

研究文章

开放访问

结肠直肠癌的准确诊断 基于组织病理学图像的使用 人工智能



K. S. 王^{1,2†}, G. 南斯拉夫^{3†}, C. 徐^{4†}, X.H. 孟^{5†}, J. 周^{1,2}, C. 郑^{1,2}, Z. 邓^{1,2}, L. 商¹, R. 刘¹, S. 苏联¹, X. 周¹,
Q. 列支敦士登¹, J. 列支敦士登¹, J. 王¹, K. 马², J. 气², Z. 胡², P. 强烈的味道或气味², J. 邓⁶, X. 邱⁷, B.Y. 列支敦士登⁷,
W.D. 沈⁷, R.P. 全⁷, J.T. 阳⁷,
L. Y. 黄⁷, Y. 肖⁷, Z.C. 阳⁸, Z. 列支敦士登⁹, S.C. 王¹⁰, H. 任^{11,12}, C. 两¹³, W. 郭¹⁴, Y. 列支敦士登¹⁴, H. 肖¹⁵, Y. 顾¹⁵,
J.P. 云¹⁶, D. 黄¹⁷, Z. 歌曲¹⁸, X. 迷¹⁹, L. 陈²⁰, X. 雁鸣²¹, Z. 列支敦士登²², Z.C. 黄³, J. 黄²³, J. 勒特雷尔绰号²⁴,
C.Y. 张²⁴, W. 周²⁵, K. 张²⁶, C. 彝语²⁷, C. 吴²⁸, H. 沈^{6,29}, Y.P. 王^{6,30}, H.M. 肖^{7*}和H. W. 邓^{6,7,29*}

摘要

背景: 准确和稳健的病理图像分析对结直肠癌 (CRC) 的诊断是耗时和知识密集型的, 但对CRC患者的治疗至关重要。目前临床/医院病理学家的繁重工作, 很容易导致基于日常图像分析对CRC的无意识误诊。

方法: 基于最先进的人工智能转移学习深度卷积神经网络, 提出了一种新的弱标记病理全切片图像 (WSI) 补丁。该方法使用了前所未有的170,099个补丁, > 14,680个WSIs, 来自> 9631名受试者, 涵盖了来自中国、美国和德国的多个独立来源的不同和具有代表性的临床病例。

结果: 我们的创新人工智能工具始终且几乎完全一致 (平均Kappa统计量为0.896) 和在诊断多中心的CRC WSIs时, 甚至往往比大多数经验丰富的病理学家更好。AI的受试者工作特征曲线下的平均面积 (AUC) 大于病理学家 (0.988 vs 0.970), 在其他AI方法应用于CRC诊断中表现最好。我们的人工智能生成的热图突出显示了癌症组织/细胞的图像区域。

(在下一页继续)

*通信信息: hmxiao@csu.edu.cn; hdeng2@tulane.edu

H.W. 邓是主要联系人

K.S. 王, G. 余, C. 徐, X.H. 孟是平等的第一作者

⁷系统生物学、数据信息和生殖健康中心,

中央基础医学院, 中央基础医学院

南方大学, 长沙410008, 中国湖南

⁶杜兰中心戴明医学系

杜兰大学医学院的生物学信息学和基因组学,

1奥尔良州运河街1440号, 1610室, 邮编70112

†信息列表可在文章的末尾获得



©作者(s)。2021开放获取本文授权在知识共享署名4.0国际许可, 允许使用、共享、适应、分发和复制在任何媒介或格式, 只要你给适当的信贷原作者(s)和来源, 提供一个链接到知识共享许可, 并表明如果有变化。本文中的图片或其他第三方材料都包含在文章的知识共享许可中, 除非在材料的信用额度中另有说明。如果材料没有包含在文章的知识共享许可中, 并且您的预期使用不被法律法规允许或超过了允许的使用, 您将需要直接获得版权所有者的许可。若要查看此许可证的副本, 请访问<http://creativecommons.org/licenses/by/4.0/>。知识共享公共领域奉献豁免 (<http://creativecommons.org/publicdomain/zero/1.0/>) 适用于本文中提供的的数据, 除非在该数据的信用额度中另有说明。

(从上一页继续)

结论：这是第一个通用的人工智能系统，能够稳定地处理大量的人工智能没有由临床病理学家普遍经历的疲劳引起的潜在偏差。这将大大减轻结直肠癌患者日常病理诊断的沉重临床负担，提高患者的治疗水平。该工具可推广到其他基于图像识别的癌症诊断。

关键词：结肠直肠癌，组织病理学图像，深度学习，癌症诊断

背景

结直肠癌（CRC）是全球发病率第三大癌症（6.1%），但死亡率第二（9.2%）。到2030年，全球新增病例和死亡人数[2]的负担预计将增加60%。准确、及时地检测结直肠癌对提高治疗效果和生存率至关重要。目前的结直肠癌诊断需要由高度专业化的病理学家进行广泛的视觉检查。诊断是通过从福尔马林固定石蜡包埋（FFPE）或冷冻组织中获得的苏木精和伊红（H&E）染色标本的数字全载玻片图像（WSIs）进行的。WSI分析的挑战包括非常大的图像大小（ $>10000 \times 10000$ 像素）、大小、形状、纹理和核染色的组织学变化，使诊断复杂和耗时。在大多数现代病理科，平均咨询工作量每年增加5-10%的~，[4]。目前的趋势表明，世界各地的病理学家短缺，包括美国[5]和中等低收入国家[6]。这导致病理学家过度劳累，这可能导致他们的日常工作上的缺陷和病理实验室的功能障碍，有更多的实验室错误[4]。胃肠病科对结肠标本检查要求高，但病理学家的培训时间长（ > 10 年）[7]。因此，必须开发可靠的病理图像分析和CRC检测工具，以提高临床效率和有效性，而不会在诊断过程中出现意外的人为偏倚。

最先进的人工智能（AI）方法，如深度学习（DL），在分类和预测方面非常强大。DL，特别是卷积神经网络（CNN）在肺[8,9]、乳腺[10,11]、前列腺癌[12-14]和皮肤[15,16]癌的WSI分析中已经有了许多成功的应用。现有的CNC WSI分析的CNN大多集中在癌症确定后的病理工作上，包括分级[17]、肿瘤细胞检测和分类[18-20]，以及生存预测[21-23]。尽管他们导致了相当高的准确性，但他们的研究样本量有限，并不能完全代表已定义的大量CRC组织学变异。这些变体包括管状的，

黏液细胞、印戒细胞和其他[24]。当应用于不同的独立样本时，这些限制会增加预测误差。同时，目前的DL模型是从单一数据源开发的，没有使用独立数据进行彻底的验证。他们只计算了补丁的准确性，而没有诊断wsi或患者。它们在各种临床环境中对CRC WSI诊断的广泛适用性，这可能涉及异质性平台和图像属性，目前仍不清楚。[25]尚未开发一种可推广于日常病理结直肠癌诊断的DL方法，可减轻病理学家的临床负担，提高诊断的准确性。

在这里，我们开发了一种新的以弱标记监督DL为中心的自动人工智能方法，用于CRC诊断的首次一般临床应用。这种AI方法使用了初始化-v3CNN架构[26]，其权值从迁移学习中初始化。弱标记监督学习在训练大量多样的数据集，而无需在对象水平上进行精确标记（e.g., 小的癌细胞），[12]。迁移学习是一种高效的图像分析DL技术，可以利用先前学习到的一般图像知识用于医学图像分析[27]。我们的工作是基于来自中国（8554例患者）、美国（1077例患者）和德国（ > 111 张载玻片）的多个独立医院/来源的wsi。本研究对提高结直肠癌诊断和治疗的有效性和效率具有很高的实用价值。它强调了人工智能在其他类型癌症的图像分析中的应用的一般意义和实用性。

方法

结肠直肠癌全幻灯片图像数据集

我们从14个独立来源收集了14,234个CRCWSIs（表1）。所有数据都被取消了识别。最大的图像集来自2010年至2018年在中南大学（中国长沙CSU）湘雅医院（XH）收治的6876名患者。XH是湖南省最大的医院，成立于1906年，与耶鲁大学[28]有密切的联系。其他独立的资料来源是美国的癌症基因组图谱（TCGA）(<https://portal.gdc.cancer.gov/>) [29]，国家肿瘤疾病中心（NCT）生物样本库

表1从多中心数据源获得的数据集的使用情况

数据源	数据集使用	样品制备	检查类型 激进的 结肠镜检查	人口*	CRC		非CRC		受试者 总数	幻灯片
					受试者	幻灯片	受试者	幻灯片		
湘雅医院（XH）	A	菲普	100% / 0%	长沙 中国	614	614	228	228	842	842
NCT-UMM（NCT-CRC-HE100 K）	B	菲普	纳	德国	纳	纳	纳	纳	纳	86
NCT-UMM（CRC-VAL-HE-7 K）	B	菲普	纳	德国	纳	纳	纳	纳	纳	25
xh	C	菲普	80% / 20%	长沙 中国	3990	7871	1849	2132	5839	10,003
xh	D	菲普	89% / 11%	长沙 中国	98	99	97	114	195	213
平矿合作医院（PCH）	C & D	菲普	60% / 40%	江西 中国	50	50	46	46	96	96
CS大学湘雅第三医院（TXH）	C & D	菲普	61% / 39%	长沙 中国	48	70	48	65	96	135
湖南省人民医院（HPH）	C & D	菲普	61% / 39%	长沙 中国	49	50	49	49	98	99
ACL	C & D	菲普	22% / 78%	长沙 中国	100	100	107	107	207	207
复旦大学上海癌症中心（FUS）	C & D	菲普	97% / 3%	劫掠去作水手 中国	100	100	98	98	198	198
广东省人民医院（GPH）	C & D	菲普	77% / 23%	广州，中国	100	100	85	85	185	185
南京鼓楼医院（NJD）	C & D	菲普	96% / 4%	南京 中国	100	100	97	97	197	197
西南医院（SWH）	C & D	菲普	93% / 7%	重庆，中国	99	99	100	100	199	199
空军第一附属医院医科大学（AMU）	C & D	菲普	95% / 5%	中国西安	101	101	104	104	205	205
中山大学癌症研究中心（SYU）	C & D	菲普	100% / 0%	广州，中国	91	91	6	6	97	97
中国人民解放军总医院（CGH）	C	菲普	纳	北京 中国	0	0	100	100	100	100
TCGA（TCGA-冷冻）	C	冻结的	100% / 0%	U. S.	631	1214	110	133	631**	1347
TCGA（TCGA-FFPE）	C	菲普	100% / 0%	U. S.	441	441	5	5	446	446
合计					6612	11,100	3129	3469	9631	14,680

*位置地图可在补充文本1中获得。a（请参见附加文件1）。**仅对于TCGA-Frozen数据，非CRC切片使用部分CRC切片上的正常肠组织

和德国曼海姆大学医学中心（UMM）病理学档案（NCT-UMM）（<https://zenodo.org/record/1214456#.XgaR00dTm00>, [22])、Adicon临床实验室（ACL）、INC和中国的11家医院（详见表1）。涉及的医院位于中国主要大都市地区服务>1.39亿人口，包括那些最负盛名的医院病理在中国：XH，复旦大学上海癌症中心（傅），中国解放军总医院（CGH），西南医院（SWH）和第一附属医院空军医科大学（AMU）；其他国家级知名医院：太阳

中山大学肿瘤中心（SYU）、南京鼓楼医院（NJD）、广东省人民医院（GPH）、湖南省人民医院（HPH）、中央湘雅第三医院（TXH）；以及地区著名的平矿合作医院（PCH）。除TCGA WSI的部分（~ 75%）来自冷冻组织[30]外，其余WSIs均来自FFPE组织。补充文本1描述了wsi的收集、质量控制和数字化的过程。a（请参见附加文件1）。

我们形成了四个数据集（表1）。数据集-a只包含来自XH的幻灯片，并用于补丁级

表2补丁级分析的数据集A（培训和测试）和数据集B（外部验证）

数据集	癌			非癌症			合计		
	受试者的幻灯片补丁			受试者的幻灯片补丁			受试者	幻灯片	小块
训练	406	406	19,940	153	153	22,715	559	559	42,655
测试	208	208	10,116	75	75	10,148	283	283	20,264
验证	纳	纳	15,550	纳	纳	91,630	纳	111	107,180*
合计	> 614	> 614 45,606		> 228	> 228	124,493	> 842	953	170,099

*有两个数据集用于验证。这个数字是这两个数据集的总和

培训和测试（表2）。我们仔细选择了WSIs，以包括所有常见的肿瘤组织学亚型。使用癌细胞/组织的不完整信息（e.g., 位置、形状和划分），病理学家弱标记wis的标记为包含或不包含癌细胞/组织。我们提供了两个弱标记的补丁作为说明性的比较示例，其中有两个完全标记的补丁作为对比部分（参见附加文件1：补充-图1）。来自同一患者的补丁都被放入相同的数据集（训练或测试），这样训练和测试数据集都是独立的。为了确保对癌症和正常组织特征的适当和全面的表现，我们将每个肿瘤样本平均包括49个补丁，每个健康样本平均包括144个补丁。包含大量癌细胞的贴片数量和只包含少量癌细胞的贴片数量是近似的平衡的，因此用于训练的贴片代表了在实践中看到的病例。

使用Dataset-B进一步验证了补丁级的性能，其中包含从NCT-UMM下载的107,180个补丁。有两个独立的亚群：86张苏木精人肿瘤组织和伊红精染色（HE）切片（NCT-CRC-HE100K）的100000张图像片和25张CRC-VAL-HE-7K）[22]的7180张图像片。补丁级别的训练、测试和外部验证的总体划分为2：1：5。所有图像为224×224像素，0.5 μm/像素。更多的描述可以在<https://zenodo.org/record/1214456#.XV2cJeg3lhF>上找到。在补丁被输入网络进行测试之前，补丁被重新调整到默认输入大小。

数据集-c用于患者水平的验证，由来自XH、其他医院、ACL、TCGA的冷冻和FFPE样本组成。考虑到SYU和CGH中癌症和非癌症载玻片的高度不平衡（表1），我们将它们合并到数据集c中。在数据集-c中，癌细胞在来自不同中心的图像中所占据的区域有所不同。来自10家医院和ACL的大部分切片（~ 72%）按区域包含10-50%的癌细胞（见附加文件1：补充图2）。

数据集d用于人类-人工智能竞赛，包含大约相等数量的幻灯片

XH，其他医院，以及ACL。每张幻灯片上平均有~ 5045补丁，超过20%的幻灯片包含< 1000补丁。补充-文本1。b总结了不同数据集中幻灯片的分配情况（见附加文件1）。

将载玻片数字化后，对癌症诊断标签进行严格、高精度的视觉验证。数据集-a和数据集-c包括超过10,000张幻灯片，由两位资深和经验丰富的病理学家通过初始和二读进行独立审查。当他们的诊断与之前的临床诊断结论一致时，然后将这些载玻片纳入数据集。如果两位专家彼此不一致或同意先前的临床诊断，则排除玻片。TCGA的载玻片标签来自

原始TCGA数据库。数据集-b的标签来自NCT-UMM。对人类-人工智能竞赛的数据集-d的二进制标签进行了更严格的检查。三名经验丰富的资深病理学家在不知道既往临床诊断的情况下，独立回顾了病理图像。如果达成共识，就包括幻灯片；否则，另外两名独立的病理学家将加入审查。经过五名病理学家的讨论，只有在达成协议后才被纳入样本；否则就被排除在外。

研究设计和管道

我们预测患者癌变状态的方法包括两个主要步骤：用于局部补丁的DL预测和用于患者水平诊断的补丁水平的结果聚合(图. 1)。预处理后的WSIs作为补丁级预测的输入。建立了一个深度学习模型来对补丁进行分析。然后，通过一种新的基于补丁簇的方法来聚合补丁水平的预测，以提供幻灯片和患者水平的诊断。斑块级预测的性能和聚合的方式在很大程度上决定了患者级诊断的准确性。我们的实证结果显示，贴片水平的~敏感性为95%，~的特异性为99%，足以达到较高的预测能力，并使用我们的方法在患者水平上控制假阳性率（FPR）

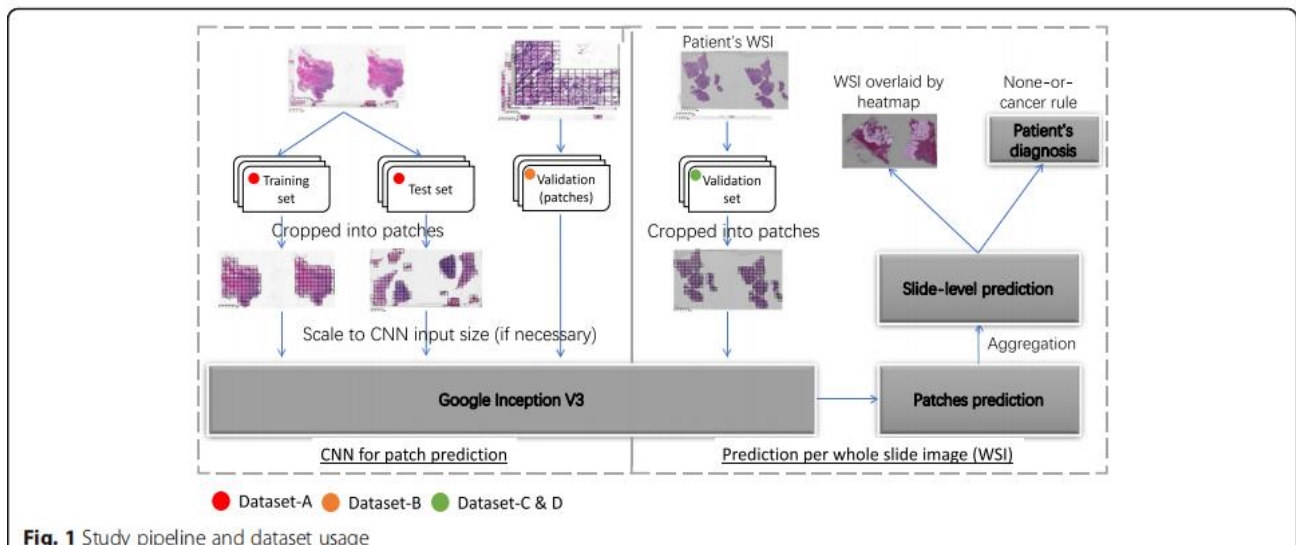


Fig. 1 Study pipeline and dataset usage

建议的聚合方法(参见附加文件1: 补充文本1. c). 此外, 还生成了热图和激活图来显示幻灯片上的信息区域。每个步骤的细节说明如下。

补丁级训练的图像预处理

图像预处理分为三个步。首先, 我们以 \times 将每个WSI与 300×300 像素贴片放大20像素, 可以很容易地转换为大多数CNN架构所需的输入大小(如初始-v3[26]所需的 299×299 输入大小, 参见附加文件1: 补充表1)。与其他使用 512×512 像素的研究相比, 使用更小的斑块尺寸将使癌症区域的边界更精确的[19]。其次, 我们根据两个标准去除非信息背景斑块: 斑块的3个颜色通道值的最大差异小于20, 或灰度[8]中50%以上的亮度小于220。结合这两个标准, 我们去除背景补丁, 并保留尽可能多的组织补丁。第三, 采用常规的图像增强程序, 如随机翻转和随机调整饱和度、亮度、对比度和色调。每个像素的颜色以每个图像的平均值为中心, 其范围从 $[0, 255]$ 转换/归一化为 $[-1, 1]$ 。

通过深度学习进行的补丁级培训

我们的DL模型使用初始空间-v3作为CNN架构, 对癌变和正常斑块进行分类。“盗梦空间”网络使用不同的内核大小, 在学习来自不同尺度的病理图像中的诊断信息方面特别强大。这个

架构在其他癌症类型[8, 15, 31, 32]的分析中取得了接近人类专家的表现。在ImageNet数据集[33]和WSIs分析[33]上, 有一些空间架构表现良好, 如空间-v1[34]、空间-v3[26]和空间-v4[35]。我们选择了《v3》是基于他们在测试集中的补丁级和患者水平的性能, 这表明《v3》中的复杂性和多尺度模块使其更适合识别组织病理学WSIs(见附加文件1: 补充文本1. d) [26, 34 - 39]。在研究过程中, 我们还测试了一些最新的算法, 如DenseNet [37]和ResNeXt [39]。初始空间-v3在补丁级CRC分类中仍然表现最好。

我们通过从ImageNet [26]中预先训练的权值进行迁移学习来初始化CNN, 该权重经过优化以捕获一般图像[27]中的结构。通过迁移学习, 我们的模型可以最有效地识别CRC诊断的关键图像特征。 300×300 像素的补丁的大小被调整为 299×299 像素。因此, 测试集中的补丁在被输入网络之前被重新调整到 299×299 像素($0.37 \mu\text{m}/\text{像素}$)。通过以下训练步骤, 对网络进行了深度微调。考虑到在聚合贴片水平的结果后可能出现的高假阳性率, 我们随机搜索最优的超参数集, 目标是达到 $> 95\%$ 的敏感性和 $> 99\%$ 的特异性。我们发现, 在贴片水平上的这个目标, 患者水平上的错误率得到了很好的控制(见附加文件1: 补充文本1. c)。使用RMSProp [40]优化器对所有层的参数进行150,000个微调后最终确定, 权重衰减为0.00004, 动量值为0.9, RMSProp衰减设置为0.9。初始学习率为

. 010, 随epoch呈指数衰减, 最终学习率为0.0001。当批大小为64时, 得到了优化结果。训练和测试程序是在一个使用NVIDIA P100 GPU的Linux服务器上实现的。我们使用Python v2.7.15和tencorflowv1.8.0来进行数据预处理和CNN模型的训练和测试。

患者诊断及假阳性对照

考虑到多个贴片水平预测积累的高假阳性率(FPR), 我们提出了一种新的基于贴片聚类的聚集方法, 基于肿瘤细胞倾向于聚集在一起(特别是在 $\times 20$ 的放大倍数下)。基于fMRI [41]的聚类推断, 我们预测, 如果在幻灯片上有几个拓扑连接的阳性斑块(由聚类大小定义), 如四个斑块为一个正方形, 则WSI为癌症阳性。否则, 我们预测这个幻灯片是负的。我们测试了不同的集群大小, 并选择了4的集群大小, 作为在测试数据集中的经验观察到的灵敏度和FPR的最佳平衡的结果(见附加文件1: 补充文本1.e)。对于有一个或多个载玻片的患者, 用 $S = \{s_1, s_2, \dots, s_l\}$, 我们结合所有患者切片的结果, 提供了患者水平的诊断 $D(S)$: $D(S) = D(s_1) \cup D(s_2) \cup \dots \cup D(s_l)$, 其中 $D(s_i) = 1$ 或0分别表示第 s_i 玻片的阳性或阴性。只要其中一张幻灯片显示诊断, 病人就会被诊断为癌症。

人工智能竞赛

6名具有不同的1至18年临床实践经验的病理学家(A-F)参加了比赛(见附加文件1: 补充表2)。病理学家在阅读了数据集-d中的WSIs后, 独立地为每个患者提供了指定癌症或非癌症的诊断。病理学家没有参与数据的收集或标记。一位独立的分析师盲目地总结和比较了人工智能和人类专家进行诊断的准确性和速度。

统计分析和可视化

我们评估了人工智能和病理学家在敏感性、特异性和准确性方面的表现

$\frac{\text{\#of correct predictions}}{\text{\#of total predictions}}$ () 用于诊断。接收器操作

计算绘制灵敏度与FPR的擦特征(ROC)曲线和相应的ROC曲线下面积(AUC)。采用配对Wilcoxon符号秩检验对多个数据集中AI和各病理医师的auc进行比较。我们检查了两两

Cohen的Kappa统计数据(K)在人工智能和病理学家之间的共识。统计分析在Rv3.5(维也纳, 奥地利)中进行, 使用软件包插入符、ggplot2、pROC和心理等。统计学显著性水平设置为alpha水平0.05。

为了在WSI中定位CRC区域, 我们根据每个补丁的置信度评分将WSI可视化为一个热图。较亮的区域表明分类器认为癌症阳性区域的信心更高。该热图是由Python (<https://www.巨蛇org/>)和由gimp覆盖的原始WSI(<https://www.gimp.org/>)。

结果

我们的模型在补丁级预测中的精度最高

我们将数据集a(表1)中的842个WSIs划分为62919个非重叠补丁(表2), 构建基于incent-v3微调的补丁级预测的CNN。平均每个WSI包含 ~ 75 个贴片, 以确保适当和全面的癌症和正常组织特征的表示。3种主要的CRC组织学亚型参与了训练和检测, 包括74.76%的管状细胞、24.59%的粘液性细胞和0.65%的印戒细胞斑块, 大致反映了其临床发生率[42]。在训练中, 有19940个(46.75%)贴片患有癌症, 有22715个(53.25%)贴片表现正常。使用另一组独立的10116(49.92%)癌症和10148(50.08%)非癌症贴片, 贴片水平预测的AI实现了98.11%的测试准确性和AUC的99.83%。9AUC在CRC诊断和预测(79.2–99.4%)方面优于以往所有的AI研究, 甚至在大多数其他类型的癌症方面也优于此类研究(82–99.9%, 见附加文件1: 补充-表3, [8、12、17、19、22、43–48])。特异性为99.22%, 敏感性为96.99%, 均比较突出。在外部验证数据集-b, 我们的模型产生了一个在NCTC-CRC-HE-100K的准确性和AUC为96.07%和98.32%, 和在CRC-VAL-HE-7K94.76%和98.45%, 匹配的性能内部数据和其他AI研究的补丁级验证分析(AUC 69.3–95.0%, 见附加文件1: 补充表3)。补丁级测试和验证结果总结如表3所示。

使用d1预测在患者水平上诊断结直肠癌小块

我们的人工智能方法对来自8594名患者的13,514张载玻片进行了患者诊断测试(数据集-c)。在XH的最大亚群(5839例患者)中, 我们的方法产生的准确率为99.02%, AUC为99.16%。2、表3)。在其他独立的多中心中

表3补丁程序级别（数据集A和数据集B）和患者级别（数据集C和数据集D）的性能总结

源敏感性特异性准确性AUC				
数据集-A（补丁级测试）				
XH	96.99%	99.22%	98.11%	99.83%
数据集-B（补丁级验证）				
NCT-CRC-HE100 K	92.03%	96.74%	96.07%	98.32%
-7k	94.24%	94.87%	94.76%	98.45%
数据集-C（患者级验证）				
xh	98.80%	99.51%	99.02%	99.16%
TCGA-冷冻	94.04%	88.06%	93.44%	%
TCGA-FFPE	97.96%	100.00	97.98%	91.05%
SYU-CGH	98.90%	%	95.43%	%
		92.45%		98.98%
				95.68%
数据集D（患者级人类人工智能竞赛）				
xh	97.96%	100%	98.97%	98.99%
同步信号单元	98.90%	100%	98.97%	99.45%
数据集d(患者级验证和人工智能比赛)				
预编译头文件	96.00%	97.83%	96.88%	97.91%
件	100%	97.92%	98.96%	%
特克斯	97.96%	97.96%	97.96%	99.20%
HPH	100%	97.96%	98.99%	%
福斯	100%	97.65%	98.91%	98.98%
GPH	92.93%	97.94%	95.41%	%
新泽西州	98.99%	97.00%	97.99%	99.99%
斯瓦	97%	97.06%	97.04%	99.15%
阿姆	100%	97.20%	98.55%	%
ACL				95.84%

在数据集上，我们的方法始终表现得很好。对于来自其他医院、TCGA-FFPE和ACL的FFPE切片，AI方法产生的平均AUC和准确性高于97.65%(图。2)。对于冷冻切片TCGA-Frozen，AI准确率和AUC分别为93.44%和91.05%。2)。对于独立数据集，我们的AUC值（范围从91.05到99.16%）高于其他基于人工智能的方法（范围从83.3到94.1%）。值得注意的是，由于大多数早期的人工智能方法都是在样本规模小得多的数据集上进行测试的（见附加文件1：补充-表3），它们的性能可能被高估了。TCGA中阴性幻灯片数量有限可能导致不平衡的问题，分类需要进一步研究，这超出了本研究的范围。TCGA-Frozen切片上的结果显示，我们的方法确实学习了癌症和正常组织的组织学形态用于癌症诊断，在FFPE和冷冻样本中都保存，尽管我们的方法是基于FFPE样本开发的。表3总结了完整的患者水平的结果。

与六名人类专家进行比赛

我们的AI方法在诊断来自独立中心的1831个WSIs方面的性能始终与病理学家相一致(数据集-d，图。3)。人工智能的平均准确率和AUC分别为98.06%（95%置信区间[CI]97.36–98.75%）和98.83%（95%CI98.15–99.51%），在7名竞争对手中排名前三(AI+6名

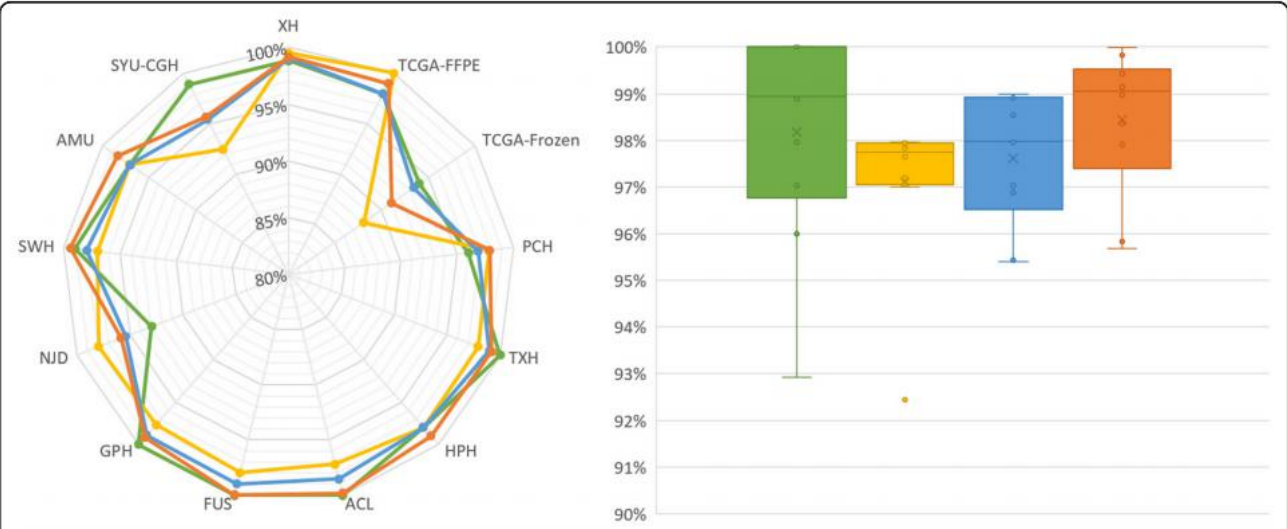


图2对来自数据集-c的12个独立数据集的患者水平测试性能。左图：数据集-c中每个数据集的灵敏度、特异性、准确性和AUC的雷达图。右图：箱线图显示了不包括XH和TCGA的数据集的敏感性、特异性、准确性和AUC的分布。方框中的水平条表示中位数，而十字架表示平均值。圆圈表示数据点

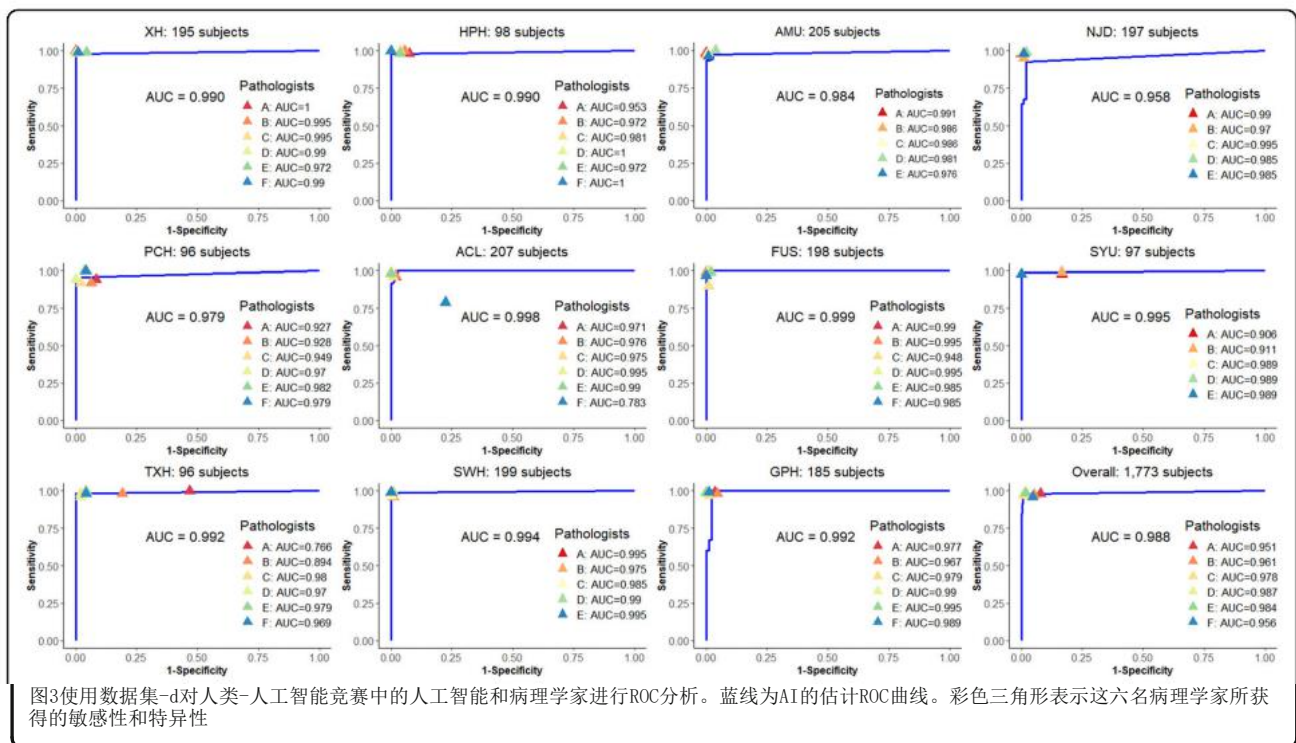


图3使用数据集-d对人类-人工智能竞赛中的人工智能和病理学家进行ROC分析。蓝线为AI的估计ROC曲线。彩色三角形表示这六名病理学家所获得的敏感性和特异性

), 高于病理学家的平均水平 (准确率97.14% (95%CI96.12–98.15%) 和AUC96.95% (95%CI95.74–98.16%))。在多中心数据集中, 对auc的配对Wilcoxon符号秩检验发现, AI与每个病理学家之间没有显著差异。相对于病理学家的平均水平 (97.47%), AI产生了最高的敏感性 (98.16%) (见附加文件1: 补充表4)。略优于AI的病理学家 (D和E) 分别有7年和12年的临床经验, 而AI分别优于其他4名病理学家, 分别有1年、3年、5年和18年的临床经验。Cohen ‘s Kappa统计数据 (K) 显示AI和每个病理学家之间有极好的一致性 ($K \geq 0.858$, 平均0.896) (见附加文件1: 补充表5)。因此, 我们的方法被证明是可推广的, 可以为潜在的CRC受试者提供诊断支持, 如一个独立的病理学家, 这可以大大减轻专业病理学家沉重的临床负担和培训成本。HumanAI比赛的详情见补充表4和表5 (见附加文件1)

病理学家都被告知要与我们的人工智能以及相互竞争; 因此, 他们的表现是在最好的条件下以最大的努力取得的, 这代表了他们的最高技能和最小的错误。然而, 由于临床工作量较大, 它们在准确性和速度方面的表现将不如AI那样稳定。人工智能的研究

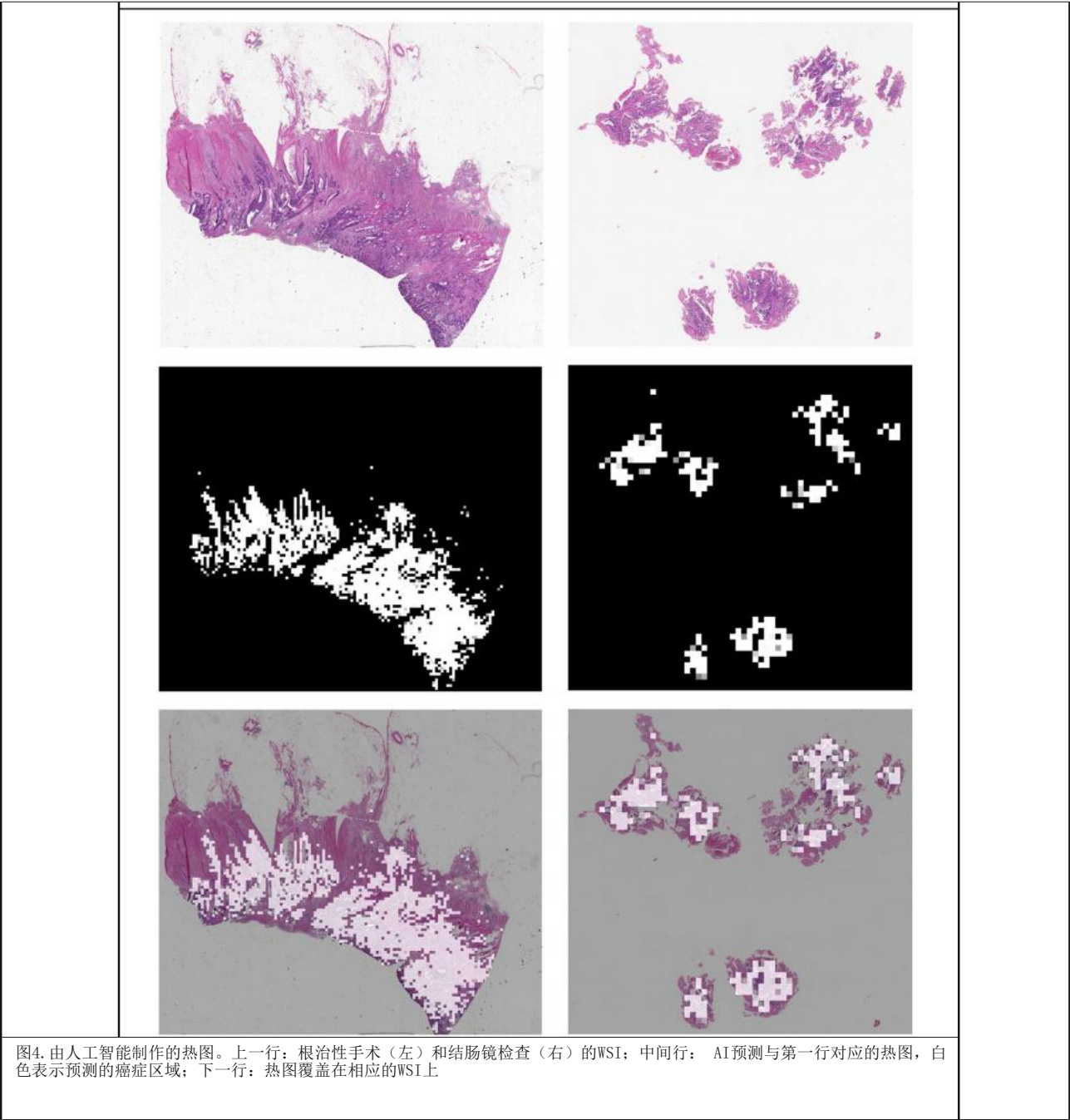
使用WSI进行癌症诊断表明, 人工智能可以准确诊断~ 20 s [8]或更少 (在我们的病例中为~13s)。随着DL技术的发展和先进的计算硬件, 人工智能可以不断改进, 并为CRC或其他癌症提供稳定、快速和准确的首次诊断。

幻灯片级热图

我们的方法提供了一个额外的独特特征: 热图突出潜在的癌症区域 (作为补丁)。在无花果. 4, 我们提出了两个WSIs, 它们与预测的热图叠加。对于根治性手术WSI和结肠镜检查WSI, 真正的癌变区域与AI获得的突出的补丁高度重叠, 这也得到了病理学家的证实。请参阅补充-图3中的更多示例 (请参见附加文件1)。此外, 为了可视化DL用于CRC检测的信息区域, 我们在补充-图4中提供了激活图 (见附加文件1)。

讨论

我们收集了高质量、全面和多个独立的人类WSI数据集, 用于训练、测试和外部验证, 方法侧重于常见临床环境下CRC的病理诊断。我们模拟了WSI分析的临床程序, 包括图像数字化、幻灯片回顾和有争议的幻灯片的专家咨询。



与其他研究[21]不同，我们在构建研究数据集时没有应用任何手动选择的幻灯片或感兴趣的区域。考虑到CRC复杂的组织学变异，我们从三种最常见的亚型中随机选择训练补丁，大致与它们的发病率成比例。大肿瘤组织和小肿瘤组织区域的图像中的斑片数量是平衡的，在斑片水平分析中很好地表示。收集到的图像经至少两名CRC病理学高级专家的同意进行标记(见附加文件1：补充文本1。b)。测试

来自中国、美国和德国不同地点的数据集作为验证和泛化的代表性池。我们的数据集很好地代表了在诊所中看到的幻灯片。因此，训练后的AI模型是ro-胸部和一般化，可推广用于分析不同生产协议的图像和图像质量。

对于一个快速增长的领域，我们意识到，在本文的研究完成后，已经提出了一些新的CNN架构，如DenseNet [37]、挤压-激励网络[38]和ResNeXt [39]。我们做了一些探索

通过比较ResNet152V2、DenseNet201和NASNetLarge相对于初始化-v3的补丁分类进行分析。DenseNet201产生了与《盗梦空间》v3相似的评估指标，而其他两种架构产生的准确性和AUC都低于《盗梦空间》v3。虽然与初始v3相比，这些新模型已被证明可以提高对ImageNet数据集的预测精度，但初始v3中的复杂性（深度和参数数量）和多尺度模块可能适用于识别CRC WSI。新架构在病理图像分析和癌症诊断方面的性能值得更多的集中研究，以进行更详细的技术比较。此外，我们还发现了其他可能扩展当前研究的技术，如半监督和无监督学习[49, 50]，它可以有效地从更多有无标签的WSIs中学习，以及多尺度决策聚合和数据增强[51]，它可以在有限数据存在的情况下工作。鉴于目前提出的方法已经取得了高度准确的性能，我们可以在未来的研究中研究这些新技术是否以及如何通过更少的数据收集和标记努力来获得当前的预测性能。

有几种组织学类型太罕见（[52]的发病率低于0.5%）不能被包括在内，如髓状、微乳头状和锯齿状。我们的人工智能方法在冷冻样本中的表现仅略低于FFPE样本。随着来自罕见类型的WSIs和更多的冷冻样本可用于未来的训练，我们希望我们的方法可以不断改进，使其更一般化。

由于不可能直接处理大尺寸的WSI，因此以往的研究大多是通过整合贴片级的识别结果来获得患者的诊断的。这种策略很难控制基于单个补丁的多个预测的累积假阳性率（FPR）。最近，库德雷等人。以正斑块的比例或所有斑块的平均概率作为WSI [8]的预测准则。虽然他们的结果在三个独立的数据集（所有小样本量（340张幻灯片）），但他们的聚合方法可能不适用于那些只有一个小区域的图像，这些图像会对癌症患者产生假阴性结果。相反，我们提出了一种新的基于贴片的WSI或患者水平预测的聚合策略，该策略很直观，可以很容易地平衡敏感性和特异性。具体来说，我们聚合了来自幻灯片上拓扑连接的补丁簇的信息，以确定癌症状态。在实践中，将聚类大小设置为4最有可能超过病理学家的平均准确性，而聚类大小为2可以用于病理筛查

根据我们的测试数据，~的平均敏感性为99.78%，~的平均特异性为72.29%（见附加文件1：补充文本1.e）。

结论

总之，我们开发了一种新的基于人工智能的组织病理学图像分类方法来进行CRC诊断，该方法在迄今为止该领域的样本量和数据源最多的情况下取得了最好的性能。我们的方法能够快速准确地地区分CRC病例与健康或炎症病例，在大规模多中心数据检测方面与病理学家相当，甚至优于病理学家。据我们所知，这是第一个为CRC初始筛查的日常临床病理诊断提供可靠、全面、可靠的辅助工具的人工智能研究。我们的方法也可以通过要求代码进行调整和应用于其他癌症类型的组织学分析。

补充信息

在线版本包含了补充材料，可在<https://doi.org/10.1186/s12916-021-01942-5>上获得。

附加文件1：补充文本1。aWSIs的收集和数字化。补充文本1。b数据集-A、b、C和D。补充-文本1。c补丁级别的性能和患者级别的准确性。补充-文本1。d补丁级不同架构的比较。补充-文本1。e比较不同集群大小的补丁级结果的聚合。补充表1。为普通CNN的输入补丁大小。补充-表2病理学家信息。补充-表3AI应用于CRC和其他癌症类型的auc列表。补充-表4人工智能和病理学家在人类-人工智能竞赛中的总体表现。补充-表5人类专家与人工智能之间一致的Cohen's Kappa系数。补充-图1弱标记和完全标记的CRC补丁。补充信息-附图2。在多个独立的WSI数据集中，癌变区域的分布，通过在WSI上包含癌细胞的斑块的比例(P)来测量。补充-图3由AI生成的热图。补充-图4由AI生成的激活图。

缩写

人工智能；人工智能；AMU：空军医科大学第一附属医院；AUC：接收机工作特征曲线下平均面积；CNN：卷积神经网络CGH：中国解放军总医院结直肠癌DL：深度学习FFPE：福尔马林固定石蜡包埋FPR：假阳性率FUS复旦大学上海癌症中心广东省人民医院HPH：湖南省人民医院湖南省人民医院卫生教育：苏木精、伊红IJD：南京鼓楼医院PCH：平矿合作医院SWH：西南医院SYU：中山大学肿瘤中心；TXH：中央大学湘雅第三医院；WSI：全片图像；XH：湘雅医院

致谢

该项目的部分计算工作是在俄克拉荷马大学（OU）的OU超级计算教育与研究中心（OSCER）进行的。作者希望感谢俄克拉荷马大学健康科学中心的兰斯·福特在编辑手稿方面的帮助。我们感谢编辑和三位匿名的审稿人，他们的评论/建议帮助改进和澄清了这份手稿。

作者的贡献

概念化, H. W.D., H.M.X., K.S.W., G.Y., C.X. 和 X.H.M.; 方法, G. Y., K.S.W., X.H.M., C.X., C.Y.Z., J.L., W.Z., K.Z., Y.P.W., s.c.w, Z.c.H., J.H., H.M.X., 和 H.W.D.; 调查, H.M.X., K.S.W., G.Y., C.X., X.H.M. 和 H.W.D.; 写作-原稿, C.X., X.H.M., G. Y. 和 K.S.W.; 写作-评论和编辑, H. W.D., H.M.X., Y.P.W., C.Y.Z., J.L., W.Z., K.Z., C.W. 和 H.S.; 资金收购, H. M.X., H.W.D., 和 K.S.W.; 资源, H.M.X., K.S.W., G.Y., 和 H.W.D.; 数据管理, K.S.W., J.Z., C.Z., Z.D., L.S., R.L., S.S., X.Z., Q.L., J.L., J.W., K.M., J.Q., Z.H., P.T., J.D., X.Q., B.Y.L., W.D.S., R.P.Q., J.T.Y., L.Y.H., Y.X., Z.C.Y., Z.L., H.R., C.L., W.G., Y.L., H.X., Y.G., J.P.Y., D.H., Z.S., X.F., L.C., X.Y., Z.L. 和 C.Y.; 监督, H. M.X. 和 H.W.D. 作者(s) 阅读并批准最终稿件。

资金

H.S. 和 H.W.D. 部分资助由美国国立卫生研究院 (R01AR059781、P20GM109036、R01MH107354、R01MH104680、R01GM109068、R01AR069055、U19AG055373、R01DK115679)、Edward G. 施里德捐赠基金和博士们。W.C. 蔡和 P. T. 杜兰大学生物统计学功夫教授。H.M.X. 国家重点研发计划项目 (2017YFC1001103、2016YFC1201805)、国家自然科学基金项目 (#81471453)、江旺教育捐赠基金资助。K.S.W. 国家自然科学基金项目 (#81972490); 湖南省自然科学基金项目 (#2015JJ2150)。Z.C.Y. 国家自然科学基金资助项目 (#81673491)。K.Z. 部分由美国国立卫生研究院资助 (2U54MD007595)。

数据和材料的可用性

由于计算/存储资源有限, 本研究中分析的数据集不能公开获得, 但在合理的要求下可以从通讯作者处获得。我们的方法的源代码可以在GitHub: <https://github.com/csu-bme/DeepPathology> 上找到CRC。

伦理批准和同意参与

不适用。

同意出版

不适用。

竞争利益

作者声明他们没有相互竞争的利益。

作者详细信息

¹中南大学湘雅医院病理科, 长沙410078, 中国湖南。²中南大学基础医学科学院病理科, 长沙410013, 中国湖南。³中南大学基础医学学院生物医学工程系, 长沙410013, 中国湖南。⁴美国俄克拉荷马大学健康科学中心生物统计学和流行病学系, 俄克拉荷马城, OK 73104。⁵湖南师范大学生命科学学院分子与统计遗传学实验室, 长沙410081。⁶杜兰大学医学院杜兰生物医学信息学和基因组学中心戴明医学系, 运河街1440号, 1610室, 新奥尔良, 美国洛杉矶70112。⁷中南大学基础医学学院基础医学学院系统生物学、数据信息与生殖健康中心, 长沙410008, 中国湖南。⁸中南大学湘雅药学院药系, 长沙410078, 中国湖南。⁹中南大学生命科学学院, 长沙410013。¹⁰湖南师范大学信息科学与工程学院, 长沙410081。¹¹中国第二军医大学工里医院病理科, 上海, 200135, 中国。¹²长治医学院附属和平医院病理科, 长治046000。¹³浙江医学实验室有限公司病理实验室, 杭州310023。¹⁴湖南省师范大学第一附属医院病理科, 湖南省人民医院, 长沙410005, 中国湖南省。¹⁵湘雅第三医院病理科,

中南大学, 长沙410013, 中国湖南。¹⁶中山大学癌症中心病理科, 广州510060。¹⁷复旦大学上海癌症中心病理科, 上海200032, 中国。¹⁸中国人民解放军总医院病理科, 北京100853。¹⁹中国南京大学医学院附属医院南京鼓楼医院病理科, 南京210008。²⁰中国空军医科大学第一附属医院病理科, 西安710032。²¹中国第三军医大学西南医院病理学和西南癌症中心研究所, 重庆400038。²²广东省医学科学院广东省人民医院病理科, 广州510080。²³中南大学基础医学科学学院解剖学与神经生物学系, 长沙410013, 中国湖南。²⁴南密西西比大学计算机科学和计算机工程学院, 哈蒂斯堡, MS 39406, 美国。²⁵密歇根理工大学计算机学院, 霍顿, 密歇根州49931, 美国。²⁶美国路易斯安那州泽维尔大学RCMI癌症研究中心生物信息学设施计算机科学系, 新奥尔良, LA 70125。²⁷美国新奥尔良市奥克斯纳医学中心病理科, 邮编: A70121。²⁸佛罗里达州立大学统计系, 塔拉哈西, 佛罗里达32306, 美国。²⁹杜兰大学医学院戴明医学系生物医学信息学和基因组学分部, 美国新奥尔良, LA 70112。³⁰杜兰大学生物医学工程系, 新奥尔良, 美国LA 70118。

收到日期: 2020年10月22日接受日期: 2021年2月16日

Published online: 23 March 2021

参考文献

1. 布雷, 费雷, 我, 西格尔, 托雷, 杰玛尔A. 2018年全球癌症统计: 全球癌症统计学对全球185个国家36种癌症的发病率和死亡率的估计。CA癌症J Clin. 2018;68(6):394–424.
2. 阿诺德M, 塞拉女士, 拉弗森M, 我, 杰玛尔A, 布雷F. 结肠直肠癌发病率和死亡率的全球模式和趋势。勇气2017; 66(4):683–91.
3. Komura D, 石川. 用于组织病理学图像分析的机器学习方法。合成结构生物技术. 2018;16:34–42.
4. 马恩R. 病理学家的工作量和患者的安全。诊断性组织抑素。2016;22(8):283–7.
5. 梅特·德姆, 科尔根TJ, 梁街, 蒂蒙斯CF, 公园JY. 2007年至2017年美国加拿大病理学家劳动力的趋势。詹姆斯Netw开放。2019;2(5):e194337.
6. 赛义德, 卢肯德, 弗莱明. 在美国国家提供病理支持。J Glob协议。2015;1(1):3–6.
7. 黑谢弗, 明天, 普里斯托斯基, 斯坦伯格. 培训病理学住院医师来实践21世纪的医学: 一个建议。Acad悲枪。2016;3:2374289516665393.
8. 库德雷, 奥坎波, 萨克拉罗普洛斯, 纳鲁拉, M, 芬约, 莫雷拉, N, 齐里戈斯. 利用深度学习对非小细胞肺癌组织病理学图像进行分类和突变预测。Nat地中海。2018;24(10):1559–67.
9. 华凯, 徐志, 海志, 郑慧, 陈勇. 利用深度学习技术对计算机断层扫描图像进行计算机辅助分类。Onco目标。2015;8:2015–22.
10. 维塔, 范迪斯特, 威廉姆斯, 王H, 王H, 克鲁斯, 冈萨雷斯, 维斯特加德, 达尔等. 乳腺癌组织病理学图像中有丝分裂检测算法的评估。医学图像肛门。2015;20(1):237–48.
11. 我, , , M. M, Manson QF等: 检测乳腺癌女性淋巴结转移的深度学习算法的诊断评估。JAMA. 2017;318(22): 2199–210.
12. 坎帕内拉, 汉娜, 杰斯拉夫, 米拉弗洛. 克劳斯·KJ: 对全幻灯片图像使用弱监督深度学习的临床级计算病理学。Nat地中海。2019;25(8):1301–9.
13. 布尔顿, 平克斯, 博文, 维克, 贝尔, 金内肯, 拉克, 赫尔斯贝尔根, 卡卡. 自动深度学习

- 使用活检的前列腺癌格里森分级系统：一项诊断研究。《柳叶刀》。2020;21(2):233–41.
14. 斯特罗姆P, 卡尔塔萨洛K, 奥尔森H. 活检中前列腺癌诊断和分级的人工智能：一项基于人群的诊断研究（第21卷，第222, 2020页）。《柳叶刀》。2020;21(2):E70.
 15. 埃斯特瓦A, 库普雷尔B, 诺沃阿拉, 科J, 斯威特SM, 布劳HM, 特伦S. 利用深度神经网络对皮肤癌的皮肤科级别分类。自然。2017;542(7639):115–8.
 16. 于磊、陈赫、袁问、秦J、衡巴。通过非常深的残余网络在皮肤镜图像中自动识别黑色素瘤。IEEE跨医学成像。2017;36(4):994–1004.
 17. Sari CT, Gunduz-Demir C. 通过深度学习的无监督特征提取进行结肠组织图像的组病理学分类。IEEE跨医学成像。2019;38(5):1139–49.
 18. 陈, 陈, 陈, 陈, 陈, 陈. 局部敏感的深度学习, 用于常规结肠癌组织学图像中的细胞核的检测和分类。IEEE跨医学成像。2016;35(5):1196–206.
 19. 哈桑, 哈达, 哈达, 哈库斯, 甜点, 皇帝. 利用卷积神经网络对多光谱结肠癌组织进行分类。J Pathol通知。2017;8:1.
 20. Chaddad A, Tanougast C. 异常细胞图像的纹理分析, 以预测结肠癌癌的连接体。肛门细胞原生质体 (Amst)。2017;2017:8428102.
 21. 比奇科夫、林德、图尔库基、诺德林、科瓦宁、维里尔、瓦利安德、伦丁、哈格伦德、伦丁J. 基于深度学习的组织分析可以预测结肠癌癌预后。科学代表。2018;8(1):3395.
 22. 卡瑟、克里萨姆、夏罗阿通、路埃德、赫佩尔、马克思、马克思、价值、费伯等。利用深度学习预测结肠癌组织学切片的生存率：一项回顾性多中心研究。Plos Med. 2019;16(1):e1002730.
 23. 例如, 德雷特、A、S、K、麦迪森、普拉丹、内斯海姆、阿尔布雷格森等。预测结肠癌预后的深度学习：一项发现和验证研究。柳叶刀。2020;395(10221):350–60.
 24. 弗莱明, 拉维拉, 塔提舍夫, 王HL. 结肠直肠癌：病理学方面。J胃肠上校。2012;3(3):153–73.
 25. 算法可以帮助专家病理学家诊断前列腺癌。癌症迪斯科2020, 10(3):0F1.
 26. 例如, 例如, 例如, 例如, 等等。重新思考计算机视觉的初始架构。日期: 2016 IEEE计算机视觉和模式识别 (CVPR) 会议: 2016年2016年6月27–30日; 2016. p. 2818–26.
 27. 柯美尼迪斯、戈德鲍姆M、蔡W、瓦伦蒂姆中国化学会、梁H、巴克斯特SL、麦基翁A、杨G、吴X、严F等。通过基于图像的深度学习来识别医学诊断和可治疗的疾病。单元2018;172(5):1122–31. e1129
 28. Li LJ, Lu GX. 中国湖南医学伦理原则在供者人工授精 (AID) 治疗中的应用: 湘雅生殖遗传医院的有效实践。J医学伦理。2005; 31(6):333–7.
 29. 格罗斯曼, 希思美联社, 费雷蒂V、瓦姆斯H, 洛伊博士, 基比瓦, 斯塔特1m. 走向对癌症基因组数据的共同愿景。N Engl J医学。2016; 375(12):1109–12.
 30. 我, 古特曼, 巴恩霍尔兹-斯隆, 委拉斯开兹维加, 兄弟DJ, 库珀小伙子。利用卷积神经网络从组织学和基因组学中预测癌症的预后。我们的课程。2018;115(13):E2970–9.
 31. 古尔山五世、彭L、科拉姆、斯图普、吴德、纳拉亚瓦米、威德纳、T夫人、卡德罗斯等。糖尿病视网膜病变的深度学习方法开发和验证视网膜眼底照片。一种检测糖尿病视网膜病变的深度学习算法的准确性。JAMA. 2016;316(22):2402–10.
 32. 例如, 例如, 例如, B, B, 桑切斯. 对医学图像分析中的深度学习的调查。医学图像肛门。2017;42:60–88.
 33. 我, 我, 我, 我. 深度卷积神经网络能够识别异构的数字病理图像。生物医学。2018;27:317–28.
 34. 蔡格迪、魏乐、杨庆、小王、里德、安格洛夫、万胡克五世、拉比诺维奇。更深入地讨论卷积。在: 2015年IEEE计算机视觉和模式识别会议 (CVPR): 2015年6月7–12日, 2015年; 2015年. p. 1–9.
 35. 艾菲, 范霍克五世, 阿勒米A. 初始网络v4, 初始网络和剩余联系对学习的影响。在: 第三十一届AAAI人工智能会议: 2017年; 2017年.
 36. 用于大规模图像识别的深度卷积神经网络。arXiv预印本arXiv: 140915562014.
 37. 黄KQ, 刘KQ, 温伯格KQ. 紧密连接的卷积网络。在: 2017 IEEE计算机视觉和模式识别会议 (CVPR) 上: 2017年7月21–26日2017; 2017. p. 2261–9.
 38. 胡杰, 沈左, 孙G. 挤压和励磁网络。在: 2018年IEEE/CVF计算机视觉和模式识别会议: 2018年6月18–23日; 2018年; 2018年. p. 7132–41.
 39. 我, 我, 我, 我. 在最小的监督下从有噪声的大规模数据集中学习。在: 2017 IEEE计算机视觉和模式识别会议 (CVPR) 上: 2017年7月21–26日2017; 2017. p. 6575–83.
 40. GSN H, 斯韦尔斯基K. 第6讲是小批量梯度下降的概述。在多伦多大学CSC321上发行的课堂讲稿; 2014年.
 41. 海勒, 斯坦利D, 耶库蒂利D, 鲁宾N, 本贾米尼Y. 基于聚类的fMRI数据分析。神经图像。2006;33(2):599–608.
 42. 刘T. 诊断病理学。3rd ed. 北京: 人民医学出版社; 2013年.
 43. 徐、贾子、王、艾、张福、赖美、张爱。通过深度卷积激活特征进行的大规模组织病理学图像分类、分割和可视化。BMC生物信息学。2017;18(1):281.
 44. 凯恩兹, 菲弗, 厄斯希勒。结肠腺的深度卷积神经网络和全变异神经网络正则化的分割和分类。PeerJ. 2017;5:e3874.
 45. 王志明, 王志明, 王志明. 使用深度卷积神经网络进行结肠直肠癌分类—一项实验研究; 2018年.
 46. 张, 张, 张, 张, 张: 用卷积神经网络自动检测浸润性导管癌。见: 医学影像2014: 数字病理学: 2014: 国际光学与光子学学会; 2014: 904103.
 47. 阿劳霍、阿雷斯塔、卡斯特罗、鲁科、阿吉亚尔、埃洛伊、波洛尼亚、坎皮略A. 利用卷积神经网络对乳腺癌组织学图像进行分类。PLoS一个。2017;12(6):e0177544.
 48. 例如, 哈内萨里, 哈比, 哈比, 哈, 我, 我. 乳腺癌的组织病理学图像分类: 一种深度学习的方法。会议日期: 2018年IEEE生物信息学和生物医学国际会议 (BIBM): 12月3–6日. 2018 2018; 2018. p. 2405–12.
 49. 冯YQ, 张林, 莫J. 深度流形保留自动编码器, 用于分类乳腺癌的组织病理学图像。我很我。2020;17(1):91–101.
 50. , 例如, 例如. 基于表示学习的无监督领域自适应, 用于乳腺癌组织病理学图像的分类。双胞胎生物医学协会。2018;38(3):671–83.
 51. 卡里米D, Nir G, 法兹利L, 黑色PC, 戈登伯格L, 萨尔库迪恩SE. 基于深度学习的前列腺癌组织病理学分级图像-多尺度决策聚合和数据增强的作用。生物医学健康信息。2020;24(5):1413–26.
 52. 博斯曼, 卡内罗, 赫鲁班, 德. 世卫组织对消化系统肿瘤的分类。4th ed. 里昂: 国际癌症研究机构; 2010年.

出版商的注意事项

施普林格自然组织对已出版的地图和机构附属机构中的管辖权主张保持中立。