Exploring the Efficacy of Generic Drugs in Treating Cancer

Ioana Baldini,[1] Mariana Bernagozzi,[2] Sulbha Aggarwal,[1,3] Mihaela Bornea,[1] Saksham Chawla,[1,4] Joppe Geluykens,[1] Dmitriy A. Katz-Rogozhnikov,[1] Pratik Mukherjee,[1] Smruthi Ramesh,[1,5] Sara Rosenthal,[1] Jagrati Sharma,[1,4] Kush R. Varshney,[1] Laura B. Kleiman,[6] Pradeep Mangalath,[6] Catherine Del Vecchio Fitz[6]

[1]IBM Research
[2]IBM Global Business Services
[3]Queens College, City University of New York
[4]Rensselaer Polytechnic Institute
[5]Northeastern University
[6]Reboot Rx

Abstract

Thousands of scientific publications discuss evidence on the efficacy of non-cancer generic drugs being tested for cancer. However, trying to manually identify and extract such evidence is intractable at scale. We introduce a natural language processing pipeline to automate the identification of relevant studies and facilitate the extraction of therapeutic associations between generic drugs and cancers from PubMed abstracts. We annotate datasets of drug-cancer evidence and use them to train models to identify and characterize such evidence at scale. To make this evidence readily consumable, we incorporate the results of the models in a web application that allows users to browse documents and their extracted evidence. Users can provide feedback on the quality of the evidence extracted by our models. This feedback is used to improve our datasets and the corresponding models in a continuous integration system. We describe the natural language processing pipeline in our application and the steps required to deploy services based on the machine learning models.

Repurposing Generic Drugs for Cancer

Each year nearly 10 million people die from cancer (Cancer Research UK 2020) and the cost of cancer diagnosis and treatment exceeds USD $1 trillion (Union for International Cancer Control 2014). Pharmaceutical research exploring new drugs to treat various cancers is an expensive and time consuming process. In contrast, there are many generic drugs available today that are inexpensive and show promising results in treating different types of cancers. Moreover, there are already several drugs that were successfully repurposed for cancer. For example, Thalidomide, a drug used to treat morning sickness in pregnant women, was proven useful for treating skin lesions and multiple myeloma. Finding new therapeutic uses for inexpensive generic drugs ("drug repurposing") could rapidly create affordable new treatments. Hundreds of non-cancer generic drugs have shown promise for treating cancer, but it is unclear which drugs to be considered for repurposing.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Scientific publications such as pre-clinical laboratory studies and small-scale clinical trials present evidence on generic drugs being used as cancer treatments. The Repurposing Drugs in Oncology (ReDO) project manually inspected articles indexed by PubMed and found anti-cancer evidence for more than 200 non-cancer generic drugs (Pantziarka et al. 2017; Bouche, Pantziarka, and Meheus 2017; Verbaanderd et al. 2017). However, PubMed indexes millions of articles and the collection is continuously updated. Therefore, manual review to identify and analyze the evidence is time-consuming and intractable at scale. It is imperative to devise (semi)automated techniques to extract and collate the existing evidence. Machine learning (ML)-powered evidence synthesis could provide a comprehensive and real-time view of drug repurposing data and enable actionable insights.

We have started an ambitious initiative to extract and synthesize the plethora of scientific evidence on generic drugs used for cancer treatment. Our goal is to identify the most promising drugs to repurpose for different kinds of cancer. Identifying drug-cancer evidence from scientific abstracts is not trivial. The articles that discuss cancer interventions use domain-specific jargon which makes the text hard to comprehend by both humans with non-expert background and machines that are not trained with domain-specific data (Lehman et al. 2019). This endeavor requires close collaboration between experts in different disciplines, such as cancer research (to provide guidance, annotate datasets, and verify results), machine learning (to collect and process data sets to be annotated, to devise machine learning models, and evaluate their performance), and software engineering (to deploy and run models as an end-to-end online application). Ultimately, implementing repurposed therapies as the standard of care in medical practice requires definitive clinical trials, new incentives and business models to fund them, and engagement by various stakeholders such as patients, doctors, payers, and policymakers.

In this paper, we highlight the key technical aspects of identifying and extracting relevant evidence from PubMed articles and describe the steps required to encapsulate, dis-

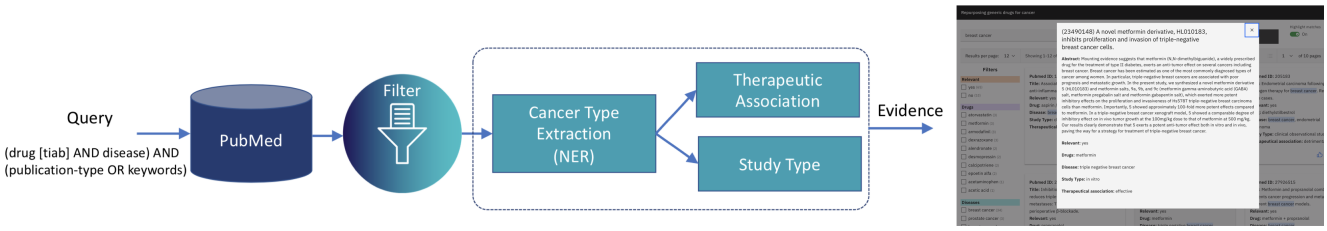

Figure 1: Method overview: The input to the evidence extraction pipeline is a list of non-cancer generic drugs, and the output is the published scientific evidence for each drug for treating various cancer types.

Retinoids can block cell proliferation and induce apoptosis in tumor cells. The antitumoral effect of synthetic retinoids like **Adapalene (ADA)** on **hepatoma** cells (HepG2, Hep1B) was investigated. **ADA** at 10(-4)M efficiently induced apoptosis, reaching 61.7% in HepG2 and 79.1% in Hep1B after 72 h incubation. This was accompanied by up-regulation of pro-apoptotic bax and caspase 3, while bcl-2 was down-regulated, shifting the bax/bcl-2 ratio to >2.3 in hepatoma cells. ***ADA*** *inhibits* ***hepatoma*** *cell growth in vitro and is a powerful inducer of* ***hepatoma*** *cell apoptosis.*

Figure 2: Sample *relevant* abstract annotation. PubMed # 15105045, *Adapalene* (in blue) is the non-cancer generic drug, used to treat *hepatoma* (in brown) cells (liver cancer). It is an *in vitro pre-clinical* study, and has an *effective* association. Evidence for association with phenotypic outcome measured is italicized.

play and access the results via a service deployed in the cloud.

NLP Pipeline for Drug-Cancer Discovery

We propose a natural language processing (NLP) pipeline that identifies the type of information in scientific publications relevant to our drug repurposing goal (Subramanian et al. 2020). A schematic view of the pipeline is presented in Figure 1. First, we query PubMed using queries inspired by the Cochrane highly sensitive search (CHSS) strategy (Dickersin, Scherer, and Lefebvre 1994) to narrow the collection of articles we analyze. For our purpose, we deem as relevant only the abstracts that discuss non-cancer generic drugs for cancer treatment and present results of phenotypic outcomes (e.g., tumor growth/reduction, patient survival, apoptosis). We start with filtering out the irrelevant documents [1] We are currently experimenting with models for abstract filtering based on different variants of BERT-based models (Devlin et al. 2019); our filtering accuracy is 90-95%, depending on the model used.

Next, cancer types and drugs are identified using named entity recognition (NER) and entity linking using ScispaCy (Neumann et al. 2019). The recall of our models is 90%. Finally, for each abstract that is deemed relevant, we classify the therapeutic association and the type of study. We

[1]Note that querying PubMed, even with a sophisticated query, may not yield only *relevant* articles. For instance, some abstracts discuss the mechanism in which the drug affects the cancer without discussing the actual effect on cancer; such documents are not relevant for our purpose.

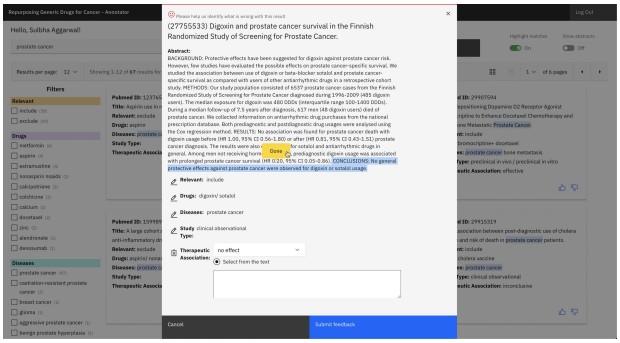

Figure 3: The interface can be used for browsing the drug-cancer evidence and also providing feedback on the quality of the predictions. In this example, a negative feedback is provided with a selection from the text as supporting evidence.

refer to the *drug, cancer, therapeutic association, and study type* as the **evidence** presented in the PubMed abstract. An example of the evidence in an abstract is shown in Figure 2.

The therapeutic association schema contains the following classes: A. *Effective*: the drug was shown to be effective for treating the cancer; B. *Detrimental*: the drug has a detrimental effect on the cancer; C. *No effect*: the drug has no effect on the cancer; D. *Inconclusive*: the results of the study are inconclusive. Due to the difficulty in finding articles that discuss studies that are non-effective (either with no effect, negative effect or inconclusive results), we collapse all initial classes into two: effective and non-effective. We are currently experimenting with models based on RoBERTa (Liu et al. 2019) and BERT (Devlin et al. 2019) variants. In the 2-class setting, the accuracy of the model varies between 75%-83% accuracy depending on the class.

The study types we consider are: 1. pre-clinical studies (in vitro, in vivo), 2. clinical case reports, and 3. clinical trials. We train random forest models with bag of words representations for the abstract text, publication type, and metadata that are available as tags for the PubMed abstracts. This task gets executed successfully with an accuracy of 95%.

Web Application

The model predictions are made available in a web application that allows browsing the abstracts with the extracted evidence (i.e., relevance, drug, cancer, therapeutic association and study type). Figure 3 shows a print screen of the

application.

A continuous integration pipeline has been built that trains each model, runs the model on a large subset of PubMed abstracts and stores the predicted information in an ElasticSearch database. To coordinate the model training and predictions by the model, we employ a Kubeflow pipeline (Kubeflow.org 2020) that has the following stages: first, the model that determines document relevance is trained; second, a validation stage that determines whether training was successful based on the performance results on a small, development dataset; third, a model for therapeutic association and the study type is trained and validated; lastly, the relevant model is used to select a large set of documents from PubMed, and the therapeutic association and the study type are predicted for this dataset. The results are stored in ElasticSearch, which acts as the application back-end.

Users can sign up to access the web application and are authenticated by a service during sign in. ElasticSearch APIs enable searching and browsing the documents in our collection using free-form queries. Users can enter search terms (drug name, cancer type) in the application search bar and the results are furnished by the ElasticSearch APIs. We allow role based access to the different components of the application; users with an 'Annotator' role can view whether the evidence from an abstract was deduced from 'annotations' or 'predictions', whereas the feature is hidden to users with a 'Practitioner' role. Based on the type of evidence (e.g., drug, cancer and its types, therapeutic association, study type), we implement different ways to filter the search results. For each abstract that we display, users can provide feedback expressing either agreement or disagreement with the extracted evidence. For providing disagreement, we further allow the user to correct any of the fields in the extracted evidence and provide the snippets from the abstract text to support their reasoning.

We intend to use this tool to not only make the automatically extracted evidence consumable by stakeholders, but also as a way to verify the information extracted by our models. We collect corrections and create a feedback loop that takes the corrected data and incorporates it back to the training datasets to be used for model retraining.

Conclusion

We showcase an end-to-end evidence discovery NLP pipeline that fetches potential candidate abstracts from PubMed for further evaluation. The goal is to identify non-cancer generic drug evidence for different cancer types. We make this evidence available in a web application that allows stakeholders to both consume the evidence and provide feedback on the quality of the extracted information.

Acknowledgments

This work was conducted under the auspices of the IBM Science for Social Good initiative and was supported in part by the Rensselaer-IBM AI Research Collaboration (part of the IBM AI Horizons Network). We are thankful for all the people that helped start the project and contributed with discussions and annotations: Shivashankar Subramanian (University of Melbourne), Diego Garcia Olano (University of Texas), Sejal Dua (Tufts University), Anastasia Spangler (Seattle University), Aleksandra Mojsilović, Karthikeyan Natesan Ramamurthy, Sushma Ravichandran, Prasanna Sattigeri (IBM Research), Tommy Bhangdia (Tufts University), Devon Crittenden (Reboot Rx), Emily Duffy (Bates College), Kelly Fan (Brown University), Sam Marchant (Colby College), Katie McKinley (Colby College), Kriti Sharda (University of Connecticut), Murshea Tuor (Tufts University), Emily van der Veen (Colby College).

References

Bouche, G.; Pantziarka, P.; and Meheus, L. 2017. Beyond Aspirin and Metformin: The Untapped Potential of Drug Repurposing in Oncology. *Eur. J. Cancer* 172: S121—-S122.

Cancer Research UK. 2020. Worldwide cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer. Accessed Nov. 30th, 2020.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*.

Dickersin, K.; Scherer, R.; and Lefebvre, C. 1994. Identifying Relevant Studies for Systematic Reviews. *BMJ* 309: 1286.

Kubeflow.org. 2020. Kubeflow pipelines. https://www.kubeflow.org/docs/pipelines/. Accessed Nov. 30th, 2020.

Lehman, E.; DeYoung, J.; Barzilay, R.; and Wallace, B. C. 2019. Inferring Which Medical Treatments Work from Reports of Clinical Trials. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 3705–3717.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Neumann, M.; King, D.; Beltagy, I.; and Ammar, W. 2019. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. arXiv:1902.07669.

Pantziarka, P.; Sukhatme, V.; Meheus, L.; Sukhatme, V.; and Bouche, G. 2017. Repurposing Non-Cancer Drugs in Oncology — How Many Drugs Are Out There? bioRxiv:197434.

Subramanian, S.; Baldini, I.; Ravichandran, S.; Katz-Rogozhnikov, D. A.; Ramamurthy, K. N.; Sattigeri, P.; Varshney, K. R.; Wang, A.; Mangalath, P.; and Kleiman, L. B. 2020. A Natural Language Processing System for Extracting Evidence of Drug Repurposing from Scientific Publications. In *IAAI*.

Union for International Cancer Control. 2014. The economics of cancer prevention and control. https://issuu.com/uicc.org/docs/wcls2014_economics_of_cancer_final?e=10430107/10454633. Accessed Nov. 30th, 2020.

Verbaanderd, C.; Meheus, L.; Huys, I.; and Pantziarka, P. 2017. Repurposing Drugs in Oncology: Next Steps. *Trends Cancer* 3(8): 543–546.