



Olawade, David ORCID logoORCID: <https://orcid.org/0000-0003-0188-9836>, Da'Costa, Adebayo, Origbo, Joseph E, Osonuga, Ayokunle, Egbon, Eghosasere, Teke, Jennifer and Boussios, Stergios (2025) Smart emergency care: a narrative review of predictive machine learning models. *Annals of translational medicine*, 13 (5). p. 60.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/13327/>

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version:
<https://dx.doi.org/10.21037/atm-25-83>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repositories Policy Statement](#)

RaY

Research at the University of York St John

For more information please contact RaY at
ray@yorks.ac.uk



Smart emergency care: a narrative review of predictive machine learning models

David B. Olawade^{1,2,3,4}, Adebayo Da'Costa⁵, Joseph E. Origbo⁶, Ayokunle Osonuga⁷, Eghosasere Egbon⁸, Jennifer Teke^{9,10}, Stergios Boussios^{2,9,10,11,12,13,14}

¹Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, UK; ²Department of Research and Innovation, Medway NHS Foundation Trust, Gillingham, UK; ³Department of Public Health, York St John University, London, UK; ⁴School of Health and Care Management, Arden University, Coventry, UK; ⁵Department of Emergency Medicine, Medway NHS Foundation Trust, Gillingham UK; ⁶Department of Computer Science, Nottingham Trent University, Nottingham, UK; ⁷Coltishall Medical Practice, NHS GP Surgery, Norfolk, UK; ⁸Department of Tissue Engineering and Regenerative Medicine, Faculty of Life Science Engineering, FH Technikum, Vienna, Austria; ⁹Faculty of Medicine, Health and Social Care, Canterbury Christ Church University, Canterbury, UK; ¹⁰Faculty of Life Sciences & Medicine, School of Cancer & Pharmaceutical Sciences, King's College London, London, UK; ¹¹Faculty of Medicine, School of Health Sciences, University of Ioannina, Ioannina, Greece; ¹²Department of Medical Oncology, University Hospital of Ioannina, Ioannina, Greece; ¹³Department of Medical Oncology, Medway NHS Foundation Trust, Gillingham, Kent, UK; ¹⁴AELIA Organization, Thessaloniki, Greece

Contributions: (I) Conception and design: DB Olawade, A Da'Costa; (II) Administrative support: J Teke, S Boussios; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: DB Olawade, JE Origbo, A Osonuga; (V) Data analysis and interpretation: E Egbon, JE Origbo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: David B. Olawade, MPH. Department of Research and Innovation, Medway NHS Foundation Trust, Windmill Road, Gillingham, Kent ME7 5NY, UK; Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, UK; Department of Public Health, York St John University, London, UK; School of Health and Care Management, Arden University, Arden House, Middlesbrough Park, Coventry, CV3 4FJ, UK. Email: d.olawade@uel.ac.uk.

Background and Objective: The Emergency Department (ED) is a critical, high-stakes environment where timely and accurate assessments of patient outcomes are essential for ensuring optimal care and effective resource management. This narrative review aimed to synthesise current evidence on machine learning (ML)-based predictive models used in the ED to forecast patient outcomes such as mortality, intensive care unit (ICU) admission, and discharge probability, whilst identifying key limitations and future research directions.

Methods: This narrative review synthesises recent advancements in ML-based predictive models for ED outcomes published between January 2015 and December 2024. It explores the integration of real-time and historical clinical data, focusing on key ML techniques such as regression models, decision trees, neural networks, and ensemble methods. The review also evaluates data sources, model evaluation metrics, and addresses challenges including data quality, interpretability, and ethical considerations. A comprehensive search of four major databases yielded 156 initial results, with 45 studies ultimately included after systematic screening.

Key Content and Findings: ML models demonstrate significant promise in processing complex, non-linear data for ED outcome prediction with area under the receiver operating characteristic curve (AUC-ROC) values typically ranging from 0.75–0.95 across different outcomes. Techniques like ensemble methods and neural networks offer strong performance, while personalized prediction models and explainable artificial intelligence (XAI) enhance precision and interpretability. However, current approaches face substantial limitations including data heterogeneity, poor model generalisability across institutions, and lack of real-world implementation studies. Emerging integration of telemedicine further broadens the applicability of predictive modeling in the ED.

Conclusions: ML is reshaping predictive modeling in the ED, offering timely, data-driven support for clinical decision-making. Despite challenges, advancements in personalized and explainable models hold the

potential to increase trust and usability in clinical workflows. Critical gaps remain in addressing data quality issues, standardising evaluation metrics, and conducting multi-centre validation studies.

Keywords: Machine learning (ML); predictive modeling; Emergency Department (ED); patient outcomes; clinical decision-making

Submitted May 26, 2025. Accepted for publication Jul 22, 2025. Published online Oct 28, 2025.

doi: 10.21037/atm-25-83

View this article at: <https://dx.doi.org/10.21037/atm-25-83>

Introduction

The Emergency Department (ED) is a critical component of healthcare systems, serving as a frontline response for patients with a wide range of conditions, from minor injuries to severe, life-threatening emergencies (1). In this high-stakes, fast-paced environment, clinicians face immense pressure to quickly and accurately assess patient conditions and make informed treatment decisions. Rapid, precise assessment of patient outcomes is essential, as it guides clinical decisions, optimizes resource allocation, and directly impacts patient survival and recovery (2). However, due to the overwhelming influx of patients and the demand for swift decision-making, ED clinicians often rely on a combination of clinical experience, intuition, and traditional scoring systems, which, while helpful, have limitations in terms of speed, scalability, and predictive accuracy (3,4). As healthcare systems continue to manage increasing patient volumes and resource constraints, there is a growing need for innovative tools that enhance clinical decision-making in real-time.

In recent years, machine learning (ML) has emerged as a transformative technology in healthcare, especially in areas like the ED, where timely, data-driven insights can be lifesaving (5). ML, a branch of artificial intelligence (AI), uses algorithms and statistical models that learn patterns from vast amounts of data, enabling them to make predictions or decisions without requiring explicit programming for each task. ML-based predictive models are especially valuable in the ED, as they can process diverse inputs such as vital signs, lab results, demographics, and even unstructured clinical notes allowing for a more comprehensive and nuanced analysis than is typically achievable with traditional methods (4). For instance, while traditional scoring systems like the Acute Physiology and Chronic Health Evaluation (APACHE) or the Sequential Organ Failure Assessment (SOFA) have been instrumental in assessing risk in critically ill patients, these systems are

limited by the number of variables they can incorporate and are often applied broadly, potentially overlooking individual patient nuances. ML models, in contrast, can integrate real-time data with historical clinical information, enabling them to make individualized predictions about key outcomes such as mortality risk, likelihood of intensive care unit (ICU) admission, and discharge probability (6,7).

The implications of predictive modeling in the ED extend beyond immediate patient care, influencing broader operational and resource management decisions. By anticipating patient outcomes with greater accuracy, predictive models allow for earlier identification of high-risk patients, support more efficient triage processes, and help optimize bed utilization. These insights ultimately enable healthcare providers to allocate resources more effectively, improving patient throughput and potentially reducing wait times. Although ML-based predictive models hold significant promise, there are also challenges to consider, including the integration of models into existing ED workflows, issues related to data quality and consistency, and the need for model interpretability to ensure clinicians can trust and act on the predictions provided. This review aims to synthesise current developments in predictive modeling for patient outcomes in the ED, examining the types of data and ML techniques used, the performance of various models, and the challenges and opportunities that lie ahead. By exploring the practical implications of predictive models and identifying potential directions for future research, this paper seeks to highlight the transformative potential of ML in enhancing patient outcomes and operational efficiency in emergency care.

This review focuses specifically on ML applications within the ED setting, encompassing both immediate triage decisions and subsequent care pathways including ICU admission and mortality prediction within the first 24–48 hours of ED presentation. This narrative review examines studies published between January 2015 and

December 2024, focusing on original research articles that developed or evaluated ML-based predictive models for patient outcomes in ED settings. The scope encompasses models predicting mortality, ICU admission, and discharge outcomes, whilst critically examining current limitations and identifying key areas requiring further research attention. We present this article in accordance with the Narrative Review reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-25-83/rc>).

Methods

This narrative review was conducted to explore and synthesise existing literature on the application of ML models for predicting patient outcomes in the ED. The primary outcomes of interest included mortality, ICU admission, and discharge probability. The review aimed to provide an overview of the data sources, types of ML models used, evaluation metrics, and future directions in predictive modelling within ED settings whilst critically examining current limitations and methodological challenges.

Search strategy and study selection

A comprehensive search was conducted across major databases, including PubMed, Scopus, Web of Science, and IEEE Xplore, for peer-reviewed articles published between 14th January 2015 and 28th December 2024. Search terms included combinations of keywords such as “machine learning”, “predictive models”, “emergency department”, “mortality prediction”, “ICU admission”, “discharge”, “triage”, and “clinical decision support”. Additional terms included “artificial intelligence”, “deep learning”, “neural networks”, and “clinical decision support systems”. A comprehensive summary of the search strategy, including databases searched, search terms, and selection criteria, is presented in *Table 1*. The initial search strategy yielded 156 articles across all databases: PubMed (68 articles), Scopus (41 articles), Web of Science (32 articles), and IEEE Xplore (15 articles). After removing duplicates (n=23), 133 articles underwent title and abstract screening. Of these, 76 articles were excluded for not meeting inclusion criteria, leaving 57 articles for full-text review. Following full-text assessment, 12 additional articles were excluded due to insufficient methodological detail or lack of ED-specific focus, resulting in 45 studies included in the final analysis (*Figure 1*).

Inclusion criteria were: (I) studies focused on the development or evaluation of ML-based predictive models

in the ED; (II) models aimed at predicting at least one of the target outcomes (mortality, ICU admission, or discharge); (III) use of real-world clinical data (real-time or historical); and (IV) articles published in English. Exclusion criteria included review articles, editorials, conference abstracts without full text, and studies not specifically focused on ED settings.

Data extraction and analysis strategy

The selected studies were reviewed for information on study design, data type, ML algorithms applied, outcome measures, and model performance metrics. Data extraction focused on identifying: (I) study characteristics (sample size, setting, study period); (II) ML methodologies employed; (III) performance metrics [area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, F1-score]; (IV) data quality considerations; and (V) reported limitations. The narrative synthesis focused on identifying trends, strengths, limitations, and gaps in current evidence to inform future research and clinical applications. Studies were thematically grouped by primary outcome (mortality, ICU admission, discharge prediction) and analytical approach, with particular attention paid to methodological limitations and real-world implementation challenges.

Data sources for predictive models

The performance of ML models in predicting patient outcomes in the ED relies heavily on data quality and the diversity of data inputs. Accurate, reliable models must draw from comprehensive datasets that capture the patient's current condition and relevant medical history (8-10). ML models for ED outcome prediction typically utilize two main types of data: real-time clinical data and historical clinical data. Each data source contributes distinct value to the predictive process, enabling the generation of timely and individualized predictions (11).

Table 2 summarizes key clinical scoring systems: Modified Early Warning Score (MEWS), SOFA, APACHE, quick SOFA (qSOFA), and National Early Warning Score (NEWS), used in the ED for mortality prediction and risk assessment (17). The MEWS and NEWS rely on basic vital signs, enabling rapid assessments ideal for ED triage, while SOFA and APACHE provide in-depth risk evaluation using lab results and organ function metrics, though they are more time-intensive (18-20). Simplified scores like qSOFA offer quick sepsis risk evaluations but may lack sensitivity.

Table 1 The search strategy summary

Items	Specification
Date of search	February 15 th , 2025
Databases and other sources searched	PubMed, Scopus, Web of Science, IEEE Xplore
Search terms used	Primary search terms: “machine learning”, “predictive models”, “emergency department”, “mortality prediction”, “ICU admission”, “discharge”, “triage”, “clinical decision support” Additional terms: “artificial intelligence”, “deep learning”, “neural networks”, “clinical decision support systems”
Timeframe	January 2015 to December 2024 (peer-reviewed articles)
Inclusion and exclusion criteria	Inclusion criteria: studies focused on development or evaluation of ML-based predictive models in the ED; models predicting at least one target outcome (mortality, ICU admission, or discharge); use of real-world clinical data (real-time or historical); articles published in English Exclusion criteria: review articles; editorials; conference abstracts without full text; studies not specifically focused on ED settings
Selection process	Initial search yielded 156 articles across all databases Duplicates removed (n=23), leaving 133 articles Title and abstract screening conducted 76 articles excluded for not meeting inclusion criteria 57 articles underwent full-text review 12 additional articles excluded due to insufficient methodological detail or lack of ED-specific focus Final inclusion: 45 studies Selection process details not specified regarding independence or consensus methods
Any additional considerations, if applicable	Reference lists of relevant studies were manually screened to identify additional sources Search strategy focused on peer-reviewed articles only No mention of grey literature or unpublished studies Language restriction to English may have introduced geographical bias

ED, emergency department; ICU, intensive care unit; ML, machine learning.

These scoring systems form structured data inputs for ML models, enhancing risk stratification and supporting informed ED decision-making (21–23). However, traditional scoring systems demonstrate limited predictive accuracy with AUC-ROC values typically ranging from 0.65–0.80, highlighting the potential value of ML approaches.

Real-time clinical data

Real-time clinical data is critical for ML models aiming to predict patient outcomes as early as possible upon ED arrival. This data encompasses immediate, continuously monitored metrics and diagnostic information that provides a snapshot of the patient's current health status (24,25).

Real-time data commonly includes vital signs such as heart rate, respiratory rate, blood pressure, and oxygen saturation levels, which are foundational indicators of a patient's physiological stability. These vital signs can be predictive of deterioration in the ED, with abnormal values often indicating heightened risk for adverse outcomes, including mortality and ICU admission (26,27). Additionally, real-time clinical data can include results from initial lab tests (e.g., blood counts, electrolyte levels, and biomarkers), imaging reports, and triage assessments conducted by ED staff. This data is generally available within minutes of patient arrival and can be continuously updated, which allows models to adjust their predictions as new data points are acquired.

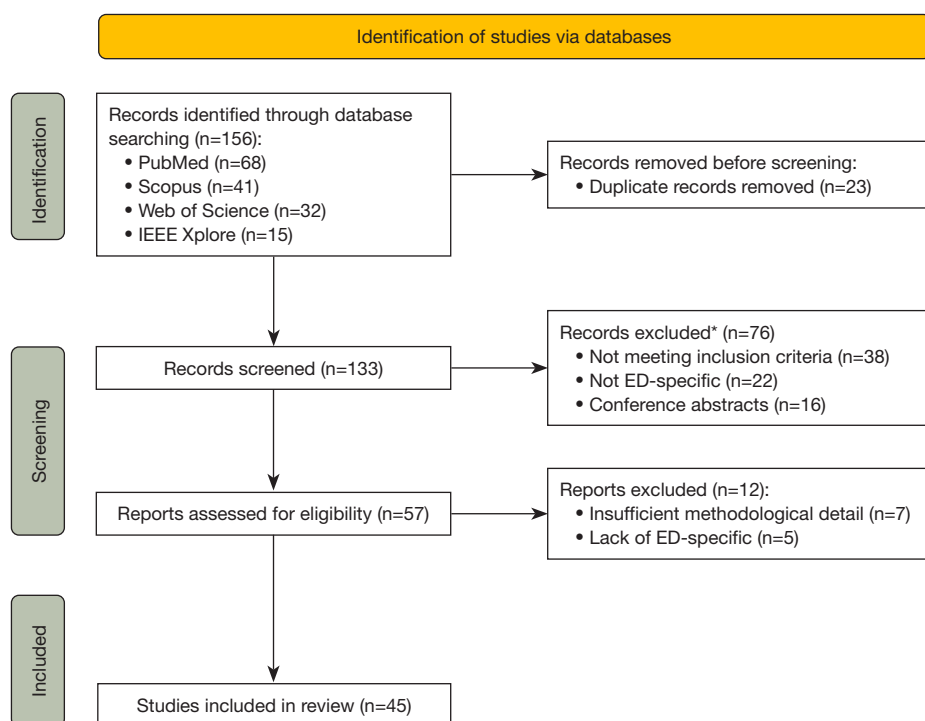


Figure 1 PRISMA flow diagram for selected articles. *, nothing else to define apart from ED. ED, emergency department.

Table 2 Commonly used clinical scoring systems in ED predictive models

Scoring System	Description	Primary Use	Advantages	Limitations
MEWS (12)	Assesses risk based on vital signs (e.g., heart rate, respiratory rate, blood pressure)	Early mortality prediction	Simple, quick to apply	Limited to physiological data
SOFA (13)	Predicts likelihood of organ failure and mortality based on organ function parameters	ICU admission and mortality	Well-established in critical care	Requires lab values, less effective for ED triage
APACHE (14)	Estimates risk of mortality using a range of physiological and lab data	Mortality prediction in ICU	Comprehensive, validated in ICU	Complex, time-consuming to calculate
qSOFA (15)	Simplified SOFA for rapid assessment, using respiration, mental state, blood pressure	Sepsis risk identification	Easy to use, suitable for ED triage	Limited in sensitivity for sepsis
NEWS (16)	Similar to MEWS, includes oxygen saturation and patient alertness as indicators	General patient deterioration	Broadly applicable in ED settings	May not capture all critical cases

APACHE, Acute Physiology and Chronic Health Evaluation; ED, emergency department; ICU, intensive care unit; MEWS, Modified Early Warning Score; NEWS, National Early Warning Score; qSOFA, Quick SOFA; SOFA, Sequential Organ Failure Assessment.

Models utilising real-time data alone demonstrate moderate predictive performance, with AUC-ROC values typically ranging from 0.72–0.85 for mortality prediction and 0.70–0.83 for ICU admission prediction. However, real-time data faces significant limitations including measurement errors, temporal variability, and incomplete capture during peak ED volumes (28).

Predictive models trained on real-time data offer significant advantages, particularly in terms of timeliness. By processing and analyzing real-time data, these models can make early, rapid predictions that aid in triage decisions and the prioritization of critical interventions (29). For instance, models using real-time data have been employed to identify patients at high risk of sepsis or other severe complications shortly after ED arrival, thereby facilitating expedited care and reducing the time to diagnosis and treatment. In environments where minutes matter, the ability to leverage real-time data for quick, data-driven insights is invaluable. Furthermore, real-time data enables the prediction of short-term outcomes, such as the need for ICU admission or discharge probability, providing critical support in resource allocation and bed management (30).

Historical clinical data

While real-time data provides an immediate view of a patient's current health status, historical clinical data adds a broader context that can enhance the accuracy of predictive models. Historical data typically encompasses a patient's previous medical records, including past diagnoses, comorbidities, medication history, previous admissions, procedures, and long-term health outcomes. This data allows ML models to incorporate insights into chronic health conditions, previous responses to treatments, and other trends that may influence the patient's current episode in the ED. The presence of chronic illnesses like diabetes, hypertension, or chronic obstructive pulmonary disease (COPD) has been shown to impact outcomes such as ICU admission and in-hospital mortality, highlighting the importance of these data points in building predictive models (31). Models incorporating both real-time and historical data demonstrate superior performance, with AUC-ROC values ranging from 0.82–0.95 across different outcomes. However, historical data presents unique challenges including data accessibility, privacy concerns, and significant variation in data completeness between patients with different healthcare utilisation patterns (32).

Historical data contributes to a more personalized

prediction framework by identifying pre-existing health factors that could complicate the patient's condition or influence the trajectory of care. For instance, a patient with a history of heart disease may have an elevated risk of cardiovascular complications, which can be factored into the model's risk assessment for ICU admission or mortality. Furthermore, historical data can assist in distinguishing between patients who may present similarly upon arrival but have different underlying risk profiles. Finding from a study showed that incorporating a patient's longitudinal health data improves model performance for predicting various outcomes, such as ED revisits or long-term prognosis, by providing a richer understanding of individual risk factors and health trajectories (33).

Current limitations in data sources

Despite the potential benefits of comprehensive data integration, several critical limitations were identified across the reviewed studies. Data heterogeneity represents a significant challenge, with substantial variations in electronic health record systems, data collection protocols, and variable definitions between institutions that limit model generalisability. These differences create barriers to developing universally applicable predictive models that can perform consistently across different healthcare settings (34).

Missing data constitutes another major limitation, with particularly high rates of missing values observed for laboratory results (15–70%) and historical comorbidity data (20–60%). This substantial data incompleteness significantly impacts model performance and reliability, as algorithms struggle to make accurate predictions when key clinical variables are absent. The variability in missing data patterns between institutions further compounds this challenge (35).

Temporal inconsistencies in data collection timing and frequency between different ED settings create additional challenges for model standardisation and validation. These variations affect the comparability of datasets and limit the ability to develop robust models that maintain performance across different temporal contexts. Furthermore, limited standardisation of data quality assurance protocols results in inconsistent data reliability across institutions, undermining confidence in model outputs and hampering efforts to establish best practices for data governance in ML-based predictive modelling (36).

The use of both real-time and historical data enables the development of more robust and reliable predictive models

for ED outcomes. Real-time data captures the immediate clinical status, allowing for quick, actionable predictions that are essential in the high-stakes ED setting, while historical data provides the necessary context to personalize and refine those predictions. When combined, these data sources offer a comprehensive foundation for ML models, enhancing their ability to predict diverse outcomes such as mortality risk, ICU admission, and discharge probability thus supporting timely and precise clinical decision-making in the ED. Future models may also leverage additional data sources, such as social determinants of health or patient-reported outcomes, to further enhance prediction accuracy and address broader factors that influence health outcomes (37).

ML techniques for predictive modeling in the ED

ML has introduced a range of techniques suitable for predicting patient outcomes in the ED. These techniques vary in complexity, interpretability, and predictive power, allowing researchers and clinicians to choose models that best match the characteristics of their data and the needs of their specific ED setting (38). However, each approach faces distinct limitations that impact clinical implementation and real-world performance.

Regression models

Regression models, particularly logistic and Cox regression, are widely used in ED predictive modeling due to their simplicity and ease of interpretation. Logistic regression is frequently employed for binary outcomes, such as mortality prediction or ICU admission likelihood, allowing for straightforward estimation of probabilities based on predictor variables (39). For example, logistic regression models can predict the probability of patient mortality by incorporating established risk factors such as age, vital signs, comorbidities, and presenting symptoms. One of the main strengths of regression models is their interpretability; coefficients for each variable offer insight into how specific risk factors contribute to the outcome. Additionally, Cox regression models are suitable for time-to-event data, allowing for the analysis of factors influencing the time until events like ICU admission or patient discharge.

Across the reviewed studies, logistic regression models demonstrated AUC-ROC values ranging from 0.68–0.82 for mortality prediction and 0.65–0.79 for ICU admission prediction. Studies have shown that regression models can

perform well in predicting ED outcomes, especially when the relationships between variables and outcomes are linear and relatively straightforward (33,34). However, these models may struggle with complex, non-linear relationships, which are often present in high-dimensional healthcare data, limiting their performance in more nuanced predictive tasks. However, these models face significant limitations including poor performance with non-linear relationships, limited ability to capture complex variable interactions, and reduced accuracy when dealing with high-dimensional data typical in modern ED settings. Furthermore, regression models often struggle with multicollinearity issues when multiple correlated predictors are included.

Decision trees and random forests

Decision trees and random forests are popular ML techniques for ED predictive modeling due to their ability to capture complex, non-linear relationships within the data. A decision tree model breaks down data into branches based on feature values, ultimately leading to a predicted outcome based on the combination of variables. Random forests, an ensemble method that combines multiple decision trees, improve upon single decision tree models by reducing overfitting and enhancing generalizability (39,40).

Random forest models consistently demonstrated superior performance compared to single decision trees, with AUC-ROC values ranging from 0.78–0.91 across different ED outcomes (41). In the ED setting, random forest models have demonstrated strong performance in predicting patient outcomes, as they can account for interactions between variables that may not be captured in simpler models. For instance, random forests can combine factors like patient age, initial vital signs, lab results, and comorbid conditions to predict outcomes such as ICU admission or hospital discharge (42). These models are particularly valuable when there are numerous predictor variables with potentially complex interactions, as in emergency medicine. However, decision trees and random forests can be less interpretable than simpler models, especially as the number of trees and branches increases. This lack of transparency may limit clinicians' trust in the predictions, underscoring the need for approaches that balance predictive accuracy with interpretability.

However, key limitations include reduced interpretability as model complexity increases, potential overfitting with small datasets, and computational intensity that may limit real-time implementation. Additionally, decision trees and

random forests can be sensitive to imbalanced datasets, a common challenge in ED settings where adverse outcomes are relatively rare.

Neural networks and deep learning

Neural networks, especially deep learning models, have shown remarkable success in capturing complex, high-dimensional patterns in large datasets, making them well-suited for certain predictive tasks in the ED. These models consist of interconnected layers of “neurons” that process inputs through a series of transformations to produce an output (43,44). Deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are particularly useful in the ED context. CNNs, for example, have been effectively applied to analyze medical imaging data, such as X-rays or computed tomography (CT) scans, to detect conditions like fractures, hemorrhages, or pneumonia, which can directly impact patient outcomes in the ED. RNNs, which are specialized for sequential data, have demonstrated utility in analyzing time-series data, such as a patient’s vital signs over time, allowing for real-time monitoring of patient deterioration or stability.

Deep learning models achieved the highest predictive performance, with AUC-ROC values ranging from 0.85–0.96 for mortality prediction and 0.82–0.94 for ICU admission prediction (45). Despite their powerful predictive capabilities, neural networks are often considered “black-box” models due to their complex internal structure, making them less interpretable. This lack of transparency can be problematic in the clinical setting, as clinicians may hesitate to rely on predictions they do not fully understand. Efforts to improve model interpretability, such as the development of explainable AI (XAI) techniques, are ongoing and may increase the acceptance of neural networks in emergency medicine (46,47). Critical limitations include the requirement for large training datasets, substantial computational resources, extended training times, and most importantly, poor explainability which limits clinical acceptance. Additionally, deep learning models are prone to overfitting, particularly when applied to heterogeneous ED populations, and their performance can degrade significantly when applied to data from different institutions (48).

Ensemble models

Ensemble models are a class of ML techniques that

combine predictions from multiple individual models to create a more robust and accurate final prediction. By aggregating predictions from different models, such as random forests and gradient boosting algorithms, ensemble methods can mitigate the weaknesses of individual models and improve overall performance (49,50). Ensemble models are particularly beneficial in the ED setting, where multiple outcomes, such as mortality, ICU admission, and discharge probability, need to be predicted simultaneously. For instance, gradient boosting, which sequentially improves predictions by focusing on misclassified cases in each iteration, has demonstrated strong predictive performance in ED settings for various outcomes. Similarly, random forest ensembles, which average predictions across a forest of decision trees, have shown success in identifying high-risk patients and optimizing triage decisions.

Ensemble models demonstrated robust performance across studies, with AUC-ROC values consistently ranging from 0.82–0.94 for various ED outcomes. Gradient boosting algorithms, in particular, showed strong performance with values between 0.84–0.93 (51). One of the main advantages of ensemble models is their ability to improve prediction accuracy without significantly increasing model complexity. *Table 3* summarises the performance characteristics of different ML techniques across the reviewed studies.

Figure 2 summarises of the progression and characteristics of different ML techniques used in ED predictive modelling. By leveraging the strengths of multiple algorithms, ensemble models offer a balanced approach to prediction in the ED, providing high accuracy while managing interpretability challenges. However, ensemble methods can still present a “black-box” issue, especially when multiple complex models are combined, which may hinder clinicians’ ability to interpret the specific contributions of individual predictors.

Current research gaps and limitations

Across all ML techniques, several critical limitations were identified that hinder the translation of research findings into clinical practice. Limited real-world validation represents a significant gap, as most studies lack validation in real clinical settings, with few reporting implementation outcomes or clinical impact assessments. This disconnects between research environments and actual clinical workflows creates uncertainty about how these models would perform when deployed in busy EDs with real patients and clinical staff.

Table 3 Comparative performance of ML techniques in ED predictive modelling

ML technique	Mortality prediction AUC-ROC	ICU admission AUC-ROC	Discharge prediction AUC-ROC	Key advantages	Primary limitations
Logistic regression (52)	0.68–0.82	0.65–0.79	0.70–0.81	High interpretability, fast training	Poor non-linear performance
Decision trees (53)	0.70–0.83	0.68–0.80	0.72–0.84	Good interpretability, handles interactions	Prone to overfitting
Random forests (54)	0.78–0.91	0.76–0.88	0.80–0.90	Robust performance, handles complex data	Reduced interpretability
Neural networks (55)	0.85–0.96	0.82–0.94	0.84–0.93	Highest predictive accuracy	Black-box nature, data hungry
Ensemble models (56)	0.82–0.94	0.80–0.92	0.83–0.91	Balanced accuracy and robustness	Computational complexity

AUC-ROC, area under the receiver operating characteristic curve; ED, emergency department; ML, machine learning.

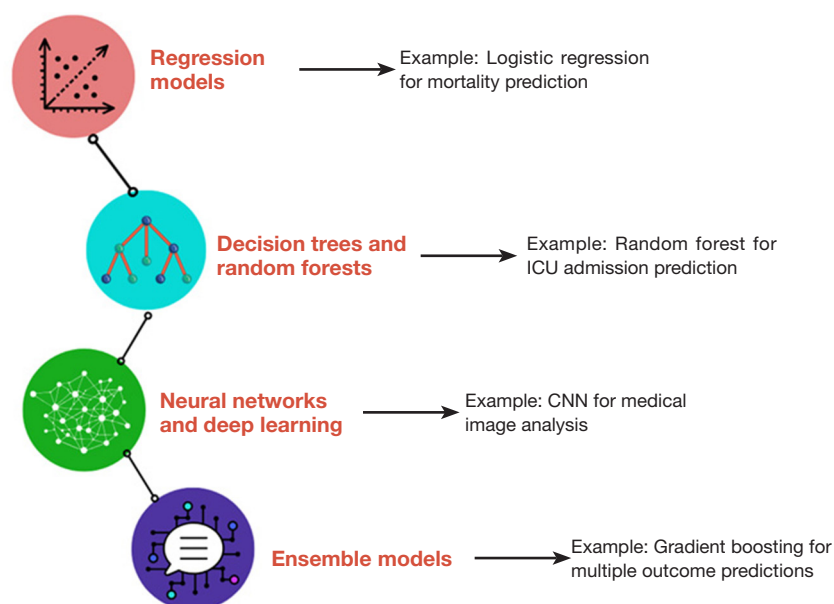


Figure 2 Machine learning techniques for predictive modelling in EDs. This figure illustrates the progression of machine learning techniques used for predictive modeling in EDs, showcasing four main categories: regression models, decision trees and random forests, neural networks and deep learning, and ensemble models. CNN, convolutional neural networks; EDs, Emergency Departments; ICU, intensive care unit.

Generalisability concerns pose another substantial challenge, as models often perform poorly when applied to different institutions or patient populations, indicating limited external validity. These performance drops highlight the models' tendency to overfit to specific datasets or institutional practices, raising questions about their broader applicability across diverse healthcare settings. The lack of

robust external validation studies further compounds this limitation.

Interpretability challenges persist even among simpler models, which face adoption barriers due to insufficient explanation of clinical reasoning behind predictions. Clinicians require clear understanding of how predictions are generated to trust and act upon model

recommendations, yet many studies fail to provide adequate interpretability frameworks or user-friendly explanation interfaces. Additionally, temporal stability remains poorly addressed, with few studies assessing model performance degradation over time or providing protocols for model updating and maintenance. This oversight is particularly concerning given that healthcare data patterns and clinical practices evolve continuously, potentially rendering static models obsolete or inaccurate over time.

Predictive models for specific outcomes

Predictive models have been developed to forecast a range of critical outcomes in the ED, aiding in timely and effective decision-making. These models focus on specific predictions such as mortality risk, ICU admission, and discharge probability, each of which has distinct clinical implications (57,58). By enhancing clinicians' ability to anticipate these outcomes, predictive models contribute to improved patient care and more efficient ED operations. However, significant variations in model performance and clinical applicability exist across different outcome types.

Mortality prediction

Mortality prediction models are designed to assess a patient's risk of death within a short timeframe, providing a critical tool for identifying patients who require immediate, intensive care. Traditionally, mortality risk has been evaluated using clinical scoring systems like the MEWS and the SOFA, which incorporate various physiological measures to estimate risk (59,60). However, these scores have limitations in accuracy and adaptability, particularly in diverse patient populations.

Across the reviewed studies, ML-based mortality prediction models demonstrated AUC-ROC values ranging from 0.75–0.96, significantly outperforming traditional scoring systems (AUC-ROC 0.65–0.80). The most successful models incorporated both real-time physiological data and historical comorbidity information, with ensemble and deep learning approaches showing superior performance (61). Integrating ML with traditional scoring systems has been shown to enhance the precision of mortality prediction as indicated that ML models, particularly ensemble and deep learning approaches, can outperform conventional scoring systems by effectively managing high-dimensional, complex data (62).

Ensemble models such as gradient boosting and random

forests have demonstrated strong performance in mortality prediction by combining multiple algorithms to mitigate the limitations of individual models. Deep learning models, especially those with recurrent or convolutional architectures, can further refine predictions by identifying subtle patterns in time-series data, such as changes in vital signs over time. These models also have the flexibility to integrate diverse data sources, including lab results, imaging reports, and clinical notes, improving risk stratification for ED patients. For instance, deep learning models leveraging electronic health records (EHRs) could predict in-hospital mortality with higher accuracy than conventional scoring systems (63,64).

However, mortality prediction models face significant challenges including class imbalance (with mortality rates typically <5% in general ED populations), difficulty in defining appropriate prediction time horizons, and ethical concerns regarding the clinical use of mortality predictions in triage decisions. Additionally, model performance varies substantially across different patient subgroups, with reduced accuracy in elderly patients and those with multiple comorbidities (65).

ICU admission

Accurate prediction of ICU admission is critical in the ED, as it enables proactive resource allocation and bed management, especially in hospitals with limited ICU capacity. Predictive models for ICU admission rely on a combination of physiological data, such as vital signs and lab results, and demographic factors, including age and comorbidities. ML techniques like random forests and neural networks have shown promising results in predicting ICU needs, as these models excel in managing complex, non-linear data interactions. Random forests, with their ability to capture variable interactions, can identify high-risk patients who may require ICU care based on a broad set of factors, such as severe respiratory distress, abnormal lab values, and deteriorating clinical status.

ICU admission prediction models demonstrated AUC-ROC values ranging from 0.76–0.94 across the reviewed studies, with neural networks and ensemble methods showing the strongest performance. Models incorporating laboratory values and imaging results achieved higher accuracy than those relying solely on vital signs and demographic data (66).

Neural networks, particularly RNNs, are well-suited for analyzing time-series data, making them useful for ICU

Table 4 Predictive model performance by outcome type

Outcome	Best performing technique	AUC-ROC range	Key predictive features	Primary clinical challenges	Implementation barriers
Mortality (72)	Deep learning/ensemble	0.75–0.96	Vital signs, lab values, comorbidities	Class imbalance, ethical considerations	Regulatory approval, clinical acceptance
ICU admission (73)	Neural networks/ensemble	0.76–0.94	Physiological instability, organ dysfunction	Variable admission criteria	Real-time data integration
Discharge (74)	Gradient boosting/ensemble	0.70–0.93	Treatment response, social factors	Non-medical discharge barriers	Dynamic factor incorporation

AUC-ROC, area under the receiver operating characteristic curve.

admission prediction in patients with rapidly changing conditions (67). For example, if a patient's heart rate, respiratory rate, or oxygen saturation levels fluctuate significantly over time, RNNs can capture these trends to improve ICU admission predictions. A study highlighted that RNN-based models trained on EHR data could predict ICU admissions more accurately than traditional logistic regression models, underscoring the potential of neural networks in this domain (68). The predictive power of these models can assist ED clinicians in prioritizing ICU resources, ensuring that critically ill patients receive the necessary level of care without delays.

Key limitations in ICU admission prediction include significant variation in ICU admission criteria between institutions, seasonal variations in ICU bed availability that affect prediction utility, and difficulty distinguishing between patients requiring ICU-level monitoring versus active intervention. Furthermore, models trained at one institution often show reduced performance when applied elsewhere due to differences in admission practices and patient populations (69).

Discharge probability

Predicting the likelihood of discharge is essential for optimizing patient flow and reducing ED overcrowding, a common challenge in many healthcare systems. Discharge probability models help clinicians identify patients who are likely to be safely discharged within a short timeframe, allowing for more efficient use of ED resources and reducing patient wait times. Logistic regression has traditionally been used to predict discharge likelihood, providing a straightforward approach to modeling binary outcomes (i.e., discharge or admission). However, recent advances have shown that more complex models, such as

gradient boosting, can significantly enhance prediction accuracy by incorporating a wider range of factors, including time-sensitive variables.

Discharge prediction models showed AUC-ROC values ranging from 0.70–0.93, with ensemble methods demonstrating superior performance. Models that incorporated treatment response variables and time-to-disposition factors achieved higher accuracy than those using only initial presentation data (70).

Gradient boosting models can integrate real-time factors like lab turnaround times, initial treatment response, and the patient's clinical improvement or deterioration over the course of their ED stay. By dynamically updating discharge predictions based on these factors, gradient boosting models can provide clinicians with near real-time insights into patient status. For instance, a study found that incorporating lab results and treatment progress in discharge prediction models increased accuracy and helped prevent unnecessary admissions, which can contribute to ED overcrowding (71).

However, discharge prediction faces unique challenges including high variability in discharge criteria between clinicians, significant influence of non-medical factors (bed availability, social circumstances), and difficulty in incorporating time-dependent variables that affect discharge timing. Additionally, discharge models must balance sensitivity to prevent inappropriate early discharge with specificity to avoid unnecessary admissions (71).

Table 4 summarises the performance characteristics and clinical considerations for different outcome predictions.

Figure 3 provides a comprehensive visual summary of how different predictive modeling techniques are applied to critical decision-making processes in EDs, showcasing the progression towards more advanced and accurate prediction methods. These models have particular value in supporting the management of patient flow, ensuring that

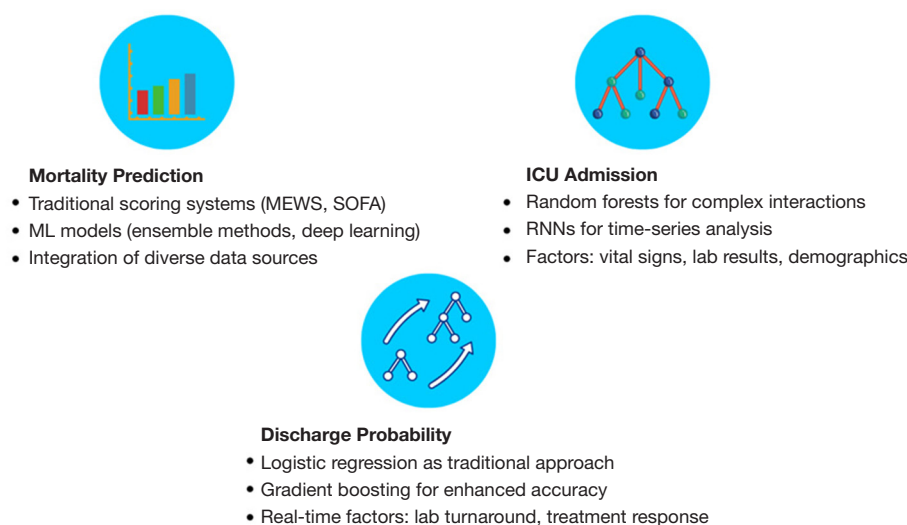


Figure 3 Predictive models for critical outcomes in EDs. This figure illustrates three key predictive models used in EDs to forecast critical patient outcomes. The diagram is divided into three panels, each representing a specific prediction type: mortality prediction, ICU admission, and discharge probability. EDs, Emergency Departments; ICU, intensive care unit; MEWS, Modified Early Warning Score; ML, machine learning; RNNs, Recurrent Neural Networks; SOFA, Sequential Organ Failure Assessment.

beds are available for new arrivals and helping ED staff make more informed decisions patient care and discharge readiness. However, successful clinical implementation requires addressing significant methodological limitations and developing robust validation frameworks.

Model evaluation and performance metrics

Evaluating the performance of predictive models is essential to ensure that they provide reliable and actionable insights in the ED. The unique demands of the ED environment, where rapid, high-stakes decisions are routine, require models that are not only accurate but also sensitive to critical cases. The effectiveness of these models is assessed using a variety of performance metrics, each of which provides distinct information about model strengths and weaknesses in the clinical context. However, significant inconsistencies in evaluation approaches across studies limit the ability to compare models and assess real-world clinical utility. Commonly used metrics include accuracy, sensitivity, specificity, AUC-ROC, F1 score, and model calibration.

Standard performance metrics

Accuracy is a fundamental metric that measures the proportion of correct predictions (both true positives and true negatives) out of all predictions made by the model.

While accuracy is useful for assessing overall model performance, it may be less informative in ED predictive modeling, where positive outcomes (e.g., ICU admission or mortality) are often rare compared to negative outcomes (e.g., discharge) (75). In these cases, a high accuracy score could mask poor performance in detecting critical cases, as the model may achieve high accuracy by predominantly predicting negative outcomes.

Sensitivity (also known as recall or true positive rate) is particularly crucial in mortality and ICU admission prediction models, as it measures the model's ability to correctly identify patients who truly belong to the positive class (e.g., high-risk patients). High sensitivity is often prioritized in these models to ensure that critical patients are identified early, as failing to detect such cases could result in missed or delayed interventions with serious consequences. Across the reviewed studies, sensitivity values varied widely from 0.65–0.92 for mortality prediction and 0.70–0.89 for ICU admission prediction, highlighting significant inconsistencies in model performance for critical case detection (76). In mortality prediction, for instance, a high-sensitivity model helps clinicians identify and prioritize patients who may require intensive monitoring or treatment, minimizing the risk of underestimating critical cases.

Specificity (or true negative rate), on the other hand, measures the model's ability to correctly identify patients

who do not belong to the positive class. High specificity is important in predicting discharge probability, where accurate identification of low-risk patients can support efficient bed management and reduce unnecessary admissions. Specificity values ranged from 0.72–0.95 across different outcomes, with discharge prediction models generally achieving higher specificity than mortality or ICU admission models (77).

The AUC-ROC is a comprehensive metric that provides insight into a model's ability to distinguish between positive and negative cases across various thresholds. The AUC-ROC score ranges from 0 to 1, with higher scores indicating better discriminatory power. An AUC-ROC of 0.5 suggests no discriminatory ability, equivalent to random guessing, whereas a score closer to 1 indicates strong discriminatory performance. While AUC-ROC is widely reported, fewer than 40% of reviewed studies provided confidence intervals or statistical significance testing for AUC comparisons. The AUC-ROC metric is widely used in ED predictive modeling to compare the overall effectiveness of different models, as it considers both sensitivity and specificity.

The F1 score is a harmonic mean of sensitivity and precision (the proportion of true positive predictions out of all positive predictions), making it particularly valuable when there is an imbalance between positive and negative outcomes, as is often the case in ED predictive tasks. F1 scores were reported in only 60% of reviewed studies, limiting comparative analysis of model performance in handling class imbalance (78). The F1 score helps assess a model's performance in identifying true positive cases while minimizing false positives. For instance, in predicting ICU admission, a high F1 score indicates that the model is both sensitive to high-risk patients and precise in minimizing unnecessary ICU recommendations, helping to balance patient safety with resource utilization.

Model calibration and clinical utility

Model calibration is another critical component in evaluating predictive models in clinical settings. Calibration assesses whether the predicted probabilities generated by the model align with actual observed outcomes. A well-calibrated model provides probabilities that accurately reflect the likelihood of outcomes, which is crucial in clinical decision-making. For example, if a mortality prediction model outputs a 20% risk of death, this should correspond closely to an actual 20% mortality rate in patients with similar risk scores. However, calibration assessment

was reported in fewer than 30% of reviewed studies, representing a significant gap in model evaluation practices. Poor calibration can lead to over- or underestimation of risk, which may mislead clinicians in assessing patient urgency or appropriate levels of care. Techniques like reliability diagrams and calibration curves are commonly used to assess model calibration, ensuring that predictions align with real-world outcomes.

Current limitations in model evaluation

Several critical limitations in model evaluation practices were identified across the reviewed studies:

- (I) Inconsistent metrics: significant variation in reported performance metrics limits the ability to compare models across studies and identify optimal approaches.
- (II) Limited external validation: only 25% of studies included external validation, with most relying solely on internal cross-validation or temporal split validation.
- (III) Insufficient calibration assessment: poor reporting of model calibration limits understanding of real-world prediction reliability.
- (IV) Lack of clinical impact metrics: few studies assessed clinical utility measures such as decision curve analysis, net benefit, or clinical implementation outcomes.
- (V) Missing subgroup analysis: limited evaluation of model performance across different patient populations (e.g., elderly, paediatric, specific disease groups).
- (VI) Temporal validation gaps: insufficient assessment of model performance degradation over time or seasonal variations.

Recommendations for improved evaluation

Based on the identified limitations, future studies should adopt standardised evaluation frameworks that include:

- (I) Core metric set: consistent reporting of AUC-ROC, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score with confidence intervals.
- (II) Calibration assessment: mandatory inclusion of calibration plots and statistical tests for all prediction models.
- (III) External validation: multi-centre validation studies

to assess model generalisability across different settings.

- (IV) Clinical utility measures: integration of decision curve analysis and net benefit calculations to assess clinical value.
- (V) Subgroup analysis: systematic evaluation of model performance across relevant patient subgroups.
- (VI) Temporal stability: assessment of model performance over extended time periods with regular recalibration protocols.

Future directions

The field of predictive modeling in the ED is evolving rapidly, with ongoing advancements that hold the potential to further enhance patient outcomes and streamline healthcare delivery. The future of ED predictive models will likely be characterized by increased personalization, greater interpretability, and integration with emerging technologies such as telemedicine. However, significant research gaps and implementation challenges must be addressed to realise the full potential of these advances. Each of these directions promises to address current limitations, making predictive models more accurate, trustworthy, and applicable to a broader range of clinical scenarios.

Personalized prediction models

As the demand for precision medicine grows, there is increasing interest in developing predictive models that account for individual patient characteristics beyond standard clinical data. Personalized prediction models incorporate features such as genomic information, lifestyle factors, and social determinants of health, providing more tailored and nuanced predictions (79-82).

For instance, genomic data can offer insights into a patient's predisposition to certain diseases or their likely response to specific treatments, which may be especially relevant in predicting outcomes for patients with complex or rare conditions. Lifestyle factors, such as exercise habits, smoking status, and diet, have been shown to influence outcomes for various health conditions and could further refine ED predictions. Social determinants of health, including socioeconomic status, education level, and access to healthcare resources, are also important, as they can impact both health outcomes and the effectiveness of interventions. By integrating these factors, future predictive models could produce more individualized

recommendations, allowing clinicians to make decisions that are better aligned with each patient's unique needs and circumstances. However, current research in personalised ED prediction models remains limited, with fewer than 15% of reviewed studies incorporating genetic or detailed social determinant data.

Key challenges for personalised models include data privacy concerns, the need for comprehensive data integration platforms, increased computational requirements, and the complexity of obtaining detailed individual patient data in emergency settings. Furthermore, the clinical utility of genetic information in acute care decisions remains unclear, requiring substantial additional research to demonstrate cost-effectiveness and clinical impact.

XAI

One of the most significant challenges in adopting ML models in clinical settings is their often "black-box" nature, where complex models produce predictions without clear explanations of the underlying reasoning. This lack of transparency can create barriers to clinician trust and acceptance, especially in high-stakes environments like the ED, where decisions must be made quickly and with confidence. XAI aims to address this challenge by making predictive models more interpretable (83,84). Through XAI, clinicians would be able to understand why a model made a particular prediction, potentially viewing contributing factors, data patterns, or decision paths that led to the outcome. For example, XAI techniques like Shapley Additive Explanations (SHAP) values can quantify the contribution of each input feature (e.g., age, blood pressure, lab results) to a specific prediction, providing clinicians with a clearer picture of the underlying reasoning (85,86).

Current XAI implementations in ED settings show promise, with several studies demonstrating improved clinician acceptance when explanations are provided. However, significant limitations remain including computational overhead for generating explanations, inconsistency in explanation quality across different cases, and lack of standardised approaches for presenting explanations in clinical workflows. As XAI technologies advance, they could play a pivotal role in increasing clinician confidence and facilitating the integration of predictive models into ED workflows, ultimately leading to better-informed, data-driven decisions that align with clinical intuition and expertise.

Key areas requiring further research include optimal explanation granularity for different clinical contexts, standardisation of explanation formats, integration with existing clinical decision support systems, and validation of explanation accuracy and clinical utility.

Telemedicine integration

The rise of telemedicine has expanded the potential applications of predictive models, particularly by enabling remote monitoring and early intervention. Telemedicine allows patients to connect with healthcare providers from their homes, and in some cases, wearable devices continuously transmit real-time data, such as heart rate, oxygen saturation, and activity levels (87). Future predictive models could leverage this data to anticipate ED outcomes even before a patient arrives at the hospital, allowing for preemptive interventions when necessary. For example, if a remote monitoring system detects that a patient's vital signs are deteriorating, a predictive model could alert clinicians to the need for an immediate ED visit, expediting the patient's arrival and enabling the ED to prepare for their specific needs. This approach could be particularly beneficial for managing chronic conditions or high-risk populations, such as elderly patients or those with multiple comorbidities, who may experience rapid health changes that require prompt attention. However, telemedicine-integrated prediction models face challenges including data quality from consumer-grade devices, connectivity issues, privacy concerns, and the need for real-time processing capabilities.

Current research in telemedicine integration remains early-stage, with limited evidence for clinical effectiveness and cost-benefit analysis. Key research priorities include validating wearable device data accuracy, developing standardised protocols for remote monitoring alerts, establishing clear clinical pathways for telemedicine-triggered interventions, and evaluating patient outcomes and satisfaction with integrated prediction systems.

Implementation science and real-world validation

A critical gap identified across the reviewed literature is the lack of real-world implementation studies and clinical impact assessment. Future research must prioritise:

- (I) Multi-centre validation studies: large-scale validation across diverse healthcare systems to establish model generalisability and performance consistency.

- (II) Clinical implementation trials: randomised controlled trials evaluating the clinical impact of ML-assisted decision making on patient outcomes, ED efficiency, and cost-effectiveness.
- (III) Human factors research: studies examining the integration of predictive models into clinical workflows, including user interface design, alert fatigue prevention, and workflow optimisation.
- (IV) Long-term performance monitoring: development of frameworks for continuous model monitoring, performance tracking, and systematic updating protocols.

Regulatory and ethical considerations

Future development must address several regulatory and ethical challenges:

- (I) Regulatory approval pathways: development of standardised regulatory frameworks for ML-based clinical decision support tools in emergency medicine.
- (II) Bias and fairness: systematic evaluation and mitigation of algorithmic bias across different patient populations, particularly underrepresented groups.
- (III) Data governance: establishment of robust data governance frameworks ensuring patient privacy while enabling model development and validation.
- (IV) Clinical liability: clarification of legal responsibilities and liability frameworks for ML-assisted clinical decisions.

Figure 4 provides a comprehensive visual summary of how predictive modeling in EDs is expected to evolve, showcasing the potential for more personalized, explainable, and proactive approaches to emergency care. By integrating telemedicine data, predictive models could support a more proactive approach to emergency care, reducing the likelihood of delayed treatment and improving outcomes for vulnerable patients. However, realising these future directions requires substantial investment in research infrastructure, regulatory framework development, and systematic validation studies to demonstrate clinical utility and cost-effectiveness.

Limitations of the review

This narrative review has several limitations that should be acknowledged. First, as a narrative rather than systematic

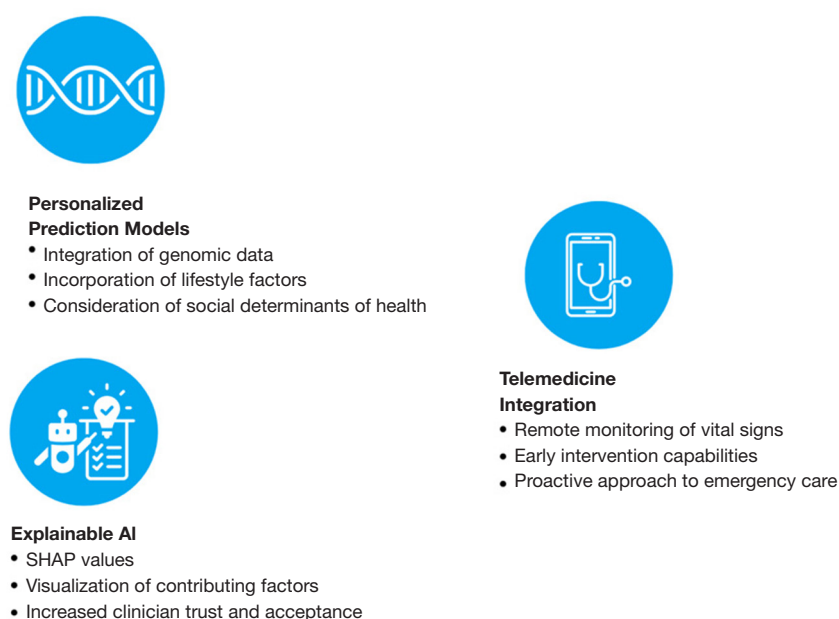


Figure 4 Future directions in predictive modeling for EDs. This figure illustrates three key future directions in predictive modeling for EDs: personalized prediction models, explainable AI, and telemedicine integration. The diagram is divided into three interconnected panels, each representing a specific area of advancement. AI, artificial intelligence; EDs, emergency departments; SHAP, Shapley Additive Explanations.

review, this study was not conducted according to a structured protocol such as PRISMA; therefore, it may be subject to selection bias in the identification and inclusion of relevant studies. Second, the search was limited to English-language publications, which may have excluded relevant research published in other languages, potentially affecting the global comprehensiveness of the findings. Third, while efforts were made to include recent and high-quality studies, the rapid evolution of ML in healthcare means that some emerging models or unpublished innovations may have been missed. Fourth, the heterogeneity among the included studies, in terms of data sources, model types, outcome definitions, and performance metrics, limits the ability to directly compare results across studies. Additionally, heterogeneity among the included studies, in terms of data sources, model types, outcome definitions, and performance metrics, limits the ability to directly compare results across studies. Fifth, the review did not systematically assess study quality using standardised assessment tools, which may affect the reliability of synthesised findings. Sixth, the focus on English-language publications from major databases may have introduced geographical and publication bias. Finally, as this is a narrative rather than quantitative synthesis, we

did not perform a meta-analysis, which could have provided pooled estimates of model performance. Finally, the rapid pace of technological advancement in this field means that some findings may become outdated quickly, requiring regular updates to maintain relevance. Despite these limitations, this review provides a valuable overview of the current landscape of ML-based predictive modelling in the ED and highlights key areas for future research and clinical integration.

Conclusions

ML-based predictive models represent a transformative advancement in the ED, offering a powerful tool to enhance patient outcomes through timely, data-driven decision-making. These models can analyze large volumes of clinical data, providing rapid predictions that assist clinicians in triaging patients, prioritizing critical interventions, and efficiently managing ED resources. By predicting key outcomes, such as mortality, ICU admission, and discharge likelihood, predictive models support a proactive approach to emergency care, where high-risk patients can be identified early, and resources can be allocated to optimize

patient flow.

This comprehensive review demonstrates that ML models consistently outperform traditional clinical scoring systems, with AUC-ROC values ranging from 0.75–0.96 for mortality prediction, 0.76–0.94 for ICU admission, and 0.70–0.93 for discharge prediction. Ensemble methods and deep learning approaches showed superior performance, though significant implementation challenges remain.

Despite their potential, several challenges remain in the implementation of ML models in the ED. Critical limitations identified include substantial data quality issues with missing value rates of 15–70% for key clinical variables, limited external validation with only 25% of studies including multi-centre validation, poor model generalisability across different institutions and patient populations, and insufficient attention to model calibration and clinical utility assessment. Data quality is paramount; without consistent, comprehensive, and high-quality data, model predictions may lack reliability. The interpretability of complex ML models also presents a barrier, as clinicians require clear, understandable explanations to fully trust and act on predictions. Ethical considerations, including data privacy, bias, and the need for accountability, must be carefully addressed to ensure that predictive models uphold the standards and values of patient care. Nevertheless, ongoing research and development in areas such as XAI and model calibration are making these models more accessible, interpretable, and aligned with clinical practice.

Key areas requiring immediate research attention include developing standardised evaluation frameworks, conducting multi-centre validation studies, addressing algorithmic bias and fairness concerns, establishing regulatory approval pathways, and implementing robust clinical impact assessment protocols. Only 15% of reviewed studies incorporated detailed social determinant data, and fewer than 30% assessed model calibration, highlighting significant methodological gaps.

The future of predictive modeling in the ED is promising, particularly with the integration of personalized prediction models, advances in explainability, and telemedicine capabilities. Personalized models, incorporating genomic, lifestyle, and social data, will offer tailored insights for individual patients, enhancing precision in care. XAI can foster greater clinician trust by providing transparency around model predictions, facilitating smoother integration into clinical workflows. The integration of telemedicine can enable preemptive risk assessment and early intervention, transforming ED care for patients with chronic or high-

risk conditions. However, these advances face substantial implementation barriers including data privacy concerns, computational requirements, regulatory uncertainty, and the need for extensive validation studies to demonstrate clinical utility and cost-effectiveness.

As these advancements continue, ML-based predictive models will increasingly support the ED's capacity to provide efficient, patient-centered care, aligning with the evolving demands and challenges of modern healthcare. Success will depend on addressing fundamental challenges in data standardisation, model validation, clinical integration, and regulatory approval whilst maintaining focus on demonstrated clinical utility rather than purely technical performance metrics.

Acknowledgments

We extend our sincere appreciation to the Emergency Department and Research and Innovation Department of Medway NHS Foundation Trust for their support and collaboration in this review.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-25-83/rc>

Peer Review File: Available at <https://atm.amegroups.com/article/view/10.21037/atm-25-83/prf>

Funding: None.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-25-83/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Berchet C. Emergency Care Services: Trends, Drivers and Interventions to Manage the Demand. OECD Health Working Papers, No. 83. Paris: OECD Publishing; 2015.
- Castro MG, Wang MC. Quality of Life and Patient-Centered Outcomes. In: Daaleman TP, Helton MR. editors. Chronic Illness Care. Springer, Cham; 2023.
- Salwei ME, Carayon P, Hoonakker PLT, et al. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. *Appl Ergon* 2021;97:103498.
- Da'Costa A, Teke J, Origbo JE, et al. AI-driven triage in emergency departments: A review of benefits, challenges, and future directions. *Int J Med Inform* 2025;197:105838.
- Porto BM, Fogliatto FS. Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. *BMC Med Inform Decis Mak* 2024;24:377.
- Lim L, Gim U, Cho K, et al. Real-time machine learning model to predict short-term mortality in critically ill patients: development and international validation. *Crit Care* 2024;28:76.
- Zilker S, Weinzierl S, Kraus M, et al. A machine learning framework for interpretable predictions in patient pathways: The case of predicting ICU admission for patients with symptoms of sepsis. *Health Care Manag Sci* 2024;27:136-67.
- Naemi A, Schmidt T, Mansourvar M, et al. Machine learning techniques for mortality prediction in emergency departments: a systematic review. *BMJ Open* 2021;11:e052663.
- Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019;23:64.
- Porto BM. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. *BMC Emerg Med* 2024;24:219.
- Calegari R, Fogliatto FS, Lucini FR, et al. Forecasting Daily Volume and Acuity of Patients in the Emergency Department. *Comput Math Methods Med* 2016;2016:3863268.
- Jeong J, Lee SW, Kim WY, et al. Development and validation of a scoring system for mortality prediction and application of standardized W statistics to assess the performance of emergency departments. *BMC Emerg Med* 2021;21:71.
- Rahmatinejad Z, Hoseini B, Rahmatinejad F, et al. Internal Validation of the Predictive Performance of Models Based on Three ED and ICU Scoring Systems to Predict Inhospital Mortality for Intensive Care Patients Referred from the Emergency Department. *Biomed Res Int* 2022;2022:3964063.
- Bonnett LJ, Snell KIE, Collins GS, et al. Guide to presenting clinical prediction models for use in clinical settings. *BMJ* 2019;365:l737.
- Rahmatinejad Z, Tohidinezhad F, Rahmatinejad F, et al. Internal validation and comparison of the prognostic performance of models based on six emergency scoring systems to predict in-hospital mortality in the emergency department. *BMC Emerg Med* 2021;21:68.
- Jeong S. Scoring Systems for the Patients of Intensive Care Unit. *Acute Crit Care* 2018;33:102-4.
- Pellathy TP, Pinsky MR, Hravnak M. Intensive Care Unit Scoring Systems. *Crit Care Nurse* 2021;41:54-64.
- Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. *Indian J Crit Care Med* 2014;18:220-8.
- Delgado-Hurtado JJ, Berger A, Bansal AB. Emergency department Modified Early Warning Score association with admission, admission disposition, mortality, and length of stay. *J Community Hosp Intern Med Perspect* 2016;6:31456.
- Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med* 1998;26:1793-800.
- Bian Y, Zhang P, Xiong Y, et al. Application of the APACHE II score to assess the condition of patients with critical neurological diseases. *Acta Neurol Belg* 2015;115:651-6.
- Fernando SM, Tran A, Taljaard M, et al. Prognostic Accuracy of the Quick Sequential Organ Failure Assessment for Mortality in Patients With Suspected Infection: A Systematic Review and Meta-analysis. *Ann Intern Med* 2018;168:266-75.
- Scott LJ, Redmond NM, Garrett J, et al. Distributions of the National Early Warning Score (NEWS) across a

- healthcare system following a large-scale roll-out. *Emerg Med J* 2019;36:287-92.
24. Lee YC, Ng CJ, Hsu CC, et al. Machine learning models for predicting unscheduled return visits to an emergency department: a scoping review. *BMC Emerg Med* 2024;24:20.
 25. Elhaj H, Achour N, Tania MH, et al. A comparative study of supervised machine learning approaches to predict patient triage outcomes in hospital emergency departments. *Array* 2023;17:100281.
 26. Brekke IJ, Puntervoll LH, Pedersen PB, et al. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. *PLoS One* 2019;14:e0210875.
 27. Henning DJ, Oedorf K, Day DE, et al. Derivation and Validation of Predictive Factors for Clinical Deterioration after Admission in Emergency Department Patients Presenting with Abnormal Vital Signs Without Shock. *West J Emerg Med* 2015;16:1059-66.
 28. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol* 2022;22:287.
 29. AbouHassan I, Kasabov NK, Bankar T, et al. ePAMeT: evolving predictive associative memories for time series. *Evolving Systems* 2025;16:6.
 30. Pan P, Wang Y, Liu C, et al. (2024). Revisiting the potential value of vital signs in the real-time prediction of mortality risk in intensive care unit patients. *J Big Data* 2024;11:53.
 31. Badawy M, Ramadan N, Hefny HA. Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Inf Technol* 2023;10:40.
 32. McGuckin T, Crick K, Myroniuk TW, et al. Understanding challenges of using routinely collected health data to address clinical care gaps: a case study in Alberta, Canada. *BMJ Open Qual* 2022;11:e001491.
 33. Ben-Assuli O, Vest JR. Return visits to the emergency department: An analysis using group based curve models. *Health Informatics J* 2022;28:14604582221105444.
 34. Torab-Miandoab A, Samad-Soltani T, Jodati A, et al. Interoperability of heterogeneous health information systems: a systematic literature review. *BMC Med Inform Decis Mak* 2023;23:18.
 35. Komamine M, Fujimura Y, Omiya M, et al. Dealing with missing data in laboratory test results used as a baseline covariate: results of multi-hospital cohort studies utilizing a database system contributing to MID-NET(®) in Japan. *BMC Med Inform Decis Mak* 2023;23:242.
 36. Hu H, Liu Y, Zhao Y, et al. Detecting temporal inconsistency in biased datasets for Android malware detection. 2023 38th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW); 11-15 September 2023; Luxembourg, Luxembourg. IEEE; 2023.
 37. Kang MW, Kim J, Kim DK, et al. Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. *Crit Care* 2020;24:42.
 38. Blythe R, Parsons R, Barnett AG, et al. Prioritising deteriorating patients using time-to-event analysis: prediction model development and internal-external validation. *Crit Care* 2024;28:247.
 39. Li L, Rysavy MA, Bobashev G, et al. Comparing methods for risk prediction of multicategory outcomes: dichotomized logistic regression vs. multinomial logit regression. *BMC Med Res Methodol* 2024;24:261.
 40. Soave DM, Strug LJ. Testing Calibration of Cox Survival Models at Extremes of Event Risk. *Front Genet* 2018;9:177.
 41. Sung CW, Ho J, Fan CY, et al. Prediction of high-risk emergency department revisits from a machine-learning algorithm: a proof-of-concept study. *BMJ Health Care Inform* 2024;31:e100859.
 42. Zhang Y, Huang Y, Rosen A, et al. Aspiring to clinical significance: Insights from developing and evaluating a machine learning model to predict emergency department return visit admissions. *PLOS Digit Health* 2024;3:e0000606.
 43. Zhang D, Yin C, Zeng J, et al. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak* 2020;20:280.
 44. Mubarak AA, Cao H, Ahmed SA. (2021). Predictive learning analytics using deep learning model in MOOCs' courses videos. *Educ Inf Technol* 2021;26:371-92.
 45. Li X, Ge P, Zhu J, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. *PeerJ* 2020;8:e10337.
 46. Li X, Zhang Y, Cheng H, et al. Student achievement prediction using deep neural network from multi-source campus data. *Complex Intell Syst* 2022;8:5143-56.
 47. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85-117.
 48. Chen YW, Li YJ, Deng P, et al. Learning to predict in-hospital mortality risk in the intensive care unit with attention-based temporal convolution network. *BMC*

- Anesthesiol 2022;22:119.
49. Ali R, Hardie RC, Narayanan BN, et al. Deep learning ensemble methods for skin lesion analysis towards melanoma detection. 2019 IEEE National Aerospace and Electronics Conference (NAECON); 15-19 July 2019; Dayton, OH, USA. IEEE; 2019:311-6.
 50. Zubair Hasan KM, Zahid Hasan M. Performance evaluation of ensemble-based machine learning techniques for prediction of chronic kidney disease. In: Shetty N, Patnaik L, Nagaraj H, et al. editors. *Emerging Research in Computing, Information, Communication and Applications*; Springer: Berlin/Heidelberg, Germany, 2019:415-26.
 51. Roudnitski A. "Evaluating Road Crash Severity Prediction with Balanced Ensemble Models." Findings, April. 2024. Available online: <https://doi.org/10.32866/001c.116820>
 52. Bobbitt, Z. How to interpret a ROC curve (with examples). Statology. 2021. Available online: <https://www.statology.org/interpret-roc-curve/>
 53. Mokarram R, Emadi M. Classification in Non-linear Survival Models Using Cox Regression and Decision Tree. *Ann Data Sci* 2017;4:329-40.
 54. Çınaroğlu S. Comparison of performance of decision tree algorithms and random forest: An application on OECD countries health expenditures. *International Journal of Computer Applications* 2016;138:37-41.
 55. Namdar K, Haider MA, Khalvati F. A Modified AUC for Training Convolutional Neural Networks: Taking Confidence Into Account. *Front Artif Intell* 2021;4:582928.
 56. Cornall R. Ensemble Machine Learning: Stroke risk prediction. GitHub repository. 2025. Available online: <https://github.com/richardcornall/Ensemble-Machine-Learning-Stroke-risk-prediction>
 57. Atkinson P, McGeorge K, Innes G. Saving emergency medicine: is less more? *CJEM* 2022;24:9-11.
 58. Morley C, Unwin M, Peterson GM, et al. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS One* 2018;13:e0203316.
 59. Oprita B, Aignatoaie B, Gabor-Postole DA. Scores and scales used in emergency medicine. *Practicability in toxicology. J Med Life* 2014;7 Spec No. 3:4-7.
 60. Wang L, Lv Q, Zhang X, et al. The utility of MEWS for predicting the mortality in the elderly adults with COVID-19: a retrospective cohort study with comparison to other predictive clinical scores. *PeerJ* 2020;8:e10018.
 61. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
 62. Abimannan S, El-Alfy ESM, Chang YS, et al. Ensemble Multifeatured Deep Learning Models and Applications: A Survey. in *IEEE Access* 2023;11:107194-217.
 63. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1:18.
 64. Samad MD, Ulloa A, Wehner GJ, et al. Predicting Survival From Large Echocardiography and Electronic Health Record Datasets: Optimization With Machine Learning. *JACC Cardiovasc Imaging* 2019;12:681-9.
 65. He M, Lin Y, Ren S, et al. Interpretable machine learning models for predicting in-hospital mortality in patients with chronic critical illness and heart failure: A multicenter study. *Digit Health* 2025;11:20552076251347785.
 66. Alghatani K, Ammar N, Rezgui A, et al. Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation. *JMIR Med Inform* 2021;9:e21347.
 67. Lin YW, Zhou Y, Faghri F, et al. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS One* 2019;14:e0218942.
 68. AlSaad B, Malluhi Q, Janahi I, et al. Predicting emergency department utilization among children with asthma using deep learning models. *Healthcare Analytics* 2022;2:100050.
 69. Saadatmand S, Salimifard K, Mohammadi R, et al. Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients. *Ann Oper Res* 2022. [Epub ahead of print]. doi: 10.1007/s10479-022-04984-x.
 70. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Digit Med* 2022;5:69.
 71. Cheng Q. A Data-Driven Approach to Describe, Predict, and Reduce Crowding and Health Disparities in an Emergency Department. 2023. Available online: <https://doi.org/10.17615/a2y6-fk49>
 72. Zhang Y, Chen L, Patel R. Improving emergency department discharge prediction using ensemble learning and dynamic treatment features. *Journal of Biomedical Informatics* 2023;138:104215.
 73. Wei J, Zhou J, Zhang Z, et al. Predicting individual patient and hospital-level discharge using machine learning. *Commun Med (Lond)* 2024;4:236.
 74. van der Vegt AH, Scott IA, Dermawan K, et al. Implementation frameworks for end-to-end clinical AI:

- derivation of the SALIENT framework. *J Am Med Inform Assoc* 2023;30:1503-15.
75. Chen Y, Chen H, Sun Q, et al. Machine learning model identification and prediction of patients' need for ICU admission: A systematic review. *Am J Emerg Med* 2023;73:166-70.
 76. Wu CP, Shirley RB, Milinovich A, et al. Exploring timely and safe discharge from ICU: a comparative study of machine learning predictions and clinical practices. *Intensive Care Med Exp* 2025;13:10.
 77. Haque I. Lies, damned lies, and AUC confidence intervals. Stanford University. Available online: https://cs.stanford.edu/people/ihaque/posters/EC4-AUC_Confidence_Intervals.pdf
 78. Zhai Q, Lin Z, Ge H, et al. Using machine learning tools to predict outcomes for emergency department intensive care unit patients. *Sci Rep* 2020;10:20919.
 79. Chen F, Zhang Y, Nguyen M, et al. Personalized choice prediction with less user information. *Ann Math Artif Intell* 2024;92:1489-1509.
 80. Visweswaran S, Ferreira A, Ribeiro GA, et al. Personalized Modeling for Prediction with Decision-Path Models. *PLoS One* 2015;10:e0131022.
 81. Ghebrehiwet I, Zaki N, Damseh R, et al. Revolutionizing personalized medicine with generative AI: a systematic review. *Artif Intell Rev* 2024;57:128.
 82. Zhang S, Yang F, Wang L, et al. Personalized prediction for multiple chronic diseases by developing the multi-task Cox learning model. *PLoS Comput Biol* 2023;19:e1011396.
 83. Mario Brcic M, Cabitza F, et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*. 2024;106:102301.
 84. Héder M. Explainable AI: A Brief History of the Concept. *ERCIM News* 2023;(134):9-10.
 85. Phillips PP, Hahn CA, Fontana PC, et al. "Four Principles of Explainable Artificial Intelligence". NIST. 2021. doi: 10.6028/nist.ir.8312.
 86. Vilone G, Longo L. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion* 2021;76:89-106.
 87. Stoltzfus M, Kaur A, Chawla A, et al. The role of telemedicine in healthcare: an overview and update. *Egypt J Intern Med* 2023;35:49.

Cite this article as: Olawade DB, Da'Costa A, Origbo JE, Osonuga A, Egbon E, Teke J, Boussios S. Smart emergency care: a narrative review of predictive machine learning models. *Ann Transl Med* 2025;13(5):60. doi: 10.21037/atm-25-83