

近年来,人工智能、大数据等技术在医学领域的研究和应用取得了长足进展,但在技术研发、应用落地、产业发展和政策监督等方面,仍然面临着诸多问题和挑战。智能医疗的发展现状如何?如何能更好地在影像识别、电子病历、手术机器人等临床领域满足需求,实现医工交叉、产学研用的融合?本期,三位医学与人工智能交叉学科的专家将从不同的角度带来深度解读。

# 当医疗遇到人工智能

►田捷 北航—首医大数据精准医疗高精尖创新中心主任,中国科学院分子影像北京市重点实验室主任,教授、博士生导师,中国医师协会超声分子影像与人工智能专委会主任。



▼詹剑锋 中国科学院计算所研究员,中国科学院大学教授。



▲赵屹 中国科学院计算所研究员、博士生导师,中国科学院大学健康医疗大数据国家研究院理事长助理。



新华社供图

## 看清医学影像的“雾中花”

□田捷

当前,很多疾病的诊疗都离不开影像工具。那么,基于人工智能和大数据,影像学又能给临床带来怎样的变化呢?

举几个关于肿瘤诊疗的例子。我们知道,目前人类还缺乏对于肿瘤治疗的有效方法。在现有的手段中,影像工具是必不可缺的,但也存在着术前影像不准确、术中治疗凭经验、术后病理滞后的问题。如果借助人工智能,我们能通过影像看到并量化肿瘤演进过程中的分子功能信息,就能大大提高临床诊断的正确性和治疗的有效性。

影像对于临床的意义,是通过从数据、图像到知识的转化来实现的。对于医生们来说,接触到的一般是从图像到知识这个过程,即通过人眼看到影像,根据自己的经验进行判断,进而对患者的病情进行诊断,制订治疗方案。这个过程过去是以人为主,以经验驱动的。而一旦有了人工智能的辅助,这个过程就将发生变化。

在这个领域,人工智能的应用有一套系统的方法,被称作影像组学。它和人类的方式最大的不同是,我们是看图像提取视觉特征,而计算机能看到图像中的高维信息。换句话说,人眼看不到或认为是噪点的元素,里面

可能恰恰含有基因蛋白等在宏观影像上无法体现的微观信息,而这些信息却能够被人工智能获取。

有人说,影像学就像是“雾里看花”,病理学则是“管中窥豹”。而人工智能却可以从影像到病理再到基因,把临床和基础研究结合在一起,用影像逼近病理,让病理前移,提升肿瘤研究和临床治疗的效果。

曾有一位放射科医生给我们提供过一个非常典型的例子。有两个肿瘤患者,年龄相仿,其肿瘤均处于I B期(早中期),肿瘤结构在影像上的呈现也非常相似。但两人的预后却差别很大:第一个患者在5年多的时间里肿瘤没有复发,另一个在1年半的时候复发了。

当时在收治这两个患者时,还没有人工智能的辅助。但现在,当我们用人工智能的方法重新获取两人肿瘤的影像时发现,图像中隐藏着大量可量化的高维信息。这其中,关于肿瘤异质性的相关信息,预后良好的患者显示呈正态分布,预后不好的患者呈杂乱分布。相关结论也发表在影像学权威杂志Radiology(《放射学》)上。

早在2007年,Nature Biotechnology(《自然-生物技术》)杂志曾发表过一篇文章。基于CT图像的28个影像学特征,利用人工智能即可重建78%的肝癌基因表达谱。换句话说,我们拍个片子,就可以借助人工智能挖掘出基因信息。

举一个脑胶质瘤的例子。过去,为了了解肿瘤的类型,我们需要在大脑中进行穿刺,这样做风险可想而知。但现在,我们做个磁共振,就可以预测肿瘤的基因,从而决定采取什么样的治疗方式。

除了帮助影像逼近病理,人工智能还可以在在一定程度上成为预测疗效的工具。按照NCCN(美国国立综合癌症网络)的指南,5厘米以下的肝癌,射频消融和手术切除的效果是差不多的。在临床上,短期效果也确实如此。

而有一位超声科医生很有心,他对所有接受超声造影检查的肝癌患者进行分组,利用人工智能分析、收集影像信息,又对患者进行随访。带着这些数据,他进行了一个非常有临床意义的研究:哪些人适合做手术切除、哪些人适合做消融。

结果显示,有17%做了消融的患者和25%做了手术切除的患者如果互换治疗方式,他们的生存期会更好。而且,随访结果与之前影像数据的预测相互印证。这提示我们,完全可以通过人工智能和大数据为患者的疗效做辅助性的预判。

再来举一个结肠癌的例子。很多结肠癌患者会选择新辅助治疗,治疗后,癌症在病理学上有没有得到完全缓解呢?临床医生通过影像无法确切掌握,患者往往还要再进行手术,拿到病理结果。

按照NCCN指南,对接受肿瘤切除的结肠癌患者,还要对淋巴组织进行清扫,防止癌症复发。其实,在医生判断的结肠癌复发状况中,有70%是假阳性。但在当下,医生也只好带着防止复发的良好初衷,按照指南要求去全部清扫,而清扫淋巴会降低患者的生活质量。

在结肠癌术后,外科医生还会采取相应措施,对癌症的远端转移进行控制。但实际上,远端转移的概率只占百分之二三十,很多患者花了钱,忍受了放化疗的痛苦,为了在他身上不太可能发生的远处转移,而做了放化疗。

上述这些临床上解决不了的问题,由人工智能介入,就可以做到对患者病情的个性化预测。转移概率小的进行随访,概率大的去做放化疗,这样就能在很大程度上避免不必要的医疗资源浪费,减轻患者和家属的痛苦和负担。

作为医生,我们不一定非要去钻研人工智能的方法,也不一定非要去搞分子生物学,我们的优势在于临床。

医生的身边有大量疑难病例,如果我们以临床上没有解决的问题作为切入点,提出问题,寻找解决方法,最后回归临床检验效果,就是一个全链条的人工智能典型临床应用场景,可以解决疗效评估、精准诊断、预后预测等问题。人工智能要在医学上绽放更多生命力,临床是最关键的一环。

虽然人工智能有很多优势,但我也想提醒大家,人工智能也有局限性,也会犯错。我们需要用人工智能来武装医生,但它无法取代医生。方法是掌握在人手上的,掌握了先进的方法,我们对患者的诊断就会更精准,治疗效果就会更好。

总量较少,难以满足医疗和研究使用。而可公开的数据,则可能存在全体的偏向特征。比如,在白种人身上广泛适用的数据,并不确定在黄种人身上应用会不会出现问题。此外,我们也缺少验证手段,难以保证某组数据集的可信程度。

令人诟病更多的是安全问题。大家普遍缺少对新技术的信任感,自己的数据由谁管理?会不会被泄露?我们也一直在研究解决方法,把数据和模型分离,这样既不用担心信息泄露,也方便数据协作。

最后,我想讲的是数据、标准与benchmarks(为不同数据和标准提供评价的体系)这三者之间的关系。

当今,我国非常重视对人工智能中的数据建立规范标准。为什么这是极具意义的一件事呢?比如一家医院,它所处的地区内某种疾病比较常见,但如果在全国范围内看是没有代表性的,那么在这个数据基础上训练的模型就无法适用于其他地方。同样,对机器人的行为制订规范与标准也非常重要。手术机器人如何评价?不能因为它现在的工作是正确的,就能保证它的设计中没有漏洞,未来不会出错。

虽然我们对于技术做了非常多的研究,也发表了很多文章,但对于如何制订标准和规则,则是接下来要思考和实践的要素。

## 今天的医疗为什么需要大数据

□赵屹

大数据,从一般定义上来讲,是指难以被传统数据库管理系统有效存储、管理、分析的复杂数据集。

虽然我们讲了很多年的大数据,可是什么样的量级才能叫作“大数据”,还没有定论。大数据产生的方式不一样,数据的结构也不一样。与传统数据相比,大数据一般有四个V的特征。第一是volume,数据量大;第二是variety,数据类型多;第三是velocity,产生的速度快;最后一个value,数据中包含着重要价值。这是大数据最重要的特征。但同时,它的价值密度低,数据之间的关系太复杂,用少量数据不太可能找到正确规律,因此,才需要庞大的数据作为基础,进行传输、存储、分析、可视化等工作。

要强调的是,作为研究使用的大数据,必须严格控制入选标准和质量。否则一旦给人工智能模型garbage in(提供“垃圾”),就一定会garbage out(结果错误)。

去年,中科院多位院士进行了一系列调研总结,发现我国拥有非常庞大的医疗健康大数据,这对进行生命健康领域原创性的研究,及对发展生物医疗产业来说,都是非常核心的竞争力。但当前的现状却不容乐观:数据碎片化严重,缺乏大数据的质量规范和标准,缺乏共享机制,没有建立大数据的综合管理体系。这一现状最直接的结果之一是患者们在不同医院就医,医院之间很难有信息共享的标准和体系。我们非常期望不同的学科及部门重视当前存在的问题,围绕医疗健康大数据进行深入研究,共同促进科技和产业的发展。

为什么强调大数据的重要性?举一个例子。2019年,有一篇发表在Nature(《自然》)杂志的文章,研究的是精神分裂症与基因遗传位点的相关性。研究者发现,当人组患者数量在6900名的时候,没有在任何基因的位点上发现跟精神分裂症的相关性;当样本人群扩大到23000人,研究发现了7个位点;而当人组人数扩大到6万多人时,可以找到22个位点;人组15万人时,可以找到108个。从这个例子可以看出,数据规模和最终得到的科学结论是正相关的。因此,大数据是开展深度学习的重要基础。

目前,我们所开展的医疗健康大数据的研究主要面向三类,分别是电子病历、临床影像以及生命组学数据。第一类,电子病历,大家都很熟悉。平常我们接触的电子病历都是医生用自然语言写的一段描述,但要让计算机进行学习和分析,对某种疾病进行诊断,还要经过数据结构化处理、数据治理等过程,把其中的关键词提出来,让计算机读懂这些信息。比如,哪些是症状,哪些是诊断。

2019年,Nature Medicine(《自然医学》)杂志发表了一篇文章。作者

应用了临床100多万册儿科门诊病历,通过结构化处理,进行深度学习,实现对儿童呼吸道疾病的诊断。最终,研究呈现出的模型,对儿童上呼吸道疾病和下呼吸道疾病的诊断率,分别能达到89%和87%。

如果再细分的话,对传染性的单核细胞增多症能达到90%的诊断率。第二类是医学影像数据。目前,医学影像数据结合人工智能已给医学发展带来了巨大变革。人工智能最早的应用就是在图片的识别上,而在医疗健康领域,早在2018年的JAMA(《美国医学会杂志》)上,就曾刊发过一篇用深度学习神经网络对糖尿病性的视网膜病变进行诊断的研究文章。

该研究使用超过12万张图片进行了训练学习。最终,完成学习的人工智能模型,在对该类疾病的诊断上,完全可以达到专业眼科医生的水准。当然,在计算机进行学习前,所有图片必须通过专家3~7次的评估和标定。学习结束后,还要通过测试来检测它的准确度。

而从时间和效率上来看,假如一位资深的病理学家看100张病理切片要花30个小时,同样一批切片,人工智能模型可能只需要一个多小时,而且还可以确保一定的准确率。

此外,人工智能在对语音的识别和理解上也能做得很好。未来,人工智能不仅可以理解我们为它输入的信号、手势动作,还可以把输入的多模态信号做一些融合,进行认知,甚至进行情感理解、语义理解最终与人进行多模态的交互。这意味着,未来从接触患者开始,就可以通过人工智能的问诊来获取信息,结构化之后再提供给医生。患者所做的检查,也可以通过数据的融合进行预诊分诊,告诉患者应该挂哪个科,考虑哪方面的疾病等等。

第三类是生命组学数据。在医疗健康领域,目前高通量低成本的基因组测序,不仅可以了解基因组,还可以对转录的RNA进行测序,甚至检测DNA不同的修饰,检测人体微生物组学,进行蛋白组的测序。每个分子组学层面的变化或异常,都会跟一些疾病发生相关。

在对这些大量的数据进行分析后,不但要把不同组学的数据综合在一起,还要把电子病历、影像数据,以及其他可穿戴设备数据等整合进来,应用人工智能技术,最终为每一个人的健康状况给出建议和诊断。

现在,我们正处于大数据驱动的新型医学研究模式之下。我们从以往只靠临床数据去研究疾病,进入到通过人工智能算法,通过计算能力,对医学和疾病进行研究的阶段。虽然人工智能时代已经来临,但它在医学健康领域的应用才刚刚起步。大数据是智能医疗的基础,人工智能只是一类工具。智慧医学的核心还是临床应用场景,其发展的关键仍是交叉学科的人才培养。

## 智慧医疗的边界在哪里

□詹剑锋

智慧医疗的科技包括很多方面的内容,我谨以脑机接口技术和医疗机器人为例,论述其伦理标准和数据库体系。

脑机接口技术分为脑电的采集、信号获取、特征提取、分类判断等方面,但其最重要的特征是它的反馈机制,也就是在收集信号以后,能够返回去作用到大脑的过程。这一点,在神经反馈、恢复瘫痪患者的运动功能等方面有很大的潜力。

脑机接口一般分为非侵入性的、

半侵入性的和侵入性的三种,从可能产生的各种潜在危害的角度来说,对人的侵入程度越轻越好。

虽然脑机接口的概念听起来很新,但其实,早在2004年,巴恩斯-犹太医院的研究人员就已经证明,从大脑表面记录的皮质电图,可以快速准确地控制一维光标。而该领域最新的进展是,通过为一些瘫痪患者植入芯片,在一些相应系统的辅助下,这些患者就能通过运动意图来完成机械臂控制,以及电脑光标控制等任务。

再说医疗机器人。医疗机器人与工业机器人之间,有很多部件是能够共享的。它的优势是,可以突破人

体的极限,完成精细动作,并实现高度自动化。目前,医疗机器人的应用主要还是以辅助为主,包括手术机器人、辅助机器人、康复机器人、服务机器人等。去年暴发的新冠肺炎疫情,也让服务机器人有了更大的发展空间。

随着智慧医疗科技的发展,新的伦理问题也随之出现。首先,数据价值无疑会对智慧医疗的定价产生影响,而这也直接涉及医疗资源的公平分配问题。

其次,脑机接口会不会成为新的“读心术”?什么情况下可以应用?什么情况下不可应用?如何制订准则?这些都显得非常重要。

再次,数据和算法也可能带来伦理问题。现在,数据库系统的数据类型有很多,如诊疗数据、健康信息、物资数据等,数据存储的方式也有本地文件存储、网络存储等。但总体而言,公开的数据集、数据