ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

Dual-phase contrast-enhanced computed tomography radiomics combined with clinical factors for predicting minimal extrathyroidal extension in papillary thyroid microcarcinoma

Xuehong He[1†], Bo Zhou[2†], Liang Lv[2], Jiang Ma[2], Ya Zou[2], Jie Huang[2*], and Zuhua Song[2*]

[1]Department of Radiology, Mianzhu People's Hospital, Mianzhu, Sichuan, China

[2]Department of Radiology, Chongqing General Hospital, Chongqing University, Chongqing, China

Abstract

Introduction: Minimal extrathyroidal extension (mETE) is an adverse feature of papillary thyroid microcarcinoma (PTMC), but its preoperative identification remains challenging.

Objective: To develop and validate a dual-phase contrast-enhanced computed tomography (CT) radiomics model and evaluate the complementary value of clinical factors for predicting mETE in PTMC.

Methods: This retrospective study included 311 patients with pathologically confirmed PTMC, randomly assigned to training (n = 217) and testing (n = 94) cohorts. Radiomics features were extracted from arterial- and venous-phase contrast-enhanced CT images and reduced using correlation analysis and least absolute shrinkage and selection operator logistic regression. Independent clinical and conventional imaging predictors were identified by multivariable logistic regression. Clinical, radiomics, and combined models were evaluated by receiver operating characteristic curve analysis, calibration, decision curve analysis, net reclassification improvement, and integrated discrimination improvement.

Results: Thirty-eight radiomics features were retained. Systemic immune-inflammation index, neutrophil-to-lymphocyte ratio, and ultrasonographic capsular contact status were independent predictors of mETE. In the training cohort, the clinical, radiomics, and combined models achieved areas under the curve (AUCs) of 0.765, 0.916, and 0.923, respectively; corresponding AUCs in the testing cohort were 0.615, 0.822, and 0.825. Both the radiomics and combined models demonstrated significantly better discrimination than the clinical model, whereas their difference was not statistically significant (p = 0.894).

Conclusion: The dual-phase CT radiomics model showed good discrimination for mETE in PTMC. Clinical factors provided complementary information but did not significantly improve discrimination. The radiomics-based model may serve as an auxiliary tool for preoperative risk assessment. However, prospective multicenter studies and independent external validation are warranted to establish its clinical utility.

Keywords: Papillary thyroid microcarcinoma; Minimal extrathyroidal extension; Radiomics; Computed tomography; Nomogram

†*These authors contributed equally to this work.*

***Corresponding authors:**
Jie Huang
(1021999827@qq.com);
Zuhua Song
(szuh2013@foxmail.com)

Citation: He X, Zhou B, Lv L, *et al*. Dual-phase contrast-enhanced computed tomography radiomics combined with clinical factors for predicting minimal extrathyroidal extension in papillary thyroid microcarcinoma. *Eurasian J Med Oncol*.
doi: 10.36922/EJMO026300370

Received: July 20, 2026

Revised: August 17, 2026

Accepted: September 29, 2026

Published online: November 16, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Thyroid malignancies have shown a marked increase in incidence worldwide, with papillary thyroid carcinoma (PTC) constituting the predominant histological subtype. Papillary thyroid microcarcinoma (PTMC), defined as having a diameter no greater than 10 mm, represents a substantial proportion of newly diagnosed thyroid cancers.[1-3] Extrathyroidal extension (ETE) is a key pathological indicator of tumor aggressiveness and is usually classified, according to the extent of invasion, into minimal ETE (mETE) and gross ETE. mETE refers to microscopic invasion of the thyroid capsule or strap muscles, whereas gross ETE denotes macroscopic extension beyond the thyroid gland, often involving the strap muscles, subcutaneous tissues, or adjacent vital organs, nerves, and vessels.[4] Although mETE is no longer incorporated into the T-stage in the eighth American Joint Committee on Cancer staging system, the updated 2025 guidelines of the American Thyroid Association (ATA) continue to recognize its association with disease progression and recurrence risk.[2] Consistently, recent studies have shown that mETE is significantly related to higher risks of lymph node metastasis, distant metastasis, and tumor recurrence, and has been identified as an independent adverse prognostic factor in PTMC.[4-9] For patients with low-risk PTMC, active surveillance can avoid unnecessary surgery, procedure-related complications, and financial burden, whereas mETE usually necessitates escalation of management, including more intensive follow-up or timely surgical intervention, to achieve better outcomes.[2,10] However, mETE is currently diagnosed mainly on postoperative histopathology, and its early and accurate assessment remains challenging. Therefore, reliable identification of mETE in PTMC is crucial for individualized clinical management, helping to prevent overtreatment in truly low-risk patients while reducing the risk of undertreatment in those with occult aggressive disease.

As the first-line imaging modality for evaluating thyroid nodules, conventional ultrasound (US) has limited diagnostic performance for detecting ETE. A systematic review and meta-analysis reported a pooled sensitivity of 76% for US in identifying ETE, whereas the specificity was only 51%. The high false-positive rate and operator dependence of US compromise diagnostic accuracy, particularly for mETE.[11] Subjective evaluation of tumor invasion into adjacent structures on conventional computed tomography (CT) signs also has clear limitations. Prior comparative evidence suggests that using CT together with US offers only a modest improvement in ETE detection,[12] while the ability of this combined approach to identify mETE remains limited. Fine-needle aspiration is invasive and cannot provide pathological information on mETE. These limitations highlight the need for objective and quantitative imaging approaches to refine risk stratification.

Radiomics enables the computational conversion of medical images into a large number of quantitative features, thereby providing information that may not be captured through routine visual interpretation. By capturing subtle textural and morphological information that reflects intratumoral heterogeneity, the tumor microenvironment, and angiogenic activity, radiomic features can serve as noninvasive surrogate biomarkers of tumor aggressiveness.[13] A recent systematic review and meta-analysis confirmed that, compared with conventional imaging features, radiomics-based models achieve higher overall accuracy in identifying ETE and other aggressive characteristics of thyroid carcinoma.[14] In particular, dual-phase contrast-enhanced CT provides dynamic information on tumor perfusion and hemodynamics, which may mirror the underlying biological invasiveness. Several studies have demonstrated that CT-based radiomics models show significantly better diagnostic performance for predicting ETE than clinical models, and that combining radiomics with clinical factors can further improve discrimination.[15,16] In addition, systemic inflammatory markers and related composite indices are effective indicators for predicting prognosis and tumor progression in cancer patients and have been shown to provide incremental diagnostic value for the preoperative risk stratification of PTC.[17] To predict mETE in PTMC, this study sought to develop and validate a dual-phase contrast-enhanced CT-based radiomics model. This model was further integrated with systemic inflammatory biomarkers and conventional imaging features to construct a nomogram, and its diagnostic performance was comprehensively evaluated.

2. Materials and methods

2.1. Patient selection

The study protocol received approval from the Institutional Ethics Committee of Chongqing General Hospital, and the requirement for written informed consent was waived. We reviewed consecutive patients with PTMC who underwent preoperative CT at our institution between December 2022 and December 2024. Patients were considered eligible based on the following criteria: (i) patients with postoperative histopathological diagnosis of PTC measuring ≤10 mm in maximum diameter, with or without mETE, and (ii) patients whose preoperative CT examination was obtained no more than two weeks before surgery. Patients were excluded for any of the following: (i) those who obtained

prior thyroid-related treatments before CT examination, such as surgery or radiofrequency ablation (n = 26); (ii) those with poor image quality (n = 56); (iii) those with incomplete or missing clinical, laboratory, or imaging data (n = 48); and (4) those with a history of other primary malignancies (n = 5). Finally, 311 patients were enrolled (Figure 1).

2.2. Imaging protocol

An IQon Spectral CT (Philips Healthcare, Netherlands) was used for preoperative neck scanning. The supine patients were scanned from the oropharynx to the superior clavicular margin. Imaging settings included a tube potential of 120 kVp, automatic current modulation, a gantry rotation time of 0.5 seconds, a detector collimation width of 0.625 mm, and a 512 × 512 reconstruction grid. Ioversol (350 mg I/mL) was administered through an antecubital venous access at 1.5 mL/kg at an injection rate of 3–4 mL/s, followed by a 20-mL saline chaser delivered at the same rate. For arterial phase (AP) acquisition, automated bolus tracking was used, with the scan initiated when attenuation in a region of interest placed in the descending aorta at the tracheal bifurcation reached 110 HU. The venous phase (VP) was obtained 30 seconds after the start of the AP.

2.3. Clinicopathologic information and conventional imaging feature assessment

Clinical variables were collected retrospectively from the hospital information system, including sex, age, and peripheral blood counts of neutrophils (N), lymphocytes (L), monocytes (M), and platelets (P). Based on these parameters, the following inflammatory indices were calculated: systemic immune inflammation index (SII) = (N × P)/L, systemic inflammation response index (SIRI) = (N × M)/L, platelet-to-lymphocyte ratio (PLR) = P / L, neutrophil-to-lymphocyte ratio (NLR) = N/L, and lymphocyte-to-monocyte ratio (LMR) = L/M. The diagnosis of mETE was established by postoperative histopathological examination, which served as the reference standard, and was defined as microscopic tumor invasion of the thyroid capsule or any anterior cervical strap muscles.

The conventional imaging features assessed on both US and CT included margin, microcalcifications, the contact status between the nodule and the thyroid capsule on US,[18] and thyroid edge interruption (defined as a partial, often "bitten cookie"-shaped defect of the normal hyperdense thyroid contour, replaced by a hypodense tumor region) on CT.[19]

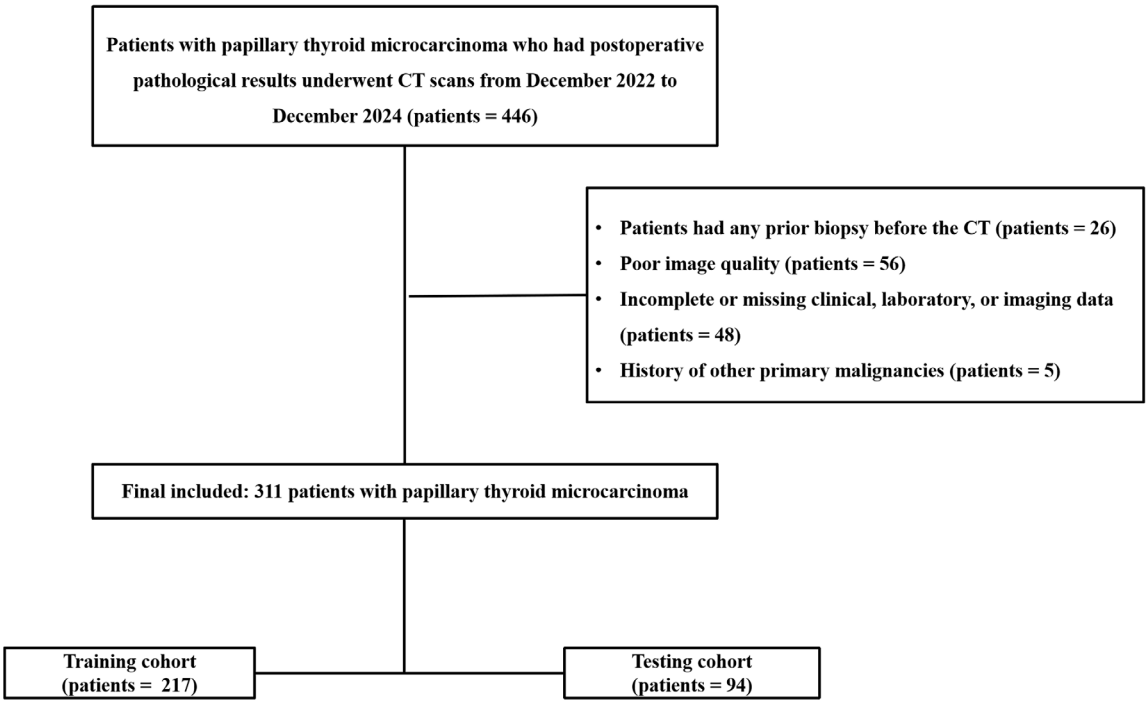


Figure 1. Flowchart of patient selection
Abbreviation: CT: Computed tomography.

2.4. Image preprocessing and tumor segmentation

Prior to radiomics analysis, image preprocessing was performed. Both AP and VP CT images were resampled to an isotropic voxel size of 0.5×0.5×0.5 mm^3 using B-spline interpolation (sitkBSpline for images, and sitkNearestNeighbor for masks to preserve binary integrity). Then, *z*-score normalization was applied using PyRadiomics (V3.0.1) settings (normalize = True, normalizeScale = 1,000) to standardize inter-patient intensity variations. We used a fixed-bin-width discretization approach, as recommended by the Image Biomarker Standardisation Initiative. Specifically, we set binWidth = 5 HU. ITK-SNAP software (version 3.8.0) was employed for three-dimensional tumor segmentation. Radiologist A1 initially delineated the entire tumor volume of interest (VOI) slice by slice on axial images, with simultaneous reference to both AP and VP images. The VOIs were subsequently reviewed and, when necessary, refined in the coronal and sagittal planes to ensure complete tumor coverage. All segmentations were then verified by radiologist A2. Any disagreement regarding the VOI contour was resolved through discussion with radiologist A3 to reach a final consensus. To assess inter-observer reproducibility, a random subset of 30 patients was independently segmented by radiologist B, who was blinded to both the original segmentations and the histopathological results. Features with an intraclass correlation coefficient greater than 0.75 were considered to demonstrate adequate consistency and were retained for further analysis.

2.5. Radiomics feature extraction and selection

Quantitative image features were derived using the PyRadiomics platform. First, we extracted features from multiple image filters, including the original image, Laplacian of Gaussian with sigma values of 1.0, 2.0, and 3.0, and Wavelet-transformed images. Radiomics features were separately extracted from the tumor VOIs on AP and VP images, including first-order statistical measures, along with texture features based on the gray-level co-occurrence matrix (GLCM), gray-level dependence matrix (GLDM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighboring gray-tone difference matrix (NGTDM), and shape-based descriptors. The AP and VP feature sets were then combined to construct a joint feature set integrating dual-phase information, yielding a composite feature set of 3,668 features for each lesion (880 GLCM, 720 first-order, 640 GLRLM, 640 GLSZM, 560 GLDM, 200 NGTDM, and 28 shape features).

Prior to dimensionality reduction, all radiomics features extracted from dual-phase CT images were standardized via *z*-score normalization. Subsequent univariate analysis using Student's *t*-test or Mann–Whitney *U* test further screened these candidates, retaining 1143 features with significant differences between mETE-positive and mETE-negative patients. Pairwise Spearman correlation analysis ($|r| > 0.9$) was first performed to eliminate highly redundant features, leaving 389 non-redundant features. Finally, least absolute shrinkage and selection operator (LASSO) logistic regression with 10-fold cross-validation was applied to the filtered features in the training cohort (Figure 2A). At the optimal regularization parameter λ (λ = 0.0256), 38 features with non-zero coefficients were retained to establish the radiomics signature. The rad-score was generated as a linear combination of these selected features weighted by their corresponding LASSO coefficients (Figure 2B), and was subsequently adopted for model construction and performance assessment. The detailed rad-score formula is provided in Appendix 1.

2.6. Model construction and evaluation

The radiomics classifier was derived from the selected dual-phase features and represented by the corresponding rad-score. Candidate clinical and conventional imaging variables were examined in the training cohort using univariable followed by multivariable logistic regression, and the retained predictors were used to establish the clinical model. For the combined model, the rad-score was entered together with these independent predictors, and the resulting regression equation was visualized as a nomogram.

Both the training and testing cohorts were utilized to evaluate model performance. Receiver operating characteristic (ROC) analysis, area under the curve (AUC), sensitivity, and specificity were used to quantify discrimination, whereas calibration plots were used to compare predicted probabilities with observed outcomes. The potential clinical usefulness of each model was examined by decision curve analysis (DCA). Between-model AUC differences were evaluated with the DeLong test. In addition, the net reclassification improvement (NRI) was calculated to assess improvements in risk reclassification, and the integrated discrimination improvement (IDI) and its statistical significance were computed to quantify the gain in overall discrimination.

2.7. Statistical analysis

We conducted all statistical analyses using Python (version 3.7.12). Continuous variables were expressed as mean ± standard deviation (SD) or median (interquartile range), while categorical variables were presented as frequencies (percentages). Univariable and multivariable logistic regression models were employed to identify independent

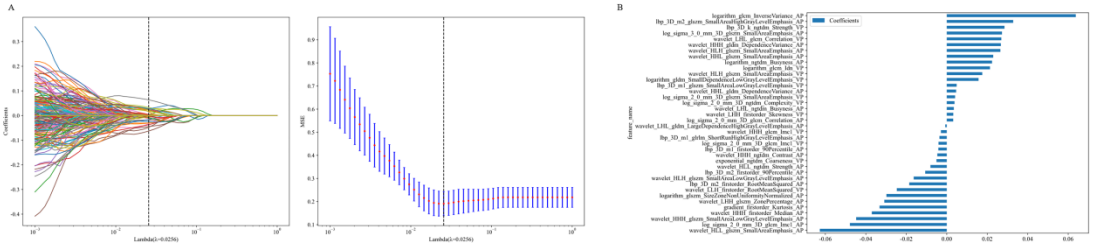


Figure 2. Radiomics feature selection using LASSO logistic regression. (A) Ten-fold cross-validation was performed to select the optimal regularization parameter **λ**. (B) LASSO coefficient profiles of the retained radiomics features.
Abbreviation: LASSO: Least absolute shrinkage and selection operator.

predictors, and the optimal threshold was determined via the Youden index. All statistical tests were two-sided, and *p*-values < 0.05 were considered statistically significant.

3. Results

3.1. Patient characteristics

Patient demographics and conventional imaging characteristics for the two cohorts are presented in Table 1. A total of 311 patients with PTMC were enrolled, including 208 patients with mETE and 103 without mETE. Using a 7:3 random split, 217 patients were allocated to the training cohort and 94 to the testing cohort. The proportions of patients with mETE were comparable between the training and testing cohorts (68.7% [149/217] vs. 62.8% [59/94]).

3.2. Clinical model development

Regression results obtained from the training cohort are summarized in Table 2. In the univariable analysis, all clinical and conventional imaging variables except sex were significantly associated with mETE (all $p < 0.05$). Among these variables, thyroid edge interruption on CT demonstrated the strongest association with mETE (odds ratio [OR] = 4.350, 95% confidence interval [CI]: 2.892–6.540), followed by the capsular contact status (OR = 3.847, 95% CI: 2.779–5.328). In the multivariable logistic regression analysis, SII (OR = 0.996, 95% CI: 0.993–0.999), NLR (OR = 1.854, 95% CI: 1.131–3.040), and capsular contact status (OR = 4.575, 95% CI: 2.638–7.933) remained predictors of mETE.

3.3. Development of the radiomics model and the combined model

Based on the dual-phase contrast-enhanced CT images, 3,668 radiomics features were initially extracted for each lesion and subsequently reduced through correlation analysis and LASSO logistic regression, ultimately yielding a 38-feature radiomics signature. These comprised 7 GLCM, 12 GLSZM, 1 GLRLM, 4 GLDM, 7 NGTDM, and 7 first-order features, while no shape features were retained. Of the 38 features, 25 were derived from the AP and 13 from the VP. This 38-feature signature was used to construct the radiomics model, which was subsequently combined with the independent clinical and imaging predictors to build the integrated combined model.

3.4. Performance of the predictive models

Model performance is detailed in Table 3 and Figure 3. The training set AUCs were 0.765 for the clinical model, 0.916 for the radiomics model, and 0.923 for the combined model. In the testing cohort, the corresponding AUCs were 0.615, 0.822, and 0.825, respectively. The optimal cutoff value for the combined model was 0.604. The combined model achieved an accuracy of 0.849 in the testing cohort, with a sensitivity of 0.966 and a negative predictive value of 0.920. Overall, both the radiomics and combined models demonstrated superior predictive performance compared with the clinical model. The DeLong test revealed that the combined model significantly outperformed the clinical model in both the training and testing cohorts ($p < 0.001$ for both). No significant difference was observed between the AUCs of the combined model and the radiomics model in either cohort ($p = 0.404$ and $p = 0.894$ for training and testing, respectively). The radiomics model also demonstrated significantly superior predictive performance compared with the clinical model in both cohorts ($p < 0.001$ and $p = 0.002$, respectively).

3.5. Development and performance of the nomogram

The combined model, which incorporated the SII, NLR, capsular contact status, and radiomics signatures from the AP and VP, was visualized as a nomogram (Figure 4). The nomogram showed robust discrimination and consistently surpassed the clinical model in predicting outcomes in both cohorts. Calibration plots (Figure 5) revealed close alignment between the predicted and actual probabilities. DCA (Figure 6) indicated that the combined model yielded a higher net benefit across clinically relevant threshold

Table 1. Baseline clinical and conventional imaging characteristics of patients in the training and testing cohorts

Characteristics	Training	Testing	*p*-value
Age (years), mean ± SD	39.77±10.36	38.95±10.62	0.406
NLR, median (IQR)	2.08 (1.63, 2.67)	1.95 (1.57, 2.64)	0.354
SII, median (IQR)	485.85 (337.74, 686.20)	459.31 (335.69, 678.05)	0.535
SIRI, median (IQR)	0.57 (0.37, 0.86)	0.58 (0.39, 0.90)	0.739
PLR, median (IQR)	138.78 (107.56, 178.16)	141.41 (112.50, 174.35)	0.949
LMR, median (IQR)	5.55 (4.15, 7.29)	5.49 (4.19, 6.92)	0.642
Sex			0.873
Female	198 (91.24)	87 (92.55)	
Male	19 (8.76)	7 (7.45)	
Thyroid edge interruption			>0.99
No	110 (50.69)	48 (51.06)	
Yes	107 (49.31)	46 (48.94)	
Margin			0.75
No	2 (0.92)	2 (2.13)	
Yes	215 (99.08)	92 (97.87)	
Microcalcifications			0.46
No	49 (22.58)	17 (18.09)	
Yes	168 (77.42)	77 (81.91)	
Capsular contact status			0.586
No	72 (33.18)	32 (34.04)	
Abutment	121 (55.76)	48 (51.06)	
Bulging	24 (11.06)	14 (14.89)	

Note: Data presented as *n*(%), unless stated otherwise.
Abbreviations: IQR: Interquartile range; LMR: Lymphocyte-to-monocyte ratio; NLR: Neutrophil-to-lymphocyte ratio; PLR: Platelet-to-lymphocyte ratio; SD: Standard deviation; SII: Systemic immune-inflammation index; SIRI: Systemic inflammation response index.

probabilities. The NRI (Figure 7) and IDI (Figure 8) further demonstrated that the radiomics model and the combined model exhibited significantly superior classification accuracy and discriminative ability compared with the clinical model.

4. Discussion

Papillary thyroid microcarcinoma is generally considered an indolent malignancy, but the presence of mETE has important implications for risk stratification and therapeutic decision-making. According to the most recent risk-stratification systems of the American Thyroid Association, mETE is associated with an increased risk of disease progression and recurrence in PTMC patients.[2] In addition, several studies have demonstrated that mETE is an independent adverse prognostic factor for lymph node metastasis, distant metastasis, and tumor recurrence in PTMC.[4-9,20] However, mETE is difficult to assess accurately on preoperative imaging and cannot be reliably identified intraoperatively, making early, precise, and noninvasive prediction crucial for tailoring treatment strategies. Therefore, in this study, we developed and validated prediction models that integrate dual-phase contrast-enhanced CT radiomics features with conventional imaging characteristics and clinical parameters to evaluate the presence of mETE in PTMC.

Table 2. Univariable and multivariable logistic regression analyses for predicting minimal extrathyroidal extension in the training cohort

Variables	Univariable logistic regression			Multivariable logistic regression		
	OR	95% CI	*p*-value	OR	95% CI	*p*-value
SII	1.001	1.001–1.001	<0.001	0.996	0.993–0.999	0.01
PLR	1.004	1.003–1.006	<0.001	1.009	1.001–1.018	0.073
Age	1.018	1.012–1.024	<0.001	0.971	0.946–0.997	0.065
LMR	1.135	1.092–1.181	<0.001	1.101	0.921–1.317	0.375
NLR	1.285	1.168–1.415	<0.001	1.854	1.131–3.04	0.04
Sex	1.714	0.784–3.747	0.257			
SIRI	1.835	1.373–2.452	<0.001	1.422	0.517–3.699	0.544
Margin	2.209	1.733–2.815	<0.001	0.529	0.083–3.35	0570
Microcalcifications	2.360	1.788–3.114	<0.001	1.044	0.541–2.016	0.915
Capsular contact status	3.847	2.779–5.328	<0.001	4.575	2.638–7.933	<0.001
Thyroid edge interruption	4.350	2.892–6.540	<0.001	1.628	0.878–3.016	0.194

Abbreviations: CI: Confidence interval; LMR: Lymphocyte-to-monocyte ratio; NLR: Neutrophil-to-lymphocyte ratio; OR: Odds ratio; PLR: Platelet-to-lymphocyte ratio; SII: Systemic immune-inflammation index; SIRI: Systemic inflammation response index.

Table 3. Predictive performance of the clinical, radiomics, and combined models in the training and testing cohorts

Model	Cohort	Accuracy (95% CI)	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Clinical	Training	0.733 (0.671–0.794)	0.765 (0.695–0.835)	0.732 (0.658–0.805)	0.735 (0.619–0.851)	0.858 (0.801–0.907)	0.556 (0.486–0.625)
Radiomics	Training	0.885 (0.836–0.925)	0.916 (0.875–0.957)	0.946 (0.897–0.975)	0.750 (0.632–0.845)	0.892 (0.832–0.935)	0.864 (0.752–0.934)
Combined	Training	0.885 (0.840–0.929)	0.923 (0.883–0.964)	0.926 (0.883–0.969)	0.791 (0.688–0.899)	0.908 (0.860–0.955)	0.831 (0.731–0.930)
Clinical	Testing	0.634 (0.536–0.733)	0.615 (0.498–0.731)	0.707 (0.589–0.824)	0.514 (0.346–0.682)	0.707 (0.614–0.799)	0.514 (0.412–0.616)
Radiomics	Testing	0.828 (0.735–0.898)	0.822 (0.730–0.915)	0.966 (0.881–0.996)	0.600 (0.435–0.748)	0.800 (0.696–0.878)	0.913 (0.733–0.986)
Combined	Testing	0.849 (0.776–0.923)	0.825 (0.729–0.921)	0.966 (0.918–0.999)	0.657 (0.498–0.817)	0.824 (0.732–0.914)	0.920 (0.812–0.999)

Abbreviations: AUC: Area under the curve; CI: Confidence interval; NPV: Negative predictive value; PPV: Positive predictive value.

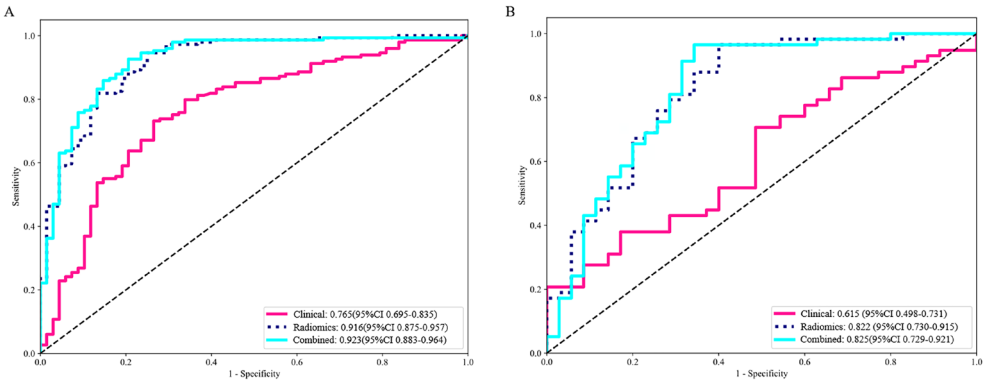


Figure 3. Receiver operating characteristic (ROC) curves for the clinical, radiomics, and combined models in the training (A) and testing (B) cohorts.

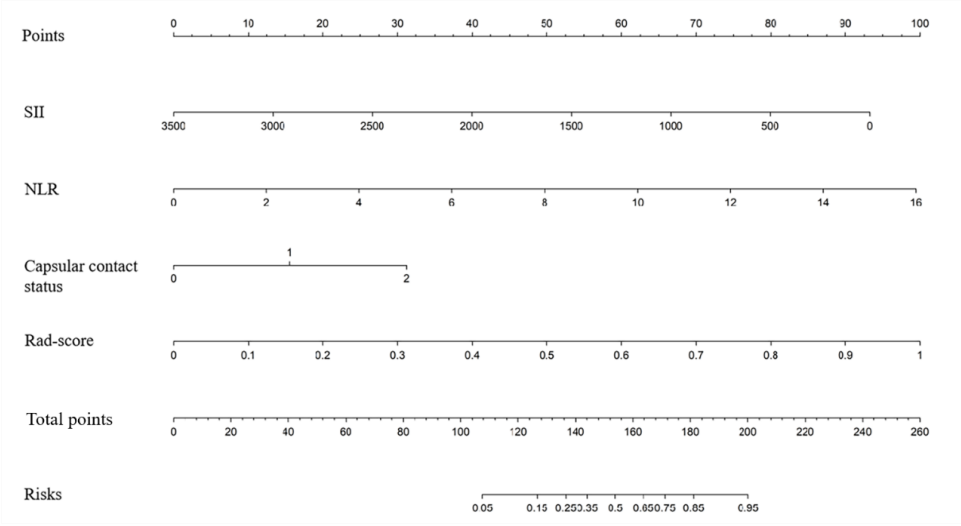


Figure 4. The nomogram for predicting the probability of minimal extrathyroidal extension in patients with papillary thyroid microcarcinoma. The nomogram integrates the rad-score with independent clinical and conventional imaging predictors, including SII, NLR, and capsular contact status on ultrasound. For capsular contact status, the three categories were coded as 0 = no contact, 1 = abutment, and 2 = bulging.
Abbreviations: NLR: Neutrophil-to-lymphocyte ratio; SII: Systemic immune-inflammation index.

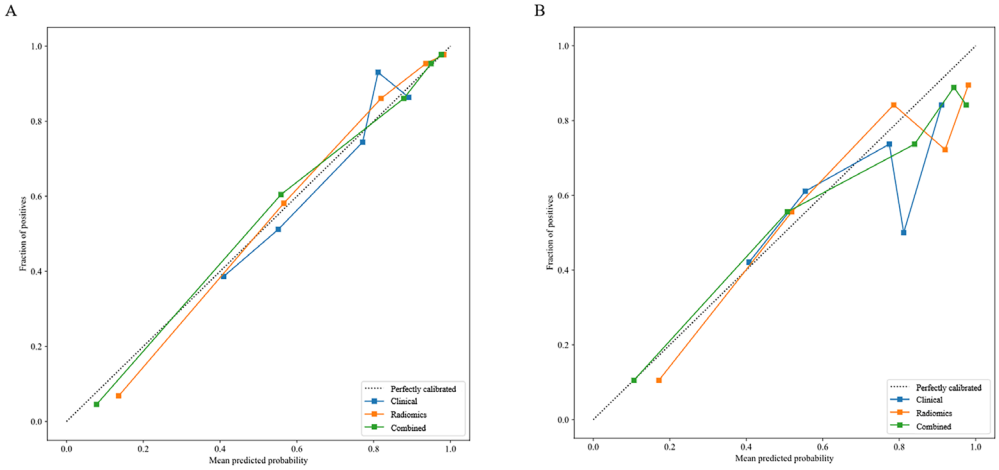


Figure 5. Calibration curves of the combined model in the training (A) and testing (B) cohorts

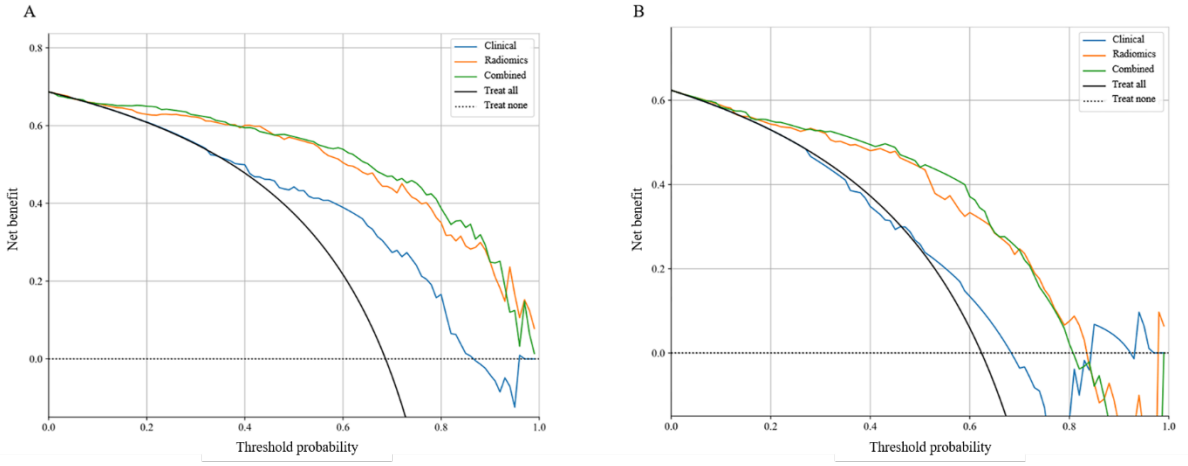


Figure 6. Decision curve analysis of the clinical model, radiomics model, and combined model in the training (A) and testing (B) cohorts

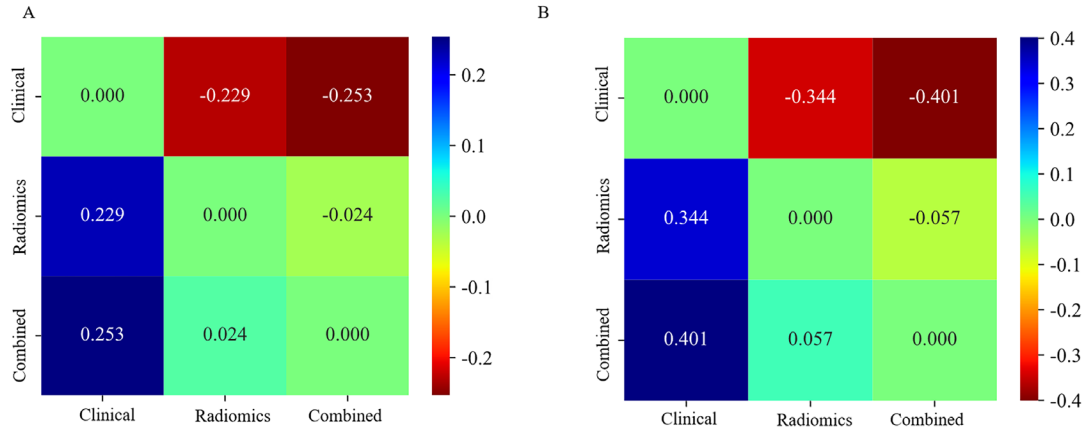


Figure 7. Net reclassification improvement (NRI) analysis comparing the predictive models in the training (A) and testing (B) cohorts

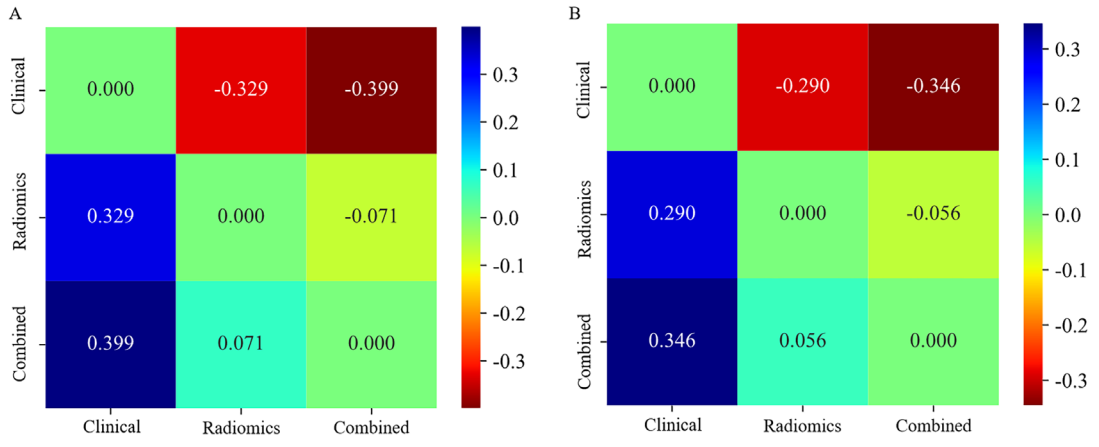


Figure 8. Integrated discrimination improvement (IDI) analysis comparing the predictive models in the training (A) and testing (B) cohorts

Recent evidence has consistently demonstrated that radiomics-based models outperform conventional imaging assessment in predicting ETE in PTC.[14] Unlike conventional imaging, which primarily relies on subjective morphological assessment, radiomics quantitatively characterizes intratumoral heterogeneity by extracting high-dimensional features that are imperceptible to the human eye. Increasing evidence suggests that intratumoral heterogeneity reflects complex biological alterations, including variations in cellular density, angiogenesis, and extracellular matrix remodeling, all of which are closely associated with tumor aggressiveness.[21] Therefore, before gross morphological changes become visually apparent, radiomics features may serve as imaging biomarkers capable of identifying subtle biological differences between tumors with and without mETE.

Lu *et al.*[22] developed a three-dimensional ultrasound-based radiomics nomogram for predicting ETE in PTC, achieving an AUC of 0.798 in the testing cohort. Chen *et al.*[15] constructed a CT-based radiomics model using non-contrast and VP CT images, which achieved an AUC of 0.772 in the testing cohort. Collectively, these studies support the ability of radiomics to characterize the invasive phenotype of PTC. In comparison, the dual-phase contrast-enhanced CT radiomics model developed in our study achieved a higher predictive performance for mETE (testing cohort AUC = 0.822) and significantly outperformed the clinical model (AUC = 0.822 vs. 0.615, $p = 0.002$). One possible explanation is that our model incorporated complementary quantitative information from dual-phase CT imaging, enabling a more comprehensive characterization of tumor perfusion, vascular heterogeneity, and tissue composition than single-phase imaging. Texture-related features accounted for 31 of the 38 selected features, including GLCM, GLSZM, GLRLM, GLDM, and NGTDM, whereas no shape features were retained. This finding is consistent with previous studies, suggesting that microscopic intratumoral heterogeneity, rather than gross tumor morphology, is more informative for identifying mETE. Texture features quantitatively characterize the spatial distribution and heterogeneity of gray-level intensities within tumors. For example, GLCM reflects local texture patterns, GLRLM characterizes the continuity of fine textures, and GLSZM represents larger homogeneous regions.[23,24] These quantitative descriptors provide complementary information regarding tumor microarchitecture that cannot be appreciated by visual assessment alone. Furthermore, 25 of the retained features were derived from the AP and 13 from the VP. AP images primarily reflect tumor perfusion and microvascular density, whereas VP images better depict contrast redistribution within the extracellular space. The complementary hemodynamic information provided by the two enhancement phases therefore enables a more comprehensive assessment of tumor biological behavior than either phase alone.[25]

Our study found that SII, NLR, and capsular contact status were predictors of mETE in patients with PTMC. SII reflects the dynamic interaction between the host immune response and the tumor microenvironment and has been associated with tumor proliferation, angiogenesis, immune evasion, and metastasis in various malignancies.[26,27] Interestingly, our multivariable analysis demonstrated an inverse association between SII and mETE (OR = 0.996, 95% CI: 0.993–0.999). To evaluate whether this paradoxical finding was attributable to collinearity, we performed Spearman correlation analysis and calculated variance inflation factors (**Appendix 2**). The results revealed a strong correlation between SII and NLR ($r = 0.860$) and elevated variance inflation factors for both SII (6.47) and NLR (5.38), indicating notable collinearity that may partially account for the coefficient reversal. Given the algebraic interdependence and collinearity among these inflammatory indices, the inverse association between SII and mETE should be interpreted with caution, and these inflammatory markers should be viewed collectively as a composite signature contributing to the predictive model. Further validation in independent cohorts is warranted. Elevated NLR reflects enhanced neutrophil-mediated inflammation together with impaired lymphocyte-mediated antitumor immunity. Neutrophils may promote tumor invasion through the release of neutrophil extracellular traps, angiogenic factors, matrix metalloproteinases, and inflammatory cytokines, thereby facilitating extracellular matrix remodeling and tumor progression.[17,28] These mechanisms may partially explain the association between elevated NLR and mETE observed in our study. Capsular contact status was also independently associated with mETE, consistent with the increased likelihood of ETE when tumors progressively abut or breach the thyroid capsule.[12] Both the radiomics and combined models outperformed the clinical model, yet integration of clinical variables yielded no statistically significant performance gains. Despite the higher sensitivity and negative predictive value of the combined model, these metrics are heavily influenced by disease prevalence. Collectively, the radiomics model demonstrated good discriminative ability for predicting mETE in patients with PTMC, while clinical factors may provide complementary information for individualized risk assessment.

To comprehensively evaluate model performance, we assessed discrimination, calibration, clinical utility, and risk reclassification using ROC curve analysis, calibration

curves, DCA, NRI, and IDI. In the testing cohort, the radiomics and combined models both exhibited significantly superior discriminative ability relative to the clinical model. Although no significant AUC difference was detected between the two models (p = 0.894), the combined model achieved higher sensitivity and negative predictive value, approaching 0.966 and 0.920, respectively. For patients with PTMC, missed mETE corresponding to false-negative outcomes can lead to undertreatment and an increased risk of disease progression. Thus, high sensitivity and negative predictive value are clinically meaningful for ruling out mETE. However, we acknowledge that this advantage is offset by a mild decline in specificity (0.657 vs. 0.600), resulting in restricted overall benefit. These findings suggest that integrating radiomics with clinically relevant predictors may further improve individualized risk assessment beyond discrimination alone. The nomogram developed in this study enables intuitive and effective prediction of the risk of mETE in patients with PTMC, which may facilitate clinical decision-making and personalized management.

This study has several limitations. First, the retrospective single-center design, together with the surgical composition of our cohort, may introduce selection bias and limit direct generalizability to broader PTMC populations—particularly those under active surveillance, given the relatively higher mETE prevalence in our surgical cohort. Second, the marked performance drop of the clinical model between cohorts, along with the high feature-to-sample ratio and the absence of nested cross-validation, suggests that our results, while promising, should be regarded as preliminary and that some degree of model optimism cannot be excluded. Additionally, the incremental value of our radiomics-based approach over routine radiologist assessment remains to be established in the absence of a dedicated reader study or external validation. Future prospective multicenter studies with larger and more representative patient populations, independent external cohorts, standardized imaging protocols, test-retest robustness checks, systematic perturbation assessments, and blinded radiologist evaluations are warranted to confirm the generalizability and refine the clinical utility of our combined model.

5. Conclusion

The dual-phase contrast-enhanced CT radiomics model demonstrated good internal discrimination for predicting mETE in PTMC. Adding clinical factors provided supplementary information without significant improvement in predictive performance. This model may serve as an adjunct tool, yet prospective multicenter studies and independent external validation are needed before clinical use.

Acknowledgments

None.

Funding

The study was funded by the Medical Research Program of the Combination of Chongqing National Health Commission and Chongqing Science and Technology Bureau, China (2024QNXM058).

Conflicts of interest

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

Author contributions

Conceptualization: Zuhua Song
Data curation: Bo Zhou, Jie Huang
Formal analysis: Liang Lv, Jiang Ma, Ya Zou
Investigation: Liang Lv
Methodology: Liang Lv, Zuhua Song
Software: Liang Lv
Supervision: Jie Huang, Zuhua Song
Validation: Jiang-Ma, Ya Zou
Writing-original draft: Xuehong He, Jie Huang
Writing-review & editing: Xuehong He, Zuhua Song

Ethics approval and consent to participate

Ethical approval for this study was granted by the institutional review board at our hospital (No. KY S2023-083-01), which waived the requirement for informed consent.

Consent for publication

Not applicable.

Availability of data

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

References

1. Bray F, Laversanne M, Sung H, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263.

   doi: 10.3322/caac.21834

2. Ringel MD, Sosa JA, Baloch Z, *et al*. 2025 American Thyroid

Association management guidelines for adult patients with differentiated thyroid cancer. *Thyroid*. 2025;35(8):841-985.

doi: 10.1177/10507256251363120

3. Wang Z, Ji X, Zhang H, Sun W. Clinical and molecular features of progressive papillary thyroid microcarcinoma. *Int J Surg*. 2024;110(4):2313-2322.

doi: 10.1097/JS9.0000000000001117

4. Zhan H, Hong Y, Zhang L, Huang K, Zheng M, Zhang F. Impact of location and size of minimal extrathyroidal extension on lymph node metastasis in papillary thyroid cancer: a retrospective analysis. *Gland Surg*. 2024;13(9):1619-1627.

doi: 10.21037/gs-24-273

5. Seifert R, Schäfers M, Heitplatz B, Kerschke L, Riemann B, Noto B. Minimal extrathyroidal extension in papillary thyroid microcarcinoma is an independent risk factor for relapse through lymph node and distant metastases. *J Nucl Med*. 2021;62(12):1702-1709.

doi: 10.2967/jnumed.121.261898

6. Marongiu A, Nuvoli S, De Vito A, Vargiu S, Spanu A, Madeddu G. Minimal extrathyroid extension (mETE) as the only risk factor in patients with papillary thyroid carcinoma (PC): its clinical impact on recurrence and outcome during long-term follow-up. *Biomedicines*. 2024;12(2):350.

doi: 10.3390/biomedicines12020350

7. Guo H, Yang H, Yang Z, Zhu M. The impact of minimal extrathyroidal extension on survival outcomes in patients with pT1 and pT2 differentiated thyroid cancer. *Surgery*. 2026;190:109820.

doi: 10.1016/j.surg.2025.109820

8. Song RY, Kim HS, Kang KH. Minimal extrathyroidal extension is associated with lymph node metastasis in single papillary thyroid microcarcinoma: a retrospective analysis of 814 patients. *World J Surg Oncol*. 2022;20(1).

doi: 10.1186/s12957-022-02629-8

9. Zuhur SS, Aggul H, Avci U, *et al*. The impact of microscopic extrathyroidal extension on the clinical outcome of classic subtype papillary thyroid microcarcinoma: a multicenter study. *Endocrine*. 2023;83(3):700-707.

doi: 10.1007/s12020-023-03533-1

10. Li G, Li R, Zhong J, *et al*. A multicenter cohort study of thyroidectomy-related decision regret in patients with low-risk papillary thyroid microcarcinoma. *Nat Commun*. 2025;16(1).

doi: 10.1038/s41467-025-57627-7

11. Issa PP, Albuck AL, Hossam E, *et al*. The diagnostic performance of ultrasonography in the evaluation of extrathyroidal extension in papillary thyroid carcinoma: a systematic review and meta-analysis. *Int J Mol Sci*. 2022;24(1):371.

doi: 10.3390/ijms24010371

12. Lee DY, Kwon TK, Sung MW, Kim KH, Hah JH. Prediction of extrathyroidal extension using ultrasonography and computed tomography. *Int J Endocrinol*. 2014;2014:351058.

doi: 10.1155/2014/351058

13. Limkin EJ, Sun R, Dercle L, *et al*. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. 2017;28(6):1191-1206.

doi: 10.1093/annonc/mdx034

14. Liu Y, Xiang L, Liu FY, *et al*. Accuracy of radiomics in the identification of extrathyroidal extension and BRAFV600E mutations in papillary thyroid carcinoma: a systematic review and meta-analysis. *Acad Radiol*. 2025;32(3):1385-1397.

doi: 10.1016/j.acra.2024.11.014

15. Chen B, Zhong L, Dong D, *et al*. Computed tomography radiomic nomogram for preoperative prediction of extrathyroidal extension in papillary thyroid carcinoma. *Front Oncol*. 2019;9:829.

doi: 10.3389/fonc.2019.00829

16. Yu P, Wu X, Li J, *et al*. Extrathyroidal extension prediction of papillary thyroid cancer with computed tomography based radiomics nomogram: a multicenter study. *Front Endocrinol (Lausanne)*. 2022;13:874396.

doi: 10.3389/fendo.2022.874396

17. Feng JW, Zhang L, Yang YX, *et al*. Development and validation of the multidimensional machine learning model for preoperative risk stratification in papillary thyroid carcinoma: a multicenter, retrospective cohort study. *Cancer Imaging*. 2025;25(1).

doi: 10.1186/s40644-025-00921-w

18. Qi Q, Huang X, Zhang Y, *et al*. Ultrasound image-based deep learning to assist in diagnosing gross extrathyroidal extension thyroid cancer: a retrospective multicenter study. *eClinicalMedicine*. 2023;58:101905.

doi: 10.1016/j.eclinm.2023.101905

19. Han ZJ, Shu YY, Lai XF, Chen WH. Value of computed tomography in determining the nature of papillary thyroid microcarcinomas: evaluation of the computed tomographic characteristics. *Clin Imaging*. 2013;37(4):664-668.

doi: 10.1016/j.clinimag.2012.12.005

20. Parvathareddy SK, Siraj AK, Qadri Z, *et al*. Microscopic extrathyroidal extension results in increased rate of tumor recurrence and is an independent predictor of patient's outcome in Middle Eastern papillary thyroid carcinoma. *Front Oncol*. 2021;11:724432.

doi: 10.3389/fonc.2021.724432

21. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol*. 2021;19(2):132-146.

doi: 10.1038/s41571-021-00560-7

22. Lu WJ, Mao L, Li J, *et al*. Three-dimensional ultrasound-based radiomics nomogram for the prediction of extrathyroidal extension features in papillary thyroid cancer. *Front Oncol*. 2023;13:1046951.

doi: 10.3389/fonc.2023.1046951

23. Aerts HJWL, Velazquez ER, Leijenaar RTH, *et al*. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5(1):4006.

doi: 10.1038/ncomms5006

24. Zhou TH, Zhou XX, *et al*. CT whole lung radiomic nomogram: a potential biomarker for lung function evaluation and identification of COPD. *Mil Med Res*. 2024;11(1).

doi: 10.1186/s40779-024-00516-9

25. Zeng D, Song Z, Liu Q, Huang J, Wang X, Tang Z. Radiomics analysis of dual-layer detector spectral CT-derived iodine maps for predicting Ki-67 PI in pancreatic ductal adenocarcinoma. *BMC Med Imaging*. 2025;25(1).

doi: 10.1186/s12880-025-01664-7

26. Tian BW, Yang YF, Yang CC, *et al*. Systemic immune-inflammation index predicts prognosis of cancer immunotherapy: systemic review and meta-analysis. *Immunotherapy*. 2022;14(18):1481-1496.

doi: 10.2217/imt-2022-0133

27. Kou J, Huang J, Li J, Wu Z, Ni L. Systemic immune-inflammation index predicts prognosis and responsiveness to immunotherapy in cancer patients: a systematic review and meta-analysis. *Clin Exp Med*. 2023;23(7):3895-3905.

doi: 10.1007/s10238-023-01035-y

28. Modestino L, Cristinziano L, Poto R, *et al*. Neutrophil extracellular traps and neutrophil-related mediators in human thyroid cancer. *Front Immunol*. 2023;14:1167404.

doi: 10.3389/fimmu.2023.1167404

Appendix

Appendix 1

The rad-score was calculated as a linear combination of the 38 selected features weighted by their corresponding LASSO coefficients. The complete formula is as follows:

Rad-score = 0.6866359447004563 - 0.033204 * gradient_firstorder_Kurtosis_AP −0.004163 * lbp_3D_m1_firstorder_90Percentile_AP − 0.003615 * lbp_3D_m1_glrlm_ShortRunHighGrayLevelEmphasis_AP − 0.010648 * lbp_3D_m2_firstorder_90Percentile_AP − 0.018542 * lbp_3D_m2_firstorder_RootMeanSquared_AP + 0.032809 * lbp_3D_m2_glszm_SmallAreaHighGrayLevelEmphasis_AP + 0.003183 * log_sigma_2_0_mm_3D_glcm_Correlation_AP − 0.047863 * log_sigma_2_0_mm_3D_glcm_Imc1_AP + 0.027288 * log_sigma_3_0_mm_3D_glszm_SmallAreaEmphasis_AP + 0.063739 * logarithm_glcm_InverseVariance_AP − 0.029819 * logarithm_glszm_SizeZoneNonUniformityNormalized_AP + 0.022358 * logarithm_ngtdm_Busyness_AP − 0.037026 * wavelet_HHH_firstorder_Median_AP + 0.026712 * wavelet_HHH_gldm_DependenceVariance_AP − 0.044771 * wavelet_HHH_glszm_SmallAreaLowGrayLevelEmphasis_AP − 0.004658 * wavelet_HHH_ngtdm_Contrast_AP + 0.004619 * wavelet_HHL_gldm_DependenceVariance_AP + 0.022943 * wavelet_HHL_glszm_SmallAreaEmphasis_AP + 0.026486 * wavelet_HLH_glszm_SmallAreaEmphasis_AP − 0.016336 * wavelet_HLH_glszm_SmallAreaLowGrayLevelEmphasis_AP − 0.062734 * wavelet_HLL_glszm_SmallAreaEmphasis_AP − 0.008059 * wavelet_HLL_ngtdm_Strength_AP − 0.030853 * wavelet_LHH_glszm_ZonePercentage_AP − 0.000738 * wavelet_LHL_gldm_LargeDependenceHighGrayLevelEmphasis_AP + 0.003485 * wavelet_LHL_ngtdm_Busyness_AP − 0.005080 * exponential_ngtdm_Coarseness_VP + 0.028521 * lbp_3D_k_ngtdm_Strength_VP + 0.004884 * lbp_3D_m1_glszm_SmallAreaLowGrayLevelEmphasis_VP − 0.004097 * log_sigma_2_0_mm_3D_glcm_Imc1_VP + 0.004100 * log_sigma_2_0_mm_3D_glszm_SmallAreaEmphasis_VP + 0.003825 * log_sigma_2_0_mm_3D_ngtdm_Complexity_VP + 0.021353 * logarithm_glcm_Idn_VP + 0.015696 * logarithm_gldm_SmallDependenceLowGrayLevelEmphasis_VP − 0.002877 * wavelet_HHH_glcm_Imc1_VP + 0.017543 * wavelet_HLH_glszm_SmallAreaEmphasis_VP + 0.003220 * wavelet_LHH_firstorder_Skewness_VP + 0.026977 * wavelet_LHL_glcm_Correlation_VP − 0.024693 * wavelet_LLH_firstorder_RootMeanSquared_VP

Appendix 2

Spearman correlation coefficients and VIF among inflammatory biomarkers

Variable	NLR	SII	SIRI	PLR	LMR	VIF
NLR	1.000	0.860	0.753	0.625	−0.552	5.38
SII	0.860	1.000	0.763	0.740	−0.513	6.47
SIRI	0.753	0.763	1.000	0.415	−0.612	4.27
PLR	0.625	0.740	0.415	1.000	−0.320	3.52
LMR	−0.552	−0.513	−0.612	−0.320	1.000	1.96

Abbreviations: CI: Confidence interval; LMR: Lymphocyte-to-monocyte ratio; NLR: Neutrophil-to-lymphocyte ratio; OR: Odds ratio; PLR: Platelet-to-lymphocyte ratio; SII: Systemic immune-inflammation index; SIRI: Systemic inflammation response index; VIF: Variance inflation factor.