Proceedings of the 11th North American International Conference on Industrial Engineering and Operations Management, Milwaukee, Wisconsin, USA, June 9-11, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260363
Published: June 9, 2026

XGBoost for Forecasting Daily Pediatric Emergency Department Visits

Ahmed Gailan Qasem
School of Health Profession
University of Alabama at Birmingham (UAB)
Birmingham, AL, USA
aqasem@uab.edu

Lori Pelletier
Connecticut Children's Medical Center
Hartford, CT, USA
LPelletier01@connecticutchildrens.org

Bunyamin Ozaydin and Abdulaziz Ahmed
School of Health Profession
University of Alabama at Birmingham (UAB)
Birmingham, AL, USA
bozaydin@uab.edu, aahmed2@uab.edu

Abstract

Accurate forecasting of daily emergency department (ED) visit volumes is essential for staffing decisions, resource allocation, and patient flow management. While machine learning has demonstrated promise in forecasting ED demand for adult populations, its use in dedicated pediatric ED settings remains underexplored, particularly with respect to incorporating temporal and weather-related predictors. This study develops an XGBoost-based framework to forecast daily pediatric ED visit volumes one week in advance using ED visits data from a pediatric hospital between October 2021 and September 2025. The model incorporates lagged visit counts, rolling-window statistics, cyclical calendar encodings, and weather-related variables to capture recurring ED utilization patterns and environmental influences on ED visit volumes. The performance of the XGBoost model is benchmarked against three statistical time-series models: autoregressive (AR), autoregressive integrated moving average (ARIMA), and seasonal ARIMA with exogenous regressors (SARIMAX). On a one-year temporal holdout, XGBoost achieves a mean absolute error (MAE) of 13.54 and a mean absolute percentage error (MAPE) of 8.47%, reducing test MAE by approximately 21% and test MAPE by approximately 37% relative to the best statistical benchmark, SARIMAX (MAE = 17.18, MAPE = 13.50%). XGBoost also exhibits the smallest train–test gap among the evaluated models, indicating strong generalization across a full year of seasonal variation. Feature importance analysis identifies recent demand history, day-of-week cyclicality, and temperature variables as the most influential predictors. These findings demonstrate that XGBoost augmented with temporal and weather-related features meaningfully improves daily pediatric ED forecasting and provides a practical tool for short-term ED operational planning.

Keywords
ED forecasting, XGBoost, SARIMAX, Pediatric

1. Introduction

Emergency departments (EDs) are a critical component of healthcare systems, serving as frontline settings for patients with urgent medical needs. In the United States, EDs recorded approximately 155 million visits in 2022, with pediatric patients representing about 20% of total ED utilization (National Center for Health Statistics, 2024). Managing this demand remains a major operational challenge as fluctuations and surges in visit volumes can strain staffing, treatment space, and clinical resources, ultimately affecting patient flow and timely access to care (Boyle et al., 2012). Accurate short-term forecasting of ED visits is therefore essential for resource allocation, patient flow management, and surge preparedness (Asheim et al., 2019; Gafni-Pappas & Khan, 2023).

Forecasting demand in pediatric EDs presents additional challenges because children's emergency care needs are shaped by age-specific, temporal, seasonal, and environmental patterns (Ozturk & Merter, 2024). Visit volumes often vary with respiratory illness seasons, school schedules, holidays, injury patterns, weather conditions, and public health interventions such as vaccination campaigns. These factors can lead to recurring weekly and seasonal trends, as well as abrupt short-term changes in demand that may differ from adult ED utilization (Alpern et al., 2006; Pakalnis & Heyer, 2016; Pines et al., 2021). Many existing ED forecasting studies have focused on overall ED attendance or adult populations, which may limit their ability to capture the unique drivers of pediatric ED utilization (Etu et al., 2025).

Existing ED forecasting studies have used a range of modeling approaches to predict fluctuations in patient volume and support resource allocation, with time-series methods among the most widely applied (Rocha & Rodrigues, 2021; Tuominen et al., 2022; Zhao et al., 2022). Traditional approaches, such as autoregressive (AR), autoregressive integrated moving average (ARIMA), and exponential smoothing, have demonstrated effectiveness for short-term forecasting of general ED arrivals. These methods are useful for capturing trends, autocorrelation, and recurring seasonal patterns in patient demand. More advanced extensions, such as seasonal autoregressive integrated moving average with exogenous regressors (SARIMAX), allow external predictors to be incorporated when modeling ED arrivals. However, while these methods are interpretable and effective for regular temporal patterns, they may be less suited to capturing nonlinear relationships, complex interactions among predictors, and abrupt changes in demand (Etu et al., 2025).

More recently, machine learning methods have gained increasing attention for ED demand forecasting. Approaches such as random forests, gradient-boosting models, and neural networks can model nonlinear patterns and complex interactions in historical visit data and have shown improved predictive performance in some settings (Etu et al., 2022; Vural et al., 2025; Zhang et al., 2022). However, evidence from pediatric emergency settings remains limited and variable. Although these methods offer greater flexibility than traditional statistical models, their performance in pediatric settings has not been consistently superior, and no single method has emerged as the best-performing approach across settings, populations, and forecasting horizons (Etu et al., 2025). Further research is therefore needed to evaluate whether the predictive benefits of machine learning generalize to pediatric ED demand forecasting, especially in settings where demand patterns are less strongly seasonal.

These limitations underscore the need for forecasting approaches tailored to pediatric ED settings. Such approaches should capture age-related demand patterns, seasonal variation, and external factors that influence pediatric arrivals, while remaining robust to irregular surges in demand. This study develops an XGBoost-based model to forecast daily pediatric ED visit volumes using four years of operational data from a pediatric hospital. The model incorporates lagged visit counts, rolling-window statistics, cyclical calendar encodings, weekend indicators, and weather-related variables. Its performance is benchmarked against conventional statistical time-series models, including ARIMA and SARIMAX. The objective is to evaluate whether a feature-rich machine learning approach can improve short-term pediatric ED demand forecasting compared with traditional statistical methods.

2. Related Work

Forecasting ED visit volume has been widely studied in healthcare operations and medical informatics. Early studies primarily relied on statistical time-series and regression-based methods, such as ARIMA and SARIMA, which remain effective for modeling temporal dependence and seasonal patterns in patient arrivals. For example, Afilal et al. (2016) used time-series models to forecast daily ED attendances, while Boyle et al. (2012) applied ARIMA to predict ED arrivals using patient demographic and clinical characteristics. Other studies extended statistical models by incorporating external predictors. Kandula et al. (2019) combined ARIMA and random forest models with Google Search trends. Duarte et al. (2021) found that ARIMA outperformed several machine learning approaches for ED attendance prediction during the COVID-19 period. Although these methods remain valuable benchmarks, their performance may be constrained when ED demand is influenced by nonlinear relationships, sudden shifts, or multiple interacting predictors.

More recent work has examined machine learning and deep learning models for ED demand forecasting. These approaches can learn complex temporal patterns and nonlinear associations from large sets of predictors. Zhang et al. (2022) used LSTM and ensemble models to forecast ED arrivals at hourly and daily intervals, while Vollmer et al. (2021) developed a framework that compared statistical and machine learning models across multiple forecasting horizons. Their findings showed that machine learning can improve forecasting accuracy in some settings, while also demonstrating that model performance depends on the dataset, prediction horizon, feature set, and operational context. Other studies have incorporated internet search behavior and epidemiological indicators to improve responsiveness to infectious disease-related demand, including google trends and RSV-related forecasting applications (Amorim et al., 2021; Etu et al., 2025; Fan et al., 2022). These studies show the potential value of exogenous predictors but also highlight the need for validating within specific clinical and operational settings.

Compared with the broader ED forecasting, pediatric ED forecasting remains relatively limited. Researchers have used both traditional time-series models and, more rarely, machine learning or deep learning to analyze the pediatric time-series data. Morzadec et al. (2021) modeled pediatric ED demand in the Paris University Hospital Trust and showed that incorporating academic schedules improved forecasts across 7-, 14-, and 28-day horizons. Guan and Engelhardt, (2019) developed recurrent neural network models for pediatric outpatient sick-visit volume and demonstrated that pediatric demand contains strong weekly and seasonal patterns. Although these studies demonstrate that advanced modeling approaches improved pediatric demand forecasting, pediatric forecasting studies remain relatively few, especially in dedicated pediatric ED settings.

This study develops a machine learning framework for forecasting daily pediatric ED visit volume. In contrast to prior studies that focus largely on general ED populations and rely primarily on statistical seasonality terms, this study uses a feature-rich XGBoost model that integrates lagged visit counts, rolling-window statistics, cyclical calendar encodings, and weather-related variables. The model is benchmarked against conventional statistical time-series approaches using four years of operational data from a pediatric hospital. By focusing on a dedicated pediatric ED setting, this study addresses the need for forecasting models that can capture temporal patterns, exogenous influences, and nonlinear demand dynamics observed in real-world pediatric ED operations.

3. Methods

The forecasting framework is developed to predict daily pediatric ED visit volumes using historical utilization patterns, calendar-based features, and weather-related predictors. Figure 1 presents the overall ED forecasting workflow. Pediatric ED visit records and weather data are first collected, aggregated, and linked by date to create a daily modeling dataset. The prepared dataset is then transformed through feature engineering and preprocessing, including the creation of lagged visit features, rolling-window demand measures, temporal and calendar-based variables, and normalized or encoded predictors. These predictors are used to train and tune the XGBoost model and conventional time-series benchmark models. Model performance is evaluated using mean absolute error (MAE) and mean absolute percentage error (MAPE), which measure forecast accuracy in terms of both absolute visit counts and relative percentage error.

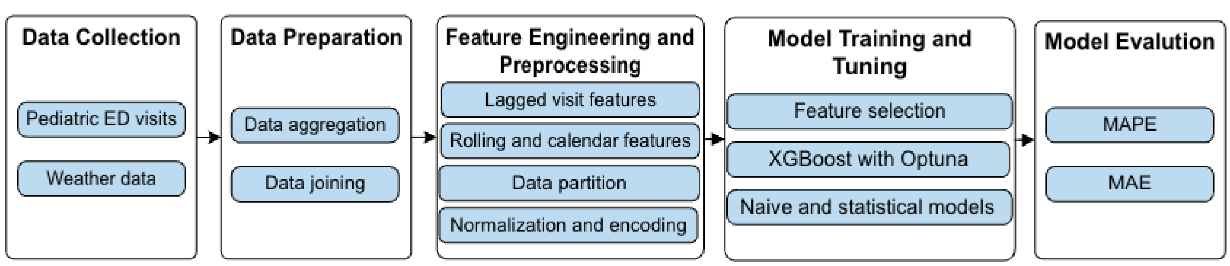

Figure 1. Overview of the ED forecasting workflow

3.1 Data Collection and Preparation

This study uses ED visit data collected from a pediatric hospital. Patient arrivals are aggregated by date to generate daily visit counts. The study period spans October 1, 2021, through September 30, 2025, yielding 1,461 daily observations and 241,839 total visits. This period is selected to reflect utilization patterns of pediatric ED after COVID-19. Weather data are obtained from the National Weather Service (NWS) and joined to the ED visits by date. The weather variables include maximum temperature, minimum temperature, average dew point, average humidity, snowfall, and average wind speed.

Figure 2 presents the daily ED visit volume over the study period. The series shows substantial day-to-day variation, with most daily volumes ranging approximately from 100 to 250 visits, excluding extreme low or high counts. Recurring seasonal patterns are also visible, with periods of higher demand occurring around winter seasons and lower-volume periods appearing during summer. In addition, intermittent surges can be observed throughout the series, reflecting short-term increases in demand that depart from the surrounding trend. These patterns highlight the need for forecasting models that can account for recent utilization trends, calendar-based variation, seasonal changes, and external factors that may contribute to short-term changes in ED volume.

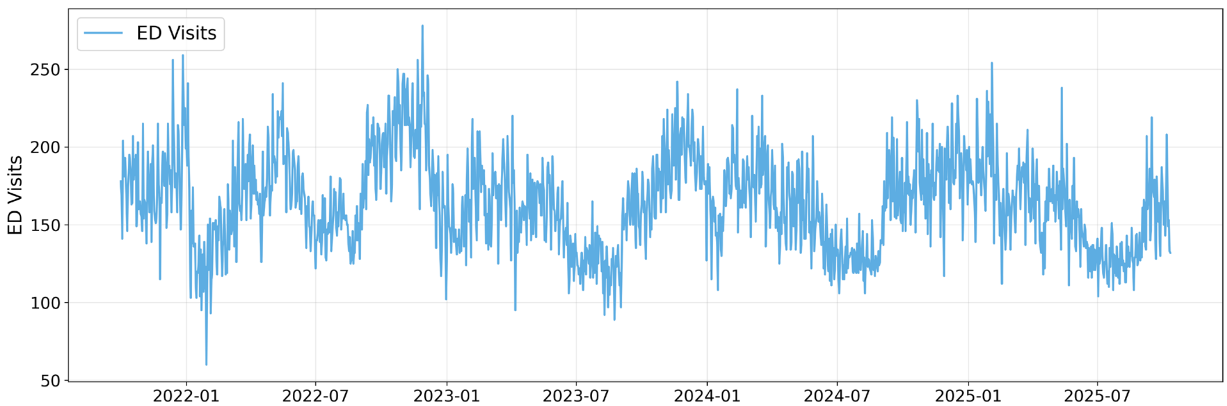

Figure 2. Daily pediatric ED visit volume from October 2021 through September 2025.

3.2 Feature Engineering

Feature engineering is conducted to capture multiple sources of variation in pediatric ED demand, including historical utilization, recent demand trends, and calendar-based patterns. Historical utilization is represented using lagged ED visit counts. Lag features are created for 1, 2, 3, 7, 14, and 30 days before each prediction date to capture short-term persistence, weekly recurrence, biweekly patterns, and longer recent demand history.

Rolling-window features are generated to summarize recent demand behavior over multiple time scales. Rolling means and rolling standard deviations are calculated over 7-, 14-, and 30-day windows. The rolling means represent recent demand levels, while the rolling standard deviations capture short-term variability in visit volume. To prevent target leakage, all lagged and rolling-window features are shifted so that only information available before the forecast date is used.

Calendar-based features are created to represent recurring temporal patterns in ED utilization. A binary weekend indicator is included, and calendar year is retained to account for broader changes across the study period. Cyclical transformations are applied to month, day of month, day of week, day of year, week of year, and quarter using sine and cosine functions, as shown in Equation (1). This encoding preserves the circular structure of calendar variables, allowing adjacent periods, such as Sunday and Monday or December and January, to be represented as close in time despite their numerical separation.

$$x_{sin} = \sin(\frac{2\pi.t}{P})\ ,\ x_{cos} = \cos(\frac{2\pi.t}{P}) \qquad (1)$$

where t represents the temporal variable and P represents its period.

Continuous numerical features are standardized using parameters estimated from the training data and then applied to the test data to avoid information leakage. Table 1 summarizes the engineered and preprocessed predictors included in the forecasting dataset.

Table 1 Summary of the exogenous variables

Feature category	**Features included**
Historical lag features	ED visits lagged by 1, 2, 3, 7, 14, and 30 days
Rolling demand features	7-, 14-, and 30-day rolling means of ED visits
Rolling variability features	7-, 14-, and 30-day rolling standard deviations of ED visits
Cyclical calendar features	Sine and cosine transformations of month of year, day of month, day of week, day of year, week of year, and quarter
Weekend indicator	Binary weekend variable
Weather-related features	Maximum temperature, minimum temperature, average dew point, average humidity, snowfall, and average wind speed

3.3 Forecasting Models

3.3.1 Statistical Benchmark Models

Three statistical benchmarks, including AR, ARIMA, and SARIMAX, are developed to capture different patterns in time-series data. The AR(p) model captures dependence on the p most recent observations. ARIMA(p,d,q) extends AR by adding d orders of differencing to address non-stationarity in time series and q moving-average terms to account for recent forecast errors. SARIMAX $(p,d,q)(P,D,Q)_s$ further introduces seasonal autoregressive, differencing, and moving-average components of order (P,D,Q) over a seasonal period s, while also allowing exogenous predictors to be included in the model.

Model orders are tuned by evaluating candidate specifications and selecting the model configurations with the best forecasting performance. The selected AR model is AR(7), which uses the previous seven daily observations to capture short-term and weekly dependence. The best performing ARIMA model is ARIMA$(7,0,7)$, which incorporates seven autoregressive terms, no differencing, and seven moving-average terms. The best performing SARIMAX configuration is SARIMA$(7,0,7)(1,0,1)_7$ with exogenous predictors. The nonseasonal component $(7,0,7)$ represents seven autoregressive terms, no differencing, and seven moving-average terms, while the seasonal component $(1,0,1)_7$ represents one seasonal autoregressive term, no seasonal differencing, and one seasonal moving-average term over a seven-day seasonal period. The weekly seasonal period reflects the recurring seven-day structure commonly observed in ED visit volumes. The SARIMAX model incorporates the same exogenous variables used in the machine learning model, including lagged visit counts, rolling-window statistics, calendar-based variables, the weekend indicator, and weather-related predictors.

3.3.2 XGBoost Model

XGBoost is a gradient-boosting algorithm that builds an ensemble of decision trees sequentially, where each new tree is trained to reduce the residual errors from the existing ensemble (Chen & Guestrin, 2016). For each observation i, the final prediction is produced by summing the outputs of multiple regression trees, as shown in Equation 2:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \tag{2}$$

Where x_i represents the predictor vector, f_k is the k-th regression tree, and K is the total number of trees in the ensemble. During training, XGBoost minimizes a regularized objective function that balances prediction error and model complexity, as shown in Equation 3:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

Where $l(y_i, \hat{y}_i)$ represents the loss function and $\Omega(f_k)$ denotes the regularization term that penalizes overly complex trees. This objective function allows the model to capture complex patterns in the data while reducing the risk of overfitting through shrinkage and tree-level regularization (Huang et al., 2021). This is particularly important for forecasting daily ED visit volumes, which are often noisy and influenced by temporal variability.

In this study, XGBoost is trained to forecast ED visit volume using lagged demand, rolling statistics, calendar variables, and weather-related features. The XGBoost model can capture nonlinear relationships and complex interactions among predictors, which may not be fully captured by traditional statistical time-series models.

3.3.2.1 Feature Selection

Recursive feature elimination (RFE) is applied to conduct feature selection prior to final tuning of XGBoost to remove uninformative predictors and reduce the risk of model overfitting and complexity. RFE starts with the full set of candidate predictors, fits the model, ranks predictors by their contribution, and iteratively removes the least informative variables until a best-performing subset is identified. The selected features are then used in the final XGBoost modeling pipeline. Figure 3 presents the retained features and their relative importance in the final model. Recent demand history dominates the predictor set, with the 7-day rolling mean and the 1-day and 7-day lagged visit counts ranking highest, followed by day-of-week and day-of-year cyclical encodings. Maximum temperature and wind speed contribute smaller but non-negligible predictive value, indicating that weather added information beyond what is captured by historical visit patterns.

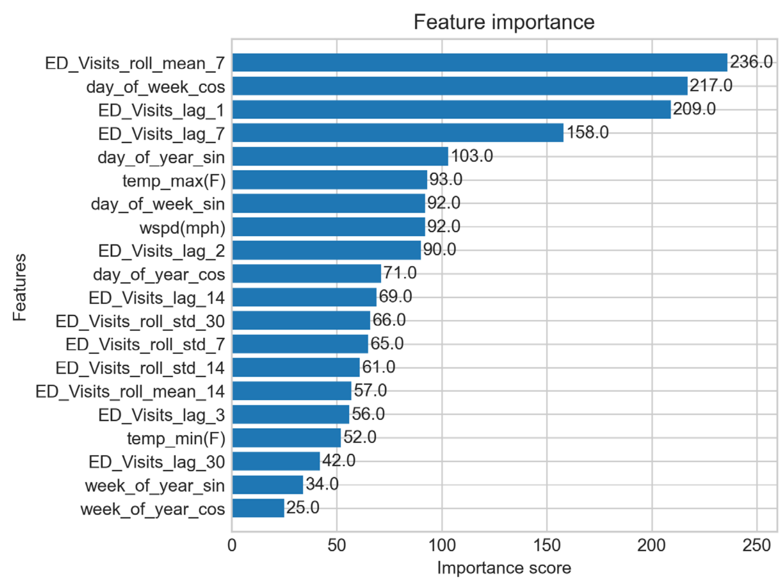

Figure 3. Feature importance of selected predictors for the final XGBoost model

3.3.2.2 Hyperparameter Tuning

Bayesian optimization with Optuna (Akiba et al., 2019) is utilized to tune XGBoost hyperparameters. In contrast to exhaustive search approaches, Bayesian optimization uses information from previous trials to guide the search toward more promising regions of the hyperparameter space. Table 2 presents the search ranges and selected values that minimize the forecast error. The tuned hyperparameters control model flexibility and overfitting. The number of estimators determines the number of boosting trees; maximum depth controls the complexity of each tree; learning rate determines how much each tree contributes to the final prediction; subsample controls the proportion of observations used to train each tree; and the alpha and lambda regularization terms penalize model complexity. Overall, the selected configuration uses shallow trees, a low learning rate, subsampling, and regularization, reflecting a conservative model structure intended to improve generalization and reduce overfitting.

Table 2. Optuna search space and selected XGBoost hyperparameters

Parameter	**Search range**	**Selected value**
Number of estimators	100 – 500	306
Maximum depth	3 – 8	3
Learning rate	0.01– 0.3	0.0129

subsample	0.6 – 1	0.6710
Regularization parameter alpha	0 – 2	1.9440
Regularization parameter lambda	0.10 – 2	0.1920

3.3.3 Experimental Setting

The dataset is divided into training and test sets using an approximately 75/25 temporal split. The training period includes the first 75% of observations from October 1, 2021, through September 30, 2024, while the test period includes the remaining 25% from October 1, 2024, through September 30, 2025. This split preserves the chronological order of the data and ensures that model performance is evaluated on a full year of unseen observations.

Time-series cross-validation is applied within the training data to support model selection and tuning while maintaining temporal order. For the statistical benchmark models, cross-validation is used to evaluate candidate model orders and select the final AR, ARIMA, and SARIMAX configurations. For the XGBoost model, cross-validation is used during feature selection and Bayesian hyperparameter optimization. In each cross-validation iteration, models are trained on earlier observations and validated on later observations. This approach enables model stability to be assessed across different temporal windows while avoiding information leakage from future observations.

For the XGBoost model, early stopping is applied during training to reduce overfitting. Training stops when validation performance does not improve after 50 boosting rounds. The test set is not used during feature selection, model-order selection, hyperparameter tuning, or early stopping, and is held only for final model evaluation.

3.4 Model Evaluation

Model performance is evaluated using mean absolute error (MAE) and mean absolute percentage error (MAPE). MAE measures the average absolute difference between observed and predicted daily visit counts as shown in Equation 4:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{4}$$

MAPE represents the average forecast error as a percentage of the observed visit volume, as shown in Equation 5:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{5}$$

Where y_i represents the observed visit count, $\hat{y}_i$ represents the predicted visit count, and n is the number of observations. MAE was used to quantify forecast error in the original unit of daily visits, while MAPE provided a scale-independent measure of forecasting accuracy that is easier to interpret for operational planning.

4. Results and Discussion

Table 3 presents the performance of the statistical time-series models and the XGBoost model for predicting daily pediatric ED visit volumes. Model performance was evaluated using MAE and MAPE across both the training and test periods. Given the average daily volume of 164 visits, MAE provides a practical measure of the typical daily forecasting error in visit counts, while MAPE expresses this error relative to observed demand.

Table 3. Training and test performance of forecasting models

Model	**Train MAE**	**Train MAPE**	**Test MAE**	**Test MAPE**
AR(7)	18.29	12.88%	28.40	22.56%
ARIMA(7,0,7)	17.70	11.55%	19.12	15.22%
$SARIMA(7,0,7)(1,0,1)_7$	16.91	10.69%	17.18	13.50%
XGBoost	11.68	7.38%	13.54	8.47%

As shown in Table 3, the AR(7) model produced the weakest test performance, with a test MAE of 28.40 visits and a test MAPE of 22.56%. Relative to the average daily volume of 164 visits, this corresponds to an average error of approximately 28 visits per day, indicating limited accuracy for operational planning. This demonstrates that relying only on the previous seven daily observations was insufficient to capture the full variation in pediatric ED demand. Although AR(7) can represent short-term autoregressive dependence and weekly patterns, it does not account for moving-average effects, seasonal patterns, calendar-based variation, or weather-related predictors. The large increase in error from the training period to the test period also indicates limited ability to generalize to the one-year holdout data.

The ARIMA(7,0,7) model improved substantially over AR(7), reducing test MAE to 19.12 visits and test MAPE to 15.22%. This indicates that incorporating moving-average terms helped account for recent forecast errors and short-term temporal dependence. However, an average error of about 19 visits per day remained relatively high compared with the mean daily volume, suggesting that autoregressive and moving-average components alone were not sufficient to fully capture variation in daily pediatric ED volume.

Among the statistical benchmark models, SARIMA$(7, 0, 7)(1, 0, 1)_7$ achieved the best performance, with a test MAE of 17.18 visits and a test MAPE of 13.50%. Compared with ARIMA, SARIMAX reduced test MAE by approximately 10% and test MAPE by approximately 11%. This improvement indicates that incorporating seasonal components and exogenous predictors added forecasting value. Nevertheless, SARIMAX still had an average error of approximately 17 visits per day and remained less accurate than XGBoost, particularly in terms of relative prediction error.

The XGBoost model achieved the best overall performance, with the lowest training and testing errors across both metrics. On the test set, XGBoost achieved an MAE of 13.54 visits and a MAPE of 8.47%. Relative to the average daily volume of 164 visits, this means the model's predictions differed from observed daily volume by approximately 14 visits per day on average. Compared with SARIMAX, XGBoost reduced test MAE by approximately 21% and test MAPE by approximately 37%. These improvements indicate that the feature-rich XGBoost model captured pediatric ED demand patterns more effectively than the statistical benchmark models.

The train-test comparison also demonstrates that XGBoost generalized well to the one-year holdout period. Its MAPE increased only modestly from 7.38% in training to 8.47% in testing with a difference of 1.09 percentage points. This was the smallest train-test increase among the evaluated models. In contrast, AR(7) showed a much larger decline in performance during testing, while ARIMA and SARIMAX also showed higher relative test errors. These findings indicate that XGBoost did not simply overfit the training data but learned relationships that transferred effectively to unseen observations.

Figure 4 provides a visual comparison of observed and predicted daily pediatric ED visits using the XGBoost model. The predicted series generally follows the observed visit volume across both the training and test periods, capturing the overall demand level, recurring temporal patterns, and broader seasonal changes. During the test period, predicted values remain close to the observed series, consistent with the low test MAPE reported in Table 3. However, the XGBoost predictions appear smoother than the observed daily series and do not fully reproduce some sharp one-day spikes or sudden drops. This is expected in daily ED forecasting because abrupt fluctuations may be driven by short-term events, localized outbreaks, school-related changes, or operational factors that are not fully captured by historical visit counts, calendar variables, and weather predictors. Therefore, the model appears most useful for forecasting expected daily demand patterns rather than predicting every extreme daily fluctuation.

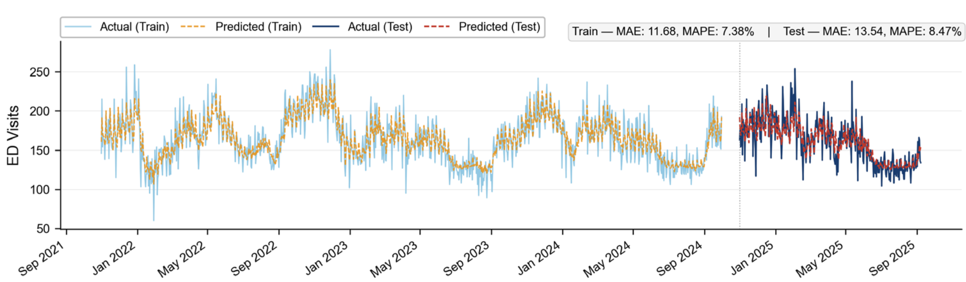

Figure 4. Observed and predicted daily pediatric ED visits across the training and test periods using the final XGBoost model.

Overall, the results indicate that incorporating engineered temporal features, rolling-window statistics, cyclical calendar encodings, weekend indicators, and weather-related variables within an XGBoost framework improved daily pediatric ED volume forecasting. The statistical models captured important time-series patterns, particularly when seasonal and exogenous components were included in SARIMAX. However, XGBoost provided stronger predictive accuracy, likely because it could model nonlinear relationships and interactions among recent demand, calendar effects, and environmental predictors. These findings support the use of feature-rich machine learning models as practical tools for pediatric ED demand planning and short-term operational decision-making.

5. Conclusion

This study developed an XGBoost-based forecasting model for daily pediatric ED visit volume and benchmarked it against three statistical time-series models, including AR(7), ARIMA(7,0,7), and SARIMA$(7,0,7)(1,0,1)_7$ using four years of post-COVID operational data from a pediatric hospital. The model incorporated lagged visit counts, rolling-window statistics, cyclical calendar encodings, a weekend indicator, and weather-related variables, and was tuned through Bayesian hyperparameter optimization. On a one-year temporal holdout, the XGBoost model achieved a test MAE of 13.54 and a test MAPE of 8.47%, corresponding to approximately a 21% reduction in test MAE and a 37% reduction in test MAPE relative to the best-performing statistical benchmark, SARIMAX. The model also provided the smallest train–test gap, indicating strong generalization across a full year of seasonal variation. Feature importance analysis identified recent demand history, particularly the 7-day rolling mean and short-horizon lagged visit counts, and day-of-week cyclicality as the dominant predictors, with temperature and wind speed contributing additional but smaller predictive value. These findings demonstrate the use of feature-rich, regularized machine learning models as practical tools for short-term pediatric ED demand planning, while also confirming that seasonal statistical models such as SARIMAX remain useful as interpretable benchmarks.

Several limitations should be acknowledged. The study used data from a single pediatric hospital, which may limit generalizability to other settings with different patient populations, catchment areas, seasonal patterns, and operational contexts. In addition, the predictor set was limited to historical utilization, calendar features, and basic weather measures, and did not include potentially important external drivers such as respiratory virus surveillance, school calendars, or community-level public health events.

Future work will extend this framework in several directions. First, forecasting will be refined to the hourly level to better support real-time staffing, patient flow, and capacity decisions. Second, the predictor set will be expanded to include epidemiological indicators, such as regional respiratory virus and RSV surveillance, syndromic data, school-calendar variables, and holiday effects. Third, probabilistic forecasting methods, including quantile regression and conformal prediction, will be incorporated to generate prediction intervals and better quantify uncertainty for capacity planning.

References

Afilal, M., Yalaoui, F., Dugardin, F., Amodeo, L., Laplanche, D. & Blua, P. *Forecasting the Emergency Department Patients Flow*. *Journal of Medical Systems*. 40(7):175, 2016, DOI: 10.1007/s10916-016-0527-0.

Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. *Optuna: A Next-Generation Hyperparameter Optimization Framework*, in *Proceedings of the 25th International Conference on Knowledge Discovery & Data Mining (KDD '19)*, pp. 2623–2631, 2019, DOI: 10.1145/3292500.3330701.

Alpern, E.R., Stanley, R.M., Gorelick, M.H., Donaldson, A., Knight, S., Teach, S.J., Singh, T., Mahajan, P., Goepp, J.G., Kuppermann, N., Dean, J.M. & Chamberlain, J.M. *Epidemiology of a Pediatric Emergency Medicine Research Network: The PECARN Core Data Project*. *Pediatric Emergency Care*. 22(10):689, 2006, DOI: 10.1097/01.pec.0000236830.39194.c0.

Amorim, A.N. de, Deardon, R. & Saini, V. *A Stacked Ensemble Method for Forecasting Influenza-Like Illness Visit Volumes at Emergency Departments*. *PLOS ONE*. 16(3):e0241725, 2021, DOI: 10.1371/journal.pone.0241725.

Asheim, A., Bache-Wiig Bjørnsen, L.P., Næss-Pleym, L.E., Uleberg, O., Dale, J. & Nilsen, S.M. *Real-Time Forecasting of Emergency Department Arrivals Using Prehospital Data*. *BMC Emergency Medicine*. 19(1):42, 2019, DOI: 10.1186/s12873-019-0256-z.

Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., Miller, P. & Fitzgerald, G. *Predicting Emergency Department Admissions*. *Emergency Medicine Journal*. 29(5):358–365, 2012, DOI: 10.1136/emj.2010.103531.

Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016, DOI: 10.1145/2939672.2939785.

Duarte, D., Walshaw, C. & Ramesh, N. *A Comparison of Time-Series Predictions for Healthcare Emergency Department Indicators and the Impact of COVID-19*. *Applied Sciences*. 11(8):3561, 2021, DOI: 10.3390/app11083561.

Etu, E.-E., Larot, J., Etu, K., Emakhu, J., Masoud, S., Tenebe, I., Huang, G., Gunaga, S. & Miller, J. *Forecasting Pediatric Emergency Department Arrivals: Evaluating the Role of Exogenous Variables Using Deep Learning Models*. *Intelligence-Based Medicine*. 12:100313, 2025, DOI: 10.1016/j.ibmed.2025.100313.

Etu, E.-E., Monplaisir, L., Masoud, S., Arslanturk, S., Emakhu, J., Tenebe, I., Miller, J.B., Hagerman, T., Jourdan, D. & Krupp, S. *A Comparison of Univariate and Multivariate Forecasting Models Predicting Emergency Department Patient Arrivals During the COVID-19 Pandemic*. *Healthcare*. 10(6):1120, 2022, DOI: 10.3390/healthcare10061120.

Fan, B., Peng, J., Guo, H., Gu, H., Xu, K. & Wu, T. *Accurate Forecasting of Emergency Department Arrivals with Internet Search Index and Machine Learning Models: Model Development and Performance Evaluation*. *JMIR Medical Informatics*. 10(7):e34504, 2022, DOI: 10.2196/34504.

Gafni-Pappas, G. & Khan, M. *Predicting Daily Emergency Department Visits Using Machine Learning Could Increase Accuracy*. *The American Journal of Emergency Medicine*. 65:5–11, 2023, DOI: 10.1016/j.ajem.2022.12.019.

Guan, G. & Engelhardt, B. *Predicting Sick Patient Volume in a Pediatric Outpatient Setting Using Time Series Analysis*, in *Machine Learning for Healthcare Conference*, pp. 271–287, 2019.

Huang, X., Yu, Z., Wei, X., Shi, J., Wang, Y., Wang, Z., Chen, J., Bu, S., Li, L., Gao, F., Zhang, J. & Xu, A. *Prediction of Vancomycin Dose on High-Dimensional Data Using Machine Learning Techniques*. *Expert Review of Clinical Pharmacology*. 14(6):761–771, 2021, DOI: 10.1080/17512433.2021.1911642.

Kandula, S., Pei, S. & Shaman, J. *Improved Forecasts of Influenza-Associated Hospitalization Rates with Google Search Trends*. *Journal of The Royal Society Interface*. 16(155):20190080, 2019, DOI: 10.1098/rsif.2019.0080.

Morzadec, T., Chappuy, H., Bouazza, N., Urien, S., Treluyer, J.-M., Foissac, F. & Beeker, N. *Forecasting the Patients Flow at Pediatric Emergency Departments*. *Journal of Medical Systems*. 45(3):29, 2021, DOI: 10.1007/s10916-021-01712-z.

National Center for Health Statistics. *Estimates of Emergency Department Visits in the United States, 2016–2022*, 2024, Available: https://www.cdc.gov/nchs/dhcs/ed-visits/index.htm.

Ozturk, C.S. & Merter, O.S. *Challenges and Facilitators in Child-Friendly Healthcare from the Perspective of Pediatric Emergency Nurses: A Qualitative Study*. *Journal of Pediatric Nursing*. 78:e180–e186, 2024, DOI: 10.1016/j.pedn.2024.07.004.

Pakalnis, A. & Heyer, G.L. *Seasonal Variation in Emergency Department Visits Among Pediatric Headache Patients*. *Headache: The Journal of Head and Face Pain*. 56(8):1344–1347, 2016, DOI: 10.1111/head.12888.

Pines, J.M., Zocchi, M.S., Black, B.S., Carlson, J.N., Celedon, P., Moghtaderi, A. & Venkat, A. *Characterizing Pediatric Emergency Department Visits During the COVID-19 Pandemic*. *The American Journal of Emergency Medicine*. 41:201–204, 2021, DOI: 10.1016/j.ajem.2020.11.037.

Rocha, C.N. & Rodrigues, F. *Forecasting Emergency Department Admissions*. *Intelligent Data Analysis*. 25(6):1579–1601, 2021, DOI: 10.3233/IDA-205390.

Tuominen, J., Lomio, F., Oksala, N., Palomäki, A., Peltonen, J., Huttunen, H. & Roine, A. *Forecasting Daily Emergency Department Arrivals Using High-Dimensional Multivariate Data: A Feature Selection Approach*. *BMC Medical Informatics and Decision Making*. 22(1):134, 2022, DOI: 10.1186/s12911-022-01878-7.

Vollmer, M.A.C., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S. & Bhatt, S. *A Unified Machine Learning Approach to Time Series Forecasting Applied to Demand at Emergency Departments*. *BMC Emergency Medicine*. 21(1):9, 2021, DOI: 10.1186/s12873-020-00395-y.

Vural, O., Ozaydin, B., Booth, J., Lindsey, B.F. & Ahmed, A. *Deep Learning-Based Forecasting of Boarding Patient Counts to Address Emergency Department Overcrowding*. *Informatics*. 12(3):95, 2025, DOI: 10.3390/informatics12030095.

Zhang, Y., Zhang, J., Tao, M., Shu, J. & Zhu, D. *Forecasting Patient Arrivals at Emergency Department Using Calendar and Meteorological Information*. *Applied Intelligence*. 52(10):11232–11243, 2022, DOI: 10.1007/s10489-021-03085-9.

Zhao, X., Lai, J.W., Ho, A.F.W., Liu, N., Ong, M.E.H. & Cheong, K.H. *Predicting Hospital Emergency Department Visits with Deep Learning Approaches*. *Biocybernetics and Biomedical Engineering*. 42(3):1051–1065, 2022, DOI: 10.1016/j.bbe.2022.07.008.

Biographies

Ahmed Gailan Qasem is an Assistant Professor of Health Informatics, School of Health Professions, University of Alabama at Birmingham. He previously worked in the healthcare industry as a Senior Data Scientist and Health Systems Engineer. He holds a B.S. and M.S. in Industrial Engineering from Jordan University of Science and Technology and a Ph.D. in Industrial and Systems Engineering from Binghamton University. His research focuses on advanced data science, simulation, and optimization to support decision making and improve healthcare operations and outcomes.

Lori Pelletier is Chief Quality and Patient Safety Officer at Connecticut Children's and The Alexis and William Popik, MD, Quality & Patient Safety Endowed Chair. She is also an Assistant Professor at UMass Medical School in the Division of Health Informatics and Implementation Science in the Department of Quantitative Health Sciences. She completed her PhD in Health Systems Engineering at Worcester Polytechnic Institute focusing on design, systems, processes and statistics. Dr. Pelletier worked for nine years at UMass Memorial Health Care building an organizational capability of improvement and operational excellence through systems design and people development. Prior to her PhD, Dr. Pelletier worked for many years as a Senior Quality Engineer and leader, Lean/Six Sigma Black Belt and ISO Lead Auditor at Solutia, Inc. Dr Pelletier lectures nationally on her current research in suicide prevention, cultural transformation, quality improvement, and the use of electronic health records.

Bunyamin Ozaydin is an Associate Professor of Health Informatics at the School of Health Professions of the University of Alabama at Birmingham. He previously worked in Departments of Ophthalmology and Anesthesiology as system and database developer and administrator. He holds a B.S. in Physics from Marmara University and an M.S. in Electrical Engineering and a Ph.D. in Computer Engineering from the University of Alabama at Birmingham. His research focuses on data infrastructures, transformation, mining, and analytics and decision support systems in healthcare.

Abdulaziz Ahmed is an Associate Professor in UAB's Health Informatics Graduate Programs. He earned his BS and MS in Industrial Engineering from Jordan University of Science and Technology and his PhD in Industrial and Systems Engineering from SUNY Binghamton. His research develops AI-driven predictive and prescriptive analytics for healthcare by integrating machine learning (including deep learning and NLP) with simulation and optimization, and he is the contact PI for an AHRQ-funded R21/R33 on mitigating emergency department crowding and has published in Expert Systems with Applications, Operations Research for Health Care, and Healthcare Management Science.