

文章编号:1001-9014(2007)04-0269-05

基于可见/近红外光谱技术的茄子叶片 灰霉病早期检测研究

吴迪¹, 冯雷¹, 张传清², 何勇¹

(1. 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029;

2. 浙江大学农业与生物技术学院, 浙江 杭州 310029)

摘要:应用可见/近红外光谱技术对茄子叶片进行灰霉病害还未在叶片表面出现病症时的早期检测. 采用化学计量学方法建立早期检测模型. 主成分分析用于对光谱数据进行降维, 得到若干个最重要的主成分. 但直接从聚类图中无法进行是否染病的鉴别. 因而将其作为变量输入BP神经网络, 从而减少了计算量, 提高了建模精度. 检测结果显示, 模型具有良好的检测效果, 能够达到100%的识别率, 正确率也能达到88%. 说明运用可见/近红外光谱技术能够实现当病症还未在叶片表面出现时的快速准确的早期检测, 为灰霉病早期检测提供了新的途径.

关键词:可见/近红外光谱; 灰霉病; 主成分分析; BP神经网络

中图分类号: S126 文献标识码: A

EARLY DETECTION OF GRAY MOLD (CINEREA) ON EGGPLANT LEAVES BASED ON VIS/NEAR INFRARED SPECTRA

WU Di¹, FENG Lei¹, ZHANG Chuan-Qing², HE Yong¹

(1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China;

2. College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, China)

Abstract: Visible and near-infrared reflectance spectroscopy (Vis/NIRS) technique was applied in the early detection of grey mold (cinerea) on eggplant leaves while the symptom had not appeared. Chemometrics was used to build the early detection model. In order to decrease the amount of calculation and improving the accuracy, principle component analysis (PCA) was executed to reduce numerous wavebands into several principle components (PCs) as input variables of BPNNS while the PCs plot of three primary PCs was failed. The performance of the BPNNS model is good with 100% recognition rate and 88% correct rate. Thus, it is concluded that the spectra technology is an available one for the early detection of grey mold on eggplant leaves while the symptom has not appeared and it provides a new method for the early detection of grey mold.

Key words: Vis/near infrared spectroscopy; grey mold (Cinerea); principal component analysis (PCA); BP neural networks (BPNNS)

引言

植物病害对于农作物生长的危害十分严重. 其不仅能够导致农产品产量和质量的下降, 同时会引起农药等药物的大量投入以及防治费用的上升, 从而增加了农业生产成本并造成严重的环境污染. 在我国每年造成近百亿元的经济损失. 因此, 植物病害的研究对于我国乃至全世界农业经济的发展具有十

分重要的意义.

由 *Botrytis cinerea* 真菌引起的灰霉病对于众多作物, 尤其是绿色大棚作物影响严重. 大棚内的气温较低、湿度较高, 对于病害的流行十分有利^[1]. 病菌主要通过种植过程或者收获后运输中造成的叶、茎、花、果实、种子等的伤口感染进行传播^[1]. 灰霉病为茄子主要病害之一, 因其发病速度快, 危害严重, 对茄子的大面积生产影响严重. 近几年, 随着大

收稿日期: 2006-08-15, 修回日期: 2007-02-09

Received date: 2006-08-15, revised date: 2007-02-09

基金项目: 国家自然科学基金(60605011)、国家十一五科技支撑计划项目(2006BAD10A04)、高等学校博士学科点专项科研基金课题(20040335034)和浙江省重大科技攻关(2005C12029)资助项目

作者简介: 吴迪(1984-), 男, 浙江杭州人, 浙江大学生物系统工程与食品科学学院博士生, 主要从事数字农业和多光谱检测技术研究.

棚栽培面积的增加,且棚内适宜的环境导致了灰霉病危害的加重,从而造成 20% 到 40% 的减产,甚至绝收. 现今灰霉病的测报预防技术还是一个薄弱的环节,主要手段也只是当病害发生后采用化学药剂防治. 由于频繁施药,结果造成病菌抗药性增强,农药污染问题严重.

现有的病害检测方法在实际农业生产中都存在一定的缺点. 肉眼观测存在着疲劳、情绪等主观因素的干扰,且只能在病症显现时进行识别;基于分子水平的检测方法,如生物测定技术、核酸序列分析技术、分子标记技术等^[2,3],虽然较为准确,但不能实时在田间操作,且需要专业人员操作,耗时,成本较高;统计学方法^[4]通过以往的病害发生数据进行一定预测的. 但是在精度方面不够高,缺少实效性,并且只适用于在一个很大的地域范围内进行粗略估计,很难对小区甚至单株植物进行. 植物学家也指出在植株出现病症了之前就能检测出染病,从而进行药物处理,效果较好. 鉴于现有的检测方法很难实现快速准确的在线早期检测,因此需要应用新的技术方法来进行病害的早期检测.

可见/近红外光谱(Visible/Near Infrared Reflectance Spectroscopy, 缩写 Vis/NIRS)分析技术作为一种快速简便且非破坏性的分析技术. 其通过采集近红外谱区包含的物质内部信息以及可见光波段包含的物质表面信息,来进行物质特性的定性和定量分析. 可见/近红外光谱技术具有快速、低成本、测量方便、非破坏性等优点,被广泛地用于各个领域^[5-8]. 作物在染病后,其内部生理结构、养素含量等先于外部形态特征,如枯叶、糜烂等发生变化,因而植株染病部分内部的光谱特性也先于外部光谱特征发生变化^[9,10]. Bravo 等人^[11]在可见光到近红外波段对小麦黄锈病进行了早期诊断. Muir 等人^[12]研究了马铃薯块茎未出现病症时的光谱反射率,以实现早期检测.

本研究对以灰霉病为主要病害的茄子作为研究对象. 利用叶片的光谱信息,通过化学计量学的方法,进行未出现病症情况下的早期检测,为灰霉病害的早期诊断提供一种新的方法.

1 材料与方 法

1.1 仪器设备与样品来源

实验使用美国 ASD (Analytical Spectral Device) 公司的便携式光谱仪 (Handheld FieldSpec), 其采样间隔 (波段宽) 1.5nm, 测定范围 325 ~ 1075nm, 分辨率 3.5nm, 探头视场角为 10°. 实验安排在室内进

行,以减少外界环境对光谱采集的干扰,实验光照采用 150W 的卤素灯模拟自然光光源,减少了因为阳光随着时间变化所造成的影响. 分析软件为 ASD ViewSpec Pro、Unscrambler 和 Data procession system (DPS).

选用茄子进行实验. 一共分 8 个品种,由浙江大学农业与生物技术学院农药与环境毒理研究所提供. 每个品种分别有 3 盆植株,每盆一株茄子. 对其中两盆植株的叶片进行了灰霉病菌的接种,另外一盆则是健康的植株,作为对照样本.

1.2 叶片样本选择与光谱数据采集

接种后一段时间,叶片样本表面情况较为复杂,存在包括灰尘、破损、霉变、腐烂、虫害等现象. 考虑到这些因素的存在可能会对数据分析和病害预测结果产生干扰,我们仅选用表面只有感染灰霉病并且已经出现明显病症以及完全健康两类叶片样本作为建模样本,而属于接种植株但表面无病症及其他现象的样本叶片则作为预测样本,一张叶片作为一个样本,共计 99 个样本.

在对光谱仪进行反射率为 100% 的标准白板光谱校正后,用光谱仪探头逐个对准每片叶片的中心进行光谱采集. 探头与叶片表面的距离控制在 15cm 左右,使光谱采集到的面积集中在叶片上,以减少背景的干扰. 每个样本重复扫描 3 遍,每遍扫描次数 30 次,并通过计算机自动保存. 整个光谱测量过程中,光谱仪、卤素灯以及样本放置的位置固定.

1.3 光谱数据预处理与化学计量学方法

采集到的光谱信号中不仅有反映样本特征的信息,同时还包含高频随机噪声、基线漂移、样本不均匀、光散射等干扰. 所以在对光谱数据进行建模之前,需要通过一定的预处理方法来减少干扰对建模的影响. 通过比较平均平滑法、一阶求导以及多元散射校正等方法的建模效果,最终选择采用平均平滑法,窗口大小选用 3. 由于系统误差样本光谱曲线在首端和末端有较大噪音 (图 1). 所以只取 400 ~ 900nm 的波段用于分析. 整个预处理程序由 Matlab 编写.

由于叶片不是纯物质,叶片光谱检测又会有噪声背景并存在统计波动性,同时近红外光谱信息强度低,在各个位置通常又有多个谱峰的重叠并且信号较弱,全部波段内的输入反射率值有 500 个,数据量太大,影响检测建模的速率和精度,因此需要从复杂、重叠的背景中提取信息. 主成分分析是一种常用的数据降维压缩的方法^[13]. 通过主成分分析降维得

到最重要的前几个主成分来代替原有的大量变量,并且挖掘光谱中对建模有关的信息.主成分分析程序采用 Matlab 编写.

人工神经网络是一种模拟人脑的信息处理系统,它主要借鉴了人脑神经系统处理信息的过程,以数学网络拓扑结构为理论基础.因其突出的非线性映照能力,倍受关注.目前最常用的是反向传输(Back Propagation, BP)模型.本研究中应用 DPS 软件建立了一个三层 BP 神经网络.通过对设定的学习样本部分进行建模,通过调整网络结点系数,反复地迭代演算,使得最终输出值的拟合残差达到预先设定的范围之内.然后通过建立好的模型对另一部分样本进行预测.本研究中,神经网络输入层节点的值来自主成分分析得到的主成分值,用以代替大量数据的原始光谱反射率值.

2 实验结果与分析

2.1 光谱图谱分析

图 1 为在同一时期测得的表面只有感染灰霉病并且已经出现明显病症以及完全健康两类叶片样本的光谱反射率曲线,分别随机选择了 2 个样本.横坐标为波长范围,纵坐标为光谱反射率.从图 1 中可以看出,出现病症的叶片光谱反射率曲线和健康叶片的相比有一定的差别,但不明显.总体上健康的叶片光谱反射率比出现病症的要略高.但是当样本数增多时两类样本的反射率曲线则会混合在一起.原因分析为影响反射率值的因素很多,不仅只是由于病害的感染引起的,相反如叶片的生长程度、各种营养成分的含量、种植环境等.这些因素都会对叶片的表明颜色、纹理等外部特征以及叶绿素含量、含水率等内部特征产生影响,从而造成反射率曲线并不反全因为染病程度而规律性分布.因此,仅仅通过光谱反射率曲线直观地进行叶片是否染病的检测是不可行的,需要通过进一步的化学计量学方法挖掘光谱反射率值中的信息.

2.2 主成分提取与聚类图分析

对叶片样本的反射率值进行主成分提取.前 8 个主成分的贡献率和累计贡献率如表 1 所列.从表 1 分析可得,主成分 1 的累计贡献率已达 96%,而前 8 个主成分的累计贡献率更是接近了 1,说明最主要的几个主成分能够很好地代表原始 500 个变量的信息.

通过选取前 3 个主成分,得到出现病症和健康共 65 个样本的三维聚类图(X 轴表示主成分 1、Y

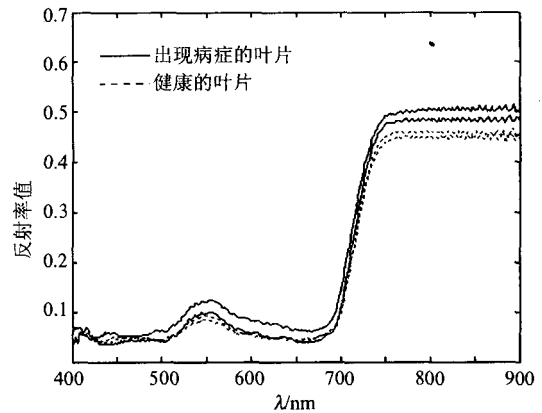


图 1 出现病症以及健康叶片的光谱反射率曲线
Fig. 1 Reflectance spectra curves of infected leaves with symptom and healthy one

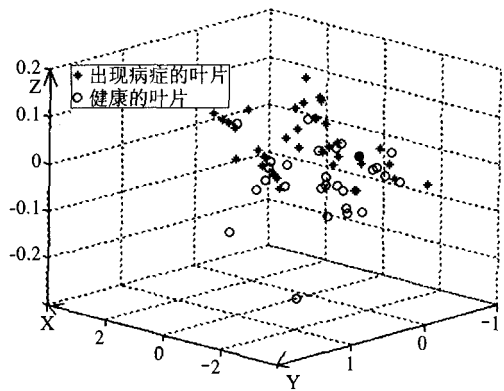


图 2 基于前 3 个主成分的三维聚类图
Fig. 2 Clustering plot of the PCA model with the first three primary PCs

表 1 前 8 个主成分的累计贡献率

Table 1 PCs and reliabilities

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
贡献率(%)	95.829	3.660	0.231	0.082	0.043	0.021	0.018	0.016
累计贡献率(%)	95.829	99.489	99.720	99.802	99.844	99.866	99.883	99.899

轴表示主成分 2、Z 轴表示主成分 3). 从图 2 中可以看到,出现病症和健康叶片的聚类样本点虽然有区别,但很难在图上区分.此外,聚类图只能在 3 个维度范围之内进行直观、定性分析,难以定量地精确区分及预测.因此,需要在主成分分析的基础上采用 BP 神经网络进行分析聚类.

2.3 茄子叶片灰霉病感染早期检测模型的建立

应用得到的主成分代替原有光谱变量进行神经网络建模.根据计算得到的主成分贡献率(表 1)分别选择 4~8 个主成分,作为人工神经网络输入结点,样本类型(出现病症和健康叶片)作为人工神经网络输出结点数值,选择表面只有感染灰霉病并且

已经出现明显病症以及完全健康两类叶片共 65 个样本作为建模样本,而属于病毒接种植株但表面无感染以及其他症状共 34 个样本叶片作为预测样本,进行人工神经网络建模.隐含层结点数参照主成分数并依据式(1)分别定为 5 到 11.

$$n_1 = \sqrt{n + m} + a, \quad (1)$$

式(1)中 n_1 为隐含结点数, n 为输入神经元个数, m 为输出神经元个数, a 为 1 到 10 之间的整数.用人工神经网络建模的样本有两类:已经出现明显病症以及完全健康的叶片,对应的输出结果也只有 2 种,因此输出结点数设置为 2 个,具体设置为:染病(10),健康(01).

此外模型最小训练速度为 0.1,数据标准化转换, Sigmoid 参数为 0.9,动态参数为 0.6,允许误差为 0.001.在光谱测量结束之后,对预测的 34 个样本所在的植株进行 10 天的培养,然后进行评估,判断是否染病,结果作为模型检测正确率的判断依据.通过比对输出值是否在阈值之内作为输出结果能够用作染病判别的依据.当输出值在阈值范围之外则为模型无法判断的样本.基于不同主成分和人工神经网络参数模型的建模结果如表 2 所列.

正确率为判断正确的个数除以样本总数,识别率为预测值在阈值范围内的样本个数除以样本总数.好的建模效果需要有高的正确率和识别率.当主成分为 4 的时候,无论从正确率还是识别率都不理想.原因分析,主要是因为主成分数量太少,虽然已经有了 99% 的信息含量,但是从结果可以看出对于病害的早期检测还是过少.当主成分为 7 和 8 的时候,效果也不够理想.分析原因,是因为主成分数过

多造成的.虽然选用主成分个数越多可以包含更多的光谱信息,但同时也含有了更多对模型建立不利的噪声.而当主成分数为 5 和 6 的时候,则模型效果较好.在分析隐含层结点数选择时可以看到,对于主成分数为 5 时,个数为 7、9 的效果较好,而对于主成分数为 6 时,个数为 5、11 时较好.阈值选择方面,较小的阈值得到的正确判断个数上会变少但判别值更加接近 1,因而比较稳定且具有说服力,但会导致识别率的下降.在本研究中,阈值变大,则正确率和识别率上升.也就是说较大的阈值,虽然判别值相对于 1 比较离散,但是可以得到更好的判别效果.综合分析,模型能够达到 100% 的识别率以及 88% 正确率.在参数选择方面,推荐较为适中的主成分个数以及较高的阈值,而隐含层结点数则需要在公式推荐的基础上尝试而定.

此外由于叶片的染病过程复杂,自身及生长环境都会发生一定的变化.如健康的叶子,在培养后发现感染,则会出现在光谱采集之前已经感染和在采集之后感染两种情况.同样在培养后依然健康的叶子,可能是一直没有感染,也有可能是光谱测量之前感染但在培养过程中自我康复.虽然上述的情况发生的概率较小,但会对模型好坏的判别产生影响.

3 结语

本文提出了应用可见/近红外光谱技术对茄子叶片灰霉病进行早期检测.采用主成分分析结合 BP 神经网络的方法,解决了光谱反射率值不适合直接应用于神经网络建模以及仅应用主成分分析鉴别能力不足的缺点,提高了模型计算的速度和精度.检测

表 2 基于不同主成分数和人工神经网络隐含层结点数的模型检测结果

Table 2 The performance of models with different parameters

n_1	阈值	主成分个数									
		8		7		6		5		4	
		C	R	C	R	C	R	C	R	C	R
5	± 0.1	71%	85%	76%	97%	82%	94%	74%	85%	47%	50%
	± 0.25	76%	91%	76%	100%	85%	97%	76%	88%	56%	65%
	± 0.4	79%	100%	76%	100%	88%	100%	82%	94%	59%	74%
7	± 0.1	65%	79%	65%	79%	71%	79%	79%	91%	53%	59%
	± 0.25	76%	94%	68%	88%	76%	85%	88%	100%	53%	62%
	± 0.4	76%	94%	71%	91%	82%	94%	88%	100%	62%	79%
9	± 0.1	62%	76%	74%	82%	68%	79%	74%	85%	50%	56%
	± 0.25	71%	88%	74%	91%	74%	85%	79%	91%	56%	65%
	± 0.4	76%	94%	76%	100%	82%	94%	88%	100%	65%	79%
11	± 0.1	68%	88%	76%	85%	85%	94%	68%	79%	53%	59%
	± 0.25	74%	94%	88%	97%	85%	97%	76%	88%	56%	62%
	± 0.4	74%	97%	88%	100%	88%	100%	82%	94%	68%	85%

注: n_1 为隐含层节点数, C 表示正确率, R 表示识别率

模型具有良好的检测效果,达到 100% 的识别率以及 88% 的正确率.说明可见/近红外光谱技术能够在病症还未在叶表出现时的早期检测,为今后光谱技术应用于植株病害早期检测提供了新的途径.

REFERENCES

- [1] O' Neill T M, Shtienberg D, Elad Y. Effect of some host and microclimate factors on infection of tomato stems by *Botrytis cinerea* [J]. *Plant Disease*, 1997, **81**(1):36—40.
- [2] Putnam M L. Evaluation of selected methods of plant disease diagnosis [J]. *Crop protection*, 1995 **14**(6):517—525.
- [3] Fox R T. The present and future use of technology to detect plant pathogens to guide disease control in sustainable farming systems [J]. *Agriculture, Ecosystems and Environment*, 1997, **64**(2):125—132.
- [4] XIONG Xue-Mei, JI Chang-Ying, Claudio M. Parametric fuzzy neural network based on genetic algorithm configured for plant disease prediction [J]. *Transactions of the Chinese society for agricultural machinery*(熊雪梅,姬长英,Claudio Moraga. 基于参数化遗传神经网络的植物病害预测方法. *农业机械学报*), 2004, **35**(6):110—114.
- [5] Gómez H A, HE Yong, Pereira G A. Non-Destructive measurement of acidity, soluble solids and firmness of satsuma mandarin using Vis/NIR-Spectroscopy techniques [J]. *Journal of Food Engineering*, 2006, **77**(2):313—319.
- [6] HE Yong, LI Xiao-Li. Discriminating varieties of waxberry using near infrared spectra [J]. *J. Infrared Millim. Waves*(何勇,李晓丽.用近红外光谱鉴别杨梅品种的研究. *红外与毫米波学报*), 2006, **25**(3):192—194.
- [7] Esteban-Diez I, Gonzalez-Saiz J M, Pizarro C. An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS [J]. *Analytica Chimica Acta*, 2004, **514**(1):57—67.
- [8] HUANG Min, HE Yong, HUANG Lin-Xia, et al. Discrimination of varieties of silkworm egg based on visible-near infrared spectra [J]. *J. Infrared Millim. Waves*(黄敏,何勇,黄凌霞,等.基于可见-近红外光谱技术的家蚕蚕种鉴别方法的研究. *红外与毫米波学报*), 2006, **25**(5):342—344, 359.
- [9] Kokaly R F, Clark R N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple liner regression [J]. *Remote Sensing of Environment*, 1999, **67**(3):267—287.
- [10] Rencet A N, Nemeth J. Defection of mountain pine beetle infection using landsat MSS and simulated thematic mapper data. canadian [J]. *Journal of Remote Sensing*, 1985, **11**(1):50—58.
- [11] Bravo C, Moshou D. Early disease detection in wheat fields using spectral reflectance [J]. *Biosystems Engineering*, 2003, **84**(22):37—145.
- [12] Muir Y, Porteous R L, Wastie R L. Experiments in the detection of incipient diseases in potato tubers by optical methods [J]. *Journal of Agricultural Engineering Research*, 1982, **27**(2):131—138.
- [13] LI Zhi-Yong, KUANG Gang-Yao, YU Wen-Xian, et al. Algorithm on small target detection base on principal component of hyperspectral imagery [J]. *J. Infrared Millim. Waves*(李智勇,匡纲要,郁文贤,等.基于高光谱图像主成分分量的小目标检测算法研究. *红外与毫米波学报*), 2004, **23**(4):286—290.