

# 基于CT影像组学模型鉴别诊断甲状腺良恶性结节的研究

王芳芳<sup>1</sup>, 秦文恒<sup>1</sup>, 李志<sup>1</sup>, 张丽娟<sup>2</sup>, 赵鲁平<sup>1</sup>, 毛森<sup>2</sup>, 侯曼曼<sup>2\*</sup>

<sup>1</sup>济宁医学院附属医院医学影像科, 山东 济宁

<sup>2</sup>济宁医学院附属医院超声医学科, 山东 济宁

收稿日期: 2026年1月13日; 录用日期: 2026年2月6日; 发布日期: 2026年2月25日

## 摘要

目的: 探讨CT影像组学模型对良恶性甲状腺结节的鉴别诊断价值。方法: 回顾性分析经手术病理学证实279例甲状腺结节患者, 其中甲状腺恶性结节133例, 良性结节146例。将患者按照8:2的比例随机分为训练集223例和测试集56例。采用联影URP组学软件基于术前平扫期、动脉期、静脉期CT图像提取影像组学特征, 后经过特征筛选, 得到影像组学评分(Radscore), 并建立影像组学随机森林、逻辑回归、支持向量机模型。采用ROC曲线下面积(Area under the curve, AUC)评估3种模型的诊断效能; 通过校准曲线和决策曲线(DCA)评估3种模型的预测效能和临床应用价值。结果: 随机森林模型训练集和测试集的曲线下面积(Area under the curve, AUC)分别为0.849 [95%可信区间(CI): 0.798~0.900]和0.788 (95% CI: 0.665~0.911), 灵敏度、特异度、准确率分别为85.8%、75.9%、79.6%和70.4%、73.9%、73.2%。逻辑回归模型训练集和测试集的AUC分别为0.817 (95% CI: 0.762~0.872)和0.803 (95% CI: 0.683~0.923), 灵敏度、特异度、准确率分别为77.8%、72.2%、75.0%和74.5%、72.2%、73.3%。支持向量机模型训练集和测试集的AUC分别为0.817 (95% CI: 0.760~0.873)和0.808 (95% CI: 0.700~0.920), 灵敏度、特异度、准确率分别为82.1%、73.9%、77.8%和77.8%、69.5%、71.4%。逻辑回归、随机森林、支持向量机三组模型的AUC两两比较的Delong检验结果分别为0.593 (随机森林 - 支持向量机)、0.751 (随机森林 - 逻辑回归)、0.831 (支持向量机 - 逻辑回归), 提示三组模型在诊断效能上表现相近; 校准曲线分析结果显示, 训练集(随机森林、逻辑回归、支持向量机) Brier Score分别为0.162、0.173、0.172; 测试集(随机森林、逻辑回归、支持向量机) Brier Score分别为0.186、0.183、0.179, 提示三组模型的预测效能均表现良好; 临床决策曲线分析提示三组模型的临床净收益显示良好。结论: CT影像组学鉴别甲状腺良恶性结节具有较高的诊断效能、较好的预测效能, 且本研究影像组学模型稳定性良好。

## 关键词

影像组学, 甲状腺结节, CT, 鉴别诊断

\*通讯作者。

# Research on the Differential Diagnosis of Benign and Malignant Thyroid Nodules Based on CT Radiomics Models

Fangfang Wang<sup>1</sup>, Wenheng Qin<sup>1</sup>, Zhi Li<sup>1</sup>, Lijuan Zhang<sup>2</sup>, Luping Zhao<sup>1</sup>, Sen Mao<sup>2</sup>, Manman Hou<sup>2\*</sup>

<sup>1</sup>Department of Medical Imaging, The Affiliated Hospital of Jining Medical University, Jining Shandong

<sup>2</sup>Department of Medical Ultrasound, The Affiliated Hospital of Jining Medical University, Jining Shandong

Received: January 13, 2026; accepted: February 6, 2026; published: February 25, 2026

## Abstract

**Objective:** To investigate the diagnostic value of a CT-based radiomics model in differentiating benign from malignant thyroid nodules. **Methods:** A retrospective analysis was conducted on 279 patients with thyroid nodules confirmed by surgical pathology, including 133 cases of malignant thyroid nodules and 146 cases of benign nodules. The patients were randomly divided into a training set (223 cases) and a test set (56 cases) in an 8:2 ratio. Using United Imaging's URP radiomics software, radiomic features were extracted from preoperative non-contrast, arterial phase, and venous phase CT images. After feature selection, a radiomics score (Radscore) was obtained, and three models—random forest, logistic regression, and support vector machine—were established based on radiomic features. The diagnostic performance of the three models was evaluated using the area under the receiver operating characteristic curve (AUC). The predictive performance and clinical application value of the three models were assessed through calibration curves and decision curve analysis (DCA). **Results:** The random forest model achieved area under the curve (AUC) values of 0.849 [95% confidence interval (CI): 0.798~0.900] and 0.788 (95% CI: 0.665~0.911) for the training and test sets, respectively. Its sensitivity, specificity, and accuracy were 85.8%, 75.9%, and 79.6% for the training set, and 70.4%, 73.9%, and 73.2% for the test set, respectively. The logistic regression model achieved AUC values of 0.817 (95% CI: 0.762~0.872) and 0.803 (95% CI: 0.683~0.923) for the training and test sets, respectively. Its sensitivity, specificity, and accuracy were 77.8%, 72.2%, and 75.0% for the training set, and 74.5%, 72.2%, and 73.3% for the test set, respectively. The support vector machine model achieved AUC values of 0.817 (95% CI: 0.760~0.873) and 0.808 (95% CI: 0.700~0.920) for the training and test sets, respectively. Its sensitivity, specificity, and accuracy were 82.1%, 73.9%, and 77.8% for the training set, and 77.8%, 69.5%, and 71.4% for the test set, respectively. Delong's test results for pairwise comparisons of the AUCs among the logistic regression, random forest, and support vector machine models were 0.593 (random forest vs. support vector machine), 0.751 (random forest vs. logistic regression), and 0.831 (support vector machine vs. logistic regression), indicating similar diagnostic performance among the three models. Calibration curve analysis revealed Brier scores of 0.162, 0.173, and 0.172 for the training set (random forest, logistic regression, and support vector machine, respectively), and 0.186, 0.183, and 0.179 for the test set (random forest, logistic regression, and support vector machine, respectively), suggesting good predictive performance for all three models. Decision curve analysis indicated favorable clinical net benefits for all three models. **Conclusion:** CT radiomics demonstrates high diagnostic efficacy and good predictive performance in differentiating benign and malignant thyroid nodules, and the radiomics models in this study exhibit excellent stability.

## Keywords

### Radiomics, Thyroid Nodules, CT, Differential Diagnosis

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,甲状腺结节的发病率不断升高,恶变率约为4%~5% [1],其中以乳头状癌最常见,约占85%~90% [2]。然而传统的影像学诊断方法在良恶性结节的鉴别中缺乏特异性,因此影像组学应运而生。荷兰学者 Lambin 等[3]首次提出,随后经历了飞速发展,目前它已普遍应用于 CT、MRI、超声以及病理学等医学领域[4] [5]。影像组学[6]主要通过获取图像、病灶分割、提取及筛选特征、构建模型等步骤实现了对疾病的诊断及疗效预测。本研究旨在较大样本量的基础上,评价基于 CT 的影像组学模型鉴别诊断甲状腺良恶性结节的研究价值。

## 2. 资料与方法

### 2.1. 一般资料

回顾性收集本院 2023 年 9 月~2025 年 6 月术前行甲状腺 CT 平扫及增强检查甲状腺结节患者,所有病例均经手术病理学证实。纳入标准:(1) 术前接受甲状腺 CT 平扫 + 增强检查;(2) 甲状腺结节直径均  $\geq 5$  mm;(3) 患者术前未接受过放化疗及穿刺。本研究经本院伦理委员会批准,所有患者签署知情同意书,允许其临床数据及影像用于研究。

### 2.2. 仪器与方法

采用 Siemens SOMATOM Definition Flash 螺旋 CT 机对 279 例患者均行甲状腺平扫 + 增强 CT 扫描检查。患者仰卧位,扫描范围口咽部至锁骨上缘。扫描参数:120 kV, 165~200 mAs, 层厚 1 mm。使用非离子碘对比剂(碘普罗胺, 300 mgI/ml, Bayer 公司)。

### 2.3. 影像组学模型的构建

由 2 名具有 6 年以上甲状腺 CT 诊断经验的放射科主治医师采用联影 URP 组学软件在 CT 图像上分别逐层手动勾画 ROI (图 1、图 2), 然后自动提取影像组学特征, 通过组内相关系数(intraclass correlation coefficient, ICC)检验保留  $ICC \geq 0.75$  的特征, 依次通过最大相关最小冗余(max-relevance and min-redundancy, mRMR)及最小绝对收缩与选择算子(least absolute shrinkage and selection operator, LASSO)算法筛选影像组学特征。行 LASSO 回归分析时, 通过 10 折交叉验证进行特征降维, 得到最优超参数  $\lambda$  值, 影像组学评分(Radscore)。最终基于随机森林、逻辑回归、支持向量机构建影像组学模型。

### 2.4. 统计学分析

采用 SPSS 26.0 进行统计分析。计量资料以  $\bar{x} \pm s$  或 M(Q1, Q3)表示, 组间比较采用 t 或 Mann-Whitney U 检验; 计数资料组间比较采用四格表  $\chi^2$  检验, 以  $P < 0.05$  为差异有统计学意义。采用联影 URP 组学软件生成随机森林、逻辑回归、支持向量机影像组学模型的受试者操作特征曲线(receiver operator

characteristic curve, ROC)、校准曲线和决策分析曲线(decision curve analysis, DCA)。通过曲线下面积(Area under the curve, AUC)评估各模型的诊断效能, 3 种模型间的 AUC 比较采用 DeLong 检验。通过校准曲线并计算 Brier 分数评估各模型的预测效能; 通过决策曲线(DCA)评估各模型的临床应用价值。

### 3. 结果

#### 3.1. 一般资料

共收集 279 例甲状腺结节患者, 其中甲状腺良性结节 146 例, 男 21 例, 女 125 例, 年龄 14~77 岁 ( $50.66 \pm 12.27$ ); 恶性结节 133 例, 男 29 例, 女 104 例, 年龄 13~82 ( $41.84 \pm 12.87$ )。经分层随机抽样分组后, 训练集 223 例, 男 40 例、女 183 例, 年龄 13~76 岁 ( $47.04 \pm 12.53$ ), 恶性 117 例、良性 106 例; 测试集 56 例, 15~82 岁 ( $44.16 \pm 15.43$ ), 男 12 例、女 44 例, 恶性 27 例、良性 29 例。两组患者性别( $\chi^2 = 0.360, P = 0.549$ )、年龄( $t = 1.480, P = 0.140$ )、良恶性病变比例( $\chi^2 = 0.008, P = 0.927$ )差异均无统计学意义。

#### 3.2. 影像组学特征及模型效能、影像组学模型比较

首先通过 ICC 检验保留  $ICC \geq 0.75$  的 1471 个特征值, 依次采用 Correlation Coefficient、mRMR 及 LASSO 算法筛选特征, 利用 10 折交叉验证选择最优模型参数  $\lambda$  ( $\lambda = 0.00599, \log \lambda = 2.222$ ), 根据公式计算出每例患者的 Radscore, 最终保留 9 个特征(图 3)并建立随机森林、逻辑回归、支持向量机影像组学模型。绘制 ROC 曲线分析影像组学模型的预测效能, 随机森林模型训练集和测试集的曲线下面积(Area under the curve, AUC)分别为 0.849 [95% 可信区间(CI): 0.798~0.900]和 0.788 (95% CI: 0.665~0.911), 灵敏度、特异度、准确率分别为 85.8%、75.9%、79.6%和 70.4%、73.9%、73.2%。逻辑回归模型训练集和测试集的 AUC 分别为 0.817 (95% CI: 0.762~0.872)和 0.803 (95% CI: 0.683~0.923), 灵敏度、特异度、准确率分别为 77.8%、72.2%、75.0%和 74.5%、72.2%、73.3%。支持向量机模型训练集和测试集的 AUC 分别为 0.817 (95% CI: 0.760~0.873)和 0.808 (95% CI: 0.700~0.920), 灵敏度、特异度、准确率分别为 82.1%、73.9%、77.8%和 77.8%、69.5%、71.4% (表 1)。随机森林、逻辑回归、支持向量机 3 种模型两两比较的 AUC 差异无统计学意义( $P > 0.05$ ) (表 2, 图 4)。

**Table 1.** Diagnostic performance of three radiomic models

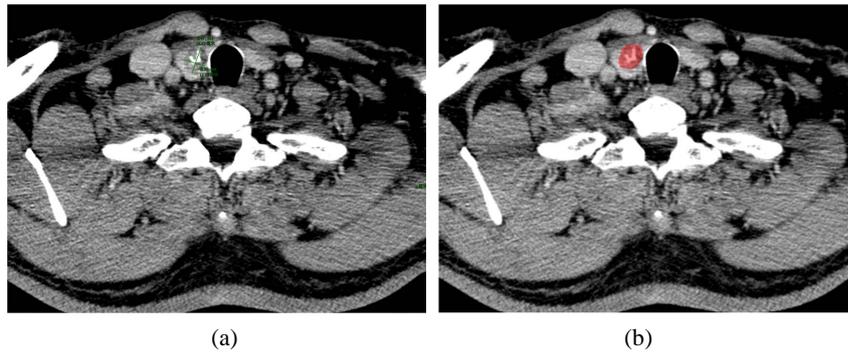
**表 1.** 3 种影像组学模型的诊断效能

模型	训练集				测试集			
	AUC (95% CI)	敏感度	特异度	准确率	AUC (95% CI)	敏感度	特异度	准确率
随机森林模型	0.849 (95% CI: 0.798~0.900)	85.8%	75.9%	79.6%	0.788 (95% CI: 0.665~0.911)	70.4%	73.9%	73.2%
逻辑回归模型	0.817 (95% CI: 0.762~0.872)	77.8%	72.2%	75.0%	0.803 (95% CI: 0.683~0.923)	74.5%	72.2%	73.3%
支持向量机模型	0.817 (95% CI: 0.760~0.873)	82.1%	73.9%	77.8%	0.808 (95% CI: 0.700~0.920)	77.8%	69.5%	71.4%

**Table 2.** Results of Delong and McNemar tests (P-values) for the three radiomic models

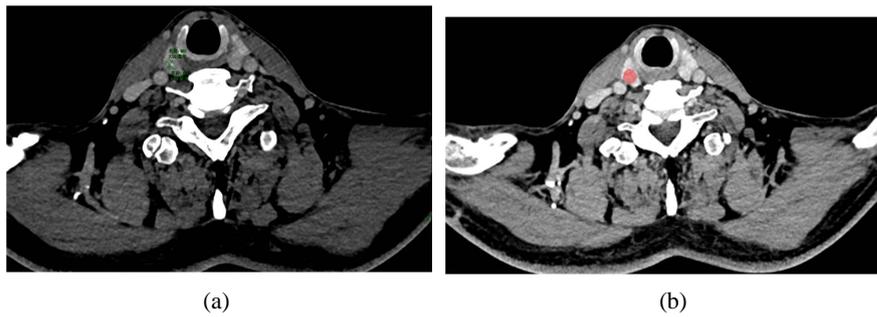
**表 2.** 3 种影像组学模型的 Delong 和 McNemar 检验结果(P 值)

模型比较	AUC	敏感性	特异性
随机森林 - 支持向量机	0.593	0.688	0.250
随机森林 - 逻辑回归	0.751	0.727	1.000
支持向量机 - 逻辑回归	0.831	1.000	0.500



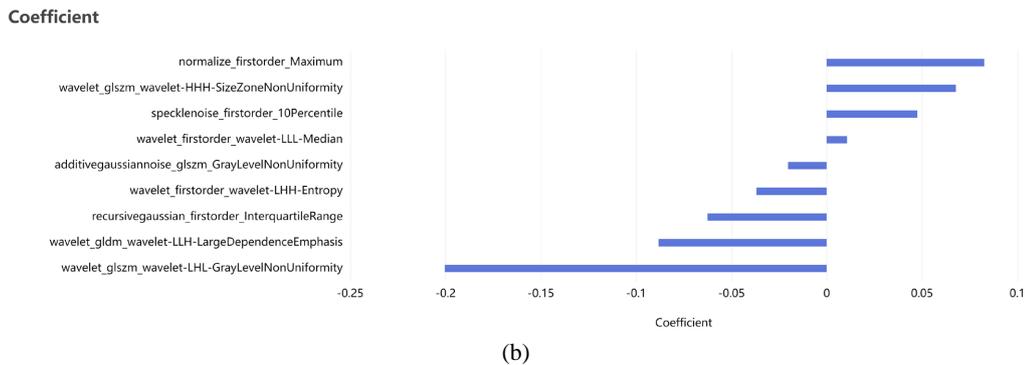
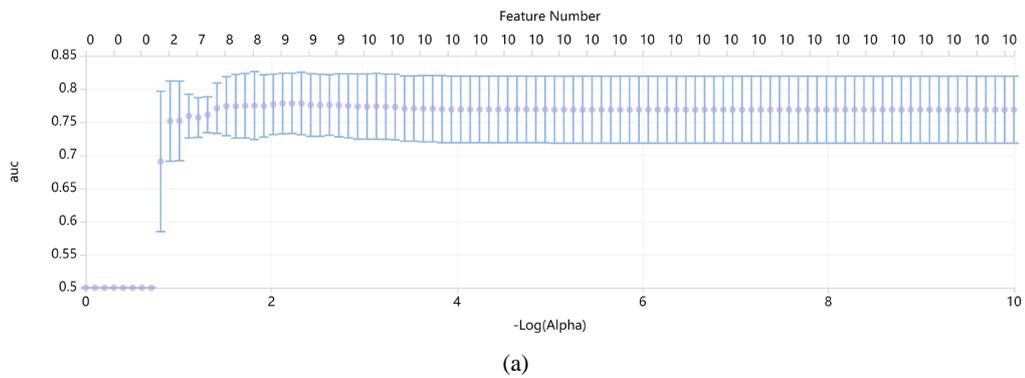
注：甲状腺增强静脉期(a)及相应 ROI 区域勾画(b)。

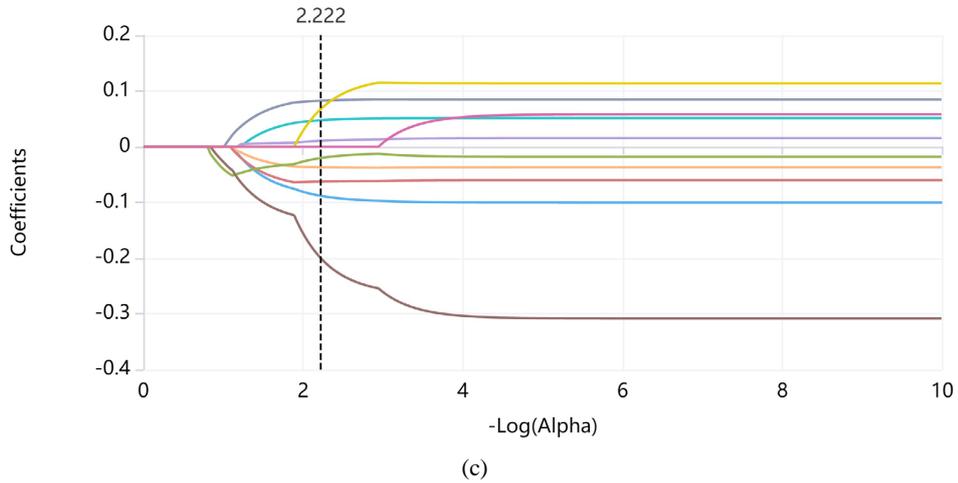
**Figure 1.** CT imaging of a 54-year-old male patient with thyroid  
**图 1.** 男性 54 岁甲状腺 CA 患者 CT 影像



注：甲状腺增强静脉期(a)及相应 ROI 区域勾画(b)。

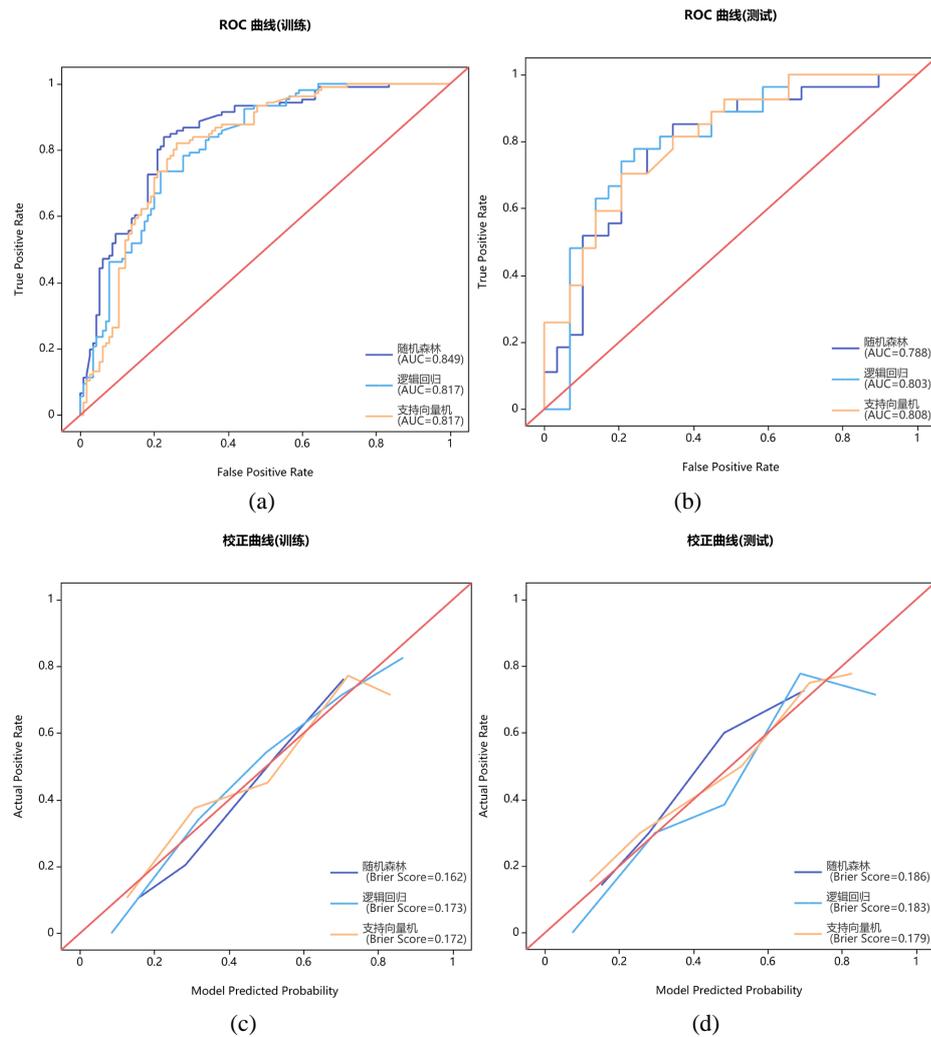
**Figure 2.** CT imaging of a 63-year-old male patient with thyroid follicular epithelial hyperplastic lesion  
**图 2.** 男性 63 岁甲状腺滤泡上皮增生性病变患者 CT 影像

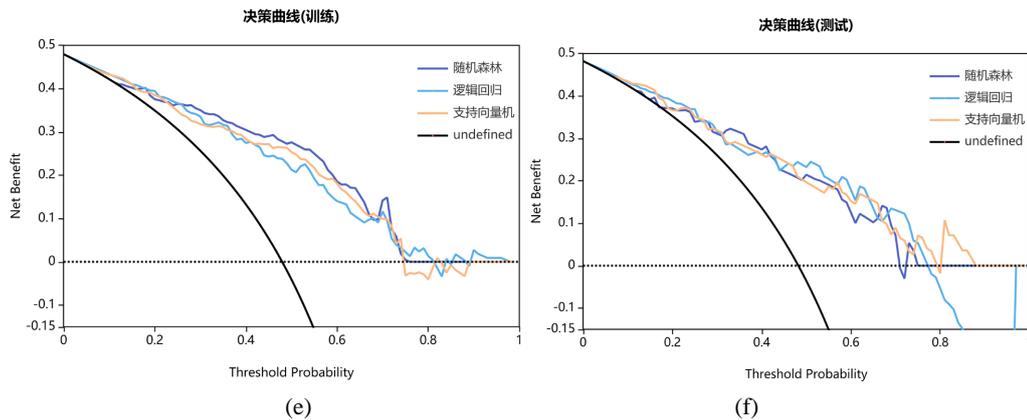




注：(a)为 LASSO 算法交叉验证曲线图；(b)为 9 个最终保留的特征值及相应回归系数；(c)为 LASSO 回归系数分布曲线图。

**Figure 3.** Feature selection procedure Using the least absolute shrinkage and selection operator (LASSO) algorithm  
**图 3.** 最小绝对收缩与选择算子(LASSO)算法特征筛选过程





(a)、(b) 随机森林、逻辑回归、支持向量机 3 种模型在训练集和测试集的 ROC 曲线。(c)、(d) 3 种模型在训练集及测试集的校正曲线。(e)、(f) 3 种模型在训练集和测试集的决策分析(DCA)曲线。

**Figure 4.** ROC curves, calibration curves, and decision curve analysis (DCA) curves for the three models: random forest, logistic regression, and support vector machine

**图 4.** 随机森林、逻辑回归、支持向量机 3 种模型的 ROC 曲线、校正曲线、决策分析(DCA)曲线

## 4. 讨论

随着居民对自身健康水平关注度及就医意愿的显著提升，甲状腺结节的检出率也逐年升高。在临床诊疗中，恶性结节通常需要手术治疗，而良性结节绝大多数仅需定期观察。由于近几年良性结节过度诊断的问题日益突出[7]，给患者带来很多不良影响，因此甲状腺结节良恶性的术前准确评估显得尤其重要。

传统的影像学诊断方法已不能完全满足目前医疗诊断的需求，因此影像组学应运而生，它[8]是从传统的影像中提取高阶特征来量化肿瘤病灶，并通过深度学习的方法挖掘与病理有潜在关联的特征，用于疾病的诊断及预后评估。陈鹏等[9]研究指出增强 CT 影像组学能够在术前有效鉴别 TI-RADS 4A 和 4B 类甲状腺微小结节的良恶性。董林娟等[10]的研究通过构建 CT 平扫、动脉期、静脉期影像组学模型评估其在甲状腺乳头状癌和结节性甲状腺肿术前鉴别诊断的价值，结果显示具有较高的鉴别诊断效能。XU 等[11]研究基于影像组学特征构建的逻辑回归模型实现了很高的诊断精度，训练组和测试组的 AUC 分别达 94.4% 和 94.2%，表明基于甲状腺 CT 的影像组学在区分良恶性甲状腺结节方面的诊断价值。以上基于 CT 影像组学的甲状腺病变研究结果均对本研究结果提供了支持。而赵泓博等[12]通过构建随机森林、逻辑回归和 SVM 机器学习模型进行分类训练，得出随机森林的 ROC 曲线下面积、准确度、灵敏度及特异度都为最佳。而本研究通过构建影像组学随机森林、逻辑回归、支持向量机三组模型，得出影像组学对甲状腺良恶性结节的鉴别诊断具有较高的预测效能。Delong 检验常用于比较不同模型间的受试者工作特征曲线(ROC)下面积(AUC)差异。本研究逻辑回归、随机森林、支持向量机三组模型的 Delong 检验结果提示提示三组模型在诊断效能上表现相近，虽然与赵泓博等研究结果有差异，却证明了本研究模型稳定性良好，可重复性强。

本研究最终保留 9 个特征，4 个均为不同滤波器变换得来的高阶统计量，5 个基于灰度直方图的一阶特征。直方图特征是图像灰度级的函数，反映像素值的分布。孔丹等[9]的研究中显示直方图特征包括相对偏度及区域百分比，能一定程度反映肿瘤内部体素分布的均匀性及强度变化，本研究结果一阶特征值包括百分位数、中位数、四分位数、最大值等，能一定程度反映肿瘤内部异质性，局部极端值(钙化、出血)，坏死/低活性区域等，分析原因可能是 ROI 包含病灶整体，未避开结节内囊变坏死及钙化，而囊变坏死及钙化包含病灶一些固有的纹理特征，与孔丹等研究结果相似。吴宇强等[13]、郭伟等[14]研究指出熵值可用于鉴别甲状腺良恶性结节，本研究结果得出的特征值也包括熵值，与其研究结果相符，熵值常与

肿瘤异质性和侵袭性相关,是预测恶性肿瘤非常有价值的特征。然而相较于形态特征、直方图特征等一、二阶统计量特征而言,高阶统计量特征能更全面地体现出原始图像的异质性,从而获得更多有价值的信息[15]。董林娟等[10]研究结果指出最终筛选的特征均是基于小波滤波的高阶特征,表明小波特征在甲状腺良恶性结节鉴别诊断中具有较大的潜力。本研究结果的9个特征值内有6个是基于小波滤波的特征值,4个高阶,2个一阶;其中4个高阶统计量中3个是基于小波滤波的高阶特征,与其研究结果相似,本研究所得小波特征既可直接反映了恶性结节中砂粒体空间分布的异质性;还可通过表明细微结构区域的大小差异,空间分布均匀性,提示肿瘤内部微环境复杂情况。部分特征值的变化(可能高也可能低)还能反映肿瘤内部均质性与异质性的平衡,提示组织结构的分化程度。这些特征值均在甲状腺良恶性结节的鉴别诊断中发挥重要作用。

本研究的局限性:本研究为回顾性研究,可能存在潜在的选择偏倚;本研究样本量较小且为单中心研究,缺乏外部验证,未选取临床因素,也未与常规形态学方法的诊断效率进行比较;未来研究可通过扩大样本量及扩展到多中心、引入深度学习或优化特征工程策略,进一步探索模型性能的潜在差异。同时仍需进一步完善临床资料,纳入临床因素及影像学形态特征。

综上所述,CT影像组学模型对于鉴别甲状腺良恶性结节具有较好的效能,能够较为准确地预测甲状腺恶性结节的概率,以辅助临床诊疗决策。

## 利益冲突

所有作者均声明不存在利益冲突。

## 基金项目

山东省济宁市科技局2023年济宁市重点研发计划项目(编号:2023YXNS089);济宁医学院附属医院苗圃科研项目(MP-MS-2023-04)。

## 参考文献

- [1] 王芳芳,徐景景,赵鲁平,等.基于CT影像特征的列线图对甲状腺良恶性结节鉴别诊断的价值[J].临床放射学杂志,2023,42(10):1561-1566.
- [2] Khatami, F. and Tavangar, S.M. (2018) A Review of Driver Genetic Alterations in Thyroid Cancers. *Iranian Journal of Pathology*, **13**, 125-135. <https://doi.org/10.30699/ijp.13.2.125>
- [3] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R.G.P.M., Granton, P., et al. (2012) Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. *European Journal of Cancer*, **48**, 441-446. <https://doi.org/10.1016/j.ejca.2011.11.036>
- [4] 曹娜,黄渤琪,刘志民,等.基于CT影像组学构建富血供超小肾癌与乏脂肪血管平滑肌脂肪瘤鉴别诊断的简易预测模型[J].实用临床医药杂志,2023,27(10):6-11.
- [5] 侯慧,朱银杏,王太宇,等.基于双参数MRI影像组学构建的支持向量机模型对乳腺癌人表皮生长因子受体-2和激素受体表达的预测效能[J].实用临床医药杂志,2024,28(4):7-13.
- [6] 韦启君,林盛才,黄福灵,等.双能CT影像组学在甲状腺结节定性诊断中的价值研究[J].中国医疗设备,2022,37(12):87-90.
- [7] 王玉春,杨斌,黄鹏飞,等.ACR TI-RADS与基于人工智能的报告系统对甲状腺结节的诊断效能及减少不必要穿刺能力的比较[J].中华超声影像学杂志,2021,30(5):408-413.
- [8] 孔丹,张建东,单文莉,等.CT影像组学模型对甲状腺良恶性结节的鉴别价值[J].中华放射学杂志,2020,54(3):187-191.
- [9] 陈鹏,张成孟,周艳艳,等.增强CT影像组学在鉴别TI-RADS 4A、4B类甲状腺微小结节良恶性的价值[J].中国CT和MRI杂志,2024,22(7):39-42.
- [10] 董林娟,资欣月,李庆文,等.基于CT影像组学列线图鉴别甲状腺乳头状癌与结节性甲状腺肿的价值[J].临床

- 放射学杂志, 2023, 42(9): 1409-1416.
- [11] Xu, H., Wang, X., Guan, C., Tan, R., Yang, Q., Zhang, Q., *et al.* (2022) Value of Whole-Thyroid CT-Based Radiomics in Predicting Benign and Malignant Thyroid Nodules. *Frontiers in Oncology*, **12**, Article ID: 828259. <https://doi.org/10.3389/fonc.2022.828259>
- [12] 赵泓博, 叶靖, 段绍峰, 等. CT 影像组学对甲状腺乳头状癌与腺瘤的鉴别[J]. 中国医学计算机成像杂志, 2021, 27(2): 102-106.
- [13] 吴宇强, 靳激扬, 冯银波. 甲状腺意外结节增强CT的纹理特征分析对良恶性鉴别的价值研究[J]. 东南大学学报: 医学版, 2016, 35(1): 112-116.
- [14] 郭炜, 罗德红, 赵燕风, 等. CT 纹理分析技术鉴别甲状腺良恶性结节可行性研究[J]. 国际医学放射学杂志, 2017, 40(1): 3-5+36.
- [15] 秦文恒, 刘慧君, 于昊, 等. 基于 CT 增强扫描静脉期的影像组学列线图预测晚期肺癌患者对培美曲塞+铂类化疗敏感性的价值[J]. 中国中西医结合影像学杂志, 2023, 21(5): 481-486.