Research

Nomogram for predicting cervical lymph node metastasis of papillary thyroid carcinoma using deep learning-based super-resolution ultrasound image

Xia Li[1] · Yu Zhao[1] · Wenhui Chen[2] · Xu Huang[1] · Yan Ding[1] · Shuangyi Cao[1] · Chujun Wang[1] · Chunquan Zhang[1]

Abstract

Objectives To investigate the feasibility and effectiveness of a deep learning (DL) super-resolution (SR) ultrasound image reconstruction model for predicting cervical lymph node status in patients with papillary thyroid carcinoma(PTC).

Methods In this retrospective study, researchers recruited 544 patients with PTC and randomly assigned them to training and test sets. SR ultrasound images were acquired using SR technology to improve image resolution, and artificial features and DL features were extracted from the original (OR) and SR images, respectively, to construct a ML, DL model. The best model was selected and aggregated with clinical parameters to construct the nomogram. The performance of the model is evaluated by ROC curves, calibration curves and decision curves.

Results In distinguishing the presence or absence of metastatic lymph nodes, the predictive performance of the SR_ResNet 101 and SR_SVM models based on SR outperformed those based on OR. In the test set, SR_SVM AUC was 0.878 (95% CI 0.8203–0.9358), accuracy 0.854, while OR_SVM AUC was 0.822 (95% CI 0.7500–0.8937), accuracy 0.665. SR_ResNet 101 AUC was 0.799 (95% CI 0.7175–0.8806), accuracy 0.793, and OR_ResNet101 AUC was 0.751 (95% CI 0.6620–0.8401), accuracy 0.713. Subsequently, Nomogram_A and Nomogram_B were constructed by integrating the SR_SVM model and SR_ResNet 101 model, respectively, with clinical parameters, while Nomogram_C was constructed solely based on clinical indicators. In the test set, Nomogram_A demonstrated the best performance with an AUC of 0.930 (95% CI 0.8913–0.9682) and accuracy was 0.829. Nomogram_B AUC 0.868 (95% CI 0.8102–0.9261) and accuracy was 0.829, while Nomogram_C AUC 0.880 (95% CI 0.8257–0.9349) and accuracy was 0.787. The DeLong test revealed that the diagnostic performance of Nomogram_A based on SR_SVM was significantly higher than that of Nomogram_B, Nomogram_C, and the level of Radiologist ($P < 0.05$). The calibration curves and Hosmer–Lemeshow tests confirmed a high degree of fit, and the decision curve analysis demonstrated clinical value and potential patient benefit.

Conclusions The predictive model constructed using SR reconstructed ultrasound images demonstrated superior performance in predicting preoperative cervical lymph node metastasis in PTC compared to OR images. The nomogram

---

Xia Li, Yu Zhao, Wenhui Chen and Xu Huang contributed equally to this work.

✉ Chunquan Zhang, jxzcq@163.com; Xia Li, lixiajxscs@163.com; Yu Zhao, yu_zhao1993@163.com; Wenhui Chen, chenwenhui1993@163.com; Xu Huang, hxhuangxu1999@163.com; Yan Ding, sqdqsh@126.com; Shuangyi Cao, csy18231213@163.com; Chujun Wang, jzcc95915@163.com | [1]Department of Ultrasound, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, No. 1, Minde Road, Donghu District, Nanchang 330006, Jiangxi, People's Republic of China. [2]Department of Hepatobiliary and Pancreatic Surgery, Ganzhou Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China.

Discover

prediction model based on SR images has the potential to enhance the accuracy of predictive models and aid in clinical decision-making.

1 Introduction

Thyroid cancer (TC) has become one of the fastest growing malignant tumors in the world in recent years, ranking among the top ten malignant tumors, especially in women, its incidence rate is among the top five. Papillary thyroid carcinoma (PTC) is the most common pathologic type [1, 2]. Almost 90.0% of cancer-related deaths are caused by metastasis [3], with cervical lymph node metastasis (CLNM) being one of the most important factors affecting the prognosis of patients with TC.

Ultrasound (US) is currently the most commonly used imaging modality for preoperative CLNM assessment of PTC. However, US has a low sensitivity for detecting CLNM [4], with studies claiming that 60.0% to 70.0% of CLNM cases remain undetected by preoperative imaging [5]. US images are often limited in spatial resolution by speckle noise, low contrast, attenuation, shadowing, and signal loss, which can make an accurate diagnosis and treatment planning difficult [6]. In addition, US diagnosis is highly operator-dependent, subtle changes are easily missed, and the quality of the US equipment used also affects the diagnosis. Therefore, accurate diagnosis of PTC CLNM remains challenge.

Recently, intelligent image processing technology based on deep learning (DL) has made significant progress, aiming to surpass the physical limitations of imaging systems, improve the spatial resolution, and enhance the reliability and efficiency of early diagnosis [7]. Super-resolution (SR) reconstruction techniques have attracted significant attention in the field of image generation, primarily encompassing traditional and DL-based methods, the latter utilizing DL models to learn the mapping between low and high resolutions, which has yielded superior reconstruction quality and precision. SR reconstruction can be used to improve the quality of images obtained from various modality such as computed tomography (CT), magnetic resonance imaging (MRI), and US, thereby enhancing the robustness and stability of radiology models [8, 9]. Promising results have been achieved in enhancing the spatial resolution of medical images [6, 10–12].
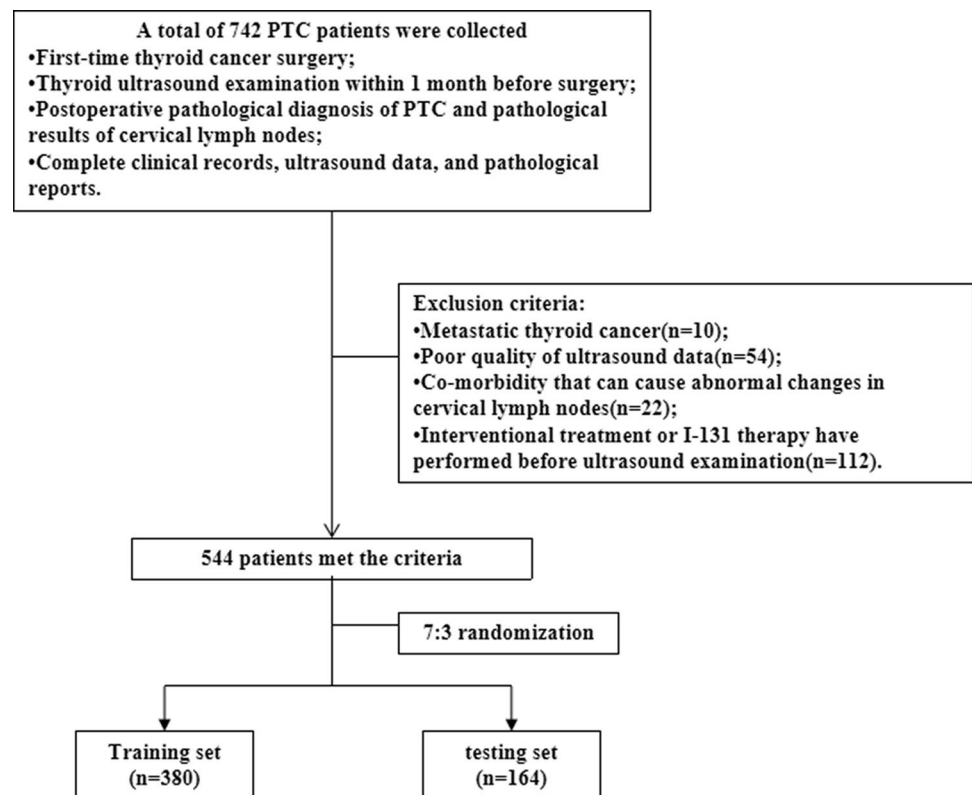
Therefore, we propose to adopt generative adversarial networks (GAN) as the basic framework for SR US images. Based on the original (OR) and SR images, we will construct models using various machine learning (ML) and DL algorithms, and develop the optimal model in combination with clinical indicators to predict the lymph node (LN) status of PTC. To our knowledge, this study is the first to clinically demonstrate the use of DL-based SR reconstruction of US images for predicting LN status in patients with PTC.

2 Materials and methods

2.1 Patient

This research protocol was approved by the Review Committee of the Second Affiliated Hospital of Nanchang University, adhering to the principles outlined in the Declaration of Helsinki and its subsequent amendments or comparable ethical guidelines. The study abstained from utilizing protected health information and ensured the de-identification of all data, thereby negating the requirement for patient informed consent. The inclusion criteria were as follows: (1) first-time TC surgery; (2) thyroid US examination within 1 month before surgery; (3) postoperative pathological diagnosis of PTC and pathological results of cervical lymph nodes (CLN); (4) complete clinical records, US data, and pathological reports. Exclusion criteria: (1) metastatic TC; (3) poor quality of US data; (4) co-morbidities that can cause abnormal changes in CLN; (5) interventional treatment or I-131 therapy have performed before US examination. In this study, 544 patients with PTC who met the criteria in our thyroid surgery department from December 2018 to January 2021 were finally randomly selected for inclusion in the study. All patients were randomly assigned to training and test cohorts according to the proportion of 7:3 (Fig. 1).

Fig. 1 Flow chart of patients enrolled in the study

A total of 742 PTC patients were collected
•First-time thyroid cancer surgery;
•Thyroid ultrasound examination within 1 month before surgery;
•Postoperative pathological diagnosis of PTC and pathological results of cervical lymph nodes;
•Complete clinical records, ultrasound data, and pathological reports.

Exclusion criteria:
•Metastatic thyroid cancer(n=10);
•Poor quality of ultrasound data(n=54);
•Co-morbidity that can cause abnormal changes in cervical lymph nodes(n=22);
•Interventional treatment or I-131 therapy have performed before ultrasound examination(n=112).

544 patients met the criteria

7:3 randomization

Training set (n=380)

testing set (n=164)

2.2 Ultrasonic image data acquisition

US examination and image acquisition were performed according to the guidelines by a experienced sonographer with 10 years of practice. The "Thyroid" mode on the Mindray US Diagnostics (Resona 7S; Shenzhen, China) was utilized for thyroid ultrasonography, following a consistent protocol for image acquisition, which included standardized settings for contrast, depth, frequency, and other parameters. Patients were supine and the anterior neck area was fully exposed. Longitudinal, transverse, and oblique thyroidectomy scans were performed with high-frequency linear array probes (5 to 14 MHz). The overall structure and echo of thyroid gland, the size, shape and internal echo of nodules and the status of CLN was evaluated. If a patient has multiple nodules at the same time, most suspicious one is selected for inclusion in the study based on physician experience. The longitudinal and transverse sections at the maximum diameter of the lesion were collected, and the lesion was located in the middle of the image. The captured US images are stored in DICOM format. Additionally, senior radiologist reassessed these images to exclude duplicates, excessively large images, images with severe artifacts, and blurry low-quality images, and further determined the presence of LNM based on the US findings.

2.3 Clinical data and pathological results

All patients underwent hemi-or total thyroidectomy and LN dissection within 1 month after the US examination. On the premise of unknown imaging results, the postoperative specimens of CLN of PTC were classified as metastatic or non-metastatic by pathologists with 10 years of experience. The patient's basic clinical data (including age, nodule size, location, US features, ACR-TIRADS [13] of nodules, and pathological findings (LN status, autoimmune thyroiditis, bilateral, multiple) were obtained through the medical record system.

2.4 Surgical methods

The indications for CLN dissection are primarily based on preoperative ultrasound and cytological examination results. When PTC is present in bilateral thyroid lobes, total thyroidectomy and bilateral central neck dissection (CND) are performed. For unilateral PTC, either unilateral thyroidectomy with isthmusectomy or total thyroidectomy is performed, accompanied by unilateral or bilateral CND (total thyroidectomy should be considered for patients with unilateral PTC

meeting one or more of the following criteria: tumor size > 4 cm, multiple tumors in a single lobe, extrathyroidal extension, or distant metastasis). Lateral neck dissection (LND) is considered when there is evidence of lateral neck LNM on preoperative examination or when suspicious lateral neck LNs are detected intraoperatively. The treatment decision for nodules < 1 cm is based on a comprehensive assessment of the patient, including the results of fine-needle aspiration biopsy, US characteristics, the patient's clinical symptoms, and the physician's judgment. Even if the nodule is less than 1 cm, surgery may still be necessary if FNA suggests malignancy or other high-risk features are present.

2.5 Ultrasound image super-resolution reconstruction

DL-based GAN network is adopted as the basic architecture. GAN is composed of a generator and discriminator network (Fig. 2A). They are trained in an adversarial manner. Generator uses the convolution layer to extract OR image features, learns residuals mapping between low resolution and high resolution through residuals blocks, and finally uses
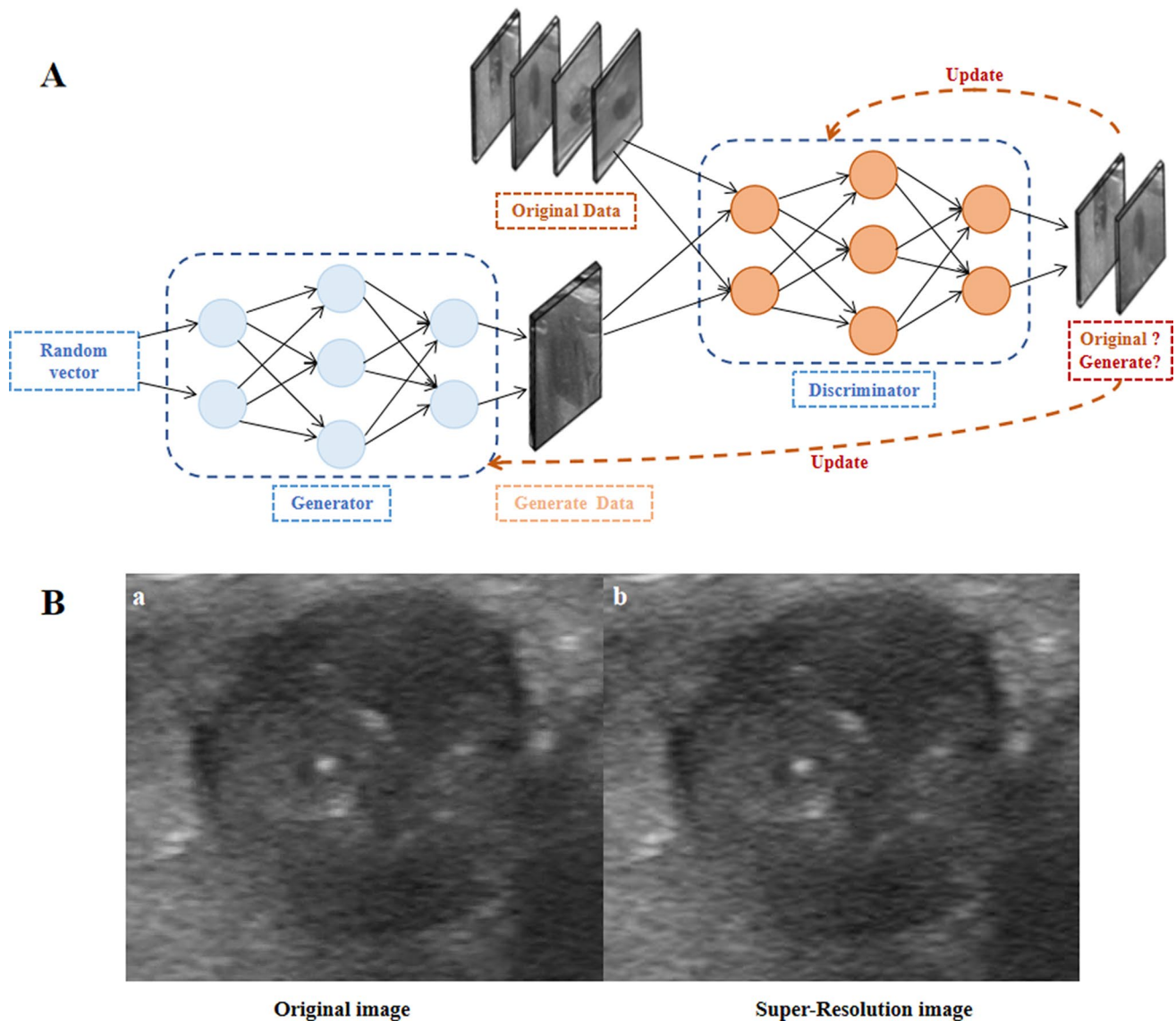

Fig. 2 Schematic diagram of a Generative Adversarial Networks (GAN) used to generate super-resolution images from original ultrasound images (**A**). The generator network randomly samples from a latent space as input and produces outputs that mimic the real data in the training set as closely as possible. The discriminator network takes either the original data or the output data from the generator network as input, and tries to distinguish the generated data from the original data, even as the generated data (**B-b**) becomes increasingly similar to the original data (**B-a**)

convolution layer to generate high-resolution images. Then, discriminator outputs a probability value through a series of convolution and pooled layers to distinguish the real and generated SR image.

In this study, transfer learning in OnekeyAI platform (version 20240616) was used to obtain SR images. On the basis of maintaining the size of the OR ultrasonic image, Gaussian noise was introduced to reduce its out-of-plane resolution to 1/4 of the OR, thereby obtaining a new low-resolution image. These low-resolution images are paired with the result-ant high-resolution images to train a lightweight GAN model. Finally, the model is applied to the OR image, and the resolution is improved by transfer learning to achieve SR reconstruction. The sagittal image reconstructed by GAN SR has a high visual similarity to the OR image. At the same time, the SR enhances the texture detail, and has a clearer edge (Fig. 2B). The technology has been evaluated on a variety of medical imaging modalities, including CT, MRI, and US, all of which have significantly improved image quality and spatial resolution. It is also compared with other state-of-the-art SR reconstruction techniques and shows superior performance [8, 11, 12].

2.6 Thyroid nodule labeling and feature extraction

The region of interest (ROI) was manually labeled by a 5-year experienced sonographer on the segmentation software ITK-SNAP (version 3.8.0, www.itksnap.org). During the annotation process of the nodules, the US physician was blinded to the pathological results and clinical characteristics of the patients. For nodules with suspicious surrounding areas, we adjusted the ROI to ensure inclusion of these areas, including halo sign, spiculation, and ill-defined margins. The anno-tations were reviewed by another physician with 10 years of experience, and any discrepancies were resolved through discussion.

The pyradiomics (version 3.0.1, http://PyRadiomics.readthedocs.io) was used to extract hand-made imaging features, including geometric, intensity and texture features, from the OR and SR images of each nodule. Geometric features describe the shape characteristics of the nodules. Intensity characteristics describe the first-order statistical distribution of voxel intensity within nodules. Texture features describe the spatial relationships and intensity distribution patterns among pixels in an image. Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone (GLSZM), Gray-Level Dependence Matrix (GLDM) and Neighbouring Gray-Tone Difference Matrix (NGTDM) is used to extract texture features.

2.7 Feature selection and radiomics model construction

After z-score regularization of all radiological features, feature selection is carried out according to the following three-step procedure: (1) Mann–Whitney U tests statistical test and feature screening were performed on all radiological features. Only features with a $p < 0.05$ were retained. (2) Spearman's rank correlation coefficient was used to evaluate the correlation between highly repeatable features, and only one of the two features with an absolute correlation coef-ficient of > 0.9 was retained. (3) The Least Absolute Shrinkage and Selection Operator (LASSO) regression model is used to construct features [12, 14], and all regression coefficients are reduced to zero by adjusting the minimum standard regularization parameter (λ), and the features with non-zero coefficients are identified. Here, we use $5 \times$ cross-validation and minimum standard adjustment λ, and the final retention coefficient is not zero to fit the regression model, forming the radiomics features. The selected features were trained using ML classifiers (support vector machine (SVM), K-Nearest Neighbors (KNN)). SR_SVM, OR_SVM, SR_KNN and OR_KNN models were finally constructed.

2.8 Deep learning model construction

The maximum cross-sectional US images of thyroid nodules were input into a pre-trained Convolutional Neural Network (CNN) through a rectangular cropping mode for transfer learning. DL features were extracted using ResNet 101 and DenseNet 121. We use the Z-score method to standardize all features and calculate the mean and variance (standard deviation) of each feature. Each column of features is subtracted from the mean, divided by variance, and converted to a standard normal distribution. Then, LASSO was used to filter out the features whose coefficients are not zero, select-ing and reducing the dimensionality of the fused features to obtain the optimal subset of fused features. Finally, the SR_ResNet 101, OR_ResNet 101, SR_DenseNet 121 and OR_DenseNet 121 models were constructed. To further explain the DL model in a human-readable format, this study utilizes Gradient-weighted Class Activation Mapping (Grad-CAM) technology to illuminate the key regions of focus for the model.

2.9 Construction of nomogram

Firstly, the clinical features with statistical significance were screened by univariate logistic regression analysis (P < 0.1), and the risk factors independently associated with the outcome were selected by multivariate logistic regression analysis. In order to intuitively and effectively evaluate the incremental prognostic value of radiomics and DL features to clinical risk factors, clinical indicators, DL signature and radiomics signature was incorporated into the diagnostic model to construct Nomogram. The diagnostic effect of Nomogram was tested in the test cohort. Calibrations were evaluated using calibration curves and the Hosmer–Lemeshow test [15]. The Receiver Operating Characteristic (ROC) curve was used to compare the diagnostic performance. Figure 3 shows the overall process of the study.

2.10 Statistical analysis

Variables are expressed as mean ± standard deviation or frequency (percent). Shapiro–Wilk test was used to assess whether the data violated the assumption of normal distribution. Student's t test or Mann–Whitney U-test was used
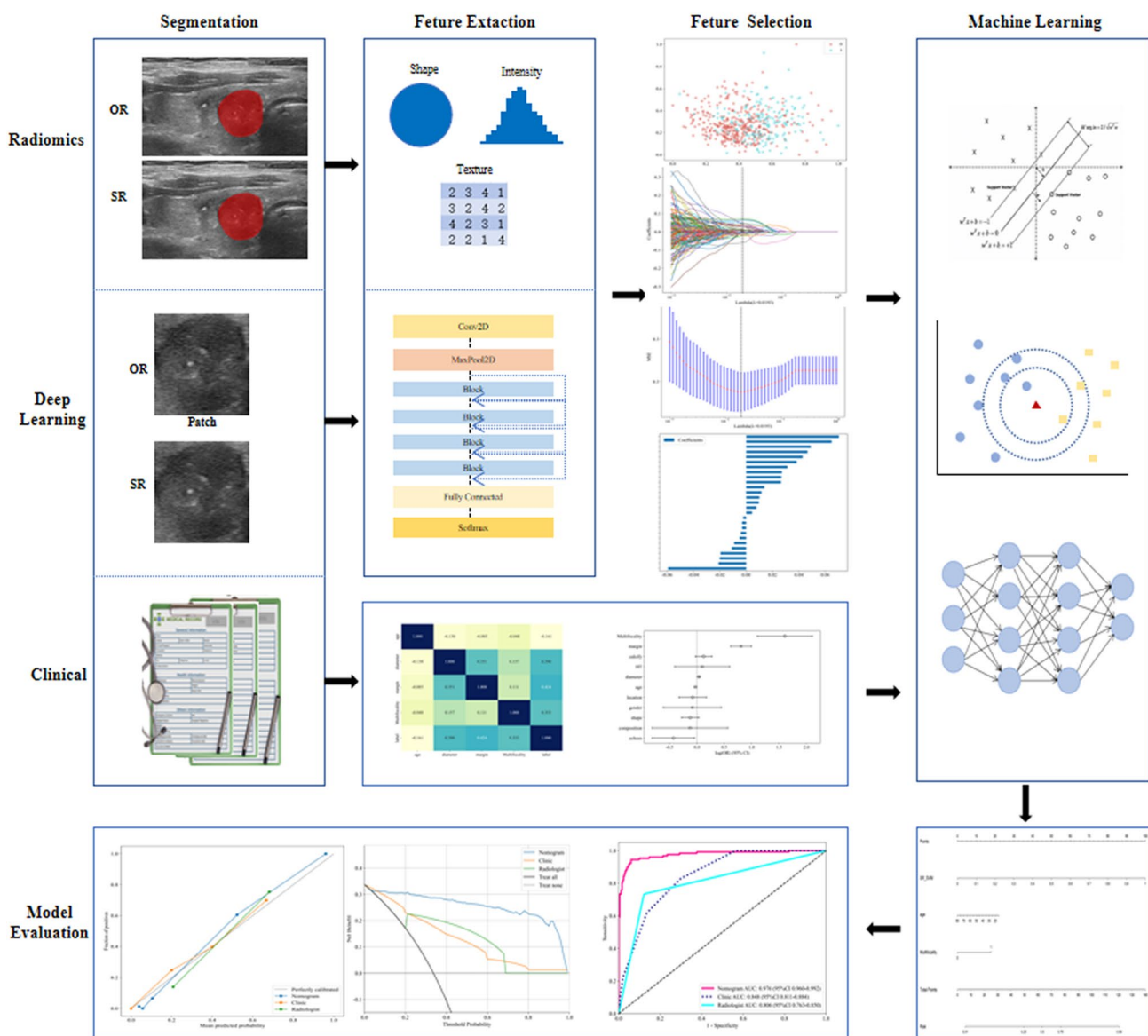

Fig. 3 The overall process of this study

to compare data between two groups. Data for categorical variables were compared using the chi-square test. Logistic regression was used to identify clinically independent correlates. DeLong test was used to compare differences between areas under the curve (AUC). Calibration curves were used and all statistical analyses were performed using R (version 4.2.3) and Python software (version 3.7.17, http://www.python).

3 Results

3.1 Baseline characteristics

In our study, 544 PTC patients with an average age of 43.03 ± 11.33 were enrolled, of which 420 (77.21%) were female and 124 (22.79%) were male, and the mean diameter was 12.22 ± 8.56 mm. There were 380 patients in the training cohort and 164 patients in the test. The US characteristics of nodules and ACR-TIRADS were not statistically different between the training and test cohort. There were 162 patients with CLNM and 382 patients without CLNM. Among the US reports, there were 182 cases with suspected LN metastasis and 362 cases without abnormal LN. Detailed data are presented in Table 1. Four cases with TR2 nodules were also included in the study, as they underwent surgery due to the nodules' significant impact on patients' lives. Intraoperative frozen section analysis indicated PTC, and based on these results, unilateral CND was performed.

The diagnostic ability of senior radiologist based on clinical indicators was evaluated in terms of AUC, which was 0.806 (95% CI 0.7626–0.8497) in the training set and 0.783 (95% CI 0.7090–0.8572) in the test set. Using clinical data, multivariate logistic regression analysis revealed that age, nodule size, margin, and multiplicity were independent risk factors for CLNM (Supplementary Table 1).

3.2 Evaluation of radiomics and deep learning model

The radiomics model evaluated 18 first-order features, 14 shape features, and 73 texture features, and after applying 17 image filters, extracted a total of 1561 manual features, i.e., $17 \times (73 + 18) + 14$. A detailed description is given in supplementary Table 2, and a comprehensive interpretation of all image features can be found on the online platform (http://pyradiomics.readthedocs.io).

The performance of the radiomics and DL model constructed is shown in Table 2. It can be found that the efficiency of the model based on SR image is generally superior to that of OR. In addition, ML model SR_SVM and DL model SR_ResNet 101 show superior performance. In the test set, the best-performing SR_SVM model had AUC values 0.56 higher than OR (p = 0.013), and the rest of the SR image-based models also had AUC values slightly higher than OR models and improved accuracy, although the differences were not statistically significant (Supplementary Fig. 1). In the training set, the AUC of the SR_SVM was 0.924 (95% CI 0.8808–0.9573), accuracy was 0.879; SR_ResNet 101 was 0.970 (95% CI 0.9510–0.9899), accuracy was 0.934; OR_SVM was 0.911 (95% CI 0.8780–0.9432), accuracy was 0.805; OR_ResNet 101 was 0.952 (95% CI 0.9284–0.9748), accuracy was 0.913. In the test set, SR_SVM was 0.878 (95% CI 0.8203–0.9358), accuracy was 0.854; SR_ResNet 101 was 0.799 (95% CI 0.7175–0.8806), accuracy was 0.793; OR_SVM was 0.822 (95% CI 0.7500–0.8937), accuracy was 0.665; OR_ResNet 101 was 0.751 (95% CI 0.6620–0.8401), accuracy was 0.713. The ROC curve of the radiomics model is shown in Fig. 4A and B, and that of the DL is shown in Fig. 4C and D. For visualizations of DL models see supplementary Fig. 2.

3.3 Nomogram construction and evaluation

The multivariate regression model revealed that age, nodule diameter, nodule margin, and multiplicity were independently associated with CLNM. By incorporating the SR_SVM model and SR_ResNet 101 model as labels with clinical parameters, Nomogram_A and Nomogram_B were derived (Figs. 5A and B), respectively. Additionally, Nomogram_C was created by solely aggregating clinical indicators (Fig. 5C). From Nomogram_A and Nomogram_B, it can be observed that clinical indicators have lower weights, indicating their relatively minor influence on prediction outcomes compared to the radiomics models.

The diagnostic performance of the nomograms was further evaluated in the test set, and a comparison was made with the diagnostic level of radiologists. As can be seen in Figs. 6A and B, within the test set, Nomogram_A demonstrated the best performance, with an AUC of 0.930 (95% CI 0.8913–0.9682) and an accuracy rate of 0.829. In contrast, Nomogram_B had

Table 1 Clinical and ultrasound characteristics of the patient

Feature name	ALL	Test	Train	P
Age (mean ± SD)	43.03 ± 11.33	43.66 ± 11.17	42.75 ± 11.40	0.425
Diameter (mean ± SD)	12.22 ± 8.56	11.57 ± 8.16	12.51 ± 8.72	0.106
Gender (%)				0.803
Female	420 (77.21)	125 (76.22)	295 (77.63)	
Male	124 (22.79)	39 (23.78)	85 (22.37)	
Bilaterality (%)				0.622
No	443 (81.43)	131 (79.88)	312 (82.11)	
Yes	101 (18.57)	33 (20.12)	68 (17.89)	
Multifocality (%)				0.590
No	408 (75.00)	120 (73.17)	288 (75.79)	
Yes	136 (25.00)	44 (26.83)	92 (24.21)	
Location (%)				0.202
Full	6 (1.10)	0	6 (1.58)	
Upper	123 (22.61)	40 (24.39)	83 (21.84)	
Middle	270 (49.63)	83 (50.61)	187 (49.21)	
Lower	114 (20.96)	36 (21.95)	78 (20.53)	
Isthums	31 (5.70)	5 (3.05)	26 (6.84)	
CLNM (%)				0.134
No	382 (70.22)	123 (75.00)	259 (68.16)	
Yes	162 (29.78)	41 (25.00)	121 (31.84)	
HT (%)				0.299
No	409 (75.18)	118 (71.95)	291 (76.58)	
Yes	135 (24.82)	46 (28.05)	89 (23.42)	
Composition (%)				0.722
Cystic and solid	13 (2.39)	5 (3.05)	8 (2.11)	
Solid	531 (97.61)	159 (96.95)	372 (97.89)	
Calcify (%)				0.797
No	228 (41.91)	68 (41.46)	160 (42.11)	
Stubby	58 (10.66)	16 (9.76)	42 (11.05)	
Peripheral	6 (1.10)	1 (0.61)	5 (1.32)	
Micro	222 (40.81)	71 (43.29)	151 (39.74)	
Stubby and micro	27 (4.96)	8 (4.88)	19 (5.00)	
Peripheral and micro	3 (0.55)	0	3 (0.79)	
Margin (%)				0.755
Smooth or ill defined	319 (58.64)	99 (60.37)	220 (57.89)	
Irregular or lobulated	158 (29.04)	44 (26.83)	114 (30.00)	
Extrathyroidal extension	67 (12.32)	21 (12.80)	46 (12.11)	
Echoes (%)				0.090
Hyper-or Isoechoic	42 (7.72)	11 (6.71)	31 (8.16)	
Hypoechoic	249 (45.77)	65 (39.63)	184 (48.42)	
Very hypoechoic	253 (46.51)	88 (53.66)	165 (43.42)	
Shape (%)				0.702
No taller-than-wide	297 (54.60)	87 (53.05)	210 (55.26)	
Taller-than-wide	247 (45.40)	77 (46.95)	170 (44.74)	
ACR_TIRADS (%)				0.688
TR2	4 (0.74)	1 (0.61)	3 (0.79)	
TR3	8 (1.47)	1 (0.61)	7 (1.84)	
TR4	102 (18.75)	33 (20.12)	69 (18.16)	
TR5	430 (79.04)	129 (78.66)	301 (79.21)	
Surgical Methods (%)				0.804
Unilateral leaf + isthmus + CDN	208 (38.24)	66 (40.24)	142 (37.37)	

Table 1 (continued)

Feature name	ALL	Test	Train	P
TT + CND	256 (47.06)	74 (45.12)	182 (47.89)	
TT + CND + LND	80 (14.71)	24 (14.63)	56 (14.74)	
Location of LNM (%)				0.840
No	379 (69.67)	117 (71.34)	262 (68.95)	
Central	114 (20.96)	33 (20.12)	81 (21.32)	
Central + Lateral	51 (9.38)	14 (8.54)	37 (9.74)	
Radiologist (%)				0.747
No	362 (66.54)	107 (65.24)	255 (67.11)	
Yes	182 (33.46)	57 (34.76)	125 (32.89)	

Continuous variables are expressed as mean ± standard (SD) deviation and categorical variables as number (percentage). CLNM: cervical lymph node metastasis; HT: Hashimoto's Thyroiditis; ACR-TIRADS: American College of Radiology Thyroid Imaging Report and Data System; TT: total thyroidectomy; CND: central neck dissection; LND: lateral neck dissection

Table 2 Performance of individual models

Model	AUC	95% CI	Accuracy	Sensitivity	Specificity	PPV	NPV	F1	Cohort
SR_ResNet101	0.970	0.9510–0.9899	0.934	0.898	0.952	0.906	0.949	0.902	Train
	0.799	0.7175–0.8806	0.793	0.689	0.832	0.608	0.876	0.646	Test
OR_ResNet101	0.952	0.9284–0.9748	0.913	0.883	0.929	0.863	0.94	0.873	Train
	0.751	0.6620–0.8401	0.713	0.733	0.706	0.485	0.875	0.584	Test
SR_DenseNet121	0.948	0.9223–0.9728	0.884	0.883	0.885	0.796	0.937	0.837	Train
	0.784	0.6988–0.8699	0.793	0.622	0.857	0.622	0.857	0.622	Test
OR_DenseNet121	0.915	0.8821–0.9484	0.853	0.867	0.845	0.74	0.926	0.799	Train
	0.814	0.7382–0.8895	0.774	0.778	0.773	0.565	0.902	0.654	Test
SR_SVM	0.924	0.8908–0.9573	0.879	0.859	0.889	0.797	0.926	0.827	Train
	0.878	0.8203–0.9358	0.854	0.644	0.933	0.784	0.874	0.707	Test
OR_SVM	0.911	0.8780–0.9432	0.805	0.945	0.734	0.644	0.964	0.766	Train
	0.822	0.7500–0.8937	0.665	0.889	0.580	0.444	0.932	0.593	Test
SR_KNN	0.914	0.8876–0.9408	0.839	0.633	0.944	0.853	0.835	0.726	Train
	0.816	0.7350–0.8971	0.811	0.578	0.899	0.684	0.849	0.627	Test
OR_KNN	0.880	0.8472–0.9134	0.829	0.641	0.925	0.812	0.835	0.716	Train
	0.737	0.6508–0.8226	0.787	0.311	0.966	0.778	0.788	0.444	Test
Radiologist	0.806	0.7626–0.8497	0.826	0.719	0.881	0.754	0.86	0.682	Train
	0.783	0.7090–0.8572	0.787	0.644	0.84	0.604	0.862	0.682	Test

AUC: area under the curve; PPV: positive predictive value; NPV: negative predictive value; SR: super resolution; OR: original

an AUC of 0.868 (95% CI 0.8102–0.9261) and the same accuracy rate of 0.829. The DeLong test revealed that the diagnostic performance of Nomogram_A, based on SR_SVM, was significantly higher than that of Nomogram_B, Nomogram_C, and the level of radiologists (P < 0.05) (Fig. 6C and D). Meanwhile, Nomogram_B, based on SR_ResNet 101, exhibited a comparable diagnostic performance to Nomogram_C, which relied solely on clinical indicators.

The calibration curve showed that Nomogram showed good consistency in diagnosing CLNM in PTC patients (Supplementary Fig. 3), and the Hosmer–Lemeshow test results showed that Nomogram plots did not deviate from perfect fit in either cohort (P > 0.05). Nomogram_A was shown by DCA to have the highest clinical value (Fig. 6E, F).
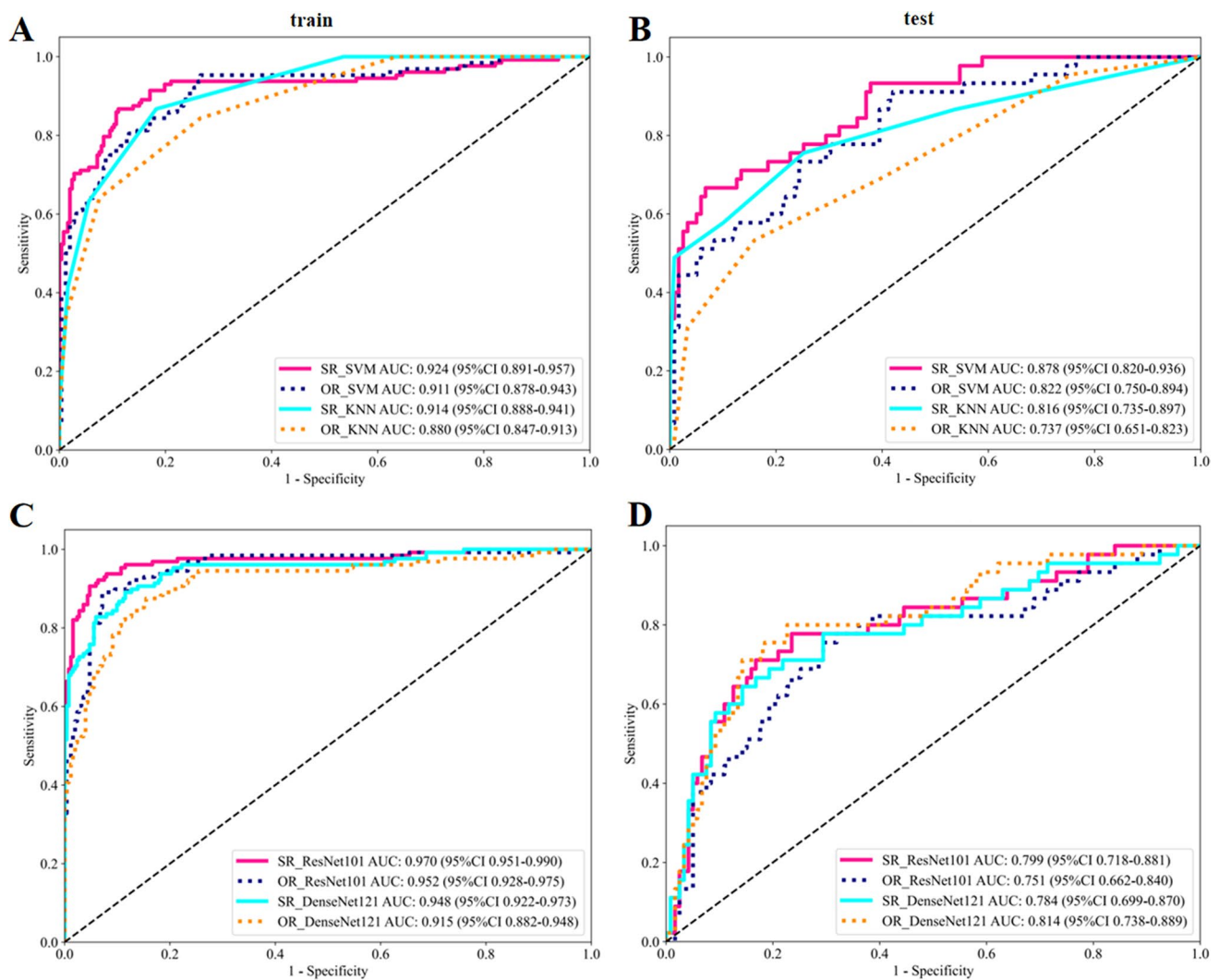
Fig. 4 ROC curves for the radiomics (**A**, **B**) and deep learning (**C**, **D**) model in the training and testing set

4 Discussion

In this study, DL-based GAN architecture was used for SR reconstruction to generate SR US images. Radiomics and DL models were developed and evaluated based on SR and OR US images, and the nomogram was further constructed combined with clinical features. The results show that SR models exhibit better performance in predicting CLNM compared to OR US-based images. The nomogram based on SR_SVM aggregation has excellent predictive performance, significantly better than clinical nomogram model and radiologist diagnostic level.

Despite PTC being an indolent tumor, CLNM may occur in some patients at an early stage, with an incidence rate ranging from 20–90% [4], mainly including central LNM and cervical lateral LNM, usually extending from the central region to the cervical lateral region, and jumping metastasis can also occur. The presence of LNM is an important reference index to determine the extent and mode of surgery, and is also an important risk factor for tumor recurrence and death [16–18]. Data published by ATA showed that the 14-year all-cause survival rate for PTC without LNM was 82%, compared with 79% for PTC with LNM ($p < 0.05$). Furthermore, the ATA guidelines categorize patients with fewer than five LNM as low-risk; however, this does not diminish the predictive value of CLNM for recurrence [19]. Studies by Sugitani et al. have demonstrated that patients with five or more LNM exhibit a significantly higher recurrence rate (19%) compared to those with fewer than five metastases (8%) [20], highlighting that even a small

Fig. 5 Summary of Nomogram. **A** Nomogram_A of obtained by aggregating SVM and clinical indicators. **B** Nomogram_B obtained by aggregating ResNet 101 and clinical indicators. **C** Nomogram_C obtained by pooling clinical factors independently associated with outcome
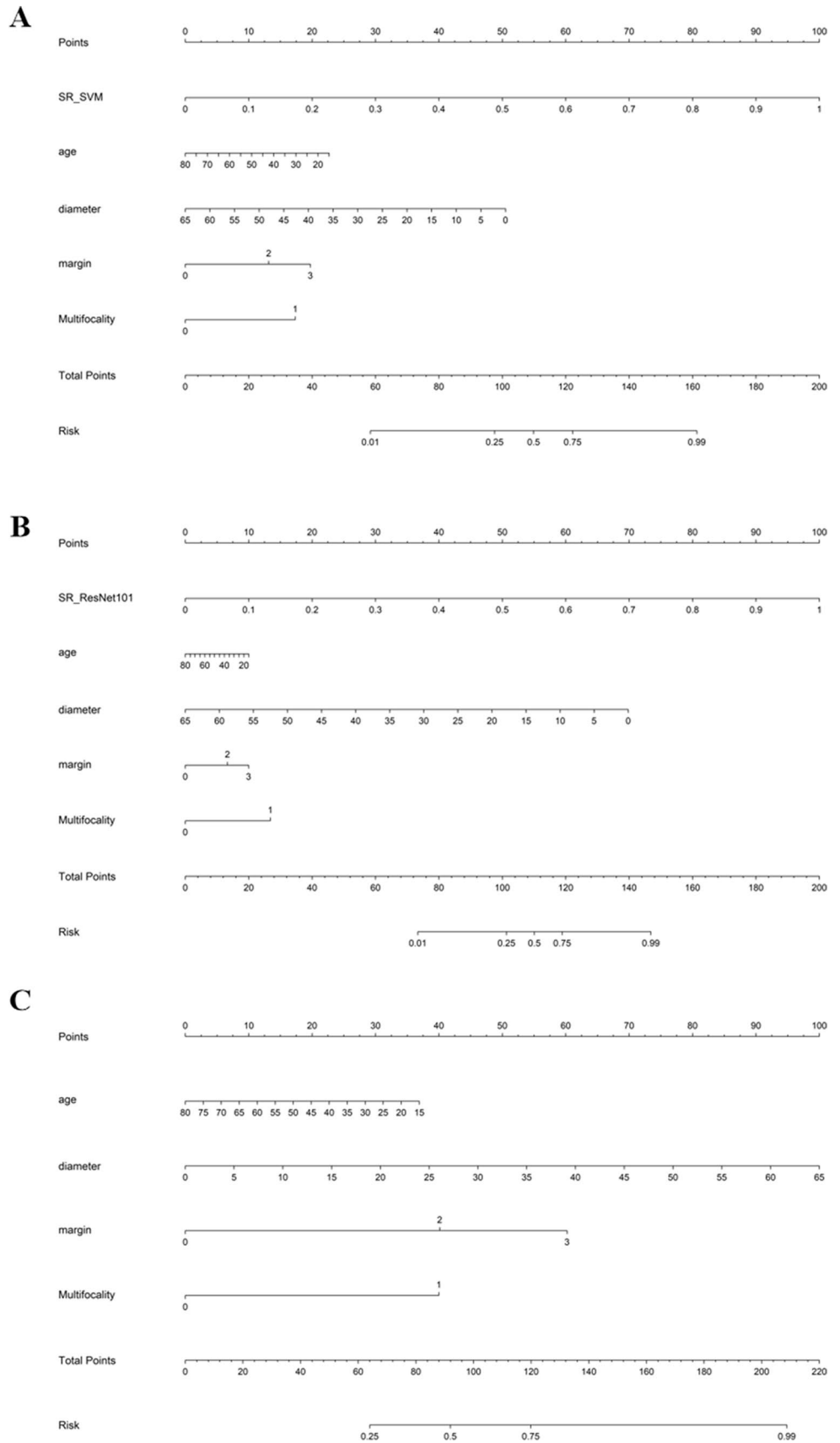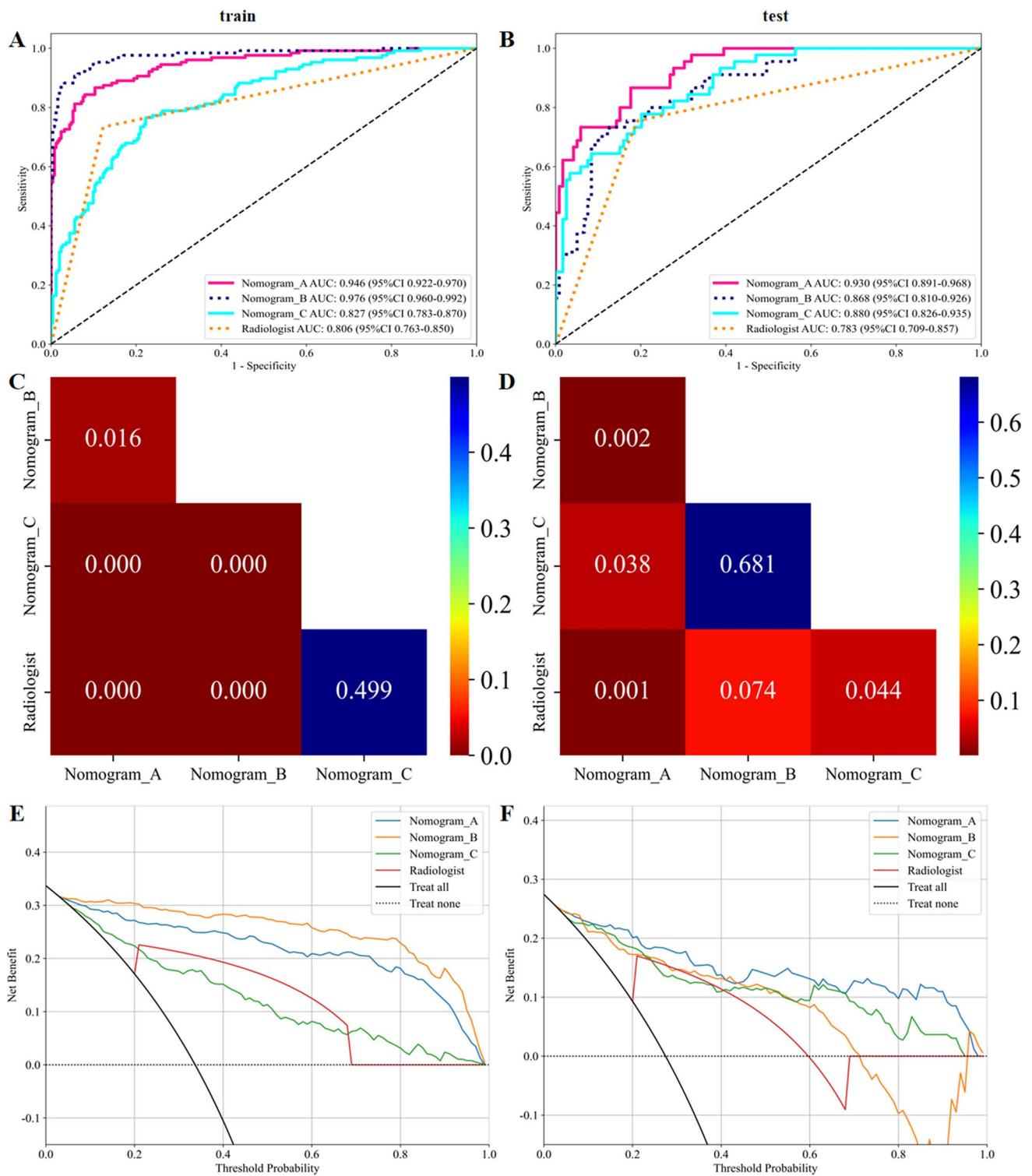
Fig. 6 Performance evaluation of the Nomograms

number of CLNM poses a risk of recurrence. Therefore, meticulous LN assessment in PTC patients is imperative to provide clinicians with a foundation for tailoring surgical plans.

Currently, US serves as the primary imaging modality for preoperative assessment of LNM in TC. However, it is plagued by low detection rates for metastatic LN and susceptibility to interference from bone and gas, which limits its effectiveness. Moreover, US often struggles to reach certain specialized regions, such as the retropharyngeal region, mediastinum,

and low-lying level IV, further constraining its utility. Studies indicate that preoperative imaging for CLNM exhibits a specificity exceeding 85% but a sensitivity ranging from only 20% to 50% [21–23]. The majority of current studies base their predictions on the primary tumor, employing traditional risk prediction models constructed from US image features, including tumor size, number, margins, Hashimoto's thyroiditis, and microcalcifications [24–26]. Our results also show that age, nodule size, margin, and multifocal are independent predictors of CLNM. In line with studies focused on PTC and LNM [27, 28], the reported nodule diameter approximates 12 mm, implying that malignant nodules may develop LNM even at smaller sizes.

However, these US signs need to be verified by experienced physicians and heavily rely on the quality of the US image. To overcome these limitations, computer-aided diagnosis (CAD) systems based on AI technologies are rapidly developing, which can transform the invisible aspects of images into values that can be read by clinicians, quantitatively analyze medical image data, and extract a large number of features such as image intensity, texture and other features from medical images by manual definition or computer automatic learning. In previous studies, AI showed good performance in predicting benign and malignant thyroid nodules and LN status [28–30]. These studies focused on gray-scale US, elastography, and contrast-enhanced US images [26, 30–32]. Due to the noise and artifacts, the low contrast between the lesion and surrounding tissue, and variability in lesion size and morphology, typical features may not always be detectable.

In recent years, SR technology improves image resolution through algorithms and restores high-resolution images from multiple low-resolution images, which has become the mainstream trend of image processing. SR can be implemented through a variety of techniques, traditional methods, or DL-based methods. DL-based SR technology is increasingly applied in medical imaging, remote sensing image and other fields [33–35]. Ryu et al. proposed that compared with the OR coronary angiography images, the image noise generated by SR is lower, the image quality is significantly higher, and the coronary stent can be displayed more clearly [36]. Farias et al. found that SR based on GAN increases the robustness of radiomic features. These findings demonstrate the promising potential of applying SR to radiomic analysis [8].

In this study, the radiomics and DL model based on SR and OR US images were established, and the nomogram was further constructed combined with clinical indicators. The results showed that SR technique significantly improved the accuracy of the prediction model (in the training set, SR_ResNet 101: AUC = 0.970, accuracy 0.934; SR_SVM: AUC = 0.924, accuracy 0.879. The test set was slightly reduced, SR_ResNet 101: AUC = 0.799 (95% CI 0.7175–0.8806), accuracy 0.793; SR_SVM: AUC = 0.878 (95% CI 0.8203–0.9358), accuracy 0.854.), suggesting that the SR images have high reliability. As the SVM was well realised in the test set, the final Nomogram_A constructed by aggregating SR_SVM, clinical indicators performed even better in predicting CLNM (AUC of 0.930 (95% CI 0.8913–0.9682), accuracy of 0.829 in the test set), and significantly outperformed the clinical nomogram model and radiologist.

The above results may be attributed to the following aspects: (1) SR technology makes the image clearer, provides more pixel information, clearer details in the image, and provides more subtle changes in nodule morphology and pathophysiology, so as to obtain more abundant and reliable feature extraction and analysis results. (2) SR images reduce image blur, improve feature recognizability, help to screen out radiomics features and DL that are strongly correlated with LNM, and improve the prediction ability of the model. (3) The prediction model constructed by logistic regression and other statistical methods can effectively integrate all valid labels to achieve accurate prediction of LNM.

In this study, we observed that the radiomics model surpassed the DL model in performance on the test set, indicating that for the specific dataset of super-resolution reconstructed PTC US images, a quantitative feature-based radiomics analysis may be more appropriate. Conversely, the DL model's high AUC value on the training set did not translate to the test set, suggesting potential issues with generalization ability, possibly due to overfitting to specific patterns in the training data. Notably, the radiomics model, with its meticulously designed features, may have more effectively captured stable features associated with LNM, rendering it less sensitive to variations in data distribution and maintaining robust predictive performance on the test set. These features may be more closely aligned with biological markers of CLNM. Nevertheless, this does not discount the potential of DL models in all cases. Their evident capabilities in image recognition and classification tasks, along with potential improvements in model architecture, data augmentation, and increased data volume, may enhance their generalization abilities. Furthermore, the development of hybrid models that integrate the strengths of both DL and radiomics models represents a promising avenue for future exploration.

This study has several limitations that must be acknowledged. Firstly, it is a single-center study, which may introduce regional and device-related differences, as well as a lack of external test. This may hinder the model's generalizability to diverse patient populations. Secondly, the retrospective nature of the study, relying on known LN biopsy results from PTC patients, and the relatively small sample size, may introduce bias into the results. Future prospective multicenter studies, with a larger sample size and the inclusion of reactive LN lesions and lymphomas as controls, are warranted. Thirdly, the use of static US images as the primary data source has inherent limitations. Future studies should consider

incorporating US videos and multi-slice images to enhance the analysis of nodule characteristics and minimize errors. Despite these limitations, the DL-based SR reconstruction method employed in this study significantly enhanced the preoperative prediction of CLNM in PTC patients.

5 Conclusion

In summary, this study found that the prediction model constructed based on SR US images outperformed the OR in predicting preoperative CLNM in patients with PTC, and demonstrate a promising application prospect. The display details of US images can be significantly improved using this technique, and the Nomogram based on SR holds great potential in enhancing the detection accuracy of prediction models and assisting clinical decision-making. It provides new insights and evidence for the application of US radiomics in the diagnosis and treatment of PTC. We believe that with continuous technological development and optimization, this method is expected to play an even greater role in clinical practice.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical statement This retrospective study was approved by the Ethics Committee of the Second Affiliated Hospital of Nanchang University and informed consent was waived (NO: IIT-O-2024-185).

Competing interests The authors declare no competing interests.

References

1. Miranda-Filho A, Lortet-Tieulent J, Bray F, Cao B, Franceschi S, Vaccarella S, Dal Maso L. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. Lancet Diabetes Endocrinol. 2021;9(4):225–34. https://doi.org/10.1016/s2213-8587(21)00027-9.
2. Xu S, Huang H, Qian J, Liu Y, Huang Y, Wang X, Liu S, Xu Z, Liu J. Prevalence of Hashimoto thyroiditis in adults with papillary thyroid cancer and its association with cancer recurrence and outcomes. JAMA Netw Open. 2021;4:7. https://doi.org/10.1001/jamanetworkopen.2021.18526.
3. Han M, Kang R, Zhang C. Lymph node mapping for tumor micrometastasis. ACS Biomater Sci Eng. 2022;8(6):2307–20. https://doi.org/10.1021/acsbiomaterials.2c00111.
4. Jiang M, Li C, Tang S, Lv W, Yi A, Wang B, Yu S, Cui X, Dietrich CF. Nomogram based on shear-wave elastography radiomics can improve preoperative cervical lymph node staging for papillary thyroid carcinoma. Thyroid. 2020;30(6):885–97. https://doi.org/10.1089/thy.2019.0780.
5. Xing Z, Qiu Y, Yang Q, Yu Y, Liu J, Fei Y, Su A, Zhu J. Thyroid cancer neck lymph nodes metastasis: meta-analysis of US and CT diagnosis. Eur J Radiol. 2020;129: 109103. https://doi.org/10.1016/j.ejrad.2020.109103.

6. Yang L, Ma Z. Nomogram based on super-resolution ultrasound images outperforms in predicting benign and malignant breast lesions. Breast Cancer: Targets Ther. 2023;15:867–78. https://doi.org/10.2147/bctt.S435510.

7. Qiu D, Cheng Y, Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: a survey. Comput Methods Progr Biomed. 2023;238: 107590. https://doi.org/10.1016/j.cmpb.2023.107590.

8. de Farias EC, di Noia C, Han C, Sala E, Castelli M, Rundo L. Impact of GAN-based lesion-focused medical image super-resolution on the robustness of radiomic features. Sci Rep. 2021;11:1. https://doi.org/10.1038/s41598-021-00898-z.

9. Zhong J, Xia Y, Chen Y, Li J, Lu W, Shi X, Feng J, Yan F, Yao W, Zhang H. Deep learning image reconstruction algorithm reduces image noise while alters radiomics features in dual-energy CT in comparison with conventional iterative reconstruction algorithms: a phantom study. Eur Radiol. 2022;33(2):812–24. https://doi.org/10.1007/s00330-022-09119-1.

10. Fan M, Liu Z, Xu M, Wang S, Zeng T, Gao X, Li L. Generative adversarial network-based super-resolution of diffusion-weighted imaging: application to tumour radiomics in breast cancer. NMR Biomed. 2020;33:8. https://doi.org/10.1002/nbm.4345.

11. Hou M, Zhou L, Sun J. Deep-learning-based 3D super-resolution MRI radiomics model: superior predictive performance in preoperative T-staging of rectal cancer. Eur Radiol. 2022;33(1):1–10. https://doi.org/10.1007/s00330-022-08952-8.

12. Wang L, Guo T, Wang L, Yang W, Wang J, Nie J, Cui J, Jiang P, Li J, Zhang H. Improving radiomic modeling for the identification of symptomatic carotid atherosclerotic plaques using deep learning-based 3D super-resolution CT angiography. Heliyon. 2024;10:8. https://doi.org/10.1016/j.heliyon.2024.e29331.

13. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS Committee. J Am Coll Radiol. 2017;14(5):587–95. https://doi.org/10.1016/j.jacr.2017.01.046.

14. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL. Computational radiomics system to decode the radiographic phenotype. Can Res. 2017;77(21):e104–7. https://doi.org/10.1158/0008-5472.Can-17-0339.

15. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. Crit Care Med. 2007;35(9):2052–6. https://doi.org/10.1097/01.CCM.0000275267.64078.B0.

16. Adam MA, Pura J, Goffredo P, Dinan MA, Reed SD, Scheri RP, Hyslop T, Roman SA, Sosa JA. Presence and number of lymph node metastases are associated with compromised survival for patients younger than age 45 years with papillary thyroid cancer. J Clin Oncol. 2015;33(21):2370–5. https://doi.org/10.1200/jco.2014.59.8391.

17. Abbasian Ardakani A, Mohammadi A, Mirza-Aghazadeh-Attari M, Faeghi F, Vogl TJ, Acharya UR. Diagnosis of metastatic lymph nodes in patients with papillary thyroid cancer. J Ultrasound Med. 2022;42(6):1211–21. https://doi.org/10.1002/jum.16131.

18. Suh YJ, Kwon H, Kim S-J, Choi JY, Lee KE, Park YJ, Park DJ, Youn Y-K. Factors affecting the locoregional recurrence of conventional papillary thyroid carcinoma after surgery: a retrospective analysis of 3381 patients. Ann Surg Oncol. 2015;22(11):3543–9. https://doi.org/10.1245/s10434-015-4448-9.

19. Haugen B, Alexander E, Bible K, Doherty G, Mandel S, Nikiforov Y, Pacini F, Randolph G, Sawka A, Schlumberger M, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid. 2016;26(1):1–133. https://doi.org/10.1089/thy.2015.0020.

20. Sugitani I, Kasai N, Fujimoto Y, Yanagisawa A. A novel classification system for patients with PTC: addition of the new variables of large (3 cm or greater) nodal metastases and reclassification during the follow-up period. Surgery. 2004;135(2):139–48. https://doi.org/10.1016/s0039-6060(03)00384-2.

21. Wada N, Duh Q, Sugino K, Iwasaki H, Kameyama K, Mimura T, Ito K, Takami H, Takanashi Y. Lymph node metastasis from 259 papillary thyroid microcarcinomas: frequency, pattern of occurrence and recurrence, and optimal strategy for neck dissection. Ann Surg. 2003;237(3):399–407. https://doi.org/10.1097/01.Sla.0000055273.58908.19.

22. Roh J, Park J, Kim J, Song C. Use of preoperative ultrasonography as guidance for neck dissection in patients with papillary thyroid carcinoma. J Surg Oncol. 2009;99(1):28–31. https://doi.org/10.1002/jso.21164.

23. Yin X, Pang T, Liu Y, Cui H, Luo T, Lu Z, Xue X, Fang G. Development and validation of a nomogram for preoperative prediction of lymph node metastasis in early gastric cancer. World J Surg Oncol. 2020;18(1):2. https://doi.org/10.1186/s12957-019-1778-2.

24. Chung SR, Baek JH, Choi YJ, Sung T-Y, Song DE, Kim TY, Lee JH. Risk factors for metastasis in indeterminate lymph nodes in preoperative patients with thyroid cancer. Eur Radiol. 2022;32(6):3863–8. https://doi.org/10.1007/s00330-021-08478-5.

25. Zhu H, Zhang H, Wei P, Zhang T, Hu C, Cao H, Han Z. Development and validation of a clinical predictive model for high-volume lymph node metastasis of papillary thyroid carcinoma. Sci Rep. 2024;14:1. https://doi.org/10.1038/s41598-024-66304-6.

26. Feng J-W, Liu S-Q, Qi G-F, Ye J, Hong L-Z, Wu W-X, Jiang Y. Development and validation of clinical-radiomics nomogram for preoperative prediction of central lymph node metastasis in papillary thyroid carcinoma. Acad Radiol. 2024;31(6):2292–305. https://doi.org/10.1016/j.acra.2023.12.008.

27. Gao Y, Wang W, Yang Y, Xu Z, Lin Y, Lang T, Lei S, Xiao Y, Yang W, Huang W, et al. An integrated model incorporating deep learning, hand-crafted radiomics and clinical and US features to diagnose central lymph node metastasis in patients with papillary thyroid cancer. BMC Cancer. 2024;24(1):69. https://doi.org/10.1186/s12885-024-11838-1.

28. Zhou L, Zeng S, Xu J, Lv W, Mei D, Tu J, Jiang F, Cui X, Dietrich C. Deep learning predicts cervical lymph node metastasis in clinically node-negative papillary thyroid carcinoma. Insights Imaging. 2023;14(1):222. https://doi.org/10.1186/s13244-023-01550-2.

29. Tian R, Yu M, Liao L, Zhang C, Zhao J, Sang L, Qian W, Wang Z, Huang L, Ma H. An effective convolutional neural network for classification of benign and malignant breast and thyroid tumors from ultrasound images. Phys Eng Sci Med. 2023;46(3):995–1013. https://doi.org/10.1007/s13246-023-01262-3.

30. Zhang C, Liu D, Huang L, Zhao Y, Chen L, Guo Y. Classification of thyroid nodules by using deep learning radiomics based on ultrasound dynamic video. J Ultrasound Med. 2022;41(12):2993–3002. https://doi.org/10.1002/jum.16006.

31. Jia W, Cai Y, Wang S, Wang J. Predictive value of an ultrasound-based radiomics model for central lymph node metastasis of papillary thyroid carcinoma. Int J Med Sci. 2024;21(9):1701–9. https://doi.org/10.7150/ijms.95022.

32. Xue J, Li S, Qu N, Wang G, Chen H, Wu Z, Cao X. Value of clinical features combined with multimodal ultrasound in predicting lymph node metastasis in cervical central area of papillary thyroid carcinoma. J Clin Ultrasound. 2023;51(5):908–18. https://doi.org/10.1002/jcu.23465.
33. Luo Y, Zhou L, Wang S, Wang Z. Video satellite imagery super resolution via convolutional neural networks. IEEE Geosci Remote Sens Lett. 2017;14(12):2398–402. https://doi.org/10.1109/lgrs.2017.2766204.
34. Zijia L, Jing H, Jiannan L, Zhicheng L, Guangtao Z. Neighborhood evaluator for efficient super-resolution reconstruction of 2D medical images. Comput Biol Med. 2024;171: 108212. https://doi.org/10.1016/j.compbiomed.2024.108212.
35. Gang Y, Li Z, Liu Aiping Fu, Xueyang CX, Rujing W. MGDUN: an interpretable network for multi-contrast MRI image super-resolution reconstruction. Comput Biol Med. 2023;167: 107605. https://doi.org/10.1016/j.compbiomed.2023.107605.
36. Jae-Kyun R, Hwan KK, Chuluunbaatar O, Som KS, Hackjoon S, Wook SJ. Improved stent sharpness evaluation with super-resolution deep learning reconstruction in coronary CT angiography. Brit J Radiol. 2024;97(1159):1286–94. https://doi.org/10.1093/bjr/tqae094.