




Prediction of Lymph Node Metastasis in Patients With Papillary Thyroid Carcinoma: A Radiomics Method Based on Preoperative Ultrasound Images

Technology in Cancer Research & Treatment
Volume 18: 1-13
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1533033819831713
journals.sagepub.com/home/tct


Tongtong Liu, MS^{1,2}, Shichong Zhou, MD^{3,4}, Jinhua Yu, PhD^{1,2}, Yi Guo, PhD^{1,2}, Yuanyuan Wang, PhD^{1,2}, Jin Zhou, MD^{3,4}, and Cai Chang, PhD^{3,4}

Abstract

Background: Papillary thyroid carcinoma is a type of indolent tumor with a dramatically increasing incidence rate and stably high survival rate. Reducing the overdiagnosis and overtreatment of papillary thyroid carcinoma is clinically emergent and important. A radiomics model is proposed in this article to predict lymph node metastasis, the most important risk factor of papillary thyroid carcinoma, based on noninvasive routine preoperative ultrasound images. **Methods:** Four hundred fifty ultrasound manually segmented images of patients with papillary thyroid carcinoma with lymph node status obtained from pathology report were enrolled in our retrospective study. A radiomics evaluation of 614 high-throughput features were calculated, including size, shape, margin, boundary, orientation, position, echo pattern, posterior acoustic pattern, and calcification features. Then, combined feature selection strategy was used to select features with the greatest ability to discriminate lymph node status. A support vector machine classifier was employed to build and validate the prediction model. Another independent testing cohort was used to further evaluate the performance of the radiomics model. **Results:** Among 614 radiomics features, 50 selected features most reflecting echo pattern, posterior acoustic pattern, and calcification showed the superior lymph node status distinguishable performance with area under the receiver operating characteristic curve of 0.753, 0.740, and 0.743 separately when using each type of features predicting the lymph node status. The results of model based on all 50 final features predicting the lymph node status shown an area under the receiver operating characteristic curve of 0.782, and accuracy of 0.712. In the independent testing cohort, the proposed approach showed similar results, with area under the receiver operating characteristic curve of 0.727 and accuracy of 0.710. **Conclusion:** Papillary thyroid carcinoma with lymph node metastasis usually shows a complex echo pattern, posterior region homogeneity, and macrocalcification or multiple calcification. The radiomics model proposed in this article is a promising method for assessing the risk of papillary thyroid carcinoma metastasis noninvasively.

Keywords

radiomics, ultrasound images, papillary thyroid carcinoma, lymph node metastasis, head and neck cancer

Abbreviations

AUC, area under the receiver operating characteristic curve; ACC, accuracy; CDFI, Color Doppler Flow Images; CT, computed tomography; DCA, decision curve analysis; FDG-PET, fluorodeoxyglucose–positron emission tomography; FNA, fine-needle aspiration; FP, false-positive; GA, genetic algorithm; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length

¹ Department of Electronic Engineering, Fudan University, Shanghai, China

² Key Laboratory of Medical Imaging, Computing and Computer-Assisted Intervention, Shanghai, China

³ Department of Ultrasound, Fudan University Shanghai Cancer Center, Shanghai, China

⁴ Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

Corresponding Authors:

Jinhua Yu, PhD, and Yi Guo, PhD, Department of Electronic Engineering, Fudan University, No. 220, Handan Road, Yangpu District, Shanghai 200433, China.
Emails: jhyu@fudan.edu.cn; guoyi@fudan.edu.cn



matrix; HT, Hashimoto thyroiditis; LN, lymph node; LND, lymph node dissection; MRI, magnetic resonance imaging; mRMR, minimum-redundancy-maximum-relevance; PTC, papillary thyroid carcinoma; PTMC, papillary thyroid microcarcinoma; SD, standard deviation; SENS, sensitivity; SPEC, specificity; SRC, sparse representation classification; SVM, support vector machine; TP, true-positive; US, ultrasound; WRST, Wisconsin rank sum test.

Received: May 22, 2018; Revised: December 6, 2018; Accepted: January 23, 2019.

Introduction

The incidence of thyroid cancer, especially papillary thyroid carcinoma (PTC), has increased dramatically over recent decades around the world; however, the mortality rate has remained stable. In papillary thyroid microcarcinoma (PTMC), defined as a tumor 1 cm or less in size, the mortality rate is even <1%.¹ Recently, the incidence and mortality rates of thyroid cancer in China have been reported as 9.0% and 0.68%, respectively. Among these cases, over 80% are PTC, whose 10-year survival exceeds 90%.¹⁻⁴ In only a few cases, PTC/PTMC will spread to cervical or lateral lymph nodes (LNs) and threaten the survival of patients.^{5,6} Due to the lack of discriminate features, it is very difficult to distinguish them when making a clinical diagnosis.

Lymph node metastasis is the most important risk factor related to recurrence and poor overall survival in patients with PTC.⁶ Due to the low detection rate of central cervical LN metastasis by ultrasound (US), fine-needle aspiration biopsy (FNA), and prophylactic lymph node dissection (LND) of central cervical LNs are performed in suspicious patients to prevent LN metastasis, and these procedures are invasive and unnecessary for most patients.^{5,7,8} Therefore, a noninvasive, efficient, and accurate way to identify patients with a high probability of LN metastasis by thyroid US examination before surgery is urgently needed.^{2,5,9,10}

Previous research has shown that in univariate and multivariate analyses, the characteristics of patients, such as age and gender, and the US features of thyroid tumors, including tumor size, capsule invasion, and microcalcification, are independent predictors of LN metastasis in patients with PTC ($P < .05$).^{2,9-16} Wu *et al* aimed to identify independent estimative factors of LN metastasis in PTC and found that size, vascularization, and coexisting Hashimoto thyroiditis (HT) were significant factors ($P = .004, .118, .016$) based on Color Doppler Flow Images (CDFIs).¹⁷ Nie *et al* compared computed tomography (CT) and US in identifying predictors of lateral LN metastasis in patients with PTC. The results showed that in the univariate analysis, age, tumor size, tumor spread, extrathyroidal extension, primary tumor location, and central LN status were significantly associated with LN metastasis ($P < .05$).¹⁸ Although these findings illustrate the feasibility of estimating LN metastasis using thyroid US images, CDFI, and CT, they are based on clinician experience and visual inspection.

Radiomics is an emerging technique that turns medical images into mineable data by extracting high-throughput features quantitatively. Mineable information that can be

discovered includes pathology, biomarkers, genomic information, and prognosis.¹⁹ For example, it has been used for estimating LN metastasis in lung cancer through a joint fluorodeoxyglucose-positron emission tomography (FDG-PET) and magnetic resonance imaging (MRI) texture-based model.²⁰ Additionally, a radiomics nomogram for the preoperative prediction of LN metastasis in patients with colorectal cancer by CT has shown good discrimination and calibration.²¹ Moreover, LN metastasis in patients with bladder cancer has been preoperatively predicted by radiomics features extracted from arterial-phase CT images with favorable estimative accuracy.²²

The aim of our study is to develop a radiomics model to predict the LN status of patients with PTC using preoperative thyroid US images. Our LN status prediction system includes 4 components: US image manual segmentation, radiomics features extraction, feature selection, and classification and analysis.

Materials and Methods

Overview

In this study, we proposed a radiomics evaluation based on preoperative US thyroid images to predict LN status noninvasively in patients with PTC. The radiomics evaluation including feature extracting of 10 type features, 3-step feature selection, and machine learning method. The workflow of our study is shown in Figure 1.

Patients

We performed a retrospective study of 1216 patients who had been diagnosed in Fudan University Shanghai Cancer Center. The study was approved by the ethics committee of the hospital. Informed consents of each patients were signed. The inclusion criteria were as follows: (1) patients having a single thyroid malignant lesion or one thyroid malignant lesion with several benign lesions, and diagnosed with PTC; (2) patients who underwent thyroid US diagnosis, LND, and US-guided FNA of the first time; and (3) patients with follow-up information and pathological examination results available in their medical records. Details of the patient enrolment are shown in Figure 2.

After that, 450 of 1216 patients with PTC were enrolled in our study. All the 450 cases were sorted by time period and were divided into a validation cohort and an independent testing cohort at a ratio of 2:1. The validation cohort consisted of

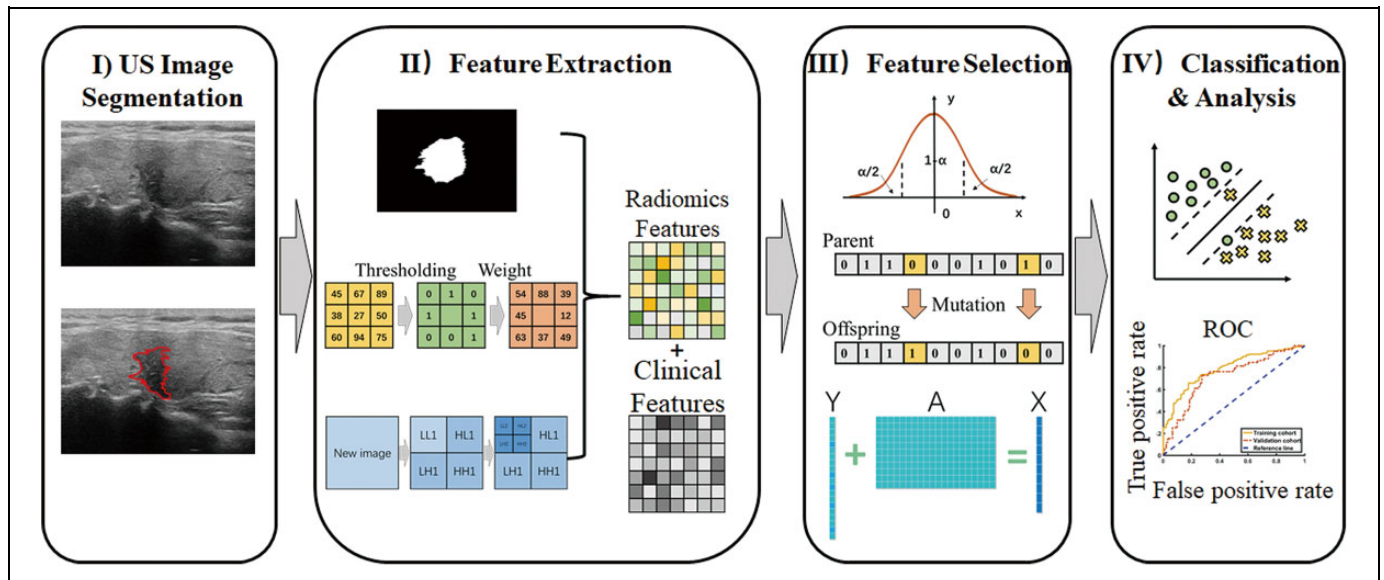


Figure 1. Lymph node (LN) status prediction system workflow. (I) US image were manually segmented. (II) Radiomics features including morphology, texture, and wavelet features were extracted from thyroid US segmented images. (III) A 3 step feature selection method was used to reduce the dimension. (IV) A support vector machine was used to build the final prediction model. US, ultrasound.

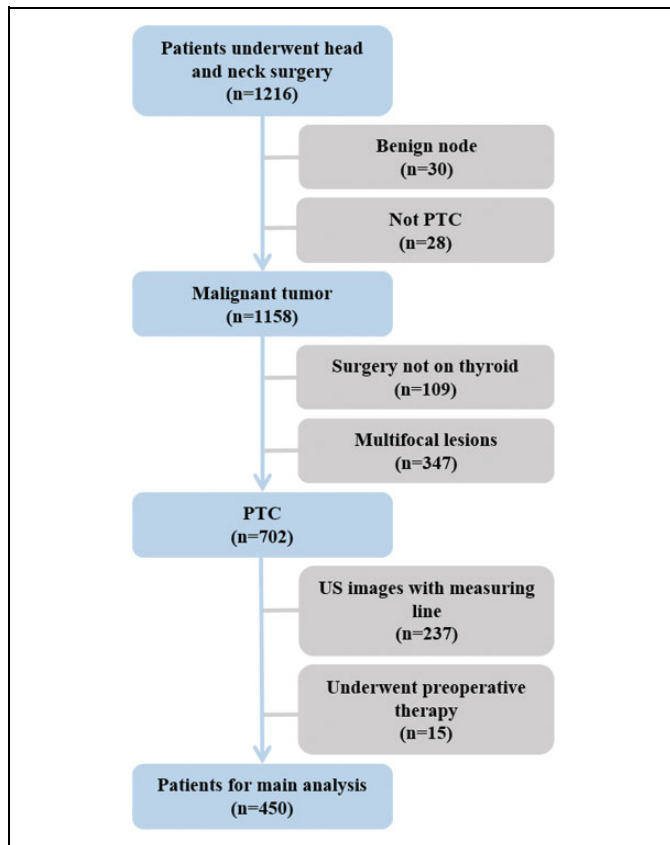


Figure 2. The results of patient enrollments. In total, 450 of 1216 patients were enrolled in this study.

300 patients collected from December 2015 to March 2016 (86 males and 214 females; mean age: 42.6 ± 11.7 years; range: 18-74 years), while 150 patients collected from March 2016 to

September 2016 formed an independent testing cohort (38 males and 112 females; mean age: 43.7 ± 12.2 years; range: 23-82 years). The independent testing cohort was used to evaluate the performance of the prediction model. A summary of the patient characteristics is shown in Table 1.

The demographic information, including age and sex, and the clinical information, including the US examination report and pathology report, were all derived from medical records.

Ultrasound Imaging and Segmentation

Real-time US examinations of thyroid glands were conducted by radiologists with at least 5 years of experience. The US thyroid images were acquired using an ultrasonic instrument from Philips (HD15), Siemens (ACUSON2000), or General Electric (LogiqE9) with 5 to 15-MHz linear transducer L12-5, 14L5, and ML6-15 separately with similar setting about gain and frequency. The spatial resolution of axial and lateral was 0.2 mm and 0.4 mm respectively. All US thyroid images were recorded, in which the tumor was manually segmented by a clinician with 10 years of experience. A MATLAB software written by our group was used in manual segmentation.

Thyroid US was performed by different senior doctors. A representative image of the lesion was saved as a DICOM file.

Standard of LN Status

Then, LND surgery and US-guided FNA were performed by doctors with at least 5 years of experience on patients with PTC who were suspected to have a malignant mass after US examination of the thyroid and LNs. Mass tissue was obtained from LND surgery and US-guided FNA of suspected lesions for histological examination, and pathology report was recorded.

Table 1. Patient Characteristics of Patients in Validation Cohort and Independent Testing Cohort.

Characteristic	Validation Cohort			Independent Testing Cohort		
	Lymphatic Node Metastasis		<i>P</i>	Lymphatic Node Metastasis		<i>P</i>
	Yes	No		Yes	No	
Age, mean (SD)	39.1 (11.2)	44.2 (11.7)	<.001	40.5 (12.6)	45.5 (11.4)	.012
Sex (Case %)			.347			.898
Male	36 (37.9%)	50 (24.4%)		21 (36.2%)	17 (18.5%)	
Female	59 (62.1%)	155 (75.6%)		37 (63.8%)	75 (81.5%)	
Tumor grade			–			–
TI-RADS 4	81	195		45	87	
TI-RADS 5	13	2		12	3	
TI-RADS 6	1	8		1	2	
Total number of LN's positive	366	0	–	314	0	–
Total number of LN's removed	1101	921	–	728	397	–
Total	95	205		58	92	
	300			150		

Abbreviations: LN, lymph node; SD, standard deviation.

The gold standard used in our study was the LN status gathered from the pathology report in the clinical records of the patients.

Fine-needle aspiration was performed for suspected lesions using a 10-mL syringe and 30-gauge needle 3 times under US guidance. One conventional smear was obtained from each aspirated sample, while the remaining aspirated material was deposited directly into a vial of alcohol-based preservative solution and sent for cytopathological diagnosis. Cells found in either the smear or liquid-based preservative solution were diagnosed by pathologists, and patients underwent surgery if they were diagnosed as category IV, V, or VI according to the Bethesda diagnostic system.⁸ Computed tomography examination of the neck was performed before surgery for each patient. Dissection of central cervical LNs was performed routinely. If CT or US indicated suspicious LNs in the lateral cervix, lateral cervical LND was performed. The pathological results of the LNs read by 2 senior pathologists with experiences more than 10 years were used as the gold standard for diagnosis and were documented in the clinical records of the patients.

Radiomics Evaluation

We proposed a radiomics evaluation for evaluating thyroid tumors. We referred to 3 guidelines, including those of the American Association Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi, and summarized their descriptors into 614 high-throughput features.⁸ The radiomics evaluation is shown in Supplemental Table E1. Features in this system can be divided into 10 categories: demographic information and tumor size, shape, orientation, position, margin, boundary, echo pattern, posterior acoustic pattern, and calcification. All features were extracted from the original US thyroid images. All image and data processing was performed in MATLAB R2015b (MathWorks, Inc, Natick, Massachusetts).

Demographic information. The demographic information, including age and sex, was gathered from patient medical records.

Size. The size of a thyroid tumor includes 3 features: the diameter of an equivalent circle, the area, and the ratio of the tumor area to that of its convex hull.²³ Because of the diversity in the depth of each US image, size features needed to be rescaled from pixel-based to real-space dimensions.

Shape. Thyroid tumors can be round, oval, or irregular.⁸ Our system used 3 features to describe the shape of a tumor: the convexity-to-tumor ratio, the compactness, and the rectangle-fitting factor.²³

Margin. Tumor margins can be smooth, irregular, lobulated, spiculate, or obscure.⁸ We used 8 features to represent margin: spiculation, extreme point number, lobule number, moment difference, edge roughness, acutance, local window mean, and acutance 2.²³

Boundary. Tumor boundaries can be characterized by a mutational interface or halo ring.⁸ Herein, the tumor boundary was described by 5 features: the deviation ratio of the inside and outside of the tumor, the mean contrast correlation coefficient of the inside and outside of the tumor, the standard deviation (SD) of the inside and outside of the tumor, and the SD and signal-to-noise ratio of the annular region.²³

Orientation. Resent research has shown that a taller-than-wide tumor shape is a very important indicator of malignancy and LN metastasis.⁸ Hence, we used 3 features, elliptical-normalized eccentricity, elliptical-normalized angle, and length-to-width ratio, to represent tumor orientation.²³

Position. Two types of position features were used, the first being the tumor location relative to the thyroid gland and the thyroid capsule.^{7,8,13,14} Thyroid tumors may appear in 7

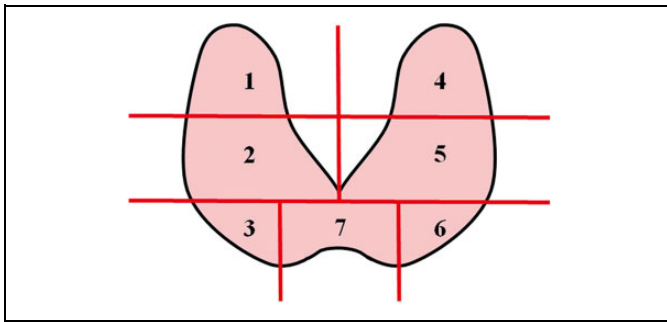


Figure 3. Diagrammatic sketch of different position in thyroid. 1: upper of the right lobe third, 2: mid of the right lobe third, 3: low of the right lobe third, 4: upper of the left lobe third, 5: mid of the left lobe third, 6: low of the left lobe third, and 7: isthmus.

different locations of the thyroid gland, as shown in Figure 3. Information regarding these 7 positions was gathered from the medical records. The second feature type comprised overlap length, overlap area, and distance to capsule. The calculations of these features are shown in Figure 4A.

Echo pattern. The echo pattern of each tumor could be hyper-echoic, isoechoic, hypoechoic, complex, or very low⁸ and can be described by 3 aspects. First, the mean echo value can be quantified by 4 features: the mean tumor contrast, mean tumor covariance, mean tumor nonsimilarity, and mean contrast correlation coefficient. Second, the relative echo intensity of the tumor compared with normal tissue can be represented by 3 features: the deviation ratio of the tumor tissue and normal thyroid gland, the relative brightness of the tumor and normal tissue, and the relative brightness of tumor and normal muscle. Regions of normal thyroid tissue and muscle were marked manually. Third, the texture complexity represented by the tumor contrast SD, tumor covariance SD, tumor nonsimilarity SD, contrast correlation coefficient SD, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size zone matrix, and neighborhood gray-tone difference matrix were also used to examine the tumor texture.²³

Posterior acoustic pattern. The posterior acoustic pattern could be indifferent, attenuated, enhanced, mixing, or shadowing or show a large or small comet-tail artifact (≤ 1.0 mm).⁸ In our system, we first defined the posterior region, as shown in Figure 4B. The features used to describe the posterior acoustic pattern can be divided into 3 parts: the mean value of the posterior region, the relative intensity of the posterior region, and the complexity of the texture, which reflects the texture distribution. The relative intensity and complexity of the posterior region were defined by comparing the posterior region with normal adjacent tissue at the same depth in terms of the relative mean, relative brightness, contrast variance, contrast coefficient, and SD. In addition, the mean and SD of the contrast, mean and SD of the covariance, and mean and SD of the nonsimilarity in the posterior region were used to represent the complexity of texture in our system.²³

Calcification. Calcification is important in analyzing thyroid tumors⁷ and can be present as microcalcification, macrocalcification, or eggshell calcification. Microcalcification appears as very small (<1 mm) hyperechoic regions, macrocalcification appears as larger (>1 mm) hyperechoic regions, and eggshell calcification is annular calcification in the tumor margin that looks like an eggshell.⁸ We used 5 features to describe the size or shape of calcification: total, minimum, and maximum calcification area, total calcification circumference, and roundness SD. Five other features were used to reflect the distribution of calcification: the number of calcification points, the SD of the calcification area, the SD of the circumference of calcification points, and the maximum and minimum distance between calcification points. These features were extracted from the fourth layer of reconstructed wavelet decomposition images, and the calcification boundary was defined by an automatic segmentation algorithm based on wavelet decomposition and threshold segmentation.²⁴

Machine Learning

Validation, feature selection, and image classification. The relation between the LN status and high-throughput radiomics features was explored by a machine learning method. To reduce bias and overfitting in our study, we used 10-fold cross-validation with 100 bootstrap in validation cohort.

A 2-dimensional visualization technique called heatmap was used to analyze the Spearman correlation matrix. In heatmap, the rows represented features, and columns represented patients. The unsupervised clustered result of high-throughput features was shown on the top row of heatmap. The real metastasis status was shown in the second row. Numerical data values were represented by colors in this graphical indication of high-throughput feature performance.²⁵ The heatmap was drawn by R (version 3.4.0 <http://www.Rproject.org>) with the “pheatmap” package.

Due to the repeatability and redundancy of radiomics features, a 3-step feature selection method was employed to reduce the dimensions of the feature set.²² We performed this technique in the validation cohorts for each bootstrap repetition. All the 3-steps feature selection and image classification were performed in MATLAB R2015b (MathWorks, Inc).

First, a 2-sided Wisconsin rank sum test (WRST) was used to select features highly related to LN status. The WRST is a nonparametric hypothesis test. It always be used to check whether or not elements in 2 independent samples correspond to the same distribution.²⁶ We used the function “ranksum” in Matlab R2015b (Mathworks, Inc) to do WRST between the data of LN positive and negative.

Second, we used a genetic algorithm (GA) combining with minimum-redundancy-maximum-relevance (mRMR) to remove the redundant features by selecting approximately half of the features with high value of fitness function.^{27,28}

Genetic algorithm, which is based on Darwinian theory of evolution, can be applied to solve optimization problems. The main concepts of GA to simulate evolution include

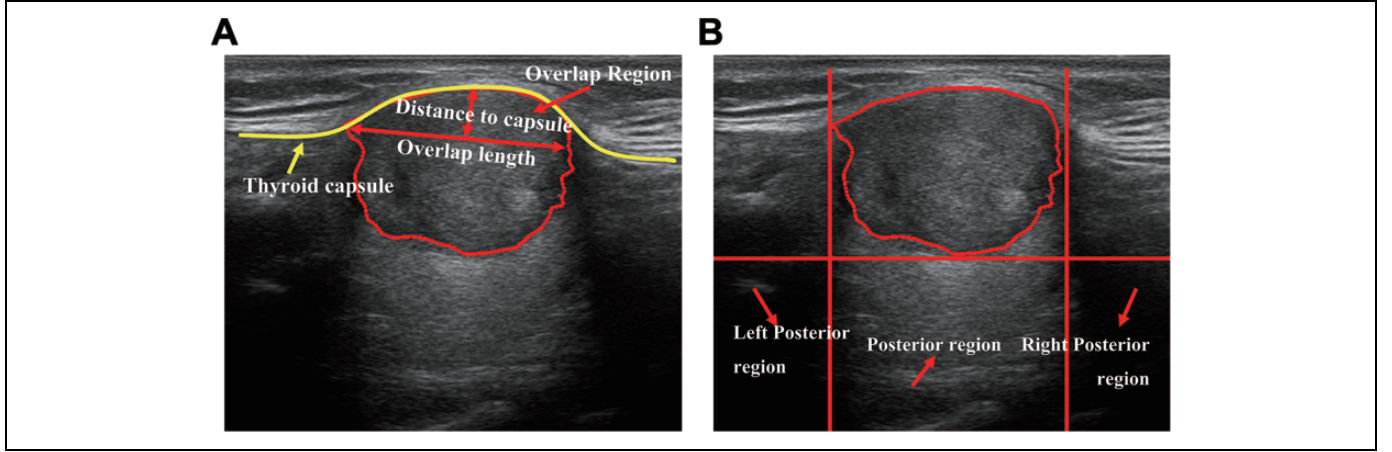


Figure 4. Diagram of 2 examples of extracting features (A) the location of tumor relative to the thyroid capsule, and (B) the location of tumor relative to the thyroid capsule.

reproduction, crossover, mutation, and so on, by rearrange the position of chromosome. Fitness function can represent the simulated living environment by the means of evolutionary simulation and suitable species generation. It can be represented as:

$$\text{Fitness}_{\text{mRMR}}(x) = \text{Accracy}(x), \quad (1)$$

where x denotes a chromosome. Afterward, a new generation with the best combination of gene can be obtained by the completion of GA according to the expressions of reproduction, crossover, and so on.

However, the fitness function in Equation 1 neglected the relationship between chromosomes. Therefore, a new objective function was proposed to evaluate the fitness of a chromosome, as followed:

$$\text{Fitness}_{\text{mRMR}}(x) = \text{Accracy}(x) + 1 - \frac{\text{Rank}(x)}{1 + \text{Accracy}(x)}, \quad (2)$$

where $\text{Rank}(x)$ denotes the sum of mRMR order number.²⁸

Third, sparse representation classification (SRC) was used to identify the most relevant features according to the importance index generated by each bootstrap iteration, in which the top 25% were preserved as remarkable features.²⁹

Because not all the features are relevant to classification, and different medical image modalities contain same texture information, features extracted from medical images are highly redundant. This problem can be solved by SRC.

The basic assumption of sparse representation is that most natural signal, which been considered as a linear combination of a few atoms from an over complete dictionary, is sparse and can be sparsely represented. The SRC are based on sparse representation. It considers the relevance between features and teachers under the influence of the other features. The model of SRC can be formulated as followed:

$$\hat{\phi} = \underset{\phi}{\text{argmin}} \|l - F\phi\|_2^2 + \eta\|\phi\|_0, \quad (3)$$

where l denotes the label set of sample, F denoted the feature set of sample, and η is regularization parameter. The absolute value of sparse representation coefficient $\hat{\phi}$ denotes the importance of the corresponding feature. After calculating the element of $\hat{\phi}$, all the values are ranked in descending order. And features corresponding to the low-ranked values will be removed as redundant features.²⁹

The remaining features were sorted by their occurrence frequency in the 100 bootstrap. The top 1% to 25% of the remaining features were input into a classifier.

Support vector machine (SVM) is one of the most commonly used classifier in medical researches. It can search for the optimal hyperplane in high-dimensional space by using kernel function.³⁰ The learning problem of SVM can be written as follows:

$$\min_{\omega, b, \xi} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_i, \quad (4)$$

$$\text{s.t. } y_i(\omega x_i + b) \geq 1 - \zeta_i, i = 1, 2, \dots, N, \quad (5)$$

$$\xi_i \geq 0, i = 1, 2, \dots, N, \quad (6)$$

where ω is weight vector, b is bias vector, ξ is slack variable, and C is penalty parameter. Besides, $x_i \in X \in R^n$, $y_i \in Y \in R^n, i = 1, 2, \dots, N$ where x_i is the i th feature and y_i is the label of x_i .

The performance was assessed by the area under the receiver operating characteristic curve (AUC), accuracy (ACC), sensitivity (SENS), and specificity (SPEC). The combination of features with the best performance was preserved as the final feature set. We used the final feature set in the validation cohort to train the prediction model by the SVM.

Independent testing. The prediction model was implemented on a separate independent testing cohort to explore its stability and generalization. The performance of the independent testing model was also assessed by the AUC, ACC, SENS, and SPEC.

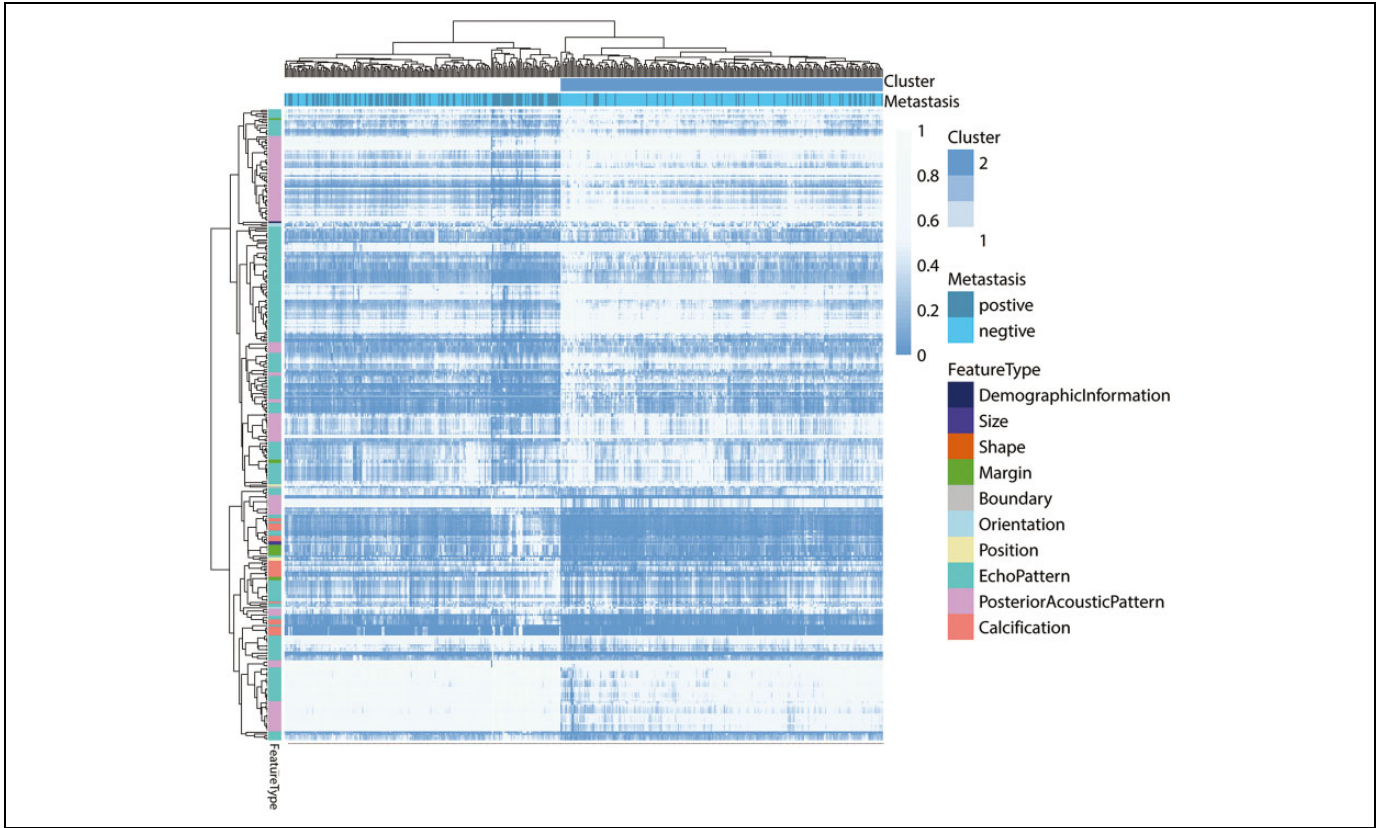


Figure 5. Heatmap of the radiomics features in our system. Cluster: result of unsupervised cluster on high-throughput features. Metastasis: the real status of LN. Feature type: the subtype of features in our radiomics evaluation. LN, lymph node.

Clinical Usefulness

The decision curve analysis (DCA) was conducted for determining the clinical usefulness of the radiomics evaluation.³¹ The y-axis of decision is net benefit, and the x-axis is different threshold probabilities. Net benefit can be calculated as followed:

$$\text{Net benefit} = \frac{\text{TP count}}{N} - \frac{\text{FP count}}{N} \left(\frac{P_t}{1 - P_t} \right), \quad (7)$$

where *TP count* and *FP count* is the number of patients with true-positive (TP) and false-positive (FP) results, *N* is the total number of patients, and P_t is the threshold probability given by predict model.

The DCA was done by R (version 3.4.0 <http://www.Rproject.org>) with the “Decision Curve” and “rmda” packages.

Results

Patient Information

The demographic and clinical data were acquired from the medical records of patients and included age, sex, tumor position, and LN status. In the validation cohort, the mean and SD age of patients with or without LN metastasis were significantly different at 39.1 ± 11.2 years and 44.2 ± 11.7 years,

respectively ($P < .001$). The same result was found in the independent testing cohort, in which the age of patients with and without LN metastasis was 40.5 ± 12.6 years and 45.5 ± 11.4 years, respectively ($P = .012$).

Regarding sex, 59 (62.1%) females and 36 (37.9%) males with LN metastasis were found in the validation cohort, while 155 (75.6%) females and 50 (24.4%) males were without LN metastasis ($P = .347$). The independent testing cohort also showed no significant difference in this characteristic ($P = .898$). Hence, sex was not related to LN status.

It was worth mentioning that there were no significant differences between the validation and independent testing sets in terms of demographic information, clinical information, or features of the radiomics evaluation.

Prediction of LN Status Based on Radiomics Evaluation in Validation Cohort

In our radiomics evaluation, 614 high-throughput features were extracted, consisted of 10 feature types. The heatmap based on the Spearman correlation matrix is shown in Figure 5. In this figure, white and gray represent stronger correlations between the features and LN status, while blue represents weaker correlations. In addition, the echo pattern and posterior acoustic pattern comprised the majority of the features and are clustered in blocks. In general, the heatmap shows some blue blocks,

Table 2. Performance of Predicting LN Status by the Radiomics Evaluation.

Cohort Type		Feature Selection Method	Feature Number	AUC	ACC	SENS	SPEC
Validation cohort	Feature selection	Origin	614	0.741 (0.690-0.791)	0.689	0.717	0.627
		WRST	mean 400	0.736 (0.685-0.787)	0.682	0.706	0.630
		GA_mRMR	mean 200	0.735 (0.685-0.789)	0.683	0.706	0.632
		SRC	75	0.754 (0.702-0.807)	0.699	0.723	0.646
	Final	–	50	0.782 (0.731-0.833)	0.712	0.674	0.730
Independent testing cohort	–	–	50	0.727 (0.653-0.801)	0.710	0.656	0.745

Abbreviations: ACC, accuracy; AUC, area under the receiver operating characteristic curve; GA, genetic algorithm; LN, lymph node; mRMR, minimum-redundancy-maximum-relevance; SRC, sparse representation classification; SPEC, specificity; WRST, Wisconsin rank sum test.
 Note: The results with Italic value had the highest AUC.

indicating the presence of a correlation between radiomics features and LN status, which requires further investigation.

In the 100 bootstrap, each time after the WRST with a significance level of .05 ($P < .05$) set as the threshold, the feature set was reduced to approximately 400. Then, by using the feature selection method of the GA combined with mRMR, the dimension was further reduced to approximately 200. The parameters of GA are as followed: population size = 50, max generation number = 30, crossover probability = 0.9, and mutation probability = 0.1. Furthermore, SRC was used to sort the features by their importance with $\tau = 0.004$, and the top 75 features were selected. The remaining features sorted by their occurrence frequency in the 100 bootstrap were processed by the SVM. We used radial basis function as the kernel function of SVM, and C was adjusted separately for the minority and the majority class by their number.

The top 50 features showed the best performance and were preserved as the final feature set. The mean performance of the final feature set implemented in the validation cohort with 100 bootstrap was AUC = 0.782, ACC = 0.712, SENS = 0.674, and SPEC = 0.730, as shown in Table 2. The receiver operating characteristic curve is shown in Figure 6. The number of each type features was shown in Figure 7. It shown that after feature selection, each type of features have at least one features remained.

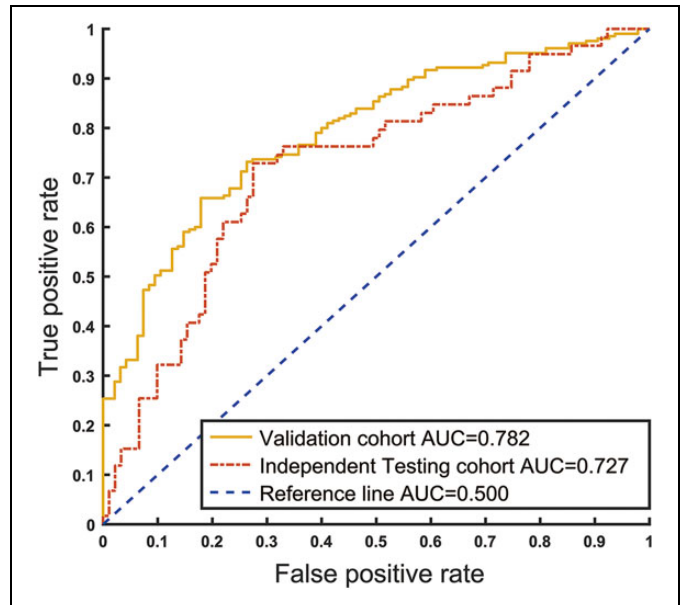


Figure 6. The ROC curves of our radiomics evaluation in validation and independent testing cohorts. The yellow line is the curve of validation cohort with AUC = 0.782, the red line is the curve of validation cohort with AUC = 0.727, and the blue line is the reference line with AUC = 0.500. AUC indicates area under the receiver operating characteristic curve; ROC, receiver operating characteristic.

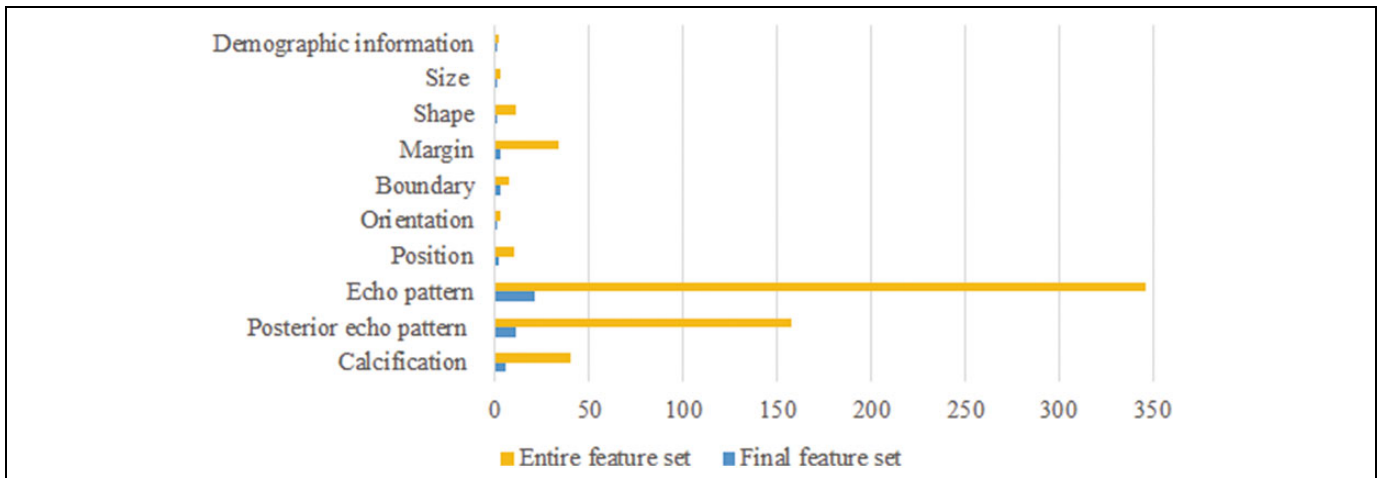


Figure 7. Number of each type features before and after feature selection.

Table 3. Performance of Predicting LN Status by Different Types of Features.

Feature Type	Feature Number	AUC	ACC	SENS	SPEC
Demographic information	1	0.610	0.565	0.674	0.515
Size	1	0.558	0.532	0.611	0.495
Shape	1	0.566	0.490	0.759	0.365
Margin	3	0.690	0.665	0.571	0.709
Boundary	3	0.625	0.702	0.393	0.846
Orientation	1	0.630	0.536	0.847	0.392
Position	2	0.644	0.683	0.432	0.800
Echo pattern	21	0.753	0.701	0.711	0.696
Posterior acoustic pattern	11	0.740	0.703	0.725	0.692
Calcification	6	0.743	0.687	0.677	0.691
All	50	0.783	0.711	0.679	0.725

Abbreviations: ACC, accuracy; AUC, area under the receiver operating characteristic curve; LN, lymph node; SPEC, specificity.

Prediction of LN Status by Feature Type

The prediction performance of each feature type in the final feature set is detailed in Table 3. The overall performance of the echo pattern was the best, with AUC = 0.753, ACC = 0.701, SENS = 0.711, and SPEC = 0.696. The sensitivity of orientation was the highest with SENS = 0.847, showing potential to identify positive cases. Regarding size, shape, demographic information, orientation, and position, these feature types only had 1 or 2 features in the final feature set; thus, their AUC values were all below 0.65 even though these features all satisfied the $P < .05$ condition. As expected, the model including all feature types outperformed those based on only one category, and the combination achieved the highest AUC of 0.783 and ACC of 0.711.

Lymph Node Status Prediction Based on the Final Feature Set in the Independent Testing Cohort

An independent testing cohort with 150 cases was used to examine the performance of our radiomics model. The final feature set and the SVM prediction model were applied to predict LN status in the independent testing cohort. The performance metrics are shown in Table 2, with AUC = 0.727, ACC = 0.710, SENS = 0.656, and SPEC = 0.745. The results indicated that our model was accurate in both the validation and independent testing cohorts.

Clinical Application

The decision curve of the radiomics evaluation on the final feature set are shown in Figure 8. The decision curve shown that if the threshold probability was 10% to 85%, using the radiomics evaluation to predict LN status added more benefit than treat all patients or treat none patients which assume all LN status as positive or negative.

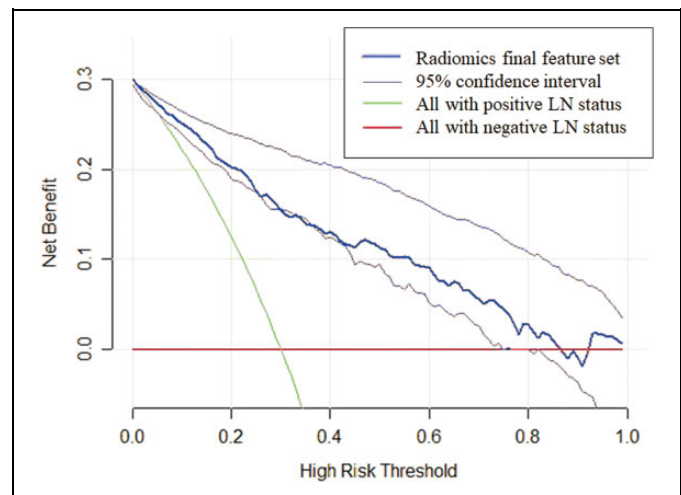


Figure 8. Decision curve analysis for the radiomics evaluation of the final feature set. The dark blue line is the decision curve of radiomics final feature set. The light blue line is the 95% confidence interval decision curve of radiomics final feature set. The green line is the decision curve for treat all patients assuming LN status as positive. The red line is the decision curve for treat none patients which assumes LN status as negative. LN indicates lymph node.

Discussion

Lymph Node Metastasis Evaluation in Current Studies

Over the past few years, the global incidence of PTC has increased significantly, while the mortality of PTC has remained stable. The sharp increase in the detection rate of PTC is because of the increased resolution of high-frequency US examination and the greater prevalence of accurate physical examination. The state of PTC in China is similar to the global situation. Although in most cases of PTC, especially PTMC, the disease progresses slowly and does not threaten the life of patients, there are still a few instances in which PTC will progress significantly, such as LN metastasis and extra-thyroidal extension. Lymph node metastasis is an important indicator in evaluating the progression of PTC, and it is a predictor of poor outcome, as determined by multivariate analysis.¹⁷ Therefore, the key to avoiding PTC progression is to distinguish the small fraction of tumors that are dangerous and have the potential for LN metastasis from all the other inert tumors that progress slowly.

However, the detection rate of central cervical LN metastasis by US examination is very low as shown in our previous study, with SENS = 0.148, SPEC = 0.940, and ACC = 0.662.³² In current clinical practice of China, the LN status is first screened and diagnosed by US examination. Then, US-guided FNA biopsy and CT are conducted for highly suspicious nodes or LNs. Subsequently, prophylactic LND of central and lateral cervical LNs is performed for patients with suspicious node during clinical diagnosing and operating to prevent serious outcomes.^{5,7,8} Nevertheless, LND may increase the risk of patients having parathyroidism and nerve injury. In addition, whether LND can improve the survival rate of patients with

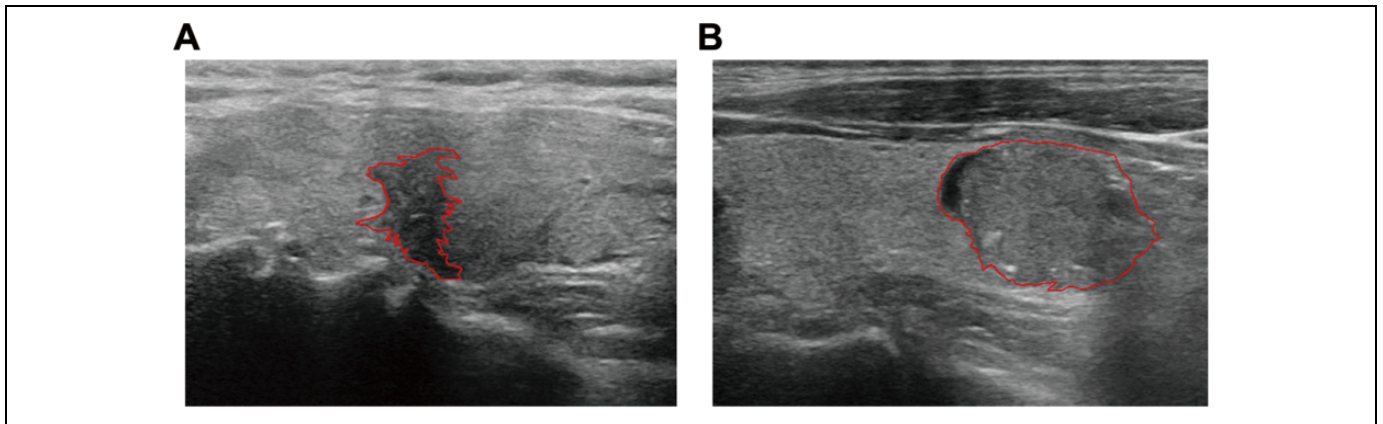


Figure 9. Examples of US thyroid images of cases with different LN status. A, US images with depth of 4.0 cm of whose LN status was positive of a 45-year-old female patient. The size of the tumor was $7 \times 11 \times 7$ mm and it was taller-than-wide, and the tumor has an irregular shape, a spiculate margin, an obscure boundary, an echo pattern heterogeneity, a mixed posterior echo pattern, multiple calcification, and located in the middle of the left lobe third. B, US images with depth of 3.5 cm of whose LN status was negative of a 44-year-old male patient. The size of the tumor was $17 \times 14 \times 18$ mm and the tumor was parallel, with an oval shape, an irregular margin, a mutation boundary, an echo pattern homogeneity, an enhanced posterior echo pattern, little calcification, and located in the low of the left lobe third. LN indicates lymph node; US, ultrasound.

PTC is debatable, necessitating clinical consideration prior to use in patients. Additionally, performing follow-up examinations for all patients with PTC is another way to detect LN metastasis earlier. However, follow-up examinations are costly and time consuming, and they are inconvenient for both patients and clinicians. All of those weakness prevent their wide use for detecting LN metastasis.

Many previous studies based on US have focused on identifying predictors for LN metastasis in patients with PTC in order to offer guidelines for clinicians. Gomez *et al* concentrated on US characteristics and found that calcification ($P = .007$) and size ($P = .003$) were associated with LN metastasis in patients with PTC.¹⁵ Wang *et al* concluded that the significant factors of a multivariate analysis were age <45 years, larger size, “wider-than-tall” shape, extrathyroid extension, and mixed flow ($P < .05$).¹⁶ These results can serve as references in clinical practice.

Although the studies mentioned above are encouraging, all of the results were assessed by P value, which is an index than can reflect relevant differences between 2 factors. In addition, and most importantly, the features used by those studies were based on clinician experience and visual inspection without repeatability.

Although US examination is the first choice for diagnosing thyroid disease in clinical practice, the examination has serious limitations. Papillary thyroid carcinoma frequently metastasizes to the lateral or central regional compartment of the LNs. Metastatic nodes in the lateral cervical compartment are easily detected by US examination, although some can be missed by radiologist carelessness. In clinical practice, if lateral LN metastasis is detected during US examination, the patient will undergo LND. However, metastatic nodes in the central cervical compartment are not so easily detected by US examination because they were obscured by thyroid and nearby tissue.

Therefore, patients with malignant FNA findings may require LND of central cervical LNs to prevent LN metastasis. Thus, LND for lateral or central cervical LN metastasis should be considered carefully before surgery to avoid overtreatment.³

Our Findings and the Advantages of Radiomics Analysis

Recently, the use of machine learning in medicine has had great success. Models built by machine learning can be considered objective observers that will consider a matter from the same viewpoint at a given level. Radiomics, which sometimes applies machine learning methods, has recently attracted the interest of many researchers for exploring the associations between diagnostic and prognostic information with quantitative medical image features.²⁰⁻²² This approach has also been used to predict LN metastasis in other types of cancer. Vallières *et al* proposed the use of an FDG-PET and MRI texture-based model for estimating LN metastasis in lung cancer.²⁰ Huang *et al* proposed a radiomics nomogram, which incorporated a 3-item radiomics signature, carcinoembryonic antigen status, and CT-reported LN status, for the preoperative prediction of LN metastasis via CT in patients with colorectal cancer.²¹ Wu *et al* proposed the preoperative prediction of LN metastasis in patients with bladder cancer through a radiomics nomogram that incorporated a radiomics signature and the CT-reported LN status.²² The above successes suggested the feasibility of applying radiomics to predict the LN status in PTC by US.

According to the analytical results of our proposed radiomics model, US images of PTC with and without LN metastasis present different radiomics signatures. Two typical US images of thyroid tumors with different LN statuses are shown in Figure 9. The results showed that LN metastasis was associated with a younger age, a larger tumor size, an oval or irregular tumor shape, a spiculate margin, an obscure

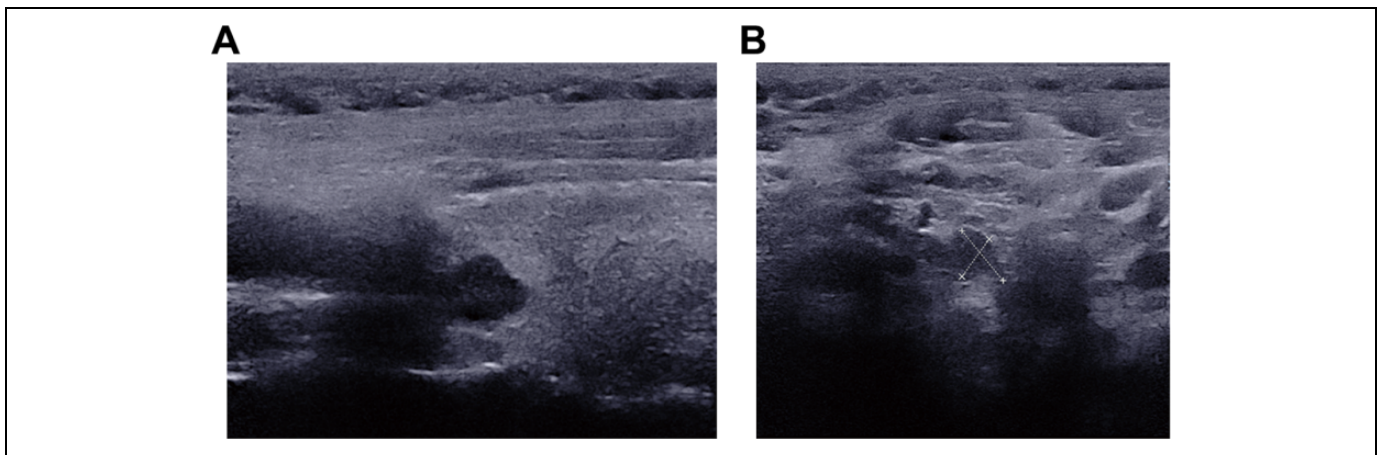


Figure 10. The US image of a 45-year-old female patient with PTC with pathological confirmed LNM after surgery. A, The primary thyroid tumor US image of the patient. B, The 0.77×0.71 cm non-enlarged lymph node US image of the patient. PTC, papillary thyroid carcinoma; US, ultrasound.

boundary, a taller-than-wide shape, thyroid invasion, a complex echo pattern, posterior region homogeneity, and macrocalcification or multiple calcification. In contrast, a negative LN status in patients with PTC was related to an older age, smaller tumor size, rounder tumor shape, a wider-than-tall tumor shape, a location inside the thyroid region, a smooth margin, a mutational boundary, echo pattern homogeneity, a complex posterior echo pattern, and little calcification. All features in the final feature set performed well in the radiomics evaluation, although their performance varied by feature type. As shown in Table 3, the performance of 3 feature types, including echo pattern, posterior acoustic pattern, and calcification, alone could distinguish between patients with and without LN metastasis. The remaining 7 feature types, including demographic information and tumor size, shape, orientation, position, margin, and boundary, did not yield a performance as high as the other ones. Details of the *P*, mean, and SD values in patients with or without LN metastasis for the 50 features in the final feature set are illustrated in Supplemental Table E2. These results are similar to those found for distinguishing malignant and benign breast cancer tumors, except for the echo pattern of the posterior region.³³

There is an example of a 45-year-old patient who were indicated no enlarged lymph nodes by US report but the radiomics model predicted metastases, later confirmed after surgery, as shown in Figure 10. Doctors couldn't determine whether the LN is malignant or not just depend on the US image of the LN and the primary thyroid tumor. But the radiomics evaluation predicted that the LN status of the patient is positive by just using the US image of the primary tumor. And the pathological report based on tissue gathered from surgery confirmed the result.

As for the stability of those radiomics features, the study of Hu *et al* shows that the segmentation results, machine models, and machine setting including gain and frequency can affect the stability of quantitative features. However, some of those features were robust, including morphological

features, intensity features, and GLCM features, while most features were insensitive to machine settings.²³ In the final feature set of 50 features in our study, some features such as GLRLM variance, were proved to be stable and cannot be easily influenced.

Our study has some limitations. First, previous studies based on multimodal techniques we mentioned above showed the applicability of multimodality images in our study. Different modalities focused on different aspects of human organs and functions, making multimodality images contain more information. Secondly, diffuse thyroid uptake such as HT often represents benign nodes but looks similar to the texture of malignants. Therefore, it could change texture feature of thyroid, and influence the stability of our model. Thirdly, recent studies shown that using multicenter data can make model more robust than just using data from single institution. Besides, big data are a required principle of radiomics for mining concealed prognosis information to avoid overfitting. Once overcome those problems, the SENS and SPEC of the model in our study may be improved.

Conclusions

The radiomics evaluation proposed in this article has potential to predict LN status noninvasively in patients with PTC based on preoperative US thyroid images. Patients with US images showing a complex echo pattern, posterior region homogeneity, and macrocalcification or multiple calcification are more likely to have LN metastasis. This LN status prediction model has the potential to facilitate early medical management and alleviate overdiagnosis.

Authors' Note

Tongtong Liu and Shichong Zhou contributed equally to this work. The study was approved by Fudan University Shanghai Cancer Institutional Review Board. The Ethics Committee reference number is No.1506169-6-NSFC.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the National Natural Science Foundation of China (61471125, 8162780) and the Science and Technology Commission of Shanghai Municipality (17411953400).

ORCID iD

Tongtong Liu, MS  <https://orcid.org/0000-0002-4112-7089>

Yi Guo, PhD  <https://orcid.org/0000-0002-7142-2871>

Supplemental Material

Supplemental material for this article is available online.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics (2017). *CA Cancer J Clin.* 2017;67(1):7-30.
2. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115-132.
3. Siddiqui S, White MG, Antic T, et al. Clinical and pathologic estimators of lymph node metastasis and recurrence in papillary thyroid microcarcinoma. *Thyroid.* 2016;26(6):807-815.
4. Toniato A, Boschini I, Casara D, Mazzarotto R, Rubello D, Pelizzo M. Papillary thyroid carcinoma: factors influencing recurrence and survival. *Ann Surg Oncol.* 2008;15(5):1518-1522.
5. Haugen BR, Alexander EK, Bible KC, 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.
6. Liu Z, Wen Z, Liu C, et al. Diagnostic accuracy of ultrasonographic features for lymph node metastasis in papillary thyroid microcarcinoma: a single-center retrospective study. *World J Surg Oncol.* 2017;15(1):32-36.
7. Frates MC, Benson CB, Charboneau JW, et al. Management of thyroid nodules detected at US: society of radiologists in ultrasound consensus conference statement. *Radiology.* 2005;237(3):794-800.
8. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for clinical practice for the diagnosis and management of thyroid nodules—2016 update. *Endocr Pract.* 2016;22(5):1-60.
9. Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Maso LD. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med.* 2016;375(7):614-617.
10. Wada N, Duh QY, Sugino K, et al. 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.
11. Roh JL, Kim JM, Park CI. Central lymph node metastasis of unilateral papillary thyroid carcinoma: patterns and factors estimative of nodal metastasis, morbidity, and recurrence. *Ann Surg Oncol.* 2001;18(8):2245-2250.
12. So YK, Son YI, Hong SD, et al. Subclinical lymph node metastasis in papillary thyroid microcarcinoma: a study of 551 resections. *Surgery.* 2010;148(3):526-531.
13. Smith VA, Sessions RB, Lentsch EJ. Cervical lymph node metastasis and papillary thyroid carcinoma: does the compartment involved affect survival? Experience from the SEER database. *J Surg Oncol.* 2012;106(4):357-362.
14. Yang Y, Chen C, Chen Z, et al. Prediction of central compartment lymph node metastasis in papillary thyroid microcarcinoma. *Clin Endocrinol.* 2014;81(2):282-288.
15. Gomez NR, Kouniavsky G, Tsai HL, et al. Tumor size and presence of calcifications on ultrasonography are pre-operative estimators of lymph node metastasis in patients with papillary thyroid cancer. *J Surg Oncol.* 2011;104(6):613-616.
16. Wang QC, Cheng W, Wen X, Li J, Jing H, Nie C. Shorter distance between the nodule and capsule has greater risk of cervical lymph node metastasis in papillary thyroid carcinoma. *Asian Pac J Cancer Prev.* 2014;15(2):855-860.
17. Wu Q, Li Y, Wang Y, Hu B. Sonographic features of primary tumor as independent estimative factors for lymph node metastasis in papillary thyroid carcinoma. *Clin Transl Oncol.* 2015;17(10):830-834.
18. Nie X, Tan Z, Ge M, Jiang L, Wang J, Zheng C. Risk factors analyses for lateral lymph node metastases in papillary thyroid carcinomas: a retrospective study of 356 patients. *Arch Endocrinol Metab.* 2016;60(5):1-8.
19. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2015;278(2):563-577.
20. Vallières M, Freeman CR, Skamene SR, Naqa IE. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastasis in soft-tissue sarcomas of the extremities. *Phys Med Biol.* 2015;60(14):5471-5496.
21. Huang Y, Liang C, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* 2016;34(18):2157-2164.
22. Wu S, Zheng J, Li Y, et al. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res.* 2017;23(22):6904-6911.
23. Hu Y, Qiao M, Guo Y, et al. Reproducibility of quantitative high-throughput BI-RADS features extracted from ultrasound images of breast cancer. *Med Phys.* 2017;44(7):3676-3685.
24. Su X. *Micro Calcification Clusters Detection Algorithms Based on SVM in Mammograms (PhD Dissertation)*. Lanzhou, China: Lanzhou University; 2010.
25. Gu J, Pitz M, Breitner S, et al. Selection of key ambient particulate variables for epidemiological studies—applying cluster and heatmap analyses as tools for data reduction. *Sci Total Environ.* 2012;435:541-550.
26. Perolat J, Couso I, Loquin K, Strauss O. Generalizing the Wilcoxon rank-sum test for interval data. *Int J Approx Reason.* 2015;56(pt A):108-121.

27. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput.* 2002;6(2):182-197.
28. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell.* 2005; 27(8):1226-1238.
29. Li Y, Namburi P, Yu Z, Guan C, Feng J, Gu Z. Voxel selection in fMRI data analysis based on sparse representation. *IEEE Trans Biomed Eng.* 2009;56(10):2439-2451.
30. Wang S, Wei J. Feature selection based on measurement of ability to classify subproblems. *Neurocomputing.* 2017;224:155-165.
31. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565.
32. Zhou S, Liu T, Zhou J, et al. Preliminary study on application of radiomics in thyroid carcinoma. *Oncoradiology.* 2017;26(2):102-105.
33. Mendelson EB, Böhm-Vélez M, Berg WA, et al. ACR BI-RADS® Ultrasound. In: Carl J. D'Orsi, ed. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System.* Reston, VA: American College of Radiology; 2013.