

Early prediction of refractory *Mycoplasma pneumoniae* pneumonia in children: development and internal validation of a combined CT radiomics–clinical machine learning model

Xiaoqiang Huang

Quanzhou Women's and Children's Hospital

Qiang Wang

Quanzhou Women's and Children's Hospital

Tianfeng Chen

Quanzhou Women's and Children's Hospital

Yan Yang

Quanzhou Women's and Children's Hospital

Risheng Huang

huang2026qz@163.com

Quanzhou First Hospital Affiliated to Fujian Medical University

Research Article

Keywords: *Mycoplasma pneumoniae* pneumonia, Refractory, Children, Radiomics, Machine learning, Nomogram, Computed tomography

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10852588/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Early prediction of refractory *Mycoplasma pneumoniae* pneumonia in children: development and internal validation of a combined CT radiomics–clinical machine learning model

Xiaoqiang Huang¹, Qiang Wang¹, Tianfeng Chen¹, Yan Yang¹, Risheng Huang^{2*}

1. Department of Radiology, Quanzhou Women's and Children's Hospital, Quanzhou 362000, Fujian, China

2. Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou 362000, Fujian, China

* **Correspondence:** Risheng Huang, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou 362000, Fujian, China. Email: huang2026qz@163.com

Abstract

Background: Refractory *Mycoplasma pneumoniae* pneumonia (RMPP) has a protracted course and frequent complications, and its diagnosis is often delayed; early identification of children at high risk is therefore crucial. This study aimed to develop and internally validate a combined machine-learning model integrating CT radiomics features and early clinical indicators to predict progression from *Mycoplasma pneumoniae* pneumonia (MPP) to RMPP in children.

Methods: This single-center retrospective study enrolled 527 children with MPP (204 RMPP, 323 non-refractory MPP [NRMPP]) between January 2022 and December 2024, randomly split into training (n = 368) and internal test (n = 159) cohorts at a ratio of approximately 7:3. From 1834 CT radiomics features, 25 selected by stability (intraclass correlation coefficient ≥ 0.85), Mann–Whitney U screening, correlation-based redundancy removal and LASSO dimensionality reduction formed a radiomics score (Rad-score); seven machine-learning classifiers were compared. A clinical model was built by multivariate logistic regression with backward stepwise selection of variables with $P < 0.1$ on univariate analysis, and a combined model integrating the clinical prediction and Rad-score was visualized as a nomogram. Performance was evaluated using the area under the receiver operating characteristic curve (AUC), calibration curves, the Hosmer–Lemeshow test and decision curve analysis (DCA).

Results: The combined model achieved AUCs of 0.876 (95% CI: 0.839–0.913) and 0.869 (95% CI: 0.815–0.923) in the training and test cohorts, respectively, significantly outperforming the clinical model (0.704 and 0.638; $P = 1.42 \times 10^{-8}$ and $P = 1.59 \times 10^{-6}$); in

the test cohort, its AUC was also higher than that of the radiomics model (0.851), although not significantly ($P = 0.060$). The final clinical model included LDH, peak body temperature and ALT. DCA demonstrated a higher net benefit of the combined model across a wide range of threshold probabilities. The Hosmer–Lemeshow test indicated some degree of miscalibration of the combined model in the test cohort ($P = 0.032$).

Conclusions: The combined machine-learning model showed good discrimination for predicting progression from MPP to RMPP (internal test AUC 0.869) and significantly outperformed the clinical model; however, its calibration was suboptimal in the test cohort (Hosmer–Lemeshow $P = 0.032$), and external validation and recalibration are required before clinical deployment.

Keywords: Mycoplasma pneumoniae pneumonia; Refractory; Children; Radiomics; Machine learning; Nomogram; Computed tomography

Background

Mycoplasma pneumoniae is one of the most common pathogens of community-acquired pneumonia (CAP) in children and the leading cause of pneumonia in Chinese children aged 6 years and above [1]. Macrolides remain the first-line treatment for M. pneumoniae pneumonia (MPP); however, the resistance rate of M. pneumoniae to macrolides in China has remained persistently high (approximately 81%) over the past two decades [1]. Owing to the spread of resistant strains and heterogeneous host immune responses, a subset of children continues to experience persistent fever and progressive clinical or radiological deterioration despite standardized macrolide therapy, thereby developing refractory M. pneumoniae pneumonia (RMPP) [1]. RMPP is characterized by a protracted disease course and is frequently complicated by plastic bronchitis [2], bronchiolitis obliterans [3], necrotizing pneumonia [4] and thromboembolism [5], and may result in permanent pulmonary sequelae. Recent meta-analyses have reported an overall prevalence of pediatric RMPP of approximately 37.8%, rising to 86.5% after the COVID-19 pandemic [6], underscoring a substantial disease burden.

According to the Evidence-Based Guideline for the Diagnosis and Treatment of Mycoplasma pneumoniae Pneumonia in Children (2023), the diagnosis of RMPP cannot be established until persistent fever and worsening pulmonary imaging findings are present for at least 7 days despite standardized macrolide treatment [1], by which time lung injury may have already progressed, leaving a narrow window for early intervention. Early identification of children at high risk of RMPP is therefore critical for the timely escalation of antimicrobial therapy, adjunctive corticosteroids or bronchoscopic intervention, and improved outcomes [7]. Previous studies have identified clinical and laboratory risk factors associated with RMPP, including prolonged fever, C-reactive protein (CRP), lactate dehydrogenase (LDH) and D-dimer [8–11], and have developed various predictive nomograms and scales based on these indicators [12–16]. However, such models rely largely on threshold-based laboratory parameters, show variable performance across populations, and rarely

incorporate imaging information, thus failing to capture the subtle heterogeneity of pulmonary lesions.

Chest CT is an essential tool for assessing the extent and severity of pulmonary involvement in pediatric MPP, and the 2023 guideline recommends lobar or segmental consolidation and pleural effusion as imaging features supporting the diagnosis [1]. Nevertheless, conventional imaging signs are mainly qualitative and frequently do not differ significantly between general MPP and RMPP, limiting their discriminative ability. Radiomics enables high-throughput extraction of quantitative gray-level, texture and shape features that are invisible to the human eye, and has demonstrated value in the diagnosis and prognostication of various diseases. Recent studies have applied CT-based radiomics and deep learning to the early prediction of RMPP, including multicenter radiomics models [17], transformer-based deep learning frameworks [18], AI-driven CT lesion quantification [19] and pulmonary vascular quantification [20]. However, most of these studies relied on a single imaging modality or employed end-to-end “black-box” models with limited interpretability; studies integrating CT radiomics with early clinical features and comparing multiple machine-learning algorithms for pediatric RMPP prediction remain scarce [21–23].

Accordingly, this study aimed to develop and internally validate a combined machine-learning model integrating CT radiomics features and early clinical indicators, visualized as a nomogram, to predict the risk of progression from MPP to RMPP in children, thereby providing an interpretable decision-support tool for risk stratification and timely clinical intervention. Given the homogeneity of the single-center population, health inequalities between sociodemographic groups were not specifically assessed.

Methods

Study design and participants

This single-center retrospective cohort study consecutively enrolled children with MPP at Quanzhou Women’s and Children’s Hospital between January 2022 and December 2024. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the hospital (approval No.: (2025) Ethics Review No. (39)). Owing to the retrospective design, the requirement for informed consent from the children and their guardians was waived by the ethics committee. All clinical data and medical images were de-identified prior to analysis.

The inclusion criteria were: (1) age > 28 days and ≤ 18 years; (2) a diagnosis of MPP meeting the criteria of the Evidence-Based Guideline for the Diagnosis and Treatment of Mycoplasma pneumoniae Pneumonia in Children (2023), with corresponding etiological evidence; (3) non-contrast chest CT performed within 48 h of the first visit or admission and before determination of the RMPP outcome; (4) thin-slice CT images with a reconstruction thickness < 1.5 mm and complete DICOM data; and (5) baseline clinical and laboratory data sufficient for the planned analyses and hospitalization or follow-up records sufficient to determine the clinical outcome. Missing values of PCT (198 cases) and D-dimer

(88 cases) were not a reason for exclusion and were handled by multiple imputation (see Statistical analysis).

The exclusion criteria were: (1) CT images with severe motion artifacts, reconstruction abnormalities or data corruption that precluded lesion segmentation; (2) congenital cardiopulmonary disease, immunodeficiency, malignancy, severe metabolic disease or other underlying conditions that could significantly affect pulmonary imaging or the disease course; (3) long-term immunosuppressive therapy before the index CT; and (4) previous lung surgery or thoracic/parenchymal structural abnormalities that significantly affected imaging analysis. Co-infection with bacteria, viruses or other pathogens was not an exclusion criterion and was not included as a predictive variable.

The first chest CT examination meeting the requirements was defined as the index CT. All predictive variables were collected before determination of the RMPP outcome. According to the 2023 guideline, RMPP was defined as persistent fever, worsening clinical symptoms or progressive pulmonary imaging findings after at least 7 days of standardized macrolide treatment [1]; the remaining children were assigned to the non-refractory MPP (NRMPP) group. All children received standardized treatment in accordance with the 2023 guideline; treatment details were not included as predictive variables. The outcome was determined from medical records by clinicians who were blinded to the model predictions.

A total of 527 children were ultimately enrolled, including 204 with RMPP and 323 with NRMPP. The cohort was randomly split into a training cohort and an internal test cohort at a ratio of approximately 7:3. The data split was performed before feature selection and model training. The patient enrollment flow is shown in Fig. 1. No formal sample size calculation was performed; all eligible consecutive children were enrolled. The final sample of 527 children (204 RMPP events) met conventional event-per-variable requirements for the clinical model, and the event-to-parameter ratio in the training cohort is discussed in the Limitations.

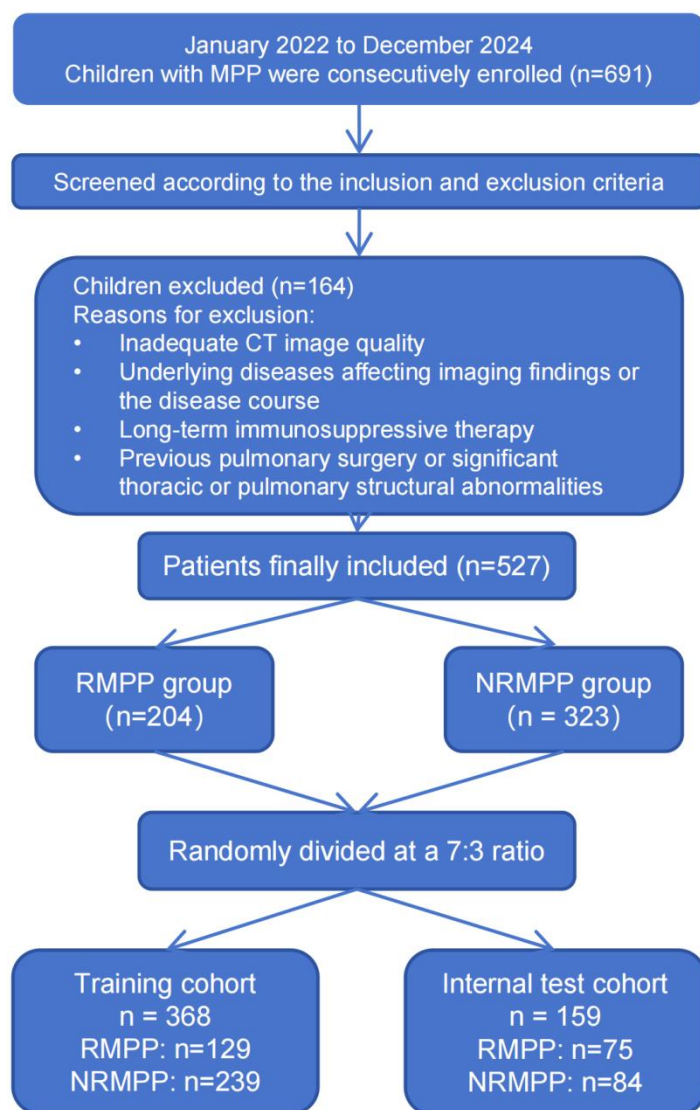


Fig. 1 Flowchart of patient enrollment.

Clinical and conventional imaging data collection

The following data were extracted from the electronic medical record system: age, sex, fever duration, peak body temperature, white blood cell (WBC) count, neutrophil percentage, lymphocyte percentage, C-reactive protein (CRP), procalcitonin (PCT), lactate dehydrogenase (LDH), alanine aminotransferase (ALT), aspartate aminotransferase (AST) and D-dimer.

Conventional imaging signs on the index CT were also recorded, including pulmonary consolidation or atelectasis, multilobar involvement, pleural effusion, and bronchial stenosis or obstruction. When a laboratory indicator had multiple measurements, the result closest to the index CT and obtained within 48 h before or after the CT examination was selected. Fever duration was defined as the time from the first body temperature $\geq$

153 37.5 °C to the index CT; peak body temperature was defined as the highest temperature
154 recorded from disease onset to the index CT.

155 Missing values of PCT and D-dimer were handled by multivariate imputation by chained
156 equations (MICE); the imputation model was fitted in the training cohort, and the same
157 imputation parameters were applied to the test cohort.

158 **CT image acquisition and preprocessing**

159 All children underwent non-contrast chest CT on an Optima CT660 helical scanner (GE
160 Healthcare). The scan parameters were as follows: tube voltage 100 kV, adaptive tube
161 current 20–400 mA, acquisition slice thickness and interval 5 mm, thin-slice lung-window
162 images reconstructed with the LUNG kernel, and thin-slice reconstruction thickness and
163 interval 0.625 mm.

164 Children were scanned in the supine position, head-first, with both arms raised as far as
165 possible. Those who could cooperate received breathing training before scanning; those
166 who could not were scanned during natural sleep or under sedation with free breathing.
167 The scan range extended from approximately 2 cm above the thoracic inlet to
168 approximately 2 cm below the posterior costophrenic angle.

169 All images underwent uniform preprocessing, including resampling to isotropic voxels of 1
170 $\times 1 \times 1$ mm³ and gray-level normalization, to reduce the influence of differences in
171 scanning parameters.

172 **Lesion segmentation and reproducibility assessment**

173 DICOM images were imported into ITK-SNAP software (version 4.2.2). Two radiologists
174 with 14 and 15 years of experience in pediatric chest imaging, respectively, manually
175 performed three-dimensional segmentation along the boundaries of pulmonary lesions
176 slice by slice. For multiple or multilobar lesions, all visible pulmonary parenchymal lesions
177 were included.

178 To evaluate segmentation reproducibility, the two radiologists independently segmented a
179 random sample of 30 children, and the intraclass correlation coefficient (ICC) of the
180 radiomics features was calculated. Features with an ICC ≥ 0.85 were considered stable and
181 were retained for subsequent analysis.

182 **Radiomics feature extraction and selection**

183 Radiomics features were extracted using the PyRadiomics software package (version
184 3.0.1), including shape features, first-order gray-level features and texture features.
185 Texture features comprised gray-level co-occurrence matrix (GLCM), gray-level run-length
186 matrix (GLRLM), gray-level size zone matrix (GLSZM), gray-level dependence matrix
187 (GLDM) and neighboring gray-tone difference matrix (NGTDM) features. In addition,
188 features were extracted from the original images and from exponential, logarithm,
189 Laplacian of Gaussian, square, square root and wavelet transformed images.

All feature selection was performed exclusively in the training cohort. First, stable features with an ICC ≥ 0.85 were retained and radiomics features were Z-score standardized; the test cohort was transformed using the mean and standard deviation of the training cohort. Features were then screened by the Mann–Whitney U test, retaining features with $P < 0.05$. For any pair of features with an absolute Spearman correlation coefficient > 0.90 , a greedy recursive strategy was used to remove the more redundant feature. Finally, the least absolute shrinkage and selection operator (LASSO) logistic regression was used for dimensionality reduction (implemented using scikit-learn), and the optimal regularization parameter λ was determined by 10-fold cross-validation using the minimum-error criterion. The features with non-zero coefficients were linearly weighted by their regression coefficients to calculate the radiomics score (Rad-score).

Model construction

Clinical model. Candidate clinical variables were first analyzed by univariate logistic regression in the training cohort; variables with $P < 0.10$ were entered into a multivariate logistic regression model to evaluate the independent associations of each candidate variable (Table 2). Backward stepwise selection was then applied to build the clinical model (clinical signature), which finally included LDH, peak body temperature and ALT. Although age reached statistical significance in the full multivariate model ($P = 0.031$), its aOR was close to 1 (1.095 per year) with limited marginal contribution, and it was not retained in the interest of model parsimony and clinical practicality; PCT ($P = 0.077$) was also not retained.

Radiomics model. The selected radiomics features were separately input into seven classifiers: logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), extremely randomized trees (ExtraTrees), XGBoost and LightGBM (LR, SVM, KNN, RF and ExtraTrees were implemented using scikit-learn; XGBoost using the xgboost library; and LightGBM using the lightgbm library). Hyperparameters were determined by 10-fold stratified cross-validation within the training cohort. The final classifier was selected based on cross-validation performance and model stability, and was evaluated once on the internal test cohort.

Combined model and nomogram. The clinical model prediction and the Rad-score were integrated using logistic regression to establish the combined model, which was visualized as a nomogram. Model discrimination was evaluated by the ROC curve and AUC, calibration by the calibration curve and the Hosmer–Lemeshow test, and clinical utility by decision curve analysis (DCA). The workflow of model construction and evaluation is shown in Fig. 2.

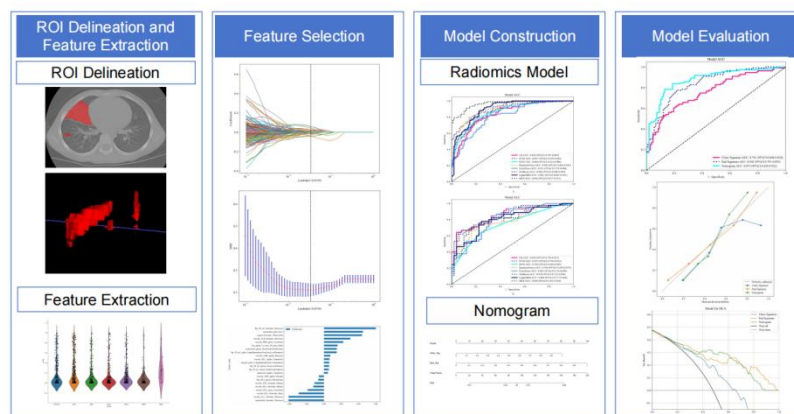


Fig. 2 Workflow of model construction and evaluation.

Statistical analysis

All statistical analyses and plotting were performed using Python 3.13, with the following key libraries: statsmodels (logistic regression and MICE multiple imputation), SciPy (t test, Mann–Whitney U test, chi-square test and Fisher’s exact test), pandas (data management) and matplotlib (plotting). Continuous variables with a normal distribution were expressed as mean $\pm$ standard deviation and compared using the independent-samples t test (scipy.stats.ttest_ind); non-normally distributed variables were expressed as median (interquartile range) and compared using the Mann–Whitney U test (scipy.stats.mannwhitneyu); categorical variables were expressed as counts and percentages and compared using the chi-square test or Fisher’s exact test. Univariate and multivariate logistic regression and the combined model were fitted using statsmodels, with results expressed as odds ratios (ORs), adjusted odds ratios (aORs) and 95% confidence intervals (CIs); missing values of PCT and D-dimer were imputed using the MICE algorithm implemented in statsmodels. The DeLong test was used to compare the AUCs of the ROC curves, implemented following the method of DeLong et al. ROC curves, calibration curves and decision curve analysis plots were generated using matplotlib. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value and F1 score were reported. All tests were two-sided, and $P < 0.05$ was considered statistically significant. This study is reported in accordance with the TRIPOD+AI statement, and the completed 27-item checklist is provided as Additional file 1.

Results

Patient characteristics

During the study period, 691 children with MPP were screened, of whom 164 were excluded for not meeting the inclusion criteria or meeting the exclusion criteria, leaving 527 children for analysis. There were 204 (38.7%) children with RMPP and 323 (61.3%) with NRMPP. According to the actual model output, the training cohort comprised 368 children (129 RMPP, 35.1%; 239 NRMPP) and the test cohort comprised 159 children (75 RMPP, 47.2%; 84 NRMPP). Baseline characteristics were broadly comparable between the

training and test cohorts; the proportion of RMPP was higher in the test cohort than in the training cohort (47.2% vs. 35.1%).

In the full cohort, compared with the NRMPP group, the RMPP group was older ($P = 0.045$) and had a longer fever duration, higher peak body temperature, and higher CRP, neutrophil percentage, PCT, LDH, ALT, AST and D-dimer levels, and a lower lymphocyte percentage (all $P < 0.05$). Sex, WBC count, pleural effusion, bronchial stenosis or obstruction, multilobar distribution and consolidation/atelectasis did not differ significantly between the two groups (Table 1).

Table 1 Baseline clinical characteristics of the full cohort

Variable	Type	NRMPP group	RMPP group	Test	P value
Age (years)	Continuous	5.00 [3.00, 7.00]	5.00 [3.00, 8.00]	Mann-Whitney U	0.045
Fever duration (days)	Continuous	5.00 [2.00, 6.00]	5.00 [4.00, 7.00]	Mann-Whitney U	0.001
Peak body temperature (°C)	Continuous	39.00 [38.50, 39.50]	39.50 [39.00, 40.00]	Mann-Whitney U	<0.001
C-reactive protein (CRP, mg/L)	Continuous	11.78 [5.39, 24.46]	16.29 [6.48, 28.45]	Mann-Whitney U	0.009
White blood cell count (WBC, $\times 10^9/L$)	Continuous	9.07 [7.30, 11.40]	8.60 [6.78, 11.12]	Mann-Whitney U	0.202
Neutrophil percentage (%)	Continuous	62.60 [53.60, 71.00]	64.85 [56.25, 73.12]	Mann-Whitney U	0.013
Lymphocyte percentage (%)	Continuous	27.65 [19.55, 35.77]	25.20 [18.05, 32.35]	Mann-Whitney U	0.009
Procalcitonin (PCT, ng/mL)	Continuous	0.28 [0.18, 0.45]	0.39 [0.24, 0.69]	Mann-Whitney U	<0.001
Lactate dehydrogenase (LDH, U/L)	Continuous	295.00 [257.50, 347.00]	375.00 [295.00, 487.50]	Mann-Whitney U	<0.001
Alanine aminotransferase (ALT, U/L)	Continuous	19.00 [13.00, 25.00]	21.00 [15.00, 31.00]	Mann-Whitney U	<0.001

Variable	Type	NRMPP group	RMPP group	Test	P value
Aspartate aminotransferase (AST, U/L)	Continuous	37.00 [30.00, 45.00]	40.00 [32.00, 51.00]	Mann-Whitney U	<0.001
D-dimer (mg/L)	Continuous	0.46 [0.32, 0.65]	0.70 [0.43, 1.42]	Mann-Whitney U	<0.001
Sex (male = 1)	Binary	184/323 (57.0%)	102/204 (50.0%)	Chi-square	0.141
Pleural effusion (imaging)	Binary	64/323 (19.8%)	41/204 (20.1%)	Chi-square	1.000
Bronchial stenosis/obstruction (imaging)	Binary	179/323 (55.4%)	123/204 (60.3%)	Chi-square	0.312
Multilobar distribution (imaging)	Binary	289/323 (89.5%)	181/204 (88.7%)	Chi-square	0.900
Consolidation/atelectasis (imaging)	Binary	300/323 (92.9%)	195/204 (95.6%)	Chi-square	0.280

264 Continuous variables are presented as median [interquartile range] (Mann–Whitney U
265 test); categorical variables as n (%) (chi-square or Fisher’s exact test).

266 **Clinical factors associated with RMPP**

267 Univariate logistic regression showed that LDH, peak body temperature, AST, ALT, D-
268 dimer, fever duration, neutrophil percentage, lymphocyte percentage, CRP, PCT and age
269 were associated with the RMPP outcome (all $P < 0.1$).

270 In the multivariate analysis, LDH (aOR = 1.005, 95% CI: 1.003–1.007, $P < 0.001$), peak body
271 temperature (aOR = 1.727, 95% CI: 1.318–2.263, $P < 0.001$), ALT (aOR = 1.015, 95% CI:
272 1.001–1.029, $P = 0.035$) and age (aOR = 1.095, 95% CI: 1.008–1.190, $P = 0.031$) were
273 independently associated with a higher risk of RMPP; PCT showed borderline significance
274 (aOR = 1.240, 95% CI: 0.977–1.573, $P = 0.077$). D-dimer, fever duration, neutrophil
275 percentage and CRP were not statistically significant in the multivariate model (Table 2).
276 On this basis, after backward stepwise selection, the final clinical model included LDH, peak
277 body temperature and ALT; age (aOR close to 1) and PCT (borderline significance) were
278 not retained.

279 **Table 2 Univariate and multivariate logistic regression analysis**

Variable	Univariable B (SE)	Univariable OR (95% CI)	P	Multivariable B (SE)	aOR (95% CI)	P
Lactate dehydrogenase	0.0055	1.005	<0.00	0.0050	1.005	<0.00

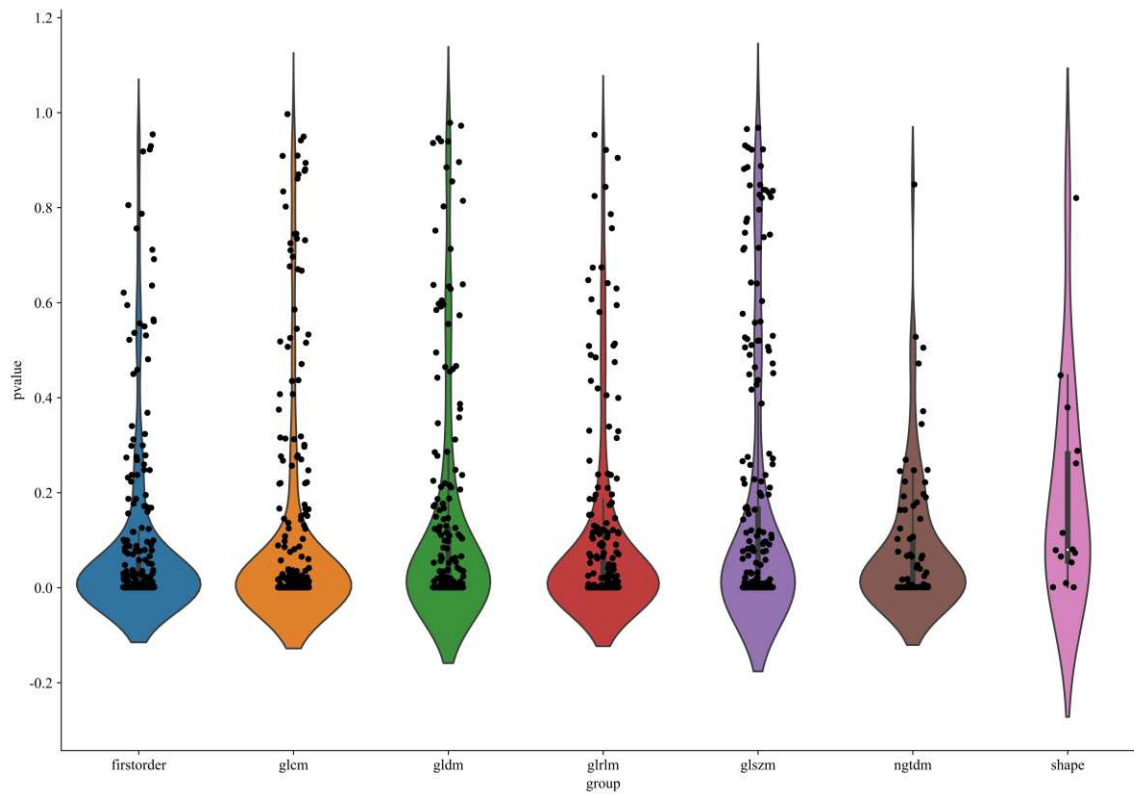
Variable	Univariable B (SE)	Univariable OR (95% CI)	P	Multivariable B (SE)	aOR (95% CI)	P
(LDH)	(0.0008)	(1.004–1.007)	1	(0.0009)	(1.003–1.007)	1
Peak body temperature	0.7630 (0.1193)	2.145 (1.697–2.710)	<0.001	0.5462 (0.1379)	1.727 (1.318–2.263)	<0.001
Aspartate aminotransferase (AST)	0.0245 (0.0063)	1.025 (1.012–1.037)	<0.001	—	—	—
Alanine aminotransferase (ALT)	0.0235 (0.0068)	1.024 (1.010–1.037)	<0.001	0.0151 (0.0071)	1.015 (1.001–1.029)	0.035
D-dimer	0.3230 (0.0970)	1.381 (1.142–1.670)	<0.001	0.0869 (0.0820)	1.091 (0.929–1.281)	0.29
Fever duration	0.1057 (0.0327)	1.111 (1.042–1.185)	0.001	–0.0129 (0.0444)	0.987 (0.905–1.077)	0.772
Neutrophil percentage	0.0190 (0.0070)	1.019 (1.005–1.033)	0.006	0.0086 (0.0088)	1.009 (0.991–1.026)	0.33
Lymphocyte percentage	–0.0217 (0.0079)	0.979 (0.963–0.994)	0.006	—	—	—
C-reactive protein (CRP)	0.0086 (0.0032)	1.009 (1.002–1.015)	0.008	–0.0038 (0.0041)	0.996 (0.988–1.004)	0.36
Procalcitonin (PCT)	0.2519 (0.0985)	1.287 (1.061–1.560)	0.011	0.2149 (0.1216)	1.240 (0.977–1.573)	0.077
Age	0.0605 (0.0331)	1.062 (0.996–1.134)	0.067	0.0910 (0.0422)	1.095 (1.008–	0.031

Variable	Univariable B (SE)	Univariable OR (95% CI)	P	Multivariable B (SE)	aOR (95% CI)	P
					1.190)	
Sex (male = 1)	-0.2805 (0.1795)	0.755 (0.531–1.074)	0.118	—	—	—
Consolidation/atelectasis (imaging)	0.5075 (0.4038)	1.661 (0.753–3.665)	0.209	—	—	—
Bronchial stenosis/obstruction (imaging)	0.2002 (0.1817)	1.222 (0.856–1.744)	0.271	—	—	—
White blood cell count (WBC)	-0.0121 (0.0198)	0.988 (0.950–1.027)	0.541	—	—	—
Multilobar distribution (imaging, validated)	-0.0771 (0.2861)	0.926 (0.528–1.622)	0.788	—	—	—
Pleural effusion (imaging)	0.0178 (0.2236)	1.018 (0.657–1.578)	0.937	—	—	—

280 The multivariate analysis included variables with $P < 0.1$ on univariate analysis; collinear
281 variables (AST and lymphocyte percentage) were excluded. PCT and D-dimer had 198 and
282 88 missing values, respectively, which were imputed using MICE multiple imputation. * $P <$
283 0.05, † $P < 0.1$.

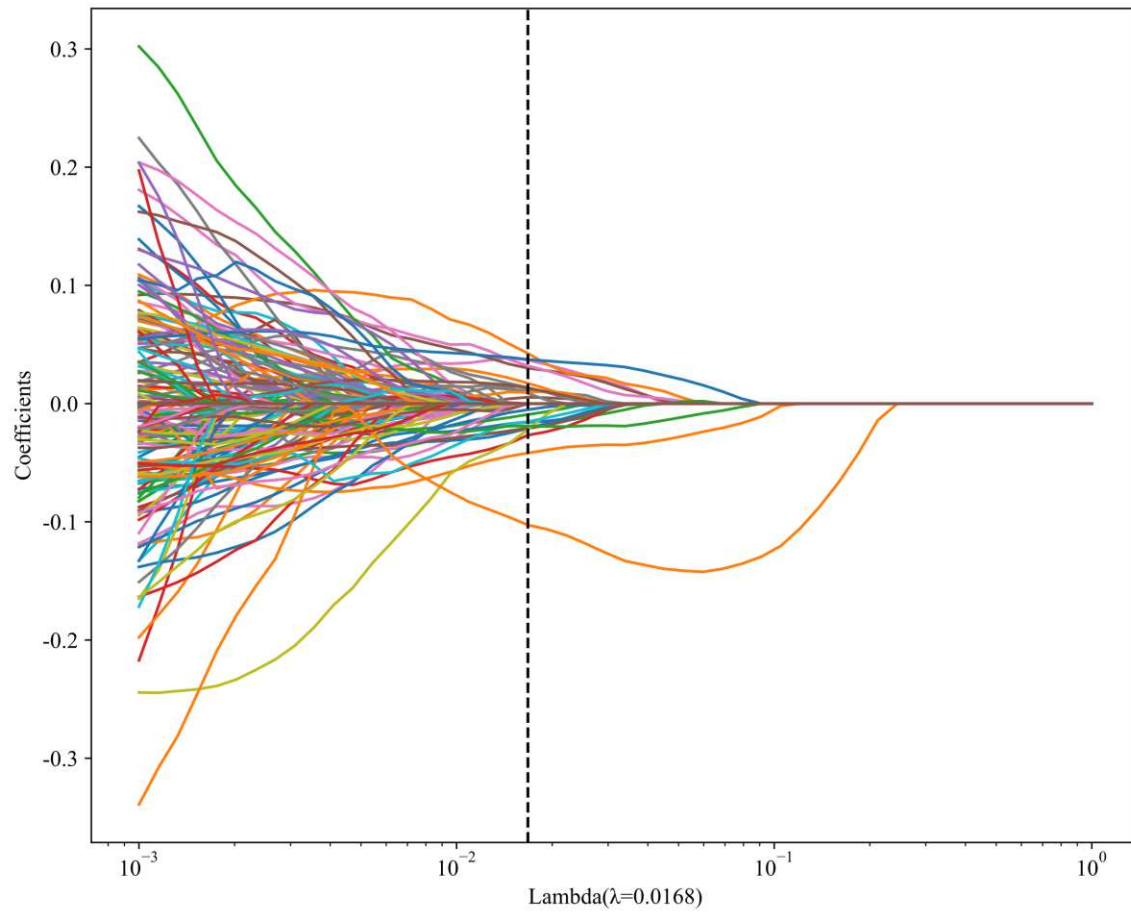
284 Radiomics feature selection and model comparison

285 A total of 1834 radiomics features were extracted from the original and transformed CT
286 images; Fig. 3 shows all features and their corresponding P values. For features with high
287 reproducibility, Spearman rank correlation coefficients were calculated, and when the
288 correlation coefficient between any two features exceeded 0.9, one of the two was retained.
289 To preserve the descriptive capability of the features as much as possible, a greedy
290 recursive deletion strategy was adopted, in which the most redundant feature in the
291 current set was removed at each step. Non-zero coefficients were selected to build the Rad-
292 score based on the LASSO logistic regression model. The coefficients and mean squared
293 error (MSE) of the 10-fold cross-validation are shown in Figs. 4 and 5. Ultimately, 25
294 radiomics features with the greatest discriminative value were selected from the high-
295 dimensional features to construct the prediction model (Fig. 6).



296

297 Fig. 3 Univariate analysis results of radiomics features and distribution of feature classes.



298

299 Fig. 4 LASSO logistic regression coefficient path plot.

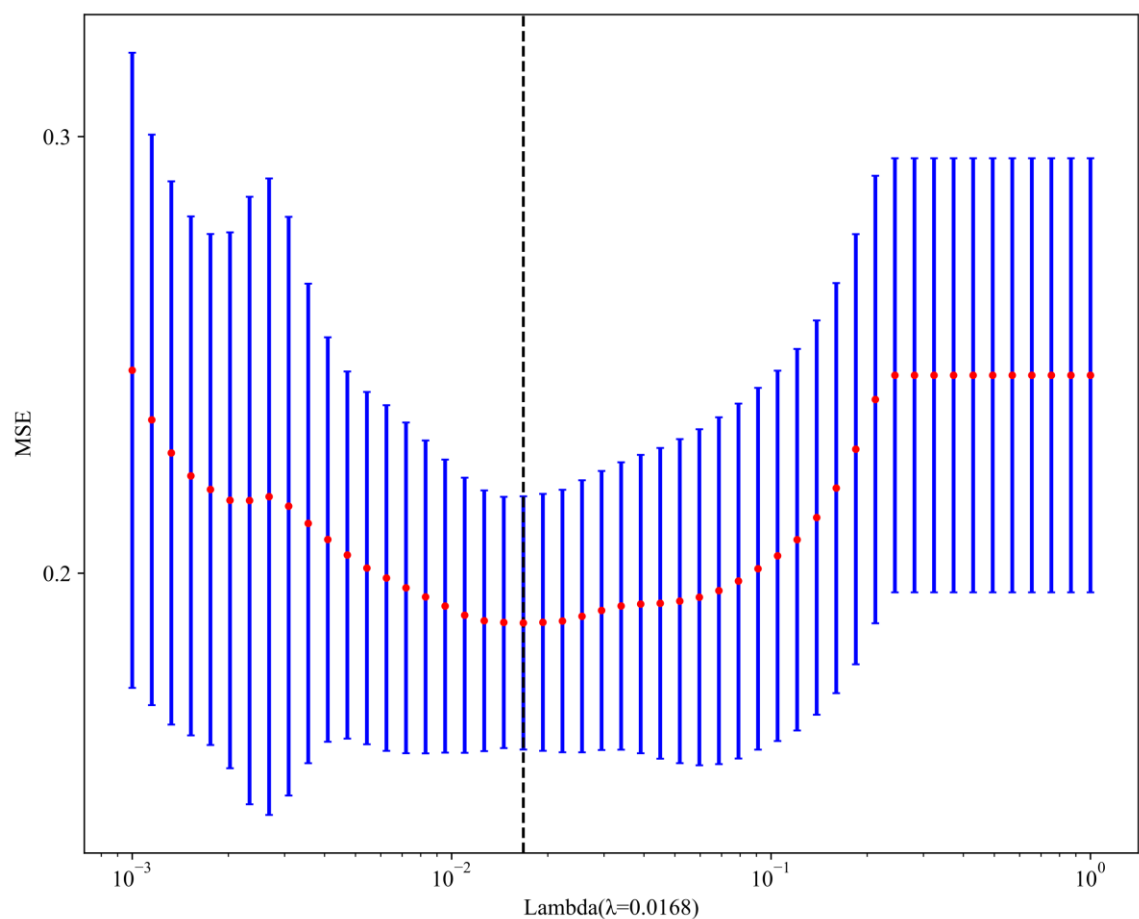


Fig. 5 Cross-validation error curve of the LASSO logistic regression.

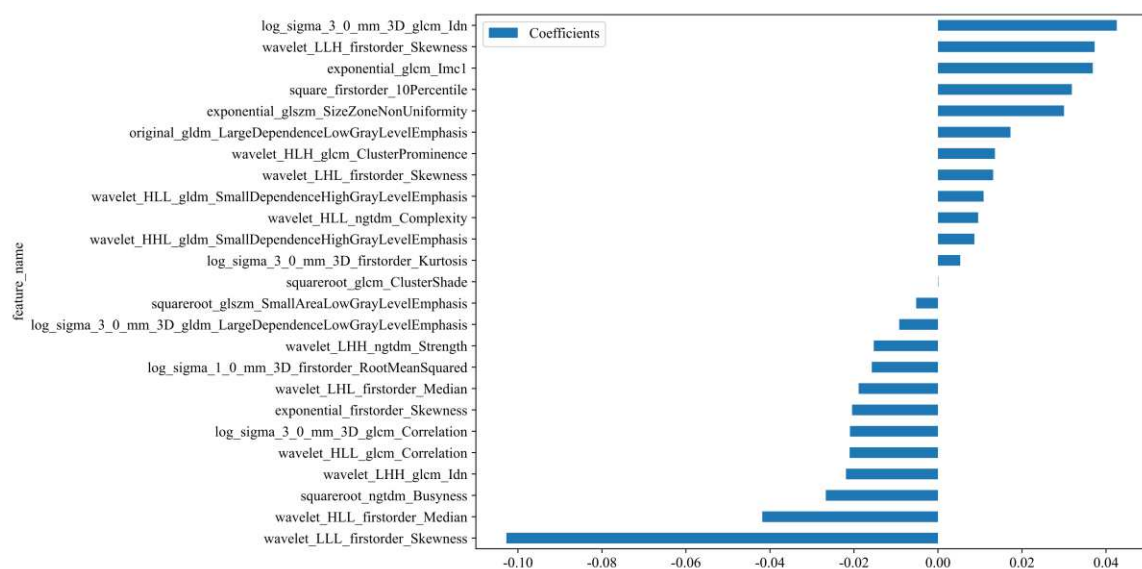


Fig. 6 Radiomics features included in the final radiomics model and their corresponding regression coefficients.

Based on the 10-fold cross-validated performance and model stability in the training cohort (i.e., the smallest gap between training and test performance and the lowest overfitting risk), LR was selected as the final radiomics model. In the internal test cohort, LR achieved the highest AUC of 0.851 (95% CI: 0.793–0.909) among the seven classifiers (SVM 0.827 [95% CI: 0.763–0.891], KNN 0.786 [95% CI: 0.717–0.855], RF 0.794 [95% CI: 0.725–0.864], ExtraTrees 0.806 [95% CI: 0.740–0.871], XGBoost 0.814 [95% CI: 0.748–0.881] and LightGBM 0.786 [95% CI: 0.715–0.856]), corroborating the robustness of the model selection. The detailed model performance is shown in Fig. 7 and Table 3.

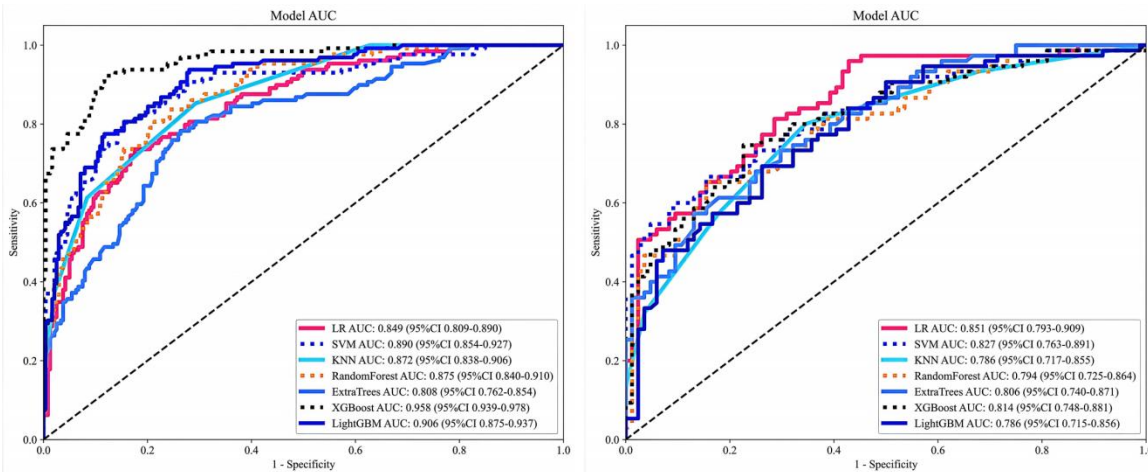


Fig. 7 ROC curves of the different radiomics models in the training (left) and internal test (right) cohorts.

Table 3 Predictive performance of the radiomics models

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.793	0.849	0.8088–0.8896	0.736	0.824	0.693	0.853	0.693	0.736	0.714	0.407	label-training
LR	0.755	0.851	0.7934–0.9095	0.96	0.571	0.667	0.941	0.667	0.96	0.787	0.231	label-test
SVM	0.832	0.809	0.8540–0.9266	0.798	0.849	0.741	0.886	0.741	0.798	0.769	0.368	label-training
SVM	0.761	0.827	0.76327–0.8910	0.667	0.845	0.794	0.74	0.794	0.667	0.725	0.477	label-test

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
KNN	0.758	0.872	0.8379–0.9064	0.853	0.707	0.611	0.899	0.611	0.853	0.712	0.4	label-training
KNN	0.723	0.786	0.7169–0.8551	0.8	0.655	0.674	0.786	0.674	0.8	0.732	0.4	label-testing
Random Forest	0.785	0.875	0.8395–0.9102	0.837	0.757	0.651	0.896	0.651	0.837	0.732	0.347	label-training
Random Forest	0.755	0.794	0.7246–0.8642	0.653	0.845	0.79	0.732	0.79	0.653	0.715	0.477	label-testing
ExtraTrees	0.753	0.808	0.7616–0.8542	0.783	0.736	0.616	0.863	0.616	0.783	0.689	0.327	label-training
ExtraTrees	0.73	0.806	0.7402–0.8714	0.573	0.869	0.796	0.695	0.796	0.573	0.667	0.447	label-testing
XGBoost	0.894	0.958	0.9390–0.9775	0.922	0.879	0.804	0.955	0.804	0.922	0.859	0.355	label-training
XGBoost	0.761	0.814	0.7477–0.8810	0.747	0.774	0.747	0.774	0.747	0.747	0.747	0.423	label-testing
LightGBM	0.845	0.906	0.8753–0.9366	0.775	0.883	0.781	0.879	0.781	0.775	0.778	0.406	label-training
LightGBM	0.717	0.786	0.7151–0.8563	0.693	0.738	0.703	0.729	0.703	0.693	0.698	0.409	label-testing

Performance of the clinical, radiomics and combined models

The clinical model achieved AUCs of 0.704 (95% CI: 0.645–0.762) and 0.638 (95% CI: 0.552–0.725) in the training and test cohorts, respectively. The corresponding AUCs of the radiomics model were 0.849 (95% CI: 0.809–0.890) and 0.851 (95% CI: 0.793–0.909).

The combined model performed best, with a training-cohort AUC of 0.876 (95% CI: 0.839–0.913) and accuracy, sensitivity and specificity of 0.818, 0.767 and 0.845, respectively; in the test cohort, the AUC was 0.869 (95% CI: 0.815–0.923) with an accuracy, sensitivity and specificity of 0.774, 0.773 and 0.774, respectively. Details are shown in Fig. 8 and Table 4.

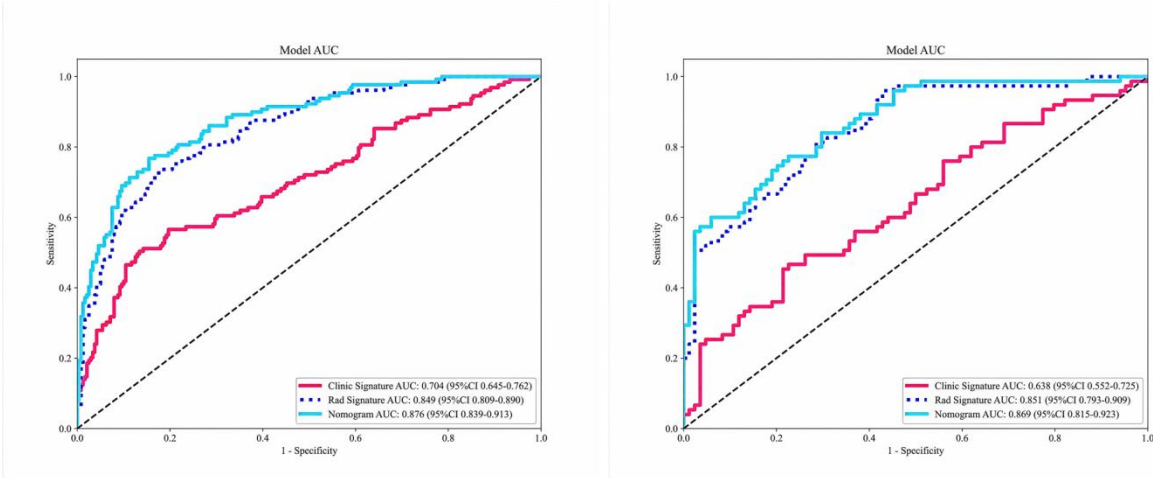


Fig. 8 ROC curves of the clinical, radiomics and combined models in the training (left) and internal test (right) cohorts.

Table 4 Evaluation metrics of the clinical, radiomics and combined models

Signature	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Cohort
Clinic Signature	0.739	0.704	0.645–0.762	0.504	0.866	0.67	0.764	0.67	0.504	0.5	0.396	Train
Rad Signature	0.793	0.849	0.809–0.8896	0.736	0.824	0.6	0.853	0.693	0.736	0.7	0.407	Train
Nomogram	0.818	0.876	0.839–0.9127	0.767	0.845	0.728	0.871	0.728	0.767	0.7	0.376	Train
Clinic Signature	0.629	0.638	0.552–0.72	0.467	0.774	0.6	0.619	0.648	0.467	0.5	0.375	Test

Signature	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Cohort
			49									
Rad Signature	0.755	0.851	0.7934–0.9095	0.96	0.571	0.667	0.941	0.667	0.96	0.787	0.231	Test
Nomogram	0.774	0.869	0.8154–0.9230	0.773	0.774	0.753	0.793	0.753	0.773	0.763	0.447	Test

The DeLong test showed that the AUC of the combined model was significantly higher than that of the clinical model in both the training and test cohorts ($P = 1.42 \times 10^{-8}$ and $P = 1.59 \times 10^{-6}$, respectively). The AUC of the combined model was also higher than that of the radiomics model in the training cohort ($P = 0.002$), but the difference did not reach statistical significance in the test cohort ($P = 0.060$). Details are shown in Table 5.

Table 5 DeLong test results

Cohort	Nomogram vs. clinical model	Nomogram vs. radiomics model
Training	1.42E-08	0.002
Test	1.59E-06	0.06

The calibration curves are shown in Fig. 9. The Hosmer–Lemeshow test results are shown in Table 6. In the training cohort, the P values of the clinical, radiomics and combined models were 0.215, 0.998 and 0.249, respectively. In the test cohort, the corresponding P values were 0.003, 0.148 and 0.032. Thus, the radiomics model did not show significant miscalibration in the test cohort, whereas the clinical and combined models showed a certain degree of miscalibration, suggesting that the probability predictions of these models may be biased in independent samples; recalibration and external validation are required before clinical deployment of the combined model (see Discussion).

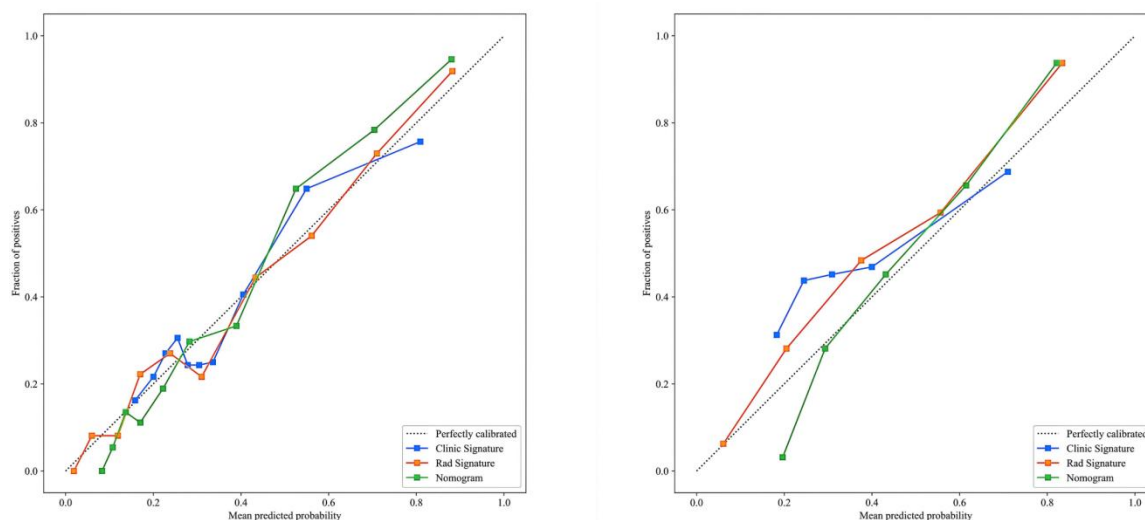


Fig. 9 Calibration curves of the clinical, radiomics and combined models in the training (left) and internal test (right) cohorts.

Table 6 Hosmer–Lemeshow goodness-of-fit test results

Cohort	Clinical model	Radiomics model	Nomogram
Training	0.215	0.998	0.249
Test	0.003	0.148	0.032

DCA is shown in Fig. 10. Within a certain range of threshold probabilities, the combined model provided a higher net benefit than the “treat-all”, “treat-none” strategies and the clinical model; its incremental clinical benefit over the radiomics model alone should be interpreted cautiously in relation to the specific threshold range.

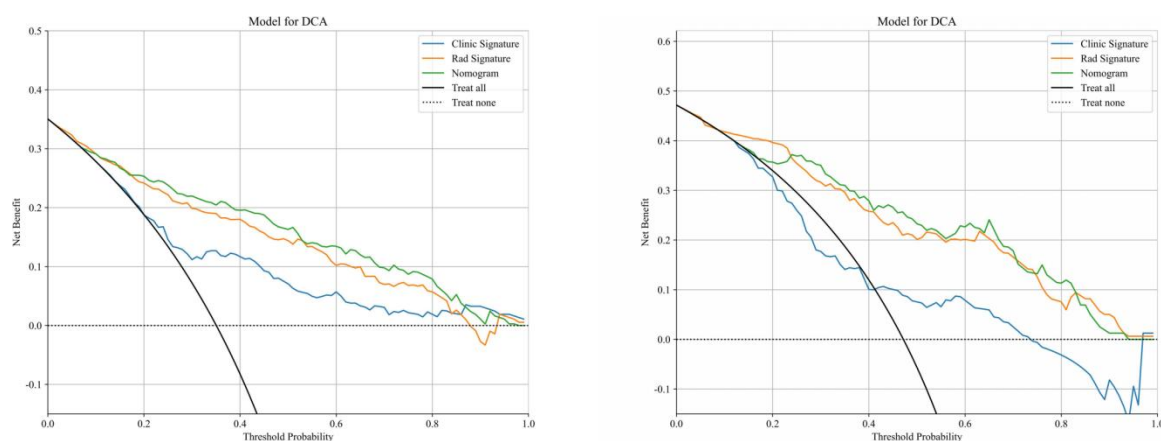


Fig. 10 Decision curve analysis of the clinical, radiomics and combined models in the training (left) and internal test (right) cohorts.

On the basis of the clinical model and the radiomics model, the clinical model prediction (clinical signature) and the Rad-score were integrated using logistic regression to construct

the combined clinical–radiomics prediction model, which was visualized as a nomogram (Fig. 11).

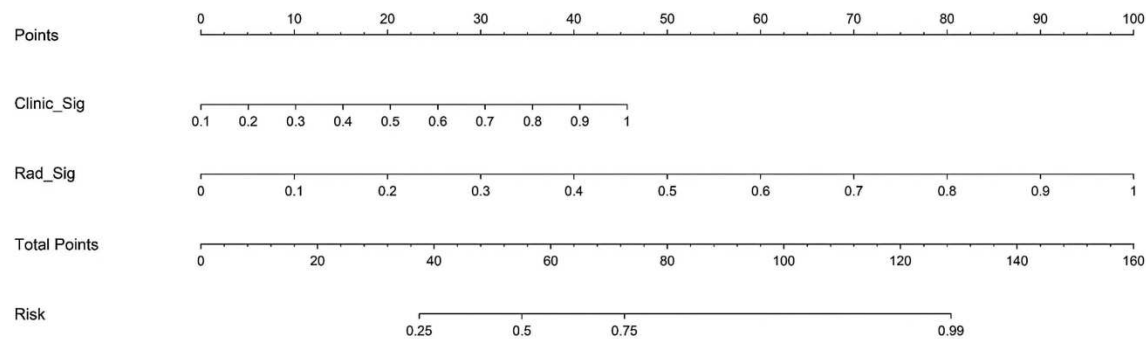


Fig. 11 Nomogram of the combined model based on the clinical model prediction (clinical signature) and the radiomics score (Rad-score).

Discussion

In this study, we developed and internally validated a combined machine-learning model integrating CT radiomics features and clinical indicators in a cohort of 527 children with MPP. The combined model achieved an AUC of 0.869 (95% CI: 0.815–0.923) in the internal test cohort, significantly outperforming the clinical model alone (AUC 0.638, $P = 1.59 \times 10^{-6}$), and provided a higher net benefit than the “treat-all”, “treat-none” strategies and the clinical model in decision curve analysis. RMPP is a severe form of pediatric MPP; according to the Evidence-Based Guideline for the Diagnosis and Treatment of Mycoplasma pneumoniae Pneumonia in Children (2023), RMPP is diagnosed when persistent fever and progressive clinical or radiological findings persist for at least 7 days despite standardized macrolide therapy [1]. The disease burden of RMPP has been increasing: a meta-analysis of 53 studies including 35,275 children reported an overall prevalence of pediatric RMPP of 37.8%, rising to 86.5% after the COVID-19 pandemic [6]. The proportion of RMPP in our cohort (38.7%) was highly consistent with these epidemiological data. Because the diagnosis of RMPP is often delayed beyond 7 days of illness, by which time pulmonary injury may have already progressed, risk stratification using information available early after admission is of considerable clinical importance [7].

Multiple clinical and laboratory risk factors for RMPP have been established. Meta-analyses have shown that fever lasting more than 10 days (OR 3.965), pleural effusion (OR 6.922), extrapulmonary complications (OR 17.762), pulmonary consolidation $\geq 2/3$ of a lobe (OR 8.245) and CRP > 40 mg/L (OR 4.975) are significantly associated with RMPP [8]; another meta-analysis of 28 studies including 6374 Chinese children further confirmed that elevated CRP, LDH, neutrophil percentage, IL-6, ESR, WBC, ALT and AST are risk factors for RMPP [9]. On the basis of these indicators, various clinical prediction tools have been developed, including a nomogram incorporating age, fever duration, lymphocyte count, D-dimer and pulmonary imaging score (development-cohort AUC 0.907) [12], a simple nomogram based on LDH, albumin, neutrophil ratio and high fever (AUC 0.884) [13], a prospectively validated early predictive scale (sensitivity 78.3%, specificity 86.2%) [14], and combined D-dimer/CRP/LDH models (AUC 0.841–0.893) [15–16]. However, such

models rely largely on threshold-based laboratory parameters and show variable performance across populations. The clinical model in our study achieved a test-cohort AUC of only 0.638, lower than previous reports, suggesting that clinical and laboratory indicators alone have limited capacity for early discrimination of RMPP. Notably, conventional CT signs (consolidation/atelectasis, pleural effusion, bronchial stenosis or obstruction, and multilobar distribution) did not differ significantly between the RMPP and NRMPP groups in our study, which is not entirely consistent with previous findings that pleural effusion and consolidation are risk factors [8]; this may be related to the high proportion of severe cases in our cohort and to the binary qualitative nature of conventional signs, which cannot reflect subtle intralesional heterogeneity. This observation provided a direct rationale for introducing quantitative radiomics.

Radiomics allows high-throughput extraction of gray-level, texture and shape features that are imperceptible to the human eye and has demonstrated unique value in disease prediction. In this study, 1834 radiomics features were extracted from CT images; after stability screening ($ICC \geq 0.85$), correlation-based redundancy removal and LASSO dimensionality reduction, 25 features were retained to construct the Rad-score, which achieved a test-cohort AUC of 0.851, markedly better than the clinical model. Recent CT-based radiomics and deep-learning approaches have also been applied to early RMPP prediction: a multicenter study constructed a clinical–imaging–radiomics integrated model with an AUC of 0.811 [17]; a transformer-based deep-learning framework achieved automated risk stratification (external test AUC 0.89) [18]; an AI-driven CT lesion quantification model reached an AUC of 0.897 [19]; and a model based on pulmonary vascular quantification parameters (BV5%, BV10%) achieved an AUC of 0.91 [20]. In addition, a lung ultrasound nomogram (AUC 0.955) [21] and bronchoscopy combined with CT scoring (AUC 0.82) [22] have been reported for RMPP prediction. Compared with end-to-end deep learning, the “manual segmentation + interpretable features + multi-classifier comparison” workflow used in our study is easier to reproduce and interpret in routine clinical settings. We selected LR from seven machine-learning classifiers (LR, SVM, KNN, RF, ExtraTrees, XGBoost and LightGBM) on the basis of cross-validated performance and model stability; it achieved a test-cohort AUC of 0.851, comparable to that of more complex models such as XGBoost reported previously [23], suggesting that after sufficient dimensionality reduction of radiomics features, a linear model can achieve robust performance while effectively avoiding the overfitting risk of complex models in small samples.

The combined model significantly outperformed the radiomics model alone in the training cohort (AUC 0.876 vs. 0.849, $P = 0.002$), but the difference was not statistically significant in the test cohort (0.869 vs. 0.851, $P = 0.060$), suggesting that the Rad-score already carries most of the predictive information and that the incremental contribution of clinical indicators was mainly observed in the training cohort, warranting confirmation in larger samples. After backward stepwise selection, the final clinical model included LDH, peak body temperature and ALT. Although age was statistically significant in the full multivariate model ($P = 0.031$), its aOR was close to 1 (1.095 per year) with limited incremental contribution, and it was not retained to preserve clinical practicality and model parsimony; although previous nomogram [12] and meta-analytic [6, 9] studies have

suggested older age as a risk factor for RMPP, its incremental predictive value has been inconsistent across populations, and its effect may be partly captured by severity indicators such as LDH. PCT showed borderline significance ($P = 0.077$) and was also not retained. The retained predictors are consistent with previous studies: LDH is the most widely recognized predictive biomarker of RMPP, and its diagnostic value has been supported by a systematic review (pooled sensitivity 0.75, specificity 0.73, AUC 0.85) [10] with a non-linear dose–response relationship with RMPP risk [24]; peak body temperature, as a clinical indicator of systemic inflammatory burden, has been incorporated into several prediction models [25–26]; and PCT and ALT have been confirmed as independent risk factors in meta-analyses of Chinese children [9]. D-dimer was significant on univariate analysis (OR 1.381) but lost statistical significance after multivariable adjustment, possibly related to collinearity with LDH and PCT and to missing-data handling; previous studies have confirmed that elevated D-dimer is closely associated with RMPP and its thrombotic complications [11, 27], and its weight in prediction models warrants further investigation.

From a translational perspective, our model may serve as a risk-stratification tool: early identification of children at high risk of RMPP after admission (when the index CT and routine laboratory tests are completed) could facilitate closer clinical monitoring and timely evaluation for treatment escalation. It should be emphasized, however, that a prediction model provides risk estimates rather than direct treatment recommendations; decisions on escalating therapy (e.g., second-line anti-*Mycoplasma* agents [28], corticosteroids [29–30], bronchoscopic alveolar lavage [31] or prophylactic anticoagulation when D-dimer is markedly elevated [1, 11]) should be made by clinicians in accordance with current guidelines and the individual patient's condition. Whether model-guided risk stratification can ultimately reduce the risk of severe complications such as plastic bronchitis [2], bronchiolitis obliterans [3], necrotizing pneumonia [4] and thromboembolism [5] warrants prospective interventional studies. Furthermore, the macrolide resistance rate of *M. pneumoniae* in China has remained high (approximately 81% over the past two decades) [1], and resistant strains are closely associated with the development of RMPP [32–33]; whether radiomics features can indirectly reflect resistance-related inflammatory–pathological phenotypes is a promising direction for future investigation.

This study has several limitations. First, the single-center retrospective design, with the training and test cohorts derived from the same center and the same CT scanner, lacks external validation, and the generalizability of the model requires confirmation in multicenter prospective studies. Second, the number of RMPP events in the training cohort (129) was relatively low relative to model complexity (event-to-parameter ratio approximately 5:1); although LASSO regularization partially mitigates overfitting, bootstrap resampling to report optimism-corrected performance is recommended. Third, the Hosmer–Lemeshow test indicated suboptimal calibration of the combined model in the test cohort ($P = 0.032$); recalibration or internal-validation correction is required before clinical use. Fourth, lesion segmentation is manual, time-consuming and subject to inter-observer variability, limiting large-scale application. Fifth, co-infection status was not incorporated as a predictive variable in this cohort; previous studies have shown that co-infections such as adenovirus are not uncommon in children with RMPP and are associated

with refractoriness [32], and the impact of co-infection on model performance and clinical interpretation warrants further evaluation. Sixth, the definition of RMPP varies across studies and guidelines (e.g., duration of macrolide therapy, fever thresholds), which may affect outcome ascertainment and cross-study comparisons [6]. Seventh, chest CT involves ionizing radiation; the indications for CT in children should be strictly controlled, and the combination of radiation-free ultrasound imaging [21] with radiomics deserves exploration. Eighth, because the cohort was split randomly rather than stratified by outcome, the proportion of RMPP differed between the training and test cohorts (35.1% vs. 47.2%); this imbalance may influence the optimal classification threshold and calibration estimates, and recalibration in the target population is recommended. In addition, model fairness across sociodemographic subgroups was not evaluated. When implementing the model, missing or poor-quality input data should be handled following the MICE framework described in the Methods, and use of the model requires a radiologist for lesion segmentation and clinicians for data entry.

Future studies should focus on multicenter external validation, deep-learning-based automated segmentation, integration of resistance genes and etiological information [34–35], and prospective interventional studies to determine whether model-guided stratified management can truly improve outcomes in children with RMPP.

Conclusions

A combined machine-learning model integrating CT radiomics features and early clinical indicators, visualized as a nomogram, showed good discrimination for predicting progression from MPP to RMPP in children (internal test AUC 0.869) and significantly outperformed the clinical model, providing an interpretable tool for early risk stratification. However, its calibration was suboptimal in the test cohort (Hosmer–Lemeshow $P = 0.032$), and external validation and model recalibration are required before clinical deployment.

List of abbreviations

ALT: alanine aminotransferase; aOR: adjusted odds ratio; AST: aspartate aminotransferase; AUC: area under the curve; CAP: community-acquired pneumonia; CI: confidence interval; CRP: C-reactive protein; CT: computed tomography; DCA: decision curve analysis; GLCM: gray-level co-occurrence matrix; GLDM: gray-level dependence matrix; GLRLM: gray-level run-length matrix; GLSZM: gray-level size zone matrix; ICC: intraclass correlation coefficient; KNN: k-nearest neighbors; LASSO: least absolute shrinkage and selection operator; LDH: lactate dehydrogenase; LightGBM: light gradient boosting machine; LR: logistic regression; MICE: multivariate imputation by chained equations; MPP: Mycoplasma pneumoniae pneumonia; NGTDM: neighboring gray-tone difference matrix; NRMPP: non-refractory Mycoplasma pneumoniae pneumonia; NPV: negative predictive value; OR: odds ratio; PCT: procalcitonin; PPV: positive predictive value; Rad-score: radiomics score; RF: random forest; RMPP: refractory Mycoplasma pneumoniae pneumonia; ROC: receiver operating characteristic; SVM: support vector machine; WBC: white blood cell; XGBoost: eXtreme gradient boosting.

Declarations

Ethics approval and consent to participate: This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Quanzhou Women's and Children's Hospital (approval No.: (2025) Ethics Review No. (39)). Owing to the retrospective design, the requirement for informed consent from participants and their guardians was waived by the ethics committee.

Clinical trial number: Not applicable (retrospective observational study).

Consent for publication: Not applicable.

Availability of data and materials: The datasets generated and/or analyzed during the current study are not publicly available owing to hospital data-sharing policy but are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: This work was supported by the 2025 Quanzhou Municipal Science and Technology Plan Project (grant number: 2025QZNY051). The funder had no role in the study design, data collection, analysis, interpretation, or writing of the manuscript.

Authors' contributions: HXQ conceived and designed the study and drafted the manuscript. HXQ, YY and CTF acquired, organized, analyzed and interpreted the data. HRS and WQ were responsible for funding acquisition, resources, supervision and project administration. All authors critically revised the manuscript for important intellectual content, approved the final version to be published, and agreed to be accountable for all aspects of the work.

Acknowledgements: Not applicable.

Study protocol and registration: No study protocol was prepared. This retrospective observational study was registered in the National Medical Research Registration and Filing Information System of China (registration No.: MR-35-25-045701); as a non-interventional retrospective study, it was not registered as a clinical trial.

Code availability: The analytical code is available from the corresponding author on reasonable request.

Patient and public involvement: No patient and public involvement was conducted.

References

1. The Subspecialty Group of Respiratory, Society of Pediatrics, Chinese Medical Association, China National Clinical Research Center of Respiratory Diseases, Editorial Board of Chinese Journal of Pediatrics. Evidence-based guideline for the diagnosis and treatment of Mycoplasma pneumoniae pneumonia in children (2023). *Pediatr Investig*. 2025;9(1):1–11. <https://doi.org/10.1002/ped4.12469>.

2. Zhao L, Zhang T, Cui X, et al. Development and validation of a nomogram to predict plastic bronchitis in children with refractory *Mycoplasma pneumoniae* pneumonia. *BMC Pulm Med*. 2022. <https://doi.org/10.1186/s12890-022-02047-2>.
3. Cheng Q, Zhang H, Shang Y, et al. Clinical features and risk factors analysis of bronchitis obliterans due to refractory *Mycoplasma pneumoniae* pneumonia in children: a nomogram prediction model. *BMC Infect Dis*. 2021. <https://doi.org/10.1186/s12879-021-06783-4>.
4. Li X, Zhao L, Cui X, et al. Development and validation of a nomogram for predicting necrotizing pneumonia in children with refractory *Mycoplasma pneumoniae* pneumonia. *Ital J Pediatr*. 2025. <https://doi.org/10.1186/s13052-025-02006-7>.
5. Fu Y, Zhang TQ, Dong CJ, et al. Clinical characteristics of 14 pediatric mycoplasma pneumoniae pneumonia associated thrombosis: a retrospective study. *BMC Cardiovasc Disord*. 2023. <https://doi.org/10.1186/s12872-022-03030-9>.
6. Huang X, Huang D, Wang W, et al. Prevalence, risk factors, and early prediction of refractory pneumonia caused by *Mycoplasma pneumoniae* in children: a systematic review and meta-analysis. *Eur J Pediatr*. 2026. <https://doi.org/10.1007/s00431-026-06802-z>.
7. Tong L, Huang S, Zheng C, et al. Refractory *Mycoplasma pneumoniae* pneumonia in children: early recognition and management. *J Clin Med*. 2022. <https://doi.org/10.3390/jcm11102824>.
8. Gong H, Sun B, Chen Y, et al. The risk factors of children acquiring refractory mycoplasma pneumoniae pneumonia: a meta-analysis. *Medicine (Baltimore)*. 2021. <https://doi.org/10.1097/MD.00000000000024894>.
9. Chen C, Chen S, Yang C, et al. Risk factors for refractory *Mycoplasma pneumoniae* in Chinese children: a meta-analysis. *Front Pediatr*. 2025. <https://doi.org/10.3389/fped.2025.1512689>.
10. Wang S, Jiang Z, Li X, et al. Diagnostic value of serum LDH in children with refractory *Mycoplasma pneumoniae* pneumonia: a systematic review and meta-analysis. *Front Pediatr*. 2023. <https://doi.org/10.3389/fped.2023.1094118>.
11. Huang X, Li D, Liu F, et al. Clinical significance of D-dimer levels in refractory *Mycoplasma pneumoniae* pneumonia. *BMC Infect Dis*. 2021. <https://doi.org/10.1186/s12879-020-05700-5>.
12. Li M, Wei X, Zhang SS, et al. Recognition of refractory *Mycoplasma pneumoniae* pneumonia among *Mycoplasma pneumoniae* pneumonia in hospitalized children: development and validation of a predictive nomogram model. *BMC Pulm Med*. 2023. <https://doi.org/10.1186/s12890-023-02684-1>.
13. Cheng S, Lin J, Zheng X, et al. Development and validation of a simple-to-use nomogram for predicting refractory *Mycoplasma pneumoniae* pneumonia in children. *Pediatr Pulmonol*. 2020. <https://doi.org/10.1002/ppul.24684>.
14. Bi Y, Zhu Y, Ma X, et al. Development of a scale for early prediction of refractory *Mycoplasma pneumoniae* pneumonia in hospitalized children. *Sci Rep*. 2021. <https://doi.org/10.1038/s41598-021-86086-5>.

15. Chen Q, Hu T, Wu L, et al. Clinical features and biomarkers for early prediction of refractory *Mycoplasma pneumoniae* pneumonia in children. *Emerg Med Int.* 2024. <https://doi.org/10.1155/2024/9328177>.
16. Shen F, Dong C, Zhang T, et al. Development of a nomogram for predicting refractory *Mycoplasma pneumoniae* pneumonia in children. *Front Pediatr.* 2022. <https://doi.org/10.3389/fped.2022.813614>.
17. Li Q, Zhang J, Song ZJ, et al. Early identification of refractory *Mycoplasma pneumoniae* pneumonia in children using CT-based radiomics: a multicenter study. *Front Med (Lausanne).* 2026. <https://doi.org/10.3389/fmed.2026.1785817>.
18. Ying Z, Hu G, Li J, et al. A CT-based deep learning model for the automated risk stratification of refractory *Mycoplasma pneumoniae* pneumonia in children. *BMC Med Imaging.* 2026. <https://doi.org/10.1186/s12880-026-02590-y>.
19. Qian Y, Tao Y, Wu L, et al. Model based on the automated AI-driven CT quantification is effective for the diagnosis of refractory *Mycoplasma pneumoniae* pneumonia. *Sci Rep.* 2024. <https://doi.org/10.1038/s41598-024-67255-8>.
20. Ying Z, Li Z, Li J, et al. Quantitative analysis of pulmonary vascular alterations in children with refractory *Mycoplasma pneumoniae* pneumonia. *Quant Imaging Med Surg.* 2025. <https://doi.org/10.21037/qims-2025-1008>.
21. Liu G, Wang G, Yang Z, et al. A lung ultrasound-based nomogram for the prediction of refractory *Mycoplasma pneumoniae* pneumonia in hospitalized children. *Infect Drug Resist.* 2022. <https://doi.org/10.2147/IDR.S387890>.
22. Lu W, Wu X, Xu Y, et al. Predictive value of bronchoscopy combined with CT score for refractory *mycoplasma pneumoniae* pneumonia in children. *BMC Pulm Med.* 2024. <https://doi.org/10.1186/s12890-024-02996-w>.
23. Jiang Y, Wang X, Li L, et al. Predicting and interpreting key features of refractory *Mycoplasma pneumoniae* pneumonia using multiple machine learning methods. *Sci Rep.* 2025. <https://doi.org/10.1038/s41598-025-02962-4>.
24. Zheng XX, Lin JL, Dai JH. Value of lactate dehydrogenase in predicting refractory *Mycoplasma pneumoniae* pneumonia in children: an evaluation based on decision curve analysis and dose-response analysis. *Zhongguo Dang Dai Er Ke Za Zhi.* 2020. <https://doi.org/10.7499/j.issn.1008-8830.2020.02.006>.
25. Xu D, Zhang AL, Zheng JS, et al. A multicenter prospective study on early identification of refractory *Mycoplasma pneumoniae* pneumonia in children. *Zhonghua Er Ke Za Zhi.* 2024. <https://doi.org/10.3760/cma.j.cn112140-20231121-00383>.
26. Fan F, Jiang F, Lv J, et al. Development of a nomogram for predicting refractory *Mycoplasma pneumoniae* pneumonia in children: a prospective study. *Transl Pediatr.* 2025. <https://doi.org/10.21037/tp-2025-407>.
27. Yue Y, Lian T, Kang L, et al. D-dimer serves as predictor of plastic bronchitis or necrotizing pneumonia in children with *Mycoplasma pneumoniae* pneumonia. *Front Pediatr.* 2025. <https://doi.org/10.3389/fped.2025.1604253>.
28. Ahn JG, Cho HK, Li D, et al. Efficacy of tetracyclines and fluoroquinolones for the treatment of macrolide-refractory *Mycoplasma pneumoniae* pneumonia in children:

a systematic review and meta-analysis. BMC Infect Dis. 2021.
<https://doi.org/10.1186/s12879-021-06508-7>.

29. Qiu JL, Huang L, Shao MY, et al. Efficacy and safety of azithromycin combined with glucocorticoid on refractory *Mycoplasma pneumoniae* pneumonia in children: a PRISMA-compliant systematic review and meta-analysis. Medicine (Baltimore). 2020. <https://doi.org/10.1097/MD.00000000000020121>.
30. Valentino MS, Pagliani C, Lovati C, et al. Taming the inflammation: the role of corticosteroids in pediatric *Mycoplasma pneumoniae*. Children (Basel). 2026. <https://doi.org/10.3390/children13030333>.
31. Yan Q, Niu W, Jiang W, et al. Risk factors for delayed radiographic resolution in children with refractory *Mycoplasma pneumoniae* pneumonia. J Int Med Res. 2021. <https://doi.org/10.1177/03000605211015579>.
32. Zhou Y, Wang J, Chen W, et al. Impact of viral coinfection and macrolide-resistant mycoplasma infection in children with refractory *Mycoplasma pneumoniae* pneumonia. BMC Infect Dis. 2020. <https://doi.org/10.1186/s12879-020-05356-1>.
33. Li G, Zhang L, Huang Y, et al. Clinical characteristics and outcomes of macrolide-resistant *Mycoplasma pneumoniae* infection in hospitalized children: a single-center retrospective cohort study in Chengdu, China. Transl Pediatr. 2025. <https://doi.org/10.21037/tp-2025-290>.
34. Zhan XW, Deng LP, Wang ZY, et al. Correlation between *Mycoplasma pneumoniae* drug resistance and clinical characteristics in bronchoalveolar lavage fluid of children with refractory *Mycoplasma pneumoniae* pneumonia. Ital J Pediatr. 2022. <https://doi.org/10.1186/s13052-022-01376-6>.
35. Xu C, Deng H, Zhang J, et al. Mutations in domain V of *Mycoplasma pneumoniae* 23S rRNA and clinical characteristics of pediatric *M. pneumoniae* pneumonia in Nanjing, China. J Int Med Res. 2021. <https://doi.org/10.1177/03000605211016376>.