

Application of Artificial Intelligence Methods to Pharmacy Data for Cancer Surveillance and Epidemiology Research: A Systematic Review

Andrew E. Grothen, PharmD¹; Bethany Tennant, PhD²; Catherine Wang, PharmD¹; Andrea Torres, PhD²; Bonny Bloodgood Sheppard, MA²; Glenn Abastillas, MA³; Marina Matatova, MA¹; Jeremy L. Warner, MD, MS⁴; and Donna R. Rivera, PharmD, MSc¹

PURPOSE The implementation and utilization of electronic health records is generating a large volume and variety of data, which are difficult to process using traditional techniques. However, these data could help answer important questions in cancer surveillance and epidemiology research. Artificial intelligence (AI) data processing methods are capable of evaluating large volumes of data, yet current literature on their use in this context of pharmacy informatics is not well characterized.

METHODS A systematic literature review was conducted to evaluate relevant publications within four domains (cancer, pharmacy, AI methods, population science) across PubMed, EMBASE, Scopus, and the Cochrane Library and included all publications indexed between July 17, 2008, and December 31, 2018. The search returned 3,271 publications, which were evaluated for inclusion.

RESULTS There were 36 studies that met criteria for full-text abstraction. Of those, only 45% specifically identified the pharmacy data source, and 55% specified drug agents or drug classes. Multiple AI methods were used; 25% used machine learning (ML), 67% used natural language processing (NLP), and 8% combined ML and NLP.

CONCLUSION This review demonstrates that the application of AI data methods for pharmacy informatics and cancer epidemiology research is expanding. However, the data sources and representations are often missing, challenging study replicability. In addition, there is no consistent format for reporting results, and one of the preferred metrics, F-score, is often missing. There is a resultant need for greater transparency of original data sources and performance of AI methods with pharmacy data to improve the translation of these results into meaningful outcomes.

JCO Clin Cancer Inform 4:1051-1058. © 2020 by American Society of Clinical Oncology

INTRODUCTION

The implementation of electronic health records (EHRs) alongside advances in precision medicine provides researchers and clinicians with access to unprecedented amounts of data. However, the volume, veracity, and variety of heterogeneous data contained in these new sources cannot be easily or quickly processed by humans.¹ This is especially true with unstructured data, such as free-text documents, which many text mining programs and data processing techniques are inefficient at accurately processing.¹ New developments in natural language processing (NLP) and other artificial intelligence (AI) approaches have begun to allow for scaled analysis in oncology.² Detailed evaluation of health data is required to answer pressing questions within cancer research, such as in cancer surveillance and epidemiology, where understanding survivorship outcomes is increasingly important.³ Pharmacy data represent one of the more voluminous data streams and have traditionally not

been captured by cancer registries. Pharmacy data captured in EHRs is specifically useful to understand prescribing and treatment patterns, adverse events, medication adherence and administration, response to treatment, and treatment outcomes. Pharmacy data could include variables related to medication use, such as drug name, dose, strength, directions (“SIGs”), cycles, and regimens. Medications are contained in both structured and unstructured formats, which can limit or control how efficiently the data are analyzed.⁴ Although formal models, or ontologies, for the representation of chemotherapy regimen data have been proposed, they are not yet widely used in production EHR environments.^{5,6}

Machine learning (ML) and NLP are advanced analytical methods capable of processing large volumes of data in a manner approximating the human learning processes and are collectively referred to as methods of AI.^{7,8} These methods are not mutually exclusive, and NLP in particular frequently uses ML techniques to

ASSOCIATED CONTENT

Appendix

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on October 9, 2020 and published at ascopubs.org/journal/cci on November 16, 2020; DOI <https://doi.org/10.1200/CCI.20.00101>

CONTEXT

Key Objective

To evaluate artificial intelligence (AI) data processing methods being leveraged in cancer pharmacy surveillance and epidemiology, including performance assessments and utility in population research.

Knowledge Generated

This systematic review, including studies published between July 17, 2008, and December 31, 2018, found a steadily increasing volume of AI methods research within cancer surveillance and epidemiology, reaching 524 publications in 2018. However, methodologic differences across studies present challenges to reproducibility, including inconsistent reporting of results and a lack of specificity in data source reporting; 45% of included studies specifically identified the data source. Further research may benefit from focusing on maximizing performance through consensus use of specific methods and reporting standards for informatics.

Relevance

These findings identify potential barriers to the effective use of emerging AI methods and critical evaluation of their performance, further establishing the need for standardization to more rapidly enhance the translation of big data in cancer research to clinical practice and patient care.

work around limitations such as inconsistent vocabulary usage. ML can be either supervised or unsupervised. Supervised techniques are defined as those methods or algorithms where idealized or preconstructed manually curated information is provided and the algorithm attempts to identify or replicate the results using the provided inputs.⁹ Unsupervised techniques are methods in which the algorithms attempt to structure the data and infer patterns in a way that maximizes outputs without providing structured or labeled information for the computational analysis^{9,10} (Appendix Table A1). Both techniques have been applied in the context of cancer research, including medical image analysis and diagnosis extraction.^{1,9} The value demonstrated through an NLP systematic review in the context of radiology, conducted by Pons et al,¹ provided the rationale to explore the use of advanced analytical methods across other clinical data types, such as for efficiently leveraging pharmacy data. Although AI is a promising approach to big data problems, there are concerns about methods, approaches, ethics, and reproducibility.^{11,12} Given the lack of previously published systematic reviews on this budding topic, the purpose of this systematic review is to evaluate the current literature on the application of NLP and ML to pharmacy data in cancer surveillance and epidemiology research.

METHODS

Data Sources and Review Phases

A literature search was conducted to locate peer-reviewed publications across PubMed, EMBASE, Scopus, and the Cochrane Library, indexed between July 17, 2008, and December 31, 2018. The precise search strategy was conducted using keywords across four content categories—cancer, pharmacy, AI methods, and population science—and restricted to English language publications with full text available.

Title/Abstract Phase

All titles and abstracts were evaluated for inclusion and exclusion by four reviewers in two teams: original search (P.F./M.M., A.E.G., C.W., and D.R.R.) and bridge search (J.L.W., A.E.G., C.W., and D.R.R.). For title/abstract phases the total number of publications was evaluated by a team of two independent reviewers. Discrepancies were discussed by the opposite team of independent reviewers until a consensus was reached. Studies that met the following criteria were selected for full text review: (1) used an observational study design, and (2) addressed pharmaceutical interventions for cancer and AI methods. Studies were excluded if human patient information was not reported, treatment information was excluded, or the treatment specified did not include a drug treatment.

Full-Text Phase

In the full-text phase, studies were reviewed by four reviewers (A.E.G., A.T., B.T., and C.W.) composed of two independent reviewer teams using the same method as in the previous phase. Individual studies were included or excluded based on the Study Design, Participants, Intervention, and Outcomes (SPIO) criteria (Table 1).

Abstraction Phase

Full-text abstraction was completed by two independent reviewers (A.E.G. and B.T.), and discrepancies were resolved by consensus. Abstracted fields included: (1) study purpose; (2) research topic; (3) cancer type or types studied; (4) study setting (ie, multisite, single practice, or other clinical setting); (5) key study characteristics, including sample size, study design, population characteristics, primary data source used, primary outcome measure, study limitations, and conclusions; (6) type and source of pharmacy data and the identification of specific drugs or drug classes; (7) type of AI methods used in the analysis; (8) characteristics of AI methods, including key findings on AI

TABLE 1. SPIO Systematic Review Study Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Study design	Any observational study design, whether as prospective or retrospective observation design	Citations that do not use primary epidemiologic data (eg, systematic reviews/meta-analyses, no interventional studies, no commentaries/narratives)
Cancer surveillance and epidemiology (participants)	Results reporting outcome prediction, incidence, prevalence, population-level data, survival outcomes, malignancies and tumors	Keywords: in vivo, in vitro, cell lines
	Human data, human health records, or patients	Nonhuman animal data
	Data abstraction, geographic regions	Nonmalignant diseases
Pharmacy data (interventions)	Indicators of pharmacy data being used in the study may include: drug names, administration routes, pharmacy claims, adverse events, survival, utilization, neoadjuvant, adherence/compliance	Radiology, screening, imaging, surgery in the absence of pharmacy keywords
AI methods: NLP/ML/AI (outcomes)	Keywords that may identify the use of NLP/ML techniques: fuzzy logic, random forest, decision tree, text mining, support vector machine, named entity recognition	Robot
	Use of the word “learning” outside the context of biologic organisms also may qualify	Keywords for behavioral interventions
Data source	SEER, NCDB, registries, claims databases, reimbursement, medical records	Based on prior exclusion criteria

Abbreviations: AI, artificial intelligence; ML, machine learning; NCDB, National Cancer Database; NLP, natural language processing; SPIO, Study Design, Participants, Intervention, Outcome.

method, novelty of the overall method, purpose, learning processes (ie, supervised, unsupervised); and (9) reported performance measures (eg, accuracy, precision, recall, F-score, area under the receiver operating characteristic curve [AUC], error analysis metrics, or other).

Reporting of Results (aggregation)

Available performance measures are reported for all studies. For Study Characteristics—Patient Data Source and Pharmacy Data—Type of Data Used, if the source was unspecified by the authors the field was populated with either “Patient Clinical Data” or “Not Specified.” Where a single method was present, the mean, median, and range (highest and lowest) scores are reported where available. In the case of multiple reported methods, the mean or median performance, where available, as well as the highest and lowest scoring methods and their metrics were reported; fields with no reported findings for a measure were populated with “NR” for “Not Reported.” For Study Limitations, where no discussion of study limitations stated by the author could be found, the field was populated with “None explicitly discussed.”

RESULTS

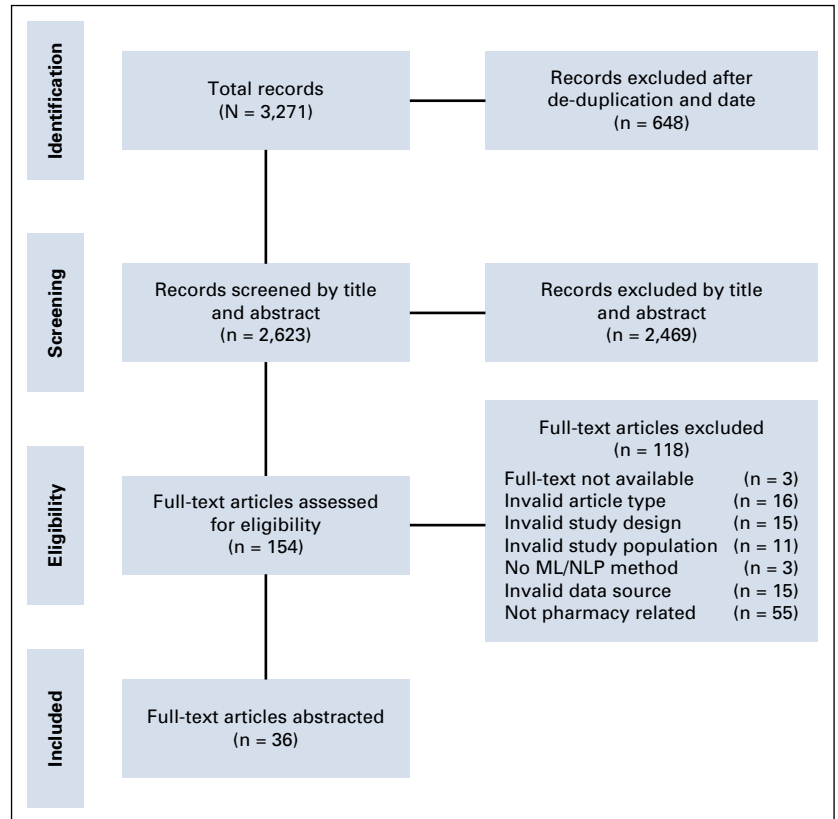
The search returned 3,271 publications in total, with 2,297 unique citations after deduplication from the original search (July 17, 2008, to July 15, 2018) and an additional 326 unique citations identified for review in the bridge search (July 16, 2018, to December 31, 2018). The results for each review phase are contained in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-

Analyses) diagram (Fig 1).¹³ The review identified 36 studies for abstraction: 29 from the original search and seven additional studies from the bridge search. There was a noticeable increase in relevant publications over time, from 115 publications in 2009 to 526 publications in 2018 (456% increase), with the highest year-over-year increase occurring between 2017 and 2018, with 153 additional new studies published in this time frame (Fig 2).

Among the included studies (Data Supplement), the study purpose varied, with the majority focused on supporting treatment selection and clinical decision making (45%). Other main purposes included predicting patient survival and prognosis (22%), collecting patient characteristics and demographic information (17%), and evaluating the incidence and reporting of adverse events (17%). Data source and methodology were varied as well. The patient data sources used in studies were most commonly unspecified (25%)—referred to as “Patient Clinical Data” in the abstractions—followed by EHR data (22%). Patient data associated with other primary sources was used in 14% of studies, and the remaining 39% of studies used miscellaneous data sources, such as public Twitter posts, patient-provider interviews, and patient-reported narratives.¹⁴⁻¹⁶ Nineteen percent of the studies used more than one data source for their patient data.

Specific to oncology, the included studies examined data across 12 cancer types. The most common type included was breast cancer (n = 11). Ten of the studies either did not specify the cancers they captured or did not limit their study or data by the cancer type recorded in the study; these have

FIG 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram: study inclusion for systematic review. ML, machine learning; NLP, natural language processing.



been referred to as “Pan-cancer” in the abstractions and were the second most common cancer type (n = 10), followed by studies of lung (n = 4) and ovarian cancer (n = 2). The following types were only specifically reported in a single study: cervical, colon, esophageal, peritoneal, prostate, renal, solid tumors not otherwise specified, and soft-tissue sarcoma. Of the 36 included studies, 16 of them specifically identified the pharmacy data source. The 20 that did not were recorded as either “Patient Data” or “Not

Specified” during abstraction as a result. Fourteen of the studies did not specify the drug agents or classes found by the researchers to any degree. Of the 22 studies that specified the drug treatments, 14 identified them categorically (eg, “Experimental,” “Platinum-Based”) and the remaining eight studies specified drug treatments captured individually.^{17,18} The highest number of specifically tracked therapies in a study was 15 individual agents.¹⁹ The most commonly administered drug class among all studies was platinum-based antineoplastics (n = 10; Data Supplement)

Among all included studies, half of them used more than one type of a given AI method, regardless of whether the methods were ML or NLP based. Twenty-five percent used NLP methods alone (n = 9), 67% used ML methods alone (n = 24), and 8% used both NLP and ML methods (n = 3). In addition, there were differences in the level of human involvement between studies; 41% of the studies used unsupervised learning (n = 15), 53% used supervised learning (n = 19), and 6% used supervised and unsupervised learning together (n = 2). ML methods differed among studies, including eight studies that used multiple ML methods for data processing, including bootstrap aggregating, random forest, support vector machines, naive Bayesian classifiers, and least absolute shrinkage and selection operator (LASSO).¹⁸ The most commonly reported outcome for AI method performance was area under the receiver operating characteristics curve, abbreviated as either AUC or AUROC (n = 14). Recall (ie, sensitivity) was

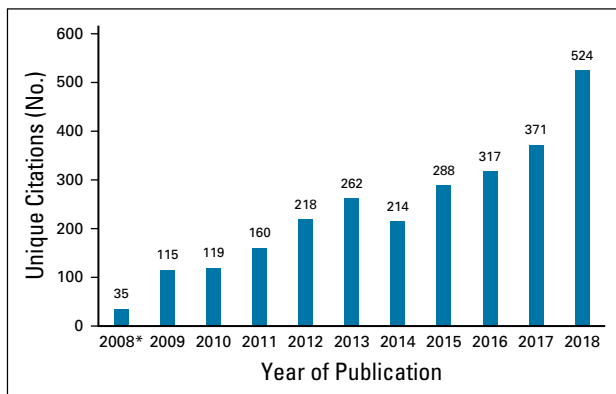


FIG 2. Pharmacy data and artificial intelligence systematic review overall article search capture for initial evaluation by year of publication. Unique publications progressively increase with time. Data show the number of unique citations published in each calendar year. (*) Unique studies published July 17, 2008, to December 31, 2008.

also a commonly reported measure ($n = 12$). Of the studies reporting precision, recall, F-score, or AUC, 10 achieved results ≥ 0.90 for at least one measure, with recall being the most commonly reported measure ≥ 0.90 in studies ($n = 7$; Data Supplement). Reporting of limitations among the abstracted studies was not consistent. There was no report of study limitations among six studies. The three most commonly stated limitations included variable or missing data ($n = 17$), followed by small sample size ($n = 8$) and clinical data from a single treatment site ($n = 6$).

DISCUSSION

The evaluation of $> 3,000$ publications in this systematic review represents more than a decade of oncology pharmacy research studies and AI scientific methods development. The rapid increase in utilization of large health care databases, including platforms for integration in health care systems and AI data methods, is observable in the increase in rate of these publications. Between 2017 and 2018, there was a 41% increase in publications, from 371 to 524, which is also the largest number of annual publications in this study. Although a yearly increase in publications of 153 from 2017 to 2018 may be modest in absolute terms, especially compared with other areas of oncology research, it represents more than the mean annual number of new publications before 2011 and nearly a five-fold increase in the number of relevant studies from the first full year of the review. Moreover, the bridge search, which was conducted to capture studies published between July 16, 2018, and December 31, 2018, returned seven additional studies eligible for final inclusion, representing almost one-fifth of all the included studies. Interest and exploration in the use of these techniques to process pharmacy data in support of cancer epidemiology and surveillance efforts appears to be increasing.

However, the integration of AI-based methods in cancer surveillance and epidemiology research for pharmacy data still consistently presents limitations. More than half of the included studies did not specify or inadequately specified the source or type of pharmacy data in their text or supplementary materials. Without this information, it is problematic for current or future researchers to replicate, test, or validate the methods or results using pharmacy data.²⁰ The value of information provided by pharmacy data is also strongly dependent on the purpose of the data collection and data provenance; medication administration records, prescription fill data, administrative claims, and prescription orders provide different data elements with varying levels of detail. Each data set must be assessed for appropriateness and quality when answering a specific research question.^{21,22} Without the ability to understand the underlying data, it is difficult to reproduce findings or determine how relevant a method may be to predicting or evaluating a given clinical context. Failure to account for these shortfalls then potentially leads to inappropriate conclusions, which results in confusion within the broader

research community and raises concerns regarding the integrity of the underlying research.²⁰

In addition, there were difficulties in comparing the performance of the AI methods reported. Fewer than half of the included studies shared any one consistent performance metric; some studies reported none of the prespecified performance metrics. Depending on the purpose of the AI method, uniformity in performance reporting is essential to meaningfully compare approaches. For example, the NLP field has adopted F-score as a default reporting metric which is used to compare performance in community tasks. The primary metrics of F-score, precision, recall, and AUC are also commonly used across many different medical AI research contexts.²³⁻²⁵ In addition, recent high-quality publications have achieved performance metrics better than 0.90²³⁻²⁷ for these measures. By comparison, 10 of the studies abstracted were able to achieve such results in at least one metric. Although performance in a given metric can be expected to vary considerably between data types and context, strong performance with these technologies achieved in previous study suggests that current research efforts between cancer surveillance, epidemiology, and pharmacy data may benefit from a focus on determining which methods perform more effectively across multiple contexts.²²

Missing or inconsistent data were listed as an explicit limitation in almost half of the studies. Although this is expected with real-world data, AI methods are understood to have advantages over more traditional methods in processing these data.²⁸ Conversely, numerous studies did not describe their study limitations, contributing to the general dilemma of a lack of reproducibility.²⁹⁻³¹ A failure to account for limitations in research prevents other researchers from appropriately replicating the findings or developing new research methods that build on the original findings. In conjunction with the wide variations in method performance and reported metrics, it also prevents outside researchers from readily interpreting the significance of the study or its performance relative to existing methods, hindering adoption of any otherwise promising methods.

The use of real-world data in AI-assisted research presents numerous ethical challenges.^{20,32} Although the use of aggregated data sources easily allows for personal identifiers to be excluded or used with matching methodologies, the development of AI techniques and algorithms requires researchers to make certain assumptions about which data are most valuable. As such, it is well known that AI systems are subject to a number of ethical risks and biases, such as difficulty assigning responsibility when a system generates incorrect or harmful outcomes or difficulty identifying how to correct or improve the system or determine the impact of a given variable because of how an AI technique reaches conclusions. Moreover, although individual patient identifiers may be relatively easy to purge from the results, the level of access AI systems demand of personal information

frequently conflicts with privacy concerns.^{32,33} In many ways, the frequent absence of discussion regarding limitations in the abstracted studies, and the attribution of poor performance to missing data, reflects the broader ethical issues that these AI methods, and the researchers using them, face when using real-world data.²⁰ A lack of transparency on the limitations of these systems in a study prevents researchers from thoroughly evaluating the ethical risks of a given AI method. The heavy attribution of their poor performance to missing data, without addressing concerns for patient privacy, data validity, or other ethical dilemmas, also reinforces potentially fallacious reasoning that poor AI performance can be resolved through compromises in protections that seek to target these issues.

There are limitations to this systematic review. The broad search strategy and multiple publication databases permitted capture of as many publications relevant to this study as possible, although the heterogeneity of the studies in terms of methods, populations, and outcomes made identifying specific risks of bias in the abstracted citations prohibitive. This heterogeneity also introduced more opportunity for inconsistencies in publication selection, which could stem from differences in the initial citation data or a greater risk of inter-reviewer discrepancy during review. As a result, it is possible that some relevant studies may have been excluded or may not have been captured in this

review. Similarly, significant and newer publications in this field frequently appear in symposia³⁴⁻³⁶ or as preprints, which presented a unique challenge in identifying appropriate publication sources. The appearance of studies outside of journal publication may delay the propagation of findings in a subject that is expanding quickly. Rapidly developing fields such as AI in medicine are also difficult to capture in a retrospective systematic review because of the time requirements necessary to conduct them. As an example, BERT methods—Bidirectional Encoder Representations from Transformers—have risen as state of the art in clinical NLP recently, but they are completely absent in our systematic review.³⁷

This study demonstrates that AI methods focusing on pharmacy data for cancer surveillance and epidemiology research are increasing in prevalence. The breadth and scope of our systematic review allowed it to capture an array of data uses, providing valuable insight into the current state of research and continuing development. A large degree of variance and heterogeneity in study design and performance metrics in this area of research is evident. There is a need for greater transparency of data sources and performance of AI methods to improve the translation of the vast repositories of data into meaningful outcomes for researchers, clinicians, and patients.

AFFILIATIONS

¹Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD

²ICF Next, Fairfax, VA

³Westat, Rockville, MD

⁴Vanderbilt University, Nashville, TN

CORRESPONDING AUTHOR

Andrew E. Grothen, 9609 Medical Center Dr, Rockville, MD 20850; e-mail: andrew.grothen@usp.org.

EQUAL CONTRIBUTION

J.L.W. and D.R.R. are co-senior authors.

AUTHOR CONTRIBUTIONS

Conception and design: Andrew E. Grothen, Glenn Abastillas, Marina Matatova, Jeremy L. Warner, Donna R. Rivera

Administrative support: Donna R. Rivera

Collection and assembly of data: Andrew E. Grothen, Bethany Tennant, Catherine Wang, Andrea Torres, Bonny Bloodgood Sheppard, Jeremy L. Warner, Donna R. Rivera

Data analysis and interpretation: Andrew E. Grothen, Bethany Tennant, Catherine Wang, Jeremy L. Warner, Donna R. Rivera

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

Jeremy L. Warner

Stock and Other Ownership Interests: HemOnc.org

Consulting or Advisory Role: Westat, IBM

Travel, Accommodations, Expenses: IBM

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank Alicia A. Livinski and Brigit Sullivan, biomedical librarians at the NIH Library, for their assistance with the literature searches and management of results. We also thank Guergana Savova and Daniel Fabbri for their insight in the state-of-the-art machine learning techniques. We thank Paul Fearn posthumously for his guidance and mentorship and for providing the original concept for this systematic review.

REFERENCES

1. Pons E, Braun LM, Hunink MG, et al: Natural language processing in radiology: A systematic review. *Radiology* 279:329-343, 2016
2. Savova GK, Danciu I, Alamudun F, et al: Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. *Cancer Res* 79:5463-5470, 2019
3. Kantarjian H, Yu PP: Artificial intelligence, big data, and cancer. *JAMA Oncol* 1:573-574, 2015
4. Duggan MA, Anderson WF, Altekruze S, et al: The Surveillance, Epidemiology, and End Results (SEER) program and pathology: Toward strengthening the critical relationship. *Am J Surg Pathol* 40:e94-e102, 2016
5. Malyt AM, Jain SK, Yang PC, et al: Computerized approach to creating a systematic ontology of hematology/oncology regimens. *JCO Clin Cancer Inform* 10.1200/CCI.17.00142
6. Warner JL, Dymshyts D, Reich CG, et al: HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. *J Biomed Inform* 96:103239, 2019
7. Esteve A, Robicquet A, Ramsundar B, et al: A guide to deep learning in healthcare. *Nat Med* 25:24-29, 2019
8. Sheikhalishahi S, Miotto R, Dudley JT, et al: Natural language processing of clinical notes on chronic diseases: Systematic review. *JMIR Med Inform* 7:e12239, 2019
9. Moore MM, Slonimsky E, Long AD, et al: Machine learning concepts, concerns and opportunities for a pediatric radiologist. *Pediatr Radiol* 49:509-516, 2019
10. Libbrecht MW, Noble WS: Machine learning applications in genetics and genomics. *Nat Rev Genet* 16:321-332, 2015
11. Char DS, Shah NH, Magnus D: Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med* 378:981-983, 2018
12. Carter RE, Attia ZI, Lopez-Jimenez F, et al: Pragmatic considerations for fostering reproducible research in artificial intelligence. *NPJ Digit Med* 2:42, 2019
13. Moher D, Liberati A, Tetzlaff J, et al: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 339:b2535, 2009
14. Freedman RA, Viswanath K, Vaz-Luis I, et al: Learning from social media: Utilizing advanced data extraction techniques to understand barriers to breast cancer treatment. *Breast Cancer Res Treat* 158:395-405, 2016
15. Zhang L, Hall M, Bastola D: Utilizing Twitter data for analysis of chemotherapy. *Int J Med Inform* 120:92-100, 2018
16. Yap KY, Low XH, Chui WK, et al: Computational prediction of state anxiety in Asian patients with cancer susceptible to chemotherapy-induced nausea and vomiting. *J Clin Psychopharmacol* 32:207-217, 2012
17. Elfiky AA, Pany MJ, Parikh RB, et al: Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw Open* 1:e180926, 2018
18. Yu KH, Levine DA, Zhang H, et al: Predicting ovarian cancer patients' clinical response to platinum-based chemotherapy by their tumor proteomic signatures. *J Proteome Res* 15:2455-2465, 2016
19. Ferroni P, Zanzotto FM, Scarpato N, et al: Risk assessment for venous thromboembolism in chemotherapy-treated ambulatory cancer patients. *Med Decis Making* 37:234-242, 2017
20. The Editors Of The Lancet Group: Learning from a retraction. *Lancet* 396:1056, 2020
21. Ma C, Smith HW, Chu C, et al: Big data in pharmacy practice: Current use, challenges, and the future. *Integr Pharm Res Pract* 4:91-99, 2015
22. Liu Y, Chen PC, Krause J, et al: How to read articles that use machine learning: Users' guides to the medical literature. *JAMA* 322:1806-1816, 2019
23. Russo DP, Zorn KM, Clark AM, et al: Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction. *Mol Pharm* 15:4361-4370, 2018
24. Sidey-Gibbons JAM, Sidey-Gibbons CJ: Machine learning in medicine: A practical introduction. *BMC Med Res Methodol* 19:64, 2019
25. Handelman GS, Kok HK, Chandra RV, et al: Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *AJR Am J Roentgenol* 212:38-43, 2019
26. Ju M, Nguyen NTH, Miwa M, et al: An ensemble of neural models for nested adverse drug events and medication extraction with subwords. *J Am Med Inform Assoc* 27:22-30, 2020
27. Wei Q, Ji Z, Li Z, et al: A study of deep learning approaches for medication and adverse drug event extraction from clinical text. *J Am Med Inform Assoc* 27:13-21, 2020
28. Steele AJ, Denaxas SC, Shah AD, et al: Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLoS One* 13:e0202344, 2018
29. Cheerla N, Gevaert O: MicroRNA based pan-cancer diagnosis and treatment recommendation. *BMC Bioinformatics* 18:32, 2017
30. Huang C, Clayton EA, Matyunina LV, et al: Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. *Sci Rep* 8:16444, 2018
31. Polak S, Skowron A, Brandys J, et al: Artificial neural networks based modeling for pharmacoeconomics application. *Appl Math Comput* 203:482-492, 2008
32. Goodman KW: Ethics in health informatics. *Yearb Med Inform* 29:26-31, 2020
33. Morley J, Floridi L, Kinsey L, et al: From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics* 26:2141-2168, 2020
34. Bhatia H, Levy M: Automated plan-recognition of chemotherapy protocols. *AMIA Annu Symp Proc* 2011:108-114, 2011
35. Lee WN, Bridewell W, Das AK: Alignment and clustering of breast cancer patients by longitudinal treatment history. *AMIA Annu Symp Proc* 2011:760-767, 2011
36. Sugimoto M, Takada M, Toi M: Comparison of robustness against missing values of alternative decision tree and multiple logistic regression for predicting clinical data in primary breast cancer. *Annu Int Conf IEEE Eng Med Biol Soc* 2013:3054-3057, 2013
37. Si Y, Wang J, Xu H, et al: Enhancing clinical concept extraction with contextual embeddings. *J Am Med Inform Assoc* 26:1297-1304, 2019



APPENDIX

TABLE A1. Glossary of Overarching Artificial Intelligence Methods

Artificial Intelligence Method	Definition ^a
Natural language processing	Algorithms used to process data in a manner comparable to human language. They can be rule based or ML based.
Decision trees	ML method that uses a preselected outcome to guide computation of different operational choices, with the algorithm identifying the performance of specific choice sequences. A major variant is random forest modeling, which randomly constructs multiple decision trees on the basis of provided data and compares their performance.
Support vector machine	A supervised ML method that uses predefined examples of two categories in a training set to model an algorithm that assigns new examples to those categories. An unsupervised variant called support-vector clustering using similar principles to identify clusters in data without predefined examples, which only works for binary categories.
Bootstrap aggregation	ML method, also known as “bagging,” which randomly constructs new, synthetic training sets on the basis of initial data in an original training set, which is then used as the basis for comparing model performance.
Kernel machine learning	An ML method that converts data into more easily categorized intermediate forms for the purpose of identifying patterns.
Artificial neural network	An unsupervised ML method that uses computational “neurons” to process input data and balance the resulting calculations in a manner that maximizes the desired outcomes.

Abbreviation: ML, machine learning.

^aSources: scikit-learn: User guide. https://scikit-learn.org/stable/user_guide.html; Goodfellow I, Bengio Y, Courville A: Deep Learning. Cambridge, MA, The MIT Press, 2016