

• 临床研究 •

# 基于光谱CT和影像组学特征的机器学习模型和列线图预测甲状腺结节良恶性

王皓<sup>1,2</sup>, 杜昱升<sup>2</sup>, 王友<sup>2</sup>, 白卓杰<sup>2</sup>, 杨帆<sup>2</sup>, 何健<sup>1,3\*</sup><sup>1</sup>南京医科大学鼓楼临床医学院, 江苏 南京 210008; <sup>2</sup>南京医科大学第四附属医院放射科, 江苏 南京 210031; <sup>3</sup>南京大学医学院附属鼓楼医院核医学科, 江苏 南京 210008

**[摘要]** 目的: 构建光谱CT参数和影像组学机器学习模型预测甲状腺结节良恶性。方法: 回顾性分析行光谱CT增强扫描的118例甲状腺结节患者(143个结节, 其中包括46例良性结节和97例恶性结节)影像及临床资料, 7:3随机分为训练集( $n=100$ )和验证集( $n=43$ )。采用差异性检验、组间一致性评估以及最小绝对收缩和选择算子算法(least absolute shrinkage and selection operator, LASSO)筛选特征并计算影像组学评分。运用决策树(decision tree, DT)、随机森林(random forest, RF)、极端梯度提升树(extreme gradient boosting, XGBoost)、支持向量机(support vector machine, SVM)、K最近邻(K-nearest neighbor, KNN)和逻辑回归(logistic regression, LR)6种机器学习算法进行建模, 筛选最佳的模型并构建列线图。结果: XGBoost模型在验证集中性能最好(曲线下面积: 0.938; 准确度: 86.05%; 灵敏度: 89.29%; 特异度: 80.00%), 标准化碘值、影像组学评分与年龄是重要且有效的预测因素, 构建的列线图具有良好的性能。结论: 结合光谱CT和影像组学的机器学习模型及列线图能够为甲状腺结节良恶性的非侵入性预测提供高准确性的参考。

**[关键词]** 影像组学; 机器学习; 光谱CT; 甲状腺结节**[中图分类号]** R814.42**[文献标志码]** A**[文章编号]** 1007-4368(2024)07-958-08**doi:** 10.7655/NYDXBNSN240049

## Machine learning models and nomograms based on spectral CT and radiomic features for predicting the benignancy and malignancy of thyroid nodules

WANG Hao<sup>1,2</sup>, DU Yusheng<sup>2</sup>, WANG You<sup>2</sup>, BAI Zhuojie<sup>2</sup>, YANG Fan<sup>2</sup>, HE Jian<sup>1,3\*</sup><sup>1</sup>Drum Tower Clinical Medical College of Nanjing Medical University, Nanjing 210008; <sup>2</sup>Department of Radiology, the Fourth Affiliated Hospital of Nanjing Medical University, Nanjing 210031; <sup>3</sup>Department of Nuclear Medicine, Gulou Hospital Affiliated to Nanjing University School of Medicine, Nanjing 210008, China

**[Abstract]** **Objective:** To establish the applicability of a predictive model based on spectral computed tomography (CT) parameters and radiomics features through machine learning for differentiating benign and malignant thyroid nodules. **Methods:** A retrospective analysis was conducted on imaging and clinical data from 118 patients with thyroid nodules who underwent spectral CT enhancement scanning (143 nodules, comprise of 46 benign and 97 malignant nodules). These patients were randomly divided into a training set ( $n=100$ ) and a validation set ( $n=43$ ) in a 7:3 ratio. Discriminative testing, intraclass correlation coefficient (ICC), and the least absolute shrinkage and selection operator (LASSO) were employed to select features and calculate a radiomics score (Radscore). Six machine learning algorithms including decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), K-nearest neighbors (KNN), and logistic regression (LR) were utilized to develop models. The optimal model was selected to construct nomograms. **Results:** The XGBoost model was demonstrated to be the best in the validation set (AUC: 0.938; accuracy: 86.05%; sensitivity: 89.29%; specificity: 80.00%), with normalized iodine concentration (NIC), Radscore, and age identified as significant predictive factors. The ensuing nomograms exhibited robust performance. **Conclusion:** The machine learning model that combines spectral CT and radiomics features with the nomograms provides a highly accurate reference for non-invasive prediction of

**[基金项目]** 江苏省青年医学重点人才(QNRC2016041); 南京医科大学科技发展基金一般项目(NMUB20230037)

\*通信作者(Corresponding author), E-mail: hjxuere@126.com

the benignity or malignancy of thyroid nodules.

[Key words] radiomics; machine learning; spectral CT; thyroid nodule

[J Nanjing Med Univ, 2024, 44(07): 958-965]

甲状腺结节是临床上常见的内分泌肿瘤,近年来其患病率呈上升趋势<sup>[1]</sup>。恶性甲状腺结节以乳头状癌和滤泡性癌为主要病理类型,其部分亚型具有较高的颈部淋巴结转移风险<sup>[2]</sup>。早期准确诊断甲状腺结节良恶性具有重要的临床意义,不仅能够减少非必要的手术及其并发症,而且对提高患者存活率及减轻心理压力有着至关重要的作用<sup>[3]</sup>。目前,多普勒超声及其引导下细针穿刺活检被广泛应用于甲状腺结节良恶性评估<sup>[4]</sup>。然而,该方法受限于检查者的主观判断,且对微小结节和混合性结节的诊断具有限制,其准确率为60%~70%<sup>[5]</sup>。传统CT在评估甲状腺结节大小、定位和淋巴结方面具有一定价值,但由于其主要依赖形态学特征并且缺乏足够的定量分析参数,在区分甲状腺结节良恶性方面存在局限性<sup>[6]</sup>。

能谱CT技术的进步扩展了CT在临床诊断中的作用,其中双层探测器光谱CT(dual detector spectral CT, DSCT)较传统能谱CT能提供更加丰富的诊断信息。现有研究表明,DSCT碘密度图反映的碘摄取量对甲状腺结节良恶性预测具有很高的价值<sup>[7]</sup>。同时,影像组学作为一门新兴学科,通过定量提取医学影像中的丰富特征<sup>[8-9]</sup>,为捕捉肿瘤的微妙变化开辟了一条新途径,而这些细节通常在常规的影像解读中可能被遗漏<sup>[10]</sup>。机器学习(machine learning, ML)作为人工智能领域的一个重要分支,在处理和析医学图像数据方面展示了其卓越能力,在诊断预测方面得到充分验证<sup>[11]</sup>。然而,DSCT、影像组学和ML三者的结合效应及其在临床决策中的实际应用价值还有待深入研究探讨<sup>[12]</sup>。本研究拟结合DSCT与影像组学,采用6种常见的ML方法建立模型来预测甲状腺结节的良恶性,并进一步构建列线图,为临床提供参考。

## 1 对象和方法

### 1.1 对象

回顾性分析2021年1月—2023年9月在南京医科大学第四附属医院经手术切除和术后病理验证为恶性(甲状腺乳头状癌、滤泡性癌)或良性(结节

性甲状腺肿、甲状腺腺瘤)的甲状腺结节患者,且于术前1周内行颈部DSCT增强扫描。排除标准:①结节呈广泛囊性病灶或大面积钙化病灶;②最大层面横径>15 mm;③病理及实验室资料不足;④年龄<18岁;⑤合并其他恶性肿瘤病史;⑥图像伪影明显,无法进行分析。最终共纳入患者118例(男22例,女96例),涵盖46个良性结节与97个恶性结节,共143个结节(25例患者存在多枚甲状腺结节)。本研究通过了南京医科大学第四附属医院伦理委员会批准(20240110-k002)。

### 1.2 方法

#### 1.2.1 扫描参数

使用DSCT(飞利浦公司,荷兰)进行颈部扫描。扫描参数:管电压120 kVp,通过Dose Right技术(index: 23)自动调整管电流(平均145 mAs),准直器宽度为64×0.625 mm,视野300 mm,螺距0.969,图像矩阵512×512,图像层厚层间距均1 mm,窗宽350 Hu,窗位60 Hu。使用双腔高压注射器注入非离子型造影剂65 mL,速率2.5 mL/s,随后以相同速率注入15 mL生理盐水。动脉期和静脉期分别在注射后25 s和60 s进行扫描,并保存光谱数据。

#### 1.2.2 光谱参数与临床因素

为了降低动脉强化的影响,提高甲状腺结节内碘摄取的可辨识度<sup>[13]</sup>,本研究使用的均为静脉期图像。采用飞利浦工作站进行光谱数据分析,选取传统图像、碘密度图和虚拟单能量图像,通过对甲状腺结节最大层面勾画感兴趣区(region of interest, ROI),覆盖横截面区域不少于结节面积的2/3,同时避开囊变与钙化,在同层面的颈静脉处也勾画ROI,测量并计算结节组织的光谱参数:标准化碘值(normalized iodine concentration, NIC)与能谱斜率(slope of spectral curves,  $\lambda_{HI}$ ),计算公式如下:  $NIC = \frac{\text{甲状腺碘值} / \text{颈总静脉碘值}}{\text{CT值}_{40 \text{ keV}} - \text{CT值}_{100 \text{ keV}}} / (100 - 40)$ <sup>[14]</sup>。通过医院信息系统检索所有患者手术前48 h内的游离三碘甲状腺原氨酸(free triiodothyronine, FT3)、游离甲状腺素(free thyroxine, FT4)、促甲状腺激素(thyroid stimulating hormone, TSH)数据。

### 1.2.3 ROI分割和特征提取

通过东华医疗的图像存储与传输系统获取所有患者的影像数据,采用3D slicer软件(版本5.2.2)对甲状腺结节影像进行逐层分割,由1位具有15年经验的放射科医师负责完成。为评估勾画的一致性,另1位拥有25年放射治疗经验的医师随机对20个良性和20个恶性病灶进行重复勾画,通过组内相关系数(interclass correlation coefficient, ICC)评估确保勾画结果的可重复性。利用pyradiomics插件执行影像组学特征提取,同时应用低通高斯滤波器对肿瘤纹理特征进行量化表征(过滤内核尺寸:3、4),以反映多尺度下的变化。共计从影像中提取出293个特征,包含形状特征(original\_shape)14个,一阶直方图特征(original\_firstorder)54个,共生矩阵特征(gray level cooccurrence matrix, GLCM)72个,灰度依赖矩阵(gray level dependence matrix, GLDM)42个,灰度游程长度矩阵(gray level run length matrix, GLRLM)48个,灰度级大小区域矩阵(gray level size zone matrix, GLSZM)48个,以及邻域灰度差分矩阵(neighborhood gray-tone difference matrix, NGTDM)15个。

### 1.2.4 影像组学特征筛选

将结节以7:3随机拆分训练集与验证集,其中训练集100个(良性:恶性=31:69),验证集43个(良性:恶性=15:28),鉴于训练集中恶性结节较多,采用合成少数类过采样技术(synthetic minority over-sampling technique, SMOTE)以倍率1.816对训练集良性结节进行过采样,以达到良恶性比例1:1(良性:恶性=69:69)。对提取后的特征数据进行Z-score标准化处理以降低特征之间数量级的差异,计算公式: $Z\text{-score}=(X-\mu)/\sigma$ ,其中, $X$ 为特征值, $\mu$ 为平均值, $\sigma$ 为标准差。通过独立样本 $t$ 检验与Mann-Whitney  $U$ 检验移除 $P \geq 0.05$ 的特征,并移除ICC值 $\leq 0.8$ 的特征,使用最小绝对收缩和选择算子算法(least absolute shrinkage and selection operator, LASSO)筛选特征,采用10折交叉相互验证选择惩罚系数 $\lambda$ 的最佳值及剩余非0的特征。采用以下公式计算每位患者的影像组学评分(radiomics score, Radscore), $\text{Radscore}=\sum(\text{特征数值} \times \text{特征权重})+b0$ (截断值)。

### 1.2.5 ML模型

6种ML分类算法:决策树(decision tree, DT)、随机森林(random forest, RF)、极端梯度提升树(extreme gradient boosting, XGBoost)、支持向量机

(support vector machine, SVM)、K最近邻(K-nearest neighbor, KNN)和逻辑回归(logistic regression, LR),在训练集中对年龄、性别、NIC、 $\lambda_{\text{th}}$ 、FT3、FT4、TSH和Radscore进行建模,采用5倍交叉验证法检验模型以降低过拟合,通过比较验证集中受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under curve, AUC)以选择最优模型。通过SHAP特征权重分布图筛选变量构建列线图以直观展示结果,并通过绘制校准曲线和决策曲线图(decision curve analysis, DCA)评估列线图模型的诊断性能。

### 1.3 统计学方法

采用R软件包(版本4.2.1)和SPSS25.0完成数据分析。符合正态分布的数据以均数 $\pm$ 标准差( $\bar{x} \pm s$ )表示并以独立样本 $t$ 检验进行比较,非正态分布数据则用中位数(四分位数)[ $M(P_{25}, P_{75})$ ]表示并以Mann-Whitney  $U$ 检验进行比较,对分类资料以频数(百分比)表示并采用卡方检验进行比较。评价所有ML模型的AUC、准确度(accuracy, ACC)、灵敏度(sensitivity, SEN)、特异度(specificity, SPE)、阳性预测值(positive predictive value, PPV)、阴性预测值(negative predictive value, NPV)、精确度(precision, PRE)、召回率(recall, REC)和F1分数(F1),通过Delong检验比较各AUC之间的差异。采用R语言中的SHAP包来解释各个特征的重要性。 $P < 0.05$ 为差异有统计学意义。

## 2 结果

### 2.1 临床资料比较

表1概述了训练集( $n=100$ )与验证集( $n=43$ )的患者基础资料。两组间在病变比例、性别、年龄、NIC、 $\lambda_{\text{th}}$ 、FT3、FT4及TSH等因素上差异均无统计学意义( $P > 0.05$ )。

### 2.2 特征筛选与影像组学评分建立

经ICC( $>0.8$ )、独立样本 $t$ 检验与Mann-Whitney  $U$ 检验( $P < 0.05$ )共筛选出特征160个。通过LASSO回归及10倍交叉验证,发现6个与预测良恶性相关的特征(图1A~C),其中包含3个一阶特征和3个纹理特征。通过公式计算Radscore,结果显示恶性组[0.574(-0.015, 1.285)]明显高于良性组[-0.580(-1.457, 0.111)],差异具有统计学意义( $Z=-4.997, P < 0.001$ ,图1D)。

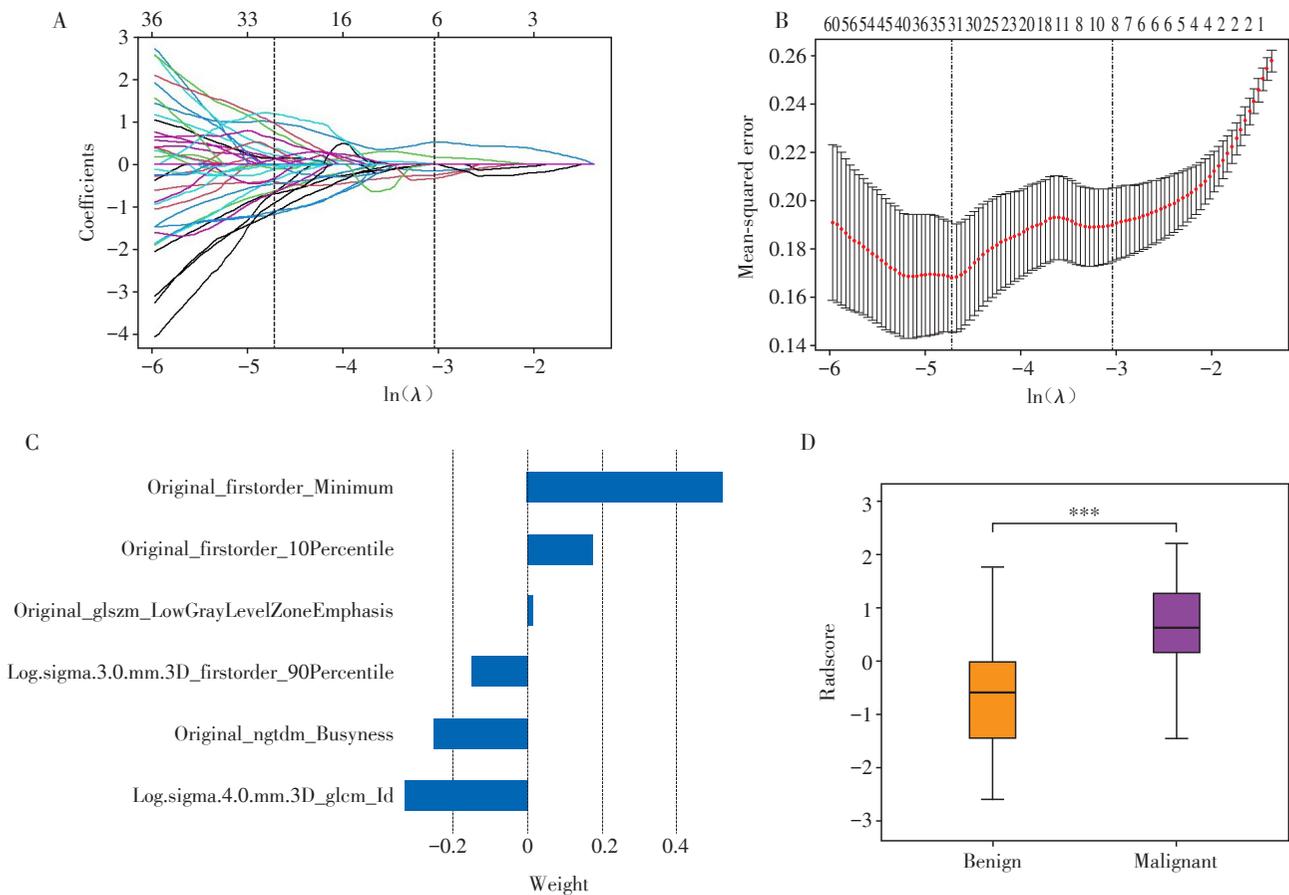
### 2.3 ML模型建立

采用DT、RF、XGBoost、SVM、KNN和LR 6种ML

表1 训练集与验证集的基本信息

Table 1 Basic information of the training and validation sets

Information	Training set( <i>n</i> =100)	Validation set( <i>n</i> =43)	Total( <i>n</i> =143)	$\chi^2/Z$	<i>P</i>
Clusters[ <i>n</i> (%)]				0.794	0.068
Benign	31(31)	15(35)	46(32)		
Malignant	69(69)	28(65)	97(68)		
Sex[ <i>n</i> (%)]				-0.104	0.747
Female	83(83)	34(79)	117(82)		
Male	17(17)	9(21)	26(18)		
Age(years, $\bar{x} \pm s$ )	53.00 $\pm$ 13.73	49.14 $\pm$ 16.63	51.84 $\pm$ 14.71	-1.339	0.185
NIC( $\bar{x} \pm s$ )	0.56 $\pm$ 0.19	0.57 $\pm$ 0.20	0.56 $\pm$ 0.19	0.383	0.705
$\lambda_{fit}[M(P_{25}, P_{75})]$	2.83(2.09, 3.64)	2.85(2.21, 3.33)	2.85(2.12, 3.58)	-0.288	0.773
FT3(pmol/L, $\bar{x} \pm s$ )	4.68 $\pm$ 0.69	4.64 $\pm$ 0.60	4.67 $\pm$ 0.66	-0.325	0.746
FT4[pmol/L, $M(P_{25}, P_{75})]$	16.45(14.17, 18.80)	15.15(14.00, 19.30)	16.30(14.10, 18.90)	-0.326	0.745
TSH[mU/L, $M(P_{25}, P_{75})]$	1.72(1.33, 2.78)	1.53(1.21, 2.64)	1.65(1.27, 2.74)	-0.564	0.573



A: LASSO feature selection and tuning, where the vertical dashed line indicated the optimal penalty coefficient  $\lambda$  corresponding to the non-zero features. B: The AUC curve plotted through 10-fold cross-validation, with the dashed lines on the left and right representing  $\lambda_{min}$  and  $\lambda_{1se}$  respectively,  $\lambda_{1se}$  was selected for this study. C: The 6 features retained after LASSO filtering and their respective weight coefficients. D: Comparison of the Radscore obtained by calculation in the training set, where the score for the malignant group was higher than that for the benign group, with a statistical significance (\*\* $P < 0.001$ ).

图1 LASSO与影像组学评分

Figure 1 LASSO and Radscore

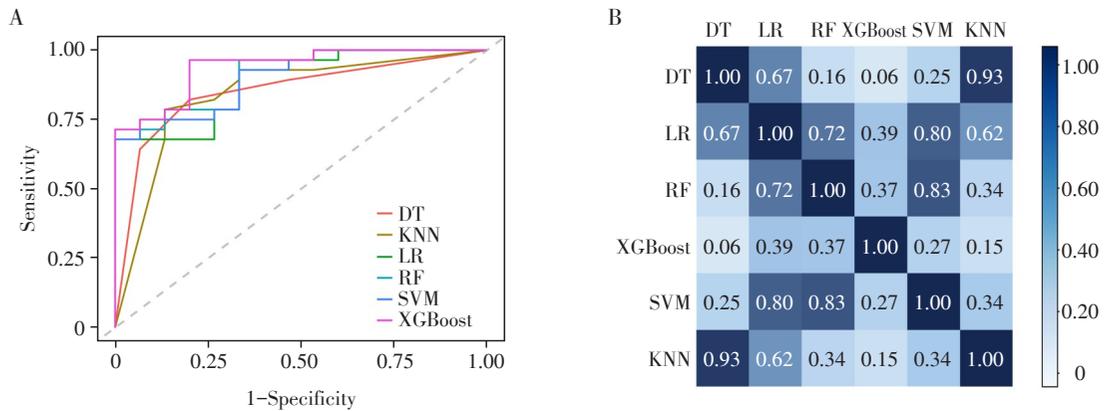
算法,纳入年龄、NIC、 $\lambda_{th}$ 、FT3、FT4、TSH和Radscore进行建模。在训练集中,AUC分别为0.935、1.000、1.000、0.953、1.000、0.956;在验证集中,AUC分别

为0.852、0.910、0.938、0.902、0.845、0.886。其中,XGBoost模型在训练集和验证集中的表现均为最优(表2、图2)。

表2 训练集与验证集中6种ML模型的预测性能

Table 2 Prediction performance of six ML models in the training and validation sets

Algorithm	ACC(%)	SEN(%)	SPE(%)	PPV(%)	NPV(%)	PRE(%)	REC(%)	F1(%)	AUC(95%CI)
Training set( $n=100$ , SMOTE: $n=38$ )									
DT	89.13	93.55	85.53	84.06	94.20	84.06	93.55	88.55	0.935(0.893–0.977)
RF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1.000(1.000–1.000)
XGBoost	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1.000(1.000–1.000)
SVM	92.03	92.75	91.30	91.43	92.65	91.43	92.75	92.09	0.953(0.915–0.990)
KNN	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	1.000(1.000–1.000)
LR	90.58	91.30	89.86	90.00	91.18	90.00	91.30	90.65	0.956(0.923–0.989)
Validation set( $n=43$ )									
DT	81.40	88.46	70.59	82.14	80.00	82.14	88.46	85.19	0.852(0.735–0.970)
RF	76.74	82.14	66.67	82.14	66.67	82.14	82.14	82.14	0.910(0.826–0.993)
XGBoost	86.05	89.29	80.00	89.29	80.00	89.29	89.29	89.29	0.938(0.870–1.000)
SVM	74.42	78.57	66.67	81.48	62.50	81.48	78.57	80.00	0.902(0.815–0.990)
KNN	81.40	89.29	66.67	83.33	76.92	83.33	89.29	86.21	0.845(0.716–0.974)
LR	74.42	78.57	66.67	81.48	62.50	81.48	78.57	80.00	0.886(0.788–0.983)



Comparison of six ML algorithms on the ROC curves (A) and confusion matrix deLong test (B) in the validation set, showing that the XGBoost had the best performance.

图2 6种ML算法预测模型比较

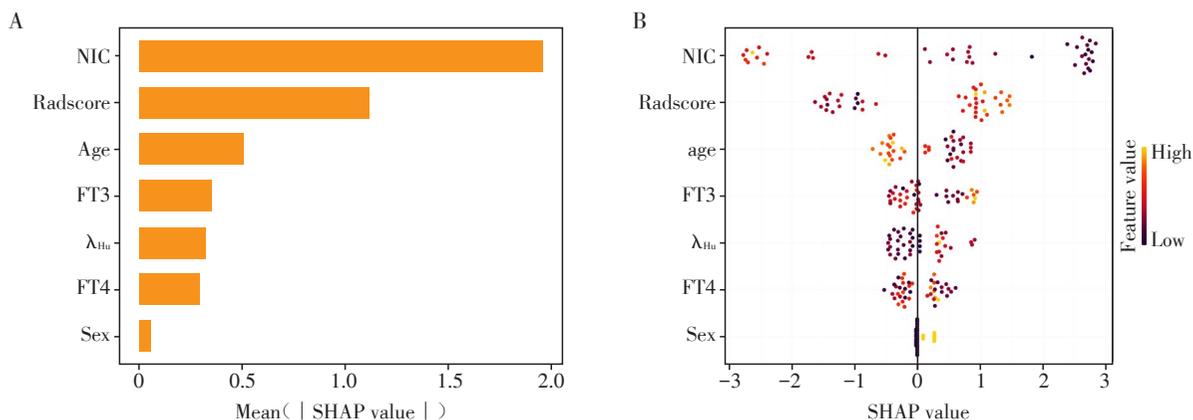
Figure 2 Compare of six ML algorithms for prediction models

### 2.4 列线图建立与验证

根据XGBoost模型SHAP特征权重分布图的排序(图3),选择前5的变量进行单因素和多因素Logistic回归分析,保留了具有统计学意义的变量NIC、Radscore和年龄( $P < 0.05$ ,表3)。基于这些变量构建列线图组合模型。通过校准曲线和DCA的评估证明,列线图是预测甲状腺良恶性结节的有效工具(图4)。

### 3 讨论

在本项回顾性研究中,利用DSCT能谱参数、患者临床资料及影像组学评分构建了一系列ML分类模型,各模型在预测甲状腺结节良恶性方面均表现出较高的性能( $AUC > 0.800$ ),以XGBoost算法的模型最为显著,其在验证集上的AUC为0.938。应用该模型可实现对甲状腺结节良、恶性状态的精准预



A: XGBoost model importance gain chart. B: XGBoost SHAP summary plot. The variables were sorted from top to bottom, and it can be observed that NIC, Radscore, age,  $\lambda_{th}$ , and FT3 are the top 5 influential features in the model.

图3 XGBoost模型变量重要性图与SHAP简要图

Figure 3 Importance plot and SHAP summary plot of XGBoost model

表3 XGBoost算法变量重要影响因素逻辑回归

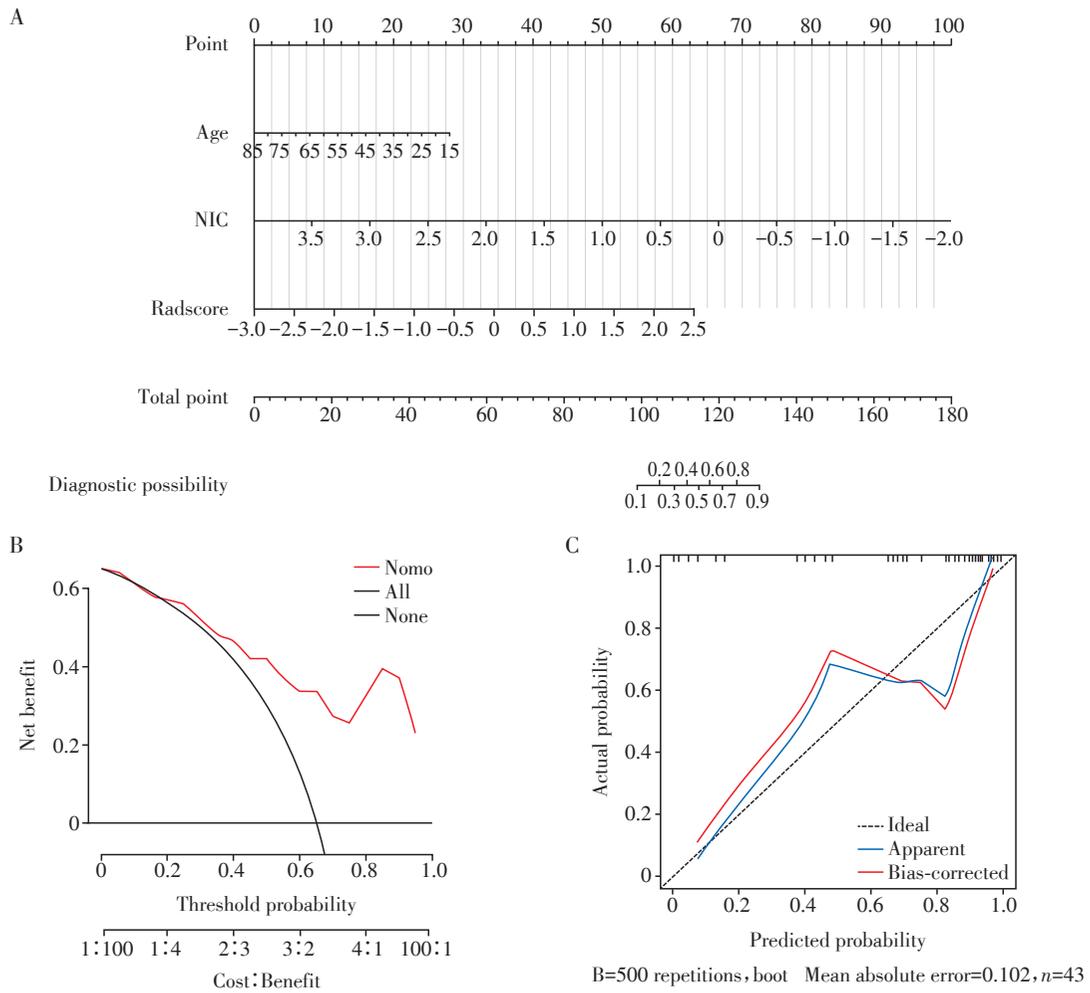
Table 3 XGBoost algorithm-variable importance factors for logistic regression

Characteristic	Univariate analysis				Multivariate analysis			
	OR	95% CI	Z	P	OR	95% CI	Z	P
NIC	0.100	0.042-0.200	-5.885	< 0.001	0.099	0.032-0.233	-4.662	< 0.001
Radscore	4.573	2.853-8.013	5.809	< 0.001	4.918	2.609-10.810	4.442	< 0.001
Age	0.954	0.926-0.981	-3.193	< 0.001	0.946	0.902-0.987	-2.465	0.014
$\lambda_{th}$	0.813	0.513-1.135	-1.045	0.296	-	-	-	-
FT3	1.061	0.764-1.485	0.355	0.722	-	-	-	-

测,进一步构建的列线图也展示了良好的预测能力和临床实用性。

光谱CT采用稀有金属钇和稀土陶瓷材质的双层探测器结构,在不更改常规扫描流程的同时可以获取能谱信息如碘密度图、能谱曲线等<sup>[15]</sup>。本研究发现恶性甲状腺结节的NIC显著低于良性结节,与先前的研究结果一致<sup>[16]</sup>。良性甲状腺结节测量出了较高的碘值,其原因主要是它们在组织结构上保留了正常甲状腺细胞的功能和更加丰富的血管供给,增强了碘造影剂的输送和积累<sup>[17]</sup>。与此相对,恶性甲状腺结节内部的纤维化、坏死或钙化以及更多的异常紊乱血管导致血流量下降,表现出碘值的降低<sup>[18]</sup>。本研究还发现,年龄是甲状腺恶性结节的预测风险因素。其原因可能与激素水平异常相关,特别是女性青春期与孕期的雌激素和孕激素波动,这些激素的显著变化可影响甲状腺细胞受体或模拟TSH的活性来改变甲状腺功能,进而促进甲状腺癌的发展<sup>[19]</sup>。此外,青年人群健康意识的提升及甲状腺超声检查的广泛应用,也可能是甲状腺结节检出率提高的原因之一。

影像组学通过对医学图像进行高通量的定量特征分析,能够揭示传统影像学中无法发现的图像特征,有助于建立诊断预测模型,为个体精准医疗提供有利的条件<sup>[20]</sup>。已有研究表明影像组学在预测甲状腺结节良恶性方面具有一定潜力, Park等<sup>[21]</sup>使用超声图像的影像组学特征进行分类得出的AUC为0.75,武欣欣等<sup>[22]</sup>结合CT的平扫与增强特征及临床数据得到了更好的AUC为0.793。本研究运用了差异性比较、组间一致性评估与LASSO的方法,最终筛选出了与良恶性相关的3个一阶特征与3个纹理特征。其中,一阶特征10Percentile与90Percentile表示直方图灰度强度的分布, Minimum则表示为图像的最小像素值, GLSZM、NGTDM与GLCM主要提供纹理的变化和细节丰富度的信息,上述特征表明良性与恶性结节在不同维度上的异质性<sup>[23]</sup>。为了提升ML预测效能,本研究比较了6种不同的ML算法,发现XGBoost表现最优,在验证集中的AUC达到了0.938。XGBoost是一种高效准确的集成学习方法,通过梯度提升和极端情况下的处理能力,能优化多个决策树以逐步减小模型的预测



A: A nomogram was constructed for the training set, incorporating Radscore, NIC, and age. B: Decision curve analysis (DCA) for the validation set. The horizontal axis represents the threshold probability, while the vertical axis indicates net benefit. The red line depicts the assumption of all cases being malignant; the green line represents the assumption of all cases being benign. The DCA of the validation set reveals that when the threshold exceeds 20%, predictions made using the nomogram offer better clinical benefits. C: The calibration curve for validation set was obtained through 500 bootstrap resampling. The 45° line represents ideal performance, and a closer proximity between the two curves indicates higher precision. The absolute error of the validation set was 0.102.

图4 列线图(A)、校准曲线(B)与决策曲线(C)分析

Figure 4 The nomogram(A), DCA(B), and the calibration(C) curves

误差,其引入了正则化算法,有效提高了鲁棒性的同时降低了过拟合<sup>[24]</sup>,在处理大规模数据和异常值等情况下表现出很好的实用价值。为了更加直观和便于临床应用,本研究通过SHAP值排列了模型中各个特征的重要性,并纳入了NIC、Radscore与年龄这3个特征建立了列线图,通过决策曲线分析,发现列线图对临床决策有明显的净效益,能够指引临床医生做出更加精确的判断。

本研究存在一些局限性:首先,回顾性的研究和较小的样本量使得结果存在一定偏倚;其次,在提升预测效能方面,尽管比较了多种ML算法的表现,但研究设备仅限于DSCT,缺少来自多中心及不

同设备的数据,需进一步扩大数据量以增强模型的性能和泛化能力;第三,即便采用了ICC以增加结果的可重复性,但由于CT增强后甲状腺肿瘤边缘的模糊现象不可避免,会对图像的手动分割准确度造成一定影响。

综上所述,本研究通过XGBoost算法整合了DSCT参数、临床因素及影像组学特征的ML模型及列线图,该模型可以无创地预测甲状腺结节的良恶性,为临床决策提供直观且可靠的参考。

[参考文献]

[1] ARAQUE D V P, BLEYER A, BRITO J P. Thyroid cancer in adolescents and young adults [J]. Future Oncol,

- 2017, 13(14): 1253-1261
- [2] BERTOL B C, MASSARO J D, DEBORTOLI G, et al. BRAF, TERT and HLA-G status in the papillary thyroid carcinoma: a clinicopathological association study[J]. *Int J Mol Sci*, 2023, 24(15): 12459
- [3] COHEN O, LAHAV G, SCHINDEL D, et al. Surgeon performed thyroid and neck ultrasound as a tool for better patient care[J]. *Harefuah*, 2020, 159(1): 128-131
- [4] 张中原, 杨艳丽, 王加林, 等. 穿刺针型号及甲状腺结节超声特征对超声引导下细针穿刺活检标本取材满意率的影响[J]. *南京医科大学学报(自然科学版)*, 2023, 43(8): 1140-1144
- [5] CHEN Z, WANG J J, GUO D M, et al. Combined fine-needle aspiration with core needle biopsy for assessing thyroid nodules: a more valuable diagnostic method?[J]. *Ultrasonography*, 2023, 42(2): 314-322
- [6] 石 灿, 彭 松, 梁 琪. 能谱CT及其联合甲状腺球蛋白检测对良恶性甲状腺结节的诊断价值[J]. *中国普通外科杂志*, 2021, 30(11): 1359-1366
- [7] MCCOLLOUGH C H, LENG S, YU L, et al. Dual- and multi-energy CT: principles, technical approaches, and clinical applications[J]. *Radiology*, 2015, 276(3): 637-653
- [8] YI R, LI T, XIE G, et al. Diagnosis of benign and malignant thyroid nodules by a dual-layer spectral detector CT-based nomogram[J]. *Front Oncol*, 2023, 13: 1132817
- [9] SCIALPI M, BIANCONI F, CANTISANI V, et al. Radiomic machine learning: is it really a useful method for the characterization of prostate cancer?[J]. *Radiology*, 2019, 291(1): 269-270
- [10] VERMA V, SIMONE C B, KRISHNAN S, et al. The rise of radiomics and implications for oncologic management[J]. *J Natl Cancer Inst*, 2018, 110(11): 1275-1276
- [11] HANDELMAN G S, KOK H K, CHANDRA R V, et al. eDoctor: machine learning and the future of medicine[J]. *J Intern Med*, 2018, 284(6): 603-619
- [12] XU H, WANG X, GUAN C, et al. Value of whole-thyroid CT-based radiomics in predicting benign and malignant thyroid nodules[J]. *Front Oncol*, 2022, 12: 828259
- [13] GAO S Y, ZHANG X Y, WEI W, et al. Identification of benign and malignant thyroid nodules by *in vivo* iodine concentration measurement using single-source dual energy CT: a retrospective diagnostic accuracy study[J]. *Medicine*, 2016, 95(39): e4816
- [14] SONG Z, LI Q, ZHANG D, et al. Nomogram based on spectral CT quantitative parameters and typical radiological features for distinguishing benign from malignant thyroid micro-nodules[J]. *Cancer Imaging*, 2023, 23(1): 13
- [15] 中华医学会放射学分会, 中国医师协会放射医师分会, 安徽省影像临床医学研究中心. 能量CT临床应用中国专家共识[J]. *中华放射学杂志*, 2022, 56(5): 476-487
- [16] LEE D H, LEE Y H, SEO H S, et al. Dual-energy CT iodine quantification for characterizing focal thyroid lesions[J]. *Head Neck*, 2019, 41(4): 1024-1031
- [17] WU W, FANG X, LI J, et al. Application of dual-source computed tomography in the diagnosis of thyroid cancer and evaluation of biological behaviors[J]. *J Clin Ultrasound*, 2023, 51(1): 195-202
- [18] CHEN D W, LANG B H H, MCLEOD D S A, et al. Thyroid cancer[J]. *Lancet*, 2023, 401(10387): 1531-1544
- [19] 吴华杰, 张 楠, 李 磊, 等. 甲状腺癌发病的危险因素及预防策略研究进展[J]. *昆明医科大学学报*, 2022, 43(5): 162-167
- [20] MAYERHOEFER M E, MATERKA A, LANGS G, et al. Introduction to radiomics[J]. *J Nucl Med*, 2020, 61(4): 488-495
- [21] PARK V Y, LEE E, LEE H S, et al. Combining radiomics with ultrasound-based risk stratification systems for thyroid nodules: an approach for improving performance[J]. *Eur Radiol*, 2021, 31(4): 2405-2413
- [22] 武欣欣, 李静静, 毛 宁, 等. 基于CT影像组学诺模图预测微小甲状腺结节良恶性[J]. *山东大学耳鼻喉眼学报*, 2020, 34(3): 32-39
- [23] 刘婷婷, 林佳璐, 娄鉴娟, 等. 多参数MRI影像组学评估浸润性乳腺癌HER-2表达状态的临床应用价值[J]. *南京医科大学学报(自然科学版)*, 2024, 44(2): 218-227
- [24] ZHANG X, YAN C, GAO C, et al. Predicting missing values in medical data via XGBoost regression[J]. *J Healthc Inform Res*, 2020, 4(4): 383-394

[收稿日期] 2024-01-14

(本文编辑: 蒋 莉)