

Development and optimisation strategies for a nomogram-based predictive model of malignancy risk in thyroid nodules

Peng He [#], Yu Liang [#], Yuan Zou ^{*}, Zhou Zou, Bo Ren, Shan Peng, Hongmei Yuan, Qin Chen

ABSTRACT

Introduction: This study aimed to develop and validate a clinical prediction model to assist radiologists in optimising the diagnostic classification of the Chinese Thyroid Imaging Reporting and Data System (C-TIRADS).

Methods: A total of 1659 patients from two hospitals were included in this study. The derivation cohort comprised 909 patients for model development and internal validation, while 750 patients formed the external validation cohort. A binary logistic regression model was constructed. Model performance in the derivation set was evaluated using receiver operating characteristic (ROC) curves and visualised with a nomogram. In the external validation set, ROC and calibration curves were used to assess discrimination and calibration.

Results: The original C-TIRADS category, abnormal cervical lymph node sonographic findings, and changes in thyroid nodule size emerged as significant predictors of C-TIRADS optimisation. The optimised nomogram demonstrated an area under the ROC curve (AUC) of 0.730 (95% confidence interval=0.697-0.762), with a sensitivity of 63.2%, specificity of 74.9%, and overall accuracy of 67.7% for predicting optimisation. Using probability thresholds of $\geq 60\%$ to recommend an upgrade and $<30\%$ to recommend a downgrade, the calibration curve showed good agreement, and decision curve analysis demonstrated a favourable net clinical benefit. External validation confirmed excellent

discrimination (AUC=0.865; 95% confidence interval=0.839-0.891).

Conclusion: An optimised C-TIRADS model that integrates imaging features of thyroid nodules with clinical risk factors may aid radiologists in improving the diagnostic efficiency and clinical utility of the TIRADS classification.

Hong Kong Med J 2026;32:Epub

<https://doi.org/10.12809/hkmj2512718>

¹ P He, MD, PhD

² Y Liang, MD

¹ Y Zou ^{*}, MD

³ Z Zou, BM

¹ B Ren, MD

⁴ S Peng, MD

¹ H Yuan, MD, PhD

² Q Chen, MD

¹ Department of Ultrasound Medicine and Ultrasonic Medical Engineering Key Laboratory of Nanchong City, Affiliated Hospital of North Sichuan Medical College, Nanchong, China

² Department of Ultrasound, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

³ Department of Orthopedics, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

⁴ Department of Rehabilitation, Second Clinical College of North Sichuan Medical College, Nanchong, China

* Corresponding author: zouyuanxiao@163.com

[#] Equal contribution

New knowledge added by this study

- This is the first study to integrate clinical risk factors with imaging features to optimise the Chinese Thyroid Imaging Reporting and Data System (C-TIRADS) classification.
- This work established a risk threshold-based decision-making framework to guide C-TIRADS classification adjustments.
- External validation demonstrated the model's generalisability across diverse clinical settings.

Implications for clinical practice or policy

- Our model improved diagnostic precision through the integration of imaging and clinical risk factors.
- This research has the potential to optimise resource allocation and reduce interobserver diagnostic variability.

Introduction

Thyroid nodules are a common clinical finding, with a prevalence of approximately 4% to 7% in the general population, and are most often detected

by ultrasonography.^{1,2} Although most thyroid nodules are benign, distinguishing malignant from benign nodules remains a clinical priority to avoid unnecessary procedures and ensure timely

intervention.³ To standardise risk stratification, various Thyroid Imaging Reporting and Data Systems (TIRADS) have been developed,^{4,5} including the ACR-TIRADS (American College of Radiology),⁶ the K-TIRADS (Korean Society of Thyroid Radiology),⁷ and the European Thyroid Association.⁸ Recognising the need for a system tailored to the Chinese healthcare context, the Chinese Artificial Intelligence Alliance for Thyroid and Breast Ultrasound proposed the Chinese TIRADS (C-TIRADS) in 2021.² However, existing TIRADS models primarily focus on sonographic characteristics and often overlook relevant clinical risk factors (eg, patient age, sex, and cervical lymph node [LN] involvement).⁹ In clinical practice, radiologists frequently incorporate such clinical information into their assessments, contributing to inconsistency and variability in TIRADS classification.

Papillary thyroid carcinoma accounts for approximately 80% to 90% of all thyroid cancers and is typically characterised by indolent behaviour.^{10,11} A substantial proportion of new cases involve papillary thyroid microcarcinoma, defined as tumours measuring less than 10 mm in diameter, which generally carry a favourable clinical prognosis.¹² Increasing recognition of the indolent nature of papillary thyroid microcarcinoma has raised concerns regarding potential overdiagnosis and overtreatment. However, current risk stratification strategies that rely solely on imaging features may either overestimate or underestimate malignancy risk, depending on the patient's broader clinical context. Approaches that incorporate clinical risk factors into TIRADS classification could address these limitations and enhance diagnostic accuracy, supporting more individualised patient management.

This study aimed to develop and externally validate a predictive model that integrates both imaging characteristics and clinical risk factors to refine the C-TIRADS classification system. To our knowledge, this is the first nomogram-based model to incorporate clinical risk factors into the C-TIRADS framework. The tool is designed to assist radiologists in improving diagnostic consistency and supporting more informed and individualised clinical decision making in the management of thyroid nodules.

Methods

Study design and population

This retrospective diagnostic study included patients with thyroid nodules who underwent surgical resection at two tertiary hospitals in China. The derivation cohort comprised patients treated at Sichuan Provincial People's Hospital from January to December 2022, while the external validation cohort was drawn from Affiliated Hospital of North Sichuan Medical College during the same period. Inclusion

甲狀腺結節惡性風險的列線圖預測模型的開發與優化策略

賀鵬、梁羽、鄒媛、鄒舟、任波、彭珊、袁紅梅、陳琴

引言：本研究旨在開發及驗證一個臨床預測模型，以協助放射科醫生優化「中國甲狀腺影像報告與資料系統」（C-TIRADS）的診斷分類準則。

方法：本研究納入來自兩所醫院共1659名患有甲狀腺結節患者，當中909名屬於模型開發及內部驗證組（推導組），其餘750名則用作外部驗證組。研究團隊構建了一個二元邏輯迴歸模型，並透過接收者操作特徵（ROC）曲線評估模型的診斷效能，並以列線圖方式視覺化結果。外部驗證則使用ROC及校準曲線來評估模型的鑑別力與準確性。

結果：原始的C-TIRADS分級、頸部淋巴結超聲波異常表現，以及甲狀腺結節大小的變化均為影響C-TIRADS分類優化的重要預測因素。優化後的列線圖模型在ROC曲線下面積（AUC）為0.730（95%置信區間=0.697-0.762），靈敏度為63.2%，特異度為74.9%，整體準確率為67.7%。當預測機率達到或高於60%時，建議提升原始分級；低於30%時則建議調低分級。使用這些門檻值所得的校準曲線顯示模型預測與實際結果吻合良好，而決策曲線分析亦顯示具良好臨床效益。外部驗證結果進一步確認該模型具備優秀的鑑別能力（AUC=0.865；95%置信區間=0.839-0.891）。

結論：優化版C-TIRADS模型結合了甲狀腺結節的影像特徵及臨床風險因素，有助放射科醫生提升診斷準確率及臨床應用效益。

criteria were: (1) thyroid nodules confirmed by postoperative pathology and (2) preoperative ultrasonography of the thyroid and cervical LNs with complete imaging and clinical records. Exclusion criteria were: (1) unclear pathological diagnosis; (2) incomplete clinical data; or (3) poor-quality ultrasound images.

Imaging evaluation and classification

Two junior radiologists, blinded to clinical and pathological information, independently classified all nodules according to the C-TIRADS criteria. Subsequently, two senior radiologists re-evaluated the cases and adjusted the classifications based on additional clinical risk factors, including patient demographics and cervical LN findings. Any modification from the initial C-TIRADS classification was defined as 'classification optimisation' (*C-TIRADS), encompassing both upgrades and downgrades.

Data collection

Structured data collection forms were used to record clinical and sonographic variables. The collected data included patient sex, age, nodule size, number of nodules, C-TIRADS classification, and the presence of abnormal cervical LNs on ultrasonography.

Predictor variables

Sonographic features that directly determine the C-TIRADS score (such as solidity, echogenicity, aspect ratio, microcalcification, and margin irregularity) were not included independently in the multivariable analysis to avoid collinearity. Based on clinical relevance and univariate regression analysis, six predictors were selected for model development, namely, patient sex, age-group (≤ 40 , 40-60, and > 60 years),^{13,14} nodule size, number of nodules (single vs multiple), presence of abnormal cervical LNs, and original C-TIRADS classification.

Model development and internal validation

A binary logistic regression model was developed using the derivation cohort from Sichuan Provincial People's Hospital ($n=909$). For categorical variables with more than two levels, dummy variables were created. The C-TIRADS category 5 was used as the reference group as it represents the highest level of suspicion and the most definitive management pathway (surgical resection), making it an appropriate clinical baseline to estimate relative malignancy risk and the need for reclassification. Model performance in the derivation cohort was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), and calibration was assessed by comparing predicted probability (PP) with observed outcomes using calibration plots.

We emphasise that the primary outcome variable for model training was the pathological diagnosis (binary: malignant vs benign). The C-TIRADS optimisation, defined as upgrading or downgrading the original category based on PP thresholds, was a post-model clinical decision rule applied to the model output, not the outcome used for model development.

Internal validation was performed using bootstrap resampling with 1000 samples to obtain bias-corrected estimates of model performance and 95% confidence intervals (95% CIs). A fixed random seed was set to ensure reproducibility. The bias-corrected C-statistic was 0.728, compared with the original apparent performance of 0.730 (a difference of 0.002), confirming the model's stable discriminative ability (online supplementary Table 1).

External validation

The final model was applied to the external cohort from Affiliated Hospital of North Sichuan Medical College ($n=750$) to evaluate its generalisability. Model discrimination was evaluated by calculating the AUC in the validation set, and calibration was assessed using calibration curves.

Nomogram construction

A nomogram was developed based on the final

multivariable regression model to provide a visual tool for clinical application. Each predictor was assigned a score, and the total score corresponded to the PP of C-TIRADS classification optimisation.

Decision curve analysis and risk thresholds

Decision curve analysis and clinical impact curves were used to evaluate the clinical utility of the nomogram by quantifying the net benefit across a range of threshold probabilities. Specifically, the nomogram generates a PP indicating whether a nodule's original C-TIRADS classification should be modified after integrating clinical information. For clinical decision making, we pre-specified probability cut-offs: PP $\geq 60\%$ (upgrade), PP $< 30\%$ (downgrade), and PP $\geq 30\%$ but $< 60\%$ (unchanged). Based on these thresholds, the model's recommendations were translated into optimised C-TIRADS categories, which were then compared with radiologists' optimisation decisions and surgical pathology findings, as appropriate. These thresholds are reported in the Results section and were applied consistently across all performance tables.

Model performance evaluation

To ensure consistent ROC analysis, all AUCs were calculated using continuous PPs rather than ordinal risk categories. For the original C-TIRADS system, the five-level ordinal classification was transformed into a continuous malignancy probability score using proportional-odds (ordinal logistic) regression. This standard statistical method was employed to model the ordered nature of the C-TIRADS categories and to derive a continuous probability of malignancy for each category, enabling fair comparison in ROC analysis against other models. For the optimised *C-TIRADS system, PPs were directly obtained from the final multivariable logistic regression model. The ROC curves and corresponding AUCs were constructed using these continuous predictions.

Statistical analysis

Statistical analyses and data visualisation were performed using SPSS (Windows version 26.0; IBM Corp, Armonk [NY], United States) and RStudio (version 2022). Categorical variables were reported as number of cases or percentages, with group comparisons conducted using Chi squared test or Fisher's exact test, as appropriate. Multivariable logistic regression analysis was conducted to identify independent predictors. Model discrimination was evaluated using ROC curves, while calibration curves were used to assess model accuracy. Clinical decision and impact curves were established to assess practical clinical utility. A two-tailed P value of < 0.05 was considered statistically significant.

Results

Baseline characteristics

All models were trained to predict pathological malignancy. The optimised ¹⁸C-TIRADS classifications presented here were derived by applying predefined probability thresholds to the model's malignancy predictions.

A total of 1659 patients with thyroid nodules were included in the study, comprising 909 patients in the derivation cohort and 750 in the external validation cohort. In the derivation cohort, 71.8% of patients were women, and the majority (90.8%) had nodules measuring ≤ 30 mm. Approximately 81.7% of patients showed no abnormal cervical LNs on ultrasonography. The rate of C-TIRADS optimisation was 60.6%. In the external validation cohort, similar distributions were observed, with a higher proportion of nodules > 30 mm (Table 1).

Univariate analysis

Univariate binary regression analysis revealed that several variables were either significantly associated ($P < 0.05$) or showed a trend towards association ($0.05 < P < 0.1$) with C-TIRADS optimisation. These variables included patient sex, age, nodule size (10-30 mm), number of nodules, solid composition, blurred margins, aspect ratio > 1 , abnormal cervical LNs, and C-TIRADS category (Table 2 and online supplementary Table 2).

Multivariable model development

A multivariable binary logistic regression model was developed to identify independent predictors associated with C-TIRADS optimisation. Six predictors were independently associated with the outcome. The key predictors of C-TIRADS optimisation were male sex, age 40 to 60 years, thyroid nodule size (per 1-mm increase), multiple thyroid nodules, presence of abnormal cervical LNs, and original C-TIRADS 4A category (online supplementary Table 3). A nomogram model was constructed based on these six independent predictors (Fig 1).

Model performance in the derivation cohort

The model demonstrated good discrimination, with an AUC of 0.730 (95% CI=0.697-0.762) in the derivation cohort (online supplementary Fig a). Internal validation using 1000 bootstrap samples yielded a bias-corrected C-statistic of 0.728, indicating stable model performance (online supplementary Table 1). Calibration curves showed good agreement between PPs and observed outcomes (online supplementary Fig b).

Diagnostic thresholds were evaluated to stratify risk. A PP of $\geq 60\%$ or $< 30\%$ was considered

TABLE 1. Patient and nodule characteristics (n=1659)*

	Derivation population (n=909)	External validation population (n=750)
Sex		
Female	653 (71.8%)	519 (69.2%)
Male	256 (28.2%)	231 (30.8%)
Age, y		
<40	419 (46.1%)	266 (35.5%)
40-60	402 (44.2%)	380 (50.7%)
>60	88 (9.7%)	104 (13.9%)
Nodule size, mm		
<10	483 (53.1%)	253 (33.7%)
10-30	343 (37.7%)	296 (39.5%)
>30	83 (9.1%)	201 (26.8%)
No. of nodules		
Single	654 (71.9%)	413 (55.1%)
Multiple	255 (28.1%)	337 (44.9%)
Composition		
Mixed	57 (6.3%)	196 (26.1%)
Solid	843 (92.7%)	536 (71.5%)
Cystic	9 (1.0%)	18 (2.4%)
Echogenicity		
Isoechoic or hyperechoic	71 (7.8%)	123 (16.4%)
Hypoechoic	566 (62.3%)	411 (54.8%)
Very hypoechoic	272 (29.9%)	216 (28.8%)
Calcification		
None	402 (44.2%)	350 (46.7%)
Micro	458 (50.4%)	358 (47.7%)
Coarse	49 (5.4%)	42 (5.6%)
Blurred margin		
No	291 (32.0%)	424 (56.5%)
Yes	618 (68.0%)	326 (43.5%)
Aspect ratio >1		
No	613 (67.4%)	140 (18.7%)
Yes	296 (32.6%)	610 (81.3%)
Abnormal cervical LN		
No	743 (81.7%)	605 (80.7%)
Yes	166 (18.3%)	145 (19.3%)
C-TIRADS category		
3	62 (6.8%)	181 (24.1%)
4A	107 (11.8%)	124 (16.5%)
4B	162 (17.8%)	104 (13.9%)
4C	518 (57.0%)	296 (39.5%)
5	60 (6.6%)	45 (6.0%)
Outcome measures		
Classification optimisation	551 (60.6%)	492 (65.6%)
Upgrade	333 (36.6%)	327 (43.6%)
Downgrade	218 (24.0%)	165 (22.0%)
Classification unchanged	358 (39.4%)	258 (34.4%)

Abbreviations: C-TIRADS = Chinese Thyroid Imaging Reporting and Data System; LN = lymph node

* Data are shown as No. (%)

TABLE 2. Predictor distribution and univariate logistic regression odds ratios for malignancy (n=909)

	Optimisation (n=551)	Unchanged (n=358)	Odds ratio (95% CI)	P value
Sex				
Female	384 (69.7%)	269 (75.1%)	1.00 (Ref)	
Male	167 (30.3%)	89 (24.9%)	1.31 (0.97-1.77)	0.08
Age, y				
<40	257 (46.6%)	162 (45.3%)	1.51 (0.95-2.40)	0.07
40-60	249 (45.2%)	153 (42.7%)	1.55 (0.97-2.47)	0.06
>60	45 (8.2%)	43 (12.0%)	1.00 (Ref)	N/A
Nodule size, mm				
<10	297 (53.9%)	186 (51.9%)	1.42 (0.89-2.27)	0.14
10-30	221 (40.1%)	122 (34.1%)	1.58 (0.97-2.56)	0.06
>30	33 (6.0%)	50 (14.0%)	1.00 (Ref)	N/A
No. of nodules				
Multiple	161 (29.2%)	94 (26.3%)	1.30 (0.96-1.76)	0.08
Single	390 (70.8%)	264 (73.7%)	1.00 (Ref)	N/A
Composition				
Solid	542 (98.4%)	301 (84.1%)	10.12 (1.12-91.34)	0.04
Cystic	1 (0.2%)	8 (2.2%)	1.30 (0.14-11.89)	0.81
Mixed	8 (1.5%)	49 (13.7%)	1.00 (Ref)	N/A
Echogenicity				
Isoechoic or hyperechoic	42 (7.6%)	29 (8.1%)	1.00 (Ref)	N/A
Hypoechoic	351 (63.7%)	215 (60.1%)	1.12 (0.86-1.86)	0.64
Very hypoechoic	158 (28.7%)	114 (31.8%)	0.95 (0.56-1.62)	0.87
Calcification type				
None	238 (43.2%)	164 (45.8%)	1.00 (Ref)	N/A
Micro	280 (50.8%)	178 (49.7%)	0.70 (0.37-1.32)	0.27
Coarse	33 (6.0%)	16 (4.5%)	0.76 (0.40-1.42)	0.39
Blurred margin				
Yes	389 (70.6%)	229 (64.0%)	0.73 (0.55-0.98)	0.04
No	162 (29.4%)	129 (36.0%)	1.00 (Ref)	N/A
Aspect ratio >1				
Yes	165 (29.9%)	131 (36.6%)	1.35 (1.01-1.79)	0.03
No	386 (70.1%)	227 (63.4%)	1.00 (Ref)	N/A
Abnormal cervical LN				
Yes	136 (24.7%)	30 (8.4%)	0.27 (0.18-0.42)	<0.01
No	415 (75.3%)	328 (91.6%)	1.00 (Ref)	N/A
C-TIRADS category				
3	6 (1.1%)	56 (15.6%)	0.15 (0.05-0.40)	<0.01
4A	80 (14.5%)	27 (7.5%)	4.14 (2.11-8.13)	<0.01
4B	114 (20.7%)	48 (13.4%)	3.32 (1.79-6.14)	<0.01
4C	326 (59.2%)	192 (53.6%)	2.37 (1.38-4.09)	<0.01
5	25 (4.5%)	35 (9.8%)	1.00 (Ref)	N/A

Abbreviations: C-TIRADS = Chinese Thyroid Imaging Reporting and Data System; LN = lymph node; N/A = not applicable; Ref = reference

* Data are shown as No. (%), unless otherwise specified

indicative of a high likelihood of classification change: a PP of $\geq 60\%$ suggested upgrading, while a PP of $<30\%$ suggested downgrading; PPs between 30% and 60% indicated that the classification was likely to remain unchanged. A detailed summary of sensitivity, specificity, and overall accuracy across these thresholds is presented in online supplementary Table 4.

External validation

When applied to the external cohort, the model achieved an AUC of 0.865 (95% CI=0.839-0.891) [online supplementary Fig c], demonstrating excellent generalisability. Calibration plots again confirmed close agreement between predicted and observed probabilities (online supplementary Fig d). At the 60% probability threshold, sensitivity was 85.0%, specificity was 69.0%, and overall accuracy was 79.7% in the external validation cohort. Diagnostic performance metrics across various risk thresholds of the final prediction model were analysed in the external validation population (online supplementary Table 5).

Clinical utility

Decision curve analysis (Fig 2a) demonstrated that the nomogram model provided greater net clinical benefit across a wide range of threshold probabilities compared with treating all or no patients. The clinical impact curve (Fig 2b) showed that the number of true positives closely approximated the predicted number across relevant thresholds. The observed distribution of histopathological outcomes was as follows: in the derivation cohort, 769 nodules (84.6%) were confirmed malignant and 140 (15.4%) were benign; in the validation cohort, 434 nodules (57.9%) were malignant and 316 (42.1%) were benign.

Comparison of diagnostic efficacy between the original C-TIRADS and optimised C-TIRADS classifications demonstrated superior performance of the optimised model in both the derivation and validation cohorts (Fig 2c and d, respectively). The optimised classification achieved higher AUC values for differentiating benign from malignant nodules (AUC=0.97 vs 0.94 in the derivation cohort; AUC=0.97 vs 0.95 in the external validation cohort). The predictive model tended to improve C-TIRADS classification by upgrading category 4A nodules to category 4B or 4C, reflecting enhanced clinical utility (Table 3 and Fig 2).

Application example of the nomogram model

A 55-year-old man underwent ultrasound examination, which revealed a solid hypoechoic thyroid nodule in the right lobe measuring approximately $7.1 \times 6.4 \text{ mm}^2$ (Fig 3a). Simultaneously, abnormal LNs were detected on the ipsilateral side of

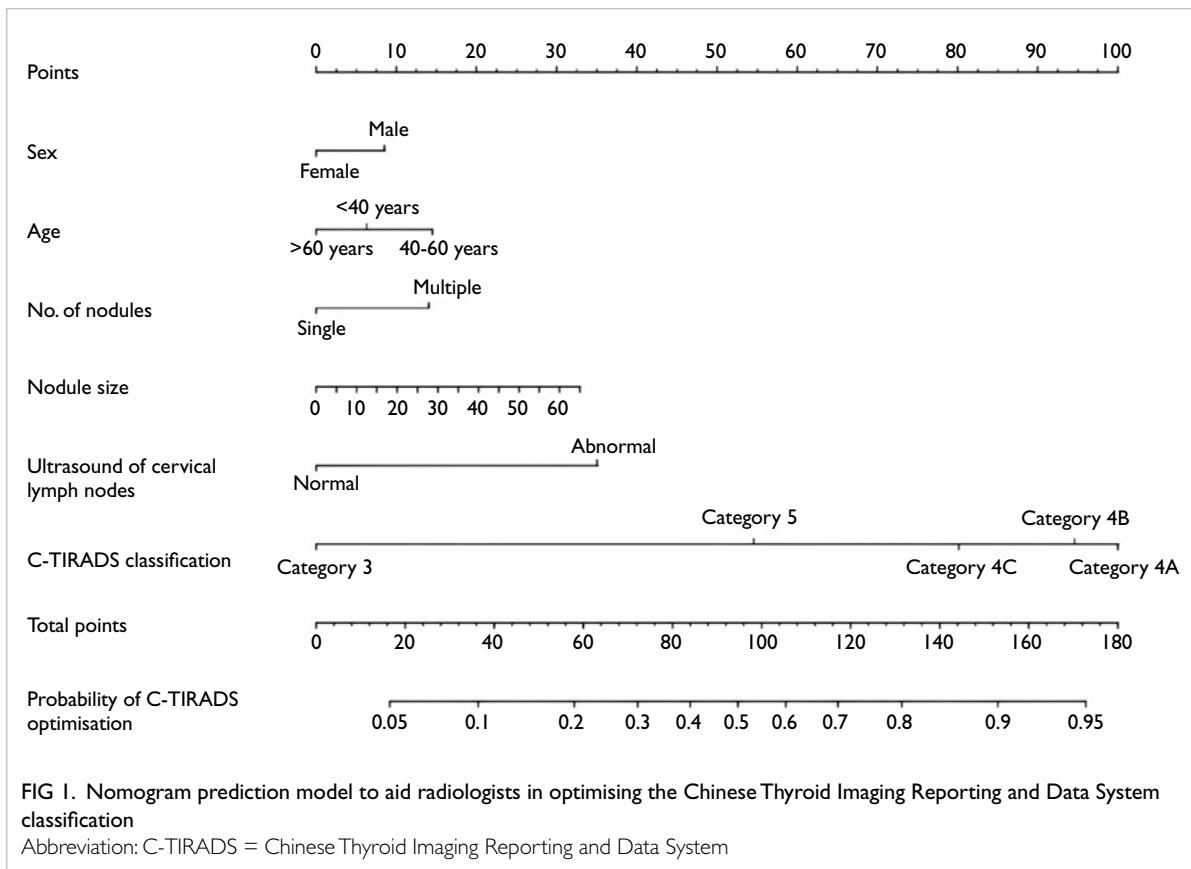


TABLE 3. Clinical diagnostic performance of the final predictive model in thyroid nodules (n=1659)*

C-TIRADS category	Derivation population (n=909)				External validation population (n=750)			
	Original C-TIRADS		Optimised C-TIRADS		Original C-TIRADS		Optimised C-TIRADS	
	Benign	Malignant	Benign	Malignant	Benign	Malignant	Benign	Malignant
3	58 (6.4%)	4 (0.4%)	100 (11.0%)	5 (0.6%)	178 (23.7%)	3 (0.4%)	193 (25.7%)	4 (0.5%)
4A	55 (6.1%)	52 (5.7%)	29 (3.2%)	45 (5.0%)	97 (12.9%)	27 (3.6%)	94 (12.5%)	9 (1.2%)
4B	22 (2.4%)	140 (15.4%)	9 (1.0%)	178 (19.6%)	30 (4.0%)	74 (9.9%)	21 (2.8%)	91 (12.1%)
4C	5 (0.6%)	513 (56.4%)	2 (0.2%)	274 (30.1%)	11 (1.5%)	285 (38.0%)	6 (0.8%)	92 (12.3%)
5	0	60 (6.6%)	0	267 (29.4%)	0	45 (6.0%)	2 (0.3%)	238 (31.7%)
AUC	0.93; P<0.01		0.97; P<0.01		0.95; P<0.01		0.97; P<0.01	

Abbreviations: AUC = area under the receiver operating characteristic curve; C-TIRADS = Chinese Thyroid Imaging Reporting and Data System

* Data are shown as No. (%), unless otherwise specified

the neck, characterised by indistinct corticomedullary differentiation and suspected microcalcifications (Fig 3b). According to the conventional C-TIRADS system, the nodule was initially classified as category 4B. However, application of the nomogram model yielded a cumulative score of 155 points, corresponding to a malignancy risk of >90%. Based on this result, the TIRADS category was optimised and upgraded to category 5 (Fig 3c). Subsequent

histopathological examination confirmed the diagnosis of papillary thyroid microcarcinoma with cervical LN metastasis.

Discussion

This study retrospectively analysed the sonographic characteristics and clinical risk factors of 1659 thyroid nodules from two large tertiary hospitals in western

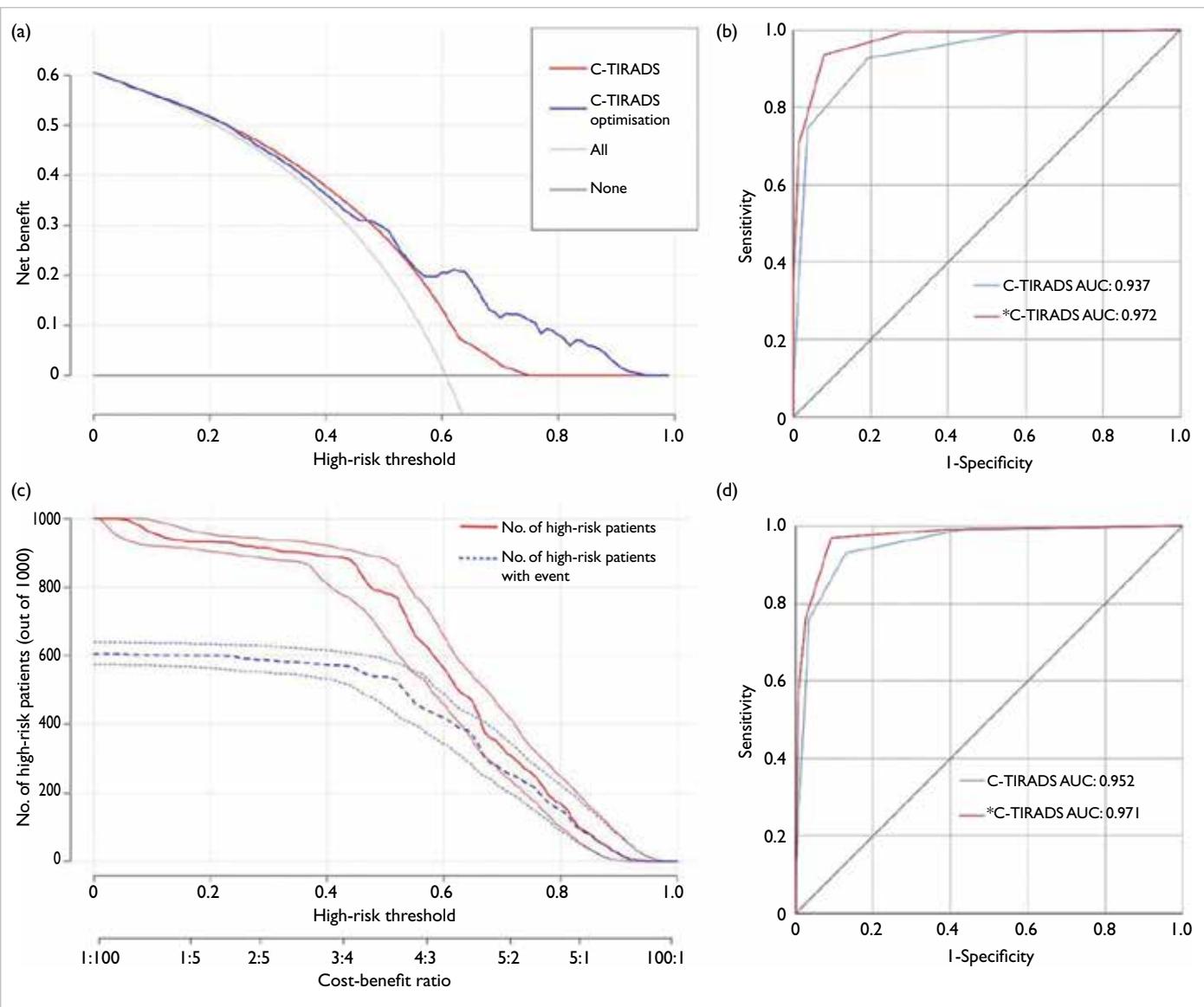


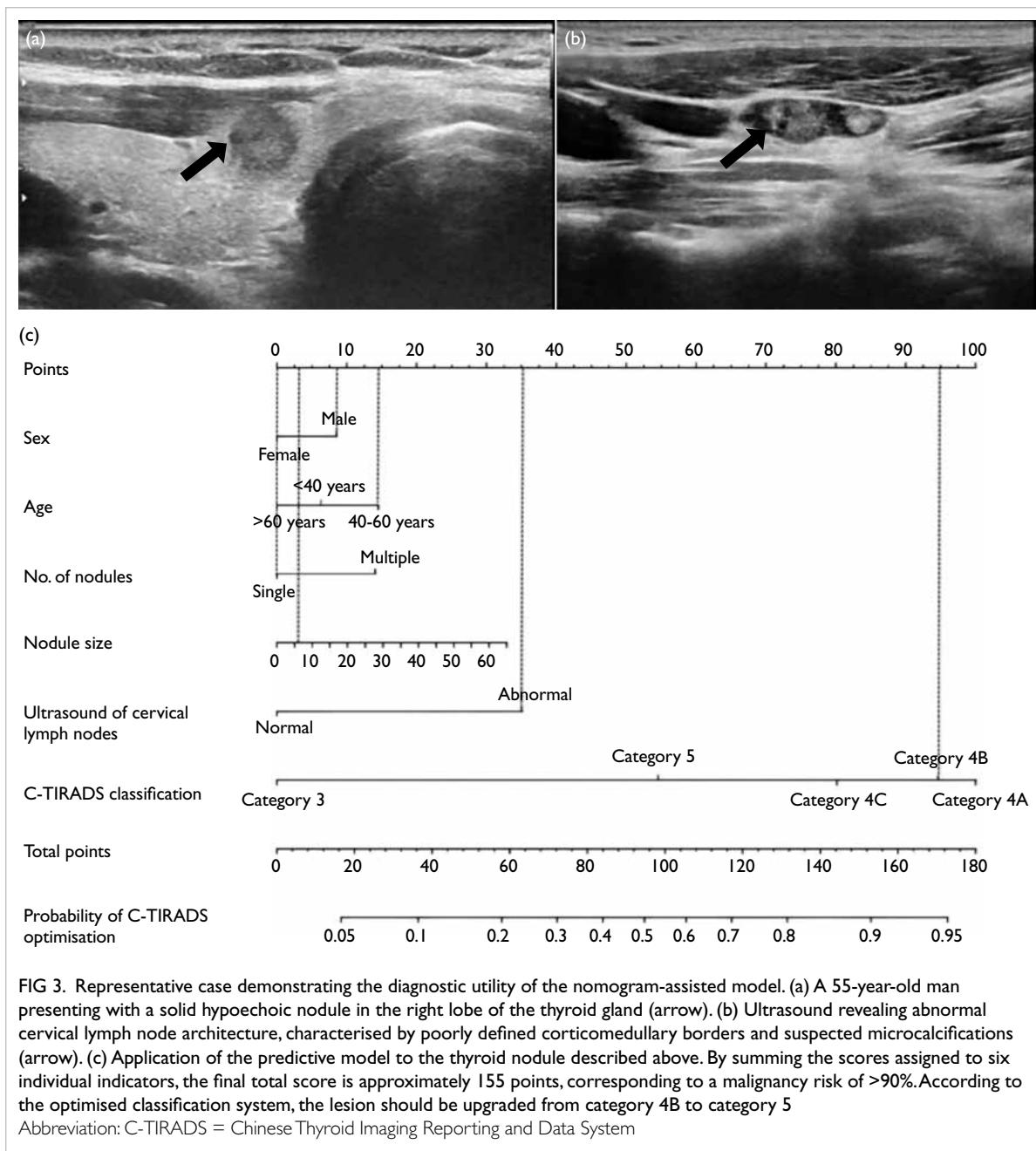
FIG 2. Comparison of the diagnostic efficacy of the Chinese Thyroid Imaging Reporting and Data System (C-TIRADS) and optimised C-TIRADS (*C-TIRADS) in the diagnosis of benign and malignant thyroid nodules. (a) Clinical decision curve of the predictive model for radiologist-optimised C-TIRADS classification in the derivation cohort. (b) Comparison of the diagnostic efficacy of C-TIRADS and *C-TIRADS for the diagnosis of benign and malignant thyroid nodules in the derivation cohort. (c) Clinical impact curves of the predictive model for radiologist-optimised C-TIRADS classification in the derivation cohort, showing the number of patients classified as high risk (solid curve) and the number of true positives among them (dashed curve) across probability thresholds. (d) Comparison of the diagnostic efficacy of C-TIRADS and *C-TIRADS for the diagnosis of benign and malignant thyroid nodules in the validation cohort

Abbreviation: AUC = area under the receiver operating characteristic curve

China, with the aim of optimising the C-TIRADS classification. A predictive model integrating clinical parameters and imaging features was developed and externally validated, demonstrating high diagnostic performance (AUC=0.865 in external validation) and clinical benefit, as evidenced by decision curve analysis.

Despite the widespread adoption of various TIRADS frameworks globally,^{2-4,8} fundamental methodological limitations persist. Current models, such as ACR-TIRADS,⁶ primarily focus

on ultrasound features and rely heavily on consensus-driven rather than statistically validated risk stratification systems.^{6,15} Although TIRADS demonstrates robust sensitivity in clinical settings, its specificity remains relatively limited.¹⁶ Interobserver variability is another key concern—radiologists' subjective interpretation of ultrasound features can result in inconsistent classification outcomes.¹⁷ To address these limitations, various strategies have been proposed, including the integration of artificial intelligence techniques to reduce



observer subjectivity.¹⁸⁻²⁰ Artificial intelligence has shown promise in matching or even surpassing the specificity achieved by radiologists; however, their clinical implementation remains constrained by challenges in interpretability and low acceptance in routine practice.

Integrating clinical risk factors may enhance risk stratification for thyroid nodules, as suggested by a growing body of evidence.²¹ In alignment with this, our study incorporated clinical variables including patient age, sex, number of nodules, and cervical LN status into the predictive model, thereby more accurately reflecting routine clinical diagnostic

workflows. While previous studies²²⁻²⁴ suggested that male patients with thyroid nodules, particularly those with indeterminate fine-needle aspiration cytology undergoing molecular testing, exhibit a higher malignancy risk,²⁵ our study did not identify a significant difference in thyroid cancer incidence between sexes. This discrepancy may be attributable to methodology differences, as molecular testing was not performed in our cohort and all diagnoses were confirmed through postoperative histopathology. The absence of statistical significance for male sex may reflect population-specific characteristics, such as regional variation in risk factor distribution

or age composition.²⁶ These methodological and demographic differences may have attenuated the observed sex-related effect. Nonetheless, male patients in our study were assigned higher risk scores, suggesting an association with malignancy risk, despite the lack of statistical significance.

Compared with previous models that primarily focused on intrinsic ultrasound features of thyroid nodules,²⁷⁻²⁹ our nomogram offers a more comprehensive assessment. Although the individual contributions of factors such as sex and age were relatively modest, they reflected subtle clinical patterns often considered by radiologists during decision making. The C-TIRADS optimisation approach demonstrated clear advantages, particularly in reducing unnecessary invasive procedures without compromising diagnostic accuracy, achieving an AUC of 0.972. Furthermore, the new model indicated that a risk threshold of $\geq 60\%$ favoured the recommendation for C-TIRADS optimisation, whereas a threshold of $< 30\%$ favoured exclusion. The integration of complex imaging data with clinical information represents a core competency for radiologists.³⁰ With appropriate standardised training and communication frameworks in place, radiologists are well positioned to leverage quantitative metrics generated by the new model into routine diagnostic workflows. This advancement holds promise for improving diagnostic consistency and accuracy in clinical practice.

Limitations

This study has several limitations that should be acknowledged. First, the optimisation of the TIRADS classification was influenced by radiologists' subjective judgement, which may have contributed to interobserver variability. Second, although data collection was conducted by trained junior radiologists, observer variation and the subjective nature of ultrasound interpretation may have affected the model's performance.³¹ Third, internal validation using bootstrap resampling may have overestimated model performance due to potential overfitting; therefore, external validation was essential to confirm generalisability. Fourth, owing to the retrospective design, only a limited set of clinical parameters (eg, sex, age, and cervical LN status) was included. Other relevant factors such as body mass index, environmental exposures, nodule location, family history of thyroid cancer, and radiation exposure history,^{32,33} were not assessed. Finally, the study cohort exclusively comprised cases confirmed by surgical pathology, resulting in a relatively low proportion of benign lesions, which may have introduced selection bias. The exclusion of patients diagnosed solely by fine-needle aspiration was intentional but may have affected the generalisability of the findings.

Future directions

To address the limitations of the present study, future research should aim to standardise the application of TIRADS by adopting unified classification frameworks and implementing regular training programmes to enhance interobserver consistency. Prospective multicentre studies involving broader and more diverse populations are warranted, incorporating a wider range of clinical risk factors to improve predictive accuracy. In particular, data regarding family history, radiation exposure, and other relevant variables across centres would support more comprehensive risk assessment and enhance the generalisability of prediction models. In addition, including patients with fine-needle aspiration-confirmed benign nodules may help achieve a more balanced representation of benign and malignant cases. The development and application of nomogram-based structured training programmes for radiologists could also be explored to further improve diagnostic consistency and clinical utility. While the widespread adoption of a revised classification system will require time, we hope that the findings of this study may contribute to that transition.

Conclusion

We developed and externally validated a nomogram-based predictive model that integrates imaging features and clinical risk factors to optimise C-TIRADS classification for thyroid nodules. The model demonstrated good discrimination and calibration across internal and external cohorts, offering a practical tool to assist radiologists in refining diagnostic assessments and improving clinical decision making. Future research incorporating additional clinical variables and prospective validation is warranted to further strengthen the model's applicability across diverse clinical settings.

Author contributions

Concept or design: Y Liang, Y Zou, P He, Q Chen.
Acquisition of data: Y Liang, Y Zou, Z Zou, B Ren.
Analysis or interpretation of data: Y Liang, S Peng, Y Zou.
Drafting of the manuscript: Y Liang, Y Zou, HM Yuan, Z Zou.
Critical revision of the manuscript for important intellectual content: P He, Y Zou.

All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Conflicts of interest

The authors have disclosed no conflicts of interest.

Declaration

This manuscript was initially posted as a preprint entitled

'Development and validation of a clinical prediction model to aid radiologists optimize thyroid C-TIRADS classification' on Research Square (DOI: 10.21203/rs.3.rs-3831900/v1). After peer feedback and extensive revisions undertaken collaboratively by the author team, the current version has substantially evolved and markedly differs from the preprint version.

Funding/support

This research was supported by Sichuan Science and Technology Program (Ref Nos.:2025ZNSFSC1751, 2026YFHZ0039), the University-Industry Collaborative Education Program (Ref No.: 250505236300920), the University-level Project of North Sichuan Medical College (Ref Nos.: CXSY24-06, CBY22-QNA48), and the Hospital-level Projects of the Affiliated Hospital of North Sichuan Medical College, China (Ref Nos.: 210930, 2023-2GC013, 2025LC010). The funders had no role in the study design, data collection/analysis/interpretation, or manuscript preparation.

Ethics approval

This research was approved by the Ethics Committee of Sichuan Provincial People's Hospital (Ref No.: ER20210347) and the Ethics Committee of Affiliated Hospital of North Sichuan Medical College, China (Ref No.: 2021ER436-1). The requirement for informed patient consent was waived by both Committees due to the retrospective nature of the research.

Supplementary material

The supplementary material was provided by the authors, and some information may not have been peer reviewed. Accepted supplementary material will be published as submitted by the authors, without any editing or formatting. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by the Hong Kong Academy of Medicine and the Hong Kong Medical Association. The Hong Kong Academy of Medicine and the Hong Kong Medical Association disclaim all liability and responsibility arising from any reliance placed on the content. To view the file, please visit the journal online (<https://doi.org/10.12809/hkmj2512718>).

References

1. Haugen BR, Alexander EK, Bible KC, 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.
2. Zhou J, Song Y, Zhan W, et al. Thyroid imaging reporting and data system (TIRADS) for ultrasound features of nodules: multicentric retrospective study in China. *Endocrine* 2021;72:157-70.
3. Trimboli P. Complexity in the interpretation and application of multiple guidelines for thyroid nodules: the need for coordinated recommendations for "small" lesions. *Rev Endocr Metab Disord* 2025;26:223-7.
4. Park JY, Lee HJ, Jang HW, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. *Thyroid* 2009;19:1257-64.
5. Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab* 2009;94:1748-51.
6. Tessler FN, Middleton WD, Grant EG, 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.
7. Shin JH, Baek JH, Chung J, 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.
8. 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.
9. Chen Z, Wang JJ, Du JB, et al. Development and validation of a dynamic nomogram for predicting central lymph node metastasis in papillary thyroid carcinoma patients based on clinical and ultrasound features. *Quant Imaging Med Surg* 2025;15:1555-70.
10. Boucail L, Zafereo M, Cabanillas ME. Thyroid cancer: a review. *JAMA* 2024;331:425-35.
11. Zhang J, Xu S. High aggressiveness of papillary thyroid cancer: from clinical evidence to regulatory cellular networks. *Cell Death Discov* 2024;10:378.
12. Ma T, Semsarian CR, Barratt A, et al. Rethinking low-risk papillary thyroid cancers <1 cm (papillary microcarcinomas): an evidence review for recalibrating diagnostic thresholds and/or alternative labels. *Thyroid* 2021;31:1626-38.
13. Kwong N, Medici M, Angell TE, et al. The influence of patient age on thyroid nodule formation, multinodularity, and thyroid cancer risk. *J Clin Endocrinol Metab* 2015;100:4434-40.
14. Pizzato M, Li M, Vignat J, et al. The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. *Lancet Diabetes Endocrinol* 2022;10:264-72.
15. Tessler FN, Middleton WD, Grant EG, Hoang JK. Re: ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 2018;15(3 Pt A):381-2.
16. Angelopoulos N, Gouliis DG, Chrisogonidis I, et al. Diagnostic performance of European and American College of Radiology Thyroid Imaging Reporting and Data System classification systems in thyroid nodules over 20 mm in diameter. *Endocr Pract* 2025;31:72-9.
17. Jin Z, Pei S, Shen H, et al. Comparative study of C-TIRADS, ACR-TIRADS, and EU-TIRADS for diagnosis and management of thyroid nodules. *Acad Radiol* 2023;30:2181-91.
18. Wildman-Tobriner B, Buda M, Hoang JK, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology* 2019;292:112-9.
19. Wu SH, Li MD, Tong WJ, et al. Adaptive dual-task deep learning for automated thyroid cancer triaging at screening US. *Radiol Artif Intell* 2025;7:e240271.
20. Trimboli P, Colombo A, Gamarra E, Ruinelli L, Leoncini A. Performance of computer scientists in the assessment of thyroid nodules using TIRADS lexicons. *J Endocrinol Invest* 2025;48:877-83.
21. Kobaly K, Kim CS, Mandel SJ. Contemporary management

of thyroid nodules. *Annu Rev Med* 2022;73:517-28.

22. Xu L, Li G, Wei Q, El-Naggar AK, Sturgis EM. Family history of cancer and risk of sporadic differentiated thyroid carcinoma. *Cancer* 2012;118:1228-35.
23. Iglesias ML, Schmidt A, Ghuzlan AA, et al. Radiation exposure and thyroid cancer: a review. *Arch Endocrinol Metab* 2017;61:180-7.
24. Saenko V, Mitsutake N. Radiation-related thyroid cancer. *Endocr Rev* 2024;45:1-29.
25. Figge JJ, Gooding WE, Steward DL, et al. Do ultrasound patterns and clinical parameters inform the probability of thyroid cancer predicted by molecular testing in nodules with indeterminate cytology? *Thyroid* 2021;31:1673-82.
26. Li X, Xing M, Tu P, et al. Urinary iodine levels and thyroid disorder prevalence in the adult population of China: a large-scale population-based cross-sectional study. *Sci Rep* 2025;15:14273.
27. Xiao J, Xiao Q, Cong W, et al. Discriminating malignancy in thyroid nodules: the nomogram versus the Kwak and ACR TI-RADS. *Otolaryngol Head Neck Surg* 2020;163:1156-65.
28. Xin Y, Liu F, Shi Y, Yan X, Liu L, Zhu J. A scoring system for assessing the risk of malignant partially cystic thyroid nodules based on ultrasound features. *Front Oncol* 2021;11:731779.
29. Zhou T, Hu T, Ni Z, et al. Comparative analysis of machine learning-based ultrasound radiomics in predicting malignancy of partially cystic thyroid nodules. *Endocrine* 2024;83:118-26.
30. Bluethgen C, Van Veen D, Zakka C, et al. Best practices for large language models in radiology. *Radiology* 2025;315:e240528.
31. He Z, Li Y, Zeng W, et al. Can a computer-aided mass diagnosis model based on perceptive features learned from quantitative mammography radiology reports improve junior radiologists' diagnosis performance? An observer study. *Front Oncol* 2021;11:773389.
32. Kim Y, Roh J, Song DE, et al. Risk factors for posttreatment recurrence in patients with intermediate-risk papillary thyroid carcinoma. *Am J Surg* 2020;220:642-7.
33. Zhao J, Wen J, Wang S, Yao J, Liao L, Dong J. Association between adipokines and thyroid carcinoma: a meta-analysis of case-control studies. *BMC Cancer* 2020;20:788.