

Fernando Chirigati
Springer Nature
One New York Plaza, Suite 4500
New York, NY 10004

November 12th, 2020

Dear Dr. Chirigati,

It is with great enthusiasm that we submit this letter to you to inquire about potential interest in publishing a new and exciting perspective article entitled “Dos and Don’ts for Deep Learning in the Life Sciences.” As you know, deep learning is exploding in popularity, and it is increasingly used for varied research purposes within the life sciences. Deep learning is a large and complex field, and its optimal application within the life sciences remains a daunting task for most individuals. By providing accessible and actionable guidance about how to best leverage deep learning to answer biological and biomedical questions, we seek to accelerate scientific progress and minimize barriers to entry.

We have assembled a diverse international team of authors, all of whom responded to an open call for contribution. These experts span an impressive global list from more than twenty academic and corporate institutions, whose research draws from relevant disciplines such as molecular biology, data science, machine learning, genomics, clinical studies, and data ethics. We affirm that our original manuscript will provide actionable, evidence-based advice for both new and experienced deep learning practitioners. In our manuscript, we draw upon key examples and guidance from our prior work—and from many other colleagues who have also made seminal contributions to the field.

Our key best practices include the following:

1. Deciding whether deep learning is appropriate for your problem
2. Using traditional methods to first establish performance baselines
3. Understanding the complexities of training deep neural networks
4. Knowing your data and your question
5. Choosing an appropriate data representation and neural network architecture
6. Tuning your hyperparameters extensively and systematically
7. Addressing deep neural networks’ increased tendency to overfit datasets
8. Making deep learning models more transparent
9. Discussing the ethics of your work
10. Acknowledging risks related to sharing models trained on sensitive data

Notably, these topics range from high-level guidance to implementation-related best practices, and they have been devised to effectively reach audiences of varying expertise. Upon notification that this manuscript might be suitable for submission to *Nature Computational Science*, we will further engage with our community of authors to coordinate a timeline; at present, our manuscript is nearly completed and pending minor internal revisions. Importantly, all authors of the manuscript meet the ICMJE and Nature authorship standards.

By providing guidance on deep learning, these powerful methods can be more properly utilized by both computational practitioners, experimental biologists, and clinical scientists. We aim not only to increase the accessibility of deep learning techniques to life sciences, but also to improve upon the overall quality and reproducibility of deep learning research in the literature.

Sincerely,
Benjamin D. Lee, on behalf of all authors