

· 共识与指南 ·

消化内镜人工智能数据采集与标注质量控制
体系专家共识意见(草案 2019, 上海)

中国医师协会消化内镜人工智能专业委员会 上海市计算技术研究所 上海市医疗器械
检测所

通信作者:李兆申, Email: zhsl@vip.163.com; 于红刚, Email: yuhonggang1968@163.com

【摘要】 随着以深度学习为代表的人工智能技术在消化内镜领域的探索和运用,作为人工智能学习材料的内镜影像,其数据质量逐渐受到关注。本共识聚焦于人工智能相关内镜数据的采集与标注,着力阐明数据采集、标注、存储、隐私保护、数据安全以及使用标准等关键问题,以期更好地服务于内镜人工智能模型的训练、调优与测评,进一步推动我国消化内镜人工智能平台的高质量发展。

【关键词】 人工智能; 数据收集; 数据分析; 质量控制

基金项目:国家重点研发计划(2018YFC1313103); 国家自然科学基金(81670473, 81873546);
上海市“曙光计划”(19SG30)

DOI: 10. 3760/cma.j.cn321463-20200509-00403

恶性肿瘤是我国居民健康的重要杀手之一,而食管胃肠癌占据我国所有恶性肿瘤的 40%^[1-2]。目前,胃癌、结直肠癌和食管癌均位列我国癌症发病率前 5 位,年发病和死亡数达全球总数的 50%,直接的年均医疗负担近千亿元^[1,3]。如不采取有效遏制措施,我国食管胃肠癌发病和死亡数将在 2020 年分别上升至 130 万和 100 万。更为严峻的,目前我国超过 85% 的食管胃肠癌发现时即为中晚期,即使经手术等综合治疗后五年生存率也不足 50%。因此,积极推进食管胃肠癌的早诊早治不仅是生命科学与医学研究的热点,也是关乎国计民生的重大社会问题^[1,3-4]。

内镜检查是食管胃肠癌早期诊断和筛查的核心技术手段和金标准,实施食管胃肠癌内镜筛查可发现和处理癌前病变,大幅降低消化道恶性肿瘤的发病率与死亡率^[4-6]。然而,内镜并不能完全检出消化道内所有病变。由于内镜医师操作规范性、病变识别能力和视觉疲劳等因素的影响,临床实践中有相当比例的病变,甚至包括进展期病变和癌前病变都可能被漏诊,直接影响到内镜检查质量与筛查效果^[7-13]。近年来,随着以深度学习为代表的新一代人工智能技术的突破,图像识别技术在医学影像与内镜等领域取得了革命性的进展,计算机辅助诊断(computer-aided diagnosis, CAD)系统有望充当内

镜医师的“第三只眼”,起到实时辅助与监督医师的作用^[14-17]。CAD 一经引入,便成为消化内镜学领域的研究热点,从单中心回顾性的模型建立与测试,到前瞻性随机对照试验(randomized controlled trial, RCT),均有大量的研究注册与发表^[18-21]。我国学者在此领域处于国际领先水平,率先报道了多项 RCT,证明 CAD 可在临床实践中协助内镜医师发现更多结肠腺瘤,并监督胃镜检查质量,减少内镜医师的检查盲区 and 病变漏诊^[19,22]。

除算法技术外, CAD 诊断性能还高度依赖于海量高质量标注数据集,以便 CAD 准确地掌握各类病变的图像特征^[14]。然而,海量高质量数据集的建立面临着一系列难题。由于胃肠道中不同部位病灶存在明显的个体差异,部分病灶与周围黏膜难以准确区分,而且内镜下观察极易受消化道内容物的影响,所以消化道病灶的高质量标注较传统影像数据更为困难,构建高质量标注数据集也成为 CAD 平台建立与优化的关键基础和首要前提^[23]。目前,高质量标注数据集的构建尚存在一些瓶颈:由于内镜数据库资料的孤岛效应和稀缺性,绝大多数 CAD 平台仅基于单中心、小样本的数据集建立,数据来源异质性较大,数据质量和标注缺少相应规范和评价,难以保证训练数据集对疾病特征的代表性;内镜数据采集质量容易受到医师操作习惯、重视程度、诊

疗规范性、存储与管理机制等多种因素的影响,产生大量非标准化数据;目前实施的研究数据标注行为多为研究者发起,其标注方法和结果的准确性与科学性缺乏必要的监督和保证,可能致使 CAD 平台存在较大的系统偏倚风险^[19,21,24-26];患者隐私及数据安全保护缺乏行业规范的监管,CAD 平台的建立可能面临道德甚至是法律风险;领域内尚未建立第三方高质量数据集,对当前种类繁多的 CAD 诊断性能不易做出统一的优效性评价^[27]。

随着 CAD 研究在内镜领域的推进,内镜数据质量与模型的规范性逐渐受到关注^[28]。国家药品监督管理局医疗器械技术审评中心最新出台的《深度学习辅助决策医疗器械软件审评要点》中明确要求重点关注数据质量控制^[29],但我国消化内镜领域 CAD 的开发缺乏相关指南或共识意见参考,亟须消化内镜与人工智能的专家共同协作,推动相关指南或共识意见的出台,以指导行业的健康、快速发展。

鉴于此,本专家共识意见聚焦于消化内镜数据采集与标注质量控制,针对消化内镜数据的采集与标注提出指导性意见,在保障患者隐私与数据安全的基础上,促进内镜领域人工智能行业的高质量发展,努力建立全行业统一的内镜数据采集、标注、存储、隐私保护、数据安全以及使用标准;同时为多中心、大样本、高质量、具有代表性的标准化第三方数据集的建立提供指导意见,更好地服务于内镜人工智能产品的训练、调优与测评,从而促进我国内镜人工智能平台的创新与发展,抢占消化内镜人工智能研发的制高点。本专家共识意见基于国内外消化内镜领域数据的采集与标注实践,经国内消化内镜与人工智能的资深专家反复讨论与修订形成,并将随着临床研究和实践的发展而进一步完善更新。

一、定义和限定范围

根据我国当前消化内镜设备与实践状况,结合目前内镜人工智能领域发展现状,建设消化内镜数据集的目的是为消化道病灶检出定位和性质鉴别的 CAD 提供高质量数据。

消化内镜数据采集范围包括前瞻性采集体检(筛查)、诊疗和随访的内镜影像及相关信息;消化内镜设备品牌、型号与类型;消化内镜影像采集过程合法合规性;消化内镜影像采集方式;消化内镜影像对比度、分辨率最低标准;消化内镜影像采集、传输、存储、共享、使用与销毁的数据安全及隐私保护机制。

二、数据采集与数据质量

数据采集应当详细记录采集的过程,包括数据的来源医疗机构、设备属性、采集时间、采集数量、操作医师、患者基线信息和疾病特征、采集形式、影像属性、完整性,以及相关的人口学基线等资料应尽可能详细,并依据国家和行业标准存储数据。

(一)数据采集

1.数据来源:数据来源应明确所属医疗机构、数据来源方式,例如直接导出、医学图像存储与传输系统(picture archiving and communication systems, PACS)导出等。

2.采集设备:为保证 CAD 能够广泛适应我国多品牌内镜共用的实际情况,采集设备应当包括(但不限于)国内市场占有率超过 5% 的消化内镜品牌,并应详细记录采集设备的具体型号以及影像留取时采集设备的相关配置参数(主机与内镜类型、冻结图像与否、特殊电子染色成像模式、有无放大、分辨率等)。此外,采集设备须确保处于设备的有效使用期内并符合质量控制要求。

3.采集过程:为避免数据选择性录入造成的偏倚,数据采集前需根据任务目标确定具体的筛选与排除标准,连续纳入病例,避免主观选择偏倚。实施消化内镜影像采集的医师须是熟练内镜操作,并通过考核的消化科、普外科或内镜科医师,具备 500 例以上的内镜操作经验;进行正式的数据采集前每名医师需经过副主任级别医师(操作经验 > 3 000 例)的培训,且学习和预录入例数分别不低于 20 例,录入结果经 3 名副主任医师评定合格后方可独立采集数据,并遵循数据采集和标注规范(下文详述)。

采集过程中,采集人员应首先详细了解采集任务的具体要求与注意事项,按照要求清理病灶表面可能存在的污渍,调整镜头及光线的位置以减少病变表面的反光点和留取不同距离与方位的病灶图像,遵循任务要求冻结图像,必要时实施染色内镜、电子染色及放大等成像模式以采集满足任务要求的病灶图像。病灶图像应当使用 JPEG 等标准格式的原始数据,一个图像有效区域最小像素矩阵不低于 230 像素×230 像素,图像矩阵不低于 512 像素×512 像素,有条件的医院可以使用 1 024 像素×1 024 像素。如基础图像不够清晰,为进一步增强系统的可推广性,可适当降低采集图像的矩阵标准,但应在采集前作出详细说明,预先制定计划。采集的病灶图像建模前不得进行任何修改、编辑,不得进行

有损压缩,且保持连续完整。

4.影像导出:病灶图像留取好后,根据实际情况,可允许下列方式导出图像:①通过内镜中心数据库后台,即从系统独立的内镜中心或 PACS 导出目标影像;②直接从消化内镜主机端口导出目标影像以保证数据质量;③通过影像记录设备从主机或显示屏端口导出目标影像。

在条件允许的情况下,建议优先采用直接和批量导出的形式,并进行内镜与病理资料的聚合,以减少人工采集的误差。值得说明的是,回顾性消化内镜数据无法适用此规范,必要时可按照任务目标筛选历史影像,但应注意回顾性数据对 CAD 平台造成的偏倚风险,具体数据要求可参考前述标准或另作说明。

5.数据交换:导出的消化内镜影像可通过手动方式上传或与医院业务系统直接对接的方式自动上传至第三方数据平台。其中与医院业务系统对接的方式须通过符合国家医疗行业标准的数据接口,并建立基于广域网的安全信息交换、安全传输机制,方可对内镜影像进行脱敏后交换和传输。

(二)数据处理

1.数据预处理:内镜影像数据的处理,包括滤波、增强、重采样、尺寸裁剪、均一化等,要求数据处理全过程应可恢复和回溯,同时明确数据处理工具的名称、版本、供应商、运行环境等信息,并充分考虑数据预处理方法对 CAD 的影响。

2.数据清洗:数据采集阶段所获取的数据,需要进行初步的筛选与清洗,剔除完全无价值数据,确保基础数据库的数据质量。常见的无价值数据情况包括:数据缺失、图像伪影、图像质量不合格、图像参数不符、非目标部位图像、图像中掺杂异物等。数据筛选和清洗的过程可通过专用工具初筛和清洗后,交由具有标注资质的医师进行二次复核,如需要剔除数据,则需详细记录剔除原因以及选用工具的名称及版本信息。

3.基础数据库:数据经预处理后所保存的基础数据库,应具备统计功能,统计项包括疾病构成、患者特征、地域分布、数据来源、采集设备、数据类型等,以明确数据分布,提高数据采集效率,确保数据分布的均衡性与代表性,并降低样本偏性对 CAD 的影响。

(三)数据质量

1.合规性:伦理批准与隐私保护。伦理批准,即

运用于 CAD 训练、优化和测试的内镜影像数据须获得医院伦理委员会批准或者豁免的临床脱敏数据。隐私保护,指应当满足相关法规要求,从志愿者、筛查项目、门诊患者、科研项目等数据库来源的数据同样应经过伦理委员会审查和批准;同时,数据采集、传输和使用的过程中须严格保护患者隐私,数据中如果含有患者隐私信息或锚定信息编码,须提前进行数据脱敏。

用于训练和测试系统的数据集存储平台应该建立电子化的数据管理系统,支持可疑数据删除或过滤功能;支持敏感数据脱敏功能,且敏感数据在使用前必须经过严密地脱敏流程,即不可逆清除患者隐私信息;同时建立严格的数据安全存储和监管机制,定期检查和报告数据的安全状态,并允许数据相关方具有了解数据安全情况的接口^[30]。

2.多样性与代表性:数据在采集和筛选过程中应避免数据采集偏倚,数据需具备广泛代表性,应优先选择多中心多层次内镜中心来源数据,地域分布尽可能广,医疗机构种类与层次尽可能丰富,患者分布(性别、年龄、民族、地域)尽可能均衡,保证每个病种或结构对于目标任务具有足够训练数据量,以保证 CAD 的泛化能力,并定期更新数据库;具体而言,每个目标病种的数据量保持均衡,每一类型样本数量应基本相同。即目标需求中的每一个细分子类的的数据,推荐数量以 1:1 的比例进行匹配,可在数据采集阶段注意控制数据类型的数量分布(有选择地多采集数量偏少类型数据等),以合理安排标注资源;由于国内内镜中心的数据存储方式方法有较大区别,推荐采集内镜影像的同时也采集患者的基本信息(性别、年龄、籍贯、民族、是否有相关疾病家族史等),可根据实际需要扩展至目前主流医院信息管理系统(hospital information system, HIS)能够提供的范围),并做脱敏处理,在保护患者隐私与数据安全的基础上,拓展数据分布的均衡性与代表性;数据的采集与标注过程中应注意避免数据重叠,以致数据结构偏倚和 CAD 过拟合。

3.数据质量控制:①数据质量要求。为减少观察者间差异,所有关于确认病灶存在与性质的结论应建立于病理诊断基础之上,或至少 3 名经验丰富(操作经验>3 000 例)的内镜医师在盲法情况下共同确认病变存在(仅限于检出和定位病灶任务),且须具有高置信度。②数据类型。尽管未来 CAD 有望具备在不同内镜模式下的广泛适用性,

但开发阶段每一种数据类型的最佳诊断效能都可能需特定方式和 CAD 参数,因此需对数据类型进行细致地分类,明确检查的部位与图像模式,以便更好地训练 CAD 及针对性调优。③数据公平性。数据的预处理过程应该注意公正地处理每份数据,数据预处理过程中不应随意删减或剔除符合质量要求的数据。如发生计划外的数据剔除,应详细说明原因。

三、数据标注与数据集构建

推荐建立针对目标任务的专用标注系统,招募符合资质的医师进行独立和重复标注,并配套数据标注标准与规范,同时落实质量监控机制,确保实现高质量的内镜数据标注。

(一)标注工作分类

研究者可根据 CAD 任务的不同选择合适的标注任务,综合考量研发成本与需求。①图片直接标注:直接赋予整张图片疾病或特征类别的标签。②病灶位置标注:从图像中定位出病灶,并圈出病灶的大致范围。③病灶性质标注:根据图像找到病灶的图像特征,将病灶定性的归为某一特征类别,或根据活检或切除后病理结果,确定病灶性质。④病灶精细轮廓标注:对在图像中找到的病灶,进行精细的病灶边缘勾画。

(二)标注资源管理

1.标注医师:标注人员应该具有超过 1 000 例且 3 年以上内镜操作经验。在保证符合资质的基础上,尽可能地纳入更广范围和更多层次的内镜医师,经验丰富者优先。

入选的标注医师应通过统一培训,掌握标注规则、标注标准以及标注工具的使用。培训结束后采用考核方式检验标注医师对标注规则和标注流程的掌握程度。考核分为理论与实际操作两个方面,以检出和分类两个任务作为评价层次,由专家组成员的评判标注结果,检查标注结果与金标准的符合度(建议>95%),测试者标注范围与金标准范围重合度(建议>80%)与达标率。必要时可将考核内容随机反复加入测试中,以确定同一病例由不同标注者或同一标注者进行重复测试的一致性。

2.标注工具:所有标注行为应基于一致且规范的标注工具,同时标注工具应当保证运行环境稳定,定期检查运行情况,并形成工作日志。

3.标注资源安全:标注数据及原始数据的上

传、交换、存储和下载须具备专门的数据安全保护措施并满足相关规范要求。标注数据及原始数据在脱敏前必须存储在具有高安全等级(不低于三级等保)的信息系统中,以确保数据不被丢失与泄露、甚至篡改;在脱敏后可以根据实际情况考虑存储在具有一般安全等级的信息系统中,数据脱敏需遵循相关标准或规范,建议在具有医疗大数据管理资质的大数据平台上管理标注数据以及原始数据。

(三)标注过程质控

1.流程设计:为保证数据标注质量,降低标注的个体差异性,建议标注应由至少 3 名医师独立完成,并通过交叉验证确认病灶的检出、分类、定位的准确性。考虑各个环节的工作量及资质要求,标注团队应至少包括标注医师、标注组长、仲裁专家三类医师,建议标注医师由至少 3 年以上内镜操作经验且操作 1 000 例以上的医师担任,标注组长由至少 5 年以上内镜操作经验且操作 3 000 例以上的医师担任,仲裁专家至少 8 年以上内镜操作经验且操作 3 000 例以上的医师担任。各批次标注任务由多位标注医师独立完成,并经标注组长、仲裁专家审核方可用于 CAD 训练。结合病灶检出分类定位流程(图 1),具体标注过程可分为以下 4 个环节(图 2)。

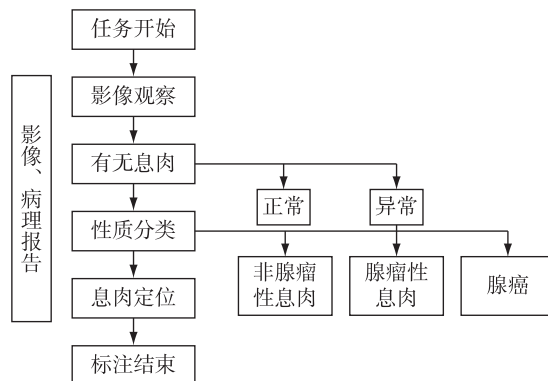


图 1 消化道病灶的检出、分类与定位流程(以结肠息肉为例)

(1)检出环节:由至少 3 名标注医师独立检出,确认内镜影像是否包含病灶,同时计算机自动统计各医师间检出的一致性,完全一致者可直接作为病灶检出结果;若有结果不一致,或有标注医师标注未能检出病变或难以确认病例(置信度较低),需经过组长与仲裁专家审核,仍不一致或不能确认者交由 3 名以上仲裁专家集体决议,以确定影像是否纳入及其可能分类。

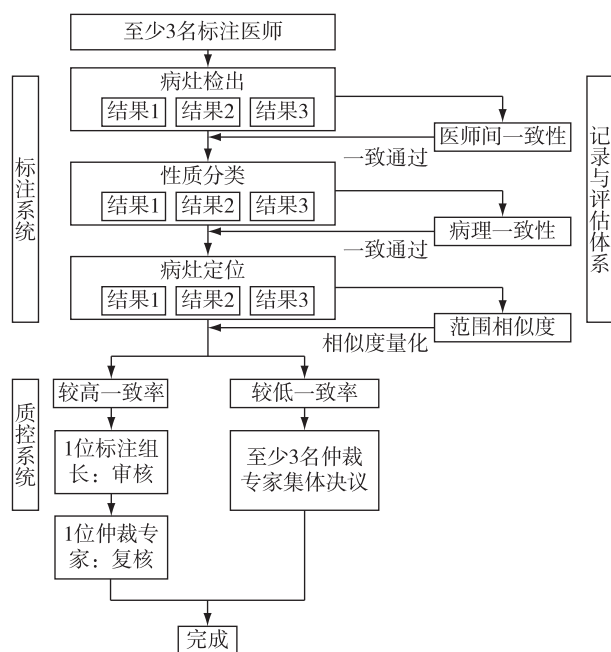


图2 消化内镜数据标注流程

(2)分类环节:由至少3名标注医师依据病理结果独立分类,确认病理性质,同时计算机自动根据病理报告确认性质分类是否正确,如无误即可作为病灶分类结果;如标注医师之间有不一致,或有标注医师标注难以分类病变(置信度较低),需经过组长与仲裁专家审核,仍不一致或不能确认者交由至少3名仲裁专家集体决议,以确定影像是否纳入及其可能分类。

(3)定位环节:由至少3名标注医师独立定位,描绘出边界清晰的病灶范围,同时计算机自动计算各医师间病灶范围相似度,如交并比(intersection over union, IOU),根据相似度水平决定审核流程,并修订病变的最终边界。

(4)审核环节:已完成标注的数据,根据定位环节的病灶范围相似度区分较高、较低一致率数据,并采取相应的审核策略。较高一致率数据由标注组长确认与修订病变边界,需至少1名仲裁专家二次复核;较低一致率或有标注医师标注难以定位病变(置信度较低)的数据,需至少3名仲裁专家集体决议病变边界甚至舍弃病例。对于检出、分类和定位环节出现的难以确认的病例,应当建立数据库亚组,以备后期进一步继续研究和处理。在条件允许的情况下,建议成立专门的专家团队,负责难以确认病例的诊断。

2.可追溯性:为保证标注结果的可追溯性,建议

在统一、规范的标注系统完成检出、分类、定位和审核环节,实现任务具象化、过程数字化,经手实名化,方便标注过程、审核环节的记录与复现。

(四)标注质量评估

标注质量高的数据应在标注正确率(类别数据)、标注精确度(坐标、时间点、个数、文字等)、标注完备性(是否漏标、重复标记)、标注一致性(不同标注人员之间、同一标注人员)方面均达到较高的标准。为保证标注数据的准确性与稳定性,正式标注过程开展之前,可进行预标注,主要目的如下:①检验标注培训效果和标注医师水平;②检验标注方案和标注策略是否可行;③为确立最终的标注策略提供数据分布及标注质量评定的参考;④预估标注耗时和成本。

(五)标注状态评估

1.标注时间监测:在培训阶段,可以通过记录测试过程中标注者对于每一个病例的标注时间,建立每一位标注者的标注时间分布。在实际标注阶段,可以通过不断比对每一个病例的标注时间在该标注人员标注时间分布中的位置来判断标注的状态。在通过设置暂停键避免不可抗力因素打断的情况下,如标注人员多次出现过长或过短的标注时间,则需要提醒和重点评估数据质量。

2.标注一致性监测:在标注过程中,随机引入部分病例,让标注人员进行二次标注,并计算两次标注之间的一致性,以确保标注人员标注一致性的水准始终能够达到项目要求,保证整体标注一致性。

3.标注准确率监测:在标注过程中,引入部分医学专家提供的带有标准标注答案的病例,计算标注人员对于带标准答案病例的准确性,用以评估标注人员的实时准确性。同时,定期更新测试数据集,保持监测系统的有效性。

(六)数据集构建

数据集可分为训练集、调优集、测试集。训练集用于算法训练、调优集用于算法调优、测试集用于算法评估。数据构建方应当明确训练集、调优集、测试集的划分方法、划分依据、分配比例,同时确保数据的分布符合临床实际情况以及训练集、调优集、测试集的样本两两无交集(同一病变的影像不被分配到不同的集合中)。

四、第三方数据库建设要求

参考国家药品监督管理局医疗器械技术审评

中心《深度学习辅助决策医疗器械软件审评要点》^[29],第三方数据库将涵盖完整数据集(包含训练集、调优集、测试集),其中测试集限于用于算法评估,不对外开放;其余子集在协商后允许有需求的企事业单位用于消化内镜人工智能算法的训练和调优。第三方数据库的建设除满足数据平台建设的通用要求(如网络与数据安全等,不再赘述)之外,还需满足相应专用要求。

(一)专业性

第三方数据库的数据标注人员、标注审核人员以及标注分歧仲裁人员应具备丰富的消化内镜实践经验。第三方数据库的创建单位应为相应临床专业领域的权威医疗机构或医学研究机构,如国家或省级相关临床医学研究中心等,省级中心应在国家中心的统一指导和认证下开展数据库建设。

(二)公共性

第三方数据库可在上述专业医疗机构的指导下,由具备医学大数据专业能力的公共服务平台建设运营。第三方大数据平台将发挥基础性作用,为数据采集、清洗、预处理、脱敏、标注、共享以及应用的全过程提供支撑服务。

(三)科学性

应当通过统计学方法,统计数据库分类样本数量,从而控制抽样误差。样本的分布情况应当符合机构目标疾病的流行病学特征。

(四)规范性

数据处理以及测评过程应当建立质量控制机制并严格遵守,数据具有可追溯性。

(五)多样性

为保证数据泛化能力,数据应来源于多中心、多类设备、不同人口属性。数据中应包含适当比例的对抗数据样本,用于评价算法的健壮性。

(六)封闭性

为客观、科学评价算法质量,应当封闭管理评测数据,测试的样本量需远大于单次测试所用数据量。

(七)动态性

为保证数据具有持续的多样性和封闭性,评测数据库按要求定期更换一定比例的数据。被更换的数据可用于构建公开数据集。

参与本共识意见制定的专家(按姓氏拼音排序):柏愚(海军军医大学长海医院)、郭强(云南省第一人民医院)、贺奇(上海市计算技术研究所)、侯晓华(华中科技大学同济医

学院附属协和医院)、胡兵(四川大学医学院华西医院)、李延青(山东大学附属齐鲁医院)、李兆申(海军军医大学长海医院)、廖专(海军军医大学长海医院)、令狐恩强(中国人民解放军总医院)、刘冰熔(郑州大学附属第一附院)、刘海涛(上海市计算技术研究所)、刘晓岗(四川省人民医院)、刘重生(上海市医疗器械检测所)、吕宾(浙江省中医院)、缪应雷(昆明医科大学附属第一医院)、任建林(厦门大学附属中山医院)、任旭(黑龙江省医院)、盛剑秋(中国人民解放军总医院第七医学中心)、王邦茂(天津医科大学总医院)、王东(海军军医大学长海医院)、王贵齐(中国医学科学院肿瘤医院)、王珏明(上海市计算技术研究所)、王良静(浙江大学医学院附属第二医院)、王洛伟(海军军医大学长海医院)、于红刚(武汉大学人民医院)、赵秋(武汉大学中南医院)、郑鹏远(郑州大学第五附属医院)、卓越(上海市医疗器械检测所)

执笔者:柏愚、赵胜兵、王树玲(海军军医大学长海医院消化内科)

利益冲突 所有作者声明不存在利益冲突

参 考 文 献

- [1] 郑荣寿,孙可欣,张思维,等. 2015 年中国恶性肿瘤流行情况分析[J]. 中华肿瘤杂志, 2019, 41(1): 19-28. DOI: 10.3760/cma.j.issn. 0253-3766. 2019. 01. 008.
- [2] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015[J]. CA Cancer J Clin, 2016, 66(2): 115-132. DOI: 10.3322/caac. 21338.
- [3] Zheng R, Zeng H, Zhang S, et al. National estimates of cancer prevalence in China, 2011[J]. Cancer Lett, 2016, 370(1): 33-38. DOI: 10.1016/j.canlet. 2015. 10. 003.
- [4] Jun JK, Choi KS, Lee HY, et al. Effectiveness of the Korean National Cancer Screening Program in Reducing Gastric Cancer Mortality[J]. Gastroenterology, 2017, 152(6): 1319-1328. e7. DOI: 10.1053/j.gastro. 2017. 01. 029.
- [5] Zhang X, Li M, Chen S, et al. Endoscopic Screening in Asian Countries Is Associated With Reduced Gastric Cancer Mortality: A Meta-analysis and Systematic Review[J]. Gastroenterology, 2018, 155(2): 347-354. e9. DOI: 10.1053/j.gastro. 2018. 04.026.
- [6] Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths[J]. N Engl J Med, 2012, 366(8): 687-696. DOI: 10.1056/NEJMoa1100370.
- [7] Wu J, Zhao SB, Wang SL, et al. Comparison of efficacy of colonoscopy between the morning and afternoon: A systematic review and meta-analysis[J]. Dig Liver Dis, 2018, 50(7): 661-667. DOI: 10.1016/j.dld. 2018.03. 035.
- [8] Aslanian HR, Shieh FK, Chan FW, et al. Nurse observation during colonoscopy increases polyp detection: a randomized prospective study[J]. Am J Gastroenterol, 2013, 108(2): 166-172.

- DOI: 10.1038/ajg.2012.237.
- [9] Lee CK, Park DI, Lee SH, et al. Participation by experienced endoscopy nurses increases the detection rate of colon polyps during a screening colonoscopy: a multicenter, prospective, randomized study [J]. *Gastrointest Endosc*, 2011, 74 (5): 1094-1102. DOI: 10.1016/j.gie.2011.06.033.
- [10] Zhao S, Wang S, Pan P, et al. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis [J]. *Gastroenterology*, 2019, 156 (6): 1661-1674. e11. DOI: 10.1053/j.gastro.2019.01.260.
- [11] Visrodia K, Singh S, Krishnamoorthi R, et al. Magnitude of Missed Esophageal Adenocarcinoma After Barrett's Esophagus Diagnosis: A Systematic Review and Meta-analysis [J]. *Gastroenterology*, 2016, 150 (3): 599-607. e7; quiz e14-15. DOI: 10.1053/j.gastro.2015.11.040.
- [12] Robertson DJ, Lieberman DA, Winawer SJ, et al. Colorectal cancers soon after colonoscopy: a pooled multicohort analysis [J]. *Gut*, 2014, 63 (6): 949-956. DOI: 10.1136/gutjnl-2012-303796.
- [13] Pickhardt PJ, Nugent PA, Mysliwiec PA, et al. Location of adenomas missed by optical colonoscopy [J]. *Ann Intern Med*, 2004, 141 (5): 352-359. DOI: 10.7326/0003-4819-141-5-200409070-00009.
- [14] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521 (7553): 436-444. DOI: 10.1038/nature14539.
- [15] Esteva A, Kuprel B, Novoa RA, et al. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks [J]. *Nature*, 2017, 546 (7660): 686. DOI: 10.1038/nature22985.
- [16] Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs [J]. *JAMA*, 2016, 316 (22): 2402-2410. DOI: 10.1001/jama.2016.17216.
- [17] Wang Z, Zhao S, Bai Y. Artificial Intelligence as a Third Eye in Lesion Detection by Endoscopy [J]. *Clin Gastroenterol Hepatol*, 2018, 16 (9): 1537. DOI: 10.1016/j.cgh.2018.04.032.
- [18] Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model [J]. *Gut*, 2019, 68 (1): 94-100. DOI: 10.1136/gutjnl-2017-314547.
- [19] Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study [J]. *Gut*, 2019, 68 (10): 1813-1819. DOI: 10.1136/gutjnl-2018-317500.
- [20] Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy [J]. *Gut*, 2019, 68 (12): 2161-2169. DOI: 10.1136/gutjnl-2018-317366.
- [21] Wang P, Xiao X, Glissen Brown JR, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy [J]. *Nat Biomed Eng*, 2018, 2 (10): 741-748. DOI: 10.1038/s41551-018-0301-3.
- [22] Alagappan M, Brown J, Mori Y, et al. Artificial intelligence in gastrointestinal endoscopy: The future is almost here [J]. *World J Gastrointest Endosc*, 2018, 10 (10): 239-249. DOI: 10.4253/wjge.v10.i10.239.
- [23] He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine [J]. *Nat Med*, 2019, 25 (1): 30-36. DOI: 10.1038/s41591-018-0307-0.
- [24] Horie Y, Yoshio T, Aoyama K, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks [J]. *Gastrointest Endosc*, 2019, 89 (1): 25-32. DOI: 10.1016/j.gie.2018.07.037.
- [25] Ahmad OF, Soares AS, Mazomenos E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions [J]. *Lancet Gastroenterol Hepatol*, 2019, 4 (1): 71-80. DOI: 10.1016/S2468-1253(18)30282-6.
- [26] Ozawa T, Ishihara S, Fujishiro M, et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis [J]. *Gastrointest Endosc*, 2019, 89 (2): 416-421. e1. DOI: 10.1016/j.gie.2018.10.020.
- [27] Vinsard DG, Mori Y, Misawa M, et al. Quality assurance of computer-aided detection and diagnosis in colonoscopy [J]. *Gastrointest Endosc*, 2019, 90 (1): 55-63. DOI: 10.1016/j.gie.2019.03.019.
- [28] de Lange T, Halvorsen P, Riegler M. Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy [J]. *World J Gastroenterol*, 2018, 24 (45): 5057-5062. DOI: 10.3748/wjg.v24.i45.5057.
- [29] 国家药品监督管理局医疗器械技术审评中心. 深度学习辅助决策医疗器械软件审评要点 [EB/OL]. [2019-07-03]. <https://www.cmde.org.cn/CL0050/19360.html>.
- [30] Price WN 2nd, Cohen IG. Privacy in the age of medical big data [J]. *Nat Med*, 2019, 25 (1): 37-43. DOI: 10.1038/s41591-018-0272-7.

(收稿日期: 2020-05-09)

(本文编辑: 唐涌进 周昊)