

引用:殷文俊,彭攀,陈星,李杰,张慧.基于高光谱成像技术对甘草饮片中甘草苷和甘草酸含量快速无损检测[J].中医导报,2026,32(5):61-68.

## 基于高光谱成像技术对甘草饮片中甘草苷和甘草酸含量快速无损检测\*

殷文俊<sup>1</sup>,彭攀<sup>1</sup>,陈星<sup>1</sup>,李杰<sup>1</sup>,张慧<sup>2</sup>

(1.株洲市中医伤科医院,湖南 株洲 412000;

2.浙江工业大学药学院,浙江 杭州 310014)

[摘要] 目的:利用高光谱成像技术结合预处理算法和特征波段筛选方法建立模型,对甘草饮片中甘草苷和甘草酸含量快速无损监测。方法:本研究利用高光谱成像技术对短波红外波段下的甘草饮片进行采集,通过高效液相色谱获取甘草饮片的甘草苷和甘草酸含量信息作为参考值,与样品光谱信息相对应。对原始光谱数据进行预处理以消除仪器的噪声干扰,依据甘草苷和甘草酸含量信息从特征波长和特征波段两个方面筛选特征波长,分别建立甘草苷和甘草酸含量的偏最小二乘回归(PLSR)、多元线性回归(MLR)和贝叶斯岭回归(BRR)定量校正模型,用预测集决定系数( $R_p^2$ )和预测集误差均方根(RMSEP)作为模型性能的评价指标。结果:高光谱甘草苷和甘草酸定量模型最优预处理方法分别为Savitzky-Golay平滑和SNV;通过对特征波段方法的筛选,CARS方法能够筛选得到与甘草苷和甘草酸含量最为相关的波长;针对甘草苷含量建立的PLSR模型和甘草酸含量建立的BRR模型对含量预测具有最优精度,模型预测集决定系数( $R_p^2$ )分别为0.958 7和0.897 3,预测集误差均方根(RMSEP)分别为0.001 8和0.003 8,表明所建立含量预测模型具有可靠的稳定性与准确性。结论:高光谱成像技术可实现甘草中甘草苷和甘草酸成分含量的快速有效检测,为中药甘草品质的快速无损检测提供了一种新的途径。

[关键词] 高光谱成像;甘草;甘草苷;甘草酸;无损检测

[中图分类号] R284.1 [文献标识码] A [文章编号] 1672-951X(2026)05-0061-08

DOI: 10.13862/j.cn43-1446/r.2026.05.010

### Nondestructive Detection of the Content of Liquirtin and Glycyrrhizic Acid in Licorice Based on Hyperspectral Imaging Technology

YIN Wenjun<sup>1</sup>, PENG Pan<sup>1</sup>, CHEN Xing<sup>1</sup>, LI Jie<sup>1</sup>, ZHANG Hui<sup>2</sup>

(1.The Orthopedics Hospital of Traditional Chinese Medicine Zhuzhou City, Zhuzhou Hunan 412000, China;

2.College of Pharmaceutical Sciences, Zhejiang University of Technology, Hangzhou Zhejiang 310014, China)

[Abstract] Objective: To establish the mathematical models for a rapid and nondestructive analytical method to monitor the content of liquirtin and glycyrrhizic acid in licorice. Methods: The contents of liquirtin and glycyrrhizic acid in licorice determined by High Performance Liquid Chromatography were set as the reference values, which associated with licorice spectra by using hyperspectral imaging technology in short-wave infrared ranges. In order to eliminate the possible instrument noise during the hyperspectral acquisition, the preprocessing methods of the spectra were carried out. Using the correlation between the content of liquirtin and glycyrrhizic acid and the reflectance data to select the effective band, quantitative calibration models were developed by partial least squares regression (PLSR), multiple linear regression (MLR) and bayesian ridge regression (BRR), and evaluated by the metrics of coefficients of determination of prediction ( $R_p^2$ ) and the root mean square errors of prediction (RMSEP). Results: The results demonstrated that the optimal preprocessing methods for hyperspectral-based quantification of liquirtin and glycyrrhizic acid were Savitzky-Golay smoothing and standard normal variate transformation (SNV), respectively. Through feature band screening, the competitive adaptive reweighted sampling (CARS) algorithm effectively identified wavelengths exhibiting the strongest correlation

\*基金项目:湖南省中医药科研计划项目(B2024136)

通信作者:张慧,女,副教授,研究方向为中药质量控制与药物分析

with liquirtin and glycyrrhizic acid contents. Specifically, PLSR model developed for liquirtin content prediction and BRR model for glycyrrhizic acid content prediction achieved the highest predictive accuracy, with  $R_p^2$  of 0.958 7 and 0.897 3, and  $RMSEP$  of 0.001 8 and 0.003 8. These findings indicate that the established content prediction models exhibit robust stability and accuracy. Conclusion: Hyperspectral imaging technology can achieve rapid and effective detection of the content of liquirtin and glycyrrhizic acid in licorice, which provides a new approach to the rapid and nondestructive detection of the quality of licorice.

[Keywords] hyperspectral imaging; licorice; liquirtin; glycyrrhizic acid; nondestructive detection

甘草具有很高的药用价值,被称为“药中国老”。甘草中的主要活性成分为甘草苷和甘草酸,2020年版《中华人民共和国药典》将甘草苷和甘草酸作为甘草的含量测定指标<sup>[1]</sup>,其中甘草苷具有抗心律失常、抗肿瘤作用;甘草酸能够在体内发挥抗炎、抗肿瘤、抗病毒的功效<sup>[2]</sup>。目前,甘草中有效成分含量的测定方法主要采用化学分析方法,如薄层色谱、高效毛细管电泳和高效液相色谱等<sup>[3]</sup>,检测过程中耗时费力,甘草的外观被破坏,容易产生化学污染并且检测成本高。高光谱成像技术将光谱信息与图像数据相结合,从而在获取图像的同时,也能得到图像中每个像素点的光谱信息,可实现原型样品光谱信息的批量同时采集,不需有机试剂提取化学成分,具有快速、无损、高效环保等优势<sup>[4-5]</sup>。目前高光谱技术已经在中药真伪鉴别、基源鉴定和含量测定等方面进行了初步的研究与应用。张蕾等<sup>[6]</sup>针对葛根与葛藤以藤代替根入药的掺伪现象,利用高光谱成像技术实现葛根和葛藤的快速、无损鉴别。潘亿瑕等<sup>[7]</sup>通过数据挖掘方法充分利用高光谱的光谱和图像数据,将光谱特征和纹理特征进行有机融合,为甘草药材的基源鉴别提供了一种新型无损的检测方法。江恩赐等<sup>[8]</sup>基于高光谱成像融合CARS特征波段筛选快速检测硫熏牛膝 $SO_2$ 含量。

本研究利用高光谱成像技术对短波红外波段下的甘草饮片进行采集,通过液相色谱测定所采集样本的甘草苷和甘草酸含量,对原始光谱进行预处理排除干扰后,依据含测数据提取特征波长,分别建立基于特征波长的定量分析模型,比较不同特征变量提取方法和定量校正模型的优劣,确定最优预测模型,从而实现甘草饮片中甘草苷和甘草酸含量快速无损检测。

## 1 材料与方法

### 1.1 材 料

1.1.1 药物与试剂 实验用甘草药材收集自内蒙古(42批次:GC1912001,GC1912002,GC1912003,GC1912004,GC1912005,GC1912006,GC1912007,GC1912008,GC1912009,GC1912010,GC1912011,GC1912012,GC1912013,GC1912014,GC1914001,GC1914002,GC1914003,GC1914004,GC1914005,GC1914006,GC1914007,GC1914008,GC1914009,GC1914010,GC1914011,GC1914012,GC1914013,GC1914014,ganc-01gc,ganc-02gc,ganc-03gc,ganc-04gc,ganc-05gc,ganc-09gc,ganc-10gc,ganc-11gc,ganc-12gc,ganc-13gc,GC171109,GC171110,

GC190201,GC190202),甘肃(34批次:GC1912015,GC1912016,GC1912017,GC1912018,GC1912019,GC1912020,GC1912021,GC1912022,GC1912023,GC1912024,GC1912025,GC1912026,GC1912027,GC1912028,GC1912029,GC1914015,GC1914016,GC1914017,GC1914018,GC1914019,GC1914020,GC1914021,GC1914022,GC1914023,GC1914024,GC1914025,GC1914026,GC1914027,GC1912028,GC1914029,ganc-06gc,ganc-07gc,ganc-17gc,ganc-18gc)和新疆(32批次:GC1912030,GC1912031,GC1912032,GC1912033,GC1912034,GC1912035,GC1912036,GC1912037,GC1912038,GC1912039,GC1912040,GC1912041,GC1914030,GC1914031,GC1914032,GC1914033,GC1914034,GC1914035,GC1914036,GC1914037,GC1914038,GC1914039,GC1914040,GC1914041,ganc-08gc,ganc-15gc,ganc-16gc,ganc-19gc,ganc-20gc,ganc-21gc,ganc-20180401,ganc-20180402),共108批次,均购自安徽亳州药材市场安徽九州方圆制药有限公司。所有收集甘草样品经浙江工业大学药学院张慧副教授鉴定为豆科植物甘草(*Glycyrrhiza uralensis* Fisch.)、胀果甘草(*Glycyrrhiza inflata* Bat.)或光果甘草(*Glycyrrhiza glabra* L.)的干燥根和根茎。乙腈(美国TEDIA公司,色谱级HPLC,批号:20085249);甲醇(美国TEDIA公司,色谱级HPLC,批号:20085008);磷酸(上海凌峰化学试剂有限公司,分析级AR,批号:20170623);无水乙醇(上海凌峰化学试剂有限公司,分析纯AR,批号:20201222);甘草苷对照品(中国食品药品检定研究院,批号:111610-201607);甘草酸对照品(中国食品药品检定研究院,批号:110731-201619)。

1.1.2 主要仪器 高效液相色谱仪(美国Agilent公司,型号:Agilent 1290);1/10万分析天平(德国Sartorius公司,型号:CPA225D);高速离心机(德国Eppendorf公司,型号:Centrifuge Mini);高速多功能粉碎机(永康市久品工贸有限公司,型号:JP-100A);数控超声波清洗器(巩义市予华仪器有限公司,型号:KQ-250DE);试验标准筛(上虞市南仪器厂,型号:50目)。高光谱成像系统由成像模块、移动平台、光源、高性能计算机与图像采集软件等部分组成。成像模块包含近红外光谱相机(英国Raptor Photonics公司,型号:OWL-640-mini);近红外成像镜头(芬兰SPECIM公司,型号:OLES22.5);成像光谱仪(芬兰SPECIM公司,型号:ImSpector-V10E);电控移动平台(东佑达机器人有限公司,型号:ETH14);光源:两个150 W石英卤素

灯(美国Illumination Technologies公司,型号:3900ER);图像采集软件(台湾五铃光学公司,Spectral Image软件)。

## 1.2 方 法

1.2.1 高光谱图像采集 高光谱图像采集前将高光谱成像仪开机预热30 min。将108批次甘草饮片,每批次选用30片放在移动平台上,甘草饮片的厚度和大小不限,在短波红外波段下进行采集。为保证所采集的高光谱图像清晰,通过调整样品和物镜之间的距离、平台移动速度、相机曝光时间,最终参数设置为:样品与物镜之间距离为29 cm,电控移动平台速度为2.6 mm/s,相机曝光时间为24 ms。获得高光谱数据立方体,其宽为640像素,长为1 350像素,光谱范围为898~1751 nm,间隔为1.67 nm,共512个波长,像素点大小为0.23 mm×0.23 mm。

### 1.2.2 HPLC含量测定

1.2.2.1 色谱条件 参照2020年版《中华人民共和国药典》高效液相色谱法(通则0512)测定。Agilent ZorBax SB C<sub>18</sub>(4.6 mm×250.0 mm, 5.0 μm)色谱柱;柱温25 ℃;流动相A为乙腈,B为0.05%磷酸水;梯度洗脱(0~8 min, 19%A; 8~35 min, 19%→50%A; 35~36 min, 50%→100%A; 36~40 min, 100%→19%A);流速为1.0 mL/min;检测波长为237 nm;进样量10 μL。

1.2.2.2 对照品溶液的制备 精密称定甘草苷和甘草酸铵对照品适量,加70%乙醇分别制成每1 mL含甘草苷20 μg、甘草酸铵0.2 mg的溶液,过0.22 μm微孔滤膜,即得对照品溶液,其中甘草酸质量=甘草酸铵质量/1.020 7。

1.2.2.3 供试品溶液的制备 将经高光谱拍摄后的甘草饮片,粉碎后过50目筛,过筛后精密称定甘草粉末0.2 g,置于具塞锥形瓶中,精密加入70%乙醇100 mL,密塞,称定质量,超声处理(功率250 W,频率40 kHz)30 min,室温放冷后,再称定质量,用70%乙醇补足减失的质量,摇匀,滤过,取续滤液,过0.22 μm微孔滤膜即得。

1.2.2.4 测定法 分别精密吸取对照品和供试品溶液各10 μL,注入液相色谱仪,测定,即得。

1.2.3 高光谱图像黑板校正 长时间使用电荷耦合器件后,仪器会发热而产生暗电流,同时不稳定的照明源会对采集得到的图像信息产生影响<sup>[9]</sup>。为消除这部分的影响,对采集得到的高光谱图像进行黑板校正。从极端照明情况下获取两个参考图像:从具有100%反射率的聚四氟乙烯材料标准白板获取得到白板参考图像;从无光源并盖上镜头盖条件下获取黑板参考图像。将后续获取的高光谱图像均用黑板参考图像进行如下校准<sup>[10]</sup>: $R_c = \frac{R_c - R_{\text{dark}}}{R_{\text{white}} - R_{\text{dark}}}$ 。其中, $R_c$ 是校准的反射率图像, $R_0$ 是原始反射率图像, $R_{\text{white}}$ 是白板参考图像, $R_{\text{dark}}$ 是黑板参考图像。

1.2.4 感兴趣区域提取 目前许多高光谱研究均通过ENVI软件中感兴趣区域(region of interest, ROI)工具箱来手动获取样品ROI图像,选取过程中样品轮廓大小难以确定,并且需要对每个样品ROI图像进行逐个保存,该方法在消耗大量时间的同时所获取的感兴趣区域不够完整。因此,将黑板校正后的图像,通过二进制阈值分割识别样品边缘,生成二进制图像,将二值图像应用于原始高光谱图像,利用Python中

“find Contours”函数自动提取所有甘草饮片的完整高光谱ROI图像。同时对短波红外波段范围内的每个ROI图像中所有像素点的光谱反射率求平均,可以获得每批次药材的平均光谱数据。将提取得到的108个高光谱样本,按照Kennard–Stone算法将样本分成校正集和预测集,其中校正集75个样本,测试集33个样本。

1.2.5 原始光谱预处理 为了剔除一些由仪器产生的无关干扰信息,增强甘草苷和甘草酸成分光谱吸收强度,建立稳定、可靠的模型<sup>[11]</sup>。使用Python程序语言编写预处理算法,将原始光谱经过多种预处理方法后,分别与液相得到的甘草苷和甘草酸含量信息一一对应,并作为输入变量,建立PLSR定量校正模型,评估模型性能,从而筛选得到最合适的预处理方法。

1.2.6 提取光谱特征波段 针对光谱信息重叠严重,模型难以捕捉到有效的甘草苷和甘草酸含量信息。直接应用原始光谱建立得到的定量模型预测能力较差,需要进行适当的特征波段筛选<sup>[12]</sup>。针对甘草苷和甘草酸的含量信息,提取相应的特征波长或者特征波段,实现在对模型进行降维的同时更能准确地反映含量信息<sup>[13]</sup>。将各批次经过预处理后的光谱分别与液相所得的甘草苷和甘草酸含量信息一一对应后,在Matlab软件中进行特征变量的筛选,将得到的特征波段作为输入变量,分别建立甘草苷和甘草酸的定量校正模型。

1.2.7 模型的评价指标 使用定量校正模型验证指标来考察模型性能,将交叉验证集决定系数( $R^2_{\text{cv}}$ )和交叉验证误差均方根(RMSECV)作为模型优化的参考指标,校正集决定系数( $R^2_c$ )、校正集误差均方根(RMSEC)、预测集决定系数( $R^2_p$ )和预测集误差均方根(RMSEP)作为模型性能的评价指标<sup>[14]</sup>。其中, $R^2_p$ 越接近1, RMSEP越小,模型的准确度和预测能力越好。

1.2.8 数据处理 本研究采用Spyder(Python 3.7)进行感兴趣区域提取,预处理,特征波长提取和定量模型的构建等数据处理分析操作;采用Matlab 2017a软件进行特征波段的筛选操作。

## 2 结 果

2.1 模型样本划分 通过Kennard–Stone算法<sup>[15]</sup>以7:3的比例划分成校正集和预测集。其中校正集75个样本,预测集33个样本。样本划分结果见表1。预测集的浓度均包含在校正集范围中,校正集和预测集的标准偏差与总样本之间无明显差异。

表1 校正集和预测集的样本划分结果

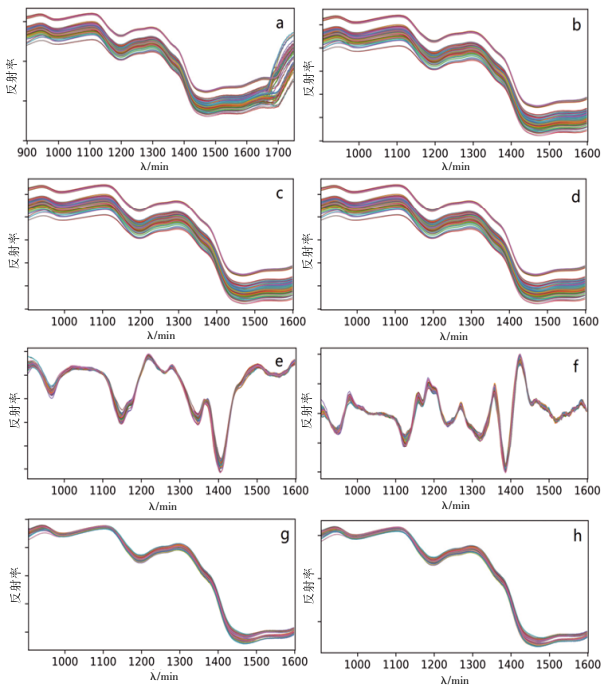
指标成分	样品组	样品数	浓度范围/(mg/mL)	平均值/(mg/mL)	标准偏差
甘草苷	校正集	75	0.001 7~0.043 0	0.012 0	0.008 8
	预测集	33	0.003 1~0.039 9	0.013 8	0.009 7
	总样本	108	0.001 7~0.043 0	0.012 6	0.009 1
甘草酸	校正集	75	0.035 5~0.104 1	0.059 4	0.012 7
	预测集	33	0.036 0~0.075 2	0.055 9	0.010 8
	总样本	108	0.035 5~0.104 1	0.058 3	0.012 2

2.2 光谱预处理方法的选择 样本平均化后的原始光谱如图1(a)所示,去除图谱中898~906 nm和1 600~1 751 nm波段前后噪声干扰比较大波段,得到906~1 600 nm(共416个波段)的光谱数据进行建模。将原始光谱经SG平滑,一阶导数(D1),

二阶导数(D2)、SNV、MSC预处理后,分别与HPLC得到的甘草苷和甘草酸含量信息建立PLSR定量校正模型,不同预处理方法的模型性能参数见表2。

表2 不同预处理方法对PLSR定量模型性能的影响

预处理方法	甘草苷				甘草酸			
	校正集		预测集		校正集		预测集	
	$R^2_c$	RMSEC	$R^2_p$	RMSEP	$R^2_c$	RMSEC	$R^2_p$	RMSEP
原始光谱	0.988 6	0.001 0	0.909 8	0.002 6	0.907 5	0.003 6	0.795 9	0.005 3
SG(9点)	0.986 9	0.001 0	0.923 4	0.002 4	0.940 3	0.002 9	0.801 6	0.005 3
SG(13点)	0.987 4	0.001 0	0.927 9	0.002 3	0.934 8	0.003 0	0.793 6	0.005 4
SG(17点)	0.986 3	0.001 0	0.932 1	0.002 2	0.929 5	0.003 1	0.782 1	0.005 5
SG(21点)	0.988 0	0.001 0	0.929 9	0.002 3	0.925 1	0.003 2	0.772 9	0.005 6
D1	0.993 0	0.000 7	0.897 5	0.002 8	0.919 4	0.003 4	0.837 0	0.004 8
D2	0.996 2	0.000 6	0.870 0	0.003 1	0.928 3	0.003 2	0.824 3	0.005 0
MSC	0.990 6	0.000 9	0.857 8	0.003 1	0.934 9	0.003 0	0.819 2	0.005 0
SNV	0.990 8	0.000 9	0.884 6	0.002 8	0.936 3	0.003 0	0.839 5	0.004 8



注:a.原始光谱;b.SG-9点光谱;c.SG-17点光谱;d.SG-21点光谱;e.一阶导数光谱;f.二阶导数光谱;g.MSC光谱;h.SNV光谱。

图1 甘草饮片光谱图

表2结果显示,预处理前后定量校正模型性能会发生变化。对于甘草苷PLSR定量校正模型来说,除SG平滑和一阶导数方法外,其他预处理方法均会对模型性能造成负面影响。这可能是由于预处理后,与甘草苷含量信息相关信号被当成噪声而改变,降低了模型的预测能力。其中,SG-17点平滑对模型性能提升最高,预测集决定系数达到了0.932 1,后续选用作为甘草苷定量校正模型的预处理方法,光谱图如图1(c)所示。对于甘草酸PLSR定量校正模型来说,除二阶导数方法外,其他预处理方法均相对于原始光谱有一定程度的提升。这说明预处理方法能有效排除背景干扰,提升模型性能。其中,SNV方法提升幅度最高,将其作为甘草酸定量校正模型的预

处理方法,光谱图如图1(h)所示。

2.3 提取特征波段方法的筛选 从特征波长和特征波段两方面入手进行筛选:针对特征波长点的筛选主要采用竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)、连续投影算法(successive projection algorithm, SPA)和随机蛙跳算法(shuffled frog leaping algorithm, SFLA);针对特征波段的筛选主要采用联合区间偏最小二乘法(synergy interval partial least squares, siPLS)和反向区间偏最小二乘法(backward interval partial least squares, biPLS)。

2.3.1 SPA法 SPA法只需少量的变量进行建模,便能输出各个变量的选择概率,完成对相关变量的筛选<sup>[16]</sup>。在甘草苷定量校正模型中,SPA算法设定波长数在1到30之间,SPA运行结果见图2(a),当模型的变量数为18时,RMSECV可以达到较低的水平,得到的特征波长在全光谱中的分布见图2(b),提取得到波长为[1 258.4、1 129.8、1 280.1、1 213.3、1 569.0、1 380.3、1 156.5、934.4、952.8、1 063.0、1 402.0、1 435.4、1 365.3、1 497.2、1 452.1、1 597.4、991.2、1 178.2 nm]。

在甘草酸定量校正模型中,设定波长数在1~30之间,SPA运行结果见图2(c),当模型的变量数为17时,RMSECV值较低,此时特征波长建立模型性能较好,提取得到的特征波长在全光谱中的分布见图2(d),提取得到波长为[1 375.3、1 388.6、906.0、1 089.7、1 154.8、1 220.0、1 131.5、932.7、1 166.5、1 298.8、1 335.2、994.5、1 408.7、1 098.1、1 183.2、1 440.4、1 517.2 nm]。

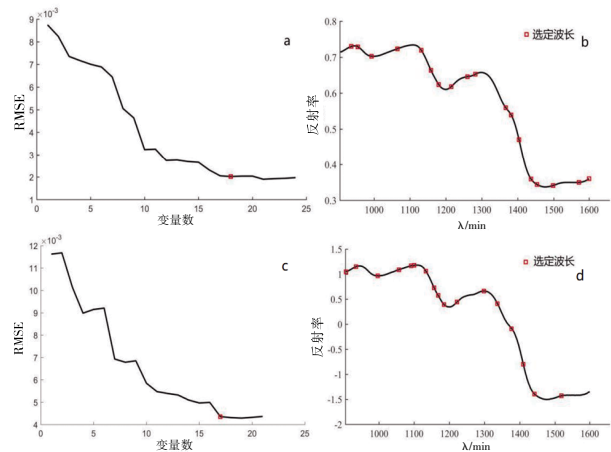


图2 甘草苷(a,b)和甘草酸(c,d)定量校正模型 SPA 运行结果

2.3.2 CARS法 CARS法通过迭代和竞争的方式进行蒙特卡罗采样,每次采样过程中以固定比例的变量子集,建立PLSR定量校正模型,采用指数递减函数和自适应加权采样技术选择较大权重变量,以交叉验证均方根误差(RMSECV)值来评估子集选择出最关键变量<sup>[17]</sup>。

在甘草苷校正模型中,设置CARS算法主成分个数为15,算法进化次数为50次,运行结果如图3(a)所示,首先运行过程中变量采用数量,随着采样次数的增加变量使用更少;其次交叉验证均方根误差值的变化趋势,在28次采样时该值最低,此时特征变量数量个数为22,提取得到波长为[906.0、921.0、1 099.7、1 101.4、1 121.4、1 123.1、1 128.1、1 129.8、1 159.8、1 209.9、1 258.4、1 281.8、1 382.6、1 400.3、1 433.7、1 475.5、

1 477.1、1 492.2、1 495.5、1 513.9、1 515.6、1 545.6 nm]。图中展示回归系数的变化路径，“\*”表示最小RMSECV所对应的采样次数，各线表示不同采样次数回归系数变化路径。在甘草酸校正模型中，CARS主成分个数设置为14，运行结果如图3(b)所示，在29次采样时可以得到最小RMSECV，此时对应的最优波长变量为20，提取得到波长为[914.4、934.4、952.8、969.5、1 026.2、1 053.0、1 129.8、1 134.8、1 291.8、1 306.8、1 336.9、1338.5、1 351.9、1 370.3、1 371.9、1 373.6、1 380.3、1 392.0、1 549.0、1 579.0 nm]。

2.3.3 SFLA法 随机蛙跳算法具有适者生存和随机搜索的特性，通过在迭代过程中产生随机子集，每个变量具有其选择概率，概率越大，相应变量将越重要，最后优选出10个选择概率最大的变量作为模型的特征波长进行建模<sup>[8]</sup>。

在甘草苷校正模型中，设置窗口大小为3，主成分数为10，迭代次数为2 000，每个波长选择概率如图4(a)所示，大多数波长的选择概率均比较低，个别波长具有较高的选择概率，最后选择波长为[1 131.5、1 156.5、1 403.7、1 201.6、1 158.2、1 098.1、1 129.8、1 126.4、1 124.8、1 199.9 nm]。在甘草酸校正模型中，窗口大小设置为3，主成分数为12，经过800次迭代后波长的选择概率如图4(b)所示，最后选择波长为[1 351.9、1 368.6、1 380.3、1 296.8、1 133.1、1 371.9、1 370.3、1 338.5、947.8、1 291.8 nm]，甘草苷模型中选择概率较高波长集中在1 100~1 200 nm之间，而甘草酸模型中集中在1 300~1 400 nm之间，两者具有明显的区别。

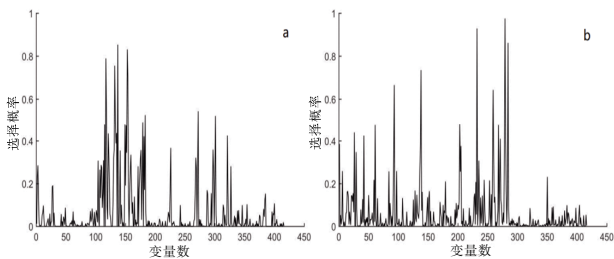


图4 甘草苷(a)和甘草酸(b)定量校正模型 SFLA 运行结果

2.3.4 siPLS法 siPLS算法将整个光谱平均分为n个子间隔，对每个子区间进行组合计算<sup>[9]</sup>。通过对联合区间和子区间数进行筛选，选择最佳的主因子数，比较PLSR模型中具有最低

RMSECV值的子区间组合作为特征波段。甘草苷和甘草酸含量联合区间偏最小二乘模型的调参结果见表3。

表3 不同联合区间数的最佳甘草苷和甘草酸模型结果

模型	联合区间数	划分区间数	所选区间	最佳主因子数	R <sup>2</sup>	RMSECV
甘草苷模型	2	9	[2,4]	10	0.935 0	0.003 2
		10	[1,4]	20	0.959 5	0.002 6
		11	[4,7]	11	0.944 9	0.003 0
	3	13	[4,13]	16	0.936 3	0.003 2
		10	[1,4,8]	17	0.965 5	0.002 4
		11	[4,5,8]	16	0.971 2	0.002 2
		12	[4,5,9]	11	0.969 6	0.002 2
	4	13	[4,6,10]	18	0.961 4	0.002 5
		6	[1,2,3,5]	14	0.972 2	0.002 1
		7	[3,4,6,7]	16	0.974 0	0.002 1
		8	[3,4,5,6]	13	0.971 3	0.002 2
9		[3,4,5,7]	13	0.970 9	0.002 2	
10		[3,4,9,10]	16	0.966 7	0.002 3	
11		[3,5,8,11]	17	0.914 0	0.003 7	
甘草酸模型	2	9	[3,6]	10	0.925 2	0.004 6
		10	[4,7]	10	0.932 5	0.004 4
		11	[4,7]	9	0.931 2	0.004 4
	3	9	[1,3,6]	11	0.936 3	0.004 3
		10	[4,6,7]	11	0.940 9	0.004 1
		11	[1,4,7]	12	0.933 9	0.004 4
	4	9	[1,3,6,7]	12	0.936 7	0.004 3
		10	[1,2,4,7]	12	0.940 6	0.004 1
		11	[1,4,7,8]	11	0.941 5	0.004 1
		12	[3,4,9,10]	11	0.936 9	0.004 3

由表3可知，在甘草苷siPLS模型中，在联合区间数为4，将光谱划为7个子区间时，主因子成分为16可获得最低RMSECV值，此时选择的波段为[1 106.4~1 303.5、1 403.7~1 600.0 nm]。在甘草酸siPLS模型中，联合区间数为4，将光谱划为11个子区间，主因子成分为11时可获得最低RMSECV值，选择的波段为[906.0~967.8、1 096.4~1 158.2、1 286.8~1 412.0 nm]。甘草苷和甘草酸siPLS模型运行结果见图5。

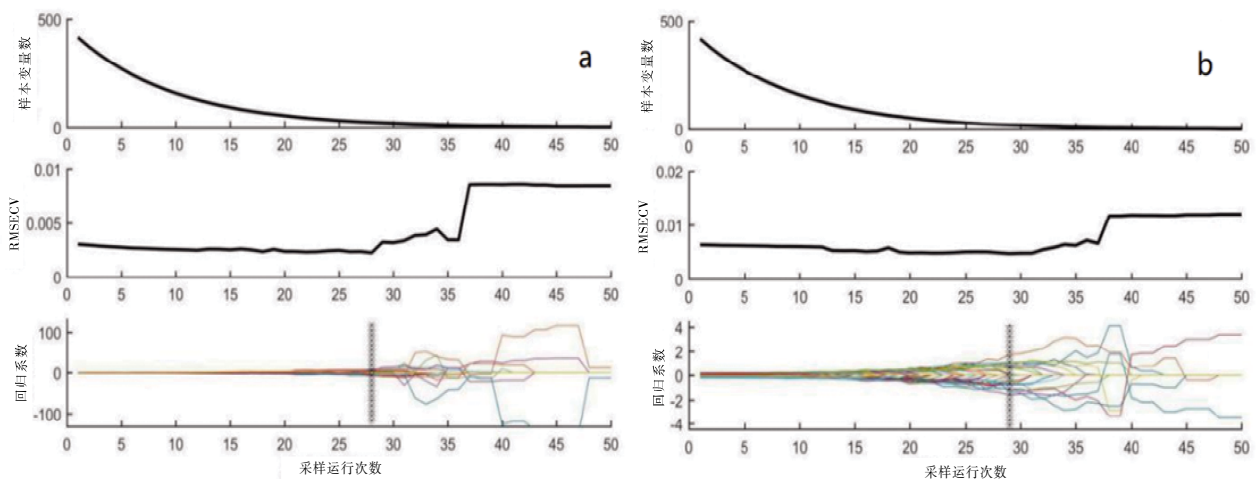


图3 甘草苷(a)和甘草酸(b)定量校正模型 CARS 运行结果

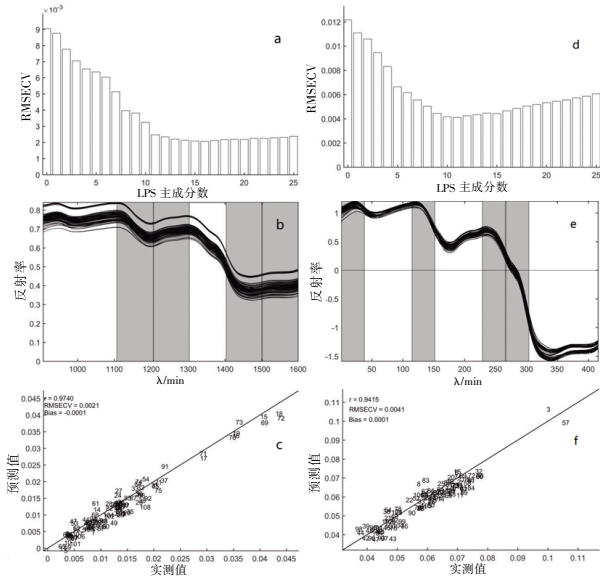


图5 甘草苷(a、b、c)和甘草酸(e、d、f)定量校正模型 siPLS 运行结果

2.3.5 biPLS法 biPLS算法是将全光谱平均分成n个子区间,首先对全光谱进行建模,之后每次剔除一个子区间,然后继续对余下的所有子区间进行建模,此步骤一直持续到只剩最后一个区间。其中,每次剔除的子区间为在回归模型建立过程中性能最差的一个<sup>[18]</sup>。甘草苷和甘草酸含量的反向区间偏最小二乘模型的调参结果见表4。

表4 不同子区间划分的最佳甘草苷和甘草酸模型结果

模型	划分区间数	所选区间	最佳主因子数	$R^2_{cv}$	RMSECV
甘草苷模型	9	[3,4,7,2,9,1]	17	0.9749	0.0020
	10	[4,2,10,9,7,1,3]	16	0.9742	0.0020
	11	[4,11,6,10,7,1]	16	0.9761	0.0020
	12	[4,12,8,10,11,2,3,5,7]	15	0.9726	0.0021
	13	[4,13,5,7,12,9,1,8,11,3]	19	0.9797	0.0018
	14	[5,7,14,9,12,11,8,1]	18	0.9800	0.0018
	15	[5,1,13]	18	0.9791	0.0018
甘草酸模型	9	[3,9,5,6,4]	13	0.9299	0.0045
	10	[4,7,2,1]	13	0.9311	0.0045
	11	[4,1,8,2]	13	0.9284	0.0046
	12	[4,1,9,2]	15	0.9270	0.0046
	13	[5,8,9,1]	11	0.9338	0.0044
	14	[2,10,5,1,3,11,13,6]	14	0.9328	0.0044
	15	[10,5,7,1,3,2,6,14]	14	0.9322	0.0044
	16	[11,6,1,3,2]	13	0.9295	0.0046
	17	[1,6,17,3,13,11,2]	13	0.9300	0.0045

由表4可知,在甘草苷biPLS模型中,当划分子区间数为14时,剔除其他子区间留下[5,7,14,9,12,11,8,1]子区间,此时最大主因子为18时,所建模型性能最佳,其 $R^2_{cv}$ 和RMSECV为0.9800和0.0018,此时选择的波段为[906.0~954.4、1106.4~1154.8、1206.6~1355.2、1407.0~1502.2、1552.3~1600.0 nm]。在甘草酸biPLS模型中,当划分子区间数为13时,所选区间为

[5,8,9,1],在最大主因子为11时模型性能最好,模型 $R^2_{cv}$ 值可达0.9338, RMSECV值为0.0044,此时选择的波段为[906.0~957.8、1119.8~1171.5、1280.1~1385.3 nm]。甘草苷和甘草酸biPLS模型运行结果见图6。

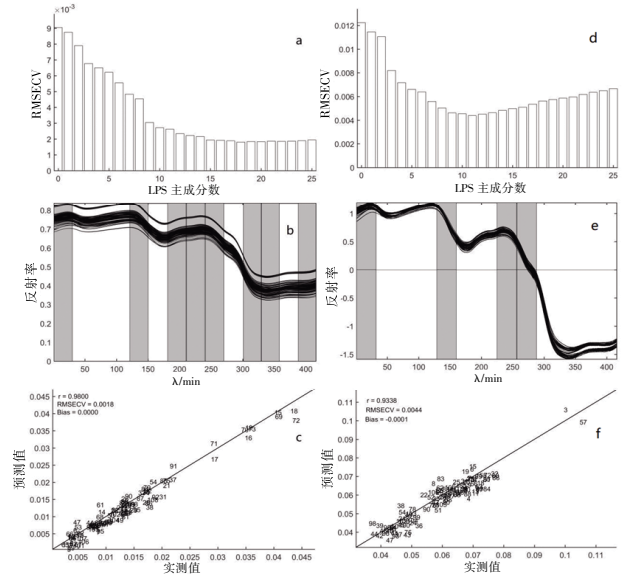


图6 甘草苷(a、b、c)和甘草酸(e、d、f)定量校正模型 biPLS 运行结果

2.3.6 定量校正模型的建立 采用多种特征波段筛选方法所建立的PLSR(偏最小二乘回归)、MLR(多元线性回归)和BRR(贝叶斯岭回归)的甘草苷和甘草酸定量校正模型性能参数如表5所示。

表5 不同变量筛选方法建模结果比较

特征波段	模型	波长数	甘草苷		甘草酸	
			校正集	预测集	校正集	预测集
			$R^2_c$	RMSEC	$R^2_p$	RMSEP
FULL	PLSR	416	0.9863	0.0010	0.9321	0.0022
	MLR		0.9999	0.0001	0.7451	0.0044
	BRR		0.9854	0.0011	0.9368	0.0021
SPA	PLSR	18	0.9673	0.0016	0.9291	0.0022
	MLR		0.9673	0.0016	0.9290	0.0022
	BRR		0.9657	0.0016	0.9262	0.0022
CARS	PLSR	22	0.9780	0.0013	0.9587	0.0018
	MLR		0.9800	0.0013	0.9455	0.0020
	BRR		0.9780	0.0013	0.9577	0.0018
SFLA	PLSR	10	0.8275	0.0037	0.7908	0.0040
	MLR		0.8369	0.0036	0.7833	0.0041
	BRR		0.7979	0.0040	0.7634	0.0043
siPLS	PLSR	237	0.9780	0.0013	0.9165	0.0025
	MLR		0.9999	0.0000	0.6951	0.0047
	BRR		0.9813	0.0012	0.9175	0.0025
biPLS	PLSR	237	0.9864	0.0010	0.9351	0.0022
	MLR		0.9999	0.0000	0.7716	0.0042
	BRR		0.9848	0.0011	0.9416	0.0020

从表5中可以看出,对于甘草苷模型,除SFLA以外,其他特征变量提取算法能在大幅度降低光谱波段使用的情况下,

基本可以与全光谱所建立模型的性能相当;其中,CARS和biPLS方法甚至能超越全光谱模型性能。CARS方法所建立的PLSR模型只需使用22个波长即可达到最好的效果,大幅度的缩短了模型的运行时间,该模型实测值和预测值散点图分布见图7(a)所示。由此可见,CARS-PLSR模型对甘草饮片甘草苷含量具有良好的预测能力,模型预测集决定系数( $R_p^2$ )和误差均方根(RMSEP)分别为0.958 7和0.001 8。

在甘草酸模型中,提取波段后均能实现在光谱维度降低的同时模型性能还能得到不错的提升,CARS方法所建立的BRM模型只需使用20个波长即可达到所有模型中最优性能。该模型实测值和预测值散点图分布见图7(b)所示。由此可见,CARS-BRM模型对甘草饮片中甘草酸含量具有良好的预测能力,模型预测集决定系数( $R_p^2$ )和误差均方根(RMSEP)分别为0.897 3和0.003 8。

3种不同类型的定量校正模型之间的比较结果显示:MLR所建立模型校正集决定系数较高,但是预测集决定系数很低,误差均方根也出现分化,说明该模型出现过拟合现象;BRM模型在模型的稳定性上有所欠缺,其不同特征波段提取方法中出现分化;而PLSR模型从整体上看,在不同的波段筛选方法下均具有不错的效果。

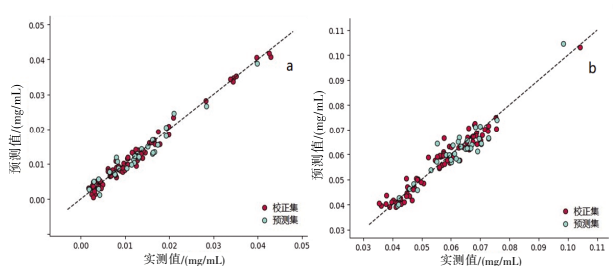


图7 甘草苷(a)和甘草酸(b)定量模型预测值与实测值相关图

### 3 结 论

3.1 技术优势分析 本研究使用的高光谱成像技术相较于传统HPLC检测方法,无需样品前处理、不消耗化学试剂,检测时间短。该技术不仅适用于实验室精准检测,更可拓展至工业生产在线质控、市场监管现场抽检等场景,能满足对中药材大规模批量检测与生产线实时监控的智能化需求,打破了传统检测方法对实验环境与专业操作的依赖,具备广泛的推广潜力。

3.2 现存问题与改进方向 尽管模型预测效果良好,但本研究仍存在一些局限性,需要在后续工作中加以改进。一是模型预测精度仍有提升空间。这主要受限于样本数量的规模与多样性。未来通过持续扩充样本库,涵盖不同产地、批次、等级的样品,将使模型更具普适性。二是存在一定的过拟合风险。特别是在使用CARS等复杂的特征波段筛选算法时,部分样本因成分分布不均、形态差异等因素可能导致过拟合现象。未来研究可通过不断扩充样本数量,改进高光谱采集过程,引入更先进的算法(如深度学习)来优化模型结构,在简化模型的同时,进一步提高其泛化能力。

3.3 产业应用建议 为推动高光谱成像技术从实验室走向产业化应用,本文提出以下建议:首先,开发专用设备。当前实验室高光谱成像系统成本高、体积大,未来需开发出成本

更低、体积更小的专用便携式高光谱检测仪,简化设备功能模块,降低采购与维护成本。其次,推动建立高光谱检测技术行业规范,明确样品预处理、检测环境参数等操作标准,建立覆盖全国主产区、多品种的中药材高光谱数据库,为模型的广泛适用性提供数据支撑,在实际生产环境中检验并优化模型,真正实现中药材质量的智能化、在线化监控。

本研究通过高光谱成像技术对短波红外波段下甘草饮片进行检测,结合液相色谱甘草苷和甘草酸含量检测结果,筛选比较5种光谱预处理方法和6种特征波段提取方法对预测模型的影响,分别建立了3种定量校正模型,从而实现对甘草指标成分甘草苷和甘草酸无损检测。SG-17点平滑和SNV预处理方法分别对应甘草苷和甘草酸定量模型能够有效剔除无关信息干扰,建立稳定、可靠的模型,特征波段提取方法能从全光谱信息中筛选得到与甘草苷和甘草酸含量特征相关的信息,从而实现模型预测性能显著提升的同时,大幅缩减模型运行的时间。其中,CARS-PLSR甘草苷含量预测模型和CARS-BRM甘草酸含量预测模型具有可靠的稳定性与准确性。应用高光谱成像技术能够实现对甘草中甘草苷和甘草酸含量的有效预测,为甘草品质的快速无损检测提供重要参考,为工业生产中中药材在线检测提供理论基础和技术支持。

### 参考文献

- [1] 国家药典委员会.中华人民共和国药典:一部[S].北京:中国医药科技出版社,2020.
- [2] 张玉珊,袁欣月,南红梅,等.基于多成分含量测定和化学计量学的不同基原甘草质量评价研究[J].中药材,2024,47(5):1208-1214.
- [3] 吴新,吴涛,王冰清,等.不同产地甘草UPLC指纹图谱的建立及6种成分定量分析[J].中草药,2025,56(7):2518-2526.
- [4] 万鑫浩,陶青,王子千,等.基于机器学习的中药制剂快速无损检测技术研究进展[J].中国中药杂志,2024,49(24):6541-6548.
- [5] 王嵩,刘诗雨,王慧,等.高光谱智能分析技术在中药材领域的研究与应用[J].中国中药杂志,2023,48(16):4320-4327.
- [6] 张蕾,赵宇平,庞银银,等.基于高光谱成像技术的葛根与葛藤鉴别[J].中国中药杂志,2023,48(16):4362-4369.
- [7] 潘忆瑕,张弘旭,颜继忠,等.基于高光谱与纹理特征融合的甘草药材基源鉴别研究[J].分析测试学报,2024,43(11):1745-1753.
- [8] 江恩赐,陈林,颜继忠,等.智能高光谱成像融合CARS特征波段筛选快速检测硫熏牛膝SO<sub>2</sub>含量[J].中国中药杂志,2022,47(7):1864-1870.
- [9] BAI R B, ZHOU J H, WANG S M, et al. Identification and classification of *Coix seed* storage years based on hyperspectral imaging technology combined with deep learning[J]. Foods, 2024, 13(3):498.
- [10] RU C L, LI Z H, TANG R Z. A hyperspectral imaging approach for classifying geographical origins of rhizoma

- atractylodis macrocephalae using the fusion of spectrum-image in VNIR and SWIR ranges (VNIR-SWIR-FuSI)[J]. *Sensors*, 2019,19(9):2045.
- [11] PAN Y X, ZHANG H X, CHEN Y, et al. Applications of hyperspectral imaging technology combined with machine learning in quality control of traditional Chinese medicine from the perspective of artificial intelligence: A review[J]. *Crit Rev Anal Chem*, 2024, 54(8):2850-2864.
- [12] ZHANG H X, PAN Y X, LIU X Y, et al. Recognition of the rhizome of red ginseng based on spectral-image dual-scale digital information combined with intelligent algorithms[J]. *Spectrochim Acta Part A Mol Biomol Spectrosc*, 2023, 297: 122742.
- [13] ZHANG L, GUAN Y, WANG N, et al. Identification of growth years for *Puerariae Thomsonii Radix* based on hyperspectral imaging technology and deep learning algorithm[J]. *Sci Rep*, 2023, 13:14286.
- [14] 殷文俊,唐建飞,郑洁,等.基于拉曼光谱实时监测甘草配方颗粒的提取过程[J].*中草药*, 2021, 52(18):5560-5568.
- [15] HU H Q, MEI Y L, WEI Y P, et al. Rapid identification of moxa wool storage period based on hyperspectral imaging technology and machine learning [J]. *Heliyon*, 2024, 10(18):e37650.
- [16] 吴龙国,马玲,张瑶,等.基于高光谱成像技术的包衣甘蓝种子色度检测[J].*分析测试学报*, 2025, 44(3):454-463.
- [17] JIN C, ZHOU X Y, HE M Y, et al. A novel method combining deep learning with the Kennard-Stone algorithm for training dataset selection for image-based rice seed variety identification[J]. *J Sci Food Agric*, 2024, 104(13):8332-8342.
- [18] 李杨,李翠玲,王秀,等.高光谱成像的黄瓜病虫害识别和特征波长提取方法[J].*光谱学与光谱分析*, 2024, 44(2): 301-309.
- [19] LI W L, SHI Y, HUANG X W, et al. Study on the diffusion and optimization of sucrose in gaido seak based on finite element analysis and hyperspectral imaging technology[J]. *Foods*, 2024, 13(2):249.

(收稿日期:2025-07-28 编辑:时格格)

(上接第25页)

- [23] 曾晶,尤昭玲.尤昭玲教授基于“生殖链”假说运用“生殖五论”治疗妇科生殖疾病的经验[J].*湖南中医药大学学报*, 2024, 44(4):643-649.
- [24] 马祯慧,李洪霖,李娜娜,等.中药调控AMPK信号通路抗消化道肿瘤的研究进展[J].*药物评价研究*, 2025, 48(7): 1992-2003.
- [25] CECCHINO G N, GARCÍA-VELASCO J A, RIAL E. Reproductive senescence impairs the energy metabolism of human luteinized granulosa cells[J]. *Reprod Biomed Online*, 2021, 43(5):779-787.
- [26] CHIANG J L, SHUKLA P, PAGIDAS K, et al. Mitochondria in ovarian aging and reproductive longevity[J]. *Ageing Res Rev*, 2020, 63: 101168.
- [27] TIOSANO D, MEARS J A, BUCHNER D A. Mitochondrial dysfunction in primary ovarian insufficiency[J]. *Endocrinology*, 2019, 160(10):2353-2366.
- [28] MASUDUL HOQUE S A M, UMEHARA T, KAWAI T, et al. Adverse effect of superoxide-induced mitochondrial damage in granulosa cells on follicular development in mouse ovaries[J]. *Free Radic Biol Med*, 2021, 163:344-355.
- [29] LABARTA E, DE LOS SANTOS M J, ESCRIBÁ M J, et al. Mitochondria as a tool for oocyte rejuvenation[J]. *Fertil Steril*, 2019, 111(2):219-226.
- [30] 赵芮竹,任鑫悦,王宇航,等.基于AMPK/SIRT1自噬通路探讨黄连解毒汤防治脓毒症小鼠急性肝损伤的作用机制[J].*中国中药杂志*, 2025, 50(2):507-514.
- [31] 卵巢储备功能减退临床诊治专家共识专家组,中华预防医学会生育力保护分会生殖内分泌生育保护学组.卵巢储备功能减退临床诊治专家共识[J].*生殖医学杂志*, 2022, 31(4):425-434.
- [32] LI Q, GENG X D, ZHENG W, et al. Current understanding of ovarian aging[J]. *Sci China Life Sci*, 2012, 55(8):659-669.

(收稿日期:2025-09-23 编辑:罗英姣)