Toxicology: Open Access

Mini Review Open Access

Environmental Toxicity Identification, Prediction, and Exploration Using Machine Learning: Problems and Perspectives

Mendeley Collins*

Department of Toxicology, UCL College of Toxicology, London, United Kingdom

Abstract

Data-driven machine learning (ML), which has gained recent popularity in environmental toxicology, has distanced itself from hypothesis-driven research during the past few decades. The application of ML in environmental toxicology is still in its infancy, however, due to knowledge gaps, technical challenges with data quality, interpretability issues with high-dimensional/heterogeneous/small-sample data analysis, and a lack of a thorough understanding of environmental toxicology. We evaluate the most current advancements in the literature and highlight cutting-edge toxicological investigations utilising ML in light of the aforementioned issues (such as learning and predicting toxicity in complicated biosystems and multiple-factor environmental scenarios of long-term and large-scale pollution).

Keywords: Machine learning; Environmental toxicology; Pollution

Introduction

New medications, industrial chemicals, and insecticides are essential to human activity. The chemical abstract service (CAS) currently has over 144 million chemical compounds and sequences registered, with over 12,000 new substances being added daily [1]. The CAS databases have shown an exponential rise in scientific publications and research in the area of discovering and studying chemicals. However, improper use and abuse of chemicals in the medical, industrial, and agricultural sectors, as well as chemical releases into the environment that were not anticipated (such as nanoparticles and micro plastics), have increased the possibility of dangers to the environment and to human health [2]. In the twenty-first century, it has grown more difficult to comprehend the intricate and interconnected chemical hazardous responses and probable pathways [3, 4, 5].

Environmental toxicology disclosure: moving from hypothesis- to data-driven

Due to the restricted number of test subjects in traditional environmental toxicology investigations, in vitro and in vivo tests must be carried out over an extended period of time, resulting in significant time and financial costs as well as possibly cruel effects. The complicated interactions of pollutants with biosystems (such as biodegradation and metabolism) as well as their distribution and transformation (such as byproducts owing to oxidation and reduction) in the environment add to the complexity of environmental toxicity prediction. Tens of thousands of chemical qualities, natural environmental elements including temperature, precipitation, hydrochemical variables, and soil components, as well as socioeconomic activity, are the influencing factors (e.g., social activity, educational possibilities, and economic status).

Therefore, it is crucial to make use of current data to create models for predicting environmental toxicity that are more accurate and affordable than conventional techniques (e.g., quantum chemical calculation). Large databases used in data mining, machine learning (ML) algorithms, and expert systems allow for more precise prediction than more conventional techniques. As a result, "toxicity prediction" has recently gained popularity in the fields of chemical discovery, synthesis, and analysis. Based on data from the Web of Science Core Collection (search data were acquired on 23 May 2022; the search topics were "toxic*" and "predict*"), there were 55,030 articles on this subject in 2022. There were 6255 papers on ML, computational, or in silico issues in these journals.

Traditional hypothesis-driven methods are being supplemented by research on data-driven methods (e.g., animal trials and in vitro cell-based models). The design and synthesis of molecular or chemical materials have been the subject of extensive research in recent years, particularly in the domains of materials science (Schmidt et al., 2019), pharmaceuticals, and biomedicine. Application of ML in numerous fields is made simple by the use of efficient algorithms and mature, widely-used programming languages. Toxicologists now have an unrivalled potential to increase the predictability of the toxicological areas thanks to new algorithms and creative models that combine the benefits of data-driven tactics with mechanistic approaches. It is impressive what machine learning has recently accomplished in terms of predicting biological reactions and chemical behaviour in the environment, including critical elements like bioactivity (as it relates to both toxicity and efficacy depending on the compound class) and fate/persistence (both in the body and in the environment). The ML algorithm can be used in some situations to identify and validate biomarkers in medicine; to predict chemical reactions; and to look for therapies for untreatable disorders. Large databases can be used as the foundation for ML algorithms that combine expert systems to provide unique, testable, and verifiable hypotheses, leading to more accurate and useful outcomes. Previous reviews concentrated on computational models for specific toxicity to coordinate the development of ML. Comprehensive applications of ML in environmental areas (e.g., recognising environmental contamination, tracing pollutant sources, researching environmental behaviour and toxicity) have recently attracted greater attention (Dávila-Santiago et al., 2022; Liu et al., 2022; Xia et al., 2022). In earlier assessments, there was little discussion of the difficulties and perspectives in the identification, forecasting, and

***Corresponding author:** Mendeley Collins, Department of Toxicology, UCL College of Toxicology, London, United Kingdom, E -mail: MendCo33@yahoo.com

Received: 03-Jan-2023, Manuscript No: tyoa-23-85612; **Editor assigned:** 05-Jan-2023, Pre-QC No: tyoa-23-85612 (PQ); **Reviewed:** 19-Jan-2023, QC No: tyoa-23-85612; **Revised:** 21-Jan-2023, Manuscript No: tyoa-23-85612 (R); **Published:** 30-Jan-2023, DOI: 10.4172/2476-2067.1000199

Citation: Collins M (2023) Environmental Toxicity Identification, Prediction, and Exploration Using Machine Learning: Problems and Perspectives. Toxicol Open Access 9: 199.

exploration of environmental toxicity [6, 7, 8].

Despite having enormous potential for solving environmental toxicity issues, ML still has difficulties in real-world settings. The two biggest problems with ML applications are inadequate data and limited model expandability. The creation of models and the prediction of new chemicals are constrained by the lack of comparable toxicological data. Data serves as the basis for ML models. To advance ML development, it is currently necessary to address the imbalance, high dimensionality, and heterogeneity of experimental chemical toxicity datasets. For novel chemicals, it is vital to develop efficient toxicity screening methods and assessment standards, especially given the lack of ML standardisation [9].

Conclusion

Our evaluation focused on state-of-the-art toxicological investigations utilising ML rather than just summarising the existing literature because the relationship between ML and environmental toxicology is still in its infancy and because knowledge gaps and technical limitations are notable. To highlight the future primary research areas and methodologies, this review primarily focuses on the present advances in environmental toxicity identification and prediction using ML. The discussion of data-driven methodologies that are required to provide insights into the mechanism research on environmental toxicology follows, from both the data and algorithmic perspectives. The main issues and viewpoints are then highlighted to lessen scientific "blind spots" for upcoming ML-based environmental toxicity assessment projects [10].

Acknowledgement

The Fundamental Research Funds for the Central Universities, the National Key Research and Development Project (2020YFC1807000 and 2019YFC1804603), the National Natural Science Foundation of China (22176103 and 21876092), and the 111 programme of the Chinese Ministry of Education all provided financial support for this work.

Declaration of competing interest

The authors affirm that they have no known financial or interpersonal conflicts that would have appeared to have an impact on the research presented in this study's the work reported in this paper.

References

1. Zhang X, Zeng Q, Cai W, Ruan W (2021) Trends of cervical cancer at global, regional, and national level: data from the Global Burden of Disease study 2019. BMC Public Health 21: 894.
2. Zhang S, Xu H, Zhang L, Qiao Y (2020) Cervical cancer: Epidemiology, risk factors and screening. Chin J Cancer Res 32: 720-728.
3. Pikala M, Burzyńska M, Maniecka-Bryła I (2019) Years of life lost due to cervical cancer in Poland in 2000 to 2015. Int J Environ Res Public Health 16: 1545.
4. Nowakowski A (2015) The implementation of an organised cervical screening programme in Poland: An analysis of the adherence to European guidelines. BMC Cancer15: 279.
5. Conceição T, Braga C, Rosado L, Vasconcelos MJM (2019) A review of computational methods for cervical cells segmentation and abnormality classification. Int J Mol Sci 20: 5114.
6. Duesing N (2012) Assessment of cervical intraepithelial neoplasia (CIN) with colposcopic biopsy and efficacy of loop electrosurgical excision procedure (LEEP). Arch Gynecol Obstet 286: 1549-1554.
7. Zhang J, Cheng K, Wang Z (2020) Prevalence and distribution of human papillomavirus genotypes in cervical intraepithelial neoplasia in China: A meta-analysis. Arch Gynecol Obstet 302: 1329-1337.
8. Sitarz K (2020) HPV infection significantly accelerates glycogen metabolism in cervical cells with large nuclei: Raman microscopic study with subcellular resolution. Int J Mol Sci 21: 2667.
9. William W, Ware A, Basaza-Ejiri AH, Obungoloch J (2018) A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. Comput Methods Progr Biomed 164: 15-22.
10. Ghoneim A, Muhammad G, Hossain MS (2020) cervical cancer classification using convolutional neural networks and extreme learning machines. Futur Gener Comput Syst 102: 643-649.