

基于太赫兹时域光谱技术的掺假川贝母检测

徐 振, 刘燕德*, 胡 军, 李茂鹏, 崔惠桢, 占朝辉

(华东交通大学机电与车辆工程学院, 南昌 330013)

摘 要: 目前川贝母粉掺假现象层出不穷, 严重影响了中药材市场的健康发展, 因此对川贝母真伪进行检测意义重大。该研究以纯品川贝母粉以及 5 种含不同掺假物的川贝母粉样品为研究对象, 探究太赫兹时域光谱技术在检测川贝母品质方面应用的可行性。利用偏最小二乘判别 (Partial Least Squares Discriminant Analysis, PLS-DA) 对纯品川贝母粉以及掺假川贝母粉建立原始光谱的二分类模型。为了同时对多种含不同掺假物的川贝母样品进行鉴别, 先对原始光谱采用多种单一预处理方法以及多种复合预处理方法进行处理, 再利用主成分分析 (Principal Component Analysis, PCA) 对数据进行降维, 最后建立支持向量机 (Support Vector Machine, SVM) 多分类模型。建立 SVM 多分类模型时, 采用网格搜索 (Grid Search) 与粒子群 (Particle Swarm Optimization, PSO) 算法两种参数优化方式, 对 SVM 的惩罚参数 (c) 与核参数 (g) 进行优化。结果显示: 6 个二分类模型的鉴别正确率均为 100%, 表明纯品川贝母粉与掺假样品的太赫兹时域光谱存在差异, 归一化-多元散射校正-PSO-SVM 多分类模型效果较为理想, 预测正确为 95.67%, 均方根误差为 0.432。该研究可为检测分析川贝母品质提供理论经验借鉴。

关键词: 光谱; 模型; 支持向量机; 网格搜索; 粒子群优化; 太赫兹时域光谱

doi: 10.11975/j.issn.1002-6819.2021.15.036

中图分类号: O433.4; O657.39; R282.5

文献标志码: A

文章编号: 1002-6819(2021)-15-0308-07

徐振, 刘燕德, 胡军, 等. 基于太赫兹时域光谱技术的掺假川贝母检测[J]. 农业工程学报, 2021, 37(15): 308-314.

doi: 10.11975/j.issn.1002-6819.2021.15.036 <http://www.tcsae.org>

Xu Zhen, Liu Yande, Hu Jun, et al. Detection of adulterated fritillariae using terahertz time domain spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(15): 308-314. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.15.036 <http://www.tcsae.org>

0 引 言

川贝母是百合科贝母属植物的鳞茎, 既是食物也是药物, 在中医药领域因其具有清热润肺、化痰止咳等药用功效, 被关注与应用。但资源稀缺、供不应求、掺假伪冒等问题严重影响其市场价值^[1-2]。川贝母常被磨粉使用, 川贝母粉掺假现象屡禁不止, 非专业人员难以对其进行准确的鉴别, 传统的“一看二闻三尝”经验鉴别法识别川贝母粉末是否掺假难度较高, 也需要丰富的实践经验。目前理化分析方法鉴别川贝母虽被广泛使用, 但此类方法样品处理流程繁杂、设备昂贵, 需要专业的技术人员^[3-5]。近红外光谱^[6]、拉曼光谱^[7-8]、紫外光谱^[9]、荧光光谱^[10]等多种光谱技术也被用于中药产地与品类的鉴别, 但是也存在吸收峰重叠严重、检测限高、响应值不稳定、对于含复杂成分的检测目标的整体分布状况较难区分等问题^[11-12]。针对当前的中药材掺假的市场现状, 急需探索出一种新的检测掺伪中药材的新手段^[13-14]。

太赫兹波频率处于 0.1~10 THz 之间, 波长介于毫米波与红外线之间, 许多生物大分子及中草药活性分子的振动及转动能级均位于此频段范围内, 样品内部成分的微小差异均可引起太赫兹图谱的变化^[15-17]。故可由太赫兹图谱对物质成分含量进行表征, 因此在食品、医药、生物化学等检测领域的应用较为广泛。刘晓庆等^[18]对比 4 种青霉素类药物在 0.2~1.4 THz 波段的太赫兹的吸收峰, 通过其质量和强度的对应关系达到检测青霉素类药品质量的目的。刘陵玉等^[19]用太赫兹时域光谱技术获取含有黄芩苷的混合物在 0.3~1.5 THz 范围内的光谱, 利用二维相关光谱分析结合支持向量机 (Support Vector Machine, SVM) 和偏最小二乘 (Partial Least Squares, PLS) 法建立两种定量检测模型, 可以快速、准确地测定混合物中黄芩苷的含量。欧阳爱国等^[20]根据玉米粉中苯甲酸在 0.5~3.0 THz 的太赫兹光谱数据, 建立 PLS、最小二乘支持向量机 (Least Squares-Support Vector Machine, LS-SVM) 和多元线性回归 (Multiple Linear Regression, MLR) 定量分析模型, 结果发现太赫兹时域光谱技术结合 LS-SVM 建立的模型在定量检测样品中苯甲酸含量时表现出优良性能。管爱红等^[21]根据红薯淀粉和明矾以及混合物的太赫兹吸收系数谱和折射率谱, 发现样品中随着明矾含量增加其吸收峰的幅度和折射率均呈下降趋势, 可由此利用太赫兹时域光谱技术对淀粉中明矾进行检测。

近些年太赫兹光谱技术在药食同源的物质检测鉴别

收稿日期: 2021-04-27 修订日期: 2021-07-16

基金项目: “十二五”国家 863 计划 (SS2012AA101906); 国家自然科学基金 (31760344); 南方山地果园智能化管理技术与装备协同创新中心项目 (赣教高字[2014]60 号)

作者简介: 徐振, 研究方向为太赫兹光谱无损检测。Email: xz2910845707@163.com
*通信作者: 刘燕德, 博士, 教授, 研究方向为光电测控技术与仪器、现代无损检测新技术及其应用。Email: jxliuyd@163.com

领域的应用效果也取得一定的突破,包括姜黄^[22]、金银花^[17]、冬虫夏草^[23]、陈皮^[24]、人参^[25]等传统中药材,表明太赫兹光谱检测技术在未来具有广阔的应用前景^[26]。但缺乏模型参数优化方法,存在鉴别准确率低等缺陷。本研究将太赫兹时域光谱技术与支持向量机相结合,并对 SVM 的参数优化方式进行探索,对含有多种掺杂物的川贝母粉进行鉴别,试图提出一种高效无损的川贝母粉掺假的定性分析方法,以期快速检测川贝母品质提供参考。

1 材料与方法

1.1 仪器与参数

本试验采用日本 Advantest 公司的 TAS7500SU 系统进行光谱采集,选择透射模式。试验所用设备的原理如图 1 所示,飞秒激光经过分束镜分为较强的泵浦光和较弱的探测光,泵浦光通过光导天线激发太赫兹脉冲,经过待测样品后携带相关信息与经过延迟系统的探测光会和,共同触发探测器,进而获得样品的时域信号。系统使用 2 个超短脉冲激光器产生太赫兹波以及探测太赫兹波。飞秒激光脉冲输出功率最大为 50 mW,中心波长 1 550 nm,重复频率 50 MHz。试验系统的扫描速度设置为 8 ms/次,扫描次数设置为 8 096 次/点,试验的环境温度维持在 20℃左右,相对湿度 25%以下。开机后利用空压机和空气干燥装置通入干燥空气,预热 30 min 后,使样品仓的空气湿度维持在 5%以下,开始采集样品的太赫兹时域光谱。采集到时域参考信号的幅值 $E_0(t)$ 以及样品时域信号的幅值 $E_{\text{trans}}(t)$ 后,利用傅里叶变换转换为频域谱的幅值 $E_0(\omega)$ 、 $E_{\text{trans}}(\omega)$,根据 Dorney 等^[27]提出的光学参数模型,得到吸收系数 $\alpha(\omega)$ 、吸光度 A 、折射率 $n(\omega)$ 等主要光学参数,其计算公式如式(1)~式(3):

$$\alpha(\omega) = \frac{2}{d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n(\omega)+1]} \right\} \quad (1)$$

$$A = -\frac{1}{d} \ln \left[\frac{E_{\text{trans}}(\omega)}{E_0(\omega)} \right]^2 \quad (2)$$

$$n(\omega) = \frac{\varphi(\omega)c}{\omega d} + 1 \quad (3)$$

式中 d 为样品厚度, mm; ω 为太赫兹波的频率, Hz; $\rho(\omega)$ 为样品信号与参考信号的振幅比; $\varphi(\omega)$ 为两者的相位差, rad; c 表示光速, m/s。

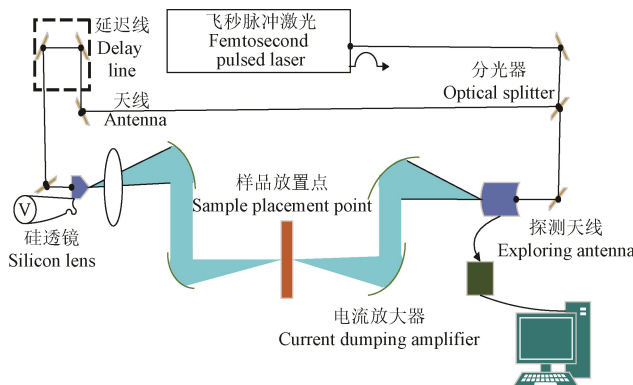


图 1 太赫兹设备原理图

Fig.1 Terahertz device schematic diagram

1.2 光谱采集

试验所用川贝母购买于康隆大药房,经检验为正品川贝母。试验过程中为保证样品的均匀性以及测量过程中的稳定性,样品制备采用粉碎压片的方法。为保证与光谱采集环境的一致性,制样环境温度设置为在 20℃,相对湿度维持在 25%左右。将纯品川贝母与 5 种掺假物(大米粉、葛粉、红薯粉、平贝母粉、小麦粉)分别烘干粉碎后,用 200 目药筛进行筛取,并按照掺假物含量 30% 的比例配置试验样品,为使样品更易压片成型且不影响太赫兹波透过样品的强度,向混合样品中加入等量聚乙烯并震荡混匀。压片之前在将样品置于恒温干燥箱中保存,以减少样品吸收的水分对太赫兹波的影响。称取每种样品 (0.150 ± 0.003) g,置于压片机模具中,压制成片,直径为 13 mm,厚度为 (1.0 ± 0.1) mm。6 种掺假的纯品各压制 10 个片,每个片采集 4 个点,共采集 240 条光谱;纯品川贝母及掺假川贝母各压制 50 个片,每个片采集 4 个点,共采集到 1 200 条纯品的光谱;本研究所有样品共采集到光谱 1 440 条。

1.3 数据处理流程

采集样品的光谱数据后,为初步判定纯品川贝母粉与掺杂样品的区别,对原始光谱建立 PLS-DA (Discriminate Analysis) 二分类判别模型。分别采用 S.G 平滑 (Savitzky-Golay Smoothing)、归一化 (Normalize) 多元散射校正 (Multiple Scatter Correction, MSC) 以及上述方法的两两结合对原始光谱进行预处理,采用主成分分析提取数据的主要变量,降低数据维度,简化计算量。采用 Kennard-Stone(K-S)方法按 1:3 的比例将光谱数据分为预测集和建模集,其中预测集 360 个光谱样本,建模集 1 080 个光谱样本^[28]。对采用多种方法预处理后的数据建立 SVM 分类模型,其中支持向量机优化方式采取网格搜索优化 (Grid Search) 以及粒子群优化 (Particle Swarm Optimization, PSO) 两种方式,并计算两种参数优化方式下的含各类掺假物的分类正确率,进行对比得出较佳的参数优化方式以及多分类方法。具体流程如图 2。

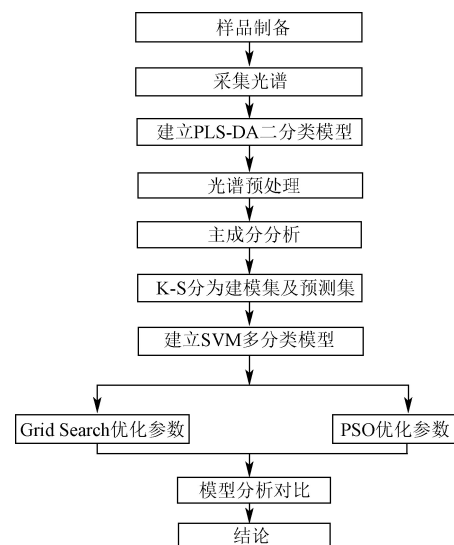


图 2 试验流程图

Fig.2 Experimental flow chart

1.4 数据处理方法

支持向量机 (Support Vector Machine, SVM) 可有效克服神经网络 (Back Propagation Neural Network, BPNN) 分类收敛难、解不稳定、推广性差等缺陷, 拥有许多传统模式识别算法不具备的优势^[29]。其最大的优点就是可以提高预测能力, 降低分类错误率。但支持向量机的运算结果很大程度上受限于惩罚参数 (c) 和核参数 (g) 的选择, 对参数随机指定难以达到最优的效果, 本研究将对比网格搜索寻优以及粒子群寻优两种参数优化方式。

网格搜索 (Grid Search) 作为一种参数寻优方法, 需要将被搜索的参数区域划分为网格, 其所有的交叉点即为参数组合 (c, g)^[30]。利用 k -fold 去测试每一组 (c, g) 对应的分类准确率, 以得到准确率最高的 (c, g) 组合作为建立模型的参数^[29,31]。

粒子群寻优 (Particle Swarm Optimization, PSO) 起始于随机解, 反复迭代寻优, 并由适应度评价解的品质^[32-33]。操作比遗传算法更简单, 跟随当前搜索到的最优值来寻找全局最优, 无需“交叉”与“变异”等的操作。粒子群优化算法有着实现容易、精度高、收敛快等优点, 在解决实际问题中具有一定的优势^[34]。

采用建模集与预测集的决定系数 R_c^2 、 R_p^2 分别评价建模集和预测集的相关程度, 越接近 1, 相关性越强。采用采用建模集均方根误差 (Root Mean Square Error of Calibration, RMSEC) 和预测集均方根误差 (Root Mean Square Error of Prediction, RMSEP) 来评价模型的预测能力, 越接近 0 预测能力越强。

2 结果与分析

2.1 原始光谱分析

图 3a 为采集到的样品的太赫兹时域光谱, 可见不同的掺假样品的时域光谱在相位与强度上均存在明显不同, 一定程度上表明含不同掺假物样品内部的成分与分子构成均与纯品川贝母有较大的区别。图 3b 为采集到的纯品川贝母以及 5 种含掺假物的样品的太赫兹时域吸收系数谱线图, 吸收系数整体上呈现上升趋势, 但是谱线存在交叉情况, 区分样品掺假的种类难度较大。图 3c 为试验样品的折射率光谱图, 6 种样品在 0.5~2.5 THz 无特别明显变化, 但总体呈现下降趋势。图 3d 为试验样品的透射比光谱图, 与吸收系数变化趋势不同, 透射比随着频率的增大逐渐减小, 但是含不同掺假物的样品光谱之间存在较为明显的交叉现象。

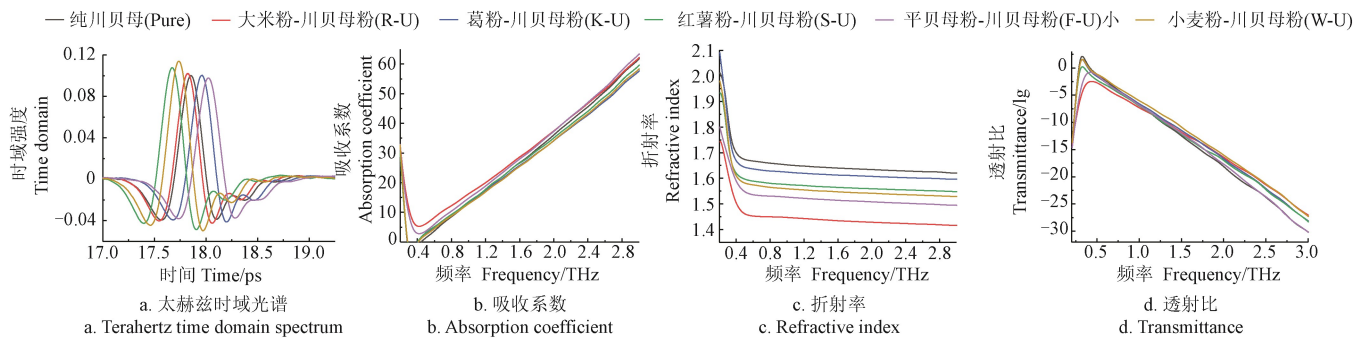


图 3 样品的太赫兹光谱图
Fig.3 Terahertz spectrum of samples

由于川贝母成分较为复杂, 通过样品的太赫兹光谱对川贝母是否掺假以及掺假的物质进行直接鉴别, 难度较大。为准确的鉴别川贝母粉是否掺假, 需要进一步利用化学计量学方法对光谱进行预处理以及建模分析。由于噪声及其他无关信息的影响, 主要对 0.5~2.5 THz 波段范围内的光谱进行分析。

2.2 掺假与未掺假二分类

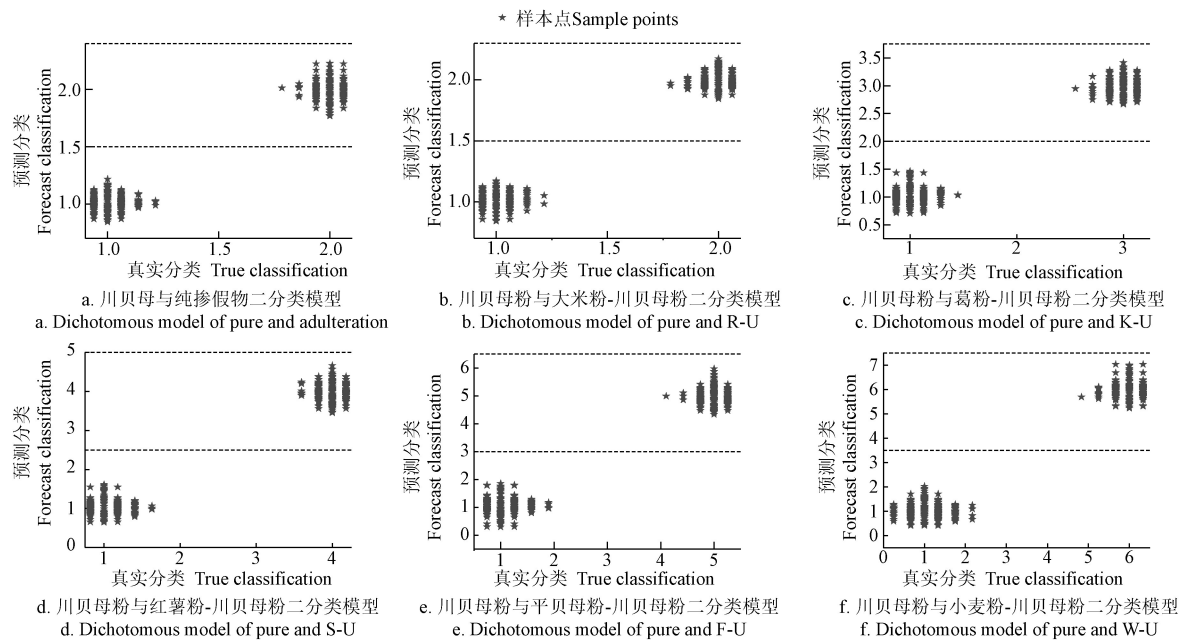
为对掺假川贝母进行准确分析, 在同时对多种掺假

川贝母进行多分类之前, 截取 0.5~2.5 THz 光谱的原始数据, 首先对纯品川贝母样品与其他 5 种掺假川贝母利用 PLS-DA 进行初步的二分类。该过程分为两步进行: 1) 鉴别区分掺假与未掺假川贝母; 2) 分别鉴别纯品川贝母粉与大米粉-川贝母粉 (纯品川贝母粉中掺杂有大米粉, 其他类似)、葛粉-川贝母粉、红薯粉-川贝母粉、平贝母粉-川贝母粉、小麦粉-川贝母粉。具体鉴别结果如表 1, 分类正确率为样本被正确分类的个数与该类样本总数中的比值。

表 1 二分类结果
Table 1 Results of dichotomy

样品类型 Sample types	R_c^2	建模集均方根误差 RMSEC	R_p^2	预测集均方根误差 RMSEP	分类正确率 Classification accuracy/%
纯川贝母粉与纯掺假样品 Pure Unibract fritillary bulb powder and adulterated sample (Pure)	0.973 5	0.081 2	0.973 0	0.082 1	100
川贝母与大米粉-川贝母粉 Rice flour-Unibract fritillary bulb powder(R-U)	0.981 8	0.067 4	0.981 5	0.068 0	100
川贝母与葛粉-川贝母粉 Kudzuvine root-Unibract fritillary bulb powder(K-U)	0.977 0	0.151 6	0.976 5	0.153 2	100
川贝母与红薯粉-川贝母粉 Sweet potato-Unibract fritillary bulb powder(S-U)	0.980 6	0.208 7	0.98 0	0.210 9	100
川贝母与平贝母粉-川贝母粉 Fritillaria ussuriensis-Unibract fritillary bulb powder(F-U)	0.977 2	0.301 7	0.976 7	0.304 9	100
川贝母与小麦粉-川贝母粉 Wheat flour-Unibract fritillary bulb powder(W-U)	0.981 4	0.340 7	0.981 0	0.344 2	100

川贝母粉掺假二分类模型如图 4。其中 PLS-DA 模型能够完全区分纯川贝母粉与纯品的掺假样品（纯大米粉、纯葛粉、纯红薯粉、纯平贝母粉、纯小麦粉），鉴别正确率能够达到 100%。同样，建立的川贝母粉与含 30%掺假物的样品的二分类模型效果也很理想。



注：1 表示纯川贝母；2 表示大米粉-川贝母粉；3 表示葛粉-川贝母粉；4 表示红薯粉-川贝母粉；5 表示平贝母粉-川贝母粉；6 表示小麦粉-川贝母粉，下同。
Note: 1 represents pure unibract fritillary bulb powder; 2 represents R-U; 3 represents K-U; 4 represents S-U; 5 represents F-U; 6 represents W-U, the same below.

图 4 川贝母粉掺假二分类模型结果

Fig.4 Results of dichotomous model of adulteration of unibract fritillary bulb powder

2.3 掺假川贝母粉的多分类研究

在进行二分类研究时，只能进行掺假贝母与未掺假贝母粉的鉴别，当多种掺假川贝母混杂在一起时，PLS-DA 模型难以进行准确的区分。为对掺假川贝母样品进行深入研究，建立 SVM 多分类模型。对原始光谱利用多种方法

进行预处理，利用主成分分析降低数据维度。建立 SVM 分类模型时，选择径向基函数作为核函数，以主成分分析后的太赫兹光谱作为输入特征，利用网格搜索（Grid Search）与粒子群优化（PSO）对 SVM 参数进行优化。

表 2 对比不同预处理情况下的支持向量机多分类模型。

表 2 川贝母粉掺假 SVM 多分类结果
Table 2 Multiple classification results of SVM adulteration of unibract fritillary bulb powder

优化方法 Parameter optimization method	预处理方法 Pretreatment method	主成分数 PCs	RMSE	分类正确率 Classification accuracy/%						平均正确率 Average accuracy
				纯川贝母 Pure	大米粉-川贝母 R-U	葛粉-川贝母 K-U	红薯粉-川贝母 S-U	平贝母-川贝母 F-U	小麦粉-川贝母 W-U	
网格搜索 Grid Search	无 None	4	0.583 1	100	100	84	82	86	84	89.33
	S.G 平滑	3	0.632 5	100	100	78	80	86	84	88.00
	Savitzky-Golay Smoothing	6	0.675 8	100	100	86	98	78	88	91.67
	归一化 Normalize	7	0.591 6	100	100	86	94	78	94	92.00
	多元散射校正 MSC	10	0.562 7	100	100	86	98	86	96	94.33
	S.G 平滑-归一化	10	0.580 2	100	100	84	98	90	92	94.00
	S.G Smoothing-Normalize	7	0.580 2	100	100	86	98	88	92	94.00
	S.G 平滑-多元散射校正	7	0.580 2	100	100	86	98	88	92	94.00
粒子群优化 PSO	归一化-多元散射校正	4	0.544 7	100	100	82	78	90	84	89.33
	无 None	3	0.547 7	100	100	80	82	92	84	89.67
	S.G 平滑	6	0.707 4	100	100	84	98	74	88	90.67
	Savitzky-Golay Smoothing	7	0.565 7	100	100	84	94	86	94	93.00
	归一化 Normalize	10	0.500 0	100	100	92	98	90	94	95.67
	S.G 平滑-归一化	10	0.506 6	100	100	84	98	90	94	94.33
	S.G Smoothing-Normalize	7	0.432 0	100	100	92	98	88	96	95.67
	S.G 平滑-多元散射校正	7	0.432 0	100	100	92	98	88	96	95.67

采用网格搜索进行参数优化时, 经过 S-G Smoothing 预处理后的多分类模型正确最低为 88.00%, 经过 S-G 平滑-归一化预处理后的多分类模型正确率较高为 94.33%, 且得到最小的均方根误差为 0.562 7。建立的 7 个多分类模型中纯川贝母与大米粉-川贝母粉的正确率均为 100%, 葛粉-川贝母粉鉴别正确率最大值为 86%, 效果不够理想, 红薯粉-川贝母粉鉴别正确率最大值为 98%, 平贝母-川贝母粉鉴别正确率最大值为 90%, 小麦粉-川贝母粉鉴别正确率最高值为 96%。采用 PSO 对参数进行优化时, 无任何预处理时正确率最低为 89.33%, 其中纯川贝母与大米粉-川贝母粉的正确率均为 100%; S-G 平滑-归一化后葛粉-川贝母粉鉴别正确率最大值为 92%, 红薯粉-川贝母粉鉴别正确率最大值为 98%, 平贝母-川贝母粉鉴别正确率最大值为 90%, 小麦粉-川贝母粉鉴别正确率最大值为 94%。其中经过 S-G S-G 平滑-归一化预处理与归一化-MS-C 预处理后的模型整体正确最高, 为 95.67%, 但经过归一化-MS-C 预处理后得到的均方根误差最小, 为 0.432 0, 故建立的归一化-MS-C-PSO-SVM 多分类模型效果最优。通过表 2 可分析出, 经过粒子群优化参数优化后的 SVM 模型分类正确率整体高于经网格搜索优化后的模型鉴别正确率。

图 5 为经主成分分析后的建立的归一化-MS-C-PSO-SVM 多分类模型结果, 相比较建立的其他多分类模型, 该模型的分类效果最好, 可得到最高的分类正确率以及最低的均方根误差。此时, 惩罚参数 (c) 为 2.550 8, 核参数 (g) 为 66.814 3。纯川贝母粉与大米粉-川贝母粉无被错误分类情况; 葛粉-川贝母粉共 4 个样品被错误分类, 1 个被错误鉴别为红薯粉-川贝母粉, 1 个被错误鉴别为平贝母-川贝母粉, 2 个被错误鉴别为小麦粉-川贝母粉; 红薯粉-川贝母粉 1 个样品被错误鉴别为大米粉-川贝母粉; 平贝母-川贝母粉被错误鉴别为大米粉-川贝母粉的样品个数为 3 个; 小麦粉-川贝母粉有 2 个样品被错误鉴别为平贝母-川贝母粉。本研究所建立的模型虽然能够以较高的准确率鉴别不同的掺假样品, 由于试验样品掺假物种均含有大量淀粉以及其他相似成分, 虽各成分含量不同, 但也会导致其太赫兹光谱信息存在微弱的相似之处, 致使少数样品被错误分类, 平均分类正确率为 95.67%。

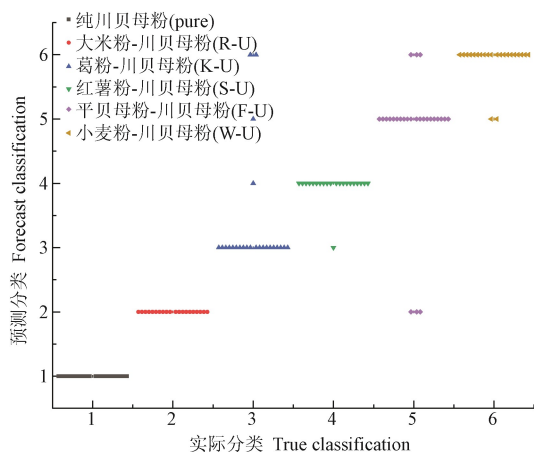


图 5 归一化-MS-C-PSO-SVM 多分类模型结果

Fig.5 Results of normalize-MS-C-PSO-SVM multi-classification model

3 结 论

本研究以纯品川贝母以及川贝母粉中 5 种常见的掺假物 (大米粉、葛粉、红薯粉、平贝母粉、小麦粉) 作为研究对象。使用建立的 PLS-DA 二分类模型鉴别纯品川贝母与各种掺假样品, 分类正确率均为 100%, 表明纯品川贝母与含掺假物的样品存在明显区别。为同时对含不同掺假物的川贝母样品进行识别, 对原始光谱采用单一预处理方法以及多种预处理方法相结合进行预处理, 并采用主成分分析降低数据维度, 建立支持向量机模型。采用网格搜索 (Grid Search) 与粒子群优化算法 (Particle Swarm Optimization, PSO) 优化支持向量机 (Support Vector Machine, SVM) 模型参数。结果表明建立的归一化多元散射校正-PSO-SVM 多分类模型效果较优, 识别 5 种掺假川贝母样品平均预测正确为 95.67%, 均方根误差为 0.432 0。本研究可为鉴别川贝母品质提供一种简洁快速无损的检测方法, 也可为后续定量检测川贝母中的掺假物含量提供基础。

[参 考 文 献]

- [1] 刘芳. 中药粉末饮片质量评价示范性研究[D]. 成都: 成都中医药大学, 2017.
Liu Fang. Study on Quality Evaluation of Herbal Powder[D]. Chengdu: Chengdu University of TCM, 2017. (in Chinese with English abstract)
- [2] 熊浩荣, 马朝旭, 国慧, 等. 川贝母野生基原植物资源分布和保育研究进展[J]. 中草药, 2020, 51(9): 2573-2579.
Xiong Haorong, Ma Zhaoxu, Guo Hui, et al. Research progress on wild source plant resources distribution and conservation of Fritillariae Cirrhosae Bulbus[J]. Chinese Traditional and Herbal Drugs, 2020, 51(9): 2573-2579. (in Chinese with English abstract)
- [3] 耿昭, 李小红, 苟琰, 等. QuEChERS 法结合气相色谱-串联质谱法测定贝母类中药中 53 种农药残留[J]. 中草药, 2020, 51(20): 5337-5347.
Geng Zhao, Li Xiaohong, Gou Yan, et al. Determination of 53 pesticide residues in different category of Fritillaria by QuEChERS and GC-MS/MS[J]. Chinese Traditional and Herbal Drugs, 2020, 51(20): 5337-5347. (in Chinese with English abstract)
- [4] 刘瑞新, 郝小佳, 张慧杰, 等. 基于电子眼技术的中药川贝母真伪及规格的快速辨识研究[J]. 中国中药杂志, 2020, 45(14): 3441-3451.
Liu Ruixin, Hao Xiaojia, Zhang Huijie, et al. A rapid identification of authenticity and specifications of Chinese medicine Fritillariae Cirrhosae Bulbus based on E-eye technology[J]. China Journal of Chinese Materia Medica, 2020, 45(14): 3441-3451. (in Chinese with English abstract)
- [5] 车朋, 刘久石, 齐耀东, 等. UPLC-ELSD 同时测定贝母类药材中 6 种生物碱的含量[J]. 中国中药杂志, 2020, 45(6): 1393-1398.
Che Peng, Liu Jiushi, Qi Yaodong, et al. Simultaneous determination of six major isosteroidal alkaloids in Beimu by UPLC-ELSD[J]. China Journal of Chinese Materia Medica, 2020, 45(6): 1393-1398. (in Chinese with English abstract)
- [6] 陈述. 基于近红外光谱技术的不同产地中药有效成分含量检测方法[J]. 激光杂志, 2020, 41(12): 22-26.
Chen Shu. Detection method of effective components of traditional Chinese medicine from different producing areas based on near infrared spectroscopy[J]. Laser Journal, 2020, 41(12): 22-26. (in Chinese with English abstract)

- [7] 王宪双, 郭帅, 徐向君, 等. 基于激光诱导击穿光谱和拉曼光谱对四唑类化合物的快速识别和分类试验研究[J]. 中国光学, 2019, 12(4): 889-896.
Wang Xianshuang, Guo Shuai, Xu Xiangjun, et al. Fast recognition and classification of tetrazole compounds based on laser-induced breakdown spectroscopy and raman spectroscopy[J]. Chinese Journal of Optics, 2019, 12(4): 889-896. (in Chinese with English abstract)
- [8] 王文娜, 陈地灵, 朱梅芳, 等. 激光拉曼光谱法无损分析鉴别川贝母[J]. 光谱学与光谱分析, 2013, 33(8): 2109-2111.
Wang Wenna, Chen Diling, Zhu Meifang, et al. The analysis and identification of fritillaria cirrhosa by raman spectra[J]. Spectroscopy and Spectral Analysis, 2013, 33(8): 2109-2111. (in Chinese with English abstract)
- [9] 赵丽茹, 李金花, 路晓玉, 等. 两头尖药材的紫外光谱鉴别研究[J]. 中西医结合心血管病电子杂志, 2019, 7(28): 1-2, 8.
Zhao Liru, Li Jinhua, Lu Xiaoyu, et al. Study on the UV spectrum identification of Anemones raddeanae Rhizoma[J]. Cardiovascular Disease Journal of Integrated Traditional Chinese and Western Medicine (Electronic), 2019, 7(28): 1-2, 8. (in Chinese with English abstract)
- [10] 樊凤杰, 轩凤来, 白洋, 等. 基于三维荧光光谱特征的中草药性模式识别研究[J]. 光谱学与光谱分析, 2020, 40(6): 1763-1768.
Fan Fengjie, Xuan Fenglai, Bai Yang, et al. Pattern recognition of traditional Chinese medicine property based on three-dimensional fluorescence spectrum characteristics[J]. Spectroscopy and Spectral Analysis, 2020, 40(6): 1763-1768. (in Chinese with English abstract)
- [11] 安思宇, 张磊, 岳洪水, 等. 基于近红外光谱的中药质量一致性控制研究进展[J]. 中南药学, 2019, 17(9): 1439-1445.
An Siyu, Zhang Lei, Yue Hongshui, et al. Research progress in quality consistency control for traditional Chinese medicine injections based on near infrared spectroscopy[J]. Central South Pharmacy, 2019, 17(9): 1439-1445. (in Chinese with English abstract)
- [12] 李冰. 紫外光谱在中药鉴别和含量测定应用中的研究[J]. 生物化工, 2018, 4(6): 131-133.
Li Bing. Ultraviolet spectrum in the application of traditional Chinese medicine identification and content determination[J]. Biological Chemical Engineering, 2018, 4(6): 131-133. (in Chinese with English abstract)
- [13] 仰铁锤, 谢慧敏, 谢慧淦, 等. 聚合酶链式反应-限制性内切酶多态法检查川贝母的掺伪情况[J]. 华西药理学杂志, 2020, 35(3): 265-269.
Yang Tiechui, Xie Huimin, Xie Huigan, et al. Research on the adulterants identification in Fritillariae cirrhosae Bulbus with PCR-RFLP method[J]. West China Journal of Pharmaceutical Sciences, 2020, 35(3): 265-269. (in Chinese with English abstract)
- [14] 杨健, 李靖, 薛维娜, 等. 实时荧光定量 PCR 法鉴别川贝母掺伪[J]. 中成药, 2020, 42(5): 1262-1268.
Yang Jian, Li Jing, Xue Weina, et al. Identification adulteration in Fritillariae Cirrhosae by real-time fluorescence quantitative PCR method[J]. Chinese Traditional Patent Medicine, 2020, 42(5): 1262-1268. (in Chinese with English abstract)
- [15] Zhang H, Li Z, Chen T, et al. Discrimination of traditional herbal medicines based on terahertz spectroscopy[J]. Optik, 2017, 138(12): 95-102.
- [16] Zhang H, Li Z. Terahertz spectroscopy applied to quantitative determination of harmful additives in medicinal herbs[J]. Optik, 2018, 156(6): 834-840.
- [17] Zhang H, Li Z, Chen T, et al. Detection of poisonous herbs by Terahertz Time-Domain spectroscopy[J]. Journal of Applied Spectroscopy, 2018, 85(1): 197-202.
- [18] 刘晓庆, 姚嘉丽, 黄凡, 等. 基于太赫兹时域光谱的青霉素类药物检测研究[J]. 光学学报, 2020, 40(6): 198-204.
Liu Xiaoqing, Yao Jiali, Huang Fan, et al. Study on detection of penicillin drugs based on Terahertz Time-Domain spectroscopy[J]. Acta Optica Sinica, 2020, 40(6): 198-204. (in Chinese with English abstract)
- [19] 刘陵玉, 常天英, 李珂, 等. 基于太赫兹辐射的黄芩苷光谱分析及定量检测[J]. 中国激光, 2020, 47(3): 313-319.
Liu Lingyu, Chang Tianying, Li Ke, et al. Spectral analysis and quantitative detection of baicalin based on Terahertz radiation[J]. Chinese Journal of Lasers, 2020, 47(3): 313-319. (in Chinese with English abstract)
- [20] 欧阳爱国, 蔡会周, 李斌, 等. 玉米粉中苯甲酸的太赫兹光谱定量检测研究[J]. 激光技术, 2020, 44(4): 478-484.
Ouyang Aiguo, Cai Huizhou, Li Bin, et al. Quantitative detection of benzoic acid in corn flour by Terahertz spectroscopy[J]. Laser Technology, 2020, 44(4): 478-484. (in Chinese with English abstract)
- [21] 管爱红, 李智, 葛宏义. 红薯淀粉中添加剂明矾的定性和定量太赫兹时域光谱技术检测[J]. 光谱学与光谱分析, 2018, 38(1): 267-270.
Guang Aihong, Li Zhi, Ge Hongyi. The Qualitative and quantitative detection of potassium alum in sweet potato starch based on Terahertz Time-Domain spectroscopy[J]. Spectroscopy and Spectral Analysis, 2018, 38(1): 267-270. (in Chinese with English abstract)
- [22] Li H, Du S Q, Xie L, et al. Identifying radix curcumae by using terahertz spectroscopy[J]. Optik, 2012, 123(13): 1129-1132.
- [23] 李辰, 魏丞昊, 王志琪, 等. 太赫兹波谱在冬虫夏草检测中的应用[J]. 深圳大学学报: 理工版, 2019, 36(2): 213-220.
Li Chen, Wei Chenghao, Wang Zhiqi, et al. Authenticity assessment of Cordyceps sinensis using terahertz spectroscopy[J]. Journal of Shenzhen University: Science & Engineering, 2019, 36(2): 213-220. (in Chinese with English abstract)
- [24] 杨少壮, 李灿, 李辰, 等. 不同贮存年限陈皮的太赫兹光谱和成像的差异分析[J]. 现代食品科技, 2019, 35(12): 258-266.
Yang Shaozhuang, Li Can, Li Chen, et al. The differences in the dried tangerine peels stored for different years revealed by terahertz spectroscopy and imaging[J]. Modern Food Science & Technology, 2019, 35(12): 258-266. (in Chinese with English abstract)
- [25] 寇天一. 基于太赫兹光谱的人参和西洋参鉴别[J]. 光学仪器, 2020, 42(5): 27-32.
Kou Tianyi. Terahertz spectroscopy for accurate identification of ginseng and panax quinquefolium[J]. Optical Instruments, 2020, 42(5): 27-32. (in Chinese with English abstract)
- [26] Zhang X, Lu S H, Liao Y, et al. Simultaneous determination of amino acid mixtures in cereal by using terahertz time domain spectroscopy and chemometrics[J]. Chemometrics and Intelligent Laboratory Systems, 2017, 164(2): 8-15.
- [27] Dorney T D, Baraniuk R G, Mittleman D M. Material parameter estimation with terahertz time-domain spectroscopy[J]. Journal of the Optical Society of America A Optics image science and vision, 2001, 18(7): 1562-1571.
- [28] Morais Camilo L M, Santos Marfran C D, Lima Kássio M G, et al. Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach[J]. Bioinformatics, 2019, 35(24): 5257-5263.
- [29] 李鑫星, 朱晨光, 白雪冰, 等. 基于可见光谱和支持向量机的黄瓜叶部病害识别方法研究[J]. 光谱学与光谱分析,

- 2019, 39(7): 2250-2256.
- Li Xingxing, Zhu Chenguang, Bai Xuebing, et al. Recognition method of cucumber leaves diseases based on visual spectrum and support vector machine[J]. Spectroscopy and Spectral Analysis, 2019,39(7): 2250-2256. (in Chinese with English abstract)
- [30] 胡晓华, 刘伟, 刘长虹, 等. 基于太赫兹光谱和支持向量机快速鉴别咖啡豆产地[J]. 农业工程学报, 2017, 33(9): 302-307.
- Hu Xiaohua, Liu Wei, Liu Changhong, et al. Rapid identification of producing area of coffee bean based on terahertz spectroscopy and support vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(9): 302-307. (in Chinese with English abstract)
- [31] 沙文, 李江涛, 鲁翠萍. 基于激光诱导击穿光谱技术寻优定量分析土壤中 Mn 元素[J]. 中国激光, 2020, 47(5): 505-513.
- Sha Wen, Li Jiangtao, Lu Cuiping. Quantitative analysis of mn in soil based on laser-induced breakdown spectroscopy optimization[J]. Chinese Journal of Lasers, 2020, 47(5): 505-513. (in Chinese with English abstract)
- [32] 陈啸, 王红英, 孔丹丹, 等. 基于粒子群参数优化和 BP 神经网络的颗粒饲料质量预测模型[J]. 农业工程学报, 2016, 32(14): 306-314.
- Chen Xiao, Wang Hongying, Kong Dandan, et al. Quality prediction model of pellet feed basing on BP network using PSO parameters optimization method[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(14): 306-314. (in Chinese with English abstract)
- [33] 彭威, 林强. 基于 PSO-SVM 算法的雷达点迹真伪鉴别方法研究[J]. 雷达科学与技术, 2020, 18(4): 429-432, 437.
- Peng Wei, Lin Qiang. An identification method of true and false plots based on PSO-SVM algorithm[J]. Radar Science and Technology, 2020, 18(4): 429-432, 437. (in Chinese with English abstract)
- [34] 王书涛, 张彩霞, 王志芳, 等. 最小二乘支持向量机在对羟基苯甲酸甲酯钠荧光检测中的应用[J]. 激光与光电子学进展, 2017, 54(7): 326-332.
- Wang Shutao, Zhang Caixia, Wang Zhifang, et al. Application of least squares support vector machine in fluorescence detection of sodium methylparaben[J]. Laser & Optoelectronics Progress, 2017, 54(7): 326-332. (in Chinese with English abstract)

Detection of adulterated fritillariae using terahertz time domain spectroscopy

Xu Zhen, Liu Yande^{*}, Hu Jun, Li Maopeng, Cui Huizhen, Zhan Chaohui

(School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: Unibract fritillary bulb, a traditional precious Chinese medicinal material, has the effects of clearing away heat, moisturizing the lungs, reducing phlegm, and relieving cough. However, the adulteration of Unibract fritillary bulbs has posed a serious threat to the medicinal effect and the healthy development of the market in recent years. Therefore, it is of great significance to accurately and rapidly detect the adulterated Unibract fritillary bulb powder. In this study, a systematic detection was conducted to distinguish the adulterated fritillariae using terahertz time-domain spectroscopy. Five samples of Fritillaria powder were used as the research objects, containing different adulterants (rice flour, Kudzu vine root powder, sweet potato powder, wheat flour, and Fritillaria Ussuriensis Maxim powder), pure Unibract fritillary bulb powder as the control group. Chemometric methods were also selected to detect the quality of Unibract fritillary bulb. The specific procedure was as follows. Firstly, adulterated samples were prepared with different types of Unibract fritillary bulbs in the same content. Then, the terahertz time-domain spectra were collected. Partial Least Squares Discriminant Analysis (PLS-DA) was also used in the range of 0.5-3.0 THz, according to the original and five adulterated Fritillaria powders. The original spectrum was used to remove the irrelevant variables and noise using the Savitzky-Golay smoothing (S-G Smoothing), Normalize, and Multiple Scatter Correction (MSC). A two-class model was established using the obtained spectral data. Thirdly, Principal component analysis (PCA) was used to reduce the dimensionality of preprocessed data, while simplifying the calculation of the model. Kennard-Stone (KS) was selected to divide the sample data into a 1:3 ratio, while the spectral data into prediction and modeling set. Finally, a Support Vector Machine (SVM) multi-classification model was established using Grid Search and Particle Swarm Optimization (PSO), where two parameters were optimized, namely, the penalty parameters (c) and the number of cores (g) of SVM. Correspondingly, the recognition accuracy rates of various samples were compared under the optimal spectral preprocessing and parameter optimization. The results showed that six binary classification models for the original spectra presented a correct identification rate of 100%, indicating a high accuracy for the pure Unibract fritillary bulb and adulterated Fritillaria. There were also great differences in the time domain spectra in the terahertz of samples. A multi-classification model was then established using Normalize combined with MSC preprocessing, further optimizing parameters using Particle Swarm Optimization (PSO). The overall accuracy of PSO optimization was higher than that of grid search optimization, where the highest accuracy rate was 100%. The lowest accuracy rate was 90%, and the average prediction accuracy was 95.67%, while the root mean square error was 0.432 when Unibract fritillary bulb powder was mixed with Fritillaria Ussuriensis Maxim powder. Consequently, Terahertz spectroscopy combined with a support vector machine can simultaneously detect a variety of Unibract fritillary bulb powder containing different adulterants. This finding can provide a theoretical experience for the detection of Unibract fritillary bulb adulteration in the field of medicine, thereby ensuring the excellent quality of Chinese medicinal materials in the trading market.

Keywords: spectroscopy; models; support vector machine; grid search; particle swarm optimization; terahertz time domain