

基于高光谱成像的甘蔗叶片早期轮斑病与锈病识别技术

黄亦其¹ 刘祥焕^{1,2} 黄震宇^{1,2} 钱万强² 刘双印³ 乔曦^{1,2}

(1. 广西大学机械工程学院, 南宁 530004; 2. 中国农业科学院(深圳)农业基因组研究所, 深圳 518120;
3. 仲恺农业工程学院信息科学与技术学院, 广州 510225)

摘要: 针对甘蔗叶片早期轮斑病与锈病发病症状相似, 难以区分, 导致在实际生产中不便对症施药的问题, 以甘蔗早期轮斑病和锈病叶片为研究对象, 探究利用高光谱成像技术来识别甘蔗叶片早期轮斑病与锈病的可行性。首先, 利用高光谱成像系统在 406 ~ 1 014 nm 光谱范围内采集甘蔗健康叶片、早期轮斑病叶片和锈病叶片的高光谱图像, 提取图像的感兴趣区域 (Region of interest, ROI) 并计算其平均光谱作为原始光谱数据, 采用一阶导数 (First derivative, FD)、Savitzky - Golay 卷积平滑 (Savitzky - Golay convolutional smoothing, SG) 和标准正态变换 (Standard normal variate, SNV) 分别对原始光谱数据进行预处理。然后, 在预处理的基础上采用主成分分析 (Principal component analysis, PCA) 算法、蚁群优化 (Ant colony optimization, ACO) 算法进行特征降维, 并将降维后的特征作为后期建模的输入变量。最后, 结合降维和不降维 2 种方式使用支持向量机 (SVM) 和随机森林 (RF) 进行识别。为了确定最优的识别模型, 对不同的预处理方法、降维方法和分类器共 18 个组合模型进行了试验。经对比发现, SG - SVM 识别模型效果最佳, 测试集准确率为 99.65%。试验结果表明, 利用高光谱成像技术进行甘蔗叶片早期轮斑病和锈病的识别可行且有效, 可为植保无人机超低空遥感病害监测提供参考。

关键词: 甘蔗叶片; 病害识别; 高光谱成像; 数据预处理; 光谱降维

中图分类号: S435.661

文献标识码: A

文章编号: 1000-1298(2023)04-0259-09

OSID:



Identification of Early Wheel Spot and Rust on Sugarcane Leaves
Based on Spectral Analysis

HUANG Yiqi¹ LIU Xianghuan^{1,2} HUANG Zhenyu^{1,2} QIAN Wanqiang² LIU Shuangyin³ QIAO Xi^{1,2}

(1. College of Mechanical Engineering, Guangxi University, Nanning 530004, China

2. Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

3. College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China)

Abstract: Aiming at the problem that the symptoms of early wheelspot disease and rust disease on sugarcane leaves are similar and difficult to distinguish, which leads to the inconvenience of prescribing the right medicine to the disease in actual production. The feasibility of using hyperspectral imaging technology to identify early wheel spot disease and rust disease on sugarcane leaves was explored. Firstly, hyperspectral images of healthy sugarcane leaves, early wheel spot leaves and rust leaves were collected by hyperspectral imaging system in the spectral range of 406 ~ 1 014 nm. The average spectral reflectance of region of interest (ROI) was extracted and its average spectrum was calculated as the raw spectral data. The first derivative (FD), Savitzky - Golay convolution smoothing (SG) and standard normal variate (SNV) were used to preprocess the original spectral data. Then on the basis of preprocessing, principal component analysis (PCA) and ant colony optimization (ACO) were used to reduce the feature dimension, and the feature after dimensionality reduction were used as the input variables in the later modeling. Finally, the support vector machine (SVM) and random forest (RF) were used for recognition by combining dimensionality reduction and non-dimensionality reduction. In order to determine the

收稿日期: 2022-07-12 修回日期: 2022-08-23

基金项目: 国家重点研发计划项目(2021YFD1400100、2021YFD1400101)、广西自然科学基金项目(2021JJA130221)和国家自然科学基金项目(61871475)

作者简介: 黄亦其(1973—),男,副教授,博士,主要从事农业信息获取与计算机视觉研究,E-mail: hyqgxu@163.com

通信作者: 乔曦(1987—),男,副研究员,博士,主要从事农业信息自动获取与处理研究,E-mail: qiaoxi@caas.cn

optimal recognition model, totally 18 combined models with different preprocessing methods, dimensionality reduction methods and classifiers were tested. By comparison, it was found that the SG – SVM recognition model had the best effect, and the accuracy of the test set was 99.65%. It was feasible and effective to use hyperspectral imaging technology to identify early wheel spot and rust on sugarcane leaves, which can provide reference for ultra-low altitude remote sensing disease monitoring of plant protection UAV.

Key words: sugarcane leaves; disease recognition; hyperspectral imaging; data preprocessing; spectral dimension reduction

0 引言

甘蔗是主要的糖料作物之一,具有重要的经济价值,其糖产量占全球 80% 左右^[1]。在甘蔗生长过程中,各种病害不仅威胁着甘蔗的发育,还会影响甘蔗的品质和榨糖量,最典型的病害有轮斑病、锈病和凤梨病等^[2-3]。甘蔗早期轮斑病和锈病发病症状相似,仅凭肉眼很难区分,传统的图像处理技术难以达到识别要求^[4]。相比之下,高光谱成像技术能够提供丰富的光谱信息,具有识别相似病害的能力^[5]。

国内外基于高光谱病害识别的研究,侧重于从光谱图像中提取目标的全光谱信息,并通过一些机器学习算法进行识别,如支持向量机^[6] (SVM) 和随机森林(RF)^[7]。全光谱数据中含有大量的冗余信息,在干扰识别的同时还会降低模型的运行速率。因此,有必要从全波段中选择可分性好的波段子集,降低数据维度^[8]。此外,在野外采集高光谱图像数据时,会受到系统噪声和外界噪声的干扰^[9]。为了得到准确的甘蔗叶片病害光谱数据并能有效地提高光谱信噪比,合理地选择高光谱预处理方法、降维方法至关重要^[10]。

目前,常用的预处理方法有一阶导数(FD)、Savitzky – Golay 卷积平滑(SG)和标准正态变换(SNV)等^[11-12]。光谱预处理可以消除光谱噪声及散射效应,前人研究已得到充分的证实并进行广泛应用^[13-14]。常见的特征降维方法有主成分分析(PCA)算法、连续投影算法、蚁群优化(ACO)算法等^[15-16]。光谱数据降低维度不仅可以缩短模型的运行时间而且在一定程度上可以提高识别效率^[17]。袁培森等^[18]采用多元散射校正对水稻细菌性条斑病光谱进行预处理,利用 RF 分类器进行识别的最高准确率为 95.24%。吴叶兰等^[19]在不同预处理方法的基础上采用 PCA 算法对柑橘病害叶片的识别率均达 84% 以上。王小龙等^[20]采用 PCA 结合 SVM 构建棉花虫害叶片的识别模型,测试集识别精度为 92.16%。这些研究表明,高光谱可用于作物病害检测,并且取得了较好的效果。但目前有关利用高光谱技术对甘蔗病害种类的识别研究鲜有报道。

本文以甘蔗叶片早期轮斑病和锈病为研究对象,以甘蔗健康叶片、早期轮斑病叶片与锈病叶片的高光谱图像为研究样本,采用高光谱预处理算法(FD、SG 和 SNV)和降维算法(PCA 和 ACO)对早期轮斑病、锈病和健康甘蔗叶片的高光谱数据进行处理,在最终的分类中,使用机器学习模型(SVM 和 RF)进行分类。将预处理算法、降维算法和分类器随机组合,探索野外甘蔗叶片早期轮斑病和锈病的最优识别模型,以期为甘蔗的病害识别提供参考依据与技术支持。

1 材料与方法

1.1 样本采集

利用四川双利合谱科技有限公司生产的 GaiaField Pro – V10 型高光谱成像光谱仪,在野外人工采集甘蔗早期轮斑病叶片、甘蔗锈病叶片以及健康甘蔗叶片 3 种类别的高光谱图像。该光谱仪波长范围为 406 ~ 1 014 nm,光谱分辨率为 3.5 ~ 4.0 nm,它可以一次性在 176 个光谱波段上进行图像拍摄,也可以根据图像上每 1 个像素点从拍摄的光谱数据文件内提取出 1 条包含 176 个光谱反射强度的光谱曲线。图像采集地点为广西壮族自治区崇左市扶绥县渠黎镇广西大学亚热带农科新城,图像拍摄于 2021 年 12 月 5—11 日,天气晴朗。图 1 为现场采样图及部分图像样例。

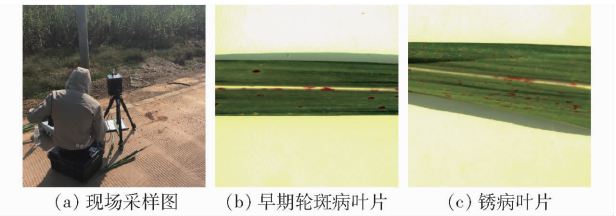


图 1 病害叶片现场采样图及部分图像样例

Fig.1 Field sampling map and partial examples of diseased leaves

拍摄时将甘蔗叶片和标准白板同时放置在平整地面上,高光谱成像光谱仪固定在叶片正上方 75 cm 处。首先调整好镜头焦距和曝光时间,使光谱仪成像清晰且甘蔗叶片的反射强度位于光谱仪感光区间内,然后使用与光谱仪配套的软件 SpecView

控制光谱仪拍摄各组样本并按顺序保存文件。由于野外光照强度变化大,样本图像采集完成后需要采集相应的白板和黑板数据。

图像采集完成后,得到的是反射强度数据文件,不同样本之间的反射强度易受到光照强度的影响而产生差异。要准确地反映被拍摄样本的光谱特性,需要将反射强度文件校正为反射率文件。高光谱图像反射校准公式为

$$R_{ci} = \frac{Sample_{ci} - dark_{ci}}{White_{ci} - dark_{ci}} \tag{1}$$

式中 R_{ci} ——校准图像反射率
 $Sample_{ci}$ ——原始图像反射强度
 $White_{ci}$ ——标准白板参考反射强度
 $dark_{ci}$ ——黑板参考反射强度

本文共采集 50 幅高光谱图像,利用 ENVI 5.3 软件以感兴趣区域 (ROI) 尺寸 10 像素 × 10 像素从样本图像中提取特征点,以每 1 个 ROI 的平均光谱作为 1 个样本,共采集 938 个数据样本,并以比例 7:3 划分为训练集和测试集。

1.2 试验方法

1.2.1 预处理

一阶导数 (FD) 可以显示出光谱微小的变化,提高光谱分辨率,有效地突出光谱波峰和波谷特征信息,经过求导的光谱数据还可以消除光谱基线平移和极限偏移带来的不利影响^[21-22]。对于分辨率高、波长点多的光谱数据,一般采用直接差分法求导,求导时对每条样本光谱曲线进行数据采样,FD 以 2 个采样点的差值除以采样点的间距获得,代表该点的光谱曲线斜率。FD 的差分求导公式为

$$X_i = \frac{x_{i+g} - x_i}{g} \tag{2}$$

式中 X_i ——FD 预处理后的反射率
 x_i ——采样点 i 反射率
 g ——采样点间距,取 9

SG 卷积平滑是一种基于局域多项式最小二乘拟合的平滑滤波方法,又称多项式平滑。它的基本思想是以波长 a 处为中心点,向左右各取 ω 个数据点,组成 1 个 $2\omega + 1$ 的窗口,然后用给定次数的多项式对窗口内的数据进行最小二乘拟合,拟合多项式在点 a 的值就是 SG 平滑后点 a 的值^[23]。其计算公式为

$$x_{a,smooth} = \frac{1}{H} \sum_{i=-\omega}^{\omega} x_{a+i} h_i \tag{3}$$

式中 $x_{a,smooth}$ ——SG 预处理后的反射率
 H ——归一化因子
 ω —— $\frac{1}{2}$ 窗口宽度

h_i ——由最小二乘拟合得到的平滑系数

本试验选择窗口宽度为 5 个数据点,多项式次数为 2。

标准正态变换 (SNV) 用来减小颗粒大小不均匀和粒子表面非特异性散射的影响,SNV 是基于高光谱矩阵的行,即对 1 条光谱进行处理^[24]。公式为

$$x_{snv} = \frac{x - \omega_1}{\sqrt{\frac{\sum_{k'=1}^m (x_{k'} - \omega_1)^2}{m - 1}}} \tag{4}$$

其中 $\omega_1 = \frac{\sum_{k'=1}^m x_{k'}}{m}$

式中 x_{snv} ——SNV 预处理后的反射率
 x ——样品反射率的平均值
 m ——波长点数

1.2.2 特征降维

主成分分析是一种常用的特征提取方法。它沿着样本矩阵协方差最大的方向由高维空间向低维空间投影,将 1 组变量通过线性变换转换到 1 个新的坐标系中^[25-26]。在新坐标系中使得任何数据投影的第 1 大方差在第 1 个坐标轴 (第 1 个主成分) 上,第 2 大方差在第 2 个坐标轴 (第 2 个主成分) 上且和第 1 个坐标轴正交,依次类推直至主成分个数与原数据中的特征数目相等。PCA 得到的主成分是原始变量的线性组合,各个主成分之间相互正交,且包含的信息互不重叠。因此,PCA 可以有效地解决光谱数据中存在的多重共线性问题,去除冗余特征^[27]。在实际中只需保留方差贡献率最大的前几个主成分就可以包含原数据中的主要信息,并使用其前几个主成分代替原始变量建立模型。

蚁群优化算法最先应用于解决 TSP 问题^[28]。ACO 具有分布性、并行性、全局性、鲁棒性强等特点,已在图像分类、无线传感器、网络性能优化、模糊控制设置等领域得到了广泛应用^[29]。基于 ACO 特征波段选择过程如下:

(1) 状态转移准则

ACO 的路径选择由概率决定,每只蚂蚁根据各条路径上的信息量独立地选择下一个转移节点,将蚂蚁 k 走过的节点记录在禁忌表中,在 t 时刻蚂蚁 k 从节点 i 转移到节点 j 的状态转移概率计算公式为

$$P_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t) \gamma_{ij}^\beta(t)}{\sum_{s \in allowed_k} \tau_{is}^\alpha(t) \gamma_{is}^\beta(t)} & (j \in allowed_k) \\ 0 & (j \notin allowed_k) \end{cases} \tag{5}$$

其中 $\gamma_{ij} = \frac{1}{d_{ij}}$ (6)

式中 τ_{ij} —— t 时刻节点 i 到节点 j 路径上的信息素浓度

γ_{ij} ——蚂蚁 k 从节点 i 到 j 的期望程度

d_{ij} ——从节点 i 到节点 j 的距离

α ——信息启发因子,反映信息素对蚂蚁选择路径的影响力

β ——期望启发因子,表示路径长度对蚂蚁的影响程度

$allowed_k$ ——蚂蚁下一步可以选择的节点

(2) 信息素更新

当蚂蚁完成 1 次循环遍历后,各路径上的信息素更新公式为

$$\tau_{ij}(t+n) = (1-\rho)\tau_{ij} + \Delta\tau_{ij}$$
 (7)

$$\Delta\tau_{ij}^k = \begin{cases} Q/L_k & (\text{蚂蚁 } k \text{ 在本次循环中经过}(i,j)) \\ 0 & (\text{蚂蚁 } k \text{ 在本次循环中不经过}(i,j)) \end{cases}$$
 (8)

其中 $\Delta\tau_{ij} = \sum_{k=1}^m \Delta\tau_{ij}^k(t)$

式中 ρ ——信息素挥发系数,反映信息素的持久性

$\Delta\tau_{ij}^k(t)$ ——本次循环完成后节点 i 到节点 j 路径上的信息素浓度增量

$\Delta\tau_{ij}^k$ ——蚂蚁 k 本次循环中在路径 (i,j) 上留下的信息素数量

L_k ——蚂蚁 k 循环 1 周的路径长度

Q ——信息素强度因子

1.2.3 识别模型建立

支持向量机(SVM)适用于解决小样本、非线性以及高维度的数据分类问题^[30]。它根据有限的样本信息在模型的复杂性之间寻求最佳折中,以获得最好的泛化能力。SVM 的基本思想是:在线性可分的情况下,利用间隔最大化的学习策略寻找一个最优分类超平面;在线性不可分的情况下,通过核函数将数据从低维特征空间映射到高维特征空间实现线性可分^[31]。本文 SVM 分类器选取的核函数为径向基函数(Radical basis function, RBF),对于每一个 SVM 模型都采用网格搜索法结合 5 折交叉验证法进行参数寻优。惩罚系数和核参数 γ 的搜索区间设定为 $2^{-5} \sim 2^5$,步长均为 0.1。随机森林(RF)是一种集成学习方法,是在 bagging 方法基础上的一种扩展变体,适用于并行计算的分类问题,对高维数据也能高效分析^[32]。RF 由 N 棵决策树组成,要求每棵树都是相互独立的,模型的预测结果由各棵决策树的“投票”决定。RF 主要通过引入随机性来保证各决策树之间相互独立,增强了模型的

泛化能力。RF 的主要参数是树的个数和节点的个数,在本文中所有 RF 模型树的个数设置为 1 000,节点个数设置为 10。

1.3 试验环境与模型评价标准

试验平台为 Window 10 系统,GPU 为 Intel(R) Core(TM) i7-7700, RAM 为 16 GB,所有算法均在 Matlab R2020a 中编写和开发。高光谱数据采用 SpecSight 和 ENVI 5.3 软件完成处理。采用准确率、平均精度、Kappa 系数和测试时间评估每个组合模型的性能^[33]。

2 结果与分析

2.1 预处理

原始光谱曲线及其预处理后的光谱曲线如图 2 所示。由图 2a 可知,在整个光谱波段范围内,锈病、早期轮斑病和健康叶片的光谱反射率分布存在大部分重叠。为了更有利于对原始光谱数据的识别,采用了 3 种预处理方法来消除噪声并突出反射率随波长的分布规律。经过 FD 处理后光谱频带的两端(400~500 nm 和 900~1 000 nm)为水平线,说明原始光谱数据明显受到了系统噪声的干扰(图 2b)。与图 2a 相比,原始光谱数据经 SG 处理后整个光谱波段范围内反射率的微小波动被消除且曲线有所平滑(图 2c)。图 2d 为原始光谱数据经过 SNV 预处理后的光谱曲线,在整个波长范围内,明显放大了样本之间的特征差异。

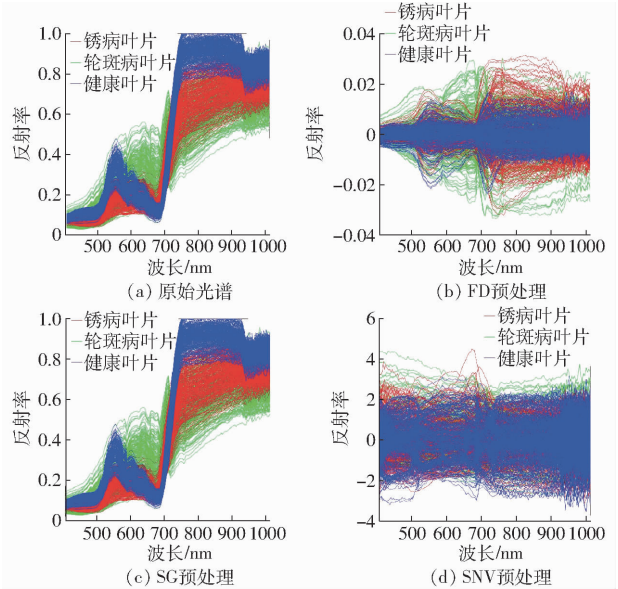


图 2 原始光谱与预处理后的光谱反射率曲线

Fig.2 Reflectance curves of original and preprocessing spectra

综上所述,3 种预处理方法都能消除原始光谱数据的部分噪声。为了确定最合适的预处理方法,下一步将分析每种预处理方法结合降维方法和分类

器对样本识别性能的影响。

2.2 主成分分析降维

在预处理结果的基础上对光谱数据进行主成分

分析。一般来说,第一主成分和第二主成分对原始数据的变化最大,938 个样本的第一和第二主成分得分如图 3 所示。

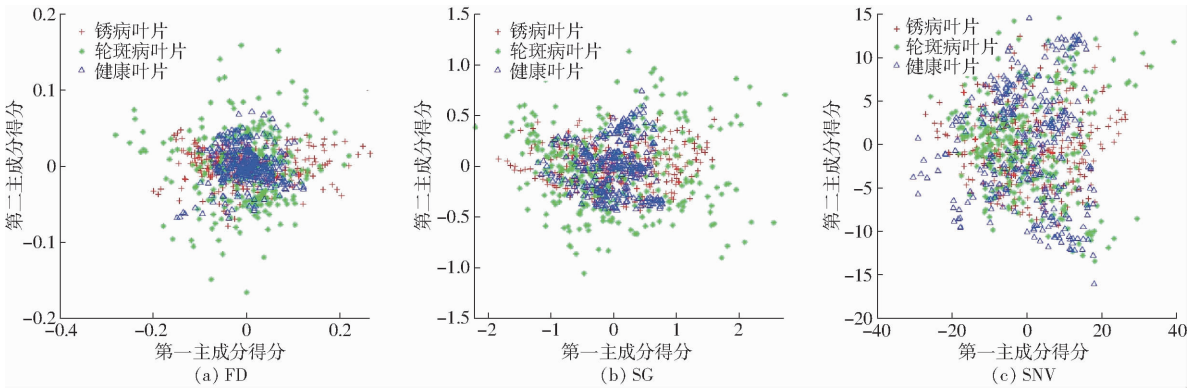


图 3 基于原始光谱数据的第一和第二主成分得分

3 种数据集都以坐标原点为中心进行聚类 and 分布,图 3 为不同预处理方法对样本聚类的影响。通过比较发现图 3a 聚类效果最好,图 3b 次之,图 3c 聚类效果最差。此外,3 种样本之间的重叠现象明显。因此,需要考虑更多的主成分来进行 3 类样本的识别。

如图 4 所示,计算了所有主成分的累积贡献率,基于 FD、SG 和 SNV 的累积贡献率曲线几乎

一致。在前 10 个主成分附近,所有曲线都已经趋于平滑且累积贡献率超过 95%。因此,前 10 个主成分能充分表示包含在 176 维原始光谱数据中的大部分信息。为了更准确地对 3 类样本进行识别,以前 k 个主成分作为模型的输入,观察模型识别准确率随主成分个数的变化关系,因为 k 的取值不仅与预处理方法有关,还与分类器相关。

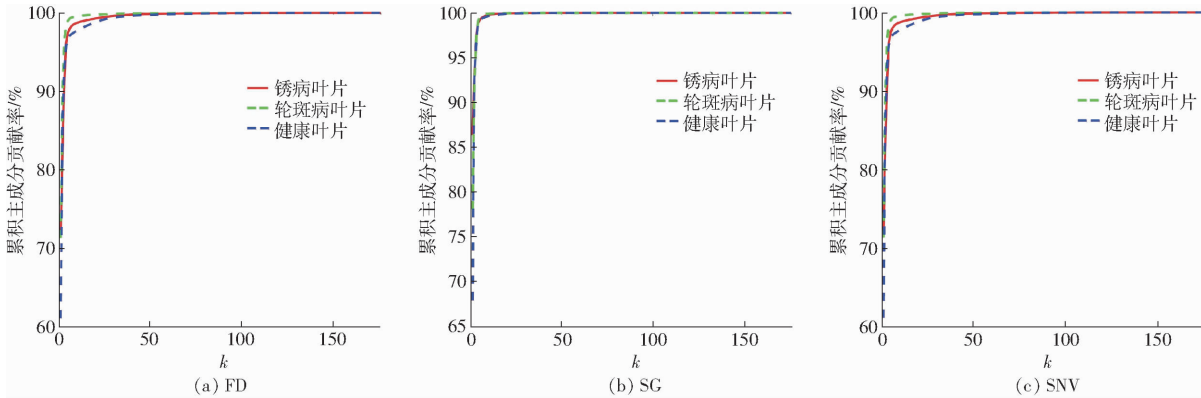


图 4 主成分累积贡献率曲线

Fig. 4 Cumulative contribution rate curves of principal component

图 5 为不同预处理方法和分类算法的识别准确率随主成分个数 k 的变化曲线。随着 k 的增大,所有曲线先上升然后逐渐平稳,最后下降,即随着 k 的增大,准确率首先迅速增大然后达到一个相对稳定值最后降低。对于 FD-PCA-SVM 和 FD-PCA-RF 模型,虽然 k 大于 100 后会获得更高的识别准确率,但随着主成分个数的增大,模型的训练时间和测试时间也会相应增大。3 种预处理方法中,FD 选择的主成分个数最多。然而这种情况是可以预测的,因为 FD 的聚类效果最好,所以需要考虑更多的主成分。所有的 PCA 模型最终选择的主成分个数如表 1 所示。

2.3 蚁群优化算法降维

ACO 是一种元启发式算法,其性能在很大程度上受到算法参数的影响,若参数设置不当,算法易陷入局部最优且无法收敛。由于 ACO 参数空间的庞大性和参数之间的关联性,很难获得最优的参数组合使优化性能最佳。ACO 没有参数选择方面的公式可循,通常根据经验而定。经过多次调试,确定 ACO 相关参数如表 2 所示。

在预处理结果的基础上基于 ACO 进行特征选取的特征序列集合如表 3 所示。

基于 FD 选择的特征个数为 8 个,SG 为 12 个,SNV 为 13 个,不同的预处理方法选择的特征个数及

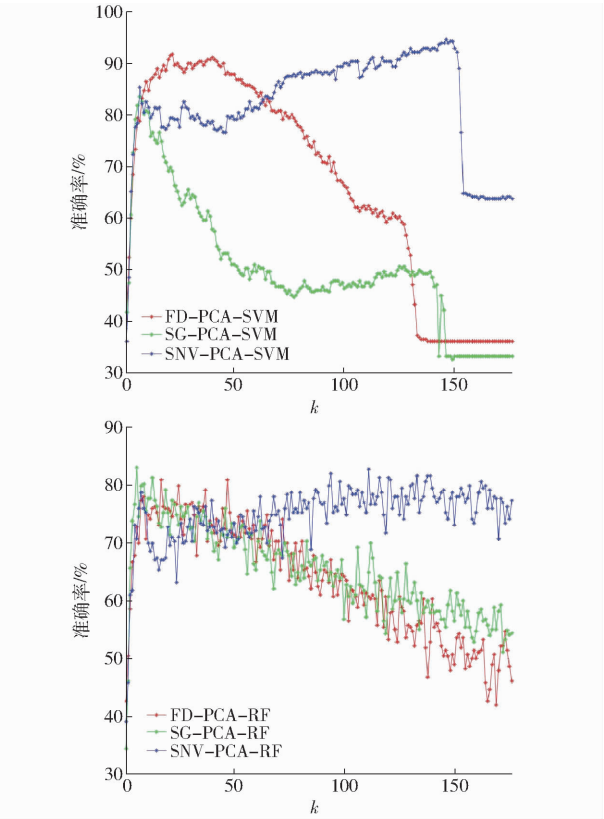


图5 第 k 个主成分的样本识别准确率曲线
Fig.5 Sample recognition accuracy curves of k principal component

表 1 基于主成分分析的特征选择个数

Tab.1 Number of feature selection based on principal component analysis

模型	特征个数
FD-PCA-SVM	22
SG-PCA-SVM	7
SNV-PCA-SVM	7
FD-PCA-RF	17
SG-PCA-RF	6
SNV-PCA-RF	8

特征序列不同。其中,SG 和 SNV 的特征序列比较连续,FD 的特征序列相对分散,但基本集中在前 40

表 2 蚁群优化算法相关参数

Tab.2 Related parameters of ant colony optimization algorithm

参数	数值
蚂蚁个数 m'	50
循环次数 t	20
α	1
β	1.5
ρ	0.3
τ_{ij}	0.7
γ_{ij}	0.1

表 3 基于蚁群优化算法的特征序列集合

Tab.3 Set of feature sequences based on ant colony optimization algorithm

预处理方法	特征个数	特征序列
FD	8	1、3、34、51、71、83、114、148
SG	12	1、2、3、4、6、7、9、29、33、77、102、152
SNV	13	1、2、3、4、5、8、11、12、13、31、63、96、121

注:特征序列中 1 代表第 1 个特征波长(406 nm),依据光谱分辨率(3.5~4.0 nm)可得到后面的特征波长。

个波段范围内(400~532 nm)。表明病变叶片内部成分变化多集中响应在此波段范围内。

2.4 基于不同组合模型的识别结果与评价

将 PCA、ACO 算法所选择的特征以及全波段特征作为 SVM 的输入变量,各 SVM 模型的特征个数、分类准确率、平均精度、Kappa 系数、测试时间以及惩罚系数 C 和核参数 γ 如表 4 所示。 γ 是 RBF 自带的一个参数,决定了数据映射到新的特征空间后的分布,控制分割超平面的形状, γ 越大,支持向量越少。 C 表示犯错成本,成本大意味着模型对误差的惩罚大,对应训练集中的误差小。Kappa 系数是一个用于一致性检验的指标,对于分类问题,一致性就是模型的预测结果和实际分类结果是否一致。从表 4 可以看出,测试集的准确率在 79.79%~99.65% 之间,平均精度在 79.48%~99.65% 之间,Kappa 系数在 0.696 8~0.994 7 之间,测试时间在

表 4 各 SVM 组合模型分类结果

Tab.4 Classification results of each SVM combined model

模型	特征个数	准确率/%	平均精度/%	Kappa 系数	测试时间/s	惩罚系数 C	核参数 γ
FD-PCA-SVM	22	91.84	92.19	0.877 7	0.006 5	2.00	2.00
SG-PCA-SVM	7	83.69	83.80	0.755 3	0.002 5	2.14	5.28
SNV-PCA-SVM	7	85.46	85.55	0.781 9	0.002 8	2.46	7.46
FD-ACO-SVM	8	83.33	83.62	0.750 0	0.002 2	16.00	6.96
SG-ACO-SVM	12	94.33	94.46	0.914 9	0.001 1	2.83	6.50
SNV-ACO-SVM	13	79.79	79.48	0.696 8	0.005 4	3.48	6.06
FD-SVM	176	92.20	92.51	0.883 0	0.028 0	6.06	0.50
SG-SVM	176	99.65	99.65	0.994 7	0.018 7	2.00	0.50
SNV-SVM	176	84.04	83.82	0.760 6	0.034 3	4.29	0.54

0.001 1 ~ 0.034 3 s 之间。在不降维模型中,SG - SVM 模型取得了最好的分类效果,准确率达到 99.65%;其次为 FD - SVM 模型,准确率为 92.20%;SNV - SVM 模型效果最差,准确率仅为 84.04%。在降维模型中,SG - ACO - SVM 模型取得了最好的识别效果,准确率达到 94.33%;其次为 FD - PCA - SVM 模型,准确率为 91.84%;SNV - ACO - SVM 模型识别效果最差,准确率仅达到 79.79%。综上所述,在所有的预处理方法中,SG 平滑效果最优,SNV 效果最差;降维算法 PCA 和 ACO 各有优劣;模型性能与预处理方法、降维算法以及分类器相关。

将 PCA、ACO 算法所选择的特征以及全波段特征作为 RF 的输入变量,各 RF 模型的特征个数、分类准确率、平均精度、Kappa 系数、测试时间如表 5 所示。从表 5 可以看出,测试集的准确率在 76.36% ~ 97.94% 之间,平均精度在 75.87% ~ 97.96% 之间,Kappa 系数在 0.643 1 ~ 0.968 1 之间,测试时间在 0.010 6 ~ 0.025 5 s 之间。在不降维模型中,SG - RF 模型分类效果最好,准确率达到 97.94%,FD - RF 模型次之,SNV - RF 模型分类效果最差,不降维模型的分类准确率均不低于 76.36%;在降维模型中,SG - ACO - RF 模型取得了最好的分类效果,准确率达到 96.52%;其次为 FD - ACO - RF 和 FD - PCA - RF 模型,准确率分别为 80.69% 和 80.20%。综上所述,在 3 种不同的预处理方法中,SG 平滑效果最优,FD 次之,SNV 效果最差;在两种不同的降维算法中,从整体上看,ACO 优于 PCA。

表 5 各 RF 组合模型分类结果

Tab.5 Classification results of each RF combined model

模型	特征 个数	准确 率/%	平均 精度/%	Kappa 系数	测试时 间/s
FD - PCA - RF	17	80.20	80.23	0.701 8	0.019 5
SG - PCA - RF	6	77.24	76.86	0.654 5	0.023 5
SNV - PCA - RF	8	79.69	79.88	0.692 1	0.020 7
FD - ACO - RF	8	80.69	81.41	0.711 8	0.019 3
SG - ACO - RF	12	96.52	96.58	0.946 8	0.010 6
SNV - ACO - RF	13	78.01	77.50	0.668 3	0.019 1
FD - RF	176	83.58	83.95	0.750 1	0.023 3
SG - RF	176	97.94	97.96	0.968 1	0.011 9
SNV - RF	176	76.36	75.87	0.643 1	0.025 5

由表 4 和表 5 可知,在所有的 PCA 降维模型以及不降维模型中,SVM 分类器的准确率明显高于 RF;基于 ACO 降维的分类模型中,SG - ACO - RF 模型准确率最高,仅使用 12 维的波段组合就能达到

接近于不降维模型 SG - RF 的识别准确率。当采用相同的分类器 SVM 时,除 SNV - PCA - SVM 模型外,所有降维模型相比于不降维模型其分类平均精度都有所降低,但是通过降维处理得到的模型运行时间明显变短,即提高了模型运行效率。在 RF 分类器中,同样将进行降维处理和不降维处理的模型进行对比,SNV - PCA - RF 和 SNV - ACO - RF 模型识别平均精度提高,其余模型都没有提高。同时,经过降维处理的模型 SG - PCA - RF 的运行时间大于不进行降维处理的模型 SG - RF,这说明通过降维处理不一定能缩短模型的运行时间。综上,并非所有的预处理方法、降维方法以及分类器的组合都能在降低维度的同时能提高识别精度。

根据准确率、平均精度和测试时间可知:在基于降维处理 PCA 建立的 SVM 模型中,FD - PCA - SVM 模型最优;在基于降维处理 ACO 建立的 SVM 模型中,SG - ACO - SVM 模型最优;在不进行降维处理的 SVM 模型中,SG - SVM 模型最优;在基于降维处理 PCA 建立的 RF 模型中,FD - PCA - RF 模型最优;在基于降维处理 ACO 建立的 RF 模型中,SG - ACO - RF 模型最优;在不进行降维处理的 RF 模型中,SG - RF 模型最优。综上所述,SG - SVM 模型性能最佳。

3 结论

(1)采用高光谱成像技术,获取目标样本的光谱数据,用 ENVI 5.3 软件提取甘蔗病害叶片的感兴趣区域,再用 FD、SG 平滑、SNV 分别对原始光谱数据进行预处理。相较而言,基于 SG 平滑的预处理效果最佳。

(2)在预处理的基础上,使用 PCA、ACO 对全波段进行特征降维,两种方法均去除了相关性小的波段,在减少模型计算量的同时还获得了较好的识别精度。

(3)在病害识别模型建立方法上,使用了 SVM 和 RF 两种分类器。为了确定甘蔗叶片早期轮斑病和锈病的最佳识别模型,分别采用了 3 种预处理方法、2 种降维方法和 2 种分类器随机组合的方法对高光谱图像数据进行处理。在 18 种组合模型中,SG - SVM 模型识别效果最佳,其测试集准确率为 99.65%,平均精度为 99.65%,Kappa 系数为 0.994 7,测试时间为 0.018 7 s。结果表明,利用高光谱成像技术进行甘蔗叶片早期轮斑病和锈病的识别可行且有效,可为植保无人机超低空遥感病害监测提供参考。

参考文献

- [1] KARAMCHANDANI B M, CHAKRABORTY S, DALVI S G, et al. Chitosan and its derivatives: promising biomaterial in averting fungal diseases of sugarcane and other crops[J]. *Journal of Basic Microbiology*, 2022, 62(5): 533–554.
- [2] THILAGAVATHI K, KAVITHA K, PRABA R D, et al. Detection of diseases in sugarcane using image processing techniques[J]. *Bioscience Biotechnology Research Communications*, 2020, 13(11): 109–115.
- [3] 李文凤, 李婕, 普家荣, 等. 低纬高原甘蔗中后期灾害性真菌病害发生为害调查分析[J]. *农学学报*, 2022, 12(6): 24–32, 59.
- LI Wenfeng, LI Jie, PU Jiarong, et al. The disastrous fungal disease at middle and late stages of sugarcane growth in low latitude plateau: investigation and analysis of occurrence and damage[J]. *Journal of Agriculture*, 2022, 12(6): 24–32, 59. (in Chinese)
- [4] XU J, MIAO T, ZHOU Y, et al. Classification of maize leaf diseases based on hyperspectral imaging technology[J]. *Journal of Optical Technology*, 2020, 87(4): 212–217.
- [5] 刘莉, 陶红燕, 方静, 等. 基于近红外高光谱的梨叶片炭疽病与黑斑病识别[J]. *农业机械学报*, 2022, 53(2): 221–230.
- LIU Li, TAO Hongyan, FANG Jing, et al. Identifying anthracnose and black spot of pear leaves on near-infrared hyperspectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(2): 221–230. (in Chinese)
- [6] 赵森, 付芸, 崔江南, 等. 高光谱的刺五加黑斑病的早期检测研究[J]. *光谱学与光谱分析*, 2021, 41(6): 1898–1904.
- ZHAO Sen, FU Yun, CUI Jiangnan, et al. Application of hyperspectral imaging in the diagnosis of *Acanthopanax senticosus* black spot disease[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(6): 1898–1904. (in Chinese)
- [7] JIANG Q, WU G, TIAN C, et al. Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features[J]. *Infrared Physics and Technology*, 2021, 118: 103898.
- [8] SALIMI A, ZIAI M, AMIRI A, et al. Using a feature subset selection method and support vector machine to address curse of dimensionality and redundancy in hyperion hyperspectral data classification[J]. *Egyptian Journal of Remote Sensing and Space Sciences*, 2018, 21(1): 27–36.
- [9] 李鑫星, 朱晨光, 白雪冰, 等. 基于可见光谱和支持向量机的黄瓜叶部病害识别方法研究[J]. *光谱学与光谱分析*, 2019, 39(7): 2250–2256.
- LI Xinxing, ZHU Chenguang, BAI Xuebing, et al. Recognition method of cucumber leaves diseases based on visual spectrum and support vector machine[J]. *Spectroscopy and Spectral Analysis*, 2019, 39(7): 2250–2256. (in Chinese)
- [10] HUANG Y, LI J, YANG R, et al. Hyperspectral imaging for identification of an invasive plant *Mikania micrantha* kunth[J]. *Frontiers in Plant Science*, 2021, 12: 626516.
- [11] 李志伟, 袁婧, 丁为民, 等. 基于高光谱成像技术识别水稻纹枯病[J]. *华南农业大学学报*, 2018, 39(6): 97–103.
- LI Zhiwei, YUAN Jing, DING Weimin, et al. Identification of rice sheath blight based on hyperspectral imaging technique[J]. *Journal of South China Agricultural University*, 2018, 39(6): 97–103. (in Chinese)
- [12] 刘爽, 谭鑫, 刘成玉, 等. 高光谱数据处理算法的小麦赤霉病籽粒识别[J]. *光谱学与光谱分析*, 2019, 39(11): 3540–3546.
- LIU Shuang, TAN Xin, LIU Chengyu, et al. Recognition of fusarium head blight wheat grain based on hyperspectral data processing algorithm[J]. *Spectroscopy and Spectral Analysis*, 2019, 39(11): 3540–3546. (in Chinese)
- [13] 汪六三, 黄子良, 王儒敬. 基于近红外光谱和机器学习的大豆种皮裂纹识别研究[J]. *农业机械学报*, 2021, 52(6): 361–368.
- WANG Liusan, HUANG Ziliang, WANG Ruijing. Identification of soybean seed coat crack based on near infrared spectroscopy and machine learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(6): 361–368. (in Chinese)
- [14] 张朱珊莹, 顾翰文, 谢凯文, 等. 基于近红外光谱的预处理及组合方法[J]. *激光与光电子学进展*, 2021, 58(16): 472–479.
- ZHANG Zhushanying, GU Hanwen, XIE Kaiwen, et al. Pretreatment and combined method based on near infrared spectroscopy[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 472–479. (in Chinese)
- [15] 刘燕德, 肖怀春, 孙旭东, 等. 柑桔叶片黄龙病光谱特征选择及检测模型[J]. *农业工程学报*, 2018, 34(3): 180–187.
- LIU Yande, XIAO Huaichun, SUN Xudong, et al. Spectral feature selection and discriminant model building for citrus leaf Huanglongbing[J]. *Transactions of the CSAE*, 2018, 34(3): 180–187. (in Chinese)
- [16] 秦立峰, 张熹, 张晓茜. 基于高光谱病害特征提取的温室黄瓜霜霉病早期检测[J]. *农业机械学报*, 2020, 51(11): 212–220.
- QIN Lifeng, ZHANG Xi, ZHANG Xiaoqian. Early detection of cucumber downy mildew in greenhouse by hyperspectral disease differential feature extraction[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(11): 212–220. (in Chinese)
- [17] 王翔宇, 朱晨光, 傅泽田, 等. 基于可见光光谱分析的黄瓜白粉病识别研究[J]. *光谱学与光谱分析*, 2019, 39(6): 1864–1869.
- WANG Xiangyu, ZHU Chenguang, FU Zetian, et al. Research on cucumber powdery mildew recognition based on visual spectra[J]. *Spectroscopy and Spectral Analysis*, 2019, 39(6): 1864–1869. (in Chinese)

[18] 袁培森, 曹益飞, 马千里, 等. 基于 Random Forest 的水稻细菌性条斑病识别方法研究[J]. 农业机械学报, 2021, 52(1): 139 – 145, 208.
YUAN Peisen, CAO Yifei, MA Qianli, et al. Identification method of rice bacterial leaf streak based on Random Forest[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1): 139 – 145, 208. (in Chinese)

[19] 吴叶兰, 陈怡宇, 廉小亲, 等. 高光谱成像的柑橘病虫害叶片识别方法[J]. 光谱学与光谱分析, 2021, 41(12): 3837 – 3843.
WU Yelan, CHEN Yiyu, LIAN Xiaoqin, et al. Study on the identification method of citrus leaves based on hyperspectral imaging technique[J]. Spectroscopy and Spectral Analysis, 2021, 41(12): 3837 – 3843. (in Chinese)

[20] 王小龙, 邓继忠, 黄华盛, 等. 基于高光谱数据的棉田虫害鉴别研究[J]. 华南农业大学学报, 2019, 40(3): 97 – 103.
WANG Xiaolong, DENG Jizhong, HUANG Huasheng, et al. Identification of pests in cotton field based on hyperspectral data [J]. Journal of South China Agricultural University, 2019, 40(3): 97 – 103. (in Chinese)

[21] ELMASRY G M, NAKAUCHI S. Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality—a comprehensive review[J]. Biosystems Engineering, 2016, 142: 53 – 82.

[22] 朱梦远, 杨红兵, 李志伟. 高光谱图像和叶绿素含量的水稻纹枯病早期检测识别[J]. 光谱学与光谱分析, 2019, 39(6): 1898 – 1904.
ZHU Mengyuan, YANG Hongbing, LI Zhiwei. Early detection and identification of rice sheath blight disease based on hyperspectral image and chlorophyll content[J]. Spectroscopy and Spectral Analysis, 2019, 39(6): 1898 – 1904. (in Chinese)

[23] CHENG Y, YANG S, LIU X, et al. Identification of wine grape varieties based on near-infrared hyperspectral imaging[J]. Applied Engineering in Agriculture, 2019, 35(6): 959 – 967.

[24] SHEN Y, LI B, LI G, et al. Rapid identification of producing area of wheat using terahertz spectroscopy combined with chemometrics[J]. Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy, 2022, 269: 120694.

[25] WU J, LI X, LIN L, et al. Fast hyperspectral discrimination of rice origin based on AlexNet convolutional neural network[J]. Journal of Chinese Institute of Food Science and Technology, 2022, 22(1): 282 – 288.

[26] XIA C, YANG S, HUANG M, et al. Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis[J]. Infrared Physics and Technology, 2019, 103: 103077.

[27] WANG L, PANG L, YAN L, et al. Nondestructive rapid identification of soybean varieties using hyperspectral imaging technology[J]. Journal of Applied Spectroscopy, 2022, 89(1): 84 – 91.

[28] DING X, LI H, ZHANG S. Optimized band selection of hyperspectral remote sensing image based on polymorphic ant colony algorithm[J]. Remote Sensing Technology and Application, 2016, 31(2): 275 – 284.

[29] DING X H, ZHANG S Q, LI H P, et al. A restrictive polymorphic ant colony algorithm for the optimal band selection of hyperspectral remote sensing images[J]. International Journal of Remote Sensing, 2020, 41(3): 1093 – 1117.

[30] WAKHOLI C, KANDPAL L M, LEE H, et al. Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics[J]. Sensors and Actuators B-Chemical, 2018, 255: 498 – 507.

[31] LI H D, XU Q S, LIANG Y Z. libPLS: an integrated library for partial least squares regression and linear discriminant analysis [J]. Chemometrics and Intelligent Laboratory Systems, 2018, 176: 34 – 43.

[32] BALLANTI L, BLESIOUS L, HINES E, et al. Tree species classification using hyperspectral imagery: a comparison of two classifiers[J]. Remote Sensing, 2016, 8(6): 445.

[33] XU H L, ZHANG H Y, HE W, et al. Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification[J]. Neurocomputing, 2019, 360: 138 – 150.

(上接第 232 页)

[28] SHI Qiang, ZHANG Jianlin, YANG Ming. Curvature adaptive control based path following for automatic driving vehicles in private area[J]. Journal of Shanghai Jiaotong University (Science), 2021, 26(5): 690 – 698.

[29] 张华强, 王国栋, 吕云飞, 等. 基于改进纯追踪模型的农机路径跟踪算法研究[J]. 农业机械学报, 2020, 51(9): 18 – 25.
ZHANG Huaqiang, WANG Guodong, LÜ Yunfei, et al. Agricultural machinery automatic navigation control system based on improved pure tracking model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 18 – 25. (in Chinese)

[30] 段贤强, 陶建峰, 覃程锦, 等. 变速条件下农业机械路径跟踪稳定控制方法[J]. 农业机械学报, 2019, 50(9): 18 – 24, 32.
DUAN Xianqiang, TAO Jianfeng, QIN Chengjin, et al. Agricultural machinery path tracking stability control method under variable speed conditions[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(9): 18 – 24, 32. (in Chinese)