·数智医学教育·

聚焦眼科学生数据标注能力培养的可信人工智能协同模式研究

刘天航[1]　张颖麟[1]　刘子莹[2]　牛妍妍[2]　胡凌溪[1]　王星月[1]　刘江[1]
[1]南方科技大学计算机科学与工程系，深圳 518055；[2]温州医科大学眼视光学院，温州 325027
通信作者：刘江，Email：liuj@sustech.edu.cn

【摘要】 目的　构建可信人工智能（artificial intelligence，AI）与眼科学生的协同标注模式，评估与研究其在眼科学生数据标注能力培养中的应用价值。**方法**　本研究共纳入30张视网膜眼底图像，按难易程度分为简单组和复杂病例组，每组各15张。招募15名眼科学生，在无AI辅助、AI辅助、AI辅助+不确定信息图3种标注场景下开展标注实验。采用平均标注时间、Dice系数（Dice similarity coefficient，DSC）及几何平均比作为评价指标，根据数据分布特征选用配对t检验或Wilcoxon符号秩检验，分析评估该协同模式对眼科学生标注表现的影响。实验结束后通过问卷收集学生对AI协同模式在效率提升、辅助价值、使用意愿及AI认知影响等方面的反馈。**结果**　在AI辅助下，标注者的眼底血管标注准确性与效率均有提升，差异具有统计学意义（P<0.05）。简单病例组平均标注耗时从1 154.79 s降至292.09 s（降幅75.13%，效应量0.867 0），DSC从66.72%升至73.98%（提升11.18%，效应量0.858 0）；复杂病例组平均标注耗时从716.90 s降至207.78 s（降幅67.30%，效应量0.859 6），DSC从62.62%升至74.41%（提升18.98%，效应量0.855 6）。同时，个体间差异显著缩小：简单病例组标注耗时人均标准差从410.65 s降至130.19 s，DSC人均标准差从3.25%降至0.53%；复杂病例组标注耗时人均标准差从292.35 s降至76.78 s，DSC人均标准差从3.18%降至0.78%。问卷调查显示，所有参与的学生均认可该模式能提升标注效率与准确性，且提升了对AI的临床应用认知与信任度。**结论**　可信AI协同模式可提升眼科学生眼底血管标注的效率与质量，有效缩小个体标注能力差异，增强对人工智能的认知与信心。该模式有望作为眼科临床教育的补充工具，为眼科学生数据标注能力培养提供初步参考路径。

【关键词】 可信AI协同模式；　医学数据标注；　能力培养
【中图分类号】 R-0；R-4；R-05
基金项目：国家自然科学基金面上项目（82272086）；国家重点研发计划（2024YFC2510800）；南方科技大学高等教育教学研究和改革项目（SJZLGC202430）
DOI: 10.3760/cma.j.cn116021-20251229-02319

Study of a trustworthy AI collaborative model focused on data annotation skill training for ophthalmology students

Liu Tianhang[1], Zhang Yinglin[1], Liu Ziying[2], Niu Yanyan[2], Hu Lingxi[1], Wang Xingyue[1], Liu Jiang[1]
[1]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China; [2]School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou 325027, China
Corresponding author: Liu Jiang, Email: liuj@sustech.edu.cn

【Abstract】 Objective　To construct a collaborative annotation model between trustworthy artificial intelligence (AI) and ophthalmology students, and to evaluate and explore its application value in cultivating the data annotation skills of ophthalmology students. **Methods**　A total of 30 retinal fundus images were included in this study and divided into a simple case group and a complex case group according to the difficulty level, with 15 images in each group. Fifteen ophthalmology students were recruited to conduct annotation experiments under three scenarios: no AI assistance, AI-assisted annotation, and AI-assisted annotation with uncertainty maps. The impact of this collaborative model on the annotation performance of

ophthalmology students was analyzed and evaluated using average annotation time, Dice similarity coefficient (DSC), and geometric mean ratio (GMR), with statistical analysis performed by paired *t*-test or Wilcoxon signed-rank test (depending on data distribution characteristics). After the experiment, a questionnaire was used to collect students' feedback on the AI collaborative model in terms of efficiency improvement, auxiliary value, willingness to use, and impact on AI cognition. **Results** With AI assistance, the accuracy and efficiency of fundus vessel annotation by the annotators were significantly improved ($P<0.05$). In the simple case group, the average annotation time decreased from 1 154.79 s to 292.09 s (a reduction of 75.13%, effect size 0.867 0), and the DSC increased from 66.72% to 73.98% (an improvement of 11.18%, effect size 0.858 0). In the complex case group, the average annotation time decreased from 716.90 s to 207.78 s (a reduction of 67.30%, effect size 0.859 6), and the DSC increased from 62.62% to 74.41% (an improvement of 18.98%, effect size 0.855 6). Meanwhile, the inter-individual variability was significantly reduced: in the simple case group, the inter-individual standard deviation of annotation time decreased from 410.65 s to 130.19 s, and the inter-individual standard deviation of DSC decreased from 3.25% to 0.53%; in the complex case group, the inter-individual standard deviation of annotation time decreased from 292.35 s to 76.78 s, and the inter-individual standard deviation of DSC decreased from 3.18% to 0.78%. The questionnaire survey showed that all participating students recognized that this model could improve annotation efficiency and accuracy, and enhanced their cognition of and confidence in the clinical application of AI. **Conclusions** The trustworthy AI collaborative model can improve the efficiency and quality of fundus vessel annotation by ophthalmology students, effectively reduce the inter-individual differences in annotation skills, and enhance students' cognition of and confidence in artificial intelligence. This model is expected to serve as a supplementary tool for clinical ophthalmology education and provides a preliminary reference path for the cultivation of data annotation skills of ophthalmology students.

【**Key words**】 Trustworthy artificial intelligence collaborative model; Medical data annotation; Skill cultivation

Fund program: General Project of National Natural Science Foundation of China (82272086); National Key Research and Development Program of China (2024YFC2510800); Southern University of Science and Technology Higher Education Teaching Research and Reform Project (SJZLGC202430)

DOI: 10.3760/cma.j.cn116021-20251229-02319

2024年，习近平总书记明确提出发展新质生产力的战略要求，这是对新一轮科技革命和生产力发展趋势的创造性概括。卫生健康领域作为科技创新的重要前沿阵地，是新质生产力培育和壮大的关键方向。加快推动人工智能技术与医疗技术、医学教育的深度融合，既是提升医疗服务水平的重要举措，也是推动医学模式创新的必然要求。人工智能技术已逐步成为优化临床流程、提升诊疗效率的重要工具，其在疾病诊断[1]、智能分析[2]、手术规划与辅助[3]等医疗场景的规模化应用已成为必然趋势。尤其是随着ChatGPT、DeepSeek及不断涌现的医学AI大模型在语言处理、知识问答及逻辑推理等方面展现出的显著优势，迫切要求医学教育与医疗从业者提高对人工智能的认识水平[4]。梁爽等[5]指出，对规范化培训医师进行AI技术赋能的培训，可提升规培医师的AI工具应用能力与素养。数据是人工智能技术的核心基石，医学数据标注是指医护人员对各类医疗影像、临床文本、生理信号等医疗数据进行标注，为人工智能模型训练提供高质量标注样本的关键过程，其标注质量直接影响模型安全性与可靠性[6]。面向AI的医学数据标注正成为医护人员必备的基础技能与素养。数据标注的错误会直接破坏模型训练的基础逻辑，导致模型输出错误诊断建议，给人工智能模型的安全性与可靠性带来极大的风险[6]。数据标注的错误来源主要包括以下3种：一是标注者对AI技术原理、数据标注规范缺乏系统认知，导致标注内容与算法识别逻辑不匹配；二是临床医护人员工作负荷重，可用于标注的时间精力有限，易出现标注疏漏与偏差；三是医学领域细分度高，医学高度碎片化的知识体系易引发跨领域判断误差。传统人工标注存在效率低、质量差异大的问题[6]，而现有医学教育中医护人员数据标注能力的系统性培养路径也不完善，如何构建科学高效的培养体系，成为人工智能与医学教育、医疗技术融合发展亟须解决的关键问题。

教育智能是指借助人工智能算法理解教育[7]，在人工智能与医学教育融合领域，国内外学者已围绕AI技术赋能医学生临床能力与职业素养培养开

展多维度探索。张绍衡等[8]针对 AI 赋能住院医师规范化培训的核心需求，从实施条件、教学体系、评价标准等维度，提出了AI辅助循证决策培训的系统化建设思路。徐赫和焦洋[9]针对人工智能驱动的临床推理教学新模式展开系统研究与展望，明确了AI技术赋能医学教育转型的核心实施路径，为高等医学教育智能化改革提供了理论参考与实践依据。谭思创等[10]基于研究生视角开展调研，证实多数医学生对AI融入医学教育持积极态度，但目前相关配套教学与系统培训仍存在显著缺口。刘江和章晓庆[11]构建了面向非计算机背景医学生的AI课程体系与个性化教学方法，为医学交叉学科人才培养提供了实践参考。张传钊等[12]将生成式AI引入临床实习教学，证实其可有效提升医学生的理论学习与临床决策能力。然而，上述研究均未聚焦医学生医学数据标注能力建设这一医学 AI 研发的核心基础环节，而高质量、规范化的数据标注既是构建高可靠性医学AI模型的核心前提，也是帮助医学生理解AI技术逻辑、提升AI应用素养的重要路径。本研究中，可信AI的核心定义为具备可解释性、不确定性量化能力与结果可重复核验的人工智能模型，其核心特征具体表现为2个方面：一是输出像素级血管分割结果，为标注者提供参考基线；二是同步输出模型预测的置信度分布，实现预测结果的可解释性与高风险区域的风险提示，满足医学场景对AI结果透明性与可靠性的核心要求。基于上述定义，本研究以视网膜眼底图像的血管标注任务为研究切入点，评估可信AI协同标注模式对眼科学生标注表现的影响，初步研究该模式在眼科学生数据标注能力培养中的应用价值。构建可信人工智能协同模式旨在解决传统人工标注的痛点，为临床医生减负增效；同时，为医学生数据标注能力培训与技能评估体系的完善提供参考；此外，为构建高质量医学影像大规模标注数据集提供了技术支撑，推动人工智能作为新质生产力，与医学教育、医疗技术深度融合，促进行业发展。

1　对象与方法

1.1　研究对象

本研究共招募15名眼科相关专业学生完成视网膜眼底图像血管标注工作，其中研究生7名、本科生8名，均无眼底图像标注经验。所有受试者需在3种预设实验条件下分批次完成标注任务，每批次标注2~3张图像，相邻批次的标注间隔为3 d。正式标注前，所有受试者均接受了标注操作与判读标准的统一培训。按照标注难度，视网膜眼底图像数据分为简单病例组和复杂病例组2组，2组图像抽样均通过Python程序实现，全程无人为筛选干预，确保抽样过程的随机性与客观性。①简单病例组。共纳入15张视网膜眼底图像，数据从STARE数据集[13]与DRIVE数据集[14]中随机选取，涵盖正常眼底图像及伴有早期轻度病理改变（如早期糖尿病视网膜病变、轻度高血压视网膜病变）的图像。该组图像具备眼底背景与血管对比度高、血管脉络清晰可辨、成像条件标准化及图像质量稳定的特征，因此血管标注任务难度较低。②复杂病例组。共纳入15张视网膜眼底图像，数据均从FIVE800[15]眼底数据集中随机采样获取，涵盖正常眼底图像及伴糖尿病视网膜病变、青光眼、年龄相关性黄斑变性等多种中重度病理改变的图像。该组图像存在显著的成像差异与质量波动，部分图像表现为血管与背景对比度低、伴噪声干扰或图像模糊，且受中重度病理改变叠加影响，血管边界模糊不清，显著增加了血管标注任务的难度。

本研究选用的STARE、DRIVE、FIVE800均为视网膜眼底血管分割领域国际公认的权威公开数据集。其中STARE、DRIVE以标准化高质量的健康及轻中度病变眼底图像为主，与简单病例组难度高度适配；FIVE800涵盖多设备、多病程、多病变类型的图像，包含大量低对比度、高噪声的中重度病变图像，与复杂病例组难度高度匹配。数据集分组贴合临床真实场景，具备良好的行业代表性。2个病例组的真实标签（ground truth，GT）由数据集自带的专家标注提供，样例数据如图1所示。

1.2　方法

1.2.1　实验设计

本研究设计了3种不同条件下的标注实验，以评估可信AI协同标注模式对眼科学生标注表现的影响，如图2所示。第一种为手工标注（无AI辅助），即完全不依赖AI模型的辅助，仅凭借自身专业知识开展标注；第二种为AI辅助，即在AI模型生成的初步分割结果基础上，针对偏差区域进行人工修正，提升标注效率；第三种为AI辅助+不确定性图提示，即在利用AI模型初始分割结果的同时，参考其输出的不确定性图以指导人工修正过程，提升标注精准度。

1.2.2　数据预处理

AI辅助的分割结果与不确定性图由一种基于贝叶斯近似的眼底图像分割模型[16]输出，该模型在公开的眼底血管分割数据集上训练得到，能够同时

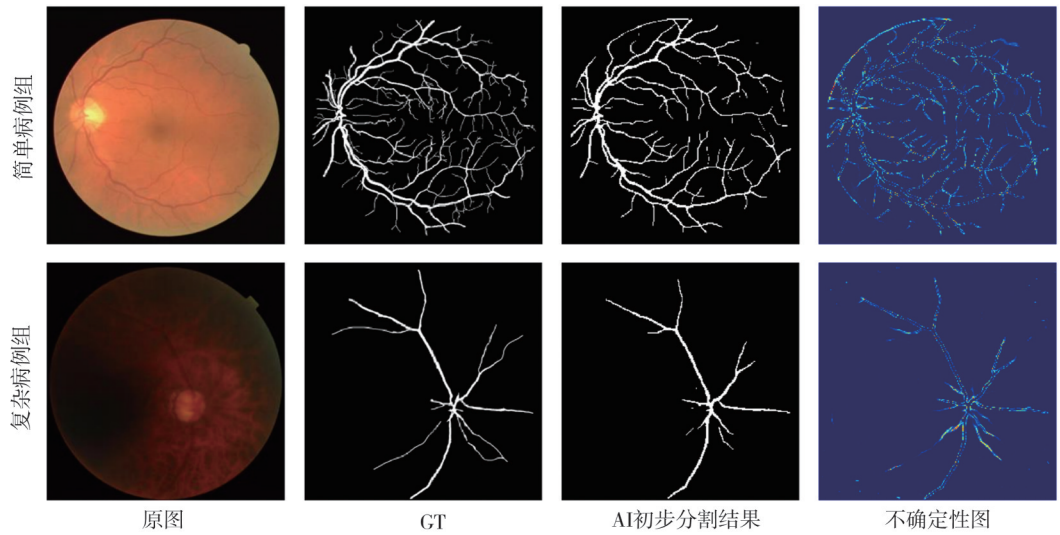


图1　2个病例组的眼底图像

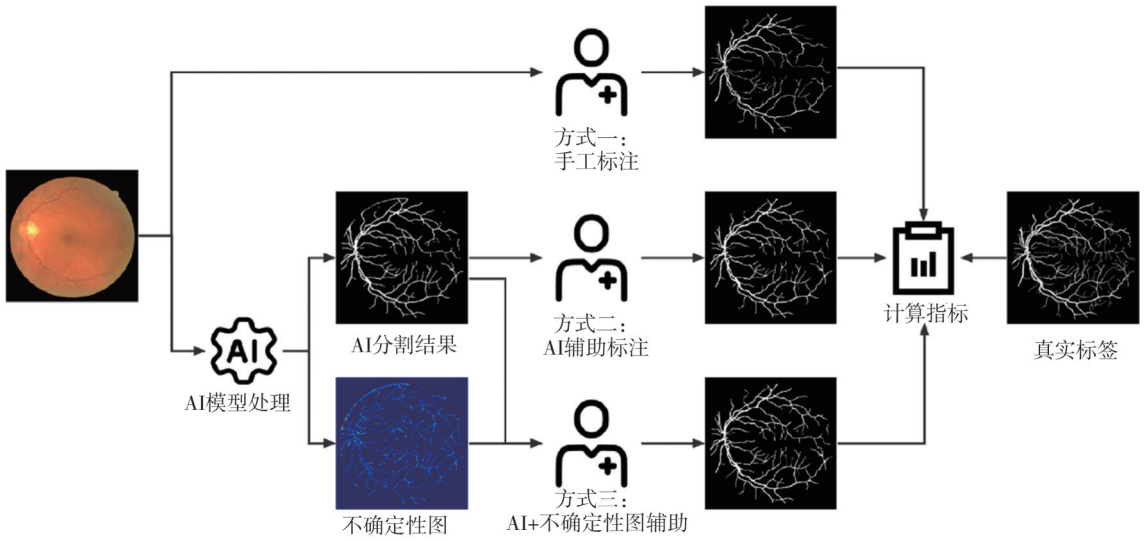


图2　标注实验设计示意图

输出血管分割结果及其像素级不确定性图。不确定性图是一种像素级概率分布方差图，用于反映模型在每个位置预测结果的置信度高低，数值较高的区域表示模型分割结果存在较大不确定性，提示标注者在该区域需重点关注与修正。不确定性图是本研究中可信AI的核心实现路径。模型通过输出不确定性图，以透明化的方式向标注者展示预测结果的置信度，提示高风险修正区域，帮助标注者理解AI的决策边界，进而提升对AI工具的信任度。在眼底血管图像中，细小血管分支、模糊的血管边缘或病灶遮挡区域常导致模型预测结果不一致，从而出现高不确定性，提示这些区域的分割结果可能存在误差，需要标注者重点校对。在图1的不确定性可视化中，使用蓝色表示低不确定性，红色表示高不确定性。

1.3　评价指标

在上述3种标注方式下，分别记录并分析医生的标注结果，设定2类评价指标以量化模型辅助效用：一是标注效率，评估平均标注耗时，如公式(1)所示；二是标注准确性，采用Dice系数(Dice similarity coefficient，DSC)作为核心指标评价标注准确性，如公式(2)所示。平均耗时的计算公式为：

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i \tag{1}$$

t_i为第i张图像的标注耗时，N表示样本数量。

DSC是衡量标注结果与真实标签之间的重叠程度，取值范围为[0，1]，值越大表示重叠越好，其计算方式如下：

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \times 100\% \tag{2}$$

其中，A和B分别表示标注结果和真实标签的像素

集合。

此外，为更直观地刻画不同标注方式在效率与准确性上的相对提升幅度，本研究还计算了几何平均比（geometric mean ratio，GMR）。GMR 计算如公式（3）所示：

$$GMR = \exp(\bar{d}), \bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i, d_i = \ln Y_i - \ln X_i = \ln\left(\frac{Y_i}{X_i}\right) \quad (3)$$

其中，样本基线方法指标为 X_i，对比方法指标为 Y_i。GMR 大于 1 视为提升，小于 1 视为降低。由此，百分比变化可表述为 $(GMR - 1) \times 100\%$。

为评估个体间标注差异，本研究除计算图像层面的标准差外，还计算了标注者标注表现的标准差，如公式（4）所示：

$$SD_{annot} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(x_k - \bar{x})^2} \quad (4)$$

其中，m 为标注者人数，x_k 为第 k 名标注者的标注表现，$\bar{x}$ 为所有人标注表现的均值。该指标用于量化标注者之间的差异，标准差越小，说明标注者的标注表现差异越不显著，标注稳定性越好。

1.4　统计学方法

数据检验分析均在 Python3、NumPy 与 SciPy 库环境下完成，描述统计数据采用均值 ± 标准差方法，并计算 95% 置信区间（95% confidence interval，95*CI*）。在统计分析过程中，首先使用 Shapiro-Wilk 检验评估样本分布的正态性。当统计量 w 接近 1 时，样本被认为近似服从正态分布。为量化正态分布数据差异的效应大小，计算 Cohen's d 效应量 d。Cohen's d 的绝对值越大，说明组间差异相对于样本波动越显著。数据符合正态分布时，采用配对样本 t 检验；若不符合正态分布，则采用 Wilcoxon 符号秩检验进行非参数比较，计算统计量 W 和效应量 r。检验水准 α=0.05。

2　结果

本研究对比了学生在手工标注、AI 辅助、AI+不确定信息图 3 种标注场景下的表现，系统评估了可信 AI 协同标注模式对眼科学生标注表现（效率与准确性）的影响，基于实验数据分析了该模式在眼科学生数据标注能力培养中的应用效果。

2.1　标注效率

不同标注方式下的平均标注耗时统计结果显示，相比于手工标注，AI 辅助的引入显著缩短了眼科学生的标注时间，且在简单病例组和复杂病例组中均表现出一致的效率提升趋势。如表 1 所示，简单病例组中，标注平均耗时从 1 154.79 s（95%*CI*=1 052.19~1 257.38 s）降至 292.09 s（95%*CI*=269.61~314.58 s），GMR 为 24.87%，降幅为 75.13%；经 Shapiro-Wilk 检验，该组数据不符合正态分布（w=0.72，α_s<0.001），因此采用 Wilcoxon 符号秩检验，差异具有统计学意义（W=0.00，P<0.05，效应量 $|r|$=0.867 0）。复杂病例组中，标注平均耗时从 716.90 s（95%*CI*=650.19~783.60 s）降至 207.78 s（95%*CI*=191.59~223.98 s），GMR 为 32.70%，降幅为 67.30%；经 Shapiro-Wilk 检验，该组数据不符合正态分布（w=0.89，α_s<0.001），因此采用 Wilcoxon 符号秩检验，差异具有统计学意义（W=132.00，P<0.001，效应量 $|r|$=0.858 0）。

2.2　标注准确性

DSC 作为评估医学图像分割准确性的核心指标，结果均显示 AI 辅助标注显著提升了标注准确性。如表 1 所示，简单病例组中，标注 DSC 由 66.72%（95%*CI*=66.04%~67.40%）升至 73.98%（95%*CI*=73.73%~74.23%），GMR 为 111.18%，提升 11.18%；经 Shapiro-Wilk 检验，该组数据不符合正态分布（w=0.97，α_s<0.001），因此采用 Wilcoxon 符号秩检验，差异具有统计学意义（W=109.00，P<0.001，效应量 $|r|$=0.859 6）。复杂病例组中，标注 DSC 由 62.62%（95%*CI*=61.14%~64.10%）升至 74.41%（95%*CI*=72.77%~76.06%），GMR 为 118.98%，提升 18.98%；经 Shapiro-Wilk 检验，该组数据不符合正态分布（w=0.98，α_s=0.02），因此采用 Wilcoxon 符号秩检验，差异具有统计学意义（W=167.00，P<0.001，效应量 $|r|$=0.855 6）。此外，图 3 对标注效率和准确性的提升进行了更为直观的可视化展示。

除整体标注准确性的提升外，AI 辅助还显著缩小了标注者之间在标注准确性上的个体差异，这一趋势通过标注者 DSC 的标准差、极差等核心统计指标的变化得到验证（表 2），且在简单与复杂病例组中呈现一致规律。简单病例组中，手工标注阶段 DSC 的人均标准差为 3.25%，极差为 10.37%，反映出不同标注者的标注效果波动较大；AI 辅助后，DSC 人均标准差降至 0.53%，极差缩小至 1.61%，分别较手工标注阶段降低 83.69% 和 84.48%，标注结果的一致性显著提升。复杂病例组中，手工标注时 DSC 人均标准差为 3.18%，极差为 10.37%；AI 辅助后人均标准差降至 0.78%，极差缩小至 2.58%，降幅分别为 75.47% 和 75.02%。图 4 和图 5 的箱线图进一步直观呈现了 AI 辅助后标注者间准确性差异缩小的分布趋势。

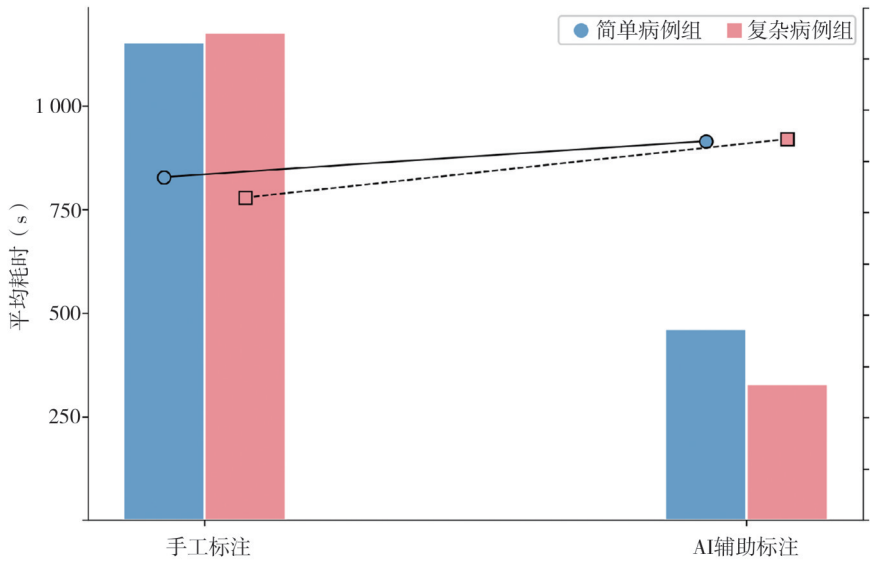


图3　不同标注方式下平均耗时（柱状图）与DSC（折线图）对比

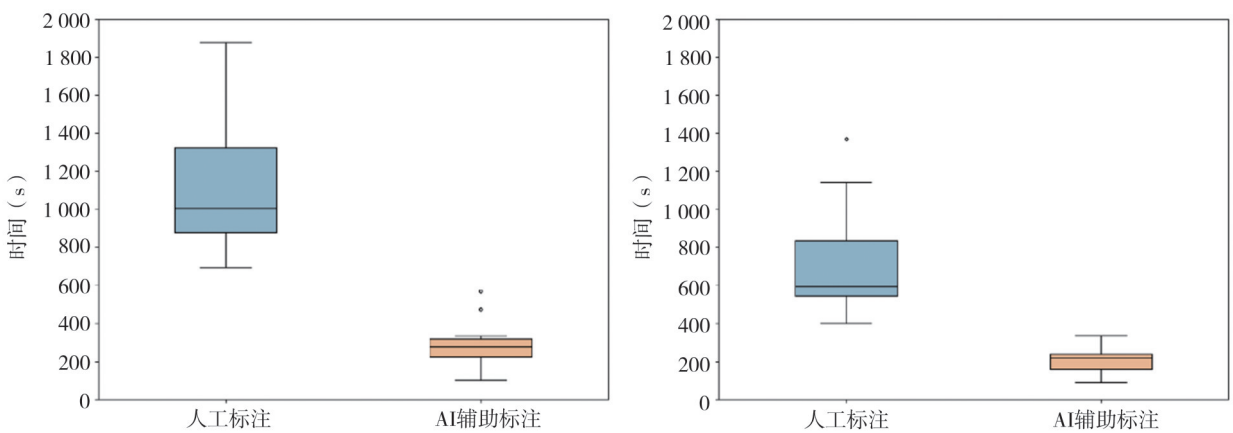
表1　独立标注与AI辅助标注的比较（$\bar{x} \pm s$）

指标	病例组	人工标注	AI辅助标注	差值	Shapiro-Wilk检验（w, α_s）	t（W）值	P值	$\lvert d \rvert$（$\lvert r \rvert$）值	GMR（%）
时间（s）	简单组	1 154.79 ± 780.91 [1 052.19, 1 257.38]	292.09 ± 171.13 [269.61, 314.58]	−862.69 ± 737.51 [−959.58, −765.80]	（0.72, 0.00）	0.00	<0.05	0.867 0	24.87
	复杂组	716.90 ± 507.75 [650.19, 783.60]	207.78 ± 123.29 [191.59, 223.98]	−509.12 ± 451.91 [−568.48, −449.75]	（0.89, 0.00）	132.00	<0.05	0.858 0	32.70
DSC（%）	简单组	66.72 ± 5.15 [66.04, 67.40]	73.98 ± 1.90 [73.73, 74.23]	7.26 ± 4.62 [6.65, 7.87]	（0.97, 0.00）	109.00	<0.05	0.859 6	111.18
	复杂组	62.60 ± 11.28 [61.14, 64.10]	74.41 ± 12.51 [72.77, 76.06]	11.79 ± 6.28 [10.96, 12.61]	（0.98, 0.02）	167.00	<0.05	0.855 6	118.98

注：方括号内为95%置信区间。

表2　标注者之间标注表现统计

指标	病例组	手工标注				AI辅助标注			
		标准差	最大值	最小值	极差	标准差	最大值	最小值	极差
时间（s）	简单组	410.65	1 875.60	694.27	1 181.33	130.19	569.80	100.20	469.60
	复杂组	292.35	1 368.27	400.87	967.40	76.78	335.13	90.33	244.80
DSC（%）	简单组	3.25	71.59	61.22	10.37	0.53	74.53	72.91	1.61
	复杂组	3.18	67.94	57.57	10.37	0.78	75.91	73.33	2.58



A. 简单病例组标注时间分布　　B. 复杂病例组标注时间分布

图4　标注时间分布箱线图

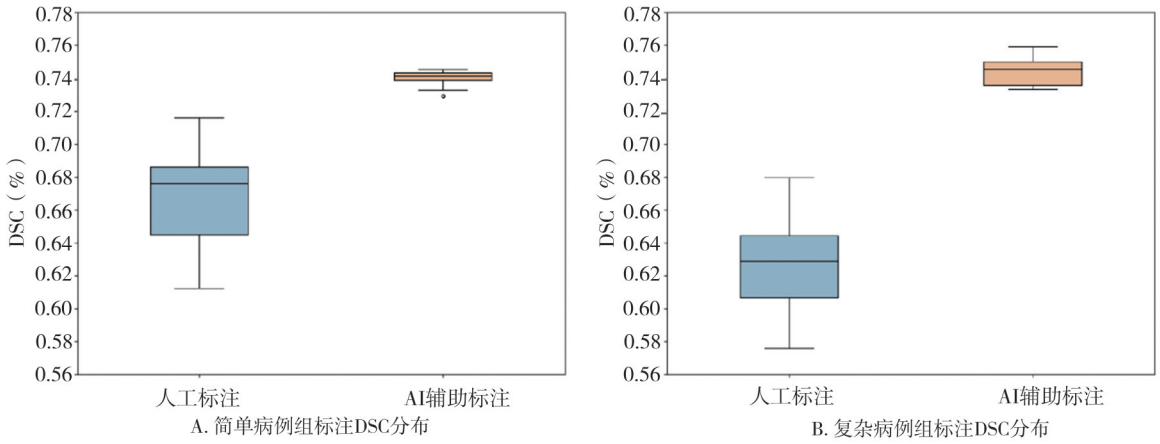


图5　标注DSC分布箱线图

表3　AI辅助与AI+不确定性图辅助在复杂病例组的准确性比较（%，$\bar{x} \pm s$）

病例组	AI辅助	AI+不确定性图辅助	差值	Shapiro-Wilk检验（w，α_s）	t（W）值	P值	$\lvert d \rvert$（$\lvert r \rvert$）值	GMR（%）
简单组	73.98 ± 1.90 [73.73，74.23]	73.87 ± 1.92 [73.61，74.12]	−0.12 ± 0.38 [−0.17，−0.07]	（0.81，0.00）	3 290.00	<0.05	0.642 6	99.84
复杂组	74.41 ± 12.51 [72.77，76.06]	73.82 ± 12.68 [72.15，75.48]	−0.60 ± 1.37 [−0.78，−0.42]	（0.61，0.00）	1 321.00	<0.05	0.776 9	99.09

注：方括号内为95%置信区间。

2.3　不确定图对AI辅助标注的影响

不确定性图的辅助效果不佳。简单病例组中，引入不确定性图后DSC略降至73.87%（95%*CI*=73.73%~74.23%），差值为−0.12%，GMR为99.84%；经Shapiro-Wilk检验，该组数据不符合正态分布（w=0.81，α_s<0.001），因此采用Wilcoxon符号秩检验，差异具有统计学意义（W=3 290.00，P<0.001，效应量$\lvert r \rvert$=0.642 6）。复杂病例组中，DSC降至73.82%（95%*CI*=72.15%~75.48%），差值为−0.60%，GMR为99.09%，整体呈现与简单病例组类似的轻微下降趋势；经Shapiro-Wilk检验，该组数据不符合正态分布（w=0.61，α_s<0.001），因此采用Wilcoxon符号秩检验，差异具有统计学意义（W=1 321.00，P<0.001，效应量$\lvert r \rvert$=0.776 9）（表3）。标注实验结束后，本研究面向15名参与实验的眼科学生发放调查问卷，通过选择题与简答题2大模块，围绕AI协同标注的效率提升、不确定性图辅助价值、同类工具使用意愿及AI认知变化4个维度，收集学生的主观评价与使用反馈。选择题结果显示，所有学生均认可AI对标注效率的提升作用；7名学生认可不确定性图的辅助价值，7名持中立态度，仅1名认为其无帮助；14名学生愿意未来继续使用同类AI工具；10名学生认为使用AI后自身对人工智能技术的理解有所增强。简答题环节，学生普遍反馈AI协同标注的核心价值在于可快速生成标注模板、精准识别图像细节，简化标注流程；同时指出AI在低质量、结构复杂的图像中标注精度不足，初始模板需人工精细化修正。学生均表示，AI能有效保障标注结果的一致性与稳定性，但在复杂标注任务中仍存在局限，人工核验是数据标注的必要环节。

3　讨论

医学数据标注是推动医疗AI发展的重要基石。传统人工标注费时费力，培训医学生利用AI工具进行高效、准确的数据标注，既是新时代医学生人工智能素养的基本要求，也是专业能力建设的重要环节。本研究通过分析15名标注者在手工标注、AI辅助、AI+不确定信息图辅助标注3种方式下的实验结果，基于预设研究目标验证了可信AI协同标注模式能够影响眼科学生的标注表现，有望为医学教育中技能培训实践提供参考。结果显示：①可信AI协同标注模式使总体标注准确性与效率均有提升，差异具有统计学意义；②可信AI协同标注模式能够缩小标注者之间在标注效率与标注准确性方面的个体差异；③可信AI协同标注模式可提高眼科学生对AI工具的认识与理解，增强其使用AI工具的信心；④可信AI协同标注模式在眼科学生标注能力培养中具有初步应用潜力。

3.1　可信AI协同标注模式可提升标注者标注效率与准确性

从标注效率来看，可信AI协同标注模式使标注者在简单病例组与复杂病例组的标注时间均实现

显著缩减(均$P<0.05$),显示出其能有效提升标注效率。具体而言,标注者的平均耗时在简单病例组中由 1 154.79 s 降至 292.09 s,降幅 75.13%;在复杂病例组中由 716.90 s 降至 207.78 s,降幅 67.30%(表 1)。结合对标注者的问卷调查,效率提升的核心原因可归为两点。一是 AI 通过预生成初步血管分割结果,替代了标注者手动勾勒血管轮廓的基础工作,这使得原本需逐像素判断血管边界的流程转变为对 AI 结果的针对性修正,大幅减少了重复劳动;二是 AI 对血管的初步定位,帮助标注者快速锁定标注范围,避免因视野遗漏导致的反复核对,进一步优化了标注流程。

在标注准确性方面,可信 AI 协同标注模式同样表现出显著提升。简单病例组的平均 DSC 由 66.72% 升至 73.98%,提升 11.18%;复杂病例组由 62.62% 升至 74.41%,提升 18.98%。其作用机制主要体现在以下 2 个方面:首先,在相对简单的标注任务中,AI 基于大规模训练数据所学习到的血管特征,能够识别因经验不足而易被忽略的细小血管结构;其次,当血管边界模糊或组织结构复杂时,AI 可利用其对高维特征的敏感性,在低对比度条件下分割人眼难以识别的血管。

综上,可信 AI 协同标注模式可有效提升眼科学生眼底血管标注任务的标注效率与准确性,可为眼科学生数据标注能力建设提供参考。

3.2　可信 AI 协同标注模式可缩小个体间标注差异

引入可信 AI 协同后,还呈现出明显的个体间效率与准确性差异缩小趋势,这一趋势通过标注者之间的标准差、极差等核心统计指标的变化得到明确验证(表 2),且在简单病例组与复杂病例组中呈现一致规律。

在标注效率方面,可信 AI 协同标注模式有效压缩了标注者之间的耗时差异。在简单病例组中,无 AI 辅助时标注耗时的人均标准差为 410.65 s,极差高达 1 181.33 s,表明不同标注者的标注速度差异大;引入 AI 辅助后,人均标准差降至 130.19 s,极差压缩至 469.60 s,分别较手工标注阶段降低 68.28% 和 60.25%。复杂病例组的优化效果更为突出,手工标注时标注耗时的人均标准差为 292.35 s,极差为 967.40 s;AI 辅助后人均标准差降至 76.78 s,极差缩小至 244.80 s,降幅分别达 73.73% 和 74.68%。这一数据变化表明,AI 协同通过标准化基础标注流程,大幅降低了个体在标注速度上的差异,使群体标注效率趋于均衡。其核心原因在于,AI 预生成的血管分割初始模板从根本上重构了标注任务的执行逻辑:将传统手工标注逐像素勾勒血管结构的高负荷基础工作,转化为仅针对 AI 预标注偏差区域的局部修正与精细化调整。与此同时,传统手工标注需根据血管管径、目标标注区域位置,频繁调整笔刷、缩放平移图片视野,才能完成血管边缘的精准勾勒;而 AI 协同模式下,标注者无需再进行大范围、高频次的软件基础操作,仅需对少数偏差区域做小范围修正,压缩了标注者间软件操作环节的时间消耗差异,从操作执行层面进一步缩小了个体标注速度的差异。通过对上述任务核心环节与操作流程的优化,初步实现了不同基础水平标注者的标注效率均衡化。

值得注意的是,作为补充辅助手段的不确定性图,其实际辅助效果仍存疑,尚未展现出稳定的正向价值。从群体数据来看,简单病例组中引入不确定性图后,DSC 较单纯 AI 辅助模式略有下降(差值 –0.12%,降低 0.16%),复杂病例组中类似趋势更为明显(差值–0.60%,降低 0.91%),这一结果与调查问卷的主观反馈形成完整印证。问卷数据显示,仅 46.67%(7/15)的参与学生认可不确定性图的辅助价值,46.67%(7/15)持中立态度,另有 6.67%(1/15)明确认为其对标注无帮助。其效果不佳并非单一的人机交互障碍,核心成因可归结为 3 点:一是像素级热图的可视化设计存在固有缺陷,易造成视觉干扰与信息过载;二是未结合眼底血管解剖结构进行对应标注,学生无法将像素级不确定性与临床标注重点建立关联;三是学生无法准确理解不确定性图的意义与应用。本次研究通过专项问卷调查与针对性访谈,系统收集了学生对不确定性图的理解程度、使用体验、认知障碍与优化建议,明确了其辅助效果不佳的具体作用机制,为该辅助工具的优化迭代提供了来自临床教学场景的一手实证数据支撑。针对上述问题,未来将从 2 个维度优化改进:一是优化可视化呈现形式,聚焦高风险区域标注,减少视觉干扰;二是配套相关前置培训,同时探索基于视觉语言大模型问答的人机交互方式,提升不确定性图的辅助价值与可学习性。

综上,可信 AI 协同标注模式通过标准化基础标注流程、减轻重复劳动负担,有效减少了标注者在眼底血管标注任务中在标注效率与标注准确性方面的个体差异,为医学教育中数据标注能力的规范化、均衡化培养提供了初步量化依据。

3.3　AI 协同标注模式可提升眼科学生对人工智能的认知与信心

标注效率与准确性的客观数据变化,直观反映

了可信 AI 协同模式对眼科学生标注实操表现的提升作用，而本研究同步开展的问卷调查结果，则从主观层面进一步验证了该模式对学生人工智能认知与应用信心的积极影响。本研究的可信 AI 协同模式，并非单纯以 AI 替代人工基础劳动，而是让学生在实操中理解 AI 的预测逻辑与能力边界，有效增强了学生对 AI 应用的信心，这也是其在医学教育场景中核心应用价值的重要组成部分。

从主观反馈结果来看，参与实验的眼科学生对可信 AI 协同标注模式整体呈现出较高的接受度与认可度，绝大多数学生明确表示愿意在后续的学习与科研工作中继续使用同类 AI 辅助工具。其主要原因在于，不同于传统课堂中对 AI 技术的理论化教学，本研究设计的人机协同标注流程使学生在实操过程中直观理解 AI 模型在眼底影像分割中的应用逻辑、技术优势与应用边界，将抽象的 AI 技术转化为可操作、可感知的实用工具，有效降低了 AI 技术的学习与应用门槛。学生的问卷反馈同时清晰呈现了对 AI 辅助工具价值与局限性的双重认知。在应用价值层面，学生普遍认可 AI 协同模式对标注流程的标准化作用，其通过预生成标注模板的方式大幅削减了重复性基础劳动，同时能精准捕捉人工易遗漏的影像细节，为标注工作提供了可靠的技术参考；而在工具局限性层面，学生也客观指出了当前 AI 模型在低质量影像、复杂病理结构场景中的精度短板，明确了人工核验与修正对保障医学标注质量的核心必要性。这种对 AI 技术的理性认知，既避免了学生对 AI 工具的盲目排斥，也杜绝了过度依赖 AI 技术的误区，帮助学生建立了正确的 AI 应用理念，这也是医学教育中人工智能素养培养的核心目标之一。此外，学生针对不确定性图辅助效果的反馈，也为可信 AI 辅助工具的优化迭代提供了来自临床教育场景的真实视角。问卷结果与客观实验数据形成了相互印证，证实不确定性图当前的可视化设计与信息呈现方式，未能有效贴合医学生的临床认知习惯，也从侧面说明，医学场景下的可信 AI 建设，不仅需要技术层面的算法优化，更需要实现技术逻辑与临床认知、教学需求的深度适配，才能真正发挥其辅助价值。

综上，本次问卷的主观评价结果证实，可信 AI 协同标注模式能在认知层面增强学生对人工智能技术的理解深度与应用信心，为眼科乃至临床医学教育中医学生数据标注能力与人工智能素养的系统性培养提供了可参考的实践路径。

3.4　可信 AI 协同模式具有临床教学转化价值与应用可行性

本研究提出的可信 AI 协同标注模式，经本次对照实验验证，可提升眼科相关专业学生的眼底图像标注表现，同时有助于学生建立对医学 AI 工具的基础认知。该模式具备向眼科临床教学场景转化的初步潜力，可为完善医学生数据标注能力培养体系提供参考。

从落地实施的基础条件来看，该模式具备初步的低成本部署基础。在 AI 工具部署成本方面，本研究采用的基于贝叶斯近似的眼底图像分割模型为开源学术模型，可直接部署于普通教学用计算机中，无需额外配置专用硬件，单台教学终端的部署成本可控制在现有医学影像教学设备的常规预算范围内，院校无需大额经费投入即可完成部署。在学生培训周期方面，该模式操作门槛与学习成本较低，针对本次纳入的眼科本科生、研究生等受试对象，仅需少量课时的理论教学和实操练习，即可使其掌握该协同标注模式的操作流程与方法，培训周期较短，具备嵌入现有眼科教学课时安排的可行性。

从教学应用的成本效益来看，该模式在本次实验场景中展现出教学投入与教学产出双向优化的潜力。在教学效率提升方面，本研究数据显示，该模式可使单张眼底图像的平均标注耗时降低 67%~75%；原本需要长时间完成的眼底图像标注实操训练，采用该模式后仅需原来时间的 1/3 至 1/4 即可完成，有望在有限的教学课时内为学生提供更多的病例实操训练机会，覆盖更多不同难度层级的眼底病例类型，提升教学容量。在师资成本节约方面，该模式可通过 AI 预标注结果与专家金标准的自动比对，实现标注结果的初步量化评价，替代教师对学生标注作业的逐一批改，教师仅需针对 AI 识别的高误差区域、学生标注的共性问题进行集中讲解，从而将更多精力投入到个性化教学与重难点知识讲解中，实现师资资源的优化配置。从体系融入路径来看，该模式可多维度嵌入现有眼科医学教育体系，并探索构建适配性教学应用范式。在课程设置方面，可将该模式作为实操模块，融入《眼科学》等本科必修课程的眼底影像教学环节，或纳入《医学人工智能》等选修课程的核心教学内容，设置 2~4 个课时的理论与实操一体化教学，构建“医学知识、AI 技能、实操应用”的一体化课程内容。在教学场景设计方面，可初步适配 3 类核心教学场景：一是基础教学场景，以正常眼底的简单病例为核心，通过 AI

协同模式让学生快速掌握眼底血管解剖结构的标注规范，夯实影像识别基础；二是进阶科研场景，以中重度病变的复杂病例为核心，通过AI预标注与提示，训练学生对复杂眼底病变的标注能力，对接临床科研的真实需求；三是规培培训场景，结合规培生的临床工作需求，开展眼底影像标注的实操训练，培养规培生利用AI工具高效完成临床数据整理、科研数据标注的能力。

综上，本研究提出的可信AI协同标注模式，依托开源模型为医学生数据标注能力与AI素养的培养提供了初步的实践方案，也为AI与医学教育的深度融合提供了阶段性实证参考。该模式在本研究中展现出眼科临床教学转化的初步潜力，有望成为眼科临床教育的补充工具。

4　结语

本研究以视网膜眼底血管标注任务为核心，初步探究了可信AI协同标注模式在眼科学生数据标注能力培养方面的效用。研究结果表明，AI协同标注模式不仅显著提高了眼科学生的标注效率和准确性，还能有效缩小个体间在标注效率和准确性上的差异，同时增强眼科学生对AI的认知与信心。该模式可作为眼科临床教育的补充工具，为构建医学生数据标注能力系统性培养路径提供借鉴与参考，推动新医科人才培养。此外，作为补充辅助手段的不确定性图，其辅助效用仍存疑。本研究作为AI协同模式应用于眼科临床教育的初步研究，仍存在一定局限性：首先，样本量小、样本来源单一，限制了结果的广泛可推广性；其次，研究仅聚焦视网膜眼底图像血管标注任务，未覆盖其他医学影像类型及更复杂的病灶标注场景；第三，不确定性图的人机交互补充设计未能充分发挥辅助作用。未来研究可从3方面开展：一是扩展样本量，纳入不同学习阶段、不同专业水平的眼科学生开展分层次研究，进一步验证该模式的普适性；二是扩展研究场景，将该模式拓展到视盘、黄斑、眼底病灶等多种眼科影像标注任务，完善多场景的教学应用范式；三是优化可信AI提示的可视化设计与临床适配性，同时开展规模化的临床教学落地实践，结合教学反馈持续优化模式设计，进一步验证该模式在眼科学生数据标注能力培养中的实用性和适应性。

利益冲突　所有作者声明不存在利益冲突

作者贡献声明　刘天航：实验设计、数据收集及处理、论文撰写；张颖麟：实验设计、论文撰写；刘子莹、牛妍妍：数据标注流程设计；胡凌溪、王星月：论文修改；刘江：研究设计、项目指导、对研究总体负责

参考文献

[1] 胥瀚文，陈墨馨，梁小乙，等. 基于面部照片的眼病智能诊断研究进展[J]. 上海交通大学学报（医学版），2025，45(9)：1249-1255. DOI: 10.3969/j.issn.1674-8115.2025.09.017.

[2] 林宇晖. 基于人工智能的血常规分析技术在临床应用中的研究进展[J]. 中国冶金工业医学杂志，2025，42(4)：384-385. DOI: 10.13586/j.cnki.yjyx1984.2025.04.066.

[3] 李海强，苏微微，李秀萍，等. 机器人辅助甲状腺手术的研究进展[J]. 江苏医药，2025，51(8)：829-833. DOI: 10.19460/j.cnki.0253-3685.2025.08.016.

[4] 门茜儒，刘林. 人工智能赋能医学教育现代化的应用挑战与现实路径[J]. 医学教育研究与实践，2025，33(5)：641-646. DOI: 10.13555/j.cnki.c.m.e.2025.05.006.

[5] 梁爽，池元伟，孟巍，等. 住院医师人工智能素养现状及其影响因素分析：基于问卷调查的实证研究[J]. 牡丹江医科大学学报，2025，46(3)：140-144. DOI: 10.13799/j.cnki.mdjyxyxb.2025.03.013.

[6] 俞立军. WB公司数据标注业务外包质量提升对策研究[D]. 呼和浩特：内蒙古工业大学，2024.

[7] 刘江，聂秋实，王星月，等. 生成式人工智能背景下的弱监督学习教学法探索[J]. 计算机教育，2025(1)：198-203，209. DOI: 10.3969/j.issn.1672-5913.2025.01.043.

[8] 张绍衡，李晓丹，谢玥，等. AI赋能的循证决策培训：内科规培智能化路径探索[J]. 中国医学教育技术，2025，39(6)：694-697，704. DOI: 10.13566/j.cnki.cmet.cn61-1317/g4.202506002.

[9] 徐赫，焦洋. 人工智能在临床推理教学新模式中的研究及展望[J]. 中华医学教育探索杂志，2025，24(10)：1372-1378. DOI: 10.3760/cma.j.cn116021-20250310-02066.

[10] 谭思创，胡琪康，王彬，等. 基于研究生视角的AI对医学作用初探[J]. 中国医学教育技术，2025，39(6)：790-795. DOI: 10.13566/j.cnki.cmet.cn61-1317/g4.202506017.

[11] 刘江，章晓庆. 面向非计算机专业的人工智能导论课程建设与探索[J]. 中国大学教学，2022(1)：46-51. DOI: 10.3969/j.issn.1005-0450.2022.01.009.

[12] 张传钊，陈月，王俊江，等. ChatGPT在肝胆胰腺外科临床实习教学中的应用[J]. 医学教育研究与实践，2025，33(4)：571-577. DOI: 10.13555/j.cnki.c.m.e.2025.04.017.

[13] Hoover A, Kouznetsova V, Goldbaum M. STARE: structured analysis of the retina dataset [D/OL]. [2025-07-30]. https://cecas.clemson.edu/~ahoover/stare/.

[14] Staal J, Abramoff MD, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina [J]. IEEE Trans Med Imag, 2004, 23(4): 501-509. DOI: 10.1109/TMI.2004.825627.

[15] Jin K, Huang X, Zhou J, et al. FIVES: a fundus image dataset for artificial intelligence based vessel segmentation [J]. Sci Data, 2022, 9: 475. DOI: 10.1038/s41597-022-01564-3.

[16] Zhang Y, Higashita R, Zeng L, et al. Online Bayesian approximation based uncertainty aware model for ophthalmic image segmentation [J]. IEEE J Biomed Health Inform, 2026, 30(3): 2368-2380. DOI: 10.1109/JBHI.2025.3593983.

（收稿日期：2025-12-29）
（本文编辑：邬小梅）