



Preoperative prediction of occult high-volume central lymph node metastasis in clinically node-negative papillary thyroid carcinoma using a multimodal practical model: development, temporal validation, and external validation

 Ziwei Zhang¹
 Cailing Lin²
 Jing Ning¹
 Xiaochen Liu¹
 Chenshan Dong¹
 Hang Ling³

¹Department of Ultrasonography, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fujian, China

²Department of Breast Surgery, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fujian, China

³Department of Pathology, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fujian, China

Handling editor: Gökçen Çoban Çiftçi

Corresponding author: Hang Ling

E-mail: ziwei_zhang2009@163.com

Received 07 April 2026; revision requested 07 May 2026; last revision requested 28 May 2026; accepted 17 June 2026.



Epub: 08.07.2026

Publication date:

DOI: 10.4274/dir.2026.264048

PURPOSE

Preoperative risk estimation of occult high-volume central lymph node metastasis (CLNM) in clinically node-negative (cN0) papillary thyroid carcinoma (PTC) remains challenging. We aimed to develop and validate a practical multimodal model integrating conventional ultrasound and radiomics for individualized preoperative risk stratification.

METHODS

In this retrospective, two-center, multicohort study, 814 patients with cN0 PTC were included and assigned to a development cohort (n = 470), a temporal validation cohort (n = 202), and an external validation cohort (n = 142). Preoperative clinical variables and grayscale ultrasound radiomics features were evaluated. Least absolute shrinkage and selection operator regression was used for feature selection. Four candidate machine-learning algorithms were compared in the development cohort, and the selected classifier was used to construct clinical, radiomics, and fusion models. Model performance was assessed in terms of discrimination, calibration, and clinical utility. SHapley Additive exPlanations (SHAPs) analysis and a web-based calculator were used to improve interpretability and applicability.

RESULTS

Three clinical predictors and 20 radiomics features were retained. Random forest was selected for final model construction. In the temporal validation cohort, the clinical, radiomics, and fusion models achieved area under the curve (AUC) values of 0.703, 0.755, and 0.845, respectively; in the external validation cohort, the corresponding AUC values were 0.676, 0.730, and 0.817. For the fusion model, sensitivity, specificity, and accuracy were 0.837, 0.695, and 0.755 in the temporal validation cohort and 0.769, 0.761, and 0.764 in the external validation cohort, respectively. SHAPs analysis identified radiomics score and maximum tumor diameter as the major contributors to model output.

CONCLUSION

A practical multimodal model integrating conventional ultrasound and radiomics showed favorable performance for preoperative risk estimation of occult high-volume CLNM in cN0 PTC and may support individualized risk stratification, pending further prospective validation.

CLINICAL SIGNIFICANCE

This multimodal model may refine preoperative risk estimation for patients with cN0 PTC who harbor occult high-volume CLNM, thereby providing an adjunctive tool for individualized perioperative risk assessment.

KEYWORDS

Papillary thyroid carcinoma, central lymph node metastasis, radiomics, ultrasound, machine learning, risk prediction

Thyroid cancer incidence has risen sharply worldwide over recent decades, and much of this increase has been driven by the expanded use of imaging, particularly thyroid and neck ultrasonography, which has facilitated the detection of small papillary thyroid carcinomas (PTCs) that might otherwise have remained clinically silent.¹ PTC now accounts for the vast majority of thyroid malignancies, and although disease-specific survival is generally excellent, cervical lymph node metastasis remains common and continues to shape operative planning, recurrence risk stratification, and follow-up intensity.^{1,2} In daily clinical practice, the challenge is therefore no longer limited to diagnosing PTC itself but extends to identifying which patients with apparently low-risk disease harbor biologically more consequential regional disease before surgery. Among patients with PTC, the central compartment is the most frequent site of nodal spread, and occult central lymph node metastasis (CLNM) is not uncommon even among patients classified as clinically node-negative (cN0) before surgery.^{2,3} This creates a familiar dilemma: Routine preoperative imaging is indispensable, yet its ability to detect central compartment disease is imperfect. In a large systematic review and meta-analysis published in *JAMA Otolaryngology–Head & Neck Surgery*, Alabousi et al.² showed that, for central compartment metastasis, computed tomography was more sensitive, whereas ultrasonography was more specific, underscoring the intrinsic diagnostic limitations of any single preoperative imaging modality in this setting. Tang

et al.³ reported in a 2023 systematic review that occult central nodal disease remains prevalent across tumor-size strata, with prevalence increasing as primary tumors enlarge. These findings support a central clinical reality: A substantial proportion of patients with cN0 PTC still carry a hidden nodal burden that is not adequately captured by routine preoperative assessment alone. However, not all occult nodal disease has the same clinical relevance. Increasing evidence suggests that nodal burden, rather than simple nodal positivity alone, is more closely aligned with recurrence-oriented risk assessment and surgical decision-making. In recent years, high-volume CLNM, commonly defined as involvement of more than five central lymph nodes, has emerged as a more meaningful end point than any-volume CLNM.^{4,5} Huang et al.⁴ specifically examined large-volume central CLNM in patients with clinical N0 PTC and identified several clinicopathologic correlates, and Zhu et al.⁵ developed a clinical predictive model for high-volume lymph node metastasis in PTC and demonstrated that this end point could be modeled preoperatively. At the same time, preoperative nomograms based on clinical and ultrasound variables have continued to improve risk estimation for central nodal metastasis.⁶ Nonetheless, important limitations remain. Many existing studies predict any CLNM rather than occult high-volume central disease, rely on single-center retrospective datasets with random internal splits rather than temporal validation, and do not incorporate quantitative imaging features that may capture latent intratumoral heterogeneity beyond conventional ultrasound assessment.^{4,6} Ultrasound radiomics offers one possible way to address this gap. By transforming grayscale ultrasound data into structured quantitative descriptors of shape, intensity, and texture, radiomics may extract tumor information that is not consistently appreciable through human interpretation alone. Several recent studies have shown that ultrasound-based radiomics can improve preoperative prediction of CLNM in PTC.^{7,8} Feng et al.,⁷ for example, developed a clinical–radiomics nomogram for preoperative prediction of CLNM in PTC and reported improved performance when radiomics features were combined with key clinical variables. Jia et al.⁸ similarly showed that an ultrasound-based radiomics model had value in identifying central nodal metastasis preoperatively. More broadly, explainable multimodal artificial intelligence (AI) models are beginning to demonstrate that integrating complementary preoperative information

can improve nodal metastasis prediction while preserving interpretability.⁹ Yet studies specifically targeting occult high-volume central CLNM in patients with cN0 PTC while simultaneously emphasizing practical preoperative inputs, temporal validation, external validation, and explainability remain limited. Against this background, we conduct a two-center multicohort study to develop and validate a practical multimodal model for the preoperative prediction of occult high-volume CLNM in cN0 PTC. We hypothesize that integrating routinely assessable conventional ultrasound features with an ultrasound–radiomics signature provides more informative preoperative risk estimation than either data stream alone. To address the limitations of prior work, we design the study around a development cohort, a temporal validation cohort, and an independent external validation cohort, further incorporating SHapley Additive exPlanations (SHAP)-based interpretability analysis together with a web-based calculator to facilitate clinical translation. In this way, the present study is intended not only to improve prediction accuracy but also to provide a clinically deployable framework for individualized nodal risk assessment before surgery.^{10,11}

Methods

Study design and participants

This retrospective, two-center, multicohort prediction study was conducted at Fujian Provincial Hospital and Fujian Union Hospital. Consecutive patients who underwent surgery for PTC during the study period were identified from institutional medical record, imaging, surgical, and pathological databases. After application of the predefined eligibility criteria, patients from Fujian Provincial Hospital constituted the internal cohort, which was chronologically divided into a development cohort and a temporal validation cohort. Patients from Fujian Union Hospital constituted the external validation cohort. Only variables available before surgery were considered candidate predictors, whereas postoperative pathological findings were used exclusively for end point ascertainment. The patient selection process is shown in Figure 1.

This study was approved by the Institutional Review Boards of Fujian Provincial Hospital and Fujian Union Hospital (approval number: K2025-04-002, date: 28.04.2025). Given the retrospective design and use of de-identified data, the requirement for informed consent was waived.

Main points

- A multimodal model combining conventional ultrasound features and ultrasound radiomics improved preoperative prediction of occult high-volume central lymph node metastasis in clinically node-negative papillary thyroid carcinoma.
- The fusion model showed higher discriminative performance than the clinical and radiomics models alone in both the temporal and external validation cohorts.
- SHapley Additive exPlanations analysis showed that radiomics score and maximum tumor diameter contributed most strongly to model output; this finding was interpreted cautiously because the radiomics signature included shape-, size-, and texture-related features.
- A web-based calculator was developed to facilitate individualized preoperative risk estimation as an adjunctive decision-support tool.

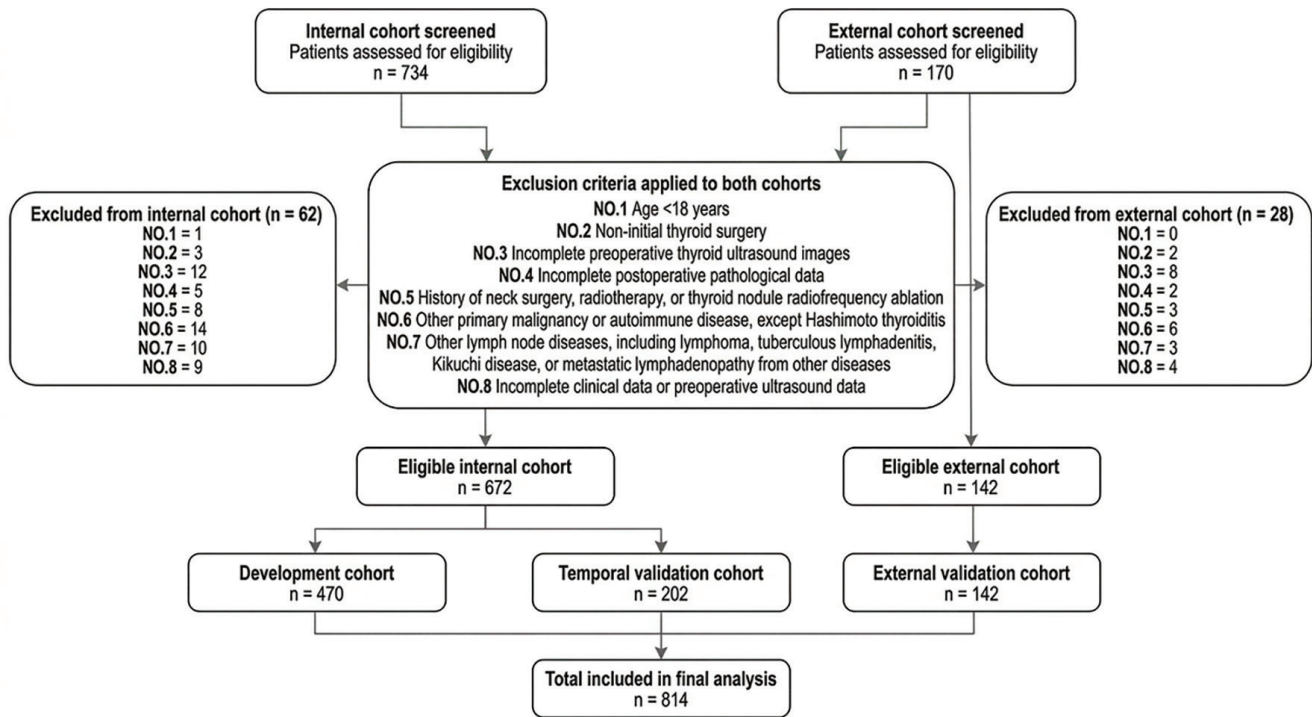


Figure 1. Flow diagram of patient selection.

Eligibility criteria

Patients were eligible for inclusion if they met all of the following criteria: (1) age $\geq$ 18 years; (2) postoperative histopathologic confirmation of PTC; (3) cN0 status before surgery, defined as no definite evidence of cervical lymph node metastasis on the original preoperative clinical assessment and neck ultrasonography reports; (4) confirmation during retrospective blinded image review that the available preoperative thyroid and neck ultrasound images did not show definite metastatic cervical lymph nodes in the assessable central and lateral neck compartments; (5) initial thyroid surgery; (6) available and evaluable preoperative conventional thyroid ultrasound images; (7) postoperative pathological data sufficient to determine central lymph node status and the number of metastatic central lymph nodes; and (8) complete clinical and preoperative ultrasound data for predictor extraction and model development.

Patients were excluded if they met any of the following criteria: (1) age < 18 years; (2) non-initial thyroid surgery; (3) incomplete preoperative thyroid ultrasound images; (4) incomplete postoperative pathological data; (5) a history of neck surgery, radiotherapy, or thyroid nodule radiofrequency ablation; (6) concomitant primary malignancies or autoimmune diseases other than Hashimoto thyroiditis; (7) concomitant lymph node diseases, including lymphoma, tuberculous

lymphadenitis, Kikuchi disease, or metastatic lymphadenopathy from other diseases; (8) documented suspicious cervical lymph nodes before surgery or indeterminate cervical lymph nodes for which metastatic disease could not be confidently excluded based on the available clinical reports, ultrasound reports, or retrospective image review; or (9) incomplete clinical or preoperative ultrasound data.

Outcome definition

The primary end point was occult high-volume CLNM, defined as postoperative pathological confirmation of more than five metastatic central lymph nodes in patients classified as cN0 before surgery. Patients with no CLNM or with five or fewer metastatic central lymph nodes were classified as negative for the primary end point. This cut-off was selected because nodal burden, rather than nodal positivity alone, is incorporated into recurrence-oriented risk stratification in differentiated thyroid cancer.

Surgical and pathological reference standard

All included patients underwent thyroid surgery with central compartment lymph node assessment sufficient for postoperative pathological determination of central nodal burden. In the participating centers, central compartment lymph node dissection was performed as part of surgical management

in the included patients rather than being inferred from imaging alone. The extent of central neck dissection was determined by tumor location, multifocality, intraoperative findings, surgeon judgment, and institutional practice and included ipsilateral or bilateral central compartment dissection when clinically indicated. The numbers of central lymph nodes examined and metastatic central lymph nodes were extracted from postoperative pathology reports. Institutional pathologists evaluated central compartment specimens, and postoperative pathology was used only for end point ascertainment. To evaluate the adequacy of the pathological reference standard, central nodal yield and metastatic central nodal counts were summarized and recorded for the development, temporal validation, and external validation cohorts.

Data collection and candidate predictors

Preoperative variables were extracted from the electronic medical records and imaging archives of each institution. Candidate predictors were selected on the basis of preoperative availability and clinical applicability and comprised demographic characteristics, comorbidities, laboratory indices, and conventional ultrasound features. The laboratory variables included routine hematologic and biochemical markers and thyroid function-related indices. The conventional ultrasound variables included tumor loca-

tion, boundary, contour, irregular margin, halo sign, composition, echogenicity, aspect ratio, focal hyperechoic foci, posterior echo feature, vascularity, Chinese Thyroid Imaging Reporting and Data System (C-TIRADS) 4 subclass, diffuse thyroid disease, multifocality, and maximum tumor diameter.

Preoperative ultrasound examination and image review

Original preoperative ultrasound examinations were performed at the participating centers by experienced thyroid ultrasound physicians or sonographers as part of routine thyroid and neck ultrasound assessment. The examinations were conducted using high-frequency linear-array probes and followed institutional thyroid and neck ultrasound protocols, which included evaluation of the primary thyroid lesion and routinely accessible cervical lymph node compartments, including the central and lateral neck compartments. Because this was a retrospective, two-center study, the examinations were not performed by the same operators across both centers, and the degree of central compartment visualization at the time of original scanning could not be fully standardized.

Preoperative cN0 status was first determined according to the original clinical assessment and neck ultrasound reports. The available preoperative ultrasound images were then independently reviewed by two senior thyroid ultrasound experts who were blinded to postoperative pathological outcomes. This retrospective review was used to confirm the absence of definite metastatic cervical lymph nodes in the assessable neck compartments and to extract conventional ultrasound features of the index thyroid lesion. C-TIRADS 4 subclass was assigned during the blinded image review according to the suspicious ultrasound features used in the C-TIRADS framework, reflecting an overall structured ultrasound suspicion category. Margin contour was recorded separately as an explicit morphologic descriptor before model development. When discrepancies occurred, the images were further reviewed by a third senior thyroid ultrasound expert, and the final classification was established by consensus. For patients with multifocal PTC, the index lesion was defined as the lesion with the largest maximum diameter on preoperative ultrasound, and this lesion was matched to the dominant postoperative tumor focus for pathological correlation. The maximum tumor diameter was measured on preoperative ultrasound and recorded in centimeters.

Radiomics workflow

Radiomics analysis was performed on preoperative conventional grayscale ultrasound images of the index lesion. Original ultrasound images were exported in Digital Imaging and Communications in Medicine format and analyzed using a prespecified Python 3.6.5–based radiomics workflow. The region of interest was manually delineated along the contour of the index lesion on the selected preoperative ultrasound image by a senior thyroid ultrasound expert. To assess feature reproducibility, a randomly selected subset of lesions was independently resegmented by a second senior thyroid ultrasound expert who was blinded to the study outcome, and interobserver agreement was quantified using the intraclass correlation coefficient. Only radiomics features with good reproducibility were retained for subsequent analysis.

Radiomics features were extracted from the segmented grayscale ultrasound images, and they included shape, first-order statistical, gray-level co-occurrence matrix, gray-level run length matrix, gray-level size zone matrix, and neighboring gray-tone difference matrix features. To avoid scale-related bias, radiomics features were standardized using z-score normalization based on the mean and standard deviation of the development cohort, and the same transformation parameters were applied to the temporal validation and external validation cohorts.

Feature selection

All feature-selection procedures were performed exclusively in the development cohort to avoid information leakage. Candidate preoperative clinical variables were entered into a least absolute shrinkage and selection operator (LASSO) regression model, and the optimal penalization parameter was determined by 10-fold cross-validation. Clinical variables with non-zero coefficients were retained for downstream modeling. According to the final LASSO model, three clinical variables were selected: maximum tumor diameter, margin contour, and C-TIRADS 4 subclass. Margin contour and C-TIRADS 4 subclass were considered partially overlapping sonographic descriptors because C-TIRADS incorporates margin-related suspicious features. Therefore, their association and the robustness of the fusion model after excluding either descriptor were evaluated in sensitivity analyses.

For radiomics feature selection, reproducible features were first retained after assessment of interobserver agreement. To

further reduce redundancy and collinearity before shrinkage, the retained radiomics features were filtered using the minimum redundancy maximum relevance algorithm. The selected candidate radiomics features were then entered into a 10-fold cross-validated LASSO model, and features with non-zero coefficients were retained to construct the radiomics signature. A total of 20 radiomics features were ultimately selected and spanned shape, first-order statistical, gray-level co-occurrence matrix, gray-level run length matrix, gray-level size zone matrix, and neighboring gray-tone difference matrix features. The radiomics signature was calculated as a linear combination of the retained features weighted by their corresponding coefficients and was defined as the radiomics score (Rad-score). Because the retained radiomics signature included shape-, size-, and texture-related descriptors, Rad-score was interpreted as an integrated ultrasound–radiomics summary rather than as a purely texture-based or latent biological marker. The Rad-score was subsequently used for radiomics model development and construction of the final fusion model. To further examine the contribution of non-shape radiomics information, a sensitivity radiomics signature was reconstructed after excluding shape- and size-related radiomics features.

Model development and algorithm comparison

All model development procedures were performed exclusively in the development cohort. After completion of clinical variable selection and construction of the radiomics signature, three model categories were prespecified: a clinical model, a radiomics model, and a fusion model. The clinical model was constructed using the retained preoperative clinical predictors, the radiomics model was constructed using the radiomics signature (Rad-score), and the fusion model was constructed by integrating the retained clinical predictors with the radiomics signature.

To compare candidate machine-learning frameworks, four classifiers were prespecified: logistic regression, random forest, XGBoost, and CatBoost. Algorithm comparison was conducted in the development cohort under the fusion setting because the fusion setting incorporated the full predictor information available before surgery. For each candidate classifier, model fitting, hyperparameter optimization, and internal resampling were completed using the development cohort only.

Hyperparameter tuning was performed using a grid-search strategy combined with 10-fold cross-validation within the development cohort. The mean cross-validated area under the receiver operating characteristic (ROC) curve (AUC) was used as the primary optimization criterion. For logistic regression, the regularization strength was tuned. For random forest, the number of trees, maximum tree depth, minimum samples required for node splitting, minimum samples required for terminal nodes, and the number of features considered at each split were tuned. For XGBoost, the number of boosting iterations, learning rate, maximum tree depth, subsampling fraction, and column subsampling fraction were tuned. For CatBoost, the number of boosting iterations, learning rate, tree depth, and L2 regularization strength were tuned. All optimized hyperparameters were fixed after model training in the development cohort and were not reestimated in the temporal validation or external validation cohorts. It should be noted that, for the internal algorithm comparison, feature selection and Rad-score construction had been completed within the development cohort before candidate classifiers were compared; therefore, this comparison was not a fully nested cross-validation procedure in which the entire preprocessing, feature selection, signature construction, tuning, and threshold selection pipeline was repeated within each training fold.

After algorithm screening, the selected classifier was used to construct the final clinical, radiomics, and fusion models in the development cohort. Model outputs were generated as predicted probabilities of occult high-volume CLNM. The classification threshold was determined in the development cohort and then held constant for temporal and external validation. To minimize information leakage, all preprocessing, feature selection, signature construction, algorithm tuning, and threshold determination steps were completed before evaluation in the validation cohorts.

Model performance was assessed at three levels. Discrimination was evaluated using ROC curves and the AUC. Calibration was assessed using calibration plots, intercepts, and slopes and Brier scores. Clinical usefulness was evaluated using decision curve analysis (DCA) and clinical impact curves (CICs) across clinically relevant threshold probabilities. The finalized models were then independently tested in the temporal validation cohort and the external validation cohort without refitting or retuning.

Temporal validation and external validation

Temporal and external validation were performed using the finalized clinical, radiomics, and fusion models and the classification threshold derived from the development cohort. In each validation cohort, AUCs were reported with 95% CIs, and sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated at the prespecified threshold.

Clinical usefulness was assessed using DCA and CICs across clinically relevant threshold probabilities.

Model interpretability and representative-case visualization

To improve the interpretability of the final fusion model, SHAP values were calculated. Global interpretability was assessed using SHAP summary plots, in which the contribution rankings of the four final input features were evaluated, and SHAP dependence plots, which were generated for Rad-score, maximum tumor diameter, margin contour, and C-TIRADS 4 subclass. Local interpretability was assessed using SHAP waterfall plots for representative true-negative, false-negative, true-positive, and false-positive cases. To further illustrate model behavior at the individual-patient level, the corresponding preoperative ultrasonographic images and postoperative histopathologic sections of these representative cases were displayed.

Web-based calculator

A web-based calculator was developed based on the final fusion model to provide individualized preoperative risk estimation of occult high-volume CLNM.

Statistical analysis

Continuous variables were expressed as mean $\pm$ standard deviation or median [interquartile range (IQR)], as appropriate, according to variable distribution. Categorical variables were expressed as counts and percentages. Baseline differences among the development, temporal validation, and external validation cohorts were assessed using one-way analysis of variance or the Kruskal–Wallis test for continuous variables and the chi-square test for categorical variables. Comparisons between patients who were outcome negative or outcome positive within the development cohort were performed using the Student t-test or Mann–Whitney U test for continuous variables and the chi-

square test or Fisher's exact test for categorical variables, as appropriate. Additional sensitivity analyses were performed to evaluate potential overlap among the final predictors. The association between margin contour and C-TIRADS 4 subclass was assessed using Cramér's V. To examine whether the fusion model depended on overlapping clinical ultrasound descriptors, two sensitivity fusion models were reconstructed in the development cohort by excluding either margin contour or C-TIRADS 4 subclass and were then evaluated without refitting in the temporal validation and external validation cohorts. To assess whether the radiomics contribution was mainly attributable to shape- or size-related features, shape- and size-related radiomics features were excluded, and a non-shape, texture-focused radiomics signature was reconstructed in the development cohort using the same feature-selection workflow. The corresponding texture-focused radiomics and fusion models were then evaluated without refitting in the temporal validation and external validation cohorts. Sensitivity models were compared with the primary models using AUC, Brier score, sensitivity, specificity, and accuracy. All analyses were two-sided, and $P < 0.05$ was considered statistically significant. Statistical analyses were performed using R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) and Python version 3.6.5 (Python Software Foundation, Wilmington, DE, USA). Model development, radiomics processing, and SHAP analyses were conducted within the prespecified Python workflow, and cohort-level statistical analyses were performed in R. This study was designed and reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Plus Artificial Intelligence framework for multivariable prediction model studies.

Results

Baseline characteristics

A total of 814 patients were included, comprising 470 patients in the development cohort, 202 in the temporal validation cohort, and 142 in the external validation cohort. The distribution of occult high-volume CLNM was similar across the three cohorts, with 213 of 470 patients (45.3%) in the development cohort, 92 of 202 patients (45.5%) in the temporal validation cohort, and 64 of 142 patients (45.1%) in the external validation cohort classified as positive ($P = 0.996$). Most baseline characteristics were compa-

rable among the three cohorts (all $P > 0.05$), except for the anteroposterior/transverse (A/T) diameter ratio ($P = 0.017$). The median age was 47.00 (37.00–55.00) years in the development cohort, 47.00 (38.25–56.00) years in the temporal validation cohort, and 48.00 (37.25–55.00) years in the external validation cohort ($P = 0.579$). Female patients accounted for 71.5%, 67.3%, and 78.2% of the three cohorts, respectively ($P = 0.089$). For aspect ratio A/T, the proportions of $A/T < 1$ were 39.8%, 51.0%, and 38.7% in the development, temporal validation, and external validation cohorts, respectively, whereas the proportions of $A/T \geq 1$ were 60.2%, 49.0%, and 61.3%, respectively ($P = 0.017$). Detailed baseline characteristics of the three cohorts are provided in Supplementary Table 1. Postoperative pathological assessment of the central compartment was available for all included patients. The median number of central lymph nodes examined was 13 (IQR, 8–17; range, 3–38) in the development cohort, 12 (IQR, 9–18; range, 4–30) in the temporal validation cohort, and 11 (IQR, 7–16; range, 3–36) in the external validation cohort. The corresponding median numbers of metastatic central lymph nodes were 3 (IQR, 0–9; range, 0–26), 3 (IQR, 0–8; range, 0–24), and 3 (IQR, 0–8; range, 0–22), respectively. Detailed surgical and pathological reference-standard information is provided in Supplementary Table 2.

Baseline characteristics in the development cohort

In the development cohort, 470 patients were included, comprising 257 outcome-negative patients and 213 outcome-positive patients. Most baseline variables did not differ significantly between the two groups (all $P > 0.05$). The median age was 48.00 (37.00–56.00) years in the outcome-negative group and 45.00 (36.00–54.00) years in the outcome-positive group ($P = 0.060$). Female patients accounted for 70.8% and 72.3% of the two groups, respectively ($P = 0.801$). Among continuous variables, red blood cell count was $4.61 (4.33–4.97) \times 10^{12}/L$ in the outcome-negative group and $4.74 (4.40–5.05) \times 10^{12}/L$ in the outcome-positive group ($P = 0.021$), thyroid-stimulating hormone was 1.80 (1.19–2.66) mIU/L and 2.01 (1.49–2.67) mIU/L, respectively ($P = 0.021$), and creatinine was 59.00 (52.00–69.00) $\mu\text{mol}/L$ and 61.00 (55.00–70.00) $\mu\text{mol}/L$, respectively ($P = 0.030$). Among categorical variables, composition differed between the two groups ($P = 0.041$). In the outcome-negative group, the proportions of predominantly cystic,

predominantly solid, and solid nodules were 0.0%, 3.1%, and 96.9%, respectively, compared with 0.9%, 7.0%, and 92.0% in the outcome-positive group. The distribution of aspect ratio A/T also differed between the groups ($P = 0.026$), with $A/T \geq 1$ observed in 65.0% of the outcome-negative group and 54.5% of the outcome-positive group. Rich vascularity was observed in 1.9% of the outcome-negative group and 8.9% of the outcome-positive group ($P = 0.001$). The distribution of C-TIRADS 4 subclass differed between the two groups ($P = 0.002$), and the proportion of C-TIRADS 4C (high suspicion) was 56.4% in the outcome-negative group and 71.8% in the outcome-positive group. Multifocality was present in 25.3% of the outcome-negative group and 34.7% of the outcome-positive group ($P = 0.033$). Detailed comparisons between outcome-negative and outcome-positive patients in the development cohort are provided in Supplementary Table 3.

Selection of clinical variables

In the development cohort, LASSO regression was applied to the candidate clinical variables. The coefficient profiles are shown in Figure 2a, and the mean cross-validated AUC across different regularization parameters is shown in Figure 2b. At the selected tuning parameter indicated by the dashed line, three clinical variables with non-zero coefficients were retained: maximum tumor diameter, margin contour, and C-TIRADS 4 subclass (Figure 2c). Because margin contour and C-TIRADS 4 subclass may partially encode overlapping suspicious ultrasound morphology, their association was further evaluated. Cramér's V values were 0.28 in the development cohort, 0.30 in the temporal validation cohort, 0.27 in the external validation cohort, and 0.29 in the overall cohort, with corresponding P values of < 0.001 , < 0.001 , 0.006, and < 0.001 , respectively. The association analysis is summarized in Supplementary Table 4.

Selection of radiomics features

In the development cohort, LASSO regression was performed for radiomics feature selection. The coefficient profiles are shown in Figure 3a, and the mean cross-validated AUC across different regularization parameters is shown in Figure 3b. At the selected tuning parameter indicated by the dashed line, 20 radiomics features with non-zero coefficients were retained (Figure 3c), including shape, first-order statistical, gray-level co-occurrence matrix, gray-level run length matrix,

gray-level size zone matrix, and neighboring gray-tone difference matrix features.

Comparison of candidate algorithms for the fusion model

Using the selected clinical and radiomics features, four candidate algorithms were developed in the development cohort: logistic regression, random forest, XGBoost, and CatBoost. The AUCs were 0.779 (95% CI, 0.740–0.823), 0.994 (95% CI, 0.990–0.998), 0.979 (95% CI, 0.969–0.987), and 0.946 (95% CI, 0.927–0.961), respectively (Figure 4a and Supplementary Table 5). Among these algorithms, random forest yielded the highest AUC.

At the optimal cut-off derived from the development cohort, random forest showed a sensitivity of 0.925 (95% CI, 0.881–0.953), a specificity of 0.984 (95% CI, 0.961–0.994), and an accuracy of 0.957 (95% CI, 0.935–0.972), with a Brier score of 0.068. XGBoost showed a sensitivity of 0.939 (95% CI, 0.898–0.964), a specificity of 0.907 (95% CI, 0.865–0.936), and an accuracy of 0.921 (95% CI, 0.893–0.942), with a Brier score of 0.077. The corresponding AUCs for CatBoost and logistic regression were 0.946 and 0.779, respectively.

Calibration curves are shown in Figure 4b, and DCA is shown in Figure 4c. Across threshold probabilities of approximately 0.05 to 0.80, the random forest, XGBoost, and CatBoost models all showed net benefit above the treat-none strategy, and the random forest curve remained at or near the uppermost position among the compared algorithms across most of the threshold range. Random forest was selected for subsequent model construction.

Model performance and clinical utility in the validation cohorts

Using the random forest algorithm, the clinical, radiomics, and fusion models were evaluated in the temporal validation cohort and the external validation cohort. In the temporal validation cohort, the AUCs were 0.703 (95% CI, 0.631–0.773) for the clinical model, 0.755 (95% CI, 0.687–0.817) for the radiomics model, and 0.845 (95% CI, 0.794–0.893) for the fusion model. In the external validation cohort, the corresponding AUCs were 0.676 (95% CI, 0.601–0.748), 0.730 (95% CI, 0.660–0.795), and 0.817 (95% CI, 0.758–0.870), respectively. For the fusion model, sensitivity, specificity, and accuracy were 0.837, 0.695, and 0.755 in the temporal validation cohort and 0.769, 0.761, and 0.764 in the external validation cohort, respectively.

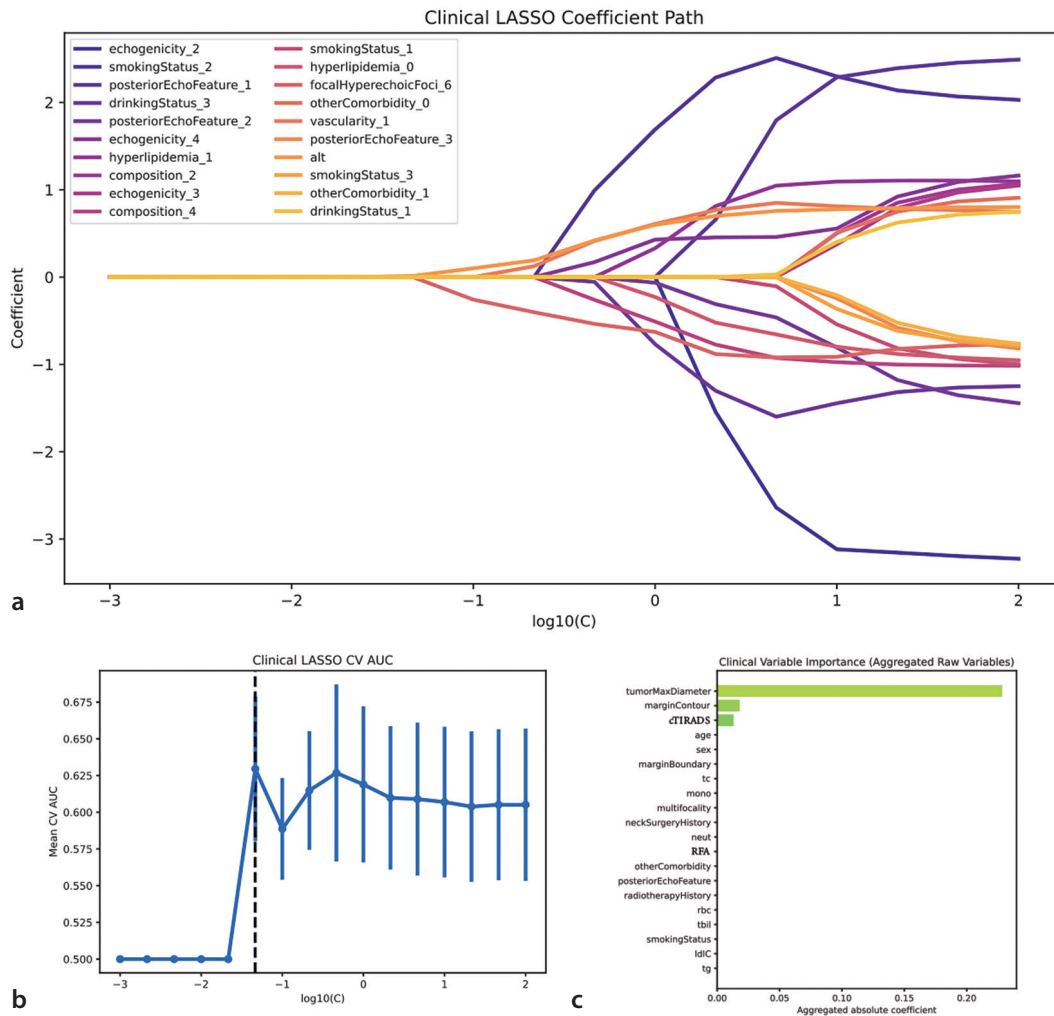


Figure 2. LASSO-based feature selection of clinical variables in the development cohort. (a) LASSO coefficient profiles of candidate clinical variables plotted against $\log_{10}(C)$. (b) Mean cross-validated area under the receiver operating characteristic (ROC) curve (AUC) across different regularization parameters; the dashed vertical line indicates the selected tuning parameter. (c) Aggregated absolute coefficients of the retained raw clinical variables. Three clinical variables were retained after LASSO shrinkage: maximum tumor diameter, margin contour, and C-TIRADS 4 subclass. LASSO, least absolute shrinkage and selection operator; C-TIRADS 4, Chinese Thyroid Imaging Reporting and Data System 4 subclass.

The Brier scores of the fusion model were 0.158 and 0.154 in the two validation cohorts. Calibration, decision-curve, and clinical-impact analyses are shown in Figure 5a–h, and detailed performance metrics are summarized in Supplementary Table 6.

Sensitivity analysis of overlapping clinical ultrasound descriptors

To further evaluate whether the fusion model depended on overlapping clinical ultrasound descriptors, two sensitivity fusion models were constructed by excluding either margin contour or C-TIRADS 4 subclass. Compared with the primary fusion model, the model without margin contour showed only a small decrease in AUC, from 0.845 to 0.833 in the temporal validation cohort and from 0.817 to 0.804 in the external validation cohort. Similarly, the model without C-TI-

RADS 4 subclass showed AUCs of 0.837 and 0.809 in the temporal and external validation cohorts, respectively. Brier scores, sensitivity, specificity, and accuracy were broadly comparable across the primary and sensitivity models. Detailed results are shown in Supplementary Table 7.

Sensitivity analysis of radiomics feature categories

A further sensitivity analysis was performed to examine whether the radiomics contribution was mainly attributable to shape- and size-related features. After excluding shape- and size-related radiomics features, a non-shape, texture-focused radiomics signature was reconstructed in the development cohort and evaluated in the validation cohorts. The texture-focused radiomics model achieved AUCs of 0.733 in the

temporal validation cohort and 0.711 in the external validation cohort. When this texture-focused radiomics signature was combined with clinical predictors, the texture-focused fusion model achieved AUCs of 0.822 and 0.798 in the temporal and external validation cohorts, respectively. Although these values were slightly lower than those of the primary fusion model, the texture-focused fusion model retained favorable discrimination and calibration. Detailed results are presented in Supplementary Table 8.

Model interpretability analysis and representative-case visualization

SHAPs analysis was performed for the fusion model. In the SHAP summary plot, the four input features were ranked by their overall contributions to model output as Rad-score, maximum tumor diameter, mar-

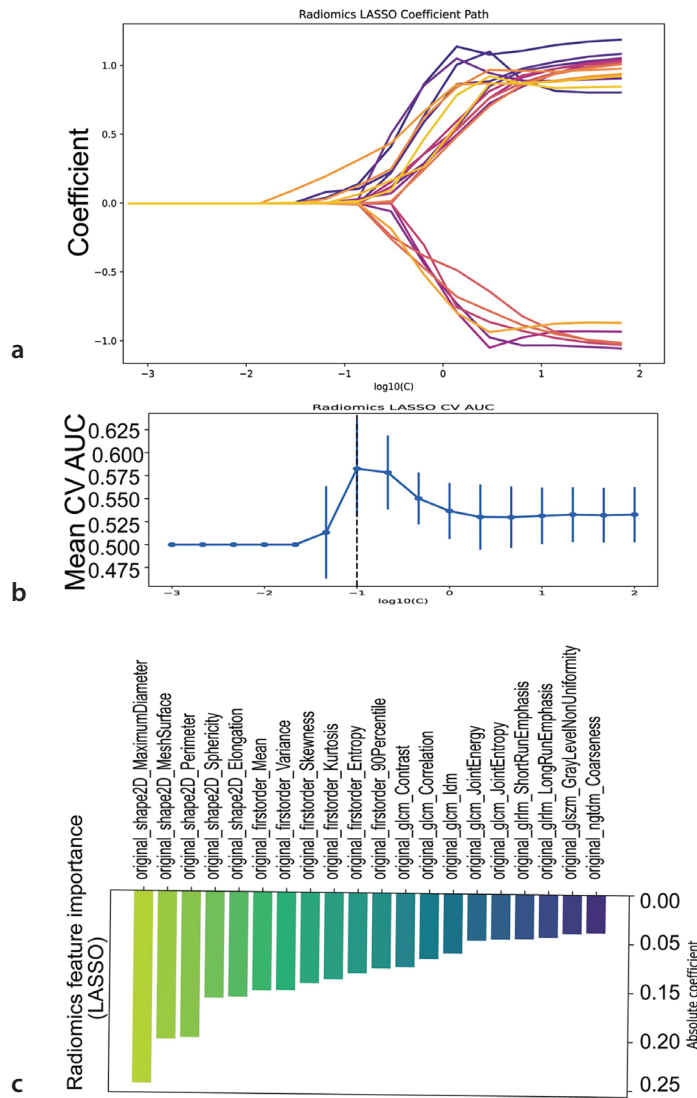


Figure 3. LASSO-based feature selection of radiomics features in the development cohort. (a) LASSO coefficient profiles of candidate radiomics features plotted against $\log_{10}(C)$. (b) Mean cross-validated area under the receiver operating characteristic curve (AUC) across different regularization parameters; the dashed vertical line indicates the selected tuning parameter. (c) Absolute coefficients of the retained radiomics features after LASSO shrinkage. A total of 20 radiomics features with non-zero coefficients were retained. LASSO, least absolute shrinkage and selection operator.

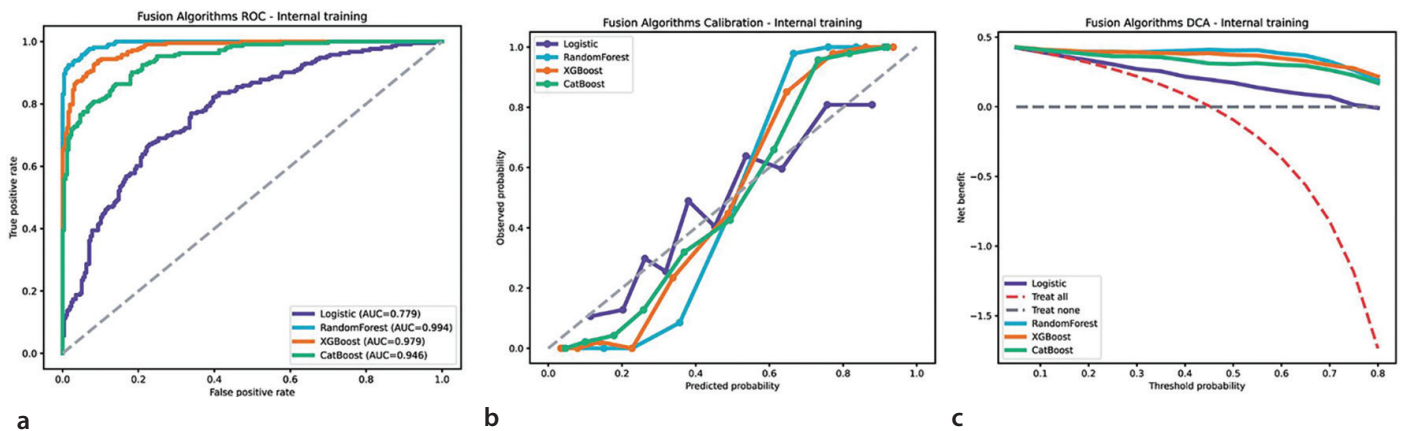


Figure 4. Comparison of candidate algorithms for the fusion model in the development cohort. (a) Receiver operating characteristic curves of the logistic regression, random forest, XGBoost, and CatBoost algorithms. (b) Calibration curves of the four algorithms. The gray dashed line indicates the ideal calibration line. (c) Decision curve analysis of the four algorithms together with the treat-all and treat-none strategies. ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis.

gin contour, and C-TIRADS 4 subclass (Figure 6a). Among these features, Rad-score showed the widest SHAP value distribution, followed by maximum tumor diameter, whereas margin contour and C-TIRADS 4 subclass showed relatively narrower distributions. In the dependence plots, SHAP values for Rad-score and maximum tumor diameter generally increased with increasing feature values, whereas those for margin contour and C-TIRADS 4 subclass were mainly distributed around zero with scattered positive and negative values (Figure 6b).

Representative SHAP waterfall plots are shown in Figure 6c. The predicted values were 0.21 for the true-negative case,

0.30 for the false-negative case, 0.676 for the true-positive case, and 0.723 for the false-positive case, compared with a baseline value of 0.452. The positive and negative contributions of each feature to the individual predictions are shown in the corresponding waterfall plots.

Representative ultrasonographic images and corresponding histopathologic sections of the true-negative, false-negative, true-positive, and false-positive cases are presented in Figure 6d.

Web-based calculator

A web-based calculator was developed based on the final fusion model for individu-

alized risk estimation of occult high-volume CLNM (Figure 7). The calculator incorporated four input variables: C-TIRADS 4 subclass, margin contour, maximum tumor diameter, and Rad-score. The initial interface of the calculator is shown in Figure 7a. Representative outputs are shown in Figure 7b and 7c. In the low-risk example, the estimated probability was 21.6%, which was below the decision threshold of 0.50. In the high-risk example, the estimated probability was 76.5%, which was above the decision threshold of 0.50. The web-based calculator is available at <https://thyroid-cancers6-gdhu6ypoguprlscktfzpk.streamlit.app/>.

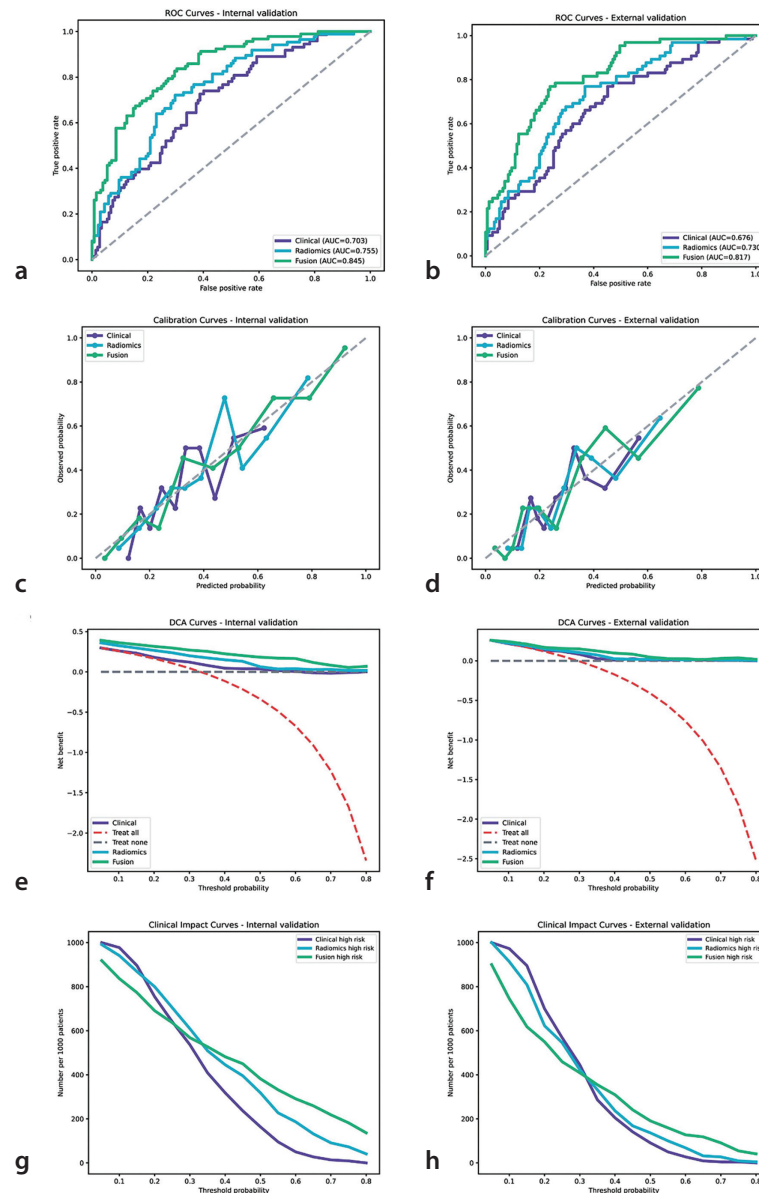


Figure 5. Performance of the clinical, radiomics, and fusion models in the temporal validation and external validation cohorts. (a, b) Receiver operating characteristic curves of the clinical, radiomics, and fusion models in the temporal validation and external validation cohorts. (c, d) Calibration curves of the three models in the temporal validation and external validation cohorts. The gray dashed line indicates the ideal calibration line. (e, f) Decision curve analysis of the three models in the temporal validation and external validation cohorts, together with the treat-all and treat-none strategies. (g, h) Clinical impact curves of the three models in the temporal validation and external validation cohorts. ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis.

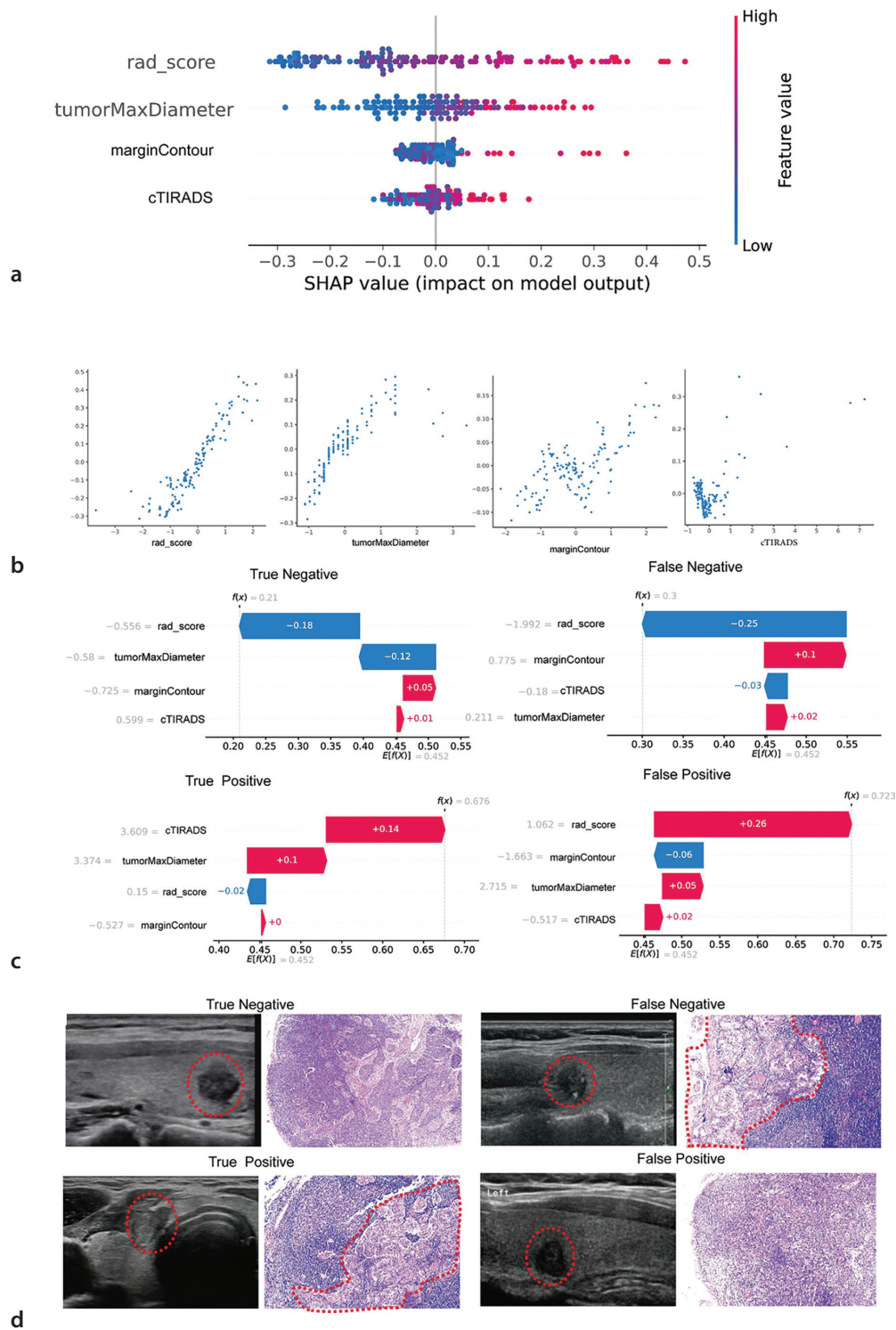


Figure 6. SHAP-based interpretability analysis and representative case visualization of the fusion model. (a) SHAP summary plot showing the distribution of SHAP values for the four input features. Features are ranked according to their overall contributions to model output. (b) SHAP dependence plots for Rad-score, maximum tumor diameter, margin contour, and C-TIRADS 4 subclass. (c) Representative SHAP waterfall plots for a true-negative case, a false-negative case, a true-positive case, and a false-positive case, showing the contribution of each feature to the individual predicted probability. (d) Representative ultrasonographic images and corresponding histopathologic sections of the true-negative, false-negative, true-positive, and false-positive cases. SHAP, SHapley Additive exPlanations; C-TIRADS 4, Chinese Thyroid Imaging Reporting and Data System 4 subclass; Rad-score, radiomics score.

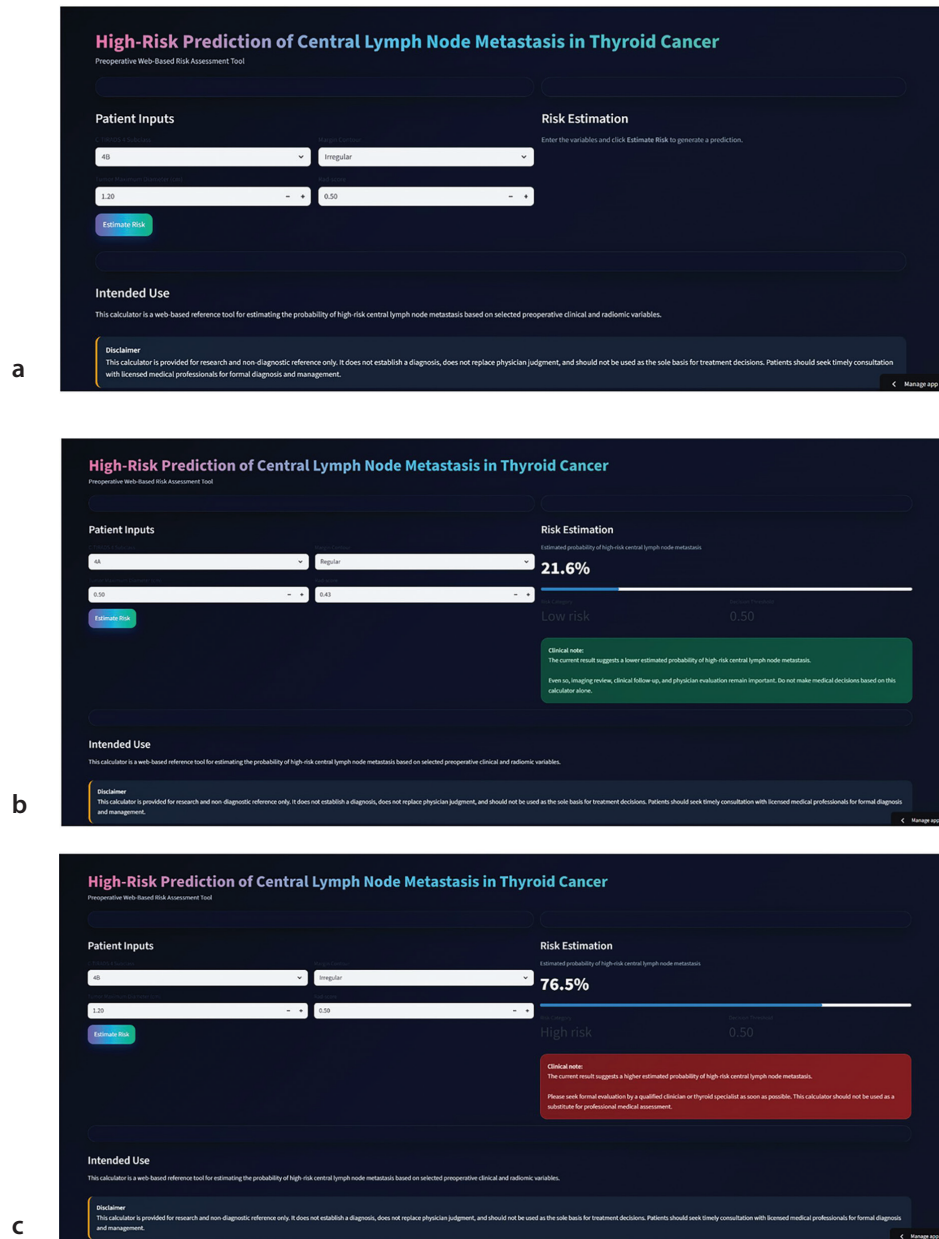


Figure 7. Web-based calculator for individualized risk estimation of occult high-volume central lymph node metastasis. (a) Initial interface of the web-based calculator showing the input fields for C-TIRADS 4 subclass, margin contour, maximum tumor diameter, and rad-score. (b) Representative low-risk output of the calculator, showing an estimated probability of 21.6%, which was below the decision threshold of 0.50. (c) Representative high-risk output of the calculator, showing an estimated probability of 76.5%, which was above the decision threshold of 0.50. C-TIRADS 4, Chinese Thyroid Imaging Reporting and Data System 4 subclass; Rad-score, radiomics score.

Discussion

The present study developed and validated a practical preoperative model for predicting the risk of occult high-volume CLNM in patients with cN0 PTC. Several aspects of this work deserve emphasis. First, rather than focusing on any-volume nodal involvement, we specifically targeted high-volume CLNM, an end point that is more closely linked to clinically meaningful nodal burden and postoperative risk stratification than incidental low-volume metastasis.^{12,13} Second,

the final fusion strategy, integrating conventional ultrasound predictors with an ultrasound–radiomics signature, showed more favorable and more stable performance than either the clinical or radiomics model alone across both temporal and external validation cohorts. Third, model interpretability analyses indicated that Rad-score and maximum tumor diameter contributed most strongly to model output, followed by margin contour and C-TIRADS 4 subclass. This result should be interpreted as a model-behavior finding rather than as evidence

that the radiomics signature purely reflects latent texture heterogeneity independent of conventional size and morphology. Among the retained predictors, maximum tumor diameter is probably the most biologically intuitive. Tumor size has consistently emerged as one of the most robust determinants of occult nodal spread in PTC, and a recent systematic review and meta-analysis by Tang et al.³ showed that the prevalence of occult central nodal metastasis increases with increasing primary tumor size. In the specific setting of high-volume central disease, Wang

et al.¹⁴ reported that larger tumor size was significantly associated with high-volume CLNM in cN0 papillary thyroid microcarcinoma, and Huang et al.¹⁵ similarly found that sonographic tumor size predicted high-volume central neck nodal metastasis. Zhu et al.⁵ also showed in a large contemporary cohort that lesion size remained an independent predictor in a clinical model for high-volume CLNM. Our findings are therefore concordant with the existing literature. Mechanistically, larger tumors may indicate longer intrathyroidal growth, a greater tumor-cell burden, more opportunities for lymphovascular permeation, and a higher probability of subclinical extracapsular or perilymphatic extension. The fact that tumor size remained important even after the addition of radiomics features suggests that conventional ultrasound still preserves essential predictive information that cannot be fully replaced by quantitative image analysis.

A second important observation is the retention of margin contour and C-TIRADS 4 subclass in the final fusion model. These variables are clinically meaningful because they encode morphologic suspiciousness that can be directly appreciated on routine ultrasound. Huang et al.,¹⁶ in a study explicitly based on C-TIRADS analysis, demonstrated that higher C-TIRADS scores, larger nodules, multifocality, and adverse sonographic features were independently associated with CLNM. Earlier sonographic work on high-volume central nodal disease also showed that suspicious ultrasound morphology of the primary lesion correlated with greater nodal burden.¹⁵ These two variables are partially overlapping because C-TIRADS 4 subclass incorporates margin-related suspicious features, whereas margin contour was recorded separately as an explicit descriptor. Their joint retention is best viewed as a practical modeling result in which a structured overall suspicion category and a specific margin descriptor both contributed to prediction in this dataset. This interpretation was supported by the weak-to-moderate Cramér's V values and by the sensitivity analyses showing only limited attenuation in performance after excluding either descriptor. Of note, some variables that were associated with the outcome at the baseline-comparison stage, such as vascularity and multifocality, were not retained in the final model. This does not necessarily indicate irrelevance; rather, their predictive information may have been captured by more informative composite variables, collinear morphologic descriptors, or the radiomics signature during shrinkage and machine-learning optimization.¹⁷⁻¹⁹

The radiomics signature ranked first in SHAP importance, but this finding also requires cautious interpretation. Ultrasound radiomics quantifies grayscale heterogeneity, shape irregularity, spatial co-occurrence, run-length distribution, and regional gray-level non-uniformity. Therefore, Rad-score in the present model should be understood as an integrated radiomics summary that includes both shape-, size-, and texture-related descriptors rather than as a pure marker of latent biological heterogeneity. Prior ultrasound-radiomics studies have repeatedly shown that quantitative image signatures can improve preoperative prediction of CLNM in PTC. Zhou et al.²⁰ established an ultrasound-radiomics nomogram for central neck lymph node metastasis and reported clinically useful discrimination in the validation cohort. Tong et al.²¹ subsequently confirmed in a multi-institutional study that ultrasound-based radiomics was valuable for predicting both central and lateral cervical nodal metastasis. More recently, Feng et al.⁷ developed a clinical-radiomics nomogram in Academic Radiology and again showed that combining radiomics with key clinical variables improved prediction of CLNM. Our findings are consistent with this broader literature. The sensitivity analysis excluding shape- and size-related radiomics features further showed that the texture-focused fusion model retained favorable performance in both validation cohorts, although its performance was slightly lower than that of the primary fusion model. This finding suggests that radiomics provided predictive information beyond conventional shape- and size-related descriptors while also indicating that shape, size, and texture jointly contributed to risk estimation. Therefore, Rad-score should be interpreted as an integrated quantitative ultrasound descriptor rather than as a purely texture-based marker or a direct surrogate of latent biological heterogeneity.

An additional issue worth discussing is why the fusion model outperformed both unimodal models and why, among the candidate algorithms, the final modeling framework achieved the most stable overall performance. In clinical terms, conventional ultrasound variables and radiomics may capture complementary but partly overlapping layers of tumor information: the former reflects structured, expert-recognizable morphology, whereas the latter summarizes quantitative image descriptors that include shape, intensity, and texture.^{22,23} Their combination may reduce information loss, but it should not be interpreted as proof that every component contributes independent biolog-

ical information. This concept is increasingly supported by modern multimodal AI studies. Shen et al.,⁹ in a recent *Nature Communications* study on thyroid cancer nodal prediction, showed that multimodal integration of ultrasound imaging with other complementary information improved predictive performance and interpretability relative to unimodal approaches. Although their end point was lateral rather than central nodal metastasis and their technical architecture was deep-learning based, the conceptual message is relevant to our findings: Multimodal fusion can improve prediction by combining complementary preoperative information. The clinical value of this study lies in its potential to support preoperative risk stratification in cN0 PTC, not to replace clinical judgment or directly detect occult nodes. In routine practice, preoperative detection of central compartment metastasis by ultrasound remains imperfect, largely because of the deep anatomic location of central nodes and shielding by the thyroid gland itself.^{2,16} This limitation becomes more important when the occult nodal burden is high because high-volume nodal metastasis has greater implications for recurrence-oriented risk stratification than low-volume microscopic disease.^{12,13} The final model uses four preoperative inputs—maximum tumor diameter, margin contour, C-TIRADS 4 subclass, and Rad-score—that are obtainable before surgery and may provide an adjunctive estimate of occult high-volume central nodal burden. However, the present study only establishes retrospective predictive performance. Prospective validation and decision-impact studies are required before determining whether model-guided risk estimation can change surgical decision-making, reduce unnecessary central neck dissection, or improve patient outcomes. Several limitations should be acknowledged. First, this was a retrospective study, and retrospective model development carries a risk of selection bias and residual confounding despite predefined eligibility criteria and blinded image assessment. The event rate of occult high-volume CLNM was approximately 45% in all cohorts, which is higher than would be expected in an unselected cN0 PTC population. This finding may reflect referral-center case mix, inclusion of surgically treated patients with postoperative pathological assessment of the central compartment, and selection related to central neck dissection. Therefore, this event rate should not be interpreted as the population-level prevalence of occult high-volume CLNM among all patients with cN0 PTC. Sec-

ond, although temporal and external validation were used to evaluate the finalized model, the internal algorithm-comparison procedure was not fully nested with respect to feature selection, radiomics signature construction, hyperparameter tuning, and threshold selection. Therefore, the very high development-cohort AUC of the random forest model may have been optimistically biased and should be interpreted cautiously. The temporal and external validation results are more clinically relevant for judging model generalizability. Third, although our study included both temporal and external validation, the two validation cohorts were derived from the same broader regional practice environment; therefore, wider geographic and ethnic validation remains necessary. Fourth, original ultrasound examinations were performed in routine clinical practice rather than under a prospective, study-specific imaging protocol, and the same operator group did not perform all examinations across centers. Central compartment assessment is inherently operator dependent and may be limited by its anatomic location; therefore, false-negative preoperative ultrasound assessment may have contributed to cN0 classification. Fifth, our end point—high-volume CLNM defined as more than five metastatic central lymph nodes—is clinically meaningful and consistent with the recurrence-oriented literature,^{12–14} but it remains influenced by the extent of central neck dissection and pathological sampling. Although central nodal yield and metastatic central nodal counts were summarized across cohorts, the extent of dissection was determined by routine clinical practice rather than by a prospective, study-specific surgical protocol. In addition, although the median central nodal yield was 13, 12, and 11 nodes in the development, temporal validation, and external validation cohorts, respectively, the range included a small number of patients with only three to four examined central lymph nodes. Because the primary end point was defined as more than five metastatic central lymph nodes, low-yield dissections may have resulted in some patients being classified as end point–negative despite possible unexamined nodal disease. Future studies using prospectively standardized central compartment assessment are needed to reduce this source of end point misclassification. Sixth, margin contour and C-TIRADS 4 subclass may contain overlapping sonographic information, and the retained radiomics features included shape- and size-related descriptors that may overlap with clinical tumor diameter and morphology. Although sensitivity

analyses excluding overlapping clinical ultrasound descriptors and shape- and size-related radiomics features were performed, these analyses were retrospective and post-hoc. Therefore, the independent contribution of individual ultrasound morphology and radiomics feature categories should be further evaluated in prospective multicenter datasets.

In conclusion, this study developed and validated a practical preoperative multimodal model for estimating the risk of occult high-volume CLNM in cN0 PTC. The fusion model integrating conventional ultrasound predictors with an ultrasound–radiomics signature showed favorable performance in the temporal and external validation cohorts. The sensitivity analyses supported the robustness of the fusion model while also indicating that the final predictors include partially overlapping clinical morphologic and radiomics shape- and size-related information. Therefore, the model should be interpreted as a practical preoperative risk-estimation tool rather than as evidence of independent biological mechanisms. Further prospective validation and decision-impact evaluation are required before routine clinical deployment.

Acknowledgements

The authors thank the Departments of Ultrasonography and Pathology of Fujian Provincial Hospital and Fujian Union Hospital for their support in data collection and case review.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This study was supported by the Fujian Provincial Natural Science Foundation (No. 2025J08042).

Supplementary Tables: <https://d2v96fxpocvxx.cloudfront.net/beb8919b-f013-4ea1-b1c8-40332e840fe1/content-images/4e9c9547-a2f3-4dd7-8790-b167d13aa5e2.pdf>

References

- Chen DW, Haymart MR. Unravelling the rise in thyroid cancer incidence and addressing overdiagnosis. *Nat Rev Endocrinol.* 2026;22(1):10–20. [\[Crossref\]](#)

- Alabousi M, Alabousi A, Adham S, et al. Diagnostic test accuracy of ultrasonography vs computed tomography for papillary thyroid cancer cervical lymph node metastasis: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg.* 2022;148(2):107–118. [\[Crossref\]](#)
- Tang L, Qu RW, Park J, Simental AA, 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(8):7335–7350. [\[Crossref\]](#)
- Huang J, Song M, Shi H, et al. Predictive factor of large-volume central lymph node metastasis in clinical N0 papillary thyroid carcinoma patients underwent total thyroidectomy. *Front Oncol.* 2021;11:574774. [\[Crossref\]](#). Erratum in: *Front Oncol.* 2024;14:1517682. [\[Crossref\]](#)
- Zhu H, Zhang H, Wei P, et al. Development and validation of a clinical predictive model for high-volume lymph node metastasis of papillary thyroid carcinoma. *Sci Rep.* 2024;14:15828. [\[Crossref\]](#)
- Li J, Sun P, Huang T, et al. Preoperative prediction of central lymph node metastasis in cN0T1/T2 papillary thyroid carcinoma: a nomogram based on clinical and ultrasound characteristics. *Eur J Surg Oncol.* 2022;48(6):1272–1279. [\[Crossref\]](#)
- Feng JW, Liu SQ, Qi GF, et al. Development and validation of clinical-radiomics nomogram for preoperative prediction of central lymph node metastasis in papillary thyroid carcinoma. *Acad Radiol.* 2024;31(6):2292–2305. [\[Crossref\]](#)
- Jia W, Cai Y, Wang S, Wang J. Predictive value of an ultrasound-based radiomics model for central lymph node metastasis of papillary thyroid carcinoma. *Int J Med Sci.* 2024;21(9):1701–1709. [\[Crossref\]](#)
- Shen P, Yang Z, Sun J, et al. Explainable multimodal deep learning for predicting thyroid cancer lateral lymph node metastasis using ultrasound imaging. *Nat Commun.* 2025;16(1):7052. [\[Crossref\]](#)
- Qu L, Zhu J, Mei X, et al. Evaluating axillary lymph node metastasis risks in breast cancer patients via Semi-ALNP: a multicenter study. *EClinicalMedicine.* 2025;85:103311. [\[Crossref\]](#)
- Yao S, Shen P, Dai F, et al. Thyroid cancer central lymph node metastasis risk stratification based on homogeneous positioning deep learning. *Research (Wash D C).* 2024;7:0432. [\[Crossref\]](#)
- 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):1–133. [\[Crossref\]](#)
- Randolph GW, Duh QY, Heller KS, et al. The prognostic significance of nodal metastases

- from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid*. 2012;22(11):1144-52. [\[Crossref\]](#)
14. Wang Z, Gui Z, Wang Z, et al. Clinical and ultrasonic risk factors for high-volume central lymph node metastasis in cN0 papillary thyroid microcarcinoma: a retrospective study and meta-analysis. *Clin Endocrinol (Oxf)*. 2023;98(4):609-621. [\[Crossref\]](#)
 15. Huang XP, Ye TT, Zhang L, et al. Sonographic features of papillary thyroid microcarcinoma predicting high-volume central neck lymph node metastasis. *Surg Oncol*. 2018;27(2):172-176. [\[Crossref\]](#)
 16. Huang W, Chen D, Zhong M, et al. Factors of central lymph node metastasis in papillary thyroid cancer based on C-TIRADS analysis. *Horm Metab Res*. 2023;55(9):585-591. [\[Crossref\]](#). Erratum in: *Horm Metab Res*. 2023;55(9):e5. [\[Crossref\]](#)
 17. Xiao F, Li JM, Han ZY, et al. Multimodality US versus thyroid imaging reporting and data system criteria in recommending fine-needle aspiration of thyroid nodules. *Radiology*. 2023;307(5):e221408. [\[Crossref\]](#)
 18. Chen J, Zhong J, Zhuang Y, et al. Multimodality ultrasound utilizing microvascular flow imaging and shear wave elastography to guide fine-needle aspiration of thyroid lesions: a prospective study validating pattern-based microvascular classification. *Thyroid*. 2025;35(5):516-526. [\[Crossref\]](#)
 19. Zhang H, Liu Q, Han X, Niu L, Sun W. TN5000: an ultrasound image dataset for thyroid nodule detection and classification. *Sci Data*. 2025;12:1437. [\[Crossref\]](#)
 20. Zhou SC, Liu TT, Zhou J, et al. An ultrasound radiomics nomogram for preoperative prediction of central neck lymph node metastasis in papillary thyroid carcinoma. *Front Oncol*. 2020;10:1591. [\[Crossref\]](#)
 21. Tong Y, Zhang J, Wei Y, et al. Ultrasound-based radiomics analysis for preoperative prediction of central and lateral cervical lymph node metastasis in papillary thyroid carcinoma: a multi-institutional study. *BMC Med Imaging*. 2022;22(1):82. [\[Crossref\]](#)
 22. Dupont J, Frau F, Robic V, et al. Beyond handcrafted radiomics in oncologic imaging: innovations in deep, explainable, multi-site and multi-omics radiomics approaches. *Diagn Interv Imaging*. 2026;107(5):165-178. [\[Crossref\]](#)
 23. Moro F, Vagni M, Tran HE, et al. Radiomics analysis of ultrasound images to discriminate between benign and malignant adnexal masses with solid morphology on ultrasound. *Ultrasound Obstet Gynecol*. 2025;65(3):353-363. [\[Crossref\]](#)