

孙晓明, 陈小龙, 余向阳, 等. 基于近红外光谱分析技术的水蜜桃产地溯源[J]. 江苏农业学报, 2020, 36(2): 507-512.
doi: 10.3969/j.issn.1000-4440.2020.02.035

基于近红外光谱分析技术的水蜜桃产地溯源

孙晓明, 陈小龙, 余向阳, 卞立平, 孙爱东

(江苏省农业科学院农产品质量安全与营养研究所/省部共建国家重点实验室培育基地——江苏省食品质量安全重点实验室, 江苏 南京 210014)

摘要: 利用近红外光谱分析技术对来自3个省份的水蜜桃进行研究, 比较主成分分析-线性判别分析(PCA-LDA)、判别偏最小二乘法(DPLS)、支持向量机(SVM)等方法对光谱数据识别的有效性差异。结果表明, SVM的准确率和召回率均高达94.47%, 明显优于PCA-LDA和DPLS, 更适用于水蜜桃产地溯源。

关键词: 水蜜桃; 产地溯源; 近红外光谱; 主成分分析-线性判别分析; 判别偏最小二乘; 支持向量机

中图分类号: TS207.7 **文献标识码:** A **文章编号:** 1000-4440(2020)02-0507-06

Traceability of honey peach origin using near infrared spectroscopy analysis techniques

SUN Xiao-ming, CHEN Xiao-long, YU Xiang-yang, BIAN Li-ping, SUN Ai-dong

(*Institute of Food Safety and Nutrition, Jiangsu Academy of Agricultural Sciences/Jiangsu Key Laboratory for Food Quality and Safety-State Key Laboratory Cultivation Base of Ministry of Science and Technology, Nanjing 210014, China*)

Abstract: In this study, honey peaches from three provinces were analyzed by near infrared spectroscopy analysis technique, and the effectiveness of principal component analysis-linear discriminant analysis (PCA-LDA), discriminant partial least squares (DPLS) and support vector machine (SVM) for spectral data recognition was compared. The results showed that the precision and recall rate of SVM were 94.47%. The SVM method was obviously better than PCA-LDA and DPLS, and it was more suitable for traceability of honey peach origin.

Key words: honey peach; geographical origin traceability; near infrared spectroscopy; principal component analysis-linear discriminant analysis; discriminant partial least squares; support vector machine

随着生活水平的提高, 人们的食品安全意识大大增强, 地域特色产品越来越受到市场的认可。这些产品通常会带来良好的经济效益。针对地域特色产品的以假乱真、以次充好现象大量存在, 既损害了消费者权利, 又打击了生产者积极性^[1-2]。因此, 开展地域特色产品产地溯源技术研究, 对于推动溯源

体系的建立与完善, 促进农产品安全有效监管具有重要意义。产地溯源技术主要分两大类^[3-6]: 一是电子信息编码技术, 即将农产品产地信息及生产、加工等环节的相关信息关联到电子信息载体, 消费者通过二维码等识别农产品的产地; 二是化学生物分析技术, 即通过检测农产品的成分来获知产地信息, 包括近红外光谱分析技术、同位素分析技术等。其中, 近红外光谱分析技术因具有操作简单、检测速度快、重现性好、绿色环保、成本低等特点^[7], 已经成为一种重要的农产品产地溯源识别技术手段。

近红外光谱主要反映有机物分子中 C-H、N-H、O-H 等含氢基团振动的倍频和合频吸收信息。不同产地的农产品, 因为品种、产地环境、加工方法等不同, 其有机物组成和含量存在较大差异, 形成近红外

收稿日期: 2019-07-09

基金项目: 国家重点研发计划项目(2017YFC1601000); 江苏省农业科技自主创新基金项目[CX(18)3054]; 国家现代农业产业技术体系桃体系项目(CARS-30-5-03)

作者简介: 孙晓明(1986-), 江苏海安人, 硕士, 助理研究员, 主要从事农业信息化、农产品质量安全研究。(E-mail) sunxiaoming2226@163.com。陈小龙为共同第一作者。

通讯作者: 孙爱东, (E-mail) idong.sun@qq.com

光谱中的差异。通过分析近红外光谱,可以对农产品进行产地溯源、真伪鉴别和品质检测^[8-16]。目前近红外光谱分析技术已得到广泛应用,李勇等采用主成分分析和线性判别分析对大米进行产地溯源分析^[17]。姜亦南等采用傅里叶变换近红外光谱法结合 SIMCA 模式识别算法对三七进行产地鉴别^[18]。对水蜜桃的相关研究较少。李剑等采用主成分分析结合最小二乘法建立了水蜜桃采摘期的鉴别模型^[19]。王铭海等建立了3个不同品种桃可溶性固形物含量预测模型^[20]。

鉴于目前应用近红外光谱分析技术识别水蜜桃产地的研究较少,本研究采用此技术,通过光谱预处理等手段结合模式识别方法进行水蜜桃产地识别研究,并在江苏、山东、福建3个省份的水蜜桃上进行试验,为水蜜桃产地溯源提供依据。

1 材料与方法

1.1 数据采集

1.1.1 水蜜桃样品采集 分别采集产自江苏省、山东省、福建省的“湖景”水蜜桃样品各50个。样品放置于空气温度25℃、湿度75%的环境中。

1.1.2 近红外光谱数据采集 利用 Thermo Antaris II 傅里叶变换近红外分析仪,采用漫反射方式采集水蜜桃样品的近红外光谱。以仪器内部空气为背景,光谱扫描波数为4 000~10 000 cm^{-1} ,扫描间隔为3.856 cm^{-1} ,每条光谱有1 557个点,扫描次数32次,4倍增益。在果实赤道上选择3个相距120°的位置测定漫反射光谱,3个位置的平均值作为该果实的光谱测定值。

1.2 数据分析

1.2.1 主成分分析-线性判别分析 主成分分析(Principal component analysis, PCA)是最常用的无监督降维方法,该方法把原始高维变量转化为几个主成分,使得每一维上的样品方差都很大^[21]。这些主成分通常表示为原始变量的线性组合,不仅能够反映原始变量的大部分信息,而且各主成分之间相互独立。线性判别分析(Linear discriminant analysis, LDA)是一种监督学习的判别分析方法,首先通过训练集建立线性判别函数,该判别函数使得类内方差最小、类间方差最大,即具有最佳的可分离性,然后利用判别函数对测试集样品进行识别^[22]。对 PCA 分析过的数据使用 LDA 进一步降维,可以将最

佳描述特征转化为最佳判别特征进行分类识别^[23]。

1.2.2 判别偏最小二乘法 判别偏最小二乘法(Discriminant partial least squares, DPLS)是一种将定量偏最小二乘法(Partial least squares, PLS)用于判别分析的方法,通过建立自变量与因变量之间的多元统计回归模型进行分析^[24]。和 PCA 方法类似,该方法同样采取了成分提取的方式,不同的是,提取成分时不仅要求成分能够较好地概括自变量的信息,同时要求对因变量的解释能力达到最大。成分通过迭代法逐步提取,成分个数通过交叉有效性检验来确定。根据 PLS 模型的预测值确定样品的类别。

1.2.3 支持向量机 支持向量机(Support vector machine, SVM)是一个经典的分类模型,主要思想是在特征空间上建立最优超平面使得类之间的间隔最大,在解决小样本、非线性及高维数据等问题中表现出许多特有的优势^[25-26]。SVM 首先通过非线性映射函数将样品从原始特征空间映射到高维特征空间,然后构造最优决策函数,综合考虑结构风险最小化和经验风险最小化,确定模型参数,同时引入间隔的概念。SVM 的优化目标为:

$$\min_{w, b, \xi_i} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^m \xi_i \quad (1)$$

$$s.t. y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, 2, \dots, m$$

式中, m 为训练样本数, x_i 为训练样本的支持向量, y_i 取值为+1或-1,表示对应样本的类别, w 为超平面的法向量, b 为位移项, c 为惩罚参数, ξ_i 为松弛变量。

求解时利用核函数代替内积运算,避免“维数灾难”。常用的核函数有线性核、多项式核、RBF核和 Sigmoid 核。

1.2.4 评价指标 评价指标主要采用准确率和召回率^[27],由分类结果的混淆矩阵 C 计算得出。混淆矩阵的行表示真实类别,列表示预测类别,元素 c_{ml} 表示真实类别为 m 、预测类别为 l 的样品数目。类别数目为 k 。准确率表示被预测为某类的样品中有多少真正为该类的,每个类的准确率计算公式如下:

$$Precision_i = \frac{c_{ii}}{\sum_{j=1}^k c_{ji}} \quad (2)$$

召回率表示每个类中样品被正确分类的比例,每个类的召回率计算公式如下:

$$Recall_i = \frac{c_{ii}}{\sum_{j=1}^k c_{ij}} \quad (3)$$

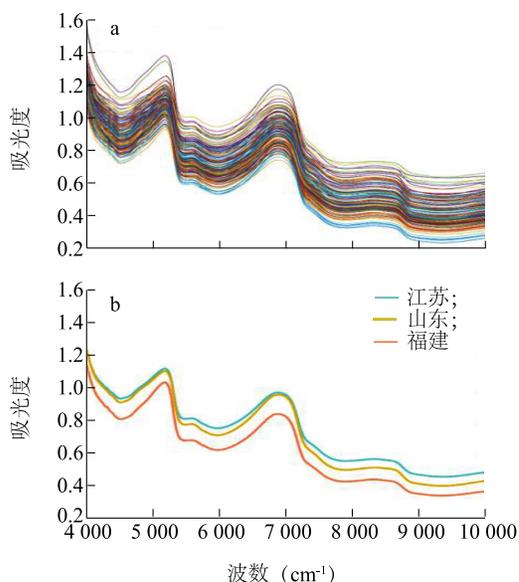
方法的准确率和召回率取类准确率和召回率的均值。数值越高,代表方法越好。

采用十折交叉验证法将样品划分为训练集和测试集,取每类的 10% 作为测试集,随机使用不同的划分 10 次,10 次十折交叉试验的均值作为最终的试验结果。所有数据分析均在 MATLAB R2016a 中编程实现。

2 结果分析

2.1 水蜜桃样品的近红外光谱

图 1a 是所有水蜜桃样品的原始光谱,图 1b 是不同省份水蜜桃样品的原始光谱均值。图 1b 中 3 个省份的均值光谱明显区分开,且江苏省和山东省的距离较近。而图 1a 中近红外光谱之间重叠严重,不能直接区分产地,存在光散射、基线漂移等问题,对试验结果影响较大,需要采取适当的光谱预处理方法去除干扰^[28],以便更好地分析。常见的光谱预处理方法有平滑(Smoothing)、一阶导数(1st Der)、二阶导数(2nd Der)、标准正态变换(SNV)、多元散射校正(MSC)、小波变换(WT)等。试验中采用 SNV 方法对光谱数据进行预处理,处理后的光谱见图 2。



a: 所有水蜜桃样品的原始光谱; b: 不同省份水蜜桃样品的原始光谱均值。

图 1 水蜜桃样品的原始近红外光谱

Fig.1 Original near infrared spectra of the honey peach samples

2.2 PCA-LDA 方法分析

首先对全波段光谱数据进行标准化处理,然后使用奇异值分解进行主成分分析。累计贡献率达到

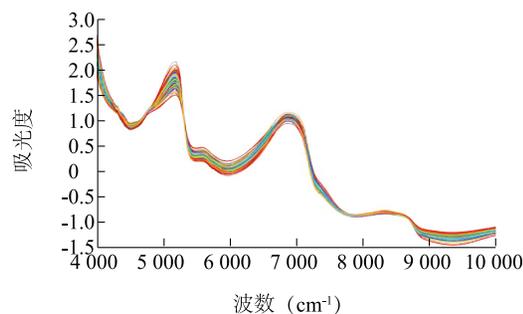


图 2 预处理后的样品近红外光谱

Fig.2 Near-infrared spectra of the samples after pretreatment

99%时即可确定主成分数目。前 10 个主成分的贡献率见表 1,主成分数目为 7 时,累计贡献率超过 99%。图 3 为前 2 个主成分的得分图,主成分 1 和主成分 2 分别解释了总方差的 65.81% 和 17.78%。由图 3 可见,利用前 2 个主成分时,山东省和福建省的水蜜桃样品在距离上基本分开,但均和江苏省的有重叠,且江苏省和山东省的重叠区域较大,仅使用 PCA 方法不能直接区分 3 个省份的样品。

表 1 前 10 个主成分的贡献率和累计贡献率

Table 1 Contribution rate and cumulative contribution rate of the top 10 principal components

主成分	贡献率 (%)	累计贡献率 (%)
1	65.81	65.81
2	17.78	83.59
3	9.45	93.04
4	2.86	95.90
5	1.49	97.39
6	1.23	98.62
7	0.42	99.04
8	0.27	99.31
9	0.19	99.50
10	0.16	99.66

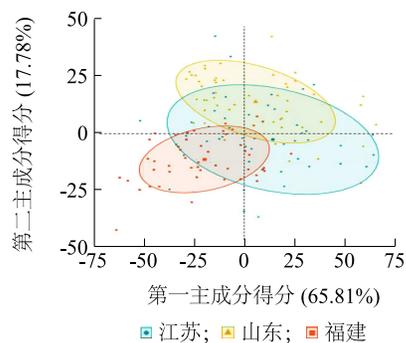


图 3 第一主成分和第二主成分得分

Fig.3 Scores of the first and second principal components

将前 7 个主成分的得分数据作为 LDA 的输入, 构建预测模型, 得到 2 个投影函数, 分别为:

$$Function1 = 0.033 \times PC1 + 0.084 \times PC2 + 0.052 \times PC3 - 0.123 \times PC4 - 0.053 \times PC5 + 0.242 \times PC6 - 0.117 \times PC7 \quad (4)$$

$$Function2 = -0.012 \times PC1 + 0.037 \times PC2 + 0.017 \times PC3 - 0.102 \times PC4 + 0.025 \times PC5 - 0.240 \times PC6 + 0.008 \times PC7 \quad (5)$$

式中, PC_i 表示样品在第 i 个主成分上的得分。预测样品类别时, 首先计算主成分得分, 然后投影到函数(4)、(5)上, 根据投影值计算属于各个类别的概率, 最大的概率对应的类别即为预测类别。

核心代码如下:

```
%1.数据标准化
[ traindata, testdata ] = DataZscore( traindata, testdata );
%2.PCA 降维, coeff 为系数矩阵, score 为得分矩阵, latent 为贡献率矩阵
[ coeff, score, latent ] = Princomp( traindata );
%3.选择主成分数目
ncomp = ChooseComp( latent );
%4.计算测试数据的得分
testdata_score = CalScore( testdata, coeff, ncomp );
%5.判别分析
testdata_predictlabel = PcaldaPredict( score, ncomp, traindata_label, testdata_score );
```

2.3 DPLS 方法分析

首先将江苏省、山东省、福建省的样品类别分别设为 -1、0、1, 然后对全波段光谱数据和类别数据标准化。标准化后的数据作为 DPLS 的输入。试验中采用的交叉有效性公式为:

$$Q_h^2 = 1 - PRESS(h) / SS(h-1) \quad (6)$$

式中, $PRESS(h)$ 表示留一交叉验证时, 提取 h 个成分, 因变量的预测误差平方和; $SS(h-1)$ 表示所有样品数据参与拟合时, 提取 $h-1$ 个成分, 因变量的预测误差平方和。取 $Q_h^2 < 1 - 0.95^2$ 时的成分个数作为最终选取的成分个数。前 8 个成分的交叉有效性 (Q_h^2) 分别为 1.000 0、0.138 0、0.404 5、0.167 9、0.001 5、0.000 6、-0.050 6、-0.009 5, $h=5$ 时满足要求。DPLS 中的成分个数小于 PCA-LDA 中的主成分个数, 这是因为 DPLS 是有监督的学习, 通常模型更简单。

使用前 5 个成分构建预测模型, 得到标准化类别预测值 (\bar{Y}) 关于成分 t 的回归函数为:

$$\bar{Y} = 0.015 8 \times t_1 + 0.028 7 \times t_2 - 0.075 0 \times t_3 + 0.026 0 \times t_4 + 0.051 8 \times t_5 \quad (7)$$

还原为原始变量, 类别预测值 y 关于光谱数据 x 的回归函数为:

$$Y = 0.005 2 \times x_1 + 0.004 3 \times x_2 + \dots + 0.077 5 \times x_{1557} + 14.097 2 \quad (8)$$

若 Y 介于 -1.5 ~ -0.5, 样品产地为江苏省; 若 y 介于 -0.5 ~ 0.5, 样品产地为山东省, 若 Y 介于 0.5 ~ 1.5, 样品产地为福建省。

核心代码如下:

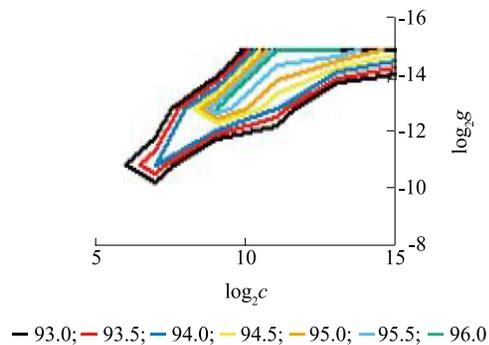
```
%1.数据标准化
traindata = Zscore( traindata );
%2.选择最佳成分个数
ndim = ChooseDim( traindata, traindata_label );
%3.使用最佳成分个数进行偏最小二乘回归
model = Plsr( traindata, traindata_label, ndim );
%4.判断测试数据分类
testdata_predictlabel = DplsPredict( model, testdata );
```

2.4 SVM 方法分析

归一化有助于提升 SVM 方法的精度和参数求解速度。首先将全波段光谱数据映射到 $[-1, 1]$ 区间内, 进行归一化处理, 其次将归一化后的数据作为 SVM 的输入。采用 RBF 核函数作为 SVM 的核函数, 公式如下:

$$K(x_i, x_j) = \exp(-g * \|x_i - x_j\|^2), g > 0 \quad (9)$$

式中, x_i, x_j 表示样品的光谱数据, g 表示核函数参数。通过交叉验证和网格搜索法确定惩罚参数 (c) 和核函数参数 (g) 的最优组合, 参数寻优过程见图 4。最优参数为 $c = 512, g = 1.220 7 \times 10^{-4}$, 交叉验证精度 96.00%。



c 表示惩罚参数, g 表示 RBF 核函数参数。参数最优组合: $\log_2 c = 9, \log_2 g = -13$ 。

图 4 SVM 中交叉验证和网格搜索法参数寻优示意图

Fig.4 Schematic diagram of cross-validation and grid search parameter optimization in support vector machine (SVM)

SVM 是一个二分类器,试验中采用一对一法,在任意两类样品之间构造一个二分类器, k 类样品共构造 $k \times (k-1)/2$ 个二分类器。使用参数最优组

合进行训练,得到 46 个支持向量,其中江苏省样品的支持向量 21 个,山东省样品的支持向量 17 个,福建省样品的支持向量 8 个(表 2)。

表 2 SVM 的支持向量、支持向量系数和位移项

Table 2 Support vectors, support vector coefficients and displacement terms of SVM

		$\alpha_i y_i$		x_i	b_{ln}
江苏	江苏 vs 山东	江苏 vs 福建			
1	42.07	0	0.39, 0.43, ..., 0.44		
⋮	⋮	⋮	⋮		2.83
21	87.50	0	-0.35, -0.30, ..., -0.50		
山东	山东 vs 江苏	山东 vs 福建			
1	-82.74	402.33	0.33, 0.35, ..., 0.40		
⋮	⋮	⋮	⋮		-2.26
17	-427.19	0	0.26, 0.20, ..., 0.39		
福建	福建 vs 江苏	福建 vs 山东			
1	-512.00	0	0.10, 0.05, ..., 0.00		
⋮	⋮	⋮	⋮		-2.45
8	-221.66	-160.96	0.36, 0.36, ..., 0.78		

i 表示支持向量数量, $\alpha_i y_i$ 表示第 i 个支持向量的系数, x_i 表示第 i 个支持向量, b_{ln} 表示判别样品属于类别 l 还是类别 n 的位移项。

训练结束后,每个二分类器都会生成一个二分类判别函数 $f_{ln}(x)$, 当 $f_{ln}(x) > 0$ 时, 样品 x 被预测为 l 类, 否则被归为 n 类。 $f_{ln}(x)$ 的计算公式为:

$$f_{ln}(x) = \sum_{i=1}^{m_l} \alpha_{lni} y_{lni} K(x, x_i) + \sum_{i=1}^{m_n} \alpha_{nli} y_{nli} K(x, x_i) + b_{ln}, \quad l < n \leq k \quad (10)$$

其中, m_l, m_n 分别为类别 l, n 的支持向量数目, $\alpha_{lni} y_{lni}, \alpha_{nli} y_{nli}$ 分别为类别 l 对类别 n 、类别 n 对类别 l 的支持向量系数, x_i 为支持向量, b_{ln} 为位移项。将 x 代入 $k \times (k-1)/2$ 个二分类判别函数进行统计, 得票最多的类别即为该样品的最终预测类别。

核心代码如下:

```
%1.训练数据和测试数据归一化
[ traindata, testdata ] = DataNormalize ( traindata,
testdata );
%2.网格法寻找最优参数
[ bestacc, bestc, bestg ] = SvmCgForClass( traindata
_label, traindata );
%3.建立 SVM 模型
model = SvmTrain( traindata_label, traindata, bestc,
bestg );
%4.判断测试数据分类
testdata_predictlabel = SvmPredict( testdata_label,
testdata, model );
```

2.5 不同分析方法的分类结果对比

不同方法的分类结果见表 3。可以发现, PCA-

LDA 方法对江苏省、山东省、福建省的水蜜桃预测正确率分别为 81.40%、87.40% 和 100.00%, DPLS 方法的预测正确率分别为 66.40%、83.60% 和 78.60%, SVM 方法的预测正确率分别为 89.20%、95.00% 和 99.20%, 3 类基本区分开, SVM 方法的分类识别正确率最高。江苏省和山东省的水蜜桃之间最容易发生误判, 可能是因为江苏省和山东省地理位置比较接近, 生长环境、气候等方面更加相似。总体上看, PCA-LDA 方法的准确率为 89.51%, 召回率为 89.60%。DPLS 方法的识别结果最不理想, 准确率仅为 79.50%, 召回率仅为 76.20%, 可能是由于没有选择合适的预处理方法导致。SVM 方法通过间隔的概念, 得到对数据分布的结构化描述, 基于结构风险最小化原则, 泛化能力强, 因此分类效果最好, 准确率达到 94.47%, 召回率达到 94.47%。综合来看, 采用近红外光谱分析技术对水蜜桃进行产地溯源是可行的, 结果令人满意。

3 结论

本研究采用近红外光谱分析技术实现水蜜桃的产地识别, 在对近红外光谱数据进行 SNV 预处理的基础上, 利用 PCA-LDA、DPLS、SVM 方法对江苏、山东、福建 3 个省份的 150 个水蜜桃样品进行产地溯源研究。试验结果表明, 3 种方法均可识别水蜜桃的产地, 其中 SVM 方法的识别结果最优, 对江苏省、

表3 PCA-LDA、DPLS和SVM方法的准确率与召回率
Table 3 Precision and recall rate of principal component analysis-linear discriminant analysis (PCA-LDA), discriminant partial least squares (DPLS) and SVM methods

模型	真实类别	预测类别(%)			准确率(%)	召回率(%)
		江苏	山东	福建		
PCA-LDA	江苏	81.40	15.60	3.00	89.51	89.60
	山东	12.60	87.40	0.00		
	福建	0.00	0.00	100.00		
DPLS	江苏	66.40	31.80	1.80	79.50	76.20
	山东	11.40	83.60	5.00		
	福建	0.00	21.40	78.60		
SVM	江苏	89.20	8.60	2.20	94.47	94.47
	山东	5.00	95.00	0.00		
	福建	0.80	0.00	99.20		

山东省、福建省的水蜜桃识别正确率分别为 89.20%、95.00%、99.20%，准确率和召回率均达到 94.47%。该方法可为农产品产地溯源研究提供参考。下一步研究将考虑提取与地域相关的特征波长，去除干扰信息，加快训练速率，进一步提高该方法的性能。

参考文献:

- [1] 管 骁,古方青,杨永健. 近红外光谱技术在食品产地溯源中的应用进展[J]. 生物加工过程, 2014, 12(2):77-82.
- [2] 钱丽丽,于 果,迟晓星,等. 农产品产地溯源技术研究进展[J]. 食品工业, 2018, 39(1):246-249.
- [3] COZZOLINO D. An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals[J]. Food Research International, 2014, 60(6):262-265.
- [4] LOHUMI S, LEE S, LEE H, et al. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration[J]. Trends in Food Science & Technology, 2015, 46(1):85-98.
- [5] 陈 璐,谷晓红,张丙春,等. 食品产地溯源技术研究进展[J]. 安徽农业科学, 2015, 43(36):109-111.
- [6] 曾楚锋,张丽芬,徐娟娣,等. 农产品产地溯源技术研究进展[J]. 食品工业科技, 2013, 34(6):367-371.
- [7] 张 勇,王 督,李 雪,等. 基于近红外光谱技术的农产品产地溯源研究进展[J]. 食品安全质量检测学报, 2018, 9(23):6161-6166.
- [8] HU X, LIU S, LI X, et al. Geographical origin traceability of cabernet sauvignon wines based on infrared fingerprint technology combined with chemometrics[J]. Scientific Reports, 2019, 9:8256-8263.
- [9] EISENSTECKEN D, STÜRZ B, ROBATSCHER P, et al. The po-

- tential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations[J]. Postharvest Biology and Technology, 2019, 147:123-131.
- [10] MANFREDI M, ROBOTTI E, QUASSO F, et al. Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics[J]. Spectrochim Acta A Mol Biomol Spectrosc, 2018, 189:427-435.
 - [11] MOSCETTI R, HAFF R P, STELLA E, et al. Feasibility of NIR spectroscopy to detect olive fruit infested by Bactrocera oleae[J]. Postharvest Biology and Technology, 2015, 99:58-62.
 - [12] 仇逊超. 红松仁脂肪的近红外光谱定量检测[J]. 江苏农业学报, 2018, 34(3):217-223.
 - [13] 廖秋红,何绍兰,谢让金,等. 基于近红外光谱的纽荷尔脐橙产地识别研究[J]. 中国农业科学, 2015, 48(20):4111-4119.
 - [14] 罗 微,杜焱喆,章海亮. PCA和SPA的近红外光谱识别白菜种子品种研究[J]. 光谱学与光谱分析, 2016, 36(11):3536-3541.
 - [15] 顾玉琦,刘瑞婷,寿国忠,等. 应用近红外光谱技术快速鉴别铁皮石斛的产地[J]. 江苏农业科学, 2016, 44(5):365-368.
 - [16] 向伶俐,李梦华,李景明,等. 近、中红外光谱法融合判定葡萄酒产地[J]. 光谱学与光谱分析, 2014, 34(10):2662-2666.
 - [17] 李 勇,严煌倩,龙 玲,等. 化学计量学模式识别方法结合近红外光谱用于大米产地溯源分析[J]. 江苏农业科学, 2017, 45(21):193-195.
 - [18] 姜亦南,蔺明焯,何 帅,等. 基于红外光谱法结合 SIMCA 模式识别不同产地三七[J]. 中医药学报, 2019, 47(1):54-57.
 - [19] 李 剑,李臻峰,宋飞虎,等. 基于近红外光谱的水蜜桃采摘期的鉴别方法[J]. 传感器与微系统, 2017, 36(10):48-50.
 - [20] 王铭海,郭文川,商 亮,等. 基于近红外漫反射光谱的多品种桃可溶性固形物的无损检测[J]. 西北农林科技大学学报(自然科学版), 2014, 42(2):142-148.
 - [21] JOLLIFFE I T. Principal component analysis[M]. New York: Springer, 2002.
 - [22] 邵圣枝,陈元林,张永志,等. 稻米中同位素与多元素特征及其产地溯源 PCA-LDA 判别[J]. 核农学报, 2015, 29(1):119-127.
 - [23] 陈 庆,黄 蕾,李雪梅. 基于主成分判别分析的高光谱遥感影像分类方法[J]. 地理空间信息, 2016, 14(1):76-78.
 - [24] 贾文坤. 基于多源信息融合的龙井茶产地鉴别研究[D]. 长春:吉林大学, 2014.
 - [25] NELLO C, JOHN S. 支持向量机导论[M]. 北京:电子工业出版社, 2004:82-98.
 - [26] 褚 璇,王 伟,赵 昕,等. 近红外光谱和特征光谱的山茶油掺假鉴别方法研究[J]. 光谱学与光谱分析, 2017, 37(1):75-79.
 - [27] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016:30.
 - [28] 褚小立. 近红外光谱分析技术实用手册[M]. 北京:机械工业出版社, 2016:115-117.

(责任编辑:张震林)