Received: 18 April 2020 | Revised: 4 September 2020 | Accepted: 22 September 2020

DOI: 10.1002/emp2.12277

JACEP OPEN | WILEY

SYSTEMATIC REVIEW–META-ANALYSIS

The Practice of Emergency Medicine

# Artificial intelligence in emergency medicine: A scoping review

Abirami Kirubarajan MSc[1,2] | Ahmed Taher MD, MPH[3] | Shawn Khan BHSc[1] | Sameer Masood MD, MPH[3,4]

[1] Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

[2] Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

[3] Division of Emergency Medicine, Department of Medicine, University of Toronto, Toronto, Ontario, Canada

[4] Toronto General Hospital Research Institute, University Health Network, Toronto, Ontario, Canada

**Correspondence**
Sameer Masood, MD, MPH, Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada.
Email: sam1472@mail.harvard.edu

**Funding and support:** By *JACEP Open* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

## Abstract

**Introduction:** Despite the growing investment in and adoption of artificial intelligence (AI) in medicine, the applications of AI in an emergency setting remain unclear. This scoping review seeks to identify available literature regarding the applications of AI in emergency medicine.

**Methods:** The scoping review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for scoping reviews using Medline-OVID, EMBASE, CINAHL, and IEEE, with a double screening and extraction process. The search included articles published until February 28, 2020. Articles were excluded if they did not self-classify as studying an AI intervention, were not relevant to the emergency department (ED), or did not report outcomes or evaluation.

**Results:** Of the 1483 original database citations, 395 were eligible for full-text evaluation. Of these articles, a total of 150 were included in the scoping review. The majority of included studies were retrospective in nature (n = 124, 82.7%), with only 3 (2.0%) prospective controlled trials. We found 37 (24.7%) interventions aimed at improving diagnosis within the ED. Among the 150 studies, 19 (12.7%) focused on diagnostic imaging within the ED. A total of 16 (10.7%) studies were conducted in the out-of-hospital environment (eg, emergency medical services, paramedics) with the remainder occurring either in the ED or the trauma bay. Of the 24 (16%) studies that had human comparators, there were 12 (8%) studies in which AI interventions outperformed clinicians in at least 1 measured outcome.

**Conclusion:** AI-related research is rapidly increasing in emergency medicine. There are several promising AI interventions that can improve emergency care, particularly for acute radiographic imaging and prediction-based diagnoses. Higher quality evidence is needed to further assess both short- and long-term clinical outcomes.

KEYWORDS

algorithm, artificial intelligence, artificial neural networks, emergency department, emergency medicine, machine learning, technology

Supervising Editor: Christian A. Tomaszewski, MD, MBA

## 1 | INTRODUCTION

The study of artificial intelligence (AI) in medicine has become increasingly popular over the last decade.[1,2] The field of AI refers to a broad subset of computer science that simulates human intelligence, including speech recognition, predictive modeling, and problem solving.[3] Machine learning (ML), a subset of AI, has recently gained popularity in medicine because of its ability to improve algorithms autonomously. Because of rapid advances in computing power and processing techniques, sophisticated ML subtypes such as deep learning have shown potential to improve patient care.[4] As such, there is growing literature that has shown that AI-based interventions can match or even outperform physician expertise.[5]

The emergency department may be uniquely situated to benefit from AI because of its potential value in prediction during triage, as well as its versatility in analyzing diverse patient factors.[6] Patients are assessed in the ED with limited information, and physicians often find themselves balancing probabilities for risk stratification and decision-making. Furthermore, there is a potential opportunity for ED flow metrics and resource allocation to be optimized through algorithm support and computerized decisionmaking.[7] This is otherwise difficult to perform with conventional computing because of the sheer number of variables involved and the constant flux of metrics.

However, there remain concerns regarding the use of AI and its implications for patient safety considering the limited body of evidence to support its implementation.[8–10] For example, unintended patient outcomes may occur if an automated system cuts corners to meet data targets, also known as reward hacking.[9,10] Additionally, it can be difficult for clinicians and medical researchers to understand the available AI-related interventions, because of the interdisciplinary nature of the field and the lack of readily available peer-reviewed research.[11] Although there have been narrative literature reviews exploring the relevance of AI in the ED, no systematic study of the research exists.[6,7,12,13] We sought to systematically search the available literature for AI interventions relevant to the ED through a scoping review and provide a snapshot of how AI can be conceptualized in contributing to emergency medicine.

## 2 | METHODS

The scoping review was conducted according to the standards and guidelines established in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis with the associated extension for Scoping Reviews (PRISMA-ScR), in addition to the fourth edition of the Joanna Briggs Institute Reviewer's Manual.[14,15] In comparison to a conventional systematic review, the scoping methodology was preferable to a conventional systematic review because of the breadth of the field and the high degree of heterogeneity of the included research. We conducted a systematic literature search of Medline-OVID, EMBASE, CINAHL, and IEEE to inform our scoping review. Our search strategy was created in consultation with a research librarian and is included in Table 1.

**TABLE 1** Search strategy. Database: Ovid MEDLINE: Epub Ahead of Print, In-Process, and Other Non-Indexed Citations, Ovid MEDLINE Daily and Ovid MEDLINE. Adapted for EMBASE, IEEE, and CINAHL

| # | Searches |
|---|---|
| 1 | (Artificial adj2 intelligen*).tw,kf. |
| 2 | ((Machine OR deep) adj0 learn*).tw,kf. |
| 3 | (Artificial neural network*).tw,kf. |
| 4 | exp Artificial intelligence/ |
| 5 | 1 OR 2 OR 3 OR 4 |
| 6 | Emergency Treatment/or Emergency Medicine/or emergency medical services/or emergency service, hospital/or trauma centers/or triage/or exp Evidence-Based Emergency Medicine/or exp Emergency Nursing/or Emergencies/or emergicent* or casualty department* or ((emergenc* or ED) adj1 (room* or accident or ward or wards or unit or units or department* or physician* or doctor* or nurs* or treatment*or visit*)).mp. or (triage or critical care or (trauma adj1 (cent* or care))).mp |
| 7 | 5 AND 6 |

### 2.1 | Eligibility criteria

Our inclusion criteria for articles were as follows:

#### 2.1.1 | Population

Treated in the out-of-hospital setting through emergency medical services (EMS), EDs, urgent care centers, and trauma bays in any country.

#### 2.1.2 | Intervention

Any computer science intervention classified as AI by the study authors, which included both supervised and unsupervised ML.

#### 2.1.3 | Comparator

No intervention, standard of care, another computer science intervention, or any other comparator.

#### 2.1.4 | Outcomes

Any outcome reported in the literature.

We examined only original, peer-reviewed literature published in the English language. The search was initially conducted on July 10, 2019 and then updated to include articles published up until February 28, 2020.

Although published conference posters, papers, and abstracts were initially eligible for inclusion, they were later excluded in a second

JACEP OPEN WILEY



**TABLE 2** Descriptions of the broad categories of artificial intelligence (AI) included in the review

| Type of AI | Common labels | Example | Definition | Study Citation |
|---|---|---|---|---|
| Supervised Machine Learning (ML) | Support vector machine (SVM) | Abedi et al[21] | ML that learns a function based on examples and previous input | Evaluation of supervised learning algorithm in emergency department using retrospective data |
| | K-nearest neighbor (KNN) | | | |
| | Naive Bayes (NB) | | | |
| | Regression techniques | | | |
| | Random forest (RF) | | | |
| | Gradient boosting (GB) | | | |
| Unsupervised ML | K principle | Farahmand et al[63] | ML that does not require human input or labeled responses to generate inferences | Artificial Intelligence-Based Triage for Patients with Acute Abdominal |
| | Linear discriminant analysis | | | Pain in Emergency Department; a Diagnostic Accuracy Study |
| | Neural networks | | | |
| | Hierarchical clustering | | | |
| Reinforcement ML | Q-learning | None included | Machine learning that trains models to make decisions based on incentives | N/A |
| | Apprenticeship learning | | | |
| Natural Language Processing (NLP) | Sentiment analysis | Pestian et al[22] | Artificial intelligence (AI) regarding human language (including language recognition, understanding, and generation) | Randomized controlled trial of natural language processing software in ED |
| | Optical character recognition | | | |
| | Natural language generation | | | |

round of screening by 2 reviewers (AK, SK). Articles were excluded if they did not self-classify as studying an AI or ML intervention within either the study title or abstract. Studies were also excluded if location or context was not the ED or out-of-hospital setting. Studies were not eligible if they used ED patient data sets but were not directly relevant to emergency medicine (eg, if ED patient visits were used to generate public health predictions). Finally, papers that did not include outcomes or evaluations were excluded.

A hand search of citations of relevant reviews was performed to ensure comprehensiveness of the search. Grey literature was not formally searched.

## 2.2 | Study selection and extraction

Study selection was completed by 2 independent, parallel reviewers (AK, AT) for both title and abstract screening and then subsequent full-text screening. A pilot test of screening was conducted for the first 100 search results. Each abstract underwent 2 rounds of evaluation by a separate reviewer. Eligible abstracts underwent full-text screening again by the 2 separate reviewers. Discrepancies in screening were resolved through consensus between the 2 reviewers (AK, AT).

Data extraction was performed independently by 2 reviewers (AK, SK), with a third (AT) resolving discrepancies via consensus. Risk of bias (ROB) for individual studies was graded using an adapted rating scale based on the Cochrane ROB tool and the ROB Assessment tool for Non-randomized Studies (RoBANS).[16,17]

## 2.3 | Categorization and analysis

We categorized the AI interventions into ML (further subdivided into supervised, non-supervised, and reinforcement ML) and natural language processing (NLP). A simplified explanation of the different AI types is provided in Table 2.

We then categorized studies by the purpose of the intervention, which were identified based on the study's reported objective. Although the list of purposes is not exhaustive, we felt that they were able to best map the intentions of the studied interventions. Purpose categories were determined through an iterative process based on frequency of results and suggestions from the literature.[6,12]

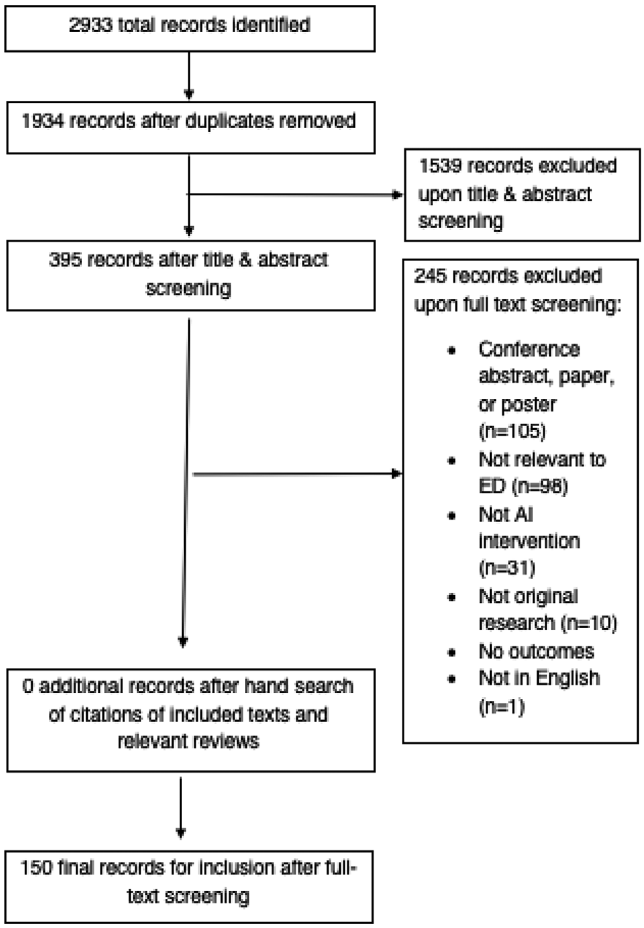


**FIGURE 1** Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow chart. AI, artificial intelligence; ED, emergency department

Data were further extracted regarding the purpose of the intervention, study type (eg, retrospective, prospective), sample size, and type of comparator (eg, human comparator, statistical model, standard of care). All outcomes were summarized descriptively.

## 3 | RESULTS

### 3.1 | Search yield

Results of the study screening process are available in the PRISMA diagram in Figure 1. Of the 2933 original database citations, 999 duplicates were removed using the Covidence platform. After screening the 1888 remaining studies, 395 were eligible for full-text evaluation. Following screening, 105 conference abstracts, posters, and papers were removed. Of the remaining articles, a total of 150 were included in the scoping review. After a hand search of relevant journals and citations, no additional studies were added. Interrater reliability for study screening for titles and abstracts was 92.1%, and for full-text review was 94.1%. The authors were in substantial agreement with a calculated kappa of 0.77 and 0.82 for abstract and full-text screening respectively.

**TABLE 3** Overview of included interventions

| Type of artificial intelligence (AI) intervention | Number of studies (N) |
|---|---|
| Supervised | N = 70 |
| Random forest | 15 |
| Support vector machine | 11 |
| Fuzzy logic | 1 |
| K-nearest neighbor | 1 |
| Decision tree | 8 |
| Supervised artificial neural network | 23 |
| Gradient-boosted algorithm | 5 |
| Classification tree | 1 |
| Other/unspecified | 6 |
| Unsupervised | N = 2 |
| Clustering | 1 |
| Neural network | 1 |
| Reinforced | N = 0 |
| Unspecified | N = 41 |
| Artificial neural network | 34 |
| Deep learning | 3 |
| Other | 4 |
| Mixed models | N = 23 |
| Natural language processing | N = 14 |
| Categories of purpose | |
| Diagnosis | 37 |
| Triage | 18 |
| Prediction | 72 |
| Decisionmaking | 3 |
| Operations | 18 |
| Other | 2 |

Note: ***(numbers do not add to 150 as several studies evaluate multiple interventions)***

### 3.2 | Article characteristics

Details of the included interventions are available in Table 3 and details of the included studies are available in Table 4. The majority of studies were rated as low ROB (n = 139, 92.6%) with the remainder rated as medium ROB (n = 11, 7.3%) (Table 5).

The majority of included studies were retrospective diagnostic accuracy designs (n = 124, 82.7%), with several prospective cohorts (n = 16, 10.7%) and before-and-after implementation designs (n = 4, 2.7%). Only 3 of the included interventions were evaluated using prospective controlled trials, 2 used simulation modeling, and one used a case-control design.

**TABLE 4** Overview of included studies and research designs

| Study setting | |
|---|---|
| Out-of-hospital | 16 |
| Emergency department/trauma bay | 134 |
| Study design | |
| Retrospective diagnostic accuracy | 124 |
| Prospective cohort | 16 |
| Case-control | 1 |
| Controlled trial | 3 |
| Before-and-after | 4 |
| Simulation modeling | 2 |
| Type of comparator | |
| None | 48 |
| Human | 24 |
| Clinical decisionmaking tool | 17 |
| Other non-AI statistical model | 27 |
| Other AI intervention | 27 |
| Standard of care (eg, before and after, set values) | 4 |
| Other (eg, cutoffs from literature, simulation) | 3 |
| | Total number of studies: 150 |

AI, artificial intelligence.

We found that 48 (32.0%) studies did not have a comparator to their AI intervention, and 27 (18.0%) studies used non-AI statistical models. Other studies compared against clinical decision tools (n = 17, 11.3%), standard of care (n = 4, 2.6%), or cutoffs from literature or simulations (n = 3, 2.0%). In 27 (18%) studies, AI tools were compared against one another. A total of 24 (16.0%) studies directly used humans (eg, physicians; EMS staff) as comparators.

## 3.3 | Intervention characteristics

AI utilization categories in emergency medicine are outlined in Figure 2. The majority of interventions centered around prediction, with 72 (48.0%) studies analyzing the predictive capabilities of AI (Table 3). Thirty-seven (24.7%) interventions aimed at improving diagnosis within the ED. Eighteen (12.0%) interventions focused on triage of emergent conditions and another 3 (2.0%) interventions focused on medical decisionmaking. Seventeen (11.3%) studies demonstrated that AI can assist with organizational planning and management within the ED. An additional 2 (1.3%) studies used AI within a research recruitment context. Among the 150 studies, 19 (12.7%) focused on diagnostic imaging within the ED. A total of 16 (10.7%) studies were conducted in the out-of-hospital environment (eg, EMS, paramedics) with the remainder occurring either in the ED or the trauma bay. The breakdown of utilization categories is outlined in Figure 3. Furthermore, we found that 70 (46.7%) studies evaluated supervised ML interventions, 2 (1.3%) studies evaluated unsupervised ML interventions, and 14 (9.3%) studies evaluated NLP interventions (Figure 4).

Of the 24(16%) studies that had human comparators (Table 5), 8 (33.3%) studies defined the performance of clinicians as the gold standard of care. The remaining 16 (66.7%) studies directly compared human versus AI performance. There were 12 (50%) studies in which AI interventions outperformed clinicians. In these studies, AI interventions were better able to diagnose acute cardiac events (including out-of-hospital cardiac arrest and myocardial infarct), identify hyperkalemia, risk stratify patients in triage, identify participants, predict wound infection, predict mortality, predict patients for clinical trials, and read imaging (including intracranial hemorrhages, otoscopic imaging, and fractures).They were non-superior to humans in 3(12.5%) studies that investigated triaging acute abdominal pain, detecting traumatic elbow effusions, and diagnosing chest X-rays. In one study, AI combined with physician judgment was superior to physician judgment alone when diagnosing myocardial infarctions via electrocardiograms (ECGs).[18]

## 3.4 | Limitations

Limitations of our review are namely due to the emerging nature of the evidence base. It is difficult to synthesize conclusions because of the heterogeneous study designs and interventions included in our review. In addition, because of the selective reporting of some studies and the lack of transparency regarding the modeling, it was not often possible to adequately critique the methodology of the studied interventions. Once more AI-related research is established, future systematic reviews may wish to assess more specific interventions in order to determine superiority and areas of improvement. Other limitations of our scoping review were that we examined only studies published in the English language, and we did not analyze patents or grey literature. Only studies that were explicitly self-classified as AI by the study authors were eligible for our analysis, and as such, studies that do not directly identify themselves as AI in their titles and abstracts may have been missed by our search strategy.

# 4 | DISCUSSION

Overall, we found that AI interventions in the ED are heterogeneous in both purpose and design. For example, supervised ML interventions included prediction models for pediatric asthma exacerbation, prediction of return visits, and stroke diagnosis.[19–21] NLP models were used to optimize resource allocation in low-resource settings, classify computed tomography (CT) imaging, and predict hospital admission using electronic medical records (EMR).[22–24] There also appears to be rapidly growing interest in the varied opportunities for AI, as most studies were published in the last 5 years. We can expect an additional

**TABLE 5** Description of studies with human comparator (n = 24)

| Author | Year | Study design | Setting/ Context | Intervention | Purpose | Details | Sample size | Outcomes | AI superiority? |
|---|---|---|---|---|---|---|---|---|---|
| Blomberg et al | 2019 | Retrospective diagnostic accuracy | EMS calling center | ML algorithms | Prediction; Out-of-hospital | Predicting out-of-hospital cardiac arrests | 108 607 | Sensitivity, specificity, PPV, NPV | Y |
| Chilamkurthy et al[57] | 2018 | Diagnostic accuracy | ED | Deep learning algorithms | Diagnosis; Imaging | Diagnosis of intracranial hemorrhages | 313 318 | AUC | N/A (physician used as gold standard) |
| Cicero et al[58] | 2017 | Diagnostic accuracy | ED | Deep convolutional neural networks | Diagnosis; Imaging | Classification of abnormalities on frontal CXRs | 35 038 | Sensitivity, specificity, AUC | N/A (physician used as gold standard) |
| Clarke et al[59] | 2002 | Mixed methods (blinded controlled observational diagnostic accuracy) | ED | Unspecified AI | Decisionmaking | Judge preference of trauma management | 97 | Ratio of judge preference | Y |
| Deleger et al[60] | 2013 | Retrospective diagnostic accuracy | ED | NLP and ML | Triage | Risk stratify appendicitis | 2100 | Recall, precision | N |
| Dipaola et al[61] | 2018 | Observational (Diagnostic accuracy study) | ED | NLP | Diagnosis; Imaging | Detecting traumatic pediatric elbow joint effusions | 901 images (882 patients) | Sensitivity, specificity, AUC | N (0.915 accuracy for fellow versus 0.907 accuracy for model) |
| England et al[62] | 2018 | Diagnostic accuracy | ED | Deep convolutional neural network | Diagnosis; Imaging | Detecting traumatic pediatric elbow joint effusions | 901 images | Sensitivity, specificity, AUC | N/A (physician used as gold standard) |
| Farahmand et al[63] | 2017 | Prospective observational accuracy | ED | Multiple ML algorithms including association rules, decision trees, clustering, LR, NN, naive Bayes | Triage | Acute abdominal pain | 215 (150 training, 75 test) | Accuracy | N/A (physician used as gold standard) |
| Forberg et al[64] | 2012 | Retrospective diagnostic accuracy | Ambulance/ critical care unit | ANN | Prediction | Predicting ST-elevated myocardial infarction with ambulance ECGs | 560 | PPV, NPV, AUC | Y |
| Heden et al[65] | 1997 | Retrospective diagnostic accuracy | ED | ANN | Diagnosis: Non-imaging | Diagnose acute myocardial infarction | 11,572 | Sensitivity, specificity | Y |
| Huesch et al[66] | 2018 | Diagnostic accuracy | ED | NLP | Diagnosis; Imaging | Diagnosis of PE | 1133 images | Sensitivity, specificity, PPV, misclassification rate | N/A (physician used as gold standard) |
| Hwang et al[67] | 2019 | Diagnostic accuracy | ED | DL | Diagnosis, Imaging | Diagnosis of chest X-rays | 1135 patients | AUC, sensitivity, specificity | Undetermined (radiology residents showed lower sensitivity but higher specificity) |

(Continues)

**TABLE 5** (Continued)

| Author | Year | Study design | Setting/ Context | Intervention | Purpose | Details | Sample size | Outcomes | AI superiority? |
|---|---|---|---|---|---|---|---|---|---|
| Jenny et al[68] | 2015 | Retrospective diagnostic accuracy | ED | Multiple ML algorithms including RF, LDA, CART, FDA, BA, NDA | Prediction | Prediction of mortality, acute morbidity, and acute infections | 1278 | Predictability, AUC | Y |
| Lammers et al[69] | 2003 | Prospective cohort | ED | ANN | Prediction | Prediction of traumatic wound infection | 5084 | Sensitivity, specificity, PPV, NPV | Y |
| Lindsey et al[45] | 2018 | Prospective diagnostic accuracy | ED | Deep neural network | Diagnosis: Imaging | Detection of fractures | 135,409 | Sensitivity, specificity | Y |
| Livingstone et al[70] | 2019 | Diagnostic accuracy (cross-sectional) | ED | Unspecified ML with ANN training | Diagnosis; Imaging | Diagnose otoscopic images | 1366 images | Accuracy | Y |
| Ni et al[54] | 2019 | Retrospective diagnostic accuracy | ED | NLP | Other: Research | Eligible patient identification | 202,795 | Workload; efficiency | Y |
| Olsson et al[71] | 2006 | Retrospective diagnostic accuracy | ED | ANN | Diagnosis: Non Imaging | Diagnosing acute coronary syndrome | 4000 (3000 training, 1000 test) | Sensitivity, specificity | Y |
| Pruitt et al[72] | 2019 | Diagnostic accuracy | ED | NLP | Diagnosis; Imaging | Extract characteristics of subdural hematoma from head CT reports | 643 images | Accuracy | N/A (physician used as gold standard) |
| Sinha et al[44] | 2001 | Prospective | ED | ANN | Diagnosis: Imaging | Diagnosis intracranial hemorrhages in closed head injuries | 351 | Sensitivity, specificity | Y |
| Sjogren et al[73] | 2016 | Controlled trial | ED | SVM | Diagnosis; Imaging | Diagnose abdominal free fluid | 20 | Sensitivity, specificity | N/A (physician used as gold standard) |
| Somoza et al.[74] | 1993 | Prospective cohort | ED | ANN | Prediction | Prediction to admission to inpatient psychiatry | 658 | Percent agreement | N/A (physician used as gold standard) |
| Spangler et al[75] | 2019 | Prospective cohort | Out-of-hospital | Multiple ML algorithms including regularized LR, SVM, RF, GB, deep neural networks | Triage | Triage out-of-hospital patients into risk scores | 38203 | Concordance | Y |
| Xue et al[18] | 2004 | Before-and-after | Out-of-hospital | Supervised ML–ANN | Diagnosis, Non-Imaging | Diagnose acute MI | 1902 ECGs | Sensitivity and specificity of a physician's judgment | Y in conjunction with physician's judgment |

ANN, artificial neural network; AUC, area under the curve; ECG, electrocardiogram; ED, emergency department; MI, myocardial infarction; ML, machine learning; N/A, not applicable; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value.

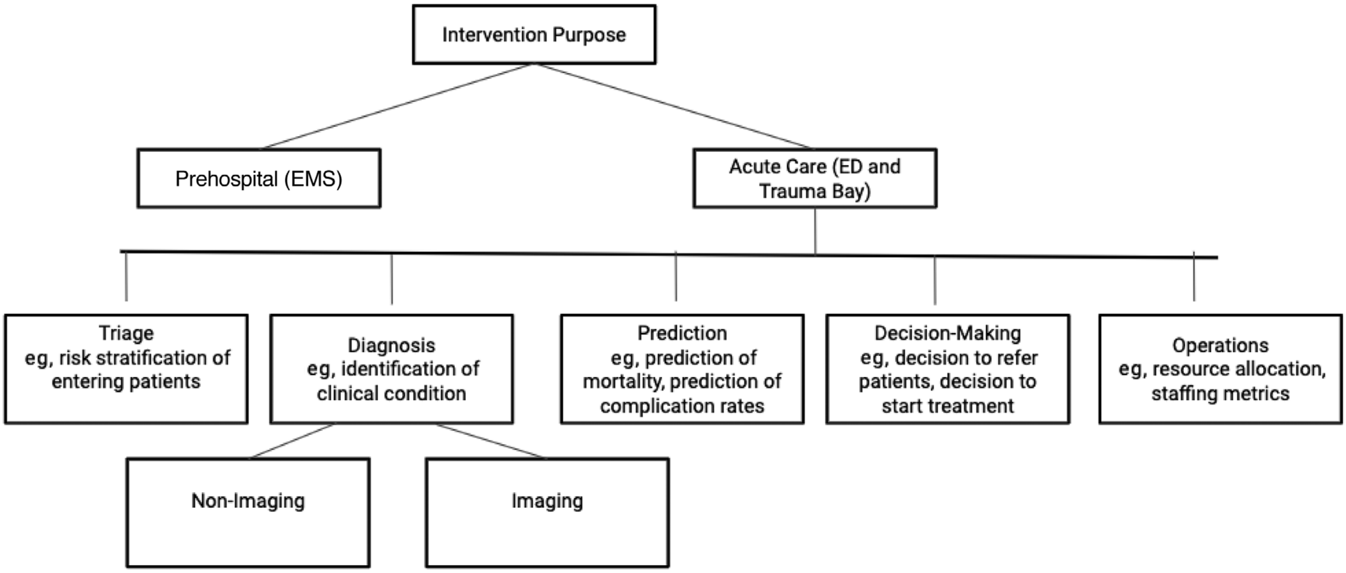


**FIGURE 2** Purpose of intervention in emergency medicine. ED, emergency department; EMS, emergency medical services

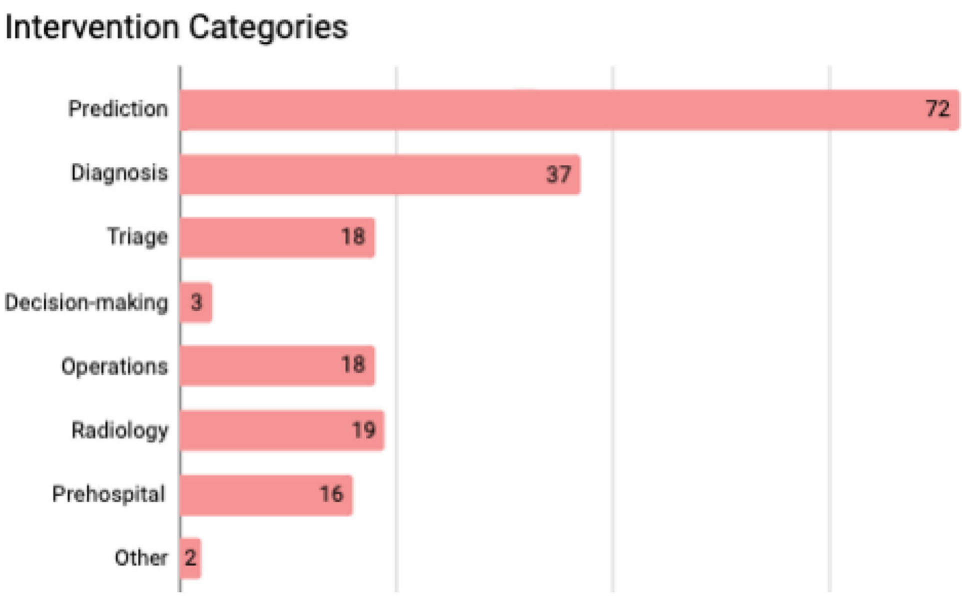


**FIGURE 3** Artificial intelligence utilization category breakdown in emergency medicine

surge of AI-related publications, as over 100 conference proceedings were excluded from our analysis.

The majority of interventions (72, 48.0%) centered around prediction, which aligns with the proposed superiority of AI in prediction modeling.[25] Several studies showed AI outperformed existing decision tools and scoring systems that were originally derived using traditional statistical modeling. Examples include the superior ability of AI to predict mortality in pneumonia as well as calculate syncope risk based on clinical criteria.[26,27] One explanation is that AI may be superior to humans in predictive modeling because of the ability to process multiple variables simultaneously across large data sets.[6] Several of our included studies showed superiority to human comparators when balancing different data points to predict complex outcomes.

As previous studies have shown, human decisionmaking is subject to potential biases and heuristics.[2,5] AI could mitigate illusory correlations and metacognition errors in medicine. For example, AI superiority in predictive modeling is hypothesized to be particularly useful in diagnosis of sepsis, a syndrome that can result in widespread organ dysfunction and high morbidity and mortality.[28] In our review, we found that 6 large cohorts took advantage of big data to predict sepsis and mortality using ML, 4 of which were multicenter studies.[29–34] The 5 studies found that ML models improved prediction among suspected sepsis patients in the ED compared to traditional tools, such as

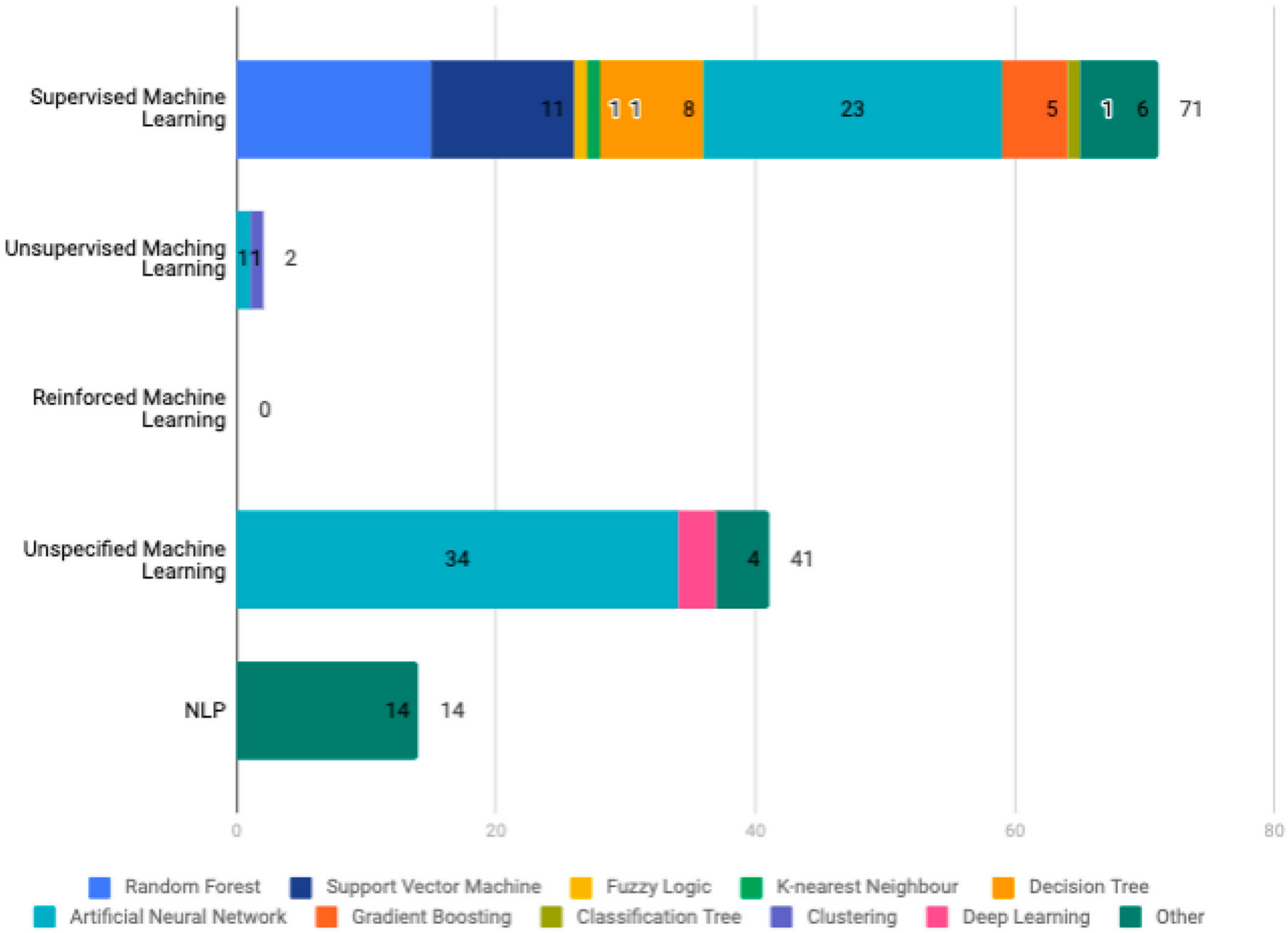


**FIGURE 4** Types of artificial intelligence interventions in emergency medicine. NLP, natural language processing

the Quick Sequential Organ Failure Assessment (qSOFA) or other early warning scores.[35]

A total of 17(9.4%) studies used AI in the out-of-hospital setting. Examples of studies in the out-of-hospital environment included demand forecast for allocation of ambulances, classification of out-of-hospital ECGs, and screening of EMS calls to recognize cardiac arrest.[36–38] Several ML algorithms used EMS data to predict outcomes for out-of-hospital cardiac arrest.[39–42] Another intervention used supervised ML to automatically link EMS electronic patient care reports to ED records.[43] The out-of-hospital setting presents a unique setting where limited clinical variables are used to make prompt decisions (for example, whether or not to transport to hospital). This is well suited to be tackled by AI given its predictive power and ability to use various data points and predict outcomes. As such, we feel that AI will likely have a significant impact in the out-of-hospital setting. Further, interventions in the out-of-hospital environment could promote interdisciplinary communication, reduce inconsistencies, and improve outcomes upon arrival to hospital.

An emerging area of interest was radiology-focused AI interventions, with 19 (12.7%) studies evaluating methods to improve imaging-based diagnosis in the ED. Of the 19 radiology studies, a total of 11(57.9%) had human comparators. For example, Sinha et al (2001) used artificial neural networks to detect intracranial hemorrhage on CT, which was more sensitive (82.2%) compared with physician prediction (62.2%).[44] In contrast, Lindsey et al used an implementation approach to combine their deep learning model with clinician readings that improved fracture detection in comparison to clinicians working alone.[45] Previous literature has noted that radiology is particularly amenable to AI interventions because of its technology-driven interface, reliance on pattern recognition, and relative wealth of data sets.[46,47] As advances in automatic lesion detection and segmentation continue, we can expect further radiology studies with a focus on emergency medicine applications.

Several studies had implications for the ED beyond clinical decision-making. Although much of the current speculation regarding AI has centered on direct patient care, our review shows an emerging interest in ED operations planning, research, and medical education. A total of 18 (12.0%) of our included studies demonstrated that AI can assist organizational planning and management within the ED, including optimization of nursing staff hours, patient satisfaction, and resource planning. Six studies used ML to predict daily patient volume and flow within the ED, including daily trauma volume.[48–53] These interventions have the potential to reduce ED wait-times through improved resource allocation and policy planning. Another 2 (1.3%) studies used ML to increase the efficiency of patient identification for ED clinical trials and research, with the goal of allowing research to be more accessible and standardized in an often fast-paced environment.[54,55]

To our knowledge, this is the first study to systematically review the body of literature regarding emergency medicine and AI. Our scoping review complements the previous literature reviews that have been conducted regarding AI in the ED. For example, Stewart et al (2018) noted that AI had diverse applications in the ED, including use for clinical image analysis, monitoring, and outcome predictions.[6] Their narrative assessment noted that most studies used retrospective data sets, which was corroborated by our systematic search. Stewart et al also noted the lack of universally recognized and standardized reporting guidelines for ML, providing a rationale for why we opted to use a scoping review to better characterize the heterogeneous literature.

Although AI appears superior to clinicians regarding predictive modeling, the current body of evidence still remains uncertain. Most studies identified did not involve a human comparator and lacked information on safety-oriented outcomes. The majority of included studies were retrospective analyses of data sets and require further validation in controlled clinical trials. In addition, most studies do not discuss the practical components associated with technology implementation, such as the convenience, training, or costs required. One obstacle to reproducibility and reliability of results is the proprietary nature of algorithms, as many interventions are not publicly available or transparently described.

Owing to the diverse and rapidly changing field of AI, a scoping review was the most suitable methodology to systematically search and evaluate the available literature. The flexible nature of the search allowed us to include a heterogeneous array of study designs and broad intervention types in order to best map the available evidence. Our scoping review followed PRISMA-ScR guidelines and systematically examined a combination of medicine, allied health, and computer science databases for a concise snapshot.[14] Whereas previous literature reviews have hypothesized different opportunities for AI implementation in emergency medicine, ours is the first to examine the available literature at a glance for physicians to understand the current field. Given the rapid development of AI technology, this review is a timely cross-section of available research. Other strengths include our 2 independent parallel reviewers, our high interrater agreement, and our analysis of interdisciplinary databases.

In comparison to prior literature reviews, our study is the first to systematically examine the scope of AI-related research in the ED. In addition, previous studies did not include out-of-hospital interventions, such as the role of AI in improving EMS or paramedic services. Other gaps in prior work include lack of information regarding the role of AI in medical education or research, as well as operations focused studies such as prediction of ED volume. Lastly, we are also the first study to categorize studies where AI interventions were balanced against human comparators in ED.

By examining the breadth of the literature, it is clear that AI shows strong promise in improving outcome prediction in the ED. AI showed superiority over human comparators in several areas, particularly when analyzing large data sets and rapidly fluctuating variables. Further research should be conducted to determine further opportunities for predictive modeling within the ED, and particularly with comparisons to existing standards of care. Studies should also consider comparing different types of ML for improved accuracy.

For effective AI implementation within hospital systems, AI-related research must progress beyond proof-of-concept. It remains difficult to determine how AI will be adopted within existing systems, as most studies do not compare their interventions to existing standards of care or human comparators. Additional challenges surrounding AI include whether physicians and healthcare staff will have difficulty interfacing with AI-based tools, and whether errors will occur as a result of poor technological literacy. Similarly, there have been concerns regarding both physician and patient uptake of AI, particularly the lack of trust in "black box" technologies that are not clearly understood.[8–10] Specific challenges include mistrust in external validity of data sets, inability of computerized tools to understand clinical context, or mistrust in programmed correlations.[8–10,56] As such, further research must be conducted regarding both physician and patient perspectives towards implementation and ethics of AI. Future research must involve prospective controlled trials in order to determine true superiority, in addition to assessing costs, feasibility, and integration.

## 5 | CONCLUSION

AI-related research is rapidly increasing in emergency medicine. Studies show promising opportunities for AI in diverse contexts, particularly regarding predictive modeling for patient outcomes. However, there remains uncertainty regarding their superiority over standard practice, and further research is needed before clinical implementation.

### CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

### FUNDING INFORMATION

The authors have no sources of funding to disclose.

### AUTHOR CONTRIBUTIONS

AK, AT, and SM conceived the study. SM supervised the conduct of the scoping review and data collection. AK and AT completed both abstract and full-text screening. AK and SK completed data extraction and analysis. AK, AT, SK drafted the article, and all authors contributed substantially to its revision.

### REFERENCES

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Med*. 2019;25(1):44.
2. Beam AL, Kohane IS. Translating artificial intelligence into clinical care. *JAMA*. 2016;316(22):2368-2369.
3. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017;69:S36-S40.

4. Jha S, Topol EJ. Information and artificial intelligence. *J Am Coll Radiol*. 2018;15(3):509-511.
5. Krittanawong C. The rise of artificial intelligence and the uncertain future for physicians. *Eur J Intern Med*. 2018;48:e13-e14.
6. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. *Emerg Med Australas*. 2018;30(6):870-874.
7. Berlyand Y, Raja AS, Dorner SC, et al. How artificial intelligence could transform emergency department operations. *Am J Emerg Med*. 2018;36(8):1515-1517.
8. MacDonald OW. *Physician Perspectives on Preventing Diagnostic Errors*. Waltham, MA: Quantia MD; 2011.
9. Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D. Concrete problems in AI safety. arXiv preprint arXiv:1606.06565. 2016.
10. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28(3):231-237.
11. Paranjape K, Schinkel M, Panday RN, Car J, Nanayakxfkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ*. 2019;5(2):e16048.
12. Liu N, Zhang Z, Wah Ho AF, Hock Ong ME. Artificial intelligence in emergency medicine. *J Emerg Crit Care Med*. 2018;2.
13. Chen N, Liu W, Bai R, Chen A. Application of computational intelligence technologies in emergency management: a literature review. *Artif Intell Rev*. 2019;52(3):2131-2168.
14. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467.
15. Chapter 11: Scoping reviews—JBI Reviewer's Manual—JBI GLOBAL WIKI [Internet]. [cited 2020 Jan 10]. https://wiki.joannabriggs.org/display/MANUAL/Chapter±11%3A±Scoping±reviews.
16. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj*. 2011;343:d5928.
17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol*. 2013;66(4):408-414.
18. Xue J, Aufderheide T, Wright RS, et al. Added value of new acute coronary syndrome computer algorithm for interpretation of prehospital electrocardiograms. *J Electrocardiol*. 2004;37:233-239.
19. Patel SJ, Chamberlain DB, Chamberlain JM. A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Acad Emerg Med*. 2018;25(12):1463-1470.
20. Hu YH, Tai CT, Chen SC, Lee HW, Sung SF. Predicting return visits to the emergency department for pediatric patients: applying supervised learning techniques to the Taiwan National Health Insurance Research Database. *Comput Methods Programs Biomed*. 2017;144:105-112.
21. Abedi V, Goyal N, Tsivgoulis G, et al. Novel screening tool for stroke using artificial neural network. *Stroke*. 2017;48(6):1678-1681.
22. Pestian JP, Grupp-Phelan J, Bretonnel CK, et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. *Suicide Life Threat Behav*. 2016;46(2):154-159.
23. Yadav K, Sarioglu E, Choi HA, Cartwright IV WB, Hinds PS, Chamberlain JM. Automated outcome classification of computed tomography imaging reports for pediatric traumatic brain injury. *Acad Emerg Med*. 2016;23(2):171-178.
24. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods Inf Med*. 2017;56(05):377-389.
25. Tiffin P, Paton L. Artificial or intelligent? Machine learning and medical selection: possibilities and risks. *MedEdPublish*. 2018;7.
26. Bae Y, Moon HK, Kim SH. Predicting the mortality of pneumonia patients visiting the emergency department through machine learning. *J Korean Soc Emerg Med*. 2018;29(5):455-464.
27. Costantino G, Falavigna G, Solbiati M, et al. Neural networks as a tool to predict syncope risk in the Emergency Department. *Ep Europace*. 2016;19(11):1891-1895.
28. Schinkel M, Paranjape K, Panday RN, Skyttberg N, Nanayakkara PW. Clinical applications of artificial intelligence in sepsis: a narrative review. *Comput Biol Med*. 2019;115:103488.
29. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med*. 2019;73(4):334-344.
30. Mao Q, Jay M, Hoffman JL, Calvert J, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*. 2018;8(1):e017833.
31. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. *Acad Emerg Med*. 201623(3):269-278.
32. Jaimes F, Farbiarz J, Alvarez D, Martínez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. *Crit Care*. 2005;9(2):R150.
33. Chiew CJ, Liu N, Tagami T, Wong TH, Koh ZX, Ong ME. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. *Medicine*. 2019;98(6):e14197.
34. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PloS One*. 2017;12(4):e0174708.
35. Marik PE, Taeb AM. SIRS, qSOFA and new sepsis definition. *J Thoracic Dis*. 2017;9(4):943.
36. Chen AY, Lu TY, Ma MH, Sun WZ. Demand forecast using data analytics for the preallocation of ambulances. *IEEE J Biomed Health Inform*. 2015;20(4):1178-1187.
37. Rad AB, Eftestøl T, Engan K, et al. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. *IEEE Trans Biomed Eng*. 2017;64(10):2411-2418.
38. Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*. 2019;138:322-329.
39. Yang Z, Yang Z, Lu W, Harrison RG, Eftestøl T, Steen PA. A probabilistic neural network as the predictive classifier of out-of-hospital defibrillation outcomes. *Resuscitation*. 2005;64(1):31-36.
40. Seki T, Tamura T, Suzuki M. SOS-KANTO 2012 Study Group. Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique. *Resuscitation*. 2019;141:128-135.
41. Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. *Resuscitation*. 2019;139:84-91.
42. Harford S, Darabi H, Del RM, et al. A machine learning based model for Out of Hospital cardiac arrest outcome classification and sensitivity analysis. *Resuscitation*. 2019;138:134-140.
43. Redfield C, Tlimat A, Halpern Y, et al. Derivation and validation of a machine learning record linkage algorithm between emergency medical services and the emergency department. *J Am Med Inform Assoc*. 2020;27(1):147-153.
44. Sinha M, Kennedy CS, Ramundo ML. Artificial neural network predicts CT scan abnormalities in pediatric patients with closed head injury. *J Trauma Acute Care Surg*. 2001;50(2):308-312.

45. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci*. 2018;115(45):11591-11596.
46. Noguerol TM, Paulano-Godino F, Martín-Valdivia MT, Menias CO, Luna A. Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning applications in radiology. *J Am Coll Radiol*. 2019;16(9):1239-1247.
47. Pakdemirli E. Artificial intelligence in radiology: friend or foe? Where are we now and where are we heading? *Acta Radiologica Open*. 2019;8(2):2058460119830222.
48. Zlotnik A, Gallardo-Antolin A, ALFARO MC, PÉREZ MC, MARTÍNEZ JM. Emergency department visit forecasting and dynamic nursing staff allocation using machine learning techniques with readily available open-source software. *CIN*. 2015;33(8):368-377.
49. Azadeh A, Yazdanparast R, Abdolhossein ZS, Keramati A. An intelligent algorithm for optimizing emergency department job and patient satisfaction. *Int J Health Care Quality Assurance*. 2018;31(5):374-390.
50. Yousefi M, Yousefi M, Ferreira RP, Kim JH, Fogliatto FS. Chaotic genetic algorithm and Adaboost ensemble metamodeling approach for optimum resource planning in emergency departments. *Artif Intell Med*. 2018;84:23-33.
51. Menke NB, Caputo N, Fraser R, Haber J, Shields C, Menke MN. A retrospective analysis of the utility of an artificial neural network to predict ED volume. *Am J Emerg Med*. 2014;32(6):614-617.
52. Dennis BM, Stonko DP, Callcut RA, et al. Artificial neural networks can predict trauma volume and acuity regardless of center size and geography: a multicenter study. *J Trauma Acute Care Surg*. 2019;87(1):181-187.
53. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159-170.
54. Ni Y, Bermudez M, Kennebeck S, Liddy-Hicks S, Dexheimer J. A real-time automated patient screening system for clinical trials eligibility in an emergency department: design and evaluation. *JMIR Med Inform*. 2019;7(3):e14185.
55. Ni Y, Beck AF, Taylor R, et al. Will they participate? Predicting patients' response to clinical trial invitations in a pediatric emergency department. *J Am Med Inform Assoc*. 2016;23(4):671-80.
56. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA*. 2018;319(1):19-20.
57. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392(10162):2388-2396.
58. Cicero M, Bilbily A, Colak E, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol*. 2017;52(5):281-287.
59. Clarke JR, Hayward CZ, Santora TA, Wagner DK, Webber BL. Computer-generated trauma management plans: comparison with actual care. *World J Surg*. 2002;26(5):536-538.
60. Deleger L, Brodzinski H, Zhai H, Li Q, et al. Developing and evaluating an automated appendicitis risk stratification algorithm for pediatric patients in the emergency department. *J Am Med Inform Assoc*. 2013;20(e2):e212-e220.
61. Dipaola F, Gatti M, Pacetti V, et al. Artificial Intelligence algorithms and natural language processing for the recognition of syncope patients on emergency department medical records. *J Clin Med*. 2019;8(10):1677.
62. England JR, Gross JS, White EA, Patel DB, England JT, Cheng PM. Detection of traumatic pediatric elbow joint effusion using a deep convolutional neural network. *Am J Roentgenol*. 2018;211(6):1361-1368.
63. Farahmand S, Shabestari O, Pakrah M, Hossein-Nejad H, Arbab M, Bagheri-Hariri S. Artificial intelligence-based triage for patients with acute abdominal pain in emergency department; a diagnostic accuracy study. *Advan J Emerg Med*. 2017;1(1):e5.
64. Forberg JL, Khoshnood A, Green M, et al. An artificial neural network to safely reduce the number of ambulance ECGs transmitted for physician assessment in a system with prehospital detection of ST elevation myocardial infarction.*Scand J Trauma Resusc Emerg Med*. 2012;20(1):8.
65. Hedén B, Ohlin H, Rittner R, Edenbrandt L. Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks. *Circulation*. 1997;96(6):1798-1802.
66. Huesch MD, Cherian R, Labib S, Mahraj R. Evaluating report text variation and informativeness: natural language processing of CT chest imaging for pulmonary embolism. *J Am Coll Radiol*. 2018;15(3):554-562.
67. Hwang EJ, Nam JG, Lim WH, et al. Deep learning for chest radiograph diagnosis in the emergency department. *Radiology*. 2019;293(3):573-580.
68. Jenny MA, Hertwig R, Ackermann S, et al. Are mortality and acute morbidity in patients presenting with nonspecific complaints predictable using routine variables?. *Acad Emerg Med*. 2015;22(10):1155-1163.
69. Lammers RL, Hudson DL, Seaman ME. Prediction of traumatic wound infection with a neural network-derived decision model. *Am J Emerg Med*. 2003;21(1):1-7.
70. Livingstone D, Chau J. Otoscopic diagnosis using computer vision: An automated machine learning approach. *Laryngoscope*. 2020;130(6):1408-1413.
71. Olsson SE, Ohlsson M, Öhlin H, et al. Decision support for the initial triage of patients with acute coronary syndromes. *Clin Physiol Funct Imaging*. 2006;26(3):151-156.
72. Pruitt P, Naidech A, Van Ornam J, Borczuk P, Thompson W. A natural language processing algorithm to extract characteristics of subdural hematoma from head CT reports. *Emerg Radiol*. 2019;26(3):301-306.
73. Sjogren AR, Leo MM, Feldman J, Gwinn JT. Image segmentation and machine learning for detection of abdominal free fluid in focused assessment with sonography for trauma examinations: a pilot study. *J Ultrasound Med*. 2016;35(11):2501-2509.
74. Somoza E, Somoza JR. A neural-network approach to predicting admission decisions in a psychiatric emergency room. *Med Decis Making*. 1993;13(4):273-280.
75. Spangler D, Hermansson T, Smekal D, Blomberg H. A validation of machine learning-based risk scores in the prehospital setting. *PloS One*. 2019;14(12):e0226518.

**How to cite this article:** Kirubarajan A, Taher A, Khan S, Masood S. Artificial intelligence in emergency medicine: A scoping review. *JACEP Open*. 2020;1:1691–1702.
https://doi.org/10.1002/emp2.12277