BMC Medical Imaging

RESEARCH

Open Access

High-risk habitat radiomics model based on ultrasound images for predicting lateral neck lymph node metastasis in differentiated thyroid cancer

Han Liu[1†], Chun-Jie Hou[1,2,3†], Min Wei[1], Ke-Feng Lu[1], Ying Liu[1], Pei Du[1], Li-Tao Sun[1*] and Jing-Lan Tang[1,2,3*]

Abstract

Background This study aims to evaluate the predictive usefulness of a habitat radiomics model based on ultrasound images for anticipating lateral neck lymph node metastasis (LLNM) in differentiated thyroid cancer (DTC), and for pinpointing high-risk habitat regions and significant radiomics traits.

Methods A group of 214 patients diagnosed with differentiated thyroid carcinoma (DTC) between August 2021 and August 2023 were included, consisting of 107 patients with confirmed postoperative lateral lymph node metastasis (LLNM) and 107 patients without metastasis or lateral cervical lymph node involvement. An additional cohort of 43 patients was recruited to serve as an independent external testing group for this study. Patients were randomly divided into training and internal testing group at an 8:2 ratio. Region of interest (ROI) was manually outlined, and habitat analysis subregions were defined using the K-means method. The ideal number of subregions ($n=5$) was determined using the Calinski-Harabasz score, leading to the creation of a habitat radiomics model with 5 subregions and the identification of the high-risk habitat model. Area under the curve (AUC) values were calculated for all models to assess their validity, and predictive model nomograms were created by integrating clinical features. The internal and external testing dataset is employed to assess the predictive performance and stability of the model.

Results In internal testing group, Habitat 3 was identified as the high-risk habitat model in the study, showing the best diagnostic efficacy among all models (AUC(CRM) vs. AUC(Habitat 3) vs. AUC(CRM + Habitat 3) = 0.84(95%CI:0.71–0.97) vs. 0.90(95%CI:0.80-1.00) vs. 0.79(95%CI:0.65–0.93)). Moreover, integrating the Habitat 3 model with clinical features and constructing nomograms enhanced the predictive capability of the combined model (AUC = 0.95(95%CI:0.88-1.00)). In this study, an independent external testing cohort was utilized to assess the model's accuracy, yielding an AUC of 0.88 (95%CI: 0.78–0.98).

[†]Han Liu and Chun-Jie Hou contributed equally to this work.

*Correspondence:
Li-Tao Sun
litaosun1971@sina.com
Jing-Lan Tang
tangjinglan_85@163.com

Full list of author information is available at the end of the article

Conclusion The integration of the High-Risk Habitats (Habitat 3) radiomics model with clinical characteristics demonstrated a high predictive accuracy in identifying LLNM. This model has the potential to offer valuable guidance to surgeons in deciding the necessity of LLNM dissection for DTC.

Clinical trial number Not applicable.

Keywords Ultrasound, Differentiated thyroid cancer, Radiomics, Habitat analysis, Lateral cervical lymph node metastasis

Background

More than 90% of malignant growths detected in the thyroid gland are categorized as differentiated thyroid carcinoma (DTC), including papillary and follicular thyroid carcinomas [1, 2]. While DTC is widespread, most patients have favorable survival rates and prognosis due to its slow-growing nature. However, some nodules display high invasiveness, leading to early local recurrence and distant metastasis [3, 4]. Studies have shown that removing central regional lymph nodes (CLMN) alongside primary lesion resection significantly influences the prognosis of DTC patients. Hence, it's recommended to routinely remove CLMN during thyroidectomy [5–7]. Accurately predicting lateral lymph node metastasis (LLNM) is crucial for determining the appropriate surgical approach and extent of clearance [8, 9].

High-frequency ultrasonography is endorsed by multiple international guidelines as the primary monitoring modality for clinical follow-up of thyroid cancer [5, 10]. It enables clear visualization of the internal structure and echogenicity of DTC nodules, allowing ultrasonographers to predict LLNM by identifying specific signs within the thyroid nodules [9, 11]. Nevertheless, this prediction method relies heavily on the sonographer's individual expertise and diagnostic proficiency, which are inherently subjective. The presence of artificial bias in the assessment of thyroid nodule characteristics by different sonographers poses a challenge in achieving consistency in clinical practice. The advancement of artificial intelligence technology has enabled image pictures to offer a number of radiomics features, thereby enhancing model accuracy [12]. The radiomics currently employed for tumor analysis quantifies features and heterogeneity by examining the interplay among intratumor voxels [13]. Recent radiomics research on LLNM originating from thyroid cancer has predominantly concentrated on traditional radiomics techniques. Nevertheless, these conventional methods are limited in their capacity to offer only a generalized radiomic characterization of the nodules within the region of interest (ROI). Consequently, this approach may fall short in capturing the intricate microscopic features and complex biological behaviors of thyroid nodules, potentially compromising the precise diagnosis of lymph node metastasis [14]. It is important to note that various microenvironments exist within the tumor, which may be influenced by factors such as complex blood supply, distinct necrotic regions, and gene expression patterns. This heterogeneity contributes to the diverse phenotypic characteristics observed in different regions of the tumor [15]. Furthermore, the interpretability of conventional radiomics features is limited, a challenge that is especially pronounced in the context of medical image analysis. While radiomics extracts high-dimensional features from medical images, offering valuable insights for disease diagnosis and prognosis, its "black box" nature poses significant challenges for clinicians attempting to comprehend the model's decision-making process [16]. A study demonstrates that deep learning surpasses traditional machine learning in performance and cross-institutional generalizability for predicting treatment outcomes in head and neck cancer through the analysis of imaging data [17]. Nevertheless, the issue of interpretability in deep learning models persists, particularly in scenarios where tumor boundaries are not well-defined, thereby complicating their application in clinical settings [17].

Habitat imaging represents an objective and non-invasive image analysis technique grounded in the machine learning principles of conventional radiomics. This approach delineates tumor subregions by analyzing the interrelationships among intratumoral voxels or pixels, thereby enabling the extraction of high-throughput features that characterize tumor heterogeneity within each subregion [18]. The analysis of habitats and segmentation of subregions elucidate a connection between the tumor microenvironment and macroscopic imaging, indicating a potential correlation between the spatial heterogeneity of images and their pathological and molecular characteristics. This correlation holds significant scientific value and promising clinical application prospects. Utilizing statistical analysis and correlation studies, it is possible to associate high-risk subregions of differentiated thyroid carcinoma (DTC) nodes, identified through habitat analysis segmentation, with lymph node metastasis in the lateral neck. These high-risk subregions can be visually distinguished by color variations. In contrast to traditional radiomics and deep learning features, which often pose challenges in clinical interpretation, the habitat analysis radiomics model offers a more intuitive approach for predicting the pattern of results. This study

will implement a novel methodology to segment tumors into distinct subregions with identical voxel sets before conducting feature screening. Its objective is to explore the correlation between different subregions and subregional features within DTC nodules and LLNM, and evaluate the diagnostic accuracy of models utilizing these subregions.

Materials and methods

Patient recruitment and selection

This retrospective study received approval from the Ethical Review Committee of Zhejiang Provincial People's Hospital (Approval No. QT2024120) and waived the requirement for informed consent. In this study, two checklists—the CLEAR checklist and the METRICS checklist—will be utilized to evaluate the quality of the radiomics studies. These checklists will be made available as supplementary information [19]. Between August 2021 and August 2023, clinical data and ultrasonographic images were collected from 332 patients pathologically confirmed to have DTC for this investigation. Furthermore, this study retrospectively gathered data from an additional cohort of 54 patients between September 2023 and January 2024, serving as an independent external testing group. Inclusion criteria comprised: (1) All patients undergoing preoperative ultrasonography. (2) Patients having comprehensive clinical data, including age, sex, thyroid hormone and antibody levels, and DTC node dimensions determined by ultrasound. (3) Thorough pathological details encompassing DTC nodal pathology and LLNM pathology (LLNM confirmation achieved through preoperative fine-needle aspiration biopsy or intraoperative pathology). (4) To address the significant disparity in sample sizes between patients without LLNM and those with LLNM, this study adjusted the sample sizes to achieve approximate equivalence between the two groups. To mitigate the risk of selection bias, patients without lymph node metastasis were randomly selected using a computer-based algorithm.

Exclusion criteria were: (1) Thyroid nodules < 2 mm in diameter (According to the guidelines established by the American Association of Clinical Endocrinologists and the American College of Endocrinology for fine-needle aspiration of thyroid nodules, and considering the specific circumstances of our study center, FNA and surgery are generally not recommended for nodules smaller than 2 mm.). (2) Prior treatments (radiotherapy, chemotherapy, or ablation surgery). (3) Images not meeting diagnostic criteria (blurry, low resolution). (4) Comorbidities with other thyroid cancers. (5) Presence of multiple cancerous nodules in the thyroid.

Following the selection process, a cohort of 214 patients was included in the study, with a mean age of 42.6 years and a standard deviation of ± 12.7 years. Among these patients, 107 had combined LLNM, while 107 did not. Subsequently, patients were randomly allocated to either the training or internal testing groups at an 8:2 ratio, as shown in Fig. 1. Furthermore, following the screening process, 43 patients were enrolled in this study as an independent external testing cohort. Among these, 23 patients were not present with lateral lymph node metastasis (LLNM), while 20 patients were diagnosed with LLNM.

Clinical data and ultrasound images

The study compiled clinical data including age, gender, maximum meridian of DTC nodules measured by ultrasound, patients' thyroid hormone levels, and antibody levels, retrieved from the electronic case system of Zhejiang Provincial People's Hospital. The laboratory data encompassed measurements of thyroid-stimulating hormone (TSH), thyroglobulin antibody (TGAB), thyroid peroxidase antibody (TPOAB), total triiodothyronine (TT3), total thyroxine (TT4), free triiodothyronine (FT3), and free thyroxine (FT4). These data were procured from the Department of Laboratory Medicine at Zhejiang Provincial People's Hospital. Based on clinical evidence and prior studies, the following thresholds were established for TSH, TGAB, and TPOAB: TSH at 2.5 ng/ml, TGAB at 100 IU/ml, and TPOAB at 35 IU/ml [20, 21]. Ultrasound images were captured using various ultrasound diagnostic instruments: Kunlun 9 (Mindray, Shenzhen, China), LOGIQ E9 (GE Healthcare, USA), and ACUSON Sequoia (Siemens, USA) ultrasound color Doppler diagnostic instruments equipped with high-frequency line array probes (probe frequency 6–15 MHz). To facilitate image analysis and feature extraction, all ultrasound images were gathered in Dicom format and then converted to nii format prior to undergoing preprocessing. Voxel intensity values of the images were discretized to mitigate noise and standardize intensity levels, ensuring consistent resolution across all images. Additionally, min-max normalization (Grayscale range: 0-255) was applied to the image signal intensities to minimize discrepancies in signal intensity arising from variations in imaging equipment and operator techniques [22, 23]. Z-score normalization was applied to the grey values of the images to mitigate the impact of heterogeneity in imaging features resulting from varying parameter settings [24]. After processing, the physical dimensions of the original image are $1260 \times 910 \times 1$ pixels, with a voxel size of $1 \times 1 \times 1$ millimeters. Then, the ROI corresponding to DTC were identified as target areas. These areas were manually segmented by a skilled thyroid-specialized ultrasonographer using ITK Snap software (version 3.8.0, http://itksnap.org), as illustrated in Fig. 2.

In this study, the ROI nodule images were initially hyper-segmented to generate multiple hyperpixels, which

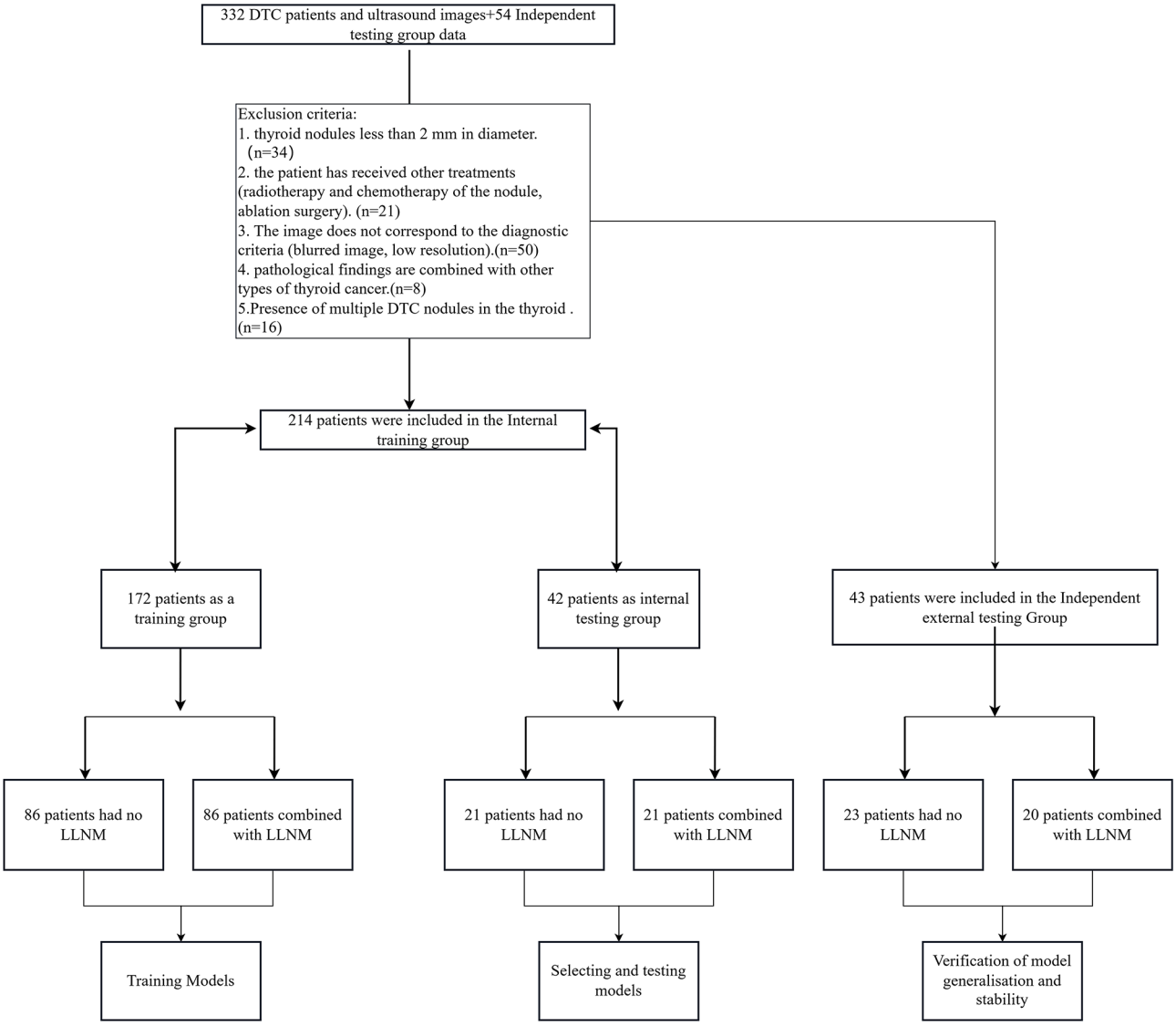

Fig. 1 Patient selection flowchart; LLNM lateral lymph node metastasis; DTC differentiated thyroid carcinoma

were then used to create different subregions. Given the challenges in visualizing subregions of DTC nodules, the K-Means method was employed for automatic segmentation [25]. This method divides the tumor volume into spatially distinct regions, visually represented in different colors. The segmentation process utilized the OKT-gen_roi_rad_features program component, based on the OnekeyAI platform (Version: 20240121, http://www.medai.icu). Calinski-Harabasz scores were computed individually for clustering iterations ranging from 2 to 10 and visualized in a line graph to aid in determining the optimal number of clusters [26]. The OKT-gen_habitat_cluster module, integrated within the OnekeyAI platform, was employed for this analysis. Results indicated that the clustering iteration yielding the highest Calinski-Harabasz score was 5. Subsequently, the ROI habitat images were clustered 5 times to serve as masks for habitat analysis within the radiological group, as illustrated in Fig. 3. Habitat analysis principles and parameter differences can be found in the supplementary documentation.

Radiomics feature extraction

In this study, features from 2D ultrasound images are extracted. This study will develop three models for comparison: a conventional tumor radiomics model for DTC nodules, a high-risk habitat model, and a combined tumor radiomics and high-risk habitat analysis model. Each of the first two models will extract 1562 radiomics features. The third model will integrate features from the first two models for screening purposes, as illustrated in Fig. 2 (b). Various methods, gray-level co-occurrence matrix (GLCM, 22.4%), gray-level run length matrix

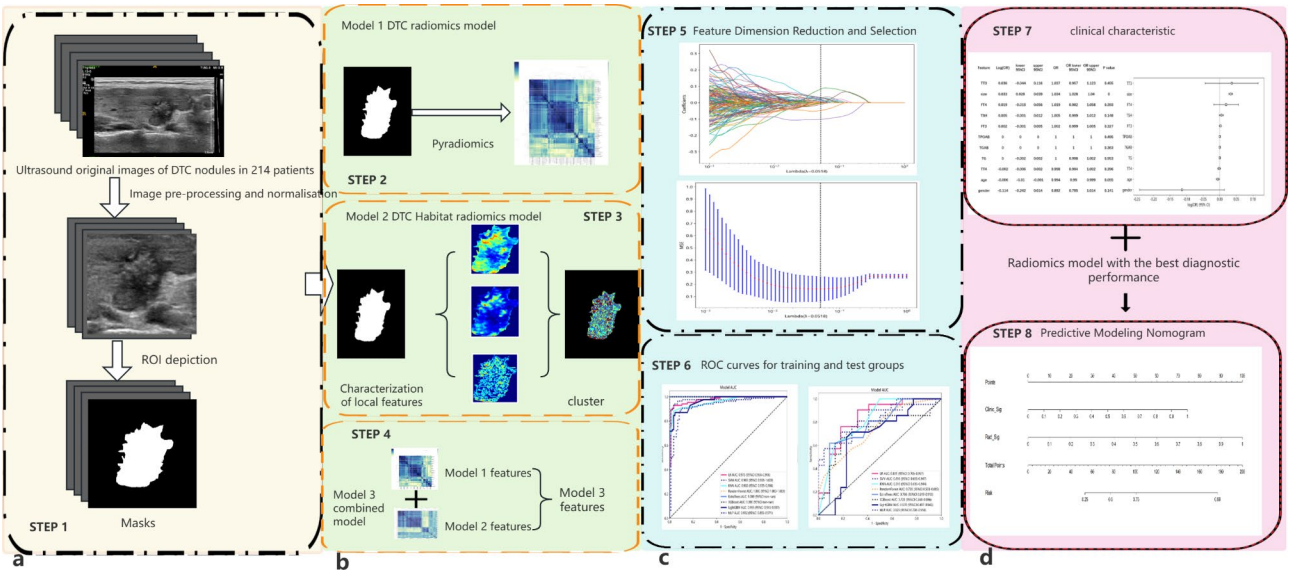

Fig. 2 Overview of the study design. (**a**) Image pre-processing and ROI selection. (**b**) Feature extraction. Extracted features of three models, including a traditional machine learning model, a habitat analysis model, and a combined model. (**c**) Feature selection and Model validation. (**d**) Combining clinical features and making nomograms

(GLRLM, 15.0%), gray level size zone matrix (GLSZM, 15.0%), neighborhood gray-tone difference matrix (NGTDM, 4.7%), grey Level Co-occurrence Matrix shape (GLDM, 13.1%) and first-order (16.8%) will be employed to calculate texture features in this study. The feature extraction method is based on Pyradiomics (version 2.2.0; http://pyradiomics.readthedocs.io). The software complies with the benchmarks and certifications established by the Imaging Biomarker Standardization Initiative (IBSI). To evaluate the reproducibility of feature extraction, ultrasound images of 50 DTC patients were randomly selected, and ROI segmentation was independently performed by two ultrasonographers using a double-blind method. The inter/intra-observer correlation coefficient (ICC) was used to evaluate inter- and intra-observer consistency of feature extraction. Intra-observer consistency was considered excellent for ICC ≥ 0.8, with a final ICC value of 0.85. This study indicates that the pertinent parameters within the code, encompassing those related to preprocessing and radiomics feature extraction, are configured to their default settings. This study encountered a limited number of missing values during the data analysis phase. Given the extensive dataset and the multitude of features, a detailed examination of the quantity and distribution of these missing values was not feasible. The approaches for addressing missing data in this study are as follows: (1) For normally distributed data, the mean imputation method is employed, wherein the mean of all values within the relevant feature column is utilized to replace missing entries. (2) For non-normally distributed data, the median imputation method is applied, whereby the median of the available values in the feature column is used for imputation. The system code autonomously computes and populates all missing values.

Feature screening and dimension reduction

To eliminate redundant and less significant features, the Mann-Whitney U test was applied to the extracted features, retaining only those radiological features with a p-value less than 0.05. For highly repetitive features, Spearman's rank correlation coefficient, Pearson correlation coefficient and Kendall correlation coefficient were employed to assess the correlation between features, and any feature exhibiting an average absolute correlation greater than 0.9 was deemed redundant and subsequently removed. The top 22 ranked features were then identified for retention using the minimum redundancy maximum relevance (mRMR) method. Finally, the most informative predictive features related to lymph node metastasis (LLNM) were identified through a 10-fold cross-validated least absolute shrinkage and selection operator (LASSO) method. In this study, the selection of regularized hyperparameters was conducted with the regularization coefficient λ serving as the criterion for determining the optimal number of features. Analysis of the LASSO plots indicated improved model convergence, while the mean squared error (MSE) plots demonstrated that the model achieved the best fit with 22 features, as illustrated in Fig. 2(C) Step 5. Subsequently, the radiomics feature scores for each patient in the training dataset were computed using the derived formula. To mitigate the risk of overfitting associated with an excessive number of features in the dataset, an independent

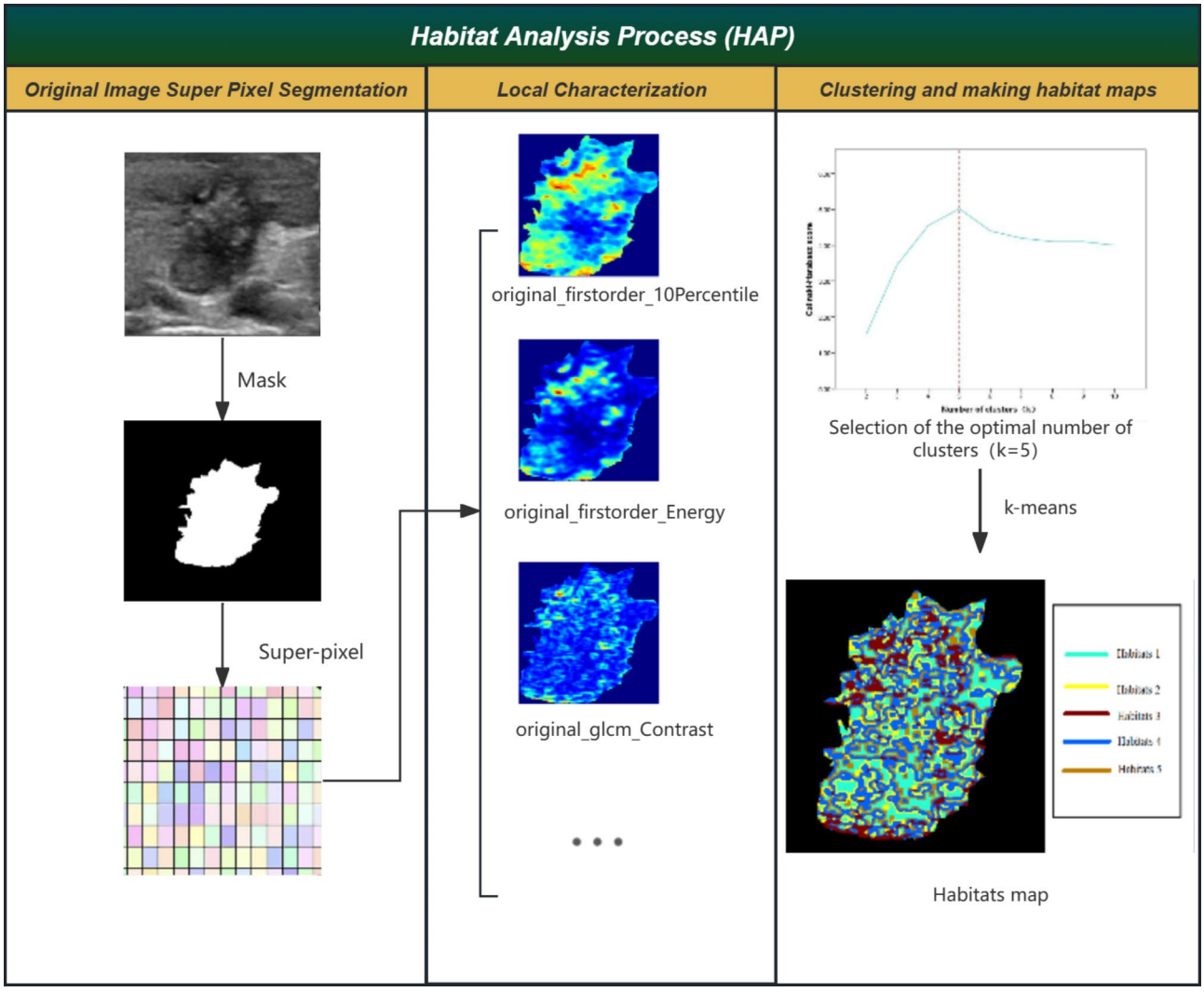

Fig. 3 Habitat analysis process. The mask obtained from each image is segmented for hyperpixel, and the computed voxels and data are used for local feature inscription and different feature images are obtained according to different computational methods. In this study, the best number of clusters is selected and clustered based on Calinski-Harabasz scores and the best number of clusters is 5. All the Masks are classified into five different subregions and models are computed

external testing cohort was incorporated into this study to assess the model's generalization capability. The processes of feature selection and dimensionality reduction for the external testing cohort were conducted using the same methodology as applied to the internal training cohort.

Radiomics modeling and comparison of model efficacy

In the internal training group of this study, eight machine learning algorithms were selected: LR, SVM, KNN, Random Forest, Extra Trees, XGBoost, LightGBM, and MLP. Subsequently, three distinct models were developed: (1) a conventional histology model of DTC nodes to predict lateral neck lymph node metastasis, (2) five prediction models based on individual habitats, with an overarching model summarizing all habitats to identify high-risk models, and (3) a traditional radiomics model of DTC nodes integrated with habitat analysis features to construct prediction models.

The internal testing group was utilized to assess the accuracy of the three models, employing the ACC curve and AUC curve to determine the most optimal algorithm and model. Subsequently, a comparative analysis was conducted. This research employed an independent external testing cohort to assess the final model's accuracy and generalizability.

Statistical analysis.

T-tests were employed for the analysis of quantitative data, while chi-square tests or Fisher's exact probability method were utilized for the comparison of count data. Multiple logistic regression equations were applied to

assess clinical influences. The predictive efficacy of models was compared using metrics such as AUC, ACC, SEN, SPE, PPV, and NPV. Statistical significance was determined by P-values less than 0.05. The study employed both post-hoc power calculation and pre-experimental sample size analysis to ensure appropriate sample size and statistical power. The post-hoc power calculation was conducted using G*Power software (version 3.1.9.7), selecting the t-test for the difference between two independent means, resulting in a power value (1-β error probability) exceeding 0.8. For the pre-experimental sample size assessment, the sample size calculator available on Calculator.net was utilized, which recommended a minimum sample size of 220. This recommendation was lower than the actual sample size of 257 used in the study. The outcomes of the calculations are presented in the supplementary file. Software tools utilized in this study included R Studio (based on R 4.2.1), Anaconda 3 (Python 3.9), SigmaPlot (version: 14.0), and the Onekey AI platform.

Results

Patient characteristics

The postoperative pathology of the patients in this study was analyzed, revealing that follicular thyroid carcinoma was diagnosed in 21 out of 257 patients (8%), while papillary thyroid carcinoma was diagnosed in 236 patients (92%). Clinical baseline characteristics, including age, gender, laboratory tests, and tumor size, were compared between patients with LLNM and those with other conditions after conducting multifactorial analysis. The only statistically significant difference observed was in the maximum measured diameter (size) of the tumors, as indicated in Table 1.

Model selection and diagnostic performance

Model algorithm selection

In this investigation, eight common machine learning algorithms were utilized, necessitating the determination of the most suitable algorithm [27]. AUC curves and accuracy curves were plotted and examined individually for the DTC nodal histology model, the high-risk habitat model (Habitats 3), and the combined model employing all eight machine learning algorithms. Subsequently, based on the AUC values and accuracy curves to avoid underfitting or overfitting, the LR algorithm was selected as the machine learning algorithm for all models in this study, as shown in Fig. 4. Based on the comparative analysis of eight distinct algorithms within the learning curve of the final model (Clinical + Habitat model), the LR model demonstrated the smallest disparity between the training set and cross-validation set, coupled with the highest AUC value. Consequently, selecting the LR model mitigates the risk of overfitting attributable to feature selection. Learning curve diagrams can be found in the supplementary document.

Habitat analysis and subregional models

The nodules underwent habitat analysis and clustering, resulting in their categorization into five distinct subregions aligned with the five habitats (habitats 1–5) depicted in the image. Each habitat could act as a mask to extract the corresponding features. ROC curves for the habitats 1–5 models, based on the LR model, were plotted and compared within the test group. The calculated AUC values for these models were 0.80(95%CI:0.67–0.94), 0.82(95%CI:0.69–0.95), 0.90(95%CI:0.80-1.00), 0.81(95%CI:0.68–0.95), and 0.81(95%CI:0.67–0.94), respectively, as shown in Table 2. Eventually, the amalgamation of habitat characteristics was used to construct a radiomics model for comprehensive habitat analysis. Subsequently, the ROC curve was generated, revealing

Table 1 Clinical baseline characteristics of LLNM

Characteristic	Training Cohort($n=172$)		*p*-value	Internal testing Cohort($n=42$)		*p*-value	External testing Cohort($n=43$)		*p*-value
	LLNM+	LLNM-		LLNM+	LLNM-		LLNM+	LLNM-	
Age(years)	40.73 ± 13.73	44.58 ± 12.27	0.07	42.90 ± 11.86	42.23 ± 10.89	0.84	40.05 ± 14.10	44.46 ± 9.93	0.24
TT3(μg/L)	1.19 ± 1.11	1.10 ± 0.19	0.43	1.05 ± 0.26	1.14 ± 0.22	0.16	1.06 ± 0.30	1.14 ± 0.16	0.09
TT4(μg/L)	88.84 ± 18.95	91.05 ± 14.76	0.84	83.76 ± 16.70	91.42 ± 14.62	0.16	85.01 ± 22.64	95.00 ± 15.99	0.19
FT3(ng/L)	6.41 ± 27.89	3.43 ± 0.43	0.89	3.28 ± 0.40	3.46 ± 0.43	0.12	3.14 ± 0.53	3.60 ± 0.49	**0.01**
FT4(ng/L)	9.02 ± 1.94	8.80 ± 1.45	0.16	8.58 ± 1.34	8.93 ± 1.09	0.25	8.64 ± 2.06	8.84 ± 1.52	0.89
Size(mm)	17.62 ± 9.82	6.09 ± 3.37	**< 0.001**	15.95 ± 8.89	6.36 ± 3.51	**< 0.001**	14.89 ± 8.06	5.96 ± 3.14	**<0.01**
Sex(%)			0.19			1.00			0.95
Male	42(48.84)	32(37.65)		8(38.10)	9(40.91)		9(47.37)	10(41.67)	
Female	44(51.16)	53(62.35)		13(61.90)	13(59.09)		10(52.63)	14(58.33)	
TSH (ng/ml)	2.17 ± 1.57	4.46 ± 14.42	0.52	2.21 ± 1.27	1.66 ± 0.98	0.31	6.58 ± 21.09	1.59 ± 0.97	0.50
TGAB (IU/ml)	76.55 ± 297.16	41.66 ± 191.67	0.41	85.97 ± 370.71	8.63 ± 25.65	1	107.22 ± 389.62	66.03 ± 255.98	0.69
TPOAB(IU/ml)	79.92 ± 208.98	59.11 ± 188.79	0.79	31.63 ± 109.01	60.27 ± 135.87	0.78	39.49 ± 116.42	90.04 ± 191.98	0.44

Significant differences are highlighted in the boldface. The p-value is the result of a one-way ANOVA. TSH, thyroid stimulating hormone; TGAB, thyroglobulin antibody; TPOAB, thyroid peroxidase antibody; TT3, total triiodothyronine; TT4, total thyroxine; FT3, free triiodothyronine; FT4, free thyroxine

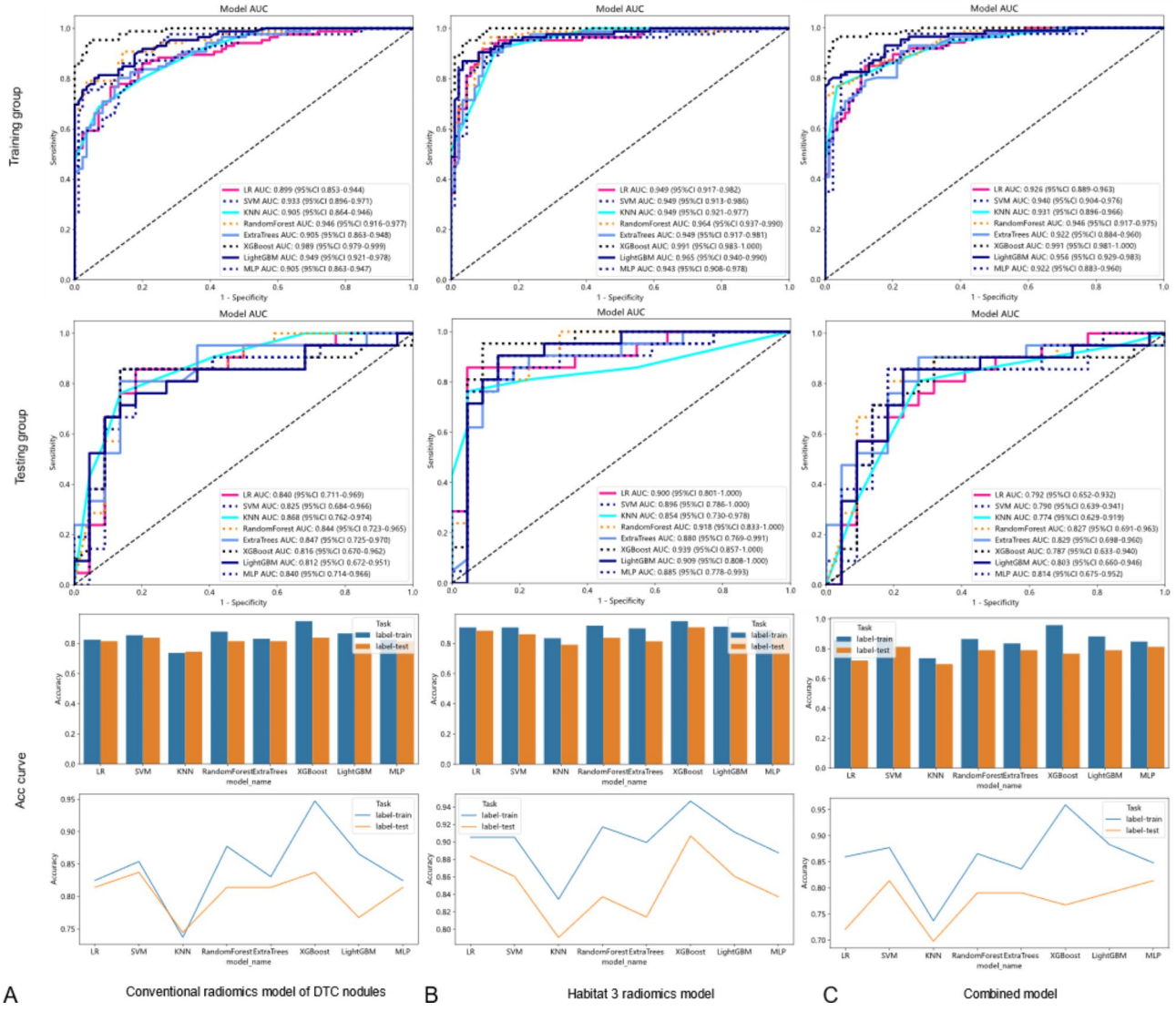

Fig. 4 Diagnostic efficacy of the three models. LR Logistic Regression; SVM Support Vector Machine; KNN, K-Nearest Neighbors; XGBoost Extreme Gradient Boosting; LightGBM Light Gradient Boosting Machine; MLP Multilayer Perceptron; AUC Area Under the Curve; ACC curve Accuracy curve. (**A**) Traditional radiomics model for DTC nodules; (**B**) Habitat 3 model; (**C**) Combined model (combining the traditional radiomics model and the Habitat 3 model)

an AUC value of 0.80(95%CI:0.67–0.94) for the combined habitat model. Following an evaluation of all habitat-related models in the study, habitats3 emerged as a high-risk habitat model based on its superior AUC and accuracy (ACC) values. Upon examination of the weights assigned to the selected features, it was determined that the feature "Small Area Low Gray Level Emphasis" in the Habitats3 model emerged as the predominant risk factor in predicting LLNM (see Fig. 5).

Comparison of performance between models

In this research, the habitats3 model will be selected and integrated with all components of the conventional radiomics model of DTC nodules to establish a composite model. This model incorporates a total of 3122 histologic features. Subsequently, ROC curves were generated for three sets of models (traditional radiomics model, habitats3 model, habitats3 + traditional radiomics model), and their AUC values were compared. The diagnostic performance of the habitats3 high-risk habitat model (internal testing group AUC: 0.90(95%CI:0.80-1.00)) surpassed that of the other two sets of models. Please see Fig. 4 for visual representation.

Clinical features combined with habitat analysis model

In this study, the identified clinical features were integrated with a high-risk habitat model and analyzed using a Nomogram. The predictive accuracy of the Nomogram was improved, achieving an AUC value of 0.97(95%CI:0.95–0.99) for the training group and

Table 2 The performance of the five intratumoral habitat models and the comprehensive habitat model

Model	ACC	AUC	95%CI	SEN	SPE	PPV	NPV
Habitats 1							
Training group	0.778	0.877	0.828–0.926	0.640	0.918	0.887	0.716
Testing group	0.744	0.803	0.665–0.940	0.810	0.682	0.708	0.789
Habitats 2							
Training group	0.840	0.915	0.875–0.955	0.847	0.833	0.837	0.843
Testing group	0.744	0.823	0.697–0.948	0.857	0.636	0.692	0.824
Habitats 3							
Training group	0.905	0.949	0.917–0.982	0.905	0.906	0.905	0.906
Testing group	0.884	0.900	0.801-1.000	0.810	0.955	0.944	0.840
Habitats 4							
Training group	0.905	0.955	0.926–0.985	0.917	0.894	0.895	0.916
Testing group	0.814	0.814	0.675–0.952	0.810	0.818	0.810	0.818
Habitats 5							
Training group	0.832	0.891	0.842–0.939	0.845	0.819	0.826	0.840
Testing group	0.767	0.805	0.667–0.944	0.762	0.773	0.762	0.773
Co-habitats model (Habitats 1–5)							
Training group	0.778	0.877	0.828–0.926	0.640	0.918	0.887	0.716
Testing group	0.744	0.803	0.665–0.941	0.810	0.682	0.708	0.789

AUC, area under the curve; CI, confidence interval; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value

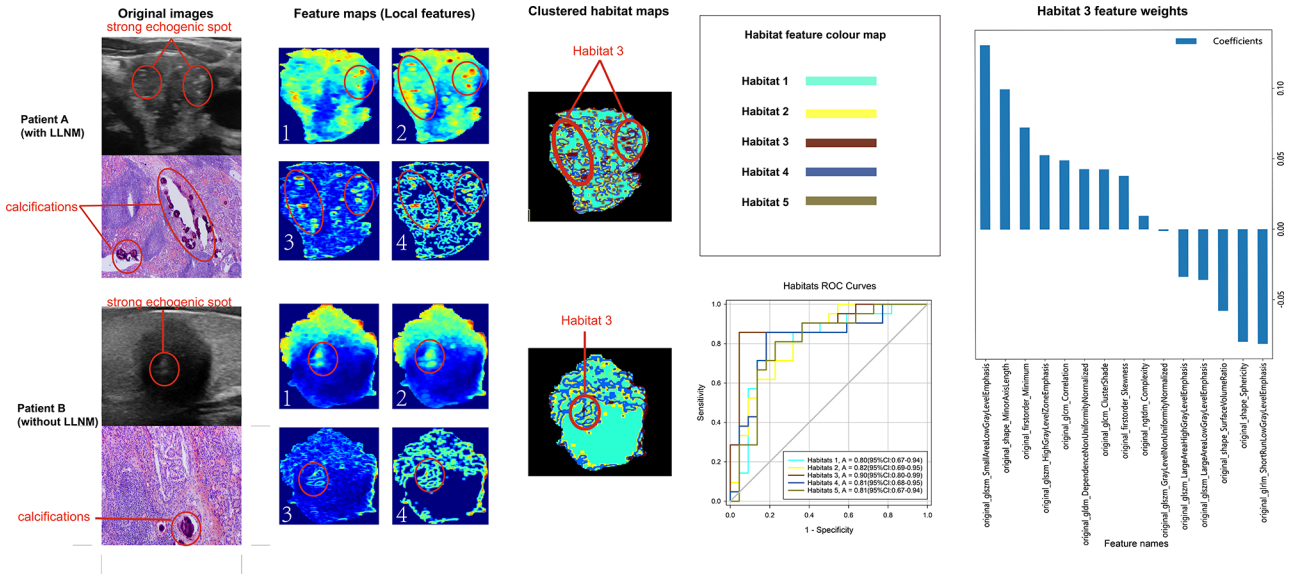

Fig. 5 Comparative analysis of receiver operating characteristic (ROC) curves for prediction. Patient A exhibited a higher quantity and greater density of strong spots, characterized as gravel-like calcifications in the pathology, on ultrasound imaging compared to Patient B. This disparity is more effectively visualized in the localized feature map. The calcifications appeared as dark red on the habitat feature map, aligning with habitat 3. Among the various habitat models tested, the habitat 3 model demonstrated superior performance in the test group, with an AUC value of 0.90. The feature "Small Area Low Gray Level Emphasis" emerged as the most influential machine learning model feature within the habitat 3 model. LLNM lateral lymph node metastasis; A area under the curve

0.95(95%CI:0.88-1.00) for the test group. The nomogram model was validated using an independent external testing cohort, resulting in an AUC of 0.88 (95%CI: 0.78–0.98). Additionally, calibration curves were generated to evaluate the model's stability and DCA curves were used to evaluate the clinical benefits of the model, as depicted in Fig. 6.

Discussion

There are several imaging methods available for observing and predicting LLNM in DTC nodes. However, ultrasonography remains the preferred method due to its cost-effectiveness, real-time capabilities, and ability to provide clear imaging of superficial organs [28, 29]. Accurately predicting LLNM preoperatively in DTC patients is crucial for surgeons to determine the appropriate extent

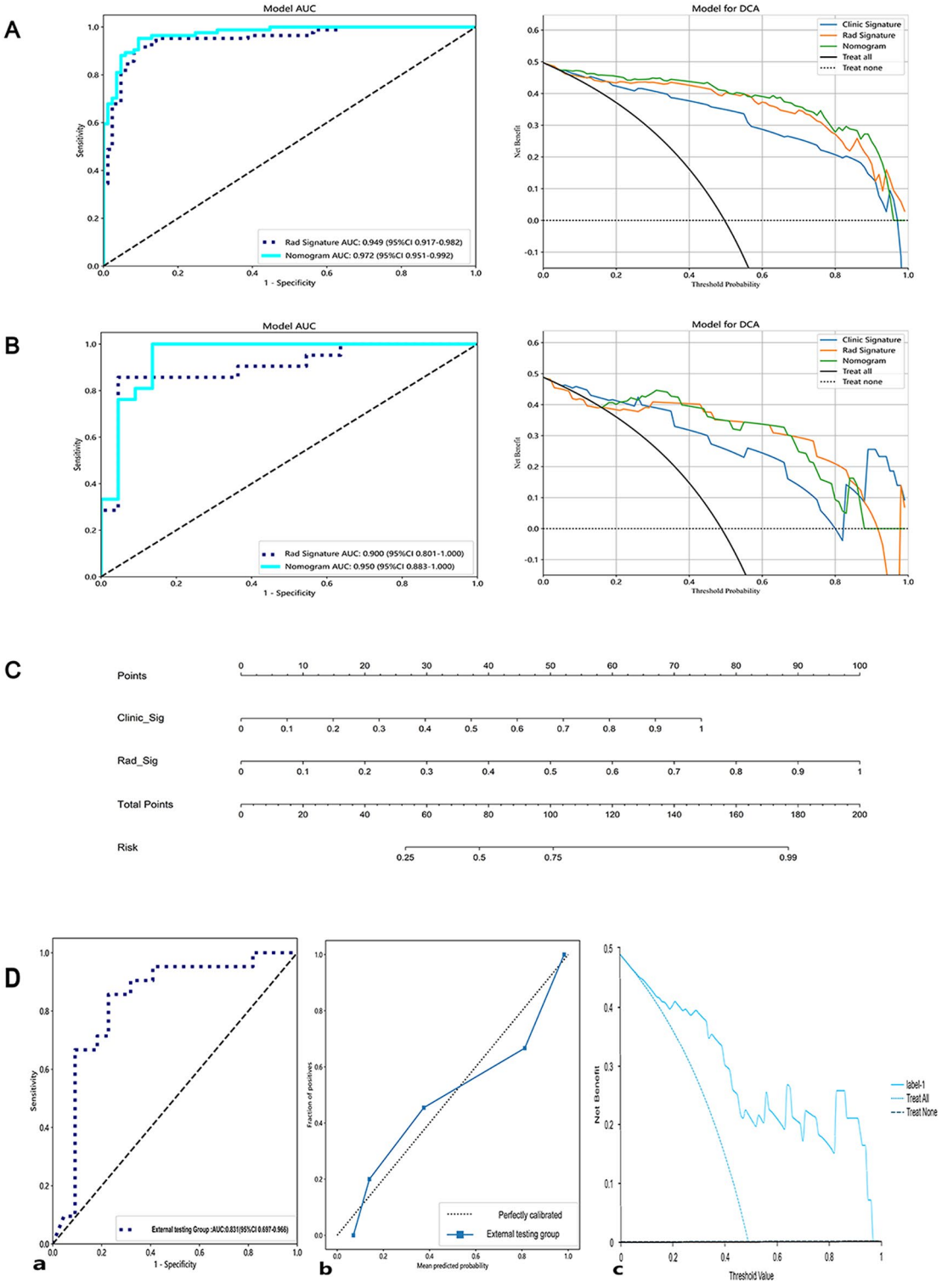

Fig. 6 Modeling results after incorporating clinical features. (**A**) AUC curve and DCA curve of the training group model; (**B**) AUC curve and DCA curve of the testing group model; (**C**) Nomograms for the Habitat 3 + clinical characterization model. Rad Signature Radiomics model signature; AUC Area under the curve; DCA Decision curve. (**D**) Results of independent external testing group data, (a) ROC curve, AUC = 0.88 (95% CI: 0.78–0.98); (b) Calibration Curve, the calibration plot demonstrated a high degree of concordance between the observed probabilities and the predicted probabilities, indicating robust predictive accuracy; (c) DCA curve, after validation, the model has a good net benefit (0.49). Label-1, Independent external testing group predictions; label, 0 = without LLNM, 1 = with LLNM

of surgical intervention. However, diagnosing LLNM on ultrasound images can vary subjectively due to factors such as neck thickness, occlusion of acoustic shadows of bony structures, and the expertise of the ultrasonographer [30, 31]. The superficial location of the thyroid gland facilitates clearer and more comprehensive ultrasound imaging, less influenced by the diagnostician's maneuvers and experience. Meanwhile, there is increasing research utilizing radiomic methods to predict LLNM in thyroid cancer nodules, with growing validation of its predictive accuracy [32].

During clinical baseline characterization, the maximum diameter of DTC nodules emerged as a significant independent predictor of LLNM following both unifactorial and multifactorial retrospective analyses, consistent with findings reported by Liu C et al. [33]. To mitigate potential subjective biases, this study utilized radiomic characteristics of ultrasound images as the defining features of DTC nodules, rather than relying solely on subjective assessments by diagnostic ultrasonographers such as morphology, echogenicity, aspect ratio, and calcification.

In the present study, a radiomics model was developed and trained to predict LLNM from ultrasound images of DTC nodules in a cohort of 214 patients. High accuracy (ACC) and validity (AUC) were demonstrated by the model in the test group, consistent with findings from a multicenter radiomics model previously developed by Yuyang Tong et al. When this study validated the model with an independent external testing group, the AUC value could also reach 0.88 (95%CI: 0.78–0.98). This underscores the strong predictive capability of the radiomics model [34]. Building upon traditional radiomics, the habitat analysis radiomics model has been noticed and used by more and more scholars due to its advantages in the study of tumor heterogeneity. In this study, habitat analysis radiomics models were included for comparative analysis. Habitat analysis involves the quantification and segmentation of distinct regions within a DTC nodule to create habitats, facilitating a more intuitive examination of the correlation between various microenvironments of a tumor in an image and specific characteristic behaviors exhibited by the tumor (e.g., tumor spread, blood supply, necrotic areas, and metastasis) [15]. A comprehensive delineation within a specific ROI is conducted by the habitat analysis model, utilizing an objective computerized delineation of quantitative heterogeneity features within the nodule. This approach allows for in-depth exploration of features within the imaging data, resulting in high diagnostic value and more rational model interpretations, as evidenced by recent studies [26, 35–37].

In the present study, the habitats3 model was found to demonstrate the highest diagnostic efficacy compared to other radiomics models. When amalgamating the features of all habitat models to construct a fused habitat analysis model and subsequently comparing its efficacy with that of the habitats3 model, it was observed that the fused habitat model did not exhibit superior diagnostic performance. Integrating habitats3 with a traditional radiomics model in this study did not result in diagnostic efficacy surpassing that of habitats3 alone. Based on the above findings, the diagnostic efficacy of the habitats3 model, characterized as a high-risk habitat model, emerged as the most prominent among all the radiomics models examined.

By analyzing the habitat features and ultrasound images in this study, it was found that the characteristics of habitat 3 significantly overlapped with the hyperechoic spots seen in the ultrasound images within the DTC nodules. Moreover, habitat 3 emerged as the most reliable predictor of LLNM among the five subregions studied. These results suggest that detecting hyperechoic spots in ultrasound images is crucial for predicting LLNM. Upon examining the important radiomic features, it was found that the "Small Area Low Gray Level Emphasis" in habitat 3 was the most critical feature, indicating that the intensity of gray levels in hyperechoic spots is vital for predicting LLNM. It's important to note that the hyperechoic spots seen in ultrasound images usually correspond to microcalcifications in pathology [38, 39]. Previous studies have shown that the size of nodules and the intensity of ultrasound grayscale in calcified spots are important factors in predicting LLNM, which is consistent with earlier research [40]. This study combines macroscopic characteristics, such as size, with relatively microscopic radiomic features, specifically habitats, to comprehensively analyze the structure and microenvironment of DTC nodules. After screening, the predictive model developed in this study, which combines clinical data with habitat 3, demonstrates higher accuracy compared to traditional radiomics models. The investigation explores the relationship between tumor heterogeneity and lymph node metastasis through a detailed examination of image data. Upon integrating the clinical features, this study observed that the presence of a thyroid nodule characterized by a large tumor diameter and multiple small, strong echoes indicates a high-risk nodule with lateral lymph node metastasis (LLNM). By synthesizing the clinical and imaging histological features, this study aims to offer valuable reference and guidance for clinical practice.

The research methodology employed in this study is both innovative and effective in evaluating the robustness and generalization capacity of the model. The segmentation of tumors into distinct subregions for feature extraction offers novel insights into DTC heterogeneity and its correlation with LLNM. Furthermore, the study utilized an independent external testing cohort to assess

the model's performance, thereby demonstrating its generalization capability. This study employed a comprehensive array of data sources to thoroughly train the model, incorporating an extensive spectrum of clinical, laboratory, and ultrasound data. This methodological approach enhanced the model's stability.

This study presents certain limitations. Primarily, despite the application of normalization and standardization techniques during image preprocessing to mitigate discrepancies among images obtained from various ultrasound devices, the variability in image acquisition across different devices continues to impact the reproducibility of radiomic features. Consequently, it is imperative that future research endeavors incorporate prospective study designs to establish uniform acquisition parameters and utilize machines from the same manufacturer, thereby enhancing reproducibility [41]. Secondly, the delineation of ROI in this study was performed manually, potentially introducing a degree of subjectivity in the selection process. The implementation of automated segmentation techniques in future research could enhance the reproducibility and objectivity of the findings. Lastly, this investigation was retrospective in nature and involved a limited sample size. Expanding the sample size in future studies could further enhance the model's generalizability and clinical applicability.

Conclusions

In this study, habitat radiomics modeling and maps were used to identify high-risk areas in DTC nodules for predicting LLNM and to uncover significant radiomics and clinical features. The developed model exhibits high diagnostic efficacy in predicting LLNM and provides precise guidance for surgical decision-making, indicating its potential for various applications.

Abbreviations

LLNM	Lateral neck lymph node metastasis
DTC	Differentiated thyroid cancer
AUC	Area under the curve
CRM	Conventional radiomics model
CLMN	Central regional lymph nodes
ROI	Regions of interest
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
ICC	Inter/intra-observer correlation coefficient
mRMR	Minimum redundancy maximum relevance
ML	Machine learning
LR	Logistic Regression
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
MLP	Multilayer Perceptron
NAC	Neoadjuvant chemotherapy
NPV	Negative predictive value
PPV	Positive predictive value
PR	Progesterone receptor
ROC	Receiver operating characteristic
SDs	Standard deviations
SENS	Sensitivity
SPEC	Specificity

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01551-1.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The study is grateful to the "onekey" platform (http://www.medai.icu) for its technical solutions.

Author contributions

H.L., J.-L.T. wrote the main manuscript text and H.C., W.M. collects patients and research materials. L.-T.S., K.-F.L., Y.L., P.D. revised the article.

Funding

This work was supported by Zhejiang Province Education Department Project (No. Y202146098) and Zhejiang Province Medical Health Science and Technology Project (No.2021KY070).

Data availability

The underlying code [and training/validation datasets] for this study are not publicly available, but the corresponding author may make them available upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Zhejiang Provincial People's Hospital (protocol code QT2024120). Since the study was a retrospective study, informed consent was waived with the approval of the Ethics Committee of Zhejiang Provincial People's Hospital. Ethical proofs can be found in the Supplementary file at the end of the article.

Consent for publication

NA.

Competing interests

The authors declare no competing interests.

Author details

[1]Cancer Center, Department of Ultrasound Medicine, Zhejiang Provincial People's Hospital (Affiliated People's Hospital), Hangzhou Medical College, No. 158 Shang tang Road, Hangzhou, Zhejiang 310011, China
[2]Key Laboratory of Endocrine Gland Diseases of Zhejiang Province, Hangzhou, Zhejiang 310014, People's Republic of China
[3]Clinical Research Center for Cancer of Zhejiang Province, Hangzhou, Zhejiang 310014, People's Republic of China

Received: 24 July 2024 / Accepted: 2 January 2025
Published online: 13 January 2025

References

1. Jongekkasit I, Jitpratoom P, Sasanakietkul T, Anuwong A. Transoral endoscopic thyroidectomy for thyroid Cancer. Endocrinol Metab Clin North Am. 2019;48:165–80.
2. Chiapponi C, et al. Radioiodine Refractory follicular thyroid Cancer and surgery for cervical relapse. Cancers (Basel). 2021;13:6230.
3. Tori M, Shimo T. Long-term efficacy of lenvatinib for recurrent papillary thyroid carcinoma after multimodal treatment and management of complications: a case report. BMC Cancer. 2018;18:698.

4. Tang T, Li J, Zheng L, Zhang L, Shi J. Risk factors of central lymph node metastasis in papillary thyroid carcinoma: a retrospective cohort study. Int J Surg (London England). 2018;54:129–32.
5. Sugitani I, et al. Indications and strategy for active surveillance of adult low-risk papillary thyroid Microcarcinoma: Consensus statements from the Japan Association of Endocrine Surgery Task Force on Management for Papillary thyroid Microcarcinoma. Thyroid. 2021;31:183–92.
6. Sawka AM, et al. A prospective mixed-methods study of decision-making on surgery or active surveillance for low-risk papillary thyroid Cancer. Thyroid. 2020;30:999–1007.
7. Tuttle RM, Alzahrani AS. Risk stratification in differentiated thyroid Cancer: from detection to final Follow-Up. J Clin Endocrinol Metab. 2019;104:4087–100.
8. Strajina V, et al. Comprehensive lateral Neck dissection in papillary thyroid carcinoma may reduce lateral Neck Recurrence Rates. Ann Surg Oncol. 2019;26:86–92.
9. Chen L, Chen L, Liu J, Wang B, Zhang H. Value of qualitative and Quantitative Contrast-Enhanced Ultrasound Analysis in Preoperative diagnosis of cervical lymph node metastasis from papillary thyroid carcinoma. J Ultrasound Med. 2020;39:73–81.
10. Haugen BR, et al. 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid Cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid Cancer. Thyroid. 2016;26:1–133.
11. Back K, Kim JS, Kim JH, Choe JH. Superior located papillary thyroid Microcarcinoma is a risk factor for lateral lymph node metastasis. Ann Surg Oncol. 2019;26:3992–4001.
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77.
13. O'Connor JP, et al. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21:249–57.
14. Fan X, Zhang H, Wang Z, Zhang X, Qin S, Zhang J, Hu F, Yang M, Zhang J, Yu F. Diagnosing postoperative lymph node metastasis in thyroid cancer with multimodal radiomics and clinical features. Digit Health. 2024;10:20552076241233244.
15. Cho HH, et al. Measurement of Perfusion heterogeneity within Tumor habitats on Magnetic Resonance Imaging and its association with prognosis in breast Cancer patients. Cancers (Basel). 2022;14:1858.
16. Raptis S, Ilioudis C, Theodorou K. From pixels to prognosis: unveiling radiomics models with SHAP and LIME for enhanced interpretability. Biomed Phys Eng Express. 2024;10(3).
17. Huynh BN, Groendahl AR, Tomic O, Liland KH, Knudtsen IS, Hoebers F, van Elmpt W, Malinen E, Dale E, Futsaether CM. Head and neck cancer treatment outcome prediction: a comparison between machine learning with conventional radiomics features and deep learning radiomics. Front Med (Lausanne). 2023;10:1217037.
18. Park JE, Kim HS, Kim N, Park SY, Kim YH, Kim JH. Spatiotemporal heterogeneity in Multiparametric physiologic MRI is Associated with patient outcomes in IDH-Wildtype Glioblastoma. Clin Cancer Res. 2021;27(1):237–45.
19. Kocak B, Akinci D'Antonoli T, Mercaldo N, Alberich-Bayarri A, Baessler B, Ambrosini I, Andreychenko AE, Bakas S, Beets-Tan R, Bressem K, et al. METhodological RadiomICs score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. Insights Imaging. 2024;15(1):8.
20. Shen CT, Zhang XY, Qiu ZL, Sun ZK, Wei WJ, Song HJ, Luo QY. Thyroid autoimmune antibodies in patients with papillary thyroid carcinoma: a double-edged sword. Endocrine. 2017;58(1):176–83.
21. Kim SS, Lee BJ, Lee JC, Song SH, Kim BH, Son SM, Kim IJ, Kim YK, Kang YH. Preoperative serum thyroid stimulating hormone levels in well-differentiated thyroid carcinoma is a predictive factor for lateral lymph node metastasis as well as extrathyroidal extension in Korean patients: a single-center experience. Endocrine. 2011;39(3):259–65.
22. Sullivan NL, Marquis KA, Rudner DZ. Recruitment of SMC by ParB-parS organizes the origin region and promotes efficient chromosome segregation. Cell. 2009;137:697–707.
23. Zhang XD, et al. A combined radiomic model distinguishing GISTs from leiomyomas and schwannomas in the stomach based on endoscopic ultrasonography images. J Appl Clin Med Phys. 2023;24:e14023.
24. Ammari S, et al. Machine-learning-based Radiomics MRI Model for Survival Prediction of recurrent glioblastomas treated with Bevacizumab. Diagnostics (Basel Switzerland). 2021;11:1263.
25. Wu J, et al. Tumor Subregion Evolution-based imaging features to assess early response and predict prognosis in Oropharyngeal Cancer. J Nucl Med. 2020;61:327–36.
26. Yang Y, et al. Spatial heterogeneity of edema region uncovers survival-relevant habitat of Glioblastoma. Eur J Radiol. 2022;154:110423.
27. Ahsan MM, Luna SA, Siddique Z. Machine-learning-based Disease diagnosis: a Comprehensive Review. Healthc (Basel Switzerland). 2022;10:541.
28. Li T, Li H, Xue J, Miao J, Kang C. Shear wave elastography combined with gray-scale ultrasound for predicting central lymph node metastasis of papillary thyroid carcinoma. Surg Oncol. 2021;36:1–6.
29. Zhang H, et al. Prediction of cervical lymph node metastasis using MRI Radiomics Approach in Papillary thyroid carcinoma: a feasibility study. Technol Cancer Res Treat. 2020;19:1533033820969451.
30. Guo L, et al. Role of ultrasonographic features and quantified BRAFV600E mutation in lymph node metastasis in Chinese patients with papillary thyroid carcinoma. Sci Rep. 2019;9:75.
31. Park VY, et al. Radiomics signature for prediction of lateral lymph node metastasis in conventional papillary thyroid carcinoma. PLoS ONE. 2020;15:e0227315.
32. Zhu H, et al. Models of ultrasonic radiomics and clinical characters for lymph node metastasis assessment in thyroid cancer: a retrospective study. PeerJ. 2023;11:e14546.
33. Liu C, et al. Risk factor analysis for predicting cervical lymph node metastasis in papillary thyroid carcinoma: a study of 966 patients. BMC Cancer. 2019;19:622.
34. Tong 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;22:82.
35. da Silva Neto OP, et al. Pathophysiological mapping of tumor habitats in the breast in DCE-MRI using molecular texture descriptor. Comput Biol Med. 2019;106:114–25.
36. Verma R, et al. Stable and discriminatory Radiomic features from the Tumor and its Habitat Associated with Progression-Free Survival in Glioblastoma: a multi-institutional study. AJNR Am J Neuroradiol. 2022;43:1115–23.
37. Natrajan R, et al. Microenvironmental heterogeneity parallels breast Cancer progression: a histology-genomic integration analysis. PLoS Med. 2016;13:e1001961.
38. Kim BK, et al. Relationship between ultrasonographic and pathologic calcification patterns in papillary thyroid cancer. Med (Abingdon). 2018;97:e12675.
39. Yin L, Zhang W, Bai W, He W. Relationship between morphologic characteristics of Ultrasonic calcification in thyroid nodules and thyroid carcinoma. Ultrasound Med Biol. 2020;46:20–5.
40. Gomez NR, et al. Tumor size and presence of calcifications on ultrasonography are pre-operative predictors of lymph node metastases in patients with papillary thyroid cancer. J Surg Oncol. 2011;104:613–6.
41. Li MD, Cheng MQ, Chen LD, Hu HT, Zhang JC, Ruan SM, Huang H, Kuang M, Lu MD, Li W, et al. Reproducibility of radiomics features from ultrasound images: influence of image acquisition and processing. Eur Radiol. 2022;32(9):5843–51.

Publisher's note