A Qualitative Exploratory Study of Emergency Medicine Clinician Perspectives on Clinical Decision Support Systems (CDSS) Rooted in Machine Learning in England

Scholarly Report submitted in partial fulfillment of the MD Degree at Harvard Medical School

Date: 5 May 2020

Student Name: Tyler F. Hayes, BS, MRes

Scholarly Report Title: A Qualitative Exploratory Study of Emergency Medicine Clinician Perspectives on Clinical Decision Support Systems (CDSS) Rooted in Machine Learning in England

Mentor: Nicos Savva, PhD

Abstract

TITLE: A qualitative exploratory study of emergency medicine clinician perspectives on Clinical Decision Support Systems (CDSS) rooted in machine learning in England

Purpose: To map the landscape of clinical decision support systems (CDSS) rooted in machine learning techniques in emergency medical care through a scoping literature review, and to explore the perspectives, experiences and attitudes of emergency medicine clinicians in England related to CDSS, particularly those rooted in machine learning.

Methods: A comprehensive scoping literature review was performed in order to capture and organize the diversity, and extent of technological validation and integration of point-of-care machine learning-based clinical decision support systems (CDSS) in emergency departments worldwide. Subsequently, semi-structured interviews with emergency medicine clinicians in England familiar with CDSS were conducted. Resulting qualitative data was analyzed to synthesize and critically compare experiences, attitudes and perspectives towards these tools.

Results: 107 publications were identified, with 50% (n = 51) coming from the US and eight from England. 3% (n = 3) were developed on training dataset without subsequent validation. 82% (n = 88) were internally validated, of which 14% (n = 12) were external validated. 3% (n = 3) were prospectively implemented in clinical practice in EDs. The most common applications were diagnosis, followed by outcome prediction, mortality, disposition planning and triage. While each of thirteen clinician interviewees was familiar with artificial intelligence and most respondents (n = 9) were familiar with machine learning and potential applications to emergency medical care, none had used a CDSS rooted in machine learning in their own practice. Themes arising from the interviews were summarized into four different factor families perceived to affect acceptance of CDSS ('facilitators' and 'barriers') – provider, practice, tool and institutional factors.

Conclusions: There exist few published machine learning-based CDSS for use at point-of-care in emergency medicine in England relative to the United States. Correspondingly, emergency medicine providers in England are not familiar with any such tools. However, they do appreciate the potential for these digital technologies to improve the provision of emergency medical care in England with advantages not only for patients but also for providers. Further work should examine the barriers to machine learning-based CDSS development and validation in England and the aspects of emergency medical care that would be most supported using these technologies. Toward this end, the value of clinicians' experiences and perspectives towards computerized support tools at the intersection of artificial intelligence should not be underestimated.

Table of Contents

Glossary

A&E: Accident & Emergency

ACS: acute coronary syndrome

AI: artificial intelligence

ANN: artificial neural network

AUC: area under the receiver operator characteristic curve

CDSS: clinical decision support system

CT: computed tomography

CTPA: computed tomography pulmonary angiogram

ECG: electrocardiogram

ED: emergency department

EHR: electronic health record

GI: gastrointestinal

ICU: intensive care unit

IRB: Institutional Review Board

LGIB: lower gastrointestinal bleed

LOS: length of stay

MI: myocardial infarction

ML: machine learning

MRI: magnetic resonance imaging

NHS: National Health Service

NICE: National Institute for Health and Care Excellence

NIH: National Institutes of Health

NPV: negative predictive value

PE: pulmonary embolism

POC: point-of-care

PPV: positive predictive value

SiM: Scholars in Medicine

SQ: search query

TAM: Technology Acceptance Model

tPA: tissue plasminogen activator

Introduction

Today, resource allocation in healthcare is under a high level of scrutiny (Guindo et al. 2012). In light of sky-rocketing healthcare costs, increased patient demand for many clinical services, growing stores of collected patient data and an increasingly complex menu of diagnostic and therapeutic options for patients, it has never been more important for key determiners of healthcare utilization to be judicious and efficient in the decisions they make (Sanders et al. 2016).

Thus, providers often turn to tools broadly termed clinical decision support systems (CDSS) to assist in the decision-making process, whether pre-, post- or during diagnosis. In addition to the growing volume of patient information collected in today's era of big data, advances in computer processing ability and statistical models, including artificial intelligence, have resulted in an ever-increasing number of computerized clinical decision support systems (CDSS).

Since the 1980s, there has been growing interest in the application of novel techniques in artificial intelligence (AI) to medical decision-making. Machine learning (ML), a subset of AI, is of particular interest to CDSS developers. Machine learning uses algorithms and statistical models in a computerized infrastructure to perform a task without explicit human instruction. Instead, 'learning' algorithms train ML tools through pattern recognition and inference from past data to make optimal predictions or recommendations on new data prospectively.

Deep learning is a subfamily of machine learning that comprises several different techniques, each relying on artificial neural networks (ANN) to model non-linear problems. ANN are composed of several layers of neurons – an input layer, output layer and one or more intermediate layers called 'hidden layers.' Similar to its biological counterpart, most simply, an artificial neuron receives inputs via connections from a preceding layer of neurons, and, if activated above a certain threshold limit, transmit outputs to one or more neurons in the subsequent layer. The ultimate output of the ANN accomplishes a task, such as identifying an image or making a prediction. During the 'learning' phase, ANN modulate the connectivity and strength of connections within the network as well as the activation threshold of individual neurons until the ultimate error rate is minimized. ANN are robust to noisy data, have the ability to represent complex functions and can learn without as in-depth an understanding of the relationships in the underlying data. However, as such, they are not as transparent to the user – how a decision is reached cannot be interrogated (Lorena et al. 2011).

Decision tree learning algorithms, a family which includes random forest and gradient boosting techniques, automatically learn an optimal decision tree structure from a dataset, which, if incomplete, can lead to inefficient decision rulemaking. However, the classification system can be interrogated, offering a level of transparency not permitted by ANN (Lorena et al. 2011; Li et al. 2012).

If the wave is caught right, Cahan et al. highlight the potential of 'the tusanmi of big data' for health, care, and personalized medicine; physicians and patients flourish when "retention, access, and analysis" is delegated to algorithms, leaving space for the human aspects of interpretation, guidance and execution to the physician (2019). At their best, CDSS rooted in big data have the potential to make healthcare safer, more affordable, swifter and more effective.

Appreciating the potential of CDSS supported by intelligent algorithms, the UK's National Health Service (NHS) released a "Code of conduct" for data-driven health and care technologies as well as an "Evidence standards framework" for digital health technologies in 2019 (Dept. of Health & Social Care, 2019; NICE UK, 2019). The 2019 NHS Long Term Plan envisages the central role of technology in improving clinical efficiency, safety and population health through a digital-ready workforce. However, digital CDSS that can bring significant insight and value to healthcare services are not being fully exploited at present (Bardsley et al. 2019).

There exist several qualitative studies which focus on potential reasons for poor acceptance, uptake, integration and impact of CDSS. Liu et al. note that the "prevailing clinical culture" at an institution; timeliness of outcome; convincingness of output; ability to change practice; and existing clinician performance are potential factors that contribute to negative clinician opinions towards CDSS (2006). Additionally, the "black-box" characteristic of many of today's CDSS, which describes the opaque nature of the datasets and inner workings of algorithms that ultimately generate a clinical recommendation, are frequent concerns raised by practitioners leading to poor acceptance of these technologies (Baro et al. 2015; Cabitza et al. 2017; Fischer et al. 2016; Verghese et al. 2018).

The uptake, integration and impact of CDSS are highly dependent on a diverse set of characteristics of the healthcare system in which they are deployed (Brailsford and Vissers 2011). Chief among these are the perspectives, beliefs and attitudes held by healthcare providers and managers towards these technologies. Mixed responses to and slow adoption of CDSS in emergency care have been attributable to several barriers at the interface of algorithm and provider (Handel et al. 2011).

If CDSS are to be implemented optimally, clinicians themselves must play a role in the development of such algorithms in order to prospectively identify challenges and avoid foreseeable mistakes (Fischer et al. 2016). Referencing the 34% of implemented digital CDSS that have failed to result in lasting improvements in clinical practice, Kawamoto et al. stress that one critical factor in improving CDSS is "describing the systems and manner in which clinicians interact with them…[the] successes and failures" (2005). To augment the clinical effectiveness and promote the use of CDSS, further research on the integration of these tools into clinical workflows is needed with attention to the characteristics of the environment into which a CDSS is deployed as well as to its intended users (Bright et al. 2012, Luxton 2019). The unmet potential of CDSS in healthcare is not closed by "pil[ing] on greater varieties of data but by combining software with the best human clinician 'hardware'" (Chen and Asch 2017).

Fundamental to NICE's 2019 "evidence standards framework" for all digital health technologies is "credibility with UK health and social care professionals," "relevance to current care pathways in the UK health and social care system," and "acceptability with users." Correspondingly, a large qualitative review of health technology assessment models points to "credibility" as the central concern of the decision-makers who use these models (Chilcott et al. 2010). With the goal of building trust and uptake of machine-led decision-making in clinical care, NHS England's 2019 "code of conduct" for digital CDSS mandates that the tools' developers "understand users, their needs and the context" and "show how they will be integrated into health and care provision and their implications on responsibility and liability."

One particular area of focus for England is the integration of modern point-of-care (POC) CDSS in its resource-strained Accident & Emergency (A&E) departments, which are projected to face increasing demand year-on-year over the next decade.

Examples of CDSS utilized in emergency care include algorithm-based tools that assist clinicians and hospital managers with hospitalization decisions for patients with low- vs. high-risk pulmonary emboli; automated interpretations of benign vs. dangerous ECG waveforms; predictions of dangerous patient-specific drug-drug or drug-disease interactions; and alerting of continuouslycaptured vital signs that may signal rapid clinical deterioration (Chen and Asch 2017; Esteva et al. 2019). These algorithms were designed with the ultimate aim of creating a clinician-computer dyad superior in decision-making to the singular clinician, or a clinician equipped with simpler decision frameworks. Today, many of these digital tools capitalize on advances in big data and machine learning with the aim of achieving more effective decision-making – both in terms of optimized resource allocation (e.g., clinicians' time, medications, diagnostic tests, hospital beds) as well as improved patient outcome (e.g., patient waiting times, diagnostic accuracy, quality of life measurers) (Sanskruti and Patel 2016).

CDSS have significant potential to improve the quality and efficiency of emergency medical care. However, Handel et al. point to the experience of several US emergency departments (ED) with CDSS and highlight "poor integration into workflow, unintended consequences, uncertainty of success, and lack of expertise" as key barriers to their usefulness and usability. Importantly, no recent qualitative study of emergency clinician perspectives on CDSS in England has been conducted (Jun et al. 2017).

Thus, the aims of this project are:

(1) To map the landscape of clinical decision support systems (CDSS) rooted in machine learning techniques in emergency medical care through a scoping literature review
(2) To explore the perspectives, experiences and attitudes of emergency medicine clinicians in England related to CDSS rooted in machine learning

To our knowledge, this study is the first to describe comprehensibly the landscape of primary published literature on point-of-care clinical decision support systems (CDSS) used in emergency medicine. It is also the first to collect and analyze qualitative data related to emergency medicine clinician experience and perspective on CDSS in England, particularly those rooted in machine learning.

Student role

I was responsible for conceiving, performing, analyzing and reporting on data generated in relation to the scoping literature review and semi-structured interview process. This study is intended to serve as a backbone for research on the generation of novel, accessible and integrable CDSS using machine learning for medical care in England.

Methods

In order to achieve the aims of the project, the overall approach comprised two stages:

(1) To complete a comprehensive scoping literature review to capture and organize the diversity, and extent of technological validation and integration of machine learning-based clinical decision support systems (CDSS) in emergency medicine departments worldwide
(2) To conduct semi-structured interviews with emergency medicine clinicians in England familiar with CDSS, and analyze resulting qualitative data to synthesize and critically compare experiences, attitudes and perspectives towards these tools

Search strategy

The search method consisted of a manual search of digital libraries known to include published literature on artificial intelligence-based tools in healthcare (Fernandes et al. 2020). These include ScienceDirect, IEEE Xplore, MEDLINEPlus (PubMed) and ISI Web of Knowledge.

Search terms

Unique search queries (SQ) were used for each digital library, given that each has a unique advanced search interface (i.e., only some libraries permit text searches of abstracts, only some permit stratification of journal by subject). For ScienceDirect, IEEE Xplore and PubMed, the SQ consisted of title or abstract searches for “algorithm”, “artificial intelligence”, “machine learning”, “deep learning”, “software”, or “clinical decision support”, and main text searches “emergency medicine” or “emergency department.” For ISI Web of Knowledge, the “topic” field was queried for the same terms. The year of publication was not constrained.

Inclusion and exclusion criteria

After duplicate publications were excluded from the digital library search, inclusion and exclusion criteria were applied, as described in Figure 1.

Organization of key characteristics from the scoping review

Publications were manually searched to identify key characteristics as presented in Table 1. A 'clinical domain', which describes the general aspect of emergency medical care provision impacted by the clinical decision support system, was ascribed to each study; domains include capacity planning, diagnosis (non-radiological), disposition, documentation, length of stay (LOS), mortality, outcome prediction, radiology, intervention and triage. For studies with applications to more specific sub-domains of clinical care, a 'clinical focus' was ascribed. The 'furthest stage of development' was also extracted from the studies; these include, in order, development (without further validation or implementation), internal validation, external validation or implementation. Implementation describes tools which were prospectively used by clinicians following external validation. Internal validation describes tools which were validated on unique data from the same pool of data used for the learning stage of algorithm development. External validation describes tools which were validated on data from a pool of data separate from that used in learning or internal validation arising from data collected from at least one separate institution. The 'country' whose data underpinned the development of the CDSS was also extracted; for all studies included, any validation or implementation stage was performed on data from the same country as the development stage. Finally, the type of 'machine learning algorithm(s)' from which the CDSS was developed was identified.

Semi-structured interviews

Semi-structured interviews with emergency medicine clinicians in England were conducted and analyzed. Semi-structured interviews were chosen because they provide enough structure to address all topics relevant to the research question while allowing enough space for the interviewee to elaborate on important issues (Britten 1995).

As is common with qualitative research, participants were selected using 'purposeful sampling', so that a broad variety of views and perspectives were elicited and explored (Sandelowski 1995). In order to capture more fully a breadth of perspectives, clinicians from at least three different A&Es in the greater London area were interviewed. Interviewees were specialist trainees in emergency medicine or consulting physicians in emergency medicine who had practiced emergency medicine within the last five years. Only Type 1 A&Es, or those that can provide the greatest scope of A&E services, were selected. Respondents were recruited through advertisements posted on the London Business School discussion board and emails to emergency medicine physicians at NHS England Trusts. Respondents were asked to recommend other potential interview subjects.

An interview guide based on study objects was developed through collaboration with a qualitative research expert familiar with semi-structured interviews with healthcare workers related to technology. Interview guide topics included the perceived benefits and disadvantages associated with CDSS, especially those rooted in machine learning; their experience, both positive and negative, in deploying CDSS; and ideas that they have for CDSS improvement in development, validation and integration into their clinical practice. The interview guide was pilot-tested with two respondents and revised to improve understanding of the questions asked and to facilitate the retrieval of information relevant to the study objectives.

As is suggested by Francis et al. regarding data saturation, an initial analysis sample of ten interviews was conducted with a stopping criterion of three (2009). In order to incentivize participation, the urgent nature and broader issue that the present study attempts to address was explained. Potential uses of interviewees' contributions (e.g., SiM report, publication) were discussed, as well as how the project might benefit their practice and the experiences of others (e.g., patients, other clinicians).

Interviews were conducted in person whenever possible. If that was not possible, interviews took place via telephone call or via video conferencing. Interviews were audio recorded and transcribed verbatim. Interviews lasted, on average, one hour. In order to identify and describe themes arising from the interviews, qualitative data analysis was followed by 'conventional content analysis' as proposed by Hsieh and Shannon (2005).

Biostatistical methods: None required.

IRB/Ethical considerations: Harvard's Office of Human Research Administration confirmed that this project does not require IRB oversight, as this investigational quality improvement study does not constitute human subjects research. Written consent was obtained from respondents prior to interviews and verbal consent was obtained prior to audio recording.

Results

Scoping literature review of machine learning-based point-of-care CDSS in emergency medicine

The scoping review, explained diagrammatically in Figure 1, identified 107 publications presenting or analyzing machine learning-based clinical decision support systems with intended or implemented applications to provider-level patient care in emergency medicine. Key characteristics of the publications arising from the review are presented in Table 1. These characteristics include one of ten 'clinical domains' within emergency medicine intended to be impacted by the CDSS; further 'clinical focuses' for those CDSS with applications to sub-domains of clinical care, such as specific clinical patient presentations or diagnoses, or discrete elements of emergency medical care workflows; the year of publication; the machine learning technique used to develop the CDSS; the country whose data was used for the development or validation of the tool or the country in which the tool was ultimately implemented; and the furthest stage of tool development, which included development without further validation, internal validation, external validation and implementation.

Firstly, the number of publications per year was assessed (Figure 2). The first tool, which used a deep learning methodology to diagnosis myocardial infarctions (MI) in emergency departments in England, was published in 1997. No publications were identified again until 2003, when one tool from the US was published using deep learning to predict lower gastrointestinal bleeds (LGIB). The next tool, which also used deep learning to diagnosis MI, was published three years later in 2006 from a group in Sweden. It was not until 2011 that the number of publications per year began to increase. The number of published articles increased at an average annual growth rate of 37% from 2011 (n = 2) to 2019 (n = 25). For comparison, the number of citations in MEDLINE increased at an average annual growth rate of 2.1% between 2011 and 2016. Extrapolating the number of articles published thus far in 2020 to the end of the year resulted in a projected further increase in articles to 42, which would represent a 68% increase from 2019.

The number of publications per year split between CDSS data sources derived from the US and other countries studies is presented in Figure 3. While only one study came from the US before 2012 versus six studies from outside the US, between 1997 and 2019, 50% (n = 51) of all publications came from the US. Apart from 2016, when four articles were published from US groups versus two from groups outside the US, since 2012 there has been no difference between the number of articles published in the US and outside of the US. For comparison, since 2012, 41-43% of publications in English on MEDLINE have come from US groups each year. 52% of tools (n = 56) were derived from data sources outside of the US (Figure 4). Of these, England was most represented country with eight published CDSS, followed by Korea (n = 7), China (n = 6), Singapore (n = 6) and Australia (n = 4).

After reviewing each of the selected articles, ten general ‘clinical domains’ of intended application were identified (Figure 5). These domains represent aspects of emergency medical care provision carried out at the provider-level. Many of these domains are interdependent and several tools have potential applications relevant to several domains, however the most salient domain of CDSS application was determined and applied to each article. Wherever possible, ‘clinical focuses’ were ascribed to CDSS (Figure 6); these focuses describe sub-domains, such as more specific diagnoses (e.g., influenza, ectopic pregnancy, stroke), patient presentations (e.g., lower back pain, trauma, acute abdomen) or elements of emergency medicine provider workflow (e.g., documentation, resuscitation).

The most represented domain was ‘diagnosis (non-radiological)’ at 21% (n = 22). This domain comprised CDSS that assisted emergency medicine physicians in identifying diagnoses at the point-of-care with inputs that included patient history, symptoms, physical exam findings, vital signs and laboratory tests but did not include or require a radiological study. The most common diagnoses were acute coronary syndrome (ACS), sepsis and infection. 17% (n = 18) of studies focused on CDSS that assist clinicians in determining the appropriateness of ordering a radiological study or identifying key findings from resulted images across all major imaging modalities (e.g., radiographs, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound); the most common clinical focuses were pulmonary embolism (PE), stroke, fracture and the identification of key diagnostic findings on chest radiographs. 15% (n = 16) of studies described CDSS that predicted a particular clinical outcome for a patient after admission to the emergency department not present on admission other than mortality, which accounted for 11% (n = 12) of articles; the most common outcomes predicted were sepsis, ACS and cardiac arrest. 12% (n = 13) of studies described CDSS whose aim was to assist emergency medicine providers with disposition planning following admission to the emergency department and clinical work-up in the ED (i.e., admission to a general medical or surgical ward, or intensive care unit (ICU); or discharge home). 11% (n = 12) of publications focused on CDSS intended to assist in the triage of patients upon arrival to the ED to the most appropriate section of the ED, which typically vary based on patient acuity and resources available (e.g., equipment, personnel). 7% (n = 7) of articles described CDSS that resulted in the recommendation of a treatment-based intervention; these included interventions for pneumonia, sepsis, ectopic pregnancy and cardiac arrest. Three articles (3%) introduced CDSS aimed at predicting hospital length of stay after admission to the ED. Two articles (2%) described tools that assisted providers with computerized documentation of encounters in electronic health records (EHR). Finally, two articles (2%) focused on daily capacity planning based on the projected number and acuity level of admissions to the ED.

CDSS were also evaluated based on the furthest stage of their development (Figure 7). 3% (n = 3) were developed on a learning or training dataset without subsequent validation on a testing or validation dataset. 82% (n = 88) were internally validated, meaning that the performance of the model was determined on a sample of subjects used to construct the model, but were not yet externally validated or implemented clinically. Cross-validation or bootstrapping methods were used across the studies for internal validation. 11% (n = 12) were external validated, by which model performance was assessed either on institutionally separate retrospective or prospective datasets. The remaining 3% (n = 3) of CDSS were prospectively implemented in clinical practice in EDs in the US. These include a tool used by emergency department clinicians to help determine the appropriateness of CT pulmonary angiogram (CTPA) studies for patients with suspected acute pulmonary emboli, which ultimately significantly decreased the use of CTPA and significantly increased the yield of ordered studies (Raja et al. 2012); a CDSS which used recorded vital signs and patient history recorded during the first stages of triage in the ED to help triage nurses identify most likely diagnoses and key correspondent patient data to be collected in light of those diagnoses (Greenbaum et al. 2019); and a tool which integrates the chest radiograph report, chief complaint, vital signs and laboratory values of a patient diagnosed with pneumonia and subsequently recommends disposition (general medical ward vs. ICU), potential follow-on diagnostic studies and a presumed most effective antibiotic (Jones et al. 2019).

CDSS were organized based on the machine learning techniques underlying their development (Figure 8). The most common methodology was deep learning, on which 36% (n = 38) of CDSS were based. Deep learning is a subfamily of machine learning that comprises several different techniques that each take advantage of artificial neural networks (ANN) to transform input data into a specific output task via an opaque decision-making process. 27% (n = 29) compared the performance of several different multiple machine learning techniques, which are often made available as an ensemble on third-party ML platforms. There was no difference in the proportion of CDSS developed via deep learning or multiple ML techniques (Figure 9). Other types of ML techniques were far less represented. Decision tree learning, including random forest and gradient boosting techniques, was the next most common method used (14%; n = 15). This was followed by logistic regression (9.3%; n = 10) and support vector machine (6.5%; n = 7). Other techniques identified were cluster analysis (0.9%; n = 1), manifold ranking (0.9%; n = 1), naïve Bayes classifier (3.7%; n = 4) and natural language processing (1.9%; n = 2).

The scoping literature review identified eight CDSS developed in England from English datasets (Table 2). Five of these studies attempted to use routinely collected patient information upon admission in order to predict the risk of severe adverse outcomes, including acute myocardial infarction, sepsis, severe bleeding, pulmonary embolism and death (Kennedy et al. 1997; Faisal et al. 2018; Faisal et al. 2018, Ayaru et al. 2015; Rucco et al. 2015). Patient information included demographic factors like age and sex; vital signs; history of present illness; medical and social histories; basic physical exam findings and routinely collected lab results. Authors of these studies cited difficulty in predicting these outcomes given current clinical frameworks and/or high time and resources costs required to improve prognostic accuracy. Techniques used included deep learning, decision tree learning, logistic regression or combinations of multiple machine learning techniques. Three of the remaining studies focused on CDSS for emergency radiology. Two used deep learning to automate detection of bony pathologies from routinely collected radiographs (Arif et al. 2018; Kim et al. 2018). One study used support vector machines to predict the development of intracranial hemorrhages following administration of tissue plasminogen activator (tPA) for ischemic stroke based on CT images and NIH stroke scores (Bentley et al. 2014).

Each of the eight CDSS described was internally validated. Four were externally validated using datasets from institutions separate from those upon which the training data was based. Data used for development, internal validation and external validation, where applicable, were all from English institutions.

Performance measures of CDSS varied across studies. The primary performance measure was the area under the receiver operator characteristic curve (AUC) in five studies and accuracy in two studies. One study focusing on image segmentation of cervical vertebrae from radiographs reported a Dice coefficient, as is commonly done for such studies (Al Arif et al. 2018). One study additionally reported sensitivity, specificity and positive predictive value (PPV), while another reported negative predictive value (NPV) in addition to those statistical measures.

There were several limitations raised by authors of these studies. Broadly speaking, these included concerns about generalizability due to characteristics of validation cohorts, both internal and external, and the practicality of integration into routine clinical practice. One of the four studies that validated its CDSS on external data saw a significant decrease in performance in the external setting (Kennedy et al. 1997). One study on predicting adverse outcomes for patients presenting to the ED with acute lower GI bleeding was limited by large differences in baseline outcome between internal and external datasets (Ayaru et al. 2015), while another examining symptomatic ICH after tPA administration for ischemic stroke was based on a highly imbalanced internal validation dataset (Bentley et al. 2014). Each study noted that inclusion criteria meant that a material number of patients were excluded from algorithm development or validation (i.e., due to missing data, exclusion of atypical presentations, conservative or narrow diagnostic definitions). Challenges to the integration of CDSS into practice that were appreciated by the authors included provider discomfort with ‘black box’ algorithms, data input that is cumbersome in clinical practice (i.e., data is not routinely collected, would take excessive time, labor or material resources to collect) or CDSS focus that is too narrow to be useful to providers who need to consider multiple differential diagnoses at once.

Facilitators and barriers: Results of semi-structured interviews with emergency medicine clinicians regarding machine learning and CDSS

The experiences, beliefs, attitudes and perspectives of thirteen emergency medicine clinicians in the greater London area towards clinical decision support systems and their intersection with machine learning were assessed using semi-structured interviews. Themes arising from the interviews were bucketed into four different factor families thought to affect adoption of CDSS (‘facilitators’ and ‘barriers’) – provider factors, practice factors, tool factors and institutional factors. These buckets are not intended to be mutually exclusive; it can be argued that several factors presented could be attributed to several families and that interdependencies exist between individual factors. Additionally, the relevance and impact of these factors will depend on real-world context. Some factors may be more important when CDSS draw from machine learning techniques while others are relevant for CDSS derived from any methodology.

Eight of the thirteen respondents were aware of CDSS as a concept. However, all acknowledged that they presently use CDSS in some capacity in clinical practice once the various applications of CDSS were explored further (e.g., computerized alerts, 'smart' provider order entry, diagnostic or therapeutic decision trees). While all of the respondents were familiar with artificial intelligence and most of the respondents (n = 9) were familiar with machine learning and potential applications to emergency medical care, none had used a CDSS rooted in machine learning in clinical practice and none were aware of any presently used in clinical practice at their institution or other institutions with which they are familiar.

Provider factors – facilitators: The prevailing sentiment among respondents was that the intention to use and promote CDSS depended on their perceived usefulness; chief among these are helping guide patient evaluations in a way that saved time, reduced the risk of adverse events and provided other value for providers (i.e., training). There should be an agreement amongst providers that the situation a CDSS tries to improve is a material problem and is an issue common enough that providers will think to turn to the CDSS for recommendations. Several interviewees stressed that tools should reduce paper charting and minimize additional documentation or duplication in assessments. Outcomes from CDSS should be automatically incorporated into documentation whenever possible. They hoped that more intelligent tools, such as those based in machine learning, would be more easily integrable into electronic health records. Interviewees expressed that prior experience using computer-based support at work would be important for CDSS acceptance, integration into practice and promotion.

Provider factors – barriers: Respondents highlighted fears that CDSS will generally reduce provider decision-making abilities with a negative spillover effect into clinical contexts that are not directly addressed by the CDSS. They expressed that CDSS should maintain provider autonomy, reflecting a desire to make their own judgments based on medical expertise. Outputs of CDSS should be perceived as recommendations not mandatory edicts. Finally, respondents expressed the importance of being able to explain to patients and other providers how a certain recommendation was arrived at; they noted that opaque methodologies such as those found within un-interrogable algorithms like deep learning could be a significant barrier.

Practice factors – facilitators: Respondents stressed that CDSS should incorporate seamlessly into clinical practice workflow for all relevant providers impacted by the tool, as they believed that perceived compatibility with workflow would be critical to CDSS adoption and promotion. Triggers to use a CDSS should be at an appropriate point in workflow and obvious. They highlighted that the tools must be available at point-of-care and point of decision-making. They should support different EHR usage patterns, as nurses, trainee physicians and consulting physicians input, view and integrate information at different points in workflow. The technical layout of the emergency department was also an important consideration; hardware necessary to execute the CDSS, like tablets at triage and computers on wheels for data collection at the patient bedside, must be readily available. Several respondents noted that the best tools will be incorporable into shared decision-making; evidence-based prediction rules should facilitate conversations with patients by providing data that supports a decision. CDSS design (e.g., patient friendly graphical outputs) should support a clinician's ability to engage a patient in a discussion about risks and benefits. Where possible, CDSS should provide statistical measure (e.g., pre- and post-test probabilities).

Practice factors – Barriers: Some respondents mentioned that the introduction of CDSS should be temporally separate from other changes to workflows (e.g., rolling out a new EHR, department policy or standard, or another tool) in order to maximize adoption. While trade-offs exist between the ability of CDSS to improve ergonomics and to raise important safety concerns and thereby avoid near misses and adverse events, respondents felt that less interruptive interventions (e.g., documentation template tools) are preferable to interventions that disrupt workflow (e.g., pop-ups).

Tool factors – facilitators: Interviewees appreciated that engaging usability experts and end user practitioners from the first phases of tool development is invaluable. Buy-in should be sought by every provider impacted by CDSS – those who collect the data, run the algorithm, interpret and communicate the outcome of the algorithm and downstream providers whose work is impacted by the tool's outcome. One participant raised that providing context-based, evidence-based information resources for providers who want to delve further into the algorithm and its recommendations could be additionally beneficial. Importantly, CDSS should operate with incomplete information, as decisions in emergency medicine often need to be based on incomplete patient histories, including comorbidities and medications; interviewees predicted that inflexibility around this point could significantly reduce the utility of CDSS. Respondents felt that tools should maximize automaticity in their operations but not result in an automated step in clinical care (e.g., sending patient information, ordering a lab test, writing a prescription) and that providers should be able to accept or override recommendations. 'Homegrown' solutions based on internal data were perceived to be most useful, as were those that were agile enough to be integrated into new EHRs and compatible with periodic software and hardware updates.

Tool factors – barriers: Several interviewees expressed that a high 'learning curve' (i.e., non-intuitive navigation) would be a significant barrier to CDSS update. Additionally, tools should not be developed for contexts for which existing tools, if present, are already perceived to be adequate. Respondents raised concerns for tools could not match human ability in complex pattern recognition; they saw potential for machine learning-, particularly deep learning-, based algorithms in this regard. One interviewee stressed that algorithms must not conflict with clinical practice guidelines. Finally, several respondents noted that CDSS should not add additional 'alert fatigue.'

Institutional factors – facilitators: Interviewees raised several factors that are relevant to the structure of the NHS in England; they noted that decisions affecting hospital operations, such as CDSS, are most often made on the Trust level. Thus, if a CDSS was to be adopted by an emergency department it would likely need to be endorsed by Trust leadership, which would potentially be facilitated by Trust leadership teams comprised of at least some members with clinical, especially emergency medicine, experience. Cases should be made to hospital administrators highlighting cost-effectiveness. If CDSS were to be approved for use in an emergency department, several respondents stressed that users would need to receive an initial and follow-up training. That training should include an elaboration on the potential outcomes that are important to each provider who interacts with the tool. Respondents felt that period updates communicating the results of the intervention would be helpful in guiding their adoption and promotion of the tools. Finally, tool use could be incentivized by hospital managers if use was tracked and incorporated into performance reviews.

Institutional factors – barriers: Respondents highlighted that emergency departments are under financial stress and that CDSS must not require additional physical or people resources for their operation or maintenance. After discussion of some of the tools published in England using machine learning, some respondents mentioned that it may be difficult to develop a tool that is felt to be relevant across all emergency departments in England given that development using data from multiple Trusts or external validation across Trusts could be difficult due to differences in norms, culture, workflow, software and hardware. Also, emergency departments in England are staffed by many locum physicians; these physicians, who may work in an emergency department for only a few weeks or months at a time, might not have opportunities to become familiar with a CDSS used at one institution.

Discussion

This study aims to assess the current landscape of published point-of-care CDSS in emergency medical care, as well as the experiences and perspectives of emergency medicine providers with these tools. At present, NHS England is investing deeply both in identifying novel digital health technologies to support their emergency departments as well as determining how these technologies can best be deployed and integrated into clinical practice. Thus, it is valuable to understand which domains of emergency medicine are attempted to being addressed by digital CDSS and the extent to which advances in machine learning approaches are assisting in this effort. It is also important to understand the perspectives of end user providers on the perceived facilitators and barriers to CDSS acceptance, integration and promotion.

The scoping literature review used in this study highlighted that nearly half of machine learning-based CDSS have been developed in the US. While the number of new CDSS published each year has increased since 2011, the gap between US and non-US studies has not narrowed. Future work might aim to address the factors that contribute to this geographic disparity in research output for these tools. Importantly, tools that were validated in one country were always developed using data from that same country. Thus, if a country is to increase adoption of machine learning-based CDSS this analysis suggests that it may be important for the tools to be developed from data within country.

The most common applications for machine learning-based CDSS were for diagnosis, followed by outcome prediction, mortality, disposition planning and triage. Diagnostic algorithms used in radiology were focused on identifying fractures and stroke; diagnostic algorithms that used patient history, presentation and collected lab values focused on sepsis, acute coronary syndrome, pulmonary embolism and pneumonia. These are common presentations to the emergency department that are potentially of high acuity with significant impacts on hospital resource requirements and patient outcome.

If CDSS are to be adopted within an institution it is important that they internally validated using patient data from that institution. The robustness of the tool can also be demonstrated by its performance when validated on an external dataset. If an externally developed tool is to be adopted within an institution, it is important that the tool maintain high performance when transferred from one setting to another. To our surprise, only 12% of CDSS in this study underwent external validation, while 82% were limited to internal validation. Only 3% of publications described tools that were further implemented into clinical practice. Thus, it appears that adoption of machine learning-based CDSS may be hampered by a lack of evidence from external validation.

Semi-structured interviewees were conducted with emergency medicine providers in the greater London metropolitan area. While interviewees were not previously familiar with CDSS based on machine learning techniques used in emergency medical practice, they were able to describe and reflect on other CDSS they have used and explore potential facilitators and barriers to computerized CDSS in general as well as unique considerations related to those based on machine learning. From these interviews, four different sets of factors related to CDSS acceptance, adoption and integration were identified – provider, practice, tool and institution factors.

Facilitating provider factors centered on perceived utility. Utility encompassed time savings, particularly regarding documentation, as well as improved patient safety. The Technology Acceptance Model (TAM) – a standard approach to technological tool evaluation – corroborates this finding (Davis 1981). The TAM states that ‘usefulness’ (or perceptions that technology supports or enhances job performance) is one of two key constructs that predict information technology acceptance. Hindering provider factors focused on perceived threats to autonomous decision-making to CDSS in general. Regarding machine learning, namely deep learning, algorithms, they expressed concerns about being unable to explain to patients or other providers how a certain decision was made due to the ‘black box’ nature of those methods.

Practice factors facilitators included tool attributes that would improve workflow and integrate with available hardware (e.g., computers on wheels, tablet computers) and software (e.g., EHR). Practice factor barriers included tool attributes that would disrupt workflow (i.e., pop-ups requiring an input from a provider). Other concurrent quality improvement initiatives could potentially also limit uptake of CDSS.

Tool-intrinsic facilitating factors centered on flexibility and generalizability – tools should be operable and reliable even with incomplete patient information, which is often a reality in emergency medicine. Respondents appreciated the potential for machine learning-based tools to supersede human ability in complex pattern recognition. In line with the TAM's axiom of 'usability' or ease of use, interviewees noted that a high learning curve for executing or interpreting CDSS would limit their acceptance amongst emergency medicine providers who already feel 'stretched' in their practice.

Finally, several interesting institutional factors emerged. Given the siloed structure of NHS Trusts, respondents pointed out that initiatives developed or considered for one emergency department are usually local decisions. An endorsement to accept a tool thus depends on the culture and norms of an individual institution. Having clinicians, especially emergency medicine providers, as part of hospital administration may improve the acceptance of CDSS. Importantly, periodic training refreshers on how to use CDSS and updates on outcomes impacted by CDSS were thought to be important. Temporary locum physicians commonly work in emergency rooms across NHS England; that a transient work force would also need to be trained on tools that tend to be specific to one particular emergency department was discussed as a considerable barrier.

Many of the considerations towards improved acceptance and integration of computerized CDSS have been previously addressed (Bates et al. 2003, Khan et al. 2015). Chief among this and prior studies are the incorporation of end users at each step of the tool development and validation process, with a focus on fitting the users' workflow and speed; maintaining autonomous medical decision-making capacity; and frequent monitoring and reporting of CDSS impact on outcomes that are meaningful to end users and patients.

Several limitations to this project exist. While databases and search terms were selected to maximize the retrieval of relevant literature, the scoping literature review might not have exhaustively captured all relevant CDSS. For example, tools developed by private companies may not have been published in academic literature. The study was also limited to publications in the English language. The quality of the included studies in the review also varied. The strength of semi-structured interviews may have also been limited by the skill of the interviewer in leading a meaningful, agile discussion and the articulacy of the respondent. True validity is impossible to ensure and there is significant subjective choice of what elements of the qualitative information should be analyzed. With self-reported behaviors and attitudes, there is a risk of social desirability bias. To abrogate some of these limitations, practice interviews and their coding were conducted alongside a researcher experienced in these methods for feedback. Selection bias among the respondents should be considered. Importantly, the generalizability of this study to emergency departments outside of the greater London area and other countries may be limited.

There exist few published machine learning-based CDSS for use at point-of-care in emergency medicine in England relative to the United States. Correspondingly, emergency medicine providers in England do not appear familiar with any such tools. However, they do appreciate the potential for these digital technologies to improve the provision of emergency medical care in England with advantages not only for patients but also for providers. Further work should examine the barriers to machine learning-based CDSS development and validation in England and the aspects of emergency medical care that would be most supported by these technologies. The value of clinicians' experiences and perspectives towards computerized support tools at the intersection of artificial intelligence should not be underestimated.

By shedding light on the perspectives of end users of clinical decision support systems in emergency medical care in England, the present study intends to inform future work related to the generation and refinement of CDSS which both capitalize on recent advances in big data and artificial intelligence and take into consideration the factors that maximize the likelihood of acceptance and uptake by healthcare providers and integration into the broader healthcare system.

References

Bardsley, Martin, et al. “Untapped potential: Investing in health and cara data analytics.” *The Health Foundation*. UK. 2019.

Baro, Emilie, et al. “Toward a Literature-Driven Definition of Big Data in Healthcare.” *BioMed Research International*, vol. 2015, 2015, pp. 1–9., doi:10.1155/2015/639021.

Bates, David, et al. “Ten Commandments for Effective Clinical Decision Support.” *Electronic Health Records*, 2013, pp. 135–156., doi:10.1201/b16306-14.

Brailsford, S, and Jan Vissers. “OR in healthcare: A European perspective.” *European Journal of Operational Research*. vol. 212, no. 2, 2011, pp. 223-234., doi:10.1016/j.ejor.2010.10.026.

Bright, TJ, Wong A, Dhurjati R, et al. “Effect of Clinical Decision-Support Systems: A Systematic Review.” *Ann Intern Med*., vol. 157, 2012, pp29–43. doi: 10.7326/0003-4819-157-1 201207030-00450.

Britten, N. “Qualitative Research: Qualitative Interviews in Medical Research.” *Bmj*, vol. 311, no. 6999, 1995, pp. 251–253., doi:10.1136/bmj.311.6999.251.

Cabitza, Federico, et al. “Unintended Consequences of Machine Learning in Medicine.” *Jama*, vol. 318, no. 6, Aug. 2017, p. 517., doi:10.1001/jama.2017.7797.

Cahan, Eli M., et al. “Putting the Data before the Algorithm in Big Data Addressing Personalized Healthcare.” *Npj Digital Medicine*, vol. 2, no. 1, 2019, doi:10.1038/s41746 019-0157-2.

Chen, Jonathan H., and Steven M. Asch. “Machine Learning and Prediction in Medicine Beyond the Peak of Inflated Expectations.” *New England Journal of Medicine*, vol. 376,no. 26, 2017, pp. 2507–2509., doi:10.1056/nejmp1702071.

Chilcott, J, et al. “Avoiding and Identifying Errors in Health Technology Assessment Models: Qualitative Study and Methodological Review.” *Health Technology Assessment*, vol. 14, no. 25, 2010, doi:10.3310/hta14250.

Davis, Fred D. “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” *MIS Quarterly*, vol. 13, no. 3, 1989, p. 319., doi:10.2307/249008.

Department of Health & Social Care, England, UK. “Code of conduct for data-driven health and care technology.” 2019.

Esteva, Andre, et al. “A Guide to Deep Learning in Healthcare.” *Nature Medicine*, vol. 25, no. 1, 2019, pp. 24–29., doi:10.1038/s41591-018-0316-z.

Fernandes, M., et al. “Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: A Review.” *Artificial Intelligence In Medicine*, vol. 102, 2020.

Fischer, T., et al. “Clinical Decision-Making and Secondary Findings in Systems Medicine.” *BMC Medical Ethics*, vol. 17, no. 1, 2016, doi:10.1186/s12910-016-0113-5.

Francis, Jill J., et al. “What Is an Adequate Sample Size? Operationalising Data Saturation for Theory-Based Interview Studies.” *Psychology & Health*, vol. 25, no. 10, 2010, pp. 1229 1245., doi:10.1080/08870440903194015.

Greenbaum, Nathaniel R., et al. “Improving Documentation of Presenting Problems in the Emergency Department Using a Domain-Specific Ontology and Machine Learning-Driven User Interfaces.” *International Journal of Medical Informatics*, vol. 132, 2019, p. 103981., doi:10.1016/j.ijmedinf.2019.103981.

Guindo, Lalla Aïda, et al. “From Efficacy to Equity: Literature Review of Decision Criteria for Resource Allocation and Healthcare Decisionmaking.” *Cost Effectiveness and Resource Allocation*, vol. 10, no. 1, 2012, p. 9., doi:10.1186/1478-7547-10-9.

Handel, Daniel A., et al. “Using Information Technology to Improve the Quality and Safety of Emergency Care.” *Academic Emergency Medicine*, vol. 18, no. 6, 2011, doi:10.1111/j.1553 2712.2011.01070.x.

Hsieh, Hsiu-Fang, and Sarah E. Shannon. “Three Approaches to Qualitative Content Analysis.” *Qualitative Health Research*, vol. 15, no. 9, 2005, pp. 1277–1288., doi:10.1177/1049732305276687.

Jones, Barbara, et al. “CDS in a Learning Health Care System: Identifying Physicians Reasons for Rejection of Best-Practice Recommendations in Pneumonia through Computerized Clinical Decision Support.” *Applied Clinical Informatics*, vol. 10, no. 01, 2019, pp. 001–009., doi:10.1055/s-0038-1676587.

Jun, Shelly, et al. “Point-of-Care Cognitive Support Technology in Emergency Departments: A Scoping Review of Technology Acceptance by Clinicians.” *Academic EmergencyMedicine*, vol. 25, no. 5, Aug. 2017, pp. 494–507., doi:10.1111/acem.13325.

Kawamoto, Kensaku, et al. “Improving Clinical Practice Using Clinical Decision Support Systems: a Systematic Review of Trials to Identify Features Critical to Success.” *Bmj*, vol. 330, no. 7494, 2005, p. 765., doi:10.1136/bmj.38398.500764.8f.

Khan, Sundas, et al. “Formative Assessment and Design of a Complex Clinical Decision Support Tool for Pulmonary Embolism.” *Evidence Based Medicine*, vol. 21, no. 1, 2015, pp. 7–13., doi:10.1136/ebmed-2015-110214.

Li, Huaxiong, et al. “An Interval Set Model for Learning Rules from Incomplete Information Table.” *International Journal of Approximate Reasoning*, vol. 53, no. 1, 2012, pp. 24–37., doi:10.1016/j.ijar.2011.09.002.

Lorena, Ana C., et al. “Comparing Machine Learning Classifiers in Potential Distribution Modelling.” *Expert Systems with Applications*, vol. 38, no. 5, 2011, pp. 5268–5275., doi:10.1016/j.eswa.2010.10.031.

Liu, Joseph, et al. “Decision Tools in Health Care: Focus on the Problem, Not the Solution.” *BM Medical Informatics and Decision-making*, vol. 6, no. 1, 2006, doi:10.1186/1472-6947-6 4.

Luxton, David D. “Should Watson Be Consulted for a Second Opinion?” *AMA Journal of Ethics*, vol. 21, no. 2, Jan. 2019, doi:10.1001/amajethics.2019.131.

National Institute for Health and Care Excellence (NICE), UK. “Evidence standards framework for digital health technologies.” 2019.

Patel, Sanskruti, and Atul Patel. “A Big Data Revolution in Health Care Sector: Opportunities, Challenges and Technological Advancements.” *International Journal of Information Sciences and Techniques*, vol. 6, no. 1/2, 2016, pp. 155–162., doi:10.5121/ijist.2016.6216.

Raja, Ali S et al. “Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department.” *Radiology* vol. 262,2 (2012): 468-74. doi:10.1148/radiol.11110951.

Sandelowski, Margarete. “Sample Size in Qualitative Research.” *Research in Nursing & Health*, vol. 18, no. 2, 1995, pp. 179–183., doi:10.1002/nur.4770180211.

Sanders, Gillian D., et al. “Recommendations for Conduct, Methodological Practices, and Reporting of Cost-Effectiveness Analyses.” *Jama*, vol. 316, no. 10, 2016, p. 1093., doi:10.1001/jama.2016.12195.

Verghese, Abraham, et al. “What This Computer Needs Is a Physician.” *Jama*, vol. 319, no. 1, Feb. 2018, p. 19., doi:10.1001/jama.2017.19198.

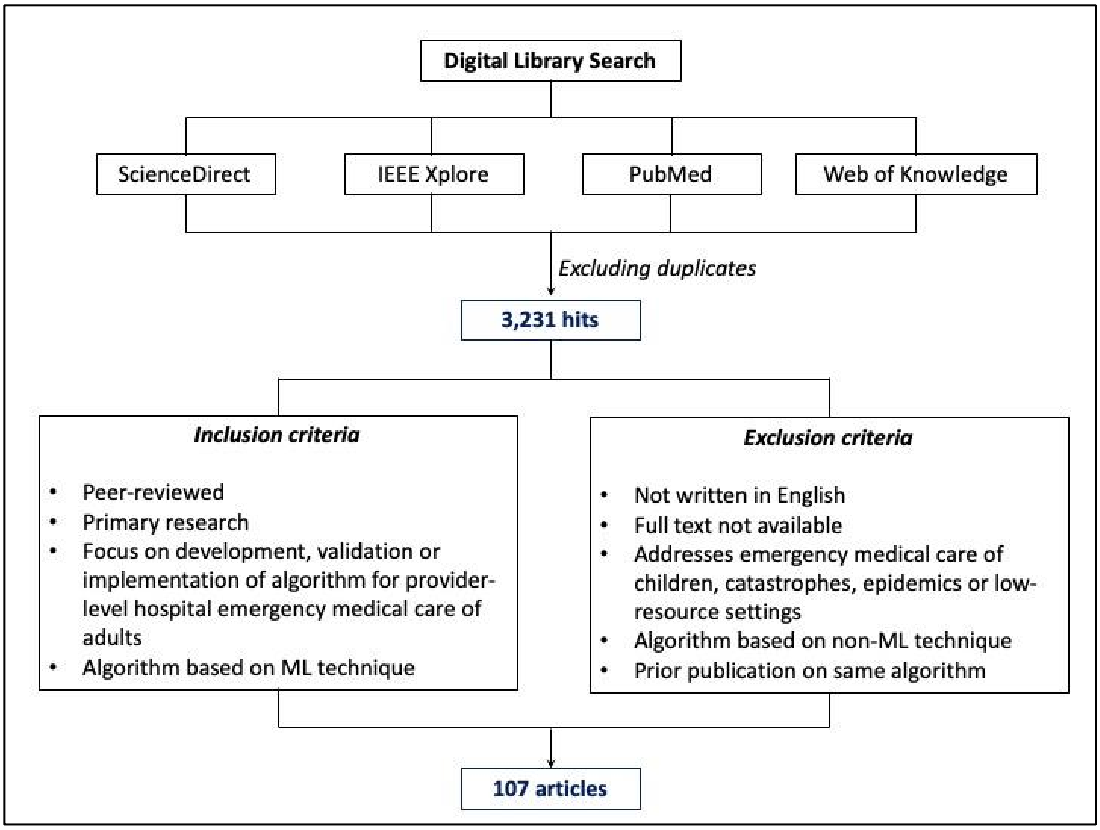

Figure 1: Flow diagram of scoping literature review, and inclusion and exclusion criteria (search date February 15, 2020).

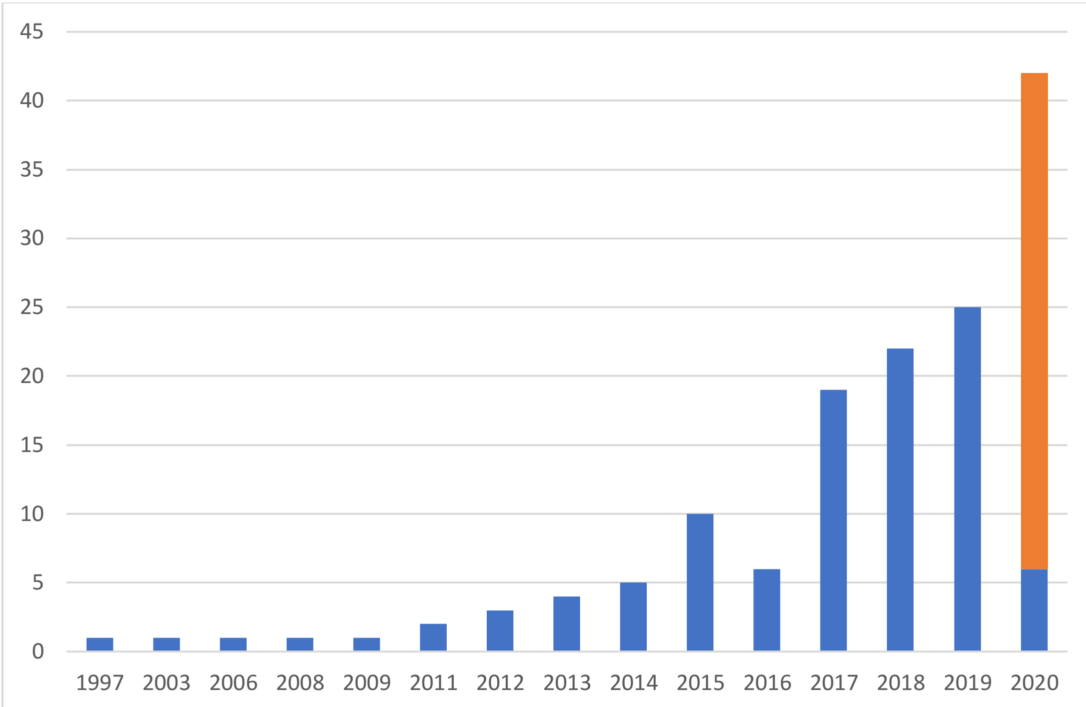

Figure 2: Number of articles published per year. The orange bar represents the number of articles projected to be published in 2020, based on an extrapolation of the articles published in 2020 at the time of the present study.

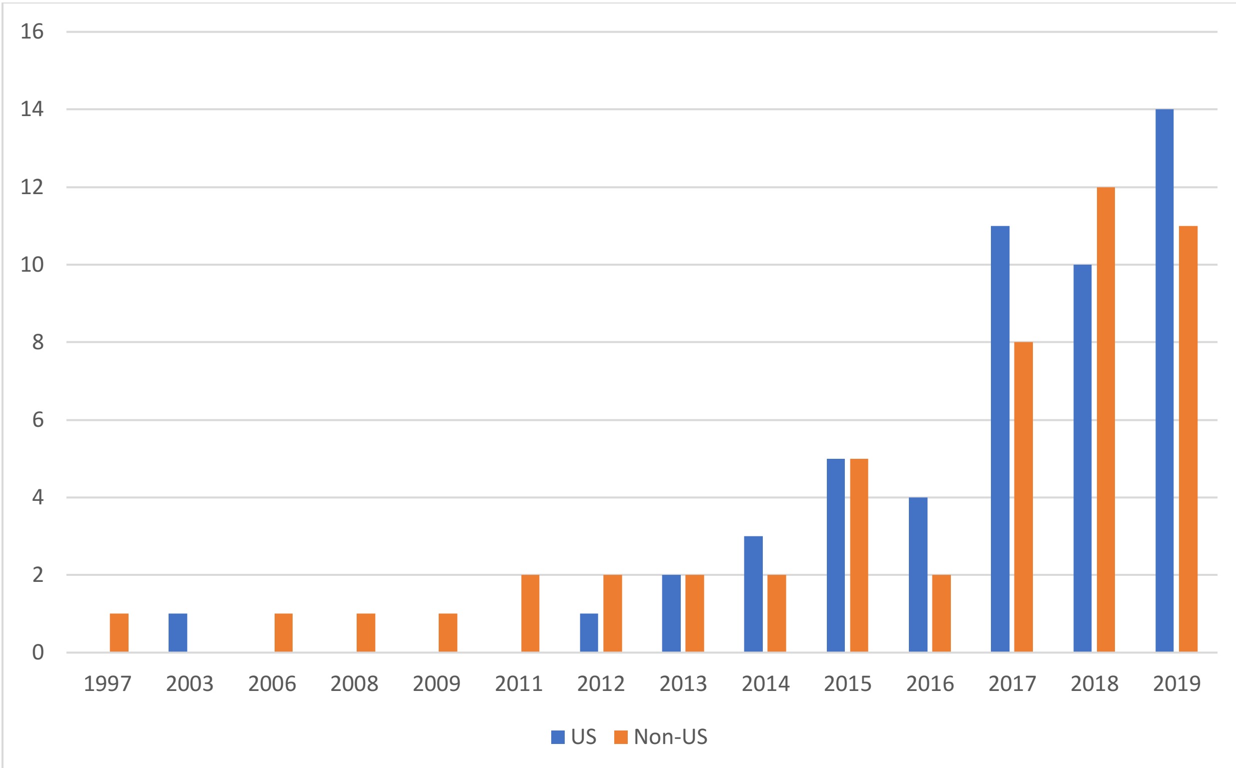

Figure 3: Number of articles published per year on CDSS based on US versus non-US training datasets.

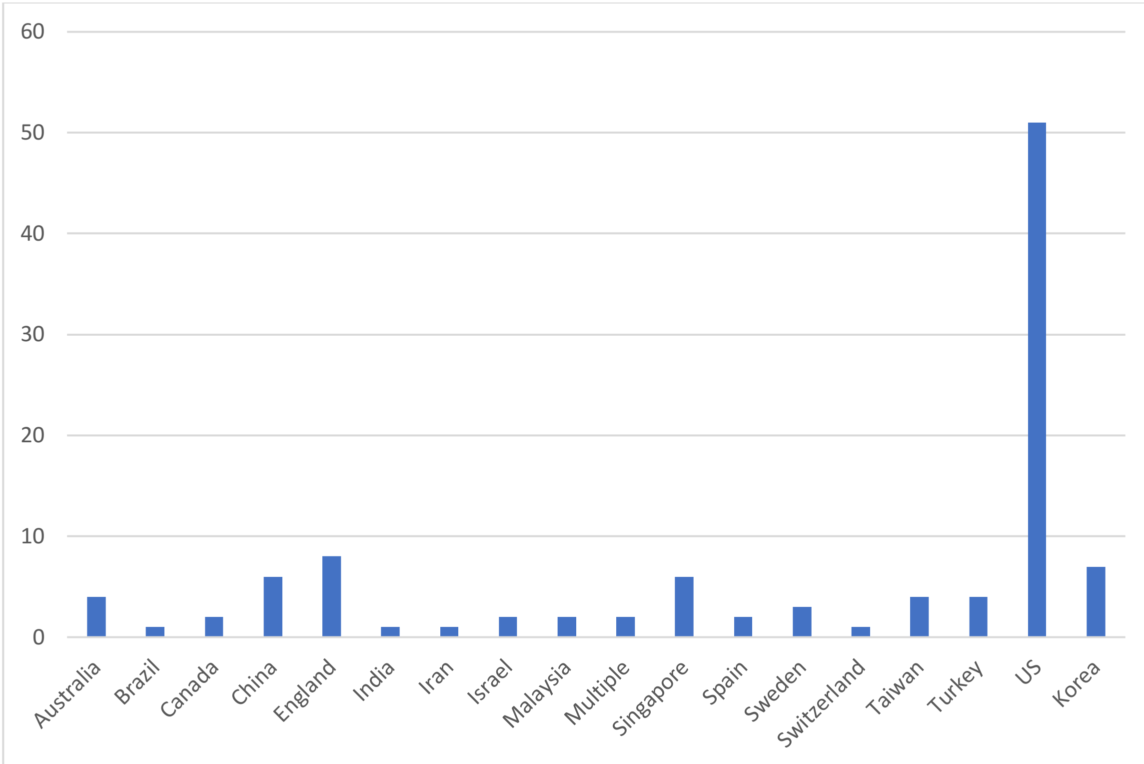

Figure 4: Number of articles published on CDSS by country of underlying training dataset.

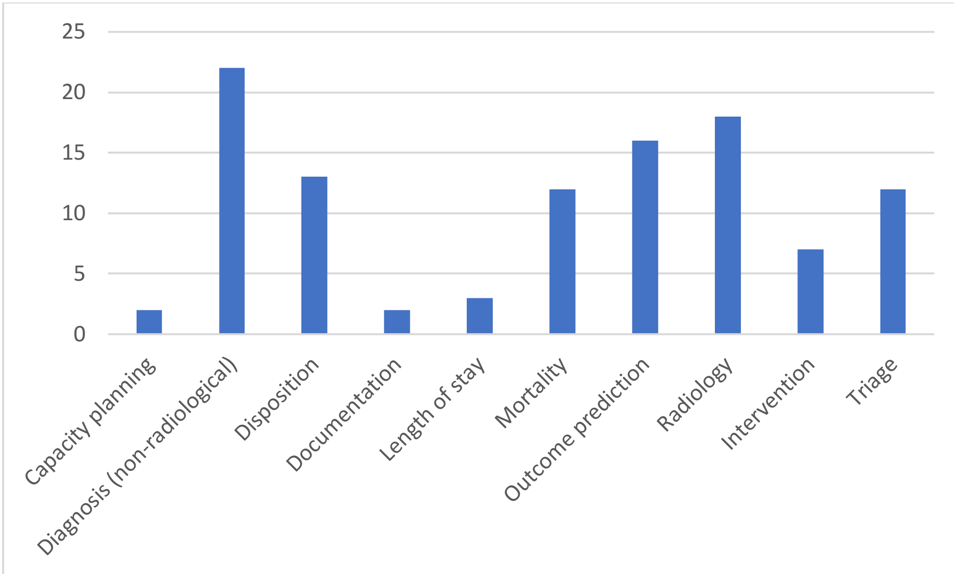

Figure 5: Number of published CDSS organized by clinical domains impacted by the published tool.

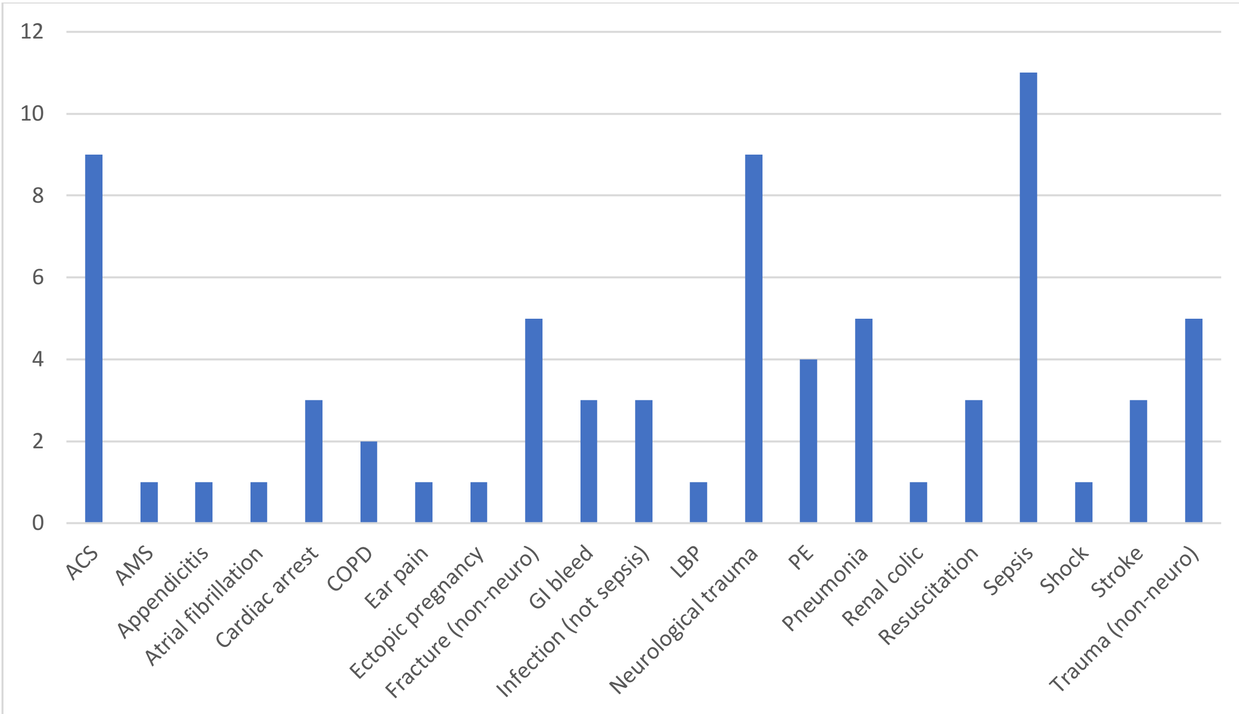

Figure 6: Number of published CDSS organized by clinical focuses impacted by the published tool. Clinical focuses represent subcategories of the 'clinical domains' presented in Figure 5. They include more specific diagnoses, patient presentations or elements of emergency medical provider workflow. *ACS* = acute coronary syndrome. *AMS* = altered mental status. *COPD* = chronic obstructive pulmonary disease. *CT* = computed tomography. *GI* = gastrointestinal. *LBP* = lower back pain. *PE* = pulmonary embolism.

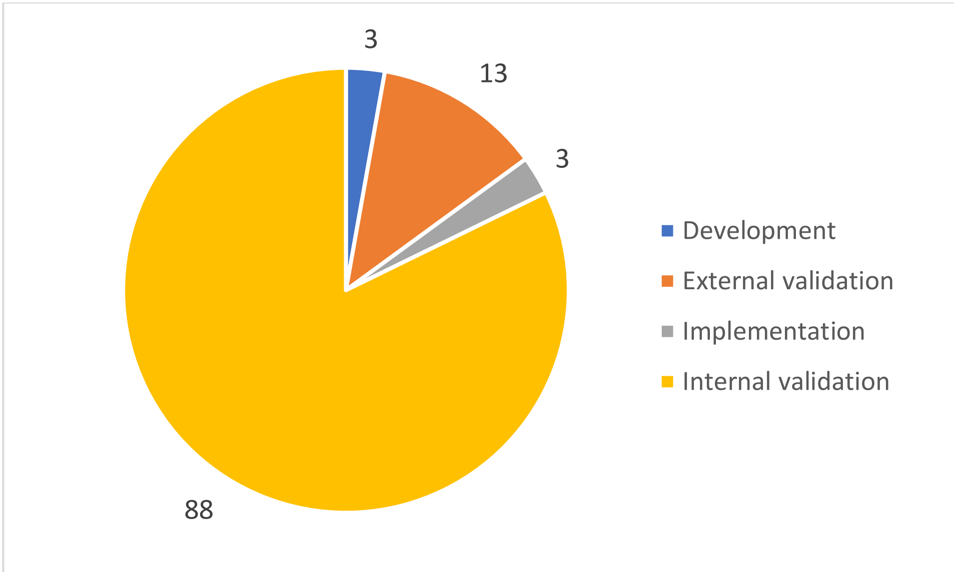

Figure 7: Number of articles by furthest stage of development of CDSS described.

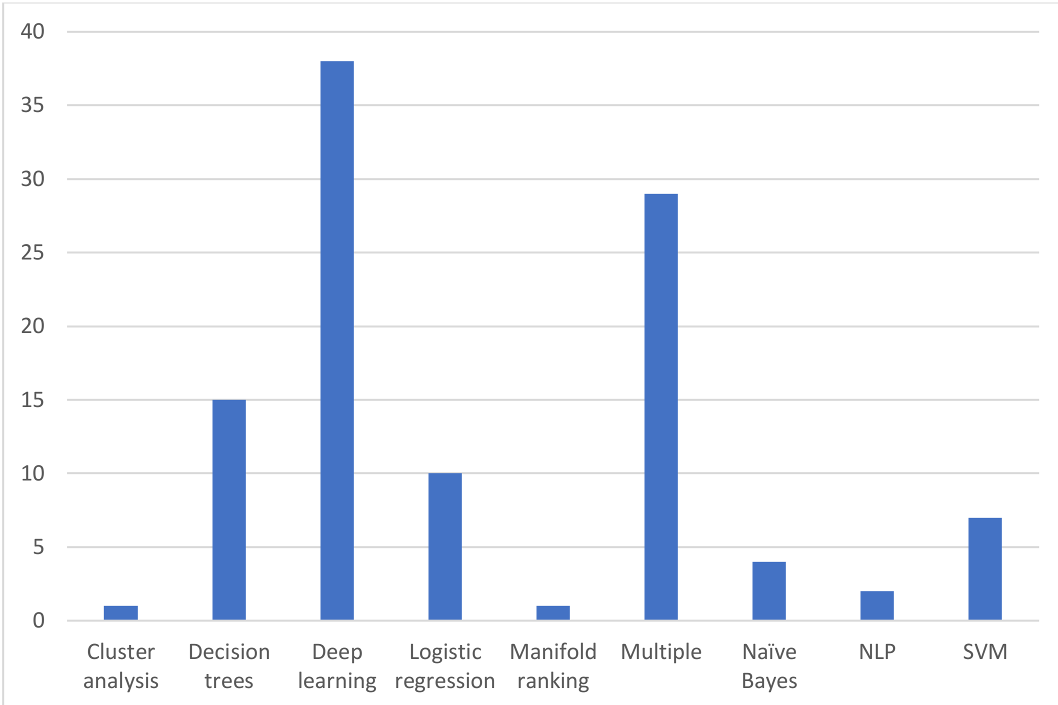

Figure 8: Number of papers by machine learning technique used for CDSS model development. *NLP* = natural language processing. *SVM* = support vector machine.

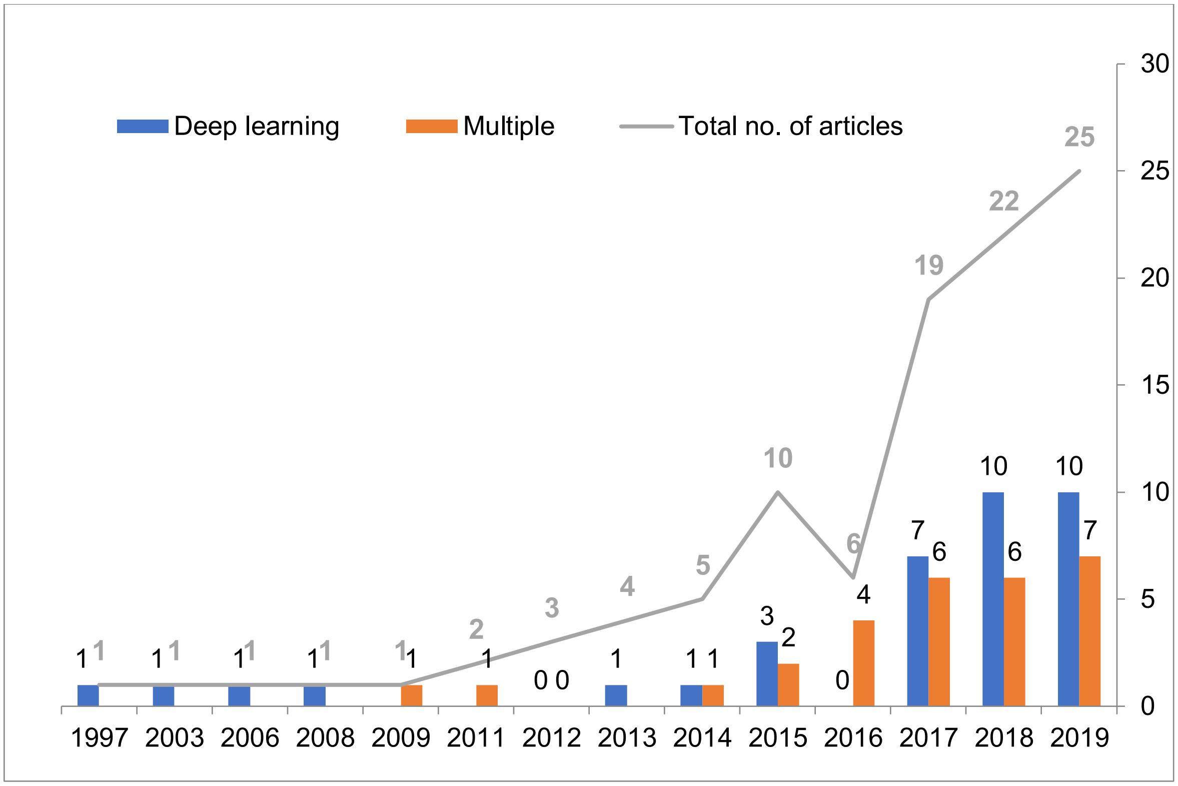

Figure 9: Number of papers per year published on CDSS using deep learning or multiple machine learning techniques in model development.

Table 1: Summary of key characteristics from scoping review. 'Country' represents the source of data or setting of the furthest development stage – which were the same for each study identified in this review. *ACS* = acute coronary syndrome. *AMS* = altered mental status*. COPD* = chronic obstructive pulmonary disease. *CT* = computed tomography. *LBP* = lower back pain*. LGIB* = lower gastrointestinal bleed. *LOS* = length of stay. *ML* = machine learning. *NLP* = natural language processing. *PE* = pulmonary embolism. *SVM* = support vector machine. *TBI* = traumatic brain injury*. UGIB* = upper gastrointestinal bleed. *UTI* = urinary tract infection.

Domain	Clinical focus	Country	Year	Furthest dev'l	ML algorithm	Citation
Capacity planning		Australia	2020	Internal validation	Random forest	Duwalage, K.I., Burkett, E., White, G., Wong, A. and Thompson, M.H. (2020), Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. Emergency Medicine Australasia. doi:10.1111/1742-6723.13481
		Turkey	2020	Internal validation	Deep learning	Nas, S., & Koyuncu, M. (2019). Emergency Department Capacity Planning: A Recurrent Neural Network and Simulation Approach. Computational and mathematical methods in medicine, 2019, 4359719. https://doi.org/10.1155/2019/4359719
Diagnosis	ACS	England	1997	External validation	Deep learning	Kennedy, R. L., Harrison, R. F., Burton, A. M., Fraser, H. S., Hamer, W. G., MacArthur, D., … Steedman, D. J. (1997). An artificial neural network system for diagnosis of acute myocardial infarction (AMI) in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. Computer Methods and Programs in Biomedicine, 52(2), 93–103. doi:10.1016/s0169-2607(96)01782-8
	Appendicitis	Taiwan	2011	Internal validation	Multiple	Hsieh, C.-H., Lu, R.-H., Lee, N.-H., Chiu, W.-T., Hsu, M.-H., & Li, Y.-C. (Jack). (2011). Novel solutions for an old disease: Diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. Surgery, 149(1), 87–93. doi:10.1016/j.surg.2010.03.023
	Sepsis	US	2017	Internal validation	SVM	Horng, S., Sontag, D. A., Halpern, Y., Jernite, Y., Shapiro, N. I., & Nathanson, L. A. (2017). Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLOS ONE, 12(4), e0174708. doi:10.1371/journal.pone.0174708
	AMS	US	2019	Internal validation	Deep learning	Obeid, J.S., Weeda, E.R., Matuskowitz, A.J. et al. Automated detection of altered mental status in emergency department clinical notes: a deep learning approach. BMC Med Inform Decis Mak 19, 164 (2019). https://doi.org/10.1186/s12911-019-0894-9
	Otolaryngology	US	2019	Internal validation	Deep learning	Livingstone, D., & Chau, J. (2019). Otoscopic diagnosis using computer vision: An automated machine learning approach. The Laryngoscope. doi:10.1002/lary.28292
	Abdominal free fluid	US	2016	Internal validation	SVM	Sjogren, A.R., Leo, M.M., Feldman, J. and Gwin, J.T. (2016), Image Segmentation and Machine Learning for Detection of Abdominal Free Fluid in Focused Assessment With Sonography for Trauma Examinations. Journal of Ultrasound in Medicine, 35: 2501-2509. doi:10.7863/ultra.15.11017
	Fracture	US	2018	Internal validation	Deep learning	Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., & Potter, H. (2018). Deep neural network improves fracture detection by clinicians. Proceedings of the National Academy of Sciences of the United States of America, 115(45), 11591–11596. https://doi.org/10.1073/pnas.1806905115

Diagnosis (cont.)	Falls	US	2019	Internal validation	NLP	Patterson, B. W., Jacobsohn, G. C., Shah, M. N., Song, Y., Maru, A., Venkatesh, A. K., Zhong, M., Taylor, K., Hamedani, A. G., & Mendonça, E. A. (2019). Development and validation of a pragmatic natural language processing approach to identifying falls in older adults in the emergency department. BMC medical informatics and decision-making, 19(1), 138. https://doi.org/10.1186/s12911-019-0843-7
	LBP red flags	Canada	2019	Development	Logistic regression	Hayden, J. A., Ogilvie, R., Stewart, S. A., French, S., Campbell, S., Magee, K., Slipp, P., Wells, G., & Stiell, I. (2019). Development of a clinical decision support tool for diagnostic imaging use in patients with low back pain: a study protocol. Diagnostic and prognostic research, 3, 1. https://doi.org/10.1186/s41512-019-0047-8
	Concussion	Canada	2018	Internal validation	Deep learning	Landry, A. P., Ting, W. K. C., Zador, Z., Sadeghian, A., & Cusimano, M. D. (2018). Using artificial neural networks to identify patients with concussion and postconcussion syndrome based on antisaccades. Journal of Neurosurgery, 1–8. doi:10.3171/2018.6.jns18607
	Infection	China	2018	Internal validation	Multiple	Tou, H., Yao, L., Wei, Z., Zhuang, X., & Zhang, B. (2018). Automatic infection detection based on electronic medical records. BMC bioinformatics, 19(Suppl 5), 117. https://doi.org/10.1186/s12859-018-2101-x
	TBI	US	2017	Internal validation	Decision trees	Peacock, W. F., 4th, Van Meter, T. E., Mirshahi, N., Ferber, K., Gerwien, R., Rao, V., Sair, H. I., Diaz-Arrastia, R., & Korley, F. K. (2017). Derivation of a Three Biomarker Panel to Improve Diagnosis in Patients with Mild Traumatic Brain Injury. Frontiers in neurology, 8, 641. https://doi.org/10.3389/fneur.2017.00641
	Influenza	US	2017	Internal validation	Naïve Bayes	Ye, Y., Wagner, M. M., Cooper, G. F., Ferraro, J. P., Su, H., Gesteland, P. H., Haug, P. J., Millett, N. E., Aronis, J. M., Nowalk, A. J., Ruiz, V. M., López Pineda, A., Shi, L., Van Bree, R., Ginter, T., & Tsui, F. (2017). A study of the transferability of influenza case detection systems between two large healthcare systems. PloS one, 12(4), e0174970. https://doi.org/10.1371/journal.pone.0174970
	ACS	Turkey	2016	Internal validation	Multiple	Berikol, G. B., Yildiz, O., & Özcan, İ. T. (2016). Diagnosis of Acute Coronary Syndrome with a Support Vector Machine. Journal of Medical Systems, 40(4). doi:10.1007/s10916-016-0432-6
	ACS	Sweden	2006	Internal validation	Deep learning	Green, M., Björk, J., Forberg, J., Ekelund, U., Edenbrandt, L., & Ohlsson, M. (2006). Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artificial Intelligence in Medicine, 38(3), 305–318. doi:10.1016/j.artmed.2006.07.006
	Atrial fibrillation	US	2019	Internal validation	Deep learning	Smith, S. W., Rapin, J., Li, J., Fleureau, Y., Fennell, W., Walsh, B. M., Rosier, A., Fiorina, L., & Gardella, C. (2019). A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation. International journal of cardiology. Heart & vasculature, 25, 100423. https://doi.org/10.1016/j.ijcha.2019.100423
	ACS	Taiwan	2019	Internal validation	Deep learning	Wu, C.-C., Hsu, W.-D., Islam, M. M., Poly, T. N., Yang, H.-C., Nguyen, P.-A. (Alex), … Li, Y.-C. (Jack). (2019). An artificial intelligence approach to early predict non-ST-elevation myocardial infarction Patients with Chest Pain. Computer Methods and Programs in Biomedicine. doi:10.1016/j.cmpb.2019.01.013

Diagnosis (cont.)	Neck injury	Turkey	2008	Internal validation	Deep learning	Bektaş, F., Eken, C., Soyuncu, S., Kilicaslan, İ., & Cete, Y. (2008). Artificial neural network in predicting craniocervical junction injury: an alternative approach to trauma patients. European Journal of Emergency Medicine, 15(6), 318–323. doi:10.1097/mej.0b013e3282fce7af.
	Renal colic	Turkey	2009	Internal validation	Multiple	Eken, C., Bilge, U., Kartal, M., & Eray, O. (2009). Artificial neural network, genetic algorithm, and logistic regression applications for predicting renal colic in emergency settings. International Journal of Emergency Medicine, 2(2), 99–105. doi:10.1007/s12245-009-0103-1 doi:10.1097/mej.0b013e3282fce7af.
	Sepsis	England	2018	External validation	Logistic regression	Faisal, M., Scally, A., Richardson, D., Beatson, K., Howes, R., Speed, K., & Mohammed, M. A. (2018). Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results*. Critical Care Medicine, 46(4), 612–618. doi:10.1097/ccm.0000000000002967
	PE	England	2015	Internal validation	Deep learning	Rucco, M., Sousa-Rodrigues, D., Merelli, E., Johnson, J. H., Falsetti, L., Nitti, C., & Salvi, A. (2015). Neural hypernetwork approach for pulmonary embolism diagnosis. BMC research notes, 8, 617. https://doi.org/10.1186/s13104-015-1554-5
	UTI	US	2018	Internal validation	Multiple	Taylor, R. A., Moore, C. L., Cheung, K. H., & Brandt, C. (2018). Predicting urinary tract infections in the emergency department with machine learning. PloS one, 13(3), e0194085. https://doi.org/10.1371/journal.pone.0194085
Disposition		US	2017	Internal validation	Deep learning	Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Patzer, A., & Schrager, J. D. (2017). Prediction of Emergency Department Hospital Admission Based on Natural Language Processing and Neural Networks . Methods of Information in Medicine, 56(05), 377–389. doi:10.3414/me17-01-0024
		US	2018	Internal validation	Multiple	Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. PloS one, 13(7), e0201016. https://doi.org/10.1371/journal.pone.0201016
		US	2015	Internal validation	Deep learning	Handly, N., Thompson, D., Li, J., Chuirazzi, D. & Venkat, A. (2014). Evaluation of a hospital admission prediction model adding coded chief complaint data using neural network methodology. European journal of emergency medicine. 22(10).
		US	2019	Internal validation	Multiple	Araz, O. M., Olson, D., & Ramirez-Nafarrate, A. (2019). Predictive analytics for hospital admissions from the emergency department using triage information. International Journal of Production Economics, 208, 199–207. doi: 10.1016/j.ijpe.2018.11.024
		Singapore	2018	Internal validation	Logistic regression	Parker, C. A., Liu, N., Wu, S. X., Shen, Y., Lam, S. S. W., & Ong, M. E. H. (2018). Predicting hospital admission at the emergency department triage: A novel prediction model. The American Journal of Emergency Medicine. doi:10.1016/j.ajem.2018.10.060
	COPD	US	2018	Internal validation	Multiple	Goto, T., Camargo, C. A., Faridi, M. K., Yun, B. J., & Hasegawa, K. (2018). Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. The American Journal of Emergency Medicine, 36(9), 1650–1654. doi:10.1016/j.ajem.2018.06.062
		Singapore	2011	Internal validation	Logistic regression	Sun, Y., Heng, B.H., Tay, S.Y. and Seow, E. (2011), Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data. Academic Emergency Medicine, 18: 844-850. doi:10.1111/j.1553-2712.2011.01125.x

Disposition (cont.)		US	2013	Internal validation	Logistic regression	Peck, J. S., Gaehde, S. A., Nightingale, D. J., Gelman, D. Y., Huckins, D. S., Lemons, M. F., … Benneyan, J. C. (2013). Generalizability of a Simple Approach for Predicting Hospital Admission From an Emergency Department. Academic Emergency Medicine, 20(11), 1156–1163. doi: 10.1111/acem.12244
		Brazil	2017	Internal validation	Multiple	Lucini, F. R., S. Fogliatto, F., C. da Silveira, G. J., L. Neyeloff, J., Anzanello, M. J., de S. Kuchenbecker, R., & D. Schaan, B. (2017). Text mining approach to predict hospital admissions using early medical records from the emergency department. International Journal of Medical Informatics, 100, 1–8. doi:10.1016/j.ijmedinf.2017.01.001
		US	2019	Internal validation	Decision trees	Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2019). Predicting 72-hour and 9-day return to the emergency department using machine learning. JAMIA open, 2(3), 346–352. https://doi.org/10.1093/jamiaopen/ooz019
		US	2019	Internal validation	Deep learning	Sterling, N. W., Patzer, R. E., Di, M., & Schrager, J. D. (2019). Prediction of Emergency Department Patient Disposition Based on Natural Language Processing of Triage Notes. International Journal of Medical Informatics. doi:10.1016/j.ijmedinf.2019.06.008
		US	2019	Internal validation	Logistic regression	Lee, S.-Y., Chinnam, R. B., Dalkiran, E., Krupp, S., & Nauss, M. (2019). Prediction of emergency department patient disposition decision for proactive resource allocation for admission. Health Care Management Science. doi:10.1007/s10729-019-09496-y
		Australia	2018	Internal validation	Multiple	Rendell, K., Koprinska, I., Kyme, A., Ebker-White, A. A., & Dinh, M. M. (2018). The Sydney Triage to Admission Risk Tool (START2) using machine learning techniques to support disposition decision-making. Emergency Medicine Australasia. doi:10.1111/1742-6723.13199
Documentation	Chief complaints	US	2018	Internal validation	Deep learning	Lee S. H. (2018). Natural language generation for electronic health records. NPJ digital medicine, 1, 63. https://doi.org/10.1038/s41746-018-0070-0
	Chief complaints	US	2019	Implementation	SVM	Greenbaum, N. R., Jernite, Y., Halpern, Y., Calder, S., Nathanson, L. A., Sontag, D., & Horng, S. (2019). Improving Documentation of Presenting Problems in the Emergency Department using a Domain-specific Ontology and Machine Learning-Driven User Interfaces. International Journal of Medical Informatics, 103981. doi:10.1016/j.ijmedinf.2019.103981
LOS		US	2015	Internal validation	Deep learning	Azari, A., Janeja, V. P., & Levin, S. (2015). Imbalanced learning to predict long stay Emergency Department patients. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). doi:10.1109/bibm.2015.7359790
		Australia	2018	Internal validation	Logistic regression	Street, M., Mohebbi, M., Berry, D., Cross, A., & Considine, J. (2018). Influences on emergency department length of stay for older people. European Journal of Emergency Medicine, 25(4), 242–249. doi:10.1097/mej.0000000000000452
		US	2019	Internal validation	Multiple	Haley S Hunter-Zinck, Jordan S Peck, Tania D Strout, Stephan A Gaehde, Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay, Journal of the American Medical Informatics Association, Volume 26, Issue 12, December 2019, Pages 1427–1436, https://doi.org/10.1093/jamia/ocz171
Mortality	Sepsis	US	2015	Internal validation	Multiple	Taylor, R. A., Pare, J. R., Venkatesh, A. K., Mowafi, H., Melnick, E. R., Fleischman, W., & Hall, M. K. (2016). Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Academic emergency medicine : official journal of the Society for

						Academic Emergency Medicine, 23(3), 269–278. https://doi.org/10.1111/acem.12876
Mortality (cont.)	ACS	Israel	2017	Internal validation	Multiple	Shouval, R., Hadanny, A., Shlomo, N., Iakobishvili, Z., Unger, R., Zahger, D., … Beigel, R. (2017). Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: An Acute Coronary Syndrome Israeli Survey data mining study. International Journal of Cardiology, 246, 7–13. doi:10.1016/j.ijcard.2017.05.067
	Trauma	US	2017	Internal validation	Decision trees	Sefrioui, I., Amadini, R., Mauro, J., El Fallahi, A., & Gabbrielli, M. (2017). Survival prediction of trauma patients: a study on US National Trauma Data Bank. European Journal of Trauma and Emergency Surgery, 43(6), 805–822. doi:10.1007/s00068-016-0757-3
	ACS	China	2017	Internal validation	Multiple	Yang, J., Li, X., Chen, T., Li, Y., Xie, G., & Yang, Y. (2018). Machine Learning Models To Predict In-Hospital Mortality For St-Elevation Myocardial Infarction: From China Acute Myocardial Infarction (Cami) Registry. Journal of the American College of Cardiology, 71(11). doi: 10.1016/s0735-1097(18)30777-0
	Sepsis	US	2014	Internal validation	Multiple	Eren Gultepe, Jeffrey P Green, Hien Nguyen, Jason Adams, Timothy Albertson, Ilias Tagkopoulos, From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system, Journal of the American Medical Informatics Association, Volume 21, Issue 2, March 2014, Pages 315–325, https://doi.org/10.1136/amiajnl-2013-001815
		Sweden	2019	External validation	Multiple	Blom, M. C., Ashfaq, A., Sant'Anna, A., Anderson, P. D., & Lingman, M. (2019). Training machine learning models to predict 30-day mortality in patients discharged from the emergency department: a retrospective, population-based registry study. BMJ open, 9(8), e028015. https://doi.org/10.1136/bmjopen-2018-028015
		Switzerland	2015	External validation	Multiple	Jenny, M. A., Hertwig, R., Ackermann, S., Messmer, A. S., Karakoumis, J., Nickel, C. H., & Bingisser, R. (2015). Are Mortality and Acute Morbidity in Patients Presenting With Nonspecific Complaints Predictable Using Routine Variables? Academic Emergency Medicine, 22(10), 1155–1163. doi: 10.1111/acem.12755
	Sepsis	Korea	2019	Internal validation	Deep learning	Perng, J. W., Kao, I. H., Kung, C. T., Hung, S. C., Lai, Y. H., & Su, C. M. (2019). Mortality Prediction of Septic Patients in the Emergency Department Based on Machine Learning. Journal of clinical medicine, 8(11), 1906. https://doi.org/10.3390/jcm8111906
		Israel	2019	Internal validation	Decision trees	Klug, M., Barash, Y., Bechler, S., Resheff, Y. S., Tron, T., Ironi, A., … Klang, E. (2019). A Gradient Boosting Machine Learning Model for Predicting Early Mortality in the Emergency Department Triage: Devising a Nine-Point Triage Score. Journal of General Internal Medicine. doi:10.1007/s11606-019-05512-7
	UGIB	Multiple	2020	External validation	Decision trees	Shung, D. L., Au, B., Taylor, R. A., Tay, J. K., Laursen, S. B., Stanley, A. J., … Laine, L. (2020). Validation of a Machine Learning Model That Outperforms Clinical Risk Scoring Systems for Upper Gastrointestinal Bleeding. Gastroenterology, 158(1), 160–167. doi: 10.1053/j.gastro.2019.09.009
	Sepsis	Singapore	2019	Internal validation	Multiple	Chiew, C. J., Liu, N., Tagami, T., Wong, T. H., Koh, Z. X., & Ong, M. (2019). Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. Medicine, 98(6), e14197. https://doi.org/10.1097/MD.0000000000014197

Mortality (cont.)		England	2018	External validation	Multiple	Faisal, M., Scally, A., Howes, R., Beatson, K., Richardson, D., & Mohammed, M. A. (2018). A comparison of logistic regression models with alternative machine learning methods to predict the risk of in-hospital mortality in emergency medical admissions via external validation. Health Informatics Journal, 146045821881360. doi:10.1177/1460458218813600
Outcome prediction	Cardiac arrest	Singapore	2012	Internal validation	SVM	Ong, M.E.H., Lee Ng, C.H., Goh, K. et al. Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score. Crit Care 16, R108 (2012). https://doi.org/10.1186/cc11396
	LGIB	US	2003	External validation	Deep learning	Das, A., Ben-Menachem, T., Cooper, G. S., Chak, A., Sivak, M. V., Gonet, J. A., & Wong, R. C. (2003). Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. The Lancet, 362(9392), 1261–1266. doi:10.1016/s0140-6736(03)14568-0
	ACS	Singapore	2014	Internal validation	Decision trees	Liu, N., Koh, Z. X., Chua, E. C.-P., Tan, L. M.-L., Lin, Z., Mirza, B., & Ong, M. E. H. (2014). Risk Scoring for Prediction of Acute Cardiac Complications from Imbalanced Clinical Data. IEEE Journal of Biomedical and Health Informatics, 18(6), 1894–1902. doi:10.1109/jbhi.2014.2303481
	Sepsis	US	2018	External validation	Decision trees	Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., Shieh, L., Chettipally, U., Fletcher, G., Kerem, Y., Zhou, Y., & Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ open, 8(1), e017833. https://doi.org/10.1136/bmjopen-2017-017833
	Shock	US	2013	Internal validation	Cluster analysis	Convertino, V. A., Grudic, G., Mulligan, J., & Moulton, S. (2013). Estimation of individual-specific progression to impending cardiovascular instability using arterial waveforms. Journal of Applied Physiology, 115(8), 1196–1202. doi: 10.1152/japplphysiol.00668.2013
	Cardiac arrest	Singapore	2015	Internal validation	Manifold ranking	Liu, T., Lin, Z., Ong, M. E. H., Koh, Z. X., Pek, P. P., Yeo, Y. K., … Liu, N. (2015). Manifold ranking based scoring system with its application to cardiac arrest prediction: A retrospective study in emergency department patients. Computers in Biology and Medicine, 67, 74–82. doi:10.1016/j.compbiomed.2015.10.001
	ACS	US	2014	Internal validation	Random forest	VanHouten, J. P., Starmer, J. M., Lorenzi, N. M., Maron, D. J., & Lasko, T. A. (2014). Machine learning for risk prediction of acute coronary syndrome. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2014, 1940–1949.
	Trauma	China	2020	Internal validation	Random forest	Li, K., Wu, H., Pan, F., Chen, L., Feng, C., Liu, Y., … Li, T. (2020). A Machine Learning–Based Model to Predict Acute Traumatic Coagulopathy in Trauma Patients Upon Emergency Hospitalization. Clinical and Applied Thrombosis/Hemostasis. https://doi.org/10.1177/1076029619897827
	Sepsis	Korea	2020	Internal validation	Multiple	Kim, J., Chang, H., Kim, D., Jang, D.-H., Park, I., & Kim, K. (2020). Machine learning for prediction of septic shock at initial triage in emergency department. Journal of Critical Care, 55, 163–170. doi: 10.1016/j.jcrc.2019.09.024
	Sepsis	US	2019	Internal validation	Naïve Bayes	Gupta, A., Liu, T., & Shepherd, S. (2019). Clinical decision support system to assess the risk of sepsis using Tree Augmented Bayesian networks and electronic medical record data. Health Informatics Journal. https://doi.org/10.1177/1460458219852872

Outcome prediction (cont.)	ACS	China	2017	Internal validation	Multiple	Hu D, Huang Z, Chan TM, et al. Acute Coronary Syndrome Risk Prediction Based on GRACE Risk Score. Studies in Health Technology and Informatics. 2017 ;245:398-402.
	Neurosurgical intervention	US	2018	Internal validation	Logistic regression	Orlando, A., Levy, A. S., Rubin, B. A., Tanner, A., Carrick, M. M., Lieser, M., … Bar-Or, D. (2018). Isolated subdural hematomas in mild traumatic brain injury. Part 2: a preliminary clinical decision support tool for neurosurgical intervention. Journal of Neurosurgery, 1–8. doi:10.3171/2018.1.jns171906
	Cardiac arrest	Korea	2019	Internal validation	Deep learning	Jang, D.-H., Kim, J., Jo, Y. H., Lee, J. H., Hwang, J. E., Park, S. M., … Chang, H. (2019). Developing neural network models for early detection of cardiac arrest in emergency department. The American Journal of Emergency Medicine. doi:10.1016/j.ajem.2019.04.006
	Sepsis	US	2017	Internal validation	Logistic regression	Danner, O. K., Hendren, S., Santiago, E., Nye, B., & Abraham, P. (2017). Physiologically-based, predictive analytics using the heart-rate-to-Systolic-Ratio significantly improves the timeliness and accuracy of sepsis prediction compared to SIRS. The American Journal of Surgery, 213(4), 617–621. doi:10.1016/j.amjsurg.2017.01.006
	LGIB	England	2015	External validation	Decision trees	Ayaru, L., Ypsilantis, P. P., Nanapragasam, A., Choi, R. C., Thillanathan, A., Min-Ho, L., & Montana, G. (2015). Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting. PloS one, 10(7), e0132485. https://doi.org/10.1371/journal.pone.0132485
	Stroke	England	2014	Internal validation	SVM	Bentley, P., Ganesalingam, J., Carlton Jones, A. L., Mahady, K., Epton, S., Rinne, P., Sharma, P., Halse, O., Mehta, A., & Rueckert, D. (2014). Prediction of stroke thrombolysis outcome using CT brain machine learning. NeuroImage. Clinical, 4, 635–640. https://doi.org/10.1016/j.nicl.2014.02.003
Radiology	Head CT	India	2018	Internal validation	Deep learning	Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., … Warier, P. (2018). Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. The Lancet. doi:10.1016/s0140-6736(18)31645-3
	Chest radiographs	US	2018	Internal validation	Deep learning	Singh, R., Kalra, M. K., Nitiwarangkul, C., Patti, J. A., Homayounieh, F., Padole, A., Rao, P., Putha, P., Muse, V. V., Sharma, A., & Digumarthy, S. R. (2018). Deep learning in chest radiography: Detection of findings and presence of change. *PloS one*, *13*(10), e0204155. https://doi.org/10.1371/journal.pone.0204155
	Chest radiographs	US	2018	Internal validation	Deep learning	Abiyev, R. H., & Ma'aitah, M. (2018). Deep Convolutional Neural Networks for Chest Diseases Detection. Journal of healthcare engineering, 2018, 4168538. https://doi.org/10.1155/2018/4168538
	Wrist fractures	England	2018	Internal validation	Deep learning	Kim, D. H., & MacKinnon, T. (2018). Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clinical Radiology, 73(5), 439–445. doi:10.1016/j.crad.2017.11.015
	PE	US	2017	External validation	Deep learning	Chen, M. C., Ball, R. L., Yang, L., Moradzadeh, N., Chapman, B. E., Larson, D. B., … Lungren, M. P. (2018). Deep Learning to Classify Radiology Free-Text Reports. Radiology, 286(3), 845–852. doi:10.1148/radiol.2017171115
	Stroke	US	2017	Internal validation	Deep learning	Prevedello, L. M., Erdal, B. S., Ryu, J. L., Little, K. J., Demirer, M., Qian, S., & White, R. D. (2017). Automated Critical Test Findings Identification and Online Notification System Using Artificial Intelligence in Imaging. Radiology, 285(3), 923–931. doi:10.1148/radiol.2017162664

Radiology (cont.)	Stroke	China	2012	Internal validation	Naïve Bayes	Li, Y.-H., Zhang, L., Hu, Q.-M., Li, H.-W., Jia, F.-C., & Wu, J.-H. (2011). Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans. International Journal of Computer Assisted Radiology and Surgery, 7(4), 507–516. doi:10.1007/s11548-011-0664-3
	Fracture	Sweden	2017	Internal validation	Deep learning	Olczak, J., Fahlberg N., Maki, A., Razavian, A.S., Jilert, A., Stark, A., Sköldenberg, O. & Gordon, M. (2017). Artificial intelligence for analyzing orthopedic trauma radiographs, Acta Orthopaedica, 88(6), 581-586, DOI: 10.1080/17453674.2017.1344459
	Fracture	Multiple	2017	Internal validation	Deep learning	Jamaludin, A., Lootus, M., Kadir, T. et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. Eur Spine J 26, 1374–1383 (2017). https://doi.org/10.1007/s00586-017-4956-3
	TBI	US	2017	Internal validation	Decision trees	Farzaneh, N., Soroushmehr, S. M. R., Williamson, C. A., Jiang, C., Srinivasan, A., Bapuraj, J. R., … Najarian, K. (2017). Automated subdural hematoma segmentation for traumatic brain injured (TBI) patients. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2017.8037505
	Abdominal ultrasound	US	2017	Internal validation	Deep learning	Cheng, P. M., & Malhi, H. S. (2017). Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images. Journal of digital imaging, 30(2), 234–243. https://doi.org/10.1007/s10278-016-9929-2
	TBI	US	2016	Internal validation	Decision trees	Molaei, S., Korley, F. K., Soroushmehr, S. M. R., Falk, H., Sair, H., Ward, K., & Najarian, K. (2016). A machine learning based approach for identifying traumatic brain injury patients for whom a head CT scan can be avoided. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2016.7591179
	PE	US	2019	External validation	Deep learning	Banerjee I, Sofela M, Yang J, et al. Development and Performance of the Pulmonary Embolism Result Forecast Model (PERFORM) for Computed Tomography Clinical Decision Support. JAMA Netw Open. 2019;2(8):e198719. doi:10.1001/jamanetworkopen.2019.8719
	PE	US	2012	Implementation	NLP	Raja, A. S., Ip, I. K., Prevedello, L. M., Sodickson, A. D., Farkas, C., Zane, R. D., … Khorasani, R. (2012). Effect of Computerized Clinical Decision Support on the Use and Yield of CT Pulmonary Angiography in the Emergency Department. Radiology, 262(2), 468–474. doi: 10.1148/radiol.11110951
	Fracture	Australia	2018	Internal validation	Deep learning	Hassanzadeh, H., Nguyen, A., Karimi, S., & Chu, K. (2018). Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection from radiology reports. Journal of Biomedical Informatics, 85, 68–79. doi:10.1016/j.jbi.2018.07.017
	Orbital blowout fractures	China	2019	Internal validation	Deep learning	Li, L., Song, X., Guo, Y., Liu, Y., Sun, R., Zou, H., … Fan, X. (2019). Deep Convolutional Neural Networks for Automatic Detection of Orbital Blowout Fractures. Journal of Craniofacial Surgery, 1. doi:10.1097/scs.0000000000006069
	Chest radiographs	Korea	2019	Internal validation	Deep learning	Hwang, E. J., Nam, J. G., Lim, W. H., Park, S. J., Jeong, Y. S., Kang, J. H., … Park, C. M. (2019). Deep Learning for Chest Radiograph Diagnosis in the Emergency Department. Radiology, 293(3), 573–580. doi: 10.1148/radiol.2019191225

Radiology (cont.)	Trauma	England	2018	Internal validation	Deep learning	Al Arif, S. M. M. R., Knapp, K., & Slabaugh, G. (2018). Fully automatic cervical vertebrae segmentation framework for X-ray images. Computer Methods and Programs in Biomedicine, 157, 95–111. doi:10.1016/j.cmpb.2018.01.006
Intervention	Defibrillation timing	US	2016	Internal validation	Multiple	Shandilya, S., Kurz, M. C., Ward, K. R., & Najarian, K. (2016). Integration of Attributes from Non-Linear Characterization of Cardiovascular Time-Series for Prediction of Defibrillation Outcomes. PloS one, 11(1), e0141313. https://doi.org/10.1371/journal.pone.0141313
	Resuscitation	US	2014	Development	Deep learning	Liu, N. T., Holcomb, J. B., Wade, C. E., Darrah, M. I., & Salinas, J. (2014). Utility of Vital Signs, Heart Rate Variability and Complexity, and Machine Learning for Identifying the Need for Lifesaving Interventions in Trauma Patients. Shock, 42(2), 108–114. doi:10.1097/shk.0000000000000186
	Ectopic pregnancy	Spain	2019	Internal validation	Multiple	Ramón Fernández, A. D., Fernández, D. R., & Prieto Sánchez, M. T. (2019). A Decision Support System for predicting the treatment of ectopic pregnancies. International Journal of Medical Informatics. doi:10.1016/j.ijmedinf.2019.06.002
	Sepsis	US	2015	Internal validation	SVM	Tsoukalas, A., Albertson, T., & Tagkopoulos, I. (2015). From data to optimal decision-making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. JMIR medical informatics, 3(1), e11. https://doi.org/10.2196/medinform.3445
	Intubation	US	2016	Internal validation	Multiple	Carlson, J. N., Das, S., De la Torre, F., Frisch, A., Guyette, F. X., Hodgins, J. K., & Yealy, D. M. (2016). A Novel Artificial Intelligence System for Endotracheal Intubation. Prehospital Emergency Care, 20(5), 667–671. doi:10.3109/10903127.2016.1139220
	Pneumonia	US	2019	Implementation	Naïve Bayes	Jones, B. E., Collingridge, D. S., Vines, C. G., Post, H., Holmen, J., Allen, T. L., Haug, P., Weir, C. R., & Dean, N. C. (2019). CDS in a Learning Health Care System: Identifying Physicians' Reasons for Rejection of Best-Practice Recommendations in Pneumonia through Computerized Clinical Decision Support. Applied clinical informatics, 10(1), 1–9. https://doi.org/10.1055/s-0038-1676587
	Pneumonia	US	2015	External validation	Logistic regression	Dean, N. C., Jones, B. E., Jones, J. P., Ferraro, J. P., Post, H. B., Aronsky, D., … Haug, P. J. (2015). Impact of an Electronic Clinical Decision Support Tool for Emergency Department Patients With Pneumonia. Annals of Emergency Medicine, 66(5), 511–520. doi:10.1016/j.annemergmed.2015.02.003
Triage		Spain	2016	Internal validation	Multiple	Zlotnik, A., Alfaro, M. C., Pérez, M. C. P., Gallardo-Antolín, A., & Martínez, J. M. M. (2016). Building a Decision Support System for Inpatient Admission Prediction With the Manchester Triage System and Administrative Check-in Variables. CIN: Computers, Informatics, Nursing, 34(5), 224–230. doi:10.1097/cin.0000000000000230
		Malaysia	2013	Internal validation	Deep learning	Azeez, D., Ali, M. A., Gan, K. B., & Saiboon, I. (2013). Comparison of adaptive neuro-fuzzy inference system and artificial neutral networks model to categorize patients in the emergency department. SpringerPlus, 2, 416. https://doi.org/10.1186/2193-1801-2-416
		US	2019	Internal validation	Multiple	Raita, Y., Goto, T., Faridi, M.K. et al. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 23, 64 (2019). https://doi.org/10.1186/s13054-019-2351-7
		Taiwan	2013	Internal validation	SVM	Wang, S.-T. (2013). Construct an Optimal Triage Prediction Model: A Case Study of the Emergency Department of a Teaching Hospital in Taiwan. Journal of Medical Systems, 37(5). doi:10.1007/s10916-013-9968-x

Triage (cont.)	Pain	Taiwan	2017	Development	Deep learning	F. Tsai, Y. Weng, C. Ng and C. Lee, "Embedding stacked bottleneck vocal features in a LSTM architecture for automatic pain level classification during emergency triage," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, 2017, pp. 313-318.
		Malaysia	2015	Internal validation	Decision trees	Azeez, D., Gan, K. B., Ali, M. A. M., & Ismail, M. S. (2015). Secondary triage classification using an ensemble random forest technique. Technology and Health Care, 23(4), 419–428. doi:10.3233/thc-150907
		US	2018	Internal validation	Random forest	Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., … Kelen, G. (2018). Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. Annals of Emergency Medicine, 71(5), 565–574.e2. doi:10.1016/j.annemergmed.2017.08.005
	COPD	US	2017	Internal validation	Multiple	Swaminathan, S., Qirko, K., Smith, T., Corcoran, E., Wysham, N. G., Bazaz, G., Kappel, G., & Gerber, A. N. (2017). A machine learning approach to triaging patients with chronic obstructive pulmonary disease. PloS one, 12(11), e0188532. https://doi.org/10.1371/journal.pone.0188532
	Acute abdomen	Iran	2017	Internal validation	Multiple	Farahmand, S., Shabestari, O., Pakrah, M., Hossein-Nejad, H., Arbab, M., & Bagheri-Hariri, S. (2017). Artificial Intelligence-Based Triage for Patients with Acute Abdominal Pain in Emergency Department; a Diagnostic Accuracy Study. Advanced journal of emergency medicine, 1(1), e5. https://doi.org/10.22114/AJEM.v1i1.11
		Korea	2020	Internal validation	Deep learning	Yu, J. Y., Jeong, G. Y., Jeong, O. S., Chang, D. K., & Cha, W. C. (2020). Machine Learning and Initial Nursing Assessment-Based Triage System for Emergency Department. Healthcare informatics research, 26(1), 13–19. https://doi.org/10.4258/hir.2020.26.1.13
		Korea	2019	Internal validation	Multiple	Choi, S. W., Ko, T., Hong, K. J., & Kim, K. H. (2019). Machine Learning-Based Prediction of Korean Triage and Acuity Scale Level in Emergency Department Patients. Healthcare informatics research, 25(4), 305–312. https://doi.org/10.4258/hir.2019.25.4.305
		Korea	2018	External validation	Deep learning	Kwon, J. M., Lee, Y., Lee, Y., Lee, S., Park, H., & Park, J. (2018). Validation of deep-learning-based triage and acuity score using a large national dataset. PloS one, 13(10), e0205836. https://doi.org/10.1371/journal.pone.0205836

Table 2: Objectives, results and limitations identified in publications on CDSS developed in England from review search. *AKI* = acute kidney injury. *AMI* = acute myocardial infarction. *AMS* = altered mental status. *ANN* = artificial neural network. *APTT* = activated partial prothrombin time. *AUC* = area under receiver operating characteristic curve. *AVM* = arteriovenous malformation. *CAD* = coronary artery disease. *CLD* = chronic liver disease. *COPD* = chronic obstructive pulmonary disease. *CP* = chest pain. *DBP* = diastolic blood pressure. *DM* = diabetes mellitus. *DRE* = digital rectal exam. *DVT* = deep vein thrombosis. *ECG* = electrocardiogram. *FC* = functional class. *GBM* = generalized boosted regression modelling. *GI* = gastrointestinal. *HCT* = hematocrit. *HLD* = hyperlipidemia. *HGB* = hemoglobin. *HTN* = hypertension. *ICH* = intracranial hemorrhage. *LR* = logistic regression. *MPAP* = mean pulmonary artery pressure. *NEWS* = National Early Warning Score. *NIHSS* = National Institutes of Health Stroke Score. *NNET* = neural network. *NPV* = negative predictive value. *NSAID* = nonsteroidal anti-inflammatory drug. *PAD* = pulmonary artery diastolic pressure. *PAS* = pulmonary artery systolic pressure. *PPI* = proton pump inhibitor. *PPV* = positive predictive value. *RF* = random forest. *RR* = respiratory rate. *RVD* = right ventricular diastolic pressure. *SBP* = systolic blood pressure. *SVM* = support vector machine.

Objective	**Author** (Year; Source)	**Algorithm** Development	**Independent Variables**	**Results by outcome**	**Limitations**
Diagnose AMI earlier using clinical and ECG data at presentation	R.L. Kennedy et al. (1997; Computer Methods and Programs in Biomedicine)	Deep learning Internal and external validation	Age, smoking, family history of CAD, DM, HTN, HLD, chest pain symptoms, symptoms associated with AMI, timing, personal history of CAD, cardiopulmonary exam findings, ECG changes	Internal validation: accuracy 91.8% (91.76%-91.84%), sensitivity 91.2% (91.16%-91.24%), specificity 90.2% (90.16%-90.24%), PPV 84.9% (84.85%-84.95%) External validation: accuracy 73.6% (93.51%-73.69%), sensitivity 52.4% (52.30%-52.50%), specificity 80.0% (79.92%-80.08%), PPV 44.0% (43.90%-44.10%)	Decrease in performance in external setting Need to incorporate rapid advances in biochemical tests ANN are relatively computationally intensive Rules governing ANN not known to operator Reliance on accurate data input (not integrated with health record or ECG)
Predict the risk of sepsis following emergency medical admission using patient's first routinely collected vital signs and blood test results	M. Faisal et al. (2018; Critical Care Medicine)	LR Internal and external validation	Age, sex, NEWS (RR, oxygen saturation, supplemental oxygen, temperature, SBP, HR, level of consciousness), DBP, albumin, creatinine, Hgb, potassium, sodium, urea, WBC count, AKI score	Internal validation: AUC 0.779 (0.772-0.786), discrimination 0.162 External validation: AUC 0.788 (0.782-0.793), discrimination 0.186	24.2% of development and validation cohorts excluded due to missing blood data, and ~5% missing NEWS data

Predict the risk of death following emergency medicine admission using patient's first blood tests and physiological measurements	M. Faisal et al. (2018; Health Informatics Journal)	Multiple (LR, RF, GBM, SVM, NNET) Internal and external validation	Age, sex, NEWS (RR, oxygen saturation, supplemental oxygen, temperature, SBP, HR, level of consciousness), albumin, creatinine, Hgb, potassium, sodium, WBC count, urea	Internal validation: AUC LR 0.8712 (0.8557-0.8868), RF 0.8569 (0.8397-0.8741), GBM 0.8719 (0.8563-0.8875), SVM 0.8724 (0.859-0.8880), NNET 0.8722 (0.8566-0.8877) External validation: AUC LR 0.8470 (0.8351-0.8589), RF 0.8044 (0.7899-0.8189), GBM 0.8483 (0.8365-0.8601), SVM 0.8470 (0.8351-0.8590), NNET 0.8475 (0.8357-0.8594)	
Automatically identify and segment cervical vertebrae in X-ray images for injury detection	S. M. M. R. Al Arif et al. (2018; Computer Methods and Programs in Biomedicine)	Deep learning Internal validation	Numerous image features	17.1% more sensitive and 70x faster than previous tests Dice similarity coefficient 0.84; shape error 1.69mm	Unable to segment C1 and C2 Outliers (i.e., implants, severe cervical injury) perform poorly Additional training data formed via image manipulation
Predict recurrent bleeding, therapeutic intervention and severe bleeding for patients presenting to the ED with acute lower GI bleeding	L. Ayaru et al. (2015; PLOS ONE)	GBM Internal and external validation	Age; gender; use of PPI, or NSAIDs or anticoagulants; alcoholism; smoking; nursing home status; history of colorectal polyp, hemorrhoids, diventicular disease, colonic AVM, syncope, CAD, HTN, stroke, COPD, DM, dementia, cancer, CLD, GI bleed; HR, SBP, DBP, AMS, abdominal pain, gross blood on DRE, Hgb, Hct, WBC count, platelet,	Accuracy, Sensitivity, Specificity, PPV, NPV (%) Internal validation: recurrent bleeding 88, 67, 91, 50, 95; therapeutic intervention 88, 80, 89, 44, 98; severe bleeding 78, 73, 80 61, 88 External validation: recurrent bleeding 88, 57, 91, 50, 94; therapeutic intervention 91, 60, 92, 27, 98;	Patients transferred not included Only 170 and 130 patients included in internal and external validation sets, respectively Datasets differed significantly in severe bleeding and therapeutic intervention outcomes Model requires the input of a high number of

			APTT, urea, creatinine	severe bleeding 83, 57, 89, 58, 90	variables for accuracy Death, an infrequent outcome, was not included
Predict which patients receiving tPA for ischemic stroke will subsequently develop symptomatic ICH using CT images	P. Bentley et al. (2014 NeuroImage: Clinical)	SVM Internal validation	NIHSS score, raw CT images	AUC 0.744 (0.738-0.748)	Definition of SICH chosen to maximize cases (in practice, more conservative definitions often used) Small sample size, highly imbalanced to non-SICH cases Image-processing steps inefficient and timely
Automate distal radius and ulna fracture detection from plain radiographs	D.H. Kim, T. MacKinnon (2018; Clinical Radiology)	Deep learning Internal Validation	Numerous image features	AUC 0.954, sensitivity 0.9, specificity 0.88	Only one type of fracture, easily identified by radiologists, included Data augmentation used to increase sample size Reference standard is radiologist read Small degree in over-fitting
Predict the risk of pulmonary embolism from non-imaging clinical variables on presentation to the ED	M. Rucco et al. (2015; BMC Res Notes)	Deep learning Internal validation	Age, number of risk factors, previous DVT, palpitations, cough, positive D-dimer, PAS, PAD, MPAP, WBC count, cancer, positive troponin, FC, PAS/FC, RVD, Wells score, Revised Geneva score, Wicki score, dyspnea, CP, pCO2, pO2, pH, hemoptysis	AUC 0.93	Dataset limited to patients with high risk of pulmonary embolism