

猕猴桃膨大果的近红外漫反射光谱无损识别*

郭文川 刘大洋

(西北农林科技大学机械与工程学院, 陕西杨凌 712100)

摘要: 在 833 ~ 2 500 nm 光谱范围内采集了 120 个猕猴桃膨大果和 120 个正常果的近红外漫反射光谱, 采用变量标准化方法对光谱进行了预处理, 基于 Kennard - Stone 方法对样本进行了划分, 分别建立了基于全光谱 (FS)、主成分分析法 (PCA) 提取的 11 个主成分和连续投影算法 (SPA) 提取的 6 个特征波长的偏最小二乘 (PLS)、支持向量机 (SVM) 和误差反向传播 (BP) 神经网络识别模型。结果说明, 所建立的 9 个模型对校正集和测试集中猕猴桃膨大果和正常果的正确识别率均分别大于 96.7% 和 93.3%。PCA 提取的主成分数和 SPA 提取的特征波长数仅是 FS 中波长数的 0.53% 和 0.29%, 建立的模型更加简单, 且识别效率较高。PLS 和 SVM 模型的识别率普遍高于 BP 神经网络模型。9 种模型中 PCA - PLS 识别率最高, 其对校正集和测试集中膨大果和正常果的正确识别率均达到 100%。该研究结果表明, 近红外漫反射光谱技术可作为一种准确、高效的方法应用于猕猴桃膨大果的无损识别中。

关键词: 猕猴桃 膨大果 近红外光谱 识别

中图分类号: O657.33; S663.4 **文献标识码:** A **文章编号:** 1000-1298(2014)09-0230-06

引言

自 20 世纪 90 年代初, 外源性植物生长调节剂 (俗称膨大剂, 如氯吡脞、吡效隆等) 作为增产的一项重要技术, 在猕猴桃生产中被大面积推广应用^[1]。膨大剂虽能使猕猴桃产量提高 30% ~ 50%, 但却使果实品质下降、不耐贮藏、烂果率高、库损严重^[2], 膨大剂对人体健康的潜在危害目前也尚不明确。识别经膨大剂处理的果实 (膨大果) 就成为猕猴桃采后加工业亟需解决的关键问题。

近红外光谱区与有机分子中含氢基团 (O—H、N—H、C—H) 振动的合频和各级倍频的吸收区一致, 通过扫描样品的近红外光谱, 可以得到样品中有机分子含氢基团的特征信息, 而且利用近红外光谱技术分析样品具有方便、快速、高效、准确和成本较低, 以及不破坏样品、不消耗化学试剂、不污染环境等优点, 因此该技术已被广泛应用于农产品 (包括水果、蔬菜、谷物、饲料等) 成分的快速定量检测以及品种识别中。在应用近红外光谱技术预测猕猴桃的内部品质, 如可溶性固形物含量 (或糖度)、硬度、干物质含量、pH 值以及损伤等方面已经取得了一定的进展^[3-9], 但是尚未见无损识别猕猴桃膨大果的研究报道。

为了探索近红外光谱技术无损识别猕猴桃膨大

果的可行性, 本文以经过氯吡脞溶液蘸泡处理的“西选二号”猕猴桃膨大果和未经处理的正常果为对象, 采用化学计量学和人工神经网络技术建立识别膨大果和正常果的定性判别模型, 分析模型识别膨大果的准确率, 为猕猴桃膨大果的工业化识别提供一种无损、快捷、可靠的方法。

1 材料与方法

1.1 实验材料

实验用猕猴桃品种为“西选二号”。在承包的猕猴桃园内, 于猕猴桃谢花后 23 d, 按照使用说明书的要求, 使用质量浓度 10 mg/L 的氯吡脞溶液蘸泡猕猴桃膨大期的嫩果, 称该组样品为膨大果。另选一组样品未进行蘸泡处理, 称该组样品为正常果。成熟的膨大和正常猕猴桃果实采摘于 2013 年 9 月 26 日下午。当天运回实验室后, 除去表皮上的杂物。选取形状比较规则、无伤痕的膨大果和正常果各 120 个, 于采摘后 9 ~ 12 h 采集所有样品的近红外漫反射光谱。

1.2 实验设备与光谱采集方法

采用德国 Bruker 公司的 MPA 型傅里叶变换近红外光谱仪采集猕猴桃的近红外漫反射光谱。光谱扫描范围: 833 ~ 2 500 nm (12 000 ~ 4 000 cm^{-1}), 共采集 2 074 个波长; 扫描次数: 32 次; 光谱分辨率:

8 cm^{-1} , 探头视场角: 90° ; 采集条件: 室温 ($23 \sim 25^\circ\text{C}$), 以仪器内置背景为参比。采集近红外漫反射光谱时, 光纤探头与猕猴桃果皮直接接触, 在果实赤道部位选取均匀分布且无缺陷的 2 个点采集光谱, 以 2 个点光谱的平均值作为样品光谱的测试结果。

1.3 猕猴桃外观参数和品质的测量方法

首先用数显游标卡尺(分辨率 0.01 mm) 测量猕猴桃果实的轴向长度和赤道部位的直径, 然后在每个猕猴桃的赤道上选取均匀分布的 4 个点, 削去所选测点的果皮后, 用压头直径为 11 mm 的 GY-3 型果实硬度计测量果肉的硬度。然后在 4 点处取适量果肉用家用压蒜器榨汁, 果汁经 3 层医用纱布过滤后, 用 PR-101 α 型折射式数字糖度计 (ATAGO, Japan) 测量果汁的可溶性固形物含量, 作为衡量猕猴桃糖度参数。再在赤道周围取完好无损的果肉约 $12 \sim 15\text{ g}$ 平均放入 3 个直径 35 mm 的铝盒中, 置于 70°C 的 101-1AB 型电热恒温鼓风干燥箱 (天津市泰斯特仪器) 中干燥 24 h , 根据干燥前后果肉的质量计算果肉的初始湿基含水率。每个样品每个品质参数多次重复的平均值作为该样品的测量结果。

1.4 样本划分方法

选择 Kennard-Stone (K-S) 法划分样本。其基本思想是在所有的样本中选择欧氏距离最远的 2 个向量对进入训练集, 在接下来的迭代过程中, 将具有最大的最小距离的待选样本选入训练集, 以此类推, 达到所要求训练集的数目后停止迭代^[10]。该方法的优点是能保证训练集中的样本按照空间距离均匀分布。

1.5 近红外漫反射光谱的预处理

由于样本在各波长点上吸光度不同, 每个样本在各波长点上形成一条不同吸光度的光谱分布曲线, 曲线满足一定分布 (如正态分布) 规律, 采用变量标准化 (Standard normalized variate, SNV) 对每条光谱曲线进行处理, 以保证处理后的光谱数据均值为 0, 标准差为 1^[11]。

1.6 数据降维方法

为了简化模型, 本文采用主成分分析法^[12] (Principal component analysis, PCA) 和连续投影算法^[13] (Successive projections algorithm, SPA) 从经过 SNV 预处理的光谱数据中提取有效的特征变量或特征波长。

1.7 建模方法

1.7.1 偏最小二乘法

偏最小二乘法 (Partial least squares, PLS) 是通过最小误差平方和找到一组数据的最佳函数匹配。该算

法能够在自变量存在严重多重相关性的条件下进行回归建模, 允许样本点个数少于变量个数^[14-16]。

1.7.2 支持向量机

支持向量机 (Support vector machine, SVM) 是一种根据统计学习理论提出的研究方法。该方法是将原始数据映射到高维空间里, 构建一个最大间隔的超平面, 在超平面两侧有两个平行的超平面, 通过调整最优超平面的空间位置, 获得两个平行超平面间的最大距离。然后假设分类器误差和平行超平面间的距离成反比关系, 从而解决常规空间里数据间线性不可分的问题^[17-18]。

1.7.3 误差反向传播神经网络

误差反向传播 (Error back propagation, BP) 神经网络是目前应用最广泛的一种用于模式识别的人工神经网络。由于 BP 神经网络具有良好的自组织、自学习和处理复杂非线性问题的能力, 因而对于复杂的、非线性的体系, 可取得更好的效果^[19-20]。

1.8 数据分析软件

光谱数据由光谱分析软件 OPUS 6.5 (Bruker Optics, Germany) 采集, 数据的预处理应用 Unscrambler 9.8 (CAMO, Norway) 完成, 特征变量或特征波长的提取和建模基于 Matlab 7.11 (The Math Works, USA) 平台。

1.9 模型的评判标准

本研究以对猕猴桃膨大果的正确识别率和对正常果的正确识别率作为评判模型精度的标准。膨大果和正常果的正确识别率越高, 模型的精度越好。以模型中输入变量的数量作为衡量模型简单与否的标准。模型的输入变量越少, 模型越简单。

2 结果与讨论

2.1 猕猴桃外观尺寸及内部品质分析

无损完整猕猴桃的长度、直径、果汁的可溶性固形物质量分数、果肉的硬度及湿基含水率的统计值及方差分析结果见表 1。由表 1 可知, 膨大果的平均长度和直径分别是正常果的 1.22 和 1.24 倍。膨大猕猴桃可溶性固形物质量分数的变化范围小于正常果, 但是两者的平均值比较接近。膨大果湿基含水率的平均值高于正常果, 但平均硬度略低于正常果。方差分析结果表明, 在 5% 的显著水平上, 膨大果和正常果的可溶性固形物质量分数没有显著差异, 但湿基含水率和硬度均有显著性差异。膨大剂对梨品质的研究结果表明, 膨大剂对果实有催熟作用^[21]。这种催熟作用使得同期采摘的猕猴桃膨大果的硬度较正常果小。该现象也发现于同期采摘的“秦美”猕猴桃中^[2]。

表1 膨大果和正常果的理化特性统计值

Tab.1 Statistic values of physical and chemical properties of kiwifruits treated by EPGR and normal ones

类别	样本数量	长度/mm		直径/mm		可溶性固形物质量分数/%		硬度/(kg·cm ⁻²)		湿基含水率/%	
		范围	均值±标准差	范围	均值±标准差	范围	均值±标准差	范围	均值±标准差	范围	均值±标准差
膨大果	120	55.3~74.2	64.7±4.37 ^a	46.3~62.6	57.0±4.5 ^a	12.3~15.9	14.6±0.9 ^a	5.4~12.6	8.7±2.6 ^a	84.1~88.9	85.8±1.1 ^a
正常果	120	44.3~65.2	52.9±3.3 ^b	41.5~50.3	45.8±4.9 ^b	13.3~17.7	14.4±1.1 ^a	6.2~14.4	9.2±2.2 ^b	81.5~85.0	83.3±0.7 ^b

* 同列不同字母表示在5%的水平上差异显著。

2.2 光谱的预处理和样本划分

对所有样品的近红外光谱曲线进行 SNV 预处理。240 个样品的原始吸收光谱曲线及经 SNV 预处理后的光谱曲线分别如图 1 和图 2 所示,图中 R 表示反射率。从图中可以看出,在 979.8、1 199.2、1 450.8、1 786.8 和 1 927.6 nm 处均有明显的吸收峰。根据杰克·沃克曼等的研究结果^[22]可知,979.8、1 199.2 和 1 450.8 nm 为水的吸收峰,而 1 786.8 nm 和 1 927.6 nm 则是猕猴桃中糖类、果胶等物质(含有 C—H、O—H 等官能团)的特征吸收峰。从波形上看,膨大果与正常果的光谱曲线相同,且 2 种类型猕猴桃果实的吸收光谱重叠严重、相似度高,很难从中分辨出膨大果。为此,需从近红外光谱中提取有效的特征变量以建立识别猕猴桃膨大果的定性判别模型。采用 K-S 法对经过 SNV 预处理后的光谱按 3:1 划分样本,得到校正集样品 180 个(膨大果和正常果各 90 个),测试集样品 60 个(膨大果和正常果各 30 个)。

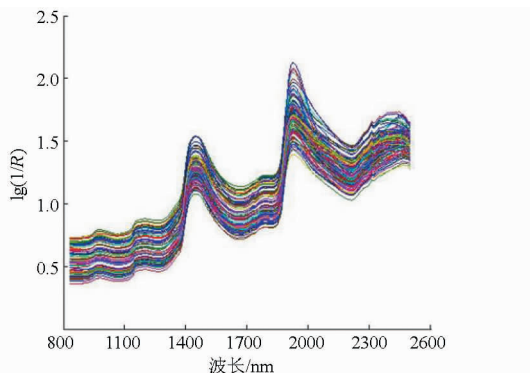


图1 240个猕猴桃样品的原始近红外光谱

Fig.1 Original near-infrared spectra of 240 kiwifruits

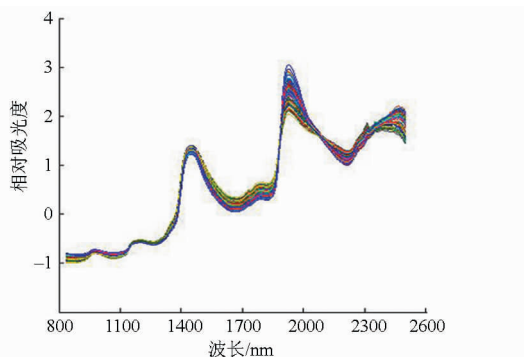


图2 SNV 预处理后的近红外光谱曲线

Fig.2 Near-infrared spectra after SNV preprocessing

2.3 光谱数据的降维

2.3.1 应用 PCA 提取主成分

本研究利用 Matlab 7.11 软件中的 Statistics Toolbox 对猕猴桃样品的近红外光谱数据进行主成分分析。前 12 个主成分的贡献率及累积贡献率见表 2。由表 2 可以看出,第 1 主成分(PC1)和第 2 主成分(PC2)的贡献率分别为 77.95% 和 13.43%,二者的累积贡献率为 91.38%。校正集中膨大果和正常果的 PC1 和 PC2 的得分图如图 3 所示。由图 3 可见,根据第 1、2 主成分得分可以区分出大部分的膨大果和正常果,但仍有一些膨大果与正常果不能区分。为充分保证模型具有较好的判别性能并有较小的计算量,本文选取累积贡献率达到 99.90% 的前 11 个主成分作为特征变量。

2.3.2 应用 SPA 选取特征波长

将 SPA 的波长数设定为 1~30,根据校正集的均方根误差选取最佳的特征波长数量。通常以均方根误差最小处的波长数作为最佳波长数量。本研究

表2 前12个主成分的贡献率和累积贡献率

Tab.2 Contribution rate and accumulative contribution rate of the first 12 principal components

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
贡献率/%	77.95	13.43	6.19	1.17	0.45	0.25	0.23	0.14	0.06	0.03	0.02	0.02
累积贡献率/%	77.95	91.38	97.56	98.73	99.17	99.42	99.66	99.79	99.85	99.88	99.90	99.92

中,随着特征波长数量的增加,均方根误差的变化过程如图 4 所示。由图 4 可见,当特征波长数量小于 6 时,随着波长数量的增加,均方根误差迅速减小,但当波长数量大于 6 时,均方根误差减小缓慢甚至保持不变。因此,本文以 6 个特征波长作为输入的

特征变量。该 6 个特征波长分别为 836、933、1 146、1 165、1 882 和 2 500 nm。

2.4 建模结果及建模方法的比较

2.4.1 建模参数的选择和设定

本文采用交叉验证的方法确定 PLS 的因子数,

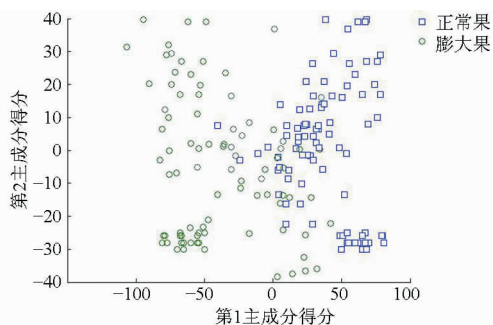


图3 校正集中膨大果和正常果的第1、2主成分得分图

Fig. 3 Score plot of PC1 and PC2 of kiwifruits in calibration set

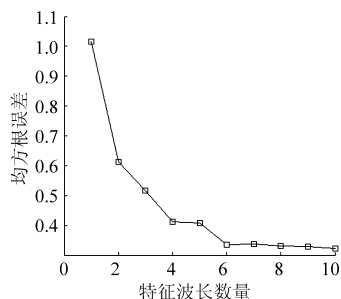


图4 均方根误差随 SPA 选择的特征波长数量的变化曲线

Fig. 4 Changed root-mean-square error with the number of selected characteristic wavelengths by SPA

根据留一交互原则选取 SVM 的惩罚因子(c)和松弛变量(g)。针对全光谱(FS)和 PCA 选取的 11 个主成分确定的 PLS 因子数和 SVM 的参数见表 3。BP 神经网络输入层传递函数为‘tansig’, 隐含层传递函数为‘logsig’, 训练函数选择‘trainrp’(回弹 BP 算法), 目标均方差设置为 0.01。由于隐含层神经元数目并没有明确的理论指导, 本文通过反复试验, 对比隐含层数量在 3~30 之间的网络预测性能确定的网络结构与参数如表 4 所示。

表3 PLS 和 SVM 建模参数的设定

Tab.3 Modeling parameters of PLS and SVM

特征变量 提取方法	PLS 因子数	SVM	
		c	g
FS	8	0.71	0.06
PCA	6	2.00	1.00
SPA	6	2.83	0.35

表4 BP 神经网络的网络结构与参数

Tab.4 Parameters and structure of BP-ANN

参数	FS - BP	PCA - BP
输入层	2 074	11
隐含层	18	5
输出层	1	1
学习率	0.3	0.2
训练次数	103	61

2.4.2 建模结果

基于 K-S 法划分的校正集样品的光谱数据, 将全光谱、经 PCA 提取的主成分以及经 SPA 提取的特征波长分别作为 PLS、SVM 以及 BP 神经网络的输入变量, 建立识别猕猴桃膨大果和正常果的模型, 并用测试集样本对模型进行验证。所建模型对校正集和预测集中膨大果和正常果的正确识别率见表 5。

表5 猕猴桃膨大果和正常果的正确识别率

Tab.5 Identification results on expanded kiwifruits and normal kiwifruits

建模方法	特征变量提取方法	校正集			测试集			%
		膨大果 (90 个)	正常果 (90 个)	平均值	膨大果 (30 个)	正常果 (30 个)	平均值	
PLS	FS	98.9	98.9	98.9	100	96.7	98.3	
	PCA	100	100	100	100	100	100	
	SPA	98.9	98.9	98.9	96.7	96.7	96.7	
SVM	FS	98.9	100	99.4	100	96.7	98.3	
	PCA	100	98.9	99.4	100	96.7	98.3	
	SPA	98.9	97.8	98.3	96.7	96.7	96.7	
BP	FS	97.8	96.7	97.2	96.7	93.3	95.0	
	PCA	97.8	97.8	97.8	96.7	96.7	96.7	
	SPA	96.7	96.7	96.7	96.7	93.3	95.0	

由表 5 可知, 在 PLS 模型中, 基于 PCA 所建的 PLS 模型对校正集和预测集中猕猴桃膨大果和正常果的正确识别率均达到 100%, 可以完全识别膨大果和正常果。基于 FS 和 SPA 所建的 PLS 模型对校正集中膨大果和正常果的正确识别率相等, 均为 98.9%, 但对于预测集, 基于 FS 所建模型的平均正确识别率为 98.3%, 略高于基于 SPA 所建模型的平均正确识别率 96.7%。在 SVM 模型中, 基于 FS 和 PCA 所建模型对校正集和预测集中样品的平均正确识别率均分别为 99.4% 和 98.3%, 略高于基于 SPA 的 98.3% 和 96.7%。在 BP 模型中, 基于 PCA 所建模型的平均正确识别率为 97.8%, 略高于 FS 的 97.2% 和 SPA 的 96.7%。对于预测集, 同样基于 PCA 的识别率 (96.7%) 高于基于 FS 和 SPA 的 95.0%。

2.4.3 建模结果比较

9 种模型对校正集和测试集中膨大果和正常果的正确识别率均分别高于 96.7% 和 93.3%, 其中 PCA-PLS 模型对校正集和测试集的识别率都达到了 100%。在各种建模方法中, 基于 PCA 所建模型的平均正确识别率最高, 其次为 FS 和 SPA。说明 PCA 提取的 11 个主成分能较全面地反映原始光谱中用于区分膨大果和正常果的有效信息。而 FS 不仅包含了有用的信息, 同时也包含了一定的噪声。SPA 虽然提取出了最少的特征波长, 但也漏掉了一些有用信息。此外, 基于 PCA 和 SPA 所建模型的输

入变量数分别只有 11 个和 6 个,仅相当于 FS 输入变量数的 0.53% 和 0.29%,很大程度上简化了判别模型。在 3 种建模方法中,当特征变量提取方法相同时,PLS 所建模型的正确识别率最高,其次为 SVM 和 BP。例如,当基于 PCA 提取的 11 个主成分建立各种模型时,PCA-PLS、PCA-SVM 和 PCA-BP 对校正集样品的平均正确识别率分别为 100%、99.4% 和 97.8%,对预测集中样品的平均正确识别率分别为 100%、98.3% 和 96.7%。

3 结论

(1) 基于全光谱、PCA 提取的 11 个主成分、SPA

优选的 6 个特征波长所建立的 PLS、SVM 和 BP 神经网络的模型对于校正集和测试集中猕猴桃膨大果的正确识别率均分别高于 96.7% 和 93.3%。

(2) 基于 PCA 所建模型的识别效果最好,其次为 FS 和 SPA。从模型简化程度上看,PCA 和 SPA 提取的特征变量数仅分别为 FS 中波长数的 0.53% 和 0.29%,模型识别效率远高于 FS。

(3) 在 3 种建模方法中,当特征变量提取方法相同时,PLS 所建模型的正确识别率最高,其次为 SVM 和 BP。其中,PCA-PLS 模型的识别率最高,其对校正集和测试集中膨大果和正常果的正确识别率均达到 100%。

参 考 文 献

- 1 蒋小平. 膨大剂在猕猴桃上应用的利弊[J]. 北方果树, 2010(3):43.
- 2 刘兴华, 郭井泉, 罗安伟, 等. 果实膨大剂对陕西省猕猴桃产业负效应的调查分析[J]. 保鲜与加工, 2004, 4(1):30-32.
Liu Xinghua, Guo Jingquan, Luo Anwei, et al. Investigations and conclusions on sprays of fruit-expander to kiwi fruit in Shaanxi Province[J]. Storage and Process, 2004, 4(1): 30-32 (in Chinese)
- 3 Meglone V A, Clark C J, Jordan R B. Comparing density and VNIR methods for predicting quality parameters of yellow-fleshed kiwifruit (*Actinidia chinensis*) [J]. Postharvest Biology and Technology, 2007, 46(1): 1-9.
- 4 蔡健荣, 汤明杰, 吕强, 等. 基于 siPLS 的猕猴桃糖度近红外光谱检测[J]. 食品科学, 2009, 30(4):250-253.
Cai Jianrong, Tang Mingjie, Lü Qiang, et al. Near infrared determination of sugar content in kiwifruits based on siPLS[J]. Food Science, 2009, 30(4): 250-253. (in Chinese)
- 5 Moghimi A, Aghkhani M H, Sazgarnia A, et al. Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit[J]. Biosystems Engineering, 2010, 106(3): 295-302.
- 6 刘卉, 郭文川, 岳绒. 猕猴桃硬度近红外漫反射光谱无损检测[J]. 农业机械学报, 2011, 42(3):145-149.
Liu Hui, Guo Wenchuan, Yue Rong. Non-destructive detection of kiwifruit firmness based on near infrared diffused spectroscopy [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3): 145-149. (in Chinese)
- 7 Chen Xiangwei, Han Wenting. Spectroscopic determination of soluble solids content of 'Qinmei' kiwifruit using partial least squares[J]. African Journal of Biotechnology, 2012, 11(10): 2528-2536.
- 8 郭文川, 王铭海, 谷静思, 等. 近红外光谱结合极限学习机识别贮藏期的损伤猕猴桃[J]. 光学精密工程, 2013, 21(10): 2720-2727.
Guo Wenchuan, Wang Minghai, Gu Jingsi, et al. Identification of bruised kiwifruits during storage by near infrared spectroscopy and extreme learning machine[J]. Optics and Precision Engineering, 2013, 21(10): 2720-2727. (in Chinese)
- 9 郭文川, 王铭海, 岳绒. 基于近红外漫反射光谱的损伤猕猴桃早期识别[J]. 农业机械学报, 2013, 44(2):142-146.
Guo Wenchuan, Wang Minghai, Yue Rong. Early recognition of bruised kiwifruit based on near infrared diffuse reflectance spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(2): 142-146. (in Chinese)
- 10 展晓日, 朱向荣, 史新元, 等. SPXY 样本划分法及蒙特卡罗交叉验证结合近红外光谱用于橘叶中橙皮苷的含量测定 [J]. 光谱学与光谱分析, 2010, 29(4):964-968.
Zhan Xiaori, Zhu Xiangrong, Shi Xinyuan, et al. Determination of hesperidin in tangerine leaf by near-infrared spectroscopy with SPXY algorithm for sample subset partitioning and monte carlo cross validation[J]. Spectroscopy and Spectral Analysis, 2010, 29(4): 964-968. (in Chinese)
- 11 石慧. 基于高光谱成像技术的对虾品质信息快速检测方法研究[D]. 杭州: 浙江大学, 2013.
- 12 Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1): 37-52.
- 13 Zhang J, Rivard B, Rogge D. The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data[J]. Sensors, 2008, 8(2): 1321-1342.
- 14 蒋红卫, 夏结来. 偏最小二乘回归及其应用[J]. 第四军医大学学报, 2003, 24(3):280-283.
Jiang Hongwei, Xia Jielai. Partial least square and its application[J]. Journal of the Fourth Military Medical University, 2003, 24(3): 280-283. (in Chinese)
- 15 赵春晖, 乔蕾. 基于改进的最小二乘支持向量机的高光谱遥感图像分类[J]. 应用科技, 2008, 35(1):44-47, 52.
Zhao Chunhui, Qiao Lei. Classification of hyperspectral remote sensing image using improved LS-SVM[J]. Applied Science and Technology, 2008, 35(1): 44-47, 52. (in Chinese)

- 16 薛佳辰, 冯钧, 雷震, 等. 基于偏最小二乘回归的性别识别[J]. 计算机工程与设计, 2013, 34(9): 3226 – 3229, 3254. Xue Jiachen, Feng Jun, Lei Ting, et al. Gender recognition using partial least squares[J]. Computer Engineering and Design, 2013, 34(9): 3226 – 3229, 3254. (in Chinese)
- 17 王晓丹, 王积勤. 支持向量机训练和实现算法综述[J]. 计算机工程与应用, 2004, 40(13): 75 – 78, 175. Wang Xiaodan, Wang Jiqin. A survey on support vector machines training and testing algorithms[J]. Computer Engineering and Applications, 2004, 40(13): 75 – 78, 175. (in Chinese)
- 18 赵杰文, 呼怀平, 邹小波. 支持向量机在苹果分类的近红外光谱模型中的应用[J]. 农业工程学报, 2007, 23(4): 149 – 152. Zhao Jiewen, Hu Huaiping, Zou Xiaobo. Application of support vector machine to apple classification with near-infrared spectroscopy[J]. Transactions of the CSAE, 2007, 23(4): 149 – 152. (in Chinese)
- 19 吴春胤, 张文昭, 欧阳庆, 等. 基于BP神经网络模型的荔枝树叶面积测定方法[J]. 农业工程学报, 2007, 23(7): 166 – 169. Wu Chunyin, Zhang Wenzhao, Ouyang Qing, et al. BP neural network model for the measurement of the leaf area of litchi[J]. Transactions of the CSAE, 2007, 23(7): 166 – 169. (in Chinese)
- 20 He Yong, Li Xiaoli, Deng Xunfei. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model[J]. Journal of Food Engineering, 2007, 79(4): 1238 – 1242.
- 21 关军锋, 马文会, 及华, 等. 套袋和膨大剂对黄冠梨钙、果胶含量和果实品质的影响[J]. 科技导报, 2008, 26(21): 73 – 76. Guan Junfeng, Ma Wenhui, Ji Hua, et al. Effects of bagging and fruit expander on Ca, pectin content and the quality of Huangguan pear fruits[J]. Science & Technology Review, 2008, 26(21): 73 – 76. (in Chinese)
- 22 杰尔·沃克曼(美), 洛伊斯·文依(美). 近红外光谱解析实用指南[M]. 诸小立, 许育鹏, 田高友, 译. 北京: 化学工业出版社, 2009: 226 – 232.

Identification of Expanded Kiwifruits by Near-infrared Diffused Spectroscopy

Guo Wenchuan Liu Dayang

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

Abstract: In order to develop a nondestructive method for identifying expanded kiwifruits, near-infrared diffused spectra of 120 expanded kiwifruits and 120 normal kiwifruits were obtained between 833 and 2 500 nm using a Fourier transformation near-infrared diffused spectrograph. Standard normal variate transformation was used to preprocess original spectra. The samples were divided into calibration set and prediction set based on Kennard – Stone method. Eleven principal components and 6 characteristic wavelengths were selected by principal component analysis (PCA) and successive projections algorithm (SPA). Partial least squares (PLS), support vector machine (SVM), and error back propagation (BP) neural network identification model were established based on full spectrum (FS), PCA, and SPA, respectively. The results showed that correct identification rates of all models were higher than 96.7% and 93.3% for calibration set and prediction set, respectively. The established models based on PCA and SPA were much simpler than those based on FS, since the variable numbers of them were only about 0.53% and 0.29% of that of FS, respectively. The identification performance of PLS and SVM were better than that of BP. The best model was PCA – PLS, whose accuracy rate reached 100% for calibration set and prediction set. The results clearly indicate that near-infrared diffused spectra technique has the potential as an efficient, accuracy and non-invasive method for distinguishing expanded kiwifruits from normal kiwifruits.

Key words: Kiwifruit Expanded fruit Near-infrared spectra Identification