AIMS Bioengineering, 8(1): 93–94.
DOI: 10.3934/bioeng.2021009
Received: 19 January 2021
Accepted: 20 January 2021
Published: 21 January 2021

http://www.aimspress.com/journal/Bioengineering

Editorial

Artificial intelligence and data science applied to bioengineering

Almudena Espín-Pérez[1], Selen Bozkurt[1], Hong Zheng[1] and Aleksandra Nivina[2]

[1] Center for Biomedical Informatics Research, Stanford University, California, USA
[2] Department of Chemistry, Stanford University, California, USA

* **Correspondence:** Email: aespin@stanford.edu.

The recent emergence of big data in chemistry and biology demands special methodologies to analyze these complex datasets. The volume, speed at which the data is created, and diversity of big data combined with the recent advancements in data storage and computer processing has facilitated the training of powerful Artificial Intelligence (AI) models to perform a wide variety of classification and prediction tasks.

AI has the potential to address global health and environmental challenges. It is increasingly used in clinical decision-making to inform precision medicine with the potential to revolutionize medical practice. However, there is a gap between published AI manuscripts and clinical deployment of them. Thus, the significance and utility of big data and AI research needs to be carefully examined. The scientific community should continue to work toward developing and refining both novel and existing methods using AI and data science to make the tools applicable in real-world scenarios.

The first challenge is to understand the fairness of AI algorithms. If the connection between the input data and the model output is indecipherable (the so-called 'black box' problem), under inappropriate management, it could lead to unexpected and unjustified decisions that can be particularly problematic in the field of medicine where lives would be on the line. We need to think critically how important it is to understand the way AI works depending on the context, and provide transparent and interpretable models that can better understand underlying mechanisms where appropriate. Moreover, we need to be aware of the extent to which AI can reliably identify causal links in data and exploit their potential to advance our knowledge and determine effective strategies. It should be also noted that AI is an umbrella term for a set of related tools and approaches, and successful outcomes in one field does not imply equivalent success in others.

The second challenge relates to the collection and preprocessing of the input data which might

cause bias in the model development. Developing strategies and tools to handle the biological heterogeneity and tease out important biological signals from technical noise such as artificial batch effects will improve our understanding and predictions of biological systems. Importantly, there is an urgent need for benchmark and quality control assessment of the current and future approaches to make reproducible and standardized pipelines.

The third challenge is to develop tools to reduce the dimensionality of the data in a contextually meaningful manner. These tools should also take into consideration the data generating process of the state of the art techniques so that assumptions embedded in these techniques are met. In addition, the availability of high volume of data, for instance in Electronic Health Records (EHRs), single-cell and/or other 'omics' fields, offers opportunities for data driven research while requiring caution for data dredging, bias, or confounding.

Importantly, caution must be exercised before deploying such algorithms for decision support, as there may be underlying societal biases and health privacy issues. AI and big data offer opportunities and challenges to support minorities' health, address the importance of consent and patient governance in data collection. Systematic biases in AI models have been a problem in other fields such as the judicial system, which showed racial biases. Instead of making the court fairer, sometimes AI introduces further bias to decision-making to the court. Similarly, in the medical field it is crucial to understand the underlying data and models and identify factors that may lead to biases when designing algorithms.

The aim of this special issue is to come up with models to process, analyze and understand the data to support and improve the performance of current tasks as well as improving our understanding of complex systems involved in biological processes. Future benefits of deploying AI models to support and enhance research will only be possible by means of interdisciplinary working teams and critical thinking. With this issue we hope to elaborate on these challenges and invite scientists to pave the ground for integrating AI to process and understand complex biological data.

AIMS Press