

# 高光谱成像的柑橘病虫害叶片识别方法

吴叶兰<sup>1</sup>, 陈怡宇<sup>1</sup>, 廉小亲<sup>1</sup>, 廖禹<sup>2</sup>, 高超<sup>1</sup>, 管慧宁<sup>1</sup>, 于重重<sup>1</sup>

1. 北京工商大学, 中国轻工业工业互联网与大数据重点实验室, 北京 100048  
2. 江西省农业科学院农业工程研究所, 江西 南昌 330200

**摘要** 为监测柑橘生长状况, 实现病虫害无损识别, 利用高光谱成像技术和机器学习方法进行柑橘病叶分类研究。使用高光谱成像仪采集 46 片柑橘正常叶、46 片溃疡病叶、80 片除草剂危害叶、51 片红蜘蛛叶和 98 片煤烟病叶的高光谱图像, 在 478~900 nm 光谱范围内对每个叶片一个或多个发病区提取 5×5 的感兴趣区域(ROI), 将 ROI 内每个像素的反射率值作为光谱信息, 则一个 ROI 得到 25 个光谱信息样本, 最终五类叶片共得到 13250 个光谱样本。利用随机法将全部样本划分为 9 938 个训练集和 3 312 个测试集。分别采用一阶求导(1<sup>st</sup>Der)、多元散射校正(MSC)和标准正态变换(SNV)三种方法对原始光谱信息进行预处理, 对不同预处理方法后的数据采用主成分分析法(PCA)提取特征波长。1<sup>st</sup>Der 预处理后得到 7 个特征波长, 分别是 520.2, 689.0, 704.8, 715.4, 731.2, 741.8 和 757.6 nm; MSC 和 SNV 预处理后得到 7 个相同的特征波长, 分别是 551.9, 678.5, 704.8, 710.1, 725.9, 731.2 和 757.6 nm; 原始光谱得到 7 个特征波长, 分别是 525.5, 678.5, 710.1, 720.7, 725.9, 757.6 和 762.9 nm。分析 PCA 后的样本分布散点图可知, 正常叶片、溃疡病叶片和红蜘蛛叶片样本有一定程度聚类, 除草剂叶片和煤烟病叶片样本有大量重叠, 仅依据 PCA 不能完成病虫害叶片的识别。对全波段(FS)和 PCA 特征波长数据在不同预处理方法下进行支持向量机(SVM)和随机森林(RF)建模, 结果表明: 数据在 1<sup>st</sup>Der 预处理方法下识别效果最佳, 1<sup>st</sup>Der-FS-SVM 模型总分类精度(OA)为 95.98%, Kappa 系数为 0.948 2, 1<sup>st</sup>Der-FS-RF 模型 OA 为 91.42%, Kappa 系数为 0.889 2, 1<sup>st</sup>Der-PCA-SVM 模型 OA 为 90.82%, Kappa 系数为 0.881 6, 1<sup>st</sup>Der-PCA-RF 模型的 OA 为 91.79%, Kappa 系数为 0.894; 对 PCA 选择的特征波长数据建模, SVM 和 RF 模型下识别率均达到 84%, 全波段下模型识别率在 88% 以上, FS 数据建模效果优于 PCA 特征波长。研究结果表明, 高光谱成像技术结合机器学习方法进行柑橘叶片分类是可行且有效的, 为柑橘病虫害的无损准确识别提供理论根据。

**关键词** 高光谱成像; 主成分分析; 支持向量机; 随机森林

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)12-3837-07

## 引言

我国是柑橘生产大国<sup>[1]</sup>。受种植环境气候影响, 柑橘易感染病虫害, 如柑橘溃疡病原属于细菌, 感染后轻则果皮呈现疤痕、果熟后不耐贮藏易腐烂, 重则导致落果; 除草剂轻则导致柑橘叶片生长缓慢, 重则导致枯萎死亡; 柑橘红蜘蛛吸食叶片、果实汁液, 引起落叶、落果; 煤烟病由多种真菌引起, 发病初期, 表面出现暗褐色点状小霉斑, 后继续扩大成绒毛状黑色或灰黑色霉层。病虫害的发生会影响柑橘的产量及质量, 严重会造成经济损失。因此, 研究一种精准的柑橘病虫害检测方法在实际生产过程中有重要意义。传统的

柑橘病虫害检测方法主要有人工检测方法, 根据种植户或专业人员的经验知识来判断患病与否, 主观性强; 病理分析法, 对柑橘的生物化学性质分析检测, 周期长, 成本高。

高光谱成像技术是结合光谱和成像技术发展起来的, 可同时获取物体的空间和光谱信息<sup>[2]</sup>。随着高光谱遥感技术的迅速发展, 高光谱成像技术用于农作物病害诊断已成为重要的研究领域, 结合现代建模和数据分析可准确地识别农作物内部属性和外部特性, 为早期病害鉴定提供有利条件<sup>[3]</sup>。刘燕德等<sup>[4]</sup>结合 LS-SVM 和 PLS-DA 两种方法对柑橘正常、缺素和不同程度黄龙病叶片进行判别, 用二阶导数处理原始数据, 采用主成分分析和连续投影算法提取特征波长, PLS-DA 模型误判率为 5.6%。梅慧兰等<sup>[5]</sup>为实现柑橘黄龙病早

收稿日期: 2020-10-26, 修订日期: 2021-01-18

基金项目: 国家重点研发计划项目(2018YFC0807903)资助

作者简介: 吴叶兰, 女, 1970 年生, 北京工商大学人工智能学院副教授 e-mail: wuyel@th. btbu. edu. cn

期鉴别及病情分级,采用一阶微分和移动窗口拟合多项式平滑两种预处理方法,结合 PLS-DA 建模,验证集相关系数均达 0.954 8 以上。兰玉彬等<sup>[6]</sup>采用 SG 平滑、一阶导数和反对数变换对柑橘黄龙病植株光谱数据进行预处理,完成主成分波长和全波段的 KNN 和 SVM 建模,得到全波段下一阶导数光谱的 SVM 模型分类准确率达 94.7%,测试集误判率为 3.36%,实现了柑橘黄龙病监测。邓小玲等<sup>[7]</sup>为了实现柑橘黄龙病植株分类,对全波段使用 BP、XgBoost 算法,对特征波段使用逻辑回归和 SVM 算法,准确率分别达 95%,95%,93%和 96%。Abdulridha 等<sup>[8]</sup>在实验室内对无症状、早期和晚期感染溃疡的柑橘叶片和未成熟果进行高光谱成像,利用神经网络径向基函数,得到不同时期下染病叶片分类准确率为 94%,96%和 100%,健康和晚期的果实分类准确度为 92%,基于无人飞行器搭载同一成像系统对室外柑橘树检测,准确率达 100%。楚秉泉等<sup>[9]</sup>对真菌感染柑橘进行高光谱成像,二次主成分分析后得到 615 和 680 nm 两个特征波长,使用特征波长的第 2 主成分图像进行柑橘腐烂部位识别,识别率达 100%。上述研究表明高光谱技术在柑橘病虫害方面的研究主要为黄龙病和溃疡病的检测与分级,结合传统机器学习方法已能取得较好结果,但目前有关其他柑橘病虫害的研究仍较少。

本文针对柑橘溃疡病、除草剂、红蜘蛛、煤烟病和正常叶片,利用高光谱成像和机器学习技术进行分类识别,得到最佳病叶分类模型,为柑橘生长状况监测及病虫害识别提供理论基础。主要研究内容有:(1)采用一阶求导、多元散射校正和标准正态变换三种预处理方法对原始光谱数据进行预处理;(2)对不同预处理方法下数据进行主成分分析以提取特征波长;(3)对全波段数据和特征波长数据分别进行支持向量机和随机森林建模,分析实验结果。

## 1 实验部分

### 1.1 样本

实验样本来自江西省新余市渝水区柑橘果园,采摘时间为 2019 年 2 月底,包括柑橘正常叶片 46 片、柑橘溃疡病叶片 46 片、柑橘除草剂危害叶片 80 片、柑橘红蜘蛛叶片 51 片、柑橘煤烟病叶片 98 片。

### 1.2 高光谱图像采集

高光谱成像系统如图 1 所示,主要由高光谱成像仪、光源、计算机、样本、载物台和电源等构成。高光谱成像仪采用 SOC710VP,光谱范围为 350~1 050 nm,光谱分辨率为 1.3 nm,光谱波段数为 128,镜头类型为 C-Mount,焦距可调。光源选用 4 个 12 V 的卤素灯,亮度可调、角度可调。

本文设定高光谱图像采集时曝光时间为 150 ms,物距为 57.6 cm,扫描速度 30 行·s<sup>-1</sup>,32 s·cube<sup>-1</sup>,采集的原始图像大小为 696×520,获取的光谱图像原始数据是像元亮度值(digital number, DN),需将像元亮度值转化成光谱反射率供后续处理。图 2 为五类柑橘叶片高光谱图像,依次为正常叶片、溃疡病叶片、除草剂叶片、红蜘蛛叶片和煤烟病叶片。

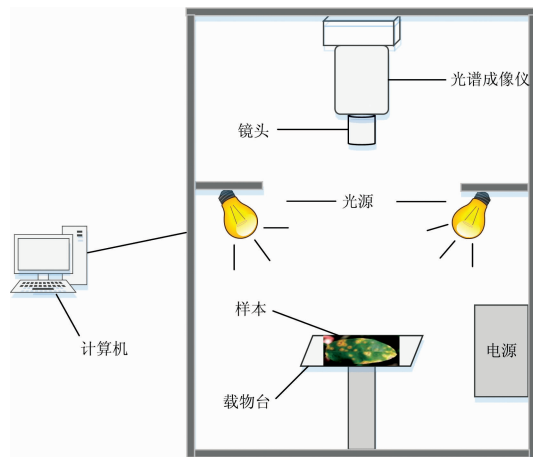


图 1 高光谱成像系统

Fig. 1 Hyperspectral imaging system



图 2 五类柑橘叶片高光谱图像

Fig. 2 Hyperspectral images of leaves of five citrus species

### 1.3 光谱数据提取及预处理

在 ENVI 中导入反射率.float 文件,去除 400 和 1 000 nm 附近受噪声影响较大的波段,选择 478~900 nm 间 81 个波段作为有效光谱范围,在每个叶片一个或多个发病区提取 5×5 的感兴趣区域(region of interest, ROI),将 ROI 内每个像素的反射率值作为光谱信息,则一个 ROI 得到的光谱信息为 25×81,五类柑橘叶片共提取 530 个 ROI,得到 13 250×81 的光谱信息矩阵。图 3 为所有样本的光谱曲线图,图 4 为五类叶片平均光谱曲线。由平均光谱图可知在 478~500 nm 的蓝光波段内光谱反射率低,形成一个波谷;在 550 nm 左右由于叶绿素反射作用呈现一个反射率峰;在红光波段内的 680 nm 附近由于叶绿素强吸收作用光谱曲线有一个波谷,700~900 nm 为高反射率区域。五类叶片光谱曲线呈现典型绿色叶片特征,而不同病虫害有不同特征,各类叶片的光谱曲线不完全一致,溃疡病叶片光谱曲线与其余四种在 680 nm 之前有明显差异;在全波段光谱(full-band spectral, FS)范围内正常叶片、除草剂叶片、红蜘蛛叶片和煤烟病叶片光谱曲线相似度高,仅反射率值存在差别。依据光谱曲线无法进行准确类别区分,还需后续建模处理。

受测量环境和高光谱仪器性能影响,获得的光谱信号易存在噪声和谐线漂移等干扰,需对其进行预处理以消除干扰<sup>[10]</sup>。利用 The Unscrambler X 10.4 软件对光谱数据分别进行一阶求导(first derivative, 1<sup>st</sup> Der)、多元散射校正(multiplicative scatter correction, MSC)和标准正态变换(standard normalized variate, SNV)预处理,对不同预处理方法下的光

谱数据建模以得到最优分类结果。

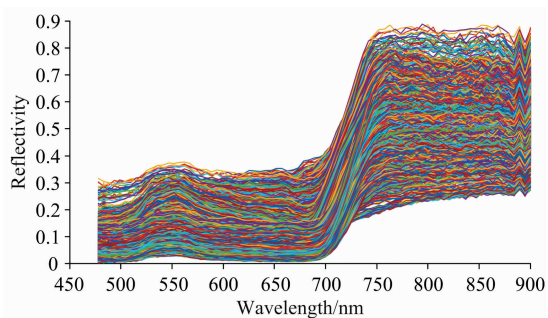


图 3 13 250 个样本光谱曲线

Fig. 3 Spectral curves of 13 250 samples

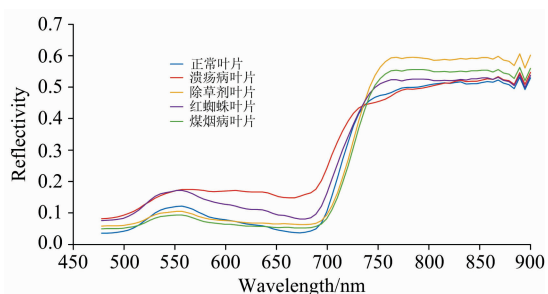


图 4 五类叶片平均光谱曲线

Fig. 4 Average spectral curves of five types of leaves

### 1.4 特征波长提取方法

高光谱数据在光谱维含有上百个连续波段，存在冗余信息，若利用这些全波段光谱数据进行实验，计算量大，影响准确率，因此可采用降维方法筛选出特征波长代入模型。主成分分析(principal component analysis, PCA)是一种正交线性变换技术，通过线性投影的方法，将目标分量根据其信息含量分布到新的坐标空间中<sup>[11]</sup>。采用 PCA 将高光谱图像信息进行压缩，去除波段间的冗余信息，实现降维。PCA 特征波长提取原理如下。

(1)光谱数据  $X$  为  $m \times n$  的矩阵， $m$  表示样本数， $n$  表示特征维度；

(2)计算  $X$  的协方差矩阵  $C(n \times n)$ ；

(3)求  $C$  的特征值和特征向量，并按从大到小的顺序排序，选择其中最大的  $k$  个，将其对应的  $k$  个特征向量分别作为列向量组成新的特征向量矩阵  $U(n \times k)$ ；

(4)将样本投影到选取的特征向量上，如式(1)

$$Y = XU \tag{1}$$

投影后的数据  $Y$  为  $m \times k$  的二维矩阵，则原始的  $n$  维特征变成了  $k$  维，达到了降维与特征波长提取目的。

### 1.5 建模方法

支持向量机(support vector machines, SVM)是核变换技术的代表算法之一，具有精度高、运算速度快、泛化能力强等优点<sup>[12]</sup>。SVM 依据结构风险最小理论，在解决小样本、非线性、高维模式等问题上具有优势，近年来广泛应用于高光谱数据处理。SVM 算法的原理是构造一个分类超平面，该

平面使得特征空间上正负样本的间隔最大化<sup>[13]</sup>；对于线性不可分情况，将低维空间问题映射到高维空间，引入核函数代替预测样本和支持向量内积，避免高维复杂运算。常用的核函数有线性核函数、RBF 核函数、多项式核函数和 Sigmoid 核函数。线性核函数是 RBF 核的一个特例，多项式核函数参数多会影响模型复杂程度，Sigmoid 核函数在某些参数上近似 RBF 核的功能，综上本文选择 RBF 核函数。

随机森林算法(random forest, RF)是决策树分类器的集合<sup>[14]</sup>，具有较高的分类准确性，不易出现过拟合现象<sup>[15]</sup>。基本思路为使用 Bootstrap 法从原始训练集中采样生成  $k$  个训练子集，构建  $k$  棵决策树，形成森林，通过每棵决策树的预测投票决定最终结果。“随机”指两方面：一是随机抽样训练子集；二是决策树构建时随机地从总的特征中选取远小于总特征个数的特征子集。随机性的引入使得每两棵树之间的差异度增加，从而使得模型的泛化能力提高，近年来随机森林算法已逐渐应用于高光谱分类。

目前基于支持向量机和随机森林算法进行柑橘病虫害分类识别的研究较少。本文选择上述两种算法建立五类柑橘叶片的分类模型，使用随机法将总样本划分为 9 938 个训练集和 3 312 个测试集，训练集与测试集之比为 3 : 1。采用测试集的总分类精度(overall accuracy, OA)和 Kappa 系数作为模型评价标准。

## 2 结果与讨论

### 2.1 特征波长的选取

主成分分析使原始光谱波段经过线性组合得到主成分，根据主成分的载荷(loadings)绝对值大小来确定特征波长。载荷是主成分与原始波长变量的相关系数，用于反映主成分和各波长变量间的密切程度。图 5 所示为原始光谱前 4 个主成分的载荷曲线，每条主成分曲线的波峰和波谷位置处载荷绝对值较大，对应的波长即为特征波长。由此原始光谱得到 7 个特征波长，分别是 525.5, 678.5, 710.1, 720.7, 725.9, 757.6 和 762.9 nm。运用相同的方法，1<sup>st</sup> Der 预处理光谱得到 7 个特征波长，分别是 520.2, 689.0, 704.8, 715.4, 731.2, 741.8 和 757.6 nm；MSC 预处理光谱和 SNV 预处理光谱得到 7 个一致的特征波长，分别是 551.9, 678.5, 704.8, 710.1, 725.9, 731.2 和 757.6 nm。由于病虫害影响，

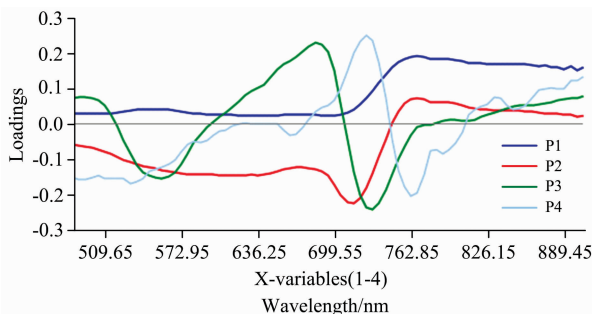


图 5 原始光谱前 4 个主成分载荷曲线

Fig. 5 The first four principal component load curves of the original spectrum

叶片表面颜色和叶绿素含量发生改变,红光波段高反射率区域受影响较大,故特征波长多落于此区间,绿光波段内受叶绿素影响原反射率峰附近也出现特征波长,所以特征波长集中在 520~760 nm 范围内。

通过主成分分析还能得到光谱样本的散点图,用来分析样本的分布情况。图 6 显示了不同预处理下光谱数据样本的散点图,编号 1—5 分别表示柑橘正常、溃疡病、除草剂、红

蜘蛛和煤烟病叶片。由图 6 可知,PC-1 和 PC-2 累计贡献率达 86% 以上,解释了原始数据的绝大部分信息,五类叶片呈现出一定聚类效果。经 MSC 和 SNV 预处理后,正常叶片、溃疡病叶片和红蜘蛛叶片有较为清晰的分类界限,除草剂和煤烟病叶片样本重叠较多。经 1<sup>st</sup> Der 预处理后聚类效果差,五类叶片样本均有大量重叠。因此还需对 PCA 后光谱数据作进一步处理以实现重叠样本的分类。

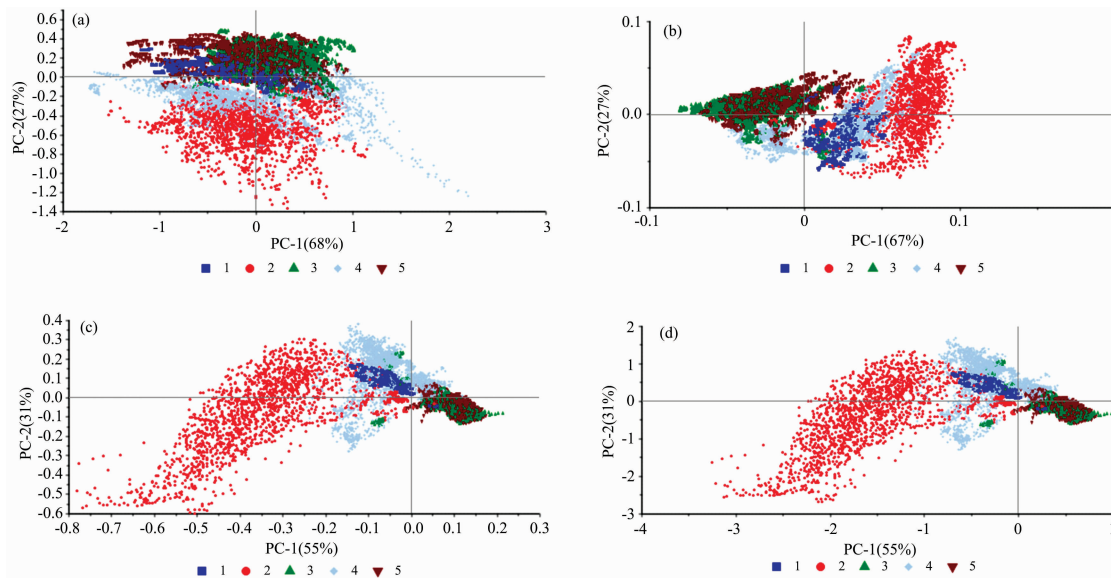


图 6 样本分布散点图

(a), (b), (c) 和 (d) 分别表示原始及经 1<sup>st</sup> Der, MSC 和 SNV 预处理后光谱得到的样本分布散点图

Fig. 6 Sample distribution scatter diagram

Figure (a), (b), (c) and (d) respectively represent the scatter plots of sample distribution using original spectra and the spectra after preprocessing by 1<sup>st</sup> Der, MSC and SNV

## 2.2 全波段数据建模

对全波段数据(FS)进行 SVM 和 RF 建模,运用网格搜索法和交叉验证法确定不同预处理后数据在 SVM 模型的最优参数。 $C$  是惩罚系数,表示对误差的宽容度,值越大,表示越不能容忍出现误差,易过拟合。 $\gamma$  是 RBF 核自带的参数,决定了数据映射到新的特征空间后的分布情况,值越大,支持向量个数越少。最终得到原始光谱数据下, $C$  为 16,  $\gamma$  为 128; 1<sup>st</sup> Der 数据下, $C$  为 8,  $\gamma$  为 8 192; MSC 数据下, $C$  为 8,  $\gamma$  为 512; SNV 数据下, $C$  为 4,  $\gamma$  为 16。RF 建模时,选择决策树颗数为 500。表 1 为不同预处理方法下全波段数据测试集的分类结果,由表 1 可知,对柑橘正常叶片、溃疡病叶片和红蜘蛛叶片,模型识别率分别可达 98%, 97% 和 96%。除草剂叶片识别率在原始数据和经 1<sup>st</sup> Der 预处理数据的 SVM 模型下可达 90%, 1<sup>st</sup> Der-FS-SVM 模型识别率为 91.05%。在进行 RF 建模后识别率均较低,最高为原始数据下 84.91%。对煤烟病叶片,均在经 1<sup>st</sup> Der 预处理数据的模型下有最高识别结果,1<sup>st</sup> Der-FS-SVM 模型识别率为 94.09%, 1<sup>st</sup> Der-FS-RF 模型识别率为 93.29%, 但与正常、溃疡病和红蜘蛛叶片相比,识别率不够稳定。除草剂叶片在五类叶片中识别结果最差,煤烟病叶片次之,其他三种叶片识别率均较稳定维持在 96% 之上。

表 1 全波段数据建模结果

Table 1 Modeling results of full-band data

模型	预处理方法	识别率/%				
		正常叶片	溃疡病叶片	除草剂叶片	红蜘蛛叶片	煤烟病叶片
SVM	原始	100.00	99.78	90.03	100.00	92.08
	1 <sup>st</sup> Der	100.00	99.78	91.05	99.58	94.09
	MSC	99.73	99.78	80.69	97.35	88.38
	SNV	100.00	99.78	79.80	98.60	87.17
RF	原始	99.46	98.44	84.91	98.74	93.19
	1 <sup>st</sup> Der	99.46	99.33	74.04	98.74	93.29
	MSC	98.91	97.10	72.76	96.23	85.97
	SNV	98.91	97.10	73.40	96.37	87.17

## 2.3 PCA 特征波长数据建模

对 PCA 特征波长数据进行 SVM 和 RF 建模识别,利用网格搜索法和交叉验证法得到不同预处理后数据在 SVM 模型的最优参数:原始数据下, $C$  为 128,  $\gamma$  为 512; 1<sup>st</sup> Der 数据下, $C$  为 4 096,  $\gamma$  为 8 192; MSC 数据下, $C$  为 4,  $\gamma$  为 4 096; SNV 数据下, $C$  为 4,  $\gamma$  为 256。RF 建模时,决策树棵数设为 500。表 2 为 PCA 特征波长数据测试集的分类结果,分析可知,对柑橘正常叶片、溃疡病

叶片和红蜘蛛叶片, 模型识别率分别可达 97%, 96% 和 93%。除草剂叶片在经 1<sup>st</sup> Der 预处理数据的模型下有最高识别结果, 1<sup>st</sup> Der-PCA-SVM 和 1<sup>st</sup> Der-PCA-RF 模型识别率分别为 79.28% 和 80.05%, 识别效果不佳, 与图 6 中除草剂叶片样本大量重叠于煤烟病叶片样本现象吻合。与除草剂叶片相同, 煤烟病叶片也在经 1<sup>st</sup> Der 预处理数据的模型下有最高识别结果, 1<sup>st</sup> Der-PCA-SVM 模型识别率为 87.27%, 1<sup>st</sup> Der-PCA-RF 模型识别率为 91.58%。综合五类叶片, 除草剂叶片识别结果最差, 煤烟病叶片次之, 其他三种叶片识别率均较稳定维持在 93% 之上。

全波段数据(FS)和 PCA 波长数据在不同预处理方法下的模型识别效果如表 3 所示。全波段光谱数据下 SVM 模型的总分类精度(OA)均高于 91%, 1<sup>st</sup> Der-FS-SVM 模型总分类精度(OA)最高, 为 95.98%, Kappa 系数为 0.948 2; RF 模型 OA 均高于 88%, 原始数据下 FS-RF 模型 OA 最高, 为 93.84%, Kappa 系数为 0.9205。对 PCA 特征波长, SVM 模型的 OA 均高于 86%, RF 模型的 OA 均高于 84%。经 1<sup>st</sup> Der 预处理数据的建模效果最佳, 1<sup>st</sup> Der-PCA-SVM 模型

OA 为 90.82%, Kappa 系数为 0.8816; 1<sup>st</sup> Der-PCA-RF 模型 OA 为 91.79%, Kappa 系数为 0.894。同时由表 3 可知各模型 Kappa 系数均维持在 0.8 以上, 表明其一致性良好。

表 2 PCA 特征波长数据建模结果

Table 2 PCA characteristic wavelength modeling results

模型	预处理方法	识别率/%				
		正常叶片	溃疡病叶片	除草剂叶片	红蜘蛛叶片	煤烟病叶片
SVM	原始	100.00	99.11	72.51	98.46	81.66
	1 <sup>st</sup> Der	99.18	99.78	79.28	98.46	87.27
	MSC	98.10	98.44	72.76	94.97	81.56
	SNV	98.10	98.66	72.51	94.83	81.16
RF	原始	98.64	96.65	74.30	98.46	85.97
	1 <sup>st</sup> Der	99.18	96.65	80.05	98.04	91.58
	MSC	97.55	97.32	68.80	93.85	80.96
	SNV	97.28	97.32	68.41	93.72	81.06

表 3 FS 和 PCA 波长数据不同预处理方法下建模结果

Table 3 Modeling results under different pretreatment methods of all-band and PCA wavelength data

模型	评价指标	FS				PCA			
		原始	1 <sup>st</sup> Der	MSC	SNV	原始	1 <sup>st</sup> Der	MSC	SNV
SVM	OA	95.23%	95.98%	91.30%	91.03%	87.53%	90.82%	86.50%	86.32%
	Kappa	0.938 5	0.948 2	0.887 9	0.884 4	0.839 2	0.881 6	0.826 0	0.823 6
RF	OA	93.84%	91.42%	88.01%	88.56%	88.77%	91.79%	84.93%	84.81%
	Kappa	0.920 5	0.889 2	0.845 3	0.852 3	0.855 0	0.894 0	0.805 6	0.804 0

综合来看, SVM 模型下识别效果优于 RF, FS 数据建模效果优于特征波长。分析 PCA 特征波长下建模结果较全波段数据差的原因: PCA 提取的特征波长大多集中在 700~770 nm, 波长范围窄, 忽略了对分类识别有效的其他波长; 提取的特征波长数量较少, 包含信息量不全面, 不能很好地解释全部数据。后续可采用其他方法提取特征波长进行分类识别。

### 3 结 论

为了识别柑橘正常、溃疡病、除草剂、红蜘蛛和煤烟病五类病虫害叶片, 利用 SOC710VP 高光谱成像仪采集五类柑橘叶片高光谱图像, 在 478~900 nm 范围内提取 ROI 得到光谱数据。采用 1<sup>st</sup> Der, MSC 和 SNV 预处理方法对原始光谱

数据进行预处理, 对不同预处理方法后数据进行 PCA 以提取特征波长。分别以 FS 和 PCA 特征波长光谱数据进行 SVM 和 RF 建模, 结果表明: (1) 一阶求导预处理方法能够优化光谱数据, 经其预处理后的数据在各模型下总分类精度达 90%, Kappa 系数达 0.88; (2) 对不同预处理后光谱数据进行 PCA, 均得到 7 个特征波长, 分别建立 SVM 和 RF 模型, 1<sup>st</sup> Der-PCA-SVM 和 1<sup>st</sup> Der-PCA-RF 模型下 OA 最高, 分别为 90.82% 和 91.79%; (3) 综合来看 FS 数据建模效果优于 PCA 特征波长, 前者模型识别率在 88% 以上, 后者识别率达到 84%, 分析可知特征波长相对集中在 700~770 nm, 范围跨度窄, 包含信息不全面, 故 PCA 模型的 OA 值较 FS 数据低。研究结果表明结合高光谱成像技术和机器学习方法可以实现柑橘病虫害叶片的分类, 为柑橘病虫害的无损准确识别提供了理论根据。

### References

[ 1 ] YUE Xue-jun, LING kang-jie, WANG Lin-hui, et al(岳学军, 凌康杰, 王林惠, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2019, 50(3): 186.

[ 2 ] Gu Q, Sheng L, Zhang T H, et al. Computers and Electronics in Agriculture, 2019, 167: 105066.

[ 3 ] Wang K J, Guo D M, Zhang Y, et al. Frontiers of Agricultural Science and Engineering, 2019, 6(2): 172.

[ 4 ] LIU Yan-de, XIAO Huai-chun, SUN Xu-dong, et al(刘燕德, 肖怀春, 孙旭东, 等). Transactions of the Chinese Society for Agricultural

- Machinery(农业机械学报), 2016, 47(11): 231.
- [ 5 ] MEI Hui-lan, DENG Xiao-ling, HONG Tian-sheng, et al(梅慧兰, 邓小玲, 洪添胜, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2014, 30(9): 140.
- [ 6 ] LAN Yu-bin, ZHU Zi-hao, DENG Xiao-ling, et al(兰玉彬, 朱梓豪, 邓小玲, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2019, 35(3): 92.
- [ 7 ] DENG Xiao-ling, ZENG Guo-liang, ZHU Zi-hao, et al(邓小玲, 曾国亮, 朱梓豪, 等). Journal of South China Agricultural University(华南农业大学学报), 2020, 41(6): 100.
- [ 8 ] Abdulridha J, Batuman O, Ampatzidis Y. Remote Sensing, 2019, 11(11): 1373.
- [ 9 ] CHU Bing-quan, ZHANG Hai-liang, LUO Wei, et al(楚秉泉, 章海亮, 罗 微, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(8): 2551.
- [10] HUO Ying-qiu, ZHANG Chen, LI Yu-hao, et al(霍迎秋, 张 晨, 李宇豪, 等). Chinese Journal of Agricultural Mechanization(中国农机化学报), 2019, 40(4): 71.
- [11] Tian X, Fan S X, Huang W Q, et al. Postharvest Biology and Technology, 2020, 161: 111071.
- [12] DU Pei-jun, XIA Jun-shi, XUE Chao-hui, et al(杜培军, 夏俊士, 薛朝辉, 等). Journal of Remote Sensing (遥感学报), 2016, 20(2): 236.
- [13] Guo A T, Huang W J, Ye H C, et al. Remote Sensing, 2020, 12(9): 1419.
- [14] Lauwers M, Cauwer D B, Nuytens D, et al. Sensors, 2020, 20(9): 2504.
- [15] LI Lei, REN Yue-mei(李 垒, 任越美). Computer Engineering and Applications(计算机工程与应用), 2016, 52(24): 189.

## Study on the Identification Method of Citrus Leaves Based on Hyperspectral Imaging Technique

WU Ye-lan<sup>1</sup>, CHEN Yi-yu<sup>1</sup>, LIAN Xiao-qin<sup>1</sup>, LIAO Yu<sup>2</sup>, GAO Chao<sup>1</sup>, GUAN Hui-ning<sup>1</sup>, YU Chong-chong<sup>1</sup>

1. Key Laboratory of Internet and Big Data in Light Industry, Beijing Technology and Business University, Beijing 100048, China
2. Institute of Agricultural Engineering, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China

**Abstract** To monitor citrus growth and realize nondestructive identification of pests and diseases, the leaf classification of citrus diseases was studied using hyperspectral imaging technology and machine learning method. Using hyperspectral imager to collect hyperspectral images of 46 normal citrus leaves, 46 canker leaves, 80 herbicide-damaged leaves, 51 red spider diseased leaves, and 98 soot diseased leaves. A  $5 \times 5$  regions of interest (ROI) were extracted from one or more diseased areas of each leaf in the 478~900 nm spectral range. Taking the reflectance value of each pixel in the ROI as the spectral information, one ROI would get 25 spectral information samples, and finally the five types of leaves get a total of 13 250 spectral samples. The samples were divided into 9938 training sets and 3 312 test sets by random method. The first derivative (1<sup>st</sup> Der), multiple scattering correction (MSC) and standard normal transformation (SNV) were used to preprocess the original spectral information, and principal component analysis (PCA) was used to extract the characteristic wavelength of the data after different preprocessing methods. After 1<sup>st</sup> Der pretreatment, 7 characteristic wavelengths were obtained, which were 520.2, 689, 704.83, 715.38, 731.2, 741.75 and 757.58nm respectively. After MSC and SNV pretreatment, 7 identical characteristic wavelengths were obtained, which were 551.85, 678.45, 704.83, 710.1, 725.93, 731.2 and 757.58 nm, respectively. The original spectrum obtained seven characteristic wavelengths, which were 525.48, 678.45, 710.1, 720.65, 725.93, 757.58 and 762.85 nm, respectively. The scatter plot of sample distribution after PCA analysis showed that there was a certain degree of clustering of normal leaves, canker leaves and starscream leaves, and a large amount of overlap between herbicide leaves and soot leaves, so the identification of pest and disease leaves could not be completed only based on PCA. Support vector machine (SVM) and random forest (RF) were used to model the all-band spectrum (FS) and PCA characteristic wavelength data under different pretreatment methods, and the results showed that: The OA of 1<sup>st</sup> Der-FS-SVM model was 95.98%, the Kappa coefficient was 0.948 2, the OA of 1<sup>st</sup> Der-FS-RF model was 91.42%, the Kappa coefficient was 0.889 2, the OA of 1<sup>st</sup> Der-FS-SVM model was 90.82%, and the Kappa coefficient was 0.881 6, OA and Kappa coefficient in 1<sup>st</sup> Der-PA-RF model was 91.79% and 0.894 respectively. For PCA characteristic wavelength data modeling, the recognition rate of SVM and RF models reached 84%, and the recognition rate of the full-band spectrum model was above 88%. The FS data modeling effect was better than that of PCA characteristic wavelength. The results show that it is feasible and effective to classify citrus leaves by hyperspectral imaging

technique combined with machine learning method, which provides a theoretical basis for the accurate and nondestructive identification of citrus pests and diseases.

**Keywords** Hyperspectral imaging; Principal component analysis; Support vector machines; Random forest

(Received Oct. 26, 2020; accepted Jan. 18, 2021)

欢迎投稿

欢迎订阅

欢迎刊登广告

## 《光谱学与光谱分析》2022年征订启事

国内邮发代码: 82-68

国外发行代码: M905

《光谱学与光谱分析》1981年创刊,国内统一刊号:CN 11-2200/O4,国际标准刊号:ISSN 1000-0593, CODEN 码:GYGFED,国内外公开发行人,大16开本,332页,月刊;是中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的学术性刊物。北京大学出版社出版,每期售价95元,全年1140元。刊登主要内容:激光光谱测量、红外、拉曼、紫外、可见光谱、发射光谱、吸收光谱、X射线荧光光谱、激光显微光谱、光谱化学分析、国内外光谱化学分析领域内的最新研究成果、开创性研究论文、学科发展前沿和最新进展、综合评述、研究简报、问题讨论、书刊评述。

《光谱学与光谱分析》适用于冶金、地质、机械、环境保护、国防、天文、医药、农林、化学化工、商检等各领域的科学研究单位、高等院校、制造厂家、从事光谱学与光谱分析的研究人员、高校有关专业的师生、管理干部。

《光谱学与光谱分析》为我国首批自然科学核心期刊,中国科协优秀科技期刊,中国科协择优支持基础性、高科技学术期刊,中国科技论文统计源刊,“中国科学引文数据库”,“中国物理文摘”,“中国学术期刊文摘”,同时被国内外的CJCR, CNKI, CSCD, SCI, AA, CA, Ei, AJ, PJK, MEDLINE, Scopus等文献机构收录。根据中国科学技术信息研究所发布信息,中国科技期刊物理类影响因子、引文量及综合评价总分《光谱学与光谱分析》都居前几位。欢迎国内外厂商在《光谱学与光谱分析》发布广告(广告经营许可证:京海市监广登字20170260号)。

《光谱学与光谱分析》的主编为高松院士。

欢迎新老客户到全国各地邮局订阅,若有漏订者可直接与《光谱学与光谱分析》期刊社联系。

联系地址:北京市海淀区学院南路76号(南院),

《光谱学与光谱分析》期刊社

邮政编码:100081

联系电话:010-62181070, 62182998

电子信箱:chnghxygpfx@vip.sina.com

修改稿专用邮箱:gp2008@vip.sina.com

网 址: <http://www.gpxygpfx.com>

