

doi:10.3969/j.issn.1005-3697.2026.04.004

✧ 智能医学研究专题 ✧

# 影像组学可能增加经验丰富医师对于甲状腺结节 ACR TI-RADS4,5 级的诊断信心

王娣<sup>1</sup>, 刘晓玲<sup>2</sup>, 姚娇<sup>2</sup>, 覃夏川<sup>3</sup>

(1. 眉山市第二人民医院; 四川 眉山 620500; 2. 首都医科大学附属北京安贞医院南充医院; 四川 南充 637000; 3. 成都市第二人民医院超声科, 四川 成都 610000)

**【摘要】目的:** 本研究探讨了影像组学在超声放射学会甲状腺影像报告与数据系统(ACR TI-RADS) 4 级和 5 级甲状腺结节诊断中的应用价值。**方法:** 回顾性分析 221 例接受甲状腺切除手术的 ACR TI-RADS 4 级和 5 级患者的临床病理和超声检查数据。图像被随机分为训练集和测试集, 使用 MaZda 软件提取影像组学特征, 并通过组内和组间相关系数(ICC), Logistic 回归和最小绝对收缩选择算子算法(LASSO)筛选特征。建立 Logistic 回归、贝叶斯(Bayes)和 K 临近(KNN)机器学习模型, 并评估模型的诊断效能。**结果:** 从超声图像中提取了 314 个特征, 筛选出 5 个最具诊断意义的特征。基于这些特征建立的机器学习模型中 KNN 模型表现最佳, 训练组和测试组的 ROC 分别为 0.849 和 0.885, 准确率分别为 0.779 和 0.761。与经验丰富的专家读片相比, KNN 模型表现更优。**结论:** 基于甲状腺超声的超声组学模型在对于甲状腺 ACR 4 级和 5 级结节的诊断中取得了突出的表现, 优于经验丰富的超声诊断医师, 为临床常规 ACR 4 级和 5 级结节鉴定提供有效的诊断参考。

**【关键词】** 甲状腺结节; ACR TI-RADS; 区分; 超声; 影像组学

**【中图分类号】** R445; R445.1; R736.1 **【文献标志码】** A

## Radiomics has the potential to increase the diagnostic confidence of experienced clinicians in assessing ACR TI-RADS category 4 and 5 thyroid nodules

WANG Di<sup>1</sup>, LIU Xiao-ling<sup>2</sup>, YAO Jiao<sup>2</sup>, QIN Xia-chuan<sup>3</sup>

(1. The Second People's Hospital of Meishan, Renshou 620500; 2. Nanchong Hospital, Beijing Anzhen Hospital, Capital Medical Hospital, Nanchong 637000; 3. Chengdu Second People's Hospital, Chengdu 610000, Sichuan, China)

**【Abstract】Objective:** To investigate the diagnostic value of radiomics in thyroid nodules of American College of Radiology-Thyroid Imaging Reporting and Data System (ACR TI-RADS) grades 4 and 5. **Methods:** A retrospective study of 221 patients with ACR TI-RADS grades 4 and 5 who underwent thyroidectomy was performed. The data on clinicopathological and ultrasound examinations of the patients were analyzed. The images were randomized into a training set and a validation set. MaZda, a software for calculating texture parameters in digitized images, was employed to manually outline the images and extract radiomics features. Features were filtered through the univariate Logistic analysis and LASSO method. The Logistic regression model, Bayes model and KNN model were trained using the selected texture features. **Results:** 314 radiomics features were extracted from each patient's ultrasound image ROI using the MaZda software package. The use of LASSO further filtered five most significant features. The KNN model, based on these features, performed the best, with ROC values of 0.849 for the training group and 0.885 for the testing group, and accuracy rates of 0.779 and 0.761, respectively. Compared to readings by experienced radiologists, the KNN model demonstrated superior performance. **Conclusion:** The ultrasound omics model based on thyroid ultrasound, which is superior to experienced physician-performed ultrasound diagnosis, has shown outstanding performance in diagnosing thyroid ACR grade 4,5 nodules and provided an effective reference for the identification of clinical routine ACR 4,5 nodules.

**【Key words】** Thyroid nodules; ACR TI-RADS; Differentiated; Ultrasound; Radiomics

随着高频超声的应用,越来越多的甲状腺结节 被发现,在成年人中具有高达 68% 的患病率<sup>[1]</sup>。大

基金项目: 四川省基层卫生事业发展研究项目(SWFZ24-C-99)

作者简介: 王娣(1998-),女,硕士研究生。E-mail:1510539913@qq.com

通讯作者: 覃夏川。E-mail:11326636@qq.com

多数结节是良性的,恶性肿瘤的风险为 7%~15%<sup>[2]</sup>。美国放射学会甲状腺成像报告和数据库系统(American college of radiology-thyroid imaging reporting and data system, ACR TI-RADS)是根据结节有无恶性特征进行评分的系统,分为 ACR TI-RADS 1 至 ACR TI-RADS 5<sup>[3]</sup>。虽然其当 TI-RADS 为 4 和 5 级时,代表了中重度可疑恶性程度,同时由于结节的特征重叠,一些良性结节也会被评分为 4,5 级<sup>[4]</sup>。在实际工作中,常常采取上下级医师会诊制度,最后由经验丰富的上级诊断医师的甲状腺诊断报告可能被作为最后的诊断结果,但是即使大多数有经验的超声医师也对一些有迷惑的肿瘤图像产生怀疑,不确定的结果很大程度会影响患者管理、随访频率、细针穿刺活检工作量、患者焦虑及甲状腺结节恶性肿瘤诊断不足/过度诊断的经济负担<sup>[5]</sup>。

影像组学是一种新兴技术,其被定义为将医学图像转化为高通量特征以定量评估肿瘤表型<sup>[6-7]</sup>。可从单个或多个医学图像中定量提取肉眼无法直接观察到的高通量特征,然后应用于机器学习方法来构建分类或预测模型。该方法可以更客观地评估肿瘤状态并区分恶性与良性肿瘤或结节<sup>[8-9]</sup>。但是本研究却并未查询到关于超声组学关于甲状腺 ACR 4,5 级结节的研究。故本研究收集由丰富甲状腺诊断医师术前诊断 4,5 级的结节,尝试使用超声组学进行分析以提高诊断医师对于此类结节的诊断信心。

## 1 资料与方法

### 1.1 一般资料

回顾性分析 2019 年 9 月至 2022 年 9 月南充市中心医院接受甲状腺切除的 221 例患者的临床资料。分析患者的临床病理和超声检查数据。纳入标准:(1)接受手术并有病理结果;(2)数据库检索了具有 8 年以上甲状腺诊断经验的医师所撰写的 ACR-TIRADS 分级为 4 级和 5 级的诊断报告。(3)术前一周内行常规超声检查;(4)最大结节直径 5~30 mm;(5)甲状腺结节有实性结构。该研究得到了机构审查委员会的批准,并免除知情同意的要求。

### 1.2 仪器与方法

本研究中用于获取图像的超声仪器包括:Min-drax Resona 7(中国深圳明德瑞生物医药电子有限公司)、Esaote My Lab(意大利)。考虑到横向和纵向评估的相应诊断结果可能会重叠,同时纵切受到的干扰更少,本研究选择结节的纵向切面。

### 1.3 图像分割和特征提取

由两名对病理结果不明的具有 5 年以上甲状腺

超声诊断工作经验的医师(医师 1, 医师 2)使用 Mazda (version 4. 6, www. eletel. pl/programy/mazda/)对图像的感兴趣区域(region of interest, ROI)进行分割及特征的提取。在分割之前,使用马自达内置的归一化方法对图像进行归一化处理,以尽量减少对比度和亮度变化的影响。医师 1 和医师 2 分别对图像进行分割和特征提取,并在 2 周后由医师 1 再次对图像进行分割和特征提取。组内和组间相关系数(intra and inter correlation coefficients, ICC) $<0.75$  在随后分析中被排除。最后,选择医师 1 对剩余的图像进行分割。

### 1.4 数据预处理、特征选择和影像组学模型的建立

IPMs 2. 4. 0(GE 医疗)用于数据预处理和功能选择。数据集以 7:3 的比例随机分配给训练队列或测试队列。在分析之前排除方差为零的变量,代之以中位数填充法。最后,采用 z score 对数据进行标准化。

首先,保留了 ICCs 为  $>0.75$  的特征,并利用单变量 Logistic 分析对  $P < 0.05$  的特征进行滤波。然后,使用最小绝对收缩选择算子算法(least absolute shrinkage and selection operator, LASSO)来进一步过滤最有用的特征。最后,使用所选特征训练了 Logistic 模型, Bayes 模型, K-nearest neighbour(KNN)模型。模型验证采用十倍分层交叉验证法。大部分样本用于模型的构建或学习过程,留下一部分样本用于模型预测的验证。

最后使用受试者工作特征(ROC)曲线下面积(AUC)、敏感度、特异度、准确度、F1 分数对模型进行评估。见图 1。

### 1.5 模型与有经验的医生的比较

在建立最佳分类算法后,由上述训练集构建预测模型,然后在测试数据集中进行验证剩余 30% 的样本,保持每个类别的比例。使用相同的测试数据集,两位专家观察者分析了图像,并对病变做出良恶性判断,如有分歧,经讨论达成共识。两位放射科医生都不了解任何患者的临床数据。将结果与最佳模型进行比较。

### 1.6 统计学分析

使用 SPSS 25. 0 (IBM, Armonk, NY, USA), R3. 5. 1 (R 统计计算研究所, 维也纳, 奥地利)进行数据分析。首先采用 Kolmogorov-Smirnov 检验来评估计量资料是否呈正态分布,符合正态分布时采用  $(\bar{x} \pm s)$  表示,组间比较采用独立样本 t 检验;不符合正态分布时,采用  $[M(P_{25}, P_{75})]$  表示,组间比较采用曼-惠特尼 U 检验。对于分类变量采用  $[n(\%)]$  表示,组间比较采用独立样本  $\chi^2$  检验或

Fisher 精确概率法检验。使用受试者工作特征 (ROC) 曲线评估各模型的准确度、敏感度、特异度

和曲线下面积 (AUC)。DeLong 检验用于比较各模型之间的差异。P<0.05 为差异有统计学意义。

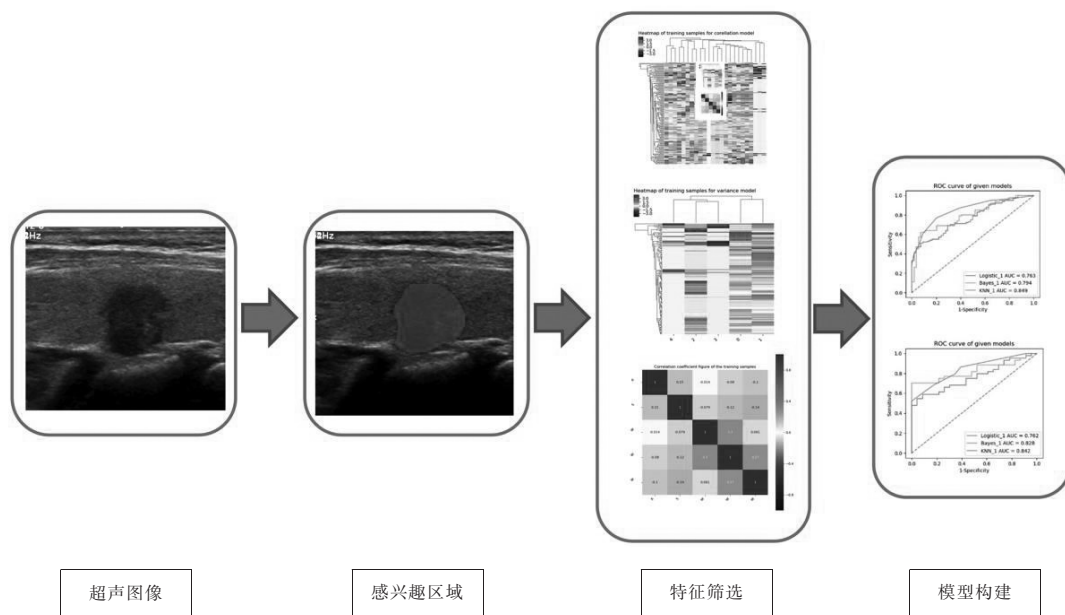


图 1 数据预处理、特征选择和影像组学模型建立的操作流程图

## 2 结果

### 2.1 两组患者一般资料比较

本研究中有 221 个 TI-RADS 4,5 类甲状腺结节,按 7:3 的比例随机分为训练集 (n=154) 和验证集 (n=67)。训练集良性 100 例,恶性 54 例;验证集良性 44 例,恶性 23 例。训练集和验证集的年龄、TI-RADS 评分、结节直径、性别、良恶性比较,差异均无统计学意义 (P>0.05)。见表 1。

表 1 训练组和测试组的临床特征 [x̄±s, M(P<sub>25</sub>, P<sub>75</sub>), n(%)]

资料	训练组 (n=154)	测试组 (n=67)	t/Z/χ <sup>2</sup> 值	P 值
年龄(岁)	47.357±13.724	46.388±13.217	0.489	0.626
性别			0.278	0.598
男	42(27.27)	16(23.88)		
女	112(72.73)	51(76.12)		
直径(mm)	11(16.8,8.325)	9.7(15.9,7.55)	1.280	0.201
ACR 分数(分)	7(9,6)	8(10,7)	1.447	0.147
ACR TI-RADS			1.105	0.294
4 级	45(29.22)	15(22.39)		
5 级	109(70.78)	52(77.61)		
病理结果			0.011	0.916
恶性	54(35.06)	23(34.33)		
良性	100(64.94)	44(65.67)		

ACR, 美国放射学会 (american college of radiology); TI-RADS, 甲状腺影像报告与数据系统 (thyroid Imaging reporting and data system)

### 2.2 特征提取, 筛选, 建立模型

使用马自达软件包从每位患者的超声图像 ROI 中提取 314 个影像组学特征。ICC>0.75 共有 296 个特征,进行多因素 Logistic 回归分析后剩下 24 个特征,再使用 LASSO 进一步筛选出了 5 个最有意义的特征: Kurtosis, Sigma, WavEnLL\_s\_7, WavEnLL\_s\_8, WavEnHL\_s\_8。最后根据这 5 个特征建立了 KNN、Logistic、Bayes 机器学习模型。三个模型测试组的平均 ROC 为 0.8, 敏感度为 0.4, 其中 KNN 的表现最好。训练组 ROC 为 0.849, 准确率为 0.779, 敏感度为 0.777, 特异度为 0.796。测试组 ROC 为 0.842, 准确率为 0.761, 敏感度为 0.727, 特异度为 0.826。见表 2、图 2 及图 3。

表 2 Logistic、Bayes 和 KNN 模型在训练组和测试组中的表现

模型类型	AUC 值	准确度	敏感度	特异度	F1 分数
Logistic					
训练组	0.763	0.649	0.500	0.926	0.649
测试组	0.762	0.701	0.591	0.913	0.722
Bayes					
训练组	0.794	0.721	0.610	0.926	0.739
测试组	0.828	0.746	0.614	1.00	0.761
KNN					
训练组	0.849	0.779	0.777	0.796	0.819
测试组	0.842	0.761	0.727	0.826	0.800

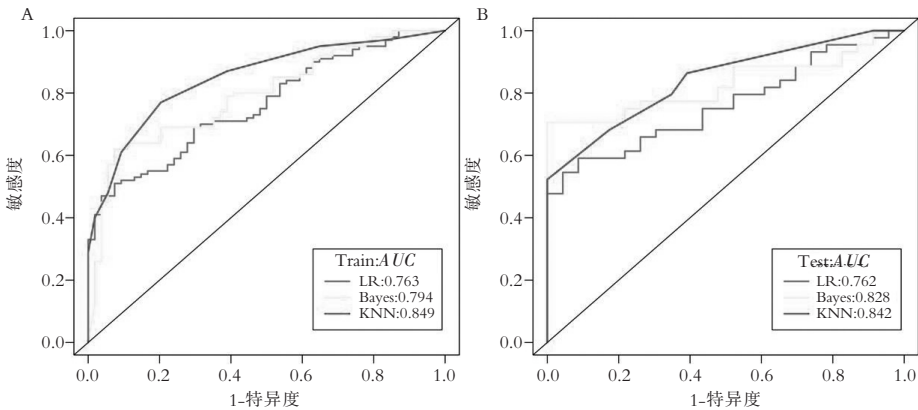


图 2 训练组和测试组中 KNN, Logistic 和 Bayes 模型的 ROC 曲线

A. 训练组三模型示意图; B. 测试组三模型示意图。

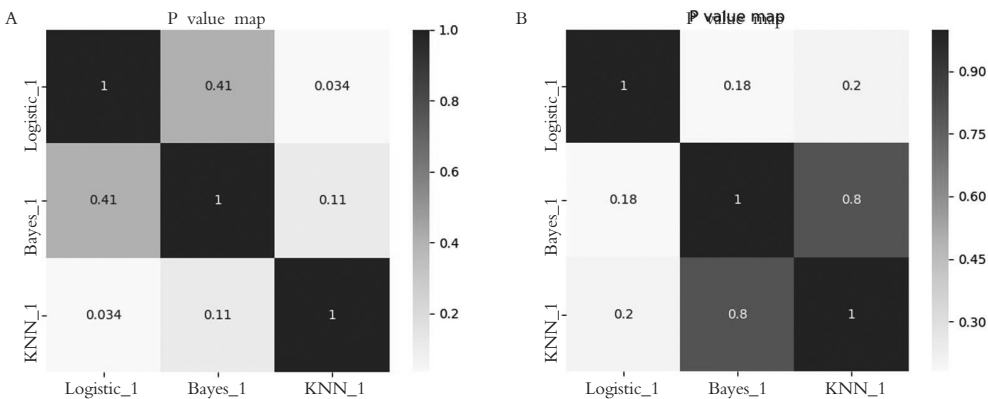


图 3 KNN, Logistic 和 Bayes 模型的 P 值示意图

A. 训练组示意图; B. 测试组示意图。两个有经验的专家读片显示, ROC 为 0.659, 准确率为 0.686, 敏感度为 0.841, 特异度为 0.313。Delong 检验结果表明 KNN 模型在训练集及验证集均优于 Logistic 模型、Bayes 模型和专家读片。

### 3 讨论

本研究表明, 在训练和测试队列中, 基于超声组学的机器学习模型能准确的诊断 ACR TI-ACR4, 5 级的结节, 其中 KNN 的表现最好, 测试组 ROC 为 0.885, 准确率为 0.761, 敏感度为 0.727, 特异度为 0.826。相比之下, 两名放射科医生的读者对于测试组的结果表现出明显较低的辨别能力 (AUC 为 0.659, 准确率为 0.686, 敏感度为 0.841, 特异度为 0.313)。因此, 基于影像组学的机器学习分类器可明显提高对甲状腺 ACR4, 5 级结节的诊断效能。

超声因其具有方便、高分辨率、安全性、无辐射等优点, 被作为是诊断甲状腺结节最重要的诊断工具<sup>[10]</sup>。ACR TI-RADS 采用定量评分的方法, 根据总分对甲状腺结节进行分类, 是临床实践中最常用的 TNs 无创危险分层工具<sup>[3]</sup>。分数范围从 0~14 分, 分数越高表示给定结节的恶性风险越大, 当其结果在 4 级和 5 级的时候, 其恶性风险的程度常常增加<sup>[11]</sup>。但在实际工作中, 不同操作者的经验和技能

会影响 TN 的准确鉴别, US 的诊断性能与执行图像采集的临床医生的经验高度相关<sup>[12-13]</sup>。有经验医师常常代表了更加准确的结果, 但是良性和恶性甲状腺结节的常规超声成像特征之间常常存在重叠。在 Koc 等<sup>[5]</sup>的研究中, 细针穿刺活检手术共发现 447 个良性结节和 45 个恶性结节, ACR-TI-RADS 获得了 59.93% 的准确率, 在近期 Thedinger 等<sup>[14]</sup>的研究中, 其使用 ACR 诊断甲状腺结节的敏感度为 73.7%, 特异度为 27.0%。一些良性结节也可能常常表现出恶性超声特征, 导致误诊及不必要的穿刺或手术治疗<sup>[15-16]</sup>。在本研究中, 团队使用影像组学诊断 4, 5 级结节, 通过 ROC, 发现准确率高干有经验丰富医师的诊断结果。

超声组学特征反映了肿瘤的纹理信息, 是肿瘤内异质性的重要标志。肿瘤内异质性可能是基因组异质性的结果<sup>[6]</sup>, 该技术可以通过计算机算法提取图像特征来定量评估高维图像特征; 因此, 它可以提取的特征比专家手动提取的特征要多得多<sup>[17]</sup>。在从数以千计的影像组学特征中选择生物标志物时,

很难以直观的方式阐明单个影像组学特征与生物行为之间的相关性。常用的方法是使用影像组学技术建立一个多特征参数用于结果估计<sup>[18]</sup>。在本研究中,团队从灰度 US 图像中提取了病变的影像组学特征。影像组学评分是基于 5 个选定的影像组学特征建立的,这些特征表明肿瘤的纹理特征。灰度甲状腺超声图像的影像组学分析在区别 ACR<sub>4,5</sub> 级结节在训练和验证队列中显示出足够的预测效率。

综上,基于甲状腺超声的超声组学模型在对于甲状腺 ACR<sub>4,5</sub> 级结节的诊断中取得了突出的表现,优于熟练的超声诊断医师,为临床常规 ACR<sub>4,5</sub> 级结节鉴定提供有效的诊断参考。

#### 参考文献

- [1] Guth S, Theune U, Aberle J, *et al.* Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination[J]. *European Journal of Clinical Investigation*, 2009, 39(8): 699–706.
- [2] Paschou SA, Vryonidou A, Goulis DG. Thyroid nodules: A guide to assessment, treatment and follow-up[J]. *Maturitas*, 2017, 96: 1–9.
- [3] 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]. *Journal of the American College of Radiology*, 2017, 14(5): 587–595.
- [4] Hekimsoy İ, Üztürk E, Ertan Y, *et al.* Diagnostic performance rates of the ACR-TIRADS and EU-TIRADS based on histopathological evidence[J]. *Diagnostic and Interventional Radiology*, 2021, 27(4): 511–518.
- [5] Koc AM, Adıbelli ZH, Erkul Z, *et al.* Comparison of diagnostic accuracy of ACR-TIRADS, American Thyroid Association (ATA), and EU-TIRADS guidelines in detecting thyroid malignancy [J]. *European Journal of Radiology*, 2020, 133: 109390.
- [6] Lambin P, Rios-Velazquez E, Leijenaar R, *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis[J]. *European Journal of Cancer*, 2012, 48(4): 441–446.
- [7] Mukherjee S, Patra A, Khasawneh H, *et al.* Radiomics-based machine-learning models can detect pancreatic cancer on prediagnostic computed tomography scans at a substantial lead time before clinical diagnosis[J]. *Gastroenterology*, 2022, 163(5): 1435–1446. e3.
- [8] Cao Y, Zhong X, Diao W, *et al.* Radiomics in differentiated thyroid cancer and nodules: explorations, application, and limitations[J]. *Cancers*, 2021, 13(10): 2436.
- [9] Li R, Li L, Xu Y, *et al.* Machine learning meets omics: applications and perspectives[J]. *Briefings in Bioinformatics*, 2022, 23(1): 466–482.
- [10] Rago T, Vitti P. Risk stratification of thyroid nodules: from ultrasound features to TIRADS[J]. *Cancers*, 2022, 14(3): 717.
- [11] Modi L, Sun W, Shafizadeh N, *et al.* Does a higher American College of Radiology Thyroid Imaging Reporting and Data System (acr ti-rads) score forecast an increased risk of malignancy? a correlation study of acr ti-rads with fna cytology in the evaluation of thyroid nodules[J]. *Cancer Cytopathology*, 2020, 128(7): 470–481.
- [12] Hoang JK, Middleton WD, Farjat AE, *et al.* Interobserver variability of sonographic features used in the American college of radiology thyroid imaging reporting and data system[J]. *American Journal of Roentgenology*, 2018, 211(1): 162–167.
- [13] Persichetti A, Stasio ED, 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.
- [14] Thedinger W, Raman E, Dhingra JK. Comparative study of ACR TI-RADS and ATA 2015 for ultrasound risk stratification of thyroid nodules[J]. *Otolaryngology-Head and Neck Surgery*, 2022, 167(1): 35–40.
- [15] Yan Y, Zhang F, Ge H, *et al.* Effect of the size of benign thyroid degenerative nodules on ACR TI-RADS categories[J]. *Journal of Medical Ultrasonics*, 2022, 49(1): 71–76.
- [16] 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.
- [17] Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542(7639): 115–118.
- [18] Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging [J]. *European Journal of Radiology*, 2009, 70(2): 232–241.

(收稿日期: 2025-01-16

修回日期: 2025-08-22)