

# 高光谱图像技术检测枇杷叶片三萜酸含量及叶面分布

邹小波, 徐艺伟, 陈武, 石吉勇, 朱瑶迪, 申婷婷

(江苏大学食品与生物工程学院, 江苏镇江 212013)

**摘要:** 枇杷叶富含三萜酸类化合物, 具有较高的药用价值。本研究首先建立枇杷叶高光谱信号与三萜酸含量的对应关系, 然后利用高光谱图像包含每个像素点的光谱信号这一独特优势, 检测枇杷叶片的三萜酸分布。通过联合区间偏最小二乘法 (siPLS) 建立三萜酸含量分析模型, 结果表明, 采用 siPLS 将全光谱均匀划分 11 个子区间, 选择 1、5、6、7 联合, 主因子数为 4, 建立的 siPLS 谱区筛选模型预测效果最佳, 其交互验证均方根误差 (RMSECV) 和预测均方根误差 (RMSEP) 分别为 3.392 mg/g 和 3.731 mg/g, 校正集和预测集相关系数分别为 0.8449 和 0.8223。根据 siPLS 模型计算叶片上所有像素点处的三萜酸含量, 并通过伪彩色方法描述叶片中三萜酸含量的分布。研究结果表明, 利用高光谱图像技术分析枇杷叶片三萜酸含量及叶面分布是可行的。

**关键词:** 三萜酸分布; 高光谱图像; 联合区间偏最小二乘; 枇杷叶片

文章编号: 1673-9078(2015)7-263-267

DOI: 10.13982/j.mfst.1673-9078.2015.9.043

## Measurement of Triterpenic Acid Content and Its Distribution in Loquat Leaves Based on the Hyperspectral Imaging Technique

ZOU Xiao-bo, XU Yi-wei, CHEN Wu, SHI Ji-yong, ZHU Yao-di, SHEN Ting-ting

(School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

**Abstract:** Triterpenic acids are abundant in Loquat leaves, which present a relatively high medicinal value. The relative relationship between hyperspectral signal and triterpenic acid content was established and the distribution of triterpenic acids in loquat leaves was measured using hyperspectral images, as they contain the spectral signal for each pixel. Synergy interval partial least squares (siPLS) was used to establish an analytical model of triterpenic acid content in loquat leaves. The results showed that the optimal model was obtained by the siPLS spectral region screening method, which separated the whole spectrum into 11 intervals and combined four intervals, including intervals 1, 5, 6, and 7, and had a factor number of 4. The model achieved a correlation coefficient of 0.8449 and root mean square error of cross validation (RMSECV) of 3.392 mg/g in the calibration set and a correlation coefficient of 0.8223 and root mean square error of prediction (RMSEP) of 3.73 mg/g in the prediction set. Every pixel in the hyperspectral image was computed and the triterpenic acid content of each pixel was obtained according to the siPLS model, followed by the creation of a pseudo-color map to describe the distribution of triterpenic acids in loquat leaves. Overall, the results sufficiently demonstrate that it is feasible to employ hyperspectral imaging technology to measure the content of triterpenic acid and its distribution in loquat leaves.

**Key words:** distribution of triterpenic acids, hyperspectral image, synergy interval partial least squares, loquat leaf

三萜酸广泛存在于植物界中, 具有减肥、降血糖、抗癌、抗肿瘤、抗肝损伤、抗菌消炎和抗病毒等多种

收稿日期: 2014-11-06

基金项目: 高新技术发展计划国家 863 项目 (2011AA100807); 全国优秀博士基金资助项目 (200968); 国家自然科学基金 (BK20130505); 新世纪优秀人才项目 (NCET-11-00986); 杰出青年基金 (BK20130010); 江苏省研究生创新基金 (KYLX\_1070)

作者简介: 徐艺伟 (1991-), 男, 研究生, 研究方向: 食品品质无损检测

通讯作者: 邹小波, 男, 教授, 博士生导师, 主要从事农产品、食品品质无损检测研究

生物学活性<sup>[1]</sup>, 三萜酸含量与中药材的品质相关。

目前, 枇杷叶中三萜酸的分析检测方法有薄层析法、高效液相色谱法、毛细管电泳法、分光光度法等, 这些方法均需要破坏叶片组织结构, 并且费时费力。为了克服常规检测手段的不足, 李蕾蕾<sup>[2]</sup>等人研究了基于近红外光谱的枇杷叶熊果酸检测方法, 得到了枇杷叶熊果酸含量和叶片吸收光谱的定量分析模型。

化学检测方法和近红外光谱检测方法只能得到采样区域的三萜酸含量情况, 难以得到三萜酸在整片

枇杷叶面上的分布情况。为了实现三萜酸叶面分布的检测,本研究提出基于高光谱图像技术的检测方法。三萜酸叶面分布图显示的结果非常直观,凭借肉眼可以直接观察三萜酸在叶片不同区域的分布情况,可以节省大量的分析时间,适用于大批量枇杷叶药材中三萜酸的快速测定。

高光谱图像不仅包含样品在每个波段下的二维图像信息,还包含样本每个像素点的光谱信息。因此,高光谱图像技术既能对研究对象的外部特征进行可视化分析,也能对内部有效成分进行定量预测。高光谱检测作为一种能多次重复对样本进行无损、快速检测的方法,越来越受到研究人员关注。利用高光谱技术检测植物中生化参量方面的研究主要包括叶片色素及叶绿素含量估测<sup>[3]</sup>、植被氮素含量测定<sup>[4]</sup>、植物中多糖含量检测<sup>[5]</sup>等。

本研究采用联合区间偏最小二乘法优化光谱特征区间,建立枇杷叶高光谱图像信号和三萜酸含量的对应关系,根据此关系,计算高光谱图像中所有像素点对应的三萜酸浓度值,将三萜酸浓度值取代对应像素点的图像信号,从而得到整个叶面的三萜酸浓度分布图。

## 1 材料与方法

### 1.1 实验材料

实验用枇杷叶采集于江苏大学校园,为了使收集的样本具有一定代表性,选择不同生长环境的不同植株进行样本采集,共采集 90 片。

### 1.2 高光谱图像的获取和标定

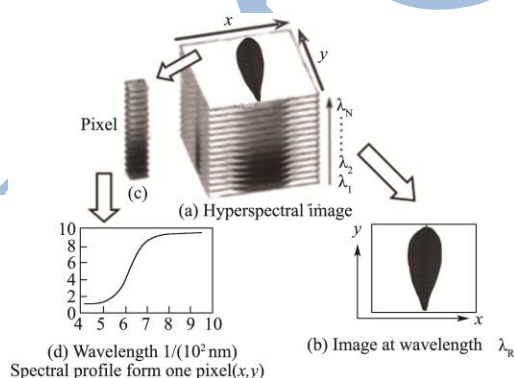


图 1 高光谱图像数据块

Fig.1 Hyperspectral imaging datacube

用基于图像光谱仪的高光谱图像采集系统采集枇杷叶片的高光谱信息。该系统由基于图像光谱仪的高光谱摄像机 (ImSpector, V10E, 芬兰), 一套 150W 光纤卤素灯 (Fiber-Lite DC950 Illuminator, DolanJenner

Industries Inc, MA, 美国), 一套包括控制装置的移动平台 (Zolix, SC30021A, 北京) 和计算机等部件组成。按照参考文献[6]的方法采集枇杷叶片的高光谱图像。

高光谱图像数据如图 1 所示。其中 a 为高光谱图像数据块的立体示意图, x、y 轴代表图像的空间位置, λ 轴代表光谱的波长。当 λ 取某个固定值时, 可以得到 λ 波长下的图像数据, 如图 1b 所示; 当 x、y 取固定值(即 a 中某个像素点)时, 可以得到单个像素点的数据块如图 1c 所示。将图 1c 中 λ 轴上所有点的图像信号值和对应的 λ 值表示成散点图, 得到像素点对应的光谱图 1d。从图 1 可见, 高光谱图像数据的特点为既有各个波长下的图像信息, 也有各个像素点对应的的光谱信息。高光谱图像数据的这一特性, 为研究三萜酸浓度叶面分布奠定了基础。

由于光源的强度分布不均匀及暗电流噪音的存在, 造成在光源强度分布较弱的波段下获得的图像含有较大的噪音。因此, 需要对所获得的高光谱图像进行黑白标定。在与样品采集相同的系统条件下, 扫描标准白色校正板得到全白的标定图像 W, 关闭相机快门进行图像采集得到全黑的标定图像 B, 完成高光谱图像的标定, 使采集得到的绝对图像 I 变成相对图像 R。

$$R_{\lambda} = (I_{\lambda} - B_{\lambda}) / (W_{\lambda} - B_{\lambda})$$

式中:  $R_{\lambda}$ 、 $I_{\lambda}$ 、 $B_{\lambda}$  及  $W_{\lambda}$  分别表示在波长 λ 下相对图像、绝对图像、黑板标定图像及白板标定图像。

### 1.3 三萜酸含量参考值的测定

完成光谱采集后, 采用化学法测定枇杷叶片三萜酸含量, 并将其作为建模的标准参照值。从矩形感兴趣区域(ROI, region of interest)选取枇杷叶片组织, 将其在 40 °C 烘干, 粉碎过 40 目得叶片干粉, 称取 0.5 g 枇杷叶干粉, 加入 20 mL 氯仿, 超声振荡处理 40 min, 中间隔 10 min, 静置过夜。过滤, 取滤液 50 μL, 60 °C 水浴挥发溶剂, 加入 5% 香草醛-冰醋酸 0.2 mL, 高氯酸 0.8 mL, 60 °C 水浴 10 min, 冷却后加入 5 mL 的冰醋酸。以试剂空白作为对比, 在 548 nm 处测定吸光值, 最后依据熊果酸浓度与吸光度的关系计算出总的三萜酸含量<sup>[7]</sup>。

## 2 结果与讨论

### 2.1 三萜酸含量参考测定结果

三萜酸含量的测定参照 1.3。一共采集了 90 个叶片样本。其中 60 个样本作为校正集用于模型校正, 另

外一组 30 个样本为预测集用于模型验证。表 1 分别列出了校正集和预测集三萜酸含量的均值、范围和标准偏差。

表 1 校正集和预测集的枇杷叶片三萜酸含量 (mg/g)

Table 1 Triterpenic acid content in loquat leaves in the calibration and validation sets

样本数	三萜酸质量		标准偏差
	分数范围	量分数均值	
校正集	60	8.83~34.96	6.36
预测集	30	9.07~34.27	6.29

取一批枇杷叶样本, 将其叶脉与叶肉分离, 分别检测叶脉与叶肉中的三萜酸含量。表 2 列出了叶脉与叶肉中三萜酸含量的检测结果。

表 2 枇杷叶片叶脉与叶肉中三萜酸含量

Tab 2 Triterpenic acid content in loquat mesophyll and vein

样品	三萜酸含量均值 (mg/g)
枇杷叶脉	27.76
枇杷叶肉	16.51

## 2.2 高光谱图像的分析

选取每片叶子上 200×200 像素的矩形感兴趣区域 (ROI, region of interest) 作为研究对象, 均选取叶片中部主叶脉两侧叶肉区域, 三萜酸含量真实值的测定区域均与所选取的 ROI 区域对应。通过计算 ROI 的各个像素点的光谱响应平均值来估算枇杷叶片的相对反射率。90 个样品共获得 90 条平均反射光谱用于分析。

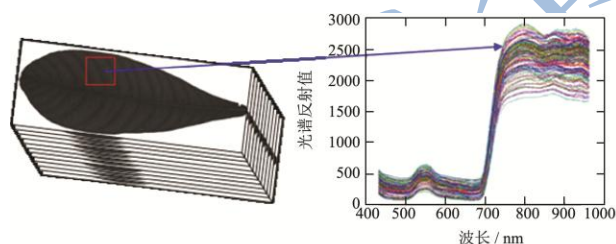


图 2 枇杷叶片高光谱图像部分区域的光谱曲线

Fig.2 Spectral profile from the hyperspectral image of some regions in the loquat leaf samples

枇杷叶高光谱图像的光谱信息提取区域和对应的光谱信息如图 2 所示。从图 2 可以看出, 光谱在波长 431~450 nm、850~962 nm 光谱范围内存在较大噪声。为了避免信噪比过低, 选取 450~850 nm 光谱范围, 总共 467 个波长的高光谱数据进行分析。随机将 90 个样本分成校正集和预测集两组, 一组 60 个样本, 另一组 30 个样本, 即校正集和预测集分别为 60×467 和 30×467 的矩阵。

## 2.3 建模及预测分析

枇杷叶高光谱数据经标准正交变换 (SNV) 预处理后, 用全光谱 (450~850 nm) 数据与三萜酸浓度进行偏最小二乘 (PLS) 建模, 当 PLS 主因子数为 7 时, 模型的均方根误差达到最小, 三萜酸浓度预测值与实测值的散点图如图 3 所示, 校正集相关系数为 0.7565, RMSECV 为 4.182 mg/g。

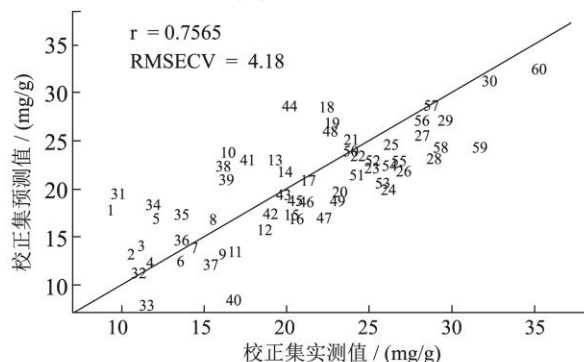


图 3 校正集 PLS 模型精度预测值与实测值的散点图

Fig.3 Predicted vs. actual in calibration model (PLS)

采用的联合区间偏最小二乘法是一种波长筛选法的改进, 该法主要用于筛选偏最小二乘法建模的波长区域。在区间偏最小二乘法 (iPLS) 建模基础上, 将同一次区间划分中精度较高的几个局部模型所在的子区间联合起来, 以联合模型的 RMSECV 值为各模型的精度衡量标准, 在所有模型中选出性能最佳者 (RMSECV 最小), 其所对应的子区间组合即为最佳组合。在最佳组合区间上所建立的联合区间偏最小二乘模型预测能力最强<sup>[8]</sup>。

表 3 分别列出了将光谱划分为 10, 11, ..., 20 个子区间后, 分别联合 2, 3, 4 个区间时, 各最优联合模型的建模区间。从表 3 可以看出将全光谱划分为 11 个子区间时, 建立于第 1、5、6、7 号 4 个子区间 (450.09~484.25 nm、592.48~628.45 nm、629.30~665.46 nm、666.32~701.78 nm) 之上的模型是最佳的。其预测能力不仅超过了全光谱模型, 更重要的是, 该模型得到了很大的简化, 其实际采用的波长个数为 170 个, 比全光谱模型少了 63.6%。这 170 个波长最终组合为 4 个主因子, 比全光谱模型少了 3 个。

熊果酸和齐墩果酸是枇杷叶中含量最高的三萜酸<sup>[9]</sup>, 两者为互同分异构体, 具有相同的官能团, 因此具有相近的红外光谱特征吸收峰。熊果酸在红外光谱中有羟基强峰 3438 cm<sup>-1</sup> 和 2940 cm<sup>-1</sup><sup>[10]</sup>, 其中 3438 cm<sup>-1</sup> 的 6 倍位于区间 450.09~484.25 nm 上, 2930 cm<sup>-1</sup> 的 5 倍位于区间 592.48~701.78 nm 上, 如图 4 所示。

证明将全光谱划分为 11 个子区间, 第 1、5、6、7 个子区间是三萜酸基本结构吸收光谱对应的特征区间。

表 3 选择不同光谱区域的 siPLS 模型的统计结果

Table 3 Results of siPLS calibration model with different spectra

划分 区间 个数	联合 2 个区间				联合 3 个区间				联合 4 个区间			
	主因 子数	入选 区间	RMS ECV	RMSEP	主因 子数	入选 区间	RMS ECV	RMSEP	主因 子数	入选 区间	RMS ECV	RMSEP
10	3	4、6	4.582	4.913	5	1、4、5	4.077	4.191	6	1、2、4、5	3.733	4.398
11	4	1、5	4.315	5.091	4	1、4、6	3.45	4.033	4	1、5、6、7	3.392	3.731
12	7	1、2	4.129	5.669	5	1、3、6	3.848	4.027	6	1、5、6、8	3.755	3.812
13	8	1、2	4.219	5.491	5	1、3、7	3.583	4.002	6	1、3、5、7	3.53	4.609
14	4	1、3	4.313	3.954	5	1、6、9	3.785	4.436	6	1、3、5、8	3.523	4.933
15	7	1、2	4.475	5.603	5	1、3、5	4.052	4.512	6	1、3、5、9	3.767	4.112
16	6	1、10	4.482	4.522	6	1、3、5	3.776	4.713	7	1、3、5、8	3.727	4.310
17	4	1、4	4.672	4.514	6	1、7、11	4.052	4.732	7	1、7、10、11	3.72	4.185
18	4	1、4	4.462	4.310	5	1、4、9	4.001	4.201	6	1、2、4、9	3.6	4.033
19	4	1、4	4.735	4.571	7	1、2、3	4.058	5.468	7	1、2、4、13	3.727	4.234
20	2	7、11	4.635	5.513	8	1、2、3	4.219	5.622	5	1、2、6、10	3.798	4.002

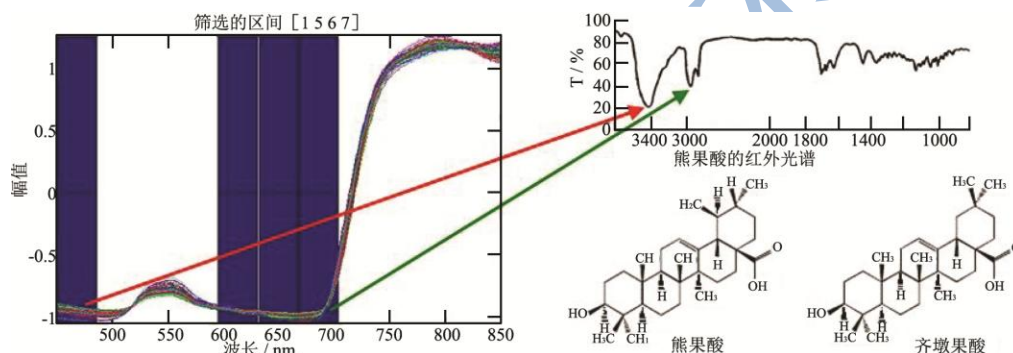


图 4 熊果酸和齐墩果酸的分子结构以及 siPLS 优选区间和熊果酸红外光谱的关系

Fig.4 Structural formulae of ursolic acid and oleanolic acid and the relationship between the intervals selected by siPLS and the IR spectra of ursolic acid

校正集和预测集 siPLS 模型三萜酸预测值与实测值的散点图如图 5, 校正集相关系数为 0.8449, RMSECV 为 3.392 mg/g, 预测集相关系数为 0.8223, RMSEP 为 3.731 mg/g。

### 2.4 三萜酸含量分布预测

高光谱数据包含高光谱图像每个像素点在各个波长下的光谱信息。而在 3.3 中, 通过 siPLS 建立了高光谱图像信号和三萜酸浓度的预测模型, 即找到了图像信号和三萜酸浓度值之间的关系。因此, 可以按如下过程检测三萜酸在整个叶面的分布: (1)采集叶片的高光谱图像, 标定后提取每个像素点对应的的光谱信息; (2)将每条光谱信息代入 2.3 中的三萜酸浓度预测模型, 计算出每个像素点对应的三萜酸浓度值; (3)将每个像素点出的三萜酸浓度值归一化并拉伸到[0,255]

范围, 得到叶面三萜酸浓度的灰度分布图; (4)将三萜酸浓度灰度分布图进行伪彩色处理, 得到颜色鲜明、直观的伪彩色分布图像。

根据上面的过程, 计算了一片枇杷叶三萜酸浓度的伪彩色分布图, 结果如图 6 所示。图中, 叶子区域中不同的颜色及颜色深浅代表了对应位置的三萜酸含量。图 6 上边的颜色标尺指明了颜色和和三萜酸浓度的对应关系。从图 6 可见, 三萜酸主要集中在叶脉附近, 该区域对应的颜色主要为黄色和橙红色, 对应的三萜酸浓度约为 30 mg/g, 叶肉部分对应的颜色主要为浅蓝色, 对应的三萜酸浓度约为 15 mg/g, 与 2.1 中分别检测叶肉和叶脉中三萜酸含量的实验结果一致, 随着距离叶脉越远, 叶肉中三萜酸浓度逐渐降低。可见, 根据预测模型所计算出来的三萜酸浓度分布图, 能较准确地反映叶面三萜酸的分布情况。

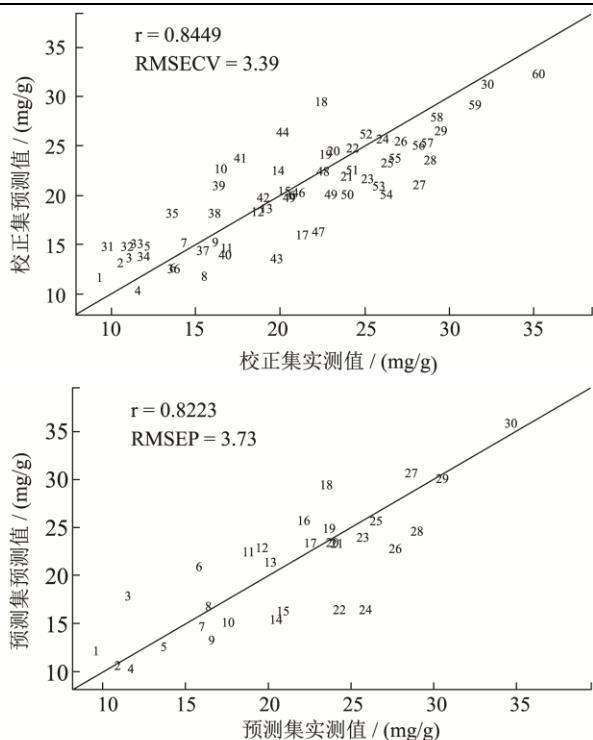


图5 校正集和预测集 siPLS 模型三萜酸浓度预测值与实测值的散点图

Fig.5 Predicted vs actual in calibration and prediction models

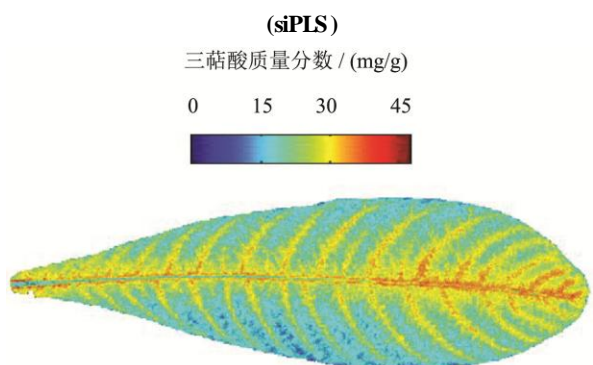


图6 基于 siPLS 模型的三萜酸含量分布

Fig.6 Distribution of triterpenic acids in the leaf based on siPLS

### 3 结论

在图像采集和图像预处理之后，提取枇杷叶片感兴趣区域的特征波长用于枇杷叶片高光谱数据同三萜酸含量关系模型的建立。将全光谱划分为 11 个子区间时，建立于第 1、5、6、7 号 4 个子区间(450.09~484.25 nm、592.48~628.45 nm、629.30~665.46 nm、666.32~701.78 nm)之上的模型是最佳的。模型对应的预测集相关系数为 0.8223，均方根误差为 3.731 mg/g。通过提取枇杷叶高光谱图像中每个像素点对应的光谱信号，结合建模的三萜酸浓度回归模型得到了叶面各个像素点对应的三萜酸浓度值，根据各个点的三萜酸浓度值，最终获得整个叶面的三萜酸浓度分布图，据

此可以轻松解读三萜酸在枇杷叶面不同部位的分布规律。本研究与常规的分析方法相比，可以节省大量的分析时间，适用于大批量枇杷叶药材中三萜酸的快速测定，为中药及其制剂的快速、有效的质量控制方法提供参考。

### 参考文献

- [1] 杨庆新,黄建安,刘仲华,等.枇杷叶中三萜酸的研究进展[J]. 食品工业科技,2008,29(3):282-285  
YAN Qing-xin, HUANG Jian-an, LIU Zhong-hua, et al. Research progress of triterpenes in loquat leaf [J]. Science and Technology of Food Industry, 2008, 29(3): 282-285
- [2] 李蕾蕾,王海霞,黄洁燕,等.近红外光谱法快速测定枇杷叶中熊果酸的含量[J].中国实验方剂学杂志,2013,19(23): 103-106  
LI Lei-lei, WANG Hai-xia, HUANG Jie-yan, et al. Rapid determination of ursolic acid content in *Eriobotrya japonica* by near-infrared spectroscopy [J]. Chinese Journal of Experimental Traditional Medical Formulae, 2013, 19(23): 103-106
- [3] Zou X B, Zhao J W, Holmes M, et al. Independent component analysis in information extraction from visible/near-infrared hyperspectral imaging data of cucumber leaves [J]. Chemometrics and Intelligent Laboratory Systems, 2010, 104(2): 265-270
- [4] 张筱蕾,刘飞,聂鹏程,等.高光谱成像技术的油菜叶片氮含量及分布快速检测[J].光谱学与光谱分析,2014,34(9):2213-2218  
ZHANG You-lei, LIU Fei, NIE Peng-cheng, et al. Rapid detection of nitrogen content and distribution in oilseed rape leaves based on hyperspectral imaging [J]. Spectroscopy and Spectral Analysis, 2014, 34(9): 2213-2218
- [5] 张银涛,葛宏立,孙贞,等.铁皮石斛多糖含量的高光谱反演研究[J].光谱学与光谱分析,2014,34(6):1645-1648  
ZHANG Yin-tao, GE Hong-li, SUN Zheng, et al. Estimating polysaccharide content with hyperspectral in *Dendrobium officinale* [J]. Spectroscopy and Spectral Analysis, 2014, 34(6): 1645-1648
- [6] 邹小波,陈正伟,石吉勇,等.基于近红外高光谱图像的黄瓜叶片色素含量快速检测[J].农业机械学报,2012,43(5): 152-156  
ZOU Xiao-bo, CHEN Zheng-wei, SHI Ji-yong, et al. Rapid detection of cucumber leaves pigments based on near infrared Hyper-spectral image technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(5):

- 152-156
- [7] 何晓文,满铭辉,纵伟.分光光度法测定枇杷叶总三萜物质含量[J].郑州轻工业学院学报(自然科学版),2007,22(1): 39-41
- HE Xiao-wen, MAN Ming-hui, ZONG Wei. Spectrophotometric methods to determine the total triterpenes in *Eriobotrya japonica* (Thumb.) Lindl [J]. Journal of Zhengzhou University of Light Industry(Natural Science), 2007, 22(1):39-41
- [8] Chen Q, Zhao J, Liu M, et al. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms [J]. J. Pharm. Biomed. Anal., 2008, 46(3): 568-573
- [9] 李子溪,江岸.枇杷叶中三萜酸的提取、分离和检测研究进展[J].楚雄师范学院学报,2013,28(3):39-43
- LI Zi-xi, JIANG An. Development of triterpenes in loquat (*Eriobotrya japonica*) leaf [J]. Journal of Chuxiong Normal University, 2013, 28(3): 39-43
- [10] 顾柏平,朱亮,柯惟忠,等.熊果酸的 Raman 光谱和红外光谱结构特征比较[J].光谱学与光谱分析,2002,22(6):970-972
- GU Bo-ping, ZHU Liang, KE Wei-zhong. Comparative study of the spectral characteristics of ursolic acid between laser raman spectra and IR spectrum[J]. Spectroscopy and Spectral Analysis, 2002, 22(6): 970-972