

人工智能超声辅助诊断系统鉴别诊断甲状腺结节良恶性的临床价值

黄奎^{1,3}, 查倩³, 应育娟³, 胡紫玥², 卢漫^{1,2}

摘要 **目的** 探讨人工智能超声辅助诊断系统(AI-UADS)鉴别诊断甲状腺结节良恶性的临床价值。**方法** 选取我院经手术病理确诊的甲状腺结节患者 489 例, 共计 630 个结节, 其中良性结节 317 个, 恶性结节 313 个。应用 AI-SONIC Thyroid 软件获取甲状腺结节恶性概率, 以此鉴别甲状腺结节良恶性, 并与超声医师诊断结果进行比较; 比较 AI-UADS 与超声医师鉴别不同大小、位置、结节与被膜关系及是否伴桥本甲状腺炎(HT)甲状腺结节良恶性的诊断效能。**结果** 630 个甲状腺结节中, AI-UADS 诊断良性结节 283 个、恶性结节 347 个, 准确诊断良性结节 268 个、恶性结节 298 个, 其鉴别诊断甲状腺结节良恶性的曲线下面积(AUC)、准确率、灵敏度、特异度、阳性预测值、阴性预测值分别为 0.939、89.8%、95.2%、84.5%、85.9%、94.7%, 均高于超声医师(0.875、83.8%、89.5%、78.2%、80.2%、88.3%), 差异均有统计学意义(均 $P < 0.05$)。AI-UADS 鉴别诊断最大径 ≤ 10 mm 甲状腺结节良恶性的准确率、特异度、阳性预测值均高于超声医师, 鉴别诊断最大径 > 20 mm 甲状腺结节良恶性的灵敏度、阴性预测值均高于超声医师, 特异度低于超声医师, 差异均有统计学意义(均 $P < 0.05$); AI-UADS 与超声医师鉴别诊断最大径 10~20 mm 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。AI-UADS 鉴别诊断左侧叶甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值均高于超声医师, 差异均有统计学意义(均 $P < 0.05$); AI-UADS 与超声医师鉴别诊断右侧叶、峡部甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。AI-UADS 鉴别诊断远离被膜的甲状腺结节良恶性的准确率、特异度、阳性预测值均高于超声医师, 鉴别诊断侵犯被膜的甲状腺结节良恶性的准确率、灵敏度、阴性预测值均高于超声医师, 差异均有统计学意义(均 $P < 0.05$); AI-UADS 与超声医师鉴别诊断紧邻被膜的甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。AI-UADS 鉴别诊断不伴 HT 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值均高于超声医师, 差异均有统计学意义(均 $P < 0.05$); AI-UADS 与超声医师鉴别诊断伴 HT 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。**结论** AI-UADS 在鉴别甲状腺结节良恶性中具有较高的诊断效能, 尤其对于位于左侧叶、最大径 ≤ 10 mm、远离被膜和侵犯被膜, 以及不伴 HT 甲状腺结节良恶性的诊断效能更高。

关键词 超声检查; 甲状腺结节; 良恶性; 人工智能

[中图法分类号] R445.1; R736.1

[文献标识码] A

Clinical value of artificial intelligence ultrasound-assisted diagnostic system in differentiating benign and malignant thyroid nodules

HUANG Kui^{1,3}, ZHA Qian³, YING Yujuan³, HU Ziyue², LU Man^{1,2}

1.School of Imaging, North Sichuan Medical College, Nanchong 637500, China. 2.Department of Ultrasound Medicine Center, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital&Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu 610000, China. 3.Department of Ultrasound Medicine, Zigong Fourth People's Hospital, Zigong 643200, China

ABSTRACT Objective To explore the clinical value of artificial intelligence ultrasound-assisted diagnostic system

基金项目: 自贡市卫健委科研项目(21yb059)

作者单位: 1. 川北医学院影像学院, 四川 南充 637500; 2. 四川省肿瘤医院·研究所 四川省肿瘤临床医学研究中心 四川省癌症防治中心 电子科技大学附属肿瘤医院超声医学中心, 四川 成都 610000; 3. 自贡市第四人民医院超声医学科, 四川 自贡 643200

通讯作者: 卢漫, Email: graceof@163.com

(AI-UADS) in differentiating benign and malignant thyroid nodules. **Methods** A total of 489 patients with thyroid nodules confirmed by surgery and pathology in our hospital were selected, with a total of 630 nodules, including 317 benign nodules and 313 malignant nodules. AI-SONIC Thyroid software was used to obtain the malignant probability of thyroid nodules for differentiating benign and malignant thyroid nodules, and its diagnostic results were compared with those of sonographers. The diagnostic efficacy of AI-UADS and sonographers in differentiating benign and malignant thyroid nodules with different sizes, locations, nodules and capsule relationships, and with or without Hashimoto's thyroiditis (HT) was compared. **Results** Among 630 thyroid nodules, 283 benign nodules and 347 malignant nodules were diagnosed by AI-UADS, and 268 benign nodules and 298 malignant nodules were accurately diagnosed. The area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value and negative predictive value for differentiating benign and malignant thyroid nodules were 0.939, 89.8%, 95.2%, 84.5%, 85.9% and 94.7%, respectively, which were higher than those of sonographers (0.875, 83.8%, 89.5%, 78.2%, 80.2% and 88.3%), with statistically significant differences (all $P < 0.05$). The accuracy, specificity and positive predictive of AI-UADS for differentiating benign and malignant thyroid nodules with a maximum diameter of ≤ 10 mm were higher than those of sonographers, the sensitivity and negative predictive value of AI-UADS for differentiating benign and malignant thyroid nodules with a maximum diameter of > 20 mm were higher than those of sonographers, and the specificity was lower than that of sonographers, the differences were statistically significant (all $P < 0.05$). There were no statistically significant differences in accuracy, sensitivity, specificity, positive predictive value, and negative predictive value for differentiating benign and malignant thyroid nodules with a maximum diameter of 10~20 mm between AI-UADS and sonographers. The accuracy, sensitivity, specificity, positive predictive value and negative predictive value of AI-UADS for differentiating benign and malignant thyroid nodules in the left lobe were higher than those of sonographers, the differences were statistically significant (all $P < 0.05$). There were no statistically significant differences in accuracy, sensitivity, specificity, positive predictive value and negative predictive value for differentiating benign and malignant thyroid nodules in the right lobe and isthmus between AI-UADS and sonographers. The accuracy, specificity and positive predictive value of AI-UADS for differentiating benign and malignant thyroid nodules far from the capsule were higher than those of sonographers, and the accuracy, sensitivity and negative predictive value of AI-UADS for differentiating benign and malignant thyroid nodules invading the capsule were higher than those of sonographers, the differences were statistically significant (all $P < 0.05$). There were no statistically significant differences in accuracy, sensitivity, specificity, positive predictive value and negative predictive value for differentiating benign and malignant thyroid nodules adjacent to the capsule between AI-UADS and sonographers. The accuracy, sensitivity, specificity, positive predictive value and negative predictive value of AI-UADS for differentiating benign and malignant thyroid nodules without HT were higher than those of sonographers, the differences were statistically significant (all $P < 0.05$). There were no statistically significant differences in accuracy, sensitivity, specificity, positive predictive value and negative predictive value for differentiating benign and malignant thyroid nodules with HT between AI-UADS and sonographers. **Conclusion** AI-UADS has high diagnostic efficacy for differentiating benign and malignant thyroid nodules, especially for thyroid nodules located in the left lobe, with a maximum diameter of ≤ 10 mm, far from the capsule and invaded the capsule, and without HT.

KEY WORDS Ultrasonography; Thyroid nodules, benign and malignant; Artificial intelligence

随着超声技术的快速发展和健康体检的普及,甲状腺结节的检出率越来越高^[1],这给超声医师带来了沉重的诊断负担。超声作为甲状腺结节首选的影像学检查方法^[2],可通过分析结节的声像图特征实现恶性风险分层。然而良恶性结节部分声像图特征重叠,导致不同声像图特征对良恶性结节的诊断灵敏度和特异度参差不齐^[3],增加了鉴别诊断难度;且诊断过程存在明显的观察者间变异和观察者内变异^[4],不同地区间医疗水平差异也会影响诊断的准确性和一致性。而基于深度学习的人工智能超声辅助诊断系统

(artificial intelligence ultrasound-assisted diagnosis system, AI-UADS)具有甲状腺结节自动分割、特征识别和良恶性预测三大核心功能。该技术通过计算机视觉算法,系统性地提取和分析传统超声图像中的多维特征,其诊断过程具有高度可重复性和同质性^[5]。本研究通过比较 AI-UADS 与超声医师鉴别不同大小、位置、结节与被膜关系及是否伴桥本甲状腺炎 (Hashimoto's thyroiditis, HT) 的甲状腺结节良恶性的诊断效能,探讨 AI-UADS 鉴别诊断甲状腺结节良恶性的临床价值。

资料与方法

一、研究对象

选取 2023 年 1 月至 2023 年 12 月自贡市第四人民医院经手术病理确诊的甲状腺结节患者 489 例,其中男 103 例,女 386 例,年龄 18~82 岁,平均 (48.9 ± 12.8) 岁。单发结节 399 例,多发结节 90 例,共计 630 个结节,最大径 3~77 mm,中位数 11.0(7.0, 21.0) mm;恶性结节 313 个,其中乳头状癌 312 个,髓样癌 1 个;良性结节 317 个,其中滤泡性结节性疾病 246 个,滤泡性腺瘤 53 个,炎性病灶 13 个(包括淋巴细胞性甲状腺炎 8 个、亚急性肉芽肿性甲状腺炎和亚急性甲状腺炎各 2 个、肉芽肿性甲状腺炎 1 个),嗜酸细胞性腺瘤 5 个。纳入标准:①目标结节有明确的术后组织病理学结果;②留存标准横切面及纵切面超声图像。排除标准:①临床及超声资料不完整;②AI-UADS 识别错误,包括任意切面未识别、多个结节识别为 1 个结节、1 个结节识别为多个结节、结节识别范围不符合等;③既往接受甲状腺结节手术或消融治疗;④甲状腺转移性肿瘤。本研究经医院医学伦理委员会批准(批准号:2022-059),为回顾性研究故免除患者知情同意。

二、仪器与方法

1. 超声检查:使用 Philips EPIQ5、迈瑞 Resona 7S、西门子 Acuson Sequoia 彩色多普勒超声诊断仪, L12-5 探头,频率 5~12 MHz; L14-5WU 探头,频率 5~14 MHz; 18L6 探头,频率 6~18 MHz。患者取仰卧位,充分暴露颈部甲状腺区域,选择预设的甲状腺检查条件,将目标结节置于图像中央,观察其数目、位置、成分、回声、形态、边缘、钙化,以及其与被膜距离及被膜是否光滑、完整;多发结节者检查时需做好位置、大小等标记。于纵切面测量结节最大径;留存静态及动态超声图像,由 3 名具有 5 年以上甲状腺检查经验的超声医师在不知病理结果及 AI-SONIC Thyroid 软件(浙江大学数理学院)诊断结果的情况下,分析结节的 5 项声像图特征(成分、回声、形态、边缘、钙化),以 ACR-TIRADS 标准^[6]分别对其进行评分;如有分歧则协商解决。

2. AI-UADS 诊断:将上述留存的静态及动态超声图像同步导入 AI-SONIC Thyroid 软件,选择内置的 ACR-TIRADS 分类诊断标准,自动输出界面取样框颜色、ACR-TIRADS 分类信息及恶性概率,恶性概率为 0~1,越接近 1,结节的恶性可能性越大。以上操作均由同一具有 14 年工作经验的超声医师完成。

3. 亚组分析:①大小亚组,根据结节最大径将结节分为 ≤ 10 mm、10~20 mm、 > 20 mm;②位置亚组,以横切面为参考,将结节分为右侧叶、峡部、左侧叶;③结节与被膜关系亚组,根据结节距被膜距离及被膜是否光滑、完整^[7],将结节分为结节侵犯被膜(结节距被膜 < 2 mm 伴被膜回声中断),紧邻被膜(结节距被膜 < 2 mm,被膜连续光滑)、远离被膜(结节距被膜 > 2 mm);④参考术后病理,依据结节背景分为伴 HT、不伴 HT。比较各亚组 AI-UADS 与超声医师鉴别诊断甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值。

三、统计学处理

应用 RStudio 软件(版本:2023.09.1+494)对数据进行统计分析。计数资料以频数或率表示,组间比较采用 χ^2 检验或 Fisher 精确检验。绘制受试者工作特征(receiver operating characteristic, ROC)曲线确定 AI-UADS、超声医师鉴别甲状腺结节良恶性的截断值并分析其诊断效能;曲线下面积(AUC)比较采用 DeLong 检验。 $P < 0.05$ 为差异有统计学意义。

结 果

一、超声医师及 AI-UADS 诊断标准

ROC 曲线分析显示,超声医师鉴别诊断甲状腺结节良恶性的截断值为 4.5 分,对应的 AUC 为 0.875(95% 可信区间:0.846~0.902, $P < 0.001$); AI-UADS 鉴别诊断甲状腺结节良恶性的截断值为 0.64,对应的 AUC 为 0.939(95% 可信区间:0.919~0.958, $P < 0.001$),高于超声医师,差异有统计学意义($P < 0.05$)。见图 1。

二、AI-UADS 及超声医师诊断情况

630 个甲状腺结节中,超声医师诊断良性结节 281 个、恶性结节 349 个,准确诊断良性结节 248 个、恶性结节 280 个;其鉴别诊断甲状腺结节良恶性的准确

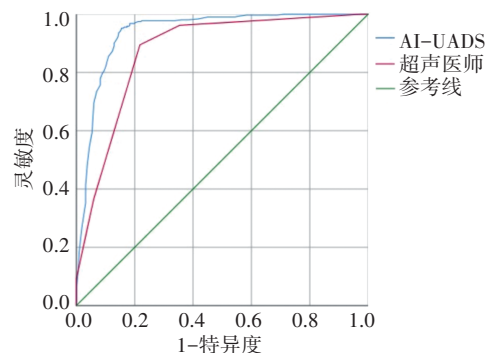


图1 AI-UADS及超声医师鉴别诊断甲状腺结节良恶性的ROC曲线图

率、灵敏度、特异度、阳性预测值、阴性预测值分别为83.8%、89.5%、78.2%、80.2%、88.3%。AI-UADS 诊断良性结节283个、恶性结节347个,准确诊断良性结节268个、恶性结节298个;其鉴别诊断甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值分别为89.8%、95.2%、84.5%、85.9%、94.7%,均高于超声医师,差异均有统计学意义(均 $P<0.05$)。见表1、2。

表1 超声医师诊断结果与病理对照 个

超声医师	病理		合计
	良性	恶性	
良性	248	33	281
恶性	69	280	349
合计	317	313	630

表2 AI-UADS诊断结果与病理对照 个

AI-UADS	病理		合计
	良性	恶性	
良性	268	15	283
恶性	49	298	347
合计	317	313	630

三、AI-UADS与超声医师鉴别不同亚组甲状腺结节良恶性的诊断效能比较

1. AI-UADS与超声医师鉴别不同大小甲状腺结节良恶性的诊断效能比较:630个甲状腺结节中,最大径 ≤ 10 mm 314个(恶性结节202个,良性结节112个),10~20 mm 151个(恶性结节69个,良性结节82个), >20 mm 165个(恶性结节29个,良性结节136个)。AI-UADS鉴别诊断最大径 ≤ 10 mm 甲状腺结节良恶性的准确率、特异度、阳性预测值均高于超声医师,鉴别诊断最大径 >20 mm 甲状腺结节良恶性的灵敏度、阴性预测值均高于超声医师,特异度低于超声医师,差异均有统计学意义(均 $P<0.05$);AI-UADS与超声医师鉴别诊断最大径10~20 mm 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。见表3和图2~5。

表3 AI-UADS与超声医师鉴别不同大小甲状腺结节良恶性的诊断效能比较 %

分类	最大径 ≤ 10 mm				
	准确率	灵敏度	特异度	阳性预测值	阴性预测值
AI-UADS	89.2	96.0	76.8	88.2	91.5
超声医师	81.8	97.0	54.5	79.4	91.0
χ^2 值	6.798	0.296	12.369	6.588	0.010
P值	0.009	0.586	<0.001	0.010	0.921

分类	最大径10~20 mm				
	准确率	灵敏度	特异度	阳性预测值	阴性预测值
AI-UADS	88.7	97.6	78.3	84.2	96.4
超声医师	84.8	92.7	75.4	81.7	89.7
χ^2 值	1.037	—	0.163	0.206	—
P值	0.308	1.000	0.687	0.650	0.272

分类	最大径 >20 mm				
	准确率	灵敏度	特异度	阳性预测值	阴性预测值
AI-UADS	92.1	82.8	94.1	75.0	96.2
超声医师	86.7	27.6	99.3	88.9	86.6
χ^2 值	2.589	17.846	—	—	8.254
P值	0.108	<0.001	0.036	0.654	0.004

—示采用 Fisher 精确检验

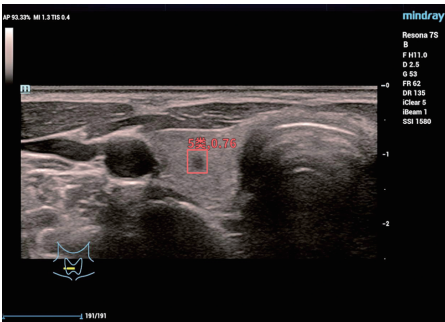


图2 甲状腺乳头状癌患者(女,52岁)二维超声图,结节位于右侧叶,最大径为3 mm,结节距被膜4 mm;AI-UADS提示ACR-TIRADS 5类,恶性概率0.76

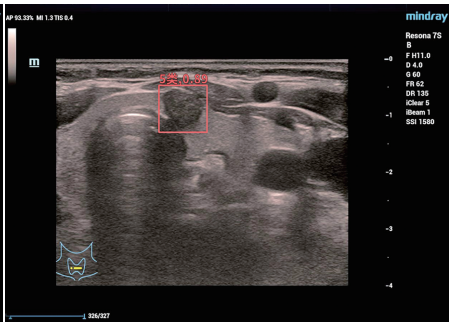


图3 甲状腺乳头状癌患者(女,50岁)二维超声图,结节位于峡部,最大径为8 mm,结节距被膜 <2 mm,被膜回声中断;AI-UADS提示ACR-TIRADS 5类,恶性概率0.89

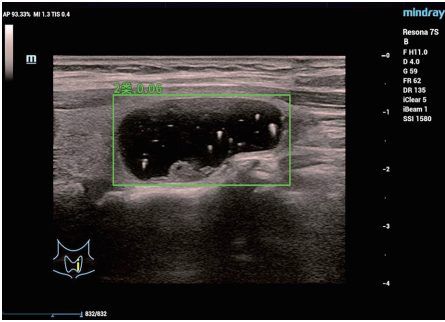


图4 甲状腺滤泡性结节性患者(女,48岁)二维超声图,结节位于左侧叶,最大径为33 mm,结节距被膜 <2 mm,被膜光滑、完整;AI-UADS提示ACR-TIRADS 2类,恶性概率0.06



图5 甲状腺乳头状癌患者(女,52岁)二维超声图,结节位于右侧叶,最大径为14 mm,结节距被膜 <2 mm,被膜中断;AI-UADS提示ACR-TIRADS 5类,恶性概率0.90

2.AI-UADS 与超声医师鉴别诊断不同位置甲状腺结节良恶性的诊断效能比较:630 个甲状腺结节中,位于右侧叶 274 个(恶性结节 126 个,良性结节 148 个),峡部 66 个(恶性结节 58 个,良性结节 8 个),左侧叶 290 个(恶性结节 129 个,良性结节 161 个)。AI-UADS 鉴别诊断左侧叶甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值均高于超声医师,差异均有统计学意义(均 $P<0.05$);AI-UADS 与超声医师鉴别诊断右侧叶、峡部甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。见表 4 和图 2~5。

3.AI-UADS 与超声医师鉴别不同被膜关系甲状腺结节良恶性的诊断效能比较:630 个甲状腺结节中,远离被膜 102 个(恶性结节 46 个,良性结节 56 个),紧邻被膜 388 个(恶性结节 137 个,良性结节 251 个),侵犯被膜 140 个(恶性结节 130 个,良性结节 10 个)。AI-UADS 鉴别诊断远离被膜的甲状腺结节良恶性的准确

率、特异度、阳性预测值均高于超声医师,差异均有统计学意义(均 $P<0.05$);AI-UADS 鉴别诊断侵犯被膜的甲状腺结节良恶性的准确率、灵敏度、阴性预测值均高于超声医师,差异均有统计学意义(均 $P<0.05$);AI-UADS 与超声医师鉴别诊断紧邻被膜的甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。见表 5 和图 2~5。

4.AI-UADS 与超声医师鉴别不同背景甲状腺结节良恶性的诊断效能比较:630 个甲状腺结节中,不伴 HT 523 个(恶性 244 个,良性 279 个),伴 HT 107 个(恶性 69 个,良性 38 个)。AI-UADS 鉴别诊断不伴 HT 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值均高于超声医师,差异均有统计学意义(均 $P<0.05$);AI-UADS 与超声医师鉴别诊断伴 HT 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。见表 6 和图 6,7。

表 4 AI-UADS 与超声医师鉴别不同位置甲状腺结节良恶性的诊断效能比较 %

分类	右侧叶					峡部					左侧叶				
	准确率	灵敏度	特异度	阳性预测值	阴性预测值	准确率	灵敏度	特异度	阳性预测值	阴性预测值	准确率	灵敏度	特异度	阳性预测值	阴性预测值
AI-UADS	87.2	92.1	83.1	82.3	92.5	93.9	96.6	25.0	96.6	75.0	91.4	97.7	86.3	85.1	97.9
超声医师	85.0	89.7	81.1	80.1	90.2	87.9	91.4	37.5	94.6	50.0	81.7	88.4	76.4	75.0	89.1
χ^2 值	0.550	0.431	0.207	0.209	0.428	1.467	-	-	-	-	11.613	8.600	5.244	4.814	8.922
P 值	0.458	0.512	0.649	0.647	0.513	0.226	0.438	1.000	0.676	0.367	0.001	0.003	0.022	0.028	0.003

-示采用 Fisher 精确检验

表 5 AI-UADS 与超声医师鉴别不同被膜关系甲状腺结节良恶性的诊断效能比较 %

分类	远离被膜					紧邻被膜					侵犯被膜				
	准确率	灵敏度	特异度	阳性预测值	阴性预测值	准确率	灵敏度	特异度	阳性预测值	阴性预测值	准确率	灵敏度	特异度	阳性预测值	阴性预测值
AI-UADS	95.1	95.7	92.9	91.8	98.1	87.1	92.7	84.1	76.0	95.5	93.6	96.9	50.0	96.2	55.6
超声医师	83.3	97.8	73.2	74.6	95.3	85.3	93.4	80.9	72.7	95.8	80.0	83.1	40.0	94.7	15.4
χ^2 值	7.337	-	7.669	5.501	-	0.531	0.057	0.882	0.495	0.020	11.242	13.846	-	0.297	-
P 值	0.007	1.000	0.006	0.019	0.585	0.466	0.812	0.348	0.482	0.887	0.001	<0.001	1.000	0.586	0.030

-示采用 Fisher 精确检验

表 6 AI-UADS 与超声医师鉴别不同背景甲状腺结节良恶性的诊断效能比较 %

分类	不伴 HT					伴 HT				
	准确率	灵敏度	特异度	阳性预测值	阴性预测值	准确率	灵敏度	特异度	阳性预测值	阴性预测值
AI-UADS	90.2	95.1	86.0	85.6	95.2	85.1	95.7	73.7	86.8	90.3
超声医师	83.4	88.5	78.9	78.5	88.7	86.0	92.8	73.7	86.5	84.8
χ^2 值	10.819	6.971	4.951	4.562	7.238	0.033	-	0.000	0.004	-
P 值	0.001	0.008	0.026	0.032	0.007	0.856	0.718	1.000	0.949	0.709

-示采用 Fisher 精确检验。HT:桥本甲状腺炎

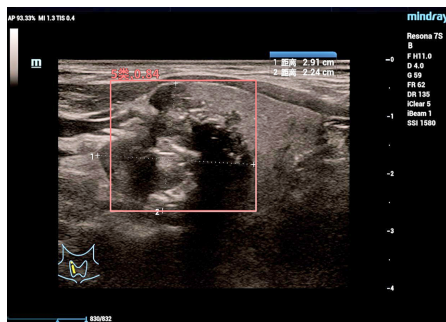


图6 甲状腺乳头状癌不伴HT患者(女,57岁)二维超声示结节位于右侧叶,最大径为29 mm,声示结节位于左侧叶,最大径为7 mm,结节距被膜<2 mm,被膜中断;AI-UADS提示 ACR-TIRADS 5类,恶性概率0.84



图7 甲状腺乳头状癌伴HT(女,39岁)二维超声示结节位于左侧叶,最大径为7 mm,结节距被膜<2 mm,被膜完整;AI-UADS提示 ACR-TIRADS 5类,恶性概率0.83

讨 论

近年来,随着高频超声技术的广泛应用,甲状腺结节的检出率显著提高。超声凭借其无创性、操作简便和经济高效等特点,已成为鉴别诊断甲状腺结节良恶性的首选影像学检查方法。恶性结节典型的声像图特征包括:边界模糊/不规则、纵横比>1、实性结构、低回声、微钙化等。然而,临床实践发现,单一特征的诊断效能具有明显局限性,表现为灵敏度与特异性差异显著(如微钙化特异性高达90%但灵敏度仅30%^[8]),且超声医师的主观判读易导致诊断偏倚。此外,结节生物学行为的异质性(包括生长位置差异、体积变化、HT背景甲状腺组织特性等)常导致良恶性结节的声像图特征出现重叠,显著增加了鉴别诊断的难度。随着人工智能在医学影像领域的快速发展,基于深度学习的计算机视觉算法能够系统提取超声图像中的多维特征信息,实现了从结节自动分割、特征量化到良恶性预测的全流程智能化分析。本研究旨在探讨AI-UADS在不同临床条件下(包括结节大小分层、解剖位置变异、被膜关系及背景甲状腺特征)对甲状腺结节良恶性的鉴别诊断效能。

超声作为甲状腺结节评估的首选影像学方法,其不仅具有较高的结节检出率,更能通过声像图特征的量化分析实现恶性风险分层的客观评估。然而,目前广泛应用的多种TIRADS分类(如ACR TI-RADS、K-TIRADS及ATA指南等)在诊断效能上各具特点。本研究采用ACR-TIRADS分类作为鉴别甲状腺结节良恶性的5项关键声像图特征(成分、回声、形态、边缘、钙化)诊断标准,结果显示AI-UADS鉴别诊断甲状腺结节良恶性的AUC、准确率、灵敏度、特异度、阳性预测值、阴性预测值分别为0.939、89.8%、95.2%、84.5%、85.9%、94.7%,均高于超声医师(0.875、83.8%、89.5%、

78.2%、80.2%、88.3%),差异均有统计学意义(均 $P<0.05$),证实了该标准的临床应用价值。AI-UADS的优势在于实现了图像特征从主观判断到客观量化的转变,通过精准提取关键图像信息显著提高了诊断的一致性和准确性。本研究中,AI-UADS成功纠正了超声医师的33个误诊(滤泡性结节性疾病28个、滤泡性腺瘤3个、炎性病灶2个)和26个漏诊(乳头状癌25个、髓样癌1个),证实AI-UADS在甲状腺结节

良恶性鉴别中具有明确的临床应用价值,其卓越的诊断效能为优化甲状腺结节管理策略提供了新的技术支撑。而这些漏误诊的病例多具有特定的解剖学特征(如左侧叶分布、被膜毗邻关系、特定大小范围、不同背景等),故本研究进一步对其进行亚组分析。

研究^[9]表明,甲状腺良恶性结节的声像图特征分布与其大小密切相关。较小结节(尤其是最大径 ≤ 10 mm)更易出现典型的恶性声像图特征(如低回声、微钙化、不规则边缘等;而随着结节增大(最大径 >20 mm),恶性征象的出现概率逐渐降低。这一现象可能与结节生物学行为及病理演变过程相关,即微小乳头状癌往往表现出典型的恶性特征,而较大结节可能因退行性改变(如囊性变、纤维化)而掩盖其恶性征象。本研究中超声医师误诊结节主要集中于较小结节(≤ 10 mm者占73.9%);漏诊结节则多见于较大结节(最大径 >20 mm者占63.6%)。相比之下,AI-UADS的误诊及漏诊分布更为均衡:误诊结节中最大径 ≤ 10 mm者占53.1%;漏诊结节中 >20 mm者占33.3%。表明AI-UADS在较小结节的误诊控制和较大结节的漏诊减少方面均优于超声医师。不同大小结节的诊断效能分析发现:①对于最大径 ≤ 10 mm结节,AI-UADS的准确率、特异度及阳性预测值均显著高于超声医师(均 $P<0.05$),提示其在鉴别诊断较小结节良恶性时更具可靠性;②对于最大径10~20 mm结节,超声医师与AI-UADS的诊断效能相当,说明此类结节声像图特征较为典型,超声医师易于识别;③对于最大径 >20 mm结节,AI-UADS的灵敏度及阴性预测值均显著高于超声医师(均 $P<0.05$),表明其能更有效识别大结节的恶性风险。

研究^[10]表明,结节位置是预测甲状腺恶性风险的独立危险因素,位于甲状腺峡部结节的恶性风险最高

($OR=2.4, P<0.05$), 而位于侧叶下份结节的恶性风险最低($OR=1.5, P<0.05$)。本研究纳入的 630 个甲状腺结节中, 位于峡部的结节恶性占比最高, 为 87.9%, 这一分布特征与上述研究结果相符, 进一步证实了结节位置与恶性风险的相关性。本研究超声医师漏诊峡部结节 5 个、误诊 3 个, AI-UADS 漏诊峡部结节 2 个、误诊 2 个; 超声医师漏诊左侧叶结节 15 个、误诊 38 个, AI-UADS 漏诊左侧叶结节 3 个、误诊 22 个; 超声医师漏诊右侧叶结节 13 个、误诊 28 个, AI-UADS 漏诊右侧叶结节 10 个、误诊 25 个。提示超声医师更易误诊左侧叶结节, AI-UADS 更易漏诊右侧叶结节。甲状腺峡部位于气管前方, 连接甲状腺左、右侧叶, 组织厚度较薄。位于峡部的甲状腺癌声像图特征表现不典型, 且易侵犯被膜及颈部淋巴结转移; 且因峡部体积较小, 此处结节更易与前后缘被膜接触, 导致边界难以辨别。故本研究超声医师与 AI-UADS 在鉴别诊断峡部甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。本研究结果显示, AI-UADS 鉴别诊断左侧叶甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值均高于超声医师, 差异均有统计学意义(均 $P<0.05$); AI-UADS 与超声医师鉴别诊断右侧叶甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。分析原因: ①解剖结构差异。左侧叶通常较右侧叶体积更大, 位置更表浅, 且左侧喉返神经走行更靠近甲状腺, 可能影响结节生长模式; 右侧叶与颈动脉鞘关系更密切, 可能干扰超声成像。②血流供应特点。甲状腺左、右侧叶动脉供血存在不对称性, 左侧叶血流灌注可能更丰富, 影响结节声像图特征表现; 且血流差异可能导致 AI 算法对不同侧叶结节的特征提取效率不同。③扫查技术因素。扫查右侧叶时探头操作受锁骨和胸骨限制更多; 扫查左侧叶时超声医师操作空间更大, 图像获取更标准; AI-UADS 对标准化图像的分析效能更高。④样本特征分布, 本研究中左侧叶较小结节(最大径 $\leq 10\text{ mm}$, 31 个)及远离被膜的结节(8 个)占比更高(14.2%), AI-UADS 诊断这类结节的特异度更高; 右侧叶侵犯被膜结节较少(10 个), 影响 AI-UADS 诊断灵敏度。

甲状腺癌腺外侵犯是 ACR-TIRADS 分类中的重要评估要素, 其诊断具有重要临床意义。研究^[11-12]表明, 被膜侵犯或突破与颈部淋巴结转移风险显著增高相关。因此, 准确识别腺外侵犯对甲状腺癌的临床管理和预后评估至关重要。本研究结果显示, 超声医师

在诊断腺外侵犯方面存在明显局限性, 共漏诊 22 个侵犯被膜的结节, 且结节多较大。分析原因: ①甲状腺疾病的多源性导致声像图表现复杂, 尤其在甲状腺癌与滤泡性结节性疾病共存时, 易将恶性结节误判为良性^[13]; ②超声医师对较大甲状腺结节的诊断灵敏度普遍较低^[8]。相比之下, AI-UADS 在腺外侵犯诊断方面表现出显著优势, 仅漏诊 4 个。本研究结果显示, AI-UADS 鉴别诊断侵犯被膜的甲状腺结节良恶性的准确率、灵敏度、阴性预测值均高于超声医师, 差异均有统计学意义(均 $P<0.05$); 超声医师误诊了 15 个远离被膜的结节, 而 AI-UADS 仅误诊 4 个, AI-UADS 诊断远离被膜的甲状腺结节良恶性的准确率、特异度、阳性预测值均高于超声医师, 差异均有统计学意义(均 $P<0.05$)。AI-UADS 与超声医师鉴别诊断紧邻被膜的甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义, 与既往研究^[14]结论一致。表明 AI-UADS 能更好地筛选出侵犯被膜及远离被膜的甲状腺恶性结节, 减少漏误诊; 而对紧邻被膜的结节构成诊断挑战。

HT 是一种自身免疫性疾病, 早期表现为弥漫性或局灶性淋巴细胞浸润; 中后期甲状腺滤泡破坏萎缩, 淋巴滤泡大量增生, 形成生发中心, 纤维组织增生形成玻璃样变; 疾病晚期, 由于大量的纤维组织增生, 腺体变硬, 并呈结节样改变, 或萎缩变小。声像图主要表现为弥漫性不均匀性减低, 伴有广泛的细线状强回声交织成网格状^[15]。既往研究^[16-17]发现, AI-UADS 能准确识别结节并判断其性质, 不受腺体实质基础病理改变的影响, 且 AI-UADS 鉴别诊断伴 HT 甲状腺结节良恶性的特异度高于超声医师($P<0.05$)。但本研究结果与之不一致, AI-UADS 与超声医师鉴别诊断伴 HT 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值比较差异均无统计学意义。分析原因: ①随着超声医师对伴 HT 甲状腺结节认识的不断深入, 其诊断特异度已提高, 缩小了与 AI-UADS 的差距^[18-19]; ②虽然伴 HT 结节更易发生甲状腺乳头状癌, 但由于慢性炎症对甲状腺实质的持续性损伤, 甲状腺实质背景粗糙不均匀, 且不同时期的病理改变可使背景回声不同, 增加了诊断难度; ③AI-UADS 对伴 HT 甲状腺结节的超声图像分割欠佳, 影响其诊断效能^[20]。本研究结果显示 AI-UADS 鉴别诊断不伴 HT 甲状腺结节良恶性的准确率、灵敏度、特异度、阳性预测值、阴性预测值均高于超声医师, 差异均有统计学意义(均 $P<0.05$)。表明 AI-UADS 在鉴别不伴 HT 甲状腺结节

良恶性中的诊断效能更高。

综上所述, AI-UADS在鉴别甲状腺结节良恶性中具有较高的诊断效能, 尤其对于位于左侧叶、最大径 ≤ 10 mm、远离被膜和侵犯被膜, 以及不伴HT甲状腺结节良恶性的诊断效能更高。但本研究为单中心研究, 且样本量小, 恶性结节病理类型较单一, 后期将扩大样本量进一步完善。

参考文献

- [1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer[J]. Thyroid, 2016, 26(1): 1-133.
- [3] Brito JP, Gionfriddo MR, Al Nofal A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and Meta-analysis[J]. J Clin Endocrinol Metab, 2014, 99(4): 1253-1263.
- [4] Persichetti A, Di Stasio E, Coccaro C, et al. Inter- and intraobserver agreement in the assessment of thyroid nodule ultrasound features and classification systems: a blinded multicenter study[J]. Thyroid, 2020, 30(2): 237-242.
- [5] Ha EJ, Lee JH, Lee DH, et al. Artificial intelligence model assisting thyroid nodule diagnosis and management: a multicenter diagnostic study[J]. J Clin Endocrinol Metab, 2024, 109(2): 527-535.
- [6] 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]. J Am Coll Radiol, 2017, 14(5): 587-595.
- [7] 巴晨曦, 徐帅娅, 张鑫茹, 等. 未侵犯被膜的甲状腺乳头状癌发生颈部淋巴结转移的危险因素分析[J]. 影像研究与医学应用, 2022, 6(23): 66-68.
- [8] 丁思悦, 丁全全, 王雁, 等. C-TIRADS与ACR TI-RADS在甲状腺结节中的诊断效能对比研究[J]. 中国超声医学杂志, 2021, 37(9): 964-967.
- [9] Qu C, Li HJ, Gao Q, et al. Alteration trend and overlap analysis of positive features in different-sized benign and malignant thyroid nodules: based on Chinese Thyroid Imaging Reporting and Data System[J]. Int J Gen Med, 2024, 17: 1887-1895.
- [10] Jasim S, Baranski TJ, Teefey SA, et al. Investigating the effect of thyroid nodule location on the risk of thyroid cancer[J]. Thyroid, 2020, 30(3): 401-407.
- [11] 胡小玲, 冉海涛. 超声影像组学评估甲状腺乳头状癌颈部淋巴结转移[J]. 中国超声医学杂志, 2022, 38(4): 367-370.
- [12] 陶玲玲, 樊金芳, 李伟伟, 等. C-TIRADS结合不同接触面积对甲状腺乳头状癌被膜侵犯的术前预测价值[J]. 中国超声医学杂志, 2024, 40(7): 733-737.
- [13] 阮健秋, 徐辉雄, 谢晓燕, 等. 甲状腺结节超声误诊原因的分析——大宗病例临床报告[J]. 影像诊断与介入放射学, 2010, 19(2): 115-117.
- [14] Fan J, Zhou W, Zhan W, et al. Clinical and ultrasonographic features of papillary thyroid carcinoma located in the isthmus[J]. Ultrasound Q, 2023, 39(1): 32-36.
- [15] 郑蓉, 张丹. 桥本甲状腺炎的诊断进展[J]. 中华临床医师杂志(电子版), 2013, 7(4): 1687-1689.
- [16] 宗方, 蒋天安, 金巧芳, 等. 桥本氏甲状腺炎合并甲状腺乳头状癌超声分析[J]. 医学影像学杂志, 2011, 21(12): 1792-1794, 1797.
- [17] Wang B, Wan Z, Zhang M, et al. Diagnostic value of a dynamic artificial intelligence ultrasonic intelligent auxiliary diagnosis system for benign and malignant thyroid nodules in patients with Hashimoto thyroiditis[J]. Quant Imaging Med Surg, 2023, 13(6): 3618-3629.
- [18] 赵永峰, 杜杰, 胡峰, 等. 3种甲状腺影像报告和数据系统对桥本甲状腺炎背景下甲状腺结节的诊断价值[J]. 中国超声医学杂志, 2023, 39(4): 365-368.
- [19] 孙鑫, 张巍, 程令刚, 等. 以7种不同甲状腺影像报告和数据系统鉴别桥本甲状腺炎背景下甲状腺良、恶性结节[J]. 中国医学影像技术, 2024, 40(2): 212-216.
- [20] Ma J, Wu F, Zhu J, et al. A pre-trained convolutional neural network based method for thyroid nodule diagnosis[J]. Ultrasonics, 2017, 73: 221-230.

(收稿日期: 2024-11-18)