Am J Cancer Res 2025;15(5):2375-2396
www.ajcr.us /ISSN:2156-6976/ajcr0163457

Original Article

Machine learning-based radiomics analysis in enhancing CT for predicting pathological subtypes and WHO staging of thymic epithelial tumors: a multicenter study

Ruoxu Zhang[1*], Xueyi Zhang[2*], Zheng Dou[1], Jiaxi Lin[3], Songbing Qin[1], Chao Xu[1], Yongbing Chen[4], Jinzhou Zhu[3], Jianping Wang[1]

[1]*Department of Radiation Oncology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China;*
[2]*Department of General Surgery, Changshu Hospital Affiliated to Soochow University, Suzhou, Jiangsu, China;*
[3]*Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China;*
[4]*Department of Thoracic Surgery, The Second Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China.*
[*]*Equal contributors.*

Received January 20, 2025; Accepted May 14, 2025; Epub May 25, 2025; Published May 30, 2025

Abstract: This study is aimed to develop predictive models for classifying thymic epithelial tumor (TET) histological subtypes (A/AB/B1, B2/B3, C) and WHO stages (I-IV) using radiomics features derived from contrast-enhanced CT scans. These models were validated on multicenter external datasets to improve preoperative diagnosis and guide treatment decisions. A total of 257 patients diagnosed with TET between January 2013 and April 2024 were retrospectively analyzed, with 181 cases from the First Affiliated Hospital of Soochow University served as the training cohort and 76 cases from the Second Affiliated Hospital used as an external test set. All patients underwent preoperative enhanced CT scans. After manual segmentation of the volume of interest (VOI), 1,038 radiomic features were extracted. Feature selection was performed using PCA and LASSO methods. Three models (clinical semantic, radiomics, and a fusion model combining both) were built using random forest algorithms. The fusion model achieved the highest performance in the external test set, with an accuracy of 0.908 and F1 score of 0.896 for histological subtype classification, and an accuracy of 0.803 and F1 score of 0.833 for WHO staging. The radiomics model shows slightly lower performance, while the clinical semantic model performs the weakest. Our findings suggest that machine learning models integrating radiomics and clinical features can effectively predict TET subtypes and stages, offering a non-invasive tool for accurate preoperative assessment with strong generalization ability.

Keywords: Thymic tumors, thymoma, machine learning, radiomics, computed tomography

Introduction

Thymic epithelial tumors are the most common tumors in the anterior mediastinal compartment [1]. The occurrence of this disease is concentrated in the middle-aged group of 40-50 years old [2]. Among all malignant tumors, the estimated incidence rate is less than 1/100,000 per year, making it a rare malignant tumor [3-5]. This disease is typically associated with autoimmune diseases such as neuromuscular disorders (myasthenia gravis, encephalitis, polymyositis), immunodeficiency diseases (hypogammaglobulinemia), hematological diseases (aplastic anemia, hemolytic anemia), collagen diseases (systemic lupus erythematosus, rheumatoid arthritis, Sjögren's syndrome), and skin diseases (pemphigus, lichen planus) [6]. The most common paraneoplastic autoimmune syndrome is myasthenia gravis (23%-47%) [6, 7]. In 2015, the WHO classified thymic epithelial tumors into the following categories based on morphology, function, genetics, and clinical evidence: A, AB, B1, B2, B3, and C (thymic carcinoma) [2, 4]. In 2017, the WHO staging system was born on the basis of the traditional Masaoka staging, and it divides cases into four stages: I, II, III, and IV, based on the integrity of the capsule, invasion of surrounding tissues, involvement of lymph nodes, and the presence of distant metastases [1, 2, 8].

In the clinical practice of TETs, the formulation and improvement of a scientifically sound treat-

https://doi.org/10.62347/STUZ8659

ment plan benefit from the joint guidance of the WHO pathological classification and the WHO staging system [8]. The standard treatment plan for thymoma should be based on the tumor's resectability at the time of initial diagnosis [3, 7]. For resectable tumors, surgery is the first choice, followed by the decision on whether to administer adjuvant radiotherapy or chemotherapy based on different pathological types and stages [7, 9, 10]. The study by Liu et al. in 2017 found that for Type A, AB, and B1 thymomas, further adjuvant treatment is often unnecessary after complete resection; however, for Stage I and II B2 and B3 thymomas, adjuvant radiotherapy is still required even after complete resection [11]. However, the diagnosis of TETs pathological classification remains a very challenging task at present. Postoperative pathology or biopsy is the main method to obtain reliable pathological and histological evidence of thymoma, but the small volume of biopsy samples may lead to a final pathological result that is not fully representative [12]. Deep biopsy is an invasive procedure with the risk of complications, while superficial tissue biopsy, such as pleural biopsy, cannot rule out the possibility of tumor implantation [5, 12]. Additionally, some methods such as biopsy under CT guidance are not routine and cannot be widely carried out and applied due to their high costs [12].

Computed tomography (CT) is the primary imaging method for the initial diagnosis and evaluation of thymic epithelial tumors and is an important basis for selecting treatment options for thymic tumors [1]. The main goal of CT examination is to detect local infiltration and identify distant metastases. This is of significant importance in the determination of N and M staging [1, 2, 8]. Infiltration, as a key basis for determining T staging categories, is difficult to accurately identify on plain CT imaging. Therefore, performing venography can optimize the assessment of vascular invasion, thereby presenting the infiltration of surrounding tissues more clearly and intuitively [7]. Multiple studies report that enhanced visual evaluation of CT images is more helpful in distinguishing different histological subtypes of TETs [13]. However, due to periodic changes and histological heterogeneity, it is difficult to make a correct diagnosis regarding the staging and typing of TETs based solely on visual inspection of CT images [5]. Moreover, there is a significant overlap in the imaging manifestations of different subtypes [14, 15], and the inter-observer variability caused by differences in the experience of radiologists is also inevitable [16], which increases the risk of misdiagnosis and the possibility of errors in treatment plans.

In response to the trend towards precision in modern medicine, especially in the field of oncology, and the need for precise and efficient treatment strategies, there is an urgent need to quantify the imaging characteristics of TETs through a technology or method [17-20]. Radiomics can perform non-invasive quantitative analysis of tumor histological heterogeneity, which currently has a high application value in clinical practice for tumor treatment, especially in the discrimination of histological subtypes and staging [21-25]. In recent years, a considerable amount of research has been conducted to non-invasively characterize different histological subtypes and stages of thymic epithelial tumors (TETs) using CT-based radiomics models, with promising results. However, most research has focused on predicting high and low risk [21-23, 25, 26], early and late stages of TETs [27]. Only a few studies have classified histological subtypes into three categories [28, 29]. This type of research has the issue of a small sample size and lacks independent external test cohorts for objective evaluation of model performance.

This study aims to establish predictive models for TETs histological subtypes (A/AB/B1, B2/B3, C) and WHO stages (I, II, III, IV) based on radiomics features from contrast-enhanced CT images, and to test and evaluate the model performance using multicenter external datasets, with the goal of quantifying TETs imaging features and refining diagnostic outcomes to guide subsequent treatment.

Methods

Patients

This study has been approved by the Medical Ethics Committee of the First Affiliated Hospital of Soochow University (Approval number: No. 176; Approval date: April 2, 2024).Due to the retrospective nature of the study, patient consent was waived. We conducted a retrospective analysis of patients diagnosed with thymoma

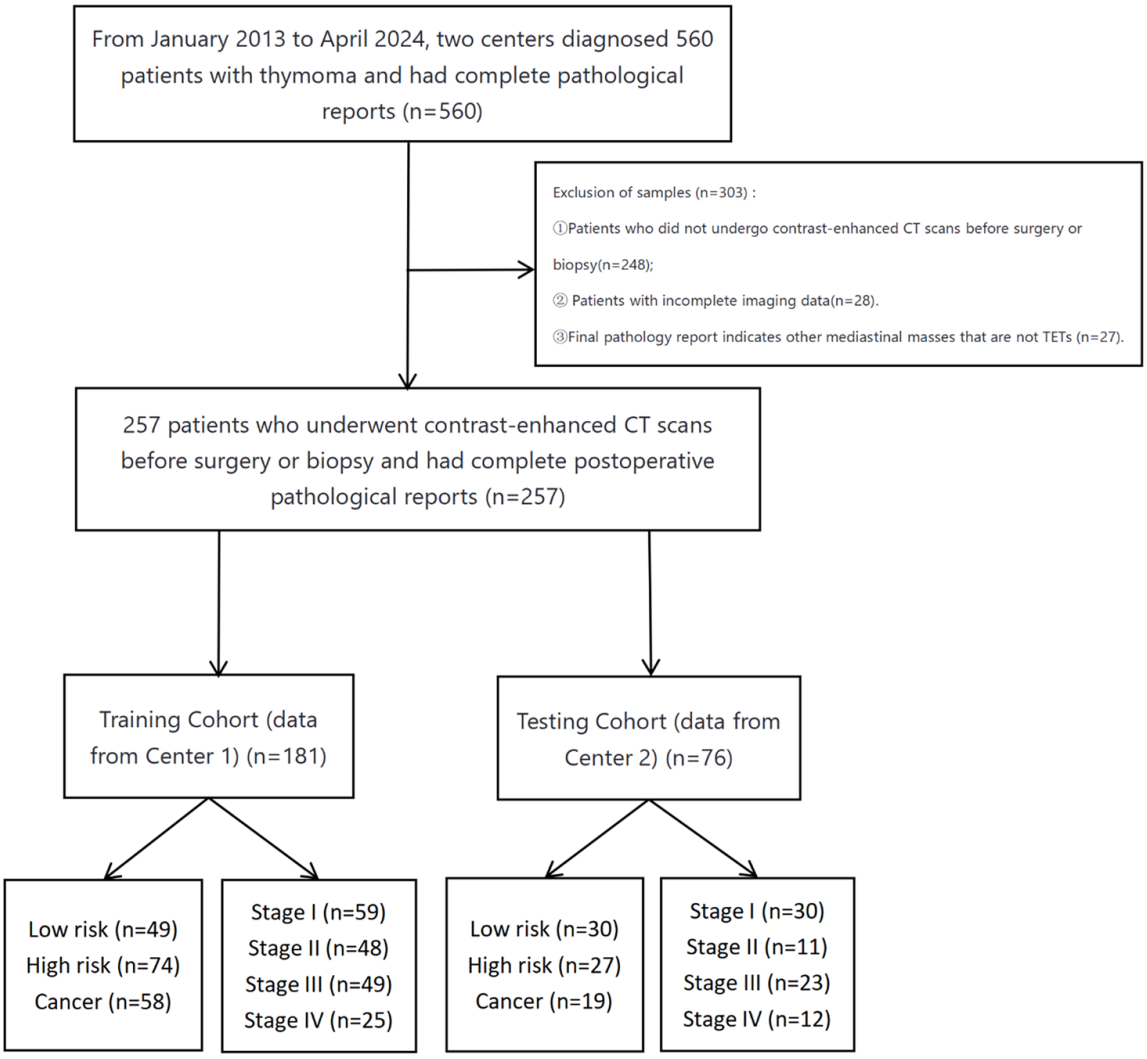


Figure 1. Flowchart for selecting cases of thymic epithelial tumors. CT, computed tomography; TETs, thymic epithelial tumor.

at the First Affiliated Hospital of Soochow University and the Second Affiliated Hospital of Soochow University from January 2013 to April 2024. Inclusion criteria: ① Patients pathologically confirmed as TETs; ② Patients with contrast-enhanced CT during the venous phase and complete postoperative pathological reports before surgery or biopsy. Exclusion criteria: ① Patients who did not undergo enhanced CT examination before surgery; ② Patients without complete pathological reports; ③ Patients with incomplete imaging data; ④ Final pathological reports indicating non-thymic epithelial tumors. The specific flowchart for including and excluding samples is as below (**Figure 1**).

Data collection

We collected patient characteristics from the inpatient system of two centers, the PACS imaging platform, and the LIS platform reports. Clinical characteristics included age, gender, cough, chest pain, chest tightness, myasthenia gravis, white blood cell count, and LDH. Semantic features were determined after reaching a consensus between two radiologists, including calcification, hemorrhage or necrosis, cystic change, ill-defined borders, adjacent lung changes, mediastinal lymph node enlargement, vascular invasion, effusion conditions, heterogeneous enhancement, tumor longitudinal diameter, and shape [30]. Two clinical physi-

cians, based on the 2015 WHO classification and the eighth edition of the TNM staging system, ultimately determined the pathological classification and staging for each patient and reached a consensus.

Image acquisition

The inspection equipment used in both institutions is CT, with the machine model being a 256-slice CT scanner (GE Revolution CT, GE, USA). Patients undergo breath-holding training before the examination. CT scan parameters: 120 kVp tube voltage, 200-450 automatic mAs tube current, pitch 0.992, slice spacing 5 mm, slice thickness 5 mm, image matrix 400*400, reconstruction slice thickness range of 1.25 mm. The contrast agent used for enhanced scanning is iodixanol 320, with an injection dose of 1.4 ml/kg, injection rate of 3 ml/s. The scan is automatically triggered at 100 HU during the arterial phase starting from the injection, and image acquisition is performed 30 seconds later to obtain venous phase images of the chest.

Image segmentation

We performed manual image segmentation using the 3D-slicer software version 5.6.2 (www.slicer.org). In this case series study, we used DICOM format (Digital Imaging and Communications in Medicine) venous phase enhanced CT images for subsequent processing. Initially, a clinical physician independently delineated the volume of interest (VOI) on each CT image, which was then reviewed and confirmed by another experienced radiologist to determine the final VOI for the next step of radiomics analysis.

Feature extraction and selection

Extract radiomics features from the VOIs of all patients' venous phase enhanced CT images using the Radiomics plugin in the 3D-Slicer software version 5.6.2. We have checked all types of features, including first-order statistical features, gray level co-occurrence matrix, gray level dependence matrix, gray level run length matrix, gray level size zone matrix, neighborhood gray tone difference matrix, shape features, two-dimensional shape features. And set the resampled voxel size to (3,3,3), indicating that the voxel size is resampled to 3 in all three dimensions. We set the kernel size of the LoG (Laplacian of Gaussian) filter to (3,4), indicating that different kernel sizes of the LoG filter are used in two dimensions. Next, we select wavelet-based features, indicating that we will use wavelet-based features for analysis. In the end, we extracted a total of 1038 features for subsequent work.

In the training dataset, continuous and categorical missing values are filled with the median and mode respectively. Normalize the continuous data in clinical features and radiomics features so that the mean is 0 and the variance is 1. For the categorical features in clinical semantic features, we adopt one-hot encoding to numerically process different categories for subsequent analysis. For feature selection in clinical semantic models, we adopt the LASSO method to select features with non-zero correlation coefficients, the specific LASSO selection process is as follows: First, the dataset is divided into 10 subsets; then each subset is used as a validation set, with the remaining 9 subsets serving as the training set, and Lasso regression is performed using different λ values. Finally, the average cross-validation error corresponding to each λ value is calculated, and the λ value that minimizes the average error (i.e., lambda.min) is selected. Subsequently, Lasso regression is run on the entire training set using this λ value to determine the final set of features (**Figures 4A**, **4D**, **5A**, **5D**). For feature selection in radiomics models, we first compress all radiomics features using PCA and select the first 12 principal components that account for 80% of the cumulative variance (**Figure 3A**). Then the LASSO regression was applied again to the 12 PCs using the same principle to select non-zero coefficient features (**Figures 4B**, **4E**, **5B**, **5E**). For feature selection of the fusion model, we performed PCA on all clinical features and 12 radiomic features, selecting the top 21 principal components using the same method (**Figure 3B**), and then we also utilize the same principle to perform LASSO regression on 21 PCs to screen for non-zero coefficient features (**Figures 4C**, **4F**, **5C**, **5F**). Since our feature selection process requires compressing all features based on the final three-classlabels or four-class-labels for relevance, therefore, when selecting the non-zero coefficient features to establish the final model, it is necessary to take the intersection

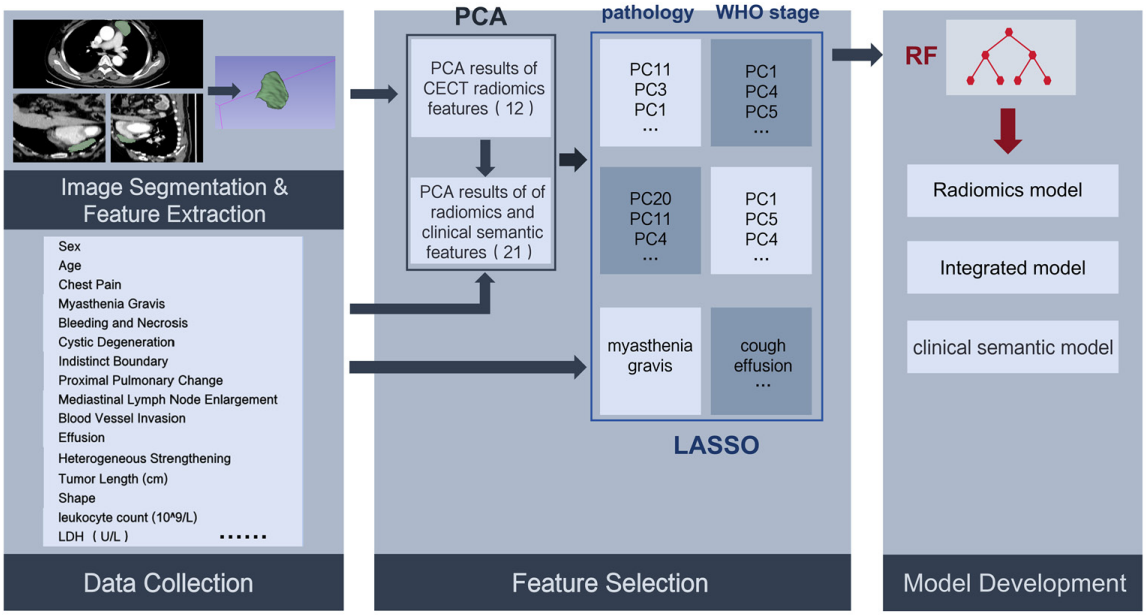


Figure 2. The technical roadmap of this study. CECT, Contrast-enhanced computed tomography; PCA, Principal Component Analysis; LASSO, least absolute shrinkage and selection operator; RF, Random Forest.

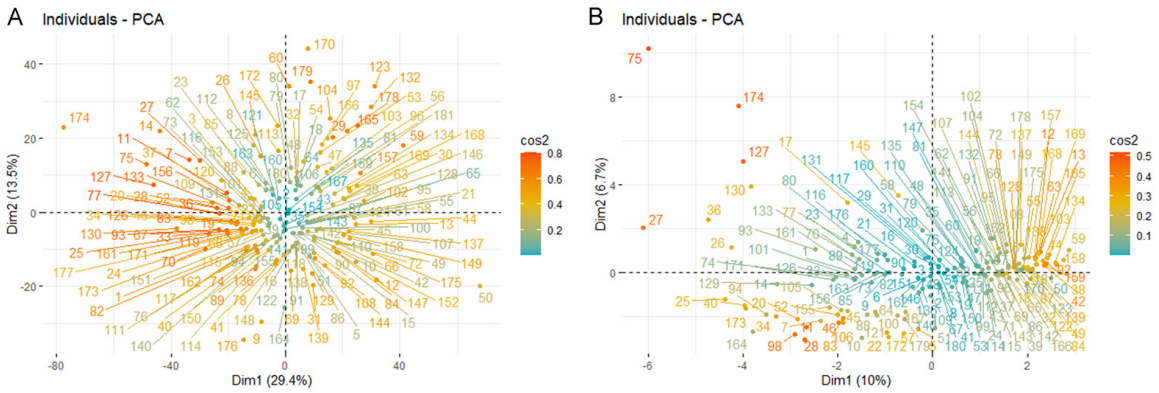


Figure 3. The PCA plot of radiomics features and fused features. The principal components can explain 80% of the variation in the original data. The plot shows the projection of individuals on the first (Dim1) and second (Dim2) principal components. The color indicates the strength of the correlation of variables with the principal components (cos^2 values), with darker colors representing stronger correlations. A. The image is the PCA result plot for radiomics features. The first 12 principal components were selected, and these components cumulatively explained 80% of the variance. B. The image is the result plot after PCA of radiomics features combined with clinical semantic features underwent PCA again. The first 21 principal components were selected, and these components cumulatively explained 80% of the variance as well.

of the compression results for each label. This results in the number of features used to build our model being more than the number indicated by the lambda.min dotted line in the cross-validation graph. The obtained feature coefficient values after compression are detailed in the supplementary materials (Tables S1, S2, S3). Afterwards, we introduce the normalization parameters of the training dataset into the test dataset, perform feature normalization on the dataset to obtain the same features for subsequent model testing (**Figure 2**).

Model developed and validation

On the training dataset, we established three classification models using the random forest classifier, which are the clinical semantic

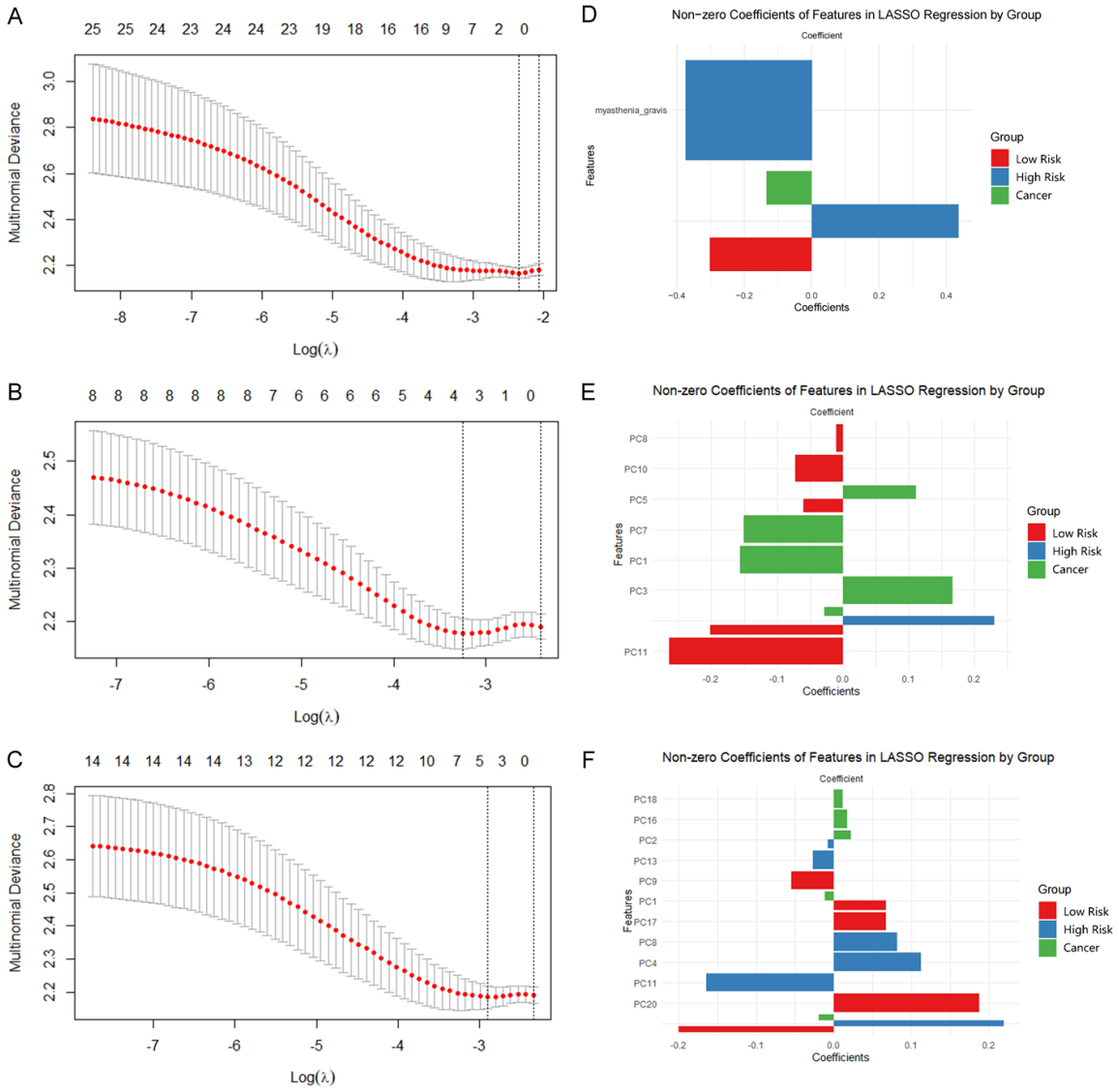


Figure 4. Training pathological classification model with LASSO cross-validation and coefficient distribution chart. The feature without annotation at the bottom is the intercept term, and the absolute value of the coefficients of each feature decreases from bottom to top. A, D. The images are the feature selection results of the clinical semantic model. B, E. The images are the feature selection results of the radiomics model. C, F. The images are the feature selection results of the fusion model.

model, the radiomics model, and the fusion model, and we used grid search to find the optimal parameter combination. Subsequently, we will train and test each model separately on the training and test datasets. Finally, we will compare the performance of these three types of models. For multi-class models, we use accuracy, precision, F1 score, recall, and confusion matrix to evaluate the overall performance of different models, and we employ the One vs Rest strategy to transform multi-class problems into binary classification problems for internal classification efficiency testing of the models, obtaining the ROC curve for each classification label, and calculating the AUC value.

Implementation and hardware

Data cleaning, feature selection, model training, and testing were all conducted in R software (version 4.4.1). The R packages used include “caret”, “glmnet”, “randomForest”, and others.

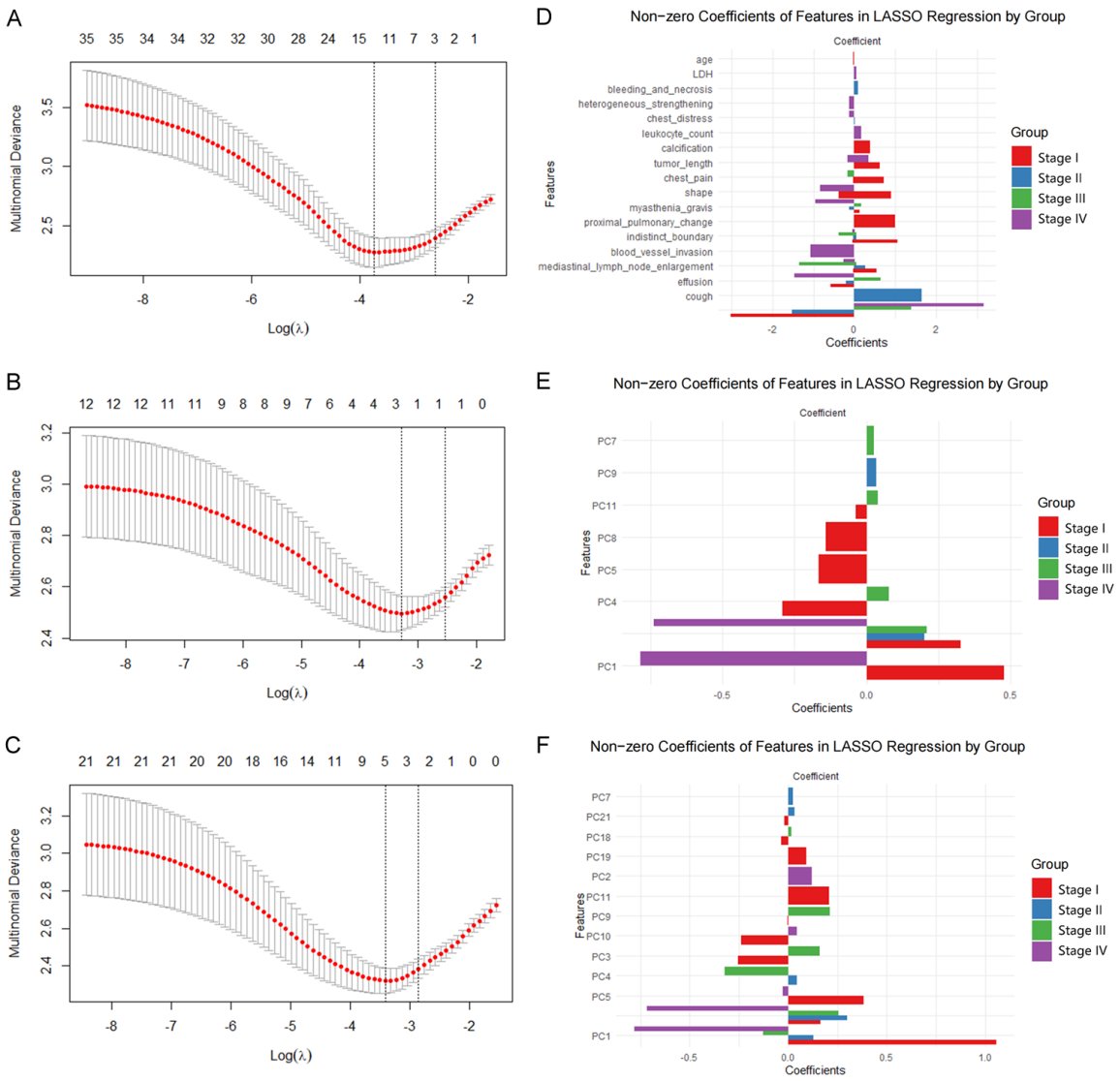


Figure 5. Training the LASSO cross-validation and coefficient distribution chart for the WHO staging model. The feature without annotation at the bottom is the intercept term, and the absolute value of the coefficients of each feature decreases from bottom to top. A, D. The images are the feature selection results of the clinical semantic model. B, E. The images are the feature selection results of the radiomics model. C, F. The images are the feature selection results of the fusion model.

Statistical analysis

Statistical analyses were performed using R and SPSS (Version 27.0). For quantitative data, the Kolmogorov-Smirnov test was first applied to assess whether continuous data followed a normal distribution. If the data conformed to a normal distribution, continuous variables were summarized as mean ± standard deviation; otherwise, they were expressed as median (interquartile range). For comparisons of continuous variables between two groups, Student's t-test or the Mann-Whitney U test was used depending on whether the data followed a normal distribution. Specifically, the t-test was employed when the data met normality assumptions; otherwise, the nonparametric Mann-Whitney U test was applied to evaluate differences between groups. Categorical variables were analyzed using the chi-square test. In cases with small sample sizes or expected frequencies below 5, Fisher's exact test was adopted to ensure result accuracy. All statistical tests were two-tailed, with values repre-

Table 1. The clinical baseline data of patients with different pathological types and WHO stages

Characteristics	Training Cohort	Testing Cohort	*P* Value
Sex			
Male	98 (54.14%)	38 (50%)	0.5436
Female	83 (45.86%)	38 (50%)	
Age	53 (44.5, 63)	54 (45.25, 62)	0.3472
Chest Distress			
Yes	26 (14.36%)	12 (15.8%)	0.769
No	155 (85.64%)	64 (84.2%)	
Chest Pain			
Yes	24 (13.26%)	15 (19.74%)	0.1866
No	157 (86.74%)	61 (80.26%)	
Cough			
Yes	17 (9.39%)	12 (15.8%)	0.1391
No	164 (90.61%)	64 (84.2%)	
Myasthenia Gravis			
Yes	23 (12.71%)	7 (9.21%)	0.4256
No	158 (87.29%)	69 (90.79%)	
Calcification			
Yes	33 (18.23%)	13 (17.11%)	0.8297
No	148 (81.77%)	63 (82.89%)	
Bleeding and Necrosis			
Yes	24 (13.26%)	5 (6.58%)	0.1224
No	157 (86.74%)	71 (93.42%)	
Cystic Degeneration			
Yes	29 (16.02%)	7 (9.21%)	0.151
No	152 (83.98%)	69 (90.79%)	
Indistinct Boundary			
Yes	31 (17.13%)	17 (22.37%)	0.3251
No	150 (82.87%)	59 (77.63%)	
Proximal Pulmonary Change			
Yes	155 (85.64%)	71 (93.42%)	0.08
No	26 (14.36%)	5 (6.58%)	
Mediastinal Lymph Node Enlargement			
Yes	44 (24.31%)	26 (34.21%)	0.1037
No	137 (75.69%)	50 (65.79%)	
Blood Vessel Invasion			
Yes	6 (3.31%)	4 (5.26%)	0.4611
No	175 (96.69%)	72 (94.74%)	
Effusion			
No effusion	151 (83.43%)	62 (81.58%)	0.951
Pleural effusion	16 (8.84%)	8 (10.53%)	
Pericardial effusion	8 (4.42%)	4 (5.26%)	
Pleural and pericardial effusion	6 (3.31%)	2 (2.63%)	
Heterogeneous Strengthening			
Yes	94 (51.93%)	31 (40.79%)	0.1028
No	87 (48.07%)	45 (59.21%)	
Tumor Length (cm)			
< 5	103 (56.91%)	47 (61.84%)	0.7229
≥ 5 and < 10	71 (39.23%)	27 (35.53%)	
≥ 10	7 (3.86%)	2 (2.63%)	

Shape			
Round	88 (48.62%)	33 (43.42%)	0.2916
Lobulated	30 (16.57%)	19 (25%)	
Irregular	63 (34.81%)	24 (31.58%)	
leukocyte count (10^9/L)	6.12 (4.98, 7.645)	5.9 (4.9, 7.875)	0.984
LDH (U/L)	180 (161.95, 205.25)	176.5 (156.5, 198)	0.3222
WHO Stage			
I	59 (32.6%)	30 (39.47%)	0.2164
II	48 (26.52%)	11 (14.47%)	
III	49 (27.07%)	23 (30.27%)	
IV	25 (13.81%)	12 (15.79%)	
Pathological Type			
Low risk	49 (27.07%)	30 (39.47%)	0.1384
High risk	74 (40.88%)	27 (35.53%)	
Cancer	58 (32.05%)	19 (25%)	
Total	181	76	

LDH is the abbreviated form for Lactate Dehydrogenase. WHO stands for World Health Organization. The values in the table represent medians or specific quantities, with values in parentheses indicating quartiles or proportions. *P*-value < 0.05 indicates significant difference.

senting medians or specific quantities, and figures in parentheses indicating interquartile ranges or proportions. A *p*-value < 0.05 was considered statistically significant. Through these methods, we systematically analyzed and interpreted differences among datasets from different centers, providing robust data support for subsequent research.

Results

Cohort and clinical characteristics

The study cohort was composed of 181 patients from Center One and 76 patients from Center Two. We used the samples from Center One for model training and the samples from Center Two for performance testing. For continuous data, we use the Mann-Whitney U test for data analysis and represent the results with medians and quartiles; for categorical data, we use the chi-square test and represent their distribution with percentages, in order to compare the characteristic distributions of the two cohorts. The clinical baseline data of patients with different pathological types and WHO stages are shown in **Table 1**. Ultimately, we found that there were no significant differences in the distribution of patients' gender, age, and all other clinical baseline characteristics between the different queues in the two centers.

Performance of the different models

To comprehensively evaluate the model's performance, we primarily conduct an overall assessment from a macro perspective and supplement this with micro indicators to evaluate the performance of each model's internal operations.

At the macro level, in the training queue, the accuracy of the fusion model for predicting WHO pathological classification is 0.923 (95% CI: 0.884-0.962), macro precision is 0.928 (95% CI: 0.738-0.942), macro recall is 0.915 (95% CI: 0.736-0.94), and macro F1 score is 0.92 (95% CI: 0.703-0.945). The accuracy of the fusion model for predicting WHO staging is 0.9 (95% CI: 0.844-0.935), macro precision is 0.883 (95% CI: 0.72-0.929), macro recall is 0.866 (95% CI: 0.723-0.917), and macro F1 score is 0.873 (95% CI: 0.756-0.924); the accuracy of the radiomics model for predicting WHO pathological classification is 0.873 (95% CI: 0.844-0.935), macro precision is 0.881 (95% CI: 0.718-0.927), macro recall is 0.862 (95% CI: 0.723-0.926), and macro F1 score is 0.869 (95% CI: 0.72-0.927), the accuracy of the radiomics model for predicting WHO staging is 0.856 (95% CI: 0.805-0.907), macro precision is 0.852 (95% CI: 0.805-0.914), macro recall is 0.839 (95% CI: 0.636-0.85), and macro F1 score is 0.841 (95% CI: 0.649-0.876);

the accuracy of the clinical semantic model for predicting WHO pathological classification is 0.663 (95% CI: 0.594-0.731), macro precision is 0.685 (95% CI: 0.519-0.75), macro recall is 0.633 (95% CI: 0.518-0.693), and macro F1 score is 0.64 (95% CI: 0.511-0.677); the accuracy of the clinical semantic model for predicting WHO staging is 0.707 (95% CI: 0.641-0.773), macro precision is 0.784 (95% CI: 0.663-0.801), macro recall is 0.667 (95% CI: 0.579-0.801), and macro F1 score is 0.684 (95% CI: 0.525-0.757). In the test queue, the accuracy of the fusion model predicting WHO pathological typing is 0.908 (95% CI: 0.843-0.973), macro precision is 0.937 (95% CI: 0.766-0.949), macro recall is 0.882 (95% CI: 0.765-0.927), and macro F1 score is 0.896 (95% CI: 0.775-0.955); the accuracy of the fusion model predicting WHO staging is 0.803 (95% CI: 0.713-0.892), macro precision is 0.878 (95% CI: 0.721-0.947), macro recall is 0.848 (95% CI: 0.723-0.948), and macro F1 score is 0.833 (95% CI: 0.667-0.888); the accuracy of the radiomics model predicting WHO pathological typing is 0.737 (95% CI: 0.638-0.836), macro precision is 0.829 (95% CI: 0.712-0.937), macro recall is 0.701 (95% CI: 0.613-0.823), and macro F1 score is 0.715 (95% CI: 0.65-0.876); the accuracy of the radiomics model predicting WHO staging is 0.75 (95% CI: 0.653-0.847), macro precision is 0.783 (95% CI: 0.645-0.847), macro recall is 0.62 (95% CI: 0.547-0.654), and macro F1 score is 0.6 (95% CI: 0.59-0.67); the accuracy of the clinical semantic model predicting WHO pathological typing is 0.658 (95% CI: 0.551-0.765), macro precision is 0.656 (95% CI: 0.605-0.703), macro recall is 0.667 (95% CI: 0.604-0.697), and macro F1 score is 0.657 (95% CI: 0.601-0.663); the accuracy of the clinical semantic model predicting WHO staging is 0.632 (95% CI: 0.523-0.74), macro precision is 0.598 (95% CI: 0.48-0.652), macro recall is 0.632 (95% CI: 0.44-0.656), and macro F1 score is 0.565 (95% CI: 0.381-0.575). The confusion matrices for the training set and test set are shown in the figure (**Figures 6**, **7**), with specific evaluation metrics presented in **Tables 2** and **3**. From this, we can see that models combining clinical semantics and radiomics perform best among all predictive models used for the same purpose, with radiomics-only models performing slightly worse than the combined models, and clinical semantics-only models having the lowest predictive performance.

At the micro level, it can be seen that the overall trend of micro-performance is consistent with the overall trend of macro-performance in both the test queue and the training queue (**Figures 8**, **9**; **Tables 4**, **5**). However, there is still a significant gap in the predictive performance of the model for different classifications. In the test queue, the fusion model of pathological typing performs best in predicting thymic carcinoma (AUC=0.733, 95% CI: 0.617-0.849), but performs worst in predicting high-risk thymoma (AUC=0.515, 95% CI: 0.377-0.653); the radiomics model of pathological typing performs better than the other two in predicting high-risk thymoma (AUC=0.705, 95% CI: 0.585-0.825). It is noteworthy that when predicting WHO stages in the test queue, the clinical semantic model's predictions for each stage are slightly better than the results of the fusion model. The fusion model performs best in predicting Stage I (AUC=0.898, 95% CI: 0.83-0.966) and Stage IV (AUC=0.841, 95% CI: 0.75-0.933), while the clinical semantic model outperforms in predicting Stage I, II, and III, but falls slightly short in predicting Stage IV (AUC=0.72, 95% CI: 0.566-0.874). However, this does not refute the conclusions drawn at the macro level; it merely indicates that the fusion model may perform poorly on certain feature combinations, especially in complex or difficult-to-distinguish stages (such as Stage II and III).

In summary, although we have demonstrated and compared the internal performance of the model at the micro level, this serves only as a reference indicator. Our main focus remains on the superiority or inferiority at the macro level.

Discussion

In this study, we developed three types of models: a prediction model based on clinical semantic features, a prediction model based on radiomic features from enhanced CT, and a prediction model that integrates clinical features with radiomic features. Each model was independently tested using an external center cohort. The prediction model that combined clinical semantic features with radiomic features achieved more accurate results in predicting pathological histological subtypes and

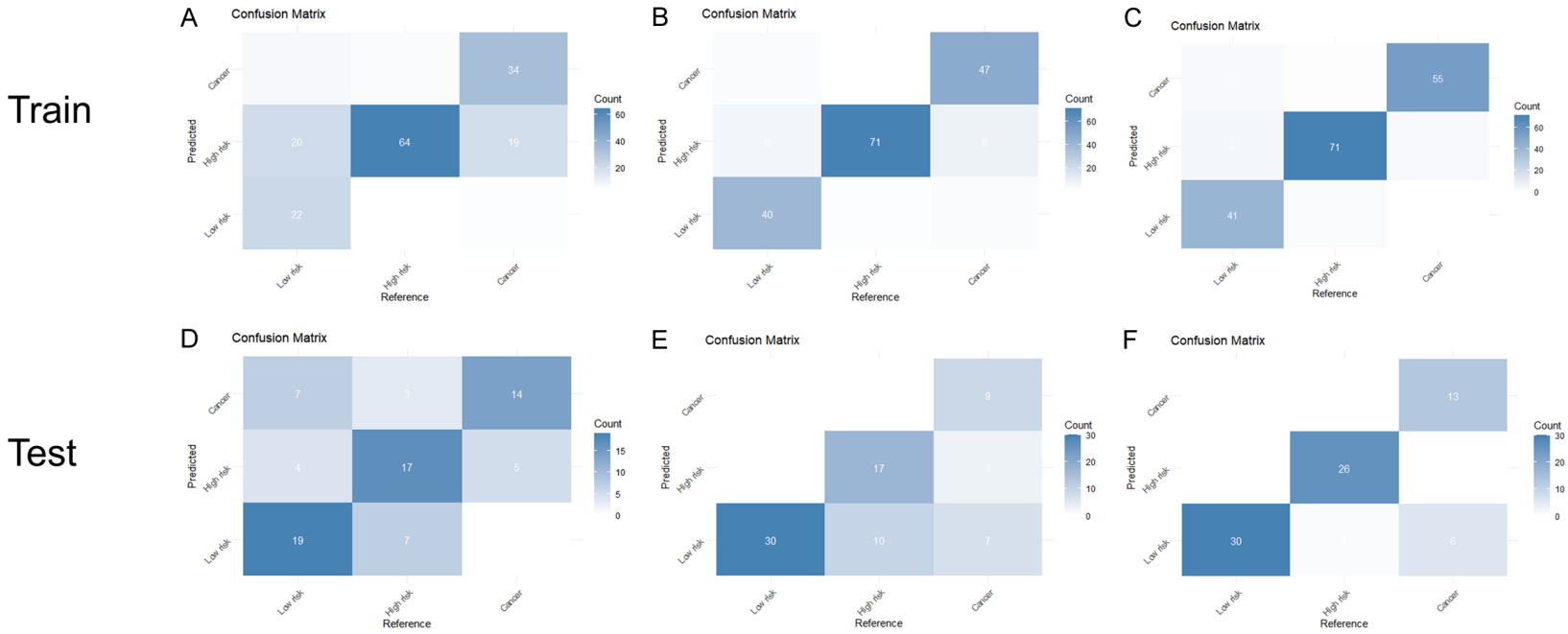


Figure 6. Confusion matrices for each pathological subtype model on the test set and training set. Each column represents the actual category (Reference), and each row represents the predicted category (Predicted). Each matrix shows the relationship between the predicted and actual values for different categories. The darker the color, the higher the count value in that cell. The test set and training set are marked on the graph. A-C. The images are the clinical semantic model, radiomics model, and fusion model, respectively. D-F. The images are similar.

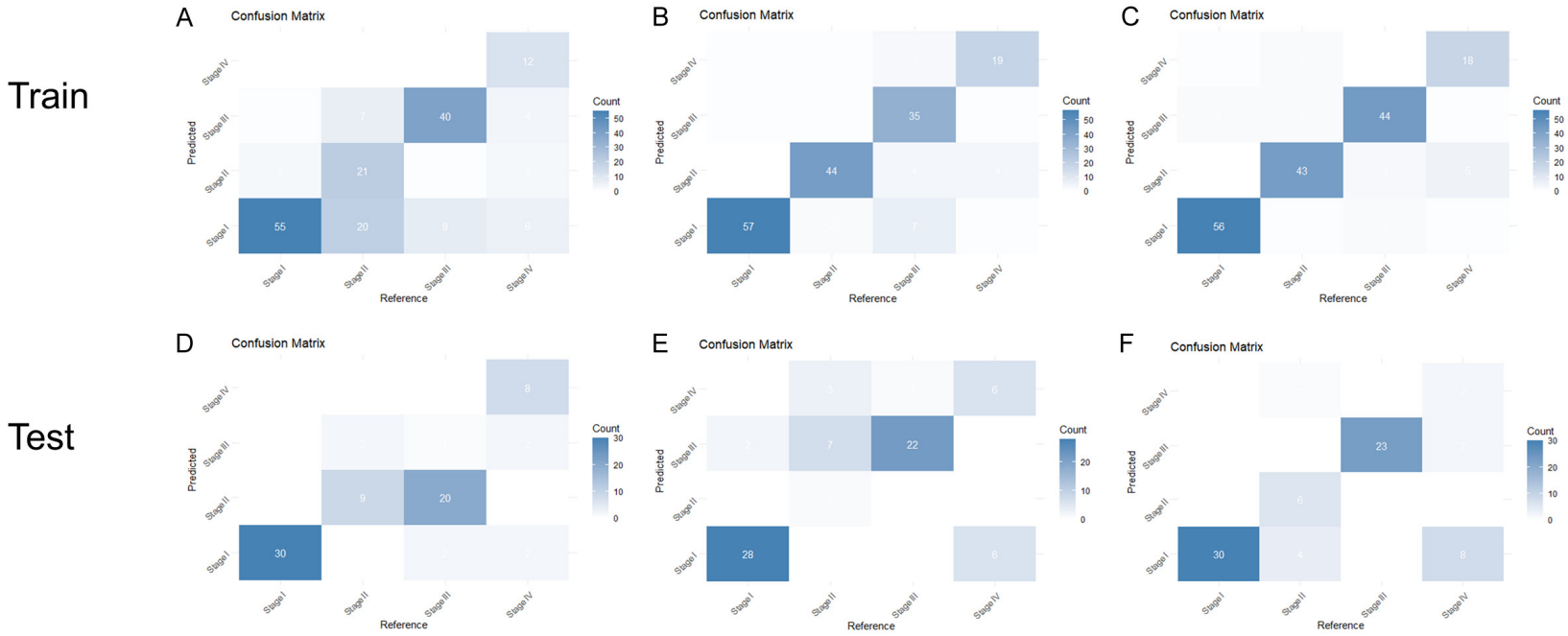


Figure 7. The confusion matrices for each WHO staging model on the test set and training set. Each column represents the actual class (Reference), and each row represents the predicted class (Predicted). Each matrix shows the relationship between the predicted and actual values for different classes. The darker the color, the higher the count value in that cell. The test set and training set are marked on the graph. A-C. The images are the clinical semantic model, radiomics model, and fusion model, respectively. D-F. The images are similar.

Table 2. The performance of the pathological classification model on the training set and test set

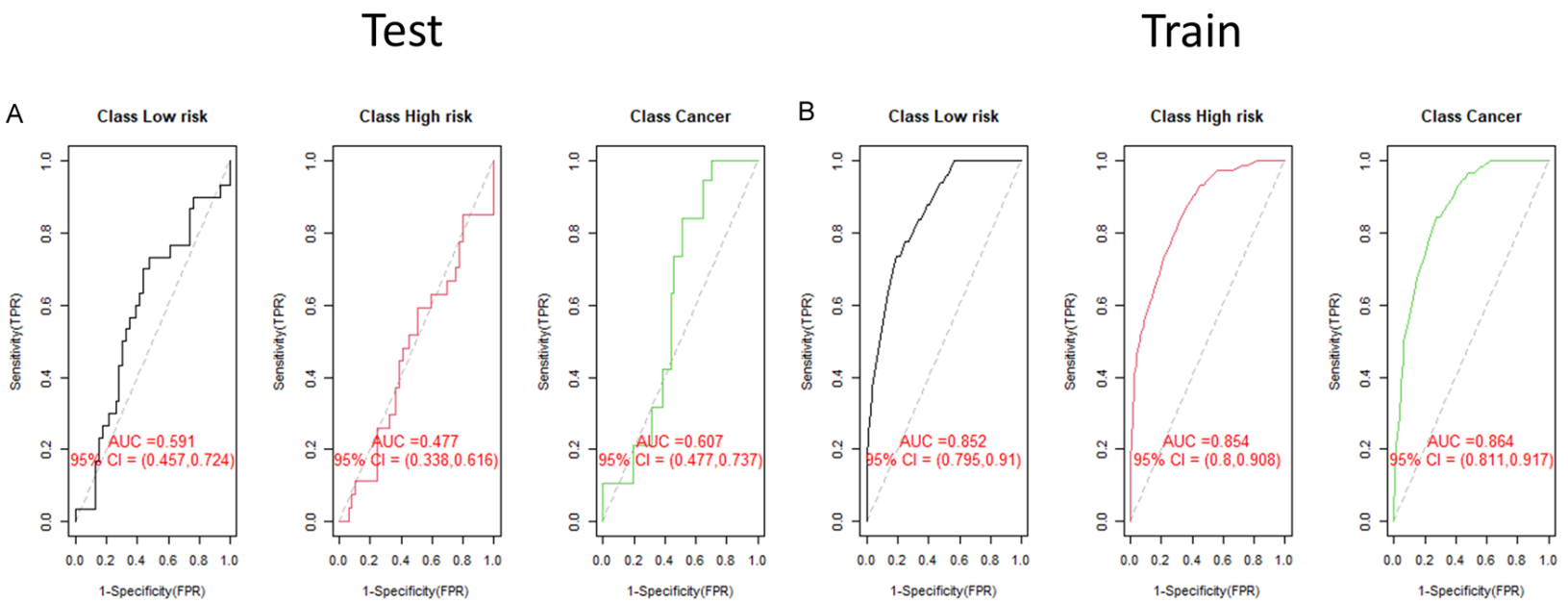
		Training Cohort			Testing Cohort		
		Clinical	Radiomics	Integrated	Clinical	Radiomics	Integrated
Pathological model	Accuracy	0.663 (95% CI: 0.594-0.731)	0.873 (95% CI: 0.844-0.935)	0.923 (95% CI: 0.884-0.962)	0.658 (95% CI: 0.551-0.765)	0.737 (95% CI: 0.638-0.836)	0.908 (95% CI: 0.843-0.973)
	Precision	0.685 (95% CI: 0.519-0.75)	0.881 (95% CI: 0.718-0.927)	0.928 (95% CI: 0.738-0.942)	0.656 (95% CI: 0.605-0.703)	0.829 (95% CI: 0.712-0.937)	0.937 (95% CI: 0.766-0.949)
	Recall Rate	0.633 (95% CI: 0.518-0.693)	0.862 (95% CI: 0.723-0.926)	0.915 (95% CI: 0.736-0.94)	0.667 (95% CI: 0.604-0.697)	0.701 (95% CI: 0.613-0.823)	0.882 (95% CI: 0.765-0.927)
	F1 Score	0.64 (95% CI: 0.511-0.677)	0.869 (95% CI: 0.72-0.927)	0.92 (95% CI: 0.703-0.945)	0.657 (95% CI: 0.601-0.663)	0.715 (95% CI: 0.65-0.876)	0.896 (95% CI: 0.775-0.955)

Precision, Recall Rate, and F1 Score are all macro indicators.

Table 3. The performance of the WHO staging model on the training set and test set

		Training Cohort			Testing Cohort		
		Clinical	Radiomics	Integrated	Clinical	Radiomics	Integrated
Staging model	Accuracy	0.707 (95% CI: 0.641-0.773)	0.856 (95% CI: 0.805-0.907)	0.9 (95% CI: 0.844-0.935)	0.632 (95% CI: 0.523-0.74)	0.75 (95% CI: 0.653-0.847)	0.803 (95% CI: 0.713-0.892)
	Precision	0.784 (95% CI: 0.663-0.801)	0.852 (95% CI: 0.805-0.914)	0.883 (95% CI: 0.72-0.929)	0.598 (95% CI: 0.48-0.652)	0.783 (95% CI: 0.645-0.847)	0.878 (95% CI: 0.721-0.947)
	Recall Rate	0.667 (95% CI: 0.579-0.801)	0.839 (95% CI: 0.636-0.85)	0.866 (95% CI: 0.723-0.917)	0.632 (95% CI: 0.44-0.656)	0.62 (95% CI: 0.547-0.654)	0.848 (95% CI: 0.723-0.948)
	F1 Score	0.684 (95% CI: 0.525-0.757)	0.841 (95% CI: 0.649-0.876)	0.873 (95% CI: 0.756-0.924)	0.565 (95% CI: 0.381-0.575)	0.6 (95% CI: 0.59-0.67)	0.833 (95% CI: 0.667-0.888)

Precision, Recall Rate, and F1 Score are all macro indicators.



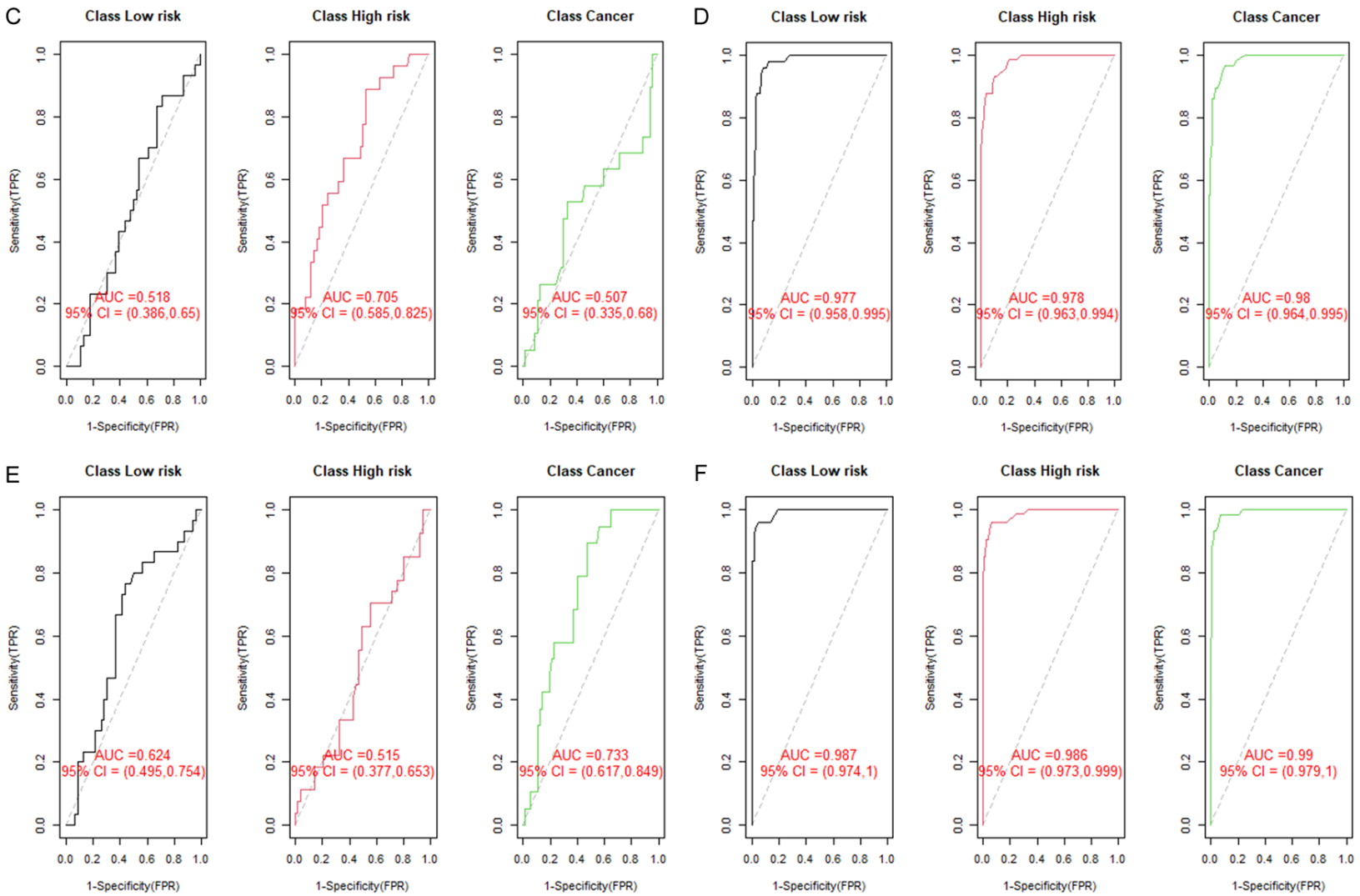
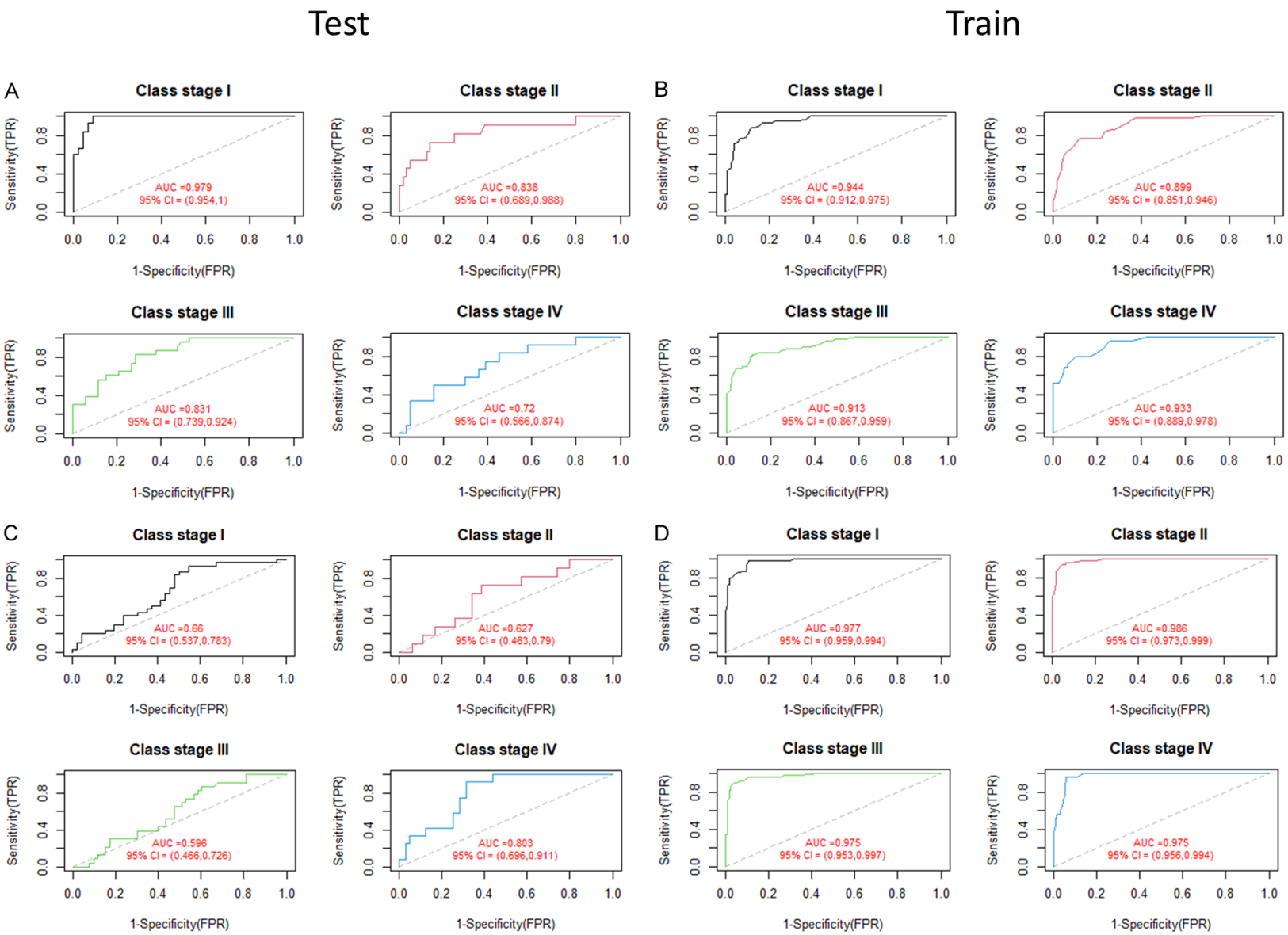


Figure 8. The ROC curves and AUC values for various pathological subtyping models on the test set and training set (One vs Rest). ROC, Receiver operating characteristic; AUC, Area under the curve of the receiver operating characteristic. The test set and training set are marked on the graph. A, B. The images are clinical semantic models. C, D. The images are radiomics models. E, F. The images are fusion models.


Test
Train
A
Class stage I
AUC =0.979
95% CI = (0.954,1)
Class stage II
AUC =0.838
95% CI = (0.689,0.988)
Class stage III
AUC =0.831
95% CI = (0.739,0.924)
Class stage IV
AUC =0.72
95% CI = (0.566,0.874)
B
Class stage I
AUC =0.944
95% CI = (0.912,0.975)
Class stage II
AUC =0.899
95% CI = (0.851,0.946)
Class stage III
AUC =0.913
95% CI = (0.867,0.959)
Class stage IV
AUC =0.933
95% CI = (0.889,0.978)
C
Class stage I
AUC =0.66
95% CI = (0.537,0.783)
Class stage II
AUC =0.627
95% CI = (0.463,0.79)
Class stage III
AUC =0.596
95% CI = (0.466,0.726)
Class stage IV
AUC =0.803
95% CI = (0.696,0.911)
D
Class stage I
AUC =0.977
95% CI = (0.959,0.994)
Class stage II
AUC =0.986
95% CI = (0.973,0.999)
Class stage III
AUC =0.975
95% CI = (0.953,0.997)
Class stage IV
AUC =0.975
95% CI = (0.956,0.994)
Sensitivity(TPR)
1-Specificity(FPR)
0.0
0.4
0.8
0.2
0.6
1.0

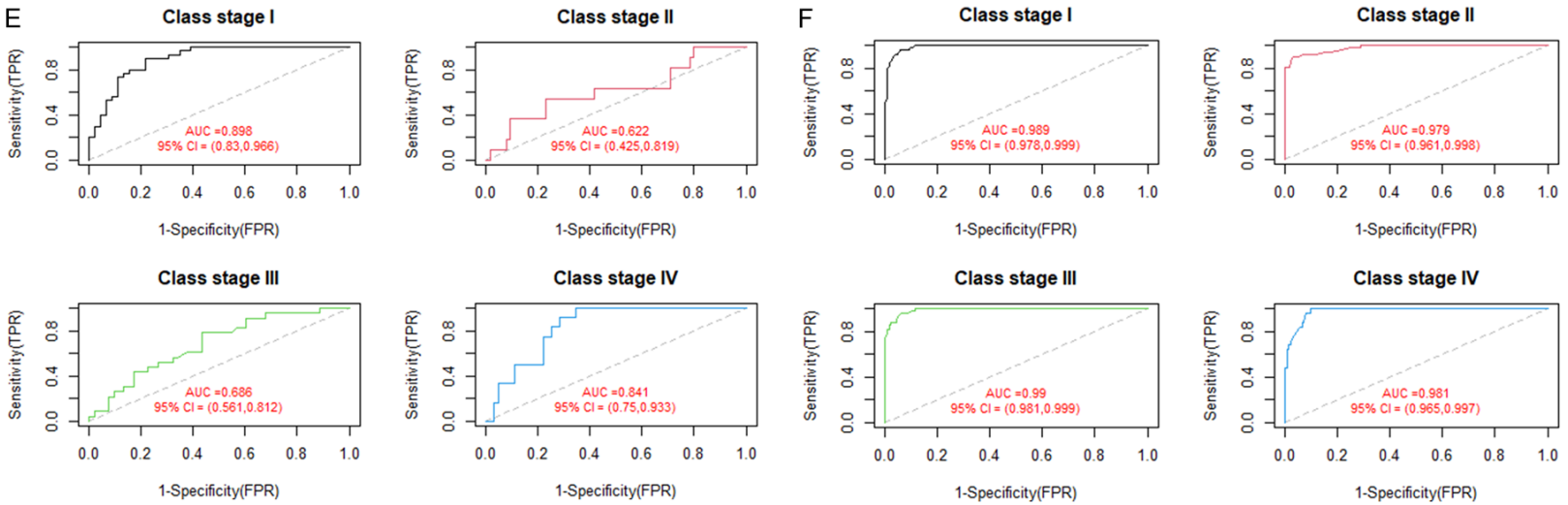


Figure 9. The ROC curves and AUC values (One vs Rest) for various WHO staging models on the test set and training set. ROC stands for Receiver operating characteristic; AUC stands for Area under the curve of the receiver operating characteristic. The test set and training set are marked on the graph. A, B. The images are clinical semantic models. C, D. The images are radiomics models. E, F. The images are fusion models.

Table 4. The AUC values for various pathological subtyping models on the test set and training set (One vs Rest)

		Training Cohort			Testing Cohort		
		Clinical model	Radiomics model	Integrated model	Clinical model	Radiomics model	Integrated model
AUC of Pathological model	Low Risk	0.852 (95% CI: 0.795-0.91)	0.977 (95% CI: 0.958-0.995)	0.987 (95% CI: 0.974-1)	0.591 (95% CI: 0.457-0.724)	0.518 (95% CI: 0.386-0.65)	0.624 (95% CI: 0.495-0.754)
	High Risk	0.854 (95% CI: 0.8-0.908)	0.978 (95% CI: 0.963-0.994)	0.986 (95% CI: 0.973-0.999)	0.477 (95% CI: 0.338-0.616)	0.705 (95% CI: 0.585-0.825)	0.515 (95% CI: 0.377-0.653)
	Cancer	0.864 (95% CI: 0.811-0.917)	0.98 (95% CI: 0.964-0.995)	0.99 (95% CI: 0.979-1)	0.607 (95% CI: 0.477-0.737)	0.507 (95% CI: 0.335-0.68)	0.733 (95% CI: 0.617-0.849)

Table 5. The AUC values (One vs Rest) for various WHO staging models on the test set and training set

		Training Cohort			Testing Cohort		
		Clinical model	Radiomics model	Integrated model	Clinical model	Radiomics model	Integrated model
AUC of Staging model	Stage I	0.944 (95% CI: 0.912-0.975)	0.977 (95% CI: 0.959-0.994)	0.989 (95% CI: 0.978-0.999)	0.979 (95% CI: 0.954-1)	0.66 (95% CI: 0.537-0.783)	0.898 (95% CI: 0.83-0.966)
	Stage II	0.899 (95% CI: 0.851-0.946)	0.986 (95% CI: 0.973-0.999)	0.979 (95% CI: 0.961-0.998)	0.838 (95% CI: 0.689-0.988)	0.627 (95% CI: 0.463-0.79)	0.622 (95% CI: 0.425-0.819)
	Stage III	0.913 (95% CI: 0.867-0.959)	0.975 (95% CI: 0.953-0.997)	0.99 (95% CI: 0.981-0.999)	0.831 (95% CI: 0.739-0.924)	0.596 (95% CI: 0.466-0.726)	0.686 (95% CI: 0.561-0.812)
	Stage IV	0.933 (95% CI: 0.889-0.978)	0.975 (95% CI: 0.956-0.994)	0.981 (95% CI: 0.965-0.997)	0.72 (95% CI: 0.566-0.874)	0.803 (95% CI: 0.696-0.911)	0.841 (95% CI: 0.75-0.933)

WHO staging compared to the other two models, demonstrating the best overall performance. This indicates that the combination of clinical semantic features and radiomic features can significantly enhance the accuracy of TETs diagnosis.

By 2012, there had been reports on the CT manifestations of thymic tumors with different pathological subtypes, but the number of cases in these reports was limited. To explore the relationship between the CT manifestations and pathological subtypes of thymic epithelial tumors (TETs), Liu et al. conducted a retrospective analysis of 105 cases of thymic tumors and concluded that there were statistically significant differences ($P < 0.05$) in tumor size, shape, necrosis or cystic change, integrity of the capsule, invasion of adjacent tissues, lymphadenopathy, and the presence of pleural effusion among different pathological types of thymoma [11]. Zhao et al.'s study confirmed the adequacy of CT manifestations in predicting tumor contours, homogeneity, degree of enhancement, peritumoral fat infiltration, mediastinal lymphadenopathy, irregular infiltration into the lung, and tumor shape based on the WHO histological classification. The study also indicated that lobulated or irregular tumor contours are characteristics predictive of a more aggressive subgroup [31]. One study found that a high white blood cell count was associated with disease recurrence in a cohort with a rich thymoma (> 90%) [32]. Compared to thymomas, thymic carcinomas and neuroendocrine tumors have lower white blood cell counts. In their study, Daniel et al. compared the white blood cell counts, circulating CRP, and LDH levels among the three major histological subgroups of thymic epithelial tumors. The final results indicated that elevated LDH levels are associated with thymic neuroendocrine tumors compared to thymomas or thymic carcinomas [32]. However, the significant variation in LDH levels within the same histological entity limits its diagnostic application [32]. Interestingly, when comparing patients with Masaoka-Koga stage III-IV thymomas to those with stage I-II, there is a significant increase in LDH levels [32]. Our study synthesized the conclusions of previous studies, collecting and summarizing all confirmed or potential histological subtype classifications and WHO staging characteristics. After feature selection, we only found a certain correlation between myasthenia gravis and high-risk thymoma, which is consistent with the findings of Cangir et al. [12], although using myasthenia gravis as a single feature for predicting pathological subtypes is not very effective. Additionally, the emergence of symptoms is more strongly associated with later WHO stages. At the same time, we also found that LDH and white blood cell count have a weak association with WHO stages, thus our study corroborates and extends the perspectives of some previous studies. Moreover, radiomic features have been shown to contribute to improving classification accuracy. By integrating these more detailed clinical features with radiomic features, the overall performance of the radiomic models has been significantly improved and enhanced.

In previous studies, scholars have focused on the application of radiomics in distinguishing different histological subtypes and stages of thymic epithelial tumors (TETs). Predictions regarding TETs histological subtypes have primarily been based on the use of preoperative imaging data to differentiate between low-risk and high-risk thymomas. In the field of traditional machine learning radiomics, Cangir et al. utilized six classifiers to construct a model based on radiomics from preoperative contrast-enhanced computed tomography (CECT) of 83 TETs patients to distinguish between low-risk and high-risk thymomas, finding that the AUC for radiomic features using the K-nearest neighbors (KNN) classifier was 0.943 [12]. Hu et al. constructed radiomics models based on preoperative CECT and UECT of 155 TETs patients using four machine learning classifiers, and ultimately found that the RF classifier performed best when UECT and CECT were used together (0.87, 95% CI: 0.80-0.92) [25]. Deep learning technology has also been widely applied in binary classification research, particularly convolutional neural networks (CNN) for extracting complex image features from preoperative CT images [13, 23]. The study by Liu et al. selected multicenter samples and created a deep learning signature (DLS) using deep learning features extracted from all lesions with convolutional neural networks. They found that the combination of subjective CT features (such as infiltration) and DLS performed best in distinguishing TETs risk status. The AUCs for the training, internal validation, external validation

1, and external validation 2 cohorts were 0.959 (95% confidence interval (CI): 0.924-0.993), 0.868 (95% CI: 0.765-0.970), 0.846 (95% CI: 0.750-0.942), and 0.846 (95% CI: 0.735-0.957), respectively [13]. Considering that thymic carcinoma is a group of heterogeneous tumors, including squamous cell carcinoma, adenocarcinoma, and undifferentiated carcinoma, these studies did not analyze thymic carcinoma and thymoma together in radiomics research [12]. However, a few scholars still include thymic carcinoma in the discussion of TETs pathological subtypes. In the field of traditional machine learning radiomics, Feng et al. used 14 machine learning models, along with different feature selection strategies, to establish a three-class radiomics model based on radiomic features, and combined with clinical variables, they established a clinical radiomics model that demonstrated superior diagnostic efficacy compared to a single radiomics model [28]. Liu et al. extracted radiomic features from the regions of interest in NECT and CECT images for each patient and compared models incorporating clinical and semantic features during the model construction process. They found that models combining radiomic features with clinical and semantic features achieved more precise predictive performance [30]. In addition, other imaging modalities are also used to construct three-class classification models, such as Xiao et al. explored the application of radiomic features based on different MRI sequences in TETs classification [29]. There are no studies on multi-classification models that have incorporated deep learning features yet. These studies have all presented meaningful conclusions, but they lack external central samples for independent testing. With the update of the staging system, it has become an inevitable trend to explore the correlation between different factors and characteristics and the various stages of the WHO, and to make predictions based on this. Yang et al. studied a preoperative staging tool that uses CT images of thymoma patients to differentiate between Masaoka-Koga (MK) stage I and stage II patients. They employed an artificial neural network (ANN) deep learning model, namely the 3D-DenseNet model, to distinguish between MK stage I and stage II thymomas. They found that deep learning has great potential in the preoperative staging of thymomas [33]. Bluthgen et al. evaluated the use of CT-derived radiomics for machine learning-based WHO staging, with RF showing good discriminative performance for early and late WHO stages (AUC, 83.8%; 95% CI, 66.9-93.4) [24]. Tian's study constructed a WHO early and late stage RF prediction model based on the radiomics data of preoperative NECT in TETs patients, with an AUC of 0.766 (95% CI, 0.642-0.886) [27]. This provides a greater practical basis for further detailed WHO staging predictions. We collected a total of 257 samples, not only establishing a larger research cohort for model building, but also including 76 external center independent samples for model testing. In the end, we constructed a three-category RF model for pathological histological subtypes and a four-category model for WHO staging, and discussed the model performance from both macroscopic and microscopic perspectives. Ultimately, our clinical semantic and CECT-based radiomics fusion model performed well in predicting pathological subtypes and WHO staging on the external test set (ACC=0.908, 95% CI: 0.843-0.973; ACC=0.803, 95% CI: 0.713-0.892). However, the performance of the staging fusion model at the microscopic level did not align with the macroscopic trend, possibly due to the difficulty in distinguishing certain radiomics features between stage II and stage III TETs, and in the one-to-rest strategy, each classification weight and the weight of macroscopic evaluation are also different [34].

Our study categorized the predictive targets into three pathological subtypes: low-risk thymoma, high-risk thymoma, and thymic carcinoma, as well as four WHO stages (I-IV). This differs from previous distinctions made between low and high-risk thymomas or between early and late-stage thymomas, as we have refined the predictive outcomes for greater accuracy. Additionally, we evaluated the model's performance using an independent external test cohort and achieved desirable results, which are also somewhat related to the method of feature fusion and selection. We used a method that more comprehensively covers different features when selecting characteristics. Principal Component Analysis (PCA) is a dimensionality reduction technique that simplifies the data structure by transforming multiple correlated variables into a few uncorrelated comprehensive variables, known as principal components, while retaining as much of the original

data's information as possible [35]. Other filtering methods, such as multivariate logistic regression, the simple lasso method may cause data loss, but this does not mean that PCA can reduce overfitting to some extent or even a great extent, therefore, PCA should not be regarded as the main method to prevent overfitting. PCA is an unsupervised learning method that does not consider labels, and therefore, important information for predicting labels may be lost during the dimensionality reduction process. Consequently, even after performing PCA on variables, we still need to use regularization terms, as this is a supervised learning paradigm that can consider label information while controlling the complexity of the model [35]. Thus, the features we obtain can significantly enhance the accuracy of the integrated model. Additionally, uniform normalization of the features of two queues may lead to data leakage from the test set, ultimately reducing the model's generalization ability. Therefore, we chose to process the test set data using the normalization parameters introduced during the training of the model, obtaining the same features. This approach can effectively prevent external data leakage from the test queue, while also allowing the variables from the training and test sets to be compared on the same scale.

This study undoubtedly has some limitations: (1) The VOI measurement location for radiomics feature segmentation is manually performed, which may lead to sampling bias. Different operators may choose different lesion locations, thereby affecting the consistency and accuracy of feature extraction. To improve the reliability of the results, future work can explore automated or semi-automated segmentation methods. (2) The VOI measurement location for radiomics feature segmentation is manually performed, which may lead to sampling bias. Different operators may choose different lesion locations, thereby affecting the consistency and accuracy of feature extraction. To improve the reliability of the results, future work can explore automated or semi-automated segmentation methods. (3) The CT scans conducted in the study were performed at several different hospitals without a standardized protocol, and different CT scanners produced by various companies were used for image acquisition. This diversity may affect the consistency and comparability of radiomic features. Future research should be conducted under unified standards for data collection to ensure the stability and reliability of the results. (4) Although our sample size is relatively large, it is still not sufficient to fully validate the generalization ability of the model. A larger sample size and a more diverse patient population will help to better assess the performance of the model. Future studies should further validate the effectiveness of the model through larger-scale prospective multicenter cohorts. (5) Limitations of a single imaging modality: Although radiomic features extracted from CT images show good prognostic value, other imaging modalities (such as MRI, PET-CT) can provide additional information. Combining multiple imaging modalities can not only enhance the predictive power of the model but also provide a more comprehensive diagnostic basis. Therefore, further research can improve the performance of the model by integrating multiple imaging modalities. (6) In the current study, the Random Forest (RF) model was adopted and achieved good results. However, existing research indicates that in certain specific situations, other machine learning models, such as Support Vector Machine (SVM) and Gradient Boosting Decision Tree (GBDT), may have better performance. Although we have not discussed these alternative models in detail in this paper, future research could consider exploring more classifiers to further optimize model performance and enhance its explanatory power. (7) Although the RF model performs excellently in predictive performance, its interpretability is relatively weak, especially when facing complex feature interactions. Future research can incorporate more interpretable models (such as logistic regression, linear regression, etc.), or use interpretability tools (such as SHAP values, LIME, etc.) to enhance the transparency and interpretability of the model [28].

Conclusion

This study provides a non-invasive imaging method to predict histological subtypes and WHO staging, avoiding the risks and discomfort associated with traditional invasive examinations such as biopsies. This represents an important advancement for patients, as it enhances the safety and comfort of diagnosis. Through the combination of radiomics and

machine learning techniques, we are able to more accurately identify different pathological histological subtypes and stages, thereby providing clinicians with more accurate diagnostic evidence. This assists in formulating personalized treatment plans and improving treatment outcomes. The good performance in the external independent test queue demonstrates that this method has strong generalization capabilities and is applicable to data from different medical institutions. This implies that the method is not limited to specific research environments and has a broad application prospect. Accurate pathological histological subtype and staging information is crucial for guiding subsequent treatment. The information obtained through imaging examinations can help doctors choose the most appropriate treatment method, thereby improving the patient's survival rate and quality of life. In summary, the radiomics-based approach proposed in this study not only provides a new tool for the diagnosis of TET patients but also demonstrates significant clinical value in improving diagnostic accuracy, reducing invasive procedures, and guiding personalized treatment.

Acknowledgements

Thanks to all the patients, healthcare professionals, and technical support experts who contributed to the study, as well as the research funding organizations that supported this project. This work was supported by the Jiangsu Provincial Medical Key Discipline (ZDXK202235).

Disclosure of conflict of interest

None.

Address correspondence to: Chao Xu and Jianping Wang, Department of Radiation Oncology, The First Affiliated Hospital of Soochow University, Pinghai Road No. 899, Suzhou 215006, Jiangsu, China. E-mail: xuchao87@suda.edu.cn (CX); wangjpsz@126.com (JPW); Jinzhou Zhu, Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Pinghai Road No. 899, Suzhou 215006, Jiangsu, China. E-mail: jzzhu@zju.edu.cn; Yongbing Chen, Department of Thoracic Surgery, The Second Affiliated Hospital of Soochow University, Sanxiang Road No. 1055, Suzhou 215004, Jiangsu, China. E-mail: chentongt@sina.com

References

[1] Roden AC, Ahmad U, Cardillo G, Girard N, Jain D, Marom EM, Marx A, Moreira AL, Nicholson AG, Rajan A, Shepherd AF, Simone CB 2nd, Strange CD, Szolkowska M, Truong MT and Rimner A. Thymic carcinomas-a concise multidisciplinary update on recent developments from the thymic carcinoma working group of the international thymic malignancy interest group. J Thorac Oncol 2022; 17: 637-650.

[2] Valavanis C, Stanc GM and Baltayiannis N. Classification, histopathology and molecular pathology of thymic epithelial tumors: a review. J BUON 2021; 26: 1198-1207.

[3] Xu C, Zhang Y, Wang W, Wang Q, Li Z, Song Z, Wang J, Yu J, Liu J, Zhang S, Cai X, Wu M, Zhan P, Liu H, Lv T, Miao L, Min L, Li J, Liu B, Yuan J, Jiang Z, Lin G, Chen X, Pu X, Rao C, Lv D, Yu Z, Li X, Tang C, Zhou C, Zhang J, Guo H, Chu Q, Meng R, Liu X, Wu J, Hu X, Fang M, Zhou J, Zhu Z, Chen X, Pan W, Pang F, Zhou Y, Jian Q, Wang K, Wang L, Zhu Y, Yang G, Lin X, Cai J, Liang L, Feng H, Wang L, Du Y, Yao W, Shi X, Niu X, Yuan D, Yao Y, Huang J, Zhang Y, Sun P, Wang H, Ye M, Wang D, Wang Z, Hao Y, Wang Z, Wan B, Lv D, Yu G, Li A, Kang J, Zhang J, Zhang C, Chen H, Shi L, Ye L, Wang G, Wang Y, Gao F, Zhou W, Hu C, Wei J, Li B, Li Z, Li Y, Liu Z, Yang N, Wu L, Wang Q, Huang W, Hong Z, Wang G, Fang M, Fang Y, Zhu X, Du K, Ji J, Shen Y, Zhang Y, Ma S, Song Y, Lu Y, Liu A, Fang W and Zhong W. Chinese expert consensus on the diagnosis and treatment of thymic epithelial tumors. Thorac Cancer 2023; 14: 1102-1117.

[4] Basse C and Girard N. Thymic tumours and their special features. Eur Respir Rev 2021; 30: 200394.

[5] Mahmoudi S, Gruenewald LD, Eichler K, Althoff FC, Martin SS, Bernatz S, Booz C, Yel I, Kinzler MN, Ziegengeist NS, Torgashov K, Mohammed H, Geyer T, Scholtz JE, Hammerstingl RM, Weber C, Hardt SE, Sommer CM, Gruber-Rouh T, Leistner DM, Vogl TJ and Koch V. Multiparametric evaluation of radiomics features and dual-energy CT Iodine maps for discrimination and outcome prediction of thymic masses. Acad Radiol 2023; 30: 3010-3021.

[6] Agrafiotis AC, Siozopoulou V, Hendriks JMH, Pauwels P, Koljenovic S and Van Schil PE. Prognostic factors and genetic markers in thymic epithelial tumors: a narrative review. Thorac Cancer 2022; 13: 3242-3249.

[7] Lichtenberger JP 3rd, Carter BW, Fisher DA, Parker RF and Peterson PG. Thymic epithelial neoplasms: radiologic-pathologic correlation. Radiol Clin North Am 2021; 59: 169-182.

[8] Xiao G, Rong WC, Hu YC, Shi ZQ, Yang Y, Ren JL and Cui GB. MRI radiomics analysis for predict-

ing the pathologic classification and TNM staging of thymic epithelial tumors: a pilot study. AJR Am J Roentgenol 2020; 214: 328-340.
[9] Muto Y and Okuma Y. Therapeutic options in thymomas and thymic carcinomas. Expert Rev Anticancer Ther 2022; 22: 401-413.
[10] Falkson CB, Vella ET, Ellis PM, Maziak DE, Ung YC and Yu E. Surgical, radiation, and systemic treatments of patients with thymic epithelial tumors: a systematic review. J Thorac Oncol 2023; 18: 299-312.
[11] Liu GB, Qu YJ, Liao MY, Hu HJ, Yang GF and Zhou SJ. Relationship between computed tomography manifestations of thymic epithelial tumors and the WHO pathological classification. Asian Pac J Cancer Prev 2012; 13: 5581-5585.
[12] Kayi Cangir A, Orhan K, Kahya Y, Özakıncı H, Kazak BB, Konuk Balcı BM, Karasoy D and Uzun Ç. CT imaging-based machine learning model: a potential modality for predicting low-risk and high-risk groups of thymoma: "Impact of surgical modality choice". World J Surg Oncol 2021; 19: 147.
[13] Chen X, Feng B, Xu K, Chen Y, Duan X, Jin Z, Li K, Li R, Long W and Liu X. Development and validation of a deep learning radiomics nomogram for preoperatively differentiating thymic epithelial tumor histologic subtypes. Eur Radiol 2023; 33: 6804-6816.
[14] Mayoral M, Pagano AM, Araujo-Filho JAB, Zheng J, Perez-Johnston R, Tan KS, Gibbs P, Fernandes Shepherd A, Rimner A, Simone II CB, Riely G, Huang J and Ginsberg MS. Conventional and radiomic features to predict pathology in the preoperative assessment of anterior mediastinal masses. Lung Cancer 2023; 178: 206-212.
[15] Gao C, Yang L, Xu Y, Wang T, Ding H, Gao X and Li L. Differentiating low-risk thymomas from high-risk thymomas: preoperative radiomics nomogram based on contrast enhanced CT to minimize unnecessary invasive thoracotomy. BMC Med Imaging 2024; 24: 197.
[16] Liu W, Wang W, Zhang H, Guo M, Xu Y and Liu X. Development and validation of multi-omics thymoma risk classification model based on transfer learning. J Digit Imaging 2023; 36: 2015-2024.
[17] Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY and Aerts HJWL. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019; 69: 127-157.
[18] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A and Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012; 48: 441-446.
[19] Syed AB and Zoga AC. Artificial intelligence in radiology: current technology and future directions. Semin Musculoskelet Radiol 2018; 22: 540-545.
[20] Ried M, Marx A, Götz A, Hamer O, Schalke B and Hofmann HS. State of the art: diagnostic tools and innovative therapies for treatment of advanced thymoma and thymic carcinoma. Eur J Cardiothorac Surg 2016; 49: 1545-1552.
[21] Liufu Y, Wen Y, Wu W, Su R, Liu S, Li J, Pan X, Chen K and Guan Y. Radiomics analysis of multiphasic computed tomography images for distinguishing high-risk thymic epithelial tumors from low-risk thymic epithelial tumors. J Comput Assist Tomogr 2023; 47: 220-228.
[22] Chen X, Feng B, Li C, Duan X, Chen Y, Li Z, Liu Z, Zhang C and Long W. A radiomics model to predict the invasiveness of thymic epithelial tumors based on contrast-enhanced computed tomography. Oncol Rep 2020; 43: 1256-1266.
[23] Liu W, Wang W, Guo R, Zhang H and Guo M. Deep learning for risk stratification of thymoma pathological subtypes based on preoperative CT images. BMC Cancer 2024; 24: 651.
[24] Blüthgen C, Patella M, Euler A, Baessler B, Martini K, von Spiczak J, Schneiter D, Opitz I and Frauenfelder T. Computed tomography radiomics for the prediction of thymic epithelial tumor histology, TNM stage and myasthenia gravis. PLoS One 2021; 16: e0261401.
[25] Hu J, Zhao Y, Li M, Liu Y, Wang F, Weng Q, You R and Cao D. Machine-learning-based computed tomography radiomic analysis for histologic subtype classification of thymic epithelial tumours. Eur J Radiol 2020; 126: 108929.
[26] Liang Z, Li J, Tang Y, Zhang Y, Chen C, Li S, Wang X, Xu X, Zhuang Z, He S and Deng B. Predicting the risk category of thymoma with machine learning-based computed tomography radiomics signatures and their between-imaging phase differences. Sci Rep 2024; 14: 19215.
[27] Tian D, Yan HJ, Shiiya H, Sato M, Shinozaki-Ushiku A and Nakajima J. Machine learning-based radiomic computed tomography phenotyping of thymic epithelial tumors: predicting pathological and survival outcomes. J Thorac Cardiovasc Surg 2023; 165: 502-516, e509.
[28] Feng XL, Wang SZ, Chen HH, Huang YX, Xin YK, Zhang T, Cheng DL, Mao L, Li XL, Liu CX, Hu YC, Wang W, Cui GB and Nan HY. Optimizing the radiomics-machine-learning model based on non-contrast enhanced CT for the simplified risk categorization of thymic epithelial tumors:

a large cohort retrospective study. Lung Cancer 2022; 166: 150-160.
[29] Xiao G, Hu YC, Ren JL, Qin P, Han JC, Qu XY, Rong WC, Yan WQ, Tian Q, Han Y, Wang WP, Wang SM, Ma J, Wang W and Cui GB. MR imaging of thymomas: a combined radiomics nomogram to predict histologic subtypes. Eur Radiol 2021; 31: 447-457.
[30] Liu J, Yin P, Wang S, Liu T, Sun C and Hong N. CT-based radiomics signatures for predicting the risk categorization of thymic epithelial tumors. Front Oncol 2021; 11: 628534.
[31] Zhao Y, Chen H, Shi J, Fan L, Hu D and Zhao H. The correlation of morphological features of chest computed tomographic scans with clinical characteristics of thymoma. Eur J Cardiothorac Surg 2015; 48: 698-704.
[32] Valdivia D, Cheufou D, Fels B, Puhlvers S, Mardanzai K, Zaatar M, Weinreich G, Taube C, Theegarten D, Stuschke M, Schuler M, Stamatis G, Hegedus B and Aigner C. Potential prognostic value of preoperative leukocyte count, lactate dehydrogenase and c-reactive protein in thymic epithelial tumors. Pathol Oncol Res 2021; 27: 629993.
[33] Yang L, Cai W, Yang X, Zhu H, Liu Z, Wu X, Lei Y, Zou J, Zeng B, Tian X, Zhang R, Luo H and Zhu Y. Development of a deep learning model for classifying thymoma as Masaoka-Koga stage I or II via preoperative CT images. Ann Transl Med 2020; 8: 287.
[34] Park D, Cho JM, Yang JW, Yang D, Kim M, Oh G and Kwon HD. Classification of expert-level therapeutic decisions for degenerative cervical myelopathy using ensemble machine learning algorithms. Front Surg 2022; 9: 1010420.
[35] D'Angelo GM, Rao D and Gu CC. Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies. BMC Proc 2009; 3 Suppl 7: S62.

Table S1. The table displays non-zero coefficients between various clinical semantic features and pathological types as well as WHO stages

Features	Pathology Types			WHO Stages			
	Low Risk	High Risk	Cancer	Stage I	Stage II	Stage III	Stage IV
Sex	NA	NA	NA	NA	NA	NA	NA
Age	NA	NA	NA	-0.035	NA	NA	NA
Chest Distress	NA	NA	NA	NA	0.019	NA	-0.125
Chest Pain	NA	NA	NA	0.698	NA	-0.175	NA
Cough	NA	NA	NA	NA	1.646	NA	NA
Myasthenia Gravis	NA	-0.375	NA	0.126	-0.126	0.166	-0.951
Calcification	NA	NA	NA	0.378	NA	NA	NA
Bleeding and Necrosis	NA	NA	NA	NA	0.082	NA	NA
Cystic Degeneration	NA	NA	NA	NA	NA	NA	NA
Indistinct Boundary	NA	NA	NA	1.001	0.037	-0.329	-0.037
Proximal Pulmonary Change	NA	NA	NA	0.995	NA	NA	NA
Mediastinal Lymph Node Enlargement	NA	NA	NA	0.509	0.266	-1.302	-0.267
Blood Vessel Invasion	NA	NA	NA	NA	NA	NA	-1.066
Effusion	NA	NA	NA	-0.582	-0.207	1.207	-1.471
Heterogeneous Strengthening	NA	NA	NA	NA	NA	NA	-0.125
Tumor Length	NA	NA	NA	0.625	NA	NA	0.171
Shape	NA	NA	NA	0.501	NA	NA	-0.832
Leukocyte Count	NA	NA	NA	NA	NA	NA	0.162
LDH	NA	NA	NA	NA	NA	NA	0.049

The leftmost column lists all clinical semantic features. The middle section lists the correlation coefficients with pathological types (low risk, high risk, and cancer). The rightmost section lists the correlation coefficients with WHO stages (Stage I to Stage IV). NA indicates that the data is not available or there is no significant correlation. Note: The table does not include the constant term.

Table S2. The table displays the non-zero coefficients between the 12 principal components obtained from PCA of different radiomics features and the pathological types as well as the WHO stages

Features	Pathology Types			WHO Stages			
	Low Risk	High Risk	Cancer	Stage I	Stage II	Stage III	Stage IV
PC1	NA	NA	-0.156	0.479	NA	NA	-0.785
PC2	NA	NA	NA	NA	NA	NA	NA
PC3	NA	NA	0.166	NA	NA	NA	NA
PC4	NA	NA	NA	-0.291	NA	0.078	NA
PC5	-0.06	NA	0.111	-0.166	NA	NA	NA
PC6	NA	NA	NA	NA	NA	NA	NA
PC7	NA	NA	-0.15	NA	NA	0.025	NA
PC8	-0.01	NA	NA	-0.141	NA	NA	NA
PC9	NA	NA	NA	NA	0.034	NA	NA
PC10	-0.073	NA	NA	NA	NA	NA	NA
PC11	-0.264	NA	NA	-0.038	NA	0.04	NA
PC12	NA	NA	NA	NA	NA	NA	NA

The leftmost column of the table lists all radiomics principal components. The middle section lists the correlation coefficients with pathological types (low risk, high risk, and cancer). The rightmost section lists the correlation coefficients with WHO stages (Stage I to Stage IV). NA indicates that the data is not available or there is no significant correlation. Note: The table does not include the constant term.

Table S3. The table displays the non-zero coefficients between the 21 principal components obtained from radiomics clinical semantic features via PCA and the pathological types as well as WHO stages

Features	Pathology Types			WHO Stages			
	Low Risk	High Risk	Cancer	Stage I	Stage II	Stage III	Stage IV
PC1	0.067	NA	-0.011	1.058	0.128	-0.128	-0.78
PC2	NA	-0.007	0.023	NA	NA	NA	0.11
PC3	NA	NA	NA	-0.256	NA	0.162	NA
PC4	NA	0.113	NA	NA	0.046	-0.321	NA
PC5	NA	NA	NA	0.382	NA	NA	-0.026
PC6	NA	NA	NA	NA	NA	NA	NA
PC7	NA	NA	NA	NA	0.024	NA	NA
PC8	NA	0.082	NA	NA	NA	NA	NA
PC9	-0.055	NA	NA	-0.003	NA	0.211	NA
PC10	NA	NA	NA	-0.239	NA	NA	0.044
PC11	NA	-0.164	NA	0.206	NA	NA	NA
PC12	NA	NA	NA	NA	NA	NA	NA
PC13	NA	-0.026	NA	NA	NA	NA	NA
PC14	NA	NA	NA	NA	NA	NA	NA
PC15	NA	NA	NA	NA	NA	NA	NA
PC16	NA	NA	0.018	NA	NA	NA	NA
PC17	0.068	NA	NA	NA	NA	NA	NA
PC18	NA	NA	0.011	-0.037	NA	0.016	NA
PC19	NA	NA	NA	0.094	NA	NA	NA
PC20	0.188	NA	NA	NA	NA	NA	NA
PC21	NA	NA	NA	-0.019	0.032	NA	NA

The leftmost column of the table lists all radiomics principal components. The middle section lists the correlation coefficients with pathological types (low risk, high risk, and cancer). The rightmost section lists the correlation coefficients with WHO stages (Stage I to Stage IV). NA indicates that the data is not available or there is no significant correlation. Note: The table does not include the constant term.