

## 基于不同算法对杏贝止咳颗粒中间体物料属性与颗粒溶解性的相关性研究

陶振<sup>1,2</sup>, 洪韵<sup>3</sup>, 安双凤<sup>3</sup>, 汪盛华<sup>3</sup>, 闫明<sup>1,2,3\*</sup>, 朱明岩<sup>3</sup>, 章晨峰<sup>1,2,3</sup>, 王振中<sup>1,2,3</sup>, 肖伟<sup>1,2,3\*</sup>

1. 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/南京中医药大学), 江苏 南京 210000

2. 南京中医药大学康缘中药学院, 江苏 南京 210023

3. 江苏康缘药业股份有限公司, 江苏 连云港 222001

**摘要:** 目的 分析比较 5 种算法对杏贝止咳颗粒(Xingbei Zhike Granules, XZG)溶解性的模型预测效果, 确定最优算法和影响颗粒溶解性的中间体粉末关键物料属性(critical material attributes, CMA), 提升 XZG 的溶解性。方法 以 25 个制剂前粉末物料属性为输入, 颗粒溶化率为输出, 对比偏最小二乘(partial least squares, PLS)、决策树(decision tree, DT)、随机森林(random forest, RF)、反向传播神经网络(back propagation neural network, BPNN)和支持向量机回归(support vector regression, SVR)算法的模型拟合与预测效果, 同时结合变量重要性与方差膨胀因子(variance inflation factor, VIF), 筛选关键物料属性。结果 5 个算法中, RF 算法的模型拟合与预测效果最好, 训练集决定系数( $R^2$ )为 0.865, 测试集  $R^2$  为 0.854, 均方根误差(root mean square error, RMSE)为 1.375, 平均绝对百分比误差(mean absolute percent error, MAPE)为 1.153%。筛选出的关键物料属性有休止角( $\alpha$ )、崩溃角( $\beta$ )、平板角( $\gamma$ )、吸湿性( $H$ )、含水量(HR)、100~212  $\mu\text{m}$  颗粒含量( $F_m$ )、长度平均径( $D_{21}$ )。结论 RF 算法建立的模型更适合预测 XZG 的溶解性; 通过控制制剂成型用主药和辅料的关键物料属性, 可以提高 XZG 的溶解性, 为提升中药颗粒剂的溶解性和品质提供新的思路。

**关键词:** 杏贝止咳颗粒; 溶解性; 关键物料属性; 物理指纹图谱; 偏最小二乘算法; 决策树算法; 随机森林算法; 反向传播神经网络算法; 支持向量机回归算法

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2024)22-7644-09

DOI: 10.7501/j.issn.0253-2670.2024.22.008

## Correlation between material properties and particle solubility of intermediate of Xingbei Zhike Granules based on different algorithms

TAO Zhen<sup>1,2</sup>, HONG Yun<sup>3</sup>, AN Shuangfeng<sup>3</sup>, WANG Shenghua<sup>3</sup>, YAN Ming<sup>1,2,3</sup>, ZHU Mingyan<sup>3</sup>, ZHANG Chenfeng<sup>1,2,3</sup>, WANG Zhenzhong<sup>1,2,3</sup>, XIAO Wei<sup>1,2,3</sup>

1. State Key Laboratory on Technologies for Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacture (Jiangsu Kanion Pharmaceutical Co., Ltd., & Nanjing University of Chinese Medicine), Nanjing 210000, China

2. Kanion School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing 210023, China

3. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China

**Abstract: Objective** To analyze and compare the model prediction effects of five algorithms on the solubility of Xingbei Zhike Granules (XZG, 杏贝止咳颗粒), to determine optimal algorithm and the critical material attributes (CMA) affecting the solubility of particles, and to improve the solubility of XZG. **Methods** Taking 25 powder material properties before granulation as input, and the solubility of particles as output. The model fitting and prediction effects of partial least squares (PLS), decision tree (DT), random forest (RF), back propagation neural network (BPNN), and support vector regression (SVR) algorithms were compared. At the same time, combined with the importance of variable and variance inflation factor (VIF), the key material properties were screened. **Results** The RF algorithm has the best model fitting and prediction effect. The training set determination coefficient is 0.865, the test set determination coefficient is 0.854, the root mean square error is 1.375, and the mean absolute percentage error is 1.153%. The CMA

收稿日期: 2024-06-15

基金项目: 国家工信部 2023 年产业基础再造和制造业高质量发展专项(TC2308068)

作者简介: 陶振(2000—), 男, 硕士研究生, 研究方向为中药制药技术与产品开发。E-mail: 2561150409@qq.com

\*通信作者: 肖伟, 中国工程院院士, 教授, 博士生导师, 研究方向为中药新药研发及过程质量控制研究。E-mail: kanionlunwen@163.com

闫明, 博士, 研究方向为中药新药研发。E-mail: ymhezhihuo@163.com

screened by had collapse  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $H$ ,  $HR$ ,  $F_m$  and  $D_{21}$ . **Conclusion** The model established by the RF algorithm is more suitable for predicting the solubility of XZG. The solubility of XZG can be improved by controlling the key material properties of the main drugs and excipients for preparation molding, which provides a new idea for improving the solubility and quality of traditional Chinese medicinal granules.

**Key words:** Xingbei Zhike Granules; solubility; critical material attributes; physical fingerprint; partial least squares algorithm; decision tree algorithm; random forest algorithm; back propagation neural network algorithm; support vector regression algorithm

中药颗粒剂是中成药常用剂型，制备工艺相对简单、载药量高、携带和服用方便，是最常用的中药复方口服固体剂型之一。溶解性是颗粒剂关键质量属性，是颗粒剂感官质量的重要考察指标。于佳琦等<sup>[1]</sup>收集了 105 批市售中成药颗粒剂，进行溶化性测试，其中符合《中国药典》2020 年版溶化性要求的批次数占比为 84.76%，部分品种的溶化性仍有待改进。杏贝止咳颗粒（Xingbei Zhike Granules, XZG）处方由麻黄、苦杏仁、桔梗、前胡、浙贝母、百部、北沙参、木蝴蝶、甘草 9 味中药材组成，提取工艺为水-醇双提，课题组前期研究发现其溶解性还有上升空间。颗粒溶解性受多种因素影响，制粒前一步中间体粉末物料属性是制粒前重要的内容，而粉末与颗粒的质量控制是保证最终产品质量稳定可控的关键<sup>[2-3]</sup>。

课题组前期研究发现，通过对多批次中药粉体物料属性的测定，能够从数字的角度探寻中药粉体物料属性与颗粒溶解性的相关性。王晴等<sup>[4]</sup>以桂枝茯苓胶囊（Guizhi Fuling Capsules, GFC）制剂成型过程中原料、中间体粉末和胶囊成品为研究对象，建立了 GFC 内容物的吸湿性预测模型，但该模型的性能需要进一步提高。陈琪等<sup>[5]</sup>同样以 80 批 GFC 生产过程中的 5 种中间体为研究对象，在以偏最小二乘（partial least squares, PLS）算法构建的吸湿性预测模型的基础上，比较了不同算法模型的预测精度，结果表明多元自适应回归样条算法建立的吸湿性预测模型更适合 GFC 的生产应用。汪盛华等<sup>[6]</sup>以 5 个水-醇双提品种的中间体粉末与颗粒为研究对象，运用 PLS 算法研究中间体粉末物理性质与颗粒溶解性的相关性，筛选出影响颗粒溶解性的关键物料属性（critical material attributes, CMA），但该算法建立的模型相对预测误差偏大。本研究尝试从算法角度提高预测模型性能，比较不同算法对 XZG 溶解性预测模型性能的影响，以期优化模型的预测性能，筛选 CMA，提升 XZG 溶解性。

决策树（decision tree, DT）、随机森林（random forest, RF）、反向传播神经网络（back propagation

neural network, BPNN）和支持向量机回归（support vector regression, SVR）等算法的原理不尽相同，但都与 PLS 算法一样，可以处理自变量与因变量之间的非线性关系，并给出预测结果。吴晓倩等<sup>[7]</sup>利用 DT 算法对未来成绩进行预测，预测准确率达到 94%。王民等<sup>[8]</sup>通过 RF 算法构建岩相识别模型，建立了岩相分布的快速识别方法。宋丽芳等<sup>[9]</sup>通过 SVR 和 BPNN 算法构建了油菜叶片含水量预测模型，为油菜水分诊断提供依据。本研究以 XZG 制剂成型工艺中喷干细粉（主药）与不同辅料所制成的混合粉为实验对象，基于粉体学性质，对混合粉与颗粒物料属性进行综合表征<sup>[10-12]</sup>。收集混合粉物料属性与颗粒溶化率数据，采用 PLS、DT、RF、BPNN、SVR 这 5 种算法进行建模并预测分析，优选出最优算法，同时筛选出影响 XZG 溶解性的 CMA，可以通过控制制剂成型用主药和辅料的 CMA，提升 XZG 的溶解性，为提升中药颗粒剂的品质提供新的思路与方法。

## 1 仪器与材料

### 1.1 仪器

BT-1001 型粉体特性测试仪、Bettersize2600 型激光粒度分布仪，丹东百特仪器有限公司；GL2-25 型实验室干法制粒机，江苏张家港市开创机械制造有限公司；Mettler Toledo 204 型万分之一电子天平，瑞士梅特勒-托利多集团公司；HWS26 型电热恒温水浴锅、DHG-9145A 型电热鼓风干燥箱、LHS-250HC-11 型恒温恒湿箱，上海一恒科学仪器有限公司；MYP11-2 型磁力搅拌器，上海梅颖浦仪器仪表制造有限公司；300 mm 型干燥器，江苏华欧玻璃有限公司。

### 1.2 材料

麦芽糊精，批号 2024033151，吉林中粮生化能源销售有限公司；XZG（批号 230801）、中间体粉末（批号 Z230801）、喷干细粉（批号 Z230801），江苏康缘药业股份有限公司；糊精（批号 202301148）、玉米淀粉（批号 20230410），曲阜市天利药用辅料有限公司；乳糖，批号 2310002，江苏道宁药业有限

公司；可溶性淀粉，批号 F2207034，湖州展望药业有限公司。

## 2 方法与结果

### 2.1 粉末与颗粒物料属性测定方法

参考文献方法<sup>[6,10,13-14]</sup>对物料的休止角( $\alpha$ )、平板角( $\gamma$ )、松装密度( $D_a$ )、振实密度( $D_c$ )、孔隙率( $I_c$ )、卡尔指数(Carr index, IC)、豪斯纳比(Hausner ratio, IH)、吸湿性( $H$ )、含水量(HR)、0~50  $\mu\text{m}$  颗粒含量(Pf)、100~212  $\mu\text{m}$  颗粒含量( $F_m$ )、50~100  $\mu\text{m}$  颗粒含量( $F_{m-1}$ )、212~355  $\mu\text{m}$  颗粒含量( $F_{m+1}$ )、相对均齐度( $I_0$ )、粒径( $D_{10}$ 、 $D_{50}$ 、 $D_{60}$ 、 $D_{90}$ )、分布范围(span)、宽度(width)和比表面积(SSA)等物理属性进行检测。

**2.1.1 崩溃角( $\beta$ )** 使用智能粉体特性测试仪测定，采用固定底面积法，底面为直径 100 mm 的圆盘，通过往进料筒里进入待测粉末，启动仪器后，样品经振动过筛后流落圆盘，待形成的锥体角度稳定后，震动 3 次，读取数据，平行测定 3 次后取其平均值计算。

**2.1.2 体积平均径( $D_{43}$ )、面积平均径( $D_{32}$ )、长度平均径( $D_{21}$ )** 使用激光粒度分布仪测定，以空气为介质，将样品加入干法分散系统的进料斗中进行测定。

**2.1.3 溶化率** 参照《中国药典》2020 年版四部 0104 颗粒剂溶化性的测定方法<sup>[15]</sup>。称取约 10 g 待测样品( $m_1$ )，加入 200 mL 热水 [(75±2) °C] 搅拌 5 min 后，1000 r/min 离心(离心半径为 10 cm) 5 min，取上清液，置恒定质量后的蒸发皿( $m_2$ )上蒸干后，于 105 °C 干燥箱中干燥至恒定质量( $m_3$ )，计算溶化率，计算公式如下。

$$\text{溶化率} = (m_3 - m_2) / m_1 \quad (1)$$

### 2.2 筛选辅料与药辅比

在实际生产中，XZG 的药辅比在 3:1 左右，本研究在此基础上扩大药辅比范围，将 XZG 喷干细粉(主药)分别与麦芽糊精、糊精、乳糖、玉米淀粉、可溶性淀粉这 5 种常用辅料，按药辅比 1:1、2:1、3:1、4:1 混匀，以 XZG 喷干细粉为实验样品，进行干法制粒预实验，确定最优制粒参数(送料速度 30 Hz、压片速度 5 Hz、制粒速度 15 Hz)，将样品在此参数下进行干法制粒，并检测颗粒的溶化率，结果见表 1。结果显示，XZG 喷干细粉(主药)与玉米淀粉的混料样品的溶化率过低，故排除玉米淀粉。

表 1 不同辅料及药辅比溶化率结果

Table 1 Results of dissolution rate of different excipients and drug-to-excipient ratio

辅料名称	溶化率/%			
	1:1	2:1	3:1	4:1
麦芽糊精	92.412	90.821	92.779	88.591
可溶性淀粉	85.983	85.516	89.556	84.462
乳糖	89.581	89.514	89.847	87.221
糊精	87.172	84.292	87.185	79.603
玉米淀粉	46.775	64.170	71.274	68.592

从图 1 中可以看出，剩余 4 种辅料中，溶化率最低的均为药辅比 4:1，溶化率最高的均为药辅比 3:1。最终选择麦芽糊精、糊精、乳糖、可溶性淀粉这 4 种辅料进行混料设计，结合实际生产的药辅比，确定药辅比为 3:1~4:1，在此药辅比下，溶化率变化较大，便于筛选 CMA。

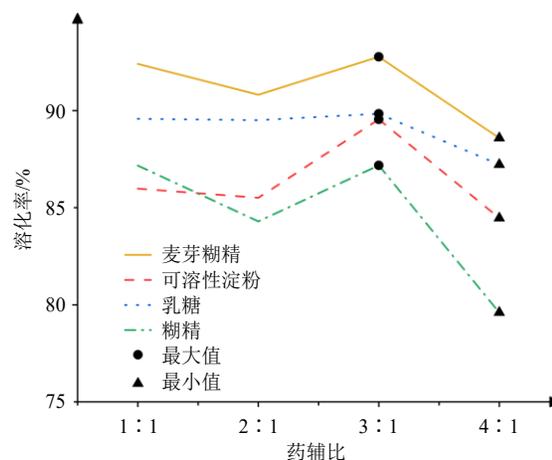


图 1 4 种辅料溶化率折线图

Fig. 1 Line chart of dissolution rate of four excipients

### 2.3 混料设计与物理指纹图谱构建

根据“2.2”项下的实验结果，使用 Design-Expert 13 软件进行混料设计，共 79 组样品，混料设计结果见表 2。按照混料设计结果制备样品，每个样品制备 2 份，按照“2.1”项下的方法检测物料属性，每个参数检测 3 次，相对标准偏差(relative standard deviation, RSD)均要求小于 3%。

为了直观反映出混料样品之间的差异性和相似性<sup>[3]</sup>，通过参考文献数据<sup>[4,16]</sup>，以稳定性、均一性、流动性、堆积性及可压性作为物理指纹图谱的一级指标，以 HR、 $H$ 、Pf、 $I_0$ 、span、width、IH、 $\alpha$ 、 $D_a$ 、 $D_c$ 、 $I_c$ 、IC 这 12 个参数为二级指标，构建混料样品的物理指纹图谱。中药物理指纹图谱是继中药化学指纹图谱<sup>[17]</sup>、中药生物指纹图谱<sup>[18]</sup>而发展起来的用

表 2 混料设计  
Table 2 Mixture design

序号	投药量/g					溶化率/ %	序号	投药量/g					溶化率/ %
	喷干细粉	麦芽糊精	糊精	乳糖	可溶性淀粉			喷干细粉	麦芽糊精	糊精	乳糖	可溶性淀粉	
1	150.00	50.00	0.00	0.00	0.00	93.292	41	160.00	0.00	0.00	26.67	13.33	91.474
2	150.00	0.00	16.67	0.00	33.33	91.016	42	160.00	13.33	13.33	0.00	13.33	91.256
3	150.00	0.00	0.00	0.00	50.00	85.527	43	157.50	13.33	0.00	26.67	0.00	91.812
4	150.00	31.25	6.25	6.25	6.25	94.685	44	157.50	0.00	26.67	0.00	13.33	81.841
5	150.00	16.67	0.00	16.67	16.67	91.553	45	150.00	40.00	0.00	0.00	0.00	93.890
6	150.00	0.00	0.00	0.00	50.00	85.524	46	150.00	26.67	0.00	13.33	0.00	89.200
7	150.00	0.00	0.00	33.33	16.67	95.171	47	150.00	0.00	40.00	0.00	0.00	94.685
8	150.00	16.67	0.00	33.33	0.00	85.808	48	160.00	5.00	5.00	25.00	5.00	91.703
9	150.00	0.00	0.00	16.67	33.33	89.146	49	155.00	0.00	0.00	40.00	0.00	94.436
10	150.00	0.00	33.33	16.67	0.00	95.361	50	150.00	0.00	0.00	13.33	26.67	85.493
11	150.00	0.00	16.67	33.33	0.00	94.480	51	160.00	0.00	26.67	13.33	0.00	91.365
12	150.00	0.00	16.67	16.67	16.67	93.688	52	160.00	26.67	0.00	0.00	13.33	87.834
13	150.00	50.00	0.00	0.00	0.00	93.296	53	160.00	10.00	10.00	10.00	10.00	88.602
14	150.00	6.25	6.25	6.25	31.25	88.781	54	160.00	0.00	13.33	0.00	26.67	85.516
15	160.00	0.00	0.00	50.00	0.00	91.660	55	160.00	5.00	25.00	5.00	5.00	87.857
16	160.00	33.33	0.00	0.00	16.67	94.213	56	160.00	13.33	0.00	0.00	26.67	84.115
17	160.00	33.33	16.67	0.00	0.00	93.812	57	155.00	0.00	0.00	0.00	40.00	88.278
18	150.00	16.67	16.67	0.00	16.67	87.045	58	150.00	26.67	13.33	0.00	0.00	94.865
19	160.00	33.33	0.00	16.67	0.00	90.415	59	150.00	25.00	0.00	0.00	25.00	91.784
20	160.00	16.67	33.33	0.00	0.00	89.976	60	150.00	5.63	25.63	5.63	5.63	88.461
21	160.00	12.50	12.50	12.50	12.50	86.004	61	150.00	0.00	50.00	0.00	0.00	87.862
22	160.00	0.00	0.00	50.00	0.00	91.660	62	150.00	0.00	25.00	0.00	25.00	88.273
23	160.00	0.00	50.00	0.00	0.00	88.473	63	150.00	50.00	0.00	0.00	0.00	90.205
24	160.00	0.00	50.00	0.00	0.00	88.475	64	160.00	0.00	25.00	25.00	0.00	84.166
25	160.00	0.00	33.33	0.00	16.67	87.683	65	150.00	25.00	25.00	0.00	0.00	89.262
26	150.00	16.67	16.67	16.67	0.00	95.919	66	150.00	0.00	0.00	50.00	0.00	80.401
27	160.00	6.25	6.25	31.25	6.25	90.258	67	160.00	0.00	40.00	0.00	0.00	82.342
28	150.00	16.67	0.00	0.00	33.33	87.202	68	160.00	15.00	0.00	15.00	15.00	79.450
29	150.00	6.25	31.25	6.25	6.25	86.221	69	160.00	0.00	25.00	0.00	25.00	87.908
30	150.00	0.00	0.00	40.00	0.00	94.430	70	150.00	0.00	0.00	0.00	40.00	84.506
31	150.00	5.00	5.00	5.00	25.00	90.542	71	150.00	0.00	0.00	0.00	50.00	85.124
32	160.00	0.00	0.00	0.00	40.00	88.275	72	150.00	5.63	5.63	25.63	5.63	85.332
33	150.00	13.33	26.67	0.00	0.00	98.263	73	150.00	0.00	0.00	0.00	50.00	85.105
34	160.00	0.00	40.00	0.00	0.00	94.684	74	160.00	0.00	0.00	25.00	25.00	80.571
35	160.00	13.33	13.33	13.33	0.00	95.210	75	160.00	5.63	5.63	5.63	25.63	84.045
36	160.00	13.33	0.00	13.33	13.33	92.362	76	160.00	45.00	0.00	0.00	0.00	88.661
37	150.00	0.00	13.33	26.67	0.00	92.780	77	155.00	40.00	0.00	0.00	0.00	89.203
38	160.00	0.00	13.33	13.33	13.33	91.911	78	150.00	11.25	11.25	11.25	11.25	85.965
39	160.00	40.00	0.00	0.00	0.00	93.893	79	150.00	0.00	25.00	25.00	0.00	84.158
40	160.00	25.00	5.00	5.00	5.00	94.225							

来表征粉末物理属性的图谱，以便更好地评价其质量一致性，预测中间体粉末与制剂的兼容性<sup>[19-20]</sup>，达到产品质量可控。

对 79 组样品的 12 个物料属性进行范围统计并转化，使其数值在 0~10，各二级指标转化公式见表 3。并最终用雷达图的形式直观可视地呈现物理指纹

图谱, 79 个样品用不同颜色线条表示, 结果见图 2, 其中内部空白部分面积越小, 代表粉末间差异性越大。基于夹角余弦值方法<sup>[21]</sup>, 计算相似度, 物理指纹图谱相似度越接近于 1, 表明各品种批次间粉末越相近。相似度结果显示, XZG 混料样品间的物料属性相似度分布在 73.911%~99.999%。溶化率的最大值为 98.263%, 最小值为 79.450%, 极差达到了 18.813%, 说明这 79 组混合粉的物料属性与颗粒的溶解性差异性较大, 便于筛选影响颗粒溶解性的 CMA。

表 3 二级物理质量指标标准化转换公式

Table 3 Standardization conversion formula of secondary physical quality indexes

一级指标	二级指标	单位	数据范围 (x)	转化公式参考
稳定性	HR	%	10~0	10-x
	H	%	30~0	10-x/3
均一性	Pf	%	100~0	10-x/10
	Iθ	-	0~0.000 2	50 000 x
	span	-	3~2	30-10 x
	width	-	60~30	(60-x)/3
流动性	IH	-	3~1	(30-10 x)/2
	α	°	60~0	10-x/6
堆积性	Da	g·mL <sup>-1</sup>	0~1	10 x
	Dc	g·mL <sup>-1</sup>	0~1	10 x
可压性	Ic	-	1~3	10 x/3
	IC	%	0~100	x/10

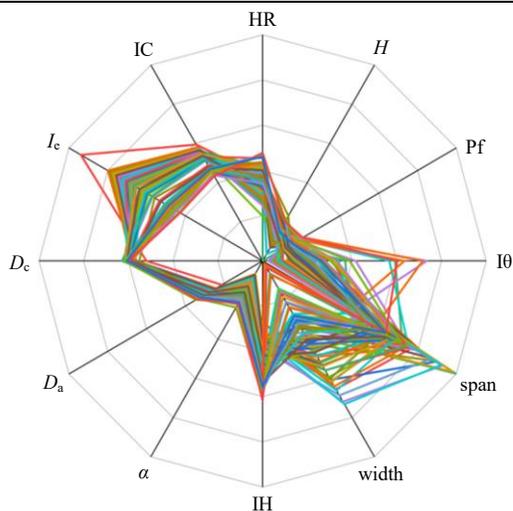


图 2 XZG 混料样品物理指纹图谱

Fig. 2 Physical fingerprint of XZG mixture sample

### 2.4 算法模型构建与性能评价

为了将粉末物料属性与颗粒溶解性的相关性进行更全面分析, 将可以获得的物料属性全部纳入模型中, 在物理指纹图谱 12 个物性指标 (HR、H、

Pf、Iθ、span、width、IH、α、Da、Dc、Ic、IC) 的基础上增加了 13 个物料属性 (β、γ、D<sub>10</sub>、D<sub>50</sub>、D<sub>60</sub>、D<sub>90</sub>、F<sub>m</sub>、F<sub>m-1</sub>、F<sub>m+1</sub>、SSA、D<sub>21</sub>、D<sub>32</sub>、D<sub>43</sub>)。以混料样品的 25 个粉末物料属性为输入, 颗粒溶化率为输出, 调用 Python 3.11.1 库中不同算法 (PLS、DT、RF、BPNN、SVR) 建立 XZG 混料样品的溶解性预测模型。用网格搜索法和经验法来选择各算法的最优超参数, 结合变量重要性和逐一剔除法筛选最优算法模型, 再结合方差膨胀因子 (variance inflation factor, VIF) 筛选 CMA。以 4 : 1 比例划分训练集和测试集, 训练集 63 组数据, 测试集 16 组数据。

以训练集和测试集决定系数 (R<sup>2</sup>)、测试集均方根误差 (root mean square error, RMSE)、测试集平均绝对百分比误差 (mean absolute percent error, MAPE) 等指标来评价模型的性能。训练集与测试集 R<sup>2</sup> 差值大于 0.2, 则模型过拟合, 结果不可靠; R<sup>2</sup> 越接近 1, 表明模型的拟合效果越好; RMSE、MAPE 越接近 0, 表明模型的预测性能越强<sup>[22-23]</sup>。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y - y_i)^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (2)$$

$$RMSE = \left[ \sum_{i=1}^n (y - y_i)^2 / n \right]^{1/2} \quad (3)$$

$$MAPE = \sum_{i=1}^n | (y - y_i) / y | / n \quad (4)$$

n 为校正集或验证集的样本数, i ∈ [1, n], y 为参考值, y<sub>i</sub> 为预测值,  $\bar{y}$  为所有样品参考值的平均值

### 2.5 算法建模结果

**2.5.1 PLS 建模结果** 25 个参数的初始模型性能指标: 训练集 R<sup>2</sup> = 0.327, 测试集 R<sup>2</sup> = 0.400, RMSE = 3.172, MAPE = 2.791%。根据模型得到的变量投影重要性 (variable importance for the projection, VIP), 将各参数按照 VIP 值从小到大依次删除并逐一建模, 从表 4 中可知当变量数为 10 时, 模型的性能最好, 此时的训练集 R<sup>2</sup> = 0.384, 测试集 R<sup>2</sup> = 0.460, RMSE = 3.010, MAPE = 2.761%。

**2.5.2 DT 建模结果** 通过网格搜索法获得最优超参数, 决策树的最大深度为 30, 节点划分最小不纯度为 0.5, 并以最优超参数建模, 得到的 25 个参数的初始模型性能指标: 训练集 R<sup>2</sup> = 0.768, 测试集 R<sup>2</sup> = 0.501, RMSE = 2.806, MAPE = 2.734%, 训练集与测试集 R<sup>2</sup> 之差大于 0.2, 模型过拟合。根据各参数的特征重要性 (feature importance, FI), 依次

从小到大剔除变量并逐一建模, 结果见表 5, 当变量数为 11 时, DT 模型的训练集  $R^2=0.780$ , 测试集  $R^2=0.687$ ,  $RMSE=2.223$ ,  $MAPE=2.191\%$ , 模型的拟合效果与预测精度达到最高。

表 4 不同变量个数建立的 PLS 模型的性能和预测性能比较

Table 4 Comparison of PLS model performance and prediction performance established with different numbers of variables

自变量 个数	训练集		测试集	
	$R^2$	$R^2$	RMSE	MAPE/%
25	0.327	0.400	3.172	2.791
24	0.325	0.392	3.198	2.803
22	0.323	0.416	3.131	2.763
14	0.302	0.404	3.168	2.842
10	0.384	0.460	3.010	2.761

表 5 不同变量个数建立的 DT 模型的性能和预测性能比较

Table 5 Comparison of DT model performance and prediction performance established with different numbers of variables

自变量 个数	训练集		测试集	
	$R^2$	$R^2$	RMSE	MAPE/%
25	0.768	0.501	2.806	2.734
22	0.693	0.537	2.703	2.596
11	0.780	0.687	2.223	2.191
6	0.628	0.600	2.512	2.269

**2.5.3 RF 建模结果** 通过网格搜索法获得最优超参数, 决策树的数量为 40, 最大生长深度为 10, 最大特征数量为 7, 并以最优超参数建模, 25 个参数得到的最初模型性能指标: 训练集  $R^2=0.886$ , 测试集  $R^2=0.671$ ,  $RMSE=1.863$ ,  $MAPE=1.722\%$ , 训练集与测试集  $R^2$  之差大于 0.2, 模型过拟合。根据 FI 值, 依次从小到大剔除变量并逐一建模, 结果见表 6。当变量数为 11 时, RF 模型的训练集  $R^2=0.865$ ,

表 6 不同变量个数建立的 RF 模型的性能和预测性能比较

Table 6 Comparison of RF model performance and prediction performance established with different numbers of variables

自变量 个数	训练集		测试集	
	$R^2$	$R^2$	RMSE	MAPE/%
25	0.886	0.671	1.863	1.722
21	0.862	0.699	1.919	1.654
17	0.873	0.653	2.059	1.652
11	0.865	0.854	1.375	1.153
8	0.886	0.671	2.005	1.768

测试集  $R^2=0.854$ ,  $RMSE=1.375$ ,  $MAPE=1.153\%$ , 模型的拟合效果与预测精度达到最高。

**2.5.4 BPNN 建模结果** 通过经验调试法获得最优超参数, 隐藏层数目为 1, 神经元数目为 10, 激活函数为整流线性单元 (rectifier linear unit, RLU), 并以最优超参数建模, 25 个物料属性得到的模型性能指标: 训练集  $R^2=0.362$ , 测试集  $R^2=0.415$ ,  $RMSE=3.230$ ,  $MAPE=2.704\%$ 。根据各参数的排列重要性 (permutation importance, PI), 依次从小到大剔除变量并逐一建模, 结果见表 7, 当变量数为 12 时, BPNN 模型的训练集  $R^2=0.353$ , 测试集  $R^2=0.426$ ,  $RMSE=3.199$ ,  $MAPE=2.685\%$ , 模型的拟合效果与预测精度达到最高。

表 7 不同变量个数建立的 BPNN 模型的性能和预测性能比较

Table 7 Comparison of BPNN model performance and prediction performance established with different numbers of variables

自变量 个数	训练集		测试集	
	$R^2$	$R^2$	RMSE	MAPE/%
25	0.362	0.415	3.230	2.704
21	0.261	0.267	3.168	3.237
17	0.299	0.233	3.701	2.949
12	0.353	0.426	3.199	2.685

**2.5.5 SVR 建模结果** 通过网格搜索法确定最优超参数, 选择径向基核函数 (radial basis function, RBF), 正则化参数为 10, RBF 的  $\gamma$  为 0.1, 并以最优超参数建模, 以 25 个物料属性得到的模型性能指标: 训练集  $R^2=0.999$ , 测试集  $R^2=0.603$ ,  $RMSE=2.733$ ,  $MAPE=2.072\%$ , 训练集与测试集  $R^2$  之差大于 0.2, 模型过拟合。根据 PI 值, 依次从小到大剔除变量并逐一建模, 结果见表 8, 当变量数为 7 时, SVR 模型的训练集  $R^2=0.999$ , 测试集  $R^2=0.660$ ,  $RMSE=2.341$ ,  $MAPE=1.873\%$ , 模型仍然过拟合。

表 8 不同变量个数建立的 SVR 模型的性能和预测性能比较

Table 8 Comparison of SVR model performance and prediction performance established with different numbers of variables

自变量 个数	训练集		测试集	
	$R^2$	$R^2$	RMSE	MAPE/%
25	0.999	0.603	2.733	2.072
15	0.999	0.619	2.624	2.005
10	0.999	0.625	2.586	1.980
7	0.999	0.660	2.341	1.873

### 2.6 不同算法模型最优性能和预测精度比较结果

不同算法建立的最优模型性能和预测精度比较结果见表 9，可以看出，RF 算法的最优预测模型在训练集和测试集上都取得了较大的  $R^2$ ，在测试集上获得了较小的各项误差，模型的稳定性和预测精度更好。RF 包含多棵决策树，从表 9 可以看出 DT 模型也有着较好的拟合效果。将 RF 第 1 棵决策树进行可视化处理，结果见图 3。决策树构建过程是从根节点开始，选择最佳的分裂特征，将数据集划分成 2 个子集，对每个子集递归地重复这个过程，直到满

表 9 不同算法模型性能和预测精度比较

Table 9 Comparison of performance and prediction accuracy of different algorithm models

算法	自变量个数	训练集 $R^2$	测试集		
			$R^2$	RMSE	MAPE/%
PLS	10	0.384	0.460	3.010	2.761
DT	11	0.780	0.687	2.223	2.191
RF	11	0.865	0.854	1.375	1.153
BPNN	12	0.353	0.426	3.199	2.685
SVR	7	0.999	0.660	2.341	1.873

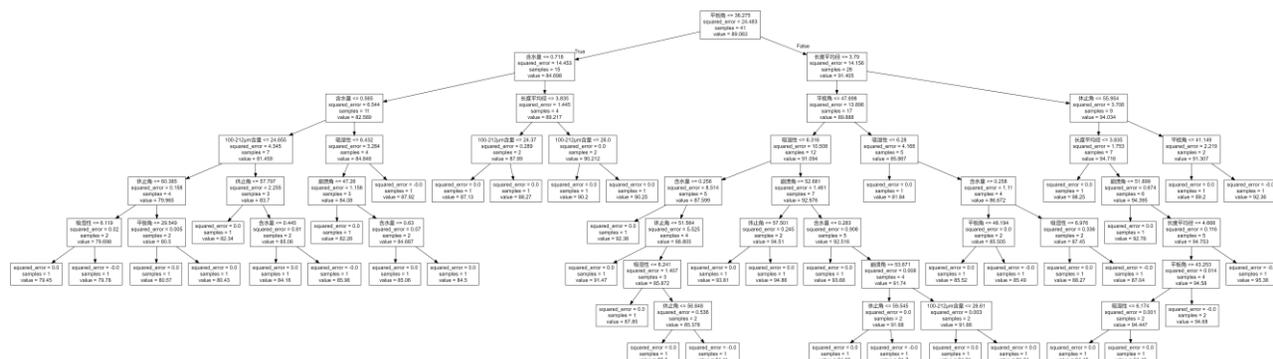


图 3 RF 第 1 棵决策树可视化图

Fig. 3 Visualization of the first decision tree in RF

足停止条件，如节点的样本数量小于阈值或决策树的深度达到预先设定的最大深度，最终形成一棵完整的决策树。

### 2.7 XZG 中间体颗粒溶解性 CMA 的筛选

当变量数为 11 时，RF 模型的拟合与预测效果最好。使用 VIF 值评价以上 11 个物料属性 ( $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $D_c$ 、 $D_a$ 、 $H$ 、 $HR$ 、 $D_{10}$ 、 $F_m$ 、 $D_{21}$ 、 $IH$ ) 的共线性程度，VIF 值大于 10，表明强共线性，结果见表 10。 $D_a$ 、 $D_c$ 、 $IH$ 、 $D_{10}$  的 VIF 值大于 10，存在较强共线性；其余参数的 VIF 值均小于 10，存在较弱共线性。以  $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $HR$ 、 $H$ 、 $F_m$ 、 $D_{21}$  为自变量，建立 RF 简化模型，RF 算法 3 个模型的性能比较见表 11。简化模型的拟合效果与预测精度都较初始模型有所提高，简化模型的拟合效果不如最优模型，但其 RMSE 有所降低，并且 MAPE 为 1.261%，简化模型仍具有较好的预测性能。

为进一步检验简化模型的预测效果，制备混料设计推荐的溶化率最优处方样品，并以其  $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $HR$ 、 $H$ 、 $F_m$ 、 $D_{21}$  为输入，根据得到的预测值计算出相对预测偏差为 1.37% (相对预测偏差 = |实际值 - 预测值| / 实际值)，说明简化模型预测性能良好。比较发现最优处方样品与该批次生产上的 XZG 中

表 10 自变量共线性评价

Table 10 Colinearity evaluation of independent variables

指标	容差	VIF	指标	容差	VIF	指标	容差	VIF
$\alpha$	0.217	4.602	$D_a$	0.049	20.246	$F_m$	0.468	2.139
$\beta$	0.143	6.972	$H$	0.389	2.574	$D_{21}$	0.101	9.912
$\gamma$	0.596	1.677	$HR$	0.635	1.574	$IH$	0.005	202.102
$D_c$	0.004	277.271	$D_{10}$	0.077	13.005			

表 11 RF 模型性能比较

Table 11 Performance comparison of RF

模型名称	训练集		测试集	
	$R^2$	$R^2$	RMSE	MAPE/%
初始模型	0.886	0.671	1.863	1.722
最优模型	0.865	0.854	1.375	1.153
简化模型	0.875	0.793	1.344	1.261

间粉末的 CMA 均有明显差异，最优处方样品的  $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $HR$  明显下降， $H$  变强， $F_m$ 、 $D_{21}$  增大，最优处方样品的溶化率较该批次 XZG 的溶化率显著提高。综上所述， $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $HR$ 、 $H$ 、 $F_m$ 、 $D_{21}$  是影响 XZG 溶解性的 CMA，可以通过控制相关物料属性，提升 XZG 的溶解性。

### 3 讨论

本研究比较了 5 种算法对 XZG 制剂成型工艺

用物料（主药和辅料）溶解性预测模型的性能。结果表明，5 种算法中 RF 算法建立的模型预测效果最好，优化了 PLS 算法预测精度较差的缺点，筛选并验证了  $\alpha$ 、 $\beta$ 、 $\gamma$ 、HR、 $H$ 、 $F_m$ 、 $D_{21}$  是影响 XZG 溶解性的 CMA，可以通过控制制剂成型处方工艺用物料（主药和辅料）的 CMA，提升 XZG 的溶解性，为提升中药颗粒剂的溶解性和品质提供新的思路与方法。

汪盛华等<sup>[6]</sup>基于 5 种制剂品种的中间体粉末与颗粒筛选出影响颗粒溶解性的 CMA 是  $D_c$ 、 $H$ 、SSA、HR、IH、 $I_c$  及 IC。其中  $H$  与 HR 与本研究结果相同。 $D_c$ 、IH、 $I_c$  及 IC 主要体现的是粉末的可压性，其研究样品为 5 个品种，品种间的可压性差异较大，本研究用的均为 XZG 的喷干细粉，差异不大，故筛选出的  $D_a$ 、 $D_c$ 、IH 因为共线性较强而排除。SSA 代表粉末的比表面积，Salish 等<sup>[24]</sup>研究发现对于具有崎岖表面的药物颗粒，大部分表面不会对溶解过程产生积极贡献，对于溶解具有意义的是能增加在溶解介质中表现为较小、完全分离的颗粒数量的特定表面积，可以理解为溶解过程与水中小颗粒的作用力和粒径有关。

本研究筛选出的  $\alpha$ 、 $\beta$ 、 $\gamma$  的大小反映了粉末粒子间的相互作用力大小， $F_m$ 、 $D_{21}$  都是与粒径有关的参数。 $\alpha$ 、 $\beta$ 、 $\gamma$  越大，则粉末粒子间的作用力越强，粉末的流动性也就越差，影响颗粒的成型<sup>[25]</sup>。若药粉粒子间的作用力过强，会影响溶解过程中水分的渗入，影响颗粒的溶解性。可以考虑添加流动性强的水溶性辅料来提升 XZG 的溶解性，既可以改善颗粒流动性，又可以充当填充剂减少药粉间的相互作用力。HR、 $H$  直接代表了粉末与水分之间的作用力，含水量越低，吸湿性越强，则越容易被水润湿，更易溶解。提示在喷干、混料等生产过程中将含水量纳入质控指标，控制喷干粉和加入辅料的含水量，进而控制制粒前一步混合粉的含水量。 $F_m$ 、 $D_{21}$  都是与粒径有关的参数，可以考虑用超微粉碎技术<sup>[26]</sup>等粉体改性技术改变粉末粒径。因此，提升 XZG 溶解性品质可以从添加辅料改性、改变粉体表面性质与改变粉体自身结构这 3 个角度出发，以期形成改善方案，为 XZG 的生产工艺变更提供参考。

在筛选药辅比的过程中发现，溶化率与药辅比曲线存在拐点，提示辅料改变颗粒溶解性可能与药辅比有关，后续将开展辅料配比与 CMA 相关性研究，以期形成控制 CMA 的辅料配比优化方案，并

从物料属性角度解释辅料改善颗粒溶解性的机制。本研究所选数据量远少于生产过程产生的大批量数据，在后续研究中还需要继续收集更多的数据来进一步验证、更新、维护和优化模型，提高模型预测精确度，提升 XZG 制剂工艺过程的智能化水平。期望在此基础上，通过对更多的中药颗粒剂品种的研究和应用，以及对引湿性、流动性等不同角度的研究，构建相关共性技术平台，为推动中药颗粒剂制剂品质的提升提供示范。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] 于佳琦, 徐冰, 黄雨妍, 等. 基于在线浊度传感器的中药颗粒剂溶化性评价和分类研究 [J]. 中国中药杂志, 2020, 45(2): 259-266.
- [2] 贾晓斌, 杨冰, 封亮, 等. 中药药剂学研究前沿: 组方制剂技术基础与关键科学问题 [J]. 药学学报, 2018, 53(12): 1943-1953.
- [3] 汪盛华, 秦春娟, 安双凤, 等. 水提干法制粒的中药配方颗粒溶化性与粉体物理属性相关性研究 [J]. 中草药, 2023, 54(5): 1439-1448.
- [4] 王晴, 徐冰, 王芬, 等. 桂枝茯苓胶囊内容物吸湿性预测建模研究 [J]. 中国中药杂志, 2020, 45(2): 242-249.
- [5] 陈琪, 徐芳芳, 张欣, 等. 基于不同算法对桂枝茯苓胶囊内容物吸湿性预测建模研究 [J]. 中草药, 2021, 52(11): 3216-3223.
- [6] 汪盛华, 闫明, 徐芳芳, 等. 基于物料粉体学性质对 5 种制剂品种颗粒溶化性的相关性研究 [J]. 中草药, 2022, 53(22): 7082-7090.
- [7] 吴晓倩, 权丽丽, 陈诚, 等. 基于大数据决策树算法的学生成绩分析与预测模型仿真 [J]. 电子设计工程, 2020, 28(24): 138-141.
- [8] 王民, 杨金路, 王鑫, 等. 基于随机森林算法的泥页岩岩相测井识别 [J]. 地球科学, 2023, 48(1): 130-142.
- [9] 宋丽芳, 廖桂平, 陈敏, 等. 基于机器学习的油菜叶片水分含量高光谱估测 [J]. 中国农业科技导报, 2024, 26(5): 110-119.
- [10] 秦春娟, 闫明, 王振中, 等. 基于多品种水-醇双提物的中药粉体性质影响颗粒吸湿性的研究 [J]. 中草药, 2023, 54(4): 1120-1126.
- [11] 熊志伟, 宁汝曦, 赵樱霞, 等. 中药制剂前物料的性质体系及其表征技术研究 [J]. 药学学报, 2021, 56(8): 2048-2058.
- [12] 李远辉, 伍振峰, 李延年, 等. 基于粉体学性质分析浸膏干燥工艺与中药配方颗粒制粒质量的相关性 [J]. 中草药, 2017, 48(10): 1930-1935.
- [13] 崔向龙, 徐冰, 张毅, 等. 质量源于设计在银杏叶片制

- 粒工艺中的应用 (I): 颗粒粉体学性质综合评价 [J]. 中国中药杂志, 2017, 42(6): 1037-1042.
- [14] 戴胜云. 中药直接压片处方智能设计方法研究 [D]. 北京: 北京中医药大学, 2019.
- [15] 中国药典 [S]. 四部. 2020: 485.
- [16] European Pharmacopeia. Dosage Forms Monographs [S]. 2013: 346-348.
- [17] 张慧, 陈燕, 汪佳楠, 等. 指纹图谱技术在中药配方颗粒质量评价及过程控制中的应用 [J]. 中国中药杂志, 2018, 43(19): 3822-3827.
- [18] 刘东方, 赵丽娜, 李银峰, 等. 中药指纹图谱技术的研究进展及应用 [J]. 中草药, 2016, 47(22): 4085-4094.
- [19] 张毅, 徐冰, 孙飞, 等. 中药提取物粉末物理指纹谱研究及应用 [J]. 中国中药杂志, 2016, 41(12): 2221-2227.
- [20] 毕映燕, 李俊江, 李季文, 等. 基于质量源于设计 (QbD) 理念的祛寒逐风颗粒制备工艺及其物理指纹图谱研究 [J]. 中草药, 2021, 52(19): 5891-5898.
- [21] 杨岩涛, 吴春英, 刘文龙, 等. 不同相似度法对当归补血汤指纹图谱分析的比较研究 [J]. 中华中医药杂志, 2013, 28(5): 1431-1435.
- [22] 刘元铭, 王振华, 王涛, 等. 热轧带钢出口凸度数据驱动建模及智能化预测分析 [J]. 中国机械工程, 2020, 31(22): 2728-2733.
- [23] 郑伟达, 张惠然, 胡红青, 等. 基于不同机器学习算法的钙钛矿材料性能预测 [J]. 中国有色金属学报, 2019, 29(4): 803-809.
- [24] Salish K, So C, Jeong S H, *et al.* A refined thin-film model for drug dissolution considering radial diffusion - simulating powder dissolution [J]. *Pharm Res*, 2024, 41(5): 947-958.
- [25] 安双凤, 闫明, 王团结, 等. 基于水-醇双提方式下中药复方粉末影响颗粒流动性的关键物料属性辨识研究[J]. 中草药, 2023, 54(03): 789-797.
- [26] 刘惠, 仇敏, 刘潜, 等. 儿童友好型小儿复方鸡内金速溶散的设计与评价 [J]. 中草药, 2024, 55(14): 4670-4678.

[责任编辑 郑礼胜]