ARTIFICIAL INTELLIGENCE

original reports abstract

Artificial Intelligence Clinical Evidence Engine for Automatic Identification, Prioritization, and Extraction of Relevant Clinical Oncology Research

Fernando Suarez Saiz, MD[1]; Corey Sanders, BS[1]; Rick Stevens, BS[1]; Robert Nielsen, BS[1]; Michael Britt, BS[1]; Leemor Yuravlivker, BS[1]; Anita M. Preininger, PhD[1]; and Gretchen P. Jackson, MD, PhD[1,2,3,4]

PURPOSE We developed a system to automate analysis of the clinical oncology scientific literature from bibliographic databases and match articles to specific patient cohorts to answer specific questions regarding the efficacy of a treatment. The approach attempts to replicate a clinician's mental processes when reviewing published literature in the context of a patient case. We describe the system and evaluate its performance.

METHODS We developed separate ground truth data sets for each of the tasks described in the paper. The first ground truth was used to measure the natural language processing (NLP) accuracy from approximately 1,300 papers covering approximately 3,100 statements and approximately 25 concepts; performance was evaluated using a standard F1 score. The ground truth for the expert classifier model was generated by dividing papers cited in clinical guidelines into a training set and a test set in an 80:20 ratio, and performance was evaluated for accuracy, sensitivity, and specificity.

RESULTS The NLP models were able to identify individual attributes with a 0.7-0.9 F1 score, depending on the attribute of interest. The expert classifier machine learning model was able to classify the individual records with a 0.93 accuracy (95% CI, 0.9 to 0.96, $P < .0001$), and sensitivity and specificity of 0.95 and 0.91, respectively. Using a decision boundary of 0.5 for the positive (expert) label, the classifier demonstrated an F1 score of 0.92.

CONCLUSION The system identified and extracted evidence from the oncology literature with a high degree of accuracy, sensitivity, and specificity. This tool enables timely access to the most relevant biomedical literature, providing critical support to evidence-based practice in areas of rapidly evolving science.

JCO Clin Cancer Inform 5:102-111. © 2021 by American Society of Clinical Oncology

BACKGROUND

The explosion of biomedical research has doubled the number of citations in PubMed from 10 million in 1992 to more than 20 million in 2012, a span of 20 years.[1,2] In the past 7 years, that number has increased again by half to roughly 30 million. This acceleration is driven by factors including the proliferation of open-access journals, streamlined peer review, and availability of electronic publications. One field with rapidly evolving science is oncology. The results of approximately 82,000 clinical trials related to oncology were published between 1963 and 2010 in PubMed, with an additional 42,000 since then. In 2019, an average of 190 clinical trial papers related to oncology were published each month.

Evidence-based medicine (EBM) relies on providing the right information for the right patient at the right time to guide clinical decision making.[3] Given the volume and expansion of the literature, and increasingly complex models of cancer, it is not surprising that clinicians often struggle to keep up with developments in oncology. Human cognitive capacity limits the amount of information that a clinician can process during clinical decision making.[4-6] Furthermore, manual extraction of data from scientific literature by humans can lead to varying interpretations.[7] Clinicians often rely on systematic reviews and guidelines to help inform decision making; however, such summaries often lag primary research reporting. Sites that publish clinical data that require human verification may further slow clinician access to information.[8]

Although chemotherapy drug and regimen information can be found on many websites such as PubMed, UpToDate, and HemOnc, manually identifying and synthesizing relevant content can be time-consuming. To fill the gap between knowledge generation and efficient incorporation into practice, clinicians need solutions that provide relevant information within their daily workflow environment.

To bridge that gap, we developed an artificial intelligence (AI)-assisted system to automate analysis of scientific literature in oncology. The system is capable

ASSOCIATED CONTENT

Appendix

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

Accepted on November 20, 2020 and published at ascopubs.org/journal/cci on January 13, 2021: DOI https://doi.org/10.1200/CCI.20.00087

CONTEXT

Key Objective

In the current study, we report the development and evaluation of an artificial intelligence (AI)-driven informatics pipeline to extract, interpret, and summarize published clinical cancer research in the context of a particular patient characteristic. The pipeline includes a mixed and integrated technique approach that includes text mining, machine learning (ML), natural language processing (NLP), and natural language understanding (NLU).

Knowledge Generated

We show that it is possible to find influential articles with high accuracy using ML trained on language used in study abstracts and titles. The system achieved 0.7-0.9 F1 score on clinical cohort characteristics using current NLP/NLU techniques, demonstrating the feasibility of developing technologies that can provide value to clinicians in real-world scenarios.

Relevance

In the current environment of exponential explosion of clinical evidence, it is necessary to develop techniques and tools that can help clinicians find and understand relevant research in a timely manner.

of ranking and filtering clinical research for a specific clinical scenario in oncology, extracting and qualifying the relevant clinical outcomes, and matching the most relevant articles to a set of patient characteristics. In this work, we describe Watson Oncology Literature Insights (WOLI), designed to assist clinicians in the practice of EBM by identifying relevant and timely information in clinical oncology research from published, peer-reviewed literature.

Our AI-assisted tool provides clinicians with targeted resources to identify, summarize, and contextualize pertinent information from the literature and other trusted resources. Although machine learning (ML) approaches have been used to enable medical evidence searches, such as Quertle[9] and, more recently, Meta,[10] they are not designed to extract specific information from citations. We accomplished this task with a combination of focused text mining, ML, natural language processing (NLP), and natural language understanding (NLU) to extract, filter, and rank information from reliable sources.

Recent work in the field of ML and NLP/NLU has focused on aspects of our system, including ML to classify abstracts,[11] medication-attribute linkage in clinical narratives,[12] the identification of rigorous clinical research evidence,[13] or PubMed-wide concept annotations.[14] WOLI automatically contextualizes information contained in research reports to a specific patient scenario or cohort to provide targeted information to clinicians.[15,16] The system circumvents the signal-to-noise problem inherent in manual searches. This manuscript describes the system architecture and presents an evaluation of its performance.

METHODS

System Description: Information Pipeline

The system's information pipeline uses metadata and annotations contained in published reports and NLP and NLU to enrich the original records (Fig 1). These enriched publication records include attributes regarding the study or publication, the patient cohort, and clinically relevant therapy and outcome statements related to therapy assessments in the text. This includes interventions (eg, chemotherapy, radiotherapy, and/or surgeries) and the criteria used to assess intervention effectiveness. The effectiveness and outcome statements are then qualified as favorable or unfavorable, much as a clinician might do during a literature review. This information is then prioritized and summarized for the user.

Corpus Definition

The system initially considers all publications available in PubMed and narrows the content to articles that describe oncology clinical trials using standard query techniques, similar to the manual curation of information for systematic reviews. The query strategy is defined in Figure 2. This strategy produces a corpus of English-language records focused on cancer clinical trials specific for a tumor type, including meta-analyses, systematic reviews, and reports of surgery, chemotherapy, and radiotherapy (hereafter referred to as the corpus).

Corpus Ranking

To rank publications by clinical relevance, the system uses references cited in oncology guidelines provided by AIM Specialty Health, eviQ (Cancer Institute of New South Wales, Australia), National Comprehensive Cancer Network (NCCN), National Cancer Institute (NCI) Physician Data Query (PDQ) summaries, and HemOnc. The overlap between the content represented by each of these databases used by the system is shown in Figure 3.

The documents in the corpus are ranked by research influence using a gradient-boosted tree ML algorithm. This algorithm, trained on text found in titles and abstracts, considers papers referenced in at least two out of five of the above guidelines as positive and/or expert (see Fig 3, purple line). WOLI uses this ML approach applied to the text after n-gram representation term frequency-inverse document

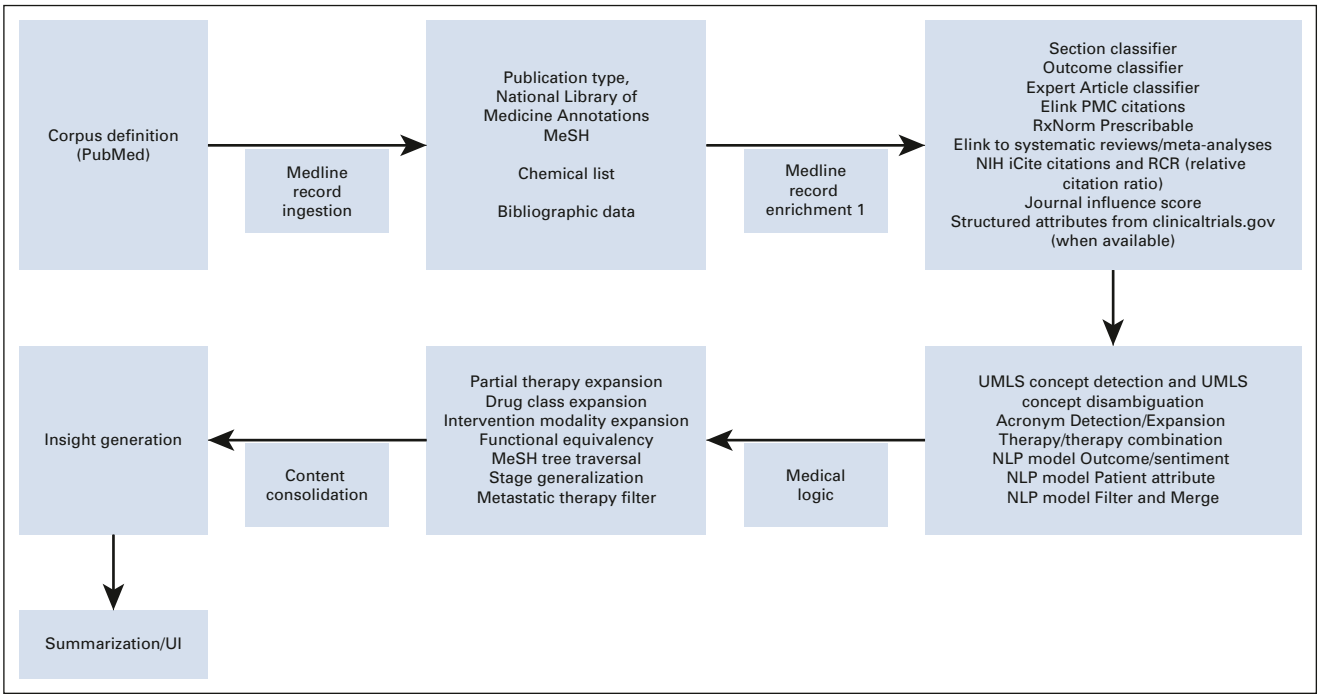

FIG 1. Graphical summary of WOLI information retrieval and record enrichment pipeline. MeSH, Medical Subject Header; PMC, PubMed Central; RCR, relative citation ratio; WOLI, Watson Oncology Literature Insights.

frequency transformation of the content in titles and abstracts.[17,18] The training and test set included papers that were part of the corpus; any papers cited in the guidelines that were absent from original corpus were added to it. The negative/nonexpert set was derived from papers published in journals with low journal influence scores, calculated as the ratio between the number of papers published by a journal and the average number of citations for any paper published in that same journal. The data set was split in an 80:20 ratio for training and testing. We used a 1:3 distribution of expert versus nonexpert publications in the data set to reflect the fact that nonexpert papers are typically more common than expert papers, taking into consideration the impact that an unbalanced data set can have on ML algorithms. The score allows the user to rank order the resulting publications.

Within the set of papers returned in the corpus, the WOLI Engine then extracts therapy characteristics (combinations of drugs, combined treatments of multiple modalities,

```
(
  ( "{TUMOR TYPE}" [mh] NOT "{TUMOR TYPE}/secondary" [mh:noexp] )
  AND (
    "clinical trial" [pt]
    OR "Meta-Analysis" [pt]
    OR "Randomized Controlled Trials as Topic" [mh]
    OR ( randomized[Title/Abstract] AND controlled[Title/Abstract] AND ( trial [tiab] OR trials
[tiab] ) )
  )
  NOT "Clinical Trial Protocol" [pt]
) OR (
  ("{TUMOR TYPE}/surgery" [majr] OR "{TUMOR TYPE}/radiotherapy" [majr])
  AND "Epidemiologic Methods" [mh]
  AND ("Meta-Analysis" [pt] OR ("Review" [pt] AND systematic [sb]) OR "Systematic Review" [pt]
  OR "Comparative Study" [pt])
)
  AND eng [la]
  AND hasabstract
  NOT "Economics" [majr]
  NOT "Retracted Publication" [pt]
  NOT "veterinary" [sh]
```

FIG 2. Query strategy used to access Medline records relevant to clinical research in oncology.

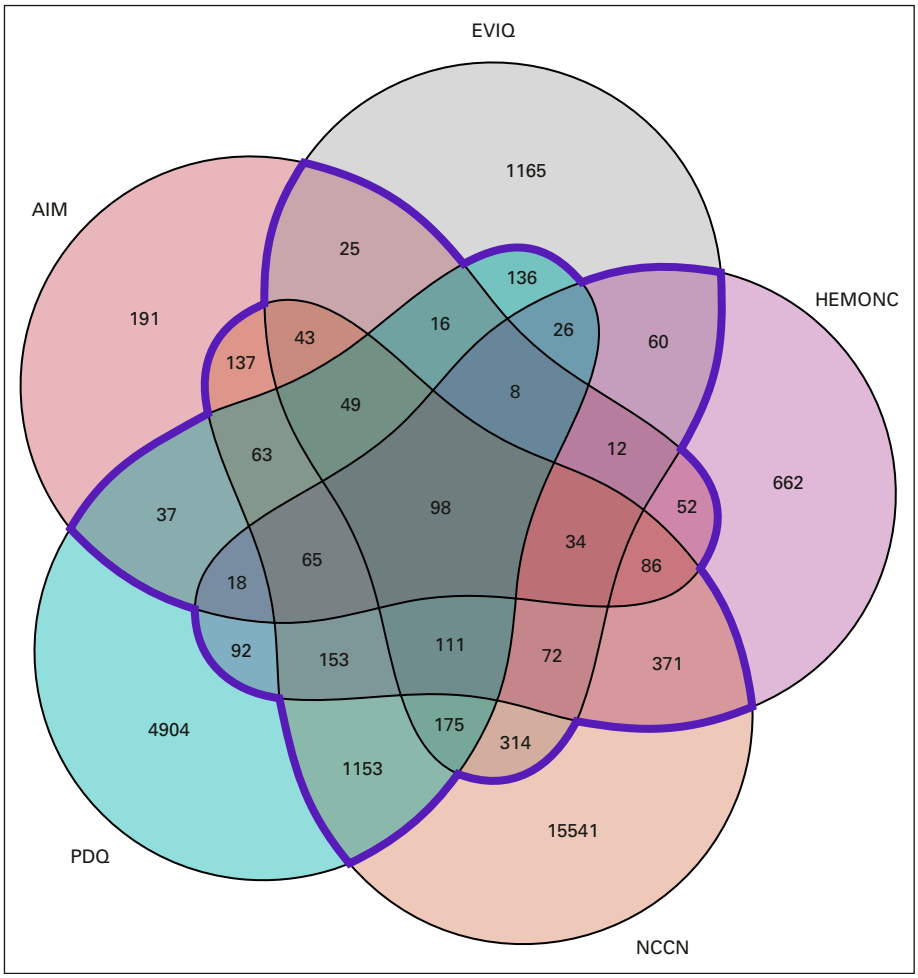

FIG 3. Venn diagram showing the size (in numbers) and overlaps between references cited in different cancer guidelines used as sources in this study, including aim (AIM Specialty Health Clinical Guidelines), hemonc (hemonc.org), pdq (National Cancer Institute [NCI] Physician Data Query summaries), NCCN (National Comprehensive Cancer Network), and eviQ (Cancer Institute of New South Wales, Australia). Papers inside purple line represent those used in training of expert paper machine learning (ML) classifier.

appropriate settings for the treatments, dosage, and administration), patient cohort definitions, and comparison statements between different treatments (primarily based on outcomes).

Patient Cohort Definitions, Therapies, and Associated Outcomes

Patient cohort definitions were determined using NLP and existing MeSH headers associated with each document in Medline. NLP was used to extract therapy details and outcome comparisons. The system used a combination of concept detection based on Unified Medical Language System (UMLS) entities and shallow parsing rules to understand grammatical context. The output of the entity extraction phase was a set of therapies (identified using UMLS identifiers for later medical logic) and comparison statements based on outcomes such as overall survival and progression-free survival. The system uses ML classifiers to assign document-level context to phrases into sections (Background, Objectives, Methods, Results, and Conclusion) and select the sentences most likely to contain useful data (Table 1).

Compared to a text-mining approach based on number of times a particular word appears in a document or how closely sets of words appear together in text, the NLP approach uses syntactic and semantic context to provide more accurate and complete entities for later analysis.

Resolution of Acronyms

Acronyms, common in medical literature, are often nonstandard and defined internally within a publication. To resolve acronyms, we combined the Schwartz and Hearst algorithm and the Abbreviations Plus P-Precision (AB3P) algorithm.[19,20] Although their performance is comparable, there are gaps in each algorithm that are filled by the other. We combined these two algorithms into an ensemble classifier to resolve long-form/short-form pairs in the medical literature.

Identification of Therapies and Treatments

After entity extraction, we use well-documented ontologies, such as SNOMED, NCI thesaurus, RxNorm, and MeSH,[21] to identify relationships between entities, including cancer types, treatments, drug classes, etc. We use the identified relationships as basic medical knowledge during interpretation. This relational knowledge can be used to expand common chemotherapy regimen names into the specific therapies included in that regimen or can be used to link

TABLE 1. List of Attributes Extracted From Medline Records and Method of Extraction

Concept	Description	Source
Publication or study attributes		
Publication type	What type of study is being reported in the abstract	MeSH
Study size	Number of patients included in the trial	CTGOV/NLP
Subject therapy	Intervention tested	CTGOV/NLP
Object therapy	Intervention control	CTGOV/NLP
Outcome	Outcome or outcomes studied (ie, survival, response, toxicity, etc)	NLP
Clinical cohort attributes		
Diagnosis	Cancer site	MeSH/NLP
Histology	Cancer morphology	MeSH/NLP
Stage	Cancer stage	NLP
Mutation	Gene alterations (ie, EGFR-positive, KRAS-negative…)	NLP
Line of therapy	Setting of the study (ie, first line, second line…)	NLP
Age	Age group studies (ie, adult, elderly…)	MeSH
HER2 status	Biomarker group (ie, positive or negative)	NLP
Estrogen receptor status	Biomarker group (ie, positive or negative)	NLP
Progesterone receptor status	Biomarker group (ie, positive or negative)	NLP
Modality	Therapy setting (ie, adjuvant or neoadjuvant)	MeSH/NLP
Node_status	Lymph node status	NLP
Menopausal_status	Menopausal status	NLP
castration_resistant	Castration resistant status	NLP
platinum_sensitivity	Platinum sensitivity status	NLP
invasion_site	Anatomic invasion site (ie, muscle invasion in bladder cancer)	NLP
met_site	Anatomic metastatic site (site of metastases liver, brain, etc)	MeSH/NLP
microsatellite_instability	MSI status	MeSH
Ecog	Clinical status	NLP
t_stage	Tumor T category (TNM)	NLP
n_stage	Tumor N category (TNM)	NLP

Abbreviations: CTGOV, clinicaltrials.gov; MeSH, medical subject headings; NLP, natural language processing.

specific treatments (such as cisplatin and doxorubicin) with general references to the same treatments (chemotherapy) in the same abstract. This relational knowledge can be used to generalize treatment parameters and identify additional treatments beyond what is explicitly stated in the document.

The ML system discriminates between different types of information in an abstract, classifying information as introductory, methodologic, data- or results-related, as well as identifying conclusions drawn from results. WOLI recognizes different types of information and assigns a relative importance to disparate pieces of information in an abstract.

Identification of Treatment-Associated Outcomes

The Population Intervention Comparator Outcome (PICO) provides a framework to assess outcomes. The NLP model in WOLI identifies therapies and associated outcomes in an abstract such as overall survival, progression-free survival, and toxicity; outcomes are qualified as favorable or unfavorable, based on information contained in the model (eg, improved progression-free survival, favorable; increased toxicity, and unfavorable). This is preferable to approaches that retrieve information and quality outcomes without extracting the intended semantic role of therapies and outcomes. Our NLP/NLU approach uses the intended semantic roles in the abstract to value and prioritize the outcome statements.

Matching Evidence to Patients and Specific Treatments

Information contained in the system is matched to patient information and treatment(s) to identify publications relevant to a given patient and treatment of interest. The system can identify publications that are relevant to a patient cohort in the absence of a preconceived treatment to define a set of therapies for further consideration by clinicians, providing information tailored to an individual patient. Publications relevant for a patient are identified based on a match between patient attributes and the study cohort

attributes. Publications are excluded when values for one or more patient attributes do not match study cohort values for that attribute.

The system uses several approaches to match publications based on a treatment of interest. Publications that refer specifically to the treatment of interest are preferred; however, the system can also correlate publications with treatments when treatments in published studies are either more general or more specific than the treatment of interest (eg, when the treatment of interest is a platinum compound and a publication refers to cisplatin, a type of platinum compound). The system can also correlate treatments with published studies that refer to compounds or entities that are members of the same class of drug as the treatment of interest and when multiple publications are required to span all parts of a multifaceted treatment plan.

Performance Evaluation

The NLP/NLU system accuracy is measured by precision, recall, and the harmonic mean of precision and recall (F1 measure) based on a ground truth encompassing more than 3,100 outcome statements in more than 1,300 abstracts. Ground truth articles with an equal distribution across cancer types of interest were randomly selected from the corpus. Priority was given to articles with a high expert similarity score; articles occurring in published guidelines were given the highest priority. The curation team consisted of four medical NLP analysts with extensive experience building information retrieval and analysis applications for various types of cancers and a medical scientist participating in independent curation and review of the ground truth generated. To account for the inter-annotator variability that can arise in NLP ground truth generation, we adopted a formal review process. Concept- and statement-level ground truths were reviewed by consensus and approved by the medical expert. The resulting ground truth was the consensus-derived, expert-reviewed, concept-level annotation for each of the statements.

Each concept was measured independently, and precision, recall, and F1 measures were calculated. The expert similarity measure was independently evaluated.

RESULTS

Expert Paper Classification

The expert similarity measure was able to classify the papers in the test set with an accuracy of 0.93 (95% CI, 0.9 to 0.96, P value $< .0001$) and sensitivity and specificity of 0.95 and 0.91, respectively.[22] Using a decision boundary of 0.5 for the positive (expert) label, the classifier demonstrated an F1 score of 92% on the test set, according to the data shown in Table 2.[22] The system identified papers with a high expert classifier score (≥ 0.85) that were absent from the training set of papers cited in clinical guidelines; these could be interpreted as papers that have all the language characteristics of an expert paper but for reasons beyond the scope of the current work were not part of the papers cited in the guidelines (Table A1).

TABLE 2. Expert Classifier Accuracy

Category	Precision	Recall	F1 Score	Support
Nonexpert	0.90	0.99	0.95	1,490
Expert	0.98	0.80	0.88	765
Average/total	0.93	0.92	0.92	2,255

Concept-Level Accuracy

For NLP concept detection, evaluation accuracy varied among attributes, with F1 scores measuring between 80 and 98 (Table 3). Complex concepts that relied on acronym expansion or that were time-dependent, such as incomplete or context-dependent therapy, had a lower accuracy as compared to those that were not, such as HER2 or ER/PR status.

DISCUSSION

This paper is one of the first to present a fully automated system capable of continuously analyzing scientific literature from Medline and identifying the most relevant literature as it relates to the efficacy of oncology therapies in the context of a patient's characteristics. WOLI is efficient enough to be used at the point of care, facilitating the translation of science into practice.

In designing WOLI, we used the Enterprise Design Thinking Framework[23] to understand clinicians' processes for staying up to date and researching specific topics within the literature. We investigated clinicians' approach to finding

TABLE 3. Concept-Level Accuracy Metrics for Selected Clinically Relevant Attributes That Rely on NLP

Concept	F1 Score
Study size	0.86
Subject therapy combined/individual components	0.88/0.94
Object therapy combined/individual components	0.81/0.87
Outcome	0.83
Diagnosis (cancer site)	0.99
Histology	0.94
Stage	0.89
Line of therapy	0.91
HER2/ER/PR	0.92
Modality	0.94
Metastatic site	0.90
Menopausal status	0.93

NOTE. Subject and object accuracy are reported both at the correct combination of therapy level (combined) and at the individual component level (individual components).

Abbreviations; ER, estrogen receptor; NLP, natural language processing; PR, progesterone receptor.

relevant literature to help address "pain points" clinicians face today. We implemented the framework by conducting interviews, beta testing, and iterative prototyping, uncovering three primary pain points leading to the development of the WOLI system. First, the time spent searching across multiple tools for relevant literature is often at the cost of time spent with patients. Second, many cancer care providers struggle to stay up to date in the face of the rapid pace of literature expansion in oncology, and third, the average workday of a cancer care provider leaves little time to conduct literature searches and read research reports on a regular cadence. These points revealed the need for an AI-based, patient-specific evidence extractor from peer-reviewed literature.

Our research suggests that highlighting influential articles for cancer treatment and centralizing these in one location can benefit clinicians. Minimizing search time and the number of tools needed to search are essential features of a system designed to automatically extract information from the literature. To inspire trust and credibility, the system needs to be transparent and provide results that are filtered according to clinical relevance in a way that can be understood by clinicians using the system. We continue to use the Enterprise Design Thinking Framework[23] to iteratively evaluate, improve, and test the tool for future enhancements.

The engine draws upon resources in peer-reviewed journals and other curated resources, such as those available at Medline and HemOnc.org. HemOnc's ontology contained more than 30,000 relationships involving drugs and drug regimen in 2018. Roughly one-third of these relationships are related to RxNorm codes, synonyms for drugs, and references to supporting literature.[24] Although HemOnc.org is a dedicated cancer resource, editing privileges are limited to hematologists and oncologists who often cite limited time to add new content.[25] This gap can be filled by automated evidence-extraction methods such as WOLI. Furthermore, tools such as WOLI could become crucial in the editorial process for expert curated sources such as guidelines, among other use cases.

The method of ranking and filtering used in our approach has several advantages over methods that define relevance based on metadata or citation- and time-dependent metrics, such as citations. Search results can be dependent upon the method used for the search, as shown in Figure 3. This figure also highlights the variation inherent in human identification of relevant research for evidence-based guidelines.

The use of targeted NLP rather than standard text mining approaches to define cohorts provided many advantages in the development of the system. Targeted NLP extracts combinations of therapy agents and understands which combinations (many times containing the same drugs) resulted in which outcomes, and the system was able to differentiate those from parallel treatments. NLP enabled the identification of elements in a regimen that changed, as well as which elements were held constant. NLP determined treatments that were the subject and the object of the comparison phrase, assigning the proper polarity to the comparison. When combining treatments, the system could assign modifiers (high-dose or low-dose or adjuvant or neoadjuvant) to treatments, matching treatment plans with improved accuracy. As with most NLP systems, it finds negated and hypothetical mentions of drug agents, recognizing when a treatment lacks supporting evidence. This work revealed that the section heading under which a mention was found had an impact on the meaning of concepts. For example, the title commonly contains important treatments, methods sections typically contain details about the treatments prescribed, and results and conclusions sections typically contain statements about outcomes. When a document lacked section headings, the system used an ML classifier to assign headings to sentences in the document.

This study has several limitations. Outcomes may be potentially biased toward research involving chemotherapy, as compared to either radiation therapy or surgical approaches, due the preponderance of reports related to chemotherapy in the training set. Broadening the training set to include other radiotherapy- or surgery-focused guidelines may help to offset this limitation. The system is limited to information available in abstracts and publicly available bibliographic databases, where the most robust and fundamental findings of a study are typically emphasized. In addition, some of the documents that the system classified as false positives contained what may have been clinically relevant information, which may occur if the algorithm had access to more recent data than the one used for training.

In conclusion, this manuscript describes a system that extracts cohort-specific oncology treatment recommendations that have proven efficacy as shown in the medical literature, contextualized against other valid treatments for the same cohort. It efficiently replicates the process of literature identification and review that must be done by clinicians to use the most recent science. The approach narrows a broadly scoped corpus into a set of articles that are most likely to be of importance to experts and extracts detailed information from those articles, including the cohort attributes, detailed therapy combinations, and study outcomes. Such tools are critical to enable the practice of EBM in clinical areas with rapidly evolving science.

AFFILIATIONS

[1]IBM Watson Health, IBM Corporation, Cambridge, MA
[2]Department of Surgery, Vanderbilt University Medical Center, Nashville, TN
[3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN
[4]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN

CORRESPONDING AUTHOR

Fernando Suarez Saiz, MD, IBM Watson Health, 75 Binney St, Cambridge, MA 02142; e-mail: fernando.suarez@ibm.com.

PRIOR PRESENTATION

Presented in part at the ASCO 2019 Annual Meeting, Chicago, IL, May 31 to June 4, 2019.

SUPPORT

Supported by IBM.

AUTHOR CONTRIBUTIONS

Conception and design: Fernando Suarez Saiz, Corey Sanders, Rick Stevens, Michael Britt, Leemor Yuravlivker, Gretchen Jackson
Administrative support: Gretchen Jackson
Collection and assembly of data: Fernando Suarez Saiz, Corey Sanders, Rick Stevens
Data analysis and interpretation: Fernando Suarez Saiz, Corey Sanders, Rick Stevens, Michael Britt, Anita M. Preininger, Gretchen Jackson
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 the 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).

Fernando Suarez Saiz
Employment: IBM
Stock and Other Ownership Interests: IBM
Patents, Royalties, Other Intellectual Property: IBM

Corey Sanders
Employment: IBM
Stock and Other Ownership Interests: IBM
Patents, Royalties, Other Intellectual Property: I am inventor on several patent applications in process related to methods for the automatic processing of published literature, data extraction techniques, and applications of the extracted knowledge.
Travel, Accommodations, Expenses: IBM

Rick Stevens
Employment: IBM
Stock and Other Ownership Interests: IBM

Robert Nielsen
Employment: IBM, IQVIA
Stock and Other Ownership Interests: IBM

Leemor Yuravlivker
Employment: IBM
Stock and Other Ownership Interests: UnitedHealth Grp Inc, CVS Health, Gilead Sciences, IBM

Anita M. Preininger
Employment: IBM
Stock and Other Ownership Interests: Merck

Gretchen Jackson
Employment: IBM, Vanderbilt University Medical Center
Leadership: IBM
Stock and Other Ownership Interests: IBM
Speakers' Bureau: IBM
Research Funding: IBM
Travel, Accommodations, Expenses: IBM

No other potential conflicts of interest were reported.

REFERENCES

1. Lu Z: PubMed and beyond: A survey of web tools for searching biomedical literature. Database (Oxford) 2011:baq036, 2011
2. Statistical Reports on MEDLINE®/PubMed®. US Department of Health and Human Services. https://www.nlm.nih.gov/bsd/licensee/2012_stats/2012_LO.html
3. Ashley EA: The precision medicine initiative: A new national effort. JAMA 313:2119-2120, 2015
4. Walsh S, de Jong EEC, van Timmeren JE, et al: Decision support systems in oncology. JCO Clin Cancer Inform 3:1-9, 2019
5. Bossaerts P, Murawski C: Computational complexity and human decision-making. Trends Cogn Sci 21:917-929, 2017
6. Halford GS, Baker R, McCredden JE, et al: How many variables can humans process? Psychol Sci 16:70-76, 2005
7. Warner JL, Anick P, Drews RE: Physician inter-annotator agreement in the quality oncology practice initiative manual abstraction task. J Oncol Pract 9:e96-102, 2013
8. Rioth MJ, Osterman TJ, Warner JL: Advances in website information resources to aid in clinical practice. Am Soc Clin Oncol Educ Book:e608-e615, 2015
9. Giglia E: Quertle and KNALIJ: Searching PubMed has never been so easy and effective. Eur J Phys Rehabil Med 47:687-690, 2011
10. A free research discovery tool from the Chan Zuckerberg Initiative, Meta.org
11. Bao Y, Deng Z, Wang Y, et al: Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes. JCO Clin Cancer Inform 3:1-9, 2019
12. Li Q, Zhai H, Deleger L, et al: A sequence labeling approach to link medications and their attributes in clinical notes and clinical trial announcements for information extraction. J Am Med Inform Assoc 20:915-921, 2013
13. Kilicoglu H, Demner-Fushman D, Rindflesch TC, et al: Towards automatic recognition of scientifically rigorous clinical research evidence. J Am Med Inform Assoc 16:25-31, 2009
14. Wei CH, Allot A, Leaman R, et al: PubTator central: Automated concept annotation for biomedical full text articles. Nucleic Acids Res 47:W587-W593, 2019

15. Kilicoglu H: Biomedical text mining for research rigor and integrity: Tasks, challenges, directions. Brief Bioinform 19:1400-1414, 2018
16. Zhang M, Del Fiol G, Grout RW, et al: Automatic identification of comparative effectiveness research from medline citations to support clinicians' treatment information needs. Stud Health Technol Inform 192:846–850, 2013
17. Friedman JH: Greedy function approximation: A gradient boosting machine. Ann Stat 29:1189-1232, 2001
18. XGboost, Extreme Gradient Boosting [Computer software]. (2020). from https://xgboost.readthedocs.io/en/latest/
19. Sohn S, Comeau DC, Kim W, et al: Abbreviation definition identification based on automatic precision estimates. BMC Bioinform 9:402, 2008
20. Schwartz AS, Hearst MA: A simple algorithm for identifying abbreviation definitions in biomedical text. Pac Symp Biocomput 451-462, 2003
21. Bodenreider O: Biomedical ontologies in action: Role in knowledge management, data integration and decision support. Yearb Med Inform 67-69, 2008
22. Suarez Saiz FJ, Sanders C, Stevens RJ, et al: Use of machine learning to identify relevant research publications in clinical oncology. Am Soc Clin Oncol 37:6558-6559, 2019
23. West DM: Design Thinking: Key to Enterprise Agility, Innovation and Sustainability (ed 4). Las Vegas, NM, Authors Press Intl, 2017
24. Malty AM, Jain SK, Yang PC, et al: Computerized approach to creating a systematic ontology of hematology/oncology regimens. JCO Clin Cancer Inform 2, 2018
25. Warner JL, Cowan AJ, Hall AC, et al: HemOnc.org: A collaborative online knowledge platform for oncology professionals. J Oncol Pract 11:e336-e350, 2015

APPENDIX

TABLE A1. Examples of Papers With a High Expert Classifier Score (≥ 0.85) That Were Not Included in the Training Set

PMID	Title	Reference
32101663	Pembrolizumab for early triple-negative breast cancer	N Engl J Med 382:810-821, 2020
31851799	Olaparib plus bevacizumab as first-line maintenance in ovarian cancer	N Engl J Med 381:2416-2428, 2019
29658848	Neoadjuvant PD-1 blockade in resectable lung cancer	N Engl J Med 378:1976-1986, 2018
18086800	Survival after adjuvant oophorectomy and tamoxifen in operable breast cancer in premenopausal women	J Clin Oncol 26:253-257, 2008