

research article

# Comparison of 2D and 3D radiomics features with conventional features based on contrast-enhanced CT images for preoperative prediction the risk of thymic epithelial tumors

Yu-Hang Yuan<sup>1</sup>, Hui Zhang<sup>1</sup>, Wei-Ling Xu<sup>1</sup>, Dong Dong<sup>1</sup>, Pei-Hong Gao<sup>1</sup>, Cai-Juan Zhang<sup>1</sup>, Yan Guo<sup>2</sup>, Ling-Ling Tong<sup>3</sup>, Fang-Chao Gong<sup>4</sup>

<sup>1</sup> Department of Radiology, The First Hospital of Jilin University, Jilin, China

<sup>2</sup> GE Healthcare, China

<sup>3</sup> Department of Pathology, The First Hospital of Jilin University, Jilin, China

<sup>4</sup> Department of Thoracic Surgery, The First Hospital of Jilin University, Jilin, China

Radiol Oncol 2025; 59(1): 69-78.

Received 31 July 2024

Accepted 27 January 2025

Correspondence to: Fang-Chao Gong, Department of Thoracic Surgery, The First Hospital of Jilin University, No. 71 Xinmin Street, Chaoyang District, Changchun 130021, Jilin, China. E-mail: fangchao\_gong@163.com

Yu-Hang Yuan and Hui Zhang contributed equally to this work.

Disclosure: No potential conflicts of interest were disclosed.

This is an open access article distributed under the terms of the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

**Background.** This study aimed to develop and validate 2-Dimensional (2D) and 3-Dimensional (3D) radiomics signatures based on contrast-enhanced computed tomography (CECT) images for preoperative prediction of the thymic epithelial tumors (TETs) risk and compare the predictive performance with conventional CT features.

**Patients and methods.** 149 TET patients were retrospectively enrolled from January 2016 to December 2018, and divided into high-risk group (B2/B3/TCs, n = 103) and low-risk group (A/AB/B1, n = 46). All patients were randomly assigned into the training (n = 104) and testing (n = 45) set. 14 conventional CT features were collected, and 396 radiomic features were extracted from 2D and 3D CECT images, respectively. Three models including conventional, 2D radiomics and 3D radiomics model were established using multivariate logistic regression analysis. The discriminative performances of the models were demonstrated by receiver operating characteristic (ROC) curves.

**Results.** In the conventional model, area under the curves (AUCs) in the training and validation sets were 0.863 and 0.853, sensitivity was 78% and 55%, and specificity was 88% and 100%, respectively. The 2D model yielded AUCs of 0.854 and 0.834, sensitivity of 86% and 77%, and specificity of 72% and 86% in the training and validation sets. The 3D model revealed AUC of 0.902 and 0.906, sensitivity of 75% and 68%, and specificity of 94% and 100% in the training and validation sets.

**Conclusions.** Radiomics signatures based on 3D images could distinguish high-risk from low-risk TETs and provide complementary diagnostic information.

Key words: thymic epithelial tumors; radiomics; computed tomography; World Health Organization; classification

## Introduction

Thymic epithelial tumors (TETs) are the most prevalent neoplasms in the anterior mediastinum. They generally exhibit low malignancy in both histologi-

cal and clinical presentations.<sup>1</sup> The WHO classification of TETs, based on lymphocyte-to-epithelial cell ratio and epithelial cell morphology, is widely adopted. According to the 2015 criteria, TETs are categorized into thymomas (with six subtypes)

and thymic carcinomas (TCs).<sup>2,3</sup> This classification serves as an independent prognostic factor and is simplified into low-risk (Type A, AB, B1) and high-risk (Type B2, B3, TCs) categories, influencing patient outcomes.<sup>4</sup> The WHO classification reflects the tumor's clinical and functional features, aiding preoperative diagnosis and treatment planning.<sup>5</sup> Accurate, noninvasive identification and subgroup classification of TETs prior to treatment hold significant clinical value.

CT imaging is the primary diagnostic tool for TETs, revealing a wide range of biological and morphological features.<sup>6,7</sup> Studies have reported specific CT characteristics of TETs.<sup>8,9</sup> While chest contrast-enhanced computed tomography (CECT) provides general morphological parameters, distinguishing between histological subgroups remains challenging due to significant overlap. Radiomics, leveraging radiomic signatures, extracts diverse, high-throughput imaging features, transforming medical images into mineable data.<sup>10,11</sup> Radiomic features can predict disease, cancer, metastasis, and prognosis.<sup>12</sup> Previous studies have used 2D radiomics to determine thymoma risk levels.<sup>13,14</sup> Wang *et al.* differentiated low-risk from high-risk thymomas and early from advanced tumors using contrast-enhanced CECT and non-enhanced CT<sup>15</sup>, but their study had a small sample size and the sample size was imbalanced.

Despite several CT-based radiomics analyses being used to classify TET risk, most studies rely on 2D imaging features. Further research is needed to explore 3D radiomic signatures, which may provide more comprehensive and accurate information for TET diagnosis and risk assessment. The available models still require refinement and validation.

This study aimed to develop conventional and 2D/3D radiomics signatures using the conventional and texture features extracted from CECT images for preoperative prediction of the WHO's TET risk classification.

## Patients and methods

### Patients

This retrospective study was approved by the Ethics Committee of the First Hospital of Jilin University, and the requirement for informed consent was waived due to the retrospective nature of the study (Study Approval Number: 2020-541). From January 2016 to December 2018, data from 175 patients who undergoing CECT examination within one week to two months before surgery who were pathologi-

cally diagnosed as TET were collected in our hospital consecutively. The inclusion criteria were: 1) underwent CECT within one week to two months before surgery without chemotherapy; 2) high CT image quality without artifacts; and 3) available clinical and surgical data. The exclusion criteria were shown in Figure 1. Finally, 149 patients with pathologically confirmed TET were enrolled in this study. The baseline characteristics of all patients including age, sex and symptoms (thoracalgia and myasthenia gravis) were collected. The patients were divided into the high-risk group (B2/B3/TCs, n = 103) and the low-risk group (A/AB/B1, n = 46) according to the WHO classification criteria of TET. All patients were randomly assigned into the training group (n = 104) and test (n = 45) group at a ratio of 7:3. Figure 1 illustrates the flow chart of the case selection process.

### CT scanning

All patients underwent routine two-phase chest CECT scans using either a 64-MDCT scanner (Definition, Siemens Healthcare, Erlangen, Germany) or a 128-MDCT scanner (iCT, Philips Healthcare, Amsterdam, Netherlands). The scanning parameters were as follows: (1) Philips iCT: tube voltage 120 kV, automatic mAs, layer thickness 5 mm, pitch 0.980; (2) GE Lightspeed VCT: tube voltage 120 kV, automatic mAs, slice thickness 5 mm, pitch 0.992. All patients were examined in a supine position, arms up, deep inspiration and scanning. The contrast agent was injected into the patient using a high-pressure syringe (Visipaque 320, Amersham Health, Cork, Ireland). A total of 60-80 mL of contrast agent was administered through the antecubital vein at a rate of 3 ml/s, which was followed by 30 ml of saline injection at the same rate.

### Conventional CT features measurement

The CT images were reviewed on the picture archiving and communication system (PACS). The conventional CT features of all patients were analyzed and recorded, including tumor's long diameter, short diameter, vertical diameter, area, perimeter and CT values. The tumor heterogeneity evaluated by the radiologist is also recorded, consisting of location (right, middle or left), morphology (lobular, /shallowly-lobulated or non-lobular), demarcation (clear, unclear or infiltrating), internal calcification and necrosis. The workflow of conventional CT feature analysis was shown in Figure 2.

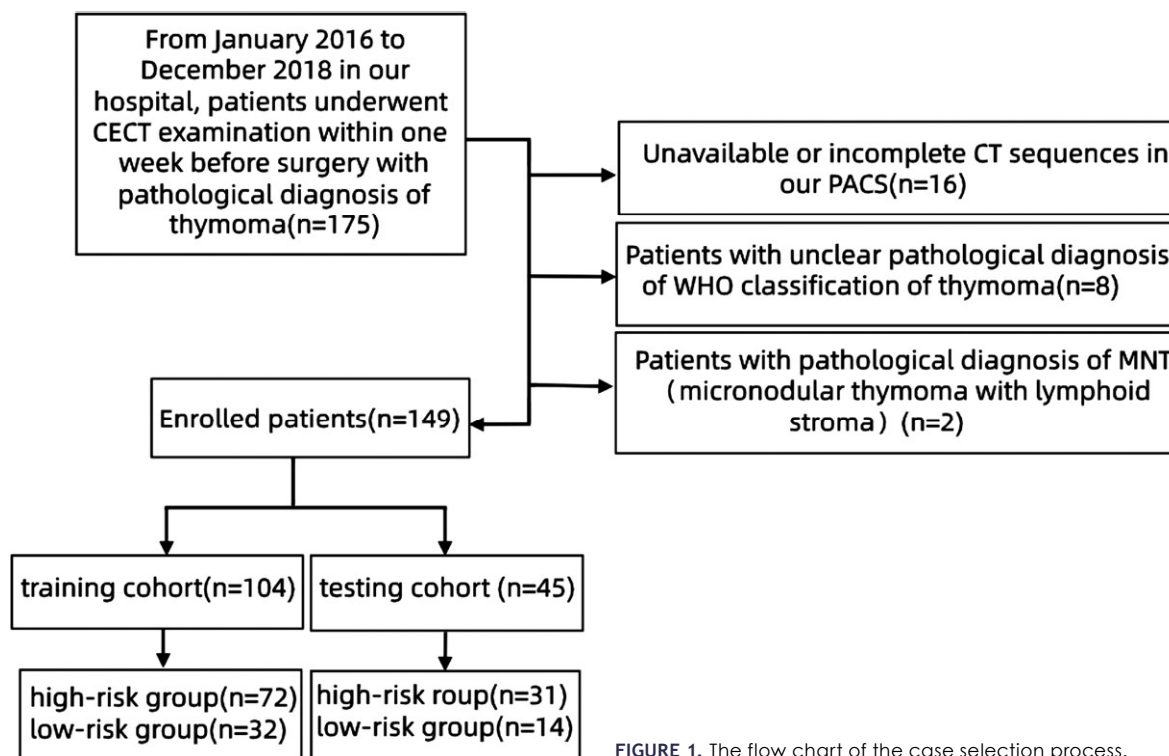


FIGURE 1. The flow chart of the case selection process.

## 2D and 3D Tumor segmentation

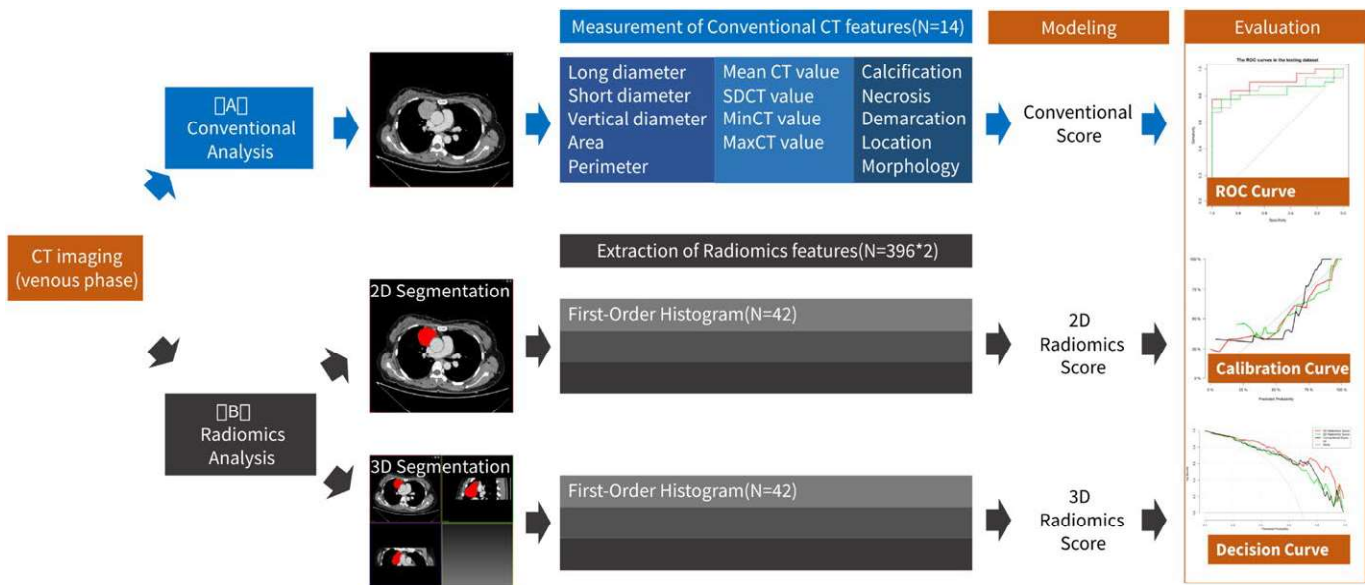
The original CT images with a thickness of 5 mm were uploaded to the A.K. software (Artificial Intelligence Kit, A.K., GE Healthcare, China) for 2D and 3D manual segmentation. First, using a linear interpolation algorithm, the raw data were resampled to a common voxel spacing of 1 mm<sup>3</sup> to construct new data points within the range of a discrete set of known data points, and the voxel would be isotropic. The data with a window width of 350 HU and a window level of 50 HU were used. The 2D and 3D regions of interest (ROIs) were delineated by two independent experienced radiologists (reader 1, XWL with 15 years of experience in chest imaging and reader 2, ZH with 10 years of experience in chest imaging) who were blinded to the pathology results. The 2D ROIs were delineated at the level of the single largest cross-sectional area around the tumor outline. The 3D ROIs were achieved from different continuous levels. To segment the ROI in axial CT images, a manual method was used on the AK software. The lesion was manually separated from the large blood vessels, lung, air, fat tissue and chest wall. The workflow of radiomics analysis was shown in Figure 2.

## 2D and 3D radiomic features extraction

According to the CECT images of each patient, a total of 396 radiomic features were extracted automatically from the 3D and 2D ROI respectively using A.K. software. The radiomic features were composed of the following classes: first-order histogram features (N = 42), second-order texture features (N = 345, including the Haralick texture (N = 10), gray level co-occurrence matrix (N = 144), gray level run length matrix (N = 180) and gray level size zone matrix (N = 11), and morphological features (N = 9). Details of the radiomics features were described in Supplementary Figure. S1.

## Features selection

To reduce the dimensionality of the conventional and 2D/3D radiomic features, the least absolute shrinkage and selection operator (LASSO) method was applied to identify the most valuable features from the training dataset. For better performance of the model, the best penalty parameters  $\lambda$  was obtained based on the loss function with the least squares during 10-fold-cross-validation procedure.



**FIGURE 2.** The flow chart of the CT imaging analysis. (A) shows the workflow of conventional analysis and 14 conventional features were recorded. (B) shows the workflow of radiomics analysis. 2D and 3D segmentation were performed on the CT images and 396 radiomics features were extracted respectively. The most predictive feature variables were selected, and the multivariate logistic regression analysis was applied to build the prediction models. The predicting abilities of the conventional and radiomics models were demonstrated by receiver operating characteristic (ROC) curves. The goodness of fit was assessed using calibration curve of the Hosmer-Lemeshow test. Additionally, decision curve analysis (DCA) was conducted to determine the clinical usefulness of the models.

## Model building and evaluation

Three models, including the conventional model, 2D radiomics model and 3D radiomics model, were established using multi-variate logistic regression based on the selected features. Then the conventional score, 2D radiomics score and 3D radiomics score were calculated for each patient via a linear combination of selected features weighted by coefficients.

The discriminative performance of the three models were evaluated by ROC curves. The area under the curves (AUC), sensitivity, specificity and the optimal cutoff value were obtained from ROC analysis. Delong test was used to identify the difference of AUC between different models, and  $P < 0.05$  indicated a significant difference. The degree of calibration was assessed in the calibration curve of the Hosmer-Lemeshow test, and  $P > 0.05$  indicated a good fit. Furthermore, decision curve analysis (DCA) was conducted to determine the clinical applicability of the models by quantifying the net benefit under different threshold probabilities.

## Statistical analysis

Continuous data with a normal distribution (according to the Kolmogorov-Smirnov test) were expressed as mean  $\pm$  standard deviation (SD) and analyzed using Student's t-test. Continuous variables with a non-normal distribution were expressed as medians (interquartile range) and analyzed using the Mann-Whitney test. Categorical variables were represented as frequencies and percentages and analyzed using the chi-square test. The statistical analysis was conducted using R software (v. 3.6.1, <http://www.R-project.org>). A two-tailed p value of less than 0.05 was considered significant.

## Results

### Baseline characteristics of the patients

There were 104 patients in the training set, among which 72 (69.2%) cases were in the high-risk group. The testing set involved 45 patients, and 31 (68.9%) cases were in the high-risk group. Age, sex, and

TABLE 1. Baseline characteristics of the patients in training and testing dataset

	Training set			Testing set		
	Low-risk (n=32)	High-risk (n= 72)	P	Low-risk (n=14)	High-risk (n=31)	P
Age, (Mean ± SD) years	53.6±11.2	52.5±11.4	0.656	54.0±10.7	56.4±8.9	0.446
Sex (male, No. (%))	14 (43.8)	37 (51.4)	0.472	7 (50.0)	18 (58.1)	0.614
Myasthenia gravis, No. (%)	7 (21.9)	24 (33.3)	0.238	0 (0.0)	8 (25.8)	0.094
Thoracalgia, No. (%)	3 (9.4)	18 (25.0)	0.067	1 (7.1)	11 (35.5)	0.104

frequency of myasthenia gravis and thoracalgia were similar between the high-risk and low-risk groups in both the training and testing sets (all  $P > 0.05$ ). The characteristics of the patients are presented in Table 1.

### Three models building

#### Conventional model

In the training set, compared with the low-risk group, the high-risk group had a lower mean HU value (62.0 *vs.* 79.5 HU,  $P < 0.001$ ), a lower maximum HU value (118.0 *vs.* 148.5 HU,  $P < 0.001$ ), a smaller short diameter 23.2 *vs.* 34.7 mm,  $P = 0.009$ ), a smaller area (628.5 *vs.* 1321.5 mm<sup>2</sup>,  $P = 0.008$ ), and a smaller perimeter (112.5 *vs.* 143.0 mm,  $P = 0.021$ ). In addition, the lesions in the high-risk group were often non-lobular ( $P = 0.010$ ), and there was no clear demarcation ( $P = 0.023$ ). The patient's CT conventional parameters are listed in Table 2.

In the testing set, the maximum HU value of the high-risk group was lower than that of the low-risk group (median HU 129.0 *vs.* 162.0,  $P = 0.002$ ). The lesions in the high-risk group tended to be non-lobular ( $P < 0.001$ ), with no clear demarcation ( $P = 0.010$ ).

After LASSO regression analysis, seven conventional CT features with non-zero coefficients were remained. The calculation formula of the conventional score was as follows:

Conventional Score =  $5.71 - 0.012 * \text{Circle.mm} - 0.051 * \text{Mean CT value} + 0.023 * \text{Minimum CT value} - 0.009 * \text{Maximum CT value} - 0.903 * \text{Necrosis} + 1.859 * \text{adjacent} - 0.595 * \text{Location}$ .

#### 2D Radiomics model

After LASSO regression analysis, five CT features including Min Intensity, Percentile75, Correlation\_angle0\_offset7, LongRunEmphasis\_angle90\_offset1, and Small Area Emphasis were retained. The formula for the calculation of the 2D radiomics score was as follows:

2D Radiomics Score =  $1.343 - 0.528 * \text{Min intensity} - 0.805 * \text{Percentile75} - 0.557 * \text{Correlation\_angle0\_offset7} + 1.343 * \text{LongRunEmphasis\_angle90\_offset1} - 0.900 * \text{SmallAreaEmphasis}$

#### 3D radiomics model

The 3D radiomics model was established using eight selected CT features, and the calculation formula of the 3D radiomics score was as follows:

3D Radiomics Score =  $1.330 - 1.731 * \text{Percentile75} - 0.915 * \text{ClusterProminence\_angle135\_offset4} - 1.266 * \text{GLCMEnergy\_angle0\_offset7} + 0.635 * \text{AngularSecondMoment} - 0.559 * \text{LongRunEmphasis\_AllDirection\_offset1} + 0.678 * \text{LowGreyLevelRunEmphasis\_AllDirection\_offset7\_SD} - 0.573 * \text{Compactness1} - 0.714 * \text{IntensityVariability}$ .

### The performance of the three models

#### Discriminative degree

In the training set, this conventional model had an AUC of 0.863 (95% confidence interval (CI): 0.786-0.940), a sensitivity of 78%, and a specificity of 88%. In the testing set, the AUC was 0.853 (95% CI: 0.740-0.965), the sensitivity was 55%, and the specificity was 100%. The 2D radiomics model yielded an AUCs of 0.854 (95% CI: 0.7-0.931) and 0.834 (95% CI: 0.714-0.954), a sensitivity of 86% and 77%, and a specificity of 72% and 86% in the training and testing set, respectively. For the 3D radiomics model, the AUC of the training and testing set was 0.902 (95% CI: 0.842-0.963) and 0.906 (95% CI: 0.820-0.991), respectively, the sensitivity was 75% and 68%, respectively, and the specificity was 94% and 100%, respectively. The ROC curves and the detailed results were shown in Figure 3 and Table 3.

As shown in Figure 3, the AUC of 3D model was usually higher compared with the conventional parameter model and 2D model. However,

TABLE 2. Distribution of conventional CT features in training and testing dataset

	Training set			Testing set		
	Low-risk (n=32)	High-risk (n= 72)	P	Low-risk (n=14)	High-risk (n=31)	P
Mean CT value (HU)	79.5 (68.4, 91.6)	62.0 (51.0, 78.8)	<b>&lt;0.001</b>	88.5 (76.0, 95.1)	67.0 (59.2, 74.0)	<b>&lt;0.001</b>
Standard deviation	18.0 (16.0, 22.6)	16.5 (14.0, 19.0)	0.050	18.0 (15.9, 26.2)	17.0 (14.2, 21.6)	0.548
Minimum CT value (HU)	-6.10±30.2	-8.5±25.0	0.678	-0.7±31.6	-7.3±17.9	0.477
Maximum CT value (HU)	148.5 (131.0, 172.1)	118.0 (105.5, 138.6)	<b>&lt;0.001</b>	162.0 (148.9, 166.1)	129.0 (105.4, 146.8)	<b>0.002</b>
Long diameter (mm)	50.6±17.0	44.1±19.4	0.106	47.8 (40.9, 57.8)	38.0 (27.7, 61.3)	0.198
Short diameter (mm)	34.7 (23.9, 41.7)	23.2 (17.7, 34.6)	<b>0.009</b>	36.5 (26.0, 45.4)	25.6 (19.0, 39.3)	0.073
Vertical diameter (mm)	48.6 (44.1, 60.2)	40.4 (29.1, 55.2)	0.204	50.5 (44.4, 63.6)	38.9 (33.1, 55.1)	0.059
Area (mm <sup>2</sup> )	1321.5 (692.0, 1889.8)	628.5 (397.4, 1409.7)	<b>0.008</b>	1024.0 (747.7, 1623.3)	651.0 (346.0, 1362.8)	0.315
Perimeter (mm)	143.0 (110.8, 167.7)	112.5 (78.6, 153.8)	<b>0.021</b>	143.5 (118.6, 255.4)	100.0 (84.1, 194.8)	0.098
Location			0.373			0.790
Right mediastinum	10 (31.3%)	33 (45.8%)		7 (50.0%)	11 (35.5%)	
Middle	8 (25.0%)	15 (20.8%)		1 (7.1%)	3 (9.7%)	
Left mediastinum	14 (43.8%)	24 (33.3%)		6 (42.9%)	17 (54.8%)	
Morphology			<b>0.010</b>			<b>&lt;0.001</b>
Lobular	5 (15.6%)	10 (13.9%)		7 (50.0%)	2 (6.5%)	
Shallowly-lobulated	15 (46.9%)	14 (19.4%)		7 (50.0%)	15 (48.4%)	
Non-lobular	12 (37.5%)	48 (66.7%)		0 (0.0%)	14 (45.2%)	
Demarcation			<b>0.023</b>			<b>0.010</b>
Clear	15 (46.9%)	17 (23.6%)		10 (71.4%)	8 (25.8%)	
Unclear	16 (50.0%)	43 (59.7%)		4 (28.6%)	17 (54.8%)	
Infiltration	1 (3.1%)	12 (16.7%)		0 (0%)	6 (19.4%)	
Internal calcification	8 (25.0%)	13 (18.1%)	0.416	4 (28.6%)	9 (29.0%)	0.746
Necrosis	12 (37.5%)	20 (27.8%)	0.321	9 (64.3%)	12 (38.7%)	0.111

Continuous variables that conformed to the normal distribution were expressed as mean ±SD. Continuous variables that did not conform to the normal distribution were represented by median values (25%, 75%). Categorical variables were expressed as No. (%).

Coarse P values represented statistically significant.

due to the limited data size, Delong test suggested that there was no statistical difference of AUCs among conventional, 2D and 3D models. The process and results of the Delong test are described in Supplementary Figure S2.

### Calibration degree

The calibration curves of the three models were illustrated in Figure 4, which demonstrated good agreement between the predictive and actual probability. All the P value of the Hosmer-Lemeshow test were more than 0.05 in both training and testing datasets for the conventional model (P = 0.735 and 0.266, respectively), 2D model (P = 0.665 and

0.492, respectively) and 3D model (P = 0.562 and 0.448, respectively).

### Clinical usefulness

DCA for the three models was performed in both training and testing datasets (Figure 5). The decision curve showed that all the three models added more benefit than using treat-all scheme (assuming all TETs are high risk) or the treat-none scheme (assuming all TETs are low risk) if the threshold probability was > 10%. When the threshold probability was between 25% to 60% or between 70% to 95%, using the 3D radiomics model to predict the TETs risk added more benefit than either the con-

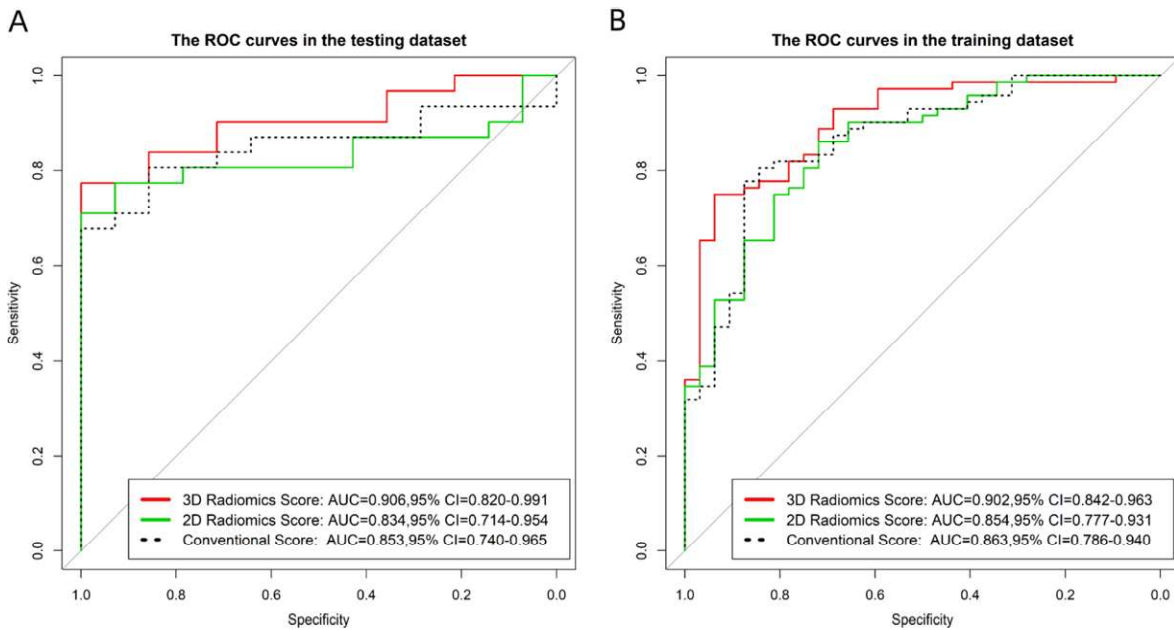


FIGURE 3. Receiver operating characteristic (ROC) curve analysis of the conventional, 2D and 3D radiomics models: (A) the training set; (B) the testing set.

ventional model or the 2D radiomics model in both training and testing datasets.

### Discussion

This study evaluates three models using conventional and radiomic signatures, revealing that radiomic features from 2D and 3D imaging serve as noninvasive biomarkers for thymic epithelial tumors (TETs). These features enable preoperative

risk stratification, guiding the surgical approach and resection extent. Importantly, risk stratification aids in personalized treatment planning for high-risk patients, often needing adjuvant therapy. A Phase II studies show that preoperative chemotherapy and radiotherapy enhance R0 resection rates.<sup>16</sup> Thus, this study highlights the significance of radiomic signatures in preoperative risk assessment and treatment planning for TETs.

In this study, three established models obtained using the conventional and radiomics signatures

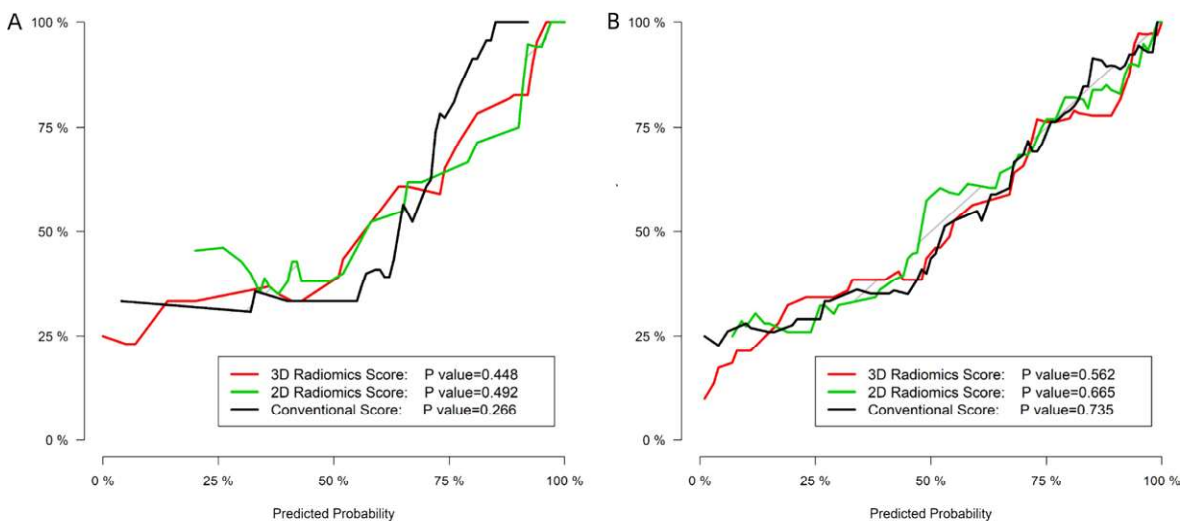


FIGURE 4. The calibration curves of the conventional, 2D and 3D radiomics models:(A) the training set; (B) the testing set.

TABLE 3. Diagnostic performance of the three models

Model	Training dataset			Testing dataset		
	Sensitivity	Specificity	AUC (95%CI)	Sensitivity	Specificity	AUC (95%CI)
Conventional models	77.8%	87.5%	0.863(0.786-0.940)	54.8%	100.0%	0.853(0.740-0.965)
2D radiomics model	86.1%	71.9%	0.854(0.777-0.931)	77.4%	85.7%	0.834(0.714-0.984)
3D radiomics model	75.0%	93.8%	0.902(0.842-0.963)	67.7%	100.0%	0.906(0.820-0.991)

AUC = area under the curve; CI = confidence interval; CT = computed tomography

were examined. The results indicated that the radiomics signatures based on conventional, 2D and 3D images could be used as noninvasive biomarkers to differentiate high-risk from low-risk TETs. Specifically, these models can provide preoperative risk stratification which informs the choice of surgical approach (minimally invasive surgery or thoracotomy) and helps define the extent of resection. By accurately predicting the risk, this model could aid in minimizing unnecessary tissue damage and reducing complications, potentially improving surgical outcomes and postoperative recovery. Since high-risk TET patients generally have poorer prognosis and often require adjuvant therapy, this stratification can guide treatment planning and improve patient outcomes by facilitating more personalized interventions.

Radiomics has shown significant potential in personalized medicine by providing non-invasive insights into tumor characteristics.<sup>10,12</sup> It has been applied in various diseases, such as predicting thyroid cancer nodules, differentiating the benign and malignant nature of pulmonary nodules, and supporting clinical decision-making for liver cancer patients. Radiomics studies using different

imaging modalities (including ultrasound, MRI, and CT) have increasingly played a crucial role in clinical decision-making, providing new technological support for prediction, diagnosis, and prognosis.<sup>17,18,19</sup>

Radiomics can be used to determine the TET classification according to WHO criteria. Previous studies have applied radiomics to classify the risk levels of thymomas and predict their invasiveness<sup>15,20,21</sup>, but the available models still need refinement and validation. Therefore, this study aimed to use radiomics based on CECT images to develop conventional and radiomics signatures, which may help the preoperative prediction of the risk of TET classified by WHO guideline.

We used a simplified classification that defines Type A, AB, and B1 thymomas as low-risk and Type B2 and B3 thymomas and TC as high-risk, which is related to patient outcome. High-risk thymic epithelial tumors reportedly have a much poorer prognosis compared with low-risk thymic epithelial tumors, the former need adjuvant therapy to improve survival. With complete resection of the tumor, the prevalence of recurrence of LTET seems low.

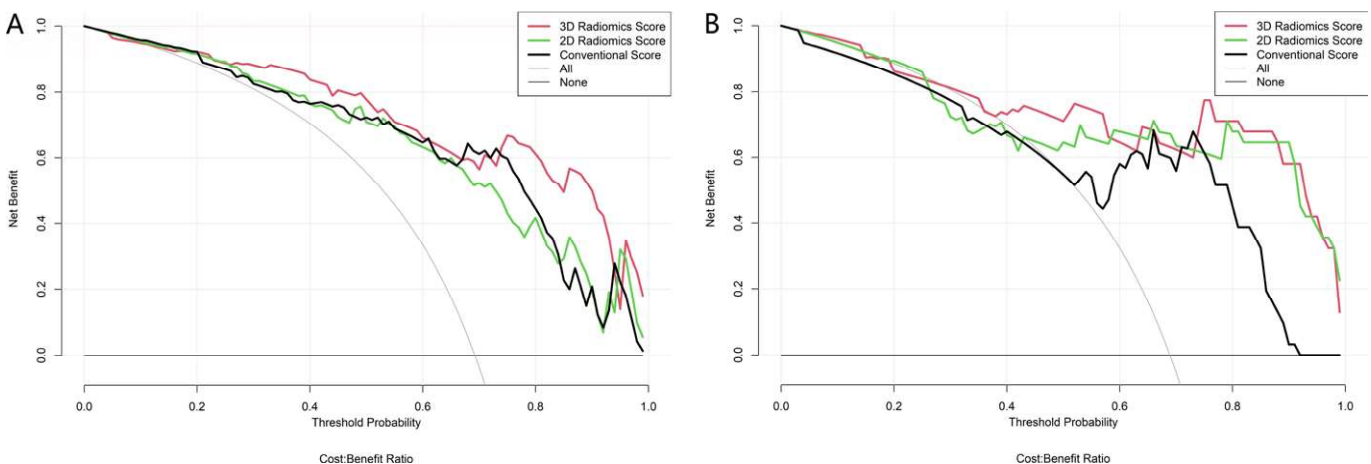


FIGURE 5. Decision curve analysis (DCA) to determine the clinical usefulness of the models by quantifying the net benefits under different threshold probabilities: (A) the training set; (B) the testing set.

Our 3D radiomics model performed better in terms of AUC, sensitivity, and specificity compared to prior models reported in the literature. For instance, Wang *et al.* reported an AUC of 0.801, with a sensitivity of 75% and specificity of 77%, while Sui *et al.* found sensitivities ranging from 71% to 74% and specificities from 65% to 74% for differentiating low-risk and high-risk TETs. The improved performance of our model may be attributed to the enhanced information captured by 3D imaging features, which provide a more comprehensive representation of tumor morphology and texture heterogeneity than 2D features. This suggests that using 3D data could offer a substantial advantage in capturing the complex structural characteristics of TETs, which is consistent with findings from previous studies on the added value of 3D features. However, it should also be noted that 3D feature extraction and analysis increase computational complexity, highlighting a potential trade-off between accuracy and practicality. In this study, the images with 5 mm thickness in CECT were used for radiomics analysis, which not only saves time but also yields satisfactory results.

Due to the small sample size, the fluctuation of individual samples affected the results greatly. The Hosmer-Lemeshow test was greater than 0.05. Although there were small fluctuations in the calibration map images of the test set, it still showed good consistency between predicted probability and true possibility without statistical deviation.

In the clinical usefulness, the 3D radiomics model was valid from 25% to 60% and from 70% to 95% for the training set, and valid from 25% to 57% and from 75% to 90% for the test set. The DCA analysis showed the 3D model to be more effective and reliable than the other two models to determine the clinical management.

Conventional features such as contour, capsule, septum and homogeneous enhancement can be used to distinguish low-risk and high-risk TETs.<sup>7-9</sup> Our study also compared the parameters obtained through the conventional CT features rather than software calculations, which also has some clinical significance for distinguishing high-risk from low-risk groups.

Except CT, magnetic resonance imaging (MRI) could also be used to differentiate the grading of TETs according to WHO criteria.<sup>22,23</sup> Nevertheless, the use and combination of multiple imaging modalities could maximize the diagnostic value, provided that all modalities are available at a given hospital, but this is not always the case, especially in developing countries like China.

This study has several limitations. First, it is a single-center, small-sample, retrospective study, which may limit the generalizability of the findings and lead to overfitting. The small sample size also means that individual sample fluctuations could have a more significant impact on the model's performance, thus affecting its robustness. To address this, we plan to expand the sample size and conduct a multi-center study for further investigation. Second, the absence of Masaoka staging information limits the model's applicability in distinguishing between different stages of thymic epithelial tumors (TETs), which is crucial for clinical decision-making. In future studies, we will incorporate this information. Third, variability in CT scan parameters and reconstruction techniques could affect the consistency and reproducibility of radiomics features, further impacting model performance. Standardizing imaging acquisition and processing methods will be essential in future studies to enhance both model robustness and clinical applicability. Additionally, AI-based automated recognition methods will be employed for more in-depth tumor analysis. Although the model demonstrates excellent performance in distinguishing risk groups, additional factors, such as tumor heterogeneity or genetic information, may further enhance predictive accuracy<sup>24</sup>.

## Conclusions

In conclusion, radiomics signatures based on conventional, 2D and 3D images could be used as non-invasive biomarkers to differentiate high-risk TETs from low-risk TETs. The 3D radiomics signature can provide complementary diagnostic information and as the most useful method to determine the clinical management.

## References

1. Engels EA. Epidemiology of thymoma and associated malignancies. *J Thorac Oncol* 2010; **5**(10 Suppl 4): S260-5. doi: 10.1097/JTO.0b013e3181f1f62d
2. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* 2015; **10**: 1243-60. doi: 10.1097/JTO.0000000000000630
3. Meurgey A, Girard N, Mervilleux du Vignaux C, Maury JM, Tronc F, Thivolet-Bejui F, et al. Assessment of the ITMIG statement on the WHO histological classification and of the eighth TNM staging of thymic epithelial tumors of a series of 188 thymic epithelial tumors. *J Thorac Oncol* 2017; **12**: 1571-81. doi: 10.1016/j.jtho.2017.06.072
4. Suster S, Moran CA. Histologic classification of thymoma: the World Health Organization and beyond. *Hematol Oncol Clin North Am* 2008; **22**: 381-92. doi: 10.1016/j.hoc.2008.03.001

5. Kondo K, Yoshizawa K, Tsuyuguchi M, Kimura S, Sumitomo M, Morita J, et al. WHO histologic classification is a prognostic indicator in thymoma. *Ann Thorac Surg* 2004; **77**: 1183-8. doi: 10.1016/j.athoracsur.2003.07.042
6. Nishino M, Ashiku SK, Kocher ON, Thurer RL, Boiselle PM, Hatabu H. The thymus: a comprehensive review-erratum. *Radiographics* 2017; **37**: 1004. doi: 10.1148/rg.2017174002
7. Sadohara J, Fujimoto K, Müller NL, Kato S, Takamori S, Ohkuma K, et al. Thymic epithelial tumors: comparison of CT and MR imaging findings of low-risk thymomas, high-risk thymomas, and thymic carcinomas. *Eur J Radiol* 2006; **60**: 70-9. doi: 10.1016/j.ejrad.2006.05.003
8. Ozawa Y, Hara M, Shimohira M, Sakurai K, Nakagawa M, Shibamoto Y. Associations between computed tomography features of thymomas and their pathological classification. *Acta Radiol* 2016; **57**: 1318-25. doi: 10.1177/0284185115590288
9. Tomiyama N, Johkoh T, Mihara N, Honda O, Kozuka T, Koyama M, et al. Using the World Health Organization Classification of thymic epithelial neoplasms to describe CT findings. *AJR Am J Roentgenol* 2002; **179**: 881-6. doi: 10.2214/ajr.179.4.1790881
10. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749-62. doi: 10.1038/nr-clinonc.2017.141
11. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol* 2016; **2**: 1636-42. doi: 10.1001/jamaoncol.2016.2631
12. Song J, Yin Y, Wang H, Chang Z, Liu Z, Cui L. A review of original articles published in the emerging field of radiomics. *Eur J Radiol* 2020; **127**: 108991. doi: 10.1016/j.ejrad.2020.108991
13. Iannarelli A, Sacconi B, Tomei F, Anile M, Longo F, Bezzi M, et al. Analysis of CT features and quantitative texture analysis in patients with thymic tumors: correlation with grading and staging. *Radiol Med* 2018; **123**: 345-50. doi: 10.1007/s11547-017-0845-4
14. Yasaka K, Akai H, Nojima M, Shinozaki-Ushiku A, Fukayama M, Nakajima J, et al. Quantitative computed tomography texture analysis for estimating histological subtypes of thymic epithelial tumors. *Eur J Radiol* 2017; **92**: 84-92. doi: 10.1016/j.ejrad.2017.04.017
15. Wang X, Sun W, Liang H, Mao X, Lu Z. Radiomics signatures of computed tomography imaging for predicting risk categorization and clinical stage of thymomas. *Biomed Res Int* 2019; **2019**: 3616852. doi: 10.1155/2019/3616852
16. Lee GD, Kim HR, Choi SH, Kim YH, Kim DK, Park SI. Prognostic stratification of thymic epithelial tumors based on both Masaoka-Koga stage and WHO classification systems. *J Thorac Dis* 2016; **8**: 901-10. doi: 10.21037/jtd.2016.03.53
17. Zhu H, Luo H, Li Y, Zhang Y, Wu Z, Yang Y. The superior value of radiomics to sonographic assessment for ultrasound-based evaluation of extrathyroidal extension in papillary thyroid carcinoma: a retrospective study. *Radiol Oncol* 2024; **58**: 386-96. doi: 10.2478/raon-2024-0040
18. Bo Z, Song J, He Q, Chen B, Chen Z, Xie X, et al. Application of artificial intelligence radiomics in the diagnosis, treatment, and prognosis of hepatocellular carcinoma. *Comput Biol Med* 2024; **173**: 108337. doi: 10.1016/j.combiomed.2024.108337
19. Warkentin MT, Al-Sawaihey H, Lam S, Liu G, Diergaarde B, Yuan JM, et al. Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches. *Thorax* 2024; **79**: 307-15. doi: 10.1136/thorax-2023-220226
20. Sui H, Liu L, Li X, Zuo P, Cui J, Mo Z. CT-based radiomics features analysis for predicting the risk of anterior mediastinal lesions. *J Thorac Dis* 2019; **11**: 1809-18. doi: 10.21037/jtd.2019.05.32
21. Marom EM, Millito MA, Moran CA, Liu P, Correa AM, Kim ES, et al. Computed tomography findings predicting invasiveness of thymoma. *J Thorac Oncol* 2011; **6**: 1274-81. doi: 10.1097/JTO.0b013e31821c4203
22. Abdel Razek AAK, Khairy M, Nada N. Diffusion-weighted MR imaging in thymic epithelial tumors: correlation with World Health Organization classification and clinical staging. *Radiology* 2014; **273**: 268-75. doi: 10.1148/radiol.14131643
23. Xiao G, Rong WC, Hu YC, Shi ZQ, Yang Y, Ren JL, et al. MRI radiomics analysis for predicting the pathologic classification and TNM Staging of thymic epithelial tumors: a pilot study. *AJR Am J Roentgenol* 2020; **214**: 328-40. doi: 10.2214/AJR.19.21696
24. Kostic Peric J, Cirkovic A, Srzentic Drazilov S, Samardzic N, Skodric Trifunovic V, Jovanovic D, et al. Molecular profiling of rare thymoma using next-generation sequencing: meta-analysis. *Radiol Oncol* 2023; **57**: 12-19. doi: 10.2478/raon-2023-0013