

Journal Pre-proofs

Forecasting emergency department overcrowding using the CEDOCS score – predictable, but not actionable?

Cornelius Born, Andreas Hein, Lilly Kämmerling, Luca Kittelmann, Martin Schmauch, Alexander Berger, Martin Gottwald, Nicola Friedhoff, Michael Dommasch, Helmut Krcmar

PII: S1386-5056(26)00346-1
DOI: <https://doi.org/10.1016/j.ijmedinf.2026.106606>
Reference: IJB 106606

To appear in: *International Journal of Medical Informatics*

Received Date: 2 June 2026
Revised Date: 9 July 2026
Accepted Date: 9 July 2026

Please cite this article as: C. Born, A. Hein, L. Kämmerling, L. Kittelmann, M. Schmauch, A. Berger, M. Gottwald, N. Friedhoff, M. Dommasch, H. Krcmar, Forecasting emergency department overcrowding using the CEDOCS score – predictable, but not actionable?, *International Journal of Medical Informatics* (2026), doi: <https://doi.org/10.1016/j.ijmedinf.2026.106606>

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 The Author(s). Published by Elsevier B.V.



An Official Journal of the International Medical Informatics Association
and the European Federation for Medical Informatics

Declaration of competing interests

The authors have no competing interests to report.

Funding

This work was supported by the German Ministry of Research, Technology and Space under grant number 16SV9048.

Highlights

- An explainable LightGBM model accurately forecasts ED overcrowding (CEDOCS) up to three hours ahead using data from more than 300,000 patient visits across two hospitals.
- SHAP analysis identifies clinically plausible drivers of overcrowding, including current CEDOCS, time of day, and ICU ambulance diversion.
- During a six-month intervention study, actions taken based on the predictions were associated with perceived tool accuracy, but not with CEDOCS or prediction error.
- These findings suggest that accurate forecasting of occupancy-based metrics alone may be insufficient for actionable clinical decision support.
- Instead, we recommend moving away from occupancy-based metrics towards models that directly capture patient-level resource demands to better reflect frontline workload.

Forecasting Emergency Department Overcrowding Using the CEDOCS Score – Predictable, but Not Actionable?

Abstract

Objective: This study aims to develop an explainable machine learning (ML) model to predict emergency department (ED) overcrowding using the CEDOCS score and evaluate its perceived actionability in clinical practice.

Methods: We retrospectively analyzed all 300,000 visits to two German EDs between 2019 and 2023. We derived electronic health records (EHR), calendar and ambulance features to predict the CEDOCS score in the next three hours. We used SHAP values to analyze feature attributions. In a six-month intervention study, we introduced the ML model at one of the EDs and assessed perceived tool accuracy, actions taken based on the system's predictions, and stress.

Results: The best-performing models achieved RMSEs of 11.76, 15.92, and 18.34 at Hospital 1 and 16.04, 21.38, and 22.91 at Hospital 2 for the one-, two- and three-hour forecasts, respectively. The current CEDOCS score, time of day, and ambulance diversion by the intensive care units had the strongest correlation with overcrowding. In the intervention study, clinician stress showed a small-to-moderate association with CEDOCS ($r = .24$, $p = .03$). Actions taken were unrelated to CEDOCS ($p = .18$) and model error ($p = .21$) but positively associated with perceived tool accuracy ($r = .38$, $p < .01$).

Conclusion: Our ML model demonstrates good predictive performance in forecasting ED overcrowding using the CEDOCS score. The SHAP analysis aligns with known causes of overcrowding. However, our intervention study suggests that accurate prediction of CEDOCS

alone may be insufficient for actionable clinical decision support, as occupancy-based metrics may not fully capture the frontline resource demands that drive clinicians' operational decisions.

Authors and Affiliations

Cornelius Born^{1*}, Andreas Hein², Lilly Kämmerling¹, Luca Kittelmann¹, Martin Schmauch¹, Alexander Berger¹, Martin Gottwald³, Nicola Friedhoff⁴, Michael Dommasch⁵, Helmut Krcmar¹

1. Technical University of Munich, Munich, Germany, School of Computation, Information and Technology
2. University of St. Gallen, St. Gallen, Switzerland, Institute of Information Systems and Digital Business
3. fortiss GmbH, Munich, Germany, Department of Machine Learning
4. AGAPLESION Evangelisches Krankenhaus Mittelhessen, Gießen, Germany, Clinic for Acute and Emergency Medicine
5. TUM Klinikum Rechts der Isar Hospital, Munich, Germany, Emergency Department

* Corresponding author: Cornelius Born (cornelius.born@tum.de)

Keywords

Emergency Department, Overcrowding, CEDOCS, Artificial Intelligence, Machine Learning, Clinical Decision Support Systems, Clinical Decision-Making

1. Introduction

Emergency department (ED) overcrowding, a situation where the demand for emergency services surpasses the available resources to provide timely care¹⁻³, is a worldwide public health problem that threatens patient safety^{4,5}. Overcrowding has been linked to adverse clinical outcomes, including increased mortality^{6,7}, as well as heightened strain on clinical staff, contributing to fatigue and burnout⁸⁻¹³. Since its recognition as a structural problem in the 1980s¹⁴, efforts to alleviate overcrowding have largely relied on increasing internal resources, e.g., hiring additional ED staff. However, these resource-driven solutions have proven insufficient amid mounting demand, staff shortages, and financial constraints^{5,15}.

With the widespread adoption of electronic health records (EHRs) that standardize clinical workflows data, hospitals are increasingly leveraging machine learning (ML) to improve healthcare operations¹⁶. In the ED context, early prediction of overcrowding could enable physicians and staffing managers to initiate targeted interventions, such as requesting support from inpatient staff, signaling ambulance diversion, and reserving inpatient beds^{17,18}. Prior studies have primarily focused on forecasting patient arrival volumes¹⁹⁻²¹ (see Supplementary Material for literature review). However, patient volume alone does not capture contextual factors such as current ED utilization and inpatient bed availability^{5,15,22}, which determine whether an incoming surge will result in overcrowding.

To address this gap, we build on the Community Emergency Department Overcrowding Scale (CEDOCS) score²³, a validated and widely adopted metric, which, along with its predecessor, has been integrated into numerous hospital information systems^{24,25}. CEDOCS quantifies the degree of overcrowding on a standardized scale (0-200) based on patient volume, acuity, treatment progression, and bed capacity. Compared with patient volume alone, this approach more closely reflects the operational logic of EDs, which requires continuous balancing of demand and supply of emergency care services²². While CEDOCS has proven useful for retrospectively documenting substantial operational burden^{26,27}, it remains uncertain whether it captures the aspects of frontline workload that are most relevant for operational decision-making. Consequently, it is unclear whether accurately predicting CEDOCS can provide clinically actionable decision support.

Therefore, this study develops an explainable ML-based early warning system for ED overcrowding and evaluates its implementation in an intervention study. For two German hospitals, i.e., Hospital 1 and Hospital 2, we augment EHR data with calendar (e.g., day of the week) and ambulance (e.g., diversion of the intensive care unit (ICU)) features. In a six-month intervention study in Hospital 2, we integrated the ML model into clinical workflows and assessed perceived tool accuracy, actions taken based on the system's predictions, and clinician stress to evaluate the model's perceived actionability. While the broader psychological evaluation of the intervention is reported separately, the present work focuses on the development and evaluation of the ML-based overcrowding prediction system.

2. Methods

2.1. Data Acquisition

This study identified all patients admitted to two German hospitals from 1 January 2019 to 31 December 2023: Hospital 1, a large urban university hospital in Bavaria with approximately 40,000 annual ED visits, and Hospital 2, a medium-sized rural hospital in Hesse with around 20,000 annual ED visits. We extracted the mode of arrival, patient age, triage urgency, responsible specialty, and discharge type from the electronic health records (EHR). In addition,

we extracted the timestamps for arrival, triage start and end, doctor arrival time, and discharge. For Hospital 2, additional timestamps for the last physician contact and treatment completion were available.

Calendar variables were derived from public datasets, including the Python holidays library, official school holiday schedules, and government-issued COVID-19 lockdowns.

We incorporated ambulance data from the IVENA system²⁸, a nationally adopted platform that facilitates coordination between hospitals and ambulance services. IVENA provides information on the assignment of ambulance patients to hospitals as well as departmental diversion statuses, whereby hospital units such as the ED or ICU can temporarily indicate limited capacity to ambulance services.

2.2. Data Preprocessing

We standardized patient-level characteristics, e.g., converting age into discrete groups. We cleaned the EHR data by removing duplicate entries and ensuring timestamp consistency to reflect the expected sequence of ED processes, e.g., triage preceding discharge. We excluded patients with a length of stay exceeding thresholds defined in consultation with lead physicians (72 h for Hospital 1; 48 h for Hospital 2), as they reflected documentation errors. In Hospital 1, timestamps for the first physician assessment were missing in 10% of cases. In Hospital 2, 5% of discharge times were missing. We imputed these values using the average duration from the previous month stratified by triage urgency. Hospital 1's ED includes a monitoring ward. As the transfer time is not captured in the EHR, we approximated it using the rolling mean of ED length of stay over the preceding hours. The resulting dataset enabled us to reconstruct patient flow through the ED and generate a continuous time series at 15-minute intervals.

We then computed the CEDOCS score for each interval based on its original formula provided in Table 1. Table 2 details the mapping between CEDOCS components and variables derivable from German hospital EHRs.

For calendar and ambulance data, we eliminated duplicates, removed outliers, and converted the data into a uniform timeline with 15-minute intervals.

Table 1: The formula of the CEDOCS score.

$$\begin{aligned}
 & -30.39 + \\
 & 3.00 \cdot A + 0.53 \cdot B + 1.16 \cdot C + 20.66 \cdot D + 0.00126 \cdot E - \\
 & (1.11e - 12) \cdot pmax(E - 18811, 0)^3 + \\
 & (8.23e - 12) \cdot pmax(E - 43012, 0)^3 - \\
 & (8.20e - 12) \cdot pmax(E - 49466, 0)^3 + \\
 & (1.08e - 12) \cdot pmax(E - 67273, 0)^3
 \end{aligned}$$

where

- A is the number of critical care patients,
- B is the longest boarding time
- C is the number of patients in the waiting room,
- D is the ratio of current ED patients to total ED beds, and
- E is the number of ED visits per year.

Table 2: Proposed calculation of the CEDOCS score based on available EHR timestamps.

CEDOCS Variable		Calculation	Explanation
<i>A</i>	Number of critical care patients	Sum of MTS ^a 1 and 2 patients awaiting or in treatment	MTS ^a 1 and 2 patients who are in the ED at the time of measurement
<i>B</i>	Longest boarding time	Maximum waiting time of patients awaiting admission	Longest duration since the first physician assessment among all patients due to be admitted to a ward
<i>C</i>	Number of patients in the waiting room	Sum of all patients in waiting room	Patients who are between admission and their first physician assessment
<i>D</i>	Ratio of current ED patients to total ED beds	Proportion of the sum of all patients in treatment to the count of beds	Patients who are between their first physician assessment and discharge divided by the fixed number of ED beds
<i>E</i>	Number of ED visits per year	Sum of patients in the previous year	Patients admitted to the ED in the previous calendar year (static annual value)

^a Manchester Triage System (MTS); interchangeable with any of the common triage systems.

2.3. Feature Engineering

Exploratory time series data analysis of the CEDOCS score using Autocorrelation Function plots revealed strong daily and weekly seasonality. We thus included the 24h- and 1-week-lagged values.

Temporal variables included hour, day of week, month, and day of year. Public holidays, school breaks, and COVID-19 lockdown periods were encoded as binary variables.

We aggregated the number of patients assigned through the ambulance over the last hour, incorporating characteristics such as age and urgency, ranging from SK1 (emergency), SK2 (inpatient), to SK3 (outpatient). Ambulance diversions are recorded at the department and urgency level. In addition to the ED, we included the ICUs and CPU. Overcrowding in these

departments can delay the transfer of critically ill patients from the ED, forcing them to remain there, i.e., ED boarding^{15,22}. Diversion events in IVENA are always registered with predefined start and end timestamps, so we calculate two features for every 15-minute interval: how long the department had already been closed (“closed since”) and how much longer the diversion is going to last (“closed until”).

The complete feature list is available in the Supplementary Material.

2.4. Model Building

Target variable

Following clinical consultation and prior studies^{29–31}, we defined the target variables as the level of ED overcrowding per the CEDOCS score, at one-, two-, and three-hour horizons. This prediction window reflects a clinically meaningful timeframe deemed sufficient for initiating operational countermeasures.

Data split

To reflect the temporal nature of the prediction task and prevent data leakage, we applied a custom time series split. The dataset was divided into training (January 2019–December 2022), validation (January–June 2023), and test (July–December 2023) sets using an 8:1:1 ratio. As the training period included the COVID-19 pandemic, we repeated the analysis using an abridged dataset beginning on 1 June 2021.

Model Development

The datasets were highly imbalanced toward non-overcrowding. To address this, we applied sample weighting during model training, assigning progressively higher weights to observations with increasing CEDOCS scores, and applied undersampling using the imblearn package with hospital-specific overcrowding thresholds (20% ratio on the training set). We derived the thresholds from interviews with lead physicians, resulting in $\text{CEDOCS} \geq 140$ for Hospital 1 and $\text{CEDOCS} \geq 100$ for Hospital 2.

Based on the literature^{32–35}, we performed feature selection separately for each hospital on the training data using a five-fold time series split with a one-year rolling training window. Within each split, features underwent sequential filtering: univariable testing ($p < .05$), multicollinearity assessment ($\text{VIF} < 10$), LASSO regression (non-zero coefficients), and multivariable regression ($p < .05$). Features selected in at least three of the five splits were retained.

We tested a range of regression models commonly used for tabular clinical data: linear regression, Random Forest, XGBoost, LightGBM, CatBoost, and NGBoost, as well as multilayer perceptrons with varying architectures. We selected the model achieving the lowest root mean squared error (RMSE) and RMSE above the overcrowding threshold on the validation set across forecasting horizons.

Model Evaluation and Interpretation

We assessed model performance using R^2 , mean absolute error (MAE), RMSE, and the ratio between RMSE and the standard deviation of the target. To evaluate predictive performance during overcrowding periods, we calculated the RMSE for observations above the overcrowding threshold and used Bland–Altman plots to assess agreement across different acuity states. We compared three model configurations: a baseline model using the last available CEDOCS value, a model using EHR and calendar features, and a final model additionally incorporating ambulance-related features. We grouped calendar variables with the EHR configuration to isolate the impact of an IVENA interface. We evaluated feature importance with SHAP^{36–38}.

2.5. Intervention Study

We implemented the best-performing model in Hospital 2 in a six-month intervention study, approved by an ethics committee. During the intervention period, the ML prototype (see Supplementary Material) was displayed on a central ED screen. Participants completed a biweekly questionnaire assessing perceived tool accuracy, actions taken based on the predictions, and experienced stress during their shifts. Perceived accuracy was measured using a single item ("The system's predictions were very accurate.") on a 5-point Likert scale. Actions taken based on the predictions were assessed using a single item ("We took actions based on the system's predictions.") on a 5-point Likert scale. Stress was measured using a 4-item scale based on Motowidlo et al.³⁹ (Cronbach's $\alpha = .90$).

We recruited 16 of the 22 ED staff members, comprising seven physicians and nine nurses. Over the six months, we collected 80 questionnaires for specific shifts. For each shift, we extracted the calculated and predicted CEDOCS scores and the time spent in ED ambulance diversion according to IVENA.

3. Results

3.1. Baseline Characteristics of Study Population

Between 2019 and 2023, a total of 216,127 patients were admitted to the ED at Hospital 1 and 94,900 patients at Hospital 2. We excluded 8,985 patients from Hospital 1, primarily because they received external treatment. These patients were referred to treatment outside the ED immediately after triage, and thus have minimal impact on overcrowding. For Hospital 2, we excluded 287 patients, primarily due to inaccurate timestamps (see Figure 1). The patient demographics are reported in the Supplementary Material.

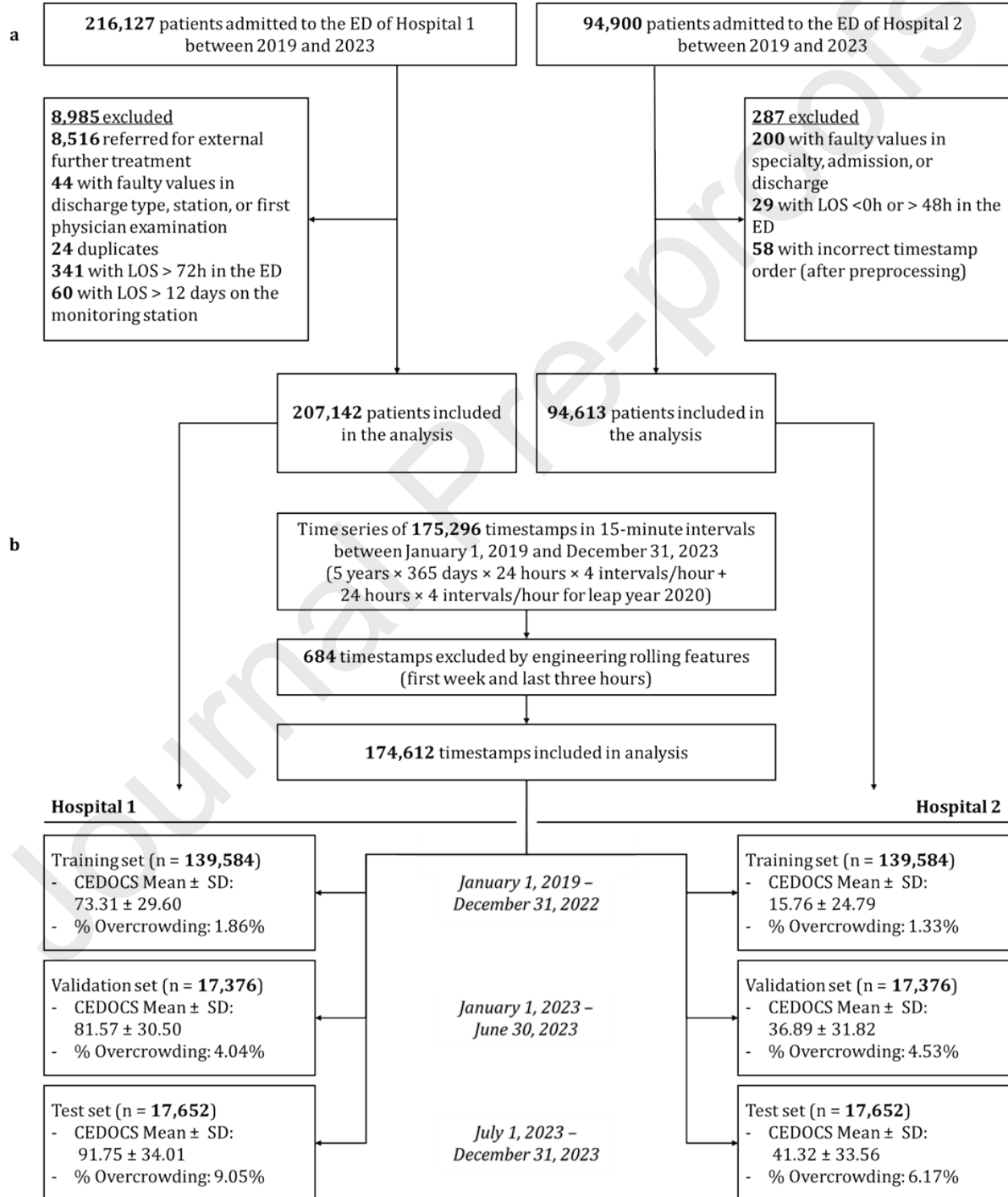


Figure 1: (a) Patient exclusion criteria. LOS: length of stay. (b) Split of timeline into training, validation, and test sets. Overcrowding defined as CEDOCS ≥ 140 for Hospital 1 and CEDOCS ≥ 100 for Hospital 2.

The dataset is then split into training (2019–2023), validation (January – June 2023), and test (July – December 2023) sets. The average CEDOCS scores and overcrowding proportion increases across sets, e.g., rising from 1.86% in training to 9.05% in the test set for Hospital 1.

3.2. Model Performance

Feature selection resulted in 15 predictors for Hospital 1 and 16 predictors for Hospital 2. The final feature sets are provided in the Supplementary Material.

LightGBM was the best-performing model on the validation set; the performance of the remaining ML models is reported in the Supplementary Material. Table 3 presents the performance of the three model configurations. For Hospital 2, the inclusion of ambulance-related variables improved predictive performance, whereas for Hospital 1 the performance impact was marginally negative. We retained the *+Ambulance* configuration for both hospitals in the final analyses for comparability. The final model achieved an RMSE of 11.76, 15.92, and 18.34 (16.04, 21.38, 22.91) in the one-, two- and three-hour forecast (Hospital 2 in brackets).

Table 3: Model performance of the best-performing model (LightGBM). Best results on validation set bold.

Outcomes		Hospital 1			
		Validation			Test
		CEDOCS	+ EHR + Calendar	+ Ambulance	+ Ambulance
R ² ↑	1h	0.83	0.85	0.85	0.86
	2h	0.61	0.74	0.73	0.76
	3h	0.34	0.65	0.64	0.68
MAE ↓	1h	9.64	8.81	8.89	9.49
	2h	14.93	11.99	12.12	12.87
	3h	19.63	13.88	14.07	14.98
RMSE ↓	1h	12.52	11.65	11.76	12.52
	2h	19.04	15.67	15.92	16.75
	3h	24.79	18.05	18.34	19.36

RMSE – Over-crowding ↓	1h	16.25	12.54	12.93	13.25
	2h	25.47	17.67	17.15	19.17
	3h	35.47	20.98	21.92	22.88
RMSE/SD Ratio ↓	1h	0.41	0.38	0.39	0.37
	2h	0.62	0.51	0.52	0.49
	3h	0.81	0.59	0.60	0.57
Hospital 2					
Validation				Test	
		CEDOCS	+ EHR + Calendar	+ Ambulance	+ Ambulance
R ² ↑	1h	0.74	0.74	0.75	0.74
	2h	0.43	0.52	0.55	0.56
	3h	0.10	0.35	0.48	0.46
MAE ↓	1h	11.50	11.49	11.39	12.15
	2h	17.41	15.93	15.53	16.40
	3h	22.18	18.58	16.80	18.25
RMSE ↓	1h	16.23	16.20	16.04	16.96
	2h	24.01	22.06	21.38	22.33
	3h	30.24	25.63	22.91	24.75
RMSE – Over-crowding ↓	1h	26.89	19.10	18.79	18.44

	2h	42.86	24.08	24.61	26.02
	3h	59.57	31.95	32.62	35.04
RMSE/SD Ratio ↓	1h	0.51	0.51	0.50	0.51
	2h	0.75	0.69	0.67	0.67
	3h	0.95	0.81	0.72	0.74

Figure 2 presents Bland–Altman analyses of the final model, which showed similar patterns across all forecasting horizons. Differences between predicted and observed scores were centered around zero. However, the model showed a tendency to underestimate severe overcrowding, e.g., with mean errors of -5.36 (-24.53) in Level 5 of the one-hour forecast (Hospital 2 in brackets).

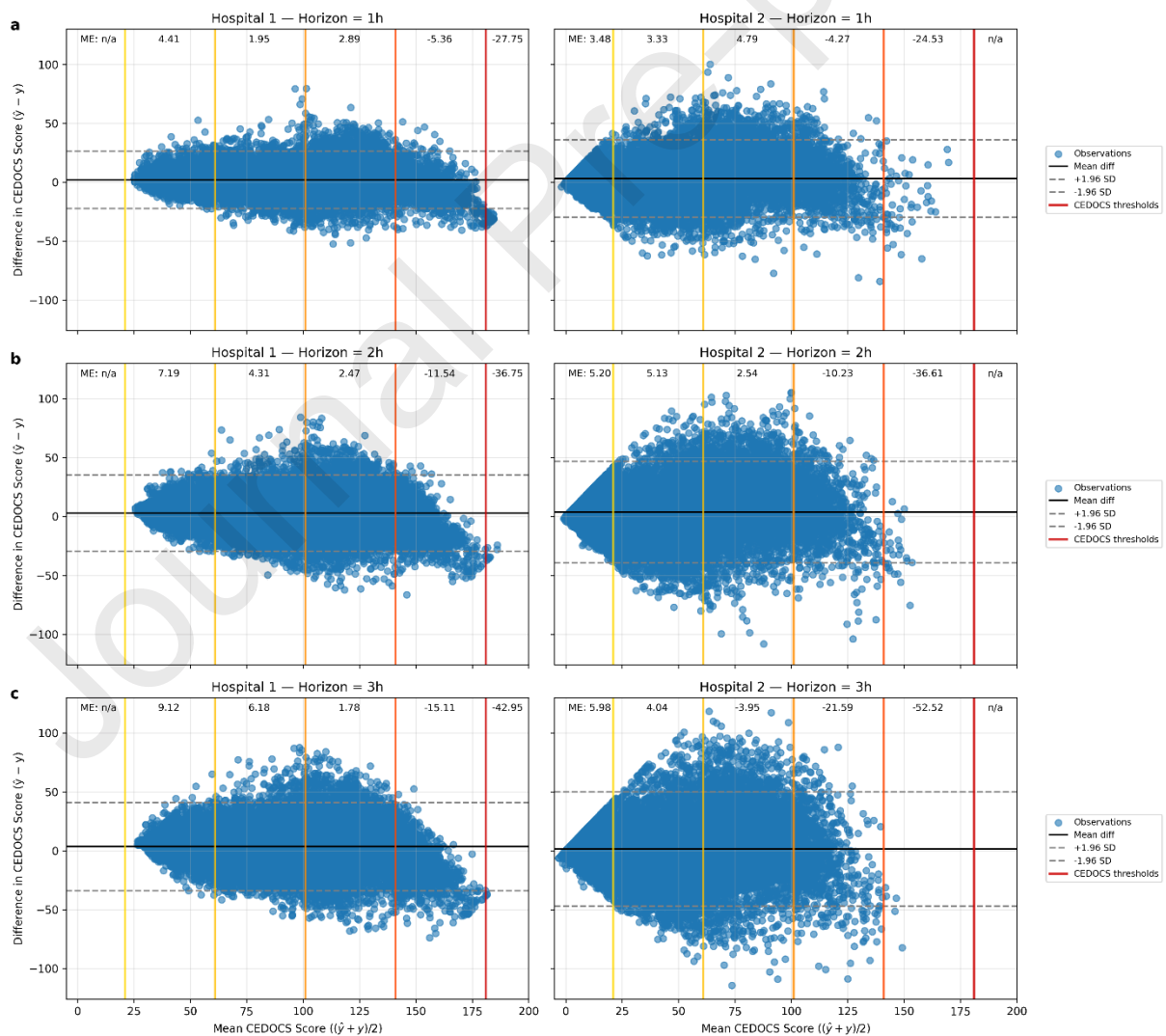


Figure 2: Bland–Altman plots of the final (+Ambulance) model on the test set for the one-hour (a), two-hour (b), and three-hour (c) forecasting horizons. Vertical colored lines indicate the CEDOCS overcrowding thresholds. The mean error

(ME) is reported separately for each CEDOCS level at the top of each panel to quantify potential directional bias across overcrowding severity.

Excluding COVID-19 data resulted in reduced model performance on the evaluation set; therefore, all available data were retained.

3.3. Model Interpretations

Figure 3 shows the ten most impactful features in the *+Ambulance* configuration for the two-hour forecasting horizon. The analysis identified the current CEDOCS score and recurring features, including its 24-hour and one-week lag values, hour of day, and day of the week, as the dominant predictors of overcrowding across both hospitals. Notably, ambulance diversion statuses of ED and non-ED emergency care departments were consistently among the most influential features. In Hospital 1, the COVID-19 lockdowns were associated with lower predicted CEDOCS values, reflecting CEDOCS's limitation in capturing operational strain, such as infection-control measures that increased workload despite reduced patient volumes.

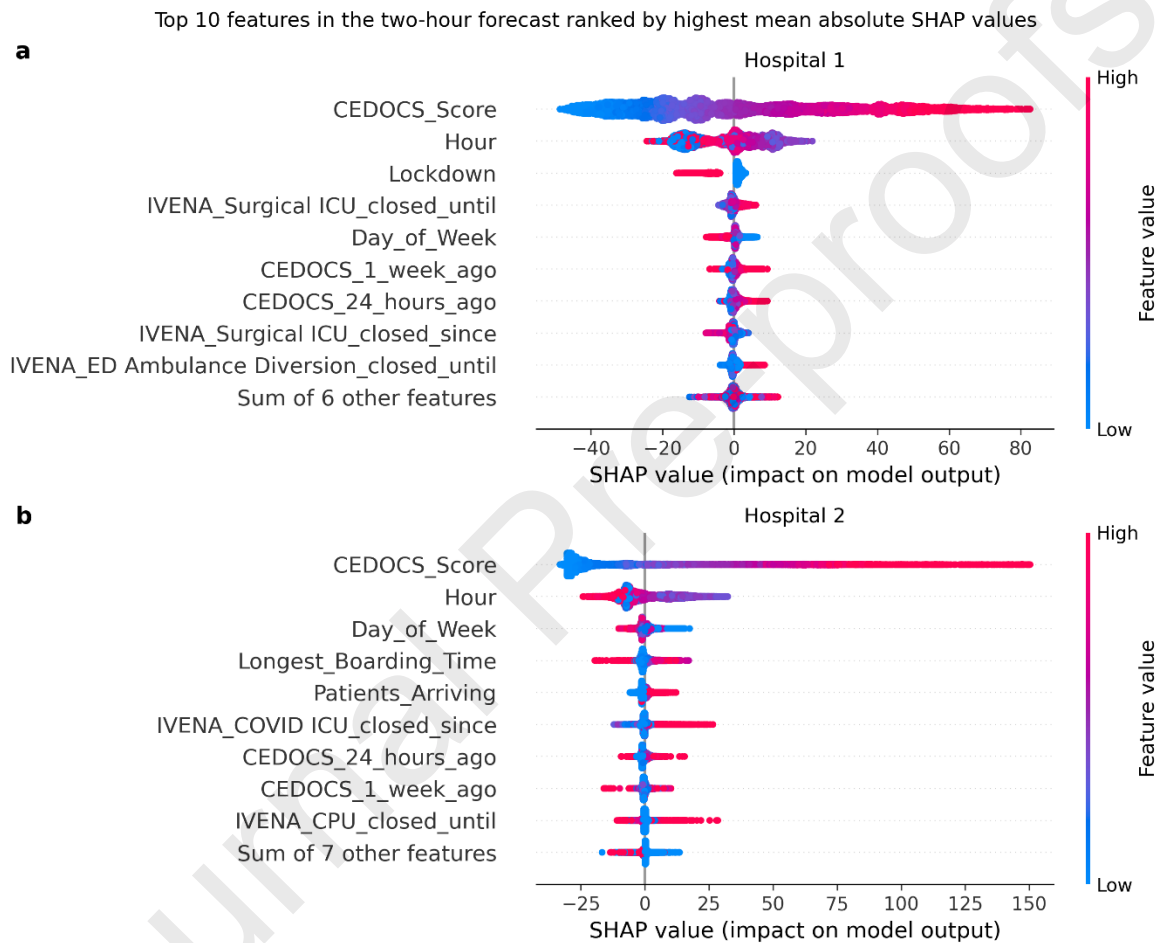


Figure 3: Top ten features ranked by highest mean absolute SHAP values for (a) Hospital 1 and (b) Hospital 2 for the two-hour forecast in the training set. The SHAP plots for the one- and three-hour forecasting horizons are provided in the Supplementary Material.

3.4. Intervention Study

In a six-month pilot intervention study, we implemented the *+EHR + Calendar* configuration at Hospital 2, as the modest performance gains achieved by incorporating IVENA ambulance data did not justify the additional monetary and technical effort required for live integration of the external interface. During this period, the model achieved RMSE values of 19.29 (1-hour), 23.23 (2-hour), and 24.93 (3-hour forecast), representing a slight decrease compared to the test set.

Table 4 summarizes the correlations between the standardized metrics for ED overcrowding with the questionnaire-based assessments. Clinician stress showed small-to-moderate associations with average CEDOCS scores during the shift (Pearson's $r = .24$, 95% CI [.02, .44], $p = .03$) and ED ambulance diversion ($r = .25$, CI [.03, .44], $p = .03$). Actions taken based on the system's predictions were not significantly associated with CEDOCS scores ($p = .18$) or model error ($p = .21$), but showed a positive association with perceived tool accuracy ($r = .38$, CI [.17, .55], $p < .01$). However, perceived tool accuracy was not significantly associated with the model error ($p = .60$).

Table 4: Pearson Correlations of Staff Stress with Overcrowding according to CEDOCS and Ambulance Diversion.
* $p < .05$, ** $p < .01$.

Variable	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.
1. Experienced Stress	3.51	1.04						
2. Average CEDOCS Score	72.03	32.10	.24*					
3. % Shift in Overcrowding	26.40	31.54	.25*	.92**				
4. % Shift in Ambulance Diversion	20.89	22.91	.25*	.46**	.45**			
5. Perceived Accuracy	2.95	1.01	-.07	-.02	.04	-.21		
6. RMSE of the 2h-Forecast	24.78	12.56	.08	.70**	.72**	.29**	-.06	
7. Actions based on Predictions	1.82	0.81	-.05	-.15	-.14	-.17	.38**	-.14

4. Discussion

This study demonstrates the feasibility of an explainable ML model to forecast ED overcrowding, while also highlighting the potentially limited actionability of CEDOCS. By integrating EHR, calendar and ambulance data from two hospitals, the model achieved RMSEs of 11.76, 15.92, and 18.34 (16.04, 21.38, 22.91) for one-, two- and three-hour forecasts (Hospital 2 in brackets). In an intervention study at Hospital 2, clinician stress showed a small-to-moderate association with CEDOCS ($r = .24$, $p = .03$). Actions taken based on the system's predictions were unrelated to CEDOCS ($p = .18$) and model error ($p = .21$), but were positively associated with perceived tool accuracy ($r = .38$, $p < .01$).

We extend the literature by predicting a capacity-aware overcrowding score rather than patient volume alone, thereby establishing a performance baseline. Only two previous studies have predicted established overcrowding scores^{29,30}. Both framed the task as binary classification and achieved high sensitivity (0.98²⁹; 0.99³⁰) but low specificity (0.10²⁹; 0.43³⁰), limiting their clinical utility through false alarms and alert fatigue^{40,41}. In contrast, our continuous predictions generally remained within the same or adjacent CEDOCS overcrowding category, given the score range (0–200) and category widths (20–40 points). However, our model tended to underestimate severe overcrowding, particularly in Hospital 2, with mean errors increasing from 2.89 (–4.27) in CEDOCS Level 4 to –5.36 (–24.53) in Level 5 (Hospital 2 in brackets). Because these peak overcrowding events are operationally most critical, improving prediction during severe overcrowding should be a priority for future research.

Our feature importance analysis reveals a strong alignment with established causes of ED overcrowding. To our knowledge, this is the first study to demonstrate these associations through ML-based feature attribution methods. Along the established input-throughput-output model for ED overcrowding²², indicators of human activity (e.g., hour of the day) emerged as important predictors for input, corroborating previous findings⁴². Staffing constraints and ED bed availability, the primary bottleneck in the throughput component^{5,15,43}, are only indirectly captured in CEDOCS. ICU ambulance diversion was among the most influential predictors across both hospitals, reinforcing the central role of access block in ED overcrowding^{14,15,43–47}. Future work could further improve prediction through integration of ambulance diversion statuses of nearby EDs, staffing and more granular ward-level bed availability data.

However, our intervention study suggests that accurately predicting CEDOCS alone may be insufficient to achieve actionability in clinical practice. Consistent with prior work^{48–51}, clinician stress was only weakly associated with CEDOCS, indicating that occupancy-based metrics capture only part of frontline workload. Likewise, reported actions based on the system's predictions were unrelated to CEDOCS or model error, but were positively associated with perceived system accuracy. Together, these findings suggest that statistical accuracy alone is insufficient for operational use. If CEDOCS only partially reflects clinicians' experienced workload, then even highly accurate forecasts may have limited operational value. This may be because CEDOCS primarily captures occupancy, whereas frontline decisions are driven by the clinical demands competing for limited staff time and attention^{52–54}. For example, older and multimorbid patients often require substantially more diagnostics, monitoring, and nursing care^{55–57}.

This study has several limitations. First, the intervention study was conducted at a single center and included only 80 questionnaires, limiting statistical power. Based on the observed effect size, approximately 259 evaluated shifts would be required to achieve 80% power for the observed stress–CEDOCS association. However, previous studies have similarly reported only limited agreement between NEDOCS/CEDOCS-type measures and subjective assessments^{48–51}. Actions based on the system's predictions were measured using a single self-report item rather than objective system use or patient outcomes. Second, although retrospective data from two hospitals were available, we intentionally developed site-specific models to account for

structural differences in infrastructure, e.g., the COVID ICU at Hospital 2, and workflows, e.g., the monitoring station of Hospital 1. Consequently, deployment at a new hospital requires local training data, limiting generalizability. Third, several methodological decisions may affect model interpretation. The CEDOCS score was calculated using the original US-derived coefficients. Although this facilitates comparison with previous work, the lack of local recalibration may introduce systematic bias. Likewise, the sample-weighting strategy may have influenced continuous performance metrics. Finally, the intervention evaluated the *+EHR + Calendar* configuration rather than the best-performing *+Ambulance* model because real-time ambulance data were unavailable during deployment. Consequently, their prospective benefit remains to be demonstrated.

In summary, this study developed and validated an explainable ML approach to forecasting ED overcrowding using the CEDOCS score. By augmenting EHR with calendar and ambulance data, the model delivers accurate predictions within a three-hour window. Our SHAP analysis reveals that the model's feature attributions align with known causes of overcrowding. However, our intervention findings suggest that accurate prediction of occupancy-based overcrowding metrics do not necessarily translate into clinical actionability.

Data availability

The datasets analyzed in this research are not openly available due to reasons of sensitivity.

Code availability

The code used in this research is available at <https://github.com/overcrowdingprediction/ed-overcrowding-prediction>

Declaration of generative AI use

We used generative AI tools (in particular ChatGPT, DeepL, and Grammarly) to assist the preparation of this publication. These tools were employed only to improve language and readability. The authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Affleck, A., Parks, P., Drummond, A., Rowe, B. H. & Ovens, H. J. Emergency department overcrowding and access block. *Canadian Journal of Emergency Medicine* **15**, 359–370 (2013).
2. Bernstein, S. L. *et al.* The Effect of Emergency Department Crowding on Clinically Oriented Outcomes. *Academic Emergency Medicine* **16**, 1–10 (2009).
3. Javidan, A. P., Hansen, K., Higginson, I., Jones, P. & Lang, E. The International Federation for Emergency Medicine report on emergency department crowding and access block: A brief summary. *Emerg Med J* **38**, 245–246 (2021).
4. Pines, J. M. & Griffey, R. T. What We Have Learned From a Decade of ED Crowding Research. *Academic Emergency Medicine* **22**, 985–987 (2015).
5. Morley, C., Unwin, M., Peterson, G. M., Stankovich, J. & Kinsman, L. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLOS ONE* **13**, (2018).
6. McKenna, P. *et al.* Emergency department and hospital crowding: causes, consequences, and cures. *Clin Exp Emerg Med* **6**, 189–195 (2019).
7. Rasouli, H. R. *et al.* Outcomes of Crowding in Emergency Departments; a Systematic Review. *Arch Acad Emerg Med* **7**, e52 (2019).
8. Baugh, J. J., Raja, A. S. & Takayesu, J. K. Help us help you: Engaging emergency physicians to identify organizational strategies to reduce burnout. *Western Journal of Emergency Medicine* **22**, 696–701 (2021).
9. Bond, K. *et al.* Frequency, determinants and impact of overcrowding in emergency departments in Canada: a national survey. *Healthcare Quarterly* **10**, 32–40 (2007).
10. de Wit, K. *et al.* A Longitudinal Survey on Canadian Emergency Physician Burnout. *Annals of Emergency Medicine* <https://doi.org/10.1016/j.annemergmed.2024.01.009> (2024) doi:10.1016/j.annemergmed.2024.01.009.
11. Johnson, T. J. *et al.* The Impact of Cognitive Stressors in the Emergency Department on Physician Implicit Racial Bias. *Academic Emergency Medicine* **23**, 297–305 (2016).
12. Nugus, P. *et al.* Work pressure and patient flow management in the emergency department: Findings from an ethnographic study. *Academic Emergency Medicine* **18**, 1045–1052 (2011).
13. Shanafelt, T. D. *et al.* Changes in Burnout and Satisfaction With Work-Life Integration in Physicians During the First 2 Years of the COVID-19 Pandemic. *Mayo Clinic Proceedings* **97**, 2248–2258 (2022).
14. Kenny, J. F., Chang, B. C. & Hemmert, K. C. Factors Affecting Emergency Department Crowding. *Emerg Med Clin North Am* **38**, 573–587 (2020).
15. Hoot, N. R. & Aronsky, D. Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine* **52**, 126–136 (2008).
16. Pianykh, O. S. *et al.* Improving healthcare operations management with machine learning. *Nat Mach Intell* **2**, 266–273 (2020).

17. Gul, M. & Celik, E. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Systems* **9**, 263–284 (2020).
18. Federal Ministry of Health. Germany's Digitalisation Strategy for Health and Care. (2023).
19. Jones, S. S. *et al.* Forecasting Daily Patient Volumes in the Emergency Department. *Academic Emergency Medicine* **15**, 159–170 (2008).
20. Ekström, A., Kurland, L., Farrokhnia, N., Castrén, M. & Nordberg, M. Forecasting Emergency Department Visits Using Internet Data. *Annals of Emergency Medicine* **65**, 436–442.e1 (2015).
21. Xu, M., Wong, T. C. & Chin, K. S. Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using artificial neural network. *Decision Support Systems* **54**, 1488–1498 (2013).
22. Asplin, B. R. *et al.* A conceptual model of emergency department crowding. *Ann Emerg Med* **42**, 173–180 (2003).
23. Weiss, S. J., Rogers, D. B., Maas, F., Ernst, A. A. & Nick, T. G. Evaluating community ED crowding: the Community ED Overcrowding Scale study. *The American Journal of Emergency Medicine* **32**, 1357–1363 (2014).
24. Dedalus. ORBIS Notaufnahme. <https://www.dedalus.com/dach/de/our-offer/products/orbis-notaufnahme/> (2023).
25. NEDOCS. Epic HL7 Interface. <https://www.nedocs.org/News/Article/Epic-HL7-Interface> (2015).
26. Pin, M., Hüfner, A., Dormann, H., Jerusalem, K. & Dodt, C. Notfallkrankenhäuser: Massive Belastungssituationen. *Deutsches Ärzteblatt* **119**, 1392–1394 (2022).
27. Hüfner, A. *et al.* German Society for Emergency Medicine (DGINA) statement on overload situations in emergency departments: Part 1. *Notf. Rettungsmed.* **28**, 171–179 (2025).
28. mainis IT-Service GmbH. IVENA eHealth - Interdisziplinärer Versorgungsnachweis. *Interdisziplinärer Versorgungsnachweis* <https://www.ivena.de/page.php?k1=main&k2=index>.
29. Davis, Z., Zobel, C. W., Khansa, L. & Glick, R. E. Emergency department resilience to disaster-level overcrowding: A component resilience framework for analysis and predictive modeling. *J. Oper. Manag.* **66**, 54–66 (2020).
30. Tuominen, J. *et al.* Early Warning Software for Emergency Department Crowding. *J Med Syst* **47**, 66 (2023).
31. Erkamp, N. S., van Dalen, D. H. & de Vries, E. Predicting emergency department visits in a large teaching hospital. *International Journal of Emergency Medicine* **14**, 34 (2021).
32. Aliyu, B. S. *et al.* Enhanced Feature Selection for Imbalanced Microarray Cancer Gene Classification Using Chaotic Salp Swarm Algorithm. *International Journal of Theoretical & Applied Computational Intelligence* 258–283 (2025) doi:10.65278/IJTACI.2025.16.
33. Karimi, M. *et al.* Feature Selection Methods in Big Medical Databases: A Comprehensive Survey. *International Journal of Theoretical & Applied Computational Intelligence* 181–209 (2025) doi:10.65278/IJTACI.2025.21.

34. Khan, M. A. *et al.* Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection. *Expert Systems* **39**, e12497 (2022).
35. Kurdi, S. Z. Machine Learning–Based Classification Framework for Human Health Care Monitoring. *International Journal of Theoretical & Applied Computational Intelligence* **2026**, 1–15 (2026).
36. Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* **2**, 749–760 (2018).
37. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* **2**, 56–67 (2020).
38. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).
39. Motowidlo, S., Packard, J. & Manning, M. Occupational Stress. Its Causes and Consequences for Job Performance. *The Journal of applied psychology* **71**, 618–29 (1986).
40. Sutton, R. T. *et al.* An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digit. Med.* **3**, 1–10 (2020).
41. Kawamoto, K., Houlihan, C. A., Balas, E. A. & Lobach, D. F. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* **330**, 765 (2005).
42. Rauch, J., Hübner, U., Denter, M. & Babitsch, B. Improving the Prediction of Emergency Department Crowding: A Time Series Analysis Including Road Traffic Flow. *Stud Health Technol Inform* **260**, 57–64 (2019).
43. Derlet, R. W. & Richards, J. R. Overcrowding in the nation's emergency departments: Complex causes and disturbing effects. *Annals of Emergency Medicine* **35**, 63–68 (2000).
44. Savioli, G. *et al.* Emergency Department Overcrowding: Understanding the Factors to Find Corresponding Solutions. *Journal of Personalized Medicine* **12**, (2022).
45. Trzeciak, S. & Rivers, E. P. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emerg Med J* **20**, 402–405 (2003).
46. Morris, Z. S., Boyle, A., Beniuk, K. & Robinson, S. Emergency department crowding: towards an agenda for evidence-based intervention. *Emerg Med J* **29**, 460–466 (2012).
47. Sartini, M. *et al.* Overcrowding in Emergency Department: Causes, Consequences, and Solutions—A Narrative Review. *Healthcare* **10**, 1625 (2022).
48. Garcia-Romero, M. *et al.* NEDOCS vs subjective evaluation, ¿Is the health personnel of the emergency department aware of its overcrowding? *Colombia Médica* **48**, 53–57 (2017).
49. Ilhan, B., Kunt, M. M., Damarsoy, F. F., Demir, M. C. & Aksu, N. M. NEDOCS: is it really useful for detecting emergency department overcrowding today? *Medicine (Baltimore)* **99**, e20478 (2020).
50. Kempf, J., Blöhm, R. & Weigl, M. Negative moralische Konsequenzen von Overcrowding bei Notaufnahmepersonal: eine explorative Befragungsstudie. *Med Klin Intensivmed Notfmed* <https://doi.org/10.1007/s00063-026-01408-7> (2026) doi:10.1007/s00063-026-01408-7.

51. Strada, A., Bravi, F., Valpiani, G., Bentivegna, R. & Carradori, T. Do health care professionals' perceptions help to measure the degree of overcrowding in the emergency department? A pilot study in an Italian University hospital. *BMC Emerg Med* **19**, 47 (2019).
52. Laxmisan, A. *et al.* The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. *International Journal of Medical Informatics* **76**, 801–811 (2007).
53. Kovacs, G. & Croskerry, P. Clinical Decision Making: An Emergency Medicine Perspective. *Academic Emergency Medicine* **6**, 947–952 (1999).
54. Croskerry, P. & Sinclair, D. Emergency medicine: A practice prone to error? *Canadian Journal of Emergency Medicine* **3**, 271–276 (2001).
55. Yordanov, Y., Beauvais, A. & Thiébaud, P.-C. Multimorbidity in emergency departments: urgent need for integrated care. *bmjmed* **3**, (2024).
56. Otto, R. *et al.* Length of stay as quality indicator in emergency departments: analysis of determinants in the German Emergency Department Data Registry (AKTIN registry). *Intern Emerg Med* **17**, 1199–1209 (2022).
57. Brick, C. *et al.* The impact of consultation on length of stay in tertiary care emergency departments. *Emerg Med J* **31**, 134–138 (2014).