·医学工程·

深度学习模型辅助基层超声医师鉴别乳腺结节良恶性的诊断效能研究

容炜[1] 谷丽萍[2*]

（1.上海市浦东新区万祥社区卫生服务中心医技科，上海 201313；
2.上海交通大学医学院附属第六人民医院超声医学科，上海 200233）

【摘要】 **目的** 评估深度学习模型辅助基层超声医师鉴别乳腺结节良恶性的诊断效能。**方法** 回顾性收集2022年9月至2024年9月在上海市第六人民医院进行常规超声检查时发现乳腺结节并诊断为BI-RADS 3类和4类的女性患者，收集患者临床资料及超声图像。构建GoogLeNet、VGG11、AlexNet、ResNet-34、ViT 5种深度学习模型，并通过对比曲线下面积（area under curve, AUC）等指标选取最佳模型。采用最佳模型对超声图像生成可视化热图，由社区卫生服务中心医师在不知病理情况下对患者的超声图像进行独立阅片，两周后再次借助最佳模型生成的可视化热图进行阅片，通过准确性、灵敏度、特异度和精确度等指标，评估和比较深度学习模型阅片、超声医师独立阅片、深度学习模型辅助超声医师阅片的诊断效能。**结果** 共纳入430例乳腺结节患者，分为训练集（n=301）和验证集（n=129）。基于乳腺超声图像构建的不同深度学习模型中，ResNet-34模型在鉴别乳腺结节良恶性方面表现最佳，AUC、准确性和灵敏度在训练集和验证集中都优于其他模型，因此选择ResNet-34模型作为最佳模型。在3种阅片方案中，基层医师独立阅片的诊断效能最低，深度学习模型在准确性、灵敏度、精确度等指标上略优于基层医师，而深度学习模型辅助超声医师阅片的诊断效能最优，所有评估指标较医师独立阅片都得到了大幅度提升。从提升幅度来看，灵敏度和F1值提升最明显，分别提高34.3%和29.9%。**结论** 基于超声图像的深度学习模型能够较好地区分乳腺结节的良恶性，结合其生成的可视化热图能够大幅提升基层超声医师鉴别乳腺结节良恶性的诊断效能，可为乳腺结节患者治疗方案的选择提供有利指导信息。

【关键词】 乳腺结节；超声；深度学习；可视化；社区

【中图分类号】 R737.9; R445.1 【文献标识码】 A 【DOI】 10.11876/mimt202603007

A study on the diagnostic efficacy of deep learning models in assisting primary ultrasound physicians in differentiating benign and malignant breast nodules

Rong Wei[1], Gu Liping[2]*

(1. Medical Technology Department, Shanghai Pudong New Area Wanxiang Community Health Service Center, Shanghai 201313, China;2. Department of Ultrasound in Medicine, Sixth People's Hospital Affiliated to Medical College of Shanghai Jiaotong University, Shanghai 200233, China)

【**Abstract**】 **Objective** Evaluate the diagnostic performance of a deep learning model in assisting primary care ultrasound doctors in differentiating benign and malignant breast nodules. **Methods** This study retrospectively collected female patients who were diagnosed with BI-RADS category 3 and 4 breast nodules during routine ultrasound examinations at Shanghai Sixth People's Hospital from September 2022 to September 2024. The clinical data and ultrasound images of these patients were collected.. Five deep learning models(GoogLeNet, VGG11, AlexNet, ResNet-34, ViT) were constructed, and the optimal model was selected

收稿日期：2025-12-29

基金项目：2024年度“浦东新区卫生系统优秀社区适宜人才培养计划”（PWRs2024-14）；“市六-临港”紧密型健康联合体临床类科研项目（ynlglht202401）；上海市浦东新区万祥社区卫生服务中心院内科研项目（WX2024-01）

第一作者：容炜，男，本科，主治医师，研究方向：超声，社区常见疾病风险分层及预警

通信作者：谷丽萍，女，硕士研究生，副主任医师，研究方向：乳腺肿瘤，甲状腺肿瘤

based on metrics such as area under curve (AUC). Visual heatmaps of the ultrasound images were generated using the optimal model. Community ultrasound physicians independently reviewed the images without knowledge of the pathological results. Two weeks later, the physicians re-reviewed the images with the assistance of the generated heatmaps. Diagnostic performance was compared among three approaches: deep learning model alone, physician independent review, and deep learning-assisted physician review, using accuracy, sensitivity, specificity, and precision as metrics. **Results** A total of 430 patients with breast nodules were included and divided into a training set (n=301) and a validation set (n=129). Among the different learning models constructed based on breast ultrasound images, the ResNet-34 model performed best in predicting benign and malignant breast nodules, with AUC, accuracy, and sensitivity superior to other models in both the training and validation sets. Therefore, the ResNet-34 model was selected as the optimal model. Among the three approaches, the diagnostic efficacy of independent review by primary care physicians was the lowest. The deep learning model slightly outperformed primary care physicians in terms of accuracy, sensitivity, and precision, while the diagnostic efficacy of the deep learning model assisting ultrasound physicians in reading was the strongest, with all evaluation indicators significantly improved. In terms of the improvement rate, the sensitivity and F1 value showed the most significant increase, rising by 34.3% and 29.9% respectively. **Conclusion** Deep learning models based on ultrasound images can effectively distinguish the benign and malignant nature of breast nodules. Combined with the generated visual heatmaps, they can significantly enhance the diagnostic efficacy of assisting primary ultrasound physicians in distinguishing between benign and malignant breast nodules, providing valuable guidance for selecting treatment plans for patients with breast nodules.

【**Key words**】 Breast nodules; Ultrasound; Deep learning; Visualization; Community

乳腺癌是威胁女性健康的重大公共卫生问题，作为女性常见恶性肿瘤之一，已成为我国20~59岁女性癌症死亡的主要原因，且呈现年轻化趋势[1]。超声检查凭借无创、经济等优势，已成为基层医疗机构乳腺癌筛查的核心手段[2]，但乳腺影像报告与数据系统（Breast Imaging Reporting and Data System，BI-RADS）分级在基层实践中面临显著挑战。基层医疗机构的超声医师尤其是低年资医师，受培训不足以及经验局限的影响，对BI-RADS分级的诊断准确性有限。乳腺结节尤其是BI-RADS 3类和4类结节，其恶性风险跨度较大，存在一定的鉴别诊断困难，易导致不必要的手术或穿刺活检[1]。因此，开发一种准确、无创的乳腺结节良恶性鉴别方法至关重要。

社区作为肿瘤防控的“前沿哨点”，其诊断准确性直接决定早期干预率。研究表明，早期乳腺癌漏检可使患者5年生存率下降30%以上[3]。这一现状凸显了提升基层诊断能力的重要性和迫切性。深度学习（deep learning, DL）是人工智能的一个高级阶段，它可以利用多层神经元级联学习方式分层抽象原始数据来模拟人类神经网络[4]，然后利用从数据中获得的特征进行分类预测。同时，深度学习的可视化工具，可以显示图像中的复杂特征，因此在医学图像分析中得到了广泛的应用[5]。

本研究旨在评估深度学习可视化模型在辅助基层超声医师鉴别乳腺结节良恶性中的应用价值，并比较深度学习模型阅片、超声医师独立阅片、深度学习模型辅助超声医师阅片的诊断效能差异，为社区乳腺癌筛查提供标准化、可复制的智能辅助方案。

1 资料与方法

1.1 一般资料

回顾性收集2022年9月至2024年9月在上海市第六人民医院进行常规超声检查时发现乳腺结节并诊断为BI-RADS 3类和4类的女性患者[6]。纳入标准：①有最终的外科手术或粗针穿刺病理诊断结果；②活检或手术前2周内进行乳腺超声检查，目标结节最大直径的二维灰阶超声图像清楚、完整，且描述性报告及临床病理资料完整[7]。③根据美国放射学会（American College of Radiology, ACR）发布的第5版BI-RADS[8]，将目标病变定为BI-RADS 3类、4A类、4B类、4C类的结节。排除标准：①病理结果不明确；②肿瘤过大导致图像显示不完全；③患者术前接受过抗癌治疗（化疗、放疗或内分泌治疗）；④图像及临床

资料不完整。本研究经医院伦理委员会审核批准（批件号：2019–027），因属于回顾性研究，豁免患者知情同意。

最终共纳入430名患者，对于多发结节的BI–RADS 3类和4类乳腺病变的患者，本研究只纳入BI–RADS类别最高的结节作为目标结节，以保证每次观察的统计独立性。将患者按照7∶3的比例随机分为训练集（n=301）与验证集（n=129）。收集患者的临床特征资料，包括年龄、肿瘤大小、肿瘤位置和病理结果。

1.2 方法

1.2.1 检查方法

采用德国Siemens公司的S2000超声诊断仪，探头检测频率为4~9 MHz[1]，由于本研究采用常规诊断超声模式，设备根据预设条件自动调节声输出，机械指数（mechanical index, MI）0.6 ~ 0.8。所有患者在活检或手术前2周内进行乳腺常规超声检查，取仰卧位，对纳入患者的乳腺目标结节进行横切面、纵切面扫描。取结节最大切面，在获得质量较好的二维图像的基础上，将图像以医学数字成像和通信（digital imaging and communications in medicine, DICOM）格式储存在硬盘中[9]，同时记录乳腺结节的大小、位置和超声特征。

1.2.2 超声图像预处理

首先挑选出符合纳入标准的乳腺结节的超声图像，将所有图像均转换为JPEG格式，并调整为1 024×1 024像素，保持空间分辨率一致。为增强模型鲁棒性并抑制过拟合，在训练阶段采用在线数据扩增策略，包括随机翻转、亮度/对比度动态扰动，以模拟临床采集图像的多样性变化[10]。然后使用开源标签工具LabelMe（http://labelme.csail.mit.edu/Release3.0），根据病理结果将恶性结节超声图像标记为“1”、良性结节标记为“0”，并生成用于模型训练的JSON格式文件。

1.2.3 主干网络选择与深度学习模型构建

本研究测试了5种典型图像分类主干网络，包括GoogLeNet、VGG11、AlexNet、ResNet–34、ViT。采用5折交叉验证，通过受试者工作特征（receiver operating characteristic, ROC）曲线评估深度学习模型的诊断效能，并在验证集中验证模型。最后绘制深度学习模型的ROC曲线，通过对比曲线下面积（area under curve, AUC）、95%置信区间、准确性、灵敏度、特异性、精准度及F1值等指标来评估模型的预测性能，选取最佳深度学习模型进行可视化。对选取的最佳模型的顶层进行定制化改造：移除原始的全连接分类层，添加全局平均池化层，将特征图压缩为512维向量；随后依次接入1个512维全连接层（ReLU激活）、Dropout层（丢弃率=0.5）以及2神经元Softmax激活层，最终输出良性及恶性结节的概率评分，二者之和为1。模型训练采用Adam优化器，初始学习率设为0.001，并配合余弦退火（Cosine Annealing）策略动态调整学习率。损失函数选用二元交叉熵（binary cross–entropy with logits），批量大小（batch size）设为32，最大训练轮次（epochs）设为100。采取动态早停（early stopping）机制调整训练周期以防止模型过拟合，当验证集损失连续10轮无下降时停止训练[10–11]。

1.2.4 深度学习模型可视化

本研究选用训练好的最佳模型对乳腺超声图像进行良恶性鉴别，并使用Grad–CAM（gradient–weighted class activation mapping）方法对模型决策进行可视化[12]。首先，对乳腺超声图像进行灰度归一化、裁剪和尺寸调整后，将图像输入最佳模型进行前向传播，获得每幅图像的良恶性预测概率。其次，对目标类别输出进行反向传播，计算其相对于最后一层卷积特征图的梯度。然后，对各通道梯度进行全局平均，得到对应的通道权重，并与特征图进行加权求和，生成初始二维热力图。随后，对热力图进行ReLU处理，仅保留对目标类别具有正贡献的区域，并通过双线性插值将热力图调整至原始图像尺寸。最后，将处理后的热力图叠加于原始超声图像上，形成彩色可视化结果，以直观显示模型关注的病灶区域。深度学习模型可视化热力图生成流程见图1，热力图计算成图过程使用Python 3.9、PyTorch 2.0及PyTorch Grad–CAM库完成。

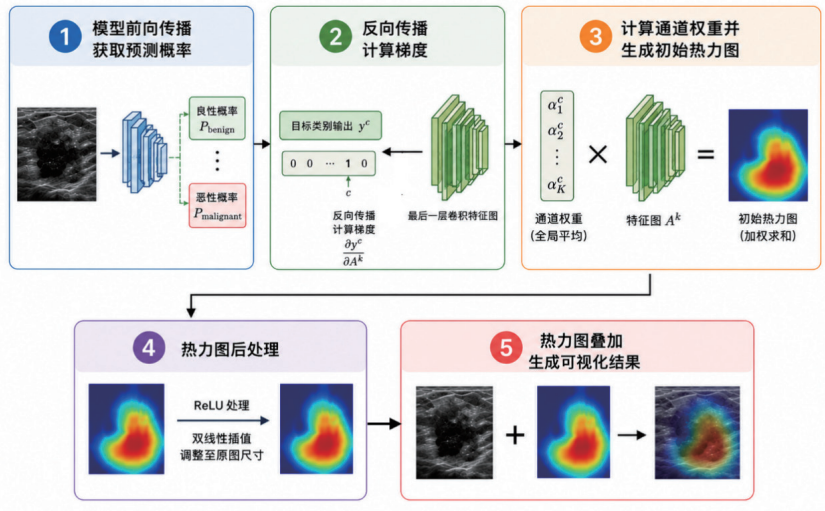


图 1 深度学习模型可视化热力图生成流程

1.2.5 图像观察及研究方法

由来自社区卫生服务中心的 1 名具备执业资格及 18 年乳腺超声诊断经验的超声医师根据 BI-RADS 分级标准[13]对乳腺结节进行评级及良恶性判断。本研究将 BI-RADS 3类和 4A 类定义为良性，BI-RADS 4B 类、4C 类定义为恶性。为评估深度学习模型辅助超声医师诊断效能，采用同一医师两次阅片方式：首先，由该名医师独立完成所有病例的超声图像解读与诊断，期间未提供任何深度学习模型辅助信息；然后，同一医师在深度学习模型可视化热图辅助下，对同一批病例进行超声图像解读与诊断。采用该设计能够最大程度减少个体间诊断水平差异对试验结果的干扰，更精准地评估深度学习辅助工具对同一医师诊断效能的提升作用。为避免前后对照设计中重复阅片导致的记忆偏倚，本研究参考同类影像学诊断研究[14]的洗脱期设置标准，在两次阅片之间设置了为期 2 周的洗脱期，该时长能够有效降低医师对初次阅片细节的记忆干扰，保障两次诊断结果的独立性与客观性。

1.2.6 模型效能评估

分别评估深度学习模型及其辅助前后超声医师阅片的诊断效能，主要评估指标如下。① AUC：AUC 越接近 1，模型预测效能越好，AUC>0.8 提示效能优秀。② 分类性能指标：包括准确性、灵敏度、特异度和精确度，准确性 = 正确预测数/总预测数，灵敏度 = 真阳性数/（真阳性数 + 假阳性数），特异度 = 真阴性数/（真阴性数 + 假阳性数），精确度 = 真阳性数/（真阳性数 + 假阴性数），F1 值 =2 × 精确度 × 灵敏度/（精确度 + 灵敏度），各指标取值 0~1，越接近 1 提示性能越优[11]。

1.3 统计学方法

采用 R 4.3.1 及 SPSS 26.0 软件对数据进行统计分析。对计量资料进行正态分布及方差齐性检验，符合正态分布的计量资料以均值 ± 标准差（$\bar{x} \pm s$）表示，组间比较采用独立样本 t 检验；不符合正态分布的计量资料以中位数（上四分位数，下四分位数）[$M(Q_1, Q_3)$] 表示，组间比较采用 Mann-Whitney U 检验。计数资料以例数（n）和百分率（%）表示，组间比较采用 χ^2 检验。以 P<0.05 表示差异有统计学意义。

2 结果

2.1 临床特征资料比较

本研究共计收集 430 例乳腺结节患者，术后病理证实良性病变为 215 例，恶性病变为 215 例。整体年龄 18~86 岁，平均年龄（47.62 ± 14.62）岁。训练集与验证集患者在年龄（P=0.716）、结节直径（P=0.671）、结节位置（P=0.292）及病理结果（P=0.674）等指标上差异均无统计学意义。具体临床特征见表 1。

2.2 最佳深度学习模型筛选

为了找到鉴别乳腺结节良恶性的最佳模型，比较了 ResNet-34、AlexNet、GoogLeNet、VGG11、ViT 模型的效能。结果表明，ResNet-34 模型的准确性、灵敏度等多数效能评估指标在训练集和验证集中优于其他模型，见表 2。ROC 曲线分析显示，ResNet-34

表 1 训练集与验证集的临床特征资料比较

特征	全部（n=430）	训练集（n=301）	验证集（n=129）	t/χ^2 值	P 值
年龄（岁，$\bar{x} \pm s$）	47.62 ± 14.62	47.45 ± 14.33	48.02 ± 15.30	18 984.5[a]	0.716
结节直径（mm）	18.93 ± 9.44	18.75 ± 9.22	19.35 ± 9.98	18 912.5[a]	0.671
结节位置［n(%)］				1.110[b]	0.292
左侧	225（52.33）	152（50.50）	73（56.59）		
右侧	205（47.67）	149（49.50）	56（43.41）		
病理结果［n(%)］				0.177[b]	0.674
良性	215（50.00）	148（49.17）	67（51.94）		
恶性	215（50.00）	153（50.83）	62（48.06）		

注：[a]为 t 值；[b]为 χ^2 值。

模型在训练集和验证集中 AUC 分别为 0.790 和 0.741，显著优于其他模型，见图 2。上述结果提示，ResNet–34 模型能够更有效地提取乳腺超声图像中的关键特征，对结节良恶性的鉴别能力更强，因而在整体诊断性能上优于其他模型。与此同时，该模型在训练集和验证集中均保持较好的表现，说明其不仅具有较高的拟合能力，也具备一定的泛化能力和稳定性。因此，筛选得到 ResNet–34 模型作为最佳模型。

表 2 不同深度学习模型诊断效能比较

模型	分组	AUC	95%*CI*	准确性	灵敏度	特异度	精确度	F1 值
ResNet–34	训练集	0.790	0.740 2 ~ 0.839 4	0.714	0.627	0.804	0.768	0.691
	验证集	0.741	0.656 1 ~ 0.826 6	0.698	0.758	0.642	0.662	0.707
AlexNet	训练集	0.617	0.554 1 ~ 0.680 4	0.591	0.608	0.574	0.596	0.602
	验证集	0.653	0.558 4 ~ 0.747 6	0.651	0.323	0.955	0.870	0.471
GoogLeNet	训练集	0.653	0.591 2 ~ 0.714 1	0.611	0.660	0.561	0.608	0.633
	验证集	0.645	0.549 1 ~ 0.741 2	0.628	0.371	0.866	0.719	0.489
VGG11	训练集	0.562	0.496 8 ~ 0.626 5	0.568	0.458	0.682	0.598	0.519
	验证集	0.672	0.577 4 ~ 0.766 4	0.667	0.371	0.940	0.852	0.517
ViT	训练集	0.567	0.501 8 ~ 0.631 7	0.565	0.582	0.547	0.571	0.576
	验证集	0.605	0.507 1 ~ 0.703 3	0.597	0.210	0.955	0.812	0.333

注：AUC 为曲线下面积。

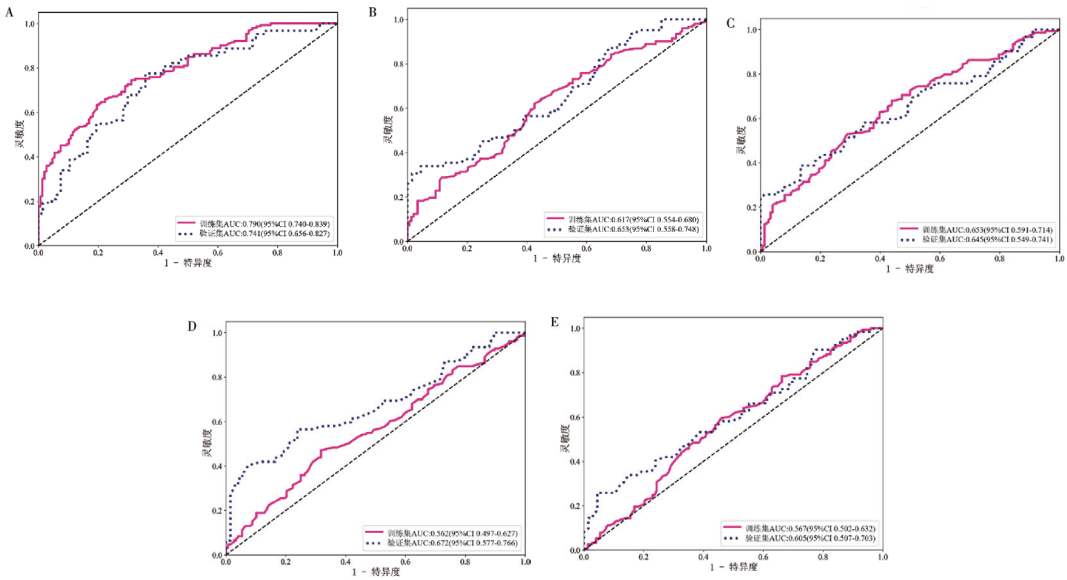


图 2 不同深度学习模型的 ROC 曲线 A 为 ResNet–34 模型；B 为 AlexNet 模型；C 为 GoogLeNet 模型；D 为 VGG11 模型；E 为 ViT 模型

注：ROC 为受试者工作特征；AUC 为曲线下面积。

2.3 深度学习模型可视化

选用ResNet–34模型，并采用Grad–CAM可视化方法对乳腺超声图像进行热图分析和可视化表达，见图3。结果显示，模型生成的热图主要集中于结节本身及其邻近关键区域，而较少关注无关的腺体背景。对于恶性结节，热图的高响应模式更多反映了内部回声分布不均、形态不规则、分叶、毛刺、钙化及与周围组织界面异常等综合征象；对于良性结节，热图呈现的响应模式则相对集中、规则，提示模型识别到的影像信息更具均质性和稳定性。上述结果表明，ResNet–34模型并非仅依赖单一局部灰度差异进行判断，而是能够综合结节整体结构特征及局部异常信息，对乳腺结节良恶性进行区分。模型提取的判别信息与临床超声诊断所关注的关键征象具有较好一致性，可为基层超声医师提供更直观的影像学参考，从而提升辅助诊断的可信度和实用性。

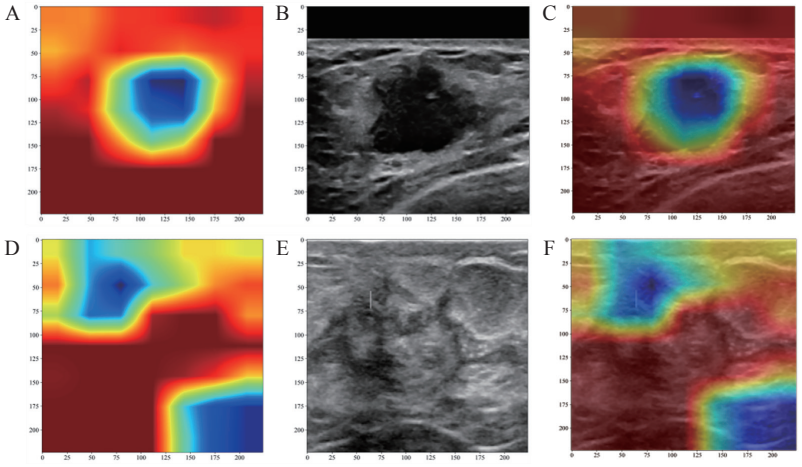


图3 ResNet–34模型生成的乳腺结节可视化热图 A为恶性结节热力图；B为恶性结节二维超声图；C为恶性结节可视化热图；D为良性结节热力图；E为良性结节二维超声图；F为良性结节可视化热图

2.4 深度学习模型辅助超声医师诊断效能

采用ResNet–34模型生成的可视化热图辅助超声医师进行第2轮阅片，并比较不同阅片方式的诊断效能，见表3。结果显示，深度学习模型单独阅片的整体诊断效能略优于超声医师第1轮独立阅片，提示该模型已具备一定的乳腺结节良恶性鉴别能力。在可视化热图辅助下，超声医师第2轮阅片的各项诊断指标均大幅度提升，诊断效能为3种方案里最高。与第1轮独立阅片相比，第2轮深度学习模型辅助阅片的AUC、准确性、灵敏度、特异度、精确度和F1值都得到提升。从提升幅度来看，灵敏度和F1值提升最明显，分别提高34.3%和29.9%，表明在深度学习模型辅助下，超声医师对恶性结节的检出能力及整体诊断均衡性均得到有效改善，见图4。上述结果表明，基于可视化热图的人机协同模式能够充分结合深度学习模型的特征提取优势和医师的临床判断能力，从而更有效地提升基层超声医师对乳腺结节良恶性的鉴别效能。

3 讨论

本研究聚焦于BI–RADS3类和4类乳腺结节的良恶性鉴别难题，通过构建并验证一种基于超声图像的深度学习可视化模型，系统评估了其在辅助诊断中的效能。研究结果显示，基于可视化热图的人机协同模式在准确性、灵敏度、特异度、精确度、F1值这些关键指标上，均显著优于基层超声医师的独立阅片表现，灵敏度和F1值提升最明显，分别提高34.3%和29.9%，证实了人工智能技术在提升乳腺结节定性诊断性能方面的巨大潜力。本研究的创新核心在于将“黑箱”式的深度神经网络转化为可视化的辅助工具，不仅实现了算法与临床场景的深度融合，还通过“人机对比”策略有效弥补了基层医师在超声影像解读经验上的不足，显著降低了漏诊与误诊风险。

近年来，深度学习在乳腺影像诊断中的价值已获广泛验证[15–21]。本研究基于ResNet–34构建的超声模型独立阅片效能略优于基层医师首轮独立诊断，与Shen等[15]报道的双模态超声深度学习放射组学模型（测试集AUC=0.825）优于全部放射科医师独立诊断的结果相一致，提示该模型已具备较可靠的良恶性判别能力。然而，单纯追求算法精度的提升并非临床转化的最终目标，本研究侧重探讨“人机协同”模式在真实阅片场景中对基层超声医师诊断效能的提升。本研究结果表明，在可视化热

表 3 深度学习模型及其辅助前后超声医师阅片的诊断效能比较

项目	AUC	95% *CI*	准确性	灵敏度	特异度	精确度	F1 值
深度学习模型	0.741	0.656 3 ~ 0.826 6	0.682	0.677	0.687	0.667	0.672
深度学习模型辅助前超声医师阅片	0.650	0.566 9 ~ 0.732 6	0.651	0.613	0.687	0.644	0.628
深度学习模型辅助后超声医师阅片	0.822	0.755 1 ~ 0.888 3	0.822	0.823	0.821	0.810	0.816

注：AUC为曲线下面积。

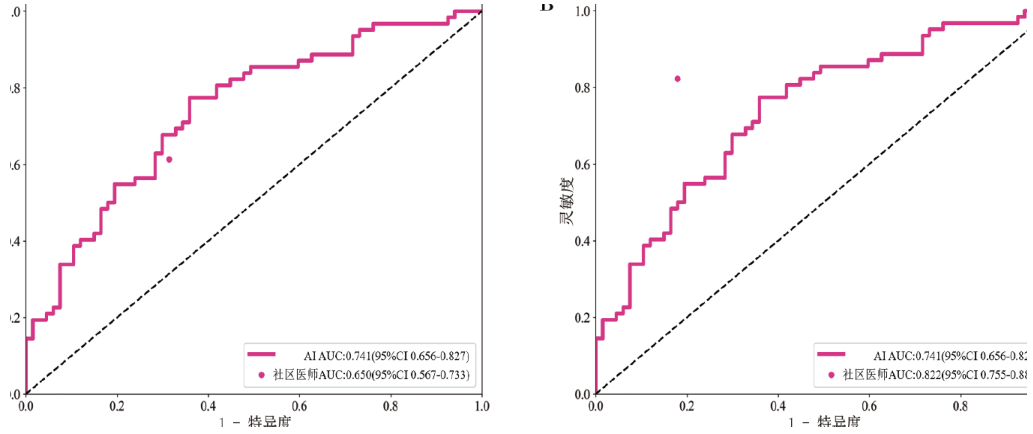


图 4 深度学习模型辅助前后超声医师阅片的 ROC 曲线 A 为深度学习模型辅助前超声医师阅片；B 为深度学习模型辅助后超声医师阅片

注：ROC为受试者工作特征；AUC为曲线下面积。

图辅助下，模型能够实质性地提升乳腺结节良恶性的鉴别效能。Jin 等[16]采用二次阅读模式，使初级放射科医师的 AUC 从 0.879 提升至 0.921，灵敏度由 0.83 提高至 0.92，不必要活检率下降 14.70%，且对≤2 cm病灶获益更明显。Kim 等[19]亦指出，无经验读者在人工智能辅助下平均灵敏度从 56.9%跃升至 87.5%，阴性预测值与准确率均显著改善，而经验丰富者因基线水平较高未见进一步提升。Ma 等[20]基于自动乳腺容积扫描的 YOLO 模型同样发现，各年资医师的 AUC 在人工智能辅助后均显著提高，有效弥补了不同经验医师间的诊断差距。本研究中，基层医师经热图辅助后诊断效能全面提升，尤其灵敏度大幅增加，再次印证人工智能对基层医院经验相对不足的医师具有更突出的增益价值[16,19–20]。除灵敏度的提升外，本研究辅助后超声医师阅片特异度亦提高 0.134，提示热图辅助在降低假阳性方面同样发挥了作用。Lu 等[17]利用深度学习放射组学模型对 BI–RADS 4A 类病灶进行风险再分层，避免了 27.7%的不必要活检，所有医师在人工智能辅助后诊断准确率和特异性均显著提高，而敏感性保持优异。Liang 等[18]开发的基于动态增强 MRI 的 BL4AS 系统在辅助医师后，假阳性率降低 27.3%，诊断准确性提升 14.3%，且不同年资医师间的判读差异缩小 24.5%。本研究以基层医师为对象，发现人机协同模式下 F1 值大幅提升，这体现了灵敏度与特异度的综合改善，与上述研究中人工智能降低假阳性、提升诊断信心的结论相互印证。

本研究所采用的可视化热图在辅助决策中发挥了关键作用。该热图直观标示模型关注区域，使基层医师在理解模型决策逻辑的同时，能够有意识地审视可疑征象。Shen 等[15]通过 Grad–CAM 热图显示模型聚焦于肿瘤内部低回声区和高增强区，Lu 等[17]与 Liang 等[18]借助热图提升模型的可解释性，为临床决策提供可视化依据。本研究中，热图辅助下的人机协同模式将深度学习特征提取优势与医师临床判断经验有机结合，最终获得的诊断效能超越单一模型或单一医师的独立表现。

值得注意的是，人工智能辅助效果可能受检查技术、病灶特征及人群乳腺背景等多种因素影响。Lee 等[22]发现 AI–CAD 系统诊断性能随乳腺密度增加而提高，在极度致密乳腺中准确性和阴性预测值最高，提示在亚洲女性等致密型乳腺高发人群中更具独特价值。Han 等[21]构建的 MRI 可解释放射组学模型针对小 BI–RADS 4 类病灶进行预测，人工智能辅助后资深与初级放射科医师的 AUC 均显著提升，且对 >10 mm病灶诊断效能更高。本研究尚未对病

灶大小、乳腺密度等进行分层分析，后续可进一步评估模型在上述亚组中的辅助价值，并探索融入超声造影等多模态信息的可能性[15]。

尽管人工智能展现出明显的辅助效能，但在临床实践中仍应严格将其定位为增强工具而非替代[23]。Fujioka等[24]在综述中强调，人工智能可提高乳腺超声诊断的精度与效率，但必须由医师负责监督应用，同时需正视算法公平性、患者接受度等挑战。Krishna NK等[23]指出，放射诊断不仅依赖于图像识别，更需要临床判断、数据整合及医患沟通，人工智能的错误或过度依赖可能削弱临床思维，尤其对经验不足者而言。本研究采用的人机协同阅片模式，要求基层医师结合热图提示与自身经验作出最终判读，既充分利用人工智能的增强作用，又避免盲目依赖，契合伦理与法律框架下人工智能辅助决策的基本原则[23-24]。未来，在推广此类模型时，需关注训练数据的多样性以规避算法偏见，并通过真实世界多中心验证筑牢其临床转化基础[23]。

本研究尚存在以下局限性：第一，本研究用于训练和验证的图像数量比较少，一定程度上会制约模型的泛化能力与稳定性，在未来的工作中，我们将收集更多的数据来验证模型的性能；第二，本研究为回顾性研究，需要更大规模的前瞻性试验来验证模型的性能；第三，本实验构建的深度学习模型仅使用超声图像的二维信息，可能遗漏了肿瘤的一些深度学习三维特征信息。

综上所述，基于超声图像的深度学习模型表现出较好的乳腺结节良恶性区分能力，在模型辅助下能够大幅提升基层超声医师鉴别乳腺结节良恶性的诊断效能，可为乳腺结节患者治疗方案的选择提供有利指导信息。

利益冲突：作者声明没有利益冲突。

[参 考 文 献]

[1] 师琳, 钟李长, 马方, 等. 瘤周超声影像组学对乳腺结节良恶性的鉴别诊断价值[J]. 肿瘤影像学, 2023, 32(6): 485–491.

[2] Li L, Niu Y, Tian F, et al. An efficient deep learning strategy for accurate and automated detection of breast tumors in ultrasound image datasets[J]. Front Oncol, 2025, 14: 1461542.

[3] Stastny I, Zubor P, Kajo K, et al. Aberrantly methylated cfDNA in body fluids as a promising diagnostic tool for early detection of breast cancer[J]. Clin Breast Cancer, 2020, 20(6): e711–e722.

[4] 张莉敏. 基于CT平扫影像组学及机器学习预测急性缺血性脑卒中常规治疗患者结局研究[D]. 重庆: 重庆医科大学, 2024.

[5] 张烁, 张荣, 张岩波. 基于深度学习的医学图像识别研究综述[J]. 中国卫生统计, 2020, 37(1): 150–156.

[6] 尚建军, 邬旭日. 超声造影对于不同大小BI–RADS 4类乳腺结节的良恶性定性诊断分析[J]. 影像研究与医学应用, 2022, 6(14): 152–154.

[7] 米拉·也尔兰, 张海见, 胡峙珩, 等. 超声影像组学对BI–RADS 4a类不规则乳腺结节良恶性的鉴别价值[J]. 分子影像学杂志, 2023, 46(1): 12–20.

[8] Merjane V, Perin D M P, Bacha P M G E, et al. Breast Imaging Reporting and Data System (BI–RADS®): a success history and particularities of its use in Brazil[J]. Rev Bras Ginecol Obstet, 2024, 46: e–rbgo6.

[9] 师琳, 钟李长, 谷丽萍. 瘤内瘤周超声影像组学预测乳腺浸润导管癌Ki–67状态[J]. 中国医学物理学杂志, 2024, 41(9): 1139–1144.

[10] 林雅丹, 阮思敏, 陈立达, 等. 面向LR–M肝癌精准诊断的超声造影时空感知深度学习系统开发[J]. 现代仪器与医疗, 2025, 31(3): 1–9.

[11] 洪森, 漆万银, 周颖俊, 等. 基于TabPFN模型的CT影像组学联合临床影像特征对亚实性肺腺癌浸润性的术前预测研究[J]. 现代仪器与医疗, 2025, 31(6): 26–35, 81.

[12] 余美慧, 袁泉, 曾书娥, 等. 基于超声图像的迁移学习模型在乳腺肿块良恶性鉴别诊断中的价值[J]. 临床超声医学杂志, 2022, 24(9): 652–656.

[13] Maani N, Westergard S, Yang J, et al. NF1 patients receiving breast cancer screening: insights from the ontario high risk breast screening program[J]. Cancers, 2019, 11(5): 707.

[14] 崔兆国, 吴昊, 汤敏, 等. 人工智能对基层影像医生筛查肺结节的辅助价值[J]. 中国现代医生, 2021, 59(1): 99–101, 105, 封3.

[15] Shen J, Chen G, Lun H, et al. Dual–modal ultrasound–based deep learning radiomics for differentiation of benign and malignant breast lesions[J]. Gland Surgery, 2025, 14(10): 2035–2050.

[16] Jin Z Y, Li J K, Niu R L, et al. Application of an artificial intelligence–assisted diagnostic system for breast ultrasound: a prospective study[J]. Gland Surgery, 2024, 13(12): 2221–2231.

[17] Lu X, Lu Y, Zhao W, et al. Ultrasound–based deep learning radiomics for multi–stage assisted diagnosis in reducing unnecessary biopsies of BI–RADS 4A lesions[J]. Quantitative Imaging in Medicine and Surgery, 2025, 15(3): 2512–2528.

[18] Liang Y, Wei Z, Dai Y, et al. An interpretable AI system reduces false–positive MRI diagnoses by stratifying high–risk breast lesions[J]. Nature Communications, 2026, 17: 2263.

[19] Kim J, Han K, Kim K W, et al. Feasibility of using an AI system for breast ultrasonography interpretation according to clinical expertise: results of a pilot study[J]. Journal of the Korean Society of Radiology, 2026, 87(2): 314–327.

[20] Ma Q, Wang J, Dong B, et al. YOLO AI model based on an automated breast volume scanner for the detection of benign and malignant breast lesions[J]. Quantitative Imaging in Medicine and Surgery, 2025, 15(10): 10156–10167.

[21] Han C, Chen J, Hong M, et al. MRI radiomics for diagnosing small BI–RADS 4 breast lesions: an interpretable model[J]. Quantitative Imaging in Medicine and Surgery, 2025, 15(6): 5060–5072.

[22] Lee J, Kim W H, Kim J, et al. Effect of a real–time artificial intelligence–assisted ultrasound system on BI–RADS C4 breast lesions based on breast density[J]. Cancers, 2026, 18(3): 536.

[23] Krishna NK R, R.S. A, K S. Artificial intelligence in radiology: augmentation, not replacement[J]. Cureus, 2025, 17(6): e86247.

[24] Fujioka T, Tsukada J, Hayashida T, et al. The future of breast cancer diagnosis in Japan with AI and ultrasonography[J]. JMA J, 2024, 8(1): 91–101.