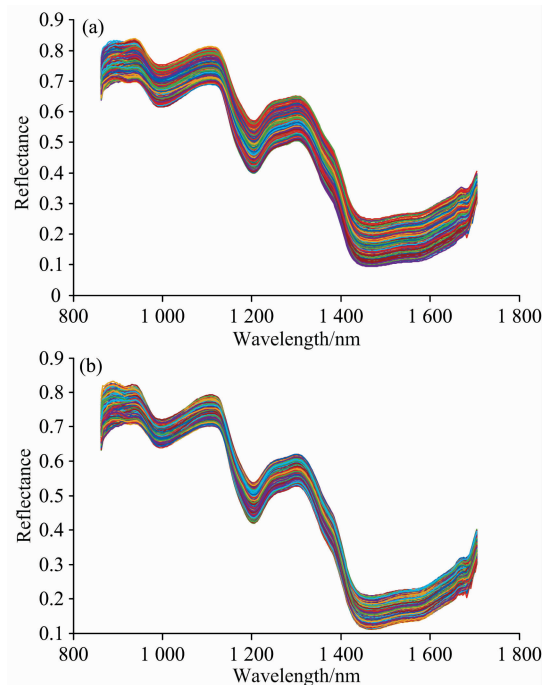


图 5 小米原始光谱曲线 (a) 和 MSC 校正后的光谱曲线 (b)

Fig. 5 Original spectral curve of millet (a) and MSC corrected spectral curve (b)

## 2.2 基于线性判别分析的样品识别

线性判别分析(linear discriminant analysis, LDA)是一种分类和降维算法, 它能将高维数据投影到最佳分类向量空间, 在新的向量空间中保证有较大的类间距离和较小的类内距离。利用 LDA 对不同产地小米样品的光谱数据进行可视化判别。首先将不同产地小米样品的光谱数据采用随机划分法划分为训练集和测试集, 测试集占比 0.3。将训练集代入 LDA 模型训练, 用训练好的模型进行预测。如图 6 画出了 LDA 在预测数据集上 2 维投影后的分类情况, 为了更清晰看出分类后的结果, 画出预测结果的混淆矩阵如图 7 所示。混淆矩阵是数据分析、数据科学和机器学习中对分类模型预测结果的一种评价方式, 它以矩阵的形式记录了样本的真实

类别和模型的预测类别之间的对应关系, 归一化后的混淆矩阵中的元素  $M(i, j)$  表示真实类别为  $i$  的样本被预测为类  $j$  的百分比。

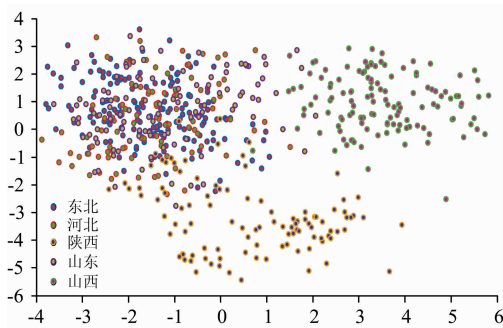


图 6 不同产地小米高光谱数据的线性判别分析

Fig. 6 Hyperspectral data LDA analysis of millet from different origins

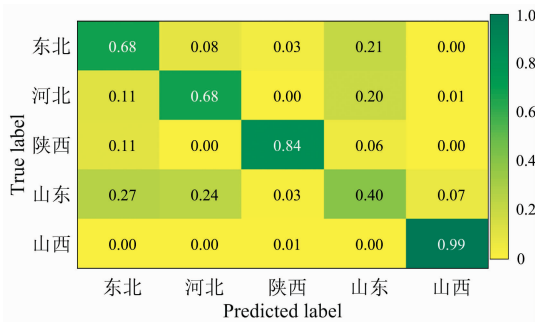


图 7 LDA 预测结果的混淆矩阵

Fig. 7 LDA prediction result confusion matrix

根据图 7 可以分析出 LDA 对山西和陕西类的预测准确率较高, 而其他三类预测准确率不理想, 其中山东类的准确率只有 0.40。LDA 虽然能较好地数据降低到二维, 但它却不能把一些冗余的特征消除, 而且在降维的过程中可能会损失一些有用的信息, 所以有必要对数据进行特征选择, 去除冗余的特征后进一步进行分析。

### 2.3 基于支持向量和逻辑回归的递归特征消除

张睿等<sup>[19]</sup>提出了一种基于一对一(one-vs-one, OVO)策略支持多分类的支持向量机递归特征消除算法(SVM-RFE), 并对比了一对多(one-vs-all, OVA)策略的 SVM-RFE 和其他几种特征选择算法, 结果表明 OVO SVM-RFE 是一种可靠有效的高光谱特征选择算法。本文采用 OVO SVM-RFE 作为其中的一个特征选择算法。逻辑回归(logistic regression, LR)是一个经典的二分类模型, 采用 OVO 策略将其扩展为多分类模型, 也将其和 RFE 结合, 与 SVM-RFE 进行对比研究。首先将训练集代入这两个模型进行训练并结合 3 折交叉验证技术选择出最优的特征子集。由于本研究为多类别分类问题, 这里采用  $F$  值的微平均作为模型的评价指标<sup>[20]</sup>。如图 8 所示, 图中横坐标代表选择的特征子集的个数, 纵坐标为模型预测结果  $F$  值的微平均, 从图 8(a)可以看出 SVM-RFE 选择的波长数为 220 个, 其模型的 Micro\_F 为 0.66; LR-

RFE 选择的波长数为 74[见图 8(b)], 其模型的 Micro\_F 为 0.59。LR-RFE 选择的波长数较少但其模型的整体分类效果不如 SVM-RFE。将选择好的特征子集应用到预测集上, 代入 SVM 和 LR 模型中, 做出两个模型预测结果的混淆矩阵如图 9(a)和(b)所示。

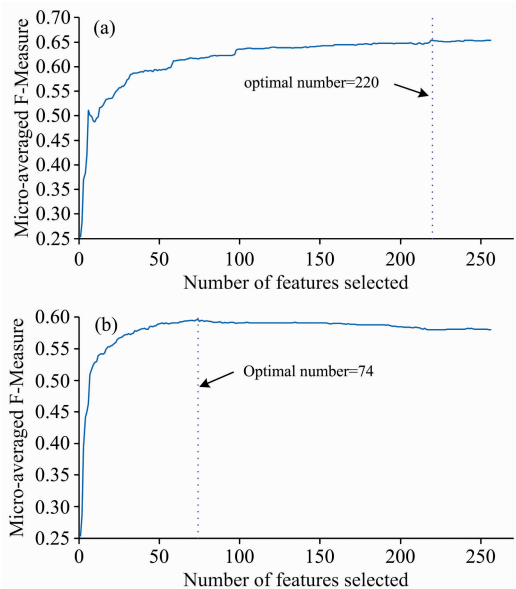


图 8 SVM-RFE 特征选择 (a) 和 LR-RFE 特征选择 (b)

Fig. 8 SVM-RFE feature selection (a) and LR-RFE feature selection (b)

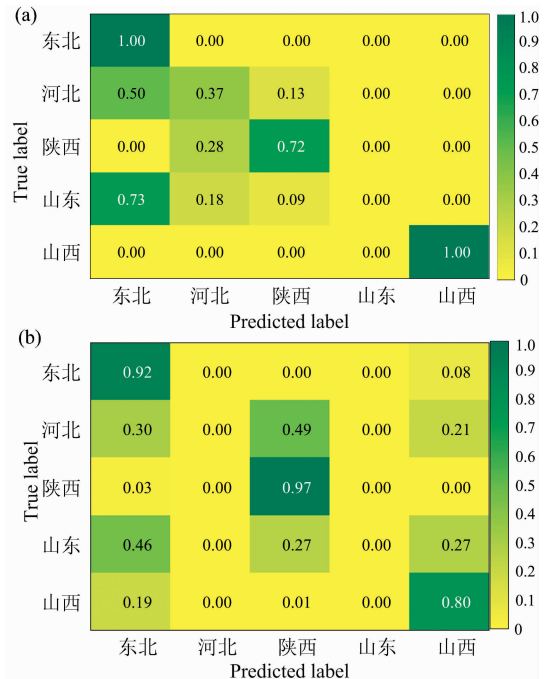


图 9 SVM 预测结果混淆矩阵 (a) 和 LR 预测结果混淆矩阵 (b)

Fig. 9 SVM prediction result confusion matrix (a) and LR prediction result confusion matrix (b)

从图 9(a)和(b)可以看出, SVM 对东北类和山西类的预测准确率为 1, 但对陕西类的预测准确率没有 LR 高, 而河北类和山东类两个模型都不能准确预测出。为了进一步验证这两个模型的性能, 分别做出两个模型的受试者工作特征曲线(ROC)如图 10(a)和(b)所示, 图中 class 0—class 4 分别代表东北、河北、陕西、山东和山西类。ROC 是反应模型敏感性和特异性连续变量的综合指标, 对验证模型的性能有很大帮助。一般采用 ROC 曲线下的面积(area under ROC curve, AUC)作为模型评价指标, 其值最大为 1, 值越大代表其模型的探测效果越好。

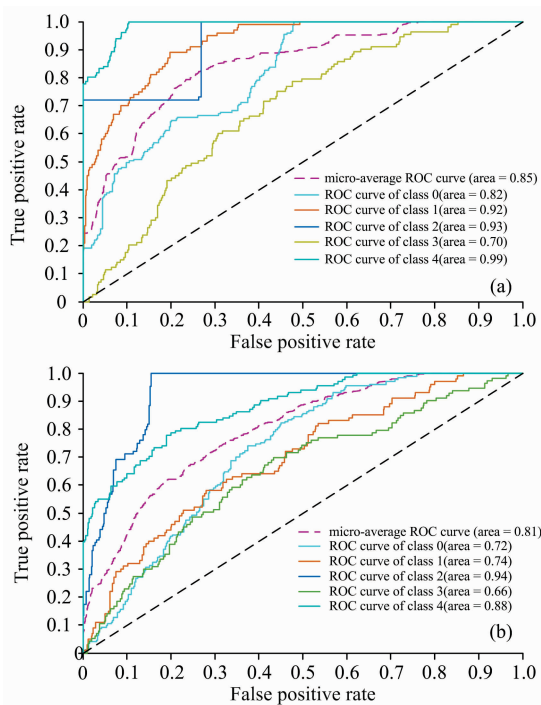


图 10 SVM ROC 曲线(a)和 LR ROC 曲线(b)

Fig. 10 SVM ROC curves (a) and LR ROC curves (b)

从图 10(a)和(b)可以看出, SVM 模型的微平均 ROC (Micro-average ROC) 曲线的 AUC 比 LR 模型高, 故可认为 SVM 模型对判别这五类小米的性能优于 LR。从 SVM 模型的 ROC 曲线可以看出, 该模型对于识别河北、陕西和山西类具有较高的性能, AUC 均在 0.9 以上; 对于识别山东和东北类该模型的性能也不低, AUC 在 0.7~0.9 之间。从 LR-RFE 模型的 ROC 曲线可以看出, 该模型对于识别陕西类具有较高的性能, AUC 为 0.94; 对于识别东北、河北和山西类的性能也不低, AUC 在 0.7~0.9 之间; 对于识别山东类的性能稍低, AUC 为 0.66。结合模型的混淆矩阵和 ROC 曲线可以分析出, 在这五类小米的判别中, SVM-RFE 模型可准确判别出山西类和东北类, 陕西类可利用 LR-RFE 模型较好地判别出, 但河北类和山东类两个模型都不能有效地判别。

### 3 结 论

(1) 利用高光谱图像系统采集五个产地小米的光谱数据, 对原始光谱进行 MSC 校正。利用 LDA 对不同产地的小米进行可视化判别分析并采用混淆矩阵对预测结果进行可视化, 结果表明 LDA 对陕西和山西类的预测准确率为 0.84 和 0.99, 而对其他三类的预测结果不甚理想。

(2) 为了去除冗余的光谱信息, 提高分类预测的准确率, 采用 SVM-RFE 和 LR-RFE 两个模型对比研究, 将模型预测结果的混淆矩阵和模型的 ROC 曲线作为评价方法。结果表明 SVM-RFE 总体分类性能优于 LR-RFE, 而对陕西类的判别 LR-RFE 要优于 SVM-RFE。

(3) 对于来源 7 个省份 5 个不同产地小米样品的鉴别分析结果表明, 可利用 SVM-RFE 准确鉴别出东北类和山西类, 利用 LR-SVM 准确鉴别陕西类, 河北和山东类两个模型都不能有效地鉴别。相比于 LDA, 这两个模型的预测准确率有了一定的提升。

### References

- [1] QIAO Ling, WANG Xin(乔玲, 王欣). Agricultural Science & Technology and Equipment(农业科技与装备), 2015, (11): 41.
- [2] ZHAO Yu, CUI Ji-han, LI Shun-guo, et al(赵宇, 崔纪菡, 李顺国, 等). Journal of Hebei Agricultural Sciences(河北农业科学), 2017, (4): 1.
- [3] CUI Ji-han, ZHAO Yu, LIU Meng, et al(崔纪菡, 赵宇, 刘猛, 等). Journal of Agricultural Science and Technology(中国农业科技导报), 2017, 19(8): 84.
- [4] LI Xing, WANG Hai-huan, SHEN Qun(李星, 王海寰, 沈群). Journal of Chinese Institute of Food Science and Technology(中国食品学报), 2017, 17(7): 248.
- [5] ZHANG Ren-tang, DONG Hao, GAO Lin, et al(张仁堂, 董浩, 高琳, 等). Food and Nutrition in China(中国食物与营养), 2012, 18(10): 22.
- [6] SONG Xue-jian, QIAN Li-li, ZHOU Yi, et al(宋雪健, 钱丽丽, 周义, 等). Food Research and Development(食品研究与开发), 2017, 38(11): 134.
- [7] Wu D, Sun D W. Innovative Food Science & Emerging Technologies, 2013, 19(1): 15.
- [8] Liu Y, Pu H, Sun D W. Trends in Food Science & Technology, 2017, 69.
- [9] Feng Y Z, Sun D W. Critical Reviews in Food Science & Nutrition, 2012, 52(11): 1039.
- [10] Dale L M, Thewis A, Boudry C, et al. Applied Spectroscopy Reviews, 2013, 48(2): 142.
- [11] WANG Qing-guo, HUANG Min, ZHU Qi-bing, et al(王庆国, 黄敏, 朱启兵, 等). Journal of Food Science and Biotechnology(食品与

- 生物技术学报), 2014, 33(2): 163.
- [12] CHU Bing-quan, ZHAO Yan-ru, HE Yong(楚秉泉, 赵艳茹, 何 勇). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(6): 1718.
- [13] WANG Wan-jiao, HE Xiao-guang, YANG Xiao-chen, et al(王婉娇, 贺晓光, 杨晓忱, 等). Food Science and Technology(食品科技), 2015, (6): 344.
- [14] Barbin D, Elmasry G, Sun D W, et al. Meat Science, 2012, 90(1): 259.
- [15] Liu D, Wang L, Sun D W, et al. Food Analytical Methods, 2014, 7(9): 1848.
- [16] Rivera N V, Gómez-Sanchis J, Chanona-Pérez J, et al. Biosystems Engineering, 2014, 122(3): 91.
- [17] Wu D, Sun D W, He Y. Innovative Food Science & Emerging Technologies, 2012, 16(39): 361.
- [18] YAO Xu, WANG Xiao-dan, ZHANG Yu-xi, et al(姚 旭, 王晓丹, 张玉玺, 等). Control and Decision(控制与决策), 2012, 27(2): 161.
- [19] ZHANG Rui, MA Jian-wen(张 睿, 马建文). Geomatics and Information Science of Wuhan University(武汉大学学报·信息科学版), 2009, 34(7): 834.
- [20] Santos A M, Canuto A M P, Neto A F. International Journal of Computer Information Systems and Industrial Management Applications, 2011, 3(1): 218.

## Discriminant Analysis of Millet from Different Origins Based on Hyperspectral Imaging Technology

JI Hai-yan<sup>1, 2</sup>, REN Zhan-qi<sup>1, 2</sup>, RAO Zhen-hong<sup>3</sup>

1. Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China
2. Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, China Agricultural University, Beijing 100083, China
3. College of Science, China Agricultural University, Beijing 100083, China

**Abstract** Hyperspectral imaging technology has been widely used in the detection of agricultural products. This paper studies the non-destructive identification of millet samples from different regions based on hyperspectral imaging and machine learning algorithms. The millet samples from seven provinces were divided into five categories according to geographical regions. They were Dongbei, Hebei, Shaanxi, Shandong, and Shanxi, respectively. A total of 23 samples were collected in these areas, including 6 samples in Dongbei, 5 samples in Shanxi, and respective 4 samples in Hebei, Shaanxi, and Shandong. Each sample was equally divided into 10 equal parts and the hyperspectral data of millet in the wavelength band from 900 to 1 700 nm was collected using a hyperspectral imager. In order to reduce the influence of uneven illumination and dark current on the experiment, the collected hyperspectral data was corrected in black and white. The ENVI software was used to select the region of interest (ROI) of millet hyperspectral image, and 9 ROIs were selected for each sample of millet. The average spectral value in the ROI was calculated, which was used as a spectrum record of the sample. Finally, a total of 2 070 spectral curves were collected, of which 540 from Dongbei, 450 from Shanxi, and several 360 from Hebei, Shandong, and Shaanxi respectively. In order to reduce the scattering phenomenon caused by the unevenness of the sample surface, which would affect the true spectral information of millet, the multivariate scatter correction (MSC) pretreatment was performed on the original spectrum. In addition, randomized division method was used to divide the corrected spectral data into training set and test set. The ratio of test set was 0.3. Linear Discriminant Analysis (LDA) was used to visualize spectral data of millet from different origins. Substituting the test set into a well-trained LDA model, and finally a confusion matrix of prediction results was created. The results showed that LDA had a prediction accuracy of 0.84 and 0.99 for Shaanxi and Shanxi, and only 0.68, 0.68, and 0.40 for Dongbei, Hebei, and Shandong. Therefore, the recursive feature elimination (RFE) was used to select useful spectral information, remove redundant information, and improve the prediction accuracy. The RFE combined with support vector machine (SVM) and Logistic Regression (LR) were used to compare and analyze the discriminant of millet from different regions. Substituting training set of millet spectral data into SVM-RFE and LR-RFE models, and the corresponding feature subsets were selected optimally by the micro-averaging of the model F-values and 3-fold cross validation technology. The results showed that the number of wavelengths selected by the LR-RFE was 74 and the Micro\_F of the model was 0.59; Meanwhile the number of wavelengths selected by the SVM-RFE was 220 and the Micro\_F of the model was 0.66. The selected feature subset was applied to the test set. Substituting the

test set into SVM and LR models respectively, and confusion matrix of model prediction results and the receiver operating characteristic curve (ROC) of the model were used as the evaluation method. The results showed that the accuracy of SVM-RFE prediction was 1, 0.37, 0.72, 0, and 1 for Dongbei, Hebei, Shaanxi, Shandong, and Shanxi, and the area under ROC curve (AUC) was 0.82, 0.92, 0.93, 0.70, and 0.99 respectively. The accuracy of LR-RFE prediction was 0.92, 0, 0.97, 0, and 0.80, and the AUC was 0.72, 0.74, 0.94, 0.66, and 0.88 respectively. It can be seen from the prediction results that the overall classification performance of SVM-RFE model was better than that of LR-RFE, while the discrimination of Shaanxi class LR-RFE was better than that of SVM-RFE. For the Hebei and Shandong categories, neither model could effectively discriminate it. Compared with LDA, the prediction accuracy of these two models had been improved.

**Keywords** Hyperspectral imaging technology; Millet; Discriminant analysis; Recursive feature elimination

(Received May 17, 2018; accepted Sep. 22, 2018)