

基于级联容错模型的桂枝茯苓胶囊软材细粉质量控制研究

钟小慧^{1,2}, 童 枫^{1,3}, 候化蕊^{1,3}, 张永超^{1,3}, 王团结^{1,3}, 章晨峰^{3,4}, 徐芳芳^{1,3*}, 刘恒旭^{1,3*}

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 211112

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

4. 江苏省海洋药物和现代中药创制重点实验室, 江苏康缘药业股份有限公司, 江苏 连云港 222047

摘要: 目的 针对桂枝茯苓胶囊(Guizhi Fuling Capsules, GFC)生产过程中软材细粉批内颜色不均可能预示的功效成分分布不均, 构建基于色差特征的“级联容错模型”, 探索一种低成本中间体质控新路径。方法 收集 5 批共 177 个软材细粉样本, 以 HPLC 法测得的芍药苷含量按 2:6:2 比例划分为低、中、高等级。以传统三分类模型为基准, 对比 126 种算法组合, 构建一级算法预警协同二级物理判向的级联容错质控模型; 引入外部验证评估泛化能力, 并依据模型结果进行质控实验; 最后利用近红外光谱(near-infrared spectroscopy, NIRS)技术进行效能上限对标。结果 相关性分析确立芍药苷为核心分级指标, 其含量与红绿色度(a^*)呈显著性正相关($r=0.64$, $P<0.001$)。分析结果表明, 基准模型准确率仅为 70.34% (RSD 为 12.70%), 难以满足质控要求; 而级联容错模型工业总准确率为 78.62%, 稳健性显著增强 (RSD 为 6.79%), 其外部验证准确率为 80.56%。质控实验显示, 混合样品的芍药苷含量 RSD, 由随机混合的 7.74% 降至 2.34%, 物料均匀度显著增加, 总混工序工时与载荷同步降低。与 NIRS (准确率为 90.68%, RSD 为 6.22%) 的对标结果表明, 色差仪性价比高, 可用于批量初筛; NIRS 更精准, 适用于高精度分级。结论 构建的级联容错模型实现了算法预警与物理纠偏的协同赋能, 有效提升了低成本光学传感器的判别稳健性, 为未来中药生产过程中批内中间体功效成分的质控提供了新策略。

关键词: 桂枝茯苓胶囊; 软材细粉; 色差特征; 级联容错; 中间体质控; 近红外光谱; 效能对标; 工业应用

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2026)18-7321-13

DOI: 10.7501/j.issn.0253-2670.2026.18.020

Research on quality control of soft material fine powder of Guizhi Fuling Capsules based on cascade fault-tolerant model

ZHONG Xiaohui^{1, 2}, TONG Feng^{1, 3}, HOU Huarui^{1, 3}, ZHANG Yongchao^{1, 3}, WANG Tuanjie^{1, 3}, ZHANG Chenfeng^{3, 4}, XU Fangfang^{1, 3}, LIU Hengxu^{1, 3}

1. State Key Laboratory of Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacture Technology, Jiangsu Kanion Pharmaceutical Co., Ltd., / Nanjing University of Chinese Medicine, Nanjing 211112, 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

4. Jiangsu Key Laboratory on Innovation for Marine Medicine and Modern Chinese Medicine, Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222047, China

Abstract: Objective To address the potential uneven distribution of active ingredients indicated by uneven color within batches of soft material fine powder during the production process of Guizhi Fuling Capsules (GFC, 桂枝茯苓胶囊), a “cascade fault-tolerant model” was constructed based on color difference characteristics to explore a new low-cost approach for intermediate quality control.

Methods A total of 177 samples of soft material fine powder from five batches were collected. The paeoniflorin content, determined by HPLC, was divided into low, medium, and high grades at a ratio of 2:6:2. Based on a traditional three-class model, a cascade fault-

收稿日期: 2026-03-31

基金项目: 国家长三角科技创新共同体联合攻关项目 (2023CSJGG1700); 连云港市重点研发计划 (产业前瞻与关键核心技术) (CG2320)

作者简介: 钟小慧, 女, 硕士研究生, 研究方向为药物制剂与产品研发。E-mail: 15680541248@163.com

*通信作者: 徐芳芳, 博士, 副主任药师, 研究方向为中药智能制造。E-mail: 879164331@qq.com

刘恒旭, 助理级工程师, 硕士研究生, 研究方向为中药制药过程新技术。E-mail: 2447566454@qq.com

tolerant quality control model was constructed by comparing 126 algorithm combinations, which involved a primary algorithm for early warning and a secondary physical direction for correction. External validation was introduced to assess the generalization ability, and quality control experiments were conducted based on the model results. Finally, near-infrared spectroscopy (NIRS) technology was used for benchmarking against the upper limit of efficacy. **Results** Correlation analysis established paeoniflorin as the core grading indicator, with its content showing a significant positive correlation with red-green chroma a^* ($r = 0.64$, $P < 0.001$). The analysis indicated that the accuracy of the benchmark model was only 70.34% (RSD was 12.70%), which was difficult to meet quality control requirements. However, the cascade fault-tolerant model achieved an overall industrial accuracy of 78.62%, with significantly enhanced robustness (RSD was 6.79%), and its external validation accuracy was 80.56%. Quality control experiments showed that the RSD of paeoniflorin content in mixed samples decreased from 7.74% in random mixing to 2.34%, indicating a significant increase in material uniformity, and a simultaneous reduction in total mixing process time and load. Benchmarking results with NIRS (accuracy of 90.68%, RSD was 6.22%) indicated that the colorimeter had a high cost-performance ratio and could be used for batch screening; NIRS was more accurate and suitable for high-precision grading. **Conclusion** The cascade fault-tolerant model constructed in this study achieved the synergistic empowerment of algorithm early warning and physical correction, effectively enhancing the discriminative robustness of low-cost optical sensors. This provides a new strategy for quality control of active ingredients in batches of intermediates during the production process of traditional Chinese medicine in the future.

Key words: Guizhi Fuling Capsules; soft material fine powder; color difference feature; cascaded fault tolerance; intermediate product quality control; near-infrared spectroscopy; performance benchmarking; industrial application

2026 年 3 月 1 日起施行的《中药生产监督管理专门规定》^[1]及相关政策解读明确鼓励企业探索生产过程中的均一化处理,通过对冲等手段缩减批间质量差异。然而,批间质量的稳态高度依赖于批内质量的均一^[2]。现行指导原则多侧重于原料饮片或中药提取物批间均化,对因复杂物理化学作用导致的制剂中间体批内不均尚缺乏明确的质控方案。

桂枝茯苓胶囊(Guizhi Fuling Capsules, GFC)作为妇科疾病的临床常用中成药,其疗效的一致性高度依赖于工艺中间体功效成分含量的稳定性^[3-5]。在湿法制粒工序中,软材细粉作为关键中间体,其是否均一直接影响最终颗粒的微观形态与整体。实际生产发现,软材细粉存在批内颜色不均现象,这往往反映功效成分或辅料分布的潜在波动。若软材细粉的含量波动过大,不仅会加大后续总混工序的生产载荷与工时,还可能超出混合工艺的调控极限。然而,传统的 HPLC 法检测滞后,难以实现“质量源于设计(quality by design, QbD)”理念的事前风险防控。因此,将均化质控逻辑由批间原料端延伸至批内制剂端,已成为中药智能制造亟待解决的课题。

“辨色论质”是中药传统质量评价的核心^[6]。色差仪通过量化明度(L^*)、红绿色度(a^*)、黄蓝色度(b^*)等参数,已在多种中药材及其饮片的质量评价中展现出良好前景^[7-11]。然而,现有研究多局限于原料药材或成品饮片的鉴别^[12],尚未有效延伸至制剂工艺中间体的监测^[13];且关联分析多集中于单一组

分^[10,14],缺乏针对复杂生产环境下的应用策略。制粒环节的软材细粉属于典型的多组分混合物料。其色差特征除受功效成分影响外,还与物料包裹形态、水分变化及颗粒大小密切相关。由于物料组成复杂、分布不均,颜色与化学成分的对应关系复杂多变,对软材细粉的质量控制提出了严苛挑战。

对此,本研究创新性地构建了一种基于色差特征的“级联容错质控模型”,该模型通过一级算法预警与二级物理判别的协同,旨在弥补传统分类模型在处理小样本、分布连续数据时稳健性不足的缺陷,实现对中间体等级的精准预判。同时,在同一评价体系下对标近红外光谱(near-infrared spectroscopy, NIRS)技术,探索低成本光学传感器与高精度光谱设备的效能边界。本研究不仅为中药中间体质控提供了新路径,也为低成本检测在工业现场的稳健应用提供了理论支撑。

1 仪器与材料

1.1 仪器

Antaris II 型傅里叶近红外变换光谱仪,配有积分球漫反射采样系统、Result 光谱采集软件,美国 Thermo 公司;CM-5 型分光测色计,日本柯尼卡美能达株式会社;ME104E 型电子天平,梅特勒-托利多仪器(上海)有限公司;Ultimate 3000 型高效液相色谱仪,赛默飞世尔科技(中国)有限公司;Milli-Q IQ7000 型超纯水系统,默克化工技术(上海)有限公司;KQ-500DE 型数控超声波清洗器,昆山超声仪器有限公司;H1650-W 型高速台式离心机,湖

南湘仪实验室仪器开发有限公司。

1.2 材料

没食子酸、芍药苷、肉桂酸、苯甲酰芍药苷、桂皮醛及丹皮酚对照品，中国食品药品检定研究院，供含量测定用，批号分别为 110831-201906、110736-202548、100419-201703、110786-201604、110710-202022、110708-201908，质量分数分别为 91.5%、98.9%、99.9%、98.8%、99.5%、99.8%；苯甲酰芍药苷对照品，批号为 DST211210-053，质量分数为 98%，成都德斯特生物技术有限公司；三氟乙酸，批号为 02007184，质量分数为 99.5%，北京迈瑞达科技有限公司；乙腈，色谱纯，西格玛里奥里奇（上海）贸易有限公司；水，超纯水，自制；其余试剂均为分析纯。

1.3 数据处理软件

采用 Unscrambler (version 11, 挪威 Camo Analytics 公司) 软件进行光谱预处理，采用 Jupyter Notebook v7.5.4 进行建模与数据处理，Jupyter Notebook 运行环境基于 Python 3.13.9 构建。

2 方法与结果

2.1 样本收集

GFC 软材细粉样本均由江苏康缘药业股份有限公司提供，采集 5 批样本（批号 Z250804、Z250805、Z250906、Z250907、Z251008）数据，5 个批次样本数量分别为 33、34、36、37、37 个，共计 177 个 GFC 软材细粉样本。

2.2 数据采集

2.2.1 色差指标的测定 使用分光测色计对软材细粉进行客观评价。测量条件：光源 D65，标准观察角 10° ，照明孔径 (Φ) 30 mm。取 10 g 左右样本，置于配备的比色皿里，轻轻压实，对仪器进行黑白板校正以后，进行样本测量^[7]，采集指标明度 (L^* ，值越大表示亮度越高)、红绿色度 (a^* ，值越大颜色越红) 和黄蓝色度 (b^* ，值越大颜色越黄^[11])。每个样本扫描 3 次，平均值用于分析。

为了放大颜色细微差异，增强与成分含量的相关性，本实验在保留原始 L^* 、 a^* 、 b^* 数据的基础上，参考数字图像技术中通过颜色参数比值、乘积扩展特征的研究范式^[15]，构建了 21 维色差指标，分别为色差值 [E^* ， $E^*=(L^{*2}+a^{*2}+b^{*2})^{1/2}$ ，代表颜色间的距离^[16]]、色度值 [C^* ， $C^*=(a^{*2}+b^{*2})^{1/2}$ ，色度值越高颜色越纯^[17]]、色调值 [H^* ， $H^*=\arctan(b/a)$ ，色调值从 0° 到 180° 依次表示紫红色 (0°)、红色、

橙红色、橙黄色、黄色 (90°)、黄绿色、绿色、蓝绿色 (180°)^[16]]、 $L^*/(L^*+a^*+b^*)$ 、 $a^*/(L^*+a^*+b^*)$ 、 $b^*/(L^*+a^*+b^*)$ 、 $L^*/(L^*+C^*+H^*)$ 、 $C^*/(L^*+C^*+H^*)$ 、 $H^*/(L^*+C^*+H^*)$ 、 $L^*\times a^*$ 、 $L^*\times b^*$ 、 $a^*\times b^*$ 、 a^*/b^* 、 a^*/L^* 、 b^*/L^* 、 L^{*2} 、 a^{*2} 、 b^{*2} 。

2.2.2 色差仪方法学考察 同一样品重复测量 6 次，其 L^* 、 a^* 、 b^* 的 RSD 分别为 0.02%、0.09%、0.08%，表明该仪器精密性良好；取样品 6 份测量，其 L^* 、 a^* 、 b^* 的 RSD 分别为 0.07%、0.16%、0.10%，表明该方法重复性良好；分别于第 0、2、4、8、12、24 h 测定，其 L^* 、 a^* 、 b^* 的 RSD 分别为 0.06%、0.08%、0.11%，表明该方法稳定性良好。

2.2.3 近红外 NIRS 光谱采集 取 10 g 左右样本，置于配备的样本杯里，轻轻压实，采用积分球漫反射方式采集 NIRS。扫描范围为 $10\ 000\sim4\ 000\text{ cm}^{-1}$ ，分辨率为 8 cm^{-1} ，2 倍增益，扫描次数 64 次，以空气为背景，每小时扫描 1 次背景。每个样本扫描 3 次，平均值用于分析。采集的 NIRS 图见图 1。

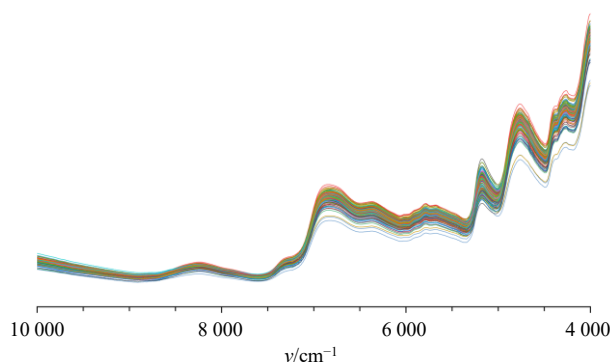


图 1 177 个软材样本的 NIRS 图

Fig. 1 NIRS spectra of 177 soft material samples

2.2.4 功效成分含量测定 李家春等^[18]依据 GFC 生产工艺，确定以没食子酸、芍药苷、肉桂酸、苯甲酰芍药苷、桂皮醛和丹皮酚共 6 个功效成分为检测指标。

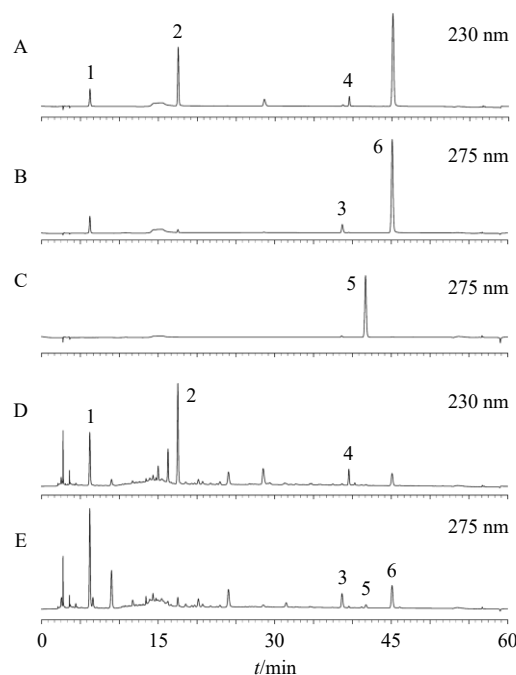
(1) 对照品溶液的制备：精密称取 5 种对照品适量，分别加入 50% 甲醇溶解并稀释制成含没食子酸 $942.45\text{ }\mu\text{g/mL}$ 、芍药苷 $1\ 978.99\text{ }\mu\text{g/mL}$ 、肉桂酸 $1\ 031.47\text{ }\mu\text{g/mL}$ 、苯甲酰芍药苷 $1\ 966.86\text{ }\mu\text{g/mL}$ 、丹皮酚 $1\ 998.00\text{ }\mu\text{g/mL}$ 的各单一对照品储备液。依次精密移取没食子酸对照品储备液 4 mL、芍药苷对照品储备液 10 mL、肉桂酸对照品储备液 1 mL、苯甲酰芍药苷对照品储备液 1 mL、丹皮酚对照品储备液 10 mL，置于 100 mL 量瓶中，加 50% 甲醇稀释，定容，摇匀，即得混合对照品溶液。

精密称取桂皮醛对照品适量, 加入乙醇制成含桂皮醛 30 190.29 $\mu\text{g/mL}$ 对照品储备液。精密移取桂皮醛对照品储备液 0.1 mL, 置于 100 mL 量瓶中, 加 50% 甲醇稀释, 定容, 摇匀, 即得桂皮醛对照品溶液。

(2) 供试品溶液的制备: 称取 0.5 g 药材细粉, 精密称定, 置于 250 mL 锥形瓶中, 用 25 mL 50% 甲醇溶解, 超声 (功率 250 W、频率 40 Hz) 30 min, 放冷, 摇匀, 12 000 r/min 条件下离心 (离心半径 6 cm) 10 min, 上清液用 0.45 μm 有机微孔滤膜滤过, 即得供试品溶液。

(3) 色谱条件: 色谱柱为 Waters Symmetry C₁₈ 柱 (250 mm $\times$ 4.6 mm, 5 μm); 流动相为 0.02% 三氟乙酸水溶液-乙腈, 梯度洗脱: 0~5 min, 5% 乙腈; 5~10 min, 5%~17% 乙腈; 10~20 min, 17%~19% 乙腈; 20~28 min, 19%~26% 乙腈; 28~50 min, 26%~48% 乙腈; 50~51 min, 48%~88% 乙腈; 51~56 min, 88% 乙腈; 56~57 min, 88%~5% 乙腈; 57~60 min, 5% 乙腈; 体积流量 1 mL/min; 柱温 30 $^{\circ}\text{C}$; 检测波长 230 nm (没食子酸、芍药苷、苯甲酰芍药苷)、275 nm (肉桂酸、桂皮醛及丹皮酚); 进样量 10 μL 。对照品及样品色谱图见图 2。

(4) 线性关系考察: 精密称取 5 种对照品适量, 分别加入 50% 甲醇溶解并稀释制成含没食子酸 94.25 $\mu\text{g/mL}$ 、芍药苷 494.75 $\mu\text{g/mL}$ 、肉桂酸 25.79 $\mu\text{g/mL}$ 、苯甲酰芍药苷 49.17 $\mu\text{g/mL}$ 、丹皮酚 499.50 $\mu\text{g/mL}$ 的混合对照品溶液 1。精密称取桂皮醛对照品适量, 加入乙醇制成含桂皮醛 75.47 $\mu\text{g/mL}$ 的对照品溶液 2。分别精密吸取混合对照品溶液 1 各 1.0、2.0、3.2、4.0、4.8、10.0 mL, 置于 10 mL 量瓶中, 加 50% 甲醇稀释至刻度, 摇匀, 即得系列混合对照品溶液。分别精密吸取对照品溶液 2 各 1.0、2.0、3.2、4.0、4.8、10.0 mL, 置于 10 mL 量瓶中, 加 50% 甲醇稀释至刻度, 摇匀, 即得系列桂皮醛对照品溶液。分别吸取上述系列对照品溶液 10 μL , 注入高效液相色谱仪, 按“2.2.4 (3)”项下色谱条件进行测定, 以对照品质量浓度为横坐标 (X), 峰面积为纵坐标 (Y) 进行线性回归, 得回归方程分别为没食子酸 $Y=402.1X-0.2067$, $r=0.9999$, 线性范围 9.42~94.25 $\mu\text{g/mL}$; 芍药苷 $Y=221.16X-0.5035$, $r=0.9999$, 线性范围 49.47~494.75 $\mu\text{g/mL}$; 肉桂酸 $Y=1580.6X-0.2817$, $r=0.9999$, 线性范围 2.58~25.79 $\mu\text{g/mL}$; 苯甲酰芍药苷 $Y=389.56X-0.0756$,



1-没食子酸; 2-芍药苷; 3-肉桂酸; 4-苯甲酰芍药苷; 5-桂皮醛; 6-丹皮酚。
1-gallic acid; 2-paeoniflorin; 3-cinnamic acid; 4-benzoylpaeoniflorin; 5-cinnamaldehyde; 6-paeonol.

图 2 混合对照品 (A、B)、桂皮醛对照品 (C) 和供试品 (D、E) 的 HPLC 图

Fig. 2 HPLC chromatograms of mixed reference substances (A, B), cinnamaldehyde reference substance (C), and test samples (D, E)

$r=0.9999$, 线性范围 4.92~49.17 $\mu\text{g/mL}$; 桂皮醛 $Y=1372.9X+1.0721$, $r=0.9993$, 线性范围 7.55~75.48 $\mu\text{g/mL}$; 丹皮酚 $Y=843.08X-4.0754$, $r=0.9998$, 线性范围 49.95~499.50 $\mu\text{g/mL}$ 。

(5) 精密度考察: 精密吸取混合对照品溶液和桂皮醛对照品溶液, 按照“2.2.4 (3)”项下色谱条件连续进样测定 6 次, 结果没食子酸、芍药苷、肉桂酸、苯甲酰芍药苷、桂皮醛和丹皮酚峰面积的 RSD 依次为 0.44%、0.38%、0.58%、0.48%、0.05%、0.41%, 结果表明该仪器精密度良好。

(6) 稳定性考察: 按照“2.2.4 (2)”项下方法制备供试品溶液 (批号 Z250805), 按照“2.2.4 (3)”项下色谱条件分别在 0、4、8、12、16、20、24 h 进样测定, 记录峰面积, 结果没食子酸、芍药苷、肉桂酸、苯甲酰芍药苷、桂皮醛和丹皮酚峰面积的 RSD 分别为 0.47%、0.14%、0.21%、0.13%、0.27%、0.10%, 结果表明供试品溶液在 24 h 内稳定性良好。

(7) 重复性考察: 平行制备 6 份供试品溶液 (批号 Z250805), 按照“2.2.4 (3)”项下色谱条件计算

样品中各成分的含量,结果没食子酸、芍药苷、肉桂酸、苯甲酰芍药苷、桂皮醛和丹皮酚质量分数的 RSD 分别为 1.06%、0.89%、0.86%、0.82%、0.84%、0.75%,结果表明该方法重复性良好,符合要求。

(8) 加样回收率考察:精密称取已测知各指标成分含量的样品 0.25 g (批号 Z250805) 置于 250 mL 锥形瓶中精密加入混合对照品溶液和桂皮醛对照品溶液,设置 50%、100%、200% 3 个加样水平,按照“2.2.4 (2)”项下制备供试品溶液,每质量浓度平行试验 3 份,按照“2.2.4 (3)”项下色谱条件进样测定,计算加样回收率,结果没食子酸、芍药苷、肉桂酸、苯甲酰芍药苷、桂皮醛和丹皮酚的平均加样回收率分别为 100.42%、99.62%、99.02%、98.91%、102.76%、99.08%,RSD 分别为 1.94%、1.97%、1.97%、1.95%、1.51%、1.49%,结果表明该实验加样回收率结果可靠。

2.3 机器学习算法

本研究分别采用支持向量机 (support vector machine, SVM)、逻辑回归 (logistic regression, LR)、K-最近邻分类 (K-nearest neighbor, KNN)、朴素贝叶斯 (naive Bayes, NB)、随机森林 (random forest, RF)、人工神经网络 (artificial neural network, ANN) 和极端梯度提升 (extreme gradient boosting, XGBoost) 共 7 种机器学习方法对软材细粉等级进行划分。SVM 作为高性能监督学习算法,借助核函数在高维空间构建最优超平面实现正负样本分类,是分类问题的经典解决方案^[19]。LR 是面向二分类任务的线性模型,通过拟合特征与标签的关联来预测样本归属于某一类别的概率^[20]。KNN 为非参数统计方法,核心是依据特征空间中 k 个最邻近样本的类别占比,判定待分类样本所属类别,也可用于回归分析^[21]。NB 基于贝叶斯概率原理,通过计算待分类项在各类别下的条件概率,将其归为概率最大的类别^[22]。RF 是集成学习类监督分类算法,通过构建多棵决策树并采用多数投票整合结果,鲁棒性强且能缓解过拟合,适配高维复杂数据集^[23]。ANN 模拟人脑神经元结构,通过多层神经元的连接与权重迭代调整,实现对复杂非线性数据模式的建模,具备自适应和自主学习特性^[24-25]。XGBoost 是高效的梯度提升框架,通过迭代训练决策树弱学习器并优化目标函数,结合正则化项控制模型复杂度,避免过拟合,泛化能力强,适用于高维非线性数据的分类与回归任务^[26]。

2.4 模型评价

混淆矩阵是评估模型分类性能的重要工具,常用准确率、F1 分数等为其评价指标^[27-29]。鉴于本研究将基准模型转化为级联模型,预警样本会在二级模型中进行 2 次判别,传统静态分类指标平等看待各类分类错误,难以体现级联容错模型在生产质控的容错价值,故在准确率、F1 分数 (F1-score) 基础上,补充构建了工业总准确率 (overall industrial accuracy, OIA) 作为场景化评价指标,形成完整评价体系。

(1) 准确率:反映模型整体识别能力的通用指标,定义为正确预测的样本数占样本总数的比例。

$$\text{准确率} = (N_A + N_C) / N \quad (1)$$

N_A 为中级样本的正确识别数, N_C 指预警样本被准确预警且判别正确的样本数, N 为样本总数

(2) F1 分数:评估模型综合性能。在级联模型中,该指标能有效权衡中级与预警样本的判别精度,避免样本不平衡对单一指标的干扰。

$$\text{F1 分数} = 2 \times \text{精确率} \times \text{召回率} / (\text{精确率} + \text{召回率})$$

(2)

$$\text{精确率} = N_C / (N_B + N_C + N_E) \quad (3)$$

$$\text{召回率} = N_C / (N_C + N_D) \quad (4)$$

N_B 指被一级模型判为预警且被二级模型判别正确方向的中含量样本数 (良性误报,后续工艺可通过总混实现质量可控), N_D 为一级模型中被误判为中级的预警样本数, N_E 为被一级模型判为预警且被二级模型判别错误方向的中含量样本数

(3) OIA: 为科学评估级联模型对制剂工艺的实际贡献,本研究构建了 OIA。

$$\text{OIA} = (N_A + N_B + N_C) / N \quad (5)$$

2.5 色差特征与功效成分的关联性分析

2.5.1 功效成分的同步消长规律分析 针对生产现场观察到的软材细粉批内颜色波动问题,本研究引入色差技术,旨在探究外观色差量化特征与内在功效成分含量的消长规律。软材细粉的统计分析见表 1,软材细粉色差指标与各功效成分含量在批内及批间均呈现不同程度的波动,且整体 RSD 普遍高于平均批内 RSD。功效成分中,即便波动最小的没食子酸,其整体 RSD (15.30%) 亦显著高于色差指标的最大波动值 (5.60%);桂皮醛等挥发性成分波动更是高达 50.89%。这表明化学含量的偏移远比表观颜色的变化剧烈,软材细粉批内和批间工艺稳定性仍有较大优化空间。

表 1 色差特征及各功效成分含量的统计分析

Table 1 Statistical analysis of color difference characteristics and content of each functional component

统计量	L^*	a^*	b^*	质量分数/(mg·g ⁻¹)					
				没食子酸	芍药苷	肉桂酸	苯甲酰芍药苷	桂皮醛	丹皮酚
极大值	53.89	14.99	32.30	8.05	30.08	0.99	3.52	0.73	3.29
极小值	40.65	11.22	27.51	3.84	9.19	0.34	1.19	0.04	0.32
极差	13.24	3.77	4.79	4.21	20.90	0.64	2.33	0.69	2.97
均值	46.07	13.31	30.13	5.96	19.20	0.68	2.34	0.27	1.51
平均批内 RSD/%	4.54	4.30	1.94	11.46	17.34	15.99	16.62	37.51	27.44
整体 RSD/%	5.60	5.56	4.05	15.30	18.80	18.76	19.29	50.89	38.17

2.5.2 色差特征与功效成分的 Pearson 相关性研究
利用 Pearson 相关性分析在 21 维色差特征中筛选出与各成分关联最紧密的前 9 位色差指标（图 3）。结果显示，除原始色度值外， a^{*2} 、 $a^* \times b^*$ 、 b^*/L^* 、 $C^*/(L^*+C^*+H^*)$ 、 $L^*/(L^*+a^*+b^*)$ 、 $b^*/(L^*+a^*+b^*)$ 、 a^*/L^* 、 C^* 、 $a^*/(L^*+a^*+b^*)$ ，等复合指标展现出更强的表征能力。其中，明度 L^* 与各功效成分含量均呈极显著负相关（ $P<0.001$ ），而红绿色度 a^* ，黄蓝色度 b^* 及多数衍生指标则表现为显著正相关，其中，丹皮酚的色差关联特性最为突出。这一规律证实了软材细粉的表现颜色漂移能灵敏映射其内在成分的

波动，为色差技术在生产过程中的高频监控提供了物理依据。

关于分级体系的建立，虽《中国药典》2025 年版对胶囊内容物总混颗粒的芍药苷及丹皮酚含量有明确限定，但对软材细粉这一中间状态尚无硬性标准；本研究通过关联分析筛选核心评价指标（图 4），结果表明芍药苷与没食子酸、肉桂酸及苯甲酰芍药苷等均呈现极显著的正相关（ $r \geq 0.92$ ， $P<0.001$ ），其表征多成分整体波动的灵敏度显著优于丹皮酚（ r 仅为 0.54~0.70）；同时作为该体系中丰度最高的成分，芍药苷的含量波动能有效映射混合

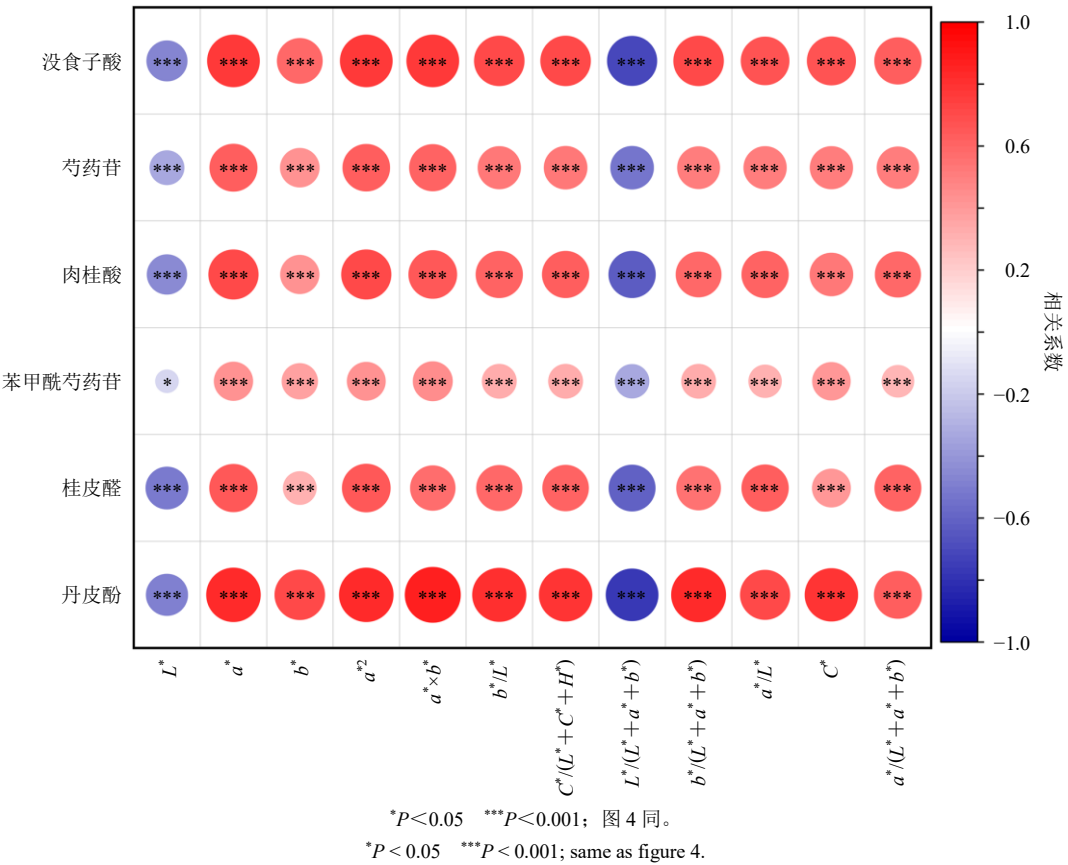


图 3 功效成分与精选色差指标的相关性热图

Fig. 3 Correlation heatmap of active ingredients and selected color difference indicators

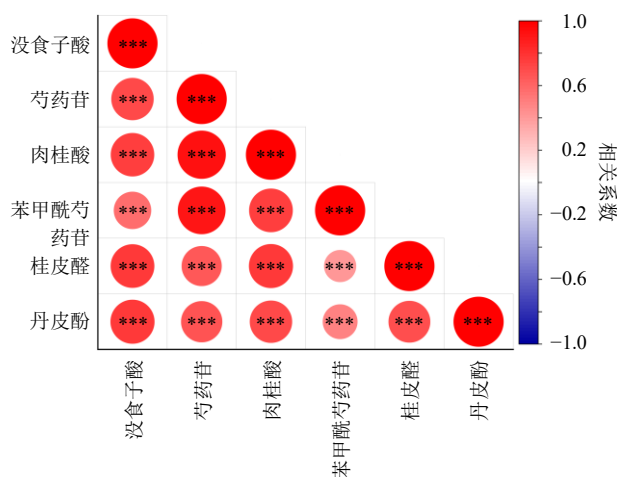


图 4 软材细粉 6 种功效成分相关性热图

Fig. 4 Heat map of correlation between six functional components of soft material fine powder

过程中的工况变化，是监测中间体批内质量一致性的理想指标。相比之下，丹皮酚受工艺干扰在软材细粉中波动剧烈且规律性较差。综上，在药典无规定检测指标成分情况下，本研究选取相对含量较高成分^[30]芍药苷作为核心分级指标，旨在通过颜色量化实现对软材细粉等级的预判与引导。

在实际工艺中，高、低含量的软材互为配比原料，而非不合格品。因此，本研究基于全体建模样本采用统计分位数法（低：中：高=2：6：2）按照芍药苷含量构建质控体系^[30]，将模型判定的低、高级软材细粉投入同一湿法制粒机混匀，将 4 机物料混合，再进行干燥总混，该方式将传统的事后检测混合均匀度前移至软材细粉中间体的质控，有效降低成品质量风险，符合中药生产全过程质量一致性要求。

2.6 传统色差模型的构建及稳健性验证

2.6.1 传统三分类基准模型的构建 本实验采用分层随机划分策略，先抽取 20% 作为外部验证集，剩余建模集按 4：1 划分为训练集与校正集，全过程保持样本类别分布均衡。为深挖色差特征对芍药苷含量分级的表征潜力，对 21 维色差特征执行了全方位的算法比选。实验设计涵盖了由 3 种预处理（标准缩放（standard scaling, SS^[31]）、幂变换（power transformation, PT^[32]）、稳健缩放（robust scaling, RS^[33]）、3 种数据均衡手段〔无处理（None）、合成少数类过采样技术（synthetic minority oversampling technique, SMOTE^[34]）、自适应合成采样算法（adaptive synthetic sampling, ADASYN^[34]〕

及 2 种特征工程策略〔全特征 Full、递归特征消除（recursive feature elimination with cross-validation, RFECV^[35]〕与 7 种主流机器学习算法（SVM、LR、KNN、NB、RF、ANN、XGB）组合而成的 126 种方案。在面对准确率相同的多组模型时，本研究优先选取对中含量样本具有更强捕获能力的模型，确保生产过程具备更高的良品通过率。结果（表 2）显示，最优模型最终锁定 PT-None-Full-LR 的组合，并将其定义为基准模型。在该组合下，模型准确率为 86.21%，其 F1 分数同步达到 0.86。

表 2 色差基准模型对比筛选结果（前 8 名）

Table 2 Comparison and selection results of color difference benchmark models (top 8)

模型组合	准确率/%	F1 分数
PT-None-Full-LR	86.21	0.86
PT-None-RFECV-LR	86.21	0.86
PT-SMOTE-Full-RF	82.76	0.83
SS-None-RFECV-KNN	82.76	0.83
SS-None-Full-SVM	82.76	0.81
SS-None-RFECV-SVM	82.76	0.81
PT-None-Full-SVM	82.76	0.81
PT-None-RFECV-SVM	82.76	0.81

深入分析最优模型的混淆矩阵（图 5），模型对中含量样本实现了 100% 的精准识别，主要的判别误差集中在中含量与高含量样本，中含量与高含量样本，未出现跨等级（如低含量误判为高含量）的严重决策失误，这为后续的工艺质控提供了良好的容错基础。

2.6.2 稳健性验证 由于色差仪仅基于 L^* 、 a^* 、 b^* 三维物理属性，在映射过程中极易受到单次随机划分的影响。为排除偶然高分，验证模型在复杂生产

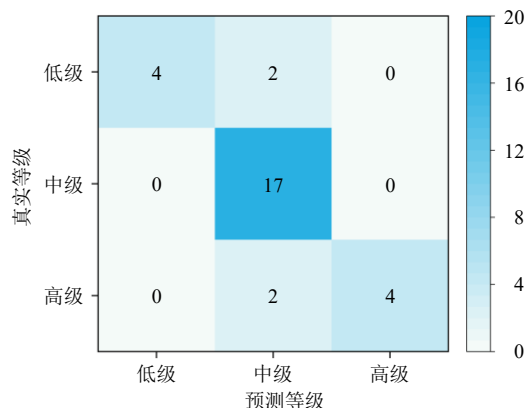


图 5 色差最优基准模型（PT-None-Full-LR）的混淆矩阵

Fig. 5 Confusion matrix of optimal color difference benchmark model (PT-None-Full-LR)

环境下的真实泛化能力,采用基于蒙特卡洛方法(Monte Carlo validation)的稳健性验证(robustness validation)^[36]。其原理是改变随机种子,产生 10 组互不相同的校正集与验证集,计算模型性能的波动情况(RSD),从而评估模型在实际工艺波动时的可靠性。结果显示,基准模型在稳健性验证中表现出特定的局限,平均准确率仅为 70.34%,其 RSD 高达 12.70%。最差情况下,准确率甚至跌至 51.72%。这一结果与单源的色差仪无法准确判别香附炮制品相互印证^[37],揭示了色差特征在三分类判别边界(中-低、中-高边界)处的严重重叠,模型极易受到重叠样本或边界样本的干扰。

2.7 级联容错质控模型的构建

2.7.1 一级预警模型的构建

基于色差基准模型稳健性不足的痛点,本研究跳出了传统的分类思路,创新性地提出了一种级联容错质控模型^[38]:将中含量样本定义为中级样本,将低、高含量样本定义为预警样本,先通过一级模型区分中级和预警样本;再将预警样本(包含真预警和误报的中含量)输入二级模型,进一步精细判别其归属低含量或高含量等级。该模型将复杂的三分类决策任务拆解为一级预警与二级判别,并融入工业容错逻辑:在制剂工艺的质控与总混环节保障下,偏向低/高含量的中含量在二级模型中被判为低/高,并不会影响最终整体质量均一,判别方向正确。依托该工业容错机制,可有效弥补单纯三分类模型受样本含量分布、分级阈值波动带来的判别局限,使模型适配实际生产场景,实现稳健的样品预警与生产质控引导。

实验保持训练集与验证集中低、中、高比例为 2:6:2,与实际生产分布一致,通过对比 3 种预处理(SS、PT、RS)、3 种数据均衡手段(None、SMOTE、ADASYN)及 2 种特征工程策略(Full、RFECV)与 7 种主流算法(SVM、LR、KNN、NB、RF、ANN、XGB)组合而成的 126 种方案。经过 GridSearch CV 全域网格搜索,在确定最优一级预警模型时(表 3),本研究不仅关注准确率与 F1 分数,更引入了级联容错逻辑下的工业评价准则:即考虑到一级模型判定的预警样本可在二级物理判别中实现良性纠偏,在性能指标相近的前提下,优先选取漏检数较低、中级数更高的组合。因此,选定 PT-SMOTE-RFECV-RF 作为级联系统的预警算法。

一级模型共筛选出 13 个特征,依次为 a^* 、 b^* 、 E^* 、 C^* 、 $L^*/(L^*+a^*+b^*)$ 、 $a^*/(L^*+a^*+b^*)$ 、 $b^*/(L^*+a^*+b^*)$ 、 $C^*/(L^*+C^*+H^*)$ 、 $L^*\times a^*$ 、 $L^*\times b^*$ 、 $a^*\times b^*$ 、 a^*/L^* 、 b^{*2} 。深入分析最优模型的混淆矩阵(图 6)可见,一级模型表现出极强的风险识别能力,在 29 个验证样本中准确预警了 12 个预警样本。虽然仍有 2 个中含量样本被误报为预警,但基于本研究提出的工业容错逻辑,这些边界样本将移交二级判别模型进行正确判别。

表 3 色差一级预警模型对比筛选结果(前 8 名)

Table 3 Comparison and screening results of color difference first-level warning model algorithms (top 8)

模型组合	准确率/%	F1 分数	漏检数	中级数
PT-SMOTE-RFECV-RF	79.31	0.77	2	13
RS-ADASYN-Full-XGB	75.86	0.74	2	12
SS-None-RFECV-RF	75.86	0.70	4	14
PT-None-RFECV-RF	75.86	0.70	4	14
RS-None-RFECV-RF	75.86	0.70	4	14
RS-ADASYN-RFECV-RF	75.86	0.70	4	14
SS-ADASYN-RFECV-RF	72.41	0.69	3	12
RS-ADASYN-RFECV-XGB	72.41	0.69	3	12

2.7.2 二级判别模型的构建

在确定最优一级预警模型后,针对被预警的 14 个预警样本(12 个预警样本与 2 个中含量样本)开展二级判别实验。重点对比了纯算法模型(方案 1:重新对低、高级样本执行特征筛选与模型比选的二分类模型判别低、高级)与物理趋势判向(方案 2:通过相关性最高的红绿色度 a^* 的训练集低、高级中位数判别低、高级)在生产决策中的表现。

在二级模型的 126 种算法比选中,多个模型(如 SVM、LR、ANN)在当前验证集上均达到了 100% 的准确率。考虑到模型架构的简洁性,最终选取 SS-None-Full-LR 组合作为方案 1 的代表。混淆矩阵显示,仅用了 1 个物理指标 a^* 的中位数的方案 2 准确率(92.86%)略低于方案 1,仅误判了 1 个样本。深入分析判别详表(表 4),对于低含量样本 z250906-33,方案 1 实现了精准预警与判别;方案 2 则由于 a^* 值高于训练集中位数,被误判为高级。

2.7.3 稳健性验证

在小样本建模中,单次划分的准确率往往具有偶然性。本研究引入了 10 次随机扰动的稳健性实验,验证模型在复杂生产环境下的可靠性。在确定级联系统参数时,决策阈值(threshold)的选择直接关系到一级预警的灵敏度与整体系统的稳健性。结果(表 5)表明,当阈值为 0.50 时,系统表现较为保守,漏检风险较高。当阈值为 0.45 时,

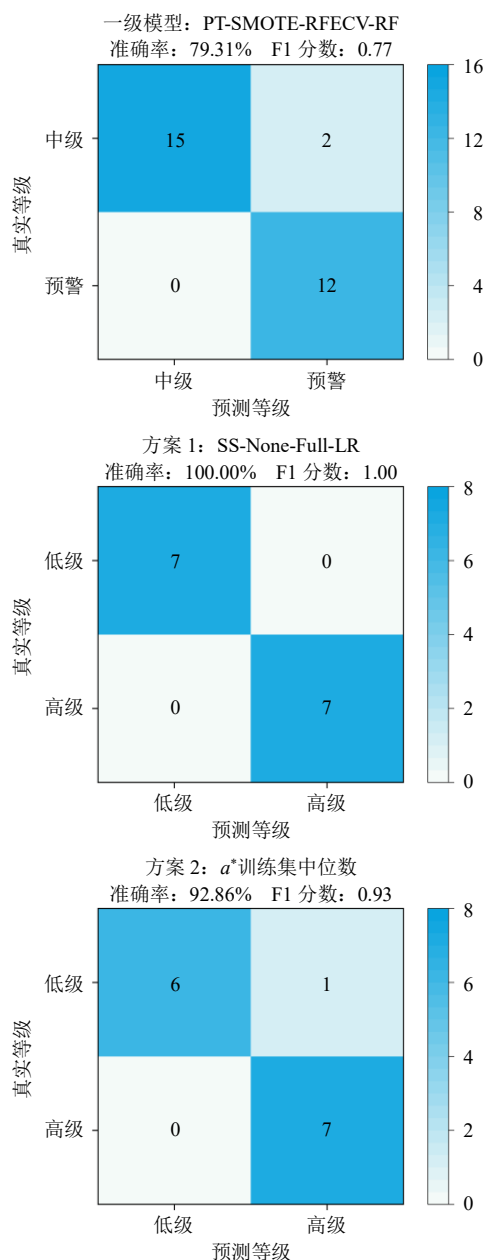


图 6 色差级联容错模型混淆矩阵

Fig. 6 Confusion matrix of color difference cascade fault-tolerant model

级联系统获得了方案 1 和方案 2 的最高平均准确率, 方案 2 的 RSD 更降低至 6.45%, 实现了精准度与稳健性的最佳平衡。

深入分析其稳健性详表(表 6), 一级模型的准确率受灵敏度原则影响降至 55.86% (RSD 为 21.34%), 将更多中级样本判为预警。但系统赋予了二级判别更广阔的纠偏空间: 方案 1 OIA 均值达到了 79.66%, RSD 仅为 6.60%。方案 2 OIA 均值达到了 78.62%, RSD 仅为 6.79%。依据奥卡姆剃刀原则 (Occam's razor), 在模型判别效果相当的情况下,

表 4 色差二级判别模型对比

Table 4 Detailed comparison table of color difference secondary discrimination models

批号	芍药苷/ (mg·g ⁻¹)	a^*	正确 等级	方案 1 等级	方案 1 预测	方案 2 等级	方案 2 预测
z250804-2	22.91	14.73	高	高	正确	高	正确
z250907-4	16.68	12.41	低	低	正确	低	正确
z250906-15	17.20	12.51	中	低	正确	低	正确
z250805-30	22.48	14.84	高	高	正确	高	正确
z251008-27	21.85	13.49	高	高	正确	高	正确
z250907-32	18.03	12.62	中	低	正确	低	正确
z250805-24	20.34	14.22	中	高	正确	高	正确
z250907-20	30.08	13.77	高	高	正确	高	正确
z250907-19	11.75	11.31	低	低	正确	低	正确
z251008-31	17.81	12.96	中	低	正确	低	正确
z250906-13	24.57	14.05	高	高	正确	高	正确
z250805-29	24.60	14.86	高	高	正确	高	正确
z250906-33	14.71	13.30	低	低	正确	高	错误
z250906-29	15.64	11.80	低	低	正确	低	正确

表 5 色差级联稳健性阈值选择

Table 5 Detailed table for selecting robustness threshold of color difference cascade

阈值	方案 1 平均 准确率/%	方案 1 RSD/%	方案 2 平均 准确率/%	方案 2 RSD/%
0.35	78.28	6.84	75.86	10.37
0.40	79.31	6.45	77.93	7.71
0.45	79.66	6.26	78.62	6.45
0.50	78.62	5.82	77.93	6.91
0.55	78.28	5.23	78.62	7.29

表 6 色差最优级联模型稳健性结果

Table 6 Robustness results of optimal cascade model for color difference

随机 种子	一级准 确率/%	一级 F1 分数	方案 1 准 确率/%	方案 1 OIA/%	方案 2 准 确率/%	方案 2 OIA/%
10	37.93	0.31	87.50	75.86	87.50	79.31
20	48.28	0.40	87.50	79.31	62.50	68.97
30	72.41	0.60	100.00	72.41	100.00	72.41
40	62.07	0.48	83.33	75.86	83.33	75.86
50	55.17	0.52	100.00	75.86	100.00	75.86
60	55.17	0.43	87.50	79.31	100.00	79.31
70	48.28	0.44	90.00	86.21	90.00	82.76
80	44.83	0.27	75.00	79.31	87.50	82.76
90	58.62	0.50	77.78	82.76	88.89	82.76
100	75.86	0.74	100.00	89.66	100.00	86.21
平均值	55.86	0.47	88.86	79.66	89.97	78.62
RSD/%	21.34	28.89	10.10	6.60	12.91	6.79

本研究最终优选方案 2 作为二级判别模型。其依据在于：方案 2 仅依赖原始单维色差特征 a^* 即可实现对低、高级的精准预判，计算逻辑简单、对算力需求极低，避免了方案 1 多维特征在小样本下可能产生的共线性干扰，具备更强的直观物理意义与工艺指导价值。

2.8 色差级联容错模型外部验证与质控实践

2.8.1 外部验证性能评估 为进一步验证色差级联系统的泛化能力及其在实际生产中的应用价值，本研究将 20% 的独立随机样本 ($N=36$ ，其规模与实际生产单批次桶数吻合) 作为外部验证集导入模型。结果显示 (图 7)，级联系统表现出良好的判别稳健性，其 OIA 达到 80.56%。

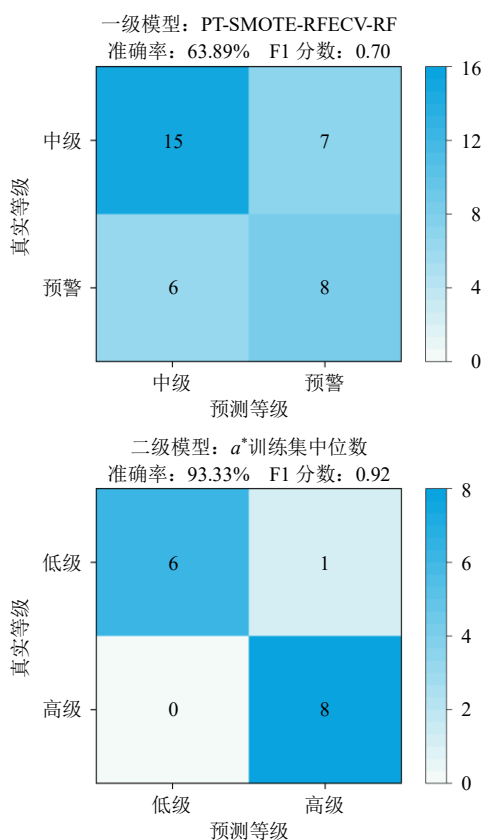


图 7 色差级联容错模型外部验证混淆矩阵

Fig. 7 Confusion matrix for external verification of color difference cascade fault-tolerant model

2.8.2 质控实践 本研究模拟实际湿法制粒生产工艺，依据级联系统的分级预警结果对外部验证的 36 个药材细粉进行针对性分配：前 3 组低、中、高级样本的个数分别为 2、5、2；第 4 组低、中、高级样本的个数分别为 1、6、2。以芍药苷含量为评价指标，将未经质控、随机混合的样本与经系统质控后的混合样本进行对比 (表 7)。结果表明，未经质

表 7 色差级联容错系统质控前后对比

Table 7 Comparison of quality control before and after in color difference cascade fault-tolerant system

批号	芍药苷/(mg·g ⁻¹)	批号	芍药苷/(mg·g ⁻¹)
质控前 1	17.67	质控后 1	19.77
质控前 2	18.30	质控后 2	19.09
质控前 3	18.83	质控后 3	18.74
质控前 4	21.04	质控后 4	18.93
平均值	18.96	平均值	19.13
RSD/%	7.74	RSD/%	2.34

控的随机混合样本含量均值为 18.96 mg/g，其 RSD 高达 7.74%，表明在无引导的生产模式下，中间体存在显著的批内质量波动风险；而经本研究构建的色差级联系统质控后，混合样本含量均值为 19.13 mg/g，其 RSD 显著降低至 2.34%，物料均匀度得到大幅提升。这一对比有力证明了基于色差级联容错模型的算法预警协同物理纠偏策略，能够以极低成本实现对药材细粉质量波动的精准干预，为中药生产全过程质控提供了可行的新策略。

为了探索药材细粉的质控上限，本研究引入了制药智能制造中主流的检测工具——NIRS 技术。以赛默飞近红外采集的光谱为对象，对比了标准正态变量变换 (standard normal variate, SNV)、多元散射校正 (multiplicative scatter correction, MSC) 及 SG 平滑一阶导数 (Savitzky-Golay first derivative, SG 1 阶导) 等预处理组合。针对 1557 维的光谱数据，利用极端随机树 (extra trees classifier, ET) 算法进行特征精炼，将维度压缩至 50~80 个核心波段，有效规避了过拟合风险。通过 GridSearch CV 对 7 种主流算法进行全域寻优，构建近红外基准模型和级联容错模型。模型标签依据芍药苷含量划分等级，其中二级判别模型基于训练集低、高级光谱特征完成建模与筛选。结果显示，近红外最优基准模型 SNV-None-RFECV-XGB 准确率为 93.10% (F1 分数 0.93) (表 8)。采用级联架构后，模型展现出显著的工业容错优势：虽然一级预警模型 SG 1st-ADASYN-Full-XGB 准确率为 89.66% (表 9)，但其产生的误报样本在二级判别中均实现了精准的方向纠偏，使整体决策效能更趋稳健 (图 8)。在多个准确率达 100.00% 的方案中，本研究并未机械沿用一级模型的集成算法 XGB，而是选择 SNV+SG 1st-None-RFECV-NB 作为二级模型。其依据在于：该组合采用标准散射校正结合一阶导数，能最大限度提

表 8 近红外基准模型对比筛选结果 (前 8 名)

Table 8 Comparison and screening results of near-infrared benchmark models (top 8)

组合	准确率/%	F1 分数
SNV-None-RFECV-XGB	93.10	0.93
SNV+SG 1st-ADASYN-Full-XGB	89.66	0.90
SNV+SG 1st-ADASYN-RFECV-XGB	89.66	0.90
SNV-None-RFECV-KNN	89.66	0.90
SNV-None-RFECV-RF	89.66	0.90
SNV-None-Full-RF	89.66	0.90
SNV-None-Full-XGB	89.66	0.89
SG 1st-SMOTE-RFECV-RF	89.66	0.88

表 9 近红外一级预警模型对比筛选结果 (前 8 名)

Table 9 Comparison and screening results of near-infrared primary early warning models (top 8)

组合	准确率/%	F1 分数	漏检数	中级数
SG 1st-ADASYN-Full-XGB	89.66	0.90	0	14
SG 1st-SMOTE-Full-XGB	86.21	0.86	1	14
SG 1st-SMOTE-RFECV-XGB	86.21	0.86	1	14
SNV+SG 1st-None-Full-RF	86.21	0.86	1	14
SG 1st-ADASYN-RFECV-RF	86.21	0.86	0	13
SNV+SG 1st-SMOTE-RFECV-RF	82.76	0.83	1	13
SG 1st-ADASYN-RFECV-XGB	82.76	0.83	0	12
Raw-None-Full-RF	79.31	0.79	2	13

取化学指纹信息；且朴素贝叶斯算法在处理二级任务这种小样本、极化分布的特征时，较复杂模型具备更高的统计稳健性与更低的过拟合风险。

2.8.3 稳健性验证 稳健性验证结果显示，近红外基准模型的平均准确率达到 78.28%，而采用级联模型后的 OIA 均值高达 90.69%，展现了极强的抗干扰能力。由此得出结论：NIRS 凭借分子层面的化学指纹信息，在软材细粉质控中具有突出的精度优势；而色差级联模型凭借 78.62% 的 OIA 及 6.79% 的稳健波动，在表观性状评价中表现出高性价比。

3 讨论

本研究针对 GFC 软材细粉批内颜色波动的痛点，构建了“级联容错质控模型”。将复杂的低-中-高 3 分类决策拆解为一级风险预警与二级物理判向，探索了一种高效、低成本的中间体质控新方案。一级模型基于多维色差特征完成风险样本初步预警；二级模型则利用红绿色度 a^* 与芍药苷的正相关性实现精准纠偏。这种算法预警协同物理纠偏的机制，规避了单一黑箱算法处理临界样本的决策风险，提升复杂工况下的判别稳定性，推动中药辨色

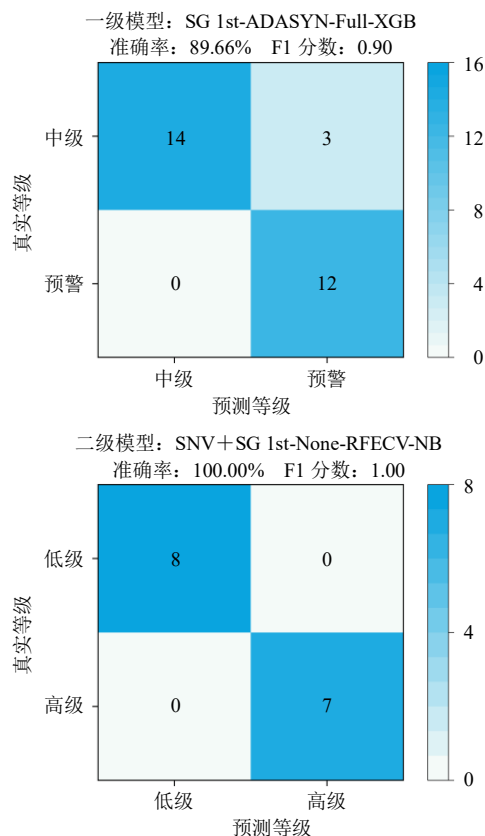


图 8 近红外级联容错模型混淆矩阵

Fig. 8 Confusion matrix of near-infrared cascade fault-tolerant model

论质由传统感官经验转向数字化管控。

研究选用红绿色度 a^* 作为二级判别核心指标。究其深层机制：芍药苷对照品呈白色，自身不会显色，但在提取浓缩与制粒工序中，其含量变化与多酚、鞣质等内源性有色物质的增减趋势高度同步。同时芍药苷热稳定性较差，含量水平可直观反映物料受热状态与有效成分富集程度，因此， a^* 值能够有效表征样品化学组分差异。同时，结合工业生产质控特点，本实验引入 OIA 作为评价指标，该指标区别于传统准确率、F1 分数，可区分漏判、误报的风险权重，有效平衡质控安全与漏报风险。测试结果显示，级联容错模型可将基准模型的 RSD 由 12.70% 降至 6.79%，判别稳定性显著提升。质控实践中，该模型可指导物料定向均一，使混合后芍药苷含量 RSD 从 7.74% 降至 2.34%，在物料均匀度显著改善的同时，降低总混工序载荷与工时，降低批次返工成本，展现了明确的增效潜力。横向与 NIRS 技术对比，后者凭借高维光谱信息，工业准确率可达 90.68%，适用于高精度检测；色差模型准确率为 78.62%，凭借低成本、检测快速的优势，更契合生

产线大批量初筛场景。

本研究虽在打乱样本模式下验证了模型的稳健性,但软材细粉受水分、粒径、辅料分布等因素干扰,色差指标存在固有上限,且模型应对跨年度、大范围工艺漂移的性能仍有待验证,后续可引入批次校正方法进一步优化。综上所述,色差仪作为一种高性价比、无损的数字化工具,在中药中间体质控方面具有广阔应用前景。本研究提出的算法标定协同物理纠偏策略,将中药批间均化研究延伸至批内质控,这不仅是过程分析技术(process analysis technology, PAT)在中药智能制造中落地应用的有益探索,也是对《中药均一化研究技术指导原则(试行)》^[39]现有批间均化政策的技术补充,为未来政策向提取液、浸膏及其他中间体均一化处理的延伸提供了先行先试的实验依据。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 国家药品监督管理局. 国家药监局关于发布中药生产监督管理专门规定的公告(2025年第79号)[EB/OL]. (2025-09-08) [2026-09-10]. <https://www.nmpa.gov.cn/xxgk/ggtg/ypggtg/ypqtggtg/20250908094819150.html>.
- [2] 孙昱,孙国祥,李焕德. 提高中药质量一致性的均化投料技术操作与控制策略[J]. 中南药学, 2020, 18(11): 1781-1785.
- [3] 邓海燕,陈文兰. 桂枝茯苓胶囊联合炔雌醇环丙孕酮片治疗多囊卵巢综合征对性激素水平的影响[J]. 临床合理用药, 2025, 18(29): 123-125.
- [4] 乔金凤,赵倩倩,尹晓梅,等. 大黄廑虫丸联合桂枝茯苓胶囊治疗子宫肌瘤的临床研究[J]. 现代药物与临床, 2026, 41(7): 2075-2080.
- [5] 周跃华,冯怡. 关于中药复方制剂均一化研究及制剂中药用物质的探讨[J]. 中草药, 2023, 54(8): 2357-2364.
- [6] 荆文光,程显隆,刘安,等. 基于“辨状论质”综合评价指数的厚朴饮片等级划分和优质优效研究[J]. 中草药, 2021, 52(8): 2285-2293.
- [7] 金月,谢静,李东辉,等. 基于色差原理的大黄化学成分含量与色度相关性分析[J]. 时珍国医国药, 2026, 37(3): 494-499.
- [8] 侯富国,桂新景,王艳丽,等. 基于智能视觉技术的白及饮片真伪快速辨识方法研究[J]. 中草药, 2023, 54(2): 509-519.
- [9] 傅饶. 丹参饮片质量识别技术与评价标准研究[D]. 南京: 南京中医药大学, 2025.
- [10] 张一凡. 基于数字智能化快速检测技术的姜炭不同炮制程度的质量分析研究[D]. 广州: 广东药科大学, 2022.
- [11] 徐红霞,吴沂芸,裴瑾,等. 红花黄酮类成分与其色度值相关性研究[J]. 中药材, 2018, 41(1): 49-54.
- [12] 李航,黎盛强,周恩丽,等. 基于神经网络优化模型的中药复方安慰剂配色模拟研究[J]. 南京中医药大学学报, 2024, 40(1): 18-25.
- [13] 缪培琪,王璧璇,司梦兰,等. 机器视觉检测技术及其在中药智能制造中的应用展望[J]. 天津中医药大学学报, 2020, 39(4): 365-373.
- [14] 雷林. 基于颜色量化与近红外光谱技术对谷芽、苍术质量快速评价的方法学研究[D]. 武汉: 湖北中医药大学, 2021.
- [15] 石晓燕. 应用数字图像技术对冬小麦叶片部分生理指标的估测[D]. 保定: 河北农业大学, 2005.
- [16] 章毅颖,张余,褚云霞,等. 利用色差仪参数分析青花菜和花椰菜叶片颜色特征[J]. 园艺学报, 2025, 52(4): 872-882.
- [17] 曹连平,王力民,李锡军,等. 色差仪的应用实践[J]. 印染, 2004, 30(24): 33-38.
- [18] 李家春,梁娜,肖伟,等. 定量指纹图谱技术在桂枝茯苓胶囊过程质量控制中的应用[J]. 中国中药杂志, 2010, 35(24): 3270-3273.
- [19] Jaganathan K, Tayara H, Chong K T. Prediction of drug-induced liver toxicity using SVM and optimal descriptor sets[J]. *Int J Mol Sci*, 2021, 22(15): 8073.
- [20] 赵文悦. 基于机器学习和网络药理学探索扛板归对糖尿病心脏病的作用及其机制[D]. 柳州: 广西科技大学, 2025.
- [21] 毋雪雁,王水花,张煜东. K最近邻算法理论与应用综述[J]. 计算机工程与应用, 2017, 53(21): 1-7.
- [22] 钱丹丹,周金海. 基于计算机视觉的中药饮片检测与分级研究[J]. 时珍国医国药, 2019, 30(1): 203-205.
- [23] 董红瑶,王弈丹,李丽红. 随机森林优化算法综述[J]. 信息与电脑, 2021, 33(17): 34-37.
- [24] 张文青,赵珊,钱文秀,等. 基于人工神经网络和支持向量机算法的药源性急性间质性肾炎计算模型的建立及在中药领域的应用[J]. 中草药, 2023, 54(2): 416-424.
- [25] 汤怡婷,陈玉鹏,孔祥英,等. 人工神经网络模型在中药复方研究中的应用进展[J]. 中华中医药杂志, 2021, 36(1): 351-354.
- [26] 何帅,周杰,张福林,等. 近红外光谱和XGBoost的流化床干燥过程水分含量在线检测[J]. 光谱学与光谱分析, 2024, 44(12): 3347-3352.
- [27] 张露文,刘雨,谢佳琦,等. 近红外光谱结合机器学习对当归及其相似品的准确鉴别[J]. 分析测试学报, 2026, 45(4): 673-679.
- [28] 刘雨. 基于近红外光谱和化学计量学的当归鉴别与定

- 量研究 [D]. 天津: 天津工业大学, 2025.
- [29] 沈晔, 王继森, 向晴, 等. 基于机器学习的中药产地溯源研究 [J]. 中草药, 2026, 57(5): 1798-1813.
- [30] 刘佳萌, 刘岩, 彭佳慧, 等. 中成药均一化投料研究策略 [J]. 中国中药杂志, 2024, 49(16): 4537-4544.
- [31] Bhattacharjee B, Madhu U, Guha S K, *et al.* Neural network approach enhancing churn prediction with categorical encoding and standard scaling [J]. *Sci Rep*, 2026, 16: 6274.
- [32] 秦艳玮, 邓小娟, 郭文雯, 等. 低剂量 CT 投影数据噪声模型研究 [C] // 第十六届中国体视学与图像分析学术会议论文集——交叉、融合、创新. 海口: 中国体视学学会, 2019: 129-130.
- [33] Oh J W, Min D J. Robust scaling strategy for frequency-domain acoustic full waveform inversion [J]. *ASEG Ext Abstr*, 2013, 2013(1): 1-4.
- [34] 孟东霞, 柳凌燕, 魏晓光. 利用二阶 k 近邻构造微簇的过采样方法 [J]. 统计与决策, 2026, 42(3): 46-51.
- [35] 姚磊, 黎振, 武川, 等. 基于互信息特征选择和递归特征消除的钢材疲劳强度预测 [J]. 机械强度, 2026, 48(1): 72-78.
- [36] 张喜科, 赵文华, 付先军, 等. 基于 XGBoost 和 SHAP 的中药寒热药性识别及寒热特征标记可视化研究 [J]. 世界科学技术—中医药现代化, 2022, 24(12): 4984-4993.
- [37] 吴鑫雨, 邱丽媛, 王又迪, 等. 基于智能感官与多源信息融合技术的香附炮制程度快速辨识方法研究 [J]. 中草药, 2023, 54(21): 7007-7016.
- [38] 王一涵, 王宗仁, 李友丽, 等. 基于 LSTM-XGBoost 的日光温室环境传感数据质量控制方法 [J]. 农业工程学报, 2026, 42(6): 254-262.
- [39] 国家药品监督管理局药品审评中心. 国家药监局药审中心关于发布《中药均一化研究技术指导原则(试行)》的通告(2020年第38号)[EB/OL]. (2020-11-04)[2026-09-10]. <https://www.cde.org.cn/main/news/viewInfoCommon/8a042d3cea539a543a2ea68b58927662>.

[责任编辑 郑礼胜]