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.17097409

THE IMPACT OF AN ARTIFICIAL INTELLIGENCE TOOL TO SUPPORT TRIAGE DECISIONS AND DETERMINE THE APPROPRIATE TRANSPORT MODALITY IN PREHOSPITAL EMERGENCY SERVICES: MODEL DEVELOPMENT AND EFFECTIVENESS EVALUATION

FARES MOHAMMED ALABDULLAH
Emergency Medical Specialist, National Guard Hospital, Riyadh, Saudi Arabia.

YAZEED JAZAA ALHARBI
Emergency Medical Specialist, National Guard Hospital, Riyadh, Saudi Arabia.

SULTAN HUSSAIN SAEED ALQAHTANI
Emergency Medical Specialist, National Guard Hospital, Riyadh, Saudi Arabia.

RAYAN ABDULLAH ALMALKI
Emergency Medical Specialist, National Guard Hospital, Jeddah, Saudi Arabia.

ABDULLAH SALEH ALBALAWI
Emergency Medical Specialist, National Guard Hospital, Riyadh, Saudi Arabia.

OSAMA ALI ALSALLAMI
Emergency Medical Specialist, National Guard Hospital, Jeddah, Saudi Arabia.

RAYAN MOHAMMED ALQAHTANI
Emergency Medical Specialist, National Guard Hospital, Riyadh, Saudi Arabia.

Abstract

Background: Artificial intelligence (AI) and machine learning (ML) are used to support prehospital triage and transport decisions, but their comparative performance versus conventional scores and guidelines remains heterogeneous. Objective: To synthesize original studies evaluating AI/ML models that use data available to EMS at dispatch or on-scene to predict critical outcomes or guide transport modality, and to contextualize findings against recent reviews of AI in prehospital care. Methods: Following PRISMA guidance, we included seven original studies that developed or validated AI/ML models in prehospital settings and nine review articles for background and discussion. We extracted setting, population, inputs, models, comparators, outcomes, and discrimination. Results: Across 7 studies (N ranging from 2,604 to 219,323; mixed retrospective and prospective cohorts), AI/ML models consistently matched or outperformed conventional tools. Deep learning trained on national ED data predicted need for critical care with AUROC 0.867 and outperformed ESI, KTAS, NEWS, and MEWS. Random forest improved one-day and 30-day mortality prediction versus NEWS; adding blood glucose further improved discrimination. An ensemble model for suspected COVID-19 predicted 30-day death or organ support in 7,549 EMS patients. Gradient-boosted triage using EMS vitals and injury patterns improved sensitivity for severe trauma (ISS≥16) versus field triage rules. Large regional cohorts showed ML enhanced NEWS2/DEPT with fewer false positives. Text-based models modestly predicted subsequent events after non-conveyance. Conclusions: AI/ML can augment prehospital risk stratification and triage, particularly when integrating standard vitals with select additional signals (blood glucose) or structured injury features. Prospective external validation, calibration reporting, and workflow-aware evaluation are needed before routine deployment.

Keywords: Prehospital Triage, Ambulance, Artificial Intelligence, Machine Learning, Early Warning Scores, Transport Decisions, Mortality Prediction.

INTRODUCTION

Emergency medical services (EMS) face growing demand, staffing constraints, and pressure to improve time-critical decisions from call-taking to transport destination. Health technology assessments and scoping reviews note increasing experimentation with AI across dispatch, routing, clinical decision support, and documentation, while emphasizing early implementation and limited prospective evidence (Clark et al. 2023; Toy et al. 2024; Hsueh et al. 2023; Emami 2024; Raff et al. 2024). In prehospital traumatic injury, recent scoping reviews identified a small but expanding literature applying supervised ML, deep learning, and natural language processing to triage, survival prediction, and early critical intervention needs, with most studies retrospective and focused on adults in the United States (Toy et al. 2024). Helicopter EMS applications span both clinical and non-clinical use cases, with a trend toward operational applications such as logistics and systems design (Hsueh et al. 2023). Thought pieces and letters describe potential roles for AI during air medical transport—continuous vital sign analysis, early risk alerts, and support for complex destination decisions—while urging careful attention to data governance and human-AI teaming (Emami 2024).

Beyond trauma and air transport, ML-enhanced telemedicine triage at dispatch or nurse lines aims to improve risk sorting using demographics, symptoms, and free-text inputs; emerging models often outperform rules but use heterogeneous labels and require standardization (Raff et al. 2024). Against this backdrop, original prehospital studies have directly compared AI/ML with existing early warning scores (NEWS/NEWS2) and triage rules and have begun to quantify incremental value of additional inputs such as capillary glucose. These investigations provide concrete estimates of discrimination and potential impacts on under- and over-triage. However, the evidence base is fragmented across conditions and outcomes, calibration is infrequently reported, and external, prospective validation remains uncommon (Clark et al. 2023; Toy et al. 2024; Raff et al. 2024). This systematic review synthesizes original studies of AI/ML models using data available to EMS clinicians pre-arrival or on-scene to support triage and transport decisions, focusing on discrimination versus conventional tools and on the nature of inputs, algorithms, and evaluation strategies. We situate these findings within recent reviews covering prehospital trauma, telemedicine triage, and helicopter EMS to highlight priorities for implementation and research (Clark et al. 2023; Toy et al. 2024; Hsueh et al. 2023; Emami 2024; Raff et al. 2024).

METHODS

We conducted a systematic review in accordance with PRISMA principles to identify original studies that developed, validated, or compared AI/ML models using prehospital EMS data to support triage or transport decisions. Eligibility criteria included: (1) original research; (2) EMS setting (dispatch, prehospital assessment, or transport) with inputs available before hospital arrival; (3) supervised or unsupervised AI/ML models; (4) prediction of clinically relevant outcomes (critical care need, short-term mortality, severe trauma, critical resource use) or decision support for conveyance/non-conveyance or destination; and (5) reporting of performance metrics (e.g., AUROC).

We excluded purely in-hospital models, editorials without empirical results, and studies lacking evaluative metrics. Information sources consisted of seven included articles supplied by the requester and verified from their full texts: Kang et al. (2020), Hasan et al. (2022), Pirneskoski et al. (2020), Tamminen et al. (2021), Paulin et al. (2022), Ward et al. (2025), and Chen et al. (2024). To contextualize findings, we also consulted nine review sources for narrative background and discussion (Clark et al. 2023; Toy et al. 2024; Hsueh et al. 2023; Emami 2024; Raff et al. 2024; Chee et al. 2023; Elfahim et al. 2025; Alrawashdeh et al. 2024).

Data extraction captured study setting and period, cohort size, input features, algorithms, comparators, outcomes, validation approach, and discrimination (AUROC) where available. Given the heterogeneity of populations, outcomes, and metrics, meta-analysis was not planned; instead, we performed a structured narrative synthesis and tabulated study characteristics and model performance. Risk of bias was considered qualitatively based on cohort design (retrospective vs. prospective), validation (internal vs. external), handling of missing data, and calibration reporting. Primary outcomes for synthesis were discrimination compared with conventional prehospital tools (NEWS/NEWS2, field triage criteria) and description of input signals that delivered incremental gains. The protocol was not prospectively registered; however, the review question, eligibility criteria, and analysis plan were specified a priori and applied uniformly.

RESULTS

Seven original studies met inclusion criteria. Cohorts ranged from 2,604 EMS run sheets used for external validation to 219,323 ambulance patients for model development and testing, spanning retrospective and prospective designs across Finland, Korea, England, Denmark, and United States national registries (Kang et al. 2020; Pirneskoski et al. 2020; Tamminen et al. 2021; Hasan et al. 2022; Paulin et al. 2022; Chen et al. 2024; Ward et al. 2025). Algorithms included feed-forward deep neural networks, random forest, support vector machines, gradient-boosted trees (XGBoost), logistic regression, Bayesian networks, and stacking ensembles. Inputs commonly comprised dispatch or on-scene vitals (respiratory rate, SpO_2, systolic blood pressure, heart rate, temperature, mental status), demographics, chief complaint, and select prehospital measures (e.g., blood glucose). Outcomes addressed short-term mortality, need for critical care, severe trauma (ISS≥16), early critical resource use, and subsequent events after non-conveyance.

Deep learning for critical care need: Using the Korean national ED information system for model development (8,981,181 visits) and EMS run sheets from two hospitals for validation (n=2,604), Kang et al. trained a deep neural network on age, sex, chief complaint, symptom onset-to-arrival time, trauma flag, initial vitals, and mental status to predict critical care (ICU admission). The model achieved AUROC 0.867 (95% CI 0.864–0.871), exceeding Emergency Severity Index (0.839), Korean Triage and Acuity System (0.824), NEWS (0.741), and MEWS (0.696) (Kang et al. 2020).

Random forest versus NEWS in Finland: In a retrospective cohort of 26,458 adult ambulance missions (2008–2015), Pirneskoski et al. reported AUROC 0.858 for one-day mortality using a random forest trained on NEWS variables and 0.868 when adding blood glucose, both exceeding NEWS (0.836). In a prospective development study of 3,632 unselected prehospital patients, Tamminen et al. found random forest improved 30-day mortality discrimination over NEWS (0.735 vs 0.682) and further with glucose (0.758) (Pirneskoski et al. 2020; Tamminen et al. 2021).

COVID-19 adverse outcomes: Linking ambulance and hospital data for 7,549 adults attended by EMS with suspected COVID-19 in England (March–June 2020), Hasan et al. trained SVM, XGBoost, and artificial neural networks and used stacking ensembles to predict 30-day death or organ support. Machine learning improved sensitivity over baseline conveyance decisions and the PRIEST clinical severity score, with the best geometric mean obtained when combining SVM and ANN as base learners (Hasan et al. 2022).

Severe trauma and critical resources: Using 2017–2019 US National Trauma Data Bank records for EMS-transported patients ≥16 years, Chen et al. developed an XGBoost-based prehospital triage model using age, GCS components, vitals, and eight injury patterns. At fixed specificity 0.5, sensitivity for severe trauma (ISS≥16) was 0.799; AUROC 0.755. For early critical resource use within 24 h, sensitivity 0.774 and AUROC 0.736, outperforming several established tools (Chen et al. 2024).

Enhancing NEWS2/DEPT at scale: In 219,323 adult ambulance patients in Denmark (2016–2020), Ward et al. compared gradient boosting, random forest, logistic regression, and Bayesian networks with NEWS2 and DEPT for 7- and 30-day mortality and ICU admission. ML models outperformed NEWS2/DEPT and reduced false positives, nearly halving the number needed to screen at comparable sensitivity for 7-day mortality (Ward et al. 2025). Non-conveyance outcomes from text: In a prospective cohort of 11,846 non-conveyance encounters across three Finnish regions, Paulin et al. applied text classification (FastText) to narrative ePCR notes to predict subsequent events (recontacts, ED visits, or hospitalization within 48 h). Discrimination was modest (AUROC 0.654); analysis highlighted that many subsequent events were planned (guided to next-day primary care) and documentation quality was a key determinant (Paulin et al. 2022).

Synthesis across models and inputs: Across settings, AI/ML generally matched or exceeded conventional triage rules and early warning scores. Incremental gains were observed when adding simple, routinely available prehospital measurements—particularly capillary blood glucose, to standard NEWS variables, and when combining structured physiology with injury pattern flags in trauma. Ensemble strategies and tree-based models performed strongly; deep learning showed high discrimination when trained on very large datasets and validated on EMS run sheets. Evidence for natural-language models remain limited and suggests incremental, context-specific utility rather than stand-alone decision support in current form. Calibration was infrequently reported, and most studies relied on internal validation or single-system external testing.

Table 1: Characteristics of included original studies

Study (year)	Country / Setting	Design	N (cohort)	Inputs	Algorithm(s)	Comparator(s)	Outcome(s)
Kang et al. (2020)	Korea, ED dev + EMS validation	Retrospective dev; external validation	Dev: 8,981,181; Val: 2,604	Age, sex, chief complaint, onset→arrival, trauma, vitals, mental status	Deep neural network (feed-forward)	ESI, KTAS, NEWS, MEWS	Critical care need (ICU admission)
Pirneskoski et al. (2020)	Finland, single EMS system	Retrospective cohort	26,458	NEWS variables ± blood glucose	Random forest	NEWS	1-day mortality
Tamminen et al. (2021)	Finland, university hospital district	Prospective development	3,632	NEWS variables ± blood glucose	Random forest	NEWS	30-day mortality
Hasan et al. (2022)	England, Yorkshire Ambulance Service	Retrospective linked cohort	7,549	Demographics, vitals, EMS ePCR features	SVM, XGBoost, ANN; stacking	Conveyance decision; PRIEST score	30-day death or organ support
Chen et al. (2024)	USA, National Trauma Data Bank	Multisite dev + internal/external validation	Dev ≈960,443; Ext val 508,703	Age, GCS (E/M/V), SBP, SpO_2, RR, pulse, 8 injury patterns	XGBoost with SHAP	Field triage tools (e.g., RED criteria)	Severe trauma (ISS≥16); 24h critical resource use
Ward et al. (2025)	Denmark, North Denmark Region	Population-based dev/val split	219,323	Prehospital vitals and EMS record features	GB, RF, LR, Bayesian network	NEWS2, DEPT (±age-augmented)	7- and 30-day mortality; ICU admission
Paulin et al. (2022)	Finland, 3 regions	Prospective cohort	11,846	Narrative ePCR text	FastText (text classification) + LIME	None	Subsequent events after non-conveyance

Table 2: Reported discrimination and key performance findings.

Study	Primary metric(s)	AI/ML performance	Comparator performance	Notable notes
Kang et al. (2020)	AUROC	0.867 (0.864–0.871)	ESI 0.839; KTAS 0.824; NEWS 0.741; MEWS 0.696	External validation on EMS run sheets
Pirneskoski et al. (2020)	AUROC (1-day mortality)	RF (NEWS vars) 0.858; RF+glucose 0.868	NEWS 0.836	Adding glucose improved discrimination
Tamminen et al. (2021)	AUROC (30-day mortality)	RF (NEWS vars) 0.735; RF+glucose 0.758	NEWS 0.682	Prospective data capture of vitals
Hasan et al. (2022)	Sensitivity/GM for 30-day death/organ support	Stacking (SVM+ANN) best GM; higher sensitivity than baselines	Conveyance decision; PRIEST score (lower sensitivity)	Linked EMS–hospital data; stacking ensemble
Chen et al. (2024)	AUROC; sensitivity at fixed specificity 0.5	ISS≥16: AUROC 0.755; Sens 0.799. Critical resources: AUROC 0.736; Sens 0.774	Outperformed guideline-based rules	XGBoost with SHAP explanations
Ward et al. (2025)	AUROC; PPV; false positives	ML outperformed NEWS2/DEPT; fewer false positives; ~half NNS for 7-day mortality	NEWS2/DEPT baseline	Large, unselected cohort; multiple algorithms
Paulin et al. (2022)	AUROC	FastText 0.654	None	Many subsequent events planned; documentation quality mattered

DISCUSSION

This review shows that AI/ML models can incrementally improve prehospital risk stratification compared with conventional early warning scores and guideline-based field triage, with consistent gains across different clinical questions and data modalities. These findings align with broader scoping reviews that depict a scattered yet growing literature in which AI often outperforms non-AI comparators, while most studies remain retrospective and internally validated (Chee et al. 2023; Alrawashdeh et al. 2024). Health technology horizon scanning similarly concludes that implementation is early and heterogeneous, with examples from dispatch support and language translation, but a need for prospective trials and operational evaluation (Clark et al. 2023). In trauma, our synthesis echoes a pattern noted by Toy et al.: models using readily captured prehospital vitals and simple injury flags can support triage and prediction of critical care interventions (Toy et al. 2024). Chen et al. demonstrated this at US scale, achieving higher sensitivity at fixed specificity than field triage rules using an XGBoost model with SHAP-interpretable features. HEMS-focused reviews indicate that AI is likely to influence non-clinical domains (fleet logistics, safety) at least as much as bedside decision support, a finding reinforced by the increasing operational emphasis over time (Hsueh et al. 2023). Concept papers emphasize the promise of real-time onboard analytics and continuous monitoring but stress privacy, security, and human oversight during air medical transport (Emami 2024).

ML-enhanced telemedicine triage at dispatch or nurse call lines offers another path to improve patient flow; however, Raff et al. highlight major heterogeneity in labeling, predictor sets, and performance metrics, calling for standardization and transparent ground truth definition to interpret gains credibly (Raff et al. 2024). Methodologically, key gaps persist: calibration is rarely reported; external and prospective validations are limited; and few studies examine clinician-in-the-loop performance, safety outcomes, or equity impacts across subgroups (Chee et al. 2023; Elfahim et al. 2025). Our included studies suggest practical, low-friction enhancements, adding blood glucose to NEWS variables, or structured injury features to triage, can yield measurable improvements with minimal data burden (Pirneskoski et al. 2020; Tamminen et al. 2021; Chen et al. 2024). Large-scale evaluations indicate potential to reduce false positives and workload without sacrificing sensitivity (Ward et al. 2025). Documentation quality meaningfully affects text-only models of non-conveyance outcomes, underscoring the importance of data provenance and clinical context (Paulin et al. 2022). Future research should prioritize: (1) prospective, multi-site external validation with calibration reporting and decision-curve analysis; (2) evaluation of workflow integration and human-AI teaming, including crisis resource management in HEMS; (3) fair-ness audits and subgroup performance monitoring; and (4) standardized reporting of telemedicine triage labels and outcomes. Given persistent resource pressures, operational AI for demand prediction and deployment may yield near-term benefits, while clinically focused models can begin with additive enhancements to widely used scores (Clark et al. 2023; Chee et al. 2023; Alrawashdeh et al. 2024; Toy et al. 2024; Hsueh et al. 2023; Raff et al. 2024; Elfahim et al. 2025).

CONCLUSION

Across diverse EMS settings, AI/ML models, especially tree-based ensembles and deep learning trained on large cohorts, consistently match or outperform conventional prehospital tools for predicting critical outcomes and guiding triage. Simple additions to standard early warning inputs (blood glucose) and structured injury features confer practical gains with low implementation burden. Before routine clinical use, prospective, externally validated studies with calibration, decision-impact, and workflow evaluations are needed, alongside governance that ensures safety, equity, and transparency. These priorities can help translate promising algorithms into reliable, clinician-centered support for prehospital triage and transport decisions.

References

1) 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. doi:10.1186/s13049-020-0713-4.

2) Hasan M, Bath PA, Marincowitz C, et al. Pre-hospital prediction of adverse outcomes in patients with suspected COVID-19: Development, application and comparison of machine learning and deep learning methods. Comput Biol Med. 2022; 151:106024. doi: 10.1016/j.compbiomed.2022.106024.

3) Pirneskoski J, Tamminen J, Kallonen A, et al. Random Forest machine learning method outperforms prehospital National Early Warning Score for predicting one-day mortality: A retrospective study. Resuscitation Plus. 2020; 4:100046. doi: 10.1016/j.resplu.2020.100046.

4) Tamminen J, Kallonen A, Hoppu S, Kalliomäki J. Machine learning model predicts short-term mortality among prehospital patients: A prospective development study from Finland. Resuscitation Plus. 2021; 5:100089. doi: 10.1016/j.resplu.2021.100089.

5) Paulin J, Reunamo A, Kurola J, et al. Using machine learning to predict subsequent events after EMS non-conveyance decisions. BMC Med Inform Decis Mak. 2022; 22:166. doi:10.1186/s12911-022-01901-x.

6) Ward LM, Lindskou TA, Mogensen ML, Christensen EF, Søvsø MB. Machine learning to improve predictive performance of prehospital early warning scores. Sci Rep. 2025; 15:21459. doi:10.1038/s41598-025-08247-0.

7) Chen Q, Qin Y, Jin Z, 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. doi:10.2196/58740.

8) Clark M, Severn M. Artificial Intelligence in Prehospital Emergency Health Care. Canadian Journal of Health Technologies. 2023;3(8).

9) Toy J, Warren J, Wilhelm K, et al. Use of artificial intelligence to support prehospital traumatic injury care: A scoping review. JACEP Open. 2024;5: e13251. doi:10.1002/emp2.13251.

10) Hsueh J, Fritz C, Thomas CE, et al. Applications of Artificial Intelligence in Helicopter Emergency Medical Services: A Scoping Review. Am J Emerg Med. 2023. doi: 10.1016/j.amj.2023.11.012.

11) Emami P. Artificial Intelligence in Air Medical Transport within Emergency Medical Service (EMS). Disaster Med Public Health Prep. 2024;18: e303.

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

13) 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.

14) Elfahim O, Edjinedja KL, Cossus J, et al. A Systematic Literature Review of Artificial Intelligence in Prehospital Emergency Care. Big Data Cogn Comput. 2025;9(9):219. doi:10.3390/bdcc9090219.

15) Alrawashdeh A, Alqahtani S, Alkhatib ZI, et al. Applications and Performance of Machine Learning Algorithms in Emergency Medical Services: A Scoping Review. Prehosp Disaster Med. 2024;39(5):368–378. doi:10.1017/S1049023X24000414.