Patterns

CellPress
OPEN ACCESS

Preview

Predicting drug toxicity at the intersection of informatics and biology: DTox builds a foundation

Matthew J. Sniatynski[1,2] and Bruce S. Kristal[1,2,*]

[1]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, 221 Longwood Ave, LM322B, Boston, MA 02115, USA

[2]Division of Sleep Medicine, Harvard Medical School, Boston, MA 02115, USA

*Correspondence: bkristal@bwh.harvard.edu

https://doi.org/10.1016/j.patter.2022.100586

Hao et al. (2022) present DTox (deep learning for toxicology), a neural network designed to predict and probe the sites and potential mechanisms underlying chemical toxicity; results provide a map to facilitate modular testing and improvements across multiple disparate applications.

Most, if not all, readers of *Patterns* would likely agree that progress in medicine in the coming years will increasingly depend on improved use of data resources. Historically, each advance in both data acquisition and analysis has improved our insight into disease and our ability to predict, diagnose, and treat it. This historical trend ranges from "simple" advances in clinical chemistry, such as the ability to measure glucose (diabetes), to modern methods that identify genetic abnormalities predisposing individuals to diseases (e.g., BRCA1 mutations and breast cancer[1]). Critically, this historical trend is equally apparent in data analysis—witness examples such as the role of statistics in seeking to establish the role of chance versus signal (e.g., clinical trials), information theory, signal processing (establishing the role of randomness and the limitations imposed by noise on detectable signal), the progressive improvements in regression-based approaches (i.e., from linear regression to least angle regression), and other modeling approaches (e.g., projection methods, trees, ensembles, SVMs, and neural-networks, including their use in deep learning and AI). Modern biomedical research would be crippled without these breakthroughs. Thus, substantial historical precedent suggests that new informatics technologies will open new insights and approaches into human health and disease. In this issue of *Patterns*, Hao et al.[2] provide just such an advance to help predict not only which potential drug candidates will have side effects but where such toxicity will manifest at the level of the individual, the organ system, and the cell.

Improving toxicity predictions has important implications for society (e.g., costs, delayed development, abandoned programs), for the individual (e.g., trial subjects), and for the technology itself, as better understanding can often be directly parlayed into improving drug candidates/pharmacophores, co-development, or other regimens. *In silico* drug screening has had successes in both increasing primary hit rates and predicting some drugs' toxicity, but the latter are often black-box models that provide little added actionable (e.g., mechanistic pathways) information. Approaches designed to identify the elements of models that contribute to class discernment (e.g., LIME,[3] Shapley[4]) are powerful for recognizing the mathematical drivers behind such classification, but they do not necessarily provide useful domain specific information—a facet of models whose importance is increasingly recognized.[5]

Hao et al. approach this problem by developing DTox, an interpretation framework for knowledge-guided neural networks designed to predict compound response to toxicity assays and infer toxicity pathways of individual compounds. DTox uses the “interpretability paradigm" explicitly as an exploratory tool; DTox not only encodes assumptions at the beginning and then moves forward but examines how these assumptions are supported or violated by the model in action. This allows knowledge to be gathered from the data in instances where the model fails as well as where it succeeds. Starting with compound chemical identifiers, DTox correctly predicts downstream transcriptional patterns of aromatase inhibitors and PXR (pregnane X receptor) agonists, and upstream events portending distinctive paths of HepG2 cytotoxicity. The pathways explored and identified are already known in general. Much of the work focuses on showing that they can recapitulate, in minutes, knowledge (i.e., specific drug side effects/mechanisms) historically gained by hard, difficult, slow, costly work. Rather than doing this by directly training on examples and simply classifying, they construct a neural network with layers informed by the knowledge of biological pathways.

The article by Hao et al. is not important because of the described method's predictive accuracy; indeed, there is little objective advantage to be gained—today—by using DTox over other toxicity predicting algorithms[6,7]; DTox barely exceeds chance performance on many key metrics (especially when taken across the entire series) and produces many false positives and false negatives. Rather, the importance lies in the demonstration that the "guts" of the method, deep-learning performed atop the visible neural network (VNN) framework, can achieve similar accuracy to other state-of-the-art methods, despite the fact that they are severely constraining the trained form that the neural network "connections" are allowed to take. This opens the door to a completely different conception of how machine learning can operate, where existing "prior" knowledge is explicitly encoded and new data/evidence is assessed against it.

Box 1. Building off DTox: A very incomplete list …

- Shift targets
  - New drug classes
  - Emphasize upstream pathways
  - Emphasize final end-stage pathways
  - Systematically examine where DTox failed
  - Predict different intermediate outcomes, e.g., metabolic changes
- Shift inputs/training data
  - Add 3D chemical descriptors
  - Use more or different assays
  - Continued exploration of compounds in training set
- Modify algorithms or neural network structure
  - Extend or alter a given domain (layer)
  - Alter, change, or subtract connections/constraints
  - Change pathway maps or usage, within or beyond Reactome
- Leverage outside improvements
  - AI/neural network advances
  - Improvements in TargetTox
  - Improved detail in biological pathways
  - Available datasets (chemistry, -omics, clinical, drug toxicity, etc.)

Indeed, it is arguably in the peculiarities of DTox's successes, in DTox's failures, and in *recognized* limitations of the biological aspects of their system where the greatest potential advances may lie. As the biological/chemical coding within and between the layers of the neural network improves (additional work, added domain knowledge; see Box 1), this type of information structure may be leveraged to identify both the upstream feeder paths and the downstream effector paths, helping to better understand the web of interactions between and within the different scales. This offers great potential for a more complete, explicit characterization of these pathways of interest and, accordingly, a better ability to recognize, characterize, understand, and prevent toxicity. Similarly, while DTox was successful for three toxicity classes, it failed to detect most of the classes tested. But, DTox demonstrated successes and an ability to elucidate both upstream and downstream effects; it is reasonable to assume that future work will enable detailed investigation of which biological/chemical/data science issues underlies its "failures." For example, training examples or the pathway coding may be inadequate, either because of the network structure or critical nodes being missing (potentially either informatics or domain-specific issues). DTox provides a foundation that can be systematically altered/expanded in many essentially modular ways to address myriad experimental questions (see Box 1).

DTox seeks domain-specific, mechanistic information, but Hao et al.'s approach may also directly fuel quantitatively improved recognition of potentially dangerous drugs. Specifically, a particular interest raised by this study is the potential use of system-level information fusion for drug discovery. It is recognized that diversity between mathematical models can be more important than accuracy for successful fusions,[8] and the importance of diversity in fusion has been specifically demonstrated for *in silico* approaches based on intact molecules[9] and binding motifs.[10] DTox is conceptually orthogonal to existing *in silico* modeling approaches; the likely synergy enables improved modeling without requiring additional primary data.

So, sit down, bring a pad—paper or electronic—and read this paper. There will be some parts you like, possibly some you don't, and more importantly, many, many parts—successes, failures, pieces present, and pieces absent—that will inspire thoughts about how to move forward.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH: R01-HL140335 (B.S.K.).

DECLARATION OF INTERESTS

M.J.S. and B.S.K. have a filed US patent application related to Sniatynski et al. (2022). The patent is not directly related to Hao et al., but the two may be useful in tandem.

REFERENCES

1. Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L.M., Haugen-Strano, A., Swensen, J., Miki, Y., et al. (1994). BRCA1 mutations in primary breast and ovarian carcinomas. Science *266*, 120–122. https://doi.org/10.1126/science.7939630.
2. Hao, Y., Romano, J.D., and Moore, J.H. (2022). Knowledge-guided deep learning models of drug toxicity improve interpretation. Patterns *3*, 100565.
3. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). Why should i trust you? Explaining the predictions of any classifier. . Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) (Association for Computing Machinery), pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.
4. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of Advances in Neural Information Processing Systems, *30*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., pp. 4765–4774.
5. Mohanty, S.D., Lekan, D., McCoy, T.P., Jenkins, M., and Manda, P. (2021). Machine learning for predicting readmission risk among the frail: explainable AI for healthcare. Patterns *3*, 100395. https://doi.org/10.1016/j.patter.2021.100395.
6. Chen, X., Roberts, R., Tong, W., and Liu, Z. (2022). Tox-GAN: an artificial intelligence approach alternative to animal studies-a case study with toxicogenomics. Toxicol Sci. *186*, 242–259. https://doi.org/10.1093/toxsci/kfab157.
7. Jiang, J., Wang, R., and Wei, G.W. (2021). GGL-Tox: geometric graph learning for toxicity prediction. J. Chem. Inf. Model. *61*, 1691–1700. https://doi.org/10.1021/acs.jcim.0c01294.
8. Sniatynski, M.J., Shepherd, J.A., Ernst, T., Wilkens, L.R., Hsu, D.F., and Kristal, B.S. (2022). Ranks underlie outcome of combining classifiers: quantitative roles for *diversity* and *accuracy*. Patterns *3*, 100415. https://doi.org/10.1016/j.patter.2021.100415.
9. Yang, J.M., Chen, Y.F., Shen, T.W., Kristal, B.S., and Hsu, D.F. (2005). Consensus scoring criteria for improving enrichment in virtual screening. J. Chem. Inf. Model. *45*, 1134–1146. https://doi.org/10.1021/ci050034w.
10. Chen, Y.F., Hsu, K.C., Lin, P.T., Hsu, D.F., Kristal, B.S., and Yang, J.M. (2011). LigSeeSVM: ligand-based virtual screening using support vector machines and data fusion. Int. J. Comput. Biol. Drug Des. *4*, 274–289. https://doi.org/10.1504/ijcbdd.2011.041415.