molecular pharmaceutics

Editorial

Cite This: *Mol. Pharmaceutics* 2018, 15, 4311−4313

pubs.acs.org/molecularpharmaceutics

Artificial Intelligence for Drug Discovery, Biomarker Development, and Generation of Novel Chemistry

The productivity of the pharmaceutical industry is on the decline. Failure rates in clinical trials exceed 90% after therapies are tested in model organisms, and the cost to develop a new drug exceeds $2.6 billion. Recent advances in artificial intelligence (AI) may help to reverse this trend and accelerate and improve pharmaceutical R&D. While the term AI and the concept of deep learning are not new, recent advances in high-performance computing, the availability of large annotated data sets required for training, and new frameworks for implementing deep neural networks (DNNs) resulted in an unprecedented acceleration of the field. Since 2014, DNNs have surpassed human accuracy in image, voice and text recognition, autonomous driving, and many other tasks.

Early presentations to the pharmaceutical industry on the advances in deep learning in 2014 and 2015 resulted in skepticism and were discarded. In 2017, many pharmaceutical companies started partnering with AI startups and academics or started internal R&D programs. From training DNNs on transcriptional response data for predicting the pharmacological properties of small molecules[1] and biomarker development,[2] to the generation of novel chemistry, deep learning techniques rapidly propagated into many areas of biomedical research.[3]

The body of knowledge and the range of applications of deep learning and other machine learning techniques has expanded quickly and permeated into many areas of drug discovery. There have been hundreds of publications deposited in peer-reviewed journals and on ArXiv. In June 2017, Molecular Pharmaceutics announced a special issue titled "Deep Learning for Drug Discovery and Biomarker Development" focused on the applications of AI in chemistry and biomedicine (Figure 1).

After a call to the most prominent scientists publishing on deep learning in the areas of computational chemistry and biology, 10 research papers were accepted.

One of the main opportunities for AI in drug discovery is in drug repurposing using abundant data sets available from high-throughput experiments with gene expression profiles. Specifically, transcriptional response profiles generated by the Broad Institute, such as the Connectivity Map.[4] The connectivity map uses gene expression signatures to connect small molecules, genes, and disease available through the LINCS Project.[5] Donner et al. used the L1000 data set to develop a new method for measuring the compound functional similarity based on gene expression data for drug repurposing.[6] The method identified drugs with shared therapeutic and biological targets even when the compounds were structurally dissimilar, thereby revealing previously unreported functional relationships between compounds.

Imaging data is among the most abundant data type available for deep learning researchers, often allowing for rapid validation of results using human visual sensory organs. Many of the tools developed for image recognition and trained and tested on simple pictures are now available for researchers working with more complex imaging data types, including computed tomography. Gao and Qian used the patch-based convolutional neural network (CNN) model combined with support vector machines to predict multidrug resistant patients with tuberculosis using a data set of 230 patients from the ImageCLEF2017 competition,[7] achieving reasonably high classification rates.[8]

Xiang and colleagues presented a multitask deep autoencoder for the prediction of the human cytochrome 450 inhibition,[9] laying the roadmap for reducing the side effects associated with inhibiting the CYP450. Lane et al.[10] compared various machine learning models for predicting hit molecules for *Mycobacterium tuberculosis* (Mtb) using a small curated data set of molecules targeting Mtb. Another article by the Ekins group[11] compared various machine learning techniques for predicting the estrogen receptor (ER) binding. In this work, Russo and co-authors compared the AdaBoost, Bernoulli Naïve-Bayes, Random Forest, support vector classification, and deep neural networks using a variety of metrics and a proprietary data set compiled from public sources to predict ER binding. Again, Random Forest outperformed the other algorithms, demonstrating the value of comparing the various algorithms, especially for simple machine learning tasks.

One of the many chemistry related machine learning challenges is the selection of the representation of molecular structure to capture as many of the relevant chemical and biological features and come as close to reality as possible. There are many representations of molecular structures, including a variety of molecular fingerprints, string-based representations, molecular graphs, and others. A molecular graph is a popular representation of the molecular structure for machine learning applications and explored by many groups working on medicinal chemistry related tasks. Hop and colleagues[12] explored the performance of geometric deep learning methods in the context of drug discovery, comparing machine learned features against the domain expert engineered features. The CNN graph outperformed the methods trained on expert engineered features on most of the data sets.

The popular deep learning techniques involving CNN are often trained on 2D and 3D images. To help facilitate f the many applications of CNNs in chemistry, Kuzminykh presented the wave transform-based representation of the 3D molecular structure.[13] The group demonstrated that the proposed representation leads to the better performance of CNN-based autoencoders than either the voxel-based representation or the previously used Gaussian blur of atoms, and it can be successfully applied to classification tasks, such as MACCS fingerprint prediction.

Deep generative models, commonly referred to as AI imagination, enabled many new applications requiring creativity and

Special Issue: Deep Learning for Drug Discovery and Biomarker Development

Published: October 1, 2018

Downloaded via 45.118.49.35 on July 27, 2025 at 04:26:34 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2018 American Chemical Society

DOI: 10.1021/acs.molpharmaceut.8b00930
Mol. Pharmaceutics 2018, 15, 4311−4313

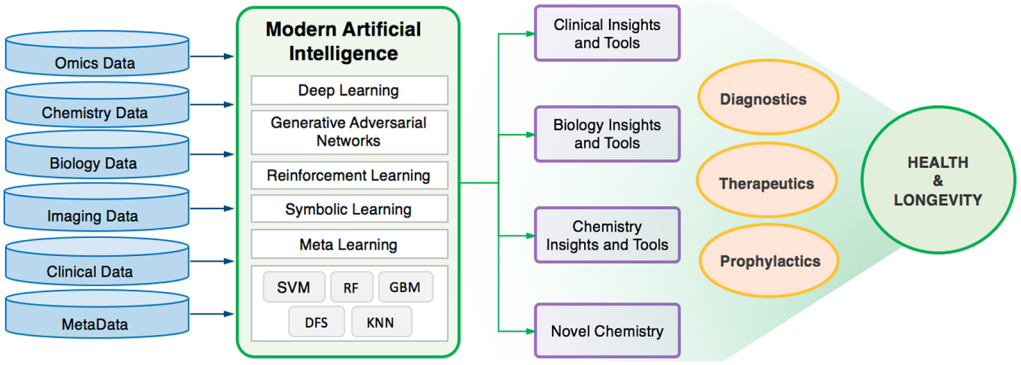

Figure 1. Potential of deep learning for drug discovery and biomarker development.

diversity. They hold a substantial promise for drug discovery, biomarker development, and even the design of novel materials. The first peer-reviewed article with an application of generative models to molecules[14] applied an adversarial autoencoder to the generation of new promising anticancer compounds. Published in a peer-reviewed journal in 2018,[15] Gómez-Bombarelli et al. described an application of the variational autoencoder (VAE) to generative chemistry with a Bayesian optimization of chemical properties. Aspuru-Guzik, one of the pioneers in the many areas of computational chemistry, published the first comprehensive review of the generative chemistry.[16] Aspuru-Gudzik and Zhavoronkov's groups jointly published the Reinforced Adversarial Neural Computer (RANC) for de Novo Molecular Design[17] in the JCIM special issue on machine learning in drug discovery.[18]

This special issue featured the two novel generative models with experimentally validated results, the entangled conditional adversarial autoencoder (ECAAE)[19] and the adversarial threshold neural computer (ATNC) for molecular de novo design.[20]

The ATNC model expanded Aspuru-Gudzik's Objective Reinforced Generative Adversarial Network for Inverse-Design Chemistry (ORGANIC)[21] and introduced the popular differential neural computer technique developed by DeepMind.[22] The generation conditions for molecules generated by RANC were confirmed using in vitro experiments.

The ECAAE is another novel generative chemistry model which generates molecular structures on the basis of various properties, such as activity against a specific protein, solubility, or ease of synthesis. The authors applied the ECAAE model to generate a novel inhibitor of Janus kinase 3, implicated in rheumatoid arthritis, psoriasis, and vitiligo. The discovered molecule was tested in vitro and showed high activity and selectivity demonstrating successful experimental validation of the de novo compound generation using the generative adversarial model with the generation conditions confirmed by experimental results.

Radinsky and Harel proposed the prototype-driven diversity network, a generative chemistry architecture based on VAE enabling the chemists to generate diverse molecules with the desired properties from a molecular prototype.[23]

This special issue on deep learning and the many special issues recently opened by chemistry and biology journals demonstrate the rapid progress and expansion of the body of knowledge in machine learning techniques for chemistry and biology applications. Several papers presented in this issue for the first time demonstrate experimental validation of the novel hybrid generative adversarial networks and reinforcement learning techniques for de novo molecular design.

This special issue demonstrated one of the limitations of the nascent field of AI for drug discovery, the lack of interpretability. None of the works on ML predictors, classifiers, and generators employed modern feature extraction and feature selection techniques in an attempt to make these tools more explainable. Modern techniques in machine learning may help demystify some of the models and expand our understanding of chemistry and biology.

While the progress in artificial intelligence for drug discovery is accelerating, the only real proof of success is the favorable testing of AI-generated molecules for the targets discovered. Using AI in clinical trials is yet to be demonstrated. The race for human validation and real cures using AI is underway.

Alex Zhavoronkov*

JHU, Insilico Medicine, Inc., 9601 Medical Center Dr, Suite 127, Rockville, Maryland 20850, United States

AUTHOR INFORMATION

Corresponding Author

*E-mail: alex@insilicomedicine.com.

ORCID

Alex Zhavoronkov: 0000-0001-7067-8966

Notes

Views expressed in this editorial are those of the author and not necessarily the views of the ACS.

REFERENCES

(1) Aliper, A.; Plis, S.; Artemov, A.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Mol. Pharmaceutics* **2016**, *13*, 2524.

(2) Putin, E.; Mamoshina, P.; Aliper, A.; Korzinkin, M.; Moskalev, A.; Kolosov, A.; Ostrovskiy, A.; Cantor, C.; Vijg, J.; Zhavoronkov, A. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging* **2017**, *8*, 1021−1033.

(3) Mamoshina, P.; Vieira, P.; Putin, E.; Zhavoronkov, A. Applications of Deep Learning in Biomedicine. *Mol. Pharmaceutics* **2016**, *13* (5), 1445−1454.

(4) *Connectivity Map (CMAP)*. Broad Institute, July 31, 2018, www.broadinstitute.org/connectivity-map-cmap.

(5) NIH LINCS Program. *NIH LINCS Program*, www.lincsproject.org.

(6) Donner, Y., Kazmierczak, S., Fortney, S.. Drug Repurposing Using Deep Embeddings of Gene Expression Profiles, *Mol. Pharmaceutics*, ASAP **2018**, DOI: 10.1021/acs.molpharmaceut.8b00284.

DOI: 10.1021/acs.molpharmaceut.8b00930
Mol. Pharmaceutics 2018, 15, 4311−4313

(7) Ionescu, B. et al. Overview of ImageCLEF 2017: Information Extraction from Images. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*; Jones, G.,et al., Eds.; CLEF 2017. Lecture Notes in Computer Science, Springer: Cham, 2017; Vol *10456*.
(8) Gao, X. W.; Qian, Y. Prediction of Multidrug-Resistant TB from CT Pulmonary Images Based on Deep Learning Techniques. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.7b00875.
(9) Li, X.; Xu, Y.; Lai, L.; Pei, J. Prediction of Human Cytochrome P450 Inhibition Using a Multitask Deep Autoencoder Neural Network. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.8b00110.
(10) Lane, T.; et al. Comparing and Validating Machine Learning Models for Mycobacterium tuberculosis Drug Discovery. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.8b00083.
(11) Russo, D.; Zorn, K.; Clark, A.; Zhu, H.; Ekins, S. Comparing Multiple Machine Learning Algorithms and Metrics for Estrogen Receptor Binding Prediction. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.8b00546.
(12) Hop, P.; Allgood, B.; Yu, J. Geometric Deep Learning Autonomously Learns Chemical Features That Outperform Those Engineered by Domain Experts. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.7b01144.
(13) Kuzminykh, D.; Polykovskiy, D.; Kadurin, A.; Zhebrak, A.; Baskov, I.; Nikolenko, S.; Shayakhmetov, R.; Zhavoronkov, A. 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.7b01134.
(14) Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The Cornucopia of Meaningful Leads: Applying Deep Adversarial Autoencoders for New Molecule Development in Oncology. *PLOS Biology*; Public Library of Science, 2016.
(15) Gómez-Bombarelli, R.; et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4* (2), 268−276.
(16) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360.
(17) Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A. Reinforced Adversarial Neural Computer for de Novo Molecular Design. *J. Chem. Inf. Model.* **2018**, *58* (6), 1194−1204.
(18) Hochreiter, S.; Klambauer, G.; Rarey, M. Machine Learning in Drug Discovery. *J. Chem. Inf. Model.* **2018**, DOI: 10.1021/acs.jcim.8b00478.
(19) Polykovskiy, D.; Zhebrak, A.; Vetrov, D.; Ivanenkov, Y.; Aladinskiy, V.; Bozdaganyan, M.; Mamoshina, P.; Aliper, A.; Zhavoronkov, A.; Kadurin, A. Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.8b00839
(20) Putin, E.; Asadulaev, A.; Vanhaelen, Q.; Ivanenkov, Y.; Aladinskaya, A.; Aliper, A.; Zhavoronkov, A. Adversarial Threshold Neural Computer for Molecular de Novo Design. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.7b01137.
(21) Sanchez-Lengeling, B.; Outeiral, C.; Guimaraes, G.; Aspuru-Guzik, A. Optimizing Distributions over Molecular Space. An Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry (ORGANIC). *CehmRxiv*. Preprint, **2017** DOI: 10.26434/chemrxiv.5309668.v3.
(22) Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwinska, A.; Gomez Colmenarejo, S.; et al. Hybrid Computing Using a Neural Network with Dynamic External Memory. *Nature* **2016**, *538*, 471.
(23) Harel, S.; Radinsky, K. Prototype-Based Compound Discovery using Deep Generative Models. *Mol. Pharmaceutics* **2018**, DOI: 10.1021/acs.molpharmaceut.8b00474.

DOI: 10.1021/acs.molpharmaceut.8b00930
Mol. Pharmaceutics 2018, 15, 4311−4313