

大健康与新医疗

BIG DATA Health
and New Medical

2023 年
第 21

上海科学技术情报研究所
上海市前沿技术发展研究中心
技术与创新支持中心(TISC)

人工智能在肿瘤诊断中的研究和应用进展

编者按

医学领域中，人工智能在药物研发、影像学诊断、病理诊断、临床治疗决策、疾病预后评估等方面均表现出重要的辅助价值。目前，恶性肿瘤发病率在全球呈逐年上升趋势，早诊早治和个体化精准治疗对于改善恶性肿瘤患者的临床预后至关重要。人工智能可协助恶性肿瘤的筛查、早期诊断、病理组织学诊断及分类，提高肿瘤的诊断准确率及效率，并可协助医生完成恶性肿瘤患者的临床诊疗决策和预后评估。本期《大健康与新医疗》简报将从研究进展、应用现状、行业动态等方面关注人工智能在肿瘤诊断中的研究和应用进展。

目 录

目 录	2
研究进展	3
癌症研究和精准肿瘤学中的人工智能	3
AI 机器学习在癌症临床诊断、预后和治疗中的进展	5
人工智能在临床肿瘤学中的研究现状及展望	8
应用现状	9
人工智能用于肺癌人群筛查	9
人工智能在甲状腺针吸细胞病理诊断中的应用	11
行业动态	13
微软联合建立 AI 癌症识别模型，训练数据集包含 400 万张图像	13
新标准	14
欧洲为癌症领域人工智能制定新标准	14

研究进展

癌症研究和精准肿瘤学中的人工智能

人工智能 (AI) 涉及癌症研究领域各个方面，包括基础研究、转化研究和临床研究。从实验室到临床，人工智能工具用于精准肿瘤学最终将加强患者管理和改善临床终点。大多数现代成功的人工智能系统都包含某种形式的深度学习。在生物医学研究文献中，普遍的共识是“人工智能”一词可以应用于深度学习方法（涉及具有数百万或数十亿参数的模型），而不是简单的统计模型。2023年5月发表在《NPJ Precis Oncol》的一篇文章，针对癌症研究领域人工智能 (AI) 各个方面的最新研究和进展进行了评论。

1、数字化与医学图像处理

医学，尤其是肿瘤学，图像数据非常丰富。例如，皮肤肿瘤可以通过视觉诊断，并且可以通过数码照片进行识别；多种肿瘤类型，例如消化道肿瘤，都是通过内窥镜诊断的；乳腺癌和许多其他肿瘤类型可通过放射成像进行筛查和检测。

肿瘤通过基于染色组织图像分析的组织病理学过程来做出明确诊断，放射

学图像也常用于癌症分期和治疗反应评估。从本质上讲，肿瘤学对图像分析的高度依赖使得此类图像的采集、解释和分析在癌症检测、表征和后续过程的所有阶段都至关重要。上个世纪，医学成像处理和存储领域发生了一场关键的革命，从放射学中的模拟探测器过渡到数字探测器。此外，病理科在肿瘤样本数字化方面的努力也得到了广泛的扩展。这种向数字化的转变为图像数据的存储、共享和使用开辟了令人兴奋的新机遇。借助数字图像数据，可以应用强大的人工智能计算方法，实现对曾经仅由医学专家主观评估的数据进行高通量分析和利用。因此，医学成像的数字化代表了一项重大突破，并有望在肿瘤学领域取得重大进展。

2012 年之前，基于计算机的图像分析是一项艰巨的挑战，高度受限于非常狭窄的领域和数据集。2012 年，AlexNet 作为第一个著名的卷积神经网络 (CNN) 的突破，深度学习已成为几乎所有图像分析任务的首选方法，因此在肿瘤学领域绝大多数人工智能解决方案都与数字图像分析有关。人工智能可以自动化人类任务，例如内窥镜视频、放射学图像或组织病理学幻灯片上的癌症检

测。目前有许多项目已获得欧洲、美国和其他市场监管部门的批准并用于患者的日常护理，然而，人工智能可以通过产生能够直接预测生存或治疗反应的基于图像的生物标志物来超越这些应用，从而从图像数据中提取比以前更多的信息。此类生物标志物可以整合到已在临床护理中使用的预后或预测模型中，从而充当综合生物标志物。

2、数字工具与语言文本处理

文本是人与人之间以及人与机器之间通信的通用接口。电子病历、患者报告的结果以及医疗保健提供者和患者之间的沟通都离不开文本。在肿瘤学中，这意味着诊断和临床工作流程的很大一部分会生成文本伪影，医生经常需要花费大量时间来阅读描述患者既往病史或诊断报告的文本文档，直到最近这一巨大的资源才开始被数字工具开发。

自然语言处理 (NLP) 领域开发了多种基于计算机的文本挖掘方法，但目前仍然没有一种方法足够稳健、有效且可推广。总结多页医疗报告、语言翻译或写作风格的转换等任务被认为非常复杂。由开创性论文“Attention is All You Need”发起的变压器神经网络的出现，刺激了 NLP 的指数级增长。通过使用具

有数百或数千亿参数的转换器（称为大型语言模型，LLM）并使用大量数据（主要是所有可用的互联网文本）对其进行训练，几乎可以在任何语言理解或创建任务中实现人类专家级的性能。在一个意想不到的历史转折中，LLM 也无意中表现出了非凡的逻辑推理能力，这项技术对社会各个方面（包括癌症研究和肿瘤学）的变革性影响不容小觑。就像互联网，LLM 的引入可能会导致该领域的整体生产力和步伐呈指数级增长，但这并不意味着癌症研究人员或肿瘤学家将被淘汰，相反，在此过程中对人工编排的需求将继续存在。

3、利用人工智能增强基因组信息

新一代测序 (NGS) 的使用彻底改变了肿瘤学领域的研究和药物开发，带来了无与伦比的精确度和成功。AI 工具为高通量基因组数据（包括全基因组、全外显子组或全转录组分析）的利用提供了巨大的可能性，不仅有望增强临床实践，而且还能获得有关癌症生物学的宝贵见解。实施基于人工智能的基因组数据分析技术可以识别在癌症中具有预后预测意义的新基因组和转录组特征，并能够对变异和基因型-表型关联的致病性进行大规模分析。重要的是，人工智能

在癌症基因组学中的应用可以将决策支持工具提升到一个新的水平，从而实现真正的个性化护理。在基因组学中采用人工智能分析仍然需要面临一些挑战并提出解决新方法。

4、肿瘤诊疗多模态的未来

变压器神经网络的巨大成功使另一个曾经难以实现的目标变得触手可及：多模态机器学习模型。这些模型可以接受不同类型数据的同时输入，例如多种图像类型、文本与图像的组合或与基因组数据集成的图像。

很明显，肿瘤决策本质上是多模态的。肿瘤学家不会仅仅根据单一图像来制定治疗方案。相反，他们会综合考虑多条信息和相应的数据，组合不同数据类型所产生的结果不仅仅是其各个部分的总和，只有结合患者的基因组成、特定的癌症基因组改变和/或病史的图像模式才变得有意义。最近证明在多模态数据上训练机器学习模型能合理应用于肿瘤学。利用变压器神经网络，这些方法变得更加强大和通用。虽然这些方法带来了挑战，但相信它们的实施迫在眉睫。

5、循证医学和以患者为中心的护理的首要地位

利用人工智能生成可靠的科学证据，从而改善癌症研究和患者护理。医学

实践的核心涉及同理心的人际互动、基于严格科学的共同决策以及对患者偏好的考虑，这些原则指导我们进行研究，因此肿瘤学领域的人工智能必须尊重并推进循证医学和以患者为中心的护理的首要地位。此外，我们必须在人工智能系统的开发、评估和临床实施中考虑道德、数据隐私问题以及潜在的偏见。

资料来源：

[1]癌症研究和精准肿瘤学中的人工智能[EB/OL].(2023.08.21).[2023.12.07].

<https://www.cn-healthcare.com/articlewm/20230821/content-1595255.html>

[2]Perez-Lopez R, Reis-Filho JS, Kather JN. A framework for artificial intelligence in cancer research and precision oncology. NPJ Precis Oncol. 2023 May 17;7(1):43

AI 机器学习在癌症临床诊断、预后和治疗中的进展

在过去的十年里，机器学习(ML)在医学，特别是在肿瘤学领域的应用呈现出快速增长。作为一类复杂的、不均一的和普遍的疾病，癌症既引发了一系列具有挑战性的诊断问题，也提供了丰富相关数据，这使得临床肿瘤学成为机器学习的一个新兴的领域。机器学习利用大数据学习数据集的模式和结构(参见机器学习入门部分对机器学习的简要介绍)。丰富的医学成像和相关分子数据促进了机器学习的应用，将这些数据源与早期癌症检测、癌症进展监测和优化治疗

方法相关联，已成为新兴的癌症研究方向。

医学成像已经成为机械学习辅助癌症诊断的有力工具。医学成像能够实现廉价和扩展性的癌症检测、定位和监视。放射学图像以及其他图像形式，如表层图像或结肠镜检查，可用于筛查和诊断组织样本中的病理图像并进行癌症诊断，辅助决策化疗或免疫治疗的方案。

由于医学成像在临床肿瘤诊断中越来越重要，成像数据的数量往往超出病理科医生的处理能力。因此，为减少临床医生的工作量，加速对时间敏感的图像的分析，并减少临床医生的错误通过自动化方法处理医学成像势在必行。计算机视觉机械学习技术的进步已经适用于医学成像，并已经在临床肿瘤学中快速准确地分析各种成像方面显示出可观的前景。

常见的机械学习模型的建立需要完成多个步骤：（A）构建随机树以进行输入对象的初步判断，基于输入特征学习模型将进行二元判断；（B）基于初步判断，学习模型将在特征空间中生成一个边界（二维度的线或多维度的超平面）以区别不同类型的数据点；（C）回归模型综合输入特征的组合，并对特定组合输入标签；（D）学习模型模拟人类神经网络，由多层神经元组成，它

们迭代计算输入的特定组合，以预测结果，如癌症的概率；（E）学习模型构建重复性神经网络，通过对输入对象中的每个对象应用相同的神经网络层，并保持它所见预测的结果；（F）卷积神经网络使用称为过滤器神经元模块来扫描图像以寻找最终模式。早期层识别边缘等低级特征，而后期层识别肿瘤形态等高级特征；（G）通过将模式中每个元素进行比较，更新相关的内部表示方法；（H）图像神经网络以图结构数据设计的，比如相邻细胞的图。它首先对图中每个节点和边的基本特征进行编码，然后神经网络层在图中传递信息来更新节点和边的表示，然后用这些表示来预测图的标签。

相应地，构建用于癌症诊断、预后和治疗的机器学习模型则需要通过三个步骤：（A）用于癌症诊断的常见分子数据集，包括循环游离 DNA (cfDNA)、甲基化水平和片段组学。目前已通过实体瘤活检的全基因组测序、单细胞转录组学和批量 RNA 测序生成了用于癌症预后的众多分子数据集。利用分子数据进行癌症治疗方向的预测和选择是一个快速发展的领域，它结合了基础分子技术和新兴方法，如空间组学。（B）分子数据通用机器学习模型的设计。（C）针对分子

数据，调整机械学习模型的选择。

机械学习算法的研发反映了肿瘤研究领域的显著进展。然而，在机械学习算法应用于患者之前，它们需要获得监管机构的批准，这需要经过已发表的学术论文更严格的临床试验和验证的测试。因此，只有部分机械学习算法最终用于患者。其中，它们通常在几个预先设定的任务中表现良好，如检测和分诊设置，并且它们在不同的患者人群中表现出可靠性和普遍适用性。

在美国，大多数机械学习算法由食品和药物管理局(FDA)作为医疗设备进行监管。在过去十年中，超过 300 个 AI 支持的医疗设备获得了 FDA 批准，其中超过 40% 自 2020 年以来获得批准。表一展示了十余种肿瘤通过审核批准、应用于临床诊断、预测和治疗的诊断机械学习模型。

然而，机械学习发展中仍然存在局限性，缺乏高质量、多样化的评估妨碍了在患者人群中评估算法真正性能的能力。一个因素是缺乏标准的测试数据集——很难进行现场验证，也很难获得患者数据，部分原因是隐私问题和限制性数据使用协议。机械学习模型的一个有充分记录的现象是，它们可以学习设备类型和演示图形中出现的伪相关性，从而在不同患者人群中评估时导致有偏倚

的性能。此外，评估测试集通常富含阳性病例，这可以产生不平衡的比较。

前瞻性试验对于衡量适当的临床结局也很重要，而不是与独立参考文献进行简单比较。例如，如果机械学习设备用于临床诊断辅助，则应通过比较使用和不使用该设备的临床医师性能来评估，而不是单独评估该设备的预测。随机化患者队列可以最大限度地减少选择测试人群时的偏倚。此外，前瞻性试验可以捕捉部署后发生的人类- AI 交互。有必要在机械学习算法获得批准后对其进行持续性能监测，并建立上市后监测机制，以确保机械学习据称的临床优势在不同的情况下保持不变。

未来，机械学习的发展依赖于生物学数据、整合影像学和组学以及机械学习方法论的发展，新的生物学和机器学习技术正在迅速出现，这将改变机器学习应用于癌症诊断的方式，并显著提高这些模型的预测能力和临床实用性。

资料来源：

[1] Cell | AI 机器学习在癌症临床诊断、预后和治疗中的进展[EB/OL].(2023.03.14).

[2023.12.07].<https://zhuanlan.zhihu.com/p/613816621>

[2] Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell. 2023 Apr 13;186(8):1772-1791

人工智能在临床肿瘤学中的研究现状及展望

在过去的十多年间，人工智能（artificial intelligence）被广泛应用于医疗健康领域，其中也包括临床肿瘤学。得益于人工智能技术的迭代更新以及癌症诊疗过程中生成的大量数字化信息的可用性，人工智能在肿瘤学领域的研究报告和临床调查呈爆炸式增长并取得重大的理论进展。人工智能可通过实现高维多模态数据的计算整合，提供更准确和高效的诊断并定制癌症患者的个性化治疗选择，协助临床医生指导临床实践。虽然人工智能当前在肿瘤学的理论研究中取得了重大进展，但其在真正的临床实践中仍处于起步阶段，距离临床实际应用仍有较大差距。因此，加强对人工智能在临床肿瘤学领域的认识，将有助于我们把握机遇、应对挑战。

题目为“Artificial intelligence for clinical oncology: current status and future outlook”的文章，对人工智能在临床肿瘤学中的最新进展进行了总结和展望。

文章概述了近年来人工智能与临床肿瘤学研究相关的重大进展，包括人工

智能在肿瘤筛查、检测、诊断、精准分型、治疗选择、预后判断和智能随访等各环节中的作用。随着人工智能技术的发展和完善，人工智能现可利用各种医学成像（如临床照片、放射学摄影图片、病理切片等）预测各种肿块的良恶性，进而改善癌症筛查、检测和诊断，并且其诊断能力接近具有多年经验的临床专家。除了改善癌症检测和诊断外，人工智能还可以辅助传统的肿瘤学TNM（tumor-node-metastasis）分期，成为癌症患者的风险分层和预后判断的工具。此外，人工智能可以检测肿瘤的分子表征，如预测癌症的驱动基因突变情况。这些技术在肿瘤组织不足或无法获取用于基因组测序的肿瘤样本时可发挥巨大作用。重要的是，人工智能最有吸引力的应用之一是预测治疗反应的能力，这对指导个性化治疗具有重要意义。该团队基于前期基础，近期发表的研究结果，阐明了人工智能在预测胃癌化疗反应的关键性和先进性。

同时，该文章指出，实现人工智能在肿瘤学临床实践中的有利影响之前，仍须克服许多限制，主要包括，打破机构数据访问的壁垒、统一数据采集的标准、实现大量高质量数据的可用性、提高人工智能的可解释性、进行前瞻性验

证和随机对照研究等。只有真正克服这些障碍后，才有理由肯定人工智能在临床肿瘤学的应用价值。

该文基于目前的研究基础，进一步给出了人工智能在临床肿瘤学研究中的未来展望。文章认为，下一代人工智能将聚焦于利用多模态数据集的互补潜力，集成临床数据、放射学图像、病理切片和基因组学特征的智能模型，最大限度地发挥精准肿瘤学的价值。此外，另一个有希望的研究方向是设计以生物学先验知识为依据的定制神经网络框架，提供一种可解释的人工智能模型以实现更好的临床应用。更重要的是，人工智能最有效的应用是辅助而不是取代临床医生，人工智能的未来应用需要从人类与人工智能的竞争导向转向人类与人工智能的协作导向，并应探索将人工智能整合到临床医生工作流程中的最佳策略，以改善癌症诊断和治疗，同时最大限度地减少伤害。相信随着技术的发展和机构合作的深入，人工智能在临床肿瘤学中将有光明的应用前景。

资料来源：

- [1]Science Bulletin | 人工智能在临床肿瘤学中的研究现状及展望 EB/OL].(2023.03.13). [2023.12.07].<https://new.qq.com/rain/a/20230313A04A2D00>
- [2] Jiang Y, Li R, Li G. Artificial intelligence for clinical oncology: current status and future outlook. Sci Bull (Beijing). 2023 Mar 15;68(5):448-451

应用现状

人工智能用于肺癌人群筛查

肺癌仍然是发展中国家癌症相关死亡的首要原因，肺癌疾病的高死亡率和发病率需要改进和修改肺癌筛查技术。不可能对所有个体进行肺癌筛查，但至少高危人群应该进行筛查。CT 筛查可以通过常规检测高危人群中的毫米级潜在恶性肺结节来提高治愈率，2010 年，对 NLST 国家肺部筛查试验数据的评估显示，与其他高危成人相比，连续三年接受低剂量计算机断层扫描 (LDCT) 年度肺癌筛查检查的高危成人组的肺癌死亡率显著降低了 20%。目前，美国和中国是唯一在各地提供肺癌筛查的国家，欧盟委员会建议对过去 15 年内戒烟、年龄在 50 至 75 岁、有 30 包年吸烟史的当前吸烟者和戒烟者进行肺癌筛查。低剂量 CT 扫描是唯一被证明可以有效降低高危患者肺癌筛查死亡率的策略。

肺癌的诊断通常依赖于肺结节的识别，肺结节是早期发现的重要放射学指标。然而，肺癌结节的治疗可能很复杂。肺癌治疗通常需要放射治疗，并随后

导致放射引起的肺部病变。这些损伤是肺实质研究的限制因素。在此背景下，计算机辅助诊断（CAD）系统对于放射科医生来说发挥着关键作用。CAD 系统减少观察错误和假阴性率并提供影像解读的第二意见以最终诊断。通过利用先进的算法和自动分析，CAD 系统有助于从结节中提取更多信息，增强诊断过程并支持更准确的决策。如今，人工智能通过 CAD 系统在肺癌筛查中进行病灶检测和分割的实用性已得到广泛认可，并且多种 CAD 系统工具已经过验证并已上市。

机器学习通过彻底改变各种任务的准确性和效率，彻底改变了放射学。专门负责解读和诊断 X 射线、CT 扫描、MRI 扫描和超声图像等医学图像的放射科医生从机器学习算法融入其实践中获益匪浅。

机器学习在放射学中的显着应用之一是图像分类和分割。通过利用卷积神经网络等技术，放射科医生可以自动对解剖结构进行分类并识别医学图像中的异常情况。事实证明，这项技术在检测肿瘤、识别肺结节和准确分割器官方面具有非常高的价值。

此外，机器学习算法有助于质量控制和图像增强：通过减少噪声、增强细

节和标准化图像采集协议，这些算法提高了医学图像的质量。这确保了一致的解释并增强了整体诊断过程。需要强调的是，虽然机器学习算法在放射学领域显示出巨大的潜力，但它们的目的是补充放射科医生而不是取代他们。放射科医生的专业知识和临床判断对于准确的诊断和有效的患者护理仍然不可或缺。

深度学习系统在肺部研究中的应用在病变检测和筛查、风险分层、结节分割、放射基因组分析、预后预测、治疗计划（病变体积、形态、关系）和反应方面似乎是准确的，并且开始在肿瘤成像中发挥关键作用。开源图像数据集的可用性使得有效的 CAD 工具的开发和验证成为可能。

资料来源：

AI 造福癌症患者：当人工智能用于肺癌人群筛查[EB/OL].(2023.10.25).[2023.12.07].

<https://www.cn-healthcare.com/articlewm/20231025/content-1615597.html>

人工智能在甲状腺针吸细胞病理诊断中的应用

细胞病理学检查是恶性肿瘤初步诊断的关键步骤。近年来，随着计算机图像处理技术和模式识别与人工智能技术的飞速发展，人工智能被广泛应用于恶性肿瘤的诊断领域中，其在细胞病理诊断中具有良好的适用性。人工智能应用于细胞病理诊断可显著提高阅片速度、降低诊断医师的劳动强度、提高工作效率

率，避免因主观因素而产生误差，是一种具有较高应用价值的恶性肿瘤筛查技术手段。

机器学习算法评估甲状腺 FNA 标本

在迄今为止数据最大的甲状腺研究中，Range 等对 908 例病例进行 Bethesda 系统分类，此研究开发了一种机器学习算法(machine learning algorithm, MLA)，通过全视野的数字化切片(whole slide image, WSI)评估甲状腺 FNA 标本。选取 908 例有明确完整病史的甲状腺 FNA 病例中最典型的涂片扫描为 WSI，其中 799 例为训练集，109 例为测试集，按照 Bethesda 系统分类，该算法在鉴别良性和恶性结节方面的敏感性为 92.0%，特异性为 90.5%。人工和 MLA 预测恶性肿瘤的曲线下面积分别为 0.931 和 0.932。MLA 预测恶性肿瘤方面的表现与人工相当。由于细胞病理专家在极端情况下(良性或恶性诊断)表现更好，因此 MLA 可以作为 FNA 活检的补充，以帮助细化不确定的类别。

数字图像的细胞核参数测量

除此之外，很多研究试图将甲状腺 FNA 诊断主观的细胞学标准转变为定量

的分类参数，大多数此类研究集中在用于鉴别乳头状癌/滤泡性癌与良性甲状腺结节细胞核的参数上。2019 年，Chain 等研究应用数字图像衍生的核测量客观地将 PTC 与良性结节和 NIFTP 区分开来，使用 Aperio 成像系统扫描每个病例的涂片，测量每个病例 125 个细胞核的长/短径，发现 PTC 的核面积大于良性，NIFTP 核面积小于 PTC。核延伸率显示相似的结果，但各组间数据重叠率更大。2019 年，Yashaswini 等使用 Bethesda 系统对 FNA 涂片中的甲状腺病变进行分类，并评估核参数在提高甲状腺恶性肿瘤预测中的意义，使用 Aperio 计算机软件对 81 例经组织学诊断确诊的甲状腺 FNA 病例进行核形态测量，对每个病例的 100 个细胞核分析了 8 个相关参数，结果表明恶性组最小核径、最大核径、核周长和核面积均高于非肿瘤组和良性组，证明了通过计算机图像分析进行的核形态测量可以用作辅助诊断工具。

总的来说，尽管在一些研究中与核形状和染色质特征相关的其他因素具有预测性，但核面积的测量似乎是鉴别 PTC 和良性结节的最可靠方法。对于滤泡癌，核大小的变化可能是一个重要的鉴别特征。

微观特征描述区分

然而，有研究未直接评估细胞形态学特征，而是根据 FNA 报告中包含的微观描述使用机器学习算法区分不同病例。2019 年，Maleki 等评估了机器学习算法 (support vector machine, SVM) 使用各种细胞形态学特征描述区分经典性 PTC 和 NIFTP/无包膜或淋巴管浸润的包裹性滤泡型甲状腺乳头状癌 (encapsulated follicular variant of papillary thyroid carcinoma, EFVPTC) 的性能，选取 59 例 (29 例经典性 PTC 和 30 例 NIFTP/EFVPTCs) 作为测试集，SVM 成功鉴别出经典性 PTC 与 NIFTP/EFVPTC 的概率为 $(76.05 \pm 0.96)\%$ ，检测经典性 PTC 的敏感性为 72.6%，特异性为 81.6%。这种方法可能对开发数据驱动评分系统有价值，可以指导细胞病理学和外科病理学诊断。大多数基于形态计量学的甲状腺 FNA 研究最大的不足是数据集太小，这些研究未来的探索需要更大的数据集，更专业的仪器设备。核形态测量研究通常需要人工识别和提取细胞核特征，费时费力。WSI 的使用有望加快未来自动化进程，促进 AI 的应用，但其所需存储空间巨大。此外，训练

过程要求具有强大的计算能力和庞大的训练集，为机器学习提供足够的数据，这可能容易出现数据过拟合。然而，计算能力、存储、成像和机器学习算法的进步表明，在不久的将来，该技术在解决甲状腺 FNA 标本的诊断不确定性方面可发挥重要作用，辅助人工显微镜诊断，并作为分子检测的潜在替代品。

资料来源：

刘颖，廉亚丽，董律吏，等. 人工智能在甲状腺针吸细胞病理诊断中的应用进展[J]. 癌变. 畸变. 突变，2023，35（5）：398-399

行业动态

微软联合建立 AI 癌症识别模型，训练数据集包含 400 万张

图像

微软宣布与数字病理学公司 Paige 合作，建立世界上最大的基于图像的癌症识别人工智能模型。该算法的训练数据集包含 400 万张图像。Paige 的 CEO 安迪·莫耶（Andy Moye）在接受媒体采访时表示：“对于癌症治疗来说，这是一个开创性的时刻。”

2023 年 8 月，人工智能乳腺癌筛查的首个临床试验结果出炉。研究人员比较了两种查看乳房 X 光片的方法：一种是由两名独立的放射科医生进行的，

另一种使用了一名放射科医生和人工智能为患者打分，使用的是从 1 到 10 的数字癌症风险评分系统。在后一组中，那些被打 10 分的人患癌风险最高，他们的图像也会由两名放射科医生查看。人工智能支持的模型减少了 44% 的工作量，检测出的癌症增加了 20%。

这听起来是件好事。从理论上讲，癌症发现得越早就更容易治疗，从而挽救生命。但数据并不总是支持这个结论。8 月下旬发表的一项研究，梳理了随机临床试验的文献，比较了两组人的死亡率（任何原因，不仅仅是癌症）：接受癌症筛查的人和没有接受癌症筛查的人。对于大多数常见类型的癌症筛查，他们没有发现显著差异。唯一的例外是乙状结肠镜检查，这是一种仅能观察结肠下部的结肠癌筛查。

为什么会这样的原因还不完全清楚。这可能是因为在研究中的设计缺陷。作者在分析中纳入的试验可能没有对参与者进行足够长的跟踪，无法看到差异。

另一种解释是，筛查对一些癌症的好处可能更大。例如，如果筛查早期发现致命的癌症，患者可能会获得宝贵的时间来成功治疗这种疾病。但是，如果筛查

发现了许多不会导致死亡的癌症，那么就不会带来特别明显的结果。

这个问题被称为过度诊断。澳大利亚一个研究小组的描述：“过度诊断不是假阳性诊断或误诊。”诊断是正确的，但它对患者的健康几乎没有益处，甚至可能造成伤害。

毫无疑问，对于如果未被发现就会导致死亡的癌症，筛查项目可以发现它们。那么，我们为什么要担心过度诊断呢？因为筛查也会造成伤害。接受结肠镜检查的患者有时会出现肠穿孔，而活检可能导致感染。放疗和化疗等治疗方法会对人们的健康造成严重风险，切除肿瘤的手术也是如此。

资料来源：

微软联合建立 AI 癌症识别模型，训练数据集包含 400 万张图像[EB/OL].(2023.09.19).
[2023.12.07].<https://finance.sina.com.cn/tech/roll/2023-09-19/doc-imznfrpa8699793.shtml>

新标准

欧洲为癌症领域人工智能制定新标准

EUCAIM(European Federation for CAncer IMages)项目是欧盟委员会发起的欧洲癌症成像倡议的基石，该倡议是欧洲战胜癌症计划核心，旨在促进癌症治疗和护理中数字技术的创新和部署，为癌症患者实现更精确、更快的

临床决策、诊断、治疗和预测医学。

近期欧盟委员会首次公开发布其平台“欧洲癌症图像”(Cancer Image Europe)，标志着 EUCAIM 发展的一个重要里程碑。首次发布的平台为欧洲各地的研究人员、临床医生和人工智能创新者带来了好处，并为癌症诊断和治疗的未来铺平了道路。

通过发布欧洲癌症图像的第一个版本，促进癌症图像和临床数据的共享，并加速欧洲各地研究人员、临床医生和人工智能创新者之间的合作。该平台旨在加快人工智能开发和其他数据密集型癌症研究活动的步伐，从而实现和增强科学突破，开启癌症诊断和治疗的未来。该项目的目标和愿景有助于实现欧盟倡议的战略目标，并促进创新和研究。

巴塞罗那超级计算中心-国家超级计算中心(BSC-CNS)是 EUCAIM 联盟的成员，BSC 研究员 Salvador Capella 表示：“EUCAIM 的早期原型将未来联邦基础设施的不同构建块结合在一起，并开始从使用它的不同用户档案中收集反馈。我们希望确保该平台——包括其联合学习能力——在设计上符合目的。”

EUCAIM 首次平台发布的主要功能和亮点包括：

来自欧盟资助的健康成像人工智能项目存储库的癌症成像数据集的公开目录，遵循共同的元数据模式。

用于了解提供程序中可用信息的联合搜索工具。

与生命科学登录验证和授权基础设施完全集成。

欧盟资助的癌症领域研究项目和基础设施的关键组成部分的重复使用和增值。

通过欧洲癌症图像基础设施，EUCAIM 项目解决了现有癌症图像存储库的碎片化问题。作为未来更新的一部分，将建立一个分布式癌症成像图集，其中包含来自 10 万多名患者的 6000 多万匿名癌症图像数据，基础设施将与欧洲健康数据空间完全一致。欧盟各地的临床医生、研究人员和创新者将可以访问这些数据，用于开发和基准测试值得信赖的人工智能工具。

BSC 与巴塞罗那大学(UB)共同领导一个技术工作包(WP6)。WP6 涉及联合学习和分析解决方案，并专注于定义利用联盟成员提供的数据的替代方法，从在每个数据提供者节点中进行计算的完全联合版本到可以在计算节点和数据

节点(包括 EUCAIM 中心节点)之间执行计算的混合模型。

此外，BSC 试图通过 EuroHPC 将可用的计算资源带给科学界。这项工作的重点是促进欧洲研究界与成员国密切协调，使用由欧盟资助的计算基础设施。

BSC 还领导该项目在 WP6 和 WP7 中的技术和科学基准测试活动，通过使用 OpenEBench 平台，该平台是在 ELIXIR 的背景下开发的，ELIXIR 是管理生命科学研究数据的泛欧基础设施。

资料来源：

欧洲为癌症领域人工智能制定新标准[EB/OL].(2023.10.09).[2023.12.07].<https://www.51cto.com/article/769076.html>



地址：上海市永福路 265 号

邮编：200031

编辑：李春霞

责编：姚恒美

编审：林鹤

电话：021-64455555

邮件：istis@libnet.sh.cn

网址：www.istis.sh.cn