



## Viewpoint

## Second-generation artificial intelligence approaches for life science research

Jürgen Bajorath

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 6, D-53115 Bonn, Germany



In computer science, artificial intelligence (AI) encompasses various disciplines including machine learning (ML), natural language processing, computer vision, expert systems, decision making, and robotics. In the life and medical sciences, AI applications are currently dominated by deep learning (DL) [1] and –to a lesser extent– robotics. While the number of ML applications using deep neural network (DNN) architectures is on the rise across the life sciences, moving forward several important issues must be taken into consideration and preparations made for the next stage of AI in life science research.

## Data heterogeneity

In terms of volumes, attributes, and complexity, the data landscape in the life sciences is extremely heterogeneous. To give just a few examples, in biology, very large data sets often originate from gene expression analysis or high content screening. The same applies to patient data combining various clinical readouts. Similarly, computational/theoretical and physical chemistry also produce some large data sets. On the other hand, in pharmaceutical research, activity data for candidate compounds or target validation data are often comparably sparse. Importantly, early-phase drug discovery typically operates in low data regimes. Furthermore, data sparseness is also a characteristic of many time series experiments in biology. Clearly, omnipresence of ‘big data’ cannot be assumed in the life sciences. Instead, volumes and complexity of available data typically vary from project to project (with no intrinsic correlation between data sparseness and complexity). This data heterogeneity naturally affects computational learning in life science research, as further discussed below.

## Models and decisions

There is a general tendency to apply computational learning methods that are too complex for prediction tasks at hand [2] and it is often incorrectly assumed that increasingly complex ML methods might be more accurate than simpler ones [2]. This also affects ML applications in the life sciences. Methodological overkill and unsubstantiated accuracy expectations work against rigorous science and cause confusion. Benchmarking of complex models, albeit necessary for basic model assessment and parameter tuning, is often misinterpreted as a validation of model

relevance. However, as long as it is not conclusively demonstrated that simpler models do not produce comparable results, increasing model complexity is not justifiable, especially in practical applications. Furthermore, it must also be considered that current ML applications do not lead to autonomous decisions beyond human reasoning. Rather, ML predictions yield hypotheses that complement decision making by experts. Accordingly, the interplay between predictions, their analysis, and conclusions drawn from them is of critical relevance for the impact of ML on life science programs. In this context, human reasoning continues to play a major role.

## Roadblocks for interdisciplinary research

There is a natural reluctance among experimentalist to rely on predictions that cannot be understood; rightly so. Lack of rationalization of predictive modeling hinders the acceptance of ML in the life sciences and limits its impact on experimental research. This especially applies to complex computational methods such as DNNs and their notorious ‘black box’ character [3]. First and foremost, impact of ML on experimental programs is achieved through prospective applications of models. Prospective use of predictions for experimental design does not only assess the ability of a model to identify novel test instances with desired properties, but also evaluates (multi-factorial) experimental processes [4]. Once results of predictions enter iterative experimental cycles, applied computational models become an integral factor in these processes, which are at least in part subjectively driven, and increasingly contribute to their validation [4]; a major goal of applied ML.

## Methodological implications

What are possible consequences of data heterogeneity, increasing use of complex DNN architectures, and limited acceptance of ML in experimental programs? To ensure further development of interdisciplinary research at the interface between computation and experiment, the focus on AI in the life sciences must, in the author’s opinion, partly shift from DL and DNNs to other AI-relevant approaches that complement predictive modeling in different ways. This shift in focus towards second-generation approaches, as they are termed herein, is thought to be critical for increasing the transparency of DL and its relevance for

E-mail address: [bajorath@bit.uni-bonn.de](mailto:bajorath@bit.uni-bonn.de)

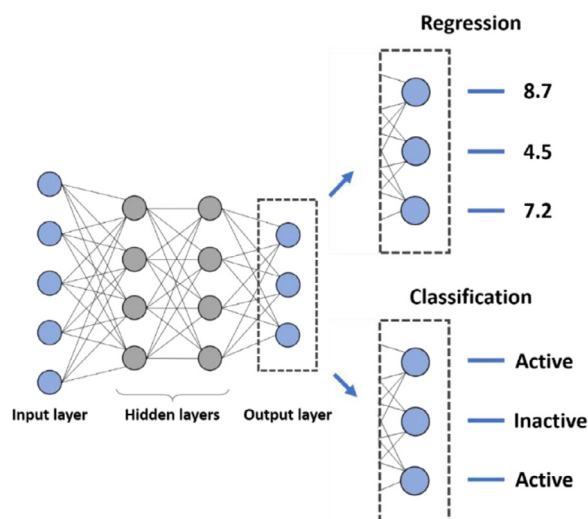
<https://doi.org/10.1016/j.ailsci.2021.100026>

Received 24 November 2021; Received in revised form 26 November 2021; Accepted 28 November 2021

Available online 1 December 2021

2667-3185/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



**Fig. 1. Deep neural networks for multi-task modeling.** Shown is a schematic representation of exemplary multi-task DNN architectures for regression or classification. Each output node corresponds to an individual prediction task.

experimental design. Such second-generation approaches do not necessarily represent novel ML paradigms. However, their potential for complementing ML/DL and increasing its impact in the life sciences might not have been sufficiently considered or is just beginning to be realized.

## Second-generation approaches

### Learning from sparse data

One of the consequences of data heterogeneity across the life sciences is that sufficiently large data sets for effective DL are often not available. Hence, an important task will be the integration of AI approaches specifically designed for small data modeling such as *transfer learning* [5], *active learning* [6], *one-shot* or *few-shot learning* [7,8], and *meta-learning* [9,10]. These techniques generally attempt to minimize the number of training instances for effective learning in related, yet distinct ways. Active learning aims at identifying smallest possible sets of training data by iterative selection of most informative training examples while transfer learning acquires knowledge from a related task for which predictive models have been derived. Similarly, one-shot learn-

ing builds upon knowledge from available models and few-shot learning incorporates data augmentation methods or meta-learning of multiple related tasks. Fig. 1 illustrates DNN architectures for multi-task learning.

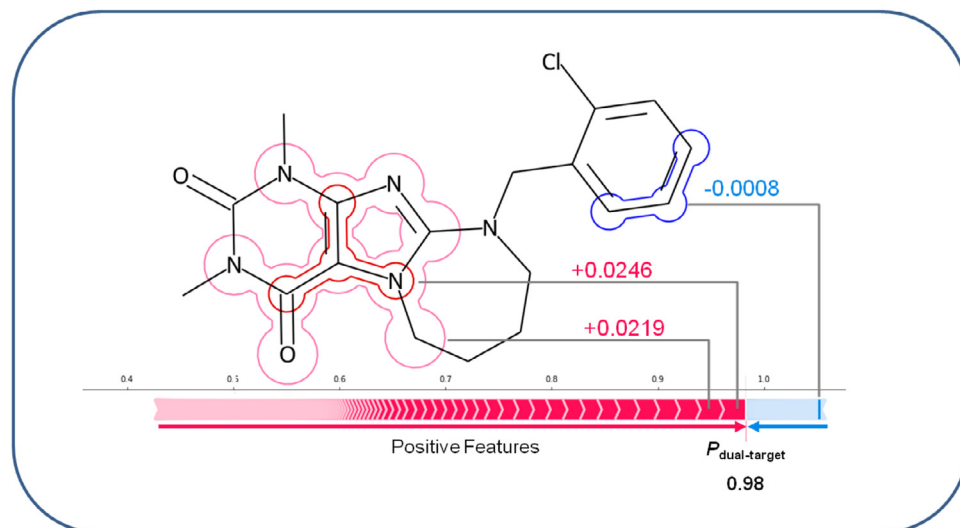
### Explainable ML

To address the reluctance of relying on black box predictions for experimental design, it will be crucial to further investigate methods for explaining ML results [11]. To explain individual predictions, various model-dependent *feature perturbation* and *weighting* methods have been introduced [11,12]. For DNNs, weight gradient analysis is often not reliable due to gradient instability across multiple layers, which often results in distinct explanations of corresponding predictions [13]. For graph convolutional neural networks, an explanation method has been introduced to identify the subgraph determining a prediction [14]. However, with the exception of such special cases, most DNN predictions cannot be rationalized. Accordingly, *model-agnostic* approaches should best be considered that are applicable to ML models generated with any algorithm, regarding of its complexity [12]. An exemplary model-agnostic approach is the *Shapley Additive exPlanations* (SHAP) methodology [15] that locally approximates the calculation of *Shapley values* from collaborative game theory [16] for ML models. SHAP represents an extension of the *Local Interpretable Model-agnostic Explanations* (LIME) method [17] and makes it possible to quantify the contribution of representation features that are present or absent in a test instance to its prediction [15]. SHAP analysis is readily applicable to DNN models [12]. Calculation of Shapley values quantifying feature importance can be combined with feature mapping and visualization to rationalize predictions [12,18], as illustrated in Fig. 2. Encouragingly, SHAP analyses of ML studies are beginning to appear in the life science literature.

Model interpretation is complemented by estimating model-inherent *uncertainty* of predictions [19]. However, ML uncertainty analysis is still in its infancy for life science applications. Importantly, quantifying uncertainty of predictions will help to prioritize model decisions, reduce black box character, and increase confidence in predictions. For deriving uncertainty estimates, model-agnostic ensemble or evidential deep learning algorithms are available [20,21], for example, which can be adapted for life science applications.

### Learning from predictions

Insights into ML predictions can also be obtained by searching for *contrastive explanations*, *counterfactuals*, or *adversarial examples* [22,23],



**Fig. 2. Shapley values and feature mapping.** For a compound with dual-target activity correctly classified using an ML model, Shapley values of representation features are calculated and features contributing to the prediction are mapped onto the structure. With one exception, feature contributions are positive and the sum of positive (red) and negative (blue) feature contributions results in a probability of 0.98 for the correct prediction. The figure was adopted from [18].

which are related to each other. Contrastive explanations aim to identify smallest feature sets whose presence or absence in a test instance is responsible for a particular class label prediction while counterfactuals represent instances with minimal feature differences leading to opposing predictions. In a similar vein, adversarial examples account for minimal feature changes converting a correct into an incorrect prediction, i.e., leading to model errors. These types of explanations or examples are not only intellectually stimulating to explore, but also help to test and better understand explanation methods. Moreover, learning from exemplary contrastive predictions or counterfactuals might reveal model characteristics that are generalizable and thus support global model analysis.

## Conclusion

The surge of DL in chemistry, biology, and medicine is exciting to witness on the one hand, but also comes along with potential caveats on the other. Current requirements of DL do not necessarily meet the data reality in the life sciences. Moreover, black box predictions using increasingly complex computational architectures may well widen the gap between ML/DL and its impact on experimental programs. This has implications for the future development of ML/DL across the life sciences. Clearly, strong emphasis must be put on prospective ML applications. Furthermore, at this stage, one should be concerned about computational concepts that complement ML/DL, positioned herein as second-generation approaches, and increase its relevance for the life sciences as well as its acceptance. Among these approaches, methods for predictive modeling in low data regimes and for rationalizing ML decisions are expected to make important contributions going forward. Integrating such approaches with DL is also expected to provide new opportunities for scientific communication at the interface between computation and experiment, which will also contribute to the further development of the field.

## Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* 2015;521:436–44.
- [2] Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models instead. *Nat Mach Intell* 2019;1:206–15.
- [3] Castellevecchi D. Can We Open the Black Box of AI? *Nature* 2016;538:20–3.
- [4] Kearnes S. Pursuing a Prospective Perspective. *Trends Chem* 2021;3:77–9.
- [5] Yang Q, Zhang Y, Dai W, Pan SJ. *Transfer learning*. Cambridge, UK: Cambridge University Press; 2020.
- [6] Cohn DA, Ghahramani Z, Jordan MI. Active Learning with Statistical Models. *J Artif Intell Res* 1996;4:129–45.
- [7] Fei-Fei L, Fergus R, Perona P. One-Shot Learning of Object Categories. *IEEE Trans Pattern Anal Mach Intell* 2006;28:594–611.
- [8] Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a Few Examples: A Survey on Few-Shot Learning. *ACM Computing Surveys (CSUR)* 2020;53:1–34.
- [9] Finn, C.; Xu, K.; Levine, S. Probabilistic Model-Agnostic Meta-Learning. *arXiv preprint*, 2018.
- [10] Zhang, Y.; Yang, Q. A Survey on Multi-Task Learning. *arXiv preprint*, 2017.
- [11] Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, Methods, and Applications in Interpretable Machine Learning. *Proc Nat Acad Sci USA* 2019;116:22071–80.
- [12] Rodríguez-Pérez R, Bajorath J. Chemistry-Centric Explanation of Machine Learning Models. *Artif Intell Life Sci* 2021;1:100009.
- [13] Ghorbani A, Abid A, Zou J. Interpretation of Neural Networks is Fragile. *Proc AAAI Conf Artif Intell* 2019;33:3681–8.
- [14] Ying R, Bourgeois D, You J, Zitnik M, Leskovec J. Gnnexplainer: Generating Explanations for Graph Neural Networks. *Adv Neur Inform Process Syst* 2019;32:9240.
- [15] Lundberg SM, Lee S. A Unified Approach to Interpreting Model Predictions. *Adv Neur Inform Process Syst (NIPS)* 2017;30:4766–75.
- [16] Shapley LS. A Value for N-Person Games. In: Kuhn HW, Tucker AW, editors. *Contributions to the theory of games*. Princeton: Princeton University Press; 1953. p. 307–17. *Annals of Mathematical Studies*.
- [17] Ribeiro MT, Singh S, Guestrin C. Why Should I Trust You? Explaining the Predictions of Any Classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016. p. 1135–44.
- [18] Feldmann C, Philipps M, Bajorath J. Explainable Machine Learning Predictions of Dual-Target Compounds Reveal Characteristic Structural Features. *Sci Rep* 2012;11:21594.
- [19] Hie B, Bryson BD, Berger B. Leveraging Uncertainty in Machine Learning Accelerates Biological Discovery and Design. *Cell Syst* 2020;11:461–77.
- [20] Lakshminarayanan B, Pritzel A, Blundell C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. *Adv Neur Inform Process Syst (NIPS)* 2017;30:6402–13.
- [21] Sensoy, M., Kaplan, L., Kandemir, M. Evidential deep learning to quantify classification uncertainty. *arXiv preprint*, 2018.
- [22] Stepin I, Alonso JM, Catala A, Pereira-Fariña M. A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence. *IEEE Access* 2021;9:11974–2001.
- [23] Goodfellow, I.J., Shlens, J., Szegedy, C. Explaining and Harnessing Adversarial Examples. *arXiv preprint*, 2014.