

结合高光谱成像和机器学习的棉种年份鉴别

段 龙¹, 鄢天荣¹, 王江丽^{2,3}, 叶伟欣¹, 陈 伟¹, 高 攀^{1,2*}, 吕 新^{2,3*}

1. 石河子大学信息科学与技术学院, 新疆 石河子 832003
2. 新疆生产建设兵团生态农业重点实验室, 新疆 石河子 832003
3. 石河子大学农学院, 新疆 石河子 832003

摘 要 棉花精量播种技术目前已经在新疆兵团全面推广, 该技术能精确实现一穴一粒的农艺技术指标, 但是也对高质量棉种的筛选提出了更高的要求。为了避免播种往年活力不足的棉种而导致发芽率降低的问题, 结合机器学习和近红外(NIR)高光谱成像技术(HSI)进行棉种年份精确鉴别, 实现棉种的快速无损筛选。采集2016年—2019年近四年外观无明显差异的棉种各360粒, 共1440粒棉种(按照3:1:1划分训练集、验证集和测试集)作为样本, 按照每批60粒采集915~1698 nm范围的棉种高光谱图像, 去除首尾两端噪声大的光谱, 保留1002~1602 nm范围的光谱为原始数据。利用Savitzky-Golay(SG)平滑算法对光谱进行预处理, 采用主成分载荷方法(PCA-loading)选取13个特征波段, 基于全部光谱数据和特征波段(± 10 nm)数据建立逻辑回归(LR)、偏最小二乘判别分析(PLS-DA)、支持向量机(SVM)、循环神经网络(RNN)、长短记忆网络(LSTM)和卷积神经网络(CNN)六种分类模型。使用全光谱数据建模时, 六种分类模型在测试集上的鉴别准确率分别为96.27%, 98.98%, 99.32%, 96.95%, 97.63%和100%, 其中CNN和SVM模型取得了较好的结果; 使用特征光谱数据建模时, 六种分类模型在测试集上的鉴别精度分别为93.56%, 97.29%, 98.30%, 95.25%, 94.24%和99.66%, 其中CNN和SVM模型仍有较好的分类结果。结果表明, 使用全光谱数据建模时, 六种分类模型都可以实现较高精度的棉种年份鉴别, 使用特征光谱数据建模时CNN和SVM模型的鉴别精度仍可达到98%; 其中深度学习方法优于传统机器学习方法, 但是传统机器学习方法仍能保持较好的鉴别准确率。因此, 结合近红外高光谱成像技术和机器学习方法能够实现棉种年份的高精度鉴别, 为棉花精量播种过程中的优质棉种选种技术提供理论依据和方法。

关键词 高光谱成像; 棉种年份鉴别; 卷积神经网络; 机器学习

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)12-3857-07

引 言

棉花种植面积不断扩大, 为深入推进棉花全程机械化战略目标, 棉花精量播种技术受到了充分重视。该技术要求精确实现一穴一粒的农艺技术指标, 因此对棉种质量提出了严格的要求。棉种破损程度和储存的年限是棉种质量的两个重要评价指标。高攀等^[1]利用近红外高光谱成像技术实现了微破损棉种的无损检测; 但对外观无明显差异, 年份却不同的棉种鉴别尚未见相关研究。储藏时间较长的棉种, 其种子活力和发芽率会严重降低^[2], 混入正常棉种中并进行精量播种会影响棉花出苗率, 最终影响棉花产量。因此, 准确有效鉴

别棉种年份尤为重要。

种子年份鉴别的传统方法包括经验鉴别和化学分析, 采用经验鉴别时, 对鉴别经验要求较高, 鉴别准确率不稳定; 化学分析因其破坏性采样和技术要求高等因素, 无法大规模检测。高光谱成像技术(hyperspectral imaging, HSI)有着近似连续的光谱信息、探测识别能力极大提高, 其在果蔬、谷物的品质检测、成分分析等方面广泛应用^[3-4]。Feng等^[5]利用HSI成功寻找到一种鉴别不同加速老化时间下玉米籽粒活力的方法, 不同老化时间下的鉴别准确率从61%到100%不等。在年份鉴别方面, Huang等^[6]基于HSI开展了不同年份玉米种子分类研究, 使用最小二乘支持向量机方法得到了精度为94.4%的分类结果。研究表明, HSI应用在分类检

收稿日期: 2020-11-14, 修订日期: 2021-02-26

基金项目: 国家自然科学基金项目(61965014), 兵团重大科技项目(2018AA00405) 资助

作者简介: 段 龙, 1998年生, 石河子大学信息科学与技术学院硕士研究生 e-mail: duanlong@stu.shzu.edu.cn

* 通讯作者 e-mail: gp_inf@shzu.edu.cn; lxshz@126.com

测、成分分析、年份鉴别等方面是可行性。

HSI 具有波段多、光谱分辨率高、信息量大、信息相关性强等特点^[7], 因此, 选取合适高效的数据分析处理方式对充分利用 HSI 尤为重要。深度学习作为机器学习的一个新的研究方向, 具有特征学习、特征提取和海量数据处理能力^[8], 并且在光谱分析和高光谱图像处理领域有所应用, Gao 等^[9]将 HSI 与深度学习相结合, 成功鉴别了沙枣的不同产地, 鉴别精度达到 97.79%, Feng 等^[10]基于 HSI 和深度学习方法成功检测了冬枣中的细微损伤, 不同产地的冬枣检测精度均在 90%~100% 之间。研究表明, 机器学习和深度学习可用于 HSI 处理, 并实现种类鉴别。

本文旨在通过采集棉种的高光谱图像, 并结合机器学习

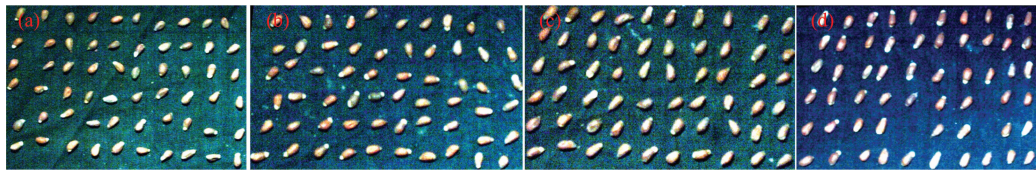


图 1 样本信息

Fig. 1 Sample information

在图 1 中, (a)—(d) 四子图中分别为 2019 年, 2018 年, 2017 年和 2016 年的样本。为了建立分类模型, 将样本随机分成训练集、验证集和测试集 (3 : 1 : 1)。具体划分如表 1 所示。

表 1 样本划分

Table 1 Sample classification

	2016 年 棉种/粒	2017 年 棉种/粒	2018 年 棉种/粒	2019 年 棉种/粒
训练集	208	208	208	208
验证集	77	77	77	77
测试集	75	75	75	75

1.2 仪器与高光谱数据采集

采用近红外 (near infrared, NIR) 高光谱成像系统拍摄了近四年不同年份的棉种表面。NIR 高光谱成像系统由四个模块组成, 包括成像模块、照明模块、升降模块和软件模块, 其具体信息如表 2 所示。

表 2 仪器信息

Table 2 Instrument information

模块	设备信息
成像模块	Surface Optics Corporation (SOC) 710 SWIR 摄像机 (Surface Optics Corporation, San Diego, California, USA)
照明模块	卤素灯, 单盏卤素灯功率为 50 W, 共使用 4 盏卤素灯
软件模块	HyperScanner V2, 用于采集数据

NIR 高光谱成像系统采集的图像尺寸为 288×512×640 (波段数×像素宽×像素长), 在拍摄过程中将 60 个同一年

和深度学习进行特征选择与特征提取, 对不同年份棉种进行快速无损鉴别, 为棉花精量播种过程中优质棉种选种技术提供理论依据和方法。

1 实验部分

1.1 材料

实验样品为新疆石河子市棉花研究所 2016 年—2019 年四年的新陆早 71 号棉种。取 4 个年份棉种各 360 粒, 共采集 1 440 份无破损、外观无明显差异的棉种样品, 样品如图 1 所示。

份的棉种样品放置在黑色面板上并调整 SOC 与样本的距离为 79 cm, 使成像模块能够捕捉到所有样本。HSI 采集后, 根据式 (1) 将原始 HSI 进行校正^[11]。

$$I_r = \frac{I_o}{2I_g} \quad (1)$$

式 (1) 中, I_r 是校正后的值, I_o 对应原始图像, I_g 是灰色 (50% 黑色和 50% 白色组合) 参考图像 (使用 SOC 提供的灰色参考板拍摄)。

1.3 数据预处理

考虑到拍摄过程中存在明显的电信号噪声, 采用 SG 平滑滤波器对图像进行预处理 (核大小为 5×5×5, 多项式阶数为 3), 以减少随机噪声。用掩模法逐个提取单粒棉种的 HSI, 计算并平均反射光谱。利用 SOC 获得的 1 100 nm 波段的图像来计算掩模, 统计含有单粒棉种的掩模的坐标。根据坐标从 HSI 中提取含有单粒棉种的子图像, 并提取出计算的平均光谱。光谱提取过程如图 2 所示。

1.4 分类模型

1.4.1 逻辑回归

逻辑回归 (logistic regression, LR) 在机器学习和统计领域是最经典的分类模型之一, 可以获得良好的鉴别结果^[12]。对于 LR 模型, 通过网格搜索算法选择合适的正则化参数、损失函数优化算法和正则化系数 C 来优化模型, 正则化参数选择 L2, 损失函数优化算法和正则化系数 C 的调参范围分别为 {newton-cg, lbfgs, liblinear, sag} 和 10^{-5} –50。

1.4.2 偏最小二乘判别分析

偏最小二乘法判别分析 (partial least squares discriminant analysis, PLS-DA), 是多变量数据分析中常用的判别分析法, 是最常用的分类模型之一^[1]。PLS-DA 具有模型参数少、解释样本观测数目少和可以减少变量间多重共线性产生的影响等优势。在本实验中, 对于 PLS-DA 模型, 通过网格搜

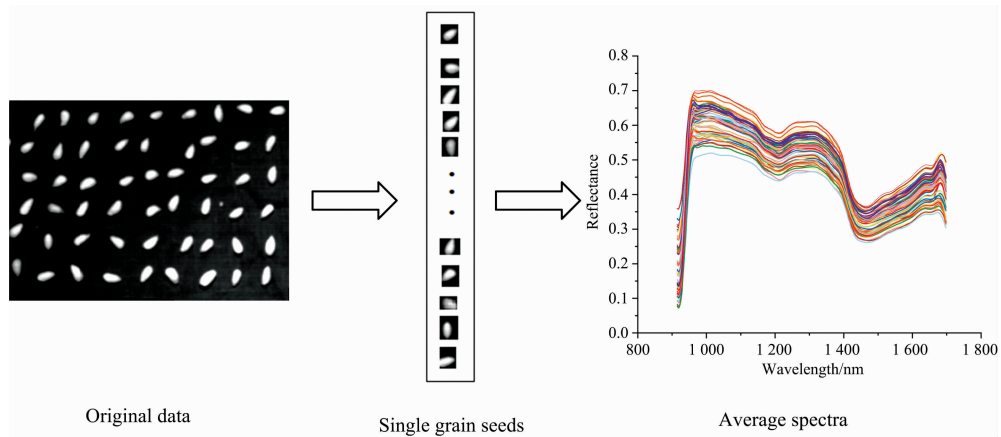


图 2 光谱提取过程

Fig. 2 Extraction process of spectra

索算法选择合适的主成分数量来优化模型, 主成分数量的调优范围为 2~20。

1.4.3 支持向量机

支持向量机(support vector machine, SVM)是常用的分类模型, 可用于定量和定性分析^[13], 其泛化能力强, 通过最大化决策边界来获取稳定的分类结果。在此, 对 SVM 模型, 通过网格搜索算法选择合适的核函数、正则化系数 C 和核系数 γ 来优化模型, 核函数、正则化系数 C 和核系数 γ 的调参范围分

别为 {poly, rbf, sigmoid}、 10^{-5} -50、 10^{-5} -50。

1.4.4 循环神经网络

循环神经网络 (recurrent neural network, RNN) 常用于处理序列化数据, 它能够有效获取序列变化数据的特征信息^[14]。实验所采集的光谱数据也为序列数据, RNN 可用作探索不同波段之间的关联关系, 理论上会有良好的分类效果。在此, 设计包含两个部分的 RNN 结构, 如图 3 所示。

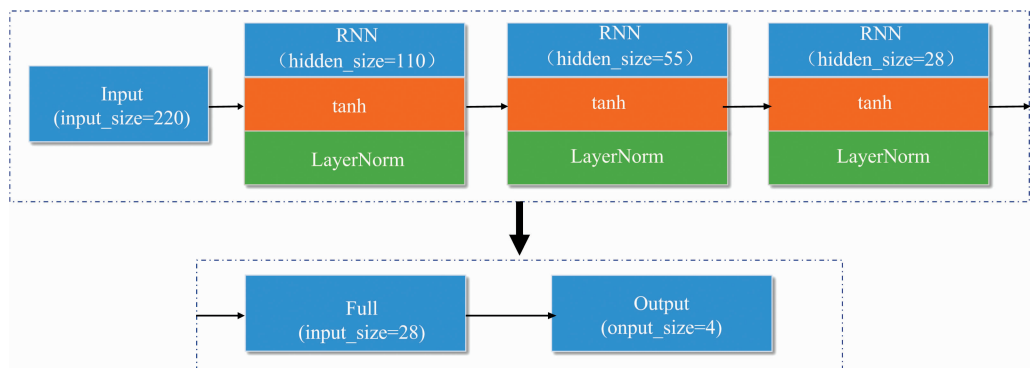


图 3 RNN 结构图

Fig. 3 RNN structure

在图 3 中, 第一部分为输入层和三个 RNN 层, 输入层的特征数为 220, RNN 层的隐藏层大小分别为 110, 55, 28, RNN 每一层包含层归一化以及激活函数, 使用 LayerNorm 进行层归一化, 选取 tanh 为激活函数; 第二部分由全连接层和输出层组成, 输出的类别包含 4 类。

1.4.5 长短记忆网络

长短记忆网络(long short-term memory, LSTM)是一种特殊的 RNN 结构, 能够更好地解决长序列训练过程中的梯度消失和梯度爆炸问题^[15], 能够在处理更长的序列化数据中有更好的表现。由于所采集的高光谱数据有大量波段, 所以采用 LSTM 在面对长序列光谱数据会有更好的分类效果。设计包含两个部分的 LSTM 结构, 如图 4 所示。

在图 4 中, 第一部分为输入层和三个 LSTM 层, 输入层

的特征数为 220, LSTM 层的隐藏层特征数分别为 110, 55, 28, LSTM 每一层都包含层归一化以及激活函数, 使用 LayerNorm 进行层归一化, 选取 tanh 为激活函数; 第二部分由全连接层和输出层组成, 输出的类别包含 4 类。

1.4.6 卷积神经网络

卷积神经网络(convolutional neural networks, CNN)在特征选择与提取、识别与分类等方面的表现优于浅层网络, 具有两点优势: (1) 局部感知, 即 CNN 仅对数据的局部元素进行感知, 然后在更高层的网络将这些局部的信息进行融合, 从而得到数据的全部表征信息; (2) 权值共享的网络结构使网络模型的复杂度降低, 使权值的数量减少。CNN 现已广泛应用于多个领域^[8-10], 但尚未发现被应用于棉种的高精度年份鉴别。常见的 CNN 结构由以下六层组成: 输入层、卷

积层、激活层、池化层、全连接层和输出层,在本实验中,所构建的 CNN 结构如图 5 所示。

在图 5 中, CNN 结构包括输入层、输出层、全连接层,以及 3 个卷积层和 3 个池化层;每个卷积层与池化层之间包含层归一化和激活函数,使用 BatchNorm 进行层归一化,选

取 ReLU 为激活函数;其中输入层为训练集数据输入,共 220 个波段点,输出层为类别的预测值,共四类。卷积层的卷积核大小为 1×3 , 每个卷积层卷积核的个数分别为 100, 50 和 20。

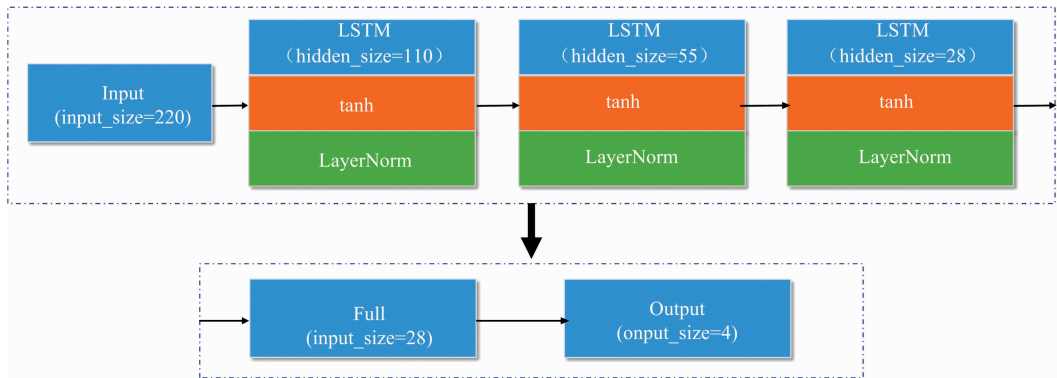


图 4 LSTM 结构图

Fig. 4 LSTM structure

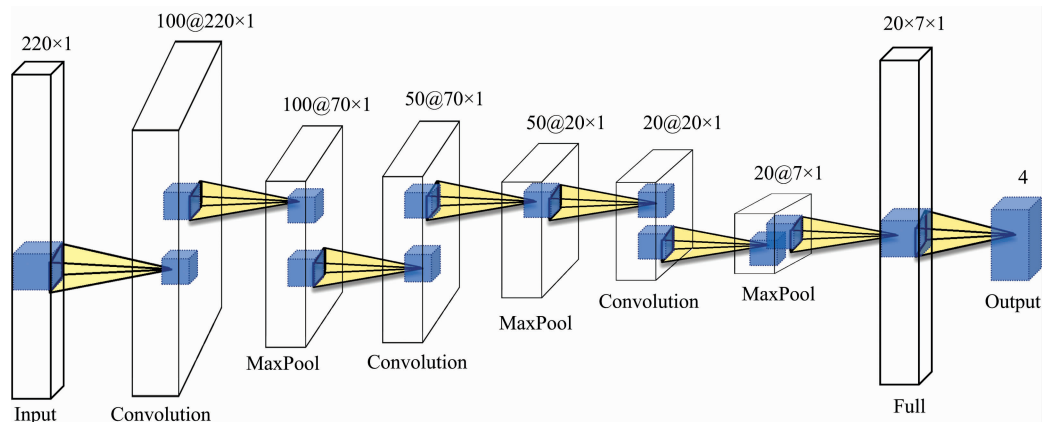


图 5 CNN 结构图

Fig. 5 CNN structure

2 结果与讨论

2.1 棉种的光谱曲线

采集的近红外光谱在 915~1 699 nm 范围内共 288 各波段。在拍摄的过程中,由于机器的启动和结束,所获取的高光谱数据起止波段部分有较为明显的噪声,为了保证数据的可靠性,截去前后两端的波段,选取 1 002~1 600 nm,其光谱数据如图 6 所示。

在图 6 全光谱曲线图中可以看到棉种的光谱反射率曲线的变化趋势具有明显的相似性,在平均光谱曲线图中不同年份棉种的光谱虽然有少许差异,但这些差异并不足以用作棉种年份的高精度鉴别。

2.2 主成分分析

主成分分析(principal components analysis, PCA)是一

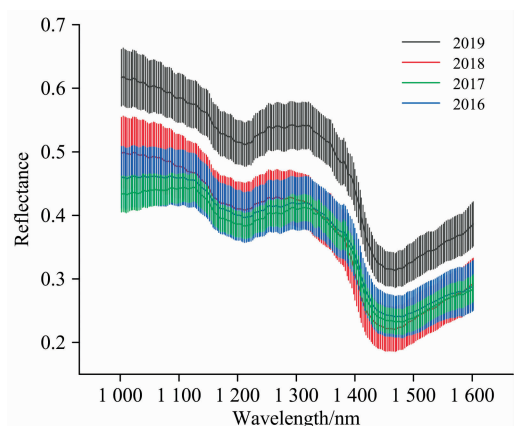


图 6 棉种光谱

Fig. 6 Spectra of cotton seeds

种通过数据降维将多变量转化为几个能够描述数据信息的主成分的统计方法。利用 PCA 对棉种的光谱数据进行分析, 选取贡献率分别为 96.8%, 2.7% 和 0.2% 的主成分 PC1, PC2 和 PC3, 贡献率累计达到 99.7%, 涵盖了绝大多数的数据信息。PC1, PC2 和 PC3 的三维得分图如图 7 所示。

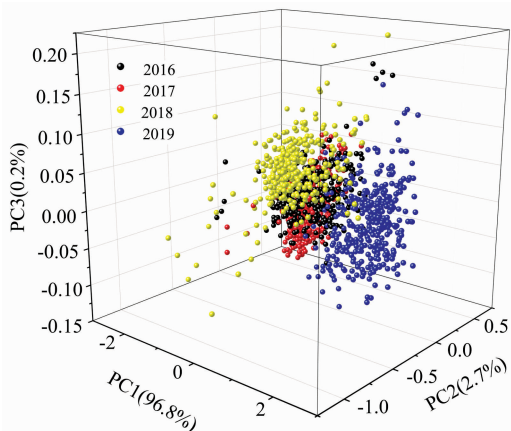


图 7 PCA 得分散点图
Fig. 7 Scores scatter plots of PCA

图 7 是 PCA 的 3 维得分图, 使用 PCA 方法降低了数据维度, 但是图中大部分数据仍然交错在一起, 加入 PCA 方法仍无法实现棉种年份的高精度鉴别。因此, 需要对棉种光谱数据进一步的分析与处理。

2.3 特征波段选择

基于全部光谱数据建立分类模型时, 常常因数据量过大而造成数据冗余和数据的共线性, 不利于有效提取光谱信息, 影响模型分类效果。为了减少数据计算量, 降低模型复杂度, 需要找到与棉种年份相关性较强的特征波段。在本实验中, 采取了 PCA-loading 方法提取特征波段。

PCA-loading 法可以反映主成分与原始光谱波段之间的相关性, 波段所对应的 loading 值越大, 表明其对主成分的贡献率越大^[16]。首先明确不同主成分的贡献率, 然后基于累积贡献率选取主成分, 最后设定阈值, 并以波段载荷图为依据筛选出波峰或波谷作为特征波段, 选择特征波段的过程如图 8 所示。

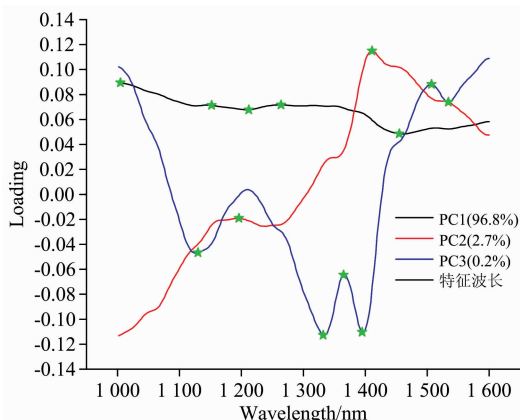


图 8 特征光谱选取
Fig. 8 Characteristic wavelength selection

经筛选, 选取了 1 005, 1 128, 1 152, 1 196, 1 212, 1 264, 1 332, 1 365, 1 395, 1 411, 1 455, 1 507 和 1 534 nm 共 13 个特征波段, 将特征波段(± 10 nm)用作模型训练。

2.4 模型建立与结果分析

建立了六种分类模型, 包括传统机器学习方法 LR, PLS-DA 和 SVM, 深度学习模型 RNN, LSTM 和 CNN。为了便于分类模型的建立, 按照划分好的数据集将去除噪音的数据作为输入数据, 所使用的 python 版本为 3.7.0, 使用 scikit-learn0.2.1 构建 LR、PLS-DA 和 SVM 模型, 使用 pytorch1.5.0 构建 CNN、RNN 和 LSTM 模型, 模型的迭代次数设置为 200 次。六种不同的分类模型分类结果如表 3 所示。

表 3 不同模型鉴别结果

Table 3 Identification results of different models

	全光谱			特征光谱		
	训练集	验证集	测试集	训练集	验证集	测试集
LR	0.967 4	0.979 0	0.962 7	0.931 3	0.930 1	0.935 6
PLS-DA	0.990 7	1.000 0	0.989 8	0.987 2	0.993 0	0.972 9
SVM	0.996 5	0.986 0	0.993 2	0.988 4	0.979 0	0.983 0
RNN	0.967 4	0.961 5	0.969 5	0.956 9	0.954 5	0.952 5
LSTM	0.980 2	0.968 5	0.976 3	0.966 2	0.965 0	0.942 4
CNN	1.000 0	1.000 0	1.000 0	1.000 0	0.986 0	0.996 6

在表 3 中, 对于棉种的全光谱近红外光谱数据, 六种模型均具有良好的性能, 训练集、验证集和测试集的分类准确率均超过 96%。SVM 和 CNN 模型的结果接近, 均超过 99%, 而 LR, PLS-DA, RNN 和 LSTM 模型的结果相对较低。LR 模型采用 L2 范式作为损失函数, 模型参数(C, solver)经调优为(48.7, lbfgs), 其在训练集、验证集和测试集中的分类准确率超过 96%。在 PLS-DA 模型中, 保留的主成分数为 7 个, 其在训练集、验证集和测试集中的分类准确率均超过 98%。对于 SVM 模型, 最终模型参数(kernel, gamma, C)经调优为(rbf, 3.2, 20.4), 训练集, 验证集和测试集的分类准确率均超过 98%。对于 RNN, LSTM 和 CNN 三个深度学习模型, CNN 的分类精度在划分的三个数据集中均达到 100%, 而 RNN 和 LSTM 仅达到 96%。由此可见, 在处理全光谱数据时, CNN 相较于其他深度学习模型有更好的表现, PLS-DA 和 SVM 相较于 LR 有更好的表现。

对于棉种的特征光谱数据, 六种模型分类效果都有不同程度的降低, 但仍具有较好的性能。SVM 和 CNN 模型的结果接近, 测试集精度均超过 98%, 相较于全光谱数据, 精度降低 1%。PLS-DA 模型精度降低 2%, RNN 精度降低了 1%, LSTM 精度降低 2%, 但是仍保持在 94% 以上, 但是 LR 模型分类精度仅有 93%, 精度降低 4%。在特征光谱数据的训练中, LR 模型采用 L2 范式作为损失函数, 模型参数(C, solver)经调优为(48.1, newton-cg); PLS-DA 模型经调优, 保留 15 个主成分; 对于 SVM 模型, 最终模型参数(kernel, gamma, C)经调优为(rbf, 22.0, 24.4)。对于 RNN, LSTM 和 CNN 三个深度学习模型, CNN 的分类精度最高,

所降低的精度也最小, 而 RNN 和 LSTM 仅达到 94% 和 95%。由此可见, 在处理特征光谱数据时, CNN 相较于其他深度学习模型有更好的表现, PLS-DA 和 SVM 优于 LR。

3 结 论

结合机器学习和近红外高光谱成像技术, 实现了棉种的年份精确鉴别。采用 SG 方法进行光谱平滑, 使用 PCA-loading 方法进行特征波段选取, 使用六种分类模型(LR, PLS-DA, SVM, RNN, LSTM 和 CNN)分别对棉种的全谱和特征光谱数据进行建模分析。结果表明, CNN 和 SVM 分类模型

的效果优于其他模型, 深度学习分类模型优于传统机器学习。由于选取特征光谱, 丢失了部分数据信息, 用特征光谱数据建模, 分类效果会有不同程度的降低, 其中 CNN 和 SVM 两个分类模型仅降低 1%, 测试集精度仍达到 98%, PLS-DA 在测试集上的分类精度达到 97%, 其他三种模型分类效果仅达到 94% 左右。因此, 在采用近红外高光谱数据进行棉种年份鉴别时, CNN 和 SVM 相较于其他四种分类模型是更好的选择。结合整体的研究结果, 将高光谱成像技术与机器学习相结合可用于棉种年份精确鉴别, 为完善棉花精量播种技术提供了理论依据与方法。

References

- [1] GAO Pan, ZHANG Chu, LÜ Xin, et al(高攀, 张初, 吕新, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(6): 1712.
- [2] FU Yuan-zhi, CHAO Mao-ni, HU Gen-hai, et al(付远志, 晁毛妮, 胡根海, 等). Seed(种子), 2018, 37(7): 31.
- [3] JIA Min, OU Zhong-hua(贾敏, 欧中华). Acta Laser Biology Sinica(激光生物学报), 2018, 27(2): 119.
- [4] Zhang C, Wang Q, Liu F, et al. Measurement, 2017, 97: 149.
- [5] Feng L, Zhu S, Zhang C, et al. Molecules (Basel, Switzerland), 2018, 23(12): 3078.
- [6] Huang M, Tang J, Yang B, et al. Computers and Electronics in Agriculture, 2016, 122: 139.
- [7] Gao F, Wang Q, Dong J, et al. Remote Sensing, 2018, 10(8): 1271.
- [8] Garcia-Gasulla D, Parés F, Vilalta A, et al. Journal of Artificial Intelligence Research, 2018, 61: 563.
- [9] Gao P, Xu W, Yan T, et al. Foods, 2019, 8(12): 620.
- [10] Feng L, Zhu S, Zhou L, et al. IEEE Access, 2019, 7: 64494.
- [11] Yan T, Duan L, Chen X, et al. RSC Advances, 2020, 10: 41936.
- [12] Wu N, Zhang C, Bai X, et al. Molecules (Basel, Switzerland), 2018, 23(11): 2831.
- [13] Okwuashi O, Ndehedehe C E. Pattern Recognition, 2020, 103: 107298.
- [14] Venkatesan R, Prabu S. Journal of Medical Systems, 2019, 43(7): 216.
- [15] Li Y, Cao H. Procedia Computer Science, 2018, 129: 277.
- [16] Jiang H, Jiang X, Ru Y, et al. Infrared Physics & Technology, 2020, 110: 103467.

Combine Hyperspectral Imaging and Machine Learning to Identify the Age of Cotton Seeds

DUAN Long¹, YAN Tian-ying¹, WANG Jiang-li^{2,3}, YE Wei-xin¹, CHEN Wei¹, GAO Pan^{1,2*}, LÜ Xin^{2,3*}

1. College of Information Science and Technology, Shihezi University, Shihezi 832003, China

2. The Key Laboratory of Oasis Eco-Agriculture, Xinjiang Production and Construction Corps, Shihezi 832003, China

3. College of Agriculture, Shihezi University, Shihezi 832003, China

Abstract At present, the technology of precision cotton seeding has been promoted comprehensively in Xinjiang Corps, which can accurately achieve the agronomic technical standards of one grain per hole, but it also sets higher demands for the screening of high-quality cotton seeds. To avoid the decrease of germination rate caused by the cotton seeds with lack of vitality in previous years, machine learning and near-infrared (NIR) hyperspectral imaging (HSI) technology can be used to identify cotton seed years with high precision and to screen cotton seeds quickly and nondestructively. A total of 1 440 cotton seeds with no difference in appearance were collected in 2016, 2017, 2018, and 2019, and 360 seeds per year (According to 3 : 1 : 1, it is divided into the training set, validation set, and test set.) as samples. Hyperspectral images of cotton seeds in the range of 915~1 698 nm were collected according to each batch of 60 seeds, and average spectra (1 002~1 602 nm) for removing obvious noise at the beginning and the end were extracted as the raw data. SavitzkyGolay (SG) smoothing algorithm was used to preprocess the spectra. The principal component analysis loading (PCA-loading) method was used to select 13 effective wavelengths. Six

classification models, including logistic regression (LR), partial least squares discriminant analysis (PLS-DA), support vector machine (SVM), recurrent neural network (RNN), long-short memory network (LSTM), and convolution neural network (CNN), were established based on full spectra and effective wavelengths. When using full spectra to build models, the identification accuracy of the six classification models on the test set was 96.27%, 98.98%, 99.32%, 96.95%, 97.63%, and 100%, respectively, among which CNN and SVM models had achieved good results. When using effective wavelengths to build models, the identification accuracy of the six classification models on the test set was 93.56%, 97.29%, 98.30%, 95.25%, 94.24%, and 99.66%, respectively, among which CNN and SVM models still had excellent classification results. The results showed that the six classification models could achieve high precision cotton seed years identification when the full spectra were used, and the identification accuracy of CNN and SVM models was still up to 98% when the effective wavelengths were used. The deep learning methods are generally better than the traditional machine learning methods, but traditional machine learning methods can still maintain good identification accuracy. Therefore, the combination of near-infrared hyperspectral imaging technology and machine learning methods can achieve high-precision identification of cotton seed years. It provides theories foundation and methods for selecting high-quality cotton seeds in the process of precision sowing.

Keywords Hyperspectral imaging; Cotton seed year-identification; Convolution neural network; Machine learning

(Received Nov. 14, 2020; accepted Feb. 26, 2021)

* Corresponding authors

本 刊 声 明

近期以来, 一些不法分子假冒《光谱学与光谱分析》期刊社名义, 以虚假网站等形式欺骗广大作者、读者。这些虚假网站公然假冒《光谱学与光谱分析》期刊名义进行大肆的征稿并骗取作者的审稿费和版面费。经部分作者及读者举报, 现有关部门已就此介入调查。本刊将通过法律途径向假冒者追究相应的责任, 维护本刊权利。

本刊官方网站已正式开通, 网址为

<http://www.gpxygpfx.com/>

在此郑重声明, 本网址为《光谱学与光谱分析》期刊唯一开通运行的官方网站。本刊从未授权任何单位或个人以任何形式(包括网上网下)代理本刊征稿、审稿等项业务。

希望广大读者和作者切实维护好自身的合法权益, 防止受骗上当。

《光谱学与光谱分析》期刊社

2019年3月15日