

人工智能系统预测早期胃癌浸润深度和分化状态的能力

——在单中心和多中心视频集中的表现

杨婷 董泽华 陶道 吴练练 于红刚

武汉大学人民医院消化内科, 武汉 430060

通信作者: 于红刚, Email: yuhonggang@whu.edu.cn

【摘要】 目的 评价 ENDOANGEL 人工智能系统在多样性测试集(多中心来源视频集)中预测早期胃癌浸润深度和分化状态的能力,并在完成系统升级后,对新系统的上述能力进行评价。方法 在已完成的 2020 年开展的单中心视频集早期胃癌诊断人机比赛基础上,于 2022 年开展来自国内 10 个省份 30 家医院 30 名内镜医师参与的人机比赛,采用国内 8 个省、市 12 家医院回顾性收集的视频作为多中心视频集。研究共分成三个阶段:第一阶段,用多中心视频集重新测试 ENDOANGEL,比较 ENDOANGEL 在单中心视频集、多中心视频集中的测试结果,并完成系统升级(ENDOANGEL-2022);第二阶段,使用多中心视频集,开展 ENDOANGEL-2022 与 30 名内镜医师参与的人机比赛,比较 ENDOANGEL-2022、ENDOANGEL 以及内镜医师在多中心视频集中的表现;第三阶段,使用 2020 年人机比赛的单中心视频集测试 ENDOANGEL-2022,比较新系统在单中心、多中心视频集中的表现。结果 相比单中心视频集中的表现,在多中心视频集中,ENDOANGEL 预测早期胃癌黏膜下浸润的敏感度明显下降[18.18% (2/11)比 70.00% (7/10), $P=0.030$],预测未分化型早期胃癌方面表现相似($P>0.05$)。在多中心视频集中,ENDOANGEL-2022 预测早期胃癌黏膜下浸润的敏感度高于 ENDOANGEL [40.00% (4/10)比 18.18% (2/11), $P=0.361$]、低于 30 名内镜医师的平均水平[40.00% 比 52.04% (95%CI: 43.70%~60.38%), $P<0.001$]、特异度低于 ENDOANGEL [82.86% (29/35)比 100.00% (34/34), $\chi^2=4.41$, $P=0.036$]、高于 30 名内镜医师的平均水平[82.86% 比 68.97% (95%CI: 60.83%~77.11%), $P=0.018$]、准确率低于 ENDOANGEL [73.33% (33/45)比 80.00% (36/45), $\chi^2=0.56$, $P=0.455$]、高于 30 名内镜医师的平均水平[73.33% 比 65.30% (95%CI: 60.61%~69.99%), $P=0.018$]; ENDOANGEL-2022 预测未分化型早期胃癌的敏感度高于 ENDOANGEL [71.43% (5/7)比 57.14% (4/7), $P>0.999$]和 30 名内镜医师的平均水平[71.43% 比 63.11% (95%CI: 55.58%~70.64%), $P=0.031$]、特异度低于 ENDOANGEL [76.32% (29/38)比 78.95% (30/38), $\chi^2=0.08$, $P=0.783$]、高于 30 名内镜医师的平均水平[76.32% 比 65.27% (95%CI: 59.10%~71.44%), $P=0.004$]、准确率与 ENDOANGEL 相同[75.56% (34/45)比 75.56% (34/45), $\chi^2=0.00$, $P>0.999$]、高于 30 名内镜医师的平均水平[75.56% 比 65.10% (95%CI: 59.96%~70.24%), $P<0.001$]。相比单中心视频集中的表现, ENDOANGEL-2022 在多中心视频集中预测早期胃癌黏膜下浸润的敏感度[40.00% 比 60.00% (6/10), $P=0.656$]、特异度[82.86% 比 93.75% (15/16), $\chi^2=0.37$, $P=0.542$]和准确率[73.33% 比 80.77% (21/26), $\chi^2=0.50$, $P=0.479$]均下降;预测未分化型早期胃癌的敏感度上升[71.43% 比 37.50% (3/8), $P=0.315$]、特异度[76.32% 比 100.00% (18/18), $\chi^2=3.48$, $P=0.062$]和准确率[75.56% 比 80.77% (21/26), $\chi^2=0.26$, $P=0.612$]下降。结论 多中心数据导致病例的异质性更大,人工智能系统容易给出错误预测,但表现优于内镜医师平均水平。

【关键词】 人工智能; 早期胃癌; 浸润深度; 分化状态

DOI: 10.3760/cma.j.cn321463-20231206-00514

收稿日期 2023-12-06 本文编辑 顾文景

引用本文:杨婷,董泽华,陶道,等.人工智能系统预测早期胃癌浸润深度和分化状态的能力:在单中心和多中心视频集中的表现[J].中华消化内镜杂志,2025,42(6):452-461. DOI: 10.3760/cma.j.cn321463-20231206-00514.



Ability of artificial intelligence system to predict invasion depth and differentiation status of early gastric cancer: performance in single-center and multi-center videos

Yang Ting, Dong Zehua, Tao Xiao, Wu Lianlian, Yu Honggang

Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan 430060, China

Corresponding author: Yu Honggang, Email: yuhonggang@whu.edu.cn

【Abstract】 Objective To evaluate the ability of ENDOANGEL artificial intelligence system to predict invasion depth and differentiation status of early gastric cancer using more diverse multi-center videos, and to test the performance of the new system upgraded from ENDOANGEL. **Methods** Based on the completed 2020 man-machine competition for early gastric cancer diagnosis using single-center videos, the second man-machine competition was conducted in 2022, involving 30 endoscopists from 30 hospitals across 10 Chinese provinces. A multi-center video cohort was retrospectively collected from 12 institutions in 8 provinces/municipalities in China. The study proceeded in 3 stages. First, the ENDOANGEL was re-tested on multi-center videos, its performance on single and multi-center videos was compared, then the ENDOANGEL was upgraded to ENDOANGEL-2022. Second, the second man-machine competition was conducted between ENDOANGEL-2022 and 30 endoscopists using multi-center videos, and the performance between ENDOANGEL-2022, ENDOANGEL and endoscopists on multi-center videos were compared. Third, the ENDOANGEL-2022 was re-tested on the single-center videos previously collected in 2020, its performance on single and multi-center videos was also compared. **Results** Compared with the performance on single-center videos, the sensitivity of ENDOANGEL for predicting submucosal invasion of early gastric cancer decreased significantly [18.18% (2/11) VS 70.00% (7/10), $P=0.030$], but demonstrated comparable ability to predict undifferentiated type of early gastric cancer ($P>0.05$). On multi-center videos, in the respect of predicting submucosal invasion of early gastric cancer, the sensitivity of ENDOANGEL-2022 was higher than that of ENDOANGEL [40.00% (4/10) VS 18.18% (2/11), $P=0.361$], but inferior to that of 30 endoscopists [40.00% VS 52.04% (95%CI: 43.70%-60.38%), $P<0.001$]. The specificity of ENDOANGEL-2022 was lower than that of ENDOANGEL [82.86% (29/35) VS 100.00% (34/34), $\chi^2=4.41$, $P=0.036$] and higher than that of 30 endoscopists [82.86% VS 68.97% (95%CI: 60.83%-77.11%), $P=0.018$], the accuracy of ENDOANGEL-2022 was lower than that of ENDOANGEL [73.33% (33/45) VS 80.00% (36/45), $\chi^2=0.56$, $P=0.455$] and higher than that of 30 endoscopists [73.33% VS 65.30% (95%CI: 60.61%-69.99%), $P=0.018$]. In the respect of predicting undifferentiated type of early gastric cancer, the sensitivity of ENDOANGEL-2022 was higher than that of ENDOANGEL [71.43% (5/7) VS 57.14% (4/7), $P>0.999$] and 30 endoscopists [71.43% VS 63.11% (95%CI: 55.58%-70.64%), $P=0.031$], the specificity of ENDOANGEL-2022 was lower than that of ENDOANGEL [76.32% (29/38) VS 78.95% (30/38), $\chi^2=0.08$, $P=0.783$] and higher than that of 30 endoscopists [76.32% VS 65.27% (95%CI: 59.10%-71.44%), $P=0.004$], the accuracy of ENDOANGEL-2022 was similar to that of ENDOANGEL [75.56% (34/45) VS 75.56% (34/45), $\chi^2=0.00$, $P>0.999$] and higher than that of 30 endoscopists [75.56% VS 65.10% (95%CI: 59.96%-70.24%), $P<0.001$]. Compared with performance in single center videos, the sensitivity [40.00% VS 60.00% (6/10), $P=0.656$], specificity [82.86% VS 93.75% (15/16), $\chi^2=0.37$, $P=0.542$] and accuracy [73.33% VS 80.77% (21/26), $\chi^2=0.50$, $P=0.479$] of ENDOANGEL-2022 for predicting submucosal invasion of early gastric cancer decreased; in predicting undifferentiated type of early gastric cancer, the sensitivity of ENDOANGEL-2022 increased [71.43% VS 37.50% (3/8), $P=0.315$], while the specificity [76.32% VS 100.00% (18/18), $\chi^2=3.48$, $P=0.062$] and accuracy [75.56% VS 80.77% (21/26), $\chi^2=0.26$, $P=0.612$] decreased. **Conclusion** Multi-center cases introduce greater heterogeneity that may reduce artificial intelligence prediction accuracy, but the artificial intelligence system still outperforms endoscopists.

【Key words】 Artificial intelligence; Early gastric cancer; Invasion depth; Differentiation status

胃癌是全球第三大癌症相关死亡原因^[1],早期诊断和治愈性治疗是减少其相关死亡的关键措施^[2]。内镜切除是早期胃癌的治疗方式之一,具有创伤小、成本-效益优越的特点^[3],但仅适用于淋巴结转移风险可忽略不计的早期胃癌,而慎用于浸润深度达黏膜下层或未分化型等淋巴结转移风险较高的早期胃癌^[4-5]。因此,准确诊断和识别早期胃癌的浸润深度和分化状态对患者的治疗和预后至

关重要。

基于计算机视觉的人工智能技术已被用于开发多种医学辅助诊断系统,然而不同研究中人工智能的诊断性能差异较大,部分原因是测试样本选择偏差,或测试样本多样性不足^[6-7],不利于准确评估人工智能的性能,阻碍其在临床中的应用。此外,之前多数研究是在静态的图片上测试或比较人工智能系统的性能^[8-9],而对于内镜诊断任务来说,人

人工智能能否处理实时数据很重要,因内镜操作是实时的过程,人工智能能否给予即时性的反馈,决定了其能否在临床使用。因此,有必要进一步探索人工智能系统在多样性案例中的表现,开发更贴合临床实践的人工智能系统。

2020年,我们开展了一项全国性的早期胃癌诊断的人机比赛,来自国内44家医院的46名内镜医师与名为ENDOANGEL的人工智能系统,就白光内镜下预测早期胃癌浸润深度和放大内镜窄带光成像下预测分化状态的能力进行比较^[10]。此次比赛的全部案例视频均来自北京大学肿瘤医院,均为临床工作中连续收集,这些视频的来源相对单一,可能缺乏多样性。因此,目前尚不清楚ENDOANGEL在更具多样性的测试集中的表现如何,是否存在潜在的挑战,以及用何种方式有助于应对潜在的挑战。为此,我们收集了来自12家医院的包含白光和染色放大观察的案例视频,将其作为多中心视频测试案例,并且开展了第二届早期胃癌人机比赛。此外,我们应用创新的训练算法,将ENDOANGEL升级为ENDOANGEL-2022,利用两届比赛收集的单中心及多中心案例,我们测试和对比了ENDOANGEL-2022与ENDOANGEL在单中心及多中心案例中的表现,并与参与人机比赛的内镜医师进行能力比较。

资料与方法

一、多中心视频收集和数据预处理

本研究从国内8个省、市的12家医院回顾性收集了一个多中心视频集。视频和病变的纳入标准:(1)患者年龄 ≥ 18 岁;(2)视频同时包含病灶白光内镜、放大内镜窄带光成像下的观察片段;(3)病变病理确诊为早期胃癌,有浸润深度和分化程度信息。排除标准:(1)既往有胃切除术病史;(2)因活动性出血、黏液致病变在放大内镜窄带光成像下难以观察;(3)因检查前有活检或溃疡愈合致放大内镜窄带光成像下难以诊断病变;(4)黏膜下异常;(5)淋巴瘤;(6)其他影响本次研究的情况。病理诊断由具有超过10年病理诊断经验的胃肠病理学家根据WHO分类进行重新审核^[11]。

一位研究助理回顾所有的视频,并将原始视频编辑成配对的白光内镜和放大内镜窄带光成像视频片段,这些片段包含了最佳的病变观察视野。所有纳入的视频都由研究助理严格保密保存。来自

于前期开展的第一届比赛中的视频集作为单中心视频集也被纳入研究。

二、研究设计

本研究属于诊断性研究,整体可分为三个阶段:第一阶段,收集多中心视频集,在多中心视频中重新测试ENDOANGEL,然后将ENDOANGEL升级并命名为ENDOANGEL-2022;第二阶段,举办人机比赛,比较ENDOANGEL-2022与内镜医师的表现;第三阶段,用之前收集的单中心视频集测试ENDOANGEL-2022。具体如下:

在第一阶段,我们用多中心视频集重新测试ENDOANGEL,比较ENDOANGEL在单中心视频集、多中心视频集中的测试结果。其中ENDOANGEL在单中心视频集中的测试结果可从之前研究中获得。然后,我们将ENDOANGEL升级并命名为ENDOANGEL-2022。

在第二阶段,我们于2022年7月29日在武汉举办了一场线下人机大赛,招募了一批内镜医师参加本次比赛,在多中心视频集中,与ENDOANGEL-2022判断早期胃癌浸润深度和分化状态的表现进行比较。所有内镜医师使用相同的笔记本电脑观看视频并在电子答题卡上填写答案。同时,对ENDOANGEL-2022、ENDOANGEL-2022在多中心视频集中的表现进行比较。

在第三阶段,我们在前一次比赛的单中心视频集中重新测试ENDOANGEL-2022,比较其在单中心、多中心视频集中的表现。

在本研究中,我们针对每一个视频(每一道“题目”)定义了一个指标,命名为错误率,其计算方式为错答的内镜医师人数/所有内镜医师人数。我们通过计算视频集中每个病变的错误率,以评估该病变的作答难度。此外,我们把50%的错误率设置为阈值,将视频中的病例分类为普通(错误率 $< 50\%$)或困难(错误率 $\geq 50\%$)。

三、参加比赛的内镜医师

我们在2021年6月发布了比赛通知,由于疫情的影响,比赛推迟到2022年7月举行。参与研究的医师的纳入标准:(1)签署知情同意书;(2)有超过1年的内镜操作经验且累积操作量超过2000例;(3)有独立操作放大内镜窄带光成像的经验;(4)在临床工作中有独立诊断早期胃癌的经验。排除标准:(1)所在医院提供了参赛视频;(2)客观原因致无法参赛或不愿参赛。根据内镜医师放大内镜窄

带光成像操作经验,将内镜医师分为初级医师(操作时间不足1年)、中级医师(1~3年)、专家(超过3年)。所有的内镜医师完成了基本信息问卷调查。

四、比赛流程

比赛包含两部分:第一部分,播放剪辑好的白光内镜视频,人工智能和内镜医师回答该病变是否为瘤变;第二部分,播放剪辑好的白光内镜和放大内镜窄带光成像视频,人工智能和内镜医师回答病变是否为早期胃癌,如病变被诊断为早期胃癌则进一步预测其浸润深度(黏膜层或黏膜下层)和分化状态(未分化或已分化)。人工智能的答案由一位工作人员及时记录。本研究仅对人工智能和内镜医师预测早期胃癌浸润深度、分化状态展开分析,人工智能或医师的预测病例数=早期胃癌病例数-早期胃癌漏诊例数。

五、ENDOANGEL-2022的构建

ENDOANGEL是一个实时人工智能系统,功能覆盖早期胃癌诊断的不同方面,包括在白光内镜下检测胃肿瘤,在放大内镜窄带光成像下识别早期胃癌,在白光内镜下预测早期胃癌浸润深度,在放大内镜窄带光成像下预测分化状态。在本研究中,我们仅就ENDOANGEL预测早期胃癌浸润深度和分化状态的功能进行评价。

ENDOANGEL-2022由ENDOANGEL升级而来。我们采用基于mean-teacher的半监督算法,对早期胃癌浸润深度(模型1)和分化状态(模型2)预测功能进行优化。使用半监督算法过程中,所有训练集图片被随机均分为10份,采用一部分有标签的图片作为新训练集,其余图片弃用。自使用10%含标签的训练集起始,每次叠加10%进行模型训练,直到使用90%含标签的训练集停止继续叠加。因此,针对每个功能,使用半监督算法会训练得到9个使用不同比例含标签训练集的模型。

针对早期胃癌浸润深度预测功能(模型1),训练集包含812例胃癌患者的5247张白光内镜图片,测试集包含62例胃癌患者的228张白光内镜图片。我们采用半监督算法对该功能进行优化,并在测试集上与传统监督算法构建的模型进行对比。最终,使用40%标记的半监督模型表现最优,其敏感度和特异度分别为96.6%和82.4%,优于其他所有的半监督模型及原始的全监督模型。因此,在本

研究中选择使用40%标记的半监督模型作为最终模型。

针对分化状态预测功能(模型2),训练集包含157例分化型早期胃癌患者的2271张放大内镜窄带光成像图片和49例未分化型早期胃癌患者的1037张放大内镜窄带光成像图片,测试集包含139例早期胃癌患者的1807张图片。采用半监督算法对该功能进行优化,并在测试集上与传统监督算法构建的模型进行对比。最终,使用50%标记的半监督模型表现最优,其敏感度和特异度分别为95.0%和91.9%,分化型、未分化型早期胃癌的最佳阈值分别为<0.15和>0.85,优于其他所有的半监督模型及原始的全监督模型。因此,在本研究中选择使用50%标记的半监督模型作为最终模型。

六、统计分析

计算人工智能和内镜医师在白光内镜下预测早期胃癌浸润深度和放大内镜窄带光成像下预测早期胃癌分化状态的准确率、敏感度、特异度、阳性预测值和阴性预测值。使用 χ^2 检验、两独立样本 t 检验比较单中心与多中心视频测试集中患者及病变的基线特征。使用 χ^2 检验、Fisher确切概率法和Mann-Whitney U 检验比较两个人工智能系统与内镜医师在单中心和多中心视频测试集中的表现。 P (双侧)<0.05为差异有统计学意义。

结 果

一、2020年单中心和2022年多中心比赛的基本情况

在2022年的比赛中,收集了来自国内8个省市的12家大型三甲医院的46例早期胃癌患者的46个早期胃癌视频,其中35个为黏膜内癌、11个为黏膜下癌,39个为分化型、7个为未分化型。在2022年的比赛中,共有来自国内10个省份40家医院的42名医师申请参加比赛,但由于时间冲突,12名医师未能参加比赛,最终来自国内10个省份30家医院的30名医师抵达武汉参加比赛。2020年单中心和2022年多中心比赛中患者及病变的基线特征见表1。

二、ENDOANGEL在多中心视频集中的表现

相比单中心视频集中的表现,在多中心视频集

中 ENDOANGEL 预测早期胃癌黏膜下浸润的敏感度显著下降 ($P=0.030$), 其他诊断指标变化不显著 ($P>0.05$), 见表 2。

相比单中心视频集中的表现, 在多中心视频集中 ENDOANGEL 预测未分化型早期胃癌的各项诊断指标未发生明显变化 ($P>0.05$), 见表 3。

表 1 2020 年单中心和 2022 年多中心比赛中患者及病变的基线特征

项目	单中心 ($n=28$)	多中心 ($n=46$)	χ^2/t 值	P 值
性别(例,男/女)	18/10	31/15	0.08	0.784
年龄(岁, $\bar{x} \pm s$)	61.36 \pm 5.81	61.15 \pm 10.59	0.11	0.915
幽门螺杆菌感染(例)			26.30	<0.001
阴性	4	19		
阳性	1	17		
未知	23	10		
病变部位(例)			6.15	0.104
胃窦	5	13		
胃体下部+胃角	4	15		
胃体中上部	5	6		
胃底+贲门	14	12		
病变巴黎分型(例)			3.71	0.628
0-I 型	2	1		
0-II a 型	5	10		
0-II b 型	3	9		
0-II c 型	9	13		
0-III 型	1	0		
混合型 ^a	8	13		
早期胃癌浸润深度 (例,黏膜内/黏膜下)	18/10	35/11	1.19	0.275
早期胃癌组织分型 (例,分化型/未分化型)	20/8	39/7	1.92	0.166

注:^a处混合型指 0-II a+II c 型和 0-II c+II a 型

表 2 ENDOANGEL 人工智能系统在多中心和单中心视频集中预测早期胃癌黏膜下浸润的表现

视频来源	敏感度	特异度	准确率	阳性预测值	阴性预测值
多中心 ^a	18.18%(2/11)	100.00%(34/34)	80.00%(36/45)	100.00%(2/2)	79.07%(34/43)
单中心	70.00%(7/10)	83.33%(15/18)	78.57%(22/28)	70.00%(7/10)	83.33%(15/18)
统计量		$\chi^2=3.34$	$\chi^2=0.02$		$\chi^2=0.00$
P 值	0.030 ^b	0.068	0.883	>0.999 ^b	0.977

注:^aENDOANGEL 漏诊了多中心视频集中 1 例早期胃癌, 因此仅对 45 例早期胃癌进行了浸润深度预测;^b使用 Fisher 确切概率法

表 3 ENDOANGEL 人工智能系统在多中心和单中心视频集中预测未分化型早期胃癌的表现

视频来源	敏感度	特异度	准确率	阳性预测值	阴性预测值
多中心 ^a	57.14%(4/7)	78.95%(30/38)	75.56%(34/45)	33.33%(4/12)	90.91%(30/33)
单中心	50.00%(4/8)	80.00%(16/20)	71.43%(20/28)	50.00%(4/8)	80.00%(16/20)
统计量		$\chi^2=0.00$	$\chi^2=0.15$		0.52
P 值	>0.999 ^b	>0.999	0.696	0.648 ^b	0.472

注:^aENDOANGEL 漏诊了多中心视频集中 1 例早期胃癌, 因此仅对 45 例早期胃癌进行了分化状态预测;^b使用 Fisher 确切概率法

三、ENDOANGEL-2022 在多中心视频集中的表现

1. 与 ENDOANGEL 比较: 在多中心视频集中, ENDOANGEL-2022 预测早期胃癌黏膜下浸润的敏感度高于 ENDOANGEL, 但差异无统计学意义 ($P=0.361$); 特异度明显低于 ENDOANGEL, 差异有统计学意义 ($P=0.036$), 见表 4。

在多中心视频集中, ENDOANGEL-2022 预测未分化型早期胃癌的敏感度高于 ENDOANGEL, 但差异无统计学意义 ($P>0.999$); 特异度、准确率、阳性预测值、阴性预测值与 ENDOANGEL 接近, 差异均无统计学意义 ($P>0.05$), 见表 5。

2. 人机大赛: 内镜医师在多中心视频集中预测早期胃癌黏膜下浸润的表现见表 6。在多中心视频集中, ENDOANGEL-2022 预测早期胃癌黏膜下浸润的敏感度低于专家组 (由 9 名内镜专家组成) ($P=0.730$)、中级医师组 (由 12 名中级内镜医师组成) ($P=0.039$)、初级医师组 (由 9 名初级内镜医师组成) ($P=0.004$) 以及 30 名内镜医师的平均水平 ($P<0.001$); 特异度高于专家组 ($P=0.730$)、中级医师组 ($P=0.178$)、初级医师组 ($P=0.050$) 以及 30 名内镜医师的平均水平 ($P=0.018$); 准确率高于专家组 ($P=0.730$)、中级医师组 ($P=0.039$)、初级医师组 ($P=0.050$) 以及 30 名内镜医师的平均水平 ($P=0.018$); 阳性预测值低于专家组 ($P=0.730$)、中级医师组 ($P>0.999$) 以及 30 名内镜医师的平均水平 ($P=0.155$), 高于初级医师组 ($P=0.004$); 阴性预测值高于专家组 ($P=0.258$)、中级医师组 ($P>0.999$)、初级医师组 ($P=0.258$) 以及 30 名内镜医师的平均水平 ($P=0.155$)。

内镜医师在多中心视频集中预测未分化型早期胃癌的表现见表 7。在多中心视频集中, ENDOANGEL-2022 预测未分化型早期胃癌的敏感度高于专家组($P=0.258$)、中级医师组($P=0.089$)、初级医师组($P=0.730$)以及 30 名内镜医师的平均水平($P=0.031$); 特异度高于专家组($P=0.050$)、中级医师组($P=0.514$)、初级医师组($P=0.050$)以及 30 名内镜医师的平均水平($P=0.004$); 准确率高于专家组($P=0.050$)、中级医师组($P=0.039$)、初级医师组($P=$

0.050)以及 30 名内镜医师的平均水平($P<0.001$); 阳性预测值低于专家组($P=0.050$)、中级医师组($P<0.001$)、初级医师组($P=0.004$)以及 30 名内镜医师的平均水平($P<0.001$); 阴性预测值高于专家组($P=0.258$)、中级医师组($P=0.039$)、初级医师组($P=0.258$)以及 30 名内镜医师的平均水平($P=0.004$)。

四、ENDOANGEL-2022 在单中心视频集中的表现

在单中心视频集中, ENDOANGEL-2022 预测

表 4 ENDOANGEL-2022 与 ENDOANGEL 人工智能系统在多中心视频集中预测早期胃癌黏膜下浸润的表现

人工智能系统	敏感度	特异度	准确率	阳性预测值	阴性预测值
ENDOANGEL-2022 ^a	40.00%(4/10)	82.86%(29/35)	73.33%(33/45)	40.00%(4/10)	82.86%(29/35)
ENDOANGEL ^a	18.18%(2/11)	100.00%(34/34)	80.00%(36/45)	100.00%(2/2)	79.07%(34/43)
统计量		$\chi^2=4.41$	$\chi^2=0.56$		$\chi^2=0.22$
P 值	0.361 ^b	0.036	0.455	0.455 ^b	0.636

注:^aENDOANGEL-2022 和 ENDOANGEL 均漏诊了多中心视频集中 1 例早期胃癌, 因此二者均仅对 45 例早期胃癌进行了浸润深度预测; ^b使用 Fisher 确切概率法

表 5 ENDOANGEL-2022 与 ENDOANGEL 人工智能系统在多中心视频集中预测未分化型早期胃癌的表现

人工智能系统	敏感度	特异度	准确率	阳性预测值	阴性预测值
ENDOANGEL-2022 ^a	71.43%(5/7)	76.32%(29/38)	75.56%(34/45)	35.71%(5/14)	93.55%(29/31)
ENDOANGEL ^a	57.14%(4/7)	78.95%(30/38)	75.56%(34/45)	33.33%(4/12)	90.91%(30/33)
统计量		$\chi^2=0.08$	$\chi^2=0.00$		$\chi^2=0.00$
P 值	>0.999 ^b	0.783	>0.999	>0.999 ^b	>0.999

注:^aENDOANGEL-2022 和 ENDOANGEL 均漏诊了多中心视频集中 1 例早期胃癌, 因此二者均仅对 45 例早期胃癌进行了分化状态预测; ^b使用 Fisher 确切概率法

表 6 内镜医师在多中心视频集中预测早期胃癌黏膜下浸润的表现

诊断指标	专家组(9名成员)	中级医师组(12名成员)	初级医师组(9名成员)	合计
敏感度	49.49% (95%CI:33.60%~65.39%)	51.92% (95%CI:38.34%~65.50%)	54.75% (95%CI:33.89%~75.61%)	52.04% (95%CI:43.70%~60.38%)
特异度	73.90% (95%CI:60.36%~87.44%)	70.75% (95%CI:57.51%~83.99%)	61.68% (95%CI:41.00%~82.35%)	68.97% (95%CI:60.83%~77.11%)
准确率	67.88% (95%CI:59.34%~76.42%)	66.47% (95%CI:58.62%~74.32%)	61.15% (95%CI:50.02%~72.28%)	65.30% (95%CI:60.61%~69.99%)
阳性预测值	43.40% (95%CI:32.50%~54.29%)	47.27% (95%CI:30.55%~63.98%)	33.10% (95%CI:20.61%~45.59%)	41.85% (95%CI:34.13%~49.57%)
阴性预测值	81.56% (95%CI:77.39%~85.78%)	82.36% (95%CI:78.97%~85.74%)	81.76% (95%CI:79.99%~83.53%)	81.95% (95%CI:80.25%~83.64%)

表 7 内镜医师在多中心视频集中预测未分化型早期胃癌的表现

诊断指标	专家组(9名成员)	中级医师组(12名成员)	初级医师组(9名成员)	合计
敏感度	64.39% (95%CI:51.66%~77.13%)	57.96% (95%CI:42.74%~73.18%)	68.70% (95%CI:54.64%~82.77%)	63.11% (95%CI:55.58%~70.64%)
特异度	64.33% (95%CI:52.16%~76.49%)	66.90% (95%CI:55.39%~78.42%)	64.04% (95%CI:51.15%~76.94%)	65.27% (95%CI:59.10%~71.44%)
准确率	64.47% (95%CI:54.40%~74.54%)	65.87% (95%CI:56.16%~75.59%)	64.70% (95%CI:54.06%~75.34%)	65.10% (95%CI:59.96%~70.24%)
阳性预测值	24.34% (95%CI:16.80%~31.88%)	27.52% (95%CI:12.63%~42.40%)	26.26% (95%CI:19.05%~33.47%)	26.19% (95%CI:20.16%~32.22%)
阴性预测值	91.77% (95%CI:89.19%~94.36%)	90.41% (95%CI:85.58%~93.24%)	92.47% (95%CI:88.99%~95.95%)	91.44% (95%CI:89.89%~92.98%)

早期胃癌黏膜下浸润的敏感度($P>0.999$)和阴性预测值($P>0.999$)低于 ENDOANGEL, 特异度($P=0.604$)、准确率($P=0.841$)和阳性预测值($P=0.603$)高于 ENDOANGEL, 见表 8。

在单中心视频集中, ENDOANGEL-2022 预测未分化型早期胃癌的特异度($P=0.107$)、准确率($P=0.422$)和阳性预测值($P=0.576$)高于 ENDOANGEL, 敏感度($P>0.999$)和阴性预测值($P>0.999$)低于 ENDOANGEL, 见表 9。

五、ENDOANGEL-2022 在单中心与多中心视频集中的表现差异

多中心视频集中 ENDOANGEL-2022 预测早期胃癌黏膜下浸润的敏感度($P=0.656$)、特异度($P=0.542$)、准确率($P=0.479$)和阳性预测值($P=0.134$)

较单中心视频集中下降, 阴性预测值较单中心视频集中上升($P>0.999$), 见表 10。

多中心视频集中 ENDOANGEL-2022 预测未分化型早期胃癌的敏感度($P=0.315$)和阴性预测值($P=0.374$)较单中心视频集中上升, 特异度($P>0.062$)、准确率($P=0.612$)和阳性预测值($P=0.275$)较单中心视频集中下降, 见表 11。

六、多中心视频集中 ENDOANGEL-2022 预测错误情况分析

单因素分析发现, 黏膜下早期胃癌更容易导致 ENDOANGEL-2022 作出错误的浸润深度预测($P=0.022$, 表 12), 内镜医师预测错误率 $>50\%$ 的病变更容易导致 ENDOANGEL-2022 作出错误的分化程度预测($P=0.021$, 表 13)。

表 8 ENDOANGEL-2022 与 ENDOANGEL 人工智能系统在单中心视频集中预测早期胃癌黏膜下浸润的表现

人工智能系统	敏感度	特异度	准确率	阳性预测值	阴性预测值
ENDOANGEL-2022 ^a	60.00%(6/10)	93.75%(15/16)	80.77%(21/26)	85.71%(6/7)	78.95%(15/19)
ENDOANGEL	70.00%(7/10)	83.33%(15/18)	78.57%(22/28)	70.00%(7/10)	83.33%(15/18)
统计量			$\chi^2=0.04$		
P 值	$>0.999^b$	0.604 ^b	0.841	0.603 ^b	$>0.999^b$

注:^aENDOANGEL-2022 漏诊了单中心视频集中 2 例早期胃癌, 因此仅对 26 例早期胃癌进行了浸润深度预测;^b使用 Fisher 确切概率法

表 9 ENDOANGEL-2022 与 ENDOANGEL 人工智能系统在单中心视频集中预测未分化型早期胃癌的表现

人工智能系统	敏感度	特异度	准确率	阳性预测值	阴性预测值
ENDOANGEL-2022 ^a	37.50%(3/8)	100.00%(18/18)	80.77%(21/26)	75.00%(3/4)	81.82%(18/22)
ENDOANGEL	50.00%(4/8)	80.00%(16/20)	71.43%(20/28)	50.00%(4/8)	80.00%(16/20)
统计量			$\chi^2=0.64$		$\chi^2=0.00$
P 值	$>0.999^b$	0.107 ^b	0.422	0.576 ^b	>0.999

注:^aENDOANGEL-2022 漏诊了单中心视频集中 2 例早期胃癌, 因此仅对 26 例早期胃癌进行了分化状态预测;^b使用 Fisher 确切概率法

表 10 ENDOANGEL-2022 人工智能系统在单中心和多中心视频集中预测早期胃癌黏膜下浸润的表现

视频来源	敏感度	特异度	准确率	阳性预测值	阴性预测值
单中心 ^a	60.00%(6/10)	93.75%(15/16)	80.77%(21/26)	85.71%(6/7)	78.95%(15/19)
多中心 ^b	40.00%(4/10)	82.86%(29/35)	73.33%(33/45)	40.00%(4/10)	82.86%(29/35)
统计量		$\chi^2=0.37$	$\chi^2=0.50$		$\chi^2=0.00$
P 值	0.656 ^c	0.542	0.479	0.134 ^c	>0.999

注:^a单中心视频集中 ENDOANGEL-2022 漏诊了 2 例早期胃癌, 因此仅对 26 例早期胃癌进行了浸润深度预测;^bENDOANGEL-2022 漏诊了多中心视频集中 1 例早期胃癌, 因此仅对多中心视频集中 45 例早期胃癌进行了浸润深度预测;^c使用 Fisher 确切概率法

表 11 ENDOANGEL-2022 人工智能系统在单中心和多中心视频集中预测未分化型早期胃癌的表现

视频来源	敏感度	特异度	准确率	阳性预测值	阴性预测值
单中心 ^a	37.50%(3/8)	100.00%(18/18)	80.77%(21/26)	75.00%(3/4)	81.82%(18/22)
多中心 ^b	71.43%(5/7)	76.32%(29/38)	75.56%(34/45)	35.71%(5/14)	93.55%(29/31)
统计量		$\chi^2=3.48$	$\chi^2=0.26$		$\chi^2=0.79$
P 值	0.315 ^c	0.062	0.612	0.275 ^c	0.374

注:^a单中心视频集中 ENDOANGEL-2022 漏诊了 2 例早期胃癌, 因此仅对 26 例早期胃癌进行了分化状态预测;^bENDOANGEL-2022 漏诊了多中心视频集中 1 例早期胃癌, 因此仅对多中心视频集中 45 例早期胃癌进行了分化状态预测;^c使用 Fisher 确切概率法

表 12 多中心视频集中影响 ENDOANGEL-2022 准确预测早期胃癌黏膜下浸润的单因素分析

变量	预测准确 (n=33)	预测错误 (n=12)	P 值
幽门螺杆菌感染(例)			>0.999
阴性	14	5	
阳性	12	5	
未知	7	2	
病变部位(例)			>0.999
胃窦	10	3	
胃体下部+胃角	10	4	
胃体中上部	4	2	
胃底+贲门	9	3	
病变巴黎分型(例)			0.260
0-I 型	1	0	
0-II a 型	7	3	
0-II b 型	4	4	
0-II c 型	9	4	
0-III 型	0	0	
混合型 ^a	12	1	
早期胃癌浸润深度(例)			0.022
黏膜内	29	6	
黏膜下	4	6	
早期胃癌组织分型			0.129
分化型	30	8	
未分化型	3	4	
错误率(例) ^b			0.077
≤50%	29	7	
>50%	4	5	

注：^a处混合型指 0-II a+II c 型和 0-II c+II a 型；^b本研究中，针对每一个视频（每一道“题目”）定义了一个指标，命名为错误率，其计算方式为错答的内镜医师人数/所有内镜医师人数

讨 论

早期胃癌患者得益于内镜切除方式的变革，因为内镜切除有整块切除、住院时间短、五年生存率与外科手术相当的优点^[12-13]。根据 JGCA 指南，评估胃癌是分化型还是未分化型是决定患者手术方式的第一步^[14]，而鉴别胃癌浸润深度是黏膜内还是黏膜下则是选择内镜切除还是外科手术最难点^[15]，因为胃癌为未分化型、浸润深度达黏膜下层者淋巴结转移风险较高，而内镜切除仅适合于淋巴结转移风险可忽略不计的早期胃癌^[4-5]。内镜下准确区分胃癌为黏膜内癌或黏膜下癌、分化型或未分化型是一个巨大的挑战，尽管有多项研究应用人工智能对其进行预测，但表现差异较大，测试集之间的异质性和多样性是潜在原因^[6-7]。因此，有必要

表 13 多中心视频集中影响 ENDOANGEL-2022 准确预测未分化型早期胃癌的单因素分析

变量	预测准确 (n=34)	预测错误 (n=11)	P 值
幽门螺杆菌感染(例)			0.170
阴性	16	3	
阳性	10	7	
未知	8	1	
病变部位(例)			0.422
胃窦	10	3	
胃体下部+胃角	9	5	
胃体中上部	4	2	
胃底+贲门	11	1	
病变巴黎分型(例)			0.290
0-I 型	1	0	
0-II a 型	9	1	
0-II b 型	4	4	
0-II c 型	9	4	
0-III 型	0	0	
混合型 ^a	11	2	
早期胃癌浸润深度(例)			>0.999
黏膜内	26	9	
黏膜下	8	2	
早期胃癌组织分型			>0.999
分化型	29	9	
未分化型	5	2	
错误率(例) ^b			0.021
≤50%	27	4	
>50%	7	7	

注：^a处混合型指 0-II a+II c 型和 0-II c+II a 型；^b本研究中，针对每一个视频（每一道“题目”）定义了一个指标，命名为错误率，其计算方式为错答的内镜医师人数/所有内镜医师人数

在一个多中心测试集中全面测试一个人工智能系统^[16]。在本研究中，我们建立了一个具有良好多样性的多中心视频集，有助于我们探索此问题。

为了探讨多中心视频集带来的差异，我们比较了人工智能从单中心应用到多中心时的表现。相比单中心视频集中的表现，在多中心视频集中，ENDOANGEL 和 ENDOANGEL-2022 预测早期胃癌黏膜下浸润的敏感度均有下降，而在预测未分化型早期胃癌方面，二者的特异度均有所下降。这些发现表明，多中心数据导致病例的异质性更大，人工智能容易给出错误的预测。

受病理图像中结直肠癌识别工作的启发，我们应用基于 mean - teacher 的半监督算法进行模型构建^[17]，它可以使用少量的标记数据，也可以使用大量的未标记数据，不仅节约了成本，而且其准确性

可优于监督学习^[17-18]。虽然在图像测试中,该方法在一定程度上提升了模型表现,然而面对异质性更强的多中心视频集时,ENDOANGEL-2022并未较前产生显著优势。这在一定程度上说明了早期胃癌深度预测和分化状态预测任务本身具备较大的难度,病例异质性对于人工智能有较大影响。最终,ENDOANGEL-2022在不同的多样性测试集中表现与ENDOANGEL相当且准确率高于内镜医师平均水平,表明半监督算法依旧有能力解决临床上标记数据不足的痛点,具备有效利用大量未标记数据来开发医疗人工智能系统的潜力。

在人工智能预测早期胃癌浸润深度、分化状态领域,存在测试集选择偏倚、在静态图片上测试等问题,降低了人工智能的临床意义。Nagao等^[19]开发了一个在白光内镜下预测胃癌浸润深度的人工智能系统,准确率、灵敏度和特异度分别为94.5%、84.4%和99.4%,然而,晚期癌症和早期胃癌都包括在测试集中。Zhu等^[20]的研究也发现了类似的结果。Hamada等^[21]的研究介绍了一个人工智能系统,其敏感度为84.9%,特异度为70.7%,但该系统只在静态图像中进行了测试。很少有研究旨在开发人工智能来识别早期胃癌的分化状态。我们的研究表明,ENDOANGEL-2022预测早期胃癌浸润深度、分化状态的准确率均显著高于30名内镜医师的平均水平,同时优于专家组医师。然而内镜医师和人工智能的表现均不可称之为理想,且人工智能升级后在同一届题目中的表现没有显著提高,其原因可能是早期胃癌浸润深度、分化状态具有多样性、不典型性,即便是经验丰富、接受过专业培训的内镜专家也很难迅速给出准确预测,普通医生则很少有机会获得这种培训,因此内镜医师在这两项任务上需要更好的训练,人工智能系统也需要进一步的训练和验证,以增强其性能。

本研究中我们还发现,在人机大赛中,ENDOANGEL-2022的特异度高于专家组、中级医师组、初级医师组以及30名内镜医师的平均水平,表明相比人工智能,内镜医师更倾向于将黏膜内早期胃癌预测为黏膜下早期胃癌,将分化型早期胃癌预测为未分化型早期胃癌,这可能是由于内镜医师担心作出相反的预测后患者选择内镜切除,导致其错过最佳外科手术时机,影响预后,为医患双方留下隐患。能够内镜切除却选择外科手术,这不仅给患者带来了额外的创伤,也增加了医疗经济负担,因此内镜医师在实时诊断时,可参考人工智能,提

高诊断特异度,为患者争取更大的效益。当然,人工智能设备研发团队也可根据临床实际需求,对人工智能设备判定的阈值作出合理调整,以更符合临床需求。

分析ENDOANGEL-2022预测错误的病例,我们仅发现病灶浸润达黏膜下、错误率大于50%更容易导致人工智能作出错误的预测,如果增大测试集中早期胃癌的样本量可能得出更多人工智能输出结果正确与否的影响因素。同时,该结果进一步说明这类病例内镜下表现特异性差,内镜医师亦容易判断错误,纳入更多相应的病例训练人工智能有望提高准确性。

本研究的局限性主要在于:(1)两届人机大赛同时测试了人工智能检测胃癌变、早期胃癌及预测早期胃癌浸润深度、分化状态四个功能,故纳入的早期胃癌病例有限,考虑继续增加测试集病例数将延长比赛时间,可能导致内镜医师疲劳而影响其发挥,以后可举办专门预测早期胃癌浸润深度、分化状态的人机大赛或培训类项目;(2)尽管我们邀请的内镜医师来自于不同级别的医院,但鲜有来自于基层医院的,以后应邀请更多来自于医疗资源不足的基层医院的医师。

总之,我们将ENDONGEL人工智能系统升级为ENDOANGEL-2022,ENDONGEL-2022预测早期胃癌浸润深度、分化状态方面的表现与ENDOANGE相当,当应用于更具多样性的多中心视频集时其表现有所下降,需对其进一步增强训练和验证;然而,人工智能的表现仍然优于内镜医师平均水平,提示早期胃癌浸润深度判断和分化状态判断两项技能仍旧是值得进一步攻克的难题。

利益冲突 所有作者声明不存在利益冲突

作者贡献声明 杨婷:分析、解释数据,撰写文章;董泽华、吴练练:酝酿和设计实验,对文章的知识性内容作批评性审阅;陶道:实施研究,采集数据;于红刚:研究指导,论文修改,经费支持

参 考 文 献

- [1] Smyth EC, Nilsson M, Grabsch HI, et al. Gastric cancer[J]. *Lancet*, 2020, 396(10251): 635-648. DOI: 10.1016/S0140-6736(20)31288-5.
- [2] Banks M, Graham D, Jansen M, et al. British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma[J]. *Gut*, 2019, 68(9):1545-1575. DOI: 10.1136/gutjnl-2018-318126.
- [3] Bourke MJ, Neuhaus H, Bergman JJ. Endoscopic submucosal dissection: Indications and application in Western endoscopy practice[J]. *Gastroenterology*, 2018, 154(7): 1887-1900. e5.

- DOI: 10.1053/j.gastro.2018.01.068.
- [4] Shiotsuki K, Takizawa K, Ono H. Indications of endoscopic submucosal dissection for undifferentiated early gastric cancer: Current status and future perspectives for further expansion[J]. *Digestion*, 2022, 103(1): 76-82. DOI: 10.1159/000519650.
- [5] Ono H, Yao K, Fujishiro M, et al. Guidelines for endoscopic submucosal dissection and endoscopic mucosal resection for early gastric cancer[J]. *Dig Endosc*, 2016, 28(1): 3-15. DOI: 10.1111/den.12518.
- [6] Yu TF, He W, Gan CG, et al. Deep learning applied to two-dimensional color Doppler flow imaging ultrasound images significantly improves diagnostic performance in the classification of breast masses: a multicenter study[J]. *Chin Med J (Engl)*, 2021, 134(4): 415-424. DOI: 10.1097/CM9.0000000000001329.
- [7] Haggemüller S, Maron RC, Hekler A, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts[J]. *Eur J Cancer*, 2021, 156:202-216. DOI: 10.1016/j.ejca.2021.06.049.
- [8] Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images[J]. *Gastric Cancer*, 2018, 21(4): 653-660. DOI: 10.1007/s10120-018-0793-2.
- [9] Niikura R, Aoki T, Shichijo S, et al. Artificial intelligence versus expert endoscopists for diagnosis of gastric cancer in patients who have undergone upper gastrointestinal endoscopy [J]. *Endoscopy*, 2022, 54(8): 780-784. DOI: 10.1055/a-1660-6500.
- [10] Wu L, Wang J, He X, et al. Deep learning system compared with expert endoscopists in predicting early gastric cancer and its invasion depth and differentiation status (with videos)[J]. *Gastrointest Endosc*, 2022, 95(1):92-104. e3. DOI: 10.1016/j.gie.2021.06.033.
- [11] Nagtegaal ID, Odze RD, Klimstra D, et al. The 2019 WHO classification of tumours of the digestive system[J]. *Histopathology*, 2020, 76(2): 182-188. DOI: 10.1111/his.13975.
- [12] Isomoto H, Shikuwa S, Yamaguchi N, et al. Endoscopic submucosal dissection for early gastric cancer: a large-scale feasibility study[J]. *Gut*, 2009, 58(3):331-336. DOI: 10.1136/gut.2008.165381.
- [13] Lee S, Choi KD, Han M, et al. Long-term outcomes of endoscopic submucosal dissection versus surgery in early gastric cancer meeting expanded indication including undifferentiated-type tumors: a criteria-based analysis[J]. *Gastric Cancer*, 2018, 21(3): 490-499. DOI: 10.1007/s10120-017-0772-z.
- [14] Japanese Gastric Cancer Association. Japanese gastric cancer treatment guidelines 2014 (ver. 4)[J]. *Gastric Cancer*, 2017, 20(1):1-19. DOI: 10.1007/s10120-016-0622-4.
- [15] Goto A, Kubota N, Nishikawa J, et al. Cooperation between artificial intelligence and endoscopists for diagnosing invasion depth of early gastric cancer[J]. *Gastric Cancer*, 2023, 26(1): 116-122. DOI: 10.1007/s10120-022-01330-9.
- [16] de Hond A, Leeuwenberg AM, Hooft L, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review[J]. *NPJ Digit Med*, 2022, 5(1):2. DOI: 10.1038/s41746-021-00549-7.
- [17] Yu G, Sun K, Xu C, et al. Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images[J]. *Nat Commun*, 2021, 12(1): 6311. DOI: 10.1038/s41467-021-26643-8.
- [18] Li J, Hu S, Shi C, et al. A deep learning and natural language processing-based system for automatic identification and surveillance of high-risk patients undergoing upper endoscopy: a multicenter study[J]. *EclinicalMedicine*, 2022, 53:101704. DOI: 10.1016/j.eclinm.2022.101704.
- [19] Nagao S, Tsuji Y, Sakaguchi Y, et al. Highly accurate artificial intelligence systems to predict the invasion depth of gastric cancer: efficacy of conventional white-light imaging, nonmagnifying narrow-band imaging, and indigo-carmin dye contrast imaging[J]. *Gastrointest Endosc*, 2020, 92(4):866-873. e1. DOI: 10.1016/j.gie.2020.06.047.
- [20] Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy[J]. *Gastrointest Endosc*, 2019, 89(4): 806-815. e1. DOI: 10.1016/j.gie.2018.11.011.
- [21] Hamada K, Kawahara Y, Tanimoto T, et al. Application of convolutional neural networks for evaluating the depth of invasion of early gastric cancer based on endoscopic images [J]. *J Gastroenterol Hepatol*, 2022, 37(2): 352-357. DOI: 10.1111/jgh.15725.