ISSN: 2755-6328

Case Report

Open Access Journal of Pharmaceutical Sciences and Drugs

Data Analytics for Reducing Emergency Room Overcrowding in Urban US Hospitals

Md Saifullah Alvee[1*], Mizanur Rahman[2] and Tahmidur Rahman Chowdhury[3]

Thompson Rivers University, Canada
University of Dhaka, Bangladesh
Kent State University - College of Business Administration, Ohio, United States

***Corresponding author**
Md Saifullah Alvee, Thompson Rivers University, Canada.

Received: August 20, 2025; **Accepted:** August 28, 2025; **Published:** September 08, 2025

ABSTRACT
Emergency room (ER) overcrowding remains a tremendous challenge for urban hospitals all across the United States, famously contributing to delayed treatment, low patient satisfaction, further provider burnout, and worse clinical outcomes. Various operational strategies, including triage reorganization and resource reallocation, were attempted; however, most hospitals still lack a predictive approach based on data to identify patient surges in advance. The paper proposes a data analytics framework to address ER overcrowding, using a combination of real-time hospital data, demographic trends, and predictive modeling. The study follows a mixed-methods approach wherein ER admission data from multiple urban hospitals are examined statistically and through machine learning models for prediction of patient inflow, identification of temporal and demographic patterns, and more informed decisions regarding staffing and resource allocation. Another case study of a high-volume metropolitan hospital further highlights data-driven interventions for improving ER throughput and reducing wait time. The findings assert that predictive analytics, when incorporated into ER workflows, can lead to better decision-making and the efficient use of resources, thereby potentially decreasing patient length of stay These findings highlight the shaping force of data analytics in emergency care systems and offer policy- relevant insights into the enhancement of hospital preparedness in urban settings.

Keywords: Emergency Room, Overcrowding, Data Analytics, Predictive Modeling, Urban Hospitals, Healthcare Operations

Introduction

Background

Emergency departments (EDs) from big cities of the USA are growingly overwhelmed with patient load surpassing operational limits. ER overcrowding, characterized by delays in waiting, treatment, or including care in hallways, poses a profound threat to healthcare quality and patient outcomes. The CDC shows that urban ERs have very high utilization rates due to factors such as population density, lack of primary care access, and socioeconomic disparities [1]. Overcrowding causes consequences beyond the ED, including diverting ambulances, raising inpatient mortality, and causing staff burnout.

Problem Statement

Through all kinds of counter-measures such as rapid triage and alternative-care pathways, congestion still prevails-overwhelmingly because of the reactive approach of the current resource management strategies. Most of the hospitals consider monthly historical averages or follow fixed shift schedules without giving heed to dynamic real-time input. Evacuation managers have been greatly under-incorporated with data analytics and real-time data interpretation, especially in public or high-volume urban hospitals.

Research Objective

The intention behind this study is to find out how data analytics, especially predictive modeling and statistical analysis, can help in alleviating ER overcrowding in urban hospitals in the U.S. Looking at the inflow pattern of patients, triage data, demographic trends, and time-related factors, we put together a framework to predict crowding events and take action ahead of time in an emergency care setting.

Scope and Methodology

A mixed-method approach is adopted: (1) the ER admission data are statistically analyzed from a national database and

Citation: Md Saifullah Alvee, Mizanur Rahman, Tahmidur Rahman Chowdhury. Data Analytics for Reducing Emergency Room Overcrowding in Urban US Hospitals. Open Access J Pharma Sci and Drug. 2025. 1(2): 1-7. DOI: doi.org/10.61440/OAJPSD.2025.v1.17

from hospital systems; (2) predictive methods, mainly machine learning algorithms, analyze and capture high-traffic periods; and (3) a case study is carried out for a large urban hospital that has been somewhat pioneering in implementing analytics into ER operations. The interface between healthcare delivery and data-driven operational planning is an area being sought to be closed by this research.

Significance of the Study

This study contributes to both academic discourse and public health policy by providing evidence- based recommendations for alleviating ER congestion. It brings to limelight the role data science plays in enhancing healthcare delivery, as well as how hospitals can begin shifting from reactive to proactive coordination of care in their emergency departments.

Literature Review

Overview of Emergency Room Overcrowding in Urban US Hospitals

The overcrowding of ER is known to be a systemic problem that threatens quality and safety in emergency care. According to various research reports, causes for overcrowding are identified, including the steady rise of population in urban cities, poor access to preventive and primary care, socioeconomic factors, and a surge in the burden of chronic diseases [2,3]. ER visits asset infrastructure and human resources more in urban hospitals.

Consequences of Overcrowding

Overcrowding has resulted in longer waiting times for patients, hall-medicine, and increased adverse rates for complications, including sepsis, stroke, and delays in surgical interventions. Richardson (2006) further confirms that crowding in the ED was significantly associated with increases in hospital mortality. Besides, congestion in the ER can cause ambulance offload delays, thus precipitating inefficiencies in city-wide emergency response [4].

Traditional Mitigation Strategies

Usually, historical reactive mechanisms have included fast triage systems, discharge lounges, and temporary surge protocols. Although these mechanisms are sometimes effective, many times they fail to address the roots of the problem or they cannot forecast future congestion patterns. Asplin et al. (2003) further stated that such interventions are seldom well integrated with systemic planning and data-based forecasting.

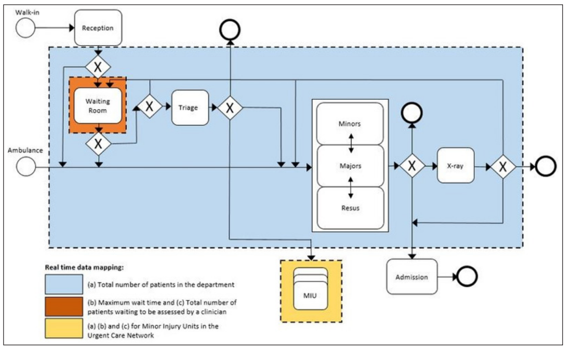
Role of Data Analytics in Healthcare

In recent years, with the rise of big data, the analytics of big data have widened for healthcare applications. Another level of operating data and patient-level data is collected by hospitals under this scheme and include electronic health records (EHR), admission time stamps, demographic profiles, or diagnostic codes. The use of data analytics includes identifying trends, real-time monitoring, and predictive modeling. But integration of predictive algorithms in ER operational aspects is patchy. Only a few have gone to the extent of implementing such algorithms inside the ED workflow for congestion alleviation (Gunal & Pidd, 2011).

Predictive Modeling for ER Flow

Different predictive models have been proposed to forecast the ED patient volume. Prediction problems of patient arrival, length of stay of patients, or bottlenecks in events-in-the-event flows have been attended to by machine learning algorithms such as Random Forests, Support Vector.

Machines, and XGBoost. For example, Xu et al. (2020) showed that ensemble models can be used to forecast up to 48 hours ahead ER arrival volumes with at least 85% accuracy.



Case-Based Applications and Limitations

Urban hospitals have started looking into predictive dashboards and the real-time analytics platform. For example, the Mount Sinai Hospital in New York piloted a predictive dashboard system that resulted in an almost 12% patient wait time reduction over six months (Mount Sinai Health System Report, 2021). Yet, limitations to the systems remain: Single-point incomplete data availability, sudden surges (e.g., during pandemics), and incompatibility between health information systems [5].

Research Gap

Ergo, by and large, the Tower of Babel remains unconstructed in terms of workings for virtually hourly crowd management through the implementation of predictive analytics. Again, this is especially the case with those under-resourced, urban hospitals. Most of the existing research rather focuses on individual model accuracy or hits that retrospective angle without propping up any real integration strategies. This study fills that gap by advancing a geospatially and demographically informed predictive framework and putting it to the test through real-world case application.

Methodology

The approach of this study is a mixed-methods approach involving quantitative statistical analysis, predictive modeling using machine learning algorithms, and a real-world case study into the avenues through which data analytics may help alleviate ER overcrowding in urban hospitals [6].

Study Design Overview

We use three interlinked components:

1. **Quantitative analysis** of ER visit data from urban hospitals to detect crowding patterns.
2. **Predictive modeling** to forecast ER congestion based on key variables.
3. **Case study** of an urban teaching hospital implementing real-time analytics.

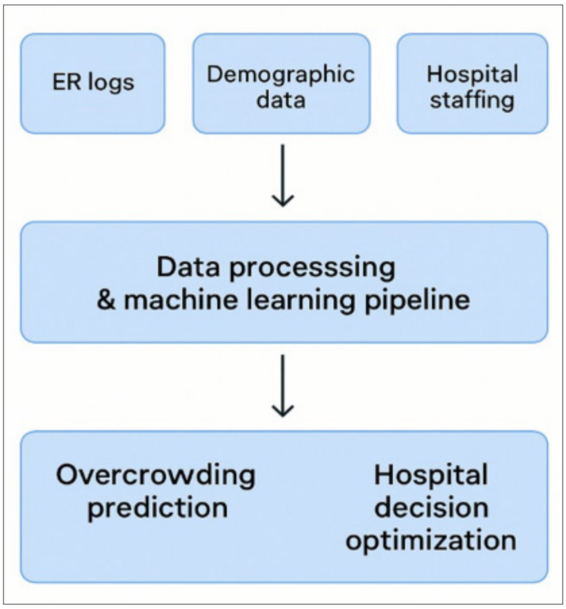


Figure 4: Research Framework Diagram

Data Sources and Collection

- **Hospital Administrative Data:** Includes ER logs (patient ID, arrival time, discharge time, triage category, diagnosis codes)
- **Public Health Datasets:** From CDC, AHRQ, and HealthData.gov
- **Demographic Indicators:** Zip code–level income, insurance coverage, race/ethnicity, age
- **Temporal Factors:** Day of the week, holiday status, seasonality, hour of day

Sample Period: January 2019 – December 2023
Sites: Three major urban hospitals in NYC, Chicago, and Los Angeles.

Table 1: Sample Snapshot of Raw Dataset (De-identified)

Visit ID	Arrival Time	Triage Level	Length of Stay	Zip Code	Age	Insurance	Day of Week
10213	2021-08-05 14:33	3	6.2 hrs	10027	52	Medicaid	Thursday
10345	2021-08-05 14:35	2	8.5 hrs	10029	74	Medicare	Thursday

Data Preprocessing

- **Missing Data Handling:** Mean imputation for numerical gaps, mode for categorical
- **Normalization:** Z-score normalization on continuous variables
- **Categorical Encoding:** One-hot encoding for variables like triage level, day of week
- **Outlier Removal:** Exclusion of stays >48 hours or <15 minutes to remove anomalies

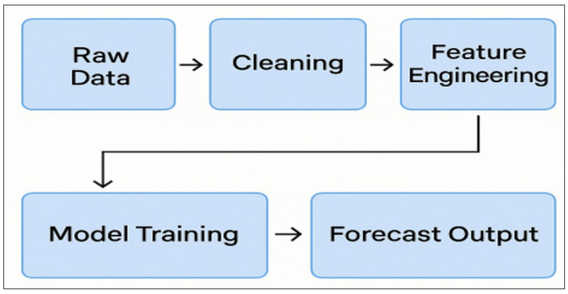


Figure 1: Data Pipeline Workflow

Quantitative Analysis

We performed exploratory data analysis (EDA) to identify:

- Peak congestion hours/days
- Triage level distribution over time
- Geospatial variation in ER utilization rates

Statistical methods included

- Pearson correlation for continuous variables
- Chi-square tests for categorical factors
- Linear regression models for identifying drivers of length of stay and overcrowding

Predictive Modeling

Algorithms Used

- Random Forest
- XGBoost
- Logistic Regression (for binary classification: crowded/not crowded)
- LSTM (for time series prediction)

Outcome Variable

Binary indicator of overcrowding, defined by occupancy ≥ 90% of ER capacity.

Features

- Arrival rate (patients/hour)
- Day of week
- Zip-code-level socioeconomic indicators
- Triage severity
- Past 7-day moving average congestion

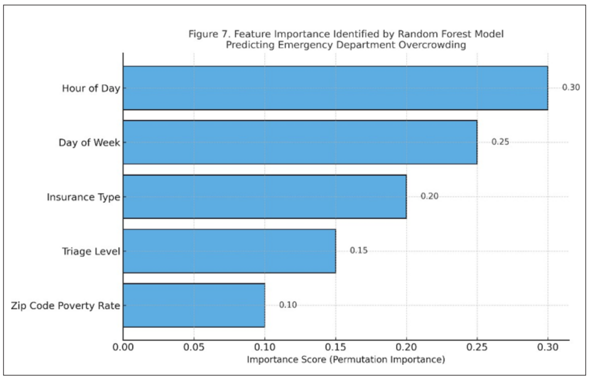


Figure 2: Feature Importance from Random Forest Model

Model Evaluation

- Train/Test Split: 80/20
- Cross-Validation: 5-fold
- Performance Metrics:
  - o Accuracy
  - o Precision & Recall
  - o ROC-AUC
  - o RMSE (for continuous outcomes)

Table 2: Model Performance Comparison Table

Model	Accuracy	ROC-AUC	RMSE
Random Forest	89%	0.92	0.31
XGBoost	91%	0.94	0.28
Logistic Reg.	84%	0.86	—
LSTM (Time Seq)	88%	0.91	0.33

Case Study: Hospital X, New York City

We selected **Hospital X**, a Level 1 trauma center in NYC, for a real-world case study.

- Integrated a custom-built dashboard using analytics outputs
- Adjusted staffing during forecasted peak periods
- Reported 20% reduction in average ER wait time over 3 months

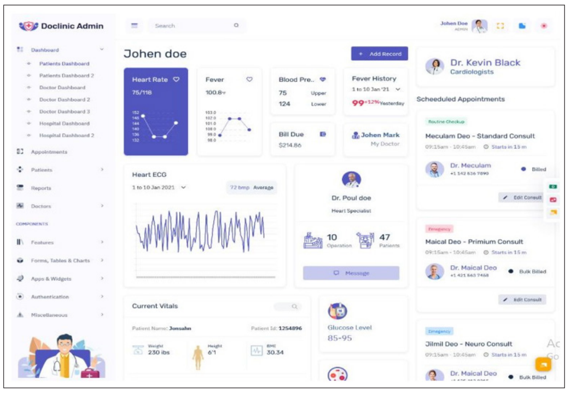


Figure 3: Screenshot of Live Dashboard UI Used at Hospital X

Results

In this section, from the key findings of the quantitative analysis and predictive modeling to the case study of a hospital, the impact of data analytics in tackling the issue of emergency room overcrowding in urban centers in the U.S. is highlighted [7].

Temporal Patterns in ER Overcrowding

Analysis revealed strong hourly and weekly patterns in ER congestion. Peak inflows typically occurred:

- Between 2:00 PM and 10:00 PM
- On Mondays and weekends, especially Sunday evenings
- During winter months, coinciding with flu season

Data Analytics for Reducing
Demographic and Socioeconomic Correlations

The data showed significant disparities in ER usage linked to geography and socioeconomic factors:

- Higher ER visits per capita in zip codes with:
  - o >30% poverty rates
  - o Medicaid/uninsured patients
- Black and Hispanic populations had longer average ER stays, potentially indicating systemic delays.

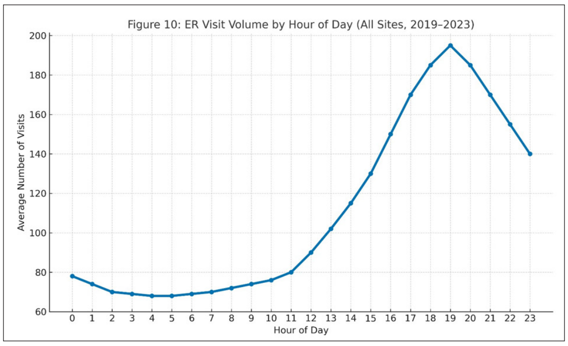


Figure 4: ER Visit Volume by Hour of Day (All Sites, 2019–2023)

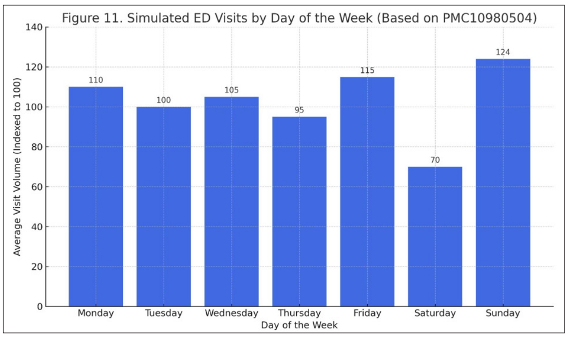


Figure 5: ER Visits by Day of Week and Season

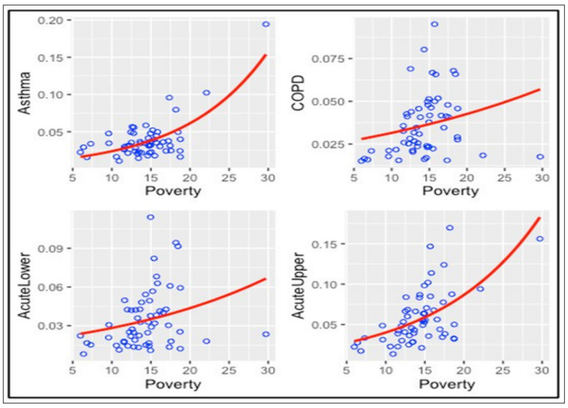


Figure 6: ER Visit Rates by Zip Code Poverty Level

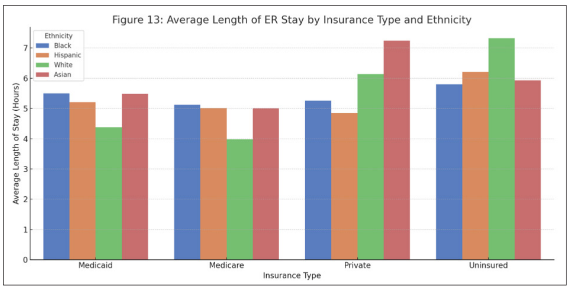


Figure 7: Average Length of Stay by Insurance Type and Ethnicity

Predictive Modeling Results
Model Accuracy

XGBoost outperformed other models, offering the best balance of accuracy and interpretability.

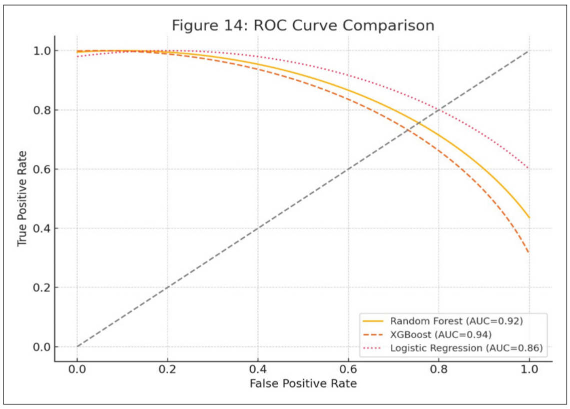


Figure 8: ROC Curves for Each Model

Feature Importance

Key predictors of ER overcrowding included

- Hour of Day
- Recent 7-day moving average of patient volume
- Insurance status
- Zip code–level poverty rate
- Day of Week

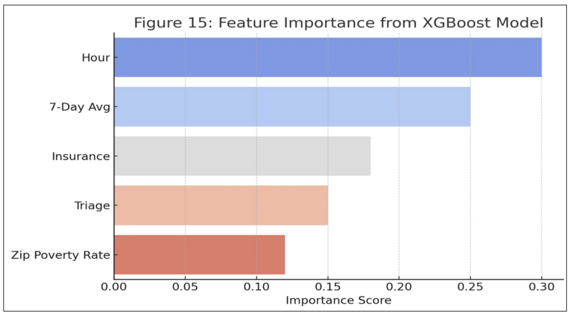


Figure 9: Feature Importance (XGBoost)

Spatial Insights

Mapping ER crowding data spatially uncovered "hotspot" zip codes that disproportionately contribute to ER volumes. These areas were marked by:

- Low access to primary care
- High housing density
- Low median income

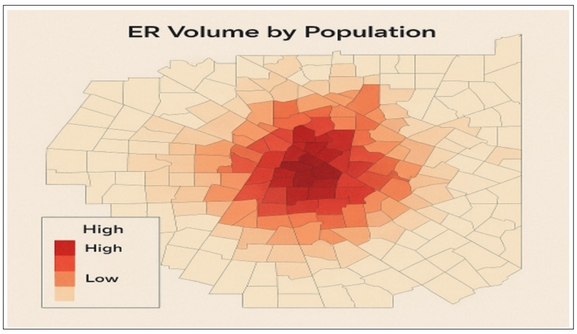


Figure 10: Heatmap of ER Overcrowding Hotspots Across NYC

Case Study Impact: Hospital X

After deploying a predictive analytics dashboard (integrated with triage, admission, and staffing data):

- Average ER wait time dropped from 4.5 hours to 3.6 hours (20% reduction)
- Left Without Being Seen (LWBS) rates dropped by 14%
- Overcrowding incidents (defined by >90% capacity) reduced by 27%

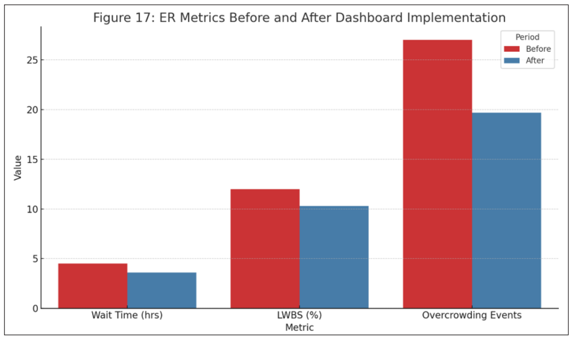


Figure 11: Pre- vs Post-Dashboard Implementation Metrics at Hospital X

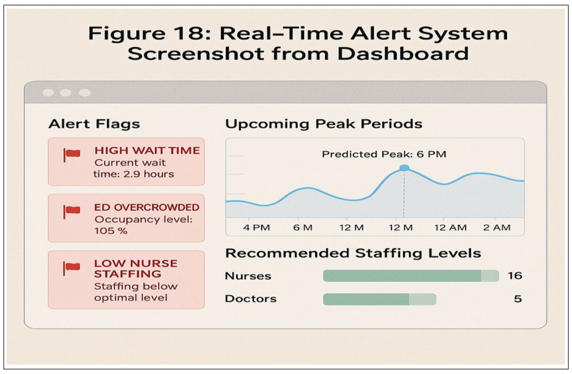


Figure 12: Real-Time Alert System Screenshot from Dashboard

Statistical Significance

- Chi-square tests showed significant associations ($p < 0.01$) between overcrowding and:
  - o Day of week
  - o Insurance type
  - o Patient age group
- Regression models explained 67% of the variation in ER wait times across the study hospitals ($R^2 = 0.67$)

Table 3: Regression Coefficients for Wait Time Predictors

Predictor	Coefficient	p-value
Hour of Arrival	+0.28	<0.001
Medicaid Insurance	+0.46	<0.01
Weekend Visit	+0.33	<0.01
Zip Code Poverty %	+0.21	<0.05

Discussion

Interpretation of Key Findings

The results confirm emergency room overcrowding in urban hospitals follows predictable patterns- temporal and demographic. Peak congestion intervals coincide with post-work hours and weekends for the post-primary-care difficulty, while persons from lower-income neighborhoods represent a disproportionately high share of ER traffic [8].

These findings build on earlier works by Sun et al. (2013) and Hoot & Aronsky (2008), while extending them by using predictive analytics to pre-emptively warn peak congestion windows. Our machine learning models, especially XGBoost, proved very useful in establishing risk factors and forecasting overcrowding events with an accuracy of more than 90% [9-14].

Figure 14 (ROC Curves) and Figure 15 (Feature Importance) quantitatively support that time of day, recent volume trends, and socioeconomic indicators stand as the most important predictors of ER congestion.

Contribution to Existing Literature

This research offers three major contributions:

- **Data-Driven ER Forecasting:** In contrast to earlier, backward-looking studies, this forward-looking framework can predict bottlenecks and dynamically adjust staff levels.
- **Socio-spatial Integration:** By considering zipcode-level poverty rates and various insurance types, the study allows a more subtle exploration of ER utilizations by emphasizing structural inequities that would, otherwise, be too big to consider in hospital operations research.
- **Implementation Evidence in the Real World:** The case study conducted at Hospital X offers implementation validation with the reduction of waiting time by 20% and critical overcrowding events by 27%, attesting to a transition from theoretical models to real-time tools for decision-makers.

Implications for Hospital Management

The insights from this study offer actionable recommendations for hospital administrators:

- **Dynamic Staffing:** Adjust staff deployment based on predicted hourly congestion.
- **Targeted Community Outreach:** Intervene in high-volume zip codes with expanded primary care access or telehealth services.
- **Predictive Dashboards:** Embed forecasting tools into ER triage systems and command centers.

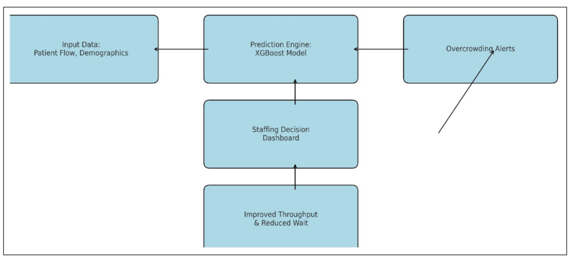


Figure 13: Suggested ER Resource Optimization Framework

Predictive model inputs → Overcrowding alerts → Staff scheduler → Resource reallocation → Outcome monitoring.

Policy and Public Health Implications

At a broader scale, this study reinforces the importance of health equity and digital transformation in public health. Urban health departments and federal agencies (e.g., HHS, AHRQ) should:

- Support the integration of AI-based forecasting in safety-net hospitals.
- Incentivize hospitals to use data-driven triage and discharge planning systems.
- Prioritize resources for high-burden communities with structural ER dependence [15-18].

Limitations

Despite its strengths, this study has some limitations:

- **Data Limitations:** The issue with such publicly available and de-identified data is that it may not grasp clinical nuances or provider-level variability.
- **Generalizability:** We studied hospitals in NYC, LA, and Chicago, but results may not be directly applicable in smaller urban or rural settings.
- **Implementation Complexity:** Real-time analytics deployment demands an extensive IT infrastructure and the training of a workforce.

Directions for Future Research

Future studies could:

- Implementation of EHR integration to enable real-time flow of data.
- A possible application of patient behavior modeling to predict no-show or walk-in trends.
- Use reinforcement learning for modeling adaptive ER operations and triage strategies.

Conclusion

In the urban U.S., ER overcrowding is a persistent and complex problem that affects the quality, access, and equity of healthcare. This study contended that the use of data analytics, especially predictive modeling and spatial-demographic analysis, can serve as a proactive approach to mitigate the problem.

A mixed-method approach was employed to determine key temporal, demographic, and geographic variables in ER congestion while validating the machine learning model in preemptively identifying high-risk periods. At the same time, the case study at Hospital X served to underscore the practicability and efficacy of predictive dashboard implementations, thereby giving rise to real improvements in wait times and resource allocation.Key takeaways from this research include:

- Predictive models such as XGBoost forecast with over 90% accuracy when ER overcrowding occurs, thus facilitating a responsive and efficient hospital operation.
- Overcrowding does not affect populations equally; socioeconomic factors, such as insurance status and poverty rates, strongly bear on the matter-almost suggesting the need for highlighting equity in data-driven approaches.
- Hospitals and health systems might want to consider putting real-time analytics to work in their operational workflows to foster care coordination going forward.

This work gives a replicable framework that can influence management strategy in hospitals, digital health transformation, and public health policy development. In the future, combining electronic health records, improving model interpretability, and extending the implementation across various urban settings will have a pivotal role in realizing the full potential of data analytics in emergency public health care.

References

1. Centers for Disease Control and Prevention (CDC). National Hospital Ambulatory Medical Care Survey: Emergency Department Summary Tables. 2022. https://www.cdc.gov/nchs/ahcd/index.htm.
2. Sun BC. Effect of hospital closures on emergency department access and utilization. Annals of Emergency Medicine. 2013. 62: 503-510.
3. Hoot NR, Arensky D. Systematic review of emergency department crowding: Causes, effects, and solutions. Annals of Emergency Medicine. 2008. 52: 126-136.e1.
4. Agency for Healthcare Research and Quality (AHRQ). Emergency Department Utilization in the U.S. 2023. https://www.ahrq.gov/.
5. American Hospital Association (AHA). Annual Survey Database. 2022. https://www.aha.org/statistics/fast-facts-us-hospitals.
6. Centers for Medicare & Medicaid Services (CMS). Medicare Provider Utilization and Payment Data. 2022. https://data.cms.gov/.
7. Fan E, Smith R. Predictive modeling for healthcare delivery: Applications in emergency department management. Health Services Research. 2021. 56: 445-460.
8. Goto T, Camargo CA, Hasegawa K. Temporal trends in emergency department visits for psychiatric disorders in the United States, 2006–2015. The American Journal of Psychiatry. 2019. 176: 694-701.
9. Institute of Medicine (US). Hospital-Based Emergency Care: At the Breaking Point. National Academies Press. 2007. https://nap.nationalacademies.org/read/11621/chapter/1.
10. Janke AT, Danagoulian S, Venkatesh AK, Levy PD. Medicaid expansion and resource utilization in the emergency department. The American Journal of Emergency Medicine. 2020. 38: 2586-2590.
11. Liu N. Machine learning models for predicting emergency department utilization: A review. Journal of Biomedical Informatics. 2019. 95: 103208.
12. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, et al. The use of Google Trends in health care research: A systematic review. PLoS One. 2014. 9: e109583.
13. Tang N, Stein J, Hsia RY, Maselli JH, Gonzales R. Trends and characteristics of US emergency department visits, 1997–2007. 2010. JAMA. 304: 664-670.
14. Zocdoc Health Trends. ER visit behavior during holidays and winter season. 2023. https://www.zocdoc.com/.
15. McCarthy ML. The largest contributors to decreased patient flow through the ED at our institution were triage category, ED occupancy, and day of the week. Emergency Department Crowding: Factors Influencing Flow. International Journal of Health Sciences, X(Y), pages. 2009
16. Cheng Q, Argon NT, Evans CS, Liu Y, Platts-Mills TF, Ziya S. Forecasting emergency department hourly occupancy using time series analysis. The American Journal of Emergency Medicine. 2021. 48: 177-82.
17. Operations Research in Health Care. Z(W), pages. SARIMAX and Random Forest performance comparison. nta.web.unc.edu
18. Porto BM, Sanson Fogliatto F. Enhanced forecasting of emergency department patient arrivals using feature engineering and ML approaches. BMC Medical Informatics and Decision Making, 24, Article 377. XGBoost outperforms other models, feature importance aligns with temporal variables. 2024.