BMJ Open

Implemented machine learning tools to inform decision-making for patient care in hospital settings: a scoping review

Andrea C Tricco ,[1,2] Areej Hezam,[1] Amanda Parker,[1] Vera Nincic,[1] Charmalee Harris,[1] Orna Fennelly,[3] Sonia M Thomas,[1] Marco Ghassemi,[1] Jessie McGowan,[4] P Alison Paprica ,[5] Sharon E Straus[1,6]

To cite: Tricco AC, Hezam A, Parker A, *et al*. Implemented machine learning tools to inform decision-making for patient care in hospital settings: a scoping review. *BMJ Open* 2023;**13**:e065845. doi:10.1136/bmjopen-2022-065845

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-065845).

Received 22 June 2022
Accepted 26 January 2023

For numbered affiliations see end of article.

Correspondence to
Dr Andrea C Tricco;
Andrea.tricco@unityhealth.to

ABSTRACT

Objectives To identify ML tools in hospital settings and how they were implemented to inform decision-making for patient care through a scoping review. We investigated the following research questions: What ML interventions have been used to inform decision-making for patient care in hospital settings? What strategies have been used to implement these ML interventions?

Design A scoping review was undertaken. MEDLINE, Embase, Cochrane Central Register of Controlled Trials (CENTRAL) and the Cochrane Database of Systematic Reviews (CDSR) were searched from 2009 until June 2021. Two reviewers screened titles and abstracts, full-text articles, and charted data independently. Conflicts were resolved by another reviewer. Data were summarised descriptively using simple content analysis.

Setting Hospital setting.

Participant Any type of clinician caring for any type of patient.

Intervention Machine learning tools used by clinicians to inform decision-making for patient care, such as AI-based computerised decision support systems or "'model-based'" decision support systems.

Primary and secondary outcome measures Patient and study characteristics, as well as intervention characteristics including the type of machine learning tool, implementation strategies, target population. Equity issues were examined with PROGRESS-PLUS criteria.

Results After screening 17 386 citations and 3474 full-text articles, 20 unique studies and 1 companion report were included. The included articles totalled 82 656 patients and 915 clinicians. Seven studies reported gender and four studies reported PROGRESS-PLUS criteria (race, health insurance, rural/urban). Common implementation strategies for the tools were clinician reminders that integrated ML predictions (44.4%), facilitated relay of clinical information (17.8%) and staff education (15.6%). Common barriers to successful implementation of ML tools were time (11.1%) and reliability (11.1%), and common facilitators were time/efficiency (13.6%) and perceived usefulness (13.6%).

Conclusions We found limited evidence related to the implementation of ML tools to assist clinicians with patient healthcare decisions in hospital settings. Future research should examine other approaches to integrating ML into hospital clinician decisions related to patient care, and report on PROGRESS-PLUS items.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ To our knowledge, this will be the first scoping review on the implementation of machine learning tools to inform decision-making for patient care in hospital settings.
⇒ Our search was limited to 2009 onwards; however, this allowed us to capture more recent/relevant machine learning tools.
⇒ A comprehensive search in multiple databases along with evidence found in published and grey literature sources allowed us to extensively map the evidence on the implementation of machine learning tools to inform decision-making for patient care in hospital settings.
⇒ Due to the large scope of the original scoping review question, we had to limit it to machine learning versus all artificial intelligence tools and inpatient hospital settings versus all settings.
⇒ In the coding of the interventions, we excluded any implementation strategies that were employed before the start of the study (not part of the machine learning intervention).

Funding Canadian Institutes of Health Research (CIHR) Foundation grant awarded to SES and the CIHR Strategy for Patient Oriented-Research Initiative (GSR-154442).

Scoping review registration https://osf.io/e2mna.

BACKGROUND

Artificial intelligence (AI) techniques have gained popularity within healthcare in recent years.[1–4] AI techniques consist of automated systems requiring 'intelligence' to perform tasks. Machine learning is an AI method that 'refers to the process of developing systems with the ability to learn from and make predictions using data'.[5 6] The use of AI in healthcare can transform clinical practice by providing aid to clinicians when interpreting data that are complex and diverse, allowing for support in clinical decision-making.

There are various ways that machine learning can be used to support clinical decision-making. One way is to assist with clinical tasks related to assessing, managing, and evaluating clinical issues and procedures.[7] Another way is to assist with epidemiological tasks, such as predicting the health needs and outcomes of specific people.[7] Machine learning can also be used for clinical administrative tasks.

Several systematic reviews have examined the use of machine learning for clinical decision-making focused on specific tasks, such as stroke and risk stratification,[1] preterm birth prediction,[2] and predicting radiation-induced neurocognitive decline.[8] However, much less focus has been on how to implement machine learning methods in hospital settings. For example, a recent scoping review found very few clinical decision support systems that used machine learning were implemented in the hospital setting.[9] This is imperative, as the successful translation of machine learning into hospital systems practice may help to improve the performance of clinical decisions related to diagnosis, prognostics and management, while saving time and improving patient outcomes.

A recent scoping review was identified that examined prognostic machine learning algorithms in paediatric chronic respiratory conditions.[10] Twenty-five studies were included and only two were implemented in a clinical setting. Furthermore, none of the included studies explicitly reported results pertaining to implementation of the machine learning algorithms.

Implementation science and practice ensures that research results are transferred and used by key knowledge users[11] and has the potential to reduce research waste.[12] An examination of the effectiveness of implemented machine learning tools is required to understand which (if any) machine learning tools have been successfully implemented in hospital settings to support decisions related to patient care and how (if at all) implementation science strategies were used to implement those machine learning tools. This examination will facilitate appropriate AI use, as well as enhance return on investment for machine learning tools. We aimed to determine strategies that have been used to implement machine learning tools to inform decision-making for patient care in hospital settings through a scoping review.

METHODS

Patient and public involvement

There was no patient or public involvement in this research.

Protocol and registration

The protocol for this scoping review was registered with Open Science Framework[13] and developed in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) Statement for protocols.[14] The JBI (formerly Joanna Briggs Institute) guidance for scoping reviews[15] informed the conduct of this scoping review. The knowledge users on the team included the Physician-in-Chief at St Michael's Hospital (SES) and former Vice President responsible for health at the Vector Institute (PAP). The knowledge users were engaged in all aspects of the review conduct. The results are reported using the PRISMA extension for scoping reviews,[16] supplemented by the updated PRISMA 2020 statement.[17]

Search strategy and selection criteria

Search strategy

An experienced librarian (JM) developed a comprehensive literature search strategy, which was peer reviewed by a second information specialist using the Peer Review of Electronic Search Strategies checklist.[18]

The following databases were searched: MEDLINE (2009–7 June 2021), Embase (2009–7 June 2021), CENTRAL (2009–7 June 2019) and the Cochrane Database of Systematic Reviews (2005–7 June 2019). A broad approach to the search question was taken to include AI terms broader than machine learning. The search was translated from the primary database MEDLINE to the other databases. All search strategies can be found in online supplemental appendix 1. Grey literature (i.e., unpublished or difficult-to-locate studies) was searched using guidance from the Canadian Agency for Drugs and Technologies in Health's Grey Matters checklist.[19] All sources for grey literature searching are available in online supplemental appendix 2. The references of all included studies and relevant reviews were scanned to identify additional potentially relevant studies for inclusion.

Eligibility criteria

The eligibility criteria are outlined according to the Population, Concept, Context mnemonic,[15] as follows:

Population

Any type of clinician caring for any type of patient.

Concept

Machine learning tools used by hospital clinicians to inform decision-making for patient care, such as AI-based computerised decision support systems or 'model-based' decision support systems were included.[20] Machine learning was defined as methods using mathematical operations to process input data, resulting in a prediction.[6] Machine learning used for decision support was defined as algorithms used to provide some form of input into human decision-making.[8] Machine learning tools used for automation without any input from the clinician were excluded. For example, machine learning tools used to predict patients at higher risk of a particular outcome without any decisions or interventions required by the clinician were excluded. Machine learning tools used for robotics (e.g., robotic surgery), interpretation of imaging such as the CT scan (if not being used to inform decision-making; for example, improving accuracy of an algorithm), medical devices (e.g., a device that monitors glucose levels and administers insulin automatically) and automatic transcribing of a clinical note for medical records were excluded. For studies that reported the use of a clinical decision support tool but did not explicitly report the use of machine learning, additional citations were identified by scanning the references to verify that machine learning was indeed used. All validation studies developing, testing or validating a machine learning model with hospital data were excluded if the model was not implemented for patient care decision-making. Only decision support tools using machine learning for clinical tasks defined as 'tasks generally performed by qualified healthcare providers related to the assessment, intervention and evaluation of health-related issues and procedures' or epidemiological tasks specified as 'tasks related to more accurately identifying the health needs and outcomes of people within a given population' were included.[8] Machine learning used for operational tasks defined as 'tasks related to activities that are ancillary to clinical tasks but necessary or valuable in the delivery of services (generally more administrative)' was excluded.[8] Studies that did not report implementation strategies for the machine learning interventions that were used for patient care were excluded.

Context

Only hospital settings were included. If a study was conducted outside of the hospital but used inpatient hospital data, then it was included.

Other criteria

Eligible study designs were primary research studies of experimental (e.g., randomised controlled trials, non-randomised controlled trials), quasi-experimental (e.g., controlled before and after studies, interrupted time series), observational (e.g., cohort studies, case-control studies, cross-sectional studies), qualitative (e.g., phenomenological, ethnography, qualitative interview) and mixed-methods (e.g., convergent parallel, embedded, explanatory sequential) design, with or without a comparator group. Studies published before 2009 were excluded to focus on the most recent evidence, along with non-English studies to increase feasibility for this project. No restrictions based on study duration were applied.

Study selection

The eligibility criteria were pilot tested by the team using 50 unique citations until >60% agreement was achieved (four training exercises with 18% agreement, 58% agreement, 30% agreement and 64% agreement). Agreement was calculated by taking the percentage of the responses of all team members to the 50 unique citations. Subsequently, the remaining titles and abstracts were screened independently by reviewers (AH, AP, VN, CH, OF, SMT, MG) working in pairs. For full-text screening, pilot testing on 20 studies was completed with 20% agreement observed across the team. The criteria were clarified, and the full-text articles were screened independently by two reviewers (AH, AP, VN, CH, OF, SMT, MG) working in pairs. All discrepancies were resolved by a third reviewer to ensure inter-rater reliability and quality checking of the screening that was completed. A clinician (SES) and methodologist (ACT) confirmed the final eligibility of all included studies. The full screening criteria for level 1 and level 2 screening are provided in online supplemental appendices 3 and 4.

Data charting

A standardised charting form was developed to chart study characteristics, population characteristics, intervention characteristics and type of outcome measure from each included article. Equity issues were abstracted using the PROGRESS-PLUS criteria.[21] The specific outcome results were not abstracted, as recommended in the JBI guide.[15] A pilot study was conducted with the team prior to charting until sufficient agreement (>60%) was achieved. All data were charted independently by two reviewers (AP, VN, OF, MG) working in pairs. All discrepancies were resolved by a third reviewer to ensure inter-rater reliability and quality checking of the screening that was completed. The data charting form is provided in online supplemental appendix 5.

Risk of bias appraisal

As recommended in the JBI guide,[15] a risk of bias appraisal was not conducted.

Analysis and presentation of results

All findings are summarised descriptively using summary tables, figures and appendices. To code the implementation strategies, the modified Effective Practice and Organisation of Care (EPOC) classification was used[22]; descriptions of categories can be found in online supplemental appendix 6. A clinician (SES) and methodologist (ACT) coded all included studies using the EPOC classification independently. Implementation barriers and facilitators were coded by one reviewer (VN) using a pre-existing framework.[23] For patient, clinician and system-level outcomes, a pre-existing framework[24 25] was also used by one reviewer (VN). All coding was conducted using simple content analysis.

Role of the funding source

The study sponsor had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

RESULTS

Literature search results

After screening 15 306 citations from the database search and 2080 citations from the grey literature search, 3474 full-text articles were obtained and screened for inclusion. Subsequently, 20 unique studies and 1 companion report[26] were included (figure 1). One included study was only available as a conference abstract.[27] A list of the studies that were very close to fulfilling the eligibility criteria but were eventually excluded is provided in online supplemental appendix 7.

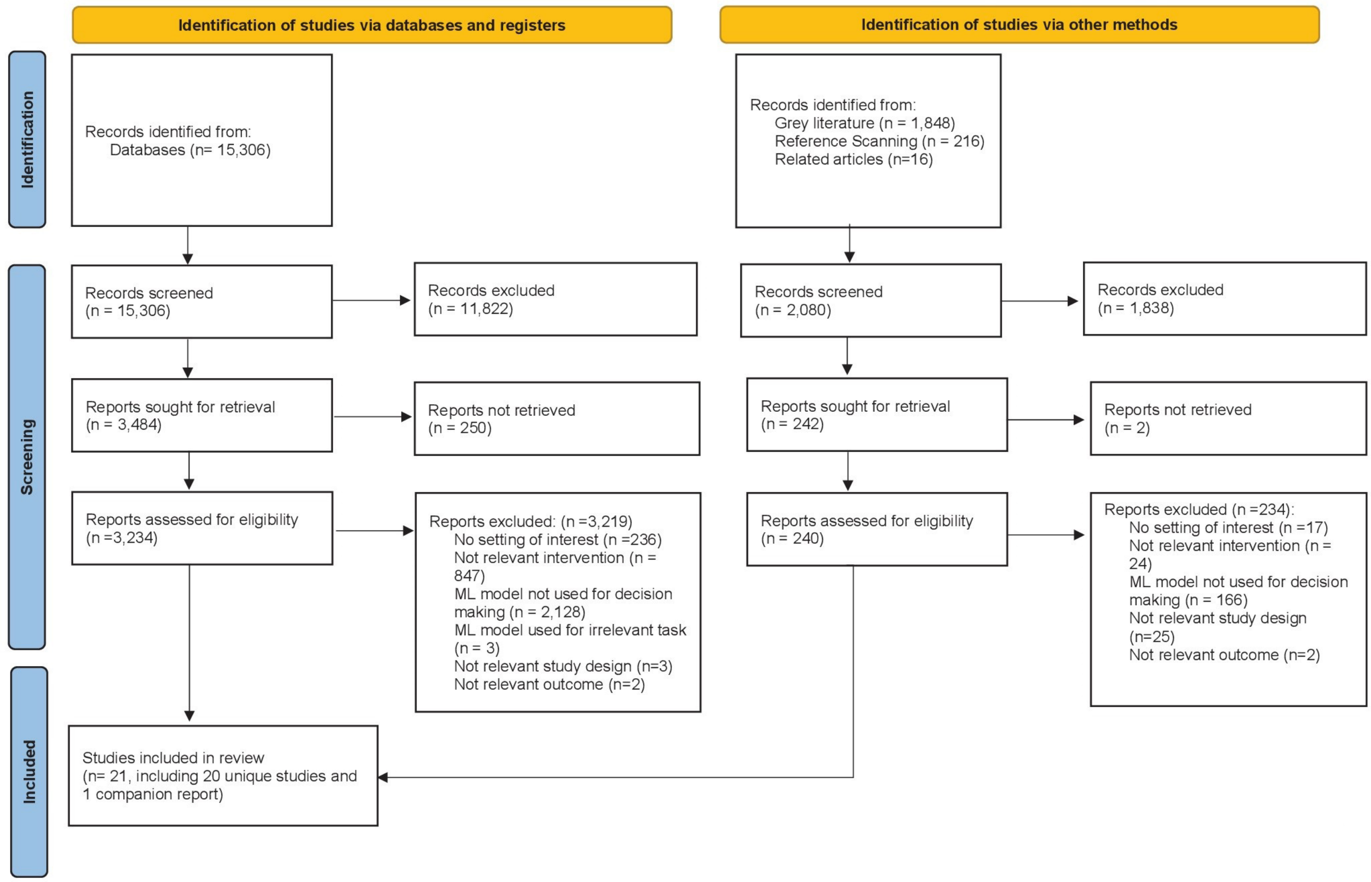

Figure 1 PRISMA study flow diagram. ML, machine learning; PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analysis.

Patient characteristics

The 20 studies included a total of 82 656 patients, with an average of 5514 patients per study (table 1 and online supplemental appendix 8). The studies included a total of 915 clinicians, with an average of 153 clinicians per study. Almost half of the patients were female (48.8%), whereas most clinicians were female (64.4%). Among studies that reported age, patients were mostly 62–72 years of age (n=3, 15.0%); this information was not reported in 60.0% of the studies. Most studies included adults at risk of infection (n=4, 20.0%; online supplemental appendix 8). Most studies did not report comorbidities (n=18, 47.4%) and when they did, the most frequent were infection (n=5, 13.2%) and congestive heart disease (n=3, 7.9%). The type of clinician involved in implementing machine learning tools was most commonly a physician (n=3, 30.0%) or nurse (n=3, 30.0%). Regarding the PROGRESS-PLUS criteria (online supplemental appendix 9), seven studies reported patient or clinician sex[27–33] and four studies reported additional PROGRESS-PLUS criteria (race, type of health insurance, rural/urban).[27 28 34 35] The demographic variables were only calculated for studies that reported them.

Study characteristics

The 20 included studies were published between 2009 and 2021, with 75.0% published 2017 onwards (table 2 and online supplemental appendix 10). Of the 20 included studies, 16 were applying algorithms which had already been trained and validated to an intervention, and 4 studies trained, validated and applied algorithm(s) as part of the same paper. North America (n=10, 50.0%) followed by Europe (n=6, 30%) were the most common continents. Most studies were cohort studies (n=8, 40.0%) followed by randomised trials (n=6, 30.0%). Most studies were conducted within 1 year (n=10, 50.0%). All studies were based on inpatient hospital data with the majority based out of a single hospital (n=14, 70.0%).

Intervention characteristics

The machine learning tools used were supervised learning (n=14, 70.0%), unsupervised learning (n=2, 10.0%) and deep learning (n=4, 20.0%) (table 3 and online supplemental appendix 11). All studies reported implementation strategies with the most common being clinician reminders (n=20, 44.4%), facilitated relay of clinical information (n=8, 17.8%) and staff education (n=7, 15.6%) (table 3 and online supplemental appendix 12). The target population of implementation strategies was most commonly healthcare providers (n=40, 88.9%).

Outcome characteristics

The outcomes reported were at the clinician level in 12 studies (table 4 and online supplemental appendix 13),

 Tricco AC, *et al*. *BMJ Open* 2023;**13**:e065845. doi:10.1136/bmjopen-2022-065845

Table 1 Summary of patient characteristics

Characteristics	Number (%)
Patient characteristics (n=20 studies)	
Total # of patients	82 656
Mean number of patients (range)	5514 (55–22 641)
Total # of clinicians	915
Mean number of clinicians (range)	152.5 (25–358)
Mean % female—patients (range)	48.8 (42–53.9)
Mean % female—clinicians (range)	64.4 (58.6–70.2)
Age (mean/median)	
≤50 years	2 (10.0)
51–61 years	2 (10.0)
62–72 years	3 (15.0)
>73 years	1 (5.0)
Not reported	12 (60.0)
Population—type	
Patients	14 (70.0)
Clinicians	3 (15.0)
Both	3 (15.0)
Type of clinician (out of 10)*	
Physician	3 (30.0)
Nurse	3 (30.0)
Pharmacist	1 (10.0)
Interns and residents	1 (10.0)
Physician assistants	1 (10.0)
Not reported	1 (10.0)
Comorbidities (out of 38)*	
Infection	5 (13.2)
Congestive heart disease	3 (7.9)
Chronic pulmonary disease	2 (5.3)
Diabetes	2 (5.3)
Other (eg, drug and alcohol abuse, other diseases)	2 (5.3)
Cerebrovascular disease	1 (2.6)
Myocardial infarction	1 (2.6)
Peripheral vascular disease	1 (2.6)
Kidney/renal disease	1 (2.6)
Solid tumour	1 (2.6)
Hypertension	1 (2.6)
Not reported	18 (47.4)

*Multiple categories reported per study.

Table 2 Summary of study characteristics

Characteristics	Number (%)
Study characteristics (n=20 studies)	
Year of publication	
2009–2016	5 (25.0)
2017–2021	15 (75.0)
Geographical region	
North America	10* (50.0)
Europe	6 (30.0)
Asia	2 (10.0)
Middle East	1 (5.0)
South America	1 (5.0)
Study design	
Cohort	8 (40.0)
RCT	6 (30.0)
Uncontrolled before–after	4 (20.0)
Cross-sectional	1 (5.0)
Mixed-methods design	1 (5.0)
Study duration	
<1 year	10 (50.0)
≥1 year	8 (40.0)
NA†	1 (5.0)
Not reported	1 (5.0)
Setting	
Single site	14 (70.0)
Multisite	6 (30.0)
# of study arms	
2	17 (85.0)
1	3 (15.0)

*Corresponding author's country was used as a proxy for one study as country of conduct was not reported.
†Cross-sectional study so study duration is NA.
NA, not applicable; RCT, randomised clinical trial.

patient level in 10 studies (online supplemental appendix 14) and health system level in 10 studies (online supplemental appendix 15). At the clinician level, the majority of the outcomes were focused on perception and satisfaction of clinicians (n=44, 71.0%), whereas at the patient level, they were mostly focused on physiological or clinical (n=46, 71.9%) outcomes. At the health system level, the most common outcomes focused on delivery of care (n=18, 45.0%). The most commonly reported implementation barriers were that the machine learning tool was perceived as being time consuming (n=3, 11.1%) and unreliable (n=3, 11.1%). In contrast, the most commonly reported implementation facilitators were that it improved time/efficiency (n=3, 13.6%) and was perceived as being useful (n=3, 13.6%).

DISCUSSION

We conducted a comprehensive scoping review on implementation strategies for machine learning tools within hospital settings. We identified only 20 studies that fulfilled our eligibility criteria. All examined implementation strategies. However, only 10 studies reported on the barriers and 14 studies reported on the facilitators to implementing machine learning tools. Across the studies, the most common implementation strategies were clinician reminders, facilitated relay of clinical information and staff education. The most common barriers were

Table 3 Summary of intervention characteristics

Characteristics	Number (%)
Intervention characteristics	
Type of machine learning in intervention arms	
Supervised learning	14 (70.0)
Deep learning	4 (20.0)
Unsupervised learning	2 (10.0)
Implementation strategies in intervention arms (out of 45)*	
Clinician reminders	20 (44.4)
Facilitated relay of info to clinicians	8 (17.8)
Staff education	7 (15.6)
Team changes	3 (6.7)
Audit and feedback	3 (6.7)
Patient education	2 (4.4)
Continuous quality improvement	1 (2.2)
Promotion of self-management	1 (2.2)
Target population for intervention (out of 45)*	
Healthcare providers	40 (88.9)
Patients/clinicians	3 (6.7)
Health system	2 (4.4)

*Multiple categories reported per study.

that the machine learning tool took time to use and was perceived as unreliable, whereas the most common facilitators were that the tool improved time or efficiency and was perceived as useful.

Our results identified several gaps in the literature. Most of the studies were conducted in high-income countries. This is understandable as machine learning tools are expensive to develop. The majority of the studies were focused on adults at risk of infection. Most of the studies reported the use of supervised machine learning tools. Few studies examined intervention strategies that have been found to be effective, such as audit and feedback.[36] A recent scoping review confirms our results with no studies reporting implementation strategies for machine learning algorithms in paediatric chronic respiratory conditions.[11]

Only nine[27–35] of the included studies provided data on any of the PROGRESS-PLUS criteria; most of the studies examined outcomes at the patient level. Examining equity in the use of AI tools is important, as some evidence suggests that machine learning algorithms can increase bias within health[37–39] through propagation of existing racial discrepancies and inequalities in socioeconomic status, gender, religion, sexual orientation or disability. This in turn further exacerbates health inequities. Future research should examine equity in relation to machine learning tools.

AI is a multibillion dollar industry, with substantial investment within health.[40] Our results suggest that very few studies are examining best strategies to implement these AI tools. Future primary research on the implementation of machine

Table 4 Summary of outcome characteristics

Characteristics	Number (%)
Outcome characteristics	
Level of outcome reported (out of 32)*	
Clinician	12 (37.5)
Patient	10 (31.3)
Health system	10 (31.3)
Clinician outcomes (out of 62)*	
Perceptions/satisfaction	44 (71.0)
Behaviours/compliance	18 (29.0)
Patient outcomes (out of 64)*	
Physiological/clinical	46 (71.9)
Delivery of care	10 (15.6)
Mortality	6 (9.4)
Life impact	2 (3.1)
Health system outcomes (out of 40)*	
Delivery of care	18 (45.0)
Costs	12 (30.0)
Hospital resource use	10 (25.0)
Implementation barriers (out of 27)*	
Time consuming	3 (11.1)
Unreliable technology	3 (11.1)
Increased work	2 (7.4)
Lack of customisable features	2 (7.4)
Requires reminders as support	2 (7.4)
Lack of integration into existing workflows	2 (7.4)
Lack of transparency	2 (7.4)
Diagnostic misinterpretation	2 (7.4)
Perceived lack of usefulness	2 (7.4)
Increased resources	1 (3.7)
Additional education required for implementation	1 (3.7)
Lack of training	1 (3.7)
Over-reliance on ML	1 (3.7)
Requires leadership support	1 (3.7)
High costs/patient insurance required	1 (3.7)
Untrusting attitude toward ML	1 (3.7)
Implementation facilitators (out of 22)*	
Improved time/efficiency of care delivery	3 (13.6)
Perceived usefulness	3 (13.6)
Organisational/leadership support	2 (9.1)
Suggestion and improvement request	2 (9.1)
Alerts and reminders as support	1 (4.5)
Cost-effectiveness	1 (4.5)
Involvement of clinicians into implementation	1 (4.5)
Enhanced communication among clinicians/care teams	1 (4.5)
Integration into existing workflows	1 (4.5)
Low alert burden and low false-alarm rate	1 (4.5)
Not increasing workload	1 (4.5)

Continued

Table 4 Continued	
Characteristics	**Number (%)**
Visual readability	1 (4.5)
Easiness of use	1 (4.5)
Timely output	1 (4.5)
Use of audit and feedback as support	1 (4.5)
Targeting specific goals	1 (4.5)

*Multiple categories reported per study.
ML, machine learning.

learning tool in hospital settings should be conducted to broaden the evidence base, including the effective implementation interventions on acceptability, appropriateness and feasibility. A future systematic review can be conducted to examine the effectiveness of the various implementation strategies to optimise the use of AI in health. This will ensure that this enormous investment is not wasted. It will also facilitate appropriate AI use within health. It is important to note that the European Union has proposed an Artificial Intelligence Act,[41] which will likely regulate sociotechnical system, and may have implications on AI implementation in healthcare settings.

Limitations

Our scoping review has limitations worth noting. Due to the large scope of the original scoping review question, we had to limit it to machine learning versus all AI tools and inpatient hospital settings versus all settings (online supplemental appendix 16). Additionally, we limited our search to 2009 onwards; however, this allowed us to capture more recent/relevant machine learning tools as technology has advanced in the last two decades as compared with older studies. We also limited our search to the English language, and studies conducted in countries where English is not the first language could have been excluded. Another potential limitation is that any implementation strategies that were employed before the start of the study (not part of the machine learning intervention) were not included in the coding of interventions.

CONCLUSIONS

In conclusion, there is a lack of evidence on implementation strategies used for machine learning tools in hospital settings. This is an urgent area of research prioritisation, given the millions of dollars invested in AI technologies within health. Future studies can report on inequities using the PROGRESS-PLUS framework. A systematic review identifying effective implementation strategies of machine learning tools to inform decision-making for patient care within hospitals would be very useful for future implementation efforts.

Author affiliations
[1]Knowledge Translation Program, St Michael's Hospital Li Ka Shing Knowledge Institute, Toronto, Ontario, Canada
[2]Epidemiology Division and Institute of Health Policy, Management and Evaluation, University of Toronto Dalla Lana School of Public Health, Toronto, Ontario, Canada
[3]Irish Centre for High End Computing (ICHEC), National University of Ireland Galway, Galway, Ireland
[4]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
[5]Institute for Health Policy, Management and Evaluation, University of Toronto Dalla Lana School of Public Health, Toronto, Ontario, Canada
[6]Department of Medicine, University of Toronto, Toronto, Ontario, Canada

Acknowledgements We thank Dr Muhammad Mamdani for providing feedback on our protocol and screening criteria, Raman Brar, Chantal Williams, Elizabeth McCarron, Jane Pearson Sharpe, and Naveeta Ramkissoon for screening citations or full-text articles, Nazia Darvesh for assisting with the protocol and screening questions, Alissa Epworth for executing the literature searches, Raymond Daniel for searching for errata and retractions, and Faryal Khan for formatting the manuscript and creating the EndNote library.

Contributors ACT interpreted the results, drafted the manuscript, and provided methodological and technical expertise. AH coordinated the review, screened citations and full-text articles, cleaned and prepared the results, revised and edited the manuscript. AP supported coordination of the review, screened citations and full-text articles, and charted data. VN, CH, OF and MG screened citations and full-text articles, and charted data. VN coded data on implementation and outcomes. SMT worked on the protocol and screened studies. JM created the literature and grey literature search strategies. PAP and SES helped conceive the study, provided methodological and content expertise throughout the project. All authors confirm that they had full access to all the data in the study and accept responsibility to submit for publication. ACT is responsible for the overall content as the guarantor and accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding This project was funded by the CIHR Foundation grant awarded to SES (grant number not available) and the CIHR SPOR initiative (GSR-154442). ACT is funded by a Tier 2 Canada Research Chair in Knowledge Synthesis (17-0126-AWA). OF was in part supported by the Health Research Board (Ireland) and the HSC Public Health Agency (CBES-2018-001) through Evidence Synthesis Ireland and Cochrane Ireland. SES is funded by a Tier 1 Canada Research Chair in Knowledge Translation (17-0245-SUB).

Disclaimer The funder of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval Ethics approval is not applicable, as this is a scoping review of pre-existing studies.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplemental information.

ORCID iDs
Andrea C Tricco http://orcid.org/0000-0002-4114-8971
P Alison Paprica http://orcid.org/0000-0001-6362-7087

REFERENCES

1 Suri JS, Paul S, Maindarkar MA, *et al*. Cardiovascular/stroke risk stratification in Parkinson's disease patients using atherosclerosis pathway and artificial intelligence paradigm: a systematic review. *Metabolites* 2022;12:312.
2 Sharifi-Heris Z, Laitala J, Airola A, *et al*. Machine learning approach for preterm birth prediction using health records: systematic review. *JMIR Med Inform* 2022;10:e33875.
3 Yang X, Mu D, Peng H, *et al*. Research and application of artificial intelligence based on electronic health records of patients with cancer: systematic review. *JMIR Med Inform* 2022;10:e33799.
4 Girolami I, Pantanowitz L, Marletta S, *et al*. Artificial intelligence applications for pre-implantation kidney biopsy pathology practice: a systematic review. *J Nephrol* 2022;35:1801–8.
5 Liu Y, Chen P-H, Krause J, *et al*. How to read articles that use machine learning: users' guides to the medical literature. *JAMA* 2019;322:1806–16.
6 Senders JT, Staples PC, Karhade AV, *et al*. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg* 2018;109:476–86.
7 Shaw J, Rudzicz F, Jamieson T, *et al*. Artificial intelligence and the implementation challenge. *J Med Internet Res* 2019;21:e13659.
8 Tohidinezhad F, Perri DD, Zegers CML, *et al*. Prediction models for radiation-induced neurocognitive decline in adult patients with primary or secondary brain tumors: A systematic review. *In Review* [Preprint] 2021.
9 Schwartz JM, Moy AJ, Rossetti SC, *et al*. Clinician involvement in research on machine learning-based predictive clinical decision support for the hospital setting: a scoping review. *J Am Med Inform Assoc* 2021;28:653–63.
10 Filipow N, Main E, Sebire NJ, *et al*. Implementation of prognostic machine learning algorithms in paediatric chronic respiratory conditions: a scoping review. *BMJ Open Respir Res* 2022;9:e001165.
11 Graham ID, Logan J, Harrison MB, *et al*. Lost in knowledge translation: time for a MAP? *J Contin Educ Health Prof* 2006;26:13–24.
12 Chan A-W, Song F, Vickers A, *et al*. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257–66.
13 Open science framework. implementation of artificial intelligence in healthcare. 2022. Available: https://osf.io/e2mna
14 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.
15 Peters MDJ, Marnie C, Tricco AC, *et al*. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth* 2020;18:2119–26.
16 Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for scoping reviews (PRISMA-scr): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
17 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:71.
18 McGowan J, Sampson M, Salzwedel DM, *et al*. PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology* 2016;75:S0895-4356(16)00058-5:40–6.:.
19 Grey matters: a practical tool for searching health-related grey literature [Ottawa: CADTH]. 2018. Available: https://www.cadth.ca/resources/finding-evidence
20 Cresswell K, Callaghan M, Khan S, *et al*. Investigating the use of data-driven artificial intelligence in computerised decision support systems for health and social care: a systematic review. *Health Informatics J* 2020;26:2138–47.
21 PROGRESS-Plus. Cochrane methods equity. 2022. Available: https://methods.cochrane.org/equity/projects/evidence-equity/progress-plus
22 Tricco AC, Thomas SM, Veroniki AA, *et al*. Quality improvement strategies to prevent falls in older adults: a systematic review and network meta-analysis. *Age Ageing* 2019;48:337–46.
23 Dodd S, Clarke M, Becker L, *et al*. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol* 2018;96:84–92.
24 Verma AA, Murray J, Greiner R, *et al*. Implementing machine learning in medicine. *CMAJ* 2021;193:E1351–7.
25 Whitelaw S, Pellegrini DM, Mamas MA, *et al*. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. *Eur Heart J Digit Health* 2021;2:62–74.
26 Burdick H, Pino E, Gabel-Comeau D, *et al*. Evaluating a sepsis prediction machine learning algorithm in the emergency department and intensive care unit: a before and after comparative study. *Clinical Trials* [Preprint].
27 Hassan AE, Ringheanu VM, Preston L, *et al*. Abstract P248: CSC implementation of artificial intelligence software significantly improves door-in to groin puncture time interval and recanalization rates. *Stroke* 2021;52(Suppl_1):Suppl
28 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;27:815–9.
29 Arboe B, Laub RR, Kronborg G, *et al*. Evaluation of the decision support system for antimicrobial treatment, treat, in an acute medical ward of a university hospital. *Int J Infect Dis* 2014;29:156–61.
30 Kofoed K, Zalounina A, Andersen O, *et al*. Performance of the treat decision support system in an environment with a low prevalence of resistant pathogens. *J Antimicrob Chemother* 2009;63:400–4.
31 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;45:48.
32 Repici A, Spadaccini M, Antonelli G, *et al*. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. *Gut* 2022;71:757–65.
33 Wijnberge M, Geerts BF, Hol L, *et al*. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the hype randomized clinical trial. *JAMA* 2020;323:1052–60.
34 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:e100109.
35 Dexheimer JW, Abramo TJ, Arnold DH, *et al*. Implementation and evaluation of an integrated computerized asthma management system in a pediatric emergency department: a randomized clinical trial. *Int J Med Inform* 2014;83:805–13.
36 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;2012:CD000259.
37 Angwin J, Larson J, Mattu S, *et al*. Machine bias. *ProPublica* 2016.
38 Barocas S, Selbst AD. n.d. Big data's disparate impact. *SSRN Journal*
39 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
40 Bresnick J. Artificial intelligence in healthcare spending to hit $36B. health IT analytics. 2022. Available: https://healthitanalytics.com/news/artificial-intelligence-in-healthcare-spending-to-hit-36b
41 The Artificial Intelligence Act. 2022. Available: https://artificialintelligenceact.eu/