

2017 American Radiology of Society thyroid imaging report and data system classification in differentiating benign and malignant thyroid nodules

CHEN Yongyu¹, HUANG Weijun^{1*}, ZHONG Minying²

(1. Department of Ultrasound and Dermatology, the First People's Hospital of Foshan, Foshan 528000, China; 2. Department of Ultrasound, People's Hospital of Foshan Nanhai, Foshan 528000, China)

[Abstract] **Objective** To explore the value of thyroid imaging report and data system (TI-RADS) recommended by American Radiology of Society (ACR) in 2017 for differentiating thyroid benign and malignant nodules. **Methods** Ultrasonic images of 497 thyroid nodules diagnosed with pathology were retrospectively analyzed using TI-RADS classification and given score and classification. The optimal critical score was determined with ROC curve, and the sensitivity, specificity and accuracy were calculated. The efficacy and consistency of senior and junior physicians in diagnosis of benign and malignant thyroid nodules using this optimal critical score were evaluated. **Results** The area under ROC curve was 0.883 ($P < 0.001$), and the optimal critical score was 5, while the sensitivity and specificity was 86.22% and 78.68%, respectively. The sensitivity, specificity and accuracy of senior physicians in diagnosis of benign and malignant thyroid nodules was 78.22% (176/225), 76.47% (208/272) and 77.26% (384/497), of junior physicians was 77.33% (174/225), 74.26% (202/272) and 75.65% (376/497), respectively, and moderate consistency was found between senior and junior physicians ($Kappa = 0.581$). **Conclusion** TI-RADS recommended by ACR in 2017 has high clinical value, and the optimal critical score for diagnosing benign and malignant thyroid nodule is 5.

[Keywords] thyroid nodule; ultrasonography; thyroid imaging report and data system

DOI: 10.13929/j.1672-8475.201809016

2017 年美国放射学会甲状腺影像报告与数据系统鉴别甲状腺良恶性结节

陈泳愉¹, 黄伟俊^{1*}, 钟敏莹²

(1. 佛山市第一人民医院超声诊疗中心, 广东 佛山 528000;
2. 佛山市南海区人民医院超声科, 广东 佛山 528000)

[摘要] **目的** 探讨 2017 年美国放射学会 (ACR) 推荐的甲状腺影像报告与数据系统 (TI-RADS) 对鉴别甲状腺良恶性结节的应用价值。 **方法** 回顾性分析经手术病理确诊的 497 个甲状腺结节的超声资料, 采用 2017 年 ACR TI-RADS 分类对结节评分和分类, 与病理结果对照, 采用 ROC 确定最佳临界评分, 计算 TI-RADS 分类诊断良恶性结节的敏感度、特异度和准确率。评价高年资与低年资医师采用此最佳临界评分诊断甲状腺良恶性结节的效能和一致性。 **结果** 采用 2017 年 ACR TI-RADS 分类诊断甲状腺恶性结节的 ROC 曲线下面积为 0.883 ($P < 0.001$), 最佳临界评分为 5 分, 敏感度 86.22%, 特异度 78.68%。以 TI-RADS 评分 > 5 分归为恶性结节, 高年资和低年资医师诊断的敏感度、特异度、准确率分

[第一作者] 陈泳愉 (1975—), 女, 广东佛山人, 学士, 副主任医师。研究方向: 甲状腺及乳腺浅表器官超声诊断。E-mail: 385333158@qq.com

[通信作者] 黄伟俊, 佛山市第一人民医院超声诊疗中心, 528000。E-mail: hwjun36@126.com

[收稿日期] 2018-09-16 **[修回日期]** 2019-03-28

别为 78.22%(176/225)、76.47%(208/272)、77.26%(384/497)和 77.33%(174/225)、74.26%(202/272)、75.65%(376/497),二者诊断良恶性结节具有中等一致性($Kappa=0.581$)。结论 2017 年 ACR TI-RADS 具有较高临床应用价值,鉴别甲状腺良恶性结节的最佳界值为 5 分。

[关键词] 甲状腺结节;超声检查;甲状腺影像报告和数据系统

[中图分类号] R581;R445.1 [文献标识码] A [文章编号] 1672-8475(2019)05-0294-05

甲状腺结节是内分泌系统常见疾病,近年来,随着超声技术的发展,甲状腺恶性结节的检出率逐年递增,但由于对甲状腺结节缺乏规范化管理,在判别良恶性结节方面存在较强的主观性。2017 年美国放射学会(American College of Radiology, ACR)根据 5 项恶性征象提出了相对简单的甲状腺影像报告与数据系统(thyroid imaging reporting and data system, TI-RADS)^[1],通过计算总分对结节分类,但并未提出鉴别良恶性结节的最佳临界值,且目前相关研究^[2-3]报道较少。本研究回顾性分析甲状腺结节的超声资料,探讨 2017 年 ACR TI-RADS 分类在鉴别诊断甲状腺良恶性结节中的价值及最佳临界值。

1 资料与方法

1.1 一般资料 回顾性收集 2015 年 10 月—2017 年 10 月在佛山市第一人民医院接受甲状腺超声检查并具有手术病理结果的 495 例患者(共 497 个结节)的声像图,男 124 例,女 371 例,年龄 14~84 岁,平均(44.0±12.5)岁;结节最大径 3~57 mm,平均(16.80±9.60)mm。

1.2 仪器与方法 采用 Siemens S2000(9L4 探头,频率 5~14 MHz)、百盛 Mylab 60(LA523 探头,频率 5~12 MHz)及 GE Logiq E9(ML6-15 探头,频率 6~15 MHz)彩色超声诊断仪。嘱患者仰卧,充分暴露颈部,全面扫查甲状腺,发现结节时留取完整声像图,描述并记录结节各项超声特征。

1.3 图像分析 由 2 名具有 15 年以上甲状腺诊断经验的医师在未知病理结果的情况下根据 2017 年 ACR TI-RADS 对结节进行分类,意见不同时共同商议得出一致结果。再由 1 名具有 10 年以上甲状腺诊断经验的高年资医师及 1 名具有 2 年甲状腺诊断经验的低年资医师在未知病理结果的情况下观察结节大小、形态、边缘、回声、内部结构及钙化等,对结节良恶性做出经验诊断;再分别于学习 2017 年 ACR TI-RADS 分类后各自对结节进行评分,做出 TI-RADS 诊断。

2017 年 ACR TI-RADS 标准从甲状腺结节的成分、回声、形状、边缘及钙化 5 个方面进行评分与分类。分类标准:成分,包括囊性(0 分)、海绵状结节(0 分)、

混合结节(1 分)、实性或几乎为实性(2 分);回声,包括无回声(0 分)、高、等回声,即回声高于甲状腺腺体、与甲状腺腺体类似(1 分)、低回声,即回声低于甲状腺腺体(2 分)、极低回声,即回声低于邻近颈部肌肉、颈前肌群(3 分);形状:非直立生长(0 分)、直立生长(3 分);边缘,包括光滑(0 分)、欠清(0 分)、不规则或分叶状(2 分)、甲状腺外侵袭(3 分);钙化(可多选),包括无或伴彗星尾征(0 分)、粗大钙化(1 分)、周围(环状)钙化(2 分)、点状强回声(3 分)。总分 0 分归为 TI-RADS 1 类(当结节成分为海绵状结节时,回声、形状、边缘各项不再计分),2 分归为 TI-RADS 2 类,3 分归为 TI-RADS 3 类,4~6 分归为 TI-RADS 4 类,≥7 分归为 TI-RADS 5 类。

1.4 统计学分析 采用 SPSS 20.0 统计分析软件,计数资料以例表示,计量资料以 $\bar{x} \pm s$ 表示。以组织病理学为金标准,根据 2 名具有 15 年以上甲状腺诊断经验的医师共同商议得出的 TI-RADS 总分构建 ROC 曲线获得最佳临界评分;采用 χ^2 检验比较高、低年资医师经验诊断及 TI-RADS 诊断的效能。采用 $Kappa$ 检验分析高、低年资医师诊断良恶性结节的一致性, $Kappa$ 值 ≤0.20 为一致性差,>0.20~0.40 为一致性一般,>0.40~0.60 为一致性中等,>0.60~0.80 为一致性好,>0.80 为一致性非常好。以 $P < 0.05$ 为差异有统计学意义。

2 结果

497 个结节中,良性结节 272 个,包括结节性甲状腺肿 261 个,滤泡性腺瘤 11 个;恶性结节 225 个,包括甲状腺乳头状癌 220 个,髓样癌 4 个,滤泡癌 1 个。TI-RADS 1~5 类结节分别占 14.08%(70/497)、10.87%(54/497)、10.46%(52/497)、23.54%(117/497)及 41.05%(204/497),TI-RADS 分类结果与病理结果比较差异有统计学意义($\chi^2 = 235.73, P < 0.001$)。不同 TI-RADS 分类结节的病理结果见表 1。

以病理诊断结果为金标准,TI-RADS 分类诊断甲状腺恶性结节的 AUC 为 0.883 [95% CI (0.854, 0.912)], $P < 0.001$,最佳临界评分为 5 分,敏感度为 86.22%(194/225),特异度为 78.68%(214/272),约登指数为 0.649,见图 1。TI-RADS 3~5 类甲状腺结节的声像图表现见图 2。

表 1 甲状腺结节 TI-RADS 分类标准及总分与病理结果比较

TI-RADS 分类	TI-RADS 总分	病理结果(个)		合计(个)	恶性百分比(%)
		良性	恶性		
1	0	68	2	70	2.86
2	2	49	5	54	9.26
3	3	46	6	52	11.54
4	4	51	18	69	26.09
5	5	9	12	21	57.14
	6	10	17	27	62.96
	7	12	39	51	76.47
	8	13	17	30	56.67
	9	6	19	25	76.00
	10	5	38	43	88.37
	11	2	14	16	87.50
	12	0	14	14	100
13	1	13	14	92.86	
14	0	10	10	100	
15	0	1	1	100	

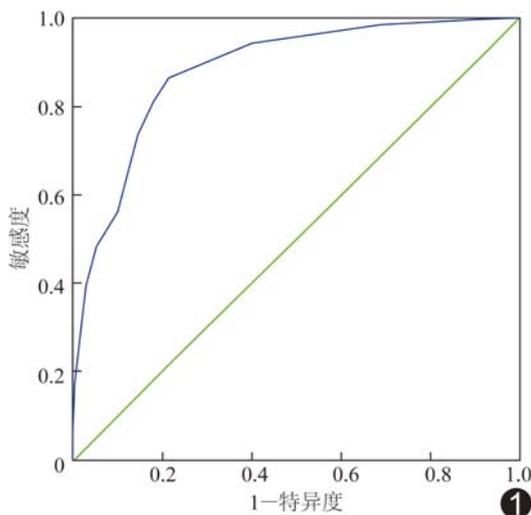


图 1 2017 年 ACR TI-RADS 鉴别甲状腺良恶性结节的 ROC 曲线

以 >5 分作为 TI-RADS 诊断甲状腺恶性结节的标准,高、低年资医师分别诊断良性结节 208、202 个,恶性结节 176、174 个,差异均有统计学意义($\chi^2 = 147.511, 131.179, P < 0.05$)。高年资医师凭经验诊断的敏感度、特异度、准确率、阳性预测值、阴性预测值与 TI-RADS 诊断差异无统计学意义(P 均 > 0.05);低年资医师 TI-RADS 诊断甲状腺恶性结节的敏感度、阴性预测值均较经验诊断更高($\chi^2 = 10.252, 7.320, P = 0.020, 0.009$),特异度、阳性预测值、准确率与经验诊断差异无统计学意义(P 均 > 0.05);见表 2。高年资医师凭经验诊断的敏感度、特异度、阳性预测值、阴性预测值、准确率均较低年资医师更高($\chi^2 = 9.565, 11.333, 11.314, 9.521, 10.345, P$ 均 < 0.05),采用 TI-RADS 诊断的敏感度、特异度、阳性预测值、阴性预测值、准确率与低年资医师差异均无统计学意义($\chi^2 = 0.051, 0.356, 0.247, 0.096, 0.312, P$ 均 > 0.05)。

以 >5 分作为 TI-RADS 鉴别良恶性结节的标准,不同年资医师诊断良恶性结节的差异有统计学意义($\chi^2 = 167.944, P < 0.001$;表 3);一致性分析结果显示,二者具有中等一致性($Kappa = 0.581$)。

3 讨论

为使甲状腺结节的管理更加规范化,参照乳腺影像报告和数据系统(breast imaging reporting and data system, BI-RADS),国外学者^[4-6]相继提出了 TI-RADS,其临床诊断价值已得到认可^[7-8],但因标准不统一,在临床应用中仍存在争议。国内学者^[9]建立了 TI-RADS 超声评价体系,认为凭借对恶性征象的权重与组合更有利于诊断恶性结节,与 2017 年 ACR TI-RADS 分类将 5 项超声恶性征象赋予不同分值以评估结节的思路相当,目的是突出恶性征象及量化甲状腺结节的评估流程。

表 2 TI-RADS 分类及经验诊断甲状腺结节的诊断效能比较[% (个)]

医师	敏感度	特异度	阳性预测值	阴性预测值	准确率
高年资医师					
TI-RADS 分类	78.22(176/225)	76.47(208/272)	73.33(176/240)	80.93(208/257)	77.26(384/497)
经验诊断	76.89(173/225)	81.25(221/272)	77.23(173/224)	80.95(221/273)	79.28(394/497)
χ^2 值	0.115	1.864	0.945	0.000	1.017
P 值	0.821	0.208	0.336	1.000	0.832
低年资医师					
TI-RADS 分类	77.33(174/225)	74.26(202/272)	71.31(174/244)	79.84(202/253)	75.65(376/497)
经验诊断	63.56(143/225)	68.75(187/272)	56.52(143/253)	69.52(187/269)	66.40(330/497)
χ^2 值	10.257	2.422	3.945	7.320	5.832
P 值	0.020	0.130	0.052	0.009	0.061



图 2 TI-RADS 3~5 类甲状腺结节声像图 A. TI-RADS 3 类结节(甲状腺滤泡腺瘤), 实性(2 分), 高、等回声(1 分), 纵横比<1(0 分), 边缘光滑(0 分), 无钙化(0 分), 总分 3 分; B. TI-RADS 4 类结节(甲状腺髓样癌), 实性(2 分), 低回声(2 分), 纵横比<1(0 分), 边缘欠清(0 分), 粗大钙化(1 分), 总分 5 分; C. TI-RADS 5 类结节(甲状腺乳头状癌), 实性(2 分), 低回声(2 分), 纵横比<1(0 分), 外侵袭(3 分), 无钙化(0 分), 总分 7 分

表 3 不同年资医师采用 TI-RADS 鉴别诊断甲状腺良恶性结节的一致性检验(个)

低年资医师	高年资医师		合计
	良性	恶性	
良性	203	50	253
恶性	54	190	244
合计	257	240	497

本研究中,以 2017 年 ACR TI-RADS 分类构建 ROC 曲线下面积为 0.883($P < 0.001$),提示其对鉴别甲状腺良恶性结节具有较高诊断价值;最佳临界评分为 5 分时,敏感度为 86.22% (194/225),特异度为 78.68% (214/272)。因此,本研究设定了 ≤ 5 分为 TI-RADS 诊断良性结节标准、 > 5 分为诊断恶性结节的标准。

低年资医师采用 TI-RADS 诊断甲状腺结节的敏感度较经验诊断有所提高(77.33% vs 63.55%),差异有统计学意义($\chi^2 = 10.252, P = 0.020$),提示 TI-RADS 分类可提高低年资医师的诊断水平。高年资医师凭借经验诊断的敏感度、特异度、阳性预测值、阴性预测值、准确率均较低年资医师更高(P 均 < 0.05),但采用 TI-RADS 诊断后,二者敏感度、特异度、准确率、阳性预测值、阴性预测值差异均无统计学意义(P 均 > 0.05),提示统一甲状腺超声报告术语的规范化管理及诊断标准对年轻医师具有很好的指导作用,可在一定程度上拉近高、低年资医师间的诊断水平。此外,本研究发现高、低年资医师的评估结果具有中等一致性($Kappa = 0.581$),提示以 > 5 分作为 TI-RADS 鉴别良恶性结节的分界点具有普遍适用性;但本研究所见

的一致性相比以往研究^[10]稍低,考虑原因在于本研究中的高年资医师积累了一定的诊断经验,对 TI-RADS 超声描述术语的理解较低年资医师更加深入。

既往研究^[11]报道,采用 TI-RADS 诊断甲状腺良恶性结节的敏感度、特异度及准确率分别为 0.912、0.851、0.893;本研究中,高年资医师 TI-RADS 诊断的敏感度、特异度及准确率分别为 78.22% (176/225)、76.47% (208/272)、77.26% (384/497),低年资医师则为 77.33% (174/225)、74.26% (202/272)、75.65% (376/497)。本研究诊断效能偏低的原因可能在于属回顾性分析,诊断医师缺乏充足的诊断信息,包括病史、体征、病灶多个切面的动态扫查、彩色血流图以及颈部淋巴结情况等,在一定程度上降低了诊断准确率。

本研究的局限性:①未纳入具有桥本甲状腺炎背景的结节;②入选结节均选自手术患者而非普通人群,可能有选择性偏倚,从而影响验证结果;③疾病谱较窄,需进一步增加样本量并进行多中心研究;④未应用弹性成像和超声造影等新技术,今后将继续完善。

综上所述,2017 年 ACR TI-RADS 分类标准可较准确地鉴别甲状腺良恶性结节,总分 ≤ 5 分可归为良性结节, > 5 分可归为恶性结节,且高、低年资医师的诊断一致性较好,对于临床诊断和治疗具有重要提示作用。

[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] 刘红,胡正明,罗海愉,等. ACR TI-RADS 分类在诊断甲状腺结节中的应用价值探究. 中国超声医学杂志, 2018, 34(8):673-675.
- [3] 王潇婧,刘利平,孙永清,等. ACR TI-RADS 分类与 2015 年 ATA 指南对甲状腺结节诊断价值的比较研究. 中华临床医师杂志(电子版), 2018, 12(4):223-226.
- [4] Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab, 2009, 94(5):1748-1751.
- [5] Park JY, Lee HJ, Jang HW, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. Thyroid, 2009, 19(11):1257-1264.
- [6] Kwak JY, Han KH, Yoon JH, et al. Thyroid imaging reporting and data system for US features of nodules: Astep in establishing better stratification of cancer risk. Radiology, 2011, 260(3):892-899.
- [7] 钟敏莹,石小红,杨丽丽,等. TI-RADS 分类系统对不同直径甲状腺结节的诊断价值. 中国超声医学杂志, 2016, 32(4):289-291.
- [8] 刘如玉,姜玉新,杨筱,等. 甲状腺结节的影像报告与数据系统分级与 2015 年美国甲状腺学会推荐超声恶性风险分层的比较研究. 中华医学超声杂志(电子版), 2017, 14(4):263-268.
- [9] 李玲,吕国荣,沈浩霖,等. TI-RADS 分类超声客观评价体系的建立和应用. 中国医学影像技术, 2017, 33(5):698-702
- [10] 郑丽萍,徐晚虹,吕国荣,等. 优化 TI-RADS 分类鉴别诊断甲状腺良恶性结节. 中国医学影像技术, 2016, 32(11):1668-1672.
- [11] 侯雪琴,范雪,杨凌菲,等. 甲状腺影像报告和数据系统分类诊断甲状腺结节良恶性的准确性. 中国医学影像学杂志, 2015, 23(7):489-493.

《中国介入影像与治疗学》投稿要求(二)

8 基金项目 书写格式应为:××基金(编号××)。基金论文投稿时应同时将基金项目批文扫描上传。本刊将优先选登基金论文。

9 第一作者 书写格式为:姓名(出生年—),性别,籍贯,学位,职称,研究方向和 E-mail。尚未毕业者学位写为“在读××”,如“在读硕士”、“在读博士”。投稿时,简介中需注明第一作者联系电话。

10 通信作者 通常为稿件所涉及研究工作的负责人,通信作者简介包括其姓名、单位、邮编和 E-mail。

11 摘要 研究论著和综述性论文均须附中、英文摘要,英文在前,中文在后;经验交流和短篇报道不设摘要。摘要应以第三人称撰写,不宜超过 350 个汉字,不宜使用“我们”、“作者”等词作为主语。研究论著类论文的摘要应按结构式摘要的特征撰写,包括目的(Objective)、方法(Methods)、结果(Results)、结论(Conclusion)四项。综述性论文的摘要应按指示性摘要的特征撰写,简要叙述文章内容,无须按“四项式”编写。摘要应具有独立性和自明性,并且拥有与文献同等量的主要信息,在有限的字数内向读者提供尽可能多的定性或定量的信息,充分反映该研究的创新之处。英文摘要在 300 个实词左右,内容和形式要求与中文摘要基本对应。

12 关键词 中文关键词应以《医学主题词注释字顺表(2002 年版)》为准,以标引能表达全文主题概念的叙词 2~5 个,尽量少用自由词。英文关键词应与中文对应,以《医学主题词注释字顺表(2002 年版)》及《Mesh》内所列的主题词为准,每个英文关键词第一个词的首字母大写,各关键词之间用分号隔开。

13 正文标题层次 如:1 资料与方法(或材料与与方法),1.1 资料(或材料),1.2 方法,2 结果,3 讨论,一般不超过三级标题,文内连序号为圈码,如①、②、③……表示。

14 医学名词 应注意规范、标准、前后统一,尽量少用缩略语,原词过长且在文中多次出现(2 次以上)者,可于括号内写出全称并加注简称,以后用简称。医学名词术语,以全国科学技术名词审定委员会审定公布的医学名词为准。