

## 人工智能在癌诊断和治疗中的应用进展

杜明月, 李学广, 左珊如, 陈紫淇, 周军华, 厉浩, 贺权源\*

(湖南师范大学 医学院 模式动物与干细胞生物学湖南省重点实验室, 湖南 长沙 410013)

**摘要:**人工智能(AI)以其远超人类的大数据整合分析能力,在病理图像分析、分子标志物发现、临床决策支持和药物开发与用药推荐等众多领域获得了广泛应用。本文总结最近几年来AI辅助癌精准诊疗的最新应用成果,分析该领域面临的主要挑战,预测未来的发展趋势。

**关键词:**人工智能;癌诊断;深度学习;精准医疗

中图分类号:R446.9 文献标志码:A

DOI:10.16352/j.issn.1001-6325.2022.08.1297

## Progress of artificial intelligence application in cancer diagnosis and treatment

DU Ming-yue, LI Xue-guang, ZUO Shan-ru, CHEN Zi-qi, ZHOU Jun-hua, LI Hao, HE Quan-yuan\*

(Key Laboratory of Model Animals and Stem Cell Biology in Hunan Province, School of Medicine,  
Hunan Normal University, Changsha 410013, China)

**Abstract:**The artificial intelligence (AI) technology with supreme capacity of big data integration and management have been used in digital diagnosis and clinical treatment, which have greatly facilitated the development of clinical oncology and precision medicine. AI technology is far superior to human beings in big data integration and analysis, so has been widely applied in pathological image analysis, molecular marker discovery, clinical decision support, drug discovery and recommendation and so on. In recent years, many remarkable achievements have been made in the field. This review discussed the latest applications of AI-assisted precision cancer diagnosis and treatment, highlighted emerging challenges, and predicted the development trend in the future.

**Key words:** artificial intelligence; cancer diagnosis; deep learning; precision medicine

目前,人工智能(artificial intelligence, AI)已被广泛应用于各种主要癌(如乳腺癌、宫颈癌、卵巢癌、脑癌、肺癌、皮肤癌、胃癌和肝癌等)的风险预测、筛查、诊断、治疗及预后的各环节。它不但极大减轻了医生的工作负担,显著提高了癌诊断的准确度,还降低了疾病治疗成本,使新一代癌诊疗变得智能化和个性化,具有重大的理论和临床应用价值。

该领域发展日新月异,近年来不断有重大成果问世。本文拟总结近3年来该领域的最新进展和成果,并对未来该领域所面临的挑战和发展趋势做出分析和预测。

### 1 人工智能概述

AI的主要目标是通过算法和计算机系统模拟

收稿日期:2021-12-01 修回日期:2022-05-05

基金项目:国家级大学生创新创业训练计划支持项目(S202010542011)

\*通信作者(corresponding author): hqyone@hunnu.edu.cn

人的认知分析能力,解决各类复杂现实问题。人工智能目前主要有3个研究领域:1)方法学研究,即以数学和计算机科学为基础的理论及算法研究;2)应用研究,即把方法学研究成果应用在各个不同领域中的研究;3)伦理学研究,研究人工智能与人类智能及社会之间的伦理关系。三者互相影响、协同发展。AI的方法学研究的主要领域是机器学习(machine learning, ML)。ML的算法则纷繁复杂,其主要分支包含集成算法、降维算法、贝叶斯算法、聚类算法、决策树算法和人工神经网络(artificial neural network, ANN)等。其中ANN在近二十年来取得了突破性进展,实现了深度学习(deep learning, DL)。DL在影像数据、序列数据和高维数据的处理等方面较其他机器学习算法有明显优势,在专家系统的构建、语音和图像识别、自然语言处理、启发式问题解决、逻辑推理和数学优化上有广泛和重要的应用。

## 2 AI在癌诊疗中应用现状

依据AI使用的数据类型和拟解决的临床问题,可大致把AI在癌诊疗中的应用分为如下几类:1)病理图像处理;2)生物标志物发现;3)临床决策支持;4)制药和转化医学。现分述如下:

### 2.1 病理数字影像的处理

在癌诊疗的各个环节都会产生大量的数字病理影像数据,这些数据是癌诊断和治疗的重要依据。数字病理影像的主要来源于:1)细胞及组织的染色涂/切片;2)放射医学影像,如计算机体层成像(computed tomography, CT)、磁共振成像(magnetic resonance imaging, MRI)、数字X线成像(digital radiography, DR);3)临床医学影像(如阴道镜照片、皮肤照片等)。传统的医学影像分析主要依靠病理医师人工依据肿瘤细胞和组织的复杂的病理学形态进行解读,这对医师的专业水平与临床经验要求极高,诊断过程耗时费力,且易受到地域、仪器精度等主客观因素影响。随着卷积神经网络(convolutional neural network, CNN)在临床病理的广泛应用,AI在病理影像处理的优异表现得到了广泛认可,已经成为了AI在临床医学应用最为成功的领域,可以实现快速癌诊断和疾病分层。近年来,通过不断提高影像数据的质量和增加其数量,优化深度学习算法,

DL模型在各类癌(如皮肤癌、肺癌)的识别和诊断的准确度可以达到或者超过专业医师的水平<sup>[1,2]</sup>。与此同时,AI病理影像处理还拓展到了肿瘤的转移能力预测<sup>[3]</sup>、患者术后预后结局预测<sup>[4]</sup>、手术的辅助决策及用药<sup>[5]</sup>和免疫治疗反应预测等领域<sup>[6]</sup>。

### 2.2 生物标志物的发现

癌生物标志物(cancer biomarkers)是在血液、其他体液或组织当中发现的与临床表型相关,能客观反映癌病理过程、或治疗效果的关键指标。根据具体应用,生物标志物主要有三种类型:预测型、预后型和诊断型。从类型上可分为遗传、转录组、表观遗传学、蛋白质组学和代谢组学生物标志物。癌生物标志物的发现是开发新型癌疗法的关键,也是精准医疗临床实践中的关键要素。新的生物标志物的测定,离不开对各组学数据进行挖掘。很多机器学习算法很早就成为了各类组学数据分析流的核心,如动态规划、聚类算法、贝叶斯算法、主成分分析(principal component analysis, PCA)。近年来利用人工智能整合各组学数据发现新癌标志物也取得较大进展。如多组学图形集成算法,该算法集成了癌患者样本提取的DNA甲基化、基因表达水平、细胞内蛋白质的相互作用以及基因突变等数据。用这些数据训练的深度学习模型可预测导致癌发展的模式和分子原理,从而发现165个新的癌基因<sup>[7]</sup>。DrBioRight是一种手机语言识别程序,医师通过与其简单对话就可以进行组学数据的分析和探索,大大降低数据分析的门槛,加速了生物标志物的发现<sup>[8]</sup>。同时,深度学习也在液体活检领域取得突破。如用33种不同类型癌的18 116例肿瘤样本的基因组图谱训练的机器学习模型,通过分析血液中微生物DNA类型就可以判断受试者(包括早期癌患者)是否患有癌以及癌类型癌<sup>[9]</sup>。一种名为DELFI的机器学习算法通过LASSO logistic regression比较血液中循环的癌细胞脱落的DNA碎片(cell-free deoxyribonucleic acid, cfDNA)片段的大小、数量以及在基因组区域的分布,准确区分肺癌患者和非肺癌患者,发现新的癌标志物<sup>[10]</sup>。

### 2.3 临床决策支持

目前,癌临床治疗决策已有一些公用的指南规范。如美国国立综合癌网络(national comprehensive cancer network, NCCN)的癌治疗指南(NCCN guild-

line), M. D. Anderson 中心的癌治疗算法 (cancer treatment algorithms) 等, 它们依据不同癌的类型和特点, 设计了类似决策树的流程, 提供了一般性的癌诊疗的决策意见。虽然这些指南每年都在更新, 内容也在不断充实, 但医师当面临复杂的临床情况时, 这类指南仍显不足。开发应用 AI 模型, 综合各类大数据, 辅助医师进行预后的预测, 选择治疗手段是 AI 医学研究的热点。如运用 AI 模型对癌患者各类临床数据进行整合, 预测卵巢癌患者对铂基类药物的应答<sup>[11]</sup>, 对肝癌患者术后生存情况进行预测, 减少肝癌术后肿瘤复发率等<sup>[12]</sup>。虽然这些研究给出 AI 辅助临床决策的范例, 但其有效性和可靠性仍然需要进一步验证。事实上, 目前要实现可移植好、可靠性高、覆盖面广的癌个性化 AI 决策支持系统仍有一定困难。其中一个著名的例子就是 2011 年启动的 IBM 的沃森肿瘤学 (Watson for Oncology) 项目, 其目标是开发能对数十种癌进行个性化诊疗的专家系统。IBM 与多家顶级癌中心合作, 在花费了数年及数亿美元后, 该系统仍不能提供安全和准确的诊疗预测结果<sup>[13]</sup>。其主要的原因因为目前癌临床数据的完备性、可靠性和健壮性还相对较低, 而大大制约了 AI 临床决策系统的开发。

## 2.4 药物开发与用药推荐

虽然 AI 的各类算法在药物开发中早有应用, 但直到 2016 年深度学习引入制药行业后, AI 对制药行业的颠覆性潜力才逐渐被意识到。由于几乎所有分子皆可用序列/类语言的形式精确表示, 自然语言处理 (natural language processing, NLP) 技术中的 RNN、transformer 等便自然而然地迁移到制药领域中来了。此外, 化合物分子也可以图的形式来表示 (原子看作顶点, 化学键当成边), 因此, 把图论和卷积论相结合图卷积神经网络也成为了一种分析药物强大工具<sup>[14]</sup>。如 2020 年采用 ANN 技术开发的 Alpha Fold2 在蛋白质结构预测上获得了惊人的进步, 其蛋白质结构预测的能力已经达到了与实验方法相媲美的程度。这不但说明 AI 技术完全有能力准确预测各类分子的结构, 同时也意味着 AI 技术具备从头设计分子和精确模拟药物与靶标之间的相互作用的潜力<sup>[15]</sup>。以利用药物数据库为训练集, 训练有向消息传递神经网络模型 (directed-message passing neural network), 并用其对天然化合物

数据库中约 1 亿种分子进行评分, 最终发现了 8 种抗生素, 其中 2 个具有广谱抗菌效果。这一研究充分展示了 AI (特别是神经网络技术) 加速药物开发的潜力<sup>[16]</sup>。除了药物开发, AI 在优化癌治疗用药策略上也取得了显著进步。如 comboFM 人工智能算法可以精确预测不同抗癌药物的组合是否可以对癌细胞形成联合杀伤作用, 从而提出更好的解决耐药性的策略, 减少单个药物的用药剂量, 减轻药物对患者的毒副作用<sup>[17]</sup>; 以及一种名为“DrugCell”的可见神经网络系统, 该系统可为不同肿瘤的患者推荐最佳药物组合, 提供优化的联合用药方案<sup>[18]</sup>。

## 3 面临的主要挑战

虽然 AI 在癌临床诊疗领域的应用成果颇丰, 但仍存在一些固有的问题急需解决。首先, 几乎在所有 AI 应用领域都存在数据缺乏, 且质量不高的问题。这一方面是由于临床数据所固有的分散性、异质性、低完备性和隐私性, 其收集、分类、整合和标准化的难度非常之大; 另一方面是由于临床数据含有大量人工解读结果。由于医生的经验差异、诊疗手段的变迁而包含隐性、无法矫正的数据偏差, 其本身含有一定错误, 这可能严重制约 AI 模型的预测的准确性, 很难保证模型预测效果的可迁移性和可重复性。其次, 需要解决“黑盒”问题, 即需要提高 AI 模型工作流程的透明度和预测结果的可解释性。只有让医生了解预测结果背后的逻辑和证据, 才能最终使 AI 被医学界认可。另外, 目前业界还没有建立对 AI 模型进行客观评价的标准, 缺乏贯穿开发、测试、应用全流程的 AI 模型研发的平台和机制, 很多发表的 AI 模型只得到了非常有限的检验反馈。最后, 对于希望回答复杂诊疗问题的人工智能宽应用, 如何整合各类数据和模型, 获得可靠的预后预测结果, 提出有效的个性化精准治疗方案还有很长的路要走。

## 4 前景与展望

总体来看, 医学 AI 系统在癌诊断领域的应用已经取得了长足的进步, 正在向更为复杂的癌预后和转化领域渗透。AI 辅助抗癌药物的研发可能处于高速发展的前夜。多种癌新技术 (如单细胞测序技术, 免疫疗法、液态活检) 为 AI 的应用提供新的研

究数据和场景。一些新数据类型如卫星图像数据<sup>[19]</sup>、互联网搜索记录<sup>[20]</sup>和可穿戴设备<sup>[21]</sup>也被引入到癌的群体风险预测和预防中,为 AI 在癌防控中的应用提出了新的探索方向。

为突破现有的数据限制,有研究者提出了一些倡议来简化和统一数据收集过程以实现数据标准化<sup>[22]</sup>。2021 年有人提出了一种名为 Swarm Learning 的临床数据合作和共享模式。其基于标准化 AI 引擎、分布式机器学习和区块链许可技术实现不同医疗机构之间医疗数据的去中心化整合。作为一个去中心化的学习系统,其有望取代目前跨机构医学研究中的中心化数据共享模式,为每个参与者提供一个平等的参与环境,帮助相关人员在 全球范围内更容易获得丰富的医学经验<sup>[23]</sup>。这些前瞻性研究为解决本领域关键挑战提供新的解决方案(图 1)。

然而,由于限制因素在短期内无法得到根本性改善,在可预测的 10~20 年,人工智能在癌诊疗中

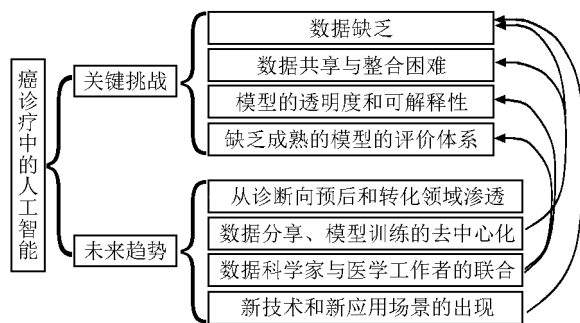


图 1 人工智能在癌诊疗中的关键挑战及未来趋势

Fig 1 The key challenges and future directions of AI applications in cancer diagnosis and treatment

的应用将集中在有较为完备的数据,针对特定癌和具体的诊疗问题,并与临床证据紧密联系,因而能获得较好的应用效果的窄任务。因此,谨慎选择、并专注于一些窄任务,由数据科学家和医学工作者建立研究联盟,以突破训练数据集的数量和质量的瓶颈,将是未来 5~10 年的主要研究模式。

## 参考文献:

- [1] Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542:115-118.
- [2] Yoo H, Kim KH, Singh R, *et al.* Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs [J]. *JAMA Netw Open*, 2020, 3: e2017135. doi: 10. 1001/jamanetworkopen. 2020.17135.
- [3] Lu MY, Chen TY, Williamson DFK, *et al.* AI-based pathology predicts origins for cancers of unknown primary [J]. *Nature*, 2021, 594:106-110.
- [4] Tahmassebi A, Wengert GJ, Helbich TH, *et al.* Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients[J]. *Invest Radiol*, 2019, 54:110-117.
- [5] 王培培,吴斌.人工智能技术在结直肠癌全程管理中的应用[J].*基础医学与临床*, 2020, 40:1570-1573.
- [6] Hashimoto DA, Rosman G, Rus D, *et al.* Artificial intelligence in surgery: promises and perils [J]. *Ann Surg*, 2018, 268:70-76.
- [7] Schulte-Sasse R, Budach S, Hnisz D, *et al.* Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms[J]. *Nat Mach Intell*, 2021, 3:513-526.
- [8] Li J, Chen H, Wang Y, *et al.* Next-generation analytics for omics data[J]. *Cancer Cell*, 2021, 39:3-6.
- [9] Poore GD, Kopylova E, Zhu Q, *et al.* Microbiome analyses of blood and tissues suggest cancer diagnostic approach[J]. *Nature*, 2020, 579:567-574.
- [10] Cristiano S, Leal A, Phallen J, *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer [J]. *Nature*, 2019, 570:385-389.
- [11] Yu KH, Hu V, Wang F, *et al.* Deciphering serous ovarian carcinoma histopathology and platinum response by convolutional neural networks[J]. *BMC Med*, 2020, 18:1-14.
- [12] Saillard C, Schmauch B, Laifa O, *et al.* Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides [J]. *Hepatology*, 2020, 72:2000-2013.
- [13] 王玉峰,蔡文杰. IBM 沃森成败录[J]. *中国工业和信*



- 息化, 2020; 78-83. doi: 10.19609/j.cnki.cn10-1299/f.2022.zl.017.
- [14] Yang X, Wang Y, Byrne R, *et al.* Concepts of artificial intelligence for computer-assisted drug discovery [J]. Chem Rev, 2019, 119:10520-10594.
- [15] Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596:583-589.
- [16] Stokes JM, Yang K, Swanson K, *et al.* A Deep learning approach to antibiotic discovery [J]. Cell, 2020, 180: 688-702.
- [17] Julkunen H, Cichonska A, Gautam P, *et al.* Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects [J]. Nat Commun, 2020, 11. doi:10.1038/S41467-020-19950-Z.
- [18] Kuenzi BM, Park J, Fong SH, *et al.* Predicting drug response and synergy using a deep learning model of human cancer cells [J]. Cancer Cell, 2020, 38:672-684.
- [19] Bibault JE, Bassenne M, Ren H, *et al.* Deep learning prediction of cancer prevalence from satellite imagery [J]. Cancers (Basel), 2020, 12:3844. doi:10.3390/cancers12123844.
- [20] Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results [J]. J Oncol Pract, 2016, 12:737-744.
- [21] Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic review [J]. JMIR Mhealth Uhealth, 2020, 8: e22006. doi:10.2196/22006.
- [22] Wilkinson MD, Dumontier M, Aalbersberg LJ, *et al.* The FAIR guiding principles for scientific data management and stewardship [J]. Sci Data, 2016, 3:1-9.
- [23] Aminoshariae A, Kulild J, Nagendrababu V. Artificial intelligence in endodontics: current applications and future directions [J]. J Endod, 2021, 47:1352-1357.

## 本刊稿件格式要求(3)

### 3.1 参考文献:应为正式发表的论文或书籍。

#### 3.1.1 引用期刊格式:

[文献序号]. 著者. 文献题名[J]. 刊名, 出版年, 卷号:起页-止页.

-外国人名书写格式:姓在前/名在后、姓全称/名缩写;3位以内作者须写全,3位以上作者列前3位,后加“等”或“*et al.*”。

-外文刊名书写格式:按照 PubMed 格式规范缩写。

例:[1] Feinstone SM, Kapikian AZ, Purcell RH, *et al.* Transfusion-associated hepatitis not due to viral hepatitis type A or B[J]. N Engl J Med, 1975, 292:767-770.

[2] 杨汀, 王辰, 庞宝森, 等. 比较地塞米松和色甘酸钠对哮喘豚鼠气道重建的作用[J]. 基础医学与临床, 2004, 24:633-636.

#### 3.1.2 引用书籍格式:

[文献序号]. 析出文献著者. 析出文献题名[M]//专著著者(或编者). 专著书名. 版次(第1版不著录). 出版地:出版者, 出版年:析出文献起页-止页.

例:[1] 章静波. 癌细胞行为的分子基础[M]//柳惠图, 王永潮, 桑建利. 分子细胞生物学. 北京:高等教育出版社, 2012:501-505.

#### 3.1.3 引用电子期刊:需注明 doi 号。

#### 3.1.4 引用毕业论文:需注明“学校所在城市:学校名称,出版年:所引用起页-止页。”。