基于 BERT 模型的医疗安全事件智能分类研究与实践*

赵从朴 袁 达 朱溥珏 周 炯 陈 政 彭 华

（中国医学科学院北京协和医院 北京 100730）

〔摘要〕 **目的/意义** 改进医疗安全事件分类评估模式，提升工作效率和时效性。**方法/过程** 选取既往医疗安全事件数据进行预处理，利用 BERT 模型进行训练、测试、迭代优化，构建医疗安全事件智能分类预测模型。**结果/结论** 利用该模型对2022 年 1—11 月临床科室上报的466 例医疗安全事件进行分类，*F*1 值达 0.66。将 BERT 模型应用于医疗安全事件分类评估辅助，可提升工作效率和时效性，有助于及时干预医疗安全风险隐患。

〔关键词〕 医疗安全事件；BERT；深度学习；智能分类

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2024.01.005

Study and Practice on Intelligent Classification of Medical Safety Incidents Based on BERT Model

ZHAO Congpu, YUAN Da, ZHU Pujue, ZHOU Jiong, CHEN Zheng, PENG Hua

Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

〔**Abstract**〕 **Purpose/Significance** To improve the classification and evaluation mode of medical safety incidents, and to improve work efficiency and timeliness. **Method/Process** The data of previous medical safety incidents are pre－processed, BERT model is used for training, testing and iterative optimization, and an intelligent classification and prediction model for medical safety incidents is built. **Result/Conclusion** The model is used to classify 466 medical safety incidents reported by clinical departments from January to November 2022, and *F*1 value reaches 0.66. The application of BERT model in the classification and evaluation of medical safety incidents can improve work efficiency and timeliness, and help timely intervene in medical safety risks.

〔**Keywords**〕 medical safety incidents; BERT; deep learning; intelligent classification

1 引言

世界卫生组织将医疗安全事件定义为：并非由疾病并发症所致，而是由医疗管理有关行为造成的伤害。美国对医疗安全事件的定义是：由医疗导致的伤害，与疾病的自然转归相反，延长了患者住院时间，导致患者残疾或二者皆有[1]。在国内，医疗安全事件是指在临床诊疗活动和医疗机构运行过程中，任何可能影响患者诊疗结果、增加患者痛苦和负担并可能引发医疗纠纷或医疗事故，以及影响医疗工作正常运行和医务人员人身安全的因素和事件[2]。医疗安全事件分类评估是建立安全预警体系、营造安全文化的重要手段。

关于医疗不良事件，国内外并没有统一的分类

〔修回日期〕 2023－10－27

〔作者简介〕 赵从朴，中级职称，发表论文 13 篇；通信作者：彭华，研究员。

〔基金项目〕 北京协和医学院中央高校基本科研业务费项目（项目编号：3332022087）。

或分级标准。美国退役军人医疗机构采用严重程度评估得分矩阵系统，将医疗不良事件分为 4 级[3]。中国医院协会根据严重程度将医疗不良事件分为 4 级[4]。上海交通大学医学院附属第九人民医院魏斌等[5]基于“有无过错事实、是否造成后果”提出 SH9 分类法，也将医疗不良事件划分为 4 级。

从中国医学科学院北京协和医院医务处安全办的工作经验来看，SH9 分类法更具备可解释性和可操作性。因此，本研究基于 SH9 分类法进行。结合实际工作情况，将医疗安全事件分为两类：医疗不良事件、患源性安全隐患事件。其中，医疗不良事件根据严重程度分为 4 个等级，患源性安全隐患事件是指“拒绝做必要检查化验”“高龄患者无家属陪护”“有暴力倾向或心理障碍”等存在安全隐患的事件。由于当前医疗安全事件分类评估工作主要依靠人工逐条标注完成，存在工作效率较低、时效性较差等问题，影响干预医疗安全风险隐患并采取措施的及时性。

随着人工智能技术的不断发展和普及，在医疗安全事件分类评估领域，基于深度学习的文本分类算法被广泛应用。双向编码器表征（bidirectional encoder representations from transformers，BERT）模型是谷歌团队于 2018 年提出的一种语言表示模型，能够获取结合上下文语境的动态词向量，在自然语言处理领域取得了良好效果[6]。到目前为止，BERT 是文本分类任务中最准确的语言模型之一[7]，在多类别大规模训练集下更能体现其分类的优越性[8]。

2 国内外研究现状

积极应对医疗安全（不良）事件是医院提高自我纠错能力、持续改进医院医疗质量管理的重要途径，是纠纷隐患提前介入干预的关键环节[9]。在美国，几乎所有医院都有医疗差错和不良事件的内部报告系统，同时还有大量医院加入了外部报告系统[10]。中国医疗安全（不良）事件报告系统起步较晚。Yao L 等[11]利用 BERT 模型和领域特定语料库对中医临床记录进行分类研究；郑承宇等[12]提出一种用于多标签医疗文本分类的深层神经网络模型，整体 *F*1 值达到 90.5%；Zhang X 等[13]利用 BERT 模型从乳腺癌医疗文书抽取实体概念及其属性。

除了类 BERT 模型，医疗安全事件文本分类还可以使用基于统计的模型方法实现。罗玮[14]对收集到的 7 009 条患者投诉文本进行人工标注，运用多种重采样方法平衡患者投诉文本数据，实现了从患者投诉文本中自动识别安全事件相关要素的模型；唐仕肖等[15]采用德尔菲法和名义群体技术，研究分析医院护理不良事件管理的应用效果，发现可明显降低不良事件发生率，同时提高护理质量和患者护理满意度；朱未等[16]应用 Reason 模型，从组织影响、不安全监督、不安全行为的前提和不安全行为 4 个方面分析医疗器械不良事件的影响因素。然而，这些基于文本统计规律进行分类的方法由于未结合大规模语言模型的语言知识上下文信息，普遍难以超越类 BERT 模型的方法，因此本研究未对基于统计的分类方法进行实验对比。

3 技术方案

3.1 数据现状和预处理

医疗不良事件分类任务具有 3 个特点：事件情况多样、样本数据有限、数据中类别不均衡现象较严重。中国医学科学院北京协和医院 2017—2021 年积累的医疗不良事件数据总量有 4 767 条，但是早期的探索实验发现其中超过 75% 的数据都是患源性安全隐患事件。由于医院 2022 年才开始以 SH9 分类法指导分类，此前数据分类标准不统一，原有的分类标签失去价值不能直接使用。因此在本研究中，将 2022 年之前的数据作为无标签数据使用。2022 年收集的数据中有 1 059 条以 SH9 分类法进行了人工标注，作为高质量的人工标注数据，是本研究的基础。

从业务中收集的医疗文本中存在医生和患者的个人信息，直接使用可能会造成医患个人隐私泄漏，且姓名、地址等信息对任务本身无意义，但可能对模型的分类造成错误引导，因此需要进行数据预处理，从而保护医患个人隐私、提升分析效率。本研究通过规则方法结合实体抽取工具，将患者与医务人员的个人信息进行识别和省略。

3.2 模型结构设计

选择 BERT 模型作为医疗安全事件智能分类的主

要训练方法，主要是因为其可以基于大规模无标注的文本数据进行预训练，通过微调来适应不同的任务和领域，还可以同时进行文本分类、命名实体识别、关系抽取等多个任务。相比之下，其他模型要从头开始训练，需要更多的数据和计算资源。而且 BERT 模型在处理文本时能够很好地捕捉上下文信息，在处理长文本方面具有优势。在医疗安全事件文本分类中，往往需要对整个文本进行分析，BERT 模型能够很好地满足这一需求。基于 BERT 通用领域预训练模型，在大规模医疗语料上做增量迁移训练，得到医疗领域的预训练模型，然后采用微调的方式，得到分类模型，进而完成医疗安全事件分类。

3.3 模型训练过程

BERT 模型用于文本分类主要包括两个步骤，见图 1。第 1 步：无监督自学习。在已有大规模语料训练得到的 BERT 模型基础上，使用医院多年积累的大规模医疗数据对其进行领域内预训练，训练数据包括医学百科、医学文献、临床指南和电子病历等高质量的文本。

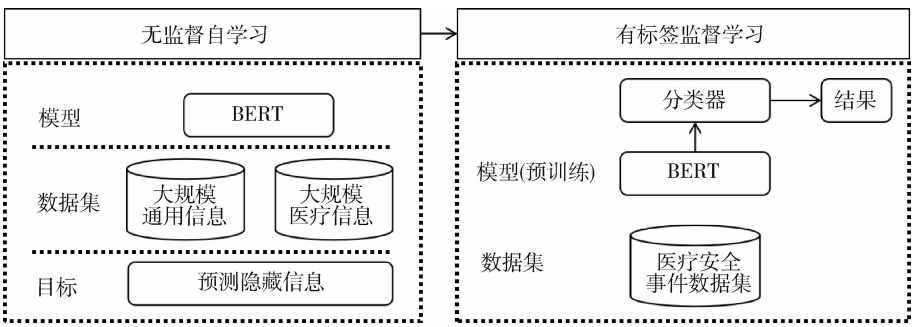

图 1 BERT 模型训练过程

第 2 步：有标签监督学习。根据事件分类任务对模型进行微调，在有标签的数据集中进行有监督训练，输出文本特征向量。通过微调过程改变模型中的权重，使模型在该数据集上学得更好，提高分类效果。在分类任务输入中，在文本序列首位加入一个特殊标记［cls］作为分类表征。将预处理后的安全事件文本输入该模型，文本经过词嵌入层将文本中的字向量化，其中包含字的嵌入向量、句子的嵌入向量以及位置的嵌入向量。例如“者”为句子第 2 个字的输入表征，其由字嵌入 $E_{者}$、句子嵌入 E_A 和位置嵌入 E_2 相加得到，见图 2。

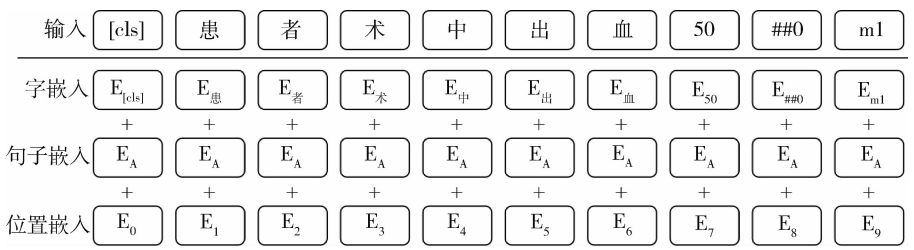

图 2 输入向量示意

BERT 模型将输入序列的嵌入向量输入到一个全连接层中，然后将输出向量传递给 softmax 分类器，经过 softmax 函数计算得到分类标签（患源性安全隐患事件、Ⅰ级事件、Ⅱ级事件、Ⅲ级事件、Ⅳ级事件）的概率，选取概率最大值所对应的标签即为模型分类结果。

3.4 数据扩充

3.4.1 步骤 为了缓解数据类别间数量不均衡的问题，需要利用早于 2022 年的无标签数据进行定向数据扩充。第 1 步：为了增加医疗不良事件数据、减少患源性安全隐患事件数据占比，使用 2022

年带有专家标注标签的 1 059 条数据，分为“患源性/有分级”两个标签，训练一个简单的 BERT 模型二分类器，对早于 2022 年的 4 767 条数据进行分类，挑选出可能是以分级数据作为标注扩充的备用数据。第 2 步：对挑选出的待标注数据进行人工标注，核验后充实到有标签数据中。第 1 步的二分类模型分类准确率约为 80%。使用该模型对无标签数据进行预测，挑选出 1 013 条待标注数据。

3.4.2 标注方案设计及试标注 为了简化标注工作、提高标注效率，设计拆分标注方案，即根据 SH9 分类法，将对医疗不良事件的分级判断转化为对事件“有过错/无过错/患源性安全隐患”和“造成结果/未造成结果/患源性安全隐患”两个判断，从而帮助标注人员更直接地对事件进行判断，提高了标注准确率，降低了标注难度。

试标注完成的 100 条数据反馈到医院医务处安全办进行结果审查。两次分别标注后组合产生的分级标签准确率为 70%，较为可用；且在试标注过程中，总结和建立了标注规范，针对专家审查过程中提出的普遍性、代表性问题，对标注人员进行专门指导和培训。经过医务处安全办审查确认，拆分标注方案合理可行，标注结果可用，见表 1。从抽样的 100 条数据可以看出，患源性安全隐患事件占比仅为 14%，说明前期的二分类筛选方法有效。

表 1 试标注审查结果

标签	试标注正确样本数（条）	试标注总样本数（条）	准确率（%）
患源性安全隐患事件	4	14	28.6
Ⅰ级事件	1	1	100
Ⅱ级事件	39	50	78
Ⅲ级事件	0	1	0
Ⅳ级事件	26	34	76.5
总计	70	100	70

3.4.3 正式标注及标注结果 基于试标注建立的标注规范，标注人员对 1 013 条数据中剩余部分进行正式标注。得到各标签结果为：“有后果”422 条、“无后果”408 条和患源性安全隐患事件 183 条；“有过错”197 条、“无过错”636 条和患源性安全隐患事件 180 条。

3.5 模型迭代优化和模型部署

在实际使用过程中，基于人工标注的少量数据，通过本研究所提出方法进行模型训练，利用得到的模型对临床科室新上报的医疗安全事件数据进行分类预测。医务处安全办工作人员对模型分类结果进行复核，相比人工分类评估，节省出来的时间可以更加专注于机器模型覆盖范围外的文本标注工作，以便进一步迭代优化。

经过多轮迭代优化后，可以得到性能可靠的医疗安全事件智能分类预测模型。该模型可通过部署深度学习模型服务方式进行调用，实现事件分类过程自动化。

4 实验及结果分析

4.1 实验数据集和实验环境

基于 2022 年全年有标签数据和扩充得到的数据，总实验数据量为 2 072 条，将其中 1 606 条划分为训练集、466 条划分为测试集进行实验。实验在使用 GPU 的硬件环境下完成，显卡型号为 NVIDIA V100，显存为 24GB，软件环境为 PyTorch1.5。

4.2 评估指标

采用准确率（accuracy）、精准率（precision）、召回率（recall）和综合评价指标 *F*1 值作为评估指标。准确率是所有样本中预测正确的样本占比，表示预测正确的样本概率。因此，准确率存在一个明显的问题，即在数据类别不均衡，特别是有极偏数据存在的情况下（如当 100 条测试样本中有 99 例患源性安全隐患事件，只有 1 例医疗不良事件时），无法客观评价算法的优劣。精准率又称查准率，是针对预测结果而言的，表示在所有被预测为正的样本中实际为正的样本占比。召回率又称查全率，是针对原样本而言的，表示在实际为正的样本中被预测为正的样本占比。精准率和召回率是一对此消彼长的度量，精准率高，会漏掉很多数据；召回率高，精准率又会降低。因此，为了兼顾二者，使用二者的调和平均数 *F*1 值。

4.3 实验方案设计

4.3.1 直接分类实验 即直接使用Ⅰ—Ⅳ级事件标签和患源性安全隐患事件标签，共 5 个分类标签，训练分类模型，并对测试集进行预测。

4.3.2 两次分类实验 根据 SH9 分类法的特点，将任务分解为对医疗不良事件进行“有过错/无过错”和“造成结果/未造成结果”两次分类，并将分类结果按照 SH9 分类法组合为Ⅰ—Ⅳ级事件。其余为患源性安全隐患标签。组合的规则，见表 2。

表 2 两次分类组合规则

标签	有过错	无过错
造成后果	Ⅰ级事件	Ⅱ级事件
未造成后果	Ⅲ级事件	Ⅳ级事件

4.4 实验结果

考虑到不同医疗安全事件类别中的样本数据分布情况，统计两种实验方法中精准率、召回率和 *F*1 值的平均值，发现使用两次分类法较直接分类法精准率、召回率和 *F*1 值均高出 2 个百分点，见图 3。

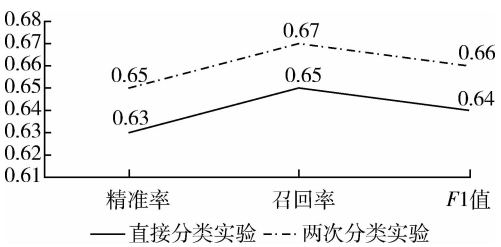

图 3 两种实验方法结果对比

4.4.1 结果分析 两次分类法针对具体场景下的任务级优化能够提升效果，说明根据医疗安全事件进行“有无过错”和“是否造成后果”的任务拆解能够提升模型分类质量。由于用于实验的有标签数据总量不大，而医疗不良事件可能涵盖的场景和事件类型等差异很大、可能性很多，类似事件少，模型难以从这些数据中学习到足够强的特征来执行分类，因此，两次分类法本质上是基于人类对该任务的理解做出的一种类似于特征工程的处理，使模型可以更有针对性地提取有利于判断分类的信息，从而获得较直接分类法更好的结果。

从各标签的数量分布和评估指标来看，仍然存在问题。一是受制于不同标签间数量的不平衡，样本数量偏少的Ⅰ级事件和Ⅲ级事件的指标表现不甚理想。通过数据定向扩充标注，虽然减少了患源性安全隐患事件在总体数据中的数量占比，但是依然难以解决医疗不良事件之间不同等级的数量不平衡问题。二是在数量比例接近的Ⅱ级事件和Ⅳ级事件之间，存在较明显的性能差异。原始数据中Ⅱ级事件的特征比较明显，一般来说会与一组后果事件存在较强关联，模型可以较容易地从训练集中学到这种关联；而Ⅳ级事件特征不明显，发生情景、事件本身内容、关联医疗资源、医疗操作等因素都有较强多样性，从而在训练数据较少的情况下难以充分训练模型，性能表现难以提升。

4.4.2 示例分析 针对两种不同的分类方法，随机选取两组示例进行分析，见表 3。

表 3 两种分类方法对比示例

医疗安全事件描述文本	直接分类	二次分类	标准答案
［示例 1］患者对治疗方式可能存在意见。患者因产后 7 周，发现宫腔残留 1 天收入院。2022 年 6 月 30 日 01：47 顺娩一活女婴，身长 51cm，体重 4 010g，术中肩难产及胎盘胎膜粘连。产后 6 周复查超声提示宫腔内见中高回声，范围约 3.9cm×2.2cm×1.4cm。无阴道出血。2022 年 8 月 26 日拟行宫腔镜检查，术中发现子宫穿孔，手术方式改变为腹腔镜检查。	患源性安全隐患	Ⅱ级事件	Ⅱ级事件
［示例 2］患者外院行手术治疗，术中大出血，盆腔纱布填塞，2022 年 11 月 12 日转我院，病情危重，医疗风险极高，患者家属期望高，存在医疗隐患。患者老年女性，2022 年 11 月 9 日因骶前肿物在外院行腹腔镜转开腹肿物探查、腹膜后肿物切除术、右附件切除术、盆腔纱布填塞术，术中见骶骨右侧一直径约 7cm 实性肿物，切除肿物过程中骶前血管出血，缝合困难，予盆腔内填塞纱布 16 块压迫止血，术中大出血，出血量约 7 350ml，予输血支持。2022 年 11 月 10 日行介入栓塞止血（骶正中动脉、右侧髂内动脉脏支主干）。	Ⅱ级事件	患源性安全隐患	患源性安全隐患

在示例1 中，使用了二次分类法将医疗事件判定为造成后果和无过错，因此得到更加准确的分类。在示例 2 中，由于受到上下文相关信息的干扰，采用直接分类方法出现分类偏差。两条示例对比表明二次分类方法能结合多维度能力判断相应类别，获得更接近事实的结果，为后续进一步优化模型提供了更好的思路。

4.4.3 实验结论 在当前数据状况和方案设计下，本研究提出的方法可以实现有价值的分类性能。同时，从医院医务处安全办实际工作来看，在相同数据量条件下，与人工逐条识别标注相比，利用上述模型辅助完成分类工作，工作时间可由“小时”级缩短至“分钟”级，在一定程度上提高了医疗安全事件分类评估工作的效率和时效性，可以更早地识别并提前介入医疗安全隐患。本研究仍然存在一定的局限性。（1）可利用的有标签数据量较有限。本研究的标注工作已经用尽 2017 年至今积累的可用数据，然而从模型实际表现来看，训练依旧不充分，模型并未学到和记忆足够多的知识以更好地解决问题。（2）数据存在严重的类别数量不均衡现象。在收集到的数据分布中，患源性安全隐患事件样本占比远大于有分级样本，同时有分级样本内部各不同等级事件数量差异较大，导致模型在不同标签上分类性能差距较大且难以平衡。上述不足可通过扩充数据获取渠道、增加标注数据量等方式进一步优化。

5 结语

本研究表明基于深度学习技术，通过 BERT 模型对临床科室上报的医疗安全事件进行智能分类，呈现出较强的管理辅助决策作用，可以帮助医务管理部门提高工作效率和时效性，有助于识别并及时干预医疗安全风险隐患。

随着类 ChatGPT 大模型技术的兴起，预训练语言模型得到进一步升级。未来可以尝试结合大模型技术，通过小样本、零样本方法利用大模型辅助人类标注，高效扩充标注数据。可利用大模型对事件进行分析，提供分类依据，辅助人类的分类决策。

利用人工智能技术赋能医疗质量安全管理是大势所趋。在本研究内容的基础上，中国医学科学院北京协和医院将进一步搭建实时在线的多院区医疗风险管理人工智能数字化平台，提升医政医管的数字化、精细化水平和能力，加强医疗质量管理，保障患者安全。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 BRENNAN T A, SOX C M, BURSTIN H R. Relation between negligent adverse events and the outcomes of medical-malpractice litigation [J]. The New England journal of medicine, 1996, 335 (26): 1963-1967.

2 沈刚，乔学斌，戴月华，等. 某三甲医院 848 例医疗不良事件的分析及对策 [J]. 江苏卫生事业管理，2021，32 (7)：885-888.

3 马旭东. 我国医疗质量安全不良事件分类的思考 [J]. 中国卫生质量管理，2021，28 (6)：46-50.

4 朱晓萍，田梅梅，施雁. 国内外医疗不良事件分类体系的研究现状 [J]. 护理研究，2013，27 (14)：1281-1284.

5 魏斌，田卓平. 医疗不良事件 SH9 分类法及其现实意义 [J]. 中国医院，2011，15 (1)：44-45.

6 焦玮，杨雪寒，孟洁，等. 针对电子医疗档案的数据分析 [J]. 微型电脑应用，2020，36 (9)：32-35.

7 DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]. Minneapolis: The 2019 Conference of the North, 2019.

8 赵旸，张智雄，刘欢，等. 基于 BERT 模型的中文医学文献分类研究 [J]. 数据分析与知识发现，2020，4 (8)：41-49.

9 白琼，李骥，李宗师. 美国医院患者安全监督及管理策略 [J]. 解放军医院管理杂志，2020，27 (6)：598-600.

10 田梅梅，施雁. 对医疗不良事件报告系统设计的思考 [J]. 中国护理管理，2011，11 (5)：3.

11 YAO L, JIN Z, MAO C, et al. Traditional Chinese medicine clinical records classification with BERT and domain specific corpora [EB/OL]. [2023-09-27]. http://dx.doi.org/10.1093/jamia/ocz164.

（下转第 38 页）

2 KRAUSE J, JOHNSON J, KRISHNA R, et al. A hierarchical approach for generating descriptive image paragraphs [C]. Honolulu: 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.

3 WANG X, PENG Y, LU L, et al. TieNet: text - image embedding network for common thorax disease classification and reporting in chest X - Rays [C]. Salt Lake City: IEEE Conference on Computer Vision and Pattern Recognition, 2018.

4 JING B, XIE P, XING E. On the automatic generation of medical imaging reports [C]. Melbourne: 56th Annual Meeting of the Association for Computational Linguistics, 2018.

5 CHEN Z, SONG Y, CHANG T H, et al. Generating radiology reports via memory - driven transformer [C]. Online: Conference on Empirical Methods in Natural Language Processing, 2020.

6 WU X, LI J, WANG J, et al. Multimodal contrastive learning for radiology report generation [J]. Journal of ambient intelligence and humanized computing, 2022, 14 (8): 11185 - 11194.

7 HOU D, ZHAO Z, LIU Y, et al. Automatic report generation for chest X - Ray images via adversarial reinforcement learning [J]. IEEE access, 2021 (9): 21236 - 21250.

8 ZHANG Y, WANG X, XU Z, et al. When radiology report generation meets knowledge graph [C]. New York: 32nd Innovative Applications of Artificial Intelligence Conference, 2020.

9 WANG X, ZHANG Y, GUO Z, et al. TMRGM: a template - based multi - attention model for X - Ray imaging report generation [J]. Journal of artificial intelligence for medical sciences, 2021, 2 (1 - 2): 21 - 32.

10 HOU B, KAISSIS G, SUMMERS R, et al. RATCHET: medical transformer for chest X - Ray diagnosis and reporting [C]. Online: International Conference on Medical Image Computing and Computer Assisted Intervention, 2021.

11 ALFARGHALY O, KHALED R, ELKORANY A, et al. Automated radiology report generation using conditioned transformers [J]. Informatics in medicine unlocked, 2021 (24): 100557.

12 NGUYEN H, NIE D, BADAMDORJ T, et al. Automated generation of accurate & fluent medical X - Ray reports [C]. Punta Cana: Conference on Empirical Methods in Natural Language Processing, 2021.

13 RAJPURKAR P, IRVIN J, ZHU K, et al. CheXNet: radiologist - level pneumonia detection on chest X - Rays with deep learning [EB/OL]. [2023 - 09 - 12]. https://arxiv.org/abs/1711.05225.

14 MCDONALD R, BROKOS G I, ANDROUTSOPOULOS I. Deep relevance ranking using enhanced document - query interactions [EB/OL]. [2023 - 09 - 12]. https://arxiv.org/abs/1809.01682.

15 PAPINENI K, ROUKOS S, WARD T, et al. BLEU: a method for automatic evaluation of machine translation [C]. Philadelphia: 40th Annual Meeting of the Association for Computational Linguistics, 2002.

16 LIN C Y. Rouge: a package for automatic evaluation of summaries [C]. Barcelona: Text Summarization Branches Out, 2004.

17 BANERJEE S, LAVIE A. METEOR: an automatic metric for MT evaluation with improved correlation with human judgments [C]. Ann Arbor: ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, 2005.

18 VEDANTAM R, ZITNICK C L, PARIKH D. Cider: consensus - based image description evaluation [C]. Boston: IEEE Conference on Computer Vision and Pattern Recognition, 2015.

(上接第 32 页)

12 郑承宇, 王新, 王婷, 等. 基于 ALBERT - TextCNN 模型的多标签医疗文本分类方法 [J]. 山东大学学报 (理学版), 2022, 57 (4): 21 - 29.

13 ZHANG X, ZHANG Y, ZHANG Q, et al. Extracting comprehensive clinical information for breast cancer using deep learning methods [EB/OL]. [2023 - 09 - 27]. http://dx.doi.org/10.1016/j.ijmedinf.2019.103985.

14 罗玮. 患者投诉中安全事件的自动识别研究 [D]. 武汉: 华中科技大学, 2019.

15 唐仕肖, 李秀云, 李文娟. NGT 与德尔菲法应用于医院护理不良事件管理系统的分析与研究 [J]. 智慧健康, 2022, 8 (28): 189 - 193.

16 朱未, 胡少科. 基于 Reason 模型的医疗器械不良事件的影响因素研究 [J]. 生物骨科材料与临床研究, 2018, 15 (3): 77 - 80.