



Papillary thyroid carcinoma: an ultrasound-based nomogram improves the prediction of lymph node metastases in the central compartment

Xiaoqi Tian^{1,2} · Qing Song^{2,3} · Fang Xie² · Ling Ren² · Ying Zhang^{1,2} · Jie Tang² · Yan Zhang² · Zhuang Jin² · Yaqiong Zhu^{1,2} · Mingbo Zhang² · Yukun Luo^{1,2}

Received: 2 October 2019 / Revised: 27 February 2020 / Accepted: 21 April 2020 / Published online: 25 June 2020
© European Society of Radiology 2020

Abstract

Objectives To develop a nomogram based on postoperative clinical and ultrasound findings to quantify the probability of central compartment lymph node metastases (CLNM).

Methods A total of 952 patients with histologically confirmed papillary thyroid carcinoma (PTC) were included in this retrospective study and assigned to three groups based on sex and age. The strongest predictors for CLNM were selected according to ultrasound imaging features, and an ultrasound (US) signature was constructed. By incorporating clinical characteristics, a predictive model presented as a nomogram was developed, and its performance was assessed with respect to calibration, discrimination and clinical usefulness.

Results Predictors contained in the nomogram included US signature, US-reported LN status and age. The US signature was constructed with tumour size and microcalcification. The nomogram showed excellent calibration in the training dataset, with an AUC of 0.826 (95% CI, 0.765–0.887) for male patients, 0.818 (95% CI, 0.746–0.890) for young females and 0.808 (95% CI, 0.757–0.859) for elder females. For male and young female patients, application of the nomogram to the validation cohort revealed good discrimination, with AUCs of 0.813 (95% CI, 0.722–0.904) and 0.814 (95% CI, 0.712–0.915), respectively. Conversely, for elderly female patients, the nomogram failed to show good performance with an AUC of 0.742 (95% CI, 0.661–0.823).

Conclusion This ultrasound-based nomogram may serve as a useful clinical tool to provide valuable information for treatment decisions, especially for male and younger female patients.

Key Points

- Age, gender, US-reported LN status and US signature were the strongest predictors of CLNM in PTC patients and informed the development of a predictive nomogram.
- Microcalcification was the strongest predictor in the US signature, as CLNM was identified in approximately 92% of patients characterised by diffuse microcalcification.
- Stratified by sex and age, this nomogram achieved good performance in predicting CLNM, especially in male and young female patients. This prediction tool may be useful as an imaging marker for identifying CLNM preoperatively in PTC patients and as a guide for personalised treatment.

Keywords Papillary thyroid carcinoma · Lymphatic metastasis · Ultrasound · Forecasting · Nomogram

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-020-06906-6>) contains supplementary material, which is available to authorized users.

✉ Yukun Luo
luoyk301@163.com

¹ Medical College of Nankai University, No.94, Weijin Road, Nankai District, Tianjin 300071, People's Republic of China

² Department of Ultrasound, Chinese PLA General Hospital, No.28, Fuxing Road, Haidian District, Beijing 100853, People's Republic of China

³ Department of Ultrasound, Seventh Medical Center of the PLA General Hospital, Beijing, People's Republic of China

Abbreviations

ATA	American Thyroid Association
CI	Confidence interval
CLND	Central compartment lymph node dissection
CLNM	Central compartment lymph node metastases
LASSO	Least absolute shrinkage and selection operator
LN	Lymph node
PTC	Papillary thyroid carcinoma

Introduction

According to the latest global cancer statistics, the incidence of thyroid cancer is rising worldwide [1]. As the most common thyroid malignancy, papillary thyroid carcinoma (PTC) accounts for approximately 85–95%, and is an indolent disease [2, 3]. However, approximately 30–80% of PTCs are associated with lymph node (LN) metastases [4–6], and some studies have shown that LN metastases are associated with disease relapse [7]. It is generally considered that central compartment lymph node dissection (CLND) is required for patients with clinically involved LNs in the central or lateral compartment [8]. Therefore, the accurate identification of central compartment lymph node metastasis (CLNM) is crucial for the optimal management of patients with PTC.

As CLNM is difficult to detect preoperatively, controversy has always existed regarding the role of routine CLND [9]. In some medical institutions, it is supported that CLND may reduce PTC recurrence, upgrade the tumour node metastasis stage and indicate risk stratification for recurrence. Accordingly, treatment regimens, such as radioactive iodine-131 therapy, may be altered. In general, CLND can allow PTC patients to receive more active treatment and less potentially hazardous reoperative surgery. In contrast, the 2015 guidelines of the American Thyroid Association (ATA) do not recommend prophylactic CLND, stating that prophylactic CLND may be considered in high-risk patients with advanced primary tumours [9]; the ATA supports the viewpoint that there is still insufficient evidence to show that prophylactic CLND is beneficial in reducing recurrence rates [10, 11]. In contrast, CLND increases potential surgical risk for patients [12, 13]. The 2014 Japanese Society of Thyroid Surgeons and Japanese Association of Endocrine Surgeons (JSTSJAES) guidelines note that in the absence of definitive data about prophylactic CLND in a large series of patients, its indication depends on the institutional policy and surgeons' skill levels [14].

As the preferred inspection method for thyroid cancer, ultrasound (US) has limited power for evaluating CLNM [15]; indeed, the sensitivity of US in diagnosing CLNM ranges from 20 to 60% [16–20]. At present, there is no uniform standard for weighing the pros and cons of prophylactic CLND. Thus, there is an urgent need for quantitative means to predict CLNM preoperatively, and attention should be focused on risk stratification.

Although several studies have reported high-risk factors relative to clinical and ultrasound features predictive of CLNM in PTC [4–6], the results have been inconsistent. In addition, some of the risk factors identified, such as extrathyroidal invasion and tumour differentiation, are only available postoperatively [21, 22] and cannot provide valuable information for aiding in pretreatment decision-making. Therefore, the development of an appropriate and noninvasive approach for assessing CLNM has been challenging.

Our study aimed not only to identify risk factors that may predict CLNM but also to develop and validate a nomogram by combining clinical and ultrasound features, an accurate and easy-to-use model for preoperatively quantifying the likelihood of CLNM in an objective manner.

Materials and methods

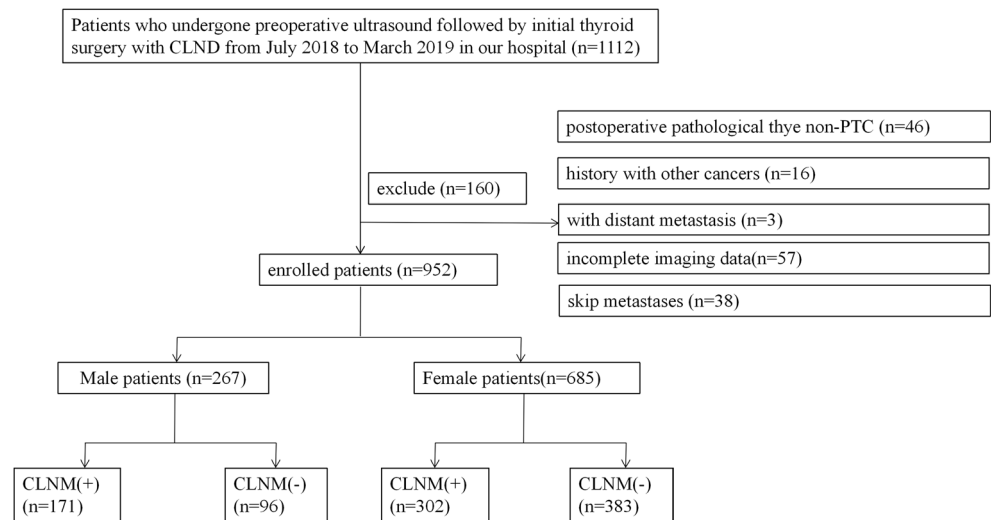
Patients

This retrospective study was approved by the Ethics Committee of the Chinese PLA General Hospital (Beijing, China), and the requirement for informed consent was waived. We retrospectively evaluated patients with histologically confirmed PTC treated in our hospital between July 2018 and March 2019. They were enrolled according to the following inclusion criteria: (1) at least one suspected malignant thyroid nodule was identified and confirmed to be malignant by US-guided puncture biopsy; (2) patients who underwent initial thyroid surgery with CLND and were confirmed as having PTC histologically; (3) no other treatment was performed before the operation; and (4) the thyroid ultrasound examination performed in our department occurred within 1 month before the operation. Patients were excluded based on the following: (1) they had distant metastases or malignant tumours in other organs; (2) they received another treatment before surgery; (3) the ultrasound imaging information was incomplete, or the quality of the images was poor; or (4) they had skip metastases [23]. Figure 1 shows the patient recruitment process. Ultimately, 952 patients (267 men and 685 women, mean age 43.2 ± 10.9 [range, 15–77] years) were included and divided into training and validation datasets according to the time of surgery. Each dataset was divided into CLNM-negative and CLNM-positive groups according to the pathology results. Baseline clinical data, including age and sex, were collected from medical records, and the threshold value of age was confirmed according to the analysis of optimal scaling regression.

Ultrasound image acquisition and assessment of ultrasound imaging features

US examinations were performed using an S2000 (Siemens Healthineers) system equipped with a 5–14-MHz linear probe

Fig. 1 Flow chart of the patients enrolled in our study



and an IU22 (Philips) system equipped with a 5–12-MHz linear probe. The US imaging characteristics of each patient were retrospectively reviewed by two independent radiologists with more than 10 years of experience in thyroid imaging; neither observer was aware of the clinical nor the pathological outcome. If the radiologists disagreed, they met to determine their final decisions by a consensus. The imaging characteristics of each nodule included tumour size, multifocality, aspect ratio (height divided by width on transverse views, A/T), tumour site, distance between the nodules and the adjacent capsule, microcalcification distribution, tumour internal vascularity and Hashimoto’s thyroiditis. Multiple images of the longitudinal and transverse axes were fully evaluated. Tumour size refers to the maximum diameter (*D*) of the nodule, as classified according to the analysis of optimal scaling regression, as follows: $D \leq 0.5$ cm, $0.5 < D$

≤ 1.0 cm, $1.0 < D \leq 1.5$ cm and $D > 1.5$ cm. If there was a suspicion of malignancy for more than one nodule, we defined it as multifocality. In multifocal cases, tumour size was classified according to the diameter of the largest tumour. The A/T was classified as ≤ 1 or > 1 . The location of the tumour was evaluated from three aspects: location 1, location 2 and location 3. These locations were divided into the following categories: upper, mid, lower, left lobe, right lobe, isthmic, inner side, outer side and middle. The microcalcification pattern was categorised as absent, present, with multiple microcalcifications and diffuse distribution within the nodules on the US image (Fig. 2a–d). The microcalcification value was defined as less than or equal to 2 mm, and the multiple microcalcification value was defined as more than five punctate high echoes within a single nodule. The relationship between the tumour and adjacent capsule was classified into three

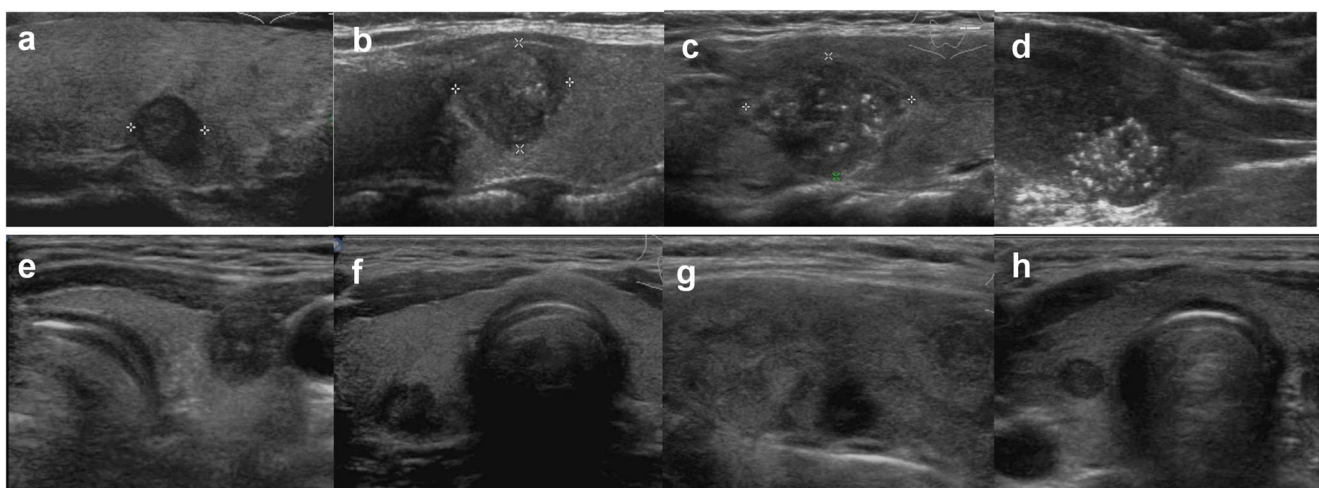


Fig. 2 Classification of the ultrasound imaging features. The microcalcification pattern was categorised as absent (a), present (b), with multiple microcalcifications (c) and diffuse distribution within the nodules (d). The distance between the tumour and the capsule was

defined as the shortest distance from the tumour border to the thyroid capsule on transverse and longitudinal views. It was classified into three categories as follows: protruding outside the thyroid capsule (e), ≤ 2 mm (f, g) and > 2 mm (h)

categories as follows: protruding outside the thyroid capsule, ≤ 2 mm (including contact with the capsule), and > 2 mm (Fig. 2e–h). Tumour vascularity was classified in accordance with the Adler criterion [24] from 0 to 3 and evaluated by colour Doppler flow imaging (CDFI). Hashimoto's thyroiditis was diagnosed on the basis of US images.

Considering that the diagnostic performance of our model depends on the accuracy of operator-reported imaging features, interobserver reproducibility for ultrasound features was assessed.

US-reported LN status

In the preoperative evaluation of cervical LNs, an LN was considered suspicious if it had one of the following features: loss of the fatty hilum, microcalcifications, hyperechoic change, a round shape or necrosis [25, 26].

Feature selection and ultrasound signature construction

We performed least absolute shrinkage and selection operator (LASSO) regression [27] to select the strongest predictive features among all the US imaging characteristics in the training cohort. The LASSO regression model operates by shrinking the coefficients of useless features to zero with the regulation parameter λ . The remaining nonzero coefficients were selected to build a logistic regression model, and the combination of these features is called the US signature in our study. Its predictive performance was assessed by ROC analysis.

Model construction and nomogram establishment in the training cohort

We performed multivariate logistic regression analysis by combining the US signature with clinical characteristics, including age, sex and US-reported LN status. A predictive model named the combined model was thus constructed. This model is presented as a nomogram that can visually and individually indicate the probability of CLNM.

Evaluation of the predictive model

The prediction formula based on the primary cohort was applied to all PTC patients in the validation cohort, and the probability of CLNM was calculated. For calibration of the model, calibration curves were plotted using pathological results and the nomogram prediction probabilities of CLNM. We evaluated the goodness of fit of the model using the Hosmer-Lemeshow test, a significant statistical method utilised to test whether the model is calibrated perfectly [28]. Nomogram discrimination was quantified using a ROC curve.

Decision curve analysis was conducted to estimate the clinical utility of the nomogram.

Statistical analysis

Statistical analysis was conducted with SPSS Statistics version 24.0 (IBM Corp.) and R software version 3.5.3 (The R Foundation for Statistical Computing). Categorical variables were reported as numbers and percentages. A chi-square test or Fisher's exact test was used to assess differences between groups. The Mann–Whitney *U* test was applied for continuous variables. The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.

Interobserver agreement was analysed for each variable using kappa (*k*) statistics. For continuous variables, the agreement of tumour size between US imaging features and pathological results was evaluated by Spearman's correlation analysis. ROC curve analysis was employed to determine the appropriate cut-off value for the probabilities of CLNM corresponding to the maximal Youden index, and the sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV) were calculated.

Results

Characteristics of patients and features of nodules

Among the 952 patients, there was a significant difference in sex between CLNM-positive and CLNM-negative patients. CLNM was found in 64.04% of male and 44.09% of female patients ($p < 0.001$). Young age was highly predictive of CLNM; for male patients, 40 years was confirmed as the threshold value for age, whereas the age distribution was particularly uneven for female patients with CLNM. To adjust for the age factor, we developed two separate nomograms for female patients, young females (≤ 35 years) and elder females (> 35 years), with CLNM probabilities of 58.25% and 38.49%, respectively. The accuracy of the subjective US-reported LN status was only 0.656 for the entire cohort, with a high specificity of 93.1% but a poor sensitivity of 39.9%. There were 284 patients who were reported to be LN negative but confirmed to have CLNM postoperatively.

The patient characteristics and US features of thyroid nodules in the training and validation cohorts are shown in Tables 1, 2 and 3. There were almost no significant differences in the characteristics between the two datasets, which justified their use as training and validation cohorts. Univariate analysis was conducted to determine differences in clinical and US characteristics between CLNM-positive and CLNM-negative groups. The agreement of ultrasound features between the two radiologists was satisfactory, with kappa coefficients between 0.81 and 0.92 (Supplementary Table 1).

Table 1 Clinical and US imaging characteristics of male patients in the training and validation datasets

Characteristic	Training dataset		P_{intra}	Validation dataset		P_{intra}	P_{inter}
	CLNM(-)(n = 63)	CLNM(+)(n = 117)		CLNM(-)(n = 33)	CLNM(+)(n = 54)		
Male	63 (35.0)	117 (65.0)		33 (37.9)	54 (62.1)		0.299
Age (years)			0.002			0.008	0.210
≤ 40	20 (23.5)	65 (76.5)		7 (20.6)	27 (79.4)		
> 40	43 (45.3)	52 (54.7)		26 (49.1)	27 (50.9)		
US-reported LN status			< 0.001			< 0.001	0.246
Negative	56 (45.9)	66 (54.1)		32 (49.2)	33 (50.8)		
Positive	7 (12.1)	51 (87.9)		1 (4.6)	21 (95.4)		
Multifocality			0.759			0.369	0.985
Absent	45 (35.7)	81 (64.3)		25 (41.0)	36 (59.0)		
Present	18 (33.3)	36 (66.7)		8 (30.8)	18 (69.2)		
Size			< 0.001			0.004	0.095
≤ 0.5 cm	10 (76.9)	3 (23.1)		7 (87.5)	1 (12.5)		
0.5–1 cm	36 (42.9)	48 (57.1)		15 (45.5)	18 (54.5)		
1–1.5 cm	14 (31.8)	30 (68.2)		9 (27.3)	24 (72.7)		
> 1.5 cm	3 (7.7)	36 (92.3)		2 (15.4)	11 (84.6)		
A/T			0.184			0.717	0.433
≤ 1	15 (27.8)	39 (72.2)		8 (34.8)	15 (65.2)		
> 1	48 (38.1)	78 (61.9)		25 (39.1)	39 (60.9)		
Location 1			0.766			0.656	0.286
Left lobe	25 (37.3)	42 (62.7)		17 (41.5)	24 (58.5)		
Right lobe	36 (34.3)	69 (65.7)		14 (33.3)	28 (66.7)		
Isthmic	2 (25.0)	6 (75.0)		2 (50.0)	2 (50.0)		
Location 2			0.890			0.035	0.088
Upper	16 (32.7)	33 (67.3)		5 (26.3)	14 (73.7)		
Mid	28 (35.0)	52 (65.0)		17 (33.3)	34 (66.7)		
Lower	19 (37.3)	32 (62.7)		11 (64.7)	6 (35.3)		
Location 3			0.655			0.813	< 0.001
Outer side	23 (38.3)	37 (61.7)		11 (39.3)	17 (60.7)		
Mid	22 (36.1)	39 (63.9)		5 (45.5)	6 (54.5)		
Inner side	18 (30.5)	41 (69.5)		17 (35.4)	31 (64.6)		
Adjacent capsule			0.039			0.877	0.347
> 2 mm	11 (61.1)	7 (38.9)		3 (37.5)	5 (62.5)		
≤ 2 mm	44 (33.3)	88 (66.7)		23 (39.7)	35 (60.3)		
Protruding capsule	8 (26.7)	22 (73.3)		7 (33.3)	14 (66.7)		
Microcalcification			< 0.001			0.026	0.764
Absent	41 (55.4)	33 (44.6)		18 (56.3)	14 (43.7)		
Present	15 (34.9)	28 (65.1)		9 (36.0)	16 (64.0)		
Multiple	7 (13.0)	47 (87.0)		6 (22.2)	21 (77.8)		
Diffused distribution	0 (0)	9 (100.0)		0 (0)	3 (100.0)		
CDFI			0.002			0.607	0.235
0	38 (46.3)	44 (53.7)		16 (43.2)	21 (56.8)		
1	16 (35.6)	29 (64.4)		11 (36.7)	19 (63.3)		
2–3	9 (17.0)	44 (83.0)		6 (30.0)	14 (70.0)		
Hashimoto’s thyroiditis			0.492			0.460	0.730
Present	8 (42.1)	11 (57.9)		4 (50.0)	4 (50.0)		
Absent	55 (34.2)	106 (65.8)		29 (36.7)	50 (63.3)		

Categorical variables are described as the number of patients with the percentage in parentheses

CLNM+ represents patients with central lymph node metastasis, and CLNM- represents patients without central lymph node metastasis. P_{intra} is the result of univariate analyses between the CLNM+ and CLNM- groups while P_{inter} represents whether a significant difference exists between the training and validation datasets

Ultrasound signature construction and diagnostic validation

LASSO regression analysis was performed to clarify the US imaging features as the strongest predictors, including tumour size and microcalcification, in the male and young female

training cohorts. Tumour vascularity was also included for the elder female patients (Fig. 3a–f). A US signature containing these CLNM-related features was constructed based on the US score. The prediction performance of the US signature was good, which was then confirmed in the validation cohort (Fig. 5a–c).

Table 2 Clinical and US imaging characteristics of young female patients (≤ 35 years) patients in the training and validation datasets

Characteristic	Training dataset		P_{intra}	Validation dataset		P_{intra}	P_{inter}
	CLNM(-)($n = 54$)	CLNM(+)($n = 76$)		CLNM(-)($n = 27$)	CLNM(+)($n = 37$)		
Young female (≤ 35 years)	54 (41.5)	76 (58.5)		27 (42.2)	37 (57.8)		0.931
Age (years)			0.057			0.151	0.860
≤ 30	20 (32.8)	41 (67.2)		9 (32.1)	19 (67.9)		
30–35	34 (49.3)	35 (50.7)		18 (50.0)	18 (50.0)		
US-reported LN status			< 0.001			0.003	0.698
Negative	48 (51.1)	46 (48.9)		24 (54.5)	20 (45.5)		
Positive	6 (16.7)	30 (83.3)		3 (15.0)	17 (85.0)		
Multifocality			0.722			0.746	0.601
Absent	44 (42.3)	60 (57.7)		21 (41.2)	30 (58.8)		
Present	10 (38.5)	16 (61.5)		6 (46.2)	7 (53.8)		
Size			0.001			0.009	0.831
≤ 0.5 cm	11 (73.3)	4 (26.7)		7 (87.5)	1 (12.5)		
0.5–1 cm	31 (50.8)	30 (49.2)		9 (47.4)	10 (52.6)		
1–1.5 cm	6 (26.1)	17 (73.9)		7 (43.7)	9 (56.3)		
> 1.5 cm	6 (19.3)	25 (80.7)		4 (19.1)	17 (80.9)		
A/T			0.127			0.355	0.401
≤ 1	17 (33.3)	34 (66.7)		10 (35.7)	18 (64.3)		
> 1	37 (46.8)	42 (53.2)		17 (47.2)	19 (52.8)		
Location 1			0.450			0.683	0.338
Left lobe	28 (45.9)	33 (54.1)		11 (44.0)	14 (56.0)		
Right lobe	25 (39.1)	39 (60.9)		16 (42.1)	22 (57.9)		
Isthmic	1 (20.0)	4 (80.0)		0 (0)	1 (100.0)		
Location 2			0.120			0.981	0.357
Upper	8 (25.8)	23 (74.2)		6 (42.9)	8 (57.1)		
Mid	24 (48.0)	26 (52.0)		15 (42.9)	20 (57.1)		
Lower	22 (44.9)	27 (55.1)		6 (40.0)	9 (60.0)		
Location 3			0.570			0.144	0.767
Outer side	17 (41.5)	24 (58.5)		7 (41.2)	10 (58.8)		
Mid	15 (35.7)	27 (64.3)		9 (64.3)	5 (35.7)		
Inner side	22 (46.8)	25 (53.2)		11 (33.3)	22 (66.7)		
Adjacent capsule			0.965			0.061	0.561
> 2 mm	4 (40.0)	6 (60.0)		4 (66.7)	2 (33.3)		
≤ 2 mm	45 (42.1)	62 (57.9)		20 (47.6)	22 (52.4)		
Protruding capsule	5 (38.5)	8 (61.5)		3 (18.8)	13 (81.2)		
Microcalcification			< 0.001			0.011	0.796
Absent	33 (67.3)	16 (32.7)		13 (68.4)	6 (31.6)		
Present	11 (39.3)	17 (60.7)		4 (44.4)	5 (55.6)		
Multiple	10 (25.6)	29 (74.4)		9 (37.5)	15 (62.5)		
Diffused distribution	0 (0)	14 (100.0)		1 (8.3)	11 (91.7)		
CDFI			0.001			0.045	0.304
0	35 (57.4)	26 (42.6)		13 (65.0)	7 (35.0)		
1	12 (36.4)	21 (63.6)		6 (31.6)	13 (68.4)		
2–3	7 (19.4)	29 (80.6)		8 (32.0)	17 (68.0)		
Hashimoto's thyroiditis			0.108			0.688	0.305
Present	19 (52.8)	17 (47.2)		7 (46.7)	8 (53.3)		
Absent	35 (37.2)	59 (62.8)		20 (40.8)	29 (59.2)		

Categorical variables are described as the number of patients with the percentage in parentheses

CLNM+ represents patients with central lymph node metastasis, and CLNM- represents patients without central lymph node metastasis. P_{intra} is the result of univariate analyses between the CLNM+ and CLNM- groups while P_{inter} represents whether a significant difference exists between the training and validation datasets

Development of the prediction model

Significant differences between CLNM-positive and CLNM-negative patients were observed for the US signature and clinical characteristics. After multivariate analysis, age, the US-reported LN status and the US signature remained independent predictors for CLNM, as shown in Table 4.

Validation of the individualised prediction nomogram

The nomogram displayed good performance for predicting CLNM in the training cohort (Fig. 4a–f). Application of the developed nomogram in the validation dataset still displayed good discrimination in the male and young female cohorts, with AUCs of 0.813 (95% CI, 0.722–0.904) and 0.814 (95%

Table 3 Clinical and US imaging characteristics of elder female patients (> 35 years) patients in the training and validation datasets

Characteristic	Training dataset		P_{intra}	Validation dataset		P_{intra}	P_{inter}
	CLNM(-)(n = 200)	CLNM(+)(n = 126)		CLNM(-)(n = 102)	CLNM(+)(n = 63)		
Elder female (> 35 years)	200 (61.4)	126 (38.6)		102 (61.8)	63 (38.2)		0.920
Age (years)			0.203			0.132	0.320
35–45	72 (58.5)	51(41.5)		37 (57.8)	27 (42.2)		
45–60	110 (61.1)	70 (38.9)		50 (60.2)	33 (39.8)		
> 60	18 (78.3)	5 (21.7)		15 (83.3)	3 (16.7)		
US-reported LN status			< 0.001			< 0.001	0.233
Negative	191 (69.9)	82(30.1)		95 (73.1)	35 (26.9)		
Positive	9 (17.0)	44 (83.0)		7 (20.0)	28 (80.0)		
Multifocality			0.034			0.832	0.392
Absent	149 (65.1)	80(34.9)		76 (62.3)	46 (37.7)		
Present	51 (52.6)	46 (47.4)		26 (60.5)	17 (39.5)		
Size			< 0.001			0.021	0.797
≤ 0.5 cm	36 (75.0)	12 (25.0)		22 (81.5)	5 (18.5)		
0.5–1 cm	128 (71.5)	51 (28.5)		56 (64.4)	31 (35.6)		
1–1.5 cm	24 (42.9)	32 (57.1)		17 (50.0)	17 (50.0)		
> 1.5 cm	12 (27.9)	31 (72.1)		7 (41.2)	10 (58.8)		
A/T			0.374			0.668	0.440
≤ 1	65 (58.0)	47 (42.0)		39 (63.9)	22 (36.1)		
> 1	135 (63.1)	79 (36.9)		63 (60.6)	41 (39.4)		
Location 1			0.542			0.966	0.233
Left lobe	85 (60.3)	56 (39.7)		44 (62.0)	27 (38.0)		
Right lobe	103 (63.6)	59 (36.4)		54 (61.4)	34 (38.6)		
Isthmic	12 (52.2)	11 (47.8)		4 (66.7)	2 (33.3)		
Location 2			0.876			0.762	0.306
Upper	45 (62.5)	27 (37.5)		23 (57.5)	17 (42.5)		
Mid	86 (62.3)	52 (37.7)		49 (62.0)	30 (38.0)		
Lower	69 (59.5)	47 (40.5)		30 (65.2)	16 (34.8)		
Location 3			0.222			0.266	0.048
Outer side	78 (66.7)	39 (33.3)		41 (65.1)	22 (34.9)		
Mid	54 (55.1)	44 (44.9)		25 (69.4)	11 (30.6)		
Inner side	68 (61.3)	43 (38.7)		36 (54.5)	30 (45.5)		
Adjacent capsule			< 0.001			0.070	0.341
> 2 mm	30 (76.9)	9 (23.1)		17(65.4)	9 (34.6)		
≤ 2 mm	145 (65.6)	76 (34.4)		72 (66.1)	37 (33.9)		
Protruding capsule	25 (37.9)	41 (62.1)		13 (43.3)	17 (56.7)		
Microcalcification			< 0.001			0.001	0.057
Absent	144 (75.4)	47 (24.6)		56 (74.7)	19 (25.3)		
Present	39 (54.2)	33 (45.8)		31 (63.3)	18 (36.7)		
Multiple	16 (27.1)	43 (72.9)		13 (38.2)	21 (61.8)		
Diffused distribution	1 (25.0)	3 (75.0)		2 (28.6)	5 (71.4)		
CDFI			< 0.001			< 0.001	0.072
0	131 (71.2)	53(28.8)		60 (73.2)	22 (26.8)		
1	46 (59.7)	31 (40.3)		27 (49.1)	28 (50.9)		
2–3	23 (35.4)	42 (64.6)		15 (53.6)	13 (46.4)		
Hashimoto’s thyroiditis			0.241			0.786	0.387
Present	44 (67.7)	21 (32.3)		24 (60.0)	16 (40.0)		
Absent	156 (60.0)	105 (40.0)		78 (62.4)	47 (37.6)		

Categorical variables are described as the number of patients with the percentage in parentheses

CLNM+ represents patients with central lymph node metastasis, and CLNM- represents patients without central lymph node metastasis. P_{intra} is the result of univariate analyses between the CLNM+ and CLNM- groups while P_{inter} represents whether a significant difference exists between the training and validation datasets

CI, 0.712–0.915), respectively. (Fig. 5a, b) The sensitivity and accuracy for the prediction of CLNM were much better than those for US detection. Comparisons of diagnostic performance are shown in Table 5.

The calibration curve for the nomogram yielded a nonsignificant statistic and suggested no departure from

a perfect fit in the validation dataset. ($p = 0.928$ for males, $p = 0.08$ for young females) (Fig. 5d, e). The combined model was presented as a nomogram (Fig. 6a, b).

However, for the elder female cohort, the performance of the prediction model was not sufficient, with

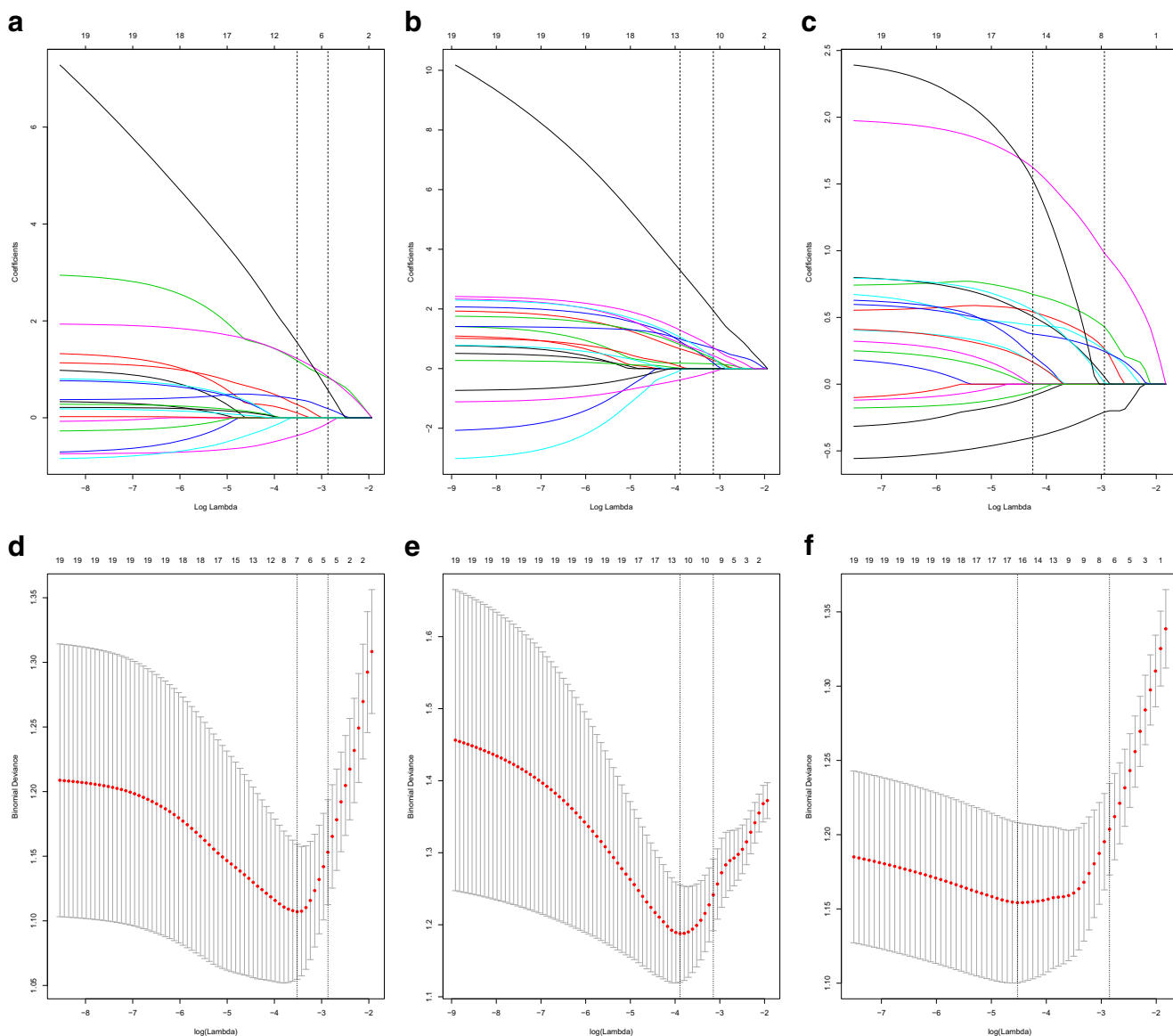


Fig. 3 LASSO coefficient profiles of the US features associated with CLNM (**a**, **b**, **c**), ultrasound feature selection by using the LASSO binary logistic regression model. Dotted vertical lines were drawn at the

optimal values by using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria) (**d**, **e**, **f**)

an AUC of 0.742 (95% CI, 0.661–0.823). (Fig. 5c). The calibration curve showed a statistically significant difference from a perfect fit in the validation dataset ($p = 0.03$) (Fig. 5f).

Clinical use

Decision curve analysis of the nomogram is presented in Fig. 6c and d. The decision curve showed that if the threshold probability was > 0.21 for males or > 0.14 for females, using the nomogram to predict LN metastases added more benefit than either the treat-all-patients scheme or the treat-none scheme.

Discussion

In this study, we developed and validated an ultrasound-based model stratified by sex and age for predicting the probability of CLNM in PTC patients. The nomogram successfully stratified patients according to their risk of CLNM and yielded excellent performance especially in male and younger female cohorts. Different from published studies, the originality of our study is that the individual probability of CLNM can be evaluated preoperatively and noninvasively [22].

Young age has been recognised as an important risk factor for predicting LN metastasis and the recurrence of PTC [29, 30]. Xu et al [31] reported that younger age (\leq

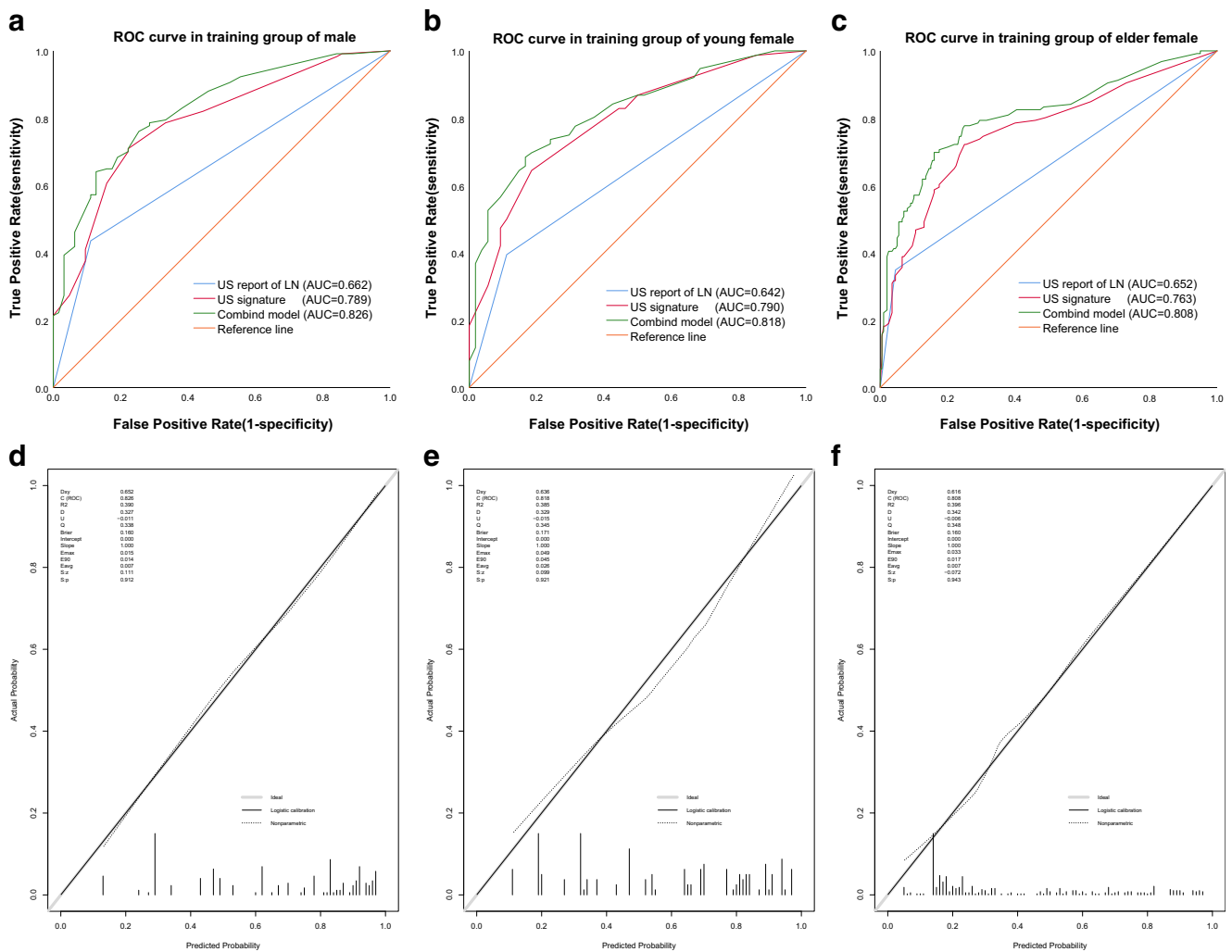


Fig. 4 ROC curve of the US-reported LN status, US signature and combined model for predicting CLNM in the training dataset of male patients (a), young female patients (b) and elder female patients (c). Calibration curve of the combined model in the training cohorts of male patients (d), young female patients (e) and elder female patients (f). The x-axis

represents the probability that the nomogram predicted CLNM, and the y-axis represents the actual rate of CLNM. The diagonal dashed line indicates ideal prediction by a perfect model, and the solid line represents the predictive power of the nomogram. The closer the solid line is to the dotted line, the better is the predictive power of the model

36 years) was an independent clinical factor for predicting CLNM. Ito et al [32] also found that patients younger than 40 years showed more tumour growth than did older patients during active surveillance, with an age threshold similar to our research findings. An increasing number of recent studies have shown an association between male sex and aggressive PTC tumour behaviours [33]. We considered that the pathogenesis and biological behaviour of tumours in PTC patients may differ according to sex and age. Therefore, when the model was stratified by age and sex, its performance was much stronger than that of the previous unstratified model. We also noted that the validation results for females older than 35 years were unsatisfactory. Based on this result, we consider that for most of the elderly patients, the tumours were discovered accidentally instead of being actively surveilled, and thus the exact

history of the tumours was unknown. As patients with a longer history of PTC who should have been in the younger group were included in the elder group, the stability of the prediction model was disrupted. We will explore deeper reasons in our future research.

In our study, a US signature was built using the strongest risk factors including tumour size and microcalcification for predicting CLNM. Tumours with a larger size on US examination were more likely to be associated with CLNM in our research, consistent with other reports [34, 35]. As a preoperative tool, the nomogram relies on a close correlation between sonographic findings and corresponding pathological results. Several studies have examined US findings of thyroid malignancies compared with pathologic results and found good agreement with regard to tumour size [36]. The discovery that merits discussion is the potential impact of microcalcification on CLNM,

Table 4 Multivariate logistic regression analysis of risk factors for CLNM

Characteristic	β	Odds ratio[95% CI]	<i>p</i>
Male			
Age (years)	0.78	2.17 [1.02–4.63]	0.045
US report of LN	1.36	3.88 [1.49–10.09]	0.005
US signature	4.40	82.04 [12.64–532.32]	< 0.001
Young female			
Age (years)	0.72	2.05 [0.86–4.85]	0.104
US report of LN	1.31	3.72 [1.25–11.04]	0.018
US signature	4.51	90.73 [13.07–630.13]	< 0.001
Elder female			
Age (years)	0.22	1.25 [0.72–2.15]	0.430
US report of LN	2.05	7.80 [3.46–17.61]	< 0.001
US signature	4.28	71.98 [21.32–243.02]	< 0.001

β is the regression coefficient, data in square brackets are 95% CIs

which was the strongest predictor in the US signature. A large number of punctate high echoes within thyroid nodules is of great predictive value for CLNM, especially with regard to the presence of diffuse distribution of microcalcification. In our study, nodules in 49 patients were characterised by diffuse microcalcification, and postoperative pathology identified CLNM in 91.8% (45/49) of these patients. To our surprise, CLNM was identified in 100% (12/12) of male patients and 96.1% (25/26) of young female patients. To the best of our knowledge, only a few studies have investigated the association between calcification and LN metastases; indeed, relevant research about the correlation between the distribution of microcalcification and CLNM has not been reported. Bai Y et al [37, 38] evaluated a group of PTC patients to determine the clinical significance of different types of calcification, finding that patients with psammoma bodies were more likely to have gross LN metastases, which is consistent with our findings,

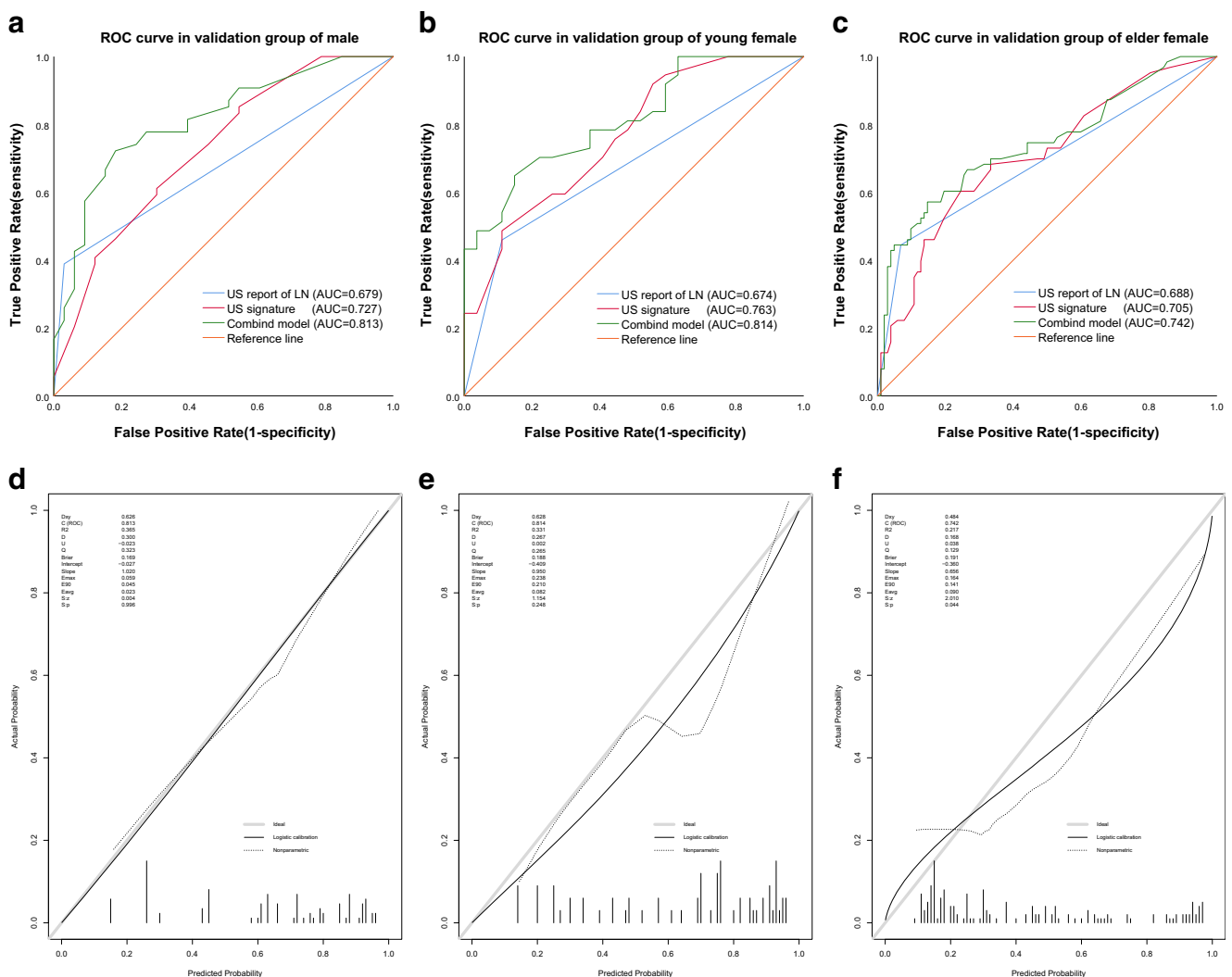


Fig. 5 ROC curve of the US-reported LN status, US signature and combined model for predicting CLNM in the validation dataset of male patients (**a**), young female patients (**b**) and elder female patients (**c**).

Calibration curve of the combined model in the validation cohorts of male patients (**d**), young female patients (**e**) and elder female patients (**f**)

Table 5 Diagnostic performance of the nomogram for predicting CLNM compared with US evaluation for LN status

Feature	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV	NPV	Cut-off
Male						0.641
US evaluation	38.9 (21/54)	96.9 (32/33)	60.9 (53/87)	95.4 (21/22)	49.2 (32/65)	
Nomogram prediction	72.2 (39/54)	81.8 (27/33)	75.9 (66/87)	86.7 (39/45)	64.3 (27/42)	
Young female						0.660
US evaluation	45.9 (17/37)	88.8 (24/27)	64.1 (41/64)	85.0 (17/20)	54.5 (24/44)	
Nomogram prediction	78.4 (29/37)	63.0 (17/27)	71.9 (46/64)	74.4 (29/39)	68.0 (17/25)	

Data are number of patients, data in parentheses are numerator/denominator of patients; *PPV*, positive predictive value; *NPV*, negative predictive value

as psammoma bodies mostly represent microcalcification on US images [39]. Nonetheless, it should be noted that other

pathological structures, such as focal fibrosis of nodular goitres, have an appearance similar to microcalcification on US images;

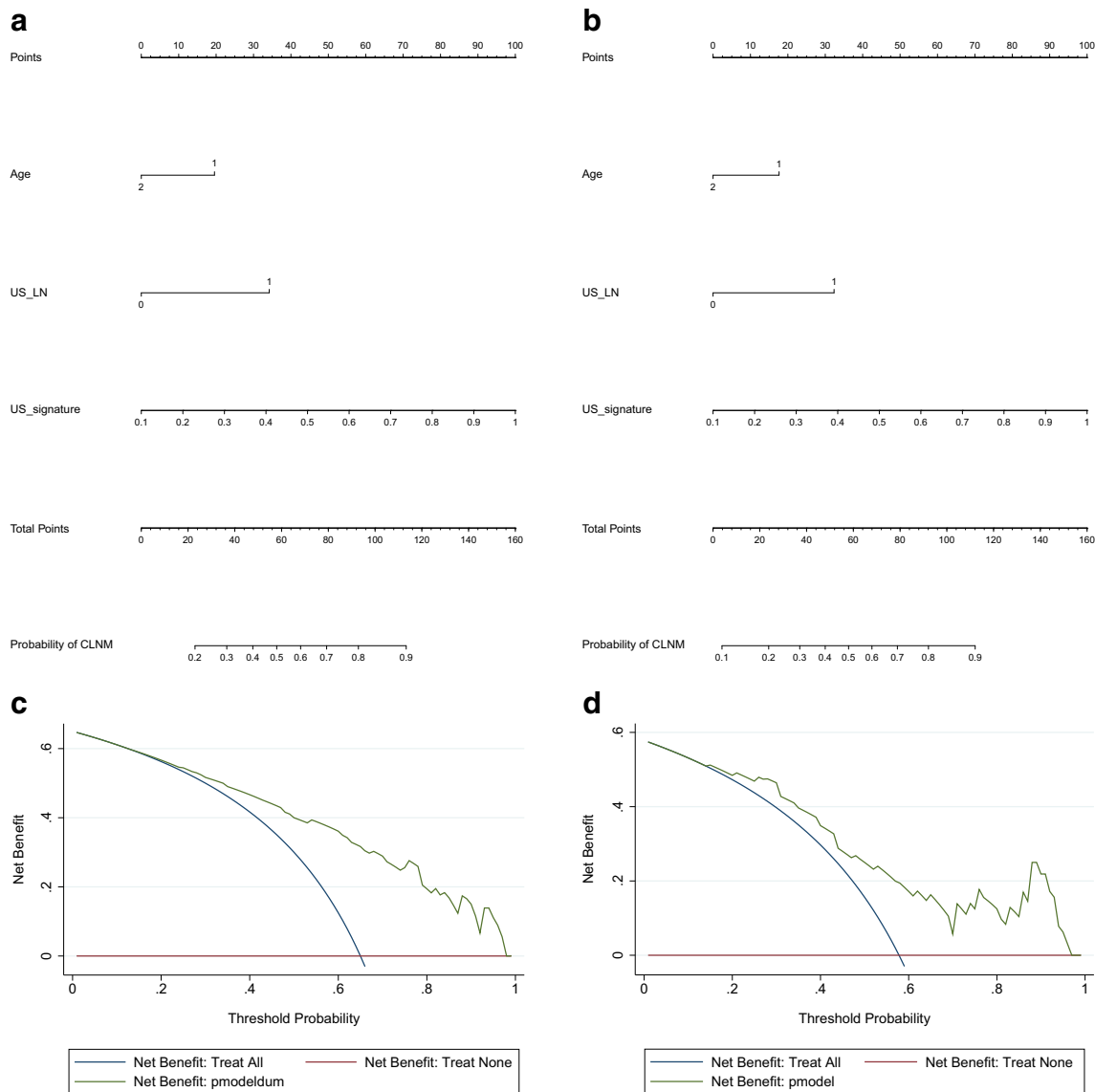


Fig. 6 Nomogram for predicting CLNM in male (a) and female patients (b). The ultrasound nomogram was developed in the training cohort, with age, US-reported LN status and US signature incorporated. The different values for each variable correspond to a point at the top of the graph;

points for all variables are added and translated into the probability of CLNM. Decision curve analysis of the ultrasound nomogram for male (c) and young female patients (d). The x-axis represents the threshold probability, and the y-axis represents the net benefit

therefore, attention should be paid to the identification of microcalcification.

In our research, CDFI was significantly different between the CLNM-positive and CLNM-negative groups; the richer the blood supply is, the higher is the probability of CLNM [40, 41]. Regardless, the evaluation of US for internal vascularity is unreliable and easily influenced by the operator and the machine. This may be why CDFI was not selected in the model building. In addition, multifocality was not included; it showed only limited statistical significance, largely because microscopic lesions smaller than 1 mm could be seen only by microscopic examination, and the identification of tumour multifocality is highly dependent on the radiologist.

Because of the heterogeneity in US image acquisition and clinical data collection in different institutions, we applied decision curve analysis [42] instead of multi-institutional prospective validation to justify its clinical usefulness.

Incorporating the US signature and clinical risk factors into an easy-to-use nomogram facilitates preoperative individualised prediction of CLNM. This nomogram may help to answer questions such as whether CLNM exists, and this may affect the surgical strategy. We suggest that patients with a high score are potential candidates for CLND. The clinical use of the nomogram may guide clinicians in stratifying patients and thereby avoid unnecessary surgery.

Despite the good results, there are still some limitations. First, the performance of our nomogram depends on the accuracy of operator-reported imaging features. The criteria used to evaluate the US signature were subjective. Nevertheless, the interobserver agreement for each feature in our study was good. Second, the pathological subtype of PTC was not taken into account. In addition, owing to its retrospective design, there was a potential for selection bias. Lastly, stringent external validation needs to be performed in larger, prospective multicentre clinical trials to obtain a more objective conclusion. The model may also be improved by the addition of more useful technologies such as elastography and computer-aided diagnosis system [43, 44], and we intend to investigate this in the future.

Conclusion

In conclusion, this study presents a nomogram based on clinical characteristics and US imaging features, and this easy-to-use scoring system can be conveniently applied to facilitate preoperative individualised prediction of CLNM in PTC patients, which is in line with the current trend towards personalised treatment.

Acknowledgements The National Natural Science Foundation of China (NSFC), the Natural Science Foundation of Beijing, China (BJNSF) and the Innovation Cultivation Fund Project of the Seventh Medical Center of the PLA General Hospital are gratefully acknowledged. Yukun Luo,

Qing Song, Ling Ren, Ying Zhang, Fang Xie, Yan Zhang, Jie Tang, Yaqiong Zhu and Zhuang Jin are gratefully acknowledged.

Funding information This study received funding from the National Natural Science Foundation of China (81771834 and 81901746); the Natural Science Foundation of Beijing, China (7194318); and the Innovation Cultivation Fund Project of the Seventh Medical Center of the PLA General Hospital (2015-LC-05).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Yukun Luo.

Conflict of interest The authors of this manuscript declare no relationship with any company whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology

- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

References

1. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. *CA Cancer J Clin* 69:7–34
2. Lee YK, Hong N, Park SH et al (2019) The relationship of comorbidities to mortality and cause of death in patients with differentiated thyroid carcinoma. *Sci Rep* 9:11435
3. Zhao L, Pang P, Zang L et al (2019) Features and trends of thyroid cancer in patients with thyroidectomies in Beijing, China between 1994 and 2015: a retrospective study. *BMJ Open* 9:e023334
4. Wu X, Li B, Zheng C, He X (2018) Risk factors for central lymph node metastases in patients with papillary microcarcinoma. *Endocr Pract* 24:1057–1062
5. Xu Y, Xu L, Wang J (2018) Clinical predictors of lymph node metastasis and survival rate in papillary thyroid microcarcinoma: analysis of 3607 patients at a single institution. *J Surg Res* 221:128–134
6. Feng JW, Yang XH, Wu BQ, Sun DL, Jiang Y, Qu Z (2019) Predictive factors for central lymph node and lateral cervical lymph node metastases in papillary thyroid carcinoma. *Clin Transl Oncol*. <https://doi.org/10.1007/s12094-019-02076-0>
7. Lee J, Song Y, Soh EY (2014) Central lymph node metastasis is an important prognostic factor in patients with papillary thyroid microcarcinoma. *J Korean Med Sci* 29:48–52
8. Haugen BR, Alexander EK, Bible KC et al (2016) 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
9. Conzo G, Tartaglia E, Avenia N et al (2016) Role of prophylactic central compartment lymph node dissection in clinically N0

- differentiated thyroid cancer patients: analysis of risk factors and review of modern trends. *World J Surg Oncol* 14:149
10. Michalaki M, Bountouris P, Roupas ND et al (2016) Low-risk papillary thyroid carcinoma patients who underwent near-total thyroidectomy without prophylactic central compartment lymph node dissection and were ablated with low-dose 50mCi RAI had excellent 10-year prognosis. *Hormones (Athens)* 15:511–517
 11. Ito Y, Miyauchi A, Kudo T, Kihara M, Fukushima M, Miya A (2017) The effectiveness of prophylactic modified neck dissection for reducing the development of lymph node recurrence of papillary thyroid carcinoma. *World J Surg* 41:2283–2289
 12. Viola D, Materazzi G, Valerio L et al (2015) Prophylactic central compartment lymph node dissection in papillary thyroid carcinoma: clinical implications derived from the first prospective randomized controlled single institution study. *J Clin Endocrinol Metab* 100:1316–1324
 13. Aydin OU, Soylu L, Ozbas S et al (2016) The risk of hypoparathyroidism after central compartment lymph node dissection in the surgical treatment of pT1, N0 thyroid papillary carcinoma. *Eur Rev Med Pharmacol Sci* 20:1781–1787
 14. Takami H, Ito Y, Okamoto T, Onoda N, Noguchi H, Yoshida A (2014) Revisiting the guidelines issued by the Japanese Society of Thyroid Surgeons and Japan Association of endocrine surgeons: a gradual move towards consensus between Japanese and western practice in the management of thyroid carcinoma. *World J Surg* 38:2002–2010
 15. Cho SJ, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH (2019) Diagnostic performance of CT in detection of metastatic cervical lymph nodes in patients with thyroid cancer: a systematic review and meta-analysis. *Eur Radiol* 29:4635–4647
 16. Lee DH, Kim YK, Yu HW, Choi JY, Park SY, Moon JH (2019) Computed tomography (CT) for detecting cervical lymph node metastasis in patients who have papillary thyroid microcarcinoma with tumor characteristics appropriate for active surveillance. *Thyroid*. <https://doi.org/10.1089/thy.2019.0100>
 17. Liu Z, Xun X, Wang Y et al (2014) MRI and ultrasonography detection of cervical lymph node metastases in differentiated thyroid carcinoma before reoperation. *Am J Transl Res* 6(2):147–154
 18. Hwang HS, Orloff LA (2011) Efficacy of preoperative neck ultrasound in the detection of cervical lymph node metastasis from thyroid cancer. *Laryngoscope* 121:487–491
 19. Stulak JM, Grant CS, Farley DR et al (2006) Value of preoperative ultrasonography in the surgical management of initial and reoperative papillary thyroid cancer. *Arch Surg* 141:489–494 discussion 494–496
 20. Yeh MW, Bauer AJ, Bernet VA et al (2015) American Thyroid Association statement on preoperative imaging for thyroid cancer surgery. *Thyroid* 25:3–14
 21. Siddiqui S, White MG, Antic T et al (2016) Clinical and pathologic predictors of lymph node metastasis and recurrence in papillary thyroid microcarcinoma. *Thyroid* 26:807–815
 22. Wang Y, Guan Q, Xiang J (2018) Nomogram for predicting central lymph node metastasis in papillary thyroid microcarcinoma: a retrospective cohort study of 8668 patients. *Int J Surg* 55:98–102
 23. Nie X, Tan Z, Ge M (2017) Skip metastasis in papillary thyroid carcinoma is difficult to predict in clinical practice. *BMC Cancer* 17:702
 24. Adler DD, Carson PL, Rubin JM, Quinn-Reid D (1990) Doppler ultrasound color flow imaging in the study of breast cancer: preliminary findings. *Ultrasound Med Biol* 16:553–559
 25. Fisher SB, Perrier ND (2018) The incidental thyroid nodule. *CA Cancer J Clin* 68:97–105
 26. Wu LM, Gu HY, Qu XH et al (2012) The accuracy of ultrasonography in the preoperative diagnosis of cervical lymph node metastases in patients with papillary thyroid carcinoma: a meta-analysis. *Eur J Radiol* 81:1798–1805
 27. Sauerbrei W, Royston P, Binder H (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 26:5512–5528
 28. Kramer AA, Zimmerman JE (2007) Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med* 35:2052–2056
 29. Thompson AM, Turner RM, Hayen A et al (2014) A preoperative nomogram for the prediction of ipsilateral central compartment lymph node metastases in papillary thyroid cancer. *Thyroid* 24:675–682
 30. Oh HS, Park S, Kim M et al (2017) Young age and male sex are predictors of large-volume central neck lymph node metastasis in clinical N0 papillary thyroid microcarcinomas. *Thyroid* 27:1285–1290
 31. Xu SY, Yao JJ, Zhou W, Chen L, Zhan WW (2019) Clinical characteristics and ultrasonographic features for predicting central lymph node metastasis in clinically node-negative papillary thyroid carcinoma without capsule invasion. *Head Neck* 41:3984–3991
 32. Ito Y, Miyauchi A, Kihara M, Higashiyama T, Kobayashi K, Miya A (2014) Patient age is significantly related to the progression of papillary microcarcinoma of the thyroid under observation. *Thyroid* 24:27–34
 33. Wang F, Zhao S, Shen X et al (2018) BRAF V600E confers male sex disease-specific mortality risk in patients with papillary thyroid cancer. *J Clin Oncol* 36:2787–2795
 34. Al Afif A, Williams BA, Rigby MH et al (2015) Multifocal papillary thyroid cancer increases the risk of central lymph node metastasis. *Thyroid* 25:1008–1012
 35. Huang XP, Ye TT, Zhang L et al (2018) Sonographic features of papillary thyroid microcarcinoma predicting high-volume central neck lymph node metastasis. *Surg Oncol* 27:172–176
 36. Zhao L, Yan H, Pang P et al (2019) Thyroid nodule size calculated using ultrasound and gross pathology as predictors of cancer: a 23-year retrospective study. *Diagn Cytopathol* 47:187–193
 37. Bai Y, Zhou G, Nakamura M et al (2009) Survival impact of psammoma body, stromal calcification, and bone formation in papillary thyroid carcinoma. *Mod Pathol* 22:887–894
 38. Ferreira LB, Lima RT, Bastos A et al (2018) OPNa overexpression is associated with matrix calcification in thyroid cancer cell lines. *Int J Mol Sci*. <https://doi.org/10.3390/ijms19102990>
 39. Pyo JS, Kang G, Kim DH, Park C, Kim JH, Sohn JH (2013) The prognostic relevance of psammoma bodies and ultrasonographic intratumoral calcifications in papillary thyroid carcinoma. *World J Surg* 37:2330–2335
 40. Jang JY, Kim DS, Park HY et al (2019) Preoperative serum VEGF-C but not VEGF-A level is correlated with lateral neck metastasis in papillary thyroid carcinoma. *Head Neck* 41:2602–2609
 41. Tu DG, Chang WW, Jan MS, Tu CW, Lu YC, Tai CK (2016) Promotion of metastasis of thyroid cancer cells via NRP-2-mediated induction. *Oncol Lett* 12:4224–4230
 42. Vickers AJ, Cronin AM, Elkin EB, Gonen M (2008) Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 8:53
 43. Xu JM, Xu XH, Xu HX et al (2016) Prediction of cervical lymph node metastasis in patients with papillary thyroid cancer using combined conventional ultrasound, strain elastography, and acoustic radiation force impulse (ARFI) elastography. *Eur Radiol* 26:2611–2622
 44. Lee JH, Ha EJ, Kim JH (2019) Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. *Eur Radiol* 29:5452–5457