



OPEN ACCESS

EDITED BY

Xin-Wu Cui,
Huazhong University of Science and
Technology, China

REVIEWED BY

Angelika Buczyńska,
Medical University of Białystok, Poland
Zhen Wang,
Shandong First Medical University, China
Yifeng Zhang,
Tongji University, China
Lianzhong Zhang,
Henan Provincial People's Hospital, China

*CORRESPONDENCE

Jundong Yao
✉ yjd2313@outlook.com

[†]These authors have contributed equally to
this work

RECEIVED 22 January 2025

ACCEPTED 14 April 2025

PUBLISHED 08 May 2025

CITATION

Cui S, Liu Q, Wang H, Li H, Li W, Li C, Bi L,
Mu Y, Guo W, Yao J and Zhang Z (2025)
The value of a combined model based on
ultra-radiomics and multi-modal ultrasound
in the benign-malignant differentiation of
C-TIRADS 4A thyroid nodules: a prospective
multicenter study.
Front. Oncol. 15:1543020.
doi: 10.3389/fonc.2025.1543020

COPYRIGHT

© 2025 Cui, Liu, Wang, Li, Li, Bi, Mu, Guo,
Yao and Zhang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

The value of a combined model based on ultra-radiomics and multi-modal ultrasound in the benign-malignant differentiation of C-TIRADS 4A thyroid nodules: a prospective multicenter study

Shuai Cui^{1†}, Qifan Liu^{1†}, Hailong Wang¹, Husha Li¹, Wei Li¹,
Chenlong Li², Leilei Bi³, Yang Mu⁴, Wenjing Guo¹,
Jundong Yao^{1*} and Zhoulong Zhang¹

¹Department of Ultrasound, The First Affiliated Hospital, College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China, ²Department of Ultrasound, General Hospital of Pingmei Shenma Medical Group, Pingdingshan, China, ³Department of Ultrasound, Xiangyang Hospital of Traditional Chinese Medicine, Xiangyang, China, ⁴Department of Ultrasound, The Fourth Affiliated Hospital of Xinjiang Medical University, Urumqi, China

Objective: To establish a combined model based on ultrasound radiomics combined with multimodal ultrasound and evaluate its value in diagnosing benign and malignant nodules classified as Chinese-Thyroid Imaging Report and Data System (C-TIRADS) 4A.

Methods: Prospective collection of data from 446 patients with thyroid nodules classified as C-TIRADS 4A between December 2023 and August 2024. Based on the enrollment timeline, patients were divided into a training set (n=312) and a test set (n=134) in a 7:3 ratio. Using clinical information, multimodal ultrasound features, and radiomics features, a radiomics model was constructed using the Random Forest (RF) machine learning algorithm. Logistic regression was employed to develop the multimodal ultrasound model and the combined model. The predictive efficiency and accuracy of these models were evaluated using Receiver Operating Characteristic (ROC) curves, calibration curves, and Decision Curve Analysis (DCA). The diagnostic efficacy of junior physicians assisted by the ultrasound radiomics model was compared with that of senior physicians. DeLong's test was performed to compare the diagnostic performance of the models.

Results: Multivariate analysis revealed that age (≤ 51 years), Sound Touch Elastography mean stiffness (STE Mean), orientation (vertical), margin (blurred), and margin (irregular) were independent risk factors for papillary thyroid carcinoma, and the multimodal ultrasound model was established. Based on 17 ultrasound radiomics features, a radiomics model was constructed using the RF machine learning algorithm. The combined model was developed by combining the two aforementioned models. In the training set, the areas under the curve (AUC) of the multimodal ultrasound model, ultrasound radiomics model, and combined model were 0.852, 0.940 and 0.956, respectively. In the test set, the AUC were 0.804, 0.832 and 0.863, respectively. DeLong's test showed that the

combined model performed best in the training set, and in the test set, the combined model outperformed the multimodal ultrasound model but showed no significant difference compared to the radiomics model. DCA indicated that the combined model achieved higher net benefits within a specific threshold probability range (0.15–0.90).

Conclusion: The combined model exhibits robust diagnostic capability in distinguishing benign from malignant thyroid nodules classified as C-TIRADS 4A.

KEYWORDS

C-TIRADS 4A, multimodal ultrasound, ultrasound radiomics, papillary thyroid cancer, benign or malignant

1 Introduction

Thyroid cancer is one of the fastest-growing malignancies in terms of incidence worldwide. Over the past three decades, its global incidence has risen significantly, currently ranking as the seventh most common cancer globally (1, 2). Papillary thyroid cancer (PTC) accounts for more than 80% of all thyroid cancer cases (3). Although its prognosis is relatively favorable, the issue of overtreatment during diagnosis and management has become increasingly prominent.

Ultrasound examination has become the preferred imaging tool for risk assessment of thyroid nodules due to its non-invasive, real-time, and reproducible advantages. The establishment of the Chinese-Thyroid Imaging Report and Data System (C-TIRADS) has provided an important framework for standardizing the risk assessment of thyroid nodule malignancy (4). However, the malignancy risk of C-TIRADS 4A nodules spans a relatively wide range (2%–10%) (5, 6), leading to a significant number of patients undergoing unnecessary fine-needle aspiration biopsies or surgeries due to diagnostic uncertainty. This not only increases healthcare costs but may also cause patient anxiety. Therefore, there is an urgent need for a more precise differentiation method to optimize the management strategy for C-TIRADS 4A nodules.

Currently, the traditional C-TIRADS primarily relies on the morphological features of grayscale ultrasound (such as margins, echogenicity, calcifications, etc.) for risk stratification. However, the introduction of multimodal ultrasound techniques, including elastography, Color Doppler Flow Imaging (CDFI), and contrast-enhanced ultrasound, has provided multidimensional information

for assessing nodule stiffness, vascular characteristics, and microcirculation. For example, Gong et al. (7) demonstrated that combining grayscale ultrasound with contrast-enhanced ultrasound improved the diagnostic the areas under the curve (AUC) from 0.844 to 0.897. Additionally, Nattabi et al. (8) further showed that nodule stiffness measured by Shear Wave Elastography (SWE) is significantly positively correlated with malignancy risk. Nevertheless, these techniques still heavily depend on physician experience, and there are notable differences in diagnostic consistency among physicians of varying experience levels, with the learning curve for junior physicians being particularly challenging.

The emergence of radiomics offers a novel approach to addressing the aforementioned challenges. This technology extracts high-throughput deep imaging features (such as texture, morphology, heterogeneity, etc.) and combines them with machine learning algorithms to construct objective quantitative models. It has already demonstrated remarkable potential in the differential diagnosis of tumors such as breast cancer and liver cancer (9–11). In the field of thyroid imaging, the application of ultrasound radiomics is still in its exploratory stages. However, preliminary studies suggest that it can effectively capture malignant features that are difficult to identify using traditional methods, such as microstructural heterogeneity within nodule (12). Nevertheless, the complementary nature of multimodal ultrasound parameters (e.g., vascular characteristics, elasticity modulus) and radiomics features has not yet been fully explored, and there is a lack of validation based on multicenter data.

This study aims to develop a machine learning model that integrates multimodal ultrasound features and ultrasound radiomics for the differentiation of benign and malignant thyroid nodules classified as C-TIRADS 4A. Through prospective multicenter data validation, the diagnostic performance of the model and its potential to assist junior physicians will be evaluated. The goal is to reduce unnecessary invasive procedures, shorten the learning curve for junior physicians, and enhance their diagnostic capabilities, thereby optimizing clinical decision-making processes.

Abbreviations: AUC, Areas under the curve; C-TIRADS, Chinese-Thyroid Imaging Report and Data System; PTC, Papillary thyroid cancer; DCA, Decision curve analysis; RF, Random Forest; ROC, Receiver Operating Characteristic; ROI, Region of interest; SD, Standard deviation; STE, Sound Touch Elastography; STQ, Sound Touch Quantify; SE, Strain Elastography; CV, Cross-validation; MSE, Mean square error; FNR, False Negative Rate; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

2 Materials and methods

2.1 Study population

Prospectively collect data from patients with thyroid nodules classified as C-TIRADS 4A by ultrasound in multicenter medical institutions from December 2023 to August 2024. The inclusion criteria were as follows: ① age ≥ 18 years; ② ultrasound diagnosis of single or multiple C-TIRADS 4A nodules; ③ undergoing fine-needle aspiration biopsy (FNA) combined with BRAF V600E gene testing or surgical resection with definitive pathological results. The exclusion criteria were as follows: ① poor-quality ultrasound images (suboptimal image quality or insufficient resolution); ② history of neck surgery, cancer, or pregnancy; ③ incomplete clinical information; ④ non-PTC malignancies. A total of 446 patients were ultimately included, with 314 cases from The First Affiliated Hospital of Henan University of Science and Technology, 53 cases from General Hospital of Pingmei Shenma medical group, 48 cases from Xiangyang Hospital of Traditional Chinese Medicine, and 31 cases from The Fourth Affiliated Hospital of Xinjiang Medical University. Based on the enrollment timeline, patients were divided into a training set and a test set in a 7:3 ratio. Figure 1 illustrates the patient enrollment flowchart. This study received approval from the Ethics Committee of the First Affiliated Hospital of Henan University of Science and Technology (2024-03-K0160). All participants provided written informed consent, in compliance with regulations of the institution and the guidelines of the Declaration of Helsinki.

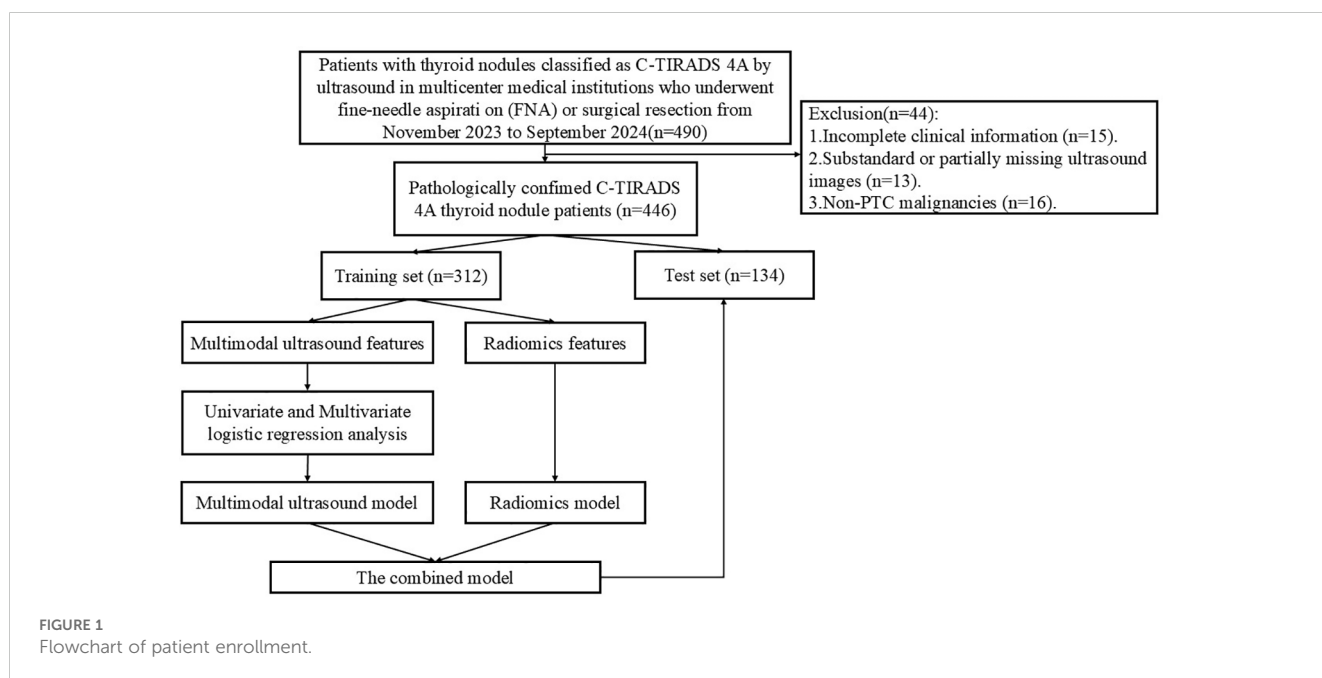
2.2 Ultrasound image acquisition

Prior to data collection, all ultrasound physician underwent centralized training on C-TIRADS 4A classification criteria,

including grayscale ultrasound features (e.g., margins, echogenicity, calcifications) and multimodal parameters (e.g., elastography thresholds). All participating centers utilized identical ultrasound systems (Mindray Resona I9) equipped with linear array probes (frequency range: 7.5–12 MHz). Imaging parameters were standardized across centers, including gain (60–70 dB), depth (2.5–3.5 cm), dynamic range (50–60 dB), and mechanical index (MI: 0.8–1.0). The patient was positioned supine, with the neck fully extended to expose the thyroid area. Grayscale images, color Doppler images, and dynamic images of nodules were collected in both transverse and longitudinal sections. Elastography settings were uniformly configured to ensure consistent Young's modulus calculations. Subsequently, three types of elastography images were acquired: Sound Touch Elastography (STE), Strain Elastography (SE), and Sound Touch Quantification (STQ). All images are stored in DICOM format.

2.3 Ultrasound image analysis

Ultrasound image feature was conducted by two doctors, Jundong Yao and Wei Li, both possessing 10 years of experience in ultrasonic diagnosis of thyroid diseases. They independently analyzed the images without access to the patients' clinical information or pathological results. In cases of differing opinions, a third doctor, Zhoulong Zhang, who has over 30 years of experience in ultrasound diagnosis, made the final decision. This study evaluates several features of nodules, including maximum diameter of the nodule, multiplicity (solitary/multiple), margin (smooth/ill-defined/irregular), halo sign (present/absent), composition (cystic-solid/solid/predominantly solid), echogenicity (isoechoic/markedly hypoechoic/hypoechoic/hyperechoic/heterogeneous), echo texture (homogeneous/heterogeneous), orientation (vertical/parallel), calcification (absent/coarse calcification/



microcalcification/indeterminate punctate echogenic foci/peripheral calcification), relationship to capsule (>2 mm/ ≤ 2 mm/extracapsular extension), and CDFI classification according to Alder (levels 0/I/II/III). The region of interest (ROI) for elastography is defined by manually tracing the contours of nodules using a specialized machine. The shell denotes a machine-generated boundary that extends 2 mm beyond the nodule following the tracing process. By outlining the nodule area, the system automatically calculates the STE mean stiffness (STE Mean), STE maximum stiffness (STE Max), STE minimum stiffness (STE Min), STE standard deviation (STE SD), SE mean stiffness (SE Mean), SE maximum stiffness (SE Max), SE minimum stiffness (SE Min), SE standard deviation (SE SD), STQ mean stiffness (STQ Mean), STQ maximum stiffness (STQ Max), STQ minimum stiffness (STQ Min), STQ standard deviation (STQ SD).

2.4 Pathological analysis

Ultrasound-guided fine-needle aspiration biopsy was performed by ultrasonologists who had completed standardized training and obtained qualification certificates. A 23G fine needle was utilized to efficiently extract samples from the thyroid lesion under ultrasound guidance. The cells were preserved in a liquid-based culture medium for subsequent genetic analysis. The test samples were then sent for pathological examination and genetic testing. A pathological diagnosis of PTC is deemed positive. The BRAF V600E mutation (a key genetic driver of PTC) is strongly associated with tumor aggressiveness, extrathyroidal extension, and lymph node metastasis (13). Its detection enhances diagnostic specificity for malignancy. If the BRAF V600E mutation is detected, it is recorded as positive; if not, it is noted as negative. All results were followed up for 6 months.

2.5 Model building

2.5.1 Construction of the multimodal ultrasound model

Univariate logistic regression analysis was conducted on the selected variables, which included gender, age, and ultrasound image characteristics such as STE Mean, STE Max, STE Min, STE SD, SE Mean, SE Max, SE Min, SE SD, STQ Mean, STQ Max, STQ Min, STQ SD, maximum diameter of the nodule, multiplicity, margin, halo sign, composition, echogenicity, echo texture, orientation, calcification, relationship to capsule and CDFI of the nodule. Variables with a p -value < 0.05 were considered risk factors. Subsequently, a multi-factor logistic regression analysis was performed, and independent risk factors were identified through stepwise forward logistic regression to construct a multimodal ultrasound model (Multi-model).

2.5.2 Construction of the ultrasound radiomics model

An ultrasound physician with 5 years of diagnostic experience (Husha Li) selected representative thyroid nodule images from each

patient who met the inclusion criteria, utilizing the RadiAnt DICOM Viewer 2021.1 software (Medixant, Poznan, Poland), and saved the selected images in DICOM format (Figure 2A). Another ultrasound physician with 5 years of experience in ultrasound diagnosis, Hailong Wang, manually depicted the ROI images of the selected patients using the polygon mode in ITK-SNAP software (www.itksnap.org) without understanding the pathological results (Figures 2B, C). The DICOM images and segmented ROIs were subsequently imported into the radiomics software PyRadiomics for feature extraction, resulting in radiomics feature. The extracted radiomics features were normalized to conform to a $N \sim (0, 1)$ distribution (Figure 2D). The Spearman correlation coefficient was employed to assess the correlation between features. For features exhibiting a correlation coefficient greater than 0.9, only one of the correlated features was retained, yielding radiomics feature selection. LASSO regression was applied for cross-validation and to determine the optimal penalty coefficient lambda. Features with a zero coefficient were excluded, and further dimensionality reduction was conducted to derive final radiomics features (Figures 2E, F). The features with non-zero coefficients were aggregated into a formula to compute the final radiomics score (see Supplementary Figure S1). The Random Forest (RF) algorithm was used to build the ultrasound radiomics model (Rad-model).

2.5.3 Construction of the combined model

Based on the Multi-model and the Rad-model, a combined model (Com-model) was established by means of Logistic regression analysis and a nomogram was subsequently plotted.

2.6 Verification and clinical application of the nomogram

Receiver Operating Characteristic (ROC) curves were generated, and the AUC was calculated to assess the discriminatory performance of the model. Calibration curves were plotted, and the Hosmer-Lemeshow goodness-of-fit test was utilized to evaluate the calibration ability of the nomogram. The Brier score was computed to assess the overall performance of the model. Decision Curve Analysis (DCA) was constructed, and the clinical utility of the nomogram was estimated by calculating the net benefit across a range of threshold probabilities.

2.7 Rad-model assisted diagnosis

Fifty thyroid nodule images were randomly selected from enrolled patients and evaluated by 10 junior physicians (ultrasound diagnosis experience ≤ 3 years) and 10 senior physicians (ultrasound diagnosis experience > 10 years) without knowledge of patient information. The 50 cases of thyroid nodules were assessed for benign and malignant conditions, as well as pathological results, and were re-evaluated one week later with the assistance of Rad-model prediction probability. The AUC was calculated to assess the diagnostic performance of both senior and junior physicians before and after the incorporation of ultrasound Rad-model assistance.

2.8 Statistical analyses

All statistical tests were conducted using SPSS 27.0, MedCalc (version 20.100), and R statistical software (version 4.0.2). We convert some continuous variables into categorical variables based on the ROC curve. Independent sample *t*-tests, chi-square tests, or Mann-Whitney *U* tests were employed to compare differences in age, nodule size, and multimodal ultrasound imaging characteristics between the training set and the test set. Logistic regression was utilized to construct Multi-model and Com-model, while the RF was adopted to establish Rad-model. ROC curves for the three models were generated using MedCalc. ROC analysis was performed, employing the Youden index to identify the optimal cut-off value for calculating the AUC, sensitivity, positive predictive value (PPV), negative predictive value (NPV) and accuracy. Internal validation of the models was performed using Bootstrap with 1000 resamplings. DeLong's test was performed to compare the diagnostic performance of the models. A difference was considered statistically significant when $P < 0.05$. We utilized the “rms” and “pec” packages to build nomograms and calibration curves, the “caret” package for bootstrap validation, and the “ramda” and “ggDCA” packages to draw clinical decision curves for analysis using R statistical software.

3 Results

3.1 Characteristics of patients

This study included a total of 446 patients out of 490 patients with thyroid nodules. The set comprised 98 males and 348 females, with an average age of 50.09 ± 12.30 years. Based on the order of enrollment, the patients were divided into a training set ($n=312$) and a test set ($n=134$) in a ratio of 7:3. Within the training set, there were 172 benign cases and 140 malignant cases, while the test set contained 79 benign cases and 55 malignant cases. The maximum diameter of nodules in the training and test set was 10.10 ± 7.60 mm and 10.72 ± 7.30 mm. The baseline characteristics of the patients are presented in Table 1. There was no statistically significant difference between the two groups ($P > 0.05$).

3.2 Construction and performance of the model

3.2.1 The Multi-model

Univariate logistic regression analyses showed that age (≤ 51 years) (OR: 2.863, 95% CI: 1.796-4.563), gender (female) (OR: 0.496, 95% CI: 0.283-0.869), STE Mean (OR: 1.046, 95% CI: 1.024-1.069), STE Max

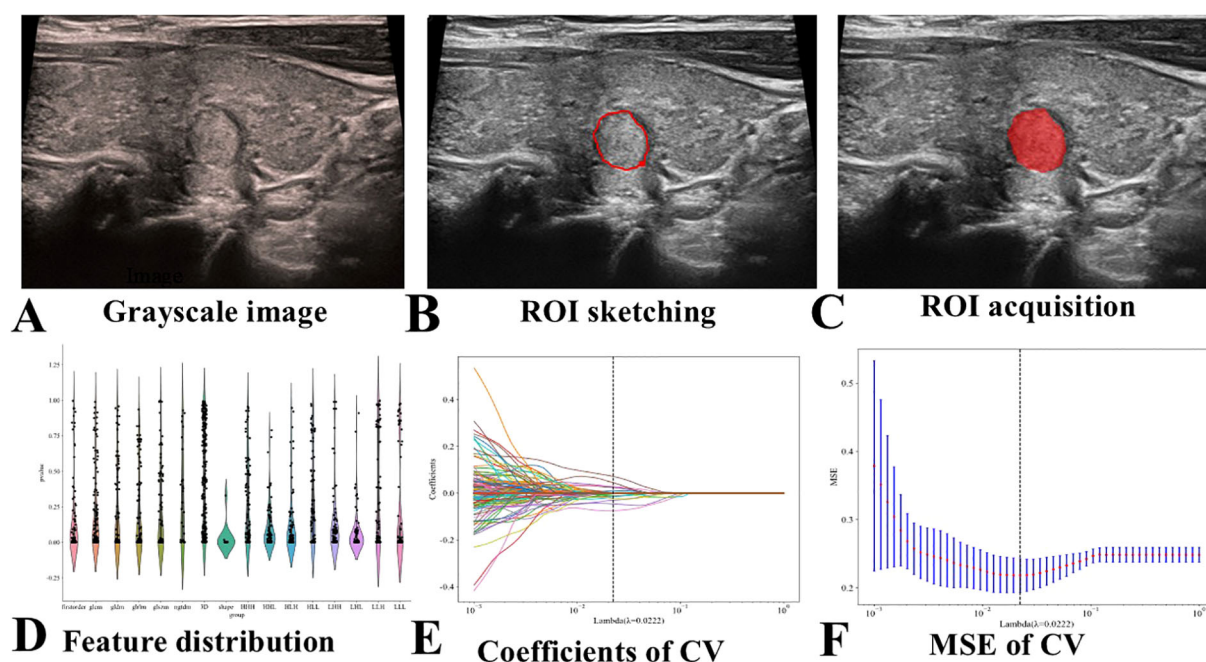


FIGURE 2

Flowchart of Ultrasound Radiomics Model Construction. CV stands for cross-validation, and MSE stands for mean square error. (A) Selection of representative thyroid nodule images. (B) Manual delineation of the ROI. (C) Example of a segmented ROI after manual delineation. (D) Normalization of the extracted radiomics features to conform to a standard normal distribution. (E) Application of LASSO regression for feature selection. (F) Final selection of radiomics features after dimensionality reduction.

TABLE 1 Baseline characteristics of patients in the training set and test set.

Characteristic	Training set (n = 312)	Test set (n = 134)	<i>P</i>
Age (years)	49.70 ± 14.04	50.79 ± 13.07	0.443
Gender			0.166
Male	63 (20.2%)	35 (26.1%)	
Female	249 (79.8%)	99 (73.9%)	
STE Mean (kPa)	32.82 ± 11.70	32.44 ± 11.94	0.445
STE Max (kPa)	63.69 ± 28.62	68.22 ± 38.03	0.476
STE Min (kPa)	14.93 ± 9.67	13.45 ± 8.78	0.353
STE SD (kPa)	9.39 ± 4.97	10.54 ± 6.59	0.391
SE Mean (%)	0.20 ± 0.11	0.19 ± 0.11	0.142
SE Max (%)	0.58 ± 0.70	0.53 ± 0.33	0.161
SE Min (%)	0.05 ± 0.06	0.06 ± 0.07	0.094
SE SD (%)	0.10 ± 0.08	0.10 ± 0.07	0.910
STQ Mean (kPa)	35.98 ± 16.12	34.62 ± 14.31	0.510
STQ Max (kPa)	61.68 ± 34.08	66.85 ± 43.08	0.524
STQ Min (kPa)	19.73 ± 13.39	17.76 ± 11.26	0.499
STQ SD (kPa)	9.43 ± 7.17	10.55 ± 8.97	0.656
Maximum diameter of nodules (mm)	10.10 ± 7.60	10.72 ± 7.30	0.063
Maximum diameter of nodules (<7.4 mm)	161 (51.6%)	57 (42.5%)	0.079
Multiplicity			0.422
solitary	272 (87.2%)	113 (84.3%)	
multiple	40 (12.8%)	21 (15.7%)	
Margin			0.970
smooth	132 (42.3%)	55 (41.0%)	
ill-defined	50 (16.0%)	22 (16.5%)	
irregular	130 (41.7%)	57 (42.5%)	
Halo sign			0.481
absent	288 (92.3%)	121 (90.3%)	
present	24 (7.7%)	13 (9.7%)	
Composition			0.557
cystic-solid	21 (6.7%)	9 (6.8%)	
solid	264 (84.6%)	109 (81.3%)	
predominantly solid	27 (8.7%)	16 (11.9%)	
Echogenicity			0.116
isoechoic	98 (31.4%)	48 (35.8%)	
markedly hypoechoic	45 (14.4%)	11 (8.2%)	
hypoechoic	164 (52.6%)	67 (50.0%)	

(Continued)

TABLE 1 Continued

Characteristic	Training set (n = 312)	Test set (n = 134)	<i>P</i>
Echogenicity			0.116
hyperechoic	1 (0.3%)	5 (3.7%)	
heterogeneous	4 (1.3%)	3 (2.3%)	
Echo texture			0.713
homogeneous	29 (9.3%)	11 (8.2%)	
heterogeneous	283 (90.7%)	123 (91.8%)	
Orientation			0.813
parallel	190 (60.9%)	80 (59.7%)	
vertical	122 (39.1%)	54 (40.3%)	
Calcification			0.774
absent	157 (50.3%)	73 (54.5%)	
coarse calcification	25 (8.0%)	13 (9.7%)	
microcalcification	101 (32.4%)	38 (28.3%)	
indeterminate punctate echogenic foci	28 (9.0%)	10 (7.5%)	
peripheral calcification	1 (0.3%)	0 (0%)	
Relationship to capsule			0.540
>2 mm	118 (37.8%)	48 (36.1%)	
≤2 mm	163 (52.2%)	68 (50.4%)	
extracapsular extension	31 (10.0%)	18 (13.5%)	
CDFI			0.398
0	135 (43.3%)	51 (38.1%)	
I	123 (39.4%)	63 (47.0%)	
II	44 (14.1%)	18 (13.4%)	
III	10 (3.2%)	2 (1.5%)	
Benign and malignant			0.455
benign	172 (55.1%)	79 (59.0%)	
malignant	140 (44.9%)	55 (41.0%)	

Data are expressed as number of patients (%) unless otherwise specified. Percentages may not sum to 100% due to rounding. All discrepancies were cross-verified with raw data and corrected.

(OR: 1.012, 95% CI: 1.004-1.021), STE Min (OR: 1.025, 95% CI: 1.001-1.050), STE SD (OR: 1.064, 95% CI: 1.015-1.115), SE Mean (OR: 0.058, 95% CI: 0.006-0.596), STQ Mean (OR: 1.026, 95% CI: 1.011-1.041), STQ Max (OR: 1.011, 95% CI: 1.003-1.019), STQ SD (OR: 1.064, 95% CI: 1.026-1.103), maximum diameter (≤7.4 mm) (OR: 1.851, 95% CI: 1.177-2.903), margin (ill-defined) (OR: 9.704, 95% CI: 4.584-20.593), margin (irregular) (OR: 13.382, 95% CI: 7.254-24.684), halo sign (present) (OR: 0.224, 95% CI: 0.075-0.670), composition (solid) (OR: 9.792, 95% CI: 2.236-42.882), echogenicity (markedly hypoechoic) (OR: 3.219, 95% CI: 1.861-5.569), echogenicity (hypoechoic) (OR: 4.809, 95% CI: 2.261-10.229), orientation (vertical) (OR: 7.875, 95%

CI: 4.697-13.203), calcification (coarse calcification) (OR: 0.218, 95% CI: 0.085-0.558), calcification (microcalcification) (OR: 0.204, 95% CI: 0.119-0.350) is a risk factor for PTC.

Multivariate logistic regression analysis showed that age (≤ 51 years) (OR: 2.752, 95% CI: 1.546-4.900), STE Mean (OR: 1.036, 95% CI: 1.010-1.062), margin (ill-defined) (OR: 6.187, 95% CI: 2.700-14.178), and margin (irregular) (OR: 7.011, 95% CI: 3.545-13.865), orientation (vertical) (OR: 3.515, 95% CI: 1.926-6.415) were independent risk factors for PTC (Table 2). The AUC of the Multi-model in the training set was 0.852 (95% CI: 0.808-0.890), sensitivity was 82.14%, specificity was 79.65%, PPV was 76.67%, and NPV was 84.57%. In the test set, the AUC of the Multi-model was 0.804 (95% CI: 0.726-0.867), sensitivity was 72.73%, specificity was 79.75%, PPV was 71.43%, and NPV was 80.77% (Figure 3 and Table 3).

3.2.2 The Rad-model

A total of 1,562 features related to radiomics were extracted initially. After applying dimensionality reduction, 17 features were selected (Figure 4), and the Rad-model was developed using the RF algorithm. The AUC of the Rad-model in the training set was 0.940 (95% CI: 0.907-0.963), sensitivity was 88.57%, specificity was 84.30%, PPV was 82.12%, and NPV was 90.06%. In the test set, the AUC of the Rad-model was 0.832 (95% CI: 0.758-0.891), sensitivity was 72.73%, specificity was 86.08%, PPV was 78.43%, and NPV was 81.93% (Figure 3 and Table 3).

3.2.3 The Com-model

The Com-model was developed by integrating the Multi-model and the Rad-model. The AUC of the Com-model in the training set was 0.956 (95% CI: 0.926-0.976), sensitivity was 90.00%, specificity was 97.79%, PPV was 96.92%, and NPV was 92.31%. Internal validation was performed using 1000 bootstrap resampling iterations, resulting in an average AUC of 0.954 for the Com-model. In the test set, the AUC of the Com-model was 0.863 (95% CI: 0.793-0.917), sensitivity was 70.91%, specificity was 89.87%, PPV was 82.98%, and NPV was 81.61% (Figure 3 and Table 3).

3.3 Verification and clinical application of the nomogram

A nomogram was created to estimate the probability of PTC. Using the nomogram derived total score, we stratified patients into low (≤ 54 , malignancy probability $\leq 15\%$), intermediate (54-68, malignancy probability 15-50%), and high-risk (> 68 , malignancy probability $> 50\%$) categories. For high-risk nodules, immediate FNA or surgery is recommended. Intermediate-risk cases may benefit from selective FNA guided by clinical factors, while low-risk nodules warrant surveillance, avoiding unnecessary biopsies. (Figure 5). The calibration plot in Figure 6 illustrates the comparison between the predicted positive rate derived from the

TABLE 2 Results based on univariate and multivariate logistic regression analysis of the training set.

Characteristic	Univariate analysis		Multivariate analysis	
	95% CI	P	95% CI	P
Age ≤ 51 (years)	1.796-4.563	<0.001	1.546-4.900	<0.01
Gender				
Male	Reference			
Female	0.283-0.869	0.014		
STE Mean (kPa)	1.024-1.069	<0.001	1.010-1.062	0.006
STE Max (kPa)	1.004-1.021	0.004		
STE Min (kPa)	1.001-1.050	0.037		
STE SD (kPa)	1.015-1.115	0.009		
SE Mean (%)	0.006-0.596	0.017		
SE Max (%)	0.386-1.246	0.221		
SE Min (%)	0.001-1.044	0.053		
SE SD (%)	0.006-2.622	0.181		
STQ Mean (kPa)	1.011-1.041	<0.001		
STQ Max (kPa)	1.003-1.019	0.004		
STQ Min (kPa)	0.998-1.032	0.091		
STQ SD (kPa)	1.026-1.103	<0.001		
Maximum diameter of nodules (mm)	0.940-1.001	0.057		

(Continued)

TABLE 2 Continued

Characteristic	Univariate analysis		Multivariate analysis	
	95% CI	<i>P</i>	95% CI	<i>P</i>
Gender				
Maximum diameter of nodules (<7.4 mm)	1.177-2.903	0.008		
Multiplicity				
solitary	Reference			
multiple	0.405-1.564	0.508		
Margin				
smooth	Reference			
ill-defined	4.584-20.539	<0.001	2.700-14.178	<0.01
irregular	7.254-24.684	<0.001	3.545-13.865	<0.01
halo sign				
absent	Reference			
present	0.075-0.670	0.007		
Composition				
cystic-solid	Reference			
solid	2.236-42.882	0.002		
predominantly solid	0.272-10.024	0.585		
Echogenicity				
isoechoic	Reference			
markedly hypoechoic	1.861-5.569	<0.001		
hypoechoic	2.261-10.229	<0.001		
hyperechoic	0	1.000		
heterogeneous	0.097-9.789	0.982		
Echo texture				
homogeneous	Reference			
heterogeneous	0.843-4.351	0.121		
Orientation				
parallel	Reference			
vertical	4.697-13.203	<0.001	1.926-6.415	<0.01
Calcification				
absent	Reference			
coarse calcification	0.085-0.558	0.001		
microcalcification	0.119-0.350	<0.001		
indeterminate punctate echogenic foci	0.228-1.256	0.151		
peripheral calcification	0	1.000		
Relationship to capsule				
>2 mm	Reference			

(Continued)

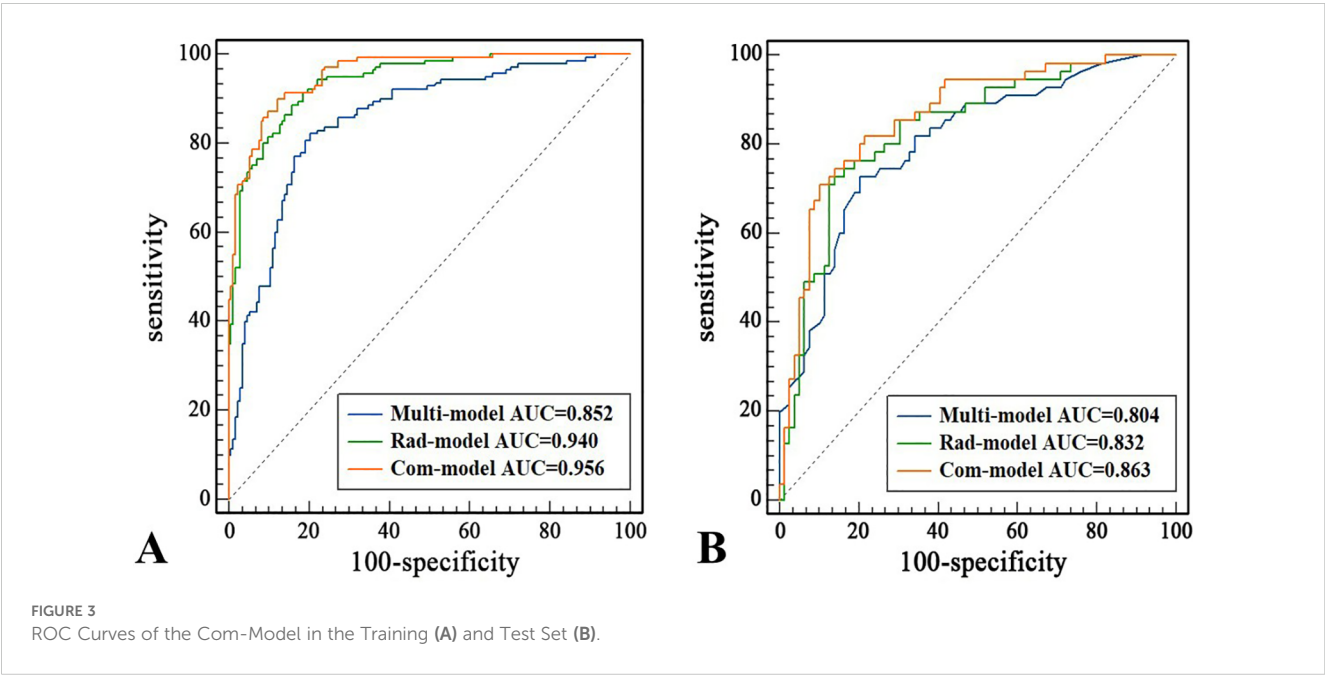
TABLE 2 Continued

Characteristic	Univariate analysis		Multivariate analysis	
	95% CI	<i>P</i>	95% CI	<i>P</i>
Relationship to capsule				
≤2 mm	0.522-1.355	0.476		
extracapsular extension	0.553-2.697	0.620		
CDFI				
0	Reference			
I	0.607-1.615	0.968		
II	0.282-1.166	0.125		
III	0.118-1.917	0.296		

nomogram and the actual observations. According to the Hosmer-Lemeshow goodness-of-fit test, both the training set ($P = 0.913$) and the test set ($P = 0.854$) show a strong fit. The Brier scores for the training and test sets are 0.08 and 0.15, respectively, indicating that our prediction model demonstrates overall robustness. The DCA shows that the model offers considerable advantages within the range of 0.15 to 0.90 (Figure 7). To further assess the effectiveness of the models, we performed statistical comparisons of the ROC curves using DeLong’s test (Table 4). The findings indicate that in the training set, the Com-model was the most effective. In the test set, the Com-model surpassed the Multi-model, with no significant difference noted when compared to the ultrasound Rad-model.

3.4 Effectiveness of the Rad-model assistance

Before utilizing the Rad-model for diagnostic assistance, the AUC for junior physicians was 0.748, while that for senior physicians was 0.837 (Table 5 and Figure 8). Following the implementation of the Rad-model, the average AUC for junior physicians significantly increased to 0.851, and for senior physicians, it improved to 0.862. DeLong’s test indicated a significant enhancement in the diagnostic performance of junior physicians after model assistance (<0.001), suggesting that the Rad-model effectively shortened their learning curve. Furthermore, there was no significant difference in AUC between senior and junior



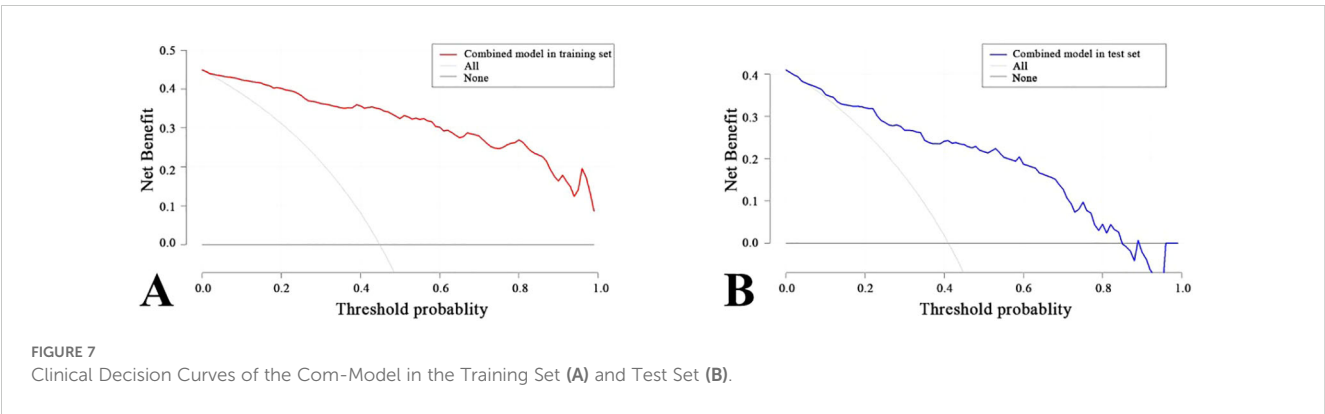
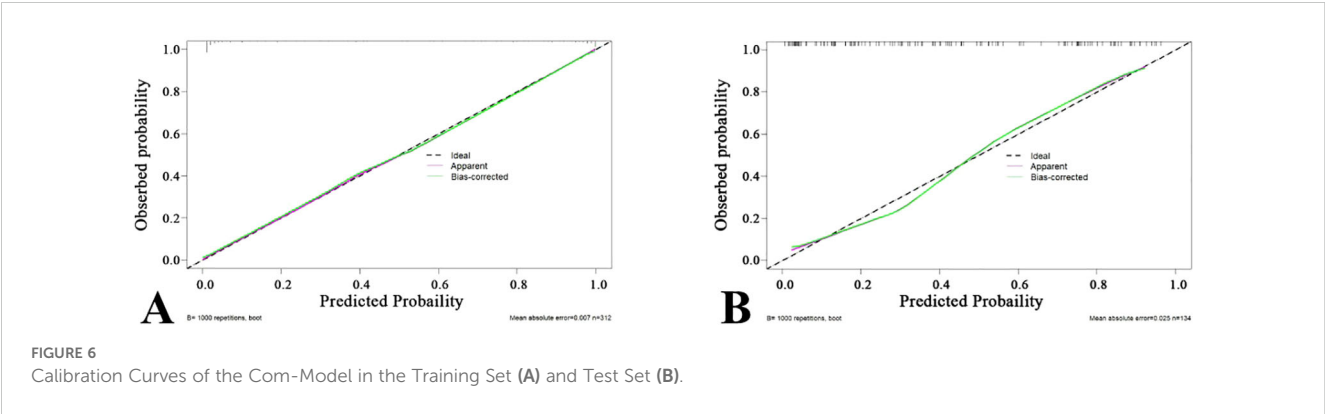
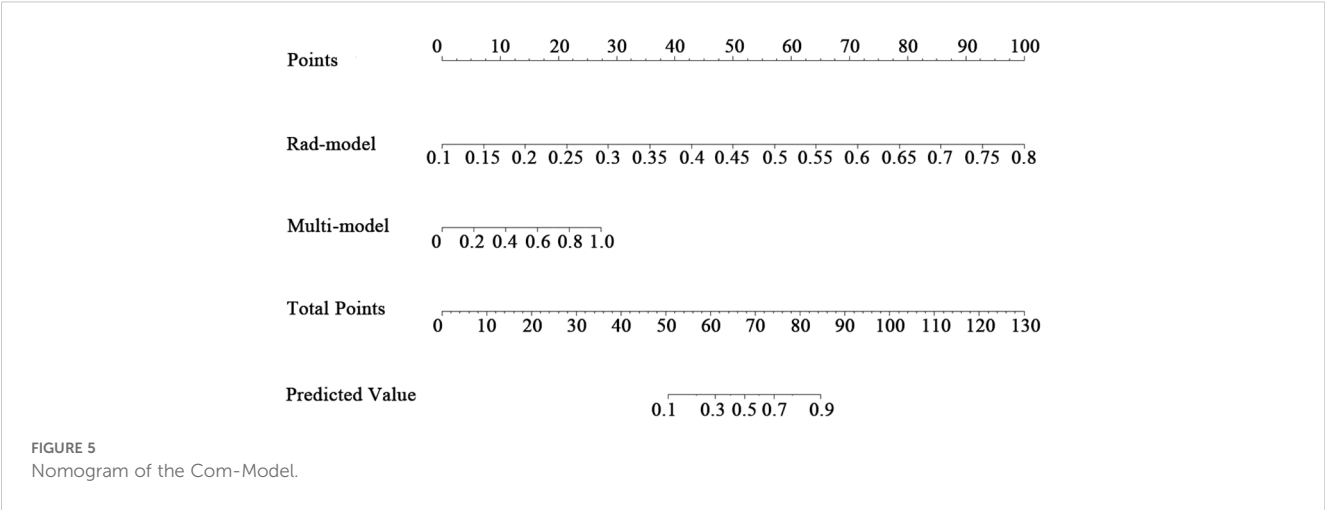


TABLE 4 Delong test of three models.

Variables	The Multi-model - The Rad-model		The Rad-model - The Com-model		The Com-model - The Multi-model	
	Training set	Test set	Training set	Test set	Training set	Test set
Z	3.64	0.65	2.42	1.55	5.34	2.16
P	<0.001	0.52	0.02	0.12	<0.001	0.03

TABLE 5 The performance of Pre- and Post-diagnostic Radiomic Model Assistance of junior and senior physicians.

Variables	Pre-diagnostic radiomic model assistance	Post-diagnostic radiomic model assistance
junior physicians (AUC)	0.748	0.851
Sensitivity (%)	80.12	85.63
Specificity (%)	82.71	87.24
PPV (%)	78.97	86.61
NPV (%)	83.82	86.98
senior physicians (AUC)	0.837	0.862
Sensitivity (%)	68.23	85.54
Specificity (%)	76.31	85.47
PPV (%)	71.43	83.02
NPV (%)	73.58	84.92

nodules. Specifically, by integrating ultrasound elastography with various other ultrasound features, the model developed in this study not only improves the accuracy of thyroid nodule characterization but also enhances the reliability of clinical decision-making, helping to reduce unnecessary invasive procedures.

Vertical orientation (where the longitudinal growth of the nodule exceeds its transverse growth) is considered an independent risk factor for malignant thyroid nodules (19, 20). From a histological perspective, this phenomenon may be attributed to the active division of tumor cells in the anterior-posterior direction within malignant nodules, while remaining relatively quiescent in other directions. This aggressive growth pattern aims to increase the tumor’s surface area, thereby facilitating more efficient nutrient acquisition and accelerating its growth process.

Additionally, blurred or irregular margins are one of the critical imaging features for assessing the nature of thyroid nodules (21). Studies have shown that the BRAF V600E mutation is the most prevalent genetic alteration in PTC, present in approximately 60–80% of cases. This mutation constitutively activates the MAPK/ERK

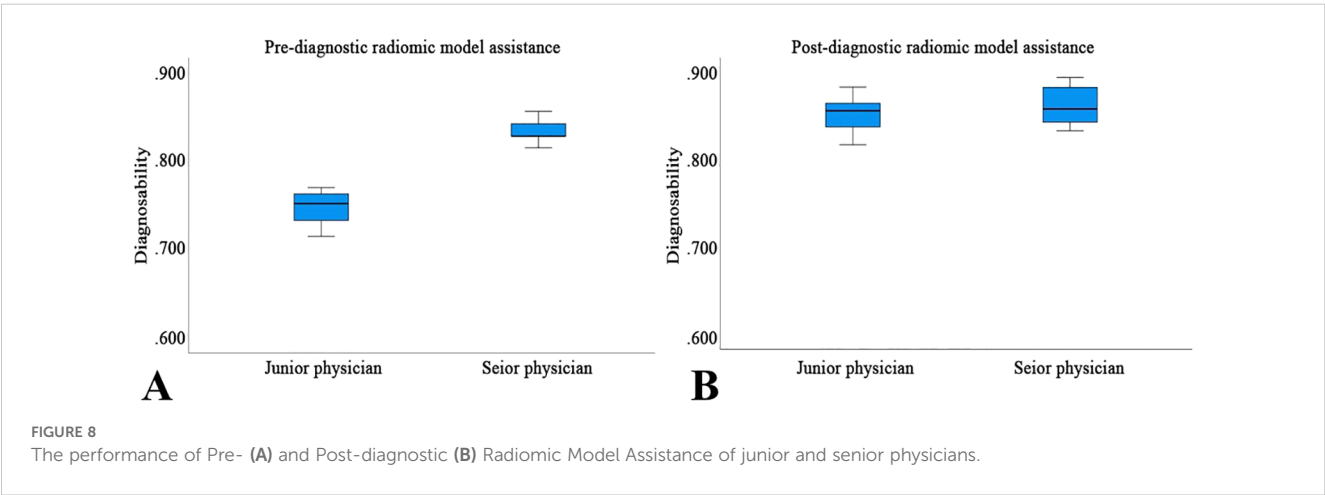
TABLE 6 The DeLong test for comparing Pre- and Post-diagnostic Radiomic Model Assistance of junior and senior physicians.

Variables	Pre- and Post-diagnostic		Post-diagnostic
	junior physicians	senior physicians	junior-senior physicians
Z	3.12	1.72	0.65
P	<0.001	0.08	0.51

signaling pathway, promoting uncontrolled cell proliferation and tumorigenesis and promotes the invasiveness and metastatic potential of tumor cells, thereby influencing the morphological appearance of the nodules. Specifically, it leads to blurred, jagged, or spiculated margins, which are often indicative of malignancy. The BRAF V600E mutation has been confirmed to be associated with the high aggressiveness and poor prognosis of PTC (21, 22).

The Com-model achieved an AUC of 0.863 in the test set, outperforming standalone C-TIRADS 4A classification [reported AUC: 0.70-0.82 in prior studies (5, 6)] and ACR TI-RADS [AUC: 0.76-0.85 (5, 18)]. For instance, Zhang et al. (18) reported an AUC of 0.834 for ACR TI-RADS in differentiating C-TIRADS 4A nodules, while our Com-model achieved higher specificity (89.87% vs. 76.5%) and comparable sensitivity (70.91% vs. 72.1%). Notably, the integration of radiomics and multimodal ultrasound features enabled our model to capture subtle malignant characteristics (e.g., microstructural heterogeneity) that conventional systems may overlook. Compared to AI-driven approaches, such as TNet [AUC: 0.865 (23)] and the CNN-based framework by Tao et al. [AUC: 0.872 (16)], our Com-model demonstrated superior generalizability across multicenter data. However, unlike deep learning models requiring large annotated datasets, our radiomics framework relies on interpretable handcrafted features, aligning better with clinical workflows.

This study compared the diagnostic performance of senior and junior physicians in differentiating thyroid nodules before and after the application of the Rad-model. Although junior physicians lack extensive experience in ultrasound diagnosis, their diagnostic accuracy significantly improved with the support of the Rad-model, reaching a level comparable to that of senior physicians. This not only greatly shortened the learning curve for junior



physicians, accelerating their path to professional proficiency, but also enhanced the overall efficiency and accuracy of thyroid nodule diagnosis. These findings suggest that the Rad-model can effectively address the challenges posed by limited clinical experience, providing robust technical support for younger physicians and thereby improving the overall quality of healthcare services.

The Com-model exhibited a false-negative rate of 10.00% in the training set and 29.09% in the test set, indicating a non-negligible risk of missing malignant cases, particularly in test set. While this rate aligns with prior studies [e.g., Zhang et al. (24)]. To mitigate the risks of false-negative results, we advocate for a tiered follow-up protocol combining short-interval ultrasound surveillance and clinical risk stratification: 6 months follow-up for intermediate-risk cases and 12 months reassessment for low-risk cases. Future iterations of the model will integrate dynamic imaging biomarkers and molecular testing to further reduce FNR, ensuring early intervention for initially missed malignancies.

Future research will focus on integrating molecular and biochemical profiling (25) with our imaging-based model to achieve a holistic assessment of thyroid nodules. For instance, combining radiomic heterogeneity with BRAF V600E or TERT promoter mutations could improve the identification of aggressive PTC subtypes, while metabolic markers of oxidative stress may refine angioinvasion risk prediction (26). Such multi-omics fusion aligns with the goals of precision oncology, enabling tailored surveillance and treatment strategies for borderline or ambiguous nodules. Additionally, liquid biopsy-derived biomarkers (e.g., ctDNA) (27) could complement ultrasound surveillance by providing real-time molecular insights into nodule dynamics. This approach may bridge the gap between static imaging assessments and the evolving biological behavior of thyroid malignancies.

To ensure consistency across multiple centers, all participating centers use the same ultrasound system and follow standardized imaging protocols. In the future, we will collaborate with medical institutions using different ultrasound systems (such as GE Logiq, Philips EPIQ, Siemens Acuson) to collect external datasets to ensure their applicability in the real world.

This study has several notable limitations. First, while it was impossible to completely eliminate all subjective factors during the analysis of multimodal ultrasound features, we addressed this by involving multiple evaluators and applying consistency tests to correct for potential subjective biases. Second, some cases relied on cytopathological reports, which inherently carry a certain rate of false negatives. To mitigate this issue, we incorporated BRAF gene testing results and conducted a 6 months follow-up observation of patients to more accurately assess the nature of the nodules.

5 Conclusion

In summary, we successfully developed a Com-model that integrates the Multi-model with the Rad-model to differentiate between benign and malignant thyroid nodules. The results demonstrate that this Com-model exhibits significant advantages in

distinguishing C-TIRADS 4A thyroid nodules, providing physicians with a rapid and accurate risk assessment tool. It effectively identifies potential malignant lesions while reducing unnecessary invasive examinations or treatments for patients with benign nodules.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by The Ethics Committee of the First Affiliated Hospital of Henan University of Science and Technology 2024-03-K0160. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SC: Writing – original draft, Writing – review & editing. QL: Writing – original draft. HW: Data curation, Writing – original draft. HL: Investigation, Writing – original draft. WL: Data curation, Writing – original draft. CL: Data curation, Writing – original draft. LB: Data curation, Writing – original draft. YM: Data curation, Writing – original draft. WG: Software, Writing – original draft. JY: Conceptualization, Project administration, Supervision, Writing – review & editing. ZZ: Formal Analysis, Data curation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Key Project of Luoyang Science and Technology Development Plan (Grant No. 2401209B).

Acknowledgments

We would like to acknowledge all participants in this study for their efforts.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1543020/full#supplementary-material>

References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2024) 74:229–63. doi: 10.3322/caac.21834
- Seib CD, Sosa JA. Evolving understanding of the epidemiology of thyroid cancer. *Endocrinol Metab Clin North Am.* (2019) 48:23–35. doi: 10.1016/j.ecl.2018.10.002
- Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European thyroid association guidelines for ultrasound Malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J.* (2017) 6:225–37. doi: 10.1159/000478927
- Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. 2020 Chinese guidelines for ultrasound Malignancy risk stratification of thyroid nodules: the C-TIRADS. *Endocrine.* (2020) 70:256–79. doi: 10.1007/s12020-020-02441-y
- Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol.* (2017) 14:587–95. doi: 10.1016/j.jacr.2017.01.046
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid.* (2016) 26:1–133. doi: 10.1089/thy.2015.0020
- Gong HY, Fu XF, He MX, Zhu J. Combination of two-dimensional ultrasound and contrast-enhanced ultrasound in the diagnosis of benign and Malignant thyroid micro-nodules. *Asian J Surg.* (2022) 45:1172–3. doi: 10.1016/j.asjsur.2022.01.091
- Nattabi HA, Sharif NM, Yahya N, Ahmad R, Mohamad M, Zaki FM, et al. Is diagnostic performance of quantitative 2D-shear wave elastography optimal for clinical classification of benign and Malignant thyroid nodules? A systematic review and meta-analysis. *Acad Radiol.* (2022) Suppl 3:S114–21. doi: 10.1016/j.acra.2017.09.002
- Zuo D, Yang L, Jin Y, Qi H, Liu Y, Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med Inform Decis Mak.* (2023) 23:276. doi: 10.1186/s12911-023-02377-z
- Din NMU, Dar RA, Rasool M, Assad A. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Comput Biol Med.* (2022) 149:106073. doi: 10.1016/j.combiomed.2022.106073
- Gao Q, Yang L, Lu M, Jin R, Ye H, Ma T. The artificial intelligence and machine learning in lung cancer immunotherapy. *J Hematol Oncol.* (2023) 16:55. doi: 10.1186/s13045-023-01456-y
- Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective. *Semin Cancer Biol.* (2023) 89:30–7. doi: 10.1016/j.semcancer.2023.01.006
- Wen SS, Wu YJ, Wang JY, Ni ZX, Dong S, Xie XJ, et al. BRAFV600E/p-ERK/p-DRP1(Ser616) promotes tumor progression and reprogramming of glucose metabolism in papillary thyroid cancer. *Thyroid.* (2024) 34:1246–59. doi: 10.1089/thy.2023.0700
- Wang L, Wang C, Deng X, Li Y, Zhou W, Huang Y, et al. Multimodal ultrasound radiomic technology for diagnosing benign and Malignant thyroid nodules of T1-Rads 4-5: A multicenter study. *Sensors (Basel).* (2024) 24:6203. doi: 10.3390/s24196203
- Liu F, Liu D, Wang K, Xie X, Su L, Kuang M, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. *Liver Cancer.* (2020) 9:397–413. doi: 10.1159/000505694
- Tao Y, Yu Y, Wu T, Xu X, Dai Q, Kong H, et al. Deep learning for the diagnosis of suspicious thyroid nodules based on multimodal ultrasound images. *Front Oncol.* (2022) 12:1012724. doi: 10.3389/fonc.2022.1012724
- Luo J, Chen J, Sun Y, Xu F, Wu L, Huang P. A retrospective study of reducing unnecessary thyroid biopsy for American College of Radiology Thyroid Imaging Reporting and Data Systems 4 assessment through applying shear wave elastography. *Arch Endocrinol Metab.* (2020) 64:349–55. doi: 10.20945/2359-3997000000267
- Zhang WB, Xu W, Fu WJ, He BL, Liu H, Deng WF. Comparison of ACR TI-RADS, Kwak TI-RADS, ATA guidelines and KTA/KSThR guidelines in combination with SWE in the diagnosis of thyroid nodules. *Clin Hemorheol Microcirc.* (2021) 78:163–74. doi: 10.3233/CH-201021
- Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health.* (2021) 3:e250–9. doi: 10.1016/S2589-7500(21)00041-8
- Xu C, Fang J, Li W, Sun C, Li Y, Lowe S, et al. Construction and validation of BRAF mutation diagnostic model based on ultrasound examination and clinical features of patients with thyroid nodules. *Front Genet.* (2022) 13:973272. doi: 10.3389/fgene.2022.973272
- Chuanji Z, Zheng W, Shaolv L, Linghou M, Yixin L, Xinhui L, et al. Comparative study of radiomics, tumor morphology, and clinicopathological factors in predicting overall survival of patients with rectal cancer before surgery. *Transl Oncol.* (2022) 18:101352. doi: 10.1016/j.tranon.2022.101352
- Trimboli P, Scappaticcio L, Treglia G, Guidobaldi L, Bongiovanni M, Giovanella L, et al. (V600E) mutation in thyroid nodules with fine-needle aspiration (FNA) read as suspicious for Malignancy (Bethesda V, Thy4, TIR4): a systematic review and meta-analysis. *Endocr Pathol.* (2020) 31:57–66. doi: 10.1007/s12022-019-09596-z
- Zhu YC, AlZoubi A, Jassim S, Jiang Q, Zhang Y, Wang YB, et al. A generic deep learning framework to classify thyroid and breast lesions in ultrasound images. *Ultrasonics.* (2021) 110:106300. doi: 10.1016/j.ultras.2020.106300
- Zhang B, Tian J, Pei S, Chen Y, He X, Dong Y, et al. Machine learning-assisted system for thyroid nodule diagnosis. *Thyroid.* (2019) 29:858–67. doi: 10.1089/thy.2018.0380
- Antony V, Sun T, Dolezal D, Cai G. Comprehensive molecular profiling of metastatic pancreatic adenocarcinomas. *Cancers (Basel).* (2025) 17:335. doi: 10.3390/cancers17030335
- Buczyńska A, Kościuszko M, Krętowski AJ, Popławska-Kita A. Exploring the clinical utility of angiogenesis markers in papillary thyroid cancer: a literature review. *Front Endocrinol (Lausanne).* (2023) 14:1261860. doi: 10.3389/fendo.2023.1261860
- Luo H, Wei W, Ye Z, Zheng J, Xu RH. Liquid biopsy of methylation biomarkers in cell-free DNA. *Trends Mol Med.* (2021) 27:482–500. doi: 10.1016/j.molmed.2020.12.011