**APPENDIX A: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist**

| **SECTION** | **ITEM** | **PRISMA-ScR CHECKLIST ITEM** | **REPORTED ON PAGE #** |
| --- | --- | --- | --- |
| **TITLE** | | | |
| Title | 1 | Identify the report as a scoping review. | 1 |
| **ABSTRACT** | | | |
| Structured summary | 2 | Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives. | 3 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach. | 5,6 |
| Objectives | 4 | Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives. | 6 |
| **METHODS** | | | |
| Protocol and registration | 5 | Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number. | 6 |
| Eligibility criteria | 6 | Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale. | 6,7 |
| Information sources\* | 7 | Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed. | 6 |
| Search | 8 | Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated. | S4-S11 |
| Selection of sources of evidence† | 9 | State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review. | 7 |
| Data charting process‡ | 10 | Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators. | 7,8 |
| Data items | 11 | List and define all variables for which data were sought and any assumptions and simplifications made. | 7,8, Table 1S (Appendix C) |
| Critical appraisal of individual sources of evidence§ | 12 | If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate). | NA |
| Synthesis of results | 13 | Describe the methods of handling and summarizing the data that were charted. | 7,8 |
| **RESULTS** | | | |
| Selection of sources of evidence | 14 | Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram. | 8, Figure 1 |
| Characteristics of sources of evidence | 15 | For each source of evidence, present characteristics for which data were charted and provide the citations. | 8-10; Tables 1, 2S, 3S; Figures 2 – 7, 1S-4S |
| Critical appraisal within sources of evidence | 16 | If done, present data on critical appraisal of included sources of evidence (see item 12). | NA |
| Results of individual sources of evidence | 17 | For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives. | NA (too many studies to do in a meaningful way at this stage) |
| Synthesis of results | 18 | Summarize and/or present the charting results as they relate to the review questions and objectives. | 8-10; Tables 1, 2S, 3S; Figures 2 – 7, 1S-4S |
| **DISCUSSION** | | | |
| Summary of evidence | 19 | Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups. | 10-12 |
| Limitations | 20 | Discuss the limitations of the scoping review process. | 11 |
| Conclusions | 21 | Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps. | 11,12 |
| **FUNDING** | | | |
| Funding | 22 | Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review. | 12 |

JBI = Joanna Briggs Institute; PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.

\* Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites.

† A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote).

‡ The frameworks by Arksey and O’Malley (6) and Levac and colleagues (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting*.*

§The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of "risk of bias" (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document).

*From:* Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. ;169:467–473. doi: 10.7326/M18-0850

**APPENDIX B: SEARCH STRATEGIES**

**Databases used:**

1. Medline-OVID
2. EMBASE
3. Cinahl
4. Cochrane Library
5. Web of Science
6. Scopus
7. IEEE Xplore
8. ACM Digital Library
9. MathSciNet
10. AAAI (<https://aaai.org/ocs/index.php/index/index/search/advanced>)
11. arXiv

## Database Searching Notes

The databases listed above have different search capabilities in terms of keywords and subject headings. We used the most rigorous approach possible for each database, whereby approaches can be broken down into three general categories:

1. **Search with keywords and subject headings:**

* Medline-OVID
  + Keywords were used to search title, abstract, and author keywords.
* Embase
  + Keywords were used to search title, abstract, and author keywords.
* Cinahl
  + Keywords were used to search title and abstract.
* Cochrane
  + Keywords were used to search title, abstract, and keywords.
* ACM Digital Library
  + No wildcard (\*), use full spellings.
  + Only used artificial intelligence subject headings (CCS); health related headings are too broad and captured too many irrelevant documents to maintain review feasibility.
  + Search “The ACM Full-Text Collection.”

*Search strategy for category 1:*

1. *Keywords and subject headings for artificial intelligence concept were searched with OR.*
2. *Keywords and subject headings for primary care concept were searched with OR.*
3. *i) and ii) were combined with AND.*
4. **Search with keywords only:**

* Web of Science
  + Keywords in “Topic” field were used to search title, abstract, author keywords, keywords plus.
* Scopus
  + Keywords were used to search title, abstract, keywords.
* MathSciNet
  + Keywords in ‘Anywhere’ field were used to search author, author/related, title, review text, journal, institution code, series, MSC primary/secondary, MSC primary, MR number, reviewer
* arXiv
  + arXiv API was be accessed using python.
  + Keywords in ‘all’ field were used to search title, author, abstract, comment, journal reference, subject category, report number, id)

*Search strategy for category 2:*

1. *Keywords for artificial intelligence concept were searched with OR.*
2. *Keywords for primary care concept were searched with OR.*
3. *i) and ii) were combined with AND.*
4. **Search with limited keywords only:**

* IEEE Xplore
  + Limited to 12 keywords.
  + Keywords were used to search metadata (abstract, index terms, bibliographic citation data.)
  + No wildcard (\*) within phrase searching, so we wrote out the 12 terms in full.
* AAAI
  + Limited to 254 characters.
  + Used only primary care concept keywords because redundant to search artificial intelligence terms in artificial intelligence proceedings.
  + Case sensitive; spellede out keywords that are most important and most likely to be capitalized differentially by different authors using upper and lower case first letter(s) and spelled out less important or less capitalization-ambiguous keywords using only lower case.

*Search strategy for IEEE Xplore:*

1. *Keywords for artificial intelligence concept were searched with OR.*
2. *Keywords for primary care concept were searched with OR.*
3. *i) and ii) were combined with AND.*

*Search strategy for AAAI:*

1. *Keywords for primary care concept were searched with OR.*

**Exact strategies:**

Available upon request.

**APPENDIX C: METHODS AND RESULTS**

**Table 1S: Data Extraction Field Characterizations.**

|  |  |
| --- | --- |
| **Field** | **Definition and Subfields** |
| *Citation Information* | Last and first name of first author, year of publication, and title |
| *Study Purpose(s)* | Three mutually exclusive overall research purposes for the AI in the study:   1. *Method Development/Adaptation:*Research that created novel AI methods or modified existing AI methods to accomplish a task relevant to PC. For example, developing a new supervised machine learning algorithm to learn a model that will predict the probability of pathological heart murmurs using digital heart sound recording data (1). This category includes studies that compare the performance of AI methods to the performance of humans or that include consultation with end users to inform tool development, as this is considered part of model testing, which may lead to further modifications before evaluating performance in the setting that the AI is intended to support (e.g. clinical practice.) 2. *Data Analysis:* Existing AI methods were used to analyze and/or extract information from data. For example, using natural language processing algorithms to identify cases of familial hypercholesteremia from electronic health records (2). 3. *Evaluation:*Research that included AI implemented in its intended setting, possibly as part of a pilot study to assess impact or usability characteristics of a tool. For example, assessing the impact of a machine learning-derived diagnosis model on reducing cervical intraepithelial neoplasia overdiagnosis in a Dutch national population screening program (3). |
| *Author Appointment(s)* | Author affiliations as presented on the manuscript, divided into 16 categories: *Biological and Biomedical Sciences, Company, Computer Science, Engineering, Epidemiology and Biostatistics, Health Sciences, Informatics, Mathematics, Medicine – Unspecified, Medicine – Specialty, Family Medicine and Primary Care, Nursing, Public Health, Statistics, Other (specified),* and *Unknown* when not enough affiliation information was provided to identify a broad discipline. When an author had multiple affiliations, all were recorded. |
| *Primary Care Function(s)* | Nine categories of PC functions or tasks that the researched AI supported or is intended to support in the future:   1. *Diagnostic Decision Support*: AI provided information to inform diagnosis, such as the probability that a patient has a particular condition. 2. *Treatment Decision Support*: AI provided information to inform treatment decisions, whereby treatment was interpreted broadly to include any management or care provided (or absence of unnecessary actions) to someone with the health condition(s) or symptom(s) of interest. 3. *Referral Support*: AI provided information to support decisions about referring patients to specialist services or AI assisted with technical aspects of the referral process. 4. *Future State Prediction*: AI provided predictions towards future events, for example utilization of emergency department, development of a health condition, or prognosis for an existing condition. 5. *Health Care Utilization Analyses*: AI provided information about interactions with or processes within health care systems, for example frequency or quantity of patient visits. 6. *Knowledge Base and Ontology Construction or Use:* Construction or use of knowledge bases or ontologies including PC concepts. 7. *Information Extraction*: AI used to extract knowledge from structured or unstructured data (e.g. electronic medical records) for further use. 8. *Descriptive Information Provision*: AI used to summarize data in a meaningful way for human interpretation, for example prevalence of a condition or patterns of patient profiles. 9. *Other*: The PC function was not represented by the above categories; specifics were recorded. |
| *Author Reported Intended End-User(s)* | People who the research or research end-product was stated as intended for, regardless of whether those intended end users were involved with the research or how close the research was to being applicable for those users in practice setting: *Patient, Physician, Nurse, Nurse Practitioner, Administrator, Researcher, Other (specified), or Unknown*. If the study was developing a deployable AI method or tool (broadly defined) but more research was needed before the AI method of interest would be ready to implement or be utilized by its intended end user, *Researcher* was included as a target end user. |
| *Target Health Condition(s)* | The health condition of interest as stated by the study authors or inferred by reviewers, or *Unknown* if no condition was stated or inferable. Conditions were extracted in full form and MZ later organized them into 27 and 10 category formats. When a study intended for AI to be applicable for all health conditions “General” was used; specifics about any test conditions were also extracted. |
| *Location of Data Source(s) or intended location of implementation* | Country or next level of granularity where data were collected, or the location where study stated implementation would occur. *Unknown* was used when the location of data source was not stated or when all data were simulated. |
| *Subfield(s) of Artificial Intelligence* | Artificial Intelligence methods were organized according to 10 subfields; a single study may include one or more subfields:   1. *Bayesian Network:* Graphical models (directed acyclic graphs) used to describe dependency relationships among variables that enable the efficient representation of multivariate probability distributions. The resulting distributions can be queried to find the probability of an event occurring given a particular set of evidence. Bayesian networks can be developed manually, such as from physician input, learned from data, or created using a combination of the two. For example, Teles et al. (2015) use a Bayesian Network to assist the diagnosis of dengue fever disease. The model includes variables for dengue fever risk factors, such as ‘Respiratory Distress’ (4). For prediction, a person’s current risk factor variable values are inputted and the conditional probability they have dengue fever is outputted (4). 2. *Computer Vision:* Includes extracting visual information and understanding it. Computer vision is distinct from image processing, which includes modifying an existing or creating a new image without focusing on the meaning of the image. For example,Zouridakis et al. (2015) present a smartphone app whereby a picture of a skin lesion is taken and computer vision is used to interpret the image and assess the likelihood of malignancy (5). 3. *Data Mining:* The process of eliciting information from collections of data, such as by finding and counting pattern occurrences using inferential algorithms; humans may then interpret these patterns. For example, Soler et al. (2015) used data mining on electronic medical records to identify relationships between reasons for encounter and diagnoses recorded for the corresponding visit (6). We did not consider extracting information in a structured way, such as using a database query to get a basic count of disease X diagnoses, to be the type of data mining that falls under the umbrella of artificial intelligence. 4. *Expert System:* Consists of two parts: 1) a knowledge base that contains facts and rules, such as if-then statements derived from medical guidelines and 2) an inference engine that uses the knowledge base to arrive at conclusions or answers to questions. For example, Lange et al. (1997) demonstrate the use of an expert system called Iliad for teaching diagnostic reasoning to Nurse Practitioner students (7). Illiad’s knowledge base is made up of medical facts and relationships. Bayesian or probabilistic and Boolean or deterministic reasoning may be used with the knowledge base to arrive at a level of confidence about a diagnosis (7). 5. *Fuzzy Models:* Rely on fuzzy logic and fuzzy sets to represent problems with uncertainty. They are often used to provide more flexibility to outcomes instead of requiring strict classification into pre-defined groups. For example, Katigari et al. (2017) used a fuzzy model as the inference engine for an expert system designed to support diagnosis of diabetic neuropathy (8). Model input includes parameters such as time with diabetes, symptom severity, and laboratory blood test values; model output is an estimate of diabetic neuropathy severity (8). 6. *Natural Language Processing:* the ability to read language used by humans and interpret it in a meaningful way; this is often accomplished by analysing syntactic and semantic characteristics of language. The input language may be audio or written. For example,Koeling et al. (2011) used natural language processing to analyse free text portions of medical records and enhance the accuracy of ovarian cancer symptom detection compared to only using the structured portion of medical records (9). 7. *Robotics:* Robotics within artificial intelligence refers to machines that can act autonomously to navigate and alter their environment. A robot may rely on other types of artificial intelligence, such as computer vision and natural language processing, to accomplish this. Robotics outside of artificial intelligence include machines that are programmed by humans to perform a defined set of actions. No examples of robotics were captured by our review. 8. *Supervised Machine Learning:* involves an algorithm learning to associate labels with observations. In the context of health, the label is often an outcome, e.g. a disease state or outcome, and the observations are often patient variables. Labels may be numeric values or categorical classifications. Supervised learning uses existing labelled data which contain a collection of observations together with their correct label to produce a model that is able to assign a label to new, previously unseen observations. Supervised machine learning techniques include Support Vector Machines, K-Nearest Neighbours, Naïve Bayes Classifier, and Random Forest Decision Trees. For example,Cox et al. (2016) used supervised machine learning to help identify undiagnosed post-stroke spasticity (10). A model was trained using a large PC database that included the outcome of interest, post stroke spasticity events, and 72 candidate variables to predict the outcome, such as demographic information, prescriptions, and medical diagnoses (10). They then used the model to identify people who had a high probability of post-stroke spasticity and checked whether the event was recorded in their records; the results of this study suggested an under recording of post stroke spasticity in PC records (10). 9. *Unsupervised Machine Learning:* algorithms learn patterns from unlabelled data (unlabelled meaning there are not defined, known outcome categories as was the case for supervised machine learning). Common unsupervised machine learning techniques include clustering data items into groups based on their similarity, association mining to identify observations that tend to occur together, autoencoders to condense data while maintaining adequate fidelity, and feature separation to examine different aspects of a dataset independently. For example, Newcomer et al. (2011) used cluster analysis on data from a health care organization to identify groups of complex patients who may benefit from targeted care strategies (11). 10. *Other:* There are additional types of AI not captured by the above, such as multi-agent systems. We did not expect a high prevalence of these methods so did not create distinct categories, but recorded details when they arose. *Other* was also used to classify studies that did not focus on any specific technique of artificial intelligence. For example, Sola et al. (2018) studied physician perceptions of artificial intelligence in general without isolating any particular artificial techniques (12). |
| *Reviewer who extracted the data* | Initials of the person who (re)read the full text article and assigned values for the seven key characteristics outlined above: JKK, ALT, or DJL. |
| *Reviewer notes* | Optional free form notes from the person extracting the data. |

**Legend:** AI: Artificial Intelligence, PC: Primary Care.

**Notes:** Subfields are ordered according to appearance in the results section of the manuscript. Cited examples are from studies captured by our scoping review.

**References:**

1. Andrisevic N, Ejaz K, Rios-Gutierrez F, Alba-Flores R, Nordehn G, Burns S. Detection of heart murmurs using wavelet analysis and artificial neural networks. Journal of Biomechanical Engineering. 2005;127:899–904.

2. Safarova MS, Liu H, Kullo IJ. Rapid identification of familial hypercholesterolemia from electronic health records: The SEARCH study. Journal of Clinical Lipidology. 2016;10:1230–9.

3. Kok MR, Boon ME, Schreiner-Kok PG, Hermans J, Grobbee DE, Kok LP. Less medical intervention after sharp demarcation of grade 1-2 cervical intraepithelial neoplasia smears by neural network screening. Cancer. 2001;93:173–8.

4. Teles G, Oliveira C, Braga R, et al. Using Bayesian networks to improve the decision-making process in public health systems. In IEEE 16th International Conference on e-Health Networking, Applications and Services, Healthcom; 2014 October 15-18; Natal, Brazil. p. 565–70. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84921824566&doi=10.1109%2fHealthCom.2014.7001904&partnerID=40&md5=2eed896c7a56df2298339bdeb208af22

5. Zouridakis G, Wadhawan T, Situ N, et al. Melanoma and other skin lesion detection using smart handheld devices. Methods in molecular biology. 2015;1256:459–96.

6. Soler JK, Corrigan D, Kazienko P, et al. Evidence-based rules from family practice to inform family practice; The learning healthcare system case study on urinary tract infections. BMC Family Practice. 2015;16(1).

7. Lange LL, Haak SW, Lincoln MJ, et al. Use of Iliad to improve diagnostic performance of nurse practitioner students. Journal of Nursing Education. 1997;36(1):36–45.

8. Katigari MR, Ayatollahi H, Malek M, Haghighi MK. Fuzzy expert system for diagnosing diabetic neuropathy. World Journal of Diabetes. 2017;8(2):80–8.

9. Koeling R, Tate AR, Carroll JA. Automatically estimating the incidence of symptoms recorded in GP free text notes. In Proceedings of the first international workshop/Managing interoperability and complexity in health systems. Conference on Information and Knowledge Management; 2011; ACM, New York, NY. p. 43-49.

10. Cox AP, Raluy-Callado M, Wang M, Bakheit AM, Moore AP, Dinet J. Predictive analysis for identifying potentially undiagnosed post-stroke spasticity patients in United Kingdom. Journal of Biomedical Informatics. 2016;60:328–33.

11. Newcomer SR, Steiner JF, Bayliss EA. Identifying subgroups of complex patients with cluster analysis. American Journal of Managed Care. 2011;17(8):E324–32.

12. Sola D, Borioli GS, Quaglia R. Predicting GPs’ engagement with artificial intelligence. British Journal of Health Care Management. 2018;24(3):134–40.

**Table 2S: Complete author appointment counts.**

|  |  |
| --- | --- |
| **Appointment Type** | **Number of studies with at least one author with the corresponding appointment** |
| Biological and Medical Sciences | 29 |
| Company | 49 |
| Computer Science | 97 |
| Engineering | 71 |
| Epidemiology and Biostatistics | 23 |
| Health Sciences | 33 |
| Informatics | 63 |
| Mathematics | 16 |
| Medicine – Other | 94 |
| Medicine – Specialty | 99 |
| Medicine – Family or Primary Care | 57 |
| Nursing | 4 |
| Public Health | 20 |
| Statistics | 15 |
| Other | 132 |
| Unknown | 110 |

**Note:** Each study fulfills one or more appointment type categories; each category is counted a maximum of one time for any given study.

**Table 3S: Detailed breakdown of location**

|  |  |
| --- | --- |
| **Location** | **Number of Studies** |
| Unknown or Not Applicable | 113 |
| United States | 79 |
| United Kingdom | 38 |
| Netherlands | 26 |
| Australia | 17 |
| Canada | 13 |
| Spain | 10 |
| Brazil | 9 |
| India | 9 |
| Iran | 8 |
| Sweden | 8 |
| China | 6 |
| Germany | 6 |
| Italy | 5 |
| Slovenia | 4 |
| Australia and France | 3 |
| Croatia | 3 |
| France | 3 |
| Greece | 3 |
| New Zealand | 3 |
| Belgium | 2 |
| Egypt | 2 |
| Finland | 2 |
| Ireland | 2 |
| Japan | 2 |
| Norway | 2 |
| Singapore | 2 |
| Taiwan | 2 |
| Austria | 1 |
| Barcelona | 1 |
| Bulgaria | 1 |
| Canada and United States and United Kingdom and Brazil and Netherlands and Australia | 1 |
| Colombia | 1 |
| Czech Republic | 1 |
| Denmark | 1 |
| Europe | 1 |
| Germany and Norway | 1 |
| Greece and Bulgaria and Albania and Fyrom and Turkey | 1 |
| Hong Kong | 1 |
| Israel | 1 |
| Kuwait | 1 |
| Malaysia | 1 |
| Malta and Netherlands | 1 |
| Portugal | 1 |
| Saudi Arabia | 1 |
| South Africa | 1 |
| Switzerland | 1 |
| Turkey | 1 |
| United Kingdom and Greece and Germany | 1 |
| United States and Panama | 1 |
| United States and United Kingdom | 1 |

**APPENDIX D: INCLUDED STUDIES**

A complete list of the 405 included references is available upon request.