Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Intratumoral and Peritumoral CT Radiomics-Clinical Feature Fusion Model for Differentiating Pulmonary Tuberculosis and Pulmonary Cryptococcosis Nodules

Chun-Yan Zhao
The Fourth People's Hospital of Nanning
Chang Song
The Fourth People's Hospital of Nanning
Hua-Wei He
The Fourth People's Hospital of Nanning
Shu-Lin Song
The Fourth People's Hospital of Nanning
Xiao-Shi Lin
Guangxi Medical University
Zhen-Tao Huang
Guangxi Medical University
Xue-Wen Huang
Guangxi Medical University
Zhou-Hua Xie
The Fourth People's Hospital of Nanning
Qing-Dong Zhu
zhuqingdong2003@163.com

The Fourth People's Hospital of Nanning

Research Article

Keywords: Pulmonary tuberculosis, Pulmonary cryptococcosis, CT radiomics, Peritumoral region, Fusion model

Posted Date: May 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9406722/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective

To investigate the clinical value of a fusion model combining intratumoral and peritumoral CT radiomics features with clinical features for differentiating pulmonary tuberculosis (PTB) from pulmonary cryptococcosis (PC) nodules, and to develop a predictive tool with strong robustness and generalizability.

Methods

A total of 214 patients with PTB or PC were consecutively enrolled and randomly divided into a training set (n = 149) and a test set (n = 65) in a 7:3 ratio. Radiomics features were systematically extracted from the intratumoral region and from peritumoral regions extending 1 mm, 3 mm, and 5 mm beyond the lesion boundary on CT images. Least absolute shrinkage and selection operator (LASSO) regression was used for optimal feature selection, and the selected features were incorporated into various machine learning algorithms, including Support Vector Machine (SVM), to construct radiomics models. Meanwhile, a clinical model was established using independent clinical risk factors. Clinical and radiomics features were then integrated to construct model. Model performance was comprehensively evaluated using Receiver Operating Characteristic (ROC) curves, calibration curves, and Decision Curve Analysis (DCA).

Results

The clinical model alone showed limited predictive performance (AUC = 0.693 in the test set) and evidence of overfitting. Among the radiomics models, the model based on 3 mm peritumoral features demonstrated superior performance (AUC = 0.911 in the test set). The combined model, integrating clinical variables with intratumoral and 3 mm peritumoral radiomics features, exhibited the highest diagnostic performance, with AUCs of 0.979 and 0.944 in the training and test sets, respectively, and a sensitivity of 0.933 and a specificity of 0.886. Calibration curves indicated that good accuracy between predicted and observed outcomes, and DeLong’s test confirmed that the combined model outperformed the individual models in most comparisons.

Conclusion

A combined model integrating intratumoral and peritumoral radiomics features together with clinical variables can compensate for the limitations of the clinical model alone. It significantly improves diagnostic performance and generalizability in differentiating PTB from PC, and may serve as a precise and robust non-invasive tool to support clinical decision-making.

1 Introduction

Pulmonary nodules are a common radiological finding in clinical practice, and have a wide range of etiologies, from benign granulomatous lesions to malignant tumors [1]. Among benign causes, infectious nodules represent an important category, with pulmonary tuberculosis (PTB) and pulmonary cryptococcosis (PC) being of particular clinical concern [2, 3]. PTB, a chronic infectious disease caused by *Mycobacterium tuberculosis* (MTB), remains a major global public health challenge. According to the latest report from the World Health Organization (WHO), there were 10.7 million new tuberculosis (TB) cases worldwide in 2024, with 1.23 million deaths attributed to the disease [4]. PC is a fungal infection caused by opportunistic pathogens such as *Cryptococcus neoformans* or *Cryptococcus gattii*. In recent years, its incidence has increased in immunocompetent individuals, accompanied by growing clinical recognition of the disease improved [3, 5]. Despite marked differences in etiology, treatment, and prognosis between PTB and PC-PTB requires standardized, long-term anti-tuberculosis therapy, whereas PC is primarily treated with antifungal agents, early differentiation remains challenging, particularly in patients with atypical clinical presentations, due to clinical manifestations and routine laboratory findings often overlap. At present, differentiation between these two diseases relies on a combination of imaging, laboratory testing, and etiological evaluation. Traditional computed tomography (CT) imaging remains the cornerstone of initial assessment. PTB nodules often show heterogeneous attenuation, irregular or coarse margins, and may be associated with satellite lesions and a tree-in-bud pattern [6]. In contrast, PC commonly presents as solitary or multiple relatively well-defined nodules or mass-like opacities. However, these morphological features lack specificity, and the risk of misdiagnosis rate increases substantially in atypical cases or when the imaging features of the two conditions overlap [7]. Etiological examination remains the diagnostic gold standard, including sputum smear or culture for acid-fast bacilli, cryptococcal capsular antigen detection, and tissue culture. However, these methods have notable limitations, including low sensitivity in some settings (e.g., sputum smear), prolonged turnaround times (e.g., culture), and the need for invasive procedures, such as bronchoscopy or percutaneous lung puncture biopsy, to obtain qualified specimens. In addition, invasive procedures may lead to complications such as hemorrhage and pneumothorax, and some patients are unable or unwilling to undergo them. Therefore, there is an urgent clinical need for a noninvasive adjunctive tool that can address the limitations of current approaches and improve diagnostic efficiency.

In recent years, radiomics has emerged as a promising medical image analysis approach, offering a novel perspective and powerful technical support for addressing this challenge. The core concept of the radiomics is the high-throughput extraction of large numbers of quantitative features from standard medical images that are invisible or difficult to detect visually, including first-order statistics, shape-based features, texture features, and higher-order filter features. By transforming images into mineable high-dimensional data, radiomics enables quantitative characterization of lesion heterogeneity, spatial complexity, and underlying biological behavior, thereby facilitating the development of objective models for disease diagnosis, prognosis prediction, and treatment response evaluation [8]. In pulmonary diseases, radiomics has been widely applied to the differentiation of benign and malignant nodules, prediction of lung cancer histological subtype and genotype, and assessment of therapeutic response,

demonstrating considerable potential [8–12]. However, most previous radiomics studies on pulmonary nodule have focused primarily on the lesion itself, namely the intratumoral region. Although this strategy has yielded encouraging results, it overlooks potentially valuable information at the interface between the lesion and the surrounding lung parenchyma, namely the peritumoral region.

Accumulating evidence suggests that the peritumoral microenvironment plays a critical role in disease initiation, progression, and biological behavior [13, 14]. In infectious diseases, the peritumoral region may reflect host immune response, inflammatory cell infiltration, and pathophysiological changes such as local edema or organization. Therefore, incorporating peritumoral radiomics features in addition to intratumoral features may provide a more comprehensive characterization of lesion biology and further improve model discrimination. Meanwhile, clinical variables, including age, gender, symptoms, comorbidities, immune status, and laboratory findings, offer complementary information from different dimensions. Effective integration of clinical and radiomics features to build multimodal diagnostic models represents an important direction in precision medicine and may help overcome the limitations of single-source diagnostic information [15, 16].

Against this background, this study aimed to construct and validate a combined model integrating intratumoral and peritumoral CT radiomics features with clinical variables for differentiating PTB from PC. This model is intended to provide clinicians with an objective, quantitative, and efficient adjunctive diagnostic tool, and to support more precise diagnosis and management of infectious pulmonary nodules.

2 Materials and Methods

2.1 Study Participants

This study retrospectively enrolled 214 patients with pulmonary nodules including 116 with PC and 98 with PTB, who were treated at the Fourth People's Hospital of Nanning between January 2016 and October 2025. The inclusion criteria were as follows: (1) pre-treatment CT of the chest was available, with complete imaging data suitable for radiomics analysis; (2) a definitive diagnosis of PC or PTB was established via histopathology, etiological culture, or diagnostic criteria consistent with current clinical guidelines; (3) due to tuberculous nodules typically show self-limited growt, the maximum diameter of the included pulmonary nodules was < 8 cm, and lesion margins were clearly delineated; and (4) complete baseline clinical data were available, including demographic characteristic, immune status, key laboratory findings, and pre-treatment imaging records. The exclusion criteria were as follow: (1) concurrent active pulmonary infection other than PTB or PC, pulmonary malignancy, or a history of ipsilateral lung surgery; (2) severe CT artifacts affecting t evaluation of the lesion or peritumoral regions; (3) missing key clinical information or incomplete imaging data; and (4) receipt of interventional treatment or systemic anti-infective therapy targeting the lesion before CT examination. The diagnosis of PTB was based on the WS 288–2017 Diagnostic Standard for Pulmonary Tuberculosis, including a positive sputum smear for acid-fast bacilli, positive sputum culture for MTB, positive molecular testing

(e.g., Xpert MTB/RIF), or histopathological evidence of tuberculous lesions. The diagnosis of PC was based on the Global Guideline for the Diagnosis and Management of Cryptococcosis: An ECMM-ISHAM-ASM Cooperated Initiative, jointly issued by the European Confederation of Medical Mycology (ECMM), the International Society for Human and Animal Mycology (ISHAM), and the American Society for Microbiology (ASM). Diagnostic criteria included positive culture of *Cryptococcus neoformans* from respiratory specimens, positive cryptococcal antigen in serum or cerebrospinal fluid, or histopathological identification of Cryptococcus spp. [17]. This study was approved by the Hospital Ethics Committee (Approval No.: [2025] 52) and conducted in accordance with the Declaration of Helsinki. Given the retrospective design and the use of anonymized data, the requirement for informed consent was waived.

2.2 Clinical Data Collection

Relevant clinical data were collected from the hospital electronic medical record system. The dataset was randomly divided into a training set and a test set at a ratio of 7:3. Continuous variables included age, body mass index (BMI), CD3, CD4, CD8, CD4/CD8 ratio, white blood cell count (WBC), red blood cell count (RBC), hemoglobin level, platelet count, neutrophil count, lymphocyte count, monocyte count, eosinophil count, basophil count, and C-reactive protein (CRP). These variables ere expressed as mean ± standard deviation. Categorical variables included gender, hypertension, heart disease, diabetes, smoking history, alcohol consumption history, and fever.

2.3 Image Acquisition

CT images were acquired using a GE LightSpeed 64-slice spiral CT scanner. All patients were scanned in the supine position and instructed to remain still and avoid speaking during image acquisition to minimize motion artifacts. Scanning was performed using a standard non-contrast protocol with a slice thickness of 5 mm. The raw data were then reconstructed to generate high-resolution CT images for subsequent analysis.

2.4 Image Processing and Region of Interest (ROI) Segmentation

CT images in Digital Imaging and Communications in Medicine (DICOM) format were exported from the imaging workstation and processed using ITK-SNAP software. To improve consistency across scans, all images underwent standardization and normalization. Specifically, voxel spacing was standardized to 1 mm × 1 mm × 1 mm, and the window width and window level were set to 1500 and − 500 Hounsfield units (HU), respectively. Regions of interest (ROIs) in the training set were manually delineated on ITK-SNAP by two radiologists. Discrepancies between the two readers were resolved by a senior radiologist with more than 20 years of experience. This manual segmentation process provided the basis for subsequent model development. To assess the contribution of the peritumoral region, each ROI was isotropically expanded by 1 mm, 3 mm, and 5 mm beyond the lesion boundary. ROIs from all slices were merged to generate a three-dimensional volume of interest (VOI).

2.5 Radiomics Feature Extraction and Selection

Hand-crafted radiomics features were categorized into three groups: geometric, intensity, and texture features. Geometric features described lesion shape and size, intensity features quantified voxel intensity distribution, and texture features characterized spatial heterogeneity using matrices such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM). In-depth radiomics analysis was performed on the VOI and its corresponding subregions. The K-Nearest Neighbors (KNN) method was employed within the clustering framework to handle unclustered regions and ensure consistent labeling of habitat subregions. Feature extraction was performed using Pyradiomics (version 3.0.1).

To ensure feature robustness, the Intraclass Correlation Coefficient (ICC) was first calculated based on segmentations performed by two radiologists, and only features with an ICC > 0.8 were retained. Pearson's correlation analysis was then used to remove highly correlated features, with a correlation coefficient threshold of 0.9. Subsequently, the Minimum Redundancy Maximum Relevance (mRMR) algorithm was further applied to reduce redundancy while preserving informative features. Final feature selection was performed using Least Absolute Shrinkage and Selection Operator (LASSO) regression. The optimal regularization parameter λ was determined via 10-fold cross-validation, and only the most predictive features were retained for model construction.

2.6 Model Construction

This study analyzed intratumoral and peritumoral features to improve diagnostic accuracy and identify imaging biomarkers associated with lesion behavior. By combining advanced image processing with machine learning, we aimed to capture subtle morphological and functional differences between PTB and PC. Peritumoral analysis was further used to explore the local microenvironment surrounding the lesion. Machine learning models were developed based on intratumoral features, peritumoral features, and fused radiomics features. Features were selected using LASSO, and hyperparameters were optimized by grid search with 5-fold cross-validation. For model evaluation, 1000 repetitions of bootstrap 5-fold cross-validation were conducted in the training set, and the intra-, peri-, and fusion models were assessed using the median Area Under the Receiver Operating Characteristic Curve (AUC). For the intra radiomics signature, selected features were incorporated into logistic regression (LR), support vector machine (SVM), and random forest (RF) algorithms to construct prediction models. For the peritumoral radiomics signature, the same feature selection strategy and machine learning algorithms were applied using only peritumoral features. For the fusion radiomics signature, selected intratumoral and peritumoral features were combined, and the same modeling pipeline was used to ensure consistency. For the clinical signature, candidate clinical variables were s screened by univariate and multivariate analysis and then entered into multivariate analysis,with $P < 0.05$ considered statistically significant. For the combined model, significant clinical predictors were integrated with the best-performing radiomics signatures identified from the peritumoral and fusion models. The final model was established using univariate analysis followed by stepwise multivariate analysis to maximize predictive performance.

2.7 Statistical Analysis

The normality of continuous clinical variables was assessed using the Shapiro-Wilk test. Depending on data distribution, between-group comparisons were performed using either the independent-samples t test or the Mann-Whitney U test. Categorical variables were compared using the chi-square (χ^2) tests or Fisher's exact test, as appropriate. Baseline characteristics were compared between the training and test sets, and $P > 0.05$ was considered to indicate no significant difference, suggesting an acceptable random spli. All analyses were performed using Python 3.7.12. Statistical analyses were performed using Statsmodels version 0.13.2. Radiomics feature were extracted using PyRadiomics version 3.0.1. Machine learning algorithms, including SVM, were implemented using Scikit-learn version 1.0.2. Model discrimination was evaluated using AUC, sensitivity, specificity, and accuracy. Differences in AUC between models were compared using the DeLong test. Calibration performance was assessed using calibration curves and the Hosmer-Lemeshow (H-L) goodness-of-fit test, $P > 0.05$ indicated good agreement between predicted and observed outcomes. DCA was performed to evaluate clinical utility by calculating the standardized net benefit across a range of threshold probabilities.

3 Results

3.1 Screening of Clinical Characteristics and Model Construction in Patients

Analysis of the baseline characteristics showed that the mean age of the enrolled patients was approximately 52 years, and 73.36% were male. No significant differences were observed between the training and test sets in age, BMI, immune cell subsets (CD3, CD4, CD8, and their ratios), complete blood count parameters (WBC, RBC, platelets, etc.), or comorbidities (including syphilis, hepatitis B, hypertension, diabetes, etc.) ($P > 0.05$). Only lymphocyte count differed significantly between the two groups ($P = 0.036$), suggesting that the dataset split was generally balanced and representative (Supplementary Table 1). Univariate analysis was performed for all clinical variables, and the odds ratio (OR) and corresponding P value were calculated for each predictor. Variables significantly associated with the outcome ($p < 0.05$) were retained and entered into multivariate analysis for construction of the clinical model. This preliminary screening step ensured that only variables with a significant individual association with the outcome were included in subsequent multivariate modeling. The results identified gender, CD4/CD8 ratio, and hemoglobin level as independent predictors (Supplementary Table 2). Based on these clinical variables, five machine learning models were constructed (Table 1). Although all models showed acceptable performance in the training set (AUC > 0.72), their performance declined significantly in the test set, indicating overfittin. Overall, the predictive performance of the clinical models remained limited and requires further improvement.

Table 1
Metrics of different model in clinical signature.

Model name	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
SVM	0.724	0.640 ~ 0.808	0.485	0.827	0.702	0.657	train
RandomForest	0.785	0.713 ~ 0.857	0.721	0.741	0.700	0.759	train
ExtraTrees	0.757	0.680 ~ 0.834	0.544	0.802	0.698	0.677	train
XGBoost	0.764	0.692 ~ 0.836	0.662	0.778	0.714	0.733	train
LightGBM	0.801	0.733 ~ 0.869	0.706	0.765	0.716	0.756	train
SVM	0.648	0.513 ~ 0.783	0.367	0.857	0.687	0.612	test
RandomForest	0.583	0.442 ~ 0.723	0.433	0.686	0.542	0.585	test
ExtraTrees	0.619	0.483 ~ 0.755	0.333	0.771	0.556	0.574	test
XGBoost	0.654	0.525 ~ 0.784	0.400	0.714	0.545	0.581	test
LightGBM	0.693	0.565 ~ 0.821	0.500	0.714	0.600	0.625	test

3.2 Radiomics Feature Extraction and Model Construction

The enrolled patients were divided into the PTB group (n = 98) and the PC group (n = 116) according to the diagnostic criteria and were randomly divided into a training set (n = 149) and a testing set (n = 65) at a ratio of 7:3. A total of 1,834 optimized radiomics features were extracted from each ROI. To identify the most discriminative features while controlling model complexity, LASSO logistic regression was employed. The optimal penalty parameter λ was determined via 10-fold cross-validation using the minimum mean squared error as the selection criterion. All hand-crafted features underwent dimensionality reduction and selection via 10-fold cross-validation and LASSO regression to identify the optimal radiomics features (Fig. 1A-L). To evaluat the predictive value of radiomics features derived from different regions, this study extracted features from the intra-tumoral region and peri-tumoral regions extending 1 mm, 3 mm, and 5 mm beyond the lesion boundary. Multiple machine learning algorithms were then used to construct prediction models, and their performance was assessed in both the training and testing sets. All feature-based models exhibited good discriminative ability in the training set. Notably, the SVM model based on 3-mm peritumoral features yielded the best performance, with an AUC of 0.958 in the training set and 0.911 in the test set, the corresponding sensitivity and specificity were 0.882 and 0.914 in the training set, and 0.767 and 0.914 in the test set, respectively. This findings indicate excellent discriminative performance and generalizability. In contrast, the model based on intra-tumoral features achieved a maximum testing set AUC of 0.922 using SVM. Furthermore, some models showed evidence of overfitting. For example, the RF model based on 1-mm peritumoral features achieved an AUC of 0.938 in the training set but only 0.776 in the test set. Overall, the combination of 3-mm peritumoral features and the SVM classifier demonstrated the best balance of predictive

performance and stability, and was therefore selected as the optimal radiomics model (Supplementary Table 3–6). ROC curves for the different models in the training and test sets are presented in Fig. 2A-H.

3.3 Construction of Clinical-Intra-Peritumoral Model

The clinical model alone showed the poorest performance, with a testing set AUC of only 0.693. All radiomics models exhibited good discriminative ability in the training set, among which the model based on 3-mm peritumoral features performed particularly well, with an AUC of 0.958 in the training set and 0.911 in the test set. Notably, the combined model integrating clinical variables, intratumoral features and 3-mm peritumoral features achieved the best predictive performance, with an AUC of 0.979 in the training set and 0.944 in the test set (Fig. 3A-B). It also showed high sensitivity (0.933) and high specificity (0.886) (Table 2). These findings suggest that integrating complementary information from intratumoral imaging features, peritumoral radiomics features, and clinical variables can generate a more robust and generalizable predictive model. The combined model not only demonstrated exceptional discriminative ability but also showed favorable stability across datasets.

Table 2
Prediction performance of habitat based rad signatures.

Signature	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
Clinical	0.801	0.7328 ~ 0.8694	0.721	0.753	0.710	0.762	train
Radiomics	0.932	0.8948 ~ 0.9694	0.750	0.963	0.944	0.821	train
Peri1mm	0.915	0.8705 ~ 0.9594	0.912	0.753	0.756	0.910	train
Peri3mm	0.958	0.9306 ~ 0.9852	0.882	0.914	0.896	0.902	train
Peri5mm	0.863	0.8038 ~ 0.9224	0.662	0.926	0.882	0.765	train
Combined	0.979	0.9626 ~ 0.9960	0.971	0.889	0.880	0.973	train
Clinical	0.693	0.5651 ~ 0.8206	0.767	0.543	0.590	0.731	test
Radiomics	0.922	0.8540 ~ 0.9898	0.867	0.914	0.897	0.889	test
Peri1mm	0.827	0.7221 ~ 0.9322	0.800	0.771	0.750	0.818	test
Peri3mm	0.911	0.8435 ~ 0.9794	0.767	0.914	0.885	0.821	test
Peri5mm	0.856	0.7626 ~ 0.9498	0.700	0.943	0.913	0.786	test
Combined	0.944	0.8881 ~ 0.9995	0.933	0.886	0.875	0.939	test

3.4 Superiority Analysis of the Combined Model

The H-L test indicated good calibration of the combined model across all study populations. The H-L test results for the training and testing sets were 0.597 and 0.619, respectively, with no significant deviation observed (all p-values > 0.05). These results reveal good agreement between predicted probabilities and observed outcomes, supporting the reliability of the model for individual risk estimation. Calibration

curves for the training and testing sets further confirmed the this agreement visually (Fig. 4A–B). The Delong test was used to compare AUC values among models. In the training set and internal validation set, the combined model performed significantly better than most of the other models ($p < 0.05$), confirming a statistically significant improvement in predictive performance (Fig. 4C–D). However, although combined model achieved a higher AUC in the test set, the difference did not reach statistical significance, suggesting that the performance gain may be more limited in independent validation.

Figure 5A–D illustrates the Integrated Discrimination Improvement (IDI) and Net Reclassification Improvement (NRI) of the radiomics-based models relative to the clinical model. In both the training and testing sets, models based on intratumoral or peritumoral features alone improved predictive performance to some extent, with the 3-mm peritumoral model showing greater improvement than the other single-region models. Notably, the combined model integrating clinical variables with multi-regional radiomics features achieved the highest IDI and NRI values, indicating superior risk reclassification and the greatest incremental predictive value over single-modality models. This result was verified in the test set, further confirming the robustness of the combined model.

3.5 Clinical Value Analysis

DCA showed that, in both the training and testing sets, the combined model integrating clinical and multi-regional radiomics features yielded a higher net benefit than either the clinical model alone or the radiomics models alone across a wide range of threshold probabilities. These findings indicate superior clinical utility and greater value for decision-making (Fig. 6A–B). Based on these results, independent predictors were further integrated to construct a nomogram, including the CD4/CD8 ratio, hemoglobin level, gender, peritumoral 3 mm radiomics features, and intratumoral radiomics features. By assigning a weighted score of each factor, the nomogram enables intuitive estimation of the total score and corresponding individual risk. It may therefore serve as a visual and individualized tool to assist clinical decision-making (Fig. 6C).

4. Discussion

In this study, we constructed a combined model based on intratumoral and peritumoral CT radiomics features together with clinical variables to address the clinical challenge of differentiating PTB from PC. The results showed that the combined model, integrating clinical factors, intratumoral radiomics features, and 3-mm peritumoral features, achieved the best predictive performance, with a testing set AUC of 0.944. This performance was superior to that of the clinical model alone (AUC = 0.693) and the individual radiomics models. These findings suggest that the proposed model may serve as a novel noninvasive adjunctive tool for the differential diagnosis of infectious pulmonary nodules and has potential clinical applicability.

We first constructed a clinical model based on independent clinical predictors, including gender, CD4/CD8 ratio, and hemoglobin level. Among these variables, the CD4/CD8 ratio, as an important indicator of immune status, deserves particular attention. CD4 + T lymphocytes, the major helper T cell

subset, play a crucial role in anti-tuberculosis immunity by activatin macrophages through cytokines such as interferon-γ, thereby helping control intracellular MTB replication. In contrast, CD8 + T cells primarily contribute to host defense against MTB by exerting cytotoxic effects on infected cells [18]. Previous studies have confirmed that patients with active PTB often exhibit reduced peripheral CD4 + T cell counts and an altered CD4/CD8 ratio, possibly as a result of MTB-induced immunosuppression or redistribution T cells to the site of infection [19]. In this study, the CD4/CD8 ratio was negatively correlated with the risk of PC (OR = 0.826), indicating that a lower ratio was associated with a relatively higher likelihood of PTB than PC. This finding is consistent with the known immunopathological differences between the two diseases and suggests that the CD4/CD8 ratio may serve as a surrogate marker of pathogen-specific host immune responses. Hemoglobin level was also negatively associated with the probability of PC (multivariate OR = 0.996, 95% CI: 0.992–0.999, P < 0.05), showing that lower hemoglobin levels were more suggestive of PTB. Anemia is a common manifestation of chronic infectious diseases and may result from inflammatory cytokine-mediated suppression of erythropoiesis, dysregulated iron metabolism, and chronic systemic consumption. As a typical chronic wasting infectious disease, PTB is frequently accompanied by anemia due to prolonged fever, anorexia, malnutrition, and inflammatory status [20, 21]. Although PC may also present with constitutional e symptoms, these manifestations are generally milder, particularly in immunocompetent hosts. Accordingly, may partly reflect the degree of chronic systemic consumption, with lower levels favoring a diagnosis of PTB, which is in line with clinical experience. Male gender was positively associated with PC (multivariate OR = 1.322, 95% CI: 1.133–1.542, P < 0.05), indicating that male patients had a higher probability of PC than female patients in this cohort. This finding is closely related to the epidemiological characteristics of the two diseases. Although tuberculosis is generally more common in males, the male predominance was even more pronounced among patients with PC in this study. Previous epidemiological studies of cryptococcosis have also identified male as an independent risk factor, possibly because of sex hormone-related immune regulation, differences in occupational exposure, and higher HIV prevalence among males [22]. The identification of male gender as an independent predictor of PC (OR > 1) in this study is consistent with previous epidemiological observations and suggests that PC should be considered more strongly in male patients during clinical differential diagnosis. Notably, smoking and alcohol consumption were significant in univariate analysis but lost significance in the multivariate model. This may reflect collinearity with ender, since smoking and alcohol consumption are more prevalent in men, or may indicate that these behaviors affect disease risk indirectly through intermediate variables such as immune status or local pulmonary defense rather than acting as direct independent predictors. Similarly, although CD4 count was significant in univariate analysis, it was not retained in the multivariate model, possibly because of its close association with the CD4/CD8 ratio. As a composite index, the CD4/CD8 ratio may better capture the balance of T-cell subsets than the absolute CD4 count alone.

This study also systematically compared the predictive performance of radiomics features derived from the intratumoral region and from peritumoral regions of different widths (1 mm, 3 mm, and 5 mm). The model based on the 3-mm peritumoral region model performed best, achieving a testing set AUC of

0.911. While the intratumoral model achieved a slightly higheAUC (0.922 vs. 0.911), the 3-mm peritumoral model showed better performance in the training set and more stable generalization overall. By contrast, the 1-mm peritumoral model exhibited clear overfitting (0.938 vs. 0.776), while the 5-mm peritumoral model showed declined performance (AUC = 0.856). These results suggest the existence of an optimal peritumoral range, a range that is too narrow may capture insufficient information and be more susceptible to noise, whereas a region that is too wide may incorporate irrelevant background lung tissue.

From a pathophysiological perspective, the superiority of the 3-mm peritumoral region may be attributable to the unique biological information contained within this zone. Infectious nodule are not isolated pathological entities, rather, their interface with the surrounding lung tissue reflects dynamic processes such as host immune response, inflammation, and tissue remodeling. In PTB, typical pathological features include caseous necrosis, granuloma formation, and fibrous encapsulation, and the peritumoral region often contains satellite lesions, tree-in-bud changes, and inflammatory infiltration. In contrast, PC ends to present as more localized gelatinousor granulomatous lesions, with less extensive surrounding tissue reaction. A 3-mm peritumoral region may therefore represent a balance between capturing the core reactive zone around the lesion, such as inflammatory infiltration and early fibrosis, and minimizing interference from unrelated lung parenchyma. This observation is in line with previous radiomics studies in oncology, which have shown that the 2–5 mm peritumoral region may provide unique value for predicting tumor invasion and lymph node metastasis. Our findings extend this concept to the differential diagnosis of infectious nodules and may provide a useful methodological reference for future studies in this field. An important question is whether the superiority of the 3-mm peritumoral region is universal or specific to the differentiation of PTB and PC. Theoretically, peritumoral pathological patterns vary across different disease types, including malignant tumors, inflammatory lesions, and benign nodules, and the optimal peritumoral range may therefore also differ. Therefore, future research should not assume that 3 mm is universally appropriate, but should instead determine the optimal range using data-driven approaches tailored to the specific clinical scenario. In this regard, the main contribution of the present study is not the proposal of a fixed universal parameter, but the demonstration of a methodological framework based on systematic comparison of multiple peritumoral ranges.

The final output of this study was a nomogram integrating CD4/CD8 ratio, hemoglobin, gender, 3-mm peritumoral radiomics features, and intratumoral radiomics features. By entering these variables, clinicians can estimate the total score and corresponding predicted risk in an intuitive manner. DCA further showed that this combined model provided greater net benefit than the clinical or radiomics models alone across a wide range of threshold probabilities, indicating good clinical utility. In practical terms, this tool may be particularly useful for patients with pulmonary nodules showing atypical CT findings, as it offers a noninvasive estimate of the probability of PTB versus PC and may assist clinicians in selecting subsequent diagnostic and therapeutic strategies, such as bronchoscopy, percutaneous lung biopsy, or empirical treatment. In this way, it may help reduce unnecessary invasive procedures and their associated complications while improving allocation of medical resources.

In summary, we constructed and validated a combined model integrating clinical factors with intratumoral, and 3-mm peritumoral CT radiomics features, which showed excellent predictive performance and favorable clinical utility for differentiating PTB from PC. The 3-mm peritumoral region appeared to be the optimal range for radiomics feature extraction in this setting, providing a useful reference for future research. Presented in the form of a nomogram, this model may offer clinicians a visual, individualized, and non-invasive adjunctive diagnostic tool to support more precise management of of infectious nodules. Several limitations should nevertheless be acknowledged. First, this was a single-center retrospective study, and although a training-test split was used, selection bias may still have been present. The generalizability of the model therefore requires further validation in larger multicenter cohorts. Second, the biological interpretability of the radiomics features remains limited. In particular, the pathological correlates of the 3-mm peritumoral region, such as inflammatory infiltration and fibrosis, were not confirmed histologically. Future studies combining imaging with pathological correlation are warranted to clarify the biological basis of these features.

Declarations

Author Contributions.

Chun-Yan Zhao: Formal analysis; resources; writing—original draft. **Chang Song:** Conceptualization; validation; writing—original draft. **Hua-Wei He:** Supervision; validation. **Shu-Lin Song:** Methodology; software. **Xiao-Shi Lin:** Formal analysis; methodology. **Zhen-Tao Huang:** Validation; writing—review & editing. **Xue-Wen Huang:** Conceptualization; methodology; software; supervision. **Zhou-Hua Xie and Qing Dong Zhu:** Funding acquisition; project administration; validation; writing—review & editing.

Funding.

This work was supported by the Guangxi Natural Science Foundation General Project (2023GXNSFAA026022) and Guangxi key research and development projects (Guike AB25069097).

Data Availability.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare no competing interests.

Ethical Approval.

The study protocol was approved by the Ethics Committee of The Fourth People's Hospital of Nanning (Ethical Approval Number: [2025] 52). The requirement for informed consent to study inclusion and the need for consent to participate were waived by the Ethics Committee of The Fourth People's Hospital of

Nanning because of the lack of study intervention in patient diagnosis and treatment and the retrospective nature of the study. This study is in accordance with the Declaration of Helsinki.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Acknowledgements

Not applicable.

References

1. Chen D, Yang L, Zhang W, Shen J, Van Schil PEY, Divisi D, et al. Prevalence and management of pulmonary nodules: a systematic review and meta-analysis. J. Thorac. Dis. 2024;16(7):4619-32. https://doi.org/10.21037/jtd-24-874.
2. Howard-Jones AR, Sparks R, Pham D, Halliday C, Beardsley J, Chen SC. Pulmonary Cryptococcosis. Journal of fungi (Basel, Switzerland) 2022;8(11). https://doi.org/10.3390/jof8111156.
3. Kahn FW, England DM, Jones JM. Solitary pulmonary nodule due to Cryptococcus neoformans and Mycobacterium tuberculosis. Am. J. Med. 1985;78(4):677-81. https://doi.org/10.1016/0002-9343(85)90413-9.
4. WORLD HEALTH ORGANIZATION. Global tuberculosis report 2025[EB/OL]. [2025-11-13]. https://www.who.int/publications/i/item/9789240116924.
5. Yamamura D, Xu J. Update on Pulmonary Cryptococcosis. Mycopathologia 2021;186(5):717-28. https://doi.org/10.1007/s11046-021-00575-9.
6. Heo JN, Choi YW, Jeon SC, Park CK. Pulmonary tuberculosis: another disease showing clusters of small nodules. AJR Am. J. Roentgenol. 2005;184(2):639-42. https://doi.org/10.2214/ajr.184.2.01840639.
7. Ogawa H, Urushibara T, Kasai H, Ikeda H, Shinozaki T. Pulmonary Cryptococcosis in a Nurse Initially Suspected of Having Pulmonary Tuberculosis. The American journal of case reports 2020;21:e919267. https://doi.org/10.12659/ajcr.919267.
8. Boubnovski Martell M, Linton-Reid K, Chen M, Aboagye EO. Radiomics for lung cancer diagnosis, management, and future prospects. Clin. Radiol. 2025;86:106926. https://doi.org/10.1016/j.crad.2025.106926.
9. 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(4):307-15. https://doi.org/10.1136/thorax-2023-220226.

10. Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. Clin. Radiol. 2020;75(1):13-9. https://doi.org/10.1016/j.crad.2019.04.017.
11. Walls GM, Osman SOS, Brown KH, Butterworth KT, Hanna GG, Hounsell AR, et al. Radiomics for Predicting Lung Cancer Outcomes Following Radiotherapy: A Systematic Review. Clin. Oncol. (R. Coll. Radiol.) 2022;34(3):e107-e22. https://doi.org/10.1016/j.clon.2021.10.006.
12. Chen M, Copley SJ, Viola P, Lu H, Aboagye EO. Radiomics and artificial intelligence for precision medicine in lung cancer treatment. Semin. Cancer Biol. 2023;93:97-113. https://doi.org/10.1016/j.semcancer.2023.05.004.
13. Duan L, Liu W, Li M, Guo L, Ren M, Dong X, et al. CT radiomics from intratumor and peritumor regions for predicting poorly differentiated invasive nonmucinous pulmonary adenocarcinoma. Sci. Rep. 2025;15(1):14434. https://doi.org/10.1038/s41598-025-99465-z.
14. Gong T, Gao Y, Li H, Wang J, Li Z, Yuan Q. Research progress in multimodal radiomics of rectal cancer tumors and peritumoral regions in MRI. Abdominal radiology (New York) 2025;50(12):5677-89. https://doi.org/10.1007/s00261-025-04965-1.
15. Zou XC, Yuan RM, Chao HC, Huang JB, Zeng T, Xu XD. Machine learning model integrating radiomics and clinical features for predicting postoperative bleeding after percutaneous nephrolithotomy. Eur. J. Med. Res. 2026;31(1):37. https://doi.org/10.1186/s40001-025-03700-2.
16. Zhao W, Hou M, Wang J, Song D, Niu Y. Interpretable machine learning model for predicting clinically significant prostate cancer: integrating intratumoral and peritumoral radiomics with clinical and metabolic features. BMC Med. Imaging 2024;24(1):353. https://doi.org/10.1186/s12880-024-01548-2.
17. Chang CC, Harrison TS, Bicanic TA, Chayakulkeeree M, Sorrell TC, Warris A, et al. Global guideline for the diagnosis and management of cryptococcosis: an initiative of the ECMM and ISHAM in cooperation with the ASM. Lancet Infect. Dis. 2024;24(8):e495-e512. https://doi.org/10.1016/s1473-3099(23)00731-4.
18. Woodworth JS, Behar SM. Mycobacterium tuberculosis-specific CD8+ T cells and their role in immunity. Crit. Rev. Immunol. 2006;26(4):317-52. https://doi.org/10.1615/critrevimmunol.v26.i4.30.
19. Yin Y, Qin J, Dai Y, Zeng F, Pei H, Wang J. The CD4+/CD8+ Ratio in Pulmonary Tuberculosis: Systematic and Meta-Analysis Article. Iran J. Public Health 2015;44(2):185-93.
20. Gelaw Y, Getaneh Z, Melku M. Anemia as a risk factor for tuberculosis: a systematic review and meta-analysis. Environ. Health Prev. Med. 2021;26(1):13. https://doi.org/10.1186/s12199-020-00931-z.
21. Barzegari S, Afshari M, Movahednia M, Moosazadeh M. Prevalence of anemia among patients with tuberculosis: A systematic review and meta-analysis. Indian J. Tuberc. 2019;66(2):299-307. https://doi.org/10.1016/j.ijtb.2019.04.002.
22. McClelland EE, Hobbs LM, Rivera J, Casadevall A, Potts WK, Smith JM, et al. The role of host gender in the pathogenesis of Cryptococcus neoformans infections. PLoS One 2013;8(5):e63632.

https://doi.org/10.1371/journal.pone.0063632.

Figures

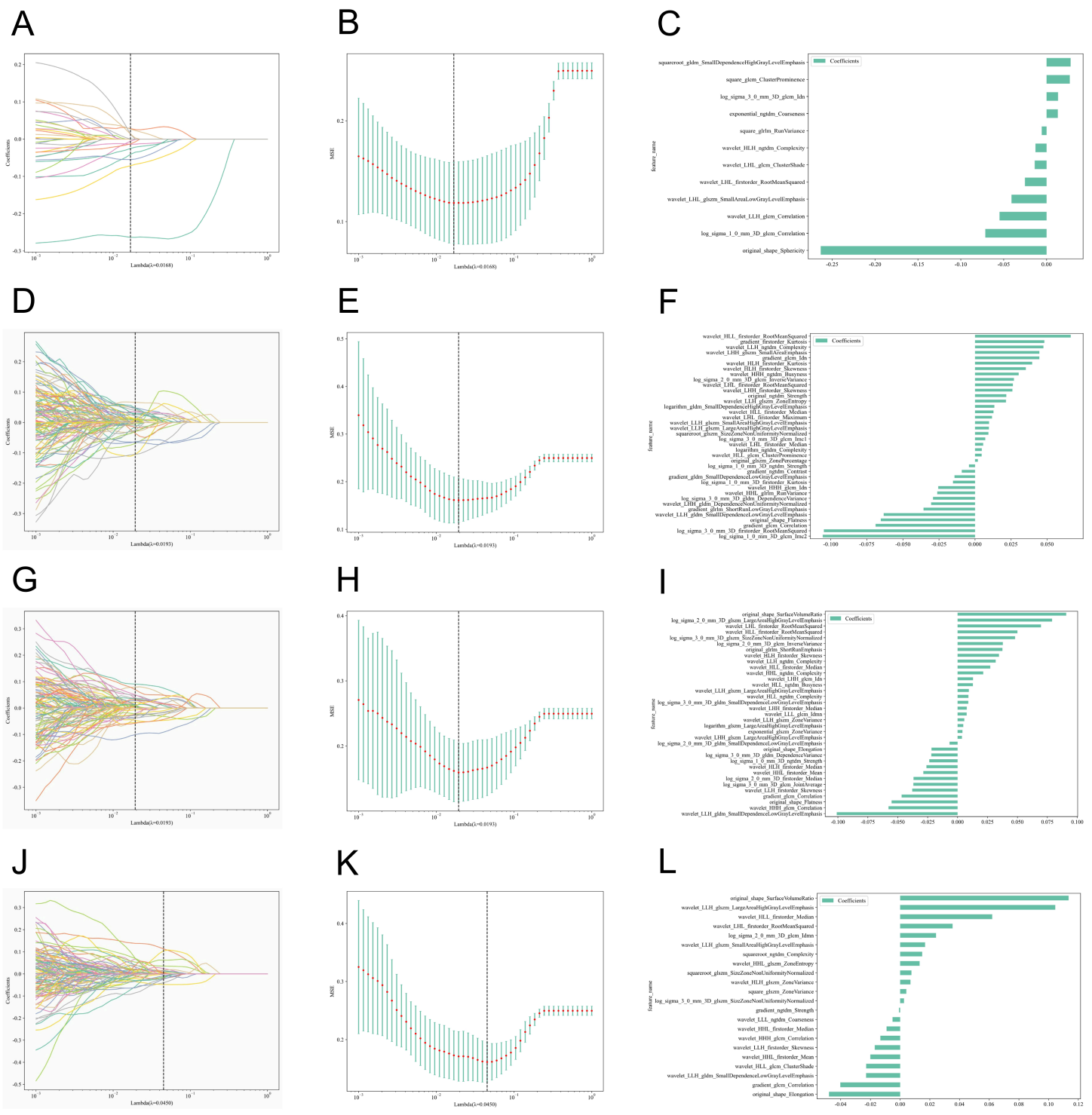


Figure 1

Workflow for extracting radiomics features (A. 10-fold cross-validation coefficient of tumor internal region features; B. Mean square error of 10-fold cross-validation of tumor internal region features; C.

Histogram of Rad score of tumor internal region features; D. 10-fold cross-validation coefficient of tumor periphery 1mm region features; E. Mean square error of 10-fold cross-validation of tumor periphery 1mm region features; F. Histogram of Rad score of tumor periphery 1mm region features; G. 10-fold cross-validation coefficient of tumor periphery 3mm region features; H. Mean square error of 10-fold cross-validation of tumor periphery 3mm region features; I. Histogram of Rad score of tumor periphery 13mm region features; J. 10-fold cross-validation coefficient of tumor periphery 5mm region features; K. Mean square error of 10-fold cross-validation of tumor periphery 5mm region features; L. Histogram of Rad score of tumor periphery 13mm region features)

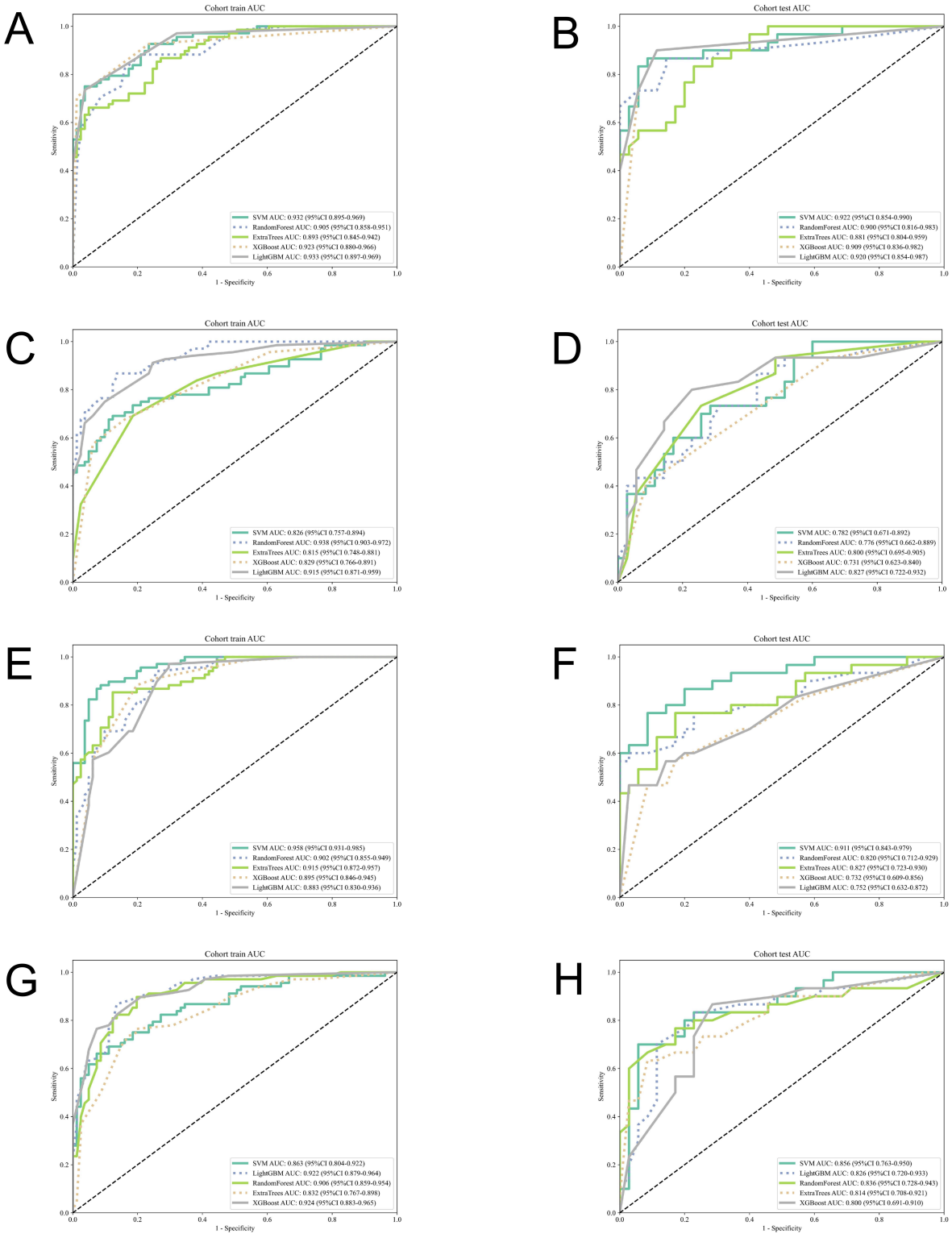


Figure 2

ROC comparison of intratumoral and peritumoral radiomics models. (A. Train cohort in intratumoral radiomics models; B. Test cohort in intratumoral radiomics models; C. Train cohort in peritumoral 1mm radiomics models; D. Test cohort in peritumoral 1mm radiomics models; E. Train cohort in peritumoral 3mm radiomics models; F. Test cohort in peritumoral 3mm radiomics models; G. Train cohort in peritumoral 5mm radiomics models; H. Test cohort in peritumoral 5mm radiomics models)

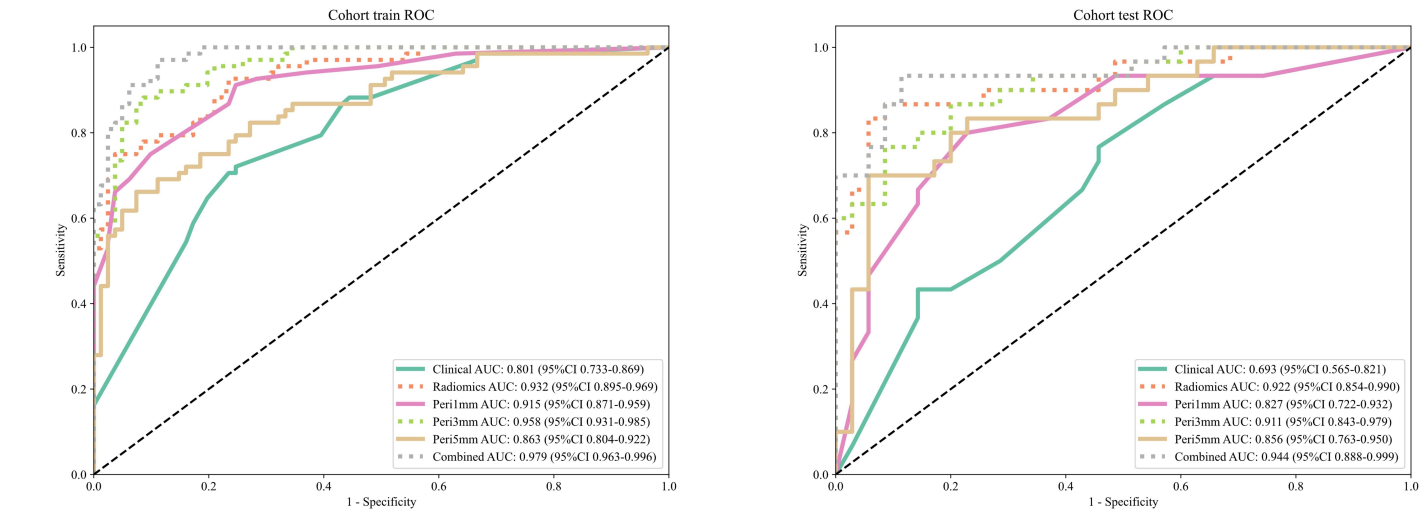


Figure 3

ROC comparison of clinical, radiomics, and fusion models. (A. Train cohort; B. Test cohort)

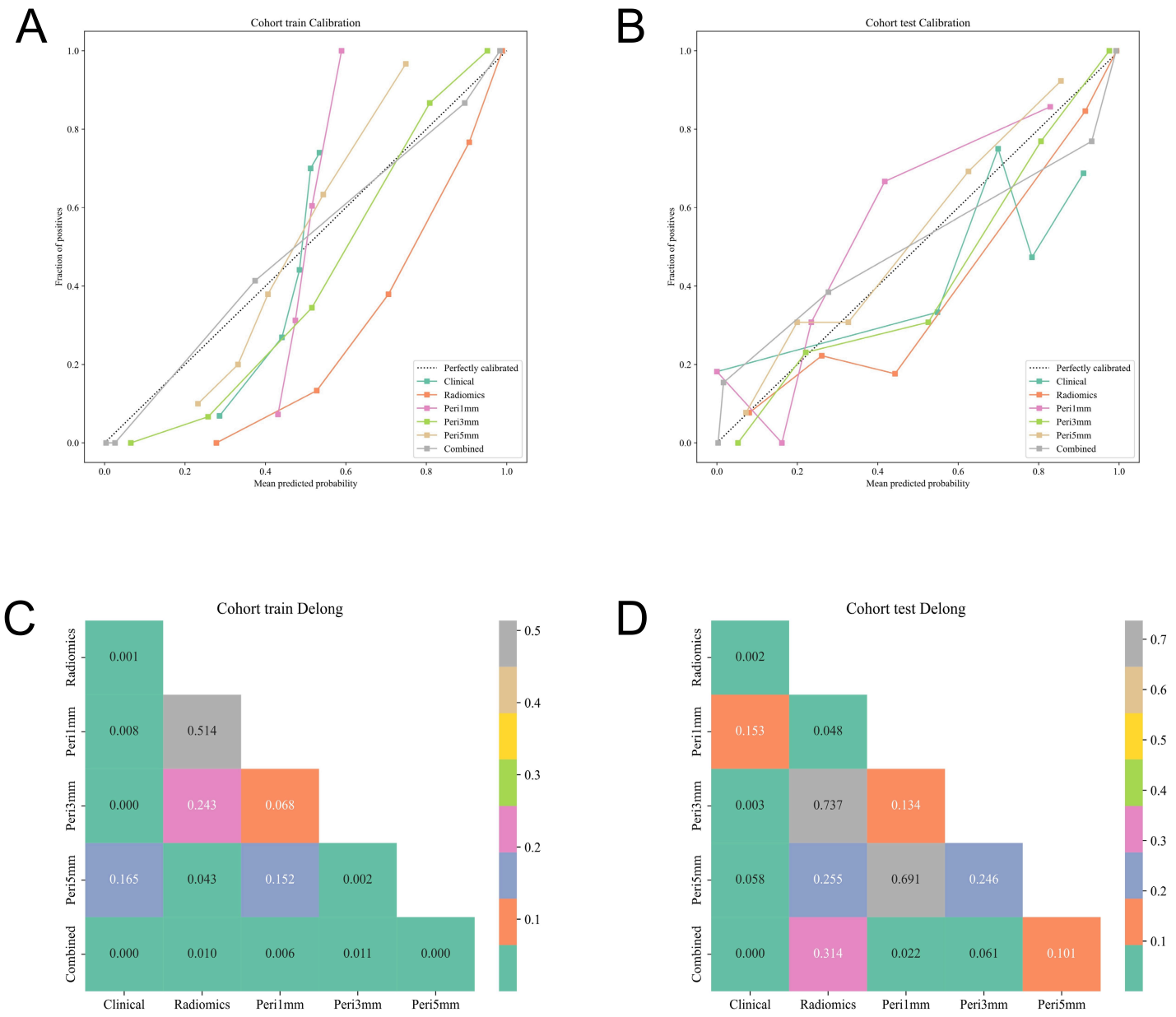


Figure 4

Calibration and Delong test. (A. Calibration in training cohort; B. Calibration in test cohort; C. Delong test in training cohort; D. Delong test in test cohort.)

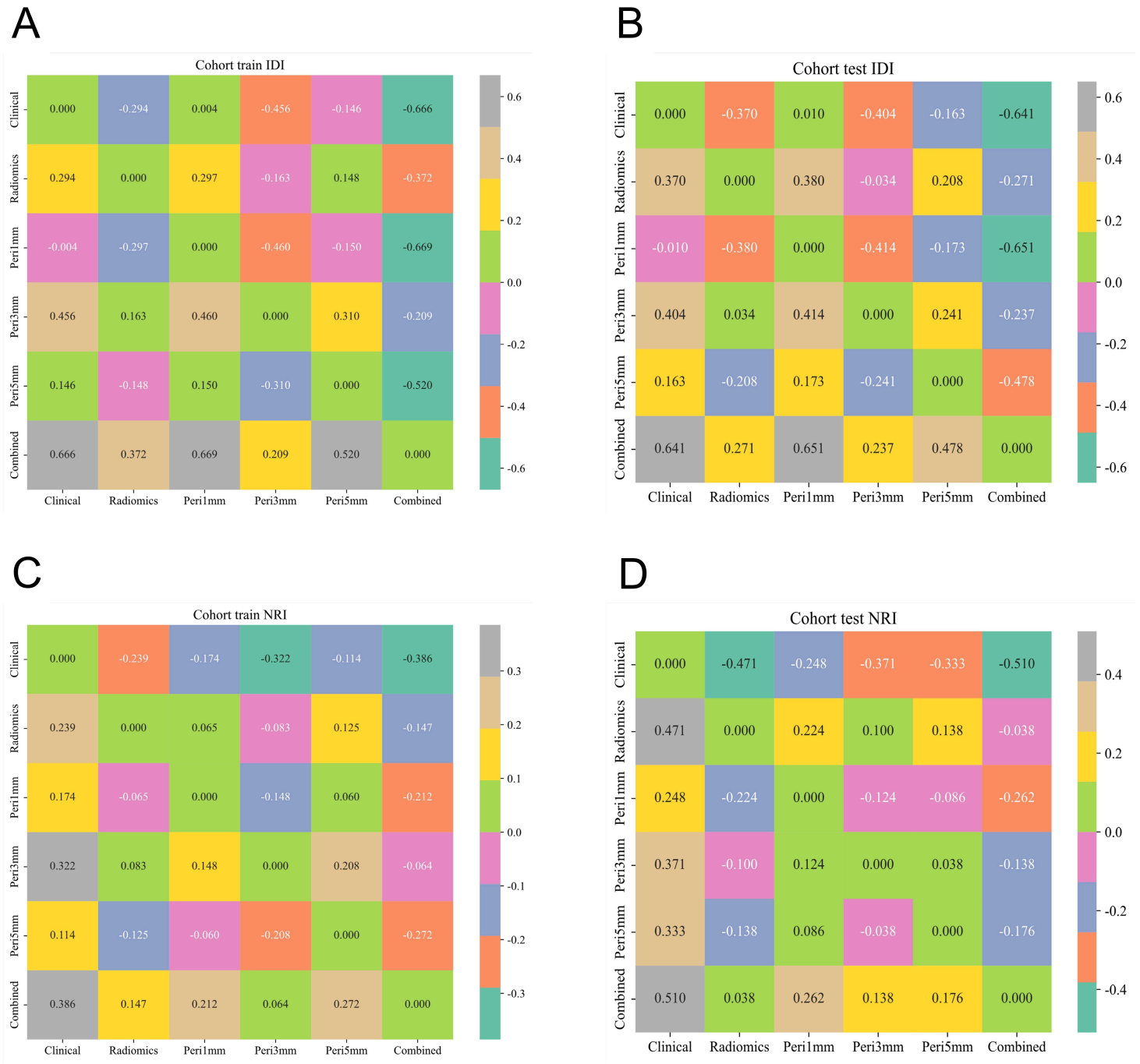


Figure 5

IDI and NRI of different models. (A. IDI in train cohort; B. IDI in test cohort; C. NRI in train cohort; D. NRI in test cohort.)

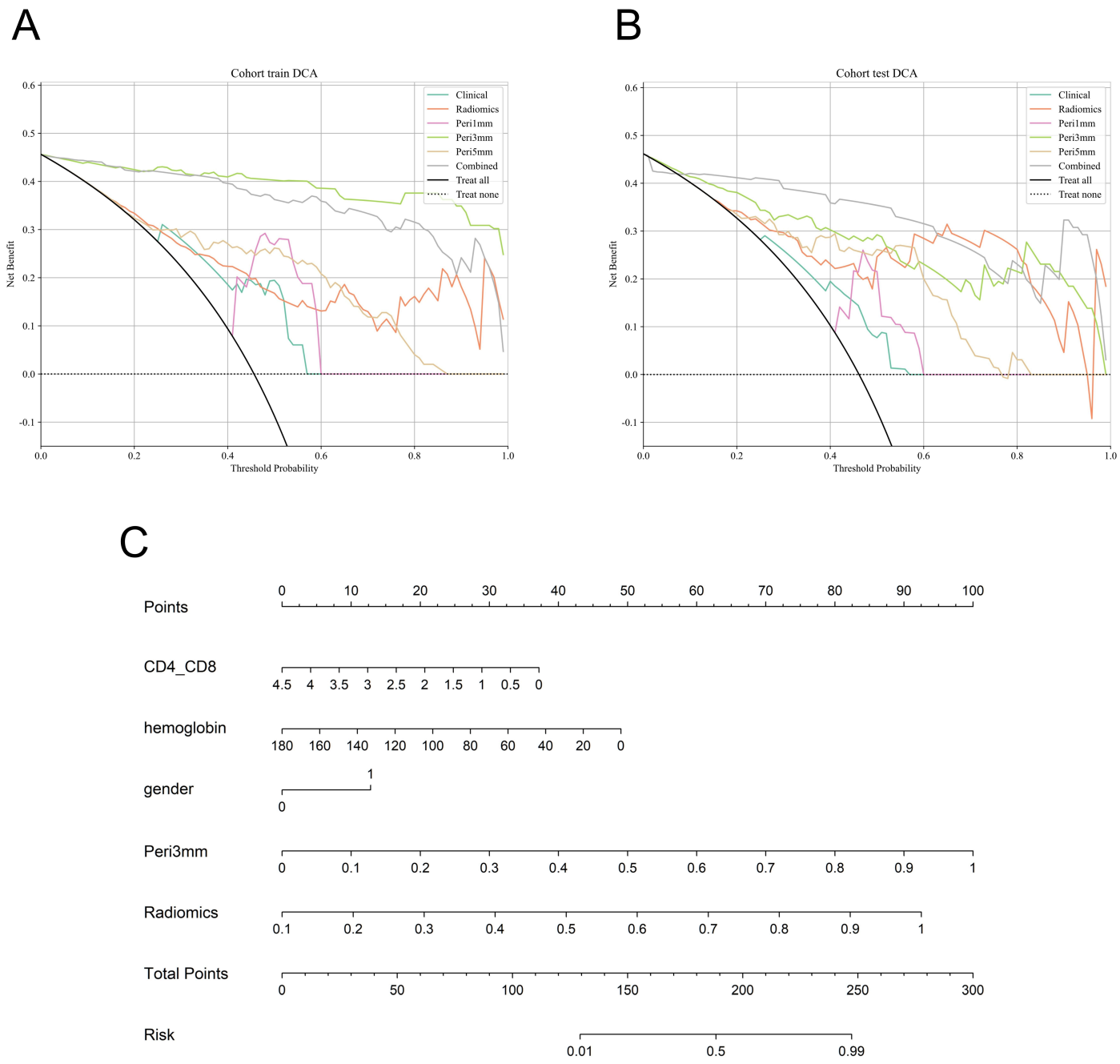


Figure 6

Clinical value analysis of the combined model. (A. DCA in training cohort; B. DCA in test cohort; C. Nomogram incorporating clinical and radiomics predictors.)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx