科学通报 2025 年 第 70 卷 第 27 期: 4619 ~ 4624
观 点
《中国科学》杂志社
SCIENCE CHINA PRESS

医学大模型的构建: 技术、挑战与发展

李岳峰, 赵韡*

国家卫生健康委统计信息中心, 北京 100044
* 联系人, E-mail: zhaowei@nhc.gov.cn

1 医学大模型的价值

2017年, 基于Transformer架构的人工智能(artificial intelligence, AI)大语言模型(large language models, LLMs)提出以来, 已经在自然语言处理、计算机视觉、多模态、音频和语音处理以及信号处理等应用领域得到广泛应用[1,2], 并建立了多模态大语言模型(multimodal large language models, MLLMs), 以大模型为代表的AI和大数据、云计算、物联网和量子计算等技术共同构成人类社会从传统“物理世界”向“数字世界”转变的关键技术群. 医学领域作为多模态大数据较为集中的领域, 利用传统的卷积神经网络(convolutional neural network, CNN)和生成对抗网络(generative adversarial networks, GAN)等机器学习方法, 进行图像的人工智能处理, 取得一些重要的临床应用成效[3]. 随着MLLMs技术的发展, 通过自注意力机制(Self-attention)捕捉输入序列中不同位置之间的依赖关系以及时间序列上的变化关系, 提高图像超分辨率, 能够更好地处理全局上下文、高频细节恢复和智能3D建模等, 可以实现机器翻译、图像识别诊断、智能推理(reasoning)、具身医用机器人(embodied medical robot)和脑机接口(brain-machine interface, BMI)应用等复杂任务, 所以医学科学和智能科学的有机结合, 为提升医疗服务效能、推动医学创新发展和促进生命健康水平, 提供了非常广泛的应用前景和价值[4]. 医用机器人在物理学机理基础上, 开发应用柔性肌肉等技术可以实现医学上类似“小脑功能”, 在此基础上通过模型推理, 将会具备医学上类似“大脑功能”. 在2024年11月, 国家卫生健康委员会、国家中医药管理局和国家疾病预防控制局发布《卫生健康行业人工智能应用场景参考指引》[5], 积极推进卫生健康行业“人工智能+”应用创新发展. 2025年1月, 美国卫生与公众服务部(United States Department of Health and Human Services, HHS)发布了“U.S. Department of Health and Human Services: Strategic Plan for the Use of Artificial Intelligence in Health, Human Services, and Public Health”, AI将全面赋能医疗研究、产品开发、临床服务、公共卫生等多个领域, 一场医疗健康领域的AI革命即将到来. 医学大模型通过技术融合与场景创新, 正在重塑医疗行业的诊断、治疗、管理与教育体系, 其价值核心在于提高诊疗精准度、降低医疗成本、实现普惠医疗. 本文就医学大模型构建的技术、挑战与趋势进行分析, 提出可行的发展路径.

赵韡 研究员, 博士生导师, 国家卫生健康委员会统计信息中心主任、党委书记, 兼任中国卫生信息与健康医疗大数据学会常务理事、中国生物医学工程学会医学人工智能分会副主任委员. 长期致力于医学信息学、健康医疗大数据及人工智能在临床医学中的创新应用研究, 聚焦医学信息技术、系统集成与智能诊疗技术开发, 取得多项系统性、原创性成果.

2 医学大模型的技术

医学大模型构建是一个系统而复杂的过程, 涉及关键技术、计算平台及医学专业术语的综合应用. 本文基于模型化、自动化和标准化的思路, 通过数据、模型、算法和工具等方面技术综合应用, 实现多模态的记忆、理解、推理和生成等类人脑的功能(图1).

2.1 总体技术框架

本文深入研究国内外主要大模型及其应用, 如ChatGPT、KiMi、DeepSeek、文心一言、天工等大模型, 提出医学大模型的技术流程架构(图2). 在医学大模型的构建中, 重点关注数据(语料)的归集, 适宜模型开发和分布式向量算力的配置等内容.

第一步, 语料收集. 利用系统或者平台间的应用程序编程接口(application programming interface, API)接口技术, 按

引用格式: 李岳峰, 赵韡. 医学大模型的构建: 技术、挑战与发展. 科学通报, 2025, 70: 4619–4624
Li Y, Zhao W. Building medical large language models: technology, challenge, and development (in Chinese). Chin Sci Bull, 2025, 70: 4619–4624, doi: 10.1360/CSB-2025-0346

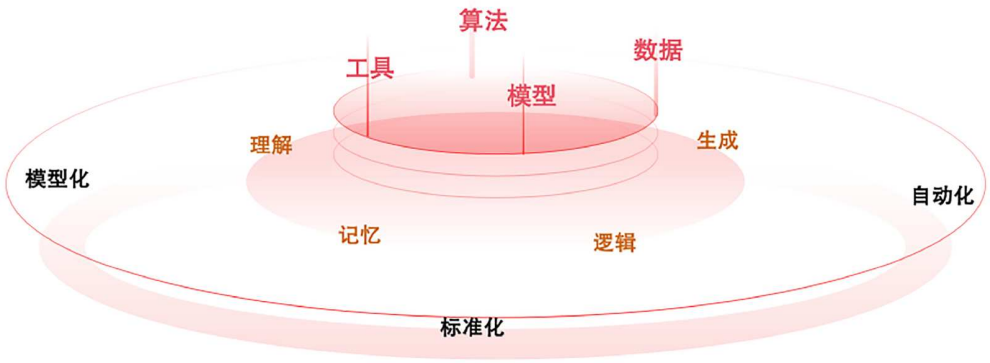

图 1 医学大模型的技术逻辑
Figure 1 The technical logic of medical large language models

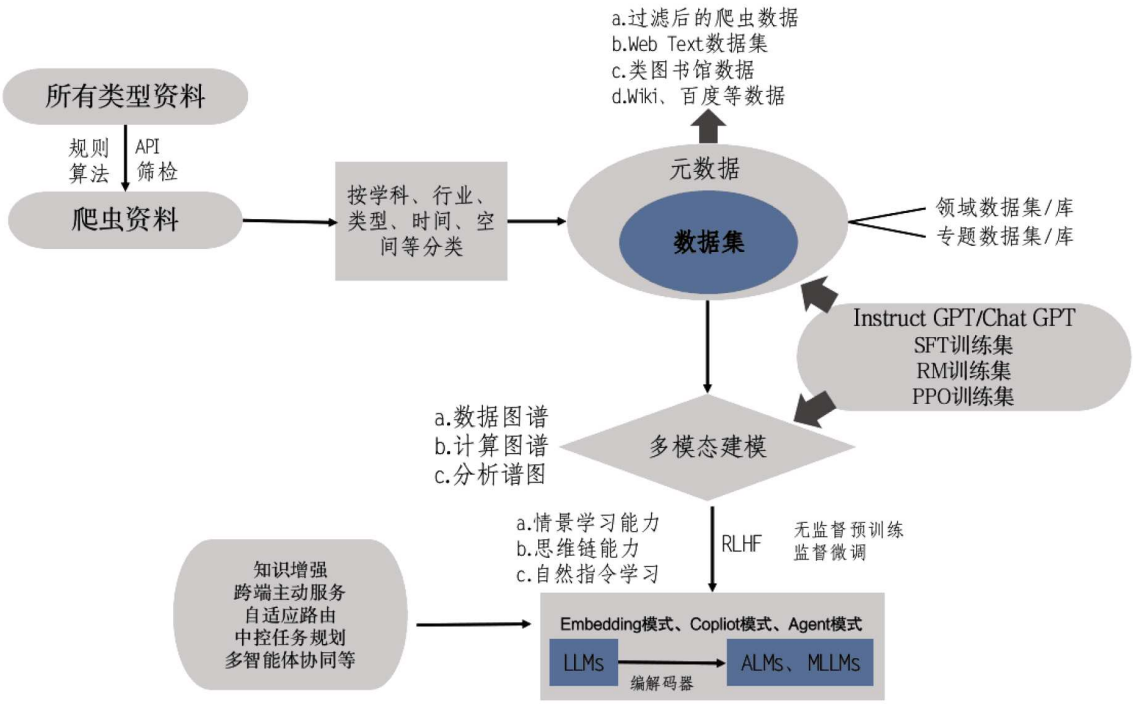

图 2 医学大模型技术架构
Figure 2 Medical large language model technology architecture

照预先制定的规则或者算法, 筛检所有类型或者模态的资料, 形成初级语料库, 即爬虫资料. 诸如过滤后的爬虫数据、Web Text数据集、类图书馆数据、Wiki、百度等数据.

第二步, 语料分类. 对这些资料按照学科、行业、类型、时间和空间等不同的维度进行分类, 利用标注技术和自然语言处理技术等, 建立由元数据刻画和管理的数据库, 在行业垂直大模型的语料库中还需要加入医学数据, 如电子病例数据、组学数据和专题调研数据等.

第三步, 建模分析. 人类的有限理性只能抓住现实中的一些重要特征模型化, 大模型能够高效地处理序列数据. 在模型预训练(Pretraining)的基础上采用微调技术(Fine-Tuning)不断调优模型. 通过将数据集分类, 建立训练数据集、奖励数据集和微调数据集, 用于模型开发的全过程.

第四步, 模型应用. 目前大模型的应用模式主要包括Embedding模式、Copliot模式和Agent模式等. 这些模型通过数据图谱、计算图谱和分析图谱等可视化展示, 运算和分析的

逻辑关系一目了然. 医学大模型目前应用模式主要有: (1) 基于知识库的智能分析、辅助决策和增强检索(retrieval-augmented generation, RAG)等; (2) 基于提示工程(prompt engineering)和长周期记忆(long-term memory)的内容生成与创作和多模态任务处理等; (3) 智能体协同应用等[6]. 目前大模型共享方式主要有全软件开源共享、软件部分功能或中间件开源和软件闭源付费使用等3种形式[7].

2.2 数据技术

数据技术是医学大模型构建的基础. (1) 数据收集和整合. 对数据资产进行收集和整合, 这可能涉及从不同来源收集数据, 并将其整合到一个统一的数据存储中, 例如数据仓库或数据湖. (2) 数据清洗和加工. 对数据进行清洗和加工, 以确保数据的质量和准确性, 包括规范数据格式、去除重复数据, 纠正错误、填充缺失值等处理. (3) 数据分析和建模. 通过数据分析和建模, 从数据资产中提取有用的信息和见解. 涉及统计分析、机器学习、数据挖掘等技术, 发现数据中的模式和趋势. (4) 价值创造. 通过对数据资产的分析和建模, 优化业务流程、改进产品和服务等, 从而创造价值. 五是数字化资产化. 通过将数据资产转化为数字形式, 并将其与其他数字资产(如软件、知识产权等)结合, 可以形成数字资产, 这些数字资产可以成为组织或个人的财产, 并为其带来价值. 医学的智慧结晶在于不断创新的医学术语, 统一概念的医学术语中可能包含多个同义词, 标准化和归一化的思路在医学大模型的构建中非常重要.

2.3 模型技术

模型技术除经典架构Transforms外, 还有混合专家模型(mixture of experts, MoE)、Mamba模型、类脑智能体(brain-inspired intelligence)、时序大模型等新型架构[8].

除编解码器技术外, 大模型还涉及提升工程、微调技术、模型压缩、模型推理、模型增强和模型对齐等技术, 每种技术都在不断地技术演进和迭代之中. 模型技术的关键在于: (1) 基于Transform架构和合适神经网络(或者多神经网络“串并联”)的选择, 并以此优化梯度和参数; (2) 各项技术的相互嵌套, 如提示工程[9]、检索增强生成[10]、参数高效微调(parameter-efficient fine-tuning, PEFT)[11]、基于人类反馈的强化学习(reinforcement learning from human feedback, RLHF)[12]、基于人工智能反馈的强化学习(reinforcement learning from AI feedback, RLAIF)[13]、多领域微调(multi-domain fine-tuning)[14]、直接偏好优化(direct preference optimization, DPO)[15]和自奖励机制(self-rewarding mechanisms)等方法可以互相补充使用, 不断提升模型性能和效率. 如DeepSeek在采用教师模型(teacher models)微调学生模型(student models)这样一种知识蒸馏技术不断优化模型性能, 较少不必要的计算量, 有效降低时间和经济成本.

2.4 算法技术

大模型是基于机器学习的算法技术, 这些方法以神经网络的方式编写于编解码器中, 编解码器中的神经网络可以采用单一性神经网络、多个神经网络并联、多个神经网络串联和混合型等4类模式. 目前医学领域大模型中, 前馈神经网络(feedforward neural networks, FNN)[16]、卷积神经网络[17]、长短期记忆网络[18]、生成对抗网络[19]、图神经网络(graph neural networks, GNN)[20]和深度强化学习网络等算法使用较多, 根据结构、功能和应用场景的不同, 可以选择合适的神经网络类型, 完成机器翻译、图像识别诊断、自然语言处理、语音识别、智能推理、具身机器人和脑机接口应用等复杂医学任务. 例如, CNN适合处理医学图像数据, RNN适合处理时间序列数据, 尤其是在队列分析和前瞻性研究中常用, Transformer在自然语言处理任务中表现出色, 可以提供基本的医学知识咨询和辅助诊断等服务, DeepSeek则将混合专家模型应用与模型微调, 降低成本提升效能. 一些医疗机构搭建不同大模型的API接口平台, 开展医学辅助诊疗和相关研究.

2.5 工具技术

大模型的运行需要强大的平台和算力技术支持. 从软件方面分析, 采用知识增强、跨端主动服务、自适应路由、中控任务规划和多智能体协同等策略, 底层技术依靠软件与算法的优化. (1) 深度学习框架的迭代. TensorFlow和PyTorch等深度学习框架的不断优化, 使得AI算法的实现更加高效. (2) 云计算与边缘计算的融合. 云计算平台(如华为云、阿里云、AWS云、Azure云、Google云等)提供了强大的AI算力支持, 同时边缘计算的兴起使得算力可以更接近数据源, 降低延迟, 提高响应速度. (3) 技术融合与创新. 多模态AI能够处理并融合文本、图像、音频和视频等多种类型的信息, 提供更加综合和丰富的交互体验. 从硬件方面分析, (1) 图形处理器(graphic processing unit, GPU)的算力支持. GPU因其强大的并行处理能力, 成为AI计算的首选硬件, 尤其在深度学习模型的训练和推理中表现突出. (2) 专用芯片的兴起. 除了GPU, 张量处理单元(tensor processing unit, TPU)、现场可编程门阵列(field programmable gate array, FPGA)和语言处理单元(language processing unit, LPU)等专用芯片也应运而生. 这些芯片针对特定算法进行了优化, 能够满足不同场景下的计算需求. (3) 神经网络处理单元(neural network processing unit, NPU)的应用. NPU在AI运算加速领域, 尤其在数据中心和终端设备中, 体现出高算力和低能耗比方面的优势.

3 医学大模型的挑战

大模型是数理逻辑推理和语言逻辑推理的完美结合. 医学领域关乎人的生命健康, 也是数据密集型的行业. 通过电子病历(electronic medical record, EHR)、计算机化医嘱录入

(computerized physician order entry, CPOE)、图像存档通信系统(picture archiving and communicate system, PACS)、临床策支持系统(clinical decisions support system, CDSS)和实验室信息系统(laboratory information system, LIS)等系统收集了大量的多模态的医疗服务数据，还有公共卫生和科学研究领域的相关数据以及各种网络平台的医学相关信息，卫生健康大数据正在以几何数量级增长，这些数据语料有助于大模型的开发应用. MedFound诊断大模型通过1760亿参数模型整合多模态医疗数据，在常见病、罕见病和跨专科诊断中表现优于专业模型，其自引导策略和偏好对齐框架提升了诊断推理的准确性，临床评估显示在病历总结、诊断修正等场景中显著辅助医生决策[21]. CHIEF模型是一款临床组织病理学成像评估基础模型，能够对源于胃、食道、肺、乳腺、前列腺、结直肠、肾、脑和肝等组织的19种癌症进行诊断，检测准确率接近94%[22]. LEDAP模型利用了基于LLM的生物文本特征编码来预测药物-疾病关联、药物-药物相互作用和药物-副作用关联. LEDAP在与其他流行的DBA分析工具相比时展示了其显著的竞争力[23]. 多模态深度学习模型EPBDXDNABERT-2使用包含690个ChP-seg实验结果的染色质免疫沉淀测序(ChP-Seg)数据进行训练, EPBDXDNABERT-2显著提高了660多个TF-DNA的预测，揭示了在全基因组关联研究中发现的与疾病相关的非编码变异的机制[24]. SSPEC导诊大模型相比人类导诊, 在事实性、安全性、共情能力均展现出明显优势, 在真实应用场景中降低了11.2%的重复沟通和5.4%的医患冲突比例[25]. 医学大模型在医疗健康领域的应用会越来越广泛, 涵盖了从临床辅助决策、医学影像诊断到药物研发、健康管理等多个方面，为提升医疗服务水平、优化医疗资源分配和改善患者健康体验提供了有力支持.

医学大模型发展还面临挑战, 主要在: (1) 大模型的适用性需要不断提升. 受到数据/语料的质量、完整性和科研论文的可得性，以及大模型快速迭代乃至开源与否的选型影响, 不同的平台所开发的行业应用模型的适用性还需要不断提升. 医学大模型需要专业的语料进行训练，获得高质量的医学论文需要付费，大模型公司如何支付这些费用及可持续的互惠机制建立，同时需要提出的是，以前的学术论文更多用于同行交流和学术水平提升，现在作为语料意义上的“种子”(seed)或者资本，必然会产生价值重估的问题. 开展行业大模型的成熟度评估显得尤为重要，这样会形成大模型开发应用评估的闭环, 提升大模型开发应用质量和效能. (2) 大模型开发应用需要业务和技术团队协作攻关. 医学大模型开发应用不仅需要人工智能专家的智慧，也需要卫生健康技术专家的合作，对专家的数学统计、人工智能、计算机技术、外语能力和专业知识需求很高, 因此, 需要利用大型平台公司、大型医院和大学科研机构的力量，加强创新复合型人才队伍的培养. (3) 及时应对大模型可能带来的算法偏见、模型幻觉、数据隐私泄露等风险. 如何提高算法的透明度和可解释性，如何提升语料的覆盖面和质量，确保智能化、规范化、可及性和公平性, 从语言智能到行为智能, 再到伦理智能, 形成负责任的人工智能还有较长的路要走.

4 医学大模型的发展

医学大模型的发展可能要更加关注卫生健康行业和人工智能大模型的技术整合. 网络平台和卫生健康行业信息平台(或系统)产生的、实时、多模态语料，进入大模型就产生要素意义的价值重估，也成为模型适用性、价值性的基础. 因此，优先推进医学术语标准化、平台互通共享和数据的综合集成能力等，产学研用机构联合攻关，形成较为完整的多模态大模型开发应用生态链. 大模型在行业中应用的扩展和深度可以从智能化、标准化、情感化和成本效益比等4个维度进行分析，从替代时序上看，越是智能化的越容易替代非智能化的业务，越是标准化的越容易替代非标准化的业务，越是成本效益比高的越容易替代成本效益低的业务，越是情感化的相对难以替代或者替代的时序上要延后一些. 大模型的逻辑在于将空间复杂性和时间复杂性压缩得更小，类似人脑思考一样. 医学大模型的发展使走向通用医学AI成为可能, 它可以通过少量或无须标注数据, 灵活应对不同的医疗需求. 通过在大规模、多样化的数据集上进行自监督学习(supervised learning)、动态任务指定(dynamic task specification)、多模态输入与输出(multimodal inputs and outputs)和医学知识的运用等，能够理解和整合来自影像、电子健康记录、组学、实验室结果等多种数据类型，生成详细的诊断报告、治疗建议甚至蛋白质设计方案[26]，在部分诊疗业务AI应用的基础上, 产生AI智能体医生, 甚至AI智能体医院, 实现创新、信任、普惠、赋能的卫生健康行业应用. 总之，AI不断提升迭代的过程中，人类自身也在不断的学习和提升过程中，呈现“双螺旋”上升，科技创新为人类社会发展赋能，人类社会发展为科技进步提供了更加广泛的空间，更精准、更有效、更便捷的生命健康赋能技术未来可期.

推荐阅读文献

1 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv:1706.03762

2 Islam S, Elmekki H, Elsebai A, et al. A comprehensive survey on applications of transformers for deep learning tasks. Expert Syst Appl, 2024, 241: 122666

3 Sathyan A, Weinberg A I, Cohen K. Interpretable AI for bio-medical applications. Complex Eng Syst, 2022, 2: 18

4 Dutta D, Chetia D, Sonowal N, et al. State-of-the-art transformer models for image super-resolution: techniques, challenges, and applications. arXiv:2501.07855

5 National Health Commission, National Administration of Traditional Chinese Medicine, National Disease Control and Prevention Administration. Reference Guide for AI Application Scenarios in the Health Industry, 2024 (in Chinese). https://www.nhc.gov.cn/guihuaxxs/c100133/202411/3dee425b8dc34f739d63483c4e5c334c.shtml [国家卫生健康委员会, 国家中医药管理局, 国家疾病预防控制局. 卫生健康行业人工智能应用场景参考指引, 2024. https://www.nhc.gov.cn/guihuaxxs/c100133/202411/3dee425b8dc34f739d63483c4e5c334c.shtml]

6 Moon J, Posada-Quintero H F, Chon K H. A literature embedding model for cardiovascular disease prediction using risk factors, symptoms, and genotype information. Expert Syst Appl, 2023, 213: 118930

7 Wong I N, Monteiro O, Baptista-Hon D T, et al. Leveraging foundation and large language models in medical artificial intelligence. Chin Med J, 2024, 137: 2529–2539

8 García-Herreros S, López Gómez J J, Cebria A, et al. Validation of an artificial intelligence-based ultrasound imaging system for quantifying muscle architecture parameters of the rectus femoris in disease-related malnutrition (DRM). Nutrients, 2024, 16: 1806

9 Carroll A N, Storms L A, Malempati C, et al. Generative artificial intelligence and prompt engineering: a primer for orthopaedic surgeons. JBJS Rev, 2024, 12

10 Posedaru B S, Pantelimon F V, Dulgheru M N, et al. Artificial intelligence text processing using retrieval-augmented generation: applications in business and education fields. Proc Int Conf Business ExCellence, 2024, 18: 209–222

11 Sonnenburg A, van der Lugt B, Rehn J, et al. Artificial intelligence-based data extraction for next generation risk assessment: is fine-tuning of a large language model worth the effort? Toxicology, 2024, 508: 153933

12 Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. arXiv.2022.2203.02155

13 Lewis F L, Vrabie D, Vamvoudakis K G. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. Control Systems IEEE, 2012, 32: 76–105

14 Liu F, Zhang T, Dai W, et al. Few-shot adaptation of multi-modal foundation models: a survey. Artif Intell Rev, 2024, 57: 268

15 Ferrag M A, Alwahedi F, Battah A, et al. Generative AI in cybersecurity: acomprehensive review of LLM applications and vulnerabilities. Internet Things Cyber-Phys Syst, 2024, (5): 1–46

16 Olivo C, Santin A O, Viegas E K, et al. Towards a reliable spam detection: an ensemble classification with rejection option. Cluster Comput, 2025, 28: 024

17 Yee J, Rosendahl C, Aoude L G. The role of artificial intelligence and convolutional neural networks in the management of melanoma: a clinical, pathological, and radiological perspective. Melanoma Res, 2024, 34: 96–104

18 Butunoi B P, Stolojescu-Crisan C, Negru V. Blood glucose prediction in type 1 diabetes based on long short-term memory. In: Advances in Computational Collective Intelligence. ICCCI 2024. Communications in Computer and Information Science. Cham: Springer, 2024, 2166: 31

19 Van Booven D J, Chen C B, Malpani S, et al. Synthetic genitourinary image synthesis via generative adversarial networks: enhancing artificial intelligence diagnostic precision. J Pers Med, 2024, 14: 703

20 Wang Y, Zhang B, Ma J, et al. Artificial intelligence of things (AIoT) data acquisition based on graph neural networks: a systematical review. Concurrency Computation, 2023, 35: e7827

21 Liu X, Liu H, Yang G, et al. A generalist medical language model for disease diagnosis assistance. Nat Med, 2025, 31: 932–942

22 Wang X, Zhao J, Marostica E, et al. A pathology foundation model for cancer diagnosis and prognosis prediction. Nature, 2024, 634: 970–978

23 Zhang Q, Ding K, Lv T, et al. Scientific large language models: a survey on biological & chemical domains. ACM Comput Surv, 2025, 57: 1–38

24 Kabir A, Bhattarai M, Peterson S, et al. DNA breathing integration with deep learning foundational model advances genome-wide binding prediction of human transcription factors. Nucleic Acids Res, 2024, 52: e91

25 Wan P, Huang Z, Tang W, et al. Outpatient reception via collaboration between nurses and a large language model: a randomized controlled trial. Nat Med, 2024, 30: 2878–2885

26 Moor M, Banerjee O, Abad Z S H, et al. Foundation models for generalist medical artificial intelligence. Nature, 2023, 616: 259–265

Summary for “医学大模型的构建: 技术、挑战与发展”

Building medical large language models: technology, challenge, and development

Yuefeng Li & Wei Zhao*

National Health Commission Statistical Information Center, Beijing 100044, China
* Corresponding author, E-mail: zhaowei@nhc.gov.cn

The medical large language models are reshaping the diagnosis, treatment, management, and education system of the medical industry through technological integration and scenario innovation. Its core value lies in improving the accuracy of diagnosis and treatment, reducing medical costs, and achieving universal healthcare. This article analyzes the technology, challenges, and trends of constructing medical large language models, and proposes feasible development paths. The construction of medical large language models is a systematic and complex process involving the comprehensive application of key technologies, computing platforms, and medical professional terminology.

This article is based on the ideas of modeling, automation, and standardization. Through the comprehensive application of technologies such as data, models, algorithms, and tools, it achieves multi-modal functions similar to the human brain, such as memory, understanding, reasoning, and generation. This paper deeply studies the major domestic and foreign large language models and their applications, including corpus collection, corpus classification, modeling analysis and model application. Data technology is the foundation of medical large language model construction, including data collection and integration, data cleaning and processing, data analysis and modeling, value creation, and digital assetization.In addition to the classic architecture of Transformers, there are also new architectures such as the Mixture of Experts (MoE), Mamba model, Brain-Inspired Intelligence, and temporal large models for modeling technology. The key to modeling technology lies in selecting appropriate neural networks (or multiple neural networks in series or parallel) based on Transformer architecture, and optimizing gradients and parameters accordingly.

The second is the mutual nesting of various technologies. The medical large language models utilize algorithmic techniques rooted in machine learning, implemented via neural network codecs. The neural networks in codecs can adopt four types of patterns: single neural networks, multiple neural networks in parallel, multiple neural networks in series, and hybrid. The operation of medical large language models requires powerful platforms and computing technology support. From a software perspective,

Software strategies include knowledge enhancement, cross-end proactive services, adaptive routing, centralized task planning, and multi-agent collaboration, underpinned by software and algorithmic optimization.

The application of medical large language models enhances medical service quality, optimizes resource allocation, and improves patient experiences However, the development of medical large language models still faces challenges: firstly, the applicability of medical large language models requires continuous refinement. The second is that the development and application of medical large language models require collaboration between business and technical teams to tackle challenges. The third is that risks such as algorithmic bias, model hallucinations, and data privacy breaches must be proactively addressed.

In summary, as AI evolves iteratively, humans concurrently learn and adapt. Technological innovation empowers our societal progress, while our societal development broadens horizons for technological advancement.

The future of more accurate, effective, and convenient life and health empowerment technologies is promising.

medical large language models, technology, challenge, development

doi: 10.1360/CSB-2025-0346