

基于 Random Forest 的水稻细菌性条斑病识别方法研究

袁培森¹ 曹益飞² 马千里¹ 王浩云¹ 徐焕良¹

(1. 南京农业大学人工智能学院, 南京 210095; 2. 南京农业大学工学院, 南京 210031)

摘要: 为了快速、准确、有效地识别发病早期的细菌性条斑病, 提出基于随机森林(Random forest, RF)算法的水稻细菌性条斑病识别方法, 利用光谱成像技术获取该病害的高光谱数据, 通过多元散射校正减少和消除噪声及基线漂移对光谱数据的不利影响。利用随机森林特征重要性指标, 选取逻辑回归(LR)、朴素贝叶斯(NB)、决策树(DT)、支持向量分类机(SVC)、k最近邻(KNN)和梯度提升决策树(Gradient boosting decision tree, GBDT)算法进行对比试验。同时筛选出12个位于450~664 nm范围内对识别模型有重要影响的光谱波段, 并与全波段进行分类结果比较。试验结果表明: RF算法的分类准确率为95.24%, 与试验选取的其他算法相比, 效果最优, 比NB准确率提高了20.97个百分点; 与全波段分类结果相比, 利用RF算法基于12个波长的识别, 波长数减少了98.05%, 识别精确率为94.66%, 召回率为99.55%, F1值为97.04%, 准确率为94.32%。虽然精确率减少了2.97个百分点、准确率减少了0.85个百分点, 但召回率增加了4.4个百分点、F1值增加了0.67个百分点, 模型精度满足要求。

关键词: 水稻表型; 随机森林; 高光谱成像; 细菌性条斑病; 病害识别

中图分类号: TP391

文献标识码: A

文章编号: 1000-1298(2021)01-0139-07

OSID:



Identification Method of Rice Bacterial Leaf Streak Based on Random Forest

YUAN Peisen¹ CAO Yifei² MA Qianli¹ WANG Haoyun¹ XU Huanliang¹

(1. College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China

2. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China)

Abstract: With the rapid development of rice phenotypic research, rice disease research has also made great progress as an important part of rice phenotypic research. In order to identify bacterial stripe disease quickly, accurately and effectively in the early stages of disease, a method for identifying bacterial stripe of rice based on a random forest algorithm was proposed. The spectral imaging technology was used to obtain hyperspectral data of the disease, and multiple noise correction was used to reduce and eliminate noise and the adverse effects of baseline drift on spectral data. Using the importance index of random forest characteristics, the logistic regression, naive Bayes, decision tree, support vector classifier, k-nearest neighbor and gradient boosting decision tree algorithms were selected for comparative test. At the same time, totally 12 spectral bands which were located in 450~664 nm had an important influence on the recognition model were screened out. The results of classification based on the whole band and the 12 important bands were compared. The experimental results showed that the classification accuracy of RF algorithm was 95.24% compared with other algorithms selected in the experiment, the accuracy was higher than that of NB algorithm by 20.97 percentage points. Compared with the whole band classification results, based on these 12 important bands, the number of bands was reduced by 98.05%, the recognition accuracy was 94.66%, the recall rate was 99.55%, the F1 value was 97.04%, and the accuracy rate was 94.32%. Although the accuracy was reduced by 2.97 percentage points, the accuracy rate was reduced by 0.85 percentage points, the recall rate was increased by 4.4 percentage points, the F1 value was increased by 0.67 percentage points, and the model accuracy was basically maintained. Although the accuracy was reduced, the model structure was more streamlined and the computational complexity was reduced. The research result showed that important bands can be used instead of full

收稿日期: 2020-04-01 修回日期: 2020-05-05

基金项目: 国家自然科学基金项目(61502236、61806097)和大学生创新创业训练专项计划项目(S20190025)

作者简介: 袁培森(1980—),男,讲师,博士,主要从事智能信息、海量数据处理与分析研究,E-mail: peiseny@njau.edu.cn

通信作者: 徐焕良(1963—),男,教授,博士,主要从事农业信息化与大数据技术研究,E-mail: huanliangxu@njau.edu.cn

bands to identify rice bacterial streak disease, which provided new ideas for the identification method of rice bacterial streak disease.

Key words: rice phenotypic; random forest; hyperspectral imaging; bacterial stripe disease; disease identification

0 引言

植物表型是植物在一定环境下可观察的形态特征,在植物保护、育种等领域具有重要应用价值,其研究涉及植物学、数据科学、机器学习等领域^[1-2]。水稻细菌性条斑病属于水稻表型研究的重要内容,是一种由稻黄单胞菌致病变种引发的水稻病害^[3],实时准确地判定水稻细菌性条斑病的分布区域和危害程度是采取有效防控措施和实现精准喷药的关键。目前,水稻细菌性条斑病常用的诊断方法主要依靠人力在田间观察,这种方法不仅费时费力、难以在大范围内展开,还受观察者自身经验的影响^[4],并且细菌性条斑病早期症状并不明显,很容易被忽略。因此,研究既能克服传统观测方法局限性,又能精确、有效诊断早期细菌性条斑病病害的技术迫在眉睫^[5]。

高光谱成像是植物病害检测中常用的监测技术,具有无污染、高效等特点^[6]。通过光谱中的每个波段记录一个图像,反映每个空间像素点的光谱信息,从而弥补了传统光谱分析与成像技术的缺点^[7-8]。研究人员对高光谱成像技术应用于植物病理分析进行了大量的研究。WU等^[9]利用高光谱成像技术结合病理分析,通过聚合酶链式反应标记不同条件下收集的稻粒真实感染状态,使用主成分分析分离健康稻谷粒和受感染稻谷粒,并使用随机森林极限学习机模型(RF-ELM)对不同感染程度的健康和感染混合粒进行分类。雷雨等^[10]对小麦叶片的高光谱图像进行掩膜处理,利用主成分分析法和最大类间方差法获取小麦叶片的条锈病病斑区域,通过面积比判定小麦的条锈病病害等级。LU等^[11]利用远程高光谱遥感监测实时和特定位置的水稻氮、磷和硫等含量。SUN等^[12]结合高光谱技术和全基因组关联分析技术对水稻种子的生化特性进行表型分析,结果表明,归一化光谱指数与蛋白质含量具有高度相关性。GHAMISI等^[13]根据LiDAR和高光谱数据的空间信息进行建模,通过使用支持向量机或随机森林分类技术获得最终的分类图。张帅堂等^[14]提出了基于高光谱成像技术和图像处理技术融合的茶叶病斑识别方法。YEH等^[15]基于高光谱成像技术,利用光谱角度映射器检测3个不同感染阶段的草莓叶状炭疽病。郭伟等^[16]通过高光谱

影像的光谱指数、比值光谱指数及归一化差值光谱指数构建偏最小二乘回归预测模型,反演冬小麦全蚀病情指数。梁琨等^[17]利用连续投影算法和自适应加权算法提取特征波段的光谱图像,识别小麦赤霉病。这些研究表明,高光谱可以用于作物病害检测,并且取得了较好的效果。但基于高光谱成像对细菌性条斑病进行早期识别的研究较少。

张智韬等^[18]探讨了分数阶微分联合支持向量机分类-随机森林模型改善高光谱监测荒漠土壤有机质含量的效果,通过筛选敏感的归一化光谱指数,建立不同分数阶微分的随机森林模型,并以不同土质中的最佳模型进行组合,构建的支持向量机分类和随机森林模型能够快速评估荒漠土壤有机质含量。竞霞等^[19]利用弗劳恩霍夫谱线三波段判别算法提取冠层日光诱导叶绿素荧光数据,结合对小麦条锈病病情严重度敏感的11种反射率微分光谱指数,基于随机森林算法和BP神经网络算法预测小麦条锈病的病情严重度。BISWAS等^[20]利用灰度共生矩阵和随机森林对感染炭疽病、白粉病和霜霉病的葡萄叶片进行分类,并确定其严重程度。目前,基于高光谱影像病害的研究多集中于提取光谱指数,再对病害进行识别和分类,基于光谱本身对病害进行识别的研究较少。

本研究以感染细菌性条斑病的水稻叶片为研究对象,利用随机森林算法基于高光谱曲线本身的光谱信息建立水稻细菌性条斑病早期检测模型,并对特征波段重要性进行筛选,以提高识别性能和效果。

1 试验材料与样本采集

1.1 试验材料

试验水稻栽培品种选用2015年被农业部评为超级稻品种的南梗9108。种植前挑选饱满种粒,浸种、催芽后,于2019年8月17日在南京农业大学植物保护学院温室播种,采用盆栽式,共80杯,每杯6~8颗种子,均匀播种,管理方法相同。水稻出苗后,搬运到室外,在自然条件下生长。试验采用的细菌性条斑病菌种由江苏省农业科学院植物保护研究所提供,菌种为致病性强的B5-16病原菌,水稻细菌性条斑病病原菌培育图如图1所示。

2019年9月21日,水稻苗长到4叶期后,接种细菌性条斑病菌。将每株水稻苗的倒二叶叶片完全

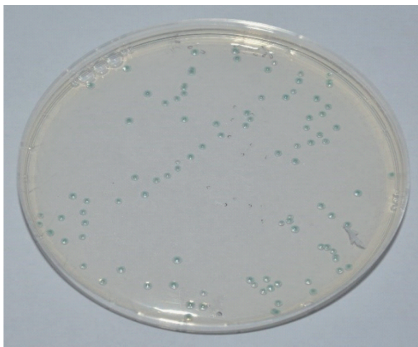


图 1 病原菌培育图
Fig. 1 Diagram of pathogen cultivation

伸展,用酒精消毒的双针头棉花注射器均匀沾取准备好的菌液,将针头口按住叶片,轻轻按动,使菌液通过压力渗入到叶片细胞间隙内^[21]。每片倒二叶首末两端各接种一次,距离保持在 8 cm 左右,水稻接种后如图 2 所示。



图 2 水稻细菌性条斑病菌接种后示意图
Fig. 2 Schematic of rice bacterial stripe inoculation

1.2 高光谱成像系统

本试验采用的推扫型高光谱成像系统(HSI-VNIR-0001, 上海五铃光电科技有限公司)如图 3 所示。采集软件为 Spectral-image 成像软件和 HSI Analyzer 分析软件。高光谱系统采集的水稻叶片光谱波长范围为 373 ~ 1 033 nm,共 616 个波段。选用 17 mm 的成像镜头,确定物距为 27 cm,亮度为 200 lx,调节曝光时间 8 ms 使分辨率达到 3 000 像素。

1.3 数据采集与校正

水稻细菌性条斑病的潜伏期为 3 d,前 3 d 接种点附近的叶片和健康叶片类似。由预试验及植保专家确定,3 ~ 7 d 为爆发期,在此期限内,细菌性条斑病会快速蔓延。由于设备与场地限制,试验在病害早期的第 3 天和第 7 天进行离体拍摄,以 6 ~ 8 片离体水稻叶片样本为 1 组放置在载物台上进行成像,水稻叶片随着载物台匀速移动,高光谱成像仪获取叶片每个像素在各个波长处的光谱信息和图像数据。共获得 1 961 个接种水稻细菌性条斑病原菌的水稻叶片样本的高光谱数据和 400 个接种蒸馏水的水稻叶片样本的高光谱数据。

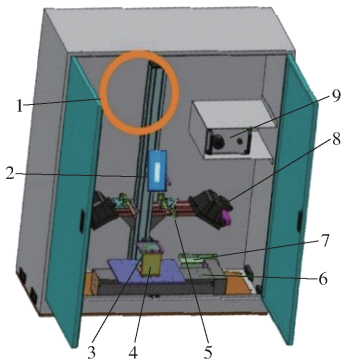


图 3 高光谱成像系统示意图
Fig. 3 Schematic of high-spectral imaging system
1. 支架 2. 成像镜头 3. 光源 4. 固定基座 5. 光源调整基座
6. 步进电机 7. 出线盒 8. 卤素灯 9. 光源底座

对高光谱仪器进行校正后再进行高光谱曲线采集,通过调节光强、图像清晰度、图像失真度来实现高光谱成像系统的校正。物镜之间的高度、电控载物台的移动速度和曝光时间都会影响图像清晰度。为了减少噪声信息对高光谱曲线的影响,进而提高光谱曲线定性或定量分析模型的稳定性和精度,需要对高光谱曲线进行黑白校正^[16],采用 HSI Analyzer 分析软件进行图像处理。先对采集的光谱图像进行校正,图像校正公式为

$$R(i) = \frac{I_r(i) - I_d(i)}{I_w(i) - I_d(i)} \tag{1}$$

式中 $R(i)$ ——校正后的图像相对反射率
 $I_r(i)$ ——试验光谱的原始图像反射率
 $I_w(i)$ ——白板校正后亮电流光谱反射率
 $I_d(i)$ ——黑板校正后暗电流光谱反射率

用水稻样本感兴趣区域 (Region of interest, ROI) 中的所有像素点的光谱平均值作为该样本的平均光谱。

1.4 光谱信息提取

通过 HSI Analyzer 加载水稻样本的高光谱图像,放大水稻叶片的高光谱图像,选择接种条斑病原菌叶片中心 10 像素 × 10 像素的圆形感兴趣区域,计算感兴趣区域内所有像素点的平均光谱,作为水稻感染条斑病叶片样本的原始高光谱数据。

2 试验方法

2.1 数据预处理

常见的高光谱数据预处理方法有 S-G 卷积平滑^[10]、多元散射校正^[14] (Multiple scattering correction, MSC)、标准正态变量变换^[15] (Standard normal variate, SNV) 和一阶导数法^[17]。比较以上 4 种方法对染病水稻数据和健康水稻数据进行平滑

处理的效果,得到经过 S-G、MSC 以及 SNV 处理后的光谱反射率基本保持了原始光谱曲线的走向,但效果不佳,而经过多元散射校正处理后的光谱曲线反射率保持在 0~0.6 的区间范围内,曲线集中、轮廓清晰,处理效果较好,预处理后的光谱曲线如图 4 所示。因此本文选定多元散射校正对数据进行预处理。

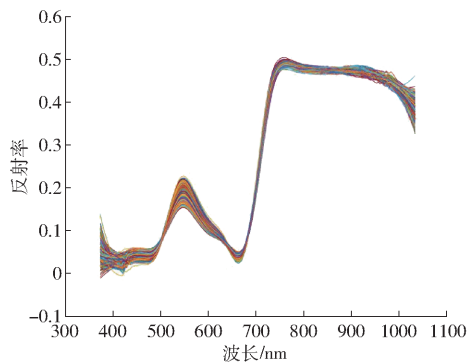


图 4 MSC 平滑效果图

Fig. 4 MSC smoothed effect diagram

2.2 随机森林模型结构及参数

随机森林(Random forest, RF)算法解决了决策树创建不稳定、过拟合等问题,在分类精度上比单株决策树有明显的提高。此外,RF 算法对比其他分类算法具有分类速度快和能够处理高维数据等特点,而且对噪声和孤立点不敏感,不存在过拟合等问题^[22-23]。

将 RF 算法用于叶片高光谱数据集,数据集中的每行数据对应 616 个反射率和标签,共 2 361 行。为保证试验数据划分的随机性和一致性,将 2 361 行水稻叶片高光谱数据以 8:2 的比例随机划分为训练集和测试集^[24],训练集共 1 889 行水稻叶片高光谱数据,用于模型的建立与优化;测试集共 472 个水稻叶片高光谱数据,用于检验。

利用 RF 算法对 1 889 个水稻叶片高光谱数据进行训练。一行叶片数据对应一个决策树,在单棵决策树进行分裂时,有 3 种方法可以选择^[25-26],分别为开平方、取对数和无处理,本模型选择将波段数开平方取整,即从 616 个反射率中随机选取 24 个反射率。用全波段反射率的熵和 24 个反射率的熵做差,算出 24 个反射率的信息增益。信息增益最大的反射率为这棵决策树的根部节点,其余反射率按信息增益大小依次排列,作为其他非叶子节点,建成一个分类决策树。按照以上步骤,生成大量的决策树,建立基于 RF 算法的水稻细菌性条斑病识别模型。在构建 RF 算法时,有 3 个影响性能和效率的重要参数^[27]:

(1) 决策树数量

设 n_{trees} 为森林中决策树数量,即评估器数量。这个参数对 RF 算法的精确性影响是单一的。较多的子树可以让模型有更好的性能,但同时会使模型运行变慢。 n_{trees} 达到一定值后,RF 算法的精确性往往不再上升或开始波动。对于这个参数,需要在训练难度和算法效果之间取得平衡。

(2) 叶节点样本数

随机森林是决策树 B 的集合 $\{T_1(X), T_2(X), \dots, T_B(X)\}$, 其中 $X = (X_1, X_2, \dots, X_p)$ 为分子描述符或与分子有关的 p 维向量。决策树 B 产生输出集合 $\hat{Y}_B = \{T_1(X), T_2(X), \dots, T_B(X)\}$, 其中 $\hat{Y}_B$ ($B = 1, 2, \dots, b$) 是第 B 棵树对一个分子的预测,对所有树的输出进行聚合以产生一个最终预测 $\hat{Y}$, 在本文水稻细菌性条斑病的识别模型中, $\hat{Y}$ 是集成的决策树的预测结果。

将水稻叶片的高光谱数据整理成一组含 n 个训练分子的数据集 $D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$, 其中 X_i ($i = 1, 2, \dots, n$) 是描述符向量, Y_i 对应的标签是 0 或 1, 训练算法步骤如下^[28]: ①从随机水稻叶片训练数据集 D 中抽取一个随机样本,即随机抽样,替换 n 个训练分子。②对于每个 bootstrap 样本,在每个节点上,随机选择随机特征数 m_{try} , 在子集中确定最佳分割位置使树长到最大尺寸(即不可能进一步分裂,停止生长)。③重复步骤①、②,直到(足够大的数量)决策树 B 长至足够大,此时叶节点的样本数为 $n_{estimators}$ 。

(3) 候选特征子集

在决策树生长过程中,所有的随机特征数都在每个节点上测试其分裂性能,而 RF 算法只测试随机特征数 m_{try} 。由于 m_{try} 通常非常小(软件中的默认值是分类描述符数量的平方根),因此运行速度非常快。为了获得具有最佳预测强度的模型复杂度,一般算法是通过交叉验证对单个决策树进行修剪,该过程计算量较大。但是,随机森林不作任何修改。所以在水稻叶片的高光谱数据量非常大的情况下,RF 算法可以比单个决策树在更少的时间内被训练。

根据 RF 算法原理^[29], 本试验取 $m_{try} = p$ 、在所有描述符中选择每个节点的最佳分割点时,RF 算法与 Bagging 相同。

在叶节点样本数 $n_{estimators} = \sqrt{m}$ 的基础上,遍历决策树 1~25, 对应的袋外误差率(OBB error rate)如图 5 所示。当决策树数量为 10 时,袋外误差率为 0.083 743, 开始缓慢下降。当决策树数量小于 5 时,袋外误差率下降幅度较大,当决策树数量等于 15 时,袋外误差率趋于平缓。决策树数量越多,模

型性能越好,但运行效率变低^[30-31]。综上所述,在兼顾分类模型精度和运行速率的条件下,令 $n_{estimators} = 24, n_{trees} = 15$ 。

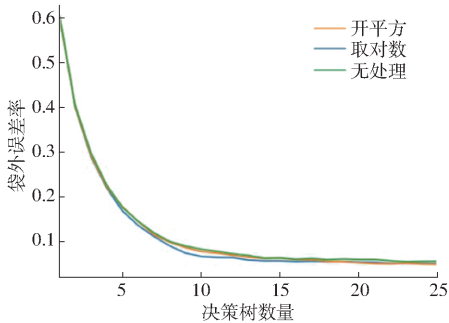


图 5 决策树数量与袋外误差率关系曲线
Fig.5 Relationship between number of decision trees and OBB error rate

2.3 光谱波段选择

水稻叶片的高光谱波长范围为 373.79 ~ 1 033.59 nm,共 616个波段,不是每一个波段对随机森林分类模型都能造成影响,对于重要性较小或者冗余波段,即使受到噪声干扰,对分类结果也没有影响。从 616 个光谱波段中筛选出对分类结果影响较大的光谱波段,用重要波段分类模型代替全波段分类模型,以此来简化随机森林分类模型,减少计算量。本文模型中对波段重要性的排序步骤如下^[26]:
①遍历 616 个波段,利用袋外数据进行预测,将每个波段的预测误差记为 $\{E_1, E_2, \dots, E_m\}$ 。
②对袋外数据的某一自变量值进行干扰,再对袋外数据进行预测,将每个波段的预测误差记为 $\{E_{11}, E_{12}, \dots, E_{1m}\}$ 。
③干扰后的预测误差与干扰前的预测误差取差值,对 m 个差值取平均值,将每个差值与平均值比较,其相差越大则该波段的重要程度越高。

2.4 水稻叶片光谱特性分析

选取经过 MSC 预处理的健康和染病水稻叶片的高光谱数据,对两种不同叶片的光谱曲线进行分析。水稻叶片样本高光谱曲线如图 6 所示。

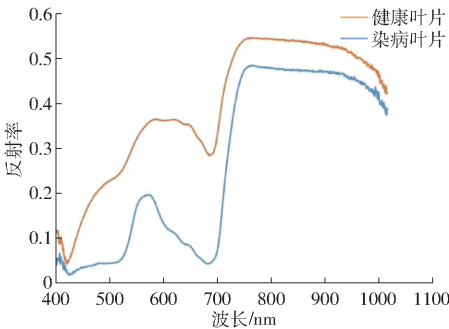


图 6 水稻叶片样本高光谱曲线
Fig.6 Hyperspectral data curves of rice leaf samples

在 616 个全波段水稻叶片高光谱数据中,对本文方法识别水稻细菌性条斑病有重要影响的波段

是 12 个,其中 10 个均匀分布在第 56 号波长 (429.27 nm) 和第 69 号波长 (442.55 nm) 之间,即蓝光波长(并不在上述范围内)附近。

剩下 2 个波长为第 301 号波长 (688.15 nm) 和第 302 号波长 (689.24 nm),验证了图 6 中水稻叶片样本的健康高光谱数据与染病数据对比分析,因为在第 275 号波长 (659.99 nm) 附近,健康水稻叶片的光谱反射率到达谷值,与同波段染病的水稻叶片光谱反射率相比,差异明显。

这 12 个波长均不在绿光波长 (550 nm) 附近。其他的 101 个波长也鲜有在绿光波长 (550 nm) 附近。

因为接种水稻细菌性条斑病病原菌的水稻叶片在发病前期,其细胞内部含有大量的叶绿素,在阳光的照射下呈墨绿色水渍斑点,与健康水稻叶片没有太大区别。在绿光波长 (550 nm) 附近,染病水稻叶片的光谱反射率与健康水稻叶片的光谱反射率也重合,充分说明了在绿光波长 (550 nm) 附近没有重要光谱的原因。

在蓝光波长 (450 nm) 附近,由于蓝光可以被水稻叶片中叶黄素吸收,转给叶绿素进行光合作用,促进水稻生长,导致大量水稻代谢产物堆积在蓝光附近,所以重要光谱基本出现在蓝光波段附近。

3 试验与结果分析

3.1 试验环境与评价指标

试验平台为 Windows10 系统,16 GB 内存,256 GB SSD,1TB HD, Intel QuadCore i7 - 8700, 4.2 GHz。采用 scikit-learn 2.2 和 Python 3.7 编写算法,高光谱数据采用 HSI Analyzer 和 Matlab 7.1 软件完成处理。

将预测正确的健康样本数量记为 T_p ,将预测错误的健康样本数量记为 T_N ,将预测正确的染病样本数量记为 F_p ,将预测错误的染病样本数量记为 F_N ^[28-29]。采用精确率 P 、召回率 R 、综合评价指标 $F1$ 值、准确率 A 作为评价指标。

3.2 波段选择结果

RF 算法筛选出的水稻叶片的高光谱波长中一共有 113 个波长对基于随机森林的水稻细菌性条斑病的分类模型有相关性影响。图 7 为波长重要程度相关性分析示意图。

由图 7 可知,排序第 1 的波长是 443.58 nm,其重要性分值为 0.046 40。重要程度排序第 1 的波长 (443.58 nm,第 70 号波长)到第 11 的波长 (432.33 nm,第 59 号波长)对 RF 算法的重要性依次缓慢递减。排序第 12 波长 (659.99 nm,第 275 号波

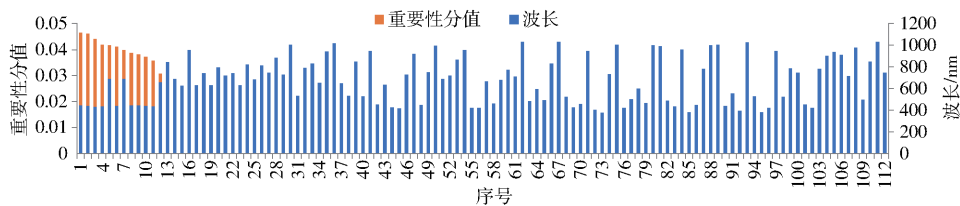


图 7 波长重要程度相关性分析示意图

Fig. 7 Correlation analysis schematic of band importance

长)到第 13 波长(845. 68 nm,第 445 号波长)对 RF 算法的重要性迅速下降。之后缓慢下降,此时的光谱波长的重要性已经很低,对 RF 算法分类水稻细菌性条斑病的影响不大。

从 113 个对本文模型有影响的波长中选取影响程度最高的前 12 个光谱波长,波长重要性排序见表 1。

表 1 波长重要性排序
Tab.1 Band importance ranking

波长序号	波长/nm	特征重要性分值
70	443. 58	0. 046 40
67	440. 51	0. 046 02
56	429. 27	0. 044 15
63	436. 41	0. 041 99
302	689. 24	0. 041 71
66	439. 48	0. 041 01
301	688. 15	0. 039 80
69	442. 55	0. 038 69
68	441. 53	0. 038 21
64	437. 44	0. 037 19
59	432. 33	0. 035 77
275	659. 99	0. 030 71

3.3 不同方法对比

试验选取逻辑回归(Logistic regression, LR)、朴素贝叶斯(Naive Bayes, NB)、决策树(Decision tree, DT)、k 最近邻(k-nearest neighbor, KNN)、支持向量分类机(Support vector classifier, SVC)和(Gradient boosting decision tree, GBDT)算法进行对比试验,相关算法参数均为 scikit-learn 2. 2 默认参数。

通过对比分析发现朴素贝叶斯准确率最低,只有 74. 27%(表 2)。因为朴素贝叶斯模型在给定输出类别的情况下,假设属性之间相互独立,这个假设在实际应用中往往不成立,当属性个数比较多或者属性之间相关性较大时,分类效果较差。相对逻辑

表 2 不同分类算法的准确率

算法	RF					
	LR	NB	DT	KNN	SVC	GBT (本文方法)
准确率/%	92. 92	74. 27	93. 30	91. 28	91. 58	93. 22

回归算法,RF 算法的准确率提高了 2. 32 个百分点,相对朴素贝叶斯算法,RF 算法的准确率提高了 20. 97 个百分点,相对决策树算法,RF 算法的准确率提高了 1. 94 个百分点,相对 KNN 算法,RF 算法的准确率提高了 3. 96 个百分点,相对 SVC,RF 算法的准确率提高了 3. 66 个百分点,相对 GBT,RF 算法的准确率提高了 2. 02 个百分点。

综上所述,基于随即森林的水稻细菌性条斑病识别方法的分类效果最优。

3.4 RF 算法结果分析

将 427 个测试水稻叶片样本数据输入训练好的水稻细菌性条斑病识别算法中,健康样本标签为 0,染病样本标签为 1。

本文方法对全波段的分类结果的精确率为 97. 63%,召回率为 95. 15%,F1 值为 96. 37%,准确率为 95. 17%。本文中的水稻叶片数据集不均衡,染病叶片远远高于健康叶片数量,F1 值较高,从侧面反映本文方法的精确性和稳定性较好。

使用随机森林算法筛选水稻叶片高光谱重要光谱波段,提高模型的预测性能,将用于分类的波段数量从 616 个减少到 12 个,数量下降了 98. 05%,大大简化了模型结构。

本文方法对 12 个波长分类的识别精确率为 94. 66%,召回率为 99. 55%,F1 值为 97. 04%,准确率为 94. 32%。与全波段分类结果相比,精确率减少了 2. 97 个百分点,召回率增加了 4. 4 个百分点,F1 值增加了 0. 67 个百分点,准确率减少了 0. 85 个百分点。虽然准确率有所下降,但是模型结构更加精简,计算复杂度下降,模型精度基本保持不变。

4 结论

(1)通过对水稻叶片全波段光谱曲线分析可得,水稻叶片光谱曲线总体呈上升趋势,患病叶片光谱曲线在 590 nm 附近,达到高峰;之后开始缓慢波动,在 650 nm 附近开始下降,在 680 nm 附近到达谷值,染病水稻叶片的光谱反射率与健康水稻叶片有明显差异;随后迅速上升,在 750 nm 附近趋于平缓。在蓝光波长(450 nm)和红光波长(664 nm)附近,染

病水稻叶片的光谱反射率与健康水稻叶片的光谱反射率差距较大。

(2) RF 算法的分类准确率最高,为 95.24%,相对 LR、NB、DT、KNN、SVC 和 GBT 算法,准确率分别提高了 2.32、20.97、1.94、3.96、3.66、2.02 个百分点。

(3) 采用 RF 算法对基于全波段和基于 12 个重要波长的分类结果进行比较。基于 12 个重要波长的识别精确率为 94.66%、召回率为 99.55%、F1 值

为 97.04%、准确率为 94.32%,与全波段分类结果相比,精确率减少了 2.97 个百分点,召回率增加了 4.4 个百分点,F1 值增加了 0.67 个百分点,准确率减少了 0.85 个百分点。虽然准确率有所下降,但是模型结构更加精简、计算复杂度下降,模型精度基本保持不变。

(4) 采用 MSC 对光谱曲线进行预处理,利用 RF 算法对细菌性条斑病高光谱影像识别的准确率均高于 94%,可以实现对细菌性条斑病的快速识别。

参 考 文 献

- [1] 唐惠燕,倪峰,李小涛,等. 基于 Scopus 的植物表型组学研究进展分析[J]. 南京农业大学学报, 2018, 41(6):169–177. TANG Huiyan, NI Feng, LI Xiaotao, et al. Analysis of the advance in plant phenomics research based on Scopus tools[J]. Journal of Nanjing Agricultural University, 2018, 41(6): 169–177. (in Chinese)
- [2] YANG W, GUO Z, HUANG C, et al. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice[J]. Nature Communications, 2014, 5:5087.
- [3] 汤帅,杨远航,潘素君,等. 水稻细菌性条斑病防治研究进展[J]. 农学学报, 2018, 8(11):16–20. TANG Shuai, YANG Yuanhang, PAN Sujun, et al. Research progress of rice bacterial leaf spot control[J]. Journal of Agronomy, 2018, 8(11): 16–20. (in Chinese)
- [4] GROBKINSKY D K, SVENSGAARD J, CHRISTENSEN S, et al. Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap[J]. Journal of Experimental Botany, 2015, 66(18):5429–5440.
- [5] NGUYEN H T, LEE B W. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression[J]. European Journal of Agronomy, 2006, 24(4):349–356.
- [6] MOSHOU D, BRAVO C, OBERTE R, et al. Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps[J]. Real-Time Imaging, 2005, 11(2): 75–83.
- [7] WANG D, DOWELL F E, LAN Y, et al. Determining pecky rice kernels using visible and near-infrared spectroscopy[J]. International Journal of Food Properties, 2002, 5(3): 629–639.
- [8] GUO L, YU Y, YU H, et al. Rapid quantitative analysis of adulterated rice with partial least squares regression using hyperspectral imaging system[J]. Journal of the Science of Food and Agriculture, 2019, 99(12):1–7.
- [9] WU N, JIANG H, BAO Y, et al. Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions[J]. Sensors and Actuators B-chemical, 2020, 308(5):1–35.
- [10] 雷雨,韩德俊,曾庆东,等. 基于高光谱成像技术的小麦条锈病病害程度分级方法[J/OL]. 农业机械学报, 2018, 49(5): 226–232. LEI Yu, HAN Dejun, ZENG Qingdong, et al. Classification method of wheat stripe rust disease degree based on hyperspectral imaging technology [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(5): 226–232. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20180526&flag=1. DOI:10.6041/j.issn.1000-1298.2018.05.026. (in Chinese)
- [11] LU J, YANG T, SU X, et al. Monitoring leaf potassium content using hyperspectral vegetation indices in rice leaves [J]. Precision Agriculture, 2020, 21(2):324–348.
- [12] SUN Dawei, CEN Haiyan, WENG Haiyong, et al. Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality[J]. Plant Methods, 2019, 15(1):1–16.
- [13] GHAMISI P, BENEDIKTSSON J A, PHINN S. Land-cover classification using both hyperspectral and LiDAR data[J]. International Journal of Image and Data Fusion, 2015, 6(3):189–215.
- [14] 张帅堂,王紫烟,邹修国,等. 基于高光谱图像和遗传优化神经网络的茶叶病斑识别[J]. 农业工程学报, 2017, 33(22): 200–207. ZHANG Shuitang, WANG Ziyang, ZOU Xiuguo, et al. Recognition of tea disease spot based on hyperspectral image and genetic optimization neural network [J]. Transactions of the CSAE, 2017, 33(22):200–207. (in Chinese)
- [15] YEH Y H H, CHUNG W C, LIAO J Y, et al. A comparison of machine learning methods on hyperspectral plant disease assessments[J]. IFAC Proceedings Volumes, 2013, 46(4):361–365.
- [16] 郭伟,朱耀辉,王慧芳,等. 基于无人机高光谱影像的冬小麦全蚀病监测模型研究[J/OL]. 农业机械学报, 2019, 50(9): 162–169. GUO Wei, ZHU Yaohui, WANG Huifang, et al. Monitoring model of winter wheat take-all based on UAV hyperspectral imaging [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(9):162–169. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20190918&flag=1. DOI:10.6041/j.issn.1000-1298.2019.09.018. (in Chinese)
- [17] 梁琨,杜莹莹,卢伟,等. 基于高光谱成像技术的小麦籽粒赤霉病识别[J/OL]. 农业机械学报, 2016, 47(2):309–315. LIANG Kun, DU Yingying, LU Wei, et al. Identification of fusarium head blight wheat based on hyperspectral imaging technology [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(2): 309–315. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20160241&flag=1. DOI:10.6041/j.issn.1000-1298.2016.02.041. (in Chinese)
- [18] 张智韬,劳聪聪,王海峰,等. 基于 FOD 和 SVMDA-RF 的土壤有机质含量高光谱预测[J/OL]. 农业机械学报, 2020, 51(1):156–167. ZHANG Zhitao, LAO Congcong, WANG Haifeng, et al. Estimation of desert soil organic matter through hyperspectra based on fractional-order derivatives and SVMDA-RF[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(1):156–167. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20200117&flag=1. DOI:10.6041/j.issn.1000-1298.2020.01.017. (in Chinese)

- [19] ZIVKOVIC Z, HEIJDEN F V D. Efficient adaptive density estimation per image pixel for the task of background subtraction [J]. Pattern Recognition Letters, 2006, 27(7): 773–780.
- [20] JU M, CHUNG Y, BAEK H, et al. Segmentation methods for a group-housed pig monitoring system [J]. Journal of Theoretical & Applied Information Technology, 2017, 95(17): 4321–4329.
- [21] 刘冬, 赵凯旋, 何东健. 基于混合高斯模型的移动奶牛目标实时检测[J/OL]. 农业机械学报, 2016, 47(5): 288–294. LIU Dong, ZHAO Kaixuan, HE Dongjian. Real-time target detection for moving cows based on Gaussian mixture model [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(5): 288–294. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20160539&flag=1. DOI: 10.6041/j.issn.1000-1298.2016.05.039. (in Chinese)
- [22] 何东健, 孟凡昌, 赵凯旋, 等. 基于视频分析的犊牛基本行为识别[J/OL]. 农业机械学报, 2016, 47(9): 294–300. HE Dongjian, MENG Fanchang, ZHAO Kaixuan, et al. Recognition of calf basic behaviors based on video analysis [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(9): 294–300. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160940&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2016.09.040. (in Chinese)
- [23] 王少华, 何东健, 刘冬. 基于机器视觉的奶牛发情行为自动识别方法[J/OL]. 农业机械学报, 2020, 51(4): 241–249. WANG Shaohua, HE Dongjian, LIU Dong. Automatic recognition method of dairy cow estrus behavior based on machine vision [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 241–249. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200428&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2020.04.028. (in Chinese)
- [24] 刘忠超, 何东健. 基于卷积神经网络的奶牛发情行为识别方法[J/OL]. 农业机械学报, 2019, 50(7): 186–193. LIU Zhongchao, HE Dongjian. Recognition method of cow estrus behavior based on convolutional neural network [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 186–193. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190719&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.07.019. (in Chinese)
- [25] OCZAK M, VIAZZI S, ISMAYILOVA G, et al. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network [J]. Biosystems Engineering, 2014, 119: 89–97.

(上接第 145 页)

- [19] 竞霞, 白宗璠, 高媛, 等. 利用随机森林法协同 SIF 和反射率光谱监测小麦条锈病[J]. 农业工程学报, 2019, 35(13): 154–161. JING Xia, BAI Zongyu, GAO Yuan, et al. Monitoring wheat stripe rust by using random forest method, SIF and reflectance spectrum [J]. Transactions of the CSAE, 2019, 35(13): 154–161. (in Chinese)
- [20] BISWAS S, SAUNSHI A, SARANGI S, et al. Random forest based classification of diseases in grapes from images captured in uncontrolled environments [C] // 2016 IEEE 13th International Conference on Signal Processing (ICSP). IEEE, 2016: 1775–1780.
- [21] 赵严, 罗登杰, 何圣贤, 等. 水稻细菌性条斑病 4 种接种方法的比较[J]. 亚热带农业研究, 2018, 14(4): 242–246. ZHAO Yan, LUO Dengjie, HE Shengxian, et al. Comparison of four inoculation methods for bacterial leaf spot of rice [J]. Subtropical Agricultural Research, 2018, 14(4): 242–246. (in Chinese)
- [22] XIA J, FALCO N, BENEDIKTSSON J A, et al. Hyperspectral image classification with rotation random forest via KPCA [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(4): 1601–1609.
- [23] CHAN C W, DESIRÉ P. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery [J]. Remote Sensing of Environment, 2008, 112(6): 2999–3011.
- [24] EDWIN R, BOGDAN Z. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images [J]. European Journal of Remote Sensing, 2017, 50(1): 144–154.
- [25] OK A O, AKAR O, GUNGOR O. Evaluation of random forest method for agricultural crop classification [J]. European Journal of Remote Sensing, 2012, 45(1): 421–432.
- [26] LABORTE A G, MAUNAHAN A A, HIJMANS R J. Opportunities for expanding paddy rice production in Laos: spatial predictive modeling using random forest [J]. Journal of Land Use Science, 2012, 7(1): 21–33.
- [27] HAM J, YANGCHI C, CRAWFORD M M, et al. Investigation of the random forest framework for classification of hyperspectral data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3): 492–501.
- [28] NADERI S, YIN T, KÖNIG S. Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups [J]. Journal of Dairy Science, 2016, 99(9): 7261–7273.
- [29] SANTOSO H, TANI H, WANG X. Random forest classification model of basal stem rot disease caused by *Ganoderma boninense* in oil palm plantations [J]. International Journal of Remote Sensing, 2017, 38(15–16): 4683–4699.
- [30] SAMAJPATI B J, DEGADWALA S D. Hybrid approach for apple fruit diseases detection and classification using random forest classifier [C] // 2016 International Conference on Communication and Signal Processing (ICCS). IEEE, 2016: 1015–1019.
- [31] STROBL C, BOULESTEIX A L, ZEILEIS A, et al. Bias in random forest variable importance measures: illustrations, sources and a solution [J]. BMC Bioinformatics, 2007, 8(1): 25–46.