




Differentiate Thyroid Follicular Adenoma from Carcinoma with Combined Ultrasound Radiomics Features and Clinical Ultrasound Features

Bing Yu¹ · Yanyan Li² · Xiangle Yu³ · Yao Ai¹ · Juebin Jin¹ · Ji Zhang¹ · YuHua Zhang² · Hui Zhu² · Congying Xie^{4,1} · Meixiao Shen³ · Yan Yang² · Xiance Jin^{1,5} 

Received: 5 August 2021 / Revised: 23 March 2022 / Accepted: 13 April 2022 / Published online: 26 April 2022
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2022

Abstract

Noninvasive differentiating thyroid follicular adenoma from carcinoma preoperatively is of great clinical value to decrease the risks resulted from excessive surgery for patients with follicular neoplasm. The purpose of this study is to investigate the accuracy of ultrasound radiomics features integrating with ultrasound features in the differentiation between thyroid follicular carcinoma and adenoma. A total of 129 patients diagnosed as thyroid follicular neoplasm with pathologically confirmed follicular adenoma and carcinoma were enrolled and analyzed retrospectively. Radiomics features were extracted from preoperative ultrasound images with manually contoured targets. Ultrasound features and clinical parameters were also obtained from electronic medical records. Radiomics signature, combined model integrating radiomics features, ultrasound features, and clinical parameters were constructed and validated to differentiate the follicular carcinoma from adenoma. A total of 23 optimal features were selected from 449 extracted radiomics features. Clinical and ultrasound parameters of sex ($p=0.003$), interior structure ($p=0.035$), edge ($p=0.02$), platelets ($p=0.007$), and creatinine ($p=0.001$) were associated with the differentiation between benign and malignant follicular neoplasm. The values of area under curves (AUCs) of the radiomics signature, clinical model, and combined model were 0.772 (95% CI: 0.707–0.838), 0.792 (95% CI: 0.715–0.869), and 0.861 (95% CI: 0.775–0.909), respectively. A final corrected AUC of 0.844 was achieved for the combined model after internal validation. Radiomics features from ultrasound images combined with ultrasound features and clinical factors are feasible to differentiate thyroid follicular carcinoma from adenoma noninvasive before operation to decrease the unnecessary of diagnostic thyroidectomy for patients with benign follicular adenoma.

Keywords Follicular neoplasm · Ultrasound · Radiomics · Classification

Bing Yu and Yanyan Li are contributed equally to this work.

✉ Meixiao Shen
smx77@sohu.com

✉ Yan Yang
yang25yan@126.com

✉ Xiance Jin
jinxc1979@hotmail.com

¹ Radiotherapy Center, Wenzhou Medical University First Affiliated Hospital, Wenzhou 32500, China

² Department of Ultrasound Imaging, Wenzhou Medical University Second Affiliated Hospital, Wenzhou 32500, China

³ Wenzhou Medical University, Wenzhou 325000, Zhejiang, China

⁴ Radiation and Medical Oncology, Wenzhou Medical University Second Affiliated Hospital, Wenzhou 32500, China

⁵ School of Basic Medical Science, Wenzhou Medical University, Wenzhou 325000, China

Introduction

Thyroid follicular neoplasm is a cytologic term that encompasses both the benign and malignant proliferation of thyroid follicular cells, which consists of follicular adenoma and carcinoma [1]. Follicular adenoma is a benign tumor while follicular carcinoma is the second most common thyroid cancer and comprises 10–20% of the thyroid cancer [2]. Follicular adenoma is a histologic niche between follicular hyperplasia and follicular carcinoma with overlapping clinical presentations, ultrasound features, and molecular biology compared with follicular carcinoma [3, 4]. Studies demonstrated that it is challenging to preoperatively differentiate follicular adenoma from carcinoma through ultrasound, fine-needle aspiration cytology, and immunohistochemistry [5]. Currently, the diagnosis of follicular carcinoma within a thyroid gland definitively is to identify capsular or vascular invasion at the periphery of the lesion among pathologic examination following diagnostic thyroidectomy [6]. However, only 15–40% of lesions classified as follicular neoplasm are malignant [7, 8]. Therefore, a noninvasive method of differentiating follicular adenoma from carcinoma preoperatively would be of great value to decrease the risks of laryngeal nerve injury and hypoparathyroidism resulted from excessive surgery for patients with follicular neoplasm [9].

Although the ability of ultrasound in differentiating thyroid follicular adenoma from carcinoma is questioned, it is still the first imaging modality of choice in the evaluation of the morphologic characteristics of thyroid nodules due to its advantages of high resolution, absence of ionization radiation, portability, and ease of use [10, 11]. Studies demonstrated that ultrasound features, such as hypoechoic, microcalcifications, and infiltrative margins, are associated with a high suspicion of malignancy [12–14]. On the contrary, studies also indicated that there is no value of ultrasound appearance in distinguishing follicular carcinoma from follicular adenoma [15, 16]. Effects on standardizing the assessment of ultrasound features have been carried out to improve the consistency of ultrasound feature-based diagnosis [12, 17].

With the emergency of radiomics, Shin et al. demonstrated that ultrasound radiomics features achieved an accuracy of 74.1% and 69.0% with artificial neural network (ANN) and support vector machine (SVM), respectively, in discriminating follicular adenoma from carcinoma on preoperative ultrasound images. Although this accuracy is higher than experienced radiologists (64.8%), it is still not accurate enough for clinical application [18]. The purpose of this study is to investigate the accuracy and feasibility of combing radiomics features with ultrasound

features and clinical parameters in the differentiating follicular adenoma from carcinoma on preoperative ultrasound images, so as to predict the malignancy of follicular neoplasm noninvasively to reduce the unnecessary surgery for patients with benign tumors.

Materials and Methods

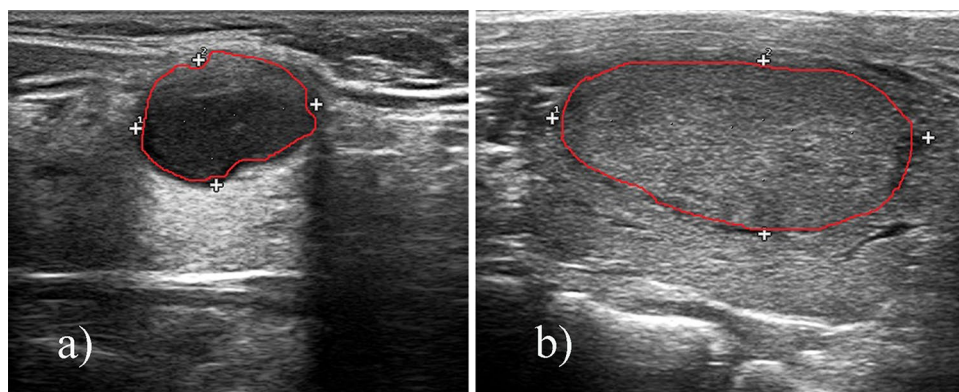
Patients

According to the electronic medical records, patients diagnosed as follicular neoplasm in authors' hospital from January 2015 to April 2020 were retrospectively reviewed. The inclusion criteria were as follows: (1) pathologically confirmed thyroid follicular carcinoma and follicular adenoma; (2) diagnosed by ultrasound images with detailed ultrasound features described. Ultrasound images, routine clinical tests, and patients' characteristics were also extracted from the electronic health records. The exclusion criteria were as follows: (1) patients with lack of digital imaging data; (2) treated with preoperative chemotherapy; and (3) with a history of other malignancies or combined malignancies. Consequently, 129 patients were enrolled in our study. This study was approved by the institutional review board and conducted in accordance with the Declaration of Helsinki (ECCR no. 2019059). Informed consent was waived by ECCR for the retrospective nature of this study.

Ultrasound Image Acquisition and Tumor Segmentations

Ultrasonography of the thyroid was conducted by trained sonographers using Hi Vision 900 system, model EUB-6500 (Hitachi Medical Corporation, Inc, Tokyo, Japan), Acuson Sequoia 512 system (Siemens Medical Solutions, Mountain View, CA), or iU22 system (Philips Healthcare, Bothell, WA) equipped with high-frequency 8–15-MHz linear transducers. Two specialists (YY Li and YH Zhang), each of whom had more than 5 years of experience in thyroid imaging, independently reviewed each set of nodule imaging findings. All images were reinterpreted based on ultrasound features, including nodule dimensions, shape (taller-wide-ratio), status (solitary or multinodular), structure (cystic, solid, or mixed), edge (smooth/unclear, lobular/irregular or outside the thyroid), echogenicity characteristics (isoechoic, hypoechoic, hyperechoic, or mixed), and presence of calcifications (absent, microcalcifications, macrocalcifications, or peripheral calcifications) were recorded, as well as age and sex of patients. Nodules on the ultrasound images were contoured by one junior radiologist and confirmed by a senior radiologist (with > 10 years of experience in thyroid sonography). A typical contour is presented in Fig. 1.

Fig. 1 A typical tumor segmentation on ultrasound images: **a** follicular adenoma, **b** follicular carcinoma



Feature Extraction and Model Building

Before radiomics feature extraction, intensity normalization was performed in ultrasound images to transform arbitrary gray intensity values into a standardized intensity range. Python (v. 3.7.0; <https://www.python.org/>) and package Pyradiomics 2.2.0 (version 2.2) were used to extract radiomics features from the manually segmented target volumes [19]. Based on different matrices that capture the spatial intensity distributions and wavelet filtering, a total of 449 radiomics features were extracted, including 90 first-order histogram statistics, 9 shape features, and 350 texture features from gray level co-occurrence matrix (GLCM), gray level dependence matrix (GLDM), gray-level run length matrix (GLRLM), and gray level size zone matrix (GLSZM). All these definitions of features were described by Imaging Biomarker Standardization Initiative (IBSI) [20].

Key radiomics features that are associated with follicular adenoma and carcinoma differentiation were selected by using Mann–Whitney U tests and the least absolute shrinkage selection operator (LASSO) [21]. Radiomics features with a $p < 0.05$ in Mann–Whitney U tests were selected as potentially informative features; then, LASSO was applied to identify optimal features for follicular adenoma and carcinoma differentiation. Ten-fold cross validation was applied to tune the elastic net parameters to reduce the redundant information and to avoid over-fitting. The elastic net penalty is controlled by α , and bridges the gap between lasso regression ($\alpha = 1$, the default) and ridge regression ($\alpha = 0$). LASSO regression model building was done using the “glmnet” package. Glmnet function in R language was applied for n cross validation ($n = 10$), which means that data was separated into 10 subsets. The model was trained with 9 subsets and tested with the remaining one subset. The glmnet algorithms use cyclical coordinate descent, which successively optimizes the objective function over each parameter with others fixed, and cycles repeatedly until convergence. Lasso regression is a regularization

technique used for more accurate prediction. A minimum standard deviation and maximum area under curves (AUCs) were achieved by tuning coefficient λ . The linear combination of selected radiomics features with respective weights makes the final radiomics signature.

Clinical Factors and Model Building

Ultrasound features of nodule dimensions, shape (taller-wide-ratio), status (solitary or multinodular), structure (cystic, solid, or mixed), edge (smooth/unclear, lobular/irregular or outside the thyroid), echogenicity characteristics (isoechoic, hypoechoic, hyperechoic, or mixed), and presence of calcifications (absent, microcalcifications, or macrocalcifications), were selected to differentiate the follicular adenoma and carcinoma. In order to investigate whether other clinical factors may differentiate the follicular adenoma and carcinoma, clinical parameters such as white blood cell (WBC), neutrophil (NEUT), lymphocyte (LYM), hemoglobin (HB), red blood cell count (RBC), platelets (PLT), alanine aminotransferase (ALT), aspartate aminotransferase (AST), albumin (ALB), blood urea nitrogen (BUN), and creatinine (CREA) were extracted from tests.

Univariate analysis was applied to select the related clinical parameters and ultrasound features in the differentiating between benign and malignant nodules. The difference of clinical parameters and ultrasound features between follicular adenoma and carcinoma was compared by using the chi-square test or by using the Mann–Whitney U test. Multivariate analyses used binary logistic regression which was applied to build the clinical model. Only the variables with a $p < 0.05$ were selected to build the clinical model with logistic regression. The combined model integrating the radiomics features and ultrasound features was built using logistic regression. A nomogram was built to further evaluate the performance of the combined model.

Model Evaluation and Statistical Analysis

The performance of differentiation models was evaluated with the receiver operating characteristic (ROC) curves. The AUCs were calculated along with a 95% confidence interval (CI) to evaluate the accuracy of these models. Since the validation group and the training group were conducted in the same group, this may overestimate the performance of the prediction model. Internal validation by bootstrap resampling with 1000 replicates was performed to correct the optimism of the model performance [22]. It extracts duplicate samples from the data set and replaces them. After estimating the original model separately in each bootstrap sample, the results of the parameters of interest in all bootstrap samples were checked. The frequency of occurrence of the variables of the final model in the bootstrap samples was used to assess the stability of the final model. Variables that occurred in more than 50% of the bootstrap models were judged to be reliable and were retained in the final model; otherwise, they were removed from the final model [23]. The goodness-of-fit of models was assessed by Nagelkerke R^2 , Akaike information criterion (AIC), and Brier score. The higher Nagelkerke R^2 indicates better calibration, and the lower AIC value and Brier score means the better of model fits. Statistical analysis was performed using R analysis platform (version 3.6.0), OriginPro2018 and MedCalc (version 19.3.0). LASSO regression model building was done using the “glmnet” package. For all tests, $p < 0.05$ was thought statically significant.

Results

Patients' Characteristics

A total of 129 patients diagnosed as thyroid follicular neoplasm were enrolled in this study with a mean age of 42.8 years (range: 4–84 years). There were 101 patients (30 males and 71 females) with pathologically confirmed follicular adenoma and 28 patients (13 males and 15 females) with follicular carcinoma, respectively. Detailed characteristics of enrolled patients are presented in Table 1. Due to the relatively small number of malignant cases, the characteristics of patients between two groups were not well balanced.

Radiomics Features and Clinical Factors

Of the 449 radiomics features, 26 were selected according to the Mann–Whitney U test with a $p < 0.05$ (Table S1). As shown in Fig. 2a, b, 23 features were further screened out from the 26 features to build the radiomics signature

using the LASSO logistic regression model. These features included 17 first-order features, and 6 Gy level run length matrix (GLRLM) features. The detail of the selected radiomics features (Table S1) and radiomics score calculation formula (Dos. S1) is shown in supplementary data, and the radiomics score for each patient was calculated.

The results of univariate analysis on preoperative clinical factors associated with histological subtypes are presented in Table 2. Diameter size and focal strong echogenicity were not associated with the differentiation of benign and malignant follicular neoplasm. As show in Table 2, for variables that showed a trend toward statistical significance in the univariate analysis, logistic regression analysis was applied in the further multivariate analysis. In the multivariate analysis, only “sex” ($p = 0.003$), “Interior structure” ($p = 0.035$), “edge” ($p = 0.02$), “PLT” ($p = 0.007$), and “CREA” ($p = 0.001$) of thyroid patient characteristics were associated with the differentiation of benign and malignant follicular neoplasm. Radiomics score was also integrated in multivariate analysis with clinical factors. Only the variables with a $p < 0.05$ were selected to build the clinical model.

Model Evaluation and Comparison

As shown in Fig. 3, the AUCs of the radiomics signature, clinical model, and combined model were 0.772 (95% CI: 0.707–0.838), 0.795 (95% CI: 0.721–0.870), and 0.861 (95% CI: 0.800–0.922), respectively. The results of radiomics score for each patient are shown in Fig. 4. The radiomics scores and predicted values in each model for patients with malignant tumor were obviously higher than those for patients with benign tumor. After internal validation, the combined model exhibited a higher goodness of fit (Nagelkerke R^2 : 0.8609; AIC: 185.02; Brier score: 0.122) and corrected performance (corrected AUC: 0.844), as shown in Table 3.

A nomogram was developed based on the combined model, as shown in Fig. 5a. Nomogram indicated that the combined model makes a better result in the differentiation of thyroid follicular adenoma and carcinoma compared with clinical model and radiomics signature alone. The calibration curve of the combined model showed the difference between the predicted probability of malignance and the actual probability. The “Ideal” line represents the perfect prediction as the predicted probabilities equal to the observed probabilities. The “Apparent” curve is the calibration of the entire cohort. The “Bias-correct” curve was the calibration created by internal validation of 1000-replicate bootstrap on the entire cohort. This shows that there is a good fit between the calibration curves and suitable for prediction, as shown in Fig. 5b.

Table 1 Characteristic of patients

Characteristic	Total (n, %) (n = 129)	Malignant (n = 28)	Benign (n = 101)	p value
Age				0.545
Mean (range)		47.04 (4–77)	41.55 (7–84)	
≤ 45	71 (55.0)	14 (50.0)	57 (56.4)	
> 45	58 (45.0)	14 (50.0)	44 (43.6)	
Sex				0.097
Males	43 (33.4)	13 (46.4)	30 (29.7)	
Females	86 (66.7)	15 (53.6)	71 (70.3)	
Diameter size (n, %)				0.024
> 1.5 cm	119 (92.2)	23 (82.1)	96 (95.0)	
< 1.5 cm	10 (7.8)	5 (17.9)	5 (5.0)	
Interior structure				0.032
Cystic and spongy	1 (0.8)	0 (0)	1 (1.0)	
Cystic solid mixture	21 (16.3)	1 (3.6)	20 (19.8)	
Solid	107 (82.9)	27 (96.4)	80 (79.2)	
Echogenicity				0.997
Iso-echoic	23 (17.8)	5 (17.9)	18 (17.8)	
Low echo	104 (80.6)	21 (75)	83 (82.2)	
Very low echo	2 (1.6)	2 (7.1)	0	
Shape				0.328
Aspect ratio < 1	127 (98.4)	27 (96.4)	100 (99.0)	
Aspect ratio > 1	2 (1.6)	1 (3.6)	1 (1.0)	
Edge				0.006
Smooth/unclear	123 (95.3)	24 (85.7)	99 (98.0)	
Lobular/irregular	5 (3.9)	3 (10.7)	2 (2.0)	
Outside the thyroid	1 (0.8)	1 (3.6)	0 (0)	
Focal strong echogenicity				0.154
Without strong echogenicity/ large tail	104 (80.6)	19 (67.9)	85 (84.1)	
Massive calcification	8 (6.2)	3 (10.7)	5 (5.0)	
Microscopic calcification	17 (13.2)	6 (21.4)	11 (10.9)	
WBC ($\times 10^9/L$)				0.021
4–10 (adult)	113 (87.6)			
5–12 (children)		28 (100)	85 (84.2)	
Abnormal range	16 (12.4)	0 (0)	16 (15.8)	
NEUT# ($\times 10^9/L$)				0.014
2.0–7.5	113 (87.6)	28 (100)	85 (84.2)	
< 2 and > 7.5	16 (12.4)	0 (0)	16 (15.8)	
LYM ($\times 10^9/L$)				0.925
0.8–4.0	124 (96.1)	27 (96.4)	97 (96.0)	
< 0.8 and > 4.0	5 (3.9)	1 (3.6)	4 (4.0)	
HB (g/L)				0.289
120–140 (children)	118 (91.5)			
110–150 (female)		27 (96.4)	91 (90.1)	
120–165 (male)				
Abnormal range	11 (8.5)	1 (3.6)	10 (9.9)	
RBC ($\times 10^{12}/L$)				0.313
3.5–5.0 (female)	107 (82.9)			
4.0–5.5 (male)		25 (89.3)	82 (81.2)	
Abnormal range	22 (17.1)	3 (10.7)	19 (18.8)	
PLT ($\times 10^9/L$)				0.328

Table 1 (continued)

Characteristic	Total (n, %) (n = 129)	Malignant (n = 28)	Benign (n = 101)	p value
100–300	109 (84.5)	22 (78.6)	87 (86.1)	
Abnormal range	20 (15.5)	6 (21.4)	14 (13.9)	
ALT (U/L)				0.899
0–40	116 (89.9)	25 (89.3)	91 (90.1)	
Abnormal range	13 (10.1)	3 (10.7)	10 (9.9)	
AST(U/L)				0.189
0–40	123 (95.3)	28 (100)	95 (94.1)	
Abnormal range	6 (4.7)	0 (0)	6 (5.9)	
ALB (g/L)				0.63
40–55	113 (87.6)	25 (89.3)	88 (87.1)	
Abnormal range	16 (12.4)	3 (10.7)	13 (12.9)	
BUN (mmol/L)				0.408
2.86–7.14	112 (86.8)	23 (82.1)	89 (88.1)	
Abnormal range	17 (13.2)	5 (17.9)	12 (11.9)	
CREA (μmol/L)				0.403
44–97 (female)	114 (88.4)			
53–106 (male)		26 (92.9)	88 (87.1)	
Abnormal range	15 (11.6)	2 (7.1)	13 (12.9)	

Categorical variables were compared by using the chi-square test; continuous variables were compared by using the Mann–Whitney *U* test

WBC, white blood cell; *NEUT*, neutrophil; *LYM*, lymphocyte; *HB*, hemoglobin; *RBC*, red blood cell count; *PLT*, platelets; *ALT*, alanine aminotransferase; *AST*, aspartate aminotransferase; *ALB*, albumin; *BUN*, blood urea nitrogen; *CREA*, creatinine

Discussion

The feasibility of combined radiomics features from ultrasound images, ultrasound features, and clinical parameters in the differentiating follicular adenoma from carcinoma was investigated in this study. A higher AUC of 0.844 was achieved with the combined model after internal validation

compared with radiomics feature (0.771) and ultrasound features (0.770) alone in the differentiating between benign and malignant follicular neoplasm.

Of the enrolled 129 patients with follicular neoplasm, 78.3% was pathologically confirmed as follicular adenoma and 21.7% was follicular carcinoma with a female-to-male ratio of 2:1. This is close to the reported ratio of

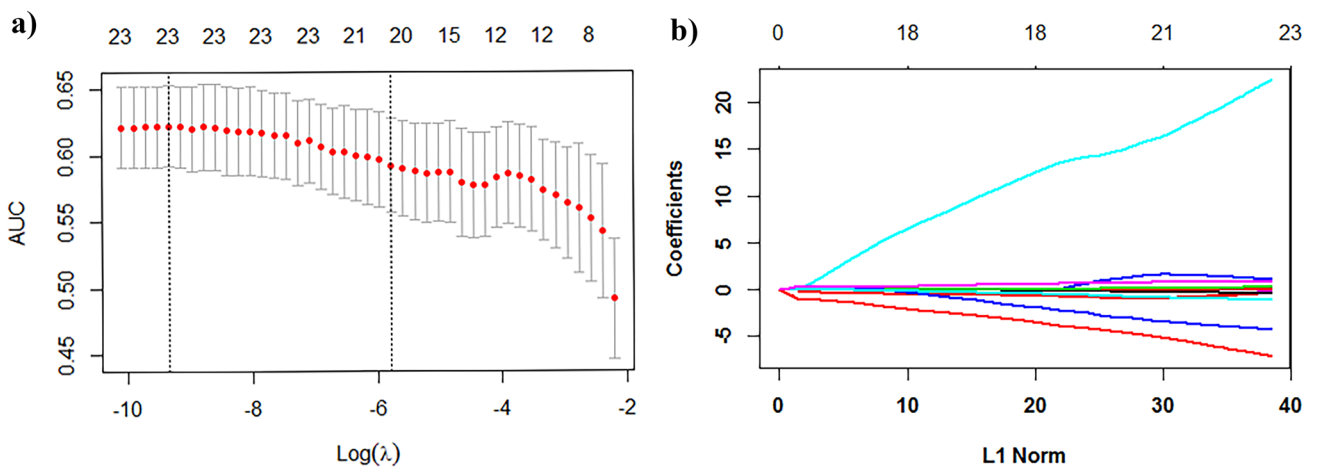


Fig. 2 Selection of histological subtype-associated radiomics features using the elastic net method. **a** Tuning parameter (λ) in the elastic net used tenfold cross validation via maximum area under curve and cri-

terion of minimum standard deviation were followed; the value of the hyperparameter alpha is 1; **b** the coefficient profiles of 23 radiomics features against the L1 norm (inverse proportional to $\log \lambda$)

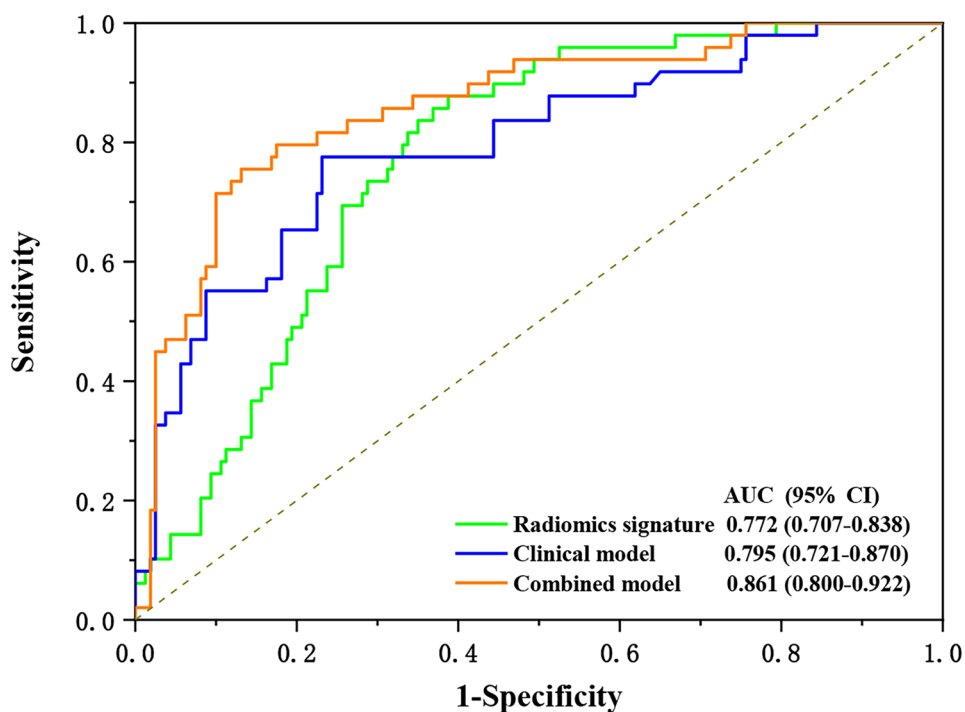
Table 2 Univariate and multivariate analysis of factors

Characteristic	Univariate analysis			Multivariate analysis						
	All			Clinical model			Combined model			
	OR	95% CI	<i>p</i>	OR	95% CI	<i>p</i>	OR	95% CI	<i>p</i>	
Age ($\leq 45, > 45$)	0.978	0.972–0.985	<0.01							
Sex	0.565	0.349–0.914	0.02	3.382	1.521–7.520	0.003	3.253	1.374–7.700	0.007	
Diameter size (> 1.5 cm, < 1.5 cm)	0.625	0.204–1.910	0.41							
Interior structure	0.557	0.47–0.66	<0.01	0.265	0.077–0.908	0.035	0.236	0.063–0.884	0.032	
Echogenicity	0.549	0.462–0.652	<0.01							
Shape (aspect ratio $< 1, > 1$)	0.302	0.219–0.417	<0.01							
Edge	1.858	0.867–3.978	0.111	0.026	0.001–0.569	0.02	0.507	0.019–0.623	0.001	
Focal strong echogenicity	0.907	0.724–1.136	0.395							
WBC ($\times 10^9/L$)	0.850	0.810–0.892	<0.01							
NEUT# ($\times 10^9/L$)	0.776	0.716–0.841	<0.01							
LYM ($\times 10^9/L$)	0.605	0.520–0.703	<0.01							
HB (g/L)	0.992	0.989–0.994	<0.01							
RBC ($\times 10^{12}/L$)	0.781	0.729–0.836	<0.01							
PLT ($\times 10^9/L$)	0.996	0.994–0.997	<0.01	1.01	1.003–1.017	0.007	1.01	1.004–1.017	0.002	
ALT (U/L)	0.951	0.935–0.966	<0.01							
AST (U/L)	0.951	0.937–0.966	<0.01							
ALB (g/L)	0.974	0.967–0.981	<0.01							
BUN (mmol/L)	0.819	0.771–0.870	<0.01							
CREA ($\mu\text{mol}/L$)	0.982	0.977–0.987	<0.01	1.057	1.024–1.091	0.001	1.067	1.034–1.102	0	
Radiomics score	1.114	1.083–1.145	<0.01				1.921	1.425–2.588	0	

Variables were compared by using the binary logistic regression

WBC, white blood cell; NEUT, neutrophil; LYM, lymphocyte; HB, hemoglobin; RBC, red blood cell count; PLT, platelets; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALB, albumin; BUN, blood urea nitrogen; CREA, creatinine

Fig. 3 Model evaluation with receiver operating characteristics curves and values of area under curves



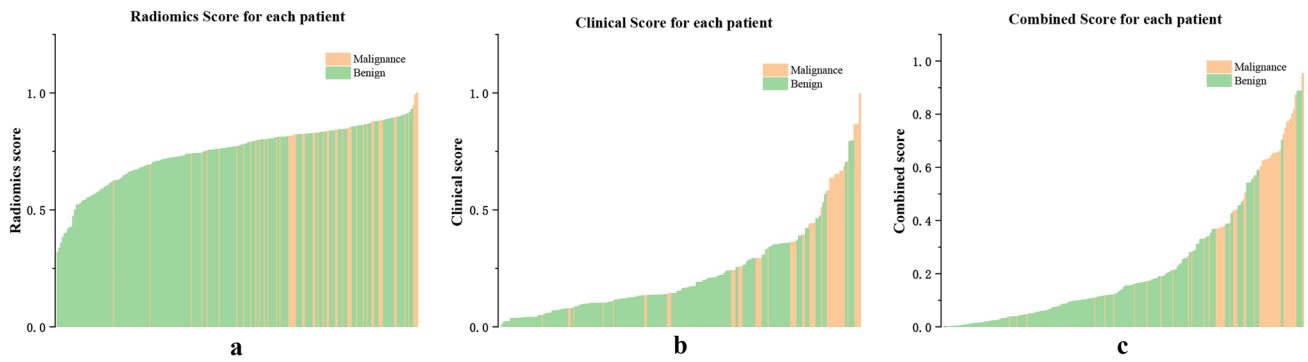


Fig. 4 Radiomics score for each patient of radiomics signature (a); of clinical model (b); of combined model (c)

approximately 80–90% of adenoma and 10–20% of carcinoma resulted from biopsy of follicular neoplasm [24, 25]. Similar, our study also indicated that follicular neoplasm occurs more often in woman than in man; however, the female-to male ratio is a bit lower than the reported 3:1. This may due to a relatively small number of patients was enrolled in this study.

The ratio of follicular adenoma to carcinoma is a clear evidence that distinguishing benign and malignant disease preoperatively is necessary to avoid overtreatment of patients with follicular adenoma. In this study, ultrasound features of interior structure, echogenicity, and shape were found to be associated with follicular carcinoma according to univariate analysis. Multivariate analysis indicated that interior structure and edge were significant in differentiating between follicular adenoma and carcinoma. Similarly, ultrasound features, such as hypoechoogenicity, noncircumscribed margins, and the presence of calcifications, were reported to be significantly associated with follicular carcinoma compared to follicular adenoma [3, 26]. In other studies, absence of internal cystic changes, lack of a perilesional halo on ultrasound, and larger diameter size have also been shown to be associated with follicular carcinoma as distinct from follicular adenoma [27, 28]. On the other hand, the ultrasound features associated with follicular carcinoma were inconsistent among different studies, and the

positive predictive values of these ultrasound features were low (ranging from 55.6 to 61.2%) [3, 27]. This might be caused largely by the inconsistency of the image quality of ultrasound across different machines and centers. It is insufficient for ultrasound features alone to distinguish a follicular adenoma from a carcinoma.

Studies indicated that cytology alone was also challenging to diagnosis of follicular neoplasm as cytologic features overlap in both benign and malignant follicular neoplasm [24]. Cytopathology results, as well as clinical variables, such as sex and age, were integrated with ultrasound images to predict the malignancy of thyroid nodules [29, 30]. Yoon et al. constructed a nomogram using ultrasound features and cytopathology results to predict the malignancy of thyroid nodules diagnosed as atypia of undetermined significance/follicular lesions of undetermined significance (AUS/FLUS) on ultrasonographic fine-needle aspiration (US-FNA) and achieved with an AUC of 0.817, compared with an AUC of 0.769 using final assessment, and an AUC of 0.779 when using the number of suspicious ultrasound features, respectively [31]. Similarly, in this study, an AUC of 0.770 after internal validation was achieved with the integration of clinical factors, such as “PLT” ($p=0.007$) and “CREA” ($p=0.001$), with ultrasound features in the differentiating between follicular adenoma and carcinoma.

Table 3 Model performance and evaluation

Model	Goodness of fit			Discrimination						Corrected performance
	Nagelkerke R^2	AIC	Brier Score	ACC	SPE	SEN	PPV	NPV	AUC	Internal validated AUC
Radiomics	0.4581	234.13	0.151	67.5	61.3	87.8	41.0	94.2	0.772	0.771
Clinical model	0.0443	195.38	0	76.1	76.9	77.6	49.4	91.7	0.795	0.770
Combined model	0.8609	185.02	0.122	81.8	86.9	75.5	58.5	92.4	0.861	0.844

Curve Internal validation was performed with 1000-replicate bootstrapping on the primary cohort

AIC, Akaike information criterion; ACC, accuracy; SPE, specificity; SEN, sensitivity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under receiver operating characteristic

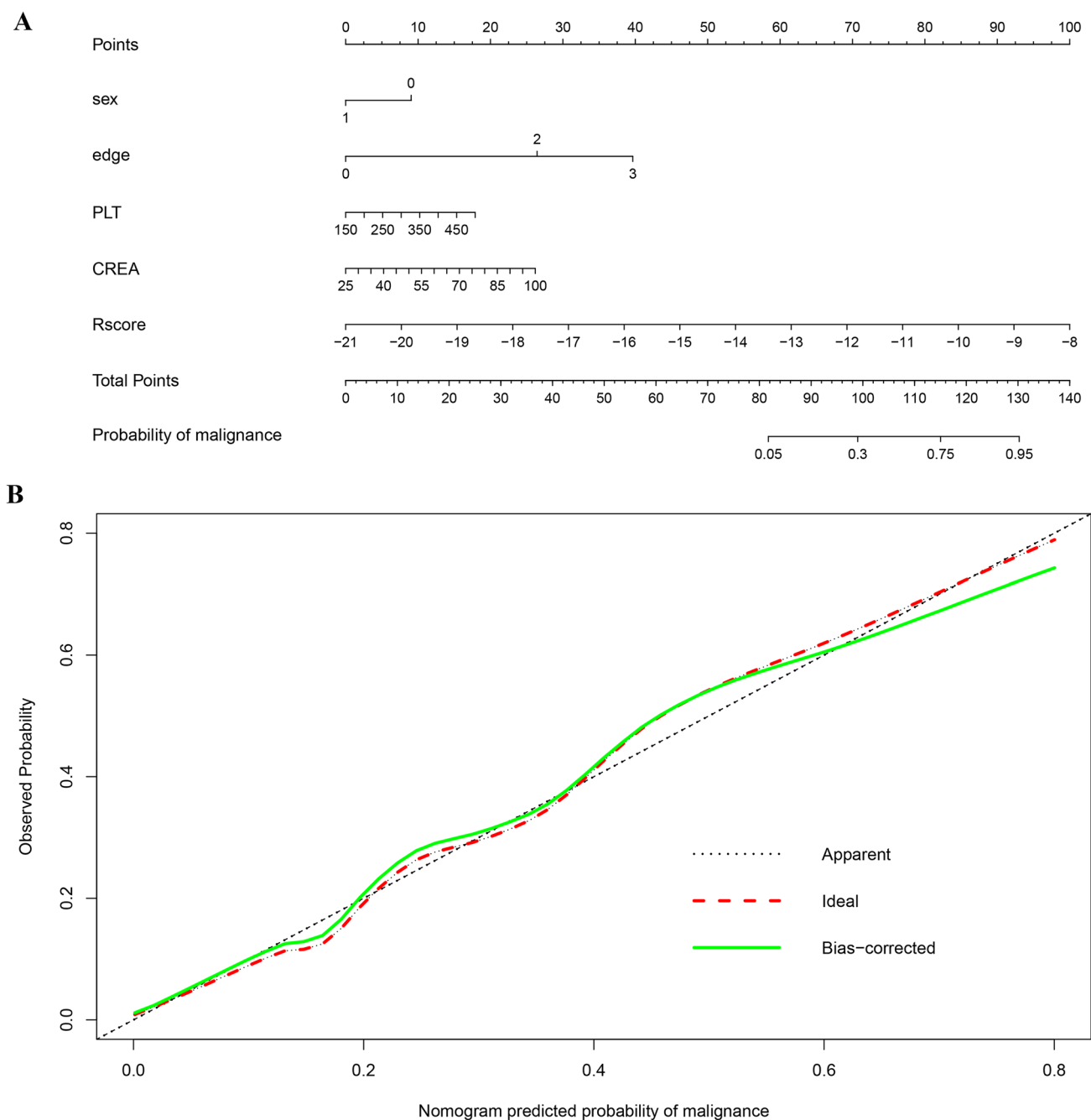


Fig. 5 **a** Nomogram developed based on the combined model; **b** calibration curve of the combined model. The “Ideal” line represents the perfect prediction as the predicted probabilities equal to the observed

probabilities. The “Apparent” curve is the calibration of the entire cohort. The “Bias-correct” curve was the calibration created by internal validation of 1000-replicate bootstrap on the entire cohort

With the emerging of deep learning and radiomics, Seo et al. conducted differentiation between follicular adenoma and carcinoma with 8-bit bitmap ultrasound images using a convolutional neural network (CNN) and achieved an AUC of 0.809 [32]. On another study, Shin et al. achieved an accuracy of only 0.741 and 0.69 using ANN and SVM, respectively, based preoperative ultrasonography in differentiating follicular adenoma from carcinoma [18]. Similarly,

in this study, we achieved an AUC of 0.771 with radiomics features alone in the differentiation of follicular adenoma from carcinoma. However, the performance of the integration model was improved by integrating radiomics features with ultrasound features and clinical parameters with an achieved AUC of 0.844.

Higher serum thyroid-stimulating hormone (TSH) levels had been reported to be an independent predictor of malignancy and

associated with increased risk of differentiated thyroid carcinoma and advanced tumor stage in elderly patients [33, 34]. Although the training-validation group method and external validation were not performed, 1000 bootstrap internal validation applied in this study demonstrated good performance [35]. Unfortunately, the TSH levels of the patients enrolled in this study were not available and did not integrate into the combined prediction model. One limitation of this study is the relatively small sample from one center. Future studies with a large sample from multiple centers in a prospective nature are urgently needed to further validate the prediction feasibility and accuracy of the combined radiomics features and clinical parameters in the differentiation of follicular adenoma and carcinoma. Another limitation of this study is the lack of real-time, prospective radiologist interpretation and comparison with models. Future work with more state-of-the-art deep learning techniques in the differentiation of follicular adenoma and carcinoma is needed.

Conclusions

Radiomics features from ultrasound images combined with ultrasound features and clinical factors are feasible to differentiate follicular carcinoma from adenoma noninvasive before operation to decrease the unnecessary of diagnostic thyroidectomy for patients with benign follicular adenoma.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10278-022-00639-2>.

Acknowledgements This work was partially funded by Radiation Oncology Basic and Translational Research Key Lab of Wenzhou (2021100848)

Funding This work was partially funded by Wenzhou Municipal Science and Technology Bureau (2018ZY016, 2019) and National Natural Science Foundation of China (No.11675122, 2016).

Availability of Data and Material Yes.

Code Availability Yes.

Declarations

Ethics Approval and Consent to Participate This study was approved by the institutional review board and conducted in accordance with the Declaration of Helsinki (ECCR no. 2019059). Informed consent was waived by ECCR for the retrospective nature of this study.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

References

1. Stolf BS, Santos MM, Simao DF, et al: Class distinction between follicular adenomas and follicular carcinomas of the thyroid gland on the basis of their signature expression. *Cancer* 2006;106.:1891–1900.
2. Howlader N, Noone AM, Krapcho M et al: SEER Cancer Statistics Review, 1975–2009 (Vintage 2009 Populations), National Cancer Institute. Bethesda, MD 2012.
3. Yoon JH, Kim EK, Youk JH, Moon HJ, Kwak JY: Better understanding in the differentiation of thyroid follicular adenoma, follicular carcinoma, and follicular variant of papillary carcinoma: a retrospective study. *Int J Endocrinol* 2014;32:1595.
4. Sobrinho-Simões M, Eloy C, Magalhães J, Lobo C, Amaro T: Follicular thyroid carcinoma. *Mod Pathol Suppl* 2011; 2:S10-8.
5. Baloch ZW, Fleisher S, LiVolsi VA, Gupta PK: Diagnosis of “follicular neoplasm”: a gray zone in thyroid fine-needle aspiration cytology. *Diagn Cytopathol* 2002;26(1):41-4.
6. McHenry CR, Phitayakorn R: Follicular adenoma and carcinoma of the thyroid gland. *Oncologist* 2011;16(5):585-93.
7. Hodak SP, Rosenthal DS. American Thyroid Association Clinical Affairs Committee Information for clinicians: commercially available molecular diagnosis testing in the evaluation of thyroid nodule fine-needle aspiration specimens. *Thyroid* 2013;23(2):131-4.
8. Baloch ZW, Seethala RR, Faquin WC, et al: Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): a changing paradigm in thyroid surgical pathology and implications for thyroid cytopathology. *Cancer Cytopathol* 2016;124(9):616-20.
9. Castro MR, Gharib H: Continuing controversies in the management of thyroid nodules. *Ann Intern Med* 2005; 142(11):926-31.
10. Moon WJ, Jung SL, Lee JH, et al: Benign and malignant thyroid nodules: US differentiation—multicenter retrospective study. *Radiology* 2008;247(3):762-70.
11. Hong YJ, Son EJ, Kim EK, Kwak JY, Hong SW, Chang HS: Positive predictive values of sonographic features of solid thyroid nodule. *Clin Imaging* 2010;34(2):127-33.
12. Friedrich-Rust M, Meyer G, Dauth N, et al: Interobserver agreement of Thyroid Imaging Reporting and Data System (TIRADS) and strain elastography for the assessment of thyroid nodules. *PLoS* 2013; One 8(10):e77927.
13. Alexander EK, Cooper D: the importance, and important limitations, of ultrasound imaging for evaluating thyroid nodules. *JAMA Intern Med* 2013;173(19):1796-7.
14. Lin JD, Hsueh C, Chao TC, Weng HF, Huang BY: Thyroid follicular neoplasms diagnosed by high-resolution ultrasonography with fine needle aspiration cytology. *Acta Cytol* 1997;41(3):687-91.
15. Koike E, Noguchi S, Yamashita H, et al: Ultrasonographic characteristics of thyroid nodules: prediction of malignancy. *Arch Surg* 2001;136(3):334-7.
16. Rago T, Di Coscio G, Basolo F, et al: Combined clinical, thyroid ultrasound and cytological features help to predict thyroid malignancy in follicular and Hürthle cell thyroid lesions: results from a series of 505 consecutive patients. *Clin Endocrinol (Oxf)* 2007;66(1):13-20.
17. 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):1748-51.
18. Shin I, Kim YJ, Han K, et al: Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. *Ultrasonography* 2020;39(3):257-265.

19. van Griethuysen JJM, Fedorov A, Parmar C, et al: Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77(21):e104-e107.
20. Zwanenburg A, Leger S, Vallières M: Steffen Image biomarker standardisation initiative. eprint arXiv 2016;1612.07003
21. Friedman J, Hastie T, Tibshirani R: Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33(1):1-22.
22. Efron B: The bootstrap and modern statistics. *J Am Stat Assoc* 2000;95:1293-6.
23. Blackstone EH: Breaking down barriers: helpful breakthrough statistical methods you need to understand better. *J Thorac Cardiovasc Surg* 2001;122:430-9.
24. Smith J, Cheifetz RE, Schneiderei N, Berean K, Thomson T: Can cytology accurately predict benign follicular nodules? *Am J Surg* 2005;189(5):592-5.
25. Carpi A, Nicolini A, Gross MD, et al: Controversies in diagnostic approaches to the indeterminate follicular thyroid nodule. *Biomed Pharmacother* 2005;59(9):517-20.
26. Kuo TC, Wu MH, Chen KY, Hsieh MS, Chen A, Chen CN: Ultra-sonographic features for differentiating follicular thyroid carcinoma and follicular adenoma. *Asian J Surg* 2020;43(1):339-346.
27. Sillery JC, Reading CC, Charboneau JW, Henrichsen TL, Hay ID, Mandrekar JN: Thyroid follicular carcinoma: sonographic features of 50 cases. *AJR Am J Roentgenol* 2010;194(1):44-54.
28. Zhang JZ, Hu B: Sonographic features of thyroid follicular carcinoma in comparison with thyroid follicular adenoma. *J Ultrasound Med* 2014;33(2):221-7.
29. Kim DW, Lee EJ, Jung SJ, Ryu JH, Kim YM: Role of sonographic diagnosis in managing Bethesda class III nodules. *AJNR Am J Neuroradiol* 2011;32(11):2136-41.
30. Gweon HM, Son EJ, Youk JH, Kim JA: Thyroid nodules with Bethesda system III cytology: can ultrasonography guide the next step? *Ann Surg Oncol* 2013;20(9):3083-8.
31. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY: A nomogram for predicting malignancy in thyroid nodules diagnosed as atypia of undetermined significance/follicular lesions of undetermined significance on fine needle aspiration. *Surgery* 2014; 155(6):1006-13.
32. Seo JK, Kim YJ, Kim KG, Shin I, Shin JH, Kwak JY: Differentiation of the follicular neoplasm on the gray-scale US by image selection subsampling along with the marginal outline using convolutional neural network. *Biomed Res Int* 2017:3098293.
33. Boelaert K, Horacek J, Holder RL, Watkinson JC, Sheppard MC, Franklyn JA: Serum thyrotropin concentration as a novel predictor of malignancy in thyroid nodules investigated by fine-needle aspiration. *J Clin Endocrinol Metab* 2006;91(11):4295-301.
34. Haymart MR, Repplinger DJ, Levenson GE, et al: Higher serum thyroid stimulating hormone level in thyroid nodule patients is associated with greater risks of differentiated thyroid cancer and advanced tumor stage. *J Clin Endocrinol Metab* 2008;93(3):809-14.
35. Brunelli A, Rocco G: Internal validation of risk models in lung resection surgery: bootstrap versus training-and-test sampling. *J Thorac Cardiovasc Surg* 2006;131(6):1243-7.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.