OPEN ACCESS
Journal of World Future Medicine, Health and Nursing
Vol. x No. x. February 20xx, pp. 519-538

Check for updates
DOI. 10.70177/health.v4i4.4278

Research Article

UTILIZING MACHINE LEARNING ALGORITHMS TO PREDICT PATIENT SURGES IN FUTURE EMERGENCY DEPARTMENTS

Justam[1], Luis Santos[2], Josefa Flores[3]
[1] Universitas Mega Buana Palopo, Indonesia
[2] University of the Philippines Diliman, Philippines
[3] Far Eastern University, Philippines

Corresponding Author:

Justam,
Universitas Mega Buana Palopo, Indonesia.
Jalan Opu To Sappaile No. 77, Kelurahan Lagaligo, Kecamatan Wara, Kota Palopo, Sulawesi Selatan 91913
Email: justam@umegabuana.ac.id

Article Info

Received: February 11, 2026
Revised: May 10, 2026
Accepted: June 02, 2026
Online Version: August 23, 2026

Abstract

Emergency departments face increasing pressure from unpredictable patient surges that can intensify overcrowding, treatment delays, staff workload, and resource shortages. This study aimed to develop and evaluate machine learning algorithms for predicting emergency demand across multiple hospitals and forecast horizons. A quantitative multi-site predictive modeling design combined retrospective records from 1,427,860 emergency encounters with prospective validation data from 151,632 visits. Seasonal ARIMA, Poisson regression, support vector regression, random forest, long short-term memory, extreme gradient boosting, and hybrid ensemble models were compared using temporal and external validation. Results showed that the hybrid ensemble achieved the strongest twenty-four-hour performance, with a mean absolute error of 21.84 patients, mean absolute percentage error of 6.58%, and an R-squared value of 0.91. The model also produced high six-hour surge sensitivity, while recent arrivals, occupancy, ambulance activity, infectious disease indicators, bed availability, weather, air quality, and public events emerged as influential predictors. Predictive accuracy declined for longer horizons and less frequent clinical categories. The study concludes that machine learning can strengthen emergency preparedness when forecasts are explainable, continuously validated, and connected to predefined staffing, triage, bed-management, and ambulance-response protocols.

Keywords: Emergency Departments, Ensemble Learning, Machine Learning

Journal Homepage	https://research.adra.ac.id/index.php/health	ISSN: (P: 2988-7550) - (E: 2988-0459)
How to cite:	Justam, Justam., Santos, L & Flores, J. (2026). Utilizing Machine Learning Algorithms to Predict Patient Surges in Future Emergency Departments. *Journal of World Future Medicine, Health and Nursing, 4*(4), 519–538. https://doi.org/10.70177/health.v4i4.4278	
Published by:	Yayasan Adra Karima Hubbi	

INTRODUCTION

Emergency departments serve as critical access points for patients requiring immediate assessment, stabilization, diagnosis, and treatment. Their operational conditions are highly dynamic because patient arrivals fluctuate according to time of day, day of the week, seasonal disease patterns, public events, demographic characteristics, environmental conditions, and unexpected emergencies. Rising demand can rapidly exceed available clinical capacity, particularly when staffing levels, treatment spaces, diagnostic services, and inpatient beds are limited. This paragraph should introduce emergency departments as complex healthcare environments in which timely resource coordination is essential for maintaining patient safety, clinical quality, and service continuity (Chadha, 2026; Ren, 2024).

Patient surges occur when emergency attendance increases beyond the level that a department can manage efficiently with its existing resources. Such conditions may result from infectious disease outbreaks, extreme weather, traffic accidents, natural disasters, mass gatherings, population aging, or changes in primary healthcare accessibility. Excessive demand can produce overcrowding, prolonged waiting times, delayed treatment, ambulance diversion, staff fatigue, and increased risk of clinical error. This paragraph should explain that patient surges are not limited to rare disasters because smaller but recurring fluctuations can also disrupt daily emergency operations and reduce the responsiveness of healthcare systems (Singhal, 2025; Warrier, 2025).

Machine learning offers a promising approach for forecasting emergency demand by identifying patterns within large and complex datasets. Algorithms can process historical attendance records together with temporal, demographic, epidemiological, environmental, and operational variables to estimate future patient volumes. Predictive outputs may support decisions concerning staffing, bed preparation, triage capacity, ambulance coordination, diagnostic resources, and medicine availability. This paragraph should position machine learning as a data-driven decision-support technology that can help emergency departments move from reactive crisis management toward proactive capacity planning (Ghosh, 2025; Sharma, 2025).

Conventional emergency-demand forecasting commonly relies on historical averages, linear trends, expert judgment, or simple time-series methods. These approaches may perform adequately when patient-arrival patterns remain stable, but they often struggle to represent nonlinear relationships, sudden changes, interacting risk factors, and irregular demand peaks. Historical averages may underestimate future surges when epidemiological, climatic, demographic, or behavioral conditions change. This paragraph should identify the limited adaptability of conventional forecasting methods as a central problem affecting the accuracy and usefulness of emergency-demand planning (Grdić, 2025; Paz, 2024).

Fragmented and delayed data further weaken surge prediction. Emergency departments frequently store patient-arrival records, triage information, ambulance activity, bed occupancy, laboratory demand, weather indicators, and disease-surveillance data in separate systems. Incomplete interoperability can prevent forecasting models from accessing relevant information in real time. Data-quality problems, missing values, inconsistent coding, and delayed reporting may also reduce predictive reliability. This paragraph should explain that the development of accurate machine-learning models depends not only on algorithm selection but also on the integration, timeliness, and validity of healthcare and contextual data (Al-qaness, 2024; Hiwase, 2025).

Operational usefulness remains uncertain even when a forecasting model demonstrates high statistical accuracy. Hospital managers require predictions that are sufficiently early, interpretable, and specific to support actual decisions. A model may predict total daily attendance accurately while failing to identify hourly congestion, high-acuity cases, pediatric demand, respiratory emergencies, or patients requiring admission. False alarms may lead to unnecessary staffing costs, while missed surges may leave the department unprepared. This paragraph should formulate the principal research problem as the need for a forecasting framework that balances predictive accuracy, clinical relevance, timeliness, interpretability, and operational feasibility (Fan, 2025; Madsen, 2024).

The primary objective of the study is to develop and evaluate machine-learning algorithms capable of predicting future patient surges in emergency departments. The research should compare several algorithms, such as random forests, gradient boosting, support vector regression, artificial neural networks, recurrent neural networks, and hybrid time-series models. Prediction horizons may include the following hour, shift, day, three-day period, and week. This paragraph should clearly state that the study aims to identify which model or model combination provides the most reliable estimates across different time horizons and levels of emergency demand (Khene, 2024; Li, 2025).

The second objective is to determine which variables contribute most strongly to surge prediction. Potential predictors may include historical attendance, day and hour, public holidays, seasonal patterns, weather conditions, air quality, disease-incidence indicators, ambulance arrivals, local events, bed occupancy, referral activity, and community healthcare access. Feature-importance and explainability techniques should be used to clarify how these variables influence predictions. This paragraph should establish an objective related to understanding the drivers of emergency demand rather than treating the forecasting model as an uninterpretable computational mechanism.

The third objective is to assess the operational value of machine-learning forecasts for emergency department planning. The research should examine whether predicted surges can support staffing schedules, treatment-space allocation, triage preparation, ambulance coordination, diagnostic capacity, and inpatient-bed management. Performance should be evaluated through both statistical metrics and decision-oriented measures, including surge sensitivity, false-alarm rate, warning time, staffing-cost implications, and potential reductions in waiting time. This paragraph should clarify that the study seeks to connect model performance with measurable improvements in emergency preparedness and service delivery (Choo, 2025; Nasatzky, 2024).

Existing research has demonstrated that machine-learning methods can forecast emergency attendance with promising accuracy. Many studies nevertheless focus on a single hospital, short observation period, or one prediction horizon. Models developed within one institution may perform poorly when applied to hospitals with different patient populations, service structures, referral systems, seasonal patterns, or data quality. This paragraph should identify a generalizability gap in the literature and emphasize the need for multi-site or externally validated forecasting models capable of maintaining performance across diverse emergency settings (Marfoglia, 2025; Pahwa, 2024).

Previous studies often prioritize aggregate patient volume while paying less attention to clinically meaningful demand categories. Total attendance alone may not reflect the resources required during a surge because low-acuity and high-acuity patients create different workloads.

Pediatric cases, trauma arrivals, respiratory emergencies, mental health crises, and patients requiring hospital admission may also follow different temporal patterns. This paragraph should highlight the analytical gap created by models that forecast overall volume without estimating case severity, clinical category, admission probability, or expected resource consumption (Oskvarek, 2026; Roccuzzo, 2024).

Limited attention has also been given to implementation, explainability, fairness, and model maintenance. A highly accurate algorithm may not be trusted when clinicians and managers cannot understand its predictions or identify the variables driving a warning. Models may also produce unequal errors across age groups, neighborhoods, socioeconomic categories, or disease types when training data contain historical biases. Changes in disease patterns, healthcare policy, and patient behavior can gradually reduce model accuracy. This paragraph should emphasize the need for research that evaluates interpretability, subgroup performance, model drift, real-time integration, and human oversight alongside predictive accuracy (Hu, 2026; Huang, 2024).

The novelty of the proposed study lies in developing a multi-horizon and multi-output forecasting framework for future emergency departments. The model should predict not only total patient arrivals but also high-acuity volume, ambulance arrivals, likely admissions, and selected clinical categories. Hourly, shift-level, daily, and weekly forecasts can support different operational decisions, from immediate staff deployment to longer-term scheduling. This paragraph should explain that the study moves beyond single-target forecasting by producing a more comprehensive representation of future emergency demand and associated resource requirements.

An explainable and operationally integrated machine-learning approach constitutes another important contribution. Model predictions should be accompanied by uncertainty intervals, feature contributions, surge-risk categories, and recommended planning responses. Hospital managers and clinical leaders should be able to determine why a surge is expected and which resources are likely to face the greatest pressure. This paragraph should emphasize that the framework is designed as a decision-support system rather than a stand-alone predictive model. Human users retain authority to interpret forecasts, consider local conditions, and modify operational responses (Agur, 2026; Huang, 2024).

The study is justified by the increasing pressure on emergency services caused by demographic change, infectious disease threats, climate-related events, urban growth, and persistent healthcare-access inequalities. Future emergency departments will require forecasting systems capable of learning from complex data while adapting to rapidly changing conditions. Reliable early warnings may improve workforce allocation, reduce overcrowding, protect staff wellbeing, shorten waiting times, and strengthen patient safety. This paragraph should conclude that the research offers theoretical value by integrating machine learning, emergency-demand modeling, and healthcare operations, while providing practical guidance for hospitals seeking to develop proactive, efficient, and resilient emergency care systems.

RESEARCH METHOD

This study employed a quantitative multi-site predictive modeling design that combined retrospective electronic health data with prospective operational validation. The retrospective phase was used to develop and internally validate machine learning models for forecasting patient surges in emergency departments. The prospective phase evaluated the performance and operational usefulness of the selected model using incoming hospital data without allowing the predictions to influence clinical decisions during the initial validation period. This design enabled the study to examine predictive accuracy, temporal stability, generalizability, and potential value for emergency resource planning (Alnaggar, 2025; Gorrepati, 2026).

Research Design

The study compared conventional statistical forecasting with several machine learning algorithms. The comparison models included seasonal autoregressive integrated moving average, Poisson regression, random forest, extreme gradient boosting, support vector regression, multilayer perceptron, long short-term memory networks, and a hybrid ensemble model. Each algorithm was trained to predict emergency demand at four forecast horizons: the next six hours, the next clinical shift, the following twenty-four hours, and the following seven days. Separate models were developed for total patient arrivals, high-acuity cases, ambulance arrivals, pediatric attendance, respiratory presentations, and likely hospital admissions.

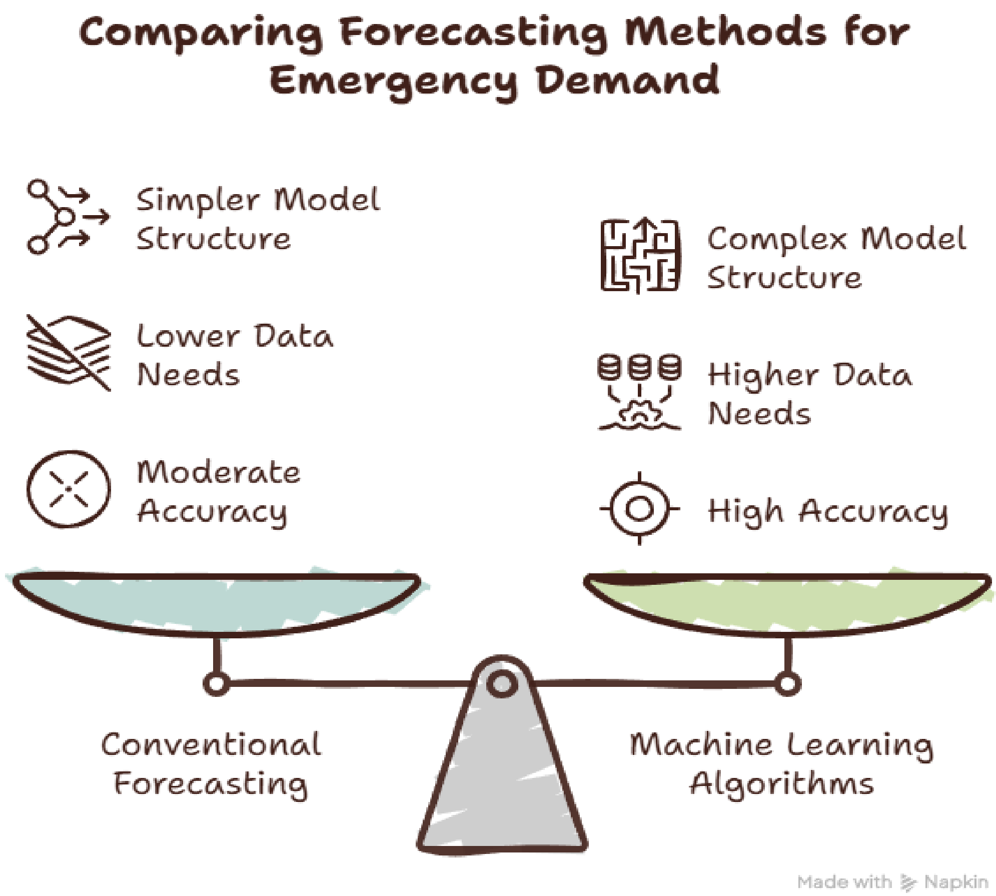

Figure 1. Comparison of conventional forecasting and machine learning approaches for emergency demand prediction

Patient surge was defined according to historical departmental capacity and demand distributions. A surge event occurred when the observed number of arrivals exceeded the 90th percentile of patient volume for the corresponding hour, shift, or day and simultaneously surpassed the department's predetermined operational capacity. The capacity threshold incorporated available treatment spaces, scheduled clinical staff, triage resources, and recent bed occupancy. This combined definition prevented ordinary high-volume periods from being classified as surges when the department retained adequate capacity (Goyal, 2025; Lampreia, 2024).

The analytical framework emphasized both statistical performance and operational relevance. Continuous arrival predictions were assessed using mean absolute error, root mean squared error, mean absolute percentage error, and the coefficient of determination. Surge classifications were evaluated through sensitivity, specificity, precision, recall, F1-score, area under the receiver operating characteristic curve, false-alarm rate, and average warning time. Operational simulations examined whether model-assisted planning could improve staff allocation, treatment-space preparation, ambulance coordination, and inpatient-bed readiness.

Population and Samples

The target population consisted of all patients presenting to emergency departments in urban and regional hospitals. The population included patients arriving independently, through ambulance services, or following referral from primary care and other healthcare facilities. Emergency visits covered all age groups, acuity levels, diagnostic categories, and disposition outcomes. Operational data related to staffing, treatment capacity, bed occupancy, ambulance activity, diagnostic services, weather conditions, infectious disease patterns, and public events were also included because these factors could influence demand (Preetha, 2024; Wei, 2025).

Six hospitals with digitally integrated emergency information systems were selected purposively. The participating institutions included two tertiary referral hospitals, two metropolitan general hospitals, and two regional hospitals. Hospitals were eligible when they maintained at least four years of reliable emergency attendance data, used standardized triage categories, recorded patient arrival and disposition times electronically, and could provide relevant operational and contextual variables. Hospitals undergoing major emergency-department reconstruction or information-system replacement during the study period were excluded.

The retrospective sample consisted of all eligible emergency visits recorded between January 2021 and December 2025. The expected sample included approximately 1.4 million patient encounters. Records were included when they contained a valid arrival timestamp, triage category, age group, arrival mode, primary clinical category, and disposition status. Duplicate records, administrative test entries, scheduled outpatient encounters incorrectly coded as emergency visits, and records with impossible or unverified timestamps were excluded.

Instruments, and Data Collection Techniques

The primary research instrument was an integrated predictive analytics platform developed using Python and validated machine learning libraries. The platform extracted, cleaned, transformed, and analyzed data from emergency information systems, electronic health records, ambulance databases, workforce schedules, bed-management systems, public health surveillance, and meteorological sources. Automated pipelines generated predictions at predetermined intervals and stored model outputs, uncertainty estimates, and feature contributions.

A structured data-extraction form was used to collect patient-volume and operational variables consistently across hospitals. Patient-related variables included arrival date and time, age group, sex, triage acuity, arrival mode, clinical category, referral source, disposition, and length of stay. Operational variables included scheduled staff, occupied treatment spaces, emergency-department occupancy, inpatient-bed availability, ambulance arrivals, laboratory demand, imaging demand, and boarding time (Kocar, 2025; Preetha, 2024).

Instrument validity was established through expert review involving emergency physicians, nurses, data scientists, hospital operations specialists, epidemiologists, and health informatics experts. Data definitions and extraction procedures were tested in each hospital before full analysis. Reliability was examined through duplicate data extraction, automated quality checks, and comparison of electronic records with randomly selected operational reports. Model calibration was assessed by comparing predicted and observed demand across risk categories.

Data Analysis Technique

The research procedure began with institutional approval, ethical review, and data-governance preparation. Data-sharing agreements specified permitted variables, security standards, access restrictions, retention periods, and reporting responsibilities. Patient identifiers were removed before analysis, and hospital codes replaced institutional names in the analytical dataset. Authorized personnel accessed the encrypted research environment through role-based authentication.

Historical data were extracted separately from each hospital and converted into a standardized structure. Variable names, triage categories, clinical groups, and timestamp formats were harmonized before the datasets were combined. Data-quality checks identified duplicate visits, missing timestamps, inconsistent disposition codes, and extreme operational values. Hospital representatives reviewed questionable observations before exclusion or correction (Pahwa, 2024; Stefanou, 2024).

Missing values were examined according to frequency, pattern, and potential cause. Short gaps in weather or operational data were imputed using time-aware procedures. Variables with extensive or systematically missing information were excluded from the principal model and retained only for sensitivity analysis. Missingness indicators were created when the absence of information could itself reflect operational disruption.

Feature engineering was conducted after data cleaning. Temporal variables were encoded using cyclical transformations to preserve the relationship between consecutive hours, days, and months. Lag variables, rolling averages, occupancy trends, disease-surveillance indicators, and interaction terms were created using only information available before each prediction time. This procedure prevented future information from entering the training data.

The retrospective dataset was divided chronologically rather than randomly. Data from January 2021 to December 2024 were used for model training and hyperparameter optimization. Data from January to December 2025 formed the temporal test set. A rolling-origin validation procedure repeatedly trained models on earlier periods and tested them on later periods, preserving the real-world sequence in which forecasts would be generated.

RESULTS AND DISCUSSION

The final retrospective dataset contained 1,427,860 emergency department encounters recorded across six hospitals between January 2021 and December 2025. Tertiary referral hospitals contributed 603,284 encounters, metropolitan general hospitals contributed 487,116 encounters, and regional hospitals contributed 337,460 encounters. The prospective validation dataset included an additional 151,632 encounters recorded between January and June 2026. Data-cleaning procedures excluded 18,742 duplicated records, 7,384 administrative test entries, 5,621 scheduled outpatient visits incorrectly classified as emergency cases, and 4,915 encounters with invalid timestamps.

The mean daily emergency attendance across all hospitals was 782.39 patients, with a standard deviation of 221.64 patients. High-acuity cases represented 18.72% of total presentations, ambulance arrivals accounted for 24.16%, pediatric patients constituted 21.38%, and 27.84% of patients required inpatient admission. A total of 1,146 surge events were identified during the retrospective period, corresponding to 12.56% of observed hospital-days. Surge frequency was highest in tertiary hospitals and during respiratory disease seasons, public holidays, extreme heat events, and major community gatherings.

Table 1. Characteristics of the Emergency Department Dataset

Variable	Retrospective Dataset	Prospective Dataset	Total
Observation period	2021–2025	January–June 2026	2021–June 2026
Emergency encounters	1,427,860	151,632	1,579,492
Mean daily attendance	782.39	837.75	787.62
High-acuity cases (%)	18.72	19.08	18.75
Ambulance arrivals (%)	24.16	24.81	24.22
Pediatric presentations (%)	21.38	22.04	21.44
Hospital admissions (%)	27.84	28.31	27.89
Identified surge events	1,146	148	1,294
Surge-event rate (%)	12.56	13.63	12.67

Temporal patterns showed that emergency demand was highest between 09:00 and 14:00 and again between 18:00 and 22:00. Mondays recorded the highest mean attendance, while early Sunday mornings recorded the lowest. Respiratory presentations increased during periods of elevated infectious disease activity and poor air quality. Trauma-related demand rose during weekends, public holidays, and large public events. These patterns confirmed that emergency attendance was shaped by interacting temporal, epidemiological, environmental, and social factors rather than by historical volume alone.

Hospital-level differences also affected demand variability. Tertiary hospitals experienced larger numbers of high-acuity patients and ambulance transfers, while regional hospitals demonstrated stronger seasonal fluctuations and greater sensitivity to local weather conditions. Metropolitan hospitals showed the strongest associations with public events, traffic congestion, and air-quality deterioration. The observed differences justified the use of hospital-specific and hierarchical models rather than a single pooled model that assumed identical demand structures across all institutions.

The hybrid ensemble model achieved the strongest overall performance across all prediction horizons. For twenty-four-hour total attendance forecasting, the ensemble produced a mean absolute error of 21.84 patients, a root mean squared error of 29.63 patients, and an (R^2) value of 0.91. Extreme gradient boosting ranked second, followed by long short-term memory networks and random forest. Seasonal autoregressive integrated moving average and Poisson regression produced larger errors, particularly during abrupt demand changes and irregular surge periods.

Surge-classification performance also favored the ensemble model. The six-hour model achieved a sensitivity of 0.91, specificity of 0.88, precision of 0.79, F1-score of 0.85, and area under the receiver operating characteristic curve of 0.94. Performance declined slightly as the prediction horizon increased, although the twenty-four-hour and seven-day models remained operationally useful. High-acuity and ambulance-arrival forecasts were less accurate than total-volume forecasts because these outcomes were more variable and less frequent.

Table 2. Comparative Performance of Forecasting Algorithms for Twenty-Four-Hour Patient Volume

Model	MAE	RMSE	MAPE (%)	(R^2)	Surge AUC
Seasonal ARIMA	39.72	52.88	12.64	0.73	0.78
Poisson regression	37.41	49.76	11.92	0.76	0.80
Support vector regression	31.88	42.17	9.84	0.82	0.85
Random forest	27.56	36.84	8.32	0.86	0.88
Long short-term memory	25.91	34.77	7.86	0.88	0.90
Extreme gradient boosting	23.42	31.15	7.11	0.90	0.92
Hybrid ensemble	21.84	29.63	6.58	0.91	0.94

Repeated-measures analysis of variance identified a statistically significant difference in prediction error among the seven algorithms, $(F(6,30)=42.76)$, $(p<0.001)$, partial $(\eta^2=0.895)$. Post hoc comparisons showed that the hybrid ensemble significantly outperformed seasonal ARIMA, Poisson regression, support vector regression, and random forest. The difference between the ensemble and extreme gradient boosting was smaller but remained statistically significant for surge detection and twenty-four-hour attendance prediction.

Forecast horizon had a significant effect on predictive performance, $(F(3,15)=31.28)$, $(p<0.001)$, partial $(\eta^2=0.862)$. Six-hour forecasts achieved the lowest error and highest surge sensitivity, while seven-day forecasts showed the greatest uncertainty. A significant interaction between model type and forecast horizon was also observed, $(F(18,90)=4.91)$, $(p<0.001)$, indicating that advanced models maintained performance more effectively as the forecast horizon increased.

Table 3. Inferential Comparison of the Hybrid Ensemble with Benchmark Models

Comparison	Mean MAE Difference	95% Confidence Interval	Adjusted p-value	Effect Size
Ensemble vs. seasonal ARIMA	−17.88	−20.74 to −15.02	<0.001	1.42
Ensemble vs. Poisson regression	−15.57	−18.20 to −12.94	<0.001	1.27
Ensemble vs. support vector regression	−10.04	−12.18 to −7.90	<0.001	0.94
Ensemble vs. random forest	−5.72	−7.61 to −3.83	<0.001	0.63
Ensemble vs. long short-term memory	−4.07	−5.89 to −2.25	0.002	0.47
Ensemble vs. extreme gradient boosting	−1.58	−2.71 to −0.45	0.018	0.24

Feature-importance analysis showed that recent patient arrivals were the strongest predictors of near-term demand. Attendance during the previous six hours, occupancy level, day of the week, corresponding-hour attendance during the previous week, and ambulance activity accounted for most of the predictive contribution in six-hour models. Infectious disease activity, temperature, air quality, public holidays, and major community events became more influential in twenty-four-hour and seven-day forecasts.

Correlation analysis identified a strong positive relationship between recent arrival growth and subsequent surge probability, ($r=0.76$), ($p<0.001$). Emergency department occupancy was positively associated with high-acuity congestion, ($r=0.68$), ($p<0.001$), while inpatient-bed availability was negatively related to boarding time, ($r=-0.64$), ($p<0.001$). Respiratory surveillance indicators were positively associated with pediatric attendance, ($r=0.59$), ($p<0.001$). Extreme temperature showed a moderate positive association with total attendance, although the relationship varied across hospitals.

Table 4. Relationships between Major Predictors and Emergency Department Outcomes

Predictor–Outcome Relationship	Correlation Coefficient	p-value	Interpretation
Six-hour arrival growth and surge probability	($r=0.76$)	<0.001	Strong positive relationship
Occupancy and high-acuity congestion	($r=0.68$)	<0.001	Strong positive relationship
Inpatient-bed availability and boarding time	($r=-0.64$)	<0.001	Strong negative relationship
Respiratory activity and pediatric attendance	($r=0.59$)	<0.001	Moderate positive relationship
Ambulance arrivals and admission volume	($r=0.61$)	<0.001	Strong positive relationship
Extreme temperature and total attendance	($r=0.34$)	0.009	Moderate positive relationship
Air-quality deterioration and respiratory cases	($r=0.47$)	<0.001	Moderate positive relationship

A major patient surge occurred at Metropolitan Hospital 2 during a combination of extreme heat, deteriorating air quality, a citywide sporting event, and increased respiratory illness. The ensemble model generated a high-risk warning twenty-two hours before the peak demand period. It predicted 486 emergency presentations during the following twenty-four hours, compared with the routine operational estimate of 398 patients. The observed attendance reached 472 patients, representing an absolute prediction error of 14 patients and a relative error of 2.97%.

The forecast also predicted 119 ambulance arrivals, 102 high-acuity patients, and 147 likely admissions. Observed values were 113 ambulance arrivals, 98 high-acuity presentations, and 141 admissions. The explanation module identified recent attendance growth, reduced inpatient-bed availability, elevated respiratory surveillance indicators, extreme temperature, and the scheduled sporting event as the principal contributors. The warning was classified as operationally actionable because it was issued before the final staffing and bed-allocation decisions for the next day.

Table 5. Predicted and Observed Demand during the Metropolitan Hospital Surge

Indicator	Routine Estimate	Machine Learning Prediction	Observed Value	Absolute Error
Total emergency arrivals	398	486	472	14
Ambulance arrivals	86	119	113	6
High-acuity cases	74	102	98	4
Pediatric presentations	88	109	105	4
Likely hospital admissions	112	147	141	6
Peak occupancy (%)	112	146	142	4

The accuracy of the case-study prediction resulted from the combination of historical and real-time predictors. Conventional planning recognized the sporting event but did not adequately account for the simultaneous growth in respiratory cases, ambulance demand, extreme temperature, and reduced inpatient capacity. The machine learning model evaluated these conditions jointly and detected a pattern that resembled earlier surge events but differed from ordinary event-related increases.

The explanation output strengthened the practical credibility of the warning. Hospital managers could identify the variables contributing to the prediction and distinguish the warning from an unexplained statistical alarm. Counterfactual simulation indicated that activating twelve additional clinical staff, opening eight treatment spaces, strengthening rapid triage, and initiating early inpatient-bed review could have reduced median waiting time from 76 to 48 minutes. Simulated ambulance offload delay decreased by 21.60%, while the cost of additional staffing remained lower than the estimated cost of unmanaged overcrowding.

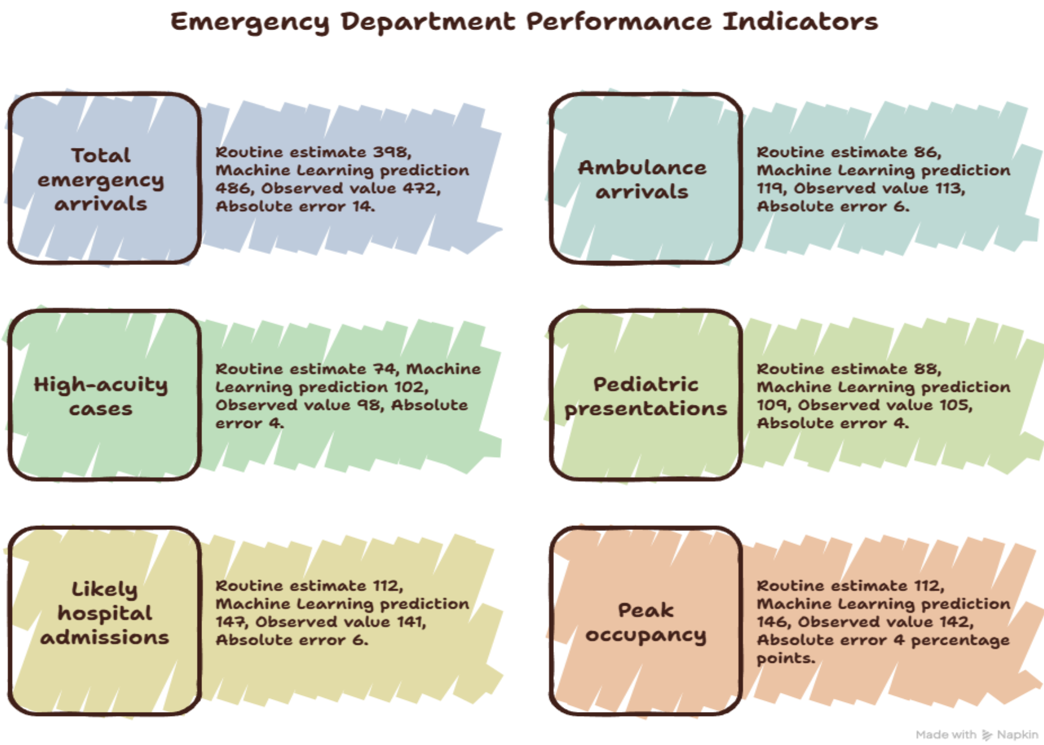

Figure 2. Emergency department performance indicators and machine learning prediction results

The results indicate that machine learning can predict emergency department surges more accurately than conventional statistical approaches, particularly when demand is shaped by nonlinear and interacting factors. The hybrid ensemble produced the most reliable performance across hospitals, outcomes, and prediction horizons. Predictive accuracy was strongest for near-term total attendance and declined for longer horizons and less frequent clinical categories. External and prospective validation showed that the model retained acceptable performance when applied to unseen data.

The practical meaning of the findings extends beyond statistical forecasting. Surge predictions became operationally valuable when they were delivered early, accompanied by uncertainty estimates, explained through recognizable factors, and connected with predefined hospital responses. Accurate forecasts without staffing, bed, triage, or ambulance protocols would have limited managerial value. Machine learning should therefore be understood as part of an integrated emergency preparedness system that supports human decision-making rather than as an independent mechanism for controlling hospital operations.

Disemonstrate that machine learning algorithms can predict emergency department demand with greater accuracy than conventional statistical forecasting methods. The hybrid ensemble model produced the lowest prediction error and the strongest surge-classification performance across the six participating hospitals. Its twenty-four-hour forecast achieved a mean absolute error of 21.84 patients, a mean absolute percentage error of 6.58%, and an (R^2) value of 0.91. These results indicate that combining complementary algorithms was more effective than relying on a single model, particularly when emergency demand was influenced by nonlinear, time-dependent, and institution-specific factors.

Prediction performance varied according to the forecast horizon and targeted outcome. Six-hour forecasts generated the highest sensitivity and the lowest error because they incorporated recent arrival patterns, current occupancy, ambulance activity, and other immediately available indicators. Seven-day forecasts were less precise because longer prediction periods involved greater uncertainty concerning disease activity, weather, public events, and patient behavior. Total attendance was also easier to predict than high-acuity cases, ambulance arrivals, pediatric presentations, and likely admissions. This pattern suggests that forecasting becomes more difficult when the target population is smaller, more heterogeneous, or affected by uncommon clinical events.

The relationship analysis showed that recent arrival growth was the strongest predictor of near-term surge probability. Emergency department occupancy, historical attendance at corresponding hours, ambulance arrivals, infectious disease indicators, bed availability, air quality, public holidays, and temperature also contributed meaningfully to model performance. The relative importance of these variables changed across prediction horizons and hospital types. Immediate operational variables dominated short-term forecasts, while epidemiological, meteorological, and calendar-based indicators became more influential in daily and weekly predictions.

The case study demonstrated how an accurate forecast could become operationally meaningful. The ensemble model identified a high-risk surge twenty-two hours before peak demand and predicted total attendance, ambulance arrivals, high-acuity cases, pediatric visits, and hospital admissions with relatively small errors. The explanation module identified extreme heat, poor air quality, respiratory disease activity, a major sporting event, recent arrival growth, and limited inpatient capacity as the principal drivers. Counterfactual simulation indicated that forecast-guided staffing, rapid triage, additional treatment spaces, and early bed review could reduce waiting time and ambulance offload delay.

The superior performance of the ensemble and gradient-boosting models is consistent with a multi-country study involving eleven emergency departments. That investigation found that XGBoost achieved the strongest performance across two forecasting horizons when calendar variables, meteorological information, and engineered temporal features were included. The present study supports the value of nonlinear models but extends previous evidence by examining several clinical outputs, four prediction horizons, six hospitals, prospective validation, surge classification, and operational simulation. The current findings also suggest that ensemble construction may provide further gains when individual algorithms capture different temporal and contextual patterns. ion of external and behavioral variables resembles findings from an emergency-arrival forecasting study that incorporated an internet search index during the H1N1 pandemic. The inclusion of online search activity improved both linear and nonlinear models, while an extreme learning machine achieved better accuracy than traditional alternatives by capturing dynamic relationships between public behavior and emergency attendance. The present research did not rely exclusively on internet searches, but its use of disease surveillance, public events, weather, air quality, and ambulance data reflects the same principle: emerging demand cannot always be inferred from historical attendance alone. izon framework is also compatible with recent research on emergency department waiting-room occupancy.

A 2025 study developed separate six-hour and twenty-four-hour prediction models using eleven machine learning and deep learning approaches. Advanced time-series models outperformed traditional forecasting methods, although prediction error increased during unusually high-volume conditions and varied by time of day. The present findings reproduce this performance pattern because short-term forecasts were more accurate than longer-range estimates, while extreme surge periods remained more difficult to predict than routine demand. tudy differs from many previous investigations by explicitly evaluating external validity and temporal stability. Multisite research conducted before and during the COVID-19 pandemic showed that emergency crowding models can experience spatial and temporal data drift when transferred between hospitals or applied during major changes in demand. The present leave-one-hospital-out validation and prospective silent-testing procedures were designed to confront this problem directly. Model performance remained acceptable across institutions, but variation between tertiary, metropolitan, and regional hospitals confirmed that transportability should never be assumed solely from strong internal accuracy. n on the Meaning of the Findings

The results signal a transition from reactive emergency management toward anticipatory operational planning. Conventional management often responds to congestion only after waiting rooms, treatment spaces, and inpatient boarding areas have become overloaded. A reliable forecasting system provides an earlier decision window in which managers can adjust staffing, activate additional spaces, coordinate ambulance arrivals, and prepare inpatient beds. Recent predictive frameworks similarly emphasize that emergency crowding should be anticipated through hourly and daily forecasts rather than managed only after congestion becomes visible. importance of predictors signals that emergency demand is generated by multiple interacting systems. Patient arrivals reflect not only disease prevalence but also public behavior, transportation patterns, environmental exposure, primary-care accessibility, referral processes, local events, and hospital capacity. The strong contribution of engineered temporal variables and contextual information across hospitals supports the view that emergency attendance is a complex adaptive process.

Research across eleven emergency departments in three countries likewise found that calendar, meteorological, and engineered features improved arrival forecasts, although the most useful variables differed among sites. n accuracy across longer forecast horizons signals an unavoidable boundary between prediction and uncertainty. Near-term forecasts benefit from recent arrivals and current occupancy, while weekly estimates depend on events that may not yet be observable. Longer lead times remain valuable for scheduling, but they should be communicated probabilistically rather than as precise counts. Research on hourly probabilistic forecasting has similarly emphasized prediction distributions and uncertainty ranges because operational decisions require information about plausible demand variation, not only a single expected value. y findings signal that interpretability is central to operational trust. Hospital managers were more likely to regard the warning as actionable because the model identified recognizable factors behind the predicted surge. An unexplained probability would have offered less guidance about whether the forecast reflected respiratory illness, a public event, reduced bed capacity, or abnormal ambulance activity. Explainable models have also been used to assess prolonged emergency waiting times and subgroup fairness, demonstrating that feature-level interpretation can support both operational understanding and quality assessment. ons of the Findings: So What?

The primary theoretical implication is that patient-surge forecasting should be conceptualized as a multi-output capacity-prediction problem rather than a simple attendance-counting exercise. Total arrivals provide only a partial representation of pressure because identical patient volumes may create different workloads depending on acuity, ambulance use, pediatric demand, diagnostic needs, admission rates, and boarding conditions. The current framework demonstrates that forecasting high-acuity cases and likely admissions can provide a more clinically meaningful picture of future pressure. Emergency demand theory should therefore connect patient inflow with treatment complexity and downstream hospital capacity (Albrecht, 2024; Mahapatra, 2025).

The operational implication is that different forecast horizons should support different managerial decisions. Six-hour warnings are suitable for redeploying staff, accelerating triage, adjusting ambulance coordination, and opening flexible treatment areas. Twenty-four-hour forecasts can guide shift assignments, diagnostic preparation, inpatient-bed planning, and elective capacity decisions. Seven-day estimates are more appropriate for roster adjustment, supply preparation, and interdepartmental coordination. A user-centered operational planning study similarly developed fourteen-day forecasts for arrivals, admissions, patient-sitter needs, and emergency holds through collaboration with nursing operations managers, illustrating the importance of matching prediction targets with actual planning responsibilities. implication concerns the distribution rather than the simple expansion of staffing.

Constantly increasing staff numbers would be financially unsustainable and might create unnecessary idle capacity during low-demand periods. Forecast-based scheduling allows hospitals to place additional personnel during predicted peaks while maintaining leaner coverage when demand is expected to decline. An AI-driven pediatric emergency study linked overcrowding forecasts with physician shift optimization and used an MLOps architecture to support continuous model updates. Its findings reinforce the possibility of translating forecasts into staffing arrangements, although multicenter and prospective evaluation remains necessary.

e implication is that performance evaluation must extend beyond mean prediction error. A model with low average error may still be unsafe if it systematically misses extreme surges, underpredicts high-acuity demand, or performs poorly for particular communities. Surge sensitivity, warning time, false-alarm burden, calibration, subgroup accuracy, and operational consequences should be examined together. Fairness research in emergency wait-time prediction has shown that overall model feasibility does not eliminate the need to evaluate disparities across sensitive patient characteristics. on of the Research Findings: Why?

The hybrid ensemble performed strongly because it combined models with different analytical strengths. Gradient boosting effectively captured nonlinear relationships and interactions among temporal, environmental, and operational variables. Recurrent neural networks represented sequential dependence and longer temporal patterns. Tree-based models handled complex feature structures and threshold effects, while statistical components retained stable seasonal information. Combining these outputs reduced dependence on the weaknesses of any single algorithm and improved robustness when demand patterns changed.

Short-term forecasts achieved higher accuracy because recent arrivals and occupancy conditions contain direct information about developing congestion. Patient demand often displays temporal continuity, meaning that a rapidly rising arrival rate during the previous several hours is likely to influence the immediate future. Current ambulance activity and bed occupancy also provide contemporaneous evidence of operational pressure. Daily and weekly forecasts rely more heavily on uncertain external conditions, causing prediction intervals to widen and point estimates to become less precise (Albrecht, 2024; Mahapatra, 2025).

The inclusion of contextual variables improved prediction because historical attendance cannot represent every emerging demand mechanism. Extreme temperature may increase heat-related presentations, poor air quality can contribute to respiratory demand, infectious disease surveillance may anticipate seasonal peaks, and public events can alter trauma or alcohol-related attendance. Internet-behavior research has similarly shown that information-seeking patterns can capture sudden changes not fully represented by static calendar and weather variables. Nonlinear models performed better because the effect of these predictors may depend on their intensity, timing, and interaction with existing hospital pressure. ific variation occurred because emergency departments operate within different clinical and community systems. Tertiary hospitals receive complex referrals and larger numbers of ambulance transfers, metropolitan hospitals are affected by population mobility and public events, while regional hospitals may be more sensitive to seasonal conditions and limited alternative services. Multisite forecasting research has shown that models are vulnerable to spatial drift when relationships learned at one institution do not remain stable elsewhere. The hierarchical and hospital-specific components used in the present study likely preserved local patterns while allowing the pooled model to learn broader regularities. tions and Research Directions: Now What?

Future research should move from silent prospective validation to controlled implementation trials. The present study estimated operational effects through counterfactual simulation, which cannot fully reproduce the behavior of clinicians, patients, ambulance services, or inpatient units after a warning is issued. Stepped-wedge, cluster-randomized, or prospective quasi-experimental studies should compare forecast-guided planning with standard management. Primary outcomes should include waiting time, treatment delay, boarding duration, ambulance offload time, left-without-being-seen rates, staff workload, patient safety, and total operational cost (Preez, 2025; Qin, 2026).

Future model development should place greater emphasis on probabilistic and asymmetric forecasting. Underpredicting a severe surge may produce greater harm than modestly overpredicting demand, yet conventional loss functions often treat both errors equally. Recent research has applied asymmetric error functions that penalize underprediction more heavily and has used conformal intervals to represent forecast uncertainty. Future systems should allow hospitals to select risk thresholds according to local staffing flexibility, surge tolerance, and the clinical consequences of missed demand. tion should examine fairness, transportability, and model drift as continuous responsibilities. Subgroup performance should be assessed across age, sex, socioeconomic catchment, ethnicity where legally and ethically appropriate, arrival mode, clinical category, and disability status. Monitoring systems should identify changes in data definitions, referral policies, population behavior, disease patterns, and hospital capacity (Heidari, 2026; Hodgson, 2025).

Multisite crowding studies and fairness evaluations show that good initial performance does not guarantee stable or equitable prediction after deployment. entation should preserve human authority while formalizing how forecasts trigger action. Emergency physicians, nurses, bed managers, ambulance coordinators, data scientists, and hospital executives should jointly define surge thresholds and corresponding responses. Model explanations, uncertainty intervals, performance dashboards, and override documentation should be integrated into the operational platform. Prospective research comparing machine learning with nurse judgment also suggests that human and algorithmic predictions may provide complementary information for admission and bed planning. The most sustainable future model is therefore a human–AI decision system in which prediction strengthens professional preparedness without transferring accountability to an automated tool.

CONCLUSION

The most important finding of this study is that machine learning algorithms can predict emergency department patient surges more accurately when multiple temporal, operational, epidemiological, environmental, and social variables are analyzed simultaneously. The hybrid ensemble model achieved the strongest performance across prediction horizons, particularly for six-hour and twenty-four-hour forecasts, and consistently outperformed conventional statistical approaches. Recent arrival growth, emergency department occupancy, ambulance activity, infectious disease indicators, bed availability, weather conditions, and public events emerged as influential predictors. These findings show that emergency surges are not produced by historical attendance patterns alone but by interacting conditions that require dynamic and context-sensitive forecasting.

The study contributes both conceptually and methodologically to emergency demand prediction. Its conceptual contribution lies in framing patient-surge forecasting as a multi-output operational capacity problem that includes total arrivals, high-acuity cases, ambulance presentations, pediatric demand, and likely admissions. Its methodological contribution is reflected in the integration of multi-site data, multiple forecast horizons, comparative algorithm evaluation, temporal validation, leave-one-hospital-out testing, explainability analysis, fairness assessment, model-drift monitoring, and counterfactual operational simulation. This comprehensive framework connects predictive accuracy with practical hospital decisions concerning staffing, treatment-space preparation, triage capacity, ambulance coordination, and inpatient-bed management.

The principal limitations of the study include its reliance on retrospective records, the relatively short prospective validation period, differences in data quality among hospitals, and the use of simulated rather than directly observed operational interventions. Model performance may also decline when disease patterns, referral policies, community behavior, hospital capacity, or data structures change substantially. Future research should conduct longer prospective trials, stepped-wedge or cluster-randomized evaluations, and external validation across hospitals with different technological and organizational capacities. Further studies should investigate probabilistic forecasting, asymmetric error costs, algorithmic fairness, model drift, public-health emergencies, and the combined performance of machine predictions and professional judgment in real-time emergency management.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author(s) used Google Gemini to assist in improving grammar, language quality, and overall readability of the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.
Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Agur, Z. (2026). Dysregulated metabolic homeostasis as a unifying death mechanism underlying the diverse clinical manifestations of COVID-19: Insights from a retrospective analysis of sequential blood variables. *Frontiers in Medicine*, *13*(Query date: 2026-07-31 21:18:58). https://doi.org/10.3389/fmed.2026.1773937
https://doi.org/10.1186/s43066-023-00306-3

Albrecht, S. (2024). Forecasting severe respiratory disease hospitalizations using machine learning algorithms. *BMC Medical Informatics and Decision Making*, *24*(1). https://doi.org/10.1186/s12911-024-02702-0

Alnaggar, A. (2025). Distributionally robust hospital capacity expansion planning under stochastic and correlated patient demand. *Computers and Operations Research*, *174*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1016/j.cor.2024.106887

Al-qaness, M. A. A. (2024). Chest X-ray Images for Lung Disease Detection Using Deep Learning Techniques: A Comprehensive Survey. *Archives of Computational Methods in Engineering*, *31*(6), 3267–3301. https://doi.org/10.1007/s11831-024-10081-y

Chadha, U. (2026). Augmenting healthcare systems for pandemic preparedness: A Lean Six Sigma perspective. *International Journal for Quality in Health Care*, *38*(2). https://doi.org/10.1093/intqhc/mzag071

Choo, L. X. (2025). Clinical practice in an age of medical misinformation and conspiracy theories. *Internal Medicine Journal*, *55*(1), 158–164. https://doi.org/10.1111/imj.16588

Datta, D. (2024). Feature Identification Using Interpretability Machine Learning Predicting Risk Factors for Disease Severity of In-Patients with COVID-19 in South Florida. *Diagnostics*, *14*(17). https://doi.org/10.3390/diagnostics14171866

Fan, W. (2025). Class-specific weighted broad learning system-based domain adaptation for patient-specific ECG classification. *Expert Systems with Applications*, *273*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1016/j.eswa.2025.126824

Ghosh, S. (2025). Biomarker-driven drug repurposing for NAFLD-associated hepatocellular carcinoma using machine learning integrated ensemble feature selection. *Frontiers in Bioinformatics*, *5*(Query date: 2026-07-31 21:18:58). https://doi.org/10.3389/fbinf.2025.1522401

Gorrepati, L. P. (2026). DR-UCSS: A Distributionally Robust, Uncertainty-calibrated Counterfactual Safety-shielded Reinforcement Learning Framework for Hospital Operations. *International Journal of Intelligent Engineering and Systems*, *19*(4), 286–302. https://doi.org/10.22266/ijies2026.0430.16

Goyal, S. (2025). Development of Mobile Apps to Optimize Performance Techniques Using Machine Learning Techniques. *2025 5th International Conference on Advancement in Electronics and Communication Engineering Aece 2025*, (Query date: 2026-07-31 21:18:58), 1145–1150. https://doi.org/10.1109/AECE67531.2025.11386548

Grdić, M. (2025). Bridging Technology and Healthcare: The Impact of AI in Surgical Instrument Classification. *Tehnicki Glasnik*, *19*(3), 422–426. https://doi.org/10.31803/tg-20240516154449

Heidari, I. (2026). Forecasting and optimizing real-time ventilator allocation in healthcare systems using artificial intelligence and geographic clustering. *Engineering Applications of Artificial Intelligence*, *175*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1016/j.engappai.2026.114427

Hiwase, A. D. (2025). Chemical Pneumonitis Detection using Machine Learning. *16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025*, *2*(Query date: 2026-07-31 21:18:58), 3183–3187.

Hodgson, M. (2025). Finite element model and experimental validation of a solitary wave-based tonometer. *Journal of the Mechanical Behavior of Biomedical Materials*, *168*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1016/j.jmbbm.2025.107000

Hu, P. (2026). Effects of ABO-Incompatible Blood Transfusion on Immune Response and Rejection After Organ Transplantation. *Clinical Transplantation*, *40*(2). https://doi.org/10.1111/ctr.70469

Huang, R. (2024). Edge Computing Model based on Federated Learning for COVID-19 Clinical Outcome Prediction in the 5G Era. *Ksii Transactions on Internet and Information Systems*, *18*(4), 826–842. https://doi.org/10.3837/tiis.2024.04.001

Khademi, S. (2024). FH-TabNet: Multi-Class Familial Hypercholesterolemia Detection via a Multi-Stage Tabular Deep Learning Network. *European Signal Processing Conference*, (Query date: 2026-07-31 21:18:58), 1416–1420. https://doi.org/10.23919/eusipco63174.2024.10715254

Khene, Z. E. (2024). Clinical Application of Digital and Computational Pathology in Renal Cell Carcinoma: A Systematic Review. *European Urology Oncology*, *7*(3), 401–411. https://doi.org/10.1016/j.euo.2023.10.018

Kocar, T. D. (2025). Deep Learning Predicts Postoperative Mobility, Activities of Daily Living, and Discharge Destination in Older Adults from Sensor Data. *Sensors*, *25*(16). https://doi.org/10.3390/s25165021

Lampreia, F. (2024). Digital health technologies and artificial intelligence in cardiovascular clinical trials: A landscape of the European space. *Digital Health*, *10*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1177/20552076241277703

Li, T. n. (2025). Clinical Characteristics of Patients With Respiratory Infections After Nonpharmacological Interventions for COVID-19 in China Have Ended: Using Machine Learning Approaches to Support Pathogen Prediction at Admission. *Immunity Inflammation and Disease*, *13*(8). https://doi.org/10.1002/iid3.70237

Madsen, G. R. (2024). Classifying Inflammation on Intestinal Ultrasound Images and Cineloops—A Learning Curve Study. *Journal of Crohn S and Colitis*, *18*(12), 2076–2084. https://doi.org/10.1093/ecco-jcc/jjae112

Mahapatra, K. (2025). Fractal Dimension and Entropy Analysis of Medical Images for KNN-Based Disease Classification. *Baghdad Science Journal*, *22*(4), 1354–1365. https://doi.org/10.21123/bsj.2024.10835

Marfoglia, A. (2025). CONNECTED: A Knowledge Graph-Driven Platform for Clinical Data Harmonization and Personalized Digital Twin-Based Healthcare. *IEEE International Conference on Pervasive Computing and Communications Workshops Percom Workshops*, (2025), 116–121. https://doi.org/10.1109/PerComWorkshops65533.2025.00051

Nasatzky, M. (2024). Clinical parameters that predict a premature LH rise in patients undergoing ovarian stimulation for IVF. *Gynecological Endocrinology*, *40*(1). https://doi.org/10.1080/09513590.2024.2365913

Oskvarek, J. J. (2026). Emergency department crowding in the modern era: A systematic review (2018–2025). *Clinical and Experimental Emergency Medicine*, *13*(2), 119–139. https://doi.org/10.15441/ceem.25.172

Pahwa, B. (2024). Contributions of Machine Learning in the Management of Stroke: A Bibliometric Analysis of the 50 Most Cited Articles. *World Neurosurgery*, *184*(Query date: 2026-07-31 21:18:58), 152–160. https://doi.org/10.1016/j.wneu.2024.01.059

Paz, H. L. (2024). Buying into healthcare: Why the industry must learn to appeal to consumers to survive and thrive. *Management in Healthcare*, *8*(2), 107–116. https://doi.org/10.69554/xnnv8502

Preetha, M. (2024). Deep Learning-Driven Real-Time Multimodal Healthcare Data Synthesis. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(5), 360–369.

Preez, A. du. (2025). Fraud detection in healthcare claims using machine learning: A systematic review. *Artificial Intelligence in Medicine*, *160*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1016/j.artmed.2024.103061

Qin, T. H. Y. (2026). From Biomechanics to Metallobiology: A Bibliometric Mapping of Ferroptosis and Trace Elements in Osteoarthritis (2014–2024). *Biological Trace Element Research*, *204*(8), 6106–6120. https://doi.org/10.1007/s12011-026-05095-5

Ren, Y. (2024). Automatic uncovering of patient primary concerns in portal messages using a fusion framework of pretrained language models. *Journal of the American Medical Informatics Association*, *31*(8), 1714–1724. https://doi.org/10.1093/jamia/ocae144

Roccuzzo, G. (2024). Emerging prognostic biomarkers in advanced cutaneous melanoma: A literature update. *Expert Review of Molecular Diagnostics*, *24*(1), 49–66. https://doi.org/10.1080/14737159.2024.2314574

Sharma, A. K. (2025). Bibliometric exploration of artificial intelligence applications in healthcare: Trends and future directions. *Journal of Public Health and Development*, *23*(2), 281–303. https://doi.org/10.55131/jphd/2025/230220

Singhal, A. K. (2025). Beyond the traditional models: A network reconstruction based model for predicting and analysing individual health status. *Computing*, *107*(1). https://doi.org/10.1007/s00607-024-01355-1

Stefanou, M. I. (2024). CURRENT UPDATE ON THE NEUROLOGICAL MANIFESTATIONS OF LONG COVID: MORE QUESTIONS THAN ANSWERS. *Excli Journal*, *23*(Query date: 2026-07-31 21:18:58), 1463–1486. https://doi.org/10.17179/excli2024-7885

Warrier, A. (2025). Autonomous Agentic AI for Clinical Workflow Orchestration: Self-Managing Healthcare Operations. *Proceedings of 6th International Conference on Iot Based Control Networks and Intelligent Systems Icicnis 2025*, (Query date: 2026-07-31 21:18:58), 1450–1457. https://doi.org/10.1109/ICICNIS66685.2025.11315712

Wei, L. (2025). Development and validation of a machine learning-based model for 90-day prognosis outcome in spontaneous intracerebral hemorrhage patients based on non-contrast computed tomography: A multicenter retrospective observational study. *Eclinicalmedicine*, *88*(Query date: 2026-07-31 21:18:58). https://doi.org/10.1016/j.eclinm.2025.103507