

· 智慧医疗 · doi:10.3969/j.issn.1671-8348.2024.17.021

网络首发 [https://link.cnki.net/urlid/50.1097.R.20240616.1730.008\(2024-06-16\)](https://link.cnki.net/urlid/50.1097.R.20240616.1730.008(2024-06-16))

基于大数据的医疗风险综合评估系统的设计*

蒋丽梅,刘 锋,杜 倩,戴黎阳,张 杨,严 敏[△]

[重庆医科大学附属第三医院(捷尔医院),重庆 401120]

[摘要] **目的** 构建基于大数据的医疗风险综合评估系统并进行一致性、效率评估。**方法** 针对住院患者风险评估现状,基于大数据的手段,应用医学自然语言处理设计医疗风险综合评估系统。系统可自动抓取患者的各项数据,通过数据挖掘和机器学习技术自动生成评分,并将风险数据发送给医护人员,从而实现医疗风险评估自动化、智能化。采用随机对照分析,分别对纳入的评分量表进行人工赋分和机器赋分,比较赋分用时,自动生成可视化风险矩阵图。**结果** 该系统纳入研究的评分系统 *Kappa* 值结果如下:Caprini 量表(外科)、Padua 量表(内科)*Kappa* 值为 1.00,NNIS *Kappa* 值为 1.00,Nomogram *Kappa* 值为 0.87,Morse 评估量表/Hendrich 模型 *Kappa* 值为 0.83,Braden *Kappa* 值为 0.80,ASA2023 *Kappa* 值为 1.00,NRS2002 *Kappa* 值为 0.90。机器赋分用时均短于人工赋分用时,差异有统计学意义($P < 0.05$)。**结论** 通过本系统构建的风险矩阵图可使评估效率和准确性大幅提升,不仅能提供精准的诊疗方案,还能缩短患者住院时间,降低医疗费用。

[关键词] 风险评估;系统设计;大数据;自然语言处理;评估自动化

[中图法分类号] R-058 **[文献标识码]** A **[文章编号]** 1671-8348(2024)17-2672-05

Design of medical risk comprehensive assessment system based on big data*

JIANG Limei, LIU Feng, DU Qian, DAI Liyang, ZHANG Yang, YAN Min[△]

[Third Affiliated Hospital (Jieer Hospital), Chongqing Medical University, Chongqing 401120, China]

[Abstract] **Objective** To construct the medical risk comprehensive assessment system based on big data, and to evaluate its consistency and efficiency. **Methods** Aiming at the current situation of risk assessment of inpatients, based on the means of big data, the medical natural language processing was used to design a medical risk comprehensive assessment system. The system can automatically capture various data of patients, automatically generate the scores by data mining and machine learning technology and send the risk data to medical staff, so as to realize the automation and intellectualization. The randomized controlled analysis was used to conduct the manual scoring and machine scoring for included the score scale. The visual risk matrix diagram was automatically generated by comparing the scoring. **Results** The *Kappa* values of the scoring system in the included study of the system were as follows: the *Kappa* value in Caprini scale (surgery) and Padua scale (internal medicine) was 1.00, NNIS *Kappa* value was 1.00, Nomogram *Kappa* value was 0.87, *Kappa* value in the Morse assessment scale/Hendrich model was 0.83, Braden *Kappa* value was 0.80, ASA 2023 *Kappa* was 1.00 and NRS 2002 *Kappa* value was 0.90. The taking time in the machine scoring all were shorter than those in the manual scoring, and the difference was statistically significant ($P < 0.05$). **Conclusion** The risk matrix graph constructed by this system could sharply increase the evaluation efficiency and accuracy, which not only provide the accuracy diagnosis and treatment regimen, but also shorten the hospitalization duration and reduce the medical costs.

[Key words] risk assessment; system design; big data; natural language processing; evaluation automation

医疗质量和患者安全是医疗服务的核心任务^[1],也是医院高质量发展的核心要求^[2],2022年“国家医疗质量安全改进目标”对各项疾病的预防率和评估率

做出了明确要求^[3-4]。通过对住院患者进行风险评估,能尽早识别疾病风险并进行干预,获得更好的医疗效果,这不仅有助于改善医疗服务,还可降低再入

院率和医疗成本^[4]。住院患者风险评估是改进医疗质量安全的重要举措^[5]，通过采用科学的程序和方法，基于患者自身情况，对重点环节、重点人群与高危人群进行监测，从而识别风险，评估患者未来发生某种特定疾病的可能性，以便在疾病和危害发生前提前实施干预^[6-7]。

随着疾病诊断相关分组 (diagnostic-related groups, DRGs) 的快速推进^[8]，医院由粗放管理转向精细化管理^[9]。为对患者的早期风险实施预警和有效预防，本院设计了综合评估系统，应用大数据、医学自然语言处理、人工智能等技术手段，结合临床工作实践，针对各项风险评估设计专用数字评估模型。在患者住院的整个诊疗过程中，通过医疗健康大数据平台自动提取各个评估模型所需数据，实时生成可视化风险矩阵图以展现评估预警结果。同时，本系统与护理、医疗信息系统联动，将评估结果信息自动返回业务系统，及时提醒医护人员进行针对性干预，将风险出现后的干预处理转变为风险出现前的积极预防。通过本系统，真正做到根据病情变化自动、实时评估，避免因主观手动评估带来的评估不及时、不准确、效率低下等问题，提升医疗风险防范能力，保障医疗安全，降低医护的工作量。

1 资料与方法

1.1 一般资料

筛选 2022—2023 年的 50~75 岁高危患者 51 例，包括非计划再次手术、住院时间超 30 d、瘫痪、超长时间手术、术中大量出血、糖尿病/高血压史、6 个月内肿瘤放疗史、6 个月内使用呼吸机、恶性肿瘤家族史、呼吸衰竭等患者，其中男 26 例，女 25 例。以医院现有的医疗信息系统、实验室信息管理系统、手术麻醉系统和护理信息系统等为基础抽取年龄、既往史、家族史、BMI、血压、症状与体征等指标数据，经过评估模型计算，得出评分用时。针对相同患者进行人工

评估并记录评分时长，两组数据采用随机对照的方法进行对比。

1.2 方法

1.2.1 选取风险评估量表

组建由医院管理、医疗、护理、信息统计中心组成的院内专家团队，通过查阅资料、调研当前临床风险评估工作的开展情况，将 7 种常见的重要风险种类纳入评估系统，包括静脉血栓栓塞症 (venous thromboembolism, VTE) 风险、手术风险、多重耐药菌感染风险、跌倒坠床风险、压疮风险、麻醉风险和营养风险。记录每一项评估开展时间、评估频次、评价关键指标、评判标准、评估后临床处理措施、评估人员等细节，收集评估量表、评价算法等资料，见表 1。

1.2.2 筛选风险评估因素形成综合评估量表

在评估量表的基础上，根据医院实际并结合临床调研、讨论，以及院内信息系统及医疗业务数据为基础，论证各项评估项目关键评价因素；通过大数据平台获取实时客观数据和计算评价结果的可行性，或通过信息化改造后满足数据提取及评估的可行性，最终筛选出 8 个一级指标、102 个二级指标作为系统研发的基础，制订风险评估表，形成评估模型。

1.2.3 设计综合评估系统架构

综合评估平台实时从医院信息系统、实验室信息管理系统、医学影像存档与通讯系统、手术麻醉系统等业务系统进行数据抽取后，对数据进行标准化清洗及转换，然后在自然语言处理^[10-11]、知识图谱^[12]、机器学习^[13]等技术的基础上，对数据进行治理，实现临床数据的标准化、归一化、结构化，将数据转化为计算机能识别的内容。根据患者的诊疗数据，触发相应的规则内容，计算患者各个风险对应的分值，对中高风险情况进行提示。医务人员收到提示后进行后续评估与处置，人为干预和处理的结果反馈和沉淀到数据中心，供系统改进，见图 1。

表 1 纳入风险评估模型量表

风险类型	评估开展时间	评估频次	评估量表
VTE 风险	入院时	动态评估	Caprini 量表(外科);Padua 量表(内科)
手术风险	术前	每次手术评估 1 次	NNIS
多重耐药菌感染风险	入院时	入院时/每 30 天评估 1 次	Nomogram
跌倒坠床风险	入院时	按风险级别	Morse 评估量表/Hendrich 模型
压疮风险	入院时	按风险级别	Braden 量表
麻醉风险	术前	每次手术评估 1 次	ASA2023
营养风险	入院时	1 次/周	NRS2002

NNIS:手术风险分级标准;Nomogram:列线图;ASA2023:2023 年版美国麻醉医师协会健康状况分级;NRS2002:2002 年版营养风险筛查。

1.2.4 数据准备与整合

整合纳入风险评估模型的相关医疗数据，包括患者基本信息、既往史、现病史、体格检查、治疗情况、检查报告结果、影像和病理检查报告结果、处方信息及治疗方案等。在数据整合中，通过以下方法对不同数

据进行归一化处理，使数据可被系统所接受。(1)对于结构化数据，从记录时间、修改权限等方面设置质量控制条件，保障数据及时、准确。(2)利用自然语言处理技术对非结构化的电子病历进行信息提取，转化为规定格式的结构化数据信息。在此过程中，开发非

结构化-结构化数据字典库,完成标准术语集制订。(3)数据清洗,通过补齐、去重和降维等方式将错误或杂乱无章的数据处理成干净、标准的数据,采用回归、贝叶斯、决策树等机器学习方法来确定填充数据的最佳值以供后续数据统计和挖掘使用。(4)数据转化,当多个系统在命名、编码、度量单位等方面存在不一致时,根据不同数据特征将其转化为可用于数据挖掘

等后续处理的形式。例如,日期格式在医院信息系统为“YYYY-MM-DD”,但在实验室信息管理系统中为“YYYY/MM/DD”,可统一将格式转化为“YYYY-MM-DD”;“性别”编码在数据库中可能为“Male/Female”“M/F”“0/1”,当数据被抽取后,可以统一将其转换为“0/1”。

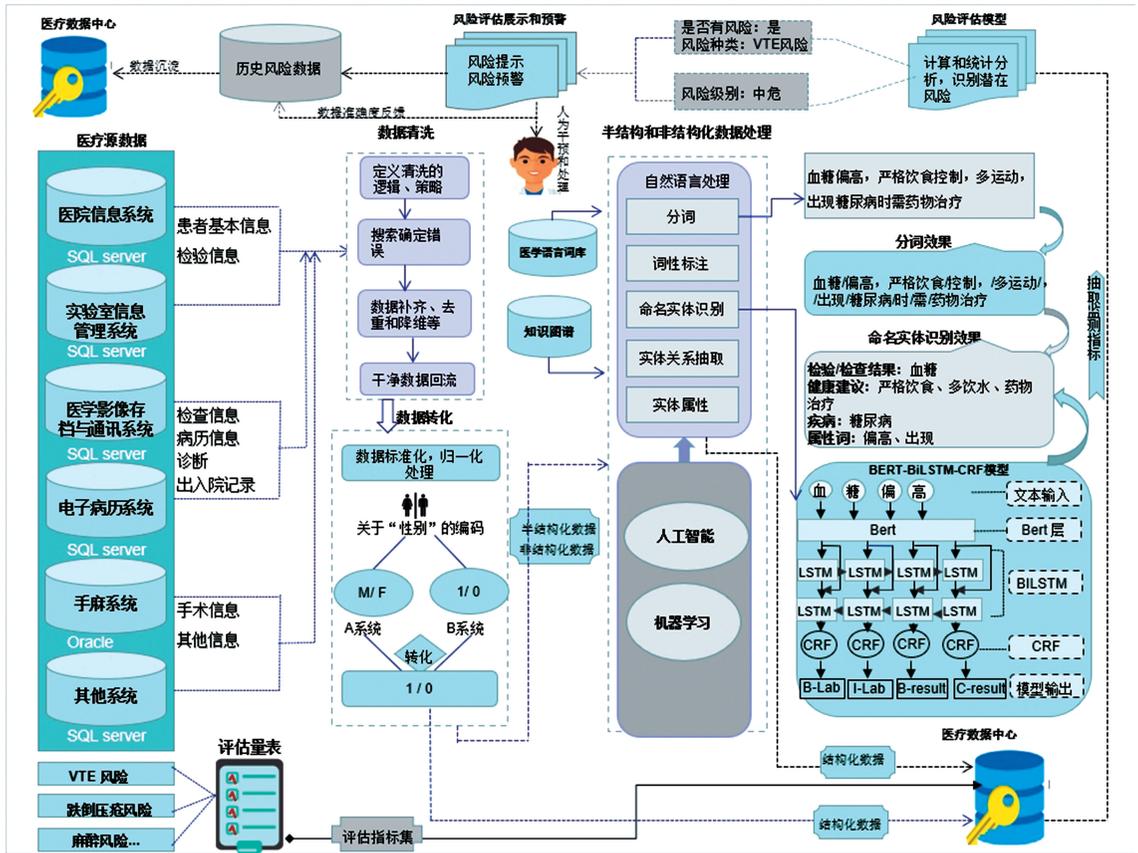


图 1 住院患者风险评估系统架构与实现过程

1.2.5 数据挖掘

对上述整合数据进行分析,采用关联规则、分类、聚类和时序分析等方式,挖掘不同因素与患者风险相关性用以支持临床决策。(1)数据分析:根据预设的目标和分析问题,选择关联规则、分类、聚类和时序分析等方式,对数据进行探测性分析,包括基础统计分析、变量筛选和特征提取等,以寻找数据中的规律和模式。(2)模型建立:根据分析结果,建立数据挖掘模型。模型建立的流程分为模型选择、模型训练和模型评价等步骤。模型选择采用分类器、回归器、聚类等技术训练模型,并根据结果进行优化。(3)模型评价:通过不同的指标和评判标准,对模型的准确性、稳定性、可解释性等进行评价。(4)模型应用:将建立的模型应用于临床风险评估中。通过对模型进行验证和测试,评估模型的有效性和实际应用效果,以提供支持临床决策的先验知识。医疗数据挖掘的具体方法是数据准备、数据分析、模型建立和模型应用的综合

过程,从大量数据中分析、挖掘出有用的信息,支持临床医生的诊断和治疗决策。

1.2.6 机器学习

利用上述挖掘的数据建立模型,预测潜在临床风险。将收集的规范化数据转化为适合机器学习算法的格式。(1)特征工程:选择重要的特征并对这些特征进行转换和降维,以提高算法的准确性。(2)选择算法:选择适合的机器学习算法,如决策树、logistic 回归、人工神经网络等^[14-15]。(3)训练模型:将收集到的数据分为训练集和测试集,使用训练集来训练机器学习模型,优化模型的参数。(4)验证模型:使用测试集来验证训练出的模型的准确性和可靠性,选择最佳模型。(5)应用模型:将训练出的模型应用于新的住院患者风险评估,验证人工评估结论。

1.2.7 综合评估系统页面展示

风险综合评估系统与住院医生站、护理系统打通,系统后台自动提取数据并根据评分规则计算患者

各类风险的级别,将各类风险中最高级别的风险显示在床位卡上。当后台计算出患者具有任意一种风险时,床位卡会显示为高、中、低。对于高危风险患者,系统将会发送详细的风险信息给管床医生和管床护士。医护人员也可以通过点击风险按钮进入风险综合评估系统查看风险详情,详情页面会推荐给医生针对性的治疗方案,以及建议完善的相关检查项目。患者入院后、手术前、手术后、患者体征变化及转科都会自动触动系统重新计算评分。医生和护士通过随机筛选部分评分数据进行人工复核,经过复核的数据将进一步沉淀入数据库,完成评分闭环。

1.2.8 可视化风险矩阵

在风险详情页面,系统将根据数据挖掘和机器学习结果,将患者的风险以可视化风险矩阵图的方式展示,同时对应给出相应的诊疗建议。以 VTE 风险为例,具体实施流程如下。(1)确认 Caprini 量表评分范围与深静脉血栓的风险等级相对应。根据 Caprini 量表评分,将风险矩阵分为 4 级:低危(0~1 分)、中危(2 分)、高危(3~4 分)、极高危(≥ 5 分),不同风险等级标记为不同颜色区域,构建出风险矩阵。低危区域为绿色,中危区域为黄色,高危、极高危区域为红色。(2)形成风险矩阵:将每个 Caprini 量表评分等级对应的风险类别放入 1 个矩阵中,以便在患者手术前的风险评估能够更好地确定风险类别和可能的预防措施,为医护人员提供清晰的可视化结构图。(3)识别预防措施:根据每个风险类别确定相应预防措施。对于中高风险患者,医护人员可以考虑通过使用某些药物来降低患者风险。其他可能的预防措施包括推荐患者

在手术前行走、穿戴静脉曲张袜或进行神经肌肉电刺激等。(4)风险评估回溯与持续改进:对实施风险应对方案的患者进行定期回访和风险评估,及时调整治疗方案,降低患者风险。根据回溯结果和实践经验,不断完善风险矩阵和风险应对方案,提高患者的安全性和治疗效果。

2 结 果

本项目将机器赋分与人工赋分用时情况进行对比(表 2):在 Caprini 量表(外科)、Padua 量表(内科)、NNIS、Nomogram、Morse 评估量表/ Hendrich 模型、Braden、ASA2023、NRS2002 评分中机器赋分用时明显低于人工赋分用时($P < 0.05$)。完成机器赋分后,根据病例获取的分值自动生成风险矩阵图,从低危、中危、高危、极高危,颜色由绿色过度到深红色,风险级别逐渐提高,见图 2。



图 2 风险评估系统自动生成风险矩阵图(以 VTE 为例)

表 2 风险评估系统与人工评分用时统计

评分量表	评估例数(n)	机器赋分用时(s)	人工赋分用时($\bar{x} \pm s$)	Kappa 值	P
Caprini 量表(外科)	6	0.62	132.62±14.04	1.00	<0.05
Padua 量表(内科)	6	0.62	105.88±11.94	1.00	<0.05
NNIS	5	0.62	164.79±11.48	1.00	<0.05
Nomogram	12	0.62	145.22±13.00	0.87	<0.05
Morse 评估量表/Hendrich 模型	8	0.62	127.85±14.00	0.83	<0.05
Braden	6	0.62	172.33±19.04	0.80	<0.05
ASA2023	7	0.63	115.88±19.52	1.00	<0.05
NRS2002	13	0.63	151.79±17.00	0.90	<0.05

3 讨 论

医护交流沟通的不足、评价能力的不同、患者信息无法及时共享等是发生医疗风险的重要原因^[16-18]。为减少医疗风险事件的发生,确保患者信息共享,本院在现有入院评估表单中新增了综合系统评估内容并形成诊疗综合评估单,通过数据支持和机制保障,使用科学化、规范化、精细化手段,提升医疗质量安全管理,保障医疗质量安全。

电子病历是医疗数据的主要来源^[19],包含了大量与疾病相关的信息,其中疾病相关和进展信息的医疗

记录对风险评估尤其重要^[20]。然而医院的病历内容多以非结构化的文本进行存储^[21],挖掘病历内容中的有效信息用于疾病的诊疗监测存在一定难度^[22]。如何准确、高效地从病历文本中抽取出有用的信息成为医学信息学研究的重点^[23]。本系统主要策略是利用信息提取和自然语言处理^[24]技术,将半结构化和非结构化文本转换为计算机可读的结构化数据。

本项目建立的综合评估系统基于《国际风险管理标准 ISO 31000》构建,将患者的诊断、病程记录与检验、检查等结果相关联,利用病程记录,检验、检查结

果记录等与疾病间的相关性进行疾病风险的综合预测,同时将疾病时间关系纳入疾病风险预测建模,最终通过风险矩阵呈现实时的可视化结果^[25]。风险矩阵又称决策矩阵风险评估技术,是一种系统方法,用于确定风险级别、比较不同风险,并确定需要首先控制的风险。其由可能性(概率、频率)和后果(严重性、影响)组成。风险优先级和资源分配取决于风险评级在风险矩阵中的位置。本项目完成了 VTE 风险、手术风险、多重耐药菌感染风险、跌倒坠床风险、压疮风险、麻醉风险和营养风险等项目的综合评估模型,医护人员通过本系统可以方便地对住院患者的风险进行统一监督和管理,降低赋分用时,自动生成风险矩阵图,使评估效率和准确性大幅提升,同时也能提供更精准的诊疗方案,缩短患者住院时间,降低医疗费用。

参考文献

- [1] 蒋立聪,蒋锋. 医务人员工作动机与工作行为与医疗质量关系的研究现状[J]. 现代医院,2023,23(1):30-32.
- [2] 蒋洁云. 信息化建设助力全过程医疗质量安全管理与实践探索研究[J]. 中国乡村医药,2023,30(4):68-69.
- [3] 国家卫生健康委办公厅. 国家卫生健康委办公厅关于印发 2022 年国家医疗质量安全改进目标的通知 [EB/OL]. [2022-03-02]. <http://www.nhc.gov.cn/yzygj/s3585/202203/ffed3474b1884058841a07c144ad094e.shtml>.
- [4] GIACOMO P, MATTEO R, EMMA M, et al. Risk analysis in healthcare organizations: methodological framework and critical variables[J]. Risk Manag Healthc Policy, 2021, 14: 2897-2911.
- [5] 许苹,周琳,马玉琴,等. 医疗风险预警预控系统框架构建研究[J]. 成都医学院学报,2023,18(5):611-619.
- [6] 杨万杰,侯晓明,孟祥飞,等. 智能计算法与传统计算法对住院患者风险评分的对照研究[J]. 中华危重病急救医学,2022,34(5):533-537.
- [7] 陈政,彭华. 北京协和医院医疗风险防控体系实践与探索[J]. 中国医院,2019,23(6):51-53.
- [8] 刘雯,龙颖,刘磊. DRG 支付背景下的手术患者住院费用分析与医用耗材成本管控探讨[J]. 保健医学研究与实践,2023,20(11):132-135.
- [9] 卿放,方浩宇. DRGs 支付方式下我国公立医院成本控制研究[J]. 会计之友,2022,34(10):89-98.
- [10] 尹思艺,庞晓燕,蔡秀军,等. 基于自然语言处理的病历智能质控系统的研究与应用[J]. 中国医药科学,2021,11(16):1-4.
- [11] 张聪品,方滔,刘昱良. 基于 LSTM-CRF 命名实体识别技术的研究与应用[J]. 计算机技术与发展,2019,29(2):106-108.
- [12] 郑增亮,蔡晓琼,黄继汉,等. 知识图谱在医学领域的应用综述[J]. 生物医学工程学杂志,2023,40(5):1040-1044.
- [13] 肖庆颖,于广军. 医疗大数据的研究与进展[J]. 上海医学,2023,46(7):420-423.
- [14] 张玉坤,刘茂福,胡慧君. 基于联合神经网络模型的中文医疗实体分类与关系抽取[J]. 计算机工程与科学,2019,41(6):1110-1118.
- [15] 朱铁兵,柏志安. 基于深度学习的医疗数据智能分析算法研究[J]. 电子设计工程,2021,29(12):84-88.
- [16] 张臻,王一如,朱宁,等. 疾病诊疗过程中的风险防范及医患沟通[J]. 医学与哲学,2021,42(4):53-56.
- [17] 秦晓强. 某三级甲等综合医院医疗风险识别评估分析[J]. 经济师,2022,37(11):258-260.
- [18] 郭东明. 基于风险管理的公立医院内部控制的优化探讨[J]. 财会学习,2021,16(31):155-157.
- [19] 任牡丹,杨柳,殷杰,等. 电子病历数据质量研究的国内外差异[J]. 中国卫生事业管理,2023,40(4):268-274.
- [20] 欧梦婵,郝学超,朱涛. 结构化信息系统在风险评估教学中的应用[J]. 教育教学论坛,2021,12(27):45-48.
- [21] 蔡翟源,陈杰,奚雪峰,等. 电子病历的关系语义实体识别[J]. 苏州科技大学学报(自然科学版),2023,40(3):62-70.
- [22] 张宸华. 数据挖掘算法在医疗管理中的应用研究[J]. 软件,2023,44(6):157-159.
- [23] 王辰,李明,马金刚. 电子病历关系抽取综述[J]. 计算机工程与应用,2023,59(16):63-73.
- [24] 朱彦华. 基于自然语言处理和知识图谱的医疗文本挖掘与知识提取[J]. 信息与电脑(理论版),2023,35(14):1-3.
- [25] MAKAJIC-NIKOLIC D. ISO 31000: risk management guidelines[M]. Berlin: Springer, 2020: 1-4.

(收稿日期:2023-09-20 修回日期:2024-06-12)

(编辑:张芑捷)