

基于高光谱技术的灰霉病胁迫下番茄叶片 SPAD 值检测方法研究

谢传奇, 何勇, 李晓丽, 刘飞, 杜朋朋, 冯雷*

浙江大学生物系统工程与食品科学学院, 浙江 杭州 310058

摘要 对灰霉病胁迫下番茄叶片中叶绿素含量 (SPAD) 的高光谱图像信息进行了研究。首先获取 380~1 030 nm 波段范围内健康和染病番茄叶片的高光谱图像, 然后基于 ENVI 软件处理平台提取高光谱图像中感兴趣区域的光谱信息, 经平滑 (Smoothing)、标准化 (Normalize) 等预处理后, 建立了基于 Normalize 预处理的偏最小二乘回归 (PLSR) 和主成分回归 (PCR) 模型。再基于 PLSR 获得的 4 个变量建立反向传播神经网络 (BPNN) 和最小二乘-支持向量机 (LS-SVM) 模型。4 个模型中, LS-SVM 的预测效果最好, 其决定系数 R^2 为 0.901 8, 预测集均方根误差 RMSEP 为 2.599 2。结果表明, 基于健康和染病番茄叶片的高光谱图像响应特性检测叶绿素含量 (SPAD) 是可行的。

关键词 高光谱图像; 叶绿素; 偏最小二乘回归; 主成分回归; BP 神经网络; 最小二乘-支持向量机; 番茄; 灰霉病

中图分类号: TP391; S432.2+2 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2012)12-3324-05

引言

叶绿素在作物生长过程中起着重要的作用, 当作物遭受病害侵染后, 叶片中叶绿素含量会降低^[1]。因此, 可以通过对叶绿素含量的测量来监测作物的生长状况, 从而及时采取有效措施, 使病害带来的损失最小。传统检测作物中叶绿素含量的方法主要有分光光度计法^[2]、高效液相色谱法^[3]、原子吸收法^[4]等。但这些方法步骤繁琐、具有破坏性, 且无法实现在线检测等。近年来, 基于可见-近红外光谱技术检测作物中化学指标的研究已有报道, Liu 等^[5]利用可见-近红外光谱技术检测油菜叶片中乙酰乳酸合酶的活性和蛋白质的含量, Min 等^[6]利用可见-近红外光谱技术检测大白菜中氮含量, Ulissi 等^[7]利用可见-近红外光谱技术检测番茄中氮含量。但对病害胁迫下作物中化学指标的检测还少有报道, 尤其是利用高光谱技术检测病害胁迫下作物中化学指标的研究更是少见。

传统的光谱技术只能提供被检测对象很小的区域且无法提供被检测对象的图像特征, 因此具有一定的局限性。高光谱图像技术突破了这方面的缺点, 它能够同时提供被检测对象的光谱信息和图像信息, 并能够通过图像获取感兴趣区域

的光谱数据。目前在农业中已广泛采用 (soil and plant analyzer development, SPAD) 值来反应叶绿素的含量^[8, 9], 但 SPAD-502 叶绿素仪只能对叶片进行逐点检测、工作量大、无法实现在线检测等。通过高光谱采集系统获取的数据所建立的模型可对叶片进行实时在线检测, 且效率高、无损、便于大规律作业。研究利用高光谱图像技术, 通过人工诱发番茄叶片感染灰霉病并测量染病和健康番茄叶片的 SPAD 值, 然后结合化学计量学方法研究灰霉病胁迫下番茄叶片的 SPAD 值与高光谱信息的响应特性。

1 实验部分

1.1 仪器和分析软件

高光谱图像采集系统 (图 1) 主要包括成像光谱仪 V10E-QE (SPECIM, Finland)、镜头 OLE-23、线光源 Schott DCR III、移动平台控制装置 SC300A、电控位移台 PSA 200-11-X、计算机等, 波长范围是 380~1 030 nm。SPAD 值的测量使用 SPAD-502 叶绿素仪 (Minlita Camera Co., 1989)。数据分析软件采用 ENVI4.6 (Research system Inc, Boulder, Co. USA)、Unscrambler9.7 (Camo, ProcessAS, Oslo, Norway)、DPS (Data Procession System), Matlab2009 (The

收稿日期: 2012-04-06, 修订日期: 2012-06-20

基金项目: 浙江省重大科技专项重点农业项目 (2009C12002), 国家 (863 计划) 课题项目 (2011AA100705), 国家自然科学基金项目 (31071332) 和中央高校基本科研业务费专项资金项目 (2012FZA6005) 资助

作者简介: 谢传奇, 1986 年生, 浙江大学生物系统工程与食品科学学院博士研究生 e-mail: cqxie@zju.edu.cn

* 通讯联系人 e-mail: yhe@zju.edu.cn

Math Works, Natick, USA)。

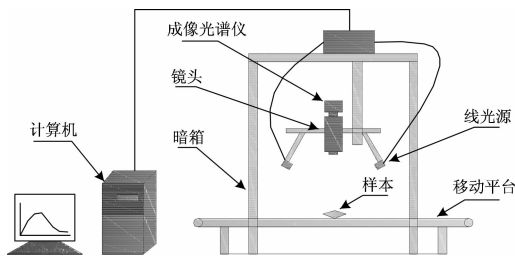


Fig. 1 Hyperspectral imaging system

1.2 样本

取 50 株番茄苗接种灰霉病菌, 每棵番茄接种 3 个叶片, 并放在自制的小区内培养, 同时另取 30 株健康番茄植株在同样的环境、不同的小区内培养。24 小时后, 接种的番茄叶片出现灰霉病病斑。最后选取 184 个叶片(健康 60, 染病 124)进行进一步研究。

1.3 高光谱图像采集与校正

由于高光谱摄像头中存在暗电流及不均匀光强产生有噪声, 因此在图像采集前还需对高光谱仪进行黑白板校正^[10]。在同样的采集环境下, 采集标准白色板得到全白的标定图像 W_λ (反射率接近 100%), 盖上镜头盖采集全黑的标定图像 B_λ (反射率接近 0%)。然后将番茄叶片平铺在移动平台上, 采集原始高光谱图像 I_λ , 最后按照式(1)对采集到的原始图像进行校正, 得到校正图像 R_λ 。

$$R_\lambda = \frac{I_\lambda - B_\lambda}{W_\lambda - B_\lambda}$$

1.4 SPAD 值测定

实验采用 SPAD-502 叶绿素仪测量番茄叶片的 SPAD 值, SPAD 值的大小反映了植物叶绿素含量的高低, 其原理是通过测量叶片在两种波长范围内的透光系数来确定叶片叶绿素含量。但 SPAD-502 叶绿素仪只能对叶片进行逐点检测, 效率较低、费时费力, 因此, 需要研究一种快速检测作物中叶绿素含量的方法。每个叶片采集完高光谱图像后, 立即用 SPAD-502 叶绿素仪测量病斑位置共 5 个点的 SPAD 值, 健康叶片采取同样的测量方法, 最后取平均值, SPAD 值见表 1。

Table 1 Statistical values of SPAD in healthy and infected leaves of tomato

Type	Number	Range	Mean	Standard deviation
Healthy	60	31.20~41.90	36.39	2.41
Infected	124	8.18~32.56	22.46	5.55
Total	184	8.18~41.90	27.00	8.09

1.5 数据分析处理

1.5.1 高光谱数据获取

基于 ENVI4.6 软件处理平台选取校正后的高光谱图像的病斑位置及健康叶片相应位置为感兴趣区域(region of interest, ROI), 算出其像素光谱反射率的平均值, 得到高光谱数据。

将 184 个染病和健康样本依次排列, 按照 2:1 的比例, 最终选取 122 个(染病 83、健康 39)样本用来建模, 62 个(染病 41、健康 21)样本用来预测, 建模集和预测集样本的 SPAD 值见表 2。

Table 2 Statistical values of SPAD in tomato leaves of calibration and prediction

Type	Number	Range	Mean	Standard deviation
Calibration	122	8.08~40.04	26.64	7.99
Prediction	62	11.36~41.90	27.71	8.30

1.5.2 高光谱数据预处理

为了消除原始光谱数据中高频随机噪声、基线漂移等干扰因素对所建模型的影响, 研究采用了 Smoothing、Normalize 等多种预处理方法。把各种预处理后的光谱数据作为自变量 X , SPAD 值作为 Y 变量, 建立 SPAD 值的偏最小二乘回归(partial least squares regression, PLSR)模型, 通过比较各个模型的预测效果从而对各种预处理方法进行评价, 最后确定基于 Normalize 的预处理方法并建立相应模型。

1.5.3 偏最小二乘回归, 主成分回归, BP 神经网络和最小二乘支持向量机

偏最小二乘回归的特点是在描述光谱矩阵 X 变量的同时也描述了浓度矩阵 Y 变量, 用矩阵 Y 变量的列参与矩阵 X 因子计算的同时, 矩阵 X 变量的列也参与矩阵 Y 因子的计算^[11]。目前 PLSR 在光谱领域里得到了广泛的应用^[12, 13]。

主成分回归(principal component regression, PCR)采用主成分分析法对样本光谱矩阵 X 进行分解, 再选取其中的主成分进行多元线性回归分析, 从而有效地克服输入变量之间的共线性问题^[14]。PCR 技术在光谱学分析中也得到了具体应用^[15, 16]。

反向传播神经网络(back propagation neural network, BPNN)是由非线性变换神经元组成的一种前馈型神经网络, 神经元的传递函数采用 Sigmoid 型函数, 可以实现输入和输出的任意非线性映射, 其结构简单、可塑性强, 可以减少光谱数据的重复性和冗余性, 具有强大的非线性预测能力^[17]。目前 BP 神经网络在光谱学分析中也得到了广泛的应用^[18-20]。

最小二乘支持向量机(least squares-support vector machine, LS-SVM)由 Suykens 等^[21]首次提出, 是对传统支持向量机的一种改进, 能够处理线性和非线性多元变量建模问题, 同时也能较快地解决多元变量的复杂性问题。LS-SVM 通过求解一系列线性方程组代替二次方程式来获得支持向量机, 提高了运算效率^[22]。LS-SVM 在光谱定性和定量分析中也得到了一些应用^[23, 24]。

以上模型效果的评价指标是预测集决定系数 R^2 (Determination Coefficient) 越大、建模集均方根误差 (root mean square error of calibration, RMSEC) 和预测集均方根误差 (root mean square error of prediction, RMSEP) 值越小且两者之间的差别越小, 则模型的预测能力越强, 效果越好^[25]。

2 结果与讨论

2.1 高光谱反射曲线分析

图 2 中(a)和(b)分别是健康和染病番茄叶片的光谱反射率曲线,图 3 是健康和染病番茄叶片光谱数值分别取平均值后得到的光谱反射率曲线,横坐标是波长,纵坐标是光谱反射率。通过图 2 可以看出健康和染病样本的光谱曲线轮廓相似,没有明显区别,在 550 和 750 nm 附近各有一个峰值,680~750 nm 光谱反射率明显增强,可见光范围内的光谱反射率低于近红外区域。但通过图 3 可以明显看出健康和染病样本在 400~1 000 nm 波段范围内存在明显差别,400~720 nm 波段范围内染病样本的光谱反射率高于健康样本,720~1 000 nm 波段范围内染病样本的光谱反射率低于健康样本,

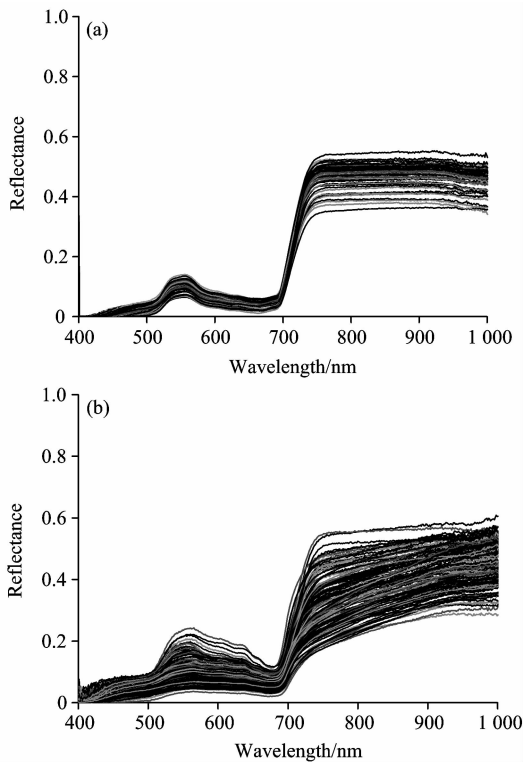


Fig. 2 Reflectance spectra curves of tomato leaves

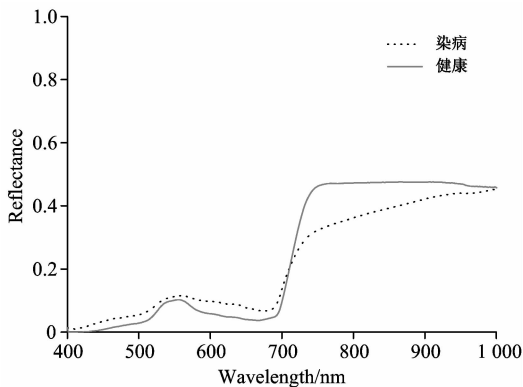


Fig. 3 Reflectance spectra curves of healthy and infected tomato leaves

在 720 nm 附近有一个交叉。这可能是因为叶片染病初期,基于植物的自我保护作用,植物的抗性增强,叶片内一些化学指标变得更加活跃,染病样本的光谱反射率也随之增大,这种现象在可见光范围内表现的较明显;而随着灰霉菌菌的侵入,其组织细胞遭到破坏,水分和叶绿素含量下降,导致光谱反射率低于健康样本,此时,这种现象在短波近红外区域表现的较明显。

由于原始高光谱在首尾波段范围内存在较多噪声,因此,实验选取 400~1 000 nm 波段范围内的光谱信息进行研究,并进一步结合化学计量学方法挖掘光谱数据中与番茄叶片 SPAD 值相关的特征信息。

2.2 高光谱数据预处理与 PLSR 模型

健康和染病番茄叶片的高光谱数据经过预处理后,建立光谱矩阵 X 与 Y 变量 SPAD 值之间的 PLSR 模型,通过不同的预处理方法所建的 PLSR 模型的预测效果不同(见表 3)。通过表 3 可以发现基于 Normalize 预处理方法建立的 PLSR 模型最优,此时预测集决定系数 R^2 最大,预测集均方根误差 RMSEP 最小, R^2 是 0.895 6, RMSEP 是 2.664 2。PLSR 模型预测值的散点图分布见图 4(a)。

Table 3 Prediction results of PLSR models with different pretreatment

Model	Pretreatment	R^2	RMSEP	Number of factors	Explained variance/%
1	Raw	0.881 5	2.838 4	5	82.467 1
2	S. M. A-5	0.882 1	2.826 7	5	82.395 6
3	Normalize	0.895 6	2.664 2	4	84.464 6
4	MSC/EMSC	0.891 9	2.724 0	3	80.172 1
5	D. G. S	0.583 2	6.202 2	4	67.724 1
6	D. S. G	0.823 9	3.483 4	2	77.661 8
7	Baseline	0.878 8	2.862 5	5	82.281 8
8	SNV	0.892 6	2.716 1	3	80.164 9

2.3 PCR, BPNN 和 LS-SVM 模型

再建立基于 Normalize 预处理后的光谱信息与 SPAD 值之间的 PCR 模型,此模型的预测集决定系数 R^2 是 0.888 3,预测集均方根误差 RMSEP 是 2.744 6,预测集散点分布图如图 4(b)。由于 PLSR 和 PCR 模型只能解释光谱数据的线性关系,无法把非线性关系纳入预测模型中,因此,研究进一步采用 BP 神经网络和 LS-SVM 算法来提高模型的预测能力。由于 Normalize 预处理方法建立的 PLSR 模型中推荐的前 4 个隐含变量是 SPAD 值的特征隐含变量,前 4 个隐含变量的累积贡献率达到最大,为 84.464 6%,故把这 4 个隐含变量作为 BP 神经网络的输入变量建立 BBNN 预测模型,此模型的预测集决定系数 R^2 是 0.893 5,预测集均方根误差 RMSEP 是 2.685 4,见图 4(c)。再用 PLSR 模型推荐的这 4 个隐含变量建立 LS-SVM 模型,LS-SVM 模型的预测集决定系数 R^2 为 0.901 8,预测集均方根误差 RMSEP 为 2.599 2,见图 4(d)。

2.4 PLSR, PCR, BPNN 和 LS-SVM 模型比较

上述分析表明利用高光谱技术检测番茄叶片 SPAD 值

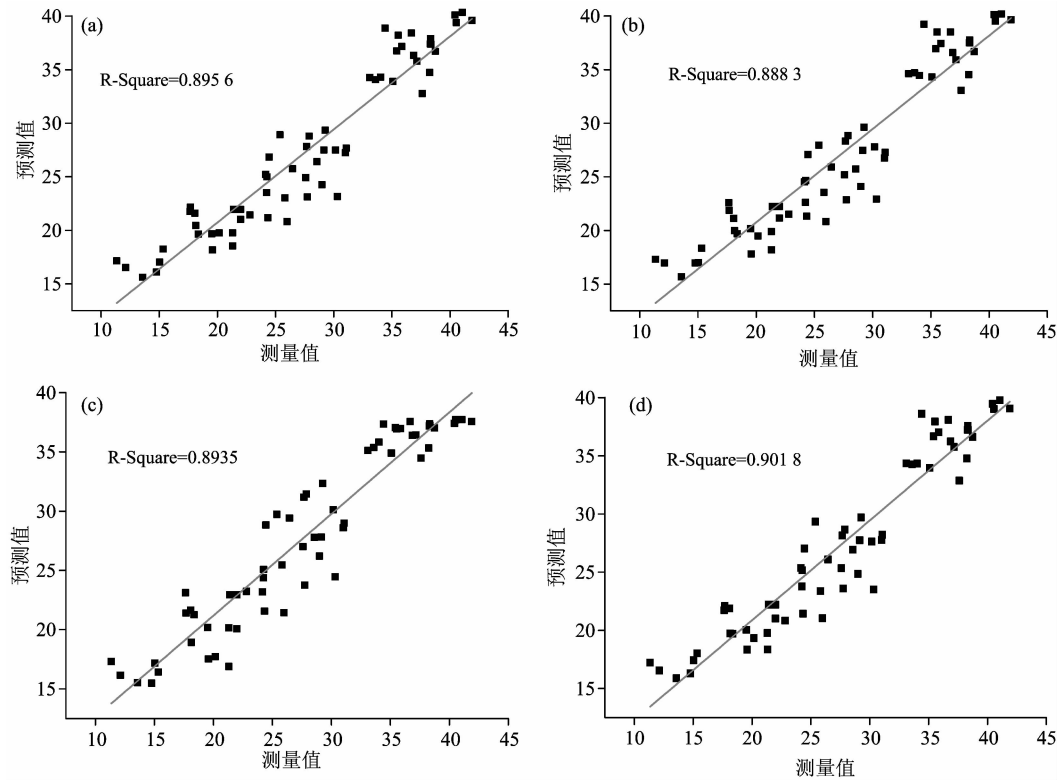


Fig. 4 Scatter plot of predicted and measured values of SPAD in tomato leaves by four different models

Table 4 Comparison among different models based on four types of algorithms

S. N. ^a	Alg. ^b	Input	Stage	Elements	R-Squ. ^c	RMSE ^d
Model 1	PLSR	470	Calibration	122	0.888 7	2.655 5
			Prediction	62	0.895 6	2.664 2
Model 2	PCR	470	Calibration	122	0.878 9	2.7699
			Prediction	62	0.888 3	2.744 6
Model 3	BPNN	4	Calibration	122	0.896 8	2.551 4
			Prediction	62	0.893 5	2.685 4
Model 4	LS-SVM	4	Calibration	122	0.890 2	2.627 4
			Prediction	62	0.901 8	2.599 2

S. N. a; sequence number. Alg. b; regression algorithm. R-Squ. c; determination coefficient. RMSE d; root mean square error

是可行的。表 4 中可以看出 4 个模型的预测效果都较好, 其中 LS-SVM 模型的效果最好, 其预测集决定系数 R^2 最大、预测集均方根误差 RMSEP 最小。PLSR 模型预测效果其次, 这可能是由于 PLSR 模型只能利用光谱数据中的线性信息, 同时较多的输入变量含有冗余信息及噪声, 导致模型预测能

力下降。相对于 PLSR 和 PCR 模型, BPNN 和 LS-SVM 模型只有 4 个输入变量, 这 4 个输入变量简化了建模步骤, 提高了运算效率。4 个模型的预测结果也说明了非线性算法在灰霉病胁迫下番茄叶片 SPAD 值预测方面优于线性算法。

3 结 论

以健康和染病番茄叶片为研究对象, 通过提取番茄叶片高光谱图像中感兴趣区域的光谱信息, 建立高光谱信息与番茄叶片 SPAD 值的相关模型。采用不同的方法对光谱数据进行预处理并建立 PLSR 模型, 通过比较, 最后确定基于 Normalize 的预处理方法并同时建立 PCR 模型, 再基于 PLSR 推荐的 4 个变量建立 BPNN 和 LS-SVM 模型。结果表明, LS-SVM 模型的预测效果最好, 预测集决定系数 R^2 为 0.901 8, 预测集均方根误差 RMSEP 为 2.599 2。本文提出的基于高光谱技术的番茄叶片叶绿素含量 (SPAD) 的检测方法, 为实现番茄植株生长状况的实时监测及番茄叶片中其他化学指标的检测提供了新的思路。

References

- [1] CAO Xue-ren, ZHOU Yi-lin, DUAN Xia-yu, et al(曹学仁, 周益林, 段霞瑜, 等). Acta Phytopathologica Sinica(植物病理学报), 2009, 39(3): 290.
- [2] Evans T, Song J C, Jameson P E. Plant Growth Regul, 2012, 66(1): 95.
- [3] Loh C H, Inbaraj B S, Liu M H, et al. Journal of Agricultural and Food Chemistry, 2012, 60(24): 6108.
- [4] Scotter M J. Food Additives and Contaminants, 2011, 28(5): 527.

- [5] Liu F, Zhang F, Jin Z L, et al. *Analytica Chimica Acta*, 2008, 629(1–2): 56.
- [6] Min M, Lee W S, Kim Y H, et al. *Hort Science*, 2006, 41(1): 162.
- [7] Ulissi V, Antonucci F, Benincasa P, et al. *Sensors*, 2011, 11(6): 6411.
- [8] Jr F A D, Center T D, Mattison E D. *Aquatic Botany*, 2012, 100: 72.
- [9] LIU Fei, WANG Li, HE Yong, et al(刘飞, 王莉, 何勇, 等). *Journal of Infrared and Millimeter Waves(红外与毫米波学报)*, 2009, 28(4): 272.
- [10] Polder G, Gerie W A M, Heijden V, et al. *Journal of Near Infrared Spectroscopy*, 2003, 11(3): 193.
- [11] XU Lu, SHAO Xue-guang(许禄, 邵学广). *Methods of Chemometrics(化学计量学方法)*. Beijing: Science Press(北京: 科学出版社), 2004. 167.
- [12] Feng J, Wang Z, West L, et al. *Anal. Bioanal. Chem.*, 2011, 400(10): 3261.
- [13] Killner M H M, Rohwedder J J. R, Pasquini C. *Fuel*, 2011, 90(11): 3268.
- [14] LU Wan-zhen(陆婉珍). *Modern Near Infrared Spectroscopy Analytical Technology(现代近红外光谱分析技术)*. Beijing: China Petrochemical Press(北京: 中国石化出版社), 2010. 39.
- [15] Mouazen A M, Kuang B, Baerdemaker J D, et al. *Geoderma*, 2010, 158(1–2): 23.
- [16] Vasques G M, Grunwald S, Sickman J O. *Geoderma*, 2008, 146(1–2): 14.
- [17] CHU Xiao-li(褚小立). *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and its Applications(化学计量学方法与分子光谱分析技术)*. Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011. 66.
- [18] Zhao X D, He Y, Bao Y D. *Sensor Letters*, 2011, 9(3): 1055.
- [19] Liu F, Ye X J, He Y, et al. *Journal of Food Engineering*, 2009, 93(2): 127.
- [20] Bierzoza M, Baker A, Bridgeman J. *Advances in Engineering Software*, 2012, 44(1): 126.
- [21] Suykens J A K, Vandewalle J. *Neural Processing Letters*, 1999, 9(3): 293.
- [22] Li J Z, Liu H X, Yao X J, et al. *Analytica Chimica Acta*, 2007, 581(2): 333.
- [23] Chauchard F, Cogdii R, Roussel S, et al. *Chemometrics and Intelligent Laboratory Systems*, 2004, 71 (2): 141.
- [24] Ferrlo M F, Godoy S C, Gerbase A E, et al. *Analytica Chimica Acta*, 2007, 595(1–2): 114.
- [25] Li X L, He Y. *Food Bioprocess Technol.*, 2010, 3: 651.

Study of Detection of SPAD Value in Tomato Leaves Stressed by Grey Mold Based on Hyperspectral Technique

XIE Chuan-qi, HE Yong, LI Xiao-li, LIU Fei, DU Peng-peng, FENG Lei*

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

Abstract Hyperspectral imaging feature of chlorophyll content (SPAD) in tomato leaves stressed by grey mold was studied in the present paper. Hyperspectral imagings of healthy and infected tomato leaves were obtained by hyperspectral imaging system from 380 to 1 030 nm and diffuse spectral response of region of interest (ROI) from hyperspectral imaging was extracted by ENVI software, then different preprocessing methods were used including smoothing and normalization etc. The partial least squares regress (PLSR) and principal component regress (PCR) models were developed for the prediction of SPAD value in tomato leaves based on normalization preprocessing method, then the back-propagation neural network (BPNN) and least squares-support vector machine(LS-SVM)models were built based on the four variables suggested by PLSR model. Among the four models, LS-SVM model was the best to predict SPAD value and the coefficient of determination (R^2) was 0.901 8 with the root mean square error of prediction (RMSEP) of 2.599 2. It was demonstrated that chlorophyll content (SPAD) in healthy and infected tomato leaves can be effectively detected by the hyperspectral imaging technique.

Keywords Hyperspectral imaging; Chlorophyll; Partial least squares regress; Principal component regress; Back-propagation neural network; Least squares-support vector machine; Tomato; Grey mold

(Received Apr. 6, 2012; accepted Jun. 20, 2012)