

# 基于卷积神经网络预测早期胃癌浸润深度的多模态AI系统开发与验证

靳春昊, 安静怡, 任琳琳, 闵丛丛, 毛涛\*

青岛大学附属医院消化内科, 山东 青岛

收稿日期: 2026年3月1日; 录用日期: 2026年3月24日; 发布日期: 2026年4月2日

## 摘要

目的: 开发并验证基于卷积神经网络的多模态AI系统, 用于预测早期胃癌(EGC)浸润深度, 辅助临床制定内镜下黏膜下剥离术(ESD)治疗决策。方法: 回顾性收集2018年10月至2025年4月青岛大学附属医院经病理确诊的902例EGC患者的11,662张内镜图像, 包括白光成像(WLI)、窄带成像(NBI)和靛胭脂染色图像。基于EfficientNet-B0架构构建CNN模型, 通过内部验证集评估不同成像模式的诊断性能, 并设计人机对比实验(图像层面和病例层面)比较AI系统与不同年资内镜医师的诊断效能。结果: 内部验证集中, WLI模型表现最优, 准确度87.6%、灵敏度89.1%、特异度85.2%、AUC 0.8836; WLI与NBI综合模型准确度85.3%, AUC 0.8634。人机对比实验中, AI系统在图像层面准确度86.65%、特异度87.47%, 全面优于内镜医师; 病例层面采用多数投票法整合信息后, AI准确度提升至92%, 灵敏度93.75%, 优势进一步扩大。结论: 本研究成功开发的多模态AI系统能精准区分EGC黏膜内与黏膜下浸润, 诊断效能显著优于不同年资内镜医师, 病例层面整合多图信息后表现更优, 具备辅助临床决策、促进诊断同质化的潜力。

## 关键词

人工智能, 卷积神经网络, 癌症的早期监测, 肿瘤分期, 胃癌

# Multimodal AI System for Predicting Early Gastric Cancer Invasion Depth Based on Convolutional Neural Networks: Development and Validation Study

Chunhao Jin, Jingyi An, Linlin Ren, Congcong Min, Tao Mao\*

Department of Gastroenterology, The Affiliated Hospital of Qingdao University, Qingdao Shandong

\*通讯作者。

文章引用: 靳春昊, 安静怡, 任琳琳, 闵丛丛, 毛涛. 基于卷积神经网络预测早期胃癌浸润深度的多模态 AI 系统开发与验证[J]. 临床医学进展, 2026, 16(4): 1110-1122. DOI: 10.12677/acm.2026.1641344

## Abstract

**Background:** Develop and test a multimodal AI system based on convolutional neural networks to predict the invasion depth of early gastric cancer (EGC) and assist clinical formulation of endoscopic submucosal stripping (ESD) treatment decisions. **Methods:** This study retrospectively collected a total of 11,662 endoscopic images of 902 patients diagnosed with early gastric cancer and adenocarcinoma by pathological diagnosis from October 2018 to April 2025 in the Affiliated Hospital of Qingdao University, including white light imaging (WLI), narrowband imaging (NBI) and indigo-carmin dye contrast images. The convolutional neural network (CNN) based on the EfficientNet-B0 architecture is used for model training and testing, and the model performance is evaluated through internal image verification and external human-computer comparison experiments (including image-based and case-based evaluation). **Results:** In the internal test set, the WLI model has the best performance, with accuracy of 87.6%, sensitivity of 89.1%, specificity of 85.2%, and AUC 0.8836; the accuracy of WLI + NBI models is 85.3%, and the AUC is 0.8634. In the human-computer comparison experiment, the AI system has an accuracy of 86.65% and a specificity of 87.47% at the image level, which is better than that of endoscopy doctors in terms of all aspects. After integrating information by majority voting at the case level, the accuracy of AI has increased to 92% and the sensitivity is 93.75%, and the advantage has been further expanded. **Conclusions:** Our study developed an AI system capable of predicting the invasion depth of EGC from multimodal endoscopic images, and the diagnostic performance superior to that of endoscopists. The model shows strong standardization and generalization capabilities, which has the potential to assist clinical decision-making and promote diagnosis homogenization.

## Keywords

Artificial Intelligence, Convolutional Neural Networks, Early Detection of Cancer, Neoplasm Staging, Gastric Cancer

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

胃癌是中国发病第三的恶性肿瘤，也是肿瘤相关死亡的第四大原因[1]。早期胃癌(Early gastric cancer, EGC)是指局限于粘膜层和粘膜下层的胃癌，不管有无淋巴结转移。不同阶段的胃癌侵袭性和预后有很大差异：早期胃癌恶性程度较低，浸润深度浅范围小，且不会出现远处转移，及时发现，早期治疗，预后好；由于绝大多数早期胃癌缺乏特异性的临床表现及特异生物标志物，这使得绝大多数患者就诊时已处于进展期[2] [3]。如果早期胃癌得以诊断，并通过手术进行根治性切除治疗，五年生存率高达 90% [4]。

内镜下黏膜下剥离术(Endoscopic submucosal dissection, ESD)是治疗早癌的主要手段，优点是低侵入性和创伤性。根据最新版的《日本胃癌治疗指南》，ESD 的适应症包括：1) 分化型黏膜内癌，无溃疡，不论病灶大小；2) 分化型黏膜内癌，有溃疡，直径小于 3 cm；3) 未分化型黏膜内癌，无溃疡，直径小于 2 cm [5]，因此准确评估侵袭深度对于决定 EGC 的适当治疗策略至关重要，尤其是需要判断病变是否已经出现黏膜下层浸润，从而确定患者是否接受 ESD。

目前尚无可靠方法可确定 EGC 的浸润深度。传统的方法包括内镜、计算机断层扫描(Computed tomography, CT)、超声内镜(Endoscopic Ultrasonography, EUS)和正电子发射断层成像(Positron emission tomography, PET)。传统内镜检查评估浸润深度的内镜下表征为: 皱襞增厚或融合、肿瘤直径  $\geq 30$  mm、明显发红、表面不规则、边缘抬高、黏膜下肿瘤样隆起边缘, 以及无延展征象[6]。新兴技术助力胃癌侵袭深度判断, 临床常用超声内镜、窄带成像技术(Narrow Band Imaging, NBI)、放大内镜(Magnifying endoscopy, ME)等。通过胃镜或超声内镜检查确定癌症是否局限于黏膜层或侵入黏膜下层的能力约为 70%~80% [7]-[9], 且内镜诊断异质性均受病变情况、仪器、内镜医生经验水平等影响, 且技术推广面临设备门槛高、操作要求严、诊断主观性强等挑战[10] [11]。因此, 降低技术使用难度、实现不同水平医师同质化诊断, 是提高早期胃癌诊断水平的关键。

人工智能(Artificial intelligence, AI)模型已在多个医学领域成功应用, 彻底改变了诊断、治疗计划制定和预后预测。卷积神经网络(Convolution neural networks, CNN)是一种成熟的深度学习模型, 由多层卷积层和池化层以及图像滤波器组成, 能够自动从图像网格状训练数据中提取特定特征, 实现整体分类[12] [13]。CNN 在图像分析方面的显著优势, 被逐步应用于胃镜检查质量控制及病灶性质判定中, 有望辅助内镜诊断[14]-[17]。例如 Nagano 等人使用不同角度和距离图像训练胃癌的浸润深度的 AI, 并使用常规白光成像(WLI)、非放大窄带成像(NBI)和靛染料对比成像(Indigo)对每个系统进行训练以预测胃癌的侵袭深度, 均显示了高于 94% 的高准确率[17]。但类似的早期研究包含进展期胃癌, 研究依然较少, 本研究希望继续探索 CNN 在早癌浸润深度的应用, 通过扩大学习不同成像条件下训练集, 以评估 CNN 技术应用于 EGC 浸润深度预测的可行性, 帮助医生准确评估肿瘤浸润深度, 并在单次内镜检查中为 EGC 患者提供治疗。

## 2. 研究对象与方法

### 2.1. 研究对象与设计

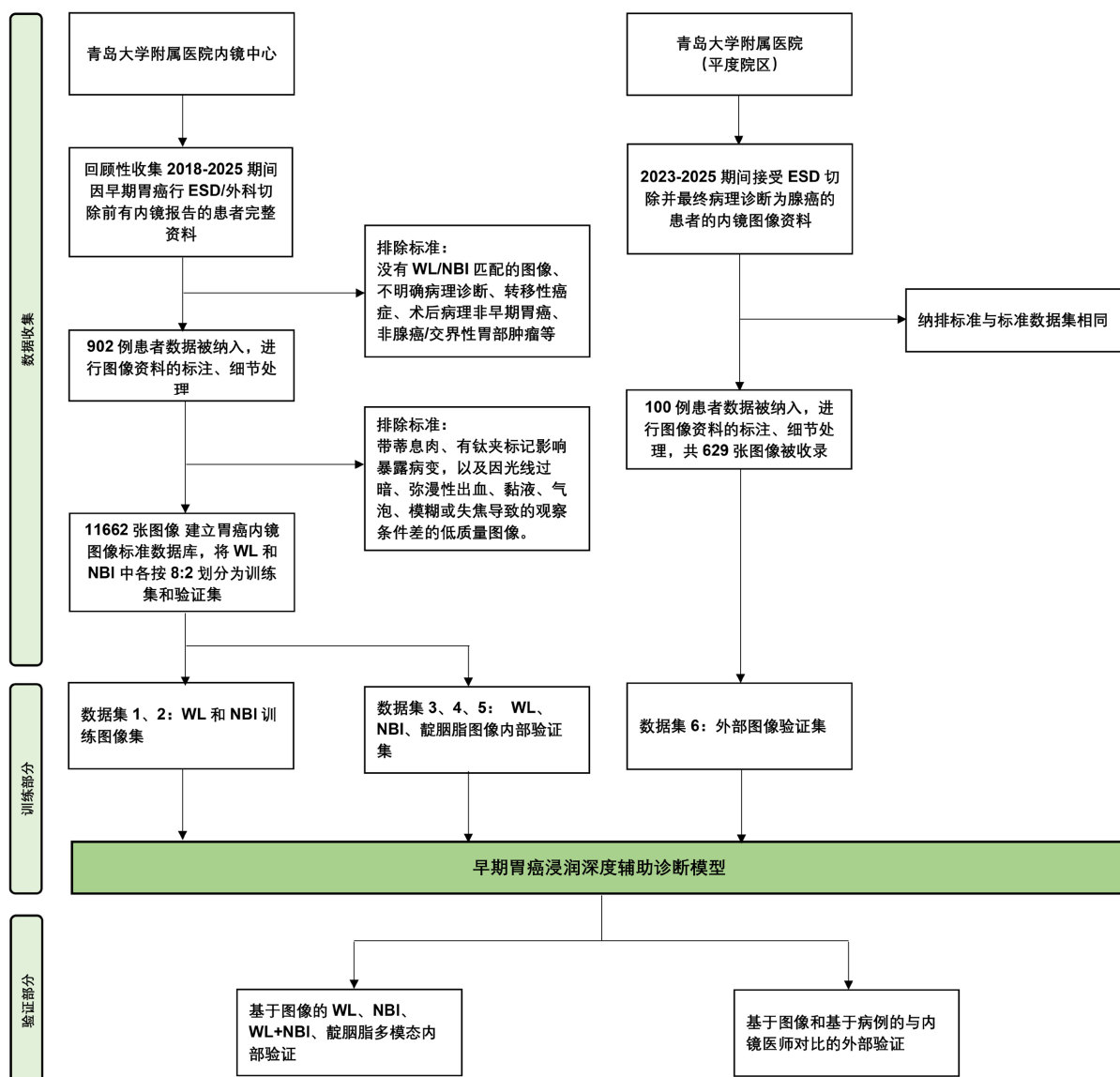
本研究是一项回顾性比较性研究。我们回顾性收集于 2018.10~2025.4 期间青岛大学附属医院以及青岛大学附属医院平度院区内镜中心术前内镜检查后因胃癌而接受 ESD 切除或外科切除, 并最终病理诊断为腺癌的病例。我们首先使用不同成像条件下的不同浸润深度的胃镜图片训练 AI 模型, 然后我们使用内部验证集初步评估了 AI 模型区分黏膜内癌和黏膜下癌图像的诊断效能。最后, 我们使用外部验证集从比较了内镜医师与 AI 模型区分黏膜内癌和黏膜下癌的诊疗水平。完整工作流程图及纳入排除参见图 1。本研究得到了青岛大学附属医院伦理委员会的审查并批准(批准号: QYFY WZLL 30799)。本研究仅使用了回顾性数据, 获取知情同意并对患者私人信息部分进行了脱敏处理。患者的排除标准为: 没有 WL/NBI 匹配的图像、不明确病理诊断、转移性癌症、术后病理非早期胃癌、非腺癌/交界性胃部肿瘤等。

### 2.2. 数据集和数据处理

#### 2.2.1. 数据收集

本研究收集的数据包括早期胃癌患者的临床基本信息、内镜报告和病理报告, 内镜操作系统中提取关键图片, 以建立内镜图像标准数据集。

临床基本信息包括年龄、性别等基线临床病理学特征。内镜检查资料关键特征为病灶大小、病灶部位、表面形态、有无溃疡等。以 ESD 或外科术后明确的病理报告诊断作为分类的“金标准”, 每个病变的病理评估由两位病理学专家医师(超过 10 年工作经验)交叉进行, 根据世界卫生组织(WHO)消化肿瘤分类第 5 版标准划分病理类型。病理报告的关键收集特征为浸润深度及分化类型, 同时还有描述大体肉眼描述、脉管淋巴浸润、切缘情况、幽门螺旋杆菌感染、免疫组化结果等。以上信息均从病历中获取。



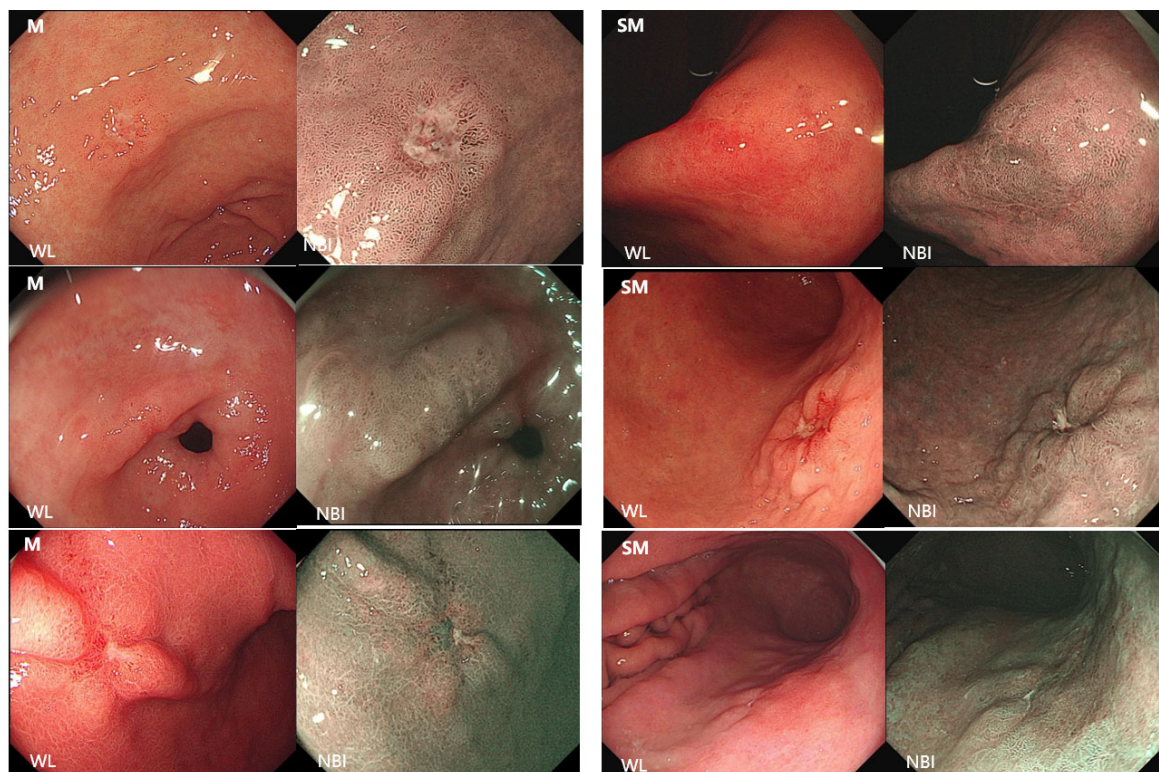
**Figure 1.** Complete workflow diagram and inclusion and exclusion criteria  
**图 1.** 完整的工作流程图以及纳入排除示意图

数据集所包含的图像类型包括白光成像(WLI)、窄带成像(NBI)、靛胭脂染色对比成像，包括同一患者多角度及多距离，具有代表性的图像如图 2 所示，收集图片格式为 JPG，分辨率  $768 \times 576$  以及  $1920 \times 1080$ ，图像中所有与患者有关的数据进行了脱敏处理，从而进一步保证数据的安全性。在收集完所有图像后，我们根据以下排除标准对图像进行了筛选：排除带有钛夹标记影响暴露病变，以及因光线过暗、弥漫性出血、黏液、气泡、模糊或失焦导致的观察条件差的低质量图像。

### 2.2.2. 数据标注与质量控制

所有图像根据内镜医师再次审核的病理记录被分为两组进行标注工作，即黏膜内组(M)和黏膜下侵犯组(SM)。在标注过程中，主要注意到以下两点：在标注质量控制方面，注重和真实病理结果的对照，每一张图由两名专家内镜医生(5000 例以上内镜操作经验，100 例以上 ESD 手术操作经验)交叉验证完成；在标注方法方面，利用多边形曲线勾画出疾病的边界，同时对涉及到建模的信息，如病变性质、大小，

浸润深度，表面情况等信息，存储在相应的可扩展表示语言文件中，供建模的计算机调用。



**Figure 2.** Representative images of EGC cases of submucosal cancer and intramucosal cancer in the endoscopic standard database, each including WL, NBI and other imaging modalities

**图 2.** 内镜标准数据库中粘膜下癌和粘膜内癌 EGC 病例的代表性图像，每个包括 WL、NBI 和其他成像模式

算法工程师将标注好的图像进行压缩、增强、滤波、去噪、图像形态学操作、重采样及均一化等处理，消除干扰，更好地展示图像特征。图像的压缩主要是为了适配模型的输入，在有限的资源下提高模型的处理速度，要求是在尽可能小的图像输入中取得尽可能高的处理精度。图像的增强主要是为了尽可能的增加图像中有用的特征的权重。图像的滤波即在尽量保留图像细节特征的对目标图像的噪声进行抑制。图像的去噪主要是为了进一步提高图像的质量。每种预处理都有若干种成型的算法，在实际应用中，我们往往会预先尝试每一种算法，最终根据实验结果选择最合适的方案。

### 2.2.3. 训练集和验证集分组信息

以最终完善收集的图像建立初始的内镜标准数据库，按照成像方式分为白光、NBI 及靛胭脂数据集。在非靛胭脂图像中，分别将白光和 NBI 中各比例为 8:2 划分为两个数据集，即训练集和内部验证集。内部测试方面将进行基于图像的验证。由于靛胭脂图像数量较少，并未将其划入训练模型中，而全部视作额外验证集进行白光模型验证。分组通过使用患者 ID 作为样本描述符的随机抽样方法完成。同一患者的所有相同类别图像均被分配至训练集或验证集。

### 2.2.4. 人工智能与内镜医师对比验证

为了进一步验证人工智能模型性能，我们进行了一项人工智能与内镜医师的识图实验以测试模型性能。一个小的外部测试数据集通过收集 2023 年至 2025 年在青岛大学附属医院平度院区接受 ESD 切除并最终病理诊断为腺癌的图像而构建，基本的患者的纳排标准及图片筛选同总的内镜数据库。此实验将同

时进行基于图像和基于病例的验证，基于病例的验证将采用多数投票原则，每一病例中多数占比的图像判定为最终病例结果。内镜医师由 2 位专家医师、2 位资深内镜医师和 2 位高级内镜医师构成，他们均未参与训练和验证集的标注。专家内镜医师被定义为在 ESD 治疗及 EGC 术前评估方面具有至少 10 年经验者，而资深内镜医师则指至少有 5 年经验者。高级内镜医师则指在胃镜 EGC 检查方面至少有 5 年经验者。

### 2.3. 人工智能模型算法

根据提供的图像及标注，算法工程师利用人工智能技术建立模型。针对区分病灶的性质这类问题，我们将问题定义成一个区分 M/SM 的二分类任务，然后选择有分类功能的网络模型实现建模。分类模型建模基本上采用同样的流程，首先是利用主干网络针对图像提取特征；然后将图像的特征输入到一个分类器中，利用函数对网络提取到的特征进行优化提取。

我们使用的是开源的 EfficientNet 算法进行建模。EfficientNet 的核心创新在于通过复合缩放 (Compound Scaling) 方法同时优化网络的深度、宽度和输入图像的分辨率。传统的深度网络通常只通过单一的维度(如深度、宽度或分辨率)进行缩放，而 EfficientNet 则结合了这三者，使得模型的计算效率和精度都得到了优化。EfficientNetB0 版本仅有 5.29M 的参数量，在保持高推理速度的情况下还能保持较高精度。算法主要是通过卷积神经网络逐步提取图像的特征，在这过程中图像的尺寸会逐步缩小，共 4 阶段，每阶段缩小 1/2。特征图越大则表示的细节特征如纹理会比较多，特征图越小所代表的语义信息会比较多。模型提取完特征后，会用全连接层将最小的特征图映射到维度为 2 的张量，再用 Softmax 激活函数将该张量里面的数值映射为 M/SM 的概率，取大的作为模型对该图像判断的类别。模型验证时的图像特征可视化识别界面如下图 3 所示。

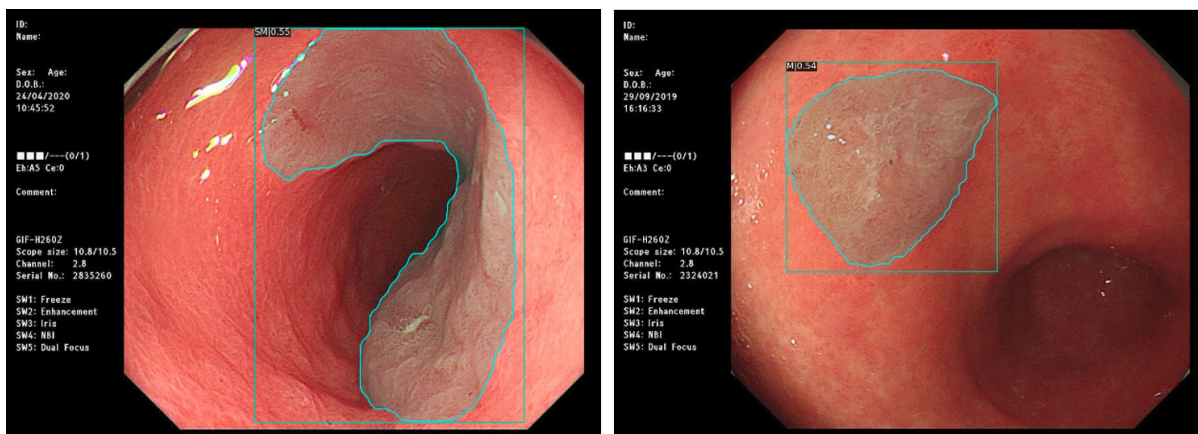


Figure 3. Visual identification interface of image features during model verification

图 3. 模型验证期间图像特征的可视化识别界面

### 2.4. 统计学方法

本研究的主要目的是测试 AI 系统在预测胃癌侵袭深度方面的有效性与准确度。主要结果评价指标为准确率，敏感性，特异性，阳性预测值 (PPV)、阴性预测值 (NPV)，通过绘制 ROC 曲线并计算 AUC 评估模型在内部验证集中的诊断效能。出于分析目的，将人工智能系统在准确诊断黏膜内 (M) 的 EGC 设定为主要目标，并诊断图像为黏膜内 (M) 定义为 1 即阳性，诊断图像为黏膜下 (SM 及更深) 为 0 即阴性，以术后病理报告作为金标准，将 AI 预测图像和病理标注均为黏膜内癌定义为真阳性，将 AI 预测图像和病理

标注均为黏膜下浸润癌定义为真阴性，以此类推构建了一个混淆矩阵。基线数据方面，由于本研究并非探索早癌与各种性状的关系，我们仅将患者内镜检查相关的关键基线特征作了 M 与 SM 组的差异性分析。最终的患者基线临床病理学特征，包括年龄、性别、组织学类型、病理肿瘤大小、肿瘤位置(胃的上、中、下三分之一)及侵袭深度分类，均从病历中获取。

分类变量以频数和百分比表示并进行 Pearson 卡方检验或 Fisher 精确检验，连续变量使用均值 ± 标准差表示并进行两独立样本 t 检验。P 值小于 0.05 被认为具有统计学显著性。使用统计软件对数据进行统计分析。数据整理与统计分析基于 SPSS 统计软件(版本 27.0; 美国 IBM 公司)和 R 软件(版本 4.3.3, R 基金会)完成。

### 3. 结果

#### 3.1. 标准内镜数据库的建立及组成

本研究最终收集到共 902 个病例，其中黏膜内组 605 个病例，黏膜下组 297 个病例，分析了来自 902 名患者的 11,662 张图像。其中按照胃癌浸润深度分为：黏膜内病变的图像占 59.0% (n = 6881)，黏膜下浸润性病变的图像占 41.0% (n = 4781)。按照成像方式分为，白光图像占 45.0% (n = 5242)，NBI 图像占 45.6% (n = 5322)，靛胭脂 9.4% (n = 1098)。患者的基线数据表如表 1 所示。表 2 描述了内镜数据库中内部训练集和验证集的分布组成。表 1 进行的差异性分析显示，相比之下，SM 组肿瘤直径显著大于 M 组(P < 0.001)。在组织学类型上，SM 期中低分化腺癌及未分化型腺癌的比例明显升高(均 P < 0.05)。大体形态方面，混合型(2a + 2c)和凹陷型(2c)有更高的黏膜下浸润风险。总之，SM 期患者主要表现为更大的肿瘤负荷、更凹陷的形态及更低分化程度，而其他基线临床病理特征如年龄、性别及肿瘤部位方面两组未呈现显著差异。

**Table 1.** Baseline clinicopathologic characteristics of patients in the training and test datasets

**表 1.** 患者在分组数据集中的基线临床病理特征表

变量	合计(n = 902)	黏膜内组(n = 605)	黏膜下组(n = 297)	统计量	P 值
年龄, 平均值 ± 标准差	62.45 ± 9.25	62.14 ± 9.54	63.08 ± 8.65	t = -1.434	0.152
肿瘤大小, 平均值 ± 标准差(cm)	2.14 ± 1.19	2.03 ± 1.09	2.36 ± 1.38	t = -3.904	<0.001
性别, 数量(%)				χ <sup>2</sup> = 0.155	0.694
男性	676 (74.94)	451 (74.55)	225 (75.76)		
女性	226 (25.06)	154 (25.45)	72 (24.24)		
位置, 数量(%)				χ <sup>2</sup> = 5.204	0.074
胃上三分之一	145 (16.08)	108 (17.85)	37 (12.46)		
胃中三分之一	235 (26.05)	160 (26.45)	75 (25.25)		
胃下三分之一	522 (57.87)	337 (55.70)	185 (62.29)		
大体分类, 数量(%)				-	<0.001
Type Is	6 (0.67)	5 (0.83)	1 (0.34)		
Type IIa	231 (25.61)	168 (27.77)	63 (21.21)		
Type IIb	161 (17.85)	121 (20.00)	40 (13.47)		

续表

Type IIc	265 (29.38)	173 (28.60)	92 (30.98)		
Type IIa + IIc	236 (26.16)	138 (22.81)	98 (33.00)		
Type III	3 (0.33)	0	3 (1.01)		
WHO 病理分类, 数量(%)				-	<0.001
高分化腺癌	148 (16.41)	108 (17.85)	40 (13.47)		
中分化腺癌	389 (43.13)	258 (42.64)	131 (44.11)		
低分化腺癌	173 (19.18)	94 (15.54)	79 (26.60)		
印戒细胞癌	84 (9.31)	61 (10.08)	23 (7.74)		
胃底腺型腺癌	49 (5.43)	29 (4.79)	20 (6.73)		
乳头状腺癌	38 (4.21)	37 (6.12)	1 (0.34)		
粘液腺癌	21 (2.33)	18 (2.98)	3 (1.01)		
日本病理分类, 数量(%)				$\chi^2 = 4.264$	0.039
分化型	624 (69.18)	432 (71.40)	192 (64.65)		
未分化型	278 (30.82)	173 (28.60)	105 (35.35)		

$\chi^2$ : 卡方检验, -: Fisher 精确检验, t: t 检验。

**Table 2.** Distributed composition of internal training set and verification set in endoscopy database

**表 2.** 内镜数据库中内部训练集和验证集的分布组成

	训练集	验证集
WL 图像集, 数量		
黏膜内组	2574	644
黏膜下组	1619	405
合计	4193	1049
NBI 图像集, 数量		
黏膜内组	2309	578
黏膜下组	1948	487
合计	4257	1065
靛胭脂对比染色图像集, 数量		
黏膜内组	-	776
黏膜下组	-	322
合计	-	1098

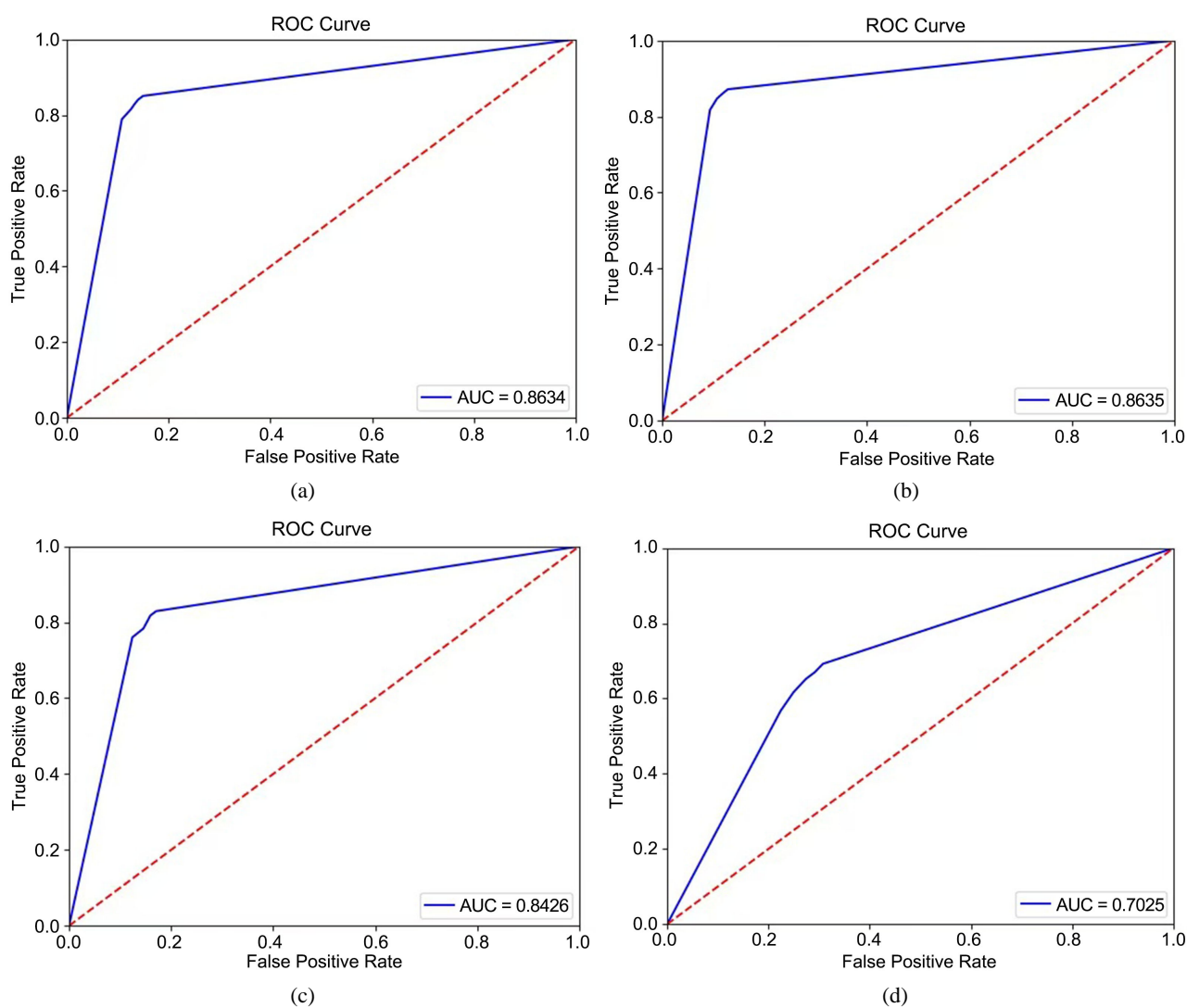
### 3.2. 模型诊断性能：基于图像的内部验证

将上述验证数据集图片识别结果进行筛检评估，以 WL+NBI 综合、WL 模型、NBI 模型、靛胭脂多模态各自验证，具体数据如表 3 所示。白光成像模型表现最优，准确率达 87.6%，敏感度 89.1%，特异度 85.2%，WL+NBI 融合图像次之，准确率 85.3%；NBI 模型准确率 83.0%；靛胭脂染色图像(Indigo-carmin dye contrast)作为单独验证集表现相对较弱，准确率仅 71.8%。绘制了测试图像概率评分的受试者工作曲

线, 如下图 4 所示。WL+NBI 综合、WL 模型、NBI 模型的 AI 系统的曲线下面积分别为 0.8634、0.8836、0.8426。

**Table 3.** Diagnostic accuracy of AI system in multi-model  
**表 3.** 多模态验证中人工智能系统的诊断准确性

	WL + NBI 综合图像集	WL 图像集	NBI 图像集	靛胭脂对比染色图像集
准确度	0.853	0.876	0.830	0.718
敏感度	0.867	0.891	0.839	0.754
特异度	0.834	0.852	0.819	0.630
阳性预测值	0.877	0.905	0.846	0.831
阴性预测值	0.820	0.831	0.811	0.515



**Figure 4.** Receiver operating characteristics curve for the artificial intelligence system. (a) WL + NBI integrated image, (b) White-light image, (c) Narrow-band image, (d) Indigo-carmin dye contrast image

**图 4.** 人工智能系统的受试者工作曲线。(a) WL + NBI 综合图像, (b) WL 图像, (c) NBI 图像, (d) 靛蓝 - 胭脂红染料对比图像

### 3.3. 人机对比试验：基于图像及病例的外部验证

人机测试的外部测试集包括 50 位 M 患者及 50 位 SM 患者，共 100 位患者，共 629 张图像，其中黏膜内病变的图像占 42.9% (n = 270)，黏膜下浸润性病变的图像占 57.1% (n = 359)。在基于图像的人机竞赛中，AI 系统在所有指标上全面领先，尤其是准确性达 0.8665 和特异性达 0.8747，具体数据如表 4 所示。内镜医师中，专家的敏感性 0.7963 相对突出，但特异性、精确性等指标表现较差。在基于图像的人机竞赛中多数指标都有所提升，AI 系统依然全面优于所有医师组，准确率达 0.92，灵敏度 0.9375，医师在病例层面的最高准确率 0.66、敏感度 0.8 也有所提高，具体数据如表 5 所示。

**Table 4.** Image-based diagnostic accuracy of AI system versus endoscopists

**表 4.** 基于图像验证中人机对比试验的诊断准确性

	内镜医师			
	AI 系统	专家内镜医师	资深内镜医师	高级内镜医师
准确度	0.8665	0.5040	0.6343	0.5167
敏感度	0.8556	0.7963	0.7185	0.6778
特异度	0.8747	0.2841	0.5710	0.3955
阳性预测值	0.8370	0.4555	0.5575	0.4575
阴性预测值	0.8895	0.6497	0.7295	0.6201

**Table 5.** Lesion-based diagnostic accuracy of AI system versus endoscopists

**表 5.** 基于病例验证中人机对比试验的诊断准确性

	内镜医师			
	AI 系统	专家内镜医师	资深内镜医师	高级内镜医师
准确度	0.92	0.57	0.66	0.54
敏感度	0.9375	0.80	0.74	0.68
特异度	0.9038	0.34	0.58	0.40
阳性预测值	0.90	0.5479	0.6379	0.5312
阴性预测值	0.94	0.6296	0.6905	0.5556

## 4. 讨论

内镜检查是诊断早期胃癌(early gastric cancer, EGC)的首选方法，而内镜下黏膜下剥离术(endoscopic submucosal dissection, ESD)是其关键治疗手段，旨在以微创方式实现早期胃癌的根治性切除。目前 ESD 的绝对适应证仅包含黏膜内癌。严格遵循适应证对于避免过度治疗或治疗不足至关重要。然而，在实际临床工作中，因浸润深度误判导致治疗方式选择不当的情况仍然存在，因此提高浸润深度的准确分期对于治疗决策具有重要意义。

在本研究中，我们构建并验证了一种多模态人工智能(AI)模型，该模型能够基于白光与窄带成像(NBI)图像准确区分 EGC 的黏膜内(M)与黏膜下(SM)浸润。内部测试集中，白光成像模型在各项诊断性能指标上表现最优，这可能是因为白光成像提供了最完整的黏膜表面信息，包括颜色、纹理、血管分布等多维度特征；AI 模型可能更好地学习了白光图像中肿瘤浸润相关的综合特征，如黏膜皱襞的僵硬、肿瘤边界的不规则性等。这也与白光成像在识别浸润深度应用的现状相契合，白光内镜是最常用的传统检查方

式, 训练数据图像标准化程度高、质量更稳定, 不同设备间差异较小。窄带成像(NBI)的优势在于可增强黏膜表面的微血管和细微结构显示, 与白光成像结合后理论上能为 AI 系统提供更丰富的结构信息。但本研究中其性能略低于白光成像, 可能的原因是联合成像后图像信息的复杂性增加, AI 模型在特征融合时存在一定的信息冗余或特征干扰。且 NBI 图像对操作者技术依赖性强、图像质量波动较大。靛胭脂染色作为仅验证数据集, 未予充分训练, 还与染料分布浓度可能不均, 掩盖某些 AI 识别的重要特征有关。后续可考虑优化多模态图像的融合策略, 如采用注意力机制突出关键特征区域, 以进一步提升联合成像的诊断性能。

为了模拟实际临床场景, 我们设计了基于图像与基于病例的双层人机对比实验。临床真实场景中, 内镜医师需结合多幅内镜图像综合判断 EGC 浸润深度, 单幅图像的诊断易受如拍摄角度、胃内气量、局部黏膜状态等干扰, 我们效仿既往研究中的“多数投票法”, 通过“多图加权”, 能更全面反映病灶的整体特征。结果表明, 在病例层面的“多数投票法”分析中, AI 系统及内镜医师的总体诊断指标较单幅图像层面均有提升。AI 模型在各项指标上均显著优于不同级别的内镜医师, 体现了 AI 在减少主观偏差、标准化判读及识别细微特征方面的优势。值得注意的是, 内镜医师普遍表现出特异度偏低的趋势, 尤其是倾向于将病变判断为可行 ESD 切除的黏膜内癌, 这可能反映了临床上为避免错失微创治疗机会的风险规避倾向。尽管专家医师的敏感度较高, 但部分有经验医师的特异度反而优于专家, 这提示可能存在过度自信偏倚, 即过度依赖直觉判断而忽视系统化特征分析。未来可针对“病例级”诊断的需求, 强化多图像融合、临床信息整合的算法模块, 提升系统对复杂病例的处理能力。

既往研究已充分证实 AI 在评估 EGC 浸润深度方面的应用潜力, 并为本研究奠定了坚实基础。Yoon、Hamada、Goto、Kim 及 Chen 等人从模型架构优化、训练数据策略、人机协作方式及动态视频分析等多方面推动了该领域的发展。Yoon 等人基于 VGG-16 的模型在 EGC 深度预测中取得 79% 敏感性、78% 特异性及 0.851 的 AUC, 并指出肿瘤分化程度显著影响 AI 对 T 分期的准确性[16]。Hamada 等人使用黏膜癌与黏膜下癌手术病例作为训练集, 使 AI 系统在图像及病灶层面的准确率分别达到 78.9% 与 83.8% [18]。Goto 等人提出 AI 与内镜医师多数投票的协作诊断策略, 使准确率提升至 78.0% [19]。Kim 等人首次采用视频分类器预测 EGC 浸润深度, AUC 达到 0.865 [20]。Chen 等人基于 YOLO V4 实现 T1a/T1b 分类, 平均准确率为 86.18% [21]。

在此基础上, 本研究通过更系统的多模态设计和更优异的模型表现进一步拓展了该领域的临床应用价值。本研究首次系统比较四种常用内镜成像模式在 AI 辅助浸润深度诊断中的差异, 弥补了既往研究中多模态对比不足的问题。在模型构建策略上, 唐德华等人采用了 ResNet50 并结合 Grad-CAM 进行可视化解释, 强调了模型的可解释性, 其内部验证集准确率为 88.2%, AUC 为 0.942, 与我们模型的诊断效能相当[22]。传统的深度网络通常只通过单一的维度(如深度、宽度或分辨率其一)进行缩放, 而本研究基于的 EfficientNet-B0 架构, 其核心理念为通过复合缩放将以上三维度结合, 在缩放时保持它们之间的比例关系, 兼顾了模型轻量化与高精度。ResNet50 通过残差连接让网络变得更深, 避免了传统深度网络中的梯度消失问题, 所以在训练非常深的网络时更为有效。而 EfficientNet-B0 采用了深度可分离卷积和复合缩放策略, 目的是在尽可能小的参数量下达到较高的效率和性能。它非常适合在资源受限的设备如移动设备上使用, 进一步为基层内镜医疗工作打下基础。

结果显示, AI 系统在白光、窄带成像、整合图像及靛胭脂染色等多种模式中均保持稳定而卓越的性能, 核心指标(敏感性、准确率及 AUC)均达到或超过既往多数研究。尤其是特异性提升至 90%, 提示本模型在降低黏膜下癌过度治疗风险方面具有更高临床价值。同时, 两院区服务人群与运营体系相对独立, 通过平度分院作为外部验证, 本研究进一步确认了模型的区域性泛化能力。本研究亦首次在山东鲁东半岛区域建立标准化胃镜 AI 数据库, 一定程度上反映了区域 EGC 的内镜学特征, 并降低区域医疗负担。

本研究存在几个局限性。首先，测试集中的患者人群并不能代表筛查人群中癌症患者的自然分布。尽管我们探讨人工智能系统对 EGC 浸润深度诊断的可行性的一种实验研究，但仍需进一步开展如临床随机研究以评估该系统在临床实践中的实际应用价值。第二，尽管本医疗机构有一定辐射范围，但训练集各来源于单一医疗机构，可能存在选择偏倚，外部测试集容量较小，并未进行较大规模多中心验证。未来有望通过已建立算法的模型进行前瞻性、多中心试验，测试模型在其他地区的适应性与有效性。最后，尽管我们进行了病例层面的分析，但目前的模型尚未整合动态视频和内镜超声等更多模态的信息，而这些信息在临床实际判断中至关重要。

## 5. 结论

综上所述，我们成功开发了一个用不同成像方式判断胃癌浸润深度的多模态 AI 系统，并通过与内镜医生的对比研究证实了其卓越的诊断效能。针对 WL + NBI 融合效果不佳的问题，未来尝试引入注意力机制(Attention Mechanism)或双分支网络(Dual-branch Network)，自主学习并提炼 WLI 中的宏观纹理与 NBI 中的微血管特征，以真正实现  $1 + 1 > 1$  的多模态增益，进一步提升模型在真实临床场景中的应用能力，更好地辅助 ESD 决策，实现精准与个体化的胃癌治疗。

## 参考文献

- [1] Siegel, R.L., Miller, K.D. and Jemal, A. (2020) Cancer Statistics, 2020. *CA: A Cancer Journal for Clinicians*, **70**, 7-30. <https://doi.org/10.3322/caac.21590>
- [2] Hamashima, C. (2018) Update Version of the Japanese Guidelines for Gastric Cancer Screening. *Japanese Journal of Clinical Oncology*, **48**, 673-683. <https://doi.org/10.1093/jjco/hvy077>
- [3] Spataro, J., Zfass, A.M., Schubert, M. and Shah, T. (2019) Early Esophageal Cancer: A Gastroenterologist's Disease. *Digestive Diseases and Sciences*, **64**, 3048-3058. <https://doi.org/10.1007/s10620-019-05817-x>
- [4] Sano, T., Coit, D.G., Kim, H.H., Roviello, F., Kassab, P., Wittekind, C., et al. (2016) Proposal of a New Stage Grouping of Gastric Cancer for TNM Classification: International Gastric Cancer Association Staging Project. *Gastric Cancer*, **20**, 217-225. <https://doi.org/10.1007/s10120-016-0601-9>
- [5] Japanese Gastric Cancer Association (2026) Japanese Gastric Cancer Treatment Guidelines 2025 (7th Edition). *Gastric Cancer*, **29**, 271-299. <https://doi.org/10.1007/s10120-025-01698-4>
- [6] Yao, K., Uedo, N., Kamada, T., Hirasawa, T., Nagahama, T., Yoshinaga, S., et al. (2020) Guidelines for Endoscopic Diagnosis of Early Gastric Cancer. *Digestive Endoscopy*, **32**, 663-698. <https://doi.org/10.1111/den.13684>
- [7] Choi, J., Kim, S., Im, J., Kim, J., Jung, H. and Song, I. (2010) Comparison of Endoscopic Ultrasonography and Conventional Endoscopy for Prediction of Depth of Tumor Invasion in Early Gastric Cancer. *Endoscopy*, **42**, 705-713. <https://doi.org/10.1055/s-0030-1255617>
- [8] Sano, T., Okuyama, Y., Kobori, O., Shimizu, T. and Morioka, Y. (1990) Early Gastric Cancer. *Digestive Diseases and Sciences*, **35**, 1340-1344. <https://doi.org/10.1007/bf01536738>
- [9] Abe, S., Oda, I., Shimazu, T., Kinjo, T., Tada, K., Sakamoto, T., et al. (2011) Depth-predicting Score for Differentiated Early Gastric Cancer. *Gastric Cancer*, **14**, 35-40. <https://doi.org/10.1007/s10120-011-0002-z>
- [10] Hyun, Y.S., Han, D.S., Bae, J.H., Park, H.S. and Eun, C.S. (2013) Interobserver Variability and Accuracy of High-Definition Endoscopic Diagnosis for Gastric Intestinal Metaplasia among Experienced and Inexperienced Endoscopists. *Journal of Korean Medical Science*, **28**, 744-749. <https://doi.org/10.3346/jkms.2013.28.5.744>
- [11] Miwata, T., Quach, D.T., Hiyama, T., Aoki, R., Le, H.M., Tran, P.L.N., et al. (2015) Interobserver and Intraobserver Agreement for Gastric Mucosa Atrophy. *BMC Gastroenterology*, **15**, Article No. 95. <https://doi.org/10.1186/s12876-015-0327-x>
- [12] Milea, D., Najjar, R.P., Jiang, Z., Ting, D., Vasseneix, C., Xu, X., et al. (2020) Artificial Intelligence to Detect Papilledema from Ocular Fundus Photographs. *New England Journal of Medicine*, **382**, 1687-1695. <https://doi.org/10.1056/nejmoa1917130>
- [13] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, **529**, 484-489. <https://doi.org/10.1038/nature16961>
- [14] Li, L., Chen, Y., Shen, Z., Zhang, X., Sang, J., Ding, Y., et al. (2019) Convolutional Neural Network for the Diagnosis

- of Early Gastric Cancer Based on Magnifying Narrow Band Imaging. *Gastric Cancer*, **23**, 126-132. <https://doi.org/10.1007/s10120-019-00992-2>
- [15] Ohmori, M., Ishihara, R., Aoyama, K., Nakagawa, K., Iwagami, H., Matsuura, N., *et al.* (2020) Endoscopic Detection and Differentiation of Esophageal Lesions Using a Deep Neural Network. *Gastrointestinal Endoscopy*, **91**, 301-309.e1. <https://doi.org/10.1016/j.gie.2019.09.034>
- [16] Yoon, H.J., Kim, S., Kim, J., Keum, J., Oh, S., Jo, J., *et al.* (2019) A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. *Journal of Clinical Medicine*, **8**, Article 1310. <https://doi.org/10.3390/jcm8091310>
- [17] Nagao, S., Tsuji, Y., Sakaguchi, Y., Takahashi, Y., Minatsuki, C., Niimi, K., *et al.* (2020) 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-Carmine Dye Contrast Imaging. *Gastrointestinal Endoscopy*, **92**, 866-873.e1. <https://doi.org/10.1016/j.gie.2020.06.047>
- [18] Hamada, K., Kawahara, Y., Tanimoto, T., Ohto, A., Toda, A., Aida, T., *et al.* (2021) Application of Convolutional Neural Networks for Evaluating the Depth of Invasion of Early Gastric Cancer Based on Endoscopic Images. *Journal of Gastroenterology and Hepatology*, **37**, 352-357. <https://doi.org/10.1111/jgh.15725>
- [19] Goto, A., Kubota, N., Nishikawa, J., Ogawa, R., Hamabe, K., Hashimoto, S., *et al.* (2022) Cooperation between Artificial Intelligence and Endoscopists for Diagnosing Invasion Depth of Early Gastric Cancer. *Gastric Cancer*, **26**, 116-122. <https://doi.org/10.1007/s10120-022-01330-9>
- [20] Kim, J., Oh, S., Han, S., Keum, J., Kim, K., Chun, J., *et al.* (2022) An Optimal Artificial Intelligence System for Real-Time Endoscopic Prediction of Invasion Depth in Early Gastric Cancer. *Cancers*, **14**, Article 6000. <https://doi.org/10.3390/cancers14236000>
- [21] Chen, T., Kuo, C., Lee, C., Yeh, T., Lan, J. and Huang, S. (2024) Artificial Intelligence Model for a Distinction between Early-Stage Gastric Cancer Invasive Depth T1a and T1b. *Journal of Cancer*, **15**, 3085-3094. <https://doi.org/10.7150/jca.94772>
- [22] 唐德华. 人工智能在早期胃癌检出及性质判断中的临床研究[D]: [博士学位论文]. 南京: 南京大学, 2021.