Am J Cancer Res 2026;16(7):2814-2826
www.ajcr.us /ISSN:2156-6976/ajcr0171885

Original Article

Association between ultrasound radiomics features and central lymph node metastasis in patients with papillary thyroid carcinoma

Zexiang Zhang, Zequn Wang, Wenjie Li, Zhaolin Yin

Department of Ultrasound, The Fourth Hospital of Hebei Medical University, Shijiazhuang 050011, Hebei, China

Received February 4, 2026; Accepted June 16, 2026; Epub July 15, 2026; Published July 30, 2026

Abstract: Objective: To develop and validate a predictive model combining ultrasound radiomics score (Rad-score) with clinical and conventional ultrasound features for preoperative assessment of central lymph node metastasis (CLNM) risk in papillary thyroid carcinoma (PTC). Methods: This retrospective study included 500 PTC patients, randomly assigned (7:3) to training (n = 350) and validation (n = 150) sets. Radiomics features underwent repeatability validation, standardization, variance filtering, correlation-based redundancy removal, and LASSO logistic regression to construct the Rad-score. Four models were established: clinical/laboratory (Model A), conventional ultrasound (Model B), radiomics (Model C), and combined (Model D). Model performance was assessed using ROC curves, calibration curves, and decision curve analysis (DCA). Results: Baseline analysis revealed that the CLNM group had a higher proportion of males (29.14% vs. 18.59%, P = 0.0286) and a significantly younger mean age (44.15±12.03 years vs. 49.25±11.80 years, P<0.001). Conventional ultrasonographic features demonstrated that the CLNM group presented higher rates of maximum tumor diameter >10 mm (57.62% vs. 31.66%, P<0.001), calcification (74.83% vs. 54.77%, P<0.001) and multifocality (51.66% vs. 35.68%, P = 0.0039). A total of 10 radiomic features were selected via the LASSO algorithm to establish the Rad-score. Univariate logistic regression analysis indicated that male gender, age ≥55 years, maximum tumor diameter >10 mm, calcification, multifocality and Rad-score were correlated with CLNM. Multivariate logistic regression analysis showed that age ≥55 years was associated with a reduced risk of CLNM (OR = 0.531, 95% CI: 0.311-0.907, P = 0.020). Tumor maximum diameter >10 mm (OR = 1.949, 95% CI: 1.192-3.187, P = 0.008) and Rad-score (OR = 35.074, 95% CI: 10.385-118.458, P<0.001) were correlated with an elevated risk of CLNM. Model performance evaluation showed that the areas under the curve (AUCs) of Model A, B, C and D were 0.632, 0.687, 0.757 and 0.764 in the training set, and 0.643, 0.702, 0.788 and 0.798 in the validation set, respectively. Conclusion: The combined model incorporating Rad-score, age stratification, maximum tumor diameter, and conventional ultrasound features effectively predicts CLNM risk in PTC patients, demonstrating good discrimination, calibration, and clinical net benefit. It has potential value for preoperative risk stratification and individualized decision support.

Keywords: Papillary thyroid carcinoma, central lymph node metastasis, ultrasound, radiomics, nomogram

Introduction

Papillary thyroid carcinoma (PTC) is a common malignant tumor of the head and neck. Its incidence has increased rapidly in recent years, posing a serious threat to global public health [1-3]. Lymph node metastasis is a frequent clinicopathological feature of PTC and affects patient recurrence and long-term prognosis [4-6]. Central lymph nodes are the most common metastatic site of PTC. Therefore, accurate preoperative assessment of central lymph node metastasis (CLNM) is crucial for developing individualized treatment regimens [7]. At present, conventional ultrasound is the main modality for evaluating CLNM in clinical practice, but its sensitivity and specificity are limited for metastatic lesions with atypical morphological changes [8]. Although ultrasound-guided fine-needle aspiration biopsy or thyroglobulin measurement in washout fluid can improve diagnostic accuracy, these are invasive exami-

https://doi.org/10.62347/SRFQ9286

nations that demand proficient operational skills, making large-scale clinical application difficult.

In recent years, radiomic techniques have developed rapidly and been widely applied in the diagnosis and prognostic evaluation of various diseases [9]. Previous studies have shown that radiomics converts medical images into quantifiable high-throughput features and captures tissue heterogeneity invisible to the naked eye, which helps improve the accuracy of risk stratification [10-12]. An increasing number of studies have focused on preoperative CLNM risk stratification of PTC based on ultrasound radiomics. However, systematic evidence is still insufficient regarding stepwise head-to-head comparisons of clinical features, conventional ultrasonographic features, radiomic features and their combined models in the same cohort, as well as the assessment of clinical net benefit via decision curve analysis (DCA) and model visualization using nomograms. Accordingly, this study aimed to construct a radiomics score (Rad-score) based on ultrasound radiomics. By combining clinical and conventional ultrasonographic features, we constructed a predictive model and nomogram to provide evidence for preoperative risk evaluation and individualized clinical decision-making of CLNM in PTC patients.

Materials and methods

Study population

This retrospective study enrolled 500 patients who received thyroid surgery at the Fourth Hospital of Hebei Medical University from January 2021 to April 2024 and were diagnosed with PTC via postoperative pathological examination. All patients were divided into a training cohort (n = 350) and a validation cohort (n = 150) at a 7:3 ratio, ensuring a similar proportion of CLNM in both cohorts. The study was approved by the Ethics Committee of the Fourth Hospital of Hebei Medical University.

CLNM determination criteria

Postoperative pathology was defined as the gold standard. The dissected central lymph nodes were fixed, embedded, sliced and stained with hematoxylin and eosin (H&E). All pathological slides were independently reviewed by two senior pathologists. Lymph node metastasis of PTC was defined as the presence of tumor cells with typical nuclear features of PTC. In case of inconsistent results, the two pathologists reviewed the slides together, or a third senior pathologist was invited to make the final judgment.

Inclusion and exclusion criteria

Inclusion criteria: (1) Pathologically confirmed PTC; (2) Underwent thyroidectomy combined with central lymph node dissection (CLND); (3) Preoperative conventional ultrasound with clear images, as well as complete clinical and laboratory data; (4) Age ≥18 years old.

Exclusion criteria: (1) History of thyroid or neck surgery; (2) Preoperative radiotherapy, radiofrequency ablation, or systemic therapy; (3) Concurrent malignant tumors of other sites or non-PTC thyroid lesions; (4) Pregnant or lactating women; (5) Lymph node skip metastasis; (6) Poor image quality.

Ultrasound examination and image analysis

A Siemens Acuson S3000 color Doppler ultrasound system equipped with an 18L6HD high-frequency linear array probe was adopted. The probe frequency ranged from 6 to 18 MHz, the mechanical index was less than 1.2, and the scanning depth was set at 3-5 cm. To minimize the influence of interpatient differences in imaging parameters on the stability of radiomic features, we kept consistent equipment model, probe frequency, gain, scanning depth and focus position during image acquisition and export. Two-dimensional grayscale ultrasound images with clear lesions, relatively complete boundaries and no obvious artifacts were selected for subsequent analysis.

Two senior ultrasonographers analyzed all images independently in a double-blind method. Ultrasound features of thyroid nodules were recorded in accordance with the terminology and classification criteria of American College of Radiology Thyroid Imaging, Reporting and Data System, including maximum tumor diameter, aspect ratio, ill-defined margins, calcification, extrathyroidal extension and multifocality. For patients with multiple lesions, data of the largest lesion were adopted [13]. Any disagreement between the two physicians was resolved

by a third senior ultrasonographer. The Kappa value for inter-observer agreement of conventional ultrasonographic features was 0.91.

Collection of clinical and laboratory data

Collected data included gender, age, BRAF V600E mutation status, thyroid antibodies and peripheral blood cell counts. Neutrophil and lymphocyte counts were recorded, and the neutrophil-to-lymphocyte ratio (NLR) was calculated. Thyroglobulin antibody (TGAb) titers presented a skewed distribution, so logarithmic transformation of ln(TGAb+1) was performed before regression analysis. Thyroid peroxidase antibody (TPOAb) >5.0 IU/mL and TGAb >4.0 IU/mL were defined as antibody positivity.

Age was analyzed as a continuous variable for baseline characteristic description and intergroup comparison. For logistic regression analysis and predictive model construction, age was dichotomized into a binary variable (age ≥55 years; yes = 1, no = 0) according to the commonly used age cutoff values for thyroid cancer staging and clinical risk stratification. To avoid duplicate inclusion of variables, continuous age was not incorporated into univariate and multivariate logistic regression analyses.

Region of interest (ROI) delineation and radiomic feature reproducibility validation

ROIs were manually delineated slice by slice along the margins of solid components by two senior radiologists or sonographers experienced in thyroid ultrasound. All operators were blinded to pathological results and CLNM status. Obvious acoustic shadowing, cystic changes, necrotic areas and adjacent normal thyroid tissue were excluded during delineation. For multifocal lesions, the largest lesion was selected for radiomic analysis.

To evaluate feature reproducibility, a random subset of images was re-delineated by the same operator after a two-week interval for intra-observer agreement assessment. Meanwhile, another operator delineated ROIs independently to evaluate inter-observer agreement. The intraclass correlation coefficient (ICC) was adopted to assess feature stability. Features with ICC >0.75 were regarded as having good reproducibility and included in the subsequent feature selection.

Radiomic feature selection and rad-score construction

Prior to radiomic feature extraction, grayscale standardization and resampling were performed on ultrasound images to reduce the effects of inter-image differences in grayscale distribution and pixel scale on feature calculation. All features were standardized in the training set, and the identical parameters were applied to the validation set.

The feature selection workflow was as follows. First, low-variance features were removed using the variance threshold method. Second, highly correlated redundant features were eliminated via correlation analysis. Third, features with good reproducibility were enrolled into the LASSO logistic regression model. LASSO regression was conducted in the training set, and 10-fold cross-validation was used to determine the optimal penalty parameter λ. For the final model, C = 0.03255, corresponding to λ = 30.722 (log λ = 3.425). A total of 10 radiomic features with non-zero coefficients were screened out under this penalty strength. The Rad-score was calculated as the weighted sum of the standardized values of the selected features and their corresponding LASSO regression coefficients. The LASSO coefficient path plot and 10-fold cross-validation curve were used to visualize the feature selection process.

Model construction

Four models were built. Model A (clinical/laboratory) was based on preoperative clinical and laboratory indicators. Model B (conventional ultrasound) was based on conventional ultrasound features. Model C (radiomics) incorporated only the Rad-score. Model D was a combined model, which integrated the Rad-score with variables that were statistically significant and clinically interpretable in the multivariate logistic regression analysis and was visualized as a nomogram. Since calcification and multifocality did not reach statistical significance in the multivariate analysis, they were excluded from the final combined model to reduce the risk of overfitting and improve model stability. These two variables were discussed as univariate correlated factors and clinical background factors in the discussion section.

The ROC curve and AUC were used to assess the discriminative ability of the models, and the 95% CI of AUC were estimated via the bootstrap method. The DeLong test was applied to compare differences in AUC values. The optimal cut-off value was determined according to the Youden index. Subsequently, sensitivity, specificity, positive predictive value, negative predictive value and accuracy were calculated, and their 95% CI were estimated using the binomial distribution method or Wilson method.

Calibration performance was evaluated by calibration curves, Brier score and Hosmer-Lemeshow test. Decision curve analysis (DCA) was performed to assess the clinical net benefit across different threshold probabilities. A nomogram was plotted for model visualization.

Statistical analysis

Statistical analyses were performed using R (version 4.2.2) and SPSS (version 26.0). Normally distributed continuous data were expressed as mean ± standard deviation (SD), and independent samples t-tests were used for between-group comparisons. Non-normally distributed continuous data were described as median (interquartile range [IQR]), and the Mann-Whitney U test was used for between-group comparisons. Categorical data were expressed as counts (percentages), and the χ^2 test or Fisher's exact test was adopted for group comparisons.

Univariate and multivariate logistic regression analyses were used to explore factors associated with CLNM and establish predictive models. Results were presented as regression coefficients (β), odds ratios (ORs) and 95% CIs. Variance thresholding, correlation analysis and LASSO logistic regression were used for radiomic feature selection, and 10-fold cross-validation was applied to determine the optimal λ for the LASSO model.

ROC curves, AUCs and the DeLong test were used to evaluate and compare model discrimination. The optimal cutoff value was determined based on the Youden index, and sensitivity, specificity, positive predictive value, negative predictive value and accuracy were calculated. Calibration curves, Brier scores, the Hosmer-Lemeshow test and DCA were used to evaluate model calibration and clinical net benefit. A two-sided P<0.05 was considered statistically significant.

Results

Comparison of baseline characteristics of the training cohort

The training cohort consisted of 350 cases, including 151 with CLNM (43.14%) and 199 without CLNM (nCLNM, 56.86%). The CLNM group had a higher proportion of males than the nCLNM group [29.14% (44/151) vs. 18.59% (37/199), P = 0.029] and a lower mean age [(44.15±12.03) years vs. (49.25±11.80) years, P<0.001]. The proportion of patients aged ≥55 years was also lower in the CLNM group [21.85% (33/151) vs. 36.18% (72/199), P = 0.005].

Continuous age was only used to compare age distribution between the two groups. Age dichotomized as ≥55 years was adopted in all subsequent regression analyses and model construction. No statistically significant differences were observed between the two groups in BRAF V600E mutation, TPOAb titer, TGAb titer, neutrophil count, lymphocyte count, or NLR (all P>0.05) (**Table 1**).

Comparison of conventional ultrasound features in the training cohort

In the training set, the rate of maximum tumor diameter >10 mm was markedly higher in the CLNM group than in the nCLNM group [57.62% (87/151) vs. 31.66% (63/199), P<0.001]. The rates of calcification [74.83% (113/151) vs. 54.77% (109/199), P<0.001] and multifocality [51.66% (78/151) vs. 35.68% (71/199), P = 0.004] were also elevated in the CLNM group.

No statistically significant differences were observed between the two groups in aspect ratio >1, ill-defined margins, extrathyroidal extension, or solid hypoechogenicity (all P>0.05) (**Table 2**).

LASSO-based screening of radiomic features and rad-score construction

Variance filtering, correlation-based redundancy removal and LASSO logistic regression were sequentially performed for radiomic feature

Table 1. Comparison of clinical and laboratory indicators in the training set

Indicators	nCLNM (n = 199)	CLNM (n = 151)	t/χ^2	P
Gender (Male)	37 (18.59)	44 (29.14)	4.792	0.029
Age ≥55 years	72 (36.18)	33 (21.85)	7.723	0.005
BRAF V600E mutation	183 (91.96)	134 (88.74)	0.698	0.403
Age (years)	49.25±11.80	44.15±12.03	3.956	<0.001
TPOAb titer	55.78±194.43	63.14±189.60	-0.356	0.722
TGAb titer	40.54±224.62	63.17±222.36	-0.939	0.349
Neutrophils (10^9/L)	3.59±1.31	3.79±1.59	-1.264	0.207
Lymphocytes (10^9/L)	1.72±0.62	1.74±0.50	-0.306	0.760
NLR	2.24±0.94	2.33±1.12	-0.811	0.418

Note: CLNM: Central lymph node metastasis; nCLNM: non-central lymph node metastasis; NLR: neutrophil-to-lymphocyte ratio; TGAb: thyroglobulin antibody; TPOAb: thyroid peroxidase antibody.

Table 2. Comparison of common ultrasound features in training set

Variable	nCLNM (n = 199)	CLNM (n = 151)	χ^2	P
Maximum diameter >10 mm	63 (31.66)	87 (57.62)	22.573	<0.001
Aspect ratio >1	75 (37.69)	43 (28.48)	2.861	0.091
Unclear margin	191 (95.98)	147 (97.35)	Fisher	0.565
Calcification	109 (54.77)	113 (74.83)	14.042	<0.001
Capsule invasion	140 (70.35)	113 (74.83)	0.652	0.419
Multifocality	71 (35.68)	78 (51.66)	8.323	0.004
Hypoechoic solid nodule	195 (97.99)	150 (99.34)	Fisher	0.395

Note: CLNM: central lymph node metastasis; nCLNM: non-central lymph node metastasis.

Table 3. Imaging biomarker features and coefficients used to construct the Rad-score

Imaging biomarker features	Coefficient (β)
original_firstorder_F006	0.020909207438205634
original_firstorder_F028	0.1196238142004828
original_glcm_F079	0.036837897004817446
wavelet_HLL_glszm_F100	0.0024639655998129665
wavelet_HLL_glszm_F103	0.0711415369677051
wavelet_HLL_glszm_F111	0.01580983630817077
logsigma_2mm_glrlm_F132	0.0378641375747964
logsigma_2mm_glrlm_F142	0.03383356145821053
logsigma_2mm_glrlm_F144	0.055928484493549604
logsigma_2mm_glrlm_F152	0.01063126845341775

selection in the training set. The penalty parameter for LASSO logistic regression was determined using 10-fold cross-validation. The final model yielded C = 0.03255, corresponding to λ = 30.722 (log λ = 3.425). Under this penalty strength, 10 radiomic features with non-zero coefficients were ultimately selected to develop the Rad-score (**Table 3**). The Rad-score was calculated as the weighted sum of the standardized values of the selected features and their corresponding regression coefficients. The LASSO coefficient profile plot and 10-fold cross-validation curve are presented in **Figures 1** and **2**.

Univariate logistic regression analysis

Univariate logistic regression analysis of the candidate variables showed that age ≥55 years was associated with a reduced risk of CLNM (OR = 0.493, 95% CI: 0.305-0.799, P = 0.004). Male gender (OR = 1.800, 95% CI: 1.091-2.971, P = 0.021), maximum diameter >10 mm (OR = 2.935, 95% CI: 1.890-4.555, P<0.001), calcification (OR = 2.455, 95% CI: 1.548-3.895, P< 0.001), multifocality (OR = 1.926, 95% CI: 1.251-2.965, P = 0.003), and Rad-score (OR = 59.284, 95% CI: 18.260-192.474, P<0.001) were associated with an increased risk of CLNM (**Table 4**).

Multivariate logistic regression analysis and combined model construction

Age ≥55 years, maximum tumor diameter >10 mm and Rad-score were included in the multivariate logistic regression to establish the final combined model (Model D). The results demonstrated that age ≥55 years was independently associated with a decreased

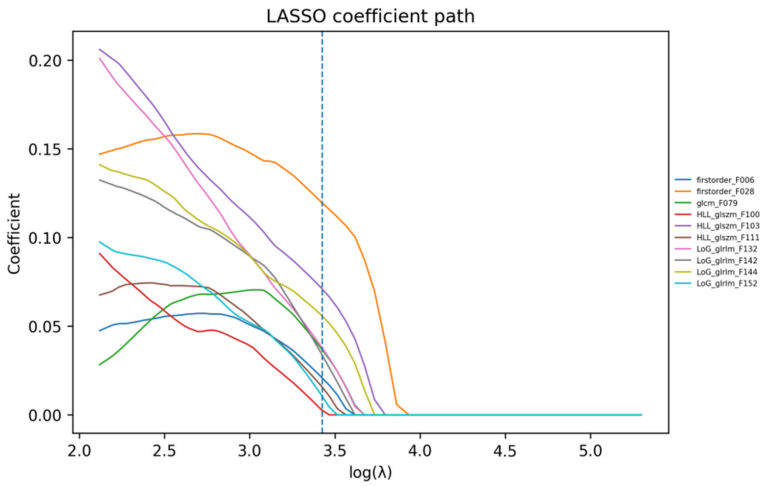

Figure 1. Path plot of LASSO coefficients.

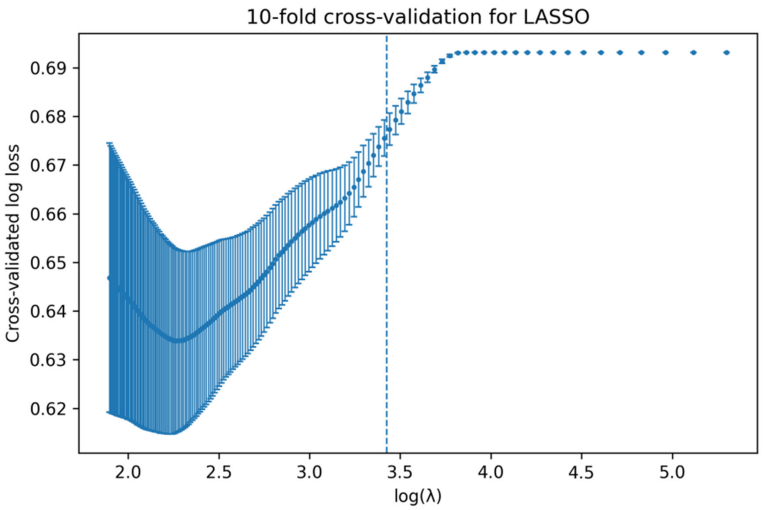

Figure 2. LASSO 10-fold cross-validation curve.

risk of CLNM (OR = 0.531, 95% CI: 0.311-0.907, P = 0.020). Tumor maximum diameter >10 mm (OR = 1.949, 95% CI: 1.192-3.187, P = 0.008) and Rad-score (OR = 35.074, 95% CI: 10.385-118.458, P<0.001) were correlated with an increased risk of CLNM, serving as independent predictors for CLNM. Although calcification and multifocality showed significant correlations with CLNM in the univariate analysis, they failed to reach statistical significance in the multivariate analysis and were therefore excluded from the final combined model (**Table 5**).

The logit equation for Model D was: logit(P) = -0.380 - 0.633 × (age ≥55 years) + 0.667 × (maximum diameter >10 mm) + 3.557 × (Rad-score), where age ≥55 years, and maximum diameter >10 mm were both assigned as binary variables, while Rad-score was a continuous variable.

ROC curve analysis

The AUC values for the four models in the training and validation cohorts are shown in **Table 6**. In the training set, the AUCs for Models A, B, C, and D were 0.632, 0.687, 0.757, and 0.764, respectively. In the validation set, the AUCs were 0.643, 0.702, 0.788, and 0.798, respectively. Model D achieved a higher AUC in both the training and validation sets, indicating that combining the Rad-score with age stratification and maximum diameter can effectively predict the risk of CLNM in PTC patients. Compared with Models A and B, Models C and D showed better discriminative ability, suggesting that radiomics features contribute substantially to CLNM risk identification (**Figure 3**).

Diagnostic performance at the optimal cutoff point

The optimal cut-off values of each model were determined based on the Youden index in the training set. Subsequently, sensitivity, specificity, positive predictive value, negative predictive value and accuracy of all models were calculated in both the training and validation sets, with 95% CI provided for each indicator.

Model D demonstrated good overall diagnostic performance in both sets, indicating not only good discriminative ability but also reasonable clinical discriminative power at the optimal cut-off point (**Table 7**).

Calibration curve analysis

Calibration curves were used to assess the agreement between predicted probabilities

Table 4. Results of univariate logistic regression analysis

Variable	β	OR (95% CI)	P
Age ≥55 years	-0.7067	0.493 (0.305-0.799)	0.004
Male	0.5880	1.800 (1.091-2.971)	0.021
ln(TPOAb+1)	0.0272	1.028 (0.916-1.153)	0.644
ln(TGAb+1)	0.1293	1.138 (0.993-1.304)	0.062
NLR	0.0876	1.092 (0.887-1.343)	0.407
BRAF V600E mutation	-0.3723	0.689 (0.336-1.413)	0.310
Maximum diameter >10 mm	1.0765	2.935 (1.890-4.555)	<0.001
Aspect ratio >1	-0.4181	0.658 (0.418-1.038)	0.072
Unclear margin	0.4313	1.539 (0.455-5.210)	0.488
Calcification	0.8983	2.455 (1.548-3.895)	<0.001
Capsule invasion	0.2257	1.253 (0.778-2.019)	0.354
Multifocality	0.6556	1.926 (1.251-2.965)	0.003
Hypoechoic solid nodule	1.1239	3.077 (0.340-27.815)	0.317
Rad-score	4.0823	59.284 (18.260-192.474)	<0.001

Note: NLR: neutrophil-to-lymphocyte ratio; TGAb: thyroglobulin antibody; TPOAb: thyroid peroxidase antibody.

Table 5. Results of multivariate logistic regression analysis (Model D)

Variable	β	OR (95% CI)	P
Age ≥55 years	-0.633	0.531 (0.311-0.907)	0.02
Maximum diameter >10 mm	0.667	1.949 (1.192-3.187)	0.008
Rad-score	3.557	35.074 (10.385-118.458)	<0.001

Note: The logit equation for Model D was: logit(P) = -0.380 - 0.633 × (age ≥55 years) + 0.667 × (maximum diameter >10 mm) + 3.557 × (Rad-score), where age ≥55 years, and maximum diameter >10 mm were both assigned as binary variables, while Rad-score was a continuous variable.

Table 6. AUC values of the four models in the training set and validation set

Model	Training set AUC (95% CI)	Validation set AUC (95% CI)
Model A	0.632 (0.574-0.690)	0.643 (0.551-0.735)
Model B	0.687 (0.633-0.741)	0.702 (0.619-0.785)
Model C	0.757 (0.707-0.805)	0.788 (0.714-0.860)
Model D	0.764 (0.712-0.813)	0.798 (0.724-0.867)

and actual observed probabilities. In both the training and validation cohorts, the calibration curve for Model D was generally close to the ideal reference line, indicating good calibration (**Figure 4**).

DCA

Decision curves were plotted based on complete case comparisons (training cohort: n = 350; validation cohort: n = 150) (**Figure 5**). The DCA results showed that in both the training and validation sets, when the threshold probability ranged from approximately 0.25 to 0.60, the combined model (Model D) provided a higher net benefit than the "treat all" strategy, the "treat none" strategy, and the other models. This suggests that Model D can offer adjunctive value for preoperative risk stratification of CLNM in PTC patients within this threshold range.

Nomogram construction

A nomogram was constructed based on the regression coefficients of Model D to facilitate individualized risk assessment. In clinical practice, the points corresponding to each predictor are summed to obtain a total score, which is then converted into the predicted probability of CLNM (**Figure 6**).

Discussion

CLNM is an important factor influencing recurrence and long-term clinical management of PTC. Although prophylactic central lymph node dissection (pCLND) can reduce the risk of local recurrence to some extent, it may lead to complications such as hypoparathyroidism and recurrent laryngeal nerve injury; thus, a one-size-fits-all surgical strategy cannot be applied to all patients in clinical practice [14-16]. Therefore, developing a preoperative CLNM risk assessment model is clinically valuable for identifying high-risk patients and facilitating the formulation of individualized surgical protocols and central lymph node management plans. Previous studies have demonstrated that low-risk PTC patients can achieve favorable long-term survival even without pCLND.

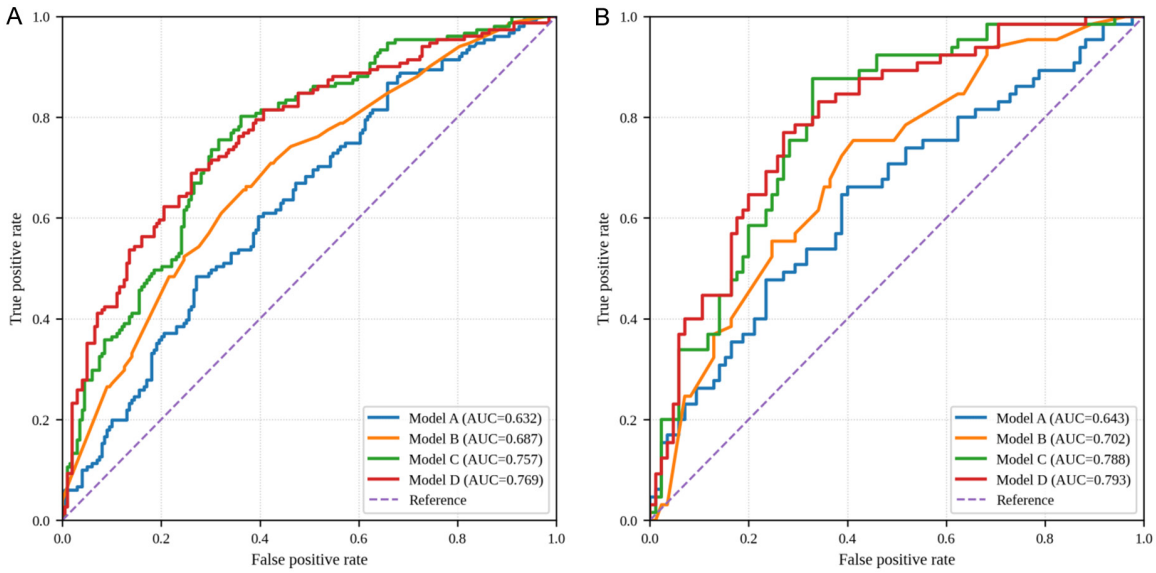

Figure 3. ROC curves for Model A-D. A: Training set, B: Validation set.

Table 7. Diagnostic performance of each model at the optimal cutoff point

Model	Dataset	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Model A	Training	0.450	0.483 (0.405-0.563)	0.729 (0.663-0.786)	0.575 (0.488-0.657)	0.650 (0.586-0.710)	0.623 (0.571-0.672)
Model A	Validation	0.450	0.538 (0.419-0.654)	0.682 (0.577-0.772)	0.565 (0.441-0.681)	0.659 (0.555-0.750)	0.620 (0.540-0.694)
Model B	Training	0.420	0.662 (0.584-0.733)	0.628 (0.559-0.692)	0.575 (0.500-0.646)	0.710 (0.639-0.772)	0.643 (0.591-0.691)
Model B	Validation	0.420	0.662 (0.540-0.765)	0.635 (0.529-0.730)	0.581 (0.467-0.687)	0.711 (0.600-0.800)	0.647 (0.567-0.719)
Model C	Training	0.362	0.801 (0.731-0.857)	0.638 (0.569-0.702)	0.627 (0.557-0.692)	0.809 (0.740-0.863)	0.709 (0.659-0.754)
Model C	Validation	0.362	0.862 (0.757-0.925)	0.671 (0.565-0.761)	0.667 (0.561-0.758)	0.864 (0.761-0.927)	0.753 (0.679-0.815)
Model D	Training	0.415	0.695 (0.618-0.763)	0.724 (0.658-0.781)	0.656 (0.580-0.725)	0.758 (0.692-0.813)	0.711 (0.662-0.756)
Model D	Validation	0.415	0.769 (0.654-0.855)	0.694 (0.590-0.782)	0.658 (0.546-0.755)	0.797 (0.692-0.873)	0.727 (0.650-0.792)

Note: The cutoff value was determined based on the Youden index in the training set and applied to both the training and validation sets. Model A: clinical/laboratory model; Model B: conventional ultrasound model; Model C: Rad-score model; Model D: combined model. PPV: positive predictive value; NPV: negative predictive value.

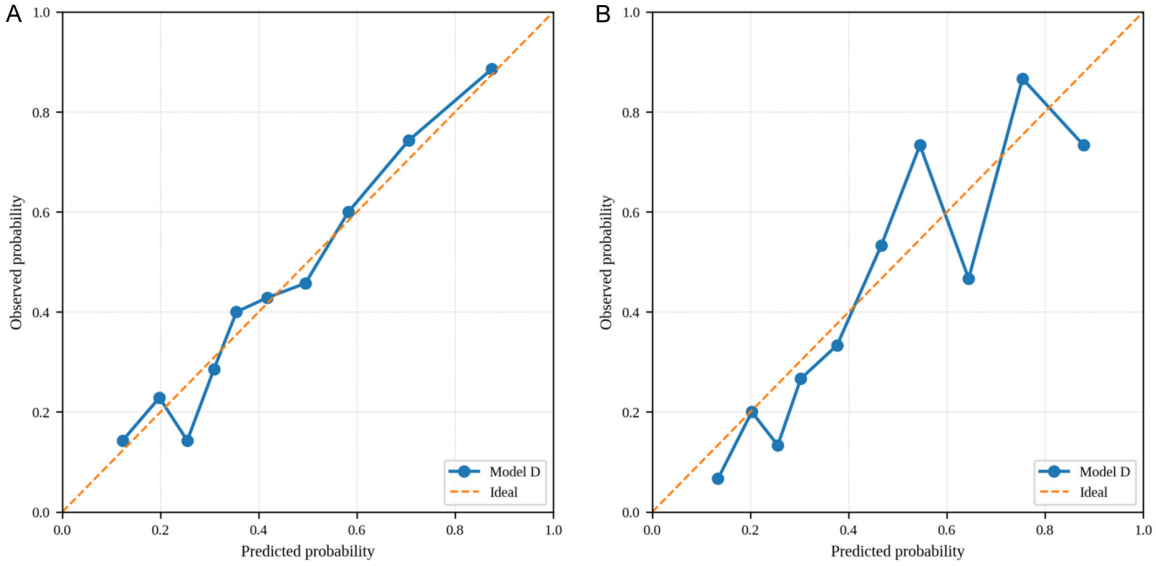

Figure 4. Calibration curves for Model D. A: Training set, B: Validation set.

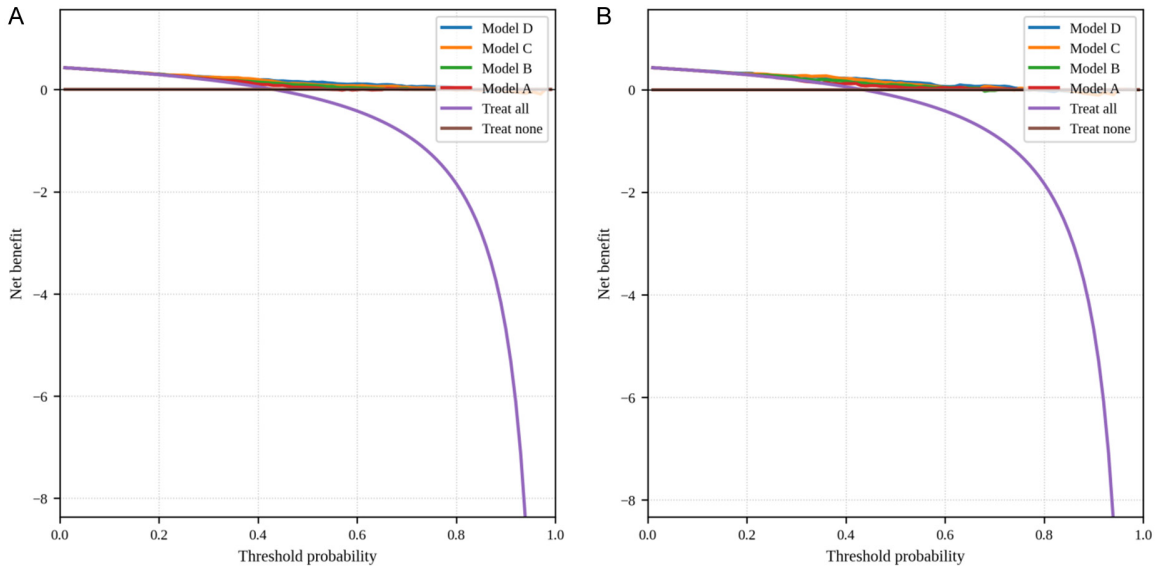

Figure 5. Decision curve analysis for Model A-D. A: Training set, B: Validation set.

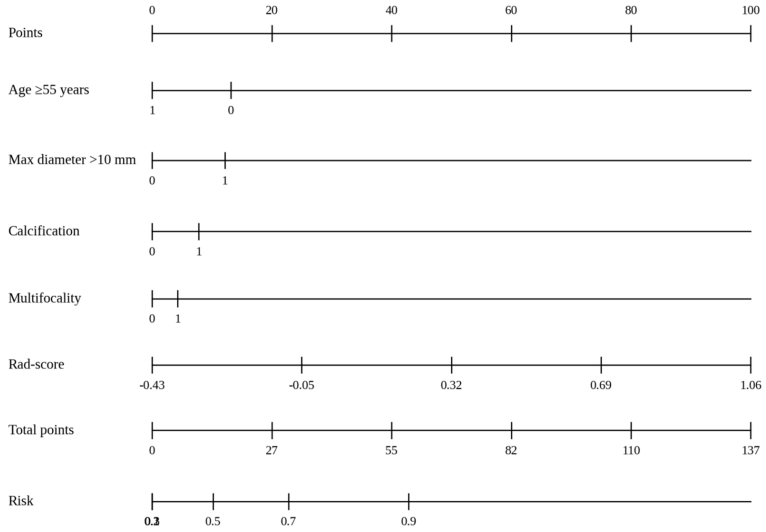

Figure 6. Nomogram for Model D.

This further highlights the need for preoperative CLNM risk stratification and the adoption of selective individualized dissection strategies [17-19].

This single-center retrospective cohort study randomly divided the enrolled cases into a training set and a validation set at a ratio of 7:3. We developed and validated a preoperative prediction model for CLNM in PTC patients based on clinical information, conventional ultrasonographic features and ultrasound radiomic features. Descriptive analyses and univariate regression in the training set revealed that male gender, maximum tumor diameter >10 mm, calcification, multifocality and Rad-score were associated with CLNM. However, only age ≥55 years, maximum tumor diameter >10 mm and Rad-score remained statistically significant in the multivariate regression analysis. Age ≥55 years was associated with a reduced risk of CLNM, while maximum tumor diameter >10 mm and Rad-score were linked to an increased CLNM risk.

Although calcification and multifocality showed statistical significance in univariate analysis and previous studies have demonstrated their associations with CLNM in PTC, they failed to exert independent predictive effects after adjustment for tumor diameter and Rad-score [20]. Following the principles of parsimony and stability for prediction model construction, the revised final combined model excluded these two variables. They were instead discussed as univariate correlated factors and clinical background factors in this section. These findings indicated that part of the risk information reflected by calcification and multifocality might have been cap-

tured by tumor diameter and radiomic features. Incorporating them into the final equation would increase model complexity and the risk of overfitting, without bringing stable incremental predictive value.

A positive correlation between tumor diameter and metastatic risk has been well recognized. Systematic reviews have confirmed significant differences in the incidence of occult CLNM across different tumor size strata, suggesting that cutoffs such as >10 mm are practically valuable for preoperative risk stratification [21, 22]. Our study further verified that maximum tumor diameter >10 mm was significantly associated with CLNM and served as an independent predictor in the multivariate model, providing a quantifiable indicator for preoperative clinical decision-making. Regarding age, age ≥55 years was identified as a protective factor in the present study, which was consistent with the general conclusion that younger patients are more susceptible to CLNM [23-26]. In our regression analyses, age was converted into a binary variable (age ≥55 years), while continuous age was only used for baseline description and intergroup comparison, so as to avoid repeated inclusion of the same variable in different forms.

Calcification, especially punctate hyperechoic foci or microcalcification, is a typical ultrasonographic sign of PTC. Its pathological basis may involve psammoma body formation, tumor stromal reaction and alterations in the local microenvironment. Prior studies have reported that microcalcification is correlated with an increased risk of cervical lymph node metastasis [27]. In our cohort, the rate of calcification was higher in the CLNM group, and it was associated with elevated CLNM risk in univariate logistic regression, indicating its clinical value for risk alert. However, it lost independent predictive significance after adjusting for age ≥55 years, maximum tumor diameter >10 mm and Rad-score. Several reasons may account for this result. First, calcification rarely occurs in isolation; it is frequently accompanied by increased tumor volume, local fibrosis, stromal reaction and enhanced tumor invasiveness. Thus, the risk information of calcification partially overlaps with conventional ultrasonographic variables such as tumor diameter. Second, the Rad-score is composed of multiple first-order grayscale features, texture features and filter-based transformed features. It can reflect ultrasound changes caused by calcification, posterior acoustic shadow and heterogeneous echo in terms of grayscale distribution, texture complexity and spatial heterogeneity, thereby covering most imaging information related to calcification. Third, calcification was analyzed as a binary variable in this study without further classification into microcalcification, coarse calcification, marginal calcification and mixed calcification, which may underestimate its independent predictive effect. Therefore, calcification is more suitable as a suggestive factor for comprehensive preoperative risk assessment, rather than an independent variable in the final prediction equation after multivariate adjustment.

Multifocality generally suggests multicentric tumorigenesis or intraglandular dissemination of PTC. Previous studies have proven its correlations with lymph node metastasis and local recurrence [28]. In our study, multifocality was associated with higher CLNM risk in descriptive comparison and univariate analysis, indicating it can serve as a clinical indicator reflecting tumor burden and potential invasiveness. However, it did not retain independent predictive power in the multivariate model, implying its predictive effect was confounded by tumor diameter and Rad-score. On the one hand, the risk related to multifocality depends not merely on the presence of multiple lesions, but also on lesion number, bilateral involvement, primary lesion size, overall tumor burden and spatial distribution. Since multifocality was only dichotomized in our analysis, the above detailed information was not quantified. On the other hand, the ROI was delineated on the largest tumor lesion for radiomic analysis. The Rad-score mainly reflects the internal texture complexity and heterogeneity of the primary lesion, while tumor diameter >10 mm represents the volume burden of the primary lesion. These two types of information could explain part of the risk associated with multifocality. Collectively, although multifocality has clinical suggestive implications, it failed to provide stable independent incremental predictive value in our cohort. For this reason, it was not included in the final model. We recommend interpreting it as a background factor in comprehensive preopera-

tive evaluation, together with age, tumor diameter and Rad-score.

Different from traditional models relying solely on clinical and conventional ultrasonographic features, the present study introduced an ultrasound-based radiomic Rad-score. Radiomics can quantitatively characterize intratumoral heterogeneity via first-order grayscale features, texture features and filter-transformed features, compensating for the limitations of naked-eye observation and subjective interpretation of conventional ultrasound [9, 29]. Pathologically, CLNM in PTC may be associated with pathological changes such as active tumor cell proliferation, aggravated stromal reaction, local fibrosis, microcalcification formation and lymphatic invasion. These microscopic pathological changes can lead to uneven grayscale distribution, increased texture complexity and abnormal spatial arrangement on ultrasound images [30, 31]. By integrating multiple radiomic features, the Rad-score can indirectly reflect the above pathological heterogeneity and aggressive biological behaviors, and thus improve the detection ability of occult preoperative CLNM.

In terms of model performance, the radiomic model (Model C) showed good discriminative ability in both training and validation sets. The revised final combined model (Model D) excluded calcification and multifocality with $P>0.05$ in multivariate analysis, and consisted of age ≥55 years, maximum tumor diameter >10 mm and Rad-score. Model D still maintained favorable discriminative power in the two datasets, suggesting that removing non-independent variables did not markedly impair predictive efficacy. Meanwhile, the model became more concise, stable and clinically interpretable. Overall, the Rad-score was the major contributor to model performance, while age stratification and tumor diameter improved its clinical applicability and interpretability. Calibration curves revealed good consistency between predicted probabilities and actual CLNM risks of Model D. DCA demonstrated a high net benefit within a certain range of threshold probabilities, supporting its application as an auxiliary tool for preoperative risk communication and individualized surgical decision-making.

Clinically, Model D can provide quantitative risk reference for the management of central lymph nodes based on preoperative available indicators including age, tumor diameter and radiomic Rad-score. For patients with low predicted risk, this model can help surgeons avoid excessive lymph node dissection. For those with high risk, clinicians are recommended to perform enhanced ultrasonographic evaluation of central lymph nodes. If necessary, comprehensive decisions can be made combining fine-needle aspiration, thyroglobulin measurement in washout fluid and intraoperative frozen section pathology. DCA results showed that Model D yielded higher net benefit than the strategies of universal intervention or no intervention within a specific threshold range, indicating its potential value for clinicians to decide whether to strengthen central lymph node management. However, this study did not directly compare Model D with existing risk stratification tools recommended by clinical guidelines or other validated prediction models. Its additional clinical value needs to be further verified in external cohorts and prospective studies.

This study has several limitations. First, this was a single-center retrospective study. Selection bias may exist in case enrollment, and the generalizability of the model requires further validation via large-sample, multicenter prospective studies. Second, the training and validation sets were randomly divided from the same cohort, which only represented internal validation. Independent external validation cohorts are still lacking. Third, although we assessed the repeatability of ROI delineation and radiomic features, ultrasound images are susceptible to equipment parameters, operator experience, probe pressure and image quality. The stability of radiomic features may decrease across different devices and medical centers. Fourth, calcification and multifocality were correlated with CLNM in univariate analysis but not in multivariate analysis, so they were excluded from the final model. Since we did not conduct detailed stratification of calcification subtypes, lesion number, bilateral involvement and overall tumor burden, further studies with larger multicenter datasets are required to evaluate their true incremental value based on radiomic models. Fifth, we did not systematically compare Model D with guideline-recommended risk stratification approaches or published prediction models, nor did we analyze its impact on clinical decision-making in real clini-

cal scenarios. Therefore, its practical clinical value remains to be confirmed.

In conclusion, this concise prediction model combining ultrasound radiomic Rad-score, age stratification and maximum tumor diameter achieves satisfactory performance in predicting preoperative CLNM risk in PTC patients. It can serve as a reference for preoperative risk stratification, individualized treatment and selective central lymph node management.

Acknowledgements

This work was supported by the Medical Science Research Project of Hebei (20250110).

Disclosure of conflict of interest

None.

Address correspondence to: Zhaolin Yin, Department of Ultrasound, The Fourth Hospital of Hebei Medical University, No. 12 Jiankang Road, Shijiazhuang 050011, Hebei, China. Tel: +86-18732-198677; E-mail: doctoryzl@126.com

References

[1] Feng JW, Liu SQ, Qi GF, Ye J, Hong LZ, Wu WX and Jiang Y. Development and validation of clinical-radiomics nomogram for preoperative prediction of central lymph node metastasis in papillary thyroid carcinoma. Acad Radiol 2024; 31: 2292-2305.

[2] Smulever A and Pitoia F. Thirty years of active surveillance for low-risk thyroid cancer, lessons learned and future directions. Rev Endocr Metab Disord 2024; 25: 65-78.

[3] Yao S and Zhang H. Papillary thyroid carcinoma with Hashimoto's thyroiditis: impact and correlation. Front Endocrinol (Lausanne) 2025; 16: 1512417.

[4] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-263.

[5] Heo DB, Won HR, Tae K, Kang YE, Jeon E, Ji YB, Chang JW, Choi JY, Yu HW, Ku EJ, Lee EK, Kim M, Choe JH and Koo BS; MASTER study group. Clinical impact of coexistent chronic lymphocytic thyroiditis on central lymph node metastasis in low- to intermediate-risk papillary thyroid carcinoma: the MASTER study. Surgery 2024; 175: 1049-1054.

[6] Kuenstner W, Alzedaneen Y and Burman KD. Update in papillary thyroid cancer. Endocrinol Metab Clin North Am 2025; 54: 329-340.

[7] Peng L, Zheng X, Xue Y, Huang C, Su X and Yu S. Central lymph nodes in frozen sections can effectively guide extended lymph node resection for papillary thyroid carcinoma. Ann Med 2023; 55: 2286337.

[8] Xu ZL, Hou JX, Zheng ZH, Deng YQ, Zeng GM, Wang SR, Zhu PS, Kang YF, Du TT, Dong J, Liu W, Li J and Cui XW. Enhancing prediction of cervical lymph node metastasis in papillary thyroid carcinoma through nodule-oriented quantification: combined S-detect and ultrasound elastography. Quant Imaging Med Surg 2025; 15: 3416-3429.

[9] Guo Y, Li T, Gong B, Hu Y, Wang S, Yang L and Zheng C. From images to genes: radiogenomics based on artificial intelligence to achieve non-invasive precision medicine in cancer patients. Adv Sci (Weinh) 2025; 12: e2408069.

[10] Albuck AL, Issa PP, Hussein M, Aboueisha M, Attia AS, Omar M, Munshi R, Shama M, Toraih E, Randolph GW and Kandil E. A combination of computed tomography scan and ultrasound provides optimal detection of cervical lymph node metastasis in papillary thyroid carcinomas: a systematic review and meta-analysis. Head Neck 2023; 45: 2173-2184.

[11] Fang M, Lei M, Chen X, Cao H, Duan X, Yuan H and Guo L. Radiomics-based ultrasound models for thyroid nodule differentiation in Hashimoto's thyroiditis. Front Endocrinol (Lausanne) 2023; 14: 1267886.

[12] Wen Q, Wang Z, Traverso A, Liu Y, Xu R, Feng Y and Qian L. A radiomics nomogram for the ultrasound-based evaluation of central cervical lymph node metastasis in papillary thyroid carcinoma. Front Endocrinol (Lausanne) 2022; 13: 1064434.

[13] Cantisani V, Bojunga J, Durante C, Dolcetti V and Pacini P. Multiparametric ultrasound evaluation of thyroid nodules. Ultraschall Med 2025; 46: 14-35.

[14] Ren T, Min S, Grodski S, Serpell J and Lee JC. Prophylactic central lymph node dissection for low-risk papillary thyroid cancer-Impact on subsequent therapy. World J Surg 2025; 49: 170-178.

[15] Qiu R and Fu J. Prophylactic central lymph node dissection in cN0 papillary thyroid cancer: a comparative study of via breast and transoral approach versus via breast approach alone. Front Endocrinol (Lausanne) 2024; 15: 1356739.

[16] Wu ZG, Zheng WT, Chen LJ, Zhu FS, Ma ZS, Cao FL, Cui BB, Xie BJ and Yan XQ. Is prophylactic central neck dissection necessary for patients with clinically node-negative papillary thyroid

microcarcinoma? A follow-up study of more than 10 years. Front Endocrinol (Lausanne) 2025; 16: 1597661.
[17] Song JL, Li LR, Xu ZL, Li JJ, Sun SR and Chen C. Long-term survival in patients with papillary thyroid cancer who did not undergo prophylactic central lymph node dissection: a SEER-based study. World J Oncol 2022; 13: 136-144.
[18] Chereau N, Christou N, Caiazzo R, Le Fouler A, Lifante JC, Maillard L, Mirallie E, Pattou F, Bouviez N, Santucci N, Brunaud L and Menegaux F. Impact of prophylactic central lymph node dissection on the complications and recurrence rates in papillary thyroid carcinoma-An AFCE (French-speaking Association of Endocrine Surgery) multicentre study based on the EUROCRINE® national data. Cancer Rep (Hoboken) 2024; 7: e1993.
[19] Papini P, Rossi L, Matrone A, De Renzis A, Morganti R, Valerio L, Ambrosini CE, Materazzi G and Elisei R. Prophylactic central neck dissection in clinically node-negative papillary thyroid carcinoma: 10-year impact on surgical and oncologic outcomes. Surgery 2025; 181: 109258.
[20] Guang Y, He W, Zhang W, Zhang H, Zhang Y and Wan F. Clinical study of ultrasonographic risk factors for central lymph node metastasis of papillary thyroid carcinoma. Front Endocrinol (Lausanne) 2021; 12: 791970.
[21] Sun W, Lan X, Zhang H, Dong W, Wang Z, He L, Zhang T and Liu S. Risk factors for central lymph node metastasis in CN0 papillary thyroid carcinoma: a systematic review and meta-analysis. PLoS One 2015; 10: e0139021.
[22] Qiu P, Guo Q, Pan K and Lin J. Development of a nomogram for prediction of central lymph node metastasis of papillary thyroid microcarcinoma. BMC Cancer 2024; 24: 235.
[23] Ozel TM, Soytas Y, Akbulut S, Celik A, Yildiz G, Karatay H and Sari S. The necessity of prophylactic central lymph node dissection in clinically no papillary thyroid carcinoma: perspective from the endemic region. Langenbecks Arch Surg 2025; 410: 109.
[24] Zhao F, Wang P, Yu C, Song X, Wang H, Fang J, Zhu C and Li Y. A LASSO-based model to predict central lymph node metastasis in preoperative patients with cN0 papillary thyroid cancer. Front Oncol 2023; 13: 1034047.
[25] Sun H and Liu M. Risk factors associated with lymph node metastasis in patients with small papillary thyroid carcinomas. Int J Clin Exp Pathol 2023; 16: 86-93.
[26] Tang L, Qu RW, Park J, Simental AA and Inman JC. Prevalence of occult central lymph node metastasis by tumor size in papillary thyroid carcinoma: a systematic review and meta-analysis. Curr Oncol 2023; 30: 7335-7350.
[27] Liu J, Jia X, Gu Y, Chen X, Guan L, Yan J, Zhai H, Zhou N, Dong Y, Zhan W, Luo X and Zhou J. Thyroid parenchyma microcalcifications on ultrasound for predicting lymph node metastasis in papillary thyroid carcinoma: a prospective multicenter study in China. Front Oncol 2021; 11: 609075.
[28] Dai Q, Liu D, Tao Y, Ding C, Li S, Zhao C, Wang Z, Tao Y, Tian J and Leng X. Nomograms based on preoperative multimodal ultrasound of papillary thyroid carcinoma for predicting central lymph node metastasis. Eur Radiol 2022; 32: 4596-4608.
[29] Jiang L, Zhang Z, Guo S, Zhao Y and Zhou P. Clinical-radiomics nomogram based on contrast-enhanced ultrasound for preoperative prediction of cervical lymph node metastasis in papillary thyroid carcinoma. Cancers (Basel) 2023; 15: 1613.
[30] Zhang XY, Zhang D, Zhou W, Wang ZY, Zhang CX, Li J, Wang L and Cui XW. Predicting lymph node metastasis in papillary thyroid carcinoma: radiomics using two types of ultrasound elastography. Cancer Imaging 2025; 25: 13.
[31] Wang L, Chen J, Yuan X, Wang J, Sun L, Jiang J, Zhang L, Liu M and Zhou Q. Lymph node metastasis of papillary thyroid carcinoma in the context of Hashimoto's thyroiditis. BMC Endocr Disord 2022; 22: 12.