

基于高光谱技术的灵芝孢子粉破壁率快速检测方法研究

潘在晨¹, 仲恂², 方玲², 祁哲晨¹, 徐靖³, 梁宗锁^{1*}, 李振皓^{2*} (1.浙江理工大学生命科学与医药学院, 浙江省植物次生代谢调控重点实验室, 杭州 310018; 2.浙江寿仙谷植物药研究院有限公司, 杭州 310012; 3.浙江省珍稀植物药工程技术研究中心, 浙江 武义 321200)

摘要: 目的 利用高光谱技术结合化学计量学建立灵芝孢子粉破壁率快速无损的检测方法。方法 采集不同破壁率灵芝孢子粉样品的高光谱图像, 选定感兴趣区域后计算获得各样品可见-短波近红外波段 (397~1 004 nm) 内的光谱数据; 比较运用标准正态变量变换、多元散射校正、Savitsky-Golay(SG) 平滑、小波变换、SG 平滑+标准正态变量变换及 SG 平滑+多元散射校正 6 种光谱预处理方法, 竞争性自适应重加权、连续投影算法、无信息变量选择、最小角回归、遗传算法 5 种特征波段提取方法以及偏最小二乘法、支持向量回归、极限学习机回归、多层感知机回归及 LightGBM 回归 5 种算法所建立的定量校正模型预测性能。结果 获得最优预测性能的算法组合为 SG 平滑+竞争性自适应重加权特征波段选择+偏最小二乘; 基于该算法组合建立的定量校正模型在破壁率区间为 90%~100% 的灵芝孢子粉样品预测集决定系数为 0.868 2, 均方根误差为 0.011 7; 将选定的最优算法组合应用于构建样品破壁率区间为 0~100% 的定量校正模型, 计算测试集决定系数为 0.973 1, 均方根误差为 0.049 3, 表现出良好的泛化能力。结论 所建立的定量检测模型可以实现对灵芝孢子粉破壁率的快速、无损检测, 为破壁灵芝孢子粉及其产品的质量控制在提供技术支撑。

关键词: 高光谱成像; 灵芝孢子粉; 破壁率; 化学计量学; 定量校正模型

中图分类号: R284.1 文献标志码: B 文章编号: 1007-7693(2024)06-0760-07

DOI: 10.13748/j.cnki.issn1007-7693.20231852

引用本文: 潘在晨, 仲恂, 方玲, 等. 基于高光谱技术的灵芝孢子粉破壁率快速检测方法研究[J]. 中国现代应用药学, 2024, 41(6): 760-766.

A Rapid, Hyperspectral-based Method for Determining Sporoderm-broken Rate of *Ganoderma Lucidum* Spore Powder

PAN Zaichen¹, ZHONG Yi², FANG Ling², QI Zhechen¹, XU Jing³, LIANG Zongsuo^{1*}, LI Zhenhao^{2*} (1.Zhejiang Provincial Key Laboratory of Plant Secondary Metabolism Regulation, College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2.Zhejiang Shouxiangu Institute of Plant Medicine, Hangzhou 310012, China; 3.Zhejiang Engineering Research Center of Rare Medicinal Plants, Wuyi 321200, China)

ABSTRACT: OBJECTIVE To establish a rapid nondestructive detection method for the sporoderm-broken rate of *Ganoderma lucidum* spore powder by hyperspectral technology combined with chemometrics. **METHODS** Hyperspectral images of *Ganoderma lucidum* spore powder samples with different sporoderm-broken rates were collected, and spectral data in the visible-shortwave near-infrared band(397–1 004 nm) range of each sample were calculated after selecting the region of interest. Compared 6 spectral preprocessing methods[standard normal variable transformation, multivariate scattering correction, Savitsky-Golay(SG) smoothing, wavelet transform, SG smoothing+standard normal variable transformation, and SG smoothing+multivariate scattering correction], 5 characteristic band extraction methods(competitive adaptive reweighting, successive projections algorithm, uninformative variables elimination, least angle regression, and genetic algorithm), and 5 algorithms(partial least squares regression, support vector regression, extreme learning machine, multilayer perceptron, and LightGBM) for constructing quantitative correction models to predicts performance. **RESULTS** The optimal combination was SG smoothing+competitive adaptive reweighted feature band selection+partial least squares. The quantitative correction model established based on the algorithm combination achieved a prediction set coefficient of 0.868 2, and a root mean square error of 0.011 7 for *Ganoderma lucidum* spore powder samples with a sporoderm-broken rate range of 90%–100%. The selected optimal algorithm combination was applied to construct a quantitative correction model with a sporoderm-broken rate range of 0–100%, the coefficient of determination for the test set was 0.973 1 and the root mean square error was 0.049 3, showing good generalization ability. **CONCLUSION** The established quantitative detection model can realize the rapid and non-destructive detection of the sporoderm-broken rate of *Ganoderma lucidum* spore powder, which provides technical support for the quality control of *Ganoderma lucidum* spore powder and its products.

KEYWORDS: hyperspectral imaging; *Ganoderma lucidum* spore powder; sporoderm-broken rate; chemometrics; quantitative calibration model

基金项目: 浙江省农业(食用菌)新品种选育重大科技专项 (2021C02073); 浙江省重点农业企业研究院项目 (2017Y20001)

作者简介: 潘在晨, 男, 硕士生 E-mail: 979137958@qq.com *通信作者: 梁宗锁, 男, 博士, 教授 E-mail: liangzs@ms.iswc.ac.cn

李振皓, 男, 博士, 副研究员 E-mail: zhenhao6@126.com

灵芝是中国传统名贵中药,具有补气安神、止咳平喘等功效,作为药品及保健食品原料广泛应用,同时已被列入食药物质试点名单。灵芝孢子是灵芝的繁殖体,研究表明其具有抗肿瘤、抗氧化、改善心血管、调节血糖血脂和增强免疫等功效^[1]。

灵芝孢子有双层孢壁结构,主要由几丁质等物质构成,质地坚韧,阻碍了多糖、三萜等有效成分的溶出。目前破壁技术已被广泛用于灵芝孢子粉的加工生产^[2-3],而破壁率是衡量破壁灵芝孢子粉质量的重要指标之一。灵芝孢子粉破壁率主要采用血球计数法、水装片结合显微技术、悬浮法结合物理技术等方法进行检测^[4],这些方法主要通过显微镜,人工计数镜检视野内破壁灵芝孢子的数量,再以相同条件下计数未破壁灵芝孢子的数量,计算破壁率,但是存在人为判断误差较大、镜检视野的局限性、样品损耗以及检测效率低等问题。因此,有必要采用先进分析仪器及技术开发快速、无损检测灵芝孢子粉破壁率的方法。

针对这一需求,本研究应用高光谱分析技术,探索其应用于灵芝孢子粉破壁率快速无损检测的可行性。高光谱成像技术作为一种能同时表征一维光谱信息和二维空间性信息的综合无损检测技术,可利用全波段或相关特征波段数据进行定量分析^[5]。近年来,高光谱检测已被应用于药材成分、真伪鉴别及产地识别等^[6-8],具有快速、准确、无损等优势。因此,本研究使用高光谱成像技术采集破壁灵芝孢子粉的高光谱数据,在对比研究多种光谱预处理、光谱特征提取及机器学习回归算法^[9-12]组合的基础上,建立了灵芝孢子粉破壁率的预测模型,开发了快速、准确、无损的灵芝孢子粉破壁率检测方法。

1 仪器与试剂

1.1 仪器

本研究所用高光谱成像系统由高光谱成像模块、移动平台、均匀光源、计算机与图像采集软件及1个用于减少环境光干扰的暗箱等部分组成。高光谱成像模块:Specim FX10相机和标准镜头 OLET15型可调节近红外聚焦透镜(Specim);移动平台:Specim LabScanner;均匀光源:3个150 W的光纤卤素灯线光源(Oriel Instruments, USA);图像采集软件:Lumo Scanner Software。

1.2 试验样品

本研究所用的破壁灵芝孢子粉(破壁率100%,通过血球计数法测定)与未破壁灵芝孢子粉(破壁率0%,通过血球计数法测定)均由浙江寿仙谷医药股份有限公司提供。

2 方法

2.1 样品的制备

破壁灵芝孢子粉和未破壁灵芝孢子粉按照一定的比例混合(90%~100%,混合间隔为1%,每份样品总重为3 g;0~100%,混合间隔为10%,每份样品总重为3 g),放入小型漩涡混合器中混匀5 min,使样品充分混匀后放入小型培养皿中铺满铺平。90%~100%内每种混合比例制备3个平行样,共计33个样品;0~100%内每种混合比例制备5个平行样,共计55个样品。

2.2 高光谱图像的采集与校正

在进行高光谱图像采集之前,应保证光源的稳定性,消除光谱仪自身的影响,因此实验前先将高光谱仪器开机预热20 min后,再进行图像采集。首先确定高光谱镜头与拍摄样品之间的最佳拍摄距离,随后调整高光谱采集数据的各项参数,具体参数设置:位移平台的速度为 $9.50\text{ mm}\cdot\text{s}^{-1}$,样品之间的距离为20 cm,电荷耦合器件相机曝光时间为26 ms。获得高光谱数据立方体,其宽为554像素,长为1 024像素,以1.67 nm的间隔从397 nm到1 004 nm的488个波长。高光谱成像系统示意图见图1。

高光谱图像黑白校正在采集得到高光谱图像后,为了减小光源不均匀,电荷耦合器件相机的暗电流以及仪器物理配置的差异对所获得的高光谱反射率图像的影响,需要对采集的原始高光谱图像进行黑白板校正,校正公式:

$$R = \frac{R_0 - B}{W - B}$$

其中 R 是校准的反射率图像, R_0 是原始反射率图像, W 是白板参考图像, B 是黑板参考图像。

2.3 感兴趣区域提取

黑白版校准后的图像,去除干扰背景和阈值分割之后,将高光谱图像转换成二进制图像。将二值图像应用于原始高光谱图像,得到破壁灵芝孢子粉的感兴趣区域图像,并对提取的感兴趣区域内每个像素点的光谱反射率进行平均处理作为

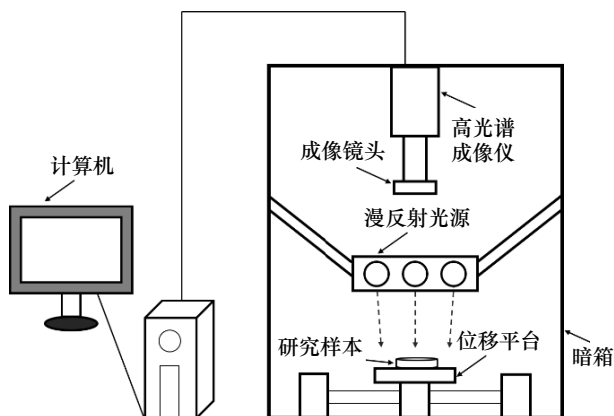


图1 高光谱成像系统示意图

Fig.1 Schematic diagram of hyperspectral imaging system

样品最终的光谱数据, 可以获得每份样品的平均光谱数据。

2.4 样品集划分

本研究采用 SPXY 法对破壁率 90%~100% 样本集按照 3 : 1 进行校正集和预测集的划分, 共 99 个样本作为校正集, 剩余共 33 个样本作为预测集; 对破壁率 0~100% 样本集按照 3 : 1 进行校正集和预测集的划分, 共 124 个样本作为校正集, 剩余共 41 个样本作为预测集。SPXY 是基于 KS 算法提出的一种改进方法^[13], KS 对数据集的划分依据是计算不同样本的 x 向量方向 (即常说的数据集的特征维度方向) 的欧氏距离; 而 SPXY 在此基础上增加了对不同样本的 y 向量方向 (即常说的数据集的真实值维度方向) 的欧氏距离的计算, 并通过正则化将 x 和 y 方向的距离结合, 更加全面地评估和划分数据集。

2.5 原始光谱的预处理

光谱数据除包含目标信息外, 往往还包含环境、仪器等因素产生的随机噪声, 为了减弱或消除无关噪声对数据的影响, 须对原始光谱数据进行预处理^[14]。本研究主要采用标准正态变量变换 (standard normal variate transformation, SNV)、多元散射校正 (multiplicative scatter correction, MSC)、Savitzky-Golay 平滑 (SG 平滑)、小波变换 (wavelet transform, WT)、SG 平滑+多元散射校正 (SG+MSC) 及 SG 平滑+标准正态变量变换 (SG+SNV) 6 种方法对原始光谱数据进行预处理。

SNV 和 MSC 作用基本相同, 用来校正样品因颗粒散射而引起的光谱的误差。SG 平滑处理可以有效平滑高频噪音, 提高信噪比。WT 主要用于数据压缩、平滑和滤波^[15]。

2.6 特征波段筛选

大量冗余的光谱信息可能掩盖与检测目标相关的关键信号, 增加建模的计算量和难度。光谱特征提取是筛选光谱数据保留其特征信息极为重要且有效的方法^[16]。本研究对比研究了竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS)^[17]、连续投影算法 (successive projections algorithm, SPA)^[18]、无信息变量选择法 (uninformative variables elimination, UVE)^[19]、最小角回归 (least angle regression, LAR)^[20] 及遗传算法 (genetic algorithm, GA)^[21] 5 种光谱特征提取方法在构建灵芝孢子粉破壁率定量校正模型中的应用效果。

2.7 模型的建立及评价指标

建立预测性能良好的数学模型是高光谱分析中的关键问题。本研究采用化学计量学方法建立高光谱定量分析模型^[22] 采用 PLSR^[23]、支持向量回归 (support vector regression, SVR)^[24]、极限学习机回归^[25]、多层感知机回归^[26] 及 LightGBM 回归^[27] 5 种方法建立定量模型。

评价模型本身预测性能大小的指标是校正集的决定系数 (R_c^2)、校正集均方根误差 (RMSEC)、交叉验证集决定系数 (R_{cv}^2)、交叉验证集均方根误差 (RMSECV)、测试集决定系数 (R_p^2)、测试集均方根误差 (RMSEP) 可用为模型预测性能评价指标, 测试集决定系数和测试集均方根误差是模型的主要参考标准。其中三者决定系数越接近于 1, 且均方根误差越低, 则模型精度越好^[28]。

2.8 数据处理

本研究采用 Python 3.10 进行光谱预处理、特征波段筛选和定量模型的构建等数据处理分析操作。

3 结果与讨论

3.1 90%~100% 破壁率最佳定量模型的构建

破壁灵芝孢子粉产品的破壁率>95% 被认定为合格产品。因此, 首先采取高光谱技术结合化学计量学方法对破壁率在 90%~100% 的灵芝孢子粉进行破壁率定量检测研究。

由于高光谱数据易受仪器噪声、周围环境等影响, 对后续建模有一定影响, 因此采用 SNV、MSC、SG、WT、SG+MSC 及 SG+SNV 共 6 种算法对原始光谱进行处理, 消除背景信息干扰, 增

强光谱与数据之间的相关性。采用 CARS、LAR、UVE、SPA 及 GA 法对破壁灵芝孢子粉预处理后的光谱数据进行特征波长提取，最终结合 PLS、

SVM、ELM、MLP 及 LightGBM 算法建立破壁灵芝孢子粉定量模型，模型比较结果见表 1。
通过特征波段的筛选能够在确保模型性能的

表 1 基于 PLS、SVM、ELM、MLP 及 LightGBM 的模型建模结果

Tab. 1 Model modeling results based on PLS, SVM, ELM, MLP and LightGBM

模型	PLSR		SVR		ELM		MLP		LightGBM	
	R_p^2	RMSEP	R_p^2	RMSEP	R_p^2	RMSEP	R_p^2	RMSEP	R_p^2	RMSEP
原始光谱	0.522 6	0.022 2	0.642 3	0.019 2	0.518 1	0.022 3	0.568 9	0.021 1	0.333 6	0.026 2
原始光谱-SNV	0.619 8	0.019 5	0.626 5	0.019 2	0.599 7	0.020 0	0.593 5	0.020 1	0.636 8	0.019 0
原始光谱-MSC	0.621 1	0.019 4	0.624 1	0.019 3	0.584 6	0.020 3	0.582 7	0.020 4	0.653 0	0.018 6
原始光谱-SG	0.592 6	0.020 5	0.709 7	0.017 3	0.633 1	0.019 5	0.539 6	0.021 8	0.332 0	0.026 2
原始光谱-WT	0.536 9	0.021 9	0.627 9	0.019 6	0.556 9	0.021 4	0.555 9	0.021 4	0.318 5	0.026 5
原始光谱-(SG+MSC)	0.562 9	0.019 9	0.753 4	0.015 3	0.611 4	0.019 1	0.643 5	0.018 3	0.715 5	0.016 3
原始光谱-(SG+SNV)	0.565 8	0.019 9	0.653 6	0.017 8	0.697 3	0.016 9	0.603 6	0.019 3	0.714 7	0.016 3
原始光谱-CARS	0.721 6	0.016 8	0.752 7	0.016 0	0.505 3	0.022 4	0.510 0	0.022 4	0.372 8	0.025 2
原始光谱-LAR	0.639 7	0.018 7	0.601 7	0.019 6	0.570 0	0.020 3	0.469 4	0.022 8	0.258 6	0.026 8
原始光谱-UVE	0.739 4	0.016 4	0.768 8	0.015 4	0.610 9	0.020 1	0.575 5	0.021 0	0.417 4	0.024 5
原始光谱-SPA	0.553 2	0.020 8	0.463 3	0.022 3	0.516 0	0.021 8	0.427 9	0.023 7	0.201 3	0.027 9
原始光谱-GA	0.537 1	0.021 7	0.689 4	0.017 7	0.571 3	0.021 0	0.492 1	0.022 6	0.372 7	0.025 3
SNV-CARS	0.863 2	0.011 4	0.772 7	0.014 1	0.549 0	0.019 2	0.509 4	0.021 5	0.620 9	0.018 3
SNV-LAR	0.590 6	0.020 5	0.607 7	0.020 2	0.558 8	0.021 4	0.510 9	0.022 6	0.629 1	0.019 6
SNV-UVE	0.754 5	0.014 8	0.727 0	0.015 4	0.659 0	0.017 8	0.613 1	0.018 4	0.631 3	0.017 9
SNV-SPA	0.653 9	0.018 6	0.628 6	0.018 6	0.799 4	0.013 4	0.268 8	0.028 1	0.616 2	0.019 8
SNV-GA	0.586 7	0.019 6	0.530 8	0.020 9	0.568 8	0.019 4	0.506 4	0.022 6	0.650 9	0.017 3
MSC-CARS	0.796 4	0.013 8	0.606 4	0.017 6	0.563 0	0.019 0	0.562 0	0.020 7	0.576 9	0.018 6
MSC-LAR	0.610 7	0.020 6	0.580 5	0.021 3	0.543 1	0.022 3	0.565 2	0.021 7	0.663 4	0.019 1
MSC-UVE	0.802 4	0.013 6	0.764 5	0.014 5	0.592 9	0.018 8	0.601 9	0.019 3	0.556 1	0.019 7
MSC-SPA	0.734 1	0.017 2	0.496 7	0.024 1	0.631 9	0.018 6	0.405 2	0.024 8	0.584 1	0.019 3
MSC-GA	0.679 1	0.018 4	0.655 8	0.018 5	0.522 6	0.021 1	0.587 4	0.020 0	0.647 3	0.018 3
SG-CARS	0.868 2	0.011 7	0.699 2	0.017 5	0.707 1	0.017 2	0.535 9	0.021 9	0.332 4	0.026 1
SG-LAR	0.707 7	0.017 5	0.623 2	0.020 0	0.657 2	0.019 1	0.570 4	0.021 4	0.492 0	0.023 2
SG-UVE	0.687 0	0.017 8	0.796 8	0.014 3	0.679 3	0.017 9	0.572 6	0.020 8	0.447 8	0.023 4
SG-SPA	0.549 7	0.023 7	0.462 9	0.022 6	0.645 6	0.020 2	0.559 4	0.021 1	0.249 6	0.027 8
SG-GA	0.513 8	0.022 3	0.738 0	0.016 4	0.623 3	0.019 4	0.560 7	0.021 0	0.410 7	0.024 5
WT-CARS	0.744 4	0.016 4	0.733 7	0.016 3	0.557 8	0.021 4	0.565 6	0.020 7	0.436 2	0.024 1
WT-LAR	0.615 4	0.019 2	0.559 4	0.020 5	0.489 9	0.022 0	0.515 2	0.021 6	0.426 5	0.023 3
WT-UVE	0.705 9	0.017 4	0.755 2	0.015 9	0.601 6	0.020 1	0.561 8	0.021 3	0.490 6	0.022 9
WT-SPA	0.441 0	0.021 5	0.423 6	0.022 3	0.608 4	0.019 4	0.394 7	0.024 8	0.303 9	0.025 4
WT-GA	0.582 1	0.020 6	0.723 4	0.016 9	0.513 4	0.022 4	0.577 8	0.020 8	0.406 9	0.024 5
(SG+MSC)-CARS	0.826 4	0.013 0	0.669 9	0.018 5	0.709 3	0.016 1	0.676 0	0.017 4	0.645 5	0.018 0
(SG+MSC)-LAR	0.595 3	0.019 2	0.579 8	0.019 5	0.516 2	0.021 0	0.559 8	0.020 0	0.579 6	0.019 6
(SG+MSC)-UVE	0.795 4	0.013 8	0.762 9	0.014 8	0.734 6	0.015 6	0.460 8	0.021 1	0.737 1	0.015 5
(SG+MSC)-SPA	0.697 7	0.017 9	0.575 1	0.021 1	0.687 0	0.018 0	0.398 5	0.024 2	0.677 8	0.017 6
(SG+MSC)-GA	0.608 6	0.018 8	0.665 9	0.017 4	0.625 3	0.018 9	0.647 7	0.018 0	0.688 1	0.016 8
(SG+SNV)-CARS	0.770 1	0.014 8	0.784 9	0.014 9	0.676 6	0.018 0	0.563 2	0.021 2	0.666 4	0.017 9
(SG+SNV)-LAR	0.625 5	0.019 0	0.526 5	0.021 2	0.524 0	0.021 2	0.490 7	0.022 1	0.633 5	0.018 8
(SG+SNV)-UVE	0.837 9	0.012 1	0.830 6	0.012 5	0.698 5	0.016 7	0.644 6	0.018 0	0.714 5	0.016 0
(SG+SNV)-SPA	0.602 7	0.017 9	0.759 2	0.016 2	0.743 0	0.016 8	0.402 0	0.024 3	0.678 2	0.018 7
(SG+SNV)-GA	0.650 6	0.018 8	0.703 9	0.016 8	0.664 8	0.017 6	0.607 3	0.019 1	0.708 5	0.016 3

基础上,大幅减少对光谱变量的使用,筛选变量后基本可以达到甚至超过全光谱变量所构建模型的性能。说明利用特征波段提取方法能够排除冗余信息,找到与指标成分密切相关的光谱变量,从而大幅减少计算量。可见-短波近红外光谱经SG预处理后,在CARS筛选后建立的PLSR模型具有最优性能,破壁率实际测量值与光谱的预测值相关性较高。其中SG-CARS-PLSR模型具有最优的预测性能,其 $R_p^2=0.8682$ 、RMSEP=0.0117,选定为构建定量校正模型的算法组合。

3.2 90%~100%破壁率区间的最优定量模型分析

破壁率区间为90%~100%的灵芝孢子粉样品在可见-短波近红外波段下经过SG平滑预处理后光谱曲线及CARS筛选的特征波段分布见图2,在使用CARS方法对SG平滑后的光谱提取特征波长时,设定蒙特卡洛采样次数为50次,并使用5折交叉验证法,最终共提取55个特征波长,占全光谱波段的12.3%。SG-CARS-PLSR的模型性能指标见表2。经过SG平滑预处理再进行CARS提取特征波长后所建立的PLSR模型效果最

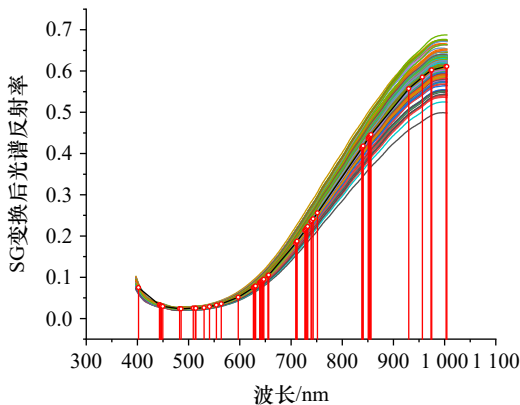


图2 经SG平滑后的破壁率区间为90%~100%的样本光谱曲线及CARS筛选的特征波段分布图

Fig. 2 Spectral curve of the sample with the sporoderm-broken rate interval after SG smoothing was 90%–100% and the characteristic band distribution map screened by CARS

表2 SG-CARS-PLSR模型预测结果

Tab. 2 Results of SG-CARS-PLSR model prediction

模型	校正集		验证集		测试集	
	R_c^2	RMSEC	R_{cv}^2	RMSECV	R_p^2	RMSEP
原始光谱-PLSR	0.875 7	0.010 9	0.642 5	0.018 6	0.522 6	0.022 2
SG-PLSR	0.851 8	0.011 4	0.739 0	0.016 0	0.592 6	0.020 5
CARS-PLSR	0.872 0	0.011 0	0.823 1	0.012 8	0.721 6	0.016 8
SG-CARS-PLSR	0.918 7	0.008 9	0.832 7	0.012 6	0.868 2	0.011 7

好,其校正集、验证集和测试集的决定系数分别为0.9187, 0.8327和0.8682,优于直接使用原始光谱所建立的定量校正模型。通过直线拟合可以发现训练集和预测集的破壁率区间为90%~100%的灵芝孢子粉破壁率测量值分别与相应的预测值有良好的线性关系,见图3。

3.3 模型泛化能力

使用破壁率区间为0~100%灵芝孢子粉样品验证所选定量校正建模算法组合的泛化能力。样品经SG平滑后光谱曲线和模型中预测值和真实值之间的关系图见图4,预测性能见表3。SG-CARS-PLSR模型的校正集和测试集的相关系数分别为0.9877和0.9731。利用SG-CARS组合建立的PLSR模型定量评价指标都有提升,通过直线拟合可以发现训练集和测试集的破壁率区间为0~100%的灵芝孢子粉破壁率测量值分别与相应的预测值有良好的线性关系。结果表明该方法适用于不同破壁率的检测,为该方法应用于破壁过程的在线/旁线检测奠定了基础。

挑选出的SG-CARS-PLSR模型在破壁率0~

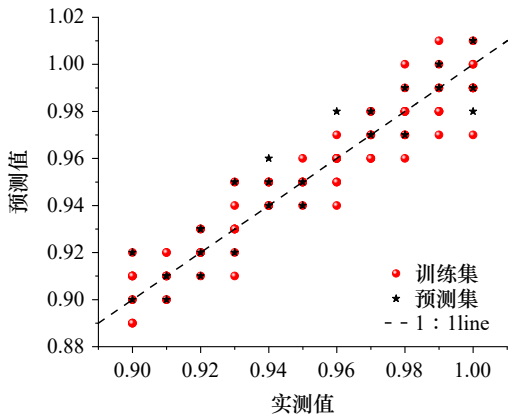


图3 破壁率90%~100%样本SG-CARS-PLSR模型中预测值和真实值之间的关系图

Fig. 3 Relationship between predicted values and true values in the SG-CARS-PLSR model with a sporoderm-broken rate of 90%–100% samples

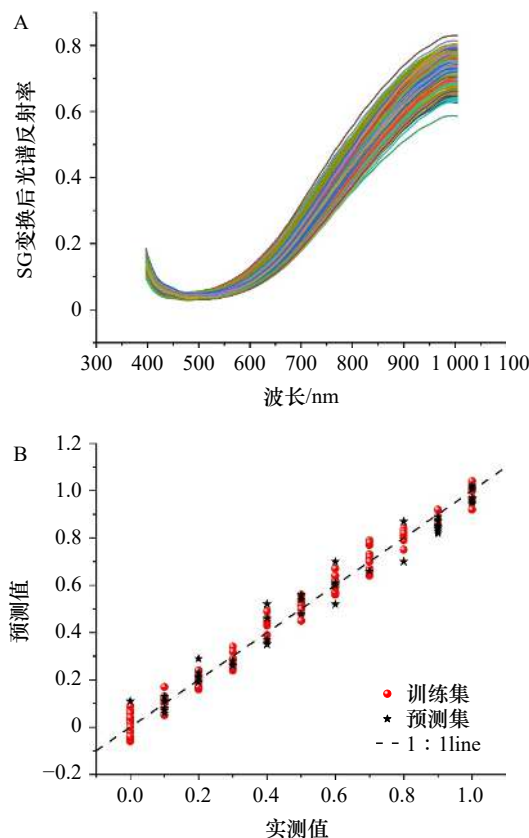


图4 破壁率 0~100% 样本在可见-近红外波段 SG 预处理后的光谱曲线 (A) 和破壁率 0~100% 样本 SG-CARS-PLSR 模型中预测值和真实值之间的关系图 (B)

Fig. 4 Relationship between predicted and true values in the SG-CARS-PLSR model for samples with 0–100% sporoderm-broken rate after SG pretreatment in the visible-near-infrared band(A) and the predicted value and true value of samples with 0–100% sporoderm-broken rate(B)

100% 的测试集决定系数为 0.973 1、均方根误差为 0.049 3。结果表明 SG-CARS-PLSR 模型具有良好的泛化能力。首先由于数据分布的原因, SG-CARS-PLSR 模型在破壁率 0~100% 的模型性能优于破壁率 90%~100%。2 种建模区间相比之下, 前者数据更为离散, 离散数据在模型中有着更好的鲁棒性和稳定性, 能够有效降低模型过拟合的风险。其次由于破壁率 90%~100% 内的相对实验误

表3 最优模型对破壁率区间为 0~100% 的样品预测性能

Tab. 3 Optimal model predicts the performance of samples with a sporoderm-broken rate range of 0–100%

模型	校正集		验证集		测试集	
	R_c^2	RMSEC	R_{cv}^2	RMSECV	R_p^2	RMSEP
原始光谱-PLSR	0.977 4	0.045 2	0.966 2	0.056 7	0.970 6	0.055 3
SG-PLSR	0.979 8	0.042 8	0.969 7	0.053 5	0.972 0	0.053 6
CARS-PLSR	0.990 1	0.031 2	0.980 1	0.044 4	0.978 0	0.045 8
SG-CARS-PLSR	0.987 7	0.034 9	0.984 0	0.039 7	0.973 1	0.049 3

差更大, 导致破壁率 90%~100% 的模型性能弱于破壁率 0~100%。

4 结论

破壁率是破壁灵芝孢子粉及相关产品的核心质量指标。本研究采用 SG 预处理结合 CARS 特征筛选方法建立的 PLSR 模型对破壁率为 90%~100% 样本有良好的预测准确率, 实现了灵芝孢子粉破壁率的定量检测。此外, 发现将 SG-CARS-PLSR 算法组合应用于破壁率 0~100% 样本中, 同样有良好的预测准确率。本研究为破壁灵芝孢子粉破壁率的快速、准确检测提供了可行的方法。该方法也有望进一步拓展应用于灵芝孢子粉破壁过程的在线检测。

REFERENCES

- [1] 中国药典. 一部[S]. 2015.
- [2] ZHAO J X, LI Z Y, ZHANG G L, et al. Knowledge mapping analysis of *Ganoderma lucidum* spore powder-related studies based on CiteSpace software[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2021, 38(12): 1416-1425.
- [3] WANG H B, HONG Z P, LI Z Y, et al. Study on the antithrombotic and cardiac function improvement effects of *Ganoderma lucidum* spore powder with different manufacturing techniques[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2021, 38(19): 2367-2373.
- [4] GAO Z C, ZHOU M, ZHANG R G, et al. Determination of wall breaking rate of *Ganoderma lucidum* spore powder[J]. Food Ind(食品工业), 2022, 43(9): 146-149.
- [5] YANG W H, HAO J W, HUANG H P, et al. Rapid determination of polysaccharide in *Pteridium aquilinum* by near infrared diffuse reflectance spectroscopy[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2023, 40(5): 597-602.
- [6] TAO Y, CHEN L, JIANG E C, et al. Application and prospects of hyperspectral imaging and deep learning in traditional Chinese medicine in context of AI and industry 4.0[J]. China J Chin Mater Med(中国中药杂志), 2020, 45(22): 5438-5442.
- [7] LI M, ZHANG X B, LIU S B, et al. Partly interpretable machine learning method of ginseng geographical origins recognition and analysis by hyperspectral measurements[J]. Spectrosc Spectr Anal(光谱学与光谱分析), 2022, 42(4): 1217-1221.

- [8] SUN M, WANG R, CHEN X H, et al. Application of hyperspectral imaging technique for *Cordyceps* powder content and true/false identification[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2017, 34(8): 1114-1117.
- [9] WOLD S, TRYGG J, BERGLUND A, et al. Some recent developments in PLS modeling[J]. Chemom Intell Lab Syst, 2001, 58(2): 131-150.
- [10] CHEN K, ZHU Y. A summary of machine learning and related algorithms[J]. Stat Inf Forum(统计与信息论坛), 2007, 22(5): 105-112.
- [11] ZHANG L, ZHANG L M, LI Y, et al. Application and improvement of partial-least-squares in Fourier transform infrared spectroscopy[J]. Spectrosc Spectr Anal(光谱学与光谱分析), 2005, 25(10): 1610-1613.
- [12] LI J X, NI L, WANG Y F, et al. Identification models of Raman spectrum for ostreae concha and haliotidis concha based on multiple algorithms[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2023, 40(4): 477-482.
- [13] LIU N, XING Z Z, QIAO L, et al. Discussion on spectral variables selection of potato chlorophyll using model population analysis[J]. Spectrosc Spectr Anal(光谱学与光谱分析), 2020, 40(7): 2259-2266.
- [14] ZHOU C, WANG H, YANG J, et al. Origin identification of Gardeniae Fructus based on hyperspectral imaging technology[J]. China J Chin Mater Med(中国中药杂志), 2022, 47(22): 6027-6033.
- [15] NIZHEN, HU C Q, FENG F. Progress and effect of spectral data pretreatment in NIR analytical technique[J]. Chin J Pharm Anal(药物分析杂志), 2008, 28(5): 824-829.
- [16] JIANG Q H, LIU F, YU D Y, et al. Rapid measurement of the pharmacological active constituents in herba epimedii using hyperspectral analysis technology[J]. Spectrosc Spectr Anal(光谱学与光谱分析), 2022, 42(5): 1445-1450.
- [17] YUAN L, SHI B, YU J C, et al. Application of different smoothing ensemble CARS algorithm in spectral discrimination of black tea grade[J]. Spectrosc Spectr Anal(光谱学与光谱分析), 2020, 40(10): 3254-3259.
- [18] ZHANG J K, RIVARD B, ROGGE D M. The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data[J]. Sensors, 2008, 8(2): 1321-1342.
- [19] YE S F, WANG D, MIN S G. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection[J]. Chemom Intell Lab Syst, 2008, 91(2): 194-199.
- [20] WANG S, HUANG M, ZHU Q B. Model fusion for prediction of apple firmness using hyperspectral scattering image[J]. Comput Electron Agric, 2012(80): 1-7.
- [21] PANDEY H M, CHAUDHARY A, MEHROTRA D. A comparative review of approaches to prevent premature convergence in GA[J]. Appl Soft Comput, 2014, 24: 1047-1077.
- [22] MA H, FENG X J, CHEN M, et al. Rapid determination of the moisture content of the original medicinal materials of lanqin oral solution by near-infrared spectroscopy coupled with chemometric algorithms[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2021, 38(23): 2932-2939.
- [23] JIANG C, SHOU D, YU Z M, et al. Rapid prediction of extractum and six active components in traditional she medicine-Melastoma dodecandrum lour. by ultraviolet spectroscopy and partial least squares regression algorithm[J]. Chin J Mod Appl Pharm(中国现代应用药学), 2020, 37(13): 1574-1579.
- [24] ZHANG C, WANG Q N, LIU F, et al. Rapid and non-destructive measurement of spinach pigments content during storage using hyperspectral imaging with chemometrics[J]. Measurement, 2017(97): 149-155.
- [25] LI B, HAN Z Y, WANG Q, et al. Research on bruise level detection of loquat based on hyperspectral imaging technology[J]. Spectrosc Spectr Anal(光谱学与光谱分析), 2023, 43(6): 1792-1799.
- [26] ZHANG C, GUO Y, LI M. Review of development and application of artificial neural network models[J]. Comput Eng Appl(计算机工程与应用), 2021, 57(11): 57-69.
- [27] KE G L, MENG Q, FINLEY T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., 2017: 3149-3157.
- [28] JIANG E C, CHEN L, YAN J Z, et al. Integrated smart hyperspectral imaging and CARS-based characteristic band selection for rapid determination of SO₂ content in sulphur-fumigated Achyranthis Bidentatae Radix[J]. China J Chin Mater Med(中国中药杂志), 2022, 47(7): 1864-1870.

收稿日期: 2023-07-03

(本文责编: 沈倩)