Journal of Asthma and Allergy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Forecasting Hospitalization for Adult Asthma Patients in Emergency Departments Based on Multiple Environmental and Clinical Factors

Hanxu Xi[1,2,*], Yudi Zhang[3,*], Rui Zuo[2], Wei Li[2], Chen Zhang[2], Yongchang Sun[1], Hong Ji[2], Zhiqiang He[3], Chun Chang[1]

[1]Department of Respiratory and Critical Care Medicine, Peking University Third Hospital, Beijing, 100191, People's Republic of China; [2]Information Management and Big Data Centre, Peking University Third Hospital, Beijing, 100191, People's Republic of China; [3]Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, 100876, People's Republic of China

*These authors contributed equally to this work

Correspondence: Chun Chang, Department of Respiratory and Critical Care Medicine, Peking University Third Hospital, No. 49 North Garden Road, Haidian District, Beijing, 100191, People's Republic of China, Email doudou_1977@163.com; Hong Ji, Information Management and Big Data Center, Peking University Third Hospital, No. 49 North Garden Road, Haidian District, Beijing, 100191, People's Republic of China, Email puh3_imc@bjmu.edu.cn

Background: Asthma is the world's second most prevalent chronic respiratory disease. Current clinical decisions regarding hospitalization for adult asthma patients in emergency departments (EDs) primarily rely on presenting clinical status, acute exacerbation severity, therapeutic response and high-risk factors. Assessing the need for hospitalization of patients with complex comorbidities remains a significant challenge.

Research Question: This study aims to develop models that integrate various environmental and clinical factors to predict the hospitalization of adult asthma patients in EDs and to interpret these models.

Study Design and Methods: A retrospective analysis was conducted utilizing data from asthma patients at a single ED from 2016 to 2023; the data included demographics, vital signs, illness severity, laboratory test results, and comorbidities, along with environmental variables. Predictive models were constructed using the extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), support vector machine (SVM), logistic regression (LR), and random forest (RF). Area under the receiver operating characteristic curve (AUC), accuracy, and F1 score were the primary metrics used to assess model performance.

Results: The analysis included 1140 ED visits. The median age was 51.0 years (interquartile range: 31.0 to 67.0 years), and 56.5% of the patients (644) were female. Overall, 21.8% of patients (249) required hospitalization after their ED visits. The AUC results for predicting hospitalization without external environmental factors were 0.8075 for XGBoost, 0.8233 for LightGBM, 0.7935 for SVM, 0.8033 for LR, and 0.8272 for RF. After integrating ambient air pollutant and meteorological features, the RF model consistently outperformed the other models, achieving an AUC of 0.8555. The most critical parameters for predicting hospitalization were found to be illness severity, oxygen saturation, age, and heart rate.

Interpretation: Machine learning (ML) models based on clinical, meteorological, and air pollution data can rapidly and accurately predict hospitalization of adult asthma patients in EDs.

Keywords: asthma exacerbation, machine learning, emergency department

Asthma is a severe global health issue and ranks as the world's second most common chronic respiratory disease,[1] affecting 1–18% of the population in various nations.[2] In China, the overall prevalence of asthma in adults over the age of 20 is estimated at 4.2%, representing 45.7 million individuals. Among these, 15.5% have visited the emergency department (ED) at least once in the past year due to worsening respiratory symptoms.[3] Following the COVID-19 pandemic, the demand for emergency medical services has surged, intensifying the challenges faced by emergency medical systems. Asthma-related visits have exacerbated the shortage of emergency medical resources. Although most

Received: 27 February 2025
Accepted: 26 May 2025
Published: 31 May 2025

asthma exacerbations are managed in outpatient settings or EDs, severe cases may require hospitalization, significantly impacting healthcare expenses.[4–7] Therefore, prompt identification of patients needing hospitalization is necessary to reduce ED boarding times,[8] optimize resource allocation, and ensure early and intensive care for those at a high risk of admission.

With the growing availability of large data sources and advancements in computational power,[9–11] predicting the risk of hospitalization for asthma has entered a new era. Zein et al[12] employed a machine learning (ML) approach, finding that the light gradient boosting algorithm was optimal for predicting hospitalization for asthma in outpatients. Patel et al[13] concluded that the gradient-boosting algorithm surpassed other algorithms in predicting pediatric asthma patient hospitalizations in the ED. Similarly, Goto et al[14] reported that the random forest (RF) algorithm demonstrated the highest discriminative ability and sensitivity in predicting hospitalization for adult asthma patients in the ED.

However, although meteorological factors and airborne pollutants are recognized as asthma risk factors,[15–18] no study has yet combined clinical, meteorological, and air pollution data to predict hospitalization for adult asthma patients in EDs. This study aims to fill this void. We believe that this approach will enable clinicians to determine quickly whether asthma patients require hospitalization, and thus will reduce ED boarding times, accelerate patient recovery, and promote rational use of emergency medical resources.

Materials and Methods

Study Design and Setting

A retrospective analysis was conducted utilizing data from the electronic health records (EHRs) of a single healthcare system. The study protocol received approval from Peking University Third Hospital Medical Science Research Ethics Committee (Approval No: IRB00006761-M2021582).

Study Samples

All ED visits by adult patients (aged ≥ 18 years) diagnosed with asthma by a physician from November 1, 2016, to July 1, 2023, were included. Pregnant women and patients without EHRs during their ED visits were excluded.

Predictors

Input variables included demographic information, vital signs, laboratory test results, comorbidities, and initial illness severity at triage. If multiple laboratory tests were conducted during ED visits, only the results from the first test were used. The severity of the initial illness was assessed using the Chinese Emergency Triage Scale (CETS), which categorizes the urgency of a patient's condition on a four-point scale, with one indicating the highest urgency.[19] Covariates were selected based on biological plausibility and evidence from prior research.

Additional input variables included external environmental factors. Concentrations of ambient air pollutants (O_3 and $PM_{2.5}$) were sourced from Tracking Air Pollution in China (TAP), a near real-time air pollutant concentration database.[20] Concentrations of other pollutants (CO, NO_2, SO_2, and PM_{10}) were obtained from ChinaHighAirPollutants (CHAP), which provides long-term, comprehensive, high-resolution, and high-quality datasets of ground-level air pollutants in China.[21,22] Meteorological variables, including temperature and relative humidity, were derived from the fifth generation of European ReAnalysis (ERA5), provided by the European Centre for Medium-Range Weather Forecasts (ECMWF).[23] Geocoding of each subject's residential street into latitude and longitude was performed using the programming interface of the mapping application Amap (AutoNavi Software). Specifically, by matching geographic coordinates and the dates of the visit to the ED, we obtained the initial 32 environmental variables, including levels of $PM_{2.5}$, O_3, CO, NO_2, SO_2, and PM_{10} at 24, 48, 168, and 336 hours prior to the patients' visits to the ED (hereafter designated as 24 h prior, 48 h prior, 168 h prior, and 336 h prior, respectively).

Data Pre-Processing

The data pre-processing procedure encompassed outlier detection, handling of missing values, standardization, and re-sampling to address class imbalances.

A multivariate model approach known as "Cook's distance"[24] was employed to identify outliers, which were removed from the dataset as necessary. Missing values were imputed using the median statistic for each respective column. To mitigate the impact of varying scales among different features on model training, the values of each continuous variable were standardized to a range of 0 to 1. After these pre-processing steps, the dataset comprised 1140 observations and 54 variables, including 22 base in-hospital variables and 32 external environmental factors. The initial dataset revealed that 21.8% of the samples belonged to the hospitalization class, indicating a class imbalance. To enhance the model's performance across both classes, a technique called SMOTEENN[25] combining undersampling and oversampling was applied to balance the dataset.

Feature Engineering

The initial input variables captured conditions at specific instants but lacked information on dynamic changes. To address this and further explore the impact of environmental conditions on asthma severity, the study incorporated feature engineering of environmental factors. Specifically, for each of the environmental factors, we calculated the difference in values at 24, 48, 168, and 336 h prior. These differential variables enabled the examination of the dynamic influences of environmental factors on asthma. Consequently, the final dataset was expanded to include 102 variables, encompassing 48 new differential variables.

Statistical Analysis

The dataset was randomly partitioned into two subsets: (1) a training dataset comprising 80% of the patients, utilized for model training, and (2) a test dataset consisting of 20% of the patients, used to evaluate the final performance of the model. The proportion of positive cases was maintained equally in both the training and test datasets. A five-fold cross-validation procedure was applied to each model within the training dataset. The model performance was assessed using the test dataset.

The performance of the models was evaluated by calculating the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Predictive models were developed using various techniques, specifically the extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), support vector machine (SVM), logistic regression (LR), and RF techniques. To enhance the interpretability of the results, the Shapley additive explanations (SHAP)[26] approach was employed. This method represents the average contribution of each feature to the overall predictions made by the model through its mean absolute SHAP value. Additionally, the cumulative effect study method was utilized to further improve model interpretability. Figure 1 provides a comprehensive overview of the procedures established for each predictive model, covering data pre-processing, feature engineering, model training, and evaluation.

The research was conducted using Python v3.12.4 with the following key dependency versions: scikit-learn v1.5.2 for ML model development and pipeline construction, XGBoost v2.1.4 for gradient boosting implementation, pandas v2.2.3 for data processing and manipulation, NumPy v2.0.2 for fundamental numerical operations, and SHAP v0.46.0 for explainable analysis.

Outcome Measures

The outcome measure in this study was the hospitalization event, defined as admission to a respiratory ward, an internal medicine ward, or an intensive care unit. A binary outcome (Yes/No) was determined based on whether a patient experienced a hospitalization event during the study period.

Results

Patient Characteristics

Between November 1, 2016, and July 1, 2023, a total of 1507 patients with asthma met the inclusion criteria for the study. From this group, exclusions were made for 293 patients who lacked electronic medical records during ED visits,

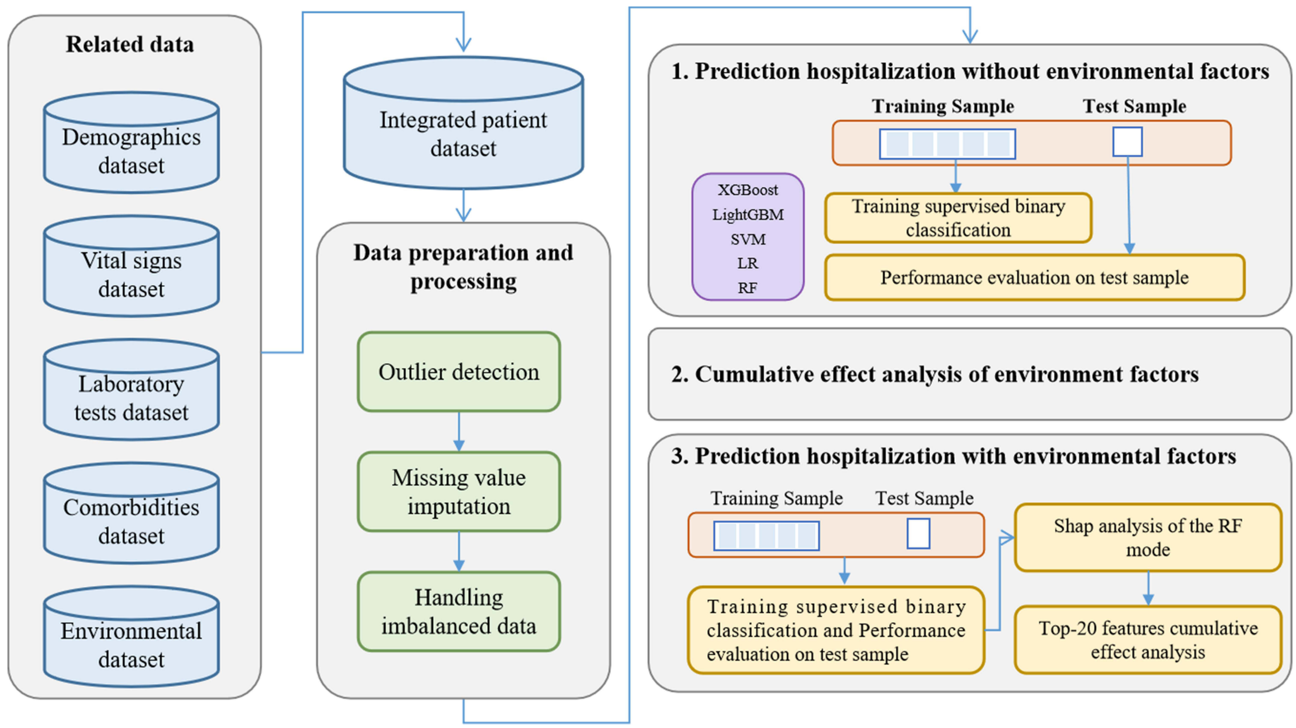

Figure 1 Overview of Model Design.

44 patients who were return visitors to the ED, and 30 pregnant patients (e-Figure 1). Consequently, 1140 ED visits were included in the analysis.

The median age of the included patients was 51.0 years, with an interquartile range of 31.0 to 67.0 years; 56.5% (644 patients) were female. After receiving ED treatment, 21.8% of the patients (249) required hospitalization. The epidemiological data of the patients included in the study are detailed in Table 1.

Table 1 Epidemiological Data of Patients

Characteristics		All Patients (N = 1140)	Hospitalization (249)	No Hospitalization (891)	P value
Age, y		51.0(67.0–31.0)	62.0(76.0–48.0)	46.0(63.0–29.0)	<0.05
Female sex		644(56.50)	153(61.40)	491(55.10)	0.074
Vital signs					
	Pulse, bpm	96.0(108.0–83.0)	104.0(117.0–91.0)	93.0(106.0–82.0)	<0.05
	Systolic blood pressure, mmHg	138.0(154.0–122.0)	147.0(164.8–131.0)	135.0(150.0–120.0)	<0.05
	Diastolic blood pressure, mmHg	81.0(90.0–71.0)	84.5(96.0–74.3)	80.0(89.0–70.0)	<0.05
Oxygen saturation, %		96.0(98.0–94.0)	94.0(97.0–90.0)	97.0(98.0–95.0)	<0.05
Triage level (CETS)					
	CETS 1	3(0.30)	3(1.20)	0(0.00)	<0.05
	CETS 2	95(8.70)	77(32.00)	18(2.10)	
	CETS 3	784(71.50)	125(51.90)	659(77.10)	
	CETS 4	214(19.50)	36(14.90)	178(20.80)	

(*Continued*)

Table 1 (Continued).

Characteristics		All Patients (N = 1140)	Hospitalization (249)	No Hospitalization (891)	P value
Laboratory tests					
	White blood cell, ×10^9/L	8.5(10.9–6.9)	8.8(11.9–7.2)	8.4(10.5–6.8)	<0.05
	Red blood cell, ×10^{12}/L	4.8(5.2–4.4)	4.7(5.1–4.3)	4.9(5.2–4.5)	<0.05
	Hemoglobin, g/L	144.0(156.0–133.0)	143.0(156.0–130.0)	145.0(156.0–134.8)	<0.05
	Platelet, ×10^9/L	241.0(288.0–200.5)	241.0(296.0–201.0)	241.0(285.0–200.0)	0.608
	Lymphocyte, ×10^9/L	1.8(2.4–1.2)	1.5(2.2–1.1)	1.8(2.4–1.3)	<0.05
	Neutrophil, ×10^9/L	5.6(7.8–4.2)	6.2(9.4–4.5)	5.5(7.4–4.1)	<0.05
	Eosinophil, ×109/L	0.3(0.6–0.1)	0.2(0.5–0.0)	0.3(0.6–0.1)	<0.05
	Monocyte, ×10^9/L	0.5(0.7–0.4)	0.5(0.7–0.4)	0.5(0.7–0.4)	0.11
Comorbidities					
	Hypertension	371(32.50)	133(53.40)	238(26.70)	<0.05
	Allergic rhinitis	397(34.80)	121(48.60)	276(31.00)	<0.05
	Eczema	203(17.80)	61(24.50)	142(15.90)	<0.05
	COPD	170(14.90)	80(32.10)	90(10.10)	<0.05
	GERD	164(14.40)	60(24.10)	104(11.70)	<0.05
	Sinusitis	132(11.60)	46(18.50)	86(9.70)	<0.05
	Bronchiectasis	106(9.30)	26(10.40)	80(9.00)	0.482
Environmental factors: Air pollution data					
24 h prior	$PM_{2.5}$, μg/m^3	36.0(59.0–19.7)	35.0(62.0–19.0)	36.0(58.0–20.0)	0.601
	O_3, μg/m^3	85.0(133.4–55.4)	85.5(133.4–55.1)	85.0(133.7–55.7)	0.918
	CO, mg/m^3	0.8(1.1–0.5)	0.8(1.1–0.5)	0.8(1.1–0.5)	0.807
	NO_2, μg/m^3	40.3(56.3–28.0)	37.7(55.2–26.3)	41.5(56.7–28.4)	0.201
	SO_2, μg/m^3	5.7(10.6–3.1)	5.4(9.2–3.1)	5.7(10.9–3.1)	0.371
	PM_{10}, μg/m^3	71.8(107.2–46.2)	70.2(109.0–44.6)	71.8(106.6–46.6)	0.814
48 h prior	$PM_{2.5}$, μg/m^3	35.0(59.0–20.0)	33.7(60.0–19.0)	36.0(59.0–20.3)	0.363
	O_3, μg/m^3	83.4(134.5–56.2)	83.3(141.4–55.6)	83.5(133.2–56.3)	0.885
	CO, mg/m^3	0.8(1.1–0.6)	0.7(1.1–0.6)	0.8(1.1–0.5)	0.586
	NO_2, μg/m^3	41.5(56.9–28.1)	41.0(56.3–28.7)	41.9(57.3–27.9)	0.668
	SO_2, μg/m^3	5.6(11.2–3.1)	5.2(10.6–3.2)	5.6(11.3–3.1)	0.782
	PM_{10}, μg/m^3	71.4(106.8–44.2)	69.2(104.6–43.0)	71.8(107.0–44.7)	0.495
168 h prior	$PM_{2.5}$, μg/m^3	35.0(56.3–18.5)	32.0(54.0–16.2)	36.0(56.8–20.0)	0.054
	O_3, μg/m^3	84.0(127.0–55.8)	86.1(126.6–56.7)	83.6(127.1–55.4)	0.662
	CO, mg/m^3	0.8(1.1–0.6)	0.7(1.0–0.5)	0.8(1.1–0.6)	<0.05
	NO_2, μg/m^3	39.8(54.9–28.5)	34.7(49.4–24.7)	41.6(56.6–29.8)	<0.05
	SO_2, μg/m^3	5.7(11.0–3.0)	4.7(8.5–3.0)	6.1(11.5–3.1)	<0.05
	PM_{10}, μg/m^3	65.8(99.8–44.7)	58.8(90.0–39.2)	68.1(102.1–47.3)	<0.05
336 h prior	$PM_{2.5}$, μg/m^3	34.0(59.0–18.0)	37.0(57.3–19.0)	33.2(59.0–18.0)	0.663
	O_3, μg/m^3	81.0(124.5–55.0)	80.8(116.1–52.3)	81.3(127.1–55.9)	0.183
	CO, mg/m^3	0.8(1.1–0.5)	0.8(1.1–0.5)	0.8(1.1–0.5)	0.642
	NO_2, μg/m^3	38.2(56.1–27.5)	37.2(58.9–26.3)	38.3(54.5–28.0)	0.964
	SO_2, μg/m^3	5.2(10.4–3.2)	4.9(10.3–3.1)	5.4(10.5–3.2)	0.379
	PM_{10}, μg/m^3	68.2(100.6–40.9)	69.6(96.4–39.6)	67.8(102.8–41.5)	0.469

(*Continued*)

Table 1 (Continued).

Characteristics		All Patients (N = 1140)	Hospitalization (249)	No Hospitalization (891)	P value
Environmental factors: Meteorological data					
24 h prior	Temperature, K	285.5(296.0–273.6)	286.2(296.2–274.1)	285.2(295.9–273.6)	0.66
	Relative humidity, %	46.5(65.0–30.3)	46.9(64.8–30.5)	46.4(65.1–30.1)	0.636
48 h prior	Temperature, K	285.3(295.7–273.4)	285.6(295.7–273.5)	285.1(295.6–273.4)	0.614
	Relative humidity, %	46.7(65.1–30.5)	44.5(63.1–30.6)	47.4(65.4–30.5)	0.792
168 h prior	Temperature, K	285.2(295.8–273.7)	285.3(296.1–273.7)	285.2(295.8–273.7)	0.665
	Relative humidity, %	46.8(65.6–29.8)	46.8(63.3–29.6)	46.8(65.8–29.8)	0.745
336 h prior	Temperature, K	285.7(295.7–273.7)	286.2(296.1–274.7)	285.4(295.6–273.6)	0.384
	Relative humidity, %	47.4(67.1–30.3)	47.4(67.8–30.2)	47.5(66.8–30.6)	0.915

Note: Data are presented as number (%) or median (interquartile range).
Abbreviations: CETS, Chinese Emergency Triage Scale; COPD, chronic obstructive pulmonary disease; GERD, gastroesophageal reflux disease; $PM_{2.5}$, particulate matter with an aerodynamic diameter ≤2.5 μm; O_3, ozone; CO, carbon monoxide; NO_2, nitrogen dioxide; SO_2, sulfur dioxide; PM_{10}, particulate matter with an aerodynamic diameter ≤10 μm; h prior, hours prior to the emergency department visit.

Hospitalization Without External Environmental Factors

The models were initially trained and evaluated using 22 in-hospital variables. Table 2 illustrates the prediction performance of five models on the training dataset, assessed via five-fold cross-validation. XGBoost achieved the highest mean sensitivity, although RF exhibited a slightly superior maximum sensitivity of 0.558 (compared with 0.5512 for XGBoost). Both SVM and LR demonstrated specificities greater than 0.96. The PPV for LR and RF exceeded 71%, while XGBoost and LightGBM presented NPVs better than 0.85. All models yielded relatively low F1-Score results, with SVM registering the lowest at 0.3. The RF algorithm achieved high accuracy and AUC (both exceeding 0.82); however, it encountered challenges in terms of high variance and slightly reduced stability.

Table 3 compares the prediction performance of the five models based on the test dataset. RF performed at high levels on all metrics except Specificity and PPV, achieving an optimal AUC of 0.8272 compared with the other models. Figure 2 presents the ROC curve for each model.

Table 2 Comparison of Receiving Operator Characteristic Curves of Five-Fold Cross-Validation Based on Training Data

	Sensitivity	Specificity	PPV	NPV	F1_score	Accuracy	AUC
XGBoost	0.4999 (0.4486–0.5512)	0.9207 (0.8990–0.9424)	0.6626 (0.5997–0.7255)	0.8580 (0.8452–0.8708)	0.5671 (0.5268–0.6074)	0.8222 (0.8062–0.8382)	0.8249 (0.8142–0.8356)
LightGBM	0.4760 (0.4380–0.5140)	0.9314 (0.9195–0.9433)	0.6805 (0.6418–0.7192)	0.8533 (0.8406–0.8660)	0.5585 (0.5344–0.5826)	0.8246 (0.8172–0.8320)	0.8202 (0.8061–0.8343)
SVM	0.2149 (0.1559–0.2739)	0.9695 (0.9569–0.9821)	0.6856 (0.5925–0.7787)	0.8020 (0.7887–0.8153)	0.3230 (0.2469–0.3991)	0.7930 (0.7789–0.8071)	0.7706 (0.7480–0.7932)
LR	0.3113 (0.2656–0.3570)	0.9602 (0.9410–0.9794)	0.7163 (0.6407–0.7919)	0.8202 (0.8045–0.8359)	0.4306 (0.3842–0.4770)	0.8082 (0.7885–0.8279)	0.7906 (0.7552–0.8260)
RF	0.4484 (0.3388–0.5580)	0.9451 (0.9292–0.9610)	0.7107 (0.6261–0.7953)	0.8487 (0.8182–0.8792)	0.5447 (0.4532–0.6362)	0.8281 (0.7981–0.8581)	0.8213 (0.7885–0.8541)

Abbreviations: XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machines; LR, logistic regression; RF, random forest; NPV, negative predictive value; PPV, positive predictive value; AUC, area under the receiver operating characteristic curve.

Table 3 Comparison of Receiving Operator Characteristic Curves Based on Test Data

	Sensitivity	Specificity	PPV	NPV	F1_score	Accuracy	AUC
XGBoost	0.4677	0.9283	0.6444	0.8625	0.5421	0.8281	0.8075
LightGBM	0.2742	0.9821	0.8095	0.8295	0.4096	0.8281	0.8233
SVM	0.1935	0.9776	0.7059	0.8134	0.3038	0.8070	0.7935
LR	0.3871	0.9596	0.7273	0.8492	0.5053	0.8351	0.8033
RF	0.5323	0.8879	0.5690	0.8722	0.5500	0.8105	0.8272

Abbreviations: XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machines; LR, logistic regression; RF, random forest; NPV, negative predictive value; PPV, positive predictive value; AUC, area under the receiver operating characteristic curve.

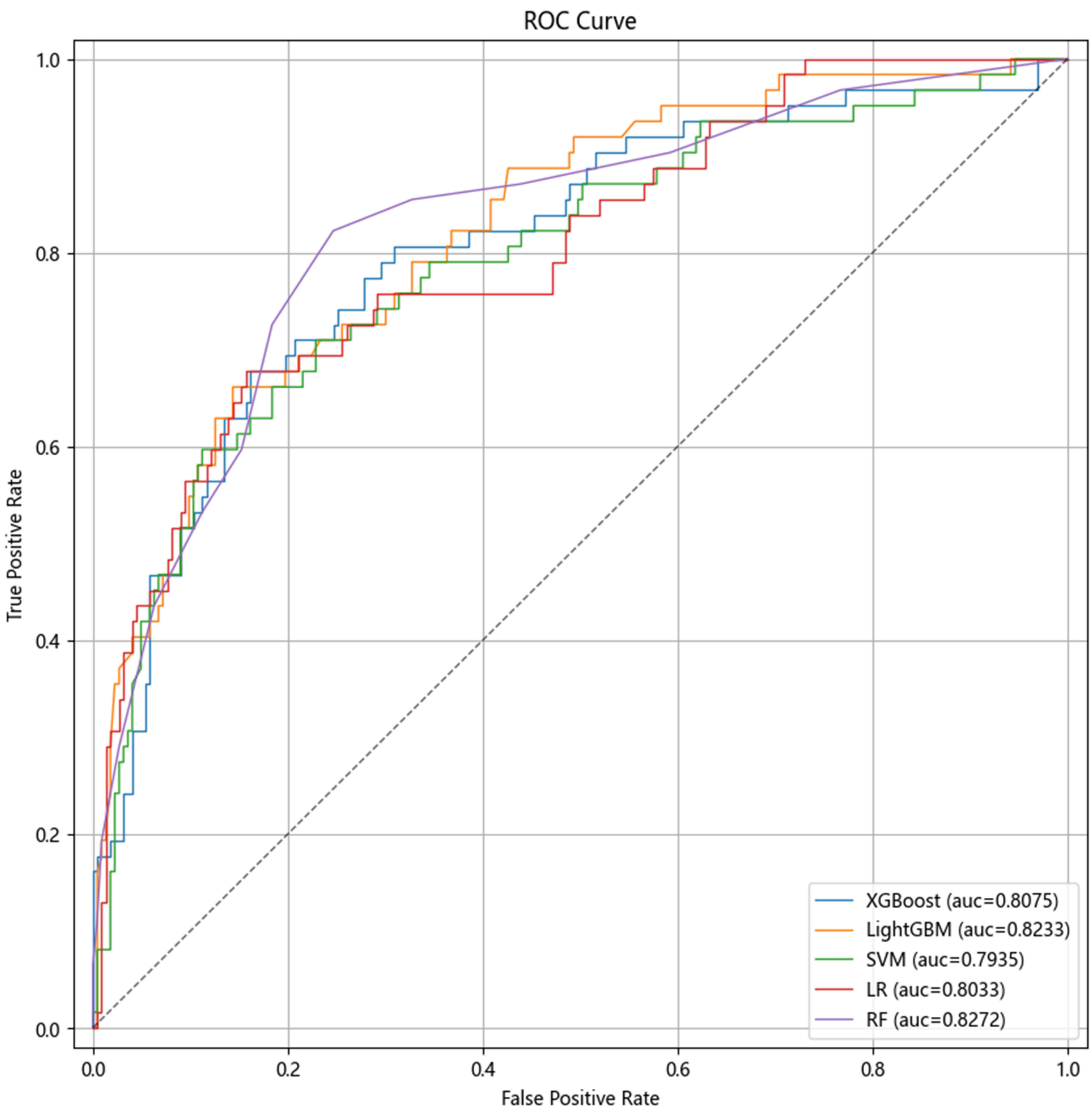

Figure 2 Receiver Operating Characteristic (ROC) Curves for Models Excluding External Environmental Factors. This figure compares the performance metrics Negative Predictive Value (NPV), Positive Predictive Value (PPV), and Area Under the Curve (AUC) for various models: Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LightGBM), Support Vector Machines (SVM), Logistic Regression (LR), and Random Forest (RF).

Cumulative Effect of Environment Variables

We sequentially integrated ambient air pollutant and meteorological features to enhance the RF model's predictive capability. Initially, the 22 in-hospital variables formed the baseline. We then progressively added environmental variables at different time points, incorporating eight variables in each phase to achieve total counts of 30, 38, 46, and 54 variables. To minimize the impact of random variation and enhance robustness, the model was trained five times for each configuration.

Figure 3 illustrates the trends in cumulative effects as environmental characteristics were integrated at various time intervals. The blue dashed line represents the mean of five AUC scores, while the light blue shaded area indicates the

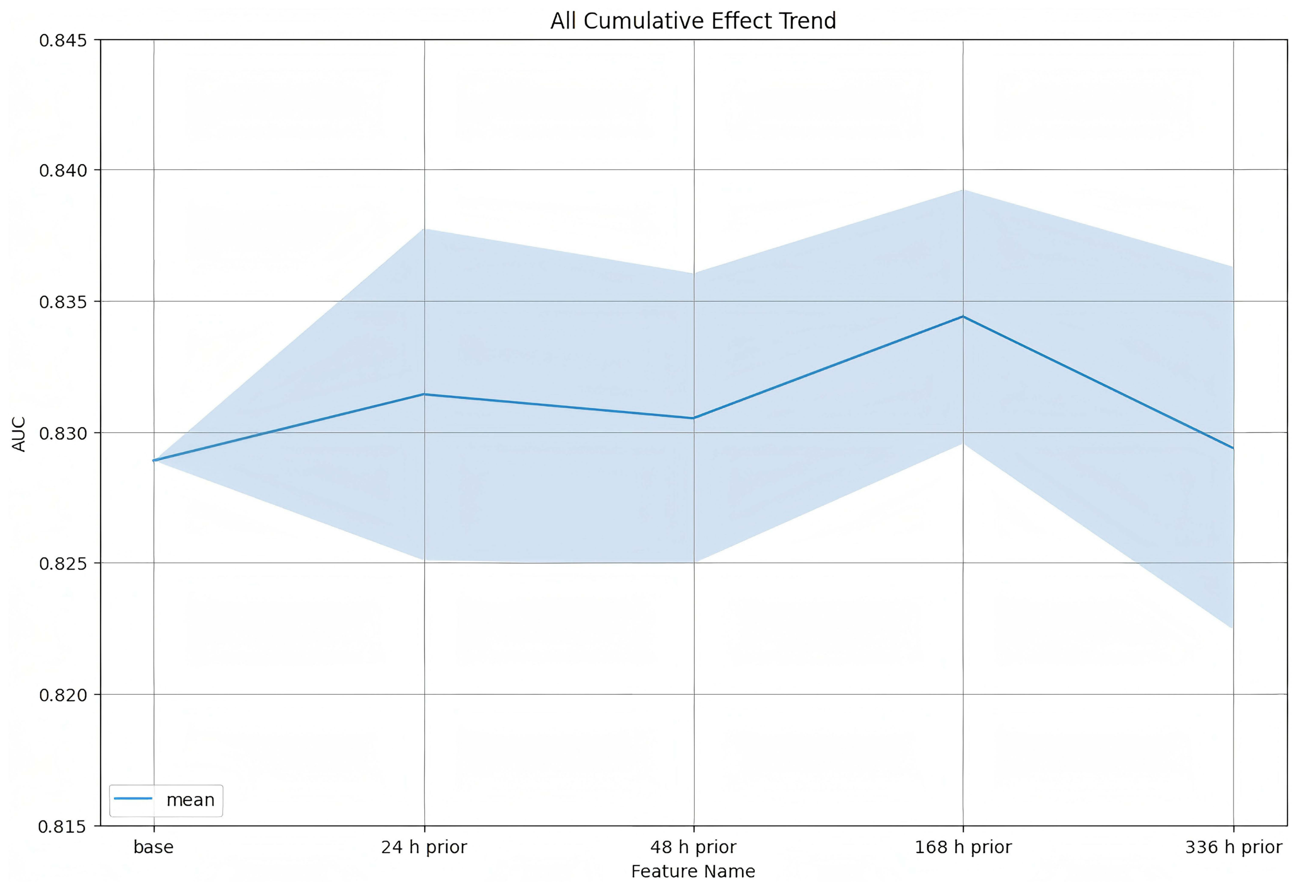

Figure 3 Trends in the Cumulative Effects of Environmental Features at Different Time Points. This figure depicts the influence of environmental factors introduced at different times on the model's performance. Baseline (Base) represents predictions without environmental data. Different times are specified as 24 h, 48 h, 168 h, and 336 h prior to the ED visit.

range of the means ± standard deviations. Notably, the cumulative effect reached its peak at 168 h prior, indicating the optimal integration of temporal environmental data for predicting hospitalization. Based on this finding, the subsequent phase of model training included a total of 46 features.

Hospitalization with External Environmental Factors

Based on the findings outlined in the previous section, the input dataset for the analysis was composed of 46 variables, which included 22 hospitalization and 24 environmental variables. Following the feature engineering processes detailed in the Data Pre-processing section, an additional 48 features were incorporated to capture the dynamics of the environmental conditions. As a result, the total number of variables used in the analysis increased to 94.

Table 4 presents a comparison of the prediction performance of five models based on the training dataset, assessed through five-fold cross-validation. Both SVM and LR exhibited the lowest values for Sensitivity, NPV, F1-Score,

Table 4 Comparison of Prediction Performance Across Five Models Using Training Data and External Environmental Factors

	Sensitivity	Specificity	PPV	NPV	F1_score	Accuracy	AUC
XGBoost	0.4856 (0.4182–0.5530)	0.9238 (0.9074–0.9402)	0.6624 (0.6252–0.6996)	0.8550 (0.8378–0.8722)	0.5568 (0.5148–0.5988)	0.8211 (0.8116–0.8306)	0.8464 (0.8382–0.8546)
LightGBM	0.4772 (0.4019–0.5525)	0.9512 (0.9264–0.9760)	0.7603 (0.6611–0.8595)	0.8562 (0.8345–0.8779)	0.5803 (0.5171–0.6435)	0.8398 (0.8157–0.8639)	0.8483 (0.8204–0.8762)
SVM	0.2100 (0.1506–0.2694)	0.9802 (0.9699–0.9905)	0.7669 (0.6667–0.8671)	0.8027 (0.7884–0.8170)	0.3257 (0.2459–0.4055)	0.8000 (0.7837–0.8163)	0.7718 (0.7423–0.8013)
LR	0.3359 (0.2995–0.3723)	0.9587 (0.9401–0.9773)	0.7205 (0.6259–0.8151)	0.8252 (0.8101–0.8403)	0.4570 (0.4083–0.5057)	0.8129 (0.7901–0.8357)	0.7979 (0.7587–0.8371)
RF	0.4982 (0.3994–0.5970)	0.9603 (0.9472–0.9734)	0.7934 (0.7329–0.8539)	0.8622 (0.8342–0.8902)	0.6072 (0.5283–0.6861)	0.8515 (0.8244–0.8786)	0.8432 (0.8167–0.8697)

Abbreviations: XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machines; LR, logistic regression; RF, random forest; NPV, negative predictive value; PPV, positive predictive value; AUC, area under the receiver operating characteristic curve.

 https://doi.org/10.2147/JAA.S512405

Table 5 Comparison of Prediction Performance Across Five Models Using Test Data and External Environmental Factors

	Sensitivity	Specificity	PPV	NPV	F1_score	Accuracy	AUC
XGBoost	0.5323	0.9327	0.6875	0.8776	0.6000	0.8456	0.8201
LightGBM	0.5161	0.9417	0.7111	0.8750	0.5981	0.8491	0.8398
SVM	0.1774	0.9821	0.7333	0.8111	0.2857	0.8070	0.8046
LR	0.3548	0.9552	0.6875	0.8419	0.4681	0.8246	0.8136
RF	0.5000	0.9462	0.7209	0.8719	0.5905	0.8491	0.8555

Abbreviations: XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machines; LR, logistic regression; RF, random forest; NPV, negative predictive value; PPV, positive predictive value; AUC, area under the receiver operating characteristic curve.

Accuracy, and AUC. Table 5 details the prediction performance comparison of the five models based on the test dataset. As shown in Table 5, the RF model consistently outperformed the others across all evaluated metrics, achieving an accuracy of 0.8491 and an AUC of 0.8555. Consequently, the RF model is established as the optimal model for this study. Figure 4 displays the ROC curve for each model, further illustrating the comparative performance of the predictive models. The hyperparameters and configurations of all models are summarized in Table 6.

SHAP Analysis of RF Model

Figure 5 displays the SHAP value distribution for the top 20 features, ranked by importance and shown on the vertical axis to the left, with the SHAP values plotted on the horizontal axis. The color bar on the right indicates the actual values of the features. The analysis reveals that CETS is the most significant feature, where lower values are strongly predictive of hospitalization. This is followed by oxygen saturation and age, then heart rate, lymphocytes, and eosinophils. Among the top 20 features, three engineered attributes are presented: the difference in NO_2 levels 48 h and 168 h prior, the difference in PM_{10} values 24 h and 48 h prior, and the difference in CO values 24 h and 48 h prior. These attributes represent changes in pollutant levels, where higher values (closer to red) correlate with increased pollution and a greater predicted likelihood of hospitalization. The inclusion of these engineered features in the top 20 highlights their practical value in enhancing the model's predictive accuracy.

Figure 6 illustrates the relative risk of hospitalization associated with the top three features and a significant air pollution attribute, as follows:

(a) CETS (Figure 6a): This is a crucial feature in assessing the risk of hospitalization for asthma. A CETS score below 2 significantly increases the likelihood of hospitalization, indicating a high risk.
(b) Oxygen Saturation (Figure 6b): This is another crucial indicator. A saturation level of 93% is identified as the critical threshold. The blue dots in the figure represent patients with oxygen saturation below this level, indicating a heightened risk of hospitalization.
(c) Age (Figure 6c): This plays a significant role in hospitalization risk, with 58 years identified as a critical turning point beyond which the probability of requiring hospitalization increases.
(d) NO_2 (Figure 6d): This feature represents the difference in the NO_2 concentrations at 48 h and 168 h prior. A differential value of 0 is the turning point, with values exceeding this mark significantly raising the probability of hospitalization. This indicates that increases in NO_2 levels over time correlate with a higher risk of hospitalization.

Top-20-Feature Cumulative Effect

To examine the impact of different variables on our prediction model, we utilized cumulative feature effect methods. We began by identifying the top 20 features through a ranking based on both SHAP values and the inherent feature importance of the RF model. The experiment involved removing these top 20 features from a set of 94, then training

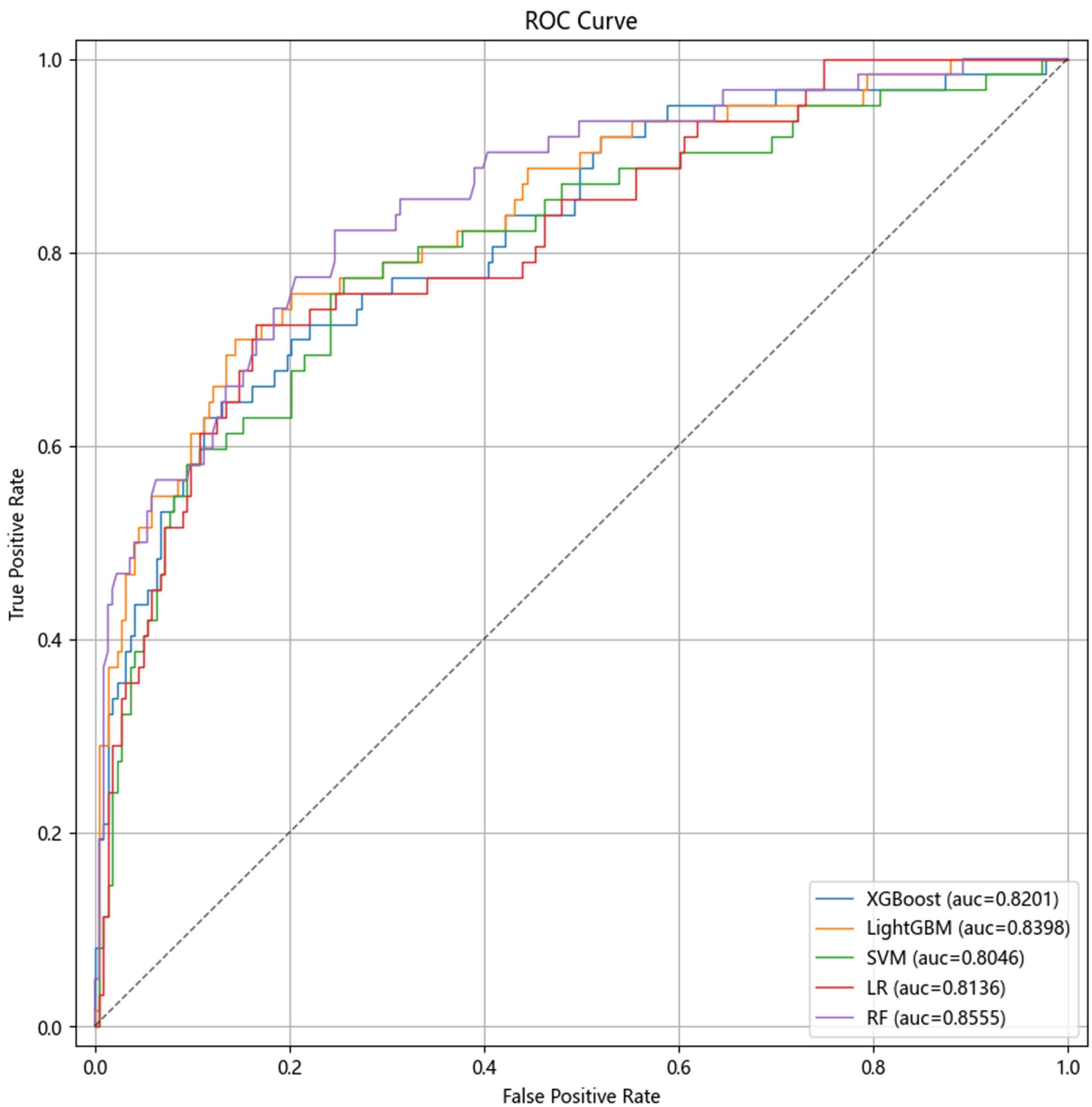

Figure 4 Receiver Operating Characteristic (ROC) Curves for Models Incorporating External Environmental Factors. This figure displays ROC curves for models enhanced with environmental data.

Abbreviations: XGB, Extreme Gradient Boosting; LightGBM, Light Gradient Boosting Machine; SVM, Support Vector Machines; LR, Logistic Regression; RF, Random Forest; NPV, Negative Predictive Value; PPV, Positive Predictive Value; AUC, Area Under the Receiver Operating Characteristic Curve.

five base models and calculating the mean values of the metrics to establish the base values. Features were incrementally reintroduced in ascending order of importance, undergoing five training sessions and subsequent metric evaluations. This approach yielded 21 sets of outcomes, with feature counts progressively increasing from 74 to 94, as depicted in Figure 7. We applied min-max scaling to normalize each metric for enhanced visualization. The normalized results were plotted along the horizontal axis, with zero as the midpoint; the left side depicted features selected based on SHAP values, and the right side depicted those based on the importance of the RF features.

The diagram shows that the top three features—CETS, oxygen saturation, and age—remained consistent across both model interpretation methodologies, highlighting their strong correlation with hospitalization for asthma. The primary

Table 6 Hyperparameter Combination of Each Model

Model	Optimal hyperparameters
XGBoost	Learning_rate=0.09, max_depth=25, n_estimator=28
LightGBM	Learning_rate=0.1, num_leaves=30, max_depth=25, n_estimators=54
SVM	C=0.9, kernel='rbf', degree=5, class_weight='balanced', probability= 'True', tol=1×10^{-3}
LR	Penalty='L2', C=0.85, class_weight='balanced', tol=1×10^{-4}
RF	n_estimators=25, max_depth=35, min_samples_leaf=1, min_samples_split=2, class_weight='balanced'

Abbreviations: XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; SVM, support vector machines; LR, logistic regression; RF, random forest.

difference between the methodologies was observed in the variable Allergic Rhinitis and in the difference in SO_2 levels at 48 h and 168 h prior. Notably, the difference was significant in the classification process of the RF model, affirming the importance of this engineered feature.

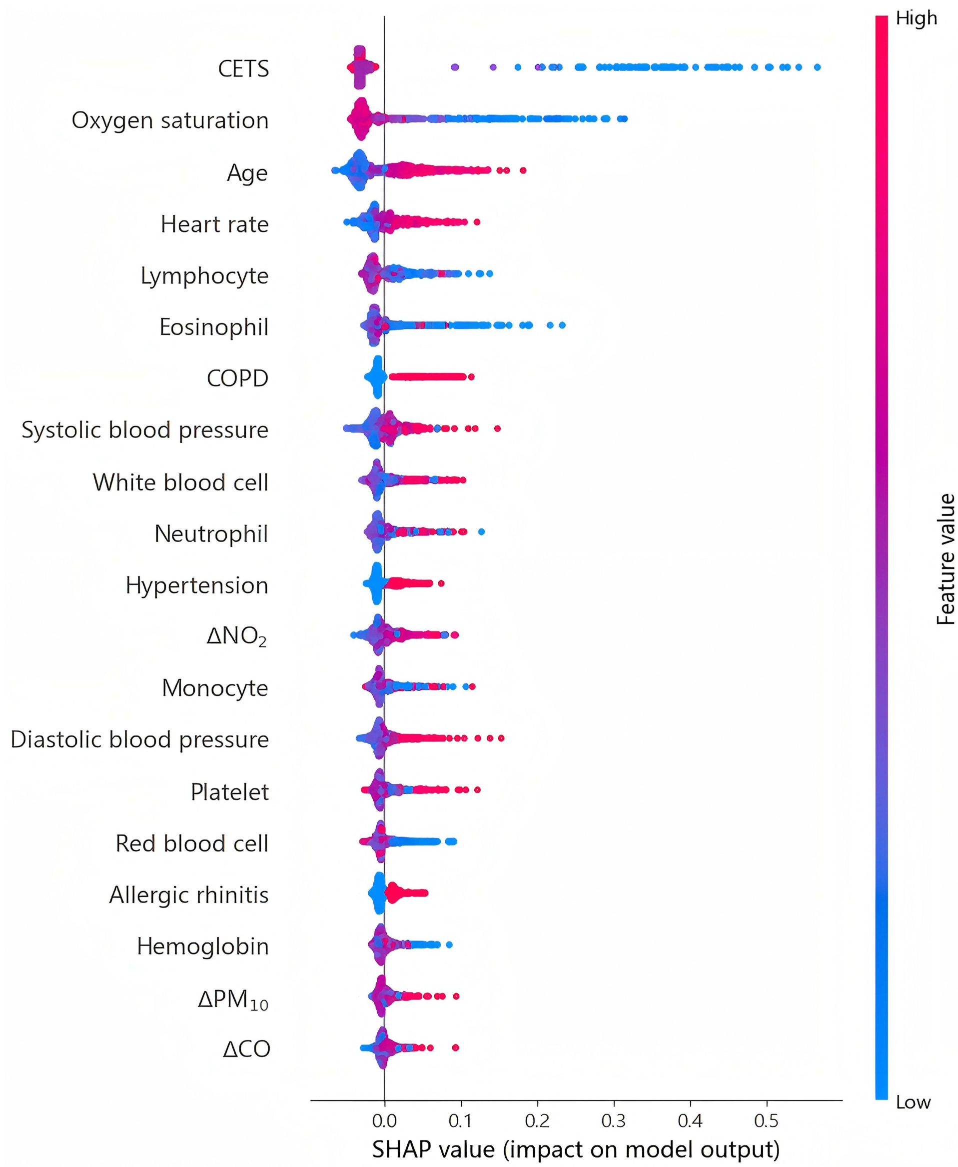

Figure 5 Prediction Model Using Random Forest (RF). This figure outlines the importance of different features in the RF model for predicting hospitalizations. Features are ranked by the magnitude of their SHAP values; higher absolute values indicate greater influence. Conditions include COPD, Hypertension, and Allergic Rhinitis (coded as 1 for present, 0 for absent).

Abbreviations: CETS, Chinese Emergency Triage Scale; COPD, Chronic Obstructive Pulmonary Disease; NO_2, nitrogen dioxide; PM_{10}, particulate matter with an aerodynamic diameter ≤10 μm; CO, carbon monoxide; ΔNO_2, difference in values of NO_2 concentrations at 48 h and 168 h before the ED visit; ΔPM_{10}, difference in values of the PM_{10} concentrations at 24 h and 48 h before the ED visit; ΔCO, difference in values of the CO concentrations at 24 h and 48 h before the ED visit.

https://doi.org/10.2147/JAA.S512405

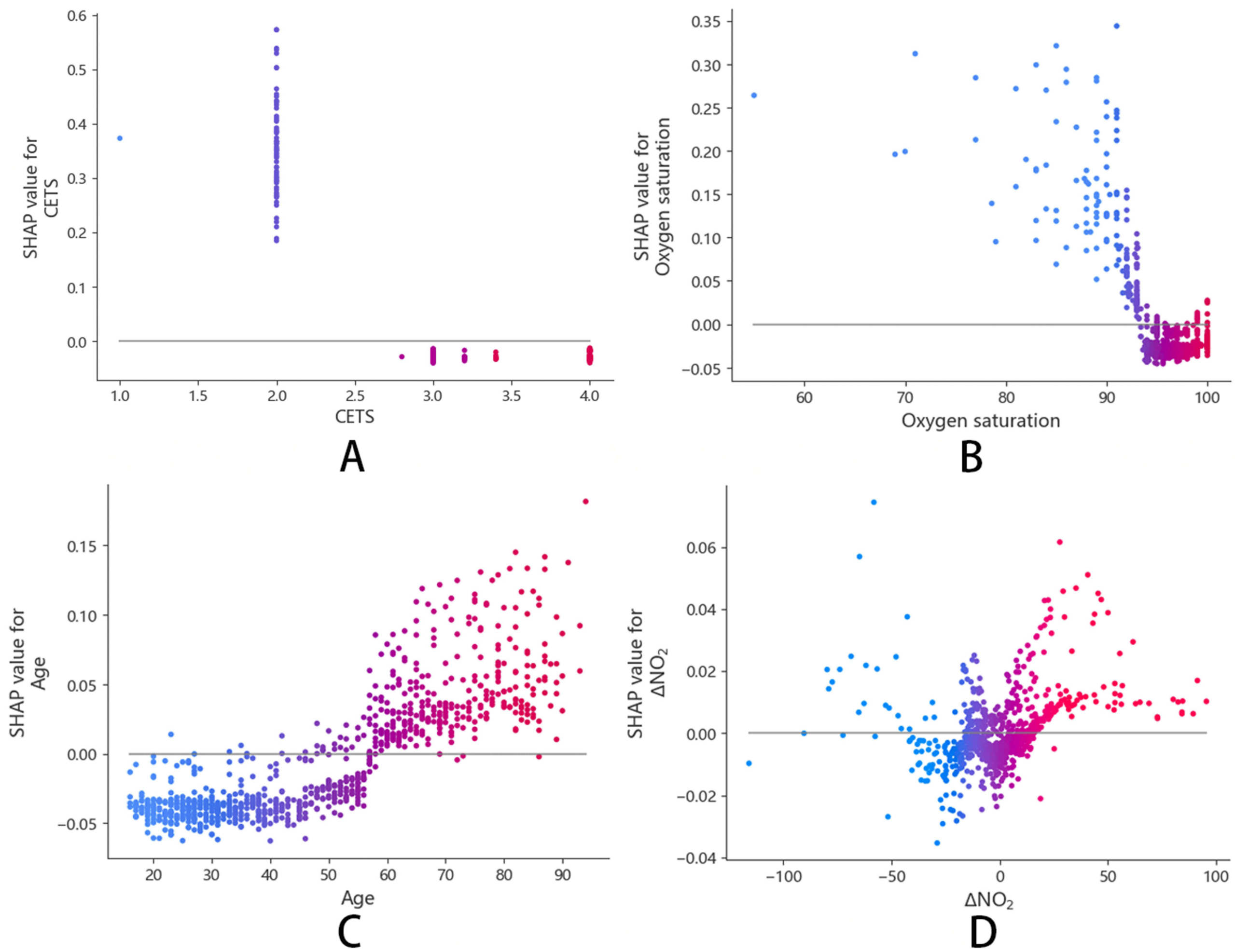

Figure 6 Relative Risk of Different Outcomes with Independent Variables in Hospitalization Prediction Models. (**A**) Relative risk of different outcomes with CETS; (**B**) Relative risk of different outcomes with Oxygen saturation; (**C**) Relative risk of different outcomes with Age; (**D**) Relative risk of different outcomes with ΔNO_2.

Abbreviations: CETS, Chinese Emergency Triage Scale; ΔNO_2, the difference in values of the NO_2 concentrations at 48 h and 168 h before the ED visit.

Discussion

In this study, prediction models were initially trained and evaluated using 22 in-hospital variables. RF performed at high levels on all metrics except Specificity and PPV, achieving an optimal AUC of 0.8272. We sequentially integrated ambient air pollutant and meteorological features to enhance the model's predictive capability. And the RF model consistently outperformed the others, achieving an AUC of 0.8555. To the best of the authors' knowledge, this is the first study to use multiple sources of data to construct predictive models that assist clinicians in identifying adult asthma patients at high risk of hospitalization.

Over the past decade, the expansion of large clinical data sources and advances in computational power have fueled the growth of ML applications in healthcare.[9] The line between ML and statistical modeling has been described as a continuum.[9] Unlike statistical models that rely on theory and assumptions, ML models learn directly and automatically from data.[27] No definitive threshold exists at which a model is considered to be a "machine learning model", as all methods exist on a continuum depending on the degree of human assumptions imposed.[9] Compared with traditional methods, ML algorithms enable a more flexible relationship between predictor variables and outcomes, rendering them preferable for predictive rather than explanatory studies.[10] Our study demonstrated that the RF algorithm exhibited

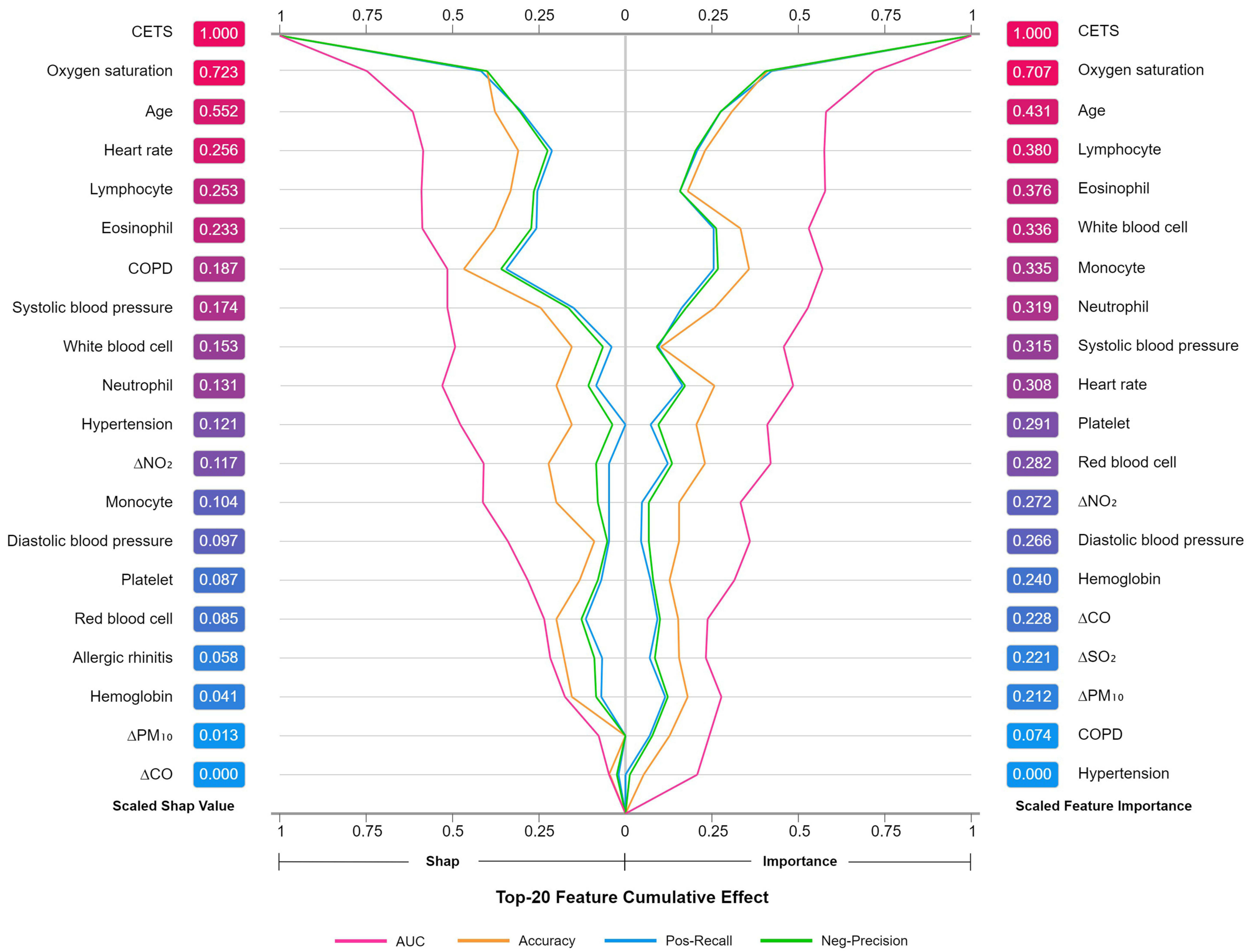

Figure 7 Top-20-Feature Cumulative Effect.

Abbreviations: CETS, Chinese Emergency Triage Scale; COPD, Chronic Obstructive Pulmonary Disease; NO_2, nitrogen dioxide; PM_{10}, particulate matter with an aerodynamic diameter ≤10 µm; CO, carbon monoxide; ΔNO_2, difference in values of the NO_2 concentrations at 48 h and 168 h before the ED visit; ΔPM_{10}, difference in values of the PM_{10} concentrations at 24 h and 48 h before the ED visit; ΔCO, difference in values of the CO concentrations at 24 h and 48 h before the ED visit; ΔSO_2, difference in values of the SO_2 concentrations at 48 h and 168 h before the ED visit.

superior performance in predicting hospitalization outcomes among adult asthma patients presenting to the emergency department. The algorithm's effectiveness stems from its ensemble approach, which utilizes multiple decision trees to capture complex hierarchical relationships and interaction effects among variables. This enables it to handle high-dimensional datasets with collinear features proficiently, as seen in our study integrating clinical, meteorological, and air pollution data.

ML models are often called "black boxes", as they input data and produce outcomes without visible internal processes.[28] To address this issue, the SHAP method was used in this study to explain the outcomes of the prediction model. Emergency triage is crucial in high-acuity EDs, and since the 1990s, several triage scales have been implemented in developed countries. In China, emergency triage faces unique challenges compared to developed countries.[19] CETS, which relies on vital complaints and objective data, is a reliable and valid method for ED triage in mainland China.[19] This study corroborated previous findings that lower CETS scores, indicating greater illness severity, were associated with a higher risk of hospitalization for asthma. Consistent with earlier research,[8] initial oxygen saturation was also a significant predictor of hospitalization.[13,14,29] Older age, which is linked to reduced lung function and a higher prevalence of chronic disease comorbidities, increased the likelihood of hospitalization.[2,30,31] Eosinophil levels typically drop during severe asthma exacerbations,[12] because the strong inflammatory response recruits eosinophils to the airways and tissues. Air pollution factors, notably NO_2 and PM_{10}, have been significantly associated with asthma exacerbations,[32,33] potentially because of oxidative stress, airway hyper-responsiveness, and remodeling.[34–38]

The rapid development of electronic medical record systems presents a unique opportunity to create clinical decision support systems (CDSS). These are computer programs that apply knowledge to data stored in EHRs to facilitate intelligent applications such as auxiliary diagnosis, treatment assistance, risk warning, medical record quality control, and infectious disease management; they have proved to be a valuable and effective tool in clinical practice.[39–41] A previous study developed a CDSS by integrating EHRs with BMJ Best Practice, achieving accuracy rates of 75.46% in first-rank diagnosis and 87.53% in top-three diagnoses;[42] this highlights the significant clinical application potential of CDSS in EHRs.[42] The predictive models developed in this study could be embedded in a CDSS to guide clinicians in making informed decisions about hospitalizing asthma patients and to enhance resource utilization.

This study has several limitations. First, the sample size was limited, and future research will require larger samples to enhance model accuracy. Second, data on ambient air pollutants and meteorological variables may not fully reflect individual exposure levels due to the omission of activities conducted away from home. Finally, the retrospective design of this study is constrained by missing data; for instance, only 31.6% of patients had recorded respiratory rate data, a factor associated with hospital admission.[13]

In conclusion, ML models based on clinical, meteorological, and air pollution data can rapidly and accurately predict hospitalization of adult asthma patients in EDs. This study lays a foundational framework for future research where similar methodologies can be applied to other chronic conditions influenced by environmental factors. Future work will focus on incorporating these models into clinical decision tools to assist clinicians in determining the need for hospitalization of asthma patients and to improve resource utilization for maximizing ED efficiency.

Abbreviation

AUC, area under the curve; CETS, Chinese Emergency Triage Scale; CHAP, China High Air Pollutants; ECMWF, European Centre for Medium-Range Weather Forecasts; ED, emergency department; ERA5, fifth generation of European ReAnalysis; LightGBM, light gradient boosting machine; LR, logistic regression; ML, machine learning; NPV, negative predictive value; PPV, positive predictive value; RF, random forest; ROC, receiver operating characteristic; SHAP, Shapley additive explanations; SVM, support vector machines; TAP, Tracking Air Pollution in China; XGBoost, extreme gradient boosting.

Ethical Statement

This study was approved by the Peking University Third Hospital Medical Science Research Ethics Committee (Approval No: IRB00006761-M2021582). The committee waived the requirement for individual patient informed consent for the review of their medical records. This waiver was granted because the study was retrospective in nature, utilizing existing medical records collected during routine clinical care, and all patient-identifiable information was removed before access and analysis by the research team. This methodological approach ensured that patient privacy and confidentiality were rigorously protected at all times. The study was conducted in compliance with the principles of the Declaration of Helsinki.

Acknowledgments

We are grateful to Yingying Ge and Shengrong Zhu for their invaluable contributions to data check and statistical analysis.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by the Capital Health Research and Development of Special (2024-3-40917), the Clinical Key Project of Peking University Third Hospital (BYSYZD2023009), and the Clinical Cohort Construction Project of Peking University Third Hospital (BYSYDL2021020).

Disclosure

All authors have read and understood the journal policy on declaration of interests and declare that they have no financial or non-financial competing interests in this work.

References

1. Soriano JB, Kendrick PJ, Gupta V, et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet Respir Med*. 2020;8(6):585–596. doi:10.1016/S2213-2600(20)30105-3
2. Global Initiative for Asthma. Global strategy for asthma management and prevention, 2024. Available from: https://www.ginasthma.org/reports. (Accessed September 15, 2024.).
3. Huang K, Yang T, Xu J, et al. Prevalence, risk factors, and management of asthma in China: a national cross-sectional study. *Lancet*. 2019;394 (10196):407–418. doi:10.1016/S0140-6736(19)31147-X
4. Krishnan V, Diette GB, Rand CS, et al. Mortality in patients hospitalized for asthma exacerbations in the United States. *Am J Respir Crit Care Med*. 2006;174(6):633–638. doi:10.1164/rccm.200601-007OC
5. Finkelstein EA, Lau E, Doble B, et al. Economic burden of asthma in Singapore. *BMJ Open Respir Res*. 2021;8(1). doi:10.1136/bmjresp-2020-000654.
6. Accordini S, Bugiani M, Arossa W, et al. Poor control increases the economic cost of asthma. A multicentre population-based study. *Int Arch Allergy Immunol*. 2006;141(2):189–198. doi:10.1159/000094898
7. Bilaver LA, Chadha AS, Doshi P, et al. Economic burden of food allergy: a systematic review. *Ann Allergy Asthma Immunol*. 2019;122(4):373–380e1. doi:10.1016/j.anai.2019.01.014
8. Sills MR, Ozkaynak M, Jang H. Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning. *Int J Med Inform*. 2021;151:104468.
9. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318. doi:10.1001/jama.2017.18391
10. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J*. 2017;38(23):1805–1814. doi:10.1093/eurheartj/ehw302
11. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233–234. doi:10.1038/nmeth.4642
12. Zein JG, Wu C-P, Attaway AH, et al. Novel machine learning can predict acute asthma exacerbation. *Chest*. 2021;159(5):1747–1757. doi:10.1016/j.chest.2020.12.051
13. Patel SJ, Chamberlain DB, Chamberlain JM. A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Acad Emerg Med*. 2018;25(12):1463–1470. doi:10.1111/acem.13655
14. Goto T, Camargo CA, Faridi MK, et al. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *Am J Emerg Med*. 2018;36(9):1650–1654. doi:10.1016/j.ajem.2018.06.062
15. C LH, M LA, Y CE, et al. The short-term association between asthma hospitalisations, ambient temperature, other meteorological factors and air pollutants in Hong Kong: a time-series study. *Thorax*. 2016;71(12):1097–1109. doi:10.1136/thoraxjnl-2015-208054
16. Newson R, Strachan D, Archibald E, et al. Acute asthma epidemics, weather and pollen in England, 1987-1994. *Eur Respir J*. 1998;11(3):694–701. doi:10.1183/09031936.98.11030694
17. G SJ, A BM, Combs V, et al. Identifying impacts of air pollution on subacute asthma symptoms using digital medication sensors. *Int J Epidemiol*. 2022;51(1):213–224. doi:10.1093/ije/dyab187
18. Zheng X, Ding H, Jiang L, et al. Association between air pollutants and asthma emergency room visits and hospital admissions in time series studies: a systematic review and meta-analysis. *PLoS One*. 2015;10(9):e0138146. doi:10.1371/journal.pone.0138146
19. Zhiting G, Jingfen J, Shuihong C, et al. Reliability and validity of the four-level Chinese emergency triage scale in mainland China: a multicenter assessment. *Int J Nurs Stud*. 2020;101:103447. doi:10.1016/j.ijnurstu.2019.103447
20. Xiao QY, Geng GN, Liu SG, et al. Spatiotemporal continuous estimates of daily 1 km PM 2.5 from 2000 to present under the Tracking Air Pollution in China (TAP) framework. *Atmos Chem Phys*. 2022;22(19):13229–13242. doi:10.5194/acp-22-13229-2022
21. Wei J, Li ZQ, Wang J, et al. Ground-level gaseous pollutants (NO 2, SO 2, and CO) in China: daily seamless mapping and spatiotemporal variations. *Atmos Chem Phys*. 2023;23(2):1511–1532. doi:10.5194/acp-23-1511-2023
22. Wei J, Liu S, Li ZQ, et al. Ground-level NO surveillance from space across China for high resolution using interpretable spatiotemporally weighted artificial intelligence. *Environ Sci Technol*. 2022;56(14):9988–9998. doi:10.1021/acs.est.2c03834
23. Hersbach H, Bell B, Berrisford P, et al. ERA5 hourly data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)[data set]. 2023.
24. Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. *Organ Res Methods*. 2013;16 (2):270–301. doi:10.1177/1094428112470848
25. Geapa B, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor*. 2004;6 (1):20–29. doi:10.1145/1007730.1007735
26. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30 (NIPS 2017). Available from: https://papers.nips.cc/paper/7062-a-unifiedapproach-to-interpretingmodel-predictions. Accessed 21 Sep 2022.).
27. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22. doi:10.1016/j.jclinepi.2019.02.004

28. The Lancet Respiratory Medicine LRM. Opening the black box of machine learning. *Lancet Respir Med*. 2018;6(11):801. doi:10.1016/S2213-2600(18)30425-9
29. Dhmdmph A, Tmph G, Kgmp M, et al. Development and internal validation of a pediatric acute asthma prediction rule for hospitalization. *Eur Res Rev*. 2014;3(2):228–235.
30. Boyd CM, Darer J, Boult C, et al. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *JAMA*. 2005;294(6):716–724. doi:10.1001/jama.294.6.716
31. Rowe BH, Villa-Roel C, Abu-Laban RB, et al. Admissions to Canadian hospitals for acute asthma: a prospective, multicentre study. *Can Respir J*. 2010;17(1):25–30. doi:10.1155/2010/178549
32. Guarnieri MMD, Balmes JRD. Outdoor air pollution and asthma. *Lancet*. 2014;383(9928):1581–1592. doi:10.1016/S0140-6736(14)60617-6
33. X-y Z, Orellano P, H-l L, et al. Short-term exposure to ozone, nitrogen dioxide, and sulphur dioxide and emergency department visits and hospital admissions due to asthma: a systematic review and meta-analysis. *Environ Int*. 2021;150:106435. doi:10.1016/j.envint.2021.106435
34. Stanek LW, Brown JS, Stanek J, et al. Air pollution toxicology—a brief review of the role of the science in shaping the current understanding of air pollution health risks. *Toxicol Sci*. 2011;120(1):S8–S27. doi:10.1093/toxsci/kfq367
35. Zhang Y, Peng L, Kan H, et al. Effects of meteorological factors on daily hospital admissions for asthma in adults: a time-series analysis. *PLoS One*. 2014;9(7):e102475. doi:10.1371/journal.pone.0102475
36. Samoli E, Nastos PT, Paliatsos AG, et al. Acute effects of air pollution on pediatric asthma exacerbation: evidence of association and effect modification. *Environ Res*. 2011;111(3):418–424. doi:10.1016/j.envres.2011.01.014
37. Zora JE, Sarnat SE, Raysoni AU, et al. Associations between urban air pollution and pediatric asthma control in El Paso, Texas. *Sci Total Environ*. 2013;448:56–65. doi:10.1016/j.scitotenv.2012.11.067
38. Mazenq J, Dubus JC, Gaudart J, et al. City housing atmospheric pollutant impact on emergency visit for asthma: a classification and regression tree approach. *Respir Med*. 2017;132:1–8. doi:10.1016/j.rmed.2017.09.004
39. Roshanov PS, Misra S, Gerstein HC, et al. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implement Sci*. 2011;6:92. doi:10.1186/1748-5908-6-92
40. Schedlbauer A, Prasad V, Mulvaney C, et al. What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? *J Am Med Inform Assoc*. 2009;16(4):531–538. doi:10.1197/jamia.M2910
41. Wang M, Lee C, Wei Z, et al. Clinical assistant decision-making model of tuberculosis based on electronic health records. *Biodata Min*. 2023;16(1). doi:10.1186/s13040-023-00328-y.
42. Tao LY, Zhang C, Zeng L, et al. Accuracy and effects of clinical decision support systems integrated with BMJ Best Practice-aided diagnosis: interrupted time series study. *JMIR Med Inf*. 2020;8(1):56–70.