

自发性脑出血临床专病数据库应用及研究进展

田泰宇¹, 李智恒¹, 朱 甜¹, 田 志^{2*}

¹吉首大学医学院, 湖南 吉首

²吉首大学临床医院, 湖南 吉首

收稿日期: 2025年4月23日; 录用日期: 2025年5月16日; 发布日期: 2025年5月27日

摘 要

随着人口老龄化的发展, 自发性脑出血的发病率居高不下, 全球年发病率高达24.6/100万。因此, 构建自发性脑出血专病数据库具有重要临床价值。通过系统收集的大量医学数据可以提供流行病学资料。与常规病历系统不同, 它支持快速生成可验证的临床假设。总之, 专病数据库为自发性脑出血的病因、临床过程、结局、预后提供了可靠的医学数据。本文重点介绍专病数据库的特点、在临床中的应用及研究进展及人工智能技术在构建数据库中的应用, 以促进多地区自发性脑出血数据库的构建、辅助医疗决策并提高精准医疗水平。

关键词

自发性脑出血, 数据库, 人工智能, 研究进展, 综述

Applications and Research Progress of Clinical Specialty Databases in Spontaneous Intracerebral Hemorrhage

Taiyu Tian¹, Zhiheng Li¹, Tian Zhu¹, Zhi Tian^{2*}

¹School of Medicine, Jishou University, Jishou Hunan

²College of Clinical Medical, Jishou University, Jishou Hunan

Received: Apr. 23rd, 2025; accepted: May 16th, 2025; published: May 27th, 2025

Abstract

The development of population ageing has resulted in a high incidence of spontaneous cerebral haemorrhage, with a global annual incidence rate of 24.6 per million. Consequently, there is a significant

*通讯作者。

文章引用: 田泰宇, 李智恒, 朱甜, 田志. 自发性脑出血临床专病数据库应用及研究进展[J]. 临床医学进展, 2025, 15(5): 1936-1941. DOI: 10.12677/acm.2025.1551577

clinical need to establish a database for spontaneous cerebral haemorrhage. The database will collate a substantial volume of medical data, providing valuable epidemiological insights. In contrast to conventional medical record systems, it will facilitate the rapid formulation of testable clinical hypotheses. In conclusion, the disease-specific database provides reliable medical data on the etiology, clinical course, outcome, and prognosis of spontaneous cerebral haemorrhage. The focus of this paper is on the characteristics of the speciality databases, their application in clinical practice and research progress, and the application of artificial intelligence technology in constructing the databases. The aim is to facilitate the construction of multi-regional spontaneous cerebral haemorrhage databases, to assist medical decision-making and to improve precision medicine.

Keywords

Spontaneous Intracerebral Hemorrhage, Database, Artificial Intelligence, Research Progress, Review

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自发性脑出血(Spontaneous Intracerebral Hemorrhage, SICH)是指非外伤情况下,脑血管急性破裂引起脑出血并造成脑损伤的疾病,全球 SICH 的年发病率为 24.6/100 万,而在发展中国家发病率更高[1]。随着人口老龄化的逐步进展, SICH 的发病率将居高不下。SICH 发展迅速,可造成永久性神经功能损伤,易合并严重并发症甚至死亡, SICH 的早期死亡率达 30%~40%,这对公众健康问题造成巨大困扰,同时也给社会带来巨大负担[2]。中国各地区人群在饮食、文化习俗、作息的差异使得 SICH 的个体化精准医疗实施困难[3]。同时传统诊疗依赖经验判断,缺乏数据支持,精准医疗的精准度差。通过 SICH 专病数据库的构建,利用数据分析挖掘与人工智能等技术掌握不同区域、不同病因自发性脑出血高危因素、发生、发展和转归特点,进而提供临床循证医学证据,提高不同区域 SICH 诊疗方案的个性化水平,对辅助提高数据驱动下的自发性脑出血医疗诊治决策有积极意义。

当前中国正努力推进国家级别数据库的建设,2023 年 3 月 7 日国务院宣布组建国家数据局,从国家层面协调推进数据基础制度建设。中国卫生健康信息化领域顶层设计体系基本形成,这一举动为我国健康医疗大数据指明发展方向[4]。然而,中国的医疗信息化发展表现出一种发展不均衡且底层数据缺少统一的行业标准,这导致原始的医疗数据需通过标准化后才可实现数据互通。同时随着精准医学和临床亚专科的兴起,针对某一类甚至一种疾病的研究愈加深入,对临床病例数据的依赖性也日益增加。提高临床科研数据可用性,是专病数据库产生的直接原因,也为后续实现互通互联提供基础[5]-[7]。

2. 自发性脑出血专病数据库的概念

自发性脑出血专病数据库是指通过将自发性脑出血相关医学专业信息标准化与计算机数据管理系统相结合,并经过数据处理、质量控制的数据仓库,是一种处于传统病例和基于人群的科学研究之间的研究[8]。数据库将自发性脑出血患者的基线资料、出血部位、病史、临床表现、体格检查、实验室检验、影像学检查及特殊检查、治疗方式、预后及随访情况等信息与计算机数据管理系统结合并应用。

3. 自发性脑出血专病数据库的分类

临床科学研究数据库主要包括临床试验数据库、病例注册数据库和专科专病数据库等。其中主要用

于按专科或病种收集病例数据和开展观察性研究当属专科专病数据库。专病数据库从参与构建数据库的中心数量出发,分类可分为单中心数据库和多中心数据库,从收集数据场所出发可分为基于医院的数据库和基于社区的数据库[7]。其建库目的主要有两类,第一类是有明确的研究问题,从研究问题出发收集相应数据,此类数据库,数据项目较少,表头相对简单。第二类是研究方向尚不清晰但有研究目的,建立数据库可以为日后的科学研究积累数据。此类数据库,数据项目广、多,表单相对复杂。

多中心专病数据库通常由数所三甲医院共同创立,它统筹的数据量大、范围广、疾病种类丰富并需要多学科共同合作,有利于疾病的全面化管理,故多中心数据库具有样本量大、多学科协作、数据共享率高、提高临床诊疗质量等优势,也因数据样本量大、多、杂,多中心数据的收集往往面临标准化和质量控制的问题,同时在数据共享中如何确保充分保护患者个人隐私的问题,还有数据库维护及地区差异等不足[9]。相较之下,单中心专病数据库虽有数据收集速度较慢,对接口需求较高等不足,但对于病例的标准化、质量控制、长期随访上具有明显的优势,且易于实施和操作,并为未来的多中心数据库打下基础。如表 1 所示。

Table 1. Classification and characteristics of disease-specific databases
表 1. 专病数据库分类及特点

单/多中心	特点	数据项及表单复杂度	数据标准及质量控制	随访
单中心	管理简单,数据一致性高 扩展性受限,外部验证困难	数据项精简 表单设计轻量化	水平高,操作简单	易
多中心	数据多样性强,结论普适性高 需解决数据异构性问题(如不同机构 电子病历系统差异)	数据项全面 表单复杂,需分层设计	样本量大、多、杂 易造成数据质量层次不齐	难

4. 国内外脑卒中数据库的应用及发展现状

病例的数据收集和数据管理是临床科学研究活动中十分重要的基础性活动。早期数据收集和管理依赖于纸质病例报告表(Case Report Form),人工数据收集、管理、核查带来了繁琐与人工误差等问题。电子化数据收集(Electronic Data Capture)和数据库系统的出现很大程度上解决了相关问题[10]。在电子化数据收集的时代背景下脑卒中专病数据库的构建应具备以下因素:1、经过标准化的数据标签;2、资金足够;3、科室或医院具备电子数据采集系统;4、拥有特定的数据管理及收集人员;5、在数据收集、数据清理、数据应用等整个流程中保护患者的个人隐私;6、团队经过相应培训[11]。

4.1.. 国外脑卒中专病数据库的应用及发展现状

1978 年美国国立卫生研究院国家神经疾病与卒中研究所开始脑卒中数据库预实验研究,建立了一套标准化的脑卒中患者电子病历资料[12],为此许多国家开始陆陆续续建立了属于自己的脑卒中数据库。如 1979 年开始录入的瑞士洛桑卒中,进行了约 25 年的随访,帮助确定了脑卒中患者的病因、危险因素等的死亡率的趋势[13]。同时世界卫生组织组建的心脑血管疾病登记中心也在 1979 年实施,目的是探索不同人群心脑血管疾病的病因和危险因素[14]。然而不足的是这种早期的数据库结构和数据类型简单。随着电子计算机技术的蓬勃发展,越来越多的科研工作者开始建立和应用脑卒中数据库。1991~1996 年期间开展的国际卒中试验,在早期服用阿司匹林和肝素对缺血性脑卒中预后的影响上做出了巨大贡献[15],但国际卒中试验仍有数据相对单一等不足。此后由 86 所美国医院共同参与的美国网络国家急性卒中治疗和操作规范化计划中心开始创立,这种多中心的数据收集和分析,大大促进了脑卒中护理技术的发展[16]。往后,这种多中心国家大型数据库成为主流。不过此时的数据库虽能很好地保存、收集脑卒中患者的临床

基线资料,但缺少对影像学资料等非结构化数据的收集。随着科研方法论的进步,前瞻性队列研究为载体的,聚焦于既定研究方向的多中心脑卒中数据库诞生。如 21 世纪初建立的 INTERACT 试验(Intensive Blood Pressure Reduction in Acute Cerebral Haemorrhage Trial)数据库[17]等。此类数据库数据项丰富是 21 世纪脑卒中数据库建设的主流之一。故专病数据库的建立不仅要聚焦于传统的通用型数据库,还要收集影像学资料等非结构化数据,以此来完善数据库。

4.2. 国内脑卒中专病数据库的应用及发展现状

1995 年由北京 7 所医院共同参与构建的多中心脑卒中数据库拉开序幕,不过其后续报道不多[18],且由于没有统一的标准以及质量控制等,此类数据库应用较为局限。目前国内最具代表性的卒中数据库是 2007 年建立的中国国家卒中数据库,填补了国内脑卒中数据库的空白。建库以来,先后发表了多项极具影响的报道。该数据库为中国卒中提供了数据支持[19]。2017 年,北京大学第一医院组织成立中国急性缺血性卒中治疗结局登记数据库,目的是探索中国收益最优的脑卒中治疗方式[20]。多中心数据库固然有数据量大、全面的优势,但数据处理困难,很大程度上数据的可靠性受到质疑,此外多中心数据库不能很好地体现各地区、各级别医院的诊疗水平和地理因素等对脑卒中造成的影响,且由于国内经济水平差异,能够参与的地区及医院也多为东方地区。同时国内自发性脑出血数据库仍是大量缺乏,故单中心的自发性专科专病数据库仍有其独特优势。

5. 人工智能技术在专病数据库上的应用

人工智能(Artificial Intelligence, AI)是一个通用术语,意味着使用计算机模拟智能行为,尽量减少人为干预[21]。通过运用 AI 技术可开展数据驱动下的健康管理及风险因素预警、病因分析、医学检验与影像学、疾病诊断、诊疗方案与手术计划、并发症发展的高危因素、术后疗效评估与康复手段等医疗诊断过程的自主学习,实现基层医院对疾病精准诊断、精准治疗。

目前,医学影像自动分割、辅助疾病诊断、药物研发、远程医疗与护理等人工智能技术愈发成熟,对于数据库的构建产生了新的应用。例如,利用人工智能的自然语言处理技术可完成数据清晰、映射、储存等工作[22],很大程度上减少了人工误差,节约了时间。这种自然语言处理能力甚至可以自动审查图表,从文字报告图像中提取结果,进一步提高了工作效率[23]。此外,医学影像自动分割对比人工手动分割更是有着显著的优越性,这使得对于脑出血、脑水肿体积的计算更为精准[24]。北京协和医院就基于此建立了一套卷积神经网络的颅内血肿自动分割模型[25]。运用卷积神经网络学习国家中心数据库中 2166 例脑出血影像。目前系统已取得显著成效,可以辅助临床医师判断脑出血常见的 5 种病因,对于各种类型的脑出血原因,其敏感度、特异度及准确性均高达 80%以上。对于动脉瘤性脑出血和高血压脑出血的准确度最高,错误率小于 6% [26]。可见 AI 在构建专病数据库各个环节都有着很大的助力,数据内容更加精确,数据可靠性更高。同时, AI 也可辅助数据库进行机器学习预测模型, Inaguma 等基于此发现慢性肾病患者肾小球滤过率下降的危险因素之一就是尿蛋白含量增加[27]。Zhi Geng 等人利用五种机器学习模型预测自发性脑出血患者 90 天内的短期预后结局,并筛选出 NIHSS 评分、AST 水平、年龄、白细胞计数、血肿体积等关键预测因子,同时也指出整合 CT/MRI 影像的自动化分割技术将对模型进一步优化[28]。Koutarou Matsumoto 等整合多模态数据利用机器学习结合 Kaplan-Meier 曲线预测院内死亡风险,专家评估下受试者工作特征曲线下面积为 0.97 (95% CI, 0.94~0.99) [29]。可见 AI 已融入现代医疗体系,在数据收集、临床影像学、病因分类、预后等多个环节上均表现出变革性潜力,推动了精准医疗前进的脚步。

6. 挑战与展望

未来, AI 和数据库将继续融合,智能化的数据收集、治理、储存将成为专病数据库构建的发展趋势。

但我国卫生信息标准起步晚,标准执行差,各医疗单位数据质量参差不齐,若有新的医学概念和治疗方式出现又如何保证其准确性和时效性。这使得数据互通、共享实现难度高,各数据库之间、医院内部之间、医院之间的可操作性差,存在“数据孤岛”问题。在互通互联的同时,患者的隐私保护也是一个巨大的问题。在数据采集上,因病例多源、多模态的特点、汉字的歧义性和原始数据不完整等因素加大了数据提取难度。对 AI 来说,在数据标注上仍有缺陷[6]。

为提供相关流行病学研究及科学提升对自发性脑出血患者的诊疗水平及生命质量,建立一个相关数据库意义重大。一方面,可以在数据库指导下掌握高危人群的地域分布特点,对具有高危因素的人群进行健康教育、实时监控及防治。另一方面,在实践中逐渐完善、丰富数据库,使之成为一个动态化、个体化并具有前瞻性的自发性脑出血防治体系,从而优化救治流程,实现超早期救治,以降低自发性脑出血的发病率、死亡率、致残率和复发率。未来应创建更多的脑出血数据库,尽量兼顾社区和医院,逐步完善各地区对脑出血的个性化防控系统,提升患者及人民幸福生活水平。

参考文献

- [1] Chamberg-Michilot, D., Brañez-Condorena, A., Alva-Díaz, C., Sequeiros, J., Abanto, C. and Pacheco-Barrios, K. (2021) Evidence-Based Appraisal of Blood Pressure Reduction in Spontaneous Intracerebral Hemorrhage: A Scoping Review and Overview. *Clinical Neurology and Neurosurgery*, **202**, Article 106497. <https://doi.org/10.1016/j.clineuro.2021.106497>
- [2] Greenberg, S.M., Ziai, W.C., Cordonnier, C., Dowlatshahi, D., Francis, B., Goldstein, J.N., *et al.* (2022) 2022 Guideline for the Management of Patients with Spontaneous Intracerebral Hemorrhage: A Guideline from the American Heart Association/American Stroke Association. *Stroke*, **53**, e282-e361. <https://doi.org/10.1161/str.0000000000000407>
- [3] Liu, L., Wang, D., Wong, K.S.L. and Wang, Y. (2011) Stroke and Stroke Care in China: Huge Burden, Significant Workload, and a National Priority. *Stroke*, **42**, 3651-3654. <https://doi.org/10.1161/strokeaha.111.635755>
- [4] 董方杰, 胡建平, 吴士勇. 我国卫生健康信息互联互通 2.0 技术特征研究[J]. 中国卫生信息管理杂志, 2023, 20(1): 1-6.
- [5] 刘迷迷, 杜国霞, 周毅, 等. 专病数据库建设与应用研究[J]. 医学信息学杂志, 2021, 42(11): 81-86, 93.
- [6] 郭强, 王丛, 衡反修. 医疗大数据平台建设机遇、挑战及其发展[J]. 医学信息学杂志, 2021, 42(1): 2-8.
- [7] 薛万国, 乔岫, 车贺宾, 等. 临床科研数据库系统的现状与未来[J]. 中国数字医学, 2021, 16(1): 2-6.
- [8] Brainin, M. (1994) Overview of Stroke Data Banks. *Neuroepidemiology*, **13**, 250-258. <https://doi.org/10.1159/000110388>
- [9] 陈亦豪, 常健博, 魏俊吉, 等. 脑卒中大型医学数据库应用及研究进展[J]. 中国现代神经疾病杂志, 2021, 21(3): 141-146.
- [10] 谢高强, 李英山, 姚晨. 电子数据采集对我国临床研究的机遇和挑战[J]. 中国新药杂志, 2013, 22(6): 620-623.
- [11] Schwamm, L., Reeves, M.J. and Frankel, M. (2006) Designing a Sustainable National Registry for Stroke Quality Improvement. *American Journal of Preventive Medicine*, **31**, S251-S257. <https://doi.org/10.1016/j.amepre.2006.08.013>
- [12] Bronstein, K., Murray, P., Licata-Gehr, E., Banko, M., Kelly-Hayes, M., Fast, S., *et al.* (1986) The Stroke Data Bank Project: Implications for Nursing Research. *Journal of Neuroscience Nursing*, **18**, 132-134. <https://doi.org/10.1097/01376517-198606000-00005>
- [13] Bogousslavsky, J., Van Melle, G. and Regli, F. (1988) The Lausanne Stroke Registry: Analysis of 1,000 Consecutive Patients with First Stroke. *Stroke*, **19**, 1083-1092. <https://doi.org/10.1161/01.str.19.9.1083>
- [14] WHO MONICA Project Principal Invest (1988) The World Health Organization Monica Project (Monitoring Trends and Determinants in Cardiovascular Disease): A Major International Collaboration. *Journal of Clinical Epidemiology*, **41**, 105-114. [https://doi.org/10.1016/0895-4356\(88\)90084-4](https://doi.org/10.1016/0895-4356(88)90084-4)
- [15] Kapral, M.K., Laupacis, A., Phillips, S.J., Silver, F.L., Hill, M.D., Fang, J., *et al.* (2004) Stroke Care Delivery in Institutions Participating in the Registry of the Canadian Stroke Network. *Stroke*, **35**, 1756-1762. <https://doi.org/10.1161/01.str.0000130423.50191.9f>
- [16] Shiotsuki, H., Ogushi, Y., Fushimi, K., *et al.* (2005) Evaluation of Applied Cases of Thrombolytic Therapy against Ultra-Acute Ischemic Stroke. Using the Japanese Standard Stroke Registry Database. *The Tokai Journal of Experimental and Clinical Medicine*, **30**, 49-62.

- [17] California Acute Stroke Pilot Registry (CASPR) Investigators (2005) Prioritizing Interventions to Improve Rates of Thrombolysis for Ischemic Stroke. *Neurology*, **64**, 654-659. <https://doi.org/10.1212/01.wnl.0000151850.39648.51>
- [18] 高晓兰, 胡长梅, 王文志, 等. 出血性卒中与缺血性卒中危险因素对比分析——多中心脑卒中数据库临床研究[J]. 中国慢性病预防与控制, 1999, 7(4): 14-16.
- [19] 刘小玲, 葛朝明. 脑卒中数据库的研究进展[J]. 中国医学创新, 2017, 14(1): 145-148.
- [20] Sun, W., Ou, Q., Zhang, Z., Qu, J. and Huang, Y. (2017) Chinese Acute Ischemic Stroke Treatment Outcome Registry (CASTOR): Protocol for a Prospective Registry Study on Patterns of Real-World Treatment of Acute Ischemic Stroke in China. *BMC Complementary and Alternative Medicine*, **17**, Article No. 357. <https://doi.org/10.1186/s12906-017-1863-4>
- [21] Hamet, P. and Tremblay, J. (2017) Artificial Intelligence in Medicine: Clinical and Experimental. *Metabolism*, **69**, S36-S40. <https://doi.org/10.1016/j.metabol.2017.01.011>
- [22] 王耀国, 李鹏, 刘迷迷, 等. 临床专病数据库建设现状与思考[J]. 医学信息学杂志, 2024, 45(3): 65-69.
- [23] Juhn, Y. and Liu, H. (2020) Artificial Intelligence Approaches Using Natural Language Processing to Advance Ehr-Based Clinical Research. *Journal of Allergy and Clinical Immunology*, **145**, 463-469. <https://doi.org/10.1016/j.jaci.2019.12.897>
- [24] Renard, F., Guedria, S., Palma, N.D. and Vuillerme, N. (2020) Variability and Reproducibility in Deep Learning for Medical Image Segmentation. *Scientific Reports*, **10**, Article No. 13724. <https://doi.org/10.1038/s41598-020-69920-0>
- [25] Chang, J.B., Jiang, S.Z., Chen, X.J., Luo, J.X., Li, W.L., Zhang, Q.H., *et al.* (2020) Consistency Evaluation of an Automatic Segmentation for Quantification of Intracerebral Hemorrhage Using Convolution Neural Network. *Chinese Journal of Contemporary Neurology and Neurosurgery*, **20**, 585-590. <https://doi.org/10.3969/j.issn.1672-6731.2020.07.005>
- [26] 潘锋. 人工智能引领神经外科医疗进入新时代[J]. 中国医药导报, 2023, 20(12): 1-3.
- [27] Inaguma, D., Kitagawa, A., Yanagiya, R., Koseki, A., Iwamori, T., Kudo, M., *et al.* (2020) Increasing Tendency of Urine Protein Is a Risk Factor for Rapid EGFR Decline in Patients with CKD: A Machine Learning-Based Prediction Model by Using a Big Database. *PLOS ONE*, **15**, e0239262. <https://doi.org/10.1371/journal.pone.0239262>
- [28] Geng, Z., Yang, C., Zhao, Z., Yan, Y., Guo, T., Liu, C., *et al.* (2024) Development and Validation of a Machine Learning-Based Predictive Model for Assessing the 90-Day Prognostic Outcome of Patients with Spontaneous Intracerebral Hemorrhage. *Journal of Translational Medicine*, **22**, Article No. 236. <https://doi.org/10.1186/s12967-024-04896-3>
- [29] Matsumoto, K., Ishihara, K., Matsuda, K., Tokunaga, K., Yamashiro, S., Soejima, H., *et al.* (2024) Machine Learning-Based Prediction for In-Hospital Mortality after Acute Intracerebral Hemorrhage Using Real-World Clinical and Image Data. *Journal of the American Heart Association*, **13**, e036447. <https://doi.org/10.1161/jaha.124.036447>