

## 基于近红外光谱技术的腰痹通胶囊中间体质量分析通用模型 建模方法研究

斯乐婷<sup>1</sup>, 张欣<sup>2</sup>, 张永超<sup>2</sup>, 张江艳<sup>2</sup>, 王钧<sup>2</sup>, 陈勇<sup>3</sup>, 刘雪松<sup>3</sup>, 吴永江<sup>3\*</sup>

(1. 浙江大学医学院附属妇产科医院药剂科, 浙江 杭州 310006; 2. 中药制药过程控制与智能制造技术全国重点实验室(江苏康缘药业股份有限公司), 江苏 连云港 222047; 3. 浙江大学药学院, 浙江 杭州 310058)

**摘要:** 采用近红外光谱技术 (near infrared spectroscopy, NIRS) 结合化学计量学建立腰痹通胶囊生产过程中中间体的质量分析通用模型, 实现三七皂苷 R1、人参皂苷 Rg1、人参皂苷 Re、人参皂苷 Rb1、人参皂苷 Rd 及水分 6 个关键质量指标的快速检测。以喷干细粉和总混颗粒为研究对象, 采用高效液相色谱法测定 5 种皂苷成分含量、烘干法测定水分含量作为参考值, 并采集近红外光谱。经蒙特卡洛交叉验证 (Monte Carlo cross validation, MCCV) 剔除异常样本后, 使用蒙特卡洛-无信息变量消除 (Monte Carlo uninformative variables elimination, MC-UVE)、竞争性自适应重加权采样 (competitive adaptive reweighted sampling, CARS) 选择特征变量, 分别采用偏最小二乘回归 (partial least squares regression, PLSR)、极限学习机 (extreme learning machine, ELM) 和蚁狮算法优化的最小二乘向量机 (ant lion optimization least squares support vector machine, ALO-LSSVM) 建立定量模型, 比较模型效果。结果表明, 经 CARS 筛选变量后建立的 ALO-LSSVM 模型效果最佳, 6 种指标成分的模型相关系数均大于 0.93, 相对标准误差控制在 6% 之内, ALO-LSSVM 更适用于数量多、信息丰富的样本, 模型预测效果和稳定性得到显著提高。本研究建立的通用模型具有良好的预测效果, 可用于腰痹通胶囊中间体的快速检测。

**关键词:** 腰痹通胶囊; 中间体; 近红外光谱技术; 通用模型; 蚁狮算法优化的最小二乘向量机

中图分类号: R917 文献标识码: A 文章编号: 0513-4870(2025)02-0471-08

## Study on the modeling method of general model of Yaobitong capsule intermediates quality analysis based on near infrared spectroscopy

SI Le-ting<sup>1</sup>, ZHANG Xin<sup>2</sup>, ZHANG Yong-chao<sup>2</sup>, ZHANG Jiang-yan<sup>2</sup>, WANG Jun<sup>2</sup>,  
CHEN Yong<sup>3</sup>, LIU Xue-song<sup>3</sup>, WU Yong-jiang<sup>3\*</sup>

(1. Department of Pharmacy, Women's Hospital School of Medicine Zhejiang University, Hangzhou 310006, China;

2. State Key Laboratory on Technologies for Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacture (Jiangsu Kanion Pharmaceutical Co., Ltd.), Lianyungang 222047, China; 3. College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China)

**Abstract:** The general models for intermediates quality analysis in the production process of Yaobitong capsule were established by near infrared spectroscopy (NIRS) combined with chemometrics, realizing the rapid determination of notoginsenoside R1, ginsenoside Rg1, ginsenoside Re, ginsenoside Rb1, ginsenoside Rd and moisture. The spray-dried fine powder and total mixed granule were selected as research objects. The contents of five saponins were determined by high performance liquid chromatography and the moisture content was

收稿日期: 2024-09-30; 修回日期: 2024-12-10.

\*通讯作者 Tel: 86-571-88208455, E-mail: yjwu@zju.edu.cn

DOI: 10.16438/j.0513-4870.2024-0955

determined by drying method. The measured contents were used as reference values. Meanwhile, NIR spectra were collected. After removing abnormal samples by Monte Carlo cross validation (MCCV), Monte Carlo uninformative variables elimination (MC-UVE) and competitive adaptive reweighted sampling (CARS) were used to select feature variables respectively. Based on the feature variables, quantitative models were established by partial least squares regression (PLSR), extreme learning machine (ELM) and ant lion optimization least squares support vector machine (ALO-LSSVM). The results showed that CARS-ALO-LSSVM model had the optimum effect. The correlation coefficients of the six index components were greater than 0.93, and the relative standard errors were controlled within 6%. ALO-LSSVM was more suitable for a large number of samples with rich information, and the prediction effect and stability of the model were significantly improved. The general models with good predicting effect can be used for the rapid quality determination of Yaobitong capsule intermediates.

**Key words:** Yaobitong capsule; intermediate; near infrared spectroscopy; general model; ant lion optimization least squares support vector machine

腰痹通胶囊为中药复方胶囊,由三七、川芎、延胡索、白芍、牛膝、狗脊、大黄、当归共八味中药经提取、浓缩、喷雾干燥、制粒、总混、填充等步骤制成,主治腰椎间盘突出症<sup>[1]</sup>。目前对腰痹通胶囊的研究主要集中在药理及临床疗效上,对其生产过程的质量控制研究较少<sup>[2-4]</sup>。喷干细粉和总混颗粒是腰痹通胶囊生产过程中喷雾干燥和总混后的两个关键中间体,其质量将直接影响到腰痹通胶囊成品的质量。皂苷类成分和水分是腰痹通中间体需要检测的共同指标,而传统分析方法(如高效液相色谱法、烘干法测水分)存在耗时长、操作繁琐等缺点,难以满足生产过程检测的需求。因此,建立一种快速、有效的中间体检测方法具有重要的应用价值。

近年来,近红外光谱技术(near infrared spectroscopy, NIRS)因其无需预处理、检测速度快、无污染等优点,在制药领域得到了广泛应用<sup>[5-7]</sup>。已有研究报道NIRS应用于苏黄止咳胶囊<sup>[8]</sup>、骨龙胶囊<sup>[9]</sup>、胆宁片<sup>[10]</sup>等中间体的快速检测,然而多数研究仅针对单个中间体建立分析模型。针对中药多个工段,通常每个工段的中间体需建立不同分析模型,如能建立通用模型则可以降低建模成本,有利于NIRS的应用推广。不同中间体样本合并后,建模信息更加复杂,适用于中药生产过程快速检测的通用模型开发存在一定难度,优选合适的建模方法成为关键。

本研究将喷干细粉和总混颗粒样本合并,基于NIRS建立三七皂苷R1、人参皂苷Rg1、人参皂苷Re、人参皂苷Rb1、人参皂苷Rc及水分6个共有关键质量指标的通用定量模型。采用蒙特卡洛-无信息变量消除(Monte Carlo uninformative variables elimination, MC-UVE)、竞争性自适应重加权采样(competitive adaptive reweighted sampling, CARS)筛选特征波长以减少无关变量,并比较偏最小二乘回归(partial least

squares regression, PLSR)、极限学习机(extreme learning machine, ELM)和最小二乘支持向量机(least squares support vector machine, LSSVM)建立的模型效果,选出每种指标成分的最优模型作为检测模型,以期为腰痹通胶囊生产过程质量控制提供方法基础。

## 材料与方法

**仪器** Thermo Scientific UltiMate3000 高效液相色谱仪(美国 Thermo Fisher 公司),配有 Corona Veo 电雾式检测器; Bruker Matrix-I 傅里叶变换近红外光谱仪(德国 Bruker 公司),内置积分球; XS105 十万分之一电子分析天平[梅特勒-托利多仪器(上海)有限公司]; FA2004B 万分之一电子分析天平(上海越平科学仪器有限公司); Milli-Q 超纯水仪(美国 Millipore 公司); DK-S26 电热恒温水浴锅; TDZ4-WS 台式低速离心机(湖南湘仪离心机仪器有限公司); DUG-9146A 电热恒温鼓风干燥箱(上海精宏实验仪器设备有限公司)。

**材料** 三七皂苷R1对照品(批号: MUST-18121705, 纯度为99.98%); 人参皂苷Rg1对照品(批号: MUST-19022811, 纯度为99.16%); 人参皂苷Re对照品(批号: MUST-19072510, 纯度为98.72%); 人参皂苷Rb1对照品(批号: MUST-19060116, 纯度为98.58%); 人参皂苷Rd对照品(批号: MUST-19060116, 纯度为98.55%); 均购自成都曼思特生物技术有限公司。乙腈和甲醇(HPLC级,北京百灵威科技有限公司); 甲醇(分析级,国药集团化学试剂有限公司); 实验用水为超纯水。

106批腰痹通喷干细粉、122批总混颗粒,在江苏康缘药业股份有限公司生产线采集。

**近红外光谱采集** 以空气为背景,采集喷干细粉和总混颗粒样品的近红外漫反射光谱。光谱采集范围为10 000~4 000  $\text{cm}^{-1}$ ,扫描分辨率16  $\text{cm}^{-1}$ ,扫描次数为64次。每个样品扫描3次,用OPUS软件计算得到

平均光谱。

**皂苷成分含量测定方法** 采用文献[11]中的高效液相色谱法测定皂苷成分含量, 取药品粉末 0.5 g, 精密称定, 置于具塞锥形瓶中, 用移液管精密加入 50 mL 甲醇, 称定重量, 加热回流 30 min, 放冷, 用甲醇补足减失的重量, 以 4 000 r·min<sup>-1</sup> 的转速离心 5 min, 取上清液, 用 0.22 μm 的滤膜过滤, 即得供试品溶液。

取适量三七皂苷 R1、人参皂苷 Rg1、人参皂苷 Re、人参皂苷 Rb1 和人参皂苷 Rd 对照品, 精密称定, 置于 25 mL 量瓶中, 用甲醇溶解并定容至刻度, 摇匀, 制成浓度分别为 0.339 2、1.089 2、0.219 2、1.030 0、0.318 8 mg·mL<sup>-1</sup> 的混合对照品储备液。

色谱柱为 XBridge Phenyl (4.6 mm × 150 mm, 3.5 μm); 以乙腈 (A)-水 (B) 为流动相, 梯度洗脱程序如下: 0~18 min, 15%~21% A; 18~30 min, 21%~23% A; 30~50 min, 23%~32% A; 50~60 min, 32%~34% A, 60~65 min, 34%~60% A。流速为 1.0 mL·min<sup>-1</sup>, 进样量为 10 μL, 柱温为 30 °C。CAD 参数设置: 雾化温度为 35 °C, 气压为 60.2 psi (1 psi ≈ 6.9 kPa), 滤波为 3.6 s, 采集频率为 5 Hz。

**液相色谱方法学考察** 总混颗粒 (即胶囊内容物) 已于文献[11]中进行液相色谱方法学验证, 不再赘述, 因此本文仅针对喷干细粉做专属性、重复性、加样回收率、稳定性考察。

取空白溶剂 (纯甲醇)、混合对照品溶液、喷干细粉供试品溶液和总混颗粒供试品溶液, 在上述色谱条件下进样分析, 比较分析色谱图以考察方法的专属性。

取喷干细粉 (批号: Z1903025-1) 6 份, 按“皂苷成分含量测定方法”项下方法平行制备供试品溶液, 在上述色谱条件下进样分析, 计算 6 份样品中 5 种皂苷含量的 RSD 值, 以考察方法的重复性。

取已知含量的喷干细粉 (批号: Z1903025-1) 6 份, 按 5 种皂苷成分含量的 100% 水平加入对照品, 按“皂苷成分含量测定方法”项下方法平行制备供试品溶液, 在上述色谱条件下进样分析, 计算 6 份样品的平均加样回收率及加样回收率的 RSD 值, 以考察加样回收率。

取喷干细粉 (批号: Z1903025-1) 的供试品溶液, 放置于室温下, 分别在 0、4、8、12、16、20、24 h 后按上述色谱条件进样分析, 记录 5 种皂苷的峰面积, 计算 RSD 值, 以考察供试品溶液的稳定性。

**水分测定** 根据 2020 年版中国药典规定, 参照烘干法 (通则 0832 第二法) 进行测定。

**异常点剔除** 采用蒙特卡洛交叉验证 (Monte Carlo cross validation, MCCV) 识别异常样本。该算法

基于蒙特卡洛采样, 同时考虑了光谱值和测量参考值<sup>[12]</sup>。本研究中采样次数设置为 1 000 次, 按 3:1 划分校正集和验证集, 建立 PLSR 模型进行预测, 模型的主成分数为 10。反复随机采样建模, 得到每个样本的预测误差分布, 计算其均值和标准差, 根据样本的均值或标准差偏离主体的程度确定异常样本, 将其剔除。

**样本划分** 采用基于 X-Y 距离样本集划分 (sample set partitioning based on joint X-Y distances, SPXY) 方法按 3:1 的比例划分校正集和验证集, 校正集用于建立模型, 验证集不参与建模过程, 可作为模型外部验证。

**变量筛选** MC-UVE 是一种改进的 UVE 算法, 采用蒙特卡洛随机抽样建立大量 PLSR 模型, 计算模型回归系数平均值与标准偏差的比值作为可信度指标 (reliability index, RI), RI 越大, 表示相应的变量越为重要, 以此选取有效光谱信息<sup>[13]</sup>。MC-UVE 算法的计算速度相较 UVE 有明显优势。CARS 与 MC-UVE 基本思想相似, 以回归系数作为衡量指标。该算法基于达尔文进化论的“适者生存”原则<sup>[14]</sup>, 从蒙特卡洛抽样运行中依次选取一系列子集来构建校准模型, 同时通过指数递减函数和自适应重加权抽样的方法强制剔除变量, 选取交叉验证误差均方根 (root mean square error of cross-validation, RMSECV) 最小的子集, 即最优变量组合。变量选择可以消除无关变量, 简化模型, 构建预测能力强的模型。本研究分别采用 MC-UVE 和 CARS 在 10 000~4 000 cm<sup>-1</sup> 波长内筛选变量, 根据模型效果选择变量筛选方法。

**定量建模方法** PLSR 算法是最经典的线性回归算法, 在 NIRS 分析中被广泛应用<sup>[15-17]</sup>。该算法完美结合了典型相关分析、多元线性回归和主成分回归, 将光谱矩阵和实测参考值矩阵的相互关系考虑在内, 对两者同时进行分解, 得到最优的定量模型, 且仅需优化潜在变量数 (latent variables, LVs) 一个参数。

ELM 算法是一种改进的单隐层前馈神经网络学习算法, 以极快的学习速度提供最佳的泛化性能<sup>[18]</sup>。该算法可以解决非线性问题, 只需简单设置激活函数和隐藏节点数两个参数。本研究将激活函数设置为 sin, 隐藏层节点数为 80。

SVM 是基于统计学理论提出的一种算法。Suykens 和 Vandewalle 对经典 SVM 进行了改进, 提出了 LSSVM 算法<sup>[19]</sup>。LSSVM 用最小二乘法线性法代替经典 SVM 中的二次规划法作为损失函数, 从而简化问题, 提高了计算速度<sup>[20]</sup>。LSSVM 模型性能主要由正则化参数 (c) 和核函数参数 (g) 决定。目前, 已有许多算法被用于选择这两个参数的最佳组合, 如粒子群算法<sup>[21]</sup>。ALO 是一种自然启发式算法, 由 Mirjalili 首次提出, 模拟蚁狮

的自然捕猎行为<sup>[22]</sup>。ALO算法在全局搜索能力、避免局部最优等方面具有优势。目前已经广泛应用于工程技术、能源、电力系统优化等领域<sup>[23]</sup>,但在中药质量控制方面应用较少。本研究采用ALO优化LSSVM的参数 $c$ 和 $g$ ,用于建立定量分析模型。

基于筛选出的特征变量,分别采用PLSR、ELM、ALO-LSSVM建立定量模型。用建立的模型预测验证集,得到预测值。

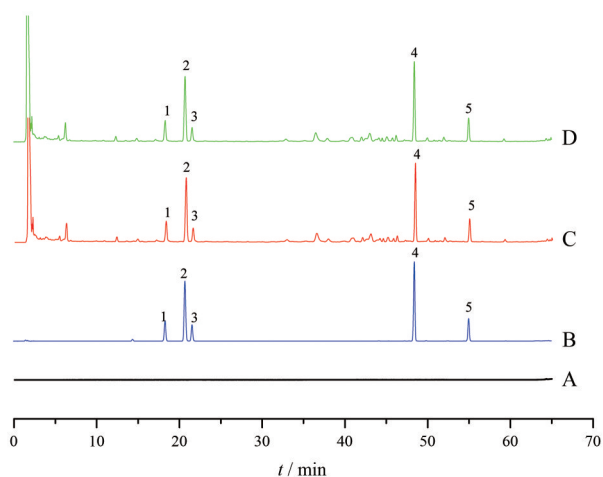
**模型评价** 校正集相关系数 (correlation coefficient of calibration set,  $R_c$ )、校正集均方根误差 (root mean square error of calibration set, RMSEC)、校正集相对标准误差 (relative standard error of calibration set, RSEC) 被用来评价校正模型的性能。验证集相关系数 (correlation coefficient of prediction set,  $R_p$ )、预测均方根误差 (root mean square error of prediction set, RMSEP) 和预测相对标准误差 (relative standard error of prediction set, RSEP)、剩余预测偏差 (residual predictive deviation, RPD) 被用来评价模型的预测能力。一般认为,相关系数 $R$ 越接近1,模型准确性越高,但 $R$ 值大小与样品浓度范围紧密相关,对于浓度范围大的样本集, $R$ 值可能接近1,但模型准确性较差,因此需要结合其他参数进行综合评价。RMSEC、RSEC越小,表明模型自测性能越好;RMSEP和RSEP越小,RPD越大,表明模型外部预测能力越好。RMSEC、RMSEP较为接近,可认为模型不存在过拟合现象。

本研究数据处理与模型建立验证在数学软件MATLAB R2021b上完成,绘图在函数绘图软件Origin Pro 8与MATLAB R2021b上共同完成。

## 结果与讨论

### 1 含量测定

**1.1 液相色谱方法学考察** 空白溶剂、混合对照品和供试品溶液的液相色谱图如图1所示,样品中其他成分及空白溶剂对测定指标成分无影响。5种皂苷成分



**Figure 1** Specificity result (A: Methanol; B: Mixed reference standards; C: Total mixed granule; D: Spray-dried fine powder). 1: Notoginsenoside R1; 2: Ginsenoside Rg1; 3: Ginsenoside Re; 4: Ginsenoside Rb1; 5: Ginsenoside Rd

的理论塔板数均大于40 000,色谱峰分离度均大于1.5,峰形对称。

6份样品中5种皂苷含量的RSD值分别为1.99%、2.04%、1.90%、1.14%、1.72%,均接近或小于2%,说明方法重复性良好。

6份样品的平均加样回收率依次为99.77%、101.7%、100.7%、97.84%、100.1%,加样回收率RSD分别为1.69%、1.54%、1.35%、0.96%、1.28%,符合药典规定。

RSD值分别为1.55%、1.68%、1.54%、1.85%、1.74%,均小于2%,说明喷干细粉的供试品溶液在24 h内稳定。

**1.2 含量测定结果** 测得106批喷干细粉和122批总混颗粒中5种皂苷及水分含量范围、平均含量及RSD如表1所示。喷干细粉与总混颗粒中指标成分的含量范围接近,为建立通用模型提供了良好的基础。

### 2 原始光谱分析

106个喷干细粉和122个总混颗粒样品的近红外光谱如图2所示。喷干细粉、总混颗粒两个中间体的

**Table 1** Results of content determination

Sample	Component	Range/%	Average/%	RSD/%
Spray-dried fine powder	Notoginsenoside R1	0.447 7-0.732 1	0.577 8	10.87
	Ginsenoside Rg1	1.585-3.291	2.186	18.15
	Ginsenoside Re	0.216 6-0.415 5	0.305 6	14.67
	Ginsenoside Rb1	1.402-2.787	1.856	14.91
	Ginsenoside Rd	0.342 3-0.604 3	0.462 8	11.96
	Moisture	4.319-7.200	5.536	10.34
Total mixed granule	Notoginsenoside R1	0.442 4-0.685 9	0.558 1	10.33
	Ginsenoside Rg1	1.618-2.695	2.136	13.46
	Ginsenoside Re	0.213 4-0.358 4	0.279 3	13.34
	Ginsenoside Rb1	1.260-2.072	1.641	13.36
	Ginsenoside Rd	0.349 3-0.567 3	0.455 0	11.42
	Moisture	4.469-7.327	5.472	8.90

光谱整体趋势相似, 合并样本后, 图谱基线增宽。腰痹通胶囊组成复杂, 光谱中包含大量物理化学信息, 谱带重叠严重, 无法直接对光谱进行分析, 因此需要结合化学计量学, 寻找合适的算法建立模型进行定量分析。

### 3 异常样本剔除与样本划分

经 MCCV 识别异常样本结果如图 3 所示, 将偏离主体较大的样本剔除。6 个指标成分分别剔除了 5、4、5、7、3、5 个异常样本。1、2、3、215、216 号样本多次被识别为异常样本, 可能是光谱异常导致。剔除异常样本后, 用 SPXY 划分校正集和验证集, 6 个指标划分的校正集和验证集样本数目分别为: 167 和 56、167 和 57、167 和 56、165 和 56、168 和 57、167 和 56。

### 4 变量筛选

MC-UVE 和 CARS 算法筛选出的三七皂苷 R1、人参皂苷 Rg1、人参皂苷 Re、人参皂苷 Rb1、人参皂苷 Rd 和水分的特征变量分布情况如图 4 所示, 以短竖线代表被筛选出的波长点, 不同颜色表示不同成分。原始光谱共包含 778 个变量, 使用 MC-UVE 提取 RI 大于 2 的光谱值, 筛选出的特征变量数分别为 172、112、181、

200、172 和 159 个, 如图 4A 所示, 5 种皂苷的特征变量分布相似, 水分的特征变量分布有明显差异; 设置 CARS 算法的采样运行次数为 100 次, 筛选出的特征变量数分别为 173、183、162、162、162 和 153, 如图 4B 所示, CARS 筛选后的变量较为分散。不同算法的筛选结果差异较大, 这与其原理不同有关。筛选出的变量分布相对分散, 可能是由于腰痹通胶囊中间体样品的组成复杂, 每个波段都存在一些有效信息。5 种皂苷含有大量 C-H、C-C 和 O-H, 其基频振动的倍频、组合频吸收峰出现在各个波段, 光谱中还存在水峰的干扰。不同方法筛选出的特征变量不同, 需要进一步建立定量模型进行比较。

### 5 模型建立与验证

基于筛选出的特征变量建立定量模型, 结果见表 2。从表中可知, 使用相同建模方法时, 基于 CARS 选择的变量建立的模型效果优于 MC-UVE。

相较于 PLSR 模型, ELM 模型的性能和预测效果均有一定提高。然而, 对于三七皂苷 R1、人参皂苷 Rd,  $R_p$  值仍小于 0.9, 与 ALO-LSSVM 相比模型效果较差。

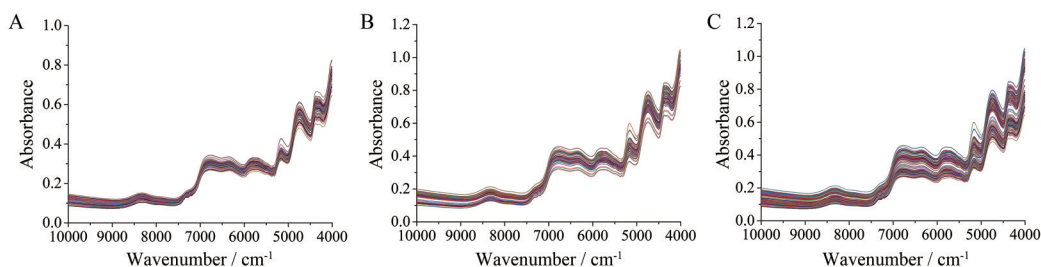


Figure 2 Raw near infrared spectra (A: Spray-dried fine powder; B: Total mixed granule; C: Intermediate set)

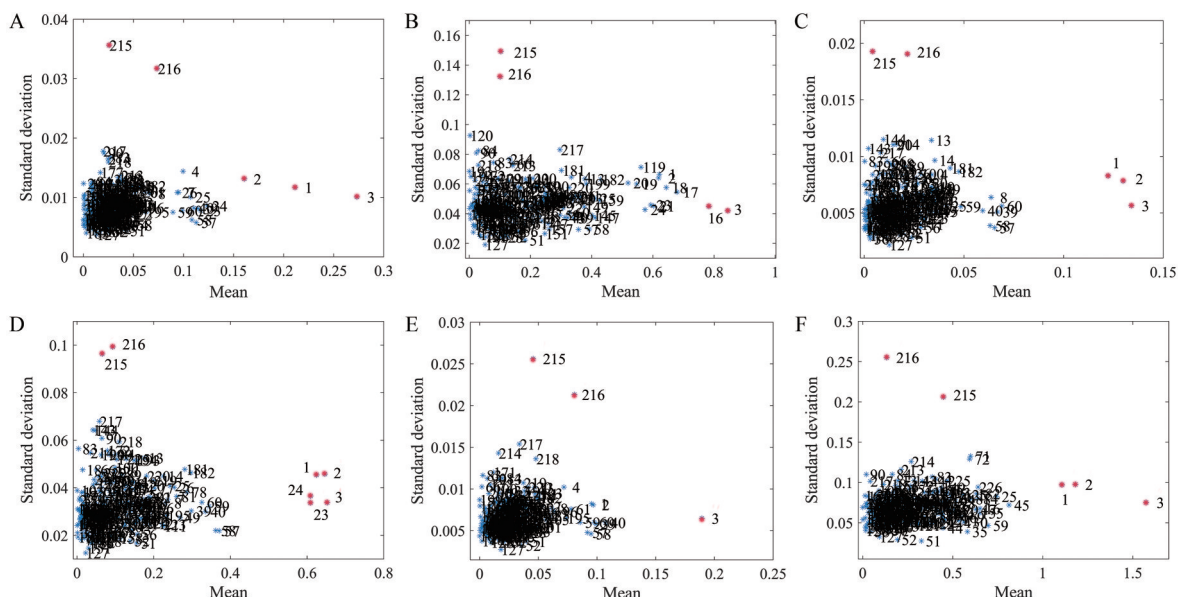
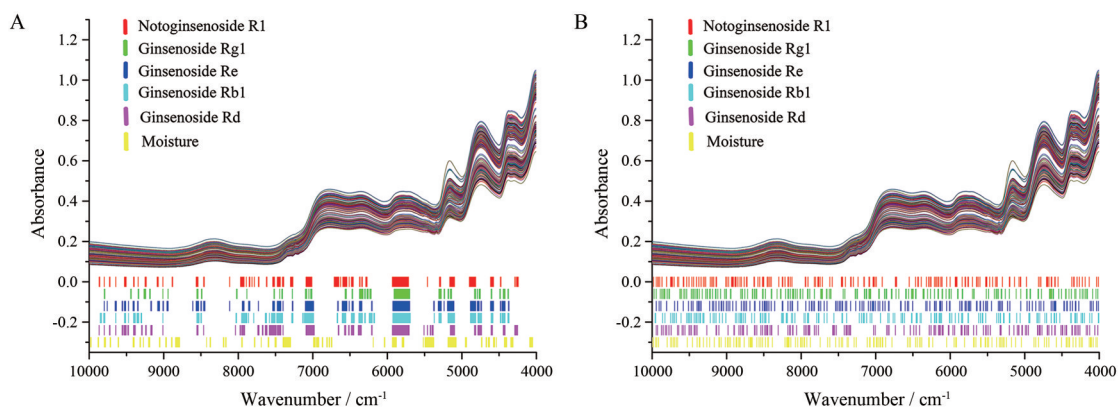


Figure 3 Mean and standard deviation distribution obtained by Monte Carlo cross validation (MCCV). A: Notoginsenoside R1; B: Ginsenoside Rg1; C: Ginsenoside Re; D: Ginsenoside Rb1; E: Ginsenoside Rd; F: Moisture



**Figure 4** The optimal subsets selected by MC-UVE (A) and CARS (B). MC-UVE: Monte Carlo uninformative variables elimination; CARS: Competitive adaptive reweighted sampling

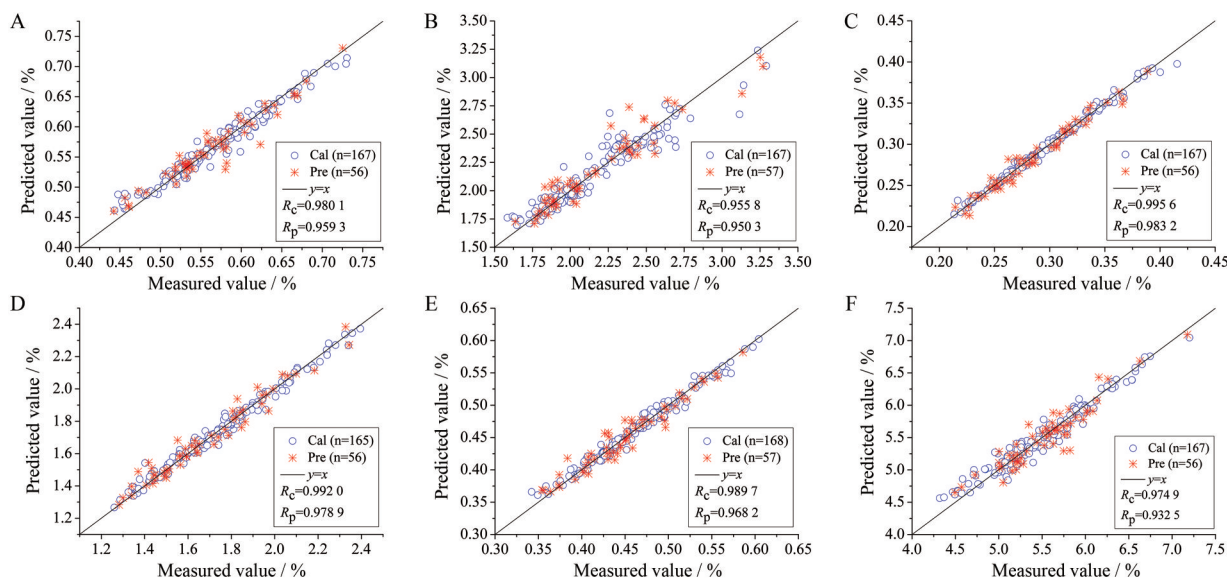
ALO-LSSVM的模型参数  $c$  和  $g$  值由 ALO 筛选得到, 6 种指标成分选择的参数差异较大。从模型性能来看, 三七皂苷 R1、人参皂苷 Rg1、人参皂苷 Re、人参皂苷 Rb1、人参皂苷 Rd、水分的 CARS-ALO-LSSVM 模型均最优,  $R_c$  值较其他两种建模方法明显提高, 依次为 0.980 1、0.955 8、0.995 6、0.992 0、0.989 7、0.974 9, 均在 0.95 以上; RMSEC、RSEC 值均有不同程度的下降, RSEC 依次为 2.05%、4.44%、1.31%、1.74%、1.61%、2.10%, 均在 5% 以内。

同时, CARS-ALO-LSSVM 模型表现出最佳的预测能力,  $R_p$  值显著提高, 依次为 0.959 3、0.950 3、0.983 2、0.978 9、0.968 2、0.932 5; 5 种皂苷模型的 RMSEP 较 PLSR 模型均降低 40% 以上, 水分降低 28%; RSEP 依次为 2.94%、5.20%、2.59%、3.02%、2.93%、3.05%, 均控制在 6% 内; 5 种皂苷模型的 RPD 较 PLSR 模型均提高

了 69% 以上, 水分提高了 38%。除此之外, RMSEP 接近 RMSEC, 说明模型不存在过拟合问题。 $R$  值大小与待测物质的浓度范围密切相关, 本研究所测的腰痹通中间体含量范围较窄, 可能导致  $R$  值低。因此, CARS-ALO-LSSVM 模型的  $R_p$  值高于 0.93 是可以接受的。

喷干细粉和总混颗粒样本合并后, 样本信息更为复杂, 经典线性算法 PLSR 在本研究中表现不佳, 难以建立符合检测要求的定量模型, LSSVM 可能更适合建立复杂系统的模型。

CARS-ALO-LSSVM 的模型预测值与实测参考值的相关性如图 5 所示, 可以更直观地看到模型的预测结果。样本点均匀分布在  $y=x$  两侧附近, 表明预测值与实测参考值较为接近, 建立的模型能够满足腰痹通中间体定量分析的需求。



**Figure 5** Correlation diagram between measured values and predicted values (A: Notoginsenoside R1; B: Ginsenoside Rg1; C: Ginsenoside Re; D: Ginsenoside Rb1; E: Ginsenoside Rd; F: Moisture). Cal: Calibration set; Pre: Prediction set

**Table 2** Comparison of model performances for Yaobitong intermediates. PLSR: Partial least squares regression; ELM: Extreme learning machine; ALO-LSSVM: Ant lion optimization least squares support vector machine;  $R_c$ : Correlation coefficient of calibration set; RMSEC: Root mean square error of calibration set; RSEC: Relative standard error of calibration set;  $R_p$ : Correlation coefficient of prediction set; RMSEP: Root mean square error of prediction set; RSEP: Relative standard error of prediction set; RPD: Residual predictive deviation

Component	Variable selection	Variable number	Method	Parameter	Calibration set			Prediction set			
					$R_c$	RMSEC	RSEC/%	$R_p$	RMSEP	RSEP/%	RPD
Notoginsenoside R1	MC-UVE	172	PLSR	LVs=15	0.829 1	0.032 9	5.77	0.827 7	0.033 1	5.84	1.78
			ELM	-	0.906 5	0.024 8	4.35	0.867 6	0.029 3	5.18	2.01
			ALO-LSSVM	c=3 738.54 g=64.08	0.949 6	0.018 4	3.23	0.910 1	0.024 4	4.31	2.42
	CARS	173	PLSR	LVs=12	0.869 8	0.029 0	5.09	0.858 4	0.030 2	5.34	1.95
			ELM	-	0.900 8	0.025 6	4.48	0.880 7	0.027 9	4.93	2.13
			ALO-LSSVM	c=25.19 g=2.93	0.980 1	0.011 7	2.05	0.959 3	0.016 6	2.94	3.54
Ginsenoside Rg1	MC-UVE	112	PLSR	LVs=15	0.831 5	0.182 3	8.38	0.814 3	0.214 4	9.70	1.72
			ELM	-	0.913 8	0.133 3	6.13	0.895 6	0.164 3	7.43	2.25
			ALO-LSSVM	c=931 192.50 g=897.19	0.918 9	0.129 5	5.95	0.912 7	0.150 9	6.83	2.45
	CARS	183	PLSR	LVs=14	0.869 9	0.161 9	7.44	0.853 2	0.192 6	8.71	1.92
			ELM	-	0.928 5	0.121 8	5.60	0.911 0	0.152 3	6.89	2.43
			ALO-LSSVM	c=4 400.65 g=76.40	0.955 8	0.096 5	4.44	0.950 3	0.114 9	5.20	3.25
Ginsenoside Re	MC-UVE	181	PLSR	LVs=16	0.902 5	0.017 9	6.06	0.897 4	0.018 1	6.27	2.28
			ELM	-	0.947 3	0.013 3	4.51	0.927 9	0.015 3	5.30	2.70
			ALO-LSSVM	c=3 359.65 g=54.48	0.976 0	0.009 0	3.07	0.965 9	0.010 6	3.68	3.87
	CARS	162	PLSR	LVs=15	0.950 3	0.012 9	4.38	0.925 1	0.015 6	5.40	2.68
			ELM	-	0.960 6	0.011 5	3.91	0.952 3	0.012 5	4.34	3.30
			ALO-LSSVM	c=41 524.48 g=59.09	0.995 6	0.003 9	1.31	0.983 2	0.007 5	2.59	5.56
Ginsenoside Rb1	MC-UVE	200	PLSR	LVs=15	0.912 3	0.099 1	5.64	0.871 3	0.123 9	7.25	2.07
			ELM	-	0.941 9	0.081 3	4.63	0.914 3	0.102 2	5.98	2.48
			ALO-LSSVM	c=27 373.06 g=177.48	0.970 8	0.058 0	3.30	0.953 6	0.076 0	4.45	3.34
	CARS	162	PLSR	LVs=14	0.944 1	0.079 8	4.54	0.921 8	0.097 9	5.73	2.59
			ELM	-	0.949 1	0.076 2	4.34	0.930 0	0.092 8	5.43	2.75
			ALO-LSSVM	c=92.67 g=4.74	0.992 0	0.030 5	1.74	0.978 9	0.051 6	3.02	4.92
Ginsenoside Rd	MC-UVE	172	PLSR	LVs=14	0.870 1	0.025 8	5.55	0.802 7	0.031 8	6.99	1.72
			ELM	-	0.919 8	0.020 5	4.42	0.884 0	0.024 9	5.48	2.17
			ALO-LSSVM	c=137.98 g=5.89	0.986 4	0.008 6	1.85	0.968 0	0.013 4	2.94	3.98
	CARS	162	PLSR	LVs=13	0.898 0	0.023 0	4.95	0.860 9	0.027 1	5.96	2.02
			ELM	-	0.909 7	0.021 7	4.67	0.897 7	0.023 5	5.16	2.29
			ALO-LSSVM	c=180.61 g=5.90	0.989 7	0.007 5	1.61	0.968 2	0.013 3	2.93	3.99
Moisture	MC-UVE	159	PLSR	LVs=15	0.851 3	0.271 5	4.94	0.809 8	0.275 5	4.96	1.72
			ELM	-	0.926 8	0.194 4	3.54	0.910 2	0.194 5	3.50	2.51
			ALO-LSSVM	c=440 390.58 g=1 811.50	0.926 8	0.194 3	3.54	0.885 4	0.218 3	3.93	2.18
	CARS	153	PLSR	LVs=15	0.924 1	0.197 7	3.60	0.864 1	0.236 4	4.26	2.02
			ELM	-	0.911 0	0.213 4	3.88	0.909 2	0.195 6	3.52	2.47
			ALO-LSSVM	c=659 861.10 g=725.83	0.974 9	0.115 1	2.10	0.932 5	0.169 6	3.05	2.79

### 结论

本研究应用NIRS结合化学计量学方法对腰痹通胶囊生产过程的两种关键中间体开展了快速检测研究, 合并喷干细粉和总混颗粒样本建立了中间体质量检测通用模型。ALO-LSSVM算法在建立通用模型过程中表现出比PLSR算法和ELM算法更好的性能, 模

型效果显著提升,  $R_c$ 值均大于0.95,  $R_p$ 值均大于0.93, RSEC和RSEP分别控制在5%和6%以内, 表明ALO-LSSVM在中药质量控制方面具有良好潜力, 所建立的定量模型能够准确预测腰痹通中间体中6个关键指标成分的含量, 满足腰痹通胶囊生产过程快速检测的需求。本研究可为NIRS结合ALO-LSSVM算法在其他

中药多工段检测应用提供参考。

**作者贡献:** 斯乐婷为第一作者, 负责实验设计和实施、数据分析和稿件编写; 张欣、张永超、张江艳、王钧负责样品收集与提供、文献研究、稿件修改; 陈勇、刘雪松提出实验思路; 通讯作者吴永江全面负责实验指导、稿件修改。

**利益冲突:** 所有作者均声明不存在利益冲突。

## References

- [1] Chinese Pharmacopoeia Committee. Pharmacopoeia of the People's Republic of China (Part 1) (中华人民共和国药典, 一部) [S]. Beijing: China Medical Science Press, 2020: 1805-1807.
- [2] Xin J, Ji ZJ, Xu HH, et al. Application value of Yaobitong capsule combined with celixib in relieving lumbar intervertebral disc degeneration and its effect on inflammatory factors [J]. Chin Arch Tradit Chin Med (中华中医药学刊), 2021, 39: 210-213.
- [3] Fan ZY, Zhang XZ, Ke ZP, et al. Network-assisted mechanism investigation of Yaobitong capsules on lumbar intervertebral discs [J]. Chin Tradit Pat Med (中成药), 2018, 40: 275-283.
- [4] Ge MD, He WQ. Clinical effect of Yaobitong capsule combined with naproxen tablets on patients with osteochondritis of lumbar endplate [J]. Chin Tradit Pat Med (中成药), 2019, 41: 2269-2271.
- [5] Jin Y, Du WJ, Liu XS, et al. Application of near infrared spectroscopy and real time release testing combined with statistical process control charts for on-line quality control of industrial concentrating process of traditional Chinese medicine "Jinyinhua" [J]. Infrared Phys Technol, 2022, 123: 104135.
- [6] Liu RH, Sun QF, Hu T, et al. Multi-parameters monitoring during traditional Chinese medicine concentration process with near infrared spectroscopy and chemometrics [J]. Spectrochim Acta Part A, 2018, 192: 75-81.
- [7] Li XM, Xu FF, Zhang X, et al. Research on prediction model of intermediate content in Jinzhen Oral Liquid based on near infrared and mid infrared spectroscopy technology [J]. Chin Tradit Herb Drugs (中草药), 2023, 54: 8007-8017.
- [8] Ma PP, Wu CL, Zhang WY, et al. Rapid detection of moisture content in intermediate Suhuang granules of Suhuang Zhike capsules by near infrared spectroscopy [J]. Cent South Pharm (中南药学), 2024, 22: 1042-1047.
- [9] Li M, Zhang CH, Liu CL, et al. Determination of moisture in the intermediate powder of Gulong Capsules by near-infrared spectroscopy [J]. Drugs Clin (现代药物与临床), 2019, 34: 2280-2282.
- [10] Tao J, Li SM, Zhan CS. Rapid prediction of emodin in Danning Tablet intermediate by model established by near infrared spectroscopy with partial least square method [J]. Phys Test Chem Anal Part B (理化检验—化学分册), 2023, 59: 833-836.
- [11] Si LT, Ni HF, Li Q, et al. Simultaneous determination of five saponins in Yaobitong capsule by HPLC-CAD [J]. Acta Pharm Sin (药学报), 2020, 55: 2449-2453.
- [12] Liu ZC, Cai WS, Shao XG. Monte Carlo cross-validation is used to identify exotic samples in near infrared spectra [J]. Sci Sin Chim (中国科学: 化学), 2008, 38: 316-323.
- [13] Cai WS, Li YK, Shao XG. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. Chemometr Intell Lab Syst, 2008, 90: 188-194.
- [14] Li HD, Liang YZ, Xu QS, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. Anal Chim Acta, 2009, 648: 77-84.
- [15] Zhang J, Guo Z, Wang SH, et al. Comparison of methods for water content in rice by portable near-infrared and visible light spectrometers [J]. Spectrosc Spectral Anal (光谱学与光谱分析), 2023, 43: 2059-2066.
- [16] Liu YD, Liao J, Li B, et al. Robustness of global model of soluble solids in Gongli pear based on near-infrared spectroscopy [J]. Spectrosc Spectral Anal (光谱学与光谱分析), 2022, 42: 2781-2787.
- [17] Sun P, Xiang CQ, Chen QW, et al. Nondestructive determination of ephedrine alkaloids and morphine in Keke tablets by near infrared spectroscopy [J]. J Instrum Anal (分析测试学报), 2024, 43: 637-639.
- [18] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks [C]//2004 IEEE International Joint Conference on Neural Networks. Budapest: IEEE, 2004: 985-990.
- [19] Suykens JAK, Vandewalle J. Least squares support vector machine classifiers [J]. Neural Process Lett, 1999, 9: 293-300.
- [20] Wang HF, Hu DJ. Comparison of SVM and LS-SVM for regression [C]//2005 International Conference on Neural Networks and Brain. Beijing: IEEE, 2005: 279-283.
- [21] Li DS, Li LN, Zhang RC. Water quality pH value determination for visible-near infrared spectroscopy based on SPA and PSO-LSSVM [J]. Laser Optoelectron Prog (激光与光电子学进展), 2023, 60: 390-395.
- [22] Mirjalili S. The ant lion optimizer [J]. Adv Eng Software, 2015, 83: 80-98.
- [23] Hu C, Cai YG, Huang JC, et al. A review of ant lion optimization algorithms research [J]. Autom Inf Eng (自动化与信息工程), 2024, 45: 1-10.