scientific reports

OPEN

Predicting hospital outpatient volume using XGBoost: a machine learning approach

Lingling Zhou[1,2], Qin Zhu[1,2], Qian Chen[1], Ping Wang[1] & Hao Huang [1]✉

Hospital outpatient volume is influenced by a variety of factors, including environmental conditions and healthcare resource availability. Accurate prediction of outpatient demand can significantly enhance operational efficiency and optimize the allocation of medical resources. This study aims to develop a predictive model for daily hospital outpatient volume using the XGBoost algorithm. Meanwhile, the forecasting performance was compared with that of the Seasonal AutoRegressive Integrated Moving Average with exogenous regressors (SARIMAX) and Random Forest (RF) models. The dataset comprises daily climate data (e.g., temperature, precipitation, PM2.5 levels), historical outpatient volume records, and the number of outpatient specialists available each day. The data range involved spans from January 1, 2014, to October 31, 2024. Data preprocessing involved addressing missing values and encoding categorical variables. Model performance was assessed using three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) , Mean Absolute Percentage Error (MAPE), and R-squared (R^2) metrics. The XGBoost model exhibited superior predictive accuracy compared to both the SARIMAX and RF models, with the lowest MAE, RMSE, MAPE, and the highest R^2, successfully capturing key relationships between climate factors, resource availability, and outpatient volume. The number of outpatient specialists, temporal variables (such as year, quarter, month, and weekday), meteorological conditions (average temperature), and air quality (PM2.5) had the most significant impact on the prediction model. This study underscores the potential of machine learning algorithms like XGBoost in effectively predicting hospital outpatient demand. The findings offer valuable insights for hospitals to make proactive adjustments to their resource allocation, thereby improving their service capacity.

Keywords XGBoost, Outpatient volume, Machine learning, Climate data, Hospital resource planning, Predictive analytics

In recent years, there has been increasing interest in the healthcare sector regarding the complex relationships between environmental factors and human health outcomes[1,2]. Among these factors, meteorological data has emerged as a key contributor to predicting and managing healthcare demand, particularly in relation to hospital outpatient visits. Understanding this relationship is not only essential for improving the efficiency of healthcare systems but also for ensuring the optimal allocation of scarce resources[3].

However, few studies have investigated the combined effects of meteorological conditions and air quality on hospital outpatient visits, while also considering the influence of physicians' professional titles. It is hypothesized that variations in weather patterns and air pollution levels may lead to fluctuations in patient demand for medical services, which could be moderated by the expertise and experience of healthcare providers.

In this study, we aim to investigate the relationship between meteorological data, air quality, and hospital outpatient volume, incorporating the number of available in outpatient specialists in the hospital. We will utilize the XGBoost model to analyze historical data from the China Meteorological Data Network, China Environmental Monitoring Stations, and medical institutions to identify key meteorological factors that significantly influence outpatient visits, and to make further predictions based on these factors.

Our findings will provide valuable insights into the complex interplay between environmental factors, healthcare resource allocation, and patient care needs. Furthermore, these results may guide policymakers and healthcare administrators in formulating strategies to optimize service delivery during periods of adverse weather or heightened pollution levels.

[1]Department of Information, Daping Hospital, Army Medical University, Chongqing 400042, China. [2]Lingling Zhou and Qin Zhu contributed equally. ✉email: zllgwy@tmmu.edu.cn

These variations in meteorological conditions can lead to predictable fluctuations in hospital outpatient volumes, highlighting the need for a thorough understanding of the underlying mechanisms. By integrating real-time and historical meteorological data, researchers and healthcare administrators can develop more advanced prediction models that account for environmental influence on patient behavior and health outcomes. This approach can enable hospitals to better anticipate demand surges, optimize staffing levels, and allocate resources more effectively[4–7].

Related work

Meteorological conditions, including temperature, humidity, precipitation, and wind speed, can significantly affect human physiology and behavior, influencing the likelihood of individuals seeking medical attention[8]. For example, fluctuations in temperature and PM2.5 levels have been associated with a higher incidence of cardiovascular diseases[9–11]. Elevated temperatures and high concentrations of gaseous pollutants are strongly linked to an increased risk of cerebrovascular disease[12]. Furthermore, higher levels of PM2.5, PM10, NO2, SO2, and O3 have been shown to raise the risk of mental disorders in China[13]. Short-term increases in PM2.5 and PM10 are also correlated with a higher incidence of upper respiratory tract infections[14]. Additionally, SO2, NO2, and PM10 may impact the daily outpatient volume of patients with dermatitis[15].

Previous researches into forecasting patient admissions have revealed that traditional statistical methods, such as linear regression, residual Error, and time-series analysis (eg.ARIMA)[4,16,17], often struggle with capturing complex, non-linear relationships and interactions between variables, particularly in the presence of multicollinearity and high-dimensional data. Additionally, traditional approaches may not fully account for the dynamic and seasonal nature of patient flows, such as day-of-week variations or long-term trends, limiting their predictive accuracy and practical utility for hospital decision-making.

Machine learning algorithms, which offer powerful predictive capabilities in big data contexts, have not yet been extensively applied in this field of research[16]. Machine learning has emerged as a transformative tool in healthcare analytics, enabling predictive models that can process large volumes of heterogeneous data[18]. Among these algorithms, XGBoost (Extreme Gradient Boosting) is particularly notable for its strong performance with structured data and its ability to model complex interactions between variables[19,20]. XGBoost's robustness to overfitting, scalability, and interpretability further enhance its suitability for healthcare applications, where accurate and actionable predictions are critical[21].

Materials and methods
Study design

This retrospective study incorporated daily meteorological data, air quality data, hospital outpatient visits, and the number of outpatient specialists from January 1, 2014 to October 31, 2024. Due to the impact of COVID-19[22] on healthcare services, data from January 23, 2020 to March 23, 2020 were excluded from the analysis. The meteorological data were sourced from the China Meteorological Data Network, air quality data from the China Environmental Monitoring Station, and hospital outpatient visit and specialist data from our hospital's information system. We developed the XGBoost model to forecast hospital outpatient visits, with its predictive performance being systematically evaluated against two benchmark models: the Seasonal AutoRegressive Integrated Moving Average with Exogenous Regressors (SARIMAX) and Random Forest (RF). Manually selecting the optimal combination of these parameters is challenging and time-consuming. To optimize the performance of the three model, we employed grid search (GS), a widely adopted parameter optimization technique, to systematically determine the optimal parameter configuration. All computational procedures and statistical analyses were implemented in Python 3.7 (Python Software Foundation, Wilmington, DE, USA), utilizing essential scientific computing libraries including NumPy (version 1.21.0) for numerical computations, pandas (version 1.3.0) for data manipulation, and scikit-learn (version 0.24.2) for machine learning implementation, thereby ensuring reproducibility and methodological rigor.

Research indicators
Meteorological data

Daily weather records, including maximum temperature (Max-T), minimum temperature (Min-T), mean temperature (Mean-T), mean wind speed (Mean-S), relative humidity (RH), precipitation (Prec), mean sea level pressure (MSL). Air quality data: Daily measurements of pollutants, including the air quality index (AQI), as well as levels of PM2.5, PM10, CO, NO2, O3 and SO2. Hospital Expert Staffing: Data sourced from hospital human resources records, detailing the number of outpatient specialists available each day (Clinc-E). Hospital outpatient data: Historical records of daily outpatient visit counts (ppl-count).

Statistical processing

First, variables with missing data below 30% were imputed using the mean values. Next, the normality evaluations employed Shapiro–Wilk (SW) tests[23] to assess distributional assumptions, while Levene's test examined homogeneity of variance across groups. Spearman's rank-order correlation coefficients (ρ) were computed to systematically evaluate the associations among continuous variables[24]. Kruskal–Wallis H tests, Dunn's statistic, and separate Mann–Whitney tests were applied to assess the explanatory power of categorical variables on the target variable's variance[25]. The categorical variables include Year, Quarter, Month, Day of month, Weekday and Weekend.

After selecting input variables based on the previous statistical test, the initial dataset was divided into two subsets: the first (covering the period from January 1, 2014, to September 30, 2024) served as the basis for model development, while the second subset (encompassing data from October 2024) was reserved for testing the

model's performance. Within the development period, for both XGBoost and RF models, the data were further divided into training and validation sets through random sampling: 80% of the cases (n = 3092) were allocated to the training set, while the remaining 20% (n = 773) were reserved for validation purposes. The SARIMAX model was trained on the entire training data (the first subset), with an external test set serving as the validation window (the second subset).

Model building

XGBoost model construction based on GS
XGBoost is an efficient and powerful ensemble learning algorithm that enhances model accuracy by iteratively building multiple decision trees and optimizing the residuals from the previous step. It excels not only in handling large datasets but also in preventing overfitting, making it widely used for classification, regression, and ranking tasks. The objective function of XGBoost at iteration *t*, denoted as $Obj(t)$, is defined as follows:

$$Obj(t) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where: *n* is the number of samples. y_i is the actual value of the i-th sample. $\hat{y}_i$ is the predicted value of the i-th sample. *K* is the number of boosted trees (weak learners). $L(y_i, \hat{y}_i)$ is the loss function that measures the difference between the actual and predicted values. f_k represents the k-th tree learned by XGBoost. $\Omega(f_k)$ is the regularization term that penalizes the complexity of the tree, which is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2 \Omega(f)$$

Here, *T* is the number of leaves in the tree, γ is a parameter that defines the minimum loss reduction required to make a further partition on a leaf, ω is the vector of scores on the leaves. λ is a parameter that controls the L2 regularization term on the leaf scores.

By optimizing this objective function, XGBoost iteratively adds new trees to the ensemble, with each tree focusing on correcting the errors made by the previous ones. This process continues until the desired level of accuracy is achieved or until a specified number of trees are built. To optimize model performance, a GS technique was applied to find for the optimal parameters.

RF model construction based on GS
A Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and control overfitting. It operates on the principle of "wisdom of the crowd," where aggregating predictions from diverse models (trees) reduces variance and enhances robustness. Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve overall model performance.

For regression problems, Random Forest makes predictions by averaging the prediction results of multiple trees. Assume there are n trees, and the prediction result of each tree is $h_m(x)$. The overall prediction result $H(x)$ of Random Forest can be represented as:

$$H(x) = \frac{1}{n} \sum_{m=1}^{n} h_m(x)$$

SARIMAX model construction based on GS
The SARIMAX(*p, d, q*)(*P, D, Q*)$_s$ model is a powerful tool for time series forecasting. It extends the ARIMA model by incorporating seasonality and exogenous variables, making it suitable for datasets with trends, seasonal patterns, and external influences. The performance of a SARIMAX model heavily depends on the selection of its hyper parameters, which include:

Non-seasonal parameters: *p* (autoregressive order), *d* (differencing order), and *q* (moving average order).

Seasonal parameters: *P* (seasonal autoregressive order), *D* (seasonal differencing order), *Q* (seasonal moving average order), and *s* (seasonal period).

Exogenous variables: External factors that influence the time series.

Model evaluation

The developed model was evaluated to assess its effectiveness in predicting outpatient volume. Model performance was measured using metrics such as R-squared (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). MAE measures the average absolute error between predicted and actual values, yielding a non-negative value. The smaller the MAE, the better the model's performance. It is less sensitive to outliers because it only considers the absolute differences between predicted and actual values (Eq. 1). Hear, y_i represents the true value, $\hat{y}_i$ is the predicted value from the model, and m is the number of samples.

$$MAE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \tag{1}$$

RMSE measures the difference between predicted and actual values by calculating the square root of the average of the square errors. The smaller the RMSE, the higher the accuracy of the model's predictions. RMSE gives more weight to larger errors, making it sensitive to outliers. It reflects the degree to which the predicted values deviate from the true values (Eq. 2).

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{2}$$

R^2 is also known as the coefficient of determination, measure how well a model fits the data. The value of R^2 is ranges 0 to 1, with a value closer to 1 indicating a better fit. When $R^2 = 1$, the model is perfectly accurate. If R^2 is less than 0, the model performs worse than the benchmark model. As shown in Eq. (3) the numerator represents the sum of the squared deviations between the true values and the predicted values, while the denominator represents the sum of the squared differences between the true values and the mean value.

$$R^2 = \mathbf{1} - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y}_i)^2} \tag{3}$$

This validation approach helps assess the model's performance across different subsets of the data, ensuring its robustness and generalizability.

MAPE is a statistical metric used to measure the accuracy of predictions by calculating the average absolute percentage error between predicted values and actual values (Eq. 4). It expresses the relative error in percentage terms, with lower values indicating more accurate predictions.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100\% \tag{4}$$

## Result
Correlation analysis
The Shapiro–Wilk tests indicated that all continuous variables deviated significantly from normality ($p < 0.05$). The statistical descriptions of continuous variables from January 1, 2014, to October 31, 2024 are shown in Table 1.

The SW correlation analysis revealed that all continuous variables ($p < 0.05$) except for NO2 exhibited significant correlations with outpatient volume, with Clinic-E showing the strongest correlation ($\rho = 0.8$) (Fig. 1). The remaining variables were ranked in descending order of correlation strength as follows: PM2.5, PM10, Max-T, Mean-T, CO, Mean-S, MSL, Min-T, O3, AQI, RH, and SO2.

Distribution of outpatient volumes under different time periods are presented in Fig. 2. All categorical variables met the normality assumption and homogeneity of variance ($p < 0.05$). The H-values for Year, Quarter, Month, Day of month, and Weekday were 245.80, 45.62, 118.62, 36.65, and 2130.10, respectively. Except for the Day of month ($p > 0.05$), other variables were found to be significant ($p < 0.05$). The Mann–Whitney test result indicated a significant difference between weekday and weekend ($p < 0.05$). For different years and months, the pairs identified by Dunn's test are presented in Tables 2 and 3. Both tables include all statistically significant

Variables	Median	Interquartile range (IQR)	Median absolute deviation (MAD)	Skewness	Peakness
Max-T (℃)	23.2	15.0–30.1	7.4	−0.05	−1.1
Min-T (℃)	15.1	8.3–20.8	6.2	−0.13	−1.1
Mean-S (km/h)	5.7	4.8–6.7	0.9	1.1	2
Mean-T (℃)	18.4	10.9–20.5	6.7	−0.03	−1.1
RH (%)	78	71–85	7	−0.58	−0.16
Prec (mm)	0.5	0–2.6	0.5	3.5	16
MSL (hPa)	1014	1006–1022	8	0.15	−0.96
AQI	65	49–89	19	1.5	3.2
PM2.5	33	22–53	14	1.9	4.3
PM10	55	36–81	21	1.6	3.8
CO	0.8	0.7–1.0	0.1	1.4	4.2
NO2	36	27–45	9	0.61	0.18
O3	63	34–107	34	0.72	−0.16
SO2	9	8–12	2	3.7	20
Clinc-E	66	33–113	35	1.3	1.2
ppl-count	7257	3559–8420	1621	−0.42	−1.1

Table 1. Statistical description and distribution characteristics of continuous variables.
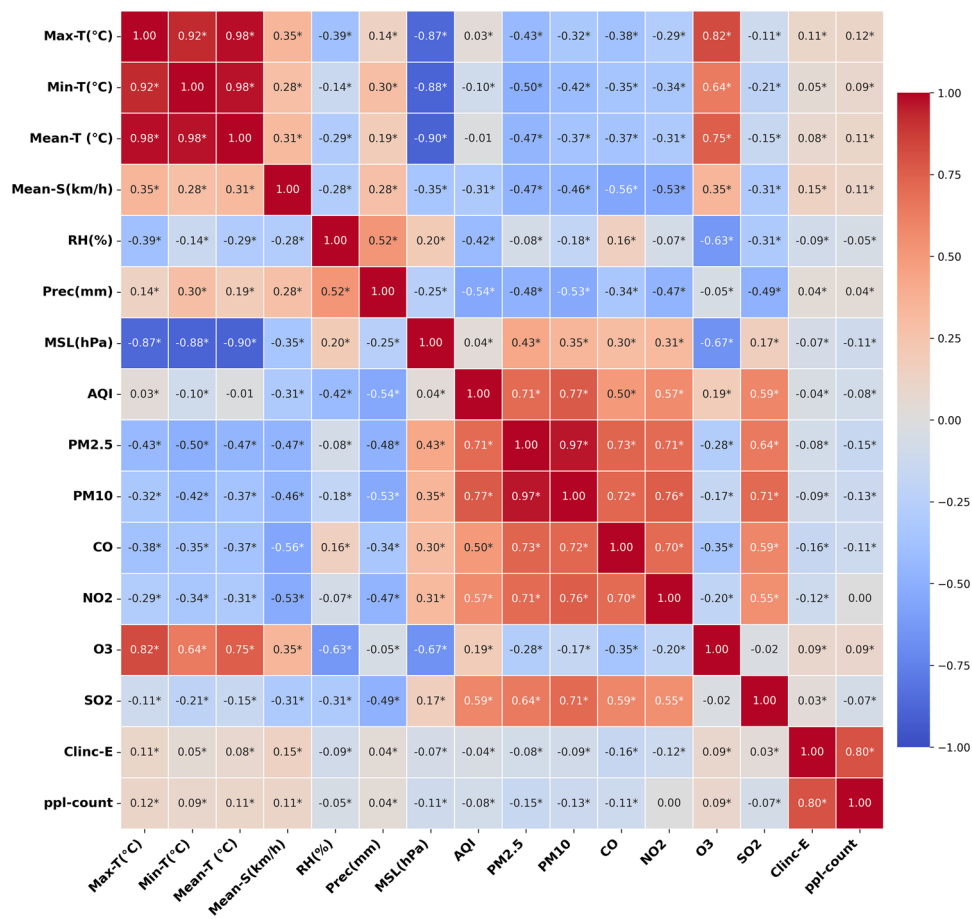
Fig. 1. The correlation heat map between outpatient volume and climatic conditions. Blue signifies negative correlations, while red indicates positive ones, with darker colors indicating stronger correlation. These '*' indicates statistical significance ($p < 0.05$).

groups. The seasonal variation analysis revealed that the second quarter (Q2) exhibits significant differences compared to both the first quarter (Q1) and the fourth quarter (Q4), while the third quarter (Q3) shows distinct variations from Q4. Additionally, the third quarter exhibited statistically significant differences from the fourth quarter ($p < 0.05$). Regarding week patterns, outpatient volumes showed significant differences across most day of week, except for Tuesday which did not differ significantly from Wednesday ($p = 0.23$).

XGboost model analysis

In the process of optimizing the XGBoost model by GS, we conducted an in-depth analysis and adjustment of its hyperparameters to achieve optimal performance. Specifically, we set the parameter colsample_bytree to 1, which means that all features will be used in each tree construction, ensuring the full utilization of feature information. The learning rate was set to 0.05. This relatively small learning rate can make the model converge more stably and reduce the risk of overfitting. By setting max_depth to 7, we limit the depth of each decision tree to prevent the model from being too complex and overfitting. At the same time, we set n_estimators to 300, indicating that 300 decision trees are used in the model, which can enhance the generalization ability and stability of the model. In addition, we set subsample to 0.7, which can effectively prevent the model from overfitting and improve the generalization ability of the model. The optimal parameter values obtained through GS are presented in Table 4. Through the above parameter adjustments, our XGBoost model achieved the best performance on the development set (Table 5).

RF model analysis

In the optimization process of the Random Forest model by GS, we conducted a detailed analysis and adjustment of its hyperparameters to achieve optimal performance. Specifically, we set the parameter max_depth to 7, which limits the depth of each decision tree. This helps prevent the model from being too complex and overfitting, while still maintaining its ability to capture the underlying patterns in the data. The max_features parameter was set to 5, meaning that each decision tree randomly selects 5 features for consideration when making splits. This introduces further randomness and diversity into the model, which can enhance its generalization ability. We also set min_samples_leaf to 1, indicating that each leaf node must contain at least one sample. This ensures that the trees are not too shallow and can make fine-grained distinctions. Additionally, min_samples_split was set to
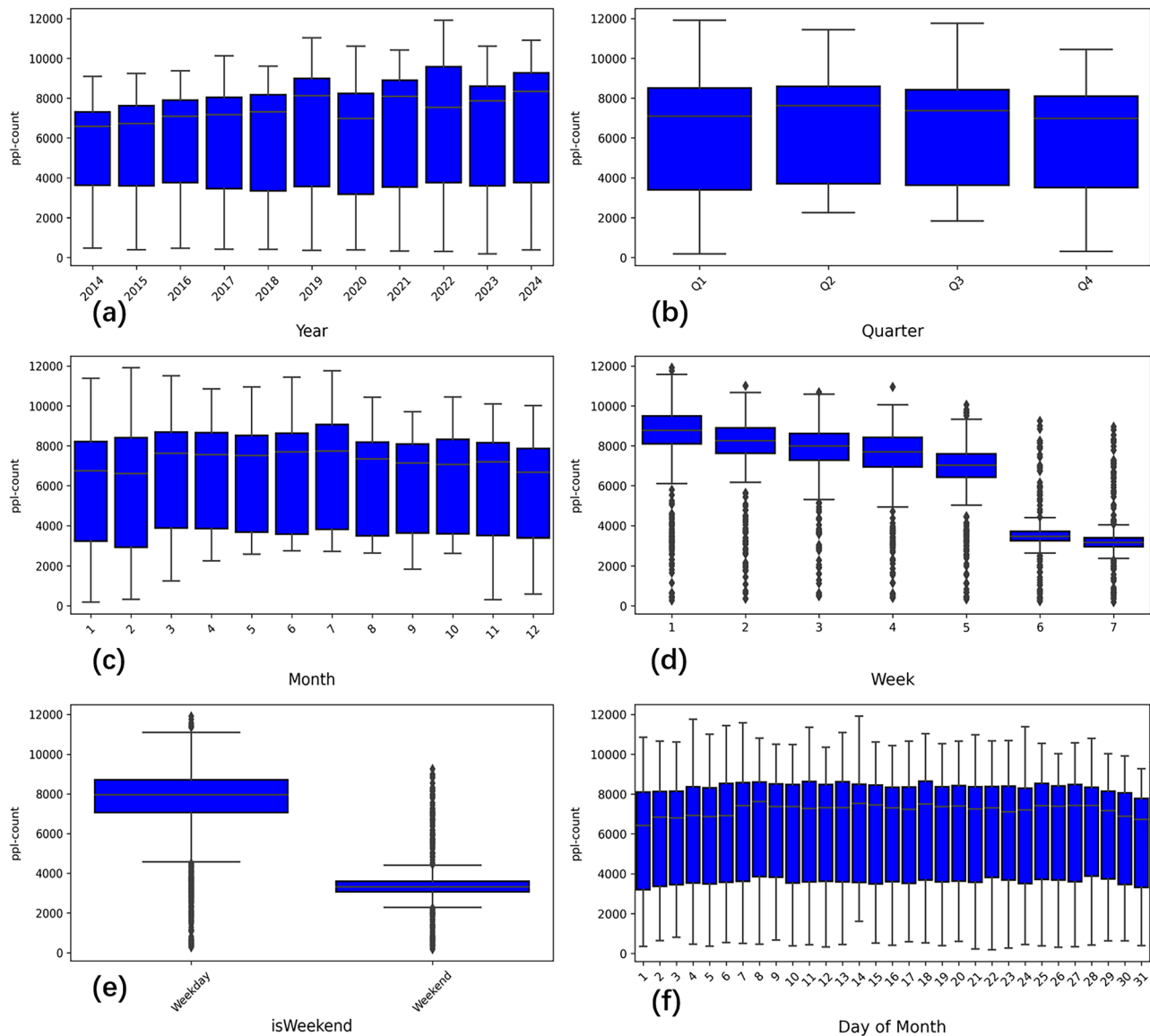
Fig. 2. Analysis of variance results for time categorical variables. (**a**), (**b**), (**c**), (**d**), (**e**), and (**f**) respectively represent the impact of Year, Quarter, Month, Weekday, Weekend, and Day on outpatient volume.

Year	2024	2023	2022	2021	2019	2018	2017
2014	0.0000	0.0000	0.0000	0.0000	0.0000	0.0197	0.0307
2015	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000
2016	0.0000	0.0007	0.0004	0.0000	0.0000	1.0000	1.0000
2017	0.0000	0.0023	0.0012	0.0000	0.0000	1.0000	1.0000
2018	0.0000	0.0038	0.0021	0.0000	0.0000	1.0000	1.0000
2020	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000

Table 2. Yearly differences of outpatient volumes identified by Dunn's post-hoc test.

4, meaning that a node must contain at least 4 samples before it can be split. This helps prevent the model from overfitting by ensuring that each split is based on a sufficient number of samples. Finally, we set n_estimators to 300, indicating that 300 decision trees are used in the model. This large number of trees further enhances the model's stability and generalization ability. Through the above parameter settings (Table 4), our RF model achieved optimal performance on the development set (Table 5).

Month	1	2	9	10	11	12
3	0.0001	0.0000	0.0132	0.1563	0.0202	0.0000
4	0.0007	0.0001	0.0716	0.6125	0.0999	0.0000
5	0.0126	0.0016	0.6357	1.0000	0.8007	0.0000
6	0.0007	0.0001	0.0686	0.592	0.096	0.0000
7	0.0000	0.0000	0.0001	0.0031	0.0002	0.0000

Table 3. Monthly Differences of outpatient volumesIdentified by Dunn's post-hoc test.

Models	Default parameters	Optimal parameters of GS
XGboost	colsample_bytree = 1 learning_rate = 0.3 max_depth = 6 n_estimators = 100 subsample = 1	colsample_bytree = 1 learning_rate = 0.05 max_depth = 7 n_estimators = 300 subsample = 0.7
RF	max_depth = "auto" max_features = "None" min_samples_leaf = 1 min_samples_split = 2 n_estimators = 100	max_depth = 7 max_features = 5 min_samples_leaf = 1 min_samples_split = 4 n_estimators = 300
SARIMAX	p = [0, 3] d = [0, 2] q = [0, 3] P = [0, 2] D = [0, 2] Q = [0, 2]	p = 2 d = 0 q = 2 P = 1 D = 0 Q = 1

Table 4. The best parameter value of the three models obtained by GS.

Model	Model error				Test error			
	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
XGBoost	324.41	496.35	8.52	0.96	578.90	852.51	12.19	0.90
RF	766.96	1066.74	16.38	0.82	996.77	1289.78	21.37	0.77
SARIMAX	432.49	664.24	10.53	0.93	905.39	1185.01	16.33	0.81

Table 5. The prediction performance of the three models.

SARIMAX model analysis

First, the Augmented Dickey-Fuller (ADF) test was conducted on the original time series, which yielded a test statistic of -6.74 ($p < 0.001$), significantly surpassing the critical values at all standard significance thresholds: -3.43 (1% level), -2.86 (5% level), and -2.57 (10% level). This indicates that the original sequence is stationary. Due to significant differences in scale among the variables, the data were standardized using Z-score normalization ($\mu = 0$, $\sigma = 1$)[26].

The seasonal periodicity (s = 7) was systematically determined based on the dataset's daily granularity and confirmed through spectral analysis, which revealed pronounced weekly cyclicity in outpatient visit patterns. Autocorrelation diagnostics further validated this selection, exhibiting significant peaks at lags (Fig. 3) corresponding to multiples of 7 (7, 14, 21 days), thereby reinforcing the weekly seasonality hypothesis. A comprehensive grid search across parametric configurations identified s = 7 as optimal, achieving a minimum Akaike Information Criterion (AIC) of 672.76 , outperforming alternatives with s = 14 (AIC = 906.74) or s = 21 (AIC = 1067.99). The final model configuration as SARIMAX (2, 0, 2) (1, 0, 1)$_7$ passed the Ljung-Box test ($p < 0.001$) and obtained the lowest MAE, RMSE, MAPE and R^2 (Table 4) in the both training and testing.

Comparing analysis

The comparative analysis of model and test errors between the observed values and predicted outputs from the three competing models are presented in Table 5, with detailed statistical metrics to quantify prediction accuracy. The XGBoost model exhibited the best performance, with the lowest MAE, RMSE, MAPE and the highest R^2 during both the modeling and testing stages.

The predictive performance of model

The development set exhibited a median of 7255 (MAD = 1607; IQR = 3559–8415). Corresponding model performance for XGBoost demonstrated robust predictive accuracy with MAE (4.47% of median), RMSE (3.84% of median). Validation on the external test set (M = 7257; MAD = 2403; IQR = 3559–8420) revealed similarly strong performance metrics with MAE (7.97% of median), RMSE (11.75% of median). These consistent error
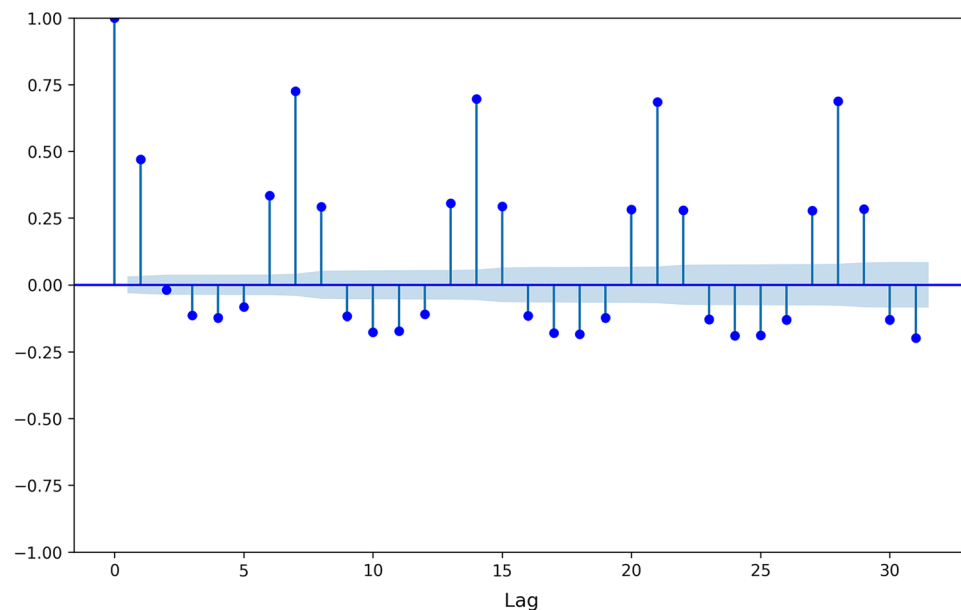
Fig. 3. Autocorrelation revealing seasonal patterns (Lags 0–31). Significant positive spikes at lags 7/14/21/28, exceeding the 95% confidence band (shaded light-blue).
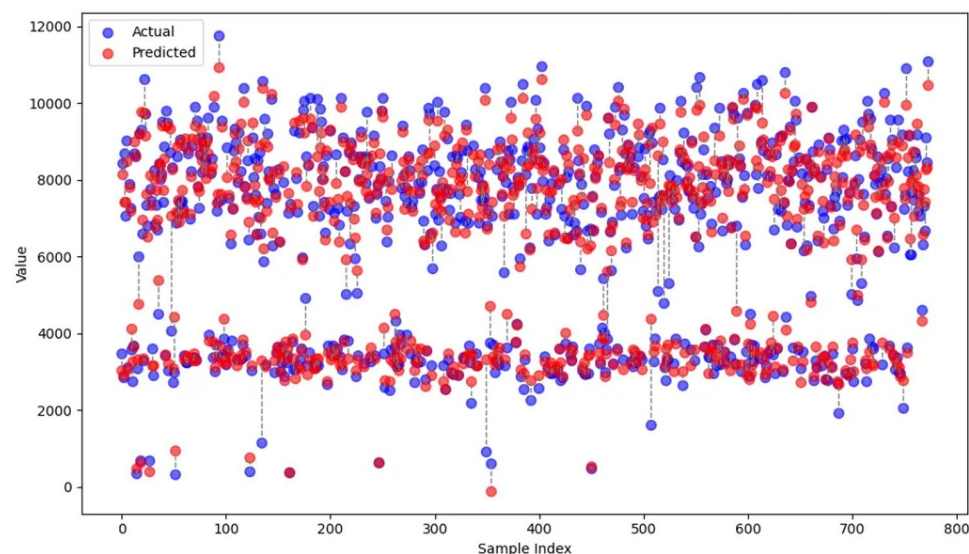

Fig. 4. Actual versus predicted values with error lines from the XGBoost development set. The red and blue dots in the graph represent the predicted and actual values, respectively, while the error bars indicate the magnitude of the difference between the them. The predicted values closely tracks observed values, confirming model reliability.

margins across development and external validation cohorts indicate substantial concordance between model predictions and observed outcomes, with relative error rates remaining below 13% for all measures. The predictive performance of the XGBoost model was rigorously validated through temporal dependency analysis and error profiling, with Fig. 4 illustrating the alignment between observed and predicted values for the development set, while Fig. 5 presents the corresponding comparison for the independent test set. The development set shows high alignment between actual and predicted outpatient volumes, with dense overlap indicating reliable predictions. For the test set, predicted values closely follow actual values, with small deviations highlighted by dashed lines, confirming robust generalization.

8
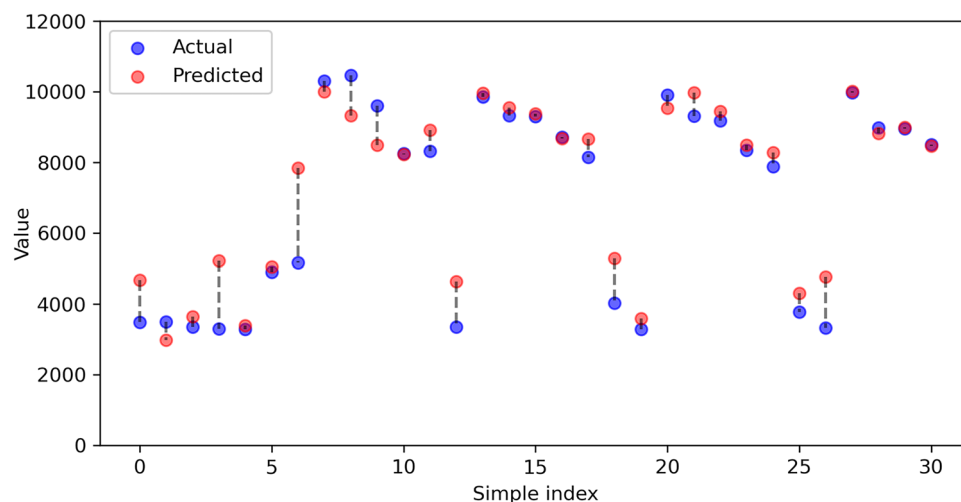
Fig. 5. Actual versus predicted values with error lines from the XGBoost test set. The predicted values are uniformly distributed around the actual values, with small errors at most time points. This indicates that the XGBoost prediction model has good fitting performance and accuracy.
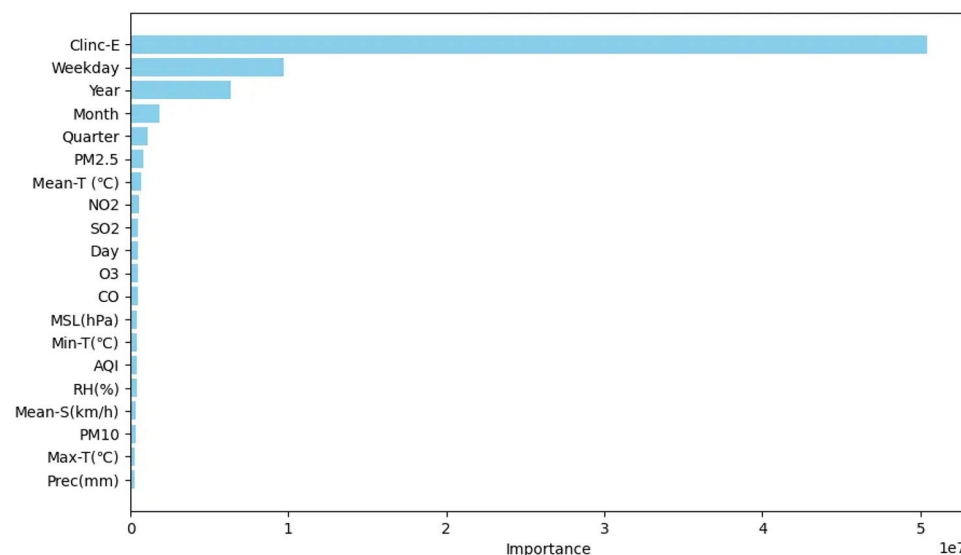

Fig. 6. Feature importance based on gain of XGBoost. The longer the blue horizontal bar, the greater the gain.

Analysis of the important influencing factors of outpatient volume

The analysis of the XGboost model's feature importance scores revealed that Clinic-E, Weekday, Year, Month, Quarter, PM2.5, and Mean-T had relatively large absolute values across the model. These findings suggest that these indicators may be key factors influencing outpatient volume (Fig. 6).

Figure 7 shows SHapley Additive exPlanations (SHAP) values for different features. Each row in the SHAP feature density scatter plot represents a specific feature, and the horizontal coordinate represents the corresponding SHAP value. The features are ordered based on the average absolute value of their SHAP values, reflecting their importance in the model. Wider areas in the plot indicate a higher concentration of samples sharing similar feature values. In Fig. 7, each dot represents an individual sample, and the color of the dot indicates the magnitude of the corresponding feature value. Redder colors indicate higher feature values, while bluer colors represent lower feature values. By examining the plot, we can deduce important insights. For instance, Clinic-E appears to be highly influential in the model, as indicated by its high importance ranking and the scarcity of individuals with high feature values. Moreover, the positive SHAP values associated with Clinic-E suggest a positive impact on the model's predictions. The scattered distribution of samples along the horizontal axis for Clinic-E further emphasizes its significant influence. Furthermore, for lower-ranking features like O3, the majority of data points are distributed around a SHAP value of 0, suggesting a limited impact on the majority of samples. However, it is worth noting that a discernible influence can be observed for a subset of individuals.
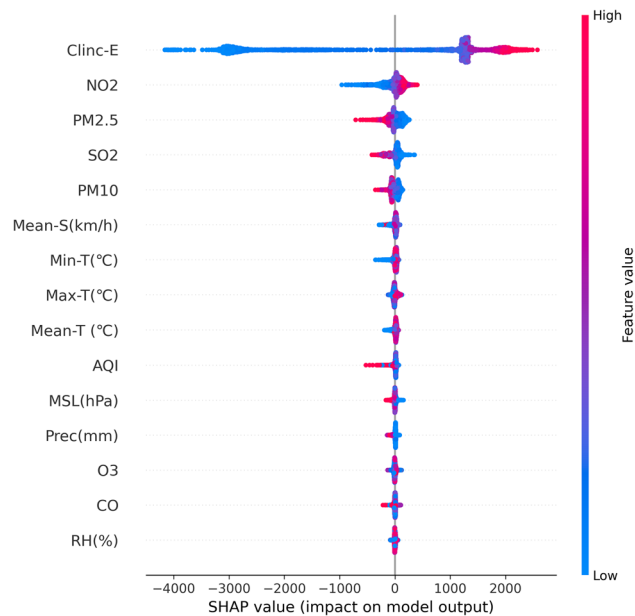
Fig. 7. SHAP values for different features of XGBoost. Each point represents a sample. The horizontal position indicates the SHAP value (contribution to the model output), where positive values (right) push predictions higher and negative values (left) push them lower. The vertical position reflects the feature importance ranking (higher ranking indicates greater overall impact). Color represents the actual feature value: redder hues denote higher values, bluer hues denote lower values.

Discussion

To the best of our knowledge, this study is the first to construct a predictive model for outpatient demand based on XGBoost, with the specific aim of guiding management strategies. This study successfully employed the XGBoost machine learning algorithm to develop a predictive model for hospital outpatient volume using meteorological data, air quality data, and the daily number of available experts.

AS shown Figs. 1, 6 and 7, natural environmental factors potentially influence patients' medical needs, which may be related to their impact on the onset and progression of diseases[27–30]. The degree of influence of these factors, as captured by the correlation analysis and the XGBoost model, is inconsistent. For instance, while NO2 demonstrated negligible statistical significance in correlation analysis, it exhibited high feature importance in SHAP analysis. This discrepancy arises from the non-linear dose–response relationship between NO2 exposure and outpatient volume, which SW correlation methods fail to adequately capture. Specifically, for the XGBoost model, the contribution of Mean-T is greater than that of Max-T. This inconsistency is also evident in the Mann–Whitney tests results. The indicator for "weekend" was removed because its relationship with "weekday" was already captured by the XGBoost model. With advancements in medical care and the influence of holidays such as the Spring Festival, May Day, and National Day, time-related factors significantly contribute to the prediction model[17].

XGBoost outperformed SARIMAX and RF due to its non-linear modeling and feature interaction handling, capturing complex meteorological-health thresholds (e.g., weekend expertise shifts) and non-Gaussian errors. Unlike SARIMAX's linear assumptions, XGBoost dynamically weights features via gradient boosting, while RF's greedysplitting (e.g., Gini impurity) amplified noise from weak features (e.g., humidity). XGBoost's L1 regularization and subsampling (subsample = 0.7) reduced overfitting versus RF's rigid hyperparameters (e.g., max_features = None), enhancing generalizability.

The predictive model uncovered several critical insights. The ability of XGboost to handle high-dimensional data and feature interactions was fully demonstrated in this application. The number of available experts was a key determinant, reflecting the crucial role of resource availability in meeting patient demand. Meanwhile, the positively correlated variable, Mean-T, and the negatively correlated variable, PM2.5, should be given special attention. Additionally, the Day of Week variable significantly contributed to outpatient visit volume, indicating that the allocation of medical resources on weekdays influences patients' choices regarding medical treatment[31–34].

These findings have significant implications for hospital resource planning and operational efficiency. On the one hand, by forecasting outpatient volumes with high accuracy, hospitals can proactively manage staffing levels, ensuring that an adequate number of experts are available to meet demand during high-volume periods. On the other hand, anticipating fluctuations in outpatient visits allows for better scheduling, reducing patient waiting times and enhancing the overall patient experience[35–37]. Meanwhile, hospitals in regions prone to extreme weather events or seasonal pollution can use these predictions to prepare for potential surges in patient admissions. For instance, during heatwaves or periods of poor air quality, hospitals can stockpile necessary medications, increase the availability of respiratory specialists, and ensure adequate cooling or air filtration systems are in

place. Furthermore, the findings can guide long-term resource allocation and infrastructure investments. For instance, during specific seasons or environmental conditions, manager might consider expanding facilities, adding specialized units, or investing in online medical consultation.

Limitations

While the model demonstrated strong predictive accuracy, several limitations should be addressed in future research. First, the model relies on historical meteorological data and specialist availability, which may not account for unexpected events, such as pandemics or natural disasters, that could significantly alter outpatient volumes. Second, other potential predictors, such as emerging health technologies, socioeconomic factors, and patient demographics, were not included in the current model but could enhance its comprehensiveness. Third, validating the model across multiple hospitals and geographic regions would help improve its generalizability and applicability. Fourth, the reliance on specialist availability data may introduce institutional bias. Due to larger hospitals typically employ more specialists and maintain higher staffing ratios. This could lead to training data from well-resourced hospitals may overrepresent complex cases handled by specialists, reducing model applicability to smaller clinics. Future studies should stratify analyses by hospital tier and include small medical institutions to validate generalizability. Fifth, excluding data from the COVID-19 period (2020–2022) may allow the model to underestimate the dynamic association of environmental factors between normal and abnormal conditions. In our follow-up work, we can combine pandemic data with anomaly detection to enhance robustness against such disruptions. Additionally, advancements in data mining technology have led to the development of models with high predictive capabilities in various fields[38,39]. It would be valuable to explore whether these models could outperform XGBoost. In the future, we would use deep learning algorithms (e.g., Transformer, LSTM) to provide potential solutions to address the limitations of traditional machine learning models in capturing complex time-dependencies and non-linear interactions[40,41].

Conclusion

This study highlights the value of integrating machine learning techniques with environmental and operational data to forecast hospital outpatient volumes. By harnessing the predictive power of XGBoost, hospitals can improve their preparedness for demand fluctuations, optimize resource allocation, and ultimately enhance patient care outcomes. The findings lay the foundation for more advanced, data-driven approaches in healthcare management.

Data availability

All data analyzed and predicted during this study are available in the Additional information.

References

1. Murtas, R., Tunesi, S., Andreano, A. & Russo, A. G. Time-series cohort study to forecast emergency department visits in the city of Milan and predict high demand: A 2-day warning system. *BMJ Open* **12**(4), e056017 (2022).
2. Peng, J. et al. Peak outpatient and emergency department visit forecasting for patients with chronic respiratory diseases using machine learning methods: Retrospective cohort study. *JMIR Med. Inform.* **8**(3), e13075 (2020).
3. Soyiri, I. N. & Reidpath, D. D. An overview of health forecasting. *Environ. Health Prev. Med.* **18**(1), 1–9 (2013).
4. Luo, L., Luo, L., Zhang, X. & He, X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Serv. Res.* **17**(1), 469 (2017).
5. Feng, H. L., Jia, Y. W., Zhou, S. Y. & Chen, H. Y. Adaptive decision support system for outpatient appointment scheduling with heterogeneous service times. *Sci. Rep.* **14**, 1–8 (2024).
6. Hadavandi, E. et al. Developing a hybrid artificial intelligence model for outpatient visits forecasting in hospitals. *Appl. Soft Comput.* **12**(2), 700–711 (2012).
7. Maninchedda, M. et al. Main features and control strategies to reduce overcrowding in emergency departments: A systematic review of the literature. *Risk Manag. Healthc. Policy* **16**, 255–266 (2023).
8. Wong, H. T., Chiu, M. Y., Wu, C. S. & Lee, T. C. Senior Citizen Home Safety Association. The influence of weather on health-related help-seeking behavior of senior citizens in Hong Kong. *Int. J. Biometeorol.* **59**(3), 373–376 (2015).
9. Lu, X. & Qiu, H. Explainable prediction of daily hospitalizations for cerebrovascular disease using stacked ensemble learning. *BMC Med. Inform. Decis. Mak.* **23**(1), 59 (2023).
10. Giorgini, P. et al. Particulate matter air pollution and ambient temperature: Opposing effects on blood pressure in high-risk cardiac patients. *J. Hypertens.* **33**(10), 2032–2038 (2015).
11. Weichenthal, S. et al. Daily summer temperatures and hospitalization for acute cardiovascular events: Impact of outdoor PM 2.5 oxidative potential on observed associations across Canada. *Epidemiology* **34**(6), 897–905 (2023).
12. Tao, J. et al. Daytime and nighttime high temperatures differentially increased the risk of cardiovascular disease: A nationwide hospital-based study in China. *Environ. Res.* **236**(Pt 1), 116740 (2023).
13. Lu, P. et al. Attributable risks associated with hospital outpatient visits for mental disorders due to air pollution: A multi-city study in China. *Environ. Int.* **143**, 105906 (2020).
14. Chen, M. J. et al. Machine learning to relate PM2.5 and PM10 concentrations to outpatient visits for upper respiratory tract infections in Taiwan: A nationwide analysis. *World J. Clin. Cases* **6**(8), 200–206 (2018).
15. Xie, L., Li, A. Q. & Li, L. Impact of air pollution and meteorological factors on outpatient visits for dermatitis: A time-series study. *Sichuan Da Xue Xue Bao Yi Xue Ban* **50**(6), 884–890 (2018).
16. Klute, B., Homb, A., Chen, W. & Stelpflug, A. Predicting outpatient appointment demand using machine learning and traditional methods. *J. Med. Syst.* **43**(9), 288 (2019).
17. Zhou, L., Zhao, P., Wu, D., Cheng, C. & Huang, H. Time series model for forecasting the number of new admission inpatients. *BMC Med. Inform. Decis. Mak.* **18**(1), 39 (2018).
18. Yang, C. C. Explainable artificial intelligence for predictive modeling in healthcare. *J. Healthc. Inform. Res.* **6**(2), 228–239 (2022).
19. Deng, A. et al. Developing computational model to predict protein-protein interaction sites based on the XGBoost algorithm. *Int. J. Mol. Sci.* **21**(7), 2274 (2020).

20. Clift, A. K. et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: Cohort study. *BMJ* **381**, e073800 (2023).
21. Li, J. L. et al. Predicting mortality in intensive care unit patients with heartfailure using an interpretable machine learning model: Retrospective cohort study. *J. Med. Internet Res.* **24**(8), e38082 (2022).
22. Xiao, H. et al. The impact of the COVID-19 pandemic on health services utilization in China: Time-series analyses for 2016–2020. *Lancet Reg. Health West Pac.* **9**, 100122 (2021).
23. Nornadiah, M. R. & Bee, Y. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analyt.* **2**(1), 21–33 (2021).
24. Schober, P., Boer, C. & Schwarte, L. A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **126**(5), 1763–1768 (2018).
25. Gu, L. S. et al. Chitosan-based extrafibrillar demineralization for dentin bonding. *J. Dent. Res.* **98**(2), 186–193 (2019).
26. Brownlee, J. Data preparation for machine learning: Data cleaning, feature selection, and data transforms in Python. *Mach. Learn. Mastery* **17**, 217–219 (2022).
27. Wang, S. et al. The impact of outdoor air pollutants on outpatient visits for respiratory diseases during 2012–2016 in Jinan, China. *Respir. Res.* **19**(1), 246 (2018).
28. Chiang, K. L., Lee, J. Y., Chang, Y. M., Kuo, F. C. & Huang, C. Y. The effect of weather, air pollution and seasonality on the number of patient visits for epileptic seizures: A population-based time-series study. *Epilepsy Behav.* **115**, 107487 (2021).
29. Hwang, H., Jang, J. H., Lee, E., Park, H. S. & Lee, J. Y. Prediction of the number of asthma patients using environmental factors based on deep learning algorithms. *Respir. Res.* **24**(1), 302 (2023).
30. Ravindra, K. et al. Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. *Sci. Total Environ.* **858**(Pt 1), 159509 (2023).
31. Chen, Y. F. et al. The magnitude and mechanisms of the weekend effect in hospital admissions: A protocol for a mixed methods review incorporating a systematic review and framework synthesis. *Syst. Rev.* **5**, 84 (2016).
32. Churchill, A. J., Gibbon, C., Anand, S. & McKibbin, M. Public opinion on weekend and evening outpatient clinics. *Br. J. Ophthalmol.* **87**(3), 257–258 (2003).
33. Feeney, C. L., Roberts, N. J. & Partridge, M. R. Do medical outpatients want 'out of hours' clinics?. *BMC Health Serv. Res.* **5**, 47 (2005).
34. Barbieri, J. S., Chu, B. & Mostaghimi, A. Sociodemographic differences associated with utilization of weekend versus week primary care visits. *J. Gen. Intern. Med.* **36**(7), 2180–2181 (2021).
35. Liu, X., Gu, F., Bai, Z., Huang, Q. & Ma, G. Forecasting of daily outpatient visits based on genetic programming. *Iran. J. Public Health* **51**(6), 1313–1322 (2022).
36. Huang, D. & Wu, Z. Forecasting outpatient visits using empirical mode decomposition coupled with back-propagation artificial neural networks optimized by particle swarm optimization. *PLoS One* **12**(2), e0172539 (2017).
37. Qiu, H. et al. Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC Med. Inform. Decis. Mak.* **20**(1), 83 (2020).
38. Ay, S., Cardei, M., Meyer, A. M., Zhang, W. & Topaloglu, U. Improving equity in deep learning medical applications with the Gerchberg-Saxton algorithm. *J. Healthc. Inform. Res.* **8**(2), 225–243 (2024).
39. Ramachandran, R. A., Koseoglu, M., Özdemir, H., Bayindir, F. & Sukotjo, C. Machine learning model to predict the width of maxillary central incisor from anthropological measurements. *J. Prosthodont. Res.* **68**(3), 432–440 (2024).
40. Zimmerman, L. et al. Context-dependent design of induced-fit enzymes using deep learning generates well-expressed, thermally stable and active enzymes. *Proc. Natl. Acad. Sci. USA* **121**(11), e2313809121 (2024).
41. Tang, H. et al. Development and validation of a deep learning model to predict the survival of patients in ICU. *J. Am. Med. Inform. Assoc.* **29**(9), 1567–1576 (2022).

Acknowledgements

Author contributions

HH designed the study. LLZ and ZQ wrote and edited the manuscript. CQ and WP contributed to the data collection and analysis. LLZ and ZQ were involved in the interpretation of the results. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

This study was approved by the Daping Hospital of Army Medical University. Informed consent was waived because this research did not involve individual data. All methods were performed in accordance with the relevant guidelines and regulations.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01265-y.

Correspondence and requests for materials should be addressed to H.H.

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

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