Healthcare Professionals' Experience of Using Artificial Intelligence: a Systematic Review with Narrative Synthesis

Abimbola Ayorinde, Daniel Opoku Mensah, Julia Walsh, Iman Ghosh, Siti Aishah Ibrahim, Jeffry Hogg, Niels Peek, Frances Griffiths

Submitted to: Journal of Medical Internet Research
on: December 22, 2023

Table of Contents

Healthcare Professionals' Experience of Using Artificial Intelligence: a Systematic Review with Narrative Synthesis

Abimbola Ayorinde[1] BSc, MSc, PhD; Daniel Opoku Mensah[1] PhD; Julia Walsh[1] PhD; Iman Ghosh[1] MSc; Siti Aishah Ibrahim[1] MSc; Jeffry Hogg[2, 3]; Niels Peek[4] PhD; Frances Griffiths[1] PhD

[1]Warwick Medical School University of Warwick Coventry GB
[2]Population Health Science Institute, Faculty of Medical Sciences Newcastle University Newcastle upon Tyne GB
[3]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester Manchester GB

Corresponding Author:
Abimbola Ayorinde BSc, MSc, PhD
Warwick Medical School
University of Warwick
Division of Health Sciences
Warwick Medical School
Coventry
GB

Abstract

Background: There has been significant increase in the development of Artificial intelligence (AI) for clinical decision support. Historically these were mostly knowledge-based systems, but recent advances include non-knowledge-based systems using some form of machine learning. The ability of healthcare professionals to trust technology and understand how it benefits patients or improves care delivery is known to be important for their adoption of that technology. For non-knowledge-based AI for clinical decision support, these issues are poorly understood.

Objective: To qualitatively synthesise evidence on the experiences of healthcare professionals in routinely using non-knowledge-based AI to support their clinical decision-making.

Methods: In June 2023 we searched four electronic databases: MEDLINE, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINHAL) and Web of Science with no language or date limit. We also contacted relevant experts and searched reference lists of included studies. We included studies of any design which reported the experiences of healthcare professionals using non-knowledge-based systems for clinical decision support in their work settings. We completed double independent quality assessment for all included studies using the Mixed Methods Appraisal Tool (MMAT). We used a theoretically informed thematic approach to synthesise the findings.

Results: After screening 7,552 titles and 182 full-text articles, we included 25 studies conducted in nine different countries. Most of the included studies were qualitative (n=14) and the remaining were quantitative (n=7) and mixed methods studies (n=4). Overall, we identified seven themes: (i) Understanding of AI applications; (ii) Level of trust and confidence in AI tools; (iii) Judging the added value of AI; (iv) Data availability and limitations of AI; (v) Time and competing priorities; (vi) Concern about governance; (vii) Collaboration to facilitate the implementation and use of AI. The most frequently occurring of these are the first three themes. For example, many studies reported that healthcare professionals were concerned about not understanding the AI outputs or the rationale behind them. There were issues with confidence in the accuracy and recommendations by the AI applications. Some healthcare professionals believed that AI provided added value and improved decision-making, some reported that it only served as a confirmation of their clinical judgment, while others did not find it useful at all.

Conclusions: Our review identified several important issues documented in various studies on healthcare professionals' use of AI in real-world healthcare settings. Opinions of healthcare professionals regarding the added value of AI for supporting clinical decision making varied widely, and many professionals have concerns about their understanding of, and trust in this technology. The findings of this review emphasise the need for concerted efforts to optimise the integration of AI in real-world healthcare settings. Clinical Trial: PROSPERO (International Prospective Register of Systematic Reviews) CRD42022336359; from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022336359

(JMIR Preprints 22/12/2023:55766)

DOI: https://doi.org/10.2196/preprints.55766

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Healthcare Professionals' Experience of Using Artificial Intelligence: a Systematic Review with Narrative Synthesis

Abimbola Ayorinde[1], Daniel Mensah[1], Julia Walsh[1], Iman Ghosh[1], Aisha Ibrahim, Jeffry Hogg[2], Niels Peek[3], Frances Griffiths[1]

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, United Kingdom

[2] Population Health Science Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[3] Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, United Kingdom

Corresponding author: Abimbola Ayorinde, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, United Kingdom. CV4 7AL; Telephone: +44(0)2476151098; Email: a.ayorinde.1@warwick.ac.uk

Abstract

Background

There has been significant increase in the development of Artificial intelligence (AI) for clinical decision support. Historically these were mostly knowledge-based systems, but recent advances include non-knowledge-based systems using some form of machine learning. The ability of healthcare professionals to trust technology and understand how it benefits patients or improves care delivery is known to be important for their adoption of that technology. For non-knowledge-based AI for clinical decision support, these issues are poorly understood.

Objective

To qualitatively synthesise evidence on the experiences of healthcare professionals in routinely using non-knowledge-based AI to support their clinical decision-making.

Methods

In June 2023 we searched four electronic databases: MEDLINE, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINHAL) and Web of Science with no language or date limit. We also contacted relevant experts and searched reference lists of included studies. We included studies of any design which reported the experiences of healthcare professionals using non-knowledge-based systems for clinical decision support in their work settings. We completed double independent quality assessment for all included studies using the Mixed Methods Appraisal Tool (MMAT). We used a theoretically informed thematic approach to synthesise the findings.

Results

After screening 7,552 titles and 182 full-text articles, we included 25 studies conducted in nine different countries. Most of the included studies were qualitative (n=13) and the remaining were quantitative (n=9) and mixed methods studies (n=3). Overall, we identified seven themes: (i) Understanding of AI applications; (ii) Level of trust and confidence in AI tools; (iii) Judging the added value of AI; (iv) Data availability and limitations of AI; (v) Time and competing priorities; (vi) Concern about governance; (vii) Collaboration to facilitate the implementation and use of AI. The most frequently occurring of these are the first three themes. For example, many studies reported that healthcare professionals were concerned about not understanding the AI outputs or the rationale behind them. There were issues with confidence in the accuracy and recommendations by the AI applications. Some healthcare professionals believed that AI provided added value and improved decision-making, some reported that it only served as a confirmation of their clinical judgment, while others did not find it useful at all.

Conclusions

Our review identified several important issues documented in various studies on healthcare professionals' use of AI in real-world healthcare settings. Opinions of healthcare professionals regarding the added value of AI for supporting clinical decision making varied widely, and many professionals have concerns about their understanding of, and trust in this technology. The findings of this review emphasise the need for concerted efforts to optimise the integration of AI in real-world healthcare settings.

Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42022336359; from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022336359

Keywords: Artificial Intelligence; healthcare; clinical decision support; health professional experiences

Introduction

Artificial intelligence (AI) refers to the ability of machines/computer programs to replicate human intelligence. AI is widely believed to have significant potential in healthcare, ranging from the transformation of many aspects of patient care to streamlining administrative processes within and between various organizations involved in healthcare [1]. Specific potential benefits include the earlier detection of disease, improved patient safety, improved estimation of capacity needs, and the facilitation of personalised medicine [1, 2]. Globally, there has been significant interest in the use of

AI in healthcare settings [3]. For example, the UK government invested £250 million in a national laboratory to boost AI in the National Health Services in 2019 [2].

AI has various applications in healthcare, but one of the most well-known involves using AI to enhance healthcare delivery to aid clinical decision making, a process known as clinical decision support systems (CDSS) [4]. Historically, CDSS matched individual patient characteristics to a manually curated, computerised clinical knowledge base [5]. These ‘knowledge-based’ CDSS comprised a set of ‘if/then’ rules with the system retrieving data from the knowledge base to evaluate the rule and then generating an output or suggested action [4]. More recently, machine learning methods have enabled the development of ‘non-knowledge-based’ systems which use complex representations (patterns, trees, networks, equations) derived from large datasets to generate outputs or actions [4]. Machine learning models are typically difficult to understand for humans because of their complexity, especially when they were derived via deep learning - the most common method nowadays [4].

The integration of this form of AI in healthcare is notably progressing more slowly than initially anticipated given the levels of investment, due to several challenges to implementation [6, 7]. Some of these challenges extend beyond specific domain of AI and reflects broader issues encountered in implementation of innovation into practice. Frameworks such as the Consolidated Framework For Implementation Research shed light on these challenges by emphasizing the multifaceted nature of the implementation process [8]. Many challenges in AI implementation are similar to the implementation of any new digital technology and several frameworks have provided comprehensive lens through which these challenges can be examined. For example, the Non-adoption, Abandonment, Scale-up, Spread and Sustainability framework [9] and the meta-analysis based Unified Theory of Acceptance and Use of Technology (meta-UTAUT) [10, 11]. However, issues specific to AI technologies, particularly the non-knowledge-based systems have been reported [7]. Healthcare professionals often struggle with concerns about the reliability, interpretability, and ethical implications of these AI systems [6, 7].

This review aimed to qualitatively synthesise the existing evidence on the experiences of healthcare professionals in using non-knowledge-based CDSS (referred to as AI in methods and results sections) that are deployed in their work setting with a focus on issues specific to non-knowledge-based CDSS. We believe that focussing on for non-knowledge-based CDSS specific issues will help with understanding the challenges that remain for the deployment of these systems in health care.

Methods

Search strategy

This reporting of this systematic review adheres to the PRISMA reporting guideline as the PRISMA-AI guideline was not available at the time of writing [12, 13]. We conducted a systematic search of Medline, Embase, CINAHL, and Web of Science for relevant articles. Searches were initially conducted in May 2022 and updated in June 2023. Overall, four key concepts and their possible variations informed the search strategy: AI (such as, artificial intelligence, machine learning and natural language processing); decision support systems (such as decision support system, computer-assisted diagnosis); healthcare professionals (such as, healthcare professionals, doctor, nurse, physician); and terms relating to experience (such as, experience, view, opinion). Terms within similar categories were combined with OR and then the results from each category were combined with AND (see Multimedia Appendix 1). The reference lists of eligible studies were also screened, and experts contacted for relevant articles that may have been missed. There were no language and date restrictions.

Study selection

The search results were imported into EndNote referencing and bibliography management software to remove duplicates. The titles and abstracts of retrieved articles were screened by one reviewer (AA, JW or DM) and a random sample of 20% were screened by a second reviewer. The title and abstract screening was completed on Rayyan, a web-based application for systematic reviews. The full-texts of potentially relevant articles were then retrieved and examined in detail for eligibility by two independent reviewers (AA and IG or JW and DM). Any discrepancies were resolved by discussion between reviewers and a third reviewer or whole project team were consulted when necessary. The most challenging aspect of the study selection was deciding whether the AI is knowledge-based or non-knowledge based. When this was unclear, we discussed as a team and deferred to the team member with data science expertise (NP) for final decision.

Inclusion and exclusion criteria

Studies were included if: (1) they targeted healthcare or social care professionals; (2) the article described a non-knowledge-based clinical decision support system deployed for healthcare; and (3) the study explored healthcare professionals' experiences of using AI in healthcare; or barriers and facilitators to AI use in the real world. All study designs were deemed acceptable for inclusion, but non-primary studies were excluded. However, we screened the references included in relevant reviews to identify primary studies that may have been missed from our searches. The in/exclusion criteria for study selection are detailed in Table 1 below.

Table 1: Eligibility criteria

	Inclusion	**Exclusion**
Population	Healthcare and social care professionals	Non-professional caregivers (such as family caregivers)
Intervention	Non-knowledge based clinical decision support systems (Sutton *et al*., 2020) deployed for healthcare.	Knowledge-based clinical decision support systems. Systems under development or testing. AI as treatment (e.g. robopets).
Comparator	Any (including usual care or no comparator)	
Primary outcomes	Healthcare professionals' experiences of the use of AI in healthcare delivery. Benefits and challenges of the use of AI in healthcare. •	We excluded studies which focused on the perception of healthcare professionals on hypothetical use rather than actual use of AI.
Secondary outcomes	Healthcare professionals understanding of AI in healthcare delivery. Perceived/actual impact of AI on decision making and the health system (e.g. the need for extra clinicians) and clinical workflow and patient pathway.	
Types of study	Any primary study (empirical study) of any design (qualitative, quantitative, or mixed methods)	Non-primary studies (such as literature reviews, opinion papers). Studies exclusively reporting the

	Studies that focused on issues relating to health care provision, health outcomes and health service/system configuration were considered for inclusion.	real-world effectiveness/ performance/diagnostic accuracy were excluded unless accompanied by results relating to healthcare professionals' experience relating to outcomes stated above.
Context	Any healthcare setting from any country	

Data extraction and study quality assessment

A data extraction form developed and piloted by the review team was used to extract relevant information from each article. The following data were extracted from eligible studies: author names, publication year, study design, study aims and objectives, geographical location, sample size, characteristics of healthcare professionals included (including any indication of how long they have been using AI), description of the AI examined, input data (image/tabular/multimodal), platform (e.g. electronic medical record based, standalone app/website, mhealth embedded), tool development (e.g. was it developed in house or by a third party), study funder, and outcomes.

Data was extracted by one reviewer and checked by a second reviewer. Disagreements were resolved by deliberation between the two reviewers and where necessary, a third reviewer or the project team was consulted.

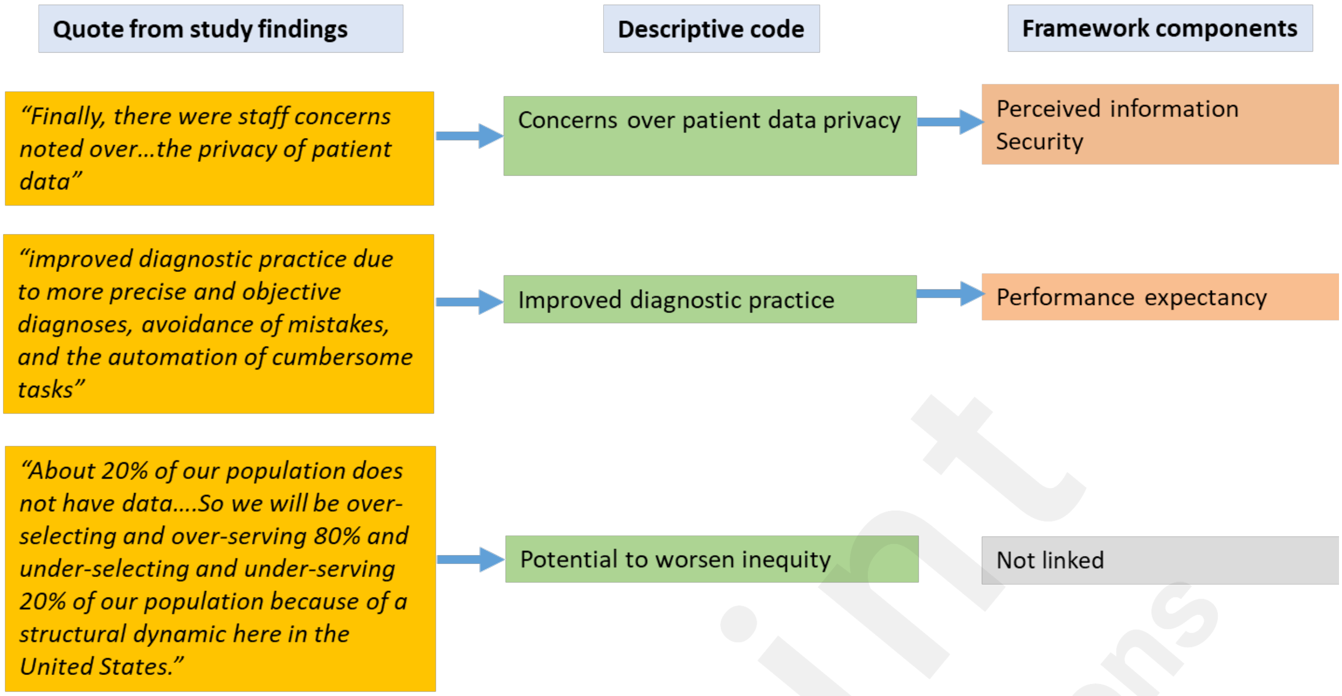
Two authors independently assessed the quality of all included studies using the Mixed Methods Appraisal Tool (MMAT) [14] which was appropriate due to the variety of study designs included in this review. Disagreements were resolved by discussion between the two assessors. A third reviewer was consulted to reach a final decision if the independent assessors were unable to reach a consensus.

Data analysis

Initially, a theoretically informed thematic synthesis was adopted to analyse the findings of the included studies [15]. The meta-UTAUT was used as a theoretical framework [10, 11]. This theoretical framework was selected as it is specific to acceptance and use of information systems and information technology, has been refined through meta-analysis and is extensively used for understanding adoption of information systems and technology [11]. Two independent reviewers coded the information from the findings of the included studies into descriptive codes and linked these codes to the components of the meta-UTAUT framework (Figure 1). During coding we remained alert to information that did not fit the framework (Figure 1). We used NVivo software to facilitate data coding process. Codes and coding were reviewed by the team and modifications discussed. Although the meta-UTAUT allowed us to understand the data, we noticed that the coding fragmented and the connections between the codes were being missed. Therefore, we decided to describe findings based on the broader themes.

After coding, we made a judgement as to whether codes were specific to AI or not by undertaking a thought experiment supported where necessary by published evidence. Team members (AA and FG) asked themselves, could this code be applied to another digital intervention including knowledge-based clinical decision support? If yes, we labelled it as non-specific. Examples of non-specific codes include problems with internet connection as this will cause a problem with any digital system requiring internet connection (e.g. video-calls); challenges with using digital systems can happen where insufficient training has been given [16]; alert fatigue is a recognised problem for any clinical decision support system [17], the importance of ensuring systems do not adversely affect workflows is a recognised issue for the design of digital interventions [18]. In our results, we provide examples of the non-specific codes. We identified and focus our analysis on the codes specific to AI. Through a process of iteration, discussion, and independent verification, we identified seven overarching themes that encapsulate the data.

Figure 1: Schematic illustration of the data coding process.



Quotes from the findings of the studies were coded into descriptive codes and linked to the components of the meta-analysis based Unified Theory of Acceptance and Use of Technology (meta-UTAUT) [10, 11]. In cases where no direct link to meta-UTAUT components existed, those codes were still retained for analysis.

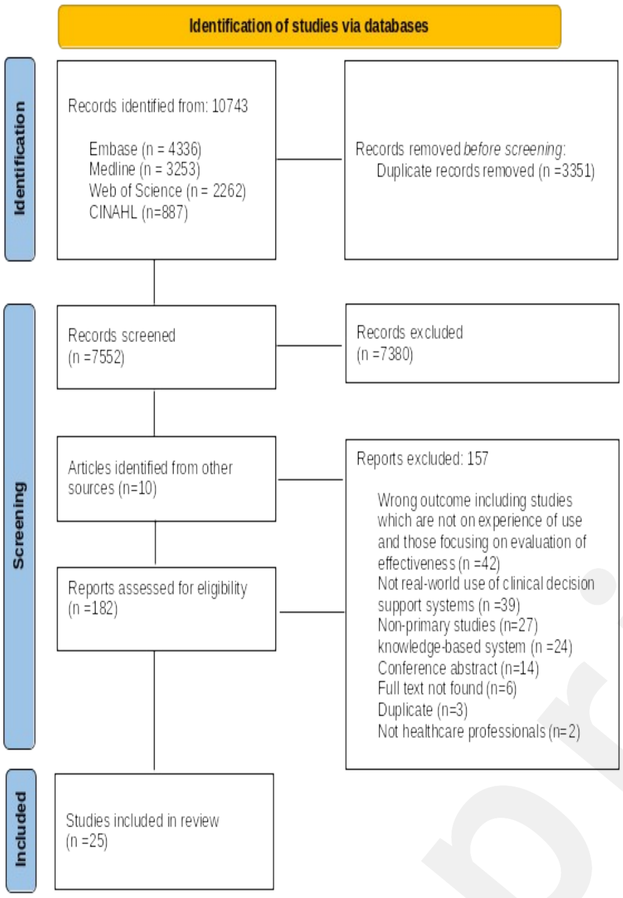
Stakeholder Engagement Workshop

We conducted a virtual stakeholder engagement workshop in August 2022 to obtain feedback on our initial findings and recommendations for future research and practice. We invited patient and public representatives, healthcare professionals, experts in AI and policy makers. Potential participants were invited through direct personal contacts, we sent out advertisements to NIHR Applied Research Collaboration (ARC) West Midlands patient and public representatives, Young Person's Mental Health Advisory Group of Kings College London, and Cross-ARC working group on AI. The workshop was a two-hour workshop over Zoom. We presented our evidence synthesis and invited attendees to provide feedback on our initial findings and recommendations of any additional research that could be conducted in this area and any practice implications from the existing findings.

RESULTS

A total of 10,743 records were retrieved from the electronic databases. After the removal of duplicates and screening based on titles and abstract, 182 full-text articles were screened. An additional 10 studies were identified from other sources. Twenty-five studies describing the experiences of healthcare professionals from differing disciplines and levels of seniority with various AI met the selection criteria and included in the final review. The process of study selection is shown in Figure 2. List of excluded studies and reasons for exclusion are presented in Multimedia Appendix 2.

Figure 2: PRISMA flow chart of study selection



Study characteristics

The characteristics of included studies are presented in Multimedia Appendix 2. The 25 included studies were published from 1990 to 2022 with 92% (n=23) of the studies published between 2019 and 2023. The studies covered nine countries, with majority from the USA (n= 12), then the Netherlands (n=2), Canada (n=2), Austria(n=2), and one from each of UK, Spain, Switzerland, Argentina, Brazil, China, and Thailand. The included studies were mainly qualitative (n=13, 52%) with 36% (n=9) quantitative and 12% (n=3) mixed methods studies. Sample size range from 12 to 724 participants, although two studies did not report this data [19, 20]. Participants of the included studies comprised various disciplines spanning various clinical specialities. The types of AI systems used in the included studies varied widely covering those used in risk prediction, diagnosis, and treatment (Table 2).

Table 2: Description of AI systems used in included studies

Study ID (Author Year)	Study type	Name of AI system	Clinical decision-making task	Setting, Country	Target population	Tool development
Beede 2020, [21]	Qualitative study	No name given	Assessment of diabetic retinopathy	Primary care clinics, Thailand	Patients (diabetic) requiring an eye examination	In house
Cruz 2019, [22]	Quantitative study	Savana System	Notifies physicians in real time about recommendations regarding the healthcare process to improve adherence rates to clinical pathways	Primary care clinics, Spain	Patients attending primary care	Commercial
Frymoyer 2020, [23]	Quantitative study	InsightRX-MIPD CDS tool	Determining the optimal dosing regimen of Vancomycin	Hospital, USA	Neonates and children for suspected or documented infections with methicillin-resistant Staphylococcus aureus (MRSA), methicillin-resistant coagulase-negative Staphylococc	Commercial

					i, and other drug-resistant gram-positive organisms.	
Ginestra 2019, [24]	Quantitative study	EWS 2.0 alert	Sepsis risk prediction	Hospital, USA	Non-Intensive Care Unit admissions	In house
Goncalves 2020, [19]	Qualitative study	Robot Laura	Early identification of sepsis	Hospital, Brazil	Patients in hospital	Commercial
Henkel 2022, [25]	Quantitative study	AI-Pathway Companion Prostate Cancer VA10B	Diagnosis and management of prostate adenocarcinoma cancer	Hospital, Switzerland	Patients undergoing prostate cancer screening	Commercial
Henry 2022, [26]	Qualitative study	Targeted Realtime Early Warning System (TREWS)	Sepsis detection and treatment management	Hospital, USA	Patients in hospital	Unclear
Jauk 2021, [27]	Mixed methods study	unclear	Predicting delirium	Hospital, Austria	Every patient admitted to one of the departments	Unclear
Jauk 2022, [28]	Quantitative study	Delirium prediction software	Delirium risk prediction	Hospital, Austria	Patients in hospital	In house
Jordan 2023, [29]	Qualitative study	KATE™	Clinical triage	Hospital, USA	Patients at emergency department	Commercial
Joshi 2022,	Qualitative	No name	Sepsis risk	Hospital,	Patients at	Commercial

[30]	study	given	prediction	USA	risk of sepsis	and in house
Kappen 2016, [31]	Mixed methods study	No name given	Predicting the risk of postoperative nausea and vomiting (PONV)	Hospital, Netherlands	All surgical patients	Unclear
Lebovitz 2022, [32]	Qualitative	No name given	Diagnosis of breast cancer, lung cancer and estimation of bone age.	Hospital, USA	Patients requiring diagnostic radiology	Unclear
Marwaha 2020, [33]	Qualitative study	Face2Gene ®	Diagnosis of children with rare genetic syndromes	Hospital, Canada	children with rare genetic syndromes	Unclear
McAdam 1990, [20]	Quantitative study	Leeds Abdominal Pain System	Diagnosis of acute abdominal pain	Hospital, UK	patients presenting with acute abdominal pain	In house
Nehme 2023, [34]	Quantitative study	GI Genius	Colorectal polyp detection	Cancer center, USA	Patients undergoing elective outpatient colonoscopy	Commercial
Rabinovich 2022, [35]	Mixed methods study	TRx	Chest x-ray interpretation	Hospital, Argentina	Patients at emergency department	In house
Romero-Brufau 2020, [36]	Quantitative study	No name given	Identify patient at risk for poor glycemic control in the next 3 months.	Primary care clinics, USA	Patients at risk of poor glycemic control	Commercial
Sandhu 2020, [37]	Qualitative study	Sepsis Watch	Sepsis risk prediction	Hospital, USA	Patients in hospital	In house

Saunders 2021, [38]	Qualitative study	The modified Hospitalised -patient One-year Mortality Risk (mHOMR) tool	Mortality risk prediction	Hospital, Canada	Patients with non-cancer serious illnesses	Unclear
Shiang 2022, [39]	Quantitative study	Artificial intelligence-based decision support system	Detection of pulmonary embolism, intracranial hemorrhage, and acute cervical spine fractures.	Hospital, USA	Patients at radiology department	Commercial
Singer 2022, [40]	Qualitative study	No name given	Readmission tool: Readmission risk prediction	Hospital, USA	All patients admitted to the hospital	Unclear
Strohm 2020, [41]	Qualitative study	BoneXpert	Bone maturity assessments based on X-rays of paediatric patients' hands.	Hospital, The Netherlands	Pediatric patients	Commercial
Sun 2019, [42]	Qualitative study	Watson for Oncology	Personalised cancer management	Hospital, China	Cancer patients	Commercial
Wang 2023, [43]	Qualitative study	No name given	Predicting risk of peripheral arterial disease	Community/ Primary care, USA	PAD patients who had at least one visit to Duke Health with a PAD related diagnosis code between January 2015 and March	In house

					2016	

Quality of included studies

The quality of included studies is presented in Multimedia Appendix 2. Overall, the quality of most of the included studies was good. Of the 25 included studies, 12 qualitative studies [19, 21, 26, 29, 32, 33, 37, 38, 40-43] fulfilled all the relevant quality criteria on MMAT while five studies fulfilled at least 80% of the quality criteria [20, 22, 24, 27, 30]. The remaining eight studies fulfilled 67% or less of the quality criteria [23, 25, 28, 31, 34-36, 39] and most of the unmet criteria were assessed as "can't tell" primarily due to lack of sufficient information.

Excluded generic issues

We identified a range of non-specific issues related to the implementation of digital technologies in healthcare (see Textbox 1 for examples). These issues have already been extensively addressed in the existing literature [8], and were therefore excluded from analysis in our study.

Textbox 1: Examples of broader issues relating to implementation reported in included studies

- Issues relating to abandonment [30, 41].
- Competing priorities, especially in cases where healthcare professionals felt the AI was not particularly useful [31, 38].
- Limited infrastructure to support the adoption of the technology [19, 21].
- Healthcare professionals' satisfaction with AI applications [19, 23, 25, 30].
- Ease of use [25, 27, 28, 35].
- Integration into the existing systems and accessibility within the existing system [23, 27, 41].
- Healthcare professionals appreciated systems that do not affect the dynamics of their workflow [22, 27].
- Concerns about alert fatigue/redundancy of alert [30, 38].
- Technological challenges relating to healthcare professionals' proficiency in the use of hardware and software [19].
- Issues relating to information security, including the necessity of secure transmission of data, [23] concerns over privacy of patient data [33] and potential

threat to national security [42].
- Lack of complete data for patients [43].
- Lack of standards for what and how data is collected [42]
- Issues with internet connections [19, 21].
- Challenges with hardware like recording equipment, smartphones and tablets [19].

Key themes

In this section we describe the experiences of healthcare professionals in using AI to support clinical decision making as reported in the included studies. We focus on the nuances we believe are specific to AI following the process described in the methods. We identified seven themes: (i) Understanding of AI applications; (ii) Level of trust and confidence in AI tools; (iii) Judging the added value of AI; (iv) Data availability and limitations of AI; (v) Time and competing priorities; (vi) Concern about governance; (vii) Collaboration to facilitate the implementation and use of AI. We provide detailed exploration of each theme below.

Understanding of AI applications

Ten studies reported concerns with healthcare professionals' lack of understanding of AI applications. This includes lack of understanding of the outputs of the AI application as well as lack of understanding of the algorithms/ the rationale for the outputs by the AI applications due to lack of transparency relating to non-knowledge based AI applications [24, 26, 27, 30, 32, 37, 38, 40-42].

> *"This lack of transparency is perceived as a major challenge; the AI technology represents as a "black box", and its users have no power to understand its mechanisms, or modify them to tackle potential problems" [42].*

For example, a study on the use of a machine learning sepsis early warning system reported:

> *"Both RRT [Rapid Response Team] nurses and ED [Emergency Department] physicians said that they lacked the knowledge and understanding required to assess the validity of the machine learning model……..Physicians also lacked knowledge about the model and the predictive nature of the model" [37].*

Another study on sepsis risk prediction reported confusion regarding understanding what the alert means. [30] In an attempt to minimise this issue, inclusion of an explanation for the alert firing was introduced in some hospitals. Some felt the explanations were helpful whilst others found it more confusing. [30]

> *"A lot of people get confused...so say you get 25, when the patient's really sick and then the number goes to twenty, does that mean the patients getting better? What do all of the subsequent numbers mean? If it goes up to 30, is the patient getting worse?" [30].*

> *"Concise content with explanations for firing was well-received. A few noted that this was not possible or was more confusing, but most felt inclusion of explanations was helpful" [30].*

Participants in a study on the use of AI tool in radiology reported:

> *"How does [the AI tool] know that this is a nodule, but this isn't?" (Dr. V) [32].*

> *"They all look identical to me" (Dr. C). They expressed frustration in their inability to understand the divergent AI results: "What is it telling me to look at? At this tissue? It looks just like the tissue over here, which is perfectly normal … I have no idea what it's thinking" (Dr. K)." [32].*

However, in a study on early warning system for sepsis, limited understanding of how the AI operates was not considered a major barrier to the use of the application [26].

Level of trust and confidence in AI tools

Findings relating to trust in accuracy and judgements/recommendations of AI tools is mixed. [19, 21, 26-28, 30-37, 39, 41, 43] Some Healthcare professionals believe that the AI performs as expected while others did not, and this was reported across various AI tools. For example, in a study on sepsis risk prediction tool, implementers were unsure whether the AI tools were able to predict sepsis with clinically meaningful specificity when compared with the traditional early warning system using systemic Inflammatory response system which rely on physiological data (vital sign and lab abnormalities). [30] Participants noted disappointment about predictive potential:

"The tool... was supposedly predictive, but we discovered...it wasn't predictive.. .it was really telling providers that they've met the criteria for severe sepsis which...is not really predictive because they've already met it. It wasn't that you were getting it before it happened so even though they were selling it as a predictive model I'm not so convinced it was predictive" [30].

In another study on the use of AI to support delirium prevention:

"s*even users (14.9%) did not believe that the application is a useful support for delirium prevention, and seven did not believe that the application can be used to detect delirium at an early stage" [27].*

Referring to the use of AI tool for lung cancer diagnosis, a study reported:

"Radiologists were deeply committed to providing judgments with maximum certainty, but they expressed difficulty feeling certain given the opacity they experienced when considering divergent AI results: "I just don't know of any radiologist who's not looking closely at the case because they have AI. Because at the end of the day, you're still responsible. How can you trust the machine that much?" (Dr. E)" [32].

In a study among healthcare professionals on the use of BoneXpert, some Healthcare professionals struggle to accept outputs from the application [41].

" Interestingly, in three hospitals, we found that the referring clinicians did not trust the output of the AI application and redid a manual bone age analysis for every scan. Thus, just like the radiologists, the referring clinicians showed varying levels of acceptance of AI applications" [41].

Divergence in opinion of healthcare professionals and AI recommendation creates uncertainty. However, by using "AI interrogation practices" radiologists are able to use AI in a way that helps them to experience less uncertainty in making final judgements [32].

> *"On the surface, it may seem that using the AI tools (and experiencing opacity) increased the overall uncertainty these radiologists experienced; however, in fact, using the AI tool resulted in radiologists experiencing less uncertainty making their final judgments. They achieved this by using "AI interrogation practices" or practices that human experts enact to relate their own knowledge claims to the AI knowledge claims [32].*

Other experiences that led to lack of trust in AI were the lack of relevant empirical evidence on AI clinical performance and of its impact on healthcare workflow and quality, and the problem for clinicians of AI results that do not match their own judgements [32, 41]. For example, in a study of the implementation of AI applications in radiology across seven hospitals in the Netherlands, authors reported:

> *"there is a lack of empirical evidence on the effect of AI applications on the radiological workflow, as well as their added value for clinical radiology practice… measuring clinical and organizational benefits of AI on a micro-level is difficult ....publications on the validation of the algorithms are based on laboratory rather than clinical settings" [41].*

Societal expectations may also influence confidence in AI use. In a study of the use of Watson tool in China, authors reported that social media and organisations frequently talk about AI and attributed "magic" qualities to AI which often leads to doctors being disappointed [42].

> *"Hospital managers/ doctors report to experiencing frustration when facing the real technology after the societal hype" [42].*

In one study, participants suggested that having a track record of healthcare professionals' action, based on AI recommendation and the corresponding patient outcomes could facilitate confidence in AI tools [43].

[43]

Judging the added value of AI

In some studies, healthcare professionals reported that AI provided added value and improved decision making [19, 20, 24, 27-30, 35, 36, 39-41]. This includes automation of burdensome tasks [41], avoidance of mistakes [32, 41], help reduce variability in clinical practice [22], improved team communication [24, 36], reduction of alert fatigue due to the combination of multiple datapoints [30], improvement in data collection [20] and team performance [19, 27], provision of more objective and precise diagnosis/prediction [41].

In some studies, some healthcare professionals found AI tools served as reminder or confirmation of their clinical judgements [23, 26, 27, 29, 32, 35, 38, 39, 42]:

> *"The prediction helps to corroborate my own estimation when seeing a patient "* [27].

However, not all users of the same tool found it useful [27],

> *"Opinions about the application's usefulness for their own work were mixed: 17 users (36.2%) reported the application to be useful for their work, while 15 users (31.9%) did not find it useful" [27].*

Healthcare professionals reported some problems relating to AI technologies not addressing issues that are deemed important to them. For example, healthcare professionals thought postoperative nausea and vomiting has low burden and should not be excessively treated [31]. In a study in a U.S. hospital among healthcare professionals using AI tools for diagnosing breast cancer, lung cancer, and bone age determinations, radiologists said AI tools identify abnormalities that they might not see but they were not clinically significant to impact on their final judgement [32]:

"Calcifications can be really little and sometimes hard to see. It [Mammo AI tool] sees those calcifications better than I do. But it also sees all kinds of calcifications that are neither here nor there" (Dr. B) " [32].

In a study of peripheral arterial disease (PAD) identification algorithm, clinicians emphasised the importance of ensuring the AI tools focus on supporting clinical decision by performing additional analysis that the clinician does not normally do [43]:

"I mean, it [the algorithm] has to solve a real problem. Like, I'm not interested in models that in a clinical sense, identify data that I could just identify in the course of my daily work." [43].

Some studies also reported that Healthcare professionals expressed uncertainty about next steps and that AI provided no actionable outputs, therefore not adding value to their clinical decision making [31, 38, 43]. Adding treatment recommendation would have been more useful. [31] For example, referring to a tool to predict the risks of postoperative nausea and vomiting a physician said:

"The intuitive use of the predicted PONV [postoperative nausea and vomiting] *risk and a stated preference for an actionable recommendation by intervention group interviewees suggested that being presented only with a predicted risk may be difficult to use in a clinical decision. Adding a risk-corresponding treatment recommendation may assist physicians in interpreting the predicted risk for a decision on PONV prophylaxis (theme 4G, quote 43)" [31].*

Healthcare professionals also voiced concerns over AI not providing actionable information:

"RES02: 'It would have been nice to have some sort of actionable items, because while the information is good to know, I was never really sure what to do with it. It's like, great, my patient has an elevated one year mortality risk. What can I do about

that? What do I do with this information?" [38].

"They gave us the risk scores, but we did not really understand what to do with this information, what do these scores mean? It took us a while to figure out what is our threshold for what are we going to consider high risk patients." [Readmission Risk Tool user]" [40].

Also, in a study of an automated mortality prediction model for identification of patients for palliative care, AI helped to reduce uncertainties especially in less experienced Healthcare professionals [38]. *[37]*In some situations, AI was reported to cause added work for healthcare professionals [21, 29, 34, 41]. For example, in a study on the use of AI for the detection of diabetic retinopathy the AI required a high standard quality data/image which may be difficult to achieve in some cases [21]. A study participant reported:

" "It gives guaranteed results, but it has some limitations. Some images are blurry, and I can still read it, but [the system] can't." P3 shared the same sentiment, "It's good but I think it's not as accurate. If [the eye] is a little obscured, it can't grade it." The system's high standards for image quality is at odds with the consistency and quality of images that the nurses were routinely capturing under the constraints of the clinic, and this mismatch caused frustration and added work" [21].

In one of the studies on the implementation of AI in diagnostic radiology, the clinical benefits or organisational goals for using AI applications are not clearly established before implementation, making it hard to assess after implementation [41].

"From an organizational perspective, clinical benefits or organizational goals that might be achieved by using AI applications are not clearly established ex-ante and therefore hard to assess after implementation" [41].

Healthcare professionals were concerned patients may not be willing to have AI used for them [42]:

> *"They [the patients] have no idea about Watson [a system for designing personalised treatment for cancer patients]. They will think: why do I need a machine to look at [my problem]? I prefer an expert doctor" [42].*

Health professionals expressed concerns about limiting treatment for palliative care patients based on the risk prediction. [38] Also, there were concerns about clinicians and regulatory agencies becoming overreliance on AI [26, 29]. For example, healthcare professionals may not actively consider their cases and they may not be able to refine their skills and maintain cultural competence [29].

Data availability and limitations of AI

A study reported the users concern about the potential for AI to worsen inequity as disadvantaged populations are not represented in health data sets [43].

> *"About 20% of our population does not have data….. So we will be over-selecting and over-serving 80% and under-selecting and under-serving 20% of our population because of a structural dynamic here in the United States" [43].*

> *"And I should also say those even worse, those 20% [who we don't have data on], or a higher proportion of them are older, coming from an economically higher social deprivation index, higher disease burden, non-white…" [43].*

Concerns over the limits of AI in considering all the aspects of patients' lives were also talked about [31, 43] For example, in the study on AI tool for PONV [31]:

> *"they [physicians] felt that a prediction model does not take into account all aspects of a specific patient. The prediction model only predicted the risk for a specific outcome and did not weigh the benefit of treatment against the expected harms and contraindications for a particular patient with specific characteristics and*

comorbidity" [31].

In the study of the use of peripheral arterial disease identification algorithm, healthcare professionals reported having access to more data about patient than AI algorithm uses [43]. There were concerns about AI tool not being able to account for factors such as patient's culture [29] and ground context and therefore limiting the potential impact of the AI [43]:

> *"Exercise, nutrition, smoking, and medication adherence are critical components of PAD [peripheral arterial disease] management and clinicians expressed concern about adopting the algorithm to make intervention recommendations without knowing more about these factors…… The algorithm's potential impact was limited by an inability to account for barriers patients faced in managing PAD"* [43].

Time and competing priorities

Healthcare professionals reported lack of time to fully utilise AI tools. For example, in a study on the use of the modified Hospitalised-patient One-year Mortality Risk (mHOMR) tool, physicians reported lacking time to address the alert as they were often focused on the acute needs of the patient [38]:

> *"PHYS07: 'The inpatient stay often is very compressed. They're in the hospital, they're getting treated, and then they're home. And so, there is not time during the inpatient stay to address these things" [38].*

In the study of the use of AI for peripheral arterial disease identification, healthcare professionals reported limited time to review high-risk patients and having difficulties in choosing which patients to prioritise [43]:

> *"The PAD [peripheral arterial disease] ML-driven CDS can be run on all patients in a population, but only a small number of high-risk patients could be reviewed each week. Stakeholders described the difficult trade-offs they had to make when considering how to pick patients to prioritize for Population Rounding" [43].*

Time pressure could be further heightened where AI recommendations differs from healthcare professionals' clinical judgment. [32] However, in a study on the use of AI in diagnostic radiology, radiologists were willing to invest additional time to relate their own knowledge to the AI knowledge claims to build understanding of the AI results and reconcile divergent views [32]:

> *"I know my limitations and I know this [CT AI] is going to help them [nodules] stand out a little better. It's worth the extra time in my mind" [32].*

Concern about governance

Healthcare professionals expressed concerns over regulatory and legal uncertainties surrounding AI use [41, 42]. For example, uncertainties regarding legal responsibilities for misdiagnosis occurring from AI was reported in a study in the Netherlands [41] and this may vary for different countries. In another study, a healthcare professionals reported that it is illegal for AI applications to make clinical decisions in China:

> *"In China it is illegal for an AI system to make a decision [1HP05]."*

> *"In our hospital, we use Watson to assist the Multidisciplinary Team (MDT). […] We discussed how to use Watson for a long time. […] As the first hospital to use Watson, we find this way [to use Watson together with the MDT]. […] But this really gives us a heavy burden! Because when we use Watson, we must have at least five doctors to work together with Watson [as required by regulation]. They [the five doctors] will sign on the report. [1HP05]" [42].*

Collaboration to facilitate the implementation and use of AI

In a study on the use of low bed tool and readmission risk tool, authors described how healthcare organisations can enable collaboration between key stakeholders (users, developers and outside experts) to facilitate the development and implementation of AI [40]. This collaboration allowed users (care management and utilization management team members) to be involved in the identification of initial needs, identification of new users, need for experts, identification of potential sources of inaccuracy and possible areas of improvement [40].

Having healthcare professionals to serve as champions/leads for AI use and having knowledge exchange platforms supported the use of AI [26, 41, 43]. For example, the study on the implementation of AI applications in clinical radiology reported that local champions (radiologists with keen interest in AI and good understanding of AI applications) are vital in encouraging implementation within their departments [41]. In another study, having a physician lead the integration of AI tool for PAD identification was reported to be beneficial in facilitating the translation to create real-world clinical impact [43]:

> *"Dr. XXX obviously was very aware, she was kind of the one that spearheaded [the integration] and wanted to get this going. So I think you know, [targeting clinical need] is her role. I'm sure with her specialty, she probably realized that it was an area that needed more attention." Operational stakeholder [43].*

Also in the study of the use of AI for early warning for sepsis, there was a new role created - sepsis watch nurse [37]. The sepsis watch nurses are the primary users of the tool and were extensively trained in the workflow and the tool. They are responsible for communicating the outputs from AI tool to physicians to facilitate the translation of the recommendations to patient's bedside. With regards to skills required to fulfil such roles, the nurses recommended:

> *"When asked about the skills and knowledge needed to be a good Sepsis Watch nurse, the RRT nurses mentioned good clinical judgment, knowledge of sepsis, and critical care experience..…….RRT nurses also explained the importance of strong communication skills to confidently speak with attending physicians whom they may not personally know……….RRT nurses thought that strong computer skills were not necessary for the role, given the simplicity of the app. RRT nurses also recommended recruiting nurses interested in the role and the need to create buy-in through continuous feedback" [37].*

In another study on early warning system for sepsis, deployment team attended staff meetings and also met with individual users, as requested, to explain the system and provide guidance on using the interface [26]. Users were also able to ask questions about system's behaviour using a feedback button [26].

In a study on the implementation of AI in diagnostic radiology, it was reported that only one of seven hospitals had a formalised innovation strategy regarding AI although three hospitals had a designated innovation manager [41]. It is worth noting that four more hospitals were reported to be developing a structured approach at the time of the study. The lack of structured implementation processes lead to substantial variations in how the application is used in different departments [41].

> *"From a workflow perspective, implementation plans do not specify how the AI application should be integrated into the clinical workflow, which leads to significant variations in the way the application is used in different departments. Furthermore, in all cases, the work done to monitor existing practices or the impact of the implementation of novel technologies on the level of the hospital is currently limited" [41].*

Key points from stakeholder engagement

The stakeholder engagement workshop was attended by 18 individuals, besides the study team and one patient and public engagement officer. This included nine patient and public representatives, seven researchers from divers related fields (including clinical researchers), one representative from an AI technology company, one policy maker in data and AI. After being presented with the study findings, participants agreed that algorithmic transparency and understanding of how AI makes predictions may be challenging with non-knowledge-based AI. They believe healthcare professionals are aware AI will be utilised more in the future, but do not feel prepared for it. Participants highlighted the need for more evidence to showcase good practice and collaborations across specialties/disciplines. They recommended early training of healthcare professionals to improve their understanding of what AI is and how it works. They acknowledged that some of the training programs may be based on theory rather than wait until AI is fully deployed in all clinical settings. They also recommended need for guidelines for AI applications to include examples of how conflicts between AI recommendations and healthcare professionals' judgments should be resolved. With regards to suggestions for future research, they suggested that more case studies and mixed methods research would be useful to provide insights into human perspectives of the quantitative research and explore healthcare professional's confidence in AI systems. Participants recommended baseline studies on clinician knowledge and experiences of AI systems to provide comparison data for future work. There were also suggestions of a need for more information on the perspectives of hospital management and patients on the real-world use of AI systems. Differences in perspectives based on individual characteristics (such as age) could also be explored.

Discussion

Principal Findings

This review describes existing empirical evidence on healthcare professionals' real-world experiences of using non-knowledge based CDSS. While some of the issues emphasised in the included studies can be attributed to implementation of any technology or intervention, we identified many issues that appear to be particularly challenging for non-knowledge based CDSS. We grouped these issues into seven themes: (i) Understanding of AI applications; (ii) Level of trust and confidence in AI tools; (iii) Judging the added value of AI; (iv) Data availability and limitations of AI; (v) Time and competing priorities; (vi) Concern about governance; (vii) Collaboration to facilitate the implementation and use of AI. Most frequently occurring issues include concerns over lack of understanding of AI outputs and lack of understanding of the algorithm/rationale for the AI output. Also, many studies highlighted the issue of trust and confidence in AI tools where some healthcare professionals believe that the AI performs function as anticipated while other hold contrary opinions. This lack of trust is further compounded when there are divergent in opinion of healthcare professional and AI recommendation. Another notable issue relates to the added value of AI. Some healthcare professionals reported that AI added value whereas, some believed it did not add value, especially when the AI technologies do not address issues that are important to the healthcare professionals or produce any actionable outputs. The challenges are interlinked, lack of understanding of the AI affects ability to use the output and so trust cannot develop and health professionals struggle to see the value in the AI and remained concerned about clinical responsibility and liability. The regulatory context and how the AI was implemented influence their approach to using the AI.

Comparison with other work

Other authors have found that the ability of healthcare professionals to trust non-knowledge based CDSS and understand how they work to benefit patient or improve care delivery are important factors in their adoption from both healthcare professionals and public perspectives [44-46]. The lack of robust empirical evidence to support the use of these systems in healthcare is thought to contribute to these issues [47, 48]. One study used the expectancy-value theory [49] and modified extended UTAUT [50] to understand how expectancy (effort expectancy and performance expectancy), trust, and perceptions of clinicians related to their intention of using an AI-based CDSS – Blood Utilisation Calculator [51]. The findings showed that expectancy and perceived risk on using the application have significant effect on trust [51]. Establishing a foundation of trust is crucial for the acceptance and effective integration of various AI technologies into healthcare practices [52].

It has been argued that for machine learning or AI systems to become trusted and accepted in healthcare, clinicians, researchers and patients need to feel that the conclusions the systems make and the way that they reach them are interpretable and explainable [53]. This includes the clinician understanding the limitations of the system due to limitations on data quality [54] and the complexity of the data [55]. However, studies suggest that whether explainability does add value to AI powered clinical decision support depends on the context it is being used and by whom in addition to technical considerations [56, 57]. In contrast, it has been argued that accuracy is sufficient for day-to-day clinical practice, with explainability a research endeavour [58].

The view that algorithmic bias can exacerbate inequalities is supported by empirical evidence [59]. These biases can be mitigated to some extent [60]. Regulators recommend that data used in training algorithms is representative of the intended patient population and an international initiative is underway to develop recommendations for the composition and reporting of datasets used in AI systems for clinical practice [61]. However, as our findings indicate, health care professionals may have data about their patient that is missing from datasets used in diagnostic CDSS [4].

The inability by clinicians to process all the information presented to them is not new but potentially worsened by digitally available information, particularly in time pressured clinical settings [62]. The need for healthcare organisations to optimize CDSS alerts is not limited to non-knowledge based systems [63]. The concern expressed about clinical responsibility and liability for non-knowledge based CDSS is widely shared [64]. Regulation is developing but solutions are not straightforward as there is interaction between AI trustworthiness and transparency and how the AI and clinician work together, which also changes the clinician-patient interaction [64, 65]. An extensive systematic review of barriers and enablers to the implementation of CDSS for chronic disease identified similar implementation issues as in our review, such as the importance of a champion for the system, the deployment of allocated personnel to use the system and the need for accessible training [66]. A recent study explored experiences of various stakeholders, including health care professionals and regulatory bodies in developing and using AI technologies [67]. The study's recommendations for promoting the deployment of AI align with our findings and the stakeholder engagement workshop. These shared recommendations emphasise the importance of establishing guidelines for AI technology development and adoption, enhancing co-creation, and providing comprehensive education and training for healthcare professionals, patients, and communities [67].

Future Directions

Many studies have examined the perceptions of healthcare professionals' on the use of AI in the healthcare settings [46, 68]. However, the majority were not focused on AI applications that are fully deployed in real world settings and even fewer have focused on non-knowledge based CDSS. This review provides more nuanced understanding of the challenges faced by healthcare professionals in contexts where non-knowledge based systems are fully deployed. Many of the challenges identified in our review could be mitigated by collaboration with various stakeholders, including healthcare professionals, patients and regulators, during the design and development stage of AI systems [69]. By involving key stakeholders in the process, developers gain invaluable insights into the practical needs and concerns of healthcare professionals, facilitating the creation of AI tools that are more aligned with the real-world requirements. Regular evaluation and validation of AI tools are essential to generate necessary data that are important for end users [70]. This will provide empirical evidence of the utility of the tools, clinical effectiveness, and generalisability [70]. Furthermore, the provision of education and training is necessary as this would equip healthcare professionals with necessary knowledge and skills to enhance their confidence and competence in using AI in their practice [67]. With these issues solved as AI development matures, further research on the experience of healthcare professionals using AI can focus on identifying unintended consequences of AI use in routine clinical practice.

Limitations

The methods used in this systematic review were established a priori. We used a comprehensive search strategy, searched four electronic databases, and contacted experts to ensure we identify relevant studies. Due to limited time and human resources, we were unable to perform independent double screening at the title and abstract screening phase. However, at least 20% were randomly screened by a second reviewer and full texts were independently screened by two reviewers. Although we dedicated substantial effort to creating a comprehensive search strategy, numerous included studies were identified through recommendations rather than the structured search. This underscores the inherent challenge in rigorously searching databases for this topic, given the diverse language used to describe CDSS. More so, it was difficult to judge sometimes whether AI is knowledge-based or non-knowledge based, particularly due to limited descriptions in some studies. Consequently, we recognise the possibility of overlooking some potentially relevant articles. Despite this limitation, our findings align with existing studies, and we believe that this review effectively captures crucial aspects of healthcare professionals' experiences in this domain.

Conclusion

This review emphasise the complex challenges faced by healthcare professionals using non-knowledge based CDSS. Issues related to understanding AI applications, trust in AI tools, and assessing added value are prominent and the interlinked nature of the challenges identified is evident. As we navigate the evolving landscape of healthcare technologies, it is important to acknowledge and address these challenges through targeted strategies to improve collaborative development, continuous evaluation/validation, education and refining the regulatory framework. This will potentially enhance the acceptance and use of AI within healthcare settings.

Funding

Project is funded by Warwick AI National Strategy Fund – Alan Turing Award

Acknowledgements

Jeffry Hogg is funded by an NIHR doctoral fellowship (NIHR301467). Abimbola Ayorinde is affiliated to the National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) West Midlands (NIHR200165). Niels Peek is funded by NIHR Applied Research Collaboration Greater Manchester (NIHR200174). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

We are grateful to Samantha Johnson, Academic Support Librarian at the University of Warwick, for her invaluable assistance in developing an effective search strategy that greatly enhanced the quality of our work. Additionally, we extend our appreciation to Niyah Campbell, Public Involvement and Engagement Officer at NIHR ARC West Midlands, for his help in coordinating the stakeholder engagement workshop. We acknowledge the contributions from members of the Warwick Medical School Artificial Intelligence Consortium who participated in the stakeholder engagement workshop: Michael Adeyemi, Katherine Barrett, Nurgul Gecer, Lukas Hughes-Noehrer, Karen Keates, Melina Kunar, Sarah Markham, Siobhan O'Connor, Theodora Oikonomidi, Busola Oronti, Kate Ryan and Sagar Jilka. Our deepest appreciation goes to all the public representatives, healthcare professionals, experts in AI and policy makers who willingly shared their insights and perspectives, enriching the depth of our research. We would like to acknowledge the Cross-Applied Research Collaboration Artificial Intelligence Working Group for their collaborative efforts and valuable contributions throughout the course of this project.

Author's contributions

AA, FG and NP were involved in the conceptualisation and funding acquisition. AA, DM, JW, IG, JH, NP and FG contributed to the design of the review. AA, DM, JW, IG, and AI contributed to data curation. AA, DM, JW, and FG performed formal analysis with input from all authors. AA wrote the original draft of the manuscript, and all authors reviewed the manuscript and approved the final version.

Conflicts of Interest

None to declare

Data Availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Abbreviations

AI	Artificial Intelligence
ARC	Applied Research Collaboration
CDSS	Clinical Decision Support Systems
CINHAL	Cumulative Index to Nursing and Allied Health Literature
ED	Emergency Department
ML	Machine Learning
MMAT	Mixed Methods Appraisal Tool
PAD	Peripheral arterial disease
PCP	Primary care providers
PONV	Postoperative nausea and vomiting
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RRT	Rapid response team
UK	United Kingdom
USA	United States of America
UTAUT	Unified Theory of Acceptance and Use of Technology

Multimedia Appendix 1: Search strategy for all databases.

Multimedia Appendix 2: Data extraction, quality assessment and list of excluded studies.

REFERENCES

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-8.

2. O'Dowd A. Government pins hopes on £250m AI centre for faster diagnosis and treatment. BMJ. 2019;366:l5106.

3. Zhang J, Whebell S, Gallifant J, Budhdeo S, Mattie H, Lertvittayakumjorn P, et al. An interactive dashboard to track themes, development maturity, and global equity in clinical artificial intelligence research. The Lancet Digital Health. 2022;4(4):e212-e3.

4. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17.

5. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj. 2005;330(7494):765.

6. Hazarika I. Artificial intelligence: opportunities and implications for the health workforce. Int Health. 2020;12(4):241-5.

7. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Services Research. 2022;22(1):850.

8. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Science. 2009;4(1):50.

9. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res. 2017;19(11):e367.

10. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD. Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model. Information Systems Frontiers. 2019;21(3):719-34.

11. Dwivedi YK, Rana NP, Tamilmani K, Raman R. A meta-analysis based modified unified theory of acceptance and use of technology (meta-UTAUT): a review of emerging literature. Curr Opin Psychol. 2020;36:13-8.

12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

13. Cacciamani GE, Chu TN, Sanford DI, Abreu A, Duddalwar V, Oberai A, et al. PRISMA AI reporting guidelines for systematic reviews and meta-analyses on AI in healthcare. Nature Medicine. 2023;29(1):14-5.

14. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information. 2018;34:285-91.

15. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology. 2008;8(1):45.

16. Upadhyay S, Hu H-f. A Qualitative Analysis of the Impact of Electronic Health Records (EHR) on Healthcare Quality and Safety: Clinicians' Lived Experiences. Health Services Insights. 2022;15:11786329211070722.

17. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, with the HI. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Medical Informatics and Decision Making. 2017;17(1):36.

18. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci. 2016;11(1):146.

19. Goncalves LS, Amaro MLD, Romero ADM, Schamne FK, Fressatto JL, Bezerra CW. Implementation of an Artificial Intelligence Algorithm for sepsis detection. Revista Brasileira De Enfermagem. 2020;73(3).

20. McAdam WA, Brock BM, Armitage T, Davenport P, Chan M, de Dombal FT. Twelve years' experience of computer-aided diagnosis in a district general hospital. Annals of the Royal College of Surgeons of England. 1990;72(2):140-6.

21. Beede E, Baylor EE, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, Vardoulakis LM. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.

22. Cruz NP, Canales L, Muñoz JG, Pérez B, Arnott I. Improving Adherence to Clinical Pathways Through Natural Language Processing on Electronic Medical Records...The 17th World Congress of Medical and Health Informatics, 25-30 August 2019, Lyon, France. Studies in Health Technology & Informatics. 2019;264:561-5.

23. Frymoyer A, Schwenk HT, Zorn Y, Bio L, Moss JD, Chasmawala B, et al. Model-Informed Precision Dosing of Vancomycin in Hospitalized Children: Implementation and Adoption at an Academic Children's Hospital. Frontiers in Pharmacology.11.

24. Ginestra JC, Giannini HM, Schweickert WD, Meadows L, Lynch MJ, Pavan K, et al. Clinician Perception of a Machine Learning-Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock. Critical Care Medicine. 2019;47(11):1477-84.

25. Henkel M, Horn T, Leboutte F, Trotsenko P, Dugas SG, Sutter SU, et al. Initial experience with AI Pathway Companion: Evaluation of dashboard-enhanced clinical decision making in prostate cancer screening. PloS one. 2022;17(7):e0271183.

26. Henry KE, Kornfield R, Sridharan A, Linton RC, Groh C, Wang T, et al. Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning

system. NPJ Digital Medicine. 2022;5(1):1-6.

27. Jauk S, Kramer D, Avian A, Berghold A, Leodolter W, Schulz S. Technology Acceptance of a Machine Learning Algorithm Predicting Delirium in a Clinical Setting: a Mixed-Methods Study. Journal of Medical Systems. 2021;45(4).

28. Jauk S, Kumar Veeranki SP, Kramer D, HÖGler S, MÜHlecker D, Eberhartl E, et al. External Validation of a Machine Learning Based Delirium Prediction Software in Clinical Routine...16th Annual Conference on Health Informatics meets Digital Health (dHealth 2022), May 24–25, 2022, Vienna, Austria. Studies in Health Technology & Informatics. 2022;293:93-100.

29. Jordan M, Hauser J, Cota S, Li H, Wolf L. The Impact of Cultural Embeddedness on the Implementation of an Artificial Intelligence Program at Triage: A Qualitative Study. Journal of Transcultural Nursing. 2023;34(1):32-9.

30. Joshi M, Mecklai K, Rozenblum R, Samal L. Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study. Jamia Open. 2022;5(2).

31. Kappen TH, van Loon K, Kappen MA, van Wolfswinkel L, Vergouwe Y, van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. J Clin Epidemiol. 2016;70:136-45.

32. Lebovitz S, Lifshitz-Assaf H, Levina N. To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. Organization Science.33(1):126-48.

33. Marwaha A, Chitayat D, Meyn MS, Mendoza-Londono R, Chad L. The point-of-care use of a facial phenotyping tool in the genetics clinic: Enhancing diagnosis and education with machine learning. American Journal of Medical Genetics Part A.185(4):1151-8.

34. Nehme F, Coronel E, Barringer DA, Romero LG, Shafi MA, Ross WA, Ge PS. Performance and attitudes toward real-time computer-aided polyp detection during colonoscopy in a large tertiary referral center in the United States. Gastrointestinal endoscopy. 2023;98(1):100-9.e6.

35. Rabinovich D, Mosquera C, Torrens P, Aineseder M, Benitez S. User Satisfaction with an AI System for Chest X-Ray Analysis Implemented in a Hospital's Emergency Setting...32nd Medical Informatics Europe Conference (MIE2022), 27-30 May, 2022, Nice, Fran. Studies in Health Technology & Informatics. 2022;294:8-12.

36. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. International Journal of Medical Informatics. 2020;137:N.PAG-N.PAG.

37. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study. J Med Internet Res. 2020;22(11):e22421.

38. Saunders S, Downar J, Subramaniam S, Embuldeniya G, van Walraven C, Wegier P. mHOMR: the acceptability of an automated mortality prediction model for timely identification of patients for palliative care. BMJ Quality & Safety. 2021;30(10):837-40.

39. Shiang T, Garwood E, Debenedectis CM. Artificial intelligence-based decision support system (AI-DSS) implementation in radiology residency: Introducing residents to AI in the clinical setting. Clin Imaging. 2022;92:32-7.

40. Singer SJ, Kellogg KC, Galper AB, Viola D. Enhancing the value to users of machine learning-based clinical decision support tools: A framework for iterative, collaborative development and implementation. Health Care Manage Rev. 2022;47(2):E21-e31.

41. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. European Radiology. 2020;30(10):5525-32.

42. Sun TQ, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly. 2019;36(2):368-83.

43. Wang SM, Hogg HDJ, Sangvai D, Patel MR, Weissler EH, Kellogg KC, et al. Development and Integration of Machine Learning Algorithm to Identify Peripheral Arterial Disease: Multistakeholder Qualitative Study. JMIR Form Res. 2023;7:e43963.

44. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Annals of Operations Research. 2020;294(1):567-92.

45. Shinners L, Aggar C, Grace S, Smith S. Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: An integrative review. Health Informatics Journal.26(2):1225-36.

46. Wu C, Xu H, Bai D, Chen X, Gao J, Jiang X. Public perceptions on the application of artificial intelligence in healthcare: a qualitative meta-synthesis. BMJ Open. 2023;13(1):e066322.

47. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Netw Open. 2022;5(9):e2233946.

48. Jungmann F, Jorg T, Hahn F, Pinto Dos Santos D, Jungmann SM, Düber C, et al. Attitudes Toward Artificial Intelligence Among Radiologists, IT Specialists, and Industry. Acad Radiol. 2021;28(6):834-40.

49. Wigfield A, Eccles JS. Expectancy–Value Theory of Achievement Motivation. Contemporary Educational Psychology. 2000;25(1):68-81.

50. Chao CM. Factors determining the behavioral intention to use mobile learning: An application and extension of the UTAUT model. Frontiers in Psychology. 2019;10(JULY).

51. Choudhury A, Asan O, Medow JE. Effect of risk, expectancy, and trust on clinicians'

intent to use an artificial intelligence system -- Blood Utilization Calculator. Appl Ergon. 2022;101:103708.

52. Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J. Implementing AI in healthcare-the relevance of trust: a scoping review. Front Health Serv. 2023;3:1211150.

53. van der Scharr M MN. van der Scharr Lab. 2021. [cited 2022]. Available from: https://www.vanderschaar-lab.com/making-machine-learning-interpretable-a-dialog-with-clinicians/.

54. Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. npj Digital Medicine. 2019;2(1):62.

55. Chin-Yee B, Upshur R. Three problems with big data and artificial intelligence in medicine. Perspectives in Biology and Medicine. 2019;62(2):237-56.

56. Amann J, Vetter D, Blomberg SN, Christensen HC, Coffee M, Gerke S, et al. To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. PLOS Digital Health. 2022;1(2):e0000016.

57. Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, Mooney C. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. Applied Sciences. 2021;11(11):5088.

58. Pierce RL, Van Biesen W, Van Cauwenberge D, Decruyenaere J, Sterckx S. Explainability in medicine in an era of AI-based clinical decision support systems. Front Genet. 2022;13:903600.

59. d'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Family Medicine and Community Health. 2022;10(Suppl 1):e001670.

60. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digital Health. 2023;2(6):e0000278.

61. Ganapathi S, Palmer J, Alderman JE, Calvert M, Espinoza C, Gath J, et al. Tackling bias in AI health datasets through the STANDING Together initiative. Nature Medicine. 2022;28(11):2232-3.

62. Sbaffi L, Walton J, Blenkinsopp J, Walton G. Information Overload in Emergency Medicine Physicians: A Multisite Case Study Exploring the Causes, Impact, and Solutions in Four North England National Health Service Trusts. J Med Internet Res. 2020;22(7):e19126.

63. Van Dort BA, Zheng WY, Sundar V, Baysari MT. Optimizing clinical decision support alerts in electronic medical records: a systematic review of reported strategies adopted by hospitals. J Am Med Inform Assoc. 2021;28(1):177-83.

64. Čartolovni A, Tomičić A, Lazić Mosler E. Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. Int J Med Inform. 2022;161:104738.

65. Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics. 2020;47(12):e3.

66. Chen W, O'Bryan CM, Gorham G, Howard K, Balasubramanya B, Coffey P, et al. Barriers and enablers to implementing and using clinical decision support systems for chronic diseases: a qualitative systematic review and meta-aggregation. Implementation Science Communications. 2022;3(1):81.

67. Nix M, Onisiforou G, Painter A. Understanding healthcare workers' confidence in AI. In: England NALHE, editor. UK2022.

68. Hogg HDJ, Al-Zubaidy M, Talks J, Denniston AK, Kelly CJ, Malawana J, et al. Stakeholder Perspectives of Clinical Artificial Intelligence Implementation: Systematic Review of Qualitative Evidence. J Med Internet Res. 2023;25:e39742.

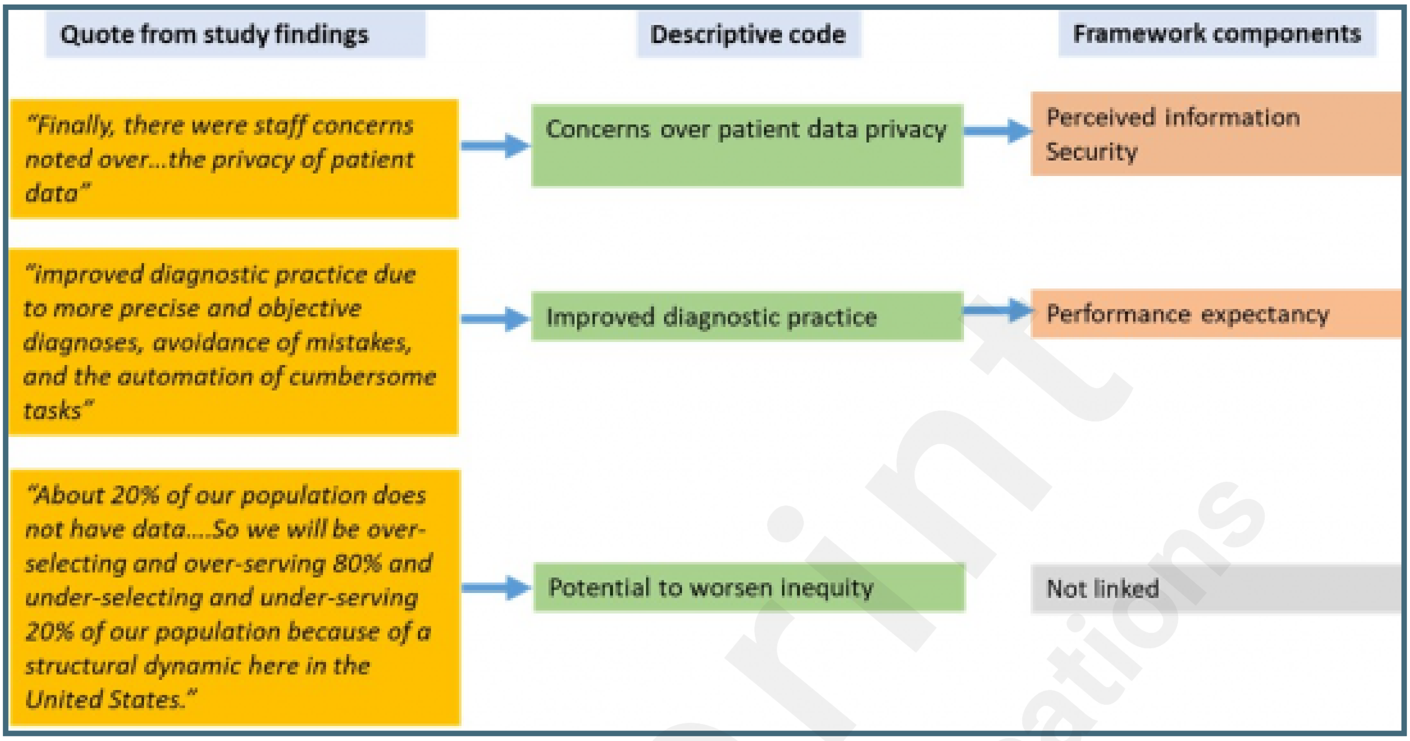
69. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021;8(2):e188-e94.

70. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, et al. A path for translation of machine learning products into healthcare delivery. EMJ Innovations 2020.

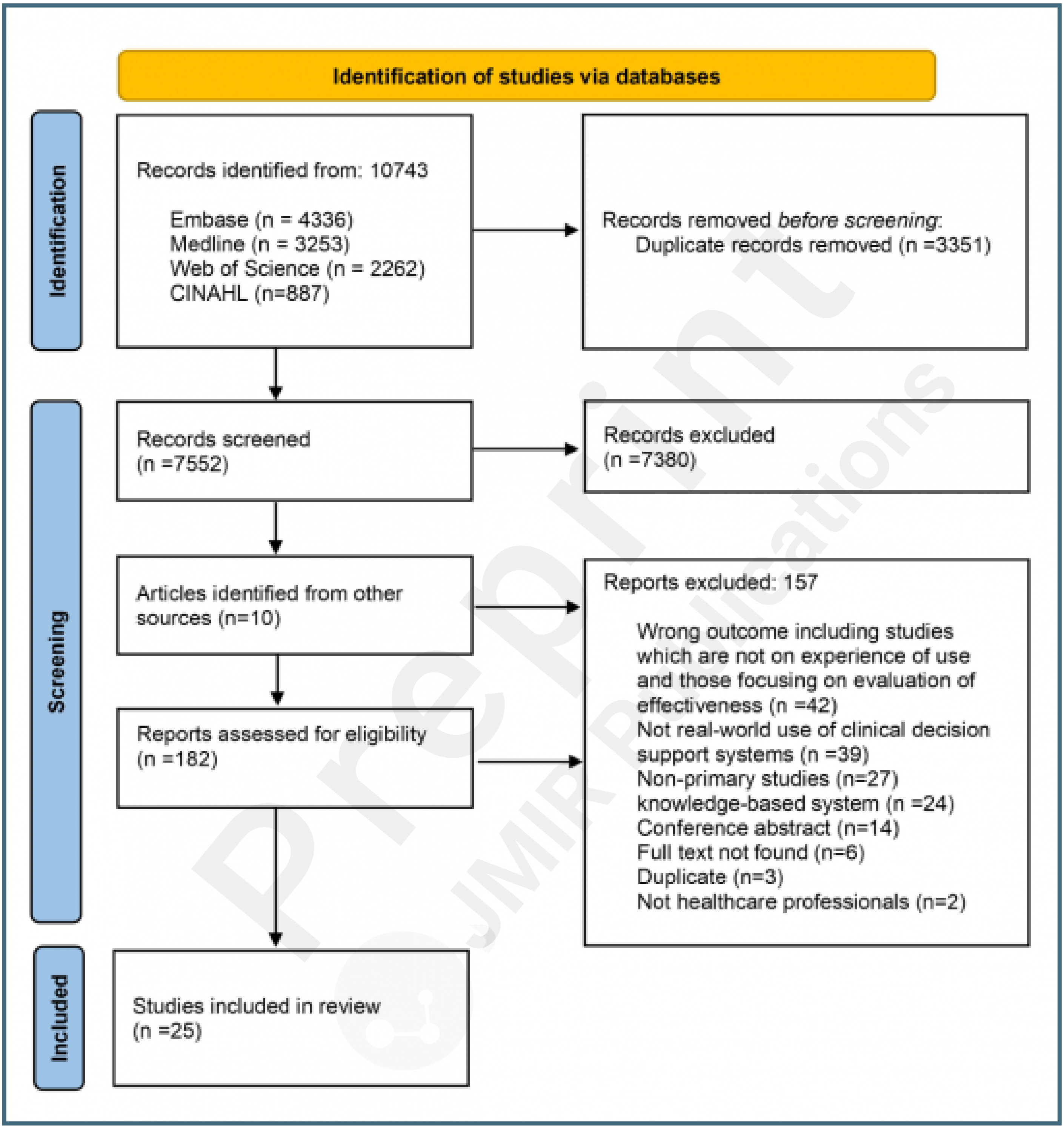
Supplementary Files

Figures

Schematic illustration of the data coding process.



PRISMA flow chart of study selection.



Multimedia Appendixes

Search strategy for all databases.
URL: http://asset.jmir.pub/assets/fd6193342157bd18a88589b6ddd87f59.docx

Data extraction, quality assessment and list of excluded studies.
URL: http://asset.jmir.pub/assets/c81cacfa2d8f9869fe5a74d839597890.xlsx

CONSORT (or other) checklists

PRISMA Checklist.
URL: http://asset.jmir.pub/assets/ee7780d71923b26d4f36b42dcf9e8769.pdf