

## Original Article

# Machine learning algorithms for the clinical diagnosis of acute atypical asthma exacerbation

Feng Ma, Weihua Zhu, Piping Jiang, Xuelian Bai, Wenya Li

*Department of Respiratory and Critical Care Medicine, Aerospace Center Hospital, Beijing 100049, China*

Received October 31, 2024; Accepted April 8, 2025; Epub April 15, 2025; Published April 30, 2025

**Abstract:** Objective: To develop a predictive model for the clinical diagnosis of acute atypical asthma attacks using machine learning algorithms and investigate the risk factors related to the diagnosis of atypical asthma. Methods: This study retrospectively collected data on characteristics, symptoms, general examinations, pulmonary functional tests, and FeNO results of patients in the Aerospace Center Hospital. Five machine learning algorithms (logistic regression, decision tree, random forest, support vector machine, extreme gradient boosting) were employed to select variables for predicting outpatient cases of atypical asthma exacerbation in routine practice. A predictive model for diagnosing atypical asthma exacerbation was then developed, optimized, and subjected to explanatory analysis. Results: After screening, 214 cases were included, with 98 diagnosed with acute exacerbation of atypical asthma and 116 undiagnosed. All patients were randomly assigned into a training set (n=149) or a validation set (n=65) at a ratio of 7:3. The predictive capabilities of five models were evaluated in the validation set. This demonstrated that all models could effectively identify patients with acute exacerbation of atypical asthma; among them, Logistic regression, random forest, and extreme gradient boosting achieved accuracies of 93.1%, and extreme gradient boosting reached 95.4%. The logistic regression model showed the best predictive performance. Model interpretation analysis revealed that FeNO, EOS, FEV1 variability, history of allergic rhinitis, and wheezing during acute attacks were significant risk factors for predicting acute exacerbations of atypical asthma. Conclusions: The application of machine learning methods for variable selection in predicting acute exacerbations of atypical asthma has shown promising results. FeNO, EOS, FEV1 variability, history of allergic rhinitis, and wheezing during acute episodes were crucial predictors of exacerbations.

**Keywords:** Atypical asthma, machine learning, diagnosis, prediction model

## Introduction

Atypical asthma (VA) is a variant of bronchial asthma characterized by symptoms such as coughing, chest tightness, and chest pain, without prominent wheezing and shortness of breath. Symptoms worsen notably upon exposure to cold air, allergens (such as pollen, dust mites), or irritants (such as perfume, cooking fumes), demonstrating significant airway hyperresponsiveness [1].

The global prevalence of bronchial asthma is as high as 358 million [2], with a prevalence of 4.2% among individuals aged over 20 years in China, and affecting a total of 45.7 million people [3]. Underdiagnosis of asthma is widespread globally, with misdiagnosis rates ranging from 19.2% to 73.3% across countries [4].

In urban areas of China, the overall asthma control rate is only 28.5%, with 71.2% of asthma patients remaining undiagnosed, a phenomenon particularly common with atypical asthma [3]. Therefore, early and accurate diagnosis of atypical asthma variants, such as cough variant asthma (CVA), chest tightness variant asthma (CTVA), and hidden asthma (HA), are crucial for improving asthma diagnosis and management [4].

Machine learning (ML), a type of artificial intelligence, has been widely used for disease diagnosis, treatment recommendations, and patient management [5, 6], providing a more precise approach for diagnosing bronchial asthma [7-9]. While the potential value of ML in diagnosing and treating other diseases is recognized, its application in diagnosing atypical

asthma remained poorly understood. Therefore, in the present study, we retrospectively collected clinical data, aiming to develop a predictive model for diagnosing acute exacerbations of atypical asthma using five machine learning algorithms. The optimal model was selected for interpretation to explore the risk factors related to acute exacerbation of atypical asthma, aiming to facilitate the rapid early diagnosis of patients experiencing acute exacerbations.

### Materials and methods

#### *Data sources*

This study enrolled patients who visited the Aerospace Center Hospital and underwent relevant examinations from June 2023 to February 2024. Inclusion criteria: ① age over 14 years; ② first-time visitors; ③ presenting with symptoms such as cough, chest tightness, chest pain, wheezing, and dyspnea. Exclusion criteria: ① non-first-time visitors or follow-up patients; ② acute exacerbation of acute or chronic respiratory tract infections; ③ acute episodes of cardiovascular diseases, hepatic or renal dysfunction, and/or other significant organ impairments; ④ history of mental illness; ⑤ significant missing clinical data. This study was approved by the Ethics Committee of the Aerospace Center Hospital (2023-004).

#### *Data collection*

The hospital's electronic medical record system was retrieved to collect patient clinical data, including: ① gender, age, body mass index (BMI), smoking history; ② history of allergic diseases; ③ pet ownership status; ④ history of chronic pharyngitis, hypertension, coronary atherosclerotic heart disease, diabetes, etc.; ⑤ disease course, cough symptoms and scoring, throat pain/dryness/itchiness, chest tightness/wheezing/dyspnea, chest pain, self-reported wheezing, nasal congestion, rhinorrhea/sneezing, itchy eyes, acid reflux/heartburn, and other clinical manifestations; ⑥ complete blood cell analysis (white blood cells, lymphocytes, neutrophils, monocytes, eosinophils, hemoglobin, platelets), routine C-reactive protein measurement; ⑦ exhaled nitric oxide (FeNO) examination; ⑧ pulmonary function test (tidal volume (VT), inspiratory capacity (IC), vital capacity (VC), forced vital capacity (FVC),

forced expiratory volume in one second (FEV1), FEV1/FVC, peak expiratory flow (PEF), maximal expiratory flow at 75% of FVC (MEF75), MEF50, MEF25, maximal mid-expiratory flow between 75% and 25% of FVC (MMEF75/25), total lung capacity (TLC), residual volume (RV), ratio of residual volume to total lung capacity (RV/TLC), diffusing capacity of the lung for carbon monoxide (single-breath method) (DLCO/SB), diffusing capacity of the lung for carbon monoxide per unit of alveolar volume (DLCO/VA) and bronchodilation test (FVC variability, FEV1 variability). Atypical asthma was diagnosed according to the "Guidelines for the Prevention and Treatment of Bronchial Asthma (2020 Edition)" [10].

#### *Data processing and model construction*

The clinical data collected were organized based on the inclusion and exclusion criteria. Patients with incomplete information, typical bronchial asthma, acute respiratory infections, and a history of mental illness were excluded. A simple random sampling method was used to divide the organized dataset into a training set (70%) and a validation set (30%). Univariate and multivariate logistic regression (LR) analyses were conducted on the training set to identify independent prognostic factors ( $P < 0.05$ ), which were used to build an LR clinical diagnostic prediction model. Additionally, four machine learning methods, decision tree (DT), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGBoost), were employed to develop clinical diagnostic models for atypical asthma. Ten-fold cross-validation was used to determine the optimal parameters for the five models, which were then validated on the validation set. Receiver operating characteristic (ROC) curves and decision curve analysis (DCA) curves were plotted to assess the discrimination and clinical utility of each model. Finally, the SHAP algorithm was used to explain and analyze the best-performing model, enhancing its interpretability and transparency.

#### *Statistical analysis*

Statistical analysis was conducted using SPSS 24.0 software. Categorical data were presented as frequencies (percentages) and compared between groups using the chi-square test or Fisher's exact test. Normally distributed continuous data were expressed as mean  $\pm$  stan-

dard deviation and compared using the independent samples t-test. Non-normally distributed continuous data were presented as median (Q1, Q3) and compared using the Mann-Whitney U test. Univariate logistic regression analysis was employed to screen relevant risk factors, and variables with a  $p$ -value  $<0.05$  were considered significant factors for the disease diagnosis. Subsequently, significant variables identified by the univariate analysis were further assessed using multivariate logistic regression to identify independent prognostic factors, which were later incorporated to construct a clinical predictive model. Model construction was performed using R 3.4.1 software, and evaluation metrics such as the area under the ROC curve (AUC), sensitivity, and specificity were calculated. Decision curve analysis (DCA) was conducted to assess clinical utility.

### Results

#### *Baseline data*

After screening, 214 cases were included in the study, comprising 98 cases diagnosed with acute exacerbation of non-typical asthma and 116 cases with undiagnosed asthma. Analysis of the baseline data (**Table 1**) revealed significant associations between non-typical asthma with various variables, including age, cough score, wheezing, nasal congestion, rhinorrhea/sneezing, pet ownership, FEV1, MEF75 (%), MEF50 (%), MEF25 (%), MMEF75/25 (%), FEV1 variability, FeNO, and eosinophils, monocyte percentage (MO), acid reflux, allergic rhinitis, DLCO/VA (%), and WBC ( $P<0.05$ ,  $P<0.01$ , or  $P<0.001$ ).

After randomly assigning patients in a 7:3 ratio, 149 individuals were allocated to the training set and 65 to the validation set. In the training set, 69 cases (46%) were diagnosed with atypical asthma, while in the validation set, 29 cases (45%) were diagnosed with atypical asthma. The balance between the training and validation sets was comparable ( $P=0.937$ ).

#### *Logistic regression analysis and validation*

By univariate logistic regression analysis, the training data (**Table 2**) were examined to identify relevant risk factors. The analysis revealed a significant correlation between disease diag-

nosis and variables such as age, cough symptom score, wheezing, nasal congestion, rhinorrhea/sneezing, history of allergic rhinitis, pet ownership, FEV1, FEV1 rate, MEF75, MEF50, MEF25, MMEF75/25, FEV1 variability, FeNO, EOS, and other factors.

Following multivariable logistic regression analysis, four variables were found to be significantly associated with acute exacerbations of atypical asthma: history of allergic rhinitis (OR=14.69, 95% CI=2.12-197.01,  $P<0.05$ ), variability in FEV1 (OR=1.51, 95% CI=1.17-2.1,  $P<0.01$ ), FeNO levels (OR=1.25, 95% CI=1.16-1.40,  $P<0.001$ ), and EOS (OR=0, 95% CI=0-0.19,  $P<0.05$ ) (**Figure 1**). These four variables were identified as independent predictive factors for acute exacerbations of atypical asthma.

The predictive model demonstrated excellent discriminative ability, as evidenced by an AUC of 0.984 (95% CI: 0.970-0.998) in the training set (**Figure 2A**) and 0.983 (95% CI: 0.960-1) in the validation set (**Figure 2B**). Results from the Hosmer-Lemeshow test indicated that the calibration curves of the training set ( $R^2=0.868$ ,  $P=0.922>0.05$ ) (**Figure 2C**) and validation set ( $R^2=0.847$ ,  $P=0.323>0.05$ ) (**Figure 2D**) were well-aligned with the ideal curve, showing no significant differences and indicating a strong calibration capability of the model.

#### *Evaluation of machine learning model performance*

The predictive capabilities of the five models were evaluated on a validation set based on a disease prediction model were constructed using logistic regression. The discriminative ability, as shown by the ROC curve (**Figure 3A**), indicated that when the model A, based on logistic regression principles (AUC=0.984), was used as a reference, the decision tree model (model B) (AUC=0.957), random forest (model C) (AUC=0.970), support vector machine model (model D) (AUC=0.968), and extreme gradient boosting model (model E) (AUC=0.982) all demonstrated good predictive discrimination.

In the detailed performance metrics for each model (**Table 3**), the XGBoost model had the highest accuracy (95.4%), outperforming both the LR model (92.3%) and RF model (93.8%) in correctly predicting samples. Sensitivity was

## Acute atypical asthma attacks

**Table 1.** Comparison of baseline clinical data between the two groups [ $M (Q_1, Q_3), \bar{x} \pm s, n (\%)$ ]

Variable	Total (n=214)	0 (n=116)	1 (n=98)	p	Statistic
Gender				0.687	0.163
Female	118 (55)	62 (53)	56 (57)		
Male	96 (45)	54 (47)	42 (43)		
Age, year	43 (34.3, 53.8)	47 (37, 58)	40 (31, 47)	<0.001	7381.5
BMI, kg/m <sup>2</sup>	24 (22, 27)	24 (21.75, 27)	24 (22, 26)	0.442	6029.5
Course of illness				0.545	1.215
Acute	85 (40)	49 (42)	36 (37)		
Subacute	50 (23)	28 (24)	22 (22)		
Chronic	79 (37)	39 (34)	40 (41)		
Cough				<0.001	38.324
0	33 (15)	18 (16)	15 (15)		
1	101 (47)	75 (65)	26 (27)		
2	59 (28)	19 (16)	40 (41)		
3	21 (10)	4 (3)	17 (17)		
Sore throat				0.916	0.011
No	146 (68)	80 (69)	66 (67)		
Yes	68 (32)	36 (31)	32 (33)		
Dyspnea				0.435	0.609
No	88 (41)	51 (44)	37 (38)		
Yes	126 (59)	65 (56)	61 (62)		
Chest pain				0.757	Fisher
No	204 (95)	110 (95)	94 (96)		
Yes	10 (5)	6 (5)	4 (4)		
Wheeze				<0.001	38.74
No	165 (77)	109 (94)	56 (57)		
Yes	49 (23)	7 (6)	42 (43)		
Stuffy nose				<0.001	13.025
No	188 (88)	111 (96)	77 (79)		
Yes	26 (12)	5 (4)	21 (21)		
Sneezing & runny nose				<0.001	13.409
No	130 (61)	84 (72)	46 (47)		
Yes	84 (39)	32 (28)	52 (53)		
Itchy eyes				0.083	Fisher
No	205 (96)	114 (98)	91 (93)		
Yes	9 (4)	2 (2)	7 (7)		
Acid reflux & heartburn				0.016	Fisher
No	207 (97)	109 (94)	98 (100)		
Yes	7 (3)	7 (6)	0 (0)		
Allergic rhinitis				0.012	6.261
No	70 (33)	47 (41)	23 (23)		
Yes	144 (67)	69 (59)	75 (77)		
Chronic pharyngitis				0.546	0.365
No	200 (93)	110 (95)	90 (92)		
Yes	14 (7)	6 (5)	8 (8)		
Hypertension & CHD				0.522	0.409
No	174 (81)	92 (79)	82 (84)		
Yes	40 (19)	24 (21)	16 (16)		
Diabetes				0.294	Fisher
No	206 (96)	110 (95)	96 (98)		
Yes	8 (4)	6 (5)	2 (2)		

## Acute atypical asthma attacks

Pet ownership				<0.001	33.996
No	173 (81)	111 (96)	62 (63)		
Yes	41 (19)	5 (4)	36 (37)		
Smoking history				0.763	0.091
No	184 (86)	101 (87)	83 (85)		
Yes	30 (14)	15 (13)	15 (15)		
VT	214.3 (165.7, 267.5)	211.9 (157.2, 258.0)	217.7 (174.2, 273.7)	0.233	5145
IC	93.2 (81.5, 103.7)	91.8 (79.8, 103.7)	94.3 (85.7, 103.5)	0.290	5205.5
VC	89.0 (84.8, 96.8)	88.8 (84.7, 95.6)	89.3 (85.1, 99.5)	0.225	5136
FVC	90.8 (86.2, 98.2)	90.9 (85.5, 97.2)	90.8 (87.2, 101.0)	0.332	5245.5
FEV1	90.3 (80.9, 98.1)	91.0 (83.2, 97.7)	87.1 (77.5, 98.1)	0.099	6428
FEV1/FVC	81.3 (76.7, 85.4)	82.7 (79.7, 86.3)	79.6 (75.3, 83.2)	<0.001	7431
PEF	88.5 (82.0, 96.7)	88.85 (81.6, 100.53)	87.95 (82.43, 95.57)	0.414	6053
MEF75	90.0 (77.5, 99.9)	93.2 (80.45, 102.5)	86 (75.75, 94.52)	<0.001	7192
MEF50	74.53±22.54	79.25±22.87	68.93±20.9	<0.001	3.444
MEF25	59.1 (47.0, 80.6)	64.3 (49.65, 85.4)	54.35 (43.5, 63.95)	<0.001	7247.5
MMEF75/25	69.1±21.2	74.1±21.78	63.1±19.04	<0.001	3.942
TLC	91.51±10.53	90.79±10.49	92.37±10.57	0.275	-1.093
RV	115.1 (101.7, 127.5)	115.05 (101.35, 129.83)	115.05 (103.67, 127.07)	0.919	5730.5
RV/TLC	122.1 (111.3, 133.3)	123.2 (110.67, 134.58)	118.9 (111.72, 128.75)	0.314	6139
DLCO/SB	82.5 (74.0, 92.9)	83.9 (74.97, 94.03)	82.2 (73.45, 92.33)	0.259	6194
DLCO/VA	95.0 (83.5, 104.8)	95.65 (88.88, 105.58)	91.15 (79.45, 104)	0.028	6677.5
FVC rate of variability	2.0 (-1.3, 4.8)	2.3 (-1.45, 5.28)	1.75 (-0.65, 4.55)	0.664	5880.5
FEV1 rate of variability	3.7 (2.1, 5.7)	3.15 (1.6, 4.9)	4.95 (2.73, 7.07)	<0.001	3842.5
FeNO	25 (15, 57)	16 (11, 19)	59.5 (41.25, 88.25)	<0.001	528.5
CRP	2 (0.97, 3.94)	1.94 (0.98, 3.88)	2.08 (0.95, 3.94)	0.781	5558
WBC	7 (5.7, 7.83)	6.84 (5.23, 7.77)	7.04 (6.53, 7.91)	0.021	4645
LY	29.4 (21.9, 33.3)	29.15 (22.92, 32.65)	29.45 (20.9, 33.75)	0.789	5805
MO	6.3 (5.4, 7.5)	6.6 (5.6, 7.8)	5.8 (5.1, 7.2)	0.002	7103
GR	59.96±9.97	61.13±9.49	58.56±10.38	0.062	1.874
EOS	4.6 (2.2, 7.38)	2.55 (1.5, 4.75)	7.1 (4.82, 9.97)	<0.001	2042
HgB	142 (133, 150.75)	141.5 (133, 147.25)	144 (133.25, 153.75)	0.060	4834
PLT	247 (215.25, 277.5)	244.5 (208.75, 275.25)	250.5 (223, 284)	0.149	5031.5

Notes: total: full dataset; 0 group: undiagnosed asthma group; 1 group: atypical asthma group; BMI: body mass index; VT: tidal volume; IC: inspiratory capacity; VC: vital capacity; FVC: forced vital capacity; FEV1: forced expiratory volume in one second; FEV1/FVC: forced expiratory volume in one second/forced vital capacity; PEF: peak expiratory flow; FEF25: forced expiratory flow at 25% of FVC; FEF50: forced expiratory flow at 50% of FVC; FEF75: forced expiratory flow at 75% of FVC; MMEF75/25: maximal mid-expiratory flow; TLC: total lung capacity; RV: residual volume; RV/TLC: residual volume/total lung capacity; DLCO/SB: diffusing capacity of the lung for carbon monoxide per square meter of body surface area; DLCO/VA: diffusing capacity of the lung for carbon monoxide/alveolar ventilation; FeNO: fractional exhaled nitric oxide; CRP: C-reactive protein; WBC: white blood cell count; LY: lymphocytes; MO: monocytes; GR: neutrophils; EOS: eosinophils; HgB: hemoglobin; PLT: platelets.

**Table 2.** Single-factor logistic regression analysis

Characteristic	B	SE	OR	CI	Z	P
PLT	0.003	0.00325	1.003	0.997-1.010	0.941	0.347
HgB	0.017	0.01200	1.017	0.994-1.042	1.420	0.156
EOS	0.419	0.07491	1.520	1.327-1.782	5.587	0.000
GR	-0.022	0.01632	0.978	0.946-1.009	-1.378	0.168
MO	-0.264	0.10112	0.768	0.623-0.927	-2.612	0.009
LY	-0.016	0.01968	0.984	0.946-1.023	-0.808	0.419
WBC	0.171	0.10673	1.187	0.966-1.472	1.606	0.108
CRP	-0.040	0.03504	0.960	0.874-1.006	-1.154	0.249
FeNO	0.156	0.02631	1.168	1.116-1.239	5.915	0.000
FEV1 rate of variability	0.182	0.05661	1.199	1.079-1.349	3.213	0.001

## Acute atypical asthma attacks

FVC rate of variability	0.006	0.03093	1.006	0.946-1.069	0.185	0.853
DLCO/VA	-0.015	0.01209	0.985	0.961-1.008	-1.272	0.203
DLCO/SB	-0.005	0.01318	0.995	0.970-1.021	-0.359	0.720
RV/TLC	0.003	0.00776	1.003	0.988-1.019	0.430	0.667
RV	0.006	0.00704	1.006	0.992-1.020	0.782	0.434
TLC	0.016	0.01558	1.016	0.986-1.048	1.012	0.312
MMEF75/25	-0.022	0.00827	0.979	0.962-0.994	-2.616	0.009
MEF25	-0.015	0.00723	0.985	0.971-0.999	-2.033	0.042
MEF50	-0.019	0.00781	0.981	0.966-0.996	-2.420	0.015
MEF75	-0.021	0.00822	0.979	0.963-0.994	-2.578	0.010
PEF	-0.012	0.01042	0.988	0.968-1.008	-1.135	0.256
FEV1/FVC	-0.065	0.02451	0.937	0.891-0.981	-2.662	0.008
FEV1	-0.019	0.01117	0.981	0.959-1.003	-1.694	0.090
FVC	-0.001	0.01231	0.999	0.975-1.024	-0.048	0.962
VC	0.000	0.01283	1.000	0.975-1.026	0.017	0.987
IC	-0.009	0.00821	0.991	0.974-1.007	-1.137	0.255
VT	0.001	0.00197	1.001	0.997-1.005	0.388	0.698
Smoking	0.388	0.46385	1.474	0.594-3.730	0.836	0.403
Pet ownership	2.379	0.57086	10.795	3.886-38.46	4.168	0.000
Diabetes	-0.266	0.92797	0.766	0.099-4.755	-0.287	0.774
Hypertension & CHD	-0.511	0.45377	0.600	0.238-1.435	-1.126	0.260
Chronic pharyngitis	0.395	0.69194	1.484	0.378-6.216	0.571	0.568
Allergic rhinitis	1.158	0.38201	3.184	1.535-6.919	3.032	0.002
Acid reflux & heartburn	-15.456	840.27417	0.000	NA-7.627	-0.018	0.985
Itchy eyes	1.312	0.83388	3.714	0.823-25.95	1.574	0.116
Sneezing & runny nose	1.120	0.35021	3.065	1.557-6.170	3.198	0.001
Stuffy nose	1.427	0.54637	4.167	1.513-13.44	2.612	0.009
Wheeze	2.864	0.63745	17.528	5.772-76.42	4.493	0.000
Chest pain	0.154	0.83355	1.167	0.210-6.488	0.185	0.853
Dyspnea	0.442	0.34060	1.556	0.801-3.057	1.297	0.195
Throat	0.046	0.34611	1.047	0.529-2.065	0.132	0.895
Cough	0.762	0.20783	2.142	1.446-3.281	3.664	0.000
Time	-0.083	0.18405	0.921	0.640-1.320	-0.449	0.653
BMI	-0.083	0.05570	0.920	0.823-1.025	-1.498	0.134
Age	-0.033	0.01232	0.967	0.943-0.990	-2.698	0.007
Gender	-0.053	0.32993	0.948	0.495-1.811	-0.162	0.872

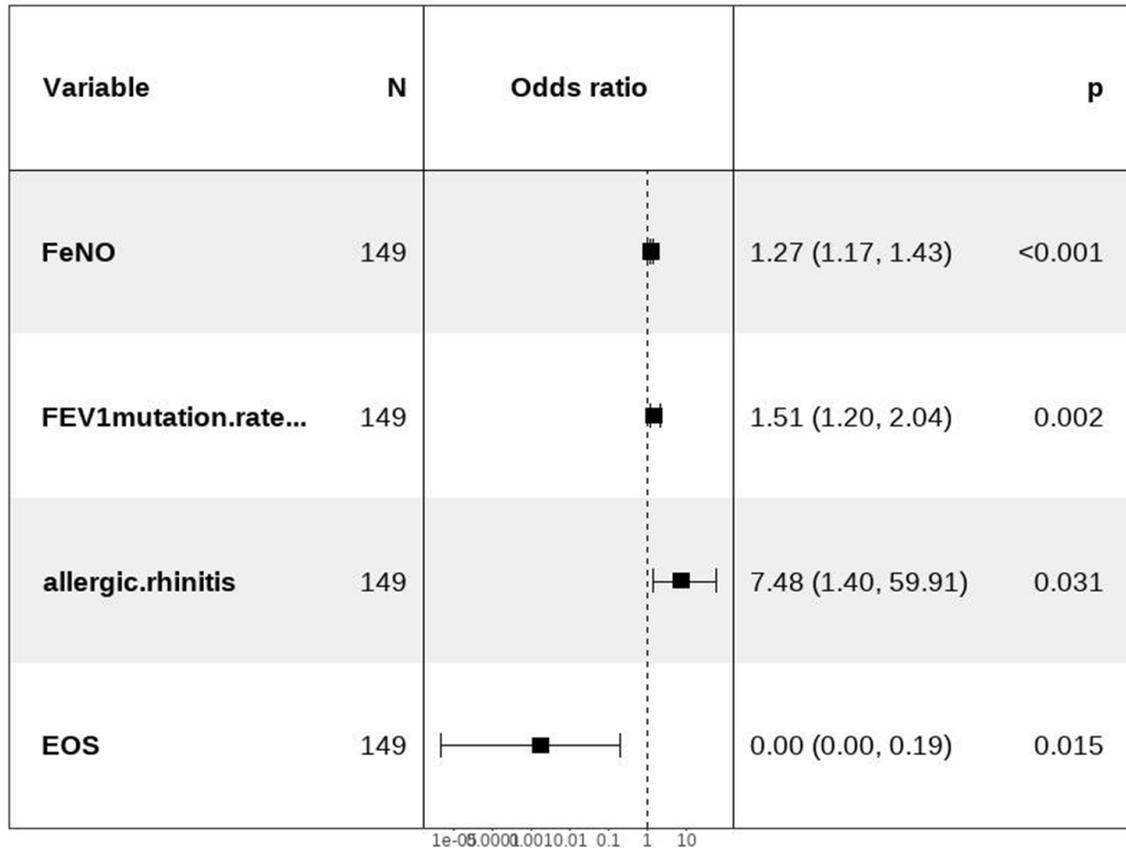
Notes: CRP: C-reactive protein; WBC: white blood cell count; LY: lymphocytes; MO: monocytes; GR: neutrophils; EOS: eosinophils; Hgb: hemoglobin; PLT: platelets; IC: Inspiratory Capacity; VC: Vital Capacity; FVC: Forced Vital Capacity; FEV1: Forced Expiratory Volume in One Second; FEV1/FVC: Forced Expiratory Volume in One Second/Forced Vital Capacity; PEF: Peak Expiratory Flow; FEF25: Forced Expiratory Flow at 25% of FVC; FEF50: Forced Expiratory Flow at 50% of FVC; FEF75: Forced Expiratory Flow at 75% of FVC; MMEF75/25: Maximal Mid - Expiratory Flow; TLC: Total Lung Capacity; RV: Residual Volume; RV/TLC: Residual Volume/Total Lung Capacity; DLCO/SB: Diffusing Capacity of the Lung for Carbon Monoxide per Square Meter of Body Surface Area; DLCO/VA: Diffusing Capacity of the Lung for Carbon Monoxide/Alveolar Ventilation; Va: Alveolar Ventilation; FeNO: Fractional Exhaled Nitric Oxide; VT: tidal volume; BMI: body mass index.

similar between the LR and RF (93.1%) models, indicating comparable identification abilities for positive cases among these three models.

In the decision curve analysis (DCA), “None” and “All” represented the extent of benefit

when patients received no clinical decision intervention and when all patients received it, respectively. **Figure 3B** showed that all five clinical prediction models significantly benefit clinical decision-making, with the logistic regression model (ModA) yielding higher net ben-

## Acute atypical asthma attacks



**Figure 1.** Multifactorial logistic regression forest diagram. FeNO: Fractional Exhaled Nitric Oxide; EOS: eosinophils; HgB: hemoglobin; FEV1: Forced Expiratory Volume in One Second.

efits across most threshold ranges compared to other models, suggesting that ModelA leads to best net clinical benefits (**Figure 3**). Considering the above indicators collectively, the logistic regression model (LR model) exhibited the best predictive performance.

### Model explanation and analysis

The logistic regression model uses traditional univariate logistic regression to screen variables, followed by multivariate logistic regression optimization to classify the target variable and construct a predictive model. The final predictive factors in the model were history of allergic rhinitis, FEV1 variability, FeNO, and EOS.

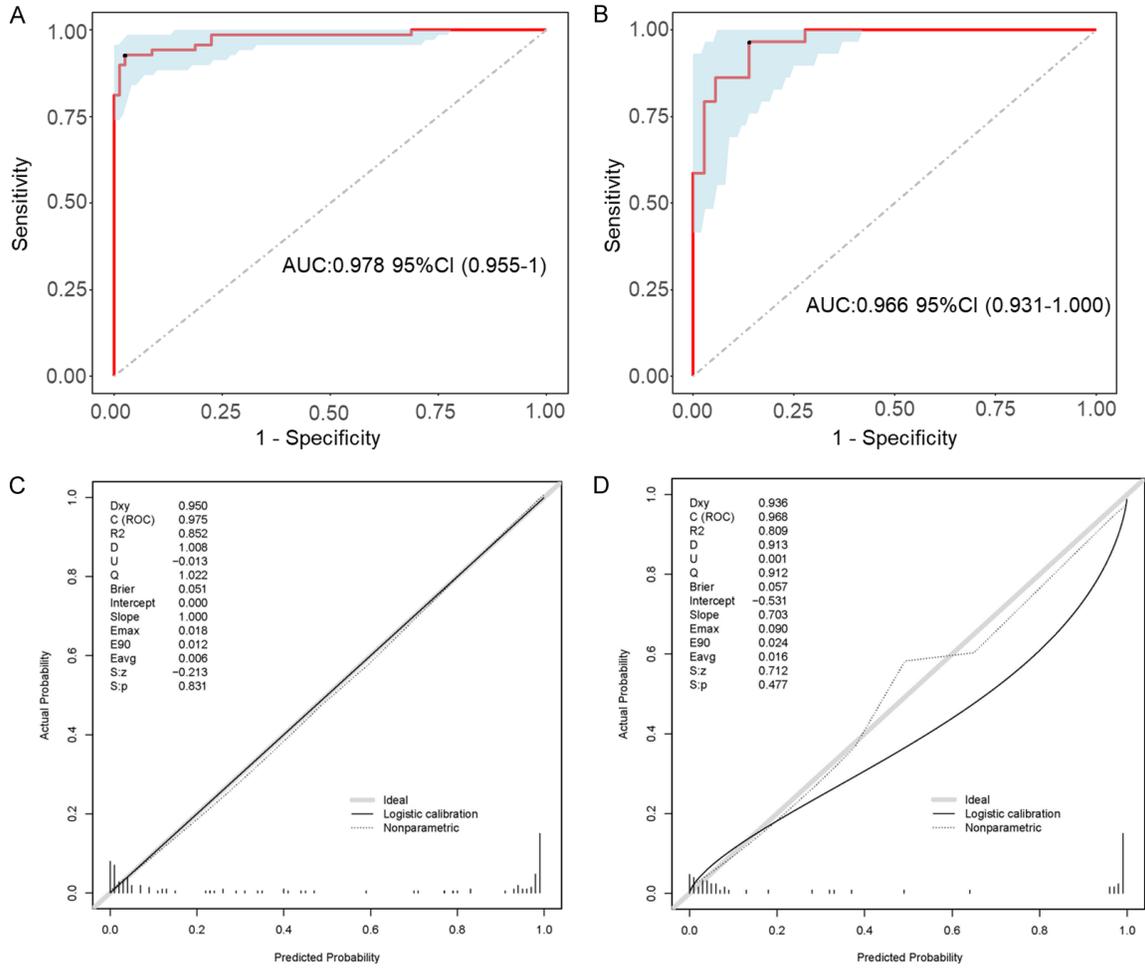
In the decision tree model (**Figure 4A**), the variable importance ranking was as follows: FeNO, EOS, wheezing during acute attacks, pet ownership, cough, and rhinorrhea/sneezing. In the random forest model (**Figure 4B**), the variable

importance ranking was as follows: FeNO, EOS, wheezing, FEV1 variability, pet ownership, and cough. In the support vector machine model (**Figure 4C**), the variable importance ranking was: FeNO, FEV1 variability, FEV1, cough, wheezing, and pet ownership. In the extreme gradient boosting model (**Figure 4D**), the variable importance ranking was: FeNO, FEV1 variability, FEV1, neutrophil percentage, monocyte percentage, wheezing, and eosinophils.

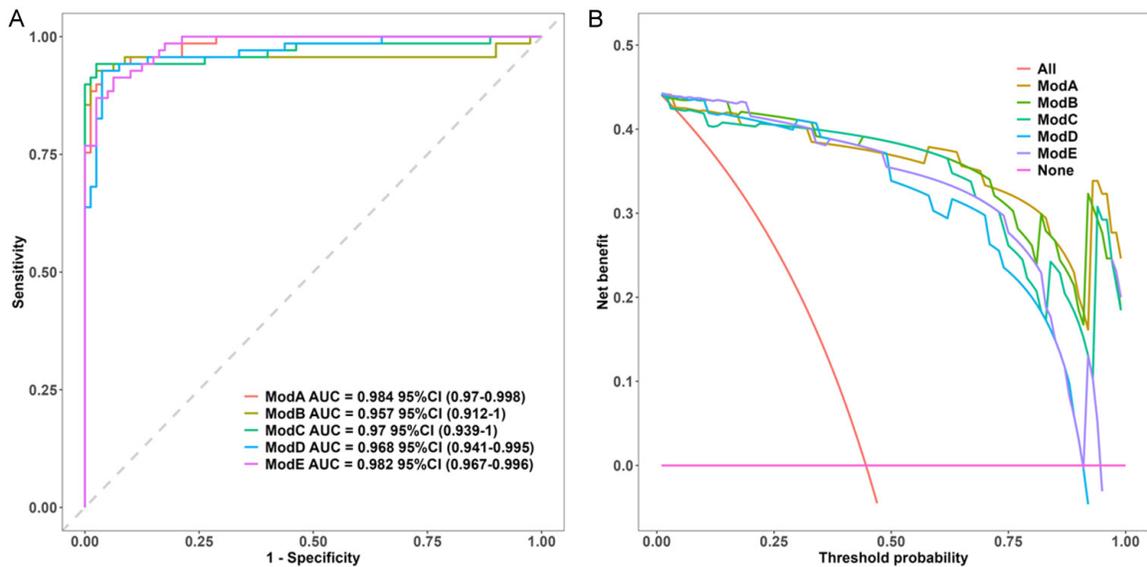
### Discussion

The Pediatric Branch of the Chinese Medical Association has invited experts from both domestic and international fields to develop a diagnostic model for predicting childhood asthma using evidence-based medicine. By integrating clinical experience, five key parameters have been identified: 1) the frequency of wheezing episodes  $\geq 4$  times; 2) presence of reversible airflow limitation; 3) personal history of allergies; 4) first-degree family history of aller-

## Acute atypical asthma attacks



**Figure 2.** The discrimination and calibration curves of the logistic regression model. A. Discriminative curve of logistic regression model in the training set; B. Discriminative curve of logistic regression model in the validation set; C. Calibration curve of logistic regression model in the training set; D. Calibration curve of logistic regression model in the validation set.



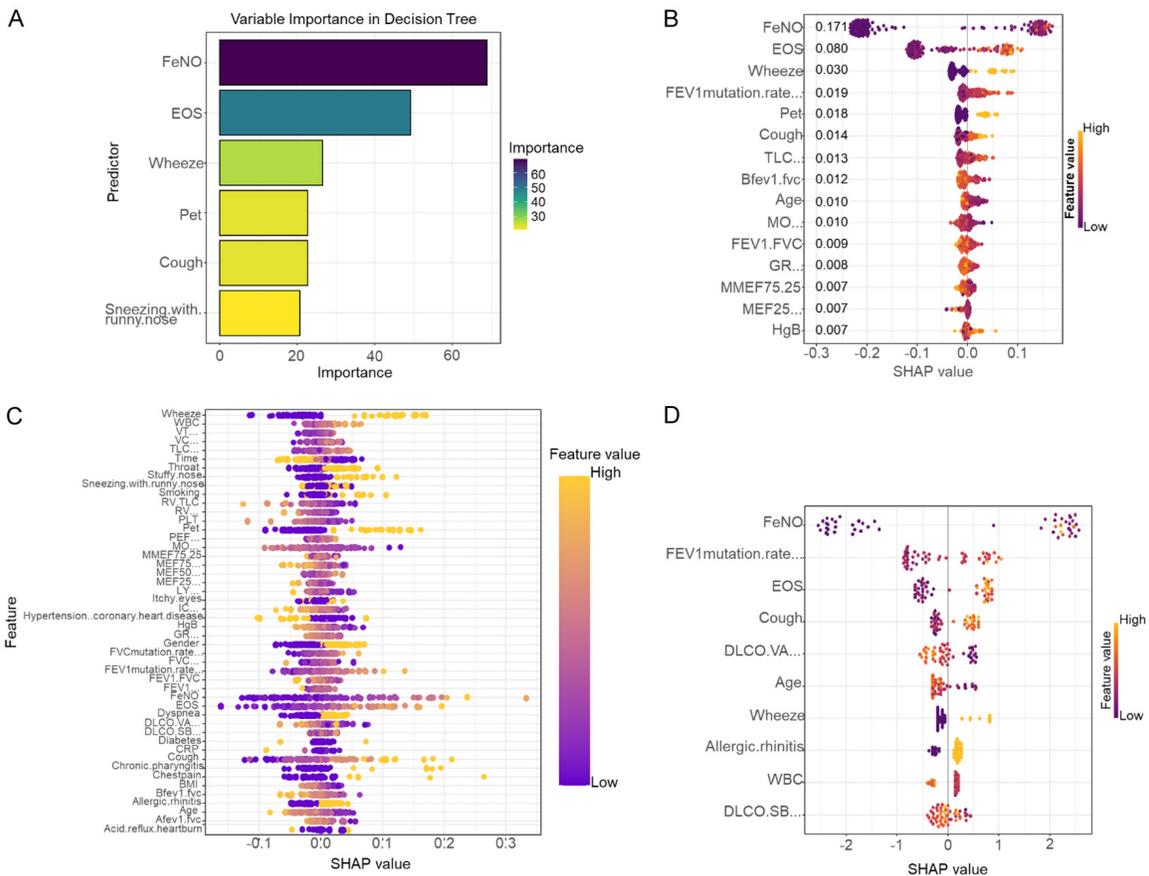
# Acute atypical asthma attacks

**Figure 3.** Performance evaluation of five machine learning models. A. Discriminative curves of 5 prediction models; B. Decision analysis curves for 5 predictive models.

**Table 3.** Comparison of performances among five predictive models

Prediction model	AUC	Accuracy rate	Sensitivity	Specificity	F1 score
LR model	0.984	92.3%	93.1%	91.7%	0.924
DT model	0.957	92.3%	89.7%	94.4%	0.920
RF model	0.970	93.8%	93.1%	94.4%	0.938
SVM model	0.968	81.5%	65.5%	94.4%	0.773
XGBoost model	0.982	95.4%	93.1%	97.2%	0.951

Notes: LR: logistic regression; DT: decision tree; FR: forest plot; SVM: support vector machine; XGboost: extreme gradient boosting.



**Figure 4.** Important variables in each model. A. Variable importance in the Decision Tree model; B. Variable importance in the Random Forest model; C. Variable importance in the Support Vector Machine model; D. Variable importance in the Extreme Gradient Boosting model.

gies; and 5) positive allergen test results [11]. However, the dynamic characteristics of children's physiologic functions and immune status limit the applicability of adult asthma diagnostic models. An ideal diagnostic prediction model should comprehensively reflect the clinical features of asthma, relevant examination

indicators, and biomarkers, while also being operationally feasible.

A prospective study [12] has shown that a four-variable model, including wheezing, FEV1, FEV1/FVC ratio, and FeNO, can effectively diagnose bronchial asthma with an AUC of 0.76.

Another prospective study [13] found that a model incorporating cough, wheezing, dyspnea, hay fever, eczema, food allergies, social class, maternal asthma, childhood passive smoking exposure, and lung function/reversibility testing achieved an internally validated AUC of 0.86. Furthermore, a retrospective study [14] identified age, gender, FEV1, eosinophil count, and FeNO as risk factors for predicting severe asthma exacerbations. This study utilized real data from public hospitals and employed structured assessments and expert diagnosis to analyze the characteristics, symptoms, general examinations, pulmonary function tests, and FeNO results of outpatient asthma patients. Five machine learning methods, including logistic regression, decision tree, random forest, support vector machine, and extreme gradient boosting, were applied to identify key variables for predicting acute exacerbations of atypical asthma. The results demonstrated that all models effectively identified acute exacerbations of atypical asthma, with the extreme gradient boosting model achieving an accuracy of 95.4%. Key variables such as FeNO, EOS, forced expiratory volume variability, history of allergic rhinitis, wheezing during exacerbations, pet ownership, and cough symptom score, were identified, consistent with previous research findings and confirming the reliability of the model. These easily accessible and interpretable feature variables can assist primary care physicians in optimizing diagnostic processes under time constraints, enhancing diagnostic efficiency for atypical asthma, and reducing misdiagnosis.

This study still has several limitations. 1) The retrospective single-center study design may have introduced selection bias. 2) The lack of external validation necessitates further prospective research to confirm the clinical applicability and utility of the model. 3) Limited clinical data and sample size may have affected the predictive performance and stability of the model.

### Conclusion

Accurately identifying atypical asthma in outpatient settings remains a significant challenge in the diagnosis and treatment of respiratory diseases. This study used five machine learning methods (logistic regression, decision tree, ran-

dom forest, support vector machine, extreme gradient boosting) to develop a clinical diagnostic prediction model. The results indicated that each model could reliably predict the acute episodes of atypical asthma with high accuracy. The selected feature variables, which are easily obtainable in clinical practice, demonstrated strong applicability and provide a basis for the timely diagnosis of acute episodes in patients with atypical asthma.

### Disclosure of conflict of interest

None.

**Address correspondence to:** Feng Ma, Department of Respiratory and Critical Care Medicine, Aerospace Center Hospital, No. 15 Yuquan Road, Haidian District, Beijing 100049, China. Tel: +86-010-59915237; E-mail: mafeng93015229@126.com

### References

- [1] Lai K. Chinese national guidelines on diagnosis and management of cough: consensus and controversy. *J Thorac Dis* 2014; 6 Suppl 7: S683-688.
- [2] GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017; 5: 691-706.
- [3] Huang K, Yang T, Xu J, Yang L, Zhao J, Zhang X, Bai C, Kang J, Ran P, Shen H, Wen F, Chen Y, Sun T, Shan G, Lin Y, Xu G, Wu S, Wang C, Wang R, Shi Z, Xu Y, Ye X, Song Y, Wang Q, Zhou Y, Li W, Ding L, Wan C, Yao W, Guo Y, Xiao F, Lu Y, Peng X, Zhang B, Xiao D, Wang Z, Chen Z, Bu X, Zhang H, Zhang X, An L, Zhang S, Zhu J, Cao Z, Zhan Q, Yang Y, Liang L, Tong X, Dai H, Cao B, Wu T, Chung KF, He J and Wang C; China Pulmonary Health (CPH) Study Group. Prevalence, risk factors, and management of asthma in China: a national cross-sectional study. *Lancet* 2019; 394: 407-418.
- [4] Asthma Workgroup; Chinese Thoracic Society; Chinese Society of General Practitioners. Chinese guideline for the prevention and management of bronchial asthma (Primary Health Care Version). *J Thorac Dis* 2013; 5: 667-677.
- [5] Lin J, Yang D, Huang M, Zhang Y, Chen P, Cai S, Liu C, Wu C, Yin K, Wang C, Zhou X and Su N. Chinese expert consensus on diagnosis and management of severe asthma. *J Thorac Dis* 2018; 10: 7020-7044.

## Acute atypical asthma attacks

- [6] Rajvanshi N, Kumar P and Goyal JP. Global initiative for asthma guidelines 2024: an update. *Indian Pediatr* 2024; 61: 781-786.
- [7] Heaney LG, Perez de Llano L, Al-Ahmad M, Backer V, Busby J, Canonica GW, Christoff GC, Cosio BG, FitzGerald JM, Heffler E, Iwanaga T, Jackson DJ, Menzies-Gow AN, Papadopoulos NG, Papaioannou AI, Pfeffer PE, Popov TA, Porsbjerg CM, Rhee CK, Sadatsafavi M, Tohda Y, Wang E, Wechsler ME, Alacqua M, Altraja A, Bjermer L, Björnsdóttir US, Bourdin A, Brusselle GG, Buhl R, Costello RW, Hew M, Koh MS, Lehmann S, Lehtimäki L, Peters M, Taillé C, Taube C, Tran TN, Zangrilli J, Bulathsinhala L, Carter VA, Chaudhry I, Eleangovan N, Hosseini N, Kerkhof M, Murray RB, Price CA and Price DB. Eosinophilic and noneosinophilic asthma: an expert consensus framework to characterize phenotypes in a global real-life severe asthma cohort. *Chest* 2021; 160: 814-830.
- [8] Hussain M and Liu G. Eosinophilic asthma: pathophysiology and therapeutic horizons. *Cells* 2024; 13: 384.
- [9] Yancey SW, Keene ON, Albers FC, Ortega H, Bates S, Bleecker ER and Pavord I. Biomarkers for severe eosinophilic asthma. *J Allergy Clin Immunol* 2017; 140: 1509-1518.
- [10] Asthma group of Chinese Throacic Society. Guidelines for bronchial asthma prevent and management(2020 edition) Asthma group of Chinese Throacic Society. *Zhonghua Jie He He Hu Xi Za Zhi* 2020; 43: 1023-1048.
- [11] Martin J, Townshend J and Brodli M. Diagnosis and management of asthma in children. *BMJ Paediatr Open* 2022; 6: e001277.
- [12] Louis G, Schleich F, Guillaume M, Kirkove D, Nekoe Zahrei H, Donneau AF, Henket M, Paulus V, Guissard F, Louis R and Petre B. Development and validation of a predictive model combining patient-reported outcome measures, spirometry and exhaled nitric oxide fraction for asthma diagnosis. *ERJ Open Res* 2023; 9: 00451-2022.
- [13] Daines L, Bonnett LJ, Tibble H, Boyd A, Thomas R, Price D, Turner SW, Lewis SC, Sheikh A and Pinnock H. Deriving and validating an asthma diagnosis prediction model for children and young people in primary care. *Wellcome Open Res* 2023; 8: 195.
- [14] Pavord ID, Holliday M, Reddel HK, Braithwaite I, Ebmeier S, Hancox RJ, Harrison T, Houghton C, Oldfield K, Papi A, Williams M, Weatherall M and Beasley R; Novel START Study Team. Predictive value of blood eosinophils and exhaled nitric oxide in adults with mild asthma: a pre-specified subgroup analysis of an open-label, parallel-group, randomised controlled trial. *Lancet Respir Med* 2020; 8: 671-680.