

Review article

Artificial Intelligence for Emergency Department Triage and Risk Stratification: A Review

Saleh Menwar Ali ^{1,*} and Zubair Sharif ²

¹ Department of Emergency, SMC Hospital, Riyadh, Saudi Arabia

² Department of Medical Laboratory Technology, Faculty of Allied Health Sciences, Superior University, Lahore, Punjab, Pakistan

*Correspondence: dr.saleh2100.sa@gmail.com

Article information:

Received: 17-09-2026

Revised: 20-10-2026

Accepted: 21-10-2026

Published: 22-10-2026

Academic Editor:

Dr. Naveed Ahmed

Keywords:

Artificial intelligence

Emergency department

Risk stratification

Machine learning

Patient safety

Clinical implementation

Abstract: Researchers increasingly investigate artificial intelligence (AI) as a tool to support triage and early risk stratification in emergency departments. However, improved predictive performance does not necessarily translate into safer decisions or better patient outcomes. This critical narrative review examines the evidence for AI in adult ED care, with particular emphasis on prospective evaluation, patient safety, and clinical implementation. The review compares conventional triage systems, early warning scores, and clinician assessment with machine learning, natural language processing, and large language models. It examines commonly used data inputs, prediction timing, model discrimination, calibration, external validation, and the limitations of outcome definitions. Attention is given to the distinction between predicting clinical deterioration, critical care need, short-term mortality, and hospital admission. Evidence is considered across retrospective development, external validation, silent prospective evaluation, clinician-visible implementation, and randomized trials. AI may help identify patients requiring closer assessment by integrating information from multiple sources. Nevertheless, its clinical value depends on data quality, timely predictions, generalizability, and healthcare professionals' ability to interpret outputs appropriately. Important risks include under-triage, automation bias, unequal performance across patient groups, and increased clinical workload. Prospective studies and randomized trials are therefore essential to establish whether AI-supported decisions improve patient outcomes compared with usual care. Future research should prioritize multicenter clinical evaluations, consistent outcome reporting, subgroup assessment, and continuous monitoring after deployment. AI implementation should also include clear accountability, human oversight, and procedures for reviewing adverse events. Evidence of predictive accuracy alone is insufficient to justify routine adoption. Clinical integration should depend on demonstrated safety and meaningful benefit for patients.

Article citation: Ali, S. M. & Sharif, Z. Artificial Intelligence for Emergency Department Triage and Risk Stratification: A Review. Electron J Med Res. 2026. 2(4): 68-80. <https://doi.org/10.64813/ejmr.2026.147>

Introduction

Emergency departments play a central role in assessing and treating patients with acute illness and injury. Emergency clinicians must make rapid decisions about the urgency of care, the need for further investigation and the likelihood of serious clinical outcomes [1]. They often make these decisions before a complete

medical history, laboratory results, or a clear diagnosis is available. At the same time, increasing patient attendance, limited staffing, overcrowding and delays in access to hospital beds place additional pressure on emergency services. Under these conditions, identifying patients who require immediate attention

while avoiding unnecessary delays for others remains a major clinical and operational challenge [2].

Triage provides an initial framework for prioritizing patients according to the urgency of their condition. Established systems, including the Emergency Severity Index (ESI), Manchester Triage System (MTS), and Australasian Triage Scale (ATS), help standardize this process and support decisions about the order of clinical assessment. However, triage is not a definitive diagnosis, and the patient's condition may change after arrival. Some patients who initially appear stable may subsequently develop serious complications, whereas others may be assigned a high level of urgency despite having a relatively low probability of an adverse outcome. These limitations highlight the need for approaches that can support early recognition of clinical risk without replacing professional judgment [3].

AI, particularly machine learning (ML) and natural language processing (NLP), has emerged as a potential means of improving risk assessment in emergency care. These methods can analyze combinations of demographic characteristics, vital signs, presenting complaints, medical history, and laboratory findings to identify patterns associated with adverse outcomes. Depending on their design, AI models may estimate the risk of short-term mortality, clinical deterioration, critical care admission or hospital admission. Some systems use structured information available at arrival, while others incorporate clinical text or update their predictions as new information becomes available. This flexibility creates opportunities to support clinical assessment at different stages of an ED visit, but it also raises important questions about when predictions are generated and whether the necessary information is available at the time of decision-making; [4].

The potential value of AI should not be judged solely by its ability to predict outcomes accurately. A model may perform well in a retrospective dataset but provide limited benefit when introduced into routine clinical practice. Its predictions may overlap with information already recognized by clinicians, depend on data that are unavailable during initial triage or reflect local admission practices rather than underlying clinical need. Differences in patient populations, healthcare resources and clinical workflows may also reduce performance when a model is transferred to another institution. Furthermore, an improvement in statistical discrimination does not necessarily translate into earlier treatment, fewer missed cases of deterioration, lower mortality or more appropriate use of hospital resources. Demonstrating clinical

usefulness therefore requires evaluation beyond model development and internal validation [5].

The distinction between retrospective performance and prospective clinical impact is particularly important in this field. Retrospective studies can establish whether a model identifies patterns associated with a defined outcome in existing data. External validation can assess whether performance is maintained in different populations or settings [6]. Prospective evaluations provide further information about how a model performs using data collected during real clinical care, while implementation studies and randomized trials can examine whether making predictions available to clinicians' changes decisions or patient outcomes. Recent work, including the MARS ED randomized trial of machine learning-based risk stratification and prospective evaluation of large language model clinical decision support in emergency care, provides an opportunity to examine this progression from prediction to clinical application. Nevertheless, the findings of individual studies must be interpreted in relation to their study design, patient population, intended use, and measured outcomes [7].

Patient safety is another essential consideration. Incorrect risk estimates may contribute to under-triage, unnecessary escalation, delayed reassessment, or inappropriate reassurance. Clinicians may also place excessive confidence in algorithmic recommendations, particularly when a system presents its output in fluent language or provides an apparently convincing explanation [8]. These concerns are especially relevant to large language models, which can generate plausible clinical recommendations but may also produce unsupported or inconsistent information. Safe implementation therefore requires clear responsibility for clinical decisions, appropriate human oversight, transparent escalation arrangements and ongoing assessment of model performance. Fairness must also be considered because differences in age, sex, language, disability, ethnicity, comorbidity and access to healthcare may affect the quality and consequences of AI-supported decisions [9].

Implementation challenges extend beyond the technical performance of a model. An AI system must fit the pace and demands of emergency care, integrate with existing information systems and provide useful information without increasing documentation burden or alert fatigue [10]. Its performance may change over time as patient characteristics, clinical practices and resource availability evolve. Continuous monitoring is therefore important, particularly when a model is used across different institutions or updated after

deployment. Evaluation should consider not only predictive performance but also clinician workload, patient experience, equity, resource use and unintended consequences [11].

Against this background, this critical narrative review examines the use of AI to support triage and early risk stratification among adults presenting to emergency departments with undifferentiated acute illness. It focuses on the clinical tasks addressed by these systems, their data inputs and prediction timing, the methods used to evaluate their performance, and the strength of evidence supporting their use in practice. Particular attention is given to prospective evaluation, clinical impact, patient safety, equity and implementation. Prediction of deterioration, critical care need, short-term mortality, and hospital admission are considered separately because these outcomes represent different clinical questions and should not be treated as interchangeable measures of success. The review aims to help emergency physicians, nurses, researchers and healthcare leaders distinguish promising predictive models from tools that have demonstrated meaningful clinical value, while identifying the evidence needed to support their safe and responsible integration into emergency care [12].

Clinical Decisions and Conventional Comparators

Emergency department triage involves prioritizing patients according to the urgency of their clinical condition, while early risk stratification estimates the likelihood of adverse outcomes and supports decisions about further assessment and treatment. Although these processes are closely related, they address different clinical questions [13]. Conventional triage systems, early warning scores and clinician assessment provide important comparators for evaluating AI. Understanding their roles is essential for determining whether AI offers meaningful benefits beyond existing clinical practice [14].

Triage, Risk Stratification and Disposition

Triage determines how quickly a patient should receive clinical assessment, based on the presenting complaint, initial observations and suspected severity of illness. Systems such as the Emergency Severity Index (ESI), Manchester Triage System (MTS) and Australasian Triage Scale (ATS) provide structured approaches to urgency classification. However, patients with similar triage categories may have

different clinical trajectories, and some initially stable patients may deteriorate during their ED stay [15].

Risk stratification addresses this uncertainty by estimating the probability of a defined outcome within a specified period. Depending on the intended application, this may include clinical deterioration, critical care admission or short-term mortality. In contrast, disposition planning concerns whether a patient should be discharged, admitted, observed or transferred to a higher level of care. These outcomes should not be treated as interchangeable. Hospital admission, for example, depends not only on clinical severity but also on local policies, bed availability and clinician judgment. Consequently, predicting admission does not necessarily establish that a model can identify patients who require urgent intervention [16].

Conventional Scores and Clinical Judgment

Early warning scores, including the National Early Warning Score 2 (NEWS2), combine physiological measurements to identify patients who may be deteriorating. Such tools provide a consistent framework for recognizing abnormal observations and supporting escalation decisions [17]. Nevertheless, physiological scores may not capture every aspect of a patient's condition, particularly when serious illness presents with subtle findings or atypical symptoms.

Clinical judgment complements these structured approaches by incorporating symptom progression, medical history, general appearance and other contextual information. However, human assessment can be affected by workload, fatigue, incomplete information and differences in experience. Conventional systems also have limitations, but they remain essential reference points because they are embedded in established clinical workflows [18].

AI models should therefore be evaluated against appropriate existing practice rather than against an unrealistic assumption that conventional assessment is consistently accurate or entirely objective. Comparisons should consider whether the model adds useful information to the triage category, physiological scores and clinician assessment already available at the relevant decision point [19].

Defining the Intended Clinical Benefit

The potential value of AI depends on the decision it is designed to support. A model that identifies patients at increased risk of deterioration may help clinician's priorities reassessment, whereas a model estimating short term mortality may support broader risk review.

Neither output should automatically determine treatment or disposition without clinical interpretation [20]. Importantly, improved predictive performance does not guarantee better patient outcomes. An AI model may identify high-risk patients more accurately without changing management because clinicians already recognize those patients or because appropriate interventions are unavailable. Conversely, a model that identifies a previously overlooked high-risk group may offer practical value even if its overall improvement in predictive accuracy is modest [21].

The MARS ED randomized trial provides a relevant example of evaluating machine learning-based risk stratification in clinical practice. Its importance lies in examining AI-supported risk assessment within a care pathway rather than relying exclusively on retrospective prediction metrics. The findings must nevertheless be interpreted according to the study population, intervention and outcomes measured [22].

Table 1. Clinical Tasks, Data Inputs, Timing, Comparators, and Outcomes for AI-Based Emergency Department Risk Stratification

Clinical task	Typical data inputs	Prediction timing	Conventional comparator	Relevant outcomes
Initial triage and urgency assignment	Age, presenting complaint, vital signs, symptoms and brief clinical notes	At arrival or initial assessment	Established triage systems and nurse assessment	Under triage, time to clinical assessment
Early deterioration prediction	Vital signs, oxygen saturation, mental status, comorbidities and presenting symptoms	At arrival and during observation	Early warning scores and clinician assessment	Subsequent deterioration, emergency intervention, unplanned intensive care
Critical care requirement	Initial clinical observations, comorbidities and early laboratory results when available	Initial assessment or early ED stay	Conventional risk scores and clinician judgment	ICU admission, urgent organ support and escalation of care
Short-term mortality prediction	Demographics, vital signs, comorbidities, laboratory results and clinical text	At arrival or following early investigations	Validated prognostic scores and clinician assessment	Defined short-term mortality, such as 7-day or 30-day mortality
Hospital admission risk	Symptoms, vital signs, medical history, initial investigations and operational variables	During initial assessment or before disposition	Clinician assessment and conventional admission prediction models	Hospital admission, discharge, return visits and subsequent adverse events
Dynamic risk reassessment	Repeated vital signs, evolving symptoms, laboratory results and treatment response	Repeatedly during ED observation	Serial clinical assessment and updated early warning scores	Changes in risk classification, escalation, delayed recognition of deterioration
Clinical-text-assisted risk assessment	Presenting complaint, triage notes, history of present illness and relevant clinical documentation	At triage or after additional history is documented	Manual documentation review and conventional NLP-based tools	Risk classification, missed high-risk presentations and decision consistency

Data Inputs and Prediction Timing

AI models used for emergency department triage and risk stratification rely on patient information available at different stages of care. The type, quality and timing of these data influence model performance and determine whether its predictions can support real clinical decisions [23].

Sources of Patient Data

Common inputs include age, sex, presenting complaints, medical history, comorbidities and initial vital signs, such as heart rate, blood pressure, respiratory rate, oxygen saturation and temperature. These variables provide information about the patient's baseline characteristics and current physiological condition. Some models also incorporate laboratory results, medication history and previous healthcare records to improve risk estimation [24].

Clinical notes represent another important data source. Natural language processing (NLP) can extract relevant information from triage notes and early clinical assessments, including symptoms that may not be adequately represented by structured fields. However, clinical documentation varies between professionals and institutions. Incomplete notes, inconsistent terminology and language differences can introduce errors or reduce the reliability of predictions [25].

Arrival Based and Dynamic Predictions

Arrival based models use information available during initial triage. Their main advantage is that they can provide early risk estimates before diagnostic investigations are completed. This may help clinicians identify patients who require closer observation or more urgent assessment [26].

Dynamic models update predictions as new information becomes available during the ED visit. For example, repeat vital signs, laboratory findings and responses to initial treatment may reveal changes that were not apparent at arrival. Although this approach may improve risk assessment, its usefulness depends on whether updated predictions arrive early enough to influence clinical decisions [27].

The distinction between these approaches is important when comparing model performance. A model using laboratory results collected several hours after arrival should not be directly compared with an arrival only model without considering the difference in available information and prediction timing [28].

Missing Data and Prediction Timing

Missing values are common in emergency care because investigations are ordered according to clinical need rather than collected uniformly for every patient. Some models may interpret missing information incorrectly, while others may produce unreliable estimates when applied to patients whose characteristics differ from the development dataset [29].

Data leakage is another concern. It occurs when a model uses information that would not have been available at the intended prediction time, potentially producing misleadingly strong results. For example, using later clinical observations to predict deterioration at arrival would not represent a valid early warning application [30].

Therefore, studies should clearly define when predictions are generated, which variables are available at that moment and how missing information is handled. These details are essential for assessing whether an AI model can provide timely, reliable support within routine emergency department workflows [31].

Model Types and Outputs

AI systems used in ED triage differ in their modelling methods, data requirements and clinical outputs. Understanding these differences is important because model complexity alone does not determine clinical usefulness. The suitability of a system depends on its intended purpose, the information available and how its predictions are presented to healthcare professionals [10].

Statistical Models and Machine Learning

Conventional statistical models, including logistic regression, estimate the relationship between patient characteristics and specific clinical outcomes. They can provide interpretable risk estimates and serve as useful comparators for more complex methods. Machine learning (ML) approaches, such as random forests and gradient boosting, can identify complex relationships among multiple variables. These models may improve prediction when clinical information contains nonlinear patterns or interactions that are difficult to represent using simpler methods [20].

However, greater complexity does not automatically produce better performance. ML models may be sensitive to differences in data quality, patient populations and clinical practice. Their predictions must therefore be compared with appropriate

conventional models and validated in settings beyond the original development dataset [32].

Natural Language Processing and Large Language Models

Natural language processing (NLP) enables AI systems to extract clinically relevant information from unstructured text, including presenting complaints and triage notes. This may help identify symptoms or contextual details that are not captured in structured electronic records [33].

Large language models (LLMs) extend these capabilities by processing clinical text and generating responses in natural language. In principle, they may support early risk assessment by organizing patient information or presenting decision support suggestions. Nevertheless, fluent language does not guarantee clinical accuracy. LLMs may generate unsupported statements, overlook important findings or produce inconsistent recommendations when given similar information. Their outputs require careful evaluation against clinically appropriate standards [34].

Risk Scores, Dynamic Predictions and Clinical Recommendations

AI systems may produce a fixed risk score, an updated probability of an adverse outcome or a recommendation for further clinical review. Fixed scores are relatively straightforward to display and interpret, whereas dynamic models can reflect changes in a patient's condition over time. Conversational systems may offer additional explanations, but these explanations should not be assumed to represent the actual reasoning behind a prediction [13].

The intended user must also be clearly defined. A triage nurse may need an immediate urgency alert, while an emergency physician may require a more detailed estimate of deterioration risk. Each output should specify the predicted outcome, relevant time horizon and threshold for clinical action [35].

Ultimately, AI models should be assessed according to their reliability, interpretability and usefulness for the intended decision. A well performing score or convincing language-based recommendation has limited value if it arrives too late, duplicates existing information or encourages inappropriate clinical action. Their role should remain supportive, with final decisions guided by clinical assessment and patient circumstances [36].

Validation and Diagnostic or Prognostic Performance

Evaluating AI models for ED triage requires more than measuring their ability to distinguish between patients who experience an adverse outcome and those who do not. Reliable evaluation must consider calibration, clinically relevant thresholds, external validity and the quality of the underlying data. These factors help determine whether a model can provide useful risk estimates in routine emergency care [3].

Discrimination and Calibration

Discrimination describes how effectively a model separates patients with an outcome from those without it. The area under the receiver operating characteristic curve (AUROC) is commonly used for this purpose. However, a high AUROC does not establish that a model provides accurate individual risk estimates or improves clinical decisions [37].

Calibration assesses the agreement between predicted risks and observed outcomes. For example, among patients assigned a predicted mortality risk of 10%, approximately 10% should experience the specified outcome if the model is well calibrated in that population. Poor calibration can lead to inappropriate reassurance or unnecessary escalation, even when discrimination remains acceptable [5].

Sensitivity, specificity, positive predictive value and negative predictive value should also be reported at clinically relevant thresholds. These measures help clarify how many high-risk patients are identified and how many patients may be incorrectly classified. Because adverse outcomes are often uncommon, predictive values may change substantially across different ED populations [38].

External and Temporal Validation

Models developed using data from one hospital may perform differently in another setting because of differences in patient characteristics, disease prevalence, documentation practices and available resources. External validation evaluates performance in an independent population, while temporal validation examines whether performance is maintained using data collected during a later period [39].

Prospective validation provides additional evidence by testing a model on patients encountered during real clinical care. Nevertheless, successful validation does not necessarily demonstrate clinical benefit. It establishes whether predictions remain reliable under

the conditions tested, not whether clinicians acting on those predictions improve patient outcomes [40].

Bias, Data Leakage and Outcome Interpretation

Data leakage occurs when a model uses information that would not have been available at the intended prediction time. This can produce misleadingly strong performance estimates. Studies should therefore specify when predictions are generated and ensure that input variables reflect information available at that point [41].

The treatment paradox also requires consideration. High risk patients may receive earlier treatment because clinicians recognize their condition, potentially reducing the adverse outcomes that a model is designed to predict. Consequently, observed outcomes do not always reflect untreated biological risk [42].

Finally, outcomes such as hospital admission and critical care use depend partly on institutional policies, bed availability and clinical practice. Researchers should distinguish these service dependent outcomes from measures of physiological deterioration or mortality [43].

Prospective Evidence and Clinical Impact

The clinical value of AI in ED triage depends on evidence that extends beyond retrospective prediction. Although retrospective studies can identify associations between patient characteristics and adverse outcomes, they cannot establish whether providing AI predictions to clinicians improves care. Evaluating the progression from model development

to prospective implementation is therefore essential for determining the practical value of these technologies [12].

From Retrospective Studies to Prospective Evaluation

Retrospective studies commonly use existing electronic health records to develop and test prediction models. They are useful for identifying promising approaches, but their findings may be affected by missing data, selection bias and differences between recorded information and real time clinical availability. External validation can establish whether performance is maintained in another population, although it still does not demonstrate that clinicians can use the model effectively [44].

Silent prospective evaluation provides a further step by testing predictions on newly encountered patients without displaying the results to clinical staff. This approach allows researchers to assess real world data quality, prediction timing and calibration without directly influencing patient management. However, because clinicians do not act on the predictions, silent evaluation cannot establish whether the system improves clinical decisions or outcomes [45].

Clinician visible implementation studies address this limitation by introducing AI outputs into actual workflows. They can examine whether clinicians respond to alerts, change their assessment or modify management plans. Nevertheless, before and after comparisons may be influenced by staffing changes, seasonal variation, patient characteristics and other developments occurring during the study period [46].

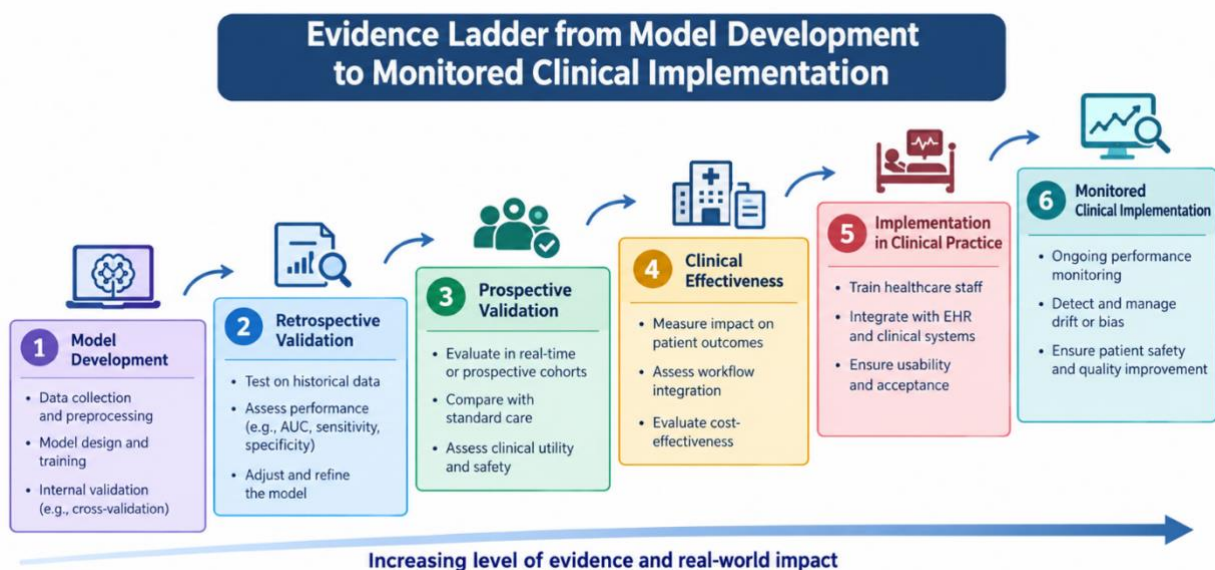


Figure 1. Evidence ladders from model development to monitored clinical implementation.

Patient Safety and Human Oversight

Patient safety is a central consideration when introducing AI into ED triage and risk stratification. Although AI models may help identify patients at increased risk, inaccurate predictions or inappropriate clinical responses can cause harm. Their implementation must therefore include clear human oversight, appropriate escalation procedures and continuous assessment of potential risks [47].

Under triage and Automation Bias

Under triage occurs when a patient is assigned a lower level of urgency than their condition requires. If an AI model underestimates the risk of serious illness, clinicians may delay reassessment or treatment, potentially worsening patient outcomes. Over-triage occurs when patients are classified as higher risk than necessary, which may lead to unnecessary investigations, inappropriate escalation, and increased pressure on limited emergency resources. The consequences depend on the clinical context and the decisions influenced by the prediction [23].

Automation bias presents another concern. Clinicians may place excessive trust in AI-generated recommendations, particularly when the system appears confident or provides a convincing explanation. This may reduce attention to important clinical findings that conflict with the model's output. Conversely, repeated inaccurate alerts may cause alert fatigue, making staff less likely to respond to subsequent warnings. Effective implementation should encourage clinicians to consider AI predictions alongside their own assessment rather than treating them as definitive instructions [48].

Language Model Errors and Clinical Accountability

Large language models introduce additional safety concerns because they can generate plausible but incorrect clinical information. They may overlook important symptoms, misinterpret clinical notes or provide recommendations that are not supported by the available evidence. A fluent explanation should therefore not be considered proof of accuracy. Systems used in emergency care require evaluation using representative clinical cases and clearly defined safety criteria [49].

Responsibility for decisions must also remain clear. Healthcare professionals should understand the intended use and limitations of the tool, while institutions should establish who is responsible for

reviewing performance, investigating incidents and responding to unsafe outputs. Clinicians must have the ability to question or override recommendations when patient findings or clinical judgment indicate a different course of action [50].

Downtime, Escalation and Incident Review

AI systems may become unavailable because of technical failures, connectivity problems or electronic health record disruptions. Emergency services must therefore be able to continue using established triage procedures without depending on AI availability. Patient care should not be delayed while staff attempt to restore a system or resolve conflicting outputs [51].

Adverse events and near misses associated with AI-supported decisions should be documented and reviewed alongside other clinical safety incidents. Reviews should examine the model's output, the information available at the time, clinician responses and relevant workflow conditions. This approach can help identify recurring problems and guide corrective action [52]. Safe AI implementation requires human oversight, clear accountability, reliable fallback arrangements and continuous safety monitoring. These safeguards are essential for ensuring that AI supports clinical judgment without creating new risks for emergency patients [53].

Equity, Workflow and Monitoring After Deployment

Implementing AI in emergency departments (EDs) must account for differences in patient populations, clinical resources, and working conditions. A model that performs well in one hospital may produce less reliable predictions in another setting. Assessing equity, workflow compatibility and performance after deployment is therefore essential to ensure that AI-supported decisions remain appropriate across different clinical environments [54].

Equity and Differences Between Patient Groups

AI models are developed using historical data that may not adequately represent every patient group. Differences in age, sex, ethnicity, disability, language and multimorbidity may influence prediction accuracy. For example, a model trained mainly on patients from one demographic group may underestimate risk in populations that were poorly represented during development. Language barriers and differences in clinical documentation may create additional

challenges when models use presenting complaints or unstructured notes [55].

Researchers should evaluate model performance across relevant subgroups using measures such as sensitivity, calibration and false negative rates. These assessments can help identify groups in which patients at risk of deterioration are more likely to be missed. However, subgroup findings must be interpreted carefully when sample sizes are small or demographic information is incomplete. Local validation is particularly important before applying a model to populations that differ substantially from the original development cohort [56].

Workflow, Staffing and Resource Availability

AI systems must fit the practical demands of emergency care. Clinicians work under time pressure,

and additional alerts or documentation requirements may increase workload rather than improve decision making. Integration with electronic health records, rapid display of risk estimates and clear presentation of actionable information may reduce these difficulties. The timing of outputs is equally important because a prediction delivered after the relevant clinical decision has been made may offer little benefit [57].

Crowding, staffing levels, laboratory turnaround times and access to inpatient beds can also influence implementation. A model may identify high risk patients accurately, but its impact will be limited if appropriate reassessment or treatment cannot be provided. Evaluation should therefore examine whether AI recommendations are feasible within the resources available at the intended institution [58].

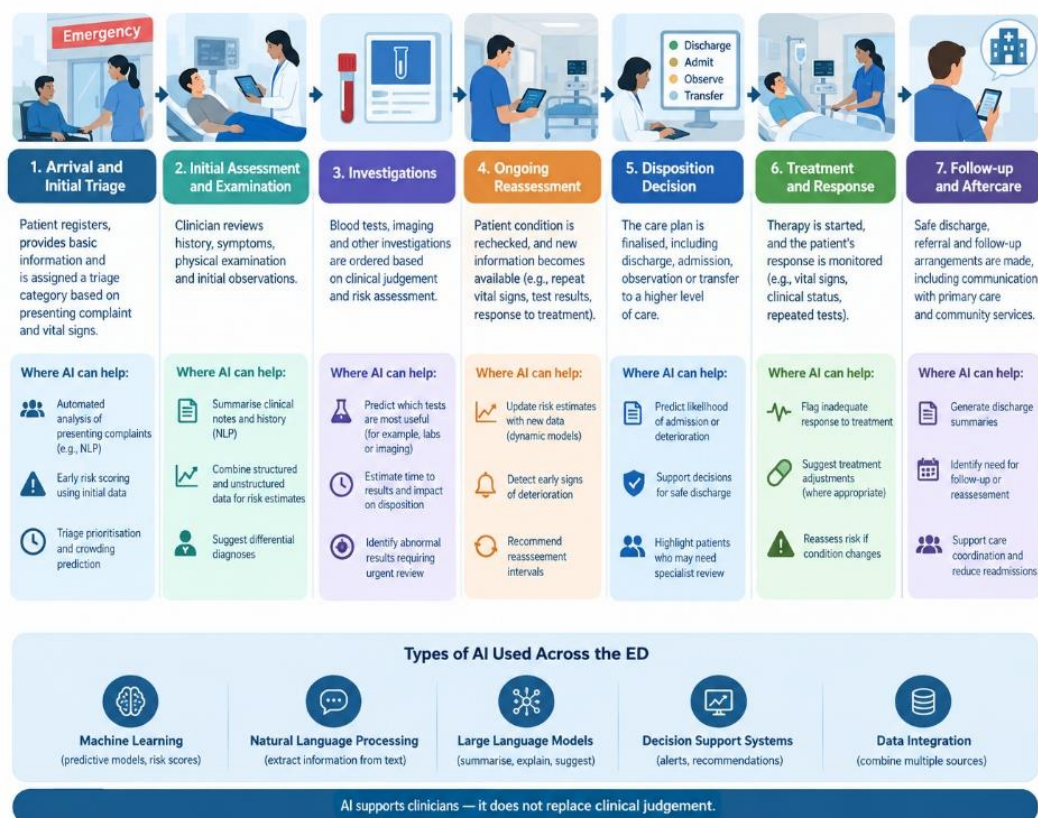


Figure 2. Emergency care workflow.

Monitoring After Deployment

Model performance may change over time as patient characteristics, disease patterns, clinical practices and data systems evolve. This phenomenon, often described as model drift, can reduce the reliability of predictions even when the model performed well during initial validation. Institutions should establish

ongoing monitoring procedures that assess calibration, discrimination, subgroup performance and clinically important errors [59].

Audit trails should record relevant model outputs, system versions and clinician responses where appropriate. Changes to algorithms should undergo suitable testing before implementation, particularly when updates may alter risk estimates or

recommendations. Responsibility for monitoring should be assigned clearly across clinical teams, technical staff and institutional leadership. Regular review of adverse events, missed deterioration and unnecessary escalation can help identify problems before they become widespread [60].

Research Priorities

Future research should establish whether AI improves ED triage and early risk stratification compared with usual clinical practice. Multicenter prospective studies are needed to assess performance across different hospitals, patient populations and healthcare systems. These studies should clearly define the intended clinical decision, prediction time and outcome rather than combining distinct endpoints such as short-term mortality, critical care admission and hospital admission [61].

Randomized controlled trials should examine whether providing AI predictions to clinicians leads to measurable improvements in patient care. Relevant outcomes include missed deterioration, time to appropriate treatment, avoidable complications, mortality and patient experience. Researchers should also measure unintended consequences, including unnecessary investigations, alert fatigue and additional workload. Where randomized trials are not feasible, carefully designed prospective implementation studies may provide useful evidence, provided that potential sources of bias are addressed [62].

Consistent reporting is another priority. Studies should describe the training population, missing data handling, validation methods, model calibration and thresholds used to support clinical decisions. Subgroup analyses should assess whether performance differs across relevant demographic and clinical groups. Economic evaluations may further establish whether the benefits of implementation justify the costs of software integration, staff training, maintenance and monitoring [63].

Research should also investigate how clinicians interact with AI recommendations. Understanding when clinicians accept, question or override predictions may help identify factors that influence safety and effectiveness. Large language models require particular attention to factual accuracy, consistency and the reliability of their recommendations under realistic clinical conditions. Independent evaluation is important because promising results from a single institution may not be reproducible elsewhere [64].

Conclusions

AI offers potential opportunities to support triage and early risk stratification in adult emergency departments by combining patient information and identifying patterns associated with adverse outcomes. Machine learning, natural language processing and large language models may assist clinicians in recognizing patients who require closer assessment. However, their usefulness depends on the intended clinical task, the quality and timing of available data, and the reliability of predictions in the population where they are applied. Evidence from retrospective studies and external validation can establish predictive potential, but it cannot independently demonstrate improved clinical care. Prospective evaluations and randomized trials are needed to determine whether AI supported decisions improve patient outcomes without introducing unacceptable risks. Particular attention should be given to under triage, automation bias, unequal performance and the practical demands of emergency workflows. The adoption of AI should therefore be guided by evidence appropriate to each intended use rather than by technological capability alone. Systems that demonstrate reliable performance and meaningful clinical benefit may support selected decisions when implemented with human oversight, clear accountability and continuous monitoring. The evidence should determine the extent of their role in emergency care, with established clinical assessment remaining central to decisions about patient urgency and treatment.

Author Contributions: Conceptualization, S.M.A.; methodology, S.M.A.; validation, Z.S.; resources, Z.S.; writing—original draft preparation, S.M.A.; writing—review and editing, Z.S.; supervision, S.M.A.; project administration, S.M.A. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This review article is based on information obtained from previously published literature and publicly available sources. No new experimental or primary datasets were generated or analyzed in this study. All relevant information and sources are appropriately cited within the manuscript and reference list. The authors are responsible for the accuracy and integrity of the information presented.

Acknowledgments: The authors gratefully acknowledge the authors and researchers whose published scientific contributions provided the foundation for this review. The authors also acknowledge the scientific databases and publicly accessible literature resources used during the preparation of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tyler, S., Olis, M., et al. Use of artificial intelligence in triage in hospital emergency departments: a scoping review. *Cureus*. **2024**. 16(5).
2. Taylor, R.A., Chmura, C., et al. Impact of artificial intelligence-based triage decision support on emergency department care. *NEJM AI*. **2025**. 2(3): A0a2400296.
3. El Arab, R.A. and Al Moosa, O.A. The role of AI in emergency department triage: an integrative systematic review. *Intensive and Critical Care Nursing*. **2025**. 89: 104058.
4. Lee, J.-T., Hsieh, C.-C., et al. Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department. *Scientific reports*. **2021**. 11(1): 19472.
5. Miles, J., Turner, J., et al. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. *Diagnostic and prognostic research*. **2020**. 4(1): 16.
6. Chen, Y.-H.J., Lin, C.-S., et al. An AI-enabled dynamic risk stratification for emergency department patients with ECG and CXR integration. *Journal of medical systems*. **2023**. 47(1): 81.
7. Sitthiprawiat, P., Wittayachamnankul, B., et al. Development and internal validation of an AI-based emergency triage model for predicting critical outcomes in emergency department. *Scientific Reports*. **2025**. 15(1): 31212.
8. Falavigna, G., Costantino, G., et al. Artificial neural networks and risk stratification in emergency departments. *Internal and Emergency Medicine*. **2019**. 14(2): 291–299.
9. Alimiri Dehbaghi, H. and Khoshgard, K. Revolutionizing emergency care: an overview of the transformative role of artificial intelligence in diagnosis, triage, and patient management. *International Journal of Emergency Medicine*. **2025**. 18(1): 242.
10. Almulihi, Q.A., Alquraini, A.A., et al. Applications of artificial intelligence and machine learning in emergency medicine triage-a systematic review. *Medical Archives*. **2024**. 78(3): 198.
11. Chen, J.-Y., Hsieh, C.-C., et al. Patient stratification based on the risk of severe illness in emergency departments through collaborative machine learning models. *The American Journal of Emergency Medicine*. **2024**. 82: 142–152.
12. Da'Costa, A., Teke, J., et al. AI-driven triage in emergency departments: A review of benefits, challenges, and future directions. *International Journal of Medical Informatics*. **2025**. 197: 105838.
13. Abualruz, H., Yasin, I., et al. The role of artificial intelligence in enhancing triage decisions in healthcare settings: a systematic review. *Applied Nursing Research*. **2025**: 152024.
14. Van Dam, P.M., Van Doorn, W.P., et al. Machine learning for risk stratification in the emergency department (MARS-ED): a randomized controlled trial. *Nature Communications*. **2025**. 17(1): 242.
15. Hosseini, M.M., Hosseini, S.T.M., et al. The aspects of running artificial intelligence in emergency care; a scoping review. *Archives of Academic Emergency Medicine*. **2023**. 11(1): e38.
16. Stewart, J., Lu, J., et al. Applications of natural language processing at emergency department triage: A narrative review. *Plos one*. **2023**. 18(12): e0279953.
17. Nord-Bronzyk, A., Savulescu, J., et al. Assessing risk in implementing new artificial intelligence triage tools—how much risk is reasonable in an already risky world? *Asian bioethics review*. **2025**. 17(1): 187–205.
18. Bhuvaneshwari, E., Prasad, K., et al. A human-centered hybrid AI framework for optimizing emergency triage in resource-constrained settings. *Intelligence-Based Medicine*. **2025**. 12: 100311.
19. Chua, M.T., Boon, Y., et al. The role of artificial intelligence in sepsis in the Emergency Department: a narrative review. *Annals of Translational Medicine*. **2025**. 13(1): 4.
20. Raita, Y., Goto, T., et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care*. **2019**. 23(1): 64.
21. Mueller, B., Kinoshita, T., et al. Artificial intelligence and machine learning in emergency medicine: a narrative review. *Acute medicine & surgery*. **2022**. 9(1): e740.
22. Tana, C., Soloperto, M., et al. Artificial intelligence for predicting lung immune responses to viral infections: from mechanistic insights to clinical applications. *Viruses*. **2025**. 17(11): 1482.
23. Porto, B.M. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. *BMC Emergency Medicine*. **2024**. 24(1): 219.
24. Akshatha, H. Predictive AI Models for Emergency Room Triage.
25. El Khatib, A., Automating Patient Triage and Disposition Prediction Using Machine Learning for Decision Support in Emergency Departments and Home Care Settings. **2025**, Université Jean Monnet (EPSCPE).
26. Hodgson, N.R., Saghafian, S., et al. Artificial intelligence-assisted emergency department vertical patient flow optimization. *Journal of*

- Personalized Medicine. **2025**. 15(6): 219.
27. Moreno-Sánchez, P.A., Aalto, M., and van Gils, M. Prediction of patient flow in the emergency department using explainable artificial intelligence. *Digital health*. **2024**. 10: 20552076241264194.
28. Araz, O.M., Olson, D., and Ramirez-Nafarrate, A. Predictive analytics for hospital admissions from the emergency department using triage information. *International Journal of Production Economics*. **2019**. 208: 199–207.
29. Kim, D., You, S., et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. *PloS one*. **2018**. 13(10): e0206006.
30. Yi, N., Baik, D., and Baek, G. The effects of applying artificial intelligence to triage in the emergency department: A systematic review of prospective studies. *Journal of Nursing Scholarship*. **2025**. 57(1): 105–118.
31. Waligora, G., Sherwin, R., and Soucy, Z. Can Application of Artificial Intelligence Improve Emergency Department Triage Performance? *Journal of Emergency Medicine*. **2025**. 78: 351–370.
32. Sánchez-Salmerón, R., Gómez-Urquiza, J.L., et al. Machine learning methods applied to triage in emergency services: A systematic review. *International Emergency Nursing*. **2022**. 60: 101109.
33. Williams, C.Y., Zack, T., et al. Use of a large language model to assess clinical acuity of adults in the emergency department. *JAMA network open*. **2024**. 7(5): e248895.
34. Masanneck, L., Schmidt, L., et al. Triage performance across large language models, ChatGPT, and untrained doctors in emergency medicine: comparative study. *Journal of medical Internet research*. **2024**. 26: e53297.
35. Fatai, A., Sattayarom, C., et al. Effectiveness of AI-assisted ESI triage on accuracy and selected outcomes in emergency nursing: a systematic review. *International emergency nursing*. **2025**. 83: 101680.
36. Menshawi, A.M. and Hassan, M.M. A novel triage framework for emergency department based on machine learning paradigm. *Expert Systems*. **2025**. 42(2): e13735.
37. Levin, S., Toerper, M., et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Annals of emergency medicine*. **2018**. 71(5): 565–574. e2.
38. Alomari, L.M., Alshammari, M.M., et al. Safety and accuracy of AI in triaging patients in the emergency department. *International journal of emergency medicine*. **2025**. 18(1): 243.
39. Soltan, A.A., Yang, J., et al. Real-world evaluation of rapid and laboratory-free COVID-19 triage for emergency care: external validation and pilot deployment of artificial intelligence driven screening. *The Lancet Digital Health*. **2022**. 4(4): e266–e278.
40. Brandao-de-Resende, C., Melo, M., et al. A machine learning system to optimise triage in an adult ophthalmic emergency department: a model development and validation study. *EClinicalMedicine*. **2023**. 66.
41. Kachman, M.M., Brennan, I., et al. How artificial intelligence could transform emergency care. *The American journal of emergency medicine*. **2024**. 81: 40–46.
42. Cha, H. and Kim, J. Ethical considerations of artificial intelligence in emergency medicine for triage and resource allocation: a scoping review. *Clinical and Experimental Emergency Medicine*. **2025**. 12(4): 306.
43. Feretzakis, G., Sakagianni, A., et al. Machine learning in medical triage: a predictive model for emergency department disposition. *Applied Sciences*. **2024**. 14(15): 6623.
44. Tortum, F. and Kasali, K. Exploring the potential of artificial intelligence models for triage in the emergency department. *Postgraduate Medicine*. **2024**. 136(8): 841–846.
45. Paslı, S., Şahin, A.S., et al. Assessing the precision of artificial intelligence in ED triage decisions: insights from a study with ChatGPT. *The American journal of emergency medicine*. **2024**. 78: 170–175.
46. Yao, L.-H., Leung, K.-C., et al. A novel deep learning-based system for triage in the emergency department using electronic medical records: retrospective cohort study. *Journal of Medical Internet Research*. **2021**. 23(12): e27008.
47. Amiot, F. and Potier, B. Artificial intelligence (AI) and emergency medicine: balancing opportunities and challenges. *JMIR Medical Informatics*. **2025**. 13(1): e70903.
48. Abdalhalim, A.Z.A., Ahmed, S.N.N., et al. Clinical impact of artificial intelligence-based triage systems in emergency departments: a systematic review. *Cureus*. **2025**. 17(6).
49. Chenais, G., Lagarde, E., and Gil-Jardiné, C. Artificial intelligence in emergency medicine: viewpoint of current applications and foreseeable opportunities and challenges. *Journal of medical Internet research*. **2023**. 25: e40031.
50. Preiksaitis, C., Ashenburg, N., et al. The role of large language models in transforming emergency medicine: scoping review. *JMIR medical informatics*. **2024**. 12: e53787.
51. Bajwa, A. AI-based emergency response systems: A systematic literature review on smart infrastructure safety. Available at SSRN 5171521. **2025**.

52. Peralta, A., Olivas, J.A., et al. Intelligent incident management leveraging artificial intelligence, knowledge engineering, and mathematical models in enterprise operations. *Mathematics*. **2025**. 13(7): 1055.
53. Richardson, N., Manikyala, A., et al. Emergency response planning: leveraging machine learning for real-time decision-making. *Emergency*. **2021**. 4: 14.
54. Hinson, J.S., Levin, S.R., et al. Enhancing emergency department triage equity with artificial intelligence: outcomes from a multisite implementation. *Annals of emergency medicine*. **2025**. 85(3): 288–290.
55. Teeple, S., Smith, A., et al. Exploring the impact of missingness on racial disparities in predictive performance of a machine learning model for emergency department triage. *JAMIA open*. **2023**. 6(4): ooad107.
56. Marques, S., Silva, P., and Andersson, H. Equity Considerations in Artificial Intelligence-Driven Triage Systems. *Bioethics*. **2025**. 3(1): 19–23.
57. Born, C., Wildmoser, J., et al. Identifying potentials for artificial intelligence-based process support along the emergency department care pathway to alleviate overcrowding. *Procedia Computer Science*. **2024**. 239: 1705–1712.
58. Boonstra, A. and Laven, M. Influence of artificial intelligence on the work design of emergency department clinicians a systematic literature review. *BMC health services research*. **2022**. 22(1): 669.
59. Tahernejad, A., Sahebi, A., et al. Application of artificial intelligence in triage in emergencies and disasters: a systematic review. *BMC Public Health*. **2024**. 24(1): 3203.
60. Wang, Y.-C., Chen, K.-W., et al. Implementation of an all-day artificial intelligence–based triage system to accelerate door-to-balloon times. in *Mayo Clinic Proceedings*. 2022. Elsevier.
61. Emre M, G., Fevzi M, E., and Efe, O. Beyond algorithm: emergency department professionals' perspectives on machine learning-based triage integration—a qualitative study. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*. **2025**. 62: 00469580251376921.
62. Aifuobhokhan, J., Agbarakwe, C.A., et al. Evaluating the Impact of Artificial Intelligence-Driven Triage Systems on Emergency Care Efficiency in Resource-Limited Settings: A Scoping Review. **2025**.
63. Goniewicz, K. Autonomous and Artificial Intelligence-Enabled Drone Systems for Medical Triage: A State-of-the-Art Review of Emerging Technologies and Swarm Applications. *Emergency Health Services Journal*. **2025**. 2(2): 26–37.
64. Picard, C., Kleib, M., et al. The use and structure of emergency nurses' triage narrative data: scoping review. *JMIR nursing*. **2023**. 6(1): e41331.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of RES Publishers and/or the editor(s). RES Publishers and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.