

基于叶绿素荧光光谱技术的茶叶藻斑病模型研究

刘燕德, 林晓东, 高海根, 王 舜, 高 雪

华东交通大学机电与车辆工程学院, 水果智能光电检测技术与装备国家地方联合工程研究中心, 江西 南昌 330013

摘 要 茶叶是我国重要的经济作物,对茶叶病害的及早发现与诊断,有利于农业生产者及时采取有效的防护措施。为了实现茶叶病害的准确判别,采用叶绿素荧光光谱对茶叶的光谱特性展开研究。实验采集了健康茶叶样本90片,藻斑病轻度病害叶片90片,藻斑病重度病害叶片90片,并根据 Kennard-Stone 算法将样本数按3:1划分训练集和预测集样本数,其中校正集为200个、验证集为70个。采用叶绿素荧光光谱采集系统对茶叶藻斑病、正常叶片进行光谱采集,其中采集参数设置为:积分时间20 ms,激光功率40 mW。分别分析了患病叶片和正常叶片的光谱响应特性,总体上看,三种叶片光谱主要存在吸收强度差异,光谱走势基本一致。在685和740 nm附近存在叶绿素的荧光峰,其差异主要表现在正常叶片光谱较另外两种叶片光谱吸收强度较高,而重度病害强度最低。然后使用多项式平滑(Savitzky-Golay)对原始光谱进行平滑和降噪处理,建立了偏最小二乘判别模型(PLS-DA),在PLS-DA建模集模型中,误判样品数为3个,误判率为3%;PLS-DA预测集模型中,误判样品个数为5个,误判率为7.1%。然后建立4种不同核函数的支持向量机模型进行比较得到,由RBF作为核函数,经主成分分析法(PCA)降维后的变量建立的SVM模型误判率最低,准确率达到95.72%,最后采用PCA结合线性判别分析方法(LDA)建立的模型效果最好,准确率达到98.9%。其中最优主成分数的选取由留一验证法取得,选用前10个主成分进行建模时,交叉验证准确率最高,达98%。通过模型对比得到PLS-DA建模集和预测集精度都达到90%以上,以四种核函数建立的支持向量机模型中,径向基核函数模型效果较优,达到95.72%。经主成分分析后建立的LDA效果最好,识别率为98.9%。该研究采用叶绿素荧光光谱结合化学计量学对茶叶病害进行识别,为茶叶病害的快速、准确预测提供一种新方法。

关键词 荧光光谱;主成分分析;偏最小二乘判别法;支持向量机;线性判别分析;藻斑病

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)07-2129-06

引 言

茶叶中含有丰富多样的化学成分,药理活性高,且富含大量的营养元素,如茶多酚、氨基酸、咖啡碱、矿物质、维生素等,可以增进人体健康,但是茶叶的品质和健康状态密切相关。从病原物种类来看,茶叶病害主要分病害(真菌、细菌、病毒等)和虫害两大类。茶藻斑病(*Cephaleuros virescens* Kunze)属菌丝病害,是茶树老叶部位的常见病害,广泛发生于各茶产区。病原藻以营养体在病叶上越冬,翌年春季在潮湿的条件下可产生游动孢子囊和可借风雨传播到茶树的游动孢子,在水中发芽的游动孢子侵入叶片角质层后,即引发茶藻斑病。并且,由于引发藻斑病的病原可在表皮细胞和角

质层之间蔓延,以后继续产生游动孢子,持续地借风雨飞溅传播,不断进行再侵染,使病害不断扩大蔓延,一般在高温多雨的季节,往往有茶藻斑病的流行^[1]。菌丝病害在染病前期并不能快速识别,导致叶枯、叶黄等现象,又具有传染性,导致一片茶园患病,对无公害茶园的影响巨大,造成巨大的经济损失。所以找到一种快速、准确、经济的检测手段是非常有必要的。

当前农作物病害问题成为了大量科研人员的关注问题。Xie^[2]采用高光谱技术研究了茄子叶部病害的光谱和纹理特征,结果表明,光谱和纹理特征在茄子叶片早期疫病检测中是有效的。Wetterich^[3]等利用中心波长分别为365,405,470及530 nm的LED作激发光源,采集黄龙病柑橘叶片在570,610,690及740 nm波长下的荧光图像,通过分割提取

收稿日期:2020-07-06,修订日期:2020-11-21

基金项目:国家自然科学基金项目(31760344),水果光电检测技术能力提升项目(S2016-90)和油茶早期炭疽病激光击穿光谱技术识别方法研究项目(GJJ190306)资助

作者简介:刘燕德,女,1967年生,华东交通大学机电与车辆工程学院教授 e-mail: jxliuyd@163.com

等步骤对图像进行处理,采用支持向量机建立病害检测模型,对病害叶片和正常叶片的区分准确率达到 90% 以上。隋媛媛^[4]等利用叶绿素荧光构建黄瓜活体叶片光谱指数研究温室黄瓜霉霜病的预测问题,分析不同染病阶段的荧光光谱强度,筛选出具有特征性的光谱指数进行定量分析。叶绿素荧光与植物光合作用密切相关,其变化早于叶绿素含量的下降及植物形态结构的变化,对植物所受到的胁迫反应灵敏,叶绿素荧光光谱在评价植物胁迫状态、生长状况的早期检测中被广泛应用。王迎旭^[5]等采用叶绿素荧光成像系统采集全植株冠层图像,以褐斑病和炭疽病胁迫下的黄瓜植株作为试验材料,分析植株生理状态,建立基于叶绿素荧光参数的病害分类和病情诊断模型。周丽娜^[6]等研究表明通过探测叶绿素荧光可实现水稻瘟病快速、准确、无损测量。竞霞^[7]等研究了日光诱导叶绿素荧光(solar-induced chlorophyll fluorescence, SIF)对冬小麦条锈病早期探测的可行性,通过相关性分析优选了遥感探测小麦条锈病早期的特征参量。万文博^[8]等用 355 nm 波长的激光光源,采用时间分辨测量法,得到完整的荧光信号分布图像,证明叶绿素荧光强度与植物生理有相对的关联。

目前叶绿素荧光光谱在农业上的运用比较新颖,但是在茶叶病害的检测方面的应用很少,而且叶绿素荧光对植物的胁迫状态关系密切。因此,本文探索利用激光诱导叶绿素荧光光谱技术为研究手段,研究健康茶叶叶片与藻斑病叶片的光谱响应并对其做出识别。

1 实验部分

1.1 样品

样品选自福建省无公害零农药有机茶基地,经过采摘分拣后立即放置在 2 °C 的温度下全程保存,低温能稳定菌丝的生长状态,不会使其腐败。其中,根据叶片的病斑程度分为轻度病害、重度病害以及正常叶片各 90 片,病害分级标准如表 1 所示。

表 1 茶藻斑病病害程度分级标准

Table 1 Classification standard of tea
Cephaleuros virescens Kunze

| 病情程度 | 受害程度 |
|------|--------------------|
| 正常叶片 | 几乎无病 |
| 轻度病害 | 叶片上有零星藻斑菌丝 |
| 重度病害 | 藻斑菌丝覆盖叶片面积 50%~75% |

1.2 叶绿素荧光光谱采集系统

搭建的测量装置如图 1 所示,该系统具体包括:激光光源、光纤、样品池、光谱仪和计算机组成。实验中使用的光谱仪是科研级别光谱仪 QE65000(Ocean Optics, 美国),分辨率为 1.5 nm,积分时间为 7 ms~15 min,工作范围是 200~1 100 nm。光谱仪侧面配有 USB 接口,与电脑相连,由电脑直接供电。光谱仪的线阵 CCD 探测器(东芝,日本)像素为 1044×64。光谱仪配有 SMA905 接口,与同样接口的 VIS-

NIR 光纤(Ocean Optics, USA)相连,光纤纤芯直径为 1 000 μm,数值孔径为 0.22,发散全角为 25.4°光源为 405 nm 的激光光源,采集时参数设定为:积分时间 20 ms,激光功率 40 mW。光源与植物叶片成 45°夹角,光纤垂直对准植物叶片,距离叶片表面高度 3.0 cm。数据由海洋光学配套的软件 SpectraSuite 采集。

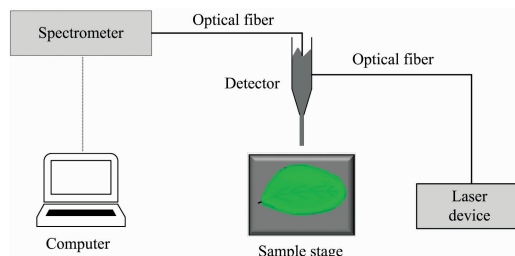


图 1 叶绿素荧光光谱原理图

Fig. 1 Schematic diagram of chlorophyll fluorescence spectrum

2 结果与讨论

2.1 茶叶叶片荧光光谱特征分析

图 2 是正常和轻度、重度藻斑病叶片的叶绿素荧光光谱,叶绿素在可见光波段有两个较强烈的吸收带,分别在 430~450 nm 蓝紫光区域和 640~660 nm 红光区域。总体上看,三种叶片光谱主要存在吸收强度的差异,光谱走势基本一致。在 685 和 740 nm 附近存在叶绿素的荧光峰,其差异主要表现在荧光峰强度的差异,原因为侵染菌丝的叶片细胞的类囊体内叶绿素含量产生了变化。正常叶片光谱较另外两种叶片光谱吸收强度较高,而重度病害强度最低,跟内部叶绿素含量的变化有关。光谱曲线分别在 570, 680 和 750 nm 附近存在突出吸收峰,在 430~600 nm 的波段中,可以看出轻度病害叶片的荧光峰强度高于正常叶片,其原因可能是儿茶素,茶黄素和花青素在病害初期的转化含量发生了了变化^[9-10]。620~780 nm(红光波段)和 780~900 nm(近红光短波段)的叶绿素荧光主要来源于光合作用的叶绿素 a,这一部分叶绿素荧光对植物生理变化反应最为灵敏,经常被用来区分植物种类和对植物所受胁迫状态进行诊断^[11]。

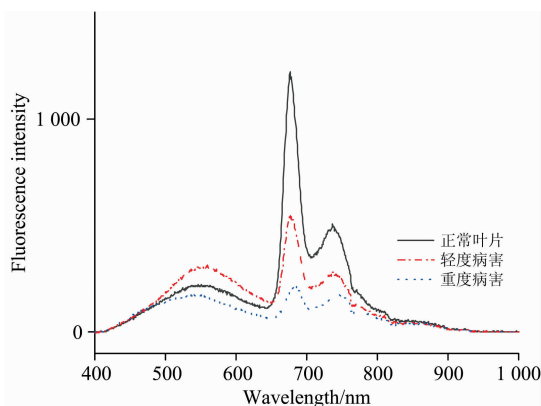


图 2 正常叶片、轻度病害和重度病害荧光光谱

Fig. 2 Fluorescence spectra of normal leaves, mild disease stage and severe disease stage

2.2 茶叶光谱信息变量降维

由于全光谱数据包含的信息量巨大, 获取的叶片光谱变量较多。光谱变量中包含着无用信息, 在建模过程中容易降低模型的预测精度。因此, 为了将光谱的有效信息能最大程度的提取出, 采用主成分分析法(PCA)以期于提高病害鉴别检测模型的预测精度与建模效率。

主成分分析是一种降低维度的方法。将原始变量进行线性变换获得少量数目的新变量, 采用新变量来表征原始变量的数据特征。首先, 对不同病害程度的叶片原始光谱进行主成分分析, 通过 PCA 压缩光谱数据降低维度, 获取能代表整个光谱的有效信息。取前三个主成分得分绘制三个不同产地样品的空间散点图。如图 3 所示, 其中第一个主成分(PC1)贡献率为 91%、PC2 贡献率为 7%、PC3 贡献率为 1%, 累积贡献率达 99%。

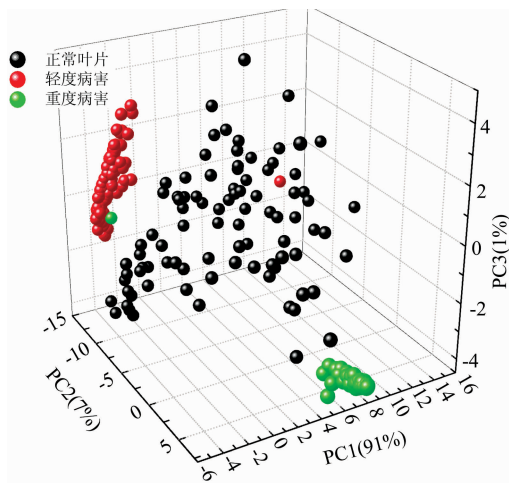


图 3 三类样本的前 3 个主成分得分分布

Fig. 3 The first three PCs score plots of spectra of all cultivars

2.3 基于光谱特征茶叶病害偏最小二乘判别模型研究

在建立模型前, 人为将正常叶片样品赋值为 1, 轻度病害样品赋值为 3, 重度病害样品赋值为 5, 取两者的中间值作为分类阈值。若预测值小于阈值 2 判定为正常叶片, 若预测值介于阈值 2.5 与 3.5 之间判定为轻度病害叶片, 若预测值大于阈值 3.5 判定为重度病害叶片。

如图 4 所示, 在 PLS-DA^[12] 建模集模型中, 误判样品数为 3 个, 误判率为 3%; PLS-DA 预测集模型中, 误判样品个数为 5 个, 误判率为 7.1%。两个判别模型的识别率都高于 90% 以上, 能够较好的将正常叶片和患病叶片分类。误判情况都是在正常叶片和轻度病害叶片上, 可能原因是在 430~600 nm 波段上, 轻度病害叶片儿茶素, 茶黄素和花青素的混合光谱强度高于正常叶片导致误判。

2.4 基于光谱特征的茶叶病害支持向量机判别模型研究

采用支持向量机(support vector machine, SVM)对茶叶叶片的病害程度进行判别分析。使用 SVM 四种核函数建模, 分别为径向基核函数(Rbf-Kernel)、线性核函数(Lin-Kernel)、多项式核(Polynomial-Kernel)和 Sigmoid 核(Sigmoid-Kernel), 如图 5。将经 PCA 方法压缩的变量作为模型的输入

变量, 获得不同变量选择方法的 SVM 模型预测结果, 如表 2 所示。

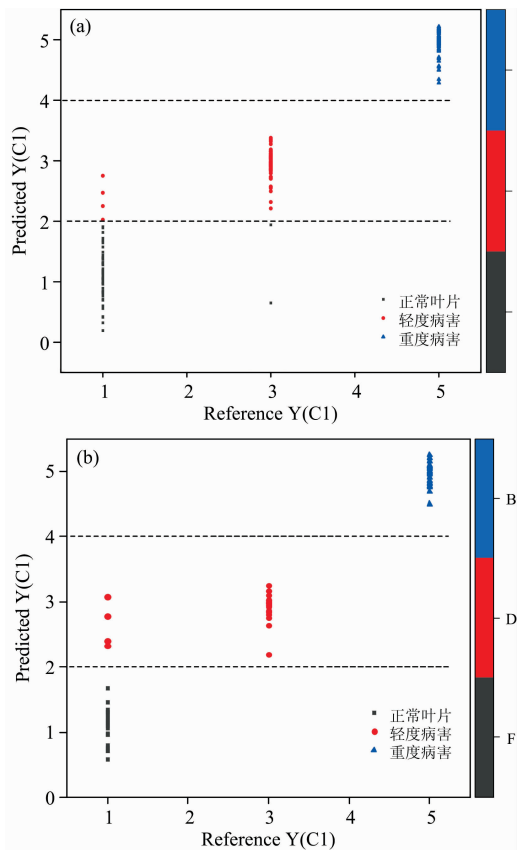


图 4 偏最小二乘判别分析模型

(a): 建模集模型; (b): 预测集模型

Fig. 4 Partial least squares discrimination model

(a): PLS-DA calibration model; (b): PLS-DA prediction model

表 2 不同核函数的支持向量机模型结果

Table 2 Support vector machine model results with different kernel functions

| 核函数 | 总误判率/% | 误判个数 | | |
|------------|--------|------|------|------|
| | | 正常叶片 | 轻度病害 | 重度病害 |
| RBF-Kernel | 4.28 | 0 | 3 | 0 |
| LIN-Kernel | 7.14 | 0 | 5 | 0 |
| Polynomial | 20 | 1 | 13 | 0 |
| Sigmoid | 5.7 | 0 | 4 | 0 |

从表 2 中可知, 采用 RBF-Kernel 的 4 种不同核函数建模, 利用筛选后的 300 个变量建立的 SVM 模型预测集误判率最低, 误判率为 4.28%。利用 Polynomial 核函数建立的模型误判率最高, 其误判率为 20%。与 PLS-DA 分析方法分析的结果相近, 判别错误基本上出现在轻度病害与正常叶片上。综合比较发现选择 RBF 作为核函数, PCA 为降维方法, 建立的 SVM 模型误判率最低, 表明定性效果最好。

2.5 基于光谱特征的茶叶病害主成分-线性判别分析(PCA-LDA)判别模型研究

在建立判别模型之前, 为消除高频随机噪声、基线漂移、

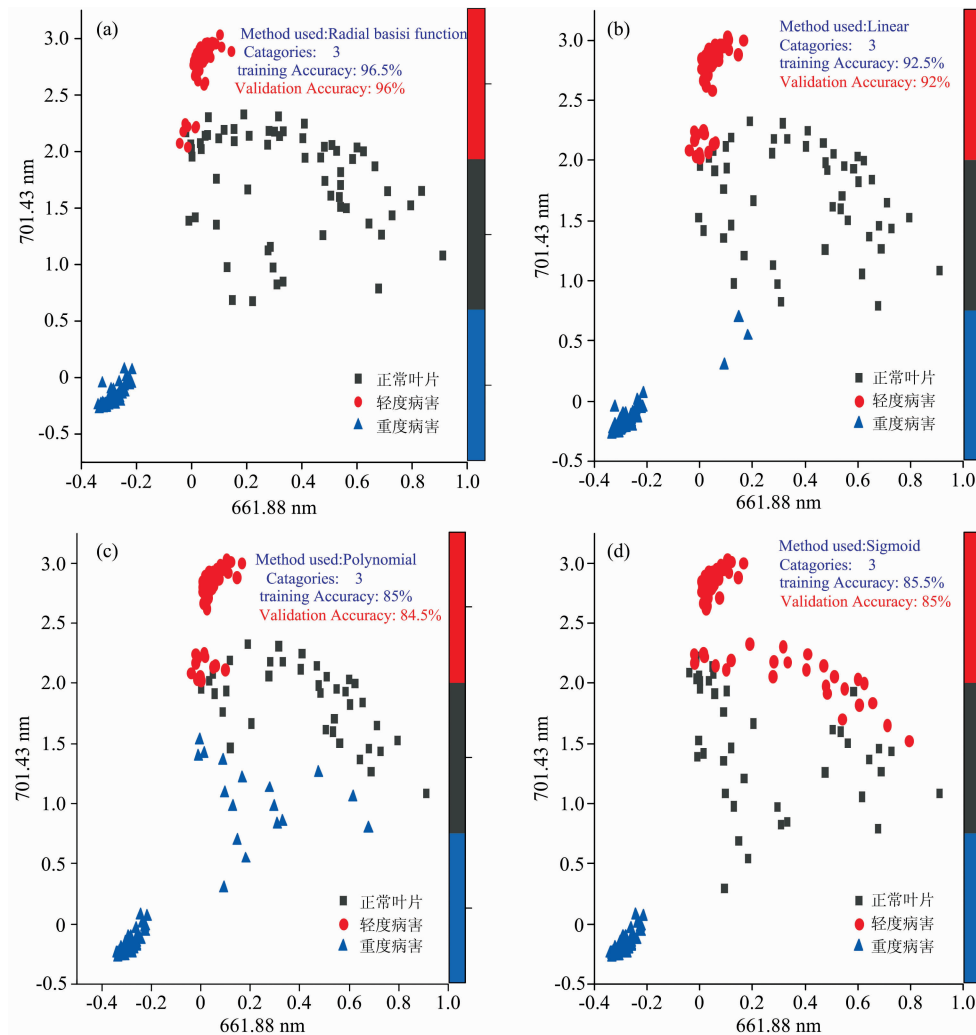


图 5 不同核函数支持向量机预测模型

(a): RBF 核函数预测模型; (b): LINEAR 核函数预测模型; (c): Polynomial 核函数预测模型; (d): Sigmoid 核函数预测模型

Fig. 5 Support vector machine prediction model with different kernel functions

(a): RBF-Kernel prediction model; (b): LIN-Kernel prediction model;

(c): Polynomial-Kernel prediction model; (d): Sigmoid-Kernel prediction model

光散射等因素对样品光谱的影响, 需要对光谱进行预处理以获取有效信息, 本研究采用标准归一化法(SNV)对光谱进行预处理。利用主成分分析法对不同病害程度的 200 个茶叶样品光谱矩阵进行解析, 如图 6 所示, 其中前 4 个主成分(principle component, PC)的累积贡献率达到 99%, 表明光谱共线性严重, 进行主成分分析十分必要。

由于主成分贡献率迅速递减, 建立模型时选取前 N 个主成分作为线性判别分析模型的输入数据相当有必要。留一交叉验证法被用于最优主成分的选取。其中主成分数据在建立模型前进行基于主成分的归一化处理。如表 3 所示, 本研究中, 选用前 10 个主成分进行建模时, 交叉验证准确率最高, 达 98%, 其中轻度病害和重度病害被误判的概率相对较大。可能由于病害的根本因素是叶片内的叶绿素含量产生的变化不能由肉眼直接分辨, 但其内部的变化还是相对较大造成的, 这可能是导致识别混淆的原因。

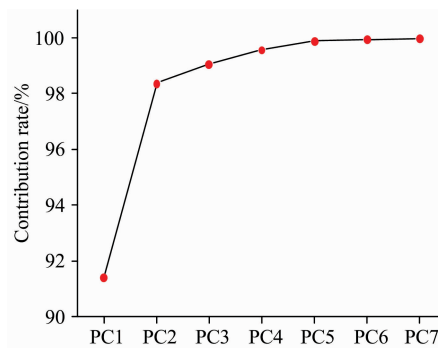


图 6 前 7 个主成分贡献率分布图

Fig. 6 Principal component contribution rate distribution

为建立模型并验证模型的预测性能, 对训练集提取的前 10 个主成分用 LDA 方法建立品种鉴别模型, 再用所得模型

表 3 前 10 个主成分建立的病害程度留一交叉验证模型结果
Table 3 The results of the cross-validation model for the degree of disease established by the first 10 principal components

| 类别 | 正常叶片 | 轻度病害 | 重度病害 | 正确识别率/% | |
|------|------|------|------|---------|-------|
| | | | | 组间 | 整体 |
| 正常叶片 | 70 | 0 | 0 | 100 | |
| 轻度病害 | 0 | 59 | 2 | 96.72 | 98.00 |
| 重度病害 | 0 | 2 | 67 | 97.10 | |

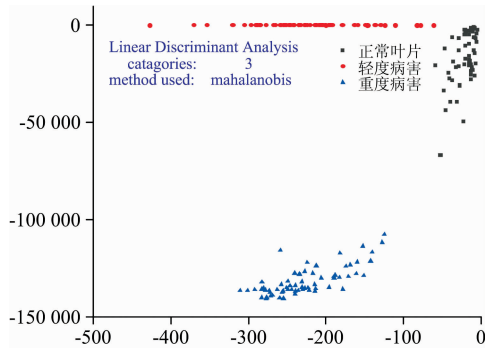


图 7 PCA-LDA 预测模型

Fig. 7 PCA-LDA prediction model

对验证集进行病害程度的外部验证, 预测效果见图 7, 正确识别率为 98.9%。说明经 PCA-LDA 模型预测效果最佳。

3 结 论

基于激光诱导叶绿素荧光光谱, 探究了一种对茶叶叶片病害的新检测技术, 研究了茶叶叶片受到病害胁迫其光谱的变化。即正常叶片与病害叶片在 600~900 nm 的波段上吸光度有明显的差异, 因为叶绿素含量的缺失是导致其产生的主要因素, 并且在荧光光谱结合化学计量学方法可以很好的区分正常、轻度病害和重度病害叶片, 采用多元散射校正, Savitzky-Golay 卷积平滑及标准化处理方法建立的支持向量机(SVM)、主成分线性判别模型(PCA-LDA)和最小二乘判别模型(PLS-DA)都能得到较好的结果: 通过模型对比得到 PLS-DA 建模集和预测集精度都达到 90% 以上; 在四种核函数建立的支持向量机模型中, 径向基核函数模型效果最好, 达到 95.72%; 经主成分分析后建立的线性判别模型(LDA)效果最好, 识别率为 98.9%。研究表明, 采用激光诱导叶绿素荧光光谱结合化学计量学方法可以较好的对茶藻斑病进行识别, 为茶叶病害识别提供了新思路。为以后遥感荧光技术结合机器学习的大面积识别应用提供了研究参考。

References

- [1] WANG Yong-chong(王永崇). Pesticide Market News(农药市场信息), 2019, (13): 66.
- [2] Xie C, He Y. Sensors, 2016, 16(5): 676.
- [3] Wetterich C B, Kumar R, Sankaran S, et al. Journal of Spectroscopy, 2013, 2013: 841738.
- [4] SUI Yuan-yuan, WANG Qing-yu, YU Hai-ye(隋媛媛, 王庆钰, 于海业). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(6): 1779.
- [5] WANG Ying-xu, SUN Ye, LI Yu-hua, et al(王迎旭, 孙 晔, 李玉花, 等). Journal of Nanjing Agricultural University(南京农业大学学报), 2020, 43(4): 770.
- [6] ZHOU Li-na, CHENG Shu-chao, YU Hai-ye, et al(周丽娜, 程树朝, 于海业, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2017, 48(2): 203.
- [7] JING Xia, LÜ Xiao-yan, ZHANG Chao, et al(竞 霞, 吕小艳, 张 超, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2020, 51(6): 191.
- [8] WAN Wen-bo, HUA Deng-xin, LE Jing, et al(万文博, 华灯鑫, 乐 静, 等). Acta Physica Sinica(物理学报), 2013, 62(19): 190601.
- [9] Lang M, Siffel P, Braunov Z, et al. Plant Biology, 1992, 105(6): 435.
- [10] Dong Y, Liu X, Mei L, et al. Journal of Food Engineering, 2014, 137: 95.
- [11] Kalaji H M, Oukarroum A, Alexandrov V, et al. Plant Physiol. Biochem., 2014, 81: 16.
- [12] Su W, Sun D. Talanta, 2016, 155: 347.

Research on Tea *Cephaleuros Virescens* Kunze Model Based on Chlorophyll Fluorescence Spectroscopy

LIU Yan-de, LIN Xiao-dong, GAO Hai-gen, WANG Shun, GAO Xue

School of Mechatronics & Vehicle Engineering, East China Jiaotong University, National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment, Nanchang 330013, China

Abstract Tea is an important cash crop in China. The early detection and diagnosis of tea diseases will help agricultural producers to take effective protective measures in time. In order to achieve accurate discrimination of tea diseases, the spectral characteristics of tea were studied using chlorophyll fluorescence spectrum. A total of 90 samples of healthy tea leaves, 90 samples of the early stage of *Cephaleuros virescens* Kunze leaf disease and 90 samples of the severe stage of *Cephaleuros virescens* Kunze leaf disease were collected in the experiment and were accordance with the Kennard-Stone algorithm divided into the training set and prediction set according to the proportion of 3 : 1 for each kind. Adopt the chlorophyll fluorescence spectrum collection system to collect the spectrum of tea leaf spot disease and normal leaves and set the collection parameters; integration time 20 ms and laser power 40 mW. The spectral response characteristics of the diseased and normal leaves were analyzed separately. In general, there are differences in the absorption intensity of the three types of leaves, and the spectrum trends are the same. There is a chlorophyll fluorescence peak near 685 and 740 nm. The difference is mainly reflected in the difference in fluorescence peak intensity. Then the polynomial smoothing (Savitzky-Golay) method was carried out for smoothing and noise reduction on the original spectral, the establishment of partial least squares discriminant model (PLS-DA), in the PLS-DA modeling set model, the number of misjudged samples is 3, the false positive rate is 3%; in the PLS-DA prediction set model, the number of false positive samples is 5, and the false positive rate is 7.1%. Then the support vector machine model established by 4 different kernel functions is compared. RBF is used as the kernel function. The SVM model established by PCA has the lowest misjudgment rate, and the accuracy rate reaches 95.72%. Finally, the model established by principal component analysis (PCA) and linear discriminant analysis (LDA) has the best effect, and the accuracy rate reaches 98.9%. The selection of the optimal number of principal components is obtained by the leave-one-out verification method. When the first 10 principal components are selected for modeling, the cross-validation accuracy rate is the highest, reaching 98%. Through model comparison, the accuracy of the PLS-DA modeling set and prediction set is more than 90%. Among the support vector machine models built with four kernel functions, the radial basis kernel function model is the best, reaching 95.72%, the linear discriminant model (LDA) established after principal component analysis has the best effect, and the recognition rate is 98.9%. This study uses chlorophyll fluorescence spectroscopy combined with chemometrics to identify tea diseases, providing a new method for rapid and accurate prediction of tea diseases.

Keywords Fluorescence Spectroscopy; Principal component analysis; Partial least squares discrimination; Support vector machines; Linear discriminant analysis; *Cephaleuros virescens* Kunze

(Received Jul. 6, 2020; accepted Nov. 21, 2020)