Artificial intelligence's essential role in the process of drug discovery

Edmon Begoli^{*,1} & Dimitri Kusnezov^{**,2}

¹Oak Ridge National Laboratory (ORNL), National Center for Computational Sciences (NCCS), 1 Bethel Valley Road, Oak Ridge, TN 37831, USA

²US Department of Energy (DOE), 1000 Independence Avenue, Washington, DC 20585, USA

*Author for correspondence: begolie@ornl.gov

**Author for correspondence: dimitri.kusnezov@hq.doe.gov

“Three specific benefits of AI-based approaches for the drug discovery process are: an automation of manual and labor intensive tasks, an ability to process, integrate and make sense of the large volumes of complex and disparate biomedical and regulatory data and an AI-enhanced process of knowledge discovery, which is especially important in the early phases of drug discovery”

First draft submitted: 21 August 2019; Accepted for publication: 25 September 2019; Published online: 15 October 2019

Keywords: artificial intelligence • deep learning • drug discovery • lead optimization • machine learning • natural language processing • oncology • precision medicine • protein folding

In its current state, the drug discovery process is a slow, expensive and complex process [1] that unfolds in a serial fashion, is dependent on the serendipity of intuition and often relies on trial-and-error methods with long and labor-intensive internal and external substages. Under the right circumstances, these characteristics are also opportunities for radical improvements through the application of appropriate artificial intelligence (AI) techniques. Note that the AI that we will focus on here is termed ‘narrow’ AI and it encompasses three main areas: deep learning [2], machine learning [3] and natural language processing [4]. AI will have a broader impact and will also be important to many other fields, ranging from smarter processes to development, but we will not comment on these aspects here.

Three specific benefits of AI-based approaches for the drug discovery process are: an automation of manual and labor intensive tasks, an ability to process, integrate and make sense of the large volumes of complex and disparate biomedical and regulatory data and an AI-enhanced process of knowledge discovery, which is especially important in the early phases of drug discovery.

Automation of manual & labor-intensive drug discovery tasks

Automating repetitive, complex and labor-intensive tasks is one of the most practical aspects of AI. In the context of drug discovery, there are plenty of areas where this quality of AI system has already shown significant results. These areas range from the enhancement and speedup of the early phases of drug discovery, such as target identification and lead optimization [5], to later, very labor-intensive phases such as contract reviews and legal negotiations [6]. In all mentioned cases, the unifying characteristic is that AI can offload menial tasks to machines and save valuable professionals’ time for more intellectually intensive tasks thereby saving significant resources by reducing processing time while minimizing errors.

Making sense of the large volumes of disparate data

The inherent challenge of interpreting and integrating an ever-expanding body of knowledge makes the drug discovery process both complex and slow. The potential combinations of molecular and system-level interactions are myriad, yet researchers are required to examine and synthesize all available published scientific research, including hundreds of thousands of scientific publications and reports as well as other knowledge sources (electronic health records, images, documentation etc.). This body of knowledge is critical; however, the scale is beyond the capacity of the human mind to fully examine, survey and synthesize. This is where AI can play a critical role. Several case

studies and available tools have demonstrated the utility of natural language processing and related text classification as well as understanding techniques for search and information discovery relevant to particular medical conditions, chemical structure properties or drug interactions and effects [7].

Fostering the process of knowledge discovery

Early phases of the drug discovery process, namely target identification and lead molecule optimization, are discovery intensive. In these phases, scientists, driven by intuition, experience and available information, first look for the appropriate targets or candidates for the application of therapies and then search for the most suitable molecular structures for the delivery of therapeutic compounds. In both instances, the process is very complex and intellectually intensive. The search for targets can take years of research or years of waiting for serendipitous intuition and it can take an equal amount of time or funds to identify and optimize the best lead molecules. In both instances, the case studies have shown that AI can accelerate the target identification process [5] and it can optimize the identification and optimization of lead molecule discovery. AI does so by searching through the past knowledge of the compounds, by examining a large variety of combinations and by recommending the most suitable leads.

To accelerate optimizing or targeting molecules and generally speed up the overall drug discovery process, AI methods hasten early-stage identification of which candidate compounds will likely prove to be effective at affecting the targeted disease. In particular, deep learning methods can explore existing data to help predict how tissue or body systems may respond to a given drug. Further, AI can play a role in the identification of target patient populations for clinical trials, in predicting disease progression based on molecular data obtained from tissue samples and in stratifying patients either in the clinical trials preparation phase or in the optimization of the treatments based on patients' responses and their individual characteristics.

In our own efforts, we operate in the partnerships that bring domain and AI expertise together, along with the use of curated and rich data modalities as the nucleus of our AI efforts. The initial steps in these partnerships are to create AI workflows that are built around the teams' disparate skills. The notable efforts are work on Traumatic Brain Injury (TBI), a public-private partnership (Accelerating Therapeutics for Opportunities in Medicine) [8] and a Department of Energy (DOE; DC, USA) - National Cancer Institute (NCI; MD, USA) partnership called JDACS4C [9]. In all of these, we have aggregated data in novel ways, developed AI workflows and created teams with skill sets that span AI, ranging from high performance computing and uncertainty quantification to medical and pharma specialists. The datasets include previously unavailable pharma data on small molecules, toxicity, animal tests, pharmacokinetics, cancer registries, genomic data, medical records, MRI and CT scans and so forth.

In the Accelerating Therapeutics for Opportunities in Medicine public-private partnership between the DOE, GlaxoSmithKline (London, UK) and the NCI, an AI-enabled workflow can learn new molecules that are better in all target metrics, including efficacy, toxicity, pharmacokinetics and developability. In the TBI effort, which is coupled with the TRACK-TBI dataset [10] from medical experts at University of California, San Francisco (USA) and AI experts at DOE, many new TBI modalities have been found. Efforts to tie-in known patient health trajectories are focused on helping to predict treatment paths that can be much more suited to individual injuries. Finally, the DOE-NCI partnership has focused on three key areas: analyzing population data from cancer registries, molecular pathways for RAS [11] and patient health trajectories from cell lines. In an integrative way, this effort is aimed at understanding the treatment efficacy at various scales and modalities of data.

The ongoing needs & anticipated challenges

Current AI approaches are all data centric. In other words, they are trained to execute their functions by using labeled data – data annotated by experts – to recognize features of the phenomena of interest. The process of labeling training data is manual, labor intensive and requires having both the right datasets and the subject matter experts to annotate them. Meeting each of these conditions is usually challenging, which is an even greater problem if we are to make this process repeatable and scalable.

A phrase often associated with AI-based decision making is 'black box'. This is due to the general lack of understanding about how deep learning architectures make the predictions that they do [12]. While these issues are less of a concern in everyday casual scenarios, in highly regulated and high-consequence domains such as precision medicine and drug discovery these issues can pose an insurmountable challenge. For that reason, there is a need for a comprehensive scientific validation and regulatory scientific framework that addresses the limitations of and quantifies the uncertainties related to AI-based decision making and that provides a set of progressive and comprehensive rules about the validation, verification and execution of an AI-enabled drug discovery process [13].

Finally, there is likely to be a significant need to overcome resistance and to facilitate the cultural adjustment in the pharmaceutical, medical, industrial and regulatory community about the role of AI. Some of this resistance will likely be due to anxiety associated with delegating any important decision making authority to AI systems, about the possible job-displacing role that intelligent machines can play and about the distrust in reliability and validity machine-produced results. All of these factors will require a comprehensive and concerted effort in communicating, educating and demonstrating the value, benefits and unique and essential role that intelligent machines can play in the process.

Promises abound

Despite the complexities at hand, the promise of AI-enabled drug discovery far outweighs the challenges described here. Clearly the process of drug discovery requires improvements, but the opportunities to do so with AI are manifest. AI-assisted automation, error reduction and knowledge discovery are all functions with a high degree of readiness for deployment and ripe for further research and development. Thus, we suggest that AI is not simply a scientific promise, but rather it is here today, quickly evolving and will become an indispensable, practical tool that is essential to the evolving drug discovery and precision medicine efforts.

Financial and competing interests disclosure

This manuscript has been in part co-authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the US Department of Energy. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. Paul SM, Mytelka DS, Dunwiddie CT *et al*. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9(3), 203–214 (2010).
2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 521(7553), 436–444 (2015).
3. Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications. *Science* 358(6370), 1530–1534 (2017).
4. Friedman C. Discovering novel adverse drug events using natural language processing and mining of the electronic health record. In: *Artificial Intelligence in Medicine. AIME 2009. Lecture Notes in Computer Science.* (5651), Combi C, Shahar Y, Abu-Hanna A (Eds). Springer, Heidelberg, Germany, 1–5 (2009).
5. Matheson R. Automating molecule design to speed up drug development. MIT News, MA, USA (2018). <http://news.mit.edu/2018/automating-molecule-design-speed-drug-development-0706#targetText=Designing%20new%20molecules%20for%20pharmaceuticals,discovery%20relies%20on%20lead%20optimization>
6. EurekaAlert. Semi-automated contracting system to accelerate partnering in AI-powered drug discovery (2019). http://www.eurekaalert.org/pub_releases/2019-07/imi-scs072319.php
7. Manica M, Cadow J. Novel AI tools to accelerate cancer research. IBM Research Blog, NY, USA (2019). <http://www.ibm.com/blogs/research/2019/07/ai-tools-for-cancer-research/>
8. ATOM Consortium (2019). <https://atomscience.org/>
9. National Cancer Institute, Center for Biomedical Informatics and Information Technology. Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) (2019). <https://datascience.cancer.gov/collaborations/joint-design-advanced-computing>
10. University of California, San Francisco. TRACK-TBI. (2019). <https://tracktbi.ucsf.edu/>
11. Downward J. Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* 3(1), 11–22 (2003).
12. Castelvechi D. Can we open the black box of AI? *Nature* 538(7623), 20–23 (2016).
13. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat. Mach. Intell.* 1(1), 20–23 (2019).