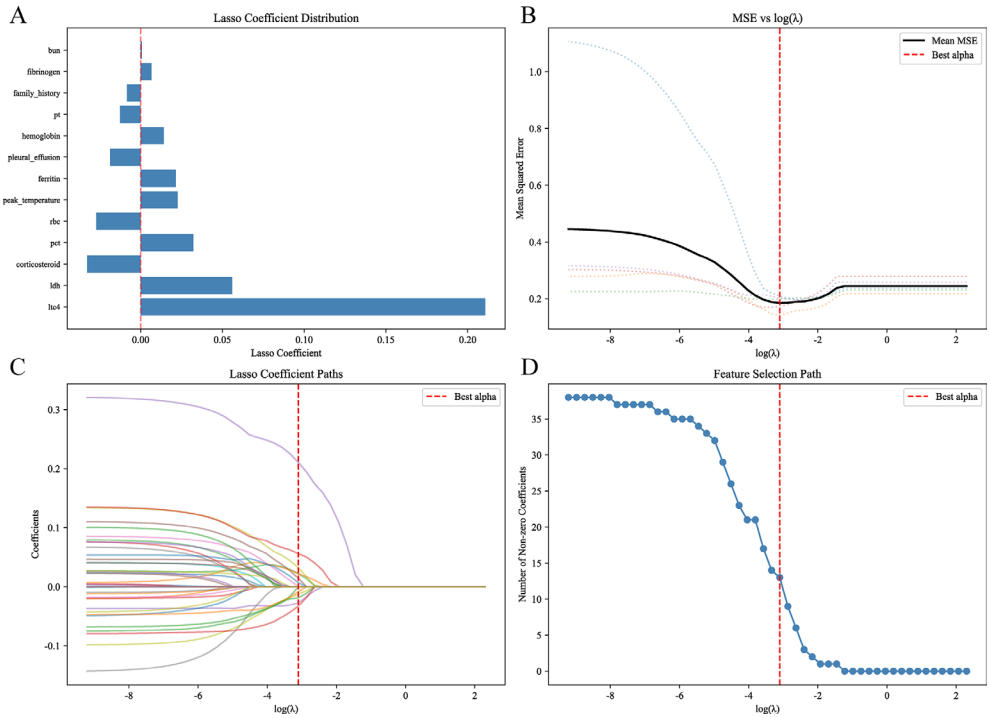

Figure S1 LASSO regression for feature selection and dimensionality reduction. (A) Distribution of LASSO coefficients for the 13 features retained at the optimal lambda. LTC4 exhibited the highest predictive weight (β=0.211), followed by LDH (β=0.056). (B) Selection of the optimal tuning parameter (lambda) using 10-fold cross-validation. The red dashed line indicates the lambda value that minimizes the MSE. (C) LASSO coefficient shrinkage paths, illustrating the profile of coefficients as the penalty term increases. (D) Feature selection flow: the model successfully reduced the feature space from 69 candidate predictors, VIF screening retained 46 features, and LASSO regression further reduced these to 13 predictors with non-zero coefficients at the optimal lambda. BUN, blood urea nitrogen; LASSO, least absolute shrinkage and selection operator; LDH, lactate dehydrogenase; LTC4, leukotriene C4; MSE, mean squared error; PCT, plateletocrit; PT, prothrombin time; RBC, red blood cell; VIF, variance inflation factor.

https://dx.doi.org/10.21037/tp-2026-0485

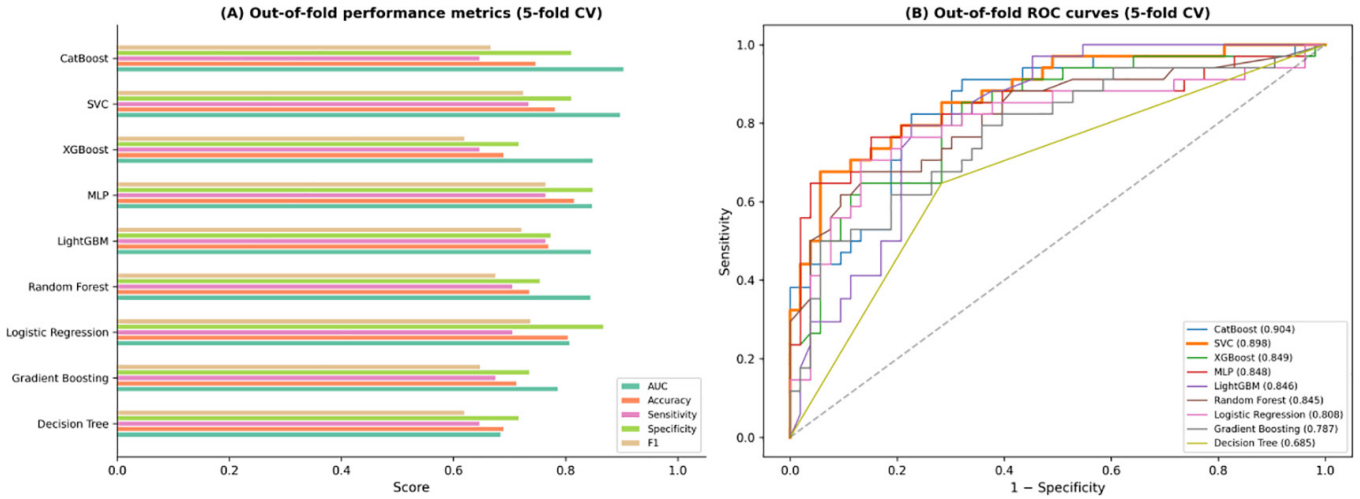

Figure S2 Performance comparison of the nine machine-learning models under 5-fold cross-validation. (A) Out-of-fold performance across five metrics (AUC, accuracy, sensitivity, specificity, and F1 score) for all candidate models under 5-fold cross-validation. (B) Corresponding out-of-fold ROC curves; the SVC offered the best balance between discrimination and generalizability. AUC, area under the curve; CV, cross-validation; LightGBM, light gradient boosting machine; MLP, multilayer perceptron; ROC, receiver operating characteristic; SVC, support vector classifier; XGBoost, extreme gradient boosting.

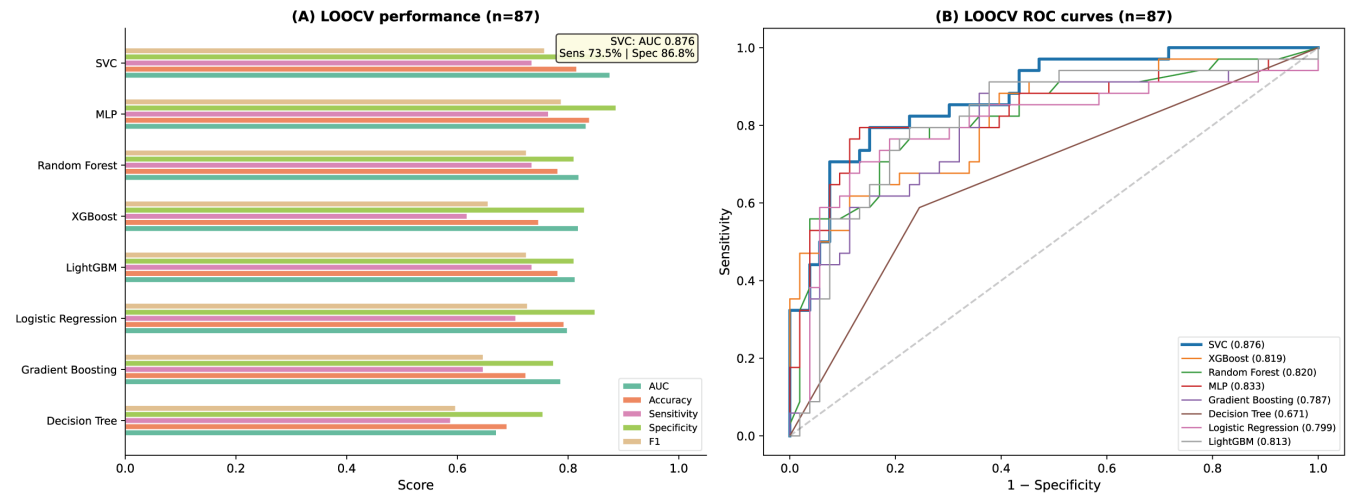

Figure S3 Robustness of the nine machine-learning models under leave-one-out cross-validation (n=87). (A) Performance of the nine models under leave-one-out cross-validation (n=87); the SVC achieved the highest AUC while maintaining balanced sensitivity (73.5%) and specificity (86.8%). (B) Corresponding LOOCV ROC curves; the parsimonious two-feature SVC performed comparably to or better than more complex ensemble models, supporting the feature-parsimonious strategy. AUC, area under the curve; CatBoost, categorical boosting; LightGBM, light gradient boosting machine; LOOCV, leave-one-out cross-validation; MLP, multilayer perceptron; ROC, receiver operating characteristic; SVC, support vector classifier; XGBoost, extreme gradient boosting.

https://dx.doi.org/10.21037/tp-2026-0485

Table S1 Detailed baseline characteristics of the overall study cohort (n=158)

Variable	Overall (n=158)
Male	79 (50.0)
Age (years)	6.71 (4.17–8.65)
BMI (kg/m^2)	14.81 (13.84–16.00)
Fever	158 (100.0)
Fever duration (days)	5.00 (3.00–7.00)
Peak temperature (°C)	39.00 (38.60–39.50)
Cough	158 (100.0)
Wheezing at baseline	49 (31.0)
Rash	18 (11.4)
Atopic history	26 (16.5)
MP-IgM positive (≥1:1600)	94 (59.5)
MP-DNA load (log_{10} copies/mL)	5.12 (4.15–6.62)
Macrolide resistance gene	32 (20.3)
FeNO (ppb)	16.29 (13.66–20.66)
Eosinophil (%)	0.70 (0.40–1.20)
BALF LTC4 (pg/mL)	41.72 (30.31–71.84)
WBC (×10^9/L)	7.83 (6.22–10.24)
CRP (mg/L)	18.07 (9.59–33.34)
IL-6 (pg/mL)	126.59 (39.53–227.42)
ESR (mm/h)	39.03 (31.00–55.00)
Number of involved lobes	1.00 (1.00–2.00)
Severe MPP	110 (69.6)
Refractory MPP	63 (39.9)
Plastic bronchitis	32 (20.3)
Mucosal necrosis	1 (0.6)
Hospital stay (days)	7.00 (7.00–9.00)

Continuous variables are presented as median (interquartile range); categorical variables are presented as n (%). Microbiologic positivity for *Mycoplasma pneumoniae* was defined as MP-IgM titer ≥1:1600. Macrolide-resistance status was determined by the presence of the 23S rRNA A2063G or A2064G point mutation. BALF, bronchoalveolar lavage fluid; BMI, body mass index; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FeNO, fractional exhaled nitric oxide; IL-6, interleukin-6; LTC4, leukotriene C4; MP, *Mycoplasma pneumoniae*; MP-DNA, *Mycoplasma pneumoniae* DNA; MP-IgM, *Mycoplasma pneumoniae* immunoglobulin M; MPP, *Mycoplasma pneumoniae* pneumonia; WBC, white blood cell.

https://dx.doi.org/10.21037/tp-2026-0485

Table S2 VIF analysis for multicollinearity assessment

Rank	Feature	VIF	Status	Final model
1	Weight	254.89536848017696	Excluded	
2	Height	206.05068858406665	Excluded	
3	AST	69.77176512937007	Excluded	
4	BMI	62.52202769163888	Excluded	
5	ALT	59.95621815619313	Excluded	
6	WBC	53.823424437805286	Excluded	
7	Neutrophil	37.59752700962274	Excluded	
8	Age	34.29979356441158	Excluded	
9	Pulmonary complication	30.98065475812496	Excluded	
10	Plastic bronchitis	27.233797677299155	Excluded	
11	Macrolide resistance gene	26.317682973228795	Excluded	
12	Refractory MPP	22.27186252339259	Excluded	
13	Resistance type	18.12884363014825	Excluded	
14	Eosinophil count	15.647371460742322	Excluded	
15	Mixed infection	15.186827263518293	Excluded	
16	Eosinophil %	13.906016502096326	Excluded	
17	CK	13.673522396069965	Excluded	
18	Lymphocyte	13.262144352122268	Excluded	
19	Severe MPP	11.643707802703007	Excluded	
20	MP-DNA load (log_{10})	10.386798020595409	Excluded	
21	Wheezing (baseline)	9.916417226242707	Retained	
22	IVIG use	9.720083335731928	Retained	
23	Corticosteroid use	9.166297116957512	Retained	
24	Ferritin	8.953476495043322	Retained	
25	CRP	8.931960834240972	Retained	
26	Plastic mucus	8.846477665857115	Retained	
27	LDH	8.039625649725846	Retained	✓
28	Extrapulmonary complication	7.484009418238921	Retained	
29	BALF LTC4	7.338038216960641	Retained	✓
30	Hospital stay (days)	7.216847633003432	Retained	
31	Pleural effusion	7.062423092008655	Retained	
32	Fibrinogen	6.270663506417231	Retained	
33	Creatinine	5.731591627211022	Retained	
34	FeNO (baseline)	5.139884764597506	Retained	
35	MP-DNA load	4.800665055293353	Retained	
36	Family allergy history	4.697701242023854	Retained	
37	Lobes involved	4.56194323065394	Retained	
38	Rash	4.470311441556796	Retained	

Table S2 (*continued*)

https://dx.doi.org/10.21037/tp-2026-0485

Table S2 (*continued*)

Rank	Feature	VIF	Status	Final model
39	IL-6	4.337420649413126	Retained	
40	PCT	4.154062044823534	Retained	
41	Atopic history	4.089987113336108	Retained	
42	Bronchoscopy frequency	4.064775095398261	Retained	
43	Hemoglobin	4.03981137583146	Retained	
44	D-dimer	3.8818554249602424	Retained	
45	Sex	3.853093149157076	Retained	
46	PT	3.711512737056931	Retained	
47	ESR	3.699190904223572	Retained	
48	RBC	3.626834781609856	Retained	
49	INR	3.6040882501962885	Retained	
50	Albumin	3.159791900806443	Retained	
51	APTT	3.132415705018962	Retained	
52	BUN	3.083485865177538	Retained	
53	Platelet	2.8976218807600054	Retained	
54	CK-MB	2.839440008697343	Retained	
55	Fever duration	2.768450429446471	Retained	
56	Peak temperature	2.6099958552084064	Retained	
57	Onset month	2.573831999420046	Retained	
58	MP-IgM high	2.3802838937565918	Retained	
59	Cough		Retained	
60	Congestion edema		Retained	
61	Mucosal necrosis		Retained	
62	Cartilage exposure		Retained	
63	Thickened texture		Retained	
64	Ground glass		Retained	
65	Patchy shadow		Retained	
66	Consolidation		Retained	
67	Pulmonary embolism		Retained	
68	Necrotizing pneumonia		Retained	
69	Asthma exacerbation		Retained	

VIF was calculated for all 69 candidate numeric predictors; 20 features with VIF >10 were excluded due to high multicollinearity and 3 additional features were excluded due to insufficient variance, leaving 46 features for subsequent LASSO regression. The final two-feature model used in the support vector classifier comprised BALF LTC4 and serum LDH (marked with ✓). ALT, alanine aminotransferase; APTT, activated partial thromboplastin time; AST, aspartate aminotransferase; BALF, bronchoalveolar lavage fluid; BMI, body mass index; BUN, blood urea nitrogen; CK, creatine phosphokinase; CRP, C-reactive protein; D-dimer, D-dimer (fibrin degradation fragment); ESR, erythrocyte sedimentation rate; FeNO, fractional exhaled nitric oxide; IL-6, interleukin-6; INR, international normalized ratio; IVIG, intravenous immune globulin; LDH, lactate dehydrogenase; LTC4, leukotriene C4; MP-DNA, *Mycoplasma pneumoniae* DNA; MP-IgM, *Mycoplasma pneumoniae* immunoglobulin M; MPP, *Mycoplasma pneumoniae* pneumonia; PCT, procalcitonin; PT, prothrombin time; RBC, red blood cell; VIF, variance inflation factor; WBC, white blood cell.

https://dx.doi.org/10.21037/tp-2026-0485

Table S3 Feature weights and coefficients from LASSO regression

Rank	Feature	Coefficient (β)	Status	Final model
1	BALF LTC4	0.211	Non-zero	✓
2	LDH	0.056	Non-zero	✓
3	Corticosteroid use	–0.033	Non-zero	
4	PCT	0.032	Non-zero	
5	RBC	–0.027	Non-zero	
6	Peak temperature	0.023	Non-zero	
7	Ferritin	0.021	Non-zero	
8	Pleural effusion	–0.019	Non-zero	
9	Hemoglobin	0.014	Non-zero	
10	PT	–0.013	Non-zero	
11	Family allergy history	–0.008	Non-zero	
12	Fibrinogen	0.007	Non-zero	
13	BUN	0.001	Non-zero	
14	Atopic history	0	Zero	
15	Rash	0	Zero	
16	Plastic mucus	0	Zero	
17	Mucosal necrosis	0	Zero	
18	Cartilage exposure	0	Zero	
19	Congestion/edema	0	Zero	
20	Consolidation	0	Zero	
21	Thickened bronchial texture	0	Zero	
22	Ground-glass opacity	0	Zero	
23	Patchy shadow	0	Zero	
24	Cough	0	Zero	
25	IVIG use	0	Zero	
26	Pulmonary embolism	0	Zero	
27	Necrotizing pneumonia	0	Zero	
28	Asthma exacerbation	0	Zero	
29	Wheezing (baseline)	0	Zero	
30	Onset month	0	Zero	
31	MP-IgM high titer	0	Zero	
32	Albumin	0	Zero	
33	FeNO (baseline)	0	Zero	
34	Platelet	0	Zero	
35	CRP	0	Zero	

Table S3 (*continued*)

https://dx.doi.org/10.21037/tp-2026-0485

Table S3 (*continued*)

Rank	Feature	Coefficient (β)	Status	Final model
36	IL-6	0	Zero	
37	ESR	0	Zero	
38	MP-DNA load	0	Zero	
39	CK-MB	0	Zero	
40	Creatinine	0	Zero	
41	Sex	0	Zero	
42	APTT	0	Zero	
43	D-dimer	0	Zero	
44	INR	0	Zero	
45	Lobes involved	0	Zero	
46	Fever duration	0	Zero	

Standardized regression coefficients (β) were obtained from LASSO logistic regression with the optimal penalty parameter (λ) selected by 10-fold cross-validation. Thirteen features had non-zero coefficients; the remaining 33 features had coefficients shrunk to zero. Recursive feature elimination subsequently retained BALF LTC4 and serum LDH (marked with ✓) as the final two predictors used in the support vector classifier. APTT, activated partial thromboplastin time; BALF, bronchoalveolar lavage fluid; BUN, blood urea nitrogen; CK, creatine phosphokinase; CRP, C-reactive protein; D-dimer, D-dimer (fibrin degradation fragment); ESR, erythrocyte sedimentation rate; FeNO, fractional exhaled nitric oxide; IL-6, interleukin-6; INR, international normalized ratio; IVIG, intravenous immune globulin; LASSO, least absolute shrinkage and selection operator; LDH, lactate dehydrogenase; LTC4, leukotriene C4; MP-DNA, *Mycoplasma pneumoniae* DNA; MP-IgM, *Mycoplasma pneumoniae* immunoglobulin M; PCT, procalcitonin; PT, prothrombin time; RBC, red blood cell.

Table S4 Performance of the nine machine-learning models under 5-fold cross-validation (out-of-fold)

Rank	Model	OOF AUC		Accuracy	Sensitivity	Specificity	PPV	NPV	F1	Brier score
		Mean	SD							
1	CatBoost	0.904	0.0447	0.747	0.647	0.811	0.688	0.782	0.667	0.16
2	SVC	0.898	0.0245	0.782	0.735	0.811	0.714	0.827	0.725	0.142
3	XGBoost	0.849	0.0525	0.69	0.647	0.717	0.595	0.76	0.62	0.211
4	MLP	0.848	0.0765	0.816	0.765	0.849	0.765	0.849	0.765	0.15
5	LightGBM	0.846	0.0316	0.77	0.765	0.774	0.684	0.837	0.722	0.175
6	Random Forest	0.845	0.0422	0.736	0.706	0.755	0.649	0.8	0.676	0.179
7	Logistic Regression	0.808	0.0643	0.805	0.706	0.868	0.774	0.821	0.738	0.166
8	Gradient Boosting	0.787	0.0476	0.713	0.676	0.736	0.622	0.78	0.648	0.248
9	Decision Tree	0.685	0.0411	0.69	0.647	0.717	0.595	0.76	0.62	0.31

Out-of-fold performance metrics for the nine candidate models obtained from 5-fold stratified cross-validation on the full analytic cohort (n=87), using BALF LTC4 and serum LDH as input features; hyperparameters were tuned by grid search. Models are ranked by mean OOF AUC. The SVC was selected for its balance of discrimination (AUC) and calibration (Brier score), and was then evaluated by leave-one-out cross-validation (*Table S5*). AUC, area under the curve; CatBoost, categorical boosting; LightGBM, light gradient boosting machine; MLP, multilayer perceptron; NPV, negative predictive value; OOF, out-of-fold; PPV, positive predictive value; SD, standard deviation; SVC, support vector classifier; XGBoost, extreme gradient boosting.

https://dx.doi.org/10.21037/tp-2026-0485