

DOI: 10.16781/j.CN31-2187/R.20230027

· 论 著 ·

基于 AI-SONIC™ Thyroid 5.3.3.0 的超声图像分析对甲状腺结节恶性风险的预测价值

郭芳琪^{1,2,3}, 刘 晟^{4,5}, 徐 磊⁶, 李勇刚^{1*}, 赵佳琦^{3*}

1. 苏州大学附属第一人民医院放射科, 苏州 215000
2. 海军军医大学(第二军医大学)第二附属医院超声诊疗科, 上海 200003
3. 同济大学附属上海市第四人民医院超声医学科, 上海 200434
4. 海军军医大学(第二军医大学)第二附属医院甲乳疝外科, 上海 200003
5. 同济大学附属上海市第四人民医院甲乳血管外科, 上海 200434
6. 浙江求是数理医学研究院, 杭州 315032

[摘要] **目的** 探讨基于超声人工智能(AI)系统 AI-SONIC™ Thyroid 5.3.3.0 的图像分析在甲状腺结节恶性风险评估中的应用价值。**方法** 选取2019年4月至2021年1月海军军医大学(第二军医大学)第二附属医院收治的453例甲状腺结节患者,共573枚甲状腺结节。以术后病理结果为金标准,通过 χ^2 检验和ROC曲线评估术前AI系统检查对不同性别分组、不同年龄分组及不同结节大小分组的甲状腺结节良恶性的鉴别诊断效能,并通过DeLong检验比较术前AI系统检查与不同年资超声医师术前应用常规超声检查鉴别诊断甲状腺结节良恶性的效能。**结果** 在术前检查的573枚甲状腺结节中,术后病理证实为恶性411枚(76.5%)、良性162枚(23.5%)。低年资超声医师应用常规超声检查鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度分别为85.2%(350/411)、55.6%(90/162)、76.8%(440/573),AUC为0.721(95%CI 0.672~0.771);高年资超声医师鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度分别为93.9%(386/411)、74.1%(120/162)、88.3%(506/573),AUC为0.865(95%CI 0.825~0.904);AI系统鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度分别为92.5%(380/411)、69.1%(112/162)、85.9%(492/573),AUC为0.809(95%CI 0.764~0.854)。DeLong检验结果显示,AI系统鉴别诊断甲状腺结节良恶性的AUC高于低年资超声医师($P=0.032$),与高年资超声医师之间差异无统计学意义($P>0.05$)。按不同性别、不同年龄分组,AI系统鉴别诊断甲状腺结节良恶性的准确度差异无统计学意义($P>0.05$);按不同结节大小分组,结节最大直径为10~<15 mm时AI系统鉴别诊断甲状腺结节良恶性的AUC最大,为0.882(95%CI 0.723~0.916)。**结论** AI-SONIC™ Thyroid 5.3.3.0可识别甲状腺结节的良性和恶性声像特征,其诊断效能接近高年资超声医师,有望成为术前预测甲状腺结节恶性风险的实用工具。

[关键词] 甲状腺结节; 超声检查; 人工智能; 计算机辅助诊断

[引用本文] 郭芳琪, 刘晟, 徐磊, 等. 基于AI-SONIC™ Thyroid 5.3.3.0的超声图像分析对甲状腺结节恶性风险的预测价值[J]. 海军军医大学学报, 2024, 45(1): 29-36. DOI: 10.16781/j.CN31-2187/R.20230027.

Ultrasound image analysis based on AI-SONIC™ Thyroid 5.3.3.0 to predict malignant risk of thyroid nodules

GUO Fangqi^{1,2,3}, LIU Sheng^{4,5}, XU Lei⁶, LI Yonggang^{1*}, ZHAO Jiaqi^{3*}

1. Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou 215000, Jiangsu, China
2. Department of Ultrasound, The Second Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai 200003, China
3. Department of Ultrasound, Shanghai Fourth People's Hospital Affiliated to Tongji University, Shanghai 200434, China

[收稿日期] 2023-02-03

[接受日期] 2023-09-26

[基金项目] 海军军医大学(第二军医大学)第二附属医院人才建设三年行动计划——“金字塔人才工程”军事医学人才项目(1009),同济大学附属上海市第四人民医院科研启动专项(SYKYQD06101),上海市虹口区卫生健康委员会医学科研课题(虹卫2302-26),上海市虹口区卫生健康委员会临床重点扶持专科项目(HKLCFC202404)。Supported by Military Medical Talent Project of “Pyramid Talent Program” of Three-year Action Plan for Talent Construction of The Second Affiliated Hospital of Naval Medical University (Second Military Medical University) (1009), Start-up Scientific Research Project of Shanghai Fourth People's Hospital Affiliated to Tongji University (SYKYQD06101), Medical Research Project of Health Commission of Shanghai Hongkou District (HW2302-26), and Clinical Key Supporting Project of Health Commission of Shanghai Hongkou District (HKLCFC202404).

[作者简介] 郭芳琪, 硕士生, 主治医师. E-mail: 1002801624@qq.com

*通信作者(Corresponding authors). Tel: 0512-67780155, E-mail: liyonggang224@163.com; Tel: 021-55603999-3550, E-mail: qiqiblue67@163.com

4. Department of Thyroid, Breast and Hernia Surgery, The Second Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai 200003, China
5. Department of Thyroid, Breast and Vascular Surgery, Shanghai Fourth People's Hospital Affiliated to Tongji University, Shanghai 200434, China
6. Zhejiang Qiushi Institute for Mathematical Medicine, Hangzhou 315032, Zhejiang, China

[Abstract] Objective To investigate the application value of artificial intelligence (AI) system AI-SONIC™ Thyroid 5.3.3.0 based on ultrasound image analysis in the risk assessment of malignant thyroid nodules. **Methods** A total of 573 thyroid nodules were selected from 453 patients who were admitted to The Second Affiliated Hospital of Naval Medical University (Second Military Medical University) from Apr. 2019 to Jan. 2021. With the postoperative pathology as the gold standard, the differential diagnostic efficacies of preoperative AI system examination for benign and malignant thyroid nodules in different gender groups, different age groups, and different nodule size groups were evaluated by χ^2 test and receiver operating characteristic (ROC) curve. The efficacies of preoperative AI system examination and conventional ultrasonography by different seniority ultrasound physicians before operation in diagnosing benign and malignant thyroid nodules were compared by DeLong test. **Results** Of the 573 thyroid nodules examined before operation, 411 (76.5%) were malignant and 162 (23.5%) were benign as confirmed by pathology after operation. The sensitivity, specificity, and accuracy of conventional ultrasonography in diagnosing benign and malignant thyroid nodules were 85.2% (350/411), 55.6% (90/162), and 76.8% (440/573), respectively, with an area under curve (AUC) of 0.721 (95% confidence interval [CI] 0.672-0.771) in the junior ultrasound physicians; the sensitivity, specificity, and accuracy were 93.9% (386/411), 74.1% (120/162), and 88.3% (506/573), respectively, with an AUC of 0.865 (95% CI 0.825-0.904) in the senior ultrasound physicians. The sensitivity, specificity, and accuracy of the AI system were 92.5% (380/411), 69.1% (112/162), and 85.9% (492/573), respectively, with an AUC of 0.809 (95% CI 0.764-0.854). DeLong test results showed that the AUC of the AI system in diagnosing benign and malignant thyroid nodules was significantly higher than that of the junior ultrasound physicians ($P=0.032$), and there was no significant difference between the AI system and senior ultrasound physicians ($P>0.05$). There was no significant difference in the accuracy of the AI system in diagnosing benign or malignant thyroid nodules among patients with different genders or different ages ($P>0.05$). For nodules of different sizes, when the maximum diameter of nodules was 10- $<$ 15 mm, the AUC of the AI system was the highest, being 0.882 (95% CI 0.723-0.916). **Conclusion** AI-SONIC™ Thyroid 5.3.3.0 can identify benign and malignant features of thyroid nodules, and its diagnostic efficiency is close to that of senior ultrasound physicians. It is expected to be a practical tool to predict the risk of malignant thyroid nodules before clinical operation.

[Key words] thyroid nodule; ultrasonography; artificial intelligence; computer aided diagnosis

[Citation] GUO F, LIU S, XU L, et al. Ultrasound image analysis based on AI-SONIC™ Thyroid 5.3.3.0 to predict malignant risk of thyroid nodules[J]. Acad J Naval Med Univ, 2024, 45(1): 29-36. DOI: 10.16781/j.CN31-2187/R.20230027.

据不完全统计,全球甲状腺结节发病率逐年增高,2020年我国甲状腺癌发病人数高达22.1万,在女性中其发病率居新发癌症第4位^[1]。高频超声检查具有实时、无创、可重复等优势,被广泛应用于甲状腺结节的筛查和评估^[2],然而由于良性与恶性甲状腺结节的超声图像特征存在重叠^[3],超声医师对两者的鉴别诊断水平受到个人主观因素、操作医师扫查技巧及甲状腺基础病变等因素干扰^[4-5],不利于甲状腺结节良恶性的精准鉴别。此外,现阶段国内使用的甲状腺影像报告和数据系统(thyroid imaging reporting and data system, TIRADS)分类版本较多,尚不能达成一致,这给临床医师解读甲状腺超声报告带来诸多困扰。2017年,

中华医学会超声医学分会浅表器官和血管学组提出建立中国(超声)甲状腺影像报告和数据系统(Chinese-TIRADS, C-TIRADS)的设想,随后制定了C-TIRADS分类系统,2020年专家委员会制定了《2020甲状腺结节超声恶性危险分层的中国指南:C-TIRADS》(以下简称C-TIRADS指南)^[6],这是中国超声医学界的首个甲状腺超声应用指南。即便有了C-TIRADS指南,甲状腺结节良恶性的鉴别诊断依然会受到不同地域医疗设备条件参差不齐、不同年资医师诊断水平悬殊等因素的影响,造成误诊或漏诊。

基于人工智能(artificial intelligence, AI)的医学影像技术近年来飞速发展,医学影像AI检测

系统目前已被应用于各种临床场景,不仅在辅助医师进行病灶识别、诊断、疗效评估等方面展示出高准确率,而且提高了医师的工作效率^[7],在克服人为因素和主观判读的局限性方面也具有较大的潜力^[8]。超声检查不同于常规X线、CT、MRI等影像检查方式,它对操作者的依赖性极强,而AI在超声图像分析和数据存储过程中能为超声医师提供精准、客观、实时的智能导向^[9],因此研发能够评估超声图像并向超声医师提供即时反馈的AI诊断模型具有重要意义。

以深度学习为代表的AI辅助诊断技术可用于甲状腺结节诊断,深度学习算法主要依赖于排列在层中计算单元的多层网络结构,它可以从输入数据中自动逐像素获取更高级别的影像特征。随着深度学习算法不断迭代更新,以数百万张图像进行训练,基于甲状腺结节数据进行深度学习,可实现甲状腺结节恶性风险的准确预测。研究表明,超声AI系统的诊断灵敏度、特异度、准确度等与高年资超声医师相当^[10-13]。超声AI系统不仅可对典型甲状腺结节的声像特征进行自动识别,还能对微小结节、不典型结节等可疑甲状腺结节的恶性风险进行细化区分。本研究通过回顾573枚甲状腺结节的术前检查资料,基于AI系统AI-SONIC™ Thyroid 5.3.3.0,评估术前AI系统对不同性别分组、不同年龄分组及不同结节大小分组的甲状腺结节良恶性的鉴别诊断效能,并与不同年资超声医师基于术前常规超声检查的诊断效能进行比较,探讨术前AI系统辅助超声检查预测甲状腺结节恶性风险的价值。

1 资料和方法

1.1 一般资料 回顾性纳入2019年4月至2021年1月于海军军医大学(第二军医大学)第二附

属医院甲乳瘤外科手术的453例甲状腺结节患者(共573枚甲状腺结节),其中男106例(23.4%)、女347例(76.6%),年龄为18~71(47.0±12.3)岁。结节直径为(14.06±14.00)cm,其中良性结节162枚(23.5%),直径为(25.52±19.40)cm;恶性结节411枚(76.5%),直径为(9.58±7.50)cm。

1.2 病例纳入与排除标准 本研究严格遵循生物医学信息学诊断试验研究中的纳入标准和排除标准^[14]。纳入标准:(1)术前接受常规超声检查的C-TIRADS分类、AI系统的风险评估;(2)经术后病理学检查结果证实;(3)观察结节具有标准完整的横、纵切面图与风险概率图。排除标准:(1)病理诊断结果不明确;(2)目标结节超出切面的观察范围;(3)患者曾接受过手术和消融等有创操作与治疗;(4)目标结节的超声图像受到测量线、彩色血流信号、弹性成像等标注的干扰。

1.3 常规超声检查与C-TIRADS分类标准 采用线阵高频(频率为7.5~12 MHz)L74M探头和Hitachi Preirus超声诊断仪(日本Hitachi公司)进行甲状腺常规超声检查。系统参数设定:增益(gain)为30 dB,时间增益补偿(time gain compensation, TGC)为零补偿,聚焦深度(depth)为2.00~3.75 cm,组织热指数(thermal index, TI)<0.4,机械指数(mechanical index, MI)为1.2。

C-TIRADS分类标准:恶性结节超声图像特征表现包括实性、极低回声、微钙化、边界不清、形态不规则及纵横比>1,而彗星尾征为良性结节的超声特征^[15]。统计恶性特征的数量,每项加1分,如表现为良性特征的彗星尾征则减去1分,根据统计的总分对甲状腺结节进行分类^[16],见表1。将C-TIRADS 2~3类结节列为偏良性,4A及以上结节列为偏恶性。

表1 基于计数法的C-TIRADS分类

Tab 1 C-TIRADS classification based on counting method

Nodule	Score	Malignancy rate/%	C-TIRADS classification
Nodule free	No score	0	Category 1: nodule free
Nodules	-1	0	Category 2: benign
	0	<2	Category 3: benign possibility
	1	2-10	Category 4A: lowly suspected malignancy
	2	10-50	Category 4B: moderately suspected malignancy
	3 or 4	50-90	Category 4C: highly suspected malignancy
	5	>90	Category 5: highly suggestive malignancy
	No score	No score	Category 6: malignancy confirmed by biopsy

C-TIRADS: Chinese-thyroid imaging reporting and data system.

1.4 AI系统检查方法及评判标准 采用AI-SONIC™ Thyroid 5.3.3.0 甲状腺结节AI辅助诊断系统 (DEMETICS®，浙江德尚韵兴医疗科技有限公司) 辅助超声检查过程进行自动检测。该系统经过20万余枚甲状腺结节穿刺及病理结果对照的深度深度学习，创新性地引入Res-Net残差神经网络架构^[17]，以解决深度学习随着网络层数增加损失函数梯度消失所导致的难以通过增加网络深度提升算法模型准确度的问题。残差神经网络通过恒等映射将浅层的参数层以跳跃链接的方式向深层映射传递，避免了损失函数梯度经多层反向传递后消失的情况，并持续优化深度学习理论和变分能量泛函结合的图像处理与分析方法。

由2名甲状腺超声医师(1名工作10年以上的高年资超声医师和1名工作5年以下的低年资超声医师)术前对患者甲状腺两侧叶及峡部进行标准化扫查。所有结节分别经过4次纵、横标准切面显示，选取结节长径显示最清楚的纵切面及结节前后径与左右径显示最清楚的横切面图进行冻结，通过DICOM格式传输至AI系统存储，并记录结节相应的风险概率值。AI系统诊断标准设定见图1，风险概率值为0~1.00，其中0~0.40为无恶性特征，用绿色包络线显示；0.41~0.60为具有恶性倾向，用橙色包络线显示；0.61~1.00为具有恶性特征，用红色包络线显示。将评分为0~0.40的结节归为良性，评分为0.41~1.00的结节归为偏恶性。

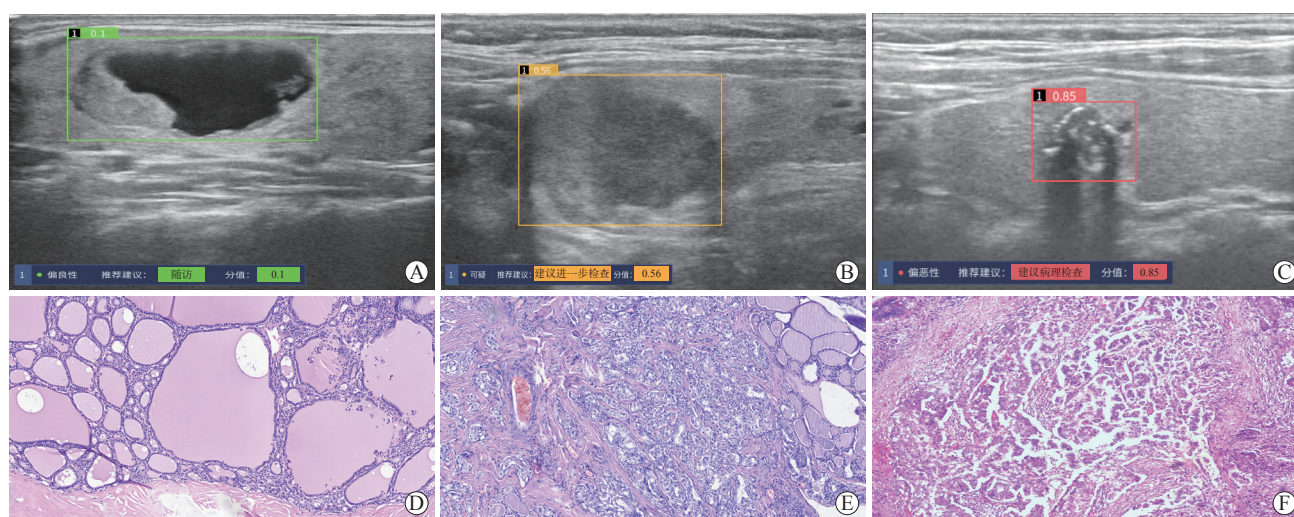


图1 AI系统对甲状腺结节风险评估声像图(A~C)及相应的病理结果(D~F)

Fig 1 Sonogram of risk assessment of thyroid nodules by AI system (A-C) and pathological classification (D-F)

A, D: The risk score of the AI system was 0.1, indicating a benign nodules, and the recommendation was to follow up (A). Pathological findings showed nodular goiter with cystic changes (D). B, E: The risk score of the AI system was 0.56, indicating a suspicious nodule, and the recommendation was to take further examination (B). The pathological findings showed papillary thyroid carcinoma (E). C, F: The risk score of the AI system was 0.85, indicating a more malignant nodule, and the recommendation was to take pathological examination (C). Pathological findings showed papillary thyroid carcinoma (F). D-F: Hematoxylin-eosin staining (10×). AI: Artificial intelligence.

将573枚甲状腺结节按患者性别分为2组，按患者年龄分为3组(18~44岁组、45~60岁组、>60岁组^[18])，按结节大小分为5组(<5 mm组、5~<10 mm组、10~<15 mm组、15~<20 mm组、≥20 mm组)，比较AI系统对不同分组的甲状腺结节良恶性的诊断价值。甲状腺结节的大小按照超声图像及报告质量管理体系^[19]进行标准化测量，在甲状腺结节长轴切面测量结节的最大直径，取3次测量结果的平均值。

1.5 诊断用时的计算 超声医师诊断用时：选取研究期间半年内对117个结节的诊断情况，记录总诊断用时，再计算每个结节的回顾性诊断用时(s)。纳入该诊断时间评估的所有图像都经过了纳入和排除标准中描述的共同预处理。AI系统诊断用时：采用同一AI系统供应商的注释软件对超声图像与相应的甲状腺结节进行分组，将平均图像处理时间乘以每个结节的平均图像数，计算出每个结节的诊断时间(s)。

1.6 统计学处理 采用 SPSS 26.0 软件进行统计学分析。计量资料以 $\bar{x} \pm s$ 表示, 组间比较采用独立样本 t 检验。计数资料以频数和百分数表示, 以术后病理结果 (2022 年 WHO 甲状腺肿瘤分类标准^[20]) 作为金标准, 计算超声医师和 AI 系统鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度。采用 χ^2 检验比较 AI 系统对不同性别、不同年龄患者甲状腺结节良恶性鉴别诊断的准确度。构建 ROC 曲线评估诊断效能, 采用 DeLong 检验^[21] 比较不同年资超声医师与 AI 系统鉴别诊断甲状腺结节良恶性的效能。检验水准 (α) 为 0.05。

2 结果

2.1 AI 系统与不同年资超声医师对甲状腺结节恶性风险的诊断效能评估 如表 2 所示, 在 573 枚甲状腺结节中, 低年资超声医师检出 151 枚偏良性、422 枚偏恶性, 高年资超声医师检出 145 枚偏良性、428 枚偏恶性, AI 系统检出 143 枚偏良性、430 枚偏恶性, 术后病理证实为恶性 411 枚 (76.5%)、良性 162 枚 (23.5%)。以术后病理结果为金标准, AI 系统诊断正确的结节共 492 枚, 其中 112 枚偏良性, 380 枚偏恶性; AI 系统误诊 81 枚结节, 其中 50 枚良性结节误诊为偏恶性, 31 枚恶性结节误诊为偏良性。在 AI 系统误诊的 81 枚结节中, 55 枚

结节高年资超声医师也同时误诊, 其中 35 枚良性结节误诊为偏恶性 (20 枚甲状腺滤泡性结节病、15 枚滤泡性甲状腺瘤), 20 枚恶性结节误诊为偏良性 (7 枚乳头状甲状腺癌、1 枚甲状腺髓样癌、12 枚滤泡性甲状腺癌); 余 26 枚结节为 AI 系统单独误诊, 其中 15 枚甲状腺滤泡性结节病误诊为偏恶性, 11 枚恶性结节误诊为偏良性 (10 枚甲状腺乳头状癌、1 枚滤泡性甲状腺癌)。

由表 2 可见, 低年资超声医师鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度分别为 85.2% (350/411)、55.6% (90/162)、76.8% (440/573), 高年资超声医师鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度分别为 93.9% (386/411)、74.1% (120/162)、88.3% (506/573), AI 系统鉴别诊断甲状腺结节良恶性的灵敏度、特异度、准确度分别为 92.5% (380/411)、69.1% (112/162)、85.9% (492/573)。ROC 曲线分析结果显示, 低年资超声医师、高年资超声医师和 AI 系统鉴别诊断甲状腺结节良恶性的 AUC 分别为 0.721 (95% CI 0.672~0.771)、0.865 (95% CI 0.825~0.904)、0.809 (95% CI 0.764~0.854), 经 DeLong 检验, AI 系统的 AUC 高于低年资超声医师 ($P=0.032$), AI 系统与高年资超声医师的 AUC 差异无统计学意义 ($P>0.05$)。

表 2 AI 系统与不同年资超声医师对恶性甲状腺结节的诊断效能

Diagnosis method	Pathological diagnosis, <i>n</i>		Sensitivity/%	Specificity/%	Accuracy/%	AUC (95% CI)
	Benign	Malignancy				
Junior ultrasound physician			85.2	55.6	76.8	0.721 (0.672, 0.771)
C-TIRADS 1-3	90	61				
C-TIRADS 4A-6	72	350				
Senior ultrasound physician			93.9	74.1	88.3	0.865 (0.825, 0.904)
C-TIRADS 1-3	120	25				
C-TIRADS 4A-6	42	386				
AI system			92.5	69.1	85.9	0.809 (0.764, 0.854)
Risk probability 0-0.40	112	31				
Risk probability 0.41-1.00	50	380				

AI: Artificial intelligence; C-TIRADS: Chinese-thyroid imaging reporting and data system; AUC: Area under curve; CI: Confidence interval.

2.2 AI 系统对不同人群甲状腺结节良恶性评估的准确度比较 在 453 例患者中, 男性 106 例, 共 127 枚结节, 111 枚与术后病理学检查结果相符, 准确度为 87.4%; 女性 347 例, 共 446 枚结节, 381

枚与术后病理学检查结果相符, 准确度为 85.4%。经 χ^2 检验, AI 系统在不同性别患者中的诊断准确度差异无统计学意义 ($P>0.05$)。

在 453 例患者中, 18~44 岁组 187 例, 共 237 枚

结节, 212枚与术后病理学检查结果相符, 准确度为89.5%; 45~60岁组209例患者, 共267枚结节, 222枚与术后病理学检查结果相符, 准确度为83.1%; >60岁组57例患者, 共69枚结节, 58枚与术后病理学检查结果相符, 准确度为84.1%。经 χ^2 检验, AI系统在不同年龄患者中的诊断准确度差异无统计学意义($P>0.05$)。

2.3 AI系统对不同大小甲状腺结节恶性风险的诊断效能 如表3所示, 结节最大直径为10~<15 mm时AI系统鉴别诊断甲状腺结节良恶性的AUC最大, 为0.882(95% CI 0.723~0.916); 结节最大直径<5 mm时灵敏度最高, 为99.0%; 结节最大直径为15~<20 mm及 ≥ 20 mm时灵敏度较低, 分别为59.1%、59.5%。

表3 AI系统对不同大小恶性甲状腺结节的诊断效能

Tab 3 Diagnostic efficacies of AI system for malignant thyroid nodules of different sizes

Maximum diameter/mm	Pathological diagnosis, <i>n</i>		Sensitivity/%	Specificity/%	Accuracy/%	AUC (95% CI)
	Benign	Malignancy				
<5			99.0	25.0	91.1	0.803 (0.652, 0.835)
Benign	3	1				
Malignancy	9	99				
5-<10			97.3	24.0	88.4	0.760 (0.609, 0.823)
Benign	6	5				
Malignancy	19	177				
10-<15			98.6	56.2	90.7	0.882 (0.723, 0.916)
Benign	9	1				
Malignancy	7	69				
15-<20			59.1	66.7	63.0	0.701 (0.742, 0.920)
Benign	16	9				
Malignancy	8	13				
≥ 20			59.5	91.8	82.0	0.773 (0.681, 0.875)
Benign	78	15				
Malignancy	7	22				

AI: Artificial intelligence; AUC: Area under curve; CI: Confidence interval.

2.4 AI系统与不同年资超声医师对甲状腺结节诊断时间的比较 AI系统诊断良、恶性甲状腺结节用时均为(0.2±0.1) s, 高年资超声医师诊断良、恶性甲状腺结节用时分别为(156.0±9.2)和(180.0±20.0) s, 低年资超声医师诊断良、恶性甲状腺结节用时分别为(216.0±12.7)和(300.0±25.6) s。经独立样本 t 检验, AI系统诊断良、恶性甲状腺结节用时均短于高年资超声医师和低年资超声医师(均 $P<0.05$), 高年资超声医师诊断良性甲状腺结节用时短于诊断恶性甲状腺结节用时($P<0.05$)。

3 讨论

近年来, 越来越多的临床研究致力于将AI应用于医学影像图像分析, 以便更高效地解决临床实际诊断问题。AI不仅可从影像图像中提取千万变化的表现特征并进行量化分析, 还可捕捉到较难识别的细微征象, 对目标区域进行自动标注、探索性

分析、定量诊断、精准评估, 辅助超声医师做出较为精确的诊断。AI辅助超声图像分析技术日渐成熟, 其诊断结果的指向性更接近病理结果, 为甲状腺结节、乳腺结节等高发疾病的初筛和评估提供了较为有益的补充^[22]。

低年资超声医师在临床工作中经验相对不足, 对结节的观察及专注能力不及高年资超声医师, 在短时间内做出准确诊断具有一定的难度。一些少见类型的恶性甲状腺结节, 如甲状腺髓样癌、未分化癌、滤泡性甲状腺癌, 发病率仅占甲状腺恶性肿瘤的2%~10%^[23-25], 其超声图像缺乏典型的恶性特征。对于恶性特征不典型的结节, 临床经验不足的低年资超声医师单纯依靠超声图像容易误判。本研究所用的AI-SONIC™ Thyroid 5.3.3.0诊断系统通过深度学习算法不断迭代升级, 创新性地引入Res-Net残差神经网络架构, 通过增加网络深度提升算法模型提高了准确度。本研究结果显示AI系统、高年资超声医师和低年资超声医师鉴别诊断

甲状腺结节良恶性的AUC分别为0.809(95% CI 0.764~0.854)、0.865(95% CI 0.825~0.904)和0.721(95% CI 0.672~0.771), AI系统的AUC高于低年资超声医师($P=0.032$), 而与高年资超声医师之间差异无统计学意义。

AI系统对不同年龄及不同性别甲状腺结节患者的诊断准确度差异无统计学意义, 说明AI-SONIC™ Thyroid 5.3.3.0在不同人群中诊断稳定性较好。本研究还分析了AI系统的诊断准确度是否受到甲状腺结节大小的影响, 结果表明AI系统诊断最大直径 <5 mm的结节灵敏度最高(99.0%)。既往研究表明, 以纵横比 >1 作为甲状腺乳头状癌的诊断标准, 其灵敏度随着结节大小的增加而降低^[26], 这与甲状腺癌的生长特性有关。早期甲状腺癌病灶往往呈纵向生长, 在声像图上表现为纵横比 >1 ^[27], 故纵横比 >1 对微小结节诊断灵敏度高; 后期癌细胞向各个方向生长, 其形态倾向椭圆形, 故纵横比 >1 在诊断大结节时灵敏度较低。既往研究报告AI系统对直径 <5 mm的小结节诊断效能与中小结节(5~ <10 mm)的诊断效能无明显差异^[12], 本研究发现2组的AUC分别为0.803、0.760。当恶性结节最大直径为10~ <15 mm时, 其边缘侵犯、形态不规则等恶性特征显著, AI系统对此类结节的鉴别诊断效能最高(AUC为0.882)。当结节最大直径为15~ <20 mm或 ≥ 20 mm时, AI系统的鉴别诊断灵敏度较低(59.1%、59.5%)。这可能是由于结节较大, 占据了甲状腺正常本底的大部分区域, 部分甲状腺结节突破前后包膜与周边组织粘连, 甚至可能超出了AI系统的识别范围, 导致AI系统难以区分结节与正常甲状腺组织; 此外, 较大的恶性结节组织分布不均、存在异质性, 部分恶性结节的局部区域还可出现坏死、液化, 都在一定程度上影响了AI系统的诊断灵敏度和准确度。

目前AI系统分析的是超声医师剪切存储的静态超声图像, 但静态图片不能全面显示结节的所有特征, 应连续、多切面、多维度观察结节特征, 综合分析后做出诊断。因此, 针对甲状腺结节的纵切面或横切面图像的诊断效能仍有较大的提升空间, 后期本团队将继续优化甲状腺结节的评估算法, 探讨如何最优地将基于单个图像得到的恶性概率转化为基于整个结节进行诊断的恶性概率, 进一步提升

AI系统的诊断效能。

本研究还统计了诊断良、恶性甲状腺结节的时间, 发现高年资超声医师诊断良性甲状腺结节用时短于诊断恶性甲状腺结节用时, 考虑因为大多数良性结节的结构特征比较明确, 而恶性甲状腺结节的结构特征相对复杂、多变, 医师需要一定的思考时间; 而AI系统基于深度学习技术, 其诊断良、恶性甲状腺结节的反馈时间是相对一致的, 并且明显短于超声医师。

笔者团队多年来针对AI-SONIC™ Thyroid系统开展了系列应用研究, 在前期已加载临床单中心数据进行了较深入的探索^[28-29], 本研究进一步验证了AI-SONIC™ Thyroid 5.3.3.0系统的临床应用场景与价值。但该系统仍存在待改进之处, 如因样本量所限, 偏良性结节、罕见结节的类型数据学习训练不够成熟, 后期需要多注重获取此类结节的资料进行整合学习并加以验证。

[参 考 文 献]

- [1] 刘宗超, 李哲轩, 张阳, 等. 2020全球癌症统计报告解读[J]. 肿瘤综合治疗电子杂志, 2021, 7(2): 1-14. DOI: 10.12151/JMCM.2021.02-01.
- [2] 周丽琴. 超声对甲状腺癌和甲状腺腺瘤的诊断和鉴别诊断价值研究[J]. 影像研究与医学应用, 2022, 6(7): 146-148. DOI: 10.3969/j.issn.2096-3807.2022.07.049.
- [3] CHOI S H, KIM E K, KWAK J Y, et al. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules[J]. Thyroid, 2010, 20(2): 167-172. DOI: 10.1089/thy.2008.0354.
- [4] DURANTE C, GRANI G, LAMARTINA L, et al. The diagnosis and management of thyroid nodules[J]. JAMA, 2018, 319(9): 914. DOI: 10.1001/jama.2018.0898.
- [5] LA VECCHIA C, MALVEZZI M, BOSETTI C, et al. Thyroid cancer mortality and incidence: a global overview[J]. Int J Cancer, 2015, 136(9): 2187-2195. DOI: 10.1002/ijc.29251.
- [6] ZHOU J, YIN L, WEI X, et al. 2020 Chinese guidelines for ultrasound malignancy risk stratification of thyroid nodules: the C-TIRADS[J]. Endocrine, 2020, 70(2): 256-279. DOI: 10.1007/s12020-020-02441-y.
- [7] 刘再毅, 石镇维. 医学影像人工智能: 进展和未来[J]. 国际医学放射学杂志, 2023, 46(1): 1-4. DOI: 10.19300/j.2023.s20494.
- [8] ZHANG Y, WU Q, CHEN Y, et al. A clinical assessment of an ultrasound computer-aided diagnosis system in differentiating thyroid nodules with radiologists of different diagnostic experience[J]. Front Oncol, 2020, 10: 557169. DOI: 10.3389/fonc.2020.557169.

- [9] 张蕊,牛丽娟.基于常规超声的深度学习技术在甲状腺结节良恶性鉴别中的研究进展[J].癌症进展,2022,20(8):757-759,765. DOI: 10.11877/j.issn.1672-1535.2022.20.08.01.
- [10] JIN Z, ZHU Y, ZHANG S, et al. Ultrasound computer-aided diagnosis (CAD) based on the thyroid imaging reporting and data system (TI-RADS) to distinguish benign from malignant thyroid nodules and the diagnostic performance of radiologists with different diagnostic experience[J]. *Med Sci Monit*, 2020, 26: e918452. DOI: 10.12659/MSM.918452.
- [11] JEONG E Y, KIM H L, HA E J, et al. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators[J]. *Eur Radiol*, 2019, 29(4): 1978-1985. DOI: 10.1007/s00330-018-5772-9.
- [12] WANG L, YANG S, YANG S, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network[J]. *World J Surg Oncol*, 2019, 17(1): 12. DOI: 10.1186/s12957-019-1558-z.
- [13] SHIN I, KIM Y J, HAN K, et al. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland[J]. *Ultrasonography*, 2020, 39(3): 257-265. DOI: 10.14366/usg.19069.
- [14] HAJIAN-TILAKI K. Sample size estimation in diagnostic test studies of biomedical informatics[J]. *J Biomed Inform*, 2014, 48: 193-204. DOI: 10.1016/j.jbi.2014.02.013.
- [15] ZHOU J, SONG Y, ZHAN W, et al. Thyroid imaging reporting and data system (TIRADS) for ultrasound features of nodules: multicentric retrospective study in China[J]. *Endocrine*, 2021, 72(1): 157-170. DOI: 10.1007/s12020-020-02442-x.
- [16] 中华医学会超声医学分会浅表器官和血管学组,中国甲状腺与乳腺超声人工智能联盟.2020甲状腺结节超声恶性危险分层中国指南:C-TIRADS[J].中华超声影像学杂志,2021(3):185-200. DOI: 10.3760/cma.j.cn131148-20210205-00092.
- [17] NIETHAMMER M, KWITT R, VIALARD F X. Metric learning for image registration[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 8455-8464. DOI: 10.1109/CVPR.2019.00866.
- [18] 许敏,韩峰,罗晓,等.计算机辅助诊断系统鉴别甲状腺结节良恶性的诊断效能及其影响因素[J].中华医学超声杂志(电子版),2019,16(4):252-256. DOI: 10.3877/cma.j.issn.1672-6448.2019.04.004.
- [19] 张勇,鲁晓,李昕,等.超声医学图像及报告质量管理体系设计与应用[J].中国卫生质量管理,2022,29(4):73-76. DOI: 10.13912/j.cnki.chqm.2022.29.04.19.
- [20] BALOCH Z W, ASA S L, BARLETTA J A, et al. Overview of the 2022 WHO classification of thyroid neoplasms[J]. *Endocr Pathol*, 2022, 33(1): 27-63. DOI: 10.1007/s12022-022-09707-3.
- [21] CARTER J V, PAN J, RAI S N, et al. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves[J]. *Surgery*, 2016, 159(6): 1638-1645. DOI: 10.1016/j.surg.2015.12.029.
- [22] AKKUS Z, CAI J, BOONROD A, et al. A survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow[J]. *J Am Coll Radiol*, 2019, 16(9): 1318-1328. DOI: 10.1016/j.jacr.2019.06.004.
- [23] PAPARODIS R D, KARVOUNIS E, BANTOUNA D, et al. Large, slowly growing, benign thyroid nodules frequently coexist with synchronous thyroid cancers[J]. *J Clin Endocrinol Metab*, 2022, 107(8): e3474-e3478. DOI: 10.1210/clinem/dgac242.
- [24] 徐文静,蔡尚,田野.甲状腺未分化癌的研究进展[J].中国肿瘤临床,2021,48(1):30-34. DOI: 10.3969/j.issn.1000-8179.2021.01.169.
- [25] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2018[J]. *CA Cancer J Clin*, 2018, 68(1): 7-30. DOI: 10.3322/caac.21442.
- [26] 宋丹琳,郑静,倪志鹏,等.超声图像人工智能辅助对甲状腺结节定性诊断价值的研究[J].中国临床医学影像杂志,2023,34(2):87-90. DOI: 10.12117/jccmi.2023.02.003.
- [27] 王洪杰,于霞,张恩东.人工智能辅助超声对甲状腺结节检出的研究进展[J].中国中西医结合影像学杂志,2020,18(4):424-426. DOI: 10.3969/j.issn.1672-0512.2020.04.027.
- [28] 郭芳琪,赵佳琦,刘晟.人工智能自动检测系统在甲状腺结节术前超声诊断中的应用[J].第二军医大学学报,2019,40(11):1183-1189. DOI: 10.16781/j.0258-879x.2019.11.1183.
- [29] 郭芳琪,赵佳琦,陈蕊,等.2.0版人工智能自动检测系统对甲状腺结节术前超声评估效能的初步探讨[J].第二军医大学学报,2020,41(10):1077-1083. DOI: 10.16781/j.0258-879x.2020.10.1077.
- GUO F Q, ZHAO J Q, LIU S. Application of artificial intelligence automatic detection system in preoperative ultrasonic diagnosis of thyroid nodules[J]. *Acad J Sec Mil Med Univ*, 2019, 40(11): 1183-1189. DOI: 10.16781/j.0258-879x.2019.11.1183.
- GUO F Q, ZHAO J Q, CHEN R, et al. Efficacy of preoperative ultrasound evaluation of thyroid nodules by artificial intelligence automatic detection system version 2.0: a preliminary study[J]. *Acad J Sec Mil Med Univ*, 2020, 41(10): 1077-1083. DOI: 10.16781/j.0258-879x.2020.10.1077.