

doi:10.6041/j.issn.1000-1298.2016.04.026

# 苹果霉心病可见/近红外透射能量光谱识别方法

雷雨 何东健 周兆永 张海辉 苏东

(西北农林科技大学机械与工程学院, 陕西杨凌 712100)

**摘要:** 针对苹果霉心病从外表无法识别的难题, 提出基于可见/近红外透射能量光谱进行快速无损识别的模型和方法。在 200 ~ 1 100 nm 波段内采集了 200 个苹果的透射能量光谱数据, 随机选取 140 个样品作为训练集, 剩余 60 个样品作为测试集。用平滑法和多元散射校正对光谱数据进行预处理。基于全光谱、连续投影算法 (SPA) 提取的 12 个特征波长、主成分分析 (PCA) 提取的 9 个主成分, 分别建立了偏最小二乘判别法、误差反向传播神经网络和支持向量机 (SVM) 识别模型。实验结果说明, 应用 PCA-SVM 建立的模型识别性能最优, 该模型对测试集和训练集中霉心病果和健康果的识别正确率分别为 99.3% 和 96.7%。基于 SPA 和 PCA 所建模型的输入变量数仅相当于基于全光谱所建模型输入变量数的 0.99% 和 0.74%, 极大降低了模型的复杂度。研究表明, 该方法是可行的且具有较高识别准确度, 为苹果在线内部品质分级和便携式苹果霉心病检测仪的研究提供了技术依据。

**关键词:** 苹果; 霉心病; 能量光谱; 连续投影算法; 主成分分析; 支持向量机

**中图分类号:** O657.33; S436.611 **文献标识码:** A **文章编号:** 1000-1298(2016)04-0193-08

## Detection of Moldy Core of Apples Based on Visible/Near Infrared Transmission Energy Spectroscopy

Lei Yu He Dongjian Zhou Zhaoyong Zhang Haihui Su Dong

(College of Mechanical and Electronic Engineering, Northwest A&amp;F University, Yangling, Shaanxi 712100, China)

**Abstract:** In order to solve the problem of identification moldy core of apples from the surface, a quick and non-destructive detection method was proposed based on visible/near infrared transmission energy spectroscopy. Visible/near infrared transmission energy spectra of 200 apples were collected in the wavelength range of 200 ~ 1 100 nm. Totally 140 samples were used for the calibration set, and 60 samples for the validation set. Smoothing method and multiple scattering correction were used to preprocess the original spectra. Totally 12 characteristic wavelengths and 9 principal components were selected by successive projections algorithm (SPA) and principal component analysis (PCA), respectively. Partial least squares discriminant analysis, error back propagation artificial neural networks, and support vector machine (SVM) measurement model were established based on SPA and PCA, respectively. The results showed that the best model was PCA-SVM, and its recognition accuracy rate reached 99.3% for the calibration set and 96.7% for the validation set. The models established based on SPA and PCA were much simpler than those based on full spectra, since the numbers of input variable of them were only about 0.99% and 0.74% of that of full spectra, respectively. The results showed that the method was available and had high identification accuracy. Meanwhile, the results would provide theoretical basis for the research and development of on-line detection of internal quality in apples and portable moldy core apple detector.

**Key words:** apples; moldy core; energy spectra; successive projections algorithm; principal component analysis; support vector machine

收稿日期: 2015-11-01 修回日期: 2015-12-21

基金项目: 国家高技术研究发展计划(863计划)项目(2013AA10230402)和陕西省科技统筹创新工程计划项目(2014KTCL02-15)

作者简介: 雷雨(1990—),男,博士生,主要从事智能化检测技术研究,E-mail: ly759059219@163.com

通信作者: 何东健(1957—),男,教授,主要从事智能化检测与控制、图像分析与计算机视觉研究,E-mail: hdj168@nwsuaf.edu.cn

## 引言

苹果霉心病又称霉腐病、心腐病,是苹果内部品质的主要病害,霉心果中含有展青霉素和棒曲霉素,具有影响生育、致畸、致癌和致突变等毒理作用<sup>[1-2]</sup>。染病果从心部逐步向外腐烂,早中期果面不表明明显相关症状,不易被发现,如果能从源头上控制染病果流入市场具有重要意义。然而因缺乏有效的检测方法,无法在生产分级流水线上剔除染病果,霉心病苹果的检测成为苹果采后加工的一大难题。

国内外学者已尝试用不同方法检测苹果霉心病<sup>[3-6]</sup>。这些方法在一定程度上检测出了苹果霉心病,但成本较高,且存在分析过程比较复杂、耗时、准确率低、技术条件复杂、需要破坏样品等缺点。

光谱技术具有分析速度快、无损检测、低成本等优点,在苹果内部品质检测方面得到广泛应用并且获得认可,但由于霉心病是从心室发病又腐烂程度不一,存在光谱的重叠,很难建立精确度高的苹果霉心病识别模型<sup>[7]</sup>。CLARK等<sup>[8]</sup>利用透射光谱检测“Braeburn”苹果内部褐变,当果轴水平放置,用PLS建模时效果最佳;MCGLON等<sup>[9]</sup>采用2种光谱扫描系统在线检测苹果褐变,预测模型具有较高精度;韩东海等<sup>[10]</sup>利用可见近红外连续透射光谱对苹果内部褐变进行了研究,样品的正确判别率达到95.65%;李顺峰等<sup>[11]</sup>利用近红外漫反射光谱技术判别苹果霉心病的可行性,基于主成分分析构建了苹果霉心病近红外漫反射光谱判别,模型的建模集正确判别率为89.9%,检测集正确判别率为87.8%;SHENDEREY等<sup>[12]</sup>采用近红外透射光谱技术对“Delicious”苹果霉心病进行检测,并建立PLS模型,检测集正确判别率为90.1%(健康果实和发病果的正确判别率分别为91.7%和88.4%)。表明可见/近红外光谱技术可识别出苹果霉心病,但上述研究一般采用通用型光谱仪在室内完成数据采集,且基于不同方法所建模型对苹果霉心病检测的影响较大,模型的准确度难以达到应用要求。本文以开发便携式苹果霉心病检测仪为目标,以陕西洛川红富士苹果为研究对象,采用便携式微型光纤光谱仪,采集样本在200~1100 nm波段内可见近红外光谱,探明霉心病病害与光谱特性的关系,研究并提出基于可见/近红外透射能量光谱进行苹果霉心病快速无损识别的模型和方法,建立多个模型进行对比分析,以确定较优的建模方法。

## 1 材料与方 法

### 1.1 实验材料

本文以红富士苹果为研究对象。2014年10月,于陕西省洛川县采集大小均匀、外表光滑、无机械损伤的疑似霉心病果(由有经验的果农选取)和好果,从中精细选取200个苹果作为实验样本,将样本清洁擦净后进行编号,并置于实验室(22℃,相对湿度55%)内,10 d后(让染病果发病)采集实验样本光谱数据。

### 1.2 透射能量光谱采集系统

供试样本光谱数据采集系统如图1所示。为减小环境光对光谱的影响,设计外侧有门、内侧涂黑的采集暗箱,位于暗箱内的载物台可沿支架调节高度,光源置于样品正上方,在样品下方光纤探头连接芯径为600 μm的QP600-1-SR光纤用于接收光,光源选用4个50 W/24 V的64445U型卤素灯(OSRAM公司),均布安装在灯架上,透过的光被位于样品下方的准直镜收集并通过光纤传输至USB2000+型便携式光谱仪(Ocean Optics公司),再通过USB数据线输入计算机,为确保只有通过样品的光射入光谱仪的光纤探头,光纤探头周围采用遮光幕帘阻挡外部光线干扰。

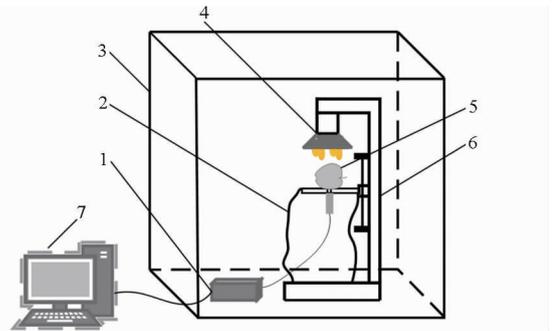


图1 可见近红外透射能量光谱采集系统

Fig. 1 Acquisition system of visible/near infrared transmission energy spectra

1. 便携式光谱仪
2. 遮光幕帘
3. 采集暗箱
4. 光源
5. 苹果样品
6. 高度可调载物台
7. 计算机

采集软件为光谱仪自带的SpectraSuite光谱采集软件。导出数据后用Unscrambler X 10.3(CAMO公司)和Matlab 2014a(MathWorks公司)进行数据分析处理。

### 1.3 光谱数据采集方法

为保证系统稳定,采集光谱数据前,提前15 min打开光源预热,在光谱数据采集软件中设置主要采集参数如下:积分时间100 ms,平均扫描次数10,平滑点数5,勾选去除暗噪声,光谱仪的光谱范围为200~1100 nm,分辨率为2.0 nm。

光谱数据采集时,将样品置于暗室内载物台上,苹果果轴与光源检测器垂直,光线从其赤道处穿过<sup>[12]</sup>。调整载物台使光源距离载物台表面分别为17、15、12 cm进行实验,经预备实验对比知,光源距离载物台表面12 cm最优。关闭暗箱门保证无外界光进入,考虑到霉心病的发病位置和形态均不同,故每个样品采集3个点的数据,各点之间约呈120°。取3个点光谱的平均值作为最终的建模光谱数据。在采集光谱数据期间,每间隔10个样品分别采集3次参考光谱和暗光谱数据,分别取其平均值作为最终的参考光谱和暗光谱数据,共采集20条参考光谱和20条暗光谱数据。

采集结束后,用刀从苹果中心线切开,切开的截面用Panasonic LX5GK型相机拍照(设置24~90 nm焦距, JPEG图片存储格式),记录苹果是否有霉心病。通过破坏性实验得出,样本中有健康果142个,霉心病果58个。分别从健康果和霉心病果中随机选取100和40个,共140个样本作为训练集,剩下的42个健康果和18个霉心病果作为测试集用于预测。

#### 1.4 原始光谱数据预处理

原始光谱数的信号是通过A/D转换出的未经加工的电压值,反映了光源的强度。为研究光源能量的稳定性,对所采集的20条参考光谱数据进行分析,结果显示每个波长点的相对标准偏差(RSD)在0.18%~0.93%范围内。可认为光源能量稳定,因此直接采用能量光谱计算建立模型是可行的<sup>[13]</sup>。

为了消除来自高频随机噪声、基线漂移、光散射、样本间本身成分含量信息的差异性等影响,提高模型的收敛性能,需对原始光谱数据进行预处理<sup>[14-15]</sup>。本文首先采用Savitzky-Golay平滑法进行平滑预处理,经预备实验对比知,平滑点数为50平滑效果最好;由于光谱数据采集受环境因素的影响,会影响识别模型的性能,所以将每间隔10个样品的原始光谱分别扣减去此段的暗光谱,以消除环境因素的影响,以波长 $\lambda$ 为例,其公式为

$$A_{\lambda} = S_{\lambda} - D_{\lambda} \quad (1)$$

式中  $S_{\lambda}$ ——在波长 $\lambda$ 时的样品光强

$D_{\lambda}$ ——在波长 $\lambda$ 时的暗光强

$A_{\lambda}$ ——样品光谱扣减去此段暗光谱的光强

然后,采用多元散射矫正(Multiple scattering correction, MSC)对光谱数据进行预处理。为了消除数据首段与末段产生的高频噪声,截取400~900 nm波段的光谱作为全光谱(FS)数据进行苹果霉心病识别分析。

#### 1.5 数据降维方法

定性分析时涉及高维光谱数据,高维数据会明显增加分析的复杂性,且每个样品的光谱数据存在谱峰重叠问题,导致光谱数据信息冗余,因此需要对光谱数据进行降维<sup>[16]</sup>。连续投影算法(Successive projections algorithm, SPA)是一种波长选取方法,通过对光谱数据投影映射选取原始光谱中的少数波长,能够有效消除众多波长变量之间的共线性影响,并且可尽可能多地概括样本光谱信息,降低模型的复杂度<sup>[17]</sup>。主成分分析(Principal component analysis, PCA)是一种将高维数据压缩,提取关键变量信息,从而通过较少的综合变量反映原始高维海量信息的多元统计分析方法<sup>[18]</sup>。本文采用SPA和PCA从经过预处理后的光谱数据中提取特征波长或特征变量。

#### 1.6 建模方法

近红外光谱定性分析常用方法有线性分析方法和人工神经网络等非线性方法。偏最小二乘判别法(Partial least squares discriminant analysis, PLS-DA)是一种广泛应用于光谱数据处理的多元统计数据分析方法,其目的是在自变量和因变量里寻找某些相关程度最大的成分线性组合,以便更好地解释因变量的变异信息<sup>[19-20]</sup>。误差反向传播神经网络(Error back propagation artificial neural networks, BP-ANN)是目前应用最广泛的一种模式识别的人工神经网络。由于其拥有较好的非线性映射能力、泛化能力和容错能力,对于高维光谱数据的处理,可取得更好的结果<sup>[21]</sup>。支持向量机(Support vector machine, SVM)是由CORTES等根据统计学习理论提出的原理性方法,可用于解决模式分类与非线性映射问题<sup>[22-23]</sup>。由于是分类问题,故采用C-SVC模型,最终的决策函数为

$$f(x) = \text{sgn} \left( \sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^* \right) \quad (2)$$

式中  $K(x, x_i)$ ——核函数类型,本文中采用RBF核函数

$\alpha_i^*$ ——Langrange系数

$y_i$ ——支持向量的类别标签

$b^*$ ——阈值

本文根据FS、SPA选择的特征波长和PCA提取的主成分,分别建立特征波长和主成分与PLS-DA、BP-ANN和SVM相结合的苹果霉心病快速无损识别模型,对样本进行定性判别,并进行对比分析,以确定较优的建模方法。

#### 1.7 苹果霉心病的评价及模型的评价标准

为判断苹果霉心病腐烂的程度,通过Matlab

2014a 对拍摄的图像进行处理,计算霉心病面积所占比例。经统计,样本的腐烂面积占横截面百分比  $M_d$  范围为 3%~80%。用  $M_d$  代表霉心病程度,经咨询苹果病害专家,规定  $M_d < 10\%$ 、 $10\% \leq M_d \leq 30\%$  和  $M_d > 30\%$  分别代表轻度、中度和重度霉心病。图 2 为典型图像。基于上述评价标准,58 个霉心病样本中,轻度、中度、重度样本个数分别为 23、20、15 个,因此,实验样本具有一定的代表性。本文拟通过霉心病的识别正确率,对识别模型进行评价。

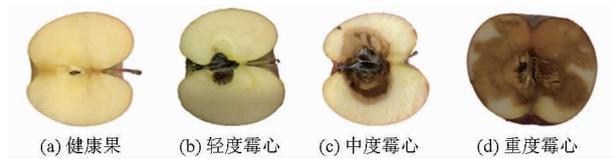


图 2 不同发病程度的苹果霉心病果样品

Fig. 2 Apple samples with different degrees of moldy core

## 2 结果与分析

### 2.1 光谱数据预处理及光谱特性分析

200 个样品的原始光谱曲线如图 3a 所示。其中,横坐标表示波长,纵坐标表示光强度,表示相应波长下光源透过苹果所发出的光通量在空间的分布密度通过 A/D 转换器转化成离散的数字量值。在 670 nm 附近有明显的叶绿素吸收峰,760 nm 附近有水或氢氧基团吸收峰(图 3a)。扣减去暗光谱后的光谱曲线及经 MSC 预处理后的光谱曲线如图 3b 和 3c 所示。由图 3c 可知,不同发病程度的光谱清楚地显示了各波长光强度的变化情况,健康果在波长 643、714、804 nm 附近有明显波峰,714 nm 附近的光强度是变化最明显的区域。健康果在 714 nm 处的透过峰最高,霉心病果的透过峰均低于健康果,且发病程度越严重,透过峰越低。进一步分析证明,受害苹果从心室开始逐渐向外扩散腐烂,病发区呈黑色或者棕灰色,对光的吸收能力较强,发病程度越重,心部腐烂面积越大,吸收能力就越强,光透过的就减少,故透过峰越低,这与王加华等<sup>[13]</sup>的研究结论相吻合。由于苹果发生霉心病病态后,无论腐烂

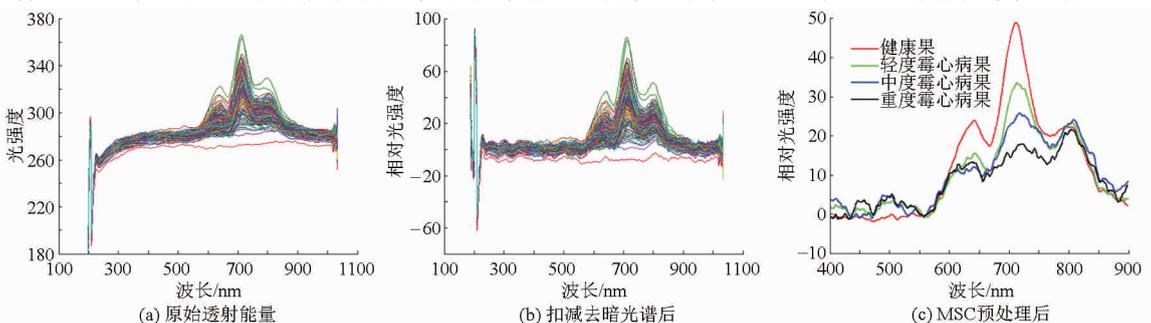


图 3 可见/近红外透射能量光谱曲线

Fig. 3 Visible/near infrared transmission energy spectra curves

程度多大,被消费者食用后都会对身体产生伤害。因此,本文仅将苹果样本分为霉心病样本和健康样本 2 类<sup>[10]</sup>。

### 2.2 光谱数据降维

#### 2.2.1 应用 SPA 提取特征波长

对随机划分后的光谱数据进行 SPA 分析,经交叉验证得到的均方根误差和选定的 12 个特征波长分布如图 4 所示。当选取 12 个特征波长时,均方根误差最小为 0.244 41(图 4a)。因此本文以 12 个特征波长的相对光强度作为模型的输入量。为验证所选波长的有效性,将 12 个特征波长的相对光强度作为 PLS-DA、BP-ANN 和 SVM 模型的输入值,建立其模型并对测试集样本进行预测。12 个特征波长以圆点表示(图 4b),选取的 12 个特征波长分别为 410.66、468.19、515.03、571.1、616.26、643、679.01、713.4、768.68、811.91、836.59、899.75 nm。在波长 643、714、804 nm 附近有明显波峰处附近都选取有特征波长点,体现了 SPA 提取的优点。

#### 2.2.2 应用 PCA 提取主成分

利用 Matlab 2014a 对样本的高维数据进行主成分分析。前 13 个主成分的贡献率及累计贡献率如表 1 所示。由表 1 可以看出,第 1 主成分(PC1)、第 2 主成分(PC2)和第 3 主成分(PC3)的贡献率分别为 51.96%、31.45% 和 7.26%,三者的累计贡献率达到 90.67%。若在建立模型中使用的主成分数过少,将会出现所建模型不能反映样本被测性质与光谱信息之间的关系,出现不充分拟合现象,而使用过多的主成分数,则会影响模型的预测精度及模型的泛化能力,出现过度拟合现象。为探明最优主成分数,用随机抽取的 140 个苹果样本作为训练集,剩余的 60 个苹果样本作为测试集,分别选择 4、5、7、9 和 13 个主成分数作为模型输入的特征变量,建立 BP-ANN 识别模型并进行对比分析,所建模型对训练集和测试集中样本的正确识别率如表 2 所示。从表 2 可知,当选择的主成分数为 9 时,模型效果最优,训练集和测试集的正确识别率分别为 93.6% 和

93.3%。表明 PCA 提取的 9 个主成分能较全面地反映原始光谱中用于区分健康果和霉心病果的有效信

息,因此本文选取累计贡献率达到 97.08% 的前 9 个主成分作为模型输入的特征变量。

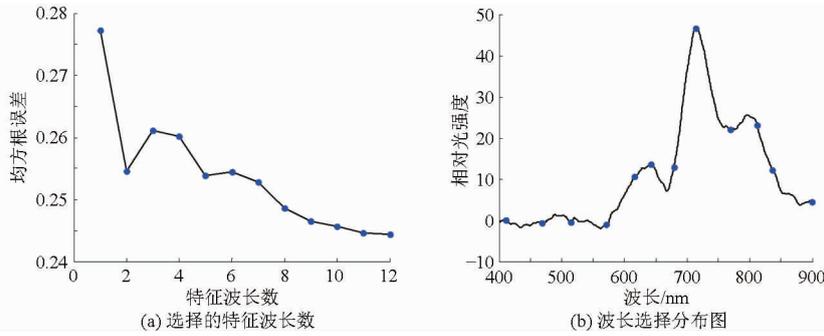


图 4 SPA 得到的均方根误差和波长选择分布图

Fig. 4 Distribution of root mean square error and selected variable wavelength obtained by SPA

表 1 PCA 前 12 个主成分的贡献率和累计贡献率

Tab. 1 Contribution rate and accumulative contribution rate of the first 12 PCA principal components

参数	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
贡献率/%	51.96	31.45	7.26	2.79	1.07	0.78	0.63	0.59	0.55	0.43	0.41	0.37	0.26
累计贡献率/%	51.96	83.41	90.67	93.46	94.53	95.31	95.94	96.53	97.08	97.51	97.92	98.29	98.55

表 2 苹果样本的 BP 结果分析

Tab. 2 BP modeling results of apples samples

主成分数	BP 网络隐含	训练集识别	测试集识别
	层个数	正确率/%	正确率/%
4	7	90.0	88.3
5	12	91.4	88.3
7	6	91.4	90.0
9	11	93.6	93.3
13	12	93.3	90.0

### 2.3 建模方法与结果

为建立鲁棒性好的识别模型,分别用 PLS-DA、BP-ANN 和 SVM 方法建立识别模型并进行实验分析。

#### 2.3.1 建模参数的选择与设定

(1)PLS-DA 参数选择:用 PLS-DA 建立模型时,主因子数个数至关重要,当主因子数较少时,无法反映样本的光谱特性,造成模型精度下降,而主因子数过多时,又会引入干扰噪声。为选择最优主因子数,本文采用交叉验证的方法来选定 PLS-DA 模型的最佳因子数,针对 FS、SPA 和 PCA 确定的 PLS-DA 主因子数如表 3 所示。

(2)BP-ANN 参数选择:经预实验知,本文采用单隐含层结构 BP 网络构建模型,包括输入层、隐含

表 3 PLS-DA 和 SVM 建模参数的设定

Tab. 3 Modeling parameters of PLS-DA and SVM

特征变量 提取方法	PLS 主因子数	SVM	
		惩罚参数 $c$	核函数参数 $g$
FS	10	1.5	0.047 4
PCA	8	4.0	0.000 5
SPA	7	4.0	0.011 0

层和输出层。本文通过经验公式,并结合试凑法确定最佳隐节点数,隐节点数的经验公式为

$$m = \sqrt{n + l} + \alpha \quad (3)$$

式中  $n$ ——输入层节点数

$l$ ——输出层节点数

$\alpha$ ——1 ~ 10 之间的常数

在此基础上网络隐含层神经元传递函数采用 ‘tansig’,输出层传递函数采用 ‘purelin’,训练函数选择 ‘trainlm’ (LM 训练法),网络训练的目标误差为 0.001,训练最大步数设为 1 000。结合经验公式范围内隐节点数的网络预测性能确定的网络结构与参数如表 4 所示。

表 4 BP-ANN 的网络结构与参数

Tab. 4 Parameters and structure of BP-ANN

参数	FS - BP	SPA - BP	PCA - BP
输入层节点数	1 210	12	9
隐含层节点数	5	14	11
输出层节点数	1	1	1
训练次数	5	12	7
均方误差	0.13	0.067	0.072

(3)SVM 参数选择:用 SVM 作分类预测时,需考虑核函数和相关参数(主要是惩罚参数  $c$  和核函数参数  $g$ )对模型性能的影响。本文选用 RBF 函数作为核函数。利用 K-折交叉验证法(K-fold cross validation, K-CV)的粗选和精选 2 个步骤优化  $c$  和  $g$  参数,参数优化结果如图 5 所示。针对 FS、SPA 和 PCA 确定的 SVM 最佳参数组合  $c$  和  $g$  如表 3 所示。

#### 2.3.2 建模结果

基于随机法划分的训练集样本的透射能量光谱

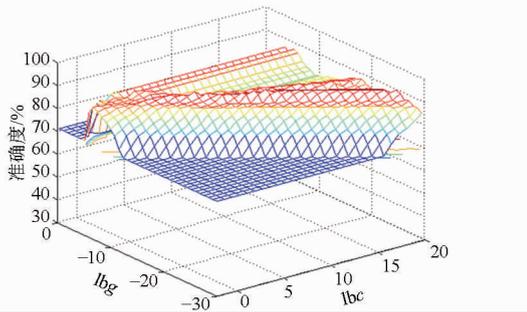
图5 SVM模型中参数 $c$ 和 $g$ 寻优过程Fig. 5 Optimization of parameters  $c$  and  $g$  for SVM

表5 霉心病果与健康果的识别正确率

Tab. 5 Identification accuracy of moldy core and healthy apples

建模方法	特征变量提取方法	运行时间/ s	训练集			测试集		
			误判数/个		识别	误判数/个		识别
			霉心病果(40个)	健康果(100个)	正确率/%	霉心病果(18个)	健康果(42个)	正确率/%
PLS-DA	FS	16.6	6	9	89.3	3	4	88.3
	SPA	8.7	2	12	90.0	1	5	90.0
	PCA	7.4	4	10	90.0	4	3	88.3
BP-ANN	FS	59.6	2	4	95.7	3	5	86.7
	SPA	7.7	4	2	95.7	3	1	93.3
	PCA	9.8	6	3	93.6	4	0	93.3
SVM	FS	2 392.5	7	13	85.7	4	6	83.3
	SPA	17.9	2	2	97.1	2	0	96.7
	PCA	12.6	0	1	99.3	2	0	96.7

模型对训练集和测试集中霉心病果和健康果的识别正确率分别为 89.3% 和 88.3%，分别有 15、7 个样本发生误判。SPA - PLS-DA 模型对训练集和测试集中样本的识别正确率相等，均为 90.0%，分别有 14、6 个样本发生误判。PCA - PLS-DA 模型对训练集和测试集中样本的识别正确率分别为 90.0% 和 88.3%，分别有 14、7 个样本发生误判。在 BP-ANN 模型中，FS - BP-ANN 模型对训练集和测试集中霉心病果和健康果的识别正确率分别为 95.7% 和 86.7%，分别有 6、8 个样本发生误判。基于 SPA 和 PCA 所建模型对训练集中样本的识别正确率分别为 95.7% 和 93.6%，分别有 6、9 个样本发生误判，对于测试集，基于 SPA 和 PCA 所建模型的识别正确率相等，均为 93.3%，误判个数均为 4 个。在 SVM 模型中，FS - SVM 模型对训练集和测试集中霉心病果和健康果的识别正确率分别为 85.7% 和 83.3%，分别有 20、10 个样本发生误判。SPA - SVM 模型对训练集和测试集中样本的识别正确率分别为 97.1% 和 96.7%。PCA - SVM 模型对训练集和测试集中样品的识别正确率分别为 99.3% 和 96.7%。预测时都仅有 2 个霉心病果误判为健康果。

### 3 讨论

可见近红外透射能量光谱定性分析时涉及高维光谱数据，这会增加分析的复杂性，且每个样品的光

数据，将 FS、经 SPA 提取的特征波长以及 PCA 提取的主成分分别作为 PLS-DA、BP-ANN 以及 SVM 模型的输入变量，建立霉心病果与健康果的识别模型，并用测试集样本对模型进行预测。理论上，输出为 1 为健康果，输出为 0 则为霉心病果，判别阈值设为 0.5，输出大于 0.5 的样品为健康果，输出小于 0.5 的样品为霉心病果。所建模型对训练集和测试集中样本的正确识别率如表 5 所示。

由表 5 可知：在 PLS-DA 模型中，FS - PLS-DA

谱数据存在谱峰重叠问题，导致光谱数据信息冗余，网络运算很容易陷入局部极小且解析困难，本文对光谱数据进行平滑和 MSC 预处理后，基于 FS、SPA 和 PCA 建立识别模型。从识别结果分析得出，9 种模型均没有欠拟合及过拟合现象发生，说明可见近红外透射能量光谱分析方法对于苹果霉心病的识别是可行的。

为分析不同建模方法的可靠性和适应性，首先对比基于 FS、SPA 和 PCA 方法所建模型用于苹果霉心病识别，基于 SPA 和 PCA 所建模型运行速度明显比基于 FS 所建模型运行速度快很多。此外，基于 SPA 和 PCA 所建模型的输入变量数分别只有 12 和 9 个，仅相当于基于 FS 所建模型输入变量数的 0.99% 和 0.74%，极大降低了模型的复杂度，运算速度明显加快，更能适应现场实时检测的应用要求。

对比 PLS-DA、BP-ANN 和 SVM 模型用于苹果霉心病识别，对于 SVM 模型，能较好地解决小样本、高维数、非线性和局部极小点等实际问题，具有较好的鲁棒性，且采用 K-CV 优化的思想可以得到最优的参数，并且有效避免过学习和欠学习状态的发生。其模型性能从优到劣依次为 SPA - SVM、SPA - BP-ANN、SPA - PLS-DA、PCA - SVM、PCA - BP-ANN、PCA - PLS-DA。SVM 模型对训练集和测试集中样品的识别正确率均分别高于 97.1% 和 96.7%，可见 SVM 模型的性能要优于 BP-ANN 和 PLS-DA 模型，

在实际应用中也更加稳定和可靠。

SPA-SVM 和 PCA-SVM 模型综合分析可知, SPA-SVM 和 PCA-SVM 模型对测试集的识别正确率均达到了 96.7%, 但对于训练集, PCA-SVM 模型的识别正确率为 99.3%, 要略高于 SPA-SVM 模型的 97.1%。说明 PCA 提取出了更全面地用于区分霉心病果和健康果有效信息, 而 SPA 提取的特征波长漏掉了一些有用信息。

与国内外研究结果比较可知, 与李顺峰等<sup>[11]</sup>研究相比, 模型训练集和测试集的正确识别率分别提高了 9.4% 和 8.9%, 由于其通过光谱漫反射检测方式判别苹果霉心病害, 而霉心病早期发生在果心及其附近, 通过近表面光谱漫反射的检测方式难以反映果心及深层果肉的特征信息, 这表明透射光谱更适合苹果霉心病的检测; 与韩东海等<sup>[10]</sup>研究相比, 识别正确率提高了 1.05%, 其判别方法采用线性判别方法, 本文建立多个模型进行对比分析, 以确定较优的建模方法, 因此模型的先进性和准确率更高; 与 SHENDEREY 等<sup>[12]</sup>研究相比, 识别正确率提高了 6.69%。

综合可知 PCA-SVM 模型具有更高的学习能力和预测能力, 能更好地解决苹果霉心病的模型识别推广能力的问题。对误判霉心病果的发病面积进行考察, 发现误判霉心病果的腐烂面积均低于

10%, 临界于健康果与病果之间, 这种发病面积小且霉菌附在水果心室里给检测带来了一定的难度。

## 4 结论

(1) 提出了一种基于可见近红外透射能量光谱进行苹果霉心病快速无损识别的模型和方法。不同发病程度的光谱清楚地显示了各波长光强度的变化情况, 健康果在 714 nm 处的透过峰最高, 霉心病果的透过峰均低于健康果, 且发病程度越重, 透过峰越低。这为苹果霉心病的光谱判别提供了理论依据。

(2) 不同光谱预处理方法对建模结果有影响。分别建立了基于 FS、SPA 和 PCA 的 PLS-DA、BP-ANN、SVM 识别模型。基于 SPA 和 PCA 所建模型的输入变量数仅相当于基于 FS 所建模型输入变量数的 0.99% 和 0.74%, 极大降低了模型的复杂度。

(3) 分析比较 9 种建模方法的识别结果, 发现 PCA-SVM 模型具有更高的学习能力和预测能力, 对训练集和测试集中的霉心病果和健康果识别正确率分别为 99.3% 和 96.7%。研究结果表明, 基于 PCA-SVM 方法应用于苹果霉心病的快速无损识别中是可行的且具有较高的识别准确度, 为苹果在线内部品质分级和便携式苹果霉心病检测仪的研究与开发提供了技术依据。

## 参 考 文 献

- 1 朱从会, 师俊玲, 杨保伟, 等. 展青霉素产生菌拮抗放线菌的分离、筛选与初步鉴定[J]. 农业机械学报, 2008, 39(3): 86-89.  
ZHU Conghui, SHI Junling, YANG Baowei, et al. Isolation screening and identification of antagonistic actinomycetes having ability in inhibiting patulin-producing penicillium[J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(3): 86-89. (in Chinese)
- 2 乌日娜, 尚洁. 水果及其制品中棒曲霉素的残留分析进展[J]. 中国国境卫生检疫杂志, 2007, 30(3): 188-190.  
WU Rina, SHANG Jie. Development of analyzing residual patulin in fruit and its product[J]. Chinese Frontier Health Quarantine, 2007, 30(3): 188-190. (in Chinese)
- 3 SCHATZKI T F, HAFF R P, YOUNG R, et al. Defect detection in apples by means of x-ray imaging[J]. Transactions of the ASAE, 1997, 40(5): 1407-1415.
- 4 CLARK C J, BIELESKI R L. Application of NMR imaging to detection of watercore in Fuji apple post-harvest[C]// Proceedings from the Sensors for Nondestructive Testing, 1997: 211-220.
- 5 李芳, 蔡骋, 马惠玲, 等. 基于生物阻抗特性分析的苹果霉心病无损检测[J]. 食品科学, 2013, 34(18): 197-202.  
LI Fang, CAI Cheng, MA Huiling, et al. Nondestructive detection of apple mouldy core based on bioimpedance properties[J]. Food Science, 2013, 34(18): 197-202. (in Chinese)
- 6 杨亮亮. 基于机器视觉和 X 射线的苹果霉心病检测方法研究[D]. 杨凌: 西北农林科技大学, 2009.
- 7 VISCARRA R A, WALVOORT D J J, MCBRATNET A B, et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties[J]. Geoderma, 2006, 131(1-2): 59-75.
- 8 CLARK C J, MCGLONE V A, JORDAN R B. Detection of brownheart in 'Braeburn' apple by transmission NIR spectroscopy[J]. Postharvest Biology and Technology, 2003, 28(1): 87-96.
- 9 MCGLONE V A, MARTINSEN P J, CLARK C J, et al. On-line detection of brownheart in braeburn apples using near infrared transmission measurements[J]. Postharvest Biology and Technology, 2005, 37(2): 142-151.
- 10 韩东海, 刘新鑫, 鲁超, 等. 苹果内部褐变的光学无损检测研究[J]. 农业机械学报, 2006, 37(6): 86-93.  
HAN Donghai, LIU Xinxin, LU Chao, et al. Study on optical-nondestructive detection of breakdown apples[J]. Transactions of the Chinese Society for Agricultural Machinery, 2006, 37(6): 86-93. (in Chinese)

- 11 李顺峰,张丽华,刘兴华,等. 基于主成分分析的苹果霉心病近红外漫反射光谱判别[J]. 农业机械学报,2011,42(10): 158-161.  
LI Shunfeng, ZHANG Lihua, LIU Xinghua, et al. Discriminant analysis of apple moldy core using near infrared diffuse reflectance spectroscopy based on principal component analysis [J]. Transactions of the Chinese Society for Agricultural Machinery,2011, 42(10): 158-161. (in Chinese)
- 12 SHENDEREY C, SHMULEVICH I, ALCHANATIS V, et al. NIRS detection of moldy core in apples[J]. Food and Bioprocess Technology, 2010, 3(1): 79-86.
- 13 王加华,孙旭东,潘璐,等. 基于可见/近红外能量光谱的苹果褐腐病和水心鉴别[J]. 光谱学与光谱分析,2008,28(9): 2098-2102.  
WANG Jiahua, SUN Xudong, PAN Lu, et al. Discrimination of brownheart and watercore of apples based on energy spectrum of visible/near infrared transmittance[J]. Spectroscopy and Spectral Analysis, 2008, 28(9): 2098-2102. (in Chinese)
- 14 刘燕德,周延睿,潘圆媛. 基于最小二乘支持向量机的辣椒可溶性固形物和维生素C含量近红外光谱检测[J]. 光学精密工程,2014,22(2):281-288.  
LIU Yande, ZHOU Yanrui, PAN Yuanyuan. Determination of soluble solid contents and vitamin C of fresh peppers based on NIR spectroscopy and least square support vector machines[J]. Optics and Precision Engineering, 2014, 22(2): 281-288. (in Chinese)
- 15 李江波,赵春江,陈立平,等. 基于可见/近红外光谱谱区有效波长的梨品种鉴别[J]. 农业机械学报,2013,44(3):153-157,179.  
LI Jiangbo, ZHAO Chunjiang, CHEN Liping, et al. Variety identification of pears based on effective wavelengths in visible/near infrared region [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(3): 153-157, 179. (in Chinese)
- 16 岳学军,全东平,洪添胜,等. 柑橘叶片叶绿素含量高光谱无损检测模型[J]. 农业工程学报,2015,31(1):294-302.  
YUE Xuejun, QUAN Dongping, HONG Tiansheng, et al. Non-destructive hyperspectral measurement model of chlorophyll content for citrus leaves[J]. Transactions of the CSAE, 2015, 31(1): 294-302. (in Chinese)
- 17 何东健,陈煦. 基于光谱分析的土壤有机质含量田间实时测定方法[J]. 农业机械学报,2015,46(2):127-132.  
HE Dongjian, CHEN Xu. Research on soil organic matter content measurement by spectral analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(2): 127-132. (in Chinese)
- 18 郭文川,刘大洋. 猕猴桃膨大果的近红外漫反射光谱无损识别[J]. 农业机械学报,2014, 45(9):230-235.  
GUO Wenchuan, LIU Dayang. Identification of expanded kiwifruit by near-infrared diffused spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(9): 230-235. (in Chinese)
- 19 KAYA-CELIKER H, MALLIKARJUNAN P K, KAYA A. Mid-infrared spectroscopy for discrimination and classification of *Aspergillus spp.* contamination in peanuts[J]. Food Control, 2015, 52: 103-111.
- 20 许文丽,孙通,吴文强,等. 可见/近红外漫反射与漫透射检测脐橙可溶性固形物的对比研究[J]. 光子学报,2013, 42(12):1486-1490.  
XU Wenli, SUN Tong, WU Wenqiang, et al. Contrast of vis/NIR reflection and diffuse transmission in detection the SSC of Gan'an navel oranges[J]. Acta Photonica Sinica,2013,42(12):1486-1490. (in Chinese)
- 21 栾郭宏,贺凯迅,程辉,等. 基于神经网络的近红外光谱辛烷值模型的研究及应用[J]. 计算机与应用化学,2014,31(1): 63-68.  
LUAN Guohong, HE Kaixun, CHENG Hui, et al. Octane model based on neural network by near-infrared spectroscopy and its application[J]. Computers and Applied Chemistry, 2014, 31(1): 63-68. (in Chinese)
- 22 CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.
- 23 祝志慧,谢德君,李婉清,等. 基于光谱技术和多分类器融合的异物蛋检测[J]. 农业工程学报,2015,31(2):312-318.  
ZHU Zhihui, XIE Dejun, LI Wanqing, et al. Abnormal eggs detection based on spectroscopy technology and multiple classifier fusion[J]. Transactions of the CSAE, 2015, 31(2): 312-318. (in Chinese)