frontiers | Frontiers in Nanotechnology

TYPE Perspective
PUBLISHED 05 March 2025
DOI 10.3389/fnano.2025.1559053

OPEN ACCESS

EDITED BY
Cristina Satriano,
University of Catania, Italy

REVIEWED BY
Vadanasundari Vedarethinam,
Cornell University, United States

*CORRESPONDENCE
Sarmistha Saha,
sarmistha_pharmacol@yahoo.com

RECEIVED 11 January 2025
ACCEPTED 19 February 2025
PUBLISHED 05 March 2025

CITATION
Saha S (2025) How nanoinformatics could pave the way to safer design of engineered nanomaterials?
Front. Nanotechnol. 7:1559053.
doi: 10.3389/fnano.2025.1559053

How nanoinformatics could pave the way to safer design of engineered nanomaterials?

Sarmistha Saha*

Department of Biotechnology, Institute of Applied Sciences & Humanities, GLA University, Mathura, Uttar Pradesh, India

Applications of nanotechnology have rapidly expanded across various fields, including materials, energy, medicine, and the environment. The toxicity of nanomaterials (NMs) and nanotechnology to living organisms, as well as potential negative environmental impacts, poses significant challenges. Laboratory approaches for assessing the negative impacts of NMs are costly, time-consuming, and frequently fall behind the development of novel materials. Therefore, intelligent systems of informatics to forecast their toxicity potentials are a possible alternative option. Research on nanotoxicology has generated extensive and diverse datasets. However, data alone does not equate to information. Since little is known about how to extract meaningful information from large streams of data, a broad discussion of potential concerns still exists. In this regard, we present a perspective on how the "big data" can change the paradigm toward data-centric computational materials research to decipher the toxicity mechanisms of NMs for their efficient risk assessment management to provide safe-by-design standards for the sustainable advancement of nanotechnology. We will also discuss the challenges of the current data-driven research into the nano realm.

KEYWORDS

nanomaterials, computational toxicity, artificial intelligence, machine learning, algorithm

1 Introduction

Across the globe, nanotechnology is driving a remarkable transformation that impacts a wide range of scientific disciplines. Yet, we must remember the principle that "with great power comes great responsibility." The nanoworld encompasses a fascinating array of structures and interactions at the nanoscale (around 1–100 nm). To visualize this scale, consider that a nanometer is about one-hundred-thousandth the diameter of a human hair (Foster, 2017). The growing popularity of nanomaterials, their distinct physicochemical properties, and increased investment in research and development could all be contributing factors to their expanding market profile. However, concerns about the impact of nanotechnologies on human health, safety, and the environment may hinder their business adoption and slow market growth. Individuals will inevitably encounter NMs during production, transportation, daily use, and disposal, as these products are designed to enhance the quality of human life. Nanomaterials can enter the body through inhalation, oral absorption, skin absorption, or medical applications. Once inside, they have the potential to accumulate in various tissues and organs. It is important to understand how these materials can navigate biological barriers, such as the skin, blood-brain barrier, reproductive system, and the interface between air and blood in the lungs (Tsuda et al.,

2019; Souza et al., 2021). Thus, it is evident that more infrastructure and resources are required for the safety assessment of next-generation nano-enabled devices. This should help to improve techniques, increase understanding, and hasten commercial adoption (Shatkin, 2020).

Harnessing the unique physicochemical properties of NMs such as size, shape, and surface chemistry offers exciting opportunities to effectively address the needs of a wide range of applications. However, it is important to consider how these alterations may influence their behavior in biological contexts, such as cellular uptake, biodistribution, and toxicity. The numerous guidelines that have already been released, specifically by the ISO, American Society for Testing and Materials (ASTM), and Organization for Economic Cooperation and Development (OECD), reflect the significant effects of physicochemical properties on the safety and effectiveness of NMs. The most important factors are undoubtedly the size distribution of the particles, surface functionalization or treatment, shape or morphology, and surface area (Cosnier et al., 2021). It has also been demonstrated that surface reactivity, agglomeration/aggregation condition, and dissolution rate are highly relevant (Gutierrez et al., 2023). There are many other NM properties that have not been tested, though, and some of them are already polydisperse after production. They may also alter over the course of their life cycle, depending on the biological medium or environment in which they are suspended or incorporated. Most of the data in nanotoxicity research comes from traditional one-at-a-time experiments conducted in various laboratory settings (Krug, 2014). The quality of this data is often compromised by insufficient or poor characterization of NMs and inadequate toxicity testing. Additionally, the processes of data collection, annotation, curation, and integration are often time-consuming, leading to further challenges in obtaining reliable data (Jeliazkova et al., 2021).

The necessary techniques may conflict, and processing and analyzing the generated data may prove difficult. The data generated by *in silico* approaches such as machine learning (ML) techniques that can create intricate networking systems that link the physicochemical qualities of nanomaterials and their destiny within the human body. Predictive modeling and *in silico* tools are essential informatic approaches for the risk assessment of nanomaterials in nanosafety. These methods decisively aim to predict the properties of nanomaterials, analyze how nanomaterials interact at cellular and molecular level, evaluate the transformations of nanomaterials caused by environmental factors or biological processes, and determine whether the byproducts of nanomaterials exhibit toxicity compared to their original forms (Lynch et al., 2021). In this context, the emerging field of nanoinformatics, which blends informatics and data from nanomaterials, is rapidly developing (Barnard et al., 2019). Risks to the environment and human health can be avoided by creating predictive modeling frameworks to forecast possible negative impacts of nanoparticles (Labouta et al., 2019).

Nanoinformatics is a developing field that sits at the intersection of multiple disciplines, including nanotechnology, medicine, biology, chemistry, physics, and informatics (which encompasses information technologies and computer science). It involves the practical application of information technologies to collect, store, retrieve, and process data, information, and knowledge related to the physicochemical properties of nanoparticles, nanomaterials, and nanodevices (Maojo et al., 2010). Nanoinformatics encompasses various applications, including the design and characterization of nanoparticles, as well as modeling and simulation. It also involves data integration and exchange, linking nanoparticle information to clinical data, and semantic annotation and retrieval. Additionally, it can contribute to the development of domain ontologies, terminologies, and standards, along with data and text mining for nanomedical research (Maojo et al., 2010).

Due to the limited availability of model-friendly databases and suitable nanodescriptors, the application of AI in nanotoxicology remains relatively uncommon (Winkler, 2020). Unfortunately, most current nanodescriptors are not specific to NMs; they were originally developed for small molecules (Karakus and Winkler, 2021). As a result, there is inadequate representation of complex NM structures and their interactions with biological systems and the environment. Additionally, challenges such as feature selection, model optimization, and interpretation arise when machine learning (ML), a subset of AI techniques, is applied in nanotoxicology (Artrith et al., 2021) (Figure 1). The risk of becoming overloaded with data and overlooking important details in nanotoxicology arises from the gathering of vast amounts of nanotoxicity data (Feng et al., 2021). Here, we will discuss how to incorporate big data into nanotoxicity databases to unravel NMs property-toxicity relationships. To understand the advanced data analysis and modeling techniques that nanoinformatics uses to design engineered nanomaterials, we will also discuss the challenges, opportunities, and solutions associated with them. We can close the gap between vast amounts of nanotoxicity data and important knowledge by using model-friendly databases, AI, and molecular simulations.

2 Informatics and nanotechnology: a powerful alliance

Nanoinformatics offers benefits in precisely assessing and forecasting biological reactions to the characteristics of known and unknown NMs, which aids in the discovery of novel functional NMs and the avoidance of adverse effects. The development of advanced predictive modeling techniques grounded in the principles of findable, accessible, interoperable, and reusable (FAIR) data has accelerated due to this concept. While the scientific community has embraced the FAIR principles, their widespread application has been limited by varying interpretations. Consequently, sophisticated methods are being developed to specifically apply the FAIR principles in the life sciences, facilitating the reuse and sharing of contemporary research data (Ammar et al., 2020).

ML techniques can accomplish big-data, high-throughput screening and reduce the amount of time and labor required for material testing, which will improve the creation and use of nanomaterials (Jia et al., 2021). To assess the impact of various physicochemical descriptors such as shape, size, zeta potential, concentration, and intracellular uptake and their influence on interactions with the cell membrane at sublethal concentrations, Singh et al. developed a ML-based approach. Their study identified nuclear area factors and cell shape index as key

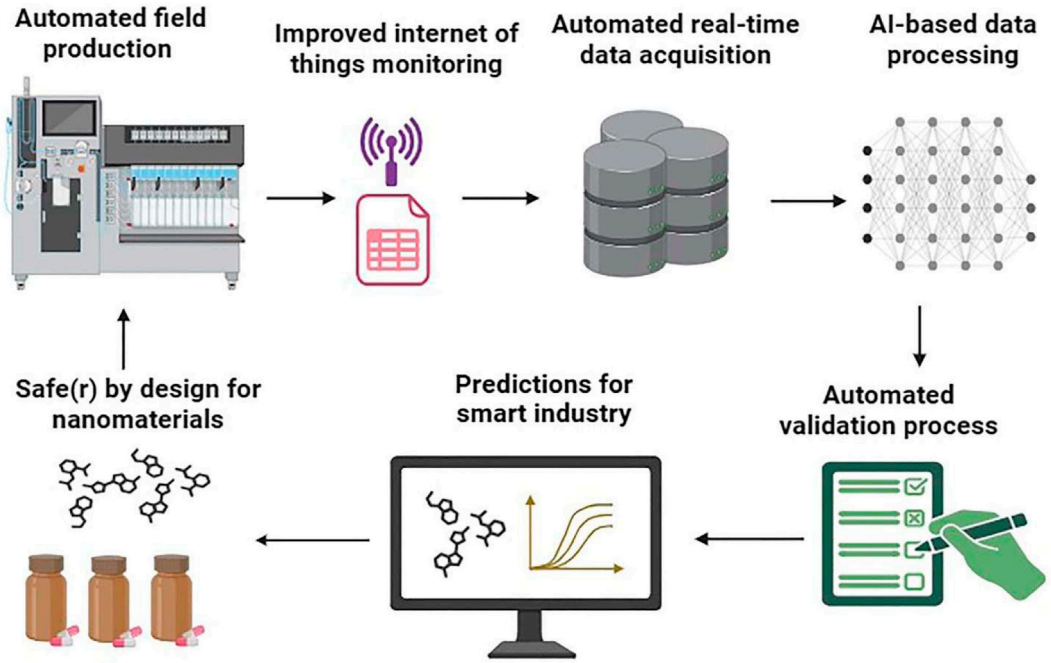

FIGURE 1
Role of AI/ML in the safe designing of nanomaterials.

descriptors that reflect the specific changes induced by nanomaterials in the epithelial Madin-Darby canine kidney (MDCK) cell line, as determined by the ML system (Singh et al., 2021).

The use of quantitative structure-activity/toxicity relationship (QSAR/QSTR)-based methods for safety and risk assessment has grown in popularity (Nyström and Fadeel, 2012; Raies and Bajic, 2016). To address regulatory concerns, several development standards have been put forth. A useful tool for predicting the molecular properties and biological interactions of pharmaceuticals is the contemporary three-dimensional quantitative structure-activity relationship (3D-QSAR) molecular docking (Lavanya and Sasipriya, 2022). Nevertheless, 3D QSAR and traditional molecular descriptors are unable to convey the nanoparticles' specificity. Furthermore, there are not many established nano-QSAR modeling techniques (Jia et al., 2021). Quantitative structure nanotoxicity relationship (QNAR) models, also known as nano-QSAR models, have been developed in this context to address certain descriptive aspects of nanomaterials (Nyström and Fadeel, 2012; Raies and Bajic, 2016; Fourches et al., 2010).

A tissue-specific classification model for predicting the neurotoxicity of nanoparticles *in vitro* systems was designed using the random forest (RF) approach (Furxhi and Murphy, 2020). The toxicity effects of metal oxide nanoparticles on *Escherichia coli* were predicted using seven machine learning methods, including linear discriminant analysis (LDA), NB, RFs, and AdaBoost. This provided a scientific foundation for the creation and manufacturing of safe NMs (Kar et al., 2021).

A tree-based approach to feature network interaction analysis and RF feature importance (TBRFA) has recently been developed (Yu et al., 2021). In addition to accurately predicting the lung burden and pulmonary immune responses of nanoparticles, our approach creates feature interaction networks that aid in the interpretation or justification of NM effects. In order to predict genetic responses to NMs under different experimental assay conditions, a perturbation theory ML (PTML)-QSTR model was also employed. This model demonstrated an accuracy of over 96%,85 indicating that it is a dependable tool for quickly and economically evaluating genetic responses to NMs.

Nanotox has recently developed a unique feature space based on the internal and external physicochemical characteristics of nanomaterials, along with data from the periodic table. This feature space is associated with an assay technique and cell source. To improve the performance of the machine learning models, the training data was adjusted to address class imbalance, the feature space was optimized for hyperparameters, and multicollinearity was minimized (Yan et al., 2019). As a result of establishing the ideal hypothesis space, the models achieved over 96% balanced accuracy. Research has shown that the toxicity of nanoparticles is influenced by several factors, including dose, exposure duration, hydrodynamic size, periodic table position, and electronegativity (Eneg). Additionally, predictors identified as nontoxic include surface area (SurfArea), conduction band energy (Ec), core size (CoreSize), and the count of oxygen atoms (NOxygen) (Yan et al., 2019).

To investigate the nanotoxicity of zigzag single-walled carbon nanotubes (SWCNTs) on mitochondrial channels, González-Durruthy et al. (2020) used molecular docking and dynamic simulations. Their work strategically targets the selective blockage of the ATP entry point in the human mitochondrial voltage-dependent anion-selective channel (hVDAC1). By employing these advanced computational models, the authors

present a compelling *in silico* approach for evaluating not only the toxicological risks but also the therapeutic potential of carbon nanotube-based nanomaterials, all while considering the crucial role of mitochondrial function.

Applications for molybdenum disulfide (MoS_2) nanomaterial in biomedicine are intriguing. It exhibits characteristics that are comparable to those of carbon-based nanoparticles, showing promise for both imaging applications suitable for cancer theranostics and medication delivery against microbes. Molecular dynamics (MD) simulations and electrophysiology experiments, however, showed that the MoS2 nanoflake might attach to a fissure at the extracellular side of the potassium channel's voltage sensor domain (VSD). This could be part of the underlying nanotoxicological mechanism and restrict the use of MoS_2 (Gu et al., 2018).

A computational pharmacokinetic model called ISD3 was developed to ensure accurate dosimetry based on experimental conditions and the physicochemical properties of the particles (Poli et al., 2020). Additionally, an intriguing database was established to create a cloud of *in vitro* and *in vivo* toxicokinetic data, which correlates concentration with time (CvT) (Sayre et al., 2020). Multiscale models have recently been integrated into an *in silico* risk assessment framework developed by the NanoSolveIT.eu project. This framework is based on the "Integrated Approach to Testing and Assessment" (IATA) and enables the implementation of a safe-by-design strategy for nanomaterials, addressing both industry and regulatory challenges (Karatzas et al., 2020; Papadiamantis et al., 2020).

Metal-organic frameworks (MOFs) with structural flaws can exhibit desirable characteristics, unlike defect-free crystals. However, identifying these flaws in MOFs can be challenging. To address this, Wu et al. (2020) created a library of 425 MOFs that feature significant missing-linker defects, which was used to train logistic regression (LR) and RF algorithms. In a couple of seconds, a deep learning system was able to extract and classify hundreds of lattice faults from scanning transmission electron microscopy pictures (Maksov et al., 2019). For the construction of innovative and promising MOFs, a computational approach that combined recurrent neural networks and a Monte Carlo tree search demonstrated notable efficiency (Zhang et al., 2020). Functionalizing MOFs with groups such as hydroxyl, thiol, cyano, amino, or nitro significantly improves their ability to capture CO_2, according to ML prediction results (Anderson et al., 2018). These predictions are notably more accurate when the types of atoms in the structure are used as descriptors, rather than the previously used building blocks (Fanourgakis et al., 2020). Data mining from a computational screening library of over 300,000 MOFs has identified strong CO_2-binding sites, achieving R^2 values greater than 0.98 (Boyd et al., 2019). The MOFs produced through this approach exhibited high adsorption capacity for CO_2 from anthropogenic sources and showed resilience against interference from nitrogen and water.

The delivery, release, and therapeutic efficacy of therapeutic drugs are significantly influenced by protein coronas (Akhter et al., 2021). It is challenging for conventional linear regression models to forecast the creation and makeup of protein coronas. Because the RF method can handle heterogeneous big data, it can resolve this issue (Findlay et al., 2018). Cellular recognition mediated by functional protein coronas was correctly predicted by the technique (Ban et al., 2020). To forecast the buildup of NMs, a supervised deep neural network was employed (Lazarovits et al., 2019). With an accuracy of up to 94%, the network was able to predict the accumulation of nanoparticles in the liver and spleen, opening the door to the development of surface chemistries tailored to the body as well as the prediction of nanoparticle behavior. The brain microenvironment and NMs characteristics including size and protein adsorption were predicted using an ANN (Curtis et al., 2019). By using this data to create predictive models of NMs transport, nanotherapeutic platforms can be designed to get past biological obstacles and accomplish localized delivery. Oh et al. (2016) conducted a meta-analysis that gathered 1,741 data samples from published literature regarding the cellular toxicity of cadmium-containing semiconductor quantum dots. There are many benefits to building datasets using meta-analysis and then using machine learning methods to examine how model animals and NMs interact. For instance, high-throughput screening phenotyping of *Caenorhabditis elegans* embryos has been accomplished using machine learning (Atakan et al., 2020).

Natural language processing (NLP) is the application of computers to useful written language activities, like information extraction and analysis from unstructured text (Jurafsky and Martin, 2014). NLP applications' utilization of human language expertise sets them apart from other data processing methods. A large number of NLP applications make use of literature that has been extracted from databases. To make sure that the documents being examined by the NLP systems contain pertinent information on manufactured nanomaterials, information retrieval, document classification, and pattern matching techniques are frequently used (Chau et al., 2006). A web-based application named the Nanomaterial Environmental Impact data Miner (NEIMiner) was developed using Drupal and a content management system (CMS) (Tang et al., 2013). NEIMiner consists of four main components including Data integration, Data management and access, Model creation, and a Nanomaterial Environmental Impact (NEI) modeling framework. This framework is similar to the Framework for Risk Analysis of Multi-Media Environmental Systems (FRAMES). Using pattern matching approaches, the Nanotoxicity Searcher is a tool for automatically annotating literature in nanomedicine and nanotoxicology (García-Remesal et al., 2013). To determine the names of nanomaterials (NANO), possible exposure routes (EXPO), target organs and/or creatures (TARGET), and forms of toxicity/damage (TOXIC), the team employed ABNER, a biomedical named entity recognizer (García-Remesal et al., 2013). The linear-chain conditional random fields (CRFs) supervised machine learning technique from Mallet, an open source, freely accessible Java-based statistical NLP toolkit, is included in ABNER.

3 Addressing challenges

Beyond conventional research, nanoinformatics adds fields like pharmacodynamics, sophisticated quantum physics and chemistry, including nanotoxicity, and new imaging models. NMs present unique challenges in nanotoxicology due to their significant size compared to organic molecules, their existence as distributions rather than discrete entities, and the intricate nature of their interactions with the surrounding corona, which forms the biologically relevant entity. This complexity hinders the effectiveness of molecular simulations applications in this field. Large computer resources are needed to specify complicated NMs systems in particular contexts and run simulations over time and length scales that are biologically relevant (Casalini et al., 2019; Foreman-Ortiz et al., 2020). Some atomistic details are unavoidably lost and the simulation precision is decreased, even though coarse-graining can enable simulation over larger time and length scales (Kmiecik et al., 2016). The other challenge is that molecular simulations are unable to adequately represent the biological medium's true complexity (Marrink et al., 2019). Most simulations view NMs as single molecules or interfaces in biological systems, however, NMs likely interact with multiple proteins simultaneously or sequentially, which is important for cell signaling. Thus, we need more studies on how NMs interact with various proteins, as well as individual proteins.

Alternatively, deep learning, however, is heavily reliant on data; the more data obtained, the better the results. Furthermore, deep learning is difficult to interpret and involves a lot of computations. It is unclear whether deep learning is superior to classical ML given the limited amount of NMs data, so a comparison between the two approaches is preferable in some situations. Models with few parameters (low complexity) and/or a prior strength should be employed when there are only limited data sets available. In this context, a "prior" can be interpreted as any presumption about the behavior of the data. For example, in linear regression, the models assume only linear interactions and the number of parameters can be readily changed.

It is important to note that Bayesian networks have been applied to datasets of all sizes, despite not being widely recommended. Effective modeling of tiny datasets is necessary due to the current scarcity of nanotoxicity data (Oksel et al., 2015). However, less complex algorithms trained with more data can outperform even the greatest algorithm taught with small datasets (Baldassi et al., 2016). Data scarcity can be addressed by integrating datasets and/or databases, which leads to the creation of new information and hypotheses (Karcher et al., 2018). The significance of data integration in nanotechnology was emphasized by Karcher et al. (2018), who also offered suggestions for furthering integration. The application of text mining algorithms is also currently hindered by the diverse and disparate data sources that offer data in various formats. Additionally, the machine's readability is limited by the lack of standards in data reporting (Plata and Jankovic, 2021). Up until recently, large-scale text mining was also constrained by paywall and copyright concerns.

Through the creation and integration of cutting-edge *in vitro* models, *in silico* techniques such as machine learning and artificial intelligence (AI)-driven predictive models, and first-principles computational modeling of NMs' properties, interactions, and effects on living systems, CompSafeNano tackles important issues in nanosafety. The creation of atomistic and quantum-mechanical descriptors for different NMs, the assessment of their interactions with biological systems, and the creation of prediction models for NM risk assessment have all advanced significantly. In order to ensure compliance with the FAIR (Findable, Accessible, Interoperable, Reusable) data principles, the CompSafeNano project has also concentrated on improving data management procedures and putting data reporting templates into place and further standardizing them.

4 Conclusion

Over the past several years, data-driven materials research has advanced rapidly, spawning an interdisciplinary field that combines data science, software, chemistry, and physics. The size and shape of crystals, along with active surface sites and the chemical environment, are key factors influencing the electrical properties and reactivity of nanomaterials. The characteristics related to size, surface, and environment such as surface energies, adsorption sites, diverse phase diagrams, and intricate reaction networks are essential for designing NMs data. We anticipate that advancements in the accessibility, quantity, correctness, and complexity of data on NMs will significantly enhance this subject as large-scale computational techniques and high-throughput automated laboratories with data curation investments become more commonplace. In addition to the current computational resources, we highlight the potential for collecting and sharing experimental results with annotated information. These approaches in nanotoxicology are still in their infancy, nevertheless, and they face many computational and experimental challenges. In summary, this perspective offers several valuable insights. 1. ML plays a vital role in identifying the key characteristics that contribute to specific hazards, enhancing our understanding of risk factors. 2. By facilitating the analysis of large omics datasets, ML supports the discovery of modes of action patterns, which are invaluable for formulating hypotheses in grouping methods. 3. Omics approaches encourage a shift from focusing solely on individual endpoints to gaining a more comprehensive mechanistic understanding across multiple endpoints derived from a single experiment. 4. Additionally, leveraging techniques from other areas of AI, such as image analysis and NLP, can significantly enhance the automation of information extraction and interlinking regarding nanomaterial toxicity. These insights collectively pave the way for more effective approaches to studying and managing nanotoxicity.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SS: Conceptualization, Data curation, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

Authors acknowledge their respective department and institutions for providing support and facilities.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akhter, M. H., Khalilullah, H., Gupta, M., Alfaleh, M. A., Alhakamy, N. A., Riadi, Y., et al. (2021). Impact of protein corona on the biological identity of nanomedicine: understanding the fate of nanomaterials in the biological milieu. *Biomed* 9 (10), 1496. doi:10.3390/biomedicines9101496

Ammar, A., Bonaretti, S., Winckers, L., Quik, J., Bakker, M., Maier, D., et al. (2020). A semi-automated workflow for FAIR maturity indicators in the life sciences. *Nanomaterials* 10 (10), 2068. doi:10.3390/nano10102068

Anderson, R., Rodgers, J., Argueta, E., Biong, A., and Gomez-Gualdron, D. A. (2018). Role of pore chemistry and topology in the CO_2 capture capabilities of MOFs: from molecular simulation to machine learning. *Chem. Mater.* 30 (18), 6325–6337. doi:10.1021/acs.chemmater.8b02257

Artrith, N., Butler, K. T., Coudert, F., Han, S., Isayev, O., Jain, A., et al. (2021). Best practices in machine learning for chemistry. *Nat. Chem.* 13, 505–508. doi:10.1038/s41557-021-00716-z

Atakan, H. B., Alkanat, T., Cornaglia, M., Trouillon, R., and Gijs, M. A. M. (2020). Automated phenotyping of *Caenorhabditis elegans* embryos with a high-throughput-screening microfluidic platform. *Microsyst. Nanoeng.* 6 (1), 24. doi:10.1038/s41378-020-0132-8

Baldassi, C., Borgs, C., Chayes, J. T., Ingrosso, A., Lucibello, C., Saglietti, L., et al. (2016). Unreasonable effectiveness of learning neural networks: from accessible states and robust ensembles to basic algorithmic schemes. *Proc. Natl. Acad. Sci. U. S. A.* 113, E7655–E7662. doi:10.1073/pnas.1608103113

Ban, Z., Yuan, P., Yu, F., Peng, T., Zhou, Q., and Hu, X. (2020). Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles. *Proc. Natl. Acad. Sci. U. S. A.* 117 (19), 10492–10499. doi:10.1073/pnas.1919755117

Barnard, A. S., Motevalli, B., Parker, A. J., Fischer, J. M., Feigl, C. A., and Opletal, G. (2019). Nanoinformatics, and the big challenges for the science of small things. *Nanoscale* 7, 19190–19201. doi:10.1039/c9nr05912a

Boyd, P. G., Chidambaram, A., Garcia-Diez, E., Ireland, C. P., Daff, T. D., Bounds, R., et al. (2019). Data-driven design of metalorganic frameworks for wet flue gas CO_2 capture. *Nature* 576, 253–256. doi:10.1038/s41586-019-1798-7

Casalini, T., Limongelli, V., Schmutz, M., Som, C., Jordan, O., Wick, P., et al. (2019). Molecular modeling for nanomaterial-biology interactions: opportunities, challenges, and perspectives. *Front. Bioeng. Biotechnol.* 7, 268. doi:10.3389/fbioe.2019.00268

Chau, M., Huang, Z., Qin, J., Zhou, Y., and Chen, H. (2006). Building a scientific knowledge web portal: the NanoPort experience. *Decis. Support Syst.* 42, 1216–1238. doi:10.1016/j.dss.2006.01.004

Cosnier, F., Seidel, C., Valentino, S., Schmid, O., Bau, S., Vogel, U., et al. (2021). Retained particle surface area dose drives inflammation in rat lungs following acute, subacute, and subchronic inhalation of nanomaterials. *Part. Fibre Toxicol.* 18, 29. doi:10.1186/s12989-021-00419-w

Curtis, C., McKenna, M., Pontes, H., Toghani, D., Choe, A., and Nance, E. (2019). Predicting *in situ* nanoparticle behavior using multiple particle tracking and artificial neural networks. *Nanoscale* 11 (46), 22515–22530. doi:10.1039/c9nr06327g

Fanourgakis, G. S., Gkagkas, K., Tylianakis, E., and Froudakis, G. E. (2020). A universal machine learning algorithm for large-scale screening of materials. *J. Am. Chem. Soc.* 142 (8), 3814–3822. doi:10.1021/jacs.9b11084

Feng, R., Yu, F., Xu, J., and Hu, X. (2021). Knowledge gaps in immune response and immunotherapy involving nanomaterials: databases and Artificial Intelligence for material design. *Biomaterials* 266, 120469. doi:10.1016/j.biomaterials.2020.120469

Findlay, M. R., Freitas, D. N., Mobed-Miremadi, M., and Wheeler, K. E. (2018). Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. *Environ. Sci. Nano* 5 (1), 64–71. doi:10.1039/c7en00466d

Foreman-Ortiz, I. U., Liang, D., Laudadio, E. D., Calderin, J. D., Wu, M., Keshri, P., et al. (2020). Anionic nanoparticle-induced perturbation to phospholipid membranes affects ion channel function. *Natl. Acad. Sci. U. S. A.* 117, 27854–27861. doi:10.1073/pnas.2004736117

Foster, B. (2017). "The world of nanotechnology," in *Nano-safety: what we need to know to protect workers*. Editors D. E. Fazarro, W. Trybula, J. Tate, and C. Hanks (Boston: De Gruyter), 1–40.

Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S. Y., Mumper, R. J., et al. (2010). Quantitative nanostructure-activity relationship modeling. *ACS Nano* 4 (10), 5703–5712. doi:10.1021/nn1013484

Furxhi, I., and Murphy, F. (2020). Predicting *in vitro* neurotoxicity induced by nanoparticles using machine learning. *Int. J. Mol. Sci.* 21, 5280. doi:10.3390/ijms21155280

García-Remesal, M., García-Ruiz, A., Pérez-Rey, D., de la Iglesia, D., and Maojo, V. (2013). Using nanoinformatics methods for automatically identifying relevant nanotoxicology entities from the literature. *Biomed. Res. Int.* 2013, 1–9. doi:10.1155/2013/410294

González-Durruthy, M., Giri, A. K., Moreira, I., Concu, R., Melo, A., Ruso, J. M., et al. (2020). Computational modeling on mitochondrial channel nanotoxicity. *Nano Today* 34, 100913. doi:10.1016/j.nantod.2020.100913

Gu, Z., Plant, L. D., Meng, X. Y., Perez-Aguilar, J. M., Wang, Z., Dong, M., et al. (2018). Exploring the nanotoxicology of MoS_2: a study on the interaction of MoS_2 nanoflakes and K^+ channels. *ACS Nano* 12 (1), 705–717. doi:10.1021/acsnano.7b07871

Gutierrez, C. T., Loizides, C., Hafez, I., Brostrøm, A., Wolff, H., Szarek, J., et al. (2023). Acute phase response following pulmonary exposure to soluble and insoluble metal oxide nanomaterials in mice. *Part. Fibre Toxicol.* 20, 4. doi:10.1186/s12989-023-00514-0

Jeliazkova, N., Apostolova, M. D., Andreoli, C., Barone, F., Barrick, A., Battistelli, C., et al. (2021). Towards FAIR nanosafety data. *Nat. Nanotechnol.* 16 (644), 644–654. doi:10.1038/s41565-021-00911-6

Jia, Y., Hou, X., Wang, Z., and Hu, X. (2021). Machine learning boosts the design and discovery of nanomaterials. *ACS Sustain. Chem. Eng.* 9, 6130–6147. doi:10.1021/acssuschemeng.1c00483

Jurafsky, D., and Martin, J. (2014). *Speech and language processing*. Englewood Cliffs, NJ, U.S.A.: Pearson.

Kar, S., Pathakoti, K., Tchounwou, P. B., Leszczynska, D., and Leszczynski, J. (2021). Evaluating the cytotoxicity of a large pool of metal oxide nanoparticles to *Escherichia coli*: mechanistic understanding through *in vitro* and *in silico* studies. *Chemosphere* 264, 128428. doi:10.1016/j.chemosphere.2020.128428

Karakus, C. O., and Winkler, D. A. (2021). Overcoming roadblocks in computational roadmaps to the future for safe nanotechnology. *Nano Futur* 5, 022002. doi:10.1088/2399-1984/abe560

Karatzas, P., Melagraki, G., Ellis, L. A., Lynch, I., Varsou, D., Afantitis, A., et al. (2020). Development of deep learning models for predicting the effects of exposure to engineered nanomaterials on *Daphnia magna*. *Small* 16, 2001080. doi:10.1002/smll.202001080

Karcher, S., Willighagen, E. L., Rumble, J., Ehrhart, F., Evelo, C. T., Fritts, M., et al. (2018). Integration among databases and data sets to support productive nanotechnology: challenges and recommendations. *NanoImpact* 9, 85–101. doi:10.1016/j.impact.2017.11.002

Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chem. Rev.* 116, 7898–7936. doi:10.1021/acs.chemrev.6b00163

Krug, H. F. (2014). Nanosafety research-Are we on the right track? *Angew. Chem. Int. Ed.* 53, 12304–12319. doi:10.1002/anie.201403367

Labouta, H. I., Asgarian, N., Rinker, K., and Cramb, D. T. (2019). Meta analysis of nanoparticle cytotoxicity via data-mining the literature. *ACS Nano* 13 (2), 1583–1594. doi:10.1021/acsnano.8b07562

Lavanya, B., and Sasipriya, G. (2022). "Study on emerging machine learning trends on nanoparticles-nanoinformatics," in *Advances in intelligent systems and computing*. Singapore: Springer, 443–458.

Lazarovits, J., Sindhwani, S., Tavares, A. J., Zhang, Y., Song, F., Audet, J., et al. (2019). Supervised learning and mass spectrometry predicts the *in vivo* fate of nanomaterials. *ACS Nano* 13 (7), 8023–8034. doi:10.1021/acsnano.9b02774

Lynch, I., Afantitis, A., Greco, D., Dusinska, M., Banares, M. A., and Melagraki, G. (2021). Editorial for the special issue from nanoinformatics to nanomaterials risk assessment and governance. *Nanomaterials* 11, 121. doi:10.3390/nano11010121

Maksov, A., Dyck, O., Wang, K., Xiao, K., Geohegan, D. B., Sumpter, B. G., et al. (2019). Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. *NPJ Comput. Mater.* 5, 12. doi:10.1038/s41524-019-0152-9

Maojo, V., Martin-Sanchez, F., Kulikowski, C., Rodriguez Paton, A., and Fritts, M. (2010). Nanoinformatics and DNA-based computing: catalyzing nanomedicine. *Pediatr. Res.* 67, 481–489. doi:10.1203/pdr.0b013e3181d6245e

Marrink, S. J., Corradi, V., Souza, P. C. T., Ingólfsson, H. I., Tieleman, D. P., and Sansom, M. S. P. (2019). Computational modeling of realistic cell membranes. *Chem. Rev.* 119, 6184–6226. doi:10.1021/acs.chemrev.8b00460

Nyström, A. M., and Fadeel, B. (2012). Safety assessment of nanomaterials: implications for nanomedicine. *J. Control. Release* 161 (2), 403–408. doi:10.1016/j.jconrel.2012.01.027

Oh, E., Liu, R., Nel, A., Gemill, K. B., Bilal, M., Cohen, Y., et al. (2016). Meta-analysis of cellular toxicity for cadmium containing quantum dots. *Nat. Nanotechnol.* 11 (5), 479–486. doi:10.1038/nnano.2015.338

Oksel, C., Ma, C. Y., Liu, J. J., Wilkins, T., and Wang, X. Z. (2015). (Q)SAR modelling of nanomaterial toxicity: a critical review. *Particuology* 21, 1–19. doi:10.1016/j.partic.2014.12.001

Papadiamantis, A. G., Jänes, J., Voyiatzis, E., Sikk, L., Burk, J., Burk, P., et al. (2020). Predicting cytotoxicity of metal oxide nanoparticles using Isalos Analytics Platform. *Nanomaterials* 10 (10), 2017. doi:10.3390/nano10102017

Plata, D. L., and Jankovic, N. Z. (2021). Achieving sustainable nanomaterial design though strategic cultivation of big data. *Nat. Nanotechnol.* 16, 612–614. doi:10.1038/s41565-021-00902-7

Poli, D., Mattei, G., Ucciferri, N., and Ahluwalia, A. (2020). An integrated *in vitro-in silico* approach for silver nanoparticle dosimetry in cell cultures. *Ann. Biomed. Eng.* 48 (4), 1271–1280. doi:10.1007/s10439-020-02449-5

Raies, A. B., and Bajic, V. B. (2016). *In silico* toxicology: computational methods for the prediction of chemical toxicity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 6 (2), 147–172. doi:10.1002/wcms.1240

Sayre, R. R., Wambaugh, J. F., and Grulke, C. M. (2020). Database of pharmacokinetic time-series data and parameters for environmental chemicals. *Sci. Data* 7 (1), 1–10. doi:10.1038/s41597-020-0455-1

Shatkin, J. A. (2020). The future in nanosafety. *Nano Lett.* 20, 1479–1480. doi:10.1021/acs.nanolett.0c00432

Singh, A. V., Maharjan, R. S., Kanase, A., Siewert, K., Rosenkranz, D., Singh, R., et al. (2021). Machine-learning-based approach to decode the influence of nanomaterial properties on their interaction with cells. *ACS Appl. Mater. Interfaces* 13 (1), 1943–1955. doi:10.1021/acsami.0c18470

Souza, M. R., Mazaro-Costa, R., and Rocha, T. L. (2021). Can nanomaterials induce reproductive toxicity in male mammals? A Historical and Critical Review. *Sci. Total Environ.* 769, 144354. doi:10.1016/j.scitotenv.2020.144354

Tang, K., Liu, X., Harper, S., Steevens, J. A., and Xu, R. (2013). NEIMiner: nanomaterial environmental impact data miner. *Int. J. Nanomed.* 8, 15–29. doi:10.2147/ijn.s40974

Tsuda, A., Donaghey, T. C., Konduru, N. V., Pyrgiotakis, G., Van Winkle, L. S., Zhang, Z., et al. (2019). Age-dependent translocation of gold nanoparticles across the air-blood barrier. *ACS Nano* 13, 10095–10102. doi:10.1021/acsnano.9b03019

Winkler, D. A. (2020). Role of artificial intelligence and machine learning in nanosafety. *Small* 16, 2001883. doi:10.1002/smll.202001883

Wu, Y., Duan, H., and Xi, H. (2020). Machine learning-driven insights into defects of zirconium metal-organic frameworks for enhanced ethaneethylene separation. *Chem. Mater.* 32 (7), 2986–2997. doi:10.1021/acs.chemmater.9b05322

Yan, X., Sedykh, A., Wang, W., Zhao, X., Yan, B., and Zhu, H. (2019). *In silico* profiling nanoparticles: predictive nanomodeling using universal nanodescriptors and various machine learning approaches. *Nanoscale* 11 (17), 8352–8362. doi:10.1039/c9nr00844f

Yu, F., Wei, C., Deng, P., Peng, T., and Hu, X. (2021). Deep Exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. *Sci. Adv.* 7, eabf4130. doi:10.1126/sciadv.abf4130

Zhang, X., Zhang, K., and Lee, Y. (2020). Machine learning enabled tailormade design of application-specific metal-organic frameworks. *ACS Appl. Mater. Interfaces* 12 (1), 734–743. doi:10.1021/acsami.9b17867