





OPEN Development of machine learning-based differential diagnosis model and risk prediction model of organ damage for severe *Mycoplasma pneumoniae* pneumonia in children

Bing He¹, Xuewen Li², Rongrong Dong¹, Han Yao¹, Qi Zhou³, Changyan Xu⁴, Chengming Shang⁵, Bo Zhao⁶, Huiling Zhou⁶, Xinqiao Yu⁶ & Jiancheng Xu¹  

Severe *Mycoplasma pneumoniae* pneumonia (SMPP) poses significant diagnostic challenges due to its clinical features overlapping with those of other common respiratory diseases. This study aims to develop and validate machine learning (ML) models for the early identification of SMPP and the risk prediction for liver and heart damage in SMPP using accessible laboratory indicators. Cohort 1 was divided into SMPP group and other respiratory diseases group. Cohort 2 was divided into myocardial damage, liver damage, and non-damage groups. The models built using five ML algorithms were compared to screen the best algorithm and model. Receiver Operating Characteristic (ROC) curves, accuracy, sensitivity, and other performance indicators were utilized to evaluate the performance of each model. Feature importance and Shapley Additive Explanation (SHAP) values were introduced to enhance the interpretability of models. Cohort 3 was used for external validation. In Cohort 1, the SMPP differential diagnostic model developed using the LightGBM algorithm achieved the highest performance with $AUC_{ROC} = 0.975$. In Cohort 2, the LightGBM model demonstrated superior performance in distinguishing myocardial damage, liver damage, and non-damage in SMPP patients (accuracy = 0.814). Feature importance and SHAP values indicated that ALT and CK-MB emerged as pivotal contributors significantly influencing Model 2's output magnitude. The diagnostic and predictive abilities of the ML models were validated in Cohort 3, demonstrating the models had some clinical generalizability. The Model 1 and Model 2 constructed by LightGBM algorithm showed excellent ability in differential diagnosis of SMPP and risk prediction of organ damage in children.

Keywords Severe *Mycoplasma pneumoniae* pneumonia, Children, Machine learning, LightGBM, Model

Mycoplasma pneumoniae pneumonia (MPP) is a common childhood respiratory disease caused by *Mycoplasma pneumoniae* (MP) infection¹. According to a 10-year study on respiratory tract surveillance in China, MP is the atypical pathogen with the highest detection rate in children aged 5 to 7 years². Although MPP is generally considered to be a self-limiting disease and most children have mild conditions with a good prognosis³, in some special cases, it can also progress into severe MPP (SMPP) or refractory MPP (RMPP) complicated by serious intrapulmonary and extrapulmonary complications⁴. Intrapulmonary complications include pleural effusion, atelectasis, necrotizing pneumonia (NP), and pulmonary embolism, etc. In severe cases, respiratory failure and hypoxemia may occur, which can be fatal⁵. In addition to causing lung damage, MPP can lead to a range of extrapulmonary complications, including myocardial injury, abnormal liver function, kidney damage, anemia, central nervous system encephalitis, and thrombosis^{6,7}.

In recent years, the incidence of SMPP has gradually increased, rising from 0.7% in 2006 to 35% in 2016⁸. The specificity of early lung manifestations in children with SMPP is poor, making it difficult to distinguish from

¹Department of Laboratory Medicine, First Hospital of Jilin University, Changchun 130021, China. ²Department of Hematology, First Hospital of Jilin University, Changchun 130021, China. ³Department of Pediatrics, First Hospital of Jilin University, Changchun 130021, China. ⁴Medical Department, First Hospital of Jilin University, Changchun 130021, China. ⁵Information center, First Hospital of Jilin University, Changchun 130021, China. ⁶Department of Laboratory Medicine, Meihoukou Central Hospital, Meihoukou 135000, China. ✉email: xjc@jlu.edu.cn

lung injury caused by other pathogens. Despite the growing availability of molecular techniques for pathogen detection, including quantitative and qualitative pathogen detection, laboratory results are usually only available after treatment decisions have been made. Moreover, with the widespread application of macrolides in children, drug-resistant MP strains have become increasingly common. In China, more than 85% of MP strains among pediatric patients have been reported as macrolide-resistant⁹. Consequently, precise prediction and identification of SMPP in clinical practice are crucial to reduce mortality and complications and to improve the prognosis of SMPP. At present, few epidemiological and comparative studies focus on the differences among viral pneumonia, bacterial pneumonia, and MPP, which may result in the abuse of antibiotics and increased microbial resistance to antibiotics¹⁰.

There are no well-defined and unified diagnostic criteria for pediatric SMPP. MPP patients with extensive lung lesions spreading over more than two thirds of the chest area in radiographic image, or those developing intra- and extra-pulmonary complications, can all be considered as SMPP. Clearly, SMPP has been recognized to present diverse clinical phenotypes, including pulmonary and non-pulmonary complication subtypes. Extrapulmonary complications can occur at any time following the onset of MP infections, even in asymptomatic cases, affecting up to 25% of those infected with MP¹¹. A study shows that cardiovascular disease accounts for approximately 27.91% of extrapulmonary complications in MPP patients¹². Factors leading to myocardial damage are significant for the early diagnosis and prevention of cardiac complications. Additionally, reports indicate that MP-related hepatitis occurs in approximately 10–30% of cases, mostly in children. This condition can manifest as asymptomatic elevation of liver enzymes, inhibition of multiple coagulation factors, or cholestasis^{13–15}. Therefore, ignoring MP-related hepatitis may delay diagnosis and lead to serious liver damage during MPP. Therefore, establishing an accurate and reliable predictive model is crucial for preventing the occurrence of related organ damage in the early of MPP.

Artificial intelligence (AI) and machine learning (ML) are increasingly recognized as powerful tools for handling complex medical tasks. Machine learning algorithms excel at exploring intricate relationships within multidimensional data, extracting hidden and valid knowledge from vast datasets, and making more accurate predictions and diagnoses of diseases¹⁶. Historically, numerous studies have developed MPP prediction models focusing on diagnosis, severity, risk factors, treatment, and prognosis^{17–20}. These studies primarily relied on constructing nomogram models for the early identification and intervention of MPP. Nomogram is a traditional calculation tool that includes variables and corresponding scoring lines. However, complex ML methods can manage a broader array of variables, often yielding more accurate and precise results than traditional modeling methods²¹. Additionally, one of the main challenges in applying MPP prediction models to clinical practice is the lack of external validation^{19–22}.

Consequently, this study aims to use machine learning algorithms, based on the first laboratory test results of SMPP patients after their visit, to construct and validate an early warning model that could be applied in early identification and intervention of SMPP, and the prevention of disease progression. Secondly, it aims to develop a risk prediction model for liver and heart damage in SMPP to improve the prognosis and outcomes of SMPP patients. The performance of the models is validated using an external validation cohort from another hospital.

Methods

Study design and study population cohorts

This multicenter observational retrospective study was conducted at two hospitals. Cohort 1 (model 1 construction and internal validation): Patients with SMPP, viral and bacterial pneumonia who first visited the First Hospital of Jilin University from 2021 to 2023 were included. They were divided into SMPP and other respiratory diseases groups to establish a differential diagnostic model for distinguishing between SMPP and other respiratory diseases. Cohort 2 (model 2 construction and internal validation): The SMPP group was divided into myocardial damage, liver damage, and non-damage groups for subgroup analysis to predict the types of extrapulmonary organ damage in SMPP early. Cohort 3 (external validation): Patients with SMPP, viral and bacterial pneumonia who first visited the Meihokou Central Hospital from 2023 to 2024 were included for external validation using the same treatments as in Cohort 1 and Cohort 2. This study was approved by the Ethics Committee of the First Hospital of Jilin University (NO. 2020-313). Due to the retrospective nature of the study, the requirement for written informed consent was waived by the Ethics Committee. The research followed the Helsinki Declaration²³.

Criteria for defining SMPP: Patients were diagnosed with SMPP according to the Guidelines for Diagnosis and Treatment of Mycoplasma Pneumoniae Pneumonia in Children (2023 Edition). MPP cases were categorized as SMPP if patients exhibited any of the followings: (1) Poor general condition; (2) Conscious disorder, cyanosis, respiratory dysfunction; (3) Hypoxemia, assisted respiration (groan, nasal fan, three concave sign), intermittent apnea, and oxygen saturation $\leq 92\%$; (4) Persistent hyperpyrexia for more than 5 days or ultra-hyperpyrexia; (5) Dehydration and food refusal; (6) Chest X-ray or CT scan showing the following findings: unilateral lung infiltration $\geq 2/3$, multi-lobe lung infiltration, pleural effusion, pneumothorax, atelectasis, lung necrosis, lung abscess; (7) Extrapulmonary complications.

Inclusion criteria: (1) Patients were younger than 18 years old; (2) Clinical data of the patients were complete. Exclusion criteria: (1) Evidence of co-infection, including bacteria, viruses, fungi, and tuberculosis, etc.; (2) Condition in the recovery period at the time of admission (patients with a disease course of more than 4 weeks, stable temperature for more than 1 week, improvement of chest imaging); (3) Pre-existing other systemic diseases, such as congenital heart disease, chronic kidney disease, chronic lung disease, connective tissue disease, hematological diseases, and tumors; (4) Patients who have received recent anticoagulant treatment or were using anticoagulants for other medical conditions; (5) Patients with a recent history of major surgery, serious trauma, or blood transfusion; and (6) Patients with an immunodeficiency or other diseases that could cause abnormal immune function, as well as those who had recently received immunotherapy or hormone therapy.

Inclusion criteria for other respiratory diseases: (1) Single infection patients with COVID-19, influenza A, and bacterial pneumonia; (2) COVID-19 and influenza A cases confirmed by reverse-transcriptase-polymerase-chain-reaction (RT-PCR); (3) Bacterial pneumonia cases confirmed by detecting at least one putative bacterial pathogen in blood or pleural fluid through culture or polymerase chain reaction (PCR). Exclusion criteria: (1) Malignant tumor; (2) Blood system disease; (3) Serious immune system disease.

Data collection

Clinical data and laboratory test results for the study subjects were collected through the laboratory information system and the electronic medical record system stored in the database. Fasting venous blood was collected within 24 h after admission for blood analysis. A chest X-ray or chest CT was performed 3 days before or within 3 days after admission, and the results were recorded. The patient ID number was used as the unique identification of the study subjects. Specifically, the following clinical data were collected for all patients: (1) Basic demographic information, including age and sex. Clinical characteristics included past medical history, clinical symptoms, length of hospital stay, type of pathogen(s), pre-treatment duration with macrolides, intrapulmonary and extrapulmonary complications, and laboratory and imaging findings. (2) Findings of physical examinations, including respiratory rate, heart rate, and breath sounds. (3) Laboratory tests included white blood cell counts (WBC), hemoglobin (HGB), absolute neutrophil count (ANC), lymphocyte count (LYC), platelet Count (PLT), C-reactive Protein (CRP), procalcitonin (PCT), lactate dehydrogenase (LDH), D-dimer, fibrinogen (FIB), activated partial thromboplastin time (APTT), prothrombin time (PT), prothrombin time activity (PTA), thrombin time (TT), alanine aminotransferase (ALT), aspartate aminotransferase (AST), albumin (ALB), blood urea nitrogen (BUN), blood electrolytes, creatine kinase (CK), MB isoenzyme (CK-MB), hydroxybutyrate dehydrogenase (HBDH), and cardiac troponin I (CTnI), etc. Serum *Mycoplasma pneumoniae* antibody and MP polymerase chain reaction (PCR) were performed to determine the presence of MP infection. We used virus antigen detection assays and RT-PCR methods to confirm pathogen identification and exclude coinfection. The diagnostic tests were determined based on clinical judgment. These tests were performed on samples taken from various sources, including nasopharyngeal swabs, throat swabs, sputum samples, pleural effusion samples, and bronchoalveolar lavage fluid. The test results of electrocardiogram (ECG), echocardiography, and abdominal ultrasound of the study subjects were collected from the image archiving and communication system.

Data cleaning and normalization

To improve data quality and ensure accuracy, consistency, and availability, we cleaned and standardized the collected raw medical data: (1) Data inspection and cleaning: After summarizing and sorting the raw data, the display formats of values, time, date, and full half-angle were integrated for consistency. Repeatability was checked, and exact duplicate data were eliminated; (2) Data normalization: The four elements of specimen type, test item name, test result unit, and test reference value were calibrated and normalized; (3) Data exclusion and padding: Tests with missing rates > 30% were excluded. For tests with missing rates ≤ 30%, a filling method representing the central tendency of the variables was selected for the measurement data: median replacement, mean replacement, or plural replacement; the count data were randomly interpolated according to the proportion of available negative and positive data.

Dealing with imbalanced datasets

In imbalanced datasets, there were differences in the number of samples between different classes, which could result in poor classification performance of the model for minority classes. To address the issue of sample imbalance, we adopted the SMOTE method to balance the datasets. The SMOTE technique was a type of oversampling method. The SMOTE technique generated randomly new examples or instances of the minority class from the nearest neighbors of line joining the minority class sample to increase the number of instances. As a result, new training datasets were generated. In Cohort 1, SMOTE increased the SMPP group from 1020 instances of the minority class to 1182 patients. In Cohort 2, SMOTE increased the liver damage and non-damage groups from 300 and 354 instances of the minority class to 366 patients.

ML algorithms and model Building

The flow chart for building models using machine learning algorithms was shown in Fig. 1. Cohort 1 and Cohort 2 used the 5-fold cross-validation method on the entire dataset, dividing it into five folds. Four of the folds were used as the training set to train the model, and the remaining fold was used as the validation set to score the model. This process was repeated five times, and the average value was taken. The risk prediction models built by Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), Light Gradient Boosting Machine (LightGBM), K Nearest Neighbors (KNN), and Random Forest (RF) were compared. The ML algorithms with the best performance were selected for subsequent model building and validation. The specific parameter values for each algorithm were as follows, XGBoost: The booster selected was gbtrees; LightGBM: The boosting type was set to gradient boosting decision trees, learning rate was set to 0.1, max depth was set to -1, number of estimator was set to 100, number of leaves was set to 31; RF: The criterion was set to gini, max impurity decrease was set to 0, number of estimators was set to 20; KNN: The number of neighbors was set to 5, weighting scheme was set to uniform; LR: Logistic regression used L2 regularization, with C equals 1, max number of iteration was set to 100. The features selected from the optimal algorithm were used for subsequent model building. To further enhance model performance, a refinement step was implemented by ranking the features based on their importance and selecting the top ten features from the best-performing model. This ensures that the models are built on the most influential features, potentially improving their predictive accuracy. Cohort 3 performs external validation of the model.

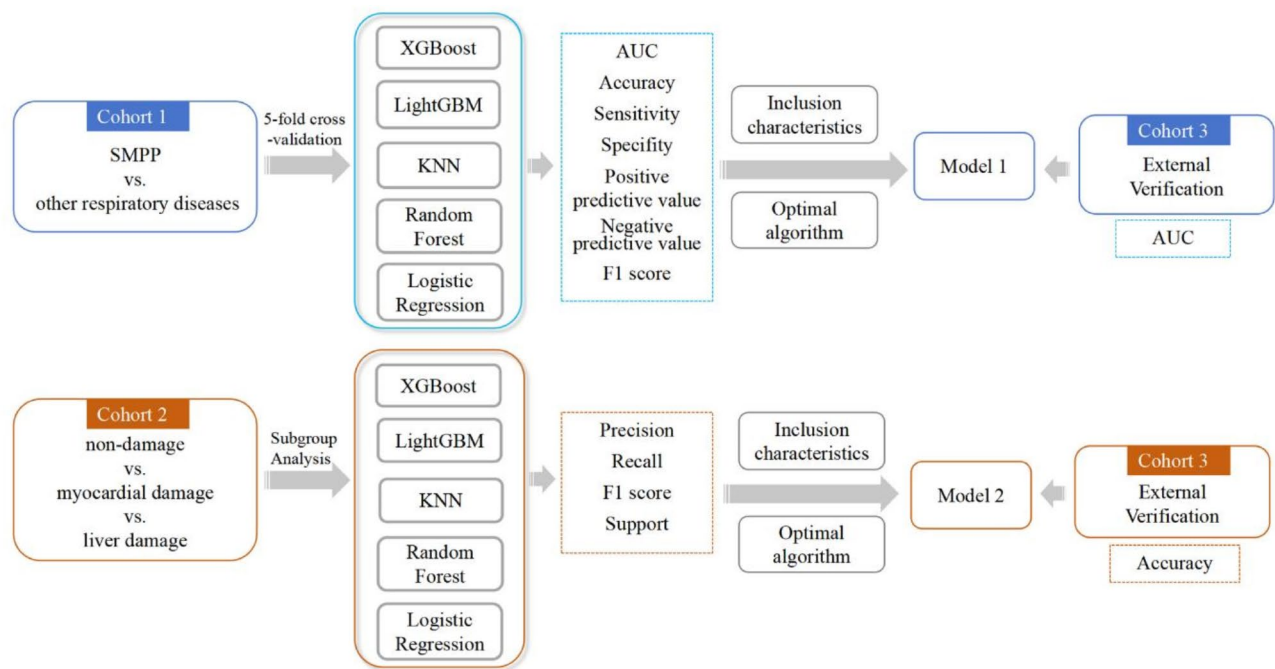


Fig. 1. Flowchart of machine learning to build the Model 1 and Model 2.

Cohort 1 and Cohort 2 datasets were divided into training and validation sets in a ratio of 7:3. Cohort 1 evaluated model performance using Receiver Operating Characteristic (ROC) curves and calculated the Area Under the Curve (AUC) for the model. Calibration curves were generated to evaluate the proximity of the model's predicted risk to the actual risk, with Decision Curve Analysis (DCA) applied to evaluate the decision utility of the model. Cohort 2 selected the LightGBM model to construct a classification model. Shapley Additive Explanation (SHAP), an interpretable tool for prediction output of machine learning models, could reflect the influence of the features in each data sample. This method could help increase visibility and interpretability of the ML model. In this study, SHAP analysis was used to explain the decision-making process of the model, including sorting the features by importance, showing the association between observed values and risk. To evaluate the performance of the model, accuracy was calculated. The macro and weighted precision, recall, and F1 scores were calculated for the three classes.

Statistical analysis

Excel 2016 was used to store and manage the data, and SPSS 22.0 was used for statistical analysis. Feature selection and model construction were carried out using the Deepwise & Beckman Coulter DxAI platform (<http://dxo.nline.deepwise.com>). In the training cohort, least absolute shrinkage and selection operator (LASSO) logistic regression analysis was utilized to rank the importance of risk factors. In LASSO regression, the beta coefficients of variables that were not strongly associated with the outcome were decreased to zero, which removed these variables from the model. Categorical variables were represented as composition ratios. Continuous variables were represented as mean \pm standard deviation (SD) or median with interquartile range (IQR), depending on the data distribution. The Student's t-test, Mann-Whitney U test, or one-way analysis of variance test was used to compare the distribution of variables between groups, as appropriate. Pearson χ^2 test or Fisher exact test was used to compare categorical variables, with a two-sided $P < 0.05$ considered statistically significant.

Results

Basic characteristics of included cohorts

Cohort 1 enrolled 1020 SMPP patients and 1182 patients with other respiratory diseases, including bacterial pneumonia, COVID-19, and influenza A patients. Cohort 2 enrolled 1020 SMPP patients, including 366 with myocardial damage, 300 with liver damage, and 354 without damage. Cohort 3 enrolled 343 SMPP patients and 313 patients with other respiratory diseases. The demographic, clinical, and laboratory characteristics of the study population were shown in Table 1. In Cohort 1, there was a significant difference in age and sex between the SMPP and other respiratory diseases groups. Patients in the SMPP group were significantly older than those in the other respiratory diseases group ($p < 0.001$). There were significant differences in laboratory indicators ($p < 0.001$) between the two groups, except for WBC, CK-MB, and LDH ($p > 0.05$). In Cohort 2, there were no significant differences in characteristics among the three groups, including sex, clinical features, APTT, and PT ($p > 0.05$). Compared to those with liver damage and non-damage, SMPP patients with myocardial damage had significantly higher CK-MB values ($P < 0.001$). Compared to those with myocardial damage and non-damage, SMPP patients with liver damage had significantly higher ALT values ($P < 0.001$). There were significant

Characteristics	Cohort 1			Cohort 2			Cohort 3				
	SMPP (n = 1020)	Other respiratory diseases (n = 1182)	P	Myocardial damage (n = 366)	Liver damage (n = 300)	Non-damage (n = 354)	P	SMPP (n = 343) Myocardial damage (n = 115) Liver damage (n = 115) Non-damage (n = 113)	P ₁	Other respiratory diseases (n = 313)	P
Age (year)	6.0 (2.0–9.0)	3.0 (2.0–7.0)	<0.001	2.0 (1.0–3.0)	8.0 (6.0–10.0)	7.0 (4.0–9.0)	<0.001	6.0 (3.0–9.0) 3.0 (2.0–5.0) 8.0 (6.0–10.0) 7.0 (6.0–9.0)	<0.001	3.0 (1.0–5.0)	<0.001
Male, n (%)	488 (47.8)	514 (43.5)	0.041	173 (47.3)	130 (43.3)	185 (52.3)	0.072	171 (49.9) 54 (47.0) 52 (45.0) 65 (57.5) 172 (50.1) 61 (53.0) 63 (55.0) 48 (42.5) 8.0 (6.0–10.0)	0.133	166 (53.0)	0.416
Female, n (%)	532 (52.2)	668 (56.5)		193 (52.7)	170 (56.7)	169 (47.7)		6.0 (5.0–8.0) 9.0 (6.0–12.0) 8.0 (6.0–12.0)		147 (47.0)	
Length of stay (days)	8.0 (6.0–10.0)	7.8 (6.0–9.0)		8.0 (6.0–10.0)	10.0 (8.0–12.0)	7.0 (5.0–9.0)		<0.001		-	-
Clinical features											
Fever, n (%)	781 (76.6)	915 (77.4)	0.270	314 (85.8)	265 (88.3)	202 (57.1)	0.086	-	-	-	-
Cough, n (%)	632 (62.0)	763 (64.6)	0.140	275 (75.1)	237 (79.0)	120 (33.9)	0.052	-	-	-	-
Sore throat, n (%)	270 (26.5)	445 (37.6)	0.076	107 (29.2)	98 (32.7)	65 (18.4)	0.152	-	-	-	-
Laboratory result											
WBC (10 ⁹ /L)	8.5 (6.6–10.6)	8.3 (6.1–11.2)	0.400	8.6 (7.4–11.2)	9.2 (6.8–12.1)	7.9 (6.1–9.3)	<0.001	8.6 (6.2–12.0) 8.3 (6.1–10.5) 9.4 (6.4–13.0) 8.9 (6.1–12.6)	0.046	-	-
LYC (10 ⁹ /L)	2.9 (2.0–3.9)	3.1 (2.1–4.5)	<0.001	3.7 (2.7–4.7)	2.4 (1.6–3.4)	2.8 (1.9–3.4)	<0.001	2.6 (1.7–3.8)	-	3.0 (1.8–4.5)	0.012
HGB (g/L)	127.0 (120.0–134.0)	122.0 (114.0–130.0)	<0.001	126.0 (117.0–132.0)	130.0 (122.0–136.0)	127.1 (122.0–133.0)	<0.001	128.0 (121.0–135.0) 17.7 (13.3–24.8) 29.0 (21.3–73.4) 16.8 (13.1–19.1) 13.4 (11.1–16.8)	-	121.0 (113.0–128.0)	<0.001
CK-MB (U/L)	18.2 (14.1–23.5)	18.9 (14.5–23.9)	0.069	25.2 (20.6–29.3)	16.5 (12.5–19.6)	15.3 (12.4–18.2)	<0.001	285.0 (240.0–361.0) 293.0 (252.0–348.0) 354.0 (283.0–451.9) 241.0 (208.0–278.0)	<0.001	-	-
LDH (U/L)	314.0 (266.0–381.0)	298.0 (254.0–345.0)	<0.001	330.1 (286.0–387.0)	357.0 (291.0–434.0)	276.0 (240.0–317.0)	<0.001	205.0 (176.0–239.0)	<0.001	287.0 (244.0–328.6)	0.206
HBbDh (U/L)	213.0 (182.0–256.0)	202.0 (170.0–232.8)	<0.001	226.0 (197.0–261.0)	234.0 (190.0–287.0)	187.0 (165.0–219.0)	<0.001	209.0 (175.0–235.0)	-	209.0 (175.0–235.0)	0.755
APTT (s)	28.5 (28.3–28.5)	28.2 (28.2–28.2)	<0.001	28.5 (28.5–28.5)	28.5 (27.1–29.8)	28.5 (28.5–28.5)	0.340	30.4 (28.5–36.0)	-	28.2 (28.2–31.7)	<0.001
Continued											

Characteristics	Cohort 1			Cohort 2			Cohort 3				
	SMPP (n = 1020)	Other respiratory diseases (n = 1182)	P	Myocardial damage (n = 366)	Liver damage (n = 300)	Non-damage (n = 354)	P	SMPP (n = 343) Myocardial damage (n = 115) Liver damage (n = 115) Non-damage (n = 113)	P ₁	Other respiratory diseases (n = 313)	P
PT (s)	11.5 (11.5–11.5)	11.7 (11.7–11.7)	<0.001	11.5 (11.5–11.5)	11.5 (11.3–11.9)	11.5 (11.5–11.5)	0.084	12.4 (11.5–13.5)	–	11.7 (11.7–13.1)	0.694
PTA (%)	93.0 (93.0–93.1)	96.2 (96.2–96.2)	<0.001	93.0 (93.0–93.0)	93.0 (83.0–95.0)	93.0 (93.0–93.0)	<0.001	86.0 (74.7–93.0)	–	96.2 (80.0–96.2)	<0.001
CRP (mg/L)	5.4 (1.7–11.7)	7.8 (2.3–14.2)	<0.001	7.7 (2.3–12.1)	5.4 (1.5–12.6)	3.7 (1.6–9.5)	<0.001	6.8 (3.6–16.3)	–	5.7 (2.0–12.7)	0.007
ALT (U/L)	17.7 (13.4–23.9)	14.9 (11.2–20.9)	<0.001	17.6 (14.3–22.8)	24.1 (20.3–27.5)	13.5 (11.0–16.4)	<0.001	20.0 (14.0–27.9)	<0.001	–	–
								17.8 (13.6–23.0)			
								27.3 (22.1–76.0)			
								15.0 (11.6–20.0)			
AST (U/L)	31.2 (23.8–39.1)	29.4 (23.0–36.2)	<0.001	37.2 (30.3–43.6)	33.9 (26.6–40.2)	24.2 (20.5–29.3)	<0.001	33.0 (26.0–43.0)	<0.001	32.4 (25.1–40.0)	0.135
								37.9 (31.2–48.0)			
								34.0 (27.1–54.2)			
								27.0 (24.0–33.0)			
ALB (g/L)	40.0 (37.5–42.5)	39.8 (37.2–42.2)	0.056	40.7 (38.4–43.5)	40.0 (37.5–42.5)	38.1 (35.5–41.0)	<0.001	40.5 ± 4.2	<0.001	–	–
								41.6 ± 4.5			
								39.2 ± 4.2			
								40.6 ± 3.7			
Glucose (mmol/L)	4.9 (4.6–5.4)	4.8 (4.5–5.1)	<0.001	4.8 (4.4–5.2)	5.1 (4.7–5.8)	5.0 (4.6–5.3)	<0.001	5.0 (4.6–5.6)	–	4.8 (4.5–5.1)	<0.001
Ca ²⁺ (mmol/L)	2.4 (2.3–2.4)	2.3 (2.2–2.4)	<0.001	2.4 (2.3–2.4)	2.3 (2.2–2.4)	2.4 (2.3–2.5)	<0.001	2.3 (2.2–2.4)	<0.001	–	–
								2.4 (2.3–2.4)			
								2.3 (2.2–2.4)			
								2.3 (2.2–2.4)			
Continued											

Characteristics	Cohort 1		Cohort 2				Cohort 3				
	SMPP (n = 1020)	Other respiratory diseases (n = 1182)	P	Myocardial damage (n = 366)	Liver damage (n = 300)	Non-damage (n = 354)	P	SMPP (n = 343) Myocardial damage (n = 115) Liver damage (n = 113) Non-damage (n = 113)	P ₁	Other respiratory diseases (n = 313)	P
CO ₂ -CP (mmol/L)	23.4 (21.9–24.9)	23.0 (21.7–24.4)	<0.001	22.1 ± 2.1	24.5 ± 2.2	23.9 ± 2.0	<0.001	22.9 (20.8–24.6) 21.0 (19.3–22.9) 24.0 (22.7–26.0) 23.0 (21.0–24.5)	<0.001	–	–

Table 1. Demographic, clinical, and laboratory characteristics. Normally distributed variables are expressed as mean ± standard deviation, and non-normally distributed variables are expressed as median and interquartile range (IQR), or n (%). SMPP Severe Mycoplasma pneumoniae pneumonia, WBC white blood cell counts, LYC lymphocyte count, HGB hemoglobin, CK-MB MB isoenzyme, LDH lactate dehydrogenase, HBDH hydroxybutyrate dehydrogenase, APTT activated partial thromboplastin time, PT prothrombin time, PTA prothrombin time activity, CRP C-reactive Protein, ALT alanine aminotransferase, AST aspartate aminotransferase, ALB albumin, Ca Calcium, CO₂-CP carbon dioxide combining power. P: statistics of differences between SMPP and other respiratory diseases groups; statistics of differences between myocardial damage, liver damage, and non-damage groups in Cohort 2. P₁: statistics of differences between myocardial damage, liver damage, and non-damage groups in Cohort 3.

differences in age and length of stay among the three groups of SMPP patients ($P<0.001$). Specifically, the age and length of stay were higher in SMPP patients with liver damage than in the other two groups. Similarly, the remaining laboratory variables differed significantly ($P<0.001$) among the three groups. In Cohort 3, there were significant differences in Age, PTA, APTT, Glucose (GLU), HGB, LYC, and CRP between the SMPP and other respiratory diseases groups ($p<0.05$), while the three subgroups of SMPP showed significant differences ($p<0.05$) in all laboratory indicators. Spearman correlation analysis method was employed to analyze the correlation between different indicators in the models (Supplementary Fig. S1).

Establishment of diagnostic model 1

In Cohort 1, LASSO regression was used to screen feature variables, resulting in 41 feature variables included as modeling features. The results of the 5-fold cross-validation of the five machine learning algorithms were shown in Table 2. Among them, the LightGBM algorithm performed the best with the highest AUC (0.975), so the model established by the LightGBM algorithm was chosen as the differential diagnostic model for SMPP and other respiratory diseases (Fig. 2). The top 10 features with the highest feature importance in Model 1 were PTA, PT, HGB, APTT, GLU, HBDH, LYC, AST, CRP, and LDH. The ROC curve results showed that Model 1 had excellent differential diagnostic ability ($AUC_{ROC}=0.975$, sensitivity=0.739, specificity=0.993) (Fig. 3A). The calibration curve indicated that Model 1’s sample probabilities were in good agreement with the predicted probabilities (Fig. 3B). Decision curve analysis demonstrated that Model 1 had a high clinical benefit (Fig. 3C). Feature importance analysis by SHAP values for the top 10 features was shown in Fig. 3D.

Establishment of classification model 2

In Cohort 2, to mitigate multicollinearity, correlation analysis was performed to identify indicators with high correlations, and univariate or multivariate analysis was used to select characteristic variables. Subsequently, LASSO regression was employed for feature variable selection. According to the importance of Lasso-selected feature variables, ALT, CK-MB, Age, AST, WBC, LDH, Length of stay, CO₂, ALB, and Ca were identified. Based on the afore-mentioned indicators, among the five machine learning algorithms XGBoost, LR, LightGBM, KNN, and RF, the LightGBM model was chosen due to its superior performance in distinguishing the three subgroups of SMPP patients. The LightGBM model was evaluated to classify the three groups in Cohort 2 (accuracy=0.814). The results, including weighted precision, recall, weighted F1 scores, and support for our classification model, were comprehensively presented in Table 3. Further elucidating the model’s intricacies, the importance of the average SHAP values of each feature in different classes were presented in Fig. 4. ALT and CK-MB emerged as pivotal contributors significantly influencing the model’s output magnitude, while Ca exhibited the lowest SHAP value, indicating a relatively lesser impact on the model’s classification. Notably, ALT was correlated with liver damage, whereas CK-MB was correlated with myocardial damage, both of which showed relatively significant contributions in the corresponding classes.

External verification

Using Cohort 3 as an external verification set, the performance of Model 1 and Model 2 were evaluated. In Cohort 3, 343 SMPP patients and 313 patients with other respiratory diseases were enrolled. The external validation results of the differential diagnostic performance of Model 1 were shown in Supplementary Fig. S2. The ROC curve results showed that the Model 1 had stable and excellent SMPP differential diagnosis ability ($AUC_{ROC}=0.884$ [95%CI, 0.860–0.909], sensitivity=0.769, specificity=0.831). The 343 SMPP patients included 115 with myocardial damage, 115 with liver damage, and 113 without damage. The overall diagnostic accuracy of Model 2 was 74.9%. Specifically, the model exhibited the highest diagnostic accuracy in identifying non-damage patients with SMPP, with a precision of 0.818. In terms of differentiating myocardial damage patients and liver damage patients, the model demonstrated precision rates of 0.757 and 0.709, respectively. The confusion matrix for external validation of Model 2 was shown in Supplementary Fig. S3.

Discussion

In recent years, the prevalence of MPP has been on the rise, with high mortality and complication rates²⁴. Although most children with MPP have a good prognosis after macrolide treatment, children with SMPP have

Model type	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95%CI)	F1 score (95%CI)
XGBoost	0.973 (0.962–0.984)	0.880 (0.869–0.892)	0.773 (0.752–0.794)	0.987 (0.981–0.993)	0.984 (0.976–0.991)	0.813 (0.799–0.828)	0.866 (0.851–0.880)
LightGBM	0.975 (0.964–0.986)	0.866 (0.858–0.874)	0.739 (0.726–0.753)	0.993 (0.989–0.997)	0.991 (0.985–0.997)	0.792 (0.784–0.801)	0.847 (0.837–0.857)
RandomForest	0.960 (0.946–0.975)	0.887 (0.873–0.902)	0.860 (0.820–0.901)	0.915 (0.888–0.941)	0.911 (0.888–0.934)	0.869 (0.836–0.903)	0.884 (0.868–0.900)
KNN	0.586 (0.536–0.636)	0.558 (0.543–0.573)	0.512 (0.477–0.546)	0.604 (0.578–0.630)	0.564 (0.548–0.579)	0.553 (0.538–0.569)	0.536 (0.512–0.560)
LR	0.735 (0.690–0.779)	0.672 (0.658–0.687)	0.652 (0.592–0.713)	0.692 (0.646–0.738)	0.681 (0.663–0.698)	0.668 (0.642–0.695)	0.664 (0.635–0.693)

Table 2. Diagnostic efficacy of five classifiers in the validation set for 5-fold cross-validation. *LR* Logistic Regression, *KNN* K Nearest Neighbors, *PPV* positive predictive value, *NPV* negative predictive value. *Sensitivity = True Positive/(True Positive + False Negative); Specificity = True Negative/(True Negative + False Positive); Accuracy = (True Positive + True Negative)/(Positive + Negative); Precision = True Positive/(True Positive + False Positive); Recall = True Positive /(True Positive + False Negative); F1 score = 2*Precision*Recall/(Precision + Recall); PPV = True Positive/(True Positive + False Positive); NPV = True Negative/(True Negative + False Negative).

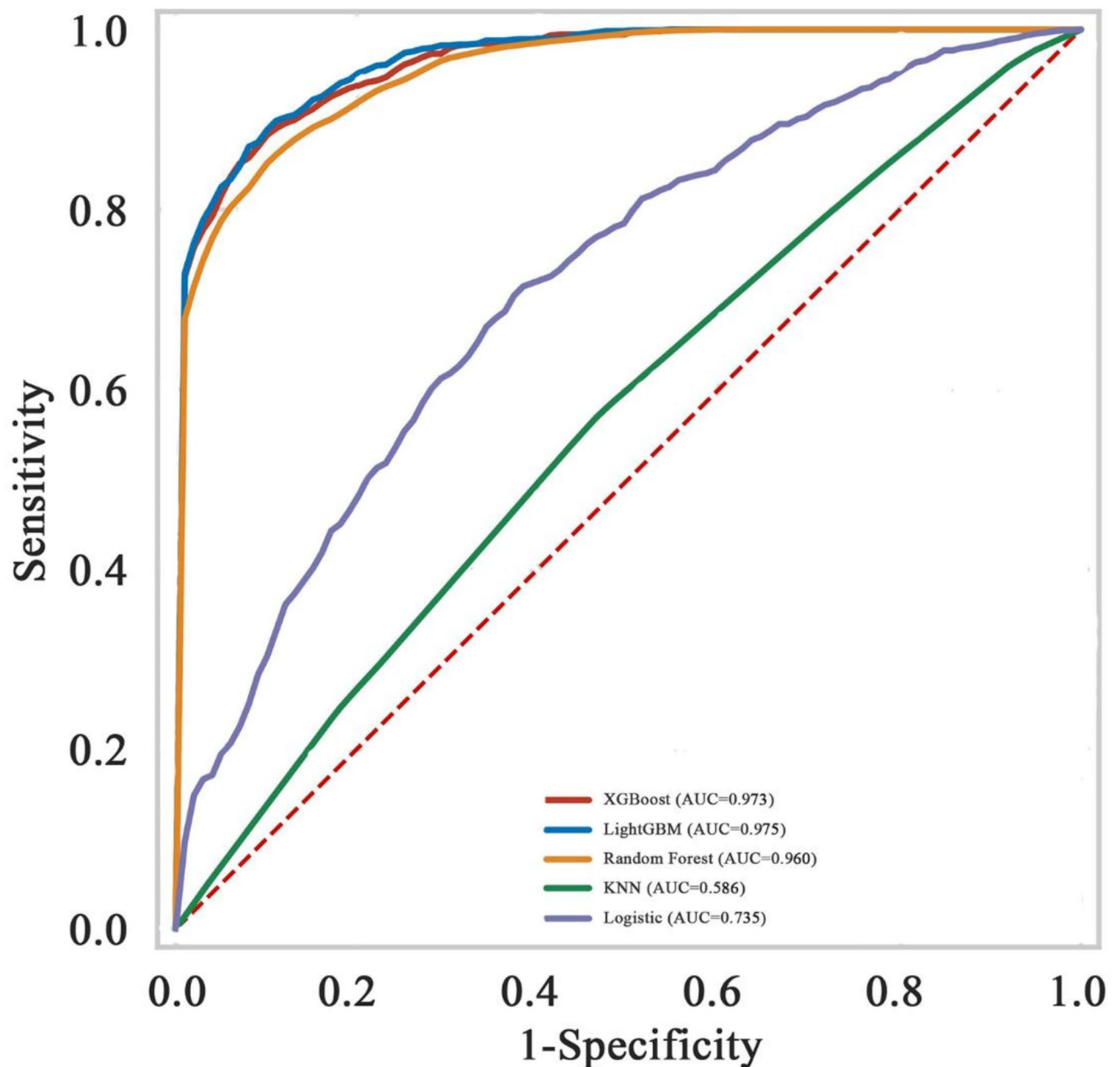


Fig. 2. Receiver operating characteristic curves (ROC) showed the predictions of the five models: XGBoost, LightGBM, Random Forest, Logistic Regression, and KNN.

a poor prognosis with macrolide treatment and a prolonged disease course. Therefore, early identification of SMPP is beneficial for rational treatment, reducing complications, and optimizing the utilization of medical resources. SMPP presents distinct diagnostic challenges due to its wide range of clinical manifestations and overlapping symptoms with other common respiratory diseases. Timely identification of pneumonia etiology can improve clinical management, including decision-making on antibiotic use.

With the development of AI, ML based on prediction models has been widely used for risk prediction and assisting diagnosis in medicine^{25–27}. Increasing studies have proven that ML algorithms have advantages over traditional statistical methods in building prediction models. In the present study, we employ five commonly used machine learning algorithms and systematically evaluate their performance. Through hyperparameter tuning and model selection, we identified the LightGBM algorithm as the best-performing model, achieving superior predictive performance compared to other algorithms. This underscores the significance of algorithm selection and parameter optimization in enhancing model performance. Compared with other diagnostic models, our model demonstrates a comparable high performance ($AUC_{ROC}=0.975$). For example, the AUC of the Nomograph model developed by Chang et al. to predict SMPP in pediatric patients by admission laboratory indicators was 0.777²⁸. Previous studies mainly used traditional calculation tools (nomograms) to establish MPP

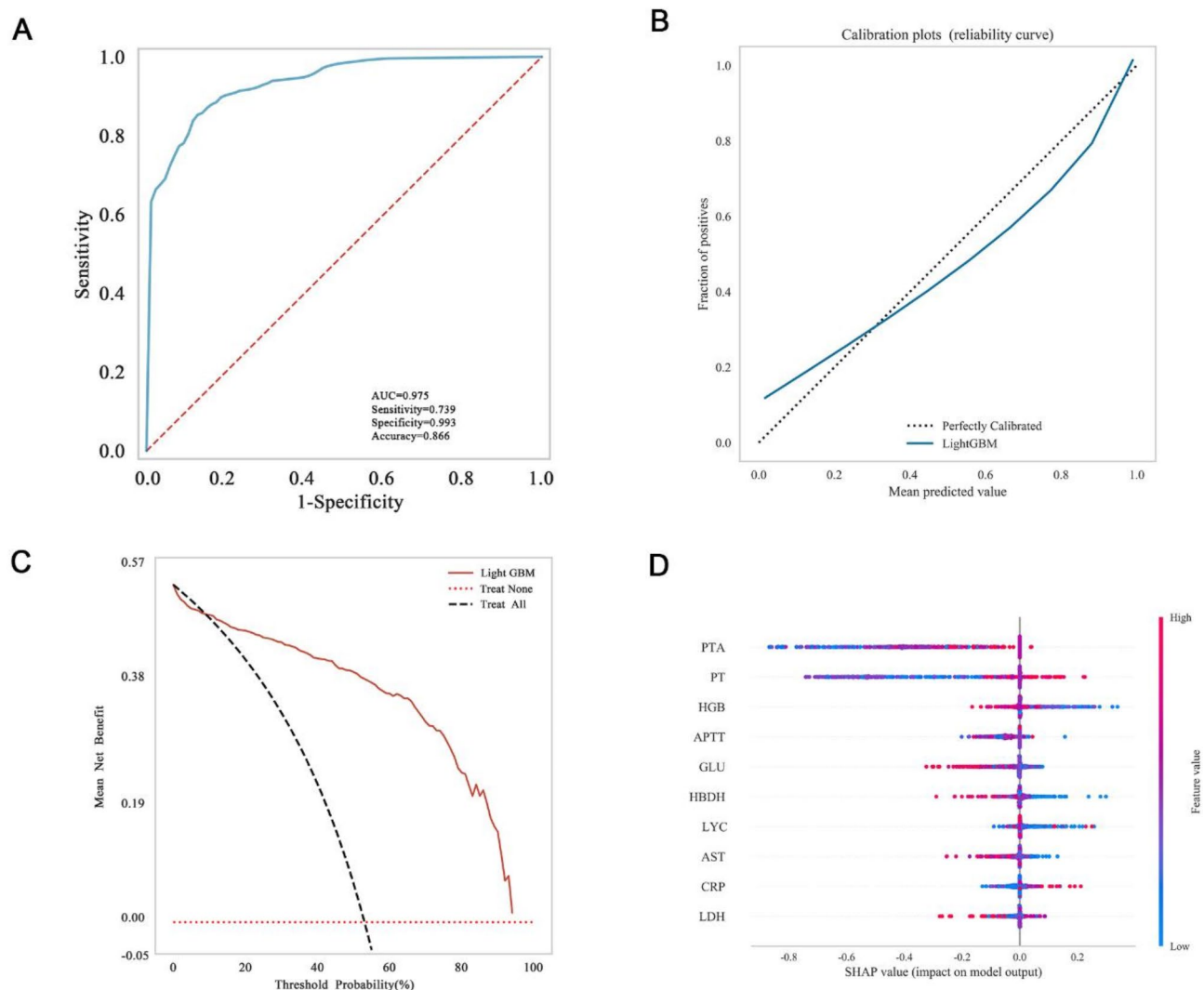


Fig. 3. In the internal validation set, the model 1 built by LightGBM algorithm was evaluated for its ability to differential diagnostic for SMPP and other respiratory diseases. **(A)** Receiver operating characteristic curve analysis; **(B)** calibration curve analysis; **(C)** decision curve analysis; **(D)** the SHAP values.

Groups	Precision	Recall	F1 score	Support
Non-damage	0.781	0.781	0.781	73
Myocardial damage	0.863	0.808	0.834	78
Liver damage	0.797	0.855	0.825	69
Accuracy	0.814	0.814	0.814	0.814
Macro avg	0.814	0.815	0.813	220
Weighted avg	0.815	0.814	0.814	220

Table 3. Performance of the built LightGBM model 2.

prediction models. Therefore, the utilization of complex ML algorithms to establish identification and prediction models for SMPP remains a relatively underexplored area of research.

Numerous studies have reported the correlation between age and MPP²⁹. Children over the age of 5 years are more susceptible to MP infection and exhibit more severe MPP symptoms. The median age of patients with viral pneumonia is significantly lower than that of MPP patients. In modeling cohort, the median age of SMPP patients is 6 years old, which is consistent with the previous finding that MP infection rarely occurs in children under 3 years of age³⁰. Lu et al.³¹ reported that age, LDH, and ESR were significant factors in predicting RMPP using logistic regression. LDH is widely distributed in various tissues of the body, including lung tissue. As a non-specific marker of tissue damage and cell death, serum LDH has long been used for the diagnosis of pulmonary infectious diseases as well as for prognosis prediction³². Consistent with previous reports^{33,34}, the current study

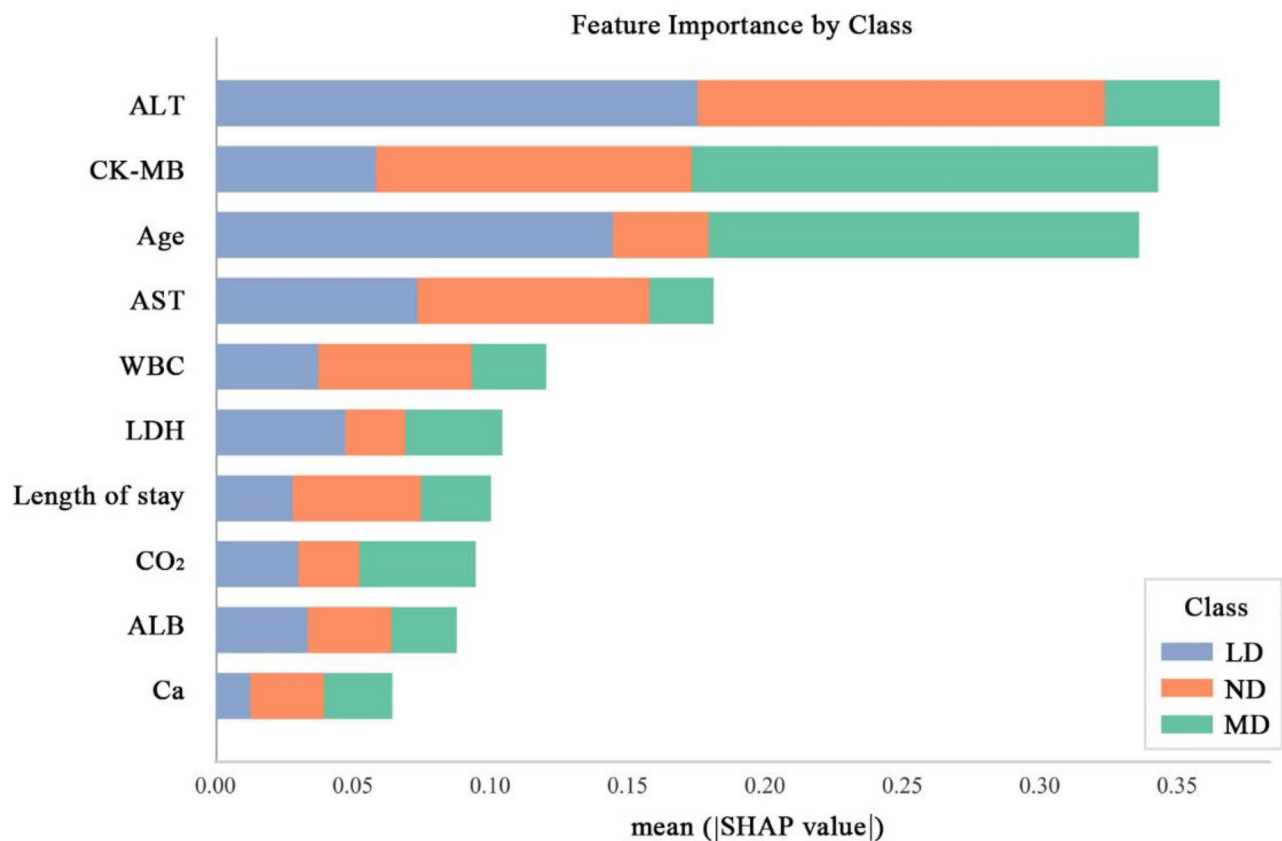


Fig. 4. Feature importance analysis by Shapley additive explanation (SHAP) values. ND non-damage, MD myocardial damage, LD liver damage.

has found that LDH is significantly increased in SMPP patients compared to patients with other respiratory diseases. CRP has also been enrolled in our diagnosis model. CRP is recognized variable of inflammation and has been shown in many studies to be significantly elevated in children with MPP and is associated with disease severity. A study has found that children with CRP > 15.49 mg/L have a higher risk of developing SMPP.

Weights are numerical parameters representing the importance of different features or inputs in a model. In the realm of clinical diseases, researching weights in relation to disease relevancy has become a hot topic. By conducting weight analysis, we find the weight of coagulation indicators, including PTA, PT, and APTT, are significantly prominent in the diagnostic model for SMPP (Model 1). This discovery provides useful references for our in-depth research and further clinical applications. Recent studies have shown that coagulation abnormalities in children with MPP are not uncommon. The specific mechanism of abnormal coagulation function in MP infection is unclear but may be related to MP inducing massive synthesis and secretion of a series of cytokines, such as interleukins, tumor necrosis factors, and chemokines, leading to local vascular damage and accumulation of metabolites in that area, resulting in vascular blockage³⁵. It is also found that abnormal coagulation may be involved in the development of SMPP and may be closely related to the development and prognosis of its complications^{36,37}. This study compared the coagulation function between the SMPP group and the other respiratory diseases group and found that the differences in PTA, PT, and APTT levels were statistically significant. Previous studies have found higher D-dimer levels in children with MPP than in healthy children and higher D-dimer levels in patients with SMPP, particularly those with SMPP with extra-pulmonary complications³⁸. However, our study shows that detecting the levels of coagulation function indices, especially PTA, PT, and APTT, can help the early identification of SMPP and other respiratory diseases patients. Given the limited content of this study, multiple controlled studies and multicentre studies can be conducted in the future to further confirm the findings of this study. It is also possible to analyse the changes in coagulation in SMPP in conjunction with other tests, such as thromboelastography; the underlying mechanisms can be explored to gain insight into the mechanisms of coagulation abnormalities and SMPP.

According to reports, the frequency of extrapulmonary symptoms associated with MP infection has increased in recent decades³⁹. The occurrence of cardiac and liver complications related to MPP has been well confirmed^{40,41}. Cardiac events and liver involvement are the two most common extrapulmonary manifestations, and multiple factors are involved in the pathogenesis of these conditions. A study found that TIM1 is associated with CK-MB, whereas TIM3 and TLR2 are associated with ALT, indicating that cardiac and liver damage caused by MP infection results from a combination of inflammatory cytokines and autoimmune reactions. In the risk predictive model of organ damage for SMPP, age emerged as an important predictor. Li et al. reported that MP

infections cause more serious myocardial damage in children aged 13–36 months and 72 months–14 years as detected using serum CK-MB concentrations⁴². In addition, our study found that the median age of SMPP children with myocardial damage was significantly lower than that of SMPP children with liver damage. These findings suggest that paediatrician should pay more attention to age-specific differences in the extrapulmonary complications associated with MP infections.

Myocardial enzymes are the main serum enzymes used for the clinical diagnosis of MPP complicated with myocardial damage, including CK, LDH, CK-MB, and AST⁴³. The fact that AST and LDH can be detected in a variety of tissues results in a lack of specificity. As a myocardium-specific enzymatic indicator, CK-MB is scarcely found in other tissues. The change in CK-MB activity is closely associated with the necrosis of myocardial cells⁴⁴. In Cohort 2 of this study, we found that the levels of CK-MB in SMPP patients with myocardial damage are significantly higher than those in patients with liver damage or non-damage. Additionally, among pediatric patients with extrapulmonary damage, except for myocardial damage, the CK-MB levels in those with liver damage are also higher than in those non-damage. A study shows asymptomatic elevation of liver enzymes in MPP patients⁴⁵. ALT levels increase significantly after infection, indicating liver involvement during the disease process. In Cohort 2, the ALT levels in patients with liver damage are significantly higher than those in patients with myocardial damage and non-damage. In this study, the feature importance analysis reveals that ALT and CK-MB are among the top contributing variables to the Model 2. Additionally, parameters such as AST, WBC, and LDH rank among the top ten in feature importance scores and serve as crucial predictors of organ damage for SMPP in children. In contrast to the prevalent binary classification methods employed by existing ML models, our Model 2 represents a pioneering effort in the realm of multiclass classification. A single model can provide the risk probabilities of SMPP associated with three distinct organ injury types, which enhances clinical operability and applicability.

This study employs readily available and simple indicators for modelling, aiming to maximize the clinical applicability. However, there are several limitations: (1) Our research cohorts are relatively homogenous, although the performance of models are promising, they may vary across different healthcare settings and patient populations. Therefore, future studies should aim to validate these models across diverse populations to explain potential confounding factors such as ethnicity. Moreover, we plan to use datasets from diverse sources and geographical locations for external validation in the future, which is crucial for evaluating the generalizability of the model and ensuring its applicability in varied clinical scenarios. (2) Due to the limited technology and resources available, this study is unable to demonstrate the ensemble learning effect of multiple machine learning models. In the future, our studies may explore using more advanced algorithms, such as deep learning or ensemble learning algorithms, to further improve prediction accuracy.

In conclusion, the machine learning algorithm based on laboratory parameters can provide clearer decision-making guidance for SMPP differential diagnosis and effectively predict the occurrence of organ damage in SMPP children, providing strong references for clinical diagnosis and treatment. It is worth further clinical research and promotion.

Data availability

The datasets used and/or analyzed during the current study can be obtained from the corresponding author on reasonable request.

Received: 28 June 2024; Accepted: 25 February 2025

Published online: 19 March 2025

References

- Kutty, P. K. et al. Mycoplasma pneumoniae among children hospitalized with community-acquired pneumonia. *Clin. Infect. Dis.* **68**, 5–12 (2019).
- Li, Z. J. et al. Etiological and epidemiological features of acute respiratory infections in China. *Nat. Commun.* **12**, 5026 (2021).
- Esposito, S., Argentiero, A., Gramegna, A. & Principi, N. Mycoplasma pneumoniae: a pathogen with unsolved therapeutic problems. *Expert Opin. Pharmacother.* **22**, 1193–1202 (2021).
- Atkinson, T. P., Balish, M. F. & Waites, K. B. Epidemiology, clinical manifestations, pathogenesis and laboratory detection of mycoplasma pneumoniae infections. *Fems Microbiol. Rev.* **32**, 956–973 (2008).
- D'Alonzo, R. et al. Pathogenesis and treatment of neurologic diseases associated with mycoplasma pneumoniae infection. *Front. Microbiol.* **9**, 2751 (2018).
- Lee, K. L. et al. Severe mycoplasma pneumoniae pneumonia requiring intensive care in children, 2010–2019. *J. Formos. Med. Assoc.* **120**, 281–291 (2021).
- Khoury, T. et al. Increased rates of intensive care unit admission in patients with mycoplasma pneumoniae: a retrospective study. *Clin. Microbiol. Infect.* **22**, 711–714 (2016).
- Gao, L. W. et al. The epidemiology of paediatric mycoplasma pneumoniae pneumonia in North China: 2006 to 2016. *Epidemiol. Infect.* **147**, e192 (2019).
- Guo, D. X. et al. Epidemiology and mechanism of drug resistance of mycoplasma pneumoniae in Beijing, China: a multicenter study. *Bosn J. Basic. Med. Sci.* **19**, 288–296 (2019).
- Ma, X. et al. Development of a Dna microarray assay for rapid detection of fifteen bacterial pathogens in pneumonia. *Bmc Microbiol.* **20**, 177 (2020).
- Saraya, T. Mycoplasma pneumoniae infection: basics. *J. Gen. Fam. Med.* 18118–18125 (2017).
- Meseguer, M. A. et al. Mycoplasma pneumoniae pericarditis and cardiac tamponade in a ten-year-old Girl. *Pediatr. Infect. Dis. J.* **15**, 829–831 (1996).
- Juarez, D., Juan, L. Z., Shrestha, S. & Ballgobin, A. Pattern and significance of asymptomatic elevation of liver enzymes in mycoplasma pneumonia in children. *Clin. Pediatr. (Phila)* **57**, 57–61 (2018).
- Poddighe, D. Mycoplasma pneumoniae-related hepatitis in children. *Microb. Pathog.* **139**, 103863 (2020).
- Song, W. J. et al. Pediatric mycoplasma pneumoniae infection presenting with acute cholestatic hepatitis and other extrapulmonary manifestations in the absence of pneumonia. *Pediatr. Gastroenterol. Hepatol. Nutr.* **20**, 124–129 (2017).

16. Obermeyer, Z. & Emanuel, E. J. Predicting the future—Big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **375**, 1216–1219 (2016).
17. Zeng, Q. et al. Epidemiological characteristics and early predict model of children Mycoplasma Pneumoniae pneumonia outbreaks after the COVID-19 in Shandong. *Sci. Rep.* **14**, 19892 (2024).
18. Zhang, X., Sun, R., Jia, W., Li, P. & Song, C. A new dynamic nomogram for predicting the risk of severe mycoplasma pneumoniae pneumonia in children. *Sci. Rep.* **14**, 8260 (2024).
19. Li, L. et al. Construction and validation of a nomogram model to predict the severity of mycoplasma pneumoniae pneumonia in children. *J. Inflamm. Res.* **17**, 1183–1191 (2024).
20. Zhang, H. & Li, H. X. Risk factors analysis and predictive model construction for liver damage in children with Mycoplasma pneumoniae pneumonia. *Infect. Dis. Info* **37**, 459–463 (2024).
21. Liu, Y. et al. Nomogram and machine learning models predict 1-year mortality risk in patients with sepsis-induced cardiorenal syndrome. *Front. Med. (Lausanne)* **9**, 792238 (2022).
22. Peng, X. et al. A preliminary prediction model of pediatric Mycoplasma pneumoniae pneumonia based on routine blood parameters by using machine learning method. *BMC Infect. Dis.* **24**, 707 (2024).
23. Bădăraș, D. O. *Declaration of Helsinki* (Mental Health Practitioner's Guide to HIV/AIDS, 2013).
24. Zhang, X., Sun, R., Jia, W., Li, P. & Song, C. Clinical characteristics of lung consolidation with mycoplasma pneumoniae pneumonia and risk factors for mycoplasma pneumoniae necrotizing pneumonia in children. *Infect. Dis. Ther.* **13**, 329–343 (2024).
25. Jayapandian, C. P. et al. Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney Int.* **99**, 86–101 (2021).
26. Segar, M. W. et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation* **143**, 2370–2383 (2021).
27. Zheng, J. et al. A multicenter study to develop a non-invasive radiomic model to identify urinary infection stone in vivo using machine-learning. *Kidney Int.* **100**, 870–880 (2021).
28. Chang, Q. et al. Prediction model for severe Mycoplasma pneumoniae pneumonia in pediatric patients by admission laboratory indicators. *J. Trop. Pediatr.* **68**, fmac059 (2022).
29. Jiang, C., Bao, S., Shen, W. & Wang, C. Predictive value of immune-related parameters in severe Mycoplasma pneumoniae pneumonia in children. *Transl. Pediatr.* **13**, 1521–1528 (2024).
30. Alvaro, V. A., Aguinaga, P. A., Navascues, O. A., Castilla, J. & Ezpeleta, B. C. Clinical characteristics of patients with mycoplasma pneumoniae infection. *Enferm Infecc Microbiol. Clin. (Engl. Ed.)* 40449–40452 (2022).
31. Lu, A., Wang, C., Zhang, X., Wang, L. & Qian, L. Lactate dehydrogenase as a biomarker for prediction of refractory mycoplasma pneumoniae pneumonia in children. *Respir. Care* **60**, 1469–1475 (2015).
32. Esteves, F. et al. (1-3)-beta-d-glucan in association with lactate dehydrogenase as biomarkers of Pneumocystis pneumonia (pcp) in hiv-infected patients. *Eur. J. Clin. Microbiol. Infect. Dis.* **33**, 1173–1180 (2014).
33. Moynihan, K. M. et al. Severe mycoplasma pneumoniae infection in children admitted to pediatric intensive care. *Pediatr. Infect. Dis. J.* **37**, e336–e338 (2018).
34. Liu, T. Y. et al. Serum lactate dehydrogenase isoenzymes 4 plus 5 is a better biomarker than total lactate dehydrogenase for refractory mycoplasma pneumoniae pneumonia in children. *Pediatr. Neonatol.* **59**, 501–506 (2018).
35. Sarathchandran, P. A. M. A., Alboudi, A. M. & Inshasi, J. Mycoplasma pneumoniae infection presenting as stroke and meningoencephalitis with aortic and subclavian aneurysms without pulmonary involvement. *BMJ Case Rep.* **2018** (2018).
36. Qiu, J., Ge, J. & Cao, L. D-dimer: the risk factor of children's severe mycoplasma pneumoniae pneumonia. *Front. Pediatr.* **10**, 828437 (2022).
37. Huang, X. et al. Clinical significance of d-dimer levels in refractory mycoplasma pneumoniae pneumonia. *Bmc Infect. Dis.* **21**, 14 (2021).
38. Li, T. et al. Evaluation of variation in coagulation among children with Mycoplasma pneumoniae pneumonia: a case-control study. *J. Int. Med. Res.* **45**, 2110–2118 (2017).
39. Lind, K. Manifestations and complications of mycoplasma pneumoniae disease: a review. *Yale J. Biol. Med.* **56**, 461–468 (1983).
40. Chen, C. J. et al. Mycoplasma pneumoniae infection presenting as neutropenia, thrombocytopenia, and acute hepatitis in a child. *J. Microbiol. Immunol. Infect.* **37**, 128–130 (2004).
41. Chang, J. H. et al. A case of acute hepatitis with mycoplasma pneumoniae infection and transient depression of multiple coagulation factors. *Yonsei Med. J.* **49**, 1055–1059 (2008).
42. Li, C. M. et al. Age-specific Mycoplasma pneumoniae pneumonia-associated myocardial damage in children. *J. Int. Med. Res.* **41**, 1716–1723 (2013).
43. Barski, L., Nevzorov, R., Horowitz, J. & Horowitz, S. Antibodies to various Mycoplasmas in patients with coronary heart disease. *Isr. Med. Assoc. J.* **12**, 396–399 (2010).
44. Youn, Y. S. et al. Difference of clinical features in childhood mycoplasma pneumoniae pneumonia. *Bmc Pediatr.* **10**, 48 (2010).
45. Qi, X., Sun, X., Li, X., Kong, D. & Zhao, L. Significance changes in the levels of myocardial enzyme in the child patients with mycoplasma pneumoniae pneumonia. *Cell. Mol. Biol.* **66**, 41–45 (2020).

Acknowledgements

The authors thank Xialin Wang and Yuming Cheng for assistance with data sorting.

Author contributions

HB contributed to data curation, formal analysis, and writing of the original draft. LXW contributed to Investigation and methodology. DRR and YH contributed to methodology. ZQ contributed to supervision. XCY, SCM, ZB, ZHL, and YXQ contributed to data collection. XJC contributed to conceptualization, reviewing and editing, and resources. All authors approved the final manuscript as submitted.

Funding

This work was supported by the Project of the Jilin Science and Technology Development Program (20220401085YY).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-92089-3>.

Correspondence and requests for materials should be addressed to J.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025