

Artificial Intelligence in Emergency Trauma Care: A Preliminary Scoping Review

Christian Angelo I Ventura ¹, Edward E Denton², Jessica A David³

¹Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD USA; Department of Allied Health, Baltimore City Community College, Baltimore, MD, USA; ²Department of Emergency Medicine, University of Arkansas for Medical Sciences, Little Rock, AR USA; Fay W. Boozman College of Public Health, University of Arkansas for Medical Sciences, Little Rock, AR, USA; ³Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, USA

Correspondence: Christian Angelo I Ventura, Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health Baltimore, Baltimore, MD, USA, Tel +1 732 372-2141, Email cventura@bcc.edu; cventu12@jhmi.edu

Abstract: This study aimed to analyze the use of generative artificial intelligence in the emergency trauma care setting through a brief scoping review of literature published between 2014 and 2024. An exploration of the NCBI repository was performed using a search string of selected keywords that returned N=87 results; articles that met the inclusion criteria (n=28) were reviewed and analyzed. Heterogeneity sources were explored and identified by a significance threshold of $P < 0.10$ or an I^2 value exceeding 50%. If applicable, articles were categorized within three primary domains: triage, diagnostics, or treatment. Findings suggest that CNNs demonstrate strong diagnostic performance for diverse traumatic injuries, but generalized integration requires expanded prospective multi-center validation. Injury scoring models currently experience calibration gaps in mortality quantification and lesion localization that can undermine clinical utility by permitting false negatives. Triage predictive models now confront transparency, explainability, and healthcare ecosystem integration barriers limiting real-world translation. The most significant literature gap centers on treatment-oriented generative AI applications that provide real-time guidance for urgent trauma interventions rather than just analytical support.

Keywords: artificial intelligence, machine-learning, emergency medicine, traumatology

Introduction

Generative artificial intelligence (AI) refers to a rapidly advancing set of machine learning techniques that can synthesize realistic artifacts such as images, text, audio, and video when supplied with basic contextual inputs.¹ Unlike discriminative models that classify inputs into existing categories, generative models create novel outputs based on patterns learned from training data. Popular examples include generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models.² These technologies have demonstrated the ability to generate highly realistic synthetic data across various modalities. At this juncture, applications within time-sensitive and high-stakes environments such as emergency trauma care remain largely conceptual.

Trauma care epitomizes the need for quick yet accurate diagnostics, triage decisions, and clinical interventions where small errors can have catastrophic consequences. Generative AI offers untapped potential to enhance human capabilities in each of these domains. However, rigorous validation is required before generative AI can be safely entrusted to inform such high-stakes trauma decisions. There remain open questions about the clinical validity, safety, explainability, and real-world viability of these technologies.² Further research into protocols for testing generative AI validity, safety compliance, transparency measures, and clinician interface designs could help unlock trauma-care-focused applications. Striking the right human–AI balance could enable better trauma resource allocation, treatment, and reduce morbidity and mortality for patients with traumatic injuries. This brief scoping review aims to map a preliminary landscape of the current research frontiers and barriers to practical usage of generative AI in emergency trauma diagnostics, triage, and care, leveraging robust methodology akin to conventional systematic review recommendations.

Methods

Search Strategy

On 26 February 2024, a PubMed National Center for Biotechnology Information (NCBI) repository search was conducted for articles published between 2014 and 2024 using the following search string: (“Generative Artificial Intelligence” [Title/Abstract] OR “Generative AI” [Title/Abstract] OR “Deep Learning” [Title/Abstract] OR “Neural Networks” [Title/Abstract]) AND (“Emergency Medicine” [MeSH Terms] OR “Emergency Care” [Title/Abstract] OR “Trauma Care” [Title/Abstract] OR “Emergency Department” [Title/Abstract]) AND (“Use” [Title/Abstract] OR “Application” [Title/Abstract] OR “Implications” [Title/Abstract]). The returned English results were uploaded to Rayyan.ai for comprehensive abstract and full-text review.³

Data Extraction

Duplicate results were assessed for and removed. Editorials, commentaries, and non-peer reviewed manuscripts were excluded. Two investigators independently reviewed abstracts to identify articles eligible for full-text review. The investigators then independently reviewed full-text articles to identify studies that met the PICOS-guided inclusion criteria.⁴ Included studies focused specifically on applications of generative artificial intelligence (AI) within emergency trauma care contexts, whether examining clinical outcomes, workflow efficiency gains, or decision-making improvements. Only original, English-language, peer-reviewed articles within the last decade were incorporated. Studies were excluded if there were concerns regarding methodical quality or integrity of the data as per the discretion of the two investigators and a third consultant. SIGN appraisal tools were used to exclude retrospective and cohort-based studies that did not meet an acceptable level of evidence for inclusion in the review.⁵ Conflicts were resolved through discussion and by mediation from a third consultant when necessary. Study methods were consistent with PRISMA recommendations, and although PROSPERO registration was not sought, the work remained faithful to conventional review standards.⁶

Analysis

Statistical analysis was conducted using Stata/BE software, focusing on the aggregate prevalence of outcomes.⁷ Studies that did not report sufficient statistical information did not undergo quantitative analysis. Heterogeneity among studies was assessed through the I^2 statistic and Chi-squared test. Significant heterogeneity was defined as a p-value of less than 0.10 or an I^2 value exceeding 50%, in accordance with the guidelines proposed in the Cochrane Handbook for Systematic Reviews of Interventions.⁸ In instances of significant heterogeneity, a fixed effects model was utilized for data analysis. Conversely, in the absence of significant heterogeneity, a random effects model was adopted to accommodate the variability among the studies included. To elucidate potential sources of heterogeneity, studies were categorized into three distinct domains based on their primary focus: triage, diagnostics, or treatment when applicable.

Ethical Considerations

Because the work did not involve the use of human research subjects, it did not require approval or review by an institutional review board or bioethics committee.

Results

The NCBI repository returned N=87 results, with n=28 articles utilized for analysis and inclusion in this study. Results were excluded if they did not satisfy the inclusion criteria. [Figure 1](#) depicts an overview of the exclusion schema. AI identified n=6 duplicative results, n=45 results were excluded due to irrelevance with respect to the area of investigation, and n=2 studies were excluded after investigators performed full-text reviews and found studies to be of unsatisfactory evidence levels in accordance with SIGN appraisal guidelines. [Table 1](#) depicts the characteristics of studies selected for inclusion.

Of the included articles, the majority were US-based (n=8) and retrospective cohort studies (n=8). Primary taxonomy schema revealed the following prevalence of data: n=21 diagnostics, n=1 treatment, n=6 triage. Deep learning

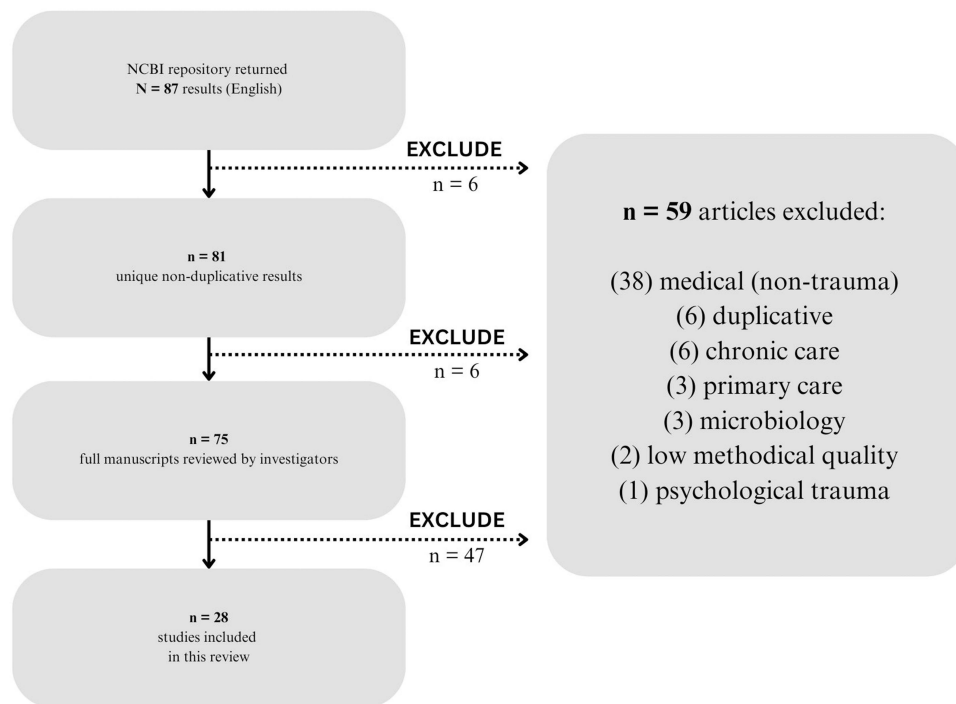


Figure 1 Study selection flow chart and overview of exclusion schema.

demonstrates high diagnostic accuracy for various traumatic injuries identifiable on medical imaging. Convolutional neural networks attained over 90% sensitivity and specificity for detecting solid organ abdominal trauma like spleen, liver, and kidney lesions on CT scans.⁹ Additional models achieved up to 97% accuracy in diagnosing distal radius fractures on radiographs,¹¹ 98% sensitivity for hip fractures on pelvic X-rays,¹⁴ and AUC exceeding 0.80 for intracranial hemorrhage detection on head CT scans.¹⁹ Deep learning also shows precision in localizing traumatic findings, with activation mapping techniques precisely pinpointing 95.9% of hip fracture lesions¹⁴ and models consistently highlighting displaced ribs on chest CTs.³²

Beyond binary classification, deep learning shows an aptitude for real-time injury severity quantification to guide downstream care decisions. One model leveraging trusted outcomes of length of stay, discharge disposition, and mortality for 176 potential injuries demonstrated comparable or superior performance to the Injury Severity Score.¹³ However, the model requires additional assessment to address mortality overestimation. Workflow and efficiency dividends represent another promising application area. Algorithmic fracture nominations focus radiologists' attention to suspicious areas, improving detection rates by 65.7% and reducing reading times.²¹ Minimizing false positives and negatives remains an open challenge.^{19,32}

Cardiac and respiratory deterioration prediction represents an emerging area harnessing deep learning's pattern recognition capabilities. An algorithm analyzed echocardiogram videos for abnormal cardiac function with AUC exceeding 0.90.¹⁸ Another model predicted cardiac arrest and respiratory failure 1–6 hours in advance using only vital signs and history, surpassing traditional early warning scores.³⁵ Considerations span collecting diverse and standardized validation data²³ to enhancing localization to avoid false negatives¹⁹ and addressing ethical concerns related to black-box recommendations and over-reliance.¹⁷ Partnership with emergency and trauma radiologists is critical for translating technical potential into clinical practice improvements.³³

Studies in the triage domain demonstrated that advanced machine learning approaches, especially deep neural network architectures, attain state-of-the-art performance across important emergency department triage outcomes. For critical care prediction, neural networks achieved a median C-statistic of 0.871, significantly outperforming conventional triage methods like vital sign cutoffs (0.832) and the Emergency Severity Index (0.809) scale.³⁰ This indicates superior discriminative ability to stratify patients likely to require critical care interventions. Deep learning methods combining

Table 1 Selected Characteristics of Studies Identified for Inclusion

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[9]	Cheng, Lin, Hsu, Chen, Huang, Hsieh, Fu, Chung, Liao	Deep Learning for Automated Detection and Localization of Traumatic Abdominal Solid Organ Injuries on CT Scans.	Retrospective Cohort	Taiwan	<ul style="list-style-type: none"> • A DLM was specifically developed to detect solid organ injuries in patients who have undergone CT scans for blunt abdominal trauma, aiming to assist in the rapid identification of life-threatening injuries. • The model was trained on 1302 CT scans (87% of the total) and tested on 194 scans (13%), taken from patients at a single trauma center between 2008 and 2017. • The DLM showed high accuracy and specificity in detecting injuries to the spleen, liver, and kidneys, indicating its potential as a tool to support clinicians in making quicker decisions regarding trauma care. 	Diagnostics
[10]	Michel, Manns, Boudersa, Jaubert, Dupic, Vivien, Burgun, Campeotto, Tsopra	Clinical decision support system in emergency telephone triage: A scoping review of technical design, implementation and evaluation.	Scoping Review	US	<ul style="list-style-type: none"> • Identified 19 CDSS for emergency telephone triage, highlighting a mix of knowledge-based[9] and data-driven[7] systems, primarily aimed at assisting nurses or non-medical staff with patient orientation and severity assessment. • Eleven CDSS were implemented in real-world settings, but only three were integrated with Electronic Health Records (EHR), indicating a gap in leveraging existing health data. • The review underscores the necessity for a hybrid, user-tailored CDSS that can interface with oral, video, and digital data, emphasizing iterative evaluation of CDSS's intrinsic characteristics and their clinical impact throughout the IT lifecycle. 	Triage

[11]	Russe, Rebmann, Tran, Kellner, Reisert, Bamberg, Kotter, Kim	AI-based X-ray fracture analysis of the distal radius: accuracy between representative classification, detection and segmentation deep learning models for clinical practice.	Comparative Observational	Germany	<ul style="list-style-type: none"> Evaluated AI models for detecting distal radius fractures in radiographs, comparing custom-trained classification, detection, segmentation models, and a commercial solution, showing high accuracies (up to 0.97) in fracture detection. A total of 2856 radiographs were analyzed, with AI models achieving high-performance metrics, including Cohen's and Fleiss' kappa values indicating strong agreement in fracture detection across models. The findings suggest that the choice of an AI tool for fracture analysis depends on the specific requirements for automation, ranging from automated classification to AI-assisted reading or minimizing false negatives. 	Diagnostics
[12]	Piliuk, Tomforde	Artificial intelligence in emergency medicine. A systematic literature review.	Systematic Review	Germany	<ul style="list-style-type: none"> This study systematically categorizes and examines existing contributions in AI applications for emergency medicine, identifying obstacles stemming from healthcare regulations and the fragmented nature of published research. Findings reveal a predominance of specialized studies in diagnostics and triage, often utilizing disparate data sources and methodologies, highlighting the necessity for standardized approaches and end-to-end solutions integrating human-machine interaction for improved generalization. 	Diagnostics

(Continued)

Table I (Continued).

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[13]	Choi, Vendrow, Moor	Development and Validation of a Model to Quantify Injury Severity in Real Time.	Retrospective Cohort	US	<ul style="list-style-type: none"> Developed and validated the Length of Stay, Disposition, Mortality (LDM) Injury Index, a deep learning model that quantifies injury severity in real-time, using three outcomes: predicted hospital length of stay, probability of discharge to a facility, and probability of inpatient mortality. The model demonstrated comparable or better performance metrics (such as AUROC, recall, and specificity) in external validation compared to the traditional Injury Severity Score (ISS), particularly in predicting facility discharge and mortality, albeit with some overestimation in mortality predictions. The model, which uses 176 potential injuries for its predictions, showed excellent calibration for predicting facility discharge but requires further study to evaluate its effectiveness at scale. 	Diagnostics
[14]	Gao, Soh, Liu, Lim, Ting, Cheng, Wong, Liew, Oh, Tan, Venkataraman, Goh, Yan	Application of a deep learning algorithm in the detection of hip fractures.	Observational Diagnostic Accuracy/ Retrospective Cohort	Singapore	<ul style="list-style-type: none"> A deep convolutional neural network (DCNN) demonstrated high accuracy (91%) and sensitivity (98%) in detecting hip fractures on plain frontal pelvic radiographs (PXR), with a low false-negative rate (2%) and an area under the receiver operating characteristic curve (AUC) of 0.98. The visualization algorithm, gradient-weighted class activation mapping (Grad-CAM), achieved a 95.9% accuracy for identifying fracture lesions, confirming the validity of the model. The study concludes that DCNNs offer an efficient and economical means for hip fracture detection and localization on PXR, potentially aiding primary physicians in emergent screening and evaluation efforts without disrupting current clinical pathways. 	Diagnostics

[15]	Sax, Warton, Sofrygin, Mark, Ballard, Kene, Vinson, Reed	Automated analysis of unstructured clinical assessments improves emergency department triage performance: A retrospective deep learning analysis.	Retrospective Cohort	US	<ul style="list-style-type: none"> • Triage models utilizing both triage variables and clinical assessments significantly outperformed models based solely on triage variables, with AUC values of 0.87 (95% CI 0.87–0.87) for both hospitalization and fast-track eligibility prediction. • Hospitalization rate among ED patients was 12.7% (n = 673,659), while 37.0% (n = 1,966,615) were deemed fast-track eligible based on discharge criteria. • Models relying solely on triage variables showed lower predictive accuracy, with AUC values of 0.77 (95% CI 0.77–0.78) for hospitalization and 0.70 (95% CI 0.70–0.71) for fast-track eligibility. 	Triage
[16]	Ouyang, Chen, Tee, Lin, Kuo, Liao, Cheng, Liao	The Application of Design Thinking in Developing a Deep Learning Algorithm for Hip Fracture Detection.	Intervention Development	Taiwan	<ul style="list-style-type: none"> • Design thinking applied to DL algorithm development improved diagnostic accuracy, sensitivity, and specificity for detecting hip fractures from pelvic plain films in trauma care. • The study identified a specific clinical need regarding femoral fracture diagnosis, leading to the development of DL models with enhanced performance. • Integration of design thinking enhanced DL algorithm performance, ensuring user-centered solutions for trauma care in healthcare settings. 	Diagnostics

(Continued)

Table I (Continued).

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[17]	Hosseini, Hosseini, Qayumi, Ahmady, Koohestani	The Aspects of Running Artificial Intelligence in Emergency Care; a Scoping Review.	Scoping Review	Iran	<ul style="list-style-type: none"> The study reveals a growing trend of AI applications in emergency medicine, covering various areas such as machine learning algorithms, pre-hospital emergency management, triage, patient disposition, disease prediction, and emergency department management. Despite the potential benefits demonstrated by AI in improving patient outcomes through predictive modeling, ethical concerns regarding AI-based decision-making transparency are highlighted. The scoping review underscores the need for an ethical framework to address the lack of transparency in AI decision-making processes within emergency medicine contexts. 	Diagnostics
[18]	He, Dash, Duanmu, Tan, Ouyang, Zou	AI-ENABLED ASSESSMENT OF CARDIAC FUNCTION AND VIDEO QUALITY IN EMERGENCY DEPARTMENT POINT-OF-CARE ECHOCARDIOGRAMS.	Prospective Diagnostic	US	<ul style="list-style-type: none"> EchoNet-POCUS, a deep learning system, was developed to assist emergency physicians (EPs) in interpreting point-of-care ultrasound (POCUS) echocardiograms and reduce operator-to-operator variability. The system achieved high accuracy in predicting abnormal cardiac function with an area under the receiver operating characteristic curve (AUROC) of 0.92 and moderate accuracy in predicting video quality with an AUROC of 0.81. EchoNet-POCUS demonstrated feasibility for real-time application on bedside echocardiogram videos using standard hardware, as evidenced by a prospective pilot study. 	Diagnostics

[19]	Abrigo, Ko, Chen, Lai, Cheung, Chu, Yu	Artificial intelligence for detection of intracranial haemorrhage on head computed tomography scans: diagnostic accuracy in Hong Kong.	Prospective Diagnostic	Hong Kong	<ul style="list-style-type: none"> • An AI algorithm developed using a large, open-access dataset of CT slices demonstrated promising utility in identifying acute intracranial hemorrhage (ICH) on non-contrast head CT scans. • The model achieved an area under the curve (AUC) of 0.842 for scan-based detection of ICH, with a pre-specified probability threshold of $\geq 50\%$ yielding 78.6% accuracy, 73% sensitivity, 79% specificity, 18.6% positive predictive value, and 97.8% negative predictive value. • Manual review of CT slices nominated by the model could reduce false-negative scans, indicating the potential for further refinement to enhance the model's localization capabilities for improved clinical application. 	Diagnostics
[20]	Sundrani, Chen, Jin, Abad, Rajpurkar, Kim	Predicting patient decompensation from continuous physiologic monitoring in the emergency department.	Observational Cohort	US	<ul style="list-style-type: none"> • A multimodal machine learning approach combining standard triage data with features from continuous physiologic monitoring accurately predicts the onset of new vital sign abnormalities in ED patients with initially normal vital signs. • The best-performing models, utilizing both engineered and transformer-derived features, achieved high predictive performance, with AUROC values of 0.836 for new tachycardia, 0.802 for new hypotension, and 0.713 for new hypoxia in a 90-minute window. • Salient features contributing to prediction include vital sign trends, PPG perfusion index, and ECG waveforms, highlighting the potential for continuous application of this approach to improve triage and predict clinical deterioration in apparently stable patients. 	Treatment

(Continued)

Table I (Continued).

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[21]	Takaki, Inoue, Maki, Furuya, Mikami, Mizutani, Takada, Okimatsu, Yunde, Miura, Shiratani, Nagashima, Maruyama, Shiga, Inage, Orita, Eguchi, Ohtori	Automated fracture screening using an object detection algorithm on whole-body trauma computed tomography.	Retrospective Observational	Japan	<ul style="list-style-type: none"> Convolutional neural network (CNN) deep learning methods were investigated for automatic localization and classification of pelvic, rib, and spine fractures in trauma care. The CNN model demonstrated promising performance with sensitivity, precision, and F1-score values of 0.786, 0.648, and 0.711, respectively, for grouped mean values of fractures. Surgeons showed improved sensitivity in fracture detection and reduced reading and interpretation time for CT scans, particularly benefiting less experienced orthopedic surgeons, suggesting potential for improved patient care and workflow efficiency in polytrauma cases. 	Diagnostics
[22]	Rashid, Zia, Rehman, Meraj, Rauf, Kadry	A Minority Class Balanced Approach Using the DCNN-LSTM Method to Detect Human Wrist Fracture.	Method Development	Pakistan	<ul style="list-style-type: none"> A fused model of deep learning, combining convolutional neural network (CNN) and long short-term memory (LSTM), was proposed to detect wrist fractures from X-ray images, providing an automated diagnosing tool as a second option for doctors. The dataset comprised 192 wrist X-ray images, and the proposed model utilized image pre-processing and data augmentation techniques to address class imbalance and enhance feature extraction. Comparative analysis with other deep learning models demonstrated that the DCNN-LSTM fusion achieved higher accuracy and exhibited potential for medical applications as a reliable second option in wrist fracture diagnosis, emphasizing its utility in reducing missed diagnoses. 	Diagnostics

[23]	Zech, Santomartino, Yi	Artificial Intelligence (AI) for Fracture Diagnosis: An Overview of Current Products and Considerations for Clinical Adoption, From the AJR Special Series on AI Applications.	Review	US	<ul style="list-style-type: none"> • Artificial intelligence (AI) and deep learning have demonstrated strong potential in accurately detecting fractures, enhancing radiologists' performance in research settings. • Despite the increasing availability of AI products for clinical use, guidance for radiologists on adopting this technology is limited. • This review outlines how AI and deep learning algorithms can assist radiologists in diagnosing fractures and provides an overview of commercially available FDA-cleared AI tools for fracture detection, along with considerations for their clinical adoption. 	Diagnostics
[24]	Wei, Li, Sing, Yang, Beeram, Puvanesarajah, Valle, Tornetta, Fritz, Yi	Detecting total hip arthroplasty dislocations using deep learning: clinical and Internet validation.	Observational Diagnostic Accuracy	US	<ul style="list-style-type: none"> • Convolutional neural networks (CNNs) were trained and evaluated for the automated detection of periprosthetic dislocations of total hip arthroplasty (THA) on radiographs, aiming to expedite diagnosis and treatment. • Multiple CNNs achieved excellent diagnostic performance, with area under the receiver operating characteristic curve (AUROC) values of 1 for both internal and external test sets, indicating high generalizability. • Class activation mapping (CAM) revealed consistent emphasis on the THA region by CNNs for both dislocated and non-dislocated cases, supporting their potential use for efficient triage in the emergency department. 	Diagnostics

(Continued)

Table I (Continued).

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[25]	Yao, Leung, Tsai, Huang, Fu	Novel Deep Learning-Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study.	Retrospective Cohort	Taiwan	<ul style="list-style-type: none"> • A deep learning-based triage system was developed using electronic medical records from emergency department (ED) patients to predict clinical outcomes after ED treatments. • The system, utilizing convolutional neural networks combined with recurrent neural networks and attention mechanisms, achieved high accuracy and area under the receiver operating characteristic curve (AUROC) values for predicting hospitalization, mortality, and admission to the intensive care unit. • Results from both the National Hospital Ambulatory Medical Care Survey and an external dataset from the National Taiwan University Hospital demonstrated superior performance compared to traditional methods, indicating its potential for implementation in real-world clinical settings to improve patient triage and resource allocation in busy EDs. 	Triage
[26]	Sanchez-Salmerón, Gómez-Urquiza, Albendín-García, Correa-Rodríguez, Martos-Cabrera, Velando-Soriano, Suleiman-Martos	Machine learning methods applied to triage in emergency services: A systematic review.	Systematic Review	Spain	<ul style="list-style-type: none"> • Machine learning (ML) methods consistently outperformed traditional triage scales/scores, such as the Emergency Severity Index, in predicting important outcomes like mortality, critical care outcomes, admission, and the need for hospitalization in the emergency department. • Among the ML models considered, XGBoost and Deep Neural Networks demonstrated the highest levels of prediction accuracy, while Logistic Regression performed comparatively worse. • The findings suggest that ML systems have the potential to significantly improve the triage process in emergency departments by accurately predicting important variables, thus aiding in more effective patient management and resource allocation. 	Triage

[27]	Dipnall, Page, Du, Costa, Lyons, Cameron, Steiger, Hau, Bucknill, Oppy, Edwards, Varma, Jung, Gabbe	Predicting fracture outcomes from clinical registry data using artificial intelligence supplemented models for evidence-informed treatment (PRAISE) study protocol.	Prospective Observational	Australia	<ul style="list-style-type: none"> • The "PRAISE" study aims to utilize artificial intelligence (AI) methods on unstructured data to describe fracture characteristics and assess if this information improves the identification of key fracture characteristics and prediction of patient-reported outcome measures and clinical outcomes following wrist fractures. • The study will include adult patients presenting with wrist fractures in four Victorian hospitals, using routine registry data from the Victorian Orthopaedic Trauma Outcomes Registry (VOTOR) and electronic medical record (EMR) information. • A multimodal deep learning fracture reasoning system (DLFRS) will be developed to reason on EMR information, and machine learning prediction models will test performance with or without output from the DLFRS, aiming to provide better prediction of clinical and patient-reported outcomes following distal radius fractures. 	Diagnostics
[28]	Kim, Jung, Park, Park, Yi, Yang, Kim, Cho, Ha	Application of convolutional neural networks for distal radio-ulnar fracture detection on plain radiographs in the emergency room.	Observational Diagnostic Accuracy	Korea	<ul style="list-style-type: none"> • Two convolutional neural network models, DenseNet-161 and ResNet-152, were evaluated for wrist fracture detection using image data collected from patients presenting with wrist trauma at the emergency department. • Performance evaluation on a test dataset showed that both models achieved high sensitivity, specificity, positive predictive value, negative predictive value, and accuracy for wrist fracture detection. • The area under the receiver operating characteristic curves (AUC) indicated excellent performance of DenseNet-161 and ResNet-152 in detecting wrist fractures, suggesting their potential utility as diagnostic aids in the emergency room setting. 	Diagnostics

(Continued)

Table I (Continued).

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[29]	Joseph, Leventhal, Grossestreuer, Wong, Joseph, Nathanson, Donnino, Elhadad, Sanchez	Deep-learning approaches to identify critically ill patients at emergency department triage using limited information.	Cross-Sectional	US	<ul style="list-style-type: none"> • Deep-learning models using only immediately available triage data demonstrated promising accuracy in identifying critically ill patients, surpassing the performance of traditional methods like vital sign thresholds and the emergency severity index (ESI). • Successively complex deep-learning models, including logistic regression, neural networks with structured data, gradient tree boosting, and neural network models with textual data, showed statistically significant improvements in accuracy, as measured by the area under the receiver-operator curve (AUC). • These findings suggest that deep-learning approaches could enhance triage processes by accurately identifying critically ill patients using readily available information, potentially leading to improved clinical and operational outcomes in the emergency department. 	Triage
[30]	Miles, Turner, Jacques, Williams, Mason	Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review.	Systematic Review	United Kingdom	<ul style="list-style-type: none"> • The review assessed the accuracy of machine learning methods in triaging patients presenting in the Emergency Care System, with a focus on calibration, discrimination, and classification statistics. • A total of 92 models from 25 studies were included, with two main triage outcomes: hospitalization (56 models) and critical care need (25 models). • Machine-learning methods, including neural networks, tree-based methods, and logistic regression, demonstrated accuracy in triaging patients, with neural networks showing the highest median C-statistic for critical care need. However, logistic regression models were noted for their transparency in reporting model performance. 	Triage

[31]	Ozkaya, Topal, Bulut, Gursoy, Ozuysal, Karakaya	Evaluation of an artificial intelligence system for diagnosing scaphoid fracture on direct radiography.	Observational Diagnostic Accuracy/ Retrospective Cohort	Turkey	<ul style="list-style-type: none"> • The study aimed to evaluate the diagnostic performance of a convolutional neural network (CNN) in detecting scaphoid fractures on anteroposterior wrist radiographs. • Results showed that the CNN had a sensitivity of 76% and specificity of 92%, with an AUC of 0.840 and an F-score of 0.826 in identifying scaphoid fractures. • The performance of the CNN was comparable to a less experienced orthopaedic specialist but better than that of the emergency department physician. However, the experienced orthopaedic specialist demonstrated the best diagnostic performance according to AUC. 	Diagnostics
[32]	Weikert, Noordtzij, Bremerich, Stieltjes, Parmar, Cyriac, Sommer, Sauter	Assessment of a Deep Learning Algorithm for the Detection of Rib Fractures on Whole-Body Trauma Computed Tomography.	Observational Diagnostic Accuracy	Switzerland	<ul style="list-style-type: none"> • A deep learning-based algorithm showed good diagnostic performance in detecting both acute and chronic rib fractures on whole-body trauma CT scans, achieving a sensitivity of 87.4% and specificity of 91.5%. • The algorithm detected 587 true-positive findings with a sensitivity of 65.7% on a per-finding level, while also identifying 97 true rib fractures not mentioned in the written CT reports. • Correct detection was particularly associated with displacement, indicating the algorithm's potential as a screening tool to prevent false-negative radiology reports in clinical settings. 	Diagnostics

(Continued)

Table I (Continued).

Reference	Author(s)	Title	Design	Country / Region	Summary of Key Findings	Domain
[33]	Jalal, Parker, Ferguson, Nicolaou	Exploring the Role of Artificial Intelligence in an Emergency and Trauma Radiology Department.	Review	Canada	<ul style="list-style-type: none"> The increasing demand for emergency radiology services, coupled with the need for accurate and timely reporting, has led to radiologists being overburdened with high imaging volumes and workload. Artificial intelligence (AI) holds promise in assisting emergency and trauma radiologists by potentially alleviating their workload through automated image analysis and interpretation. This article aims to provide an evidence-based discussion on the evolving role of AI in emergency and trauma radiology departments, addressing technical processes, challenges in algorithm training and validation, ethical considerations, and the pivotal role of emergency radiologists in implementing AI-guided systems to improve patient care and reduce radiologist burnout. 	Diagnostics
[34]	Hwang, Nam, Lim, Park, Jeong, Kang, Hong, Kim, Goo, Park, Kim, Park	Deep Learning for Chest Radiograph Diagnosis in the Emergency Department.	Retrospective Diagnostic Accuracy	Korea	<ul style="list-style-type: none"> The deep learning (DL) algorithm exhibited a high diagnostic performance in identifying abnormal chest radiographs in the emergency department (ED), with an area under the receiver operating characteristic curve (AUC) of 0.95, a sensitivity of 88.7% at a high-sensitivity cutoff, and a specificity of 69.6% at the same cutoff. When compared to on-call radiology residents, the DL algorithm demonstrated higher sensitivity (88.7% vs 65.6%, $p < 0.001$) and lower specificity (69.6% vs 98.1%, $p < 0.001$) at the high-sensitivity cutoff. Reinterpretation of chest radiographs using the algorithm's outputs led to an improvement in sensitivity among residents (73.4%, $p = 0.003$) but a reduction in specificity (94.3%, $p < 0.001$), suggesting that the algorithm aided in improving the sensitivity of resident evaluations. 	Diagnostics

[35]	Kim, Chase, Chang, Kim, Park	Predicting Cardiac Arrest and Respiratory Failure Using Feasible Artificial Intelligence with Simple Trajectories of Patient Data.	Retrospective Cohort	Korea	<ul style="list-style-type: none"> • Developed to predict cardiac arrest or acute respiratory failure 1 to 6 hours in advance, FAST-PACE utilizes a concise set of patient features from ICU data of 29,181 patients, excluding lab results. • The AI model achieved an AUC of 0.886 for predicting cardiac arrest and 0.869 for respiratory failure 6 hours before occurrence, surpassing existing warning scores like MEWS and NEWS. • The study demonstrates that FAST-PACE, relying solely on simple vital signs and patient history, outperforms traditional warning scores, offering a feasible and accurate tool for early prediction of adverse events in emergency situations. 	Diagnostics
[36]	Landry, Ting, Zador, Sadeghian, Cusimano	Using artificial neural networks to identify patients with concussion and postconcussion syndrome based on antisaccades.	Prospective Cohort	Canada	<ul style="list-style-type: none"> • Significant differences observed in prosaccade error rate and median antisaccade latency between control and concussion/PCS groups. • Artificial neural networks (ANNs) achieved accuracies of 67% for concussion and 72% for PCS compared to controls. • ANNs were unable to differentiate between concussion and PCS, indicating persistence of eye movement abnormalities in patients with PCS. 	Diagnostics

convolutional and recurrent neural networks attained even higher predictive accuracy with AUCs exceeding 0.95 for outcomes like hospitalization, mortality, and ICU admission.²⁵ The high AUC values reflect the precise delineation of patients at low versus high risk for adverse outcomes. Beyond binary classification, machine learning methods also show strong calibration for predicting length of stay with mean absolute errors averaging around only 24 to 48 hours.³⁰ Overall, the studies validate machine learning, especially modern deep neural networks, as valuable clinical decision support tools for improved patient acuity assessment and risk segmentation early in the emergency care process.

While machine learning methods strongly outperform standard triage approaches, simpler methods like logistic regression still demonstrate utility. Despite attaining lower median predictive accuracy than neural networks, linear models provide transparency into how different clinical variables are weighted and combined for overall risk estimations.³⁰ This interpretability promotes clinician trust and understanding of model recommendations, a key element influencing real-world adoption. Future triage systems should explore ensembling complex deep learning components with explainable modeling techniques to optimize performance and explicability.

Findings also revealed missed opportunities for leveraging diverse patient data, both structured and unstructured, to further enhance predictive insights. For instance, models utilizing only structured triage data (eg, vital signs, demographics) achieved AUCs of 0.77 and 0.70 for hospitalization and fast-track predictions, respectively. However, models also fed long-format clinical notes attained substantially higher discrimination with an AUC of 0.87 for both outcomes.¹⁵ This underscores the wealth of nuanced clinical information within free-text assessments. Structuring and embedding these heterogeneous data into neural networks could strengthen patient acuity evaluations. Few systems currently draw data directly from electronic health records, representing another untapped data source.¹⁰

Discussion

Most of the investigated studies demonstrated that CNNs can achieve exceptional diagnostic accuracy, with sensitivity and specificity exceeding 90% for diverse traumatic injuries identifiable on CT, X-ray, and MRI.^{9,11,14,19,24} However, reliance on single-center, retrospective data risks optimism bias and overfitting, requiring further multi-institutional prospective validation encompassing diverse patient populations and scanners before responsible clinical integration.^{9,11,14,19,21,24} Additional gaps emerge in standardized injury quantification and localization. Injury severity scoring models demonstrate comparable performance to validated classifications like ISS but currently suffer from mortality overestimation, requiring refinements to improve calibration.¹³ Enhancing lesion localization also remains critical for avoiding missed diagnoses, with hybrid human-AI workflows emphasizing fracture nominations showing particular promise to improve interpreter sensitivity.^{19,21,32} Regarding operational efficiency, while studies hypothesize expedited interpretations and reduced reading volumes, robust quantifications through metrics like interpretation times, reporting throughput, protocol adherence rates, or workload reductions are lacking.^{21,33} Methodical workflow studies leveraging process mining techniques could elucidate AI's concrete efficiency dividends.

Advanced deep learning models achieve impressive triage predictions, but considerable IT ecosystem integration barriers persist, with only 11.7% of identified clinical decision support systems interfacing with EHRs.¹⁰ Implementation research fusing predictive models with health information exchanges could strengthen risk analytics through expanded data interoperability. Another persistent gap emerges between predictive prowess and model transparency, a key element for clinical adoption. While complex neural networks boast strong performance, simpler regression models provide greater visibility into risk calculations.³⁰ Hybrid human-AI approaches blending complex and interpretable models could balance performance and explicability. Additionally, prediction-centric studies dominate over evidence confirming meaningful care pathway improvements, with assessments of tangible patient or system-level benefits remaining sparse. Significant translational research confirming clinical decision support tools that safely utilize enhanced predictions to guide impactful protocols is urgently required. Application-focused treatment investigations represent the largest literature void, with only one study examining an AI-enabled point-of-care ultrasound tool for cardiac assessments.¹⁸ Studies evaluating real-time AI guidance for urgent trauma interventions like ventilator titrations, smart wound analytics, and specialty consultation recommendations are sparse but sorely needed. Robust treatment-oriented research assessing AI's downstream therapy optimization potential will be critical for patient outcome improvements.

Limitations

This review has several limitations worth noting. First, the literature extraction was conducted in only one database, which poses the risk of excluding relevant studies indexed elsewhere. Expanding the search across more databases could reduce this risk. Second, limiting the date range from 2014 onward may omit important prior foundational research. Third, while attempts were made to appraise study quality, meta-analyses innately rely upon the methodical rigor of the included works. Variability between study designs is a common source of heterogeneity. Fourth, manual screening and data extraction introduce the potential for human error or bias, which could be mitigated through dual independent reviewer methods for all stages. Fifth, parameters like the confidence interval and statistical tests for assessing heterogeneity and publication bias, while standard, still involve some subjectivity. Finally, the qualitative synthesis was designed to summarize key themes but was not a fully comprehensive overview of all variables and outcomes reported across heterogeneous trauma-focused generative AI research. A more exhaustive quantification of specific clinical endpoints could better direct practical applications and implementations.

Conclusion

While studies demonstrate that artificial intelligence and deep learning models can achieve impressive diagnostic, prognostic, and workflow efficiency capabilities within emergency trauma contexts, substantial research gaps remain before widespread clinical integration. Diagnostically, CNNs attain exceptional sensitivity and specificity for diverse injury detection, but reliance on retrospective single-center data risks optimism and overfitting biases, necessitating expanded prospective multi-institutional validation on heterogeneous scanners and populations. Injury severity scoring requires calibration refinements to address mortality overestimation, while enhanced lesion localization techniques can heighten model utility by reducing false negatives. Robust workflow studies leveraging process mining methods could better quantify efficiency gains beyond conjectured reading or interpretation time reductions. Prognostically, deep neural networks boast predictive accuracy for triage acuity metrics but confront integration, transparency, and translation obstacles limiting clinical adoption. Significant investigation confirming improved care pathways utilizing enhanced predictions is lacking.

Most prominently, the literature lacks application-oriented treatment studies evaluating real-time AI guidance for augmenting urgent trauma interventions through analytical techniques like inferential sensor fusion, physiology-based parameter customization, and multi-modal feature extraction. Limited ultrasound point-of-care analysis signifies an early use case, but substantial outcome-centric research is urgently required. In conclusion, realizing artificial intelligence's radiant potential necessitates confronting these gaps through rigorous expansion of prospective, multicenter studies as well as an investigational emphasis on model explainability, systems integration, and therapy-centric assessments linking AI utilization to patient benefits uniquely salient within emergency trauma settings.

Acknowledgments

The work is solely that of the authors and does not necessarily represent the views, policies, or opinions of their affiliated institutions, employers, or partners. It was not reviewed or endorsed by any specific institution in particular.

Author Contributions

All authors contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Funding

The work is not funded by any specific source.

Disclosure

The authors report no known conflicts of interest, financial or otherwise in this work.

References

- Martinelli DD. Generative machine learning for de novo drug discovery: a systematic review. *Comput Biol Med.* 2022;145:105403. doi:10.1016/j.combiomed.2022.105403
- Paladugu PS, Ong J, Nelson N, et al. Generative adversarial networks in medicine: Important considerations for this emerging innovation in artificial intelligence. *Ann Biomed Eng.* 2023;51(10):2130–2142. doi:10.1007/s10439-023-03304-z
- Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A; Ouzzani et. al Rayyan — a web and mobile app for systematic reviews. *Syst Rev.* 2016;5(210). doi:10.1186/s13643-016-0384-4
- Amir-Behghadami M, Janati A. Population, intervention, comparison, outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emerg Med J.* 2020;37(6):387. doi:10.1136/emered-2020-209567
- Methodology Checklist 1: Systematic Reviews and Meta-Analyses. Scottish Intercollegiate Guidelines Network, Available from: <https://www.sign.ac.uk/what-we-do/methodology/checklists/>. Accessed May 18, 2024.
- Moher D, Liberati A, Tetzlaff J, Altman, P DG. Group Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.
- Med P, Harris JD, Quatman CE, Manning MM, Siston RA, Flanigan DC. How to write a systematic review. *Am J Sports Med.* 2014;42(11):2761–2768. doi:10.1177/0363546513497567
- Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for systematic reviews of interventions version 6.2 Cochrane; 2021.
- Cheng CT, Lin HH, Hsu CP, et al. Deep Learning for automated detection and localization of traumatic abdominal solid organ injuries on CT scans. *J Imaging Inform Med.* 2024. doi:10.1007/s10278-024-01038-5
- Michel J, Manns A, Boudersa S, et al. Clinical decision support system in emergency telephone triage: a scoping review of technical design, implementation and evaluation. *Int J Med Inform.* 2024;184:105347. doi:10.1016/j.ijmedinf.2024.105347
- Russe MF, Rebmann P, Tran PH, et al. AI-based X-ray fracture analysis of the distal radius: accuracy between representative classification, detection and segmentation deep learning models for clinical practice. *BMJ Open.* 2024;14(1):e076954. doi:10.1136/bmjopen-2023-076954
- 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
- Choi J, Vendrow EB, Moor M, Spain DA. Development and validation of a model to quantify injury severity in real time. *JAMA network open.* 2023;6(10):e2336196. doi:10.1001/jamanetworkopen.2023.36196
- Gao Y, Soh NYT, Liu N, et al. Application of a deep learning algorithm in the detection of Hip fractures. *iScience.* 2023;26(8):107350. doi:10.1016/j.isci.2023.107350
- Sax DR, Warton EM, Sofrygin O, et al. Automated analysis of unstructured clinical assessments improves emergency department triage performance: a retrospective deep learning analysis. *J Am Coll Emerg Physicians Open.* 2023;4(4):e13003. doi:10.1002/emp2.13003
- Ouyang CH, Chen CC, Tee YS, et al. The application of design thinking in developing a deep learning algorithm for hip fracture detection. *Bioengineering.* 2023;10(6):735. doi:10.3390/bioengineering10060735
- Masoumian Hosseini M, Masoumian Hosseini ST, Qayumi K, Ahmady S, Koohestani HR. The aspects of running artificial intelligence in emergency care; a scoping review. *Arch Acad Emerg Med.* 2023;11(1):e38. doi:10.22037/aaem.v11i1.1974
- He B, Dash D, Duanmu Y, Tan TX, Ouyang D, Zou J. Ai-Enabled Assessment Of Cardiac Function And Video Quality In Emergency Department Point-Of-Care Echocardiograms. *J Emerg Med.* 17:2023. doi:10.1016/j.jemermed.2023.02.005
- Abrigo JM, Ko KL, Chen Q, et al. Artificial intelligence for detection of intracranial haemorrhage on head computed tomography scans: diagnostic accuracy in Hong Kong. *Hong Kong Med J.* 2023;29(2):112–120. doi:10.12809/hkmj209053
- Sundrani S, Chen J, Jin BT, Abad ZSH, Rajpurkar P, Kim D. Predicting patient decompensation from continuous physiologic monitoring in the emergency department. *NPJ Digit Med.* 2023;6(1):60. doi:10.1038/s41746-023-00803-0
- Inoue T, Maki S, Furuya T, et al. Automated fracture screening using an object detection algorithm on whole-body trauma computed tomography. *Sci Rep.* 2022;12(1):16549. doi:10.1038/s41598-022-20996-w
- Rashid T, Zia MS, Najam-Ur-Rehman M, Rauf T, Kadry S HT, Kadry S. A minority class balanced approach using the DCNN-LSTM method to detect human wrist fracture. *Life.* 2023;13(1):133. doi:10.3390/life13010133
- Zech JR, Santomartino SM, Yi PH. Artificial Intelligence (AI) for fracture diagnosis: an overview of current products and considerations for clinical adoption, from the ajr special series on ai applications. *AJR Am J Roentgenol.* 2022;219(6):869–878. doi:10.2214/AJR.22.27873
- Wei J, Li D, Sing DC, et al. Detecting total Hip arthroplasty dislocations using deep learning: clinical and Internet validation. *Emerg Radiol.* 2022;29(5):801–808. doi:10.1007/s10140-022-02060-2
- Yao LH, Leung KC, Tsai CL, Huang CH, Fu LC. 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
- Sánchez-Salmerón R, Gómez-Urquiza JL, Albendín-García L, et al. Machine learning methods applied to triage in emergency services: a systematic review. *Int Emerg Nurs.* 2022;60:101109. doi:10.1016/j.ienj.2021.101109
- Dipnall JF, Page R, Du L, et al. Predicting fracture outcomes from clinical registry data using artificial intelligence supplemented models for evidence-informed treatment (PRAISE) study protocol. *PLoS One.* 2021;16(9):e0257361. doi:10.1371/journal.pone.0257361
- Kim MW, Jung J, Park SJ, et al. Application of convolutional neural networks for distal radio-ulnar fracture detection on plain radiographs in the emergency room. *Clin Exp Emerg Med.* 2021;8(2):120–127. doi:10.15441/ceem.20.091
- Joseph JW, Leventhal EL, Grossestreuer AV, et al. Deep-learning approaches to identify critically ill patients at emergency department triage using limited information. *J Am Coll Emerg Physicians Open.* 2020;1(5):773–781. doi:10.1002/emp2.12218
- Miles J, Turner J, Jacques R, Williams J, Mason S. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. *Diagn Progn Res.* 2020;4:16. doi:10.1186/s41512-020-00084-1
- Ozkaya E, Topal FE, Bulut T, Gursoy M, Ozuysal M, Karakaya Z. Evaluation of an artificial intelligence system for diagnosing scaphoid fracture on direct radiography. *Eur J Trauma Emerg Surg.* 2022;48(1):585–592. doi:10.1007/s00068-020-01468-0
- Weikert T, Noordtzi LA, Bremerich J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol.* 2020;21(7):891–899. doi:10.3348/kjr.2019.0653
- Jalal S, Parker W, Ferguson D, Nicolaou S. Exploring the role of artificial intelligence in an emergency and trauma radiology department. *Can Assoc Radiol J.* 2021;72(1):167–174. doi:10.1177/0846537120918338

34. Hwang EJ, Nam JG, Lim WH, et al. Deep Learning for Chest Radiograph Diagnosis in the Emergency Department. *Radiology*. 2019;293(3):573–580. doi:10.1148/radiol.2019191225
35. Kim J, Chae M, Chang HJ, Kim YA, Park E. Predicting CARDIAC ARREST AND RESPIRATORY FAILURE USING FEASIBLE ARTIFICIAL INTELLIGENCE WITH SIMPLE TRAJECTORIES OF PATIENT DATA. *J Clin Med*. 2019;8(9):1336. doi:10.3390/jcm8091336
36. Landry AP, Ting WKC, Zador Z, Sadeghian A, Cusimano MD. Using artificial neural networks to identify patients with concussion and postconcussion syndrome based on antisaccades. *J Neurosurg*. 2018;1–8. doi:10.3171/2018.6.JNS18607

Medical Devices: Evidence and Research

Dovepress

Publish your work in this journal

Medical Devices: Evidence and Research is an international, peer-reviewed, open access journal that focuses on the evidence, technology, research, and expert opinion supporting the use and application of medical devices in the diagnosis, monitoring, treatment and management of clinical conditions and physiological processes. The identification of novel devices and optimal use of existing devices which will lead to improved clinical outcomes and more effective patient management and safety is a key feature of the journal. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/medical-devices-evidence-and-research-journal>