

## 多光谱成像的玛咖掺伪定性鉴别和定量分析

张宏蕊<sup>1,2</sup>, 刘长虹<sup>1</sup>, 张九凯<sup>2</sup>, 韩建勋<sup>2</sup>, 陈颖<sup>2</sup>, 郑磊<sup>1\*</sup>

1. 合肥工业大学食品与生物工程学院, 安徽 合肥 230009

2. 中国检验检疫科学研究院, 北京 100176

**摘要** 玛咖(*Lepidium meyenii* Walp.)是生长在高海拔地区的十字花科独行菜属一年生或两年生的草本植物,具有丰富的营养价值和生物保健功效。自2011年被列为新资源食品后,玛咖产业得到了迅速发展,价格不断上涨。由于芜菁(*Brassica rapa* L.)外形与玛咖极为相似,受经济利益驱使,不法商家常将芜菁冒充玛咖,制成玛咖粉、玛咖片和玛咖饮品等牟取暴利,这给玛咖健康产业的有序发展带来了严重的负面影响。因此玛咖的真伪鉴别是非常必要的,但目前对于玛咖的真伪鉴别多为传统方法,快速检测方法较少。采用多光谱成像技术,建立了一种玛咖与芜菁真伪的快速无损鉴别的新方法。主要包括玛咖切片和玛咖粉真伪鉴别两部分内容;其一是针对玛咖切片的真伪鉴别,通过 Videometer A/S 多光谱成像仪对240片玛咖和芜菁切片(玛咖和芜菁切片各120片)进行数据采集,波长范围在可见光区域和近红外区域,分别为405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 780, 850, 870, 890, 910, 940和970 nm共19个波段,为了有效鉴别玛咖和芜菁,首先进行了主成分分析(principal component analysis, PCA),然后结合支持向量机(support vector machines, SVM),遗传算法优化支持向量机(genetic algorithm and support vector machine, GA-SVM)和反向传播神经网络(back propagation neural network, BPNN)算法建立了分析模型,校正集与预测集的样品数量比值为3:1。研究发现PCA分析可以明显地将玛咖和芜菁区分,SVM模型对于玛咖和芜菁切片的预测正确率分别为98.33%,100%,GA-SVM和BPNN模型对玛咖和芜菁切片的鉴别正确率均为100%。其二是针对玛咖粉的真伪鉴别,选择120份玛咖粉,向其中掺入20%,40%,60%,80%,4个不同掺假水平(W/W)的芜菁粉进行多光谱数据采集,结合偏最小二乘(partial least squares, PLS)和最小二乘支持向量机(least squares-support vector machines, LS-SVM)对芜菁的掺假比例进行了定量预测。研究发现,PLS和LS-SVM模型对于玛咖粉中芜菁粉的掺入比例的预测决定系数( $R_p^2$ )分别为0.992和0.994,预测均方根误差(RMSEP)分别为2.718%和2.675%,相对预测误差(RPD)分别为12.782和12.987。相比较而言,LS-SVM模型具有更高的 $R_p^2$ ,RPD较低的RMSEP,对于玛咖粉中掺入芜菁粉比例的预测性能更好。为玛咖真伪的快速无损鉴别提供了一种新方法。

**关键词** 玛咖;芜菁;真伪鉴别;多光谱成像;无损检测

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)01-0152-05

## 引言

玛咖原产于南美秘鲁海拔3500 m以上的安第斯山脉<sup>[1]</sup>,在秘鲁当地食用和药用历史已有2000多年,具有“南美人参”的美誉。它含有多种次生代谢产物如玛咖酰胺、玛咖烯、芥子油苷和甾醇等<sup>[2]</sup>,具有提高生育力、抗疲劳、神经保护和缓解更年期综合征<sup>[3]</sup>等功效。2002年在我国云南、

西藏等地相继引种成功<sup>[4]</sup>,2011年以来产业高速发展,但随着玛咖价格的上涨,利润空间加大化,掺假现象愈发严重,产品质量参差不齐。

目前玛咖真伪鉴别方法主要包括液相色谱、质谱和DNA条形码等。例如,Zhang等基于液相色谱和液相色谱质谱联用分析了不同产地玛咖的芥子油苷,为玛咖的质量评估提供了依据<sup>[5]</sup>;李绍辉等利用超高效液相色谱串联质谱法检测玛咖中的非法添加物<sup>[6]</sup>;Chen等通过DNA条形码对玛咖

收稿日期:2018-11-13,修订日期:2019-03-15

基金项目:国家重点研发计划项目(2017YFF0211004)资助

作者简介:张宏蕊,1993年生,合肥工业大学食品与生物工程学院硕士研究生 e-mail: zhanghongrui\_18@163.com

\* 通讯联系人 e-mail: lei.zheng@aliyun.com

及其掺假物进行了鉴别<sup>[7]</sup>。然而色谱、质谱技术样品预处理复杂,设备昂贵,需要消耗大量的化学试剂,DNA 条形码需要专门的数据库作为支撑。近几年红外光谱和电子鼻技术等快速检测技术开始用于玛咖的分类鉴别,如王元忠等分别利用红外光谱和近红外光谱对玛咖的产地进行了鉴别分类<sup>[8-9]</sup>;党艳婷等基于玛咖中芥子油苷含量通过电子鼻技术对玛咖进行了品质等级分类<sup>[10]</sup>,实现了玛咖品质的快速鉴定。但对于玛咖的掺假研究尚少。作为一种快速无损检测技术,多光谱成像在质量监控方面有着很大的潜力,应用于农产品无损检测领域,主要有三个方向:一是农产品品质的预测分级,如 Martina 等利用多光谱成像成功地将被镰刀菌感染的小麦种子与未感染的小麦种子区分,为小麦种子的质量评估和筛选提供了依据<sup>[11]</sup>;二是加工过程中的质量监测,如脱水胡萝卜片的颜色和水分含量是其质量监测的重要指标,Liu 等用多光谱实现了实时监测胡萝卜切片在热空气脱水情况下颜色和水分的变化情况<sup>[12]</sup>;三是真伪掺假的在线鉴别,如 Liu 等通过多光谱成像有效地鉴别出注水牛肉,而且检测出了牛肉样本中增加的水分含量<sup>[13]</sup>;Xiong 等采用多光谱成像结合多种化学计量学方法成功辨别了辐照虾和非辐照虾,为辐照水产品的检测提供了方法<sup>[14]</sup>。本研究利用多光谱成像结合 SVM, GA-SVM 和 BPNN 模型对玛咖及其掺假物茺菁切片进行了分类鉴别;利用 PLS 和 LS-SVM 模型对玛咖粉中掺入茺菁粉的比例进行了定量预测,旨在为玛咖的掺伪定性定量分析提供一种方法。

## 1 实验部分

### 1.1 材料与设备

玛咖干果(采购于云南宁蒗县,西藏,经云南省农科院科学院药用植物研究所鉴定),茺菁为市售。

Videometer A/S 多光谱成像仪(丹麦);QY-2B 型多功能中药切片机(温州顶历医疗器械有限公司);HC-100T 型多功能粉碎机(河城工贸有限公司);bencbtopk 冷冻干燥机(美国 VirTis 公司)。

### 1.2 方法

#### 1.2.1 样品制备

玛咖和茺菁干果利用自动切片机切成 3 mm 厚度的薄片,冷冻干燥 24 h。玛咖和茺菁切片分别来自不同的玛咖和茺菁干果,各 120 片,其中任意选择 180 片为校正集,60 片为预测集。进行多光谱数据采集。切片实验完成后,模拟市场上玛咖粉掺假,将切片粉碎成细粉末,过 60 目筛。然后茺菁粉以 20%,40%,60%,80%,共 4 个掺假水平(W/W)掺到玛咖粉中。另外,留有纯的玛咖粉和茺菁粉用于实验。将两种粉末充分搅拌混合均匀,将混匀后的样品置于培养皿中,使培养皿表面平整,然后利用多光谱测量仪对不同掺假水平的样品进行光谱测定,每份样品由不同的玛咖和茺菁干果磨成的粉组成,校正集 90 份样品,预测集 30 份样品。

#### 1.2.2 多光谱图像采集

多光谱成像仪可以测量 405,435,450,470,505,525,570,590,630,645,660,700,780,850,870,890,910,

940 和 970 nm 共 19 个波段的光谱反射值,覆盖了可见光以及近红外波长。多光谱成像采集系统由摄像头、LED 灯、积分球等部分组成,测量时将样品放于积分球内,积分球表面涂有白色颜料,光线可以均匀传播,光谱敏感摄像头置于积分球的顶部,LED 灯置于积分球边缘位置。多光谱图像是含有光谱数据和形态信息值的三维图像,原始图像里包含了噪声和冗余信息,需要进行预处理后再提取光谱值。

#### 1.2.3 数据分析

采用的化学计量学方法包括 PCA, SVM, GA-SVM, LS-SVM, BPPN, PLS, 其中需要的软件有 Origin 8.5 和 Matlab 2011a。

PCA 是一种常用的无监督的多元统计学方法,通过运用少量的特征对样品进行描述,达到降低特征空间维数的目的。PLS 是一种线性回归的多元校准模型,BPPN 和线性模型相比,更能准确地解决复杂的问题,在模式识别中有广泛的应用。SVM 是一个可以进行非线性分类、函数估计和核密度估计的一种新型算法,具有较好的泛化能力,广泛应用于统计分析和回归分析中。LS-SVM 是在 SVM 算法基础上作了改进,可以同时进行线性和非线性建模分析,可以快速准确地进行多元建模<sup>[15]</sup>。GA-SVM 是将支持向量机和遗传算法结合起来的一种混合算法,遗传算法提取最优的特征子集,支持向量机确定特征子集的合适值。

#### 1.2.4 模型校准

模型构建后需要评价样本测量值和预测值间的回归模型,判断模型是否准确可靠和稳定,主要通过校准均方根误差(RMSEC)、预测均方根误差(RMSEP)、偏差(bias)、校准( $R_c$ )和预测( $R_p$ )中的决定系数( $R^2$ )以及相对预测误差(RPD)来进行评估。RPD 值越大则预测化学成分准确率越高,一般 RPD 大于 3 可以用于质量控制。通常来说一个好的模型应该具有较高的  $R_c^2$ ,  $R_p^2$  和 RPD 以及较低的 RMSEC, RMSEP 和 bias<sup>[16]</sup>。

## 2 结果与讨论

### 2.1 玛咖及其掺假物茺菁的光谱分析

玛咖及其掺假物茺菁在 405~970 nm 波长范围内的平均反射光谱变化趋势相似,两者随着波长的增加,反射值逐渐增加。但是存在一些差异,玛咖的反射值始终高于茺菁的反射值[图 1(a)],这反映了玛咖及茺菁中物理和化学成分存在差异。在 405~700 nm 可见光范围内,玛咖反射值高于茺菁的反射值主要是和样品的颜色有关。在 940~970 nm 范围内玛咖和茺菁光谱值差异主要是由于 O—H 伸缩或弯曲模式引起的。在不同比例茺菁粉和玛咖粉平均反射光谱中[图 1(b)],所有样品的平均光谱呈相似的变化趋势,随着波长的增加呈上升趋势。样品的反射值与掺入茺菁粉比例呈正相关增长,其中掺入 60%茺菁粉、掺入 80%茺菁粉和茺菁粉 3 组之间反射值递增的梯度较为明显,另外在 700~850 nm 范围内 6 组反射值差异最为显著。由于玛咖中成分复杂,仅依靠化学成分来定性分析是困难的,因此应用化学计量学方法进一步分析光谱数据以获取更好的玛咖鉴别结果。

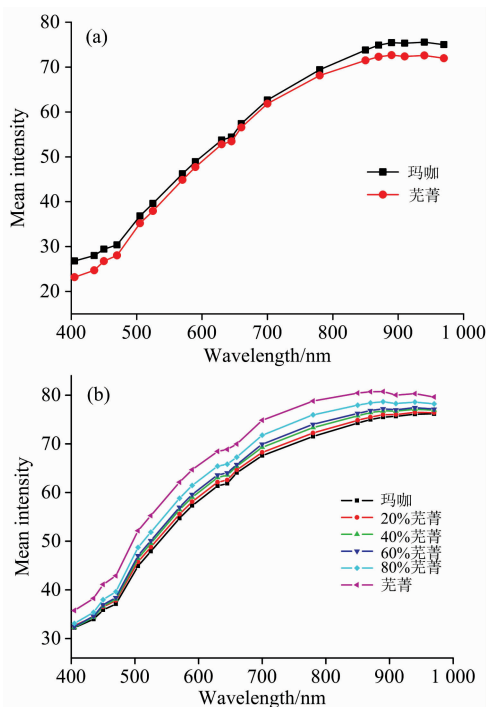


图 1 玛咖和芜菁的平均反射光谱

(a): 玛咖和芜菁切片; (b): 玛咖粉及掺入不同比例的芜菁粉

Fig. 1 Average reflectance spectra of Maca and turnip

(a): Slices of Maca and turnip;

(b): Maca powder with different adulterated levels of turnip powder

## 2.2 玛咖及其掺假物芜菁切片的分类分析

为了对玛咖和芜菁进行有效鉴别, 首先采用 PCA 分析以观察主成分空间中玛咖及其掺假物芜菁样品之间的变化和区分能力。玛咖及其掺假物芜菁的主成分三维得分图结果显示, 三个主成分 PC1, PC2, PC3 总得分达 99.63%, 得分依次为 93.46%, 4.23%, 1.94%。可清晰地将玛咖及其掺假物芜菁分成两组(图 2)。结果表明玛咖和芜菁可以被区分并且样品的不同光谱属性和样品的特征是相关的。在 PCA 可视化分析基础之上建立 SVM, GA-SVM, BPNN 等模型进一步准确验证。240 个玛咖和芜菁样品(玛咖、芜菁各 120), 分为 2 组, 其中校正集 180 个, 预测集 60 个。3 个模型鉴别结果如表 1 所示, 在 SVM 模型的预测集中, 只有 1 个玛咖样品没有正确分类, 玛咖和芜菁的预测正确率分别为 98.33%, 100%。在 GA-SVM 和 BPNN 模型中, 玛咖和芜菁样品全部正确分类, 预测正确率均为 100%。3 个模型相比, GA-SVM 和 BPNN 模型更适于玛咖和芜菁的鉴别, 表明多光谱成像结合 GA-SVM 或 BPNN 模型可以有效实现玛咖及其掺假物芜菁的鉴别。

## 2.3 掺假玛咖粉中芜菁粉的定量预测

在光谱分析中, 最佳校正模型的选择是非常重要的。利用 PLS, LS-SVM 算法建立校准模型对向玛咖粉中掺入 20%, 40%, 60%, 80% 的芜菁粉进行了预测, 共 120 份样品, 其中校正集 90 份, 预测集 30 份。两个模型预测的结果如表 2 所示。根据最低 PRESS 值确定了 PLS 模型的潜变量

个数为 5, 通过留一交叉验证法得到 LS-SVM 模型预测芜菁粉比例的最佳参数( $\gamma, \delta^2$ )为(1 024, 0.008 9)。通过评估  $R_c^2$ ,  $R_p^2$ , RMSEC, RMSEP, RPD 和 bias 选出最佳的分析模型, 最佳模型应具有较高的  $R_c^2$ ,  $R_p^2$  和 RPD 以及较低的 RMSEC, RMSEP 和 bias。表 2 中 LS-SVM 模型的决定系数  $R_c^2$ ,  $R_p^2$ , 分别为 0.997, 0.994, 略高于 PLS 模型( $R_c^2$ ,  $R_p^2$  分别为 0.996, 0.992), 预测集的 RMSEP 和 bias 分别为 2.675%, 0.032%, 低于 PLS 模型的 RMSEP 2.718% 和 bias 1.635%。另外, RPD 作为验证模型准确性的重要参数, 在 LS-SVM 模型中为 12.987 高于 PLS 模型。通常 RPD 值大于 8, 被认为模型实际应用的预测能力非常好。因此 LS-SVM 模型预测玛咖粉中掺入芜菁粉比例的能力较高, 相比于 PLS 模型更适于玛咖粉中芜菁粉掺入比例的预测。PLS 模型预测能力稍弱的原因可能是因为它线性回归的方法, 而 LS-SVM 是非线性的算法, 有更强的自学和调整能力, 而且玛咖成分复杂, 非线性方法更适合。

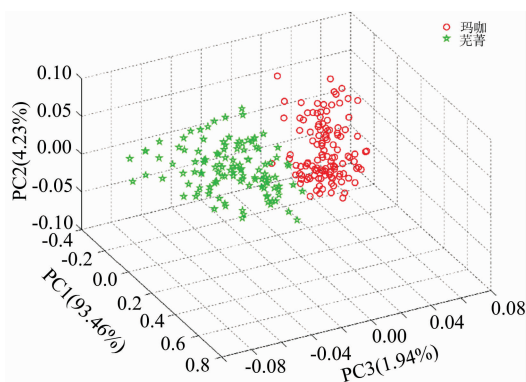


图 2 玛咖和芜菁光谱的三维主成分分析图

Fig. 2 3-D principal component analysis of the Maca and turnip spectra

表 1 用于鉴别玛咖及芜菁切片的 SVM, GA-SVM, BPNN 模型结果对比

Table 1 Comparison of SVM, GA-SVM, and BPNN model results for identifying Maca and turnip slices

	SVM		GA-SVM		BPNN	
	MS <sup>a</sup>	Accuracy/%	MS <sup>a</sup>	Accuracy/%	MS <sup>a</sup>	Accuracy/%
校正集						
玛咖(n=90)	0	100	0	100	0	100
芜菁(n=90)	0	100	0	100	0	100
预测集						
玛咖(n=30)	1	98.33	0	100	0	100
芜菁(n=30)	0	100	0	100	0	100

<sup>a</sup> Misclassified samples.

目前市场上向玛咖粉中掺入芜菁粉的量多少不一, 而且分布范围较广, 实时鉴别难度较大。本研究的主要目的是应用多光谱成像实现玛咖粉中掺假芜菁粉检测的可能性, 通过对 20%, 40%, 60%, 80% 的掺假比例研究后发现, 掺假玛咖样本的多光谱信号随掺假比例呈现规律性变化, 证明了该方法进行芜菁掺假定量分析是可行的。在将来的研究中, 需

要对模型预测的最低检测水平进行研究, 并对定量分析的稳定性进行优化, 使得建立的模型能够更好地应用于市场上玛

咖掺假的快速无损检测。总之, 多光谱成像结合多元统计学方法在玛咖真伪鉴别方面有很大的发展空间和应用前景。

表 2 用于预测玛咖粉中茛菪粉掺假比例的 PLS 和 LS-SVM 模型结果对比  
Table 2 Comparison of PLS and LS-SVM models for predicting the percentage of adulteration of turnip powder in Maca powder

方法	校正集				预测集			
	$R_c^2$	RMSEC/%	RPD	bias/%	$R_p^2$	RMSEP/%	RPD	bias/%
PLS	0.996	2.132	16.079	$-9.144 \times 10^{-13}$	0.992	2.718	12.782	1.635
LS-SVM	0.997	1.536	22.366	0.099	0.994	2.675	12.987	0.032

注:  $R_c^2$ : 校准绝对系数(coefficient of determination in calibration);  $R_p^2$ : 预测绝对系数(coefficient of determination in prediction); RMSEC: 校准均方根误差(root mean square error of calibration); RMSEP: 预测均方根误差(root mean square error of prediction); RPD: 相对预测误差(residual predictive deviation); bias: 偏差

### 3 结 论

本研究的创新性主要在于利用多光谱成像结合不同化学计量学方法, 包括 PCA, SVM, GA-SVM, BPNN, PLS 和 LS-SVM 对玛咖及茛菪切片进行了鉴别以及对玛咖粉中茛菪

粉的掺入比例进行了预测。结果表明, 对玛咖及茛菪切片的鉴别, 两者可以成功区分, GA-SVM 和 BPNN 模型有较好的校准和预测能力, 准确率为 100%; 对玛咖粉中掺入茛菪粉比例的预测, LS-SVM 模型的预测性能最佳。多光谱成像结合化学计量学方法可以为玛咖掺伪定性定量分析提供一种快速无损有效的新方法。

### References

- [ 1 ] Zha S, Zhao Q, Chen J, et al. Carbohydr Polym, 2014, 111: 584.
- [ 2 ] Chen L F, Li J Y, Fan L P, et al. Journal of Food Quality, 2017, (2): 1.
- [ 3 ] Stojanovska L, Law C, Lai B, et al. Climacteric, 2015, 18(1): 1.
- [ 4 ] YANG Min, HUANG Shao-jun, ZHU Yan-qin, et al(杨敏, 黄绍军, 朱艳琴, 等). Food Chemistry(食品科学), 2016, 37(2): 132.
- [ 5 ] Zhang L M, Cao J, Hao L M, et al. Food Analytical Methods, 2017, 10(7): 2013.
- [ 6 ] LI Shao-hui, WU Han-qiu, XU Xiu-li, et al(李绍辉, 吴寒秋, 许秀丽, 等). Science and Technology of Food Industry(食品工业科技), 2016, 37(11): 310.
- [ 7 ] Chen J J, Zhao Q S, Liu Y L, et al. Chinese Journal of Natural Medicines, 2015, 13(9): 0653.
- [ 8 ] WANG Yuan-zhong, ZHAO Yan-li, ZHANG Ji, et al(王元忠, 赵艳丽, 张霁, 等). Food Chemistry(食品科学), 2016, 37(4): 169.
- [ 9 ] WANG Yuan-zhong, ZHAO Yan-li, ZHANG Ji, et al(王元忠, 赵艳丽, 张霁, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(2): 394.
- [ 10 ] DANG Yan-ting, YUAN Peng, XIA Kai, et al(党艳婷, 苑鹏, 夏凯, 等). Food Chemistry(食品科学), 2018, 39(6): 291.
- [ 11 ] Martina V, Merete H O, René G, et al. Plos One, 2016, 11(3): 1.
- [ 12 ] Liu C H, Liu W, Lu X Z, et al. Food Chemistry, 2016, 195: 110.
- [ 13 ] Liu J X, Cao Y, Wang Q, et al. Food Chemistry, 2016, 190: 938.
- [ 14 ] Xiong C W, Liu C H, Liu W, et al. Journal of Food Engineering, 2016, 175: 85.
- [ 15 ] Liu C H, Liu W, Chen W, et al. Food Chemistry, 2015, 173: 482.
- [ 16 ] Liu C H, Hao G, Su M, et al. Journal of Food Engineering, 2017, 215: 78.

# Qualitative Identification and Quantitative Analysis of Maca Adulteration Based on Multispectral Imaging Technology

ZHANG Hong-ru<sup>1, 2</sup>, LIU Chang-hong<sup>1</sup>, ZHANG Jiu-kai<sup>2</sup>, HAN Jian-xun<sup>2</sup>, CHEN Ying<sup>2</sup>, ZHENG Lei<sup>1\*</sup>

1. School of Food and Biological Engineering, Hefei University of Technology, Hefei 230009, China

2. Chinese Academy of Inspection and Quarantine, Beijing 100176, China

**Abstract** Maca (*Lepidium meyenii* Walp.), an annual or biennial herb of Brassicaceae family, grows at high altitudes and contains rich nutritional value and bio-health benefits. After being listed as a new resource food in 2011, Maca is gradually becoming familiar to the public, the Maca industry has developed rapidly and price has risen steadily. Due to the fact that the shape of turnip (*Brassica rapa* L.) is very similar to that of Maca, driven by economic interests, illegal businessmen often pass turnip off as Maca to make Maca powder, slices and drinks in order to make exorbitant profits, which has brought serious negative impact on the orderly development of Maca healthy industry. Therefore, the authenticity identification of Maca is very necessary, but most of methods for the authenticity identification of Maca are traditional, and there are few rapid detection methods. In this study, a new method for rapid and non-destructive identification of Maca and turnip was established by using multispectral imaging technology. The experiment mainly focuses on the authenticity identification of Maca slices and Maca powder. One is to identify the authenticity of Maca slices. A total of 240 Maca and turnip slices (120 Maca slices and 120 turnip slices, respectively) were selected to collect data by the Videometer Lab equipment, which acquired the multispectral images at 19 different wavelengths from the visual region to the lower wavelengths of the NIR region and the detailed information of the measured wavelength were 405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 780, 850, 870, 890, 910, 940 and 970 nm. In order to identify Maca and turnip effectively, the principal component analysis (PCA) was first performed. Then the qualitative analysis model was generated using support vector machine (SVM), genetic algorithm optimization support vector machine (GA-SVM) and back propagation neural network (BPNN) algorithm, and the ratio of the the calibration set to the prediction set is 3 : 1. The results demonstrated that clear differences between Maca and turnip could be easily visualized by PCA. The predictive accuracies by SVM model for Maca and turnip slices were 98.33% and 100%, respectively, and the predictive accuracies by GA-SVM and BPNN model could be as high as 100%. The other is the identification of Maca powder. 120 samples of Maca powder were selected and 20%, 40%, 60%, 80%, 4 different adulterated levels (W/W) of turnip powder were mixed for multispectral data acquisition, Combining partial least squares (PLS) and least squares support vector machine (LS-SVM), the adulteration ratio of turnip was quantitatively predicted. The study found that the prediction coefficient ( $R_p^2$ ) of PLS and LS-SVM models were 0.992 and 0.994, the predicted root mean square error (RMSEP) were 2.718% and 2.675% and the relative prediction error (RPD) were 12.782 and 12.987, respectively. In comparison, the LS-SVM model had higher  $R_p^2$ , RPD, lower RMSEP, so it was considered to have better predictive performance for the proportion of turnip powder adulterated to Maca powder. In conclusion, the research results provide a method for the rapid and non-destructive identification of Maca authenticity.

**Keywords** Maca; Turnip; Authentication; Multispectral imaging; Non-destructive detection

(Received Nov. 13, 2018; accepted Mar. 15, 2019)

\* Corresponding author