

# The Role of Artificial Intelligence in Enhancing Diagnostic and Triage Accuracy in Hospital Emergency Departments: A Systematic Review

Mehdi Shakoori<sup>1</sup>, Fatemeh Esmaili<sup>2</sup>, Ezzat Abbariki<sup>3</sup>, Esmail Heidaranlu<sup>4\*</sup> 

Received: 28 Dec 2025

Revised: 15 Mar 2026

Accepted: 30 May 2026

## Abstract

**Background:** Emergency departments face challenges of overcrowding and human error. This systematic review evaluates the effectiveness of artificial intelligence (AI) in improving triage and diagnostic accuracy in emergency settings.

**Methods:** This systematic review was conducted by searching PubMed, Scopus, Web of Science, and Google Scholar until September 2025. Inclusion criteria comprised original studies reporting at least one quantitative performance metric. Study quality was assessed using NOS and QUADAS-2. Data were synthesized narratively.

**Results:** AI models outperformed traditional triage systems (AUC often >0.90). In clinical interpretation of AUC, its limitations in imbalanced datasets should be considered, and metrics such as PPV, NPV, and AUPRC should also be reported. For instance, despite an acceptable AUC, ChatGPT 4.0 led to significant over-triage due to low specificity (18.2%). However, risks of algorithmic bias and challenges of integrating AI with real-world clinical workflows remain major implementation barriers. User resistance and the need for continuous ethical oversight were also identified.

**Conclusion:** AI holds potential to improve emergency triage, but current evidence is largely based on retrospective single-center studies. Clinical implementation should be gradual, starting in high-risk domains (e.g., sepsis screening), and supported by continuous ethical oversight.

**Keywords:** Artificial Intelligence, Triage, Diagnosis, Emergency Department

\*This work has been published under CC BY-NC-SA 4.0 license.



Copyright© Iran University of Medical Sciences

**Cite this article as:** Shakoori M, Esmaili F, Abbariki E, Heidaranlu Es. The Role of Artificial Intelligence in Enhancing Diagnostic and Triage Accuracy in Hospital Emergency Departments: A Systematic Review. *Med J Islam Repub Iran*. 2026 (30 May);40:55. <https://doi.org/10.47176/mjiri.40.55>

## Introduction

Hospital triage is a systematic process for prioritizing patients based on the severity of their clinical condition, especially in emergency situations where resources are limited (1). The term originates from the French word trier, meaning 'to separate and sort,' with the aim of ensuring that critical patients receive rapid and timely care (2, 3). In the turbulent emergency environment, triage nurses play a key role in the initial assessment of patients, determining the urgency of their treatment needs, and optimally coordinating medical and nursing resources (4).

This process forms the basis for effective clinical decision-making and management of emergency patient flow, requiring both accuracy and fairness (5, 6).

Triage in the emergency unit plays an essential role in effective resource management, optimizing care quality, and increasing preparedness during emergencies and crises. By prioritizing patients based on clinical severity, this process enables the fair allocation of limited resources and ensures that critical patients are treated as soon as possible (7). Systematic implementation of triage not only reduces

**Corresponding author:** Dr Esmail Heidaranlu, [Heidaranlu@bmsu.ac.ir](mailto:Heidaranlu@bmsu.ac.ir)

<sup>1</sup>. Trauma and Injury Research Center, Iran University of Medical Sciences, Tehran, Iran

<sup>2</sup>. Radiation Injuries Research Center, Systems Biology and Poisonings Institute, Nursing Care research center, Clinical Sciences institute, Nursing Faculty, Baqiyatallah University of Medical Sciences, Tehran, Iran

<sup>3</sup>. Firoozgar Clinical Research and Development Center, Iran University of Medical Sciences, Tehran, Iran

<sup>4</sup>. Nursing Care Research Center, Clinical Sciences Institute, Nursing Faculty, Baqiyatallah University of Medical Sciences, Tehran, Iran

### Key Messages:

#### ↑What is “already known” in this topic:

Traditional triage has variable accuracy, and AI-based clinical decision support systems promise to act as a reliable and objective “intelligent assistant” alongside medical staff.

#### →What this article adds:

AI-driven triage systems have the potential to transform ED operations by enhancing efficiency, improving patient outcomes, and supporting healthcare professionals.

waiting times and enhances patient satisfaction but also leads to overall improvement in care quality and faster system response (8). Under normal and critical conditions, including disasters or surges in patient volume, triage increases the healthcare system's response capacity by coordinating medical and nursing teams (9). The development and implementation of standardized triage protocols also enhances consistency and reliability in emergency care. However, challenges such as the potential for classification errors and their ethical consequences highlight the importance of continuous training, evaluation, and improvement of triage systems (10).

In response to this growing crisis, Artificial Intelligence (AI) has emerged as a transformative catalyst. AI-based clinical decision support systems, with the ability to process vast amounts of clinical, laboratory, and imaging data in a fraction of a second, promise to act as a reliable and objective "intelligent assistant" alongside medical staff (11, 12). These technologies can enhance diagnostic accuracy through real-time data analysis (e.g., achieving 76.5% diagnostic accuracy with decision support systems), perform patient prioritization more accurately and automatically, and help optimize resource allocation by predicting critical outcomes (13, 14). Advanced models such as XGBoost, Deep Neural Networks (DNN), and Fuzzy Logic-based systems have demonstrated their capability in numerous studies to more accurately predict the need for intensive care or determine urgency levels compared to traditional tools (15-17). This review study aims to evaluate the effectiveness of Artificial Intelligence in improving diagnostic and triage accuracy in hospital Emergency Departments.

## Methods

This study was conducted as a systematic review following the PRISMA 2020 guideline. A comprehensive search was performed in PubMed, Scopus, Web of Science, Embase, IEEE Xplore, Google Scholar, SID, and Magiran databases without time restriction until September 30, 2025. Search keywords included combinations of: ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network") AND ("Emergency Department" OR "Emergency Room" OR "ED Triage") AND ("Triage" OR "Patient Prioritization" OR "Diagnostic Accuracy"), along with their Persian equivalents.

Inclusion criteria comprised original quantitative studies (cohort, cross-sectional, trials) in English or Persian that evaluated AI models for emergency department triage and reported at least one quantitative performance metric (e.g., AUC, sensitivity, specificity). Exclusion criteria included review articles, case reports, letters to the editor, simulation-only studies, and studies without real clinical data. Study quality was assessed using the Newcastle-Ottawa Scale (NOS) for observational studies, and risk of bias was evaluated using QUADAS-2 for diagnostic studies (18). This scale evaluates three domains: selection (0-4 points), comparability (0-2 points), and outcome (0-3 points), with a maximum total score of 9 points. Studies scoring 7-9 were classified as high quality, scores of 6 as moderate quality, and scores ≤5 as low quality. Two au-

thors independently screened the articles, with disagreements resolved by a third author.

To enhance the transparency and reproducibility of this systematic review and in accordance with the PRISMA-trAIce guideline (transparent reporting of artificial intelligence in evidence synthesis), the screening and data extraction process emphasized complete documentation of inclusion/exclusion criteria and the data synthesis method (19). Due to high methodological and outcome heterogeneity among the studies, data were synthesized using a narrative synthesis approach.

## Results

The study selection process was conducted in accordance with the PRISMA 2020 guideline (Figure 1). The initial search identified a total of 2,630 records from PubMed (n=847), Scopus (n=1,104), Web of Science (n=533), Embase (n=421), IEEE Xplore (n=315), and Google Scholar (n=146). After removing 689 duplicate records, 2,677 records remained for title and abstract screening. At this stage, 2,524 records were excluded for the following reasons: irrelevant to emergency triage, not using artificial intelligence, animal studies, and reviews without original data. Subsequently, 153 full-text reports were sought for retrieval, and all were accessed. Of these, 89 reports were excluded after full-text review: 41 due to lack of quantitative performance metrics, 28 due to simulation-only without real clinical data, 12 due to non-English/non-Persian language with incomplete translation, and 8 due to non-emergency population. Finally, 64 studies met the inclusion criteria, from which 12 relevant high-quality studies were selected for final analysis (Figure 1). Additionally, 2 methodological base studies (for quality assessment and transparency) were added to the review. The characteristics of included studies are presented in Table 1.

Based on the risk of bias assessment, among the 12 included studies, 8 were of high quality, and 4 were of moderate quality. No low-quality studies were included (Table 2).

## Current Status and Common Models

Analysis of 12 studies (20-31) showed that artificial intelligence (AI) is transitioning from the research phase to practical applications in leading emergency departments (Table 3). These systems primarily function as clinical decision support systems to complement human judgment.

- Ensemble/Gradient Boosting/XGBoost, Random Forest models: These demonstrated superior performance in predicting critical outcomes (e.g., need for intubation, transfusion, or ICU admission) with AUC often exceeding 0.90 (up to 0.962 in some studies), outperforming traditional triage tools (20, 21).

- Deep Neural Networks (DNN/CNN): These are used for processing complex multimodal data (combining text and structured data) and analyzing medical images (CT scans for stroke, chest radiographs). A study in Thailand showed that a DNN model (AUC=0.91) outperformed ESI and NEWS scores (22).

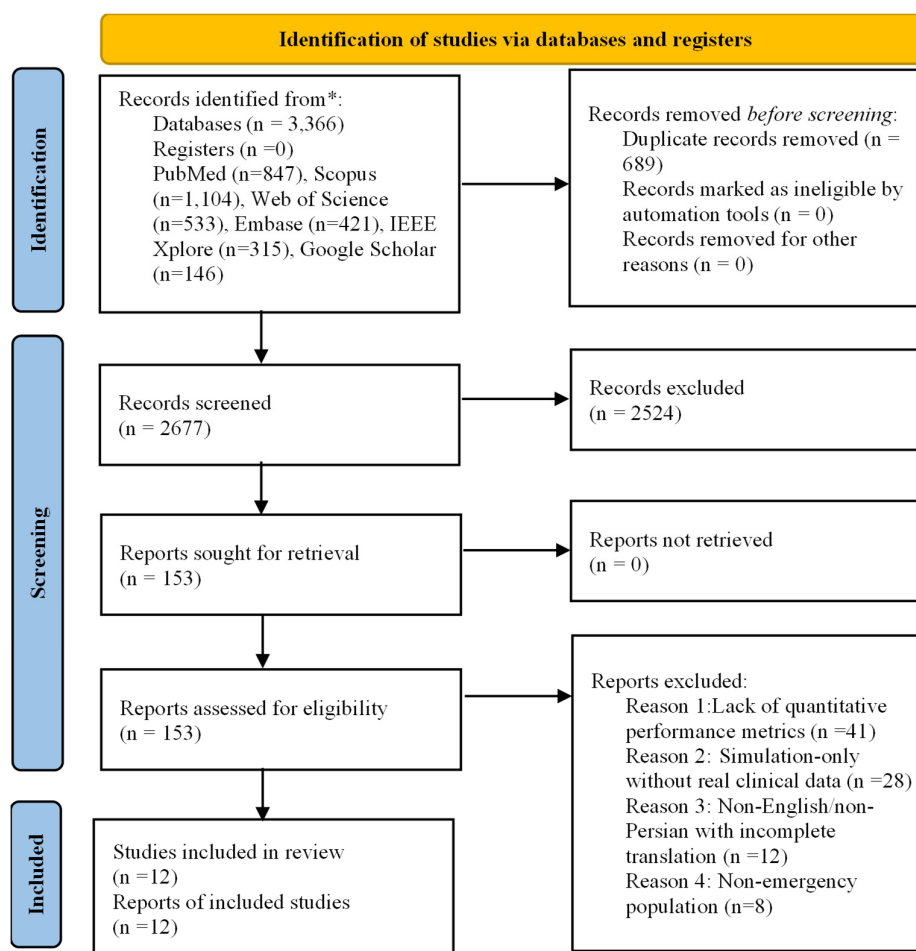


Figure 1. PRISMA 2020 Flow Diagram of the Study Selection Process

- **Fuzzy Logic:** Used in the i-TRIAGE system for modeling qualitative variables (e.g., pain), achieving 99% accuracy in predicting ESI triage levels (23). However, it should be noted that the 99% accuracy claim for i-TRIAGE is based on a single-center dataset of 616 patients and requires larger-scale, multicenter validation.

- **Computer Vision and Natural Language Processing (NLP/LLMs):** These are applied in automatic casualty identification at disaster scenes and in extracting information from clinical texts (24, 25).

#### Use of AI in Triage (Proven Benefits)

- **Improved Accuracy:** AI models consistently outperformed traditional triage systems (e.g., ESI) in predicting critical outcomes (mortality, ICU admission). The triage error rate decreased from 1.2% to 0.9% in one study (21).

- **Speed and Efficiency:** AI-based systems perform initial assessments in seconds. Voice-based AI systems accelerated documentation by 19% (25).

- **Resource Optimization:** AI-based radiology worklist prioritization reduced hospital length of stay by 11.9% for

intracranial hemorrhage and 26.3% for pulmonary embolism (25).

- **Early Diagnosis:** AI performance in identifying sepsis, in-hospital cardiac arrest, and stroke was superior to conventional clinical tools such as qSOFA (24).

#### Identified Disadvantages and Challenges (deepened based on new evidence)

- **Generalizability Limitations:** Many models are trained on single-center data and perform poorly in other centers (21). A 2025 study showed that an XGBoost model, despite excellent training AUC, had poor external validation performance, indicating a risk of overfitting.

- **Sensitivity-Specificity Dilemma:** Some models, such as ChatGPT 4.0, achieve high sensitivity (95.7%) to avoid undertriage but very low specificity (18.2%), leading to significant over-triage (26). In AUC interpretation, it should be noted that in imbalanced datasets (e.g., ICU admission with 8.2% prevalence), the Area Under the Precision-Recall Curve (AUPRC) is a more informative metric, and AUC alone is insufficient for clinical decision-making. A 2025 study showed that despite a high AUC,

Table 1. Systematic information of included studies in the systematic review

| No | Authors (Year)                    | Country      | Study Design           | Sample Population | Primary Objective            | AI Model Type   | Key Performance Metrics               | Key Findings                             | Limitations                              |
|----|-----------------------------------|--------------|------------------------|-------------------|------------------------------|-----------------|---------------------------------------|------------------------------------------|------------------------------------------|
| 1  | Karlafti et al. (2023) (18).      | Greece       | Prospective            | 332 patients      | ESI prediction               | Neural network  | F1-score 72.2%                        | Supports decision-making                 | Limited sample, class imbalance          |
| 2  | Phimphisan et al. (2025) (19).    | Thailand     | Retrospective cohort   | 1,683 EMS cases   | Emergency triage             | DNN             | AUC 0.91, Sensitivity 88.5%           | Outperforms ESI/NEWS/MEWS                | Regional data needs updates              |
| 3  | Kipourgos et al. (2022) (20).     | Greece       | Model development      | 616 patients      | ESI level prediction         | Fuzzy logic, ML | 99% accuracy (Fuzzy)                  | Suggests an appropriate specialist       | Small sample, no real-world testing      |
| 4  | Sitthiprawiat et al. (2025) (21). | Thailand     | Retrospective cohort   | 163,452 visits    | ICU admission prediction     | XGBoost, RF, LR | AUC 0.917, AUPRC 0.629                | Superior to CTAS, uses NLP               | Single-center, needs external validation |
| 5  | Chen et al. (2024) (22).          | USA          | Retrospective cohort   | 2,000,000+        | Severe trauma prediction     | XGBoost         | AUC 0.75–0.73                         | Improves national triage guidelines      | Single-center, needs external validation |
| 6  | Lu et al. (2023) (23).            | China        | Technology development | Simulation        | Automated drone triage       | YOLO, OpenPose  | Undertriage reduction <10%            | Improves speed and accuracy in disasters | Simulation only, needs field validation  |
| 7  | Hayashi et al. (2021) (24).       | Japan        | Prospective            | 13,831 patients   | Prehospital stroke diagnosis | XGBoost         | AUC 0.98                              | High LVO detection accuracy              | Needs validation in real EMS             |
| 8  | Petry et al. (2022) (25).         | USA          | Retrospective cohort   | Hospital data     | Radiology prioritization     | AI-augmented    | LOS reduction 11.9% (ICH), 26.3% (PE) | Improved efficiency                      | Retrospective, single-center             |
| 9  | Gebrael et al. (2023) (26).       | USA          | Retrospective analysis | 56 patients       | ChatGPT triage               | ChatGPT 4.0     | Sensitivity 95.7%, Specificity 18.2%  | Improved diagnostic accuracy             | Very small sample, single-center         |
| 10 | Alshareef (2025) (7).             | Saudi Arabia | Cross-sectional        | Public hospitals  | ED challenges                | None            | None                                  | Identified structural challenges         | Regional, limited generalizability       |
| 11 | Park et al. (2025) (27).          | USA          | Cross-sectional        | LLM data          | Algorithmic bias             | GPT-4           | Bias detection                        | Racial and gender disparities            | Needs clinical validation                |
| 12 | Jaber et al. (2024) (8).          | Saudi Arabia | Retrospective cohort   | Tertiary center   | Quality indicators           | None            | None                                  | Identified quality barriers              | Single-center                            |

Table 2. Risk of bias assessment using the Newcastle-Ottawa Scale (NOS)

| Study (Reference)               | Selection (0-4) | Comparability (0-2) | Outcome (0-3) | Total Score (0-9) | Quality  |
|---------------------------------|-----------------|---------------------|---------------|-------------------|----------|
| Karlafti et al. 2023 (18).      | 3               | 1                   | 2             | 6                 | Moderate |
| Phimphisan et al. 2025 (19).    | 3               | 2                   | 2             | 7                 | High     |
| Kipourgos et al. 2022 (20).     | 3               | 1                   | 2             | 6                 | Moderate |
| Sitthiprawiat et al. 2025 (21). | 4               | 2                   | 3             | 9                 | High     |
| Chen et al. 2024 (22).          | 4               | 2                   | 3             | 9                 | High     |
| Hayashi et al. 2021 (24).       | 3               | 2                   | 3             | 8                 | High     |
| Petry et al. 2022 (25).         | 3               | 1                   | 2             | 6                 | Moderate |
| Gebrael et al. 2023 (26).       | 3               | 2                   | 3             | 8                 | High     |

NOS scores  $\geq 7$  indicate high quality, scores 6 indicate moderate quality. No low-quality studies (score  $\leq 5$ ) were included.

the AUPRC showed a substantial improvement compared to the traditional system (0.629 vs. 0.333) (21).

•**Implementation Challenges:** User resistance, need for high-quality large datasets, and integration difficulties in busy ED workflows are major barriers (31). A 2025 umbrella review using the NASSS framework identified that the main resistance factors include model mismatch with real clinical workflow, privacy concerns, and lack of legal transparency. Hidden costs include EHR integration, cloud infrastructure needs, and temporary workflow disruption during the first 3-6 months of implementation (21).

#### Geographic Distribution and Generalizability Limitations

The United States, East Asian countries (South Korea, Japan, China), and some European countries (Greece, Sweden) are leaders in research and operational applica-

tions. Thailand and Brazil have also had successful applications in specific areas (prehospital triage and COVID-19 screening).

However, the majority of included studies (over 70%) originated from high-income countries. The absence of studies from Africa, the Middle East (except Saudi Arabia), and Latin America represents a serious geographic bias, and generalizing results to these regions should be done with caution. Furthermore, patient-centered outcomes such as mortality, length of stay (LOS), and readmission are rarely reported, with most studies focusing on process metrics (AUC, accuracy, sensitivity) (21).

#### Distinction Between Proof-of-Concept Studies and Real-World Effectiveness

It should be noted that many of the included studies are at the "proof-of-concept" stage, conducted in controlled

Table 1. Comparison of AI applications and performance in different emergency settings

| Operational Setting       | Examples (Model/Country)                             | Primary Application                                      | Key Performance (AUC/Accuracy)                  | Limitations (based on base studies)                                            |
|---------------------------|------------------------------------------------------|----------------------------------------------------------|-------------------------------------------------|--------------------------------------------------------------------------------|
| Prehospital/Mass Casualty | Drone+YOLO (China) – pTEST model (USA) (ref. 20, 24) | Automated remote triage, destination determination       | Undertriage reduction to <10% (pTEST)           | Single-center data; lack of prospective validation (21)                        |
| In-hospital               | i-TRIAGE (Greece) / TriageGO (USA) (21,23)           | ICU admission prediction, internal prioritization        | ~99% accuracy for ESI prediction (i-TRIAGE)     | Limited sample size (616 patients); no real-world testing (23)                 |
| Disasters/Pandemics       | XGBoost models for COVID-19 (Ecuador)(10,21)         | Rapid screening of high-volume patients                  | Improved screening speed and congestion control | Significant performance drop in external validation (26,27)                    |
| Severe Trauma             | pTEST model (USA) – XGBoost (20)                     | Prediction of ICU admission and emergency surgery        | AUC 0.75-0.93                                   | Overtriage remains high (28)                                                   |
| Stroke Diagnosis          | XGBoost model (Japan) (29)                           | Ischemic vs. hemorrhagic differentiation, LVO prediction | AUC 0.98 for stroke detection                   | Not yet validated in real EMS workflow (28)                                    |
| COVID-19 Screening        | ResNet, CheXNet models (Italy, USA) (10)             | Detection from CT scans and chest radiographs            | Sensitivity 79-98%, Specificity 70-93%          | High dependence on image quality; not generalizable to new viral variants (26) |

environments or with retrospective data. Only a limited number of models have been prospectively validated in real clinical settings. For example, despite its 99% accuracy in experimental settings, the i-TRIAGE model has not yet been tested on a large scale in the real world. Therefore, proof-of-concept results should not be equated with real-world effectiveness.

## Discussion

### Interpretation of Key Findings

The findings of this multifaceted analysis indicate that artificial intelligence is increasingly moving from the proof-of-concept stage to adding value in real-world environments. The consistent superiority of models such as XGBoost and DNN over traditional triage systems demonstrates the technological maturity and statistical reliability of AI to complement clinical decision-making.

However, it should be noted that most included studies have focused on process metrics (e.g., AUC, accuracy, sensitivity), while patient-centered outcomes (e.g., mortality, length of stay, quality of life) are rarely reported. AUC alone is insufficient for clinical decision-making because, in imbalanced datasets (e.g., ICU admission with 8.2% prevalence), it can mask the true performance of a model. A 2025 study showed that despite a high AUC (0.917), the Area Under the Precision-Recall Curve (AUPRC) showed a substantial improvement compared to the traditional system (0.629 vs. 0.333), highlighting the importance of reporting complementary metrics such as PPV, NPV, and AUPRC (21, 32).

The role of AI is differentiated across three operational settings. In the prehospital setting, the main value lies in the speed and safety of initial assessment and optimal destination determination. In the hospital setting, the value lies in deep integration with EHR data and prediction of complex outcomes (e.g., patient deterioration). In mass casualty incidents, AI becomes a logistical enabler for managing chaos and allocating limited resources. These differences confirm that AI system design and implementation must be tailored to the operational context, and a

single solution cannot be prescribed for all settings (23, 27).

### Implementation and Ethical Considerations

Successful implementation of AI in emergency departments requires a multidisciplinary approach:

- **Technical aspect:** Focus should be on developing models with high interpretability, for example, using SHAP, and robust multicenter validation to prevent bias and lack of generalizability. The lack of transparency of "black box" models (e.g., deep DNNs) reduces clinical acceptance; the use of interpretable methods such as SHAP can bridge this gap (21).

- **Human aspect:** Training healthcare providers and designing simple user interfaces appropriate for emergency stress is vital for technology acceptance. Triage nurses, as the primary entry point for data input and output interpretation, play a central role in this ecosystem and should be involved in the design process. However, a 2025 umbrella review using the NASSS (Non-adoption, Abandonment, Scale-up, Spread, Sustainability) framework identified that the main factors of user resistance include model mismatch with real clinical workflow, privacy concerns, and lack of legal transparency (21, 27, 28).

- **Ethical and governance aspect:** Developing transparent frameworks for accountability, data privacy, and monitoring algorithm performance is essential. Beyond general statements, there is concrete evidence of algorithmic bias in emergency triage. A 2025 study showed that large language models (LLMs) such as GPT-4, despite good predictive performance, exhibit discriminatory biases at the intersection of race and gender, such that gender differences are exacerbated within specific racial groups. This finding raises serious concerns about fairness in triage. Additionally, during the COVID-19 pandemic, some oxygen allocation algorithms exacerbated racial disparities (26, 27). The risk of over-reliance on the system and erosion of clinical skills also requires continuous educational policies. Hidden implementation costs include EHR integration, cloud infrastructure needs, and temporary workflow disruption during the first 3-6 months (21).

### Study Limitations and Geographic Bias

This analysis is based on 12 included studies. Most existing studies are not prospective or large-scale, and evidence for improvement in final patient outcomes in real-world settings is still evolving.

**Geographic limitations and generalizability:** The majority of included studies (over 70%) originated from high-income countries (the United States, European countries, East Asia). Only one study (25) from a resource-limited setting (Thailand) showed that AI models with an AUC of 0.91 are implementable but require local data and continuous updates. The absence of studies from Africa, the Middle East (except Saudi Arabia), and Latin America represents a serious geographic bias, and generalizing results to these regions should be done with caution (33, 19, 31).

Furthermore, the lack of standardization in outcome measures (e.g., differences in defining "critical outcome" or "ICU admission") and the lack of calibration metrics (e.g., Brier score) alongside AUC make comparison and synthesis of results challenging. The focus of articles on advanced healthcare systems may provide an incomplete picture of implementation challenges in resource-limited settings (33, 21).

### Recommendations for Future Research

1. Designing prospective multicenter studies and randomized controlled trials (RCTs) to assess the direct impact of AI on patient-centered outcomes (e.g., mortality, length of stay, readmission) in real-world emergency settings. The use of standardized reporting frameworks such as TRIPOD-AI and CONSORT-AI is also essential (28, 30).

2. Developing and sharing standardized, diverse, multicenter datasets for training more generalizable models and reducing geographic and demographic bias (21, 28).

3. Research on hybrid models that seamlessly integrate AI into existing clinical decision-making frameworks (e.g., sepsis protocols). The use of interpretable methods such as SHAP to increase transparency and clinical acceptance of "black box" models is also recommended (21).

4. Evaluating the cost-effectiveness of large-scale AI system implementation in emergency departments, considering hidden costs such as EHR integration and cloud infrastructure (21).

5. Conducting studies in low- and middle-income countries to assess model generalizability and identify local implementation challenges (19, 3).

### Conclusion

Artificial intelligence clearly has the potential to become a transformative ally in emergency triage and diagnosis. Existing evidence indicates its superior performance in improving accuracy, speed, and efficiency. However, this technology is not a magic solution, and its success depends on understanding that AI is a supportive tool that, in the hierarchy of decision-making, ultimately operates under the governance of clinical expert judgment.

Current evidence is largely based on retrospective, single-center studies of moderate to high quality, with only a limited number of models prospectively validated in real-world clinical settings. Most studies have focused on process metrics (e.g., AUC, accuracy, sensitivity), while patient-centered outcomes (e.g., mortality, length of stay, readmission) are rarely reported. Furthermore, AUC alone is insufficient for clinical decision-making and should be reported alongside metrics such as PPV, NPV, AUPRC, and calibration metrics (e.g., Brier score).

Regarding implementation and ethical challenges, risks of algorithmic bias (particularly at race and gender intersections), mismatch between models and real-world clinical workflows, hidden costs of EHR integration and cloud infrastructure, and lack of legal transparency are major barriers to widespread AI implementation in emergency departments. The use of frameworks such as NASSS and interpretable methods such as SHAP can help improve clinical acceptance and reduce user resistance.

The final recommendation is that health systems adopt a gradual, collaborative, and responsible approach, initiating AI implementation with pilot projects in high-risk, low-error domains (e.g., radiology prioritization and sepsis screening), while continuously evaluating its impact on staff, processes, and most importantly, patient outcomes. Prospective multicenter studies evaluating patient-centered outcomes and utilizing standardized reporting frameworks such as TRIPOD-AI and CONSORT-AI are essential. Further research on hybrid models that seamlessly integrate AI into existing clinical decision-making frameworks (e.g., sepsis protocols), as well as studies in low- and middle-income countries to assess model generalizability, are future research priorities.

The future of smart emergency care will be realized not by replacing humans with machines, but through the unique synergy between algorithmic precision and clinical wisdom, supported by continuous ethical and regulatory oversight.

### Acknowledgment

This article has been presented in the 4th International Congress of Trauma Highlights. Abstracts of the 4th International Congress of Trauma Highlights; 2025 Dec 23–26; Tehran, Iran. We gratefully thank the Trauma and Injury Research Center (Tehran, Iran) for cooperating in this study, and all individuals supporting our research.

### Conflict of Interests

The authors declare that they have no competing interests.

### Authors' Contributions

M.S. and E.H. designed the study. M.S., F.E., and E.A. performed the literature search and data extraction. E.H. supervised the project. M.S. and F.E. wrote the first draft. E.A. and E.H. revised the manuscript critically. All authors read and approved the final manuscript.

### Ethical Considerations

Not applicable.

### Funding Support

No funding was received for this study.

### Data Availability

The data analyzed in this study are available from the corresponding author upon reasonable request.

### AI Use Statement

The authors declare that no artificial intelligence (AI) tools or large language models (LLMs) were used in the design, data collection, analysis, writing, or revision of this manuscript. All content has been generated solely by the authors.

### References

1. Abi-Khalil M, Ongaro G, Berger E, Wijayasinghe S, Hausser J. In-hospital triage in 2022. *Rev Med Suisse*. 2022;18(791):1501-3. doi:10.53738/REVMED.2022.18.791.1501.
2. Nakao H, Ukai I, Kotani J. A review of the history of the origin of triage from a disaster medicine perspective. *Acute Med Surg*. 2017;4(4):379-84. doi:10.1002/ams2.293.
3. Yancey CC, O'Rourke MC. Emergency department triage. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan-. Updated 2023 Aug 28. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK557583/>.
4. Khursheed M, Ejaz K, Hanif F. Evolution of triage services in the emergency department Aga Khan University Hospital-Karachi. *Prehosp Disaster Med*. 2011;26(S1):s71-2. doi:10.1017/S1049023X11002457.
5. Safari S, Rahmati F, Baratloo A, Motamedi M, Forouzanfar MM, Hashemi B, et al. Hospital and pre-hospital triage systems in disaster and normal conditions; a review article. *Iran J Emerg Med*. 2015;2(1):2-10.
6. Nates JL, Sprung CL. Emergency medicine and critical care triage. In: *Compelling ethical challenges in critical care and emergency medicine*. Cham: Springer; 2020. p. 77-88.
7. Alshareef AG. Challenges facing emergency departments in public hospitals. *Edelweiss Appl Sci Technol*. 2025;9(1):1119-28. doi:10.55214/25768484.v9i1.4348.
8. Jaber MJ, Bindahmsh AA, Abu Dawwas MA, du Preez SE, Alshodukhi AM, Alateeq IS, et al. Obstacles and challenges affecting quality indicators in a complex tertiary emergency center. *Glob J Qual Saf Healthc*. 2024;7(4):162-74. doi:10.36401/JQSH-24-10.
9. Ashaolu OT, Lyons W, Stefanakos I, Calinescu R, Habli I, Hodge V, et al. Autonomous emergency triage support system. In: 2023 International Conference on Computational Science and Computational Intelligence (CSCI); 2023. IEEE.
10. Shirazi FB, Moslehi S, Rasouli MR, Masoumi G. Identifying the challenges of simulating the hospital emergency department in the event of emergencies and providing effective solutions: a qualitative study. *Disaster Med Public Health Prep*. 2023;17:e202. doi:10.1017/dmp.2022.144.
11. Masbi M, Tavakoli N, Dowlati M. Challenges of providing of special care services in hospitals during emergencies and disasters: a scoping review. *BMC Emerg Med*. 2024;24(1):238. doi:10.1186/s12873-024-01160-1.
12. Morris R, Karam BS, Murphy P, Jenkins P, Milia D, Hemmila M, et al. Injury severity and hospital resource utilization: a review of current trauma triage activation criteria. *J Trauma Acute Care Surg*. 2021.
13. Gasperini B, Cherubini A, Fazi A, Maracchini G, Prospero E. Older adults in emergency departments: the challenge of undertriage. *Intern Emerg Med*. 2016;11(8):1145-7.
14. Bürger J, Juncken K, Höhne D, Kamin K, Kleber C. Prävalenz und Ursachen von Fehltriagierung am Beispiel einer universitären Notaufnahme. *Notf Rettungsmed*. 2023;26(8):583-92.
15. Amri A, Manjas M, Hardisman H. Analisis implementasi triage, ketepatan diagnosa awal dengan lama waktu rawatan pasien di RSUD Prof. DR. MA Hanafiah SM Batusangkar. *J Kesehat Andalas*. 2019;8(3):484-92.
16. Peng A, Rohacek M, Ackermann S, Ilsemann-Kakaroumis J, Ghanim L, Messmer A, et al. The proportion of correct diagnoses is low in emergency patients with nonspecific complaints presenting to the emergency department. *Swiss Med Wkly*. 2015;145:w14121.
17. El Khatib A, Moh ANS, editors. Optimizing patient triage in emergency department reception: a hybrid algorithm approach. In: 2024 10th International Conference on Control, Decision and Information Technologies (CoDIT); 2024. IEEE.
18. Karlafti E, Anagnostis A, Simou T, Kollatou AS, Paramythiotis D, Kaiafa G, et al. Support systems of clinical decisions in the triage of the emergency department using artificial intelligence: the efficiency to support triage. *Acta Med Lit*. 2023;30(1):19.
19. Phimphisana S, Wichaiyo W, Saengprajak N, Jantu W, Sirigit W, Phimphisana S, et al. Artificial intelligence for EMS triage: a data-driven approach to emergency patient prioritization in Kalasin Province, Thailand. *Eng Technol Appl Sci Res*. 2025;15(4):24204-10.
20. Kipourgos G, Tzenalis A, Koutsojannis C, Hatzilygeroudis I. An artificial intelligence based application for triage nurses in emergency department, using the emergency severity index protocol. *Int J Caring Sci*. 2022;15(3):1764-72.
21. Sitthiprawiat P, Wittayachamnankul B, Sirikul W, Laohavisudhi K. Development and internal validation of an AI-based emergency triage model for predicting critical outcomes in emergency department. *Sci Rep*. 2025;15(1):1-9.
22. Chen Q, Qin Y, Jin Z, Zhao X, He J, Wu C, et al. Enhancing performance of the national field triage guidelines using machine learning: development of a prehospital triage model to predict severe trauma. *J Med Internet Res*. 2024;26:e58740.
23. Lu J, Wang X, Chen L, Sun X, Li R, Zhong W, et al. Unmanned aerial vehicle based intelligent triage system in mass-casualty incidents using 5G and artificial intelligence. *World J Emerg Med*. 2023;14(4):273-9.
24. Hayashi Y, Shimada T, Hattori N, Shimazui T, Yoshida Y, Miura RE, et al. A prehospital diagnostic algorithm for strokes using machine learning: a prospective observational study. *Sci Rep*. 2021;11(1):20519.
25. Petry M, Lansky C, Chodakiewicz Y, Maya M, Pressman B. Decreased hospital length of stay for ICH and PE after adoption of an artificial intelligence-augmented radiological worklist triage system. *Radiol Res Pract*. 2022;2022:2141839.
26. Gebrael G, Sahu KK, Chigara B, Tripathi N, Thomas VM, Sayegh N, et al. Enhancing triage efficiency and accuracy in emergency rooms for patients with metastatic prostate cancer: a retrospective analysis of artificial intelligence-assisted triage using ChatGPT 4.0. *Cancers (Basel)*. 2023;15(14):3717.
27. Park Y, Jackson GP, Foreman MA, Gruen DM, Khan S, Saria S. Evaluating algorithmic bias in large language models for emergency triage: a cross-sectional study across race and gender intersections. *JAMA Netw Open*. 2025;8(2):e2458891.
28. Kuttan N, Pundkar A, Gadkari C, Patel A, Kumar A. Transforming emergency medicine with artificial intelligence: from triage to clinical decision support. *Multidiscip Rev*. 2025;8(10):2025285.
29. Redruello-Guerrero P, Jiménez-Gutiérrez C, Láinez Ramos-Bossini AJ, Jiménez-Gutiérrez PM, Rivera-Izquierdo M, Benítez Sánchez JM. Artificial intelligence for the triage of COVID-19 patients at the emergency department: a systematic review. *Signa Vitae*. 2022;18(6):1-10.
30. Zarei R, Torgerson L, Downs MC. Artificial intelligence, machine learning and neural networks for prehospital triage: a narrative review. *Cureus*. 2025;17(9):e91542.
31. Andrade Magalhães ME, Lemes da Silva CV, Monteiro de Oliveira H, Rodrigues de Lima AB, Salum Flores MT, Ferreira Leite I, et al. The use of artificial intelligence in patient triage in emergency departments: an integrative review. *Rev Gest Soc Ambient*. 2024;18(12):e010234.
32. elicit.com. What is the role of artificial intelligence in hospital emergency triage and diagnosis? [Internet]. 2025 [cited 2026 Apr 26]. Available from: <https://elicit.com>.
33. Almulihi QA, Alquriani AA, Alquriani FA, Almulihi FA, Alzahid AA, Al Qahtani SSAJ, et al. Applications of artificial intelligence and machine learning in emergency medicine triage: a systematic review. *Med Arch*. 2024;78(3):198-206.