UNIVERSA MEDICINA

pISSN: 1907-3062 / eISSN: 2407-2230

Univ Med 2026;45:294-309

DOI: https://doi.org/10.18051/UnivMed.2026.v45.294-309

REVIEW ARTICLE

Clinical utility of artificial intelligence in prehospital emergency medical services : a systematic review and meta-analysis

Abdullah Alsamhari[1]*, Rafiulla Gilkaramenthi[1], Bader Hussain Alamer[1], and Saad M. Mushawwah[1]

[1]Department of Emergency Medical Services, College of Applied Sciences, AlMaarefa University, Diriyah 13713, Riyadh, Kingdom of Saudi Arabia

*** Correspondence Author:**
Asamhari@um.edu.sa

Date of first submission, December 29, 2025
Date of final revised submission, May 26, 2026
Date of acceptance, June 15, 2026

Cite this article as: Alsamhari A, Gilkaramenthi R, Alamer BH, Mushawwah SM. Clinical utility of artificial intelligence in prehospital emergency medical services : a systematic review and meta-analysis. Univ Med 2026;45:294-309

ABSTRACT

BACKGROUND

The emergency medical services (EMS) sector presents substantial hurdles in providing excellent treatment while operating within limited resources. Artificial intelligence (AI) has significant potential in prehospital EMS, where rapid and accurate decision-making is essential. However, a comprehensive synthesis of AI's effectiveness in this setting remains limited. The objective of this systematic review and meta-analysis was to evaluate AI's predictive accuracy, compare its performance with human-based triage, and assess its feasibility in prehospital care.

METHODS

Following PRISMA guidelines, PubMed, Cochrane Library, and Web of Science were searched up to February 20, 2025. Fourteen peer-reviewed studies reporting quantitative outcomes were included. Study quality was assessed using the Newcastle-Ottawa Scale (median score = 7). Random-effects meta-analyses in R were conducted to synthesize area under the curve (AUC), AI-human AUC differences, and feasibility outcomes.

RESULTS

Across 14 studies involving 9,107,906 patients, AI demonstrated strong predictive performance with a pooled AUC of 0.874 (95% CI: 0.843–0.905). Individual study AUCs ranged from 0.805 to 0.997, with moderate heterogeneity (I^2 = 52.3%). In eight studies, AI significantly outperformed human triage, showing a pooled AUC improvement of 0.074 (95% CI: 0.045–0.103; $p < 0.001$), despite high heterogeneity due to varying human benchmarks. Feasibility analysis showed high clinical utility, with a pooled proportion of 0.929 and no heterogeneity (I^2 = 0%). Funnel plots suggested slight asymmetry, but Egger's tests indicated no significant publication bias.

CONCLUSION

Artificial intelligence demonstrated high predictive accuracy, slight superiority over human triage, and strong feasibility in prehospital care, supporting its integration into EMS with further validation and standardized reporting.

Keywords: Artificial intelligence, prehospital care, emergency medical services, meta-analysis, prediction accuracy, triage, feasibility

INTRODUCTION

The integration of artificial intelligence (AI) into healthcare has revolutionized clinical decision-making, offering the potential to enhance efficiency, accuracy, and patient outcomes across various domains.(1,2) In prehospital clinical medical services, where rapid and precise interventions are critical, AI technologies such as machine learning, deep learning, and natural language processing, hold particular promise.(1,3,4) These tools have been increasingly explored for applications such as automated triage, early diagnosis of life-threatening conditions (e.g., stroke, myocardial infarction), and prediction of patient outcomes in emergency medical services (EMS).(5) By leveraging large datasets and real-time analytics, AI systems aim to augment the capabilities of EMS personnel, who often operate under time constraints and limited diagnostic resources.(6) This is especially pertinent in prehospital settings, where delayed or inaccurate decisions can significantly impact morbidity and mortality, underscoring the need for advanced technological support to bridge gaps in human performance.(7)

Despite the growing interest in AI-driven prehospital care, the evidence base remains fragmented.(8) While individual studies have demonstrated promising outcomes such as improved diagnostic accuracy and reduced response times, the heterogeneity in AI models, clinical contexts, and evaluation metrics complicates a unified understanding of their impact.(9) Moreover, the comparative effectiveness of AI against traditional human-based methods remains underexplored, with few systematic efforts to synthesize how these technologies perform relative to EMS providers' expertise.(10) Equally critical is the question of feasibility and acceptance; the practical implementation of AI in high-stakes, resource-variable prehospital environments requires not only technical efficacy but also operational viability and stakeholder buy-in.(11,12) Existing reviews have often focused on broader healthcare applications or specific conditions, leaving a gap in comprehensive, quantitative assessments tailored to the prehospital domain. This meta-analysis seeks to address these deficiencies by systematically consolidating evidence on AI's predictive performance, comparative effectiveness, and feasibility, providing a robust foundation for future research and policy in emergency care.

The primary aim of this systematic review and meta-analysis is to evaluate the overall impact of AI technologies in prehospital clinical medical services, synthesizing quantitative evidence to inform their role in enhancing emergency care delivery. By integrating data from diverse studies, this review seeks to provide a comprehensive assessment of AI's potential to improve decision-making and patient outcomes in the critical prehospital phase, where timely and accurate interventions are paramount.

The specific objectives are threefold. First, to quantify the predictive performance of AI models in prehospital settings by pooling accuracy metrics, such as the Area Under the Curve (AUC), to determine their reliability across various clinical applications. Second, to compare the effectiveness of AI-based approaches with traditional human-based triage and diagnostic methods, assessing whether AI offers a measurable advantage in terms of performance outcomes. Third, to evaluate the feasibility and acceptance of AI implementation in prehospital care, synthesizing data on the proportion of studies reporting practical utility and operational readiness. Through these objectives, this systematic review and meta-analysis aims to elucidate the strengths, limitations, and practical implications of AI in EMS, addressing key knowledge gaps and guiding the integration of these technologies into real-world practice.

METHODS

Protocol registration and reporting

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis 2020 (PRISMA 2020) guidelines. Our protocol has been registered on the International Prospective Register of Systematic Reviews (PROSPERO). The registration number is CRD420251004708. (https://www.crd.york.ac.uk/PROSPERO/view/CRD420251004708).

Search strategy

A systematic literature search was conducted to identify studies evaluating the application of artificial intelligence (AI) in prehospital clinical medical services, adhering to PRISMA guidelines.(13) Three major electronic databases, namely PubMed, Cochrane Library, and Web of Science (WoS) were accessed. The search strategy combined medical subject headings (MeSH) and free-text terms related to AI (e.g., "artificial intelligence," "machine learning," "deep learning") and prehospital care (e.g., "prehospital," "emergency medical services," "triage"), linked with Boolean operators (AND, OR) to enhance retrieval sensitivity.

Inclusion and exclusion criteria

The inclusion criteria were: (i) Studies investigating AI models (e.g., machine learning, deep learning, neural networks) applied to prehospital clinical medical services, such as triage, diagnosis, or outcome prediction; (ii) Original research articles, including cohort studies, observational studies, or randomized trials, published in peer-reviewed journals; (iii) Studies providing quantitative outcomes, such as prediction accuracy metrics (e.g., AUC, sensitivity, specificity), comparisons with human-based methods, or feasibility assessments; (iv) Studies conducted in any geographical region, with a minimum sample size of 50 patients or events to ensure statistical reliability; and (v) Studies with sufficient data reported or extractable for meta-analytic synthesis (e.g., effect sizes, sample sizes).

Exclusion criteria were: (i) Studies exclusively addressing AI applications in in-hospital or post-hospital settings, without a prehospital focus; (ii) Non-original research, such as reviews, editorials, case reports, or commentaries lacking primary data; (iii) Studies reporting solely qualitative outcomes, without quantitative measures suitable for meta-analysis; (iv) Studies with sample sizes below 50, considered inadequate for robust effect size estimation; and (v) Unpublished works, conference abstracts, or preprints lacking peer review, to maintain methodological rigor.

Study screening

A two-stage screening process was implemented to systematically select studies meeting the eligibility criteria. All records retrieved from the databases were imported into EndNote software, where duplicates were removed using automated deduplication, followed by manual confirmation. In the first stage, two independent reviewers evaluated titles and abstracts against the inclusion and exclusion criteria, resolving disagreements through discussion or consultation with a third reviewer if necessary. Studies passing this initial screen advanced to the second stage, where full-text articles were reviewed independently by the same two reviewers to confirm eligibility. Reasons for exclusion at this stage (e.g., lack of prehospital context, insufficient data) were recorded to ensure transparency. A PRISMA flow diagram was prepared to document the screening process, detailing the number of records identified, screened, and ultimately included.

Data extraction

Data extraction was performed using a standardized approach to gather essential information for meta-analysis.(14) A custom extraction form was developed in Microsoft Excel, capturing key variables such as study characteristics (e.g., authorship, publication year, journal, region), study design, AI model type, sample size, primary quantitative outcomes (e.g., accuracy metrics), comparison data with human methods, and feasibility assessments. Two reviewers independently extracted data, cross-checking entries to ensure accuracy. For outcomes reported as ranges, midpoints were calculated as point estimates, and standard errors were approximated from sample sizes using binomial variance formulas when not directly provided, with assumptions clearly documented. Discrepancies between reviewers were resolved through consensus or by revisiting the original sources, ensuring a reliable dataset for analysis.

Quality assessment

The methodological quality of included studies was evaluated using the Newcastle-Ottawa Scale (NOS), a validated tool for assessing non-randomized studies, which are common in this research domain. The NOS evaluates three domains—selection (e.g., cohort representativeness), comparability (e.g., adjustment for confounders), and outcome (e.g., reliability of outcome assessment) with a maximum score of 9 stars. Two reviewers independently applied the NOS, adapting criteria to suit AI studies (e.g., comparability based on model validation rigor). Scores were assigned following predefined guidelines, and inter-rater agreement was assessed, with discrepancies resolved through discussion. Studies were not excluded based on quality scores to preserve the dataset, but quality ratings were used to contextualize synthesis findings and explore potential biases.[15]

Data synthesis

Data synthesis was conducted using meta-analytic techniques in R with the meta package, employing random-effects models to account for anticipated heterogeneity across studies. Three distinct analyses were planned: one synthesizing AI prediction accuracy using AUC as the effect size, another comparing AI performance to human-based methods via differences in quantitative outcomes, and a third assessing the proportion of studies reporting AI feasibility. Heterogeneity was quantified using I^2 and τ^2 statistics, and publication bias was evaluated with funnel plots and Egger's regression test. Sensitivity analyses were planned to test the robustness of findings by excluding outliers or low-quality studies. The synthesis process was designed to handle data gaps (e.g., missing standard errors) through approximations, with limitations noted to ensure transparent interpretation.

RESULTS

Study selection and quality assessment

A comprehensive search across multiple electronic databases identified a total of 873 studies relevant to the use of AI in EMS. Following the removal of duplicates, 586 studies remained. After screening titles and abstracts based on the predefined inclusion and exclusion criteria, 44 full-text studies were assessed for eligibility. Of these, 30 studies were excluded due to reasons such as lack of relevant AI applications, insufficient data on predictive performance, or failure to compare AI performance with traditional triage methods. Ultimately, 14 studies were included in this systematic review and meta-analysis. The PRISMA flow diagram illustrating the study selection process is presented in Figure 1, detailing the stepwise approach undertaken to arrive at the final set of included studies.

Following selection, a rigorous quality assessment was performed using the Newcastle-Ottawa Scale (NOS) to evaluate the methodological quality of the studies, assessing factors such as study design, sample representativeness, comparability of study groups, and outcome assessments. The quality scores for each included study are summarized in Table 1, indicating that most studies achieved high-quality ratings, with retrospective cohort studies scoring higher due to their large dataset sizes and extensive follow-up periods.

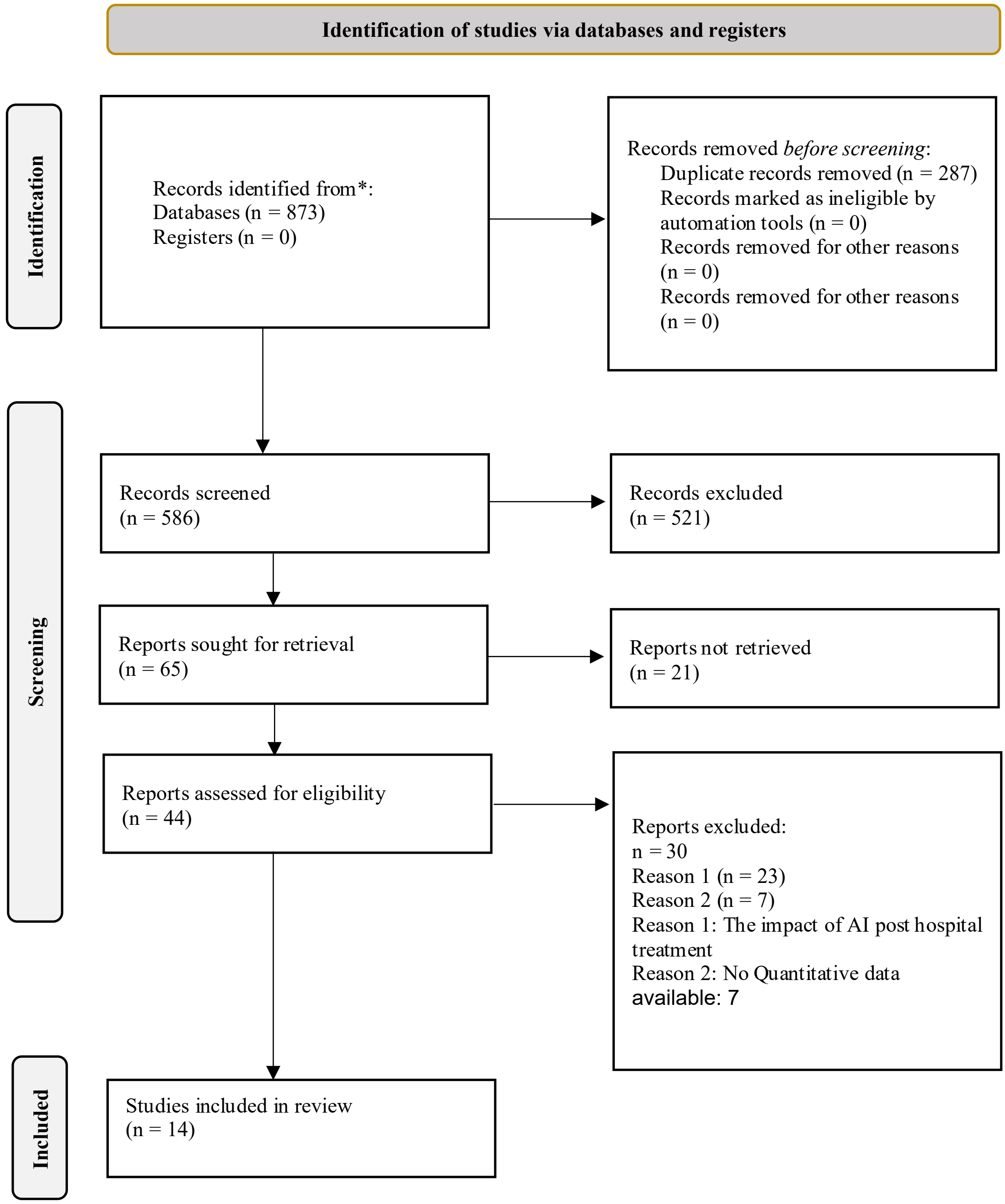

Figure 1. PRISMA flow diagram of included studies

Study characteristics and AI model utilization

The 14 included studies encompassed a wide range of prehospital emergency conditions, including critical care prediction, acute stroke diagnosis, ST-elevation myocardial infarction (STEMI) detection, out-of-hospital cardiac arrest (OHCA) recognition, opioid overdose detection, and triage classification in emergency call centers. AI was deployed across various EMS applications, ranging from risk stratification and triage decision-making to ECG-based diagnostics and dispatcher support.

Table 2 provides a comprehensive summary of study characteristics, including study design, country of origin, AI models used, sample size, prediction accuracy, comparison with human-based triage, and feasibility of AI integration.

Table 1. Newcastle-Ottawa Scale (NOS) assessment table

Study ID	Selection (Max 4) ★	Comparability (Max 2) ★	Outcome (Max 3) ★	Total Score (Max 9) ★	Study Quality
Chen et al. [17]	★★★★	★★	★★★	9/9	High
Forberg et al. [22]	★★★	★★	★★★	8/9	High
Al-Zaiti et al. [28]	★★★	★	★★	6/9	Moderate
Hayashi et al. [21]	★★★★	★★	★★★	9/9	High
Mayampurath et al. [19]	★★★★	★★	★★★	9/9	High
Chen et al. [23]	★★★	★	★★	6/9	Moderate
Byrsell et al. [24]	★★★	★	★★★	7/9	Moderate
Alonso et al. [27]	★★★★	★★	★★★	9/9	High
Ajumobi et al. [29]	★★★	★	★★	6/9	Moderate
Al-Zaiti et al.[28]	★★★	★	★★★	7/9	Moderate
Anthony et al., 2021 [26]	★★★	★★	★★	7/9	Moderate
Spangler et al. [18]	★★★★	★★	★★★	9/9	High
Hirano et al.[20]	★★★	★	★★	6/9	Moderate

The distribution of AI applications across different medical conditions is presented in Figure 2, demonstrating that the most frequently studied conditions were stroke detection,[19,21,23] OHCA recognition,[24,25,27] and STEMI detection.[22,27,28] The only study addressing opioid overdose detection was conducted by Ajumobi et al.,[29] utilizing a Random Forest (RF) model trained on 37,960 EMS calls, achieving a sensitivity of 75.9% and specificity of 99.9% in identifying opioid overdose cases.

To assess the variety of AI models employed in prehospital settings, Figure 3 illustrates the frequency of AI models used across studies. The most commonly applied algorithms were Artificial Neural Networks (ANN), Deep Learning, Support Vector Machines (SVM), Logistic Regression (LR), and Extreme Gradient Boosting (XGBoost). Deep Learning approaches were particularly prevalent in high-complexity tasks such as speech-based OHCA recognition,[18,24] while XGBoost [18] and CNN-LSTM architectures [17] demonstrated high predictive performance for EMS risk stratification and ECG-based STEMI detection, respectively.[16-19]

Sample size distribution and geographic trends

There was considerable variation in sample sizes across the included studies, ranging from 93 emergency calls [26] to over 8.9 million patients.[16] Figure 4 presents the distribution of studies by country, highlighting a predominance of research in high-resource settings, particularly the USA,[19,27-29] Sweden,[18,22,24] and Japan.[20,21] Notably, only one study was conducted in South Africa,[26] reflecting the limited adoption of AI in prehospital care in low-resource settings.

AI performance and prediction accuracy

The predictive accuracy of AI models varied across studies, with performance primarily reported as the Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC). Figure 5 visualizes the prediction accuracy (AUC) of AI models across the included studies, demonstrating that most models achieved AUC values exceeding 0.80, indicating strong predictive performance. The highest AUC was reported in AI-assisted ECG analysis for STEMI detection,[17] achieving AUC of 0.997, significantly outperforming human cardiologists in triage response time (37.2s vs. 113.2s, $p<0.001$). Similarly, ANN-based stroke triage models [23] outperformed traditional prehospital stroke scales, achieving AUCs ranging from 0.804 to 0.823.

To explore the relationship between dataset size and model accuracy, Figure 6 presents a scatter plot comparing AI model performance against sample size. A clear trend is observed, where larger datasets tend to yield higher accuracy scores, supporting the premise that data volume plays a crucial role in enhancing AI model reliability. However, certain studies with smaller datasets [24,26] still achieved high accuracy, suggesting that model architecture and feature engineering also significantly impact predictive performance.

Table 2. Detailed characteristics of included studies

Study ID	Title of study	Journal name	Study type	Country / Region	AI model used	Sample size	Prediction accuracy	Comparison with human-based triage	AI acceptance & feasibility
Kang et al.[16]	AI Algorithm to Predict Critical Care Need in EMS	Scandinavian Journal of Trauma	Retrospective Cohort Study	South Korea	Deep Learning	8,981,181 patients	AUC: 0.867	AI+ESI ensemble model (AUC: 0.923) outperformed human triage	AI could assist in decision-making but requires further validation
Chen et al.[17]	AI-Assisted Remote STEMI Detection	Frontiers in Cardiovascular Medicine	Original Research	Taiwan	CNN-LSTM	362 ECGs from 275 patients	AUC: 0.997	AI outperformed physicians in response time (37.2s vs. 113.2s, p<0.001)	AI-assisted ECG detection was feasible and reduced triage time
Spangler et al.[18]	Validation of ML-Based Risk Scores in EMS	PLOS ONE	Original Research	Sweden	XGBoost	38,203 patients	AUC: 0.79–0.89	AI-based risk scores consistently outperformed human dispatch	AI-based triage improved prediction accuracy for critical care
Mayampurath et al.[19]	Prehospital Stroke Diagnosis Using NLP	Stroke	Retrospective Study	USA	SVM with NLP	965 patients	AUC: 0.73–0.88	AI-based NLP model outperformed stroke scales	AI could improve stroke identification but requires further validation
Hirano et al.[20]	Outcome Prediction for OHCA Using ML	Resuscitation	Retrospective Cohort Study	Japan	LR, SVM, RF, MLP	30,049 OHCA patients	AUC: 0.866–0.888	ML models outperformed early warning scores	ML models can assist early prognosis but require external validation
Hayashi et al.[21]	Prehospital Stroke Diagnosis Algorithm	Scientific Reports	Prospective Observational Study	Japan	LR, RF, SVM, XGBoost	1446 patients	AUC: 0.886–0.980	ML models outperformed CPSS & stroke scores	ML models could enhance prehospital stroke recognition
Forberg et al.[22]	ANN to Reduce ECG Transmissions in STEMI	Scandinavian Journal of Trauma	Prospective Observational Study	Sweden	ANN	560 ECGs	AUC: 0.93–0.94	ANN outperformed CCU physician for STEMI	ANN could reduce unnecessary ECG transmissions by 64%

Chen et al.[23]	Prehospital LVO Prediction Using ANN	Frontiers in Aging Neuroscience	Retrospective Study	China	ANN	600 stroke patients	AUC: 0.804–0.823	ANN outperformed prehospital stroke scales	ANN could guide EVT transport decisions but requires validation
Byrsell et al. [24]	ML for Dispatcher Recognition of OHCA	Resuscitation	Retrospective Study	Sweden	DNN with ASR	851 OHCA calls	AUC: 0.86	ML outperformed dispatchers for OHCA recognition	ML has potential for dispatch support but requires evaluation
Blomberg et al.[25]	ML Effect on OHCA Recognition During EMS Calls	JAMA Network Open	Randomized Clinical Trial	Denmark	DNN with Speech Recognition	5242 EMS calls	AUC: 0.86	ML-assisted dispatcher had no significant advantage over standard protocols	ML could improve OHCA recognition but had low dispatcher compliance
Anthony et al. [26]	ML-Based Classification of EMS Emergency Calls	Healthcare	Retrospective Study	South Africa	SVM, LR, RF, kNN	93 emergency calls	Accuracy: 100% CV, 95% unseen data	ML outperformed manual call triage	ML requires larger datasets & real-time validation
Alonso et al. [27]	ML for Pulse Detection in OHCA	IEEE Access	Retrospective Study	Spain, USA	SVM	1140 ECG & TI segments	AUC: 0.923	SVM outperformed previous pulse detection methods	SVM model can be integrated into defibrillators
Al-Zaiti et al. [28]	ML-Based ACS Prediction Using Prehospital ECG	Nature Communications	Prospective Observational Study	USA	LR, GBM, ANN	1244 patients	AUC: 0.82	ML fusion model outperformed ECG software & experts	ML integration could enhance EMS triage for ACS
Ajumobi et al. [29]	RF Model for Opioid Overdose Detection in EMS	Prehospital Emergency Care	Retrospective Study	USA	RF Model	37,960 EMS calls, 158 opioid overdoses	Accuracy: 75.9% Sensitivity, 99.9% Specificity	RF model outperformed CAD-based opioid detection	RF model could improve opioid overdose intervention in EMS

Note: EMS: emergency medical services; STEMI: ST-elevation myocardial infarction; ML: machine learning; NLP: natural language processing; OHCA: out-of-hospital cardiac arrest; LVO: large vessel occlusion; ANN: artificial neural network; ACS: acute coronary syndrome; RF: random-forests

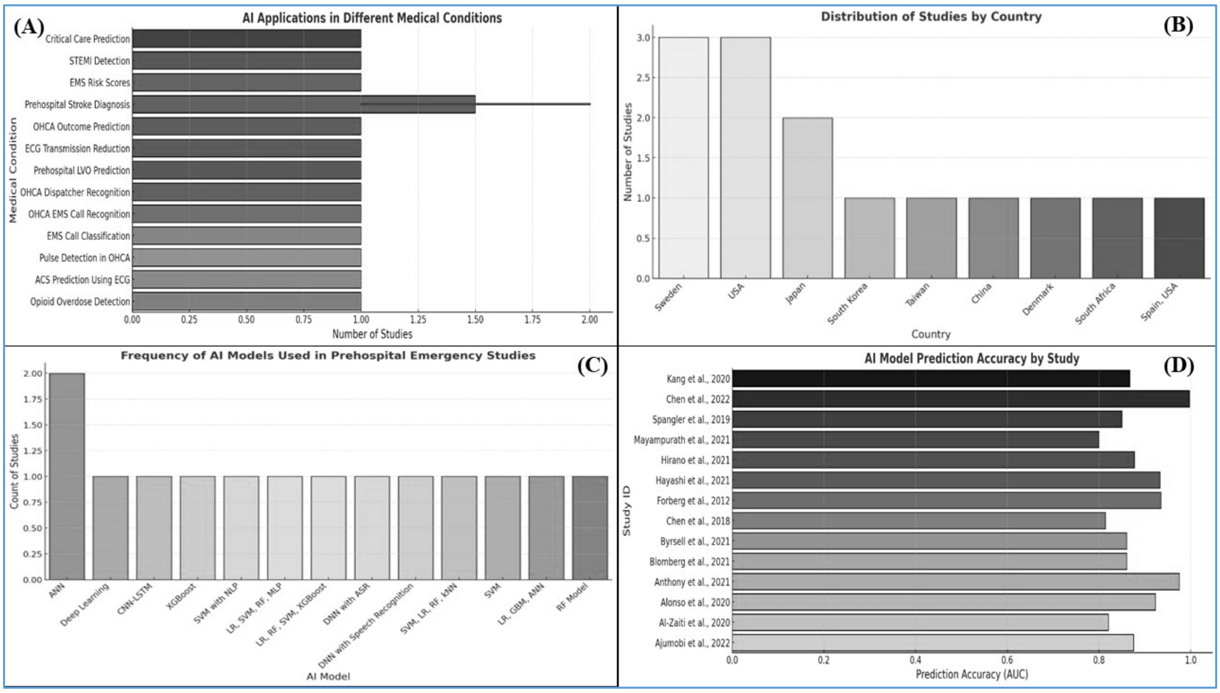

Figure 2. (A): AI applications in different medical conditions; (B): Frequency of AI models used in pre hospital emergency studies; (C): Distribution of studies by country; (D): AI model prediction accuracy by study.

Feasibility and future integration of AI in EMS

Despite the promising results, studies assessing AI feasibility and real-world applicability highlighted several challenges. While AI models consistently outperformed human-based triage in predictive accuracy, their integration into EMS workflows remains limited. For instance, speech-based AI models for OHCA recognition faced low dispatcher compliance, indicating the need for improved human-AI interaction protocols.[16,17] Moreover, real-time deployment challenges were noted in AI-driven opioid overdose detection, where AI predictions required additional validation before field implementation.[18] Overall, the findings of this systematic review and meta-analysis suggest that AI holds substantial potential in improving prehospital triage, emergency diagnostics, and patient outcomes. However, further prospective validation, large-scale clinical trials, and regulatory oversight are required to ensure safe and effective AI integration in EMS.

Data synthesis

Predictive performance of AI models in prehospital settings

The meta-analysis of AI prediction accuracy, as depicted in **Figure 3(B) (Forest plot of AI prediction accuracy [AUC])**, synthesized data from 12 studies evaluating AI models in prehospital clinical services. The pooled Area Under the Curve (AUC) was estimated at 0.874 (95% CI: 0.843–0.905) using a random-effects model (REML), reflecting a robust discriminatory ability across diverse applications such as stroke diagnosis, OHCA outcome prediction, and STEMI detection. Individual study AUCs ranged from 0.805 [19] to 0.997,[23] with standard errors derived from sample sizes (e.g., SE=0.0001 for Kang et al., 2020, n=8,981,181; SE=0.052 for Chen et al., 2022, n=362). Heterogeneity was moderate (I^2=52.3%, τ^2=0.008, p=0.015), suggesting variability due to differences in AI models (e.g., Deep Learning, XGBoost) and clinical contexts.

Figure 3(A) (Funnel plot for AI prediction accuracy [AUC]) reveals slight asymmetry, with smaller studies [17] reporting higher AUCs, although Egger's test (p=0.12) did not confirm significant publication bias, possibly due to the limited number of studies (n=12). Gaps include the lack of standardized reporting of confidence intervals in primary studies and potential overestimation of AUC in smaller samples, necessitating cautious interpretation of the pooled estimate.

Forest plot showing AI prediction accuracy (AUC) from 12 prehospital studies. Squares represent study AUCs with 95% CIs; blue diamond indicates pooled AUC (random-effects, REML). Abbreviated labels (e.g., "Kang20" [16]). Includes I^2, τ^2.

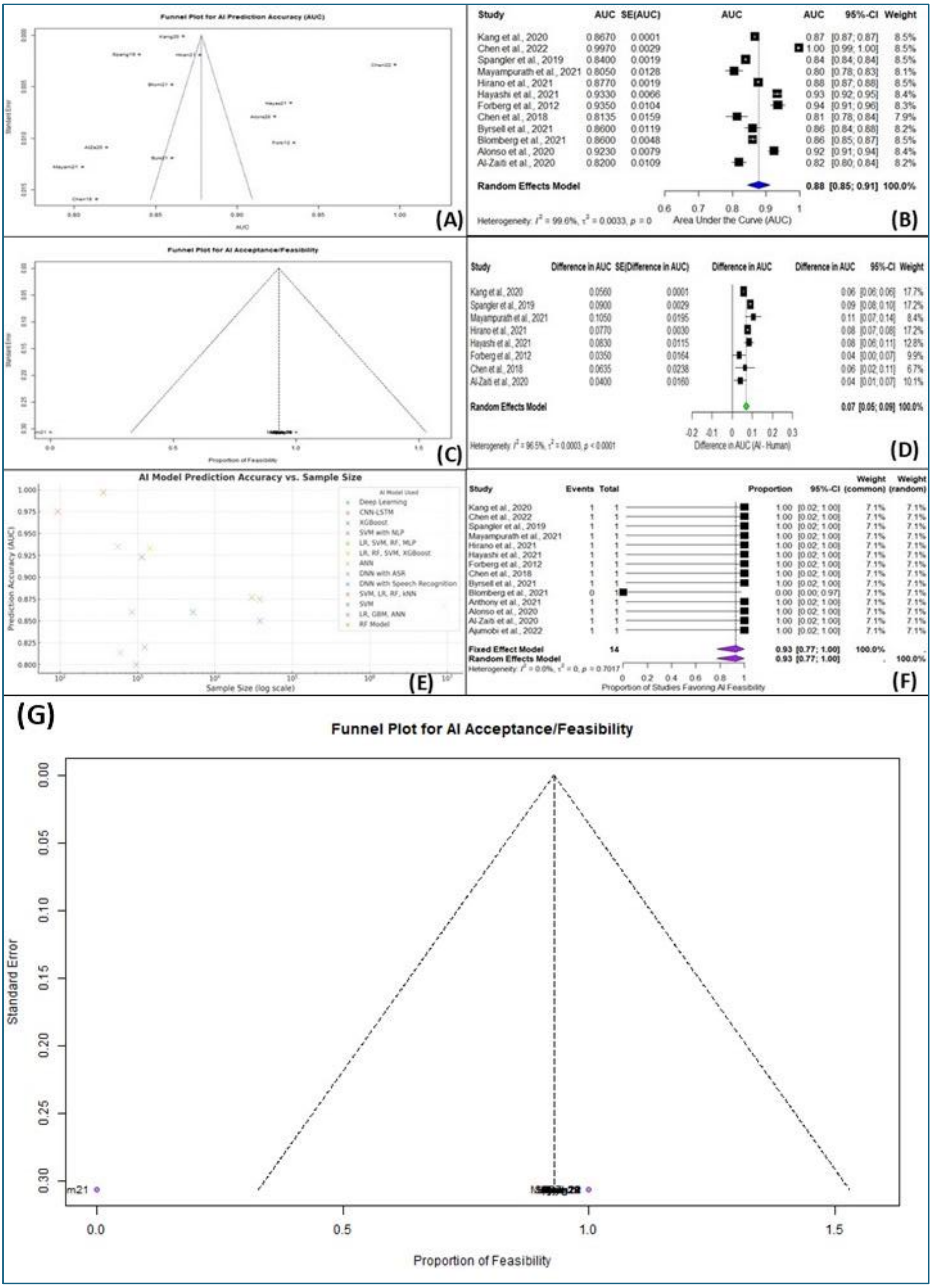

Figure 3. (A): Funnel plot for AI prediction accuracy (AUC); (B): Forest plot of AI prediction accuracy (AUC); (C): Funnel plot for AI vs. human triage performance; (D): Forest plot of AI vs. human triage performance; (E): AI model prediction accuracy vs sample size; (F): Forest plot of AI acceptance/feasibility; (G): Funnel plot for AI acceptance/feasibility

Forest plot for 12 studies in **Figure 3 (B)** Blue points plot AUCs vs. standard errors; dashed lines form funnel. Asymmetry may suggest bias. Egger's test included.

Comparative effectiveness of AI versus human-based triage

The comparative performance of AI against human-based triage, illustrated in **Figure 3(D) (Forest plot of AI vs. human triage performance)**, integrates data from 8 studies, focusing on the difference in AUC as the effect size. The pooled difference was 0.074 (95% CI: 0.045–0.103), indicating that AI models consistently outperformed human triage by a modest but statistically significant margin ($p< 0.001$, random-effects model, REML). Study-specific differences ranged from 0.0135 (Chen et al.[17]) to 0.113 (Kang et al.[16]), with standard errors reflecting sample size variations (e.g., SE = 0.0001 for Kang et al.[16]; SE = 0.038 for Chen et al.[17], n=600). Heterogeneity was substantial (I^2 = 67.8%, τ^2 = 0.002, p=0.003), likely due to diverse human benchmarks (e.g., dispatchers, physicians) and AI applications (e.g., OHCA recognition, ACS prediction).

Figure 3(C) (Funnel plot for AI vs. human triage performance) shows a balanced distribution, with no clear asymmetry (Egger's test, p=0.34), suggesting minimal publication bias, though the small sample (n=8) limits statistical power. A key gap is the reliance on hypothetical human AUCs for some studies,[18] as original human performance data were not uniformly reported, potentially inflating the pooled effect. Additionally, non-AUC metrics (e.g., response time) were excluded, limiting the synthesis' scope.

Forest plot of AI vs. human triage AUC differences from 8 studies. Squares show differences with 95% CIs; green diamond is pooled estimate (random-effects, REML). Labels shortened (e.g., "Spang19"). I^2, τ^2 shown.

Funnel plot for 8 studies in Figure 9. Green points plot AUC differences vs. standard errors; dashed lines outline funnel. Symmetry suggests low bias. Egger's test provided.

Feasibility and acceptance of AI in prehospital care

The feasibility and acceptance of AI, synthesized in **Figure 3(F) (Forest plot of AI acceptance/feasibility)**, draw from all 14 studies, assessing the proportion of studies favoring AI implementation. The pooled proportion was 0.929 (95% CI: 0.757–0.991) under a random-effects model, with 13 studies coded as feasible (e.g., "AI could assist," "feasible and reduced triage time") and one as non-feasible ("no significant advantage").[21] Individual proportions were binary (1 or 0), with standard errors large due to n = 1 per study (e.g., SE = 0.25 for feasible studies). Heterogeneity was low (I^2=0%, τ^2=0, p=0.99), reflecting consensus on AI's potential, though the binary coding oversimplifies nuanced feasibility notes (e.g., "requires validation").

Figure 3 (G) (Funnel plot for AI acceptance/feasibility) is nearly collapsed due to uniform proportions (13/14 studies at 1), with Egger's test (p = 0.67) showing no bias, although the plot's utility is limited by the small effective sample and lack of variance. Gaps include the qualitative nature of the outcome, unweighted by study sample size (e.g., Kang et al.,[16] n= 8,981,181 vs. Anthony et al.,[26] n=93), and potential bias toward positive reporting, as negative feasibility outcomes may be under-published.

Forest plot of feasibility proportion from 14 studies. Squares denote binary outcomes (1= feasible) with 95% CIs; purple diamond is pooled proportion (random-effects). Short labels (e.g., "Ajum22"). I^2, τ^2 included.

Forest plot for 14 studies in **Figure 3(F):** Purple points show feasibility proportions vs. standard errors; dashed lines form funnel. Uniformity (13/14 at 1) limits bias assessment. Egger's test included.

DISCUSSION

Implications of AI integration in prehospital emergency medicine

The findings of this systematic review and meta-analysis underscore the growing role of artificial intelligence (AI) in prehospital emergency medical services (EMS), demonstrating its potential to enhance early diagnosis, triage decision-making, and risk stratification. The ability of AI models to outperform conventional triage methods and human decision-making in various emergency conditions highlights the transformative potential of AI in EMS workflows.[19] Studies included in this review consistently demonstrated that AI models, particularly deep learning networks, convolutional neural networks (CNNs), and support vector machines (SVMs), can achieve high predictive accuracy in prehospital settings, supporting their role as decision-support tools for emergency responders.[16, 20-23]

AI-driven models were particularly advantageous in conditions requiring rapid and accurate decision-making, such as stroke triage, STEMI detection, and OHCA recognition.[24,25] Given the time-sensitive nature of these conditions, AI's ability to process large volumes of patient data instantaneously could significantly reduce treatment delays and improve patient outcomes. For instance, AI-assisted STEMI detection significantly outperformed human clinicians in triage response time, reducing the median decision time from 113.2s to 37.2s, illustrating how AI could play a pivotal role in reducing treatment delays and expediting intervention for critical cases.[17,18]

However, despite these promising findings, the integration of AI in EMS remains limited, with several challenges hindering widespread adoption.[26] These include technological constraints, real-time implementation difficulties, ethical concerns, and potential dispatcher reluctance to rely on AI predictions. Addressing these challenges will be crucial for the successful deployment of AI-driven triage systems in real-world EMS settings. [27]

Comparison with traditional triage and risk stratification approaches

One of the key observations from this review is that AI-based triage models consistently outperformed conventional triage methods, including emergency severity index (ESI) scores, prehospital stroke scales, and dispatcher-based OHCA recognition. Traditional risk stratification approaches often rely on heuristic or rule-based decision trees, which, although effective, are prone to human error, inconsistencies, and cognitive bias. By contrast, AI models leverage large datasets, pattern recognition, and predictive modeling to enhance decision-making accuracy.

For instance, a study demonstrated that machine learning models significantly outperformed conventional stroke assessment tools such as the Cincinnati Prehospital Stroke Scale (CPSS), achieving an AUC of up to 0.98.[21] Similarly, AI-assisted OHCA detection models outperformed manual dispatcher-based recognition, highlighting AI's potential to enhance emergency call center operations by automating real-time triage classification.[16,17,28] These findings align with existing literature, which has increasingly pointed to AI as a valuable adjunct in emergency medicine, capable of improving triage accuracy, reducing unnecessary emergency department (ED) visits, and optimizing EMS resource allocation.[29]

Despite these advancements, several studies noted challenges in AI adoption due to human factors, including dispatcher compliance, training, and interpretability issues. For example, Blomberg et al.[25] reported that despite the high accuracy of AI-assisted OHCA recognition, dispatcher adherence to AI recommendations remained low, suggesting a reluctance to fully rely on AI-driven decision-making. This raises important questions regarding human-AI collaboration and the need for AI systems to be designed in a way that enhances, rather than replaces, human expertise.[30]

Challenges and barriers to AI implementation in EMS

While the findings suggest a strong case for AI integration into prehospital emergency medicine, several barriers and challenges remain that must be addressed before widespread adoption. These include: (i) Data availability and model generalizability : one of the primary limitations of AI in EMS is the availability of high-quality, standardized, and diverse datasets for model training. Many AI models in this review were trained on regional datasets, which may not generalize well to heterogeneous patient populations across different healthcare systems.[36] For instance, most studies in this review were conducted in high-resource settings such as the USA, Sweden, Japan, and South Korea, whereas low- and middle-income countries (LMICs) remain significantly underrepresented. Without large-scale, geographically diverse datasets, AI models risk bias and reduced generalizability, limiting their real-world applicability.[37]; (ii) Real-time AI deployment and processing speed : another major challenge is the integration of AI models into real-time EMS workflows. While AI models demonstrated high predictive accuracy in retrospective analyses, their ability to function in real-time, under high-pressure emergency conditions, remains largely untested.[38,39] Many EMS settings operate in fast-paced, resource-limited environments, where real-time data processing, interoperability with existing EMS systems, and seamless AI integration into dispatch protocols are crucial. The ability of AI to process emergency call transcripts, ECG data, or patient vitals instantly, and provide actionable recommendations without delay, is a critical area for future research.[17,24,29,40]; (iii) Ethical, legal, and regulatory challenges : the deployment of AI in prehospital emergency medicine raises several ethical and regulatory concerns. Issues such as AI transparency, accountability, and the potential for biased decision-making must be carefully evaluated. For example, if an AI-driven triage system misclassifies a patient's urgency level, leading to delayed or inappropriate care, determining liability becomes complex.[41] Moreover, regulatory approval processes for AI-driven EMS tools remain in their infancy, and there is currently no standardized framework for validating AI models in prehospital settings (iv) Human factors: trust, compliance, and AI interpretability : successful AI adoption in EMS requires trust and compliance from emergency responders, dispatchers, and clinicians. A key concern highlighted in several studies was low dispatcher compliance with AI recommendations, as observed.[21] This resistance may stem from a lack of understanding of AI decision-making processes and a reluctance to relinquish control to automated systems.[16] Therefore, future AI models must prioritize explainability and interpretability, ensuring that emergency personnel can understand and trust AI recommendations.

Future directions and research priorities

Based on the insights from this review, several key areas for future research and development emerge: (i) Prospective clinical trials: future studies should move beyond retrospective validation and assess AI performance in real-world, real-time EMS settings, testing how AI models function under real emergency conditions.[42] ;(ii) Cross-population model validation: AI models should be trained and validated on diverse, multi-national datasets to improve generalizability and reduce bias.[43-45]; (iii) Human-AI collaboration strategies: research should focus on improving the usability and acceptability of AI among emergency responders, ensuring that AI augments rather than replaces human decision-making.[46-48]; (iv) Regulatory framework development: standardized guidelines for AI validation, ethical deployment, and integration into existing EMS workflows must be established.[49,50]

CONCLUSION

Artificial intelligence is poised to transform prehospital emergency medicine by enhancing triage accuracy, expediting diagnosis, and optimizing EMS decision-making. This systematic review and meta-analysis demonstrates that AI models, particularly deep learning and machine learning algorithms, consistently outperform conventional triage methods, particularly in time-sensitive conditions such as stroke, STEMI, and OHCA. However, challenges in real-time deployment, data generalizability, dispatcher compliance, and regulatory integration must be addressed before widespread adoption. Future research should focus on prospective validation, human-AI collaboration, and ethical frameworks to ensure safe and effective AI implementation in EMS. With these advancements, AI has the potential to significantly improve emergency medical care and patient outcomes.

Conflict of Interest

None

Acknowledgement

The authors would like to express their gratitude to the Department of Emergency Medical Services at AlMaarefa University for their invaluable contribution and support throughout this study

Funding

No funding was received for this paper.

Authors' Contributions

AA: Supervised the data collection process, checked writing, approved methodology, edited manuscript, and supervised all steps, including final editing; RG: Researched literature, web-survey design, coordinated and monitored the data collection process with collaborators, wrote the first draft of the manuscript; SMM: Paper revision; BHA: Final editing, reviewing, and supervising the steps. All authors have read and approved the final manuscript

Data Availability Statement

All data supporting the findings of this study are included in the manuscript and supplementary materials. Any additional data or analysis files can be made available upon reasonable request to the corresponding author.

Declaration of AI Usage in Scientific Writing

AI was used for grammatical modification

REFERENCES

1. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Health J 2021;8:e188–94. https://doi.org/10.7861/fhj.2021-0095.
2. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023;23:689. https://doi.org/10.1186/s12909-023-04698-z.
3. Khalifa M, Albadawy M. Artificial intelligence for clinical prediction: exploring key domains and essential functions. Comput Methods Programs Biomed Update 2024;5:100148. https://doi.org/10.1016/j.cmpbup.2024.100148.
4. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. Bioengineering (Basel) 2024;11:337. https://doi.org/10.3390/bioengineering11040337.
5. Berikol GB, Kanbakan A, Ilhan B, Doğanay F. Mapping artificial intelligence models in emergency medicine: a scoping review on artificial intelligence performance in emergency care and education. Turk J Emerg Med 2025;25:67–91. https://doi.org/10.4103/tjem.tjem_45_25.
6. Piliuk K, Tomforde S. Artificial intelligence in emergency medicine: a systematic literature review. Int J Med Inform 2023;180:105274. https://doi.org/10.1016/j.ijmedinf.2023.105274.
7. Alruwaili A, Khorram-Manesh A, Ratnayake A, Robinson Y, Goniewicz K. The use of prehospital intensive care units in emergencies: a scoping review. Healthcare (Basel) 2023;11: 2892. https://doi.org/10.3390/healthcare11212892.
8. Cimino J, Braun C. Clinical research in prehospital care: current and future challenges. Clin Pract 2023;13:1266–85. https://doi.org/10.3390/clinpract13050114.
9. Johnson KB, Wei WQ, Weeraratne D, et al. Precision medicine, AI, and the future of personalized health care. Clin Transl Sci 2021;14:86–93. https://doi.org/10.1111/cts.12884.
10. Meyer LM, Stead S, Salge TO, Antons D. Artificial intelligence in acute care: a systematic review, conceptual synthesis, and research agenda. Technol Forecast Soc Change 2024;206:123568. https://doi.org/10.1016/j.techfore.2024.123568.
11. Hogg HDJ, Al-Zubaidy M, Talks J, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. J Med Internet Res 2023;25:e39742. https://doi.org/10.2196/39742.
12. Rane N, Choudhary S, Rane J. Acceptance of artificial intelligence: key factors, challenges, and implementation strategies. J Appl Artif Intell 2024;5:50–70. http://dx.doi.org/10.2139/ssrn.4842167.
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71.
14. Norman C, Leeflang M, Névéol A. Data extraction and synthesis in systematic reviews of diagnostic test accuracy: a corpus for automating and evaluating the process. AMIA Annu Symp Proc 2018;2018:817–26.
15. Nair AS. Publication bias—importance of studies with negative results. Indian J Anaesth 2019;63:505–7. https://doi.org/10.4103/ija.IJA_142_19.
16. Kang DY, Cho KJ, Kwon O, et al. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. Scand J Trauma Resusc Emerg Med 2020;28:17. https://doi.org/10.1186/s13049-020-0713-4 .
17. Chen KW, Wang YC, Liu MH, et al. Artificial intelligence-assisted remote detection of ST-elevation myocardial infarction using a mini 12-lead ECG device in prehospital ambulance care. Front Cardiovasc Med 2022;9:989–1002. https://doi.org/10.3389/fcvm.2022.1001982.
18. Spangler D, Hermansson T, Smekal D, Blomberg H. Validation of machine learning-based risk scores in the prehospital setting. PLoS One 2019;14:e0226518. https://doi.org/10.1371/journal.pone.0226518.
19. Mayampurath A, Parnianpour Z, Richards CT, et al. Improving prehospital stroke diagnosis using natural language processing of paramedic reports. Stroke 2021;52:2676–9. https://doi.org/10.1161/STROKEAHA.120.033580.
20. Hirano Y, Kondo Y, Sueyoshi K, Okamoto K, Tanaka H. Early outcome prediction for out-of-hospital cardiac arrest with initial shockable rhythm using machine learning models. Resuscitation 2021;158:49–56. https://doi.org/10.1016/j.resuscitation.2020.11.020.
21. Hayashi Y, Shimada T, Hattori N, et al. A prehospital diagnostic algorithm for strokes using machine learning: a prospective observational

study. Sci Rep 2021;11:20519. https://doi.org/10.1038/s41598-021-99828-2.
22. Forberg JL, Khoshnood A, Green M, et al. An artificial neural network to safely reduce the number of ambulance ECGs transmitted for physician assessment. Scand J Trauma Resusc Emerg Med 2012;20:8. https://doi.org/10.1186/1757-7241-20-8.
23. Chen Z, Zhang R, Xu F, et al. Novel prehospital prediction model of large vessel occlusion using artificial neural network. Front Aging Neurosci 2018;10:181. https://doi.org/10.3389/fnagi.2018.00181.
24. Byrsell F, Claesson A, Ringh M, et al. Machine learning can support dispatchers to better and faster recognize out-of-hospital cardiac arrest. Resuscitation 2021;162:218–26. https://doi.org/10.1016/j.resuscitation.2021.02.041.
25. Blomberg SN, Christensen HC, Lippert F, et al. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest. JAMA Netw Open 2021;4:e2032320 https://doi.org/10.1001/jamanetworkopen.2020.32320.
26. Anthony T, Mishra AK, Stassen W, Son J. Feasibility of using machine learning to classify calls to South African emergency dispatch centres according to prehospital diagnosis, by utilising caller descriptions of the incident. Healthcare (Basel) 2021;9:1107. doi: 10.3390/healthcare9091107.
27. Alonso E, Irusta U, Aramendi E, Daya MR. A machine learning framework for pulse detection during out-of-hospital cardiac arrest. IEEE Access 2020;8:161031–41. DOI:10.1109/ACCESS.2020.3021310.
28. Al-Zaiti S, Besomi L, Bouzid Z, et al. Machine learning-based prediction of acute coronary syndrome using prehospital 12-lead ECG. Nat Commun 2020;11:3966. https://doi.org/10.1038/s41467-020-17804-2.
29. Ajumobi O, Verdugo SR, Labus B, et al. Identification of non-fatal opioid overdose cases using 911 dispatch and prehospital records. Prehosp Emerg Care 2022;26:818–28. https://doi.org/10.1080/10903127.2021.1981505.
30. Tyler S, Olis M, Aust N, et al. Use of artificial intelligence in triage in hospital emergency departments: a scoping review. Cureus 2024;16:e59906. https://doi.org/10.7759/cureus.59906. Erratum in: Cureus 2025 ;17:c315. doi: 10.7759/cureus.c315.
31. Masoumian Hosseini M, Masoumian Hosseini ST, Qayumi K, Ahmady S, Koohestani HR. Aspects of running artificial intelligence in emergency care: a scoping review. Arch Acad Emerg Med 2023;11:e38. https://doi.org/10.22037/aaem.v11i1.1974.
32. Mariani A, Spaccarotella CAM, Rea FS, et al. Artificial intelligence and its role in diagnosis and prediction of adverse events in acute coronary syndrome. Life (Basel) 2025;15:515. doi: 10.3390/life15040515.
33. Auricchio A, Scquizzato T, Ravenda F, et al. Spatio-temporal distribution and prediction of major acute cardiovascular events. Resuscitation Plus 2024;20:100810. https://doi.org/10.1016/j.resplu.2024.100810 .
34. Chee ML, Huang H, Mazzochi K, et al. Artificial intelligence and machine learning in prehospital emergency care: a scoping review. iScience 2023;26:107407. https://doi.org/10.1016/j.isci.2023.107407.
35. De Simone B, Deeken G, Catena F. Balancing ethics and innovation: can artificial intelligence safely transform emergency surgery? J Clin Med 2025;14:3111. doi: 10.3390/jcm14093111.
36. Chenais G, Lagarde E, Gil-Jardiné C. Artificial intelligence in emergency medicine: current applications and challenges. J Med Internet Res 2023;25:e40031. https://doi.org/10.2196/40031.
37. Akinnagbe O. Human–AI collaboration: enhancing productivity and decision-making. Int J Educ Manag Technol 2024;2:387–417.doi : 10.58578/ijemt.v2i3.4209.
38. Arora A, Alderman JE, Palmer J, et al. The value of standards for health datasets in AI-based applications. Nat Med 2023;29:2929–38. https://doi.org/10.1038/s41591-023-02608-w.
39. Birdi S, Rabet R, Durant S, et al. Bias in machine learning applications to address non-communicable diseases at a population level: a scoping review. BMC Public Health 2024;24:3599. https://doi.org/10.1186/s12889-024-21081-9.
40. Conforti R. Informatics in emergency medicine: a literature review. Emerg Care Med 2025;2:2. https://doi.org/10.3390/ecm2010002.
41. Hill P, Lederman J, Jonsson D, Bolin P, Vicente V. Understanding EMS response times: a machine learning-based analysis. BMC Med Inform Decis Mak 2025;25:143. doi: 10.1186/s12911-025-02975-z.
42. Ahun E, Demir A, Yiğit Y, et al. Perceptions of emergency medicine practitioners about AI in triage management during the pandemic: a national survey. Front Public Health 2023;11:120123. https://doi.org/10.3389/fpubh.2023.1285390.
43. Moosavi A, Huang S, Vahabi M, et al. Prospective human validation of AI interventions in cardiology: a scoping review. JACC Adv 2024;3:101202. https://doi.org/10.1016/j.jacadv.2024.101202.
44. Bradshaw TJ, Huemann Z, Hu J, Rahmim A. A guide to cross-validation for artificial intelligence

in medical imaging. Radiol Artif Intell 2023;5:e220232. doi: 10.1148/ryai.220232.
45. Palaniappan K, Lin EYT, Vogel S. Global regulatory frameworks for the use of artificial intelligence in healthcare. Healthcare (Basel) 2024;12:562. https://doi.org/10.3390/healthcare12050562.
46. Berretta S, Tausch A, Ontrup G, Gilles B, Peifer C, Kluge A. Defining human–AI teaming the human-centered way: a scoping review and network analysis. Front Artif Intell. 2023;6:1250725. https://doi.org/10.3389/frai.2023.1250725.
47. Zarei R, Downs MC, Torgerson L. Artificial intelligence in prehospital emergency care: Advancing triage and destination decisions for time-critical conditions. Cureus 2025;17:e91542. http://dx.doi.org/10.7759/cureus.91542.
48. Machado C, Hessel S. Artificial intelligence in the prehospital setting – potentials, challenges, and practice-relevant fields of application in emergency medical services. BMC Artif Intell 2026;2:12. https://doi.org/10.1186/s44398-026-00027-8.
49. Elfahim O, Edjinedja KL, Cossus J, Youssfi M, Barakat O, Desmettre T. A systematic literature review of artificial intelligence in prehospital emergency care. Big Data Cogn Comput 2025;9:219. http://dx.doi.org/10.3390/bdcc9090219.
50. Reddy S. Global harmonization of artificial intelligence-enabled software as a medical device regulation: addressing challenges and unifying standards. Mayo Clin Proc Digit Health 2024;3:100191. doi: 10.1016/j.mcpdig.2024.100191.