

EDITORIAL

Deep learning in bioinformatics and biomedicine

Deep learning is a subfield of machine learning that considers computational models with multiple processing layers [1, 3, 6]. At the core of all deep learning approaches lies ‘representation learning’: the models automatically learn a representation of the input data without the explicit guidance of a domain expert. Low-level features (such as edges in image data) are moved forward to the next layer where higher level, more abstract features (such as shapes) are extracted. This feature extraction is based on nonlinear functions in the processing units. Thereby, deep learning is capable of discovering the intrinsic hierarchies in the training data that can be exploited for a variety of analytical tasks. Most deep learning methods are designed for supervised classification, that is tasks for which an input–output mapping has to be learned from labeled training data. The success and failure of predictive models crucially depend on how well the hierarchical structures can be captured. ‘Deep neural networks’, such as multilayer feed-forward perceptrons, convolutional neural networks and recurrent neural networks, are particularly good at this task.

Deep learning requires truly large amounts of training data, which became widely available with the dawn of the big data era. The advances in experimental capabilities have led to increasing amounts of complex data in the life sciences, too. The advent of high-throughput technologies brought a paradigm shift from traditionally hypothesis-driven to data-driven research. Studies in biology and medicine rests on ever-growing volumes of complex data that characterize phenomena across a wide range of physical and organizational scales, from molecules to environments [10, 11]. More data hold the promise of a better understanding of the mechanisms underlying many biological structures and (patho-)physiological processes [7], which might ultimately lead to improved therapies for patients. As many problems in the life sciences require the decoding of complex interactions between entities (e.g. genes, proteins), and given the wealth of multimodal data, deep learning methods are believed to have a transformative impact on biomedicine [2].

The aim of this special issue is to provide the readers with a set of reviews that describe the latest concepts, innovations, approaches and technologies in the area of deep learning in bioinformatics, computational biology and systems medicine.

In their article titled ‘Biological network analysis with deep learning’, Muzio *et al.* review the state of the art of deep learning methods for the analysis of graph data. Biological data are often represented in the form of graphs, such as protein interaction networks and gene regulatory networks. However,

neural networks commonly operate on vector and matrix data, not graph-structured data, so adaptations are required. Graph neural networks are a relatively new type of deep neural networks that can learn hierarchical nonlinearities and neighborhood information from graph data. Muzio *et al.* review a variety of applications, including protein function prediction, polypharmacy prediction in drug discovery and development, metabolic pathway prediction, drug–target prediction, prediction of drug properties, drug–drug interaction prediction and disease diagnosis.

In ‘Deep learning meets metabolomics: A methodological perspective’, Sen *et al.* survey deep learning approaches for metabolomics, that is the large-scale analysis of metabolites in biological systems. The article presents a comprehensive overview of the potential and limitations of deep learning for the data analysis pipeline, from data acquisition and processing to metabolite identification, metabolic phenotyping and biomarker discovery.

The article ‘Deep learning in systems medicine’ by Wang *et al.* focuses on the applications of deep learning in predictive, preventive and precision medicine. A key component of systems medicine is the integration of heterogeneous, multiscale data. After a short review of the fundamentals of deep learning, Wang *et al.* discuss some successful applications of deep learning for systems medicine.

One area of systems medicine with deep learning is brain disorders, for which Burgos *et al.* give a review in ‘Deep learning in brain disorders: from data processing to disease treatment’. This review covers image reconstruction, signal enhancement, cross-modality image synthesis, disease biomarker discovery, diagnosis of neurodegenerative diseases, prediction of their clinical outcomes and disease understanding.

In ‘Deep learning approaches for neural decoding across architectures and recording modalities’, Livezey and Glaser discuss applications of deep learning in neuroscience, with a focus on approaches to neural decoding, i.e. how signals from the brain can be used to predict behavior, perception or cognitive states.

In their article ‘Unsupervised and self-supervised deep learning approaches for biomedical text mining’, Nadif and Role give an appraisal of deep learning methods that overcome the lack of labeled data in biomedical text mining, enabling significant improvements in text classification, text clustering, named entity recognition, question–answering tasks and relation extraction.

Although all six articles conclude that deep learning systems hold great promise for scientific advances in the life sciences, there is also a general consensus that the lack of ‘interpretability’ represents a limitation to their deployment in clinical practice. Informally, ‘interpretability’ can be defined as that property of an artificial intelligence system that allows a human domain expert to understand the inner workings of the system and why it made—or did not make—certain predictions. Deep neural networks, however, are opaque black boxes. Among the set of models with similar performance, some might obey additional constraints, such as sparsity and interpretability and could therefore be preferable to black box models [9]. The relatively young subfield of ‘eXplainable and Interpretable Machine Learning’ focuses on issues of algorithmic transparency, interpretability, accountability and explainability of algorithmic decisions [4].

And there is another problem: deep neural networks excel in domains where huge amounts of (labeled) training data are available and where prior knowledge is practically irrelevant, for example image recognition tasks [5]. However, the sciences are immensely knowledge-rich domains. Integrating domain knowledge into deep learning systems is largely uncharted territory [8]. It is also certainly not the case that deep learning systems can always simply learn an appropriate representation from the raw input data without prior preprocessing by a data scientist with relevant background knowledge.

Despite these caveats, the articles of this special issue showcase a wide and fascinating range of deep learning applications in the life sciences. We wish to express our gratitude to the authors, the reviewers and the editorial staff for giving us the opportunity to compile this timely special issue.

Daniel Berrar

Data Science Laboratory, Department of Information and Communications Engineering, Tokyo Institute of Technology, Ookayama, Tokyo 152-8550, Japan; Tel.: +81-3-5734-3088.
Email: daniel.berrar@ict.e.titech.ac.jp

Werner Dubitzky

Werner Dubitzky is a freelance data scientist. His research interests are data science and bioinformatics.

Daniel Berrar is an Associate Professor in the School of Engineering, Tokyo Institute of Technology. His research interests are machine learning, data science, and their applications in the life sciences.

References

1. Bengio Y. Learning deep architectures for AI. *Found Trends Mach Learn* 2009; 2:1–127.
2. Ching T, Himmelstein D, Beaulieu-Jones B, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018; 15(141):20170387.
3. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press, 2016.
4. Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. *WIREs Data Min Knowl Discov* 2019; 9(4):e1312.
5. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges C, Bottou L et al. (eds). *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, 1097–105.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521:436–44.
7. Leoneli S. *Data-Centric Biology: A Philosophical Study*. Chicago: The University of Chicago Press, 2016.
8. Marcus G. Deep learning: a critical appraisal. *CoRR*, abs/1801.006312018;1–27.
9. 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.
10. Stephens Z, Lee S, Faghri F, et al. Big data: astronomical or genetical? *PLoS Biol* 2015; 13(7):e1002195.
11. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. *Big Data* 2013; 1(2): 85–9.