Artificial Intelligence and Primary Care Research: A Scoping Review

Jacqueline K. Kueper, MSc1
Amanda L. Terry, PhD2
Merrick Zwarenstein, MBChB, PhD3
Daniel J. Lizotte, PhD4
1Departments of Epidemiology & Biostatistics and Computer Science, Western University, London, Ontario, Canada
2Departments of Epidemiology & Biostatistics, Family Medicine, Schulich Interfaculty Program in Public Health, Western University, London, Ontario, Canada
3Departments of Epidemiology & Biostatistics and Family Medicine, Western University, London, Ontario, Canada
4Departments of Epidemiology & Biostatistics, Computer Science, Schulich Interfaculty Program in Public Health, Statistical & Actuarial Sciences, Western University, London, Ontario, Canada

ABSTRACT

PURPOSE Rapid increases in technology and data motivate the application of artificial intelligence (AI) to primary care, but no comprehensive review exists to guide these efforts. Our objective was to assess the nature and extent of the body of research on AI for primary care.

METHODS We performed a scoping review, searching 11 published or gray literature databases with terms pertaining to AI (eg, machine learning, bayes* network) and primary care (eg, general pract*, nurse). We performed title and abstract and then full-text screening using Covidence. Studies had to involve research, include both AI and primary care, and be published in English. We extracted data and summarized studies by 7 attributes: purpose(s); author appointment(s); primary care function(s); intended end user(s); health condition(s); geographic location of data source; and AI subfield(s).

RESULTS Of 5,515 unique documents, 405 met eligibility criteria. The body of research focused on developing or modifying AI methods (66.7%) to support physician diagnostic or treatment recommendations (36.5% and 13.8%), for chronic conditions, using data from higher-income countries. Few studies (14.1%) had even a single author with a primary care appointment. The predominant AI subfields were supervised machine learning (40.0%) and expert systems (22.2%).

CONCLUSIONS Research on AI for primary care is at an early stage of maturity. For the field to progress, more interdisciplinary research teams with end-user engagement and evaluation studies are needed.


INTRODUCTION

Artificial intelligence (AI) research began in the 1950s, and public, professional, and commercial recognition of its potential for adoption in health care settings is growing.1-7 This application includes primary care,8-10 defined by Barbara Starfield as “The level of a health service system that provides entry into the system for all new needs and problems, provides person-focused (as opposed to disease-oriented) care over time, provides care for all but very uncommon or unusual conditions, and coordinates or integrates care provided elsewhere or by others.”11(p8-9)

Given the recent surge in uptake of electronic health records (EHRs) and thus availability of data,12,13 there is potential for AI to benefit both primary care practice and research, especially in light of the breadth of practice and rapidly increasing amounts of information that humans cannot meaningfully condense and comprehend.1,2,4-11,14-20

AI’s immediate usefulness is not guaranteed, however: EHRs were predicted to transform primary care for the better, but led to unanticipated outcomes and encountered barriers to adoption.12,21-23 AI could also harm, for example, by exaggerating racial, class, or sex biases if models are built with biased data or used with new populations for whom performance may be poor. Liability, trust, and disrupted workflow are further concerns.3

AI initially focused on how computers might achieve humanlike intelligence and how we might recognize this.24-25 Two approaches emerged,
rule centric and data centric. Rule-centric methods capture intelligence by explicitly writing down rules that govern intelligent decision making, whereas data-centric methods learn specific tasks using previously collected data.24 Examples of health applications are presented below.

MYCIN was the first rule-based AI system for health care, developed in the 1970s to diagnose blood infections using more than 450 rules derived from experts, textbooks, and case reports.24,26 Although met with initial enthusiasm, rule-centric methods faltered when faced with increasing complexity. As availability of EHRs increased, AI shifted toward data-centric, machine learning methods designed to automatically capture complex relationships within health data. Machine learning methods are now used in health research to predict diabetes and cancer from health records,16,27-29 and together with computer vision have been applied to skin cancer diagnosis based on skin lesion images.30,31 Machine learning and natural language processing methods extract structured information from unstructured text data,15 which could potentially remove some of the EHR-associated burden from clinicians.6,32,33

These examples predominantly come from referral care settings, not from primary care, where the spectrum of illness is wider, and clinicians have fewer diagnostic instruments or tests available. Despite optimism for using AI to benefit primary care, there is no comprehensive review of what contribution AI has made so far, and thus little guidance on how best to proceed with research. To address this gap, our objective was to identify and assess the nature and extent of the body of research involving AI and primary care.

METHODS

We performed a scoping review according to published guidelines whereby a systematic search strategy identifies literature on a topic, data are extracted from relevant documents, and findings are synthesized.34-36 We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews (PRISMA-ScR) Checklist (Supplemental Appendix 1, http://www.AnnFamMed.org/content/18/3/250/suppl/DC1),37 and registered our protocol with the Open Science Framework (osf.io/w3n2b).

Search Strategy

We developed our search strategies (Supplemental Appendix 2, http://www.AnnFamMed.org/content/18/3/250/suppl/DC1) iteratively and in collaboration with a medical sciences librarian for health sciences, computer science, and interdisciplinary databases. Strategies included key words and, where possible, subject headings around the concepts of AI and primary care. Terms were identified through searches of the National Library of Medicine MeSH Tree Structures and by discipline experts on our review team. Supplemental Appendix 2 contains an overview of the search strategy development process (Figure 1S) and final strategies for the 11 published or gray literature databases: Medline-OVID, EMBASE, CINAHL, Cochrane Library, Web of Science, Scopus, Institute of Electrical and Electronics Engineers Xplore, Association for Computing Machinery Digital Library, MathSciNet, Association for the Advancement of Artificial Intelligence, and arXiv. Retrieved references were uploaded into Covidence.38 Where possible, English-language limits were set, to estimate the amount of literature missed, searches were rerun for a subset of the databases (Medline-OVID, CINAHL, Web of Science) with language limits reset to accept all non-English languages. Each search retrieved fewer than 10 documents. We used Covidence38 to remove duplicate results and facilitate the screening process.

Study Selection

Title and Abstract Screening

For preliminary screening, 2 reviewers (J.K.K., D.J.L.) independently rated document titles and abstracts as to whether they met our eligibility criteria: (1) reported on research, (2) mentioned or alluded to AI, and (3) mentioned primary care data source, setting, or personnel. We pilot-tested the first 25 and next 100 documents, discussing disagreements to ensure mutual understanding of the eligibility criteria and capture of relevant literature. A third reviewer (A.L.T.) resolved remaining initial disagreements. If 2 reviewers rated a document as meeting the above criteria, the document progressed to full-text screening. A large number of documents on computerized cognitive behavioral therapy (37 documents) were excluded because underlying methods were often unclear and reviews on these systems already exist.19,41

Full-Text Screening

For our full-text screening, 2 reviewers (J.K.K., D.J.L.) independently reviewed the full text of each document for the following eligibility criteria: (1) was a research study, (2) developed or used AI (Table 1S, Supplemental Appendix 3, http://www.AnnFamMed.org/content/18/3/250/suppl/DC1, contains subfield definitions), (3) used primary care data and/or study was conducted in a primary care setting and/or explicitly mentioned study applicability to primary care. Documents were excluded if they were narratives or editorials, did not apply to primary care, or were not...
accessible in English language full text. As for title and abstract screening, we performed pilot-testing and refined the eligibility criteria. Disagreements were resolved by discussion until consensus was reached.

A notable challenge arose from authors’ use of terminology that overlaps with AI when the methods used are not considered AI, we excluded these studies. We also excluded 34 studies because there was insufficient information to determine whether AI was involved, even after consulting references cited in methods. For example, 1 study referred to simple string matching as natural language processing.44

Data Extraction and Synthesis
We developed the data extraction sheet iteratively to ensure relevant and consistent information capture, performing pilot-testing and revisions for 3 and then 5 randomly selected articles.31,45-51 Remaining documents were split alphabetically and extracted independently (100 by A.L.T., 50 by D.J.L., 250 by J.K.K.). We extracted the following information: publication details, study purpose(s), author appointment(s), primary care function(s), author-intended target end user(s), target health condition(s), location of data source(s) (if any), AI subfield(s), the reviewer who performed extraction, and any reviewer notes. We agreed on definitions for each data extraction field (Table 1S, Supplemental Appendix 3). For fields except publication details, author appointments, and additional notes, we predefined categories based on the pilot testing and on content knowledge; studies could belong to multiple categories. An “other” category captured specifics of studies that did not fit into a predefined category, and an “unknown” category was used if not enough information was provided for category selection. We summarized results as categorical variables for 7 data extraction fields and performed selected cross-tabulations.

RESULTS

Searches
We retrieved 5,515 nonduplicate documents for title and abstract screening; 727 met the eligibility criteria for full-text screening and 405 met the final criteria as shown in Figure 1. (Supplemental Appendix 4, available at http://www. AnnFamMed.org/content/18/3/250/suppl/DC1, contains a list of the 405 references.) The AI and primary care study with the earliest date of publication, 1986, developed a supervised machine learning method to support abdominal pain diagnoses.52 Studies are summarized below according to the 7 key data extraction categories mentioned above.

Study Purpose
The majority of studies (270 studies, 66.7%) developed new or adapted existing AI methods using secondary data. The second most common study purpose (86 studies, 21.2%) was analyzing data using AI techniques, such as eliciting patterns from health data to facilitate research. Few (28 studies, 6.9%) evaluated AI application in a real-world setting.

Some series of studies reported on multiple stages of a project, from AI development to pilot-testing; these projects included intended end users located in a primary care setting.53-60 A small minority of studies (21 studies, 5.2%) had multiple purposes. Figure 2 presents all combinations.

Figure 1. PRISMA flow diagram.

7,900 Documents imported for screening
(Completed April 6, 2018)

2,385 Duplicates

5,515 Underwent title and abstract screening
(Completed May 28, 2018)

4,788 Excluded

727 Underwent full-text screening
(Completed September 14, 2018)

322 Excluded

177 Not AI
42 Not primary care
60 Not research studies
20 Full text not accessible
21 Duplicates
2 Not in English language

405 Met all eligibility criteria
(Data extraction completed January 23, 2019)

AI = artificial intelligence; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

*Not primary care* used as exclusion criterion when multiple criteria applied.
Author Appointment

We categorized author appointments into 4 categories: (1) technology, engineering, and math (TEM) discipline, meaning an author appointed in a department of mathematics, engineering, computer science, informatics, and/or statistics; (2) primary care discipline, meaning an author appointed in a department of family medicine, primary care, community health, and/or other analogous term; (3) nursing discipline; and (4) other. Authors were predominantly from TEM disciplines with 214 studies (52.8%) having at least 1 author with a TEM appointment compared with just 57 studies (14.1%) having at least 1 author with a primary care appointment. Twenty-three studies (5.7%) had a primary care–appointed author listed first and 27 (6.7%) had one listed last. These patterns remained when unspecified or general medical appointments (ie, nonspecialist) were counted as primary care appointments. Four studies had authors with nursing appointments. Cross-tabulations between study purpose and author appointment categories did not suggest that author appointment types differed by study purpose. Table 1 presents a summary of the body of literature broken into primary care and TEM author disciplines; Table 2S (Supplemental Appendix 3) breaks down author appointments into 16 categories.

Primary Care Function

Diagnostic decision support was the most common primary care function addressed in studies (148 studies, 36.5%), followed by treatment decision support (56 studies, 13.8%), and then using AI for extracting information from data sources such as EHRs (49 studies, 12.1%). The most frequent combination of functions was information extraction and description (21 studies, 5.2%). Figure 3 summarizes primary care function counts; Figure 2S (Supplemental Appendix 3) presents more detail.

Reported Target End User

The majority of studies reported physicians as a target end user, either alone or in combination with other target end users (243 studies, 60%). There appeared to be no positive association between having physicians as a target end user and having at least 1 author with a medical appointment: the percentage of studies with at least 1 author with any kind of medical appointment was similar between studies with physician and exclusively nonphysician target end users (51.9% and 46.3%, respectively). Twenty-six studies (6.4%) stated that their research was intended for patients, 25 (6.2%) for administrative use, and 9 (2.2%) for nurses or nurse practitioners, either alone or in combination with other end users. Figure 3S-A (Supplemental Appendix 3) shows the number of studies that included each of the target end user categories; Figure 3S-B (Supplemental Appendix 3) presents all combinations on a per-study basis.

Health Condition

About one-quarter of studies (108 studies, 26.7%) focused on developing, using, or analyzing AI so that it would be relevant for most health conditions seen in primary care settings. Of studies that targeted a particular condition, chronic physical conditions were more frequent than acute or psychiatric conditions. We condensed target health conditions into 10 categories, with study distribution shown in Figure 4; Figure 4S (Supplemental Appendix 3) expands them into 27 categories.

<table>
<thead>
<tr>
<th>Table 1. Appointments of Study Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author Appointment Category</strong></td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Primary care and TEM</td>
</tr>
<tr>
<td>Primary care and no TEM</td>
</tr>
<tr>
<td>TEM and no primary care</td>
</tr>
<tr>
<td>Neither TEM nor primary care</td>
</tr>
</tbody>
</table>

TEM = technology, engineering, and math.

Note: To be included in a row count, a study must have had at least 1 author with an appointment in the category or categories indicated.
**Geographic Location**
The location of most data source(s) used in a study or the intended location of AI implementation was higher-income countries belonging to the Organisation for Economic Co-operation and Development. Low- and middle-income countries were poorly represented. Most studies used data from a single country, with the United States being the most common source (79 studies, 19.5%). Figure 5S (Supplemental Appendix 3) summarizes location counts and per capita rates; Table 3S (Supplemental Appendix 3) contains a more detailed breakdown.

**AI Subfield**
Most studies (363 studies, 89.6%) used methods within a single subfield of AI, and of these, supervised machine learning was the most common (162 studies, 40.0%), followed by expert systems (90 studies, 22.2%), and then natural language processing (35 studies, 8.6%). There were no articles on robotics. Expert systems had the earliest median year of publication (2007); data mining had the most recent (2015). Figure 6S-A (Supplemental Appendix 3) presents frequencies and median year of publication for 10 subfields of AI used by studies captured in our literature review; all AI subfield combinations are presented in Figure 6S-B (Supplemental Appendix 3).

**DISCUSSION**

**Key Findings**
We identified and summarized 405 research studies involving AI and primary care, and discerned 3 predominant trends. First, regarding authorship, the vast majority of studies did not have any primary care involvement. Second, in terms of methods, there was a shift over time from expert systems to supervised machine learning. And third, when it came to applications, studies most often developed AI to support diagnostic or treatment decisions, for chronic conditions, in higher-income countries. Overall, these findings show that AI for primary care is at an early stage of maturity for practice applications, meaning more research is needed to assess its real-world impacts on primary care.

The dominance of TEM-appointed authors and AI methods development research is congruent with the early stage of this field. An AI-driven technology needs to be working well before real-world testing and implementation. Good performance is achieved through methods development research, which is further reflected by most studies specifying researchers as an intended end user alongside clinicians—more work is required before implementing the AI in a practice setting. On the other hand, research focused on AI for analyzing health data is distinct and at a later stage of maturity. These AI applications are not intended for everyday clinical practice, so although their methodologic performance is important, longer-term health or workflow outcomes may not need to be assessed before real-world use.

The dominant subfields of AI identified by our review mirror trends in AI advances and align with other characteristics of
the included studies. Expert systems comprise a substantial portion of the literature but are now less common (median publication year 2007 vs 2014 for supervised machine learning), reflecting a general shift in AI research from expert systems and rule-centric AI methods to machine learning and data-centric AI methods.\textsuperscript{63} The latter are amenable to providing diagnostic and treatment recommendations as well as predicting future health, which supports primary care activities such as primary prevention and screening. This trend also aligns with the focus on physicians as target eventual end users.

Underlying drivers of AI research, and by extension maturation, are data availability and quality, particularly after the shift toward data-driven machine learning methods. The United States is the single dominant country in the field, which is unsurprising given its population, wealth, and research resources and output.\textsuperscript{64-67} The high standing of the United Kingdom and Netherlands despite smaller populations may be attributable to primary care data availability,\textsuperscript{68,69} facilitated by high adoption rates of EHRs,\textsuperscript{70} and strong information technology academics and industries.\textsuperscript{71,72} Investments in data generation, quality, and access will increase future possibilities for AI to be used to strengthen primary care in the corresponding region.

Strengths and Limitations

Strengths of our review include a comprehensive search strategy, without date restriction, with use of inclusive eligibility criteria and conducted by an interdisciplinary team. Limitations include multiple reviewers extracting data without double coding, English language restriction, and the lack of single widely accepted definitions for primary care or AI to guide screening. Proprietary research would not be captured by our review, nor would research completed after our search date.

Future Research

Our next steps include further assessing the quality of the included studies and summarizing exemplary research projects. We additionally recommend a review on AI for the broader primary health care system that includes clinicians beyond physicians and nurses (eg, social workers, physiotherapists).

For the field to mature, future research studies should have interdisciplinary teams with primary care end user engagement. Value must be placed both on developing rigorous methods and on identifying potential impacts of the developed AI on care delivery and longer-term health outcomes. Inclusion of nurses, patients, and administrators needs to increase—identifying relevant nonphysician end user activities that could be augmented by AI is an outstanding research endeavor on its own.

We expect future AI methods development to shift toward a middle ground between rule-centric and data-centric methods because interpretable models better support decisions and trust in the health care setting. For example, explainable AI is a paradigm whereby one can understand what a model is doing or why it arrives at a particular output.\textsuperscript{73-75} Interpretability of models is additionally important from an equity lens to be able to identify and then avoid AI reproduction of biases in data, which is a present concern with data-driven methods.\textsuperscript{76} It is also important to remember that AI is not always a superior solution: a recent review found no benefit overall of machine learning compared with logistic regression analysis for clinical prediction rules.\textsuperscript{77}
Conclusions

Ours is the first comprehensive, interdisciplinary summary of research on AI and primary care. Two fundamental aims in the body of research emerged: providing support for clinician decisions and extracting meaningful information from primary care data. Overall, AI for primary care is an innovation that is in early stages of maturity, with few tools ready for widespread implementation. Interdisciplinary research teams including frontline clinicians and evaluation studies in primary care settings will be crucial for advancement and success of this field.

To read or post commentaries in response to this article, see it online at http://www.AnnFamMed.org/content/18/3/250.

Submitted August 19, 2019; submitted, revised, November 11, 2019; accepted November 21, 2019.

Key words: artificial intelligence; health information technology; health informatics; electronic health records; big data; data mining; primary care; family medicine; decision support; diagnosis; treatment; scoping review

Funding support: The Canadian Institutes of Health Research and Innovations Strengthening Primary Health Care through Research supported this research through funding for J. Kueper’s doctoral studies (CGS-D and TUTOR-PHC Fellowship).

Disclaimer: The views expressed are solely those of the authors and do not necessarily represent official views of the authors’ affiliated institutions or funders.


Supplemental materials: Available at http://www.AnnFamMed.org/content/18/3/250/suppl/DC1/.

References


63. Hao K. We analyzed 16,625 papers to figure out where AI is headed next. *MIT Technology Review,* 2016; 5:361.


