

·论著·

# 基于多特征拟合诊断胃褪色调肿瘤性病变的人工智能系统的构建和验证

曾晓铨 董泽华 李艳霞 邓云超 于红刚 陈明锴

武汉大学人民医院消化内科 消化系统疾病湖北省重点实验室, 武汉 430060

通信作者: 陈明锴, Email: kaimingchen@163.com

**【摘要】** **目的** 构建一个基于多特征拟合的人工智能诊断系统,用于白光内镜下诊断胃褪色调肿瘤性病变,并验证其诊断效能。**方法** 收集武汉大学人民医院和中国人民解放军总医院第七医学中心2012年12月至2021年7月的胃镜图像,选取267例患者的823张胃部褪色调病灶图像。通过文献检索选取5个与褪色调肿瘤性病变有关的白光下的镜下特征:病灶位置、病灶边界是否清晰、病灶表面是否粗糙、病灶是否近圆形、病灶是否凹陷,把带有手工标注特征的图像输入机器学习算法中训练,选出最优模型作为多特征拟合诊断系统,系统确定各个病灶特征的诊断权重;用同样的数据集训练测试传统单一深度学习模型。并对两个模型的诊断效能进行比较,并邀请8名不同级别的内镜医师进行人机比较。**结果** 多特征拟合诊断系统诊断胃褪色调肿瘤性病变的准确率、灵敏度、特异度分别是82.11%(101/123)、78.43%(40/51)和84.72%(61/72)。病灶特征按权重由高到低依次是病灶是否凹陷(权重0.71)、病灶位置(权重0.11)、病灶表面是否粗糙(权重0.08)、病灶边界是否清晰(权重0.06)和病灶是否近圆形(权重0.04)。多特征拟合诊断系统的诊断准确率显著高于非专家内镜医师平均诊断准确率(82.11%比74.31%, $Z=-2.785$ , $P=0.008$ ),与专家内镜医师水平相当(82.11%比83.20%, $Z=-0.696$ , $P=0.700$ )。多特征拟合诊断系统与传统深度学习模型诊断准确率差异无统计学意义[82.11%(101/123)比82.93%(102/123), $P=1.000$ ]。**结论** 本研究构建的基于多特征拟合的人工智能诊断系统,在白光下诊断胃褪色调肿瘤性病变具有较高的准确率。

**【关键词】** 机器学习; 人工智能; 内镜; 胃褪色调病变

基金项目:湖北省消化疾病微创诊治临床医学研究中心(2023CCB005)

## Construction and Validation of an artificial intelligence system based on multi-feature integration for diagnosing gastric whitish neoplastic lesions

Zeng Xiaquan, Dong Zehua, Li Yanxia, Deng Yunchao, Yu Honggang, Chen Mingkai

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

Corresponding author: Chen Mingkai, Email: kaimingchen@163.com

**【Abstract】** **Objective** To construct and validate an artificial intelligence diagnostic system based on multi-feature integration for diagnosing gastric whitish neoplastic lesions under white-light endoscopy. **Methods** Gastroscopic images from Renmin Hospital of Wuhan University and The Seventh Medical Center of Chinese PLA General Hospital were collected from December 2012 to July 2021. A total of 823 images of gastric whitish lesions from 267 patients were finally selected. Five white-light endoscopic features associated with gastric whitish lesions were selected through a literature search, including lesion location, boundary clarity, surface texture, roundness, and depression status. Images with manually annotated features were used to train machine learning models, with the optimal model selected as the multi-feature fitting diagnostic system, which assigned diagnostic weights to each feature. A conventional

DOI: 10.3760/cma.j.cn321463-20240805-00209

收稿日期 2024-08-05 本文编辑 许文立

引用本文:曾晓铨,董泽华,李艳霞,等.基于多特征拟合诊断胃褪色调肿瘤性病变的人工智能系统的构建和验证[J].中华消化内镜杂志,XXXX,XX(X):1-6. DOI: 10.3760/cma.j.cn321463-20240805-00209.



deep learning model was trained with the same dataset. The diagnostic performance of the two models were compared, and eight endoscopists of varying expertise were invited to participate in human-machine comparisons. **Results** Accuracy, sensitivity, and specificity of the multi-feature fitting diagnostic system were 82.11% (101/123), 78.43% (40/51), and 84.72% (61/72), respectively. Feature weights in descending order were depression (0.71), lesion location (0.11), surface roughness (0.08), boundary clarity (0.06), and subcircular shape (0.04). The diagnostic accuracy of the system was significantly higher than that of non-expert endoscopists [82.11% (101/123) VS 74.31%,  $Z=-2.785$ ,  $P=0.008$ ] and comparable to that of expert endoscopists [82.11% (101/123) VS 83.20%,  $Z=-0.696$ ,  $P=0.700$ ]. There was no significant difference in accuracy between the multi-feature fitting diagnostic system and the traditional deep learning model [82.11% (101/123) VS 82.93% (102/123),  $P=1.000$ ]. **Conclusion** The feature-weighted artificial intelligence diagnostic system for gastric whitish neoplastic lesions demonstrated clinically relevant diagnostic accuracy under white-light endoscopy.

**【Key words】** Machine learning; Artificial intelligence; Endoscopy; Gastric whitish lesions

**Fund program:** The Project of Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision (2023CCB005)

胃癌是全球第五大常见癌症,第三大常见癌症死亡原因,全球每年诊断胃癌超过 100 万例,造成巨大的医疗负担<sup>[1-2]</sup>。早期胃癌五年生存率达 90% 以上,而进展期胃癌不足 50%<sup>[3-4]</sup>。胃镜广泛用于筛查和诊断胃癌,可以直视黏膜病变并对可疑病变进行组织学活检<sup>[5-8]</sup>,其中白光胃镜最为常用<sup>[9]</sup>。然而,内镜医师严重短缺,不同地区医院的医疗资源分配不均衡<sup>[10-11]</sup>,有数据显示西方国家的上消化道内镜检查中胃癌的漏诊率为 5%~10%<sup>[12-13]</sup>。更糟糕的是,早期未分化型胃癌尤其是早期印戒细胞癌,可在白光下仅表现为平坦或轻微凹陷的褪色病变,即使是经验丰富的内镜医师也容易漏诊<sup>[14-17]</sup>。而且未分化型早期胃癌的内镜治愈率通常低于分化型早期胃癌<sup>[18]</sup>,因此,如何提高胃部褪色调的肿瘤性病变的检出率,降低漏诊率对改善患者预后至关重要。

近年来人工智能技术,包括深度学习和机器学习等算法模型,在医学领域的应用越来越广泛<sup>[19]</sup>。然而深度学习的“黑盒”特性依然存在,影响了其可信度和可接受度,因此增强人工智能系统在医学领域的可解释性十分重要<sup>[20]</sup>。已有研究把医学领域知识纳入人工智能模型的构建用于诊断疾病如 CT 的良恶性结节识别等<sup>[21]</sup>,然而并未有针对胃的褪色调肿瘤性病变相关的研究。因此,本文建立了一个基于机器学习的多特征拟合人工智能诊断系统,纳入了与胃褪色调的肿瘤性病变相关的特征知识,用于诊断白光下胃部褪色调肿瘤性病变,并与基于传统深度学习模型的人工智能系统的性能进行比较。

## 资料与方法

### 一、研究材料

回顾性收集武汉大学人民医院和中国人民解放军总医院第七医学中心 2012 年 11 月至 2021 年 7 月共来自于 267 例患者的 823 张白光下胃褪色调病灶图像,其中 700 张图像(数据集 1:包括来自于 95 例患者的 242 张肿瘤性病灶图像;来自于 124 例患者的 458 张非肿瘤性病灶图像)作为模型的训练集;123 张图像(数据集 2:包括来自 18 例患者的 51 张肿瘤性病灶图像;来自 30 例患者的 72 张非肿瘤性病灶图像)作为模型的测试集。所有图像均来自白光非放大模式,所使用的内镜来自富士和奥林巴斯。图像选取的标准是白光下的褪色调局灶病变,排除具有多个病灶的图像和因为光影泡沫等导致模糊的图像。本文中的肿瘤性病变是指经组织病理学诊断为高级别上皮内瘤变、黏膜内癌、黏膜下癌、进展期癌等病变,而非肿瘤性病变是指经组织病理学诊断为慢性炎、萎缩性胃炎、肠上皮化生等良性病变。

本研究得到了武汉大学人民医院和中国人民解放军总医院第七医学中心伦理委员会的批准(伦理批号:2019K-K094)。由于是回顾性收集的数据,机构审查委员会豁免了知情同意。

### 二、研究方法

#### 1. 构建多特征拟合诊断模型

通过文献检索与白光下胃褪色调病变相关的指南或研究<sup>[6, 15, 17, 22]</sup>,选取 5 个与褪色调肿瘤性病变有关的白光下镜下特征:(1)病灶位置(胃中上部/胃下部/部位无法判断);(2)病灶边界是否清晰

(清晰/不清晰);(3)病灶表面是否粗糙(粗糙/光滑);(4)病灶是否近圆形(近圆形/非近圆形);(5)病灶是否凹陷(凹陷/非凹陷)。诊断相关特征的典型图片如图 1 所示。数据集所有图像 5 个特征由 2 名内镜经验大于 5 年的内镜专家进行标注,若 2 名专家的标注有不一致,则由第三名专家进行最终决定,最终使训练集和测试集的每张图片都同时带有专家标注的 5 个病灶特征标签。然后把带有特征标签的数据集 1 的图片输入决策树(decision tree, DT)机器学习算法中进行训练,得到多特征拟合诊断模型,在测试集中测试拟合诊断模型的测试精度。决策树算法确定了各个病灶特征的诊断权重,权重越大表示与诊断的相关性越大。

## 2. 构建传统单一深度学习模型

为比较多特征拟合诊断模型和传统单一深度学习模型诊断胃褪色调肿瘤性病变的性能,基于 Resnet 50 算法,使用相同的训练集构建单一深度学习模型。构建和验证深度学习模型所使用的训练集和测试集与多特征拟合诊断模型的相同。

## 3. 人机比较试验

使用图片测试集进行人机比较测试。共 8 位内镜医师参加人机比较实验,3 位专家内镜医师(独立胃镜操作经验 $\geq 5$ 年),5 位非专家内镜医师(独立胃镜操作经验 $< 5$ 年)。参与人机比较实验的 8 位内镜医师事先对测试集的病灶情况和病理信息均未知。8 位内镜医师对测试集的每张病灶图像独立做出诊断是否为肿瘤性病灶。对比模型和不同年资内镜医师的诊断水平。

## 4. 数据统计分析

所有分析使用 SPSS 26.0(IBM, USA)进行。分别将多特征拟合诊断模型、单一深度学习模型的测试结果做出混淆矩阵,计算准确率、灵敏度和特异度。评估多特征拟合诊断模型、单一深度学习模型和内镜医师诊断效能。McNemar 检验用于比较模型之间的准确率、敏感度和特异度。Mann-Whitney  $U$  检验用于比较模型和内镜医师之间的表现。 $P < 0.05$  为差异有统计学意义。

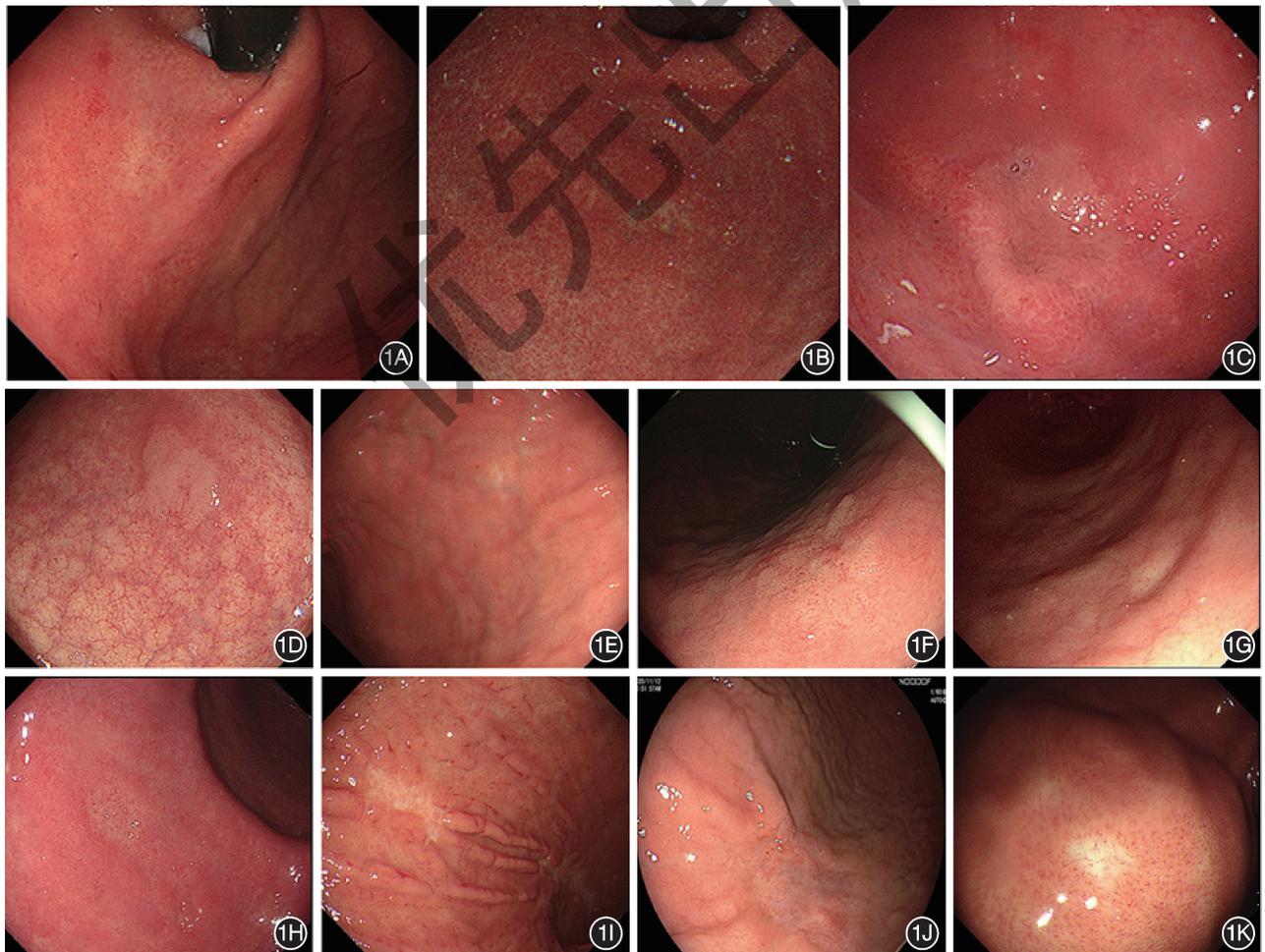
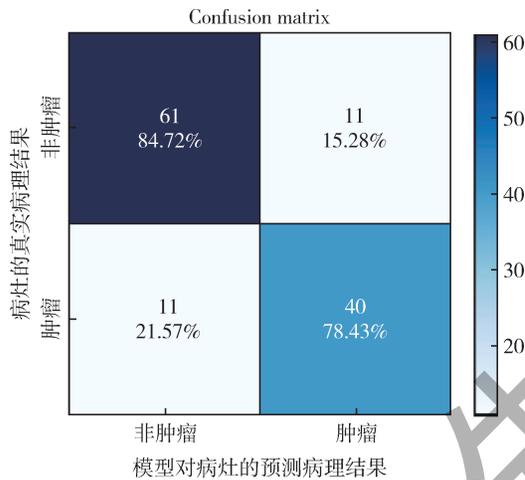


图 1 5 个诊断相关特征的典型图片 1A:病灶的位置位于胃中上部;1B:病灶的位置位于胃下部;1C:病灶的位置无法判断;1D:病灶边界清晰;1E:病灶边界不清晰;1F:病灶表面粗糙;1G:病灶表面光滑;1H:病灶是近圆形;1I:病灶非近圆形;1J:病灶凹陷;1K:病灶非凹陷

### 结 果

#### 1. 多特征拟合诊断模型的诊断效能

基于DT算法的多特征拟合诊断模型在测试集的诊断准确率、灵敏度、特异度分别是 82.11% (95%CI: 74.40%~87.88%)、78.43%(95%CI:64.30%~88.15%)、84.72% (95%CI: 73.88%~91.77%)，多特征拟合诊断模型的混淆矩阵见图 2。病灶特征按权重占比由高到低依次是病灶是否凹陷(权重 0.71)、病灶位置(权重 0.11)、病灶表面是否粗糙(权重 0.08)、病灶边界是否清晰(权重 0.06)和病灶是否近圆形(权重 0.04)。



注：准确率=(40+61)/123×100%=82.11%；灵敏度=40/(40+11)×100%=78.43%；特异度=61/(61+11)×100%=84.72%

图2 多特征拟合诊断模型的混淆矩阵结果

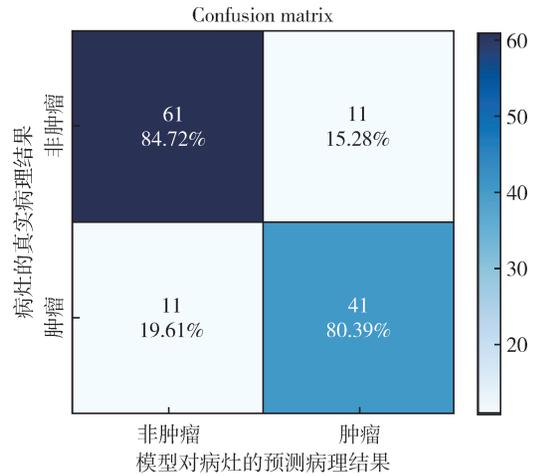
#### 2. 单一深度学习模型的诊断效能

在测试集中,单一深度学习模型的诊断准确率、灵敏度和特异度分别是 82.93% (95%CI: 75.31%~88.56%)、80.39% (95%CI:66.45%~89.71%)、84.72%(95%CI:73.88%~91.77%)，单一深度学习模型的混淆矩阵见图 3。

多特征拟合诊断模型的准确率(82.11%比 82.93%,  $P=1.000$ )、灵敏度(78.43%比 80.39%,  $P=1.000$ )、特异度(84.72%比 84.72%,  $P=1.000$ )和传统单一深度学习模型差异均没有统计学意义。

#### 3. 模型与内镜医师的诊断精度比较

专家内镜医师的平均诊断准确率、灵敏度和特异度分别为 83.20%、85.62% 和 81.48%；非专家内镜医师的平均诊断准确率、灵敏度和特异度分别为 74.31%、75.69% 和 73.33%。多特征拟合诊断模型的诊断性能与专家内镜医师相当(准确率 82.11%



注：准确率=(41+61)/123×100%=82.93%；灵敏度=41/(41+10)×100%=80.39%；特异度=61/(61+11)×100%=84.72%

图3 单一深度学习诊断模型的混淆矩阵结果

比 83.20%,  $Z=-0.696, P=0.700$ ；灵敏度 78.43% 比 85.62%,  $Z=0.696, P=0.700$ ；特异度 84.72% 比 81.48%,  $Z=-0.696, P=0.700$ ；多特征拟合诊断模型的准确率显著高于非专家内镜医师(82.11%比 74.31%,  $Z=-2.785, P=0.008$ )，灵敏度和特异度均稍优于非专家内镜医师(灵敏度 78.43%比 75.69%,  $Z=-0.559, P=0.690$ ；特异度 84.72%比 73.33%,  $Z=-1.677, P=0.151$ )，但差异无统计学意义。

### 讨 论

本研究构建了一个基于多特征拟合的人工智能诊断系统,在白光下诊断胃褪色调肿瘤性病变,表现出较高精度和可解释性的结果,与传统单一深度学习模型表现相当。在人机比较中,多特征拟合诊断模型表现出与专家相当的诊断效能,并显著高于非专家年资医师的诊断水平。

胃癌是我国常见的癌症之一,早期胃癌和晚期胃癌的生存率差异极大<sup>[3-4]</sup>,因此胃癌的早期诊断和早期治疗对于改善患者预后和提高生存质量至关重要。食管胃十二指肠镜是广泛用于筛查、诊断和治疗胃癌的工具,其中白光胃镜是最常用、最简便的工具<sup>[9]</sup>。虽然经胃镜检查后胃癌的检出率有明显改善,但仍不可避免出现漏诊、误诊情况,有研究表明,胃镜检查后的胃癌漏诊率可达 25.8%<sup>[23]</sup>,可能导致严重后果。因此如何通过胃镜检查提高胃癌检出率十分重要。

人工智能近些年发展十分迅猛,尤其在医学领域的应用越来越广泛,已有研究将领域知识纳入人

工智能模型构建用于乳腺癌的诊断、良恶性肺结节的鉴别诊断等<sup>[21, 24]</sup>,然而并未有针对胃褪色调肿瘤性病变相关研究,因此本研究建立一个基于机器学习的多特征拟合人工智能系统用于胃褪色调肿瘤性病变的诊断,并纳入与胃褪色调肿瘤性病变相关的特征知识,通过查阅文献,选取与褪色调肿瘤性病变相关的内镜下特征:病灶位置、病灶表面是否粗糙、病灶边界是否清晰、病灶是否近圆形、病灶是否凹陷。再对图像的病灶特征进行人工标注,标注好的图像输入机器学习算法进行拟合训练,得到最优拟合诊断模型。将领域特征知识输入机器学习模型中进行训练得到多特征拟合诊断模型,这种方法类似于医师学习过程,通过学习和归纳总结典型病例特征来提高疾病的诊断水平。本研究初步验证了通过结合领域知识和使用多特征拟合的方法,可以使用相对有限的数量来训练模型,达到与传统单一深度学习模型相当的精度,为仅有有限数据集的疾病的模型构建提供一种新的思路。

本研究褪色调病灶的白光镜下特征由内镜专家达成一致后人工标注,保证了训练集和测试集各个病灶特征的准确性和标准一致性,内镜医师遇到新的案例,只需要在人工智能辅助诊断系统中输入病灶的几个特征,就可以利用模型输出诊断该褪色调病变是否为肿瘤性病变。但在面对临床中大量的图像数据时仍需要花费一定的人力和时间,因此未来可考虑构建一系列特征提取模型,用于自动提取褪色调病灶图像的镜下特征,降低人力和时间消耗,可在一定程度上提高流程的智能化程度和自动化程度,但模型自动提取的特征可能与测试集标签中的病灶特征出现偏差,从而导致不合理的诊断结果。流程智能化自动化在提升效率和便捷性方面确实具有较大优势,但同时也可能带来一些潜在风险,比如数据特征判断和诊断准确性方面。在未来的研究中是否有必要进一步研究自动化? 这些问题值得我们深思,并在未来的研究实践中不断探索。

本研究有以下局限性:(1)本研究仅回顾性地收集2家医院的图像数据用于构建模型,今后需要纳入更多医院、前瞻性数据来提高和验证模型的精度。(2)本研究只对模型进行了图片测试,未在视频层面进行测试,这是因为褪色调局灶病变在临床中比较少见,其中的肿瘤性病变更少见,今后我们会纳入更多不同地区的医院针对性地收集视频进行视频测试,以增加模型的临床实际使用性。

总之,本研究构建了一个基于多特征拟合的人工智能诊断系统在白光下诊断胃褪色调肿瘤性病变,具有较好的准确率,未来有望用于临床辅助内镜医师提高对胃褪色调肿瘤性病变的检出率。

**利益冲突** 所有作者声明不存在利益冲突

**作者贡献声明** 陈明锴、于红刚监督实验和提供工作支持;曾晓铨:设计实验、实施实验采集和分析数据、文章撰写;董泽华、李艳霞、邓云超:采集数据;董泽华:分析数据

## 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] Thrift AP, El-Serag HB. Burden of gastric cancer[J]. *Clin Gastroenterol Hepatol*, 2020, 18(3):534-542. DOI: 10.1016/j.cgh.2019.07.045.
- [3] Ajani JA, D'Amico TA, Bentrem DJ, et al. Gastric cancer, Version 2.2022, NCCN clinical practice guidelines in oncology [J]. *J Natl Compr Canc Netw*, 2022, 20(2): 167-192. DOI: 10.6004/jnccn.2022.0008.
- [4] Sexton RE, Al Hallak MN, Diab M, et al. Gastric cancer: a comprehensive review of current and future treatment strategies[J]. *Cancer Metastasis Rev*, 2020, 39(4):1179-1203. DOI: 10.1007/s10555-020-09925-3.
- [5] 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.
- [6] Yao K, Uedo N, Kamada T, et al. Guidelines for endoscopic diagnosis of early gastric cancer[J]. *Dig Endosc*, 2020, 32(5): 663-698. DOI: 10.1111/den.13684.
- [7] Wang FH, Zhang XT, Li YF, et al. The Chinese Society of Clinical Oncology (CSCO): clinical guidelines for the diagnosis and treatment of gastric cancer, 2021[J]. *Cancer Commun (Lond)*, 2021, 41(8): 747-795. DOI: 10.1002/cac2.12193.
- [8] Lordick F, Carneiro F, Cascinu S, et al. Gastric cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up[J]. *Ann Oncol*, 2022, 33(10): 1005-1020. DOI: 10.1016/j.annonc.2022.07.004.
- [9] 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.
- [10] Haruma K. What is the appropriate interval between endoscopies for the early detection of gastric cancer? [J]. *Endoscopy*, 2022, 54(9):859-860. DOI: 10.1055/a-1808-6382.
- [11] 王天骞, 王贵齐. 中国胃癌内镜诊治现状[J]. *中国实用内科杂志*, 2018, 38(4):271-274. DOI: 10.19538/j.nk2018040102.
- [12] Pimenta-Melo AR, Monteiro-Soares M, Libânio D, et al. Missing rate for gastric cancer during upper gastrointestinal endoscopy: a systematic review and meta-analysis[J]. *Eur J Gastroenterol Hepatol*, 2016, 28(9):1041-1049. DOI: 10.1097/MEG.0000000000000657.
- [13] Hernanz N, Rodríguez de Santiago E, Marcos Prieto HM, et al. Characteristics and consequences of missed gastric cancer: a multicentric cohort study[J]. *Dig Liver Dis*, 2019, 51(6):

- 894-900. DOI: 10.1016/j.dld.2019.02.006.
- [14] Phalanusitthepha C, Grimes KL, Ikeda H, et al. Endoscopic features of early-stage signet-ring-cell carcinoma of the stomach[J]. *World J Gastrointest Endosc*, 2015, 7(7):741-746. DOI: 10.4253/wjge.v7.i7.741.
- [15] Watari J, Tomita T, Ikehara H, et al. Diagnosis of small intramucosal signet ring cell carcinoma of the stomach by non-magnifying narrow-band imaging: a pilot study[J]. *World J Gastrointest Endosc*, 2015, 7(12):1070-1077. DOI: 10.4253/wjge.v7.i12.1070.
- [16] Law TT, Tong D, Wong SW, et al. Helicobacter pylori-negative gastric mucosa-associated lymphoid tissue lymphoma: magnifying endoscopy findings[J]. *Hong Kong Med J*, 2015, 21(2):183-186. DOI: 10.12809/hkmj134208.
- [17] Kanesaka T, Sekikawa A, Tsumura T, et al. Absent microsurface pattern is characteristic of early gastric cancer of undifferentiated type: magnifying endoscopy with narrow-band imaging[J]. *Gastrointest Endosc*, 2014, 80(6): 1194-1198. e1. DOI: 10.1016/j.gie.2014.08.021.
- [18] Inuyama M, Horiuchi Y, Yamamoto N, et al. Usefulness of magnifying endoscopy with narrow-band imaging for diagnosing mixed poorly differentiated gastric cancers[J]. *Digestion*, 2021, 102(6):938-945. DOI: 10.1159/000517970.
- [19] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence[J]. *Nat Med*, 2019, 25(1): 44-56. DOI: 10.1038/s41591-018-0300-7.
- [20] Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine[J]. *Wiley Interdiscip Rev Data Min Knowl Discov*, 2019, 9(4): e1312. DOI: 10.1002/widm.1312.
- [21] Xie Y, Xia Y, Zhang J, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT[J]. *IEEE Trans Med Imaging*, 2019, 38(4): 991-1004. DOI: 10.1109/TMI.2018.2876510.
- [22] Lee HH, Lee SY, Yoon HY, et al. Change of mucosal color in early gastric cancer[J]. *J Dig Dis*, 2012, 13(10):510-516. DOI: 10.1111/j.1751-2980.2012.00620.x.
- [23] Hosokawa O, Hattori M, Douden K, et al. Difference in accuracy between gastroscopy and colonoscopy for detection of cancer[J]. *Hepatogastroenterology*, 2007, 54(74):442-444.
- [24] Hsu SM, Kuo WH, Kuo FC, et al. Breast tumor classification using different features of quantitative ultrasound parametric images[J]. *Int J Comput Assist Radiol Surg*, 2019, 14(4): 623-633. DOI: 10.1007/s11548-018-01908-8.
- [25] Li YF, Guo LZ, Zhou ZH. Towards safe weakly supervised learning[J]. *IEEE Trans Pattern Anal Mach Intell*, 2021, 43(1): 334-346. DOI: 10.1109/TPAMI.2019.2922396.