

基于高光谱成像法对大豆种子的品种分类鉴别

王盛楠¹, 谭 勇^{1*}, 刘春宇¹, 宋少忠², 李 政¹

¹长春理工大学物理学院, 吉林 长春

²吉林工程技术师范学院信息工程学院, 吉林 长春

收稿日期: 2022年3月10日; 录用日期: 2022年4月4日; 发布日期: 2022年4月13日

摘 要

本文针对东北大豆品种分类鉴别的需求, 利用高光谱成像技术获取了6种大豆样品在392.38~1011.01 nm高光谱图像, 提取感兴趣(ROI)区域数据, 获得大豆种子样品的反射光谱曲线。经过卷积平滑(S-G)预处理, 再根据大豆光谱曲线差异选取455.54 nm、479.3 nm、604.04 nm、657.46 nm、705.72 nm、856.89 nm、918.07 nm、953.54 nm作为特征波段, 分别输入极限学习机(ELM)和随机森林(RF)模型, 得到的分类正确率分别为78.22% 和98.89%, 模型预测时间分别为11 s和12 s。研究表明, 经卷积平滑和高光谱波段优化的特征波段, 运用随机森林(RF)模型分析是分类准确率最高、预测时间较快的分类方法, 高光谱成像法可有效对大豆品种做出分类鉴别。

关键词

高光谱成像, 极限学习机, 随机森林

Variety Classification and Identification of Soybean Seeds Based on Hyperspectral Imaging Method

Shengnan Wang¹, Yong Tan^{1*}, Chunyu Liu¹, Shaozhong Song², Zheng Li¹

¹College of Physics, Changchun University of Science and Technology, Changchun Jinlin

²School of Information Engineering, Jilin Engineering Normal University, Changchun Jinlin

Received: Mar. 10th, 2022; accepted: Apr. 4th, 2022; published: Apr. 13th, 2022

*通讯作者。

文章引用: 王盛楠, 谭勇, 刘春宇, 宋少忠, 李政. 基于高光谱成像法对大豆种子的品种分类鉴别[J]. 传感器技术与应用, 2022, 10(2): 177-186. DOI: 10.12677/jsta.2022.102022

Abstract

According to the needs of soybean variety classification and identification in Northeast China, the hyperspectral images of six soybean samples at 392.38~1011.01 nm were obtained by hyperspectral imaging technology, the region of interest (ROI) data were extracted, and the reflection spectrum curves of soybean samples were obtained. After convolution smoothing (S-G) pretreatment, 455.54 nm, 479.3 nm, 604.04 nm, 657.46 nm, 705.72 nm, 856.89 nm, 918.07 nm and 953.54 nm were selected as the characteristic bands according to the difference of soybean spectral curves. The classification accuracy was 78.22% and 98.89% respectively, and the prediction time of the model was 11 s and 12 s respectively. The results show that the random forest (RF) model analysis is the classification method with the highest classification accuracy and faster prediction time. Hyperspectral imaging method can effectively classify and identify soybean varieties.

Keywords

Hyperspectral Imaging, Extreme Learning Machine, Random Forest

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

大豆作为我国主要的农业产品之一，在我国农业上占有重要地位。传统方法大多根据大豆种子的形态、纹理和大小上的差异对其品种分类。但由于多数大豆种子品种间的外观差异不甚明显，容易出现误判，因此传统方法较难对大豆种子的品种做出较为准确的分类鉴别。有必要研究出一种准确、快速的大豆种子品种分类鉴别方法。

高光谱成像法具有无损样品、检测速度快和可批量检测等优点，近年来在食品安全、医学诊断和航天领域应用广泛[1] [2] [3] [4]。2019年，Zhu等[5]人采集了10个品种的大豆的高光谱图像，对图像进行了预处理后，采用不同模型对其进行了分类，结果表明经多元散射校正MSC——竞争性自适应重加权算法CARS——集成学习组合的方式识别效果最高，交叉验证后精度达到99.8%。同年，Wang等[6]人基于高光谱图像技术对10个品种的大豆进行了分类，采用主成分分析(PCA)对数据降维，从PCA选取的特征图像提取纹理参数(熵、能量等)，结合遗传算法优化BP神经网络(GA-BP)的预测模型准确率达到92%。2020年，李亚婷[7]基于高光谱技术采集了不同程度老化的大豆种子，对数据进行降噪后采用堆叠自动编码器(SAE)选取了59个特征谱段，并建立支持向量机(SVM)模型。实现了高光谱图像对大豆种子的活力鉴别，验证集的准确率为98.67%。

以上研究表明，高光谱成像法对大豆种子的品种的分类鉴别具有可行性。上述已发表的研究在大豆品种、活力鉴别等方面取得了一定成果，但针对我国东北地区的大豆研究还不够深入，所应用的分类模型方法也不够全面。本文以东北地区的大豆种子作为实验样品，600粒大豆样品作为研究对象，运用高光谱成像法获取其大豆种子的反射光谱均值曲线，并结合极限学习机(ELM)和随机森林(RF)两种模型对其做出品种分类，丰富了大豆品种分类的模型方法，实现了基于高光谱成像法对大豆种子品种的分类鉴别。

2. 实验部分

2.1. 实验样品

选取 6 个品种的大豆种子为实验样品, 6 个品种分别为东生 9、绥农 42、绥农 26、黑农 51、黑农 61 和合农 75, 所有大豆样品均来自永吉县农业技术推广站, 且 6 个品种大豆均被广泛培育种植, 具备代表性和普遍性, 能够客观地说明实验结果。6 种大豆如表 1 所示。

Table 1. Soybean seeds of 6 different varieties

表 1. 6 个不同品种大豆种子

	合农 75	东生 9	黑农 61	黑农 51	绥农 26	绥农 42
审定编号	国审豆 20170005	国审豆 2013004	国审豆 2014003	吉审豆 2011022	黑审豆 2008013	黑审豆 2016005

每个品种选取 100 粒豆体匀称, 色泽均匀, 完整无损, 无感染病害的大豆种子作为实验样品, 由于大豆种子表面光滑且呈椭球状, 非常容易滚动, 导致高光谱数据采集困难, 而且后续提取数据困难, 为方便大豆种子的高光谱数据提取的准确和便捷, 故采用如图 1 所示的大豆筛作为大豆样品摆放的磨具, 大豆筛共有 100 个孔位, 刚好满足本实验的要求。为避免其他因素对光谱图像产生影响, 实验时将大豆样品放入磨具, 再将模具中大豆平稳地轻放在由黑色摄影布(反射率 $\leq 0.5\%$)包裹的硬卡纸上, 减少杂散光对实验的影响。



Figure 1. Soybean sieve

图 1. 大豆筛

2.2. 实验设备

高光谱以纳米级高分辨率, 光谱采集范围从 392.38 nm 到 1011.01 nm, 共 462 个波段, 且每个波段都有其相应的光谱图像, 具有辨识度高, 表征数据类型多样, 采集到的信息量大优点, 在遥感探测方面的卓越贡献, 更加被广泛运用于其他领域范畴, 拥有更好的发展前景。实验采用 RESONON 台式高光谱成像仪为研究平台, RESONON 台式高光谱成像系统主要由照明系统, 摄像系统, 控制系统构成。图 2 是高光谱的实验装置示意图。

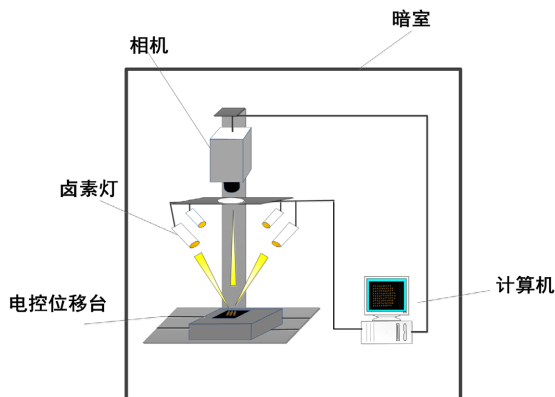


Figure 2. Schematic diagram of hyperspectral experimental device
图 2. 高光谱的实验装置示意图

2.3. 数据采集

考虑到光强分布不均会对图像带来噪声影响,采集高光谱图像数据之前须对实验设备进行黑白板的反射校正。其校正后的图像由以下公式计算获得,公式中原始图像为 I ,扫描白板得到图像为 W ,然后用黑色镜头盖遮住镜头,扫描得到全黑色的图像 B ,最后得到校正后的图像 R 。

$$R = \frac{I - B}{W - B}$$

在采集高光谱图像数据时,为获取样本数据不失真,影像清晰,大小适中。反复调节实验设备参数,总结多次实验经验,经过调整,最后确定的实验参数如下:光谱范围为 392.38~1011.01 nm,镜头到样品间的距离 45 cm,扫描速度为 0.11 cm/s,归位速度为 2.50079 cm/s,点动速度为 1.0 cm/s,帧率为 9.8 Hz,积分时间为 46.701 ms。

实验共采集 6 种大豆种子,每种各 100 粒的高光谱图像。由于所得大豆样品的高光谱图像数据共有 462 个波段,数据量庞大,故只展示各大豆样品在 653.45 nm 波段的高光谱图像,如图 3 所示。

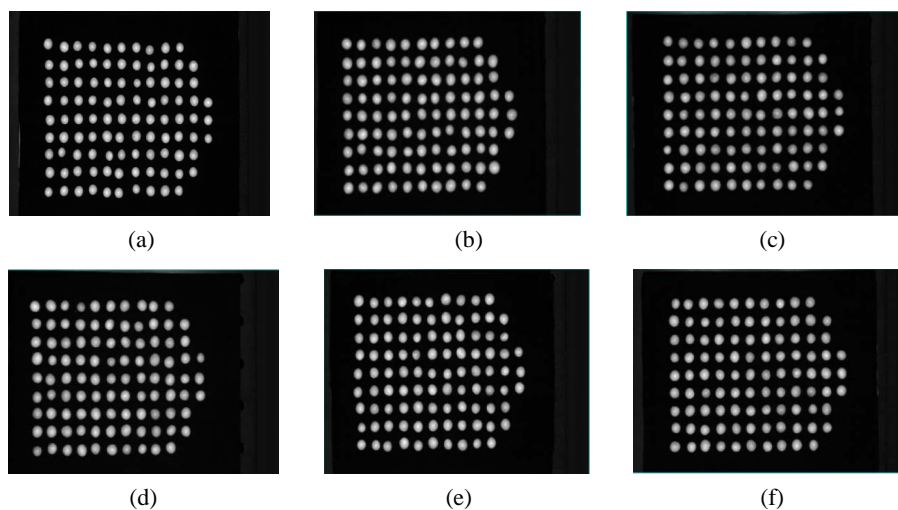


Figure 3. Hyperspectral image of soybean sample in 653.45 nm band. (a) Dongsheng 9; (b) Henong 75; (c) Heinong 51; (d) Heinong 61; (e) Suinong 26; (f) Suinong 42

图 3. 大豆样品在 653.45 nm 波段的高光谱图像。(a) 东生 9; (b) 合农 75; (c) 黑农 51; (d) 黑农 61; (e) 绥农 26; (f) 绥农 42

3. 数据处理

3.1. 感兴趣区域选取

感兴趣区域 ROI 提取就是将研究对象从整个高光谱图像中提取出来, 从而去除背景或其他因素对研究对象光谱数据造成的不利影响[8]。在每粒大豆样品的居中部位选取 30×30 pixels 大小的感兴趣区域, 以该区域全部像素光谱平均值作为该样品的反射光谱均值光谱曲线, 提取所有大豆种子样本的平均反射光谱。图 4 为大豆样品东生 9 中某两粒大豆的感兴趣区域的提取图示, 其余所有大豆样品的感兴趣区域提取皆采用此方法。

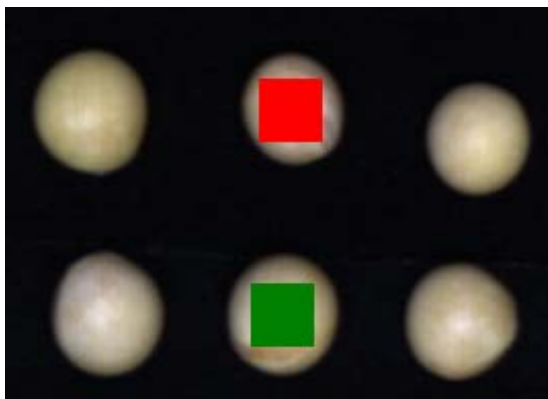
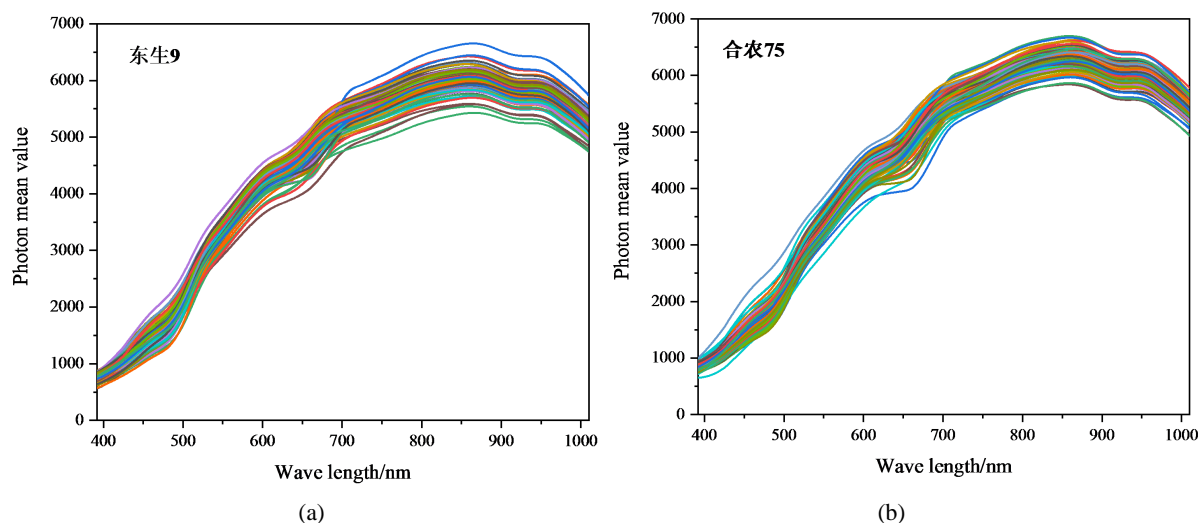


Figure 4. Extraction diagram of region of interest
图 4. 感兴趣区域的提取图示

3.2. 卷积平滑处理(Savitzky-Golay)

SG 平滑即 Savitzky-Golay 滤波平滑, 其采用多项式进行了数据, 基于最小二乘法, 能够保留分析信号中的有用信息, 消除随机噪声[9]。选择卷积平滑的预处理方法, 提高大豆样品的反射光谱的平滑性, 并降低噪音的干扰, 使大豆的有效高光谱数据更加得以凸显。图 5 为大豆样品在 S-G 平滑后的反射光谱均值光谱图。由图可知, 光谱数据经过卷积平滑处理后, 光谱曲线光滑, 各品种大豆的反射光谱曲线对比较为清晰, 为后续数据处理分析奠定了良好的基础。



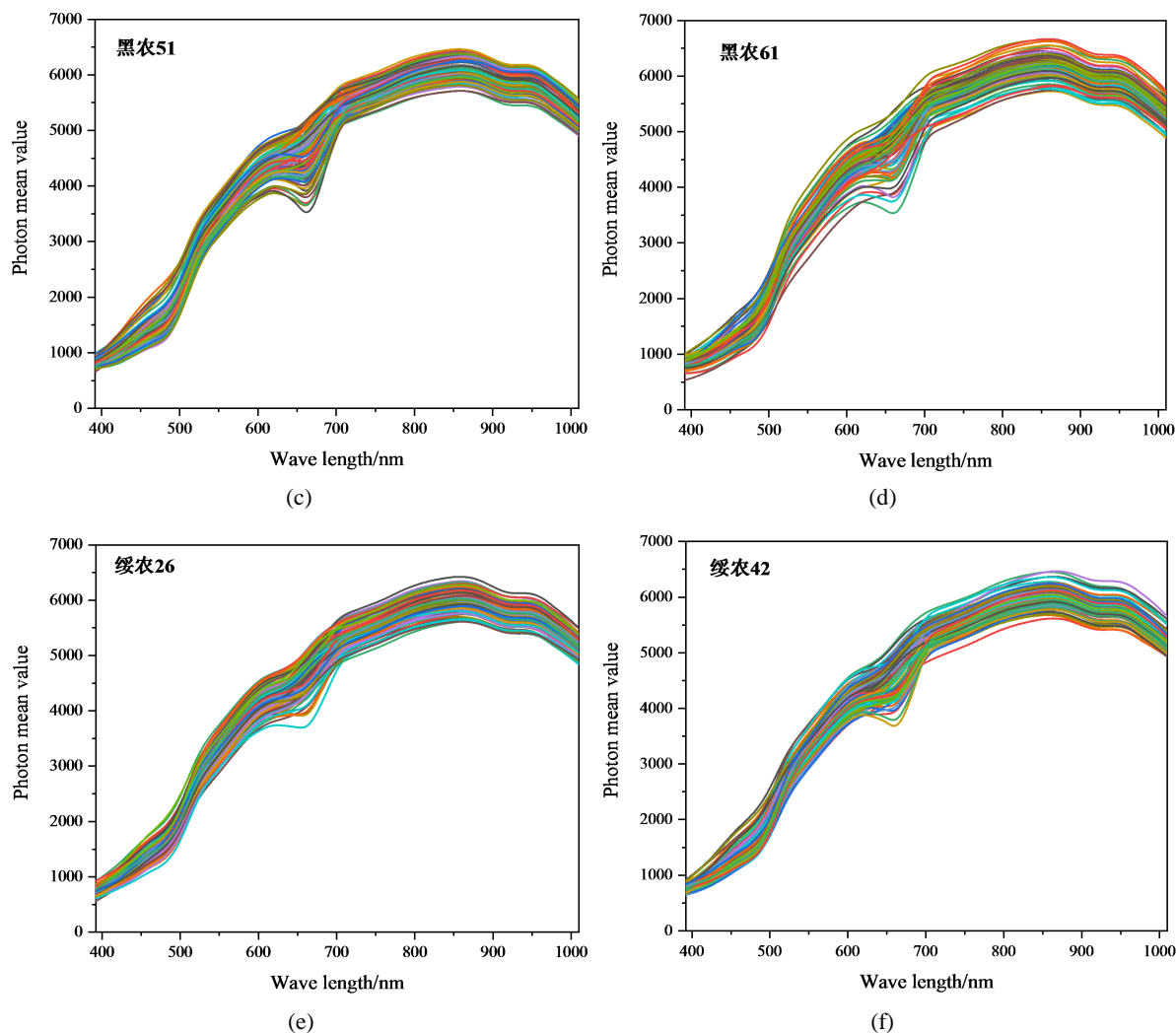


Figure 5. Average reflection spectrum of soybean sample after Savitzky-Golay. (a) Dongsheng 9; (b) Henong 75; (c) Heinnong 51; (d) Heinnong 61; (e) Suinong 26; (f) Suinong 42

图 5. 大豆样品 S-G 平滑后的反射光谱均值图。(a) 东生 9; (b) 合农 75; (c) 黑农 51; (d) 黑农 61; (e) 绥农 26; (f) 绥农 42

4. 高光谱特征波段选择

此次实验使用的成像光谱仪的光谱范围为 392.38~1011.01 nm, 共有 462 个波段。由图 5 的大豆样品 S-G 平滑后的反射光谱均值图可知, 采集到的 462 个波段范围内没有明显的噪声波段。因此, 所采集的 462 个波段上的反射光谱信息均对大豆品种分类鉴别有效。不过由于高光谱信息繁杂, 数据量非常大, 导致存在对高光谱信息建模难度大, 模型准确率较低且建模时间较长等问题。优化高光谱特征波段选择, 选取反射光谱中对分类最有价值的特征波段变得尤为重要。所以本文将每个品种的 100 组大豆样品分别做平均处理, 得到 6 种大豆样品的平均反射光谱曲线的对比图, 如图 6。由图可知, 在 420~520 nm、550~630 nm、640~742 nm、810~980 nm 波段范围处, 6 种大豆样品的平均反射光谱曲线差异最为明显, 由此选择波段中筛选出差异最大的几个波段, 即四个波段范围的波峰和波谷处, 分别为: 455.54 nm、479.3 nm、604.04 nm、657.46 nm、705.72 nm、856.89 nm、918.07 nm、953.54 nm。大大地减少了反射光谱中的重复、冗余的数据, 且保留了反射光谱中最能突出大豆品种差异的波段, 最大程度上还原原光谱数据对于大豆

品种分类的有效信息。从而弱化鉴别模型复杂度，达到提高模型分类正确率并缩短模型预测时间的效果。

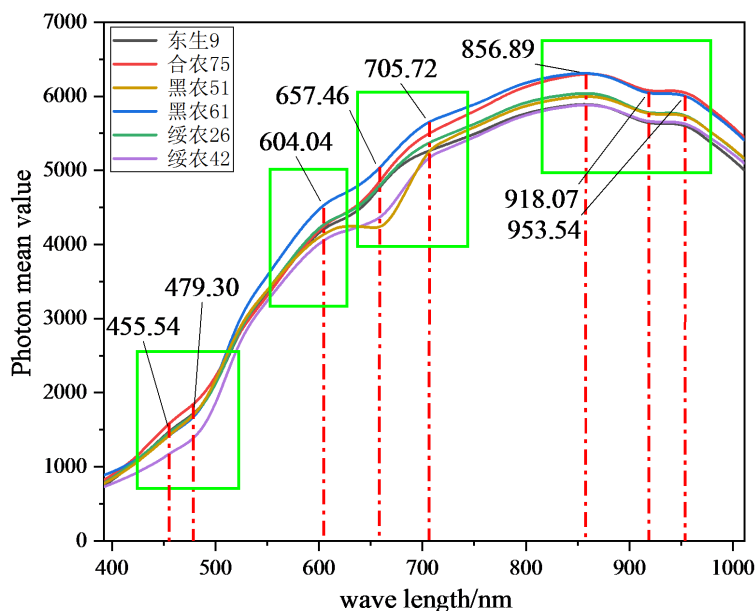


Figure 6. Comparison of average reflectance spectra of six soybean samples
图 6. 6 种大豆样品的平均反射光谱曲线的对比图

5. 大豆光谱数据的分类建模方法

经过感兴趣区域(ROI)和卷积平滑(Savitzky-Golay)处理后得到的全波段光谱曲线，再经过高光谱特征波段优化，选择出 6 个品种大豆在 455.54 nm、479.3 nm、604.04 nm、657.46 nm、705.72 nm、856.89 nm、918.07 nm、953.54 nm 波段处的反射光谱作为高光谱的特征波段输入，再分别建立极限学习机(ELM)和随机森林(RF)的分类模型，所得结果如下。

5.1. 极限学习机(ELM)

输入极限学习机(ELM)模型[10]的训练集大豆种子样品反射光谱为 420 个，测试集大豆种子样品反射光谱为 180 个。图 7 为特征波段反射光谱的 ELM 的大豆品种分类结果图，图中红色记号代表大豆种子的预测类别，蓝色记号代表大豆种子的实际类别，其中红色记号与蓝色记号的重合部分代表大豆品种预测正确，未重合部分代表大豆品种预测错误。红色记号与蓝色记号重合越多，预测正确率越高，反之预测正确率越低。在图 7 中，红色记号与蓝色记号重复较多，但仍有部分红色记号与蓝色记号距离较远，代表分类正确率较低。图 8 为特征波段反射光谱的隐含层神经元个数对 ELM 性能影响图，其中红色记号代表预测集正确率，蓝色记号代表训练集正确率，横坐标为隐含层神经元个数，纵坐标为分类正确率。由图可知，其训练集分类正确率和预测集分类正确率较低。输出结果显示，大豆种子样品的特征波段反射光谱经过极限学习机模型分类，其分类正确率为 78.22%，经过计时，该模型预测时间为 11 s。

5.2. 随机森林(RF)

输入随机森林(RF)模型[11]的训练集大豆种子样品反射光谱为 420 个，测试集大豆种子样品反射光谱为 180 个。图 9 为特征波段反射光谱的随机森林的大豆品种分类结果图，图中红色记号代表大豆种子的

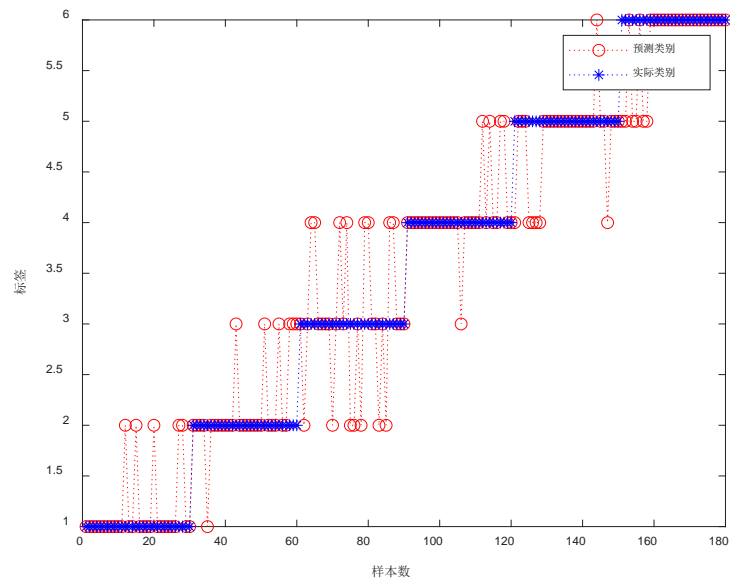


Figure 7. Soybean variety classification results of ELM with characteristic band spectrum

图 7. 特征波段光谱的 ELM 的大豆品种分类结果图

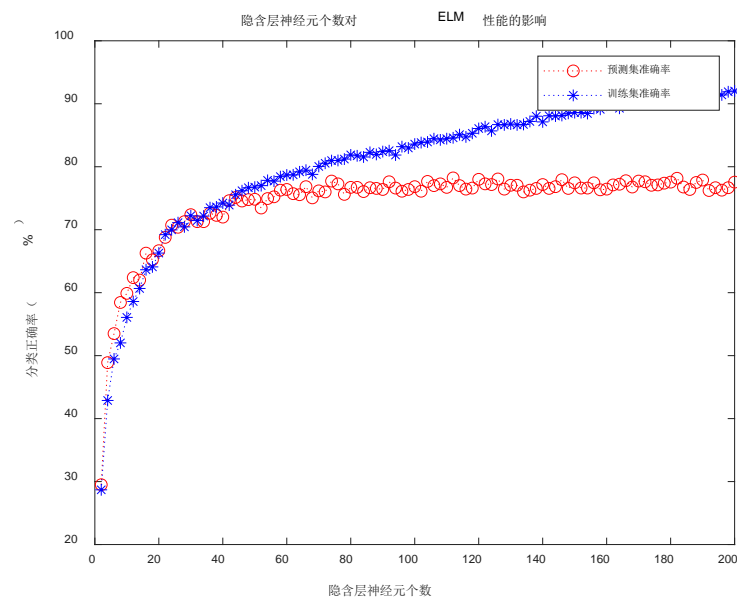


Figure 8. Effect of the number of hidden layer neurons of characteristic band spectrum on ELM performance

图 8. 特征波段光谱的隐含层神经元个数对 ELM 性能影响图

预测类别，蓝色记号代表大豆种子的实际类别，其中红色记号与蓝色记号的重合部分代表大豆品种预测正确，未重合部分代表大豆品种预测错误。红色记号与蓝色记号重合越多，预测正确率越高，反之预测正确率越低。在图 9 中，红色记号与蓝色记号基本全部重合，代表分类正确率较高，预测效果较好。图 10 为特征波段反射光谱的随机森林中决策树棵树对性能的影响图，其中红色记号代表预测集正确率，蓝色记号代表训练集正确率，横坐标为随机森林中决策树棵树，纵坐标为分类正确率。由图可知，其训练

集分类正确率和预测集分类正确率均较高, 预测效果较好。输出结果显示, 大豆种子样品的特征波段反射光谱经过随机森林模型分类, 其分类正确率为 98.89%, 经过计时, 该模型预测时间为 12 s。

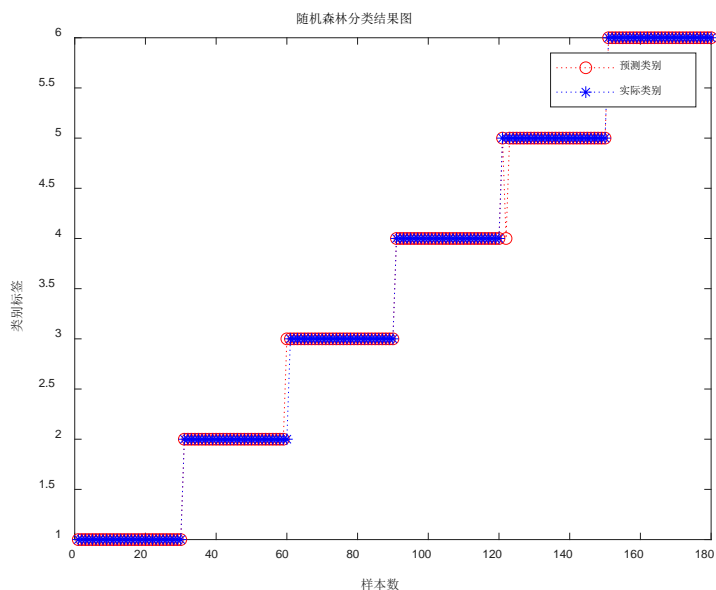


Figure 9. Classification results of soybean varieties in random forest with characteristic band spectrum

图 9. 特征波段光谱的随机森林的大豆品种分类结果图

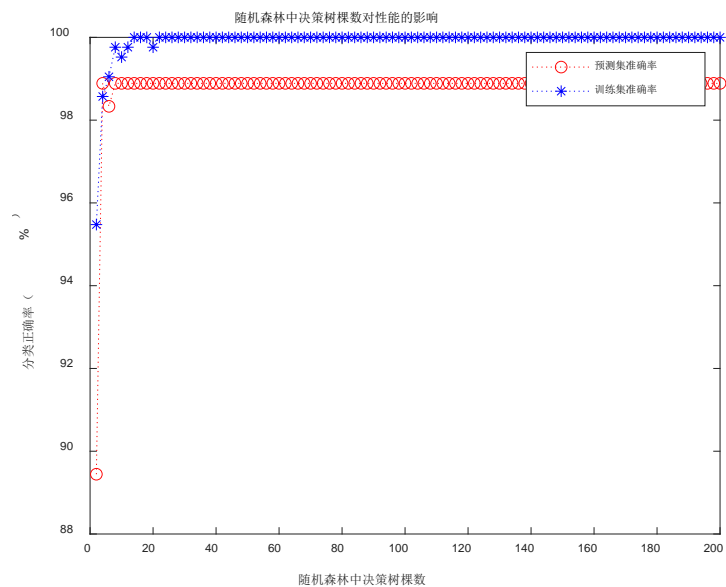


Figure 10. Effect of decision tree on performance in random forest with characteristic band spectrum

图 10. 特征波段光谱的随机森林中决策树棵数对性能的影响图

5.3. 大豆品种分类鉴别的方法比较

对比特征波段的反射光谱建立的极限学习机(ELM)和随机森林(RF)的模型正确率和模型预测时间, 得

到 RF 模型的分类正确率远高于 ELM 模型, ELM 模型的预测时间仅比 RF 模型的预测时间快 1 s。综合比较, 经过 S-G 平滑和高光谱波段优化的特征波段运用随机森林(RF)模型是分类准确率较高、预测时间较快的分类方法, 正确率达到 98.89%, 预测时间为 12 s, 可完成对大豆种子品种的分类鉴别。

6. 结论

采用高光谱成像法, 获取了 6 种大豆样品在 392.38~1011.01 nm 的高光谱图像, 提取了其反射光谱曲线。根据不同种类大豆的反射光谱曲线差异选取了 8 个特征波段, 采用随机森林(RF)和极限学习机(ELM)进行了分类, 结果 RF 所得的分类正确率为 98.89% 高于 ELM 的准确率 78.22%, 且模型预测时间分别为 12 s 和 11 s。本文得到了大豆品种分类准确率尽可能高、尽可能快的高光谱数据处理分析手段, 为大豆种子品种分类鉴别和智慧农业提供了参考。

基金项目

吉林省自然科学基金项目(2020122348JC), 吉林省发改委创新能力建设项目(2020C019-6)。

参考文献

- [1] 雷裕, 胡新军, 蒋茂林, 黄志雄, 马小燕, 田建平, 黄丹. 高光谱成像技术应用于畜禽肉品质研究进展[J]. 食品安全质量检测学报, 2021, 12(21): 8404-8411.
- [2] Ma, J., Sun, D.W., Pu, H.B., *et al.* (2019) Advanced Techniques for Hyperspectral Imaging in the Food Industry: Principles and Recent Applications. *Annual Review of Food Science and Technology*, **10**, 197-220. <https://doi.org/10.1146/annurev-food-032818-121155>
- [3] Dietrich, M., Marx, S., Forst, M., *et al.* (2021) Bedside Hyperspectral Imaging Indicates a Microcirculatory Sepsis Pattern—An Observational Study. *Microvascular Research*, **136**, Article No. 104164. <https://doi.org/10.1016/j.mvr.2021.104164>
- [4] 白玉, 刘丽娜, 张宁, 林晨, 宋维, 朱新忠. 基于改进 RX 增量学习的高光谱图像异常检测[J]. 电光与控制, 2022, 29(2): 16-19+48.
- [5] Zhu, S.L., Chao, M.N., *et al.* (2019) Identification of Soybean Seed Varieties Based on Hyperspectral Imaging Technology. *Sensors*, **19**, Article No. 5225. <https://doi.org/10.3390/s19235225>
- [6] Wang, R., Tan, K., Li, M., *et al.* (2019) Discriminating Soybean Seed Varieties Using Hyperspectral Imaging and Machine Learning. *Journal of Computational Methods in Sciences and Engineering*, **19**, 1001-1015. <https://doi.org/10.3233/JCM-193562>
- [7] 李亚婷. 基于高光谱图像技术的大豆种子活力与生活力无损检测研究[D]: [硕士学位论文]. 镇江: 江苏大学, 2020.
- [8] 于明, 邳艳芹. 一种图像的感兴趣区域提取方法[J]. 电子设计工程, 2012, 20(7): 160-162.
- [9] 董涛. 基于表面增强拉曼光谱的土壤农残快速检测方法研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2019.
- [10] 陈诚, 廖桂平, 李锦卫, 等. 基于 GA、BP 神经网络和多元回归的集成算法研究[J]. 计算技术与自动化, 2011, 30(2): 89-95.
- [11] 饶姗姗, 冷小鹏. 基于组合特征选择的随机森林信用评估[J]. 计算机系统应用, 2022, 31(3): 345-350.