Using artificial intelligence methods for systematic review in health sciences: a systematic review. [Appendices]

BLAIZOT, A., VEETTIL, S.K., SAIDOUNG, P., MORENO-GARCIA, C.F., WIRATUNGA, N., ACEVES-MARTINS, M., LAI, N.M. and CHAIYAKUNAPRUK, N.

2022

Online Supplementary file

The use of artificial intelligence methods for systematic reviews in health sciences: A systematic review

Aymeric Blaizot, Sajesh K. Veettil, Pantakarn Saidoung, Carlos Francisco Moreno-Garcia, Nirmalie Wiratunga, Magaly Aceves-Martins, Nai Ming Lai, Nathorn Chaiyakunapruk

Appendix 1

Appendix Table 1.1 Search strategy in Medline

Ovid MEDLINE(R) and In-Process, In-Data-Review & Other Non-Indexed Citations and Daily 1946 to April 22, 2021

# No	Search	Results
1	exp Artificial Intelligence/	110887
2	(artificial adj2 intelligence).ti,ab.	11071
3	AI.ti,ab.	29542
4	exp Machine Learning/	26270
5	(machine adj2 learning).ti,ab.	38324
6	(deep adj2 learning).ti,ab.	15481
7	exp Robotics/	30303
8	robot$.ti,ab.	50127
9	exp Neural Networks, Computer/	32551
10	(neural adj2 network$).ti,ab.	58083
11	meta-analysis.pt.	130037
12	systematic review.pt.	150849
13	or/1-10	230182
14	or/11-12	217022
15	13 and 14	**1831**

Appendix Table 1.1 Additional Search strategy in Medline
Ovid MEDLINE(R) 1946 to May 14, 2021

# No	Search	Results
1	ASreview.tw.	0
2	Abstrackr.tw.	10
3	DistillerSR.tw.	18
4	(SWIFT-review or SWIFT-Active).tw.	6
5	Rayyan.tw.	37
6	Colandr.tw.	1
7	RobotReviewer.tw.	6
8	RCT tagger.tw.	0
9	(National Centre for Text Mining or NACTeM).tw.	2
10	RobotAnalyst.tw.	3
11	(ExaCT Adj3 extract$).tw.	65
12	Lingo3G.tw.	0
13	GAPscreener.tw.	1
14	Trial2Rev.tw.	0
15	Systematic review.pt.	150884
16	Or/1-14	141
17	15 and 16	35

Appendix Table 1.2 Inclusion and Exclusion criteria

Inclusion Criteria	Exclusion Criteria	Identified Tools
-All systematic reviews with or without meta-analysis that was aided by any AI as part of their methodology. - Any form of AI method, including machine learning, deep learning, neural network, or any other application that are used to enable full or semi-autonomous performance of one or more stages in the development of evidence synthesis. This includes rapid review, umbrella review, evidence gap map, evidence mapping, and scoping review.	- Studies which used any tools for data management only (eg. COVIDENCE). - Studies which are protocols for systematic reviews. - Studies which analyze effectiveness of systematic review software. -Studies not published in English	ASreview Abstrackr SWIFT-review/ SWIFT-Active Screener EPPI-reviewer Rayyan RobotReviewer RCT tagger National Centre for Text Mining (NaCTeM) tools RobotAnalyst ExaCT Lingo3G GAPscreener Trial2Rev

Appendix 2 Quality Assessment

Appendix Table 2.1 AMSTAR-2 Evaluation

Evaluations	**Russell Viner 2021**(1)	**M.J. Giummarra 2020**(2)	**Goldkuhle M. 2019**(3)	**Pinna 2021**(4)	**Gaskins 2020**(5)	**Riley 2020**(6)	**Siqueira 2020**(7)	**Nascimento 2021**(8)	**Xiong 2021**(9)
Overall appraisal	Moderate	Moderate	Moderate	Moderate	Critically Low	High	Critically Low	Critically Low	Low
Did the research questions and inclusion criteria for the review include the components of PICO?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did the report of the review contain an explicit statement that the review methods were established before the conduct of the review, and did the report justify any significant deviations from the protocol?	Yes	Yes	Yes	No	Yes	Yes	No	Partial Yes	No
Did the review authors explain their selection of the study designs for inclusion in the review?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did the review authors use a comprehensive literature search strategy?	Yes	Yes	Yes	Yes	Yes	Yes	Partial Yes	Partial Yes	No
Did the review authors perform study selection in duplicate?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did the review authors perform data extraction in duplicate?	Yes	Yes	Yes	No	Yes	Yes	No	No	No
Did the review authors provide a list of excluded studies and justify the exclusions?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did the review authors describe the included studies in adequate detail?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
Did the review authors report on the sources of funding for the studies included in the review?	No	No	No	No	No	No	No	No	No
If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	NA	NA	NA	NA	NA	Yes	NA	NA	Yes
If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	NA	NA	NA	NA	NA	Yes	NA	NA	Yes

Did the review authors account for RoB in individual studies when interpreting/ discussing the review results?	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
Did the review authors provide a satisfactory explanation for, and discuss, any heterogeneity observed in the review results?	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes

Note: AMSTAR-2 is designed to evaluate systematic reviews. Therefore, it would not be appropriate to evaluate rapid reviews as they were not designed for those standards.

Appendix 3 Additional Information of Included Studies

Appendix Table 3.1 Additional Information of Included Studies

Author, Year, Country and Design	Category: Research Question	Tool	Stage of Process that AI involved	Number of articles that went through AI	Method of Validation	Description of Use	Reported Concerns	Human interventions required when AI performed not as expected
Russell Viner, 2021(1) UK, Italy, USA SR	School closures on Physical & mental health of children	EPPI-Reviewer 4	T&A screening	No RCTs identified. Will be applied if future trials were identified	Decisions about inclusion were independently reassessed by the senior authors (SH/ JP). But not validated the excluded articles.	16778 articles identified. The ML algorithm was trained on the first 1500 articles and then a classifier model built to rank subsequent studies and identify a threshold below which studies were highly likely to not be relevant.	NA	NA
M.J. Giummarra, 2020(10) Australia, UK SR	The association between fault or blame-related attributions and procedures after transport injury and health and work-related outcomes	Abstrackr	T&A screening	1,157	Any disagreements regarding eligibility were resolved through discussion, and consultation with the other authors. Any disagreements regarding eligibility were resolved through discussion, and consultation with the other authors.	1 independent reviewer manually + 1 reviewer coded citations for relevance until no further studies were predicted to be relevant.	the risk of missing a very small number of relevant studies; Manual T&A screening identfied more studies for full-text screening; a separate evaluation of the study methods suggest that the methods were not detrimental	NA

		Wordstat and QDA Miner	Full-text screening	200		1 independent reviewer manually + 1 reviewer used text mining for initial idenfication of relevant articles for full-text screening follwed by manual screening.		NA
Goldkuhle M, 2018(3) Germany, Austria, US Rapid review	Nivolumab for adults with Hodgkin's lymphoma	Robot Reviewer	Data extraction	1,260	The two review authors resolved any discrepancies by discussion; had they not reached consensus, they planned to consult a third review author, but this was not necessary. If required, they would have contacted the authors of specific studies for supplementary information. No RCTs were identified	2 indepentdently extracted data and compared result with RobotReviewer.	The software recognised correctly that one study was not an RCT and therefore did not extract any data of this study. However, the two remaining studies were falsely labelled to be RCTs.	Yes, "Although our search did not identify any eligible RCTs, we uploaded available full-texts of the three included studies into the software RobotReviewer. The software recognised correctly that one study was not an RCT and therefore did not extract any data of this study. However, the two remaining studies were falsely labelled to be RCTs. The extraction results on both studies could not be used in the further review process. Yet some characteristics of included trials were given sufficiently. This is especially concerning the data regarding included participants and study interventions. We did not consider the 'Risk of bias' function of the software, since it was based on the criteria of the Cochrane 'Risk

								of bias' tool for RCTs only."
			ROB assessment	101	NA	1 independent reviewer manually + 1 reviewer used RobotReviewer and compared results.	NA	Yes. Since the tool is designed for RCTs it was not feasible for the review. Hence, reviewers had to do this step.
Pinna, 2021(4) Italy, Canada SR	The Impact of Alexithymia on Treatment Response in Psychiatric Disorders	Rayyan	T&A screening	495	After uploading, screening of the literature was performed in blind by two investigators (MM and PP). Disagreement between reviewers was resolved by joint discussion with a third senior investigator (FP). The quality of evidence was assessed using the Newcastle Ottawa Scale (NOS)	Used to expedite T&A screening	NA	NA
Gaskins, 2020(5) UK SR	Factors influencing implementation of aerobic exercise after stroke	Rayyan	Post-protocol screening	4,177	Reviewer re-screened relevant articles given by Rayyan	Rayyan autonomously retrospectively screened for relevant articles.	NA	NA
Riley, 2020(6) US	Interventions for increasing colorectal cancer	Rayyan	T&A screening	2,267	NA	Used to expedite T&A screening	NA	NA

MA	screening uptake among African-American men							
Siqueira, 2020(7) Brazil Integrative Review	Use of serious games for health students to learn about cardiopulmonary resuscitation	Rayyan	T&A screening	297	During the second stage of the selection, it was verified whether the studies selected by the researchers were the same	Used to expedite T&A screening	NA	NA
Nascimento, 2021(8) Brazil Integrative Review	The use of Bloom's taxonomy for developing competence in nursing professionals and students in clinical simulation.	Rayyan	T&A screening	16,941	The studies were selected by two professionals, independently, who read titles and abstracts, applied the Rayyan review, which eliminates duplicate articles and facilitates triage,	Used to expedite T&A screening	NA	NA
Xiong, 2021(9) New Zealand, China, Canada MA	The Relative Risk of Atrial Fibrillation in Patients With Diabetes Mellitus	K-means clustering algorithm	T&A screening	No RCTs identified. Will be applied if future trials were identified	Conducted a duplicate manual screening	The algorithm was trained on a set of relevant studies. Then, from 4177 articles, it made 14 clusters. Cluster #5, containing 416 articles, was the most similar to the studies it was trained on.	NA	NA
Lam, 2019(11) US	low-calorie sweeteners (LCS) with respect to	SWIFT-Active	T&A screening	1,157	NA	500 Abstracts were screened in duplicate to calibrate the reviewers. Single-screening was then employed until	NA	NA

rEM(rapid Evidence Mapping)	human dietary exposures and health outcomes.	Screener				SWIFT-Active Screener estimated that 95% of relevant articles had been included.		
		SWIFT-Review	Data extraction	200	NA	SWIFT-Review was used for data-extraction, though information on length and sample size categories had to collected manually. SWIFT-Review was also used to help generate the evidence map.	SWIFT-Review could not automate all aspects of data extraction. A human intervention was required to manually extract study sample size and review of automated tagging for each category as SWIFT-Review did not perform those tasks effectively.	"This greatly reduced the amount of required manual extraction of data from the articles, although some aspects were still manual, such as the manual data extraction was required for study sample size and review of automated tagging for each category."
Deng, 2019(12) US MA	Validation of a semiautomated natural language processing based procedure for quantifying the risk of cancer associated with pathogenic mutations in germline	Semi-automated natural language processing	Abstract classification and filtering and text mining	1,260	Conducted a duplicate manual screening	Used NLP for abstract review and filtration with human review employed in between steps. Reference retrieval was done manually.	NLP missed one paper out of ten critical to the review	NA

	cancer susceptibility							
Aali, 2020(13) UK Scoping Review	Post-stroke Fatigue	Robot Reviewer	ROB assessment	8	One of the reviewers (GA) also double-checked and revised RobotReviewer's assessment and corrected the data where necessary.	1 reviewer double-checked and modified RobotReviewer's assement of ROB of the included studies.	NA	NA

Appendix 4 General Characteristics AI tools

Appendix Table 4.1 General Characteristics of tools used

Name	Used by	Description	Stage of Review	ML	NLP	Availability	Comments
Abstrackr(14) http://abstrackr.cebm.brown.edu/account/login	M.J. Giummarra	Abstrackr helps you upload and organize the results of a literature search for a systematic review. It also makes it possible for your team to screen, organize, and manipulate all of your abstracts in one place.	Abstract screening	Yes	No	Yes	Records are uploaded and screened as "relevant","borderline", or "irrelevant" by the reviewer. The reviewer can also tag terms that are indicative of their relevance or irrelevance.
SWIFT-review/ SWIFT-Active Screener(15) https://www.sciome.com/swift-review/	Lam	“Sciome Workbench for Interactive computer-Facilitated Text-mining” is a freely available interactive workbench which provides numerous tools to assist with problem formulation and literature prioritization.	search, categorize, and prioritize large (or small) bodies of literature in an interactive manner	Yes	Text Mining	Yes	Technically it is machine learning, although it appears that the developers have compiled a "dictionary" of search strategies that ca ben downloaded and plugged in into the tool to perform the review.
EPPI-reviewer(16) https://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4&	Russell Viner	application for all types of literature review, including systematic reviews, meta-analyses, 'narrative' reviews and meta-ethnographies	manages references, stores PDF files and facilitates qualitative and quantitative analyses such as meta-analysis and thematic synthesis	Yes	No	No, EPPI-Reviewer fees are based on a subscription model	It appears to be more of a reference manager. The scientific paper describing the tool is not accessible.
Rayyan(17) https://www.rayyan.ai/	Pinna ,Gaskins, Riley, Siqueira, Nascimento	Free web and mobile app	Screening	Yes	Yes	Free to start	The tool also contains a graph visualization of the interactions between papers
RobotReviewer(18)	Goldkuhle M. Aali	Automatic extraction of data from clinical trial reports	Data extraction ('PICO', study design, and whether there	Yes	Yes	Yes, even downloadable from Github (open source)	Uses state of the art NLP (Word embeddings, BERT, etc.) and machine learning (SVM, CNN) methods

https://www.robotreviewer.net/			is a risk of bias)				
K-means clustering(19) N/A	Xiong	Method of vector quantization that computes centroids and iterates until it finds the optimal centroid	Screening	Yes	No	NA	Implemented by reviewers for data mining
Natural Language Processing(12) https://github.com/YujiaBao/PubmedClassifier	Deng	Automatically retrieves abstracts and applies a classifier	Screening	Yes	Yes	Yes from Github	Developed and implemented by reviewers

Appendix 5 Characteristics of Crossingham *et al.*

Appendix Table 5.1 Crossingham *et al.*

Author, Year, Country and Design	Category and Health Science Area	Tool	Stage of Process that AI involved	ML	NLP	Method of use	Description of Methods	Validation	Advantages
Crossingham, 2021(20) UK, Australia SR	Effectiveness Review Respiratory	Cochrane RCT Classifier	Identifying Randomized controlled trials	Yes	No	Human in the loop	Used the Cochrane RCT Classifier as part of the Screen4me methodology to alleviate the work of excluding non-RCTs.	No	NA

References

1. Viner R, Russell S, Saulle R, Croker H, Stansfeld C, Packer J, et al. Impacts of school closures on physical and mental health of children and young people: a systematic review. medRxiv. 2021 Feb 12;2021.02.10.21251526.

2. Giummarra MJ, Lau G, Gabbe BJ. Evaluation of text mining to reduce screening workload for injury-focused systematic reviews. Inj Prev J Int Soc Child Adolesc Inj Prev. 2020 Feb;26(1):55–60.

3. Goldkuhle M, Dimaki M, Gartlehner G, Monsef I, Dahm P, Glossmann J, et al. Nivolumab for adults with Hodgkin's lymphoma (a rapid review using the software RobotReviewer). Cochrane Database Syst Rev. 2018 Jul 12;2018(7):CD012556.

4. Pinna F, Manchia M, Paribello P, Carpiniello B. The Impact of Alexithymia on Treatment Response in Psychiatric Disorders: A Systematic Review. Front Psychiatry [Internet]. 2020 [cited 2021 Jul 18];0. Available from: https://www.frontiersin.org/articles/10.3389/fpsyt.2020.00311/full

5. Gaskins NJ, Bray E, Hill JE, Doherty PJ, Harrison A, Connell LA. Factors influencing implementation of aerobic exercise after stroke: a systematic review. Disabil Rehabil. 2019 Dec 25;0(0):1–15.

6. Rogers CR, Matthews P, Xu L, Boucher K, Riley C, Huntington M, et al. Interventions for increasing colorectal cancer screening uptake among African-American men: A systematic review and meta-analysis. PLOS ONE. 2020 Sep 16;15(9):e0238354.

7. Siqueira TV, Nascimento J da SG, Oliveira JLG de, Regino D da SG, Dalri MCB. The use of serious games as an innovative educational strategy for learning cardiopulmonary resuscitation: an integrative review. Rev Gaucha Enferm. 2020;41:e20190293.

8. Nascimento J da SG, Siqueira TV, Oliveira JLG de, Alves MG, Regino D da SG, Dalri MCB. Development of clinical competence in nursing in simulation: the perspective of Bloom's taxonomy. Rev Bras Enferm [Internet]. 2021 Mar 24 [cited 2021 Jul 18];74. Available from: http://www.scielo.br/j/reben/a/zgmY8gmZF3Q98JrxzLdCLrC/?lang=en

9. Xiong Z, Liu T, Tse G, Gong M, Gladding PA, Smaill BH, et al. A Machine Learning Aided Systematic Review and Meta-Analysis of the Relative Risk of Atrial Fibrillation in Patients With Diabetes Mellitus. Front Physiol [Internet]. 2018 [cited 2021 Jul 18];0. Available from: https://www.frontiersin.org/articles/10.3389/fphys.2018.00835/full

10. Giummarra MJ, Lau G, Grant G, Gabbe BJ. A systematic review of the association between fault or blame-related attributions and procedures after transport injury and health and work-related outcomes. Accid Anal Prev. 2020 Feb 1;135:105333.

11. Lam J, Howard BE, Thayer K, Shah RR. Low-calorie sweeteners and health outcomes: A demonstration of rapid evidence mapping (rEM). Environ Int. 2019 Feb 1;123:451–8.

12. Deng Z, Yin K, Bao Y, Armengol VD, Wang C, Tiwari A, et al. Validation of a Semiautomated Natural Language Processing-Based Procedure for Meta-Analysis of Cancer Susceptibility Gene Penetrance. JCO Clin Cancer Inform. 2019 Aug;3:1–9.

13. Aali G, Drummond A, das Nair R, Shokraneh F. Post-stroke fatigue: a scoping review. F1000Research. 2020 Aug 25;9:242.

14. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. Syst Rev. 2018 Mar 12;7(1):45.

15. Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, et al. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int. 2020 May 1;138:105623.

16. Park SE, Thomas J. Evidence synthesis software. BMJ Evid-Based Med. 2018 Aug;23(4):140–1.

17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016 Dec 5;5(1):210.

18. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide. Res Synth Methods. 2018;9(4):602–14.

19. Kalra M, Lal N, Qamar S. K-Mean Clustering Algorithm Approach for Data Mining of Heterogeneous Data. In: Mishra DK, Nayak MK, Joshi A, editors. Information and Communication Technology for Sustainable Development. Singapore: Springer; 2018. p. 61–70. (Lecture Notes in Networks and Systems).

20. Crossingham I, Turner S, Ramakrishnan S, Fries A, Gowell M, Yasmin F, et al. Combination fixed-dose beta agonist and steroid inhaler as required for adults or children with mild asthma. Cochrane Database Syst Rev [Internet]. 2021 [cited 2021 Nov 24];(5). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013518.pub2/full