An ultrasound model for predicting recurrence of papillary thyroid carcinoma after complete endoscopic resection

Bin Lu, Yibo Zhou, Xiaofeng Lu, Wenchao Weng, Shengye Wang, Jianlin Lou

1 Department of Ultrasound, Affiliated Jinhua Hospital, Zhejiang University School of Medicine, Jinhua, Zhejiang Province, China
2 Department of Breast and Thyroid Surgery, Affiliated Jinhua Hospital, Zhejiang University School of Medicine, Jinhua, Zhejiang Province, China
3 Department of Radiation Oncology, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang, China
4 Department of Head and Neck Surgery, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang, China

Abstract

Introduction: Papillary thyroid cancer (PTC) is one of the most common malignancies involving the endocrine system. The aim of this study was to explore the clinical value of ultrasound-based radiomics for predicting the recurrence of PTC after complete endoscopic resection.

Material and methods: The general data of 361 PTC patients were collected. They were randomly assigned to the modeling group (n = 253) and the validation group (n = 108) according to the ratio of 7:3. In the modeling group, the PyRadiomics package was applied to extract radiomic features from preoperative ultrasound images, and least absolute shrinkage and selection operator (LASSO) was used to screen and construct a radiomics score (Rad-score). Independent prognostic predictors were identified using the Cox proportional hazards model, and a nomogram prediction model was constructed by R software.

Results: Using the LASSO regression model, 7 radiomic features were screened and then the Rad-score was calculated. Based on the Rad-score, modeling and validation groups were divided into low- and high-risk groups, and the 10-year recurrence-free survival rates were 94.7% vs. 95.9%, 83.6% vs. 80.0%, and 50.0% vs. 66.6%, respectively (p < 0.001). Multivariate analysis revealed that age, lymph node metastasis and Rad-score were independent predictors for recurrence-free survival (p < 0.05).

Conclusions: The ultrasound-based radiomics score can effectively predict the postoperative recurrence-free survival in patients with PTC. The nomogram prediction model is superior to the AJCC staging system in terms of predictive accuracy and consistency.

Key words: thyroid cancer, ultrasonography, radiomics, prognosis, nomogram.

Introduction

Papillary thyroid cancer (PTC) is one of the most common malignancies involving the endocrine system, with a rising incidence rate in recent decades [1–3]. Long-term survival can be obtained for most patients with PTC after radical surgery, but recurrence may result in treatment failure [4, 5]. Thus, accurately identifying the cases at high risk of recurrence contributes importantly to formulation of treatment strategies.
individualized treatment and follow-up protocols by clinicians [6]. The tumor-node-metastasis staging system proposed by the American Joint Committee on Cancer (AJCC) is currently well recognized to evaluate the prognosis of TC patients [7]. However, patients with an identical pathological stage and therapeutic schedule may have markedly different clinical outcomes [8]. In recent years, radiomics, a novel technology capable of extracting numerous quantitative feature data from images through automatic algorithms, has been applied in the diagnosis, staging and prognostic evaluation of diseases [9–11].

**Aim**

The aim of this study was to investigate the efficacy of ultrasound-based radiomics for predicting the prognosis of patients with PTC who underwent complete endoscopic resection, and to compare it with the traditional AJCC staging system.

**Material and methods**

**General clinical data**

The general clinical data of patients who underwent complete endoscopic resection for PTC in our hospital between January 2010 and December 2014 were collected for retrospective case-control analysis. The inclusion criteria were as follows: a) patients who received thyroid ultrasonography before surgery, with available images, b) those diagnosed with PTC by postoperative histopathology, c) those who underwent thyroid surgery for the first time, and d) those with complete pathological data.

The exclusion criteria involved: a) patients with malignancies in other organs, b) those with evidence of distant metastasis revealed by preoperative examination, c) those with tumors incompletely resected, d) those who received thyroid radiofrequency, microwave therapy or head and neck radiotherapy in the past, or e) those with non-papillary cancer components revealed by postoperative histopathology.

A total of 361 patients were included in the present study, including 72 males and 289 females aged 42 (21–79) years on average. Among them, there were 239 patients with tumor diameter \( \leq 2 \) cm and 122 patients with tumor diameter \( > 2 \) cm. Extra-glandular invasion occurred in 74 patients and lymph node metastasis was present in 119 patients. Based on the AJCC staging system (8th edition), there were 289, 54 and 18 patients in stages I, II and III, respectively. According to the ratio of 7 : 3, the patients were assigned to the modeling group \((n = 253)\) and the validation group \((n = 108)\) using a random number table. The baseline data of the two groups were similar \((p > 0.05)\) (Table I).

**Ultrasoundography**

Thyroid ultrasonography was conducted for all patients before the operation using ultrasonic instruments Philips (HD15), Siemens (ACUSON2000) and General Electric (LogiqE8) with 10-15 MHz linear array probes of L12-5, 14L5 and 11L-D respectively. Thyroid lesion images were stored in the Digital Imaging and Communications in Medicine (DICOM) format.

**Construction and evaluation of radiomics prediction model**

**Lesion area delineation**

Using ITK-SNAP software (http://www.itksnap.org), an eligible ultrasound image was selected from each patient by an attending physician who had worked in the ultrasound department for more than 8 years, followed by lesion area delineation. The images should meet the following criteria: a) the lesion with most typical malignant features, b) the relationship between lesion and capsule was displayed, c) the lesion with a maximum diameter, and d) without measuring marks or Doppler imaging. For patients with multiple lesions, delineation should be performed for the largest one [10]. After all lesion areas were delineated, a chief physician with more than 20 years of working experience in the ultrasound department was responsible for evaluating images. Any objection should be processed through re-delineating corresponding areas by the attending physician after consultation. Both physicians were unaware of the patients’ information.

**Extraction of radiomic features**

The PyRadiomics open-source platform (v2.2.0, https://pyradiomics.readthedocs.io/) was employed to extract radiomic features from the lesion area of each patient, including first-order statistics, 2D shape, texture and wavelet features [12]. In total,
1,209 radiomic features were extracted from each lesion area.

Screening of radiomic features and modeling

In the modeling group (n = 253), the most representative radiomic features were selected using the three-step method. Firstly, univariate Cox regression analysis was employed to identify 119 features significantly associated with recurrence-free survival (p < 0.05). Secondly, Pearson’s correlation coefficient between every two features was calculated. After excluding the features hardly affecting recurrence-free survival among those with strong correlations (correlation coefficient: > 0.9), 33 features were screened. Thirdly, 7 radiomic features of the highest prognostic value were identified using least absolute shrinkage and selection operator (LASSO) regression analysis [13, 14]. The radiomics score (Rad-score) was the sum of each radiomic feature multiplied by the corresponding weight coefficient. X-tile software (v3.6.1) was utilized to analyze the optimal cut-off value of Rad-score for predicting recurrence-free survival. Furthermore, a nomogram prediction model was constructed by R software (v4.1.1) combined with Rad-score and clinical pathological factors [15].

Evaluation of prediction model

Harrell’s concordance index was applied to evaluate the prognostic discrimination ability of the prediction model [16]. The calibration curve was plotted to display the consistency between the predicted and actual survival rates. The Akaike information criterion (AIC) was used to compare the predictive ability of models, and a lower AIC value corresponded to higher predictive accuracy [17]. The likelihood ratio χ² test was employed to evaluate the homogeneity of the prediction model, and
a higher likelihood ratio $\chi^2$ value represented better homogeneity [18]. Net reclassification improvement (NRI) was utilized to access the predictive accuracies of different models [19]. Additionally, a decision curve was drawn to compare the clinical benefits of prediction models [20].

**Postoperative follow-up**

Follow-up was carried out through outpatient review or telephone to obtain the information about whether the disease recurred after surgery. The patients were reviewed by physical, laboratory and imaging examinations once every 3–6 months within 1 year after surgery, and then once every 6–12 months until December 2021. Overall survival was defined as the time from the operation to the last follow-up or death.

The endpoint of this study was progression-free survival, which was defined as the time from the operation to the first postoperative recurrence, or the time from the operation to the last follow-up or death if the disease did not recur during follow-up.

**Statistical analysis**

SPSS 22.0 software and R (v4.1.1) software were used for statistical analysis. Numerical data were expressed as percentages, and the $\chi^2$ test was employed for comparison between groups. A Kaplan-Meier survival curve was plotted to compare the recurrence-free survival, and the log-rank test was utilized to determine the significance. The Cox proportional hazards model was used for univariate and multivariate analyses. $P < 0.05$ represented statistically significant differences.

**Results**

**Construction of ultrasound-based radiomics score in modeling group**

In the modeling group ($n = 253$), 7 radiomic features were screened using the LASSO regression model, including original shape 2D PerimeterSurfaceRatio, original shape 2D Elongation, original glcm Autocorrelation, original glszm LGLZE, wavelet-LH glcm ClusterProminence, wavelet-HL ngtdm Contrast, and wavelet-HH glszm GLNN (Figure 1). As a result, the features such as shape, margin and echo pattern of tumor lesions were elucidated, with corresponding weight coefficients of 0.17132882, 0.09810219, 0.11553653, 0.15965165 and 0.09748579, respectively. The Rad-score ranged from –2.14 to 3.31.

**Recurrence-free survival**

The median follow-up time of the 361 included patients was 108 (8–137) months. The 5- and 10-year recurrence-free survival rates of modeling and validation groups were 92.9% vs. 95.4% and 87.4% vs. 89.8%, respectively.

According to X-tile software, the optimal cutoff values of Rad-score for predicting postoperative recurrence were 0.15 and 0.52, and the patients in modeling and validation groups were further divided into the low-risk group (Rad-score < 0.15, $n = 244, 67.5$%), medium-risk group (Rad-score: 0.15–0.52, $n = 80, 22.2$%) and high-risk group (Rad-score > 0.52, $n = 37, 10.3$%). Kaplan-Meier survival analysis showed that the 10-year recurrence-free survival rates were 94.7% vs. 95.9%, 83.6% vs. 80.0%, and

![Figure 1. Screening of radiomic features based on the LASSO regression model](image)

**LASSO – least absolute shrinkage and selection operator.**
50.0% vs. 66.6%, respectively. Different risk groups had significantly different recurrence-free survival ($\chi^2 = 15.805, 63.590, p < 0.001$) (Figure 2).

Factors influencing recurrence-free survival

Univariate analysis revealed that age, tumor diameter, extra-glandular invasion, lymph node metastasis and Rad-score were significantly associated with the recurrence-free survival of the modeling group ($p < 0.05$). Multivariate analysis showed that age, lymph node metastasis and Rad-score were independent influencing factors for the recurrence-free survival of the modeling group ($p < 0.05$) (Table II).

Construction and evaluation of nomogram prediction model

Based on the above-mentioned findings of multivariate analysis, age, lymph node metastasis and Rad-score were selected as the independent prognostic predictors to construct a nomogram model for predicting the prognosis of the modeling group (Figure 3). Harrell’s concordance indices of the model for modeling and validation groups were 0.829 and 0.845, respectively, indicating high predictive accuracies. The calibration curve indicated that the recurrence-free survival predicted by the nomogram model was close to the actual value, suggesting high consistency (Figure 4).

Comparison between nomogram prediction model and AJCC staging system (8th edition)

The AIC values of the nomogram prediction model for modeling and validation groups were 287.02 and 70.35, respectively, which were superior to those of the AJCC staging system (8th edition) (321.54 and 83.92). The likelihood ratio $\chi^2$ values of the nomogram prediction model for modeling and validation groups were 56.07 and 24.65, respectively, also exceeding those of the AJCC staging system (8th edition) (21.56 and 11.08). The results of NRI analysis suggested that compared with the AJCC staging system (8th edition), the predictive accuracies of the nomogram prediction model for modeling and validation groups were augmented by about 65.4% and 43.9%, respectively. Furthermore, the nomogram prediction model was better than the AJCC staging system (8th edition) in terms of clinical benefits (Figure 5).

Discussion

In recent years, the incidence rate of TC has been rising worldwide, and it is predicted that TC will re-
place colorectal cancer as the fourth most common malignancy in 2030 [1, 3]. At present, TC is mainly treated by surgical resection. Since endoscopic thyroidectomy was first completed by Hüscher et al. [21] in the 1990s and thyroid malignancy was removed by Shimizu and Tanaka [22] through the subclavian approach with a small incision for the first time, the safety and feasibility as well as advantages such as minimally invasive technique and aesthetic effects of complete endoscopic thyroidectomy have been demonstrated gradually [23]. PTC is the most common type of TC. Although the prognosis of most patients with PTC is satisfactory, a few tumors are still highly invasive and distantly metastasize after radical resection [24, 25]. Consequently, much attention has been paid to identifying the high-risk factors for postoperative recurrence in such patients and formulating individualized regimens.

Age is recognized as a primary prognostic factor of TC, and tumor-specific death readily occurs in older patients [26, 27]. TC is the only malignancy that includes age in the AJCC staging system, verifying the importance of age for prognostic evaluation [7]. So far, the correlations of age with postoperative recurrence and prognosis of PTC remain elusive, which may be attributed to the association with the BRAF gene carried by patients, as reported by Shen et al. [28]. The disease mortality of patients carrying the BRAF V600E mutation was found to progressively rise with aging. Likewise, age was also an independent predictor of postoperative recurrence in patients with PTC in this study, suggest-

### Table II. Univariate and multivariate Cox regression analysis results of factors influencing recurrence-free survival in 253 patients undergoing complete endoscopic resection of PTC

| Item                        | 10-year recurrence-free survival rate (%) | Univariate analysis | Multivariate analysis |
|-----------------------------|------------------------------------------|---------------------|-----------------------|
|                             |                                          |                     |                       |
| Age [years]:                |                                          |                     |                       |
| < 55                        | 91.7%                                    | 0.001               | 0.955 0.346 7.640 2.599 1.320–5.117 0.006 |
| ≥ 55                        | 76.3%                                    | 0.111               |                       |
| Gender:                     |                                          | 0.111               |                       |
| Male                        | 80.7%                                    |                     |                       |
| Female                      | 89.0%                                    |                     |                       |
| Tumor diameter:             |                                          |                     |                       |
| ≤ 2 cm                      | 91.6%                                    | 0.003               | -0.138 0.404 0.118 0.871 0.395–1.921 0.732 |
| > 2 cm                      | 78.8%                                    |                     |                       |
| Multifocality:              |                                          | 0.432               |                       |
| No                          | 88.1%                                    |                     |                       |
| Yes                         | 84.3%                                    |                     |                       |
| Extra-glandular invasion:   |                                          |                     |                       |
| No                          | 92.0%                                    | 0.001               | 0.570 0.390 2.137 1.769 0.823–3.800 0.144 |
| Yes                         | 68.6%                                    |                     |                       |
| Lymph node metastasis:      |                                          |                     |                       |
| No                          | 91.0%                                    | 0.002               | 1.133 0.351 10.423 3.106 1.561–6.181 0.001 |
| Yes                         | 80.0%                                    |                     |                       |
| Rad-score:                  |                                          |                     |                       |
| Low-risk group              | 94.7%                                    | < 0.001             | 1                     |
| Medium-risk group           | 83.6%                                    | 1.622               | 0.412 15.513 5.064 2.259–11.351 < 0.001 |
| High-risk group             | 50.0%                                    | 2.686               | 0.403 44.390 14.677 6.660–32.348 < 0.001 |

CI – confidence interval, OR – odds ratio, PTC – papillary thyroid cancer, Rad-score – radiomics score.
Figure 3. Nomogram prediction model including age, lymph node metastasis and Rad-score.

Rad-score – Radiomics score.

Figure 4. Calibration curve analysis of nomogram prediction model for modeling and validation groups.

Lymph node metastasis has been closely associated with the postoperative recurrence of PTC [29, 30]. Compared with patients who are pathologically N0 (pN0), the recurrence rate remains significantly higher even in those with micro-infiltration of tumor cells in lymph nodes [31]. In this study, lymph node metastasis was discovered by postoperative pathology in about 1/3 of patients with PTC, and their 10-year recurrence-free survival rate was markedly lower than that of the cases without lymph node metastasis (80% vs. 91%). Hence, standardized and thorough lymph node dissection should be conduct-
An ultrasound model for predicting recurrence of papillary thyroid carcinoma after complete endoscopic resection

Intratumor heterogeneity involves multiple spatial concepts from gene, protein, metabolism to physiology and anatomy, leading to various sensitivities of different individuals or even the same individual to the same treatment regimen [31]. Radiomics can extract numerous imaging features from medical images, allowing clinicians to quantitatively evaluate intratumor heterogeneity from a macroscopic perspective [32, 33] without requiring tissue biopsy or interventional surgery. These image heterogeneity parameters, as special biological markers, can predict the prognosis and treatment outcomes of patients. As reported by Xiong et al. [34], the ultrasound image-based radiomics prediction model was able to effectively predict recurrence-free survival rate in patients with invasive breast cancer. In a study conducted by Jiang et al. [35], the CT image-based radiomics prediction model was remarkably correlated with the clinical outcomes and chemotherapy sensitivity of patients with gastric cancer. In the present study, a radiomics prediction model (Rad-score) based on preoperative thyroid color Doppler ultrasonography was also constructed, according to which the patients were further divided into three risk groups with significant prognostic differences. The nomogram prediction model constructed with age, lymph node metastasis and Rad-score was prominently superior to the AJCC staging system in terms of predictive efficiency.

Conclusions
The ultrasound-based radiomics score is an important predictor for postoperative recurrence in patients with PTC undergoing complete endoscopic resection. By combining the radiomics score with other high-risk clinical factors, the nomogram prediction model can be utilized to formulate individualized treatment and follow-up protocols for patients at different risks. Regardless, this study is limited. This is a retrospective study with a small sample size, so the results may be biased. Further prospective studies with larger sample sizes are still needed to validate the results herein.

Acknowledgments
Bin Lu and Yibo Zhou contributed equally to this study.
This study was financially supported by Zhejiang Public Welfare Technology Application Research Project (No. LGF20H160003).

Conflict of interest
The authors declare no conflict of interest.

References
1. Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the united states. Cancer Res 2014; 74: 2913-21.
1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016; 66: 115-32.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.
3. Shaha AR. Recurrent differentiated thyroid cancer. Endocr Pract 2012; 18: 600-3.
4. MeLeod DS, Sawka AM, Cooper DS. Controversies in primary treatment of low-risk papillary thyroid cancer. Lancet 2013; 381: 1046-57.
5. Zhou SC, Liu TT, Zhou J, et al. An ultrasound radiomics nomogram for predicting disease-free survival of invasive breast cancer. Front Oncol 2021; 11: 621993.
6. van Dien J, van Hagen P, Grotenhuis BA, et al. Comparison of the 6th and 7th Editions of the UICC/AJCC TNM Classification for Esophageal Cancer. Ann Surg Oncol 2012; 19: 2142-5.
7. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008; 8: 53.
8. MeLeod DS, Sawka AM, Cooper DS. Controversies in primary treatment of low-risk papillary thyroid cancer. Lancet 2013; 381: 1046-57.
9. Wang C, Dai L, Wu X, et al. A nomogram for predicting overall-specific survival in thyroid cancer patients with total thyroidec- tomy: a SEER database analysis. Gland Surg 2021; 10: 2546-56.
10. Zhou H, Jin Y, Dai L, et al. Differential diagnosis of benign and malignant thyroid nodules using deep learning radiomics of thyroid ultrasound images. Eur J Radiol 2020; 127: 108992.
11. Xiong L, Chen H, Tang X, et al. Ultrasound-based radiomics analysis for predicting disease-free survival of invasive breast cancer. Front Oncol 2021; 11: 621993.
12. van Dien J, van Hagen P, Grotenhuis BA, et al. Comparison of the 6th and 7th Editions of the UICC/AJCC TNM Classification for Esophageal Cancer. Ann Surg Oncol 2012; 19: 2142-5.
13. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997; 16: 385-95.
14. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J Royal Statist Soc Series B-Statistical Methodol 2011; 73: 273-82.
15. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008; 26: 1364-70.
16. Gonen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika 2005; 92: 965-70.
17. Awad AM. Properties of the Akaike information criterion. Microelectronics Reliability 1996; 36: 457-64.
18. Talsma K, van Hagen P, Grotenhuis BA, et al. Comparison of the 6th and 7th Editions of the UICC/AJCC TNM Classification for Esophageal Cancer. Ann Surg Oncol 2012; 19: 2142-8.
19. Pencina MJ, D’Agostino Sr RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011; 30: 11-21.
20. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008; 8: 53.