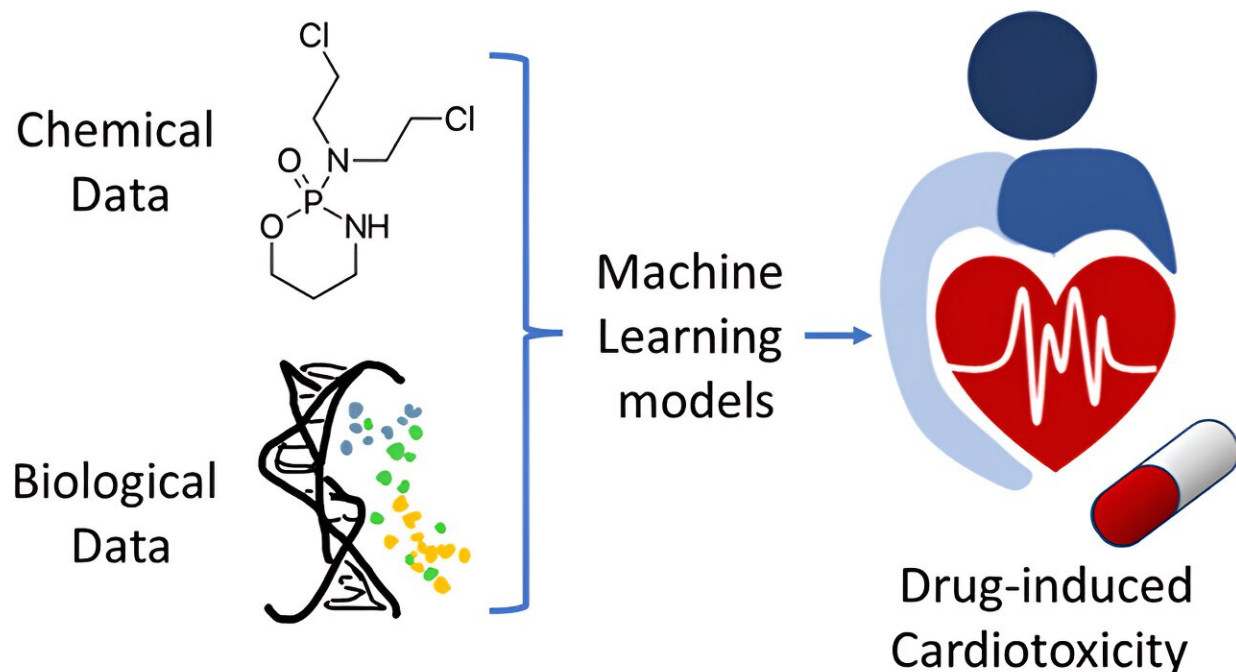


De-risking drug discovery with predictive AI

July 18 2024, by Claire Hendershot



DICTrank Predictor

Credit: *Journal of Chemical Information and Modeling* (2024). DOI: 10.1021/acs.jcim.3c01834

Developing a new drug can take years of research and cost millions of dollars. Still, more than 90% of drug candidates fail in clinical trials, with even more that never make it to the clinical stage. Many drugs fail because they simply aren't safe.

Researchers at the Broad Institute of MIT and Harvard have developed AI models that can screen the potential biological effects of drugs before they ever enter a living organism.

Srijit Seal, a visiting scholar at the Carpenter-Singh Lab in the Broad's Imaging Platform, trained multiple predictive machine learning models to identify chemical and structural drug features likely to cause toxic effects in humans.

Together, the tools estimate how a drug may impact diverse outcomes of interest to drug developers: general cellular health, pharmacokinetics, and heart and liver function. As of now, papers describing three of these machine learning tools have been published, in the [*Journal of Chemical Information and Modeling*](#), [*Molecular Biology of the Cell*](#), and [*Chemical Research in Toxicology*](#). A fourth is in the works.

Predictive models don't eliminate laboratory experiments, but they can help researchers narrow the selection pool of potential drugs, allocating more time and resources to experiment on the more promising candidates.

Seal began this work after wondering if more toxicology insights could be gleaned from a drug candidate's chemical structure. Drug toxicity can be an issue even after FDA approval; drug-induced cardiotoxicity (DICT) and drug-induced liver injury (DILI) each contribute to a significant percentage of post-market drug withdrawals.

To better understand the complex biological mechanisms that make drugs toxic to [human organs](#), the FDA has curated categorical lists of drugs' likelihood to cause toxic effects in the heart and liver.

"Since the FDA released these datasets, we wondered if we could use them to predict toxicity using machine learning," said Seal.

Seal used these FDA-curated lists as training data for two toxicity-predicting machine learning models: one for cardiotoxicity and one for liver injury. With additional inputs of chemical structure, physicochemical properties, and pharmacokinetic parameters, the models learned to identify features that contribute to [drug toxicity](#). The cardiotoxicity predictor, DICTrank Predictor, is the first predictive model of the FDA's DICT ranking list.

Often structurally similar compounds have different effects on liver function in animals and humans, and this is why DILIPredictor had the extra challenge of needing to differentiate toxicity between species. DILIPredictor correctly predicted when compounds would be safe in humans, even if the same compounds were toxic in animals.

Drug developers also assess pharmacokinetic effects, or how an organism absorbs, distributes, metabolizes, and clears a drug. It's crucial to determine these properties as early as possible: drugs that don't distribute to the desired target aren't efficacious, whereas drugs that stay in the body for too long can induce toxic effects.

Pharmacokinetic modeling is difficult, time-consuming, and requires expensive instruments and software. Predictive machine learning could provide a way for researchers to "fail faster" and focus their experimental efforts on the drugs with the best bioavailability. To help achieve this, Seal has been working with collaborators to develop a predictive pharmacokinetic modeling tool.

"Machine learning in pharmacokinetics is becoming popular," said Seal. "We wondered if we could design a predictive model and compare it to industry models, for now at least as a proof-of-concept.

"Drug design needs some kind of feedback loop to ensure that what you're designing is actually going to work in the human body and not

cause unintended toxicity," he added. This suite of predictive machine learning tools, if applied in early drug discovery, could provide the framework for that loop.

Another aspect of drug toxicology is related to cell health. When [machine learning](#) models predict a potential impact for a compound, researchers often want more detail, such as the mechanism by which the compound is impacting cells. Seal then turned to features extracted by CellProfiler, an open-source imaging software for interpreting cellular morphological features.

"CellProfiler looks at the physical features of cells as image-based data and tries to predict how they have changed with respect to a control," Seal explained. "When we asked industry biologists how they worked with CellProfiler data, they told us that sometimes they didn't know how to interpret these image-based features in a biological context."

To make CellProfiler data more biologically interpretable, Seal developed BioMorph, a deep learning model that combines CellProfiler's imaging data with data on cell health, such as the rates at which cells grow and multiply.

Training on two complementary datasets allows BioMorph to infer how a particular compound's mechanism of action could affect cell health. When BioMorph was tested on data outside of its training set, the model correctly matched compounds with the cellular features affected by that particular compound.

"BioMorph provides further detail that scientists can read and understand from a biological point of view," said Seal. "We're looking forward to hearing people's feedback on using BioMorph for their individual test cases."

More information: Srijit Seal et al, Insights into Drug Cardiotoxicity from Biological and Chemical Data: The First Public Classifiers for FDA Drug-Induced Cardiotoxicity Rank, *Journal of Chemical Information and Modeling* (2024). [DOI: 10.1021/acs.jcim.3c01834](https://doi.org/10.1021/acs.jcim.3c01834)

Srijit Seal et al, From pixels to phenotypes: Integrating image-based profiling with cell health data as BioMorph features improves interpretability, *Molecular Biology of the Cell* (2024). [DOI: 10.1091/mbc.E23-08-0298](https://doi.org/10.1091/mbc.E23-08-0298)

Srijit Seal et al, Improved Detection of Drug-Induced Liver Injury by Integrating Predicted In Vivo and In Vitro Data, *Chemical Research in Toxicology* (2024). [DOI: 10.1021/acs.chemrestox.4c00015](https://doi.org/10.1021/acs.chemrestox.4c00015)

Provided by Broad Institute of MIT and Harvard

Citation: De-risking drug discovery with predictive AI (2024, July 18) retrieved 18 July 2024 from <https://phys.org/news/2024-07-de-drug-discovery-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.