

高光谱成像的三七粉质量等级无损鉴别

张付杰, 史磊, 李丽霞*, 赵浩然, 朱银龙

昆明理工大学现代农业工程学院, 云南 昆明 650500

摘要 三七粉是三七的主要消费和商品形式, 市场上存在以次充好、甚至是掺假的现象, 由于是粉状物料, 难以用肉眼判别, 为了实现对不同质量等级的三七粉进行无损鉴别。将30头、40头、60头和80头的三七主根研磨成粉, 制备样本。采用可见近红外高光谱成像系统(400.68~1 001.61 nm)采集4种不同头数三七粉, 共计384个样品的高光谱图像, 提取高光谱图像感兴趣区域(ROD)的平均光谱值作为样本原始光谱。将384个三七粉样本按2:1的比例划分训练集和测试集。采用卷积平滑(SG)、多元散射校正(MSC)和标准正态变量变换(SNV)3种预处理方法对三七粉样本光谱信息进行预处理并建立支持向量机(SVM)分类模型, 通过比较基于3种预处理方法的SVM模型测试集分类准确率, 确定SNV为最优预处理方法。采用迭代保留信息变量(IRIV)、变量组合集群分析(VCPA)和变量组合集群分析混合迭代保留信息变量(VCPA-IRIV)3种特征选择方法提取SNV预处理后光谱的特征波长并建立基于特征光谱和原始光谱的SVM分类模型, 通过比较基于3种特征选择方法得到的特征波长建立的SVM模型测试集分类准确率, 发现将VCPA与IRIV相结合的VCPA-IRIV为最优特征选择方法。VCPA-IRIV提取了18个特征波长代替全光谱数据参与建模, 该算法在降低模型复杂度的同时保持了模型分类精度。为了提高模型分类精度, 采用引力搜索算法(GSA)对SVM模型中惩罚因子 c 和核参数 g 进行寻优, 并与网格搜索(GS)的结果进行比较, 结果表明, VCPA-IRIV-GSA-SVM模型分类效果最好, 测试集分类准确率达到100%。可见, 利用可见近红外高光谱成像对三七粉进行质量等级无损鉴别是可行的, 为市场上三七粉的质量等级鉴别提供了参考。

关键词 可见近红外高光谱成像; 三七粉; 特征选择; 支持向量机; 引力搜索算法

中图分类号: TP391 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)07-2255-07

引言

三七为五加科植物三七 *Panax notoginseng* (Burk.) F. H. Chen 的干燥根和根茎, 含有许多营养成分, 如皂苷、黄酮、氨基酸、多糖和许多其他微量元素^[1]。三七粉是三七的主要消费和商品形式, 不同质量等级的三七粉用肉眼很难分辨, 外加市场上不同质量等级的三七粉价格差异较大, 因此对三七粉进行质量等级鉴别具有十分重要的意义。

现在常用的三七粉质量等级鉴别方法为人工检测。但人工检测需要检测人员有过硬的技术和丰富的经验, 不具备普适性。随着检测技术的进步, 高效液相色谱、近红外光谱等技术能够用于三七粉质量等级鉴定。Meng^[2]等利用高效液相色谱与化学模式识别相结合的方法对三七进行分类, 可以

清晰地区分“春七”和“冬七”。Li^[3]等利用高效液相色谱法结合PCA-MD成功鉴别了三七粉的真伪性。Yang^[4]等将近红外光谱和红外光谱数据融合并建立PSO-SVM模型对掺假物比例不同的三七粉进行分类, 分类正确率分别达到了96.65%和96.97%。Zhou^[5]等人利用多传感器红外光谱结合高层次多传感器信息融合策略的RF-Bo模型对不同产地的三七进行分类, 分类正确率达到了95.6%。但高效液相色谱检测试验成本高、周期长, 而且具有破坏性。近红外光谱技术不能提供目标图像上每个像素的光谱细节, 检测精度不够高。因此需要寻找一种更高准确率、更高效率的无损检测方法来实现三七粉质量等级的鉴别。

高光谱成像将光谱技术和成像技术相结合, 不仅可以反映样本的外部特征, 还可以反映样本内部生化信息^[6], 已经广泛应用于农产品检测领域。孙婷^[7]等利用高光谱成像将光

收稿日期: 2021-06-10, 修订日期: 2021-10-12

基金项目: 国家重点研发计划项目(2017YFC1702503), 云南省重大科技专项计划项目(202102AA310048), 云药之乡项目(202102AA310045)资助

作者简介: 张付杰, 1977年生, 昆明理工大学现代农业工程学院副教授 e-mail: 20030031@kust.edu.cn

* 通讯作者 e-mail: lilixia2012@kust.edu.cn

谱和图像信息结合并构建 SVM 模型对 11 类酿酒高粱进行分类, 准确率达到 91.8%; 孙俊^[8]等利用高光谱成像结合 GA-PNN 神经网络对江苏、安徽、山东三个品种的红豆进行鉴别, 识别准确率达到 97.5%; Wang^[9]利用高光谱图像对不同成熟的玉米种子进行分类, 从胚乳侧选择特征波长结合 PLS-DA, 准确率达到 100%; Weng^[10]等利用高光谱成像结合主成分分析网络对水稻品种进行分类, 准确率达到 98.57%; Jennifer Dumont^[11]利用高光谱图像结合 SVM 模型成功鉴别了正常的挪威云杉种子、被 *Megastigmussp* 侵染的挪威云杉种子和空壳的挪威云杉种子, 准确率达到 93.8%。但是目前利用高光谱图像技术实现三七粉质量等级鉴别的研究还鲜有报道。

三七作为中药材, 内部药用成分众多, 而不同成分在光谱中的吸收波段不同。相关研究表明, 多糖在可见光光谱范围内存在吸收波段, 皂苷和水分在近红外光谱范围内存在吸收波段。不同质量等级的三七粉内在成分含量比例不同^[12], 在可见光光谱和近红外光谱范围内具有不同的光谱特征, 因此本研究基于高光谱成像技术对三七粉进行质量等级鉴别研究。本研究以四种不同质量等级的三七粉作为研究对象, 基于不同预处理算法、特征选择算法对三七粉的高光谱数据进行处理, 并建立分类模型, 以实现三七粉质量等级的无损鉴别。

1 实验部分

1.1 三七粉样本准备

选择来自云南文山的 30 头、40 头、60 头、80 头的三七, 将不同头数的三七主根研磨成粉, 制备样本, 根据三七主根的头数把三七粉分为 4 个质量等级。每个样本称量 20 g, 总共制备了 384 个三七粉试验样本(每个质量等级 96 个)。将所有样本按 2:1 的比例划分训练集和测试集, 其中训练集

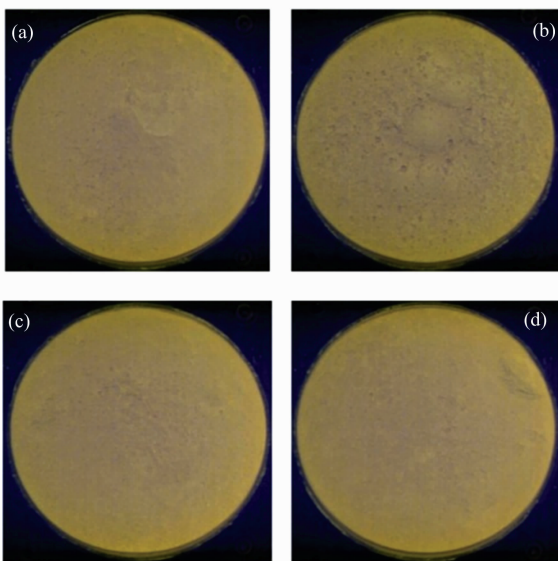


图 1 四种质量等级的三七粉

Fig. 1 Four quality grades of *panax notoginseng* powder

256 个样品, 测试集有 128 样品。4 个质量等级的三七粉如图 1 所示, 不同质量等级的三七粉存在一定的差异, 但难以用肉眼判别。

1.2 高光谱成像系统

高光谱成像系统是由高光谱图像摄影仪(VNIP-HIS-MOS)、卤素灯光源(UBer LED100 型, IT, USA)、分光模具(V10E 型, SPECIM, Finland)、光纤和电控平移台等组成。高光谱图像摄影仪是由 CCD 相机(Zyla4.2 型, Andor, UTKL)、光谱仪等组成, 光谱范围为 400.68~1 001.61 nm, 光谱分辨率为 2.8 nm, 图像分辨率为 1 024×478 像素。使用的高光谱图像采集系统的主要结构如图 2 所示。

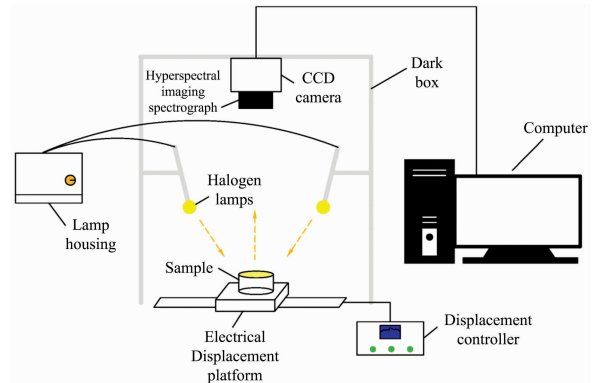


图 2 高光谱成像系统结构图

Fig. 2 Diagram of hyperspectral imaging system

1.3 高光谱图像采集和光谱数据提取

在试验前对高光谱成像系统进行 20 min 的预热和黑白板标定。黑白板标定可以减小暗电流噪声和光源强度分布不均匀对试验造成的影响, 白板的反射率 100%, 盖上 CCD 相机镜头获得反射率为 0% 的黑板环境。白板环境下, 设定高光谱图像相机的曝光时间为 7 ms; 黑板环境下, 设定高光谱图像相机的曝光时间为 17 ms; 设定平移台的速度为 1.99 mm·s⁻¹。依次采集四种不同质量等级的三七粉样本的高光谱图像。

采用“矩形区域法”在每个高光谱图像中手动选择 60×60 像素的正方形作为感兴趣区域(region of interest, ROI), 然后将 ROI 中像素的平均值作为每个样本的光谱值。

1.4 数据预处理

在客观环境下难以避免噪声的干扰, 高光谱成像器械难以避免发生基线漂移^[13]。为了减少它们对试验结果的影响, 需要对高光谱数据进行预处理。采用卷积平滑(savitzky-golay, SG)、多元散射校正(multiplication scatter correction, MSC)和标准正态变量变换(standard normalized variable, SNV)^[14]这 3 种方法分别对高光谱数据进行预处理, 并对比其效果, 选出最优的预处理方法。

1.5 特征选择

预处理后的光谱数据分布在高维空间, 但有些维度与建模无关。为了减少光谱数据中一些无用的维数, 采用特征选择的方法对光谱数据进行降维。采用迭代保留信息变量(iteratively retains informative variables, IRIV)^[15]、变量组合

集群分析 (variable combination population analysis, VC-PA)^[16] 和变量组合集群分析混合迭代保留信息变量 (variables combination population analysis and iterative retained information variable, VCPA-IRIV)^[17] 分别从全光谱数据中提取特征波长变量。

1.6 建模方法

1.6.1 支持向量机

支持向量机 (support vector machine, SVM) 是一种高维信息处理的重要工具。SVM 以其良好的泛化能力在光谱数据的分类中得到了广泛的应用。相关研究表明, 惩罚因子 c 和核参数 g 的选择对 SVM 的性能起着至关重要的作用。因此, 有必要对 SVM 的参数进行优化来提升分类效果^[18]。

1.6.2 引力搜索算法优化支持向量机

引力搜索算法 (gravitational search algorithm, GSA) 是一种种群优化算法。用 GSA 对 SVM 的参数 c 和 g 进行寻优, 具体流程如下^[19]:

- (1) 对参数进行初始化: 随机产生质点位置的数目 (群体规模) 和最大迭代次数;
- (2) 设置质点的移动范围 (参数 c 和 g 的搜索范围);
- (3) 设置样品训练集的交叉验证数, 并通过计算粒子的适应度值确定最优质点;

(4) 计算质点质量、质点在各维数上的加速度, 对质点进行位置更新;

(5) 重复步骤 (3) — (4), 当达到最大迭代次数时, 迭代停止, 获得最佳的 (c, g) ;

(6) 将参数的最优值代入 SVM 模型中进行预测。

1.7 数据处理软件

使用五铃光学公司高光谱成像系统 HSI Analyzer 软件进行图像校正和感兴趣区域提取, 使用 The Unscrambler X 10.4 软件进行预处理, 使用 Matlab 2016a 软件进行特征选择和数据建模。

2 结果与讨论

2.1 样本的光谱特征及最优预处理方法的选择

高光谱数据容易受噪声和仪器的干扰, 可能会影响后续建模的精度, 因此对高光谱数据进行预处理。原始光谱与 SG, MSC 和 SNV 预处理后的光谱如图 3 所示。观察图 3(a) 可以发现, 在采集开始时样本数据受噪声的影响较大。对比图 3(a) 和图 3(b) 可以发现, SG 预处理后的光谱曲线比原始光谱曲线更平滑。对比图 3(a) 和图 3(c) 可以发现, MSC 预处理使各高光谱曲线间差距变小, 表明 MSC 预处理对光谱采

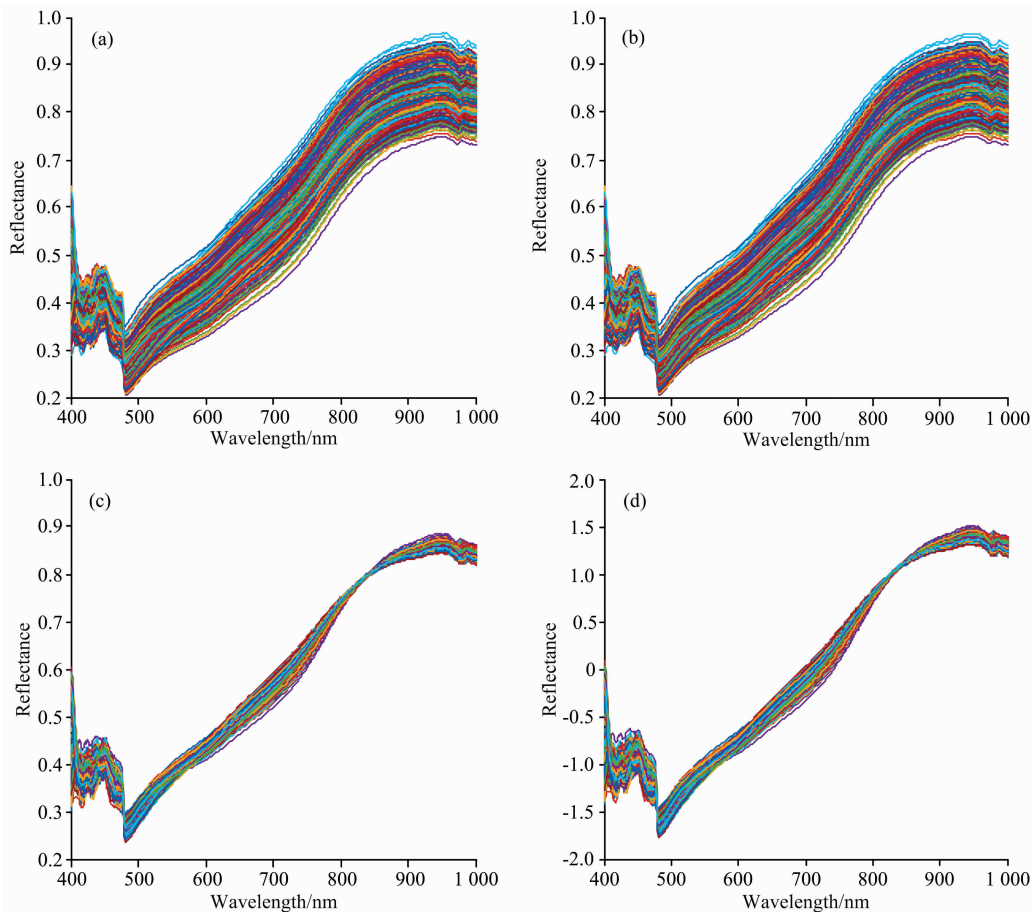


图 3 (a)原始光谱; (b)SG 预处理后光谱; (c)MSC 预处理后光谱; (d)SNV 预处理后光谱
 Fig. 3 (a) Original spectra; (b) Spectra after SG pretreatment; (c) Spectra after MSC pretreatment; (d) Spectra after SNV pretreatment

集过程中因散射现象等引起的光谱误差起到了校正作用。观察图 3(d)可以发现, SNV 预处理效果与 MSC 相似, 对光谱采集过程中因散射现象等引起的光谱误差起到了校正作用, 但 SNV 预处理的实质是对高光谱数据的标准正态化。

为了选择最优的预处理方法, 采用 SVM 对预处理后的高光谱数据进行建模, 核函数选择径向基核函数, 设置参数 c 和 g 为默认值, SVM 建模结果如表 1 所示。通过三种预处理方法对三七粉原始光谱的处理对比分析, 发现 SNV 预处理后的光谱具有最优的三七粉质量等级的预测能力。

表 1 不同预处理方法 SVM 建模结果
Table 1 SVM modeling results of different pretreatment methods

预处理方式	训练集准确率/%	测试集准确率/%
SG	74.609 4	89.062 5
MSC	90.625 0	95.312 5
SNV	90.625 0	96.875 0

2.2 特征选择

2.2.1 迭代保留信息变量

在 IRIV 特征选择的过程中, 采用 5 折交叉验证的方法建立了偏最小二乘(PLS)模型。然后以交互验证均方误差(RMSECV)作为评价指标来选择特征波长。在每次迭代中, 剔除一些不相关和干扰的波长变量, 保留特征波长变量。从图 4 中可以看出, 在第 6 次迭代之前, 变量数量迅速减少, 从 478 个减少到 69 个。第 10 次迭代后, 完全剔除了无信息变量和干扰信息变量。最终, 反向消除后保留了 30 个有效波长, 分布在图 5 的平均光谱上。

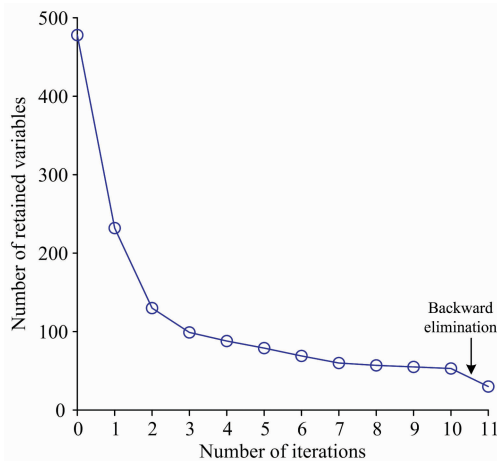


图 4 IRIV 选择的过程

Fig. 4 Process of IRIV selection

2.2.2 变量组合集群分析

在 VCPA 特征选择的过程中, 指数递减函数(EDF)和二进制矩阵采样(BMS)运行次数分别设置为 50 次和 1 000 次, BMS 初始采样权值设置为 0.5, 最优子集的比例设置为 10%, EDF 运行后剩余变量数设置为 14, 采用 5 折交叉验证的方法建立 PLS 模型, 计算所有子集的 RMSECV, 以 RM-

SECV 作为评价指标筛选出最优子集; 利用 EDF 剔除子集中贡献率较低的变量。迭代 50 次, 剩余 14 个变量。最后, 计算这 14 个变量之间所有变量组合的 RMSECV, 提取 RMSECV 最小的变量组合。最终选择了 11 个特征波长, 分布在图 6 的平均光谱上。

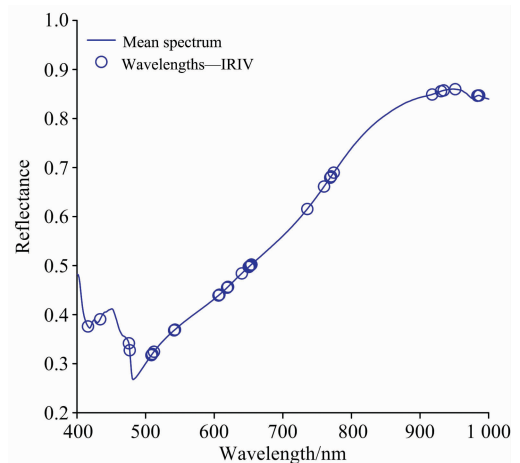


图 5 IRIV 所选择的波长

Fig. 5 Wavelength selected by IRIV

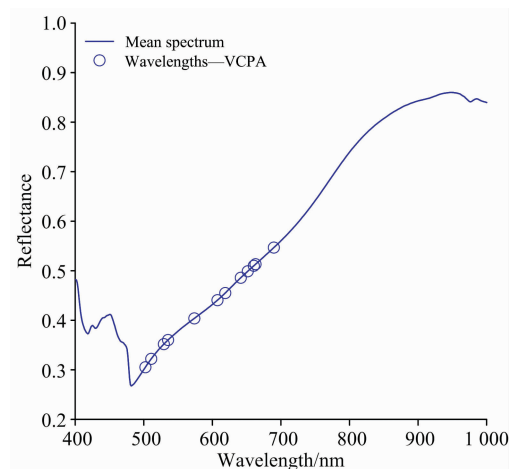


图 6 VCPA 所选择的波长

Fig. 6 Wavelength selected by VCPA

2.2.3 变量组合集群分析混合迭代保留信息变量

VCPA-IRIV 将 VCPA 与 IRIV 相结合, 先通过 VCPA 缩小变量空间, 再通过 IRIV 进一步优化剩余的变量。与 VCPA 和 IRIV 相比, VCPA-IRIV 消除了 VCPA 中最差子集对特征选择的不利影响; VCPA-IRIV 通过 EDF 消除贡献小的变量, 剩余的变量相对集中且优化, 使得 IRIV 更容易、更好的选择最优变量子集。在 VCPA-IRIV 特征选择过程中, 设置 EDF 运行后剩余变量数为 100, 其余参数与 2.2.2 节中相同。首先进行 VCPA, 利用 PLS 计算所有子集的 RMSECV, 筛选出 100 个最优子集, 再利用 EDF 剔除 100 个子集中贡献率较低的变量, 迭代 50 次, 剩下 100 个变量。再对这 100 个变量进行 IRIV, 剔除不相关和干扰的波长变量, 经多次迭代直至完全剔除了无信息变量和干扰信息变量。最终选

择了 18 个特征波长，分布在图 7 的平均光谱上。

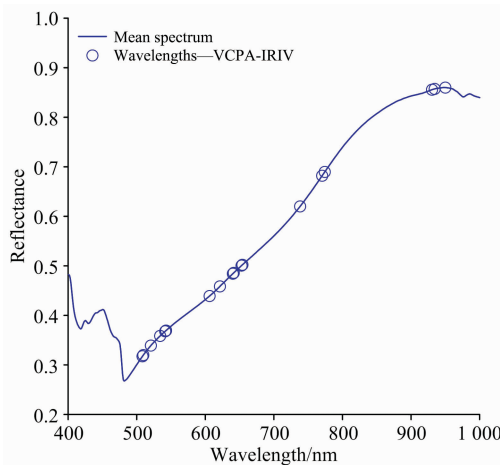


图 7 VCPA-IRIV 所选择的波长

Fig. 7 Wavelength selected by VCPA-IRIV

2.3 基于特征选择的 SVM 建模

建模试验由 2 个部分组成，第一，基于全光谱数据建立 SVM 模型。第二，基于 3 种特征波长数据建立 SVM 模型。SVM 模型参数同 2.1 节中相同，试验结果如表 2 所示。

表 2 基于全光谱和特征光谱的 SVM 建模结果

Table 2 SVM modeling results based on full spectral data and feature wavelengths

特征选择方式	波长数	训练集准确率/%	测试集准确率/%
无	478	90.625 0	96.875 0
IRIV	30	84.375 0	92.187 5
VCPA	11	86.328 1	92.968 8
VCPA-IRIV	18	90.234 4	96.875 0

观察表 2 可以发现，基于全光谱和特征光谱建立的 SVM 分类模型均取得了较好的效果，这 3 种特征选择方法均保存了三七粉的有效信息。对图 5、图 6 和图 7 进行对比分析，IRIV 提取的特征波长集中在 415.985~986.13 nm 范围内，部分波长在受噪声影响区域内；VCPA 提取的特征波长集中在 510.85~689.636 nm 范围内，忽视了近红外光谱区域的有效信息；VCPA-IRIV 提取的特征波长分布在 476.838~995.163 nm 范围内，保存了三七粉在可见光光谱和近红外光谱区域内的有效信息，也没有受到噪声的影响，SVM 模型分类准确率最高。因此认为 VCPA-IRIV 是最优的特征选择方法。在模型复杂度方面，VCPA-IRIV 算法简化了模型，降低了计算复杂度。在建模精度方面，VCPA-IRIV-SVM 模型的测试集分类准确率与全光谱 SVM 模型测试集分类准确率相同。由于参数 c 和 g 的选择对 SVM 的分类精度起着重要的作用，因此引入智能优化算法 GSA 对 SVM 中参数 c 和 g 进行寻优，并与网格搜索(grid search, GS)的结

果进行比较。

在 GSA 中，参数 c 和 g 的搜索范围分别设置为 $[0.01, 100]$ 和 $[0.1, 10]$ ，最大迭代次数设置为 100，群体规模设置为 20。在 GS 中，参数 c 和 g 的搜索范围均设置为 $[2^{-8}, 2^8]$ ，两种优化算法均采用 5 折交叉验证方式。建模结果如表 3 所示。

表 3 基于 VCPA-IRIV 特征选择方法的 GSA-SVM 和 GS-SVM 建模结果

Table 3 GSA-SVM and GS-SVM modeling results based on VCPA-IRIV

模型	c	g	训练集分类准确率/%	测试集分类准确率/%
VCPA-IRIV-GSA-SVM	15.868 0	1.048 0	100	100
VCPA-IRIV-GS-SVM	90.509 7	0.044 1	98.828 1	99.218 8

GS 具有更快的收敛速度，但由于搜索点固定，也错过了最优解，因此分类准确率低于 GSA。相比之下，VCPA-IRIV-GSA-SVM 模型性能最好，训练集和测试集的分类准确率均达到了 100%，最终选择 VCPA-IRIV-GSA-SVM 模型作为三七粉质量等级的分级模型。

3 结 论

市场上三七粉以次充好现象严重，为了保证三七粉质量，基于高光谱成像技术对不同质量等级的三七粉进行无损鉴别。首先，采集样本的高光谱图像，通过选择 ROI 得到 60×60 像素的光谱信息，然后分别用 SG, MSC 和 SNV 对光谱数据进行预处理，反映三七粉质量等级有效信息的特征波长分别由 IRIV, VCPA 和 VCPA-IRIV 提取。然后分别建立基于全光谱和特征光谱的 SVM 模型，并引入 GSA 和 GS 对 SVM 模型中的 c 和 g 进行优化。结果表明：

(1) 分别建立基于 SG, MSC 和 SNV 的分类模型并进行比较。结果表明，适当的预处理方法(SNV)可以使模型具有良好的性能。

(2) 分别建立 IRIV-SVM, VCPA-SVM, VCPA-IRIV-SVM 和全光谱的 SVM 模型，并进行比较。结果表明，适当的特征选择方法(VCPA-IRIV)可以在降低模型复杂度的情况下保持模型的性能。

(3) 引入 GSA 和 GS 对 SVM 模型中的 c 和 g 进行优化，以模型的测试集分类准确率作为评价指标。VCPA-IRIV-GSA-SVM 分类模型性能最优，训练集和测试集分类准确率均达到了 100%，因此将 GSA 作为优化原模型的智能算法。

综上，利用可见近红外高光谱成像技术对三七粉进行质量等级鉴别是可行的，该方法为市场上三七粉的质量等级鉴别提供了参考。

References

- [1] Chinese Pharmacopoeia Commission(国家药典委员会). The Pharmacopoeia of the People's Republic of China: Part 1(中华人民共和国药典: 一部). Beijing: China Medical Science and Technology Press(北京: 中国医药科技出版社), 2015. 11.
- [2] Meng Zhe, Huang Yang, Wang Lijun, et al. Separation Science Plus, 2020, 3(6): 200.
- [3] Li Chao, Qin Yunhua, Yang Qianxu, et al. Journal of Pharmaceutical and Biomedical Analysis, 2020, 182: 113127.
- [4] Yang Xiaodong, Song Jie, Peng Lin, et al. Infrared Physics and Technology, 2019, 103: 103101.
- [5] Zhou Yuhou, Zuo Zhitian, Xu Furong, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 226: 117619.
- [6] Cong Sunli, Sun Jun, Mao Hanping, et al. Journal of the Science of Food and Agriculture, 2018, 98(4): 1453.
- [7] SUN Ting, TIAN Jian-ping, HU Xin-jun, et al(孙 婷, 田建平, 胡新军, 等). Food and Fermentation Industries(食品与发酵工业), 2021, 47(5): 186.
- [8] SUN Jun, LU Xin-zi, ZHANG Xiao-dong, et al(孙 俊, 路心姿, 张晓东, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2016, 47(6): 215.
- [9] Wang Zheli, Tian Xi, Fan Shuxiang, et al. Infrared Physics and Technology, 2021, 112: 103596.
- [10] Weng Shizhuang, Tang Peipei, Yuan Hecai, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 234: 118237.
- [11] Jennifer Dumont, Tapani Hirvonen, Ville Heikkinen, et al. Computers and Electronics in Agriculture, 2015, 116: 118.
- [12] LIU Da-hui, XU Na, GUO Lan-ping, et al(刘大会, 徐 娜, 郭兰萍, 等). China Journal of Chinese Materia Medica(中国中药杂志), 2016, 41(5): 776.
- [13] Shrestha S, Matej K, Zibrat U, et al. Sensors and Actuators B: Chemical, 2016, 237: 1027.
- [14] YIN Wen-jun, RU Chen-lei, ZHENG Jie, et al(殷文俊, 茹晨雷, 郑 洁, 等). China Journal of Chinese Materia Medica(中国中药杂志), 2021, 46(4): 923.
- [15] Yao Kunshan, Sun Jun, Zhou Xin, et al. Journal of Food Process Engineering, 2020, 43(7): e13422.
- [16] SUN Zong-bao, WANG Tian-zhen, LIU Xiao-yu, et al(孙宗保, 王天真, 刘小裕, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2020, 40(10): 3224.
- [17] WANG Sheng-lin, YANG Chong-shan, LIU Zhong-yuan, et al(王盛琳, 杨崇山, 刘中原, 等). Journal of Tea Science(茶叶科学), 2021, 41(2): 251.
- [18] Li Yating, Sun Jun, Wu Xiaohong, et al. Journal of Food Science, 2019, 84(8): 2234.
- [19] CHEN Yu-xin, YIN Xiao-chuan, TAN Ren(陈玉鑫, 殷肖川, 谭 韧). Journal of Air Force Engineering University • Natural Science Edition(空军工程大学学报 • 自然科学版), 2018, 19(5): 78.

Study on Nondestructive Identification of *Panax Notoginseng* Powder Quality Grade Based on Hyperspectral Imaging Technology

ZHANG Fu-jie, SHI Lei, LI Li-xia*, ZHAO Hao-ran, ZHU Yin-long

Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China

Abstract *Panax notoginseng* powder is the main consumption and commodity form of *panax notoginseng*. There are shoddy or even adulterated phenomena in the market. As *panax notoginseng* powder is a powdery material, it is not easy to distinguish with the naked eye. In order to identify the quality grade of *panax notoginseng* powder, visible near-infrared hyperspectral imaging technology was used to identify the *panax notoginseng* powder with different quality grades. The taproots of *panax notoginseng* of 30 heads, 40 heads, 60 heads and 80 heads were ground into powder to prepare samples. The hyperspectral image of 384 samples of four quality grades was acquired by using a visible near-infrared hyperspectral imaging system(400.68 ~ 1 001.612 nm). Region of interest (ROI) was extracted from the hyperspectral image, and the average spectral value of samples was calculated. 384 samples of *panax notoginseng* powder were divided into training sets and test sets in a ratio of 2 : 1. The original spectra of *panax notoginseng* powder were preprocessed using multiplication scatter correction (MSC), Savitzky-Golay (SG) and standard normal variable (SNV), and the support vector machine (SVM) was employed to form the classification models based on MSC, SG and SNV. By comparing the classification accuracy of SVM models based on MSC, SG and SNV, it was found that SNV had the best effect on preprocessing. Iterative reserved information variable (IRIV), variable combined

cluster analysis (VCPA) and variable combined cluster analysis and iterative reserved information variable (VCPA-IRIV) were adopted to extract feature wavelengths from the spectra after SNV pretreatment, and the SVM was employed to form the classification models based on feature spectra and original spectra. By comparing the range of feature wavelengths and the classification accuracy of SVM models based on IRIV, VCPA and VCPA-IRIV, it was found that VCPA-IRIV, which combines VCPA and IRIV, had the best effect on feature selection. VCPA-IRIV extracted 18 feature wavelengths to participate in the modeling instead of the full spectra, and the algorithm can reduce the complexity of the model while maintaining the model's classification accuracy. In order to improve the classification accuracy of the model, the gravitational search algorithm (GSA) was introduced to search the optimal parameters (c, g) in the SVM model and compared with Grid Search (GS). The results indicated that the VCPA-IRIV-GSA-SVM model has the best classification effect, and the classification accuracy of the test set reached 100%. Thus, it is feasible to use visible near-infrared hyperspectral imaging technology to identify the quality grade of *panax notoginseng* powder. This method references the quality grade identification of *panax notoginseng* powder in the market.

Keywords Visible near-infrared hyperspectral imaging; *Panax notoginseng* powder; Feature selection; Support vector machine; Gravitational search algorithm

(Received Jun. 10, 2021; accepted Oct. 12, 2021)

* Corresponding author

本刊声明

近期以来,一些不法分子假冒《光谱学与光谱分析》期刊社名义,以虚假网站等形式欺骗广大作者、读者。这些虚假网站公然假冒《光谱学与光谱分析》期刊名义进行大肆的征稿并骗取作者的审稿费和版面费。经部分作者及读者举报,现有关部门已就此介入调查。本刊将通过法律途径向假冒者追究相应的责任,维护本刊权益。

本刊官方网站已正式开通,网址为

<http://www.gpxygpx.com/>

在此郑重声明,本网址为《光谱学与光谱分析》期刊唯一开通运行的官方网站。本刊从未授权任何单位或个人以任何形式(包括网上网下)代理本刊征稿、审稿等项业务。

希望广大读者和作者切实维护好自身的合法权益,防止受骗上当。

《光谱学与光谱分析》期刊社

2019年3月15日