

儿童和青年甲状腺结节的超声诊断：影像医师整体印象、ACR TI-RADS 评分和深度学习算法的诊断效能比较

摘要

背景：在目前的临床实践中，儿童甲状腺结节的评估通常基于影像医师对超声图像的整体印象。

目的：为了比较影像医师整体印象、美国放射学院甲状腺影像报告和数据系统（ACR TI-RADS）和深度学习算法在超声鉴别儿童和青年甲状腺良恶性结节的诊断效能。

方法：这项回顾性研究纳入 2004 年 1 月 1 日至 2020 年 9 月 18 日期间年龄 ≤ 21 岁且超声诊断甲状腺结节的患者，以细针穿刺和/或手术切除的明确病理结果作为参考标准，最终纳入了 139 例患者（中位年龄 17.5 岁；女性 119 例，男性 20 例）。每位患者选择一个甲状腺结节，获取该结节的单张横向和纵向图像进行进一步评估。三位影像医师根据其整体印象（良性 vs 恶性）和 ACR TI-RADS 评分独立地对结节进行评价。同时利用先前开发的深度学习算法确定每个结节的恶性可能性，用于推导风险水平。计算不同方法诊断恶性肿瘤的敏感性和特异性。采用 kappa 系数对一致性进行评估。

结果：影像医师整体印象的诊断敏感性为 32.1%~75.0%（平均 58.3%，95% CI：49.2~67.3%），特异性为 63.8%~93.9%（平均 79.9%，95% CI：73.8~85.7%）。ACR TI-RADS 的诊断敏感性为 82.1%~87.5%（平均 85.1%，95% CI：77.3~92.1%），特异性为 47.0%~54.2%（平均 50.6%，95% CI：41.4~59.8%）。深度学习算法的敏感性为 87.5%（95% CI：78.3~95.5%），特异性为 36.1%（95% CI：25.6~46.8%）。两两组合的观察者间一致性，整体印象的 kappa 值为 0.227~0.472，ACR TI-RADS 的 kappa 值为 0.597~0.643。

结论：与整体印象相比，ACR TI-RADS 和深度学习算法都具有较高的敏感性，但特异性较低。深度学习算法具有与 ACR TI-RADS 相近的敏感性，但特异性较低。ACR TI-RADS 的观察者间一致性高于影像医师整体印象。

临床意义：ACR TI-RADS 和深度学习算法或许可以作为指导儿童甲状腺结节进行细针穿刺决策的潜在替代策略。