Patterns



Editorial

Al in drug discovery: Applications, opportunities, and challenges

Artificial intelligence (AI) is an umbrella term for a whole host of different approaches and technologies that revolve around the use of advanced computational approaches that can emulate (parts of) human decision making. Today, the discipline within AI that receives most attention and acclaim is machine learning (ML), focused on enabling computers to learn from experience as opposed to following a direct programing that instructs them what to do. Over the past years, AI (and especially ML) has increasingly found its way into healthcare in general, and drug discovery and development in particular. From drug-target interaction modeling to drug repositioning, AI augments the capabilities of human researchers, often outperforming alternative methods. Yet, access to large amounts of high-quality data for model training is a key bottleneck that needs to be addressed to unlock the full potential of AI drug discovery.

A first important caveat is that there is not one AI/ML approach to drug discovery, but dozens to hundreds—there are many different branches of ML (e.g., supervised and unsupervised learning, reinforcement learning), different algorithmic approaches within those branches (e.g., linear regression, neural networks, support vector machines), and different application areas (e.g., target identification, generative chemistry, patient stratification), which can then be applied across indication areas from rare and ultra-rare all the way to highly prevalent diseases. In other words, it is a vast space, which also evolves very rapidly.

In this context, we created a special collection, "The rise of artificial intelligence in drug discovery: Challenges and opportunities," and brought together researchers from academia and industry to share their cross-boundary Al-driven solutions in the format of original research and review articles that assist the drug discovery field in arriving at a new era. The special collection can be visited here. This special collection brings together a selection of publications that exemplify the progress made in this space and highlight areas of active research. Given the background of Patterns, these publications all lean very heavily on the data side, investigating in depth how best to leverage the power contained within biological and clinical data. This is probably the most important lesson we can learn from the articles in this collection: any AI or ML approach can only ever be as good as the underlying data. Just like human doctors and scientists, any artificially intelligent system also (or even more so) relies on high-quality, reliable, and representative data that enable robust decision making.

Four publications (*Patterns* 100396, 100496, 100441, and 100433) focus on the application of Al tools for drug repurposing/repositioning. Identifying potential indications where existing drugs could have a clinical effect benefiting patients is a particularly attractive and de-risked proposition, as it considerably shortens the time to approval—given that the drugs under investigation for repositioning are already approved. Moreover, the high costs of developing an entirely new therapeutic agent can be significantly reduced when the same agent can be used for

multiple indications. The COVID-19 pandemic saw several high-profile attempts at drug repurposing, not all of which have been successful, highlighting the need for innovative approaches with robust positive predictive value. The COVID International Research Team used two ML approaches: one to rank broad-spectrum antivirals in their likely activity against SARS-CoV-2 itself and another to rank FDA-approved drugs in their likely activity against parts of the human (i.e., host) interactome that play a role in mediating the SARS-CoV-2 effect. Their novel methods and tool help the exploration of drugs and their molecular targets using heterogeneous sources of information with multiple layers of interconnection. It is particularly encouraging to see several drugs predicted by their work to already be under clinical investigation. Although they have focused their analysis on COVID-19 (due to obvious reasons considering the time of publication), their proposed methods and tool can prospectively be applied in any drug repurposing context, thereby creating more ubiquitous AI solutions and tools that can be adapted for multiple purposes in different domains.

A team from PrecisionLife, an Oxford-based biotechnology company, focused on leveraging high-resolution patient stratification to identify opportunities for repositioning, using an approach based on multi-gene disease signatures (as opposed to single mutations) that can identify a second indication for a given drug-indication pair. On the other hand, a study led by Ping Zhang from Ohio State University uses a different approach for repurposing: computing gene expression changes caused by more than 10,000 existing drugs based on their SMILES codes, and comparing these profiles with those belonging to known treatments, they were able to identify and partially experimentally validate potential treatment options for pancreatic cancer, a neoplasm with dismal survival rates overall. The authors proposed this approach as a new perspective for precision drug discovery. Another disease where progress on new treatment options has been very slow is Alzheimer's disease. Showing the synergy between dry and wet lab approaches, researchers at the Fraunhofer Institute for Algorithms and Scientific Computing in Germany used a biological systems approach that includes different steps from literature mining and subgraph detection to target selection and screening. They used this approach to identify potentially druggable modulators of the phosphorylation of microtubule-associated tau protein as targets, then used a conventional screen to identify chemical hits that lead to inhibition of their chosen target, HDAC6.

Changing gears, a group of researchers from both academia and industry prepared a thorough review of ML applications for genomics data, which is growing rapidly thanks to the significant reduction in sequencing costs. They identified and classified 22 applications for ML in genomics across the entire therapeutics pipeline, from discovering novel targets and therapeutics to clinical and post-market studies. They argue that the key to unlocking the potential of genomics data will be to combine them with







other types of data, moving from silos to an interconnected, integrated perspective on biomedical data, providing levels of insight that otherwise would be impossible to obtain. We can find that this principle—of moving from silos to interconnected data-holds true in other contexts as well: when modeling drug-target-interactions, Ruan et al. found that moving from simplified one target-one drug models to higher-order connections (that allow for more than pairwise connections and for additional associations between targets, drugs, and diseases) significantly improves the predictive capacity of these models. Their study used hypergraph and neural networks, leading to an improved performance compared with other computational models. However, the authors do comment on the scarcity of high-quality, large-scale datasets being an impeding factor in further developing and improving their models.

Finally, a group at Imperial College London studied the use of bond-to-bond propensity, a method that enables the modeling of an effect on one part of the protein (especially the catalytically active or orthosteric site that drugs are traditionally developed for) on other parts of the protein-where allosteric modulators can bind and exert their effects on the protein. Allostery-based drugs offer several advantages over those developed for the orthosteric site-above all, they act on less evolutionarily conserved regions, thereby reducing the likelihood of off-target effects, and they can be used to modulate their target of inter-

est's activity, as opposed to only leading to complete inhibition or activation. Research into allostery is growing rapidly, and studies like this will enable researchers around the world to find allosteric modulators faster.

In summary, in this special collection we have seen a range of topics and techniques, all of which will hopefully see widespread adoption in both academia and industry - even more so in a time where technology-enabled and Al-driven drug discovery are gaining more and more traction and starting to have a tangible impact on drug discovery processes. Clearly, our increasing ability to develop, deploy, and continuously improve AI methodologies at scale bodes well for the future of Al-driven drug discovery, leading to better drugs to improve patients' lives globally.

DECLARATION OF INTERESTS

M.I.B. is a co-founder, shareholder and director of Arctoris.

Martin-Immanuel Bittner

Arctoris and the Young Academy of the German National Academy of Sciences

Sahar Farainia

Scientific editor. Patterns

https://doi.org/10.1016/j.patter.2022.100529