doi:10.6041/j.issn.1000-1298.2015.03.046

番茄叶片早疫病近红外高光谱成像检测技术*

谢传奇[1,2]　方孝荣[3]　邵咏妮[1]　何　勇[1]

(1. 浙江大学生物系统工程与食品科学学院, 杭州 310058; 2. 佛罗里达大学农业与生物工程系, 盖恩斯维尔 32611;
3. 金华职业技术学院成人教育学院, 金华 321017)

摘要: 提出了基于格拉姆斯密特(MGS)模型和贝叶斯罗蒂斯克回归(BlogReg)的近红外高光谱成像技术检测番茄叶片早疫病的方法。利用高光谱图像采集系统获取波长 874 ~ 1 734 nm 范围内 70 个染病和 80 个健康番茄叶片的高光谱图像,选取染病和健康叶片 30 像素 ×30 像素感兴趣区域的光谱反射率。建立了番茄叶片早疫病的最小二乘-支持向量机(LS - SVM)识别模型,再通过 MGS 和 BlogReg 提取特征波长(EW),分别得到 5 个(911、1 409、1 511、1 609、1 656 nm)和 9 个(901、905、908、915、918、1 123、1 305、1 460、1 680 nm)特征波长,并建立 EW -LS - SVM 和 EW - LDA 模型。在所有模型中,建模集的正确识别率为 93% ~98%,预测集的正确识别率为 96% ~100%。结果表明,近红外高光谱成像技术检测番茄叶片早疫病是可行的,MGS 和 BlogReg 都是有效的特征波长提取方法。

关键词: 番茄　早疫病　近红外光谱　格拉姆斯密特模型　贝叶斯罗蒂斯克回归　最小二乘-支持向量机

中图分类号: TP391; S436.412　　**文献标识码**: A　　**文章编号**: 1000-1298(2015)03-0315-05

Detection of Early Blight on Tomato Leaves Using Near-infrared Hyperspectral Imaging Technique

Xie Chuanqi[1,2]　Fang Xiaorong[3]　Shao Yongni[1]　He Yong[1]

(1. *College of Biosystems Engineering and Food Science*, *Zhejiang University*, *Hangzhou* 310058, *China*
2. *Department of Agricultural and Biological Engineering*, *University of Florida*, *Gainesville FL*32611, *USA*
3. *Adult Education College*, *Jinhua Polytechnic*, *Jinhua* 321017, *China*)

Abstract: Early detection of early blight on tomato leaves using NIR hyperspectral imaging technique based on modified gram-schmidt (MGS) model and Bayesian logistic regression (BlogReg) were studied. Hyperspectral images of 70 infected and 80 healthy tomato leaves were acquired by hyperspectral imaging system in the spectral wavelength of 874 ~ 1 734 nm. Spectral reflectance of 30 × 30 pixels from region of interest (ROI) of hyperspectral image was extracted. Least squares - support vector machine (LS - SVM) model based on the full wavelength was established to detect early blight. Five (911 nm, 1 409 nm, 1 511 nm, 1 609 nm, 1 656 nm) and nine wavelengths (901 nm, 905 nm, 908 nm, 915 nm, 918 nm, 1 123 nm, 1 305 nm, 1 460 nm, 1 680 nm) were selected by MGS and BlogReg, respectively. Then, LS - SVM and linear discriminant analysis (LDA) models were built based on these effective wavelengths. Among these models, the correct classification rates were 93% ~98% in calibration set and 96% ~100% in prediction set, respectively. The result indicated that it was feasible to detect early blight on tomato leaves by using NIR hyperspectral imaging technique.

Key words: Tomato　Early blight　Near-infrared spectroscopy　Modified gram-schmidt model　Bayesian logistic regression　Least square - support vector machines

收稿日期: 2014 - 04 - 04　修回日期: 2014 - 04 - 19

* 国家高技术研究发展计划(863 计划)资助项目(2013AA102301)、高等学校博士学科点专项科研基金资助项目(20130101110104)、教育部留学回国人员科研启动基金资助项目和中央高校基本科研业务费专项资金资助项目(2014FZA6005)

作者简介: 谢传奇,博士生,主要从事光谱及光谱检测技术研究,E-mail: cqxie@zju.edu.cn

通讯作者: 何勇,教授,博士生导师,主要从事精细农业与农业物联网技术研究,E-mail: yhe@zju.edu.cn

引言

我国是农业大国,农作物病害是制约农业生产稳定发展的最重要因素之一,实时、灵敏、可靠的作物病害检测、预报和防治是进行科学的作物生产管理的基础。我国的农业工程水平较低,虽然国内研究学者利用生理生化技术在植物病害检测方面做了一些研究并且取得了进展,但是已有利用光谱和光谱成像技术对植物病害检测的模型,无论是单一病害诊断模型,还是相似病害诊断模型,都缺乏深入研究,很难在农业工程实际过程中得到应用。目前,这是影响植物病害快速、准确、早期检测技术研究及应用的瓶颈问题。如何突破传统,运用工程学、计算机等现代技术等手段解决这个问题在现代农业中具有深远的意义。

番茄叶片很容易感染早疫病,尤其在高温高湿的环境中,当植株感染病害后,病斑会迅速蔓延,直接影响植株的健康生长及番茄果实的品质和产量。因此,研究一种无损且有效的番茄叶片早疫病检测技术极其重要。高光谱成像技术结合了光谱和图像技术,具有多波段、高分辨率、谱图合一等优点[1]。通过光谱和图像特征的有效融合,可以避免仅依靠图像或光谱特征的不足,有效地提高农作物病害检测的准确率。目前高光谱成像技术已经在作物病害检测中得到广泛应用[2-7]。但利用近红外高光谱成像技术结合格拉姆斯密特和贝叶斯罗蒂斯克回归算法提取特征波长识别番茄叶片早疫病的研究还未见报道。

本文利用近红外高光谱成像技术建立番茄叶片早疫病的识别模型,基于格拉姆斯密特(Modified gram-schmidt, MGS)和贝叶斯罗蒂斯克回归(Bayesian logistic regression, BlogReg)提取特征波长,并比较用全波段和特征波长所建模型的识别效果,从而确定番茄叶片早疫病识别的最优模型和最佳波长。

1 材料与方法

1.1 仪器设备

近红外高光谱图像采集系统(图 1)主要包括计算机、成像光谱仪(N17E 型, Specim, Finland)、OLE-23型镜头、线光源(Schott DCR Ⅲ型, Oriel Instruments, USA)、SC300A 型移动平台控制装置、PSA 200-11-X 型电控位移台等。波长范围为 874~1 734 nm,光谱分辨率为 5 nm,满帧像素数为 320 像素×256 像素,计算机接口为 USB。数据分析软件为 ENVI 4.7(Research System Inc., Boulder Co., USA)、Unscrambler V9.7(Camo, Process AS, Oslo, Norway)和 Matlab R2009a(The Math Works Inc., Natick, MA, USA)。

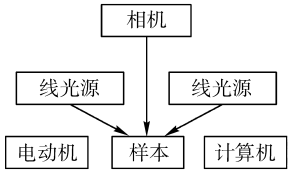

图 1 近红外高光谱图像采集系统

Fig. 1 NIR hyperspectral imaging system

1.2 试验材料

试验用番茄品种为浙杂 809,共 40 盆,每盆 1 棵,将其中 20 棵接种早疫病菌,每棵番茄接种5 片叶子,每片叶子上接种一块底面直径为 5 mm 的圆柱形菌丝块,剩余 20 棵作为健康对照,将全部番茄植株放置于可控人工气候箱中培养。为了利于早疫病菌的侵染,将可控人工气候箱设置为 2 段运行,2 段的温度分别是 28℃和 26℃,相对湿度都是 86%,分别保持 12 h 光照和 12 h 黑暗,2 段交替工作。约 24 h 后接种叶片出现微小病斑,剔除未侵染样本和试验过程中损坏的样本,最终选取 70 个染病叶片和80 个健康叶片进行进一步研究。从染病和健康样本中随机选取 2/3 即 100 个(健康:53,染病:47)作为建模集,剩下的 1/3 即 50 个(健康:27,染病:23)作为预测集,用来验证模型的准确性,健康样本的 Y 变量设为 1,染病样本的 Y 变量设为 2。

1.3 试验流程

首先利用高光谱图像成像仪采集番茄叶片的高光谱图像,然后对原始图像进行黑白板校正,从每个校正后的高光谱图像中提取 30 像素×30 像素的感兴趣区域(Region of interest, ROI)的光谱反射率信息。基于全波段(波长 900~1 700 nm)建立最小二乘-支持向量机(Least squares - support vector machine, LS-SVM)识别模型,同时分别用格拉姆斯密特和贝叶斯罗蒂斯克回归对全波段光谱进行特征波长提取,将得到的特征波长(Effective wavelength, EW)作为 X 变量分别建立 EW-LS-SVM 和 EW-LDA(Linear discriminant analysis)识别模型,并比较各模型的鉴别效果,选取最优识别模型。整个流程如图 2 所示。

1.4 图像采集与校正

首先调整近红外高光谱成像仪的各项参数,主要包括调节光强度、图像清晰度、镜物垂直高度、移动平台的速度和曝光时间。经过反复尝试,物镜垂直高度、曝光时间和移动平台的速度分别设置为 29 cm、4.8 ms 和 24 mm/s。将采下的番茄叶片迅

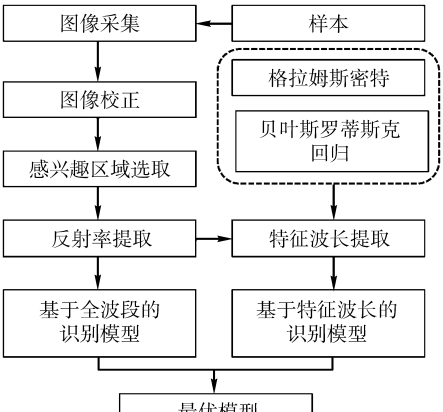

图 2　试验流程图

Fig. 2　Main steps of experiment

速、依次平铺在移动平台上，采集 874 ~ 1 734 nm 波段范围内的高光谱图像。为了消除原始高光谱图像在某些波段范围内的噪声，需要对原始高光谱图像进行黑白板校正，白板为聚四氟乙烯漫反射材料（CAL - tile200 型，200 mm × 25 mm × 10 mm）。采集标准白色板得到全白的标定图像 I_w（反射率接近 100%），关闭相机镜头盖、切断光源，采集全黑的标定图像 I_d（反射率接近 0%）。然后对原始图像 I_r 进行校正，R 为校正后的高光谱图像，即

$$R = \frac{I_r - I_d}{I_w - I_d} \tag{1}$$

1.5　识别模型

LS - SVM 模型能够有效地处理分类和回归的建模问题，同时也能较快地解决多元变量的复杂性问题[8-9]，已被广泛用于光谱研究领域[10-11]。线性判别分析是一种功能强大的识别方法，可用在模式识别、统计学和机器学习等领域，目前 LDA 算法已在光谱领域得到了广泛的应用[12-14]。本文将提取的特征波长作为新的输入变量分别建立 LS - SVM 和 LDA 识别模型。

1.6　特征波长提取方法

MGS 算法的基本思想是根据投影原理在原有正交基的基础上构造一个新的正交基。先根据相关性从特征矩阵 $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$ 中选取一个特征向量作为第 1 个正交基 $\boldsymbol{x}_{z1}$，其中 $\boldsymbol{x}_i$ 为 n 维向量。计算其他向量在该正交基上的投影 $\boldsymbol{x}_p = \frac{(\boldsymbol{x}_i, \boldsymbol{x}_{z1})}{\| \boldsymbol{x}_{z1} \|^2} \boldsymbol{x}_{z1}$，其他向量与投影的差 $\boldsymbol{x}_i' = \boldsymbol{x}_i - \boldsymbol{x}_p$ 与正交基 $\boldsymbol{x}_{z1}$ 正交，得到与第 1 个正交基正交的 $n-1$ 个特征向量。同样，类标签向量 $\boldsymbol{Y}$ 在第 1 个正交基上的投影 $\boldsymbol{Y}_p = \frac{(\boldsymbol{Y}, \boldsymbol{x}_{z1})}{\| \boldsymbol{x}_{z1} \|^2} \boldsymbol{x}_{z1}$，$\boldsymbol{Y}$ 与投影的差 $\boldsymbol{Y}' = \boldsymbol{Y} - \boldsymbol{Y}_p$ 与正交基 $\boldsymbol{x}_{z1}$ 正交。至此，便选择出了第 1 特征向量，并通过 MGS 消除了第 1 个向量对后续特征选择的影响，既选择出了相关特征向量，又消除了冗余。更新 $\boldsymbol{x}_i = \boldsymbol{x}_i', \boldsymbol{Y} = \boldsymbol{Y}', n = n - 1$，重复以上程序直至达到设定的终止准则。具体步骤为：

（1）对特征矩阵 $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$ 的各特征向量进行单位化，$\boldsymbol{x}_{ni} = \frac{\boldsymbol{x}_i}{\| \boldsymbol{x}_i \|}$（$i = 1, 2, \cdots, n$）。

（2）将类标签 $\boldsymbol{Y}$ 与各单位化特征向量进行内积 $\boldsymbol{y}_{p_i} = (\boldsymbol{Y}, \boldsymbol{x}_{ni})$（$i = 1, 2, \cdots, n$），选择使 $\boldsymbol{y}_{p_i}$ 最大的特征向量 $\boldsymbol{x}_{ns}$（$s \in 1, 2, \cdots, n$），并作为标准正交基。

（3）将特征矩阵 $\boldsymbol{X}$ 的各特征向量投影到标准正交基 $\boldsymbol{x}_{ns}$ 上，$\boldsymbol{x}_p = \frac{(\boldsymbol{x}_i, \boldsymbol{x}_{ns})}{\| \boldsymbol{x}_{ns} \|^2} \boldsymbol{x}_{ns}$，并计算特征向量与投影的差 $\boldsymbol{x}_i' = \boldsymbol{x}_i - \boldsymbol{x}_p$。

（4）将类标签向量 $\boldsymbol{Y}$ 投影到标准正交基 $\boldsymbol{x}_{ns}$ 上，$\boldsymbol{Y}_p = \frac{(\boldsymbol{Y}, \boldsymbol{x}_{ns})}{\| \boldsymbol{x}_{ns} \|^2} \boldsymbol{x}_{ns}$，并计算 $\boldsymbol{Y}$ 与投影的差 $\boldsymbol{Y}' = \boldsymbol{Y} - \boldsymbol{Y}_p$。

（5）更新变量，$\boldsymbol{X} = [\boldsymbol{x}_1', \cdots, \boldsymbol{x}_{s-1}', \boldsymbol{x}_{s+1}', \cdots, \boldsymbol{x}_n']$，$\boldsymbol{Y} = \boldsymbol{Y}', n = n - 1$。

（6）若已选择的特征向量数达到要求，则终止循环，否则继续执行步骤（2）~（6）。

贝叶斯罗蒂斯克回归也是一种有效的波长提取方法，目前已经在医学领域得到应用[15]。

2　结果与讨论

2.1　样本光谱特征

试验选取校正后的高光谱图像中 30 × 30 个像素点为感兴趣区域（ROI）。然后将 ROI 的所有像素点在 900 ~ 1 700 nm 波长范围内的反射率进行平均，最后每个样本得到 256 个波段信息（$\boldsymbol{X}$ 变量），通过提取 ROI 区域内多个点的平均光谱来代替单点光谱，这样能够减少单点光谱的偶然性和差异性，有助于提高模型的准确性和稳定性[16]。图 3 是健康和染病番茄叶片在 874 ~ 1 734 nm 波段范围内的光谱反射率分别取平均后得到的光谱反射率曲线。通过图 3 可以看出，2 类样本的光谱反射率曲线相似，没有较大区别，曲线首尾两端存在噪声，大致走势是先增大后减小，之后再增大，最后减小，染病和健康样本曲线分别在 1 453 nm 和 1 456 nm 附近处各有一个波谷。但在某些波段范围内染病和健康样本存在差别，在 900 ~ 1 155 nm 波段范围内，染病样本的光谱反射率低于健康样本，在 1 155 ~ 1 700 nm 波段范围内染病样本的光谱反射率明显高于健康样本，在 1 155 nm 附近处有一个交叉。为了更好地识别出染病样本，减少某些波段范围内噪声对所建模型精确度和稳健性的影响，本文只采用 900 ~ 1 700 nm 波段

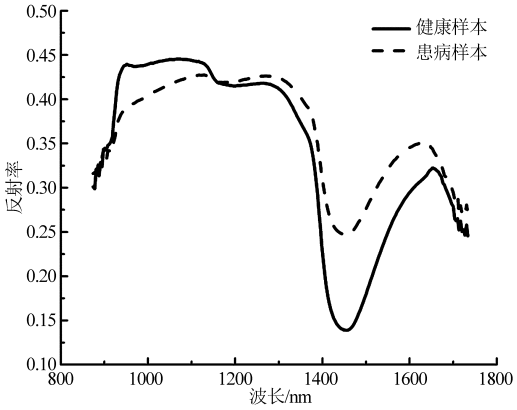

图 3 健康和染病叶片的光谱反射率曲线

Fig. 3 Spectral reflectance curves of healthy and infected tomato leaves

范围内的光谱信息。

2.2 特征波长选取

较多的输入变量不仅降低了运算效率、不便于实时监测仪器的开发,同时可能含有大量冗余信息,从而影响模型的识别效果。试验采用 MGS 和 BlogReg 2 种方法分别对全谱进行特征波长提取。通过 MGS 得到了 5 个特征波长(911、1 409、1 511、1 609、1 656 nm),通过 BlogReg 获得 9 个特征波长(901、905、908、915、918、1 123、1 305、1 460、1 680 nm)。与全谱相比,MGS 和 BlogReg 推荐的特征波长数分别只占全波段波长数 238 的 2.10% 和 3.78%,然后将这些得到的新特征变量代替原始全波段变量分别建立 EW - LS - SVM 和 EW - LDA 识别模型。

2.3 基于全波段和特征波长的识别模型

将全波段光谱反射率作为 $\boldsymbol{X}$ 变量建立 LS - SVM 识别模型,建模集和预测集的识别率分别是 98% 和 96%。基于全波段建立的识别模型,虽然取得了较好的效果,但较多的输入变量(238 个)不仅降低了运算效率,而且可能含有大量冗余信息,从而影响模型的识别效果。研究进一步将 MGS 和 BlogReg 推荐的 5 个和 9 个特征波长作为输入变量分别建立 MGS - LS - SVM、MGS - LDA、BlogReg - LS - SVM 和 BlogReg - LDA 模型,4 个模型的识别效果都较好,建模集的识别率为 93% ~98%,预测集的识别率为 96% ~100%,所有模型的结果如表 1 所示,本文识别率均为正确识别率。MGS 推荐的输入变量只有 5 个,BlogReg 推荐的输入变量只有 9 个,较少的输入变量不仅消除了某些无效波段的冗余信息,而且也提高了建模效率。

表 1 不同模型的识别结果

Tab. 1 Correct classification rates of different models

模型	输入变量数	建模集 样本数	建模集 误判数	建模集 识别率/%	预测集 样本数	预测集 误判数	预测集 识别率/%
LS - SVM	238	100	2	98	50	2	96
MGS - LS - SVM	5	100	5	95	50	0	100
MGS - LDA	5	100	6	94	50	2	96
BlogReg - LS - SVM	9	100	2	98	50	1	98
BlogReg - LDA	9	100	7	93	50	2	96

2.4 特征波长识别率

本文将 2 种方法获取的共 14 个特征波长作为输入变量,分别建立 LS - SVM 和 LDA 识别模型,建模集的结果如图 4a 所示,预测集的结果如图 4b 所示。建模集和预测集中,每个波长点处 LS - SVM 模型的识别率均不小于 LDA 模型的识别率,每一个模型的识别率总体趋势为先升高后下降。建模集中 LS - SVM 模型识别率最高(88%)的波长是 1 460 nm,识别率最低(64%)的是 918 nm,LDA 模型识别率最高(85%)的波长是 1 460 nm,识别率最低(36%)的是 901 nm;预测集中 LS - SVM 模型识别率最高(94%)的波长是 1 460 nm,识别率最低(62%)的是 1 123 nm,LDA 模型识别率最高(90%)的波长是 1 409、1 460 和 1 511 nm,识别率最低(24%)的是 911 nm。因此,近红外区域波长 1 460 nm 处的光谱

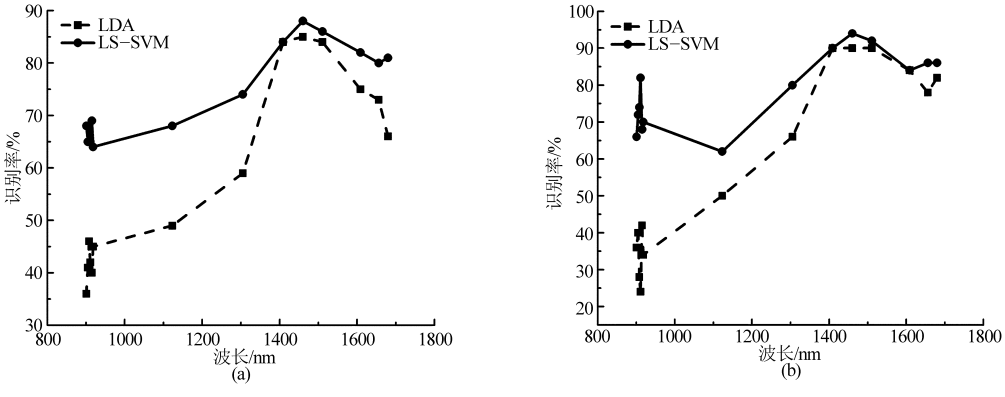

图 4 特征波长的识别率

Fig. 4 Correct classification rates based on each selected wavelength

(a) 建模集 (b) 预测集

反射率对番茄叶片早疫病害的识别具有重要作用，这为病害检测仪器的开发提供了理论支持。

3　结论

（1）采用近红外高光谱成像技术，结合不同识别模型和特征提取方法，实现了对番茄叶片早疫病的识别研究。

（2）分别以全波段光谱、MGS 和 BlogReg 提取的特征波长建立了不同的识别模型，结果可以看出，在所有模型中，BlogReg－LS－SVM 模型的效果最优，建模集和预测集的识别率均达到了 98%，且输入变量仅为 9 个，这样不仅简化了模型，而且提高了运算效率。其他一些模型中建模集的识别率略低于预测集，这可能是因为样本的随机选择和部分轻微染病样本被误判为健康样本等原因造成的。

（3）波长 1 460 nm 处的光谱反射率有利于病害的检测。利用近红外高光谱成像技术检测番茄叶片早疫病是可行的，而且 MGS 和 BlogReg 2 种方法能够有效地提取特征波长。

参　考　文　献

1　Kamruzzaman M, Elmasry G, Sun D W, et al. Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis[J]. Analytica Chimica Acta, 2012,714:57－67.

2　Bauriegel E, Giebel A, Geyer M, et al. Early detection of fusarium infection in wheat using hyper-spectral imaging[J]. Computers and Electronics in Agriculture, 2010,75(2): 304－312.

3　Fiore A D, Reverberi M, Ricelli A, et al. Early detection of toxigenic fungi on maize by hyperspectral imaging analysis[J]. International Journal of Food Microbiology, 2010,144(1): 64－71.

4　Williams P J, Geladi P, Britz T J, et al. Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis[J]. Journal of Cereal Science, 2012,55(3): 272－278.

5　Mahlein, A K, Rumpf T, Welke T,et al. Development of spectral indices for detecting and identifying plant diseases[J]. Remote Sensing of Enviorment, 2013, 128 :21－30.

6　Rumpf T, Mahlein A K, Steiner U, et al. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance[J]. Computers and Electronics in Agriculture, 2010,74(1):91－99.

7　Yang C H, Everitt J H, Fernandez C J. Comparison of airborne multispectral and hyperspectral imagery for mapping cotton root rot [J]. Biosystems Engineering, 2010,107(2): 131－139.

8　Wu D, Feng L, He Y, et al. Variety identification of Chinese cabbage seeds using visible and near-infrared spectroscopy[J]. Transactions of the ASABE, 2008,51(6) : 2193－2199.

9　Liu F, He Y. Classification of brands of instant noodles using Vis/NIR spectroscopy and chemometrics[J]. Food Research International, 2008,41(5):562－567.

10　Nie P C, Wu D, Sun D W, et al. Potential of viaible and near infrared spectroscopy and pattern recognition for rapid quantification of notoginseng powder with adulterants[J]. Sensors, 2013, 13(10):13820－13834.

11　Wu D, Chen J Y, Lu B Y, et al. Application of near infrared spectroscopy for rapid determination of antioxidant activity of bamboo leaf extract[J]. Food Chemistry, 2012,135(4):2147－2156.

12　Wei X, Liu F, Qiu Z J, et al. Ripeness classification of astringent persimmon using hyperspectral imaging technique[J]. Food and Bioprocess Technology, 2014,7(5): 1371－1380.

13　Riovanto R, Cynkar W U, Berzaghi P, et al. Discrimination between Shiraz wines from different Australian regions: the role of spectroscopy and chemometrics[J]. Journal of Agricultural and Food Chemistry, 2011,59(18): 10356－10360.

14　Kamruzzaman M, Elmasry G, Sun D W, et al. Application of NIR hyperspectral imaging for discrimination of lamb muscles[J]. Journal of Food Engineering, 2011,104(3): 332－340.

15　Gawley G C, Talbot N L C. Gene selection in cancer classification using sparse logistic regression with Bayesian regularization [J]. Bioinformatics, 2006,22(19): 2348－2355.

16　Zhang C, Liu F, Zhang H L, et al. Identification of varieties of black bean using ground based hyperspectral imaging[J]. Spectroscopy and Spectral Analysis, 2014,34(3): 746－750.