

University of Groningen

Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care

Medic, G; M, Kosaner Klie; Atallah, L; Weichert, J; Panda, S; Postma, M; EL-Kerdi, A

Published in:
 F1000Research

DOI:
[10.12688/f1000research.20498.1](https://doi.org/10.12688/f1000research.20498.1)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Medic, G., M, K. K., Atallah, L., Weichert, J., Panda, S., Postma, M., & EL-Kerdi, A. (2019). Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review [version 2; peer review: 2 approved]. *F1000Research*, 8(1728). <https://doi.org/10.12688/f1000research.20498.1>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



SYSTEMATIC REVIEW

REVISED Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review [version 2; peer review: 2 approved]

Goran Medic ^{1,2}, Melodi Kosaner Kließ³, Louis Atallah⁴, Jochen Weichert⁴, Saswat Panda³, Maarten Postma^{2,5,6}, Amer EL-Kerdi⁴

¹Health Economics, Philips, Eindhoven, Noord-Brabant, 5621JG, The Netherlands

²Department of Pharmacy, Unit of Pharmacotherapy, -Epidemiology & -Economics, University of Groningen, Groningen, 9700 AB, The Netherlands

³Global Market Access Solutions Sàrl, St-Prex, 1162, Switzerland

⁴Philips, Cambridge, MA, 02141, USA

⁵Department of Health Sciences, University Medical Centre Groningen, University of Groningen, Groningen, 9700 AB, The Netherlands

⁶Department of Economics, Econometrics & Finance, University of Groningen, Groningen, 9700 AB, The Netherlands

v2 First published: 08 Oct 2019, 8:1728 (<https://doi.org/10.12688/f1000research.20498.1>)
 Latest published: 27 Nov 2019, 8:1728 (<https://doi.org/10.12688/f1000research.20498.2>)

Abstract

Background: Clinical decision support (CDS) systems have emerged as tools providing intelligent decision making to address challenges of critical care. CDS systems can be based on existing guidelines or best practices; and can also utilize machine learning to provide a diagnosis, recommendation, or therapy course.

Methods: This research aimed to identify evidence-based study designs and outcome measures to determine the clinical effectiveness of clinical decision support systems in the detection and prediction of hemodynamic instability, respiratory distress, and infection within critical care settings. PubMed, ClinicalTrials.gov and Cochrane Database of Systematic Reviews were systematically searched to identify primary research published in English between 2013 and 2018. Studies conducted in the USA, Canada, UK, Germany and France with more than 10 participants per arm were included.

Results: In studies on hemodynamic instability, the prediction and management of septic shock were the most researched topics followed by the early prediction of heart failure. For respiratory distress, the most popular topics were pneumonia detection and prediction followed by pulmonary embolisms. Given the importance of imaging and clinical notes, this area combined Machine Learning with image analysis and natural language processing. In studies on infection, the most researched areas were the detection, prediction, and management of sepsis, surgical site infections, as well as acute kidney injury. Overall, a variety of Machine Learning algorithms were utilized frequently, particularly support vector machines, boosting techniques, random forest classifiers and neural

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
REVISED		
version 2	report	report
published 27 Nov 2019		
version 1		
published 08 Oct 2019	report	report

1 **Stavros Nikolakopoulos** , University Medical Center Utrecht, Utrecht, The Netherlands

2 **Milena Kovacevic** , University of Belgrade, Belgrade, Serbia

Any reports and responses or comments on the article can be found at the end of the article.

networks. Sensitivity, specificity, and ROC AUC were the most frequently reported performance measures.

Conclusion: This review showed an increasing use of Machine Learning for CDS in all three areas. Large datasets are required for training these algorithms; making it imperative to appropriately address, challenges such as class imbalance, correct labelling of data and missing data.

Recommendations are formulated for the development and successful adoption of CDS systems.

Keywords

sepsis, hemodynamic instability, respiratory distress, infection, machine learning, clinical trials, critical care.

Corresponding author: Goran Medic (goran.medic@philips.com)

Author roles: **Medic G:** Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation; **Kosaner Kließ M:** Data Curation, Formal Analysis, Methodology, Project Administration, Validation, Writing – Review & Editing; **Atallah L:** Writing – Original Draft Preparation, Writing – Review & Editing; **Weichert J:** Writing – Review & Editing; **Panda S:** Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing; **Postma M:** Conceptualization, Supervision, Writing – Review & Editing; **EL-Kerdi A:** Conceptualization, Funding Acquisition, Methodology, Supervision, Validation, Writing – Review & Editing

Competing interests: PM has no conflicts of interest. MG, AL, WJ and ELKA are the employees of Philips. KKM and PS are the employees of Global Market Access Solutions Sàrl. Global Market Access Solutions Sàrl. Received funding from Philips to perform systematic literature review. PM is the employee of the University of Groningen, The Netherlands who provided scientific oversight for the whole project and did not receive any financial support.

Grant information: The study was supported by funding from Philips.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Medic G *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Medic G, Kosaner Kließ M, Atallah L *et al.* **Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review [version 2; peer review: 2 approved]**

F1000Research 2019, 8:1728 (<https://doi.org/10.12688/f1000research.20498.2>)

First published: 08 Oct 2019, 8:1728 (<https://doi.org/10.12688/f1000research.20498.1>)

REVISED Amendments from Version 1

All comments from the Reviewers were addressed in the updated version. We could not address the layout issue that Reviewer 1 made as this is the Journal's decision how tables are made in the PDF.

The question of Reviewer 2 regarding the rationale for including the studies predicting AKI within the Infection/sepsis results section is addressed here:

Severe infection is a major cause of AKI in ICU patients, while conversely, AKI patients are at increased risk for infection [1]. Sepsis is an important cause of AKI, and AKI is a common complication of sepsis [2]. We felt that given this relationship, CDS for AKI fits well under this section. The reviewer is correct to propose the link between AKI and shock, however, not all AKI cases lead to shock- so we felt it matched this section more.

[1] Vandijck DM, Reynvoet E, Blot SI, Vandecasteele E, Hoste EA. Severe infection, sepsis and acute kidney injury. *Acta Clin Belg.* 2007;62 Suppl 2:332-6.

[2] Steven J. Skube, Stephen A. Katz, Jeffrey G. Chipman, and Christopher J. Tiganelli. *Surgical Infections*. <http://doi.org/10.1089/sur.2017.261> Volume: 19 Issue 2: February 1, 2018

Any further responses from the reviewers can be found at the end of the article

Introduction

Critical care, including intensive and emergency care, is the most expensive and human resource intensive area of in-hospital care. Despite having the most technologically advanced devices, it is the area associated with the highest morbidity and mortality rates¹. Decision-making for clinical teams in this area is complex due to variability in procedures and data-overload from the plethora of existing devices. In fact, misdiagnosis in the intensive care unit (ICU) is 50% more common than other areas², and errors, especially medication errors which account for 78% of serious medication errors³, can have a long lasting effect even after patients are discharged.

Computerized decision support (CDS) systems have emerged as tools providing intelligent decision making based on patient data to address many of the challenges of critical care. CDS systems can be based on existing guidelines or best practices; and can also utilize machine learning as a means of compiling several data inputs to provide a diagnosis, recommendation, or therapy course. CDS systems can improve medication safety by providing recommendations relating to dosing⁴⁻⁶, administration frequencies⁵, medication discontinuation⁶ and medication avoidance⁵. Moreover, these novel systems can improve the quality of prescribing decisions by triggering alerts or warning messages on drug duplication, contraindications, drug interaction errors⁷, side-effects and inappropriate medication orders⁵. CDS system notifications can be applied during the prescribing, administering or monitoring stages to detect and prevent medication errors⁸. These systems can also target patients to facilitate shared decision-making to empower as well as to motivate them⁹⁻¹¹. The need for such systems stems from hospitals having to deal with strict guidelines to improve outcomes, document care cycles (raising the need for administrative tasks) and reduce readmissions. This is combined with the need to cope with financial constraints, such as staff shortages and increased pressure to reduce the length of stay^{12,13}.

Strategies for bringing CDS to clinics have been the topic of several workshops, conferences and focus groups¹⁴. Factors for success in designing CDS include providing measurable value, producing actionable insights, delivering information to the user at the right time, and demonstrating good usability principles¹⁴.

Early warning systems (EWS) are CDS systems designed for initial assessment and identification of patients at risk of deterioration in in-patient ward areas¹⁵⁻¹⁷. These systems have shown that they can enable caregivers and rapid response teams to respond earlier – in time to make a difference¹⁸. By alerting clinicians to higher risk patients, treatments can be administered early or harmful medications can be stopped, potentially leading to improved outcomes. Early recognition and timely intervention are also critical steps for the successful management of shock¹⁹, cardiorespiratory instability²⁰ and severe sepsis. In sepsis management, adequate timing of administration of antibiotics is directly associated with survival rates²¹, and incidence, severity and duration of infections.

According to the Society of Critical Care Medicine (SCCM)²², the five primary ICU admission diagnoses for adults are respiratory insufficiency/failure with ventilator support, acute myocardial infarction, intracranial hemorrhage or cerebral infarction, percutaneous cardiovascular procedures, and septicemia or severe sepsis without mechanical ventilation. SCCM also highlights other conditions involving high ICU demand such as poisoning and toxic effects of drugs, pulmonary edema and respiratory failure, heart failure and shock, cardiac arrhythmia and renal failure. Given the above, three high-impact areas were selected for the current research where early detection and treatment could impact outcomes for patients in the ICU. The first is that of hemodynamic instability, where early detection could help patients prevent deterioration into shock. The second is that of respiratory distress, affecting many ventilated patients (up to 40% are ventilated according to SCCM)²². The third area selected is that of infection, with a focus on sepsis. Sepsis is the most common cause of death among critically ill patients, with occurrence rates varying from 13.6% to 39.3%^{23,24}. All three areas are major areas of concern with relatively high prevalence in critical care having long term effects on patients.

The study focuses on both detection, which alerts the clinician to the presence of these specific conditions, as well as prediction of deterioration by alerting the clinician in advance that a patient will deteriorate into one of these disease states. The aims of this study were to perform and report a systematic review of the utilization of CDS systems in the three selected disease areas and summarize the methodological aspects of identified studies.

Methods

Search strategy

A systematic literature review was carried out to identify evidence-based study designs, methods and outcome measures that have been used to determine the clinical effectiveness of CDS systems in the detection and prediction of three populations representing the variety and majority of morbid conditions in a critical care setting: Shock (hemodynamic (in-)stability),

respiratory distress/failure and infection/sepsis. The search strategy combined ‘intervention terms’ and ‘disease terms’ to identify primary research evaluating the diagnostic performance of CDS systems and other machine learning algorithms in three different populations of any age, sex, and race. Systematic literature reviews were also included for locating further relevant primary research. The search was conducted in MEDLINE (PubMed), ClinicalTrials.gov and Cochrane Database of Systematic Reviews (CDSR); and limited to studies published or registered between January 1, 2013 and November 8, 2018 and reported in English. Publication dates were limited to focus results on the most recent developments in this fast-evolving research domain. Another method to ensure up-to-date results was to include conference abstracts from 2017 onwards regardless of whether or not they were followed up with a detailed publication. Ongoing studies identified in the clinical trials register were also kept in the review. Study protocols identified from bibliographic databases were, however, excluded assuming that final study results would be available and identified

elsewhere. The strategy employed in PubMed is provided as *Extended data*, Table 1–Table 3^{25–27}.

Studies conducted in US, Canada, UK, Germany or France with more than 10 subjects per arm were included. These countries were selected because they are known to be active in CDS development. The inclusion and exclusion criteria for selecting abstracts and subsequent full-text publications were based on the population, interventions, comparators, outcomes, and study design (PICOS). These criteria are listed in **Table 1**.

Study selection and data extraction

Study selection and data extraction was carried out by a single reviewer (MKK or SP). In cases of uncertainty, a second, or even third reviewer, was consulted. Data extraction was performed using a standard data extraction form (DEF). Key data from each additional eligible study were extracted by recording data from original reports into the DEF. The DEF included information on study design, inclusion/exclusion criteria, sample

Table 1. Study selection criteria for the systematic literature review.

Criteria		Inclusion	Exclusion
STUDY DESIGN	Abstract selection	Randomized controlled trials (RCT) Observational (retrospective and prospective) studies In-hospital settings: Acute care, Intensive care unit (ICU), Emergency department (ED), Medical Surgery, General ward Geography: US, Canada, Europe	Systematic Literature Reviews or meta-analyses* Review papers, newsletters and opinion papers where treatments of interest are only discussed Methodology studies or protocols Case studies (sample size of 1 patient) Studies with less than 10 patients per arm; Conference abstracts published only as abstracts in 2013, 2014, 2015 and 2016 Geography**: All countries and regions except: US, Canada, UK, Germany, France Publications without an abstract
	Full-text selection	Randomized controlled trials (RCT) Observational (retrospective and prospective) studies In-hospital settings: Acute care, Intensive care unit (ICU), Emergency department (ED), Medical Surgery, General ward Geography**: US, Canada, UK, Germany, France Conference abstracts published only as abstracts in 2017 and 2018	Systematic Literature Reviews or meta-analyses* Review papers, newsletters and opinion papers where treatments of interest are only discussed Methodology studies or protocols Case studies (sample size of 1 patient) Studies with less than 10 patients per arm; Geography**: All countries and regions except: US, Canada, UK, Germany, France Publications published only as abstracts in 2013, 2014, 2015 and 2016 (which were not superseded by full-text publication).
POPULATION	Abstract and full-text selection	Studies that include humans only – adults, children and neonates (or (electronic) medical records) Both sexes are included Patients with or at risk of developing shock (hemodynamic (in-stability)) Patients with or at risk of developing respiratory distress/failure Patients with or at risk of developing infection or sepsis Healthy people only; Healthy people and patients	<i>In-vitro</i> studies Animal studies

Criteria		Inclusion	Exclusion
TREATMENT / INTERVENTION	Abstract and full-text selection	Artificial intelligence Machine learning (i.e. Deep learning models) Clinical decision support Computer aided detection Early Warning System	Automatic diagnosis systems (i.e. ELISA tests) Screening tests (i.e. Automated analysis of portable oximetry) Sequencing tests Mathematical models*** - which model the predictability of disease or treatment/ intervention (i.e. Modelling studies have been widely used to inform human papillomavirus vaccination policy decisions) Multivariable hierarchal logistic regression models*** (models which are based only on statistics - but there is no machine learning)
COMPARATOR	Abstract and full-text selection	All comparators	No selection will be made regarding comparator
OUTCOMES	Abstract and full-text selection	Detection and/or prediction outcomes, such as: <ul style="list-style-type: none"> • Sensitivity (SD) (%) • Specificity (SD) (%) • NPV (%) • PPV (%) • Likelihood ratio • Accuracy (SD) (%) • Prevalence of disease (%) • OR; 95% CI; p-value • HR; 95% CI; p-value • Median (IQR); p-value • ROC AUC <p>For all outcomes (if reported): Measure of variability (i.e. Standard error of mean (SE), Standard deviation (SD)); measure of uncertainty (i.e. 95% CI)</p> <p>The outcomes should be reported in the following manner:</p> <ul style="list-style-type: none"> • per arm (study group vs. control group) individually; • difference between 2 arms. 	Studies not reporting detection and/or prediction outcomes Studies discussing interventions of interest, but no outcomes are reported

* Systematic Literature Reviews and (network) meta-analysis are excluded from data extraction since the pooled results cannot be used in our analysis. However, good quality (network) meta-analysis and systematic literature reviews (i.e. Cochrane reviews) will be used for cross-checking of references if the search did not omit any articles.

** If studies are conducted in multiple countries and at least 1 of the included countries is included – the study will be included in the selection.

*** Mathematical and logistic regression models – can be used to validate and evaluate Interventions of interest (that are listed as included intervention), but the texts discussing these models without any “learning potential” or artificial intelligence potential will be excluded. Therefore, these models can be the foundation of the included listed interventions but will not be included in the Data Extraction Files unless they have also machine learning or artificial intelligence or some other form of “learning potential” on top of the statistical mathematical model. Researchers will pay special attention and caution when screening these abstracts and/or full-text articles.

AUC = Area under the curve; ED = Emergency department; ELISA = Enzyme-linked immunosorbent assay; HR = Hazard ratio; ICU = Intensive care unit; IQR = interquartile range; NPV = Negative predictive value; OR = Odds ratio; PPV = Positive predictive value; RCT = Randomized controlled trial; ROC = Receiver Operating Characteristic; SD = Standard deviation; SE = Standard error; UK = United Kingdom; US = United States.

size and characteristics, interventions, outcome measures (measures of predictability like: sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), likelihood ratio, accuracy (percentage of correctly identified cases in relation to the whole sample), odds ratio (OR), hazard ratio (HR), median, receiver operating characteristic (ROC) area under the curve (AUC); and length of hospitalization among others).

Studies identified from the ClinicalTrials.gov registry that did not report results were also included in the extraction to give some indication of the outcomes being collected.

Study quality appraisal

This research was not aimed at summarizing study results and assessing the relative effectiveness of CDS systems. Therefore, an appraisal of study quality was not deemed necessary.

Results

Shock (hemodynamic (in-)stability)

The search yielded 1588 hits. Screening the titles and abstracts led to 1502 being excluded. The full texts of the remaining 86 titles were obtained and assessed against the PICOS criteria. Studies were excluded due to irrelevant study design (n=22), population (n=1), intervention (n=5), and outcomes (n=38). A total of 20 studies were finally included in this systematic literature review. This included 5 trials identified from ClinicalTrials.gov. The study selection process is depicted in Figure 1.

Study characteristics. Of the 15 published studies, five were conducted by research groups outside the USA²⁸⁻³². Ten studies

were conducted in the US^{19,33-41}. Thirteen studies were retrospective^{19,28-33,35,37-41} and only two were prospective^{34,36}. Nine studies were single-center^{28,30,31,33,37-41} and six studies were multi-center^{19,29,32,34-36}. Five studies were time-series^{28,30-32,40} and nine were case-series^{19,29,33-35,37-39,41}.

Across all studies, three had sample sizes ≤ 100 ^{29,30,36}; three had sample sizes of 101-1000^{28,31,32}; four studies had sample sizes of 1001-10,000^{19,33,34,37,42}; and another five studies, four retrospective single-center studies and one multi-center, had sample sizes larger than 10,000^{35,38-41}. The three largest studies included patients admitted to various wards of a specified hospital. The majority of the studies did not restrict their sample to a specific in-patient hospital setting. Five studies reported on patients

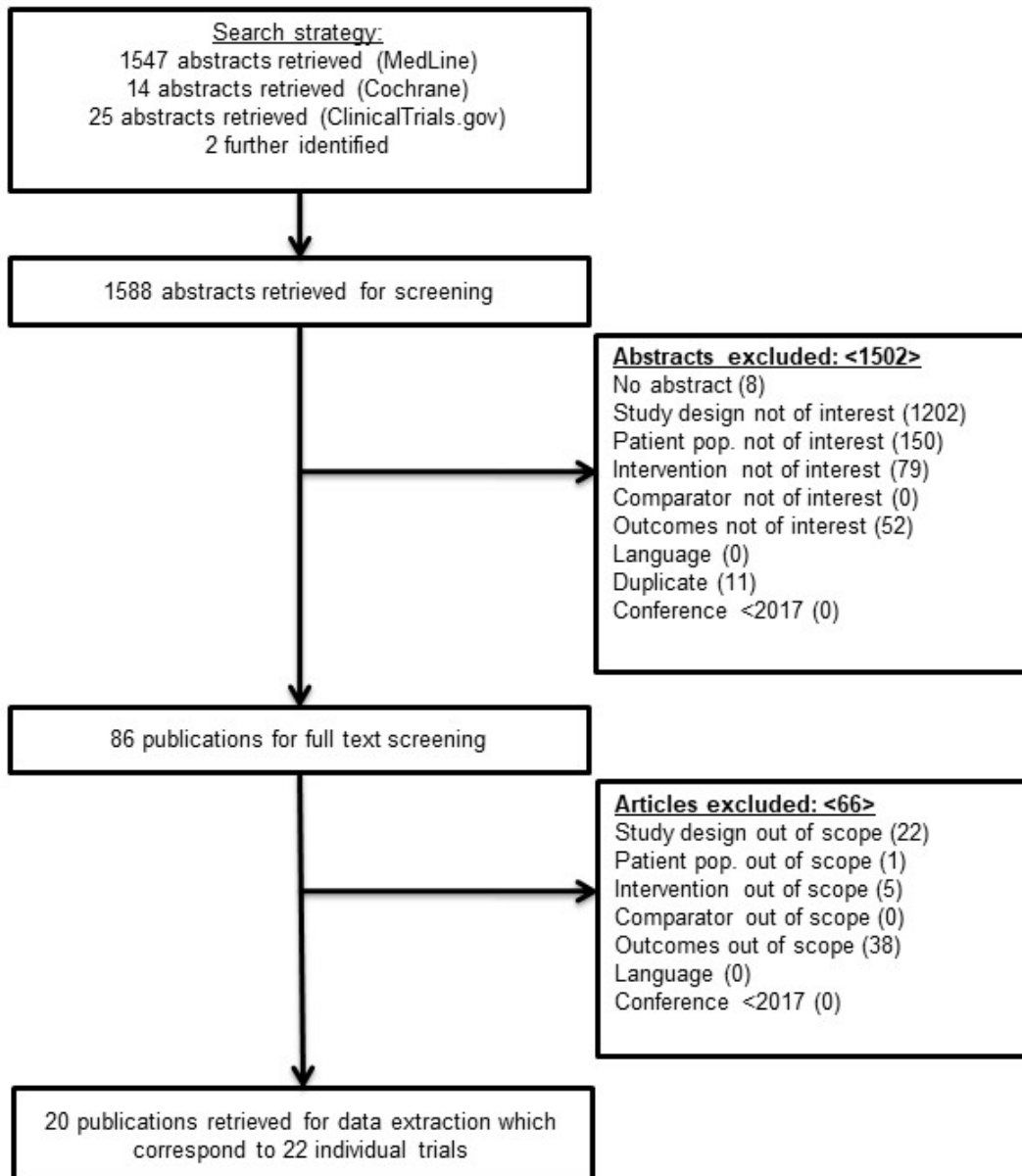


Figure 1. Study selection – Shock. Pop. = Population.

in the ICU^{19,28,32,40,41} and one study reported on patients admitted to the surgical ward³³.

The characteristics of the published studies are summarized in [Table 2](#).

CDS systems. Machine learning algorithms were developed to detect or predict septic shock^{28,33,35,40,41}, various heart arrhythmias^{29,30,34}, heart failure^{37–39}, hemodynamic instability and hypovolemia^{19,36}, myocardial infarction³¹, as well as hypotension³².

All studies, except one, trained a single algorithm. Ebrahimzadeh *et al.* 2018³⁰ trained and compared support vector machine (SVM), instance-based and neural network models to predict paroxysmal atrial fibrillation. SVMs were the most frequently used algorithms, followed by least absolute shrinkage and selection operator (LASSO) regularization. In one study, the SVM was trained using sequential minimal optimization³⁷.

Machine learning models were trained and validated in 14 studies and subsequently tested in an independent dataset in 3 studies^{19,35,37}. In one study an algorithm trained to classify arrhythmias was not validated but compared to physician's manual classifications³⁴.

An overview of the investigated machine learning algorithms is presented in [Table 3](#).

Outcome measures. Three of the 15 papers measured a single outcome of model performance. In two studies the preferred measure was accuracy^{28,34}; whereas in another study this was the ROC AUC. This study was large and based their algorithm on EHRs³³. Across all studies, accuracy was reported in about half of the instances and the ROC AUC was one of the most frequently reported outcomes.

Sensitivity and specificity were reported together in 10 studies. Blecker *et al.* 2016³⁸ reported sensitivity together with PPV. Sensitivity and specificity were not measured in the study by Sideris *et al.* 2016³⁷, instead model accuracy and the ROC AUC were preferred. This study was concerned with developing an alternative 'comorbidity' framework based on disease and symptom diagnostic codes to cluster individuals at low to high risk of developing chronic heart failure.

PPVs were reported in six studies and accompanied with negative predictive values in two studies. These studies developed and validated machine-learning algorithms for the early detection of less investigated health conditions, these being hemodynamic instability in children¹⁹ and acute decompensated heart failure³⁹. The highest number of outcome measures, including likelihood ratios, was observed in Calvert *et al.* 2016⁴⁰ who investigated an under-represented population of patients with Alcohol Use Disorder.

The outcomes measured are summarized in [Table 4](#).

Ongoing studies. Five studies are currently ongoing, one in Germany⁴³ and the others in the USA^{44–47}. Two studies are

prospective case series^{44,47}, two studies are prospective cohort studies^{43,45} and one is a RCT⁴⁶. Two of the studies are concerned with developing prediction models, and the others are concerned with implementing machine learning algorithms into clinical practice as early warning systems.

The details of these trials are summarized in [Table 5](#).

Respiratory distress/failure

The search yielded 1279 hits. Screening the titles and abstracts lead to 1142 being excluded. The full texts of the remaining 137 titles were obtained and assessed against the PICOS criteria. Studies were excluded due to irrelevant study design (n=42), population (n=6); intervention (n=18) and outcomes (n=47), and conference proceeding from before 2017 (n=2). A total of 22 studies were finally included in this systematic literature review. None of the trials retrieved from ClinicalTrials.gov were included. The study selection process is depicted in [Figure 2](#).

Study characteristics. Of the included studies, 17 were conducted in the US^{33,48–63}. Five studies were conducted outside the US; two in Canada^{64,65} by the same research group, two in France^{66,67} and one in the UK⁶⁸. In total, 17 studies were retrospective^{33,48–50,52–55,58–66} and five were prospective^{51,56,57,67,68}. Of these studies, 12 were single-center^{33,48,49,51,52,54,55,58,59,64–66} and 10 studies were multi-center^{50,53,56,57,60–63,67,68}. Five studies were time-series^{48,52,55,56,64}, 14 studies were case-series^{33,49,51,53,54,57–62,65,66,68}, one was case-control⁵⁰ and one was case/time series study⁶³.

The smallest sample of 100 patients came from two single-center retrospective studies^{48,66}. Ten studies had sample sizes of 101–1000^{33,49–53,57,63,67,68}; seven studies had sample sizes of 1001–10,000^{54,55,59,60,62,64,65}; and three had sample sizes larger than 10,000^{56,58,61}. The largest study included more than 50,000 patients admitted to the ED of two centers over a 3-year period⁶¹. Several published studies did not report their in-patient setting. When reported, some evaluated data from different wards^{56,59,64,65,68}, and some included patients admitted only to the ED^{53,54,61,63}, the ICU^{48,60,67} and the surgical ward^{33,51,55}.

The characteristics of all published studies are given in [Table 6](#).

CDS systems. About half of the studies developed machine-learning algorithms, whereas the other half focused on natural language processing (NLP) algorithms. One study differed from the rest by developing a computer-aided detection (CAD) system to measure the axial diameter of the right and left pulmonary ventricles, aiding in the diagnosis of pulmonary embolisms⁴⁹. Many learning algorithms were concerned with detecting pulmonary embolisms and deep vein thrombosis^{53,54,58,59,64–67} as well as pneumonia^{33,48,57,60–63}. Three studies developed machine-learning algorithms to detect COPD^{50,56,69}. One study developed a machine learning algorithm to detect acute respiratory distress syndrome⁵²; while other studies developed machine learning algorithms to detect respiratory distress or failure following a pressure support ventilation trial⁶⁷, cardiovascular surgery⁵⁵ and pediatric tonsillectomy⁵¹.

Table 2. Design aspects of published studies on shock.

Study	Study Design	Country and institution(s)	Number of patients (records)	Population/disease definition	In-patient setting	Collected data
Ghosh 2017	Retrospective time series single center	Australia University of Technology Sydney & The University of Melbourne	209	Sepsis or severe sepsis	ICU	(mean arterial pressure), heart rate, respiratory rate
Hu 2016	Retrospective case series single center	USA, Minnesota University of Minnesota	NR (8909)	NR	Surgery	EHRs
Li 2014	Retrospective case series multi-centric (3 centers)	UK, Oxford University of Oxford & Mindray	NR (67)	Ventricular flutter, fibrillation and tachycardia	NR	Electrocardiography
Mahajan 2014	Prospective case series multi-centric (4 centers)	USA University of Southern California, Mayo Clinic-Rochester, University of North Carolina, Sanger Heart & Vascular Institute & Boston Scientific	410 (908)	Ventricular fibrillation, ventricular tachycardia and other arrhythmias	NR	Electrograms
Mao 2018	Retrospective case series multi-centric (5 centers)	USA University of California, Stanford Medical Centre, Oroville Hospital, Bakersfield Heart Hospital, Cape Regional Medical Centre, Beth Israel Deaconess Medical Center	359,390	NR	various	Vital signs
Reljin 2018	Prospective case-control multi-centric (2 centers)	USA University of Connecticut, Campbell University School of Medicine, University of Massachusetts Medical School, Yale University School of Medicine & Worcester Polytechnic Institute	36 (94)	Traumatic injury, healthy controls	NR	Photoplethysmographic signals
Sideris 2016	Retrospective case series single center	USA, Los Angeles University of California	1948	Primarily heart failure	various	EHRs
Blecker 2016	Retrospective case series single center	USA, New York New York-Presbyterian Hospital & New York University	NR (47,119)	NR	various	EHRs
Blecker 2018	Retrospective case series single center	USA, New York New York University	NR (37229)	NR	various	EHRs

Study	Study Design	Country and institution(s)	Number of patients (records)	Population/disease definition	In-patient setting	Collected data
Calvert 2016	Retrospective time series single center	USA, California Dascena Inc. & University of California	29083	NR	ICU	vital signs
Donald 2018	Retrospective time series + Prospective time series multi-centric (22 centers)	Europe	173	Traumatic brain injury	ICU	Demographic, clinical and physiological data
Ebrahimzadeh 2018	Retrospective time series single center	Iran University of Tehran, Iran University of Science and Technology, University of Sheikhbahae & Payame Noor University of North Tehran	53 (106)	Paroxysmal atrial fibrillation	NR	Electrocardiography
Potes 2017	Retrospective case series multi-centric (2 centers)	USA, California & UK, London Children`s Hospital Los Angeles, St. Mary`s Hospital, London & Philips	8022	NR	ICU	Vital signs, laboratory values, and ventilator parameters.
Henry 2015	Retrospective case series single center	USA, Maryland John Hopkins University	16234	NR	ICU	EHRs
Strothoff 2018	Retrospective time series single center	Germany, Berlin Fraunhofer Heinrich Hertz Institute & University Medical Center Schleswig-Holstein, Kiel	200 (228)	Myocardial infarction and healthy controls	NR	Electrocardiography

USA: United States of America. UK: United Kingdom. NR: Not reported. ICU: Intensive care unit. EHR: Electronic health records.

The classifiers used in the NLP-based studies were various. However, some commonalities emerged between the studies developing machine-learning algorithms. Multiple studies applied SVM, logistic regression, random forests, K- nearest neighbor (kNN), gradient boosting and neural network models. Various classifiers were explored in 5 studies.

Machine learning and NLP-based algorithms were trained and validated in 20 studies and subsequently tested in an independent dataset in 6 studies^{52,56,60–62,67}. The CAD system mentioned above and an electronic pulmonary embolism severity index were trained and compared to a reference dataset classified by physicians^{49,53}.

An overview of the developed learning algorithms is provided in [Table 7](#).

One study, Reamoroon *et al.* 2018⁵², used a novel sampling technique to accommodate for inter-dependency in longitudinal

data. Model accuracy and ROC AUC with this method was <5% better than random sampling and 4–11% better than no sampling.

Outcome measures. The majority of the studies reported multiple outcome measures of model performance. The most frequently reported outcome measure was sensitivity, followed by specificity and ROC AUC. Likelihood ratios, on the other hand, were only reported in one study: Silva *et al.* 2017⁶⁷ reported eight outcome measures of their novel machine learning model to predict post extubation distress. The outcomes measured across all studies are summarized in [Table 8](#).

Many of the studies that developed NLP-based algorithms reported negative and positive predictive values, as well as sensitivity and specificity. In contrast, the ROC AUC was the most frequently reported outcome measure of machine learning

Table 3. Overview of the algorithms developed to detect shock.

Study	Predicted disease	Learning algorithm												
		CHMM	Decision trees	LR, LASSO regularisation	LR, not specified	SVM	kNN	RF	gradient tree boosting	Adaptive boosting	Bayesian neural network	convolutional neural network	Multilayer perceptron	mixture of expert
Ebrahimzadeh 2018	paroxysmal atrial fibrillation					✓	✓						✓	✓
Li 2014	Ventricular fibrillation and tachycardia					✓								
Mahajan 2014	heart arrhythmias					✓								
Strothoff 2018	myocardial infarction										✓			
Sideris 2016	heart failure					✓								
Blecker 2016	heart failure			✓										
Blecker 2018	heart failure			✓										
Reljin 2018	Hypovolemia					✓								
Potes 2017	hemodynamic instability									✓				
Donald 2018	Hypotension										✓			
Ghosh 2017	septic shock	✓												
Hu 2016	septic shock			✓										
Mao 2018	septic shock							✓						
Calvert 2016	septic shock				✓									
Henry 2015	septic shock			✓										

CHMM: clustered hidden Markov model. LR: Logistic regression. SVM: Support vector machine. kNN: k nearest neighbor. RF: Random forest. Conv.: Convolutional.

Table 4. Overview of measured outcomes in studies on shock.

Study	Sensitivity	Specificity	NPV	PPV	Negative LR	Positive LR	Accuracy	Prevalence	OR	RR	ROC AUC
Ghosh 2017							✓				
Hu 2016											✓
Li 2014	✓	✓					✓				✓
Mahajan 2014							✓				
Mao 2018	✓	✓									✓
Reljin 2018	✓	✓					✓				
Sideris 2016							✓				✓
Blecker 2016	✓			✓							✓
Blecker 2018	✓	✓	✓	✓							✓
Calvert 2016	✓	✓			✓	✓	✓		✓		✓
Donald 2018	✓	✓		✓							✓
Ebrahimzadeh 2018	✓	✓		✓			✓				
Potes 2017	✓	✓	✓	✓		✓					✓
Henry 2015	✓	✓									✓
Strothoff 2018	✓	✓		✓							

NPV: Negative predictive value. PPV: Positive predictive value. LR: Likelihood ratio. OR: Odds ratio. RR: Risk ratio. ROC AUC: Receiver operating characteristic area under the curve.

Table 5. Overview of ongoing studies on shock.

Identifier code	Study Design	Countries and study centers	Hospital setting	Intervention	Sample characteristics	Outcome(s)
NCT03582501	Prospective case series Year of study: 2019–20 Duration: 12 months	USA Mayo Clinic Arizona, Florida & Rochester	NR	Lower body negative pressure to simulate hypovolemia	Estimated: 24 Age: 18–55 Definition: Healthy non-smoker, no history of hypertension, diabetes, CAD and neurologic diseases	<u>Primary outcome</u> Blood pressure <u>Secondary outcome</u> Heart rate
NCT02934971	Prospective cohort study Year of study: 2017–19 Duration: 24 months (up to 6 months follow-up)	Germany, Aachen Aachen University Hospital	Out-patient	Chemotherapy or no chemotherapy	Estimated: 400 Age: ≥ 18 Definition: Patients scheduled for chemotherapy at increased risk of cardiotoxicity and age-matched controls	<u>Primary outcome</u> change in left ventricular ejection fraction
NCT03235193	Prospective cohort study Year of study: 2017 Duration: 3 months	USA, West Virginia Dascena Inc. & University of California	ED, ICU	The InSight algorithm used as an EWS to detect sepsis and severe sepsis detection from EHRs compared to severe sepsis detection from EHRs alone	Estimated: 1241 Age: ≥ 18 Definition: All admitted patients	<u>Primary outcome</u> in-hospital mortality <u>Secondary outcomes</u> length of stay in hospital and ICU, hospital readmission
NCT03644940	RCT Year of study: 2020–21 Duration: 6 months	USA, California Dascena Inc. & University of California	Cardiology, GI, ICU, Medicine, Oncology, Surgery, Transplant and ED	subpopulation-optimized version of InSight compared to the original version used as an early warning system to identify patients at high risk of severe sepsis; followed by physician assessment of sepsis	Estimated n: 51645 Age: >18 Definition: NR	<u>Primary outcomes</u> in-hospital SIRS-based mortality <u>Secondary outcomes</u> in-hospital severe sepsis/shock-coded mortality; SIRS-based hospital length of stay; Severe sepsis/shock-coded hospital length of stay
NCT03655626	Single-arm trial up to Year of study: 2018–19 up to Duration: 6 months	USA, North Carolina Duke University Hospital	ED	machine learning algorithm to predict sepsis, custom dashboard and monitoring	Estimated n: 3200 Age: >18 Definition: NR	<u>Primary outcome</u> rate of CMS bundle completion for patients with sepsis <u>Secondary outcomes</u> time to sepsis diagnosis; number of patients developing sepsis; number of patients developing sepsis and not treated; length of stay in ED and hospital; inpatient mortality; ICU requirement rate; time from sepsis onset to blood culture, antibiotics, IV fluids, lactate, CMS bundle completion; rate of lactate complete; number of sepsis diagnostic codes per month

USA: United States of America. NR: Not reported. ED: Emergency department. ICU: Intensive care unit. GI: Gastroenterology.

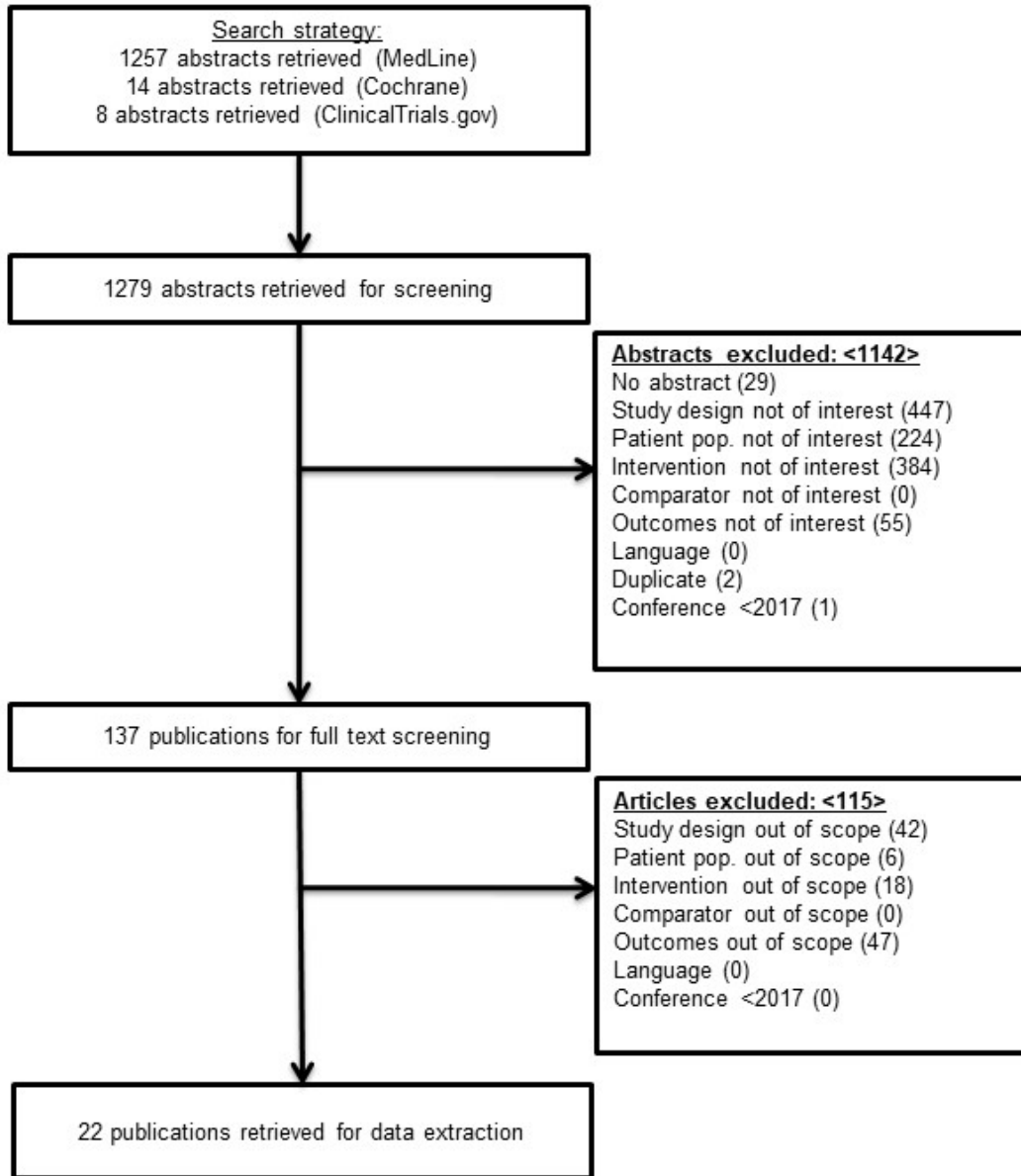


Figure 2. Study selection - Respiratory distress-failure. Pop. = Population.

Table 6. Design aspects of published studies on respiratory distress or failure.

Study	Study Design	Countries and institution(s)	Number of patients (records)	Population/disease definition	In-patient setting
Bejan 2013	Retrospective time series single center	USA, Washington University of Washington	100	NR	ICU
Kumamaru 2016	Retrospective case series single center	USA, Massachusetts Brigham and Women's Hospital	125	acute pulmonary embolism	NR
Bodduluri 2013	Retrospective case-control multi-center (national data)	USA, Iowa The University of Iowa	153	smokers with or without COPD and non-smokers	NR
Biesiada 2014	Prospective case series single center	USA, Cincinnati Children's Hospital Medical Center & University of Cincinnati	347	current tonsillitis, adenotonsillar hypertrophy or obstructive sleep apnea	Surgery
Reamaroon 2018	Retrospective time series single-center	USA, Michigan University of Michigan	401	mild hypoxia and acute hypoxic respiratory failure	NR
Vinson 2015	Retrospective case series multi-center (4 centers)	USA, California the Kaisers Permanente CREST Network	593	acute pulmonary embolism	ED
Huesch 2018	Retrospective case series single center	USA, Pennsylvania Milton S. Hershey Medical Center	1133	individuals suspected of pulmonary embolism	ED
Mortazavi 2017	Retrospective time series single center	USA, Connecticut Yale University	5214	patients undergoing cardiovascular procedures: CABG, PCI and ICD procedures	Surgery
Pham 2014	Retrospective case series single center	France CHU de Caen, Caen & Hôpital Européen Georges-Pompidou, Paris	NR (100)	individuals suspected of having Venous thromboembolism	NR
Rochefort 2015	Retrospective time series single center	Canada, Quebec McGill University	1649 (2000)	individuals suspected of having Venous thromboembolism	various
Silva 2017	Prospective before-after multi-center (3 centers)	France University Teaching Hospital of Purpan, Toulouse; Hôpital Dieu Hospital, Narbonne; Saint Eloi Hospital, Montpellier	136	hemodynamic instability, respiratory failure, multiple trauma, nontraumatic coma, and postoperative complication of abdominal surgery	ICU
Gonzalez 2018	Prospective time series multi-center, multi- national	USA Binham and Women's Hospital (on behalf of the COPD and ECLIPSE Study investigators)	11655	smokers with or without COPD	various
Tian 2017	Retrospective case series single center	Canada, Quebec McGill University	2819 (4000)	individuals suspected of having Venous thromboembolism	various

Study	Study Design	Countries and institution(s)	Number of patients (records)	Population/disease definition	In-patient setting
Choi 2018	Prospective case series multi-center (3 centers)	USA Mayo Clinic, Scottsdale; National Jewish Health, Denver; University of Washington Medical Center, Seattle & Veracyte Inc.	139 (403)	suspected interstitial lung disease	NR
Yu 2014	Retrospective case series single center	USA, Massachusetts Brigham, and Women's Hospital & Harvard Medical School,	NR (10,330)	individuals suspected of pulmonary embolism	NR
Swartz 2017	Retrospective case series single center	USA, New York New York University & Mount Sinai St. Luke's Hospital	NR (2400)	individuals suspected of having Venous thromboembolism	various
Liu 2013	Retrospective case series multi-center (21 centers)	USA, California Kaiser Permanente	NR (2466)	NR	ICU
Haug 2013	Retrospective case series multi-center(2 centers)	USA, Utah LDS Hospital and Intermountain Medical Centre	NR (362,924)	NR	ED
Dublin 2013	Retrospective case series multi-center (regional data)	USA, Seattle Group Health Research Institute & University of Washington	NR (5000)	NR	NR
Phillips 2014	Prospective case series multi-center	UK, Llaneli Swansea University, Aberystwyth University & Hywel Dda University Health Board	181	with and without COPD	various
Hu 2016	Retrospective case series single center	USA, Minnesota University of Minnesota	NR (8909)	NR	Surgery
Jones 2018	Retrospective case/time series multi-center (number of centers unknown)	USA, Utah & Washington VA Salt Lake City Health Care System, University of Utah & George Washington University	NR (911)	individuals suspected of pneumonia	ED

NA: Not applicable. NR: Not reported. USA: United States of America. COPD: Chronic obstructive pulmonary disease. ECLIPSE: Evaluations of COPD Longitudinally to Identify Predictive Surrogate Endpoints. UK: United Kingdom. CABG: Coronary artery bypass grafting. PCI: Percutaneous coronary intervention. ICD: Implantable cardioverter defibrillator. ICU: Intensive care unit. ED: Emergency department.

algorithm performance. It was also the single preferred outcome in three studies^{33,50,55}. About half of the studies additionally reported sensitivity, specificity, and accuracy. One study reported specificity with sensitivity set at 90% and 95% to ensure that few disease positive cases were missed⁵². The single study that developed a CAD system measured the ROC AUC and model accuracy⁴⁹.

Infection or sepsis

The search yielded 2659 hits. Screening the titles and abstracts lead to 2562 being excluded. The full texts of the remaining 97 titles were obtained and assessed against the PICOS criteria. Studies were excluded due to irrelevant study design (n=41), population (n=4); intervention (n=6) and outcomes (n=14).

A total of 31 studies were finally included in this systematic literature review. Four of these were ongoing trials. The study selection process is depicted in [Figure 3](#).

Study characteristics. Of the included studies, 24 were conducted in the US. Three studies were conducted outside the US; one in France; one in the Netherlands and one in the UK. In total, 21 studies were retrospective^{33,35,70–88} and six were prospective^{89–94}. There were 21 single-center studies^{33,70–75,77–83,86–88,90–92,94} and six multi-center studies^{35,76,84,85,89,93}. Seven studies were time series^{71,78,82,84–86,92}, 18 studies were case series^{33,35,70,72–76,80,81,83,87–91,93,94}, one was a case-control⁷⁷ and one was a matched-controlled study⁷⁹.

Table 7. Overview of the algorithms developed to detect respiratory distress or failure.

Study	Predicted disease	Learning algorithm																					
		NLP	assertion classification	symbolic classifiers	rule or probability based	KNN	ONXX	RF	LR, LASSO penalized	LR, LASSO regularization	LR, not specified	gradient (descent) boosting	Maximum Entropy	SVM	Partial least-squares regression	NeGX	hierarchical classification	Bayesian network	neural network	J48	JRIP	PART	
Reamaroon 2018	ARDS						✓			✓				✓									
Gonzalez 2018	COPD, ARDE																	✓					
Bodduluri 2013	COPD				✓																		
Phillips 2014	COPD																			✓			✓
Bejan 2013	Pneumonia	✓	✓																				
Dublin 2013	Pneumonia	✓				✓																	
Haug 2013	Pneumonia	✓																✓					
Hu 2016	Pneumonia	✓					✓																
Liu 2013	Pneumonia	✓			✓																		
Choi 2018	Pneumonia	✓					✓				✓			✓								✓	
Jones 2018	Pneumonia	✓												✓									
Siwa 2017	Postintubation distress														✓								
Mortazavi 2017	Postoperative respiratory failure						✓				✓												
Vinson 2015	Pulmonary embolism	✓			✓																		
Yu 2014	Pulmonary embolism	✓							✓														
Huesch 2018	Pulmonary embolism	✓			✓																		
Kumamaru 2016	Pulmonary embolism*																						
Pham 2014	Pulmonary embolism, DVT	✓																					
Rochefort 2015	Pulmonary embolism, DVT																						
Swartz 2017	Pulmonary embolism, DVT	✓																					
Tian 2017	Pulmonary embolism, DVT	✓		✓																			
Biesiada 2014	Respiratory depression						✓																

*A computer aided detection system was developed for measuring the right ventricular/left ventricular axial diameter ratio and detecting pulmonary embolism. ARDS: Acute respiratory distress syndrome. ARDE: Acute respiratory disease events. COPD: Chronic obstructive pulmonary disease. DVT: Deep vein thrombosis.

Table 8. Overview of measured outcomes in studies predicting respiratory distress or failure.

Study	Algorithm	Sensitivity	Specificity	NPV	PPV	negative LR	positive LR	Accuracy	Prevalence	OR	RR	ROC AUC	Diagnostic yield
Kumamaru 2016	CAD							✓				✓	
Bodduluri 2013	ML											✓	
Hu 2016	ML											✓	
Mortazavi 2017	ML											✓	
Rocheffort 2015	ML	✓	✓	✓	✓							✓	
Silva 2017	ML	✓	✓	✓	✓	✓	✓	✓				✓	
Vinson 2015	ML	✓	✓	✓	✓			✓					
Biesiada 2014	ML	✓	✓					✓	✓		✓		
Choi 2018	ML	✓	✓									✓	
Gonzalez 2018	ML							✓	✓	✓		✓	
Phillips 2014	ML	✓	✓					✓				✓	
Reamaroon 2018	ML		✓					✓				✓	
Bejan 2013	NLP	✓	✓	✓	✓			✓					
Dublin 2013	NLP	✓	✓	✓	✓								
Haug 2013	NLP											✓	
Liu 2013	NLP	✓	✓	✓	✓								
Pham 2014	NLP	✓			✓								
Swartz 2017	NLP	✓	✓	✓	✓								✓
Tian 2017	NLP	✓	✓	✓	✓								
Yu 2014	NLP			✓	✓							✓	
Huesch 2018	NLP	✓	✓	✓	✓			✓					
Jones 2018	NLP	✓	✓	✓	✓							✓	

NLP: Natural language processing. ML: Machine learning. CAD: Computer aided detection. NPV: Negative predictive value. PPV: Positive predictive value. LR: Likelihood ratio. OR: Odds ratio. RR: Risk ratio. ROC AUC: Receiver operating characteristic area under the curve.

The smallest studies included patients with leukemia⁸⁹ and combat casualty patients⁹⁰. Four studies had a sample size below 1000^{70,72,73,79}, three had a sample size between 1001–10,000^{33,71,87} and 12 had a sample size larger than 10,000^{35,74,77–78,80–82,84–87,88}. Eight studies had samples even larger than 50,000^{35,74,77,78,82,84,85,88}. Large samples were achieved by less restrictive inclusion criteria where all patients admitted to specific ward(s) or hospital(s) over a given time were defined.

Majority of the published studies evaluated data from different wards; several studies included patients admitted only to the

ICU^{70,72,81,84–86,93} and surgical ward^{73,76,78,87,91,92}, less often the General ward³³ and Emergency Department⁷⁴. Of these, 23 studies included data collected at their own hospital; and four utilized previously collated databases^{76,81,84,86}.

The characteristics of all published studies are given in Table 9.

CDS systems. The machine learning algorithms evaluated in the studies were developed to predict a range of diseases. These included sepsis^{33,35,72,78,81,85,93,94}, acute kidney injury^{70,78–80,82,84,91}, surgical site infections^{33,73,76,87,92}, central line-associated

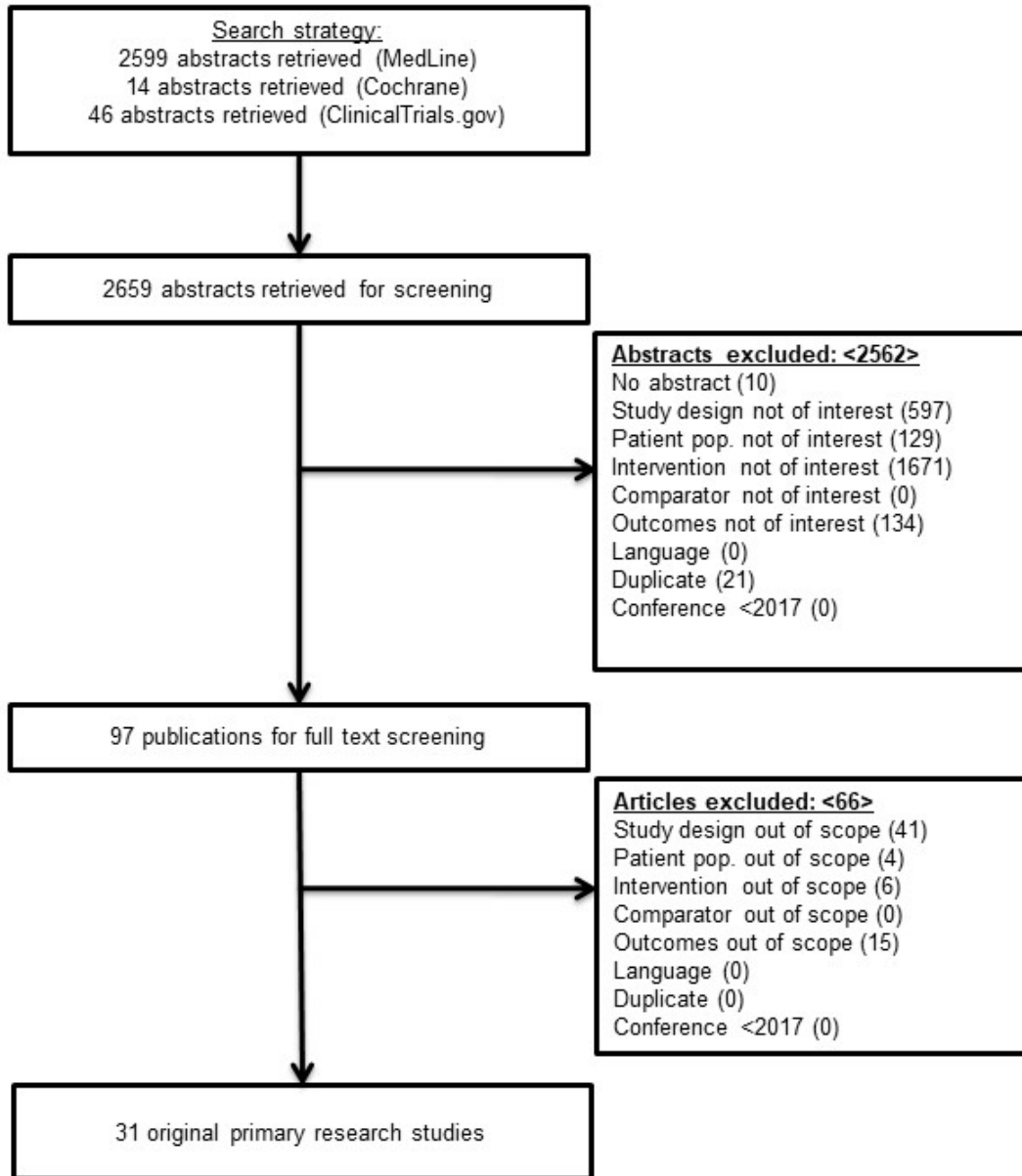


Figure 3. Study selection - infection or sepsis. Pop. = Population.

Table 9. Design aspects of published studies on infection or sepsis.

Study	Study Design	Country and institution(s)	Number of patients (records)	Population/disease definition	In-patient setting
Ahmed 2015	Retrospective case series single center	USA, Minnesota Mayo Clinic Rochester	944	NR	ICU
Brasier, 2015	Prospective case series multi-center (3 sites)	USA, Texas Aspergillus Technology Consortium & University of Texas	57	Leukemia	NR
Dente, 2017	Prospective case series single center	USA, Maryland Emory University, Walter Reed National Military Medical Centre	73	Combat casualty patients	NR
Hu, 2016	Retrospective case series single center	USA, Minnesota University of Minnesota	NR (8,909)	NR	General
Konerman, 2017	Retrospective time series single center	USA, Michigan University of Michigan	1,233	Chronic hepatitis c	NR
Legrand, 2013	Prospective case series single center	France, Paris Hôpital Européen Georges Pompidou Assistance Publique-Hopitaux de Paris	202	Infective endocarditis	Surgery
Mani, 2014	Retrospective case series single center	USA, New Mexico University of New Mexico	299	Sepsis	ICU
Mao 2018	Retrospective case series multi-center (5 centers)	USA University of California, Stanford Medical Centre, Oroville Hospital, Bakersfield Heart Hospital, Cape Regional Medical Centre, Beth Israel Deaconess Medical Center	359,390	NR	various
Sanger, 2016	Prospective time series single center	USA, Washington University of Washington	851	Open-abdominal surgery patients	Surgery
Scicluna, 2017	Prospective case series multi-center (2 sites + national database)	Netherlands & UK Amsterdam Academic Medical Center, Utrecht University Medical Center & UK Genomic Advances in Sepsis study	787	Sepsis	ICU
Sohn, 2016	Retrospective case series single center	USA, Minnesota Mayo Clinic Rochester	751	Colorectal surgery patients	Surgery
Taylor, 2018	Retrospective case series single center	USA, Connecticut Yale University School of Medicine,	55,365 (80,387)	Suspected urine tract infection	ED
Hernandez 2017	Retrospective case series single center	UK, London Imperial College Healthcare NHS Trust	> 500,000	NR	NR

Study	Study Design	Country and institution(s)	Number of patients (records)	Population/disease definition	In-patient setting
Bartz-Kurycki 2018	Retrospective case series multi-center (national database)	USA, Texas University of Texas	13,589	NR	Surgery
Beeler 2018	Retrospective case-control single center	USA, Indiana Indiana University Health Academic Health Center	NR (70,218)	Central venous line with or without central line-associated bloodstream infections	NR
Bihorac 2018	Retrospective time series single center	USA, Florida University of Florida Health	51,457	NR	Surgery
Chen 2018	Retrospective matched pairs (1:1 case matching) single center	USA, Kansas University of Kansas Health System	358	Stage 3 AKI and non-AKI controls	NR
Cheng 2017	Retrospective case series single center	USA, Kansas University of Kansas Medical Center	33,703 (48,955)	NR	NR
Desautels 2016	Retrospective case series single center	USA, California Dascena Inc. & University of California	NR (21,176)	NR	ICU
Koyner 2015	Retrospective time series single center	USA, Chicago University of Chicago	NR (121,158)	NR	NR
LaBarbera 2015	Retrospective case series single center	USA, Pennsylvania Pinnacle Health Hospital, Harrisburg	198	Clostridium difficile infection	NR
Mohamadlou 2018	Retrospective time series multi-center (2 sites)	USA Dascena Inc., University of California & Stanford University	68,319	NR	ICU
Nemati 2018	Retrospective time series multi-center (3 sites)	USA, Georgia Emory University School of Medicine & Georgia Institute of Technology	69,938	NR	ICU
Parreco 2018	Retrospective time series single center	USA, Florida University of Miami	NA (22,201)	NA	ICU
Taneja 2017	Prospective case series single center	USA, Illinois University of Illinois	444	Suspected sepsis	NR
Weller 2018	Retrospective case series single center	USA, Minnesota Mayo Clinic Rochester	1,283	Colorectal surgery patients	Surgery
Wiens 2014	Retrospective case series single center	USA single center not specified	NR (69,568)	NR	various

NA: Not applicable. NR: Not reported. USA: United States of America. UK: United Kingdom. ICU: Intensive care unit. ED: Emergency department. AKI: Acute kidney injury.

bloodstream infections^{77,86}, *Clostridium difficile*^{83,88}, pulmonary aspergillosis⁸⁹, bacteremia⁹⁰, fibrosis⁷¹, urine tract infection^{33,74} and infections in general⁷⁵.

Almost half of the studies compared different machine learning algorithms, while the others focused only on Bayesian algorithms^{73,92}, decision tree algorithms⁸⁴, ensemble algorithms^{35,71,82,83,90,93}, regression algorithms^{33,78,85}, regularization algorithms^{81,88} and rule learning⁷⁰. The most frequently applied model was random forest (15 studies) followed by logistic regression (10 studies), support vector machines (5 studies), naïve Bayes (5 studies) and gradient tree boosting (5 studies).

One study compared three different sampling methods for handling class imbalance; under-sampling the majority class (RANDu), over-sampling the minority class (RANDo) and synthetic minority over-sampling (SMOTE). This was a very large study including more than 500,000 patients to predict the onset of infections⁷⁵. The authors found that SMOTE outperformed the other techniques and improved model sensitivity. Two other very large studies used the RANDu method⁸⁰ and mini-batch stochastic gradient descent with backpropagation⁸⁵. No other studies were concerned with imbalance in disease positive and negative classification.

Machine learning models were trained and validated in 26 studies and subsequently tested in an independent dataset in four studies^{35,72,75,77}.

The machine learning algorithms used are illustrated in [Table 10](#).

Outcome measures. The most frequently reported outcome measure was the ROC AUC. Three studies did not report this measure: Ahmed *et al.* 2015⁷⁰ developed an algorithm based on decision rules; Legrand *et al.* 2013⁹¹ was primarily interested in identifying risk factors of AKI after cardiac surgery; and Scicluna *et al.* 2017⁹³ was primarily concerned with identifying genetic biomarkers of sepsis.

Sensitivity and specificity were reported together in 14 studies^{35,70–72,74,75,78,81–84,87,90,92}. When specificity was not reported, sensitivity was reported together with PPV; and when sensitivity was not reported, this was due to sensitivity being set at a fixed value to report other diagnostic performance measures. In relation to the prior observation, more studies reported PPV than NPV. Four studies reporting likelihood ratios reported both negative and positive likelihood ratios^{70,74,81,84}.

An overview of measured outcomes is illustrated in [Table 11](#).

Ongoing studies. Four trials are currently ongoing, one in Germany and the others in the USA, all concerned with the prediction of sepsis. Three of them are prospective studies and one is retrospective. The retrospective study aims to develop a prediction algorithm based on claims data, EHRs, risk factors and survey data of an estimated 50,000 adult patients admitted to the ED. The German study [NCT03661450](#)⁹⁵ is a single-arm trial evaluating the utility of a CDS system to identify SIRS or sepsis

from EHRs in a pediatric ICU population. Another single-arm trial [NCT03655626](#)⁴⁷ is concerned with implementing a sepsis prediction algorithm in clinical practice as an early warning system. [NCT03644940](#)⁴⁶ is comparing two versions of InSight introduced into clinical practice as an early warning system.

Discussion and conclusions

This systematic literature review shows that over the last 2 decades, there has been an increased interest in CDS as means of supporting clinicians in acute care. CDS has been investigated for several applications ranging from the detection of health conditions^{60,61}, to the prediction of deterioration or adverse events^{40,55,76,81,83,84}. Applications also include therapy guidance, as well as updating clinicians on new or changed recommendations⁹⁶. CDS can also provide guidance by predicting clinical trajectories for different patient profiles over time⁹⁷.

From rule-based algorithms and simple regression models, CDS has evolved to encompass a multitude of techniques in Machine-Learning⁹⁸. These techniques can be dependent on the problem selected and the data types used. Across the three disease areas investigated, the frequent use of random forest classifiers (28.1%), support vector machines (21.9%), boosting techniques (20.3%), LASSO regression (18.8%) and unspecified logistic regression models (10.9%) were observed. The use of more complex modeling such as maximum entropy, Hidden Markov Models (for temporal data analysis) as well as Convolutional Neural Networks has also emerged over the last few years. In the respiratory distress area, the use of NLP models is more common as radiology reports and clinical notes are the main source of input. Different image analysis techniques have been developed to aid in the prediction and diagnosis of respiratory events from radiology images.

Typical measures of NLP model performance include sensitivity, specificity and predictive values. In measuring ML algorithm performance, sensitivity, specificity and ROC AUC are more common. A wide range of outcome measure were reported in research on less-investigated health conditions^{40,67}; and also when uncommon, more complex algorithms were compared to basic algorithms^{74,78,81,84}. This is not surprising given the novelty of these applications.

Many of the ML algorithms and all of the NLP models covered in this work were based on medical data collected in certain clinical sites rather than publicly available data. Datasets from national audits, completed studies or other online sources can additionally play a role, particularly in model validation and testing. This could aid in the adoption and wider use of CDS systems. In this SLR, publicly available datasets were mainly utilized for developing prediction models of heart arrhythmias^{29–31}, hypotension³², septic shock^{28,33,40,41}, COPD⁵⁰, pneumonia³³ and a range of infections^{33,76,78,81,84,86}. In only three cases were they used for testing model performance in sepsis and septic shock prediction; this included the InSight algorithm^{35,85,93}.

Most of the studies identified in this SLR were retrospective and originated in the USA where electronic health records (EHR)

Table 10. Overview of machine learning algorithms evaluated in studies on infection or sepsis.

Study	Predicted disease	Machine learning algorithm																																
		Rule learning	NB	tree augmented NB	AODE	lazy Bayesian rules	Bayesian GLM	Bayesian network analysis	CART	decision tree classifier	neural network	RF	(extreme) gradient boosting	adaptive boosting	ensemble classifier	k nearest neighbor	MARS	GPS	Lasso penalized LR	LR, not specified	SVM	generalized additive model	GLM	stepwise regression	polynomial linear model	polynomial spline regression	Weibull PH model	L2-regularised LR	elastic net regularization					
Ahmed 2015	AKI	✓																																
Legrand, 2013	AKI						✓			✓	✓	✓										✓				✓				✓				
Cheng 2017	AKI											✓		✓																				
Koyner 2015	AKI												✓																					
Bihorac 2018	AKI, sepsis																					✓												
Mohamadlou 2018	AKI, Stage 2/3								✓																									
Chen 2018	AKI, Stage 3								✓	✓	✓					✓																		
Dente, 2017	bacteremia									✓																								
Beeler 2018	CLABSI									✓	✓																							
Parreco 2018	CLABSI									✓			✓																					
LaBarbera 2015	clostridium difficile									✓																								
Wiens 2014	clostridium difficile																																	
Konerman, 2017	fibrosis										✓																							
Hernandez 2017	infection		✓						✓	✓	✓																							
Brasier, 2015	pulmonary aspergillosis																																	
Mani, 2014	sepsis		✓							✓	✓																							
Mao, 2018	sepsis												✓																					

Study	Predicted disease	Machine learning algorithm																															
		Rule learning	NB	tree augmented NB	AODE	lazy Bayesian rules	Bayesian GLM	Bayesian network analysis	CART	decision tree classifier	neural network	RF	(extreme) gradient boosting	adaptive boosting	ensemble classifier	k nearest neighbor	MARS	GPS	Laaso penalized LR	LR, not specified	SVM	generalized additive model	GLM	stepwise regression	polynomial linear model	polynomial spline regression	Weibull PH model	L2-regularised LR	elastic net regularization				
Sciicluna, 2017	sepsis									✓																							
Desautels 2016	sepsis																													✓			
Nemati 2018	sepsis																																
Taneja 2017	sepsis		✓								✓			✓																			
Sanger, 2016	SSI		✓																		✓												
Sohn, 2016	SSI																																
Bartz-Kurycki 2018	SSI										✓																						
Weller 2018	SSI		✓								✓			✓																			
Hu 2016	SSI, UTI, pneumonia, sepsis																				✓												
Taylor, 2018	UTI										✓		✓																				

AKI: Acute kidney injury. SSI: Surgical site infection. UTI: Urinary tract infection. LABSI: Central line-associated bloodstream infections. NB: Naive Bayes. AODE: Averaged one dependence estimators. CART: Classification and regression tree. RF: Random forest. MARS: Multivariate Adaptive Regression Splines GPS: Generalized path seeker algorithm. LR: Logistic regression. SVM: Support vector machine. GLM: Generalized linear model. PH: Proportional hazards.

Table 11. Overview of measured outcomes in studies predicting sepsis or infection.

Study	Sensitivity	Specificity	NPV	PPV	negative LR	positive LR	Accuracy	Prevalence	OR	RR	ROC AUC
Ahmed 2015	✓	✓	✓	✓	✓	✓			✓		
Brasier, 2015							✓				✓
Dente, 2017	✓	✓					✓				✓
Hu, 2016											✓
Konerman, 2017	✓	✓	✓	✓				✓			✓
Legrand, 2013									✓		
Mani, 2014	✓	✓	✓	✓							✓
Mao 2018	✓	✓	✓	✓			✓				✓
Sanger, 2016	✓	✓	✓	✓			✓				✓
Scicluna, 2017								✓			
Sohn, 2016											✓
Taylor, 2018	✓	✓			✓	✓	✓				✓
Hernandez 2017	✓	✓									✓
Bartz-Kurycki 2018											✓
Beeler 2018											✓
Bihorac 2018	✓	✓	✓	✓			✓	✓		✓	✓
Chen 2018	✓			✓							✓
Cheng 2017	✓			✓							✓
Desautels 2016	✓	✓			✓	✓	✓		✓		✓
Koyner 2015	✓	✓	✓	✓							✓
LaBarbera 2015	✓	✓		✓							✓
Mohamadlou 2018	✓	✓			✓	✓	✓		✓		✓
Nemati 2018		✓					✓				✓
Parreco 2018	✓	✓	✓	✓			✓				✓
Taneja 2017											✓
Weller 2018											✓
Wiens 2014	✓			✓							✓

NPV: Negative predictive value. PPV: Positive predictive value. LR: Likelihood ratio. OR: Odds ratio, RR: Risks ratio. ROC AUC: Receiver operator curve area under the curve.

are commonly used. This makes it easier to access and compile large amounts of patient-level information. Many of the studies on shock and infection/sepsis based their models on data extracted from EHRs and utilized large sample sizes. The diversity in the identified CDS systems makes it challenging to draw conclusions on methodology. The lack of comparisons between different classifiers within studies, especially for the

indication of shock, adds to this challenge. To assess the effectiveness of ML algorithms, future research should evaluate multiple algorithms on standard well-labeled datasets.

Class imbalance can be an important issue when training classifiers on datasets for the conditions highlighted in this work. Unequal distributions can arise naturally between disease negative

and positive classes when forming validation sets, particularly when disease prevalence is low⁷⁵. We refer the reader to several machine learning reviews that have addressed this issue^{99–101}. Another important issue in forming disease positive classes relates to the analysis of repeated-measures within subjects, for example, when clinical records are available for each hospitalization day. Several studies have approached this by selecting the first record indicating positive for a health condition. Few researchers have utilized all records and corrected for within-subject variation. An example is the selection of cases depending on observed correlation decay⁵².

In all three areas investigated, the number of retrospective studies exceeded by far the number of prospective studies conducted in a clinical setting. This highlights the challenges in substantiating clinical performance while bringing new clinical decision tools to routine in-hospital patientcare. Examples of algorithms that can be integrated in clinical practice include InSight^{45,46} and Sepsis Watch⁴⁷ which are intended for predicting sepsis and septic shock.

The current systematic literature review did not search multiple bibliographic databases or clinical trial registers; and focused on diagnostic performance rather than other outcomes. In fact, during study screening, trials that evaluated the impact of early warning systems on measures of clinical workflow, rate of re-admissions and/or mortality were discarded as they are somehow out of the focus of this work. This implies that there may be more CDS systems used in practice for the three populations investigated within this research, where the outcomes measured are different. Limiting the search to publications in English and to studies conducted in particular countries; and the exclusion of study protocols identified from the bibliographic database search without checking for later publications from the same authors may have further limited the studies selected. Nevertheless, studies identified within each population represented a diverse range of models applied in different hospital settings trained to predict a range of health conditions. The most widely researched conditions were sepsis and septic shock, venous thromboembolisms, acute kidney injury and surgical site infections.

Specific challenges were identified in collecting sufficient data for training CDS systems on hemodynamic instability. Patients who are, for example, at risk of hemorrhage due to a traumatic injury need to be carefully monitored; and the speed by which they reach a critical state may influence data and study management. It may also be difficult to find healthy volunteers who are willing to undergo procedures like lower body negative pressure which can be unpleasant³⁶. Identification of cases in need of hemodynamic interventions can lend towards larger sample size¹⁹. Other conditions that need further attention are clostridium difficile and CLABSI. Prediction models were driven by almost perfect specificity and very low (<10%) sensitivity^{77,83,86,88}. Considering that these studies used a wide range of features from the EHRs and a large number of patients, except LaBarbera,

Nikiforov⁸³, there is a need to better understand the risk factors to improve sensitivity.

Based on the literature reviewed in this work, as well as several recent surveys and workshops, we would recommend the following points to be addressed when bringing a new CDS tool to critical care^{14,102–104}:

- Integrating CDS in clinical workflows without adding unnecessary extra work to busy clinical teams. The CDS101 toolbox by HIMMS highlights the “CDS five rights”, which are certainly applicable to critical care¹⁰⁵: Providing the right information in the right intervention format, to the right person at the right point in their workflow, and through the right channel.
- Developing tools and concrete proof-points able to assess CDS efficacy in the clinic. This also highlights the importance of providing continuous feedback to clinicians.
- The importance of easy to use user interfaces and focusing on human-computer interaction during deployment.
- Efficient training that is available when needed.
- Being aware of alert or alarm fatigue and not overloading clinicians with alerts due to CDS. The intensive care unit is already plagued with alarms, and if anything, CDS should help in reducing alarms by bundling alerts according to underlying conditions.
- Displaying the rationale for decisions as well as the underlying data to clinical users would lead to improved adoption.
- Understanding ethical challenges for CDS, as well as a careful risk assessment in every site before deployment¹⁰⁶.
- Being able to repeat/standardize implementation across organizations – most prospective studies reviewed in this work covered single centers. Only a few were multi-center studies.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required

Extended data

Figshare: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 1-Search strategy for shock (hemodynamic (in-stability) in MEDLINE. docx. <https://doi.org/10.6084/m9.figshare.9892109.v125>.

Figshare: Working title: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended

data - Table 2-Search strategy for respiratory distress or respiratory failure in MEDLINE.docx. <https://doi.org/10.6084/m9.figshare.9892112.v1>²⁶.

Figshare: Working title: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 3-Search strategy for infection or sepsis in MEDLINE.docx. <https://doi.org/10.6084/m9.figshare.9892115.v1>²⁷.

Reporting guidelines

Figshare: PRISMA checklist for 'Evidence-based Clinical Decision Support Systems for the prediction and detection of three

disease states in critical care: A systematic literature review'. <https://doi.org/10.6084/m9.figshare.9894107.v1>¹⁰⁷.

Data are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Acknowledgments

We would like to thank Mark Connolly from Global Market Access Solutions Sàrl for his contribution during the whole project.

References

- Molina JA, Seow E, Heng BH, *et al.*: **Outcomes of direct and indirect medical intensive care unit admissions from the emergency department of an acute care hospital: a retrospective cohort study.** *BMJ Open.* 2014; 4(11): e005553. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Winters B, Custer J, Galvagno SM Jr, *et al.*: **Diagnostic errors in the intensive care unit: a systematic review of autopsy studies.** *BMJ Qual Saf.* 2012; 21(11): 894–902. [PubMed Abstract](#) | [Publisher Full Text](#)
- Rothschild JM, Landrigan CP, Cronin JW, *et al.*: **The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care.** *Crit Care Med.* 2005; 33(8): 1694–700. [PubMed Abstract](#) | [Publisher Full Text](#)
- Donovan JL, Kanaan AO, Thomson MS, *et al.*: **Effect of clinical decision support on psychotropic medication prescribing in the long-term care setting.** *J Am Geriatr Soc.* 2010; 58(5): 1005–7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Field TS, Rochon P, Lee M, *et al.*: **Computerized clinical decision support during medication ordering for long-term care residents with renal insufficiency.** *J Am Med Inform Assoc.* 2009; 16(4): 480–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kennedy CC, Campbell G, Garg AX, *et al.*: **Piloting a renal drug alert system for prescribing to residents in long-term care.** *J Am Geriatr Soc.* 2011; 59(9): 1757–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tamblyn R, Egualde T, Buckridge DL, *et al.*: **The effectiveness of a new generation of computerized drug alerts in reducing the risk of injury from drug side effects: a cluster randomized trial.** *J Am Med Inform Assoc.* 2012; 19(4): 635–43. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marasinghe KM: **Computerized clinical decision support systems to improve medication safety in long-term care homes: a systematic review.** *BMJ Open.* 2015; 5(5): e006539. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quinn CC, Clough SS, Minor JM, *et al.*: **WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction.** *Diabetes Technol Ther.* 2008; 10(3): 160–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Coiera E, Lau AY, Tsafnat G, *et al.*: **The changing nature of clinical decision support systems: a focus on consumers, genomics, public health and decision safety.** *Yearb Med Inform.* 2009; 84–95. [PubMed Abstract](#) | [Publisher Full Text](#)
- Agoritsas T, Heen AF, Brandt L, *et al.*: **Decision aids that really promote shared decision making: the pace quickens.** *BMJ.* 2015; 350: g7624. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vincent JL, Einav S, Pearse R, *et al.*: **Improving detection of patient deterioration in the general hospital ward environment.** *Eur J Anaesthesiol.* 2018; 35(5): 325–333. [PubMed Abstract](#) | [Free Full Text](#)
- Cox JC, Sadiraj V, Schnier KE, *et al.*: **Higher Quality and Lower Cost from Improving Hospital Discharge Decision Making.** *J Econ Behav Organ.* 2016; 131(B): 1–16. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tcheng JE, Bakken S, Bates DW, *et al.*: **Optimizing Strategies for Clinical Decision Support.** In: *The Learning Health System Series*, N.A.o. Medicine, Editor. Washington DC USA. 2017. [Reference Source](#)
- Duncan H, Hutchison J, Parshuram CS: **The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children.** *J Crit Care.* 2006; 21(3): 271–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Parshuram C, Duncan HP, Joffe AR, *et al.*: **Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children.** *Crit Care.* 2011; 15(4): R184. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chapman SM, Wray J, Oulton K, *et al.*: **'The Score Matters': wide variations in predictive performance of 18 paediatric track and trigger systems.** *Arch Dis Child.* 2017; 102(6): 487–495. [PubMed Abstract](#) | [Publisher Full Text](#)
- Philips. [Accessed 2nd July 2018]. 2018. [Reference Source](#)
- Potes C, Conroy B, Xu-Wilson M, *et al.*: **A clinical prediction model to identify patients at high risk of hemodynamic instability in the pediatric intensive care unit.** *Crit Care.* 2017; 21(1): 282. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hravnak M, Devita MA, Clontz A, *et al.*: **Cardiorespiratory instability before and after implementing an integrated monitoring system.** *Crit Care Med.* 2011; 39(1): 65–72. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gaieski DF, Mikkelsen ME, Band RA, *et al.*: **Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department.** *Crit Care Med.* 2010; 38(4): 1045–53. [PubMed Abstract](#) | [Publisher Full Text](#)
- Critical care statistics. [cited 2019 September 10]. [Reference Source](#)
- Mayr FB, Yende S, Angus DC: **Epidemiology of severe sepsis.** *Virulence.* 2014; 5(1): 4–11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sakr Y, Jaschinski U, Wittebole X, *et al.*: **Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit.** *Open Forum Infect Dis.* 2018; 5(12): ofy313. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Medic G, Kließ MK, Atallah L, *et al.*: **Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 1-Search strategy for shock (hemodynamic (in-)stability) in MEDLINE.docx.** *figshare.* Dataset. 2019. <http://www.doi.org/10.6084/m9.figshare.9892109.v1>
- Medic G, Kließ MK, Atallah L, *et al.*: **Working title: Evidence-based Clinical**

- Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 2-Search strategy for respiratory distress or respiratory failure in MEDLINE.docx.** *figshare*. Dataset. 2019.
<http://www.doi.org/10.6084/m9.figshare.9892112.v1>
27. Medic G, Kließ MK, Atallah L, *et al.*: **Working title: Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. Extended data - Table 3-Search strategy for infection or sepsis in MEDLINE.docx.** *figshare*. Dataset. 2019.
<http://www.doi.org/10.6084/m9.figshare.9892115.v1>
 28. Ghosh S, Li J, Cao L, *et al.*: **Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns.** *J Biomed Inform.* 2017; 66: 19–31.
[PubMed Abstract](#) | [Publisher Full Text](#)
 29. Li Q, Rajagopalan C, Clifford GD: **Ventricular fibrillation and tachycardia classification using a machine learning approach.** *IEEE Trans Biomed Eng.* 2014; 61(6): 1607–13.
[PubMed Abstract](#) | [Publisher Full Text](#)
 30. Ebrahimzadeh E, Kalantari M, Joulani M, *et al.*: **Prediction of paroxysmal Atrial Fibrillation: A machine learning based approach using combined feature vector and mixture of expert classification on HRV signal.** *Comput Methods Programs Biomed.* 2018; 165: 53–67.
[PubMed Abstract](#) | [Publisher Full Text](#)
 31. Strodthoff N, Strodthoff C: **Detecting and interpreting myocardial infarction using fully convolutional neural networks.** *Physiol Meas.* 2018; 40(1): 015001.
[PubMed Abstract](#) | [Publisher Full Text](#)
 32. Donald R, Howells T, Piper I, *et al.*: **Forewarning of hypotensive events using a Bayesian artificial neural network in neurocritical care.** *J Clin Monit Comput.* 2018; 33(1): 39–51.
[PubMed Abstract](#) | [Publisher Full Text](#)
 33. Hu Z, Melton GB, Moeller ND, *et al.*: **Accelerating Chart Review Using Automated Methods on Electronic Health Record Data for Postoperative Complications.** *AMIA Annu Symp Proc.* 2016; 2016: 1822–1831.
[PubMed Abstract](#) | [Free Full Text](#)
 34. Mahajan D, Dong Y, Saxton LA, *et al.*: **Performance of an automatic arrhythmia classification algorithm: comparison to the ALTITUDE electrophysiologist panel adjudications.** *Pacing Clin Electrophysiol.* 2014; 37(7): 889–99.
[PubMed Abstract](#) | [Publisher Full Text](#)
 35. Mao Q, Jay M, Hoffman JL, *et al.*: **Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU.** *BMJ Open.* 2018; 8(1): e017833.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. Reifin N, Zimmer G, Malyuta Y, *et al.*: **Using support vector machines on photoplethysmographic signals to discriminate between hypovolemia and euvoolemia.** *PLoS One.* 2018; 13(3): e0195087.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 37. Sideris C, Pourhomayoun M, Kalantarian H, *et al.*: **A flexible data-driven comorbidity feature extraction framework.** *Comput Biol Med.* 2016; 73: 165–72.
[PubMed Abstract](#) | [Publisher Full Text](#)
 38. Blecker S, Katz SD, Horwitz LI, *et al.*: **Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data.** *JAMA Cardiol.* 2016; 1(9): 1014–1020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Blecker S, Sontag D, Horwitz LI, *et al.*: **Early Identification of Patients With Acute Decompensated Heart Failure.** *J Card Fail.* 2018; 24(6): 357–362.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 40. Calvert J, Desautels T, Chettipally U, *et al.*: **High-performance detection and early prediction of septic shock for alcohol-use disorder patients.** *Ann Med Surg (Lond).* 2016; 8: 50–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. Henry K, Hager DN, Pronovost PJ, *et al.*: **A targeted real-time early warning score (TREWScore) for septic shock.** *Sci Transl Med.* 2015; 7(299): 299ra122.
[PubMed Abstract](#) | [Publisher Full Text](#)
 42. Panahiazar M, Taslimithrani V, Pereira N, *et al.*: **Using EHRs and Machine Learning for Heart Failure Survival Analysis.** *Stud Health Technol Inform.* 2015; 216: 40–4.
[PubMed Abstract](#) | [Free Full Text](#)
 43. NCT02934971: **Optimized Multi-modality Machine Learning Approach During Cardio-toxic Chemotherapy to Predict Arising Heart Failure (MERMAID).**
[Reference Source](#)
 44. NCT03582501: **Measurement of Hemodynamic Responses to Lower Body Negative Pressure (LBNP).**
[Reference Source](#)
 45. NCT03235193: **Predictive algorithm for Evaluation and Intervention in Sepsis (PREVISE).**
[Reference Source](#)
 46. NCT03644940: **Subpopulation-Specific Sepsis Identification Using Machine Learning.**
[Reference Source](#)
 47. NCT03655626: **Implementation and Evaluations of Sepsis Watch.**
[Reference Source](#)
 48. Bejan CA, Vanderwende L, Evans HL, *et al.*: **On-time clinical phenotype prediction based on narrative reports.** *AMIA Annu Symp Proc.* 2013; 2013: 103–10.
[PubMed Abstract](#) | [Free Full Text](#)
 49. Kumamaru KK, George E, Aghayev A, *et al.*: **Implementation and Performance of Automated Software for Computing Right-to-Left Ventricular Diameter Ratio From Computed Tomography Pulmonary Angiography Images.** *J Comput Assist Tomogr.* 2016; 40(3): 387–92.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 50. Bodduluri S, Newell JD Jr, Hoffman EA, *et al.*: **Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework.** *Acad Radiol.* 2013; 20(5): 527–36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 51. Biesiada J, Chidambaram V, Wagner M, *et al.*: **Genetic risk signatures of opioid-induced respiratory depression following pediatric tonsillectomy.** *Pharmacogenomics.* 2014; 15(14): 1749–1762.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 52. Reamaron N, Sjöding MW, Lin K, *et al.*: **Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome.** *IEEE J Biomed Health Inform.* 2019; 23(1): 407–415.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 53. Vinson DR, Morley JE, Huang J, *et al.*: **The Accuracy of an Electronic Pulmonary Embolism Severity Index Auto-Populated from the Electronic Health Record: Setting the stage for computerized clinical decision support.** *Appl Clin Inform.* 2015; 6(2): 318–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 54. Huesch MD, Cherian R, Labib S, *et al.*: **Evaluating Report Text Variation and Informativeness: Natural Language Processing of CT Chest Imaging for Pulmonary Embolism.** *J Am Coll Radiol.* 2018; 15(3 Pt B): 554–562.
[PubMed Abstract](#) | [Publisher Full Text](#)
 55. Mortazavi BJ, Desai N, Zhang J, *et al.*: **Prediction of Adverse Events in Patients Undergoing Major Cardiovascular Procedures.** *IEEE J Biomed Health Inform.* 2017; 21(6): 1719–1729.
[PubMed Abstract](#) | [Publisher Full Text](#)
 56. González G, Ash SY, Vegas-Sánchez-Ferrero G, *et al.*: **Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography.** *Am J Respir Crit Care Med.* 2018; 197(2): 193–203.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 57. Choi Y, Liu TT, Pankratz DG, *et al.*: **Identification of usual interstitial pneumonia pattern using RNA-Seq and machine learning: challenges and solutions.** *BMC Genomics.* 2018; 19(Suppl 2): 101.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 58. Yu S, Kumamaru KK, George E, *et al.*: **Classification of CT pulmonary angiography reports by presence, chronicity, and location of pulmonary embolism with natural language processing.** *J Biomed Inform.* 2014; 52: 386–93.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 59. Swartz J, Koziatsek C, Theobald J, *et al.*: **Creation of a simple natural language processing tool to support an imaging utilization quality dashboard.** *Int J Med Inform.* 2017; 101: 93–99.
[PubMed Abstract](#) | [Publisher Full Text](#)
 60. Liu V, Clark MP, Mendoza M, *et al.*: **Automated identification of pneumonia in chest radiograph reports in critically ill patients.** *BMC Med Inform Decis Mak.* 2013; 13: 90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 61. Haug PJ, Ferraro JP, Holmen J, *et al.*: **An ontology-driven, diagnostic modeling system.** *J Am Med Inform Assoc.* 2013; 20(e1): e102–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 62. Dublin S, Baldwin E, Walker RL, *et al.*: **Natural Language Processing to identify pneumonia from radiology reports.** *Pharmacoepidemiol Drug Saf.* 2013; 22(8): 834–41.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 63. Jones BE, South BR, Shao Y, *et al.*: **Development and Validation of a Natural Language Processing Tool to Identify Patients Treated for Pneumonia across VA Emergency Departments.** *Appl Clin Inform.* 2018; 9(1): 122–128.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 64. Rochefort CM, Verma AD, Eguale T, *et al.*: **A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data.** *J Am Med Inform Assoc.* 2015; 22(1): 155–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 65. Tian Z, Sun S, Eguale T, *et al.*: **Automated Extraction of VTE Events From Narrative Radiology Reports in Electronic Health Records: A Validation Study.** *Med Care.* 2017; 55(10): e73–e80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 66. Pham AD, Névéal A, Lavergne T, *et al.*: **Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings.** *BMC Bioinformatics.* 2014; 15: 266.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 67. Silva S, Ait Aissa D, Cocquet P, *et al.*: **Combined Thoracic Ultrasound Assessment during a Successful Weaning Trial Predicts Postextubation Distress.** *Anesthesiology.* 2017; 127(4): 666–674.
[PubMed Abstract](#) | [Publisher Full Text](#)
 68. Phillips C, Mac Parthaláin N, Syed Y, *et al.*: **Short-Term Intra-Subject Variation in Exhaled Volatile Organic Compounds (VOCs) in COPD Patients and Healthy**

- Controls and Its Effect on Disease Classification.** *Metabolites.* 2014; 4(2): 300–18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
69. Phillips R, Williams D, Bowen D, *et al.*: **Reaching a consensus on research priorities for supporting women with autoimmune rheumatic diseases during pre-conception, pregnancy and early parenting: A Nominal Group Technique exercise with lay and professional stakeholders [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2018; 3: 75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
70. Ahmed A, Vairavan S, Akhouni A, *et al.*: **Development and validation of electronic surveillance tool for acute kidney injury: A retrospective analysis.** *J Crit Care.* 2015; 30(5): 988–93.
[PubMed Abstract](#) | [Publisher Full Text](#)
71. Koneran MA, Lu D, Zhang Y, *et al.*: **Assessing risk of fibrosis progression and liver-related clinical outcomes among patients with both early stage and advanced chronic hepatitis C.** *PLoS One.* 2017; 12(11): e0187344.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
72. Mani S, Ozdas A, Aliferis C, *et al.*: **Medical decision support using machine learning for early detection of late-onset neonatal sepsis.** *J Am Med Inform Assoc.* 2014; 21(2): 326–36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
73. Sohn S, Larson DW, Habermann EB, *et al.*: **Detection of clinically important colorectal surgical site infection using Bayesian network.** *J Surg Res.* 2017; 209: 168–173.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
74. Taylor RA, Moore CL, Cheung KH, *et al.*: **Predicting urinary tract infections in the emergency department with machine learning.** *PLoS One.* 2018; 13(3): e0194085.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
75. Hernandez B, Herrero P, Rawson TM, *et al.*: **Supervised learning for infection risk inference using pathology data.** *BMC Med Inform Decis Mak.* 2017; 17(1): 168.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
76. Bartz-Kurycki MA, Green C, Anderson KT, *et al.*: **Enhanced neonatal surgical site infection prediction model utilizing statistically and clinically significant variables in combination with a machine learning algorithm.** *Am J Surg.* 2018; 216(4): 764–777.
[PubMed Abstract](#) | [Publisher Full Text](#)
77. Beeler C, Dbeibo L, Kelley K, *et al.*: **Assessing patient risk of central line-associated bacteremia via machine learning.** *Am J Infect Control.* 2018; 46(9): 986–991.
[PubMed Abstract](#) | [Publisher Full Text](#)
78. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, *et al.*: **MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery.** *Ann Surg.* 2019; 269(4): 652–662.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
79. Chen W, Hu Y, Zhang X, *et al.*: **Causal risk factor discovery for severe acute kidney injury using electronic health records.** *BMC Med Inform Decis Mak.* 2018; 18(Suppl 1): 13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
80. Cheng P, Waitman LR, Hu Y, *et al.*: **Predicting Inpatient Acute Kidney Injury over Different Time Horizons: How Early and Accurate?** *AMIA Annu Symp Proc.* 2018; 2017: 565–574.
[PubMed Abstract](#) | [Free Full Text](#)
81. Desautels T, Calvert J, Hoffman J, *et al.*: **Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach.** *JMIR Med Inform.* 2016; 4(3): e28.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
82. Koyner JL, Carey KA, Edelson DP, *et al.*: **The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model.** *Crit Care Med.* 2018; 46(7): 1070–1077.
[PubMed Abstract](#) | [Publisher Full Text](#)
83. LaBarbera FD, Nikiforov I, Parvathani A, *et al.*: **A prediction model for Clostridium difficile recurrence.** *J Community Hosp Intern Med Perspect.* 2015; 5(1): 26033.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
84. Mohamadlou H, Lynn-Palevsky A, Barton C, *et al.*: **Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data.** *Can J Kidney Health Dis.* 2018; 5: 2054358118776326.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
85. Nemati S, Holder A, Razmi F, *et al.*: **An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU.** *Crit Care Med.* 2018; 46(4): 547–553.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
86. Parreco JP, Hidalgo AE, Badilla AD, *et al.*: **Predicting central line-associated bloodstream infections and mortality using supervised machine learning.** *J Crit Care.* 2018; 45: 156–162.
[PubMed Abstract](#) | [Publisher Full Text](#)
87. Weller GB, Lovely J, Larson DW, *et al.*: **Leveraging electronic health records for predictive modeling of post-surgical complications.** *Stat Methods Med Res.* 2018; 27(11): 3271–3285.
[PubMed Abstract](#) | [Publisher Full Text](#)
88. Wiens J, Campbell WN, Franklin ES, *et al.*: **Learning Data-Driven Patient Risk Stratification Models for Clostridium difficile.** *Open Forum Infect Dis.* 2014; 1(2): ofu045.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
89. Brasier AR, Zhao Y, Spratt HM, *et al.*: **Improved Detection of Invasive Pulmonary Aspergillosis Arising during Leukemia Treatment Using a Panel of Host Response Proteins and Fungal Antigens.** *PLoS One.* 2015; 10(11): e0143165.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
90. Dente CJ, Bradley M, Schobel S, *et al.*: **Towards precision medicine: Accurate predictive modeling of infectious complications in combat casualties.** *J Trauma Acute Care Surg.* 2017; 83(4): 609–616.
[PubMed Abstract](#) | [Publisher Full Text](#)
91. Legrand M, Pirracchio R, Rosa A, *et al.*: **Incidence, risk factors and prediction of post-operative acute kidney injury following cardiac surgery for active infective endocarditis: an observational study.** *Crit Care.* 2013; 17(5): R220.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
92. Sanger PC, van Ramshorst GH, Mercan E, *et al.*: **A Prognostic Model of Surgical Site Infection Using Daily Clinical Wound Assessment.** *J Am Coll Surg.* 2016; 223(2): 259–270.e2.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
93. Scicluna BP, van Vught LA, Zwinderman AH, *et al.*: **Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study.** *Lancet Respir Med.* 2017; 5(10): 816–826.
[PubMed Abstract](#) | [Publisher Full Text](#)
94. Taneja I, Reddy B, Damhorst G, *et al.*: **Combining Biomarkers with EMR Data to Identify Patients in Different Phases of Sepsis.** *Sci Rep.* 2017; 7(1): 10800.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
95. NCT03661450: **Evaluation of the Accuracy of a Clinical Decision-Support System (CDSS) to Support Detection of SIRS and Sepsis in Paediatric Intensive Care Patients Compared to Medical Specialists.** [Reference Source](#)
96. Van de Velde S, Korteisto T, Spitaels D, *et al.*: **Development of a Tailored Intervention With Computerized Clinical Decision Support to Improve Quality of Care for Patients With Knee Osteoarthritis: Multi-Method Study.** *JMIR Res Protoc.* 2018; 7(6): e154.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
97. Pinaire J, Azé J, Bringay S, *et al.*: **Patient healthcare trajectory. An essential monitoring tool: a systematic review.** *Health Inf Syst Syst.* 2017; 5(1): 1.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
98. Middleton B, Sittig DF, Wright A: **Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision.** *Yearb Med Inform.* 2016; (Suppl 1): S103–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
99. Longadge R, Dongre S: **Class Imbalance Problem in Data Mining Review.** *arXiv e-prints.* 2013.
[Reference Source](#)
100. Buda M, Maki A, Mazurowski MA: **A systematic study of the class imbalance problem in convolutional neural networks.** *Neural Netw.* 2018; 106: 249–259.
[PubMed Abstract](#) | [Publisher Full Text](#)
101. Nanni L, Fantozzi C, Lazzarini N: **Coupling different methods for overcoming the class imbalance problem.** *Neurocomputing.* 2015; 158: 48–61.
[Publisher Full Text](#)
102. Kindle RD, Badawi O, Celi LA, *et al.*: **Intensive Care Unit Telemedicine in the Era of Big Data, Artificial Intelligence, and Computer Clinical Decision Support Systems.** *Crit Care Clin.* 2019; 35(3): 483–495.
[PubMed Abstract](#) | [Publisher Full Text](#)
103. Rumsfeld JS, Joynt KE, Maddox TM: **Big data analytics to improve cardiovascular care: promise and challenges.** *Nat Rev Cardiol.* 2016; 13(6): 350–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
104. (NQF). **NQF: Driving Quality and Performance Measurement—A Foundation for Clinical Decision Support: A Consensus Report.** 2010; NQF: Washington, DC.
[Reference Source](#)
105. HIMMS: **Clinical Decision Support 101.** 2019; [cited 2019].
[Reference Source](#)
106. Char DS, Shah NH, Magnus D: **Implementing Machine Learning in Health Care - Addressing Ethical Challenges.** *N Engl J Med.* 2018; 378(11): 981–983.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
107. Medic G, Kließ MK, Atallah L, *et al.*: **Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review.** *PRISMA Checklist.* *figshare.* Dataset. 2019. <http://www.doi.org/10.6084/m9.figshare.9894107.v1>

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 04 December 2019

<https://doi.org/10.5256/f1000research.23644.r57156>

© 2019 Nikolakopoulos S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **Stavros Nikolakopoulos** 

Department of Biostatistics & Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

My comments were adequately addressed in the text. No further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 28 November 2019

<https://doi.org/10.5256/f1000research.23644.r57155>

© 2019 Kovacevic M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **Milena Kovacevic** 

Department of Pharmacokinetics and Clinical Pharmacy, Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia

Thank you for addressing my comments. I have no further comments to make.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Pharmacokinetics and Clinical Pharmacy; Patient outcomes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 18 November 2019

<https://doi.org/10.5256/f1000research.22530.r56200>

© 2019 Kovacevic M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Milena Kovacevic 

Department of Pharmacokinetics and Clinical Pharmacy, Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia

The review summarizes the utilization of clinical decision support (CDS) systems in three selected states in critical care – shock/hemodynamic (in-)stability; respiratory distress/failure; and infection/sepsis. The background of the study has a strong rationale.

The study comprised the results from primary sources, describing models/algorithms used to detect and alert clinicians to the presence of these conditions, as well as models/algorithms developed to predict deterioration in an individual patient state, leading to these selected conditions.

The systematic review was performed and the findings are presented in line with the PRISMA guidelines. Variables for which data were sought were clearly stated (PICOS) in Table 1.

Specific comments:

- What I found especially beneficial for the readers and future research in this area, is Table 2 with the presented collected data used for training algorithms.
- It would be beneficial to provide additional information whether an internal or external validation was performed - within Table 4 (measured outcomes in studies on shock), Table 8 (measured outcomes in studies on respiratory distress/failure) and Table 11 (measured outcomes in studies on infection/sepsis).
- What was the rationale for including the studies predicting acute kidney injury within the Infection/sepsis results section? If it is about the decline in glomerular filtration rate due to hypotension seen in sepsis, it might have been presented within the Shock section.
- Table 7: include the abbreviations for ARDS (Acute respiratory distress syndrome), ARDE (Acute respiratory disease events) and DVT (deep vein thrombosis) below the Table.
- Table 9: include the abbreviation for AKI (Acute kidney injury) below the Table.

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Not applicable

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Pharmacokinetics and Clinical Pharmacy; Patient outcomes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 06 November 2019

<https://doi.org/10.5256/f1000research.22530.r56199>

© 2019 Nikolakopoulos S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stavros Nikolakopoulos 

Department of Biostatistics & Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

The authors report on a systematic review in order to assess the state-of-the-art in the field of Clinical Decision Support (CDS) systems in the last 5 years (2013-2018). They review and report on study designs, outcomes and methods employed in CDS in the scientific literature as well as in study databases (like [Clinicaltrials.gov](#)).

The paper is clearly written and organized. The methodology for the systematic review is solid and comprehensive. The topic is also very relevant and timely. I do have some concerns which are mentioned below:

- The authors could potentially include in the study (as described by the inclusion criteria), conference abstracts that were published only as abstracts in 2017 or 2018, even without subsequent publication. I assume they do that in order to somehow keep up with later developments even if they are not published elsewhere, given the very fast pace of the research area. However, they exclude protocols of studies that were published in the same (or more extended) time frame, which seems slightly inconsistent. Some discussion concerning this choice would be enlightening.
- There seems to be some confusion with terminology, with unknown consequences on the review's results. The authors seem to separate "machine learning" methods, from "statistical" methods (

Table 1: "Multivariable hierarchical logistic regression models*** (models which are based only on statistics - but there is no machine learning)", as an exclusion criterion). This is clearly not the suitable platform to resolve this issue, but, the distinction between machine learning and statistics is not at all that clear. Specifically, under the term "supervised learning", any regression method (statistics) could be classified. So, logistic regression IS a machine learning method. So is LASSO and several other methods reported. Again, this is not the appropriate place for going into further details, but there is certainly some confusion, especially when in the results Logistic regression keeps appearing as a preferred method.

- Again concerning terminology, the term "accuracy" appears often in the results section. Sometimes it is reported as a different outcome than i.e. ROC AUC, sensitivity and specificity. All the latter methods are quantifying "accuracy" in some way and some clarification is needed.

Minor comments:

- Table 1: Treatment/Intervention, a parenthesis is missing.
- Tables 7 & 10: Maybe reverse the orientation of the column titles, it is impossible to read on a screen.

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Not applicable

Are the conclusions drawn adequately supported by the results presented in the review?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research