

本地化大语言模型在胃癌术前药物重整中的应用模式构建与实践^Δ

朱宇轩^{1,2*}, 张冀中^{1,2}, 孙雨豪^{1,2}, 温佳瑜^{1,2}, 刘欣¹, 魏继福¹, 黄凌莉^{1#} (1. 江苏省肿瘤医院药学部/江苏省肿瘤防治研究所/南京医科大学附属肿瘤医院/江苏省恶性肿瘤先进诊疗重点实验室, 南京 210009; 2. 中国药科大学基础医学与临床药学学院, 南京 211198)

中图分类号 R969.3;R735.2 文献标志码 A 文章编号 1001-0408(2026)08-1062-06
DOI 10.6039/j.issn.1001-0408.2026.08.16



摘要 **目的** 构建本地化大语言模型(LLM)辅助胃癌术前药物重整模式,并进行效果评价。**方法** 回顾性纳入2024年1月至2026年1月江苏省肿瘤医院胃外科249例入院前存在持续用药史的胃癌患者。根据时间先后将患者划分为训练集(154例)和验证集(95例)。基于指南、药品说明书等证据,构建标准化药物重整流程与结构化知识库,并在院内私有化部署DeepSeek-V3 LLM,结合检索增强生成技术,实现对用药信息的自动整合、风险筛查以及个体化建议生成。采用机器评分(BERT Score和ROUGE-1、2、L)与人工评分[七维指标(7DI)]评价LLM生成建议的质量;运用Spearman相关分析探究机器评分与人工评分的相关性;采用Cronbach's α 系数检验人工评分结果的内部一致性;比较不同难易程度(简单、中等、高难度3个等级)药物重整任务的人工与LLM药物重整耗时。**结果** 最终构建了涵盖8大类药物、能够覆盖常见及高风险术前用药场景的结构化知识库。机器评分方面,BERT Score的精确率为 0.783 ± 0.033 ,召回率为 0.811 ± 0.038 ,F1分数为 (0.796 ± 0.028) 分;ROUGE-1、ROUGE-2和ROUGE-L 3个层级的F1分数分别为 (0.566 ± 0.067) 、 (0.338 ± 0.076) 和 (0.468 ± 0.082) 分。3名人工评分者的7DI评分为32.06~33.45分。机器评分的F1分数与人工评分的7DI评分均呈显著正相关(最高决定系数=0.611, $P < 0.001$),且人工评分内部一致性良好(Cronbach's $\alpha = 0.876$)。在效率方面,与人工药物重整耗时比较,LLM药物重整耗时在简单组、中等组、高难度组中均减少90%以上($P < 0.001$)。**结论** 基于本地化LLM与结构化知识库构建的药物重整模式,在胃癌术前复杂用药场景中具有较高的准确性、一致性和临床可用性,能够提升药物重整效率,同时降低潜在用药风险。**关键词** 药物重整;人工智能;大语言模型;胃癌;术前用药;用药安全

Construction and practice of application model for localized large language model in preoperative medication reconciliation for gastric cancer

ZHU Yuxuan^{1,2}, ZHANG Jizhong^{1,2}, SUN Yuhao^{1,2}, WEN Jiayu^{1,2}, LIU Xin¹, WEI Jifu¹, HUANG Lingli¹ (1. Dept. of Pharmacy, Jiangsu Cancer Hospital/Jiangsu Institute of Cancer Research/Nanjing Medical University Affiliated Cancer Hospital/Jiangsu Key Laboratory of Innovative Cancer Diagnosis and Therapeutics, Nanjing 210009, China; 2. School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 211198, China)

ABSTRACT **OBJECTIVE** To construct a preoperative medication reconciliation model assisted by a localized large language model (LLM) for gastric cancer and evaluate its clinical efficacy. **METHODS** A total of 249 gastric cancer patients with a history of continuous medication before admission in the Gastric Surgery Department of Jiangsu Cancer Hospital were retrospectively enrolled. Patients were divided into training set (154 cases) and validation set (95 cases) based on the order of time. Based on guidelines, drug package inserts, and other evidence, a standardized medication reconciliation process and a structured knowledge base were constructed. DeepSeek-V3 LLM was deployed privately in the hospital, combined with retrieval-augmented generation technology, to achieve automated integration of medication information, risk screening, and generation of personalized recommendations. The quality of LLM-generated recommendations was evaluated using automatic metrics (BERT Score and ROUGE-1, 2, L) and manual scoring [seven-dimensional index (7DI)]. Spearman correlation analysis was performed to explore the correlation between automatic scores and manual scores. Cronbach's α coefficient was used to test the internal consistency of manual scoring results. The time consumed by manual and LLM-assisted medication reconciliation was compared across tasks of different difficulty levels (simple, moderate, and high). **RESULTS** A structured knowledge base covering 8

^Δ **基金项目** 江苏省药学会“药”研新声药学科科研项目(No. 202564052);江苏省肿瘤医院科技发展基金项目(No.ZYGL202402, No.ZYGL202502);江苏省肿瘤医院高水平医院学科建设“群峰计划”项目(No.DFXK202501)

* **第一作者** 硕士研究生。研究方向:医院药学。E-mail: yuxuanzhu_123@163.com

通信作者 主管药师,硕士。研究方向:医院药学。E-mail: huang_lingli@163.com

major drug categories was finally established, covering common and high-risk preoperative medication scenarios and providing structured retrieval support for the LLM. For automatic evaluation, the precision, recall, and F1-score of BERT Score were 0.783 ± 0.033 , 0.811 ± 0.038 , and 0.796 ± 0.028 , respectively. The F1-scores of ROUGE-1, ROUGE-2 and ROUGE-L were 0.566 ± 0.067 , 0.338 ± 0.076 and 0.468 ± 0.082 , respectively. The 7DI scores from three manual raters ranged from 32.06 to 33.45. The F1-score of automatic scoring was significantly positively correlated with the 7DI score of manual scoring (maximum coefficient of determination=0.611, $P < 0.001$), and the internal consistency of manual scoring was good (Cronbach's $\alpha = 0.876$). In terms of efficiency, LLM-assisted medication reconciliation reduced time consumption by more than 90% compared with manual reconciliation in the simple, moderate, and high-difficulty groups ($P < 0.001$). **CONCLUSIONS** The medication reconciliation model constructed based on a localized LLM and structured knowledge base shows high accuracy, consistency, and clinical applicability in complex preoperative medication scenarios for gastric cancer. It can improve the efficiency of medication reconciliation and reduce potential medication risks.

KEYWORDS medication reconciliation; artificial intelligence; large language model; gastric cancer; preoperative medication; medication safety

药物重整是指在患者就诊、转科、手术或出院等医疗过渡阶段,对其既往及当前用药信息开展系统、精准且可追溯的收集、核对与优化工作,旨在降低用药差错以及药物相关不良事件的发生风险。世界卫生组织已将药物重整列为提升用药安全的关键策略之一,着重强调其在多药联合治疗以及复杂临床路径管理中的核心地位^[1]。然而,在实际临床工作中,药物重整仍存在诸多实施难点与风险暴露点,例如患者用药史信息分散,以及针对抗凝、降糖等高风险药物的术前调整方案复杂且缺乏统一执行标准。

胃癌是全球范围内发病率和死亡率均较高的恶性肿瘤之一,中国人群中,45岁及以上为胃癌的高发阶段,且发病率会随年龄的增长而上升^[2],该类人群常合并有高血压、糖尿病、冠心病等多种慢性疾病,故需长期服用多种药物^[3]。目前,可切除胃癌患者的标准治疗以手术为主,术前阶段是患者药物治疗方案发生显著变动的时期,极易出现用药疏漏、重复、相互作用、停药或续药遗漏等问题^[3-5]。因此,药物重整显得尤为重要。国内外研究均表明,实施规范化的药物重整,尤其是有职业药师参与的药物重整,能够显著减少住院患者的药物差异和处方错误,进而提升用药安全性^[6-7]。然而,在实际工作中,由于药物重整通常依靠人工核对,其流程涉及医院电子病历、患者访谈、家属反馈等多方面;使得信息碎片化、核对效率低下、人员经验参差不齐等问题普遍存在^[8]。

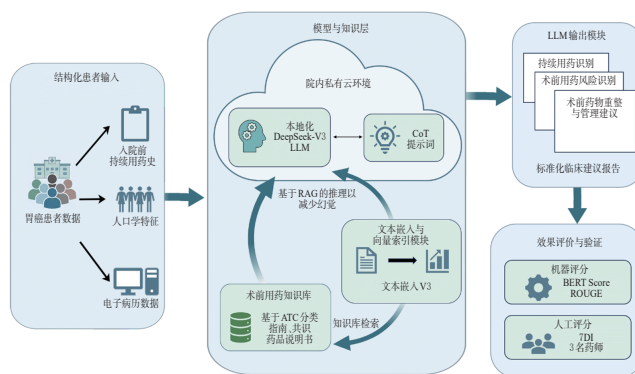
随着大型语言模型 (large language model, LLM) 在临床知识理解与临床推理能力方面取得显著进展,该模型已被证实具备对病历信息进行结构化、术语归一化以及临床知识检索的潜在能力^[9-10]。作为开源模型,DeepSeek理论上能够在机构内部进行部署,可以以安全且低成本的方式开展训练与应用,这是其相较于其他 LLM 的优势^[11-12]。目前,尚未有研究将本地化 LLM 应用于肿瘤患者的药物重整与用药管理。为此,本研究以胃癌患者术前用药为切入点,通过在院内私有化部署 LLM 并整合术前用药知识库,构建了一套术前智能化药物重整模式,以实现自动整合、风险筛查以及个

体化建议生成,旨在为提高用药重整的效率,减少人为差错,提升患者用药安全性提供可推广的技术路径。

1 本地化 LLM 辅助胃癌术前药物重整模式的构建

1.1 研究设计与整体技术流程

本研究聚焦于江苏省肿瘤医院(以下简称“本院”)胃癌术前复杂的用药场景,构建了本地化 LLM 辅助胃癌术前用药重整模式,具体技术路径见图 1。如图 1 所示,本研究首先针对入院前有持续用药史的胃癌患者开展结构化数据整理工作。在院内私有化部署的 LLM 环境中,结合术前用药知识库与 RAG 技术,实现对持续用药的自动整合、风险筛查以及个体化建议生成。随后,通过机器评分与人工评分,对 LLM 生成建议的质量和临床可用性进行系统评价。



ATC:解剖学治疗学化学分类系统;CoT:思考链;RAG:检索增强生成;BERT Score:基于语义嵌入的文本生成质量评价指标,用于衡量生成文本与参考文本在语义层面的相似度;ROUGE:基于n-gram重叠的自动文本评价指标,用于衡量生成文本与参考文本在关键信息覆盖程度的相似性;7DI:七维指标。

图1 本地化 LLM 辅助胃癌术前药物重整模式的技术路径

1.2 资料来源与分组

回顾性收集 2024 年 1 月至 2026 年 1 月在本院胃外科住院的 249 例胃癌患者作为研究对象。本研究的纳入标准为——(1)年龄 ≥ 18 岁;(2)经病理学确诊为胃恶性肿瘤;(3)初诊且拟行手术治疗的患者;(4)有长期服药

史:术前至少连续4周服用 ≥ 1 种药物;(5)具备完整的术前评估资料。本研究的排除标准为:(1)拒绝手术或最终未行手术的患者;(2)无法获取完整用药史的患者;(3)行急诊手术的患者。收集患者的基础信息、既往病史、诊疗记录等资料。所有涉及患者隐私的信息均予以剔除,并整理成结构化、标准化的输入格式,以构建标准化的输入格式数据集,作为LLM应用的基础任务语料库。本研究方案已获得本院伦理委员会批准(批件号:YK-2025-189),针对去标识化的患者数据开展回顾性分析,豁免患者知情同意。

将2024年1月至2025年9月收治的154例患者划分为训练集,2025年10月至2026年1月收治的95例患者划分为验证集。

1.3 药物重整场景机器人构建与提示词设计

本研究依托本院的本地化LLM部署环境展开,以FastGPT平台(V 4.9.0)作为基础平台,采用私有化部署方式,运行于私有云服务器中。通过严格的技术隔离与访问控制策略,确保所有数据的存储、处理和调用均局限于该环境内部,不会向外部流转。本研究部署了本地化的DeepSeek-V3版本LLM(671B),设定温度为0、Top_p值为0.9,研究团队在此基础上开展场景应用模式构建。

完成基础部署后,进一步构建面向术前药物重整任务的对话代理。于FastGPT平台(V 4.9.0)新建“药物重整辅助工作台”,并对对话代理进行配置。针对术前药物重整的任务特性,选用CoT设计提示词,将术前药物重整任务拆分为若干具有明确临床逻辑的认知阶段。相关提示词的组成要素与设计目的见表1。

表1 基于CoT的术前药物重整提示词设计

认知阶段	提示词核心内容	设计目的
角色与任务定义	明确模型以胃外科临床药师身份执行药物重整与用药管理任务	将模型行为约束为专业药师视角,避免通用文本生成
信息整理与标准化	规范患者信息与用药清单,统一药品通用名,对缺失剂量或频次明确标注“未提及”,不进行假设	确保用药史核对准确,降低模型幻觉风险
药物问题识别	要求对每一种药物逐一与用药原则进行比对,识别潜在风险与管理问题	防止遗漏用药,保证评估的系统性
术前用药重整辅助决策	针对识别问题生成具体可操作的术前用药管理建议,包括停药或重启、桥接治疗及监测要点,复方制剂作为整体处理	模拟真实临床药师决策逻辑,提升用药建议的临床可行性
结构化结果输出	按固定格式输出标准化用药清单与分条管理建议	提高结果一致性、可读性及后续评价的可重复性

1.4 知识库的构建与更新

在模式构建阶段,训练集主要用于流程与知识体系的构建和优化。依据国内外围手术期指南中关于术前用药调整的内容^[13-15]以及药品说明书,构建胃癌术前药物重整的标准化流程框架;同时,结合真实临床处方数据,验证流程与规则的覆盖性和可执行性,并持续完善药物目录及规则描述,该目录基于训练集数据整理而成。本研究聚焦于胃癌患者入院前有持续用药史的药物重整。因此,在知识库构建过程中,虽参考了涵盖术前、术中、术后全程的围手术期用药指南,但重点提取并纳入的是与术前用药管理直接相关的内容。依据ATC

对药物进行分类,构建适用于术前药物重整的常用药物目录。接着,通过专家(包括中级及以上职称且工作年限 ≥ 5 年的临床医师、药师、护师)咨询,确定药品目录的范围能够涵盖临床常见用药场景。将所有文档导入FastGPT知识库模块,借助text-embedding-v3索引模型构建文本向量化索引,形成一个可供LLM自动检索的知识库。知识库搜索参数设置引用上限为20 000 tokens,最低相关度为0.7。

在上述知识库初步构建完成后,进一步结合训练集数据,对知识库内容以及相关决策逻辑展开迭代优化。首先对训练集中的154例胃癌患者的术前用药目录进行系统梳理,形成初始药物目录;随后在真实的临床场景中,对本地化LLM辅助胃癌术前药物重整模式开展场景化验证,识别其在复方制剂成分拆解、商品名标准化、抗血栓药物的停药与重启时机判断、降糖药物分类管理这4个关键环节中存在的偏差和不足。针对上述问题,对知识库内容和规则体系进行有针对性的补充与修订,同时对提示词引导的决策逻辑进行同步优化。经过多轮迭代后,逐步形成能够覆盖常见及高风险术前用药场景的结构化知识体系,为后续模式应用与效果评价奠定基础。

为确保知识库与术前用药临床实践及循证医学证据实现同步更新,本研究构建了具备分层级、可操作性的知识库动态更新与治理机制,具体如下——(1)指南与共识更新对接机制:临床药师按照固定周期对知识库进行更新维护,每3个月对国内外相关指南、专家共识中的新内容开展系统检索与筛选。纳入流程采用“初筛量表+AGREE II(临床指南研究与评价工具)质量评价+德尔非法专家共识”的三阶段机制,以确保入库文献在权威性、循证质量以及本院适用性方面均满足临床要求。(2)新药与药品信息动态录入机制:针对新获批药物以及药品说明书的变更信息,由临床药师协同药事管理部门按季度进行梳理。提取围手术期相关的适应证、禁忌证、停药和重启时机、替代方案以及风险提示等关键要素,经结构化标注后纳入知识库;同时,对已入库的药品信息实施版本管理,标注更新日期和证据来源,避免使用过期知识。

经过多轮迭代优化,最终构建了能够覆盖常见及高风险术前用药场景的结构化知识库。此知识库共涵盖8大类药物、19类亚组、68个药物亚类,包含326种同一通用名药物,具体目录见表2。

2 本地化LLM辅助胃癌术前药物重整模式的效果评价

2.1 效果的评价实施

在完成本地化LLM辅助胃癌术前药物重整模式构建后,将该模式应用于验证集患者,以评价其在真实临床场景中的应用效果。由本地化LLM辅助胃癌术前药物重整模式自动生成药物重整建议,作为生成文本;由临床药师进行人工药物重整,作为参考文本。

表2 胃癌患者术前用药知识库

药物种类	亚组	纳入亚组 数量/类	药物品种数/ 种	具体亚类	
心血管系统用药	抗高血压药	7	51	血管紧张素转化酶抑制剂、血管紧张素Ⅱ受体阻滞剂、β受体阻滞剂、钙通道阻滞剂、利尿剂、α受体阻滞剂、复方制剂	
	抗心绞痛药	4	24	硝酸酯类、β受体阻滞剂、钙通道阻滞剂、抗血小板药	
	降脂药	4	16	他汀类、烟酸和苯氧酸衍生物、胆汁酸螯合剂、胆固醇吸收抑制剂	
	血液及造血系统用药	抗血栓药	3	11	维生素K拮抗剂、新型口服抗凝药、抗血小板药
	消化与代谢系统用药	降糖药	9	46	双胍类、磺脲类、格列奈类、α-糖苷酶抑制剂、噻唑烷二酮类、DPP-4抑制剂、SGLT-2抑制剂、胰岛素、复方制剂
全身激素类用药	营养治疗药	3	10	整蛋白肠内营养制剂、短肽型肠内营养制剂、整蛋白纤维型肠内营养制剂	
	止吐药	3	15	5-羟色胺3型受体拮抗剂、多巴胺受体拮抗剂、神经激肽-1受体拮抗剂	
	止泻药	3	7	阿片类受体激动剂、肠道物理吸附剂、脑啡肽酶抑制剂	
	抑酸药	2	11	质子泵抑制剂、H2受体拮抗剂	
	促胃肠动力药	2	4	多巴胺受体拮抗剂、5-羟色胺4型受体激动剂	
神经系统用药	甲状腺激素	1	2	甲状腺激素类	
	抗甲状腺药	1	3	硫脲类	
呼吸系统用药	抗精神病药	1	6	非典型抗精神病药	
	抗焦虑药	4	14	苯二氮卓类、氯螺环癸二酮类、抗组胺药、氨基甲酸类	
	抗抑郁药	5	21	选择性5-羟色胺再摄取抑制剂、5-羟色胺-去甲肾上腺素再摄取抑制剂、三环类抗抑郁药、单胺氧化酶抑制剂、非典型抗抑郁药	
肌肉骨骼系统用药	平喘药	4	15	吸入性β ₂ 受体激动剂、吸入性抗胆碱能药、吸入性糖皮质激素、白三烯抑制剂	
	风湿性疾病药	4	29	非甾体抗炎药、改善病情抗风湿药、糖皮质激素、免疫抑制剂	
全身用抗感染药	抗痛风药	3	13	抑制尿酸生成药、促尿酸排泄药、抗痛风发作药	
	抗菌药	5	28	头孢菌素类、氟喹诺酮类、氨基糖苷类、硝基咪唑类、克林霉素类	

DPP-4:二肽基肽酶-4;SGLT-2:钠-葡萄糖协同转运蛋白2。

2.2 评价指标与方法

2.2.1 评价指标

采用机器评分与人工评分相结合的方式,对本地化LLM辅助胃癌术前药物重整模式自动生成的药物重整建议的质量进行评价。(1)机器评分:运用BERT Score和ROUGE(包括ROUGE-1、ROUGE-2、ROUGE-L)分别计算相应的精确率、召回率和F1分数,将生成文本和参考文本进行对比^[16-17]。本研究运用Python的Jieba库开展文本分词工作,借助BERT Score/ROUGE库进行机器评分计算,选用“bert-base-chinese”预训练模型,以精准计算中文文本的BERT Score。(2)人工评分:借鉴电子病历人工评审方法,采用7DI评分体系^[18],根据研究内容对其中部分指标进行调整,涉及信息完整性、术语使用规范性、逻辑一致性、语法结构、文体风格、问题识别全面性、建议可行性与针对性7个维度。所有生成的用药建议均由3名经验丰富的药师(编号1~3)进行评价。

为量化评估本地化LLM辅助胃癌术前药物重整相较于人工药物重整的效率优势,本研究引入“药物重整

耗时”作为过程性结局指标。依据患者疾病及术前用药的复杂程度,将药物重整任务划分为简单、中等、高难度3个等级(由2名药师在不知晓LLM输出结果的前提下进行分级,若存在分歧则通过讨论达成一致)。人工药物重整耗时指参与评价的临床药师在模拟真实工作流程的情况下完成药物重整的时间,即从开始核对患者既往用药信息并检索相关指南资料起,至形成完整的用药建议为止,所记录的总耗时为人工药物重整耗时。LLM自动生成的药物重整耗时由系统后台自动记录,即从接收完整结构化患者信息输入开始,至生成完整药物重整建议结束,所记录的系统运行时间为LLM药物重整耗时。根据药物重整任务划分的难易程度进行分组,分别计算人工药物重整耗时与LLM药物重整耗时,以比较两种模式在不同任务复杂程度下的效率差异。

2.2.2 统计学方法

采用Python(3.13版本)和Excel 2021进行数据处理。本研究中,符合正态分布的计量资料以 $\bar{x} \pm s$ 表示,不符合正态分布的则以 $M(P_{25}, P_{75})$ 表示。运用Spearman相关分析探究机器评分与人工评分的相关性;采用Cronbach's α 系数检验3名药师评分结果的内部一致性, $\alpha \geq 0.80$ 表示内部一致性良好^[19]。采用Mann-Whitney U检验比较人工药物重整耗时和LLM药物重整耗时的差异。检验水准 $\alpha = 0.05$ 。

2.3 本地化LLM生成药物重整建议的评价结果

2.3.1 机器评分

BERT Score平均得分如下:精确率为 0.783 ± 0.033 ,召回率为 0.811 ± 0.038 ,F1分数为 (0.796 ± 0.028) 分。在ROUGE-1、ROUGE-2和ROUGE-L 3个层级上的精确率分别为 0.542 ± 0.099 、 0.317 ± 0.087 、 0.434 ± 0.096 ;召回率分别为 0.613 ± 0.101 、 0.378 ± 0.097 、 0.530 ± 0.113 ;F1分数分别为 (0.566 ± 0.067) 、 (0.338 ± 0.076) 、 (0.468 ± 0.082) 分。结果见图2。

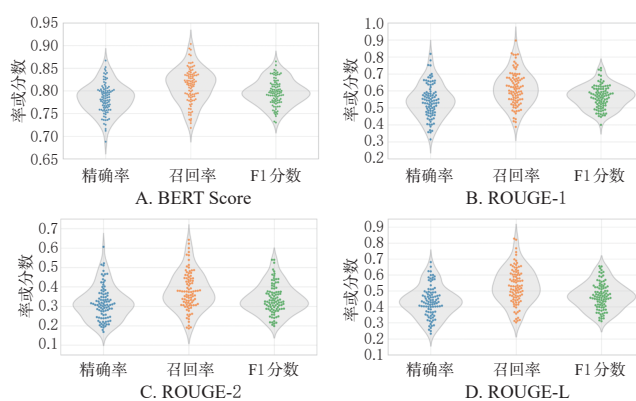


图2 生成文本与参考文本的BERT Score与ROUGE评分小提琴图

2.3.2 人工评分

3名药师的7DI评分分别为 (33.45 ± 1.49) 、 (32.06 ± 1.84) 、 (32.48 ± 1.65) 分。Spearman相关分析结果显示,机器评分的F1分数与3名药师的7DI评分均呈显著正

相关,其中,与药师(编号2)的评分结果的相关性最高[决定系数(R^2)=0.611, $P<0.001$],其次是药师(编号1)的评分结果($R^2=0.538$, $P<0.001$),与药师(编号3)的评分结果的相关性相对较低($R^2=0.487$, $P<0.001$)。内部一致性检验结果显示,Cronbach's $\alpha=0.876$,表明3名药师的内部一致性良好($\alpha\geq 0.80$)。

2.3.3 药物重整耗时

简单组、中等组、高难度组药物重整任务分别有65、21、9份。与人工药物重整耗时比较,LLM药物重整耗时在简单组、中等组、高难度组中均显著减少(减少90%以上, $P<0.001$)。结果见表3。

表3 不同难度组的人工药物重整耗时与LLM药物重整耗时比较[M(P_{25} , P_{75}),s]

组别	药物重整任务份数	人工药物重整耗时	LLM药物重整耗时	U	P
简单组	65	185.00(138.00,242.00)	11.15(9.78,12.84)	4 225.00	<0.001
中等组	21	349.00(291.00,414.00)	14.68(13.24,16.31)	441.00	<0.001
高难度组	9	733.00(697.00,914.00)	19.29(17.98,21.00)	81.00	<0.001

3 讨论

3.1 本地化LLM辅助胃癌术前药物重整模式的可行性与准确性

本研究依托本院私有化部署LLM环境,针对胃外科术前用药场景,构建了胃癌术前药物重整模式,结果显示,在验证集中BERT Score的F1分数为0.796,这表明该模式在语义内容覆盖与忠实度方面表现良好,具有较强的临床语义理解与复现能力^[16]。其生成的药物重整建议的人工评分(7DI评分)在32.06~33.45分之间,这表明LLM的输出在信息完整性、问题识别全面性等临床维度上具有较高质量,能够满足临床评价要求。进一步分析结果显示,机器评分的F1分数与3名药师的7DI评分均呈显著正相关(R^2 最高为0.611),且3名药师的内部一致性良好(Cronbach's $\alpha=0.876$)。以上结果表明,机器评分与药师人工评分在整体趋势上高度一致,既验证了自动评价指标在本研究中的适用性,也为后续大规模应用中的质量监控提供了可替代人工的量化工具。

3.2 本地化LLM辅助胃癌术前药物重整模式在关键用药场景中的优化路径与效果

术前药物重整的核心挑战在于应对临床用药的复杂性与异质性。用药信息不一致、未停药或误停药、药物相互作用以及剂量不当等状况,均可能引发术后出血、心血管并发症、感染等不良事件^[20]。本研究基于训练集对模型进行系统验证,识别出模型在药物重整环节存在的4类核心偏差,并通过针对性优化,形成了一套能够覆盖常见及高风险用药场景的辅助决策流程。

复方制剂与商品名识别是术前用药重整的基础环节。训练集验证表明,当患者使用复方制剂(如氯沙坦钾氢氯噻嗪),或电子病历仅记录商品名时,模型初期对活性成分的识别存在遗漏或错误。针对这一问题,本研究完善了药物术语标准化规则,在提示词中明确要求对复方制剂进行成分拆解、将商品名映射至通用名,并通过知识库补充常用商品名与通用名的对应关系。抗血

栓药物的停药与重启时机的判断,是术前用药管理中风险最高的环节之一。华法林、氯吡格雷、阿司匹林等药物会抑制凝血功能,若术前未及时停用或调整,可能导致术中出现难以控制的出血情况^[21]。这些药物的重整处理需综合考量药物半衰期、手术出血风险和患者血栓栓塞风险,传统上高度依赖临床经验。本研究借助提示词,能够依据药物种类自动识别不同药物的干预点,进而提高抗血栓管理的精准度。降糖药物的分类管理同样存在显著的类别差异,模型初期对部分高风险降糖药物的药物重整处理要点有所遗漏。本研究通过补充不同类别降糖药物的管理要点,在知识库中明确标注了术前停药时间、术中替代方案以及术后重启指征。优化后,模型对降糖药物的处理建议更为系统全面,有助于降低围手术期血糖波动相关风险。

3.3 本地化LLM辅助胃癌术前药物重整模式的创新与效率优势

目前,LLM在医疗领域具有较大潜力,但多集中在诊断、文档撰写、问答支持等较为通用的任务层面^[22-24]。本研究在药物重整和用药管理这一细分任务中,尤其是在术前用药的临床环境下,尝试将LLM进行知识库化并实现机器人化,建立了仅服务于单个任务的LLM。该模式不仅达成了智能化的信息整合,改善了传统模式下药师需手动在多个系统间切换、查阅纸质资料的耗时状况,使药师从重复性劳动中解放,而且显著提高了工作效率,促使药师能够将宝贵的时间与精力投入到更高级别的临床服务中。在药物重整中最为耗时且高度依赖知识检索与综合判断的是“用药建议生成环节”,本研究结果显示,在不同难易程度任务中,LLM药物重整耗时仅为人工药物重整耗时的10%以内。这种效率提升在高复杂度病例中尤为明显,这表明该模式在真实的临床场景中具有显著节省时间的潜力,进而可能间接优化药师人力资源配置并提升服务可及性。

3.4 本地化LLM与知识库的优势

既往研究显示,在医疗环境中直接运用通用LLM存在可解释性欠佳、知识不透明、与本地临床流程脱节等风险^[25]。本研究在本院可控的环境下运用LLM,并构建本院的知识库,而非直接调用通用大模型。此举有助于提升模型建议的临床适用性,降低通用模型因脱离具体场景而产生的偏差或幻觉风险,还能有效规避诸多数据隐私问题^[26]。本研究依托本地化知识库,在胃癌术前用药场景中得以应用,为该场景下的药物重整实践提供了初步的实证参考;同时,本研究采用真实患者数据,而非使用仿真或模拟数据,增强了研究的现实意义。

3.5 研究局限性

本研究仍存在一定局限性。首先,研究样本量与时间跨度有限,这可能会对LLM建议效果的稳定性造成影响。其次,知识库是根据近2年的用药目录构建而成的,尽管已初步建立更新机制,但尚未对其稳定性进行长期验证;且低频用药类别(如抗精神病药、免疫抑制

剂)由于真实病例覆盖不足,缺乏充分的情景化验证。此外,尽管本研究构建的“本地化LLM+知识库+RAG技术路径”具有良好的可复制性,但由于面临医院信息系统异构性、临床流程差异以及知识库本地化适配等问题,其在真实世界医疗场景中的落地仍面临一定挑战。需要说明的是,这种局限性并不意味着该技术路径不可推广,而是临床人工智能系统在跨机构应用中的普遍特征。

综上所述,基于本地化LLM与结构化知识库构建的药物重整模式,在胃癌术前复杂用药场景中具有较高的准确性、一致性和临床可用性,能够提升药物重整效率,同时降低潜在用药风险。

参考文献

- [1] World Health Organization. Medication safety in polypharmacy: technical report[EB/OL]. [2025-11-11]. <https://www.who.int/publications/i/item/WHO-UHC-SDS-2019.11>.
- [2] SUNDAR R, NAKAYAMA I, MARKAR S R, et al. Gastric cancer[J]. *Lancet*, 2025, 405(10494):2087-2102.
- [3] 李苗钰,刘凯,张维汉,等. 全球及中国胃癌的流行病学特点及趋势:2018—2022《全球癌症统计报告》解读[J]. *中国普外基础与临床杂志*, 2024, 31(10):1236-1245.
- [4] LERTKOVIT S, SIRIUSAWAKUL A, SURARUNSUMRIT P, et al. Polypharmacy in older adults undergoing major surgery: prevalence, association with postoperative cognitive dysfunction and potential associated anesthetic agents[J]. *Front Med*, 2022, 9:811954.
- [5] 马晓波,马小桐,马建清,等. 1990—2021年中国肿瘤疾病负担分析[J]. *医学新知*, 2025, 35(8):885-891
- [6] YAMADA Y, KOBAYASHI R, YAMAMOTO T, et al. Medication reconciliation by pharmacists for pre-admission patients improves patient safety[J]. *J Pharm Health Care Sci*, 2024, 10(1):19.
- [7] MUELLER S K, SPONSLER K C, KRIPALANI S, et al. Hospital-based medication reconciliation practices: a systematic review[J]. *Arch Intern Med*, 2012, 172(14):1057-1069.
- [8] KENNELTY K A, CHEWNING B, WISE M, et al. Barriers and facilitators of medication reconciliation processes for recently discharged patients from community pharmacists' perspectives[J]. *Res Soc Adm Pharm*, 2015, 11(4):517-530.
- [9] HE K, MAO R, LIN Q K, et al. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics[J]. *Inf Fusion*, 2025, 118:102963.
- [10] SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge[J]. *Nature*, 2023, 620(7972):172-180.
- [11] TEMSAH A, ALHASAN K, ALTAMIMI I, et al. DeepSeek in healthcare: revealing opportunities and steering challenges of a new open-source artificial intelligence frontier[J]. *Cureus*, 2025, 17(2):e79221.
- [12] SANDMANN S, HEGSELMANN S, FUJARSKI M, et al. Benchmark evaluation of DeepSeek large language models in clinical decision-making[J]. *Nat Med*, 2025, 31(8):2546-2549.
- [13] 中华医学会糖尿病学分会. 中国2型糖尿病防治指南:2013年版[J]. *中华糖尿病杂志*, 2014, 6(7):447-498.
- [14] 唐密密,许世伟,葛杰,等. 胃癌患者围手术期药学监护专家共识[J/OL]. *今日药学*, 2025:1-26[2026-01-10]. <https://link.cnki.net/urlid/44.1650.R.20250110.1457.007>.
- [15] OPREA A D, KESHOCK M C, O'GLASSER A Y, et al. Preoperative management of medications for psychiatric diseases society for perioperative assessment and quality improvement consensus statement[J]. *Mayo Clin Proc*, 2022, 97(2):397-416.
- [16] ZHANG T Y, KISHORE V, WU F, et al. BERTScore: evaluating text generation with BERT[EB/OL]. [2025-11-11]. <https://arxiv.org/pdf/1904.09675>.
- [17] LIN C Y. ROUGE: a package for automatic evaluation of summaries[EB/OL]. [2025-11-11]. <https://www.microsoft.com/en-us/research/publication/rouge-a-package-for-automatic-evaluation-of-summaries/>.
- [18] DING H L, XIA W J, ZHOU Y J, et al. Evaluation and practical application of prompt-driven ChatGPTs for EMR generation[J]. *NPJ Digit Med*, 2025, 8(1):77.
- [19] GEORGE D, MALLERY P. IBM SPSS statistics 29 step by step: a simple guide and reference[M]. 18th Ed. New York:Routledge, 2024:251-262.
- [20] YE J C. Patient safety of perioperative medication through the lens of digital health and artificial intelligence[J]. *JMIR Perioper Med*, 2023, 6:e34453.
- [21] 中国心胸血管麻醉学会非心脏麻醉分会,中国医师协会心血管内科医师分会,中国心血管健康联盟. 抗血栓药物围手术期管理多学科专家共识[J]. *中华医学杂志*, 2020, 100(39):3058-3074.
- [22] CASCELLA M, MONTOMOLI J, BELLINI V, et al. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios[J]. *J Med Syst*, 2023, 47(1):33.
- [23] CHINTAGUNTA B, KATARIYA N, AMATRIAIN X, et al. Medically aware GPT-3 as a data generator for medical dialogue summarization[EB/OL]. [2025-11-11]. <https://arxiv.org/pdf/2110.07356>.
- [24] 张微微,唐婧,冀召帅,等. 人工智能在临床药学中的应用进展[J]. *药学学报*, 2025, 60(10):3045-3059.
- [25] LI H Y, FU J F, PYTHON A. Implementing large language models in health care: clinician-focused review with interactive guideline[J]. *J Med Internet Res*, 2025, 27:e71916.
- [26] WIEST I C, FERBER D, ZHU J F, et al. Privacy-preserving large language models for structured medical information retrieval[J]. *NPJ Digit Med*, 2024, 7(1):257.
(收稿日期:2026-01-16 修回日期:2026-03-13)
(编辑:舒安琴)