Journal of the American Medical Informatics Association, 2023, **30(12)**, 2050–2063
https://doi.org/10.1093/jamia/ocad180
Advance access publication 30 August 2023
Review

Review

Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review

Anindya Pradipta Susanto [1,2,*], David Lyell [1], Bambang Widyantoro [2,3], Shlomo Berkovsky [1], Farah Magrabi [1]

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia, [2]Faculty of Medicine, Universitas Indonesia, Jakarta, DKI Jakarta 10430, Indonesia, [3]National Cardiovascular Center Harapan Kita Hospital, Jakarta, DKI Jakarta 11420, Indonesia

*Corresponding author: Anindya Pradipta Susanto, MD, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia (anindya.susanto@hdr.mq.edu.au, anindya.susanto@ui.ac.id)

Abstract

Objective: This study aims to summarize the research literature evaluating machine learning (ML)-based clinical decision support (CDS) systems in healthcare settings.

Materials and methods: We conducted a review in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta Analyses extension for Scoping Review). Four databases, including PubMed, Medline, Embase, and Scopus were searched for studies published from January 2016 to April 2021 evaluating the use of ML-based CDS in clinical settings. We extracted the study design, care setting, clinical task, CDS task, and ML method. The level of CDS autonomy was examined using a previously published 3-level classification based on the division of clinical tasks between the clinician and CDS; effects on decision-making, care delivery, and patient outcomes were summarized.

Results: Thirty-two studies evaluating the use of ML-based CDS in clinical settings were identified. All were undertaken in developed countries and largely in secondary and tertiary care settings. The most common clinical tasks supported by ML-based CDS were image recognition and interpretation ($n = 12$) and risk assessment ($n = 9$). The majority of studies examined assistive CDS ($n = 23$) which required clinicians to confirm or approve CDS recommendations for risk assessment in sepsis and for interpreting cancerous lesions in colonoscopy. Effects on decision-making, care delivery, and patient outcomes were mixed.

Conclusion: ML-based CDS are being evaluated in many clinical areas. There remain many opportunities to apply and evaluate effects of ML-based CDS on decision-making, care delivery, and patient outcomes, particularly in resource-constrained settings.

Key words: artificial intelligence; automation; clinical decision support; evaluation; machine learning.

Background and significance

Contemporary clinical decision support (CDS) systems are increasingly being embedded with artificial intelligence (AI), especially machine learning (ML) models trained on a wide variety of clinical datasets.[1,2] Like the previous generation of CDS which were largely based on human-engineered rules, *ML-based CDS* can support clinicians in tasks such as disease diagnosis, treatment selection, patient monitoring, and risk stratification for primary prevention.[3] While many studies have demonstrated the performance of ML models for specific clinical tasks,[4,5] little is known about the use of ML-based CDS in healthcare settings as well as their effects on decision-making, care delivery, and patient outcomes.[6] In contrast, rule-based CDS have been shown to be effective in improving care delivery and patient outcomes in a wide-variety of clinical tasks such as computerized provider order entry (CPOE) and electronic prescribing, diagnostic assistance, and for preventive care reminders.[7,8]

Previous reviews have examined the application of AI in specific health conditions, such as colonoscopy, stroke, and sepsis.[9–11] This scoping review aims to take a broader view by summarizing the research literature about the evaluation of ML-based CDS in clinical settings. Here, ML models must perform well on real-world populations and ML-based CDS need to be seamlessly integrated with clinical workflows as well as the existing information technology (IT) infrastructure.[12] To better understand the role of ML-based CDS in clinical tasks, we examined the level of system autonomy[13] and summarized effects on decision-making, care delivery, and patient outcomes.[14] As ML-based CDS operates within a human-technology system, clinician interaction with CDS influences how they make decisions that then affect care delivery and patient outcomes. Previous reviews that have examined AI in clinical settings have not considered the level of system autonomy or the specific role of the CDS in clinical tasks.[15–17]

Received: May 10, 2023; **Revised:** August 1, 2023; **Editorial Decision:** August 20, 2023; **Accepted:** August 23, 2023

Materials and methods

Given that the research literature about the evaluation of ML-based CDS appears young and heterogeneous, a scoping review was undertaken focusing on studies reporting the evaluation of ML-based CDS systems in clinical settings and their effects on decision-making, care delivery, or patient outcomes. The review was conducted using the methodology outlined by Arksey and O'Malley[18] and refined by Levac et al.[19] This framework consists of 5 steps: setting the research question; searching relevant studies; selecting the study based on inclusion and exclusion criteria; extracting the data; and collating, summarizing, and reporting the results. Reporting was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta Analyses extension for Scoping Reviews) standard.[20]

Search strategy

Bibliographic databases, including PubMed, Medline, Embase, and Scopus were searched in April 2021. The search query used was ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("clinical decision support" OR "computer-assisted") AND (predict* OR screen* OR diagno* OR treat* OR manag* OR detect* OR prescri* OR prognosis OR triage OR monitor*) AND (inpatient* OR in-patient OR "clinical setting" OR "hospital setting" OR "primary care"). Appropriate vocabulary terms were included (Table S1), and the retrieval set was limited to articles published in 2016 or later. In addition to the 4 databases, we searched ClinicalTrials.gov and manually searched the reference lists of the retrieved articles using a forward-backward snowballing approach.

Study selection

The search identified a total of 1255 studies (Figure 1). After removal of duplicates, the title and abstract of 1111 articles were screened independently by 2 reviewers (A.P.S. and F.M.) to identify relevant studies according to inclusion and exclusion criteria (Table S2).[21] Articles were limited to literature published in commercial bibliographic databases, including primary studies published over a 5-year period (2016-2021). Our search was limited to 5 years as the application of ML-based CDS and medical device approvals were mostly since 2016.[13,15] Studies about systems not used by healthcare professionals (ie, consumer facing systems without clinician supervision) or those reporting development or validation of models on retrospective datasets were excluded. Non-English articles and conference abstracts were also excluded leaving 45 studies for further assessment. Full-length articles were retrieved and assessed independently against the inclusion criteria by 2 reviewers (A.P.S. and F.M.). Articles not meeting the inclusion criteria were excluded and any disagreements about inclusion or exclusion of an article were resolved by consensus.

Data extraction and synthesis

For each included study, descriptive information about the clinical task, care setting, study design, CDS users, CDS task, ML type, and method were extracted. Key findings about CDS effects on clinical decision-making, care delivery, and patient outcomes were also identified. The following data were extracted:

Geo-economic setting

We examined the countries where the study took place and classified them using the World Bank's classification by income.[22]

CDS task

CDS systems can assist with a variety of clinical tasks. We categorized CDS tasks into[1]: evidence retrieval, CPOE and electronic prescribing, diagnostic assistance, therapy planning and critiquing, risk assessment, process support systems, image recognition and interpretation including computer aided diagnosis, and expert laboratory information systems.

ML type and method

Where available the type and method of ML were examined. ML type was categorized into supervised learning, unsupervised learning, and reinforcement learning. ML methods were reported as extracted from the literature. A study could be assigned to 1 or more method(s).

Level of autonomy

The level of autonomy was examined using a previously published 3-level classification based on how clinical tasks are divided between the clinician and CDS[13]:

1) Autonomous information: These CDS systems are characterized by a separation between what CDS and clinician contribute to the task, where CDS contributes information that clinicians can use to make decisions, for example, an imaging system provides a colored display to enhance clinician's perception to differentiate human tissue images.
2) Assistive: These CDS are characterized by overlap in what clinician and CDS contribute to the task, but where clinicians provide the decision on the task. Such overlap or duplication occurs when clinicians need to confirm or approve CDS provided information or decisions; for example, a CDS assists clinicians to detect osteoarthritis from a knee X-ray image with a disclaimer that the system should be used in lieu of full patient evaluation.
3) Autonomous decision: Here CDS provides the decision for the clinical task which can then be enacted by clinicians or the CDS itself, for example, a CDS provides screening for diabetic retinopathy in primary practice where the result is directly used as referral decisions.

To determine the level of autonomy, we examined the CDS task, the stage of human information processing that was automated by the CDS,[23] as well as the CDS input and output. The clinician task and CDS task were subsequently compared to assess whether clinicians needed to verify decisions provided by the CDS system (assistive) or could rely on the CDS information or decisions (autonomous). Classification of the stage of automation and level of autonomy was performed by A.P.S. and reviewed by D.L.

Effects on decision-making, care delivery, and patient outcomes

CDS effects were examined using an established framework called the *information value chain*, which shows that multiple steps are necessary from using ML-based CDS to impacting patient outcomes including clinicians interacting with CDS

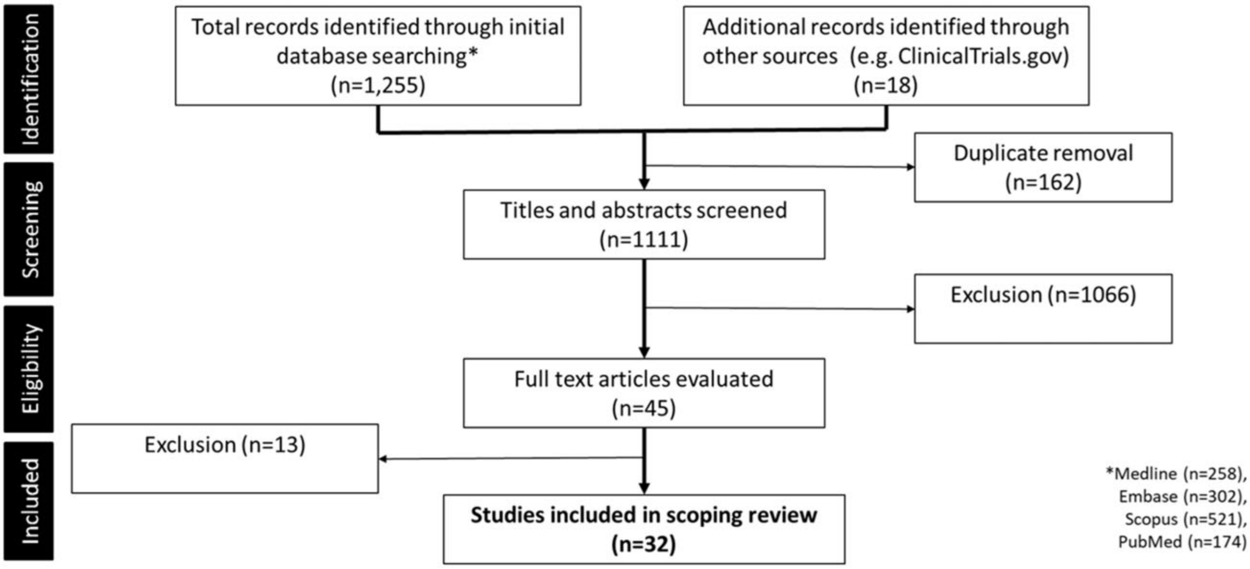

Figure 1. Article search and retrieval process flow diagram.

systems, receiving new information that then alters decisions, the care delivery process, and outcomes.[14,24] For instance, a clinician may interact with a CDS system to receive important new information based on which they decide to implement (decision-making) an intervention (care delivery) that leads to changes in the patient's clinical condition (patient outcomes). Thus, effects of CDS systems were examined based on changes in decision-making, care delivery, or patient outcomes. When studies did not evaluate these effects, we summarized their key findings.

A narrative synthesis then integrated findings into descriptive summaries for each level of autonomy. We focused on the clinical task assisted by CDS and reported effects on decision-making, care delivery, or patient outcomes.

Results

Descriptive analysis of all studies

We identified 32 studies describing the evaluation of ML-based CDS in healthcare settings (Figure 1, Table 1). The majority were prospective cohort studies ($n=18$; 56%) or randomized controlled trials (RCTs) ($n=9$; 28%) in secondary-tertiary care settings ($n=29$; 91%). Only 14 (44%) reported the clinical trial registration. Most studies were conducted after 2019 ($n=27$; 84%) in high-income nations ($n=24$; 75%; Figure 2).

Of the 32 studies reviewed, the most common task supported by ML-based CDS was image recognition and interpretation ($n=12$; 38%) followed by risk assessment ($n=9$; 28%) where sepsis ($n=5$) was the predominant risk to be evaluated (Table 2). Most CDS were based on supervised learning ($n=28$; 88%) using a wide variety of methods, including random forest ($n=7$) and convolutional neural networks ($n=7$). While convolutional neural networks were mainly used for image recognition ($n=6$), classic ML methods such as random forest ($n=4$) and gradient boosting ($n=2$; Table 3) were utilized for risk assessment. A summary of the studies by the level of autonomy of the CDS is given in the following sections (Table 4).

Assistive: CDS assists human decisions

Most studies examined ML-based CDS that were assistive ($n=23$; Table 2). Here clinicians needed to confirm or approve CDS provided information or decisions such as recommendations, alerts, or risk scores for diagnosis or further actions. The following sections summarize these studies by the CDS task.

Risk assessment

Eight studies related to CDS that assisted clinicians in assessing the risk of complications during hospitalization. Of these, 5 alerted clinicians about sepsis risk by text message or phone call providing them with a predictive risk score for interpretation, triggering re-assessment and further action should the clinician observe heightened risks.[27–29,42,43] One study used the learning health system framework to integrate a deep learning sepsis CDS into routine care.[42] In 2 other studies, the majority of clinicians (55-62%) did not change their perceptions about sepsis risk and reported no change in care delivery, although few orders were significantly increased such as intravenous bolus, hematology, and metabolic blood tests.[28,29] However, 2 studies, including 1 RCT by Shimabukuro et al., demonstrated that the use of a CDS to predict sepsis risk shortened hospital stay and reduced mortality.[27,43]

Risk of delirium, another complication of hospitalization, was predicted by 1 CDS using a random forest-based algorithm achieving a sensitivity of 74% and a specificity of 82%.[32,33] Although most clinicians indicated that the information about delirium and its early detection provided by the CDS was useful ($n=47$; 68%), only 1 in 3 (33%) reported using the system and considering its recommendations in their clinical decisions (19%).[32] Another study that compared the accuracy of perioperative risk assessment between physicians and CDS found physicians changing their risk-assessment score in more than 75% of cases ($n=150$).[26]

Image recognition and interpretation

The next most common CDS task was image recognition and interpretation ($n=8$). Six studies examined polyp detection during colonoscopy, where CDS assisted clinicians to

Table 1. Characteristics of studies reporting evaluation of ML-based CDS in healthcare settings ($n = 32$).

Characteristics	n	%
Study design		
Experimental, randomized controlled trial	10	31
Experimental, nonrandomized trial	1	3
Experimental, cross-over trial	1	3
Observational, prospective cohort	17	53
Observational, cross-sectional	2	6
Mixed-methods	1	3
Year of publication		
2021	8	25
2020	13	41
2019	6	19
2018	2	6
2017	2	6
2016	1	3
Clinical trial registration	14	44
Geo-economic setting		
High-income countries (HIC)	24	75
Upper middle-income countries (UMIC)	8	25
Lower middle-income countries (LMIC)	–	–
Low-income countries (LIC)	–	–
Clinical setting		
Primary care	2	6
Secondary-tertiary care	29	91
Community care	1	3
CDS task		
Image recognition & interpretation	12	38
Risk assessment	9	28
Diagnostic assistance	5	16
Treatment planning & critiquing	3	9
Process support system	2	6
CPOE & electronic prescribing	1	3
Evidence retrieval	–	–
Expert laboratory information system	–	–
ML type		
Supervised learning	28	88
Reinforcement learning	1	3
Unsupervised learning	–	–
Not reported	3	9
ML method[a]		
Support vector machine	3	9
Random forest	7	22
Logistic regression	2	6
Gradient boosting	3	9
Convolutional neural network	7	22
Artificial neural network	3	9
(Unspecified) deep learning	3	9

Abbreviations: CDS, clinical decision support; CPOE, computer processed order entry; ML, machine learning.
[a] CDS could be assigned to 1 or more ML method(s).

distinguish between cancerous adenoma lesions and benign polyps.[30,34,35,38,46,47] Of these, 5 RCTs demonstrated superiority of CDS in assisting clinicians.[30,34,38,46,47] CDS-assisted colonoscopy was found to increase the adenoma finding rate leading to better clinical outcomes. In another prospective study, the real-world performance of a colonoscopy image recognition system was shown to be comparable to expert gastroenterologists.[35]

Another study evaluated a radiology double reading system whereby a CDS was used to detect discrepancies between CT scans and the interpretation reports provided by radiologists.[45] Here, the CDS combined ML algorithms with natural language processing to process lung CTs and provide interpretation reports. Out of the 104 potential pulmonary cancer nodules flagged by the system, decisions on 7 cases (7%) were subsequently corrected by re-issuing results about clinically significant nodule requiring follow-up care. Another CDS that assisted clinicians in magnetic resonance imaging segmentation for radiotherapy was shown to shorten segmentation time while maintaining the safety of organs-at-risk.[40]

Diagnostic assistance

Three CDS assisted clinician in diagnosis. A CDS that provided a probability score, based on web-based patient questionnaires, predicted differential diagnosis of various pediatric neuromuscular diseases.[31] An antibiotic selection CDS used blood and microbiological data to predict the probability of bacterial infection.[37] The CDS provided recommendations for prescription of antibiotics within 72 hours of patient admission. Another RCT of a CDS assisting emergency dispatchers in diagnosing out of hospital cardiac arrest demonstrated higher sensitivity when dispatchers were assisted by CDS (85%) compared to dispatchers alone (77%).[25] However, this study observed lower specificity in CDS-assisted dispatchers versus dispatchers alone (97.4% vs 99.6%, $P < .001$) and no difference in dispatchers' decision to recognize cardiac arrest.

Treatment planning

Two studies compared the use of ML-based CDS in treatment planning with conventional techniques. In the first, an ML algorithm that assisted with planning of prostate cancer brachytherapy was shown to shorten planning time and maintain safe dosimetry levels.[36] The second involved computed tomography-guided radiotherapy planning where CDS used reduced treatment time and the radiotherapy dose by 42%.[44]

Process support

One study examined a CDS that identified patients at high-risk for glycemic control and recommended socio-clinical interventions to assist clinicians in managing diabetic patients in a primary care setting.[39] Undesirably, clinicians reported that the system was unhelpful in identifying the right interventions (median score 11 out of 0-100 on helpfulness scale).

CPOE & Electronic prescribing

One study evaluated an ML-enabled order entry system designed to provide both synchronous and asynchronous alerts about prescription errors using outlier detection.[41] The system was shown to have high accuracy (85%) and was considered to be clinically useful with 43% of alerts resulting in the subsequent modification of orders. Clinicians changed prescriptions in response to system flags about bradycardia, elevated liver function tests, and hypotension.

Autonomous information: CDS provides information to make decisions

Only 2 studies examined ML-based CDS that contributed information for clinicians to make decisions. In the first, a CDS that provided an objective prediction of the patient's dry weight for hemodialysis prescription was shown to be effective in lowering or stopping antihypertensive treatments in 29% of cases compared to a subjective estimation by clinicians.[49] The second study examined use of a CDS that provided body mass index (BMI) prediction based on 190

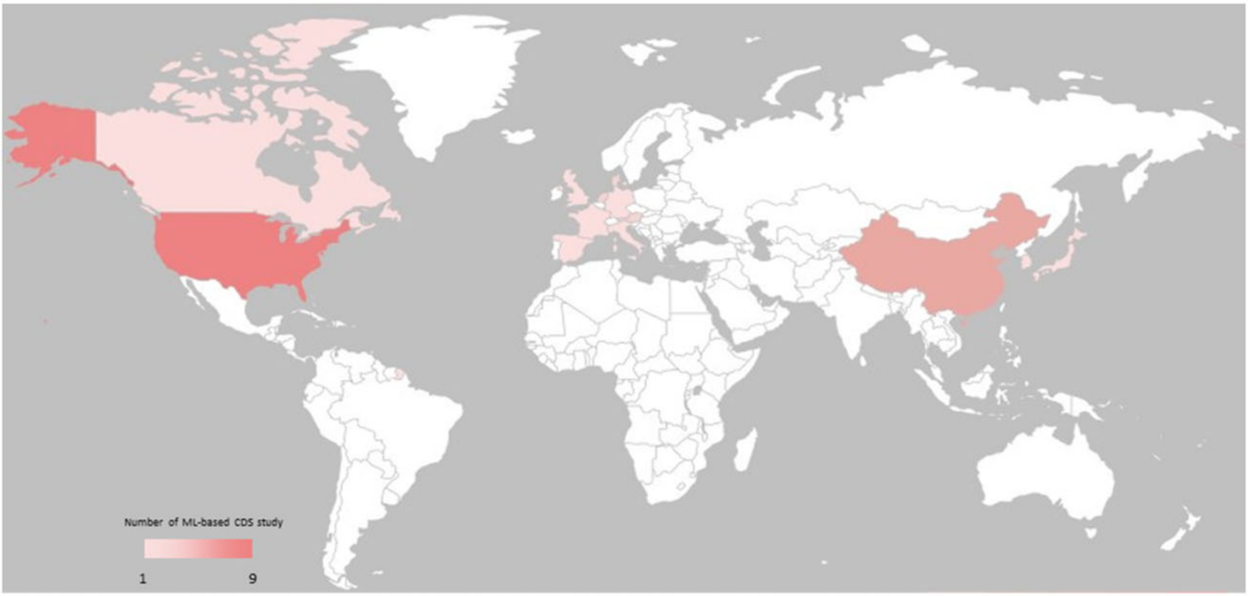

Figure 2. Geographical distribution of the included studies.

Table 2. Included studies by CDS task and level of system autonomy.[13]

	Level of autonomy	Assistive	Autonomous information	Autonomous decision	Total
CDS task	Image recognition & interpretation	8	–	4	12
	Risk assessment	8	–	1	9
	Diagnostic assistance	3	1	1	5
	Treatment planning	2	1	–	3
	Process support system	1	–	1	2
	CPOE & electronic prescribing	1	–	–	1
		23	2	7	32

Abbreviations: CDS, clinical decision support; CPOE, computerized provider order entry.

Table 3. Included studies by ML method and CDS task.

	CDS task	CPOE & e-prescribing	Diagnostic assistance	Treatment planning	Risk assessment	Process support	Image recognition	Total
ML method	Support vector machine	–	2	–	–	–	1	3
	Random forest	–	2	–	4	–	1	7
	Logistic regression	–	2	–	–	–	–	2
	Gradient boosting	–	1	–	2	–	–	3
	Convolutional neural network	–	–	1	–	–	6	7
	Artificial neural network	–	2	–	1	–	–	3
	Unspecified deep learning	–	1	–	–	–	2	3

Abbreviations: CDS, clinical decision support; CPOE, computerized provider order entry.

variables related to genetic, social, diet, and other risk factors, identifying 4 important predictive variables for early intervention in childhood obesity and predicting disease risk, including nutri-status perception, energy expenditure, mother's BMI, and father's BMI.[48]

Autonomous decision: CDS decides in place of human

Seven studies examined ML-based CDS that provided decisions for clinical tasks that could be enacted by clinicians. Four of these involved image recognition and interpretation CDS. A CDS which processed clinical dermatology photographs to distinguish between fungal onychomycosis and noninfectious onychodystrophy was found to reduce prescriptions for unnecessary antifungal medications.[52] Another study involving an RCT with 300 patients examined the diagnosis of childhood cataract based on anterior ocular images.[53] Here, the CDS provided recommendations for surgery or conservative follow-up based on the diagnosis. Although real-world performance was degraded compared to expert clinicians, patients managed in the CDS group reported faster service (2.79 vs 8.53 minutes, $P < .001$) and better

Table 4. Studies reporting evaluation of ML-based CDS in clinical settings grouped by the level of autonomy ($n = 32$).

Author Country	Clinical task	Care setting; study design	CDS users	CDS task[1]	Stage of human information processing[23]	Effects on decision-making, care delivery and patient outcomes, or other key findings
Assistive: CDS assists human decisions ($n = 23$)						
Blomberg et al. (2021)[25] Denmark	Identifying patients suspected of outside of hospital cardiac arrest in emergency phone calls	Secondary-tertiary care; experimental randomized controlled trial	Emergency dispatchers, Emergency physicians	Diagnostic assistance	Decision selection	*Decision*: No significant difference in dispatchers' decision to recognize cardiac arrest. Performance of CDS-assisted dispatchers versus dispatchers alone: sensitivity (85% vs 78%, $P < .001$) and specificity (97% vs 100%)
Brennan et al. (2019)[26] United States	Risk assessment for 6 postoperative complications	Secondary-tertiary care; experimental non-randomized trial	Surgeons, anesthesiologists	Risk assessment	Decision selection	*Decision*: Physicians changed their risk-assessment score in more than 75% cases ($n = 150$)
Burdick et al. (2020)[27] United States	Predicting sepsis in hospitalized patients and providing alerts to clinician	Secondary-tertiary care; observational prospective cohort	Doctors/Critical Care	Risk assessment	Decision selection	*Outcomes*: Decreased hospital mortality by 39.5%, hospital length of stay by 32.3%, and 30-day readmission rate for sepsis-related patients by 22.7%
Giannini et al. (2019)[28] United States	Predicting sepsis in hospitalized patients and providing alerts to clinician	Secondary-tertiary care; observational prospective cohort	Doctors/Critical Care	Risk assessment	Information analysis	*Care*: Slight increase in lactate testing, administration of IV fluid boluses, and blood test (complete blood count and metabolic panel) *Outcomes*: Shortening time to ICU transfer. No significant effect on hospital length of stay and mortality
Ginestra et al. (2019)[29] United States	Predicting sepsis in hospitalized patients and providing alerts to clinician	Secondary-tertiary care; observational prospective cohort	Nurses and doctors	Risk assessment	Decision selection	*Decision*: Majority of clinicians (55% nurses and 62% providers) reported no change in perception of the patient's risk for sepsis *Care*: Some clinicians (30% nurses and 9% providers) reported change of care upon alert from CDS
Gong et al. (2020)[30] China	Detecting colonoscopy insertion-withdrawal time and endoscope slipping to aid clinician distinguish between adenomatous colorectal cancer and benign polyps	Secondary-tertiary care; experimental randomized controlled trial	Gastroenterologist/ Endoscopist	Image recognition & interpretation	Information analysis	*Outcomes*: Increased adenoma detection rate over the control group (odds ratio 2.30, 95% CI 1.40–3.77; $P = .0010$)
Grigull et al. (2016)[31] Germany	Diagnosing neuromuscular diseases (7 diagnosis) based on questionnaire responses	Secondary-tertiary care; observational prospective cohort	Primary care physicians	Diagnostic assistance	Decision selection	Real world accuracy reached 89% in diagnosing different neuromuscular diseases

(continued)

Table 4. (continued)

Author Country	Clinical task	Care setting; study design	CDS users	CDS task[1]	Stage of human information processing[23]	Effects on decision-making, care delivery and patient outcomes, or other key findings
Jauk et al. (2021)[32] Austria	Predicting the risk of delirium among hospitalized patients	Secondary-tertiary care; mixed-method	Physicians and nurses	Risk assessment	Decision selection	*Decision*: Majority of clinician agreed that the CDS provided additional information (68%), but only 19% considered the output of CDS in clinical decisions and 33% used the CDS regularly
Jauk et al. (2020)[33] Austria	Predicting the risk of delirium among hospitalized patients	Secondary-tertiary care; observational prospective cohort	Physicians and nurses	Risk assessment	Decision selection	CDS performed with sensitivity of 74% and specificity of 82% *Care*: High-risk patient obtained nonpharmacological preventive treatment (eg, reinforcement of visual and hearing aids, hydration, sleep time management, clear communication).
Liu et al. (2020)[34] China	Detecting polyps in colonoscopy and distinguishing between adenomatous colorectal cancer and benign polyps	Secondary-tertiary care; experimental randomized controlled trial	Gastroenterologists/ Endoscopists	Treatment planning & critiquing	Decision selection	*Outcomes*: Increase in adenoma detection rate versus the control group (39% vs 24%, $P < .001$)
Mori et al. (2020)[35] Japan	Detecting polyps in colonoscopy and distinguishing between adenomatous colorectal cancer and benign polyps	Secondary-tertiary care; observational prospective cohort	Gastroenterologists/ Endoscopist	Image recognition & interpretation	Decision selection	In clinical setting, CAD performed as good as expert gastroenterologist in distinguishing between adenomatous colorectal cancer and benign polyps. NPV = 96%
Nicolae et al. (2020)[36] Canada	Planning treatment for low-dose-rate brachytherapy to be input into a Treatment Planning System (TPS)	Secondary-tertiary care; experimental randomized controlled trial	Urologists	Image recognition & interpretation	Information analysis	*Care*: Shorter planning mean planning time for the PIPA arm (2.38 + 0.96 minutes) compared with the conventional (43.13 + 58.70 minutes) $P \gg .05$ *Outcomes*: Excellent safety by the CDS with no significant differences was observed in preimplant or day 30 dosimetry
Rawson et al. (2018)[37] United Kingdom	Diagnosing a community-acquired bacterial infection within 72 hours of admission, leading to decision of antibiotic prescription	Secondary-tertiary care; observational prospective cohort	Physicians	Diagnostic assistance	Decision selection	*Care*: Ordering microbiological laboratory tests for individuals with high possibility of bacterial infection prior to antibiotic prescription
Repici et al. (2020)[38] Italy	Detecting polyps in colonoscopy and distinguishing between adenomatous colorectal cancer and benign polyps	Primary care; experimental randomized controlled trial	Gastroenterologists/ Endoscopists	Process support system	Information analysis	*Outcomes*: Increase in adenoma detection rate over the control group (55% vs 40%, $P < .001$)

(continued)

Table 4. (continued)

Author Country	Clinical task	Care setting; study design	CDS users	CDS task[1]	Stage of human information processing[23]	Effects on decision-making, care delivery and patient outcomes, or other key findings
Romero-Brufau et al. (2020)[39] United States	Predicting high glycemic risk in patient with diabetes and providing recommendation	Primary care; observational cross-section	Primary care physicians, registered nurses, licensed practical nurses, social workers	CPOE & electronic prescribing	Information analysis	*Care*: Clinician felt that care was better coordinated ($P < .001$)
Savenije et al. (2020)[40] Netherlands	Segmenting organs-at-risk for prostate radiotherapy while manual segmentation is time consuming	Secondary-tertiary care; observational prospective cohort	Radiotherapists	Image recognition & interpretation	Information analysis	Faster inference time by CDS compared to conventional methods (14 minutes vs 60 seconds) while maintaining good performance by dice similarity coefficient (0.92-0.98 for various organs)
Segal et al. (2019)[41] Israel	Preventing prescription error based on irregularities	Secondary-tertiary care; observational prospective cohort	Physicians	Risk assessment	Decision selection	*Decision*: 43% of the alerts caused changes in subsequent medical orders, 39% of the erroneous medication orders were modified during the order of medication (synchronous flags), and 61% were modified during monitoring phase (asynchronous flags) following a change in clinical indicators
Sendak et al. (2020)[42] United States	Predicting sepsis in hospitalized patients and providing alerts to clinicians	Secondary-tertiary care; observational prospective cohort	Intensivists, emergency department clinicians, rapid response team, nurses	Risk assessment	Decision selection	Learning health system framework was used to integrate system to routine care
Shimabukuro et al. (2017)[43] United States	Predicting sepsis in hospitalized patients and providing alerts to clinicians	Secondary-tertiary care; experimental randomized controlled trial	Critical care doctors	Risk assessment	Information analysis	*Outcomes*: Decreased length of stay from 13.0 to 10.3 days ($P = .042$). Decreased in-hospital mortality by 12% ($P = .018$)
Sibolt et al. (2021)[44] Denmark	Generating pre-treatment plans for adaptive guided radiotherapy in bladder cancer	Secondary-tertiary care; observational prospective cohort	Radiotherapists	Treatment planning & critiquing	Decision selection	*Care*: Patients obtained faster treatment adaptive CDS with median duration of 17.6 minutes from preparation to treatment delivery *Outcomes*: Adaptive CDS reduced high-dose planning target volume in bladder by 42% compared to the scheduled conventional technique
Tan et al. (2021)[45] Singapore	Recognizing pulmonary nodules in CT scan as a double read safety system	Secondary-tertiary care; observational prospective cohort	Radiologists, patient safety officers	Image recognition & interpretation	Information analysis	*Decision*: Seven cases of 104 flagged images were deemed clinically significant and clinicians were informed to change subsequent management

(continued)

Table 4. (continued)

Author Country	Clinical task	Care setting; study design	CDS users	CDS task[1]	Stage of human information processing[23]	Effects on decision-making, care delivery and patient outcomes, or other key findings
Wang et al. (2019)[46] China	Detecting polyps in colonoscopy and distinguishing between adenomatous colorectal cancer and benign polyps	Secondary-tertiary care; experimental randomized control trial	Gastroenterologists/ Endoscopists	Image recognition & interpretation	Information analysis	*Outcomes*: Increased adenoma detection rate versus the control group (29% vs 20%, $P<.001$)
Wang et al. (2020)[47] China	Detecting polyps in colonoscopy and distinguishing between adenomatous colorectal cancer and benign polyps	Secondary-tertiary care; observational perspective cohort	Gastroenterologists/ Endoscopists	Image recognition & interpretation	Information analysis	*Outcomes*: Decreased in adenoma miss rate versus the control group (14% vs 40%, $P<.0001$)
Autonomous information: CDS provides information to make decisions ($n=2$)						
Marcos-Pasero et al. (2021)[48] Spain	Calculating predicted body mass index (BMI) to predict childhood obesity and give early intervention	Community; observational cross-section	Clinical nutritionists	Diagnostic assistance	Information analysis	This study collected information by self-questionnaire. Five top predictors were nutrition perception, difference between energy intake and expenditure, father's BMI, mother's BMI, and mother's meals
Niel et al. (2018)[49] France	Estimating dry weight in hemodialysis patient to avoid side effects of hypertension due to underestimation or overestimation of dry weight	Secondary-tertiary care; experimental cross-over trial	Nephrologists	Treatment planning & critiquing	Information analysis	*Care*: discontinuation or decrease in antihypertensive prescription. *Outcomes*: significant decrease in systolic blood pressure or better hypertension control.
Autonomous decision: CDS decides in place of human ($n=7$)						
Chen et al. (2020)[50] United States	Deciding whether to accept or ignore alerts and following up with shingles vaccination	Secondary-tertiary care; observational prospective cohort	Physicians	Process support system	Decision selection	*Care*: Weekly counts of shingles vaccination remained stable after activation of suppression system versus control group (326.3 vs 331.3, $P=.38$).
Isma'eel et al. (2017)[51] Lebanon	Predicting the presence of cardiac ischemia	Secondary-tertiary care; observational prospective cohort	Cardiologists, primary care physicians	Risk assessment	Decision selection	*Decision*: Change physician's decision to perform downstream stress testing. CDS performed better than conventional risk model with higher discriminatory power 1.61, NPV 98%, sensitivity 91%
Kim et al. (2020)[52] South Korea	Diagnosing onychomycosis by clinical photograph, leading to decision of antifungal prescription	Secondary-tertiary care; observational prospective cohort	Dermatologists, other physicians (nondermatologist)	Image recognition & interpretation	Decision selection	*Decision*: Change physician's prescription of antifungal medication (82%)

(continued)

Table 4. (continued)

Author Country	Clinical task	Care setting; study design	CDS users	CDS task[1]	Stage of human information processing[23]	Effects on decision-making, care delivery and patient outcomes, or other key findings
Lin et al. (2019)[53] China	Diagnosing childhood cataract, providing comprehensive evaluation of the disease, and recommending option of surgery	Secondary-tertiary care; experimental randomized controlled trial	Ophthalmologists, primary care physicians	Image recognition & interpretation	Decision selection	Faster time in receiving diagnosis by CDS (2.79 minutes) compared to experts (8.52 minutes), $P < .001$ CDS had inferior accuracies compared to experts, both in determining cataract diagnosis (87%) and treatment recommendation (71%)
Wintjens et al. (2020)[54] Netherlands	Diagnosing COVID-19 by recognizing the pattern of volatile organic compound from a breath analyzer to exclude infected patients for elected surgery	Secondary-tertiary care; observational prospective cohort	Physicians	Diagnostic assistance	Decision selection	The CDS demonstrated real world sensitivity of 86% and NPV of 92% to be used as triage tool for patients undergo elected surgery
Xiao (2021)[55] China	Diagnosing hepatobiliary diseases using ocular images	Secondary-tertiary care; observational prospective cohort	Ophthalmologists, hepatobiliary surgeons	Image recognition & interpretation	Decision selection	The ROC were 0.93 (0.91-0.94) for slit lamp and 0.68 for fundus images
Yao et al. (2021)[56] United States	Identifying patients with ventricular dysfunction to recommend for further supporting examination	Primary care; experimental randomized controlled trial	Primary care physicians	Image recognition & interpretation	Decision selection	*Outcomes*: Increase diagnosis of low ejection fraction within 90 days of the ECG (2.1% in intervention arm vs 1.6% in control group OR 1.32 $P = .007$)

Abbreviations: CDS: clinical decision support; CPOE: computerized provider order entry; CT: computed tomography; ECG: electrocardiogram; NPV: negative predictive value; ROC: receiving operating characteristic.

satisfaction. The third study demonstrated real-world performance of a CDS to screen patients for 7 hepatobiliary diseases based on changes in eye appearance and color using slit lamp and fundus images (AUC = 0.74).[55] The fourth study involved a CDS that identified ventricular dysfunction from analysis of electrocardiogram (ECG) images and was found to maintain referrals rate for echocardiography (18% control vs 19% CDS intervention, $P = .17$), while increasing the case finding at the same time (odds ratio = 1.32; $P = .007$).[56]

The fifth study evaluated a CDS that automated referrals for costly stress test and noninvasive imaging based on the risk of cardiac ischemia from clinical data.[51] The CDS was shown to perform better than a conventional risk model with potential to reduce 59% of unnecessary tests ($n = 486$). The sixth study related to a CDS that automated the diagnosis of COVID-19 to screen patients for elective surgery.[57] The system utilized biosensors for breath analysis and was demonstrated to have a real-world sensitivity of 0.86 and negative predictive value of 0.92. The final study examined CDS to suppress irrelevant alerts from another CDS about shingles vaccination.[50] The CDS deployment was able to reduce 44% of inappropriate alerts and maintained similar vaccination counts.

Effects of ML-based CDS on decision-making, care delivery, and clinical outcomes

Only 8 studies (25%) examined the effects of ML-based CDS on decision-making in healthcare settings reporting mixed results. Of these, 5 reported benefits and improvements in a variety of decisions including surgery risk assessment,[26] new clinically significant CT scan interpretation,[45] reducing further examinations,[51] and better prescription decisions.[41,52] In the other 2 studies, CDS implementation did not influence diagnostics of cardiac arrest[25] as well as predicting risk of sepsis[29] and delirium.[32]

The effects of ML-based CDS on care delivery were examined in 10 studies (31%). Patients predicted to be at high risk of hospital complications were given more aggressive treatments. Patients with high-risk bacterial infection and sepsis prediction were ordered more blood tests and administered more IV fluid bolus.[28,37] Nonpharmacological preventive treatments, such as hydration and sleep management, were given to high-risk delirium patients.[32,33] Use of CDS in treatment planning was also shown to reduce time and improve care coordination.[36,39,44] However, in 1 study that evaluated a CDS to predict sepsis, no changes were observed in the 12 out of 15 care process measures assessed.[29]

Discussion

Despite rapid growth in the development of ML-based CDS, few studies have evaluated these CDS in clinical settings to examine their effects on decision-making, care delivery, or patient outcomes. We found that most CDS evaluated in clinical settings were assistive ($n = 23$), requiring clinicians to confirm or approve CDS provided information or decisions, and where the responsibility for the final decision generally rests with the clinician. The most common use of assistive CDS was in risk assessment for sepsis[27–29,42,43] and for interpreting cancerous lesions in colonoscopy.[30,34,35,38,46,47] While ML-based CDS are being applied in many clinical areas, further observational studies are required to understand how they are used by clinicians as well as their effects on decision-making, care delivery, and patient outcomes.

Clinical decisions assisted by CDS

The level of autonomy determines how clinicians interact with and use CDS. Our findings are consistent with the findings of previous reviews of ML-based medical devices[13] and nursing CDS[58] which found that most contemporary systems were assistive.

Only 1 in 4 studies (26%) examined effects on decision-making which were mixed. While 3 studies demonstrated improvements with the use of assistive CDS,[26,41,45] 3 others found no improvement.[25,29,32] Although such observational studies provide an opportunity to examine effects in real-world settings, controlled experiments which enable patient- and risk-free evaluation are useful to better understand specific effects on decision-making.[59]

Clinicians typically integrate many different sources of information including CDS advice to make decisions.[60] In this context, assistive CDS may often require clinicians to consider "black-box" CDS advice against their own decisions. For instance, when predicting the risk of surgical complications, to validate CDS advice clinicians must independently assess risk based on history taking, physical examination, laboratory results, and other supporting examinations on top of using the CDS.[26] Such a double up in the clinical workflow may require more time and delay decisions, especially if the clinician's decision does not agree with the CDS, potentially increasing risks to patient safety.[61]

Still, early demonstration of benefits supports the use of assistive CDS in clinical settings. Of the 23 studies we examined, 5 demonstrated improvements in care delivery such as increasing preventive orders in high-risk patients,[28,37] shorter time to treatment,[36,44] and improved care coordination.[39] Six studies showed enhanced patient outcomes including higher case finding,[30,34,38,46] shorter hospital length of stay, and lower mortality.[27,43] Such benefits strengthen the case for using these systems and are likely to increase implementation.

Tasks supported by ML-based CDS

Different to rule-based CDS that are mostly focused on supporting prescribing and care reminders,[7,8] we found that the most common task supported by ML-based CDS was in image recognition and interpretation followed by risk assessment. This shift in clinical tasks supported by CDS could be attributed to the wide availability of deep learning ML methods such as convolutional neural networks[34,38,40,52,53,56] as well as the large amounts of standardized data readily available in imaging and electronic health record data in risk assessment. For instance, a CDS for sepsis risk assessment was trained with data from 684 443 encounters over 5 years from across 6 institutions.[27,62] While rule-based CDS have been shown to improve care delivery (57%) and patient outcomes (30%) especially in drug ordering and preventive care,[7] the effects of ML-based CDS were mixed.

Evaluating ML-based CDS in clinical settings

Most studies in this review come with high levels of heterogeneity in outcome measures making comparison difficult. We observed many different study designs ranging from RCTs to qualitative interviews, using a wide variety of outcome measures to examine effects on decision-making, care delivery, and patient outcomes. The reporting of evaluation studies is likely to improve with recent publication of reporting standards, such as DECIDE-AI for early-stage clinical evaluation[63] and CONSORT/SPIRIT-AI for larger clinical trials.[64,65]

Only a few of the studies we reviewed examined effects on decision-making (25%), care delivery (31%), and clinical outcomes (38%). To that end we have demonstrated the use of 2 frameworks. The first is the *level of autonomy*, a 3-level classification based on how clinical tasks are divided between the clinician and CDS that allows examination of the specific role of the CDS in clinical tasks.[13] The second is an established *information value chain* framework, which separates the multiple steps from system use to impacting clinical outcomes—interacting with CDS, receiving new information, decision-making, and care delivery.[9] Were these to be used as standard templates in future studies, it would be possible to make comparative assessments between studies.

There is also a need to examine the actual use of CDS by clinicians in the real world. None of the studies examined patterns of use, and only 10 (31%) reported the number of clinicians involved in evaluations. The number of clinicians, their expertise, and experience with CDS are important variables affecting adoption and use.[63] To that end, mixed-method studies are particularly valuable to measure actual use and to understand factors affecting acceptance of ML systems.[32]

ML-based CDS in resource-constrained settings

Most studies were undertaken at secondary and tertiary settings (91%) in the developed world (100%). Despite the potential for AI to improve health services in resource-constrained settings,[66–68] our review has identified a gap in evaluating ML-based CDS in such settings.[69,70] Resource-constrained settings are characterized by limited medical expertise and infrastructure leading to the suboptimal delivery of services.[71] Such conditions are not bound by the economic conditions as resource limitations are also experienced by developed countries, such as in rural or regional areas with limited access to expertise and modest IT infrastructure. In addition, there may be a greater role for AI to provide specific clinical expertise in primary care with the shift toward promotive and preventive care which is increasing the demand for primary health services.

These facts, combined with the main finding of this review that most ML-based CDS were assistive and still required experts to oversee decision-making, make it necessary to examine the appropriateness of ML-based CDS in resource-constrained settings as these systems may increase the burden on clinicians and even increase risks. Accordingly, we argue that autonomous CDS may be more suitable for resource-constrained settings. For instance, 3 studies in this review proposed potential use of their autonomous CDS to screen hepatobiliary disease,[55] childhood cataract,[53] and onychomycosis[52] in remote or poorly serviced areas.

Limitations

There are several limitations. First, this review is limited to the published research literature about ML-based CDS in clinical settings. We did not include gray literature, such as white papers and reports. Second, our analysis of the level of CDS autonomy was limited to the CDS information that was reported in the papers and prior validation studies. This information was less structured than the indications of use in medical device approval documents which informed the development of the level of autonomy.[13] Finally, there was considerable heterogeneity in the study designs and outcome measures which prevented quantitative examination of the effects on decision-making, care delivery, and patient outcomes.

Conclusion

ML-based CDS are being applied in a variety of clinical areas, but evaluation of their effects on decision-making, care delivery, and patient outcomes is limited. There remain opportunities to evaluate the feasibility of using ML-based CDS in clinical settings, especially in resource-constrained contexts, to support clinical decisions where there is a lack of specialist expertise and sophisticated medical equipment.

Acknowledgments

We wish to acknowledge the invaluable contributions of Ying Wang, who is providing technical advice about ML methods and their applications.

Author contributions

A.P.S. conceived this study, designed, and conducted the analysis with advice and input from F.M. and D.L. A.P.S. and F.M. selected the studies. A.P.S. and D.L. assessed the stage of automation and level of autonomy. A.P.S. drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have approved the final manuscript.

Supplementary material

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

Funding

This study was supported by a Macquarie University doctoral scholarship awarded to A.P.S. (iMQRES 20201869) and the NHMRC Centre for Research Excellence (CRE) in Digital Health (APP1134919). The funding sources did not play any role in the study design, in the collection and analysis, and interpretation of data, or in the writing of the report.

Conflicts of interest

None declared.

Data availability

All data relevant to the analysis are included in the article and online supplementary material. There are no new data associated with this article.

References

1. Coiera E. Clinical decision support system. In: Coiera E, ed. *Guide to Health Informatics*. 3rd ed. Taylor & Francis Group; 2015:417.
2. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17.

3. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2(10):719-731.
4. Coiera E. On algorithms, machines, and medicine. *Lancet Oncol*. 2019;20(2):166-167.
5. Vasey B, Ursprung S, Beddoe B, et al. Association of clinician diagnostic performance with machine learning-based decision support systems: a systematic review. *JAMA Netw*. 2021;4(3):e211276.
6. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb Med Inform*. 2019;28(1):128-134.
7. Jaspers MW, Smeulers M, Vermeulen H, et al. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011;18(3):327-334.
8. Kwan JL, Lo L, Ferguson J, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *BMJ*. 2020;370:m3216.
9. Li JW, Wang LM, Ang TL. Artificial intelligence-assisted colonoscopy: a narrative review of current data and clinical applications. *Singapore Med J*. 2022;63(3):118-124.
10. Yeo M, Kok HK, Kutaiba N, et al. Artificial intelligence in clinical decision support and outcome prediction - applications in stroke. *J Med Imaging Radiat Oncol*. 2021;65:518-528.
11. Wu M, Du X, Gu R, et al. Artificial intelligence for clinical decision support in sepsis. *Front Med (Lausanne)*. 2021;8:665464.
12. Coiera E. The last mile: where artificial intelligence meets reality. *J Med Internet Res*. 2019;21(11):e16323.
13. Lyell D, Coiera E, Chen J, et al. How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices. *BMJ Health Care Inform*. 2021;28(1):e100301.
14. Coiera E. Assessing technology success and failure using information value chain theory. *Stud Health Technol Inform*. 2019;263:35-48.
15. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. *J Med Internet Res*. 2021;23(4):e25759.
16. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. *J Med Internet Res*. 2019;21(4):e12286.
17. Knop M, Weber S, Mueller M, et al. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence-enabled clinical decision support systems: literature review. *JMIR Hum Factors*. 2022;9(1):e28639.
18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19-32.
19. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5(69):1-9.
20. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-473.
21. Peters MD, Godfrey CM, Soares PMCB, et al. Joanna Briggs Institute Reviewers' Manual. 2015 edition/supplement. Adelaide: The Joanna Briggs Institute; 2015.
22. The World Bank. World Bank Country and Lending Groups. 2021. Accessed February 15, 2021. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
23. Parasuraman R, Sheridan T, Wickens C. A model for types and levels of human interaction with automation. *IEEE Trans Syst Man Cybern B*. 2000;30(3):286-297.
24. Coiera E. A new informatics geography. *Yearb Med Inform*. 2016;(1):251-255.
25. Blomberg SN, Christensen HC, Lippert F, et al. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical trial. *JAMA Netw Open*. 2021;4(1):e2032320.
26. Brennan M, Puri S, Ozrazgat-Baslanti T, et al. Comparing clinical judgment with the MySurgeryRisk algorithm for preoperative risk assessment: a pilot usability study. *Surgery (United States)*. 2019;165(5):1035-1045.
27. Burdick H, Pino E, Gabel-Comeau D, et al. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. *BMJ Health Care Inform*. 2020;27(1):e100109.
28. Giannini HM, Ginestra JC, Chivers C, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med*. 2019;47(11):1485-1492.
29. Ginestra JC, Giannini HM, Schweickert WD, et al. Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. *Crit Care Med*. 2019;47(11):1477-1484.
30. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol*. 2020;5(4):352-361.
31. Grigull L, Lechner W, Petri S, et al. Diagnostic support for selected neuromuscular diseases using answer-pattern recognition and data mining techniques: a proof of concept multicenter prospective trial. *BMC Med Inform Decis Mak*. 2016;16:31.
32. Jauk S, Kramer D, Avian A, et al. Technology acceptance of a machine learning algorithm predicting delirium in a clinical setting: a mixed-methods study. *J Med Syst*. 2021;45(4):48.
33. Jauk S, Kramer D, Großauer B, et al. Risk prediction of delirium in hospitalized patients using machine learning: an implementation and prospective evaluation study. *J Am Med Inform Assoc*. 2020;27(9):1383-1392.
34. Liu WN, Zhang YY, Bian XQ, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol*. 2020;26(1):13-19.
35. Mori Y, Kudo SE, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy a prospective study. *Ann Intern Med*. 2018;169(6):357-366.
36. Nicolae A, Semple M, Lu L, et al. Conventional vs machine learning-based treatment planning in prostate brachytherapy: results of a Phase I randomized controlled trial. *Brachytherapy*. 2020;19(4):470-476.
37. Rawson TM, Hernandez B, Moore LSP, et al. Supervised machine learning for the prediction of infection on admission to hospital: a prospective observational cohort study. *J Antimicrob Chemother*. 2019;74(4):1108-1115.
38. Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology*. 2020;159(2):512-520.e7.
39. Romero-Brufau S, Wyatt KD, Boyum P, et al. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int J Med Inform*. 2020;137:104072.
40. Savenije MHF, Maspero M, Sikkes GG, et al. Clinical implementation of MRI-based organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy. *Radiat Oncol*. 2020;15(1):104.
41. Segal G, Segev A, Brom A, et al. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. *J Am Med Inform Assoc*. 2019;26(12):1560-1565.
42. Sendak MP, Ratliff W, Sarro D, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med Inform*. 2020;8(7):e15182.
43. Shimabukuro DW, Barton CW, Feldman MD, et al. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res*. 2017;4(1):e000234.
44. Sibolt P, Andersson LM, Calmels L, et al. Clinical implementation of artificial intelligence-driven cone-beam computed tomography-

guided online adaptive radiotherapy in the pelvic region. *Phys Imaging Radiat Oncol*. 2021;17:1-7.
45. Tan JR, Cheong EHT, Chan LP, et al. Implementation of an artificial intelligence-based double read system in capturing pulmonary nodule discrepancy in CT studies. *Curr Probl Diagn Radiol*. 2021;50(2):119-122.
46. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68(10):1813-1819.
47. Wang P, Liu P, Glissen Brown JR, et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. *Gastroenterology*. 2020;159(4):1252-1261.e5.
48. Marcos-Pasero H, Colmenarejo G, Aguilar-Aguilar E, et al. Ranking of a wide multidomain set of predictor variables of children obesity by machine learning variable importance techniques. *Sci Rep*. 2021;11(1):1910.
49. Niel O, Bastard P, Boussard C, et al. Artificial intelligence outperforms experienced nephrologists to assess dry weight in pediatric patients on chronic hemodialysis. *Pediatr Nephrol*. 2018;33(10):1799-1803.
50. Chen J, Chokshi S, Hegde R, et al. Development, implementation, and evaluation of a personalized machine learning algorithm for clinical decision support: case study with shingles vaccination. *J Med Internet Res*. 2020;22(4):e16848.
51. Isma'eel H, Sakr G, Serhan M, et al. Artificial neural network-based model enhances risk stratification and reduces non-invasive cardiac stress imaging compared to Diamond–Forrester and Morise risk assessment models: a prospective study. *J Nucl Cardiol*. 2018;25(5):1601-1609.
52. Kim YJ, Han SS, Yang HJ, et al. Prospective, comparative evaluation of a deep neural network and dermoscopy in the diagnosis of onychomycosis. *PLoS One*. 2020;15(6):e0234334.
53. Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine*. 2019;9:52-59.
54. Wintjens A, Hintzen KFH, Engelen SME, et al. Applying the electronic nose for pre-operative SARS-CoV-2 screening. *Surg Endosc*. 2021;35(12):6671-6678.
55. Xiao W, Huang X, Wang JH, et al. Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study. *Lancet Digit Health*. 2021;3(2):e88-e97.
56. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*. 2021;27(5):815-819.
57. Abedin Z, Hoerner R, Habboushe J, et al. Implementation of a fast healthcare interoperability resources-based clinical decision support tool for calculating CHA(2)DS(2)-VASc scores. *Circ Cardiovasc Qual Outcomes*. 2020;13(2):e006286.
58. Akbar S, Lyell D, Magrabi F. Automation in nursing decision support systems: a systematic review of effects on decision making, care delivery, and patient outcomes. *J Am Med Inform Assoc*. 2021;28(11):2502-2513.
59. Carayon P, Hoonakker P, Hundt AS, et al. Application of human factors to improve usability of clinical decision support for diagnostic decision-making: a scenario-based simulation study. *BMJ Qual Saf*. 2020;29(4):329-340.
60. van Baalen S, Boon M, Verhoef P. From clinical decision support to clinical reasoning support systems. *J Eval Clin Pract*. 2021;27(3):520-528.
61. Lyell D, Magrabi F, Raban MZ, et al. Automation bias in electronic prescribing. *BMC Med Inform Decis Mak*. 2017;17(1):28.
62. 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.
63. Vasey B, Nagendran M, Campbell B, et al.; DECIDE-AI Expert Group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. 2022;377:e070904.
64. Cruz Rivera S, Liu X, Chan AW, et al.; SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351-1363.
65. Liu X, Cruz Rivera S, Moher D, et al.; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374.
66. Alami H, Rivard L, Lehoux P, et al. Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Glob Health*. 2020;16(1):52.
67. Wahl B, Cossy-Gantner A, Germann S, et al. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health*. 2018;3(4):e000798.
68. Kiyasseh D, Zhu T, Clifton D. The promise of clinical decision support systems targetting low-resource settings. *IEEE Rev Biomed Eng*. 2020;15:354-371.
69. Fraser HSF, Zahiri K, Kim N, et al. The Global Health Informatics landscape and JAMIA. *J Am Med Inform Assoc*. 2023;30(4):775-780.
70. Ciecierski-Holmes T, Singh R, Axt M, et al. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *NPJ Digit Med*. 2022;5(1):162.
71. van Zyl C, Badenhorst M, Hanekom S, et al. Unravelling 'low-resource settings': a systematic scoping review with qualitative content analysis. *BMJ Glob Health*. 2021;6(6):e005190.