



Editorial

First-generation themed article collections

Artificial Intelligence in the Life Sciences has completed the first round of four Themed Article Collections (TACs) focusing on different topics related to AI in the life sciences. Summaries of the TAC papers are presented below. Some publications qualified for more than one TAC. The *Journal* is grateful to the Guest Editors for their dedicated efforts in organizing and editing these TACs and to our authors for their valued contributions.

Women in AI in the life sciences

Edited by Steve Gardner, Carolina Horta Andrade, Melissa Landon, and Raquel Rodríguez-Pérez.

The TAC *Women in AI in the Life Sciences* highlights relevant investigations of AI applications to solve biomedical, healthcare, or chemistry problems. This collection contains seven papers, including three Research Articles, a Short Communication, a Perspective, and two Reviews. Two research papers focus on machine learning in the field of cheminformatics. Yang et al. presented machine learning classification models for JAK1 inhibitors and structure-activity relationship analyses based on models' outputs. Stroobants et al. evaluated proteochemometric (PCM) models for drug-kinase affinity prediction in a pharmaceutical industry setting. Key limitations of PCM models were detected for this application, which highlighted the need for more robust evaluations. Related to affinity predictions toward protein systems of interest, Ochoa et al. provided three computational protocols with open code to assist in structural and dynamic analyses of protein-peptide complexes. Data was also made publicly available by Chavez-Hernandez and Medina-Franco, who reported a selection and characterization of diverse compounds from natural products. In Volkamer et al., eight scientists from different pharmaceutical companies and academic organizations presented a perspective on machine learning for small molecule drug discovery, in which they discussed the model life cycle and highlighted the main challenges, differences, and potential for collaborations in industry and academia. Silva-Mendonça et al. introduced the concept of few-shot learning and its application to drug discovery, giving an in-depth analysis of the latest research papers in the field. Finally, El-Taraboulsi et al. presented a review of deep learning for cardiac image segmentation, where they summarized the theoretical basis of different deep learning algorithms, highlighted top-performing models, and evaluated key limitations in current research.

AI in the life sciences by Latin Americans

Edited by Carolina Horta Andrade and José L. Medina-Franco.

This article collection features contributions by Latin American

scientists to advance the development and use of artificial intelligence in the life sciences. The TAC includes three papers that describe freely available tools to conduct various tasks relevant to drug discovery, such as performing structural and dynamical analysis of protein-peptide complexes and obtaining natural product subsets from large compound databases. Prada Gori et al. discuss the suite LIDeB Tools, a collection of cheminformatics applications for clustering small molecules, decoy generation, druggability assessment, and data standardization and visualization. Sánchez-Cruz presents an analysis of deep learning methods developed for molecular docking, including opportunities for future development. In a different but relevant application of deep learning in the life sciences, Merchan et al. showed that Siamese and triplet neural networks are practical tools for classifying quickly and precisely MALDI-TOF-generated mass spectra of neotropical and public-health-relevant arthropod species. Finally, a group of Latin American scientists provided a perspective on cheminformatics topics investigated over the past 12 years, highlighting the contributions to open-access tools and educational-related resources and events. The authors also discuss the challenges of cheminformatics research and education in Latin America.

AI techniques advancing the research of bioactive molecules in industry

Edited by Johannes Kirchmair and Floriane Montanari.

This TAC comprises six articles highlighting the diversity of applications of AI in pharmaceutical companies. The TAC begins with a report by Zhao et al. from Bayer AG on a novel deep learning model for predicting the bioconcentration factor, a key parameter in environmental risk assessment. A highlight of this work is a new explanation method that enables atom-level interpretations of individual predictions. Model interpretability also takes center stage in the second contribution by Mora et al. from AstraZeneca, which introduces new multi-task convolutional neural network (CNN) and graph convolutional neural network (GCNN) models for predicting metabolic clearance parameters from molecular images. The CNN pixel intensities are shown to correlate with clearance predictions, and both CNN and GCNN interpretations complement each other, demonstrating their high potential for guiding medicinal chemistry design.

The third contribution, from the MELLODDY consortium, explores the use of conformal efficiency as a proxy for assessing the applicability domain of models. The work shows that the applicability domain of the models can be further extended by federated learning. Another outcome of the MELLODDY consortium, detailed in the fourth contribution, discusses the challenges of benchmark models in the context of evolving

<https://doi.org/10.1016/j.ailsci.2023.100088>

Received 10 November 2023; Accepted 10 November 2023

Available online 15 November 2023

2667-3185/© 2023 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/>).

chemical reference spaces, with a particular emphasis on proteochemometric models.

As the pharmaceutical industry explores novel modalities beyond small molecules, the fifth contribution of this TAC highlights PROTAC modeling. This work by Rao et al. from Celeris Therapeutics GmbH proposes a novel method for modeling model ternary complexes formed by the target protein, the PROTAC, and ligase using Bayesian optimization.

Finally, the sixth contribution aims to bring machine learning closer to experimentalists. Here, researchers from Sanofi propose using a CNN to classify images of dose-response curves to reduce or alleviate the manual work of high-throughput screening (HTS) biologists and speed up HTS data analysis.

Early career researchers in AI

Edited by Francesca Grisoni and Filip Miljković.

This article collection was envisioned to support the professional development and scientific contributions of young researchers from both academia and industry. Early career researchers were required to submit their scientific proposals as the corresponding authors, thus gaining first-hand experience heading a publication project.

The work by Panzarella et al. discusses some of the currently available biomedical ontologies and frameworks (i.e., ontology development environments) applied in the life sciences. It provides a comparison concerning usability, scalability, stability, integration, documentation, and originality. As scientific literature represents a wealth of unstructured data that is often still inadequately and insufficiently utilized, life science ontologies are continuously being developed to address the challenges of chemical and biological knowledge communication, framework definition for understanding concept similarities and differences, and appropriately positioning research work in the scientific context.

In another work, Gadiya et al. applied a formerly devised patent

enrichment tool to extract, integrate, and analyze patent literature for Alzheimer's disease and rare diseases. Their study identified novel drug repurposing activities across the two disease areas and suggested novel potential therapeutic approaches applicable to these indications, thus demonstrating the expanded use of patent documents in drug discovery research.

The third contribution by Sánchez-Cruz et al. explored experimental uncertainty pertaining to protein-ligand binding affinity prediction models, as well as the prospects of combining data of different origins for machine learning. As demonstrated, the performances obtained by state-of-the-art machine learning models tend to be overly optimistic, often causing a misperception that predictions of such models reach experimental accuracy.

Given the high quality and conceptual diversity of studies published in these TACs, *Artificial Intelligence in the Life Sciences* will launch a second round of TACs in 2024 focusing on different scientific topics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Jürgen Bajorath^{*}, Steve Gardner, Francesca Grisoni, Carolina Horta Andrade, Johannes Kirchmair, Melissa Landon, José L. Medina-Franco, Filip Miljković, Floriane Montantari, Raquel Rodríguez-Pérez

^{*} Corresponding author.

E-mail address: bajorath@bit.uni-bonn.de (J. Bajorath).