Evaluation of Radiomics Models Based on Computed Tomography for Distinguishing Between Benign and Malignant Thyroid Nodules

Dan Kong, MM,* Jiandong Zhang, MM,* Wenli Shan, MM,* Shaofeng Duan, PhD,† and Lili Guo, MD*

Aim: The aim of the study was to investigate the diagnostic value of radiomics models based on computed tomography (CT) in distinguishing between benign and malignant thyroid nodules.

Materials and Methods: We conducted a retrospective analysis of the clinical and imaging data of 172 patients with pathology-confirmed thyroid nodules (83 benign nodules and 89 malignant nodules). All patients underwent a plain CT scan + arterial and venous contrast enhancement before the operation. Using the stratified random sampling method, patients were divided into a training group (121 cases) and a test group (51 cases) at a ratio of 7:3. A K software was used to extract radiomics features from the preoperative CT images, and minimum redundancy maximum relevance and least absolute shrinkage and selection operator regression analyses were then used for feature screening and model construction. Receiver operating characteristic (ROC) curves were constructed for the training and test groups to verify model performance and evaluate the efficacy of the radiomics features in identifying benign and malignant thyroid nodules. We then used the most efficient models to construct a nomogram. For the training group, 1-way analysis of variance and multivariate logistic regression analysis were used to screen statistically significant clinical features, and the radiomics scores were combined to construct a radiomics nomogram. We used ROC curve analysis to evaluate the predictive performance of the model.

Results: Screening yielded 21 radiomics features that were used to construct a model for differentiating between benign and malignant thyroid nodules. For the training group, the area under the ROC curve of the preoperative CT images, and minimum redundancy maximum relevance and least absolute shrinkage and selection operator regression analyses were then used for feature screening and model construction. Receiver operating characteristic (ROC) curves were constructed for the training and test groups to verify model performance and evaluate the efficacy of the radiomics features in identifying benign and malignant thyroid nodules. We then used the most efficient models to construct a nomogram. For the training group, 1-way analysis of variance and multivariate logistic regression analysis were used to screen statistically significant clinical features, and the radiomics scores were combined to construct a radiomics nomogram. We used ROC curve analysis to evaluate the predictive performance of the model.

Conclusions: The radiomics nomogram constructed by combining radiomics characteristics and clinical risk factors was efficacious for distinguishing benign and malignant thyroid nodules.

Key Words: tomography, x-ray computer, thyroid nodules, diagnosis, differentiation, radiomics

(Original Article)

The incidence of thyroid nodules has increased in recent years. Approximately 4% to 5% of thyroid nodules are malignant, mainly consisting of papillary carcinoma. Regular follow-up is needed for benign thyroid nodules, among which a small proportion need surgical treatment (either nodulectomy or less invasive ablation surgery), whereas malignant nodules should be operated on as soon as possible. In addition to lesion removal, some patients also need to undergo corresponding regional lymph node dissection. Ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are currently used for nodule detection. Ultrasound is the most widely used technique, but the diagnosis is overly reliant on experience, resulting in high subjectivity. By comparison, CT and MRI do not produce clear and reliable signs that can be used to differentiate between benign and malignant nodules. The nonspecificity of the diagnosis of benign and malignant nodules based on CT and MRI images results in a misdiagnosis rate as high as 40% to 70%. Therefore, it is very important to develop an accurate method for the differential diagnosis of benign and malignant nodules in follow-up treatment and improve prognosis.

Radiomics is an emerging image analysis technology in cancer treatment, in which high-throughput features are extracted from traditional images to quantify tumor lesions, and deep learning methods are used to mine features that are potentially associated with pathology for disease diagnosis and prognostic evaluation. Zhou et al. predicted lymph node metastasis of thyroid cancer using ultrasound-based radiomics, resulting in an accuracy of 73.1%, a sensitivity of 71.4%, and a specificity of 74.0%. Another study showed no significant difference between ultrasound and multislice spiral CT in the diagnosis of benign and malignant thyroid nodules. Few reports have described the application of thyroid radiomics for CT examinations. We applied CT radiomics to distinguish benign from malignant thyroid nodules in this study.

MATERIALS AND METHODS

Clinical Data
This retrospective study was approved by the hospital ethics committee, and the need for informed consent was waived. We retrospectively collected data from patients with surgically and pathologically confirmed thyroid nodules from May 2019 to August 2020. The following clinical data were obtained for each nodules in follow-up treatment and improve prognosis. Ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are currently used for nodule detection. Ultrasound is the most widely used technique, but the diagnosis is overly reliant on experience, resulting in high subjectivity. By comparison, CT and MRI do not produce clear and reliable signs that can be used to differentiate between benign and malignant nodules. The nonspecificity of the diagnosis of benign and malignant nodules based on CT and MRI images results in a misdiagnosis rate as high as 40% to 70%. Therefore, it is very important to develop an accurate method for the differential diagnosis of benign and malignant nodules in follow-up treatment and improve prognosis.

Radiomics is an emerging image analysis technology in cancer treatment, in which high-throughput features are extracted from traditional images to quantify tumor lesions, and deep learning methods are used to mine features that are potentially associated with pathology for disease diagnosis and prognostic evaluation. Zhou et al. predicted lymph node metastasis of thyroid cancer using ultrasound-based radiomics, resulting in an accuracy of 73.1%, a sensitivity of 71.4%, and a specificity of 74.0%. Another study showed no significant difference between ultrasound and multislice spiral CT in the diagnosis of benign and malignant thyroid nodules. Few reports have described the application of thyroid radiomics for CT examinations. We applied CT radiomics to distinguish benign from malignant thyroid nodules in this study.

MATERIALS AND METHODS

Clinical Data
This retrospective study was approved by the hospital ethics committee, and the need for informed consent was waived. We retrospectively collected data from patients with surgically and pathologically confirmed thyroid nodules from May 2019 to August 2020. The following clinical data were obtained for each
patient: age, sex, and CT characteristics. These latter data included the following: the presence or absence of microcalcification: calcification with a cross-sectional diameter of 2 mm or less; regular or irregular lesion morphology: most benign thyroid nodules grow expansively, and their boundaries are regular and round, oval, or thyroid shape, while most malignant nodules grow infiltratively, and as a result, the limitations of growth rate and surrounding tissues vary with the position, and the lesions are mostly irregular in shape; positive/negative edge interruption sign: also known as bite cake sign, it has been recognized by some scholars for its diagnostic value in thyroid cancer and refers to the spanning of the maximum diameter of the tumor within the junction area between the tumor and the thyroid or outside the thyroid, indicating that the tumor involves or is close to the thyroid capsule; clear/fuzzy lesion boundary after enhancement, defined by the difference in CT values between the lesion and adjacent normal structures: if the difference increases after enhancement, the boundary is clear, and otherwise, it is blurred; abundant cells or follicles, less nodular infarction and vitreous degeneration, obvious enhancement of the tumor body, reduced density difference with the thyroid, and reduced/blurred low-density area of the tumor body for malignant thyroid nodules; richness in fibers, infarcts, and hyaline degeneration for benign nodules: these components have low or no enhancement, the density difference between thyroid and benign thyroid nodules is increased, and the low-density area of the tumor body is clear; positive/negative cystic change ratio greater than 50%: the volume of cystic components accounts for more than one-half of the whole tumor body. There can be varying degrees of fibrosis and fibrous membranes around benign thyroid nodules, which can affect the blood supply of the nodules, resulting in bleeding, necrosis, and cystic change, while malignant nodules, particularly papillary carcinoma, have rich, qualitative fibers and slow growth. Although larger tumors can also be necrotic and cystic, the volume of cystic change is less than 50%, accounting for only 3%. The CT features were reviewed and recorded by 2 experienced imaging physicians. Any disagreement was resolved by consensus.

The following inclusion criteria were used: (1) a thyroid noncontrast scan + enhanced examination conducted within 1 to 15 days before the operation; (2) no history of other tumors; (3) a grade 2 to 4 lesion according to the thyroid imaging report and data system of the preoperative ultrasound examination of the thyroid; (4) lesion diameter greater than 1 cm; and (5) a new diagnosis, with the lesion having never been subjected to biopsy, radiotherapy, or chemotherapy. A total of 278 patients were initially recruited and collected. The exclusion criterion was a previous puncture or operation (25 patients); in addition, 33 patients were complicated with other tumors, and 48 patients had a lesion diameter less than 1 cm. A total of 172 patients were finally included, including 83 with benign nodules and 89 with malignant nodules.

Computed Tomography Examination Method

A Siemens definition 64-slice CT scanner from Germany was used to perform routine noncontrast scans and 2-phase enhanced scans. The tube voltage was 120 kV, and CARE Dose 4D was used. The slice thickness was 2 mm, and the pitch was 0.8. A volume of 60 to 70 mL of the contrast agent ioversol (containing 320 mgI/mL of iodine) was injected into the median elbow vein at an injection rate of 3.0 mL/s, followed by injection of 15 mL of normal saline. The aortic arch was monitored by the contrast agent bolus tracking method, where the trigger threshold was 100 HU, the arterial phase delay time was 10 seconds, and the venous phase was scanned after 25 seconds. Before the scan, patients were instructed to breathe in and then hold their breath with the arms placed at the sides of the body. During the scan, the patients were asked to maintain a supine position with the neck leaning backward and maximally lowered shoulders, and swallowing was prohibited. The scanning range extended from the mandible to the base of the neck. If the thyroid extended behind the sternum, the scanning range was expanded to cover the entire thyroid.

Image Segmentation and Feature Extraction

We used the stratified random sampling method to divide patients in a 7:3 ratio into training and test groups. The training group data were used for feature screening and model construction.

Region of Interest Segmentation

Two experienced radiologists (A and B) used ITK-SNAP (www.itksnap.org) software to delineate the edge of the targeted lesion layer-by-layer on the CT noncontrast scan, arterial-phase images, and venous phase images to synthesize a 3-dimensional (3D) region of interest (ROI). The ROIs were constructed twice by radiologist A over 3 to 6 days and once by radiologist B.

Feature Screening

A.K. software (Version: 3.2.0. R, Artificial Intelligence Ki, GE Healthcare) was used to standardize the triple-phase CT images of the patients, followed by feature extraction.

We used R software (http://www.Rproject.org, Version: 3.4.4) to analyze the data. We first performed consistency tests within and between the observer data sets; that is, we calculated the intraclass correlation coefficient (ICC) between the features extracted from the 2 ROIs constructed by radiologist A and the ICC between the features extracted from the first ROI constructed by radiologist A and the ROI constructed by radiologist B. Features with an ICC greater than 0.75 in both calculations were retained, and the features extracted from the first ROI constructed

| TABLE 1. Clinical Characteristics of Patients in the Training and Test Groups |
| --- |
| Variable | Training (n = 121) | Test (n = 51) | P |
| Age, mean ± SD, y | 46.4 ± 13.4 | 49.5 ± 13.5 | 0.16 |
| Sex, n (%) | | | |
| Male | 26 (21.5%) | 44 (86.3%) | 0.33 |
| Female | 95 (78.5%) | 7 (13.7%) | 0.62 |
| Microcalcification, n (%) | | | |
| Absent | 48 (39.7%) | 23 (45.1%) | |
| Present | 73 (60.3%) | 28 (54.9%) | 1.00 |
| Lesion morphology, n (%) | | | |
| Regular | 89 (73.6%) | 37 (72.5%) | |
| Irregular | 32 (26.4%) | 14 (27.5%) | |
| Edge interruption sign, n (%) | | | |
| Absent | 13 (10.7%) | 9 (17.6%) | 0.32 |
| Present | 108 (89.3%) | 42 (82.4%) | 0.07 |
| Clear after enhancement, n (%) | | | |
| Absent | 37 (30.6%) | 8 (15.7%) | |
| Present | 84 (69.4%) | 43 (84.3%) | 1.00 |
| Cystic-ratio-50p, n (%) | | | |
| Absent | 24 (19.8%) | 10 (19.6%) | |
| Present | 97 (80.2%) | 41 (80.4%) | |
| Radscore | −0.1 (−1.1 to 0.8) | 0 (−1.0 to 1.1) | 0.59 |
by radiologist A were used for subsequent analysis. First, the minimum redundancy maximum relevance was used to remove redundant and irrelevant features. The retained features were imported into a least absolute shrinkage and selection operator (LASSO) regression model. A 10-fold cross-validation method was used to identify the hyperparameter $\lambda$ of the LASSO regression model. The $\lambda$ corresponding to the smallest model error was selected, and features with nonzero coefficients were retained. Regression and dimensionality reduction were used to screen out features with good generalizability, and a prediction model was established. The imaging Radscore of each patient in the training set was calculated as the sum of the product of the features and the corresponding coefficients, and the Wilcoxon rank-sum test was used to compare the imaging Radscore between benign and malignant nodules.

Screening of Clinical Features and Establishment of the Radiomics Nomogram Model

The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the efficacy of the 3-phase scan in identifying benign and malignant thyroid nodules. The optimal phase combined with clinical features was selected to establish the corresponding radiomics nomogram. In the training group, 1-way analysis of variance was used to screen independent clinical features ($P < 0.05$). Multivariate logistic regression was then used in conjunction with radiomics tags to screen the final predictors and construct a radiomics nomogram.

Statistical Analysis

R software was used for statistical analysis. The calibration curve was used to evaluate the radiomics score. All AUCs were reported with 95% confidence intervals (CIs). Continuous variables were tested using the t test or the Wilcoxon rank-sum test, whereas categorical variables were analyzed using the Pearson $\chi^2$ test or Fisher exact test. The Hosmer-Lemeshow goodness-of-fit test was used to create a calibration curve to evaluate the nomogram. A 2-sided $P$ value less than 0.05 was considered significant.

RESULTS

General Information

A total of 172 patients were enrolled, including 33 males and 139 females aged 25.0 to 64.0 (45.3 ± 11.2) years. There were 83 patients with benign nodules, including 50 with thyroid adenoma (24 follicular adenoma, 16 eosinophilic adenoma, 9 atypical adenoma, and 1 with a hyaline variable beam tumor), 29 patients with nodular goiter, 2 patients with granulomatous thyroiditis, and 2 patients with hemangioma. The remaining 89 patients had malignant nodules, including 55 with papillary thyroid carcinoma, 21 with follicular carcinoma, 7 with medullary carcinoma, 1 with undifferentiated carcinoma, 4 with metastasis, and 1 with lymphoma. The size of the included lesions ranged from 1 to 2.4 cm.

Using stratified random sampling to divide the cases in a 7:3 ratio, 121 patients were placed in the training group, and 51 were placed in the test group. There were no significant differences in the sex, age, CT characteristics, or distribution of benign and malignant nodules between the training and test groups ($P > 0.05$; Table 1).

Feature Selection and Radiomics Signature Construction

The 3-phase images of the patients were extracted. First, minimum redundancy maximum relevance was used to remove redundant and irrelevant features. The retained features were imported into a least absolute shrinkage and selection operator (LASSO) regression model. A 10-fold cross-validation method was used to identify the hyperparameter $\lambda$ of the LASSO regression model. The $\lambda$ corresponding to the smallest model error was selected, and features with nonzero coefficients were retained. Regression and dimensionality reduction were used to screen out features with good generalizability, and a prediction model was established. The imaging Radscore of each patient in the training set was calculated as the sum of the product of the features and the corresponding coefficients, and the Wilcoxon rank-sum test was used to compare the imaging Radscore between benign and malignant nodules.

The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the efficacy of the 3-phase scan in identifying benign and malignant thyroid nodules. The optimal phase combined with clinical features was selected to establish the corresponding radiomics nomogram. In the training group, 1-way analysis of variance was used to screen independent clinical features ($P < 0.05$). Multivariate logistic regression was then used in conjunction with radiomics tags to screen the final predictors and construct a radiomics nomogram.

STATISTICAL ANALYSIS

R software was used for statistical analysis. The calibration curve was used to evaluate the radiomics score. All AUCs were reported with 95% confidence intervals (CIs). Continuous variables were tested using the t test or the Wilcoxon rank-sum test, whereas categorical variables were analyzed using the Pearson $\chi^2$ test or Fisher exact test. The Hosmer-Lemeshow goodness-of-fit test was used to create a calibration curve to evaluate the nomogram. A 2-sided $P$ value less than 0.05 was considered significant.

RESULTS

General Information

A total of 172 patients were enrolled, including 33 males and 139 females aged 25.0 to 64.0 (45.3 ± 11.2) years. There were 83 patients with benign nodules, including 50 with thyroid adenoma (24 follicular adenoma, 16 eosinophilic adenoma, 9 atypical adenoma, and 1 with a hyaline variable beam tumor), 29 patients with nodular goiter, 2 patients with granulomatous thyroiditis, and 2 patients with hemangioma. The remaining 89 patients had malignant nodules, including 55 with papillary thyroid carcinoma, 21 with follicular carcinoma, 7 with medullary carcinoma, 1 with undifferentiated carcinoma, 4 with metastasis, and 1 with lymphoma. The size of the included lesions ranged from 1 to 2.4 cm.

Using stratified random sampling to divide the cases in a 7:3 ratio, 121 patients were placed in the training group, and 51 were placed in the test group. There were no significant differences in the sex, age, CT characteristics, or distribution of benign and malignant nodules between the training and test groups ($P > 0.05$; Table 1).

Feature Selection and Radiomics Signature Construction

The 3-phase images of the patients were extracted. First, minimum redundancy maximum relevance was used to remove redundant and irrelevant features. The retained features were imported into a least absolute shrinkage and selection operator (LASSO) regression model. A 10-fold cross-validation method was used to identify the hyperparameter $\lambda$ of the LASSO regression model. The $\lambda$ corresponding to the smallest model error was selected, and features with nonzero coefficients were retained. Regression and dimensionality reduction were used to screen out features with good generalizability, and a prediction model was established. The imaging Radscore of each patient in the training set was calculated as the sum of the product of the features and the corresponding coefficients, and the Wilcoxon rank-sum test was used to compare the imaging Radscore between benign and malignant nodules.

The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the efficacy of the 3-phase scan in identifying benign and malignant thyroid nodules. The optimal phase combined with clinical features was selected to establish the corresponding radiomics nomogram. In the training group, 1-way analysis of variance was used to screen independent clinical features ($P < 0.05$). Multivariate logistic regression was then used in conjunction with radiomics tags to screen the final predictors and construct a radiomics nomogram.

FIGURE 1. Radiomics features selected for the training group. A, The optimal tuning parameter ($\lambda$) was selected using 10-fold cross-validation (CV) with the LASSO regression model. B, $\lambda$ was used to obtain 21 nonzero coefficient radiomics features, where the different colored lines represent the change tracks for the feature coefficients. Figure 1 can be viewed online in color at www.jcat.org.

FIGURE 2. The final selected features extracted from plain, arterial, and venous contrast-enhanced CT images used to distinguish benign nodules from malignant nodule patients are shown in (A, B, and C), respectively. Figure 2 can be viewed online in color at www.jcat.org.
redundant and irrelevant features, and 30 features were retained. Then, the optimized feature subset was selected using the LASSO algorithm to build the model (Fig. 1). The model error on the vertical axis was minimized by adjusting the value of $\lambda$ on the horizontal axis, and the log($\lambda$) value corresponding to the best value of $\lambda$ on the vertical dotted line was selected as the minimum standard. The optimal $\lambda$, 0.047 (log($\lambda$) = −0.049) was obtained by 10-fold cross-validation and used to obtain 21 nonzero coefficient radiomics features. Among the 21 features that were screened out, 8 were in the noncontrast scan phase, 7 were in the arterial phase, and 6 were in the venous phase. The 21 features consisted of 2 first-order features and 19 texture features, including 5 gray-level run-length matrix (GLRLM) features, 4 gray-level size zone matrix (GLSZM) features, 7 gray-level dependence matrix (GLDM) features, and 3 neighboring gray tone difference matrix (NGTDM) features (Fig. 2). The Radscore was calculated by summing the selected features weighted by their coefficients:

\[
N: Radscore = 1.127 \times \text{wavelet HHH glrlm RunPercentage} + 0.099 \times \text{wavelet HHL firstorder Energy} + 0.202 \times \text{wavelet LLIH glszm LargeAreaHighGrayLevelEmphasis} - 0.128 \times \text{wavelet LHH glszm LargeAreaLowGrayLevelEmphasis} + 0.277 \times \text{wavelet LLIH glszm ZoneVariance} - 0.038 \times \text{wavelet HHL gldm DependenceEntropy} + 0.257 \times \text{wavelet HHH gldm SmallDependenceEmphasis} + 0.073 \times \text{wavelet HHL glrlm RunLengthNonUniformityNormalized} - 0.191
\]

\[
A: Radscore = 0.439 \times \text{wavelet HHH gldm SmallDependenceLowGrayLevelEmphasis} - 0.201 \times \text{original glrlm RunEntropy} + 0.277 \times \text{wavelet HHH gldm SmallDependenceEmphasis} + 0.299 \times \text{wavelet HHH gldm DependenceEntropy} + 0.475 \times \text{wavelet LLL gldm DependenceEntropy} + 0.082 \times \text{wavelet LLIH firstorder Energy} + 0.073 \times \text{wavelet HHL glrlm RunLengthNonUniformityNormalized} + 0.191
\]

\[
V: Radscore = -0.038 \times \text{wavelet HHL glszm ZoneVariance} - 0.955 \times \text{wavelet LLIH gldm DependenceEntropy} + 0.617 \times \text{wavelet HHL gldm SmallDependenceEmphasis}
\]

This formula was used to calculate the imaging Radscores of the training and test groups. The scores for the malignant nodules were significantly higher than those of the benign group, and the difference was statistically significant ($P < 0.001$, Wilcoxon rank-sum test), as shown in Figure 3.

**TABLE 2. Comparison of Prediction Model Results Between the Training and Test Groups**

| Phase       | Training Group | Test Group |
|-------------|----------------|------------|
|             | ACC, % | SEN, % | SPE, % | PPV, % | NPV, % | ACC, % | SEN, % | SPE, % | PPV, % | NPV, % |
| Plain phase  | 0.78   | 0.59   | 0.99   | 0.97   | 0.69   | 0.63   | 0.35   | 0.92   | 0.82   | 0.58   |
| Arterial phase | 0.84   | 0.81   | 0.86   | 0.86   | 0.81   | 0.77   | 0.73   | 0.80   | 0.79   | 0.74   |
| Venous phase | 0.83   | 0.78   | 0.88   | 0.88   | 0.79   | 0.75   | 0.69   | 0.80   | 0.78   | 0.71   |

ACC indicates accuracy; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity.

---

**Diagnostic Efficiency of the Radiomics Model**

The AUCs of the radiomics feature models (Radscores) in the training group that were used to identify benign and malignant thyroid nodules in the noncontrast scan and the arterial and venous phases were 0.86 (95% CI, 0.79–0.92), 0.89 (95% CI, 0.83–0.95), and 0.88 (95% CI, 0.82–0.94), respectively. The corresponding AUCs of the 3-phase model in the test group were 0.76 (95% CI, 0.63–0.90), 0.78 (95% CI, 0.65–0.91), and 0.76 (95% CI, 0.62–0.90). The results are shown in Table 2.

**Clinical Feature Screening and Nomogram Construction**

The clinical feature model consisted of age, sex, and CT characteristics (presence or absence of microcalcifications, regular/irregular lesion morphology, positive/negative negative edge interruption sign, clear/fuzzy lesion boundary after enhancement, positive/negative cystic change ratio >50%). The analysis of variance results for the training group showed that the presence or absence of microcalcifications, regular/irregular lesion morphology, and positive/negative cystic change ratio greater than 50% were clinically independent risk factors ($P < 0.05$). The multivariate logistic regression results showed that regular/irregular lesion morphology and positive/negative cystic change ratio greater than 50% were independent factors for differentiating benign and malignant nodules ($P < 0.05$). The individualized prediction model was composed of the independent clinical factors and the Radscore and is represented by a nomogram (Fig. 4).
Nomogram Evaluation and Verification

The ROC curves of the 3 models—the radiomics nomogram, radiomics score, and clinical features—of the training and test groups are shown in Figure 5. The AUCs for the training group were as follows: radiomics nomogram, 0.93 (95% CI, 0.88–0.98); radiomics score, 0.89 (95% CI, 0.83–0.95); and clinical characteristics, 0.74 (95% CI, 0.67–0.82). The AUCs for the test group were as follows: radiomics nomogram, 0.84 (95% CI, 0.73–0.95); radiomics score, 0.78 (95% CI, 0.65–0.91); and clinical characteristics, 0.76 (95% CI, 0.64–0.87). These results show that the radiomics nomogram had a higher discrimination efficiency than the radiomics model and clinical characteristics alone and is therefore better (Table 3).

The decision curve analysis of the nomogram for individual predictions is illustrated in Figure 6. The decision curve shows that if the threshold probability of a patient or doctor was greater than 10%, the radiomics nomogram for predicting benign and malignant thyroid nodules would be more beneficial than the clinical features model (without the Radscore).

The nomogram calibration curve for identifying the probability of benign and malignant nodules in patients showed that the nomogram prediction was in excellent agreement with actual observations for the training (A) and test (B) groups (Fig. 7).

DISCUSSION

Current Research Status of CT Imaging of the Thyroid Gland

Computed tomography examination exposes a patient to a controlled radiation dose, and the thyroid is relatively sensitive to radiation. However, CT examination can compensate for the inadequacy of ultrasound examination in revealing calcification, and CT-enhanced scans can show the characteristics of the internal blood supply of a lesion. With the more frequent use of low-dose CT over the past few years, the clinical application of CT examination for thyroid diseases within the safe dose range has increased. With traditional imaging, visualization is mainly used to determine lesion morphology. This method no longer meets the needs of contemporary precision medicine, however. The heterogeneity of malignant tumors is closely related to the aggressiveness and prognosis of these tumors. Therefore, accurately distinguishing between benign and malignant and heterogeneous tumors before surgery has important clinical value for the treatment and prognosis of the patient. Dutch scholars proposed the concept of radiomics in 2012, whereby a large quantity of imaging information is extracted using traditional imaging methods, such as CT and MRI, followed by segmentation, feature extraction, and model establishment to identify tumors. The results are then used in conjunction with mining deeper information hidden in the image to guide personalized clinical decision making and treatment planning. Computed tomography radiomics has been studied for differentiating benign and malignant rectal, prostate, and lung nodules and breast masses but has rarely been applied to the thyroid.

Characteristics of the Radiomics Method Used in This Study

The ROIs in this study were all manually drawn. Features were retained only when the intraobserver and interobserver ICCs
Radiomics Features

In this study, thyroid lesion images were obtained for a noncontrast, arterial phase, and venous phase scans. Feature extraction yielded a total of 21 features (8 in the noncontrast scan, 7 in the arterial phase scan, and 6 in the venous phase scan), of which 2 were first-order features and 19 were texture features (5 GLRLM, 4 GLSZM, 7 GLDM, and 3 NGTDM). First-order features quantitatively describe the voxel intensity distribution in CT images based on common basic metrics. Textural features describe the spatial distribution characteristics of the voxel intensity based on the joint probability distribution of specific pixels. Abnormal angiogenesis, cell permeability changes, and necrosis in malignant tumors often produce heterogeneity, mixed components, roughness within nodules, and a complex grey scale distribution in tumor images. The previously mentioned variations are not easily detectable by the naked eye but can be detected as textural features that are not affected by subjective factors. Gray-level run-length matrix contains high-order statistics of image grey scale histograms that indicate the randomness and uncertainty of the run length; GLSZM indicates the uncertainty and randomness of the regional grey scale distribution, which is related to the texture heterogeneity of a 3D region; GLDM quantifies an image based on a grey scale correlation; and NGTDM produces a neighborhood grey scale difference matrix that reflects heterogeneity in the spatial distribution of lesions. Considering the extracted features (GLRLM, GLSZM, GLDM, first-order, and NGTDM) and the diagnostic features for the 3-phase images in this study, there are differences in the voxel intensity, texture gray length, and distribution of heterogeneity between benign and malignant thyroid nodules.

The results of this study show that the highest discrimination efficiency was obtained using the CT arterial phase scan, for which the AUC and accuracy for the test group were 0.78 and 0.76, respectively. The corresponding AUC and accuracy of the venous phase scan were approximately the same but slightly lower at 0.76 and 0.74, respectively. The corresponding AUC and accuracy of the noncontrast scan were relatively low at 0.76 and 0.63, respectively. This result was obtained because a noncontrast scan mainly reflects the heterogeneity in the tissue composition and cell density in a tumor, whereas an enhanced CT scan can also indicate changes in the tumor blood supply. Thus, an enhanced CT scan can detect and describe the heterogeneity within a tumor more accurately than a noncontrast scan. Wu et al used a CT imaging nomogram to predict benign and malignant small thyroid nodules. However, the 19 imaging omics features that were extracted were all from the noncontrast scan period, which is inconsistent with the results of this study. The reason for this inconsistency may be the selection of nodules measuring below 1 cm in the study by Wu et al because enhanced CT offers no discernible advantage for small lesions.

TABLE 3. Comparison of the Results of the Radiomics Nomogram, Radiomics Score, and Clinical Feature Prediction Model

|                  | Training Group | Test Group |
|------------------|----------------|------------|
|                  | ACC, % | SEN, % | SPE, % | PPV, % | NPV, % | ACC, % | SEN, % | SPE, % | PPV, % | NPV, % |
| Radiomics nomogram | 0.88   | 0.87   | 0.88   | 0.80   | 0.86   | 0.80   | 0.81   | 0.80   | 0.79   | 0.74   |
| Radiomics score  | 0.84   | 0.81   | 0.86   | 0.86   | 0.81   | 0.77   | 0.73   | 0.80   | 0.79   | 0.74   |
| Clinical features | 0.68   | 0.98   | 0.62   | 0.95   | 0.79   | 0.65   | 0.96   | 0.32   | 0.60   | 0.89   |

ACC indicates accuracy; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPE, specificity.

Construction and Effectiveness of the Nomogram

In this study, we selected the optimal phase, the arterial phase, to further establish and verify a nomogram based on radiomics and clinical characteristics to improve prediction performance. The radiomics nomogram constructed by Wu et al, which included radiomics characteristics and clinical risk factors, was more effective in predicting benign and malignant small thyroid nodules (AUC, 0.69) than the use of radiomics characteristics (AUC, 0.79) and pure clinical risk factors alone (AUC, 0.70). A nomogram constructed by Tang et al and other radiomics models based on CT images have demonstrated high efficacy in...
distinguishing between follicular thyroid carcinoma and follicular thyroid adenoma (AUC, 0.91). The calibration curves of the pathological results were in good agreement with those for the predicted results. Using the radiomics nomogram constructed in this study by combining radiomics and clinical features differentiated between benign and malignant thyroid nodules (AUC, 0.84; 95% CI, 0.73–0.95) more effectively than using simple radiomics features (AUC, 0.78; 95% CI, 0.65–0.91) and pure clinical features (AUC, 0.76; 95% CI, 0.64–0.87).

Ultrasound-guided fine-needle aspiration biopsy (FNAB) is currently widely used in clinical practice. National guidelines have established FNAB as a category A recommendation for the diagnosis of thyroid nodules. However, some nodules still cannot be diagnosed using a single FNAB. Even after repeated FNABs, approximately 9.9% to 47.8% of nodules cannot be diagnosed or unambiguously identified. Tumor tissues demonstrate both temporal and spatial heterogeneity; the small part of a tumor tissue that is extracted in a biopsy cannot comprehensively represent the tumor tissue characteristics. Radiomics is a noninvasive imaging technique for extracting images from an ROI. The quantitative features obtained from these images may not be discernible by the human eye but can more objectively reflect the heterogeneity in tumors. The radiomics nomogram used in this study is a noninvasive predictive tool for analyzing overall tumor characteristics, thereby comprehensively reflecting the heterogeneity in the tumor and providing a novel method for the differential diagnosis of benign and malignant thyroid nodules.

Limitations and Prospects

Radiomics has good application prospects but is still in its infancy. Thus, the technology is not yet fully developed, and the results have poor reproducibility. Furthermore, the features extracted from data collected by multiple and single centers are different. Parekh and Jacob\(^{20}\) found differences in the image results obtained using different scanners, as well as between manual and automatic separation. This study was performed on a relatively small sample obtained at a single center. In future studies, larger samples from multiple centers need to be collected to verify the repeatability and predictive power of the proposed method.

To ensure the consistency of the results, all thyroid nodule patients in our department underwent the same CT scan with the same examination parameters. Noncontrast CT can detect intralesional calcification and is a necessary part of the routine examination. Yao\(^{21}\) found that a radiomics model based on noncontrast CT had good value in differentiating benign and malignant thyroid nodules, and Chen et al\(^{22}\) did likewise using texture feature analysis. These studies only investigated the results of single-phase scans and did not perform multiple phase comparisons. Therefore, a 3-phase scanning method was designed in this study to find the optimal phase; however, this increased the dose of radiation delivered to the patient and is not a common scanning method for the thyroid gland/head and neck in daily practice.

In this study, a single thyroid nodule was selected to avoid mutual interference from multiple nodules and accurately analyze its pathological features and internal features. It is not easy to find thyroid nodules smaller than 1 cm on CT, and thus, diagnoses are easily missed. However, for nodules larger than 1 cm, most can be found by noncontrast-enhanced scanning. Therefore, our inclusion criteria were a single thyroid nodule larger than 1 cm. However, a high percentage of patients with thyroid nodules actually had multiple lesions, and thus, in future studies, we will study the effect of multiple lesions in greater depth.

In conclusion, this study provides a valuable model for identifying benign and malignant thyroid nodules based on a combined radiomics model and clinical risk factors. A multicenter prospective study will be carried out in the future to further improve the clinical application value and differential diagnosis level of the proposed method.

REFERENCES

1. Bomeli SR, LeBeau SO, Ferris RL. Evaluation of a thyroid nodule. Otolaryngol Clin North Am. 2010;43:229–238.
2. LiVolsi VA. Papillary thyroid carcinoma: an update. Mod Pathol. 2011;24(suppl 2):S1–S9.
3. Xie BK, Guan YB, Yuan XP, et al. Relationship between computed tomography (CT) manifestations and pathology in thyroid carcinoma [in Chinese]. Ai Zheng. 2003;22:192–197.
4. Zhou S, Liu T, Zhou J, et al. Preliminary study on application of radiomics in thyroid carcinoma. Onco Targets Ther. 2017;26:102–105.
5. Lu F, Ma Z. The application value of multi-slice spiral CT and ultrasound in the differential diagnosis of thyroid nodules. J Med Targets. 2014;24:628–631.
6. Wu CW, Dionigi G, Lee KW, et al. Calcifications in thyroid nodules identified on preoperative computed tomography: patterns and clinical significance. Surgery. 2012;151:464–470.

7. Zhu MP, Zhou XY, Han ZJ. The value of CT features and the combination in diagnosing papillary thyroid carcinoma. Chin Clin Med Imaging. 2014; 5:840–843,855.

8. Han ZJ, Shu YY, Lai XF, et al. Value of computed tomography in determining the nature of papillary thyroid microcarcinomas: evaluation of the computed tomographic characteristics. Clin Imaging. 2013;37:664–668.

9. 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;26:1–133.

10. Daghir-Wojtkowiak E, Wiczling P, Bocian S, et al. Least absolute shrinkage and selection operator and dimensionality reduction techniques in quantitative structure retention relationship modeling of retention in hydrophilic interaction liquid chromatography. J Chromatogr A. 2015;1403:54–62.

11. Head and Neck Science Group, Radiology Branch of Chinese Medical Association. Expert consensus on imaging examination procedures for thyroid nodules. Chin J Radiol. 2016;50:911–915.

12. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366:883–892.

13. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–577.

15. Sun K, Huang SH, Wong DS, et al. Design and application of a variable selection method for multilayer perceptron neural network with LASSO. IEEE Trans Neural Netw Learn Syst. 2017;28:1386–1396.

16. Zhang L, Fang M, Zang Y, et al. Development and application of radiomics. Chin J Radiol. 2017;51:75–77.

17. Wu XX, Li JJ, Mao N, et al. A radiomics nomogram based on computed tomography for predicting benign and malignant thyroid nodules. J Otolaryngol Ophthalmol of Shandong University. 2020;34:32–39.

18. Tang P, Ren C, Shen L, et al. Development and validation of a diagnostic nomogram for the preoperative differentiation between follicular thyroid carcinoma and follicular thyroid adenomas. J Comput Assist Tomogr. 2021;45:128–134.

19. Li CL, Zhou L, Sun H. Application status and progress of fine-needle aspiration biopsy of thyroid nodules in our country. Chin J Pract Surg. 2020;40:195–198.

20. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. Expert Rev Precis Med Drug Dev. 2016;1:207–226.

21. Yao XX. Construction and validation of two-level CT-based radiomics models used for thyroid cancers screening-CHKD doctoral dissertation. MM: WanNan Medical College; 2021.

22. Chen JJ, Fang ZH, Zhong WJ, et al. CT texture analysis in differentiating benign and malignant thyroid nodules and predicting lymph node metastasis[J]. Chin J Med Imaging. 2021;37:5–39.