

doi:10.6040/j.issn.1673-3770.0.2023.470

深度学习技术在咽喉内镜应用中的研究进展及前景分析

程卓^{1,2}, 梁辉², 邢鲁民³

1. 山东第一医科大学(山东省医学科学院)耳鼻咽喉头颈外科, 山东 济南 250021

2. 山东第一医科大学第一附属医院(山东省千佛山医院)耳鼻咽喉头颈外科, 山东 济南 250014

3. 山东第一医科大学第一附属医院(山东省千佛山医院)信息中心, 山东 济南 250014

摘要: 深度学习的出现对医疗水平特别是医学检查的进步起到了巨大的推动作用, 耳鼻咽喉头颈外科部分领域亦因此获益, 基于深度学习的咽喉内镜检查数据分析领域近 5 年来做出了极有成效的尝试。本文以近 5 年基于深度学习的咽喉内镜应用及相关研究作为讨论主体, 分析该领域的研究进程并将其发展阶段划分为神经网络萌芽阶段、神经网络与医学的交融和适用性发展的神经网络阶段三个阶段; 以临床、样本信息、其他三个方面分别讨论现阶段研究瓶颈, 并阐述了未来可能的解决方案及发展前景, 指出了当前咽喉内镜中深度学习应用的主要障碍, 并给出了未来多中心研究、多任务学习、高水平信息数据采集等可能的发展趋势展望。

关键词: 咽喉内镜; 人工智能; 深度学习; 耳鼻咽喉头颈外科学; 计算机辅助诊断

中图分类号: R762; TP18 **文献标志码:** A **文章编号:** 1673-3770(2026)01-0112-08

引用格式: 程卓, 梁辉, 邢鲁民. 深度学习技术在咽喉内镜应用中的研究进展及前景分析[J]. 山东大学耳鼻喉眼学报, 2026, 40(1): 112-119. CHENG Zhuo, LIANG Hui, XING Lumin. Research progress and prospect analysis of deep learning technology in the application of pharyngeal and laryngeal endoscopy[J]. Journal of Otolaryngology and Ophthalmology of Shandong University, 2026, 40(1): 112-119.

Research progress and prospect analysis of deep learning technology in the application of pharyngeal and laryngeal endoscopy

CHENG Zhuo^{1,2}, LIANG Hui², XING Lumin³

1. Department of Otorhinolaryngology & Head and Neck Surgery, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan 250021, Shandong, China

2. Department of Otorhinolaryngology & Head and Neck Surgery, The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, Jinan 250014, Shandong, China

3. Department of Information, The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, Shandong, Jinan 250014, Shandong, China

Abstract: The emergence of deep learning has played a huge role in the promotion of medical quality, especially in the advancement of medical examination, and some areas of otolaryngology and head and neck surgery have benefited from it. On the basis of deep learning, the field of endoscopic analysis of the pharynx and larynx has made very effective attempts in the past five years. This article discusses the research and related research based on deep learning in pharyngeal and laryngeal endoscopic application in the past five years, analyzes the research progress in this field, and divides its development stage into three stages: the stage of neural network germination, the blending of neural network and medicine, and the development of neural network applicability. Based on clinical, sample information and other three aspects, this article discusses current research bottlenecks, expounds possible solutions and development prospects in the future, points out the main obstacles in the application of deep learning in current pharyngeal and laryngeal endoscopic research, and gives a possible development trend outlook in multiple aspects such as multicenter research, multitask learning, high-level data information collection in the future.

Key words: Pharyngeal and laryngeal endoscopy; artificial Intelligence; deep learning; Otorhinolaryngology head and neck surgery; Computer-assisted diagnosis

喉癌、鼻咽癌等头颈恶性肿瘤的早期筛查与发现是此类疾病的治疗关键之一, 咽喉内镜检查是目

前发现、检出此类疾病以及相当一部分咽喉疾病的主要方式, 也是基本手段。提高咽喉内镜的疾病检

收稿日期: 2023-12-15

通信作者: 梁辉。E-mail: onlinelh@163.com

出效率和能力,是提高头颈恶性肿瘤及部分常规咽喉疾病临床治疗预后的关键点之一。人工智能(artificial intelligence, AI)的主要目标是使机器能够胜任一些通常需要人类智能才能完成的复杂工作,机器学习(machine learning, ML)的研究极大推动了人工智能的进步,而深度学习(deep learning, DL)则是被引入以帮助机器学习更接近其最初的目标-人工智能的崭新研究方向,其在医学领域尤其展现出极强的适用性及巨大的开发潜力。随着多组学数据的频繁使用,医疗检查数据规模迅速膨胀,深度学习已成为医疗数据分析的强大助力。通过对大量数据进行分析,提取有医学意义的抽象内涵并将其转化为有效医学提示,AI 可以显著提高疾病诊疗的效率、准确性和安全性^[1-2],从而深刻影响治疗结果和患者预后及愈后生活质量^[3]。

咽喉内镜是耳鼻咽喉头颈外科常规检查中的一种,可用于日常筛查咽喉良恶性病变、癌前病变,以及实施部分咽喉疾病的病理活检。目前结合深度学习的咽喉内镜检查主要应用于咽喉病变检测分类,癌前及恶性病变的风险评估,相关研究呈逐年上升趋势。咽喉部病变范围较广,最严重的疾病莫过于咽喉恶性肿瘤。由于咽喉恶性肿瘤早期发现治疗的生存率较晚期有明显改善^[4-5],且早期检出癌前病变和恶性病变对于保护咽喉所承担的器官功能如:呼吸、吞咽、发音等尤为重要,特别是咽喉恶性肿瘤的治疗如咽部切除、颈部淋巴结清扫以及放疗常对咽喉器官功能有较大影响,较早期的病变发现与医疗干预可保证相当程度患者咽喉功能的留存和术后社区适应能力及患者个人尊严的保护^[6]。因此这一领域的深度学习应用研究意义重大而深远。

本次综述以近 5 年发表在“Web of Science、PubMed、Cochrane Library、Embase、arXiv”等数据库的部分研究及综述为基础,讨论咽喉内窥镜结合 DL 在临床中的应用研究进展,分析目前存在的部分瓶颈,提出未来发展趋势。

我们检索前述数据库,初步筛选条件为:重复研究、非 DL 应用相关研究、咽喉内窥镜检查不相关研究,二次筛选去除结果中接触式内窥镜、睡眠内窥镜相关研究,对剩余 41 项咽喉内窥镜检查相关研究进行了分析,其主要研究方向分布,见图 1。这里需要一提的是,深度学习相关研究本身仍处于开发阶段,因此部分研究并不能确切予以界定为单纯病变检测研究

还是方法/技术开发,部分研究方向界定可能存在争议。相关 DL 研究讨论涉及概念、参数汇总可参照 Sampieri 等^[7]综述。

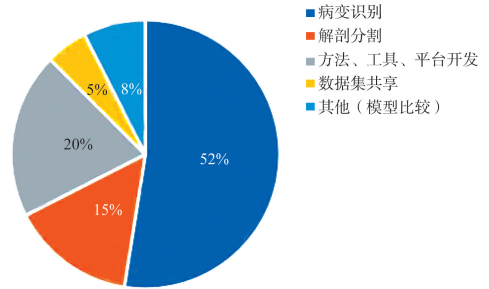


图 1 41 项咽喉内窥镜检查相关研究方向
Figure 1 41 research directions related to pharyngeal endoscopy examination

1 研究进展

自 2006 年 DL 概念提出^[8],其深刻改变了人工智能在医学领域的应用结构。医学领域相关研究不胜枚举,最开始集中出现于计算机视觉(CV)领域^[9],特别是图像分析领域,以完成相对较简单的分类任务为主。

在耳鼻咽喉头颈外科领域,现阶段 DL 的应用目的主要包括区分正常声带与病变声带、咽喉病变的检测分类、鼻咽病变分类等。2017—2018 年间,基于人工智能的咽喉内镜自动图像分析登上医学人工智能应用舞台,基于 DL 的人工智能医学融合开始踏入初步尝试阶段。Mascharak 等^[10]首次将机器学习应用于咽喉内窥镜检查的自动评估中,目标是检测口咽癌,这是咽喉内窥镜领域的首次应用尝试,虽然样本数仅 30 例,但其团队开发的分类模型仍表现出优秀的开发潜力与良好的适用性,表明人工智能在耳鼻咽喉科学领域有极大的应用可能。

以此为基础,DL 技术迅速跟进。依据近 5 年相关数据库中检索到的研究文献,2018、2019 两年间,基于 DL 的研究纷至沓来。研究样本人群主要分布于中国、欧美地区,研究方向主要集中于口、咽恶性肿瘤的识别检测^[11]、声带识别与运动评估^[12]。

2 数据集的发展

DL 在咽喉内窥镜检查的应用发展速度极为迅速,首先表现为数据集的增大和开源数据集的出现,见表 1。

表 1 BAGLS 与 Laryngoscope8 数据集对比
Table 1 Comparison between the BAGLS dataset and the Laryngoscope8 dataset

数据集	发表年	样本来源	是否标注	标注	内容	部位	作者
BAGLS	2020	欧盟和美国的七个机构的临床样本	有	临床诊断	包含 59 250 个高速视频内窥镜帧,数据包含发音健康和紊乱的样本,表现为功能性和器质性发音障碍,且被不同医生标记。	声门区	Gómez 等 ^[13]
Laryngoscope8	2021	解放军总医院第六医学中心临床样本	有	病理诊断及临床诊断	50 个独立个体的 3 057 张图像,且图像被不同医生标记(包括七个病理标签和一个正常标签)	喉部区域	Yin 等 ^[14]

近 5 年相关研究数量发表趋势,见图 2。

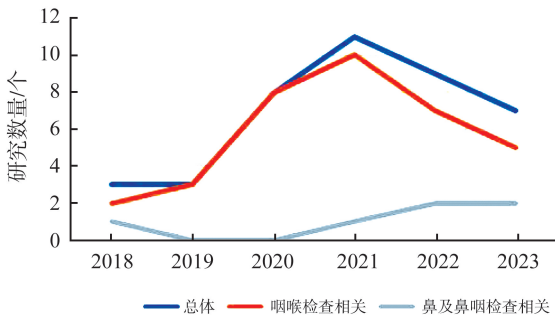


图 2 近 5 年 DL 咽喉内镜相关研究趋势图
Figure 2 Trend chart of DL pharyngeal endoscopy related research in the past 5 years

同时研究中样本数量也呈极高速增长趋势,陆续有超过 10 000 张^[15-18]甚至 20 000 张^[5,13,19-21]图像样本的研究出现,大型数据集的相继投入也得益于 BAGLS、Laryngoscope8 大型数据集的分享使用。Gómez 等^[13]于 2020 年公开 BAGLS 数据集致力于推动声门自动分割纳入日常临床常规,并刺激相关领域创新、合作和研究进展;随后 Yin 等^[14]于 2021 年公开 Laryngoscope8 数据集,希望借此刺激推动喉部疾病自动分类及相关研究的发展。两个数据集的公开代表着新的研究阶段的开启,此后的许多研究都是基于或部分基于这两个数据集,如 Kruse 等^[21]以及 Kist 等^[22-23]研究。开源数据集的出现推动了咽喉内镜领域各项研究发展进程,部分受到数据集相关因素限制的研究得以借助开源数据集进行实验与探索。

3 模型的进步

随着大量研究的投入以及数据处理模型的优化升级,应用于咽喉内镜的 DL 模型的多数研究呈现与初期应用简单机器学习相比更加优秀的效能,2022 年 Zurek 等^[6]发表的一项 META 分析中,人工智能模型的总体准确性很高(从 0.806~0.997),并表现出良好的敏感性及特异性。

从学习模型的进步以及医疗数据采集器材的迭

代中受益,咽喉内镜的智能应用不再限于初期单纯、局部病变的检测和识别,开始向着多种疾病的筛查与疑难疾病(如声带运动异常^[24-25])的辅助诊断转变。在样本类型方面,见图 3(部分研究图像样本、视频样本均涉及),从 2019 年开始,对于视频样本的研究也开始崭露头角。

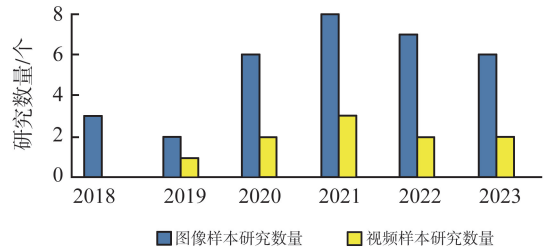


图 3 2018—2023 年 DL 相关研究样本类型
Figure 3 Sample types for DL related studies from 2018 to 2023

咽喉内镜领域常用的机器学习算法有神经网络、支持向量机和 k 近邻等,并且在使用神经网络的研究中,绝大多数是在最近 5 年发表的。我们把这一阶段分为神经网络萌芽阶段、神经网络与医学的交融和适用性发展的神经网络阶段。

3.1 神经网络萌芽阶段

Mascharak 等^[10]和 Moccia 等^[26]在 2018 年发表的研究分别使用了贝叶斯分类机和支撑向量机完成了分类任务,且均使用了包含 NBI 窄带内镜(narrow band imaging, NBI)图像——一种有利于早期发现喉癌及其前体病变的描述肿瘤特异性新生血管的新成像技术^[27-29]学习集。尽管样本数量不足且应用模型尚处尝试阶段,前者(准确性 65.9%、敏感性 66.8%、特异性 64.9%)和后者(信息帧的 91% 召回率)仍表现出不俗的性能,这两项研究不仅给后续的 NBI 相关研究^[30-32]提供了良好思路,也为迎接神经网络的来临打下坚实的基础。

3.2 神经网络与医学的交融

2018 年至 2021 年间,借着信息工程领域“信息革命”带来的技术提升的推动力,空前的计算机算力提升推动神经网络的医学尝试迅速崛起。这期间

的研究以分类和分割为主,研究使用的网络模型种类繁多,其中包括不限于完全卷积网络(fully convolution networks, FCNs)、RCNN、F-RCNN、InceptionNET、VGGNET、U-NET、ResNet等。如Li等^[5]一项研究应用了完全卷积网络(FCNs)构建学习模型eNPM-DM用于检测鼻咽恶性肿瘤并获得了总准确率为88.7%(95%可信区间(CI)为87.8%~89.5%)优秀效能,后来2019年Lin等^[33]发表的研究成功将FCN与标示感兴趣区(region of interest, RoI)结合应用于视频样本的分析,其团队借助学习模型提取了重要的喉部运动信息,证明其可用于辅助喉闭合的诊断。

这一阶段的中后段(2019~2021年间)以InceptionNET、ResNet以及U-NET模型为主,VGGNET紧随其后,其中InceptionNET、ResNet模型分类任务使用较多,且在部分研究中常被作为候选模型用以对比效能,如Cho等^[20],Patrini等^[34]以及Cho等^[35];在Cho等^[20]中识别声带的ResNet和Inception分别达到0.985和0.971的特异性和0.865和0.892的敏感性,而在Patrini等^[34]研究中构建的ResNet和Inception两者的表现则低于VGG16,这也说明了对于不同应用条件模型适应性存在较大差距。

另一方面,U-NET、ResNet多在分割任务中被采用,如Laves等^[36]一项研究中U-NET被用于喉软组织的语义分割,后来借助U-NET及其扩展模型Fehling等^[16]用于声门区的自动分割,Parker等^[37]用于拔管后患者喉部病变的识别,Kist等^[38]则将模型用于搭建一个名为OpenHSV的开放研究平台,其研究证明该平台提供了一个相对有效的、流程式的高速视频内窥镜数据采集和分析工具。ResNet及其衍生模型被用于声带、气管的分类识别^[39]、喉病变的识别^[19],鼻息肉和倒置乳头瘤的鼻内镜图像自动检测^[40],优化咽喉内窥镜视频帧自动选择的学习策略^[34],帮助高速视频内窥镜(high-speed videendoscopy, HSV)开发相关数据分析工具软件^[22]以及基于医院采集的诊断图像的自动检测喉部肿块算法开发^[41]。

值得一提的是,借助提高的精度与速度以及端到端的目标检测框架(F-RCNN)和能够处理物体检测的同时还能处理语义分割问题且能够满足图像语义分割的准确度要求(MASK-RCNN),F-RCNN以及MASK-RCNN在这一阶段开始被重视,如2020年Matava等^[39]进行的一项研究开发了基于F-RCNN的一种能够在视频喉镜检查或支气管镜检查时实时分类声带和气管解剖的模型,在其研究中

与同时构建的ResNet相比,基于F-RCNN的Inception模型特异性稍低,灵敏度更高,总体置信度更高,因此凭借优秀的持续性能和更好的视频效果Inception模型被采纳并进一步训练;另一边,Kono等^[42]于2021年进行的一项研究则基于MASK-RCNN开发了一种具有相对高灵敏度和一般程度特异性的用于实时诊断咽癌的人工智能系统,其AI系统每张图像的转换速度为0.03s,检测癌症的敏感性、特异性和准确性分别为92%、47%和66%,虽然其限于应用目的特异性相对较差,但仍然在敏感度与准确性上体现出其模型在视频分析应用中一定程度的开发潜力。

3.3 适用性发展的神经网络

这一阶段(2022—2023)实际上是与前一阶段相对应的,在这一阶段与之前最明显的不同之处在于相当数量的研究采用了并非当时开源数据集训练效果最好的模型,而是更加注重研究需要来选择模型甚至开发适应研究需求的特殊模型。

Ay等^[43]2022年发表的一项研究中通过多个分类法(KNN、SVM、DT等)以及CNN的两种结构(3-Channel CNN结构和1-Channel(灰色比例)CNN结构)对比证明了CNN入路,特别是3-Channel的CNN结构,达到了研究中最高的精度水平,其研究表明使用适当的图像处理技术与合适的DL网络可帮助研究人员在识别鼻息肉方面获得高准确性的建议。

在这一阶段研究模型繁杂,如YOLOv5及其相关系列模型^[18,44-45]被用于鼻咽癌与喉癌的检测分类,DenseNet169用以识别舌癌患者^[46],Xception用于确定咽喉解剖位置^[47]。

另一方面,在深度学习应用于咽喉内窥镜检查的初期,限于器械条件、数据分析能力、模型适应性等多方面原因,对于运动状态的分析如声带运动是相对较困难的。近2年来,随着检查设备的更新、模型优化的发展以及新型学习“迁移学习”的引进如Zhao等^[48]的研究,声带运动分析、声带运动障碍等问题迎来了新的解决办法。

2022年,Kist等^[23]基于其团队自行优化的U-NET模型架构提出了新的声门自动分割技术,证明了单一的通道在编码器-解码器结构的潜伏空间足以进行声门区域的重建。后来,2023年发表的两项研究^[21,49]均基于U-NET的优化扩展架构,Kruse等^[21]开发了多任务的深层神经网络来检测声门中线,Pedersen等^[49]提出了一种精确计算声带间距的新方法,在特定的位置上用声带间界线作为计算声

带间距的基线,实现声门的间隙的定位、定量,Sakthivel 等^[50]则引入了新的入路 ODL-VFDDC 以识别高速视频内窥镜下振动时的声带边界。这些研究表明,随着外部条件的升级换代,在咽喉内镜领域的应用研究逐渐突破了动态研究的障碍,向更加全面的研究领域迈进。

虽然尚处研究探索阶段,神经网络应用一定程度上提高了部分疾病检测效率。但在某些研究方面仍存在极大的开发潜力,如我们支持 Sampieri 等^[7]综述中提出的观点:对于部分研究,预测任务越复杂,预测的准确性就越低这一观点。事实上,对于喉部病变分类任务的复杂性,训练需要数量庞大和特征多样化的数据集。一定程度上可以认为,单独引入二元分类工具来筛选哪些病变具有高恶性风险是深度学习临床实践的具有深远意义的转折点,同时我们认为,下一里程碑可能在多疾病筛选与多组学分析这一实践领域,届时深度学习的能力与用途将进一步扩展。

4 讨论

4.1 目前研究的瓶颈

深度学习的医学应用是建立在数据分析的进步之上的,可以说是技术的革新成就了“深度医学”。虽然在多个医学领域中,深度学习展现出前所未有的数据处理分析能力和较为优越的实用性,研究瓶颈也逐渐明显,主要瓶颈集中在数据集的数量与质量(特别是标签数据集)、模型优化水平和相应高规格仪器设备的缺乏,模型优化的水平是信息技术进步与医学多组学发展交融的标志,其不仅受到信息工程领域的推动和制约也受到医学数据处理与应用水平的影响,而仪器设备的发展水平在一定程度上影响着实际研究样本的质量,间接影响着训练集;另将其他瓶颈(包含数据集)分为临床、样本信息及其他三方面,在下文详细分析。

临床方面主要包括:①数据采集人员专业水平的影响,如消化内窥镜、电子喉镜图片、影像数据采集质量极易受到操作者专业水平的影响,较优秀的操作者或专业医师在条件限制较多的情况下甚至可以规避拍摄角度、体液或分泌物、烟雾等因素对检查、操作和数据图像采集的影响;②采集人员疲劳度,临床人员的疲劳程度将在一定程度上影响样本采集水平;③样本采集过程中的患者、受检查者的配合程度,配合度较好的样本其清晰度,视角都对训练效果有一定影响;④人体解剖结构影响,如解剖变异、咽喉内镜检查时部分结构异常(如会厌上翘角

度)会影响数据质量;⑤体液因素对数据采集的影响,部分已罹患相关疾病或部分术后患者如上呼吸道感染表现强烈时鼻分泌物较多或气管切开术后患者的痰液较多可影响图像采集。

样本信息方面:①训练集数据量少,特别是有标签训练集数据量少(针对目前主流研究策略,即有监督学习和半监督学习),数据集的大小在一定程度上影响着模型学习最终水平;②地域因素的差异可在一定程度上影响训练集的训练成果,导致模型适应性的降低。此类因素包括地区发展水平(一线城市与偏远地区生活水平的差距、工作环境差异等因素可影响采集数据)、医院设备条件(如三级甲等医院与下级医院的设备差距较大,设备限制采集数据如图片分辨率、视频帧率、清晰度以致影响整体训练效果),同时,患者体质及民族特殊性等因素也对训练数据有一定程度的影响;③针对高速视频、照片的高质量影像捕捉、照片抓取能力的欠缺。由于患者并不能时刻保持静止状态,而咽喉部位的运动存在一定的不可控性,因此迅速而准确地抓取瞬时照片以拍摄符合要求的病变或可疑区域的高清晰度照片是必要而迫切需要的。

其他因素:①受法律及伦理要求等多方面约束,各研究中心无法将医学数据建立相关的大数据库,数据限制影响研究进展(值得高兴的是,近年来这层壁垒在逐渐瓦解,如 Yan 等^[51]的研究);②数字鸿沟,即部分人群对于深度学习等先进技术理解不足,在临床研究阶段不愿接受相关试验参与或拒绝配合深度学习应用的检查,导致研究模型推广和进展受阻;③国家间科技水平差异、工业制造业等制造水平差距导致的医学发展程度,医疗器械设备的应用水平,医学应用、医学发展投入的限制。

4.2 未来

深度学习在咽喉内镜领域发展突飞猛进,上述部分影响因素已有许多解决策略,如可借助更加优化的图片处理模型可将训练集数量增加,以增强训练效果;通过联合多中心研究以规避地域性差异因素引起的数据问题;同时,多任务模型(如分类、识别模型的共同参与)的应用可以帮助完成更多的目标或较复杂的任务,并提供更强的适用性与实用性;先进的设备或数据采集流程与技术技巧也可帮助更好得采集数据(如通过在电子咽喉内镜镜头上涂擦水溶性凝胶以避免镜头起雾)。

实际上,未来首先应解决的难题仍然是训练集与模型优化这两个重点。随着人工智能系统的更新迭代,更先进的数据分析能力有望帮助去噪(如图

片或视频中的烟、雾、血、分泌物等),从而提高训练集的样本质量,其次以国家联系、大学或医院为核心纽带推动的多中心研究可以推进研究数据(主要是临床数据)的数量提升和质量优化,促进研究普遍性与代表性的提升,同时,线上数据共享,能够提高疾病检出和推进大数据库的建立。在研究本身,提升数据集的样本数量和标签样本质量,可以在一定程度解决训练集瓶颈,这也是我们团队接下来的研究努力方向,通过提升样本数量与标签样本质量,同时积极与数据信息处理专家团队合作,提高模型适用性与实用性,以期获得更加适合医疗体系的辅助诊断系统。模型优化方面,增加医学、信息工程领域交流、合作学会、讲座等,帮助搭建交叉领域沟通桥梁和交流机遇,让双方理解彼此需要什么,技术发展的空缺在何处,可以为深度学习的医学应用模型发展提供更多的机会和动力。

其他方面,影像、电子病历以及各种辅助检查输出的数据分析优化和辅助诊断的多组学因素参与可以尽量减少模型偏倚,包括部分研究中的主观偏倚,而多个地区的医生下乡、线上医疗的推广和医疗知识普及等惠民政策可以促进数字鸿沟的消除。

基于深度学习的医疗辅助诊断系统可以帮助临床大夫脱离枯燥繁复的重复检查,帮助大夫节省出更多的精力着眼于疑、难、重病患者的诊治及疾病的研究,可以通过迅速而及时的辅助提示帮助医生提高诊断决策的果断力,补全信息获取缺口,减少漏诊,并实现较前更有效、及时的活检效率,推动疾病的早发现、早治疗,在多个医学领域,发挥更加重要的作用。

参考文献:

- [1] Zhong NN, Wang HQ, Huang XY, et al. Enhancing head and neck tumor management with artificial intelligence: integration and perspectives [J]. *Semin Cancer Biol*, 2023, 95: 52-74. doi:10.1016/j.semcancer.2023.07.002
- [2] 朱志玲, 李松, 管国芳. 人工智能在耳鼻咽喉头颈外科的运用及展望[J]. *山东大学耳鼻喉眼学报*, 34(2): 115-120. doi:10.6040/i.issn.1673-3770.0.2019.598
ZHU Zhiling, LI Song, GUAN Guofang. Application and prospect of artificial intelligence in otolaryngology [J]. *Journal of Otolaryngology and Ophthalmology of Shandong University*, 34(2): 115-120. doi:10.6040/i.issn.1673-3770.0.2019.598
- [3] 刘佳钰, 樊慧明, 邹游, 等. 人工智能在鼻咽癌诊断与治疗中的应用研究进展[J]. *山东大学耳鼻喉眼学报*, 2023, 37(2): 135-142. doi:10.6040/j.issn.1673-3770.0.2022.089
LIU Jiayu, FAN Huiming, ZOU You, et al. Research progress on the application of artificial intelligence in the diagnosis and treatment of nasopharyngeal carcinoma [J]. *Journal of Otolaryngology and Ophthalmology of Shandong University*, 2023, 37(2): 135-142. doi:10.6040/j.issn.1673-3770.0.2022.089
- [4] Abe S, Oda I. Real-time pharyngeal cancer detection utilizing artificial intelligence: journey from the proof of concept to the clinical use [J]. *Dig Endosc*, 2021, 33(4): 552-553. doi:10.1111/den.13833
- [5] Li CF, Jing BZ, Ke LR, et al. Development and validation of an endoscopic images-based deep learning model for detection with nasopharyngeal malignancies [J]. *Cancer Commun*, 2018, 38(1): 59. doi:10.1186/s40880-018-0325-9
- [6] Żurek M, Jasak K, Niemczyk K, et al. Artificial intelligence in laryngeal endoscopy: systematic review and meta-analysis [J]. *J Clin Med*, 2022, 11(10): 2752. doi:10.3390/jcm11102752
- [7] Sampieri C, Baldini C, Azam MA, et al. Artificial intelligence for upper aerodigestive tract endoscopy and laryngoscopy: a guide for physicians and state-of-the-art review [J]. *Otolaryngol Head Neck Surg*, 2023, 169(4): 811-829. doi:10.1002/ohn.343
- [8] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507. doi:10.1126/science.1127647
- [9] Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare [J]. *Nat Med*, 2019, 25(1): 24-29. doi:10.1038/s41591-018-0316-z
- [10] Mascharak S, Baird BJ, Holsinger FC. Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning [J]. *Laryngoscope*, 2018, 128(11): 2514-2520. doi:10.1002/lary.27159
- [11] Tamashiro A, Yoshio T, Ishiyama A, et al. Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks [J]. *Dig Endosc*, 2020, 32(7): 1057-1065. doi:10.1111/den.13653
- [12] Wang YY, Hamad AS, Lever TE, et al. Orthogonal region selection network for laryngeal closure detection in laryngoscopy videos [C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Montreal, QC, Canada. IEEE, 2020: 2167-2172. doi:10.1109/EMBC44109.2020.9176149
- [13] Gómez P, Kist AM, Schlegel P, et al. BAGLS, a multi-hospital benchmark for automatic glottis segmentation [J]. *Sci Data*, 2020, 7(1): 186. doi:10.1038/s41597-020-0526-3

- [14] Yin L, Liu Y, Pei MT, et al. laryngoscope8: Laryngeal image dataset and classification of laryngeal disease based on attention mechanism [J]. *Pattern Recognit Lett*, 2021, 150(C): 207-213. doi:10.1016/j.patrec.2021.06.034
- [15] Xiong H, Lin PL, Yu JG, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images [J]. *EBioMedicine*, 2019, 48: 92-99. doi:10.1016/j.ebiom.2019.08.075
- [16] Fehling MK, Grosch F, Schuster ME, et al. Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network [J]. *PLoS One*, 2020, 15(2): e0227791. doi:10.1371/journal.pone.0227791
- [17] Dunham ME, Kong KA, McWhorter AJ, et al. Optical biopsy: automated classification of airway endoscopic findings using a convolutional neural network [J]. *Laryngoscope*, 2022, 132(Suppl 4): S1-S8. doi:10.1002/lary.28708
- [18] Wang SX, Li Y, Zhu JQ, et al. The detection of nasopharyngeal carcinomas using a neural network based on nasopharyngoscopic images [J]. *Laryngoscope*, 2024, 134(1): 127-135. doi:10.1002/lary.30781
- [19] Ren JJ, Jing XP, Wang J, et al. Automatic recognition of laryngoscopic images using a deep-learning technique [J]. *Laryngoscope*, 2020, 130(11): E686-E693. doi:10.1002/lary.28539
- [20] Cho WK, Choi SH. Comparison of convolutional neural network models for determination of vocal fold normality in laryngoscopic images [J]. *J Voice*, 2022, 36(5): 590-598. doi:10.1016/j.jvoice.2020.08.003
- [21] Kruse E, Dollinger M, Schutzenberger A, et al. Glottis-NetV2: temporal glottal midline detection using deep convolutional neural networks [J]. *IEEE J Transl Eng Health Med*, 2023, 11: 137-144. doi:10.1109/JTEHM.2023.3237859
- [22] Kist AM, Gómez P, Dubrovskiy D, et al. A deep learning enhanced novel software tool for laryngeal dynamics analysis [J]. *J Speech Lang Hear Res*, 2021, 64(6): 1889-1903. doi:10.1044/2021_JSLHR-20-00498
- [23] Kist AM, Breining K, Dörrich M, et al. A single latent channel is sufficient for biomedical glottis segmentation [J]. *Sci Rep*, 2022, 12(1): 14292. doi:10.1038/s41598-022-17764-1
- [24] Adamian N, Naunheim MR, Jowett N. An open-source computer vision tool for automated vocal fold tracking from videoendoscopy [J]. *Laryngoscope*, 2021, 131(1): E219-E225. doi:10.1002/lary.28669
- [25] Wang TV, Adamian N, Song PC, et al. Application of a computer vision tool for automated glottic tracking to vocal fold paralysis patients [J]. *Otolaryngol Head Neck Surg*, 2021, 165(4): 556-562. doi:10.1177/0194599821989608
- [26] Moccia S, Vanone GO, Momi E, et al. Learning-based classification of informative laryngoscopic frames [J]. *Comput Methods Programs Biomed*, 2018, 158: 21-30. doi:10.1016/j.cmpb.2018.01.030
- [27] Ni XG, Zhang QQ, Wang GQ. Narrow band imaging versus autofluorescence imaging for head and neck squamous cell carcinoma detection: a prospective study [J]. *J Laryngol Otol*, 2016, 130(11): 1001-1006. doi:10.1017/S0022215116009002
- [28] Kraft M, Fostiropoulos K, Gürtler N, et al. Value of narrow band imaging in the early diagnosis of laryngeal cancer [J]. *Head Neck*, 2016, 38(1): 15-20. doi:10.1002/hed.23838
- [29] Yang Y, Liu J, Song F, et al. The clinical diagnostic value of target biopsy using narrow-band imaging endoscopy and accurate laryngeal carcinoma pathologic specimen acquisition [J]. *Clin Otolaryngol*, 2017, 42(1): 38-45. doi:10.1111/coa.12654
- [30] He YR, Cheng YD, Huang ZG, et al. A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis [J]. *Ann Transl Med*, 2021, 9(24): 1797. doi:10.21037/atm-21-6458
- [31] Xu JW, Wang J, Bian XZ, et al. Deep learning for nasopharyngeal carcinoma identification using both white light and narrow-band imaging endoscopy [J]. *Laryngoscope*, 2022, 132(5): 999-1007. doi:10.1002/lary.29894
- [32] Weng JJ, Wei JZ, Wei YZ, et al. Diagnosis of nasopharyngeal carcinoma with convolutional neural network on narrowband imaging [J]. *Lin Chuang Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, 2023, 37(6): 483-486. doi:10.13201/j.issn.2096-7993.2023.06.015
- [33] Lin JY, Walsted ES, Backer V, et al. Quantification and analysis of laryngeal closure from endoscopic videos [J]. *IEEE Trans Biomed Eng*, 2019, 66(4): 1127-1136. doi:10.1109/TBME.2018.2867636
- [34] Patrini I, Ruperti M, Moccia S, et al. Transfer learning for informative-frame selection in laryngoscopic videos through learned features [J]. *Med Biol Eng Comput*, 2020, 58(6): 1225-1238. doi:10.1007/s11517-020-02127-7
- [35] Cho WK, Lee YJ, Joo HA, et al. Diagnostic accuracies of laryngeal diseases using a convolutional neural network-based image classification system [J]. *Laryngoscope*, 2021, 131(11): 2558-2566. doi:10.1002/lary.29595
- [36] Laves MH, Bicker J, Kahrs LA, et al. A dataset of

- laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation [J]. *Int J Comput Assist Radiol Surg*, 2019, 14(3): 483-492. doi:10.1007/s11548-018-01910-0
- [37] Parker F, Brodsky MB, Akst LM, et al. Machine learning in laryngoscopy analysis: a proof of concept observational study for the identification of post-extubation ulcerations and granulomas [J]. *Ann Otol Rhinol Laryngol*, 2021, 130(3): 286-291. doi:10.1177/0003489420950364
- [38] Kist AM, Dürr S, Schützenberger A, et al. OpenHSV: an open platform for laryngeal high-speed videoendoscopy [J]. *Sci Rep*, 2021, 11(1): 13760. doi:10.1038/s41598-021-93149-0
- [39] Matava C, Pankiv E, Raisbeck S, et al. A convolutional neural network for real time classification, identification, and labelling of vocal cord and tracheal using laryngoscopy and bronchoscopy video [J]. *J Med Syst*, 2020, 44(2): 44. doi:10.1007/s10916-019-1481-4
- [40] Girdler B, Moon H, Bae MR, et al. Feasibility of a deep learning-based algorithm for automated detection and classification of nasal polyps and inverted papillomas on nasal endoscopic images [J]. *Int Forum Allergy Rhinol*, 2021, 11(12): 1637-1646. doi:10.1002/alr.22854
- [41] Kim GH, Sung ES, Nam KW. Automated laryngeal mass detection algorithm for home-based self-screening test based on convolutional neural network [J]. *Biomed Eng Online*, 2021, 20(1): 51. doi:10.1186/s12938-021-00886-4
- [42] Kono M, Ishihara R, Kato Y, et al. Diagnosis of pharyngeal cancer on endoscopic video images by Mask region-based convolutional neural network [J]. *Dig Endosc*, 2021, 33(4): 569-576. doi:10.1111/den.13800
- [43] Ay B, Turker C, Emre E, et al. Automated classification of nasal polyps in endoscopy video-frames using handcrafted and CNN features [J]. *Comput Biol Med*, 2022, 147: 105725. doi:10.1016/j.combiomed.2022.105725
- [44] Azam MA, Sampieri C, Ioppi A, et al. Deep learning applied to white light and narrow band imaging video-laryngoscopy: toward real-time laryngeal cancer detection [J]. *Laryngoscope*, 2022, 132(9): 1798-1806. doi:10.1002/lary.29960
- [45] Wellenstein DJ, Woodburn J, Marres HAM, et al. Detection of laryngeal carcinoma during endoscopy using artificial intelligence [J]. *Head Neck*, 2023, 45(9): 2217-2226. doi:10.1002/hed.27441
- [46] Heo J, Lim JH, Lee HR, et al. Deep learning model for tongue cancer diagnosis using endoscopic images [J]. *Sci Rep*, 2022, 12(1): 6281. doi:10.1038/s41598-022-10287-9
- [47] Nakajo K, Ninomiya Y, Kondo H, et al. Anatomical classification of pharyngeal and laryngeal endoscopic images using artificial intelligence [J]. *Head Neck*, 2023, 45(6): 1549-1557. doi:10.1002/hed.27370
- [48] Zhao Q, He YQ, Wu YD, et al. Vocal cord lesions classification based on deep convolutional neural network and transfer learning [J]. *Med Phys*, 2022, 49(1): 432-442. doi:10.1002/mp.15371
- [49] Pedersen M, Larsen CF, Madsen B, et al. Localization and quantification of glottal gaps on deep learning segmentation of vocal folds [J]. *Sci Rep*, 2023, 13(1): 878. doi:10.1038/s41598-023-27980-y
- [50] Sakthivel S, Prabhu V. Optimal deep learning-based vocal fold disorder detection and classification model on high-speed video endoscopy [J]. *J Healthc Eng*, 2022; 4248938. doi:10.1155/2022/4248938
- [51] Yan PK, Li SH, Zhou Z, et al. Automated detection of glottic laryngeal carcinoma in laryngoscopic images from a multicentre database using a convolutional neural network [J]. *Clin Otolaryngol*, 2023, 48(3): 436-441. doi:10.1111/coa.14029

(编辑:李纬)