

# 嗓音分析与内镜技术结合人工智能在咽喉病变诊疗中的应用和发展

宋琦<sup>1</sup> 李晓明<sup>1</sup>

**[摘要]** 耳鼻喉科咽喉病变诊疗中,将嗓音分析或内镜技术与人工智能结合的应用和发展迅猛。本文通过文献复习,回顾了嗓音分析或内镜技术与人工智能结合的历史和原理,对其应用和发展现状进行了总结,概括其优势在于强大的学习和判读能力、惊人的速度和耐性以及稳定的复制和拓展性。目前制约其发展的关键是机器学习过程中的不确定性、小样本引起的误差和伦理上的哲学思考。未来的发展方向应该是耳鼻咽喉头颈外科医生在掌握过硬的专业知识基础上,学习流行病学、经典统计学领域相关知识,加强与机器学习开发人员之间的交流合作,最终使先进的科学技术能在临床真正地使用,最大程度造福广大患者。

**[关键词]** 咽喉疾病;嗓音分析;内镜技术;人工智能

**DOI:**10.13201/j.issn.2096-7993.2022.08.017

**[中图分类号]** R766.5 **[文献标志码]** A

## Application and development of voice analysis and endoscopic technology combined with artificial intelligence in the diagnosis and treatment of throat disease

SONG Qi LI Xiaoming

(Department of Otolaryngology Head and Neck Surgery, the 980th Hospital of the Joint Logistics Support Unit of the Chinese PLA, Shijiazhuang, 050082, China)

Corresponding author: SONG Qi, E-mail: 13315100789@126.com

**Summary** In the diagnosis and treatment of throat disease, the application and development of combining voice analysis or endoscopic technology with artificial intelligence has developed rapidly. This paper reviews the history and principles of the combination of voice analysis or endoscopic technology with artificial intelligence, summarizes its status of application and development, and sums up its advantages that lie in the strong learning and interpretation ability, amazing speed and tolerance, and stable replication and expansion. The key to restrict its development is the uncertainty in the process of machine learning, the error caused by small samples, and the ethical philosophical thinking. Future development direction should be that the surgeons in otolaryngology head and neck department on the basis of excellent professional knowledge, learn related knowledge of epidemiology, classic statistics, strengthen the exchanges and cooperation with machine learning developers. Eventually, advanced science and technology can be truly used in clinical practice to maximize the benefit of the majority of patients.

**Key words** throat disease; voice analysis; endoscopic technology; artificial intelligence

<sup>1</sup> 中国人民解放军联勤保障部队第九八〇医院耳鼻咽喉头颈外科(石家庄,050082)  
通信作者:宋琦,E-mail:13315100789@126.com

**引用本文:**宋琦,李晓明.嗓音分析与内镜技术结合人工智能在咽喉病变诊疗中的应用和发展[J].临床耳鼻咽喉头颈外科杂志,2022,36(8):647-650. DOI:10.13201/j.issn.2096-7993.2022.08.017.

- [35] Zhang S, Wang L, Gong S, et al. Auditory and speech performance after unilateral cochlear implantation for cochlear nerve canal stenosis[J]. Ear Nose Throat J, 2021;1455613211045563.
- [36] Tahir E, Çınar BÇ, Özkan HB, et al. Successful Use of a Cochlear Implant in a Patient with Bony Cochlear Nerve Canal Atresia[J]. J Int Adv Otol, 2020, 16(2): 271-273.
- [37] Zainol Abidin Z, Mohd Zaki F, Kew TY, et al. Cochlear nerve canal stenosis and associated semicircular canal abnormalities in paediatric sensorineural hearing

- loss: a single centre study[J]. J Laryngol Otol, 2020, 134(7):603-609.
- [38] Ocak E, Kocaöz D, Acar B, et al. Radiological Evaluation of Inner Ear with Computed Tomography in Patients with Unilateral Non-Pulsatile Tinnitus[J]. J Int Adv Otol, 2018, 14(2):273-277.
- [39] Rah YC, Yoon YS, Chang MY, et al. Facial nerve stimulation in the narrow bony cochlear nerve canal after cochlear implantation[J]. Laryngoscope, 2016, 126(6):1433-1439.

(修回日期:2021-11-29)

耳鼻咽喉头颈外科是医学中最古老的分支,现代机器学习技术应用其中独具潜力。1991年, Freeman<sup>[1]</sup>报道采用神经网络对听觉脑干反应波形进行自动识别;2000年, Schönweiler 等<sup>[2]</sup>采用人工神经网络对语音声学特征进行分类;2004年, Whipple 等<sup>[3]</sup>报道口腔鳞状细胞癌基因组预测结果;这些都是机器学习应用于耳鼻咽喉头颈外科临床的早期报道。此后,耳鼻喉科关于机器学习的应用文献急剧增加。尤其在咽喉病变诊疗中,将嗓音分析或内镜技术与人工智能结合,其应用和发展迅猛。因此,耳鼻咽喉头颈外科医生应该审慎地了解这些分析工具的能力和局限性。

## 1 历史和原理

咽喉病变的诊断方法主要有以下3种:①嗓音分析,利用数字信号处理技术分析语言声学参数,具有无创特点,可对嗓音质量进行客观评估;②电子喉镜,通过影像确定病变位置并初步判断病变性质;③喉肌电图,通过喉部肌肉运动情况诊断声带病变。这几种方法,过去主要依赖于医生的临床经验和主观判断。随着计算机技术的普及和提高,人工智能开始与这些方法相结合。机器学习算法擅长提取声音、图像和文本数据。无论是嗓音的声学分析,还是内镜下图像数据的处理,人工智能技术都表现出巨大的应用前景。

嗓音是通过肺部呼出气流引发声带规律性振动,产生特征性的声波,经口鼻共鸣后形成。当咽喉发生病变时,病态嗓音信号的声学参数会与正常状态下产生一定偏差,通过声音采集设备采集的声音,经计算机预处理和提取声学特征后,输入到分类器或识别机中进行评估。目前人工智能主要用于对病态嗓音二分类(区别正常或病理)和声带类病理嗓音细分类(具体病因)识别研究。常用的模式识别算法有高斯混合模型(gaussian mixture model)<sup>[4]</sup>、人工神经网络(artificial neural network)<sup>[5]</sup>、支持向量机(support vector machine)<sup>[6]</sup>、随机森林(random forest)<sup>[7]</sup>、深度神经网络(deep neural network)<sup>[8]</sup>、卷积神经网络(convolutional neural network)<sup>[9]</sup>等。

喉镜图像是利用光学信号对咽喉部组织成像,其颜色和纹理等信息丰富,可以进行深度挖掘。常用的咽喉内镜包括纤维喉镜、电子喉镜、频闪喉镜及 NBI 内镜。利用人工智能将机器学习与图像处理技术结合,对其图像中高维度量特征(主要是灰度和纹理特征)进行提取,输入构建好的模型进行分析判断,将临床辅助诊断问题转化为模式识别问题。其流程主要包括:图像获取→图像预处理→感兴趣区分割→特征提取与分类→模型构建。

## 2 应用和发展现状

### 2.1 嗓音分析与人工智能结合

2017年,庞宇峰等<sup>[10]</sup>采用荷兰阿姆斯特丹大

学语音科学研究所研制的 Praat 软件进行嗓音采集,应用 SPSS Modeler 软件,通过人工神经网络中多层感知器的方法建模,其神经网络病态嗓音识别率达到 75.7%。2019年, Fang 等<sup>[11]</sup>报道了神经网络、支持向量机与高斯混合模型在语音样本中诊断常见语音障碍方面的能力:作者使用中国台湾数据库训练算法,包括 60 个正常的语音样本和 402 个来自各种病理声带的标记语音样本,疾病包括:声带小结、肿瘤、声带萎缩、声带麻痹/瘫痪、声带肌张力障碍和声带沟等,其神经网络在验证马萨诸塞州眼耳医院的外部语音障碍数据库时,准确率达到 99.2%。

### 2.2 内镜技术与人工智能结合

2014年, Huang 等<sup>[12]</sup>开发的声带疾病智能识别系统,分析了声带正常、声带瘫痪、声带息肉、声带囊肿四种声带模式的特征值,使用支持向量机分类器对声带疾病进行识别,准确率达到 98.75%。2015年, Unger 等<sup>[13]</sup>基于声带波形中断预测声门恶性肿瘤,使用神经网络来识别频闪喉镜下正常受试者和 T1a 声门癌患者声带波形态的差异,从视频样本中识别有关振动测量参数,其识别声带恶性病变灵敏度和特异度均 100%。2018年, Mascharak 等<sup>[15]</sup>使用多光谱窄波段成像和白光内镜来量化评估口咽的淋巴上皮组织,结果显示,基于颜色上的差异,能够很好地区分口咽部鳞状细胞癌和健康黏膜(准确率 65.9%,而在白光下的准确率为 52.3%)。Song 等<sup>[16]</sup>开发了基于智能手机的口腔内双通道成像平台,利用卷积神经网络基于自身荧光和白光图像,对口腔黏膜潜在病灶和恶性病变进行分类,结果准确率近 86.9%,但训练样本相对较小(66 个正常样本和 64 个可疑样本)。Moccia 等<sup>[17]</sup>提出了一种基于强度、关键点和图像空间内容特征的 NBI 内镜帧分类方法,采用支持向量机与径向基函数和一对一方案区别模糊、唾液或镜面反射、欠曝光等情况,可用于喉镜下喉部病变的人工智能识别与诊断;其基于深度学习从 33 例喉鳞状细胞癌患者的内镜视频中提取了 1320 个肿块,构成 4 种喉组织类别的数据库,该模型在早期喉癌组织分类中准确率达 93%<sup>[18]</sup>。2019年, Xiong 等<sup>[19]</sup>开发了一种在喉镜图像中自动检测喉癌的深度学习卷积神经网络,在所有病变和正常组织中检测喉癌和癌前病变的敏感度为 73.1%,特异度为 92.2%,AUC 为 0.922,总体准确率为 86.7%,相当于拥有 10~20 余年工作经验的专家。2020年, Kuo 等<sup>[20]</sup>从喉内镜图像中自动筛选清晰图像并利用喉特征结构进行区域自动分割,量化评估咽喉反流引起的喉部变化,判断和分析其严重程度,采用 Fisher 线性判别法确定特征和分类性能,采用支持向量机作为分类器对咽喉反流进行判断,其准确率

为 97.16%, 敏感度为 98.11%, 假阳性率为 3.77%; 图像结合反流症状指数评分, 采用神经网络对严重程度进行分级, 准确率为 96.08%。Dunham 等<sup>[21]</sup>使用卷积神经网络自动分类喉镜检查结果, 良性疾病正确诊断率为 80.9%, 恶性疾病正确诊断率为 93%。Ren 等<sup>[22]</sup>开发一种基于深度学习卷积神经网络的计算机辅助诊断系统, 用于区分喉部肿瘤, 总体准确率达到 96.24%; 对于白斑、良性、恶性、正常和声带小结, 敏感度和特异度分别为 92.8% 和 98.9%、97.0% 和 99.7%、89.0% 和 99.3%、99.0% 和 99.4%、97.2% 和 99.1%。2021 年, Xu 等<sup>[14]</sup>基于白光成像 (WLI) 和窄带成像 (NBI) 鼻咽镜图像开发一种基于深度学习的自动诊断系统, 以识别鼻咽癌 (NPC) 和非癌症 (炎症和增生), 综合应用总体准确率达到 94.9%。Hu 等<sup>[9]</sup>使用卷积神经网络模型与专科医生的准确率相比, 在区分正常嗓音、声带萎缩、单侧声带麻痹、器质性声带病变和内收肌痉挛性发声困难方面的敏感度为 66%, 特异度为 91%, 总体准确率为 66.9%; 而 2 名喉科医生的总体准确率分别为 60.1% 和 56.1%, 2 名普通耳鼻喉科医生的总体准确率分别为 51.4% 和 43.2%。胡蓉等<sup>[23]</sup>使用基于深度卷积神经网络的人工智能在喉鳞状细胞癌窄带成像中进行辅助诊断, 人工智能验证集的准确率为 90.91%, 敏感度为 90.12%, 特异度为 91.53%, 与耳鼻咽喉头颈外科专家判读结果相当; 其判读速度明显高于耳鼻咽喉头颈外科专家。

Yao 等<sup>[24]</sup>对人工智能在喉镜检查中的应用进行综述, 发现目前研究主要集中在声带振动分析 (43%)、病变识别 (24%) 和声带运动检测 (19%); 最常见的自动化任务是识别声带小结 (19%)、息肉 (14%)、瘫痪 (11%)、麻痹 (8%) 和囊肿 (7%); 成像方式包括高速喉视频 (45%), 频闪镜 (29%) 和窄带成像内镜 (7%)。这些文献主要由科学、技术、工程和数学专家撰写 (占 76%), 其中只有 30 项研究 (31%) 涉及数学专家和耳鼻喉科专家合作; 数据库多来自单一机构 (84%), 其中德国 23%、美国 16%、西班牙 9%、意大利 8% 和中国 8%。

### 3 潜在的价值和意义

人工智能在医学上的应用主要分为两个部分: 一是图像识别; 二是深度学习。这两个部分均是基于医学信息 (声音、图像和文本等) 大数据所进行的数据挖掘和应用, 其中深度学习是核心环节。

传统人工分析的缺点包括但不限于: ①判断主观、经验性, 缺乏精确量化标准, 易误判; ②人为原因出现误差及疲劳; ③海量信息易漏诊; 此外, 医疗从业人员数量短缺和职业培训的时间成本, 严重制约专科医师的成长和进步; 而且, 目前医学信息数据年增长率约为 30%, 远远超出人员增长的势头。

人工智能与医学信息的融合, 恰好可在多学科发展拾遗补缺。其强大的学习和判读能力、惊人的速度和耐性以及稳定的复制和拓展性毋庸置疑。发展成熟后可以节约时间, 提高效率, 辅助专家诊断, 缩短医生培训时间; 网络化后还可以实现基层迅速同质推广。

### 4 存在的问题和面临的挑战

已报道的很多模型性能令人印象深刻, 其敏感度和特异度均超过 95%。但至今仍未看到其广泛的部署和实施。

部分原因在于机器学习技术本身。机器学习算法输入因素影响输出结果<sup>[25]</sup>。人工智能系统本身存在一定程度的不确定性: 我们无法确定机器学习算法使用了哪些规则并怎样生成输出结果。因此, 需要正确把握机器学习算法中使用的有影响力的变量或特征。这就需要临床医生参与, 既在经典专科方面接受过培训, 也拥有广泛的数学知识并延伸到流行病学和经典统计学领域。此外, 临床医生与机器学习开发人员之间的交流也非常重要。增加合作、交流和临床医生参与能增加这些技术在医学上的相关性、使用和受益。

另一个原因是“输入垃圾, 输出垃圾”的问题。机器学习算法依靠尽可能真实准确且代表性好的输入数据进行推断。目前数据多为单中心小样本量, 存在不同程度的误差, 这些误差可能在机器学习的过程中被放大<sup>[26]</sup>。

此外, 还有人担忧机器是否会接管人类的判断和决策<sup>[27]</sup>。应该说, 在现有的技术下, 仍然需要人工输入和算法审计来监督机器学习。临床医生需要与数据科学家合作, 优化医疗数据库的效用<sup>[28]</sup>。毕竟, 医学是科学和人文的交叉学科; 在诊疗过程中, 医生不断对患者的体征、现象、发生的事件进行思考和总结。从哲学角度出发, 医学以人为本, 强调在医疗过程中对人的关心、关怀和尊重。这就决定了人工智能在未来很长一段时间内, 无法像医生一样运用自身的专业知识和经验, 去解决纷繁复杂的患者状况。

### 5 展望

大数据通常定义 3 个 V: 即数据量、数据流速度和数据类型<sup>[29]</sup>。将机器学习技术应用于医疗的优势在于算法能够处理由无数异质性变量组成的大型复杂数据库。在信息时代, 耳鼻咽喉头颈外科医生除掌握过硬的专业知识外, 还需要学习流行病学、经典统计学领域的相关知识, 加强与机器学习开发人员之间的交流合作。最终目标是使先进的科学技术能在临床真正地使用, 从而最大程度地造福广大患者。

利益冲突 所有作者均声明不存在利益冲突

### 参考文献

[1] Freeman DT. Computer recognition of brain stem au-

- ditory evoked potential wave V by a neural network [J]. *Proc Annu Symp Comput Appl Med Care*, 1991; 305-309.
- [2] Schönweiler R, Hess M, Wübbelt P, et al. Novel approach to acoustical voice analysis using artificial neural networks [J]. *J Assoc Res Otolaryngol*, 2000, 1(4):270-282.
- [3] Whipple ME, Mendez E, Farwell DG, et al. A genomic predictor of oral squamous cell carcinoma [J]. *Laryngoscope*, 2004, 114(8):1346-1354.
- [4] Godino-Llorente JI, Gómez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters [J]. *IEEE Trans Biomed Eng*, 2006, 53(10):1943-1953.
- [5] Ritchings RT, McGillion M, Moore CJ. Pathological voice quality assessment using artificial neural networks [J]. *Med Eng Phys*, 2002, 24(7/8):561-564.
- [6] Saeedi NE, Almasganj F, Torabinejad F. Support vector wavelet adaptation for pathological voice assessment [J]. *Comput Biol Med*, 2011, 41(9):822-828.
- [7] Hemmerling D, Skalski A, Gajda J. Voice data mining for laryngeal pathology assessment [J]. *Comput Biol Med*, 2016, 69:270-276.
- [8] Chen L, Chen J. Deep Neural Network for Automatic Classification of Pathological Voice Signals [J]. *J Voice*, 2022, 36(2):288.
- [9] Hu HC, Chang SY, Wang CH, et al. Deep Learning Application for Vocal Fold Disease Prediction Through Voice Recognition; Preliminary Development Study [J]. *J Med Internet Res*, 2021, 23(6):e25247.
- [10] 庞宇峰, 黄娟, 徐蓓峥, 等. 病态嗓音的定量分析及人工神经网络识别 [J]. *临床耳鼻咽喉头颈外科杂志*, 2017, 31(2):100-102.
- [11] Fang SH, Tsao Y, Hsiao MJ, et al. Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach [J]. *J Voice*, 2019, 33(5):634-641.
- [12] Huang CC, Leu YS, Kuo CF, et al. Automatic recognizing of vocal fold disorders from glottis images [J]. *Proc Inst Mech Eng H*, 2014, 228(9):952-961.
- [13] Unger J, Lohscheller J, Reiter M, et al. A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis [J]. *Cancer Res*, 2015, 75(1):31-39.
- [14] Xu J, Wang J, Bian X, et al. Deep Learning for nasopharyngeal Carcinoma Identification Using Both White Light and Narrow-Band Imaging Endoscopy [J]. *Laryngoscope*, 2021.
- [15] Mascharak S, Baird BJ, Holsinger FC. Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning [J]. *Laryngoscope*, 2018, 128(11):2514-2520.
- [16] Song B, Sunny S, Uthoff RD, et al. Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning [J]. *Biomed Opt Express*, 2018, 9(11):5318-5329.
- [17] Moccia S, Vanone GO, Momi E, et al. Learning-based classification of informative laryngoscopic frames [J]. *Comput Methods Programs Biomed*, 2018, 158:21-30.
- [18] Moccia S, De Momi E, Guarnaschelli M, et al. Confident texture-based laryngeal tissue classification for early stage diagnosis support [J]. *J Med Imaging (Bellingham)*, 2017, 4(3):034502.
- [19] Xiong H, Lin P, Yu JG, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images [J]. *EBioMedicine*, 2019, 48:92-99.
- [20] Kuo CJ, Kao CH, Dlamini S, et al. Laryngopharyngeal reflux image quantization and analysis of its severity [J]. *Sci Rep*, 2020, 10(1):10975.
- [21] Dunham ME, Kong KA, McWhorter AJ, et al. Optical Biopsy: Automated Classification of Airway Endoscopic Findings Using a Convolutional Neural Network [J]. *Laryngoscope*, 2020.
- [22] Ren J, Jing X, Wang J, et al. Automatic Recognition of Laryngoscopic Images Using a Deep-Learning Technique [J]. *Laryngoscope*, 2020, 130(11):E686-E693.
- [23] 胡蓉, 钟琦, 徐文, 等. 基于深度卷积神经网络的人工智能在喉鳞状细胞癌窄带成像辅助诊断中的应用 [J]. *中华耳鼻咽喉头颈外科杂志*, 2021, 56(5):454-458.
- [24] Yao P, Usman M, Chen YH, et al. Applications of Artificial Intelligence to Office Laryngoscopy: A Scoping Review [J]. *Laryngoscope*, 2021.
- [25] Chowdhury NI, Smith TL, Chandra RK, et al. Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks [J]. *Int Forum Allergy Rhinol*, 2019, 9(1):46-52.
- [26] Mullainathan S, Obermeyer Z. Does Machine Learning Automate Moral Hazard and Error? [J]. *Am Econ Rev*, 2017, 107(5):476-480.
- [27] Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine [J]. *JAMA*, 2017, 318(6):517-518.
- [28] Sataloff RT. Data Scientists: They know what we don't know that we don't know about Big Data [J]. *Ear Nose Throat J*, 2016, 95(8):302-305.
- [29] Risoud M, Bonne NX, Vincent C. Big Data: Coming soon to ENT [J]. *Eur Ann Otorhinolaryngol Head Neck Dis*, 2016, 133(3):157.

(收稿日期:2021-12-14)