

· 论 著 ·

人工智能 ChatGPT-4V 在乳腺超声病灶良恶性鉴别中的诊断效能

贺玉卿¹, 吴梓政¹, 郭 帅¹, 高庆壮², 李 慧², 杨百胜²

1. 秦皇岛市第一医院超声医学科, 河北 秦皇岛 066001;

2. 秦皇岛市卫生学校, 河北 秦皇岛 066001

[摘要] 目的: 评估 ChatGPT-4V 在乳腺超声病灶良恶性判读中的诊断效能, 并与低年资及高年资医师进行比较, 探讨其辅助诊断的可行性。方法: 回顾并纳入 2024 年 1 月—2025 年 6 月秦皇岛市第一医院乳腺病变患者, 以病理学检查结果为金标准, 由 ChatGPT-4V、2 名低年资 (3~5 年工作经验) 及 2 名高年资医师 (>10 年工作经验), 采用盲法独立判读超声图像。记录灵敏度、特异度、准确度、受试者工作特征 (receiver operating characteristic curve, ROC) 曲线的曲线下面积 (area under curve, AUC), 并以乳腺影像报告和数据库 (Breast Imaging Reporting and Data System, BI-RADS) 为标准评估形状、边界、回声类型、后方回声、钙化特征识别准确度。采用 McNemar 检验比较准确度, DeLong 检验比较 AUC, 临床决策曲线评估净收益。结果: ChatGPT-4V 诊断效能接近低年资医师 (准确度 $P>0.05$), 但低于高年资医师 ($P<0.05$), 临床决策曲线显示低阈值净收益接近低年资医师。ChatGPT-4V 与低年资医师比较, 在回声类型 ($P=0.012$)、后方回声 ($P=0.018$) 方面的识别准确度显著更低, 钙化特征识别差异无统计学意义 ($P=1.000$); ChatGPT-4V 与高年资医师比较, 在形状、边界、回声类型、后方回声及钙化所有超声特征的识别上均显著不足 ($P<0.05$)。ChatGPT-4V 误判 24 例 (16.0%), 恶性误为良性多见于边界光整, 良性误为恶性多见于不规则形状。结论: ChatGPT-4V 接近低年资医师效能, 适合基层筛查辅助, 但在复杂特征识别方面需改进, 未来可继续优化以提升临床应用价值。

[关键词] 乳腺癌; 乳腺结节; 大型语言模型; 人工智能; 超声; 诊断性能

中图分类号: R737.9; R445.1 文献标志码: A

DOI: 10.19732/j.cnki.2096-6210.2026.01.008

基金项目: 河北省医学科学研究课题计划 (20231893); 秦皇岛市科学技术研究与
发展计划 (202301A199)。

利益冲突: 作者声明无利益冲突。

伦理批件: 2023KZ041。

知情同意: 豁免。

引用本文: 贺玉卿, 吴梓政, 郭 帅, 等. 人工智能 ChatGPT-4V 在乳腺超声病灶良恶性鉴别中的诊断效能 [J]. 肿瘤影像学, 2026, 35(1): 57-63.

Funding: Medical Science Research Project of Hebei (20231893); Qinhuangdao S&T Plan Program (202301A199).

Conflicts of interest: authors declare no conflicts of interest.

Ethical approval: 2023KZ041.

Informed consent: exemption.

Cite this article: HE Y Q, WU Z Z, GUO S, et al. Diagnostic performance of artificial intelligence ChatGPT-4V in differentiating benign and malignant breast lesions on ultrasound [J]. Oncoradiology, 2026, 35(1): 57-63.

Diagnostic performance of artificial intelligence ChatGPT-4V in differentiating benign and malignant breast lesions on ultrasound HE Yuqing¹, WU Zizheng¹, GUO Shuai¹, GAO Qingzhuang², LI Hui², YANG Baisheng² (1. Department of Ultrasound, The First Hospital of Qinhuangdao, Qinhuangdao 066001, Hebei Province, China; 2. Qinhuangdao Health School, Qinhuangdao 066001, Hebei Province, China)

Correspondence to: WU Zizheng E-mail: wzzh890415@163.com

[Abstract] Objective: To evaluate the diagnostic performance and feature recognition accuracy of ChatGPT-4V in classifying benign and malignant breast lesions on ultrasound, compared with junior and senior physicians, and to explore its feasibility as an auxiliary diagnostic tool. **Methods:** A retrospective study included patients with breast lesions from The First Hospital of Qinhuangdao between January 2024 and June 2025. With pathological examination results as the gold standard, ChatGPT-4V, two junior physicians (3-5 years of experience), and two senior physicians (>10 years of experience) independently interpreted ultrasound images in a blinded manner. Sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC) were recorded. The accuracy of identifying ultrasound features, including shape, margin, echo pattern, posterior acoustic features, and calcifications, were evaluated against

the criteria of the Breast Imaging Reporting and Data System (BI-RADS). The McNemar test was used to compare diagnostic accuracy among different interpreters, the DeLong method was employed to compare AUC values, and decision curve analysis (DCA) was performed to assess the net benefit across varying threshold probabilities. **Results:** ChatGPT-4V's diagnostic performance was comparable to that of junior physicians (accuracy $P>0.05$) but inferior to senior physicians ($P<0.05$). DCA showed similar net benefit to junior physicians at low thresholds. Compared with junior physicians, ChatGPT-4V had significantly lower accuracy in identifying echo pattern ($P=0.012$) and posterior features ($P=0.018$), with no statistical difference in calcification recognition ($P=1.000$). Compared with senior physicians, ChatGPT-4V showed significantly insufficient accuracy in recognizing all ultrasound features, including shape, margin, echo pattern, posterior features, and calcification ($P<0.05$). ChatGPT-4V misdiagnosed 24 cases (16.0%), with malignant-to-benign errors often linked to smooth margins and benign-to-malignant errors to irregular shapes. **Conclusion:** ChatGPT-4V demonstrates diagnostic performance close to junior physicians, making it a potential auxiliary tool for breast ultrasound screening in primary care. However, its limitations in complex feature recognition require improvement through targeted optimization to enhance clinical utility.

[**Key words**] Breast cancer; Breast nodule; Large language model; Artificial intelligence; Ultrasound; Diagnostic performance

乳腺癌是全球女性最常见的恶性肿瘤,其早期诊断对提高患者生存率和改善患者预后至关重要^[1]。乳腺超声检查因其无创、经济和实时成像等优势,成为乳腺病灶良恶性鉴别的重要手段^[2]。通过分析病灶的形状、边界、回声类型、钙化等特征,结合第6版乳腺影像报告和数据系统(Breast Imaging Reporting and Data System, BI-RADS),超声检查可为临床诊断提供可靠依据^[3]。然而,超声诊断高度依赖于医师的工作经验,低年资医师因技术不足易出现误判,高年资医师虽诊断准确度较高,但面临高强度工作压力,难以满足大规模筛查需求^[2]。因此,开发智能化辅助诊断工具以提升乳腺超声的诊断效能和一致性具有重要临床价值。

近年来人工智能(artificial intelligence, AI)技术在医学影像领域的应用取得显著进展,深度学习模型如卷积神经网络(convolutional neural network, CNN)在乳腺超声病灶分类中展现出接近或优于影像科医师的性能^[4]。然而,传统AI模型需大量标注数据以及训练流程复杂,限制了其在资源有限环境中的推广。ChatGPT-4V作为多模态大型语言模型,具备图像处理和自然语言理解能力,可通过优化提示语直接分析超声图像,无需专门训练,展现出潜在的临床应用前景^[5]。尽管ChatGPT-4V在胸部X线、计算机断层成像(computed tomography, CT)等影像学诊断中已有初步探索,其在乳腺超声良恶性判读中的诊断效能及与不同经验水平医师的比较研究仍处于空白。

本研究基于2024年1月—2025年6月秦皇岛市第一医院150例乳腺病变患者的超声图像,以病理学检查结果为金标准,旨在评估ChatGPT-4V在乳腺超声病灶良恶性鉴别中的灵敏度、特异度等,并与低年资和高年资医师比较,探讨其作为辅助诊断工具的可行性。本研究将为AI在乳腺超声诊断中的应用提供实证依据,并为提示语优化和模型改进提供参考。

1 资料和方法

1.1 研究对象

回顾并纳入2024年1月—2025年6月于秦皇岛市第一医院行乳腺超声检查并经病理学检查确诊的乳腺病变患者。纳入标准:①年龄18~80岁;②乳腺超声发现明确病灶;③具有完整病理学结果(经穿刺活检或手术切除确诊);④超声图像质量清晰,满足分析要求。排除标准:①病理学检查结果不明确;②图像存在严重伪影或分辨率不足;③合并其他乳腺恶性肿瘤或全身转移性疾病。

本研究经医院伦理委员会批准(2023KZ041),伦理审查文件明确通过匿名化流程:所有患者数据均移除姓名、身份证号、住院号等可识别身份的信息,仅保留病灶相关的超声图像及病理学特征等医学数据,且图像预处理过程中进一步裁剪无关区域,确保无法通过任何信息追溯至个体。

1.2 研究设计

本研究为回顾性诊断试验研究，旨在评估ChatGPT-4V在乳腺超声病灶良恶性判读中的诊断效能，并与低年资及高年资医师比较。以病理学检查结果为金标准，采用盲法设计，ChatGPT-4V及医师在不知晓病理学检查结果的情况下独立判读超声图像。

1.3 数据收集与图像预处理

超声检查采用法国Supersonic Imagine公司的AixPlorer诊断仪，配备4~15 MHz高频线阵探头，图像存储为DICOM或JPEG格式，分辨率不低于512×512像素。每例患者病灶选取最具代表性的横切面和纵切面图像各1张，共300张图像。图像预处理：①匿名化处理，移除患者身份信息；②裁剪无关区域，保留病灶及周围2 cm组织；③统一图像对比度和亮度，确保判读一致性。该操作基于临床实践中超声图像常因设备参数、操作手法差异存在亮度/对比度波动的特点，调整目的为减少非病灶因素干扰（如背景噪声、设备差异），且调整幅度严格控制临床常规阅片可接受范围内——以病灶的形状、边界、回声类型等核心诊断特征清晰可辨为标准，未对病灶本身的原始特征进行实质性修改。临床资料（年龄、临床表现）由电子病历系统提取，病理学检查结果由病理科2名高级医师复核确认。

1.4 判读流程

使用国内具备相关运营资质与授权的平台接入OpenAI 2025版ChatGPT-4V服务，该平台通过技术手段，在符合国家网络安全与数据合规要求的前提下，完成与ChatGPT-4V的对接工作，确保研究过程中数据传输与使用的合法性与安全性。具体操作如下，研究人员在该指定平台界面，通过API接口输入经过预处理的图像。输入时，统一使用规范的提示语：“分析以下乳腺超声图像，判断病灶为良性或恶性，提供恶性概率（0%~100%），并基于形状、边界、回声类型、后方回声、钙化特征说明依据。”为保证结果可靠性，针对每例患者图像，按照上述流程重复判读3次，随后取恶性概率的均值作为最终输出结果。

医师判读选取2名低年资医师（从业3~5年，

乳腺超声诊断500~1 000例）和2名高年资医师（从业>10年，诊断>5 000例）参与判读。医师在专用工作站上独立阅片，盲法评估每例病灶的良恶性，并提供恶性概率（0%~100%），依据美国放射学会（American College of Radiology, ACR）BI-RADS第6版指南^[6]，每名医师判读完成后记录诊断结果及依据（形状、边界等特征）。2名同资历医师的诊断意见不一致时，商议后达成共识。

1.5 观察指标

主要观察指标：ChatGPT-4V、低年资医师及高年资医师的灵敏度、特异度、准确度、阳性预测值、阴性预测值及受试者工作特征（receiver operating characteristic curve, ROC）曲线的曲线下面积（area under curve, AUC）。次要观察指标：①各组在超声特征（形状、边界、回声类型、后方回声、钙化）识别上的准确度，以病理学检查结果或BI-RADS指南为标准；②临床决策曲线（decision curve analysis, DCA）的净收益。

1.6 统计学处理

采用SPSS 26.0和R 4.5.1软件分析数据，所有连续变量先通过Shapiro-Wilk检验进行正态性检验、Levene检验进行方差齐性检验，满足正态性和方差齐性的连续变量 $\bar{x}\pm s$ 表示，两组间比较采用独立样本 t 检验，不满足正态性或方差齐性的连续变量以 $M(P_{25}, P_{75})$ 表示，两组间比较采用Wilcoxon秩和检验，多组间比较使用Kruskal-Wallis秩和检验；分类变量以 $n(\%)$ 表示，组间比较采用 χ^2 检验；灵敏度、特异度等诊断效能指标计算95% CI，AUC的比较采用DeLong检验，不同判读主体的准确度比较采用配对设计的McNemar检验，通过DCA评估各组在不同阈值概率下的净收益，所有检验均为双侧， $P<0.05$ 为差异有统计学意义。

2 结 果

2.1 患者一般资料与病理学类型

共纳入150例患者，患者年龄20~75岁，平均（45.6±12.3）岁；病理学检查结果显示良性病灶75例（50%，包括纤维腺瘤、囊肿等），恶性病灶75例（50%，包括浸润性导管癌、导管内癌

等)。150例乳腺病变患者的病理学类型见表1, 研究流程见图1。

表1 150例乳腺病变患者的病理分布

Tab.1 Pathological distribution of 150 patients with breast lesions

病理学类型	n (%)
良性病灶	75 (50.0)
纤维腺瘤	40 (26.7)
囊肿	20 (13.3)
乳腺增生	10 (6.7)
其他良性病灶	5 (3.3)
恶性病灶	75 (50.0)
浸润性导管癌	50 (33.3)
导管内癌	15 (10.0)
其他恶性病灶	10 (6.7)
总计	150 (100.0)

2.2 诊断效能比较

ChatGPT-4V在乳腺超声病灶良恶性判读中的诊断效能接近低年资医师,但在灵敏度和准确度方面低于高年资医师。ChatGPT-4V的AUC为0.76 (95% CI 0.66~0.85),与低年资医师的AUC [0.79 (95% CI 0.70~0.88)]比较,差异无统计学意义 ($Z=0.856, P=0.388$),但显著低于高年资医师的AUC [0.86 (95% CI 0.78~0.93)],差异有统计学意义 ($Z=2.234, P=0.025$)。DCA曲线显示ChatGPT-4V在低阈值区间净收益接近低年资医师,具体诊断效能指标及ROC曲线、DCA曲线见表2和图2。

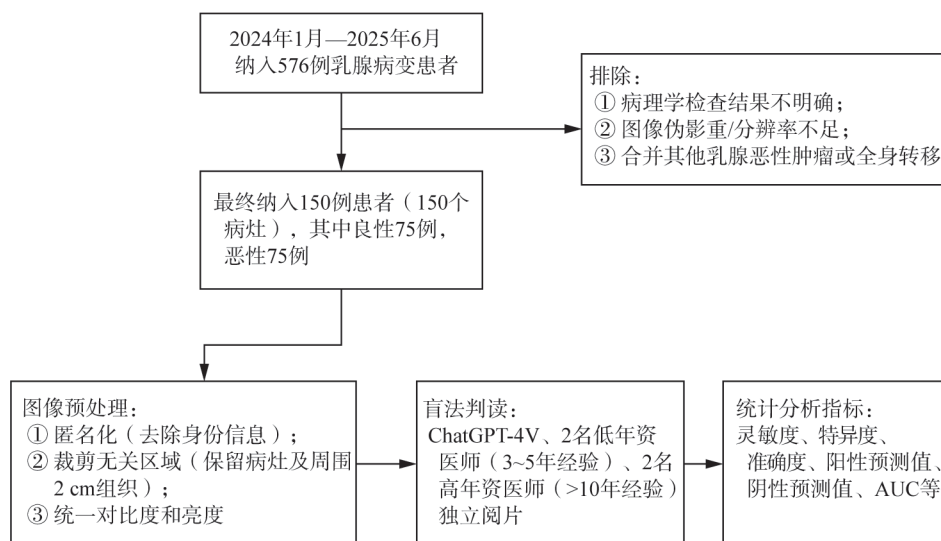


图1 研究流程图

Fig.1 Flow chart of the study

表2 ChatGPT-4V与医师在乳腺超声病灶良恶性判读中的诊断效能

Tab.2 Diagnostic performance of ChatGPT-4V and physicians in differentiating benign and malignant breast lesions on ultrasound

组别	灵敏度 (95% CI) /%	特异度 (95% CI) /%	准确度 (95% CI) /%	阳性预测值/%	阴性预测值/%	AUC	P值
ChatGPT-4V	72.0 (60.8~81.6)	84.0 (74.1~91.0)	78.0 (70.7~84.2)	82.1	74.5	0.76	—
低年资医师	74.7 (63.5~83.9)	88.0 (79.0~94.1)	81.3 (74.2~87.1)	86.2	77	0.79	0.388
高年资医师	89.3 (80.6~95.0)	92.0 (83.8~96.9)	90.7 (84.8~94.9)	90.5	90	0.86	0.025

注: P值为与ChatGPT-4V比较,准确度采用McNemar检验, AUC采用DeLong检验。

2.3 超声特征识别准确度

ChatGPT-4V、低年资医师和高年资医师在超声特征 (形状、边界、回声类型、后方回声、钙化) 识别上的准确度,以ACR BI-RADS第6版指南^[6]为标准, ChatGPT-4V在回声类型 ($P=$

0.012)和后方回声 ($P=0.018$)识别上低于低年资医师,在所有特征上低于高年资医师 ($P=0.001\sim 0.003$),钙化识别差异无统计学意义 ($P>0.05$,表3)。

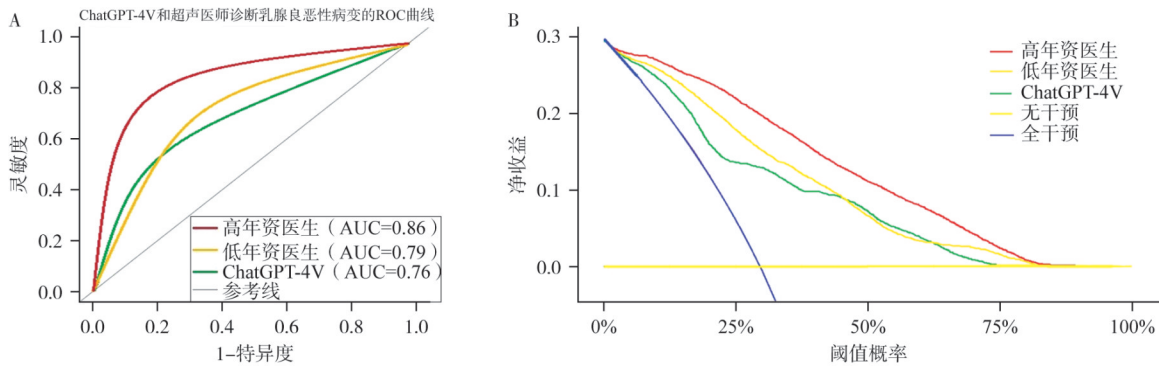


图2 ChatGPT-4V和超声医师诊断乳腺良恶性病变的ROC曲线 (A) 和DCA曲线 (B)

Fig.2 ROC curves (A) and DCA curves (B) of ChatGPT-4V and ultrasound physicians in diagnosing benign and malignant breast lesions

表3 ChatGPT-4V与医师在乳腺超声病灶特征识别准确度比较

Tab.3 Comparison of accuracy in identifying features of breast ultrasound lesions between ChatGPT-4V and physicians

特征	准确度/%			P值*	P值#
	ChatGPT-4V	低年资医师	高年资医师		
形状	62.7	72.0	81.3	0.048	0.001
边界	68.0	74.7	88.0	0.146	0.002
回声类型	52.0	68.0	81.3	0.012	<0.001
后方回声	55.3	72.0	78.7	0.018	0.003
钙化	82.7	82.7	88.0	1.000	0.134

*: ChatGPT-4V与低年资医师比较, #: ChatGPT-4V与高年资医师比较, P值用McNemar检验。

2.4 误判患者分析

ChatGPT-4V共误判24例(16.0%),包括恶性误为良性14例(18.7%)和良性误为恶性10例(13.3%)。恶性误为良性患者多见于边界光整或

无微钙化的病灶,良性误为恶性患者多见于不规则形状或伴钙化。低年资医师误判18例(12.0%),高年资医师误判9例(6.0%)。误判详情见表4, ChatGPT-4V判读例图见图3。

表4 ChatGPT-4V在乳腺超声病灶良恶性判读中的误判患者分析

Tab.4 Analysis of misdiagnosed cases by ChatGPT-4V in differentiating benign and malignant breast lesions on ultrasound

误判类型	例数	常见特征	低年资医师判读正确例数	高年资医师判读正确例数
恶性误为良性	14	边界光整、无微钙化	11	13
良性误为恶性	10	不规则形状、伴钙化	9	10

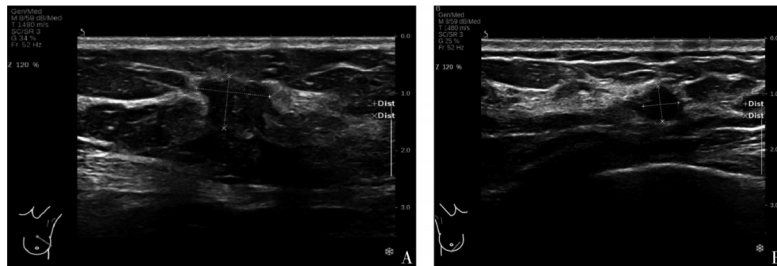


图3 ChatGPT-4V判读例图

Fig.3 Example images interpreted by ChatGPT-4V

A: ChatGPT-4V判读为恶性,恶性概率85%,依据形状不规则,边界不光整,低回声,后方回声衰减,伴微钙化。病理学检查结果示(左侧乳腺)浸润性导管癌。B: ChatGPT-4V判读为良性,恶性概率20%,依据形状不规则,边界光整,低回声,后方回声无明显衰减,无微钙化。病理学检查结果示(右乳腺)浸润性导管癌。

3 讨 论

本研究评估了 ChatGPT-4V 在乳腺超声病灶良恶性判读中的诊断效能及超声特征识别能力, 并与低年资及高年资医师比较。结果表明, ChatGPT-4V 在乳腺超声病灶良恶性判读中的诊断效能接近于低年资医师, 但与高年资医师相比仍存在差距, 尤其在灵敏度及整体准确度方面。这与文献中关于 AI 在医学影像诊断中的表现一致。例如, 有学者^[7]建立了一种基于深度学习的乳腺超声 AI 模型, 其诊断性能与中级医师相当, 但低于经验丰富的影像科医师, 原因在于 AI 模型缺乏对复杂患者的综合判断能力。本研究中, ChatGPT-4V 作为通用视觉-语言模型, 未针对乳腺超声进行专门优化, 可能限制了其在细微特征识别上的表现, 而高年资医师凭借丰富经验, 能更准确地判读非典型病灶。

DCA 分析显示, ChatGPT-4V 在低风险阈值区间的净收益与低年资医师相近, 提示其在初步筛查中具有潜在价值。既往研究^[8-9]表明, AI 辅助系统在乳腺超声筛查中可减少不必要的活检, 尤其在低阈值场景下表现优异。这表明 ChatGPT-4V 可作为低年资医师的辅助工具, 特别是在医师资源有限的基层医疗机构, 可协助提高诊断效率。然而, 其在高风险患者中灵敏度不足, 可能导致恶性病灶漏诊, 限制了其独立应用的可靠性。

基层医疗机构常面临医师资源不足、低年资医师占比高的问题, 结合本研究结果, ChatGPT-4V 诊断效能接近低年资医师, 提示其可承担基础初筛任务, 为低年资医师提供参考, 弥补其对复杂特征的识别不足, 这未来需通过前瞻性研究验证该模式的实际效能, 为基层提供更具体的应用方案。

以 BI-RADS 指南依据标准, ChatGPT-4V 识别超声特征(如形状、边界、回声类型、后方回声、钙化)的能力总体上逊于高年资医师, 部分特征(如回声类型、后方回声)甚至不及低年资医师, 但在钙化识别方面表现较好。这可能与钙化特征在超声

图像中的高对比度有关, 易被视觉模型检测。研究^[10]表明, 专门训练的 AI 模型在乳腺超声微钙化检测中表现出色, 而通用模型对复杂特征(如回声类型)的判读能力较弱, 可能是因为缺乏针对性训练数据。本研究中, ChatGPT-4V 在回声类型和后方回声方面识别不足, 可能源于其对超声图像声学特性的理解有限, 而这些特征需要结合 BI-RADS 标准^[6]和临床经验进行综合分析。

高年资医师在所有特征识别上均表现优异, 反映了其系统培训和长期临床经验的的优势。Dan 等^[11]指出, 经验丰富的医师通过结合 BI-RADS 指南和临床背景, 能更准确地识别复杂特征, 如不均匀回声或后方衰减。ChatGPT-4V 的特征识别能力虽接近于低年资医师, 但在处理非典型患者时仍需改进, 未来可通过针对性训练提升其性能。

ChatGPT-4V 的误判患者主要包括恶性病灶误判为良性和良性病灶误判为恶性, 前者多见于边界光整或无微钙化的病灶, 后者多见于不规则形状或伴钙化的病灶。低年资医师的误判率低于 ChatGPT-4V, 高年资医师误判最少。这与既往研究^[12]中关于 AI 误判模式的结果一致。Gu 等^[13]指出, AI 对不规则形状的良性病灶可能因形状特征的过度敏感而误判为恶性。高年资医师的低误判率得益于其综合分析能力, 包括结合患者临床信息(如年龄、家族史)和 BI-RADS 指南。研究^[14-15]表明, 经验丰富的医师通过整合超声和临床特征, 可显著降低误诊率。ChatGPT-4V 未输入临床信息, 仅依赖图像特征, 可能导致误判增加。这提示未来需开发多模态 AI 模型, 整合超声图像和临床数据以提高判读准确度。

研究存在一定局限性: ① 样本量与单中心局限, 150 例单中心样本可能受地域患者特征(如病理学类型构成)及超声设备单一性影响, 结论外推至其他中心时需谨慎, 未来需多中心大样本研究验证。② 与专业 AI 模型的定位差异, 本研究未纳入专业乳腺超声 AI 模型对比, 专业模型经专项训练后在复杂特征方面识别可能表现更优, 但依赖标注数据且部署成本高; 而 ChatGPT-4V 作为通用模型, 其优势在于无需专项训练即

可快速应用,更适合基层轻量化筛查,两者适用场景不同(精准诊断vs初步筛查),未来需开展头对头研究明确效能边界。

总之,ChatGPT-4V在乳腺超声病灶判读中表现出接近于低年资医师的诊断效能,适合作为基层医疗的辅助工具,尤其在初步筛查和钙化检测中具有应用潜力。然而,其在复杂特征识别和恶性病灶检测中的不足限制了独立应用价值。未来通过针对性优化和多模态整合,ChatGPT-4V有望在乳腺超声诊断中发挥更大的作用,为临床提供更高效的支持。

第一作者:

贺玉卿 (ORCID: 0009-0002-2840-4798), 硕士, 主治医师。

通信作者:

吴梓政 (ORCID: 0000-0001-8253-0352), 博士, 副主任医师, E-mail: wzzh890415@163.com。

作者贡献声明:

贺玉卿: 设计论文框架, 起草论文; 吴梓政: 研究过程实施; 郭帅: 数据收集; 高庆壮: 指导撰写文章。李慧、杨百胜: 统计学分析, 绘制图表。

[参 考 文 献]

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2021, 71(3): 209-249.
- [2] GUO Y L, LI N, SONG C H, et al. Artificial intelligence-based automated breast ultrasound radiomics for breast tumor diagnosis and treatment: a narrative review [J]. *Front Oncol*, 2025, 15: 1578991.
- [3] YAN H J, DAI C C, XU X J, et al. Using artificial intelligence system for assisting the classification of breast ultrasound glandular tissue components in dense breast tissue [J]. *Sci Rep*, 2025, 15(1): 11754.
- [4] LEE J, KIM W H, KIM J, et al. Efficacy of a real-time artificial intelligence ultrasound system with computer-aided detection and diagnosis for breast cancer: a feasibility study [J]. *J Breast Cancer*, 2025, 28(3): 206-214.
- [5] MAHANT S S, VARMA A R. Artificial intelligence in breast ultrasound: the emerging future of modern medicine [J]. *Cureus*, 2022, 14(9): e28945.
- [6] American College of Radiology. ACR BI-RADS atlas: breast imaging reporting and data system (6th edition) [M]. Reston, VA: American College of Radiology, 2021: 42-98.
- [7] XIANG H L, WANG X, XU M, et al. Deep learning-assisted diagnosis of breast lesions on US images: a multivendor, multicenter study [J]. *Radiol Artif Intell*, 2023, 5(5): e220185.
- [8] LIU H X, CUI G Z, LUO Y, et al. Artificial intelligence-based breast cancer diagnosis using ultrasound images and grid-based deep feature generator [J]. *Int J Gen Med*, 2022, 15: 2271-2282.
- [9] 王 琪, 党晓智, 许 磊, 等. 超声在乳腺癌筛查中的应用现状与未来 [J]. *中华医学超声杂志(电子版)*, 2024, 21(4): 429-433.
- [9] WANG Q, DANG X Z, XU L, et al. Current status and future perspectives of application of ultrasound in breast cancer screening [J]. *Chin J Med Ultrasound Electron Ed*, 2024, 21(4): 429-433.
- [10] PESAPANE F, TRENTIN C, FERRARI F, et al. Deep learning performance for detection and classification of microcalcifications on mammography [J]. *Eur Radiol Exp*, 2023, 7(1): 69.
- [11] DAN Q, XU Z T, BURROWS H, et al. Diagnostic performance of deep learning in ultrasound diagnosis of breast cancer: a systematic review [J]. *NPJ Precis Oncol*, 2024, 8(1): 21.
- [12] 何奕宗, 姚振强, 何小娜, 等. 具备视觉功能的ChatGPT对乳腺超声图像病变的识别能力和诊断价值初探 [J]. *中国超声医学杂志*, 2025, 41(1): 13-16.
- [12] HE Y Z, YAO Z Q, HE X N, et al. Preliminary exploration of ChatGPT with vision in the recognition and diagnostic value of breast ultrasound lesions [J]. *Chin J Ultrasound Med*, 2025, 41(1): 13-16.
- [13] GU Y, XU W, LIN B, et al. Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study [J]. *Insights Imaging*, 2022, 13(1): 124.
- [14] QIAN X J, PEI J, ZHENG H, et al. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning [J]. *Nat Biomed Eng*, 2021, 5(6): 522-532.
- [15] 程妙仙, 曾令红, 吴 忧, 等. 人工智能与大数据在超声医学实践中的应用进展 [J]. *肿瘤影像学*, 2023, 32(1): 78-82.
- [15] CHENG M X, ZENG L H, WU Y, et al. Application progress of artificial intelligence and big data in ultrasound medicine practice [J]. *Oncoradiology*, 2023, 32(1): 78-82.

(收稿日期: 2025-08-15 修回日期: 2026-01-05)