

DOI: 10.13652/j.spjx.1003.5788.2024.80130

# 基于随机森林的原位质谱法快速鉴别 铁棍山药真伪

钟恒艳<sup>1</sup> 陈 春<sup>1</sup> 欧阳永中<sup>1,2</sup> 周 林<sup>3</sup> 郭伟清<sup>1,4</sup>

(1. 佛山大学, 广东 佛山 528000; 2. 佛山职业技术学院, 广东 佛山 528137; 3. 广东一方制药有限公司, 广东 佛山 528244; 4. 有研(广东)新材料技术研究院, 广东 佛山 528000)

**摘要:** [目的] 建立一种快速无损的铁棍山药真伪鉴别方法。[方法] 在常温常压环境下, 运用大气压化学电离质谱技术对不同产地的铁棍山药(TG)和非铁棍山药(FTG)化学成分进行检测, 每种铁棍山药样品和非铁棍山药样品各采集 200 组数据, 共获得 3 600 个质谱数据点, 利用主成分分析(PCA)和随机森林(RF)算法对获得的一级质谱数据进行模式识别分析, 建立铁棍山药与非铁棍山药真伪鉴别模型。[结果] 通过 HS-APCI-MS 获得的铁棍山药样品与非铁棍山药样品的一级质谱图差异明显, 根据主成分累计方差贡献图, 前 7 个主成分的累计方差贡献率为 85.63% ( $\geq 85\%$ ); 当决策树数量为 25 时, 训练集、检测集准确率均达到 100%, 所建立的铁棍山药原位质谱分析结合 RF 算法对铁棍山药的鉴别效果显著, RF 的分类效果优于 PCA。[结论] 应用原位质谱分析技术结合 RF 算法可以快速无损鉴别铁棍山药真伪。

**关键词:** 铁棍山药; 大气压化学电离源; 主成分分析; 随机森林算法; 鉴别

## Rapid identification of the authenticity of iron rod yam by in-situ mass spectrometry based on random forest algorithm

ZHONG Hengyan<sup>1</sup> CHEN Chun<sup>1</sup> OUYANG Yongzhong<sup>1,2</sup> ZHOU Lin<sup>3</sup> GUO Weiqing<sup>1,4</sup>

(1. School of Foshan University, Foshan, Guangdong 528000, China; 2. Foshan Polytechnic, Foshan, Guangdong 528137, China; 3. Pharmacy Department, Guangdong Yifang Pharmaceutical Co., Ltd., Foshan, Guangdong 528244, China; 4. GRINM (Guangdong) Institute for Advanced Materials and Technology, Foshan, Guangdong 528000, China)

**Abstract:** [Objective] To establish a fast and nondestructive analysis method for identifying iron rod yam. [Methods] Atmospheric pressure chemical ionization mass spectrometry (APCI-MS) was employed to detect the chemical constituents of iron rod yam(TG) and non-iron rod yam (FTG) from different origins under ambient temperature and pressure, With 200 sets of data collected from each type of TG and FTG, and a total of 3 600 mass spectrometry data points were obtained. Subsequently, the initial level of the mass spectrometry data obtained was analyzed using Principal Component Analysis (PCA) and the random forest (RF) algorithm. Pattern recognition analysis established a model to differentiate between TG and FTG based on their chemical compositions. [Results] The difference between the first-level mass spectra obtained by HS-APCI-MS was obvious between TG samples and FTG samples. The cumulative variance contribution plot of the principal components showed that the first seven principal components accounted for 85.63% ( $\geq 85\%$ ) of the variance. The accuracy of the training set and detection set reached 100% when the number of decision trees was 25. HS-APCI-MS combined with RF algorithm had a significant identification effect on TG, and the classification effect of RF was superior to that of PCA. [Conclusion] Atmospheric pressure chemical ionization mass spectrometry, combined with the RF algorithm, can rapidly and non-destructively identify TG and FTG, providing a new technical method for authenticating TG.

**Keywords:** iron rod yam; atmospheric chemical ionization source; principal component analysis; random forest algorithm; identification

基金项目: 中药复杂体系原位质谱分析方法研究(项目批号 2021ZDZX2060); 佛山市南海区重点领域科技攻关专项

通信作者: 欧阳永中(1980—), 男, 佛山大学教授, 博士。E-mail: ouyang7492@163.com

郭伟清(1989—), 女, 有研(广东)新材料技术研究院工程师, 博士。E-mail: guoweiqing@grinm.com

收稿日期: 2024-02-03 改回日期: 2024-10-11

铁棍山药(*Dioscorea opposita* cv. 'Tiegun')原产地为河南焦作温县铁棍山一带<sup>[1]</sup>,故名“铁棍山药”。其在《神农本草经》《雷公炮制药性解》《本草经集注》等古籍中均有记载,含有皂苷、黏液质、胆碱等多种活性成分,药食兼优<sup>[2]</sup>。但在过去很长一段时间内,铁棍山药的种植面积逐年减少<sup>[2]</sup>。近年在中医药振兴发展行动的推崇下<sup>[3]</sup>,铁棍山药相应市场需求愈发增大。中国铁棍山药产量少,价格高,常有不良商贩将麻山药、太谷山药、其他杂交山药等普通山药作为铁棍山药售卖的情况发生,严重扰乱了市场秩序。此外,普通山药与铁棍山药外观相似,无法通过肉眼进行区分。

目前,鉴别山药的常规分析方法有性状鉴别<sup>[4]</sup>、显微鉴别<sup>[5]</sup>、红外光谱鉴别<sup>[6]</sup>、薄层色谱鉴别<sup>[7]</sup>、DNA 测序鉴别<sup>[8]</sup>、核磁共振代谢组学技术<sup>[9]</sup>等,这些方法均存在前处理复杂、试验过程操作繁琐、耗时长、结果准确性低、无法实现大批量样品的快速检测等缺点。顶空进样结合大气压化学电离质谱技术(HS-APCI-MS)是一种新型的无损原位质谱分析技术,无需复杂样品前处理,检测速度快,在检测过程中不会破坏样品中的化合物,具有无损、绿色、高效、耗时短、操作简单等特点。

化学计量法作为一种数据处理方法,与 HS-APCI-MS 结合应用,可高效处理庞大且复杂的质谱数据,建立数据模型,筛选出重要的特征生物标志物并进行结构表征,从而达到快速鉴别的目的。主成分分析(PCA)是一种多变量统计方法<sup>[10-12]</sup>,用于考察多个变量之间的相关性,是目前应用最广泛的多维数据分析工具之一。随机森林算法(RF)是一种高度灵活的机器学习算法,也是一种基于 Bagging 的重要的集成算法<sup>[13]</sup>,该算法在组合多棵决策树的基础上进行优化,主要用于解决分类和回归问题<sup>[14]</sup>,RF 算法能够有效处理海量高维质谱数据,并能够从海量高维数据中快速筛选出特征化合物信息<sup>[15]</sup>。

研究拟针对山药基质复杂、组分含量差异大、挥发性成分较多等特点,构建基于顶空进样结合大气压化学电离质谱法(HS-APCI-MS),并对仪器装置参数和试验条件进行优化,得到铁棍山药样本原位质谱分析的最佳试验条件。利用 HS-APCI-MS 方法获取的一级质谱样本数据,结合主成分分析(PCA)和随机森林算法(RF)构建识别模型<sup>[16-17]</sup>,快速区分铁棍山药与非铁棍山药,以期为铁棍山药的道地属性鉴别提供一种准确有效的方法参考。

## 1 材料与方法

### 1.1 材料、仪器与试剂

山药:采集不同产地的铁棍山药及非铁棍山药(包括

广山药、太谷山药、麻山药),详细信息见表 1,且均经专家鉴定;

线性离子阱质谱仪:LTQ-XL21922 型,美国 Thermo Fisher Scientific 公司;

顶空进样结合大气压化学电离源:HS-APCI 型,实验室自主搭建;

电热恒温水浴锅:DK-98-II 型,天津市泰斯特仪器有限公司;

电子天平:JJ124BF 型,常熟市双杰测试仪器厂;

氮气、氦气:纯度 99.99%,液化空气(广东)工业气体有限公司;

甲醇:HPLC 级,塞默飞世尔科技(中国)有限公司。

### 1.2 试验方法

准确称取山药固体样品 2.000 g,置于顶空进样瓶中水浴 10 min。设置最佳试验参数,将两根气体导管接入顶空瓶中,使载气能够在顶空瓶和导管中自由流通,实现动态顶空进样。利用氮气作为载气将待测物质吹扫至电离区域进行分子-离子反应,目标化合物被离子化后进入质谱中进行检测。以干净且无样品的顶空瓶作为空白背景信号,所有谱图扣除空白背景,得到多组不同样本的一级指纹图谱。每个样本提取 200 组数据进行模式识别分析,获得铁棍山药与非铁棍山药的鉴别模型,实现两者的鉴别。

HS-APCI 离子源条件:载气压力 1.0 MPa,平衡时间 10 min,平衡温度 70 °C,样品出口与质谱进样口距离 1.0 cm,高压放电针与出样口之间的夹角 180°,两者在同一水平面上。

质谱仪参数:质谱扫描范围  $m/z$  50~500,正离子模式,电晕针电压 4.5 kV,离子传输管温度 155 °C。其他值采用默认参数以及系统自动优化。

### 1.3 数据处理及分析

使用 Xcalibur4.2 软件将质谱数据导入 Excel 中,利用 Matlab 软件对质谱数据进行主成分分析(PCA)以区分铁棍山药样品与非铁棍山药样品;利用 Matlab 内部开发的 RF 对数据进行分析,以分类鉴别铁棍山药与非铁棍山药。

## 2 结果与讨论

### 2.1 铁棍山药与非铁棍山药的质谱指纹图谱分析

采用正离子模式,在优化后的 HS-APCI-MS 条件下,分别对铁棍山药样品和非铁棍山药样品进行原位质谱检测分析,得到各自的一级质谱指纹图谱。由图 1 和图 2 可知,铁棍山药与非铁棍山药的一级质谱指纹图谱的信号峰种类和信号强度存在差异。

表1 山药样品信息

Table 1 Sample information of yam

样品类型	拉丁学名	采集地	基源物种
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市温县(前期)	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市温县(中期)	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市温县(后期)	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市温县武德镇乡宋庄村	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南焦作市武陟县西陶镇大甬张村	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市武陟县郭乡西镇村	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市温县番田镇西留石村	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市温县招贤乡单庄村	薯蓣
铁棍山药(冻干粉)	<i>Dioscorea opposita</i>	河南省焦作市武陟县大封乡驾部2村	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市郑章镇郑章村	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市明官店乡	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市明官店乡	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市蠡县大曲堤乡	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	山西运城市新绛县龙兴镇桥东村	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市北段村乡瓦子里村	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市石佛镇路景村	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市石佛镇石佛村	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市蠡县	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市安国市	薯蓣
麻山药(冻干粉)	<i>Dioscorea opposita</i>	河北省保定市博野县	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	山西省运城市永济市蒲州镇	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	河南省焦作市解放区丰收路	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	河南省焦作市孟州市	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	河南省焦作市温县	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	山东省菏泽市定陶县陈集镇	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	山东省菏泽市定陶县陈集镇七一村	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	山东省菏泽市定陶区中华路	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	河南省焦作市温县余村	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	河南省焦作市温县陆庄	薯蓣
铁棍山药(饮片)	<i>Dioscorea opposita</i>	河南省焦作市温县南张伦	薯蓣
太谷山药(饮片)	<i>Dioscorea opposita</i>	山西省晋中市平遥县道备村头道街	薯蓣
太谷山药(饮片)	<i>Dioscorea opposita</i>	山西省晋中市平遥县中都中街49号	薯蓣
太谷山药(饮片)	<i>Dioscorea opposita</i>	山西吕梁市岚县东村镇新建路	薯蓣
麻山药(饮片)	<i>Dioscorea opposita</i>	河北省衡水市安平县	薯蓣
麻山药(饮片)	<i>Dioscorea opposita</i>	河北省保定市蠡县大曲堤乡高佐村	薯蓣
麻山药(饮片)	<i>Dioscorea opposita</i>	河北省保定市蠡县大曲堤乡北绪口村	薯蓣
麻山药(饮片)	<i>Dioscorea opposita</i>	河北省保定市蠡县大曲堤乡东庞果庄村	薯蓣
广山药(饮片)	<i>Dioscorea opposita</i>	广西壮族自治区贵港市桂平市金田镇	薯蓣
广山药(饮片)	<i>Dioscorea opposita</i>	广西壮族自治区玉林市陆川县	薯蓣

## 2.2 主成分分析

PCA作为一种常见的数据降维技术,能够将高维数据转换为低维数据并保留数据集中的关键信息。试验分

别扫描和采集铁棍山药与非铁棍山药样本  $m/z$  50~500 区域内的 HS-APCI-MS 一级质谱数据,对海量高维原始质谱数据进行 PCA 分析,得到各主成分的得分图、主成分累

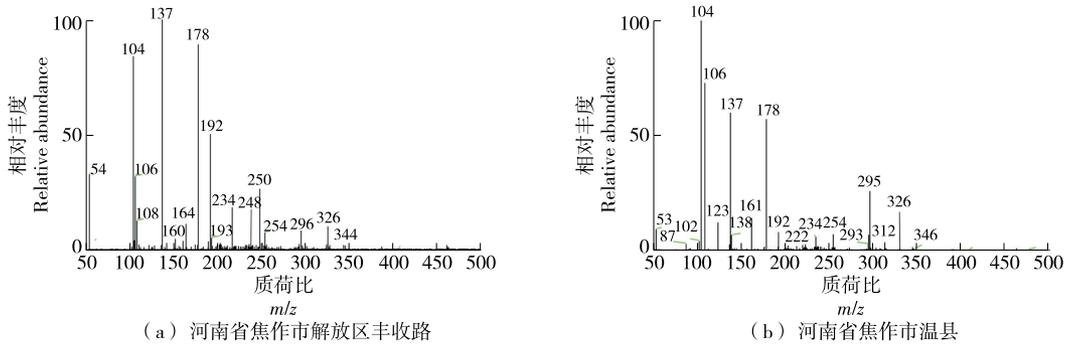


图 1 正离子模式下部分铁棍山药样品的质谱指纹图谱  
Figure 1 Mass spectrometry fingerprints of TG under the positive ion mode

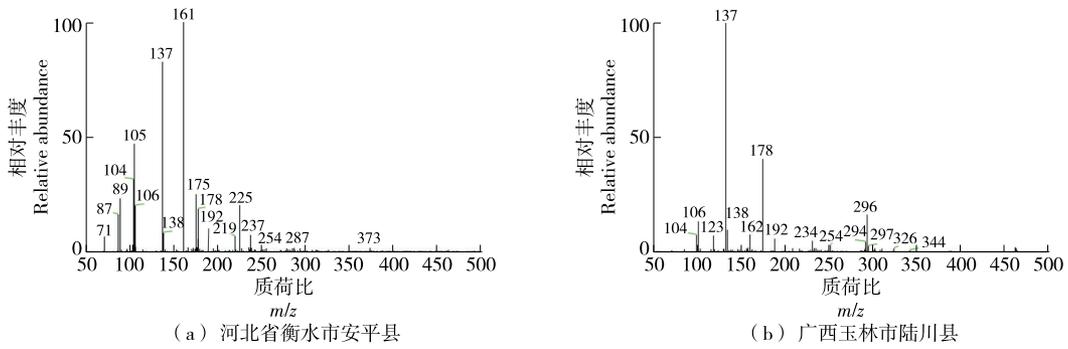


图 2 正离子模式下部分非铁棍山药样品的质谱指纹图谱  
Figure 2 Mass spectrometry fingerprints of FTG under the positive ion mode

计方差贡献率图和相应的载荷图。

由图 3(a)可知,两组山药样本质谱原始数据存在一定的差异,其中代表铁棍山药样本的蓝色正方形与代表非铁棍山药样本的红色圆形能够被区分开。PCA 得分图显示,铁棍山药样品与非铁棍山药样品分布在相对独立的区域,可得到较为明显的区分,说明样品成分的种类和相对含量存在差异。铁棍山药样本与非铁棍山药样本存在部分重叠的区域,说明该鉴别模型分类效果一般。由图 3(b)可知,前 7 个主成分的累计方差贡献率为 85.63% ( $\geq 85\%$ ), PC1、PC2、PC3 的贡献率分别为 42.48%、12.73%、11.89%,前 3 个主成分的解释方差达到了 67.10%,包含了绝大部分样本信息。由图 3(c)可知,PC1 主要受  $m/z$  104, 106, 137, 178 的影响,PC2 主要受  $m/z$  142, 172, 178 的影响,PC3 主要受  $m/z$  106, 109, 123, 178 的影响。

综上,将获得的山药样品质谱数据进行融合后所构建的 PCA 鉴别模型出现交叉聚集现象,区分效果一般,说明使用该鉴别模型会出现少部分的假阳性现象。这可能是由于 PCA 算法作为一种无监督算法,在样品质谱数据中包含了大量其他差异信息时无法很好地区分两种山药,从而导致鉴别出现假阳性现象。此外,不同产地、不

同种类的山药样品可能存在相似性,有着相同的挥发性物质,或者某些产地的铁棍山药样本特征成分含量低,造成样本区分时存在部分重叠的现象,导致分类效果不显著。

### 2.3 随机森林算法

2.3.1 RF 鉴别模型的构建 将获得的铁棍山药样本与非铁棍山药样本原始数据集按 1:1 比例随机分为训练集和检验集,训练集用于训练和建立分类模型,检验集用于检验和评估所建立的模型的性能。由于铁棍山药样本来自不同的产地,各个产地的地理位置、土壤条件、海拔、气候等因素有所差异,加上各地栽培方法、培育时间以及采收方式的不同,导致铁棍山药样本之间的各成分及其含量不同,可通过优化 RF 模型提高对铁棍山药与非铁棍山药的鉴别能力。

由图 4 可知,当建模分类树数量为 0~25 时,随着决策树数量的增加,袋外数据误差值(OOB)随之降低。当决策树数量为 25~500 时,OOB 值低于 0.005 且趋于稳定,为最优模型,故选取决策树数量值为 25 建立模型。

由图 5 可知,铁棍山药样品和非铁棍山药样品之间呈现明显的分隔,无重叠部分,表明模型具有优秀的分类能力。由表 2 可知,HS-APCI-MS 方法结合 RF 可以有效地

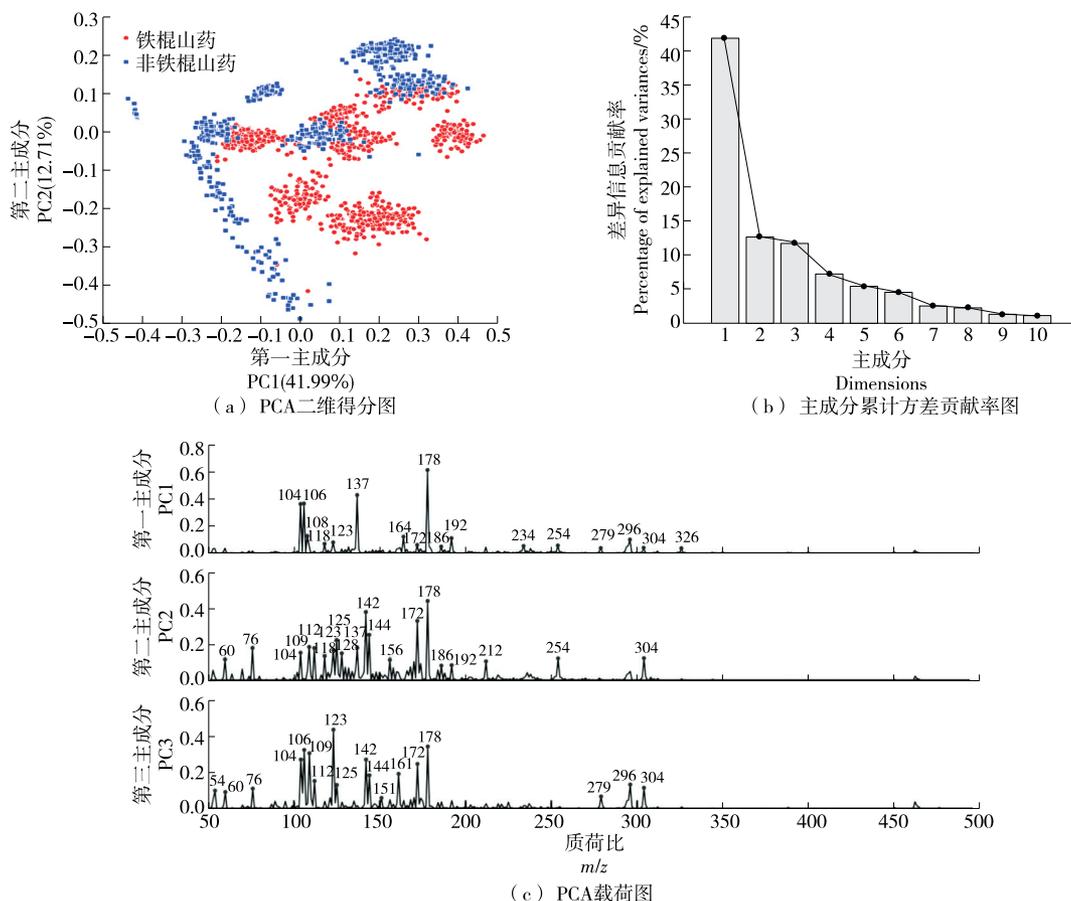


图3 铁棍山药和非铁棍山药样本数据的PCA分析结果

Figure 3 TG and FTG samples data PCA analysis

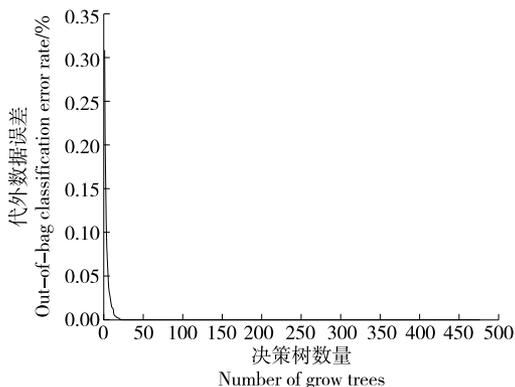


图4 随机森林决策树目优化过程图

Figure 4 Parameter number of trees selected by the Random Forest after finetuning the model

对铁棍山药样本和非铁棍山药样本进行区分。

综上,与PCA法相比,RF对于铁棍山药和非铁棍山药的鉴别效果更优。

2.3.2 RF的特征变量重要性评估 RF的预测误差率OOB相比原来特征置换前会增大,特征重要程度越高,模

表2 混淆矩阵分类结果

Table 2 Result of classificaton of confusion matrix

实际类别	预测类别			
	训练集		检验集	
	铁棍山药	非铁棍山药	铁棍山药	非铁棍山药
铁棍山药	100	0	100	0
非铁棍山药	0	100	0	100

型预测误差率OOB变化值越大<sup>[21]</sup>。因此,RF算法可通过调整变量顺序得到OOB预测错误率来衡量特征变量的重要性<sup>[13]</sup>。由图6可知,VIP得分越高,该特征碎片离子对模型准确性的贡献率越大。铁棍山药样本与非铁棍山药样本的重要离子碎片变量有 $m/z$  60, 109, 148, 120, 78, 53等,特征化合物可视为鉴别铁棍山药与非铁棍山药的潜在生物标志物,是区分铁棍山药样本与非铁棍山药样本的重要标志物。说明HS-APCI-MS结合RF可快速、高效获得其他大部分特征生物标志物,在食品、中药材等复杂体系样品的道地属性鉴别和质量控制等研究中具有潜在的应用价值。

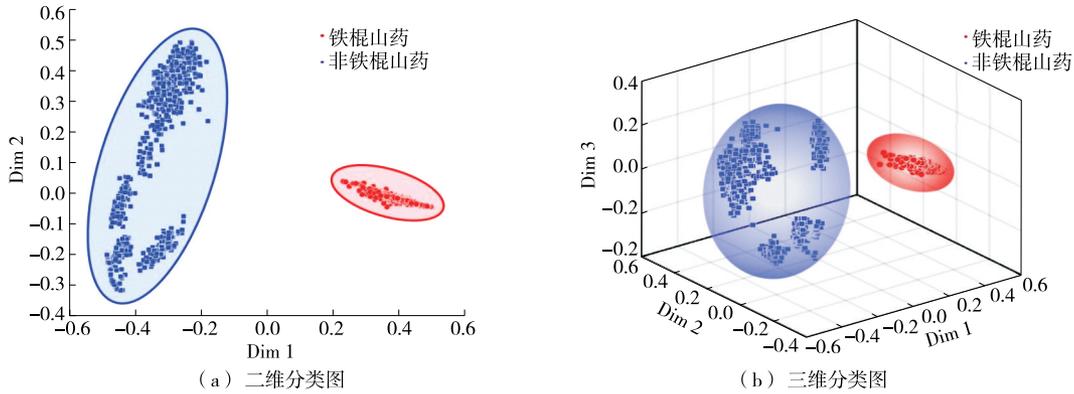


图 5 样本数据的随机森林多维标度分类图

Figure 5 Proximity of RF for distinguishing TG from yam samples using multi-dimensional scaling (MDS)

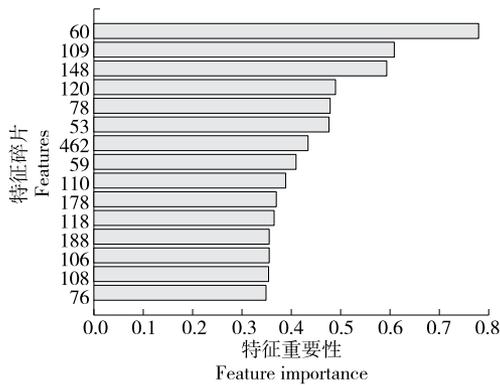


图 6 铁棍山药样本区分非铁棍山药样本的变量重要性排序

Figure 6 Variables importance ranking for distinguishing TG from yam samples

### 3 结论

以铁棍山药和非铁棍山药(包括太谷山药、广山药、麻山药)为研究对象,通过顶空进样结合大气压化学电离质谱法方法和模式识别技术(包括主成分分析和随机森林算法)对样本的一级质谱数据进行了分析,建立了区分铁棍山药与其他山药的数据模型。结果表明,顶空进样结合大气压化学电离质谱法结合主成分分析或随机森林算法均可用于区分铁棍山药样品和非铁棍山药样品。然而,使用顶空进样结合大气压化学电离质谱法结合主成分分析构建的鉴别模型在区分效果方面表现一般;相比之下,随机森林算法在鉴别铁棍山药与非铁棍山药方面效果更佳,准确率可达 100%,验证了随机森林算法模型在鉴别铁棍山药与非铁棍山药方面的有效性。此外,利用随机森林算法分析的变量重要性评分,筛选出了区分两者的潜在差异性化合物。研究表明,原位质谱方法结合随机森林算法能够有效鉴别铁棍山药的真伪。基于

此,未来的研究可以通过扩大样本量来构建更稳健的模型,并探索其他更优的模式识别方法,以对数据进行更深入的研究,从而进一步提升鉴别的准确性和适用性。

### 参考文献

[1] WANG Q, WANG X, WU X, et al. H<sup>-1</sup>NMR-based metabolic profiling approach to identify the geo-authentic Chinese yam (*Dioscorea polystachya* Turczaninow cv. Tiegung)[J]. Journal of Food Composition and Analysis, 2021, 98(1): 103805.

[2] 訾帅朋, 田大伟, 常青晓. 温县铁棍山药产业发展调研报告[J]. 河南农业, 2022(7): 4.

ZI S P, TIAN D W, CHANG Q X. Research report on the development of Wen County's Tiegung industry[J]. Henan Agriculture, 2022(7): 4.

[3] 安莉, 汪红, 马婧玮, 等. 基于 UPLC-Q/TOF-MS/MS 和生物信息学探讨铁棍山药皮中化学成分的药用和营养价值[J]. 食品工业科技, 2023, 44(2): 1-9.

AN L, WANG H, MA J W, et al. Medicinal and nutritional value of the chemical compositions of *Dioscorea opposita* Thunb. cv. tiegun peel based on UPLC-Q/TOF-MS/MS and bioinformatics [J]. Science and Technology of Food Industry, 2023, 44(2): 1-9.

[4] 杨燕梅, 郑学东, 王海涛, 等. 山药和广山药的生药鉴别研究[J]. 中兽医医药杂志, 2020, 39(3): 87-89.

YANG Y M, ZHENG X D, WANG H T, et al. Determination of colchicine content in *Lilium lancifolium* Thunb and its bulbs[J]. Journal of Traditional Chinese Veterinary Medicine, 2020, 39(3): 87-89.

[5] 国家药典委员会. 中华人民共和国药典: 四部[S]. 2020 版. 北京: 中国医药科技出版社, 2020: 230-231.

Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China: 4 volumes[S]. 2020 ed. Beijing: China Medical Science Press, 2020: 230-231.

[6] 李长滨, 牛畅炜, 苏丽, 等. 不同产地山药的近红外鉴别和差异分析[J]. 食品研究与开发, 2022, 43(15): 175-181.

- LI C B, NIU C W, SU L, et al. Identification and variance analysis of chinese yam from different origins by near infrared spectroscopy[J]. Food Research and Development, 2022, 43 (15): 175-181.
- [7] 唐庆, 王然, 杨德泉, 等. 土炒山药的薄层色谱鉴别及尿囊素含量测定研究[J]. 药品评价, 2022, 19(14): 853-857.
- TANG Q, WANG R, YANG D Q, et al. Study on thin layer chromatographic identification and determination of urantoin content of soil fried yam[J]. Drug Evaluation, 2022, 19(14): 853-857.
- [8] 孟啸龙, 孟乡, 周国富, 等. DNA条形码在山药及其混伪品鉴定中的应用[J]. 辽宁中医杂志, 2020, 47(4): 163-166.
- MENG X L, MENG X, ZHOU G F, et al. Application of DNA barcode in identification of *D. oppositifolia* and its adulterants [J]. Liaoning Journal of Traditional Chinese Medicine, 2020, 47 (4): 163-166.
- [9] 魏亮, 张超楠, 李科研, 等. 基于核磁共振代谢组学技术比较不同产地铁棍山药的化学成分差异[J]. 分析测试学报, 2022, 41(5): 774-780.
- WEI L, ZHANG C N, LI K Y, et al. Comparison of chemical constituents in *Dioscorea polystachya* Truczaniow cv. tiegun planted in different counties by <sup>1</sup>H-NMR metabonomics approach[J]. Journal of Instrumental Analysis, 2022, 41(5): 774-780.
- [10] KOH J, KIM B J, QU Y, et al. Mass spectral profiling of caseinomacropptide extracted from feeding material and jejunal fluid using three methods-ethanol precipitation, perchloric acid precipitation, and ultrafiltration[J]. Food Chemistry, 2023, 398(1): 133864.
- [11] 孙强, 王丽艳, 荆瑞勇, 等. 不同品种亚麻籽8种生物活性物质含量及其主成分分析[J]. 食品与机械, 2022, 38(9): 40-45, 145.
- SUN Q, WANG L Y, JING R Y, et al. Determination and principal component analysis of 8 kinds of bioactive[J]. Food & Machinery, 2022, 38(9): 40-45, 145.
- [12] 黄盛蓝, 杜木英, 周先容, 等. 发泡泡菜品质及风味物质主成分分析[J]. 食品与机械, 2017, 33(12): 36-44.
- HUANG S L, DU M Y, ZHOU X R, et al. Quality indifferent softness of pickles and its principal components analysis of volatiles[J]. Food & Machinery, 2017, 33(12): 36-44.
- [13] BREIMAN L. Random forest[J]. Machine Learning, 2001, 45 (1): 5-32.
- [14] 陈卓, 丁利, 曹天红, 等. 基于层次分析法和随机森林回归算法的谷物资源风险评估模型[J]. 食品与机械, 2021, 37(12): 58-66.
- CHEN Z, DING L, CAO T H, et al. Risk assessment model of grain resource based on AHP and random forest regression algorithm[J]. Food & Machinery, 2021, 37(12): 58-66.
- [15] SIMIC V, TORKAYESH A E, MAGHSOODI A I. Locating a disinfection facility for hazardous healthcare waste in the COVID-19 era: a novel approach based on Fermatean fuzzy ITARA-MARCOS and random forest recursive feature elimination algorithm[J]. Annals of Operations Research, 2023, 328(1): 1 105-1 150.
- [16] 欧阳永中, 曾玉庭, 郭伟清, 等. 基于随机森林算法的原位质谱快速鉴别肺癌的方法研究[J]. 分析化学, 2020, 48(8): 1 012-1 022.
- OUYANG Y Z, ZENG Y T, GUO W Q, et al. Rapid identification of lung cancer by in situ mass spectrometry based on the random forest algorithm[J]. Chinese Journal of Analytical Chemistry, 2020, 48 (8): 1 012-1 022.
- [17] 丁莎, 申涛榕, 张艳飞, 等. 基于随机森林算法的烟草提取物类别识别模型研究[J]. 分析测试学报, 2023, 42(11): 1 510-1 516.
- DING S, SHEN T R, ZHANG Y F, et al. Study on classification model for tobacco extracts based on random forest algorithm[J]. Journal of Instrumental Analysis, 2023, 42 (11): 1 510-1 516.
- [18] 武炜杰, 张景祥. 融合分类信息的随机森林特征选择算法及应用[J]. 计算机工程与应用, 2021, 57(17): 147-156.
- WU W J, ZHANG J X. Random forest feature selection algorithm based on categorization information and application [J]. Computer Engineering and Application, 2021, 57(17): 147-156.