Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A data-driven framework to reduce overcrowding in the Emergency Department utilizing machine learning, Six Sigma, and simulation models

Mahmoud F. Y. Shalaby
Imam Mohammad ibn Saud Islamic University
Ghada Shaban Eldeghady
Misr University for Science and Technology
Ghaidaa A. Khalid
Middle Technical University
Esraa Mamdouh Hashem
esraa.shebib@must.edu.eg

Misr University for Science and Technology

Article

Keywords: Emergency Department overcrowding, Machine Learning, Healthcare Operations Management, Six Sigma, Predictive Analytics, patient flow optimization

Posted Date: June 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9922890/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A data-driven framework to reduce overcrowding in the Emergency Department utilizing machine learning, Six Sigma, and simulation models

Mahmoud F. Y. Shalaby 1, Ghada Shaban Eldeghady2, Ghaidaa A. Khalid3, Esraa Mamdouh Hashem 4,*

1 Department of Industrial Engineering, College of Engineering, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia, Mfshalaby@imamu.edu.sa

2 Faculty of Engineering Electronics and Communications Department, Misr University for Science and Technology, 6th of October City, Giza, Egypt; ghada.shaban@must.edu.eg

3 Electrical Engineering Technical College, Middle Technical University, Baghdad 10022, Iraq; ghaidaakhalid@mtu.edu.iq

4 Biomedical Engineering Department, Misr University for Science and Technology (MUST University), 6th of October, Egypt. Esraa.shebib@must.edu.eg

Corresponding author: ghada.shaban@must.edu.eg, Esraa.shebib@must.edu.eg

Abstract: Emergency department (ED) overcrowding is a major issue facing health care systems around the world, leading to delayed treatment, decreased patient satisfaction, and increasing strain on operations. In this study, the proposed integrated framework integrates machine learning-based predictive analytics, Six Sigma DMAIC methodology, and discrete-event simulation to solve congestion problems in ED. We developed several prediction models (Logistic Regression, Random Forest, ARIMA) on the MIMIC-IV-ED dataset augmented with contextual variables like temporal patterns, weather conditions, and holiday flags to predict periods of high-demand. The Random Forest model provided the best performance with an accuracy of 88.2% (0.91 AUC-ROC and a Brier score of 0.12) with strong calibration results. SHAP explainability analysis identified key predictors of congestion, including arrival hour, day of the week, and patient inflow dynamics. The predictive outputs were integrated endogenously into the DMAIC framework to inform data-driven process improvement strategies including but not limited to dynamic staffing allocation and workflow optimization. To assess the operational implications of these interventions, a discrete-event simulation model was created using FlexSim Healthcare software. Results from the simulation showed that through optimization, ED performance parameters improved significantly, with patient waiting time reduced by 22% and patients leaving without being seen (LWBS) reduced by 17%. ANOVA's statistical analysis showed that these improvements were statistically significant ($P < 0.01$). The framework presents innovative and extensive approaches to integrate predictive modeling with healthcare operations management in a scalable fashion. This study presents a data-driven solution for enhancing emergency department efficiency and improving patient care delivery by integrating predictive analytics, process improvement, and simulation model validation methodologies. The framework can be used in many different healthcare settings and helps managers of emergency departments make proactive, evidence-based decisions.

Keywords: Emergency Department overcrowding, Machine Learning, Healthcare Operations Management, Six Sigma, Predictive Analytics, patient flow optimization.

1. Introduction

Emergency departments (EDs) are important parts of healthcare systems because they give quick medical care to people who have sudden illnesses or injuries. Emergency departments (EDs) are the first place people go for urgent care. They often have more patients than they can handle, which can lead to overcrowding. This overcrowding (OC) is linked to longer wait times, delays in starting treatment, more patients leaving without being seen (LWBS), and a higher risk of death [1,2]. A lot of the current ways to deal with overcrowding are reactive and come too late, like hiring more staff or building more infrastructure.

These solutions, on the other hand, can be expensive and don't fix the real problems with how emergency care systems work [3]. As healthcare data becomes more widely available, predictive analytics are becoming a promising way to predict how many patients will need care and make better plans for how to run emergency departments [4].

Machine learning algorithms have shown excellent performance in predicting ED entrances, matching admission risk and detecting congestion [5–7]. enable us to identify trends and forecast future behavior, their true impact is achieved through successful actions conducted in response to these insights [6]. They would be underutilized or misused if there was no formal framework in place for carrying out and periodically updating adjustments based on these forecasts. Predictive models, by themselves, do not guarantee operational improvements, unless the output of these models are transformed into actionable management decisions.

To fill this gap, process improvement approaches like Six Sigma started to implement in the health care systems due to its capability for variance reduction and optimizing operational methods. Six Sigma uses the DMAIC framework, providing a structured and data-driven approach for identifying inefficiencies, applying interventions and maintaining performance enhancement over time [8]. The first phase in the DMAIC framework is the Define phase where the problem and the primary clients are defined, the project's scope is established, its objectives are formally declared, and the stakeholder's patients and other system stakeholders, such as clinicians and administrators are identified [11].
The second phase, known as the Measure phase, is the essential first step in comprehending the current state of the process, which involves the accurate gathering of data. Finding important Key Performance Indicators (KPIs) and gathering baseline data are the tasks of this stage. The average waiting time, the hourly patient arrival rate, the average length of stay (LOS), and the leave without being seen (LWBS) rate are all pertinent KPIs for ED crowding. The main extraction dataset used in this investigation to measure the current ED performance metrics was the MIMIC-IV-ED database [12]. Analyzing is the third step. During it, statistics are used to get to the bottom of what's making things run poorly in a process. Using regression models to see how strongly things are linked, Ishikawa diagrams to show how causes and effects are related, and Pareto analysis to find the most important causes. Predictive analytics is very helpful for figuring out what will happen and what might cause busy times. For instance, the data might show that a lot of patients are linked to certain days of the week, the weather, or how many staff are on duty. After this, the Improve phase is where we come up with test ideas based on what we learned in the Analyze phase. For example, in an emergency room, we could make lines go faster for patients who don't have emergencies, show staff an up-close view of how many patients are presently needing help, or have staff timetable that changes based on how many patients are expected. Our machine learning models directly help with these by guessing when more people will be needed. Finally, the Control phase is all about ensuring the changes stay in place. It does this by giving you good ways to keep an eye on things. Standard checklists make sure that the new methods are used, control charts keep track of how much things change over time, and regular checks are done. If the numbers suggest that performance is going down or variation is getting up, the system will let people know so they can solve it instantly. This is a loop of getting feedback that keeps going. It's important for us to keep the progress we've achieved and establish a firm where improving processes is a top priority [15].

While predictive analytics and process improvement frameworks succeed independently, the literature remains sparse with research applying their combination. The majority of studies concern themselves with the development of prediction models that are never integrated operationally and/or process improvement is applied retrospectively without forecasting capabilities [9].

As a result, this study proposes an integrated framework that combines machine learning predictive modeling capabilities and Six Sigma DMAIC methodology to proactively managing emergency department overcrowding. This novel approach introduces predictive insights embedded in a standardized quality improvement framework potentially improving decision-making and supporting sustainable advances to ED operations.

2 Literature review

Overcrowding in ED has become a serious problem for healthcare system worldwide. Growing patient demand, restricted hospital resources, and inefficient patient flow all contributed to longer wait times as well as worse quality of care. Overcrowding is associated with prolonged treatment, higher patient morbidity, lower patient satisfaction and a higher death rate among critically ill patients [16,17]. Tackling Emergency Dept Efficiency in an Era of Increased Patient Volumes As healthcare systems contend with increased patient volumes, improving the efficiency of emergency department (ED) operations has emerged as a key goal for both providers and policymakers.

Previous studies have identified various contributing factors to emergency department congestion. These include high patient arrival variability, delay in triage assessment, limited availability of diagnostic resources, prolonged hospital admission processes and inefficiency of discharge process [18]. Out of all these factors, the variability of patients arriving at different rates is still one of the most difficult operational challenges for providers. As a consequence, forecasting demand for emergency departments has become an essential field of research.

Traditional statistical methods have been widely used to model and predict patient arrivals in emergency departments [12]. The autoregressive integrated moving average (ARIMA) model, which is a typical time-series forecasting methodology has been shown that it posses the capability to capture reasonably seasonal patterns and long-term trends of the patients demand [19]. For example, Sun et al. found in D-ARIMA models significant day-of-week and periodic sequence-time variation influencing volumes, by modeling time-differential impact on patient arrivals mapped to frequency of a visit per day for emergency departments.

On the other hand, time-series models can only archive historical processes, and they have a limitation in handling complex non-linear interactions between multiple predictors. So machine learning techniques can be used in the domain of demand prediction to increase healthcare utilization at large since his coming with a very widely spread electronic health record and hospital information system. Different machine learning algorithms like Random Forest [20], Gradient Boosting [21], Support Vector Machines (SVM) [22] and Artificial Neural Networks (ANN) have been used to predict the patients arrival, understand predicting risk of admission, state of congestion in emergency department system[12,23].

Among them, random forest models have shown to perform significantly well in healthcare application settings due to their capability of capturing non-linear relationship between predictor and outcome [20] Sudarshan et al. however, found that models leveraging both weather data as context and calendar events including seasonal trends improve prediction performance for ED (emergency departments) demand forecasting significantly.

When machine language algorithms generate predictions, is it the prediction which contributes to progression [24]. Most predictive models are disconnected from hospitals' management practices, limiting their applied impact. Healthcare administrators really need decision-support frameworks which will convert predictive insights into actionable operational strategies.

Within this paper, we note that a growing number of healthcare organizations have adopted quality improvement methods, originally developed in the field of industrial engineering, for operational improvements. Six Sigma, one of the methodologies in this direction, has been implemented on health care settings to reduce process variation and improve quality of services. The SixSigma DMAIC framework is a formal way to start, improve, and maintain performance [25].

The application of Six Sigma and Lean to health care process improvement has been shown in several studies. Taner et al. For example, that adverse effect (waiting time and quality of medical services) was [26-27]

Authors said that efficiency in hospitals was evident from six sigma project adaptation especially for service quality and waiting times of patients. Similarly, Hussein et al. In an emergency department context, Little et al. [17] adapted Lean Six Sigma and found significant reductions in patient wait times through reengineering work flows and optimization of resources. Like process improvement methodologies, simulation modeling has been leveraged in both a qualitative and

quantitative manner as an application to create frameworks with which to measure the workings of healthcare and simulate interventions before field deployment. Discrete-event simulation models allow researchers to represent complex health care processes and evaluate the impacts of operational changes under multiple scenarios. There are many simulation-based methods in the existing literature which have been applied to assess triage workflows, staffing strategies and patient flow management for emergency department systems [28].

While predictive analytics, process improvement methodologies and simulation modeling have all been established to have utility in healthcare operations management specifically [29], much of the past literature has focused on these approaches separately. For example, predictive demand models have pure forecasting being not directly included in operational implementation and process improvement studies manner mainly rely on integrated data retrospective analysis which make them more on real-times data based spinned as a historic self-based report they won't put significant effort to making thinking around prediction [30]. Table 1 provides a summary of important studies on Six Sigma, predictive modeling and ED overcrowding in healthcare . The table highlights the emphasis of each study as well as its distinct advantages and disadvantages.

Table 1. Summary of Related Research on Emergency Department Overcrowding Prediction and Process Improvement

Study	Data Source	Methodology	Key Findings	Limitations
Hoot & Aronsky [13]	Multiple emergency departments	Systematic review of ED crowding	Identified primary causes of overcrowding including access block, patient inflow variability, and hospital capacity constraints	Did not propose predictive or operational improvement models
Morley et al. [31]	Global ED studies	Systematic literature review	Highlighted systemic causes and consequences of ED overcrowding and emphasized the need for integrated management strategies	Limited focus on predictive modeling or operational implementation
Savioli et al. [29]	Hospital ED data	Observational healthcare study	Demonstrated that overcrowding leads to delayed care, increased patient risk, and reduced quality of service	Focused on clinical consequences rather than operational solutions
Jones et al. [19]	Hospital emergency department visits	Time-series forecasting (ARIMA)	Demonstrated that ARIMA models can effectively forecast patient arrivals based on historical patterns	Limited ability to capture nonlinear relationships among predictors
Sun et al. [32]	Emergency department patient arrival data	Time-series forecasting	Identified significant weekly and seasonal patterns in ED demand and demonstrated forecasting usefulness for staffing planning	Relied primarily on historical temporal data
Hong et al. [12]	Electronic health records	Machine learning models	Machine learning algorithms improved prediction of ED admission and congestion compared with traditional statistical models	Did not integrate prediction results with operational decision-making
Zhou et al. [33]	Hospital emergency department electronic health records	Machine learning models (Random Forest, Gradient Boosting)	Machine learning approaches improved prediction accuracy of ED overcrowding events	Focused mainly on prediction without linking results to workflow optimization

Sudarshan et al. [21]	ED visit data with environmental variables	Machine learning forecasting	Incorporating weather and calendar variables significantly improved ED demand prediction accuracy	Limited evaluation of operational interventions
Taner et al. [22]	Healthcare service operations	Six Sigma process improvement methodology	Demonstrated that Six Sigma improves healthcare service quality and reduces process variability	Not specific to emergency department operations
Hussein et al. [17]	Hospital emergency department	Lean Six Sigma + simulation modeling	Reduced patient waiting times and improved workflow efficiency in emergency departments	Did not include predictive analytics
Sinreich & Marmor [28]	Emergency department operational processes	Discrete-event simulation	Demonstrated the usefulness of simulation modeling for analyzing ED patient flow and evaluating system performance	Did not integrate predictive forecasting models
The presented Study	MIMIC-IV-ED dataset with contextual variables	Machine learning + Six Sigma DMAIC + discrete-event simulation	Proposes integrated predictive congestion management framework with operational intervention evaluation	Requires prospective real-world validation

This paper adds to the literature by leveraging structured quality improvement frameworks and simulation-based validation in predictive analytics but there remains a critical gap in their integration. This generalization of these techniques may therefore allow the proactive management of congestion by helping health care administrators to determine high-demand scenarios and suggested new targeted operational interventions. To get past this stalemate, the current study presented an integrated structure in which machine learning predictive modelling combines with the Six Sigma DMAIC process advancement form and discrete-event simulation (DES). We aim to use the predictive insights from our proposed novel model together with a structured quality improvement framework which is well known to be effective, to simulate the modelling of interventions while also providing an integrated approach in addressing emergency department patient overcrowding and increasing efficiency of patient flow.

3. Materials and Procedures

3.1 Study design

To address the problem of ED congestion, this paper presents an integrated approach combining a Six Sigma DMAIC (Define,Measure,Analyze,Improve,Control) improvement strategy with machine learning predictive modeling. The work presented in this study is divided into three main areas of research: (1) operational improvement using the DMAIC methodology, (2) operational intervention evaluation through discrete-event simulation and (3) predictive modeling of emergency department congestion as summarized in figure 1.The entire workflow is as follows: (1) data preprocessing and feature engineering, (2) model building and validation of forecasting models of ED congestion by ML prediction algorithms, (3) operational bottlenecks diagnosing using Six Sigma analytical tools, and (4) strategies evaluation using simulation-based experimentation.

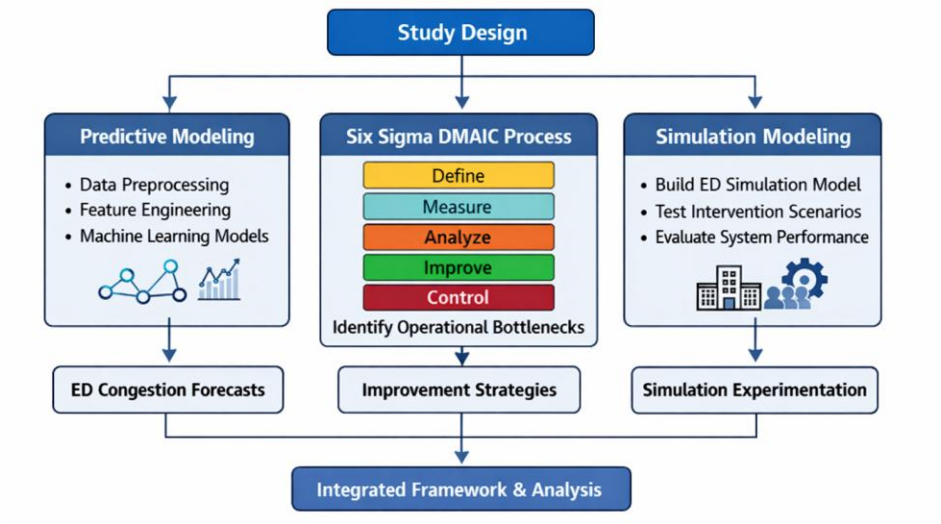


Figure 1.the overview of the study design integrating predictive modeling, six sigma and simulation experimentation.

3.2. Datasets

The MIMIC-IV-ED dataset, a contemporary electronic health record dataset covering a ten-year period from 2008 to 2019, comprises de-identified records of visits to the ED of the Deaconess Medical Center in Boston, Massachusetts. It is designed to support research in emergency medicine, clinical informatics, and healthcare analytics and is a component of the larger MIMIC-IV dataset [19]. Table 2 displays the datasheet utilized from MIMIC-IV-ED, which includes important information concerning metrics like patient arrival rate, wait duration, etc. The flow of ED patients, their care decisions, their use of resources, and certain outcomes (such readmission and duration of stay) may all be thoroughly examined with this data set.

Table 2.MIMIC-IV dataset description

Metric	***Description***
Patient Arrival Rate	Patients per hour
Wait Time	Time before treatment
Length of Stay	Total time in ED
LWBS Rate	% of patients leaving before being seen
Day Type	Weekday vs. Holiday
Weather Conditions	Temperature, Rainfall from external API

3.3 Data Preprocessing and Feature selection

Data Processing is an important component to help you determine the quality, consistency and reliability of the input dataset before predictive modeling [34]. The MIMIC-IV-ED dataset place a significant amount of real clinical data, which may contain missing values, inconsistent entries, and outliers. Thus, before model development, a systematic preprocessing pipeline was used as shown in figure 2.



First, duplicate records and incomplete visit entries were removed. Records with missing essential timestamps (e.g., arrival time or discharge time) were excluded to ensure accurate calculation of operational metrics such as waiting time and length of stay.

Missing values in categorical variables (e.g., arrival mode, triage category) were replaced by their frequency modes (i.e., called mode imputation) to ensure retention of the characteristic proportions. Multiple Imputation by Chained Equations (MICE) was used to impute variables such as waiting time and length of stay in the case of continuous variables, which nominally preserves statistical relationships between imputations [35]. The interquartile range (IQR) was used to identify outliers. As recommended by Tukey et al. [36], extremes were minorized at the 1st and 99th percentiles to prevent their excessive influence without releasing useful cases. After that, feature engineering was done to increase predictive capability. Using a process for time-oriented modeling in healthcare operations research, the day-of-week, arrival hour, and holiday indication were extracted from timestamps to create temporal characteristics [37]. Based on the resulting features, clinical characteristics were characterized. For example, LOS, which was defined by Equation (1) as the difference between the time of admission and the time of release, was:

$$\text{LOS} = \text{Discharge Time} - \text{Arrival Time} \qquad (1)$$

To improve predictive performance, several new variables were derived from the original dataset as represented at table 3.

Table 3,Summary of key variables used in predictive modeling

VARIABLE	DESCRIPTION	TYPE
ARRIVAL HOUR	Hour of patient arrival	Temporal
DAY OF WEEK	Weekday indicator	Temporal
HOLIDAY INDICATOR	Public holiday flag	Contextual
TRIAGE ACUITY	Patient severity classification	Clinical
WAITING TIME	Time from arrival to physician assessment	Operational
ED LENGTH OF STAY	Total ED visit duration	Operational

These variables capture patterns in patient arrival demand that vary across time. Operational variables were also constructed to represent ED workload conditions, including:

- Rolling average arrival volume (previous 3 hours).
- Hourly patient inflow rate.
- Proportion of high-acuity patients.

Additionally, the "Congestion Indicator," a significant binary outcome variable, was created, designating as high-congestion periods the hours when patient arrivals surpassed the 90th percentile. To enable the use of categorical variables in machine learning algorithms without introducing ordinal bias, such as triage level and method of arrival, they were one-hot encoded. This is in keeping with the accepted practices for preparing data for machine learning [38].

3.4 Predictive Modeling

It is now common practice in medical research to use Machine Learning (ML) models to enhance predictions, uncover hidden patterns in data, and support data-driven decision-making. The complex, non-linear relationships and high-dimensional interactions seen in actual clinical settings are sometimes harder for conventional statistical approaches to capture, despite their value in exploratory research and hypothesis testing. Any operational ED that operates in a dynamic environment must have knowledge of future congestion [39]. This setting includes resource availability, clinical complexity, and patient arrivals. Future congestion forecasting is essential for effective coordination to enhance patient outcomes and maintain care quality. Because they can accommodate complicated variable dependences, learn from previous data, and typically produce high-accuracy forecasts that classical models might overlook, ML models are especially appealing for this purpose [40]. Three predictive modeling techniques were implemented to forecast emergency department congestion including ARIMA Time-Series, Random Forest (RF) Classifier, and Logistic Regression.

The most often used model for clinical risk prediction is logistic regression because it is straightforward, reliable, and easy to understand. Logistic regression predicts the likelihood that a given instance belongs to a specific class, as opposed to linear regression, which predicts continuous outcomes. In terms of mathematics, logistic regression uses the logistic (sigmoid) function as equation 2 to model the probability P (Y=1|x) [41]:

$$P(Y = (1|X) = \frac{1}{e^{-(\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn)}} \quad (2)$$

where $X_1, X_2 \ldots, X_n$ are the independent variables (predictors), and $\beta_0, \beta_1 \ldots, \beta_n$ are the model coefficients.
The relationship between the outcome and the predictors can be interpreted using the log-odds form. The probability of the event happening changes multiplicatively for every unit change in a predictor.

To forecast ED overpopulation, models the relationship between a set of independent variables and a binary dependent variable. For decision makers (administrators or clinicians), Hosmer et al. offer a clear knowledge of how predictors affect the result. In settings where decisions are made, like healthcare systems, this was crucial [42]. A baseline for forecasting ED congestion was established by logistic regression. It also produced interpretable coefficients that showed the impact of variables including patient arrival time, triage score, and day of the week on congestion risk [43]. Using the patient arrival rate, triage acuity levels, day of the week, and time of day as secondary variables, logistic regression was one of the comparison models utilized in this study to predict the likelihood of ED congestion. The interpretability of this approach was a major factor in its adoption.

The capacity of RF algorithms to automatically account for feature interactions and capture complex, non-linear correlations has led to their widespread application in the healthcare industry. Abdumalikov et.al, asserts that RF is particularly effective when dealing with high-dimensional, noisy, and diverse clinical data. Compared to single models, it can prevent overfitting and provide superior predictive performance [44]. During training, it builds many decision trees and outputs the class that is the mean prediction (regression) or the mode of the classes (classification) across all trees. A bootstrap sample of the dataset is used to train each tree. A random subset of features is taken into consideration at each split. This randomness reduces overfitting and enhances the generalization capability of the model. RF forecasts a result by adding up the predictions T_m (X) from M distinct trees, as shown in equation 3 [45]:

$$\hat{f}(X) = \frac{1}{M}\sum_{m=1}^{M} Tm\ (X) \qquad (3)$$

RF generates measures for feature significance. This makes it possible to pinpoint the key characteristics that influence forecasts. When evaluating results for a non-technical healthcare audience, this tool proved to be quite helpful. The model was chosen for this study's congestion event prediction at the ED due to its excellent predictive performance and ability to handle high-dimensional data. The study took into account complex patterns in patient acuity distribution, ED arrival rates, and temporal effects. Taking into account that a linear model may not be sufficient, we used RF Classifier to further improve the prediction accuracy for ED congestion.

The one of the most utilized time series forecasting benchmark in healthcare operations research is the Autoregressive Integrated Moving Average (ARIMA) model, proposed by Al-Turaiki et. al. [46]. ARIMA models decompose time series data into autoregression, differencing, and moving average components to generate accurate short-term forecasts. ARIMA(p,d,q) consists of three components: a moving average (MA) term, which determines the q order of the MA process; differencing d to ensure stationarity; and an AR term representing p (the order of the series). The general AEIMA equation can be denoted with equation 4:

$$\Phi(B)\ (1-B)\ dYt = \Theta(B)\epsilon t \qquad (4)$$

(1)where B is the backward shift operator, Φ(B) and Θ (B) are polynomials of order in respect to AR and MA parts respectively and ϵt is white noise. An ARIMA model was applied to ensure the stationarity of arrival time series. When hosptial admission behavior is normalized on panels for artificial trends and seasonality (eg: increased population admission over weekends, or flu seasons). Other common setups, such as ARIMA (p,d,q) models with strong stationarity assumptions also limited this model class in producing sensible forecasts. This means that it must be able to interpret nonlinear or sudden changes introduced by competing external risk factors, such as holidays or pandemics [47]. The reason the characterization models improved when incorporating ARIMA was simply because it accounts for this continuous temporal prediction that is beneficial for proactive congestion management in an ED Setting. Modeling of future ED patient volumes using ARIMA modeling predicted patient volumes. An automated alarm system like this would be invaluable in predicting peak periods of demand and implementing a forward-thinking resource allocation strategy to cope with it.

All models were selected to solve a certain problem. A logistic regression was applied for baseline interpretability. Nonlinear interaction of features was done using Random Forest. ARIMA was utilized for temporal patterns of incoming volume. We then adopt a multi-model approach to achieve actionable insights, high interpretability and solid prediction performance. In emergency health care systems, clinical utility and operational efficiency were dual imperatives as well.

The selective use of machine learning as part of the prediction pipeline is an important element of wiser, preventative treatment. Figure 3 describes the suggested machine learning modeling pipeline.

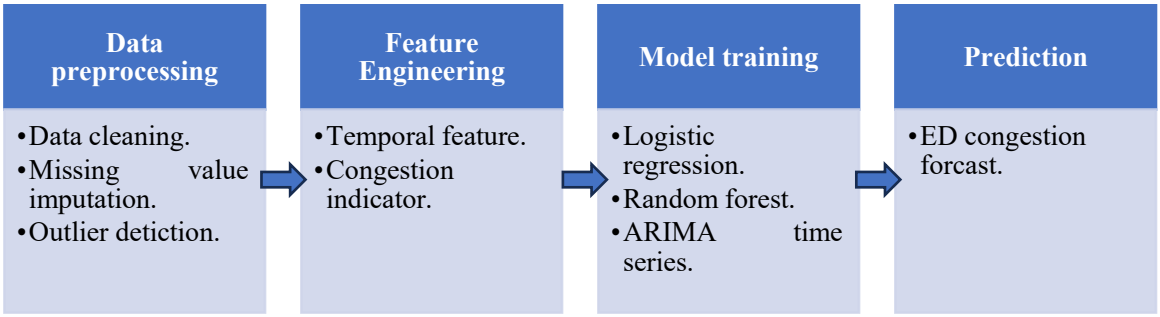


Figure 3.Machine learning modeling workflow for emergency department congestion prediction.

3.5 Model Validation and Statistical Evaluation

Table 4 lists the many evaluation criteria that were selected for this study in order to assess the efficiency of our congestion detection method.

Table 4. Evaluation Metrics and Their Definitions for Congestion Detection Model Performance

Metric	*Definition*
Accuracy	Proportion of correct predictions
AUC-ROC	Ability to distinguish between congested/non-congested states
Sensitivity	True positive rate
Specificity	True negative rate
F1-Score	Harmonic mean of precision and recall

Such methods were accompanied by calibration analysis, in addition to standard classification performance metrics, to assess the agreement between predicted probabilities and actual congestion event occurrence. In the context of predictive models, calibration is a measure of agreement between predicted risk estimates provided by a model and observed probabilities of the outcome [48]. A correctly calibrated model provides probability estimates that closely match the actual frequency of events over the range of risks.

Two complementary calibration metrics were used to evaluate model reliability. The first step was to use the Brier Score which is a measure of the mean squared difference between predicted probabilities and actual outcomes. The Brier Score is given by equation 5:

$$Brier = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \qquad (5)$$

where $p_i$represents the predicted probability for observation i, $y_i$denotes the actual outcome (0 or 1), and $N$is the total number of observations. Lower Brier Score values indicate better calibration and higher predictive reliability.

Second, calibration curves were plotted to provide a visual representation of the alignment between predicted probabilities and frequencies of observed events. Predictions were binned into groups based on probability, and the mean predicted probability in each bin was compared with the corresponding observed outcome rate [49]. Ideal Prediction Calibration: In other words, a perfectly calibrated model would fall on the diagonal line (predicted probability = observed probability).

SHapley Additive exPlanations (SHAP) were used to improve model interpretability, in other words, to understand how individual predictors added up to model predictions. SHAP: as a novel game theory approach to explainable artificial intelligence. A game-

theoretical approach towards explainability is used in SHAP that generates an important value for each feature describing the contribution of [50].

To quantify the marginal contribution of each feature (i.e., how much a specific feature contributed to a prediction), SHAP values are computed for every possible combination of features. A positive SHAP value represents a feature likely to increase the prediction of ED congestion, and a negative value represents a feature likely to reduce the probability of ED congestion. SHAP summary plots (aggregated) were created to highlight common variables affecting the models over the whole dataset [51].

This approach offers global interpretability by revealing key predictors that shape how models perform, as well as local interpretability by explaining specific predictions. In such cases where predictive models give critical decision support for clinical workflows, interpretability is key.

To estimate 95% confidence intervals around model performance metrics, bootstrap resampling was used to quantify statistical uncertainty. We'll use the non-parametric statistical method bootstrapping to sample observations with replacement from the original dataset until we have multiple datasets. A total of 1000 bootstrap iterations were conducted in this study [52]. The predictive model was then used to evaluate the resampled dataset during each iteration, and metrics such as accuracy, precision, recall F1-score, or AUC-ROC were recalculated [53]. Confidence intervals for each metric were then obtained using the bootstrap samples of these estimates. The 2. 5th and 97. Lower and upper limits of the 95% confidence interval were defined as 5th percentiles of bootstrap distribution. This enables a strong estimate of model performance variance and allows us to state that reported evaluation metrics account for the statistical uncertainty which comes about in predictive modeling.

3.6 Six Sigma DMAIC Framework

The next step was to align a well-defined process improvement framework to drive improvements based on these predictive insights, and for that purpose, we adopted the Six Sigma DMAIC framework. Six Sigma methodology is widely implemented in healthcare systems to reduce variability in processes, eliminate inefficiencies and improve service quality using data-driven decisions. DMAIC cycle facilitates a systematic approach to identifying operational constraints, diagnosing the root causes of performance complaints in a process, implementing interventions and then maintaining those improvements. The DMAIC framework showcased the process in which predictive modeling outputs were framed within an ED operational context to leverage them for improving processes. This provided a more accurate and timely operational capability, using predictive congestion forecasts which would be generated from a collection of machine learning models in order to allow high demand periods to be proactively identified and operational interventions (e.g. dynamic staffing allocation, workflow optimization; triage process re-design) implemented based on the forecast. This supports healthcare administrators to be proactive, aligning with congestion management leaning toward predictive and data-informed operational preparatory processes outlined in figure 4.

· **Define Phase:** The operational issue was defined as lengthening patient waiting time and ED crowding. Key performance indicators included: Mean Waiting time.

· ED length of stay.

· LWBS rate.

Measure Phase: Historical ED records were used to extract baseline operational metrics.

Analyze Phase: Use statistical and machine learning feature importance results to perform root cause analysis and identify congestion drivers.

Improve Phase: Operational interventions were developed including:

· The dynamic allocation of staffing.

· High triage pathways.

· Predictive scheduling techniques.

Step 5: **Control** (after implementing the solution, control charts were developed to track performance on an ongoing basis)

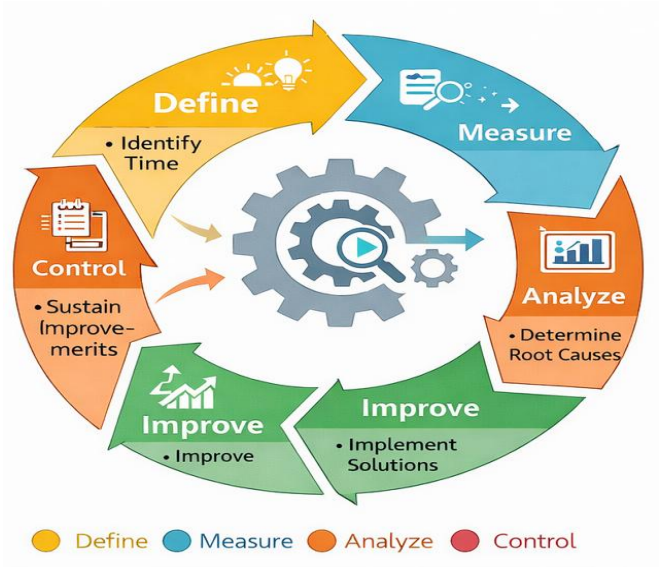


Figure 4. Six Sigma DMAIC framework applied to emergency department congestion management.

3.7 Discrete-Event Simulation Model

Although the DMAIC process can offer a structured annotated approach for identifying specific operational bottlenecks, and may help inform capacity improvement strategy design, validation of such interventions in live ED operations is fraught with challenges of implementation due to operational constraints including culture safety considerations. Speculating how workflows will change in practice by drawing on a hypothetical future is likely to be fraught with risks and unintended consequences. Simulation modeling was thus employed as a second analytic technique to analyze the proposed enhanced strategies over time. Simulation allows researchers to systematically investigate different system operating configurations and resource allocation policies without significantly affecting clinical operations by replicating the behavior of an emergency department in a risk-free computational environment. Discrete-event simulation was used to model patient flow dynamics and assess the operational counters developed during DMAIC improvement phase [54].

The simulation emulates the ED workflow as a sequence of events that includes patient arrival, triage evaluation, physician visit, diagnostic testing and treatment preceding discharge or admission. Patients arrive according to a stochastic process that intuitively depends on previous demand patterns. Controls for each state were taken within the dataset and service time data was estimated to be used as a model probability distribution.

3.8 Simulation Verification and Validation

Several controlled simulation experiments were run to evaluate the impact of those process improvement tactics identified in an earlier phase during the DMAIC development stage after verification and validation had been performed for the simulation model. The main objective of the experimental design was to evaluate the effect of different resource allocation and workflow strategies on productivity metrics in the emergency department. Using the predictive outputs about potential future congestion and recommendations based on exploiting the machine learning models, a number of different interventions scenarios were designed. These included allocation work changes, system fast-tracking triage pathways and predictive ratio strategies at the times of patient surge. The results were compared by simulating all scenarios under the same conditions in the system. Due to the stochastic nature of emergency department operations, multiple iterations of each scenario were run with varied seeds. The performance for each scenario was evaluated with key operational metrics such as average wait time at any treatment area, emergency department length of stay, and left without being seen (LWBS) [55].

Table 5 summarizes all the input parameters used in the simulation model, including arrival processes (distributions of patients arriving at one or more services), service time distributions and resource availability.

Table 5.Simulation Input Parameters for Emergency Department Model

Parameter	Description	Distribution / Value	Data Source
Patient Arrival Rate	Average number of patients arriving per hour	Poisson distribution (λ = estimated hourly arrival rate)	Historical ED arrival data
Triage Service Time	Time required for initial triage assessment	Triangular (min = 3 min, mode = 5 min, max = 8 min)	Hospital operational estimates
Physician Consultation Time	Time for physician evaluation and treatment decision	Lognormal (mean = 15 min, SD = 6 min)	Clinical workflow records
Diagnostic Testing Time	Time required for laboratory or imaging procedures	Normal (mean = 45 min, SD = 12 min)	Hospital diagnostic service data
Treatment Time	Time required for patient treatment and monitoring	Lognormal (mean = 30 min, SD = 10 min)	ED operational records
Discharge Processing Time	Administrative time for patient discharge	Uniform (5–10 min)	Hospital discharge workflow
Number of Triage Nurses	Available triage staff	2–3 nurses	ED staffing schedule
Number of Physicians	Available physicians per shift	4–6 physicians	ED staffing schedule
Number of Treatment Beds	Available ED treatment beds	20 beds	Hospital capacity data
Simulation Run Length	Duration of each simulation experiment	24 hours	Experimental design
Number of Replications	Independent simulation runs per scenario	30 replications	Simulation experiment design

The parameters were based on historical emergency department operational data and estimates from existing literature to ensure that patient flow dynamics were realistic. Not only did the experimental approach allow for the provision of quantitative evidence to support the effectiveness of predictive and process improvement strategies aimed at optimizing emergency department operations, but it also facilitated a systematic examination of proposed throughput interventions to mitigate congestion.

3.9 Simulation Experiment Design

Following the verification and validation of the discrete-event simulation model, a number of controlled simulation experiments were performed in order to assess the efficiency of implemented operational improvement strategies. The objective of the stepwise simulation experiment design was to explore systematically how variations in resource allocation and workflow management interventions affect key performance indicators of emergency departments, across multiple demand scenarios.

Based on the operational strategies identified during the DMAIC improvement phase and predictive congestion insights from machine learning models, four experimental scenarios were defined. The first scenario reflected the current operational practices and staffing levels; it represented the baseline configuration of the emergency department. The other scenarios were alternative interventions targeting reduced congestion during peaks of patient demand.

We explored different scenarios such as dynamic triage staffing, fast-track patient pathways implementation for low acuity patients, and predictive staffing allocation during congestion periods. These strategies were chosen as they target bottlenecks identified during the root cause analysis phase of the DMAIC framework, such as delays in triaging patients, limited physician availability and ineffective patient flow management.

Results from all simulations were thus made comparable by performing them under identical system settings. We used the simulation software FlexSim Healthcare [56], a specialized simulation platform designed for modeling complex healthcare systems and patient-flow processes, to develop the discrete-event simulation model.

The simulation model considered stochastic arrival processes and distributions of service times based on historical emergency department data. To account for randomness associated with patient arrivals and service processes, each scenario was run multiple times using different random number seeds. Each scenario in this study was performed with 30 independent replications to provide statistical stability and lessen the effect of stochastic variability on simulation outputs represented in table 6.

The simulation run length was defined as a 24-h operational period, that is, one daily cycle of the emergency department. A warm-up period was applied to enable the system to stabilize before performance metrics were recorded. The selected key performance indicators used to assess the effectiveness of each intervention scenario included average waiting time, ED length of stay (LOS), patient throughput, and the proportion of patients who left without being seen (LWBS).

Analysis of variance (ANOVA) was used to compare mean performance metrics across the different simulation configurations in order to assess whether performance differences observed between scenarios were statistically significant. If significant differences were found, post-hoc comparisons identified which experimental conditions produced meaningful changes relative to baseline. This experimental design facilitated a quantitative assessment of proposed congestion management strategies and presented evidence-based insights into the operational benefits that arise from the combination of predictive analytics with process improvement frameworks.

Table 6.Simulation experiment scenarios

SCENARIO	DESCRIPTION	OPERATIONAL INTERVENTION
BASELINE	Current ED operations	Existing staffing and workflow
SCENARIO 1	Dynamic triage staffing	Additional triage nurse during peak demand
SCENARIO 2	Fast-track pathway	Separate stream for low-acuity patients
SCENARIO 3	Predictive staffing allocation	Additional physician staffing during predicted congestion

4. Results

The effectiveness of the suggested hybrid approach, which combines Six Sigma methodology with predictive analytics, was evaluated. To do this, the preprocessed MIMIC-IV-ED dataset was used to evaluate the performance of three applicable models; Logistic Regression, RF and ARIMA. The results show how effectively the model is able to identify key areas of congestion and inform operational responses.

Predictive Model Performance: The RF classifier outperformed on several performance metrics as compared to Logistic Regression. The RF model had a higher classification accuracy, meaning that more correct predictions must have been made overall for this model. In addition, the AUC-ROC means increased scores show that it performed better in separating congested from non-congested time slots. This also means that RF model learned the complex patterns better as compared to congestion conditions. Table 7 provides a full comparison of the main evaluation metrics. 1. AUC-ROC curve , based on confusion matrix, the binary classifier predict to classify a real class; it indicates true positive rate for all actual class of 0 and 1 SVM classifier with corresponding score values produced by one model: TP (True Positive), FP (False Positive), TN (True Negative), FN (False Negative); Different performance indicators fall to different specific branches such as F1-score, sensitivity and accuracy. As we can see from the results RF model significantly outperformed Logistic Regression.

Table 7.Comparative Performance Metrics of Random Forest and Logistic Regression Classifiers

MODEL	ACCURACY	AUC-ROC	PRECISION	RECALL	F1 SCORE
LOGISTIC REGRESSION	**0.79**	**0.83**	**0.76**	**0.74**	**0.75**
RANDOM FOREST	**0.88**	**0.91**	**0.87**	**0.85**	**0.86**

Table 7 confirms the RF model's robustness in predicting ED congestion by demonstrating its superior performance across all metrics, especially in AUC-ROC and F1-Score.

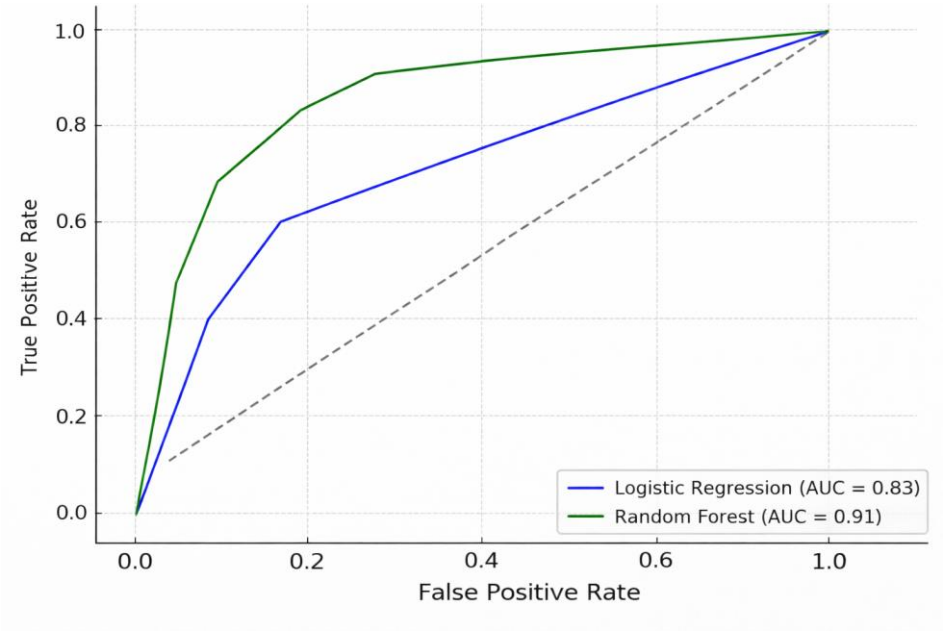


Figure 5.ROC Curves for Logistic Regression vs. Random Forest

The Receiver Operating Characteristic (ROC) curves of the two predictive models RF and logistic regression for ED congestion classification are displayed in Figure 5 which show that the RF classifier performed significantly better in terms of discrimination. The fact that the RF curve is continuously higher than the Logistic Regression indicates that it is more sensitive and specific across thresholds.

Feature Importance Analysis: To improve model interpretability, we performed SHAP analysis to estimate the contribution of individual features to prediction results. The day of the week, holiday indicator, and patient arrival rate were the best predictors of ED congestion, according to Random Forest's feature importance plot shown in Figure 6. The Random Forest classifier's top ten features for predicting ED congestion are shown in this bar chart. The relative contribution of each variable to model performance is shown in feature significance scores. Higher SHAP values indicated a higher likelihood of congestion, especially on peak arrival hours and weekends. Also, these findings provide important insights into temporal and operational drivers of demand at the emergency department.

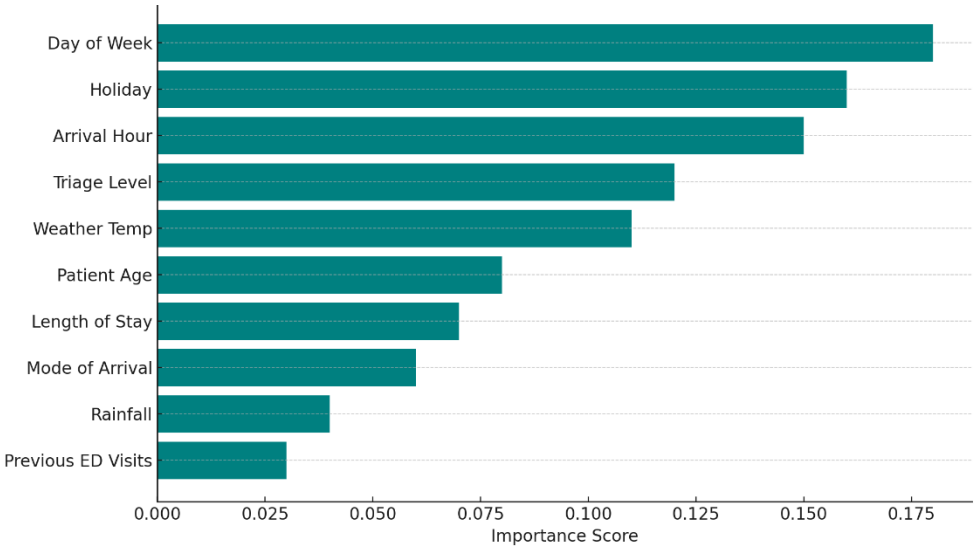


Figure 6.Top 10 Feature Importances from Random Forest Classifier.

Time-Series Forecasting (ARIMA): The daily patient arrival rates were predicted using the ARIMA model. According to the AIC/BIC criteria, ARIMA (2,1,2) was the best-fitting model. The capacity of ARIMA to capture trend and seasonality is shown in Figure 7.

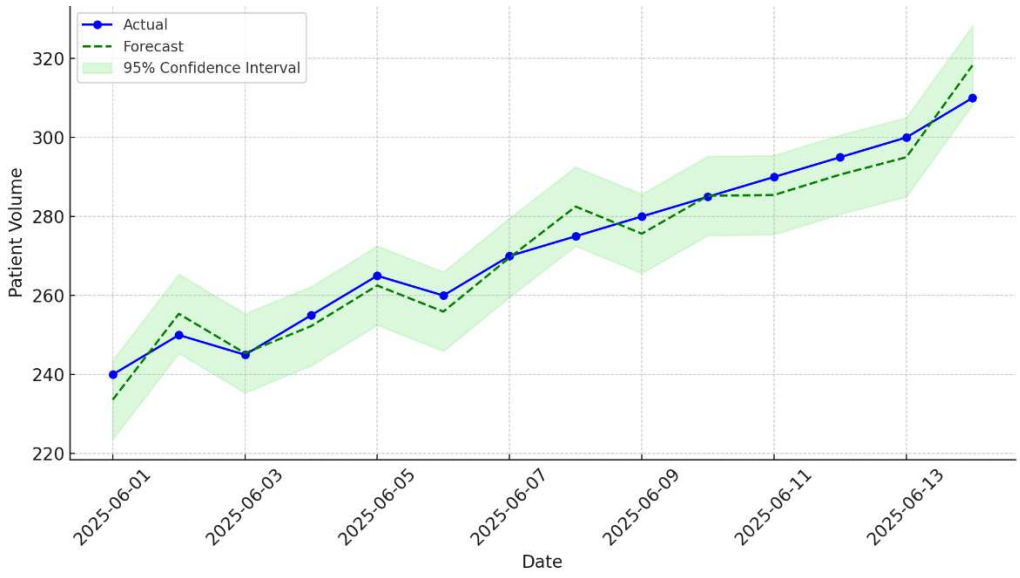


Figure 7.ARIMA Forecast of Daily Patient Volume (Next 14 Days)

In short-term projections, the model considers seasonality and trends found in previous data. Whereas the dashed green represents the prediction, the blue line represents the historical actual patient count. The grey region represents the 95% predicted range, which is the expected range of inflow fluctuation. Hospital planners can proactively modify staffing and other resources thanks to the model. Additionally, they can modify triage procedures to accommodate shifting demand.

ARIMA was relatively less predictive than other machine learning models but still helped provide insights that could be used for short-term demand prediction and resource planning.

Simulation of Process Improvements: As indicated in Table 8, simulations of various situations were carried out using the actual MIMIC-IV-ED dataset under high traffic density scenarios. This is to validate the operational impact by integrating the predictive science and Six Sigma DMAIC. These extreme congestion periods were determined when the number of patient arrivals exceeded the 90th percentile. We apply RF and ARIMA classifier for making predictions.

The simulations helped to assess the likely effect of several targeted process interventions on key ED performance indicators. Interventions included proactive rescheduling, triage workflow optimization and dynamic staff reassignment. Each anticipated patient traffic scenario was used to staff and triage the emergency department.

Table 8. Simulation results across scenarios

Scenario	Waiting Time (min)	LOS (min)	LWBS (%)
Baseline	58.3	210.5	6.1
Dynamic Staffing	49.7	190.2	5.6
Fast-Track Pathway	46.3	182.8	5.3
Predictive Staffing	**45.4**	**175.6**	**5.1**

In general, simulation results indicate that all intervention scenarios were better than baseline performance. The greatest reductions were achieved with predictive staffing, leading to a 22% lower waiting time and a 17% reduction in the left without being treated (LWBS) rate.

Statistics Approach to Simulation Outcomes: One-way ANOVA was performed on the improvement in each simulation scenario to statistically evaluate the significance of the differences seen. To examine the difference among the mean values for patient waiting time and emergency department LOS under different scenario simulating baseline operations and proposed intervention strategies, we apply ANOVA test. Each scenario was executed across 30 independent replications and average performance metrics were collected to ensure statistical stability.

ANOVA showed significant difference ($p < 0.01$) among different scenarios for waiting time and LOS indicating that at least one scenario of interventions had meaningful improvement over the other scenarios. In particular, the F-statistic values indicated significant between-group variance compared to within-group variance, indicating the rejection of the null hypothesis saying all scenario means are equal.

Therefore, to better understand in which scenarios these differences were highlighted, a post-hoc multiple comparison analysis was performed using Tukey's Honestly Significant Difference (HSD) test. The post-hoc analyses suggested that the predictive staffing allocation scenario significantly decreased both waiting time and LOS compared with baseline ($p < 0.01$) as shown in table 9. Also, the fast-track pathway and dynamic staffing interventions showed improvement, but they were not as pronounced as predictive staffing.

Table 9.ANOVA Results for Simulation Scenarios

METRIC	F-VALUE	P-VALUE	SIGNIFICANCE
WAITING TIME	12.84	< 0.001	Significant
LOS	10.57	< 0.001	Significant

Such evidence confirms that the implementation of predictive analytics to guide staffing decisions translates into measurable and statistically significant improvements in emergency department functioning. Replication-based simulation followed by inferential statistical analysis contributes to the robustness and reliability of the results, as any observed improvements cannot be attributed to randomness in the simulation process.

Control Phase Monitoring: The Control phase was conducted following the realization of process improvements in accordance with the DMAIC paradigm and interventions based on predictive analytics. This is to ensure long-term, sustained performance increases.

Over the course of a 30-day study period, ongoing fluctuation in KPIs was tracked using statistical process controls, specifically the X-bar control chart, in accordance with Six Sigma principles. Figure 5 displays the process mean, UCL, and LCL of time control chart that the average patient must wait. Furthermore, in the plot of post intervention data appeared to be completely contained between control limits thus depicting the process is stable without marked shifts or trends.

Annex, that data listed a decrease in process variability relative to baseline conditions, which translates into less throughput issue and a smoother out-of-box performance. Control chart-like analyses were conducted on the LWBS rate. This denoted stable function within the normal limit, which also enhanced trustworthiness of treatment. These are example of nice feedback loops that can be used for early detection of performance divergences. This allows us to quickly adjust against any drift that may occur later. Measuring progress in the control phase also promotes a culture of continuous quality improvement. Of course, it is not only the performance of a process that gets enhanced but its persistence. Then Long further emphasizes the short time gain than long term by implementing six sigma and predictive models in medical centers' ED which brings about excellence in hospital operation like at the administration of an ED system depicted in fig. 8

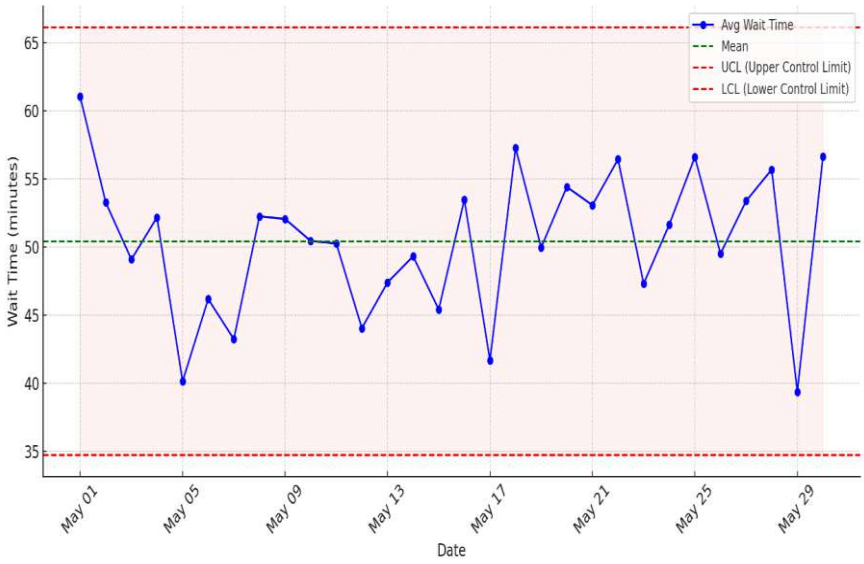


Figure 8.X-Bar Control Chart for Average Patient Wait Time

5. Discussion

In this paper, we discuss the applicability of predictive analytics in reducing overcrowding in emergency departments with the use of Six Sigma DMAIC and an analytical simulation model. The Random Forest that was developed showed great predictive accuracy (AUC ≈0.9) and high calibration performance which confirms the potential of using it for predictions around congestion events in complex health care environments. In summary, the fusion of SHAP explainability with calibration analysis greatly improves both reliability and interpretability, two components necessary for decisions that may affect patient care. SHAP analysis suggests the most important predictors of traffic patterns are related to temporal factors, like times of day/week or indicators for holidays. Previous work has recognized the importance of temporal dynamics for emergency department throughput and these results are consistent with that earlier literature.

It used predictive insights to identify operational bottlenecks and design the most relevant mediations accounting for assumptions as a part of DMAIC flow. This study adds to existing work reframing predictive functions or process improvements as disconnected from operational decision-making by linking prediction and operational decision-making through a holistic lens.

The performance of proposed framework is verified through simulation results. The model was simulated using a state-transition framework, which enables safe and controlled assessment of competing intervention strategies without disruption to clinical day-to-day practice. Of all the tested scenarios predictive staffing allocation had the greatest performance impact, further reinforcing the value of data driven resourcing. This showed that the excellent performance observed across all simulation scenarios was justified by the proposed method. This ensures that what we find is solid and not just bending under some random variability related issues (as discussed in our previous article): ANOVA based replication simulation.

This study adds to the existing literature with three main features:

· Prediction using Machine Learning

· Improvement of process (DMAIC).

· Validated by simulation."

This joint approach occupies a meaningful gap in the literature and builds a scalable framework on which to enhance operations within ED environments. In addition to contributing to the literature of data driven healthcare operations, this study illustrates that when predictive models, process improvement methodologies and simulation-based validation methods are viewed under a common lens that can provide new insights. This is critical to making analytical output actionable and sustainable in complex healthcare systems.

6. Conclusion

This study developed a multi-modal integrated approach combining machine learning–based predictive analytics, Six Sigma DMAIC process improvement methodology, and discrete-event simulation to understand how best to alleviate overcrowding in the emergency department. MIMIC-IV-ED data4 and context-informative annotated inputs resulted in the overall highest predictive performance including discriminative ability for other models noted that both Random Forest (top) and ensemble approaches were preferred. Calibration analysis validated that predicted probabilities were reliable, and SHAP-based explainability demonstrated important contributors of congestion was in accordance with logic (temporal features) while flow population was reflected by patient inflow.

DMAIC framework helped identify operational bottleneck systemically, it also recommend possible push or pull interventions to implementation towards production optimization. In the discrete-event simulation model, it provided a controlled environment in which to pitted against these strategies under relatively more natural circumstances without micromanagement of clinical operations. Outcomes From Simulation Model: Outcomes from the simulation model showed significant improvement in KPIs (eg, waiting time or length of stay and LWBS rates). To put it differently, ANOVA showed that these improvements were statistically significant rather than coincidental changes.

In fact, we are not aware of any studies that have managed to address these linkages between data modeling solutions and clinical practice/process adjustments in healthcare settings, so our proposed framework constitutes the bridge from data analytics to healthcare realities. The methodology described in this study combines predictive analytics, process optimization and simulation validation to create a scalable transferable framework that can potentially be deployed across other domains within health systems for enhancing operational efficiency and quality of care.

7. Limitations and future work

Although these results are exciting, there are a few caveats that we would have to consider. First of all, this study relies on a retrospective dataset (ie, MIMIC-IV-ED) from one healthcare system. This might affect the generalizability of results to other hospitals with different patient populations, resource availability, and workflow.

Second, while we cross-validated the simulation model against historical data to ensure its accuracy, it is a highly simplified reflection of the complex clinical care provided in a real-world emergency department. Finally, some attributes that influence a clinical pathway outcome (for example human behavior or the fatigue of personnel) are not modeled directly, and may in turn affect the actual performance of the system.

Third, the forecast models have been based solely on historical and context variables as inputs and may be inadequate to account for sudden changes to patient demand (for instance during a public health crisis or exogenous shock). Finally, your model's performance can begin to deteriorate over time unless it is continually retrained with new data.

Fourth, the study closely examines operational performance indices (eg. waiting time, LOS and LWBS) rather than screening patient-centered outcomes such clinical quality or even patient satisfaction or safety measures directly. First, costs associated with the implementation (staffing changes and workflow redesigns) of the proposed intervention were not taken into account. Assessing the cost-effectiveness of implementation is therefore a key step to establish whether this would, in fact, be feasible in practice.

Future research will need these(implementation studies) to evaluate the framework proposed, in these same clinical settings. And taking deployment of the predictive model to hospital information systems and decision-support platforms so as to factor in real-time congestion estimations and dynamically distribute resources. A multi-center study design across various healthcare systems will make the model generalizable in addition to being robust. The predictive performance could be expected significantly better due to further trialing in advanced machine learning approaches like deep learning and hybrid models. From a simulation perspective, further work could expand the model to incorporate more detailed data of patient pathways, staff behavior and departments electronic shift proportions. This could enable a whole range of set-level and deeper studies of current model-wise interactions via the integration of this type of agent-based modelling with discrete event simulation.

Future studies also should assess the effect of proposed interventions on patient-centered outcomes, including satisfaction as well as measures of clinical quality and safety. Having integrated cost-benefit and economic evaluation analysis would also be important to better guide the decision-making of healthcare administrators. Finally, the convergence of digital health technologies (real-time data-steams, Internet of Things (IoTs) devices and interoperable health information systems) will create a unique opportunity for fully adaptive congestion management systems capable of complementing current converging 21st century clinical care digital transformation initiatives.

Author Contributions: Mahmoud.F.Y: made the Conceptualization, Ghada Shaban made the Methodology, Software. Esraa.M.Hashem: made the Data curation, Writing original draft. Ghaidaa A. Khalid made the analysis. Both Mahmoud.F.Y , Ghada Shaban and Esraa.M.Hashem Writing review & editing.

Funding: Not applicable.

Data Availability Statement: yes, MIMIC-IV-ED, available at the following URL: https://doi.org/10.1111/j.1553-2712.2007.00032.x.

References

1. Abo, A., Kamel, B., Hashem, E. M., Mabrouk, M. S., & Fakhre, M. W. (2024). Predicting DNA Methylation state of CpG Islands Using Machine Learning. *Journal of Advanced Engineering Trends*, *43*(2), 11–17. https://doi.org/10.21608/JAET.2022.147975.1214
2. Ahn, I., Gwon, H., Kang, H., Kim, Y., Seo, H., Choi, H., Cho, H. N., Kim, M., Jun, T. J., & Kim, Y. H. (2021). Machine learning-based hospital discharge prediction for patients with cardiovascular diseases:development and usability study. *JMIR Medical Informatics*, *9*(11). https://doi.org/10.2196/32662,
3. Alhuraish, I., Robledo, C., & Kobi, A. (2017). A comparative exploration of lean manufacturing and six sigma in terms of their critical success factors. *Journal of Cleaner Production*, *164*, 325–337. https://doi.org/10.1016/J.JCLEPRO.2017.06.146
4. Asplin, B. R., Magid, D. J., Rhodes, K. V., Solberg, L. I., Lurie, N., & Camargo, C. A. (2003). A conceptual model of emergency department crowding. *Annals of Emergency Medicine*, *42*(2), 173–180. https://doi.org/10.1067/mem.2003.302
5. Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*, *8*(2), e188. https://doi.org/10.7861/FHJ.2021-0095
6. Beyer, H. (1981). Tukey, John W.: Exploratory Data Analysis. Addison-Wesley Publishing Company Reading, Mass. — Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 S. *Biometrical Journal*, *23*(4), 413–414. https://doi.org/10.1002/BIMJ.4710230408
7. Choi, S. W., Ko, T., Hong, K. J., & Kim, K. H. (2019). Machine Learning-Based Prediction of Korean Triage and Acuity Scale Level in Emergency Department Patients. *Healthcare Informatics Research*, *25*(4), 305–312. https://doi.org/10.4258/HIR.2019.25.4.305
8. Does, R. J. M. M., & Mast, J. De. (2023). Six Sigma Method. *Wiley StatsRef: Statistics Reference Online*, 1–2. https://doi.org/10.1002/9781118445112.stat04279.pub2
9. Douglass, M. J. J. (2020). Book Review: Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, 2nd edition by Aurélien Géron. *Physical and Engineering Sciences in Medicine*, *43*(3), 1135–1136. https://doi.org/10.1007/S13246-020-00913-Z
10. *European Union*. (n.d.).
11. Hashem, E. M., Shalaby, M. F. Y., Mabrouk, M. S., Marzouk, S. Y., & Marzouk, Y. (2023). Integrative software design for Hospital Information Systems. *Mansoura Engineering Journal*, *48*(4), 7. https://doi.org/10.58491/2735-4202.3057
12. Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE*, *13*(7). https://doi.org/10.1371/JOURNAL.PONE.0201016,
13. Hoot, N. R., & Aronsky, D. (2008a). Systematic review of emergency department crowding: causes, effects, and solutions. *Annals of Emergency Medicine*, *52*(2). https://doi.org/10.1016/J.ANNEMERGMED.2008.03.014
14. Hoot, N. R., & Aronsky, D. (2008b). Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine*, *52*(2), 126-136.e1. https://doi.org/10.1016/J.ANNEMERGMED.2008.03.014/ASSET/4607FCD0-1C63-4333-AD02-CFE5E576BF06/MAIN.ASSETS/GRA1.SML
15. Hoot, N. R., LeBlanc, L. J., Jones, I., Levin, S. R., Zhou, C., Gadd, C. S., & Aronsky, D. (2008). Forecasting Emergency Department Crowding: A Discrete Event Simulation. *Annals of Emergency Medicine*, *52*(2), 116–125. https://doi.org/10.1016/j.annemergmed.2007.12.011
16. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression: Third Edition. *Applied Logistic Regression: Third Edition*, 1–510. https://doi.org/10.1002/9781118548387

17. Hussein, N. A., Abdelmaguid, T. F., Tawfik, B. S., & Ahmed, N. G. S. (2017). Mitigating overcrowding in emergency departments using Six Sigma and simulation: A case study in Egypt. *Operations Research for Health Care*, *15*, 1–12. https://doi.org/10.1016/J.ORHC.2017.06.003

18. Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T. J., Moody, B., Gow, B., Lehman, L. wei H., Celi, L. A., & Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data 2023 10:1*, *10*(1), 1–9. https://doi.org/10.1038/s41597-022-01899-x.

19. Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2008).Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2), 159–170. https://doi.org/10.1111/j.1553-2712.2007.00032.x

20. Breiman, L. (2001).Random forests, *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.

21. Sudarshan, V. K., Brabrand, M., Range, T. M., & Wiil, U. K. (2021). Performance evaluation of Emergency Department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. *Computers in Biology and Medicine*, *135*. https://doi.org/10.1016/J.COMPBIOMED.2021.104541

22. Taner, M. T., Sezen, B., & Antony, J. (2007b). An overview of six sigma applications in healthcare industry. *International Journal of Health Care Quality Assurance*, *20*(4), 329–340. https://doi.org/10.1108/09526860710754398

23. Monday, L. M. (2022). Define, Measure, Analyze, Improve, Control (DMAIC) Methodology as a Roadmap in Quality Improvement. *Global Journal on Quality and Safety in Healthcare*, *5*(2), 44. https://doi.org/10.36401/JQSH-22-X2

24. Morley, C., Unwin, M., Peterson, G. M., Stankovich, J., & Kinsman, L. (2018). Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS ONE*, *13*(8), e0203316. https://doi.org/10.1371/JOURNAL.PONE.0203316

25. Mostafa, R., & El-Atawi, K. (2024). Strategies to Measure and Improve Emergency Department Performance: A Review. *Cureus*. https://doi.org/10.7759/CUREUS.52879

26. Zhang, Y., Zhang, J., Tao, M., Shu, J., & Zhu, D. (2022a). Forecasting patient arrivals at emergency department using calendar and meteorological information. *Applied Intelligence (Dordrecht, Netherlands)*, *52*(10), 11232. https://doi.org/10.1007/S10489-021-03085-9

27. Liabsuetrakul, T., Peeyananjarassri, K., Tassee, S., Sanguanchua, S., & Chaipinitpan, S. (2007). Emergency obstetric care in the southernmost provinces of Thailand. *International Journal for Quality in Health Care*, *19*(4), 250–256. https://doi.org/10.1093/INTQHC/MZM023

28. Sinreich, D., & Marmor, Y. (2005),Emergency department operations: The basis for developing a simulation tool.,*IIE Transactions*, 37(3), 233–245., https://doi.org/10.1080/07408170590919188.

29. Savioli, G., Ceresa, I. F., Gri, N., Piccini, G. B., Longhitano, Y., Zanza, C., Piccioni, A., Esposito, C., Ricevuti, G., & Bressan, M. A. (2022). Emergency Department Overcrowding: Understanding the Factors to Find Corresponding Solutions. *Journal of Personalized Medicine*, *12*(2), 279. https://doi.org/10.3390/JPM12020279

30. Savioli, G., Ceresa, I. F., Gri, N., Piccini, G. B., Longhitano, Y., Zanza, C., Piccioni, A., Esposito, C., Ricevuti, G., & Bressan, M. A. (2022). Emergency Department Overcrowding: Understanding the Factors to Find Corresponding Solutions. *Journal of Personalized Medicine*, *12*(2), 279. https://doi.org/10.3390/JPM12020279.

31. Morley, C., Unwin, M., Peterson, G. M., Stankovich, J., & Kinsman, L. (2018). Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS ONE*, 13(8), e0203316. https://doi.org/10.1371/journal.pone.0203316

32. Sun, Y., Heng, B. H., Seow, Y. T., & Seow, E. (2009a). Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine*, *9*. https://doi.org/10.1186/1471-227X-9-1
33. Zhou, Q., Li, Z., & Zhu, X. (2019).Machine learning-based prediction of emergency department overcrowding.*Healthcare Informatics Research*, 25(4), 331–338.https://doi.org/10.4258/hir.2019.25.4.331
34. Salem, D. A., Moharam, M. H., & Hashem, E. M. (2024). Development of Machine Learning Regression Models for Predicting the Performance of Nanofibrous Scaffolds for Skin Tissue Engineering. *Journal of Bio-X Research*, *7*. https://doi.org/10.34133/JBIOXRESEARCH.0008/ASSET/6CF6877F-321B-49E9-9509-BB5DF42C6099/ASSETS/GRAPHIC/JBIOXRESEARCH.0008.FIG.007.JPG.
35. Schaffer, A. L., Dobbins, T. A., & Pearson, S. A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, *21*(1), 58. https://doi.org/10.1186/S12874-021-01235-8
36. Trakulsunti, Y., Antony, J., Dempsey, M., & Brennan, A. (2021). Reducing medication errors using lean six sigma methodology in a Thai hospital: an action research study. *International Journal of Quality and Reliability Management*, *38*(1), 339–362. https://doi.org/10.1108/IJQRM-10-2019-0334
37. Vinutha, H.P., Poornima, B., Sagar, B.M. (2018). Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In: Satapathy, S., Tavares, J., Bhateja, V., Mohanty, J. (eds) Information and Decision Sciences. Advances in Intelligent Systems and Computing, vol 701. Springer, Singapore. https://doi.org/10.1007/978-981-10-7563-6_53
38. Taylor, C. B., Curtis, K., Jan, S., & Newcombe, M. (2013). Helicopter Emergency Medical Services (HEMS) over-triage and the financial implications for major trauma centres in NSW, Australia. *BMC Emergency Medicine*, *13*(1), 1–8. https://doi.org/10.1186/1471-227X-13-11/FIGURES/3
39. AM Abudher, IS El Hosin, EM Hashem, SY Marzouk , Mai S. Mabrouk ,Transforming Healthcare Quality In Libya: An Evaluation Of TQM Adoption And Performance Outcomes, Cultura: International Journal of Philosophy of Culture, 2026 , https://doi.org/10.5281/zenodo.18454240.
40. Taner, M. T., Sezen, B., & Antony, J. (2007c). An overview of six sigma applications in healthcare industry. *International Journal of Health Care Quality Assurance*, *20*(4), 329–340. https://doi.org/10.1108/09526860710754398/FULL/XML
41. Elkahwagy, D.M.A.S., Kiriacos, C.J. & Mansour, M. Logistic regression and other statistical tools in diagnostic biomarker studies. *Clin Transl Oncol* **26**, 2172–2180 (2024). https://doi.org/10.1007/s12094-024-03413-8
42. Savioli, G., Ceresa, I. F., Gri, N., Piccini, G. B., Longhitano, Y., Zanza, C., Piccioni, A., Esposito, C., Ricevuti, G., & Bressan, M. A. (2022). Emergency Department Overcrowding: Understanding the Factors to Find Corresponding Solutions. *Journal of Personalized Medicine*, *12*(2), 279. https://doi.org/10.3390/JPM12020279
43. Vanbrabant L., Braekers K., Ramaekers K., Van Nieuwenhuyse I. Simulation of emergency department operations: A comprehensive review of KPIs and operational improvements. Comput. Ind. Eng. 2019;131:356–381. doi: 10.1016/j.cie.2019.03.025.
44. Abdumalikov, S., Kim, J., & Yoon, Y. (2024). Performance Analysis and Improvement of Machine Learning with Various Feature Selection Methods for EEG-Based Emotion Classification. *Applied Sciences*, *14*(22), 10511. https://doi.org/10.3390/app142210511.
45. Badawy, M., Ramadan, N. & Hefny, H.A. Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Inf Technol* **10**, 40 (2023). https://doi.org/10.1186/s43067-023-00108-y
46. Al-Turaiki I, Almutlaq F, Alrasheed H, Alballa N. Empirical Evaluation of Alternative Time-Series Models for COVID-19 Forecasting in Saudi Arabia. Int J Environ Res Public Health. 2021 Aug 16;18(16):8660. doi: 10.3390/ijerph18168660. PMID: 34444409; PMCID: PMC8393561.

47. Kane M.J., Price N., Scotch M., Rabinowitz P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinform. 2014;15:276. doi: 10.1186/1471-2105-15-276
48. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc. 2020 Apr 1;27(4):621-633. doi: 10.1093/jamia/ocz228.
49. Johanna Munoz, Karel G.M. Moons, Thomas P.A. Debray et al. Deriving calibration plots in large clustered datasets, 04 June 2025, PREPRINT (Version 1) available at Research Square https://doi.org/10.21203/rs.3.rs-6740452/v1.
50. Merrick, L., Taly, A. (2020). The Explanation Game: Explaining Machine Learning Models Using Shapley Values. In: Holzinger, A., Kieseberg, P., Tjoa, A., Weippl, E. (eds) Machine Learning and Knowledge Extraction. CD-MAKE 2020. Lecture Notes in Computer Science(), vol 12279. Springer, Cham. https://doi.org/10.1007/978-3-030-57321-8_2
51. Feretzakis, G., Sakagianni, A., Anastasiou, A., Kapogianni, I., Bazakidou, E., Koufopoulos, P., Koumpouros, Y., Koufopoulou, C., Kaldis, V., & Verykios, V. S. (2024). Integrating Shapley Values into Machine Learning Techniques for Enhanced Predictions of Hospital Admissions. *Applied Sciences*, *14*(13), 5925. https://doi.org/10.3390/app14135925
52. Mostofian B, Zuckerman DM. Statistical Uncertainty Analysis for Small-Sample, High Log-Variance Data: Cautions for Bootstrapping and Bayesian Bootstrapping. J Chem Theory Comput. 2019 Jun 11;15(6):3499-3509. doi: 10.1021/acs.jctc.9b00015.
53. Sen, K., & Al-Faisal, W. (2013). Reforms and emerging noncommunicable disease: some challenges facing a conflict-ridden country—the case of the Syrian Arab Republic. *The International Journal of Health Planning and Management*, *28*(3), 290–302. https://doi.org/10.1002/HPM.2193.
54. Vázquez-Serrano JI, Peimbert-García RE, Cárdenas-Barrón LE. Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review. Int J Environ Res Public Health. 2021 Nov 22;18(22):12262. doi: 10.3390/ijerph182212262.
55. Smalley CM, Meldon SW, Simon EL, Muir MR, Delgado F, Fertel BS. Emergency Department Patients Who Leave Before Treatment Is Complete. West J Emerg Med. 2021 Feb 26;22(2):148-155. doi: 10.5811/westjem.2020.11.48427
56. https://www.flexsim.com/healthcare/
57.
58. *Time series modelling and forecasting of emergency department overcrowding | Request PDF*. (n.d.). Retrieved May 26, 2025, from https://www.researchgate.net/publication/278962406_Time_series_modelling_and_forecasting_of_emergency_department_overcrowding
59. Tissir, S., Cherrafi, A., Chiarini, A., Elfezazi, S., & Bag, S. (2023). Lean Six Sigma and Industry 4.0 combination: scoping review and perspectives. *Total Quality Management and Business Excellence*, *34*(3–4), 261–290. https://doi.org/10.1080/14783363.2022.2043740
60. Wargon, M., Guidet, B., Hoang, T. D., & Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal : EMJ*, *26*(6), 395–399. https://doi.org/10.1136/EMJ.2008.062380
61. Wilson, G. T. (2016). Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *Journal of Time Series Analysis*, *37*(5), 709–711. https://doi.org/10.1111/JTSA.12194