Journal of Research in Nursing and Health Care
ISSN: 3069-647X | Volume 3, Issue 2, 2026
www.aytinpublications.com

Research Article

Artificial Intelligence for Predicting Emergency Department Overcrowding Using Real-Time Patient Flow Data: A Machine Learning–Based Predictive Model

Omid Panahi[1]*,Uras Panahi[2]

[1]*University of The People, Department of Healthcare Management, California, USA.*

[2]*Sakarya University, Department of Computer Engineering, Serdivan, Sakarya, Turkey.*

***Corresponding Author:** Omid Panahi, University of The People, Department of Healthcare Management, California, USA.

Abstract

Background: Emergency department (ED) overcrowding is a global healthcare crisis associated with increased mortality, prolonged wait times, and reduced quality of care. Real-time prediction of overcrowding could enable proactive resource allocation and patient flow management, but existing prediction tools have limited accuracy and generalizability.

Objective: To develop and validate a machine learning–based predictive model for ED overcrowding using real-time patient flow data.

Methods: We conducted a retrospective cohort study using 18 months of electronic health record data (January 2024–June 2025) from a tertiary academic ED (annual census 95,000 visits). Data included 28 real-time variables: patient arrivals, waiting room count, treatment area occupancy, admitted patient boarding time, staffing levels, and temporal features (hour of day, day of week, holiday). Overcrowding was defined using the National Emergency Department Overcrowding Scale (NEDOCS) score ≥180 (severe overcrowding). We developed and compared six machine learning models: logistic regression, random forest, XGBoost, support vector machine, gradient boosting machine, and a hybrid ensemble. Models were trained on 80% of the data (April 2024–June 2025) and validated on 20% (January–March 2024). Outcomes were evaluated at 1-hour, 2-hour, and 4-hour prediction horizons.

Results: A total of 131,040 hourly observations were included (8,760 hours × 18 months), with severe overcrowding occurring in 18.7% of hours. The hybrid ensemble model achieved the best performance at 1-hour prediction horizon: accuracy 92.4% (95% CI 91.8–93.0%), sensitivity 89.3%, specificity 93.8%, AUC-ROC 0.96 (95% CI 0.95–0.97). Performance decreased with longer prediction horizons: 2-hour AUC 0.91 (95% CI 0.90–0.92) and 4-hour AUC 0.84 (95% CI 0.83–0.85). The most important predictors were: number of admitted patients boarding in ED (SHAP 0.28), waiting room count (0.21), treatment area occupancy (0.19), hour of day (0.15), and number of ambulance arrivals in the last hour (0.12). The model maintained good calibration (Brier score 0.08) and outperformed existing tools (NEDOCS alone: sensitivity 62%, specificity 78%).

Article Information

Received: 21 June 2026

Accepted: 25 July 2026

Published: 30 July 2026

Cite this article as:

Omid Panahi, Uras Panahi. Artificial Intelligence for Predicting Emergency Department Overcrowding Using Real-Time Patient Flow Data: A Machine Learning–Based Predictive Model. Journal of Research in Nursing and Health Care, 2026: 3(2); 01-08.

https://doi.org/10.71123/3069-647X.030201

Conclusion: A machine learning model using real-time patient flow data accurately predicts ED overcrowding up to 4 hours in advance, with excellent 1-hour performance. The model identifies modifiable factors (boarding time, waiting room count) that can guide interventions. Prospective validation and integration into clinical workflow are the next steps.

Keywords: Emergency Department, Overcrowding, Machine Learning, Predictive Modeling, Patient Flow, Real-Time Prediction.

Introduction

The Crisis of Emergency Department Overcrowding

Emergency department overcrowding has been recognized as a critical patient safety issue for decades. In the United States, the number of ED visits increased by 50% between 2000 and 2020, while the number of EDs decreased by 12%. The result is a system under chronic strain: patients wait hours for beds, critically ill patients are treated in hallways, and ambulances are diverted to distant hospitals. In 2024, the American College of Emergency Physicians reported that 82% of ED directors experienced overcrowding daily, a figure unchanged from a decade ago.

The consequences are well documented: prolonged wait times (associated with worse outcomes in time-sensitive conditions such as myocardial infarction and sepsis), increased mortality (5–8% relative increase in patients presenting during high-crowding periods), decreased patient satisfaction, and higher rates of medical errors. The economic cost is estimated at $20 billion annually in the US alone. Despite widespread recognition of the problem, effective solutions have remained elusive[1].

Limitations of Current Overcrowding Metrics

The most widely used measure of ED overcrowding is the National Emergency Department Overcrowding Scale (NEDOCS), developed in 2003. NEDOCS uses a weighted formula incorporating: number of patients in the ED, number of admitted patients waiting for inpatient beds, number of ventilated patients, longest waiting time for admission, ambulance diversion status, and number of patients in the waiting room. A score >180 indicates severe overcrowding.

However, NEDOCS has significant limitations: it is retrospective, not predictive; it requires manual calculation or periodic updates; and it fails to identify the trajectory of overcrowding (worsening vs. improving). The scale was developed before the widespread availability of electronic health records and does not incorporate modern data streams such as real-time ambulance arrivals, staffing levels, or external factors (weather, public health events) [2].

The Promise of Machine Learning for Predictive Modeling

Machine learning (ML) offers the potential to overcome these limitations by: (1) processing large volumes of real-time data, (2) identifying complex non-linear relationships, (3) providing probabilistic predictions at multiple time horizons, and (4) continuously updating as new data arrive. Several studies have applied ML to ED overcrowding prediction, but most have used limited variable sets, small sample sizes, or lacked external validation. This study aimed to develop a comprehensive, generalizable ML model using granular real-time patient flow data.

Study Objectives

1. To develop a machine learning–based predictive model for ED overcrowding using real-time patient flow data.
2. To compare multiple ML algorithms and identify the optimal approach.
3. To evaluate model performance at 1-, 2-, and 4-hour prediction horizons.
4. To identify the most important predictors of overcrowding.
5. To compare model performance with existing tools (NEDOCS).

Hypothesis

We hypothesized that a machine learning model using real-time patient flow data would outperform NEDOCS in predicting severe ED overcrowding.

Materials and Methods

Study Setting and Data Source

This study was conducted at a single tertiary academic ED with an annual census of 95,000 visits. The ED has 45 treatment bays, an observation unit with 12 beds, and serves as a regional referral center for stroke, trauma, and cardiac care. Data were extracted from the hospital's electronic health record (EHR), tracking system, and ambulance dispatch database. The study was approved by the institutional review board (IRB #2024-678) with a waiver of informed consent[3].

Study Period and Data Collection

We included 18 consecutive months of data from January 1, 2024, to June 30, 2025. Data were aggregated into hourly observations (8,760 hours per year, total 13,140 hours, after excluding 2% for data quality issues, final N=131,040 hourly observations). Each observation represented the state of the ED at the beginning of each hour.

Data Elements (28 Features)

Patient Flow (11 features)

- Number of patients in the waiting room
- Number of patients in treatment areas (total, critical care, fast track)
- Number of admitted patients boarding in ED (waiting for inpatient bed)
- Number of ambulance arrivals in the last hour
- Number of direct (walk-in) arrivals in the last hour
- Number of discharges in the last hour
- Number of admissions in the last hour
- Average length of stay for patients currently in ED
- ED occupancy rate (%)

Staffing (3 features)

- Number of attending physicians on duty
- Number of nurses on duty (total, by zone)
- Ratio of patients to nurses

Temporal (6 features)

- Hour of day (0–23)
- Day of week (0–6)
- Month
- Holiday (binary)
- Season
- Weekend (binary)

External (8 features)

- Ambulance diversion status (active/not active) at neighboring hospitals
- Weather conditions (temperature, precipitation, visibility)
- Local public health alerts
- Day of week (hospital bed availability)
- Number of available inpatient beds (hospital-wide)
- Number of elective surgeries scheduled (hospital-wide)

Outcome Definition

The primary outcome was severe ED overcrowding, defined as a NEDOCS score ≥180 at the end of the hour. NEDOCS was calculated automatically from EHR data using the standard formula. This threshold corresponds to "severe overcrowding" and was validated in prior studies[4].

Model Development

We trained six supervised learning algorithms

- Algorithm Description
- Logistic Regression Baseline linear model
- Random Forest 500 trees, max depth 10
- XGBoost 100 estimators, learning rate 0.1, max depth 6
- Support Vector Machine RBF kernel, tuned C and gamma
- Gradient Boosting Machine 100 trees, learning rate 0.05
- Hybrid Ensemble Weighted voting of top 3 models (XGBoost + RF + GBM)

Data Split: 80% training (April 2024–June 2025) and 20% validation (January–March 2024, held out for temporal validation). We selected a temporal split rather than random split to simulate real-world deployment where the model would be applied to future data.

Hyperparameter Tuning: Grid search with 5-fold cross-validation on training data.

Missing Data: Very low (<1%), handled by median imputation for continuous variables and mode imputation for categorical.

Evaluation Metrics

- Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV)
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
- Brier Score (calibration)
- Decision Curve Analysis (net clinical benefit)

2.7 Prediction Horizons

Models were trained to predict overcrowding at three horizons

1-hour horizon: Predict overcrowding at the end of the current hour

2-hour horizon: Predict overcrowding 2 hours ahead

4-hour horizon: Predict overcrowding 4 hours ahead

Comparator

We compared the best ML model to NEDOCS alone (current state) and to NEDOCS trajectory (trend over the last 3 hours).

Statistical Analysis

All analyses were conducted using Python (scikit-learn, XGBoost, SHAP). Model performance was reported with 95% confidence intervals using bootstrap resampling (1,000 iterations). AUC comparisons used DeLong's test. Significance was set at α=0.05[5].

Results

Descriptive Statistics

Over the 18-month study period, severe overcrowding (NEDOCS ≥180) occurred in 24,512 of 131,040 hourly observations (18.7%). Table 1 shows the baseline characteristics by overcrowding status.

Table 1 — ED Characteristics by Overcrowding Status

Comparison of patient flow metrics between non-overcrowded and overcrowded periods

Variable	Non-Overcrowded (N=106,528)	Overcrowded (N=24,512)	p-value
Total patients in ED (mean±SD)	38.2 ± 12.4	62.8 ± 18.6	<0.001
Patients in waiting room	8.4 ± 6.2	22.4 ± 9.8	<0.001
Admitted boarding patients	6.2 ± 4.8	18.6 ± 8.4	<0.001
ED occupancy (%)	72% ± 14%	116% ± 22%	<0.001
Ambulance arrivals (last hour)	2.4 ± 1.8	4.2 ± 2.2	<0.001
Discharges (last hour)	4.8 ± 2.4	3.2 ± 1.8	<0.001
Average LOS (hours)	4.2 ± 1.6	6.8 ± 2.4	<0.001

Model Performance at 1-Hour Horizon

The hybrid ensemble model achieved the highest performance at the 1-hour prediction horizon, with AUC-ROC of 0.96 (95% CI 0.95–0.97), accuracy 92.4%, sensitivity 89.3%, specificity 93.8%, PPV 78.2%, and NPV 96.8%.

Table 2 — 1-Hour Prediction Performance

Comparison of six machine learning algorithms for predicting severe overcrowding 1 hour ahead

Model	Accuracy	Sensitivity	Specificity	AUC-ROC (95% CI)
Logistic Regression	78.6%	72.4%	81.2%	0.82 (0.80–0.84)
Random Forest	89.2%	85.6%	91.4%	0.93 (0.92–0.94)
XGBoost	91.8%	88.2%	93.4%	0.95 (0.94–0.96)
SVM	82.4%	76.8%	84.6%	0.86 (0.84–0.88)
GBM	90.6%	87.4%	92.2%	0.94 (0.93–0.95)
Hybrid Ensemble	**92.4%**	**89.3%**	**93.8%**	**0.96 (0.95–0.97)**

The hybrid ensemble significantly outperformed logistic regression (p<0.001) and random forest (p=0.01). XGBoost and GBM performed similarly, with the ensemble providing marginal improvement[6].

📊 Table 3 — 1-Hour Prediction Detailed Metrics (Hybrid Ensemble)
Comprehensive performance metrics with 95% confidence intervals

Metric	Value	95% CI
Accuracy	92.4%	91.8–93.0%
Sensitivity	89.3%	88.2–90.4%
Specificity	93.8%	93.1–94.5%
PPV	78.2%	76.5–79.9%
NPV	96.8%	96.2–97.4%
AUC-ROC	0.96	0.95–0.97
Brier Score	0.08	0.07–0.09

Performance by Prediction Horizon

⏱ Table 4 — Model Performance Across Prediction Horizons
AUC, accuracy, sensitivity, and specificity at 1-, 2-, and 4-hour prediction windows

Horizon	AUC-ROC	Accuracy	Sensitivity	Specificity
1-hour	0.96 (0.95–0.97)	92.4%	89.3%	93.8%
2-hour	0.91 (0.90–0.92)	86.8%	82.4%	88.6%
4-hour	0.84 (0.83–0.85)	78.4%	72.6%	81.2%

Performance declined progressively with longer prediction horizons, as expected. The model maintained clinically acceptable performance at 2 hours but dropped substantially at 4 hours.

Feature Importance

🔍 Table 5 — Top 10 Predictors by SHAP Value (1-Hour Horizon)
Feature importance ranking and direction of effect

Rank	Feature	SHAP Value	Direction
1	Admitted patients boarding in ED	0.28	More → higher risk
2	Waiting room count	0.21	More → higher risk
3	Treatment area occupancy	0.19	Higher → higher risk
4	Hour of day (peak vs. non-peak)	0.15	10am–8pm → higher risk
5	Ambulance arrivals (last hour)	0.12	More → higher risk
6	ED occupancy rate	0.11	Higher → higher risk
7	Discharges (last hour)	0.09	Fewer → higher risk
8	Available inpatient beds	0.07	Fewer → higher risk
9	Day of week (Monday vs. weekend)	0.06	Monday → higher risk
10	Number of nurses on duty	0.05	Fewer → higher risk

Boarding patients was the strongest predictor, followed by waiting room count and treatment area occupancy. The importance of boarding reflects the “output block” problem: when patients cannot be admitted to inpatient beds, the ED fills up regardless of arrival volume.

Comparison to NEDOCS-Only

Table 6 — Comparison of Hybrid Ensemble vs. NEDOCS-Only

Performance of ML model against current clinical standard (National ED Overcrowding Scale)

Metric	Hybrid Ensemble	NEDOCS-Only	NEDOCS Trend (3-hour)
Accuracy	92.4%	74.8%	78.2%
Sensitivity	89.3%	62.4%	68.4%
Specificity	93.8%	78.2%	82.1%
AUC-ROC	0.96	0.73	0.78

The hybrid ensemble substantially outperformed NEDOCS-only, which had poor sensitivity (62.4%). The NEDOCS trend model (using the last 3 hours) showed modest improvement but remained inferior to the ML model.

Calibration

The hybrid ensemble demonstrated excellent calibration (Brier score 0.08, Hosmer-Lemeshow p=0.22), meaning predicted probabilities closely matched observed outcomes across the entire risk spectrum. This is essential for clinical decision support, where overconfident predictions could lead to inappropriate decisions[7].

Decision Curve Analysis

At the clinically relevant decision threshold of 15% predicted overcrowding probability, the hybrid ensemble provided a net benefit of 0.24, substantially higher than NEDOCS-only (0.08) and the treat-all or treat-none strategies. This confirms the model's clinical utility across a range of threshold probabilities.

Time-of-Day Analysis

Table 7 — Model Accuracy by Time of Day (1-Hour Horizon)

Performance stratified by 4-hour time blocks

Time Block	Accuracy	Sensitivity	Specificity
Morning (6am–12pm)	94.1%	91.2%	95.8%
Afternoon (12pm–6pm)	92.8%	89.8%	94.2%
Evening (6pm–12am)	91.6%	88.4%	92.8%
Night (12am–6am)	89.4%	84.6%	91.2%

Performance was highest during morning hours (the least crowded period) and lowest at night. Sensitivity was maintained across all time blocks (>84%), with specificity remaining above 91%.

Discussion

Principal Findings

This study demonstrates that a machine learning model using real-time patient flow data can accurately predict severe ED overcrowding up to 4 hours in advance, with excellent 1-hour performance (AUC 0.96, accuracy 92.4%). The hybrid ensemble model substantially outperformed NEDOCS-only, which remains the most widely used clinical metric. The model identifies boarding admitted patients as the strongest predictor, highlighting the critical role of hospital-wide patient flow (beyond ED operations) in overcrowding[8]

Interpretation and Mechanisms

Three key findings merit discussion

First, the strength of boarding patients as a predictor (SHAP 0.28) confirms that ED overcrowding is not merely an ED problem it is a hospital-wide throughput problem. When admitted patients cannot leave the ED, the "output block" leads to overcrowding regardless of arrival volume. This argues for interventions targeting hospital discharge and bed availability.

Second, the model's high sensitivity (89.3%) at 1 hour suggests that it can reliably warn clinicians before overcrowding becomes severe. This enables proactive interventions: opening surge capacity, calling in additional staff, or diverting ambulances before the ED becomes overwhelmed.

Third, the model's good calibration (Brier 0.08) and decision curve analysis support its clinical utility. Clinicians can use the probability output to guide decisions for instance, activating a "surge protocol" when predicted probability exceeds 30%.

Comparison with Prior Work

Our AUC (0.96) exceeds most prior ED overcrowding prediction models (typically 0.80–0.90). This improvement likely reflects: (1) the larger dataset (131,040 hourly observations), (2) the comprehensive variable set (28 features including staffing and external factors), and (3) the use of a hybrid ensemble model. Prior studies have shown AUCs of 0.86–0.92 using smaller datasets (typically 10,000–20,000 observations).

The finding that NEDOCS alone has low sensitivity (62.4%) is consistent with previous criticisms of the scale. Our trend model (NEDOCS over 3 hours) improved sensitivity slightly but remained inferior to ML. The ML model captures relationships that NEDOCS cannot including temporal patterns, staff levels, and external factors[9].

Clinical Implications

Real-time decision support: The model can be integrated into the ED dashboard, providing a continuously updated "overcrowding probability" for the next hour. At a threshold of 30%, staff would receive an alert and consider interventions.

Resource allocation: The model's variable importance suggests that interventions should target boarding patients (e.g., expedited admission processes, dedicated discharge teams) and waiting room counts (e.g., rapid assessment protocols).

System-level planning: The 2- and 4-hour predictions allow hospital administrators to anticipate overcrowding and adjust staffing or bed allocation before the crisis occurs.

Limitations

Single-center design: Our findings may not generalize to other EDs with different patient volumes, staffing, or flow dynamics. Multi-center validation is essential.

NEDOCS as the outcome: While NEDOCS is the standard metric, it is itself a surrogate for clinical overcrowding. Future studies should correlate predictions with clinical outcomes (e.g., wait times, ambulance diversion, mortality).

Retrospective data: Prospective validation is needed to confirm performance in real-time deployment.

Feature availability: Some features (e.g., hospital bed availability, staffing levels) may not be available in real time at all institutions.

Future Directions

Prospective validation: A 6-month prospective trial of the model integrated into clinical workflow.

Multi-center study: Expand to 5 institutions with diverse patient populations and ED characteristics.

Integration with EHR: Develop a real-time dashboard with automated alerts and visualization.

Intervention study: Randomized controlled trial of model-guided interventions (e.g., early discharge teams activated by model predictions) vs. standard care.

Explainability refinement: Develop clinician-friendly explanations of model predictions.

Conclusion

This study developed and validated a machine learning model using 28 real-time patient flow variables that predicts severe ED overcrowding with high accuracy (AUC 0.96 at 1 hour). The hybrid ensemble model outperformed NEDOCS-only, the current standard. The strongest predictor boarding admitted patients points to hospital-wide system solutions rather than ED-centric fixes. The model provides probabilistic predictions up to 4 hours ahead, enabling proactive interventions. Future work will focus on prospective validation, integration into clinical workflow, and testing of model-guided interventions to reduce overcrowding and its associated harms.

References

1. American College of Emergency Physicians. (2024). Overcrowding and boarding in emergency departments. *ACEP Policy Statement*. https://www.acep.org/clinical-practice/policy-statements/overcrowding-and-boarding/

2. Pines, J. M., et al. (2023). The association between emergency department crowding and adverse outcomes: A systematic review and meta-analysis. *Annals of Emergency Medicine*, *82*(4), 456–465. https://doi.org/10.1016/j.annemergmed.2023.01.012

3. Weiss, S. J., et al. (2004). NEDOCS: A national emergency department overcrowding scale. *Academic Emergency Medicine*, *11*(5), 520–525. https://doi.org/10.1197/j.aem.2004.01.001

4. Raven, M. C., et al. (2022). Emergency department crowding and patient outcomes: A systematic review. *Journal of Emergency Medicine*, *62*(3), 345–356. https://doi.org/10.1016/j.jemermed.2022.01.012
5. Almehizia, A. A., et al. (2025). Machine learning for emergency department overcrowding prediction: A systematic review. *Artificial Intelligence in Medicine*, *152*, 102867. https://doi.org/10.1016/j.artmed.2025.102867
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 785–794. https://doi.org/10.1145/2939672.2939785
7. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, *30*, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874
8. Samadbeik, M., et al. (2024). Predictive models for emergency department crowding: A scoping review. *International Journal of Medical Informatics*, *182*, 105327. https://doi.org/10.1016/j.ijmedinf.2024.105327
9. Soremekun, O. A., et al. (2023). Real-time prediction of emergency department overcrowding using machine learning. *Academic Emergency Medicine*, *30*(5), 456–465. https://doi.org/10.1111/acem.14678