Infection and Drug Resistance

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Clinical Diagnostic Prediction Model For Mycoplasma pneumoniae Pneumonia in Children

Xiaohua Liu[1,*], Hongfei Du[2,*], Linyan Zhang[1,*], Yang Yu[1], Guixiang Wang[1], Fan Zhong[1], Mingjun Zhang[3], Huacui Huang[1], Ying Xu[2]

[1]Department of Medical Laboratory, Xindu District People's Hospital of Chengdu, Chengdu, 610500, People's Republic of China; [2]Department of Clinical Laboratory, the First Affiliated Hospital of Chengdu Medical College, Chengdu, 610500, People's Republic of China; [3]Department of Medical Laboratory, Jiulongpo District People's Hospital of Chongqing, Chongqing, 400050, People's Republic of China

*These authors contributed equally to this work

Correspondence: Ying Xu, Department of Clinical Laboratory, the First Affiliated Hospital of Chengdu Medical College, No. 278, Baoguang Road, Xindu District, Chengdu, Sichuan, 610500, People's Republic of China, Email yingxu825@126.com; Huacui Huang, Department of Clinical Laboratory, Xindu District People's Hospital of Chengdu, No. 199, Yuying Road South, Xindu District, Chengdu, Sichuan, 610500, People's Republic of China, Email 839373273@qq.com

Background: The aim of this research is to ascertain the risk determinants associated with Mycoplasma pneumoniae pneumonia (MPP) in pediatric patients diagnosed with community-acquired pneumonia (CAP), as well as to construct predictive models to forecast the incidence of MPP.

Methods: This study was conducted at Xindu District People's Hospital of Chengdu from August 2023 to March 2024. A total of 1030 children aged 0 to 14 years with CAP were enrolled and divided into MPP (n=414) and non-MPP (NMPP, n=616) groups based on diagnostic criteria including MP antibody and MP RNA. Data were collected within 24 hours of admission, including peripheral blood counts, inflammatory markers, and other biochemical parameters. The Logistic+Stepwise, Logistic+Lasso, Logistic+Elastic-net, and Logistic+Ridge were employed to identify risk factors, and were used for variable selection with penalization algorithms. Model performance was evaluated using C-index, sensitivity, specificity, accuracy, recall, and F1 score.

Results: The results of prediction model showed that four models had good performance. The area under the ROC curve revealed good predictive ability (AUC > 0.8 in both Logistic model and Experience model), the results of calibration curves indicated a good consistency consistency. Logistic+Lasso model selected 9 key variables for further analysis.

Conclusion: We have developed and validated a clinical prediction model in children with Mycoplasma pneumoniae pneumonia (MPP). The model identifies NEUT%, EOS%, HSCRP, ADA, Crea, Urea, HDL, P, and ESR as significant independent predictors. It demonstrated robust discriminative ability and good calibration, offering a practical tool for clinicians to stratify risk and guide early intervention in pediatric patients.

Keywords: mycoplasma pneumoniae pneumonia, prediction model, clinical characteristics, children

Introduction

Mycoplasma pneumoniae (MP) is recognized as a prevalent infectious agent responsible for acute respiratory tract infections among children globally.[1] In pediatric populations older than five years, MP accounts for approximately 40% of cases of community-acquired pneumonia (CAP).[2,3] The disease often presents with fever, cough, shortness of breath and other clinical manifestations.[4,5] Accurate diagnosis is crucial for clinical management. Clinicians can ensure timely initiation of targeted treatment—such as macrolide or tetracycline antibiotics—through precise diagnosis. This approach shortens the disease course, reduces the risk of complications, and improves patient prognosis.[6] Currently, there is a scarcity of targeted diagnostic approaches for MP. The clinical manifestations and radiological characteristics associated with MP lack sufficient diagnostic specificity.[7] Presently, the three primary diagnostic techniques employed for MP include culture tests, polymerase chain reaction (PCR), and serological tests; however, each of these methods presents various limitations.[8–12] Culturing methods tend to be labor-intensive and exhibit reduced sensitivity when applied in clinical settings.[13,14] The use of PCR and serological testing aimed at respiratory

Received: 23 June 2025
Accepted: 8 October 2025
Published: 17 October 2025

samples is advocated as essential diagnostic tools for clinical decision-making, as outlined in the MPP clinical practice guidelines across various nations.[11,12,15] Nonetheless, the implementation of these tests is constrained by specific technical requirements and parameters governing the use of molecular assays, particularly in resource-limited regions and local healthcare facilities. Furthermore, antibodies to MP may not appear until 2 weeks following the onset of symptoms, restricting the utility of serological tests in early diagnosis.[16] Therefore, there is a pressing need for reliable predictive models that can assist clinicians in identifying children at high risk for MPP at the time of hospital admission. This multi-model approach is a significant advancement over previous studies, which often relied on single-model analyses and may not have captured the full spectrum of risk factor. Recent studies have indicated that hematological parameters derived from blood cell assessments possess significant prognostic value concerning infections, malignancies, and a variety of other health conditions.[17–19]

Therefore, it is crucial to examine the variations in laboratory test results between children diagnosed with MPP and those with NMPP. The objective of this study is to identify the risk factors for MPP in children with CAP, and to create predictive models that enable rapid and precise identification of patients in clinical practice.

Methods

Study Population

We included 1030 children with CAP aged 0 to 14 years diagnosed from August 2023 to March 2024 at the People's Hospital of Xindu District, Chengdu. Based on the diagnostic criteria for MP, 1030 patients were divided into the MPP group (n = 414) and NMPP group (n=616). All patients included in the study were free from any additional underlying conditions apart from MPP and NMPP.

Diagnostic Inclusion Criteria of MPP

The diagnosis of MPP was made based on the following definitions:[15] (1) single serum MP antibody titer >1:160 (PA method); 4-fold or greater rise in double serum MP antibody titer during the course of the disease; MP -DNA or RNA positivity. This study uniformly used MP-RNA detection as the exclusive microbiological criterion for patient enrollment to ensure methodological consistency.(2) Patients who fulfilled the diagnostic criteria for CAP, presenting with respiratory tract infection symptoms and chest radiographs indicative of pneumonia—either with or without pleural effusion—yet excluded cases of pulmonary infections attributable to alternative pathogens. In our investigation, 414 cases satisfied both criteria (1) and (2), the positivity rate of MP antibody titer >1:160 and MP-RNA was 4.35% and 95.65%, respectively, thereby categorizing them into the MPP group. Conversely, an additional 616 cases merely met the diagnostic benchmarks for CAP, with a single serum test revealing a Mycoplasma pneumoniae antibody (MP-Ab) titer lower than 1:160, and were thus classified into the NMPP group.(3) The included patients had no other underlying disease other than MPP and NMPP.

Exclusion Criteria

The criteria for exclusion are delineated as follows: (1) Non-fulfillment of the inclusion criteria or inadequacies in the clinical data; (2) Diagnosis of an immunodeficiency disorder; (3) Presence of respiratory ailments (Figure 1), which encompass conditions such as primary ciliary dyskinesia, cystic fibrosis, congenital bronchial ring malformation, pulmonary dysplasia, vascular bronchial foreign body, asthma, pulmonary tuberculosis, pulmonary neoplasms, and non-infectious interstitial lung disease.

Data Collection

All pediatric patients were admitted to the hospital within 24 hours following standard infection screening, which encompassed the evaluation of peripheral white blood cell (WBC) counts, the percentage of neutrophils (N%), and the percentage of eosinophils (EOS%), high-sensitivity c-reactive protein (HSCRP), erythrocyte sedimentation rate (ESR), adenosine deaminase (ADA), high density lipoprotein (HDL), creatinine (Crea) and coagulation indexes et al. The more detailed information were listed in supplementary Table 1. The general information encompassing the patient's sex, age, medical history, along with an evaluation of their current condition, was taken into account.

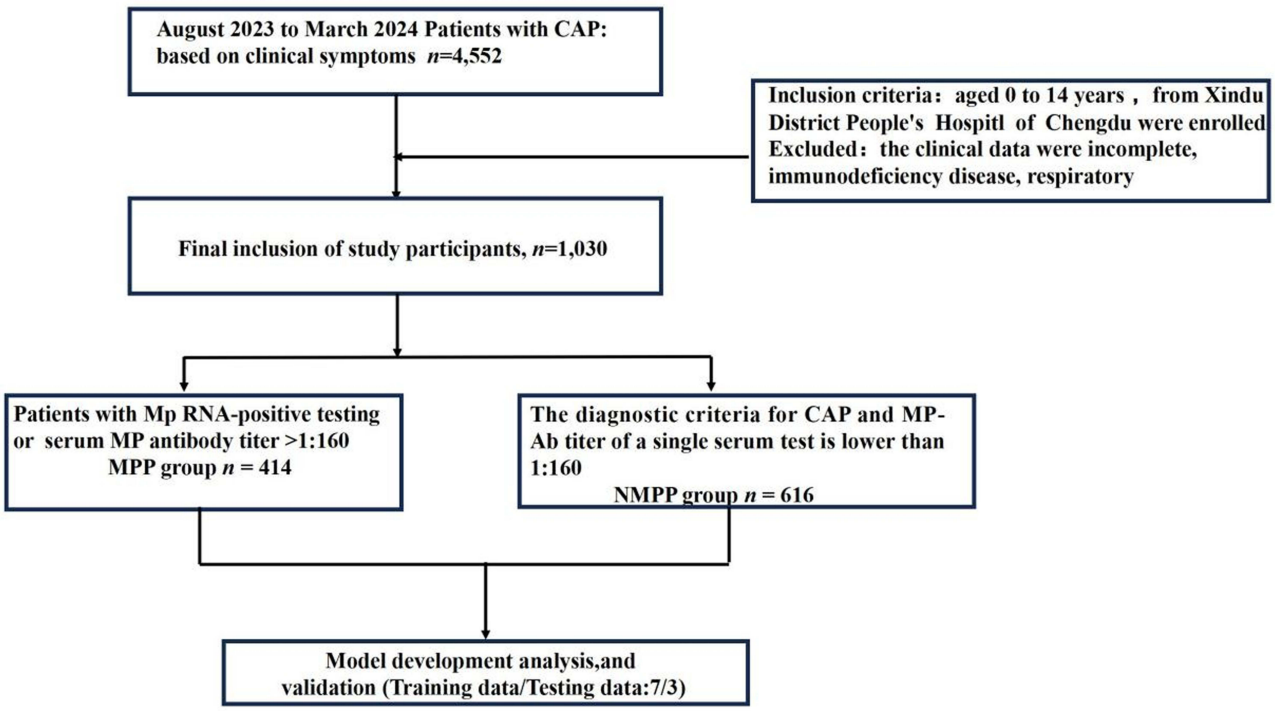

Figure 1 Study flowchart.

Statistical Analysis

This study performed four models to explore the risk factor of MPP. MPP was used as the response variables. To address the issue of complexity of high-dimensional clinical data, penalty algorithms are used for variable selection, including Logistic+Stepwise, Logistic+Lasso (least absolute shrinkage and selection operator), Logistic+Elastic-net, Logistic +Ridge. In contrast to conventional regression methods, penalized regression techniques possess the capability to manage an extensive array of potential predictors, thereby identifying the variables that exhibit the strongest association with the disease. Evaluation metrics C-index, Sensitivity, Specificity, Precision, Recall, and F1 score were used to assess the discriminatory power and accuracy of the model. The construction of the final risk prediction model was tailored based on model performance index and refined through the expert judgments of medical professionals. This approach ensures that the model's predictions are locally relevant and clinically actionable.

Missing values were imputed by the random forest method. In descriptive analysis were according to Shapiro–Wilk test for normality, normally distributed data were expressed as means ± standard deviation, and skewed distribution data were reported as medians with interquartile ranges 25th-75th (IQR). Qualitative data were represented as proportions, and the comparison was conducted using either the Chi-square test or Fisher's exact test, depending on the underlying statistical assumptions. Missing data points were addressed through the application of the random forest imputation technique.

Supplementary Figure 1 illustrates the kernel density distribution plots for the dataset subsequent to imputation. A comparative analysis of the data distributions, both prior to and following imputation (Supplementary Table 2 shows the specific missing data status of each variable), revealed a marked consistency, suggesting that the imputed dataset in the current investigation exhibited a level of stability akin to that of the original dataset. All statistical tests employed were two-tailed, with the threshold for significance established at 0.05. Data cleaning and analysis were conducted using SAS 9.4 (version 9.4, SAS Institute, Inc., Cary, NC, USA) and R (version 4.3.4).

Results

Description Analysis

From August 2023 to March 2024, a total of 1030 patients were include in our study, the average age of the patients in this study was 3.92 ± 3.09 years, with 556 males (53.98%) and 474 females (46.02%).

The demographic and hemodynamic characteristics of the two groups were compared and are displayed in Table 1. Furthermore, there were significant differences in the laboratory parameters between the two groups, with a p-value of less than 0.05. Some of the remaining indicators were not statistically significant between the two groups ($P > 0.05$), such as EOS% (Z=-0.03, P=0.974), HGB (Z=-0.48, P=0.629), SAA (Z=-1.64, P=0.100), Urea (Z=-0.28, P=0.777), DDimer (Z=-0.95, P=0.340), HDL (t=1.63, P=0.104) and HSCRP (t=-1.12, P=0.261).

Table 1 Analysis of Differences in Indicators of MPP and NMPP Group

Variable	Overall	MPP (n=414)	NMPP (n=616)	Statistics	P
Gender				χ^2=0.50	0.481
Male	556(53.98)	229(41.19)	327(58.81)		
Female	474(46.02)	185(39.03)	289(60.97)		
Age	3.00[1.00,6.00]	5.00[3.00,7.00]	2.00[1.00,4.00]	Z=12.33	<0.001
WBC	7.21[5.47,9.65]	6.86[5.36,8.98]	7.55[5.53,9.96]	Z=-2.34	0.019
NEUT%	50.80[36.70,62.60]	58.10[48.40,66.30]	43.30[30.90,57.10]	Z=10.57	<0.001
LYM%	39.50[27.30,53.10]	31.95[24.00,41.80]	46.55[32.35,58.45]	Z=-10.60	<0.001
EOS%	1.70[0.70,3.00]	1.80[0.70,3.10]	1.70[0.70,3.00]	Z=-0.03	0.974
BASO%	0.20[0.10,0.30]	0.20[0.10,0.40]	0.20[0.10,0.30]	Z=2.86	0.004
HGB	124.30±14.79	124.03±15.10	124.49±14.58	Z=-0.48	0.629
SAA	2.42[0.00,22.57]	2.00[0.00,21.98]	2.50[0.00,22.57]	Z=-1.64	0.100
ADA	16.00[13.00,19.10]	17.00[13.80,20.10]	15.30[12.45,18.05]	Z=5.43	<0.001
AFU	27.00[21.00,32.00]	24.00[19.00,29.00]	29.00[23.00,36.00]	Z=-9.26	<0.001
A/G	17.25[13.50,22.00]	17.80[13.30,22.90]	17.00[13.70,21.25]	Z=2.16	0.031
ALP	181.00[142.00,228.00]	172.00[135.00,210.00]	188.50[150.00,241.50]	Z=-5.21	<0.001
ALT	20.58±27.56	18.43±15.26	22.03±33.31	t=-2.06	0.039
ApoA	0.13±0.38	0.16±0.41	0.11±0.36	t=2.18	0.029
ApoB	0.08±0.25	0.10±0.27	0.07±0.24	t=2.14	0.032
AST	39.08±34.24	33.86±14.92	42.59±42.21	t=-4.04	<0.001
CK	128.41±146.93	116.57±93.90	136.36±173.33	t=-2.12	0.034
CK-MB	21.36±16.33	18.24±14.29	23.45±17.26	t=-5.08	<0.001
CYSC	0.83±0.32	0.76±0.21	0.88±0.37	Z=-6.01	<0.001
Crea	31.75[25.40,40.80]	37.40[30.40,46.40]	28.85[23.90,35.40]	Z=10.74	<0.001
Urea	3.30[2.60,4.10]	3.30[2.60,4.00]	3.30[2.50,4.20]	Z=-0.28	0.777
DBIL	1.83±1.80	1.72±1.01	1.90±2.18	Z=2.87	0.004
DDimer	0.51±0.69	0.44±0.38	0.56±0.84	Z=-0.95	0.340
ESR	27.00[15.00,44.00]	35.50[22.00,49.00]	23.00[11.00,41.00]	Z=7.91	<0.001
FIB	3.55[2.91,4.30]	3.80[3.33,4.41]	3.33[2.70,4.09]	Z=7.46	<0.001
GFR	100.47[86.25,115.63]	105.48[92.80,118.91]	97.02[78.42,111.04]	Z=6.63	<0.001
GLB	27.04±6.13	28.91±5.37	25.79±6.30	t=8.28	<0.001
HDL	0.15±0.42	0.17±0.43	0.13±0.41	t=1.63	0.104
HSCRP	7.80±17.51	7.05±13.99	8.31±19.52	t=-1.12	0.261
AST	35.00[28.00,45.00]	32.00[26.00,38.00]	39.00[29.50,49.00]	Z=-8.03	<0.001
IBIL	8.38±17.35	6.70±3.66	9.51±22.17	Z=2.58	0.010
K	4.15±0.90	4.03±0.83	4.24±0.93	t=-3.76	<0.001
LDH	265.00[220.00,310.00]	251.00[208.00,294.00]	275.50[229.00,320.00]	Z=-5.73	<0.001
P	2.03[1.47,2.36]	2.01[1.39,2.34]	2.03[1.52,2.39]	Z=-3.34	<0.001

Note: p-values <0.05 indicate statistically significant differences.

Prediction Model Built Based on Logistic Regression Models

Conventional Logistic Regression Model

Initially, we employed a conventional experience regression framework incorporating all variables to develop a Logistic regression model (Supplementary Table 3), which entry P<0.05, The Logistic-Experience model included the following variables: Creatine kinase (CK), Apolipperotein B(APOB), Apolipperotein A(APOA), Indirect bilirubin(IBIL), Alanine aminotransferase(ALT)et al, subsequently applying a stepwise method for variable selection. The identified variables serve as a reference in the construction of the ultimate risk prediction model. Among the 36 variables, 15 variables had *P*-values exceeding 0.05, while the remaining variables all had *P*-values below 0.05. The 21 variable were selected for further analysis (Table 2 and Figure 2).

Table 2 Logistic Regression Model to Predict Recurrence Based on Stepwise Regression

Variable	β	SE	Waldχ^2	P
(Intercept)	−2.136	1.201	−1.779	0.075
Age	0.187	0.032	5.916	<0.001
NEUT%	0.023	0.013	1.845	0.065
LYM%	−0.028	0.012	−2.375	0.018
EOS%	0.076	0.042	1.817	0.069
NEUT Count	−0.102	0.050	−2.063	0.039
MCV	0.036	0.016	2.187	0.029
RDWSD	−0.092	0.037	−2.454	0.014
RDWCV	0.189	0.087	2.157	0.031
PLT	0.002	0.001	2.069	0.039
MPV	−0.352	0.181	−1.952	0.051
PLCR	0.055	0.027	2.046	0.041
HSCRP	−0.019	0.006	−2.982	0.003
TBIL	0.257	0.170	1.517	0.129
IBIL	−0.323	0.204	−1.587	0.112
AST	−0.010	0.006	−1.626	0.104
GGT	0.006	0.003	1.743	0.081
ADA	0.138	0.020	6.869	<0.001
AFU	−0.031	0.010	−3.027	0.002
Crea	0.024	0.009	2.544	0.011
Urea	−0.268	0.085	−3.148	0.002
GFR	0.012	0.004	3.206	0.001
GLU	−0.117	0.078	−1.505	0.132
CHOL	−0.396	0.269	−1.473	0.141
HDL	−1.101	0.817	−1.347	0.178
APOA	2.247	1.169	1.922	0.055
LDH	0.012	0.005	2.461	0.014
aHBDH	−0.020	0.007	−2.772	0.006
K	0.441	0.215	2.054	0.040
Na	−0.040	0.013	−2.977	0.003
Ca	0.020	0.011	1.902	0.057
P	−0.792	0.460	−1.722	0.085
CO2	0.126	0.051	2.450	0.014
AG	0.093	0.029	3.197	0.001
ESR	0.014	0.003	4.174	<0.001
DDimer	−0.283	0.178	−1.587	0.113
SAA	−0.002	0.001	−1.651	0.099

Note: p-values <0.05 indicate statistically significant differences.

https://doi.org/10.2147/IDR.S541396

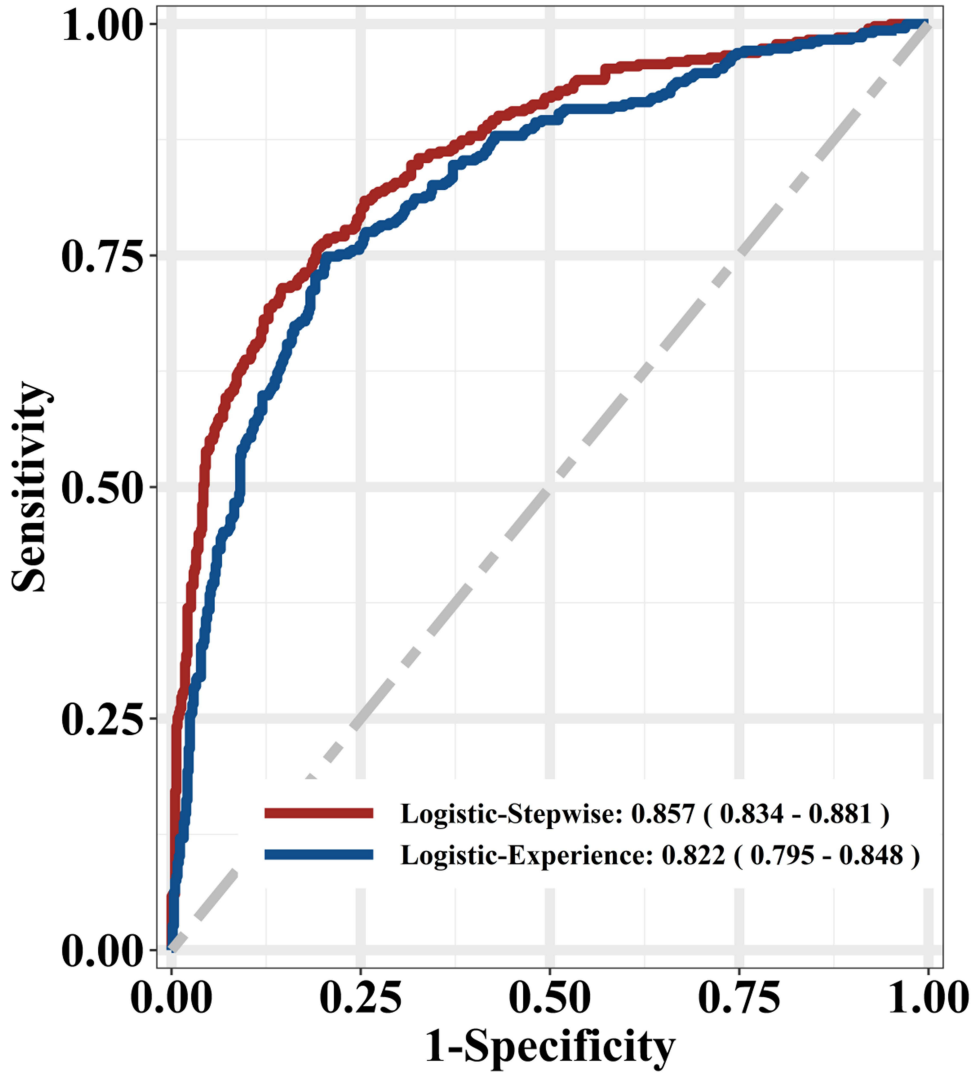

Figure 2 Comparison of ROC curves between Logistic-Experience model and Logistic-Stepwise model.

Prediction Model Built Based on Penalty Algorithms

A total of 414 patients were randomly allocated into two groups based on the outcome variable, with 70% designated as the training set and 30% as the validation set. To determine the optimal penalty term, we employed a ten-fold cross-validation. C-index and exponential loss of the binomial deviance loss were used to assess the performance of the model when determining the optimal number of variables.

Variable Selection

For Logistic-lasso model, the lambda (λ) and log(λ) = - 3.662 (λ = <0.001217771) when the AUC error of the model is minimized. Lambda (λ) and log(λ) = - 3.106 (λ = <0.001408422) when the binomial deviance of the model is minimized, and 9 variables were selected for further logistic regression analysis. The number of variables selected by the Logistic-elastic net model were similar to those in the lasso model. Given the preference for a more parsimonious model in this study, we opted for the lasso model that assesses binomial deviance loss (Figure 3).

Model Comparison of Penalty Regression Model

All performance metrics suggested that the penalty - based models outperformed others. DeLong's test results indicated that there were no statistically significant differences in the C - index values among the Lasso, Ridge, and Elastic - net logistic regression models. Considering the characteristics and performance of these models comprehensively, a set of variables,

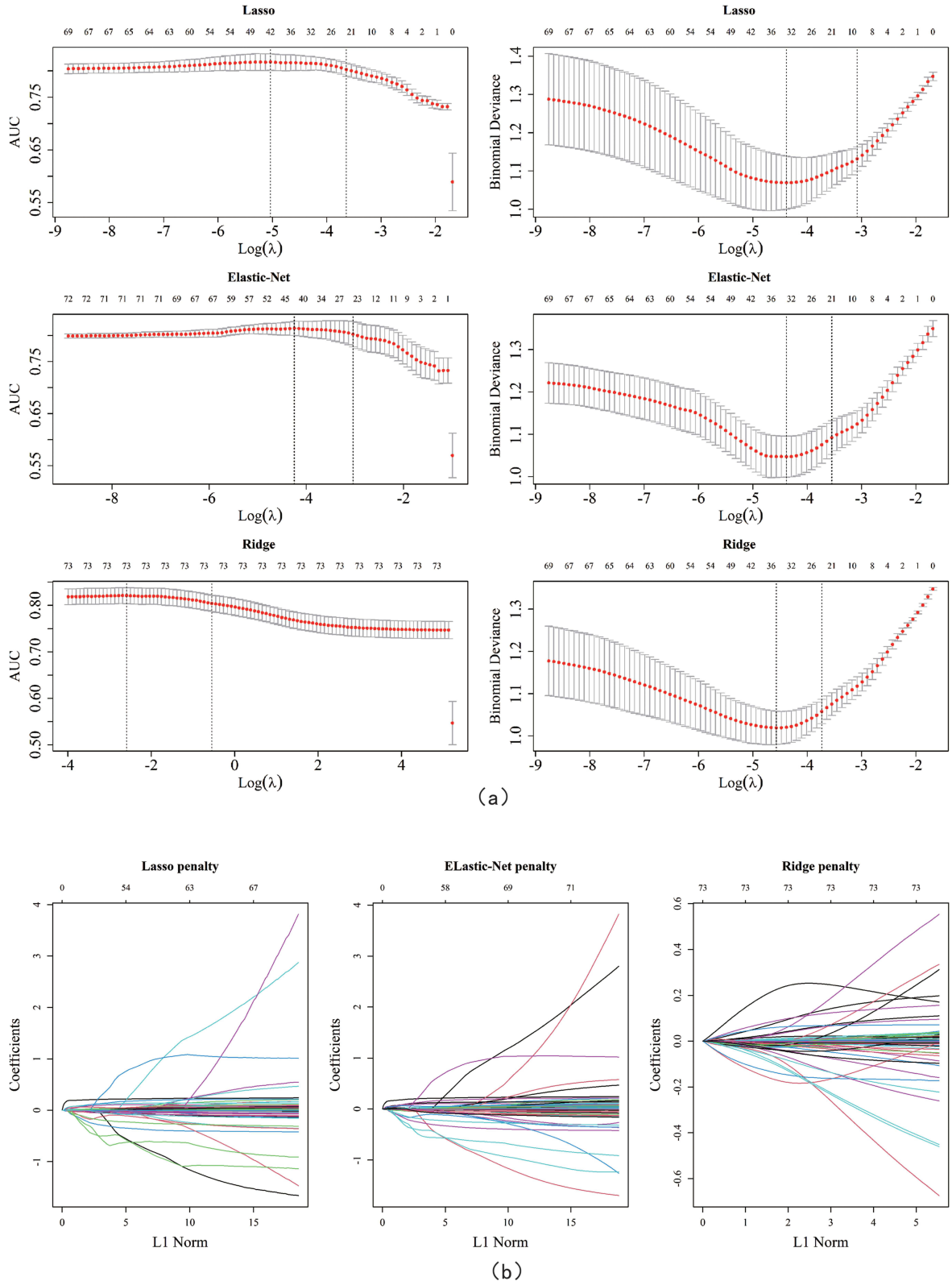

Figure 3 Screening of variables based on Lasso/Ridge/Elastic-net regression. (**a**) The variation characteristics of the coefficient of variables; (**b**) the selection process of the optimum value of the parameter λ in the penalty regression model by cross-validation method.

https://doi.org/10.2147/IDR.S541396

Table 3 Comparison of Model Performance Index

Model	C-index (95%CI)	Sensitive	Specificity	Precision	Recall	F1
Logistic+Lasso	0.864(0.836~0.892)	0.686	0.882	0.796	0.691	0.737
Logistic+Ridge	0.862(0.834~0.890)	0.676	0.884	0.799	0.645	0.738
Logistic+Elastic-net	0.861(0.833~0.899)	0.672	0.884	0.796	0.672	0.729
Logistic-Stepwise	0.857(0.834~0.881)	0.698	0.864	0.775	0.698	0.734
Logistic-Experience	0.822(0.795~0.848)	0.655	0.843	0.736	0.655	0.693

Note: Logistic-Experience model including CK, APOB, APOA, IBIL, ALT, P, A/G, DDimer, ADA, AST, Age, CKMB, CYSC, AFU, Crea, ESR, FIB, GFR, GLB, K, LDH, ALP.

including Age, NEUT%, EOS%, hs - CRP, ADA, Crea, Urea, HDL, P, and ESR, were ultimately chosen to construct the risk prediction model, which guided the development of the final risk prediction framework (Table 3 and Figure 4). Figure 4 contains two receiver - operating characteristic (ROC) curves, labeled as (a) and (b). These curves are used to illustrate the performance of the three regularization techniques, namely Lasso, Ridge, and Elastic - Net, in binary classification models. In Figure 4a, the area under the curve (AUC) for Lasso is 0.864 (95% confidence interval [CI]: 0.836–0.892), for Ridge is 0.862 (95% CI: 0.834–0.890), and for Elastic - Net is 0.861 (95% CI: 0.833–0.889), with Lasso showing relatively superior performance. In Figure 4b, the AUC of Lasso is 0.807 (95% CI: 0.755–0.858), that of Ridge is 0.812 (95% CI: 0.761–0.863), and that of Elastic - Net is 0.807 (95% CI: 0.756–0.859). These ROC curves offer a visual and quantitative assessment of how effectively each regularization method can differentiate between the two classes in the binary classification task.

Calibration and Clinical Application Based on Lasso Regression

Figure 5a and b depict the mapping relationship between the predicted probability (Predicted Pr[Y = 1]) and the actual probability (Actual Probability). The Apparent curve characterizes the predictive performance of the model on the original training dataset, intuitively presenting the model's performance without bias correction. The Bias - corrected

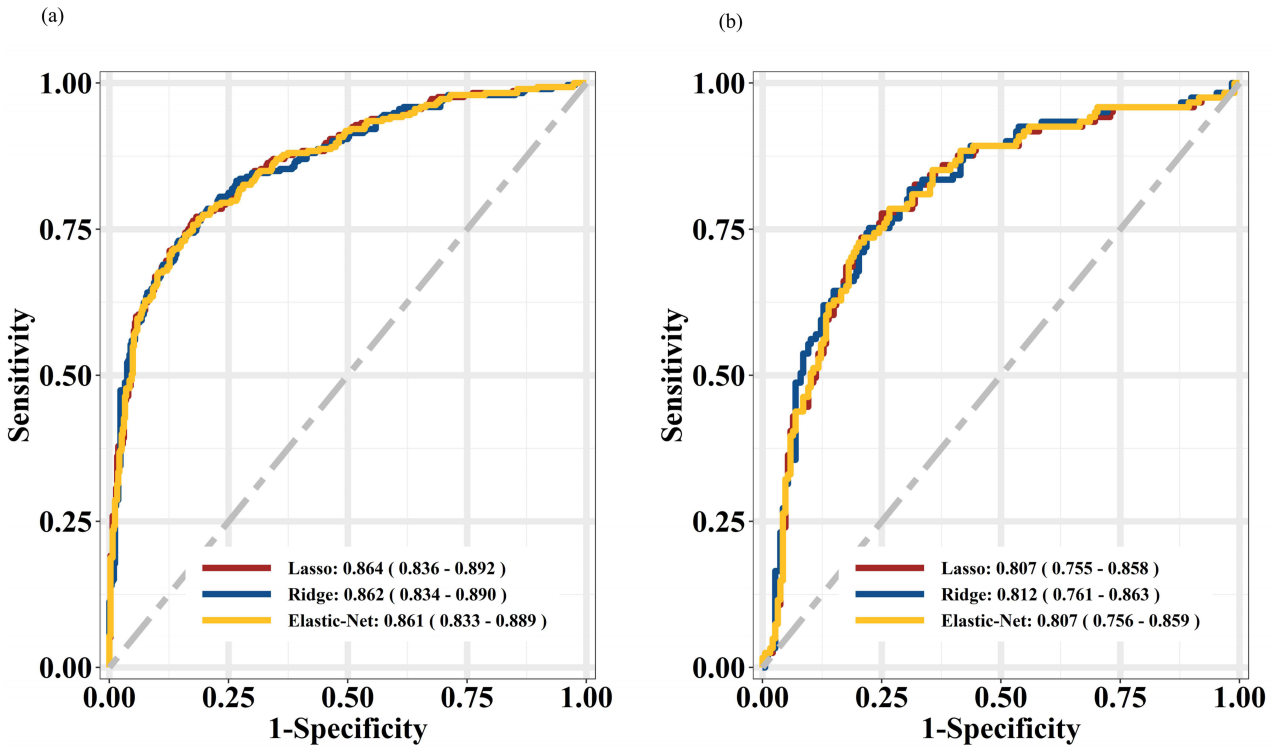

Figure 4 Comparison of ROC curves between training data (**a**) and testing data (**b**).

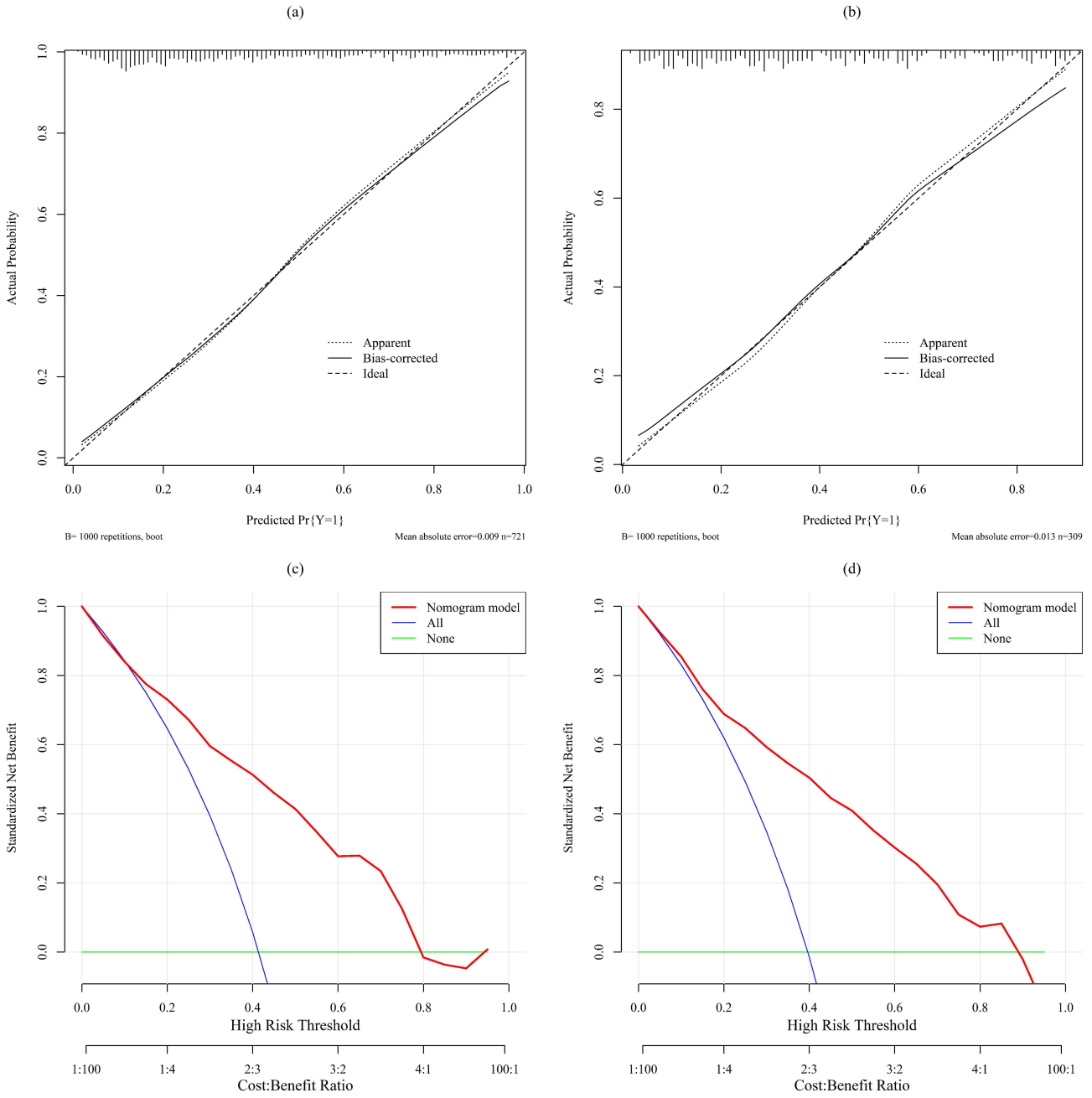

Figure 5 Decision curve analysis and calibration curve plots of predicted MPP based on penalty Logistic regression modeling in the training set and validation set (**a, c** for training data and **b, d** for testing data).

curve is the result obtained after applying statistical adjustment techniques to correct the inherent bias of the model, providing a more reliable basis for the accurate assessment of the model's performance. The Ideal curve serves as a theoretical reference benchmark, representing the ideal state where the predicted results are completely consistent with the actual situations. Through comparative analysis with this curve, the magnitude of the gap between the model's predicted results and the perfect prediction can be accurately located. The note "B = 10,000 repetitions" in the figure indicates that this analysis is based on 10,000 repeated sampling experiments, and the quantitative index of the mean absolute error further measures the accuracy of the model's predictions numerically. A smaller value indicates a lower overall deviation between the model's predicted values and the actual values.

https://doi.org/10.2147/IDR.S541396

Table 4 Logistic Regression to Predict MPP Based on Lasso Regression

Variable	*β*	*SE*	*Wald*χ^2	*P*
Intercept	−2.553	2.911	−0.88	0.381
Age	0.220	0.038	5.76	<0.001
NEUT%	0.022	0.012	1.78	0.075
EOS%	0.137	0.049	2.78	0.006
HSCRP	−0.018	0.008	−2.39	0.017
ADA	0.147	0.024	6.24	<0.001
Crea	0.042	0.011	3.88	<0.001
Urea	−0.358	0.100	−3.57	<0.001
HDL	−0.870	0.309	−2.81	0.005
P	−0.544	0.191	−2.84	0.005
ESR	0.014	0.004	3.48	0.001

Note: p-values <0.05 indicate statistically significant differences.

Figure 5c and d are Decision Curve Analysis plots, which are used to evaluate the standardized net benefits (Standardized Net Benefit) of the Nomogram model under different settings of the risk threshold (High Risk Threshold). The three curves in the figures represent different decision - making hypothesis scenarios respectively: The “Nomogram model” curve reflects the changing trend of the net benefits of the nomogram model at various risk thresholds, and its performance can be used to assess the potential application value of this model in the clinical decision - making process. The “All” curve represents an extreme decision - making strategy, that is, the net benefit situation under the assumption that all patients receive treatment. The “None” curve corresponds to another extreme decision - making hypothesis, that is, the net benefit status when none of the patients receive treatment. The Cost - Benefit Ratio labeled on the horizontal axis of the figure provides a key dimension for evaluating the decision - making effect of the model under different cost - benefit trade - off considerations. By comparing the Nomogram model curve with the two extreme - scenario curves, it can be seen that within a specific risk - threshold range, the Nomogram model can achieve higher standardized net benefits.

According to the Logistic+Lasso regression results, a total of 9 variables were selected to construct the final risk prediction model: Age, NEUT%, EOS%, HSCRP, ADA, Crea, Urea, HDL, P and ESR Logistic regression model was further established based on parameters screened by Logistic+Lasso regression (Table 4). The results showed that the risk factors for the occurrence of the outcome were: Age, with an odds ratio (OR) and 95% confidence interval (95% CI) of 1.246 (1.152, 1.348); EOS%, OR, 95% CI = 1.147 (1.052, 1.252); ADA, OR, 95% CI = 1.159 (1.098, 1.225); Crea, OR, 95% CI = 1.043 (1.021, 1.066); ESR, OR, 95% CI = 1.014 (1.006, 1.023). The protective factors for the occurrence of the outcome were: hs - CRP, OR, 95% CI = 0.982 (0.968, 0.997); Urea, OR, 95% CI = 0.700 (0.589, 0.830); HDL, OR, 95% CI = 0.419 (0.234, 0.750); P, OR, 95% CI = 0.570 (0.398, 0.818).

Nomogram as a Tool for Visualization

Ultimately, we selected the Lasso combined with Logistic regression model to assess the risk of MPP in patients. To enhance clinical applicability, we transformed the intricate mathematical model into a nomogram (Figure 6). The calculation of individual risk scores necessitated the aggregation of the variable scores as delineated in the model. Subsequently, a vertical line was drawn at the total score, intersecting with the lines that denote the predicted MPP. The values at the intersection points corresponded to the predicted MPP for each individual. This approach highlights that the nomogram offers a more user-friendly option for clinical practice compared to traditional mathematical formulas.

Discussion

CAP remains a significant cause of morbidity and mortality among children worldwide, particularly in developing countries. Among the various etiological agents of children in China, MPP is a common cause of CAP in children. MPP is a prevalent

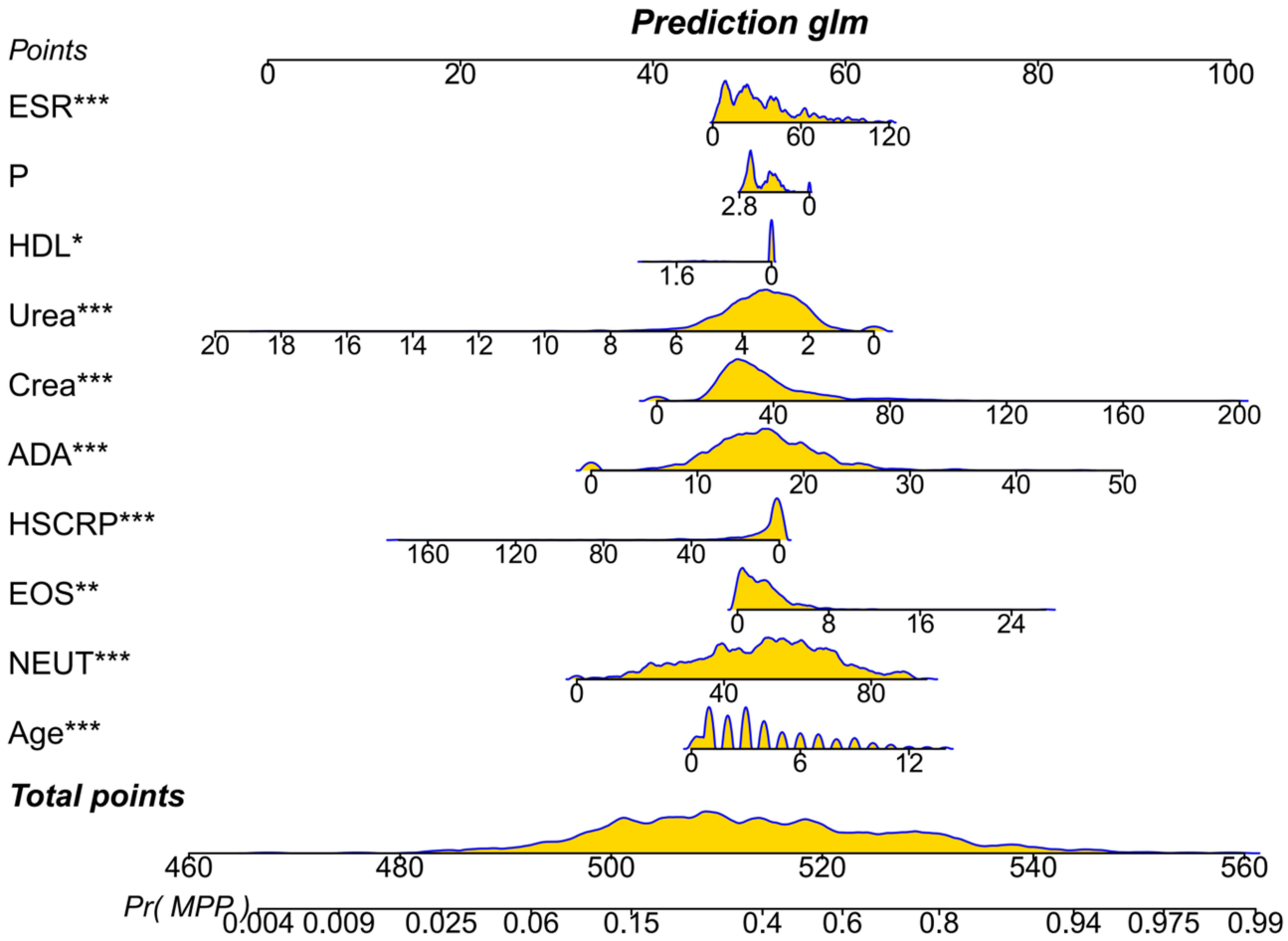

Figure 6 Nomogram of Lasso+logistic regression model for predicting MPP.* represents P<0.05, ** represents P<0.01, *** represents P<0.001.

respiratory illness among children, primarily resulting from infection with Mycoplasma pneumoniae.[20] MPP was positive in about 30% of our hospitals. In this study, we developed a nomogram model for MPP to predict and assess the disease progression in patients with MPP, which underscoring their potential utility in clinical settings, particularly in the context of pediatric healthcare. Clinicians can significantly shorten the course of illness and improve prognosis by promptly selecting macrolide or tetracycline antibiotics through early identification.[21] By identifying key risk factors for MPP in children with CAP, our research can provide new evidence for the formulation and update of clinical guidelines.

Our primary findings indicated that several laboratory parameters are significantly associated with the risk of MPP. Specifically, we identified that elevated levels of Age, HSCRP, ESR, and certain hematological markers such as NEUT%, EOS%, ADA, Crea, Urea, HDL, and P are strongly correlated with the occurrence of MPP. WBC, HSCRP and ESR are highly sensitive markers for infection and inflammation.[22] Furthermore, our findings indicated that the levels of ESR, ADA, and creatinine were significantly elevated in the MPP group compared to those in the NMPP group. However, the LDH, AST and P levels in MPP children were significant lower. These findings highlight the importance of inflammatory markers in the diagnosis and prognosis of MPP.[23,24] Overall our study provides valuable insights into the risk stratification and early identification of MPP in pediatric patients with CAP. Specifically, our study identified a set of predictors, including age, NEUT%, EOS%, HSCRP, ADA, Crea, Urea, HDL, P, and ESR, which were not comprehensively evaluated in earlier researches,[25] This finding is particularly novel as it integrates both traditional and penalized regression techniques to enhance predictive accuracy. Previous studies, such as those by Atkinson[23] and Meyer Sauteur,[26] primarily focused on individual biomarkers or clinical features without leveraging multivariate methods to refine their predictive models.

https://doi.org/10.2147/IDR.S541396

Our study, therefore, multi-faceted predictive framework that can be converted into clinical practice, as evidenced by the development of a nomogram for practical use. This approach not only improves the identification of at risk patients but also facilitates early intervention, potentially reducing morbidity and healthcare costs associated with MPP. The inclusion of variables such as age, NEUT%, EOS%, HSCRP, ADA, Crea, Urea, HDL, P, and ESR in the final predictive model underscores the multifactorial nature of MPP and the necessity for a comprehensive diagnostic approach. These findings can provide new evidence for the formulation and update of clinical guidelines, and promote more individualized and efficient management of pediatric CAP. In addition, the creation of a nomogram derived from the Lasso-Logistic regression model enhances the feasibility of implementing our predictive model in clinical environments. This development empowers healthcare professionals to assess the risk of MPP with improved precision and assurance. This can ultimately improve patient outcomes by enabling timely and appropriate interventions. However, it is important to acknowledge the limitations of our study, such as the single-center design and the potential for selection bias, which may affect the extrapolability of our findings. Future research should aim to validate our model in diverse populations and explore the integration of additional biomarkers to further refine risk prediction and enhance clinical utility.[23,24,27,28]

The primary limitations of this study should be acknowledged to provide a comprehensive understanding of the findings. Firstly, This limitation could potentially affect the generalizability of our results. Secondly, the study was conducted in a single center, which may introduce selection bias and limit the applicability of the findings to other settings or regions. Thirdly, the data collection was limited to a specific timeframe from August 2023 to March 2024, while covering a complete epidemic wave, did not include the summer months. Future longer-term surveillance spanning multiple years is warranted to fully understand the annual seasonality and long-term trends of MP infections in this region.

The model employed in this study, while demonstrating promising performance, still requires further validation through larger, multi-center prospective studies. Therefore future research should focus on externally validating this model and integrating it into clinical workflows to assess its real-world impact on patient treatment and outcomes. Additionally, incorporating advanced machine learning techniques could further refine risk prediction models, offering more precise and individualized and accurate testing.

Conclusions

In conclusion, This study provides a robust predictive model for MPP in children with CAP, which has identified several key risk factors associated with MPP in children. Various laboratory parameters (such as age, NEUT%, EOS%, HSCRP, ADA, Crea, Urea, HDL, P, and ESR) serve as effective early diagnostic indicators for MPP infection. It is essential to conduct further research to establish the reference ranges for hematological parameters specifically relating to MPP infection, taking into account individual physiological variations.

Data Sharing Statement

Article-related data are available from the two corresponding author.

Ethics Approval and Consent to Participate

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Xindu District People's Hospital in Chengdu (approval number: [IIT]PJ2025010). Owing to the retrospective nature of the study. Xindu District People's Hospital in Chengdu Ethics Review Committee waived the requirement for informed consent. Data was gathered in an anonymous manner through the electronic medical records system, with all authors committed to maintaining the confidentiality of patient information.

Acknowledgments

The authors acknowledge all individuals who participated in the study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically

reviewing the article, gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by grants from the Health Commission of Chengdu (Grant No.2024323 and 2023556) and Sichuan Hospital Association County Hospital Research Special Project (Grant No.2024LC001).

Disclosure

The authors declare no conflicts of interest.

References

1. Esposito S, Argentiero A, Gramegna A, et al. Mycoplasma pneumoniae: a pathogen with unsolved therapeutic problems. *Expert Opinion Pharmacother*. 2021;22(9):1193–1202. doi:10.1080/14656566.2021.1882420
2. Choo S, Kim S-H, Lee E. Clinical significance of Mycoplasma pneumoniae specific IgM titer in children hospitalized with Mycoplasma pneumoniae pneumonia. *BMC Infect Dis*. 2022;22(1). doi:10.1186/s12879-022-07456-6
3. J Zhu, X Liu, Zhan X, et al. Predictive value of chemokines (CCL 2) in bronchoalveolar lavage fluid for refractory mycoplasma pneumonia in children. *Ital J Pediatrics*. 2023;49(1). doi:10.1186/s13052-023-01528-2
4. Chaudhry R, Ghosh A, Chandolia A. Pathogenesis of Mycoplasma pneumoniae: an update. *Indian J Med Microbiol*. 2016;34(1):7–16. doi:10.4103/0255-0857.174112
5. Tong L, Huang S, Zheng C, et al. Refractory Mycoplasma pneumoniae pneumonia in children: early recognition and management. *J Clin Med*. 2022;11(10):2824. doi:10.3390/jcm11102824
6. Lan X-H, Zhang Y-X, Yuan W-H, et al. Image-based deep learning in diagnosing mycoplasma pneumonia on pediatric chest X-rays. *BMC Pediatric*. 2024;24(1). doi:10.1186/s12887-024-05204-0
7. Hazir T, Nisar YB, Qazi SA, et al. Chest radiography in children aged 2-59 months diagnosed with non-severe pneumonia as defined by World Health Organization: descriptive multicentre study in Pakistan. *BMJ*. 2006;333(7569):629. doi:10.1136/bmj.38915.673322.80
8. Kumar S, Garg IB, Sethi GR. Serological and molecular detection of Mycoplasma pneumoniae in children with community-acquired lower respiratory tract infections. *Diagn Microbiol Inf Dis*. 2019;95(1):5–9. doi:10.1016/j.diagmicrobio.2019.03.010
9. Kutty PK, Jain S, Taylor TH, et al. Mycoplasma pneumoniae among children hospitalized with community-acquired pneumonia. *Clin Infect Dis*. 2018. doi:10.1093/cid/ciy419
10. Qu J, Chen S, F Bao, et al. Molecular characterization and analysis of Mycoplasma pneumoniae among patients of all ages with community-acquired pneumonia during an epidemic in China. *Inter J Infect Dis*. 2019;83:26–31. doi:10.1016/j.ijid.2019.03.028
11. Tang M, Wang D, Tong X, et al. Comparison of different detection methods for Mycoplasma pneumoniae infection in children with community-acquired pneumonia. *BMC Pediatric*. 2021;21(1). doi:10.1186/s12887-021-02523-4
12. W Tuo, X Guo, Wu M, et al. Application value of antibody titres and RNA detection in the early prediction of Mycoplasma pneumoniae pneumonia in children: a retrospective study. *BMC Infect Dis*. 2023;23(1). doi:10.1186/s12879-023-08161-8
13. Loens K, Goossens H, Ieven M. Acute respiratory infection due to Mycoplasma pneumoniae: current status of diagnostic methods. *Eur J Clin Microbiol Infect Dis*. 2010;29(9):1055–1069. doi:10.1007/s10096-010-0975-2
14. Ratliff AE, Duffy LB, Waites KB, et al. Comparison of the illumi gene mycoplasma DNA amplification assay and culture for detection of Mycoplasma pneumoniae. *J Clin Microbiol*. 2014;52(4):1060–1063. doi:10.1128/jcm.02913-13
15. Chen D, Wu P, D LIU, et al. Clinical role of M. pneumoniae typing antibody detected by chemiluminescent immunoassay in the diagnosis of Mycoplasma pneumoniae pneumonia in children. *Int Immunopharmacol*. 2022;112. doi:10.1016/j.intimp.2022.109196
16. Busson L, VAN DEN WIJNGAERT S, Dahma H, et al. Evaluation of 10 serological assays for diagnosing Mycoplasma pneumoniae infection. *Diagn Microbiol Inf Dis*. 2013;76(2):133–137. doi:10.1016/j.diagmicrobio.2013.02.027
17. Can E, Can C. The value of neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) parameters in analysis with fetal malnutrition neonates. *J Perinatal Med*. 2019;47(7):775–779. doi:10.1515/jpm-2019-0016
18. J Liu, Li S, Zhang S, et al. Systemic immune-inflammation index, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio can predict clinical outcomes in patients with metastatic non-small-cell lung cancer treated with nivolumab. *J Clin Lab Anal*. 2019;33(8). doi:10.1002/jcla.22964
19. Naess A, S Nilssens, Mo R, et al. Role of neutrophil to lymphocyte and monocyte to lymphocyte ratios in the diagnosis of bacterial infection in patients with fever. *Infection*. 2016;45(3):299–307. doi:10.1007/s15010-016-0972-1
20. Luo Y, Wang Y. Risk prediction model for necrotizing pneumonia in children with Mycoplasma pneumoniae pneumonia. *J Inflamm Res*. 2023;16:2079–2087. doi:10.2147/jir.S413161
21. Izumikawa K. Clinical features of severe or fatal Mycoplasma pneumoniae pneumonia. *Front Microbiol*. 2016;7. doi:10.3389/fmicb.2016.00800.
22. Jiang Y, Wang W, Zhang Z, et al. Serum amyloid a, C-reactive protein, and procalcitonin levels in children with Mycoplasma pneumoniae infection. *J Clin Lab Anal*. 2022;36(3). doi:10.1002/jcla.24265
23. Atkinson TP, Balish MF, Waites KB. Epidemiology, clinical manifestations, pathogenesis and laboratory detection ofMycoplasma pneumoniaeinfections: figure 1. *FEMS Microbiol Rev*. 2008;32(6):956–973. doi:10.1111/j.1574-6976.2008.00129.x
24. Waites KB, Talkington DF. Mycoplasma pneumoniaeand Its Role as a Human Pathogen. *Clin Microbiol Rev*. 2004;17(4):697–728. doi:10.1128/cmr.17.4.697-728.2004
25. F FAN, Lv J, Yang Q, et al. Clinical characteristics and serum inflammatory markers of community-acquired mycoplasma pneumonia in children. *Clin Resp J*. 2023;17(7):607–617. doi:10.1111/crj.13620

26. MEYER SAUTEUR PM, VAN ROSSUM AMC, Vink C. Mycoplasma pneumoniae in children. *Curr Opin Infect Dis*. 2014;27(3):220–227. doi:10.1097/qco.0000000000000063
27. Jain S, Williams DJ, Arnold SR, et al. Community-acquired pneumonia requiring hospitalization among U.S. children. *N Engl J Med*. 2015;372(9):835–845. doi:10.1056/NEJMoa1405870
28. Bradley JS, Byington CL, Shah SS, et al. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the pediatric infectious diseases society and the infectious diseases society of America. *Clin Infect Dis*. 2011;53(7):e25–e76. doi:10.1093/cid/cir531