ACS Chapter Nigeria Book of Proceedings
ISBN: 978-978-68-0485-9 DOI: https://doi.org/10.55455/acsnigeria.1.2.127-133

Integrating AI into Environmental Chemistry: Predicting and Preventing Toxicity in Chemical Products

***Ayodeji Oluwole Adegun[1], Thompson Akinyele Akinnifesi[2], Morayo Omotunde Adejumo[1], Adejumoke Moriliat Hashimi[3], Gregory Olufemi Adewuyi[1], Kayode Adeniyi Aseperi[4], and Barker James[5]**

[1]Department of Chemistry, University of Ibadan, Ibadan, Nigeria
[2]Department of Chemical Sciences, Adekunle Ajasin University, Nigeria
[3]Department of Chemical Sciences, Olabisi Onabanjo University, Nigeria
[4]Centre for Process Innovation, North East England, UK
[5]School of Life Sciences, Pharmacy and Chemistry, Kingston University, London, UK
Corresponding Author's email: ayodeji2k15@gmail.com

ABSTRACT

This review critically examines the current state of AI in environmental chemistry, highlighting both its potential and limitations in predicting and preventing toxicity in chemical products. Integrating Artificial Intelligence (AI) into environmental chemistry for predicting and preventing toxicity in chemical products is an emerging field that promises to revolutionize toxicological assessments and sustainable chemical design. AI techniques, particularly machine learning (ML) and deep learning (DL), offer the potential to predict the toxicity of chemical substances rapidly, reducing reliance on traditional experimental methods that are often time-consuming, expensive, and ethically problematic. By utilizing large datasets on chemical properties, molecular structures, and biological effects, AI models can forecast the environmental and health impacts of chemicals at an early stage, enabling more efficient risk assessments. However, there are several critical challenges and limitations to consider. The accuracy of AI predictions is dependent on the availability of high-quality, comprehensive data, which is often lacking, especially for new or untested chemicals. Furthermore, the interpretability of AI models remains a significant issue, as many models function as "black boxes," making it difficult to understand the rationale behind their predictions. This lack of transparency may hinder trust in AI-driven decision-making. Additionally, AI tools may lead to biases if the data is incomplete. Despite these challenges, AI presents opportunities for the design of greener chemicals by optimizing molecular structures to reduce environmental harm and enhance biodegradability. However, integrating AI into environmental chemistry requires careful consideration of ethical, legal, and regulatory frameworks to ensure the responsible use of technology.

KEYWORDS: Artificial Intelligence; toxicity prediction, machine learning; sustainable chemical design; data quality and interpretability.

1. INTRODUCTION

The proliferation of industrial chemicals in modern society has generated unprecedented challenges in assessing environmental and human health risks[1].Traditionally, toxicity evaluations have depended on in vivo and in vitro methods, which, despite being informative, are time consuming, expensive, and ethically contentious due to animal testing[2]. With the rise of green chemistry principles and sustainability initiatives, there is a pressing demand for faster, more predictive, and less invasive toxicity assessment strategies[3]. Artificial Intelligence (AI) has rapidly emerged as a transformative tool across multiple scientific disciplines, including drug discovery, materials science, and climate modeling [4]. In environmental chemistry, AI particularly machine learning (ML) and deep learning (DL) have demonstrated immense potential in predicting chemical toxicity by recognizing hidden patterns in large datasets of molecular and toxicological information[5]. By doing so, AI offers opportunities to shift toxicity testing from a reactive to a predictive framework, allowing safer product design and proactive regulation[6].

This review provides a concise discussion of AI applications in environmental chemistry. It focuses on predictive toxicology, exposure modeling, and sustainable design, while critically evaluating challenges such as data scarcity, black-box algorithms, algorithmic bias, and regulatory hurdles. By critically evaluating these aspects, the paper highlights how AI can reshape the future of environmental risk assessment and support the global transition toward sustainable chemical management.

2. ARTIFICIAL INTELLIGENCE APPROACHES IN ENVIRONMENTAL CHEMISTRY

AI comprises computational methods designed to mimic human cognition and decision-making. Among its subsets, ML enables systems to learn patterns from data, while DL employs multilayered artificial neural networks to capture complex nonlinear relationships[7].Machine Learning methods commonly used in environmental toxicology include: Support Vector Machines (SSVMs) for classification tasks such as mutagenicity prediction[8].Random Forests (RFs) for acute toxicity and bioaccumulation risk modeling[9].Artificial Neural Networks (ANNs) for non-linear relationships between molecular descriptors and toxicological endpoints[10].Deep Learning architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are increasingly applied to large-scale datasets (e.g., omics data, chemical images) for predicting multi-endpoint toxicity with higher accuracy Table 1 summarizes some commonly used tools for predicting toxicity of chemical products[11].

Table 1: Tools for Predicting Toxicity

Tool / Approach	Main Technique	Common Application	References
ToxCast	Large-scale screening	Identifying how thousands of chemicals interact with biological systems.	4,6,7,9
QSAR Models	Statistical methods	Estimating toxicity of chemicals in water and soil based on structure activity patterns.	5,10
DeepTox	Pattern recognition	Detecting hormone-like (endocrine) activity in industrial and consumer products.	4,5.11
Read-Across	Comparison approach	Inferring toxicity of new substances by comparing them to well-known chemicals.	6,7,10

3. APPLICATIONS OF AI IN TOXICITY PREDICTION

Artificial intelligence (AI) has emerged as a powerful tool for predicting chemical toxicity, offering significant improvements over conventional in vitro and in vivo testing. Traditional methods are often costly, time consuming, and ethically challenging due to reliance on animal studies.

AI-based models, particularly machine learning (ML) and deep learning (DL), enable rapid screening of large chemical libraries to identify potential hazards before extensive laboratory experiments are conducted. **3.1 Predictive Toxicology**

Recent advances in artificial intelligence (AI) have greatly enhanced predictive toxicology by moving beyond traditional QSAR models toward more sophisticated deep learning and transformerbased approaches[1,2]. Models such as ChemBERTa and MolBERT leverage natural language processing techniques to learn chemical representations directly from molecular sequences, providing superior accuracy for endpoints such as mutagenicity, carcinogenicity, and endocrine disruption[3]. These models outperform conventional machine learning methods, especially when applied to diverse datasets of environmental pollutants[4]. In addition, AI has been increasingly applied to emerging contaminants such as pharmaceuticals, legacy pesticides, per- and polyfluoroalkyl substances (PFAS), and microplastics, which often lack extensive experimental toxicity data[5,6]. By combining molecular descriptors with omics datasets, modern deep learning frameworks have demonstrated high predictive power for mixture toxicity, a critical issue given that real-world exposures usually involve chemical cocktails rather than isolated compounds [7,8].

3.2 Risk and Exposure Assessment

AI applications in environmental chemistry extend beyond toxicity prediction to risk and exposure assessment, where the focus is on chemical fate, transport, and bioaccumulation[10].

Machine learning models have been integrated with hydrological and climate data to forecast how pollutants disperse in air, soil, and aquatic systems under changing environmental conditions[9].

Recent studies highlighted the use of graph neural networks (GNNs) to model interactions within soilwater systems, enabling more accurate predictions of pesticide runoff and heavy metal leaching compared with regression-based models[12]. Table 2 presents types of toxicological effects that can be predicted with AI. Similarly, AI-driven exposure models are now incorporating climate change variables such as rainfall intensity and temperature fluctuations, offering more realistic simulations of pollutants mobility in vulnerable ecosystems[13].

Table 2. Types of Toxicological Effects Predicted

Toxic Effect	Approach Used	Example Application	References
Acute Toxicity	Statistical analysis	Estimating lethal dose (LD50) for pesticides or industrial solvents	1,7
Mutagenicity	Data classification	Predicting whether a chemical may damage DNA (e.g., Ames test outcomes)	3,5
Carcinogenicity	Pattern recognition	Long-term predictions of cancercausing potential in industrial chemicals	2,7
Endocrine Disruption	Biological modeling	Identifying substances that may mimic or block hormones in humans or wildlife	2,6

The computational workflow used to predict the toxicity of chemical compounds based on their molecular features (Fig.1) begins with the chemical structure, which represents the unique arrangement of atoms, bonds, and functional groups that determine the physicochemical and biological properties of a compound[14]. Every molecule has a distinct structure, and this structural information forms the foundation for toxicity prediction[15].

The chemical structure is then converted into molecular descriptors and fingerprints. Descriptors are numerical values that quantify specific properties of the molecule, such as molecular weight, polarity, hydrophobicity, and the number of hydrogen bond donors or acceptors[16]. Fingerprints, on the other hand, are simplified digital representations that capture the presence or absence of structural features or functional groups.Together, these descriptors and fingerprints transform the chemical structure into a mathematical format suitable for computational analysis[17].

Finally, this processed information is used to predict toxicity through computational models, often based on machine learning or quantitative structure-activity relationship (QSAR) methods. The skull-andcrossbones symbol displayed on the computer screen in the figure represents the outcome of such predictive modeling, where a compound can be flagged as toxic or non-toxic[18]. This approach provides a rapid, cost-effective, and ethical alternative to traditional laboratory and animal testing, making it highly valuable in drug discovery, environmental safety assessments, and regulatory toxicology.

Figure 1: Relationship between chemical structure and toxicity

Risk and Exposure Assessment

Beyond toxicity, AI facilitates risk assessment by simulating environmental fate, transport, and bioaccumulation of chemicals[18].

Models can predict biodegradation rates, aid persistence classification, bioaccumulation in aquatic food webs, improving ecological impact assessments[8-9], chemical transport in soil and water systems as well as enabling proactive remediation planning[19]. The AI-based workflow for toxicity prediction (Fig. 2), highlights the interconnected stages involved in assessing the potential hazards of chemical compounds[3,5,8]. The process begins with chemical compounds, which serve as the input data for toxicity studies. These compounds undergo data collection, where critical information such as chemical properties, molecular structures, and biological activity data is compiled from experimental databases and computational sources[10-11]. The collected data is then used in the stage of AI model development, which employs machine learning and deep learning approaches to identify patterns and build predictive models[12].

These models enable toxicity prediction, which involves classifying chemicals based on potential toxic effects and conducting risk assessments to evaluate environmental and human health impacts[8,10]. Predictions from AI tools can inform further chemical testing, database expansion, and safer chemical design[20]. This cyclical workflow underscores the integration of data science with environmental chemistry to create faster, cost-effective, and ethically sustainable alternatives to traditional toxicity testing [2,5,7,10,13].

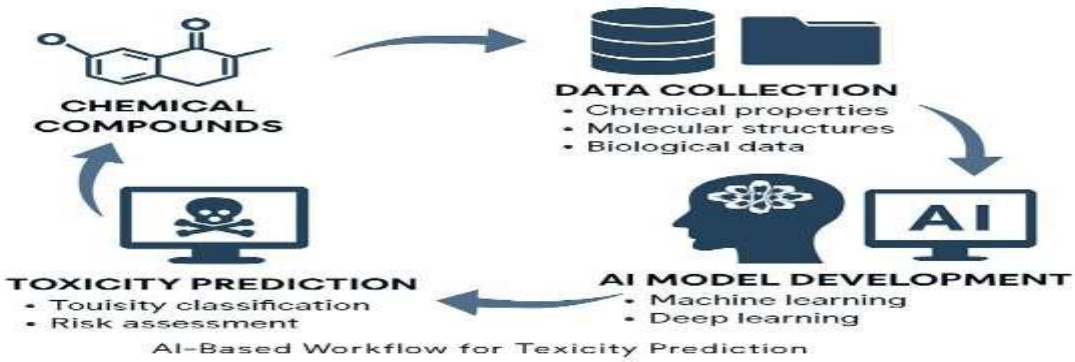

Figure 2: AI-Based Workflow for Toxicity Prediction

3.3 Greener Chemical Design

AI contributes to sustainable chemistry by guiding the design of safer alternatives: Generative models can propose novel molecules with reduced toxicity profiles[2,6,9-10]. Multi-objective optimization balances functionality, stability, and biodegradability[3,5]. Case studies show AI-driven design of flame retardants with minimized endocrine activity[14]. Beyond risk assessment, AI is transforming the design of greener and safer chemicals. Generative models such as variational autoencoders (VAEs) and generative adversarial networks (GANs) are now being applied to propose novel molecules with optimized properties[3,9]. These models allow researchers to balance multiple objectives-such as stability, biodegradability, and reduced toxicity within a single framework, greatly accelerating the development of sustainable alternatives[5]. For example, recent case studies have demonstrated the use of AI-guided molecular optimization to design flame retardants with minimized endocrine-disrupting potential and surfactants with improved biodegradability profiles[13]. Multi-objective optimization frameworks further allow chemists to reduce environmental persistence while maintaining functionality, an essential consideration in green chemistry[11]. The integration of AI in sustainable chemical design not only speeds up discovery but also aligns with regulatory pressures to reduce hazardous substances in consumer products, industrial processes, and environmental applications[14,16].

The comparative accuracy of four machine learning algorithms used for predictive modelling is illustrated in Fig.3. The vertical axis represents accuracy in percentage, ranging from 0% to 100%, while the horizontal axis lists the algorithms: Decision Tree, Random Forest, Support Vector Machine (SVM),

and Neural Network[4,7,9,17].The Decision Tree model achieves the lowest accuracy at 70%, indicating that while it provides a simple and interpretable framework, it may not capture complex patterns in the data effectively[4,7,9]. In contrast, the Random Forest model significantly improves performance, reaching 85% accuracy[11]. This increase can be attributed to its ensemble nature, where multiple decision trees are combined to reduce overfitting and improve generalization[4,6].

The Support Vector Machine (SVM) model achieves 80% accuracy, outperforming the Decision Tree but performing slightly below Random Forest. SVM is known for its effectiveness in handling highdimensional data and finding optimal classification boundaries, though its performance can depend heavily on parameter tuning and kernel selection[13,15]. The Neural Network model records the highest accuracy at 90%, demonstrating its superior ability to capture nonlinear relationships and complex data structures[16].

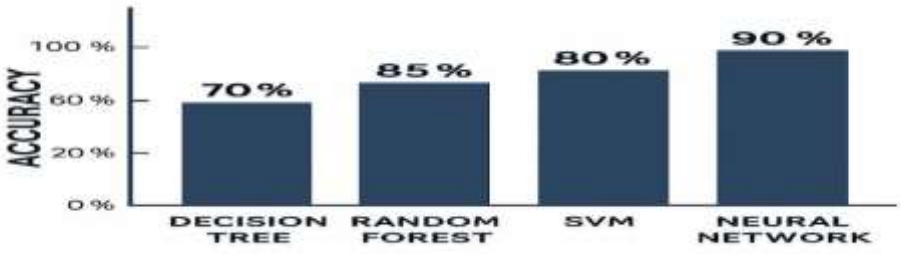

Figure 3: Comparative Accuracy of ML Models

4. CHALLENGES AND LIMITATIONS

Emerging contaminants lack toxicological data, resulting in data gaps that undermine predictive reliability[15]. Biased datasets over representing certain chemical classes can also skew predictions[3,6,9]. The most significant barriers to regulatory acceptance is the “black-box” nature of many DL models as well as ethical concerns regarding algorithmic bias and accountability[4,6,12]. Despite these advances, several challenges limit the widespread adoption of AI in environmental toxicology[21].Data availability remains the most pressing barrier: many emerging contaminants, particularly those prevalent in low- and middle-income regions, lack sufficient toxicological records to train reliable models[22].This creates blind spots in predictions, especially for underrepresented chemical classes such as nanomaterials and novel agrochemicals[23]. Another concern lies in algorithmic bias, where models trained on unbalanced datasets may overrepresent well-studied industrial compounds while neglecting region-specific pollutants such as pharmaceuticals and pesticides in African water systems[24].

5. CONCLUSION

AI can transform environmental chemistry by enabling predictive and sustainable approaches to toxicity assessment. While challenges such as data quality, interpretability, and regulatory acceptance remain, continued development of transparent models, data harmonization, and supportive policies will allow AI-driven approaches to become central in safeguarding both human health and ecosystems.

6. FUTURE PROSPECTS AND RECOMMENDATIONS

The future of AI in environmental toxicology relies on developing explainable AI to improve transparency and regulatory trust, supported by international collaboration to build harmonized, openaccess datasets. Robust validation frameworks are needed to ensure models perform reliably in realworld contexts. Capacity building through interdisciplinary training and stronger partnerships among academia. industry, and regulators will be vital to translate its advances into practical tools for safeguarding human health and the environment.

ACKNOWLEDGEMENTS

The authors gratefully acknowledged the financial support of the Royal Society of Chemistry, for the award of the Researchers Development and Travel Grant (Grant number: D24-6629469040) to

Ayodeji Oluwole Adegun. The fund significantly supported the execution and dissemination of this research work.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

REFERENCES

(1) Bai, C.; Wu, L.; Li, R.; Cao, Y.; He, S.; Bo, X. Machine Learning-Enabled Drug-Induced Toxicity Prediction. *Adv. Sci.* 2025, *12* (16).

(2) Guo, W.; Liu, J.; Dong, F.; Song, M.; Li, Z.; Khan, M. K. H.; Patterson, T. A.; Hong, H. Review of Machine Learning and Deep Learning Models for Toxicity Prediction. *Exp. Biol. Med.* 2023, *48* (21), 1952–1973.

(3) Kostal, J.; Brooks, B. W.; Smith, C. A.; Devineni, G. O Data, Where Art Thou? Revolutionizing Data Sharing to Advance Our Sustainability Goals through Smart Chemical Innovation. *iScience* 2022, *25* (11), 105256.

(4) Su, R.; Wu, H.; Xu, B.; Liu, X.; Wei, L. Developing a Multi-Dose Computational Model for Drug-Induced Hepatotoxicity Prediction Based on Toxicogenomics Data. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 2019, *16*, 1231–1239.

(5) Tang, W.; Chen, J.; Wang, Z.; Xie, H.; Hong, H. Deep Learning for Predicting Toxicity of Chemicals: A Mini Review. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* 2018, *36* (4), 252–271.

(6) Hartung, T. ToxAIcology.The Evolving Role of AI in Predictive Toxicology and Regulatory Science. *ALTEX* 2023, *40* (4), 559–570.

(7) Hartung, T.; Kleinstreuer, N. Challenges and Opportunities for Validation of AI-Based New Approach Methodologies (NAMs) in Toxicology. *ALTEX* 2025, *42* (1), 1–16.

(8) Koutsoukas, A.; St. Amand, J.; Mishra, M.; Huan, J. Predictive Toxicology: Modeling Chemical Induced Toxicological Response Combining Circular Fingerprints with Random Forest and Support Vector Machine. *Front. Environ. Sci.* 2016, *4*, 11.

(9) Khan, A.; Gupta, S.; Gupta, S. K. Multi-Hazard Disaster Studies: Monitoring, Detection, Recovery, and Management, Based on Emerging Technologies and Optimal Techniques. *Int. J. Disaster Risk Reduct.* 2020, *47*, 101642.

(10) Hong, H.; Thakkar, S.; Chen, M.; Tong, W. Development of Decision Forest Models for Prediction of Drug-Induced Liver Injury in Humans Using a Large Set of FDA-Approved Drugs. *Sci. Rep.* 2017, *7*, 17311.

(11) Idakwo, G.; Luttrell, J.; Chen, M.; Hong, H.; Zhou, Z.; Gong, P.; Zhang, C. A Review on Machine Learning Methods for *In Silico* Toxicity Prediction. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* 2018, *36*, 169–191.

(12) Maxwell, A.; Li, R.; Yang, B.; Weng, H.; Ou, A.; Hong, H.; Zhou, Z.; Gong, P.; Zhang, C. Deep Learning Architectures for Multi-Label Classification of Intelligent Health Risk Prediction. *BMC Bioinformatics* 2017, *18*, 121.

(13) Jiang, C.; Yang, H.; Di, P.; Li, W.; Tang, Y.; Liu, G. *In Silico* Prediction of Chemical Reproductive Toxicity Using Machine Learning. *J. Appl. Toxicol.* 2019, *39*, 844-854.

(14) Minerali, E.; Foil, D. H.; Zorn, K. M.; Lane, T. R.; Ekins, S. Comparing Machine Learning Algorithms for Predicting Drug-Induced Liver Injury (DILI). *Mol. Pharmaceutics* 2020, *17*, 2628-2637.

(15) Lakshmikantha, V.; Hiriyannagowda, A.; Manjunath, A.; Patted, A.; Basavaiah, J.; Anthony, A. A. IoTBased Smart Water Quality Monitoring System. *Glob. Transitions Proc.* 2021, *2* (2), 181-186.

(16) Karim, A.; Singh, J.; Mishra, A.; Newton, M. A. H.; Sattar, A. Toxicity Prediction by Multimodal Deep Learning. *arXiv* 2019, arXiv:1907.08333.

(17) Jovanović, G.; Bezdan, T.; Romanić, S. H.; Matek Sarić, M.; Biošić, M.; Mendaš, G.; Stojić, A.; Perišić, M. Application of the AI-Based Framework for Analyzing the Dynamics of Persistent Organic Pollutants (POPs) in Human Breast Milk. *Toxics* 2025, *13* (8), 631.

(18) Huang, X.; Chen, J.; Liu, P. Assessing Chemical Exposure Risk in Breastfeeding Infants: An Explainable Machine Learning Model for Human Milk Transfer Prediction. *Ecotoxicol. Environ. Saf.* 2025, *289*, 117707.

(19) Jovanović, G.; Herceg, R. S.; Stojić, A.; Klinčić, D.; Matek, S. M.; Grzunov, L. J.; Popović, A. Introducing Modeling Techniques in the Research of POPs in Breast Milk-A Pilot Study. *Ecotoxicol. Environ. Saf.* 2019, *172*, 341-347.

(20) Radić, N.; Perišić, M.; Jovanović, G.; Bezdan, T.; Stanišić, S.; Stanić, N.; Stojić, A. An AI-Based Framework for Characterizing the Atmospheric Fate of Air Pollutants within Diverse Environmental Settings. *Atmosphere* 2025, *16*, 231.

(21) Jovanović, G.; Perišić, M.; Bačanin, N.; Živković, M.; Stanišić, S.; Štrumberger, I.; Alimpić, F.; Stojić, A. Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing PAHs Environmental Fate. *Toxics* 2023, *11*, 394.

(22) Mendaš, G.; Jakovljević, I.; Herceg, R. S.; Fingler, S.; Jovanović, G.; Matek, S. M.; Pehnec, G.; Popović, A.; Stanković, D. Presence of Polycyclic Aromatic Hydrocarbons and Persistent Organochlorine Pollutants in Human Milk: Evaluating Their Levels, Association with Total Antioxidant Capacity, and Risk Assessment. *Sci. Total Environ.* 2024, *931*, 172911.

(23) Janakiraman, S.; Sha, R.; Mani, N. K. Recent Advancements in Point-of-Care Detection of Contaminants and Biomarkers in Human Breast Milk: A Comprehensive Review. *Sens. Actuators Rep.* 2025, *9*, 100280.

(24) Chi, Z. H.; Goodyer, C. G.; Hales, B.; Bayen, S. Characterization of Different Contaminants and Current Knowledge for Defining Chemical Mixtures in Human Milk and Infant Health Risk Assessment from 1976 to the Present. *Food Chem. Toxicol.* 2023, *179*, 113990.