

病理报告中自然语言处理技术对病历质控和编码影响的探索

陈宏 许海峰 郝赫 孙红霞*

通辽市人民医院

DOI:10.12238/bmtr.v7i3.14446

[摘要] 病理报告包含大量非结构化文本(如形态学描述、诊断结论),传统人工病案质控及编码效率低且易遗漏,肿瘤形态学编码也容易张冠李戴。而自然语言处理技术通过语义识别、实体抽取和逻辑校验,可自动化完成病理分级及分期的提取,能够对良性、恶性或转移性质冲突进行逻辑化校验;与肿瘤形态学(WHO最新版本)编码进行标准化编码映射,提高病案质控效率及编码质量,为医院高质量发展、重点学科建设提供优质数据支撑,为加速临床研究发挥重要作用。

[关键词] 病理报告; 病案质控及编码; 自然语言处理技术

中图分类号: R36 文献标识码: A

Exploration of the Impact of Natural Language Processing Techniques on Medical Record Quality Control and Encoding in Pathological Reports

Hong Chen Haifeng Xu He Hao Hongxia Sun*

Department of Medical Records, Tongliao People's Hospital

[Abstract] Pathology reports contain a large amount of unstructured text, such as morphological descriptions and diagnostic conclusions. Traditional manual chart quality control and coding are inefficient prone to omissions. Tumor morphology coding is also prone to confusion. Natural language processing technology can automatically complete the extraction of pathological grading and staging through semantic recognition, entity extraction logic verification, and can logically verify benign, malignant or metastatic conflicts. It can also perform standardized coding mapping with tumor morphology coding (latest WHO version). It is correctly in the DRG/DIP disease group, which improves the efficiency of chart quality control and coding, and provides high-quality data support for the high-quality development of hospitals the construction of key disciplines, and plays an important role in accelerating clinical research.

[Key words] pathology report; case quality control and coding; natural language processing technology

国家卫生健康委办公厅于2021年开始出台“国家医疗质量安全改进目标”,并配套“国家医疗质量安全改进目标说明”,其中提高肿瘤治疗前临床TNM分期评估率,至2025年已经连续5年都被列入十大改进目标内容之一,目的是为充分发挥引导工作方向、激发行业内生动力、实现医疗质量安全改进的积极作用,提升医疗质量安全管理水平^[1]。

随着无纸化病历的实施,病历质控人员及编码员工作由人工翻阅病案拓展到在无纸化电子病历中浏览,电子病历左侧设置标题栏实现快速精准阅读,质控耗时从40分钟/份(包括回收、整理时间)降至8分钟/份;病理报告中设置完整性校验,强制字段缺失提醒(如未标注切缘状态);一致性校验:免疫组化结果与病理形态学编码矛盾预警(如CD117+、CD34+、DOG1+但未标注“恶性胃肠道间质瘤”或“胃肠道间质肉瘤”,只单纯标注为“胃肠道间质瘤”);病理报告中术语更加标准

化,与WHO分类更新同步,并将更新内容进行展示,提醒临床医生、质控员、编码员变化内容,提高形态学及疾病编码正确率;通过分析NLP技术在病理报告及病历首页附页结构化处理、关键信息提取和自动编码方面的表现,达到医疗质量安全目标中肿瘤病历质控目标及保证编码的准确性。而自然语言处理(natural language processing, NLP)是人工智能领域中的一个重要研究方向,主要研究实现人与计算机之间用自然语言进行有效通信的各种理论和方法^[2]。某三甲医院病案科、信息科、病理科联合运用NLP技术完成质控及编码工作。

1 肿瘤病历质控及编码方式的现状分析

现阶段,随着科学技术的发展,信息技术逐渐在社会全面领域被应用,其中包括医疗服务行业,并且新型的医疗模式逐渐向信息化靠拢,而其中病历电子化便是重要的改革产物,其能够有效替代以往的纸质版病历,不仅更加有利于查询、管理,

同时保存时间更久^[3]。而将电子病历进行结构化后可以更加有利于管理。

五线城市的某三级甲等医院在推行无纸化病历之前，质控员及编码员查看文档文书及回收的纸质病历中的病理报告，归档后病理结果仍然未出具的结合PACS系统，病案管理系统将医学文本转化成信息化服务，且处于基础工具提供阶段。仅能完成简单的形式质控，病历终末质控几乎全靠专业质控人员抽样检查后反馈科室进行整改，存在病历质点覆盖率低、效率低下、缺乏医疗全过程质控等问题^[4]。

2018年，国家卫健委发布《关于进一步推进以电子病历为核心的医疗机构信息化建设的通知》，指出要重点发挥电子病历信息化作用，促进医疗管理水平提升和智慧医院发展，对电子病历数据质控提出了更高的要求。2019年国务院办公厅发布的《关于加强三级公立医院绩效考核工作的意见》，将住院病案首页数据和电子病历应用水平纳入三级公立医院绩效考核，要求医院加强住院病案首页质量管理，提升首页数据质量。国家关于公立医院改革及高质量发展示范项目亦明确指出，电子病历高质量发展是公立医院高质量发展的重要环节之一，而电子病历的高质量发展要依托于高质量的病历环节与终末内涵质控^[5]。以电子病历为载体的医学文本信息系统重要性不断攀升，临床医护人员及病案、信息、医务等相关职能部门工作要求加码，数据上传时限有具体要求，无纸化病历的推行加速发展，2年时间实现了无纸化病历全院所有科室全面推行，病历质控及编码方式转变为电子形式。

随着国家信息产业的推动，人工智能(Artificial Intelligence, AI)与大数据分析应用加速落地，AI在医疗行业中针对流程优化、预测性分析和预测性决策方面有了诸多应用，尤其是2025年1月15日DeepSeek的正式上线“带飞”医疗行业，铺天盖地的多家医院接入DeepSeek，加速推动了医疗智能化，但是对于五线城市医疗机构过多投入信息化建设对DIP支付费用造成成本加成，故对肿瘤病历专项质控及编码质量，病理报告应用传统自然语言处理技术仍然是医疗资源节省的措施之一。

2 肿瘤病历质控及编码的意义

2.1 肿瘤分类的历程

起始于19世纪60-70年代，世界卫生组织(WHO)为了便于肿瘤临床诊疗和临床研究，尽其所能地汇编了人类迄今为止遇到的所有肿瘤类型。主流的肿瘤编码规则分部位编码和形态学编码两种，WHO沿用了形态学编码，来源于国际疾病分类(ICD-0)。作这样选择概因肿瘤是机体内细胞在内部和外界致癌因素长期作用下诱发，并且演化发展的生长物，形态学编码不仅区分了肿瘤的大致发病部位，还能够准确锁定其发展阶段，为临床诊疗和研究提供了便利。WHO的肿瘤分类目录与时俱进，各个疾病系统历经第一版到第五版的分类变化。

2.2 肿瘤病历的质控

恶性肿瘤在我国位于居民死因排序首位。全面科学评估肿瘤患者病情，是肿瘤规范化治疗的基础。提高肿瘤患者治疗前完

成临床TNM分期评估的比例可以提高肿瘤患者诊疗方案的科学性、合理性，提升肿瘤患者诊疗效果和生存率。病历质控员依据肿瘤病历的质控标准，极大地推动肿瘤临床诊疗工作的规范化、科学化进程，促进癌症的预防与控制工作，提高恶性肿瘤的诊治水平。

2.3 肿瘤病历的诊断编码

组织学编码是肿瘤编码的基础，前者失之毫厘，后者谬以千里，国际疾病分类对编码员的综合知识要求很全，不仅要掌握解剖学、诊断学、病理学等基础医学知识，并且还要掌握医学的前沿发展动态，如：疾病发生机理(诊断编码)、新技术新业务(手术编码)、病理学新的分期分型(病理编码)等。尤其是肿瘤编码，专业性强，难度大，编码时，应认真查阅病案，对患者的诊疗情况进行综合分析总结，同时与ICD-10和ICD-0编码原则有机结合，切勿生搬硬套，机械编码，编码的过程就是不断学习与积累临床经验的过程，如此，才能掌握疾病诊断与病理诊断的实际内涵，保证新发肿瘤编码的准确性^[6]。

3 实际应用探索

(1)在病案首页中增加附页，首次治疗前临床分期评估、首次非手术治疗前病理学诊断、首次靶向/免疫治疗前分子病理检测、手术患者术中淋巴结清扫范围均设为选项填写。同时，采用数据治理和自动化清洗技术，确保数据的完整性和一致性相比之前浏览整份病历查找分期及评估内容节省平均32分钟/份；

医院信息科工程师在医疗辅助决策信息系统中增加统计报表，质控员登录系统对比报表及电子病历内容质控，信息系统利用自然语言处理技术^[7]，对电子病历、检查报告、病理报告等非结构化数据进行自动化处理，转化为结构化信息，依据高频词生成问题提示列表^[8]在辅助决策系统展示。

病历质控员每天在辅助决策系统根据统计报表，汇总出当天出院病历的肿瘤病历，耗时30秒，登录电子病历系统，系统左侧设置标题栏进行快速精准阅读，对肿瘤治疗前临床TNM分期评估中缺陷内容在质控系统中反馈给临床，耗时1分钟，按照本院病历质控标准，运用电子形式对病历其他内容进行质控及反馈，耗时5分钟，质控耗时从40分钟/份(包括回收、整理时间)降至8分钟/份。

(2)病理报告中设置完整性校验，强制字段缺失提醒(如未标注左右侧，切缘状态等)；一致性校验：免疫组化结果与病理形态学编码矛盾预警(如CD117+、CD34+、DOG1+未标注“恶性胃肠道间质瘤”或“胃肠道间质肉瘤”，只单纯标注为“胃肠道间质瘤”)，病理报告结构化内容包含肿瘤来源、发生部位、具体位置、组织学类型、肿瘤大小等^[9]。

基于抽取式及生成式自动摘要相结合的方式，从病理报告中部位及组织学分类、形态学分类等内容，同时从非结构化诊断描述中抽取关键句子，如免疫组化阳性内容，将病理科人员提供WHO最新版本的病理组织学分类融入生成模型，为肿瘤诊断编码提供关键信息，通过抽取与生成结合、专注临床术语，生成的摘要更贴近实际需求，提高了编码的质量。

4 总结

电子病历是与临床工作关系最为紧密的信息化产物。临床病历具有文书数量多、文本长、非结构化、语义表达主观等特征,使得电子病历质量难以实现全面、深度数字化的内涵管理。医院病案科、病理科与信息科合作,在目前医院信息系统基础上,为节省成本,在不投入资源的情况下,结合deepseek理念,创建病理知识库,应用传统自然语言处理技术完成肿瘤病历的质控及编码,实现简便,计算资源需求低;保持医学文本原意和语境,语法忠实度高;快速提取关键信息,提升质控及编码效率,实现医疗服务质量整体提升。

自然语言处理技术(natural language processing, NLP)通过使用定义的语言规则和相关领域知识,从自然语言自由文本中执行句法处理、提取信息、捕获含义、检测关系任务,可帮助病历质控员及编码员进行辅助决策支持^[10]、临床记录信息提取^[11]和疾病风险预测^[12]

NLP技术在病理报告无纸化管理和肿瘤病历质控中表现出显著效果,能够有效提高数据处理效率和编码准确性,为医疗质量管理和大数据应用提供可靠技术支持。未来应进一步加强多模态数据融合和深度学习模型优化,提升系统智能化水平。

[基金项目]

2024年内蒙古医学科学院公立医院科研联合基金科技一般项目(2024GLLH0909)。

[参考文献]

[1] 卫生健康委办公厅国家卫生健康委办公厅关于印发2021年国家医疗质量安全改进目标的通知[Z]. 国卫办医函(2021)76号

[2] 刘丽莉,李明,罗晓兰,等.基于自然语言处理智能技术的中医术语研究文献计量分析[J].上海中医药杂志,2024,58(7):1-6,14.

[3] 李泓璇,孟伊霏,刘红梅,等.基于HBM理论的慢性肾脏病3~5期病人个性化饮食管理方案的构建[J].循证护理,2023,9(14):2549-2554.

[4] 刘祉呈,陈爱芬,黄颂列.基于AI技术的电子病历质控平

台构建与实践[J].医院管理论坛,2021,38(8):40,83-84.

[5] 曹凯迪,荆莹,郭建军.基于AI的电子病历全流程质控管理平台的构建与应用[J].江苏卫生事业管理,2025,36(3):419-426.

[6] 张萍,张渝,陈学涛.新发肿瘤疾病分类编码质量分析[J].中华肺部疾病杂志(电子版),2020,(3):434-436.

[7] 左丹妮,刘欣华,归纯漪.基于自然语言处理技术的五官科患者在线咨询问题提示列表构建与科普AI助手开发[C].2024HINC第六届上海国际护理大会.

[8] 杨丽冰,郭超,姜会珍.李单青人工智能辅助肺癌数据库构建[J].中国胸心血管外科临床杂志2025,32(2):167-174.

[9] 王坚.胃肠间质瘤的规范性病理诊断[J].临床肿瘤学论坛,2017,(3):17-20.

[10] PETERSON B,HERNANDEZ E J,HOBBS C,et al.Automated prioritization of sick newborns for whole genome sequencing using clinical natural language processing and machine learning[J].Genome Med,2023,15(1):18.

[11] FU S,THORSTEINSDOTTIR B,ZHANG X,et al.A hybrid model to identify fall occurrence from electronic health records[J].Int J Med Inform,2022,162.

[12] 雪峰豪,蒋海波,唐聃.深度学习在健康医疗中的应用研究综述[J].计算机科学,2023,50(4):1-15.

作者简介:

陈宏(1970--),女,汉族,内蒙古通辽人,本科,职称:主任护师,单位:通辽市人民医院,主要研究方向:病案质控。

许海峰(1988--),男,汉族,内蒙古通辽人,本科,职称:中级,单位:通辽市人民医院,主要研究方向:医学信息化。

郝赫(1988--),男,汉族,内蒙古通辽市人,本科,副高,通辽市人民医院,主要研究方向:医学信息化。

*通讯作者:

孙红霞(1973--),女,蒙古族,内蒙古通辽人,硕士研究生,职称:主任医师/主任技师,单位:通辽市人民医院,主要研究方向:病案信息学-国际疾病分类及病历质控。