

人工智能驱动甲状腺癌超声诊断与风险分层的临床实践

伍晓妮

湘雅常德医院 湖南 常德 415000

【摘要】：甲状腺结节的超声评估已经形成较成熟的风险分层体系，但实际判读仍受到切面选择、图像质量和操作者经验影响。人工智能进入甲状腺超声后，应用重点已由单幅图像分类逐步转向征象识别、TI-RADS 分级、穿刺选择及颈部淋巴结评估。现有文献表明，多中心训练、多视图分析和多模态信息整合能够补充传统超声判读，不过模型在不同设备、特殊病理类型和动态扫查场景中的稳定性仍需验证。本文结合甲状腺结节的实际诊疗环节，对 AI 在图像识别、风险分层、穿刺决策和术前评估中的应用进行梳理，并讨论模型解释、外部验证和临床部署中的关键问题。现阶段，AI 更适合作为超声医师的辅助工具，其结果仍需与实时扫查、病理结果及患者个体风险共同分析。

【关键词】：甲状腺癌；人工智能；超声；深度学习；TI-RADS；风险分层

DOI:10.12417/2811-051X.26.12.084

1 超声风险分层的技术挑战

甲状腺超声的任务并不止于发现结节。成分、回声、纵横比、边界和强回声灶等征象最终要落实到风险等级，再进一步决定随访或细针穿刺。ACR TI-RADS 已经给出了较清晰的分级框架^[1]，但同一结节在不同操作者、不同切面下仍可能得到不完全一致的描述。动态扫查本身也是影响判断的重要环节，静态图像很难完整保留探头移动过程中获得的信息。

AI 更容易介入的是那些可重复识别的图像征象。例如结节轮廓、径线、回声和钙化等特征均可被模型量化，但这些结果并不能脱离原始图像单独解释。对于边界模糊、体积较小或位置特殊的结节，重新扫查往往仍是确认结果的必要步骤。由此看，AI 对甲状腺超声的价值更接近“增加一次结构化复核”，而不是改变现有诊断标准。

2 多中心与多视图模型的结合应用

Chen 等采用多中心数据训练深度学习模型，用于甲状腺实性结节成分分类和良恶性鉴别^[2]。与单中心数据相比，这类设计能够在一定程度上增加病例和设备来源的多样性。Vahdati 等则将同一结节的多个超声切面同时纳入模型分析^[3]。但两项技术均无法解决甲状腺超声依赖于单图诊断问题。

多视图分析尤其贴近超声检查方式。医师通常需要从横切面、纵切面及其他角度反复观察结节，最大径、包膜关系和边界状态也可能随切面改变。对多发结节而言，AI 还可用于病灶定位、径线记录和随访匹配，减少前后检查中结节对应错误。自动测量能够提高效率，但只要切面选择发生变化，测量结果仍可能出现偏差，因此人工复核不能省略。

3 TI-RADS 分级与穿刺决策

如果模型只输出“良性”或“恶性”，它与临床流程之间仍存在距离。TI-RADS 分级不仅反映恶性风险，还直接关联是否随访、是否穿刺以及检查间隔。Zhou 等将 AI 用于甲状腺结节超声征象识别和细针穿刺决策^[4]，说明算法输出可以与现有分级体系直接对接。

这类系统是否好用，取决于结果能否被医师核对。单独给出恶性概率时，风险来源并不清楚；若系统同时显示结节轮廓、主要可疑征象和建议分级，医师可以回到原始图像逐项确认。AI 与人工分级不一致时，也不必简单选择其中一方。对结节重新扫查并检查分歧征象，更符合超声诊断的工作方式。

穿刺与否并不是单纯依据模型给出的恶性概率。结节大小、患者年龄以及随访过程中病灶是否发生变化，都会影响后续处理。对于体积较小且超声风险较低的结节，即使模型给出较高风险值，超声医师仍需结合现有穿刺指征进行判断。模型结果在这里更多是增加一项参考信息，而不是直接决定是否操作。

4 原发灶及淋巴结的多模态识别

乳头状甲状腺癌明确后，颈部淋巴结也成为术前超声检查的重要内容。Wang 等汇总了超声 AI 预测颈部淋巴结转移的相关资料，不同文献在病例数量、模型设置以及验证方法上并不一致^[5]。临床扫查时，模型结果可以作为寻找可疑淋巴结的一项补充，但淋巴结形态、内部回声、血流情况以及必要时的穿刺结果仍需由超声医师结合实际图像分析。

除了常规灰阶图像，已有工作开始加入弹性、血流等其他超声信息。Li 等比较了多参数超声与 AI 辅助灰阶超声，在部

作者简介：伍晓妮，女（1992—），硕士，主治医师。研究方向：超声医学。

项目基金：2025 年度常德市科技创新指导性项目，2025ZD199。

分实性结节中观察到多参数超声具有较好的诊断表现^[6]。Athreya 等采用了另一种思路,将超声影像和分子检测同时用于不确定结节的风险分析^[7]。两项工作的技术路线并不相同,但都没有把判断局限在单一灰阶图像上。

桥本甲状腺炎背景下,腺体本身可能出现弥漫性回声改变,结节边界和内部回声的判断随之变得复杂。滤泡性肿瘤、囊实性结节以及术后颈部结构改变,也与常见乳头状癌图像存在差别。这些病例在训练数据中所占比例如果较低,模型在总体病例中的表现就不能直接代表其对特殊病例的识别情况,因此仍有必要分别观察不同亚组的结果。

5 临床部署更需要看稳定性和可用性

模型进入医院以后,设备厂家、图像后处理方式和病例构成都可能影响实际表现。单中心内部验证良好,只能说明模型在相近数据环境下有效。真正推广前,还应进行独立外部验证,并以患者为单位划分训练集和验证集,避免同一患者的不同图像进入不同数据集。

临床评价也应与使用场景对应。除了灵敏度、特异度和

AUC,还可比较不同 TI-RADS 类别中的漏诊率、穿刺减少比例以及医师使用 AI 前后的分级一致性。 workflow 同样不能忽略。若使用 AI 需要额外上传图像、重复勾画或长时间等待结果,算法性能再好,也可能难以长期应用。

设备升级和病例构成变化还会使后续数据逐渐偏离原始训练集。医院部署后可按 TI-RADS 类别和病理类型定期抽样复核,发现误判增加时再进行校准。涉及穿刺、手术等重要决策,最终判断仍应由临床医师结合完整超声检查和病理证据完成。

6 结语

甲状腺超声 AI 目前已经覆盖结节识别、TI-RADS 分级、穿刺选择、淋巴结评估和多模态风险判断,但不同环节的成熟程度并不一致。图像识别较容易标准化,涉及动态扫查、特殊病理类型和治疗决策时,人工判断仍不可替代。后续工作的重点不宜只放在提高模型指标,还应验证其不同设备、不同病例构成和真实诊疗流程中的稳定性,并观察 AI 介入后是否真正改善随访、穿刺和术前评估。

参考文献:

- [1] Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol*, 2017, 14(5): 587-595.
- [2] Chen C, Jiang Y, Yao J, et al. Deep learning to assist composition classification and thyroid solid nodule diagnosis: a multicenter diagnostic study. *Eur Radiol*, 2024, 34(4): 2323-2333.
- [3] Vahdati S, Khosravi B, Robinson KA, et al. A multi-view deep learning model for thyroid nodules detection and characterization in ultrasound imaging. *Bioengineering*, 2024, 11(7): 648.
- [4] Zhou T, Xu L, Shi J, et al. US of thyroid nodules: can AI-assisted diagnostic system compete with fine needle aspiration? *Eur Radiol*, 2024, 34(2): 1324-1333.
- [5] Wang X, Qi Y, Zhang X, et al. Ultrasound-based artificial intelligence for predicting cervical lymph node metastasis in papillary thyroid cancer: a systematic review and meta-analysis. *Front Endocrinol*, 2025, 16: 1570811.
- [6] Li Y, Li X, Luo Y, et al. Multiparametric ultrasound techniques are superior to AI-assisted ultrasound for assessment of solid thyroid nodules: a prospective study. *Endocrine*, 2025, 90(1): 225-236.
- [7] Athreya S, Melehy A, Suthahar SSA, et al. Combining ultrasound imaging and molecular testing in a multimodal deep learning model for risk stratification of indeterminate thyroid nodules. *Thyroid*, 2025, 35(5): 590-594.