Dr. Scott Markel

Dassault Systèmes BIOVIA

5005 Wateridge Vista Dr.

San Diego, CA 92121

December 6th, 2018

Dear Dr. Markel,

It is with great enthusiasm that we are writing to ask your interest in publishing a Ten Simple Rules article entitled "Ten Simple Rules for Deep Learning in Biology." Deep learning (DL) is exploding in popularity and increasingly used for biological data analysis. Alone, DL is large and complex field, and its correct application to biological research remains a daunting task for most. By providing accessible and actionable guidance on how to best use DL to answer biological questions, we seek to accelerate scientific progress and minimize barriers to entry.

Inspired by *Opportunities and obstacles for deep learning in biology and medicine* and *Ten Simple Rules for Writing a PLOS Ten Simple Rules Article* (Rule 5: Collaborate), we have already assembled a team of expert authors in bioinformatics and DL through an open call for contributions. To that end, we have created a GitHub repository to host the discussion and drafting process. Further, we propose to write the article with Manubot, a collaborative manuscript authoring platform based on GitHub. By writing the manuscript in the open on GitHub, which is extensively used by those actively conducting DL research, and soliciting input from the wider scientific community, we are convinced that this manuscript will provide actionable, DL-specific advice for both new and experienced DL practitioners.

In preparation for this letter of inquiry, the following ten rules have been proposed:

1. Concepts that apply to machine learning also apply to deep learning
2. Understand the complexities of training deep neural networks
3. Know your data and your question
4. Choose an appropriate neural network architecture and data representation
5. Tune your hyperparameters extensively and systematically
6. Address deep neural networks' increased tendency to overfit the dataset

7. Use traditional methods to establish performance baselines
8. Do not necessarily consider a DL model as a black box
9. Interpret predictions in the correct manner
10. Don't share models trained on sensitive data

These rules range from high-level guidance to implementation best practices and they have been devised to effectively reach audiences of varying expertise. Upon notification that this manuscript will be suitable for submission to PLOS Computational Biology, we will further engage with the community to solicit feedback and contributions. All authors of the paper will meet the ICMJE authorship standards.

By providing guidance on DL, these powerful methods can be more properly utilized by both computational and experimental biologists. Here, we aim to increase the accessibility of DL techniques to biology and thereby improve the overall quality and reproducibility of DL in the literature.

Sincerely,

Benjamin Lee, on behalf of all contributors