中华医学会系列杂志

ISSN 1007-5232
CN 32-1463/R

中华消化内镜杂志®

ZHONGHUA XIAOHUA NEIJING ZAZHI

2025年2月 第42卷 第2期

CHINESE JOURNAL OF DIGESTIVE ENDOSCOPY

Volume 42 Number 2
February 2025

ISSN 1007-5232
02
9 771007 523250

中华消化内镜杂志®

CHINESE JOURNAL OF DIGESTIVE ENDOSCOPY

月刊　1996年8月改刊　第42卷　第2期　2025年2月20日出版


微信:xhnjxw


新浪微博

主　管
中国科学技术协会

主　办
中华医学会
100710,北京市东四西大街42号

编　辑
中华消化内镜杂志编辑委员会
210003,南京市紫竹林3号
电话:(025)83472831,83478997
传真:(025)83472821
Email:xhnj@xhnj.com
http://www.zhxhnjzz.com
http://www.medjournals.cn

总编辑
张澍田

编辑部主任
唐涌进

出　版
《中华医学杂志》社有限责任公司
100710,北京市东四西大街42号
电话(传真):(010)51322059
Email:office@cmaph.org

广告发布登记号
广登32010000093号

印　刷
江苏省地质测绘大队

发　行
范围:公开
国内:南京报刊发行局
国外:中国国际图书贸易集团有限公司
(北京399信箱,100048)
代号 M4676

订　购
全国各地邮政局
邮发代号 28-105

邮　购
中华消化内镜杂志编辑部
210003,南京市紫竹林3号
电话:(025)83472831
Email:xhnj@xhnj.com

定　价
每期25.00元,全年300.00元

中国标准连续出版物号
ISSN 1007-5232
CN 32-1463/R

2025年版权归中华医学会所有

未经授权,不得转载、摘编本刊文章,不得使用本刊的版式设计

除非特别声明,本刊刊出的所有文章不代表中华医学会和本刊编委会的观点

本刊如有印装质量问题,请向本刊编辑部调换

目　　次

述　　评

教练理论:一套临床人工智能系统开发与验证理论体系 …… 85
于红刚　蔺蓉

共识与指南

超声内镜人工智能系统临床应用专家共识(2024,武汉) …… 94
中华医学会消化内镜学分会大数据协作组

专家论坛

《胰腺体外震波碎石术专家共识》要点解读 …… 104
衣津慧　胡良峰

论　　著

基于深度学习的人工智能技术在肠道准备评估中的应用 …… 109
王雯　姚理文　熊慧珍　李秋成　陈泓磊　于红刚

肠道准备过程中不良反应的相关危险因素分析 …… 115
苏巧玉　张英杰　王泽宇　康晓宇　潘阳林

利那洛肽联合复方聚乙二醇电解质散肠道准备改良方案的随机对照研究 …… 120
唐丽欣　郭雨栋　李士学　唐秀芬

早期食管癌合并多原发癌的内镜和病理特征及危险因素分析 …… 125
刘昊玺　张倩　张阳　张昌昊　邢洁　李鹏　张澍田

内镜治疗Ⅰ~Ⅲ度内痔的双中心随机对照研究 …… 131
何安玲　马超　肖勇　朱珂　刘书中　陈明锴

床旁胶囊内镜在急重症消化道出血患者中的应用价值 …… 137
张琪琦　张洁　陈路　林壁涛　陈振煜　王馨珂　万田漠　丰希成　王振疆　李爱民　吴保平　刘思德　罗晓蓓

剪状刀内镜黏膜下剥离术治疗无蒂隆起型结直肠上皮源性肿瘤的有效性及安全性 …… 142
高岩　王叶　蓝宇　贾纯增

短篇论著

双镜联合精准穿刺引导扩张治疗结直肠癌术后吻合口闭锁初探(含视频) …… 148
王东洋　陈明　张昊　李浩　翟会专　孙明娟　王海鹏　李增军

病例报道

附加通道辅助内镜黏膜切除术治疗直肠神经内分泌瘤1例(含视频) …… 151
刘朝晖　吴瑞暖　黄伟英　邢艳杰　孙大勇　李锐
内镜黏膜下剥离术治疗跨越幽门环的早期胃癌1例 …… 154
李劼　郑庆芬　赵丽霞　刘冰熔
内镜下食管肌层切开联合主动脉带膜支架隔绝术治疗食管异物穿孔伴主动脉损伤1例 …… 156
詹磊磊　陈桂权　梁瑞茵

综　　述

超声内镜在血管介入诊疗方面的研究进展 …… 159
万笑雨　刘萍　曹新广　郭长青
粪菌移植的发展趋势、适应证领域、移植途径和政策管理:基于CiteSpace分析 …… 164
赵睿　董晨　鲍艳凤　孙驰　顾志峰

读者·作者·编者

《中华消化内镜杂志》对来稿中统计学处理的有关要求 …… 130
《中华消化内镜杂志》2025年可直接使用英文缩写的常用词汇 …… 147
《中华消化内镜杂志》2025年征订启事 …… 147
中华医学会杂志社关于在论文写作过程中使用生成式人工智能技术的有关规定 …… 163
插页目次 …… 103

本刊稿约见第42卷第1期第82页

本期责任编辑　顾文景　唐涌进

本刊编辑部工作人员联系方式

唐涌进,Email:tang@xhnj.com　　朱　悦,Email:zhuyue@xhnj.com
周　昊,Email:zhou@xhnj.com　　钱　程,Email:qian@xhnj.com
顾文景,Email:gwj@xhnj.com　　许文立,Email:xwl@xhnj.com

本刊投稿方式

登录《中华消化内镜杂志》官方网站http://www.zhxhnjzz.com进行在线投稿。


唐涌进


周　昊


顾文景


朱　悦


钱　程


许文立

(扫码添加编辑企业微信)

·述评·

教练理论：一套临床人工智能系统开发与验证理论体系

于红刚[1]　蔺蓉[2]

[1]武汉大学人民医院消化内科，武汉 430060；[2]华中科技大学同济医学院附属协和医院消化内科，武汉 430022

通信作者：于红刚，Email：yuhonggang1968@163.com



于红刚，武汉大学人民医院消化内科主任，一级主任医师，二级教授，博士生导师，中华医学会消化内镜学分会常务委员，中华医学会消化内镜学分会大数据协作组组长，湖北省医学会消化内镜分会主任委员，湖北省消化内镜介入诊疗质量控制中心主任



蔺蓉，华中科技大学同济医学院附属协和医院消化内科主任，主任医师，教授，博士生导师，中华医学会消化病学分会委员、胃肠激素与神经内分泌肿瘤学组副组长，中国医师协会内镜医师分会常委，中国抗癌协会肿瘤内镜分会常委、神经内分泌肿瘤分会常委，湖北省医师协会消化病分会候任主委

【提要】 近年来，人工智能技术蓬勃发展，在图像识别、特征提取等方面的优势使之有机会被广泛应用于医学影像和医疗文书等领域，作为辅助诊疗活动的重要工具。然而，临床医学科学家往往并非人工智能领域的专家，在选择构建或评估模型时需要与计算机领域的工程师专家沟通合作，不充分的沟通或不恰当的合作模式会导致资源的利用效率降低，无法实现高效率跨学科合作。对于临床医师，人工智能模型像一个语言不通的学徒，临床医师无法将自身的理论与实践经验直接传授给学徒，而需要一位懂得模型语言的翻译官。本文旨在规范临床人工智能系统开发和验证流程，提出一种医工合作的新模式，以指导临床人工智能系统的开发与应用，构建具有较高诊疗质量、较广泛适用场景的人工智能辅助诊疗系统。本文提出了临床人工智能系统开发的8个阶段，其中，临床专家医师作为教练员（coach）的角色，计算机工程师作为翻译官（translator）的角色，人工智能模型作为学徒（apprentice）的角色，以教练-翻译-学徒的模式（coach-translator-apprentice）进行高质量临床人工智能系统的开发与验证。

【关键词】 人工智能；　医工合作；　教练模式；　辅助诊疗

基金项目：湖北省重点研发计划项目（2023BCB153）

DOI：10.3760/cma.j.cn321463-20241115-00524

收稿日期　2024-11-15　　**本文编辑**　朱悦　唐涌进

引用本文：于红刚，蔺蓉. 教练理论：一套临床人工智能系统开发与验证理论体系[J]. 中华消化内镜杂志，2025，42(2)：85-93. DOI：10.3760/cma.j.cn321463-20241115-00524.

www.zhxhnjzz.com

A theoretical system for the development and validation of the clinical artificial intelligence systems

Yu Honggang[1], Lin Rong[2]

[1] Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan 430060, China; [2] Department of Gastroenterology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

Corresponding author: Yu Honggang, Email: yuhonggang1968@163.com

背　景

近年来，人工智能技术蓬勃发展，在肿瘤筛查、疾病诊断、风险预测和辅助治疗等医学领域带来了广泛且深远的技术革新，不仅显著提高了医疗服务水平，同时也改变了传统的医学模式。我国政府出台了一系列政策文件，大力推广人工智能技术在医疗各细分领域的应用[1-2]。这意味着应用人工智能系统已成为我国现代医疗领域发展的必然趋势。然而，临床医师往往并非人工智能领域的专家，在选择构建或评估模型时需要与计算机领域的工程师专家沟通合作，不充分的沟通或不恰当的合作模式会导致资源利用效率下降，低质量的人工智能系统在模型构建和验证上可能存在许多不足，如数据选择与处理偏差、系统准确性与泛化性缺乏临床验证等，这可能导致诊断错误、临床决策误导、伦理道德困境等一系列问题，从而降低人工智能系统的临床效能[3-4]。对于临床医师，人工智能模型像一个语言不通的学徒，临床医师无法将自身的理论与实践经验直接传授给学徒，而需要一位懂得模型语言的翻译官。

目前人工智能相关规范或理论体系大多着眼于系统构建和应用过程中的安全性与伦理规范性[5-7]，对于人工智能系统开发和验证的完整流程尚缺乏完善的理论体系规范，建立规范的人工智能系统开发和验证体系是保证人工智能在医疗健康领域发展与应用的基础和前提。本理论体系提出一种医工合作的新模式，以指导临床人工智能系统的开发与应用，其中临床医师作为教练员（coach）角色，计算机工程师作为翻译官（translator）角色，人工智能模型作为学徒（apprentice）角色。本文将构建高质量临床人工智能系统的过程分为 8 个阶段，并详细阐述了各阶段临床医师、计算机工程师和人工智能模型的角色与作用，促进医学与人工智能的深度融合和跨学科的高效合作（图 1）。

阶段一：发现科学问题

随着人工智能技术的快速发展，构建高质量临床人工智能系统已成为提升医疗服务质量的重要任务。在这一过程中，科学问题的发现是系统构建的基石，直接影响后续的研究设计、数据收集和模型构建，临床医师在该环节占据主导地位。

1. 观察临床现象

在医疗实践中，临床医师常接触到多种临床现象，其中不乏具有研究潜力的现象。以胃肠道肿瘤为例，其晚期病死率居高不下，而早期诊断率偏低，已成为亟待解决的重大公共卫生问题。早期诊断可显著提高患者的生存率：早期患者的五年生存率超过 95%，而晚期的五年生存率不足 20%[8]。然而，5%~10% 的早期胃肠道肿瘤在内镜检查中未能被发现，反映出当前早期诊断的缺陷[9-10]。通过对这些现象的细致观察，研究者可捕捉到潜在的研究方向，为后续问题的提出奠定基础。

2. 提出临床问题

在观察到特定的临床现象后，研究者需将其转化为具体的科学问题。如胃肠道肿瘤的早期诊断率偏低的原因是什么。问题的提出应基于临床经验和相关生物医学知识，确保其科学性和实际意义，从而引导后续研究的方向。

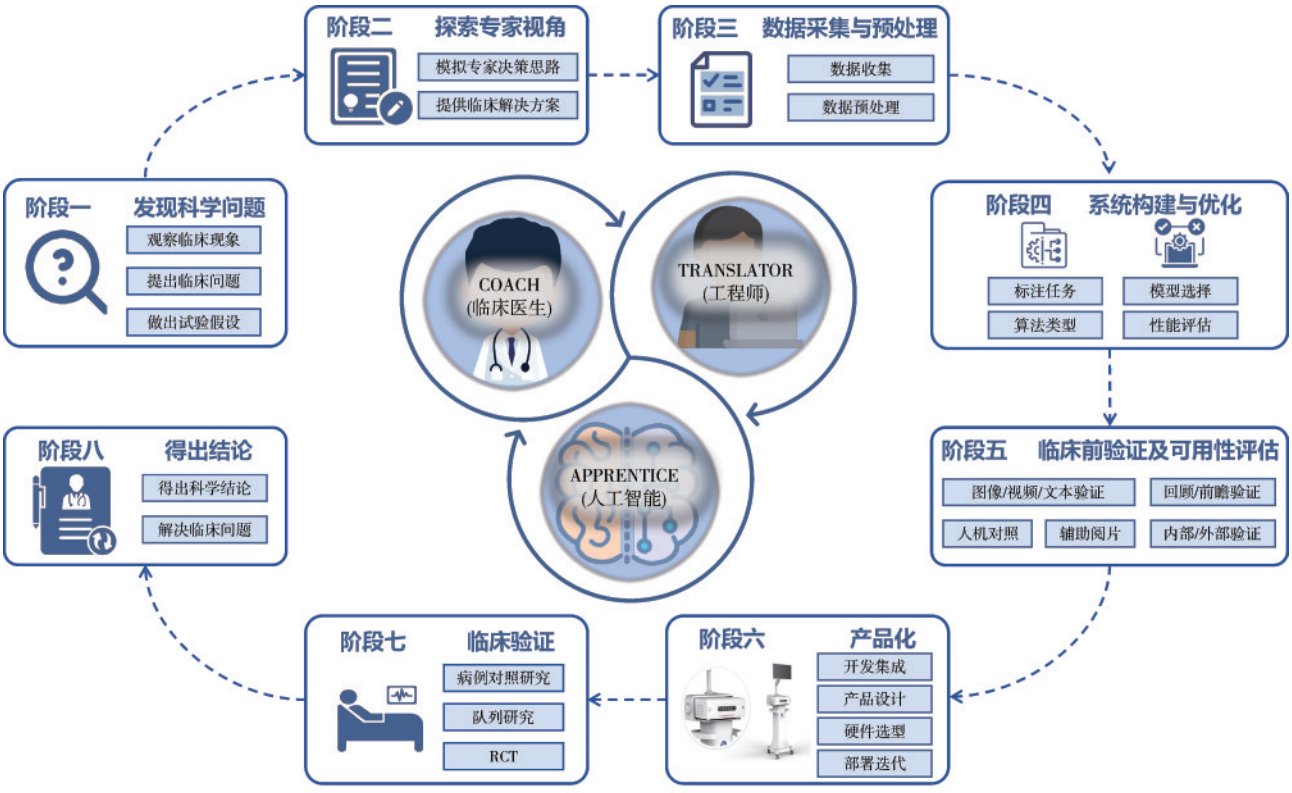
3. 做出试验假设

假设是对研究问题的初步解释，是科学研究中的关键环节。针对胃肠道肿瘤早期诊断率偏低的问题，可假设其原因与消化内镜诊断质量的差异有关。消化内镜在早期胃肠道肿瘤的检出和诊断中发挥着关键作用[11-13]，而由于内镜操作中出现的盲区及医师对疾病认知的不足，可能导致早期病灶的漏诊。研究假设应基于已有的研究成果、理论框架以及临床经验，为后续的实验设计提供方向，也为数据收集、分析和技术路线的选择奠定基础。

阶段二：探索专家视角

为了确保人工智能在临床中的有效性并提高模型的可解释性，计算机工程师需要作为翻译官的角色，将作为教练员的临床专家的先验专业知识（如早癌特征、临床决策思路等）传达给模型，以构建更贴合临床实际需求的系统。本阶段中，临床专家需要与工程师密切沟通，充分呈现解决科学问题

www.zhxhnjzz.com



注：RCT指随机对照试验

图1　临床人工智能系统开发与验证流程图

的临床思路，以便工程师明确目标任务类型，为后续的数据集构建与数据预处理提供导向。

针对阶段一提出的胃肠道肿瘤早期诊断率偏低可能与消化内镜诊断质量差异有关的临床假设，为了减少医师在内镜检查中的漏诊和误诊，可从“看得全”和“认得清”两方面入手，以临床专家的决策思路调整模型设计。

实现检查过程中的实时质量控制是“看得全”的关键。在上消化道内镜检查中，内镜医师根据解剖学特征判断已检查与未检查的部位。因此，可基于相关临床理论体系提出的标准化方案和图谱[14-16]，将上消化道分为多个部位作为质控站点，在内镜检查过程中以拼图的形式标识已检查部位并提示未检查部位，从而减少上消化道内镜检查过程中的盲区[17]。

提高对病灶特征的辨识能力是“认得清”的关键。临床医师主要通过学习教材、理论体系、图谱等文献资料，以及大量的临床实践积累，提升对于疾病的诊断能力，对于诊疗方案的精准选择高度依赖医师的操作经验和临床诊疗能力。基于以上诊断逻辑及临床痛点，工程师将这些临床知识“翻译”给模型，使模型不仅能够识别这些特征，还能对特征的重要性分配合适的权重，从而在辅助诊断时输出明确的诊断依据，提升模型的客观性和可解释性。基于该思路，工程师利用临床专家总结的11个早期胃癌内镜诊断特征，构建了一个兼具准确性、可解释性和较高医师满意度的早期胃癌诊断系统[18]。

阶段三：数据采集与预处理

在数据的收集与预处理阶段，在明确研究方案、确定所需要的数据集后，通常需要经过伦理委员会的批准才能进行下一步的数据采集。在许多情况下，会使用现有数据进行回顾性研究，这类研究通常可以申请知情同意豁免。而在前瞻性研究情况下，知情同意则是必要的。鉴于医疗数据的重要性与保护患者隐私的原则，在工程师接触到医疗数据前应由医师团队进行内部的脱敏处理，如去除与可识别身份特征信息相关的数据以及不影响疾病诊断的敏感个人信息，如电话号码、家庭住址等。工程师可协助提供一些特定的小程序工具，如裁切或遮蔽医学图像中显示的患者隐私信息，以提高数据脱敏效率[19]。

在工程师取得对匿名化医疗数据的访问权限后，可以对数据进行适当的预处理，但应当保留一份

www.zhxhnjzz.com

不被编辑的原始数据。常见的处理包括数据清洗以及缺失值的处理[20]。数据清洗主要关注数据质量问题，包括数据重复、不完整、异常等。临床医师需要根据临床问题制定完整的数据清洗计划，如识别相似、重复、不完整、异常的记录以及逻辑错误，并将需求告知工程师；由工程师确定具体的技术细节，譬如在医师要求的分辨率、帧率下，模型是否能满足实时检测的性能，并验证清理后的数据是否适合训练任务。而缺失值的处理可以采用删除、加权调整、数据插补等方式，也可以根据临床真实场景制定[21]。

此外，在整个研究过程中，数据采集与预处理是反复、多批次的，应充分考虑数据规模、采集效率、成本周期等多个重要制约因素，需根据任务需求、产品技术特点、目标用户人群等进行动态调整和平衡，不能一味地追求数据规模[22]。

阶段四：系统构建与优化

构建与优化系统是整个流程的关键环节。我们提倡采用逻辑和行为拟人化的策略进行系统的设计与构建。临床医师与工程师需要协同根据数据模态、任务导向设计合适的技术路线。一套完整的人工智能系统的构建可能需要多个模型的协同参与。而临床医师需要基于临床问题，参考专家范式，通过逻辑和行为拟人的策略来设计技术路线。例如，对于胃镜下早期胃癌诊断困难的临床问题，由医师参考专家的诊断策略，选择以图像中病灶的形态、色调、粗糙度等特征作为模型特征提取的目标，进而设计多特征拟合诊断系统，从而实现自动化可解释的早期胃癌诊断[23]。如果系统构建过程中没有采用逻辑拟人策略，应当采用如类激活映射等方式对结果进行分析与解释，以提高人工智能系统的可解释性与接受度。

而工程师作为翻译官，需要完成医师设计系统的实施。具体而言，需要工程师根据所构建的数据集选择技术路线中每个步骤所需的合理算法，并将每步的算法进行整理与关联，从而尽可能满足实际临床任务要求，同时合理平衡需求与成本。例如，在医疗文本生成中，大语言模型具有组织自由文本的优势，但考虑到训练大模型的高成本与不稳定性[24]，仅需简单结构文本的任务则不适宜采用大语言模型。另外，工程师也需要与医师协同确定系统的实验室评价指标，如对具体任务的准确率、灵敏度、特异度等，在实验室层面对技术路线的关键步骤进行测试，从而给后续整体验证提供参考。

阶段五：临床前验证与可用性评估

临床前验证是决定一项智能诊疗技术是否能够顺利进入产品化阶段的关键步骤。这个阶段通常需要对模型进行全面测试，以确保其性能指标均能达到预期要求。此外，还需要进行临床前的人机对比验证或辅助能力测试，例如回顾性验证与前瞻性验证、内部验证与外部验证等。若验证结果理想，即可进一步推进到产品化阶段，反之，则需要对存在的不足进行修正与优化。依据临床前验证中系统的应用与对比条件，主要可分为图像与视频测试、人机对照测试、辅助阅片测试3个步骤。

1. 图像与视频测试

训练并验证模型后，可以针对选定的任务，对模型进行图像与视频测试。一般来说，视频测试的要求高于固定图像测试。以内镜检查为例，由于实际临床环境中，医师的进镜、退镜等操作会伴随检查画面的实时变动，因此，实时反馈就显得格外重要。能够在视频测试中获得优异表现的模型，更贴合临床实际工作环境，也更容易被临床医师接受[25]。

2. 人机对照测试

在人工智能相关的临床研究中，研究者们往往不仅会关注模型的表现，更会关注模型与人类医师相比解决特定任务的能力如何，此时需要进行人机对照研究。一般需要构建人机对照数据集，由多名人类医师和人工智能系统分别独立完成对应任务，从而对比人类医师和人工智能系统的能力[26]。一般情况下，若人工智能系统针对特定任务的表现能够显著优于人类医师的平均水平，则认为针对特定任务，该模型有潜力在实际临床环境中发挥较大作用。若人工智能系统针对特定任务的表现能够与专家医师相当，则认为针对该任务，该模型有潜力在实际临床环境中发挥重大作用。

3. 辅助阅片测试

现阶段，人工智能系统大多扮演辅助诊疗的角色，最终临床诊断与治疗决策由临床医师完成。因此，除了关注针对特定任务下人工智能系统自身的表现、人工智能系统与人类医师相比孰优孰劣，研究者们往往会重点关注人类医师在人工智能系统辅助下能否有显著提升。针对特定任务，若人类医师在人工智能系统的辅助下，其能力有显著性提升，或新手医师在人工智能系统的辅助下，其能力与专家医师相当，则认为针对该任务，该模型有潜力在实际临床环境中发挥重大作用[27]。

www.zhxhnjzz.com

阶段六：产品化

医学人工智能模型的产品化是一个多阶段的过程，涵盖了从概念到产品成型的各个层面，包含了产品原型设计、硬件选型、开发和集成、部署与迭代等多个关键阶段。作为国内外首个人工智能消化内镜质控和辅助诊疗系统，内镜精灵产品的研发经历了从理论到产品落地的全整流程，成为教练理论指导人工智能模型产品化的典型实例。

1. 产品原型设计

产品原型设计是教练员发挥主导作用的最重要阶段，即由专家医师发起，以临床需求为导向研发临床人工智能系统。本阶段重点在于将医学人工智能模型的概念转化为一个可应用于临床的人工智能产品雏形，包括核心功能定义、用户界面设计、数据接口，以及如何与现有的医疗工作流程相结合。针对临床消化内镜检查中存在的漏诊率高、早期癌识别困难等临床实际痛点，本团队将临床需求分解为两步，同时构想了人工智能系统应具有的核心功能，首先要实现质量控制，然后做到辅助诊断。进一步地，主导设计了如用虚拟胃模型显示已观察部位、文字提示未监测部位、仪表盘显示退镜速度、红框标记高危病灶等人机交互的用户界面，使产品真正贴合临床，易于使用者操作和理解。

2. 硬件选型

硬件选型阶段涉及到为医学人工智能模型选择合适的硬件平台。这包括评估模型的计算需求，选择能够提供足够处理能力的中央处理器、图形处理器或其他专用硬件。同时，还需要考虑数据存储、内存、网络连接等硬件规格，以及硬件平台的功耗、稳定性、耐用性、可扩展性和成本效益。硬件选型应该支持模型的长期发展和潜在的性能升级。

3. 开发和集成

开发和集成阶段是将医学人工智能模型从原型转化为实际可用产品的过程。这包括编写代码、实现算法、开发用户界面，并将模型集成到人工智能系统中。开发过程中需要遵循软件工程的最佳实践，确保代码的质量和可维护性。集成过程中需要确保产品能够与电子健康记录（EHR）、医疗影像系统等其他医疗系统无缝对接。

4. 部署与迭代

部署与迭代阶段涉及到将医学人工智能模型部署到实际的医疗环境中，并根据用户反馈进行持续的优化和改进。部署过程中需要确保模型的稳定性和安全性，以及在高负载情况下的性能。迭代过程中需要收集用户反馈，对模型进行调整和升级，以满足不断变化的医疗需求和标准。此外，还需要定期更新模型，以纳入最新的医疗研究成果和数据。

从临床出发，遵循产品化关键步骤，医学人工智能模型可以成功地从研究概念转变为实际的医疗产品，为患者提供更好的医疗服务。

阶段七：临床验证

临床验证是评估人工智能系统是否能够有效运作于真实临床环境中的直接检验，可以为得出科学结论提供证据等级较高的数据支撑。临床验证的结果能够直接反映出人工智能系统在真实世界中的适用性与局限性，同时也可以获取临床的真实反馈，进而评估其解决特定临床问题的能力。现阶段对于人工智能系统的临床验证主要集中在两个层面，依据是否干预正常的诊疗过程可分为观察性队列研究和随机对照临床试验。

1. 队列研究

队列研究（cohort study）是一种前瞻性或回顾性的观察性研究设计，常用于评估特定暴露因素与特定结局之间的关联。在这种研究中，研究者首先根据是否暴露于某个特定因素将研究对象分为两个或多个组别，然后随访这些组别，记录并比较不同组别中发生特定结局（如疾病、特定事件或死亡）的频率。人工智能相关前瞻性观察性队列研究中，不干预医师的诊疗流程，待完成诊疗流程并采集相关试验数据后，由人工智能系统进行验证并与人类医师的诊疗能力进行对比。

2. 随机对照研究

随机对照试验（randomized controlled trial，RCT）一般被认为是评价新系统或药物诊疗效果的金标准，RCT 严格控制试验入组与排除标准和其他条件，并进行随机化分组，因此能够最大限度地减少影响因果推断的因素，使得研究结论较为确定，所形成的证据可靠性也较高。单中心 RCT 的证据等级高于临床前试验（如图片与视频测试、人机对抗研究、辅助阅片研究等），前瞻性多中心大样本 RCT 的证据等级最高。人工智能相关 RCT 设计时，一般会针对特定任务，设定两组试验条件，对照组为人类医师独立完成诊疗活动，试验组为人类医师在人工智能系统辅助下完成诊疗活动。若针对特定任务，应用人工智能系统辅助组医师的表现显著优于人类医师独立诊疗组，则可以认为，在符合临床试验的条件下，该人工智能系统能够显著提升医师针对该临床任务的诊疗能力[12,28]。

www.zhxhnjzz.com

阶段八：得出科学结论

通过前述的7个步骤，我们需要总结工作内容，得出科学结论。在科学研究中，阐述实验结果、得出科学结论是至关重要的部分，它直接影响到研究的可信度、科学性和应用价值。基于此，我们提出以下建议。

1. 研究结果要基于数据，避免过度解读或者轻视阴性统计结果。结果的阐述应严格基于实验数据，不夸大或过度解读。例如，如果某一差异达到了统计学意义，但效应量较小或者实际差异并不明显，那么应明确说明它在实际应用中的有限性。使用“可能”“倾向于”等词汇来描述趋势，避免用“证明”“确立”等带有过度确定性的词语。对差异有统计学意义的结果要仔细解释其临床意义。然而，当某些数据差异未达到统计学意义，应避免单纯解读为“无效”或“没有价值”，可以指出“不显著”并探讨其可能的原因。

2. 阴性结果的研究价值。针对阴性结果，首先要考虑是否选择了合适的统计指标和效应量，不能单纯依赖*P*值。同时，要考虑阴性结果的潜在可能依据，进行充足的亚组分析和敏感性分析，让科学结论更严谨，更具备合理性。

3. 科学结论尽量得到其他研究的支持。研究人员应将结果与已有研究文献进行对比，探讨相似或不同之处。若结果与先前的研究存在显著差异，则可尝试解释可能的原因，如实验设计、样本特征、技术差异等。

4. 讨论实验局限性。研究需要明确指出实验的局限性，包括样本量、实验设计、数据来源等。清晰的局限性讨论有助于表明研究者对结果的严谨态度。阐述局限性并不意味着削弱结果的价值，反而可以明确研究结论的适用条件，增加结论的可信度。

5. 避免推测性结论。对结果讨论部分通常建议不进行过多、过度推测。如果确实有合理的推测，则应在“未来展望”或“建议”部分进行说明。

6. 得出科学结论的建议结构。建议按照如下顺序阐述科学结论：总结主要发现、解释结果的意义、讨论限制性因素、提出未来研究方向。

科学结论的阐述也可以参考各行业领域的理论体系，在得出科学结论时，避免主观色彩，基于数据进行实事求是的分析和解释。做到客观、严谨、谨慎地阐述研究结果，既展示了研究的科学价值，也避免了结果的过度解读。

理论实例

1. 胃镜检查中的质量控制模型

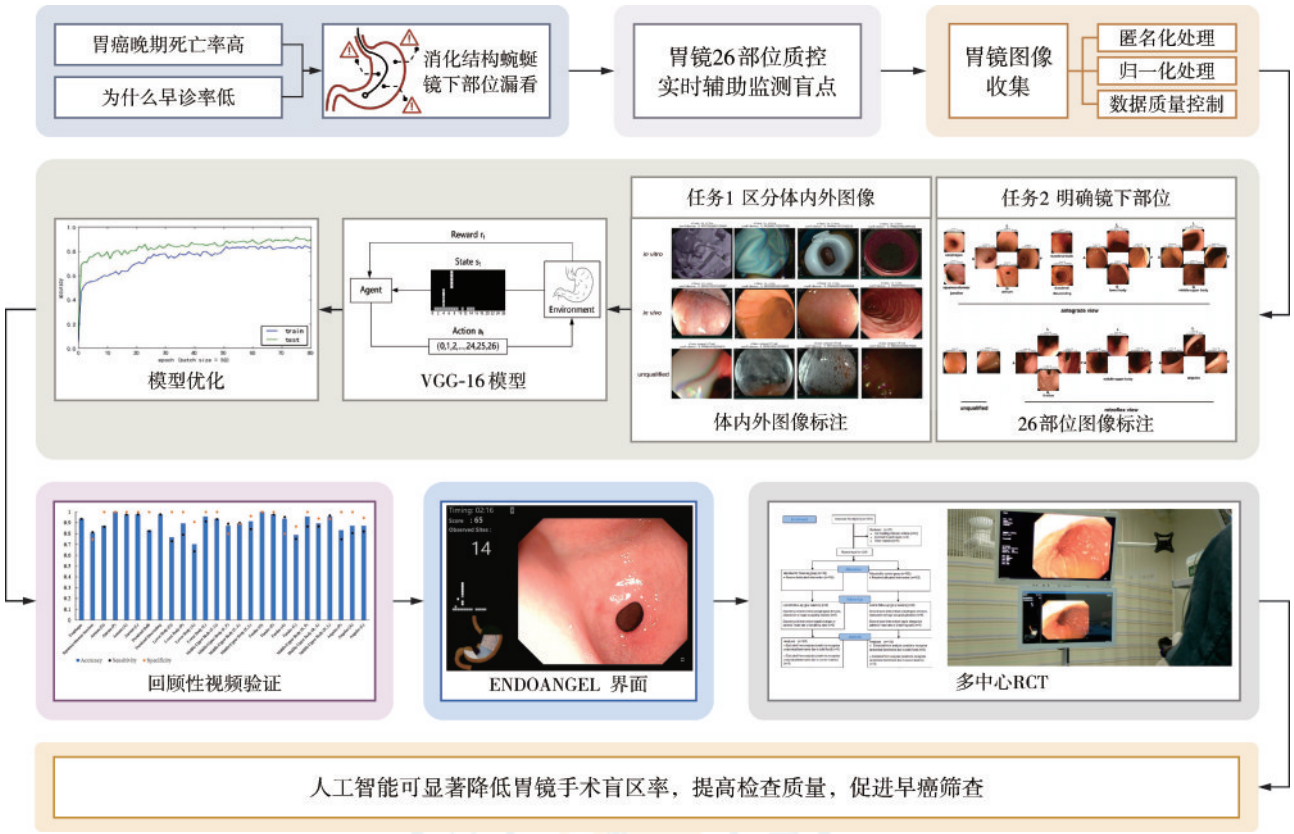
胃镜检查是早期发现上消化道肿瘤的重要手段，全面观察上消化道是早期诊断病变的基本前提。胃系统筛查方案建议胃镜检查应至少包含22个部位，以避免遗漏可疑的癌性病变[15]。然而，内镜医师可能因主观因素或操作水平有限而忽略某些区域，导致早期病变的漏诊（观察现象，明确问题）。因此，需构建一个实时胃镜检查质量改进系统——WISENSE，用于协助内镜医师完善对上消化道各部位的观察，减少胃镜检查盲区，从而降低病变的漏诊率（专家视角）。

WISENSE的训练和测试来自3个数据集：数据集1包括12 220张体外、25 222张体内和16 760张不合格的胃镜图像，由两名博士生完成对图像的标记。基于欧洲胃肠内镜学会的胃镜检查协议和胃系统筛查方案[29]，34 513张合格的图像由2名具有1~5年内镜检查经验的医师和3名专家级内镜医师将图像独立标记为26个不同部位或不可识别部位的图像。此外，使用30个胃镜视频以获取深度强化学习（deep reinforcement learning，DRL）的最佳状态，并使用107个胃镜视频来测试WISENSE在临床环境中的性能（数据准备）。

WISENSE的构建和优化基于深度卷积神经网络（deep convolutional neural network，DCNN）和DRL：DCNN 1模型用于识别图像是来自体内或体外；DCNN 2模型用于分类上消化道的26个部位；DRL模型用于优化内镜检查过程中的决策（系统构建和优化）。

在真实胃镜视频中，WISENSE监测盲区的平均准确率、敏感度和特异度分别为90.02%、87.57%和95.02%（回顾性验证）。此外，本团队前瞻性招募了324例胃镜检查患者，根据是否在WISENSE的协助下接受胃镜检查进行随机分组。结果显示，WISENSE组的盲区率显著低于对照组（5.86%比22.46%，*P*<0.001），同时，WISENSE组生成的照片记录完整性显著高于对照组内镜医师生成的照片记录完整性（90.64%比79.14%，*P*<0.001）（产品化与临床验证）。WISENSE作为基于深度学习的内镜质量改进系统，可在胃镜检查期间实时监测盲区和辅助图像采集。它作为一个强大的辅助工具，可以减轻内镜医师之间的技能差异，并提高日常内镜检查的质量（科学结论）[28]（图2）。

www.zhxhnjzz.com



注：RCT指随机对照试验

图2　教练理论应用实例1：胃镜26部位盲区监测系统的构建与验证　步骤一、临床专家基于晚期胃癌死亡率高的现象，提出临床问题：为什么胃癌的早期诊断率偏低；步骤二、临床专家提出可能的痛点：胃腔结构蜿蜒曲折，部分部位在胃镜检查中易漏诊，并提出解决方案，构建一套质量控制策略，要求胃镜监测需观察到26个胃腔结构；步骤三、工程师依据临床专家的方案，明确目标任务类型，对收集的临床影像数据进行预处理；步骤四、逻辑与行为拟人化的系统构建过程；步骤五、回顾性视频验证；步骤六、内镜精灵ENDOANGEL产品化；步骤七、随机对照临床试验；步骤八、得出科学结论，解决临床问题

2. 超声内镜检查质量控制系统

胆道和胰腺疾病的诊断常面临复杂性和高技术要求的挑战。超声内镜检查（endoscopic ultrasonography，EUS）在诊断胆总管结石、胆管梗阻、胆管癌等方面的灵敏度接近内镜逆行胰胆管造影，但其操作依赖于医师技能和经验[30]（观察现象，明确问题）。胆道评估也要求医师具备较高的专业技能。因此，开发超声内镜实时辅助系统以提高EUS的应用效率和准确性尤为重要。为此，开发了一个深度学习辅助的线性EUS系统——胆胰大师（BP MASTER）[31]，旨在帮助识别胰腺标准站、对胆道进行分割和直径测量并指导医师操作（专家视角）。分别用10 681幅和2 529幅图像建立了分类和胆道分割模型，并进行内外部图像验证、视频验证和人机竞赛以确保其可靠性（数据准备与回顾性验证）。该系统包括4个DCNN模块：DCNN 1用于筛选EUS图像，DCNN 2负责筛选合格图像，DCNN 3用于识别胰腺的标准站点，DCNN 4则执行胆管分割（系统构建）。研究数据显示，DCNN 1、DCNN 2和DCNN 3的图像分类准确率分别为100%、87.4%和93.3%。在外部验证中，该系统站点分类准确率为82.6%，在人机竞赛中与专家的分类准确率接近。通过该系统的培训，医师的准确率从60.8%提高至76.3%，显著缩短了学习时间并提升了操作效率（回顾性验证）。

考虑到胰腺疾病诊断的特殊挑战，本团队进一步开发了EUS-IREAD系统[32]。该系统基于前期关于胰腺EUS分站的工作，将胰腺的识别增多至8个站点和24个解剖结构（专家视角）。其中DCNN 1用于站点识别，DCNN 2、DCNN 3和DCNN 4分别用于胰腺定位、血管定位和胆胰管定位，DCNN 5用于肾脾分类。团队进一步验证了该系统在降低EUS图像解读难度时的能力（系统优化）。在测试中，DCNN 1的分类准确率为95%，DCNN 2、DCNN 3和DCNN 4的Dice系数分别为0.80、0.76和0.68，DCNN 5的准确率为93%。在随机对照研究中，EUS-IREAD辅助的内镜医师在肝门、胆总管、胆胰管等部位的漏检率显著降低（产品化与临床验证）。

www.zhxhnjzz.com

此研究表明，EUS-IREAD 系统提升了 EUS 质量，减少了医师技能差异，是 EUS 的有效辅助工具（科学结论）（图 3）。

总结与讨论

本理论体系提供了一套分步骤的临床人工智能系统开发与验证流程，提出了一种新的医工交叉合作模式，以指导临床人工智能系统的开发与应用。由于临床医师具有丰富的临床实践经验，因此对于亟待解决的临床痛点和科学问题有更深刻的理解和更准确的把握，故而临床医师在整个系统开发过程中需处于主导地位。必须特别强调的是，本文提出的观点无意降低计算机工程师在整个人工智能系统构建中的地位与作用；相反的，由于临床医师大多缺乏足够的计算机领域知识储备，在系统的构建与优化过程中，非常依赖于计算机工程师对于任务类型的判断与模型构建方案的选择。有经验的工程师能够依据临床痛点，准确选择合适的训练方法与应用逻辑，事半功倍地应对复杂系统的开发与优化。

本文提出的合作模式，明确了临床人工智能系统开发的基本步骤，以及各个步骤中临床医师、计算机工程师和人工智能系统的分工与合作方式，形成了以临床医师为主导，计算机工程师密切配合，分步骤分阶段对智能系统进行构建、验证与优化的理论体系，旨在为构建具有较高诊疗质量、较广泛适用场景的人工智能辅助诊疗系统提供流程规范。

利益冲突　所有作者声明不存在利益冲突

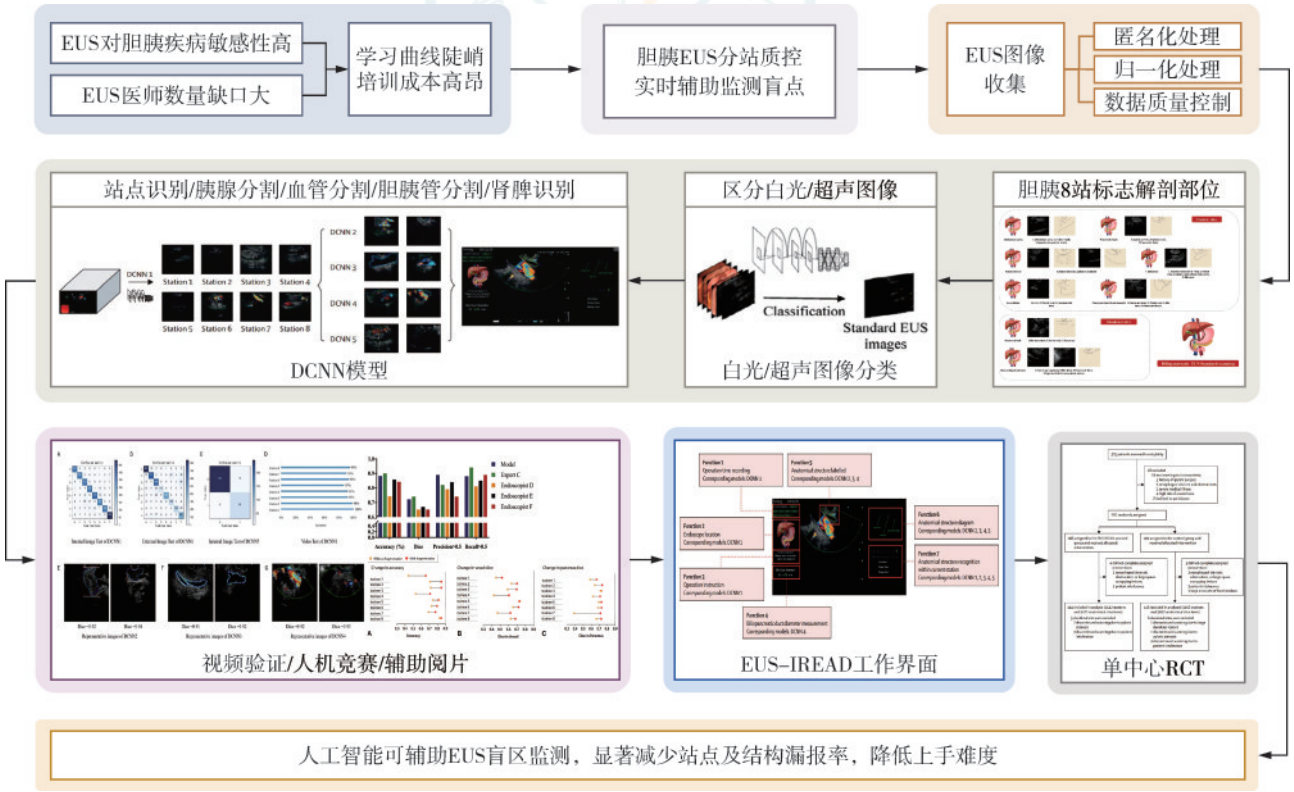
参　考　文　献

[1] 国家标准化管理委员会. 国家新一代人工智能标准体系建设指南[Z].(2020-08-04)[2024-11-12].https://www.sac.gov.cn/cms_files/filemanager/1907776593/attach/20233/P020211018573104094770.pdf.

[2] 中共中央，国务院.“健康中国 2030”规划纲要[Z].(2016-10-25) [2024-11-12]. https://www. gov. cn/zhengce/2016-10/25/content_5124174.htm.

[3] Cai Y, Cai YQ, Tang LY, et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review[J]. BMC Med, 2024,22(1):56. DOI: 10.1186/s12916-024-03273-7.

[4] Pachouly J, Ahirrao S, Kotecha K, et al. A systematic literature review on software defect prediction using artificial intelligence: datasets, data validation methods, approaches,



注：EUS 指超声内镜检查；DCNN 指深度卷积神经网络；RCT 指随机对照试验

图 3　教练理论应用实例 2：超声内镜质量控制系统的构建与验证　步骤一、临床专家基于超声内镜胆胰疾病诊断面临挑战这一现象，提出临床问题：如何提升超声内镜学习效率；步骤二、临床专家针对步骤一中的问题，提出解决方案，构建一套辅助监测系统，实时辅助医师进行超声内镜诊疗；步骤三、工程师依据临床专家的方案，明确目标任务类型，对收集的临床影像数据进行预处理；步骤四、逻辑与行为拟人化的系统构建过程；步骤五、临床前验证；步骤六、胆胰大师 EUS-IREAD 产品化；步骤七、随机对照临床试验；步骤八、得出科学结论，解决临床问题

www.zhxhnjzz.com

and tools[J]. Eng Appl Artif Intell, 2022, 111: 104773. DOI: 10.1016/j.engappai.2022.104773.

[5] World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models [Z]. (2024-01-18) [2024-11-12]. https://www. who. int/publications/i/item/9789240084759.

[6] 国家互联网信息办公室. 生成式人工智能服务管理暂行办法 [Z]. (2023-07-13) [2024-11-12]. https://www. cac. gov. cn/2023-07/13/c_1690898327029107.htm.

[7] 国家药监局. 国家药监局关于发布人工智能医用软件产品分类界定指导原则的通告(2021年第47号)[Z]. (2021-07-08) [2024-11-12]. https://www. nmpa. gov. cn/xxgk/ggtg/ylqxggtg/ylqxqtggtg/20210708111147171.html.

[8] Katai H, Ishikawa T, Akazawa K, et al. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001-2007) [J]. Gastric Cancer, 2018, 21(1): 144-154. DOI: 10.1007/s10120-017-0716-7.

[9] Januszewicz W, Witczak K, Wieszczy P, et al. Prevalence and risk factors of upper gastrointestinal cancers missed during endoscopy: a nationwide registry-based study[J]. Endoscopy, 2022,54(7):653-660. DOI: 10.1055/a-1675-4136.

[10] Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis[J]. Endosc Int Open, 2014,2(2):E46-50. DOI: 10.1055/s-0034-1365524.

[11] Arnold M, Abnet CC, Neale RE, et al. Global burden of 5 major types of gastrointestinal cancer[J]. Gastroenterology, 2020,159(1):335-349.e15. DOI: 10.1053/j.gastro.2020.02.068.

[12] Rutter MD, Senore C, Bisschops R, et al. The European Society of Gastrointestinal Endoscopy Quality Improvement Initiative: developing performance measures[J]. Endoscopy, 2016,48(1):81-89. DOI: 10.1055/s-0035-1569580.

[13] Cheung D, Evison F, Patel P, et al. Factors associated with colorectal cancer occurrence after colonoscopy that did not diagnose colorectal cancer[J]. Gastrointest Endosc, 2016,84(2): 287-295.e1. DOI: 10.1016/j.gie.2016.01.047.

[14] Bretthauer M, Aabakken L, Dekker E, et al. Requirements and standards facilitating quality improvement for reporting systems in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) position statement[J]. Endoscopy, 2016, 48(3): 291-294. DOI: 10.1055/s-0042-100186.

[15] Yao K. The endoscopic diagnosis of early gastric cancer[J]. Ann Gastroenterol, 2013,26(1):11-22.

[16] Torkamani A, Andersen KG, Steinhubl SR, et al. High-definition medicine[J]. Cell, 2017,170(5):828-843. DOI: 10.1016/j.cell.2017.08.007.

[17] Wu L, He X, Liu M, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial[J]. Endoscopy, 2021,53(12):1199-1207. DOI: 10.1055/a-1350-5583.

[18] Li J, Zhu Y, Dong Z, et al. Development and validation of a feature extraction-based logical anthropomorphic diagnostic system for early gastric cancer: a case-control study[J]. EClinicalMedicine, 2022, 46: 101366. DOI: 10.1016/j.eclinm.2022.101366.

[19] Aryanto KY, Oudkerk M, van Ooijen PM. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy[J]. Eur Radiol, 2015, 25(12): 3685-3695. DOI: 10.1007/s00330-015-3794-0.

[20] García S, Ramírez-Gallego S, Luengo J, et al. Big data preprocessing: methods and prospects[J]. Big Data Analytics, 2016,1(1):9. DOI:10.1186/s41044-016-0014-0.

[21] Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning[J]. Nat Med, 2020, 26(3): 364-373. DOI: 10.1038/s41591-020-0789-4.

[22] Wu Z, Yao L, Liu W, et al. Development and validation of a deep learning-based histologic diagnosis system for diagnosing colorectal sessile serrated lesions[J]. Am J Clin Pathol, 2023, 160(4):394-403. DOI: 10.1093/ajcp/aqad058.

[23] Dong Z, Wang J, Li Y, et al. Explainable artificial intelligence incorporated with domain knowledge diagnosing early gastric neoplasms under white light endoscopy[J]. NPJ Digit Med, 2023,6(1):64. DOI: 10.1038/s41746-023-00813-y.

[24] Zhou L, Schellaert W, Martínez-Plumed F, et al. Larger and more instructable language models become less reliable[J]. Nature, 2024, 634(8032): 61-68. DOI: 10.1038/s41586-024-07930-y.

[25] He X, Wu L, Dong Z, et al. Real-time use of artificial intelligence for diagnosing early gastric cancer by magnifying image-enhanced endoscopy: a multicenter diagnostic study (with videos)[J]. Gastrointest Endosc, 2022,95(4):671-678.e4. DOI: 10.1016/j.gie.2021.11.040.

[26] Wu L, Shang R, Sharma P, et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomised controlled trial[J]. Lancet Gastroenterol Hepatol, 2021, 6(9): 700-708. DOI: 10.1016/S2468-1253(21)00216-8.

[27] Du H, Dong Z, Wu L, et al. A deep-learning based system using multi-modal data for diagnosing gastric neoplasms in real-time (with video)[J]. Gastric Cancer, 2023,26(2):275-285. DOI: 10.1007/s10120-022-01358-x.

[28] 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.

[29] Bisschops R, Areia M, Coron E, et al. Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative[J]. Endoscopy, 2016, 48(9): 843-864. DOI: 10.1055/s-0042-113128.

[30] Wani S, Keswani R, Hall M, et al. A prospective multicenter study evaluating learning curves and competence in endoscopic ultrasound and endoscopic retrograde cholangiopancreatography among advanced endoscopy trainees: the rapid assessment of trainee endoscopy skills study [J]. Clin Gastroenterol Hepatol, 2017, 15(11): 1758-1767. e11. DOI: 10.1016/j.cgh.2017.06.012.

[31] Yao L, Zhang J, Liu J, et al. A deep learning-based system for bile duct annotation and station recognition in linear endoscopic ultrasound[J]. EBioMedicine, 2021, 65: 103238. DOI: 10.1016/j.ebiom.2021.103238.

[32] Wu HL, Yao LW, Shi HY, et al. Validation of a real-time biliopancreatic endoscopic ultrasonography analytical device in China: a prospective, single-centre, randomised, controlled trial[J]. Lancet Digit Health, 2023, 5(11): e812-820. DOI: 10.1016/S2589-7500(23)00160-7.

www.zhxhnjzz.com

FUJIFILM
Value from Innovation

清晰诊疗 健康相伴

广告

New Generation Endoscope System

NEW

ELUXEO 7000

新 一 代 内 镜 系 统

LCI BLI

LCI：联动成像技术
BLI：蓝光成像技术

新定义 新选择

NEW DEFINITION NEW CHOICE



沪械广审(文)第271130-61740号

富士胶片株式会社
FUJIFILM Corporation
东京都港区西麻布二丁目26番30号

富士胶片(中国)投资有限公司
FUJIFILM (China) Investment Co., Ltd.
上海市浦东新区平家桥路100弄6号晶耀前滩T7，6楼

⚠ 禁忌内容或注意事项详见说明书。

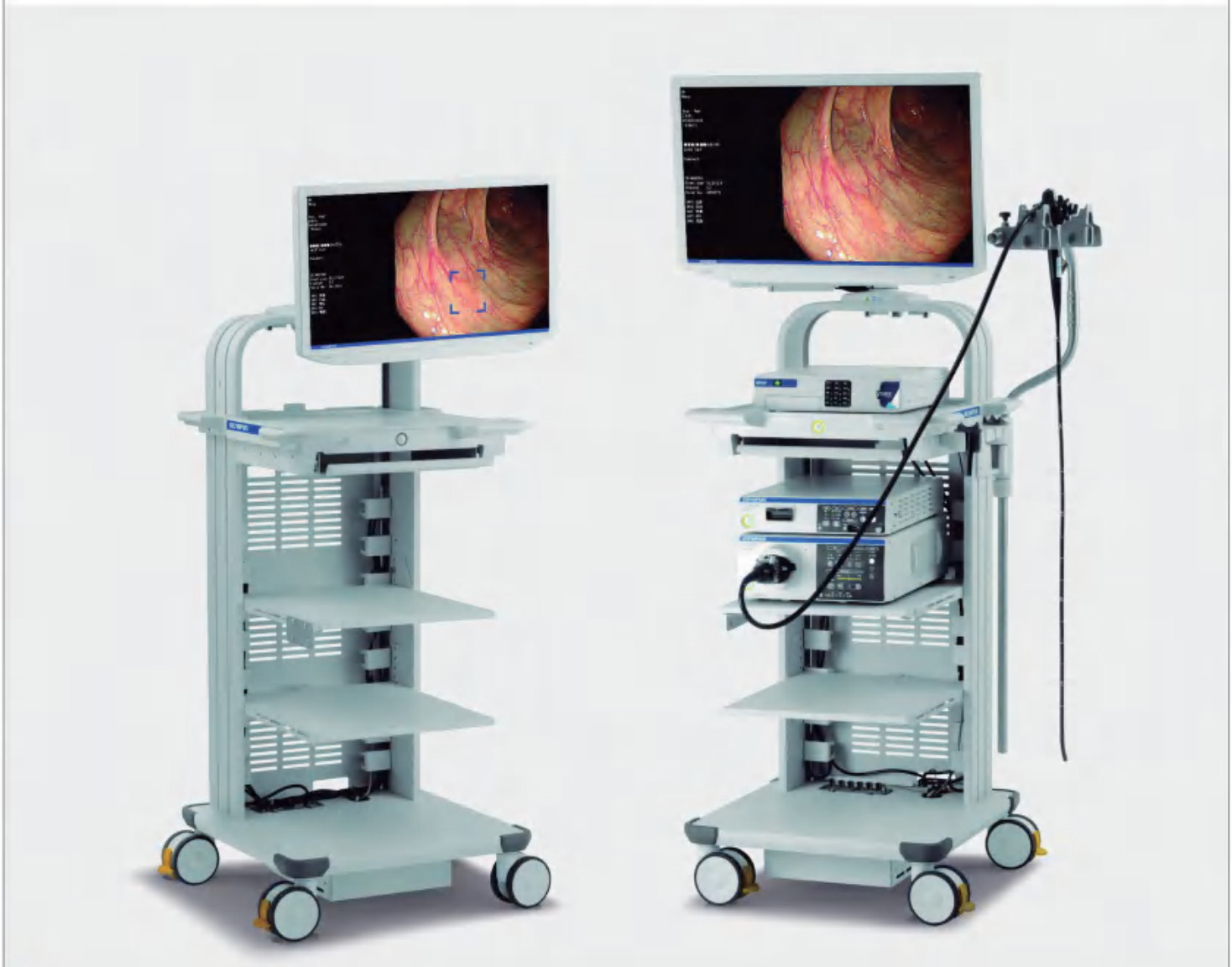
ELUXEO7000为VP-7000与BL-7000的统称
VP-7000：电子图像处理器 国械注进 20172062462
BL-7000：医用内窥镜用冷光源 国械注进20182060487

商标 FUJIFILM 和产品标识均为日本富士胶片株式会社持有。

OLYMPUS

广告

影诺医疗 INNOVISION



人工智能 助力诊疗

肠息肉电子下消化道内窥镜图像辅助检测软件

影诺鹰眼®

厦门影诺医疗科技有限公司

产品名称：肠息肉电子下消化道内窥镜图像辅助检测软件

注册证号：国械注准20233211648

生产企业：厦门影诺医疗科技有限公司

奥林巴斯(北京)销售服务有限公司为厦门影诺医疗科技有限公司授权经销商。

北京总部：北京市朝阳区新源南路 1-3号 平安国际金融中心 A 座 8 层

代表电话: 010-58199000

医疗内镜室户服务中心：400-820-2084

广告批准文号：闽械广审(文)第281107-04403号

本资料仅供医学专业人士阅读。

禁忌内容或注意事项详见说明书。

所有类别均基于本公司产品，特此说明。

规格、设计及附件如有变更，请以产品注册信息为准。

AD0075SV V01-2312

ISSN 1007–5232　CN 32–1463/R　邮发代号：28–105　国内定价：25.00 元