

Review Article

Inam Ur Rehman, Inam Ullah, Habib Khan, Mohammed Seghir Guellil, JaKeoung Koo*, Jakyoung Min, Shabana Habib, Muhammad Islam, and Mi Young Lee*

A comprehensive systematic literature review of ML in nanotechnology for sustainable development

<https://doi.org/10.1515/ntrev-2024-0069>

received March 17, 2024; accepted July 10, 2024

Abstract: The rapid expansion of nanotechnology has transformed numerous sectors, with nanoproducts now ubiquitous in everyday life, electronics, healthcare, and pharmaceuticals. Despite their widespread adoption, concerns persist regarding potential adverse effects, necessitating vigilant risk management. This systematic literature review advocates for leveraging artificial intelligence (AI) and machine learning (ML) methodologies to enhance simulations and refine safety assessments for nanomaterials (NMs). Through a comprehensive examination of the existing literature, this study seeks to explain the pivotal role of AI in boosting NMs sustainability efforts across six key research themes. It

* **Corresponding author: JaKeoung Koo**, School of Computing, Gachon University, 1342 Seongnam-daero, Seongnam-si 13120, Republic of Korea, e-mail: jakeoung@gachon.ac.kr

* **Corresponding author: Mi Young Lee**, Office of the Research, Chung-Ang University, Seoul, South Korea, e-mail: miylee@cau.ac.kr

Inam Ur Rehman: Masaryk Institute of Advanced Studies, Czech Technical University, Prague, Praha 160 00, Czechia; Department of CS & IT, University of Malakand, Chakdara 18800, KP, Pakistan, e-mail: inam.btk@gmail.com

Inam Ullah: Department of Computer Engineering, Gachon University, AI Building, Seongnam, 13120, Republic of Korea, e-mail: inam@gachon.ac.kr

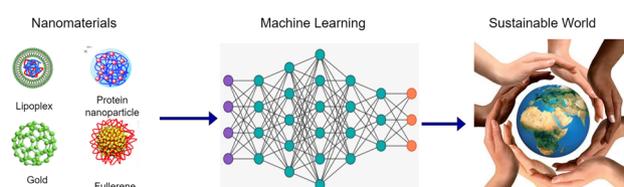
Habib Khan: School of Computing, Gachon University, 1342 Seongnam-daero, Seongnam-si 13120, South Korea, e-mail: habibkhan@ieee.org

Mohammed Seghir Guellil: Faculty of Economics, Business and Management Sciences, University of Mascara, MCLDL Laboratory, Mascara, 29000, Algeria, e-mail: m.guellil@univ-mascara.dz

Jakyoung Min: Department of Design Innovation, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul, 05006, Republic of Korea, e-mail: mj@sejong.ac.kr

Shabana Habib: Department of Information Technology, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia, e-mail: s.habibullah@qu.edu.sa

Muhammad Islam: Department of Electrical Engineering, College of Engineering, Qassim University, Buraydah, 52571, Saudi Arabia, e-mail: muha.khan@qu.edu.sa



Graphical abstract

explores their significance in advancing sustainability, hazard identification, and their diverse applications in this field. In addition, it evaluates the past sustainability strategies for NMs while proposing innovative avenues for future exploration. By conducting this comprehensive analysis, the research aims to illuminate the current landscape, identify challenges, and outline potential pathways for integrating AI and ML to promote sustainable practices within nanotechnology. Furthermore, it advocates for extending these technologies to monitor the real-world behaviour of NMs delivery. Through its thorough investigation, this systematic literature review endeavours to address current obstacles and pave the way for the safe and sustainable utilization of nanotechnology, thereby minimizing associated risks.

Keywords: nanomaterial, machine learning, artificial intelligence, nanosafety, nanomaterial sustainability

1 Introduction

Nanomaterials (NMs) are a significant breakthrough in global science and technology, typically ranging from 1 to 100 nm in size and exhibiting exceptional chemical, physical, and biological properties. These characteristics make NMs crucial across various fields [1]. According to MarketSandMarkets, the global market for NMs is expected to reach \$75.64 billion by 2025, with a strong annual growth rate of 13.2% from 2020 to 2025 [2]. NMs are integrated into everyday products such as cosmetics, sunscreen, food packaging, medicines, water filtration systems, and energy production. The development of nanomedicines and

nanotechnology has brought numerous benefits to humanity, a trend likely to continue. However, it is increasingly important to address the potential negative effects associated with their use. Currently, many diseases arise from regular exposure to harmful chemicals or materials, whose risks are often hidden by their perceived benefits [3]. NMs enter our lives through food additives, industrial processes, processed foods, packaging materials, cigarettes, cosmetics, medications, paints, and propellants. These substances can threaten human health, causing diseases like Parkinson's, cancer, Alzheimer's, asthma, bronchitis, emphysema, arrhythmia, urticaria, vasculitis, dermatitis, Crohn's disease, podocytosis, thrombosis, and hypertension [4].

To ensure safety and well-being, comprehensive nanotoxicological testing is crucial. The rapid growth of the nanoparticle (NP) market poses significant challenges for regulatory oversight and safety assessment using traditional *in vitro* and *in vivo* techniques. The increasing speed and volume of data make it difficult to evaluate chemicals with conventional methods, particularly given the numerous chemical toxicological endpoints [5]. Therefore, a shift towards computational modelling methods is needed to assess and predict the risks associated with these particles. Such approaches can expedite assessments, save resources, and help develop a framework for better health outcomes. The growing importance, complexity, and versatility of machine learning (ML)-based information techniques have brought them to the forefront of various fields [6–10]. However, acquiring substantial data for these techniques is increasingly challenging [7,10]. Typically, educational data devices are used to create broadly applicable patterns but often have lower predictive accuracy in specific scenarios [11]. This trend has significantly impacted nanotechnologies and NMs, and this influence is expected to continue [10]. High-throughput NM manufacturing and characterization have received relatively less attention despite their potential for thorough analysis using inspection methods [10]. Future advancements in high-throughput synthesis, characterization, and ML modelling are anticipated to significantly enhance the chemical industry workforce [12].

It is essential to monitor recent advancements in ML reported by scientists dedicated to developing safer NMs [10]. These advancements should be adopted by those following in their footsteps [10]. Fuxhi *et al.* [13] demonstrate a growing preference for nonlinear modelling techniques, moving away from the prevalent use of linear regression [14]. While these methods often confirm quantitative relationships between shape and properties, there is a noticeable shift towards understanding nanospecific capabilities through physicochemical interpretations rather than purely theoretical descriptors [10].

Despite ongoing debates and the lack of widespread consensus on statistical preprocessing strategies, regulatory

guidelines and validation strategies for modelling sets are underused, despite their common application in quantitative structure–activity relationship (QSAR) studies [10]. Templates for reporting versions in regulatory risk analyses of engineered NMs are suggested to provide well-defined and organized model descriptions [14]. These templates cover QSAR model reporting and are relevant to physiologically based kinetic (PBK) and environmental exposure models for NMs, offering a comprehensive overview of the computational model landscape for hazard assessment [10]. The absence of effective analytical methods complicates the verification of models and the assessment of their application and identity in risk evaluations [10]. Recent research from the EU has explored algorithms for measuring the residence times of engineered NMs, providing guidance in the REACH regulation [14].

Various sources have extensively discussed trends in partial and widespread automation relevant to nanotechnology, highlighting the automation of nanoscience, investigation of additives, and the development of self-contained experimental devices [10]. Techniques like pulsed laser deposition, chemical vapour deposition, biomolecular templating, and electrodeposition have been utilized for the automated or semi-automated synthesis of inorganic NMs [15]. Notably, the application of flow chemistry showcases the potential for automating the generation of assemblies for colloidal NP solutions, challenging previously held beliefs [16]. Interestingly, the article discussing NPs designed for medicinal purposes intentionally excludes references to automation, informatics, or ML modelling. The use of ML to assess potential negative consequences of NM datasets, following a “safe by design” approach, is gaining momentum, particularly with the collaboration of NM regulators [17]. Similarly, ML is used to model the toxicological properties of NMs by leveraging omics data. Recent reviews have highlighted the potential of combining integrative omics with ML to profile NMs and model their biological impacts for safety and risk assessments [18–20]. Over the past decade, various computational models and ML techniques have been developed to predict the toxicological properties and adverse effects of NPs [21]. Among these, QSARs and quantitative structure–property relationships (QSPRs) are the most commonly used methods [22]. Recently, the European Commission has launched several projects to explore the potential of modelling NP toxicity and properties [23,24].

The novelty of this article lies in its thorough systematic literature review (SLR) review of existing research on using ML and artificial intelligence (AI) to improve the sustainability and safety of NMs (NMs). Unlike other reviews, this study covers almost all relevant literature, showcasing the significant progress and potential of ML and AI in making NMs more sustainable. It also tackles interdisciplinary challenges and suggests new research directions, highlighting the

role of these technologies in creating safer and more sustainable nanotechnology applications. In addition, the article offers detailed recommendations for future research at the intersection of nanotechnology and ML.

Up to the author's knowledge, the present study possesses unique characteristics that distinguish it from earlier surveys:

- 1) **Comprehensive SLR:** Unlike previous studies, this article conducts an extensive SLR, covering nearly all existing literature on the use of AI and ML for evaluating the sustainability of NMs.
- 2) **Rigorous methodology:** It employs a predefined method for searching, selecting, and extracting data from articles, which reduces bias and ensures a thorough and systematic review process.
- 3) **In-depth analysis:** The article offers a detailed examination of the significance and application of AI and ML in enhancing the sustainability of NMs, highlighting key advancements and the potential benefits of these technologies.
- 4) **Interdisciplinary challenges and future research:** It addresses interdisciplinary challenges and suggests innovative future research directions, underscoring the essential role of AI and ML in fostering safer and more sustainable nanotechnology practices.
- 5) **Detailed recommendations:** The article provides specific recommendations for future research at the intersection of nanotechnology and ML, offering guidance for upcoming studies and progress in this field.

1.1 Research goals

This research aims to comprehensively investigate and analyze existing studies concerning the application of AI and ML in the realm of sustainability for NMs. This study is centred around six distinct research questions.

RQ1: What is the importance of AI and ML techniques in NM sustainability?

RQ2: What is the role of AI and ML tools in the identification of hazards associated with NMs?

RQ3: What are the applications of AI and ML in NM sustainability?

RQ4: What are the past NM sustainability roadmaps?

RQ5: What are the future NM sustainability roadmaps?

RQ6: What are the recommendations for future research in the intersection of nanotechnology and ML?

1.2 Contribution and design

In this SLR, we aimed to contribute to ongoing research on the sustainability of NMs and to enhance understanding of

their applications across various scientific domains. Our key findings are as follows:

- 1) We conducted an SLR focusing on the sustainability of NMs, where we pinpointed 180 primary research articles from a pool of 395 pertinent studies.
- 2) We highlighted the significance of AI and ML techniques in NM sustainability.
- 3) We outlined the applications of AI and ML in enhancing NM sustainability.
- 4) We discussed both past and future roadmaps for NM sustainability.
- 5) We presented a detailed examination of the role of ML in assessment processes related to NM sustainability.
- 6) We provided specific recommendations for future research in the intersection of nanotechnology and ML.

Table 1 presents a comparison between our systematic literature review (SLR) and existing relevant surveys, considering four key criteria: methodology, number of database sources searched, time interval, and study focus. As shown in the table, our study is unique as it is the first to conduct a comprehensive SLR specifically focused on the integration of AI and ML in the sustainability and safety assessment of NMs. This highlights the novelty and thoroughness of our research approach, which stands out from previous surveys that either did not follow a systematic methodology, searched a limited number of databases, covered shorter time spans, or addressed broader, less specific topics.

This research aims to aid scholars in expanding their knowledge of NM sustainability. Moreover, by exploring the use of advanced technologies like AI and ML in this context and thoroughly assessing their pros and cons, we hope to lay the groundwork for further advancements in the field.

The article is organized as follows: Section 2 describes the research methodology. Section 3.3 presents the results of the study, Section 4 discusses future research recommendations, and finally, Section 5 presents the conclusion.

2 Research methodology

The research methodology employed in this study is the SLR, which is a well-established and commonly used approach [29]. This method involves systematically analyzing existing literature by following a predetermined set of steps to identify, assess, and interpret relevant research related to the research questions [29,30]. Figure 1 illustrates the entire process of the SLR.

Table 1: Comparison of the proposed SLR with existing reviews

Reference	Title	Year	Methodology	Databases searched	Time interval	Study focus
[25]	A ML examination of NM safety	2020	Ordinary review	ScienceDirect, Web of Science, Google Scholar and PubMed	2010–2019	ML methodologies used to predict toxicological outcomes
[22]	Practices and trends of ML application in nanotoxicology	2020	Ordinary review	Google Scholar, Elsevier (Scopus and ScienceDirect), Web of science and PubMed	2013–2022	ML application in nanotoxicology
[11]	Role of AI and ML in nanosafety	2020	Ordinary review	—	—	Roadblocks to the application of ML to predict potentially adverse effects of NMs
[26]	Nanotechnology for a sustainable future: addressing global challenges with the international network4sustainable nanotechnology	2021	Ordinary review	—	—	Global challenges to sustainable nanotechnology
[27]	Overcoming roadblocks in computational roadmaps to the future	2021	Ordinary review	—	2000–2020	Roadmaps for the future of safe nanotechnology
[19]	ML-integrated omics for the risk and safety assessment of NMs	2021	Ordinary review	—	—	ML in risk and safety assessment of NMs
[10]	ML algorithms are applied in NM properties for nanosecurity	2022	Ordinary review	—	—	Roadblocks to ML in nanosafety
[28]	Using ML to make NMs sustainable	2023	Ordinary review	—	2001–2022	ML techniques for environmental risk assessment to ensure sustainability
Our proposed SLR	A comprehensive systematic literature review of ML in nanotechnology for sustainable development	2024	Systematic literature review	Springer, IEEE Xplore, ACM, Science Direct, Wiley, and Google Scholar	2013–2023	Significance of ML in advancing sustainability, hazard identification, and their diverse applications in NMs, evaluates past sustainability strategies for NMs while proposing innovative avenues for future exploration, and recommendations for future research

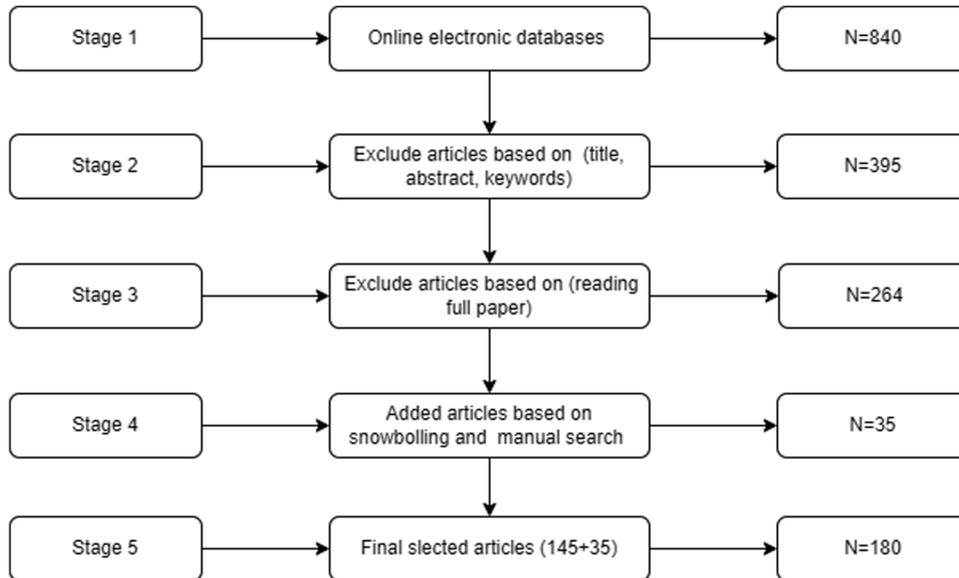


Figure 1: The entire process of selecting articles for the SLR.

2.1 Data analysis tools and methods

To ensure a robust and reproducible analysis of the collected data, a variety of statistical tools and methods were employed. The approach involved several key steps:

2.1.1 Software utilized

Microsoft access: This software was used for database management, enabling efficient storage, retrieval, and manipulation of large datasets. Custom queries were created to filter and aggregate data as needed.

draw.io: This tool was employed to create detailed diagrams and flowcharts, which helped visualize data relationships, processes, and analytical workflows. This enhanced the clarity and understanding of the analysis steps.

2.1.2 Visualization

Data visualizations were generated using draw.io, to highlight key findings and trends. draw.io was specifically utilized to design flowcharts and diagrams that mapped out the data analysis process, providing a clear visual representation of the steps and methodologies used.

By integrating these statistical tools and methods, the analysis aimed to extract meaningful insights from the data, ensuring the results were accurate, reliable, and

reproducible. The use of Microsoft Access for effective data management and draw.io for clear visualization of processes further enhanced the transparency and replicability of the study.

2.2 Searching article

Conducting searches in online repositories stands as an essential phase in conducting an SLR. It involves the initial formulation of a search string, which was made based on the recommendations provided for the SLR process [29]. This procedure entailed generating a search query involving key terms and their appropriate alternatives, applying different Boolean operators. Below is the whole search query used to choose the article.

((NM sustainability OR NM safety OR nanosafety) AND (artificial intelligence OR AI) AND (machine learning OR ML))

To locate relevant articles based on the stated search criteria, we performed searches on six distinct online digital databases: Springer, IEEE Xplore, ACM, Science Direct, Wiley, and Google Scholar (Table 2). These widely recognized online repositories are expected to provide extensive literature coverage on the subject matter. In addition, we employed manual searching and the snowballing technique, as described by Alzoubi *et al.* [30], to uncover additional pertinent articles. By using this search approach, we located a sum of 840 articles within the stated online repositories.

Table 2: Results of the articles search

Databases	First filtration	Second filtration	Third filtration	Final selection	Selected articles (%)
IEEE Xplore	28	21	5	1	0.56
Google Scholar	184	48	29	42	23.33
Science Direct	175	91	78	45	25.00
ACM	138	89	45	16	8.89
Springer	149	52	35	14	7.78
Wiley	166	94	72	27	15.00
Snowballing			0	25	13.89
Manual search				10	5.56
Total	840	395	264	180	100

2.3 Criteria for article selection

Various filters were used to select pertinent research articles (Table 3). Initially, a set of predetermined search terms was used to query particular online digital databases for research articles, leading to the discovery of 840 articles across six databases. Afterwards, during the second stage, an in-depth review of these articles took place focusing on the significance of the article title, abstract, and keywords. A total of 185 articles were picked in this stage, with others being excluded. Finally, in the third stage, complete articles were read, following the specific inclusion and exclusion conditions:

- 1) The research articles related to the sustainability of NMs.
- 2) The research articles were relevant to AI and ML techniques.
- 3) We included research articles that underwent peer review and were written in English.
- 4) We incorporated the latest editions of research studies.
- 5) Duplicate articles were excluded from the analysis.

We selected 145 most relevant articles at the current stage. Subsequently, in the fourth stage, thirty-five articles were identified through manual and snowballing methods. In conclusion, a complete set of 180 relevant articles was finally selected for inclusion in this study (Figure 2). The search processes were performed on Oct 25, 2023, and all relevant research up to that date was incorporated.

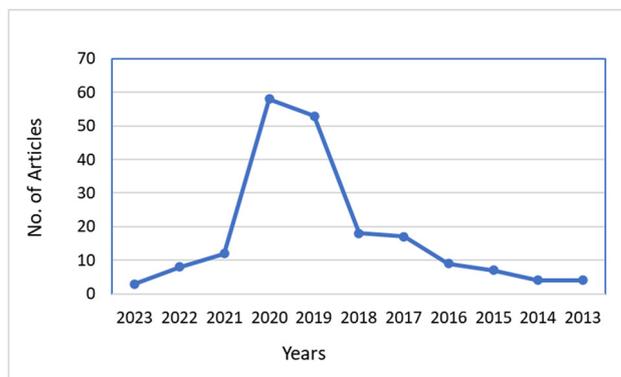
Table 3: Stages of article filtering

Filtration	Technique	Criteria for evaluation
Filtration 1	Finding pertinent research articles from an online database	Keywords
Filtration 2	Excluding research solely relying on the title, abstract, and keywords	Title, abstract, and keywords
Filtration 3	Including papers based on their complete content	Full paper
Final filtration	Final chosen articles	Final selected papers

3 Results

3.1 RQ1: What is the importance of AI and ML techniques in NM sustainability?

The process of exploring the biological characteristics of NMs involves several steps. First, the materials are synthesized and then subjected to relevant biological assays, generating a dataset used to train ML models [11]. Physicochemical attributes of NMs are transformed into mathematical descriptors, which are selected features from a pool and used in ML methods to build predictive models for desired properties [31]. These models are validated by predicting

**Figure 2:** Number of articles reference to each year.

results for separate materials not involved in the model-building process.

In the realm of AI, AI encompasses tasks demonstrating human intellectual characteristics, while ML, as a subset of AI, accesses data, identifies patterns, and generates insights [11,32,33]. ML’s appeal lies in its versatility and platform technology functionality. Unlike traditional software, ML methods offer a more flexible approach, where the same algorithm and software program can be applied in various modelling situations, covering different materials and outcomes [34,35]. ML algorithms learn patterns from data similarly to human learning but with faster speed and an enhanced capacity to manage data with higher dimensions [36].

3.1.1 Traditional ML methods

Traditional ML methods encompass a variety of techniques, including regression, Bayesian networks, artificial neural networks (ANNs), genetic algorithms, decision trees, and support vector machines (SVMs) [37,38]. These techniques have been crucial in generating valuable models elucidating the biological properties of NMs, as extensively documented in the literature. While specific examples are discussed below, detailed explanations of these methods are omitted here due to their widespread usage and familiarity. For a thorough understanding, recent reviews and references specific to each method are recommended. Most computational models in the literature, illuminating the

correlation between NM properties and biological endpoints, have predominantly utilized traditional statistical techniques. Particularly, conventional ML techniques with simple architectures, have been prominently featured in these models [39]. However, ML models [35] often lack awareness of complex environmental factors that can limit their operational effectiveness. With the increasing prominence of ANNs in broader scientific and technological fields, there is renewed interest in their application across various domains, including nanosafety and drug discovery [40]. This resurgence underscores a heightened recognition of the potential and versatility of ANNs, representing a notable evolution in their application within specialized domains.

3.1.2 Deep learning (DL) techniques

DL methods involve neural networks with multiple hidden layers and complex architectures. They have significantly impacted various fields of science and technology, particularly excelling in recognizing features in speech, images, and decision-making processes [41]. Unlike earlier simplistic approaches, contemporary DL methods have the advantage of automatically producing valuable descriptors during modelling. Researchers suggested in DL studies that deep and attention-based DL networks can achieve excellent performances [42–44]. Common DL-driven approaches are based on convolutional neural networks (CNNs) [45,46], recurrent neural networks (RNNs) [47], transformer models [36], generative adversarial networks (GANs) [48], and

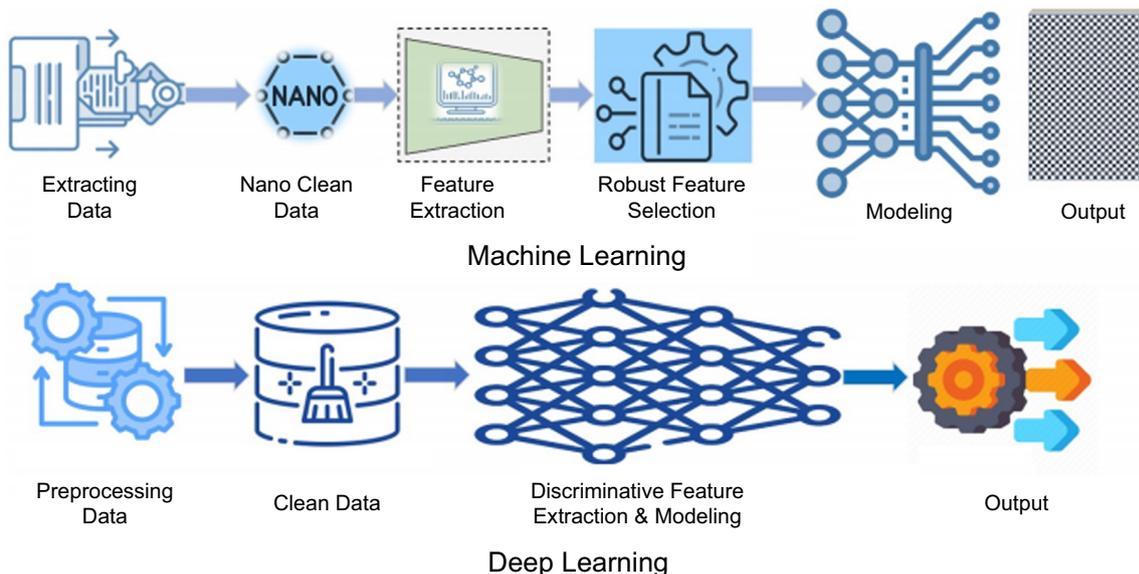


Figure 3: The traditional ML techniques need feature engineering (top). The DL techniques automatically perform feature extraction and analysis (bottom).

autoencoders [49]. CNNs, a type of supervised ML, are adept at identifying image features based on spatial correlations, focusing on local relationships in data and exhibiting resilience to minor alterations [12] (Figure 3).

A cascade of deep neural networks (DNNs) comprises two distinct networks: the first matches each material with an outcome, while the second links the outcome to one or more materials. This configuration, often referred to as an autoencoder, serves to both reduce dataset dimensionality and predict materials with specific properties based on the trained ML model. Generative adversarial networks GANs address the inverse mapping problem in materials design. GANs consist of a generator that produces trial structure–property models and a discriminator that evaluates the quality of these models against available unlabelled data. Originally designed to generate structures autonomously, GANs have proven effective in this capacity [12]. Similarly, active learning employs ML to select experiments efficiently to achieve predetermined goals, akin to directed evolution processes involving sequential generations of contestant structures undergoing selection, modification, and testing [12].

3.2 RQ2: What is the role of AI and ML tools in the identification of hazards associated with NMs?

ML techniques have become increasingly important for identifying hazards linked to NMs, with a strong emphasis on methods enabling thorough analysis, especially for managing large datasets and detailed content. This includes studying coronas of the protein around NMs when they interact with bodily liquids. ML shows significant promise in addressing read-across challenges, particularly in swift *in vitro* studies related to material properties, as indicated by a notable increase in attention (Table 4). The growing emphasis on computationally intensive network analysis further underscores its importance [50–53].

3.2.1 *In vitro* approaches

In vitro systems have traditionally been employed primarily in human health studies, but there is a growing trend in their application in environmental research is also dominant [28]. ML can optimize the process of learning new concentration-response curves through the use of nonlinear methodologies, potentially enhancing efficiency and accuracy in this aspect of research. Moreover, in hazard assessment,

Table 4: Application of ML and AI in hazard identification

Reference	Year	Key info	ML Tools	Applications
[50,52,53]	2019, 2019, 2020	Hazard identification	High-throughput, network analysis	NMs hazard identification, protein corona analysis
[89]	2020	Detecting malformations	Random forest	Predicting NM toxicity in zebrafish embryos
[90]	2019	Prediction of toxicity	ML graph theory	Acute ecotoxicity prediction
[91]	2021	ML correlation of NM parameters with <i>in vivo</i> toxicity in soil organisms	ANN-type recurrent self-organizing maps	Analyzing earthworm movements
[92]	2020	ML-based graph theory for acute ecotoxicity prediction	ML bioaccumulation models	Studying bioaccumulation in conventional chemicals
[118]	2020	ML method for analyzing earthworm movements	Random forests	Predicting HC50 values for conventional chemicals
[94]	2019	ML models for bioaccumulation and trophic transfer in conventional chemicals	Evolutionary/genetic algorithms	Improving species distribution models
[95]	2020	ML techniques improving environmental species distribution models	ML media effects understanding	Understanding media effects on nanotoxicity
[96]	2019	ML understanding of media effects on nanotoxicity	ML contaminant identification	Identifying contaminants in sewage sludge
[97]	2020	ML methods for identifying contaminants in sewage sludge		
[140]	2011			

test platforms are often selected with the aid of integrated approaches to testing and assessment (IATA) or intelligent testing strategies (ITS) [54]. However, one notable limitation of ITS and IATA is their independence from the surrounding system, overlooking potentially crucial additional factors. The continuous evolution of sophisticated models suggests that future approaches may involve integrative ML multitasking methods. These methods could predict multiple toxicities simultaneously, mitigating the overfitting and employing techniques such as network analysis or forest trees [55,56]. Understanding mechanisms of action is crucial for comprehending hazards and facilitating risk evaluation. Omics data, including transcriptomics, metabolomics, and proteomics, serve this purpose but pose challenges due to their extensive variables [57,58]. ML offers an attractive approach to address the complexity and difficulty in identifying causal relationships within omics datasets [59–62]. While ML techniques find extensive application in omics, their application to materials and environmental risk is limited. Clustering techniques in omics analysis, while effective in finding patterns, lack explanatory power for mechanisms. Refined techniques like model-based approaches, especially when structured with known classes, provide more information on potential mechanisms [59,63]. The challenge lies in identifying causal relationships between omics expression levels and phenotypic outcomes, which ML techniques, mechanism-based or employing logic networks, can address, led by training data for causality verification [64–68]. The use of ML should align either with a highly mechanistic approach for identifying mechanisms or as a means to verify consistency with existing mechanistic understanding. In another method, ML approaches are utilized as supplementary investigative instruments rather than definitive identifiers [69–73]. Special attention must be given to the input data in dealing with complex omics data, as biases in input data can propagate through automated algorithms, impacting the final output and posing challenges when employing ML methods. This challenge is not limited to ML techniques alone; rather, it is crucial to confront them to guarantee impartial and dependable outcomes. In addition to omics approaches dedicated to the inner workings of cells, various other omics methodologies play an important role in environmental research, such as genetic barcoding for species diversity analysis, provide ideal datasets for ML applications [74–76]. In summary, environmental omics generates extensive data necessitating automated analysis to consistently identify underlying mechanisms of action.

In environmental studies, particularly those involving (smart-)NMs, *in vitro* techniques are gaining prominence [77]. ML approaches effectively enhance the usefulness of *in vitro* techniques by swiftly producing large datasets

using high-content and high-throughput methods. This capability allows for the concurrent investigation of multiple materials under comparable circumstances [78,79]. ML approaches such as NNs and RF algorithms can be used with the obtained datasets to find similarities between materials by detecting patterns [80–82].

ML techniques for image analysis are also applicable to these datasets, enabling the quantification of processes during experiments, such as agglomeration, which significantly influences exposure [83–86]. However, ML image evaluations require a large and well-labelled dataset for training. Furthermore, *in vitro* studies can be used to determine the makeup of NMs protein coronas (PCs) [87]. For instance, Findlay *et al.* [88] employed existing data on bacteria and yeast to assess the effectiveness of different ML methods in foreseeing the formation of PC on basic Ag NPs, and how it correlates with the NPs' chemical and physical attributes. Their research revealed that ML approaches allow the prediction of PC composition through conventional experimental data and readily available details about the biophysical behaviours of proteins. In summary, ML plays a vital role in uncovering attributes in *in vitro* analysis, and automation can contribute to enhancing repeatability.

3.2.2 *In vivo* approaches

In environmental hazard identification, initial tests typically involve *in vivo* experiments, which are considered a key aspect in many testing guidelines [28]. While ML techniques have been widely used to assess the toxicity of traditional chemicals and address environmental concerns, their application to smart NMs has been limited. Notable studies in this area include Karatzas *et al.* [89], who used an ML approach to automatically detect abnormalities and classify affected *Daphnia* organisms after exposure to silver and titanium NMs. To *et al.* [90] developed an ML technique using RF for predicting the harmful effects of metal-containing NMs on zebrafish embryos based on their physical and chemical properties. Similarly, Gomes *et al.* [91] employed ML to establish a link between various parameters of NMs and the toxicity experienced by soil organisms when exposed to different types of titanium dioxide (TiO₂) materials infused with iron (Fe).

A range of ML methods is used to forecast the potential toxicity of traditional chemicals within living organisms. For instance, ML-driven graph theory has been employed to anticipate severe ecotoxicity, drawing insights from datasets such as AIST-MeRAM and ECTOX for estimation and assessment purposes [92]. In a different context, Ji *et al.* [93] implemented an ANN-based recurrent self-

organizing map to examine the locomotion patterns of earthworms after their exposure to copper sulphate (CuSO_4). This method holds significance in assessing the behavioural responses of worms following exposure to various copper-based (smart-)NMs. Trophic transfer and bioaccumulation, essential endpoints in environmental risk assessment, have been studied using linear models and ML techniques for traditional chemicals in invertebrates and fish [94]. Although ML has not been extensively applied to these endpoints in NMs, its potential is evident. In addition, ML models like random forests have been successful in predicting hazard concentration levels (HC50) for traditional chemical materials, suggesting the possibility of extending these models to NMs [95].

Environmental species distribution models, crucial for identifying potential effects, can be enhanced using ML techniques, as demonstrated by studies using evolutionary/genetic algorithms [96]. ML can also aid in understanding the effects of nanotoxicity, drawing insights from studies on conventional chemicals [97]. A significant source of NM emissions called “Sewage sludge,” could benefit from ML applications in identifying important contaminants and understanding their impact on soil and crops [98]. While previous studies focused on conventional chemicals, similar ML methods could be applied to smart NMs. Beyond direct applications to current environmental risk assessment practices, ML shows promise in handling larger datasets and integrating more complex tests [28]. Examples include using ML in animal species classification, predicting avian field biodiversity, and using DL for continuous camera monitoring to identify invertebrate species [28]. It is widely applicable in real-time analysis like in surveillance applications [99]. They suggest that ML can facilitate the combination of diverse datasets and predict NM toxicity under various scenarios, highlighting its relevance in hazard assessment and risk management [28].

3.2.3 Exposure analysis

Critical components of analyzing the fate and exposure of substances include evaluating environmental releases, forecasting substance levels in the environment, and understanding accumulation in organisms, which can be done at various scales [28]. The soil compartment is highlighted as the main sink for (smart-)NMs [28]. Various statistical and logical models have been utilized for predicting environmental concentrations, often requiring intensive data calculations [28]. Studies like Wigger *et al.* [100] have shown that combining various modelling techniques with ML can significantly improve the effectiveness of these models. Wikle employed ML, particularly NNs and deep hierarchical

models, to predict changes in soil moisture in large geographic areas [101].

Moreover, ML has been utilized to a restricted degree in simulating the destiny and potential exposure of NMs [28]. Notably, Goldberg *et al.* [102] pioneered the use of ML to develop advanced fate models for NMs. Their work demonstrated that decision trees, including RF regression and classification, proved effective in identifying previously concealed key descriptors crucial for huge models. The authors highlighted the advantage of ML approaches, emphasizing their relative lack of restrictions due to the absence of prior linearity assumptions. However, a significant hurdle reported by the authors was the scarcity of data, encompassing both quantity and quality concerns. Subsequent studies, such as that by Babakhani *et al.* [103], leveraged ANN to investigate the transport of NMs in porous media. The study revealed that the ANN-based correlation technique outperformed existing empirical approaches in forecasting continuum model attributes and breakthrough models. These approaches hold promise in supporting other NM-specific fate models like Simplebox. Developing ML-based models specific to NM species becomes crucial, particularly because (smart-)NMs deviate from the fate and exposure models applied to traditional chemicals. Unlike chemicals assuming a steady-state condition, NMs exhibit characteristics like agglomeration/aggregation influenced by surface properties [104].

Fazeli-Sangani *et al.* [105] examined the effect of various features of soil and experimental variables on the movement of TiO_2 in actual soil environments using data analysis techniques. Several methods such as multiple linear regression, random forest, classification and regression trees, and ANN were utilized for this purpose. Their results showed that ANN outperformed RF in predicting TiO_2 transport, despite RF needing less input data. MLR techniques generally performed poorly, and CART had good calibration but poor validation. In addition, various researchers have used ML to predict environmental parameters important for fate and exposure modelling, especially in studies on soil carbon and nitrogen distributions. Pinheiro *et al.* [106] utilized ML, specifically regression trees, to predict soil texture through digital soil mapping, with applications in remote sensing data. These environmental ML techniques are crucial for developing local and regional predicted environmental concentration models, essential for environmental risk assessment. In addition, ML has potential to predict exposure and toxicity throughout a material's life cycle, though this application remains underexplored. Furthermore, since organisms in nature are mostly exposed to mixtures of chemicals and materials, mixture modelling is another vital area in risk assessment where

computational methods, including ML, could provide valuable insights.

3.2.4 Risk classification

During the risk assessment phase of an environmental risk assessment, the primary objective is to systematically amalgamate data related to material features, exposure, and hazards to establish a robust environmental evaluation [28]. Traditionally, this process employs tools featuring decision points based on material description, exposure, utilizing fate, and hazard data, often incorporating IATA [28]. Consequently, decisions are required regarding data selection and integration, frequently employing simple decision trees with predefined structures and binary decisions. The risk measuring paradigm aligns with ML approaches; however, ML presents the possibility of employing more intricate decision tree structures to substitute traditional IATAs [28]. In other domains, ML-based methodologies have been explored for risk assessments, such as the evaluation of flooding risks [107] and soil quality based on polycyclic aromatic hydrocarbon (PAH) pollution [108].

As highlighted earlier, the pursuit of automation is a crucial objective in risk assessment, driven by the need for efficiency and cognitive relief for human risk assessors facing time constraints and cognitive overload due to numerous assignments requiring critical decisions without sufficient time for exhaustive data review [109–111]. In a real risk assessment where the combination of exposure, fate, and hazard data significantly influences the outcome, the choice of data and data treatment techniques, such as ranking, logical choice, ranking, or statistical techniques, play a crucial role [112]. Some ML-based methods opt for predictive approaches, sidestepping conventional explaining statistics [112]. The potential incorporation of various data and possibly new methods implies that upcoming risk assessment experts and scientists will need to navigate and select from diverse analytical frameworks, leading to varying approaches in evaluating statistical significance and importance [113–116]. This holds relevance in both the middle and ultimate phases of a risk assessment [117]. Considering the likelihood of future risk assessment methods incorporating network analysis [118], it becomes clear that risk assessors may face increasingly complex situations. In these cases, despite using advanced DL techniques, it may be not easy for anyone to fully understand the entire data processing and its outcomes, effectively making the assessment a black-box system. This challenge, familiar to those assessing conventional chemicals, is expected to intensify with the added complexity of smart-NMs [28]. This issue is particularly

evident when data-intensive models, such as most ML/AI models, are applied to smaller datasets, even with validated background models [28].

However, by using appropriate data and a well-designed model-flow, new automated models are expected to deliver faster and more standardized calculations, enabling more comprehensive risk analyses than previously possible [119–121]. These models can, for example, connect exposure and toxicity parameters to global weather patterns, sea currents, soil–water flow models, and known biodiversity distributions using local/global models [122,123]. This advancement could enable the rapid integration of additional data, allowing chemometric, biometric, and risk data to be collaboratively processed with ML techniques [124]. Employing various ML techniques, such as symbolic regression, may also uncover additional biological or ecological “laws” [125].

Nonetheless, it is important to note that ML methods may encounter reproducibility challenges, especially when handling highly complex data and models [28]. A fundamental concern underlying environmental risk assessment involves the uncertainty and sensitivity associated with the outcome, represented as a value with related statistics [28]. Uncertainty arises from various sources, such as the lack of adequate input data or the selection of potentially “incorrect” input data, hindering the models from effectively addressing the risk [126]. Incorporating M does not alleviate this problem; in fact, it may further obscure the issue, particularly when using hidden layers, making it less certain which data influences the outcome. ML is generally acknowledged as data-hungry, emphasizing the heightened importance of acquiring sufficient data of high quality [28]. Sensitivity, a crucial aspect of ERA concerning how well it can capture a true effect, is also relevant to ML techniques, encompassing features like specificity and accuracy, integral to predictiveness training through training sets [28]. However, determining how an ML model is derived, especially when using deep or reinforcement learning methods involving temporary virtual nodes, can be challenging [28]. This issue can be partially addressed by employing robust adversarial controls [127]. It is also important to recognize that correlations are often misinterpreted as causality in traditional data analysis.

In the context of risk management, Hino *et al.* [128] presented that ML techniques can enhance the effectiveness of public agencies in meeting regulatory objectives. They trained a regression forest model to predict which facilities were at risk of failing inspections and used the model’s risk scores to propose alternative allocations of Clean Water Act inspections. These alternative strategies nearly doubled the detection rate of violations during inspections, with an increase of over 600% under the

most aggressive reallocation scheme. In addition, Duan *et al.* [129] applied ML within a multimedia system to compare traditional environmental chemistry with ML methodologies. Forest *et al.* [130] highlighted the importance of selecting the appropriate endpoint in ML systems for assessing human health hazards. These studies underscore the potential for ML-based techniques from a regulatory perspective. A widely discussed topic in nano environmental health and safety (nano EHS) revolves around the potential of current ML techniques, such as read-across, classification, or ranking tools, to facilitate the grouping of materials that would otherwise be challenging to classify [111,131]. As previously mentioned, these techniques also hold promise for identifying new materials with specific properties [132,133].

ML tools must be designed to seamlessly interface with the Internet of Things (IoT) to be effectively applicable in various contexts [134]. The integration of IoT with ML applications in environmental settings [135] has the potential to facilitate comprehensive online risk management, such as precision agriculture and farming practices utilizing drone technologies [136]. This encompasses tasks like automatic regulation and enhancement of production sites, machinery, agricultural robots, and irrigation systems for the implementation of water quality measures, soil moisture control [137], and promotion of behavioural changes [138]. Nevertheless, the extensive capabilities afforded by such integration pose significant legal and ethical challenges [139] and carry the risk of potential large-scale misuse, such as through hacking. The integration of IoT remains a crucial consideration in contemporary ML implementations, enabling ML systems confined in virtual spaces to directly influence the physical world on a significant scale without human intervention [28].

3.3 RQ3: What are the applications of AI and ML in NM sustainability?

NMs are ubiquitous in modern life, present in food additives, industrial processes, cosmetics, medications, and more, with the potential to cause a range of health issues from Parkinson's disease to cancer (Figure 4) [2,4,141]. However, evaluating their safety through conventional methods is challenging due to the rapid proliferation of NPs and the complexity of the data involved. To address this, a shift towards computational modelling methods is crucial, offering the promise of faster assessments, resource conservation, and the establishment of a unified framework for promoting better health outcomes.

Numerous reviews on the use of ML techniques in nanotoxicology were conducted over the past decade, as evidenced by various sources [142]. We present key examples of published research to illustrate the diverse ML approaches utilized for modelling NM properties across various applications (Table 5). Our selection aims to showcase the breadth of ML applications in predicting NM hazards and outcomes in “safe-by-design” projects, emphasizing technologies that hold promise in achieving predefined milestones.

Puzyn *et al.* [143] demonstrated the utilization of ML and statistical techniques to forecast the adverse characteristics of NMs. They developed a simple linear regression model with a single parameter to forecast the cytotoxicity of 17 distinct metal oxide NPs against *Escherichia coli*. The model used descriptors acquired from quantum chemical calculations. Similarly, Epa *et al.* [144] employed Bayesian regularized neural networks and linear regression for the prediction of biological effects associated with 51 metal oxide NPs including different 109 metal oxide NPs and metal cores having distinct cores but differing surface modifiers. These models assisted numerical forecasts concerning the apoptosis of flat muscle cells and the uptake of NP concentration by human pancreatic cancer cells and umbilical vein epithelial cells during *in vitro* experiments. They exhibited predictive accuracy in a separate test batch of NMs, with an error of <3 for apoptosis and within a twofold range for cellular uptake in the specified cell lines. Fourches *et al.* [145] utilized SVM classification and k-nearest neighbours (kNNs) regression to characterize the identical dataset. Their models proved an external prediction accuracy of up to 73% on classification while attaining an r^2 value of 0.72 for regression models predicting corresponding properties. Liu *et al.* [146] presented ML techniques to assess the impacts of 44 iron oxide core NPs, frequently used in molecular imaging, across four different cell types, comprising aortic endothelial cells, vascular smooth muscle cells, macrophages/monocytes, and hepatocytes. Self-organizing map-based clustering was applied to characterize types by employing four different biological analyses and adjusting NP concentrations. The classification involved the application of a logistic regression, Bayesian classifier, and NN classifiers. These classifiers were trained based on descriptors such as size, spin-spin, zeta potential, and spin-lattice relaxivity. Remarkably, the two class models showed comparatively best accuracies, surpassing 78%.

Papa *et al.* [147] analyzed the ZnO and TiO₂ NPs' cytotoxicity, utilizing experimental descriptors and testing different concentrations to assess their impact on disrupting lipid membranes in cells, measured through lactate dehydrogenase (LDH) levels. They developed models employing

SVMs, linear regression, and neural networks. These models exhibited accurate predictions of LDH levels in the testing while errors ranged from 8 to 17% compared to control cells. Notably, the nonlinear ML techniques surpassed the performance of the near-regression models. To overcome the limitation of available toxicity data, Chen *et al.* [148] used ML techniques for the classification of NM ecotoxicity. They applied read-across attributes calculated from various sources of ecotoxicity data to develop models tailored to specific species and a joined model capable of predicting toxicity across various species. Employing various tree algorithms, including decision tree, functional tree, and random tree techniques, they achieved accurate classifications for LC50 global models, correctly identifying over 70% samples in training and testing. Remarkably, the functional tree demonstrated high accuracies of 93 and 100% for predicting the toxicity of metallic NPs to *Danio rerio* in training and testing, respectively.

Gernand *et al.* [149] conducted an investigation utilizing classification and regression techniques (regression trees and random forest) to forecast the pulmonary toxicity of 17 varieties of carbon nanotubes. The utilized model incorporated descriptors associated with the nanotubes' categories and dimensions, coverage time and dose, concentrations of metallic impurities, and features of the visible rodents. The pulmonary toxicity measures consisted of assessing the levels of macrophages and neutrophils, as well as concentrations of LDH and the total number of proteins. The models exhibited strong predictive capabilities for the four pulmonary endpoints, achieving r^2 values in the range of 0.88 and 0.96. Key attributes influencing carbon nanotube pulmonary toxicity encompassed metallic impurity quantities and identities, nanotube surface area, length and diameter, and cumulative size. To address NP agglomeration, a pivotal factor in NM biological effects, surface

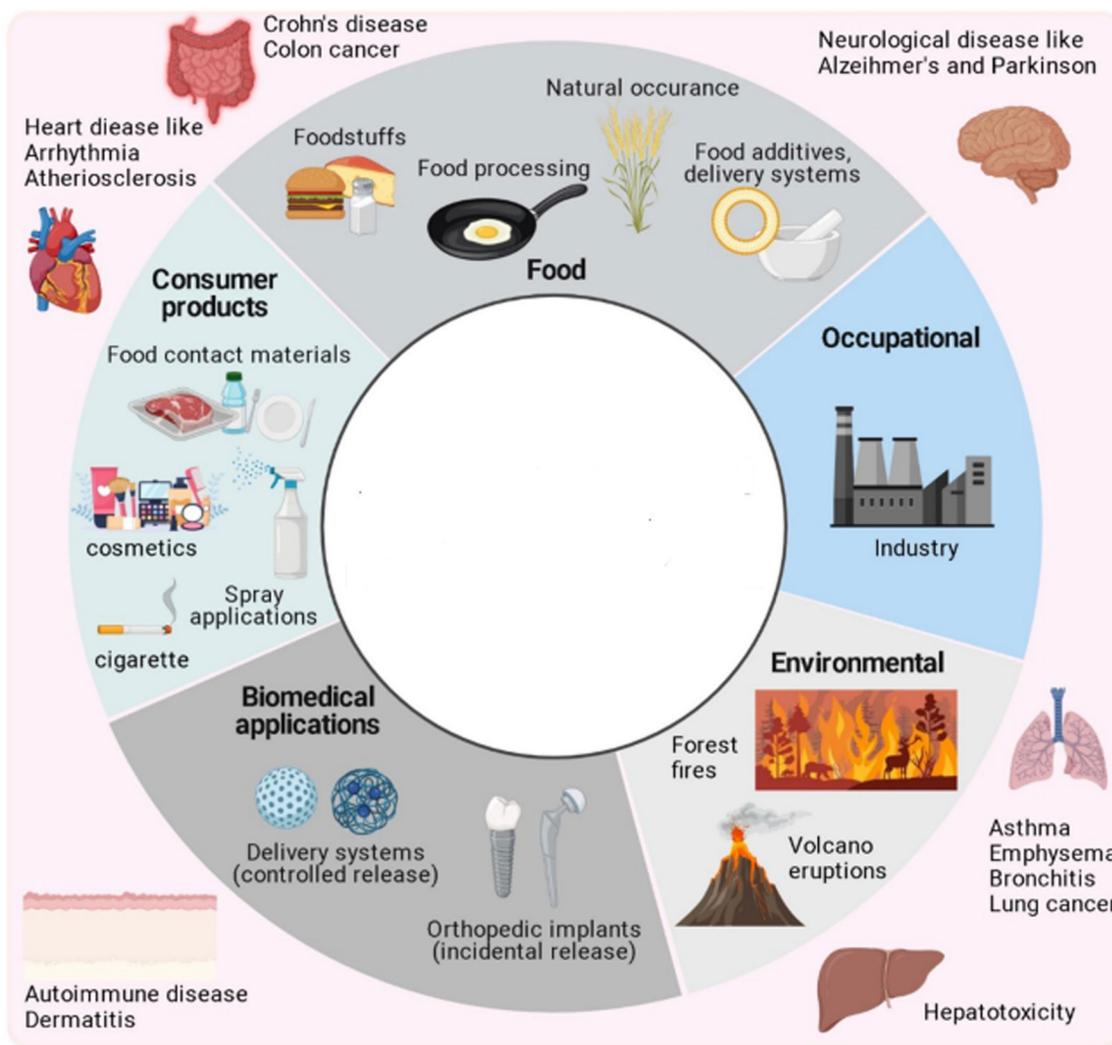


Figure 4: The potential origins (inner circle) and the resultant health effects (outer circle) stemming from exposure to NMs [2].

Table 5: Application of ML and AI in nanosafety

Reference	Year	NM type	ML techniques	Biological effects studied	Performance metrics	Key findings
Puzyn <i>et al.</i> [143]	2011	Metal oxide NPs	One-parameter linear regression	Cytotoxicity against <i>Escherichia coli</i>		Developed a linear regression model to predict cytotoxicity
Epa <i>et al.</i> [144]	2012	Metal oxide NPs	Linear regression, Bayesian neural networks	Smooth muscle cell apoptosis, NP uptake	Standard error, Predictive accuracy	Created predictive models for accurate predictions for apoptosis and uptake
Liu <i>et al.</i> [146]	2013	Iron oxide core NPs	Self-organizing map, Bayesian classifier, logistic regression, linear discriminate analysis, nearest neighbor classifiers	Impact on four cell types, using various assays	Relatively high accuracies exceeding 78%	Effective classification models for assessing the impact of iron oxide core NPs on different cell types
Germand <i>et al.</i> [149]	2014	Carbon nanotubes	Classification and regression trees, random forest	Pulmonary toxicity	r^2 values ranging from 0.88 to 0.96	Demonstrated robust predictive capabilities for pulmonary toxicity, identifying key attributes contributing to toxicity
Mikolajczyk <i>et al.</i> [150]	2015	Metal oxide NPs	Linear regression	Zeta potential	RMSE error, r^2 value	Predicted zeta potentials with RMSE error of 1.25 mV and r^2 of 0.87
Papa <i>et al.</i> [147]	2015	TiO ₂ and ZnO NPs	Multiple linear regression, neural networks, support vector machines	Cytotoxicity assessed through lactate dehydrogenase levels	Errors ranging from 8 to 17% relative to control cells	Nonlinear ML methods outperformed near-regression models, accurately predicted LDH levels
Chen <i>et al.</i> [148]	2016	NMs	Various tree algorithms (functional tree, decision tree, random tree)	Ecotoxicity classification for multiple species	Accuracies of 93 and 100% for predicting toxicity to Danio rerio in training and test sets.	Achieved accurate classifications for LC50 global models
Fourches <i>et al.</i> [39]	2016	Carbon nanotubes	k-Nearest neighbors, random forest, Support vector machines	Various biological endpoints including acute toxicity	Predictive accuracies of up to 77%	Predicting biological endpoints and screening potential surface ligands
Le <i>et al.</i> [152]	2016	ZnO NPs	Linear regression, Bayesian neural network	Biological responses in human cells (cell viability, membrane integrity, oxidative stress)	High r^2 values. Low standard errors	Nonlinear ML models outperformed linear ones, modelling entire NP dose-response curve
Oksel <i>et al.</i> [153]	2016	NMs Genetic programming-based decision tree (GPTree)	Nanostructure-activity relationship models	Training accuracy = 98-100%, Test accuracy = 86-100%	GPTree provided accurate nanoSAR models	
Concu <i>et al.</i> [154]	2017	Metal, metal oxide, and silica NPs	Linear neural networks, radial basis functions, multilayer perceptrons, probabilistic neural networks	Ecotoxicity and cytotoxicity across various organisms	Area under ROC curve values of 0.999 for the training set and 0.998 for the test set. 98% two-class accuracy in the test set	Exceptional predictive accuracy for diverse NPs
Wang <i>et al.</i> [155]	2017	Gold NPs	k-Nearest neighbors	Cellular uptake, induction of oxidative stress, hydrophobicity	High cross-validated predictivity	Demonstrated efficacy in predicting various NP properties
Fourches <i>et al.</i> [142]	2017	Various NPs	Support vector machine, kNNs-based regression	Various toxicological properties	External prediction accuracy	73% external prediction accuracy for classification models and r^2 of 0.72 for regression models
Kovalishyn <i>et al.</i> [156]	2018	Inorganic NMs	k-nearest neighbours, random forest, neural networks			Developed models for diverse endpoints, providing predictive

(Continued)

Table 5: Continued

Reference	Year	NM type	ML techniques	Biological effects studied	Performance metrics	Key findings
Hatamina <i>et al.</i> [157]	2019	Iron oxide NPs	Neural network	Toxicological or ecotoxicological properties and physicochemical properties Toxicity on kidney cells	q^2 values (0.58 to 0.80) for cross-validated regression models, r^2 values (0.49 to 0.78) for test sets Specific statistical values not provided in the graphs	power for toxicological/ecotoxicological properties High predictability in modelling toxicity, detailed analysis of input parameters on kidney cell viability Complex protein pattern on NP surfaces predicts organ uptake
Lazerovits <i>et al.</i> [67]	2019	NPs	Supervised DNN	Adsorption of blood proteins, blood clearance, organ accumulation	Accuracy $\leq 94\%$ for predicting NP spleen and liver accumulation	

charge effects on dispersed NPs were considered to prevent cohesion. In a separate study, Mikolajczyk *et al.* [150] utilized ML techniques to estimate the zeta potential, which indicates surface charge, for a set of 15 metal oxide NPs, using both computed descriptors and image based. Despite the application of linear regression methods, their models demonstrated zeta potential predictions with a root-mean-square error (RMSE) of 1.25 mV in the test dataset and an r^2 coefficient of determination of 0.87.

A revolutionary study by Fourches *et al.* [39] implemented the principles of “safe-by-design” by employing ML techniques to assess biological outcomes. The investigation centred on an 83-member library of surface-modified carbon nanotubes having consistent dimensions. Various biological activities, including carbonic anhydrase, bovine serum albumin, chymotrypsin, hemoglobin activity, and *in vitro* analyses for immune toxicity and acute toxicity, were meticulously examined. By leveraging conventional ML models such as kNNs, SVM, and random forest, they achieved impressive predictive accuracies, reaching 75% for protein binding and 77% for acute toxicity outcomes. The study revealed chemical surface features associated with specific biological actions, leading to the virtual screening of a huge library having 240,000 impending carbon nanotubes’ surface ligands. Experimental validations of these predictions demonstrated remarkable accuracy. This systematic approach, involving the modification of physicochemical properties, comprehensive biological evaluation, and computational examination, aligns seamlessly to forecast nanotoxicology outcomes and the principle of designing safety into NMs. It significantly contributes to advancing the automated comprehension of nanobio interactions and aids the development of quantitatively predictive and vigorous models for NM characteristics within applicable domains [151].

In a similar context, Le *et al.* [152] investigated 45 ZnO NPs, systematically altering particle dimensions, doping type, shape proportions, dopant density, and surface finishing. The biological response data resulting from this experimentation was modelled using Bayesian regularized NN and linear regression ML techniques. The biological analyses encompassed evaluations of membrane integrity, cell viability, and oxidative stress, describing the damage caused to human liver carcinoma cells (HepG2) or human umbilical vein endothelial cells by ZnO NPs. The nonlinear ML models demonstrated superior performance compared to the linear counterparts, offering predictions for an external test set with remarkable r^2 values and low standard errors.

Oksel *et al.* [153] introduced a new strategy employing a genetic programming-based decision tree (GPtree) to depict NM properties. This method, known for its robustness and

automation, excels in constructing precise nanostructure–activity relationship (nanoSAR) models, especially when confronted with limited datasets. GPtree autonomously identifies relevant descriptors, confirming model accuracy while improving interpretability. Indicating its adaptability, the approach effectively developed precise nanoSAR models for four different datasets. Particularly, these models showed parsimony by utilizing solely thirteen (13) predictors from a wide collection of descriptors, attaining high accuracies going from 98 to 100% during training and 86 to 100% during testing, correspondingly. The decision trees generated by GPtree provided visually intuitive representations of decision thresholds associated with each descriptor, considerably improving the interpretability of models predicting the structure–activity link of NPs.

Concu *et al.* [154] conducted an extensive study involving 260 silica, metal, and metal oxide NPs encompassing 31 chemical compositions from various literature sources. Their investigation included measurements of ecotoxicity and cytotoxicity across a spectrum of organisms, and bacteria, spanning algae, fungi, plants, fishes, crustaceans, and mammalian cell lines. Through random combinations, they generated 54,371 pairs of NPs, with training (40,804 pairs) and test sets (13,567 pairs) formed randomly. Employing linear NNs, radial basis functions, multilayer perceptrons, and probabilistic NNs, the models demonstrated outstanding predictive precision, achieving an area in the receiver operator characteristic (ROC) curve values of 0.99 and 0.998 for the training and testing, respectively. The two-class accuracy for the test set reached 98%. These models also provided insights into the descriptors influencing adverse biological responses. Despite employing Y-scrambling to assess overfitting, the authors noted the challenge posed by unrealistically improved accuracy and ROC values, raising concerns about potential chance correlation problems in the models.

Wang *et al.* [155] analyzed NPs of gold, exploring variations in size, surface modifications, and surface coverage. Their NP collection comprised 34 particles, and by employing 29 descriptors along with the kNN algorithm, they created predictive models for the absorption of substances by human kidney cells and lungs, the capacity to activate oxidative stress and hydrophobicity. Both models, incorporating 11 or fewer descriptors, demonstrated excellent predictability during cross-validation, with r^2 values of 0.99, 1.0, 0.99, and 0.97, respectively, evaluated through tenfold cross-validation. This study showcased the efficacy of their approach in predicting various NP properties for small descriptors set.

Kovalishyn *et al.* [156] compiled a dataset from 128 literature sources, incorporating 964 data points for inorganic NMs. Their analysis encompassed toxicological/ecotoxicological

properties (EC50, MIC, LC50, and mortality rate) as well as physicochemical characteristics for metal and metal oxide NPs within the size range of 1–90,000 nm. Utilizing random forest, kNN, and NN techniques, models were developed for these endpoints across various species. Evaluation of model predictive power through cross-validation and test sets yielded q^2 values of 0.58–0.80 for cross-validated regression techniques and r^2 values of 0.49–0.78 for test sets. However, the encoding method for species within the models remains unclear.

Hataminia *et al.* [157] utilized an NN to simulate the nephrotoxicity induced by iron oxide NPs on kidney cells. The model, which underwent training based on NP size, concentration, incubation duration, and surface charge, demonstrated high predictability, although specific statistical values were not provided in the graphs. The study extensively analyzed the impact of these four input factors on the viability of kidney cells. While DL algorithms have made a substantial impact in various fields, including materials science, their application to modelling NM properties relevant to nanosafety has been limited [158]. The DNNs can automatically produce higher features for ML models. Surprisingly, this capability has not been widely explored in predictive nanotoxicology, likely due to the shortage of nanospecific descriptors hindering progress.

Most applications of DL algorithms involve utilizing their image recognition capabilities for extracting microscopy image data of NPs/cells. Similarly, Coquelin *et al.* [159] used CNNs to estimate particle size distribution from scanning electron microscopy (SEM) images of aggregated TiO₂ particles, aiming to automate SEM measurements. Context encoding was introduced to forecast absent components of combined NPs. Rusk *et al.* [160] applied the CNNs techniques for the segmentation of transmission electron microscopy images for different reliable size distribution calculations. Ilet *et al.* [161] employed image analysis utilizing open-source software tools, CellProfiler, along with a CNN algorithm named ilastik, to study NP distributions.

Lazerovits *et al.* [67] presented a significant utilization of ML in the area of nanosafety. Their study examined how blood proteins adhere to NPs after intravenous injection. Through the employment of data from protein mass spectrometry as inputs and correlating them with blood clearance and organ accumulation as outputs, they effectively trained a supervised DNN. The results from the network accurately anticipated the accumulation of NPs in the liver and spleen, revealing the considerable impact of the intricate protein pattern on the NP surface on organ uptake. This valuable insight has delivered a basis for the advance of NPs with obviously reduced accumulation in the liver and spleen.

3.4 RQ4: What are the past NM sustainability roadmaps?

Over the past decade, several significant roadmaps for the safe utilization of NMs have been proposed. The primary comprehensive set of milestones was established at a workshop sponsored by COST (European Cooperation in Science and Technology) in 2011, focusing on Quantitative Nanostructure Toxicity Relationships held in Maastricht [162]. The milestones acknowledged in this roadmap and subsequent reviews emphasized crucial aspects for advancing NM research, including:

- Improved material characterization for biological activity trials.
- Creation of NM-oriented descriptors.
- Development of high output *in vitro* methods.
- Enhanced tracking techniques for NPs in the body.
- Improved data storage/sharing.
- Production of *in vivo* data.
- Development of models for predicting NP behaviour.

These milestones, originally research-driven, have significantly influenced research agendas in Europe and the United States. The Maastricht roadmap, in particular, played a pivotal role in securing European Union funding and was used in governmental policies on NM safety, including the UK Parliamentary Office of Science and Technology [162]. The roadmap's impact is further reflected in the funding of numerous associated COST Actions, as well as now Horizon projects (2020) focusing on nanosafety [52,86,109,163,164,165]. Despite the passage of time, many milestones from the initial Maastricht workshop remain pertinent due to slower-than-expected progress in achieving them. Factors contributing to this slower progress include the delayed adoption of high-throughput characterization and synthesis technologies, resulting in a scarcity of big and robust datasets for ML model training and a lack of interpretable and efficient NMs descriptors.

To address these challenges, Worth *et al.* [166] proposed a recent US and EU Roadmap of nanoinformatics up to 2030, with a specific emphasis on using computational methods to predict potential adverse effects of NMs. This roadmap aligns with the previously identified milestones and emphasizes the need for continued efforts in nanosafety research. Different milestones of 2-, 5-, and 10-year duration have not been fulfilled, emphasizing the importance of ongoing efforts. To tackle these challenges, significant initiatives financed by the EU Horizon 2020 program aim to establish joint networks of scientists dedicated to crafting computational solutions for predicting nanosafety. These projects focus on generating relevant data, implementing nanoinformatics-driven

decision-support strategies, and developing methodologies for safe-by-design NMs.

Challenges faced by regulators include adapting to a fast-varying system of industrial NMs. The NanoReg2 initiative deals with this problem by linking structure-based design to the regulatory procedure, verifying new NMs through established grouping methods. NanoSolveIT employs a nanoinformatics-centred decision-support methodology, utilizing advanced *in silico* techniques to recognize key features of NMs accountable for unfavourable effects.

Horizon 2020 includes 67 toxicology-related projects, including 11 focused on NM toxicity, underscoring the commitment to advancing research in this field [www.fabiadisconzi.com/open-h2020/per-topic/toxicology/list/index.html]. Recent literature also emphasizes the need for sustainable nanotechnology system governance for European society [167], aligning with the efforts of regulatory and standardization communities, such as FDA, EPA, ECHA, ISO, and OECD, in developing approved techniques for measuring inherent and external attributes of NMs.

3.4.1 Data

ML techniques hold significant potential for modelling and predicting both advantageous and detrimental characteristics of NMs. Their appeal arises from their rapid processing, independence from intricate mechanistic models, utilization of open-source and accessible technologies, and demonstrated effectiveness in discerning patterns within intricate datasets across various domains [168]. As methods grounded in data-driven principles, their efficacy is inherently tied to the information employed in their creation. The abundance, quality, diversity, and scope of data play pivotal roles in crafting resilient and anticipatory ML models that can be widely applied across diverse domains.

3.4.2 Quantity of data

The progress of nanosafety research projects faces significant challenges due to the scarcity of substantial datasets, primarily attributed to the time, expenses, and ethical concerns related to collecting data involving animals and humans. Unlike the assessment of properties in isolated, well-defined organic pristine or molecules NMs unaffected by biological environments, evaluating the NMs features that exhibit variations with time or in diverse biological fluids lacks standardization and poses technical complexities [169]. To address these limitations, read-across techniques computational models and computational models are

employed to bridge data gaps [167]. However, the optimal approach is unquestionably the expansion of available datasets. In the past, advancements have been made in the high-throughput synthesis of NMs [170], accelerated physicochemical features [171], and swift *in vitro* toxicity screening [172,173]. The ongoing development, widespread practice, and use of these high-throughput approaches hold great promise for substantially augmenting the volume of NMs in the upcoming years.

3.4.3 Quality of data

In addition to challenges associated with the scale of datasets, similar to other material types, issues frequently arise regarding the data reliability of experiments on the biological and physicochemical impacts of NMs. Constraints related to time, cost, and ethical considerations not only limit the quantity of available data but also impact the feasibility of conducting multiple experimental replicates. These constraints have repercussions on the quality of data, signal-to-noise ratio, and the identification and handling of outliers. The reproducibility of scientific findings has emerged as a critical concern in recent years, with specific attention directed towards NMs. Various research groups have contributed methodologies aimed at enhancing the reproducibility of NM synthesis [174–177]. Addressing the reproducibility of NM bioassays, Petersen *et al.* [178] have made significant contributions, while Baer *et al.* [179] have explored the influence of materials provenance on reproducibility.

Accompanying experiments on NMs offers greater challenges as compared to small organic molecules, mainly due to intrinsic differences in their shape and size, tendencies to agglomerate, and the creation of corona. Identifying the complexity of these features, researchers are actively striving to improve experimental duplicability. Baer *et al.* [179] and Petersen *et al.* [178] have explored different features of NM bioassays and the impact of material sources on reproducibility, respectively. Notably, Galmarini *et al.* [180] studied the sophisticated effects of corona formation on experimental reproducibility.

3.4.4 Diversity

In conclusion, it is imperative to develop broadly applicable models for predicting potential negative characteristics of NMs. These models should be trained using data obtained from a diverse array of NMs and encompass a comprehensive range of relevant biological endpoints. Recent papers have underscored the importance of conducting systematic

studies that explore various NM morphologies, types, and endpoints [151,181].

3.4.5 Description

NPs exhibit various attributes, such as agglomeration, dissolution, photocatalytic properties, and the formation of persistent complexities with biological ions. These characteristics significantly impact the biological behaviour of NMs [126,182,183], although advancements have been achieved in discerning the location, timing, and methodology for characterizing NMs, it remains challenging due to the dynamic nature of NM characteristics, which can vary considerably under various circumstances and with time. Further, these varying characteristics are highly responsive to the environment, with interactions among properties adding complexity to the characterization process.

Ideally, it is preferable to evaluate the physicochemical features of NMs within comparable biological situations, as the original characteristics of NMs tend to vary when they come into connection with biological liquids [184]. NM surfaces rapidly develop a biological layer, making a bio-corona that modifies material properties like surface charge, particle dimensions, and aggregation status quickly [185]. The corona experiences dynamic changes over time, with primarily attached macromolecules being gradually replaced by those with stronger affinities. Additional investigation is essential to fully recognize how NMs' physicochemical properties, including surface chemistry, size, aggregation, and dissolution, are adapted in different biological settings. It is crucial to determine whether these modifications make NMs more or less toxic compared to their original state. This information is crucial for regulators to more accurately measure the risks posed by NMs to human/environmental health.

Encouragingly, advancements are being achieved in the characterization of the NP corona based on the physicochemical properties of NMs and the surrounding environment. Recent research has presented a range of mass spectrometric, spectroscopic, and chromatographic techniques aimed at evaluating NP coronas. Among these methods are some that allow the analysis of the evolving properties of biocoronas over time (Figure 5) [186–194]. These results in the creation of descriptive and potentially predictive models for the conformation of NP corona [129,195–197] and its active behaviour. These models lay the basis for the rational design and progress of NMs aimed at reducing or fully avoiding corona formation [198].

Considering solubility and biopersistence is paramount in the assessment of nanosafety, as these factors directly

impact internal exposure to NMs. The main source of biological activity in soluble NMs may arise primarily from the liquefied substance rather than the NP itself, as exemplified by ZnO. The dissolution rate is subjective to various physical characteristics of NPs (size, particle shape, coating, shape, core doping), allowing potential opportunities for rationally controlling dissolution to reduce hostile biological effects [152]. Regardless of its critical significance, there remains uncertainty regarding which features of NMs impact solubility and, more specifically, biopersistence [199]. Further research is needed to comprehend how the dissolution of NMs in biological liquids or their perseverance in the body influences biodistribution and biological effects, and how these processes can be improved through formulations, coatings, or other approaches.

3.4.6 Descriptors

Molecular descriptors, capturing the necessary structural and physicochemical qualities of NMs, are pivotal components in computational modelling studies across various scientific domains, including environmental/medicinal chemistry, pharmacology, toxicology, genomics, and drug design [200,201]. These descriptors can be created either through theoretical techniques or standard experiments. Theoretical descriptors offer diverse chemical property sources and wide coverage of the chemical features space, providing an extensive representation of NMs. However, experimentally derived descriptors, such as shape, size, solubility, and agglomeration, have limited applicability in ML models for biological effects due to their resource-intensive nature and unavailability for hypothetical materials [202]. Numerous theoretically sound descriptors can be calculated directly from molecular structures using different software packages [203]. Nevertheless, existing molecular descriptors and fingerprints primarily

result from pharmaceuticals specifically designed for organic small molecules, lacking specificity for NMs. The unique challenges posed by NMs, such as polydispersity, complex interactions, and incomplete characterization, necessitate the development of nanospecific descriptors. DL, coupled with imaging methods, offers a promising avenue for generating such descriptors. Russo *et al.* [204] utilized CNNs to digitize complex nanostructures, allowing the training of predictive models for NP biological properties.

The interpretability of NM descriptors is crucial for understanding the mechanisms behind biological responses and for the rational design of improved materials. Traditional descriptors often lack interpretability, prompting the need for descriptors that are easily related to chemically recognizable features. De *et al.* [205] and Sizochenko *et al.* [206] proposed efficient and interpretable methods for encoding NM properties. De *et al.* [205] utilized 23 descriptors derived from the periodic table, while Sizochenko employed a “liquid drop” model and a simple representation of molecular structure for NPs.

3.4.7 Domain

ML models are essential tools in predicting NM properties, yet their efficacy is constrained by the limited datasets available, covering only a limited range of biological and physicochemical attributes [207]. The scarcity of diverse and large datasets represents a major impediment to the widespread adoption of ML techniques for NMs’ biological characteristics [208]. Computational models determine greater efficacy when operating within the multidimensional space described by different descriptors and biological reactions contained in the training data, and their reliability diminishes as predictions move further from this domain [37].

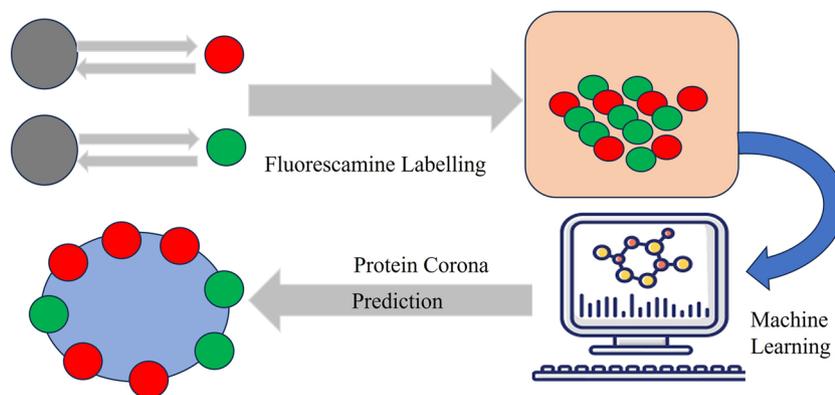


Figure 5: Fluorescence levels to correlate with the number of associated proteins in the corona of various NMs.

Novel modelling techniques, such as nearest neighbor Gaussian processes [37], and current ML approaches leveraging Bayesian statistics [208], offer avenues to estimate the prediction reliability. As larger and more diverse datasets become available, coupled with the adoption of ML methods that better gauge prediction uncertainties, the challenges associated with “out of domain” predictions are expected to decrease [37]. The prediction associated with *in vivo* responses to NMs remains a persistent challenge, primarily due to ethical, time, and cost constraints limiting experimental *in vivo* data generation [209,210]. By employing effective design of experiment techniques to minimize *in vivo* experiments and leveraging ML models, predictions within the models’ domains of applicability can be made. However, constraints on experimental *in vivo* testing are likely to persist, prompting the exploration of surrogate bioassays like toxicogenomics or new phenotypic assays [209,210].

The rapid advancement of high-throughput “omics” techniques, encompassing proteomics, transcriptomics, and metabolomics, allows for the generation of extensive datasets on cellular or organism responses to NM exposure [209,210]. ML modelling methods can exploit these data by using expression profiles as descriptors to enhance predictions of *in vivo* responses and by employing measured nanodescriptors to model associations between NM omics signatures or pathways and physicochemical properties identified using expression profiling [211]. Furchi *et al.* [211] employed physicochemical descriptors alongside experimental conditions to model biological results from genome-wide studies, demonstrating the potential of Bayesian networks. The integration of omics data with ML further contributes to modelling toxicological properties of NMs for risk assessment and safety [18–20]. Metabolomics, offering insights into metabolic dysregulation caused by NM exposure, is a valuable tool for risk assessment [212]. Metabolic and genomic fingerprints, coupled with sparse attribute selection and feature importance techniques, facilitate the training of predictive ML models for evaluating potential adverse biological characteristics of NMs.

Hybrid ML modelling techniques, combining *in vitro* analysis results and nanodescriptors to model limited *in vivo* data, hold promise [212]. Lee *et al.* [213] demonstrated the success of merging molecular descriptors for medications with basal cytotoxicity analyses *in vitro* to better predict human critical toxicity compared to relying only on descriptors or *in vitro* outcomes. Recent reviews by Collins *et al.* [172] comprehensively cover advancements in high-throughput alternative techniques for evaluating the toxicity of NMs.

3.5 RQ5: What are the future NM sustainability roadmaps?

The removal or reduction of the aforementioned challenges above in nontoxicity research would provide access to a substantial and diverse array of pertinent datasets, along with a set of efficient nanospecific descriptors. Numerous well-validated ML algorithms are available for training on such data and descriptors, enabling the development of robust models for critical toxicological endpoints. Recent assessments have underscored the significance of particular ML methodologies in the modelling of nanotoxicity information [5,11,13,40,214,215]. Brown *et al.* [12] conducted a review highlighting ML’s application in analyzing large nanoscience datasets, accelerating material discovery, and exploring the science behind memristive devices for ML hardware implementation. Ahmad *et al.* [19] reviewed the use of DL techniques in modelling NM properties relevant to nanosafety. However, there is a shortage of scientists with the necessary multidisciplinary knowledge to apply new technological advancements to computational nanotoxicology. The training of researchers in the skills required to create robust models for potential negative effects of NMs and probe mechanistic aspects of NM-biology interactions is still limited. The dominant paradigm in many scientific areas, including nanosafety, involves large multidisciplinary teams. Initiatives like the H2020 projects in the EU and similar efforts in the United States and Asia are working to build critical mass in this field [216].

Advancements in accurately detecting little particles in intricate samples have unveiled that several procedures and instruments produce NPs in the environment. Although certain NPs can traverse biological membranes, presenting novel opportunities in diagnostics and therapies, concerns regarding potential risks to human health from exposure to NMs have gained heightened importance. Experimental nanotoxicology, alongside its computational counterpart, contributes significantly to understanding NM interactions with biological systems, predicting potential adverse effects, developing safe-by-design techniques for industry, and delivering data and models for regulatory decisions.

Making predictions, particularly about the future, is challenging, especially considering the inability to achieve the milestones set almost a decade ago. Despite these challenges, it is plausible to develop a roadmap within the next ten years that effectively addresses the existing challenges in NMs research [27]. Notably, the advancements in ML methods, particularly in DL, hold promise in overcoming some of the obstacles outlined in the review. While these methods demand substantial data, they have the potential to simplify the creation of nanospecific descriptors and may introduce innovative ways of utilizing raw data.

Critical to the progress is the transparent sharing of negative outcomes from toxicological studies, as this contributes significantly to the cumulative knowledge in nanoscience [27]. Recognizing that a lack of adverse effects is a positive outcome in nanosafety, models need to discern which NP features lead to adverse effects and which do not. This approach aligns with the vision of achieving rational “safe-by-design” NMs. The speed at which these promising solutions to computational nanotoxicology challenges are adopted remains speculative, raising the intriguing possibility of a forthcoming decade marked by a golden age of safe nanotechnology use that enhances the quality of life for all [27].

3.6 Latest findings and their novelty in AI and ML integration for NM sustainability

Table 6 presents a collection of the most recent discoveries and their importance in the integration of AI and ML for the sustainability of NMs.

These findings underscore the pivotal role of AI and ML in advancing nanotechnology, particularly in enhancing the safety and sustainability of NMs. The integration of these technologies offers significant potential to improve predictive modelling, risk assessment, and the overall management of NMs, thereby addressing key concerns related to their use and impact.

4 Future research recommendations

4.1 RQ6: What are the recommendations for future research in the intersection of nanotechnology and ML?

Below are some specific suggestions for future research at the intersection of nanotechnology and ML:

4.1.1 Advanced predictive models development

- Utilize ML algorithms to forecast properties and behaviours of novel NMs based on their atomic structure, composition, and synthesis methods.
- Explore DL techniques to effectively handle complex datasets and nonlinear relationships.

4.1.2 Enhanced NM design

- Investigate ML-driven approaches for crafting NMs with customized properties for specific applications like drug delivery, catalysis, and energy storage.
- Integrate multi-objective optimization techniques to balance conflicting material properties.

Table 6: Latest findings and their significance in AI and ML integration for NM sustainability

Research theme	Latest findings	Significance
Toxicological modelling	Utilized ML to model the toxicological properties of NMs using omics data; reviewed potential of combining omics with ML for safety assessment	Enhanced predictive accuracy in safety evaluations of NMs, crucial for risk management and regulatory compliance
Predictive modelling techniques	Developed various computational models, including QSARs and QSPRs, to predict NP toxicity	Facilitates the early identification of potentially hazardous NMs, streamlining safety testing procedures
European commission initiatives	Launched several projects to explore modelling NP toxicity and properties	Demonstrates institutional support for advancing NM safety research using AI/ML methodologies
Integration with omics data	Explored integrative omics combined with ML to profile NMs for biological impact modelling	Provides comprehensive insights into the biological interactions and impacts of NMs, improving risk assessments
Advancements in ML techniques	Reviewed various advanced ML techniques for improving the prediction of NM behaviour and properties	Highlights the ongoing evolution and refinement of ML techniques to enhance the reliability of NM assessments
Safety and risk assessment	Emphasized the use of AI/ML to enhance simulations and safety checks for NMs	Promotes the development of more robust and accurate safety protocols for NM usage in various applications

4.1.3 Toxicity prediction and risk assessment

- Create ML models for predicting NM toxicity and evaluating their environmental and biological impacts.
- Explore integrative methods that merge experimental data with computational predictions to enhance accuracy and reliability.

4.1.4 Nanomedicine and therapeutics

- Utilize ML algorithms to personalize medicine by analyzing patient data to design targeted nanotherapeutics and optimize drug delivery systems.
- Investigate reinforcement learning for adaptive drug delivery and treatment optimization.

4.1.5 NM characterization and imaging

- Develop ML-based techniques to interpret data from advanced imaging and characterization methods like electron microscopy and spectroscopy.
- Explore unsupervised learning for automated feature extraction and classification.

4.1.6 Smart nanodevices and sensors

- Investigate ML approaches to design intelligent nanodevices and sensors capable of real-time monitoring and feedback across various applications.

4.1.7 Materials discovery and exploration

- Employ ML algorithms to expedite the discovery of novel NMs with desired properties such as high strength and conductivity.
- Explore generative models for automated materials synthesis and virtual screening.

4.1.8 Interdisciplinary collaboration

- Foster collaboration among nanotechnologists, materials scientists, chemists, biologists, and computer scientists to address complex challenges.

4.1.9 Ethical and societal implications

- Consider ethical, societal, and regulatory implications including data privacy, security, environmental impact, and equitable access to technology.

5 Conclusion

The combination of nanotechnology and AI marks a crucial point in addressing safety and toxicity concerns associated with nanoproducts. As nanotechnology progresses across various industries, the need for strong safety assessments becomes increasingly clear. Although nanoproducts are widely used, worries about their potential harmful effects remain, necessitating careful risk management. This SLR advocates using AI and ML to enhance simulations and improve safety checks for NMs. By thoroughly examining existing research, the study highlights how AI and ML can promote sustainability in NMs in five key areas. It discusses how these technologies can help identify issues, innovate new applications for NMs, and improve their management. The review also assesses past efforts to ensure the safety of NMs and suggests new ideas for the future. By doing this detailed review, the study aims to show the current state of affairs, identify ongoing challenges, and propose ways to use AI and ML to make nanotechnology safer and more environmentally friendly. It also suggests using these technologies to monitor the real-world use of NMs. The study aims to address current problems and promote safer and more sustainable use of nanotechnology.

Facing the challenges identified in this review, the collaboration between nanotechnology and ML not only addresses current limitations but also sets a path for future advancements. The partnership between these innovative fields ensures the sustainable and responsible application of nanotechnology, fostering innovation while reducing potential risks. As we stand on the verge of an era marked by the merging of nanotechnology and AI, it is crucial to embrace these advancements thoughtfully, ushering in a new era of technological integration that prioritizes safety, effectiveness, and long-term sustainability.

While the SLR on the integration of AI and ML in the sustainability of NMs provides valuable insights, it has several limitations. Publication bias may skew results since studies with significant outcomes are more likely to be published, potentially overestimating AI and ML effectiveness. The scope of literature might be incomplete, excluding relevant studies in non-English languages or less accessible journals. The quality of included studies varies, with some having methodological weaknesses, affecting the overall conclusions. The heterogeneity of study designs, methodologies, and outcomes complicates unified conclusions and generalizable recommendations. As both AI/ML and nanotechnology are rapidly evolving fields, the review's findings might quickly become outdated. There are also limited data on long-term effects and real-world applications, with a lack of longitudinal studies. Interdisciplinary challenges in

integrating AI/ML with nanotechnology may have limited the depth of some studies. In addition, the findings may not be universally applicable across all NMs or industries due to the specific contexts of the studies. Future research should strive to include a diverse range of studies, improve methodological quality and consistency, conduct longitudinal studies, foster interdisciplinary collaborations, and regularly update reviews to reflect the latest advancements and trends.

Acknowledgments: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R111A1A01055652).

Funding information: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R111A1A01055652).

Author contributions: Inam Ur Rehman: conceptualization, analysis, design, and writing original draft. Inam Ullah: visualization, synthesis, and analysis. Habib Khan: literature search, data collection, visualization, and analysis. Mohammed Seghir Guellil: investigation, literature search, and reviewing. Shabana Habib: investigation, formal analysis, and reviewing. Muhammad Islam: data validation, reviewing, and visualization. Jakyoungh Min: writing, funding, and review. Mi Young Lee: study design, conceptualization, and visualization. JaKeoung Koo: reviewing, editing, and supervision.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: All data generated or analysed during this study are included in this published article.

References

- [1] Bayda S, Adeel M, Tuccinardi T, Cordani M, Rizzolio F. The history of nanoscience and nanotechnology: from chemical-physical applications to nanomedicine. *Molecules*. 2019;25(1):112.
- [2] Singh AV, Varma M, Laux P, Choudhary S, Datusalia AK, Gupta N, et al. Artificial intelligence and machine learning disciplines with the potential to improve the nanotoxicology and nanomedicine fields: a comprehensive review. *Archives Toxicol*. 2023;97(4):963–79.
- [3] Domingues C, Santos A, Alvarez-Lorenzo C, Concheiro A, Jarak I, Veiga F, et al. Where is nano today and where is it headed? A review of nanomedicine and the dilemma of nanotoxicology. *ACS Nano*. 2022;16(7):9994–10041.
- [4] Asati S, Sahu A, Jain A. Nanotoxicity: the dark side of nanoformulations. *Curr Nanotoxicity Prev*. 2021;1:6–25.
- [5] Singh AV, Ansari MHD, Rosenkranz D, Maharjan RS, Kriegel FL, Gandhi K, et al. Artificial intelligence and machine learning in computational nanotoxicology: unlocking and empowering nanomedicine. *Adv Healthc Mater*. 2020;9(17):1901862.
- [6] Amiri Z, Heidari A, Navimipour NJ, Unal M, Mousavi A. Adventures in data analysis: a systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. *Multimedia Tools Appl*. 2024;83(8):22909–73.
- [7] Parashar A, Parashar A, Ding W, Shabaz M, Rida I. Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches. *Pattern Recognit Lett*. 2023;172:65–73.
- [8] Norozpour S, Darbandi M. Proposing new method for clustering and optimizing energy consumption in WSN. *Talent Development Excellence*. 2020;29:12.
- [9] Darbandi M, Haghighoo S, Hajiali M, Khabir A. Prediction and estimation of next demands of cloud users based on their comments in CRM and previous usages. In: 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT). IEEE; 2018. p. 81–6.
- [10] Prasad K, Srinivasa Rao V, Harini P, Mukiri RR, Ravindra K, Vijaya Kumar D, et al. Machine learning algorithms are applied in nanomaterial properties for nanosecurity. *J Nanomaterials*. 2022;2022:1–14.
- [11] Winkler DA. Role of artificial intelligence and machine learning in nanosafety. *Small*. 2020;16(36):2001883.
- [12] Brown KA, Brittan S, Maccaferri N, Jariwala D, Celano U. Machine learning in nanoscience: big data at small scales. *Nano Lett*. 2019;20(1):2–10.
- [13] Furxhi I, Murphy F, Mullins M, Poland CA. Machine learning prediction of nanoparticle in vitro toxicity: A comparative study of classifiers and ensemble-classifiers using the Copeland index. *Toxicol Lett*. 2019;312:157–66.
- [14] Lakshmiopathy R, Balaji G, Rico ILR. Removal of Pb²⁺ ions by ZSM-5/AC composite in a fixed-bed bench scale system. *Adsorption Sci Technol*. 2021;2021:1–8.
- [15] Reddy PL, Deshmukh K, Kovářk T, Reiger D, Nambiraj NA, Lakshmiopathy R, et al. Enhanced dielectric properties of green synthesized Nickel Sulphide (NiS) nanoparticles integrated polyvinylalcohol nanocomposites. *Materials Res Express*. 2020;7(6):064007.
- [16] Bensemlali M, Bajolle F, Laux D, Parisot P, Ladouceur M, Fermont L, et al. Neonatal management and outcomes of prenatally diagnosed CHDs. *Cardiol Young*. 2017;27(2):344–53.
- [17] Yoon SA, Hong WH, Cho HJ. Congenital heart disease diagnosed with echocardiogram in newborns with asymptomatic cardiac murmurs: a systematic review. *BMC Pediatrics*. 2020;20:1–10.
- [18] Ahmad F, Wang X, Li W. Toxicometabolomics of engineered nanomaterials: progress and challenges. *Adv Funct Materials*. 2019;29(51):1904268.
- [19] Ahmad F, Mahmood A, Muhmood T. Machine learning-integrated omics for the risk and safety assessment of nanomaterials. *Biomaterials Sci*. 2021;9(5):1598–608.
- [20] Ahmad F, Abubshait SA, Abubshait HA. Untargeted metabolomics for Achilles heel of engineered nanomaterials risk assessment. *Chemosphere*. 2021;262:128058.

- [21] Furxhi I, Murphy F, Mullins M, Arvanitis A, Poland CA. Nanotoxicology data for in silico tools: a literature review. *Nanotoxicology*. 2020;14(5):612–37.
- [22] Furxhi I, Murphy F, Mullins M, Arvanitis A, Poland CA. Practices and trends of machine learning application in nanotoxicology. *Nanomaterials*. 2020;10(1):116.
- [23] Haase A, Klaessig N, EU US roadmap nanoinformatics 2030. Project Report. Zenodo, EU Nanosafety Cluster; 2018.
- [24] Puzyn T, Jeliaskova N, Sarimveis H, Robinson RLM, Lobaskin V, Rallo R, et al. Perspectives from the NanoSafety Modelling Cluster on the validation criteria for (Q) SAR models used in nanotechnology. *Food Chem Toxicol*. 2018;112:478–94.
- [25] Furxhi I. A machine learning examination of nanomaterial safety. Ireland: University of Limerick; 2020.
- [26] Pokrajac L, Abbas A, Chrzanowski W, Dias GM, Eggleton BJ, Maguire S, et al. Nanotechnology for a sustainable future: Addressing global challenges with the international network4-sustainable nanotechnology. Washington, US: ACS Publications; 2021.
- [27] Karakus CO, Winkler DA. Overcoming roadblocks in computational roadmaps to the future for safe nanotechnology. *Nano Futures*. 2021;5(2):022002.
- [28] Scott-Fordsmand JJ, Amorim MJ. Using machine learning to make nanomaterials sustainable. *Sci Total Environ*. 2023;859:160303.
- [29] Zuhaib M, Shaikh FA, Tanweer W, Alnajim AM, Alyahya S, Khan S, et al. Faults feature extraction using discrete wavelet transform and artificial neural network for induction motor availability monitoring-internet of things enabled environment. *Energies*. 2022;15(21):7888.
- [30] Alzoubi YI, Gill AQ, Al-Ani A. Empirical studies of geographically distributed agile development communication challenges: A systematic review. *Inform Manag*. 2016;53(1):22–37.
- [31] Magsi AH, Ghulam A, Memon S, Javeed K, Alhussein M, Rida I. A machine learning-based attack detection and prevention system in vehicular named data networking. *Comput Mater Contin*. 2023;77(2):1445–65.
- [32] Huang R, Xiao R, Zhu W, Gong P, Chen J, Rida I. Towards an efficient real-time kernel function stream clustering method via shared nearest-neighbor density for the IIoT. *Inform Sci*. 2021;566:364–78.
- [33] Deng Y, Teng S, Fei L, Zhang W, Rida I. A multifeature learning and fusion network for facial age estimation. *Sensors*. 2021;21(13):4597.
- [34] Fei L, Zhang B, Xu Y, Tian C, Rida I, Zhang D. Jointly heterogeneous palmprint discriminant feature learning. *IEEE Trans Neural Netw Learn Syst*. 2021;33(9):4979–90.
- [35] Khan H, Haq IU, Munsif M, Mustaqeem, Khan SU, Lee MY. Automated wheat diseases classification framework using advanced machine learning technique. *Agriculture*. 2022;12(8):1226.
- [36] Khan H, Hussain T, Khan SU, Khan ZA, Baik SW. Deep multi-scale pyramidal features network for supervised video summarization. *Expert Syst Appl*. 2024;237:121288.
- [37] Burden FR, Winkler DA. Relevance vector machines: sparse classification methods for QSAR. *J Chem Inform Model*. 2015;55(8):1529–34.
- [38] Le TC, Winkler DA. Discovery and optimization of materials using evolutionary approaches. *Chem Rev*. 2016;116(10):6107–32.
- [39] Fourches D, Pu D, Li L, Zhou H, Mu Q, Su G, et al. Computer-aided design of carbon nanotubes with the desired bioactivity and safety profiles. *Nanotoxicology*. 2016;10(3):374–83.
- [40] Reyes KG, Maruyama B. The machine learning revolution in materials? *MRS Bulletin*. 2019;44(7):530–7.
- [41] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.
- [42] Liu R, Madore M, Glover KP, Feasel MG, Wallqvist A. Assessing deep and shallow learning methods for quantitative prediction of acute chemical toxicity. *Toxicol Sci*. 2018;164(2):512–26.
- [43] Anand A, Bedi J, Aggarwal A, Khan MA, Rida I. Authenticating and securing healthcare records: A deep learning-based zero watermarking approach. *Image Vision Comput*. 2024;145:104975.
- [44] Munsif M, Khan SU, Khan N, Baik SW. Attention-based deep learning framework for action recognition in a dark environment. *Hum Centric Comput Inf Sci*. 2024;14:1–22.
- [45] Hwang B, Kim J, Lee S, Kim E, Kim J, Jung Y, et al. Automatic detection and segmentation of thrombi in abdominal aortic aneurysms using a mask region-based convolutional neural network with optimized loss functions. *Sensors*. 2022;22(10):3643.
- [46] Albalawi E, Thakur A, Dorai DR, Bhatia Khan S, Mahesh T, Almusharraf A, et al. Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach. *Front Comput Neurosci*. 2024;18:1418546.
- [47] Salim MM, Singh SK, Park JH. Securing smart cities using LSTM algorithm and lightweight containers against botnet attacks. *Appl Soft Comput*. 2021;113:107859.
- [48] Karbhari Y, Basu A, Geem ZW, Han GT, Sarkar R. Generation of synthetic chest X-ray images and detection of COVID-19: A deep learning based approach. *Diagnostics*. 2021;11(5):895.
- [49] Jeon S, Seo JT. A synthetic time-series generation using a variational recurrent autoencoder with an attention mechanism in an industrial control system. *Sensors*. 2023;24(1):128.
- [50] Bahl A, Hellack B, Balas M, Dinischiotu A, Wiemann M, Brinkmann J, et al. Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*. 2019;15:100179.
- [51] Gajewicz A, Puzyn T, Odziomek K, Urbaszek P, Haase A, Riebeling C, et al. Decision tree models to classify nanomaterials according to the DF4nanoGrouping scheme. *Nanotoxicology*. 2018;12(1):1–17.
- [52] Lamon L, Asturiol D, Vilchez A, Ruperez-Illescas R, Cabellos J, Richarz A, et al. Computational models for the assessment of manufactured nanomaterials: Development of model reporting standards and mapping of the model landscape. *Comput Toxicol*. 2019;9:143–51.
- [53] Papadiamantis AG, Jänes J, Voyiatzis E, Sikk L, Burk J, Burk P, et al. Predicting cytotoxicity of metal oxide nanoparticles using Isalos analytics platform. *Nanomaterials*. 2020;10(10):2017.
- [54] Rovida C, Alépée N, Api AM, Basketter DA, y Bois F, Caloni F, et al. Integrated testing strategies (ITS) for safety assessment. *ALTEX-Alternatives Animal Experiment*. 2015;32(1):25–40.
- [55] Barabási AL. The network takeover. *Nature Phys*. 2012;8(1):14–6.
- [56] Jeong J, Choi J. Development of AOP relevant to microplastics based on toxicity mechanisms of chemical additives using ToxCast™ and deep learning models combined approach. *Environ Int*. 2020;137:105557.
- [57] Peng T, Wei C, Yu F, Xu J, Zhou Q, Shi T, et al. Predicting nanotoxicity by an integrated machine learning and metabolomics approach. *Environ Pollut*. 2020;267:115434.
- [58] Vishnoi S, Matre H, Garg P, Pandey SK. Artificial intelligence and machine learning for protein toxicity prediction using proteomics data. *Chem Biol Drug Design*. 2020;96(3):902–20.

- [59] Ciaramella A, Staiano A. On the role of clustering and visualization techniques in gene microarray data. *Algorithms*. 2019;12(6):123.
- [60] David R. The promise of toxicogenomics for genetic toxicology: past, present and future. *Mutagenesis*. 2020;35(2):153–9.
- [61] Ewald JD, Soufan O, Crump D, Hecker M, Xia J, Basu N. EcoToxModules: custom gene sets to organize and analyze toxicogenomics data from ecological species. *Environ Sci Technol*. 2020;54(7):4376–87.
- [62] Soufan O, Ewald J, Viau C, Crump D, Hecker M, Basu N, et al. T1000: a reduced gene set prioritized for toxicogenomic studies. *PeerJ*. 2019;7:e7975.
- [63] Silva JCF, Teixeira RM, Silva FF, Brommonschenkel SH, Fontes EP. Machine learning approaches and their current application in plant molecular biology: A systematic review. *Plant Sci*. 2019;284:37–47.
- [64] Federico A, Serra A, Ha MK, Kohonen P, Choi JS, Liampa I, et al. Transcriptomics in toxicogenomics, part II: preprocessing and differential expression analysis for high quality data. *Nanomaterials*. 2020;10(5):903.
- [65] Hemmerich J, Ecker GF. In silico toxicology: From structure-activity relationships towards deep learning and adverse outcome pathways. *Wiley Interdiscipl Rev Comput Mol Sci*. 2020;10(4):e1475.
- [66] Kinaret PAS, Serra A, Federico A, Kohonen P, Nymark P, Liampa I, et al. Transcriptomics in toxicogenomics, part I: experimental design, technologies, publicly available data, and regulatory aspects. *Nanomaterials*. 2020;10(4):750.
- [67] Lazarovits J, Sindhvani S, Tavares AJ, Zhang Y, Song F, Audet J, et al. Supervised learning and mass spectrometry predicts the in vivo fate of nanomaterials. *ACS Nano*. 2019;13(7):8023–34.
- [68] Lopatkin AJ, Collins JJ. Predictive biology: modelling, understanding and harnessing microbial complexity. *Nature Rev Microbiol*. 2020;18(9):507–20.
- [69] Chen L, Meng X, Gu J, Fan W, Abdlli N, Peparh FA, et al. Silver nanoparticle toxicity in silkworms: Omics technologies for a mechanistic understanding. *Ecotoxicol Environ Safety*. 2019;172:388–95.
- [70] Halder AK, Melo A, Cordeiro MND. A unified in silico model based on perturbation theory for assessing the genotoxicity of metal oxide nanoparticles. *Chemosphere*. 2020;244:125489.
- [71] Krewski D, Andersen ME, Tyshenko MG, Krishnan K, Hartung T, Boekelheide K, et al. Toxicity testing in the 21st century: progress in the past decade and future perspectives. *Archiv Toxicol*. 2020;94:1–58.
- [72] Liu Z, Huang R, Roberts R, Tong W. Toxicogenomics: a 2020 vision. *Trends Pharmacol Sci*. 2019;40(2):92–103.
- [73] Scala G, Federico A, Fortino V, Greco D, Majello B. Knowledge generation with rule induction in cancer omics. *Int J Mol Sci*. 2019;21(1):18.
- [74] Cordier T, Lanzén A, Apothéloz-Perret-Gentil L, Stoeck T, Pawlowski J. Embracing environmental genomics and machine learning for routine biomonitoring. *Trends Microbiol*. 2019;27(5):387–97.
- [75] Cordier T, Alonso-Sáez L, Apothéloz-Perret-Gentil L, Aylagas E, Bohan DA, Bouchez A, et al. Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. *Mol Ecol*. 2021;30(13):2937–58.
- [76] Pencheva I, Esteve M, Mikhaylov SJ. Big data and AI-A transformational shift for government: So, what next for research? *Public Policy Administration*. 2020;35(1):24–44.
- [77] Selmani A, Ulm L, Kasemets K, Kurvet I, Erceg I, Barbir R, et al. Stability and toxicity of differently coated selenium nanoparticles under model environmental exposure settings. *Chemosphere*. 2020;250:126265.
- [78] Ma Y, Wang J, Wu J, Tong C, Zhang T. Meta-analysis of cellular toxicity for graphene via data-mining the literature and machine learning. *Sci Total Environ*. 2021;793:148532.
- [79] Poulsen KM, Pho T, Champion JA, Payne CK. Automation and low-cost proteomics for characterization of the protein corona: experimental methods for big data. *Analytic Bioanalytic Chem*. 2020;412:6543–51.
- [80] Bilal M, Oh E, Liu R, Breger JC, Medintz IL, Cohen Y. Bayesian network resource for meta-analysis: cellular toxicity of quantum dots. *Small*. 2019;15(34):1900510.
- [81] Cao J, Pan Y, Jiang Y, Qi R, Yuan B, Jia Z, et al. Computer-aided nanotoxicology: risk assessment of metal oxide nanoparticles via nano-QSAR. *Green Chem*. 2020;22(11):3512–21.
- [82] Price E, Gesquiere AJ. An in vitro assay and artificial intelligence approach to determine rate constants of nanomaterial-cell interactions. *Sci Rep*. 2019;9(1):13943.
- [83] Oktay AB, Gurses A. Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images. *Micron*. 2019;120:113–9.
- [84] Rajesh K, Lalitha Bhaskari D. Automatic image annotation: A review of recent advances and literature. In: *Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics*. Vol. 2. Springer; 2020. p. 271–81.
- [85] Reker D, Bernardes GJ, Rodrigues T. Computational advances in combating colloidal aggregation in drug discovery. *Nature Chem*. 2019;11(5):402–18.
- [86] Varsou DD, Afantitis A, Tsoumanis A, Melagraki G, Sarimveis H, Valsami-Jones E, et al. A safe-by-design tool for functionalised nanomaterials through the Enalos Nanoinformatics Cloud platform. *Nanoscale Adv*. 2019;1(2):706–18.
- [87] Ban Z, Yuan P, Yuuuu F, Peng T, Zhou Q, HUUU X. Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles. *Proc Nat Acad Sci*. 2020;117(19):10492–9.
- [88] Findlay MR, Freitas DN, Mobed-Miremadi M, Wheeler KE. Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. *Environ Sci Nano*. 2018;5(1):64–71.
- [89] Karatzas P, Melagraki G, Ellis LJA, Lynch I, Varsou DD, Afantitis A, et al. Development of deep learning models for predicting the effects of exposure to engineered nanomaterials on *Daphnia magna*. *Small*. 2020;16(36):2001080.
- [90] To KT, Truong L, Edwards S, Tanguay RL, Reif DM. Multivariate modelling of engineered nanomaterial features associated with developmental toxicity. *NanoImpact*. 2019;16:100185.
- [91] Gomes SI, Amorim MJ, Pokhrel S, Mädler L, Fasano M, Chiavazzo E, et al. Machine learning and materials modelling interpretation of in vivo toxicological response to tio 2 nanoparticles library (uv and non-uv exposure). *Nanoscale*. 2021;13(35):14666–78.
- [92] Takata M, Lin BL, Xue M, Zushi Y, Terada A, Hosomi M. Predicting the acute ecotoxicity of chemical substances by machine learning using graph theory. *Chemosphere*. 2020;238:124604.
- [93] Ji CW, Park YS, Cui Y, Wang H, Kwak IS, Chon TS. Analyzing the response behaviour of *Lumbricus variegatus* (Oligochaeta:

- Lumbriculidae) to different concentrations of copper sulfate based on line body shape detection and a recurrent self-organizing map. *Int J Environ Res Public Health*. 2020;17(8):2627.
- [94] Miller JW, Paul MJ, Obenour DR. Assessing potential anthropogenic drivers of ecological health in Piedmont streams through hierarchical modelling. *Freshwater Sci*. 2019;38(4):771–89.
- [95] Hou P, Jolliet O, Zhu J, Xu M. Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. *Environ Int*. 2020;135:105393.
- [96] Gobeyn S, Mouton AM, Cord AF, Kaim A, Volk M, Goethals PL. Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecol Model*. 2019;392:179–95.
- [97] Shadrin D, Pukalchik M, Kovaleva E, Fedorov M. Artificial intelligence models to predict acute phytotoxicity in petroleum contaminated soils. *Ecotoxicol Environ Safety*. 2020;194:110410.
- [98] Perez-Alonso D, Pennnna-Tejedor S, Navarro M, Rad C, Arnaiz-González A, Diiiiiiez-Pastor JF. Decision Trees for the prediction of environmental and agronomic effects of the use of Compost of Sewage Sludge (CSS). *Sustainable Production Consumption*. 2017;12:119–33.
- [99] Parashar A, Parashar A, Abate AF, Shekhawat RS, Rida I. Real-time gait biometrics for surveillance applications: A review. *Image Vision Comput*. 2023;138:104784.
- [100] Wigger H, Kägi R, Wiesner M, Nowack B. Exposure and possible risks of engineered nanomaterials in the environment-Current knowledge and directions for the future. *Rev Geophys*. 2020;58(4):e2020RG000710.
- [101] Wikle CK. Comparison of deep neural networks and deep hierarchical models for spatio-temporal data. *J Agricult Biologic Environ Stat*. 2019;24(2):175–203.
- [102] Goldberg E, McNew C, Scheringer M, Bucheli TD, Nelson P, Hungerbuhler K. What factors determine the retention behaviour of engineered nanomaterials in saturated porous media? *Environ Sci Technol*. 2017;51(5):2729–37.
- [103] Babakhani P, Bridge J, Doong Ra, Phenrat T. Parameterization and prediction of nanoparticle transport in porous media: A reanalysis using artificial neural network. *Water Resour Res*. 2017;53(6):4564–85.
- [104] Peijnenburg W, Praetorius A, Scott-Fordsmand J, Cornelis G. Fate assessment of engineered nanoparticles in solids dominated media-Current insights and the way forward. *Environ Pollution*. 2016;218:1365–9.
- [105] Sangani MF, Owens G, Nazari B, Astaraei A, Fotovat A, Emami H. Different modelling approaches for predicting titanium dioxide nanoparticles mobility in intact soil media. *Sci Total Environ*. 2019;665:1168–81.
- [106] Pinheiro HSK, Carvalho Wd, Chagas CdS, Anjos LHCD, Owens PR. Prediction of topsoil texture through regression trees and multiple linear regressions. *Revista Brasileira de Cieceencia do Solo*. 2018;42:e0170167.
- [107] Wang Z, Lai C, Chen X, Yang B, Zhao S, Bai X. Flood hazard risk assessment model based on random forest. *J Hydrol*. 2015;527:1130–41.
- [108] Olawoyin R, Nieto A, Grayson RL, Hardisty F, Oyewole S. Application of artificial neural network (ANN)-self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. *Expert Syst Appl*. 2013;40(9):3634–48.
- [109] Pikula K, Zakharenko A, Chaika V, Kirichenko K, Tsatsakis A, Golokhvast K. Risk assessments in nanotoxicology: Bioinformatics and computational approaches. *Current Opinion Toxicol*. 2020;19:1–6.
- [110] Aziz S, Dowling M. Machine learning and AI for risk management. New York: Springer International Publishing; 2019.
- [111] Thomas RS, Bahadori T, Buckley TJ, Cowden J, Deisenroth C, Dionisio KL, et al. The next generation blueprint of computational toxicology at the US environmental protection agency. *Toxicol Sci*. 2019;169(2):317–32.
- [112] Boulesteix AL, Schmid M. Machine learning versus statistical modelling. *Biometric J*. 2014;56(4):588–93.
- [113] Greenland S. Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values. *Amer Stat*. 2019;73(sup1):106–14.
- [114] Hicks DJ, Magnus P, Wright J. Inductive risk, science, and values: A reply to MacGillivray. *Risk Anal*. 2020;40(4):667–73.
- [115] MacGillivray BH. Null hypothesis testing ≠ scientific inference: A critique of the shaky premise at the heart of the science and values debate, and a defense of value-neutral risk assessment. *Risk Anal*. 2019;39(7):1520–32.
- [116] Richter AN, Khoshgoftaar TM. A review of statistical and machine learning methods for modelling cancer risk using structured clinical data. *Artif Intell Med*. 2018;90:1–14.
- [117] Teschendorff AE. Avoiding common pitfalls in machine learning omic data science. *Nat Mater*. 2019;18(5):422–7.
- [118] Ij H. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233.
- [119] Bioglio L, Rho V, Pensa RG. Ranking by inspiration: a network science approach. *Mach Learn*. 2020;109(6):1205–29.
- [120] Thompson WH, Wright J, Bissett PG, Poldrack RA. Dataset decay and the problem of sequential analyses on open datasets. *Elife*. 2020;9:e53498.
- [121] Xu C, Jackson SA. Machine learning and complex biological data. Ireland: Springer Nature; 2019.
- [122] Bauer P, Dueben PD, Hoefler T, Quintino T, Schulthess TC, Wedi NP. The digital revolution of Earth-system science. *Nature Comput Sci*. 2021;1(2):104–13.
- [123] Garduño-Jiménez AL, Durán-Álvarez JC, Gomes RL. Meta-analysis and machine learning to explore soil-water partitioning of common pharmaceuticals. *Sci Total Environ*. 2022;837:155675.
- [124] Chapman J, Truong VK, Elbourne A, Gangadoo S, Cheeseman S, Rajapaksha P, et al. Combining chemometrics and sensors: Toward new applications in monitoring and environmental analysis. *Chem Rev*. 2020;120(13):6048–69.
- [125] Cardoso P, Branco VV, Borges PA, Carvalho JC, Rigal F, Gabriel R, et al. Automated discovery of relationships, models, and principles in ecology. *Front Ecol Evolut*. 2020;8:530135.
- [126] Forest V. Biological significance of the nanoparticles protein corona. In: *Nanoparticle-Protein Corona*. UK: Royal Society of Chemistry; 2019. p. 31–60.
- [127] Neftci EO, Averbeck BB. Reinforcement learning in artificial and biological systems. *Nature Machine Intell*. 2019;1(3):133–43.
- [128] Hino M, Benami E, Brooks N. Machine learning for environmental monitoring. *Nat Sustain*. 2018;1(10):583–8.
- [129] Duan Y, Coreas R, Liu Y, Bitounis D, Zhang Z, Parviz D, et al. Prediction of protein corona on nanomaterials by machine learning using novel descriptors. *NanoImpact*. 2020;17:100207.
- [130] Forest V, Hochepeid JF, Pourchez J. Importance of choosing relevant biological end points to predict nanoparticle toxicity with computational approaches for human health risk assessment. *Chem Res Toxicol*. 2019;32(7):1320–6.
- [131] Sizochenko N, Syzochenko M, Fjodorova N, Rasulev B, Leszczynski J. Evaluating genotoxicity of metal oxide

- nanoparticles: Application of advanced supervised and unsupervised machine learning techniques. *Ecotoxicol Environ Safety*. 2019;185:109733.
- [132] Chen C, Zuo Y, Ye W, Li X, Deng Z, Ong SP. A critical review of machine learning of energy materials. *Adv Energy Materials*. 2020;10(8):1903242.
- [133] Horton M, Dwaraknath S, Persson K. Promises and perils of computational materials databases. *Nature Comput Sci*. 2021;1(1):3–5.
- [134] Reyna A, Martín C, Chen J, Soler E, Díaz M. On blockchain and its integration with IoT. Challenges and opportunities. *Future Generation Comput Syst*. 2018;88:173–90.
- [135] Jameel SM, Hashmani MA, Rehman M, Budiman A. An adaptive deep learning framework for dynamic image classification in the internet of things environment. *Sensors*. 2020;20(20):5811.
- [136] Kour VP, Arora S. Recent developments of the internet of things in agriculture: a survey. *IEEE Access*. 2020;8:129924–57.
- [137] Rivero AJL, Alayón CAM, Ferro R, de la Iglesia DH, Secades VA. Network traffic modelling in a wi-fi system with intelligent soil moisture sensors (Wsn) using iot applications for potato crops using arima and sarima time series. *Appl Sci (Switzerland)*. 2020;10(21):1–19.
- [138] Zhu Z, Ng DWH, Park HS, McAlpine MC. 3D-printed multifunctional materials enabled by artificial-intelligence-assisted fabrication technologies. *Nature Rev Materials*. 2021;6(1):27–47.
- [139] Scheibner J, Jobin A, Vayena E. Ethical issues with using Internet of Things devices in citizen science research: a scoping review. *Front Environ Sci*. 2021;9:629649.
- [140] Cortet J, Kocev D, Ducobu C, Džeroski S, Debeljak M, Schwartz C. Using data mining to predict soil quality after application of biosolids in agriculture. *J Environ Quality*. 2011;40(6):1972–82.
- [141] Ahamed N. Ecotoxicity concert of nano zero-valent iron particles-a review. *J Crit Rev*. 2014;1(1):36–39.
- [142] Fourches D, Lougee R. Quantitative nanostructure-activity relationships: methods, case studies, and perspectives. *Bioactivity of Engineered Nanoparticles*. Singapore: Springer; 2017. p. 361–76.
- [143] Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari TP, Michalkova A, et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnol*. 2011;6(3):175–8.
- [144] Epa VC, Burden FR, Tassa C, Weissleder R, Shaw S, Winkler DA. Modeling biological activities of nanoparticles. *Nano Letters*. 2012;12(11):5808–12.
- [145] Fourches D, Pu D, Tassa C, Weissleder R, Shaw SY, Mumper RJ, et al. Quantitative nanostructure-activity relationship modelling. *ACS Nano*. 2010;4(10):5703–12.
- [146] Liu R, Rallo R, Weissleder R, Tassa C, Shaw S, Cohen Y. Nano-SAR development for bioactivity of nanoparticles with considerations of decision boundaries. *Small*. 2013;9(9–10):1842–52.
- [147] Papa E, Doucet J, Doucet-Panaye A. Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. *SAR QSAR Environ Res*. 2015;26(7–9):647–65.
- [148] Chen G, Peijnenburg WJ, Kovalishyn V, Vijver MG. Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. *RSC Adv*. 2016;6(57):52227–35.
- [149] Gernand JM, Casman EA. A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. *Risk Anal*. 2014;34(3):583–97.
- [150] Mikolajczyk A, Gajewicz A, Rasulev B, Schaeublin N, Maurer-Gardner E, Hussain S, et al. Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. *Chem Materials*. 2015;27(7):2400–7.
- [151] Bai X, Liu F, Liu Y, Li C, Wang S, Zhou H, et al. Toward a systematic exploration of nano-bio interactions. *Toxicol Appl Pharmacol*. 2017;323:66–73.
- [152] Le TC, Yin H, Chen R, Chen Y, Zhao L, Casey PS, et al. An experimental and computational approach to the development of ZnO nanoparticles that are safe by design. *Small*. 2016;12(26):3568–77.
- [153] Oksel C, Winkler DA, Ma CY, Wilkins T, Wang XZ. Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology*. 2016;10(7):1001–12.
- [154] Concu R, Kleandrova VV, Speck-Planche A, Cordeiro MND. Probing the toxicity of nanoparticles: a unified in silico machine learning model based on perturbation theory. *Nanotoxicology*. 2017;11(7):891–906.
- [155] Wang W, Sedykh A, Sun H, Zhao L, Russo DP, Zhou H, et al. Predicting nano-bio interactions by integrating nanoparticle libraries and quantitative nanostructure activity relationship modelling. *ACS Nano*. 2017;11(12):12641–9.
- [156] Kovalishyn V, Abramenko N, Kopernyk I, Charochkina L, Metelytsia L, Tetko IV, et al. Modelling the toxicity of a large set of metal and metal oxide nanoparticles using the OCHEM platform. *Food Chem Toxicol*. 2018;112:507–17.
- [157] Hataminia F, Noroozi Z, Eslam HM. Investigation of iron oxide nanoparticle cytotoxicity in relation to kidney cells: A mathematical modelling of data mining. *Toxicol in Vitro*. 2019;59:197–203.
- [158] Horwath JP, Zakharov DN, Megret R, Stach EA. Understanding important features of deep learning models for transmission electron microscopy image segmentation. 2019. arXiv: <http://arXiv.org/abs/arXiv:191206077>.
- [159] Coquelin L, Fischer N, Feltn N, Devoille L, Felhi G. Towards the use of deep generative models for the characterization in size of aggregated TiO₂ nanoparticles measured by Scanning Electron Microscopy (SEM). *Materials Res Express*. 2019;6(8):085001.
- [160] Rusk N. Deep learning. *Nat Methods*. 2016;13(1):35–35.
- [161] Ilet M, Wills J, Rees P, Sharma S, Micklethwaite S, Brown A, et al. Application of automated electron microscopy imaging and machine learning to characterise and quantify nanoparticle dispersion in aqueous media. *J Microscopy*. 2020;279(3):177–84.
- [162] Winkler DA, Mombelli E, Pietrousti A, Tran L, Worth A, Fadeel B, et al. Applying quantitative structure-activity relationship approaches to nanotoxicology: current status and future potential. *Toxicology*. 2013;313(1):15–23.
- [163] Shatkin JA. *The future in nanosafety*. Washington, DC: ACS Publications; 2020.
- [164] Stueckle TA, Roberts JR. *Perspective on current alternatives in nanotoxicology research*. New York: Mary Ann Liebert, Inc., Publishers 140 Huguenot Street, 3rd Floor; 2019.
- [165] Winkler DA. Issues for and examples of computational design of “safe-by-design” nanomaterials. In: *Computational Nanotoxicology*. London: Jenny Stanford Publishing; 2019. p. 507–33.
- [166] Worth A, Aschberger K, Asturiol D, Bessems J, Gerloff K, Graepel R, et al. Evaluation of the availability and applicability of computational approaches in the safety assessment of nanomaterials.

- Publications Office of the European Union, Luxembourg. 2017;10:248139.
- [167] Helma C, Rautenberg M, Gebele D. Nano-lazar: read across predictions for nanoparticle toxicities with calculated and measured properties. *Front Pharmacol.* 2017;8:377.
- [168] Baeck SH, Jaramillo TF, Kleiman-Shwarscstein A, McFarland EW. Automated electrochemical synthesis and characterization of TiO₂ supported Au nanoparticle electrocatalysts. *Measurement Sci Technol.* 2004;16(1):54.
- [169] Gleiter H. Nanostructured materials: basic concepts and microstructure. *Acta Materialia.* 2000;48(1):1–29.
- [170] Hiszpanski AM, Gallagher B, Chellappan K, Li P, Liu S, Kim H, et al. Nanomaterial synthesis insights from machine learning of scientific articles by extracting, structuring, and visualizing knowledge. *J Chem Inform Model.* 2020;60(6):2876–87.
- [171] Wang A, Marinakos SM, Badireddy AR, M. Powers C, A. Houck K. Characterization of physicochemical properties of nanomaterials and their immediate environments in high-throughput screening of nanomaterial biological activity. *Wiley Interdiscipl Rev Nanomed Nanobiotech.* 2013;5(5):430–48.
- [172] Collins AR, Annangi B, Rubio L, Marcos R, Dorn M, Merker C, et al. High throughput toxicity screening and intracellular detection of nanomaterials. *Wiley Interdiscipl Rev Nanomed Nanobiotech.* 2017;9(1):e1413.
- [173] Qiu TA, Nguyen THT, Hudson-Smith NV, Clement PL, Forester DC, Frew H, et al. Growth-based bacterial viability assay for interference-free and high-throughput toxicity screening of nanomaterials. *Analytic Chem.* 2017;89(3):2057–64.
- [174] Sims KR, Maceren JP, Strand AI, He B, Overby C, Benoit DS. Rigor and reproducibility in polymer nanoparticle synthesis and characterization. *RSC Adv.* 2020;10(5):2513–8.
- [175] Núñez RN, Veglia AV, Pacioni NL. Improving reproducibility between batches of silver nanoparticles using an experimental design approach. *Microchem J.* 2018;141:110–7.
- [176] Liz-Marzán LM, Kagan CR, Millstone JE. Reproducibility in nanocrystal synthesis? Watch out for impurities!. Washington, DC: ACS Publications; 2020.
- [177] Cockram AA, Bradley RD, Lynch SA, Fleming PC, Williams NS, Murray MW, et al. Optimization of the high-throughput synthesis of multiblock copolymer nanoparticles in aqueous media via polymerization-induced self-assembly. *Reaction Chem Eng.* 2018;3(5):645–57.
- [178] Petersen EJ, Hirsch C, Elliott JT, Krug HF, Aengenheister L, Arif AT, et al. Cause-and-effect analysis as a tool to improve the reproducibility of nanobioassays: four case studies. *Chem Res Toxicol.* 2019;33(5):1039–54.
- [179] Baer DR, Munusamy P, Thrall BD. Provenance information as a tool for addressing engineered nanoparticle reproducibility challenges. *Biointerphases.* 2016;11(4):04B401.
- [180] Galmarini S, Hanusch U, Giraud M, Cayla N, Chiappe D, von Moos N, et al. Beyond unpredictability: the importance of reproducibility in understanding the protein corona of nanoparticles. *Bioconjugate Chem.* 2018;29(10):3385–93.
- [181] Basant N, Gupta S. Multi-target QSTR modelling for simultaneous prediction of multiple toxicity endpoints of nano-metal oxides. *Nanotoxicology.* 2017;11(3):339–50.
- [182] Zeng L, Gao J, Liu Y, Gao J, Yao L, Yang X, et al. Role of protein corona in the biological effect of nanomaterials: Investigating methods. *TrAC Trends Analytic Chem.* 2019;118:303–14.
- [183] Xu L, Xu M, Wang R, Yin Y, Lynch I, Liu S. The crucial role of environmental coronas in determining the biological effects of engineered nanomaterials. *Small.* 2020;16(36):2003691.
- [184] Wills JW, Summers HD, Hondow N, Soorash A, Meissner KE, White PA, et al. Characterizing nanoparticles in biological matrices: tipping points in agglomeration state and cellular delivery in vitro. *ACS Nano.* 2017;11(12):11986–2000.
- [185] Shannahan J. The biocorona: a challenge for the biomedical application of nanoparticles. *Nanotechnol Reviews.* 2017;6(4):345–53.
- [186] Tiwari A, Bhatia P, Randhawa JK. Systematic spectroscopic investigation of structural changes and corona formation of bovine serum albumin over magneto-fluorescent nanoparticles. *Dalton Trans.* 2020;49(35):12380–9.
- [187] Szekeres GP, Fernández-Iglesias N, Kneipp J, Montes-Bayón M, Bettmer J. Mass spectrometric approach for the analysis of the hard protein corona of nanoparticles in living cells. *J Proteomics.* 2020;212:103582.
- [188] Shah J, Singh S. Nanoparticle-protein corona complex: composition, kinetics, physico-chemical characterization, and impact on biomedical applications. UK: Royal Society of Chemistry; 2019.
- [189] Pinals RL, Yang D, Rosenberg DJ, Chaudhary T, Crothers AR, Iavarone AT, et al. Quantitative protein corona composition and dynamics on carbon nanotubes in biological environments. *Angewandte Chem Int Edn.* 2020;59(52):23668–77.
- [190] Patel P, Kumar A. Factors affecting a nanoparticle's protein corona formation. In: *Nanoparticle-Protein Corona.* UK: Royal Society of Chemistry; 2019. p. 61–79.
- [191] Pareek V, Bhargava A, Bhanot V, Gupta R, Jain N, Panwar J. Formation and characterization of protein corona around nanoparticles: a review. *J Nanosci Nanotech.* 2018;18(10):6653–70.
- [192] Panchal D, Patel R, Siddheshwari M, Rahaman E, Patel V, Karakoti AS. NP-protein Corona interaction: characterization methods and analysis. UK: Royal Society of Chemistry; 2019.
- [193] Ovais M, Nethi SK, Ullah S, Ahmad I, Mukherjee S, Chen C. Recent advances in the analysis of nanoparticle-protein coronas. *Nanomedicine.* 2020;15(10):1037–61.
- [194] Mohammad-Beigi H, Hayashi Y, Zeuthen CM, Eskandari H, Scavenius C, Juul-Madsen K, et al. Mapping and identification of soft corona proteins at nanoparticles and their impact on cellular association. *Nature Commun.* 2020;11(1):4535.
- [195] Yin MM, Chen WQ, Luuuu YQ, Han JY, Liu Y, Jiang FL. A model beyond protein corona: thermodynamics and binding stoichiometries of the interactions between ultrasmall gold nanoclusters and proteins. *Nanoscale.* 2020;12(7):4573–85.
- [196] Wheeler K, Findlay M, Freitas D, Mobed-Miremadi M, Eramo S, Baumgartner K. Toward predictive analysis of nanoparticle protein corona populations. In: *Abstracts of Papers of the American Chemical Society.* vol. 258. American Chemical Society 1155 16th st, NW, Washington, DC 20036 USA; 2019.
- [197] Baweja L. Computer simulations for understanding nanoparticle-biomolecule corona formation. Vol. 40. London; Royal Society of Chemistry; 2019.
- [198] Alberg I, Kramer S, Schinnerer M, Hu Q, Seidl C, Leps C, et al. Polymeric nanoparticles with neglectable protein corona. *Small.* 2020;16(18):1907574.
- [199] Laux P, Riebeling C, Booth AM, Brain JD, Brunner J, Cerrillo C, et al. Biokinetics of nanomaterials: The role of biopersistence. *NanoImpact.* 2017;6:69–80.

- [200] Wyrzykowska E, Jagiello K, Rasulev B, Puzyn T. Descriptors in nano-QSAR/QSPR modelling. In: *Computational Nanotoxicology*. London: Jenny Stanford Publishing; 2019. p. 245–302.
- [201] Mikulskis P, Alexander MR, Winkler DA. Toward interpretable machine learning models for materials discovery. *Adv Intell Syst*. 2019;1(8):1900045.
- [202] Oksef C, Ma CY, Wang XZ. Structure-activity relationship models for hazard assessment and risk management of engineered nanomaterials. *Proc Eng*. 2015;102:1500–10.
- [203] Karelson M. *Molecular descriptors in QSAR/QSPR*. 2000.
- [204] Russo DP, Yan X, Shende S, Huang H, Yan B, Zhu H. Virtual molecular projections and convolutional neural networks for the end-to-end modelling of nanoparticle activities and properties. *Analytic Chem*. 2020;92(20):13971–9.
- [205] De P, Kar S, Roy K, Leszczynski J. Second generation periodic table-based descriptors to encode toxicity of metal oxide nanoparticles to multiple species: QSTR modelling for exploration of toxicity mechanisms. *Environ Sci Nano*. 2018;5(11):2742–60.
- [206] Sizochenko N, Rasulev B, Gajewicz A, Kuz'min V, Puzyn T, Leszczynski J. From basic physics to mechanisms of toxicity: The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. *Nanoscale*. 2014;6(22):13986–93.
- [207] DiFranzo A, Sheridan RP, Liaw A, Tudor M. Nearest neighbor gaussian process for quantitative structure-activity relationships. *J Chem Inform Model*. 2020;60(10):4653–63.
- [208] Winkler DA. Sparse QSAR modelling methods for therapeutic and regenerative medicine. *J Comput-Aid Mol Design*. 2018;32:497–509.
- [209] Saquib Q, Xia P, Siddiqui MA, Zhang J, Xie Y, Faisal M, et al. High-throughput transcriptomics: An insight on the pathways affected in HepG2 cells exposed to nickel oxide nanoparticles. *Chemosphere*. 2020;244:125488.
- [210] Martins C, Dreij K, Costa PM. The state-of-the art of environmental toxicogenomics: challenges and perspectives of “omics” approaches directed to toxicant mixtures. *Int J Environ Res Public Health*. 2019;16(23):4718.
- [211] Furxhi I, Murphy F, Poland CA, Sheehan B, Mullins M, Mantecca P. Application of Bayesian networks in determining nanoparticle-induced cellular outcomes using transcriptomics. *Nanotoxicology*. 2019;13(6):827–48.
- [212] Rusyn I, Sedykh A, Low Y, Guyton KZ, Tropsha A. Predictive modelling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol Sci*. 2012;127(1):1–9.
- [213] Lee S, Park K, Ahn HS, Kim D. Importance of structural information in predicting human acute toxicity from in vitro cytotoxicity data. *Toxicol Appl Pharmacol*. 2010;246(1–2):38–48.
- [214] Wang M, Wang T, Cai P, Chen X. Nanomaterials discovery and design through machine learning. *Small Methods*. 2019;3(5):1900025.
- [215] Wei J, Chu X, Sun XY, Xu K, Deng HX, Chen J, et al. Machine learning in materials science. *InfoMat*. 2019;1(3):338–58.
- [216] Duschl A, Windgasse G. A survey on the state of nanosafety research in the European Union and the United States. *J Nanoparticle Res*. 2018;20:1–5.