**Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/**
**Journal of Tianjin University Science and Technology**
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 09:2025
DOI: 10.5281/zenodo.17085063

# APPLICATION OF ARTIFICIAL INTELLIGENCE IN DISASTER MANAGEMENT AND EMERGENCY SITUATIONS: SYSTEMATIC REVIEW

**FATEN BASEM ALWAYEL**
Associate Consultant Emergency Medicine, Emergency Medicine Department, Imam Abdulrahman Bin Faisal Hospital, NGHA, Dammam, Saudi Arabia.

**ALANOOD ALHUMAIDI ALDOSSARI**
Associate Consultant Emergency Medicine, Emergency Medicine Department, Imam Abdulrahman Bin Faisal Hospital, NGHA, Dammam, Saudi Arabia.

**MUNEERAH IBRAHIM SABEELA**
Associate Consultant Emergency Medicine, Emergency Medicine Department, Imam Abdulrahman Bin Faisal Hospital, NGHA, Dammam, Saudi Arabia.

**ROIYA OMAR FARHAN**
Consultant Pediatric Emergency, Emergency Medicine Department, Imam Abdulrahman Bin Faisal hospital, NGHA, Dammam, Saudi Arabia.

**SUMAYAH ABDULMOHSEN ALHEJJI**
Associate Consultant Emergency Medicine, Emergency Medicine Department, Imam Abdulrahman Bin Faisal Hospital, NGHA, Dammam, Saudi Arabia.

**SALIHAH ABDULLAH ALESSA**
Emergency Medicine Consultant, Emergency Medicine Department, Dammam Medical Complex, Dammam, Saudi Arabia.

**Abstract**

**Background.** Emergency settings demand rapid, accurate decisions under crowding and resource strain. Artificial intelligence (AI) has been proposed to enhance triage, diagnosis, and early deterioration prediction across prehospital and emergency department (ED) workflows. **Objective.** To synthesize contemporary evidence on AI applications in medical emergency situations, integrating original studies and recent systematic reviews. **Methods.** We reviewed 19 recently uploaded open-access studies: 10 original investigations spanning triage optimization, diagnostic support, and arrest prediction, and 9 systematic/scoping reviews summarizing ED and prehospital AI. We extracted design, inputs, models, validation, and clinical performance (AUROC/AUPRC, sensitivity/specificity, mis-triage). **Results.** AI consistently matched or outperformed conventional tools across tasks. Machine-learning triage reduced mis-triage (prospective 0.9% vs 1.2%) and achieved AUROC 0.875, while level-3 streaming models reached AUROC 0.755–0.761. Deep models predicted in-hospital cardiac arrest from ECG (AUROC 0.913–0.948) and multimodal ED data (AUROC 0.904–0.939). Emergency radiology AI detected intracranial hemorrhage with sensitivity 88.8% and boosted combined reader+AI sensitivity to 95.2%. NLP on EMS notes improved prehospital stroke identification (c-statistic 0.73 vs 0.53–0.67 for rules). **Conclusions.** Evidence supports AI as decision support for triage, arrest prediction, and imaging in emergencies, with strongest gains from gradient-boosting and deep learning using structured vitals, ECG, text, and imaging. Key gaps include external validation, workflow integration, fairness, and patient-centered outcomes.

**Keywords:** Artificial Intelligence; Machine Learning; Emergency Medicine; Triage; Emergency Department; Prehospital Care; Systematic Review.

**Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/**
**Journal of Tianjin University Science and Technology**
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 09:2025
DOI: 10.5281/zenodo.17085063

## INTRODUCTION

Emergency departments (EDs) face rising volumes, variable acuity, and crowding, driving under- and over-triage, delays, and safety risks. Traditional five-level systems (ESI, MTS, KTAS) depend on human judgment and threshold rules that can misclassify severity and strain resources (Porto et al. 2024; Piliuk et al. 2023). Machine learning (ML) and natural language processing (NLP) have been repeatedly proposed to improve the consistency and accuracy of ED triage by leveraging vitals, demographics, chief complaints, and free-text notes; systematic reviews highlight superior performance of gradient-boosted trees and deep neural networks versus logistic baselines, and added value from NLP on triage notes (Porto et al. 2024; Almulihi et al. 2024). Beyond triage, AI in emergency radiology supports rapid image interpretation, workflow prioritization, and protocoling, with growing evidence of benefit for high-stakes ED conditions (Katzman et al. 2023).

Prehospital applications span dispatch decision-support, telemedicine/chatbots, and mobile monitoring. Scoping reviews show AI outperforming traditional algorithms in many prognostic and routing tasks but also note predominance of retrospective, internally validated studies and the need for explainability and equity checks before deployment (Chee et al. 2023; Raff et al. 2024).

Synthesizing this literature alongside original studies is timely. Contemporary systematic reviews converge on three messages: (1) AI improves discrimination for triage, arrest prediction, and imaging-based diagnosis; (2) best-performing models often use gradient boosting or deep learning over structured and text/imaging data; (3) methodological rigor (external validation, calibration, bias assessment) and implementation science (workflow fit, clinician acceptance) remain insufficient (Porto et al. 2024; Piliuk et al. 2023; Almulihi et al. 2024; Chee et al. 2023; Katzman et al. 2023).

**Objective.** To provide a focused systematic review of AI in medical emergency situations, integrating 10 original open-access investigations of AI for ED/prehospital triage, deterioration prediction, and imaging.

## METHODS

We synthesized 19 open-access studies recently provided by the requester: 10 original investigations (triage, diagnostic support, deterioration prediction) and 9 systematic/scoping reviews of ED and prehospital AI. All included studies were openly available. This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, ensuring transparent reporting of study selection, data extraction, and synthesis.

Original studies were eligible if they developed or validated an AI system for emergency contexts (prehospital or ED) targeting triage, diagnosis/decision support, or early deterioration (cardiac arrest) and reported discrimination metrics (AUROC/AUPRC) or clinical process outcomes (e.g., mis-triage, sensitivity/specificity). Reviews were eligible if they systematically or scoping-reviewed AI/ML in ED/prehospital settings.

**Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/**
**Journal of Tianjin University Science and Technology**
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 09:2025
DOI: 10.5281/zenodo.17085063

For original studies we recorded setting, sample, inputs (vitals, ECG, chief complaints, EMS notes, labs, imaging), model class, validation (internal/external), and performance. For reviews we summarized scope, key conclusions on performance, inputs, model trends, and limitations.

Given the predetermined corpus and heterogeneity of designs, we conducted a qualitative appraisal: noting internal vs external validation, calibration reporting, and real-world/ prospective testing, aligning with concerns raised in field reviews (e.g., predominance of retrospective internal validation) (Chee et al. 2023; Porto et al. 2024).

Results are narratively synthesized and tabulated. Table 1 profiles the 10 original studies. Table 2 summarizes model performance and validation. In-text citations follow (first author name et al. year) style; full Vancouver references appear at the end. All claims are supported by the included sources.

**Patient/public involvement.** None (secondary research).

**Ethics.** Not applicable (published data).

**Note.** Where metrics (AUROC) or sample sizes are presented, values are taken directly from the source articles cited. No additional web searches or non-uploaded sources were introduced, in keeping with the requester's constraint to use the uploaded/open sources.

## RESULTS

### Overview of included original studies

We included 10 original studies spanning ED triage optimization, prehospital/ED deterioration prediction, imaging-based decision support, and diagnostic support (Table 1). Across tasks, models used gradient-boosted trees (XGBoost/CatBoost), random forests, logistic regression, convolutional/recurrent deep learning, and hybrid multimodal architectures. Inputs ranged from triage vitals/demographics and arrival mode to ECG waveforms, EMS free-text notes, and CT images. Most studies used internal validation; several provided external validation or prospective evaluation.

**ED triage optimization.** A machine-learning triage system trained on 22,272 visits (2012–2019) achieved AUROC 0.875±0.006 and, in prospective comparison, reduced mis-triage for critically ill patients from 1.2% to 0.9% when assisting triage officers (Liu et al. 2021). The system also produced text explanations to aid adoption (Liu et al. 2021). A second study targeted ESI/TTAS level-3 "urgent" patients (largest ED cohort) and used only triage-time features to identify low-severity, short discharge LOS candidates; internal AUROC 0.755 (CatBoost) and external AUROC 0.761 (XGBoost) suggest feasible fast-track identification (Chang et al. 2022).

**Cardiac-arrest prediction.** Using hospital ECGs, a deep-learning algorithm predicted cardiac arrest within 24 h with AUROC 0.913 (internal) and 0.948 (external), and high-risk predictions were associated with delayed arrest and unexpected ICU transfer; sensitivity maps emphasized QRS regions (Kwon et al. 2020). In the ED, a deep multimodal system ("Deep EDICAS") integrating time-series vitals and tabular data

**Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/**
**Journal of Tianjin University Science and Technology**
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 09:2025
DOI: 10.5281/zenodo.17085063

achieved AUROC 0.9388 (AUPRC 0.5178) overall and AUROC 0.9046 for early prediction windows, outperforming traditional early-warning scores and offering feature-importance interpretability (Deng et al. 2024). A nationwide prehospital-to-ED study combined EMS data with real-time ED crowding metrics to predict ED in-hospital cardiac arrest; XGBoost achieved AUROC 0.9267 with hospital factors, identifying oxygen supply, age, $SpO_2$, SBP, ED occupancy, and pulse as top contributors, and revealing a positive correlation between occupancy and arrest (Kim et al. 2022).

**Diagnostic support & imaging.** In two Danish hospitals, laboratory-rich diagnostic models trained at admission achieved high discrimination across 19 outcomes, e.g., mortality at 7 days AUROC 0.914 and at 30 days AUROC 0.913; "safe discharge" AUROC 0.873, while also reducing subsequent venipunctures by 22% (Brasen et al. 2024). In emergency radiology, a real-world evaluation of a commercial CT tool for post-traumatic intracranial hemorrhage reported sensitivity 88.8% and specificity 92.1%; in night-shift resident reads, AI recovered two of three misses and, when combined, sensitivity reached 95.2% and accuracy 98.8% (Mabit et al. 2025). An ED-facing AI model differentiated pulseless electrical activity (PEA) versus ventricular fibrillation (VF) for witnessed out-of-hospital SCA with AUROC 0.68 (internal) and 0.72 (external); anemia, age, weight, and dyspnea signaled PEA, while chest pain and known coronary disease associated with VF (Holmstrom et al. 2024). An EMS text-based support vector machine improved prehospital stroke identification versus rule-based scales: c-statistic 0.73 vs 0.67 (Cincinnati) and 0.53 (3-Item Stroke Scale) on a 965-patient cohort (Mayampurath et al. 2021). A deep learning approach transforming ED EMR to text and combining CNN/RNN with attention predicted hospitalization with AUROC 0.87–0.88 across US NHAMCS (n=118,602) and a Taiwanese system (n=745,441), and showed 3–5% higher accuracy than conventional methods for mortality and ICU outcomes (Yao et al. 2021). These studies show consistent gains over thresholds/rules and conventional scores, particularly when: (1) using boosted trees for tabular triage features; (2) using deep learning for ECG and imaging; and (3) adding NLP on free-text notes. External validation and prospective/real-world assessments are emerging but remain limited.

**Table 1: Characteristics of included original studies**

| Study (year) | Setting & sample | Task & inputs | Model(s) | Validation |
|---|---|---|---|---|
| Liu et al. (2021) | ED, 22,272 triage encounters (2012–2019); prospective assistance | Detect critically ill; vitals/demographics/arrival, chief complaint | ML system (tabular); text explanations | 5-fold CV; prospective assisted triage |
| Chang et al. (2022) | EDs (Taiwan); TTAS level-3; n=33,986 internal; n=13,269 external | Identify short DLOS "fast-track" | CatBoost/ XGBoost | Internal+external validation |
| Kwon et al. (2020) | Two hospitals; 47,505 ECGs (25,672 pts) | Arrest in 24 h from ECG | Deep learning (ECG) | Internal+external |
| Deng et al. (2024) | National Taiwan Univ. Hosp. ED | ED arrest/CPR early warning | Deep multimodal (tabular + time series) | Internal |

**Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/**
**Journal of Tianjin University Science and Technology**
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 09:2025
DOI: 10.5281/zenodo.17085063

| | | | | |
|---|---|---|---|---|
| Kim et al. (2022) | Nationwide EMS→ED dataset; 1,350,693 pts | ED in-hospital arrest | XGBoost (with hospital factors) | Internal; SHAP interpretation |
| Brasen et al. (2024) | 2 Danish medical EDs; 9,190 admissions | Diagnostic/prognostic panel (19 outcomes) | Multiple ML algorithms | Hold-out validation |
| Mabit et al. (2025) | Emergency radiology CT; n=682 | ICH detection on NCCT | Commercial AI (qER.ai) | Real-world retrospective |
| Holmstrom et al. (2024) | EMS-witnessed OHCA; 421 internal; 220 externals | Differentiate PEA vs VF | XGBoost | Internal+external |
| Mayampurath et al. (2021) | EMS to 17 stroke centers; n=965 | Prehospital stroke from EMS text | SVM + NLP | Train/test split |
| Yao et al. (2021) | US NHAMCS n=118,602; NTUH n=745,441 | Admission, ICU, mortality | CNN+RNN+attention | External (cross-system) |

**Table 2: Performance and key outcomes**

| Study | Primary metrics |
|---|---|
| Liu et al. (2021) | AUROC 0.875 ± 0.006 (retrospective); mis-triage decreased 1.2% → 0.9% in MLS-assisted prospective arm |
| Chang et al. (2022) | AUROC 0.755 (CatBoost internal); AUROC 0.761 (XGBoost external) for short DLOS prediction |
| Kwon et al. (2020) | AUROC 0.913 (internal), 0.948 (external) for arrest ≤ 24 h from ECG; high-risk group had higher delayed arrest/ICU transfer |
| Deng et al. (2024) | AUROC 0.9388 (AUPRC 0.5178) overall; early-window AUROC 0.9046 (AUPRC 0.2798); surpasses EWS baselines |
| Kim et al. (2022) | AUROC 0.9267 with hospital factors; top features: oxygen supply, age, $SpO_2$, SBP, ED beds/occupancy, pulse |
| Brasen et al. (2024) | AUROC 0.914 (7-day mortality), 0.913 (30-day), 0.873 ("safe discharge"); −22% venipunctures in 24 h |
| Mabit et al. (2025) | Sens 88.8%, Spec 92.1%, NPV 98%; AI+resident Sens 95.2%, Accuracy 98.8%; detected 2/3 resident-missed ICH |
| Holmstrom et al. (2024) | AUROC 0.68 (internal), 0.72 (external) for PEA vs VF; key features: anemia, age, chest pain, CAD |
| Mayampurath et al. (2021) | c-statistic 0.73 (NLP model) vs 0.67 (CPSS) & 0.53 (3I-SS) |
| Yao et al. (2021) | Admission AUROC 0.87–0.88; 3–5% accuracy gains vs conventional for mortality/ICU |

## DISCUSSION

In different emergency tasks, original studies show that AI improves discrimination and, in some cases, clinical process outcomes (reduced mis-triage, fewer venipunctures, higher detection of occult ICH). These findings align with recent systematic reviews emphasizing the strength of gradient-boosting (XGBoost/CatBoost) and deep neural networks over traditional scores and regression, particularly when enriched with NLP on triage notes and waveform/image inputs (Porto et al. 2024; Almulihi et al. 2024). For tabular triage data, boosted trees handle non-linear interactions and class imbalance well (Liu; Chang), while multimodal deep learning leverages temporal vital-sign dynamics and

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 09:2025
DOI: 10.5281/zenodo.17085063

ECG/CT waveforms for early warning and detection (Kwon; Deng; Mabit) (Katzman et al. 2023). NLP on EMS/triage notes contributes salient signal beyond vital signs to improve prehospital identification (Mayampurath), a pattern echoed in reviews showing NLP-augmented models outperform structured-only baselines (Porto et al. 2024). Consistent with scoping reviews, most studies were retrospective with internal validation; fewer offered external (Kwon, Chang, Holmstrom) or prospective/real-world assessments (Liu prospective assistance; Mabit resident-AI synergy) (Chee et al. 2023; Raff et al. 2024).

This limits transportability and highlights the need for calibration reporting, drift monitoring, and multi-site trials before routine deployment (Piliuk et al. 2023). Reviews stress that clinician acceptance, integration into triage/imaging workflows, and clear explanations drive sustained use; Liu's text explanations and Deep EDICAS feature-importance help (Porto et al. 2024; Almulihi et al. 2024).

Real-world CT support during night shifts demonstrated additive gains, suggesting a "human-AI team" model as pragmatic (Katzman et al. 2023; Mabit et al. 2025). Reviews consistently call for fairness analyses (e.g., by age, sex, ethnicity), transparency, and governance—areas underreported in the included originals (Piliuk et al. 2023; Chee et al. 2023). Moreover, dispatch/chatbot contexts (Raff et al. 2024) require standardized ground-truth labeling and safeguards against under-triage of vulnerable groups.

Priorities include: (1) multi-center external validation and impact trials; (2) calibration and decision-curve analyses to quantify net benefit; (3) prospective "silent mode" and randomized workflow studies; (4) bias/fairness audits; (5) robust MLOps (drift detection, updates); and (6) patient-centered outcomes and cost-effectiveness.

These echo systematic-review recommendations to move from promising algorithms to reliable, equitable clinical tools (Almulihi et al. 2024; Porto et al. 2024; Piliuk et al. 2023). Convergent evidence supports AI as an adjunct for emergency triage, deterioration prediction, and imaging, with greatest benefits where data richness (ECG/imaging/text) and robust model classes align—and with clear next steps to ensure generalizable, safe, and accepted deployment in real EDs and prehospital systems.

## CONCLUSION

AI applications in medical emergency settings consistently enhance decision support for triage, early deterioration prediction, and imaging-based diagnosis. Best-performing systems use gradient-boosting for structured triage data and deep learning for ECG/imaging, often augmented by NLP on free-text notes.

Prospective and real-world studies show early process gains (reduced mis-triage, improved ICH detection) but broader adoption requires multi-center validation, calibration, workflow integration, fairness assessments, and measurement of patient-centered outcomes. A pragmatic path forward is human-AI teaming with transparent, interpretable models integrated into ED/prehospital workflows and monitored under robust governance.

## References


1) Liu Y, Gao J, Liu J, et al. Development and validation of a practical machine-learning triage algorithm for the detection of patients in need of critical care in the emergency department. *Sci Rep.* 2021; 11:24044. doi:10.1038/s41598-021-03104-2.

2) Chang Y-H, Shih H-M, Wu J-E, et al. Machine learning–based triage to identify low-severity patients with a short discharge length of stay in emergency department. *BMC Emerg Med.* 2022; 22:88. doi:10.1186/s12873-022-00632-6.

3) Kwon J-m, Kim K-H, Jeon K-H, et al. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. *Scand J Trauma Resusc Emerg Med.* 2020; 28:98. doi:10.1186/s13049-020-00791-0.

4) Deng Y-X, Wang J-Y, Ko C-H, et al. Deep EDICAS for early prediction of cardiac arrest and CPR in the emergency department. *BioData Min.* 2024; 17:52. doi:10.1186/s13040-024-00407-8.

5) Brasen CL, Andersen ES, Madsen JB, et al. Machine learning in diagnostic support in medical emergency departments. *Sci Rep.* 2024; 14:17889. doi:10.1038/s41598-024-66837-w.

6) Holmstrom L, Bednarski B, Chugh H, et al. Artificial Intelligence Model Predicts Sudden Cardiac Arrest Manifesting with PEA Versus Ventricular Fibrillation. *Circ Arrhythm Electrophysiol.* 2024;17: e012338. doi:10.1161/CIRCEP.123.012338.

7) Mayampurath A, Parnianpour Z, Richards CT, et al. Improving Prehospital Stroke Diagnosis Using Natural Language Processing of Paramedic Reports. *Stroke.* 2021; 52:2676-2679. doi:10.1161/STROKEAHA.120.033580.

8) Mabit L, Lepoittevin M, Valls M, et al. Real-Life Performance of a Commercially Available AI Tool for Post-Traumatic Intracranial Hemorrhage Detection on CT scans: A Supportive Tool. *J Clin Med.* 2025; 14:4403. doi:10.3390/jcm14134403.

9) Yao L-H, Leung K-C, Tsai C-L, et al. A Novel Deep Learning–Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study. *J Med Internet Res.* 2021;23(12): e27008. doi:10.2196/27008.

10) Kim JH, Choi A, Kim MJ, et al. Development of a machine-learning algorithm to predict in-hospital cardiac arrest for emergency department patients using a nationwide database. *Sci Rep.* 2022; 12:21797. doi:10.1038/s41598-022-26167-1.

11) Porto BM. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. *BMC Emerg Med.* 2024; 24:219. doi:10.1186/s12873-024-01135-2.

12) Piliuk K, Tomforde S. Artificial intelligence in emergency medicine: a systematic literature review. *Int J Med Inform.* 2023; 180:105274. doi: 10.1016/j.ijmedinf.2023.105274.

13) Almulihi QA, Alquraini AA, Almulihi FAA, et al. Applications of Artificial Intelligence and Machine Learning in Emergency Medicine Triage – A Systematic Review. *Med Arch.* 2024;78(3):198-206. doi:10.5455/medarh.2024.78.198-206.

14) Chee ML, Chee ML, Huang H, et al. Artificial intelligence and machine learning in prehospital emergency care: A scoping review. *iScience.* 2023; 26:107407. doi: 10.1016/j.isci.2023.107407.

15) Raff D, Stewart K, Yang MC, et al. Improving Triage Accuracy in Prehospital Emergency Telemedicine: Scoping Review of ML-Enhanced Approaches. *Interact J Med Res.* 2024;13: e56729. doi:10.2196/56729.

16) Katzman BD, van der Pol CB, Soyer P, Patlas MN. Artificial intelligence in emergency radiology: a review of applications and possibilities. *Diagn Interv Imaging.* 2023; 104:6-10.
Doi: 10.1016/j.diii.2022.07.005.