

# 评估可解释的人工智能技术在多种成像方式下解释乳腺癌诊断的有效性

李 禄<sup>1</sup>, 罗浩军<sup>1,2\*</sup>

<sup>1</sup>重庆医科大学附属第二医院乳腺甲状腺外科, 重庆

<sup>2</sup>重庆市第五人民医院, 重庆

收稿日期: 2025年1月27日; 录用日期: 2025年2月21日; 发布日期: 2025年2月28日

## 摘 要

乳腺癌持续位居全球女性癌症发病与致死的主要原因之列。早期且精确的诊断对于优化患者预后具有举足轻重的地位。乳房X线摄影、超声检查及磁共振成像(Magnetic Resonance Imaging, MRI)等影像学技术在乳腺癌的诊断中扮演着至关重要的角色。然而, 这些技术手段面临着准确性波动、操作员依赖性显著及结果阐释困难等多重挑战。在此背景下, 人工智能(Artificial Intelligence, AI), 尤其是可解释人工智能(Explainable Artificial Intelligence, XAI)的融入, 已成为提升诊断精确度及增强信任度的革命性途径。本综述聚焦于XAI技术在乳腺癌诊断领域内, 于不同成像模式中的应用效果比较。深入探讨了核心的XAI方法, 诸如Shapley加性解释(SHAP)、局部可解释模型无关解释(LIME)以及基于梯度的类激活映射(Grad-CAM), 着重阐述了它们在增进模型可解释性及提升临床实用性方面的具体成效。综述不仅分析了XAI技术在乳房X线摄影、超声及MRI应用中的优势与局限, 还特别强调了其在提高AI辅助预测透明度方面的贡献。此外, 本文亦评估了XAI在应对假阳性、假阴性问题以及多模态成像数据整合挑战中的效能。该评论的核心价值在于, 它全面剖析了XAI在缩小AI技术进展与临床实际应用之间鸿沟的潜力。通过提升透明度, XAI技术能够增强临床医生对AI的信任度, 促进其更顺畅地融入诊断工作流程, 从而助力个性化医疗实践的推进及患者治疗成效的改善。综上所述, 尽管XAI在提升AI模型可解释性与准确性方面取得了显著进展, 但在计算复杂度控制、普遍适用性拓展及临床接纳度提升等方面仍面临诸多挑战。未来研究应着重于优化XAI方法、促进跨学科间的深度合作, 并开发标准化的框架体系, 以确保XAI技术能在多样化的临床环境中实现可扩展性与可靠性的双重提升。

## 关键词

乳腺癌诊断, 可解释人工智能(XAI), SHAP, LIME, Grad-CAM, 影像学技术, 个性化医疗, 人工智能

\*通讯作者。

# Evaluating the Effectiveness of Explainable AI Techniques in Interpreting Breast Cancer Diagnoses Across Multiple Imaging Modalities

Lu Li<sup>1</sup>, Haojun Luo<sup>1,2\*</sup>

<sup>1</sup>Department of Breast and Thyroid Surgery, The Second Affiliated Hospital of Chongqing Medical University, Chongqing

<sup>2</sup>Chongqing Fifth People's Hospital, Chongqing

Received: Jan. 27<sup>th</sup>, 2025; accepted: Feb. 21<sup>st</sup>, 2025; published: Feb. 28<sup>th</sup>, 2025

## Abstract

Breast cancer remains one of the leading causes of cancer incidence and mortality among women worldwide. Early and accurate diagnosis plays a pivotal role in optimizing patient prognosis. Imaging techniques such as mammography, ultrasound, and magnetic resonance imaging (MRI) play crucial roles in the diagnosis of breast cancer. However, these techniques face multiple challenges, including accuracy fluctuations, significant operator dependency, and difficulties in result interpretation. In this context, the integration of Artificial Intelligence (AI), especially Explainable Artificial Intelligence (XAI), has become a revolutionary approach to improving diagnostic accuracy and enhancing trust. This review focuses on the comparative application of XAI technologies across different imaging modalities in breast cancer diagnosis. It delves into core XAI methods such as Shapley Additive Explanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM), with an emphasis on their effectiveness in enhancing model interpretability and improving clinical utility. The review analyzes not only the advantages and limitations of XAI in mammography, ultrasound, and MRI applications but also highlights its contribution to increasing the transparency of AI-assisted predictions. Additionally, the review evaluates the performance of XAI in addressing issues related to false positives, false negatives, and the challenges of multimodal imaging data integration. The core value of this review lies in its comprehensive analysis of the potential of XAI in bridging the gap between advancements in AI technology and clinical application. By enhancing transparency, XAI can boost clinicians' trust in AI, facilitating its smoother integration into diagnostic workflows, thereby promoting personalized medical practices and improving patient treatment outcomes. In conclusion, despite significant progress made by XAI in improving AI model interpretability and accuracy, challenges remain in terms of computational complexity, general applicability, and clinical acceptance. Future research should focus on optimizing XAI methods, fostering interdisciplinary collaboration, and developing standardized frameworks to ensure the scalability and reliability of XAI technologies in diverse clinical environments.

## Keywords

Breast Cancer Diagnosis, Explainable AI, SHAP, LIME, Grad-CAM, Imaging Modalities, Personalized Medicine, Artificial Intelligence



## 1. 引言

乳腺癌是全球女性最常见且致命的癌症之一。2020 年约 230 万例新诊断的乳腺癌病例, 占有癌症死亡的 15% [1] [2]。乳腺癌尤其在低收入和中等收入国家造成了严重公共卫生挑战[3]。早期发现在提高生存率方面起着至关重要的作用, 乳房 X 线摄影、超声和磁共振成像(Magnetic Resonance Imaging, MRI)等各种筛查方式通常用于检测该疾病。尽管筛查取得了重大进展, 但挑战仍然存在, 包括假阳性、过度诊断和筛查技术获取方面的差异。研究表明, 乳房 X 线摄影筛查的有效性可能因年龄、乳房密度和风险因素而存在很大差异, 这凸显了对个性化的检测方法的需求[4]-[6]。随着乳腺癌筛查研究的发展, 提高诊断准确性并尽量减少不必要的干预仍然是临床医生和研究人员的首要任务[7] [8]。

人工智能(Artificial Intelligence, AI)的出现改变了医学成像, 提供了可帮助临床医生更准确、更有效地解释复杂医疗数据的工具。在乳腺癌方面, 人工智能在提高诊断准确性方面表现出了良好的效果, 尤其是在乳房 X 线摄影和 MRI 方面[4] [9]。然而, 尽管取得了进步, 但人工智能技术通常以“黑匣子”的形式运行, 这使得医疗保健专业人员难以理解模型如何做出决策。这导致了可解释人工智能(Explainable Artificial Intelligence, XAI)的发展, 旨在使人工智能决策过程更加透明和可解释[10] [11]。XAI 在医学领域尤为重要, 因为了解诊断背后的原因可以直接影响临床决策和患者结果。已经引入了显著图(Saliency maps, 显著图是一种基于弱监督的归因方法, 其核心在于通过一个分类模型生成图像的显著特征图, 为深度学习模型的决策过程提供了一种可视化的理解方式。)和规则基于模型(rule-based models 是一种通过制定一系列规则和约束, 对数据进行处理和分析的方法。它主要依赖于事先设定的规则来描述和解释现象, 从而实现对数据的挖掘和预测, 以实现对复杂数据的有效处理和深度分析。)等技术来提高透明度, 使临床医生能够信任人工智能输出并将其有效地整合到临床工作流程中[12] [13]。

本综述旨在评估各种 XAI 技术在不同成像模式解释乳腺癌诊断的比较效果。随着乳腺癌筛查中 AI 应用的不断发展, 了解哪些 XAI 方法提供最可靠和最透明的结果对于将 AI 融入临床实践至关重要。通过回顾现有文献, 本综述将评估不同 XAI 方法的优势和局限性, 重点关注它们在保持诊断准确性的同时增强可解释性的能力。此外, 它将强调在临床环境中采用 XAI 所面临的挑战, 并提出未来研究的潜在领域, 以改善 AI 在乳腺癌诊断中的应用[14] [15]。

## 2. XAI 技术概述

在医学成像领域, 尤其是乳腺癌诊断中, 可解释人工智能(XAI)技术已成为提升 AI 模型透明度与信任度的关键工具。XAI 技术, 诸如 Shapley 加性解释(SHAP)、局部可解释模型无关解释(LIME)以及基于梯度的类激活映射(Grad-CAM)等, 旨在提高 AI 模型的透明度和可解释性, 从而助力临床医生验证 AI 驱动的诊断结果[16] [17]。这些技术不仅能够帮助理解 AI 的行为, 还能让临床医生验证 AI 驱动的结果。

SHAP 作为一种基于合作博弈理论的 XAI 技术, 通过量化每个特征对模型预测的贡献来进行解释[18]。SHAP 提供了一种数学上一致的方式来观察特征对结果的影响值, 这对于理解医疗影像中每个特征(例如病变的大小、形状或纹理)对临床决策的重要性至关重要[19]。然而, SHAP 计算上可能比较复杂, 尤其是在处理大数据集或高度复杂的模型时, 可能会限制其在实时应用中的可扩展性[20]。尽管如此, SHAP 凭借其强大的理论基础和可靠性, 依然是开发透明 AI 模型的基石。

LIME 方法则通过构建简洁且具备解释性的局部模型来逼近复杂模型的预测行为, 为单个预测结果提供直观的阐释[21]。其核心优势在于模型无关性, 即能够适配并解析任何类型的机器学习模型, 展现出广泛的跨领域应用灵活性。这一特性使得 LIME 在临床决策支持系统中尤为适用, 医生可通过理解 AI 模型的预测逻辑来指导特定患者的治疗决策[22]。然而, LIME 生成的解释存在不稳定性, 源于其对输入数据扰动的依赖性, 可能导致替代模型无法始终如一地精确反映原始模型的全局特性。尽管如此, LIME 在处理速度和实施便捷性方面的优势, 使其在需要对每个数据实例提供个性化解释时仍被广泛采纳。

Grad-CAM 作为一种专为卷积神经网络(CNN)设计的可视化技术, 在医疗影像分析领域的应用日益广泛[23]。该技术通过分析模型输出对各层特征图的梯度敏感性, 构建出能够突出显示对模型决策起到关键作用的图像区域的热图。这一特性极大地促进了临床医生对 AI 模型预测依据的理解, 特别是在乳腺 X 光、MRI 等医学影像中识别可疑肿块等关键信息方面[24]。然而, 在处理低分辨率或高噪声水平的图像时, Grad-CAM 可能难以提供精确的特征定位[25]。尽管如此, Grad-CAM 凭借其强大的可视化功能, 已成为临床实践中解释深度学习模型在医疗影像分析领域决策过程的关键工具。

此外, 其他 XAI 方法, 如深度学习重要特征(DeepLIFT)和积分梯度(Integrated Gradients)等, 也在提升 AI 模型可解释性方面展现出显著效果。这些方法为 AI 模型的可解释性提供了额外的层次和视角, 但在医疗影像分析领域, SHAP、LIME 和 Grad-CAM 凭借其准确性、易用性以及临床相关性方面的显著优势, 依然是最为广泛采用的 XAI 技术。

### 3. 乳腺癌诊断中的影像学方法及可解释人工智能(XAI)

乳腺癌的诊断依赖于多种影像学方法, 每种方法均展现出独特的优势和局限性。乳房 X 线摄影、超声波以及磁共振成像(MRI)等成像技术在乳腺癌不同阶段的检测与诊断中发挥着至关重要的作用。近年来, AI 与 XAI 的融合进一步强化了这些成像技术的效能, 不仅提升了诊断的准确性, 还增强了临床决策的科学性和可信度。

#### 3.1. 乳房 X 线摄影

乳房 X 线摄影作为乳腺癌筛查的基石, 通过 X 射线成像能力有效检测出早期肿瘤, 尤其擅长识别肿块和微钙化等早期恶性肿瘤指征。然而, 对于乳腺组织致密的女性, 乳房 X 线摄影的检测效果可能受限[26]。为应对这一挑战, AI 技术被引入乳房 X 线摄影领域, 旨在提高检测效能。深度学习技术训练的 AI 模型在提升乳房 X 线摄影的敏感性和特异性、减少假阴性和假阳性方面展现出巨大潜力[27]-[29]。SHAP 和 LIME 等 XAI 技术的集成, 使临床医生能够更深入地理解 AI 驱动的决策逻辑, 进而增强对 AI 模型的信任度, 为临床决策提供有力支持[30]。这些进展对于优化临床工作流程、改善患者预后具有重要意义。

#### 3.2. 超声成像

超声成像常与乳房 X 线摄影联合使用, 尤其适用于评估乳房 X 线摄影难以清晰成像的致密乳房组织。超声成像利用高频声波创建实时图像, 具有非侵入性、无辐射性等优点, 为乳腺病变的评估提供了另一种可靠选择[31]。超声成像在区分良恶性病变方面尤为擅长, 可通过评估病变的形状、大小、回声性等特征进行诊断。AI 在超声成像中的应用, 通过自动分析这些特征, 显著提高了诊断准确性, 并能根据病变特征进行精确分类[32]。此外, Grad-CAM 等 XAI 工具能够突出显示影响 AI 预测结果的超声图像特定区域, 从而增强了 AI 模型的可解释性[17] [24]。这种透明度对于确保超声模型的临床可操作性至关重要, 有助于提升临床医生对 AI 结果的信任度[33]。AI 驱动的超声系统为临床医生提供了宝贵支持, 增强了决策信心, 尤其在面对其他成像方式结果模糊的复杂情况时。

### 3.3. 磁共振成像(MRI)

MRI 在诊断复杂或疑似乳腺癌病例中展现出独特优势, 尤其适用于乳房致密的患者。其卓越的软组织对比度能够精确检测肿瘤, 并评估肿瘤大小、扩散情况及周围组织受累状况[34] [35]。当乳房 X 线摄影或超声成像无法提供明确结果时, MRI 常作为首选检查手段。AI 在 MRI 中的应用, 通过自动执行病变分割、分类等任务, 显著提升了诊断准确性, 实现了更高效、准确的图像解释[22]。此外, AI 模型还能通过分析纹理模式和动态增强图像, 帮助区分良恶性病变[25] [36]。然而, MRI 图像的复杂性和高维性给解释带来挑战, AI 算法通过在大型数据集上训练, 提供自动分析并引导放射科医生关注关键特征, 从而简化了这一过程[37]。将 SHAP 和 LIME 等 XAI 技术集成到基于 MRI 的 AI 模型中, 进一步提高了可解释性, 确保了临床医生能够理解 AI 决策背后的逻辑并将其融入临床实践[22] [37]。这些进展对于增强 MRI 在临床工作流程中的作用、改善诊断准确性和患者护理质量具有重要意义。

### 3.4. 新兴成像模式

多模态成像等新兴成像方式正通过整合多种成像技术的优势, 重塑乳腺癌诊断的格局。PET/MRI 等多模态系统结合了功能和解剖成像, 提供了关于肿瘤生物学、大小和代谢活动的更全面数据, 有助于更深入地了解癌症生物学特性[38]。AI 与多模态成像系统的集成提高了诊断准确性, 因为它能够融合来自多个来源的数据, 提供更精确的评估。此外, AI 还被应用于对比增强成像, 以分析造影剂在肿瘤组织中的行为, 根据血管分布和灌注差异区分良恶性病变[23]。分子成像则是另一种受益于 AI 集成的新兴成像方式, 它利用示踪剂观察组织中的生物变化, 为肿瘤行为提供有价值见解, 为更加个性化的治疗方案铺平道路[22]。

## 4. XAI 技术在不同影像模式下的比较效果分析

### 4.1. 比较方法论框架

为了全面比较 XAI 技术(即可解释人工智能)在不同乳腺癌影像模式下的应用效果, 本研究确立了以下几个核心评估维度: 可解释性、准确性、临床医生信任度以及模型稳健性。可解释性在提升临床医生对 AI 决策的理解与信任方面扮演着关键角色, 尤其是在处理如 MRI 等具有高维度和复杂数据结构的影像模式时[13] [39]。准确性则聚焦于 XAI 技术在精确识别真阳性病例并有效减少假阳性方面的效能, 这在乳腺 X 线摄影中尤为重要, 因假阳性结果在该领域较为普遍[12]。临床医生信任度是衡量 AI 技术临床接受度的核心指标, 研究表明, 当临床医生能够直观理解预测背后的逻辑时, 他们更倾向于信赖 AI 的输出结果[10] [13] [40]。模型稳健性则用于评估 XAI 方法在不同影像条件及多样数据集下的稳定性和可靠性, 这对于集成多源数据的多模态影像系统尤为重要[8]。

### 4.2. 不同影像模式下的 XAI 技术效果比较

XAI 技术在不同影像模式下的应用效果呈现出显著差异, 每种技术针对特定数据类型均展现出独特的优势和局限性。

在乳腺 X 线摄影领域, SHAP 与 LIME 技术已被证实能够显著提升 AI 模型的可解释性, 通过精确解析与恶性肿瘤密切相关的关键特征(如微钙化和肿块), 进而辅助放射科医生更准确地识别可疑病灶, 增强诊断信心[4] [30]。此外, Grad-CAM 在乳腺 X 线摄影中的应用同样取得了显著成效, 通过生成热力图直观标注感兴趣区域, 为临床医生验证 AI 预测提供了有力支持[12]。

在超声影像领域, XAI 技术如 LIME 和 Grad-CAM 在识别并阐释区分良性与恶性病变的关键特征(如

不规则边界和低回声区)方面展现出卓越效能。鉴于超声影像高度依赖于操作员的技术水平, XAI 技术在提升诊断一致性和透明度方面显得尤为重要[17][24]。然而, 超声影像的准确性仍受限于图像质量和操作员技能, 这在一定程度上制约了 XAI 技术在该领域的广泛应用。

在 MRI 领域, Grad-CAM 在可视化乳腺中影响最为显著的区域方面展现出重要价值, 为临床医生提供了关于肿瘤位置的深刻洞察。尽管 MRI 以其高度敏感性著称, 但其较低的特异性和复杂的数据结构给 XAI 方法带来了严峻挑战。在 MRI 应用中, XAI 技术需有效处理大数据集, 并提供清晰、实用的洞察, 以避免过度信息干扰临床医生的判断[41]。此外, 结合 MRI 与 PET 或超声的多模态影像技术在 XAI 领域的应用也提出了新的挑战, 要求 XAI 方法能够整合并合理解释来自多个不同来源的数据[5][12], 从而进一步提升了技术实现的复杂性和对算法鲁棒性的要求。

## 5. XAI 技术的优势与局限

XAI 技术在乳腺癌诊断中展现出显著优势。通过 SHAP、LIME 及 Grad-CAM 等方法, XAI 技术为临床医生提供了清晰的预测洞察, 大大提升了 AI 模型的可解释性, 并增强了临床医生对 AI 结果的信任度。特别是 Grad-CAM 技术, 其生成的热力图能够直观显示乳腺影像中的关键区域, 帮助临床医生更有效地验证 AI 预测结果。通过这一技术, 放射科医生能够明确知道 AI 模型在乳腺 X 光片或 MRI 图像中关注的区域, 从而提升其决策的可信度和准确性[17][42]。此外, XAI 技术通过聚焦肿瘤的大小、质地等关键特征, 显著提高了乳腺癌诊断的准确性, 有效减少了假阳性和假阴性率, 尤其在乳腺 X 线摄影中, XAI 方法帮助放射科医生识别肿瘤的微小特征, 从而减少错误诊断。

尽管 XAI 技术在乳腺癌诊断领域展现出诸多优势, 其应用仍面临若干关键局限性。首要问题在于, 部分 XAI 技术, 诸如 Grad-CAM 和 SHAP, 在计算层面较为复杂, 特别是在处理诸如 MRI 等高维数据时, 计算复杂度显著攀升, 从而制约了其在临床环境中的即时应用潜力。具体而言, SHAP 需对每一特征执行详尽的贡献度分析, 这在面对特征数量庞大的深度学习模型时, 计算开销与时间延迟成为影响临床应用效率的关键因素, 此问题在高维数据集及大规模影像数据的处理场景下尤为凸显, 可能削弱 XAI 在临床决策支持系统中实现实时反馈的能力[43]。其次, XAI 所提供的解释性信息存在准确性及关联性不足的问题。例如, Grad-CAM 可能强调与诊断非直接相关的图像区域, 特别是在图像质量欠佳或噪声水平较高的情况下, 这可能导致临床医师对 AI 结果的误解甚至误判。此类不一致性可能削弱 XAI 的实用价值和用户信任度。尽管 Grad-CAM 通过热力图直观展示感兴趣区域, 但非相关区域的不当高亮仍可能对临床决策构成干扰, 特别是在肿瘤体积小或图像结构复杂的病例中[44]。

最后, XAI 技术在不同数据集及影像模态间的泛化性能尚待增强。不同影像模态(如乳腺 X 线摄影、超声、MRI 等)及数据来源的多样性可能导致 XAI 解释的有效性和一致性出现波动。因此, XAI 方法需具备跨多种影像模态和数据集的优化能力, 以确保其在复杂多变的临床环境中实现广泛应用与稳定表现[45]。以 MRI 与正电子发射断层成像(PET)数据融合的多模态影像分析为例, XAI 技术需有效处理并整合不同来源的数据, 生成统一且可解释的输出, 这对 XAI 方法的适应性和鲁棒性提出了更高要求。

### 5.1. XAI 技术在乳腺癌诊断中的发展方向

XAI 正迅速崛起为提升医疗诊断精确度的核心驱动力, 尤其在乳腺癌识别领域展现出巨大潜力。展望未来, XAI 的发展将聚焦于数个核心方向, 其中个性化与定制化应用尤为关键。当前, XAI 的应用多聚焦于标准的模型训练与预测任务, 然而, 随着个性化需求的日益增长, XAI 技术将深度融合患者的特定属性(诸如年龄、乳腺密度、既往病史等), 定制化地调整 AI 输出解释, 从而为临床医生提供更加精确且个性化的诊断辅助[33]。例如, 通过整合患者的个性化特征信息, SHAP 能够为每位患者提供专属的特

征贡献分析,这不仅使 XAI 能够阐释单一病例的预测依据,还能助力临床医生依据患者的具体状况灵活调整诊断策略。

此外, XAI 技术的演进将深度融合深度学习算法与可解释性框架,如注意力机制,旨在提升 AI 模型的预测精度同时保持高度透明度[44] [46]。这一融合策略将强化 AI 在识别肿瘤形态、大小及微钙化等关键特征上的能力,进一步巩固乳腺癌诊断的准确性。随着深度学习与 XAI 的紧密结合,未来的 AI 模型不仅将实现精度上的飞跃,还将在保持透明度的基础上,实现更为复杂的疾病特征提取与解析。

交互式解释技术的发展同样不容忽视,它赋予临床医生动态查询 AI 模型并调整输入参数的能力,从而直观理解特定特征如何影响诊断结果[13]。这种互动性不仅极大地增强了 AI 的可解释性,还推动了个性化医疗的快速发展,使得治疗策略能够依据患者的特定属性进行量身定制。这一技术的革新使得 AI 从“黑箱”转变为可与临床医生进行高效互动的智能辅助系统。

随着多模态 AI 系统的兴起, XAI 技术的应用面临着前所未有的挑战与机遇。在融合乳腺 MRI、钼靶、PET、超声等多种影像技术的背景下, XAI 需进一步发展以适应复杂数据集的综合分析,为临床医生提供整合且易于理解的诊断解释。未来, XAI 技术还需克服多模态数据融合中的不一致性问题,确保模型能够高效处理来自不同来源的数据,并为临床提供一致且可靠的解释[45]。此外,标准化 XAI 方法在不同影像模式与数据集中的应用,将是未来发展的关键所在,旨在提升 XAI 在不同医疗环境中的可扩展性与泛化能力。

## 5.2. 跨学科合作的不可或缺性

要实现 XAI 在乳腺癌诊断中的最大潜能, AI 研究者、临床专家及伦理学家的紧密合作至关重要。临床专家的实践洞察为 AI 模型设计提供了坚实的现实基础,而 AI 研究者的技术创新则推动了模型性能与透明度的双重提升。面对个性化和多模态数据融合的需求, AI 研究者在模型设计时需兼顾精度、透明性和解释性,尤其是在临床环境中的可接受性[33]。伦理学家则负责监督整个过程的公正性,解决数据隐私、算法偏见及临床接受度等关键问题,确保 XAI 技术的伦理应用[16]。随着 AI 技术在医疗领域的广泛渗透, XAI 不仅要追求可解释性,还要确保在不同患者群体中的公平性。例如, AI 系统在处理不同种族、性别或年龄的患者数据时,如何避免偏见并保障其公平性,已成为 XAI 伦理设计的重要议题。跨学科合作能够确保 AI 技术的应用不仅是技术层面的突破,更是社会和伦理层面的共识,进一步推动 XAI 在乳腺癌诊断中的深入应用与持续发展。

这种跨学科协作机制对于构建符合临床需求、保障患者权益的 XAI 工具具有举足轻重的作用。未来, XAI 的研发将不再局限于提升模型的准确性和可解释性,还将通过与临床、伦理等领域的深度融合,构建一个更加全面、透明且伦理可接受的医疗 AI 系统[45]。这种协作机制将为 AI 技术在乳腺癌诊断中的广泛应用奠定坚实的基础,推动医疗智能化进程迈向新的高度。

## 6. 结论

XAI 技术在乳腺癌诊断领域的应用,显著提高了诊断的准确性和可解释性,增强了临床医生的信任度。SHAP、LIME 及 Grad-CAM 等 XAI 方法通过直观展示关键特征,有效辅助了临床决策,特别是在乳腺 X 线摄影、超声及 MRI 等影像模式下。尽管 XAI 在计算复杂性和跨数据集泛化能力方面仍面临挑战,但其提供的个性化见解为精准医疗开辟了新途径。

未来, XAI 技术的持续进步与跨学科合作的深化,将是克服现有局限、推动其在乳腺癌诊断乃至更广泛医疗领域广泛应用的关键。随着技术的不断成熟, XAI 有望彻底改变癌症诊断与治疗的面貌,为全球患者带来更加精准、高效且人性化的医疗服务。

## 参考文献

- [1] Lima, S.M., Kehm, R.D. and Terry, M.B. (2021) Global Breast Cancer Incidence and Mortality Trends by Region, Age-Groups, and Fertility Patterns. *E Clinical Medicine*, **38**, Article 100985. <https://doi.org/10.1016/j.eclim.2021.100985>
- [2] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [3] Siegel, R.L., Miller, K.D., Fuchs, H.E. and Jemal, A. (2022) Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, **72**, 7-33. <https://doi.org/10.3322/caac.21708>
- [4] Sechopoulos, I., Teuwen, J. and Mann, R. (2021) Artificial Intelligence for Breast Cancer Detection in Mammography and Digital Breast Tomosynthesis: State of the Art. *Seminars in Cancer Biology*, **72**, 214-225. <https://doi.org/10.1016/j.semcancer.2020.06.002>
- [5] Ren, W., Chen, M., Qiao, Y. and Zhao, F. (2022) Global Guidelines for Breast Cancer Screening: A Systematic Review. *The Breast*, **64**, 85-99. <https://doi.org/10.1016/j.breast.2022.04.003>
- [6] Farkas, A.H. and Nattinger, A.B. (2023) Breast Cancer Screening and Prevention. *Annals of Internal Medicine*, **176**, ITC161-ITC176. <https://doi.org/10.7326/aitc202311210>
- [7] Huntley, C., Torr, B., Sud, A., Houlston, R.S., Hingorani, A.D., Jones, M.E., *et al.* (2023) The Impact of Risk Stratification by Polygenic Risk and Age on Breast Cancer Screening in Women Aged 40-49 Years: A Modelling Study. *The Lancet*, **402**, S54. [https://doi.org/10.1016/s0140-6736\(23\)02103-7](https://doi.org/10.1016/s0140-6736(23)02103-7)
- [8] Pashayan, N., Antoniou, A.C., Ivanus, U., Esserman, L.J., Easton, D.F., French, D., *et al.* (2020) Personalized Early Detection and Prevention of Breast Cancer: ENVISION Consensus Statement. *Nature Reviews Clinical Oncology*, **17**, 687-705. <https://doi.org/10.1038/s41571-020-0388-9>
- [9] Zhang, J., Wu, J., Zhou, X.S., Shi, F. and Shen, D. (2023) Recent Advancements in Artificial Intelligence for Breast Cancer: Image Augmentation, Segmentation, Diagnosis, and Prognosis Approaches. *Seminars in Cancer Biology*, **96**, 11-25. <https://doi.org/10.1016/j.semcancer.2023.09.001>
- [10] Aung, Y.Y.M., Wong, D.C.S. and Ting, D.S.W. (2021) The Promise of Artificial Intelligence: A Review of the Opportunities and Challenges of Artificial Intelligence in Healthcare. *British Medical Bulletin*, **139**, 4-15. <https://doi.org/10.1093/bmb/ldab016>
- [11] Borys, K., Schmitt, Y.A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C.M., *et al.* (2023) Explainable AI in Medical Imaging: An Overview for Clinical Practitioners—Saliency-Based XAI Approaches. *European Journal of Radiology*, **162**, Article 110787. <https://doi.org/10.1016/j.ejrad.2023.110787>
- [12] Karim, M.R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., *et al.* (2023) Explainable AI for Bioinformatics: Methods, Tools and Applications. *Briefings in Bioinformatics*, **24**, bbad236. <https://doi.org/10.1093/bib/bbad236>
- [13] Gurmessa, D.k. and Jimma, W. (2024) Explainable Machine Learning for Breast Cancer Diagnosis from Mammography and Ultrasound Images: A Systematic Review. *BMJ Health & Care Informatics*, **31**, e100954. <https://doi.org/10.1136/bmjhci-2023-100954>
- [14] Gao, Y., Lin, J., Zhou, Y. and Lin, R. (2023) The Application of Traditional Machine Learning and Deep Learning Techniques in Mammography: A Review. *Frontiers in Oncology*, **13**, Article 1213045. <https://doi.org/10.3389/fonc.2023.1213045>
- [15] Jin, X., Liu, C., Xu, T., Su, L. and Zhang, X. (2020) Artificial Intelligence Biosensors: Challenges and Prospects. *Bio-sensors and Bioelectronics*, **165**, Article 112412. <https://doi.org/10.1016/j.bios.2020.112412>
- [16] Abas Mohamed, Y., Ee Khoo, B., Shahrimie Mohd Asaari, M., Ezane Aziz, M. and Rahiman Ghazali, F. (2025) Decoding the Black Box: Explainable AI (XAI) for Cancer Diagnosis, Prognosis, and Treatment Planning-A State-of-the Art Systematic Review. *International Journal of Medical Informatics*, **193**, Article 105689. <https://doi.org/10.1016/j.ijmedinf.2024.105689>
- [17] Song, D., Yao, J., Jiang, Y., Shi, S., Cui, C., Wang, L., *et al.* (2023) A New Xai Framework with Feature Explainability for Tumors Decision-Making in Ultrasound Data: Comparing with Grad-CAM. *Computer Methods and Programs in Biomedicine*, **235**, Article 107527. <https://doi.org/10.1016/j.cmpb.2023.107527>
- [18] Rodríguez-Pérez, R. and Bajorath, J. (2019) Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *Journal of Medicinal Chemistry*, **63**, 8761-8777. <https://doi.org/10.1021/acs.jmedchem.9b01101>
- [19] Zaric, O., Hatamikia, S., George, G., Schwarzhans, F., Trattinig, S. and Woitek, R. (2023) AI-Based Time-Intensity-Curve Assessment of Breast Tumors on MRI. *European Radiology*, **34**, 179-181. <https://doi.org/10.1007/s00330-023-10298-8>
- [20] Khalid, A., Mehmood, A., Alabrah, A., Alkhamees, B.F., Amin, F., AlSalman, H., *et al.* (2023) Breast Cancer Detection

- and Prevention Using Machine Learning. *Diagnostics*, **13**, Article 3113. <https://doi.org/10.3390/diagnostics13193113>
- [21] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Benetot, A., Tabik, S., Barbado, A., *et al.* (2020) Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*, **58**, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
  - [22] Li, J., Sheng, D., Chen, J., You, C., Liu, S., Xu, H., *et al.* (2023) Artificial Intelligence in Breast Imaging: Potentials and Challenges. *Physics in Medicine & Biology*, **68**, 23TR01. <https://doi.org/10.1088/1361-6560/acfade>
  - [23] Woitek, R. and Brindle, K.M. (2023) Hyperpolarized Carbon-13 MRI in Breast Cancer. *Diagnostics*, **13**, Article 2311. <https://doi.org/10.3390/diagnostics13132311>
  - [24] Dong, F., She, R., Cui, C., Shi, S., Hu, X., Zeng, J., *et al.* (2021) One Step Further into the Blackbox: A Pilot Study of How to Build More Confidence around an AI-Based Decision System of Breast Nodule Assessment in 2D Ultrasound. *European Radiology*, **31**, 4991-5000. <https://doi.org/10.1007/s00330-020-07561-7>
  - [25] Zhang, J., Cui, Z., Shi, Z., Jiang, Y., Zhang, Z., Dai, X., *et al.* (2023) A Robust and Efficient AI Assistant for Breast Tumor Segmentation from DCE-MRI via a Spatial-Temporal Framework. *Patterns*, **4**, Article 100826. <https://doi.org/10.1016/j.patter.2023.100826>
  - [26] Lowry, K.P., Coley, R.Y., Miglioretti, D.L., Kerlikowske, K., Henderson, L.M., Onega, T., *et al.* (2020) Screening Performance of Digital Breast Tomosynthesis vs Digital Mammography in Community Practice by Patient Age, Screening Round, and Breast Density. *JAMA Network Open*, **3**, e2011792. <https://doi.org/10.1001/jamanetworkopen.2020.11792>
  - [27] Oh, J., Kim, H., Lee, K.M. and Ryu, C. (2022) Reliable Quality Assurance of X-Ray Mammography Scanner by Evaluation of the Standard Mammography Phantom Image Using an Interpretable Deep Learning Model. *European Journal of Radiology*, **154**, Article 110369. <https://doi.org/10.1016/j.ejrad.2022.110369>
  - [28] Schaffter, T., Buist, D.S.M., Lee, C.I., *et al.* (2020) Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Network Open*, **3**, e200265.
  - [29] Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., *et al.* (2019) Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison with 101 Radiologists. *Journal of the National Cancer Institute*, **111**, 916-922. <https://doi.org/10.1093/jnci/djy222>
  - [30] Sun, J., Sun, C., Tang, Y., Liu, T. and Lu, C. (2023) Application of SHAP for Explainable Machine Learning on Age-Based Subgrouping Mammography Questionnaire Data for Positive Mammography Prediction and Risk Factor Identification. *Healthcare*, **11**, Article 2000. <https://doi.org/10.3390/healthcare11142000>
  - [31] Berg, W.A., Bandos, A.I., Mendelson, E.B., Lehrer, D., Jong, R.A. and Pisano, E.D. (2015) Ultrasound as the Primary Screening Test for Breast Cancer: Analysis from ACRIN 6666. *Journal of the National Cancer Institute*, **108**, djv367. <https://doi.org/10.1093/jnci/djv367>
  - [32] Zhang, B., Vakanski, A. and Xian, M. (2023) BI-RADS-NET-V2: A Composite Multi-Task Neural Network for Computer-Aided Diagnosis of Breast Cancer in Ultrasound Images with Semantic and Quantitative Explanations. *IEEE Access*, **11**, 79480-79494. <https://doi.org/10.1109/access.2023.3298569>
  - [33] Shamir, S.B., Sasson, A.L., Margolies, L.R. and Mendelson, D.S. (2024) New Frontiers in Breast Cancer Imaging: The Rise of AI. *Bioengineering*, **11**, Article 451. <https://doi.org/10.3390/bioengineering11050451>
  - [34] Feig, S. (2010) Cost-Effectiveness of Mammography, MRI, and Ultrasonography for Breast Cancer Screening. *Radio-logic Clinics of North America*, **48**, 879-891. <https://doi.org/10.1016/j.rcl.2010.06.002>
  - [35] Hirsch, L., Huang, Y., Makse, H.A., Martinez, D.F., Hughes, M., Eskreis-Winkler, S., *et al.* (2024) Early Detection of Breast Cancer in MRI Using AI. *Academic Radiology*, **20**, 1-8. <https://doi.org/10.1016/j.acra.2024.10.014>
  - [36] Diakogiannis, F.I., Waldner, F., Caccetta, P. and Wu, C. (2020) ResuNet-A: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **162**, 94-114. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>
  - [37] Salim, M., Liu, Y., Sorkhei, M., Ntola, D., Foukakis, T., Fredriksson, I., *et al.* (2024) AI-Based Selection of Individuals for Supplemental MRI in Population-Based Breast Cancer Screening: The Randomized Screentrustmri Trial. *Nature Medicine*, **30**, 2623-2630. <https://doi.org/10.1038/s41591-024-03093-5>
  - [38] Qin, Y., Wu, F., Hu, Q., He, L., Huo, M., Tang, C., *et al.* (2023) Histogram Analysis of Multi-Model High-Resolution Diffusion-Weighted MRI in Breast Cancer: Correlations with Molecular Prognostic Factors and Subtypes. *Frontiers in Oncology*, **13**, Article 1139189. <https://doi.org/10.3389/fonc.2023.1139189>
  - [39] Zhang, Y., Weng, Y. and Lund, J. (2022) Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics*, **12**, Article 237. <https://doi.org/10.3390/diagnostics12020237>
  - [40] Lo Gullo, R., Marcus, E., Huayanay, J., Eskreis-Winkler, S., Thakur, S., Teuwen, J., *et al.* (2024) Artificial Intelligence-Enhanced Breast MRI. *Investigative Radiology*, **59**, 230-242. <https://doi.org/10.1097/rli.0000000000001010>

- 
- [41] Meyer - Base, A., Morra, L., Tahmassebi, A., Lobbes, M., Meyer-Base, U. and Pinker, K. (2020) AI-Enhanced Diagnosis of Challenging Lesions in Breast MRI: A Methodology and Application Primer. *Journal of Magnetic Resonance Imaging*, **54**, 686-702. <https://doi.org/10.1002/jmri.27332>
  - [42] Brunetti, N., Calabrese, M., Martinoli, C. and Tagliafico, A.S. (2022) Artificial Intelligence in Breast Ultrasound: From Diagnosis to Prognosis—A Rapid Review. *Diagnostics*, **13**, Article 58. <https://doi.org/10.3390/diagnostics13010058>
  - [43] Borys, K., Schmitt, Y.A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C.M., *et al.* (2023) Explainable AI in Medical Imaging: An Overview for Clinical Practitioners—Beyond Saliency-Based XAI Approaches. *European Journal of Radiology*, **162**, Article 110786. <https://doi.org/10.1016/j.ejrad.2023.110786>
  - [44] Wang, X., Chou, K., Zhang, G., Zuo, Z., Zhang, T., Zhou, Y., *et al.* (2023) Breast Cancer Pre-Clinical Screening Using Infrared Thermography and Artificial Intelligence: A Prospective, Multicentre, Diagnostic Accuracy Cohort Study. *International Journal of Surgery*, **109**, 3021-3031. <https://doi.org/10.1097/js9.0000000000000594>
  - [45] Pfeuffer, N., Baum, L., Stammer, W., Abdel-Karim, B.M., Schramowski, P., Bucher, A.M., *et al.* (2023) Explanatory Interactive Machine Learning. *Business & Information Systems Engineering*, **65**, 677-701. <https://doi.org/10.1007/s12599-023-00806-x>
  - [46] Wang, W. and Wang, Y. (2023) Deep Learning-Based Modified YOLACT Algorithm on Magnetic Resonance Imaging Images for Screening Common and Difficult Samples of Breast Cancer. *Diagnostics*, **13**, Article 1582. <https://doi.org/10.3390/diagnostics13091582>