新型人脸识别软件可助力罕见病诊断

作者：赵广立 李惠钰 来源：中国科学报

本文原地址：https://www.iikx.com/news/progress/3747.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！



新型人脸识别软件可助力罕见病诊断

新型人脸识别软件可助力罕见病诊断。将人工智能(AI)用于医疗辅助诊断早已经不是什么新鲜事，仅AI医学影像辅助诊疗一项就有非常多的应用场景。然而，近日来自美国波士顿一家名为FDNA的数字医疗公司的研究人员大开脑洞，提出用深度学习技术仅仅通过看面相，就能帮助医生识别罕见基因疾病，并付诸实践。他们的研究于1月7日登上了《自然—医学》杂志，引起业内关注。

中医讲究望闻问切，其中望即有望五官知五脏望痰辨病等含义。如果AI能学会这本领，相信很容易获得人们青睐。不过，深度学习是一项比较吃数据的技术，而罕见病恰恰是临床医学上数据贫乏的领域。FDNA公司研究人员试图依靠AI望五官辨识基因异常引发的罕见病，行得通吗?

猜测准确率达65%

发表在《自然—医学》上的这篇文章提出了一种深度学习算法Deep Gestalt，它通过一款名为Face

2Gene的App在智能终端发挥作用。据论文介绍，该App依靠深度学习算法和类脑神经网络来区分人类照片中与先天性和神经发育障碍有关的独特面部特征，然后利用从照片中推断出的模式，定位出可能的诊断结果，并提供可能的综合征遗传病选项。

在实验开始之前，FDNA公司首席技术官Yaron Gurovich带领研究者们首先训练人工智能系统来区分Cornelia de Lange综合征和Angelman综合征，这两种疾病患者都有区别于其他疾病的明显面部特征;同时，研究人员还训练该模型对Noonan综合征的不同基因形式进行分类。接下来，研究者们给该算法模型输入了涵盖216种不同综合征的17000多张确诊病例的图像。

论文显示，在对502 张不同图像上的新面孔进行测试时，Face2Gene的最佳诊断猜测准确率达到65%。如果考虑多个预测结果，则Face2Gene 的top10准确率可以达到约90%。

这个结果显然有其一定的医学意义。FDNA 首席医疗官、该论文的共同作者之一Karen Gripp使用该算法帮助诊断了一名4岁女童的Wiedemann Steiner 综合征——这是由一种叫作KMT2A 的基因变异引起的罕见疾病。Gripp在研读了相关病例报告之后得知，该疾病会导致儿童牙齿过早生长。

由于年龄较小，这名小患者除了掉落大部分乳牙并长出多个恒牙，很多典型的躯体症状尚未显现。Gripp将该女童的照片上传至Face2Gene后，软件上出现了Wiedemann Steiner综合征;接着，Gripp 用靶向DNA 测试进一步确诊了这一结果。

虽然Gripp最终依靠的是DNA检测进行确诊，但这名医生认为，该AI方法可以帮助医生缩小病症范围，节约了昂贵的多基因检测费用。

目前，Face2Gene这款App处于开放状态，全世界的医疗专家都可以免费使用它。当然，这也是FDNA公司想要获取更多数据的策略，它需要更多数据来训练模型，使Face2Gene能够驾轻就熟地帮助医生辨别患者可能的基因遗传病。

Yaron Gurovich介绍，现在该项目的数据库中已有大约15万张照片。而随着更多医疗专家将病人的照片上传到该App，该项目的准确率也得到略微提升。

质疑：结论并不令人信服

并非没有质疑。在记者的采访中，长年致力于罕见病治疗研究的中科院生物物理研究所研究员刘光慧告诉《中国科学报》，罕见病种类太多，AI难以达到100%识别。

一般的罕见病都有特殊面容，比如儿童早衰症及成人早衰症患者，面容都很特殊，通过AI识别有一定道理，这是一个很好的方向。刘光慧转而说道，但也有很多罕见病没有特殊面容或面容特征不明显，那么AI可能还不能做到100%识别。

言外之意，如果在上述4岁女童案例中，换一种面容特征不那么明显的罕见病来测试，Face2Gene则可能会失灵;而如果Face2Gene一开始就判断错了，那么最终到底能不能省下多基因检测费，则又另当别论了。

更重要的是，该论文中的结论并不十分令人信服。

502个图的测试集，分了200个类，top10 的准确率达到90%——假设数据平均，每个类仅两张图片可测试，这样的结果我不认为能得到‘ 识别遗传病准确率91% ’的结论。有质疑者认为，该论文结论稍显草率。

更有专业人士指出，该结论有过拟合之嫌，在实测中不一定可以达到这样的准确率。通常，在一些统计中，为了得到一致假设而使假设变得过度严格，就会出现过拟合，而要避免过拟合的发生，通常需要使用增大数据量和测试样本集的方法对分类样本进行评价。

对此，中科院自动化所研究员、生物识别与安全技术研究中心主任李子青(注：该文引用了李子青团队的一篇深度学习研究论文)对《中国科学报》记者谈了他的看法。他指出，这篇文章提供了三个实验结果：其中两个实验是从已确诊综合征的人脸图像中，识别出一种特定的综合征;另一个是区分Noonan综合征的两个子类型。可喜的是，三个都超过了门诊专家的准确性。但这篇文章所解决的问题，是从已确诊的人脸图像进一步分类或区分子型;从深度学习和模式识别角度来看，都不算难。

李子青认为，真正有挑战的是综合征筛查，即从普通人群发现有综合征的人，并对大规模人群诊断达到一定的准确度。这个问题更加有临床诊断意义，并且技术上比该文所处理的问题的难度要大很多。

对于深度学习算法而言，李子青认为只要有数据，总可以训练出一个东西，而至于效果如何，首先是取决于算法——在合理算法基础上，就是拼数据和计算资源。至于如何判断实效?则要看其推广性及泛化性。他表示认同记者的这一观点：罕见病病例少，数据难获取，模型的推广性不容易保证，这种方法的大众普查中应用还不成熟。

技术风险仍存

一般而言，只有训练数据集够好，算法才足够好用。从这个评价尺度上，Face2Gene的技术存在风险。对此，牛津大学计算生物学家Christoffer Nellker很是认同。他认为，尤其是涉及到那些全球患者人数极少的罕见疾病时，数据集的封闭化、商业化会威胁到这项技术的主要潜在优势。

此外，训练数据过于集中导致的种族偏见(大部分为白人)也是一大问题。2017年的一份儿童智力障碍研究表明，Face2Gene
对唐氏综合征的识别率在比利时白人小孩中为80%，而在刚果黑人小孩中仅为37%。对此，Yaron Gurovich意识到这个问题需要解决，但他也认为随着训练数据集变得更加多样化，算法对非洲面孔的识别准确率会随之提升，偏见会越来越少。

除了数据量的问题外，数据维度也非常重要。汇医慧影创始人兼CEO柴象飞在接受《中国科学报》记者采访时说，有时更多维度的数据比简单更大的数据更能帮助算法模型更精准地预测结果。

不过，科大讯飞智慧医疗事业部总经理陶晓东向《中国科学报》提出了类似算法在临床上常被忽略掉的另一个技术风险，这一波人工智能过度依赖于数据，忽略了很多数据之外的信息。

深度学习解决问题的基本思路，到目前为止还都没有什么太大的突破。陶晓东对记者说，在医疗领域里尤其如此，你不可能像ImageNet那样获得几百万、上千万的训练数据。因此，最基础的医学理论，应该被考虑在内。

连基本的解剖信息都没有用在深度学习的框架里。陶晓东说，这是值得目前如火如荼的AI+医疗深思的。

柴象飞对此也表示赞同，他对记者说，一些基于经验或常识的推理对于模型更精准地接近真相也非常重要，这或许对于数据样本偏少的罕见病模型意义更大。

论文相关信息：DOI:10.1038/s41591-018-0279-0

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发