

doi: 10.12452/j.fxcxb.25110701

血府逐瘀浸膏质量智能评价体系构建:基于物理指纹图谱、机器视觉与大语言模型的联合应用

沙鑫^{1,2}, 常昊^{1,2}, 宋纹³, 余河水^{1,2,4}, 李正^{1,2,4}, 李文龙^{1,2,4*}

(1. 天津中医药大学 中药制药工程学院, 天津 301617; 2. 现代中医药海河实验室, 天津 301617; 3. 天津宏仁堂药业有限公司, 天津 300380; 4. 天津市中药智能制药与绿色制药重点实验室, 天津 301617)

摘要: 为解决中药浸膏生产中依赖单一指标或经验判断浓缩终点的局限, 该研究以血府逐瘀浸膏为对象, 创新性提出“物理指纹图谱+机器视觉+大语言模型”的多模态融合策略智能质控体系。通过筛选密度、pH值等7项核心物理指标, 并进行标准化处理构建物理指纹图谱; 然后结合马氏距离剔除异常样本, 使用“CRITIC+熵权+TOPSIS法”计算综合得分, 并以0.6为合格阈值, 将100批样本分为62批合格与38批不合格样本; 同时采集并提取浸膏图像特征构建了6种机器学习模型, 对比结果显示, XGBoost模型的性能最优, 准确率、精确率、召回率、F1分数、AUC分别达0.9333、1.0000、0.8333、0.9091、0.9630。在此基础上进一步开发集成了大语言模型的智能评估平台, 可在10s左右完成分析并生成可操作的工艺建议。该体系为血府逐瘀浸膏提供了“客观分级+快速预测+智能建议”一体化质控方案, 也为中药浸膏类中间体的多维度质量评价提供了可复用的技术范式, 将有力推动中药生产向“数据驱动”转型。

关键词: 血府逐瘀浸膏; 物理指纹图谱; 机器视觉; 大语言模型; 质量控制

中图分类号: O657; R284 **文献标识码:** A **文章编号:** 1004-4957(2026)06-0001-09

Construction of Intelligent Quality Evaluation System for Xuefu Zhuyu Extract: Combined Application of Physical Fingerprint, Machine Vision and Large Language Model

SHA Xin^{1,2}, CHANG Hao^{1,2}, SONG Wen³, YU He-shui^{1,2,4}, LI Zheng^{1,2,4}, LI Wen-long^{1,2,4*}

(1. College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China; 2. Haihe Laboratory of Modern Chinese Medicine, Tianjin 301617, China; 3. Tianjin HongRenTang Pharmaceutical Co., Ltd., Tianjin 300380, China; 4. Tianjin Key Laboratory of Intelligent and Green Pharmaceuticals for Traditional Chinese Medicine, Tianjin 301617, China)

Abstract: In order to solve the limitation of relying on a single index or experience to judge the concentration end point in the production of traditional Chinese medicine extract, this study took Xuefu Zhuyu extract as the object, and innovatively proposes an intelligent quality control system of “physical fingerprint + machine vision + large language model” multi-modal fusion strategy. Seven core physical indexes such as density and pH value were screened and standardized to construct physical fingerprint. Then, the abnormal samples were eliminated by Mahalanobis distance, and the comprehensive score was calculated by “CRITIC+entropy weight+TOPSIS” method. With 0.6 as the qualified threshold, 100 batches of samples were divided into 62 batches of qualified samples and 38 batches of unqualified samples. At the same time, the extract image features were collected and extracted, and six machine learning models were constructed. Among them, the XGBoost model had the best performance, with accuracy, precision, recall, F1 score, and AUC reaching 0.9333, 1.0000, 0.8333, 0.9091, and 0.9630, respectively. On this basis, an intelligent evaluation platform integrated with large language model was further developed, which could complete the analysis and generate operable process suggestions in 5–6 minutes. This system provides an integrated quality control scheme of “objective grading+rapid prediction+intelligent” suggestion for Xuefu

收稿日期: 2025-11-07; 修回日期: 2025-11-26

基金项目: 国家重点研发计划(2023YFC3504502); 现代中医药海河实验室-天津宏仁堂药业有限公司合作项目(XMH2022004)

* 通讯作者: 李文龙, 博士, 研究员, 研究方向: 中药质量控制技术研究, E-mail: wshlwl@tjutcm.edu.cn

Zhuyu extract, and also provides a reusable technical paradigm for multi-dimensional quality evaluation of Chinese medicine extract intermediates, which effectively promotes the transformation of Chinese medicine production to “data-driven”.

Key words: Xuefu Zhuyu extract; physical fingerprint; machine vision; large language model; quality control

中药浸膏作为中药制剂生产的核心中间体,其质量一致性直接决定了最终产品的药效稳定性与临床用药安全性,是中药现代化生产中质量控制的关键环节^[1]。血府逐瘀制剂作为经典的活血化瘀类中成药,由桔梗、赤芍、甘草、桃仁等11味中药材组成,具有活血祛瘀、行气止痛等功效,在心脑血管疾病治疗中应用广泛^[2]。该制剂的核心中间体——血府逐瘀浸膏的物理化学属性(如动力黏度、粒径分布等)若发生波动,将直接威胁下游生产工艺的稳定性(例如:黏度过高易导致喷雾干燥喷枪堵塞,粒径不均则影响流化床制粒的均一性)及终产品质量(如关键有效成分溶出度不一致),最终可能导致临床疗效的不可预测性^[3-4]。然而,当前企业在实际生产过程中,常依赖于单一指标(如密度)或传统经验方法(如“挂旗”、“滴水成珠”)来判断浓缩终点,这些方法主观性强、误差大,难以全面、客观地反映浸膏的复杂质量属性^[5]。因此,构建一套科学、全面、基于客观指标的血府逐瘀浸膏质量评价体系,对于精准控制其生产过程、保障制剂质量均一性与临床疗效的可靠性,具有至关重要的意义。

当前,中药浸膏的质量控制主要依赖两类方法,但各自存在明显局限。一方面,部分企业或研究仍沿用单一物理指标(如密度或固含量)进行浓缩终点判定,如黄玉普等^[6]在进行浸膏粉干燥工艺优选时,以相对密度为1.05和1.3作为清膏和稠膏浓缩终点的判断,然而单一指标难以全面反映浸膏整体物理特性对其后续制剂工艺(如干燥、制粒、压片)适应性的影响;另一方面,化学指纹图谱,如高效液相色谱法(HPLC),虽被广泛用于表征浸膏中多组分的化学组成,如李柯等^[7]利用HPLC指纹图谱对复方感冒灵流浸膏进行质量评价与药效物质研究,但该方法通常存在检测周期长、前处理复杂、仪器成本高、专业依赖性强等问题,且所获化学信息难以直接关联浸膏的平均粒径、黏附性等关键下游加工性能。随着科学技术的发展,物理指纹图谱因能快速表征物质多维度物理属性、反映批次间整体一致性,在食品、化工等领域的质量评价中已展现优势^[8],例如杨婉等^[9]利用物理指纹图谱对升陷汤标准煎液进行质量一致性评价,对比发现该方法可以表征煎液内在质量变化;而机器视觉技术凭借非接触、高通量的特点,可通过提取样品外观特征(颜色、纹理、形态)间接关联内在质量,实现快速分类和检测。例如You等^[10]利用机器视觉结合深度学习对碾茶干燥过程中的含水量进行了预测;Pérez-Calabuig等^[11]利用机器视觉技术检测无麸质啤酒中的麸质,实现了快速预筛选;同时随着学科的交叉应用,人工智能驱动的中药生产过程智能化在中药工业中具有高效率、低成本和科学化的特点^[12]。因此“物理指纹图谱+机器视觉+机器学习与大语言模型等AI技术”的融合,将会为浸膏质量的快速判别提供新路径。

基于此,本研究针对性弥补现有方法局限,提出“物理指纹图谱+机器视觉+大语言模型”的多模态融合智能质控策略:首先筛选密度、pH值等7项核心物理指标,经标准化处理构建物理指纹图谱并结合主成分分析,通过多维度物理属性的系统整合,弥补单一物理指标难以全面表征浸膏整体特性的缺陷,实现批次一致性精准评估;同时采用马氏距离剔除异常样本,借助“CRITIC+熵权+TOPSIS”组合方法完成浸膏质量分级,确保评价的客观性与科学性;其次,借助机器视觉技术提取浸膏颜色、纹理、形态等25项图像特征,构建随机森林、K近邻算法(KNN)、XGBoost、支持向量机(SVM)、人工神经网络及LightGBM六类代表性机器学习模型,凭借非接触、高通量优势快速实现质量分类预测,弥补化学指纹图谱检测周期长、难以关联下游加工性能的不足;尤为关键的是,本研究进一步融合基于大规模稀疏MoE架构和Transformer架构的大语言模型(Doubao,字节跳动公司),构建一体化智能评价体系,将AI判别结果与多维物理特征深度融合,自动生成面向生产一线人员的可操作自然语言建议。本文旨在通过上述设计,克服现有方法主观性强、维度单一、解释性差的局限,实现浸膏质量的多维度、快速、客观评价,弥合“黑箱模型”与“工艺认知”之间的鸿沟,使质量判定不仅“判得准”,更能“说得清、用得上”。同时为中药浸膏类中间体的多维度质量评价体系构建提供可复用的技术参考,推动中药生产从“经验控制”向“数据驱动”转型。

1 实验部分

1.1 材料与仪器

血府逐瘀浸膏来自天津宏仁堂药业有限公司。Bettersize2600型激光粒度分析仪(丹东百特仪器有限公司)、SevenDirect SD30型电导率仪(瑞士METTLER TOLEDO公司)、SNB-2型数字式粘度计(上海精密科学仪器有限公司)、PZ-100型接触角测量仪(北京品智创思精密仪器有限公司)、101-2EBS型电热鼓风干燥箱(北京市永光明医疗仪器有限公司)、PHS-25型精密酸度计(上海雷磁科技有限公司)。

1.2 实验方法

1.2.1 物理指纹图谱的构建 浸膏常作为制剂中间体，其下游加工过程主要依赖喷雾干燥和流化床干燥工艺。结合前期调研结果以及对干燥过程的深入分析，本研究选取密度(ρ)、pH值、固含量(SC)、表面张力(γ)、平均粒径(d_{50})、电导率(σ)、动力黏度(μ)7项物理指标作为血府逐瘀浸膏的关键属性，并参照(2020版)《中国药典》以及相关文献资料对其进行测量^[13-14]。筛选依据和检测方法如表1所示。

表1 血府逐瘀浸膏关键物理指标筛选依据及检测方法

Table 1 Screening reasons and detection methods of key physical indicators for Xuefu Zhuyu extracts

Physical indicator	Screening reason	Detection method
表面张力	影响喷雾液滴尺寸，决定液-固润湿行为及成核机制，进而控制颗粒形态与大小	接触角测量仪测定
黏度	影响喷雾液滴尺寸与雾化均匀性，进而影响颗粒形成效率与最终粒度分布	数字黏度计测定
平均粒径	反映浸膏中不溶性颗粒大小，过大颗粒易堵塞喷雾枪雾化器，导致制粒中断或失败	激光粒度分析仪测定
固含量	决定浸膏中固体物质总量，是计算制粒所需辅料比例的关键依据	烘干法测定
密度	物质的固有属性，影响单位体积物料的质量，是工艺参数(如进料速率)设定的参考	质量体积法测定
电导率	反映电解质离子浓度，影响液滴在流化床静电场中的荷电与结合行为	电导率仪测定
pH值	影响物料理化性质(如溶解度、稳定性)及颗粒形态，需保持批次间稳定以确保工艺重现性	pH计测定

为验证血府逐瘀浸膏批次间的质量一致性，本研究通过构建物理指纹图谱开展批次一致性分析。由于7项物理指标的量纲与数值范围差异较大(如pH值范围1.0~14.0，固含量范围0~100%)，为消除量纲对后续分析的影响，根据《中国药典》(2025版)和文献调研确定了浸膏各项指标的范围，采用线性转化的方式将各参数转化为0~10区间的标准化值^[15]。7项关键物理指标的数值范围和转换公式如表2所示。基于标准化值，分别采用Origin2022软件和SIMCA 14.1软件绘制物理指纹图谱、PCA得分图和OPLS-DA得分图。

1.2.2 浸膏质量分级 在浓缩过程中，血府逐瘀浸膏其各项物理指标的变化并非均与浓缩时间呈正相关，例如pH值会随着浓缩时间的增加而减小。不仅如此，浸膏的物理属性并非越大越好，如浸膏黏度过大时呈稠浸膏状态，会直接影响后续的喷雾干燥、流化床制粒等下游加工过程的顺利开展。因此，浸膏的各项物理指标应是越接近理想值越好。然而，目前尚无资料明确界定血府逐瘀浸膏各物理参数的理想值，需通过科学方法确定。

考虑到样本参数测量时可能会存在随机误差，易导致出现异常批次。因此首先对100批浸膏的物理属性标准化值进行预处理，并采用方差分析判断样本的离散程度。由于浸膏浓缩过程中的7项关键物理指标联系紧密，传统基于欧氏距离的检测方法难以准确识别异常样本。为此，本研究采用基于马氏距离的多元统计方法，其优势在于能够有效度量样本点在多元相关空间中的统计距离，反映数据协方差结构的优点，从而更可靠地识别并剔除潜在的离群点，确保后续分析基于高质量、一致性的数据集^[16]。随后结合四分位距和中位数确定各参数的理想值。计算各参数指标与理想值差的绝对值作为逆向指标(差值绝对值越小，指标越优)。采用CRITIC权重法和熵权法确定各指标权重，再通过组合赋权法得到最终的组合权重，最后结合TOPSIS法计算浸膏样品的综合得分，并进行合格与不合格样品的划分。

1.2.3 质量判别模型构建 通过已有的理论知识和前期预实验，本研究总结了血府逐瘀浸膏图像特征(颜色特征、纹理特征、面积、圆整度)与关键物理属性(固含量、动力黏度等)的联系，如图1所示，

表2 血府逐瘀浸膏关键物理指标的标准化换算方法
Table 2 Standardized conversion method of key physical indicators of Xuefu Zhuyu extracts

Physical indicator	Scope	Formula
ρ	0~2	5X
pH	1~6	2X-2
σ	3 000~10 000	X/1 000
μ	0~500	X/50
d_{50}	0~10	X
γ	0~50	X/5
SC	40~60	X/6

它们之间呈正相关或负相关。

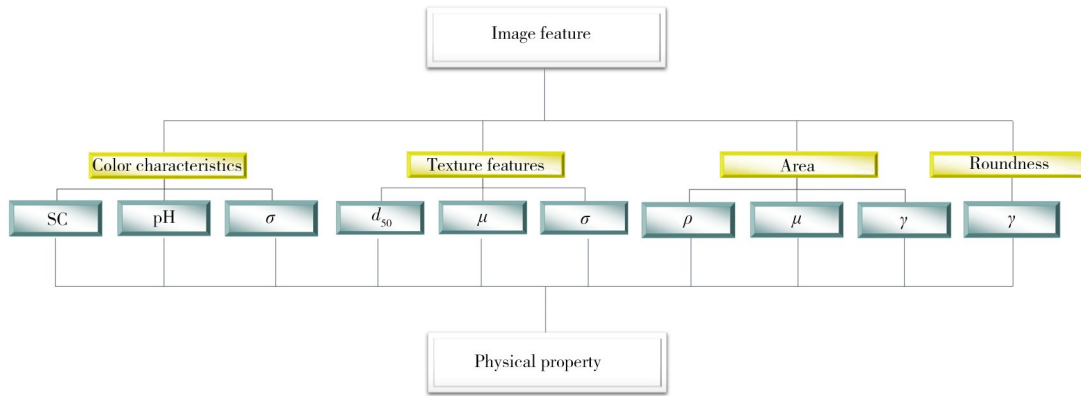


图1 血府逐瘀浸膏关键物理指标和图像特征的联系

Fig. 1 The correlation between key physical indicators and image characteristics of Xuefu Zhuyu extracts

本研究的图像数据采集采用以下方式：使用智能手机拍摄 1 g 浸膏样本的静态图像，对获取的原始图像进行预处理(具体流程详见图 2)。首先对原始图像进行中值滤波去噪，以消除图像采集过程中引入的噪声干扰，基于预处理后的图像，利用颜色矩的原理提取包含一阶矩(均值)、二阶矩(方差)和三阶矩(偏度)在内的 9 项颜色特征，以量化表征浸膏的整体颜色分布特性；在去噪图像的基础上分别执行边缘锐化和拉普拉斯算子平滑两种预处理步骤，最后根据灰度共生矩阵在边缘锐化后的图像上提取 14 项纹理特征(如能量、熵、对比度、同质性等)，用于描述浸膏表面的粗糙度、均匀性等纹理属性；结合形态学方法(阈值分割)提取浸膏区域的几何属性，包括面积以及圆整度特征。对从 3 个维度提取到的 25 项图像特征数据进行归一化处理，消除不同特征量纲和数值范围差异的影响，确保尺度一致性，为后续质量判别模型的建立提供标准化数据。

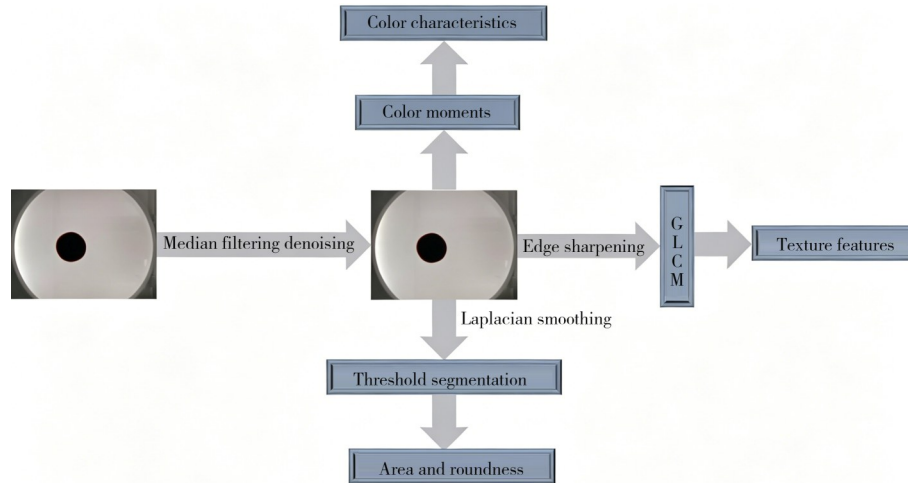


图2 图像特征提取步骤

Fig. 2 Image feature extraction steps

将 100 批样本按照 7 : 1.5 : 1.5 的比例采用随机分层抽样方法划分为训练集(70 批)、验证集(15 批)和测试集(15 批)。其中，训练集和验证集用于模型的训练和参数校准，测试集用于模型性能验证。采用 Python3.9 软件分别构建随机森林、KNN、XGBoost、SVM、人工神经网络、LightGBM 6 种机器学习模型。选取准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 分数、AUC 以及混淆矩阵作为模型性能的评价指标。

1.2.4 构建“AI 视觉预测 + 豆包大模型建议”智能评估平台 在完成 XGBoost 模型对浸膏浓缩终点高精度预测的基础上，本研究进一步开发了一套面向生产现场的浸膏质量智能评估 Web 系统。系统采用前后端分离架构，前端构建了用户友好的交互界面，支持图像上传与结果可视化；后端以 Flask 框架为核心，集成前期完整的图像分析过程(包括去噪、锐化、颜色矩提取、GLCM 纹理分析、轮廓形态学

分割等), 调用已训练的归一化器和XGBoost模型进行“合格/不合格”二分类判定。创新性地引入基于大规模稀疏MoE架构和Transformer架构的大语言模型(Doubao, 字节跳动公司)的API, 选择该模型核心基于3点考量: 其一, 其融合双架构的设计兼具千亿级参数规模与高效推理能力, MoE的稀疏激活机制可在保证专业度的同时降低计算延迟, Transformer架构则能精准捕捉物理特征与工艺建议的语义关联, 适配工业现场快速响应需求; 其二, 该模型在中文专业文本生成与行业技术方案输出任务中表现优异, 对中药浸膏生产的专业术语把控精准, 能将高维物理特征(如RGB均值、轮廓面积等)转化为一线工艺人员易懂的操作指引; 其三, 支持灵活的API调用方式, 可与后端Flask框架无缝对接, 快速整合模型预测结果与多维特征数据。基于上述优势, 系统根据模型预测结果与提取的多维物理特征(如RGB均值、轮廓面积、圆整度、纹理对比度等), 动态生成专业建议: 若判定为“不合格”, 系统提示浓缩工艺调整方向(如调节温度、真空度或时间); 若为“合格”, 则推荐下游加工路径(如制丸、胶囊填充或冻干)。该系统可实现从“图像输入→特征提取→智能判定→专家级建议”的端到端闭环, 推动数据驱动模型与领域知识深度融合。

2 结果与讨论

2.1 物理指纹图谱的构建

根据“1.2.1”所述方法, 对100批血府逐瘀浸膏的7项物理指标进行测定和标准化, 得到各项指标的实测值与标准化值。结果显示, 各参数实测值之间差异较大, 经标准化处理后, 所有参数均转化为0~10区间内的值, 成功消除了量纲, 为后续的质量一致性评价奠定了基础。

100批样品标准化参数的相关性热图如3所示, 该热图聚焦于批次间相关性分析。在热图中, 红色表示样品间正相关, 蓝色表示负相关, 颜色深浅代表相关性强度。由图可知, 红色区域分布广泛, 占比超过80%, 表明大多数样品之间存在正相关关系, 反映出不同批次血府逐瘀浸膏在物理属性上具有较高的一致性, 这可能与企业生产过程中采用相对统一的浓缩工艺有关。值得注意的是, S19、S20、S33、S34等批次与其他批次间的相关性差, 甚至出现负相关, 推测原因可能是这些批次浓缩未达到规定终点或在数据采集过程中存在误差。

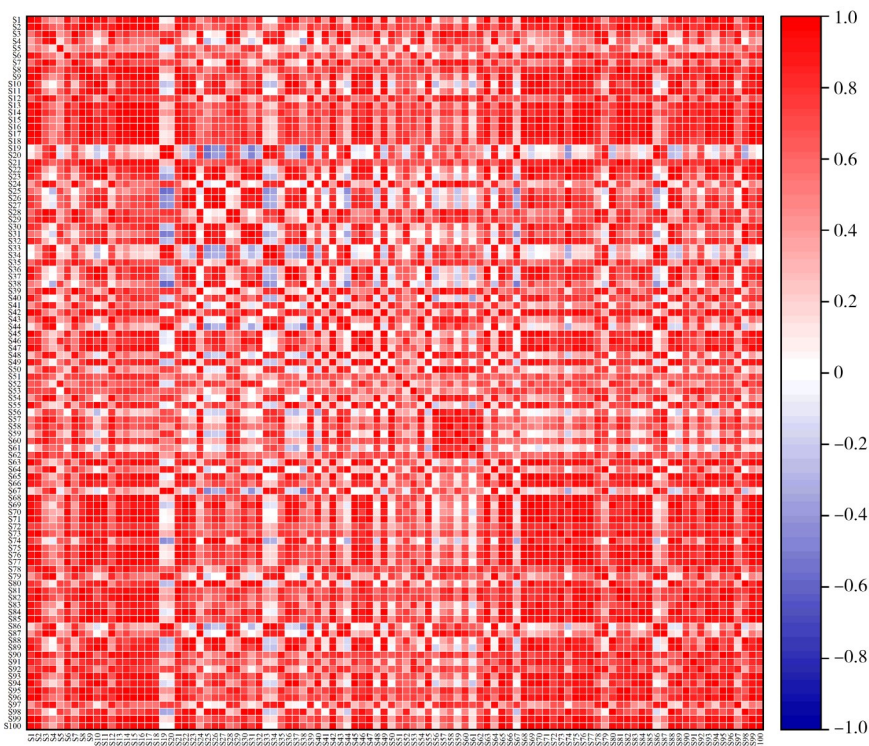


图3 血府逐瘀浸膏的相关性热图

Fig. 3 Correlation heatmap of Xuefu Zhuyu extracts

血府逐瘀浸膏的主成分(PCA)得分图如图4所示。PCA分析结果显示PC1解释了47.5%的方差,PC2解释了20.0%的方差,两者累计贡献率达67.5%,可基本反映样本的主要物理特征差异。从图4可以看出,多数样本点都集中在PC1(-2~2)、PC2(-2~2)范围内,说明这些样本在PC1和PC2对应的物理参数上具有相似的特征,整体聚类效果较好。但在95%置信区间内,样本点分布稀疏,甚至存在偏离中心区域的情况,如S40、S53、S61等批次,这可能是由于样本在主成分所对应的原始物理变量上存在与其他样本不同的特性,需进一步关注。

为了验证PCA的准确性,采用SIMCA 14.1软件建立了OPLS-DA得分图,如图5所示。OPLS-DA分析结果显示 $t[1]+t[2]$ 轴解释了约79.4%的变量方差($R^2X[1]=0.567$ 、 $R^2X[2]=0.227$),说明OPLS-DA对变量的覆盖度优于PCA;但从样本分布看,OPLS-DA中不同样本未呈现明显的组间分离趋势,结合PCA的降维结果可知,无论采用无监督(PCA)还是有监督(OPLS-DA)的降维方法,样本的组间差异均不显著;同时OPLS-DA的Hotelling T^2 椭圆(95%置信区间)显示大部分样本均处于正常分布范围内,个别样本如S61、S53、S56等在置信区间外或边缘,与PCA中样本的整体分散特征一致。

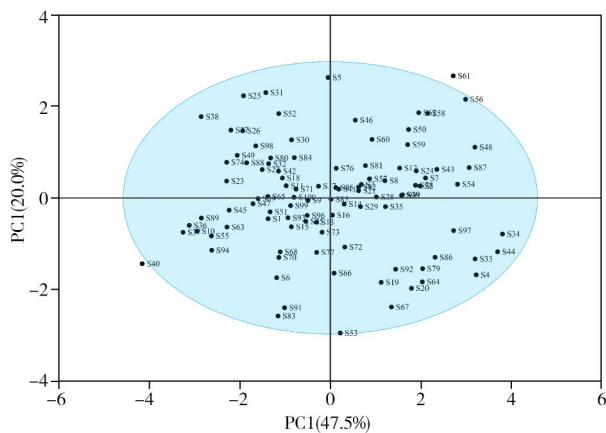


图4 血府逐瘀浸膏的PCA得分图

Fig. 4 PCA score plot of Xuefu Zhuyu extracts

血府逐瘀浸膏的物理指纹图谱如图6所示,该图可直观呈现出样品在7项物理指标维度上的差异与分布情况。结果显示,100批浸膏的密度指标高度集中,表面张力和固含量差异较小,而黏度、电导率、pH值、平均粒径差异较大。推测原因是企业在判断浸膏浓缩终点时,常以密度作为单一指标(控制密度在 $1.0\sim 1.2\text{ g/cm}^3$),导致其他物理属性缺乏有效控制,造成下游加工条件不统一(如黏度过大导致喷雾干燥时喷嘴堵塞,平均粒径过大导致制粒不均匀)。

2.2 浸膏质量分级

基于“1.2.2”方法,首先计算了100批样本7项标准化参数的方差,结果如表3所示。其中 σ 、 μ 、 d_{50} 、pH值4项指标的方差均大于0.5,整体波动范围较大,因此需先剔除异常样本。采用马氏距离法对100批样品进行异常值检验。结果显示S5、S19、S25、S26、S27、S38、S41、S46、S47、S49、S51、S61共12个批次被判定为异常样本,将其剔除后保留88批有效样本用于后续理想值计算。

计算88批有效样本的标准化参数四分位距,确定各参数科学合理的理想阈值范围,最终以各参数的中位数为理想值,结果如表3所示。随后,以各参数与理想值差的绝对值为逆向指标,分别计算

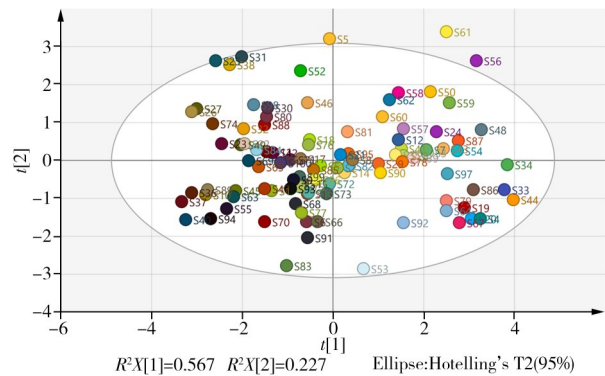


图5 血府逐瘀浸膏的OPLS-DA得分图

Fig. 5 OPLS-DA score plot of Xuefu Zhuyu extracts

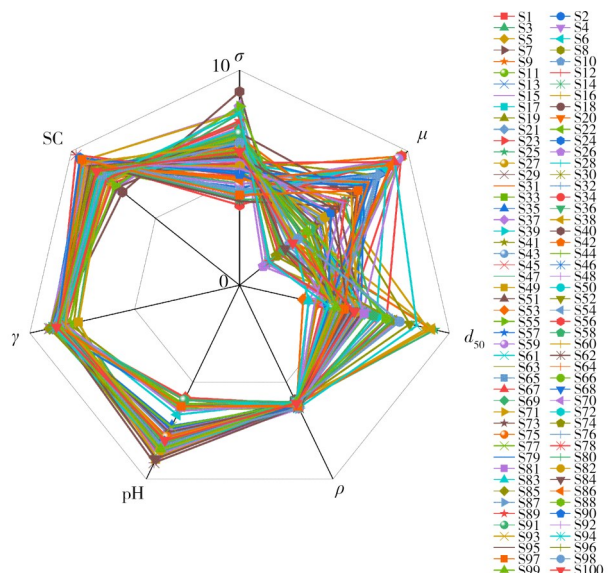


图6 血府逐瘀浸膏的叠加物理指纹图谱

Fig. 6 The combined physical fingerprint spectra of the Xuefu Zhuyu extracts

CRITIC 权重、熵权及组合权重。由图7可知，动力黏度的组合权重最高，其次是平均粒径，这表明这两项参数对浸膏质量的影响最大；值得注意的是，密度的组合权重甚至为0，说明其对浸膏质量几乎没有影响，这与企业当前以密度为唯一检测指标的现状形成鲜明对比。该权重分配也符合中药浸膏质量控制的实际需求，如动力黏度影响浸膏的后续制粒、压片。

基于组合权重与TOPSIS法计算了100批浸膏的综合得分，并以0.6为合格线划分浸膏质量等级。其中合格样本62批，不合格样本38批。需要说明的是，100批样本均已通过《药品管理法》、《中国药典》等国家标准规定的检测方法与合格阈值判定，为厂家合规放行产品。而本研究中对38批不合格样本的判定是基于构建的“多维度物理属性+综合评价模型”得出的结果，核心目标是实现浸膏质量的“精细化分级”而非单纯的“合格/不合格判定”，其评价指标覆盖了传统标准未涉及的工艺适配性相关物理属性，判定标准严于现行的国家标准，因此对于不合格样本的判定更为严格，导致不合格率相对较高。

2.3 模型性能分析

分别采用6种机器学习模型建立血府逐瘀浸膏质量分类模型，各模型性能参数如表4所示，混淆矩阵如图8所示。由表4可知，XGBoost模型的综合性能最优，其准确率、精确率、召回率、F1分数、AUC分别达到0.9333、1.0000、0.8333、0.9091、0.9630；随机森林、KNN、LightGBM 3种模型的性能相近，准确率均为0.8000，精确率为0.8000，F1分数为0.7273，其中KNN和LightGBM的AUC分别为0.8704、0.9074，证明模型的可靠性优于随机森林；神经网络模型和SVM的召回率最高(1.0000)，但二者精确率较低，导致其F1分数也较低。

从混淆矩阵(图8)可以看出，XGBoost模型将测试集中的一批不合格样本判定为合格样本，这对于浸膏的生产加工过程是不允许的。经核查发现其判错的样品为S23，其综合得分为0.5958(约等于合格阈值0.6)，因此推测误判原因可能是因为综合得分过于接近阈值导致特征差异不明显，所以本研究认为其模型的误判是可以理解的，可通过增加样本量进一步优化模型性能。而随机森林、KNN、LightGBM模型均存在3批样品误判的情况；神经网络模型虽未将不合格样品误判为合格品，但有4批合格样品被误判为不合格品；SVM模型同样未出现不合格样品误判为合格品，有3批合格样品被误判为不合格品。综合来看，XGBoost模型在血府逐瘀浸膏质量分类预测中具有较高的可靠性，可用于浸膏质量的快速判别。

表4 不同分类模型的性能参数

Table 4 Performance parameters of different classification models

Model	Accuracy	Precision	Recall	F1 Score	AUC
Random Forest	0.8000	0.8000	0.6667	0.7273	0.8421
KNN	0.8000	0.8000	0.6667	0.7273	0.8704
Neural Network	0.7333	0.6000	1.0000	0.7500	0.8704
XGBoost	0.9333	1.0000	0.8333	0.9091	0.9630
SVM	0.8000	0.6667	1.0000	0.8000	0.9074
LightGBM	0.8000	0.8000	0.6667	0.7273	0.9074

表3 血府逐瘀浸膏关键物理指标的方差和理想值

Table 3 The variances and ideal values of the key physical indicators of Xuefu Zhuyu extracts

Physical indicator	Variance	Ideal value
ρ	0.00965	6.16
pH	0.79895	8.20
σ	1.38411	6.11
μ	5.67420	4.46
d_{50}	1.80024	4.79
γ	0.06913	8.70
SC	0.33544	8.57

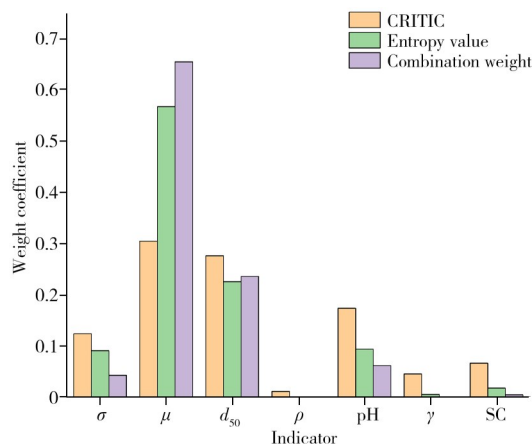


图7 血府逐瘀浸膏关键物理指标的3种权重系数图

Fig. 7 The three weight coefficient diagrams of the key physical indicators of Xuefu Zhuyu extracts

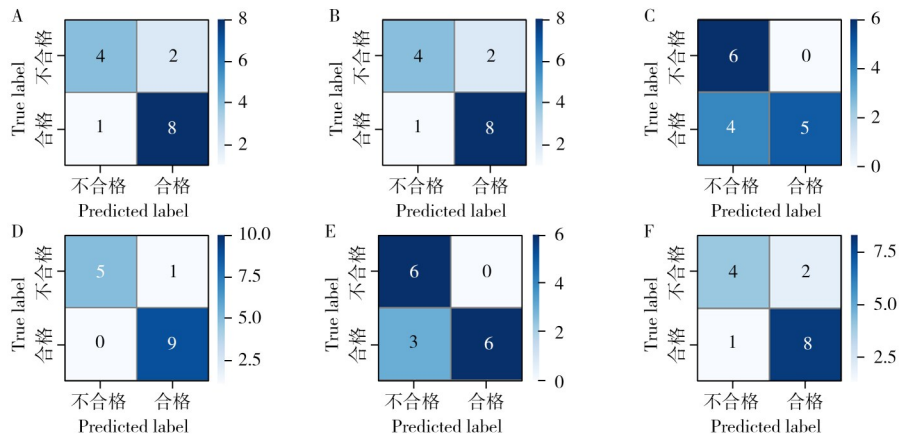


图8 不同模型的混淆矩阵图

Fig. 8 Confusion matrix diagrams of different models

A: random forest; B: KNN; C: neural network; D: XGBoost; E: SVM; F: LightGBM

2.4 智能评估平台的应用讨论

本研究完成了智能评估平台的构建并将其部署至公网服务器，地址为 <https://shaxin.pythonanywhere.com>。通过实际运行发现，系统可在 10 s 左右完成单张浸膏图像的全流程分析，并生成结构化报告与自然语言建议。豆包大模型基于真实物理特征生成的建议内容专业、具体、可操作，例如：“轮廓面积偏小(435.91 mm²)，圆整度较低(0.806)，建议适当降低浓缩温度并提高真空度以改善流动性”或“浸膏颜色均匀、形态规整，推荐采用喷雾干燥制粒后压片”。该设计在保持 XGBoost 模型原有高判别性能(准确率 93.33%)的同时，显著提升了结果的可解释性与实用性，有效弥合了“黑箱 AI”与“工艺人员”之间的认知鸿沟，使 AI 输出不仅“判得准”，更能“说得清”，便于一线人员理解与执行。相较于传统依赖比重计或经验观察的方法，本系统实现了无损、客观、多维、智能的质控升级，验证了“机器视觉+大模型”在中药中间体质量控制中的巨大潜力，为中药智能制造提供了可推广的“感知-决策-建议”一体化技术范式。

3 结论

本研究针对血府逐瘀浸膏质量评价中“单一指标代表性不足、化学检测效率低”的痛点，构建了“物理指纹图谱+机器视觉+大语言模型”多模态融合策略的浸膏质量智能评价体系，实现了从“数据感知”到“智能决策”再到“知识输出”的闭环。研究筛选 7 项核心物理参数构建物理指纹图谱，分析显示浸膏批次间物理属性整体一致性较好。经马氏距离剔除异常样本后，通过“CRITIC+熵权+TOPSIS法”实现浸膏质量的客观分级。在机器视觉建模方面，通过提取浸膏静态图像特征，构建 6 种机器学习分类模型，结果显示 XGBoost 模型性能最优，准确率达 0.933 3、AUC 为 0.963 0，仅 1 批综合得分 0.595 8(接近合格线)的样品误判，其优势源于模型对“物理参数-图像特征”关联关系的精准捕捉(如纹理特征与黏度、平均粒径强相关)。尤为关键的是，本研究进一步开发了集成大语言模型(Doubao, 字节跳动公司，基于大规模稀疏 MoE 与 Transformer 架构)的交互式智能评估网站：系统完成机器学习模型的判别后，能基于 AI 对多维特征自动生成专业、可操作的工艺建议，整套分析流程只需 10 s。

该联合评价体系不仅为血府逐瘀浸膏提供了“客观分级+快速预测+智能建议”的一体化质量控制方案，也为中药浸膏类中间体(如丹参浸膏、黄芪浸膏)的多维度质量评价提供了可复用的技术框架。未来可进一步拓展至中药制剂全链条质量控制，通过建立“浸膏物理属性-制剂工艺参数-成品药效”的关联模型，推动中药生产从“经验控制”向“数据驱动”转型。

但本研究也存在潜在局限性，如体系普适性需通过跨浸膏类型进行验证、模型性能依赖高质量数据标注，以及不同浸膏核心物理属性与外观特征可能存在差异，需针对性适配模型。对此，拟通过模块化框架设计(仅需替换核心指标与训练数据即可适配)、数据增强与鲁棒性处理、后续扩大样本覆盖范围等方式应对。

参考文献:

- [1] Gao H, Wan Q, Li Y H, Yang M, Qi Y R. *Chin. J. Pharm.* (高欢, 万琴, 李远辉, 杨明, 齐娅汝. 中国医药工业杂志), **2024**, 55(4): 475-481.
- [2] Dai X Y, Yang J, Ye Y L, Wang C J, Liu Y B, Zhang M, Chai X, Wen X L, Wang Y F. *J. Ethnopharm.*, **2025**, 348: 119-873.
- [3] Ballesteros Martínez M Á, Gaukel V. *Drying Technol.*, **2025**, 43(11/12): 1824-1832.
- [4] Wan Q, Gao H, Qi Y R, Qiu X R, Li Y S, Yang M, Li Y H, Wu Z F. *Chin. Tradit. Herb. Drugs*(万琴, 高欢, 齐娅汝, 邱勋荣, 李玉珊, 杨明, 李远辉, 伍振峰. 中草药), **2023**, 54(23): 7884-7894.
- [5] Wu S J, Zhang X Y, Guo S B, Zhou G F, Chen W, Li W L. *J. Instrum. Anal.* (吴思俊, 张晓阳, 郭抒博, 周桂芳, 陈伟, 李文龙. 分析测试学报), **2023**, 42(8): 1012-1025.
- [6] Huang Y P, Wu D Z, Tang Y M, Wang S, Wang Y. *China Pharm.* (黄玉普, 吴大章, 汤玉梅, 王森, 王玉. 中国药房), **2022**, 33(22): 2748-2752.
- [7] Li K, Chen D, Fu G C, Hou Q, Shen B B, Gong X B, Li R C. *Chin. J. Mod. Applied Pharm.* (李柯, 陈丹, 符国成, 侯茜, 沈冰冰, 龚小兵, 李若存. 中国现代应用药学), **2019**, 36(15): 1887-1891.
- [8] Wang D D, Chen X P, Bai S X, Wu Z M, Dong J Y, Yu X T. *China Pharm.* (王单单, 陈雪平, 白淑贤, 吴作敏, 董敬远, 于晓涛. 中国药房), **2025**, 36(11): 1329-1334.
- [9] Yang W, Zhou H Y, Qiu Z D, Yang J, Chen X. *Chin. Tradit. Herb. Drugs* (杨婉, 邹海英, 邱智东, 杨晶, 陈新. 中草药), **2023**, 54(6): 1804-1813.
- [10] You J, Li D, Wang Z, Chen Q, Q O Y. *J. Sci. Food Agric.*, **2024**, 104(9): 5486-5494.
- [11] Pérez-Calabuig A M, Pradana-López S, Cancilla J C, López-Ortega S, Torrecilla J S. *Food Control*, **2026**, 179: 111557.
- [12] Li Z, Cheng N T, Zhao X P, Tao Y, Xue Q L, Gong X C, Yu Y, Zhu J Q, Wang Y. *China J.Chin. Mater. Med.* (李正, 程宁涛, 赵筱萍, 陶益, 薛启隆, 龚行楚, 于洋, 朱捷强, 王毅. 中国中药杂志), **2025**, 50(12): 3197-3203.
- [13] Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China*. Part I. 2020 ed. Beijing: China Medical Science Press (国家药典委员会. 中华人民共和国药典. 一部. 2020版. 北京: 中国医药科技出版社), **2020**: 894-985.
- [14] Zhu Y P, Liu Y X, Shao M Q, Wu L. *Chin. J. Hosp. Pharm.* (朱亚苹, 刘玉鑫, 邵孟其, 吴磊. 中国医院药学杂志), **2025**, 45(12): 1349-1355.
- [15] Teng K X, Fu H, Wang Z C, Shen Y Q, Xie X Y, Zhao J, Qu H B. *Chin. Tradit. Herb. Drugs* (滕凯旋, 傅豪, 王中昌, 沈逸群, 谢欣媛, 赵洁, 瞿海斌. 中草药), **2022**, 53(3): 712-719.
- [16] Yuan B, Huang X T. *Compt. Appl.* (袁博, 黄宪通. 计算机应用), **2025**, 45(12): 3964-3969.

(责任编辑: 盛文彦)