

Explainable multimodal deep learning for predicting thyroid cancer lateral lymph node metastasis using ultrasound imaging

Received: 29 October 2024

Accepted: 9 July 2025

Published online: 01 August 2025

 Check for updates

Pengcheng Shen^{1,10}, Zheyu Yang^{2,10}, Jingjing Sun^{3,10}, Yun Wang⁴, Cheng Qiu⁵, Yirou Wang⁶, Yongyong Ren^{1,7}, Sheng Liu⁸, Wei Cai^{1,7,9} & Siqiong Yao^{1,7}

Preoperative prediction of lateral lymph node metastasis is clinically crucial for guiding surgical strategy and prognosis assessment, yet precise prediction methods are lacking. We therefore develop Lateral Lymph Node Metastasis Network (LLNM-Net), a bidirectional-attention deep-learning model that fuses multimodal data (preoperative ultrasound images, radiology reports, pathological findings, and demographics) from 29,615 patients and 9836 surgical cases across seven centers. Integrating nodule morphology and position with clinical text, LLNM-Net achieves an Area Under the Curve (AUC) of 0.944 and 84.7% accuracy in multicenter testing, outperforming human experts (64.3% accuracy) and surpassing previous models by 7.4%. Here we show tumors within 0.25 cm of the thyroid capsule carry >72% metastasis risk, with middle and upper lobes as high-risk regions. Leveraging location, shape, echogenicity, margins, demographics, and clinician inputs, LLNM-Net further attains an AUC of 0.983 for identifying high-risk patients. The model is thus a promising tool for preoperative screening and risk stratification.

Thyroid cancer is a prevalent malignancy worldwide, with an increasing incidence reported globally^{1,2}. Cervical lymph node metastasis, encompassing both central and lateral compartments, is a critical factor affecting patient prognosis, with an incidence rate of 20–50%³, and increasing the risk of mortality by 46%^{4,5}. The central compartment is widely recognized as the station for lymph node metastasis⁶, and numerous related studies have been conducted^{7–9}. In contrast, studies on lateral compartment metastasis are scarce, primarily due to: (1) the complexity of the anatomical structure¹⁰, with the distribution

pathways and connection patterns of lymphatic vessels around the thyroid varying among individuals¹¹; (2) the dispersed distribution of lateral lymph nodes, complicating statistical analysis and potentially leading to omissions¹²; (3) limitations of research methods—lateral lymph nodes are often located in deep cervical tissues, and their small size and depth make accurate detection and research challenging, with ultrasound sensitivity in predicting lateral lymph node metastasis (LLNM) being only 62%³; and (4) difficulty in sample acquisition. Biopsy of lateral lymph nodes requires a high level of clinician expertise and

¹Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, PR China. ²Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, PR China. ³Department of Ultrasound, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai, PR China. ⁴Hepatobiliary Pancreatic Center, Xuzhou Central Hospital, Xuzhou, Jiangsu Province, PR China. ⁵Medical college, Nantong University, Nantong, Jiangsu, PR China. ⁶Department of Endocrinology and Metabolism, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, PR China. ⁷SJTU-Yale Joint Center of Biostatistics and Data Science, Technical Center for Digital Medicine, National Center for Translational Medicine, Shanghai Jiao Tong University, Shanghai, PR China. ⁸Department of Thyroid and Breast Surgery, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai, PR China. ⁹Institute of Bioinformatics, Shanghai Academy of Experimental Medicine, Shanghai, PR China. ¹⁰These authors contributed equally: Pengcheng Shen, Zheyu Yang, Jingjing Sun. ✉e-mail: tocter@msn.com; cawei@shsmu.edu.cn; huilu@sjtu.edu.cn; yaosiqiong@sjtu.edu.cn

poses certain risks to patients, limiting the scale and depth of related studies¹³. Furthermore, central compartment lymph nodes are routinely dissected, whereas prophylactic dissection of lateral lymph nodes is not typically considered standard procedure in some countries and regions, leading to a significant shortage of available samples for research³.

Preoperative prediction of Lateral Lymph Node Metastasis (LLNM) is crucial for surgical planning and prognostic management in thyroid cancer. A positive LLNM result typically indicates that the tumor has begun to spread more extensively, prompting physicians to adopt a more aggressive treatment strategy, which may include expanding the surgical field and considering postoperative radiotherapy or chemotherapy¹⁴. Literature^{15,16} has shown that LLNM is associated with a worse prognosis; the recurrence rate is significantly higher compared to patients with central compartment lymph node metastasis (60% vs. 30%, $P = 0.007$). Disease-free survival and average recurrence time are also markedly shorter (30 months vs. 52 months, $P = 0.035$, and 7 months vs. 44 months, $P = 0.004$, respectively)¹⁵. Therefore, effectively predicting LLNM enables physicians to develop more appropriate treatment plans, reduce the risk of cancer progression due to missed dissections, and more accurately assess patient prognosis, providing more comprehensive support and care^{3,16}. In some cases, patients may present with skip metastases—negative central compartment and positive lateral compartment^{17,18}—which are prone to being missed during preoperative evaluation and surgery¹⁹. Moreover, the prognosis of skip metastases varies among different tumor types^{20,21}, suggesting that clinicians should consider the specific biological characteristics of the tumor and the anatomical pathways of the lymphatic system¹³. Currently, preoperative lymph node biopsy is the standard method for evaluating LLNM³; however, ultrasound-guided fine-needle aspiration has limitations, such as inaccurate or missed punctures^{3,22}. Given the particular importance of LLNM, there is an urgent need for preoperative evaluation methods that can effectively predict LLNM, assisting clinicians in determining the nature of the disease and taking appropriate measures, thereby contributing significantly to improving patient survival rates.

Current research indicates that cervical lymph node metastasis is closely associated with the histological morphology and location of the primary tumor^{6,23,24}. Specifically, it often results from the growth and spread of the primary tumor (thyroid nodule), with different tumor types exhibiting varying metastatic tendencies. For instance, abnormally enlarged tumors may be an important sign of lymph node metastasis²⁵, while tumors with abnormal morphology or texture may also suggest the possibility of metastasis^{6,26}. Moreover, the growth location of the primary tumor affects the invasion pathway of tumor cells into the lymphatic system, increasing the likelihood of superior pole metastasis^{23,24} and influencing the risk and prognosis of lymph node metastasis in different regions^{3,15}.

Ultrasound imaging, known for its non-invasive, real-time, and convenient features, is one of the most common diagnostic methods for thyroid cancer. It helps physicians detect early tumor abnormalities such as increased size, irregular shape, and abnormal internal structure, thereby playing a predictive role in LLNM^{27–29}. However, it suffers from low inter-organ contrast and poor image quality, and evaluation results heavily depend on the physician experience. Deep learning techniques can enhance tumor recognition by learning image features such as tumor morphology, size, and calcification^{6,30–33}. Techniques like foreground-background algorithms and graph convolutional networks statistically analyze positional information⁷, assisting physicians in preoperative diagnosis and prognosis assessment, including tumor malignancy grading, subtype evaluation, and prediction of cervical lymph node metastasis^{34,35}. Yet, there is a lack of large-scale cohort studies and efficient intelligent tools for precise analysis of LLNM^{16,27,36–38}, with conclusions often lacking qualitative/quantitative explanations^{23,24}.

Here we show the LLNM-Net, a bidirectional attention architecture that integrates multimodal data for preoperative LLNM prediction. As illustrated in Fig. 1a, we employ foreground-optimized segmentation^{39,40} and Central Point Distance Transformation (CPDT)⁴¹ to extract tumor morphology and precise location. Our Thyroid Multimodal Deep Learning (TMDL) transformer⁴² (Fig. 1b) fuses imaging features with clinical reports and demographic data via bidirectional attention exchange^{43–45}. We generate 3D risk heatmaps through diffeomorphic registration and perform attention-based gradient analysis to interpret metastasis mechanisms. Evaluated on 39,451 patients from seven institutions (Fig. 2), LLNM-Net provides quantitative preoperative assessment to guide surgical planning and prognosis management.

Results

Data description

We collected pathological diagnoses, preoperative ultrasound images, radiology reports, and demographic information from a cohort of 39,451 patients (Table 1). Notably, the median age was 43 years, with female patients outnumbering male patients by more than twofold, and approximately 91% of the cohort identified as Han ethnicity. The most represented categories in the Kwak Thyroid Imaging Reporting and Data System (Kwak-TIRADS)⁴⁶ were 4B (48%) and 4C (28%). Patients with thyroid nodules smaller than 10 mm accounted for 73%, and the rate of LLNM-positive patients was 52%. The subtypes collected included 35,804 cases of papillary thyroid carcinoma (PTC), 2845 cases of follicular thyroid carcinoma (FTC), and 802 cases of medullary thyroid carcinoma (MTC). We trained the model using 80% of the 29,615 patients from two hospitals, with the remaining 20% used for model validation. The external test sets comprised 9836 patients from five multicenter sites. More detailed information can be found in Tables S1 and S2.

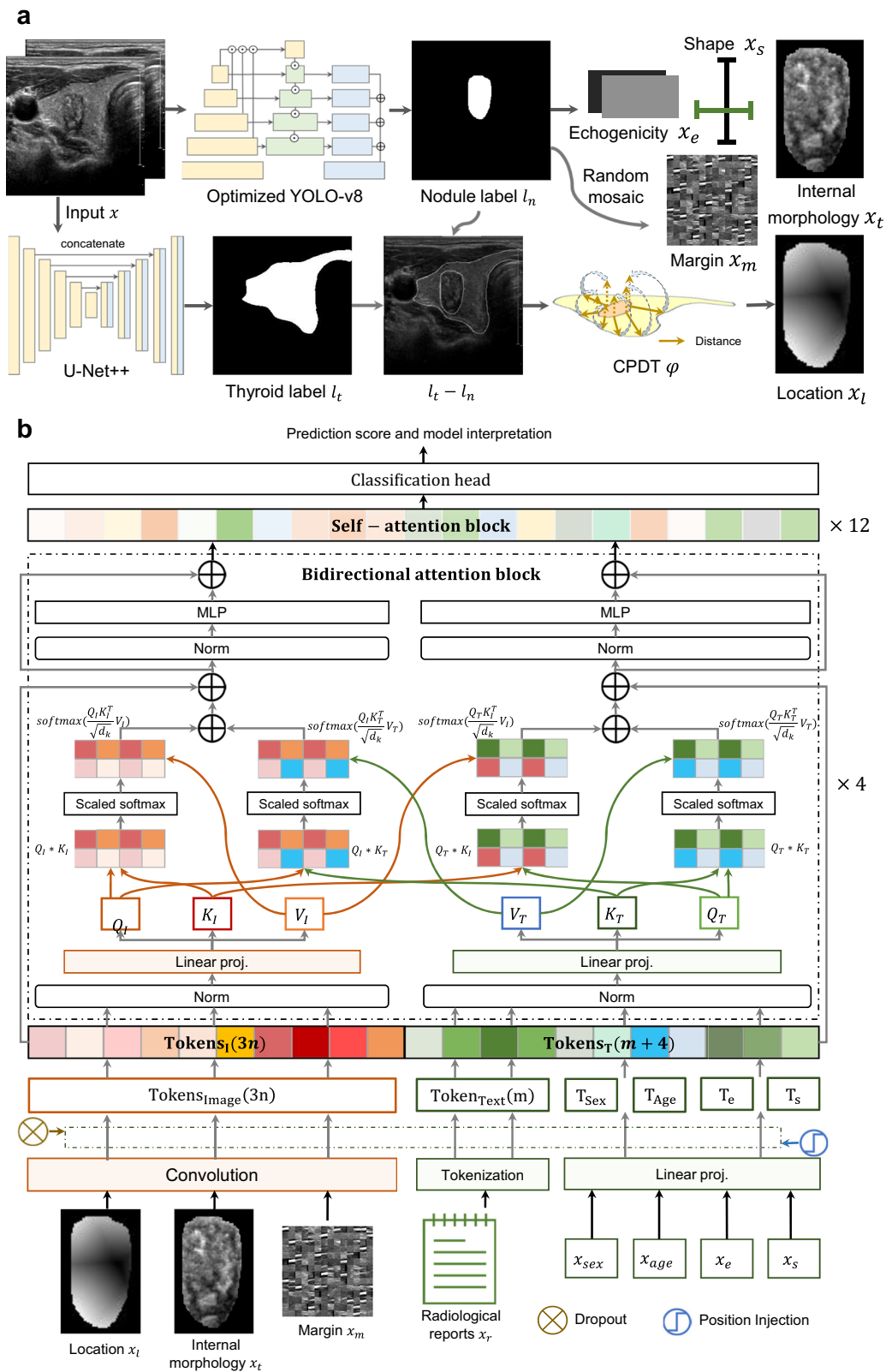
Prediction performance of models and human experts

Figure 3a shows that LLNM-Net exhibits significantly superior predictive performance, with an Area Under the Curve (AUC) of 0.948 (95% CI: 0.946–0.950) on the validation set. Furthermore, the AUC on the external test set is 0.944 (95% CI: 0.942–0.945), with an accuracy of 0.847 (95% CI: 0.840–0.854). In the comparative test, LLNM-Net (accuracy = 0.875) is significantly higher than the predictive accuracy of human experts (accuracy = 0.643, paired t -test, $t = 1.998$, $P = 0.0473$) (Fig. 3a, b). Additionally, the high specificity and PPV demonstrated in the results can more effectively prevent missed diagnoses and enhance the screening performance for LLNM. The accuracy of the segmentation module is presented in Table S4, and the accuracy results for FTC and MTC can be found in Table S5. Comparative experimental results with the latest published AI methods are provided in Table S7.

The study recruited 108 imaging experts, including 42 senior radiologists (with over five years of clinical experience, 21 female and 21 male) and 66 junior radiologists (with three to five years of clinical experience, 33 female and 33 male). These physicians demonstrated high accuracy in diagnosing the malignancy of thyroid nodules (Fig. 3b), with an average accuracy of 0.883, specificity of 0.899, and sensitivity of 0.868. However, in the experiments predicting LLNM classification, both senior and junior physicians exhibited lower predictive capabilities, with an accuracy of 0.643, specificity of 0.642, and sensitivity of 0.644 (Fig. 3b, c). This indicates that the ability to predict LLNM based on traditional clinical knowledge and subjective experience is limited, whereas artificial intelligence models can learn important factors contributing to LLNM from a large number of features.

Qualitative and quantitative analysis for predicting LLNM

Figure 4 provides qualitative and quantitative analyses for predicting LLNM in the locational dimension. Figure 4a displays the visualization results of nodule locational information. We defined the central point



and calculated the minimum distance from all points within the nodule region, obtaining the location feature image through the transformation ϕ . The model learns iteratively and computes the attention heatmap through gradient-weighted calculations⁴⁷, with the heatmap indicating areas identified by the model that have a greater impact on LLNM metastasis. Figure 4b presents a statistical analysis of the minimum distance from the nodule to the thyroid capsule. The results

show that as the minimum distance decreases, the probability of LLNM correspondingly increases. When the minimum distance is less than 0.25 cm, the average probability of LLNM occurrence exceeds 72%.

Figure 4c illustrates the three-dimensional risk heatmap model for LLNM occurrence in thyroid regions, statistically analyzed based on the diffeomorphic affine transformation method^{48,49}. This model is publicly available at: <https://snowinbio.github.io/LLNM-Net/>. The

Fig. 1 | Network architecture of LLNM-Net. **a** Data feature extraction process. For the ultrasound image input x , we used the optimized YOLO-v8 model to segment the nodule label l_n . Subsequently, the U-Net ++ model is utilized to segment the thyroid label l_t . Following this, l_n is combined with x to derive the features including shape x_s , echogenicity x_e , internal morphology x_i , and obtaining margin x_m through random mosaic method. The combination of l_t and l_n yields the merged label $l_t - l_n$, which is processed through the CPDT ϕ , converting distance information into image grayscale values. Thereby we obtained the location information x_l . **b** Overall workflow of TMDL. The input data consists of 8 features, including: morphological features (x_t, x_m, x_e, x_s), locational information x_l ,

radiological reports x_r , demographics (sex x_{sex} and age x_{age}). In the two-layer initial embedding layers, convolutional layers are used to encode the image-type features x_l, x_t and x_m into a sequence of image patch tokens ($Tokens_I$). The encoder encodes $x_{sex}, x_{age}, x_e, x_s$ and unstructured data x_r into $Tokens_T$. The two types of tokens are combined into unified tokens and then input into the bidirectional attention block. This block consists of two normalization layers (Norm), a bidirectional multimodal fusion layer, and a multi-layer perceptron (MLP). The attention between the two types of data is exchanged and computed. This block is stacked into four layers, followed by 12 layers of self-attention blocks.

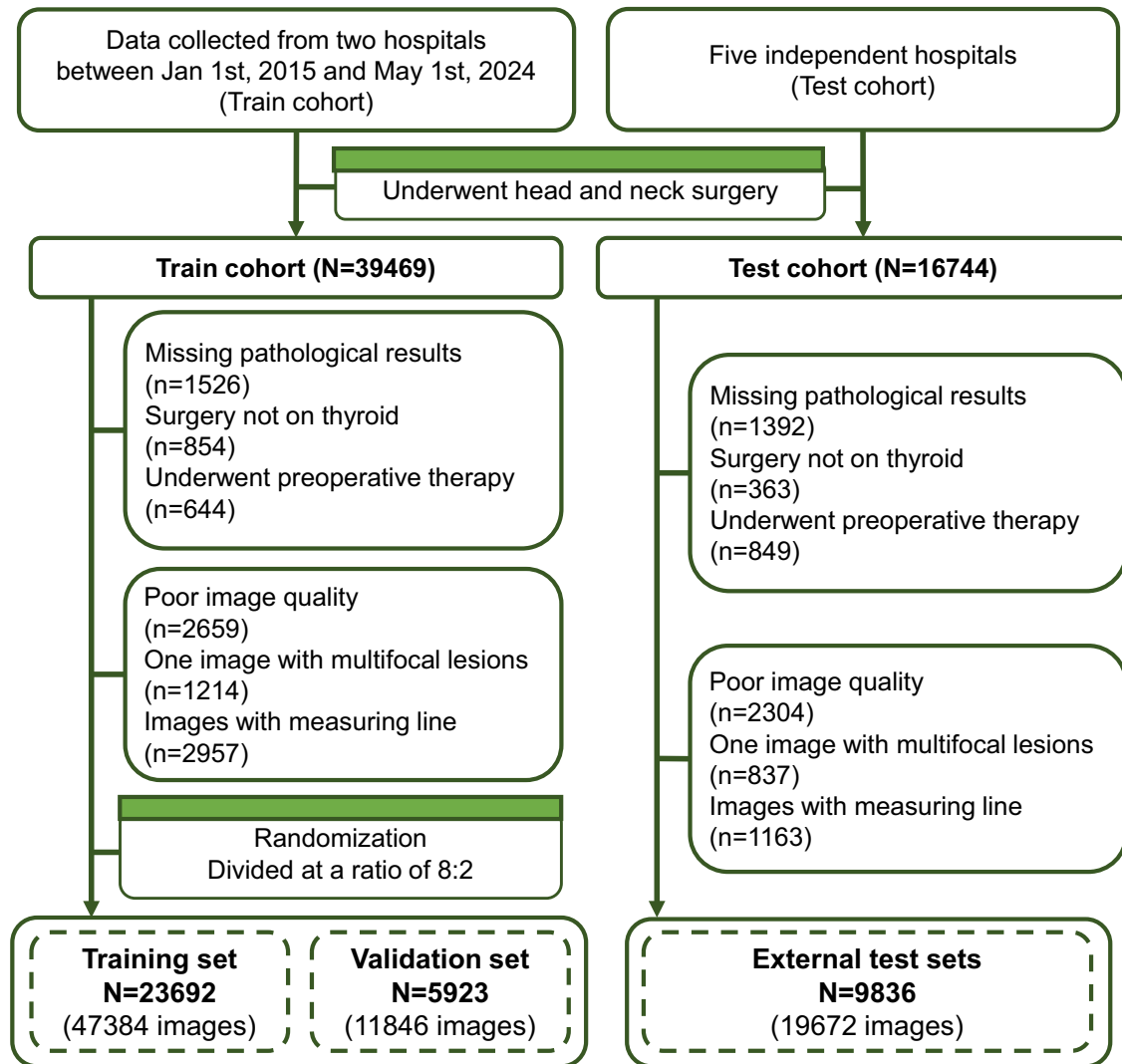


Fig. 2 | Flow diagram summarizing the inclusion of patients. This study consisted of 23,692 patients in the training set, 5923 patients in the validation set, and 9836 patients in external test sets. Poor image quality includes blurred nodular

areas, image jitter, and incomplete imaging of the nodular region. Source data are provided as a Source Data file.

results indicate that the upper central region of the left lobe of the thyroid, as well as the medial and lateral regions of the upper part of the right lobe, are key areas of concern for LLNM occurrence, likely closely related to lymphatic metastasis pathways. Figure 4d showcases an example application of the heatmap model. When provided with the growth location of a thyroid nodule, the model can automatically generate the probability value for LLNM occurrence at that location.

In Fig. 5, we present an example of LLNM-Net. The model takes both imaging data and user information as inputs and automatically outputs a multi-feature score. It also provides the contribution ratio of

each feature, as illustrated by the Sankey diagram on the right. This multi-feature metastatic score helps physicians understand how the model predicts metastasis. To demonstrate that our interpretation aligns with clinical knowledge, we compared it with TIRADS. It can be concluded that there is a correlation between tumor malignancy and LLNM, as shown in Figs. S1, S2.

Figure 6a shows the contribution ratio of all features to LLNM, indicating that the growth location of the thyroid nodule is the most significant factor, accounting for 48.8%. This is followed by morphological features, which account for 29.9%. Among the morphological features, shape and internal morphology

Table 1 | Characteristics of the patients at baseline^a

Characteristic	Training set	Validation set	External test sets	Total
Participants, <i>n</i>	23,692	5923	9836	39,451
Cancer participants, <i>n</i> (%)				
LLNM	12,036 (51)	3158 (53)	5490 (56)	20,684 (52)
Only LLNM	1769 (7)	558 (9)	826 (8)	3153 (8)
CLNM and LLNM	10,267 (43)	2600 (44)	4664 (47)	17,531 (44)
No LLNM	11,656 (49)	2765 (47)	4346 (44)	18,767 (48)
Only CLNM	6695 (28)	1512 (26)	1965 (20)	10,172 (26)
No LNM	4961 (21)	1253 (21)	2381 (24)	8595 (22)
Age, years, median (IQR)	43 (34–54)	43 (35–53)	45 (36–51)	43 (35–52)
Sex, <i>n</i> (%)				
Female	16,773 (71)	4079 (69)	7638 (78)	28,490 (72)
Male	6919 (29)	1844 (31)	2198 (22)	10,961 (28)
Nation, <i>n</i> (%)				
Han nationality	22,491 (95)	5767 (97)	7569 (77)	35,827 (91)
Minority nationality ^b	1201 (5)	156 (3)	2267 (23)	3624 (9)
Kwak TIRADS, <i>n</i> (%)				
1–3	116 (0)	34 (1)	141 (1)	291 (1)
4 (4A–4C)	21,473 (91)	5360 (90)	7648 (78)	34,481 (87)
4A	1285 (5)	294 (5)	3044 (31)	4623 (12)
4B	12,871 (54)	3440 (58)	2661 (27)	18,972 (48)
4C	7317 (31)	1626 (27)	1943 (20)	10,886 (28)
5–6	2103 (9)	529 (9)	2047 (21)	4679 (12)
Metastatic number, <i>n</i> (%)				
0	11,656 (49)	2765 (47)	4346 (44)	18,767 (48)
1–5	10,102 (43)	2535 (43)	4867 (49)	17,504 (44)
6–10	1248 (5)	454 (7)	442 (4)	2144 (6)
>10	686 (3)	169 (3)	181 (2)	1036 (3)
Max tumor size, mm, <i>n</i> (%)				
<10	17,404 (73)	4220 (71)	7079 (72)	28,703 (73)
>10	6288 (37)	1703 (29)	2757 (28)	10,748 (27)
Histological subtypes, <i>n</i> (%)				
Papillary Thyroid Cancer	21,977 (93)	5171 (87)	8656 (88)	35,804 (91)
Follicular Thyroid Cancer	1351 (6)	635 (11)	859 (9)	2845 (7)
Medullary Thyroid Cancer	364 (2)	117 (2)	321 (3)	802 (2)

^aPercentages may not add up to 100 due to rounding.

^bThe ethnic groups listed in the table corresponded to those indicated on the patients' identification documents for the surgical procedure. Among the represented ethnic minorities were Mongolians, Manchus, and Huis.

contribute 7.5% and 64.1%, respectively, suggesting that clinical attention should focus on nodules with abnormal shapes or significant enlargement. Text information from clinical reports contributes 19.7% overall. Gender is an important factor in demographics (accounting for 56.7% in demographics). Figure 6b displays the attention heatmaps for LLNM-negative and LLNM-positive cases, showing that the network focuses on the edges and various internal texture features. Figure S3 illustrates examples of the association between imaging reports and internal morphological features.

The decision curve analysis (Fig. 6c) illustrates the clinical benefit of strategy (the benefit of effective LLND minus the cost of LLND). Under different cost-benefit ratio settings, our model-based strategy outperforms the naïve strategies of performing lateral lymph node (LLN) fine-needle aspiration (FNA) for all or no patients. Figure 6d shows the clinical impact of our model-based strategy within our cohort. At a threshold of 0.5, the strategy improves the identification of 47.4% of patients who should undergo LLN FNA, with only 5.3% of LLNM-negative patients undergoing unnecessary

LLN FNA. Additionally, we conducted a reverse cognitive test to evaluate the interpretability of our model. The results indicate that clinicians' understanding of the model outcomes improved by 25.4% compared to general AI (Fig. S4).

Predicting high-risk lymph node metastasis patients

The model is capable of predicting the stage of lateral lymph node metastasis. Based on clinical guidelines^{3,50–52}, we classified lymph node metastasis into three stages according to the number and size of metastatic nodes:

Stage 1: Low risk. Five or fewer micro-metastases (<0.2 cm in largest dimension).

Stage 2: Medium risk. More than five metastatic lymph nodes, and any metastatic lymph node <3 cm in largest dimension.

Stage 3: High risk. Any metastatic lymph node >3 cm in largest dimension.

Table 2 demonstrates that the model exhibits good predictive performance, achieving an average AUC of 0.971 in external test sets.

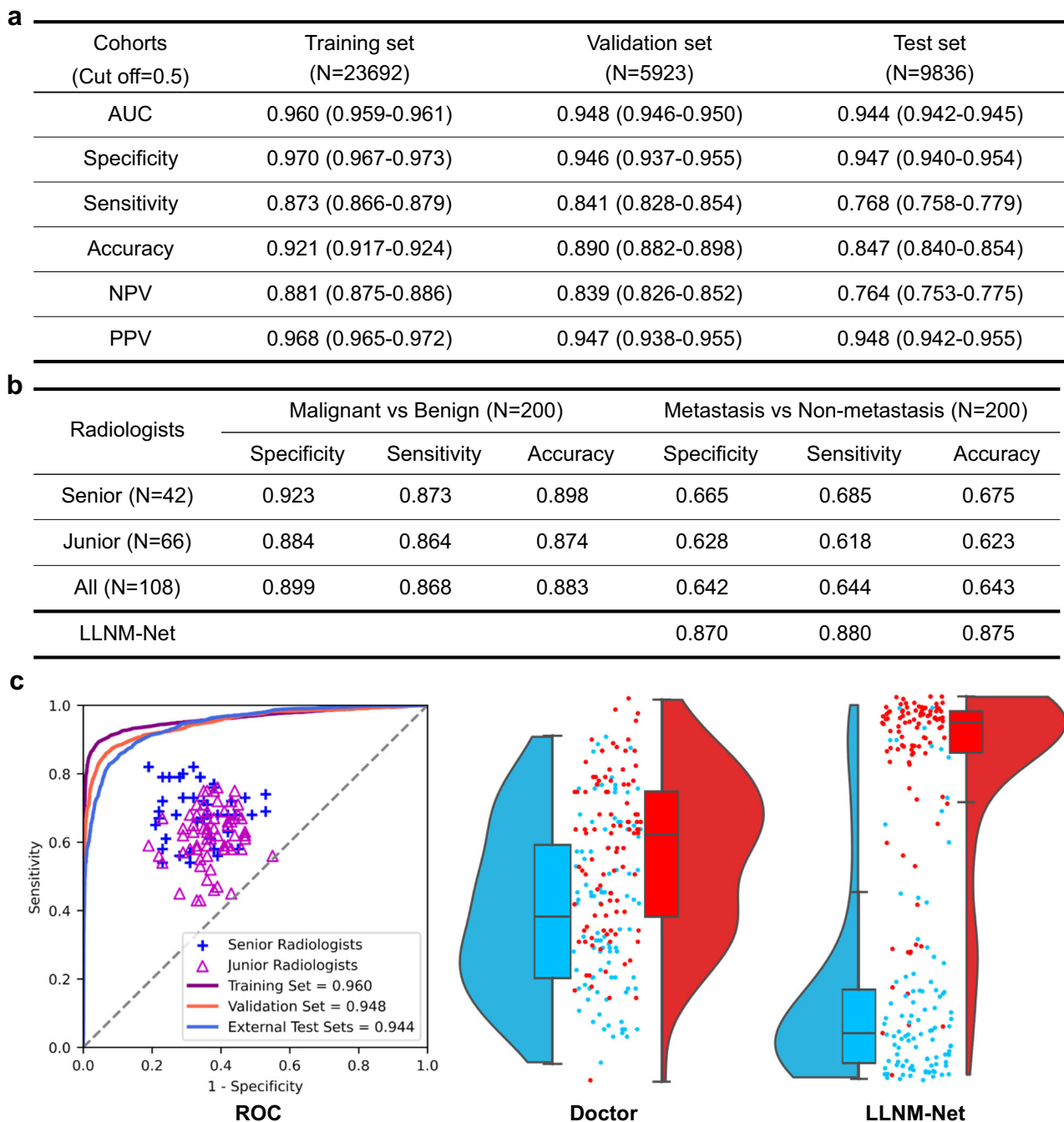


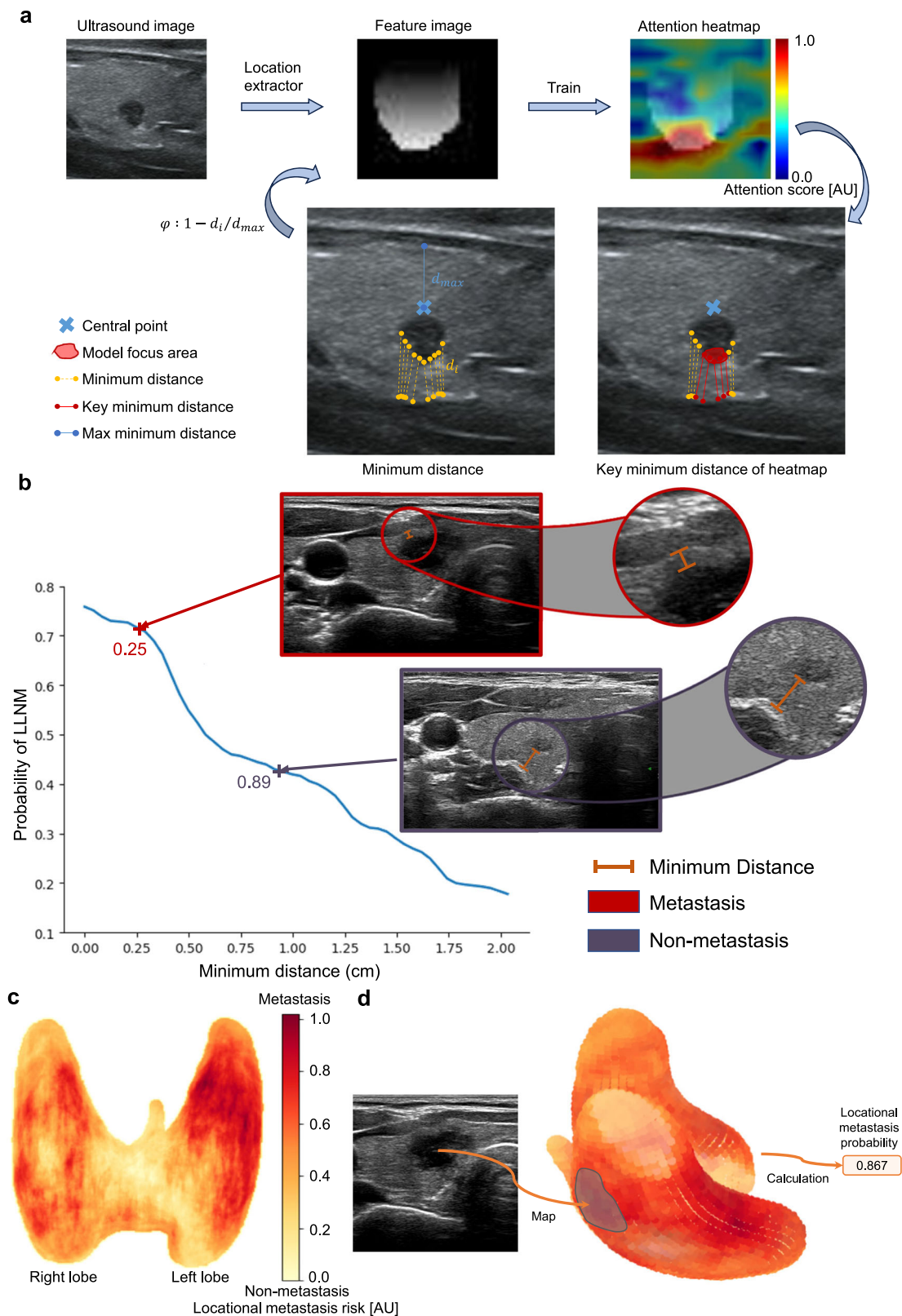
Fig. 3 | Prediction performance of model and human. a Prediction performance of model. The table displays the AUC, specificity, sensitivity, accuracy, negative predictive value (NPV) and positive predictive value (PPV) prediction performance of LLNM-Net for patients across the three datasets. Source data are provided as a Source Data file. **b** Prediction performance of the comparative test between radiologists and LLNM-Net. Radiologists performed well in the malignant classification test but poorly in the LLNM classification test. In contrast, LLNM-Net performed better on the same test dataset. Source data are provided as a Source Data file. **c** ROCs of LLNM-Net on the training set, validation set, and external test sets, as well as the predictive performance of senior and junior radiologists. The AUC results are presented as mean values, and 95% confidence intervals are derived from $n = 100$ experimental replicates for each task setting.

In each replicate trial, real patient input data are selected via bootstrap sampling from the real dataset. We used a two-sample two-sided unadjusted Kolmogorov-Smirnov (KS) test for goodness of fit to examine the predictive distribution values of radiologists and LLNM-Net. Raincloud plots with violin and box diagrams are used to show the comparison of individual-level prediction probabilities between the radiologists (Doctor raincloud plot, mean accuracy of 108 radiologists, $n = 200$) and the LLNM-Net (LLNM-Net raincloud plot, $n = 200$, $KS = 0.385$, $P < 1 \times 10^{-12}$). Each boxplot includes a box representing the median value and interquartile range (IQR). The whiskers extend from the box to the maximum and minimum values, with their length not exceeding 1.5 times the IQR. The red color indicates LLNM-positive samples, while the blue color represents LLNM-negative samples. Source data are provided as a Source Data file.

This suggests that LLNM-Net can precisely identify individuals at medium to high risk, offering physicians reference advice for FNA testing, and ensure regular follow-ups for low-risk individuals, thereby optimizing the efficient allocation of medical resources.

Application of LLNM-Net in clinical practice

Figure 7 illustrates the traditional clinical guidelines³ and guidelines with LLNM-Net. In Fig. 7a, when patients undergo thyroid imaging, physicians assess the malignancy of nodules based on subjective



experience and decide whether FNA is necessary. If needed, a cervical lymph node ultrasound examination is considered to evaluate the likelihood of LLNM and consider whether to perform FNA for the lateral lymph nodes. Then the decision to proceed with thyroidectomy and LLND is made based on the FNA results. However, whether the physician conducts a cervical ultrasound examination depends on subjective experience, leading to potential missed diagnoses.

Incomplete coverage of the detection area during the ultrasound examination can also result in missed diagnoses, affecting the accuracy of FNA results for lymph nodes and influencing the decision to perform LLND.

In Fig. 7b, the integration of LLNM-Net aids in preoperative LLNM risk prediction, helping doctors determine whether to conduct cervical ultrasound examinations, thereby reducing missed diagnoses. During

Fig. 4 | Qualitative and quantitative assessment of nodule locational information. **a** Visualization of nodule locational information. We obtained the location feature image using the CPDT method. Subsequently, we computed the model parameters to generate a heatmap that maps the key minimum distances the model focuses on, which are crucial for assessing the probability of LLNM occurring in relation to the nodule’s location. In the figure, a higher attention score indicates that the model pays more attention to the region. The color bar is labeled as “Attention score [AU]”, ranging from 0.0 (blue) to 1.0 (red). **b** Quantitative assessment of nodule locational information. The curve illustrates that as the minimum distance between the thyroid nodule and the thyroid capsule decreases, the probability of LLNM gradually increases. Two examples are

presented: LLNM-positive cases (red contour) and LLNM-negative cases (purple contour). When the distance is less than 0.25 cm, the probability of LLNM increases by 72%. **c** Qualitative assessment of nodule locational information. This presents risk heatmaps illustrating the likelihood of LLNM occurrence when thyroid nodules are located in different regions. The depth of color represents the locational metastasis risk. When the risk value is 1.00, the risk of metastasis in the locational dimension reaches the maximum. The metastasis risk is at the minimum when the risk value is 0.00. The color bar is labeled as “Locational metastasis risk [AU]”, with values from 0.0 (white) to 1.0 (red). **d** Example of a risk heatmap. Given the nodule’s location, the probability value for LLNM occurrence can be automatically generated.

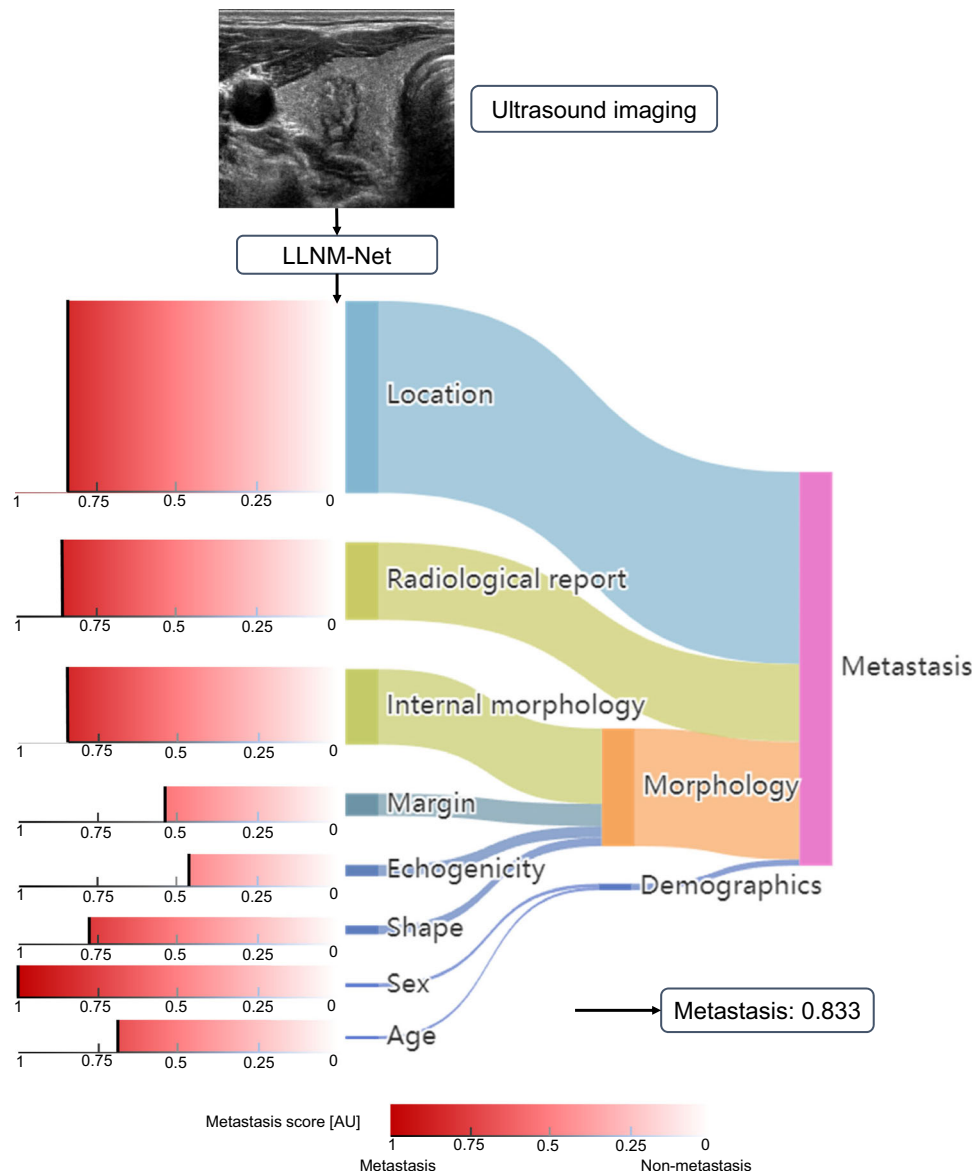
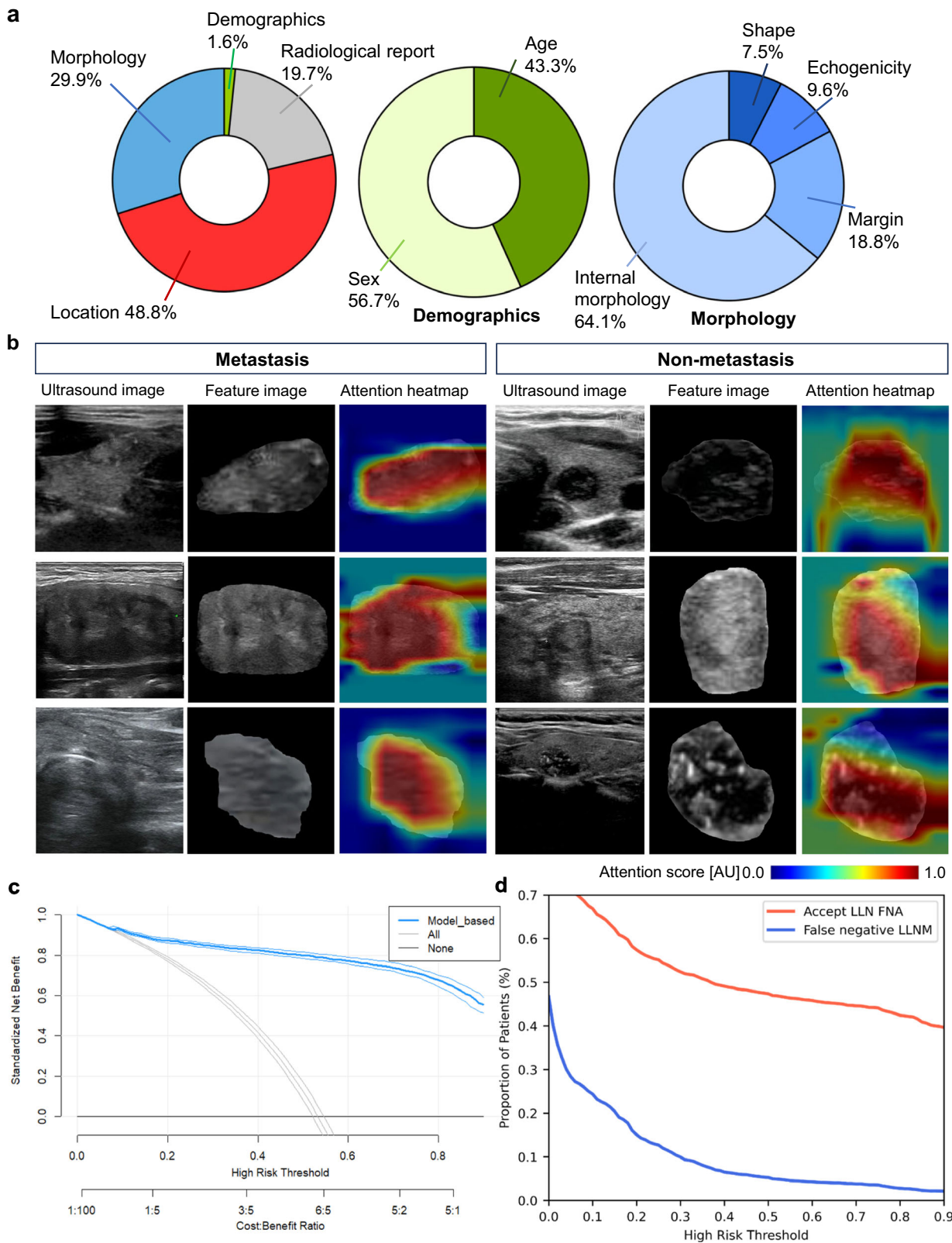


Fig. 5 | An illustrative example of LLNM-Net results. The results output from LLNM-Net. On the left, the prediction scores for each feature response are given, while the Sankey diagram on the right shows the contribution ratio to metastasis. The prediction scores are integrated according to their contribution to finally

obtain the metastasis prediction probability. In this example, the final metastasis score is 0.833, which is LLNM positive, consistent with the actual outcome. The color bar is labeled as “Metastasis score [AU]”, with values from 0.0 (white) to 1.0 (red). Source data are provided as a Source Data file.

cervical ultrasound examinations, LLNM-Net highlights patients at risk of LLNM, decreasing the rate of missed diagnoses. It also recommends FNA for lymph nodes for medium and high risk LLNM patients, reducing missed diagnoses caused by incomplete ultrasound examination areas or physician judgment errors. In summary, LLNM-Net helps improve the diagnostic and treatment process for LLNM.

In this diagnostic study, we developed an interpretable multi-modal deep learning model that can be implemented as an AI support system for LLNM risk assessment based on thyroid ultrasound images. This model provides qualitative and quantitative clinical explanations for predictions based on the fusion transformer method. This study addresses the lack of effective methods for preoperative diagnosis of



LLNM, providing clinical insights while accurately screening high-risk populations, and significantly improving patient survival rates and societal welfare.

Discussion

We developed LLNM-Net, achieving precise preoperative prediction of LLNM and demonstrating strong generalizability across five

external centers, with predictive performance reaching an AUC of 0.944 (95% CI: 0.942–0.945). Currently, ultrasound has a low sensitivity of 0.623 for detecting lateral neck regions. Our model’s AUC improved by 7.4% compared to the best existing model, and its accuracy was 20.4% higher than that of human physicians. In high-risk population screening, the AUC reached 0.971. Furthermore, compared to general AI scores, our model improved clinicians’

Fig. 6 | Attention analysis and decision curve analysis. **a** Contribution analysis of features. The location information of thyroid nodules is the most significant factor influencing LLNM (accounting for 48.8%). Clinical findings accumulated by physicians, hidden in the imaging report, account for 19.7%. The internal morphology feature accounts for 64.1% in morphology, while gender accounts for 56.7% in demographics. Source data are provided as a Source Data file. **b** Visualization of the attention for Internal morphology. The columns display the ultrasound image, the corresponding feature image, and the attention heatmap in LLNM-Net. The hotspots in the heatmap indicate the areas where the model focuses. The higher attention score indicates that the model pays more attention to the area. All nodules shown in the figure are malignant. In the

morphological feature panel, the model focuses more on the overall morphological characteristics of the nodules. This suggests that the internal texture features of the nodules contribute more significantly to the model's prediction of LLNM. The color bar is labeled as "Attention score [AU]", with values from 0.0 (blue) to 1.0 (red). **c** Decision curve analysis, the Standardized net benefit (The benefit of effective LLND and the cost of LLND) of the model-based strategy, the strategy that do LLND for "All patients" and "None patients". Source data are provided as a Source Data file. **d** Clinical impact of the model-based strategy. Red line: The proportion of patients that accept LLN FNA under model-based strategy. Blue line: The proportion of false negative LLNM cases. Source data are provided as a Source Data file.

Table 2 | Performance of multi-level risk of lymph node metastasis^a

Cohorts (Cut off = 0.5)	Training set (N = 13166)	Validation set (N = 3458)	External test sets (N = 4060)
AUC	0.984 (0.982–0.987)	0.994 (0.992–0.996)	0.971 (0.964–0.979)
Specificity	0.962 (0.958–0.967)	0.967 (0.960–0.974)	0.938 (0.930–0.948)
Sensitivity	0.908 (0.899–0.918)	0.903 (0.887–0.920)	0.863 (0.845–0.884)
Accuracy	0.938 (0.932–0.944)	0.945 (0.936–0.955)	0.910 (0.896–0.923)
NPV	0.961 (0.956–0.965)	0.976 (0.971–0.981)	0.953 (0.946–0.961)
PPV	0.904 (0.895–0.914)	0.914 (0.897–0.932)	0.894 (0.878–0.912)

^aThe table displays the average values of multi-classification metrics.

understanding of the results by 25.4%. This significantly addresses the gap in LLNM research, enhances preoperative predictions for more precise treatments, and guides physicians in early patient stratification for closer monitoring and treatment, thereby improving patient quality of life and survival rates. This is crucial for enhancing fairness in AI-based clinical diagnosis.

Our study provides guidance for clinical detection area research, revealing findings: nodules located in the upper middle region of the left lobe and the upper middle region of the right lobe are high-risk areas for LLNM, suggesting that clinicians should focus on these regions. This may be closely related to lymphatic metastasis pathways. Statistics show that when the minimum distance between the nodule and the capsule is less than 0.25 cm, the average probability of LLNM exceeds 72%, indicating a correlation between the minimum distance, growth region, and LLNM. We quantified the risk areas and feature contributions for each patient.

We addressed the challenge of small foregrounds and high intra-class variance in backgrounds in ultrasound images by using an attention-based foreground optimization segmentation network for precise nodule segmentation. To tackle the multi-scale issue in ultrasound images, we proposed a CPDT method to accurately extract the precise location information of thyroid nodules. To efficiently integrate ultrasound images and clinical information, we designed a multimodal deep learning approach based on a bidirectional attention exchange mechanism, extracting local interconnected information between report text and imaging features and uncovering clinicians' latent findings. This end-to-end multimodal feature extraction method can be widely applied for efficient tumor detection, growth location analysis, and nature assessment, with the potential to solve most imaging diagnostic tasks in various diseases.

In traditional diagnostic processes, physicians rely on subjective experience to decide whether to perform cervical ultrasound examinations, leading to potential missed diagnoses. During the cervical ultrasound examination, incomplete coverage of the detection area may also result in missed diagnoses, affecting the accuracy of FNA results for lymph nodes and influencing the decision for LLND. Using LLNM-Net can optimize this process by helping reduce the missed detection rate of cervical ultrasounds and minimizing missed diagnoses caused by incomplete examination areas, thus improving clinical guidelines.

There are still some limitations in our study. First, the number of collected cases and disease types is limited. In the future, we plan to collect more extensive data from a broader population, including more subtypes, countries, and regions, to enhance LLNM-Net's generalizability and applicability. Second, the actual clinical benefits of our model have not yet been validated. We plan to design prospective experiments for validation and explore the model's real-world effectiveness across different institutions and regions.

Methods

Ethics approval

All clinical data, including demographics, operative procedures, pathology, and complications, were retrospectively collected. This study was approved by the local Ethics Committee and the Institutional Review Board (IRB) of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine Hospital, and undertaken according to the Declaration of Helsinki. Informed consent from patients with thyroid cancer and controls was exempted by the IRB because of the retrospective nature of this study.

Data collection

We conducted a retrospective analysis by gathering preoperative thyroid ultrasound images, radiological reports, and clinical information from patients undergoing thyroidectomy. The criteria for patient inclusion in our study were as follows: (a) patients must be 18 years of age or older, (b) they should have undergone thyroid ultrasound examination with clear ultrasound images available, (c) a diagnosis of thyroid malignant nodule following thyroidectomy, (d) patients were required to have undergone central lymph node dissection with a total of at least 5 lymph nodes removed, and (e) there must be a pathologic assessment of FNA for LLNM. Exclusion criteria were as follows: missing pathological reports, surgery not on thyroid, and patients who had received preoperative treatment. To maintain a high standard of image quality, we implemented rigorous control measures, which involved excluding cases with poor image quality, one image with multifocal lesions, and images with measuring lines. Concurrently, as part of our data collection, we gathered extensive patient demographic information, radiological reports, Kwak Thyroid Imaging Reporting and Data Systems (Kwak-TIRADS) grade⁴⁶, postoperative pathology results, and details regarding LLNM. Each patient's data

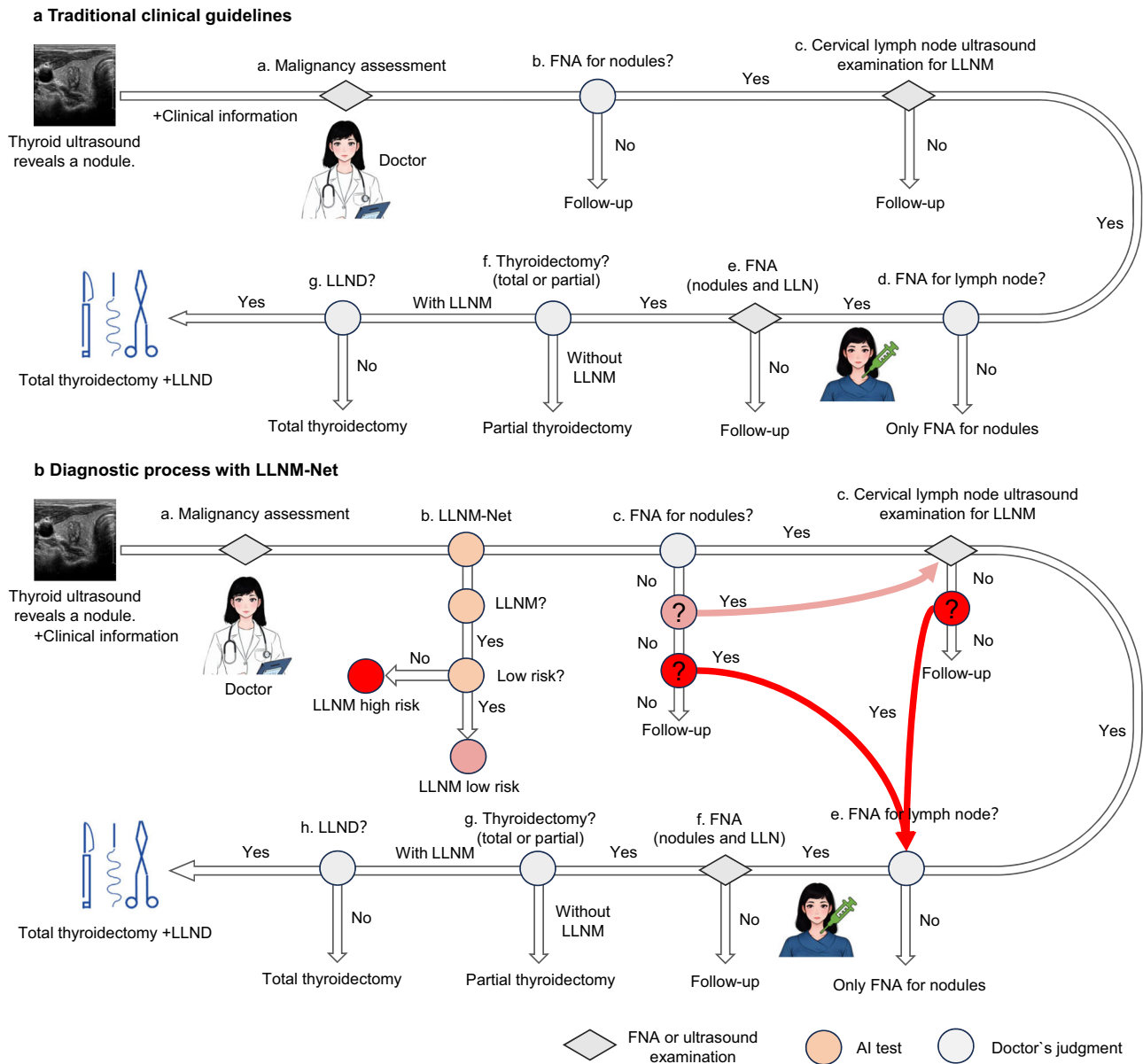


Fig. 7 | Clinical application of LLNM-Net. **a** Traditional clinical guidelines for LLNM. The process primarily involves the following steps: Doctors assess the malignancy of thyroid nodules via ultrasound and clinical information; Determine whether to perform FNA for nodules; Cervical lymph node ultrasound examinations for LLNM; Determine whether to perform FNA for LLN; FNA for lymph nodes

and LLN, among other procedures to assess the necessity of thyroidectomy and LLND. **b** Clinical guidelines for LLNM with LLNM-Net. LLNM-Net assists in pre-operative risk prediction for LLNM, guiding the detection of cervical lymph nodes in low-risk patients and recommending FNA for lymph nodes for medium and high risk patients.

includes two clear ultrasound images from different orientations, a complete ultrasound report, and clinical information. Sex information was collected through self-reporting. However, the primary objective of this study is to predict lateral lymph node metastasis in thyroid cancer, and no differential results were found for sex characteristics, so no further differentiation is made.

We collated patient data from seven hospitals to form training and test cohorts spanning from January 2015 to May 2021. Figure 1 delineates the process of patient inclusion and exclusion. Furthermore, the training cohort from two hospitals was subdivided into a training set and a validation set, while the test cohort from five additional hospitals was designated as external test sets. This methodical strategy guaranteed that our study population was a representative sample of diverse individuals across various geographical and ethnic strata in China. For patient privacy protection, the researchers were granted access solely to anonymized data.

For the classification of patients as LLNM positive, we selected individuals in whom at least one positive lymph node was identified among those excised during surgery. In terms of meticulous quality control for data annotation, we implemented a two-step process:

- (1) Differentiation of malignant nodules: All malignant nodules were diagnosed based on pathological reports. Independent ultrasound physicians with over 5 years of experience were assigned to reassess the images. In cases where discrepancies between their evaluation and the original report were identified, we sought expert judgment to resolve the differences.
- (2) Pathological annotation of nodules with LLNM: In managing patients with multiple nodules, determining which nodule metastasized to the lymph nodes posed a challenge during the annotation process. To address this complexity, three ultrasound radiologists were engaged to meticulously compare the ultrasound images with the corresponding pathological reports

for each patient. Their objective was to select the images of nodules most likely to have metastasized, taking into account factors such as nodule location and degree of malignancy.

Data quality control principles

We obtained preoperative thyroid ultrasound images from seven hospitals. To analyze the ultrasound images, we first removed all patient, institution, and device information from the images. Then we trained an image cropping model to crop images from different institutions and devices, applying a standardized brightness range to achieve uniform images. To preserve the morphology of the nodules (particularly the aspect ratio), we did not use any scaling methods throughout the process. There are two types of clinical text data: unstructured imaging reports (containing the physician's expertise) and structured demographic data (age and gender). We set the maximum length for imaging report data to 50 characters: if the report length exceeded 50 characters, we used only the first 50 characters; otherwise, we applied zero-padding to meet the length requirement.

LLNM-Net architecture

To effectively predict LLNM preoperatively, we have developed the LLNM-Net. This model combines segmentation, distance transformation, and intra-model attention exchange modules to achieve an integrated analysis of the tumor's morphological and locational information. It also incorporates demographic information and clinical reports to provide a comprehensive prediction of LLNM. Additionally, it performs qualitative and quantitative analysis of the metastasis mechanism through attention-based feature analysis (Fig. 1).

Figure 1a shows the process of extracting independent features within the model. Ultrasound images of thyroid nodules present challenges such as small detection targets (foreground) and high intra-class variance in the background. We employed a foreground optimization segmentation network^{39,40} based on an attention mechanism to achieve precise segmentation of thyroid nodules, simultaneously extracting morphological features such as internal morphology, edges, echogenicity differences, and shape. To address the issue of multi-scale input in ultrasound images, we proposed a CPDT method⁴¹ to accurately extract the precise location information of thyroid nodules.

The ultrasound report contains verbal descriptions by medical experts regarding nodule characteristics, such as "normal size and volume," "heterogeneous echogenicity," and "diffuse changes"⁵³⁻⁵⁵. We designed a TMDL transformer⁴² based on a bidirectional attention exchange mechanism⁴³⁻⁴⁵ to efficiently integrate imaging features, report text information, and patient demographic data (Fig. 1b). The TMDL consists of two embedding layers, four bidirectional attention blocks, and twelve self-attention blocks. The embedding layers convert inputs into image and text tokens, which are then processed through the bidirectional attention blocks. In these blocks, attention exchange is used to compute intermodal attention among tokens across different modalities, uncovering potential local interconnections between report text and imaging features, providing advantages over non-integrated models. The computed multimodal representations are then fed into the twelve self-attention blocks for efficient learning.

We conducted qualitative and quantitative analyses of key factors related to LLNM. Using a flexible diffeomorphic registration method⁴⁸, we created a risk heatmap from a three-dimensional perspective showing the likelihood of LLNM occurrence in different thyroid regions. Additionally, we used attention-based gradient-weighted calculations⁴⁷ to analyze the relationship between various features and the prediction outcomes.

Data feature extraction process

For the input image data x , we used an optimized YOLO-v8 model^{39,40} for segmentation, obtaining the nodule label l_n . And we used the U-Net

++ network^{56,57} to obtain the thyroid label l_t . Based on the obtained label l_n and image x , we applied a cropping operation to extract the texture feature x_t , and calculated the length and width to obtain the shape⁵⁸ feature x_s . Besides, we derived x_e by calculating the difference in the mean echo values inside and outside the nodule boundary^{30,59}. Using a random mosaic method⁶⁰, we minimized the influence of nodule morphology on edge blurring, independently extracting the edge feature x_m . Then we obtained the merged label $l_t - l_n$.

We designed the CPDT method to convert positional features. For a point l_i within the nodule region, where $l_i \in l_n$, and the point l_j within the thyroid region, where $l_j \in l_t$, the following equation applies:

$$d_i(l_i, l_t) = \min_{l_j \in l_t, l_t} \left(\sqrt{(p_{l_j} - p_{l_i})^2 + (q_{l_j} - q_{l_i})^2} \right) \quad (1)$$

Where p and q represent the horizontal and vertical coordinates of a point in the image. d_i represents the minimum distance from point l_i to the thyroid capsule.

We have defined the central point C , which is the point within the thyroid region that has the maximum distance to the thyroid capsule. The maximum distance, d_{\max} , from any point inside the thyroid to the capsule is defined as follows:

$$d_{\max}(l_t) = \min_{l_j \in l_t, l_t} \left(\sqrt{(p_{l_j} - p_{l_t})^2 + (q_{l_j} - q_{l_t})^2} \right) \quad (2)$$

When evaluating the risk of metastasis based solely on distance metrics, the point C has the minimum risk of distance-related metastasis. Meanwhile, points on and beyond the thyroid capsule can be considered to have the maximum risk of distance-related metastasis. Therefore, we designed the distance transformation φ to represent the risk of distance-related metastasis for a given point:

$$\varphi(l_t, l_i) = 1 - d_i(l_i, l_t) / d_{\max}(l_t) \quad (3)$$

Through the φ , the grayscale value of each point in l_n is converted to its minimum distance from the thyroid, thereby representing the relative locational information of the nodule and thyroid. Ultimately, we obtained the locational information x_l . By calculating the distance ratio between the nodule region and point C , we can extract information about the different positions of the nodule within the thyroid region. For instance, when the nodule is on the left and right sides of point C , the resulting x_l will be different, even if the distance to the edge of the thyroid is the same.

TMDL module

In practice, we pass multimodal input data (i.e., medical images and clinical text information) to the TMDL module to compute prediction logits, where binary cross-entropy is chosen as the loss function. TMDL is a unified Transformer module. Its structure mainly includes: two initial embedding layers that embed tokens from input images and text respectively; four stacked bidirectional multimodal attention blocks that learn intermediate representations of fused features by capturing interactions between tokens from the same modality and different modalities; 12 stacked self-attention blocks that learn the overall multimodal representation and enhance its discriminative power, and a classification head for generating prediction logits.

In TMDL, the multimodal input data consists of eight components: image data includes location x_l , texture x_t and margin x_m , as well as imaging reports x_r , echogenicity x_e , shape x_s and each patient's gender x_{sex} and age x_{age} . We combine x_l , x_t and x_m and pass them through a convolutional layer, which generates a series of visual tokens. Next, we add standard learnable 1D positional embeddings^{61,62} and dropout to each visual token, resulting in a series of image patch tokens

$Tokens_{Image}(3n)$, where n is the length of a single image patch. At the same time, we use a tokenization encoder to encode each word in x_r . Specifically, we use a pre-trained BERT model⁶² to generate embedding feature vectors for each word in x_r , producing a series of word tokens $Tokens_{Text}(m)$, where m is the maximum length set for the text. We linearly project x_{sex}, x_{age}, x_e and x_s to obtain encoded feature vectors $Tokens_{Sex}, Tokens_{Age}, Tokens_e$ and $Tokens_s$. We then concatenate $\{Tokens_{Text}(m), Tokens_{Sex}, Tokens_{Age}, Tokens_e, Tokens_s\}$ to generate a series of clinical text tokens $Tokens_T(m+4)$. In practice, we set mmm to 50.

The combined tokens are fed into four stacked bidirectional multimodal attention blocks. Assume that the input to the first bidirectional multimodal attention block consists of $Tokens_I^l$ and $Tokens_T^l$, where $l (=0)$ denotes the layer index, $Tokens_I^0 = Tokens_{Image}(3n)$ represents the set of image patch tokens, and $Tokens_T^0 = Tokens_T(m+4)$ represents the set of clinical text tokens. In the bidirectional multimodal attention block, the process of generating the query, key, and value matrices for each modality is as follows:

$$Q_I^l, K_I^l, V_I^l = LP(Norm(Tokens_I^l)) \tag{4}$$

$$Q_T^l, K_T^l, V_T^l = LP(Norm(Tokens_T^l)) \tag{5}$$

Where $LP(\cdot)$ and $Norm(\cdot)$ represent linear projection and layer normalization, respectively. The forward pass within the bidirectional multimodal attention block can be summarized as follows:

$$T_I^l = Attention(Q_I^l, K_I^l, V_I^l) + \alpha Attention(Q_T^l, K_T^l, V_T^l) \tag{6}$$

$$T_T^l = Attention(Q_T^l, K_T^l, V_T^l) + \alpha Attention(Q_I^l, K_I^l, V_I^l) \tag{7}$$

Among them, $Attention(Q_I^l, K_I^l, V_I^l)$ and $Attention(Q_T^l, K_T^l, V_T^l)$ capture intra-modal connections within the image and text modalities, respectively. $Attention(Q_I^l, K_T^l, V_T^l)$ and $Attention(Q_T^l, K_I^l, V_I^l)$ explore inter-modal connections between the image and text. Next, the intra-modal and inter-modal connections are encoded into latent representations T_I^l and T_T^l . After some preliminary experiments, we set α to 1.0. $Attention(Q, K, V)$ consists of two matrix multiplications followed by a scaled *softmax* operation:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}V\right) \tag{8}$$

Here, T denotes the matrix transpose operator, and d_k is a scaling hyperparameter, which we set to 64. We then introduce residual learning and pass the resulting T_I^l, T_T^l to the next normalization layer and MLP:

$$Tokens_I^{l+1} = MLP(Norm(T_I^l)) + Tokens_I^l \tag{9}$$

$$Tokens_T^{l+1} = MLP(Norm(T_T^l)) + Tokens_T^l \tag{10}$$

$Tokens_I^{l+1}$ and $Tokens_T^{l+1}$ are passed as inputs to the next bidirectional multimodal attention block, producing $Tokens_I^{l+2}$ and $Tokens_T^{l+2}$. This operation is repeated until the fourth layer, generating $Tokens_I^{l+4}$ and $Tokens_T^{l+4}$. Then we concatenate the tokens from $Tokens_I^{l+4}$ and $Tokens_T^{l+4}$ to form a unified sequence of tokens, which are passed to the subsequent self-attention blocks. We also allocate 12 multiple heads⁴² in the bidirectional multimodal attention and self-attention blocks. This multi-head

mechanism allows the model to perform attention operations simultaneously across multiple representation subspaces and subsequently aggregate the results.

Finally, we apply average pooling to the unified tokens generated from the last self-attention block to obtain the overall multimodal representation used for predicting LLNM. This representation is passed through a two-layer MLP to produce the final prediction logits. During the training phase, we compute the binary cross-entropy loss between these logits and the lymph node metastasis labels, as given by the following formula:

$$loss = -\frac{1}{N} \sum_{i=1}^N [Y_i \cdot \log(P(Y_i)) + (1 - Y_i) \cdot \log(1 - P(Y_i))] \tag{11}$$

Here, N represents the number of samples in the training set, Y_i denotes the label of a sample, and $P(Y_i)$ is the probability value predicted by the LLNM-Net output. A patient has two sets of imaging data from different directions but shares the same clinical information. Each set of patient data results in a loss value calculation, so there are two loss values per patient. We apply average pooling to these values, taking the mean, and then pass it to the two-layer MLP and loss function.

Model interpretation method

We used a standard attention analysis method for feature analysis. For each layer in LLNM-Net, we computed the average attention weights across multiple heads. Considering the residual connections, we added an identity matrix to each attention matrix and normalized the resulting weight matrix. Next, we recursively multiplied the weight matrices from different layers of LLNM-Net. Finally, we obtained an attention map that includes the similarity between each input token and the CLS token. Since the CLS token is used for diagnostic prediction, these similarities indicate the correlation between the input tokens and the prediction outcome, which can then be used for visualization. We used Grad-CAM++⁴⁷ to visualize the model parameters.

To provide a qualitative interpretation of position, we employed a symmetric diffeomorphism-based algorithm⁴⁸ for registration. Thyroid images are not spatially aligned due to individual variability and factors during image acquisition. Therefore, we needed to map all thyroid data onto a unified standardized template. We defined the registration process as R . This process entailed interacting with an atlas feature matrix referred to as x_l and a target feature matrix indicated by x_l^R , both expressed as functions $x_l, x_l^R : R$. The algorithm posits that the diffeomorphism ϑ is established within the domain of the feature matrix Ω , connecting these feature matrices so that $x_l^R = x_l \cdot \vartheta^{-1}$. The boundary point $\vartheta = \mathfrak{D}_1$ of the curve $\vartheta = \mathfrak{D}_t, t \in [0, 1]$ adheres to the ordinary differential equation (o.d.e.):

$$\frac{d\mathfrak{D}_t}{dt} = v_t(\mathfrak{D}_t), \mathfrak{D}_0 = Fd, t \in [0, 1] \tag{12}$$

In this context, $\vartheta_0 = Fd$ presents the identity transformation, while v_t signifies the time-varying, smooth velocity field, which is defined as $v_t : \Omega \rightarrow R, t \in [0, 1]$. The computation of ϑ is performed as indicated below: $\vartheta = \mathfrak{D}_1 = \int_0^1 v_t(\mathfrak{D}_t) dt$ with $\vartheta_0 = Fd$. Here, we determine the optimal v_t by solving the standard Large Deformation Diffeomorphic Metric Matching (LDDMM)⁶³ equation:

$$\hat{v} = \underset{v: \frac{dv}{dt} = v_t(\vartheta_t)}{\operatorname{argmin}} \left(\int_0^1 \int_{L^2} |Lv_t|^2 dt + \frac{1}{\sigma^2} \int_{L^2} |x_l \cdot \vartheta^{-1} - x_l^R|^2 \right) \tag{13}$$

Where L is the smoothness operator defined by equation: $L = -\alpha \nabla^2 + \gamma x$, where ∇^2 is the Laplacian operator. We used linear interpolation for image transformation. Mutual information served as

the optimization metric during the registration process, and the final evaluation index employed was the mean square error (MSE).

$$MSE = \frac{1}{MN} \sum_{q=1}^M \sum_{p=1}^N [x_i(p, q) - x_i^r(p, q)]^2 \quad (14)$$

where M and N represented the row and column dimension of the matrix, respectively.

We calculated the sum of $x_i^r(Meta)$ for all metastatic patients and subtracted the sum of $x_i^r(Non)$ for all non-metastatic patients. Then we scaled the matrix to the 0–1 range, obtaining the metastasis risk distribution map $Risk_{Meta}$.

$$Risk_{Meta} = Norm \left(\sum x_i^r(Meta) - \sum x_i^r(Non) \right) \quad (15)$$

Model evaluation and radiologist competing test

We evaluated the performance of the predictive model using the AUC of the receiver operating characteristic (ROC) curve, as well as its sensitivity, specificity, accuracy, NPV, and PPV. To compare the predictive effectiveness of AI and human experts for LLNM, we designed an ultrasound physician test experiment. During the recruitment of physicians, a rule of equal representation of male and female experts was followed. Sex information was collected through self-reporting. However, the sex of experts is not used as a variable in the analysis of this study. All participating physicians were required to complete two tasks:

Task 1: Physicians were asked to diagnose thyroid cancer based on 200 ultrasound images, which included 100 benign and 100 malignant nodules.

Task 2: Physicians were required to predict LLNM based on 200 cases using ultrasound images, imaging reports, and clinical information, including 100 LLNM-positive and 100 LLNM-negative cases.

The purpose of Task 1 was to assess the participating physicians' expertise in detecting thyroid lesions on ultrasound images. Task 2 was designed to evaluate the physicians' ability to predict LLNM using a combination of images and clinical information.

Statistical analysis

We estimated the 95% confidence intervals (CI) for the performance metrics pertaining to our classification results using bootstrapping, which encompassed AUC, sensitivity, specificity, accuracy, NPV and PPV. The method we used involved implementing n-out-of-n bootstrap sampling with replacement at the image level for our datasets. For each bootstrap sample (100 samples), we calculated and retained the performance metrics specific to that sample. This process was carried out 1000 times. Subsequently, we established the 95% CIs by taking the 2.5th and 97.5th percentiles from the distribution of each metric's empirical data. All computations and statistical analyses were conducted using Python, version 3.9 (Python Software Foundation).

NPV is the probability that a person testing negative for a disease truly does not have the disease. In other words, it's the percentage of negative results that are correct. The formula for NPV is:

$$NPV = \frac{True\ negative}{True\ negative + False\ negative} \quad (16)$$

PPV is the probability that a person testing positive for a disease truly has the disease. It's the percentage of positive results that are correct. The formula for PPV is:

$$PPV = \frac{True\ positive}{True\ positive + False\ positive} \quad (17)$$

Sensitivity (also referred to as the true positive rate and the recall) is the proportion of positives that are correctly identified as follows:

$$Sensitivity/TPR = \frac{True\ positive}{True\ positive + False\ negative} \quad (18)$$

Specificity (also known as the true negative rate), which measures the proportion of correctly identified negatives, was calculated as follows:

$$Specificity/TNR = \frac{True\ negative}{True\ negative + False\ positive} \quad (19)$$

AUC, standing for Area under the ROC Curve, measuring the entire two-dimensional area underneath the entire ROC curve (think integral calculus) from (0,0) to (1,1), was calculated as:

$$A = \int_0^1 TPR(FPR^{-1}(x))dx \quad (20)$$

Where $FPR = 1 - TNR = 1 - Specificity$. Given two bounding boxes b_1 and b_2 , their IoU could be computed as:

$$IOU(b_1, b_2) = \frac{A(b_1 \cap b_2)}{A(b_1 \cup b_2)} = \frac{A(b_1 \cap b_2)}{A(b_1) + A(b_2) - A(b_1 \cap b_2)} \quad (21)$$

Where $A(\cdot)$ was the area of the shape. The calculation of IoU could therefore be formulated as a problem involving the computation of the area of each spherical rectangle and the intersection of two spherical rectangles.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the research findings in this study are available in the article and its supplementary information files. The minimum dataset required to interpret, verify, and extend the study results on patients with lateral lymph node metastasis has been deposited in Hugging Face with the access link: https://huggingface.co/datasets/Snowinbio/LLNM_Multimodal_dataset. This includes: (1) Pre-processed cropped imaging data (ultrasound images with anonymized meta-data). (2) Corresponding ultrasound imaging reports for patients, including professional descriptions of the images by physicians, characteristics of nodules, etc. (3) Clinical characteristic information of patients, including age and sex. The source data file containing detailed LLNM-Net outputs and key evaluation metrics can also be obtained via the following link: <https://github.com/Snowinbio/LLNM-Net/blob/main/Source%20Data.xlsx>. Due to ethical restrictions and patient confidentiality agreements, the full dataset (such as raw imaging data, detailed imaging reports, and patient clinical records) cannot be publicly available. This is because even after de-identification, detailed patient clinical records and high-resolution imaging data may still pose a risk of re-identification due to the unique characteristics of thyroid cancer cases. Researchers wishing to access additional data for non-commercial academic purposes may submit a formal application to the corresponding author. Applications will be reviewed by the institutional ethics committee and data custodians. The applicable conditions are as follows: (1) Purpose: The data may only be used for research purposes consistent with the original study objectives. (2) Access restrictions: Requestors must sign a data use agreement prohibiting re-identification or redistribution. (3) Data retention: Approved data will be available for 2 years from the date of publication. The data for each figure in this study are included in the "Source

Data” section, with the file name Source Data.xlsx. This file can also be downloaded from the following link: <https://github.com/Snowinbio/LLNM-Net/blob/main/Source%20Data.xlsx>. Source data are provided in this article. Source data are provided with this paper.

Code availability

The primary code for the project is accessible at: <https://github.com/Snowinbio/LLNM-Net.git>. Installation instructions are provided in the repository. We have provided a permanent reference⁶⁴ for the specific code version used in this study. The code is released under the Apache License 2.0, which permits free use, modification, and redistribution under its terms. The model implementation is built upon multiple publicly available open-source projects. We have retained all original license information and copyright notices in the corresponding source files. Specifically, we acknowledge the following contributions: IRENE (Apache 2.0): <https://github.com/RL4M/IRENE>, YOLOv8 by Ultralytics (AGPL-3.0): <https://huggingface.co/Ultralytics/YOLOv8>.

References

- Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **74**, 229–263 (2024).
- Miranda-Filho, A. et al. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. *Lancet Diab Endocrinol.* **9**, 225–234 (2021).
- Haugen, B. R. 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* **26**, 1–133 (2016).
- Zaydfudim, V., Feurer, I. D., Griffin, M. R. & Phay, J. E. The impact of lymph node involvement on survival in patients with papillary and follicular thyroid carcinoma. *Surgery* **144**, 1070–1077 (2008).
- Smith, V. A., Sessions, R. B. & Lentsch, E. J. Cervical lymph node metastasis and papillary thyroid carcinoma: does the compartment involved affect survival? Experience from the SEER database. *J. Surg. Oncol.* **106**, 357–362 (2012).
- Yu, J. et al. Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat. Commun.* **11**, 1–10 (2020).
- Yao, S. et al. Thyroid cancer central lymph node metastasis risk stratification based on homogeneous positioning deep learning. *Research* **7**, 0432 (2024).
- Wang, Y. et al. Risk factors and a prediction model of lateral lymph node metastasis in CNO papillary thyroid carcinoma patients with 1–2 central lymph node metastases. *Front. Endocrinol.* **12**, 716728 (2021).
- Lin Y., Cui N., Li F., Wang Y., Wang B. The model for predicting the central lymph node metastasis in cNO papillary thyroid microcarcinoma with Hashimoto’s thyroiditis. *Front. Endocrinol.* **15**, 1330896 (2024).
- Machens, A., Hauptmann, S. & Dralle, H. Prediction of lateral lymph node metastases in medullary thyroid cancer. *Br. J. Surg.* **95**, 586–591 (2008).
- Mohebati, A. & Shaha, A. R. Anatomy of thyroid and parathyroid glands and neurovascular relations. *Clin. Anat.* **25**, 19–31 (2012).
- Greene, F. L. et al. *AJCC cancer staging handbook: TNM classification of malignant tumors*. (Springer Science & Business Media, 2002).
- Liddy, W., Bonilla-Velez, J., Triponez, F., Kamani, D. & Randolph G. 31 - Principles in Thyroid Surgery. In: *Surgery of the Thyroid and Parathyroid Glands (Third Edition)* (ed Randolph GW). (Elsevier, 2021).
- Ito, Y. et al. Preoperative ultrasonographic examination for lymph node metastasis: usefulness when designing lymph node dissection for papillary microcarcinoma of the thyroid. *World J. Surg.* **28**, 498–501 (2004).
- de Meer, S. G. A. et al. Not the number but the location of lymph nodes matters for recurrence rate and disease-free survival in patients with differentiated thyroid cancer. *World J. Surg.* **36**, 1 (2012).
- Ruan, J. et al. Lateral lymph node metastasis in papillary thyroid microcarcinoma: a study of 5241 follow-up patients. *Endocrine* **83**, 414–421 (2024).
- Hu, D. et al. Risk factors for and prediction model of skip metastasis to lateral lymph nodes in papillary thyroid carcinoma. *World J. Surg.* **44**, 1498–1505 (2020).
- Feng, J.-W. et al. Predictive factors for lateral lymph node metastasis and skip metastasis in papillary thyroid carcinoma. *Endocr. Pathol.* **31**, 67–76 (2020).
- Wu, X., Li, B., Zheng, C. & He, X. Risk factors for skip metastasis in patients with papillary thyroid microcarcinoma. *Cancer Med.* **12**, 7560–7566 (2023).
- Liu, J., Liu, Q., Wang, Y., Xia, Z. & Zhao, G. Nodal skip metastasis is associated with a relatively poor prognosis in thoracic esophageal squamous cell carcinoma. *Eur. J. Surg. Oncol.* **42**, 1202–1205 (2016).
- Prenzel, K. L. et al. Skip metastasis in nonsmall cell lung carcinoma. *Cancer* **100**, 1909–1917 (2004).
- Tee, Y. Y., Lowe, A. J. & Brand, C. A., Judson, R. T. Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: a comprehensive literature review. *Ann. Surg.* **246**, 714–720 (2007).
- Qubain, S. W., Nakano, S., Baba, M., Takao, S. & Aikou, T. Distribution of lymph node micrometastasis in pNO well-differentiated thyroid carcinoma. *Surgery* **131**, 249–256 (2002).
- Back, K., Kim, J. S., Kim, J.-H. & Choe, J.-H. Superior located papillary thyroid microcarcinoma is a risk factor for lateral lymph node metastasis. *Ann. Surg. Oncol.* **26**, 3992–4001 (2019).
- Ywata de Carvalho, A., Kohler, H. F., Gomes, C. C., Vartanian, J. G. & Kowalski, L. P. Predictive factors for recurrence of papillary thyroid carcinoma: analysis of 4,085 patients. *Acta Otorhinolaryngol. Ital.* **41**, 236–242 (2021).
- Liu, J. et al. Follicular variant of papillary thyroid carcinoma. *Cancer* **107**, 1255–1264 (2006).
- Sheng, L. et al. Predicting factors for central or lateral lymph node metastasis in conventional papillary thyroid microcarcinoma. *Am. J. Surg.* **220**, 334–340 (2020).
- Huang, J. et al. Developing and validating a multivariable machine learning model for the preoperative prediction of lateral lymph node metastasis of papillary thyroid cancer. *Gland Surg.* **12**, 101–109 (2023).
- Yao, J. et al. DeepThy-Net: a multimodal deep learning method for predicting cervical lymph node metastasis in papillary thyroid cancer. *Adv. Intell. Syst.* **4**, 2200100 (2022).
- Yao, S. et al. Human understandable thyroid ultrasound imaging AI report system—a bridge between AI and clinicians. *iScience* **26**, 106530 (2023).
- Yao, S. et al. Enhancing the fairness of AI prediction models by Quasi-Pareto improvement among heterogeneous thyroid nodule population. *Nat. Commun.* **15**, 1958 (2024).
- Ren, Y. et al. BMAP: a comprehensive and reproducible biomedical data analysis platform. *bioRxiv*, 2024.2007.2015.603507 (2024).
- Deng, L., Wu, Y., Ren, Y., Lu, H. Autonomous self-evolving research on biomedical data: the DREAM paradigm. *Adv. Sci.* **12**, 2417066 (2025).
- Ha, E. J. et al. Artificial intelligence model assisting thyroid nodule diagnosis and management: a multicenter diagnostic study. *J. Clin. Endocrinol. Metab.* **109**, 527–535 (2024).

35. Zhou, L.-Q. et al. Deep learning predicts cervical lymph node metastasis in clinically node-negative papillary thyroid carcinoma. *Insights Imaging* **14**, 222 (2023).
36. Tong, Y. et al. Ultrasound-based radiomic nomogram for predicting lateral cervical lymph node metastasis in papillary thyroid carcinoma. *Acad. Radiol.* **28**, 1675–1684 (2021).
37. Xing, Z. et al. Thyroid cancer neck lymph nodes metastasis: Meta-analysis of US and CT diagnosis. *Eur. J. Radiol.* **129**, 109103 (2020).
38. Dai, F. et al. Improving AI models for rare thyroid cancer subtype by text guided diffusion models. *Nat. Commun.* **16**, 4449 (2025).
39. Zheng, Z., Zhong, Y., Wang, J., Ma, A. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020).
40. Redmon, J. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2016).
41. Borgefors, G. Distance transformations in digital images. *Comput. Vis. Graph. Image Process.* **34**, 344–371 (1986).
42. Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems* (NeurIPS, 2017).
43. Zhou, H.-Y. et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat. Biomed. Eng.* **7**, 743–755 (2023).
44. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. Association for Computational Linguistics (2019).
45. Kolesnikov, A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: *International Conference on Learning Representations* (2021).
46. Kwak, J. Y. et al. Thyroid imaging reporting and data system for us features of nodules: a step in establishing better stratification of cancer risk. *Radiology* **260**, 892–899 (2011).
47. Selvaraju, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization (2017).
48. Avants, B. B., Tustison, N. & Song, G. Advanced normalization tools (ANTS). *Insight J.* **2**, 1–35 (2009).
49. Tustison, N. J. et al. The ANTsX ecosystem for quantitative biological and medical imaging. *Sci. Rep.* **11**, 1–13 (2021).
50. Ito, Y. et al. Prognosis of patients with papillary thyroid carcinoma having clinically apparent metastasis to the lateral compartment. *Endocr. J.* **56**, 759–766 (2009).
51. Zaydfudim, V., Feurer, I. D., Griffin, M. R. & Phay, J. E. The impact of lymph node involvement on survival in patients with papillary and follicular thyroid carcinoma. *Surgery* **144**, 1070–1078 (2008).
52. Ricarte-Filho, J. et al. Papillary thyroid carcinomas with cervical lymph node metastases can be stratified into clinically relevant prognostic categories using oncogenic BRAF, the number of nodal metastases, and extra-nodal extension. *Thyroid* **22**, 575–584 (2012).
53. Ramundo, V. et al. Is thyroid nodule location associated with malignancy risk? *Ultrasonography* **38**, 231–235 (2019).
54. Zhang, F. et al. Thyroid nodule location on ultrasonography as a predictor of malignancy. *Endocr. Pract.* **25**, 131–137 (2019).
55. Jasim, S., Baranski, T. J., Teefey, S. A. & Middleton, W. D. Investigating the effect of thyroid nodule location on the risk of thyroid cancer. *Thyroid* **30**, 401–407 (2020).
56. Ronneberger, O., Fischer, P., Brox, T. U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (eds Navab N., Hornegger J., Wells W. M., Frangi A. F.) (2015).
57. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. & Liang, J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **39**, 1856–1867 (2020).
58. Alexander, E. K. et al. Thyroid nodule shape and prediction of malignancy. *Thyroid* **14**, 953–958 (2004).
59. Xie, C., Cox, P., Taylor, N. & LaPorte, S. Ultrasonography of thyroid nodules: a pictorial review. *Insights Imaging* **7**, 77–86 (2016).
60. Xu, P., Ding, J., Zhang, H. & Huang, H. Discernible image mosaic with edge-aware adaptive tiles. *Comput. Vis. Media* **5**, 45–58 (2019).
61. Zhang, K. et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat. Biomed. Eng.* **5**, 533–545 (2021).
62. Akselrod-Ballin, A. et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology* **292**, 331–342 (2019).
63. Beg, M. F., Miller, M. I., Trounev, A. & Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* **61**, 139–157 (2005).
64. Shen, P. Explainable multimodal deep learning for predicting thyroid cancer lateral lymph node metastasis using ultrasound imaging: LLNM-Net. v1.0.0 edn. Zenodo (2025).

Acknowledgements

Funding. This work was partially supported by the National Natural Science Foundation of China (grant No. 62406191), Shanghai Municipal Education Commission (No.2024AIYB010), the Fundamental Research Funds for the Central Universities (YG2025LC03), the Interdisciplinary Program of Shanghai Jiao Tong University (YG2024QNA02), the National Natural Science Foundation of China (82403192) and the Shanghai Anticancer Association (SACA-CY23C07), the Startup for Young Faculty (SFYF) at SJTU(24×010500175), the Science and Technology Commission of Shanghai Municipality (STCSM) (Grant No. 23JS1400800&23JS1400700), SJTU Transmed Awards Research (STAR) Grant No. 20210106 (HL) and Neil Shen's SJTU Medical Research Fund. The content of this article does not reflect the view of the funding sources.

Author contributions

P.S., S.Y., H.L. and W.C. developed the concept for the manuscript. P.S. and S.Y. contributed to drafting of the manuscript. P.S. and S.Y. designed the model and analysis the data. P.S. and Z.Y. contributed to data cleaning. W.C., S.L., Z.Y., J.S., Yun Wang, Q.C., Yirou Wang and Y.R. contributed to providing medical data and clinical advice.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62042-z>.

Correspondence and requests for materials should be addressed to Sheng Liu, Wei Cai, Hui Lu or Siqiong Yao.

Peer review information *Nature Communications* thanks Qicheng Lao and Nikita Pozdeyev for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025