DOI: 10.16781/j.CN31-2187/R.20250580

·专题报道·

# 人工智能辅助乳腺超声区分良恶性肿块并减少非必要穿刺活检的回顾性临床研究

徐　倩，刁雪红*，杜　宇，杨雪婷，柳　燕，吴婷婷
上海交通大学医学院附属第一人民医院超声医学科，上海 200080

［摘要］ **目的**　评估人工智能（AI）辅助超声诊断系统在乳腺肿块中的诊断价值，重点分析其在优化乳腺影像报告和数据系统（BI-RADS）分类及减少非必要穿刺活检方面的作用。**方法**　回顾性纳入 222 例患者共 243 枚经病理确诊的乳腺肿块，由低年资医师先进行常规超声检查（US）并判读 BI-RADS 分类，再使用 AI 系统采集图像并依据 BI-RADS 标准进行自动化分类。将 US 和 AI 评估的 BI-RADS 2～3 类视为良性、4～5 类视为恶性，以病理结果为金标准，比较 US 与 AI 对乳腺肿块良恶性的诊断效能。**结果**　病理结果显示，243 枚乳腺肿块中恶性 126 枚（51.85%）、良性 117 枚（48.15%）。ROC 曲线分析显示，AI 诊断乳腺肿块良恶性的 AUC 为 0.874，高于 US 的 0.839（$P<0.001$）。AI 诊断乳腺肿块良恶性的特异度（79.49%）和阳性预测值（83.33%）均优于 US（分别为 70.94%、78.21%），但灵敏度（95.24%）和阴性预测值（93.94%）略低于 US（分别为 96.83%、95.40%）。在 BI-RADS 4a 类肿块中，AI 的诊断准确度（73.91%）高于 US（57.50%）。AI 将 15 枚 US BI-RADS 4a 类肿块降级，其中 13 枚（86.67%）为良性，从而避免了穿刺活检；但出现 2 枚（13.33%）假阴性。**结论**　AI 辅助乳腺超声诊断系统可提高低年资医师对乳腺肿块的诊断效能，并有助于对 US BI-RADS 4a 类肿块降级，从而减少不必要的穿刺活检。然而，AI 在识别非典型或隐匿性病变方面仍存在局限性，需结合临床信息综合判断。

［关键词］ 超声检查；人工智能；乳腺肿块；乳腺穿刺
［引用本文］ 徐倩，刁雪红，杜宇，等. 人工智能辅助乳腺超声区分良恶性肿块并减少非必要穿刺活检的回顾性临床研究［J］. 海军军医大学学报，2026，47（1）：46-52. DOI: 10.16781/j.CN31-2187/R.20250580.

## Artificial intelligence-assisted breast ultrasound for discriminating malignant and benign breast masses and reducing unnecessary biopsies: a retrospective clinical study

XU Qian, DIAO Xuehong*, DU Yu, YANG Xueting, LIU Yan, WU Tingting
Department of Ultrasound, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China

［**Abstract**］ **Objective**　To evaluate the diagnostic value of an artificial intelligence (AI)-assisted breast ultrasound system in the assessment of breast masses, with a focus on analyzing its role in optimizing Breast Imaging Reporting and Data System (BI-RADS) classification and reducing unnecessary biopsies. **Methods**　A total of 243 pathologically confirmed breast masses from 222 patients were retrospectively included. Conventional ultrasonograpny (US) and BI-RADS classification were performed by junior physicians, and an AI system was used to collect images and perform automatic classification according to BI-RADS standards. BI-RADS 2-3 assessed by US and AI were regarded as benign, while BI-RADS 4-5 were defined as malignant. Pathological results were used as the gold standard to compare the diagnostic efficacy of US and AI. **Results**　Pathological examination revealed 126 (51.85%) malignant and 117 (48.15%) benign lesions among the 243 breast masses. Receiver operating characteristic (ROC) curve analysis showed that the AI system achieved an area under curve (AUC) of 0.874 for discriminating malignant and benign breast masses, significantly higher than the 0.839 of US ($P<0.001$). The specificity and positive predictive value of AI were 79.49% and 83.33%, respectively, which were better than the 70.94% and 78.21% of the US, whereas the sensitivity (95.24% vs 96.83%) and negative predictive value (93.94% vs 95.40%)

［收稿日期］ 2025-08-26　［接受日期］ 2025-12-08
［基金项目］ 国家自然科学基金（82071931），国家重点研发计划（2022YFC3602400）. Supported by National Natural Science Foundation of China (82071931) and National Key Research and Development Program (2022YFC3602400).
［作者简介］ 徐　倩，初级技师. E-mail: sycskxq@163.com
*通信作者（Corresponding author）. E-mail: xuehong_d@126.com

were slightly lower. In BI-RADS 4a masses, the diagnostic accuracy of AI (73.91%) was higher than that of US (57.50%). AI downgraded 15 cases of BI-RADS 4a masses, of which 13 cases (86.67%) were benign, thus avoiding unnecessary biopsy; yet there were 2 cases (13.33%) of false negatives. **Conclusion**　This AI-assisted breast ultrasound system can improve the diagnostic performance of junior physicians and enable accurate downgrading of BI-RADS 4a masses, thereby reducing unnecessary biopsies. However, it still has limitations in identifying atypical or occult lesions and requires comprehensive judgment based on clinical information.

［**Key words**］ ultrasonoraphy; artificial intelligence; breast masses; breast biopsy

［**Citation**］ XU Q, DIAO X, DU Y, et al. Artificial intelligence-assisted breast ultrasound for discriminating malignant and benign breast masses and reducing unnecessary biopsies: a retrospective clinical study[J]. Acad J Naval Med Univ, 2026, 47(1): 46-52. DOI: 10.16781/j.CN31-2187/R.20250580.

乳腺癌是全球女性发病率最高的恶性肿瘤，其死亡率高居女性癌症相关死亡第2位，且发病年龄呈现显著年轻化趋势[1-5]。超声检查因具备无创、便捷等优势，已成为乳腺癌筛查与诊断的重要方法。然而，常规超声诊断结果高度依赖医师个人经验，不同操作者对乳腺影像报告数据系统（Breast Imaging Reporting and Data System，BI-RADS）分类的判断常存在准确性不一致、可重复性较低的问题[6]。近年来，医学与工程技术的深度融合推动了人工智能（artificial intelligence，AI）在乳腺影像诊断中的应用，基于AI的计算机辅助诊断系统在乳腺肿块鉴别诊断中展现出重要价值[7-10]。本研究采用的AI辅助诊断系统是一套基于深度学习的超声影像智能分析系统，它通过整合BI-RADS关键形态学特征与病理诊断结果，实现对乳腺肿块的智能化分析与分类。该系统具有标准化、智能化、可重复的特点，能够有效减少人为因素干扰，从而提升诊断效率与一致性[11-13]。本研究旨在评估该AI系统在辅助低年资医师进行乳腺肿块诊断中的临床应用价值，并重点探讨其在优化临床决策、减少不必要穿刺活检方面的潜在作用。

## 1　资料和方法

1.1　*研究对象*　本研究为回顾性临床研究，纳入2023年6月至2023年12月在上海交通大学医学院附属第一人民医院接受乳腺超声检查，并经穿刺活检或手术病理确诊的女性乳腺肿块患者。纳入标准：（1）经高分辨率超声检查发现乳腺肿块，由低年资超声医师（工作年限＜3年）先进行常规超声检查（conventional ultrasonography，US）并判读BI-RADS分类，再用搭载AI辅助诊断系统的设备进行图像采集及自动化分析；（2）肿块经穿刺活检或手术切除，且病理诊断明确；（3）肿块超声图像符合AI系统的识别标准；（4）临床及影像学资料完整，可进行后续分析。排除标准：（1）超声影像资料不完整或图像质量不符合分析要求；（2）病理标本无法明确诊断或取材不充分；（3）妊娠期或哺乳期患者；（4）受检者配合度差，影响图像采集质量；（5）非肿块型乳腺病变（如钙化、导管扩张等）。本研究获得我院伦理委员会审批。

1.2　*仪器与方法*　由低年资医师先进行US并判读BI-RADS分类，再使用AI系统采集图像并依据BI-RADS标准进行自动化分级。US检查采用Aplio 900超声诊断设备（日本佳能医疗系统，配备5～18 MHz高频线阵探头）。患者取仰卧位，双上肢自然外展，充分暴露双侧乳房及腋窝区域[14]。将超声诊断仪器参数调整至最佳条件，选择高频探头，将耦合剂涂抹于乳房表面及周围，保持用量适中、涂抹均匀，轻压探头，检查期间避免过度加压。采用斜切声束，自腋前线至胸骨缘行辐射状扫查后，再于第2肋至第6肋间自上而下行系列横切扫查，探查到可疑肿块后进行多切面扫查，观察肿块位置、大小、边界、形态、内部回声、后方回声、血流及钙化情况等。所有US图像均由2名不知晓病理结果的低年资医师依据BI-RADS分类标准独立评估，若出现分歧，以较高分类作为最终结果。

2名低年资医师完成US后，采用搭载VAid Breast乳腺自动检测系统的VINNO ULTIMUS 9E超声诊断设备（飞依诺科技股份有限公司，配备6.5～15 MHz高频线阵探头）采集图像并进行AI辅助诊断。患者的体位与US时一致，扫查前调节仪器增益、深度和扫查的方向以获取清晰图像。

在肿块最大径纵切面、横切面、最具恶性征象切面及血流最丰富切面上，由系统自动勾画出结节边界，若自动描绘边界与肿块实际边界不匹配，操作者可手动调整、勾画肿块边界。选择最匹配的边界后，系统自动列出肿块的各种特征（大小、深度、形态、边界、内部回声等）以及BI-RADS分类。VAid Breast乳腺自动检测系统采用多阶段串联的AI模型架构实现病灶的良恶性鉴别。首先使用轻量化的Mobile-Former网络对超声扫查部位进行快速分类，判别是否为乳腺组织；随后采用嵌入Ghost模块的PPlite-seg模型精准分割乳腺腺体组织；接着使用改进的YOLOv8模型检测核心病灶，通过C2f模块增强特征表达能力，并引入DAttention模块使模型能自适应聚焦于可疑病灶区域；然后基于DDRNet架构并结合注意力机制的语义分割模型，对检测出的病灶进行精细的前后景分割以勾画边界；最后综合提取病灶的形状、纹理等特征，依据BI-RADS标准进行自动化分级，并生成包括良恶性判断在内的结构化诊断报告与临床建议。

本研究将BI-RADS 1～3类乳腺肿块归类为良性病变，BI-RADS 4类和5类归类为恶性病变[15]。

1.3　统计学处理　采用SPSS 25.0软件进行数据分析。符合正态分布且方差齐的计量资料以$\bar{x}\pm s$表示，组间比较采用独立样本$t$检验；计数资料以频数和百分数表示，组间比较采用$\chi^2$检验或Fisher确切概率法。采用ROC曲线分析计算AUC，通过$Z$检验比较不同方法的AUC差异，并计算灵敏度、特异度、阳性预测值、阴性预测值，评价US和AI两种方法及各类超声恶性征象对乳腺肿块良恶性的诊断效能。检验水准（$\alpha$）为0.05。

## 2　结　果

2.1　患者与肿块基线特征　本研究最终纳入222例女性患者，包括243枚肿块（左侧116枚，右侧127枚）。患者年龄为17～90（54.2±15.2）岁。病理结果显示，243枚乳腺肿块中恶性126枚（51.85%）、良性117枚（48.15%）。肿块最大径为（18.91±16.83）mm，其中恶性肿块的最大径为（23.14±20.91）mm，良性肿块的最大径为（14.35±8.90）mm，差异有统计学意义（$P<0.001$）。恶性肿块的病理类型包括浸润性乳腺癌89枚，导管原位癌24枚，实体乳头状癌10枚，腺癌2枚，乳腺小叶原位癌1枚；良性肿块病理类型包括纤维腺瘤69枚，硬化性腺病29枚，导管内乳头状瘤16枚，叶状肿瘤1枚，乳腺囊肿2枚。

单因素分析结果显示，在年龄、最大径、边界、形态、内部回声、血流、钙化、后方回声、纵横比及腋下淋巴结方面，良性肿块组与恶性肿块组间差异均有统计学意义（均$P<0.05$），见表1。

2.2　各类超声恶性征象的诊断效能评价　在各类超声恶性征象中，形态不规则诊断乳腺恶性肿块的灵敏度和阴性预测值最高，分别为92.06%和78.72%；而纵横比的特异度和阳性预测值最高，均为100.00%。各类超声恶性征象的诊断效能见表2。

2.3　US与AI系统诊断效能比较　将BI-RADS 2～3类视为良性、4～5类视为恶性，以病理结果为金标准，分析US和AI对乳腺肿块良恶性的诊断效能。ROC曲线分析显示，AI的AUC为0.874，大于US的0.839，差异有统计学意义（$P<0.001$）；US诊断乳腺肿块良恶性的灵敏度、特异度、阳性预测值、阴性预测值分别为96.83%、70.94%、78.21%、95.40%，AI的灵敏度、特异度、阳性预测值、阴性预测值分别为95.24%、79.49%、83.33%、93.94%，AI诊断乳腺肿块良恶性的特异度和阳性预测值均优于US，但灵敏度和阴性预测值略低于US。见表3。

对比US与AI对不同分类乳腺肿块良恶性的诊断效能，AI在BI-RADS 4a、4b及4c类肿块诊断中均优于US。其中，在BI-RADS 4a类肿块中，AI诊断准确度（73.91%）高于US（57.50%）；对于恶性可能性更高的BI-RADS 4b类与4c类肿块，AI诊断准确度均达到100.00%，而US分别为70.37%与96.43%。见表4。

2.4　AI系统对US BI-RADS分类结果的降级效能　AI系统对US不同BI-RADS分类肿块的重分类评估结果见表5，典型图像见图1。对于US诊断为BI-RADS 4a类的肿块，AI将其中的15枚降级为BI-RADS 3类。术后病理证实，这15枚降级肿块中，13枚（86.67%）为良性（包括纤维腺瘤4枚、硬化性腺病4枚、导管内乳头状瘤5枚），从而避免了不必要的穿刺活检。然而，在降级肿块中出现2枚（13.33%）假阴性，病理结果分别为导管内乳头状瘤伴导管原位癌、导管内乳头状瘤伴周围乳腺低级别导管原位癌。此外，在AI维持诊断为4a类的21枚肿块中，有4枚（19.05%）术后病理证实为良性病变。

**表1　乳腺良性肿块组和恶性肿块组一般临床资料比较**

**Tab 1　Comparison of baseline characteristics between benign and malignant breast mass groups**

*n* (%)

| Characteristic | Benign mass $N=117$ | Malignant mass $N=126$ | $\chi^2$ value | *P* value |
|---|---|---|---|---|
| Patient age | | | 6.662 | 0.010 |
| ＜40 years | 30 (25.64) | 16 (12.70) | | |
| ≥40 years | 87 (74.36) | 110 (87.30) | | |
| Largest diameter | | | 14.290 | ＜0.001 |
| ＜2 cm | 91 (77.78) | 69 (54.76) | | |
| ≥2 cm | 26 (22.22) | 57 (45.24) | | |
| Location | | | 1.555 | 0.212 |
| Left | 51 (43.59) | 65 (51.59) | | |
| Right | 66 (56.41) | 61 (48.41) | | |
| Margin | | | 58.747 | ＜0.001 |
| Clear | 77 (65.81) | 22 (17.46) | | |
| Unclear | 40 (34.19) | 104 (82.54) | | |
| Shape | | | 21.819 | ＜0.001 |
| Regular | 37 (31.62) | 10 (7.94) | | |
| Irregular | 80 (68.38) | 116 (92.06) | | |
| Internal echogenicity | | | 8.686 | 0.003 |
| Homogeneous | 82 (70.09) | 65 (51.59) | | |
| Heterogeneous | 35 (29.91) | 61 (48.41) | | |
| Blood flow | | | 19.455 | ＜0.001 |
| Absence | 83 (70.94) | 54 (42.86) | | |
| Presence | 34 (29.06) | 72 (57.14) | | |
| Calcification | | | 19.013 | ＜0.001 |
| Absence | 97 (82.91) | 72 (57.14) | | |
| Presence | 20 (17.09) | 54 (42.86) | | |
| Posterior echo | | | 21.135 | ＜0.001 |
| Absence | 101 (86.32) | 86 (68.25) | | |
| Shadowing | 2 (1.71) | 21 (16.67) | | |
| Enhancement | 13 (11.11) | 12 (9.52) | | |
| Mixed | 1 (0.85) | 7 (5.56) | | |
| Aspect ratio | | | 18.051 | ＜0.001 |
| ＜1 | 117 (100.00) | 108 (85.71) | | |
| ≥1 | 0 | 18 (14.29) | | |
| Suspected axillary lymph node | | | 22.889 | ＜0.001 |
| Absence | 116 (99.15) | 101 (80.16) | | |
| Presence | 1 (0.85) | 25 (19.84) | | |

**表2　各类超声恶性征象对乳腺肿块良恶性的诊断效能**

**Tab 2　Diagnostic performance of various ultrasonographic malignant characteristics in discriminating benign and malignant breast masses**

% (*n*/*N*)

| Malignant characteristic | Sensitivity | Specificity | Positive predictive value | Negative predictive value |
|---|---|---|---|---|
| Unclear margin | 82.54 (104/126) | 65.81 (77/117) | 72.22 (104/144) | 77.78 (77/99) |
| Irregular shape | 92.06 (116/126) | 31.62 (37/117) | 59.18 (116/196) | 78.72 (37/47) |
| Heterogeneous | 48.41 (61/126) | 70.09 (82/117) | 63.54 (61/96) | 55.78 (82/147) |
| With blood flow | 57.14 (72//126) | 70.94 (83/117) | 67.92 (72/106) | 60.58 (83/137) |
| With calcification | 42.86 (54/126) | 82.91 (97/117) | 72.97 (54/74) | 57.40 (97/169) |
| Aspect ratio≥1 | 14.29 (18/126) | 100.00 (117/117) | 100.00 (18/18) | 52.00 (117/225) |
| Suspected axillary lymph node | 19.84 (25/126) | 99.15 (116/117) | 96.15 (25/26) | 53.46 (116/217) |
| Posterior echo | | | | |
| Shadowing | 16.67 (21/126) | 98.21 (115/117) | 91.30 (21/23) | 52.27 (215/220) |
| Enhancement | 9.52 (12/126) | 11.11 (13/117) | 48.00 (12/25) | 47.71 (104/218) |
| Mixed | 5.56 (7/126) | 99.15 (116/117) | 87.50 (7/8) | 49.36 (116/235) |

表 3　US 和 AI 对乳腺肿块良恶性的诊断效能

Tab 3　Diagnostic performance of US versus AI in discriminating benign and malignant breast masses

| Method | Sensitivity/% (*n*/*N*) | Specificity/% (*n*/*N*) | PPV/% (*n*/*N*) | NPV/% (*n*/*N*) | PLR | NLR | Accuracy/% (*n*/*N*) |
|---|---|---|---|---|---|---|---|
| US | 96.83 (122/126) | 70.94 (83/117) | 78.21 (122/156) | 95.40 (83/87) | 3.33 | 0.04 | 84.36 (205/243) |
| AI | 95.24 (120/126) | 79.49 (93/117) | 83.33 (120/144) | 93.94 (93/99) | 4.64 | 0.06 | 87.65 (213/243) |

US: Conventional ultrasonography; AI: Artificial intelligence; PPV: Positive predictive value; NPV: Negative predictive value; PLR: Positive likelihood ratio; NLR: Negative likelihood ratio.

表 4　US 与 AI 对不同 BI-RADS 分类乳腺肿块的诊断效能

Tab 4　Diagnostic performance of US and AI for breast masses of different BI-RADS categories

% (*n*/*N*)

| Method | Pathologic result | BI-RADS category | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4a | 4b | 4c | 5 |
| US | Benign | 100.00 (2/2) | 95.29 (81/85) | 42.50 (17/40) | 29.63 (16/54) | 3.57 (1/28) | 0 (0/34) |
| | Malignant | 0 (0/2) | 4.71 (4/85) | 57.50 (23/40) | 70.37 (38/54) | 96.43 (27/28) | 100.00 (34/34) |
| AI | Benign | 92.86 (13/14) | 94.12 (80/85) | 26.09 (24/92) | 0 (0/24) | 0 (0/23) | 0 (0/5) |
| | Malignant | 7.14 (1/14) | 5.88 (5/85) | 73.91 (68/92) | 100.00 (24/24) | 100.00 (23/23) | 100.00 (5/5) |

US: Conventional ultrasonography; AI: Artificial intelligence; BI-RADS: Breast Imaging Reporting and Data System.

表 5　AI 对 US BI-RADS 分类乳腺肿块的重分类评估结果

Tab 5　Reclassification of breast mass US BI-RADS categories by AI

*n*

| BI-RADS category | US result | AI downgraded | | AI unchanged | | AI upgraded | |
|---|---|---|---|---|---|---|---|
| | | Benign[a] | Malignant[a] | Benign[a] | Malignant[a] | Benign[a] | Malignant[a] |
| 2 | 2 | 0 | 0 | 2 | 0 | 0 | 0 |
| 3 | 85 | 4 | 0 | 69 | 1 | 8 | 3 |
| 4a | 40 | 13 | 2 | 4 | 17 | 0 | 4 |
| 4b | 54 | 16 | 25 | 0 | 7 | 0 | 6 |
| 4c | 28 | 1 | 21 | 0 | 5 | 0 | 1 |
| 5 | 34 | 0 | 32 | 0 | 2 | 0 | 0 |

[a]: Pathologic result. AI: Artificial intelligence; US: Conventional ultrasonography; BI-RADS: Breast Imaging Reporting and Data System.

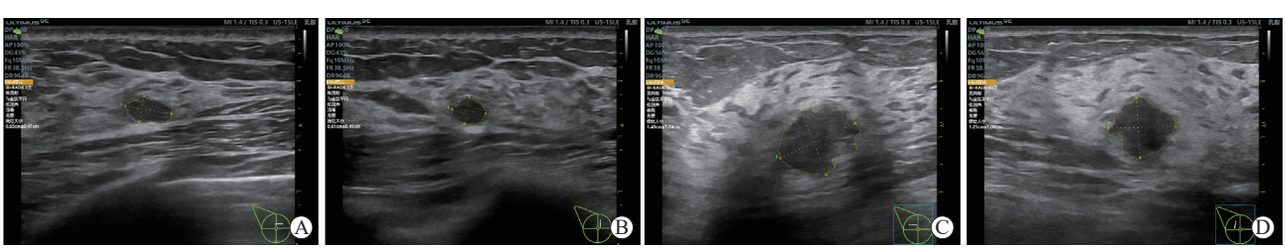


图 1　US BI-RADS 分类经 AI 降级或升级的乳腺肿块典型声像图

Fig 1　Typical sonographic images of US BI-RADS categories downgraded or upgraded by AI in breast masses

A, B: A case of US BI-RADS 4a mass that was downgraded to BI-RADS 3 by AI, with a pathological result of fibroadenoma (A: Long-axis view of the breast mass; B: Short-axis view of the breast mass); C, D: A case of US BI-RADS 3 mass that was upgraded to BI-RADS 4a by AI, with a pathological diagnosis of invasive ductal carcinoma (C: Long-axis view of the breast mass; D: Short-axis view of the breast mass). US: Conventional ultrasonography; BI-RADS: Breast Imaging Reporting and Data System; AI: Artificial intelligence.

## 3　讨　论

乳腺癌的早期精准诊断对改善患者预后及制定个体化治疗方案具有重要意义。乳腺超声是一种无创、无辐射、成本低廉的检查方法[6]，也被推荐为我国女性乳腺癌的首选筛查手段[11]。尽管超声检查广泛应用于临床，但其诊断准确性高度依赖医师经验，尤其低年资医师在 BI-RADS 分类判读中易出现主观差异，可能导致误诊或过度干预[10]。本研究通过比较 AI 辅助诊断系统与 US 对乳腺肿

块的诊断效能，证实了AI技术在提升诊断效能、优化临床决策及减少不必要穿刺活检方面具有重要价值。

本研究结果显示，AI鉴别乳腺良恶性肿块的AUC为0.874，高于US的0.839（$P<0.001$），提示AI系统在整体诊断能力上更具优势。进一步分析发现，AI在特异度（79.49% vs 70.94%）和阳性预测值（83.33% vs 78.21%）上优于US，表明AI可更精准地识别良性病变，从而降低假阳性率，减少不必要的穿刺活检。US容易产生假阳性结果[16]，而AI在降低乳腺肿块假阳性率方面具有优势，展现出AI在辅助低年资医师进行BI-RADS分类评估中的重要潜力。

US对BI-RADS 4a类肿块的处理常面临两难困境：过度穿刺（阳性率仅2%～10%）或潜在漏诊风险[17-18]。本研究显示，在BI-RADS 4a类结节中，AI诊断准确度达73.91%，显著高于US的57.50%。特别值得关注的是，AI对4a类肿块的降级诊断展现出重要临床价值。在本研究中15枚被AI降级为3类的肿块中，13枚经病理证实为良性，包括纤维腺瘤、硬化性腺病等常见良性病变。这一结果表明AI系统可有效减少BI-RADS 4a类肿块被过度分类的情况，从而避免对良性病变的不必要穿刺。此外，在AI维持4a类的肿块中，假阳性率仅19.05%，远低于US的42.50%。这一发现与Ju等[15]的研究结论一致，该研究证实乳腺AI诊断系统可避免32.5%的BI-RADS 4a类良性病变的穿刺活检。

需要说明的是，本研究采用的是一款成熟的AI辅助诊断系统，其核心算法本身并非本研究首创，本工作的主要创新点和价值在于其临床应用的即时性与集成性。与多数仍处于实验室阶段的AI模型不同，该软件直接集成于超声设备中，实现了临床工作流程中的实时AI辅助诊断，显著提升了其在真实临床环境中的可用性，这种集成化应用模式有助于提高低年资超声医师对乳腺结节良恶性诊断的准确度，同时对优化医疗资源、减轻患者心理负担及提升整体诊疗效率具有重要意义。

然而，在临床实践中应用AI辅助诊断系统时，需审慎看待其升降级决策的局限性。本研究中，AI系统降级的肿块中出现2枚（13.33%）假阴性结果，病理结果均为导管内乳头状瘤伴导管原位癌。此类病变的恶性成分往往呈局灶性、隐匿性分布，且未形成明显的肿块结构，其超声图像缺乏典型恶性特征，如形态显著不规则、毛刺征或明显微钙化。当前的AI模型主要基于对肿块宏观形态学特征（如边界、形态、回声）的深度学习，对于仅在病理层面才明确显现的“镶嵌式”癌变，其识别能力存在固有局限。这也反映出AI模型的诊断性能受限于训练数据中所涵盖的病变类型和特征多样性。若训练集中此类不典型、隐匿性癌变的样本不足，模型便难以捕捉其细微的影像学表现，这一发现为后续模型的优化指明了方向。

在AI维持BI-RADS 4a类诊断的肿块中，出现4枚（19.05%）假阳性，包括硬化性腺病2枚、复杂纤维腺瘤1枚和良性叶状肿瘤1枚，这些误诊病例均具有形态不规则或边界不清等恶性征象，AI基于形态学分析将其判读为可疑恶性，从而导致分类偏高。本研究使用的AI系统在鉴别诊断中存在以下局限性：（1）对于病理表现复杂的交界性病变，其判别能力仍有待提升；（2）在面对不典型影像特征时，系统可能过度依赖单一模式的分析策略；（3）尽管AI能够输出BI-RADS分类，但其决策过程缺乏透明性，医师难以了解“降级”或“升级”所依据的具体影像特征。基于上述发现，我们建议对所有判读为BI-RADS 3类及以上的病灶，同步启用AI系统进行分析，若医师与AI分类一致，则按常规流程处理；若出现分歧，则启动复核机制。特别是对于AI建议将4a类降级为3类的病例，必须由高年资医师复核确认降级依据充分，且无其他临床疑虑时，方可采纳降级建议、避免活检。该标准化流程可在发挥AI减少不必要活检优势的同时，通过人工复核将漏诊风险降至最低。

本研究的局限性主要包括以下几方面：（1）样本量有限，且为单中心研究，结果仅反映在我院接受进一步诊疗的人群特征，外推性尚需通过多中心、大样本研究加以验证。（2）AI系统仍存在一定的假阴性风险，尤其对于病灶较小、恶性特征不典型或伴有微浸润的恶性结节易出现误判，今后需结合病灶历史变化、分子影像等多模态数据进一步优化模型。（3）当前特征分析仍偏向单一影像学特征，例如，形态不规则虽具有较高的灵敏度（92.06%），但特异度较低（31.62%）；而纵横比、腋下淋巴结肿大、混合回声特异度高（分别为100.00%、99.15%、99.15%），灵敏度却偏低（分

别为14.29%、19.84%、5.56%），这提示未来模型应致力于深度融合多维度影像特征与临床信息，以提升综合判别能力。

综上所述，本研究表明基于深度学习的超声AI辅助诊断系统能显著提高对乳腺肿块的诊断效能，尤其在BI-RADS 4a类肿块的良恶性鉴别中价值突出，可为低年资医师提供可靠的决策支持，而且可通过对部分病变的精准降级有效减少不必要的穿刺活检。后续研究应进一步扩大样本规模、增强模型对不典型与隐匿性癌变的识别能力、开发并整合可解释性工具，推动AI技术与临床诊疗路径的系统性整合。

## ［参 考 文 献］


［1］ BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74(3): 229-263. DOI: 10.3322/caac.21834.

［2］ GIAQUINTO A N, SUNG H, MILLER K D, et al. Breast cancer statistics, 2022[J]. CA Cancer J Clin, 2022, 72(6): 524-541. DOI: 10.3322/caac.21754.

［3］ LOIBL S, POORTMANS P, MORROW M, et al. Breast cancer[J]. Lancet, 2021, 397(10286): 1750-1769. DOI: 10.1016/s0140-6736(20)32381-3.

［4］ SIEGEL R L, GIAQUINTO A N, JEMAL A. Cancer statistics, 2024[J]. CA Cancer J Clin, 2024, 74(1): 12-49. DOI: 10.3322/caac.21820.

［5］ XIONG X, ZHENG L W, DING Y, et al. Breast cancer: pathogenesis and treatments[J]. Signal Transduct Target Ther, 2025, 10(1): 49. DOI: 10.1038/s41392-024-02108-4.

［6］ BERG W A, BLUME J D, CORMACK J B, et al. Operator dependence of physician-performed whole-breast US: lesion detection and characterization[J]. Radiology, 2006, 241(2): 355-365. DOI: 10.1148/radiol.2412051710.

［7］ LÅNG K, JOSEFSSON V, LARSSON A M, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study[J]. Lancet Oncol, 2023, 24(8): 936-944. DOI: 10.1016/S1470-2045(23)00298-X.

［8］ RESCH D, GULLO R L, TEUWEN J, et al. AI-enhanced mammography with digital breast tomosynthesis for breast cancer detection: clinical value and comparison with human performance[J]. Radiol Imaging Cancer, 2024, 6(4): e230149. DOI: 10.1148/rycan.230149.

［9］ BAHL M, CHANG J M, MULLEN L A, et al. Artificial intelligence for breast ultrasound: AJR expert panel narrative review[J]. AJR Am J Roentgenol, 2024, 223(6): e2330645. DOI: 10.2214/AJR.23.30645.

［10］ SCHIAFFINO S, GIROMETTI R. Artificial intelligence in breast imaging: from theoretical promise to clinical mirage?[J]. Eur Radiol, 2025. DOI: 10.1007/s00330-025-11828-2.

［11］ HE P, CHEN W, BAI M Y, et al. Deep learning-based computer-aided diagnosis for breast lesion classification on ultrasound: a prospective multicenter study of radiologists without breast ultrasound expertise[J]. AJR Am J Roentgenol, 2023, 221(4): 450-459. DOI: 10.2214/AJR.23.29328.

［12］ MANGO V L, SUN M, WYNN R T, et al. Should we ignore, follow, or biopsy? Impact of artificial intelligence decision support on breast ultrasound lesion assessment[J]. AJR Am J Roentgenol, 2020, 214(6): 1445-1452. DOI: 10.2214/AJR.19.21872.

［13］ LAI Y C, CHEN H H, HSU J F, et al. Evaluation of physician performance using a concurrent-read artificial intelligence system to support breast ultrasound interpretation[J]. Breast, 2022, 65: 124-135. DOI: 10.1016/j.breast.2022.07.009.

［14］ 中国医师协会超声医师分会. 中国浅表器官超声检查指南[M]. 北京：人民卫生出版社，2017：326.

［15］ JU Y, ZHANG G, WAN Y, et al. Integration of AI lesion classification, age, and BI-RADS assessment to reduce benign biopsies on breast ultrasound[J]. Eur Radiol, 2025, 35(9): 5658-5670. DOI: 10.1007/s00330-025-11467-7.

［16］ BERG W A, RAFFERTY E A, FRIEDEWALD S M, et al. Screening algorithms in dense breasts: AJR expert panel narrative review[J]. AJR Am J Roentgenol, 2021, 216(2): 275-294. DOI: 10.2214/AJR.20.24436.

［17］ JATOI I, PINSKY P F. Breast cancer screening trials: endpoints and overdiagnosis[J]. J Natl Cancer Inst, 2021, 113(9): 1131-1135. DOI: 10.1093/jnci/djaa140.

［18］ GU Y, XU W, LIU T, et al. Ultrasound-based deep learning in the establishment of a breast lesion risk stratification system: a multicenter study[J]. Eur Radiol, 2023, 33(4): 2954-2964. DOI: 10.1007/s00330-022-09263-8.


［本文编辑］ 孙　岩