

OPEN ACCESS

EDITED BY

Horatiu Silaghi,
University of Medicine and Pharmacy Iuliu
Hatieganu, Romania

REVIEWED BY

Jeehee Yoon,
Chonnam National University Bitgoeul
Hospital, Republic of Korea
Aixia Sun,
Michigan State University, United States

*CORRESPONDENCE

Ying Huang
✉ huangying712@163.com

RECEIVED 22 September 2023

ACCEPTED 21 March 2024

PUBLISHED 03 April 2024

CITATION

Chen J-h, Zhang Y-Q, Zhu T-t, Zhang Q,
Zhao A-x and Huang Y (2024) Applying
machine-learning models to differentiate
benign and malignant thyroid nodules
classified as C-TIRADS 4 based on 2D-
ultrasound combined with five contrast-
enhanced ultrasound key frames.
Front. Endocrinol. 15:1299686.
doi: 10.3389/fendo.2024.1299686

COPYRIGHT

© 2024 Chen, Zhang, Zhu, Zhang, Zhao and
Huang. 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.

Applying machine-learning models to differentiate benign and malignant thyroid nodules classified as C-TIRADS 4 based on 2D-ultrasound combined with five contrast-enhanced ultrasound key frames

Jia-hui Chen, Yu-Qing Zhang, Tian-tong Zhu, Qian Zhang, Ao-xue Zhao and Ying Huang*

Department of Ultrasound, Shengjing Hospital of China Medical University, Shenyang, China

Objectives: To apply machine learning to extract radiomics features from thyroid two-dimensional ultrasound (2D-US) combined with contrast-enhanced ultrasound (CEUS) images to classify and predict benign and malignant thyroid nodules, classified according to the Chinese version of the thyroid imaging reporting and data system (C-TIRADS) as category 4.

Materials and methods: This retrospective study included 313 pathologically diagnosed thyroid nodules (203 malignant and 110 benign). Two 2D-US images and five CEUS key frames ("2nd second after the arrival time" frame, "time to peak" frame, "2nd second after peak" frame, "first-flash" frame, and "second-flash" frame) were selected to manually label the region of interest using the "Labelme" tool. A total of 7 images of each nodule and their annotations were imported into the Darwin Research Platform for radiomics analysis. The datasets were randomly split into training and test cohorts in a 9:1 ratio. Six classifiers, namely, support vector machine, logistic regression, decision tree, random forest (RF), gradient boosting decision tree and extreme gradient boosting, were used to construct and test the models. Performance was evaluated using a receiver operating characteristic curve analysis. The area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy (ACC), and F1-score were calculated. One junior radiologist and one senior radiologist reviewed the 2D-US image and CEUS videos of each nodule and made a diagnosis. We then compared their AUC and ACC with those of our best model.

Results: The AUC of the diagnosis of US, CEUS and US combined CEUS by junior radiologist and senior radiologist were 0.755, 0.750, 0.784, 0.800, 0.873, 0.890, respectively. The RF classifier performed better than the other five, with an AUC of 1 for the training cohort and 0.94 (95% confidence interval 0.88–1) for the test cohort. The sensitivity, specificity, accuracy, PPV, NPV, and F1-score of the RF model in the test cohort were 0.82, 0.93, 0.90, 0.85, 0.92, and 0.84, respectively. The RF model with 2D-US combined with CEUS key frames achieved equivalent performance as the senior radiologist (AUC: 0.94 vs. 0.92, $P = 0.798$; ACC: 0.90

vs. 0.92) and outperformed the junior radiologist (AUC: 0.94 vs. 0.80, $P = 0.039$, ACC: 0.90 vs. 0.81) in the test cohort.

Conclusions: Our model, based on 2D-US and CEUS key frames radiomics features, had good diagnostic efficacy for thyroid nodules, which are classified as C-TIRADS 4. It shows promising potential in assisting less experienced junior radiologists.

KEYWORDS

thyroid nodules, ultrasound, contrast-enhanced ultrasound, machine learning, radiomics features, key frames, radiologists

1 Introduction

Thyroid nodules are a common clinical condition. In recent decades, the use of high-resolution ultrasound has rapidly increased worldwide (1, 2). The detection rate of thyroid nodules can reach 67%; however, only 5–15% of them are malignant (3, 4). In clinical practice, many patients suffer some complications after surgical thyroidectomy (5, 6). Moreover, the status quo of overdiagnosis and overtreatment has added unnecessary burdens to patients. In 2020, Chinese experts developed the Chinese version of the thyroid imaging reporting and data system (C-TIRADS) to evaluate the characteristics of thyroid nodules, providing a more practical and concise tool for daily clinical practice (7). Most nodules classified as C-TIRADS 3 or 5 can be quickly distinguished accurately using two-dimensional ultrasound (2D-US) alone; however, there is a wide range of malignancy rates among thyroid nodules classified as C-TIRADS 4 (2–90%). Moreover, some hypoechoic Hashimoto nodules with blurred margins can be classified as C-TIRADS 4 (8). and mummified nodules with internal necrotic components may also exhibit marked hypoechogenicity (9). Distinguishing these from malignant nodules poses challenges, leading to the low specificity of 2D-US and warranting fine needle aspiration (FNA), an invasive procedure (2). Thus, there is a need to explore new methods for a more precise diagnosis of thyroid nodules which are classified as C-TIRADS 4.

Contrast-enhanced ultrasound (CEUS), which describes focal microcirculation perfusion status by distinguishing acoustic features of tissue backgrounds, plays an essential role in the diagnosis of thyroid nodules and differentiation of necrotic benign nodules from malignant ones to avoid FNA procedures (10). Additionally, CEUS is utilized in the field of interventional ultrasonography, which includes assisting biopsy and FNA procedures and estimating therapeutic conditions after ablation (11, 12). Despite not being recommended as part of the guidelines for diagnosing thyroid nodules, numerous studies have demonstrated that CEUS exhibits a sensitivity and specificity of discriminating malignant nodules from benign nodules that could reach 0.87 and 0.83, respectively (13, 14). The consensus on the

qualitative and quantitative analysis of CEUS recommends that malignant characteristics include later wash-in, heterogeneous hypoenhancement, earlier wash-out, and centripetal perfusion (15–17). Machine learning (ML) is an algorithm based on representational learning of data, except for computer vision, natural language processing, and speech recognition, and has played a prominent role in the medical field (18–21). ML can significantly limit interobserver variations (22). With the rapid development of artificial intelligence (AI), radiomics has recently attracted the attention of researchers. Radiomics can transform pixels in medical images into high-dimensional features and quantitative data that can be calculated, which could show intratumor heterogeneity and texture features (23, 24). ML algorithms can be used to develop predictive models and calculate their performances. In the field of thyroid nodules, ML is mostly based on 2D-US images, with an accuracy (ACC) of approximately 0.88–0.92 (25, 26). To our knowledge, only two studies have used CEUS images to build AI models for diagnosing thyroid nodules (27, 28). Wan et al. used deep learning (DL) to build a diagnostic model based on dynamic CEUS video and obtained relatively high performance (27). Guo et al. used logistic regression to build ML models based on US and CEUS features, while only included a single frame of CEUS images (28). Our study aimed to explore the useful information of CEUS images for diagnosing C-TIRADS 4 thyroid nodules. Herein, we combined 2D-US with five CEUS key frames as an import for further radiomics feature extraction and ML model development, aimed at examining the value of ML model based on 2D-US and CEUS key frames in the differential diagnosis of benign and malignant nodules which are classified as C-TIRADS 4.

2 Materials and methods

2.1 Patients

This retrospective study was conducted between September 2019 and February 2023. Data from 313 thyroid nodules in 300

patients which underwent FNA or thyroid surgery at our hospital were included in this study. The inclusion criteria were: (1) patients aged ≥ 18 years; (2) nodule classified as C-TIRADS 4 (with at least one malignant sign); (3) some suspicious malignant nodules that needed CEUS examination to exclude mummified nodules before FNA, and some cystic-solid nodules which were classified as C-TIRADS 3 but the most component were the solid and were eccentric distribution; (4) CEUS examination procedures that contained “double-flash” at 40s and 60s, respectively; and (5) patients who signed an informed consent form and obtain pathological results of thyroid nodules after the CEUS examination. The exclusion criteria were: (1) allergy to any of the components in the ultrasound contrast agent; (2) nodules with macrocalcification during B-mode ultrasound examination; (3) FNA pathological results incomplete or categorized as Bethesda I, III, and IV; and (4) CEUS videos with severe motion. The patients were separated at a ratio of 9:1. Our study was approved by Medical Ethics Committee of Shengjing Hospital of China Medical University (2023PS967K). The PASS.15 software (NCSS LLC, Kaysville, UT, USA) was used to calculate the sample size, with parameters set to ensure the power of 0.90 and level α was set at bilateral 0.05. Based on our expected results, the receiver operating characteristic (ROC) curve was set to 0.90. The false-positive rate was limited from 0 to 1. The group allocation was set at 2. The number of nodules included in the training cohort was 144 in the malignant group and 72 in the benign group (total = 216), with an additional 10% for dropouts. Hence, the final result was 158 and 80 nodules in the malignant and benign groups, respectively (total = 238).

2.2 US, CEUS examinations and images selection

An L14-3U transducer (frequency: 3–9 MHz) from the Resona 9 device (Mindray, Shenzhen, China) and an L12-5 transducer (frequency: 5–12 MHz) from the iU22 device (Philips, Amsterdam, The Netherlands) were used. 2D-US was performed by two radiologists, one with 3 years of experience in thyroid ultrasound and the other with >10 years of experience in thyroid ultrasound. We measured the thyroid size, nodule numbers, nodule size, nodule location, component, echogenicity, shape, margin, and the presence or absence of Hashimoto's background and microcalcification. We then recorded following the C-TIRADS guidelines. In patients with multiple nodules, the ones most suspicious for malignancy were selected for observation and subsequent CEUS examination. The C-TIRADS classification was recorded, and nodules with inconsistent C-TIRADS results were reevaluated and decided upon. Subsequently, CEUS was performed by an experienced radiologist, who then selected the largest section of the nodule, including the surrounding normal thyroid tissue. The mechanical index was set to 0.06–0.08, and the gain, depth, acoustic window, and focal zone were adjusted. The probe stabilized, and the CEUS mode was initiated. For this procedure, 59 mg of contrast agent (SonoVue; Bracco, Milan, Italy) was mixed with 5 mL of saline to prepare a suspension. The suspension (1.5 mL) was injected rapidly through the superficial

vein of the elbow, followed by a 5 mL saline flush. The timer was started simultaneously with the time of injection. The term “flash” means when the microbubbles had been blown up, the remaining microbubbles would reperfuse after the “flash” without the bolus's influence, making good efforts to observe reperfusion status. The radiologist pressed the contrast agent click-button in the 40th and 60th seconds, defined as “first-flash” and “second-flash,” respectively. The entire dynamic recording lasted 80 seconds and was recorded in “AVI” format. Two experienced radiologists immediately diagnosed patients. CEUS observation parameters, including wash-in pattern (earlier, synchronous, and later), enhanced intensity (hypo-, iso-, and hyperintensity), enhanced homogeneity (homo- and heterogeneous), enhanced method (centripetal and centrifugal), and wash-out pattern (earlier, synchronous, and later), were recorded. The nodules with inconsistent results were examined and discussed. According to the previous studies (17, 29–32), nodules with “later wash-in”, “heterogeneous hypointensity”, “centripetal enhancement” and “earlier wash-out” were malignant parameters for thyroid nodules. In our study, we defined nodules with at least two of the among parameters as malignant nodules, the others were defined as benign nodules.

Furthermore, the nodule's largest transverse and longitudinal sections were selected in 2D-US after rotating the probe 90° clockwise. Regarding CEUS, the perfusion of the contrast agents gradually changes with changes in brightness during CEUS examinations, which could reveal the blood supply of the nodule. Many previous studies have also suggested wash-in or -out patterns of contrast agents, and the enhanced intensity in the nodule area compared to the surrounding normal thyroid tissues was the most helpful parameter for diagnosing malignant nodules (11, 31, 33, 34). The “double-flash,” identified as a new CEUS quantitative parameter in our previous study, indicated that the diagnostic accuracy in distinguishing malignant and benign thyroid nodules could reach 88.4% (24). Therefore, based on these principles and results, five CEUS key frames were finally selected: the “2nd second after the arrival time” frame, “time to peak” frame, “2nd second after peak” frame, “first-flash” frame, and “second-flash” frame.

2.3 Nodule segmentation

The 80-second CEUS video of each patient was converted to 1120 images (14 images every second) using Python code. One radiologist (with 3 years of CEUS experience) browsed the images and found five key CEUS frames. The radiologist manually delineated the boundary of the region of interest (ROI) on seven images (two from 2D-US and five from CEUS key frames) using “Labelme” in an Anaconda (<http://anaconda.org>) environment. The second radiologist (with 8 years of CEUS experience) checked the segmentations. If there were any inconsistencies, the results were jointly discussed, and further modifications were made until a consensus was reached. Finally, the patient images and labels were imported into the Darwin Research Platform (<https://arxiv.org/abs/2009.00908>) for feature extraction and model establishment. The workflow scheme is illustrated in [Figure 1](#). The nodule segmentation process is described in the [Supplementary Materials](#).

2.4 Feature extraction and selection

After nodule segmentation, feature extraction was performed using the “PyRadiomics” package for Python (Python Software Foundation, Beaverton, OR, USA). Radiomics features include first-order, shape, and texture. First-order features can be obtained using a simple metric procedure to clarify the distribution of voxel intensities, such as mean range, variance, and kurtosis. Texture features are used to describe the heterogeneity of the lesion, including the gray-level cooccurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level dependence matrix (GLDM), neighboring gray-tone difference matrix (NGTDM), and gray-level size zone matrix (GLSZM). Eight kinds of filters were applied in our study to transform the original images: exponential, gradient, local binary pattern- two dimensional (Lbp-2D), logarithm, square, square root, wavelet, and Laplacian of Gaussian (LoG). First-order shape and texture features were extracted from the derived images. However, since a single image contained 1125 features, seven images from one patient produced 7875 features in total. We extracted all features and subsequently selected them. Feature selection is an important ML procedure because it reduces computational complexity and trains classifiers more accurately. Maximum absolute normalization was used to scale the numerical value to the unit length within a range of -1 to 1 . The variance threshold can remove all low-variance features. To reduce overfitting and find definitive correlation features, only F values equal to 0 were excluded from this study. The classifiers also contain algorithms that iteratively calculate the importance of the features. Finally, the decision tree (DT) classifier was used to determine the most relevant feature rankings (Figure 2).

2.5 Model development

Six ML models, namely support vector machine (SVM), logistic regression (LR), DT, random forest (RF), gradient boosting decision tree (GBDT), and extreme gradient boosting (XGBOOST) were used to determine the best diagnostic performance. The radial basis function was used in the SVM classifier, and the penalty coefficient C was used

to set the tolerance for misclassified samples (from 0.0001 to 1,000). LR was based on an elastic net, and the I1 ratio was set to 0.5. For RF, DT, GBDT, and XGBOOST, the maximum depth of the tree was set to 5 to avoid overfitting. If values were missing, we chose the mean value as a supplement. The 10-fold crossvalidation was used to inspect the accuracy of the models. The ROC curve and area under the curve (AUC) were used to compare the performance of the six ML models, and the sensitivity, specificity, accuracy, F1-score, positive predictive value (PPV), and negative predictive value (NPV) were calculated.

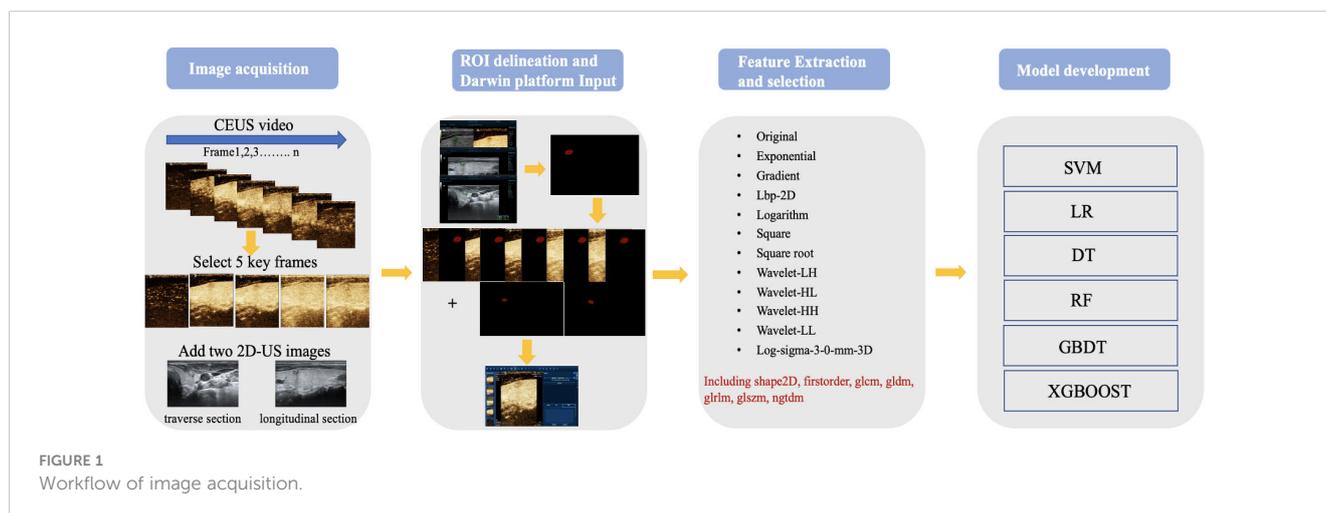
2.6 Statistical analysis

Statistical analysis was performed using the SPSS software (version 26.0; IBM Corp., Armonk, NY, USA). Count data were recorded as frequencies and rates. The measurement data that confirmed a normal distribution were recorded as mean \pm standard deviation, while data that were not consistent with a normal distribution were recorded as the median (interquartile range). Furthermore, measurement data between groups were compared using the independent t -test and Mann-Whitney U test. Count data (clinical data, 2D-US and CEUS data) were analyzed using chi-square or Fisher’s exact tests. Radiomics analyses were performed using Python (version 3.6). Delong’s test was used to test whether there were any differences in AUC among the six ML models and between the ML model and human readers. A calibration curve demonstrated the consistency between the prediction model and the actual situation. Decision curve analysis (DCA) was used to determine whether this model had net clinical benefits. Statistical significance was set at $P < 0.05$.

3 Results

3.1 Clinical and sonographic data

A total of 313 nodules were enrolled in our study, with 282 in the training cohort and 31 in the test cohort. The training cohort included 100 benign and 182 malignant nodules, while the test



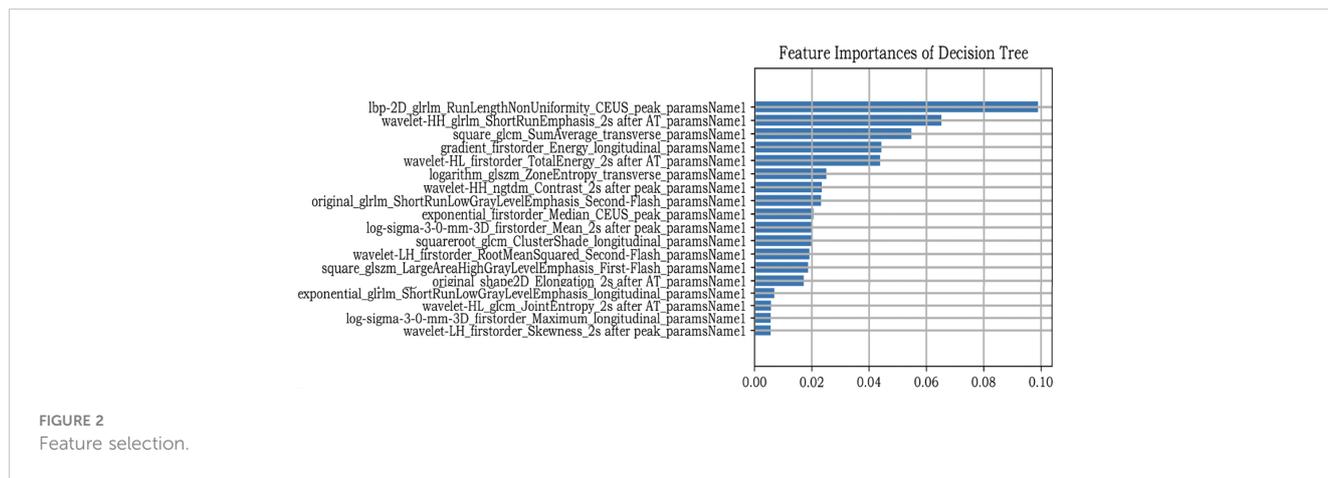


FIGURE 2
Feature selection.

cohort included 10 benign and 21 malignant nodules. In our data, 89 nodules were classified as C- TIRADS 4a, 128 nodules were classified as C- TIRADS 4b, 96 nodules were classified as C- TIRADS 4c. The malignancy rate of C- TIRADS 4a, 4b and 4c were 34.8% (31/89), 70.3% (90/128) and 85.4% (82/96), respectively. The characteristics of the nodules are listed in Table 1, and the patient inclusion flowchart is shown in Figure 3. In the training cohort, the clinical and sonographic variables between the malignant and benign groups showed significant differences in age, number, size, solid composition, microcalcification, shape, margin, enhanced intensity, homogeneity, and wash-in patterns (all $P < 0.05$). However, no significant difference was found in sex, location, Hashimoto’s background, echogenicity, centripetal

enhancement, and wash-out patterns (all $P > 0.05$). There was no statistically significant difference in the distribution of patients between the training and test cohorts ($P > 0.05$).

3.2 The US and CEUS analysis by human reader

Each nodule was evaluated simultaneously by a junior radiologist (3 years of CEUS experience) and a senior radiologist (8 years of CEUS experience). The parallel method was used for combined diagnosis of C- TIRADS and CEUS. That is to say, if both C- TIRADS and CEUS were benign, the final diagnosis was recorded

TABLE 1 Clinical and sonographic characteristics.

Characteristics	Training cohort (n=282)				Test cohort (n = 31)	P
	Total (n = 282)	Benign(n = 100)	Malignant (n = 182)	p		
Age (years)	44.57 ± 12.31	48.12 ± 12.5	42.62 ± 11.79	0.000*	45.52 ± 11.47	0.683
Sex				0.429		0.497
Female	224 (79.4%)	82 (82.0%)	142 (78.0%)		23 (74.2%)	
Male	58 (20.6%)	18 (18.0%)	40 (22.0%)		8 (25.8%)	
Number				0.000*		0.691
Single	99 (35.1%)	19 (19.0%)	80 (44.0%)		12 (38.7%)	
Multiple	183 (64.9%)	81 (81.0%)	102 (56.0%)		19 (61.3%)	
Size (mm)				0.000*		0.779
Maximum diameter	10.67 ± 9.07	15.1 ± 12.14	8.24 ± 5.52		10.2 ± 7.83	
Location				0.133		0.595
Upper pole	64 (22.7%)	23 (23.0%)	41 (22.5%)		10 (32.3%)	
Middle	112 (39.7%)	37 (37.0%)	75 (41.3%)		10 (32.2%)	
Subthyroid pole	77 (27.3%)	34 (34.0%)	43 (23.6%)		7 (22.6%)	
Isthmus	29 (10.3%)	6 (6.0%)	23 (12.6%)		4 (12.9%)	
Hashimoto Background				0.917		0.377
Yes	46 (16.3%)	17 (17.0%)	29 (15.9%)		7 (22.6%)	
No	236 (83.7%)	83 (83.0%)	153 (84.1%)		24 (77.4%)	
Solid composition				0.007*		1.000
Yes	278 (98.6%)	96 (96.0%)	182 (100%)		31 (100%)	
No	4 (1.4%)	4 (4.0%)	0		0	

(Continued)

TABLE 1 Continued

Characteristics	Training cohort (n=282)				Test cohort (n = 31)	P
	Total (n = 282)	Benign(n = 100)	Malignant (n = 182)	p		
Very low echogenicity				0.150		0.860
Yes	16 (5.7%)	3 (3.0%)	13 (7.1%)		2 (6.5%)	
No	266 (94.3%)	97 (97.0%)	169 (92.9%)		29 (93.5%)	
Microcalcification				0.043*		0.075
Yes	89 (31.6%)	24 (24.0%)	65 (35.7%)		5 (16.1%)	
No	193 (68.4%)	76 (76.0%)	117 (64.3%)		26 (83.9%)	
Shape (Aspect ratio)				0.000*		0.971
>1	81 (28.7%)	12 (12.0%)	69 (37.9%)		9 (29.0%)	
<1	201 (71.3%)	88 (88.0%)	113 (62.1%)		22 (71.0%)	
Margin				0.000*		0.279
Regular	144 (51.1%)	72 (72.0%)	72 (39.6%)		19 (61.3%)	
Irregular	138 (48.9%)	28 (28.0%)	110 (60.4%)		12 (38.7%)	
Enhanced intensity				0.000*		0.148
Hyperenhancement	54 (19.2%)	36 (36.0%)	18 (9.9%)		10 (32.3%)	
Iso-enhancement	149 (52.8%)	32 (32.0%)	47 (25.8%)		16 (51.6%)	
Hypoenhancement	79 (28.0%)	32 (32.0%)	117 (64.3%)		5 (16.1%)	
Homogeneity				0.000*		0.150
Homogeneous	108 (38.3%)	56 (56.0%)	52 (28.6%)		16 (51.6%)	
Heterogeneous	174 (61.7%)	44 (44.0%)	130 (71.4%)		15 (48.4%)	
Centripetal enhancement				0.699		1.000
Yes	28 (9.9%)	9 (9.0%)	19 (10.4%)		3 (10%)	
No	254 (90.1%)	91 (91.0%)	163 (89.6%)		28 (90%)	
Wash-in				0.001*		0.180
Synchronous	133 (47.2%)	51 (51.0%)	82 (45.1%)		20 (64.5%)	
Later	116 (41.1%)	29 (29.0%)	87 (47.8%)		9 (29.0%)	
Earlier	33 (11.7%)	20 (20.0%)	13 (7.1%)		2 (6.5%)	
Wash-out				0.337		0.731
Synchronous	180 (63.8%)	64 (64.0%)	116 (63.7%)		22 (71.0%)	
Later	44 (15.6%)	12 (12.0%)	32 (17.6%)		4 (12.9%)	
Earlier	58 (20.6%)	24 (24.0%)	34 (18.7%)		5 (16.1%)	

*Represents P <0.05. Numerical data are presented as mean ± standard deviation. Categorical data are presented as numbers (%).

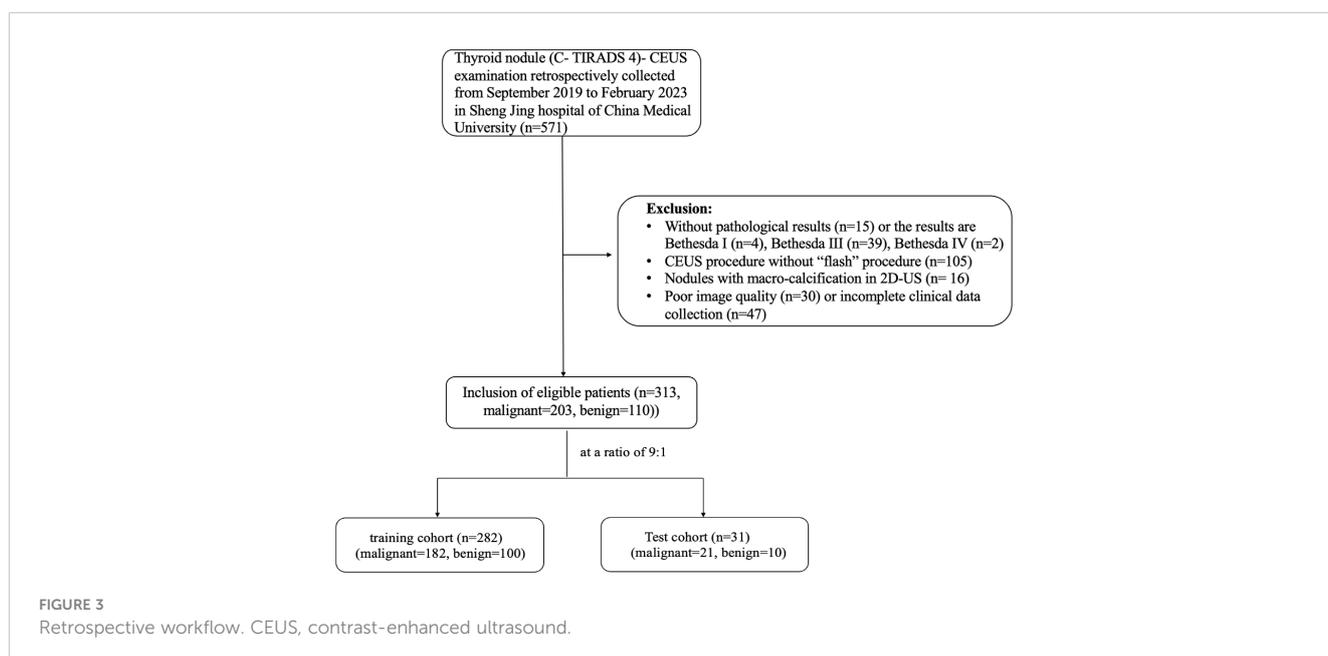
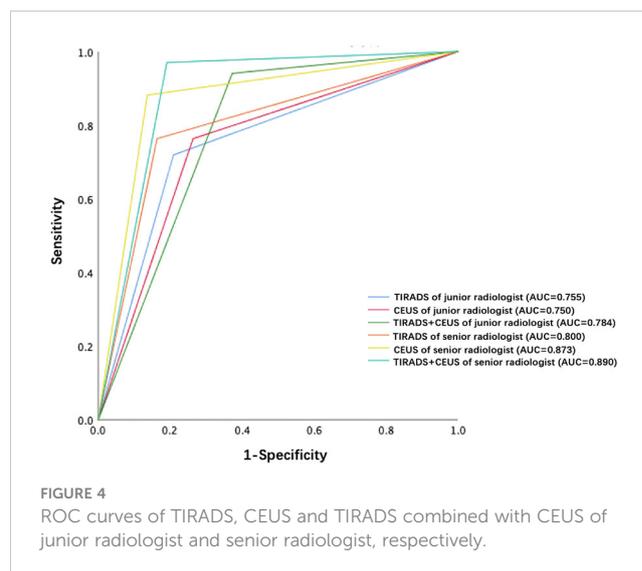


FIGURE 3 Retrospective workflow. CEUS, contrast-enhanced ultrasound.

as benign, while one of the C-TIRADS or CEUS was malignant, the final diagnosis was recorded as malignant. As shown in Table 2; Figure 4, the AUCs of junior radiologist observing US for C-TIRADS classification, CEUS videos, and the combined diagnosis of the two methods were 0.755, 0.750, 0.784, respectively. Except from the specificity and the PPV, the sensitivity, NPV and accuracy of combining US and CEUS by junior radiologist were higher than using US and CEUS alone, which were 0.941, 0.852, 0.831, respectively. The AUC of senior radiologist observing US for C-TIRADS classification, CEUS video, and combined diagnosis of the two methods were 0.80, 0.873 and 0.890 respectively. Except from the specificity and the PPV, the sensitivity, NPV and accuracy of combining US and CEUS by senior radiologist were higher than using US and CEUS alone, which were 0.970, 0.937, 0.914, respectively.

3.3 Prediction performance of ML models based on 2D-US combined with CEUS key frames

The six classifiers (SVM, LR, DT, RF, GBDT, and XGBOOST) and their performance are listed in Table 3. AUCs for SVM, LR, DT, RF, GBDT, and XGBOOST in the training cohort were 0.75, 0.87, 1.00, 1.00, 1.00, and 0.92, respectively. In the test cohort, AUCs of SVM, LR, DT, RF, GBDT, and XGBOOST were 0.74, 0.81, 0.84, 0.94, 0.92, and 0.92, respectively. The ROC curves of the six ML models are shown in Figure 5. The results of the Delong test showed that in the test cohort, the difference between AUC of SVM, LR, and DT was not statistically significant ($P > 0.05$). Similarly, the difference in AUC between RF, XGBOOST, and GBDT was not statistically significant ($P > 0.05$). RF, GBDT, and XGBOOST had comparable predictive effectiveness. The differences in AUC between GBDT, LR, and DT were not statistically significant ($P > 0.05$); however, AUCs of RF and XGBOOST were statistically significant compared to those of SVM, LR, and DT, respectively (all $P < 0.05$). Notably, AUC of RF was the highest in the test cohort



(0.94). Additionally, the calibration and DCA curves of RF showed favorable consistency with reality (Figure 6). The cases in test cohorts were presented in Figures 7, 8.

3.4 Comparison with human readers

A senior radiologist (8 years of CEUS experience) and a junior radiologist (3 years of CEUS experience) independently reviewed the transverse and longitudinal sections of the test cohort's 2D-US and CEUS videos of each nodule. Both groups were blinded to clinical characteristics and pathological results, and a definitive diagnosis of whether each nodule was benign or malignant was provided. The diagnostic performances of the best-performing RF model and human readers are summarized in Table 4; Figure 9. As shown, the RF model achieved an equivalent performance to that of the senior radiologist ($P = 0.799$) and gained more specificity. The RF model outperformed the junior radiologist ($P = 0.039$) and showed greater sensitivity, specificity and NPV.

TABLE 2 The US and CEUS analysis by human readers.

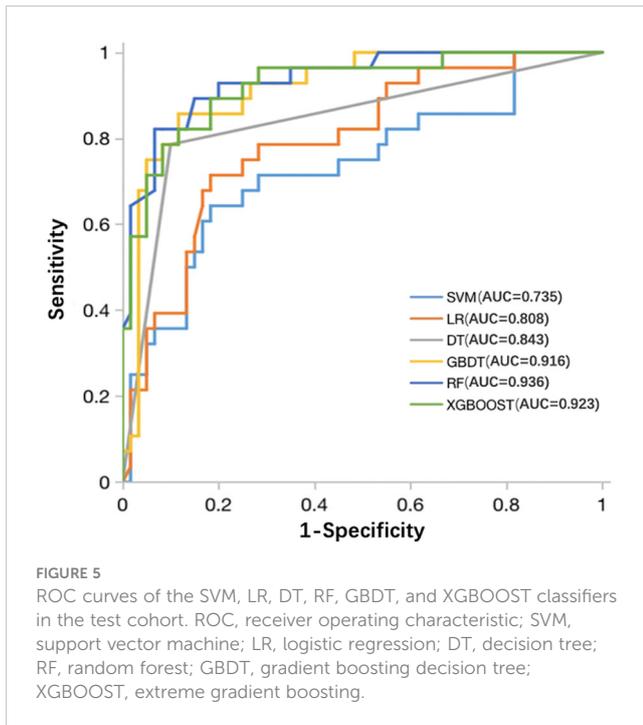
Models	SEN	SPE	PPV	NPV	Accuracy	AUC
Junior radiologist C- TIRADS	0.720 (0.651, 0.779)	0.791 (0.701, 0.860)	0.864 (0.801, 0.910)	0.604 (0.519, 0.683)	0.744	0.755 (0.698, 0.812)
Junior radiologist CEUS	0.764 (0.698, 0.819)	0.736 (0.642, 0.814)	0.842 (0.780, 0.890)	0.628 (0.538, 0.710)	0.754	0.750 (0.692, 0.808)
Junior radiologist C- TIRADS+CEUS	0.941 (0.897, 0.968)	0.627 (0.529, 0.716)	0.823 (0.767, 0.869)	0.852 (0.752, 0.918)	0.831	0.784 (0.698, 0.812)
Senior radiologist C- TIRADS	0.768 (0.703, 0.823)	0.836 (0.751, 0.898)	0.897 (0.839, 0.936)	0.662 (0.576, 0.739)	0.792	0.800 (0.747, 0.853)
Senior radiologist CEUS	0.882 (0.827, 0.921)	0.864 (0.782, 0.920)	0.923 (0.873, 0.955)	0.800 (0.713, 0.864)	0.875	0.873 (0.828, 0.918)
Senior radiologist C- TIRADS+CEUS	0.970 (0.934, 0.988)	0.809 (0.721, 0.875)	0.904 (0.855, 0.938)	0.937 (0.862, 0.974)	0.914	0.890 (0.844, 0.936)

C- TIRADS, Chinese version of thyroid imaging reporting and data system; CEUS, contrast-enhanced ultrasound; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value.

TABLE 3 Predictive performance of six machine learning models based on 2D-US and CEUS key frames.

Parameter	SVM		LR		DT		RF		GBDT		XGBOOST	
	Training cohort	Test cohort	Training cohort	Test cohort	Training cohort	Test cohort	Training cohort	Test cohort	Training cohort	Test cohort	Training cohort	Test cohort
AUC	0.746 (0.707–0.786)	0.735 (0.615–0.854)	0.867 (0.839–0.895)	0.808 (0.709–0.907)	1	0.843 (0.757–0.929)	1	0.936 (0.884–0.988)	0.999 (0.998–1)	0.916 (0.854–0.978)	1	0.923 (0.864–0.984)
ACC	0.741	0.75	0.791	0.807	1	0.864	1	0.898	0.99	0.864	1	0.841
SEN	0.671 (0.61–0.726)	0.643 (0.458–0.793)	0.813 (0.761–0.857)	0.679 (0.493, 0.821)	1 (0.985–1)	0.786 (0.605–0.898)	1 (0.985,1)	0.821 (0.644–0.921)	0.988 (0.966–0.996)	0.857 (0.685–0.943)	1 (0.985–1)	0.893 (0.728–0.963)
SPE	0.773 (0.736–0.807)	0.8 (0.682–0.882)	0.781 (0.744–0.814)	0.867 (0.758, 0.931)	1 (0.993–1)	0.9 (0.799–0.953)	1 (0.993,1)	0.933 (0.841–0.974)	0.991 (0.978–0.996)	0.867 (0.758–0.931)	1 (0.993–1)	0.817 (0.701–0.894)
PPV	0.581 (0.523–0.636)	0.6 (0.423–0.754)	0.635 (0.581–0.685)	0.704 (0.515–0.841)	1 (0.985–1)	0.786 (0.605–0.898)	1 (0.985–1)	0.852 (0.675–0.941)	0.98 (0.955–0.992)	0.75 (0.579–0.867)	1 (0.985–1)	0.694 (0.531–0.82)
NPV	0.834 (0.798–0.864)	0.828 (0.711–0.904)	0.899 (0.869–0.923)	0.852 (0.743–0.92)	1 (0.993–1)	0.9 (0.799–0.953)	1 (0.993–1)	0.918 (0.822–0.964)	0.994 (0.984–0.998)	0.929 (0.83–0.972)	1 (0.993–1)	0.942 (0.8440.98)
F1-Score	0.741	0.621	0.713	0.691	1	0.786	1	0.836	0.984	0.8	1	0.781

SVM, support vector machine; LR, logistic regression; DT, decision tree; RF, random forest; GBDT, gradient boosting decision tree; XGBOOST, extreme gradient boosting; AUC, area under curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value.

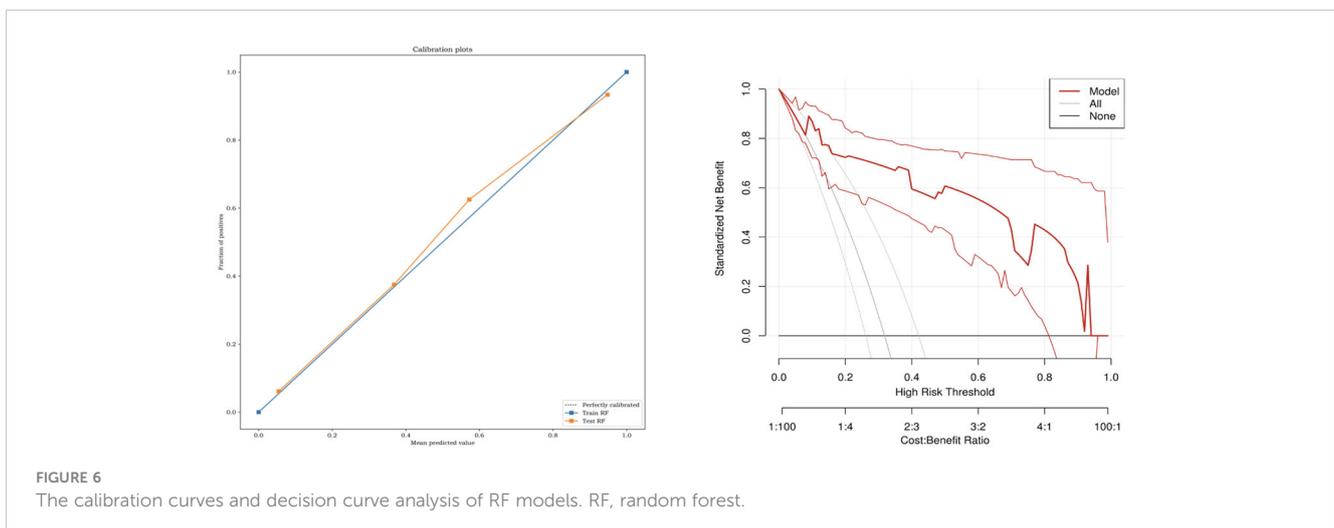


4 Discussion

In this study, we constructed six ML models using 2D-US images combined with five CEUS keyframes. The ROC curves showed that the diagnostic performance of our models was desirable, with all AUC values >0.80 in the test cohort (except SVM [0.74]). Moreover, we compared our best model with human readers (senior and junior radiologists) and found that the best ML model achieved equivalent performance to that of the senior radiologist and outperformed the junior radiologist.

Traditional diagnostic methods for thyroid nodules, such as 2D-US, color Doppler flow imaging (CDFI), elastography, and FNA, have many disadvantages (35–37); the main ones are severe overdiagnosis and overtreatment (38, 39). For example, some Hashimoto's nodules may show hypoechogenicity with blurred

margins on 2D-US, which may be classified as TIRADS >4 and require unnecessary FNA according to the guidelines (7, 40, 41). CEUS, as a novel noninvasive microangiography technology, can reveal microvasculature with a smaller diameter ($>40\ \mu\text{m}$) than that by CDFI ($>100\ \mu\text{m}$) and is helpful in the detection of malignant thyroid nodules (42, 43). Recent studies have indicated that CEUS could modify the current TIRADS to create a new risk stratification that may reduce unnecessary biopsies (42–46). Our team had published one CEUS- TIRADS model to differentiate thyroid nodules (C-TIRADS 4) by combining CEUS with C-TIRADS (46), which had high clinical practicability in clinic. Additionally, CEUS images may contain valuable information that has not received sufficient attention in daily clinical practice. In recent years, AI, especially radiomic features, has demonstrated promising potential for evaluating the characteristics of thyroid nodules (47, 48). Radiomics has also been used to diagnose cytologically uncertain nodules (49–51), lymph node metastases (52, 53), and extrathyroidal extension (54). Many studies employing AI for evaluating the thyroid are mainly based on 2D-US images (48, 55, 56). In 2015, LeCun introduced the principles of deep learning and convolutional neural networks (CNNs) (18), attracting the interest of many researchers. The principle of machine or deep learning is that CNNs are trained using a large number of 2D-US images with known corresponding pathological results. A specific algorithm is used to segment US images. After several calculation iterations, the CNNs can capture and analyze thyroid nodules and suggest risk stratification. Studies on ML based on 2D-US to distinguish malignant thyroid nodules from benign nodules could reach a diagnostic accuracy of approximately 90%. Peng et al. developed a deep learning AI model based on 2D-US to diagnose thyroid nodules that outperformed 12 radiologists (AUC: 0.922 vs. 0.839, $P < 0.05$) (37). Conversely, a study conducted by Sun et al., also based on 2D-US, indicated that the experts achieved better performance (AUC: 0.881 vs. 0.819) (57). Gong et al. reported that an AI-assisted diagnostic system combined with CEUS could significantly improve the diagnostic sensitivity and NPV in diagnosing thyroid nodules classified as American College of Radiology Thyroid Imaging (ACR-TIRADS) 4 (58). However, to



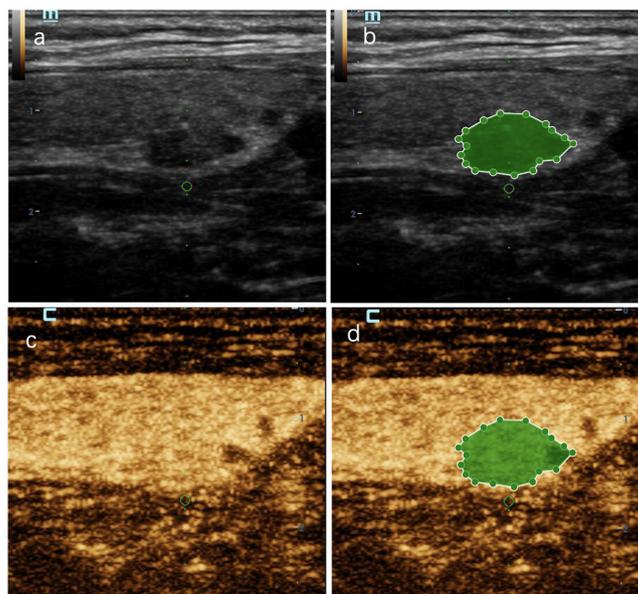


FIGURE 7

A thyroid nodule in left lobe in a 46-year-old woman in test cohort. (A) 2D-US image; (B) the mask image corresponding to 2D-US image; (C) CEUS image at peak time; (D) the mask image of CEUS image at peak time. The nodule was solid, hypoechoic, blurred margin, aspect ratio less than 1, with microcalcification and was categorized as C-TIRADS 4c. CEUS showed “later wash-in, heterogeneous enhancement” and “later wash-out”, and was diagnosed as malignant. RF model classifies it as malignant. Histologic analysis revealed papillary microcarcinoma (PTMC).

our knowledge, few researchers have developed AI models based on CEUS images. To date, only two studies have proposed AI diagnostic models based on CEUS image information (27, 28). Wan et al. used DL to build a diagnostic model based on dynamic CEUS video and obtained AUC of 0.92 (27), which was lower than

ours (AUC: 0.94); ACC in their study was substantially lower than that in ours (0.80 vs. 0.90). Guo et al. used logistic regression to build ML models based on US and CEUS features, while as for CEUS features, only a single frame of CEUS images was used (28). Our studies extracted radiomics features five key CEUS frames and

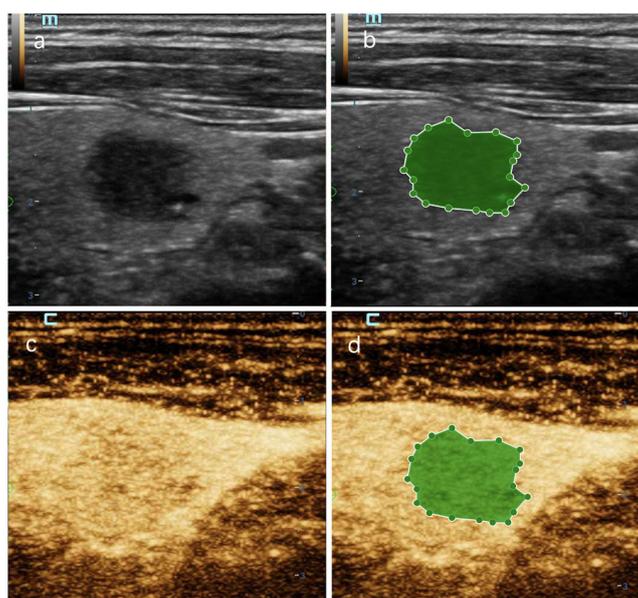


FIGURE 8

A thyroid nodule in right lobe in a 56-year-old man in test cohort. (A) 2D-US image; (B) the mask image corresponding to 2D-US image; (C) CEUS image at peak time; (D) the mask image of CEUS image at peak time. The nodule was solid with blurred margin and was categorized as C-TIRADS 4b. CEUS showed “later wash-in” and “with hypointensity”, and was diagnosed as malignant. RF model classifies it as benign. Histologic analysis revealed nodular goiter with granuloma formation.

TABLE 4 Diagnostic performance of the RF model compared to human readers in the test cohort.

	Models	SEN	SPE	PPV	NPV	Accuracy	AUC	P
Test cohort	RF model	0.821 (0.644–0.921)	0.933 (0.841–0.974)	0.852 (0.675–0.941)	0.918 (0.822–0.964)	0.898	0.936 (0.884–0.988)	
	Senior radiologist	0.965 (0.868–0.994)	0.839 (0.655–0.939)	0.917 (0.808–0.968)	0.929 (0.750–0.988)	0.920	0.923 (0.854–0.991)	0.799
	Junior radiologist	0.817 (0.691–0.901)	0.786 (0.585–0.910)	0.891 (0.771–0.955)	0.667 (0.481–0.814)	0.807	0.801 (0.696–0.906)	0.039*

RF, random forest; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve.

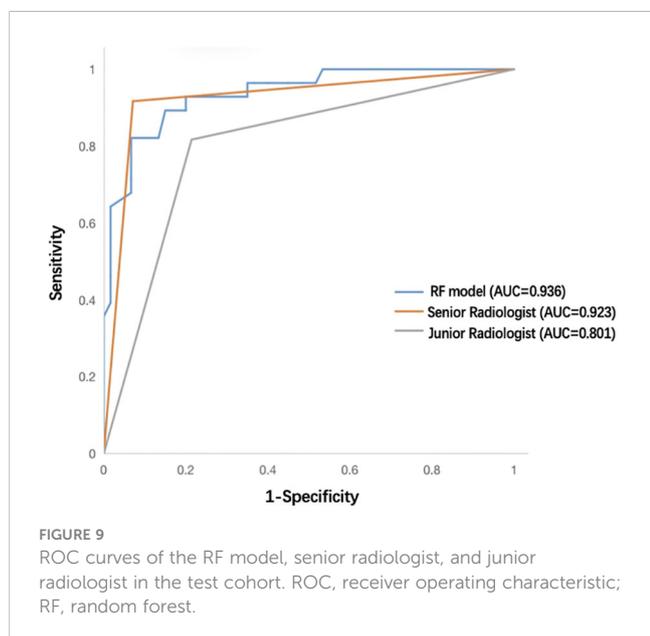
the sample size of our study is bigger (313 vs. 123). And our study aimed at thyroid nodules which are classified as C-TIRADS 4, which are relatively hardly differentiated in clinic. Therefore, this was the first study to provide the highest value of radiomics information from CEUS images in thyroid nodules (C-TIRADS 4) evaluation, offering a promising, noninvasive, fast, feasible, and reliable method.

In our study, none of the patients experienced complications during CEUS and FNA. By comparing the FNA and surgical pathological results from January 2016 to June 2021 in our hospital, we found that the success rate and diagnostic accuracy of FNA were 96.6% and 93.3%, respectively (59). The accuracy of FNA was much higher than that in most previous studies, indicating that the pathological results from FNA at our institution were reliable. Our study also demonstrated that malignant thyroid nodules commonly occurred in younger people ($P < 0.05$). The statistical differences between malignant and benign nodules in the training cohort were also significant for nodule number, nodule size, nodule composition, the presence of microcalcifications, shape, margin, enhanced intensity of CEUS, homogeneity of CEUS, and wash-in patterns of CEUS (all $P < 0.05$). Regarding CEUS patterns, the malignant nodules in our data mostly showed hypoenhancement (117/182; 64.3%), heterogeneous enhancement (130/182; 71.4%), and later wash-in (87/182;

47.8%), which is consistent with previous studies (12, 33, 34). This may be attributed to the peripheral blood vessels of malignant nodules being damaged by malignant growth, hindering contrast agent entry. When the nodule is small, the number of new blood vessels, branches, and arteriovenous fistulas is not relatively large, and the inside of the nodule will be closely related to the poor blood supply and uneven distribution of blood vessels within the malignant nodules. In the present study, the mean maximal diameter of malignant nodules was smaller than that of benign nodules ($P < 0.05$), which may indicate that the direction of perfusion of contrast agents was difficult to observe, which explains the lack of statistical significance in the enhancement methods and wash-out patterns. And in our data, the diagnostic AUC and accuracy of both junior and senior radiologist of using US combined with CEUS were higher than those of US or CEUS alone.

In this study, we first extracted nearly all radiomics features as published in the present literature. Subsequently, we adopted maximum abs normalization to preprocess the data. Many data normalization methods are used in ML, such as Z-score standardization, max abs normalization, min-max normalization, robust scaling, and median absolute deviation. The advantage of max abs normalization lies in its ability to retain data distribution without centralizing it, preserving the sparsity of large-scale data such as ours. We then used the variance threshold to eliminate outliers from the data. DT is a nonparametric method. Thus, it does not make any assumptions regarding the spatial distribution or categorical structure of the data, making it suitable for our study. The best feature selection is based on the DT classifier. Wavelet features accounted for the largest proportion of radiomics features (6/18). High-dimensional wavelet features are texture features that show lesion heterogeneity (60). Fan et al. used ML to predict the aggressiveness of prostate cancer, and wavelet features accounted for the largest proportion of their models (61). Meng et al. and Aerts et al. reached similar conclusions (60, 62). Additionally, CEUS frames played a substantial role in feature selection (12/18), illustrating the importance of CEUS images. Moreover, among the selected features, the top-ranked one was the “time to peak” frame. This may be because the image is brightest at the peak time, and the number of microbubbles in the nodule area is the highest, which can probably provide more information.

Classifiers play a crucial role in ML procedures. Our study uses six classifiers for model development (SVM, LR, DT, RF, GBDT, and XGBOOST). Support vector machine (SVM) is a kind of generalized linear classifier for binary classification of data according to supervised learning, which is more suitable for dealing with complex nonlinear



equations than logistic regression. Compared with SVM, LR can be used for multivariate classification and is more suitable for small data volume. Decision tree (DT) is a basic classification and regression method and defined as a conditional probability distribution on feature space and class space. Both random forest (RF) and gradient boosting decision tree (GBDT) are based on DT. RF is an extension of a parallel ensemble learning method, and “random” means the randomness of the selected partition attributes. GBDT is a decision tree model trained with gradient boosting strategy, which performs well in screening features (63). XGBOOST is a kind of basic GBDT, but compared with GBDT, it can support custom loss functions and add more regular terms, handling of missing value and column sampling. Among the four models based on DT, RF can converge to a lower generalization error than the traditional DT. What is more, DT selects the optimal partition attribute from all attribute sets, while RF selects the partition attribute only in a subset of the attribute set, so the training efficiency is higher. And each tree of RF only chooses part of samples and features, breaking through the “overfitting” defect of DT. Compared with GBDT and XGBOOST, the performance of RF is more stable, the parameter adjusting is relatively less complicated, the operation time is short, and the universality is stronger. Compared with SVM and LR, RF randomly selects samples and features for each tree, removes noise variables, increases noise resistance and provides more stable performance. Moreover, unlike SVM, as the number of observed samples and features increases, SVM firstly needs to spend much time to find a suitable kernel function during the calculation. RF has no such weakness. The results of our study also proved that RF was the optimal classifier for our model. In our data, the RF, GBDT, and XGBOOST classifiers generally performed better than the SVM, LR, and DT classifiers. The RF model performed the best (AUC: 0.94, 95% CI: 0.884–0.988; ACC: 0.90). In the test cohort, our RF model obtained an equivalent performance to that of the senior radiologist (AUC: 0.94 vs. 0.92, $P = 0.798$; ACC: 0.90 vs. 0.92) and was considerably higher in specificity than both the senior (0.93 vs. 0.84) and junior (0.93 vs. 0.79) radiologists. The good performance of our model also indicated that during the CEUS process, the radiologists could pay more attention to those five time points: “2nd second after the arrival time,” “time to peak” frame, “2nd second after peak” frame, “first-flash” frame, and “second-flash” frame, especially the peak time. This not only achieves comparable performance in diagnosing thyroid nodules, which are classified as C- TIRADS 4, but also saves radiologists time compared to watching the entire CEUS video.

This study had some limitations. First, this was a single-center retrospective study; our institution is a referral center, and the malignancy risk of thyroid nodules is relatively high, which may have led to selection bias in our samples. Second, this study lacked external verification, requiring a multi-center, multi-hospital, multi-region study to augment the robustness and generalizability of our results. Third, the ROI lines of the nodules were all manually delineated, and key-frame selection was also observed and operated by radiologists, although we had obtained rather good performance; however, these two procedures are time-consuming and prone to errors, and their efficiency and accuracy could potentially be improved with the implementation of a mature automated artificially intelligent system.

5 Conclusion

Our study established six ML models based on two 2D-US images and five CEUS key frames to distinguish malignant from benign thyroid nodules which were classified as C-TIRADS 4. Our study highlighted the information of CEUS image extracted by ML that could not be seen by human eyes, indicating that CEUS may have great potential in the field of thyroid nodules. The RF model, as the optimal ML algorithm, may provide a noninvasive, convenient, feasible, and highly accurate method for invasive FNA and assist junior radiologists in diagnosis or preoperative prediction models. Further studies will address these limitations, making it possible to improve clinical diagnostic and therapeutic strategies.

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 Medical Ethics Committee of Shengjing Hospital, China Medical University. 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

J-hC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Y-QZ: Conceptualization, Methodology, Software, Validation, Visualization, Writing – review & editing. T-tZ: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Writing – review & editing. QZ: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review & editing. A-xZ: Data curation, Formal analysis, Writing – review & editing. YH: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by grants from the 345 Talent Project of Shengjing Hospital of China Medical University; Liaoning Province Bai Qian Wan Talents Program; Liaoning Province “Xingliao Talent Plan” Medical Master Project (YXMJ-LJ-10) and Liaoning Provincial

Science and Technology Program Combined Program (Key R&D Program Projects).

Acknowledgments

We would like to thank Yizhun Medical AI Technology Co., Ltd., who kindly provided the Darwin research platform and technical support.

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.

References

- Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med.* (2016) 375:614–7. doi: 10.1056/NEJMp1604412
- 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
- Batawil N, Alkordy T. Ultrasonographic features associated with Malignancy in cytologically indeterminate thyroid nodules. *Eur J Surg Oncol.* (2014) 40:182–6. doi: 10.1016/j.ejso.2013.11.015
- Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology.* (2011) 260:892–9. doi: 10.1148/radiol.11110206
- Ma YH, Yue T, He QQ. Tracheal injury following robotic thyroidectomy: A literature review of epidemiology, etiology, diagnosis, and treatment and 3 case reports. *Asian J Surg.* (2023) 10:039. doi: 10.1016/j.asjsur
- Haddou N, Idrissi N, Ben Jebara S. Analysis of voice quality after thyroid surgery. *J Voice.* (2023) S0892-1997(23)00208-4. doi: 10.1016/j.jvoice.2023.06.027
- 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
- Zhu TT, Zhuang LT, Ma XF, Zhao AX, Huang Y. Differential diagnosis of Malignant and Hashimoto thyroid nodules by conventional ultrasound combined with contrast-enhanced ultrasound. *Chin J Med Imaging Technol.* (2021) 37:1789–93. doi: 10.13929/j.issn.10033289.2021.12.007
- Chen S, Tang K, Gong Y, Ye F, Liao L, Li X, et al. Value of contrast-enhanced ultrasound in mummified thyroid nodules. *Front Endocrinol (Lausanne).* (2022) 13:2022.850698. doi: 10.3389/fendo.2022.850698
- Yin T, Zheng B, Lian Y, Li H, Tan L, Xu S, et al. Contrast-enhanced ultrasound improves the potency of fine-needle aspiration in thyroid nodules with high inadequate risk. *BMC Med Imaging.* (2022) 22:83. doi: 10.1186/s12880-022-00805-6
- Zhang M, Luo Y, Zhang Y, Tang J. Efficacy and safety of ultrasound-guided radiofrequency ablation for treating low-risk papillary thyroid microcarcinoma: A prospective study. *Thyroid.* (2016) 26:1581–7. doi: 10.1089/thy.2015.0471
- Wang Y, Dong T, Nie F, Wang G, Liu T, Niu Q. Contrast-enhanced ultrasound in the differential diagnosis and risk stratification of ACR TI-RADS category 4 and 5 thyroid nodules with non-hypovascular. *Front Oncol.* (2021) 11:2021.662273. doi: 10.3389/fonc.2021.662273
- Zhang J, Zhang X, Meng Y, Chen Y. Contrast-enhanced ultrasound for the differential diagnosis of thyroid nodules: An updated meta-analysis with comprehensive heterogeneity analysis. *PLoS One.* (2020) 15:e0231775. doi: 10.1371/journal.pone.0231775
- Wan Q, Cao P, Liu J. Meta-analysis of contrast enhanced ultrasound in judging benign and Malignant thyroid tumors. *Comput Math Methods Med.* (2021) 2021:2577113. doi: 10.1155/2021/2577113
- Wu Q, Wang Y, Li Y, Hu B, He ZY. Diagnostic value of contrast-enhanced ultrasound in solid thyroid nodules with and without enhancement. *Endocrine.* (2016) 53:480–8. doi: 10.1007/s12020-015-0850-0

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/fendo.2024.1299686/full#supplementary-material>

- Zhang Y, Luo YK, Zhang MB, Li J, Li J, Tang J. Diagnostic accuracy of contrast-enhanced ultrasound enhancement patterns for thyroid nodules. *Med Sci Monit.* (2016) 22:4755–64. doi: 10.12659/msm.899834
- Sorrenti S, Dolcetti V, Fresilli D, Del Gaudio G, Pacini P, Huang P, et al. The role of CEUS in the evaluation of thyroid cancer: from diagnosis to local staging. *J Clin Med.* (2021) 10(19):4559. doi: 10.3390/jcm10194559
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539
- Wongkoblap A, Vadillo MA, Curcin V. Modeling depression symptoms from social network data through multiple instance learning. *AMIA Jt Summits Transl Sci Proc.* (2019) 2019:44–53.
- Jing X, Wielema M, Cornelissen LJ, van Gent M, Iwema WM, Zheng S, et al. Using deep learning to safely exclude lesions with only ultrafast breast MRI to shorten acquisition and reading time. *Eur Radiol.* (2022) 32(12):8706–15. doi: 10.1007/s00330-022-08863-8
- Zheng Y, Zhou D, Liu H, Wen M. CT-based radiomics analysis of different machine learning models for differentiating benign and Malignant parotid tumors. *Eur Radiol.* (2022) 32(10):6953–64. doi: 10.1007/s00330-022-08830-3
- Almberg SS, Lervåg C, Frengen J, Eidem M, Abramova TM, Nordstrand CS, et al. Training, validation, and clinical implementation of a deep-learning segmentation model for radiotherapy of loco-regional breast cancer. *Radiother Oncol.* (2022) 173:62–8. doi: 10.1016/j.radonc.2022.05.018
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature.* (2020) 577:89–94. doi: 10.1038/s41586-019-1799-6
- Mao N, Yin P, Wang Q, Liu M, Dong J, Zhang X, et al. Added value of radiomics on mammography for breast cancer diagnosis: A feasibility study. *J Am Coll Radiol.* (2019) 16:485–91. doi: 10.1016/j.jacr.2018.09.041
- Bai Z, Chang L, Yu R, Li X, Wei X, Yu M, et al. Thyroid nodules risk stratification through deep learning based on ultrasound images. *Med Phys.* (2020) 47:6355–65. doi: 10.1002/mp.14543
- Nguyen DT, Kang JK, Pham TD, Batchuluun G, Park KR. Ultrasound image-based diagnosis of Malignant thyroid nodule using artificial intelligence. *Sensors (Basel).* (2020) 20(7):1822. doi: 10.3390/s20071822
- Wan P, Chen F, Liu C, Kong W, Zhang D. Hierarchical temporal attention network for thyroid nodule recognition using dynamic CEUS imaging. *IEEE Trans Med Imaging.* (2021) 40:1646–60. doi: 10.1109/tmi.2021.3063421
- Guo SY, Zhou P, Zhang Y, Jiang LQ, Zhao YF. Exploring the value of radiomics features based on B-mode and contrast-enhanced ultrasound in discriminating the nature of thyroid nodules. *Front Oncol.* (2021) 11:738909. doi: 10.3389/fonc.2021.738909
- Jin ZQ, Yu HZ, Mo CJ, Su RQ. Clinical study of the prediction of Malignancy in thyroid nodules: modified score versus 2017 american college of radiology's thyroid imaging reporting and data system ultrasound lexicon. *Ultrasound Med Biol.* (2019) 45:1627–37. doi: 10.1016/j.ultrasmedbio.2019.03.014
- Sidhu PS, Cantisani V, Dietrich CF, Gilja OH, Saftoiu A, Bartels E, et al. The EFSUMB guidelines and recommendations for the clinical practice of contrast-enhanced ultrasound (CEUS) in non-hepatic applications: update 2017 (Long version). *Ultraschall Med.* (2018) 39:e2–e44. doi: 10.1055/a-0586-1107

31. Radzina M, Ratniece M, Putrins DS, Saule L, Cantisani V. Performance of contrast-enhanced ultrasound in thyroid nodules: review of current state and future perspectives. *Cancers (Basel)*. (2021) 13(21):5469. doi: 10.3390/cancers13215469
32. He Y, Wang XY, Hu Q, Chen XX, Ling B, Wei HM. Value of contrast-enhanced ultrasound and acoustic radiation force impulse imaging for the differential diagnosis of benign and Malignant thyroid nodules. *Front Pharmacol*. (2018) 9:1363. doi: 10.3389/fphar.2018.01363
33. Pang T, Huang L, Deng Y, Wang T, Chen S, Gong X, et al. Logistic regression analysis of conventional ultrasonography, strain elastosonography, and contrast-enhanced ultrasound characteristics for the differentiation of benign and Malignant thyroid nodules. *PLoS One*. (2017) 12:e0188987. doi: 10.1371/journal.pone.0188987
34. Yu D, Han Y, Chen T. Contrast-enhanced ultrasound for differentiation of benign and Malignant thyroid lesions: meta-analysis. *Otolaryngol Head Neck Surg*. (2014) 151:909–15. doi: 10.1177/0194599814555838
35. Torigian DA, Li G, Alavi A. The role of CT, MR imaging, and ultrasonography in endocrinology. *PET Clin*. (2007) 2:395–408. doi: 10.1016/j.cpet.2008.05.002
36. Jiang L, Zhang D, Chen YN, Yu XJ, Pan MF, Lian L. The value of conventional ultrasound combined with superb microvascular imaging and color Doppler flow imaging in the diagnosis of thyroid Malignant nodules: a systematic review and meta-analysis. *Front Endocrinol (Lausanne)*. (2023) 14:2023.1182259. doi: 10.3389/fendo.2023.1182259
37. Chambara N, Lo X, Chow TCM, Lai CMS, Liu SYW, Ying M. Combined shear wave elastography and EU TIRADS in differentiating Malignant and benign thyroid nodules. *Cancers (Basel)*. (2022) 14(22):5521. doi: 10.3390/cancers14225521
38. Takano T. Overdiagnosis of juvenile thyroid cancer: time to consider self-limiting cancer. *J Adolesc Young Adult Oncol*. (2020) 9:286–8. doi: 10.1089/jayao.2019.0098
39. Acosta GJ, Singh Ospina N, Brito JP. Overuse of thyroid ultrasound. *Curr Opin Endocrinol Diabetes Obes*. (2023) 30:225–30. doi: 10.1097/med.0000000000000814
40. Zhao T, Xu S, Zhang X, Xu C. Comparison of various ultrasound-based Malignant risk stratification systems on an occasion for assessing thyroid nodules in hashimoto's thyroiditis. *Int J Gen Med*. (2023) 16:599–608. doi: 10.2147/ijgm.S398601
41. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol*. (2016) 17:370–95. doi: 10.3348/kjr.2016.17.3.370
42. Xiao F, Li JM, Han ZY, Liu FY, Yu J, Xie MX, et al. Multimodality US versus thyroid imaging reporting and data system criteria in recommending fine-needle aspiration of thyroid nodules. *Radiology*. (2023) 307:e221408. doi: 10.1148/radiol.221408
43. Zhou P, Chen F, Zhou P, Xu L, Wang L, Wang Z, et al. The use of modified TI-RADS using contrast-enhanced ultrasound features for classification purposes in the differential diagnosis of benign and Malignant thyroid nodules: A prospective and multi-center study. *Front Endocrinol (Lausanne)*. (2023) 14:2023.1080908. doi: 10.3389/fendo.2023.1080908
44. Zhu T, Chen J, Zhou Z, Ma X, Huang Y. Differentiation of thyroid nodules (C-TIRADS 4) by combining contrast-enhanced ultrasound diagnosis model with chinese thyroid imaging reporting and data system. *Front Oncol*. (2022) 12:2022.840819. doi: 10.3389/fonc.2022.840819
45. Cheng H, Zhuo SS, Rong X, Qi TY, Sun HG, Xiao X, et al. Value of contrast-enhanced ultrasound in adjusting the classification of chinese-TIRADS 4 nodules. *Int J Endocrinol*. (2022) 2022:5623919. doi: 10.1155/2022/5623919
46. Ruan J, Xu X, Cai Y, Zeng H, Luo M, Zhang W, et al. A practical CEUS thyroid reporting system for thyroid nodules. *Radiology*. (2022) 305:149–59. doi: 10.1148/radiol.212319
47. Zhu YC, Jin PF, Bao J, Jiang Q, Wang X. Thyroid ultrasound image classification using a convolutional neural network. *Ann Transl Med*. (2021) 9:1526. doi: 10.21037/atm-21-4328
48. 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–e9. doi: 10.1016/s2589-7500(21)00041-8
49. Alabrak MMA, Megahed M, Alkhouly AA, Mohammed A, Elfandy H, Tahoun N, et al. Artificial intelligence role in subclassifying cytology of thyroid follicular neoplasm. *Asian Pac J Cancer Prev*. (2023) 24:1379–87. doi: 10.31557/apjcp.2023.24.4.1379
50. Hirokawa M, Niioka H, Suzuki A, Abe M, Arai Y, Nagahara H, et al. Application of deep learning as an ancillary diagnostic tool for thyroid FNA cytology. *Cancer Cytopathol*. (2023) 131:217–25. doi: 10.1002/cncy.22669
51. Cui Y, Fu C, Si C, Li J, Kang Y, Huang Y, et al. Analysis and comparison of the Malignant thyroid nodules not recommended for biopsy in ACR TIRADS and AI TIRADS with a large sample of surgical series. *J Ultrasound Med*. (2023) 42:1225–33. doi: 10.1002/jum.16132
52. Wang Z, Qu L, Chen Q, Zhou Y, Duan H, Li B, et al. Deep learning-based multifeature integration robustly predicts central lymph node metastasis in papillary thyroid cancer. *BMC Cancer*. (2023) 23:128. doi: 10.1186/s12885-023-10598-8
53. Abbasian Ardakani A, Mohammadi A, Mirza-Aghazadeh-Attari M, Faeghi F, Vogl TJ, Acharya UR. Diagnosis of metastatic lymph nodes in patients with papillary thyroid cancer: A comparative multi-center study of semantic features and deep learning-based models. *J Ultrasound Med*. (2023) 42:1211–21. doi: 10.1002/jum.16131
54. Lu WJ, Mao L, Li J, OuYang LY, Chen JY, Chen SY, et al. Three-dimensional ultrasound based radiomics nomogram for the prediction of extrathyroidal extension features in papillary thyroid cancer. *Front Oncol*. (2023) 13:2023.1046951. doi: 10.3389/fonc.2023.1046951
55. Liu Z, Zhong S, Liu Q, Xie C, Dai Y, Peng C, et al. Thyroid nodule recognition using a joint convolutional neural network with information fusion of ultrasound images and radiofrequency data. *Eur Radiol*. (2021) 31:5001–11. doi: 10.1007/s00330-020-07585-z
56. Gomes Ataide EJ, Ponugoti N, Illanes A, Schenke S, Kreissl M, Friebe M. Thyroid nodule classification for physician decision support using machine learning-evaluated geometric and morphological features. *Sensors (Basel)*. (2020) 20(21):6110. doi: 10.3390/s20216110
57. Sun C, Zhang Y, Chang Q, Liu T, Zhang S, Wang X, et al. Evaluation of a deep learning based computer-aided diagnosis system for distinguishing benign from Malignant thyroid nodules in ultrasound images. *Med Phys*. (2020) 47:3952–60. doi: 10.1002/mp.14301
58. Gong ZJ, Xin J, Yin J, Wang B, Li X, Yang HX, et al. Diagnostic value of artificial intelligence-assistant diagnostic system combined with contrast-enhanced ultrasound in thyroid TI-RADS 4 nodules. *J Ultrasound Med*. (2023) 42:1527–35. doi: 10.1002/jum.16170
59. Ma XF ZT, Zhuang LT, Huang Y. Retrospective study and interrupted time series analysis of ultrasound guided thyroid fine needle aspiration. *J China Clinic Med Imaging*. (2022) 33:837–41. doi: 10.12117/jccmi.2022.12.001
60. Bhattacherjee S, Kim CH, Park HG, Prakash D, Madusanka N, Cho NH, et al. Multi-features classification of prostate carcinoma observed in histological sections: analysis of wavelet-based texture and colour features. *Cancers (Basel)*. (2019) 11(12):1937. doi: 10.3390/cancers11121937
61. Fan X, Xie N, Chen J, Li T, Cao R, Yu H, et al. Multiparametric MRI and machine learning based radiomic models for preoperative prediction of multiple biological characteristics in prostate cancer. *Front Oncol*. (2022) 12:2022.839621. doi: 10.3389/fonc.2022.839621
62. Meng X, Xia W, Xie P, Zhang R, Li W, Wang M, et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. *Eur Radiol*. (2019) 29:3200–9. doi: 10.1007/s00330-018-5763-x
63. Zhang Z, Jung C. GBDT-MO: gradient-boosted decision trees for multiple outputs. *IEEE Trans Neural Netw Learn Syst*. (2021) 32:3156–67. doi: 10.1109/TNNLS.2020.3009776