# Application of machine learning in nanotoxicology: a critical review and perspective

Zhou, Y.; Wang, Y.; Peijnenburg, W.J.G.M.; Vijver, M.G.; Balraadjsing, S.; Dong, Z.; ... ; Fan, W.

Environmental Science & Technology

pubs.acs.org/est

Critical Review

# Application of Machine Learning in Nanotoxicology: A Critical Review and Perspective

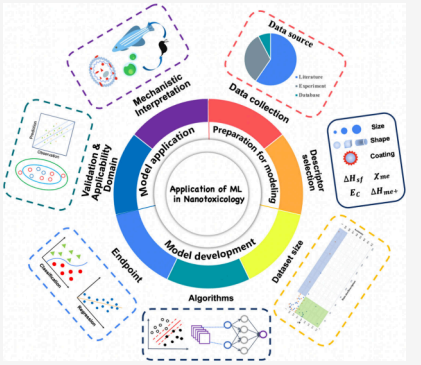
Yunchi Zhou, Ying Wang,* Willie Peijnenburg, Martina G. Vijver, Surendra Balraadjsing, Zhaomin Dong, Xiaoli Zhao, Kenneth M. Y. Leung, Holly M. Mortensen, Zhenyu Wang, Iseult Lynch, Antreas Afantitis, Yunsong Mu, Fengchang Wu, and Wenhong Fan*

Cite This: *Environ. Sci. Technol.* 2024, 58, 14973−14993

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The massive production and application of nanomaterials (NMs) have raised concerns about the potential adverse effects of NMs on human health and the environment. Evaluating the adverse effects of NMs by laboratory methods is expensive, time-consuming, and often fails to keep pace with the invention of new materials. Therefore, *in silico* methods that utilize machine learning techniques to predict the toxicity potentials of NMs are a promising alternative approach if regulatory confidence in them can be enhanced. Previous reviews and regulatory OECD guidance documents have discussed in detail how to build an *in silico* predictive model for NMs. Nevertheless, there is still room for improvement in addressing the ways to enhance the model representativeness and performance from different angles, such as data set curation, descriptor selection, task type (classification/regression), algorithm choice, and model evaluation (internal and external validation, applicability domain, and mechanistic interpretation, which is key to ensuring stakeholder confidence). This review explores how to build better predictive models; the current state of the art is analyzed via a statistical evaluation of literature, while the challenges faced and future perspectives are summarized. Moreover, a recommended workflow and best practices are provided to help in developing more predictive, reliable, and interpretable models that can assist risk assessment as well as safe-by-design development of NMs.




**KEYWORDS:** *nanomaterials, computational toxicity, machine learning, algorithm, classification/regression, prediction*



## 1. INTRODUCTION

Nanomaterials (NMs) have been used in various fields such as medicine, food, electronics and all kinds of technology because of their unique physicochemical properties.[1,2] However, the rapid development of nanotechnology and its wide use in products causes unintended emissions which raised concerns about the potential adverse effects of NMs on human health and the environment. Researchers have found that exposure to specific NMs may cause allergies or neurotoxicity to humans.[3,4] Moreover, NMs released into the aquatic environment may pose risks to aquatic organisms like fish, crustaceans, algae and bacteria.[5−9] The toxic effects of NMs are believed to be related to physicochemical properties such as particle size,[10] shape,[11] surface area, chemical composition and stability.[12] Given the complexity and heterogeneity of NMs, *in vitro* and *in vivo* toxicity assessments require extensive time and resources.[13] This could easily result in the rate at which new advanced NMs are being developed (currently the fourth generation of materials), outpacing safety testing and regulatory assessment of these materials and their products.[14] Therefore, *in silico* methods, which comply with the 3Rs (Replacement, Reduction and Refinement) principles, have been developed to minimize and prioritize experimental testing efforts while ensuring human and environmental safety.

An increasing number of *in silico* methods have been developed over the years to predict the toxic properties and adverse effects of NMs. Quantitative structure−activity relationships (QSARs) and quantitative structure−property relationship (QSPRs) are the most popular *in silico* methods.[15] QSAR and QSPR methods establish relationships between the structures of NMs and their activity or physicochemical properties using mathematical statistics or machine learning (ML) algorithms. Puzyn et al.[16] discussed the role of nano-QSAR methods for predicting the hazards of NMs as early as 2009. After that, the application of *in silico* methods for predicting the toxicity of NMs has made rapid progress in various aspects: (1) The algorithms used in models have

Received: April 3, 2024
Revised: July 24, 2024
Accepted: July 24, 2024
Published: August 7, 2024

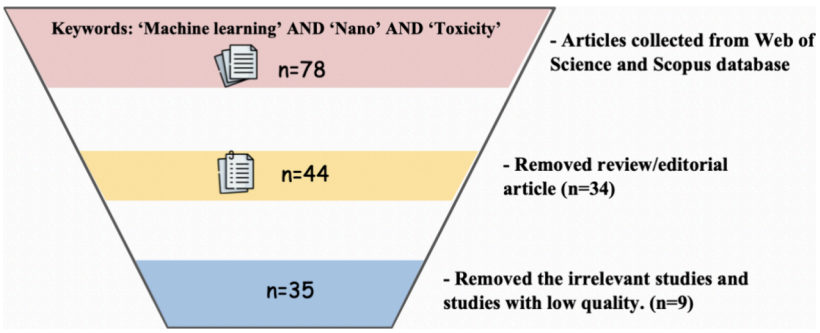



**Figure 1.** Workflow of the literature research undertaken for the present analysis.

evolved from single-descriptor linear models to various nonlinear ML models; (2) The scope of prediction has expanded from metal and metal oxides NMs to include carbon-based NMs and organic NMs.,[17,18] although predictive model development beyond the first generation of NMs is still notoriously challenging; (3) In terms of data sources, databases of significant sizes have been developed,[19] although significantly more effort in strengthening the databases and knowledgebase is needed; (4) the selection of descriptors has advanced from considering only structural parameters and physicochemical properties of NMs to incorporating information on exposure conditions and different species of biota,[20] and to considering whole NM descriptors and descriptors as distributions rather than absolute values.[21]

Another important *in silico* approach risk assessment of NMs is quantitative read-across, a data gap filling technique used on unknown or untested NMs.[22−25] The technique is applied within analog and category approaches. Briefly, an analog approach involves a target and a source substance whereas a category approach includes 2 or more source substances. The field of read-across is extensively described in the peer reviewed literature as well as in technical guidance published by the Organisation of Economic Co-operation and Development (OECD),[26] the European Chemicals Agency (ECHA),[27] and the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC).[28] Machine learning, mainly unsupervised learning (e.g., self-organizing map, clustering) plays an important role in the development of nanoread-across.[29−32] It is also clear that there are still many challenges around the justified use of read across.[33]

As a newly developing robust nonparametric approach, ML empowers computer systems to learn from the data and enhance their performance without being explicitly programmed.[34] ML algorithms are efficient in dealing with heterogeneous data and in finding relationships between complex factors,[35] and thus have been applied to complex problems such as toxicity prediction, environmental risk assessment, water quality assessment and identification of pollution sources.[36] Specifically in the field of nanotoxicology, ML had been applied to predict the cytotoxicity of NMs and was shown to produce efficient and powerful predictive tools, as exemplified by Ma et al.[17] These authors developed models to predict graphene cytotoxicity using random forest (RF) learning algorithms with a good performance, for which the squared correlation coefficient ($R^2$) > 0.8. In another study, an artificial neural network (ANN) was used to build QSAR models to predict the cytotoxicity of metal oxide NMs to *E. coli*[37] or to predict mortality of metal-based NM on *D. magna*.[46] The models performed well in developing both quantitative and categorical models, for which the squared cross validated correlation coefficient ($Q^2$) of leave-one-out validation was >0.8, and the classification accuracy equaled 100%. However, it should be noted that ML models commonly become less interpretable as the complexity of the models increases.[38] Therefore, in addition to research in improving model performance, researchers are also seeking ways to enhance the interpretability of ML nanotoxicity models.[39,40] For example, Yu et al.[41] proposed a feature importance and feature interaction network analysis to interpret the developed RF models for immune responses and the lung burden of NMs, demonstrating mechanistic insights from the ML model.

A number of reviews have summarized the sources of data, modeling techniques, approaches for interpretation of models, existing challenges and future perspectives on application of ML in nanotoxicology. Winkler et al.[42] summarized the existing methods and anticipated the future development of nano-QSAR models. More recently, Furxhi et al.[15] presented the techniques and procedures of existing models that researchers can adopt to assemble their own nanotoxicological *in silico* studies, and reviewed specifically the data used for modeling, the techniques for data preprocessing and the algorithms applied for model implementation.[43] Ji et al.[44] summarized the ML models used for predicting the cytotoxicity of NMs using different algorithms. An overview on the existing databases was provided by Basei et al.[45] and *in silico* models were evaluated using criteria inspired by the Organization for Economic Co-operation and Development (OECD) principles for QSAR model validation.[26] Li et al.[46] reviewed the general procedures for the construction and application of nano-QSAR models of metal-based and metal-oxide NMs and an overview of available databases and common algorithms. More recently, Jia et al.[38] focused on the applications of interpretable ML in computational toxicology, including toxicity feature data, model interpretation methods, use of knowledgebase frameworks and recent applications.[47]

Acknowledging these comprehensive reviews of previous studies, it is evident that ML holds great promise for the future in the field of nanotoxicology. However, existing literature reviews still lack depth in discussing how to enhance the representativeness (in this case: how well the samples or model results represent a larger set of untested NMs) and performance of the developed ML models from various aspects, such as data set size, descriptor selection, endpoint and algorithm selection, and model evaluation. Therefore, we aim to discuss several crucial issues about the application of ML in nanotoxicology based on a critical analysis of existing literature in this review article considering: (i) the curation of data sets, (ii) the selection of descriptors, (iii) approaches for


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

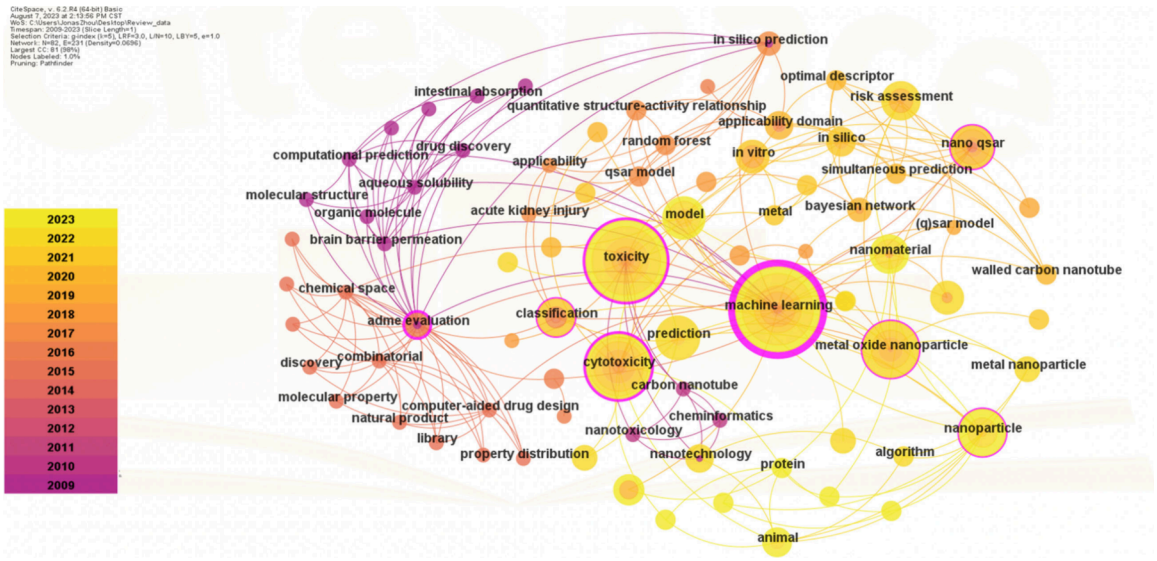



**Figure 2.** Visualization of analysis result of CiteSpace, the nodes represent the keywords of each article, and the lines stand for the relationships between them.

choosing the appropriate type of model task (classification/regression), (iv) selecting appropriate algorithms, and (v) the evaluation of model performance (internal- and external validation, definition of applicability domain and mechanistic interpretation to enhance model utility for use in regulatory decision-making and for safe-by-design of NMs in the context of product design). The current research status is analyzed, and the challenges that have yet to be addressed and overcome are summarized. Moreover, we provided a recommended workflow and best practices, as well as future perspectives for implementation of ML in nanotoxicology.

## 2. LITERATURE SURVEY

In order to provide an overview of the status of the application of ML models in the domain of nanotoxicology, we collected articles that implement ML models to predict the toxicity of NMs, using the Web of Science and Scopus' advanced searches, with the keywords: "Machine Learning" AND "Nano" AND "Toxicity" up to September 2023 (Figure 1).

In total 78 articles were identified after the preliminary literature search. Based on this, CiteSpace (version 6.2.4)[48] was used to analyze the connections between the articles and to visualize the results at an interval of one year. The "keyword" was chosen as the node type, the selection standard was "TOP 5.0%", with 50 being the maximum number of items selected for each interval (Figure 2). The larger the circles in Figure 2, the more research has been done in that area. The nodes are composed of circles of different colors, and the colors of the circles correspond to the years in which the keyword appears. In general, these keywords can be divided into three categories: different type of NMs (i.e., metal oxide nanoparticle, metal nanoparticle, walled carbon nanotube), terms related to ML modeling (i.e., nano-QSAR, classification, random forest), and terms about toxicity experiments (i.e., cytotoxicity, acute kidney injury). The results show that ML application is most often used to build QSAR-based classification models to predict the cytotoxicity of the first generation of NMs like metal oxide nanoparticles (NPs) and metal NPs. From the colors of the nodes in Figure 2, we can see that carbon-based materials, such as carbon nanotubes, have gained attention in recent years. This is because analytical techniques are improving, and also society is making increased use of carbon-based materials, including graphene and graphene oxide.

After that, an examination on title, abstract, and keywords was applied in order to exclude the review and editorial articles. Thereafter, the articles were manually filtered by the following rules, some of which are in accordance with the OECD principles for the creation of QSARs:[26] (I) the article should focus on the implementation of a ML model; (II) the ML model should have a clear data source, an unambiguous algorithm, and a well-defined endpoint; and (III) the model performance should be measured appropriately (at least cross-validation, goodness-of-fit, predictivity). Studies that did not meet any of the aforementioned criteria were excluded. As summarized in Figure 1, 35 articles implementing ML models for nanotoxicity prediction were finally identified and adopted in this study after filtering using this structured approach.

The following aspects of information are the objects of our literature review collection: (1) data set, including data source and volume, (2) descriptor information, such as descriptor(s) used, their classification and importance rank, (3) study design information (e.g., NMs, test organisms, endpoint, algorithms used), (4) post-treatment (model validation and applicability domain), and (5) mechanistic interpretation.

## 3. PREPARATION FOR MODELING: DATA COLLECTION AND DESCRIPTOR SELECTION

### 3.1. Data Collection: Source, Volume, and Quality.

The first step for the creation of ML models is data collection. Depending on the study design, the sources of data mainly include literature, databases, experiments and integration of multiple sources.[49] In total, 52 records were generated for statistical analysis from the articles reviewed (Table S1).

In addition to the collection of data from the published literature, public databases were also mined. Table 1 describes the databases from which data were retrieved. These databases provided data as generated from experiments performed by authors who provided their data to the database operators and data collected from literature. These databases are commonly



https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

Table 1. Databases commonly used for nanotoxicity modeling

| Name | Address | Description |
|---|---|---|
| eNanoMapper | http://search.data.enanomapper.net/ | A Database and Ontology Framework for Nanomaterials Design and Safety Assessment |
| OCHEM[71] | http://ochem.eu/ | A web-based platform that aims to automate and simplify the typical steps required for QSAR modeling |
| NanoReg2 database | https://enanomapper.adma.ai/projects/nanoreg2/ | A set of eNanoMapper database instances and an aggregated search index used for data management |
| Omics database instance | https://enanomapper.adma.ai/about/omics/ | A separate eNanoMapper database instance used to gather metadata that links to nanosafety-relevant omics data |
| ECOTOX[181] | https://cfpub.epa.gov/ecotox/ | A comprehensive Knowledgebase providing single chemical environmental toxicity data on aquatic and terrestrial species. |
| NBI | http://nbi.oregonstate.edu/ | A database of NM physicochemical properties and their biological interactions, and a computational tool to predict the hazards of newly designed NMs |
| NanoCommons Knowledge Base[182] | https://ssl.biomax.de/nanocommons/cgi/login_bioxm_portal.cgi | An ontology-linked semantic database for NMs and their transformations, that provides interoperability for nanosafety data sources and tools, on both semantic and technical levels, and integrates a number of predictive models and descriptor generating tools. |
| NanoPharos | https://db.nanopharos.eu/Queries/Datasets.zul?datasetID=1 | A database of ready-for-modeling data sets that can be directly imported into computational workflows. Data sets can be enriched with a wide range of structural, molecular and atomistic NMs descriptors to increase the potential for ML model development. |

used for modeling of nanotoxicity, but not all are specific for NMs.[50−53]

Figure 3A demonstrates the relationship between data source, data set size, the number of descriptors and the type of ML task (Classification/Regression). We can infer that most researchers tend to collect data from literature/published data and experiments, as most common databanks contain limited NMs and not all crucial descriptors. The data volume used for modeling depends also on the type of task: classification tasks use significantly higher data volumes than those used for regression tasks. It should be pointed out that classification and regression tasks can use overlapping data, and the data requirements and preprocessing are the same. The difference mainly lies in the target variable: the endpoint of classification is categorical, and the endpoint of regression is continuous.[54] Generally, categorical data are more readily available than continuous data, and the continuous endpoint can be converted to categorical, but the reverse requires a natural ordinal relationship of class labels. This explains why classification tasks use more data than regression tasks.

Generally, most studies building ML models for nanotoxicology use data collected from different articles, where experiments often follow different protocols. The heterogeneity of data and missing values pose difficulty to pretreatment and thus, inevitably lower the overall data quality. Data heterogeneity encompasses both the wide variety of categorical data and the huge range of variation in numerical data, and thus the handling methods are different. Methods of handling categorical data include one-hot coding (which converts categorical values to numeric ones),[51,55] and methods of handling numerical data include normalization,[56] standardization[57] and logarithm-scaling.[58] Commonly used methods for handling missing values, from simple to complex, include: (1) removal of entire observations with missing data; (2) filling the missing gaps with representative values, such as mean, median, mode and expert knowledge,[20] (3) interpolation using neighbor values;[59] and (4) imputation using ML-based methods.[23,60,61] It is recommended to explore the imputation methods to fill the data gaps, as it has the least error and results in the best prediction accuracy.[62] For example, in the study of Sizochenko et al.,[23] a decision tree (DT) classification model to evaluate the genotoxicity of metal oxide NMs was developed, combining supervised and unsupervised learning methods to fill in data gaps, with the results showing that the model achieved 75% accuracy and 25% error rate in the test data set.

Given that there is no universally accepted scheme for assessing data sets, several solutions have been proposed to assess the quality and completeness of the data set.[63] For example, Lubinski et al.[64] presented a criteria to evaluate the usefulness and quality of data in nano-QSPR/nano-QSAR: the data source, data volume, experimental protocol and NP characterization were comprehensively considered to rank a data set. Trinh et al.[65] also proposed their own criteria (PChem score) to evaluate the quality and completeness of data, which focused on the data source and measurement method of NMs physicochemical properties, meaning that experimentally measured data with commonly used or standardized methods will get a high score under this criteria. More recently, Basei et al.[66] proposed an approach to assess the completeness and quality of the data automatically, whereby the criteria of evaluation were divided into four dimensions: Completeness (the degree to which all required


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

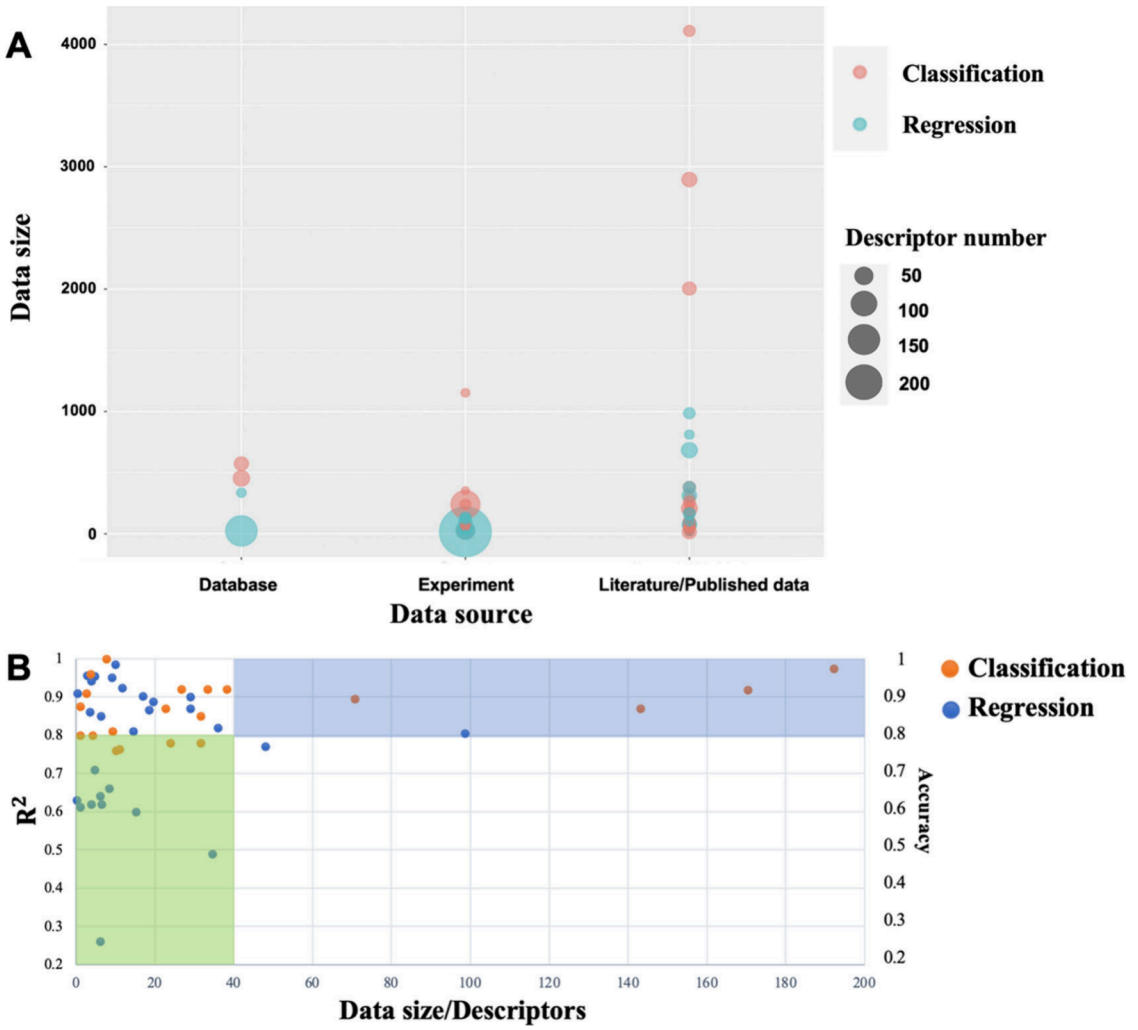



**Figure 3.** Descriptive statistics of the model information. (A) The relationships between the data source, data set size, the number of descriptors and the task type. (B) The relationships between the model performance and the ratio between data set size and the number of descriptors.

data in a data set in available), Reliability (if a study was conducted in a reliable manner), Relevance (if a study was conducted using standard protocols/procedures) and Adequacy (the usefulness of the data for risk assessment purposes). Although these methods still require efforts to become widely recognized standards for evaluating data set quality, the attention given to data source and reliability is something that cannot be overlooked.

**3.2. Identification of Descriptors for Modeling.** The basic idea of ML models is to establish a relationship between independent variables (like physicochemical properties of NMs, exposure conditions and organism information) and the dependent endpoints of interest. A reasonable taxonomy of descriptors is a prerequisite for descriptor selection. Based on the source of descriptors, they may be divided into experimental and theoretical types.[43] Experimental descriptors refer to data acquired via experimental methods. Nanospecific descriptors such as shape, size, aggregation state, and surface area can be obtained from images taken by techniques such as transmission electron microscopy (TEM), scanning electron microscope (SEM), and nanoparticle tracking analysis (NTA).[67−69] Test organisms-related descriptors can be obtained from images, as well as from endpoint measurements like growth, reproduction, biomarkers and mortality scorings, or experimental methods like permeability of membranes, and any early warning indicators/biomarkers.

Theoretical descriptors are obtained by calculation on the basis of structural information about molecules using several software packages and platforms, e.g., ADRIANA, Cerius2, MOPAC,[70] OCHEM platform.[71] Simplified molecular input-line entry system (SMILES) is a typical theoretical descriptor, it converts a chemical structure into a string that can be easily read by humans and processed by computers,[72] it also serve as the basis for calculating other theoretical descriptors with the help of software.[73] However, SMILES-based descriptors ignore NMs physical properties (i.e., size, shape, coating) which is crucial in nanotoxicology modeling.[74] Quantum chemical-based descriptors are parameters derived from quantum mechanical calculations that describe the electronic structure, reactivity, and properties of molecules.[75] They are applied in QSAR models to predict the adverse effects of NMs,[76,77] but on the other hand they are computationally expensive and require expert knowledge to interpret the relationship between nanostructure and toxicity.[74] As an successful alternative to quantum-chemical descriptors, periodic table-based descriptors (i.e., molecular weight, electronegativity, valence of metal) have the advantages of being both convenient for calculation and easy to understand.[20,58] Theoretical organism-related


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

descriptors from commonly used laboratory-reared species used in ecotoxicity can be taken from trait databases like Add-my-pet (www.bio.vu.nl/thb/deb/deblab/add_my_pet/index.html) or FishBase.[77] The descriptors related to cell type features can be found in the database of 10x Genomics (www.10xgenomics.com). Note that, although saving time and cost, the addition of theoretical descriptors does not always improve the model performance, because NMs behave differently than conventional materials and calculated theoretical descriptors are conventionally more collected, trained and thus more applicable to soluble chemicals[60] and non-nanomaterials.[59,78] Therefore, it is necessary to develop a simple and appropriate method in order to convert NMs structural information into computer language.[79]

In addition, the toxic effect of NMs are the result of complex interactions between the properties of the NMs themselves, the exposure conditions and the test organisms.[80] From this perspective, we can also divide the descriptors into three categories: NMs-related, exposure conditions, and organism-related descriptors (Table S2), as based on our previous study,[20] good practice should include all three types of descriptor in order to fully reflect the processes and transformations of NMs in the environment, and the context-dependent nature of many NMs properties.[81,82]

As the cornerstone of the nano-QSAR model theory, the NMs-related descriptors are the most diverse and most extensively studied. Among them, the crystalline form, band gap energy, electronegativity, and enthalpy of cation formation are commonly used in models, because they are closely related to the release of ions from the surface of NMs and to the photochemical effects of NMs,[61,83] which have been proven to be a critical part of the mechanisms underlying nanotoxicity, especially oxidative stress and inflammation.[84] For the experimental descriptors, size,[85] shape,[86,87] coating,[88] and surface area,[89] are frequently considered as important factors, as they directly influence the physical contact of NMs with the test organisms.

Features of the exposure conditions describe the experimental environment, including the exposure dose, exposure duration, zeta potential, medium composition, temperature, and illumination. The exposure dose and duration influence the toxic effect of NMs. Connell et al.[90] have established models to describe the relationship between the exposure duration and $LC_{50}$, while Choi et al.[52] developed a quasi-QSAR model to explore the influence of exposure dose on cell viability of human lung (BEAS-2B) and skin (HaCaT) cells. Zeta potential reflects the stability of the NM suspension and the surface morphology of NMs, which is a crucial part of the most common mechanisms of nanotoxicity, including charge related lysosomal swelling and oxidative stress.[91] Zeta potential is also considered as an endpoint of prediction but it should be noted that zeta potential is an extrinsic NM property which depends on the composition of the NM and on the composition of the exposure medium simultaneously.[92] Illumination is an environmental factor that cannot be ignored, especially when the NMs are photoactive ($TiO_2$NPs, AgNPs): the exposure to ultraviolet (UV) light will enhance the appearance and dissolution of NPs, thus influencing their toxic potency.[93−95]

Organism-related descriptors are receiving less attention within the literature. For *in vivo* toxicity prediction models, species type is not considered as a descriptor because the models only focus on a single species. For cell viability prediction models, cell line, cell species, cell source, and cell type are commonly applied as categorical variables.[17] Compared to the NMs-related descriptors, there exists a clear imbalance of descriptor categories in current toxicity prediction models.[20] As organisms' selection is related to the chemistry of the exposure medium and thus subsequently to the fate of the NMs. Converting the differences between species into computer language is an interesting and challenging topic, but it is worth exploring because it will allow us to gain insight into the mechanisms of toxicity of NMs and to build accurate multispecies models. Allosteric scaling is one approach used in PBPK modeling to account for differences in sizes of test organisms,[96] it can be used for NMs-induced impacts in lungs and even for aquatic cladoceran species by having the smallest species being more sensitive to CuNP compared to the larger species.[97,98] Another way of approaching is classifying species according to trophic level. In an earlier attempt,[20] this was ineffective because it lacked crucial information, such as the morphology, ecological traits and intrinsic sensitivities. Currently, the inclusion of such species characteristics remains an obstacle because trait databases always contain generic and averaged values for restricted species characteristics, this neglects that species within different stages of the life will have different sensitivity due to altered morphological and energy related traits, and feeding behavior. Considering the similarities and differences between species from the perspective of the functional genome is likely also a promising approach. For example, an ancestral gene pathway triggered in response to NMs exposure has recently been elucidated, involving cell stress responses, protein misfolding and chromatin remodeling and immunomodulation, that utilizes zinc finger proteins and is conserved across humans, zebrafish, daphnids, earthworms and wheat.[99] Applying the ratio of certain genes in the whole genome is, for instance, a possible solution in this respect as it reflects the differences in antitoxic ability of different species,[100] and is at the same time easy to convert as a numerical variable. This is not commonly utilized in nanotoxicology at present because of our insufficient understanding of the relationship between gene expression and apical endpoints.

The number of descriptors should be matched to the data set size, meaning that the data volume should increase as the number of descriptors grows. A limited data set size, or number of descriptors, will affect the predictive power, interpretability and the applicability domain of the model. The data set size also has an influence on the ML algorithm used. More complex algorithms such as neural networks will, in general, benefit from larger data sets. Even though there is currently no specification for the minimum data set size required for model training, the standard for the data volume required for modeling should be matched with the actual situation. Zhu et al.[60] proposed in their review that the minimum requirement of data volume to descriptor ratio for simple classification problems should be at least 10, which is recommendable for nanotoxicity modeling as the generation of reliable data is time-consuming. Figure 3B demonstrates the relationship between model performance (i.e., accuracy for classification, and $R^2$ for regression), and the data volume to descriptor ratio. We can infer that models with high ratios (>60) generally perform better.[101] However, the vast majority of studies with a ratio between 20 and 40 also performed well, and some failures were more specific to improper data preprocessing methods.[53] This suggests that it would be


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

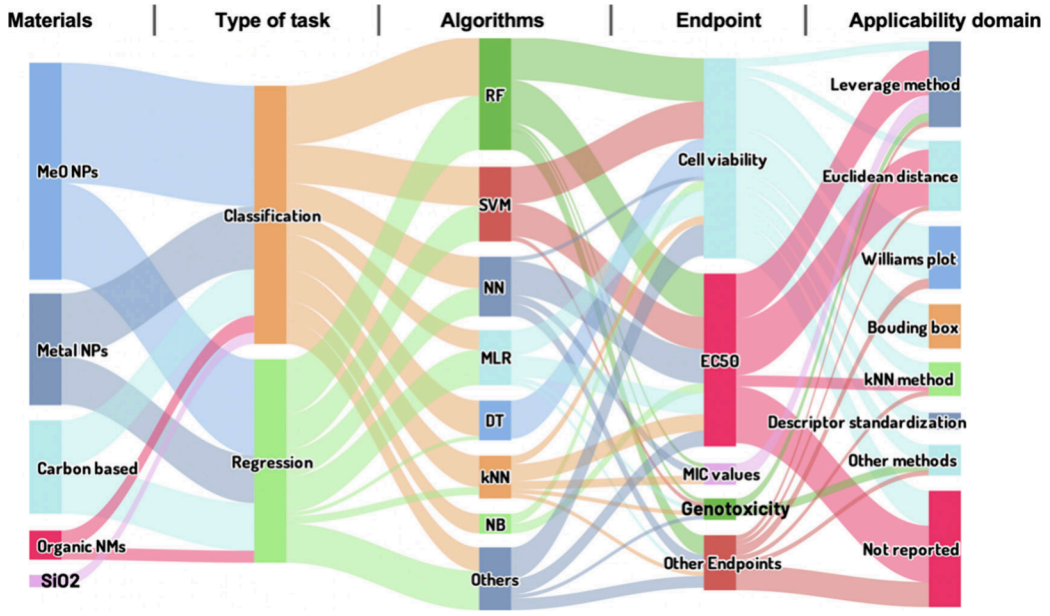



**Figure 4.** Sankey Figure about the study design information using the"highcharts" platform (https://www.hcharts.cn). The width of the flows is proportional to the number of cases. RF: random forest, SVM: support vector machine, NN: neural network, MLR: multiple linear regression, DT: decision tree, kNN: k-nearest neighbor, NB: naïve Bayes, others: other algorithms.

optimal for a model to learn from sufficient data to reduce the probability of poor model performance, but when getting more data is likely impossible or would pose a significant burden, setting the ratio between 20 and 40 would be a good compromise. In fact, this effect may be even greater given that researchers may selectively publicize the results of their better models. For example, Hager et al.[56] developed a model to predict the cytotoxicity of several organic and inorganic NMs. They collected 2896 data samples, and examined 15 descriptors, including four NMs-related features, seven cell-related features, three methodological parameters and the exposure time. As a result, the data volume to descriptor ratio was 170, indicating a good learning potential. In contrast, in a study using ML to predict the developmental toxicity of NMs, a RF model using 1000 trees was applied to deal with only 90 data points and 14 descriptors. We believe the hyperparameter of the model was too high, and the data size was too small to support a ML application, and it turned out that the model performance was not satisfactory enough.

## 4. MODEL DEVELOPMENT: ALGORITHMS, END POINTS, AND DOMAIN OF APPLICABILITY

By analogy to the first three of the five principles of QSARs development and validation for regulatory purposes as proposed by the OECD,[26] each model should be associated with (i) a well-defined endpoint, (ii) an unambiguous algorithm, and (iii) a defined domain of applicability. The Sankey diagram presented in Figure 4 demonstrates the important information flow of modeling: What materials are chosen, which type of the prediction (classification or regression), the algorithm used, the endpoint, and the method used to define the applicability domain (AD). These are discussed in further detail below.

**4.1. The Materials.** To date, nano-QSAR models have been primarily developed for metal/metal oxide materials.[16] However, with the emergence of novel materials, the prediction scope has expanded to carbon-based materials (graphene, multiwalled carbon nanotubes, etc.) as well as organic NMs (polymeric nanoparticles).[55,102] As shown in Figure 4, although single element first generation NMs being mostly metal/metal oxide NMs still remain the mainstream focus of research, analytical development to characterize the third and fourth generation of NMs (e.g., carbon based NMs, hybrid NMs, and self-forming nanostructures) is in its infancy.[103]

Currently, one of the major challenges is to consider more new NMs into existing predictive models which usually have insufficient toxicity data and appropriate descriptors. By their nature, these novel NMs have been studied for a relatively short time, meaning that there is limited available data on their physicochemical properties, toxicities, risks to humans and the environment for modeling. Additionally, when a new material is added for prediction, researchers need to identify descriptors that can appropriately capture its physicochemical properties. For example, Wang et al.[57] developed a set of geometrical descriptors for graphene based on the nanostructure annotation techniques. In another study of Sengottiyan et al.,[92] the atomic molecular weight and the number of hybridized carbon atoms were used to describe the structure of the core and coating of organic NMs. Integrating these data sets is not straightforward: harmonization, interconversion, and collaboration with NMs designers are all useful methods to enhance the NMs risk assessment.

**4.2. The Algorithms.** In terms of the algorithms used in NMs toxicity modeling, the OECD principles require full model structure and accurate values for all the model parameters to be specified.[26] Figure 4 demonstrates the algorithms used for classification and regression. As can be seen, commonly used algorithms such as RF, support vector machine (SVM) and neural network (NN) are frequently used in both type of tasks, while other algorithms were preferentially used in one type of task. For the classification model, researchers preferred decision tree (DT), k-nearest neighbor (k-NN) and Naïve Bayes (NB), while multiple linear


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

regression is more often used for regression tasks. It is recommended to compare at least two algorithms to justify the algorithm selection. In our statistics, 38.4% of the studies applied a single algorithm without comparison with other methods, but a majority of studies (61.6%) used at least two methods and performed a comparison. For example, Yang et al.[75] assessed various algorithms, including DT, RF, locally weighted learning, k-NN, logistic regression, Bayes classifier, and SVM, for the prediction of toxicity of NMs in immune cells. Based on sensitivity (SE), specificity (SP), overall predictive accuracy (ACC), F1 score, Matthews' correlation coefficient (MCC), and the area under curve (AUC) of a receiver operating characteristic (ROC) curve (a graphical plot that illustrates the performance of a binary classifier model at varying threshold values), k-NN, RF, Bayesnet and DT exhibited better performances than other models.[75] Based on the comparison, the four best-performing classifiers were selected to contrast the resulting consensus models. The creation of models based on multiple algorithms can also be beneficial as each individual model can be combined (ensemble methods or stacking) into meta-models often improving predictive performance.[47]

The hyperparameters of models are often overlooked during modeling, and in fact, for some algorithms whose performance is strongly influenced by hyperparameters, i.e., a parameter which specifies details of the learning process itself, such as the learning rate or the choice of optimizer. Thus, interpreting model results without hyperparameter optimization is questionable.[60] Researchers are, therefore, advised to report the hyperparameters of the models they used in their papers (i.e., the number of trees and number of features selected at each node for RF, the kernel function used for SVM, the number of neurons and the learning rate for NN, the number of nearest neighbors for kNN etc.) and perform the hyperparameter optimization.[104,105]

*4.2.1. Algorithms Used for Classification.* The RF is the most popular algorithm to date, as 29 out of 52 studies applied RF to develop models and generate predictions. RF is a type of supervised algorithm that works by constructing a multitude of shallow decision trees to avoid errors and make a more reliable prediction than a single deep decision tree. The advantages of RF lie in its ability to maintain high accuracy and robustness and work with missing data and categorical data, while dealing with a large number of input features. However, it is worth noting that RF can be less interpretable and requires more training time and memory if the number of trees and the depth of the trees are high (e.g., number of trees >1000, depth of trees >10). RF shows good performance in both classification and regression tasks, and often outperforms other algorithms.[17,20,80,89,104]

The SVM is the second most popular algorithm, applied by 19 out of 52 models. SVM is good at handling high-dimensional problems. It projects data as points into a high-dimensional space and then searches for the hyperplane using a kernel function that can best separate the classes of data.[106] In the study by Liu et al., SVM was applied to build a classification model for the recognition of the toxicity of engineered NMs to zebrafish embryo.[56] The accuracy of the model was 97.4 ± 0.95% in a 10-fold cross-validation. SVM has advantages of high accuracy and resistance to overfitting with the appropriate kernel function.[107] However, it should be noted that the performance of SVM is sensitive to the optimization of the algorithmic parameters.[44]

Sixteen studies in our literature analysis applied NN. NN is a collection of nodes which are interconnected and divided by multiple layers, the node of each layer works by receiving, processing, and delivering signals to the next layer, similar to neurons in brains. In the study performed by Sizochenko et al.,[108] a NN model was developed to estimate the zeta potential of 208 silica- and metal oxide NPs in different media. NN is good at processing large and complex data sets, and building nonlinear relationships between dependent and independent variables. Disadvantages include the fact that NN is memory intensive and prone to overfitting, and the poor (mechanistic) interpretability is also a shortcoming of NN. However, with the continuous advancement of research, the potential of NNs is being increasingly explored. Deep learning models, which break the limitations of layers of traditional NNs, have been developed and applied on larger-scale data training.[109,110]

DT is a nonparametric supervised ML algorithm, each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. It works by recursively splitting the training samples into two groups based on the feature that best divides them.[111] The DT algorithm is easy to interpret and can handle a mixture of data types, but during our investigation, it is found to have, in general, a lower classification accuracy compared to other classifiers, and DT models are prone to overfitting.[112,113]

KNN is a popular instance-based learning method, that works by comparing an input data point to k-nearest neighbors based on a distance measure in the training data set.[114] In the study by Varsou et al.,[85] a KNN based model was developed to explore the effects of a panel of freshly dispersed and environmentally aged Ag and $TiO_2$ NMs on immobilization (acute toxicity) of *Daphnia magna*. The model reached accuracies of 89.5% and 76.5% on test and training samples, respectively. In another study, the Enalos implementation of the KNN methodology was applied to produce a predictive regression model for cytotoxicity of NMs.[50] The model was evaluated and found to have a value of $R^2$ of 0.91. The KNN method is simple to implement and requires no prior knowledge of the data distribution, but its accuracy decreases significantly when the number of dimension increases.

The NB classifier is a supervised ML algorithm which is used for classification tasks. It simplifies a classification problem by calculating a single probability for each variable.[115] NB classifier is less complex yet nevertheless efficient, and performs well particularly with small data set volumes. In the study of Simeone et al.,[116] a naïve Bayesian classifier was developed to determine the most probable level of toxicity of a NM given its composition. However, the NB classifier requires the fulfillment of the assumption that predictors are conditionally independent, and that all the features contribute equally to the outcome. These conditions will not always hold in real scenarios, especially for nanotoxicity data where lots of variables interact with, and are dependent on, others.

*4.2.2. Algorithms Used for Regression.* Multiple linear regression (MLR) is a type of statistical model constructed by estimating the coefficients of the independent variables through the use of a least-squares method, which works by seeking a relationship that minimizes the sum of the squared residuals between the observed values of the dependent variable and the predicted values of the independent variables.[15] In the study of Milolajczyk et al.,[113] a MLR


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

model was built to evaluate the *in vitro* cytotoxicity of 29 $TiO_2$-based multicomponent NMs. The linear MLR model successfully described the cytotoxicity of the NMs, and a linear relationship between the endpoint ($EC_{50}$) and the additive electronegativity ($\chi_{mix}$) of the NMs was established. The $R^2$ of the model equaled 0.87, and the root-mean-square error (RMSE) was 0.20. Zhou et al.[117] applied MLR to predict the cytotoxicity of 17 metal oxide NMs to *E. coli*, results demonstrated that MLR model presented high reliability and good predictive performance ($R^2$ greater than 0.84). In another study, MLR was used to evaluate the cytotoxicity of metal oxide NMs toward RAW 264.7 cells,[118] along with an intelligent consensus prediction (ICP) to select the best combination of different MLR models. The model achieved not only good results but also successfully elucidated the mechanism of cellular toxicity by the metal oxide NMs. MLR is one of the most commonly applied regression-based QSAR modeling techniques. It has the key advantages of transparency and ease of interpretability, whereby the importance of each independent variable can be easily identified. But unfortunately, MLR cannot detect nonlinear causal relationships, and does not guarantee causal inference.[44]

Aside from MLR, other regression algorithms were used in the investigated studies, such as partial least-squares regression (PLSR), kernel ridge regression (KRR), and Gaussian process regression (GPR).[57,77] PLSR is a method that combines principal component analysis and multiple regression.[119] It is suitable for the modeling of small data set sizes and large numbers of descriptors, because it performs a descriptor dimension reduction and constructs a set of components that accounts for as much as possible of the total descriptors variance in the data set, which can avoid multicollinearity and model overfitting.[120,121]

KRR is a regression algorithm which combines ridge regression with the kernel trick, it introduces a regularization hyperparameter to prevent the model from overfitting the training data. Furthermore, the kernel function transforms the data into higher dimensions, which allows the model to solve nonlinear problems.[122] GPR is a probabilistic approach to regression modeling, that estimates a probability distribution over a possibly infinite number of functions that fit the data. The main advantage of GPR compared to other algorithms is that it provides a probabilistic interval for the prediction, which means it generates both the predicted value and the confidence interval of the prediction.[123]

The algorithms used for classification do not undergo significant changes when applied to regression tasks. For NN, the only variation required is addition of the activation function used in the final layer and the corresponding loss function applied during backpropagation. For RF, it exhibits more substantial disparities when applied in regression tasks. The construction of individual trees follows a slightly different splitting criteria than that used in classification models. Furthermore, the aggregation of predictions also differs. In classification tasks, it can include taking the mode of the trees' hard classifications or the mean of the trees' soft classifications, while in regression tasks the predictions are simply averaged.

*4.2.3. More Advanced Algorithms for the Future.* The past few years have witnessed the great achievements of deep learning (DL) algorithms in speech recognition, image processing, and language understanding. Although DL has not yet been widely applied in nanotoxicology modeling owing to the limitedly availability of nanobioactivity/toxicity data, several attempts have been made to implement DL models into prediction of NMs activity. For example, in the study of Russo et al.,[124] an image processing convolutional neural network (CNN) was built to predict the cellular uptake of 77 NMs in A549 cells, and their lipophilicity and zeta potential in water. The NM structure information was transformed into "virtual molecular projections", which are multidimensional digital data that represent and include all components of the NMs' structure. The model achieved good results in the prediction of cell uptake ($R^2$ = 0.845) and lipophilicity ($R^2$ = 0.89) but did not perform well in zeta potential prediction.

Due to their ability to handle massive, complex and often ill-understood data, DL models enable us to address the challenge of the explosive growth of data in the field of nanotoxicology. However, the complex structure of DL models make them less interpretable compared with classic ML models.[38] Methods such as backpropagation-based methods and connection weight-based methods, have been developed to improve the interpretability of DL models.[125,126]

Nanomaterials can be developed in virtual with any shape, size and composition currently. Additionally, they are dynamic in behavior (fate) at the interface of exposure and the test organisms (related to uptake and effects), and able to showcase an enormous set of different types of toxic effects. This results in a thousand of combinations that can be modeled and the dynamics and variation challenge every model to make accurate predictions. Thus, the application of DL models in computational toxicology holds great promise for the future because it can deal with complex data sets, while efforts still need to be made to improve their interpretability based on processes that are agreed upon in the scientific nanoexperts community.

In response to the limitation of nanotoxicology data set size, researchers have proposed a series of solutions to improve model performance and interpretability. Transfer learning is a possible solution for limited data set size; it works by exploiting the knowledge gained from a previous task to improve the generalization about another task.[127] By applying transfer learning, the need for large amounts of nanotoxicity data can be reduced and predictions can be made more efficient and accurate. For example, in the study of Zhong et al.,[128] a knowledge transfer model was developed to predict the reactivity of organic contaminants toward four oxidants; the result showed that the knowledge transfer model outperformed multitask learning and image-based transfer learning. However, it should be noted that the application of new methods should be combined with understanding of the complexities of NMs' properties and toxicity mechanisms, otherwise the model will have an interpretability problem.[60]

**4.3. The End Point.** According to the definition of the OECD, the endpoint refers to any physicochemical property, biological effect or environmental parameter related to chemical structure that can be measured and modeled.[26] In the studies reviewed, the predicted endpoints were clearly identified, 47 out of 52 studies applied toxicological endpoints. Cell viability and the median effective concentration affecting 50% of individuals ($EC_{50}$) are the most chosen endpoints (Figure 4), as they are the most commonly used regulatory toxicity measures.

The minimum inhibitory concentration (MIC) and genotoxicity were also used as prediction endpoints. The MIC value refers to the lowest concentration of the toxicant needed to produce an inhibitory effect (on microbes).[53] The


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

Table 2. Methods commonly applied to define the AD of a model

| AD Methods | | Hypothesis | Flaws |
|---|---|---|---|
| **Ranges in descriptor space** | Bounding Box | Modeled descriptors are considered with a uniform distribution | Empty regions in the interpolation space cannot be identified; Correlation among descriptors cannot be considered |
| **Distance-based** | Leverage approach | The AD space is defined as a squared area with the ±3 bands for standardized residuals ($\sigma$), the leverage threshold is defined as $h^* = 3(p + 1)/n$, where $p$ is the number of descriptors and $n$ is the number of molecules | — |
| | Euclidean-based | Euclidean method calculates the distance from every other point to a particular point in the data set | The variables must be statistically independent |
| | k-NN | The theory is based on similarity search for a new chemical entity with respect to the space generated by the training set compounds | — |
| **Geometrical** | Convex Hull | The approach recognizes the boundary of the data set considering the degree of data distribution | The complexity swiftly amplifies in higher dimensions |
| **Miscellaneous** | Standardization approach | The data follows an ideal data distribution | Does not consider intercorrelation among descriptors, nor the relative contribution of descriptors |

genotoxicity was expressed in a binary form, but the definition differs in review studies: Kotzabasaki et al.[129] measured genotoxicity by DNA strand breaks (*in vivo*) and gene mutation in mammalian cells (*in vitro*), while Sizochenko et al.[23] analyzed genotoxicity using the% of DNA in the tail, tail olive moment and tail intensity (via the Comet assay). These metrics can more accurately reflect the specific toxic effects and mechanisms of toxicity of NMs on cells.

**4.4. The Applicability Domain.** The task of an AD is to define the space of descriptors in which a model can make reliable predictions.[130] Therefore, a well-defined AD space is a "must have" characteristic for *in silico* prediction systems. AD definition techniques vary according to different hypothesis (Table 2): Ranges in descriptor space (i.e., Bounding box, PCA Bounding Box, etc.), distance-based (i.e., Leverage approach, Euclidean distance, k-NN approach, etc.), geometrical (i.e., Convex Hull), and miscellaneous (i.e., Standardization approach, Kernel-based, etc.).[130]

Forty out of 52 studies defined the AD of their models. Among them, distance-based methods, which focus on the "distance-to-centroid" principle, are the most frequently employed approach (23 out of 52).[58,59,131] Leverage approach is a main branch of distanced-based methods: the AD space of the model is defined as a squared space within ±3 bands for standardized residuals ($\sigma$). the leverage threshold is defined as Equation 1:

$$h^* = 3(p + 1)/n \quad (1)$$

where $p$ is the number of descriptors, $n$ is the number of data points. The leverage values ($h$) are calculated and plotted ($X$-axis) vs cross-validated standardized residuals ($\sigma$) ($\gamma$-axis), and the resulting plot is called the Williams plot.[132] A Williams plot illustrates the distribution of data points, and any points outside the leverage value and $\pm 3\sigma$ are considered as less reliable.[133,134]

In addition to the definition of the AD based on the leverage method, Euclidean-based methods can also be used to detect outliers. Euclidean-based methods calculate the distance from every other point to a particular point in the data set. By creating a boundary area normalized mean distance scores of the training set, a data point of the test set which resides inside the domain covered by the training set is considered as reliable.[130] The KNN approach is another way to define the AD. This approach calculates the similarity between data points by taking the distance of a point from the nearest training compound or its distances from the KNN in the training set.[135] If the calculated distance values of a data point are within the threshold set by the training set, the prediction of these molecules are considered as reliable.

Although the AD's distance approach can evaluate the confidence levels by drawing iso-distance contours in the interpolation space, one of the disadvantages is the hypothesis that the data should follow a normal distribution. For uniformly distributed data and independent descriptors, the Bounding Box is an appropriate approach to define the AD: it defines an n-dimensional hyperrectangle. Each side of this rectangle ranges from the minimum to maximum values of each descriptor, and data points outside of these particular ranges are considered to be outside of the AD.[136]

The geometrical method estimates the direct coverage of an n-dimensional set utilizing the convex hull calculation performed based on complex efficient algorithms.[137] However, convex hull performs well only on two and three dimensions, the complexity swiftly amplifies in higher dimensions, making it difficult to apply in practice. For miscellaneous approaches, the most used was the standardization technique: Briefly, a data point was considered an outlier if all normalized descriptors for the data point were greater than 3, otherwise it was a nonoutlier.[138] The standardization technique is simple to understand, and is applied by several studies among the literature analyzed,[75,76] but it does not consider intercorrelation among descriptors and the relative contributions of descriptors. Therefore, researchers are advised to justify the reason for the selection of the AD determination method based on different available hypotheses.

## 5. MODEL APPLICATION: MODEL VALIDATION AND INTERPRETATION

**5.1. Model Validation.** The fourth OECD principle for QSAR model validation includes the need for goodness-of-fit, robustness, and predictability measures.[26] The model validation can be divided into internal and external validation. The internal validation focuses on the goodness-of-fit and robustness, while the external validation mainly focuses on the predictability.

Of the studies reviewed, with the exception of the unsupervised model, almost all the models (47 out of 52) performed internal validation to test the goodness-of-fit and robustness of the model predictions. Goodness-of-fit is a measure of how well the predicted values fit the observed values, and the $R^2$ is a common index to test the quality of fit.[139] The robustness of a model refers to its ability to perform


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

well when faced with perturbation. The RMSE and $Q^2$ are common indicators to evaluate the robustness of a model and these indicators are also a crucial component to include in model validation in our view.[55,134] The *k*-fold cross-validation (*k*-fold CV), leave-one-out cross-validation (LOOCV) and leave-multiple-out cross-validation (LMOCV) are the most frequently used resampling methods to evaluate the model robustness.[57,140] *k* -fold cross-validation splits the data set into *k* subsets, one as the test set and the other as a training set, this process is repeated *k* times to obtain the average value of performance score (e.g., $R^2$).[85] LOOCV is a special case of *k*-fold CV, where *k* equals to the training sample size to maximize the number of CV folds.[77] LMOCV is a generalization of LOOCV, where instead of leaving out a single data point, multiple data points are left out each time and the remaining data points are used for training.[85] *k*-fold CV can effectively use all data points, and is more computationally efficient for large data sets. Compared to *k*-fold CV, LOOCV provides the least biased estimate of model accuracy for small data sets. However, when it comes to large data sets, LOOCV is not recommended as it is computationally expensive, and the variance could be high since the training set is evaluated on a single data point. LMOCV, in this case, provides a balance between *k*-fold CV and LOOCV as it can be more flexible and less computationally intensive. It is recommended that researchers consider the size of the data set, the computational resources available, and requirements of model validation to choose an appropriate method.

Y-randomization is another valuable but less applied tool in internal validation of models: the model is tested with the randomly shuffled endpoint Y values, and the result ($R^2$, Z-score, etc.) is compared to the original model in order to determine if the model is overfitted.[139,141] It is worth noting that Y-randomization can also help identify critical features which determine the endpoint, and thus can support model explainability.

Only 18 out of 52 models performed external validation, which is much less than the models that performed internal validation. The goal of external validation is to evaluate if the model can perform well on new data. We believe that both are necessary for model development, but if the model will be deployed in practical application and need to meet regulatory standards, external validation will be more necessary and reliable. As an important process to assess the model predictability, external validation is more and more being considered as a necessary step in model validation.[142] The predictability of a model can be evaluated by means of the mean squared error (MSE),[105] the $Q^2_{ext}$ value,[113] the mean absolute error (MAE) and the $RMSE_{ext}$.[77]

For classification models, the sensitivity (recall), specificity, accuracy, Matthews correlation coefficient (MCC), Receiver Operating Characteristic (ROC) curve and area under the curve (AUC) are the most applied tools to evaluate the model performance.[76,143] These metrics are calculated based on the numbers of true negatives (TN), true positives (TP), false negative (FN), and false positives (FP). The indicators are the basis for the calculation of the sensitivity (SE = TP/[TP +FN]), specificity (SP = TN/[TN+FP]), and accuracy (ACC = [TP + TN]/[TP + FP + TN + FN] of the models, and for the calculation of the MCC[144] (Equation 2):

$$\left(MCC = (TP \times TN - FP \times FN) \Big/ \sqrt{(TP + FP)(TP + FNeqn)(TN + FP)(TN + FN)}\right) \quad (2)$$

Sensitivity, also known as recall, measures the proportion of actual positives that a model correctly identified, high sensitivity means that the model correctly identifies the positive instances. Specificity, on the other hand, represents the model's ability to identify the negative instances. Accuracy measures the model's ability to make correct predictions (TP +TN), it is useful but can be misleading when the data set is imbalanced.[76,143] The MCC value ranges from −1 to +1, a higher value of the MCC indicates that the binary predictor is able to correctly predict the majority of positive and negative cases.[144] The ROC curve illustrates the diagnostic ability of a binary classifier system, it is constructed by plotting the TP rate versus the FP rate at various threshold settings.[145] The area under the ROC curve (AUC) is a measure of how well a model distinguishes positive and negative data points:[146] a larger AUC indicates higher model predictivity, but an AUC of 0.5 means that the model predicts randomness. The above-mentioned metrics are commonly simultaneously used because they address different aspects of model (i.e., the overall accuracy, the balance between different types of errors, the robustness to class imbalance), and they are all essential for performance evaluation and comparison between models. Therefore, it is recommended to evaluated the model rather than just one metric.[15]

**5.2. Mechanistic Interpretation.** According to the fifth OECD principle, a (Q) SAR model should be associated with a "mechanistic interpretation" wherever such an interpretation can be made.[26] The intent of this principle is to ensure that there is an association between the descriptors used and the endpoint predicted: If a descriptor in the model is considered as important, that means it may act as a crucial factor in the toxicity of the specific NMs.[20] For the time being, the exploration of model interpretability in most studies is based on the method of calculating the feature importance, such as weight of importance in RF-based models, SHAP (Shapley Additive exPlanations), PDPs (Partial Dependence Plots) and LIME (Local Interpretable Model-agnostic Explanations).[17,80,104,147,148] SHAP explain the output of a model by calculating the average contribution of each feature to the prediction across all possible subsets of features.[149] SHAP can be used in both local and global model performance and is not algorithm specific, but can be computationally expensive. PDPs focus on the relationship between a single feature and the endpoint, it provides a visualization of how a feature influences the prediction result.[150] Nonetheless, PDPs require that the features are independent of each other, which is not always the case. LIME explains the local predictions by approximating the model with an interpretable model, and can be used for any model.[151] However, the result of LIME can be unstable due to random generation of training data, and introducing Variables Stability Index (VSI) and Coefficients Stability Index (CSI) can solve this problem.[147]

In addition to the above-mentioned interpretability method, there are some studies that have explored model interpretability more deeply: Sizochenko et al.[84] introduced the causation inference method, which is based on conditional probability and directed and undirected graphs, to elucidate the underlying structure of the nanotoxicity data set. They


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

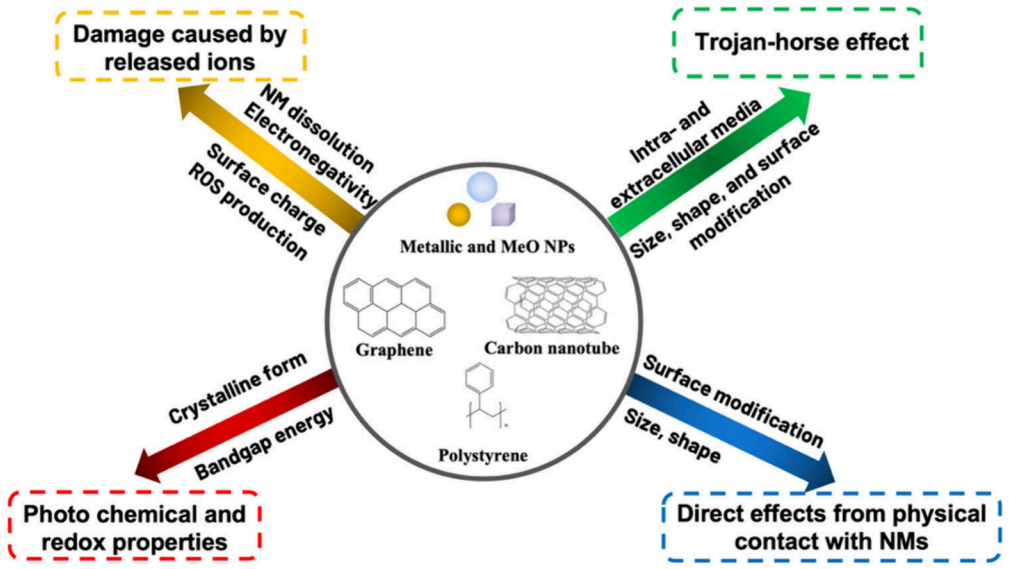



**Figure 5.** Overview of the mechanisms of toxicity, as deduced from the reviewed literature, next to the arrow are the factors influencing the specific mechanism of toxicity.

applied this method to a nano-QSAR model of metal oxides' toxicity toward bacteria *E. coli* and found that not every descriptor that was statistically correlated with the toxicity endpoint. This is in essence a general rule "Statistical significant aspects do not necessarily mean biological significant effects" that is easily misinterpreted when having too suggestive interpretation of findings.[152]

More recently, Yu et al.[41] improved the interpretability of a RF model by applying multiple feature importance analysis methods (i.e., increased MSE value, node purity increase, mean minimal depth, and *P* value). Moreover, they created a feature interaction network to explore the joint effects of multiple features of NMs to immune response and organ burden of NMs in rat, and the result explained the critical roles of specific surface area (SSA) and diameter in the NM-induced immune response and organ burden, respectively.

While there have been advances in methods to improve the interpretability of models, it should be recognized that the methods mentioned above are in the mathematics domain and are not necessarily related to the biochemical mechanisms of the problem being modeled.[60] Therefore, in order for the interpretability of a model to be consistent with the "real" mechanism behind the toxicity effects of NMs, expert knowledge is indispensable to understand the causal relationship between the features and the toxicity endpoint in question.

Based on the collected literature, Figure 5 summarizes four main mechanisms of NMs toxicity and the factors influencing the extent of toxicity: Damage caused by released ions, photochemical and redox properties, direct effects from physical contact with NMs, and the Trojan-horse effect.[83] These main mechanisms are consistent with the proposal of Sizochenko et al.:[61] (1) direct damage caused by released ions, as also seen in[58,61] (2) oxidative stress induced by the excessive production of reactive oxygen species (ROS) by released ions or by the surface of NMs; (3) adsorption of biologically active molecules onto the NM surface, which is influenced by surface modification;[65,92] (4) molecular structure-related effects that result in photochemical and redox properties, as also seen in,[20,80] and (5) Trojan horse effects, indicating NMs act as a vector for the uptake of contaminants to organisms.[153]

One way that the released ions cause toxicity is through interference with biochemical processes, disruption of cell membranes, or activation of immune system responses leading to inflammation.[154] The released ions can also generate ROS, which can cause damage to cell components such as DNA, proteins, and lipids.[23] The mechanism of oxidative stress generated by NMs varies according to their type. For example, for most representative metal oxides (e.g., $Co_3O_4$, $Cr_2O_3$, $Ni_2O_3$), the oxidative stress injury is due to the overlap of metal oxide conduction band energies with the cellular redox potential. For other highly soluble NMs like ZnO and CuO, the generation of oxidative stress is related to the extent of NM dissolution.[155−157]

The toxicity from ion release is influenced by various factors, which can be divided into three categories: the physicochemical properties of the NMs (e.g., surface charge, surface area, size), the exposure conditions such as the dose and the concentration gradient between the particle surface and the bulk solution, and the type of cell it interacts with.[51] For example, Zhang et al. proved in their study that it was possible to predict the oxidative stress of various metal oxide NPs based on band gap energy levels, size and particle dissolution.[158] Moreover, in the study of Choi et al.[51] the cell origin was considered as an important attribute, without which the model could not make precise predictions.

The physical interaction between NMs and cells is influenced by surface adsorption phenomena.[159] Adsorption of small molecules and proteins is an important part of the environmental behavior of NMs.[160] Biomolecule binding is crucial to the toxicity of NMs because adsorption can alter the physicochemical properties of the NMs, like for example the surface charge.[161] Adsorption of biomolecules can result in the formation of a protein corona, which can enhance the interaction between NMs and cell membrane receptors, and subsequently modulate their toxic effects.[50] The ease of adsorption to NMs can be measured by the adsorption energy; the higher the adsorption energy, the easier it is for atoms to separate from the material surface (as the adsorption energy is


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

expressed as a negative value). The adsorption energy is considered as the crucial factor in the model for predicting the cell viability proposed by Sang et al.[104] and Regonia et al.,[77] with other theoretical descriptors, such as absolute electronegativity and lowest orbital energy also being considered important.

The structure of molecular compounds can also be a cause of toxicity. Titanium dioxide ($TiO_2$) in the anatase form is known for its photoactivity due to its unique electronic band structure, and studies have proven that exposure to UV light can cause $TiO_2$ to generate ROS and be toxic to aquatic species.[162] The relevance of illumination and nanotoxicity has been confirmed in our previous study, indicating the need to consider exposure conditions in future modeling studies.[20]

The "Trojan-horse effect" represents the mechanism by which NMs act as carriers[133,163] or that bind external contaminants[135,164] and then—because of the nanospecific features penetrates a cell membrane and then release toxic chemicals at higher local concentrations than they would achieve otherwise.[153] Multiple factors have been proven to be related to the Trojan horse effect of NMs, including their size, shape, surface charge and surface modification.[165] However, as the Trojan-horse effect often appears in conjunction with other toxicity mechanisms, it is necessary to examine the components (e.g., molecules, ions, proteins) in both intra- and extracellular media to confirm their presence.[165] This complexity is the major reason why there are almost no models that consider explaining Trojan horse effects, as the relevant data is difficult to obtain. In fact, our previous research has found that the bioaccumulation mechanism of metals in *Daphnia magna* changed in the presence of $TiO_2$ NPs and the bioaccumulation of more than 85% of the tested metals increased in the presence of $TiO_2$ NPs.[166]

## 6. CHALLENGES IN THE IMPLEMENTATION OF ML AND POSSIBLE SOLUTIONS

In the present study, we investigated the application status of various technologies in the field of ML as applied to nanotoxicology. Compared to other domains where ML is more established, application of ML in nanotoxicology is still in the exploratory stage.[43] Most studies tend to extract data from literature and experiment with the data volumes ranging from a minimum of 15[129] to a maximum of 19,404.[105] As a rough estimation, about 80% of the modeling approaches in the studied literature employed less than 500 samples, which is much less than that in chemical or drug toxicity assessment (from thousands to hundreds of thousands).[167] This indicates that there is still a need to pay attention to the performance of ML models on small sample sizes. Few studies investigated the quality of the data set for modeling, with the consequence that poor data quality may impact model performance and interpretability.[168] Therefore, quality of the experimental data should be implemented to during data collection, e.g., NanoCRED, GUIDENano.[169] The relationship between the data volume and the model performance is also an interesting topic, although a small data volume does not necessarily cause poor performance, but a sufficiently large number of samples can make sure that the models perform better (Figure 3B).

The extraction of descriptors for modeling is closely related to the data set composition, especially for experimental descriptors. Within the literature collected, the descriptors that have been previously identified as important in experiments are often used in modeling, while newly developed descriptors are continually added to the models to test their performance. Meanwhile, the boundaries of descriptors are expanding along with the advancement of ML technologies, and many new types of descriptors have been developed and applied in experiments (e.g., graph presentation,[56] virtual molecular projections,[124] geometrical descriptors[170]), yielding promising results. However, the selection of model descriptors is currently largely dependent on empirical knowledge, which means subjective biases could occur. Additionally, relying on empirical knowledge may overlook potentially important features and useful descriptors. Therefore, it is advisible to adopt a more systematic and comprehensive approach to descriptor selection.

Within the reviewed studies, RF, SVM, and NN were abundantly used as compared to other algorithms (e.g., MLR, DT, kNN, NB), these ML algorithms often show good performance and robustness on small data sets, and they are at the same time easy to interpret.[112,171,172] However, the implementation of more advanced algorithms is currently limited by the paucity of available data, such as is the case for DL. Compared with other algorithms, DL requires more data and computational resources but less human intervention.[44] It is, thus, a challenge to determine whether DL is able to realize its immense potential in the prediction of (cyto)toxicity of NMs with the enrichment of NM-related data.

Despite being one of the OECD principles, the importance of the assessment of the AD is relatively less recognized. The definition of the AD has led to numerous approaches based on different hypotheses, among which distance-based methods are the most commonly used ones.[130,132] Researchers should acknowledge that a well-defined applicability domain is also a necessary requirement for modeling.

The mechanistic interpretation should be adapted to the needs of specific modeling approach and endpoint.[173] First it is important to have that process-based knowledge at the exposure-uptake interface is known and second how NMs are biodistributed, which target gives effects, and what the chain of effects is (i.e., adverse outcome pathways; AOPs). This knowledge is needed to have the correct descriptors in the databases vice versa helps to interpret findings. In future, an incorporation of molecular initiating events (MIEs), key events (KEs), and adverse outcomes (AO) into the AI models (https://aopkb.oecd.org/), this will further enable us to make better prediction of toxicity endpoints of untested NMs in future. This is useful information for (eco)toxicologists to use in for instance risk assessment,[41,174] and for NMs designers to use in the development of safer NMs. However, the expansion of the descriptor scope (e.g., structure information, exposure environment, test organisms) increases the difficulty in identifying critical features and explaining underlying mechanisms. Moreover, challenges exist in how to evaluate the levels of interpretability or to compare interpretability of different models[175,176] and how to do justified read across to untested NMs (especially think of third and fourth generations of NMs) and untested organisms including protected or endangered species.

## 7. RECOMMENDATIONS, BEST PRACTICES, AND OUTLOOK

Given the above interpretation of the collected results and its checks against OECD criteria, it is evident that several challenges need to be addressed before *in silico* tools can be accepted by regulators as an informative alternative to


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

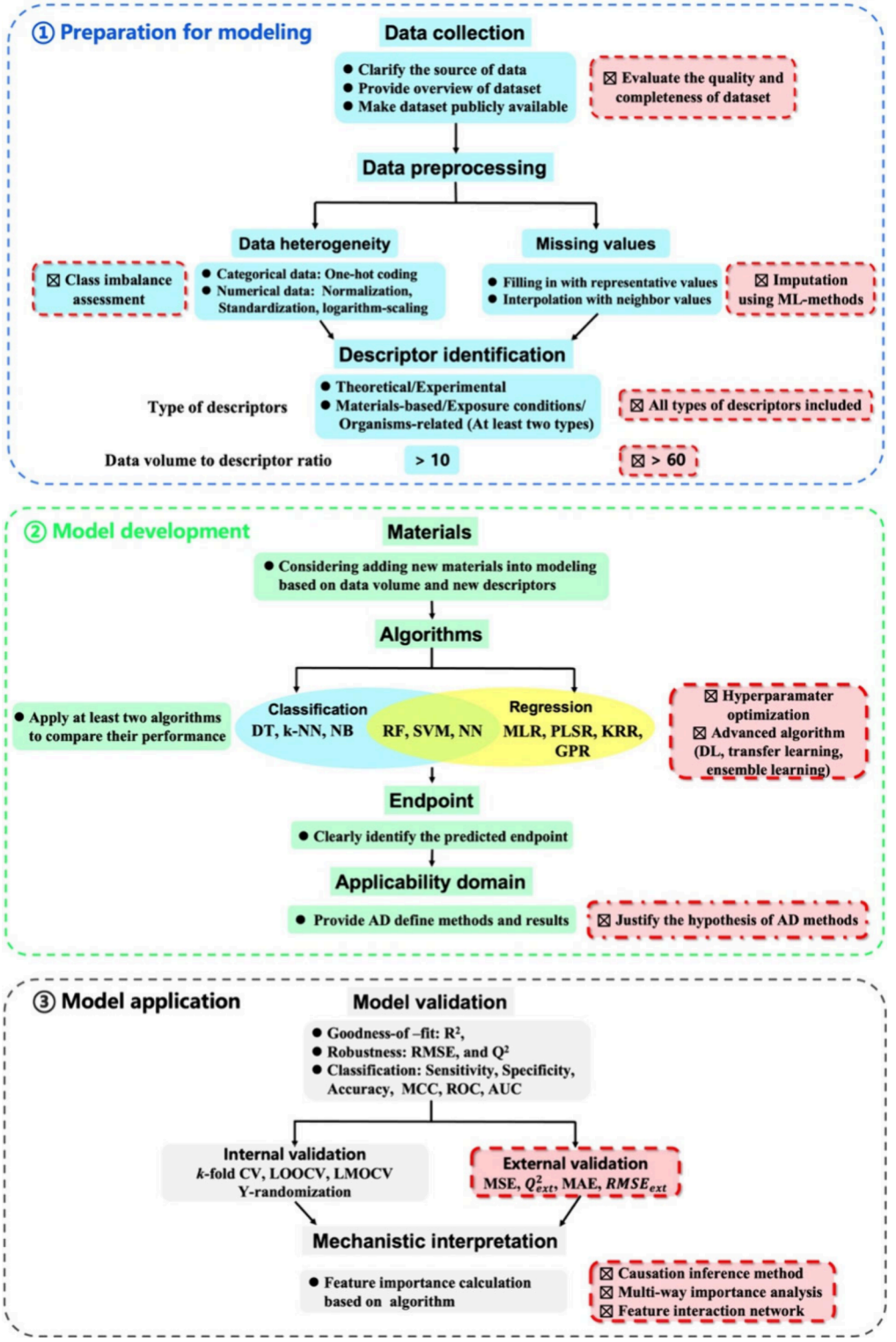



**Figure 6.** Recommended workflow and the best practices for the application of ML in nanotoxicology; best practices are indicated in the red dotted box.

conventional experimental toxicity testing. Therefore, we propose a recommended workflow as well as best practices for the application of ML in nanotoxicology (Figure 6).

Given the current limitations in data availability, the establishment of large curated databases that are specifically tailored for modeling in nanotoxicology remains a top priority for the development of relevant research. Considerable efforts are already being made in this regard (e.g., eNanomapper, NanoPharos and others listed in Table 1), in addition to making nanotoxicity data more FAIR (Findable, Accessible, Interoperable and Reusable). Natural language processing (e.g., large language models) could be leveraged in this context for the mining of data from published literature at a large scale to aid in the generation of curated data sets.[47,177,178] However,


https://doi.org/10.1021/acs.est.4c03328
Environ. Sci. Technol. 2024, 58, 14973−14993

more effort is needed in standardization of data formats, ontology, quality control measures, and the integration of diverse data types to enhance the reliability and validity of ML applications in the field. The reporting of data within the literature should also be improved upon, as the abiotic conditions (or other crucial parameters) used during experiments are frequently described inadequately, which complicates the use of such data for modeling. Reporting on models themselves also needs to be improved, in particular in terms of the choice of descriptors, the internal and external validation and the selection of the hypothesis used to determine the applicability domain.

When selecting descriptors, it is important to appropriately choose theoretical and experimental descriptors according to the following specific conditions. Size nor composition is enough to assess NMs properties and induced biological effects. Experimental descriptors should provide more information for the specific case-dependent outcome, while theoretical descriptors offer information for the representative population. Feature elimination during model training can help in reducing irrelevant descriptors in addition to getting a better descriptor combination which simplifies model interpretation.

With the increasing application of ML in nanotoxicology, researchers need not only to choose appropriate models based on the task type (classification or regression), but also need to be able to convert complex nanotoxicity problems into ones that can be handled by ML.[36] More advanced algorithms are worth trying in future studies, such as DL, multimodal data modeling, self-supervised pretraining modeling, etc. But also, algorithms that are more effective at using smaller data sets could be considered, such as few-shot, zero-shot and transfer learning.[128,179] Moreover, the authors call for increased model availability via user-friendly web applications and cloud platforms, such as that provided for the models developed in the NanoSolveIT project in which >50 nanoinformatics models were deployed (https://www.enaloscloud.novamechanics.com/nanosolveit.html). This involves assessing the feasibility and effectiveness of deploying ML models related to nanotoxicology on various cloud platforms, with a focus on the accessibility and compatibility of these platforms with specific computational needs and data structures. FAIRification of nanoinformatics models is also increasingly being recognized as a critical step in maximizing uptake (by industry and regulators) and reusability of models.[180]

Improving the interpretability of models is a necessary condition for regulatory acceptance, and computer literacy and expert knowledge in nanotoxicology are both indispensable for progress in this area. Considering the interdisciplinary nature, developing Graphical User Interfaces (GUI) that simplify the interaction with complex ML models for researchers in nanotoxicology who may not have in-depth computational expertise is critical.

The present study advocates that the development of more predictive and more reliable models is well possible with more advanced data mining and ML practices. Furthermore, combined with expert nanotoxicology knowledge, ML remains a powerful tool in unraveling the mechanisms of toxicity of NMs for their effective risk assessment management, and is poised to offer safe-by-design criteria for generating "greener" products of NMs, thereby enabling sustainable development of the nanotechnology.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c03328.

Overview of the literature information collected, including NMs category, number of models, data size and number of descriptors, three categories of important descriptors identified in reviewed studies, and the corresponding data acquisition method (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

**Ying Wang** − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China;* orcid.org/0000-0002-1424-0422; Email: yingw@buaa.deu.cn, wy2012bnu@126.com

**Wenhong Fan** − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China; Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing 100191, China;* orcid.org/0000-0001-5047-1714; Email: fanwh@buaa.edu.cn

### Authors

**Yunchi Zhou** − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China; Ecole Centrale de Pékin/School of General Engineering, Beihang University, Beijing 100191, China;* orcid.org/0000-0001-6603-2177

**Willie Peijnenburg** − *Institute of Environmental Science, Leiden University, Leiden 2300 RA, The Netherlands; National Institute of Public Health and the Environment, Center for Safety of Products and Substances, Bilthoven 3720BA, The Netherlands*

**Martina G. Vijver** − *Institute of Environmental Science, Leiden University, Leiden 2300 RA, The Netherlands*

**Surendra Balraadjsing** − *Institute of Environmental Science, Leiden University, Leiden 2300 RA, The Netherlands;* orcid.org/0000-0003-2167-8163

**Zhaomin Dong** − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China; Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing 100191, China;* orcid.org/0000-0002-0251-2355

**Xiaoli Zhao** − *State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China;* orcid.org/0000-0002-5671-389X

**Kenneth M. Y. Leung** − *State Key Laboratory of Marine Pollution and Department of Chemistry, City University of Hong Kong, Hong Kong 999077, China*

**Holly M. Mortensen** − *Public Health and Integrated Toxicology Division, Center for Public Health and Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, United States*

**Zhenyu Wang** − *Institute of Environmental Process and Pollution Control, School of Environment & Ecology, Jiangnan University, Wuxi 214122, China;* orcid.org/0000-0002-5114-435X

**Iseult Lynch** − *School of Geography, Earth and Environmental Sciences, University of Birmingham,*

*Edgbaston, Birmingham B15 2TT, United Kingdom;* orcid.org/0000-0003-4250-4584

**Antreas Afantitis** − *Department of ChemoInformatics, NovaMechanics Ltd., Nicosia 1046, Cyprus;* orcid.org/0000-0002-0977-8180

**Yunsong Mu** − *School of Environment & Natural Resources, Renmin University of China, Beijing 100872, China*

**Fengchang Wu** − *State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China;* orcid.org/0000-0001-8595-2243

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c03328

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 42330710 and 42177240), the National Key R&D Program of China (No. 2022YFC3204800), and the Fundamental Research Funds for the Central Universities. The authors of University Leiden (MGV, WJGMP, SB) are paid by ERC Consolidator grant EcoWizard (Grant Agreement No. 101002123). IL and AA acknowledge Horizon 2020 Marie Skłodowska-Curie Actions project CompSafeNano (Grant Agreement No. 101008099). KMYL acknowledges the support of Innovation and Technology Commission (ITC) of the Hong Kong SAR Government to the State Key Laboratory of Marine Pollution (City University of Hong Kong) (No. 9448002).

## ■ REFERENCES

(1) Hansen, S. F.; Hansen, O. F. H.; Nielsen, M. B. Advances and challenges towards consumerization of nanomaterials. *Nat. Nanotechnol* **2020**, *15* (12), 964−965.

(2) Hofmann-Amtenbrink, M.; Grainger, D. W.; Hofmann, H. Nanoparticles in medicine: Current challenges facing inorganic nanoparticle toxicity assessments and standardizations. *Nanomedicine: Nanotechnology, Biology and Medicine* **2015**, *11* (7), 1689−1694.

(3) Maynard, A. D.; Aitken, R. J.; Butz, T.; Colvin, V.; Donaldson, K.; Oberdörster, G.; Philbert, M. A.; Ryan, J.; Seaton, A.; Stone, V.; et al. Safe handling of nanotechnology. *Nature* **2006**, *444* (7117), 267−269.

(4) Nel, A.; Xia, T.; Mädler, L.; Li, N. Toxic Potential of Materials at the Nanolevel. *Science* **2006**, *311* (5761), 622−627.

(5) Sharma, V. K. Aggregation and toxicity of titanium dioxide nanoparticles in aquatic environment—A Review. *Journal of Environmental Science and Health, Part A* **2009**, *44* (14), 1485−1495.

(6) Gousiadou, C.; Marchese Robinson, R. L.; Kotzabasaki, M.; Doganis, P.; Wilkins, T. A.; Jia, X.; Sarimveis, H.; Harper, S. L. Machine learning predictions of concentration-specific aggregate hazard scores of inorganic nanomaterials in embryonic zebrafish. *Nanotoxicology* **2021**, *15* (4), 446−476.

(7) Ribeiro, F.; Gallego-Urrea, J. A.; Jurkschat, K.; Crossley, A.; Hassellöv, M.; Taylor, C.; Soares, A. M. V. M.; Loureiro, S. Silver nanoparticles and silver nitrate induce high toxicity to Pseudokirchneriella subcapitata, Daphnia magna and Danio rerio. *Science of The Total Environment* **2014**, *466−467*, 232−241.

(8) Park, S.-Y.; Choi, J.-H. Geno- and Ecotoxicity Evaluation of Silver Nanoparticles in Freshwater Crustacean Daphnia magna. *Environmental Engineering Research* **2010**, *15* (1), 23−27.

(9) Connolly, M.; Fernandez-Cruz, M. L.; Quesada-Garcia, A.; Alte, L.; Segner, H.; Navas, J. M. Comparative Cytotoxicity Study of Silver Nanoparticles (AgNPs) in a Variety of Rainbow Trout Cell Lines (RTL-W1, RTH-149, RTG-2) and Primary Hepatocytes. *International Journal of Environmental Research and Public Health* **2015**, *12* (5), 5386−5405.

(10) Pan, Y.; Neuss, S.; Leifert, A.; Fischler, M.; Wen, F.; Simon, U.; Schmid, G.; Brandau, W.; Jahnen-Dechent, W. Size-dependent cytotoxicity of gold nanoparticles. *Small* **2007**, *3* (11), 1941−1949.

(11) Jia, Y.-P.; Ma, B.-Y.; Wei, X.-W.; Qian, Z.-Y. The in vitro and in vivo toxicity of gold nanoparticles. *Chin. Chem. Lett.* **2017**, *28* (4), 691−702.

(12) Gatoo, M. A.; Naseem, S.; Arfat, M. Y.; Mahmood Dar, A.; Qasim, K.; Zubair, S. Physicochemical Properties of Nanomaterials: Implication in Associated Toxic Manifestations. *BioMed. Research International* **2014**, *2014*, 498420.

(13) Chen, R.; Qiao, J.; Bai, R.; Zhao, Y.; Chen, C. Intelligent testing strategy and analytical techniques for the safety assessment of nanomaterials. *Anal. Bioanal. Chem.* **2018**, *410* (24), 6051−6066.

(14) Roco, M. C. The long view of nanotechnology development: the National Nanotechnology Initiative at 10 years. *J. Nanopart. Res.* **2011**, *13* (2), 427−445.

(15) Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C. A. Practices and Trends of Machine Learning Application in Nanotoxicology. *Nanomaterials* **2020**, *10* (1), 116.

(16) Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnol.* **2011**, *6* (3), 175−178.

(17) 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. *Science of The Total Environment* **2021**, *793*, 148532.

(18) Sengottiyan, S.; Mikolajczyk, A.; Jagiełło, K.; Swirog, M.; Puzyn, T. Core, Coating, or Corona? The Importance of Considering Protein Coronas in nano-QSPR Modeling of Zeta Potential. *ACS Nano* **2023**, *17*, 1989.

(19) Juganson, K.; Ivask, A.; Blinova, I.; Mortimer, M.; Kahru, A. NanoE-Tox: New and in-depth database concerning ecotoxicity of nanomaterials. *Beilstein J. Nanotechnol* **2015**, *6*, 1788−1804.

(20) Zhou, Y.; Wang, Y.; Peijnenburg, W.; Vijver, M. G.; Balraadjsing, S.; Fan, W. Using Machine Learning to Predict Adverse Effects of Metallic Nanomaterials to Various Aquatic Organisms. *Environ. Sci. Technol.* **2023**, *57*, 17786.

(21) Wyrzykowska, E.; Mikolajczyk, A.; Lynch, I.; Jeliazkova, N.; Kochev, N.; Sarimveis, H.; Doganis, P.; Karatzas, P.; Afantitis, A.; Melagraki, G.; et al. Representing and describing nanomaterials in predictive nanoinformatics. *Nat. Nanotechnol.* **2022**, *17* (9), 924−932.

(22) Oomen, A. G.; Bleeker, E. A. J.; Bos, P. M. J.; Van Broekhuizen, F.; Gottardo, S.; Groenewold, M.; Hristozov, D.; Hund-Rinke, K.; Irfan, M.-A.; Marcomini, A.; et al. Grouping and Read-Across Approaches for Risk Assessment of Nanomaterials. *International Journal of Environmental Research and Public Health* **2015**, *12* (10), 13415−13434.

(23) 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. *Ecotoxicology and Environmental Safety* **2019**, *185*, 109733.

(24) Wigger, H.; Nowack, B. Material-specific properties applied to an environmental risk assessment of engineered nanomaterials - implications on grouping and read-across concepts. *Nanotoxicology* **2019**, *13* (5), 623−643.

(25) Gajewicz, A. What if the number of nanotoxicity data is too small for developing predictive Nano-QSAR models? An alternative read-across based approach for filling data gaps. *Nanoscale* **2017**, *9* (24), 8435−8448.

(26) OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; OECD, 2014. DOI: 10.1787/9789264085442-en.

(27) ECHA. Guidance on information requirements and chemical safety assessment. In *Guidance Documents*; ECHA, 2008.

(28) ECETOC. *Technical Report 116 Category approaches, read-across, (Q)SAR*; 2012. https://www.ecetoc.org/publication/tr-116-category-approaches-read-across-qsar/.
(29) Banerjee, A.; Kar, S.; Pore, S.; Roy, K. Efficient predictions of cytotoxicity of $TiO_2$-based multi-component nanoparticles using a machine learning-based q-RASAR approach. *Nanotoxicology* **2023**, *17* (1), 78−93.
(30) Chatterjee, M.; Banerjee, A.; De, P.; Gajewicz-Skretna, A.; Roy, K. A novel quantitative read-across tool designed purposefully to fill the existing gaps in nanosafety data. *Environmental Science-Nano* **2022**, *9* (1), 189−203.
(31) Gajewicz, A. How to judge whether QSAR/read-across predictions can be trusted: a novel approach for establishing a model's applicability domain. *Environmental Science: Nano* **2018**, *5* (2), 408−421.
(32) 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. *Ecotoxicology and Environmental Safety* **2019**, *185*, 109733.
(33) Patlewicz, G.; Lizarraga, L. E.; Rua, D.; Allen, D. G.; Daniel, A. B.; Fitzpatrick, S. C.; Garcia-Reyero, N.; Gordon, J.; Hakkinen, P.; Howard, A. S.; et al. Exploring current read-across applications and needs among selected U.S. Federal Agencies. *Regul. Toxicol. Pharmacol.* **2019**, *106*, 197−209.
(34) Singh, A. V.; Ansari, M. H. D.; Rosenkranz, D.; Maharjan, R. S.; Kriegel, F. L.; Gandhi, K.; Kanase, A.; Singh, R.; Laux, P.; Luch, A. Artificial Intelligence and Machine Learning in Computational Nanotoxicology: Unlocking and Empowering Nanomedicine. *Adv. Healthcare Mater.* **2020**, *9* (17), No. 1901862, DOI: 10.1002/adhm.201901862.
(35) Brown, K. A.; Brittman, S.; Maccaferri, N.; Jariwala, D.; Celano, U. Machine Learning in Nanoscience: Big Data at Small Scales. *Nano Lett.* **2020**, *20* (1), 2−10.
(36) Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J. G.; Gu, A.; Li, B.; Ma, X.; Marrone, B. L.; Ren, Z. J.; Schrier, J.; et al. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environ. Sci. Technol.* **2021**, *55* (19), 12741−12754.
(37) Fjodorova, N.; Novic, M.; Gajewicz, A.; Rasulev, B. The way to cover prediction for cytotoxicity for all existing nano-sized metal oxides by using neural network method. *Nanotoxicology* **2017**, *11* (4), 475−483.
(38) Jia, X.; Wang, T.; Zhu, H. Advancing Computational Toxicology by Interpretable Machine Learning. *Environ. Sci. Technol.* **2023**, *57*, 17690.
(39) Kuz'min, V. E.; Polishchuk, P. G.; Artemenko, A. G.; Andronati, S. A. Interpretation of QSAR Models Based on Random Forest Methods. *Molecular Informatics* **2011**, *30* (6−7), 593−603.
(40) Marchese Robinson, R. L.; Palczewska, A.; Palczewski, J.; Kidley, N. Comparison of the Predictive Performance and Interpretability of Random Forest and Linear Models on Benchmark Data Sets. *J. Chem. Inf Model* **2017**, *57* (8), 1773−1792.
(41) Yu, F. B.; Wei, C. H.; Deng, P.; Peng, T.; Hu, X. G. Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. *Science Advances* **2021**, *7* (22), No. eabf4130, DOI: 10.1126/sciadv.abf4130.
(42) Winkler, D. A.; Mombelli, E.; Pietroiusti, A.; Tran, L.; Worth, A.; Fadeel, B.; McCall, M. J. Applying quantitative structure-activity relationship approaches to nanotoxicology: Current status and future potential. *Toxicology* **2013**, *313* (1), 15−23.
(43) Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C. A. Nanotoxicology data for in silico tools: a literature review. *Nanotoxicology* **2020**, *14* (5), 612−637.
(44) Ji, Z.; Guo, W.; Wood, E. L.; Liu, J.; Sakkiah, S.; Xu, X.; Patterson, T. A.; Hong, H. Machine Learning Models for Predicting Cytotoxicity of Nanomaterials. *Chem. Res. Toxicol.* **2022**, *35* (2), 125−139.
(45) Basei, G.; Hristozov, D.; Lamon, L.; Zabeo, A.; Jeliazkova, N.; Tsiliki, G.; Marcomini, A.; Torsello, A. Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of in silico tools: A critical review. *NanoImpact* **2019**, *13*, 76−99.
(46) Li, J.; Wang, C.; Yue, L.; Chen, F.; Cao, X.; Wang, Z. Nano-QSAR modeling for predicting the cytotoxicity of metallic and metal oxide nanoparticles: A review. *Ecotoxicology and Environmental Safety* **2022**, *243*, 113955.
(47) Yan, X. L.; Yue, T. T.; Winkler, D. A.; Yin, Y. G.; Zhu, H.; Jiang, G. B.; Yan, B. Converting Nanotoxicity Data to Information Using Artificial Intelligence and Simulation. *Chem. Rev.* **2023**, *123* (13), 8575−8637.
(48) Chen, C.; Hu, Z.; Liu, S.; Tseng, H. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. *Expert Opinion on Biological Therapy* **2012**, *12* (5), 593−608.
(49) Peng, T.; Wei, C.; Yu, F.; Xu, J.; Zhou, Q.; Shi, T.; Hu, X. Predicting nanotoxicity by an integrated machine learning and metabolomics approach. *Environ. Pollut.* **2020**, *267*, 115434.
(50) Papadiamantis, A. G.; Janes, J.; Voyiatzis, E.; Sikk, L.; Burk, J.; Burk, P.; Tsoumanis, A.; Ha, M. K.; Yoon, T. H.; Valsami-Jones, E. Predicting Cytotoxicity of Metal Oxide Nanoparticles Using Isalos Analytics Platform. *Nanomaterials* **2020**, *10* (10), 2017.
(51) Choi, J.-S.; Ha, M. K.; Trinh, T. X.; Yoon, T. H.; Byun, H.-G. Towards a generalized toxicity prediction model for oxide nanomaterials using integrated data from different sources. *Sci. Rep.* **2018**, *8* (1), 6110.
(52) Choi, J.-S.; Trinh, T. X.; Yoon, T.-H.; Kim, J.; Byun, H.-G. Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. *Chemosphere* **2019**, *217*, 243−249.
(53) Kovalishyn, V.; Abramenko, N.; Kopernyk, I.; Charochkina, L.; Metelytsia, L.; Tetko, I. V.; Peijnenburg, W.; Kustov, L. Modelling the toxicity of a large set of metal and metal oxide nanoparticles using the OCHEM platform. *Food Chem. Toxicol.* **2018**, *112*, 507−517.
(54) Loh, W.-Y. Classification and Regression Trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2011**, *1*, 14−23.
(55) Romeo, D.; Louka, C.; Gudino, B.; Wigstrom, J.; Wick, P. Structure-activity relationship of graphene-related materials: A meta-analysis based on mammalian in vitro toxicity data. *Nanoimpact* **2022**, *28*, 100436.
(56) Liu, R.; Lin, S.; Rallo, R.; Zhao, Y.; Damoiseaux, R.; Xia, T.; Lin, S.; Nel, A.; Cohen, Y. Automated Phenotype Recognition for Zebrafish Embryo Based In Vivo High Throughput Toxicity Screening of Engineered Nano-Materials. *PLOS One* **2012**, *7* (4), No. e35014, DOI: 10.1371/journal.pone.0035014.
(57) Wang, T.; Russo, D. P.; Bitounis, D.; Demokritou, P.; Jia, X.; Huang, H.; Zhu, H. Integrating structure annotation and machine learning approaches to develop graphene toxicity models. *Carbon* **2023**, *204*, 484−494.
(58) Kar, S.; Pathakoti, K.; Tchounwou, P. B.; Leszczynska, D.; Leszczynski, J. Evaluating the cytotoxicity of a large pool of metal oxide nanoparticles to *Escherichia coli*: Mechanistic understanding through *In Vitro* and *In Silico* studies. *Chemosphere* 2021; Vol. *264*. DOI: 10.1016/j.chemosphere.2020.128428.
(59) Balraadjsing, S.; Peijnenburg, W. J. G. M.; Vijver, M. G. Exploring the potential of in silico machine learning tools for the prediction of acute Daphnia magna nanotoxicity. *Chemosphere* **2022**, *307*, 135930.
(60) Zhu, J.-J.; Yang, M.; Ren, Z. J. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environ. Sci. Technol.* **2023**, *57* (46), 17671−17689.
(61) Sizochenko, N.; Mikolajczyk, A.; Jagiello, K.; Puzyn, T.; Leszczynski, J.; Rasulev, B. How the toxicity of nanomaterials towards different species could be simultaneously evaluated: a novel multi-nano-read-across approach. *Nanoscale* **2018**, *10* (2), 582−591.
(62) Waljee, A. K.; Mukherjee, A.; Singal, A. G.; Zhang, Y.; Warren, J.; Balis, U.; Marrero, J.; Zhu, J.; Higgins, P. D. Comparison of

imputation methods for missing laboratory data in medicine. *BMJ. Open* **2013**, *3* (8), e002847.
(63) Marchese Robinson, R. L.; Lynch, I.; Peijnenburg, W.; Rumble, J.; Klaessig, F.; Marquardt, C.; Rauscher, H.; Puzyn, T.; Purian, R.; Aberg, C.; et al. How should the completeness and quality of curated nanomaterial data be evaluated? *Nanoscale* **2016**, *8* (19), 9919−9943.
(64) Lubinski, L.; Urbaszek, P.; Gajewicz, A.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Leszczynska, D.; Leszczynski, J.; Puzyn, T. Evaluation criteria for the quality of published experimental data on nanomaterials and their usefulness for QSAR modelling. *SAR and QSAR in Environmental Research* **2013**, *24* (12), 995−1008.
(65) Trinh, T. X.; Ha, M. K.; Choi, J. S.; Byun, H. G.; Yoon, T. H. Curation of datasets, assessment of their quality and completeness, and nanoSAR classification model development for metallic nanoparticles. *Environmental Science-Nano* **2018**, *5* (8), 1902−1910.
(66) Basei, G.; Rauscher, H.; Jeliazkova, N.; Hristozov, D. A methodology for the automatic evaluation of data quality and completeness of nanomaterials for risk assessment purposes. *Nanotoxicology* **2022**, *16* (2), 195−216.
(67) Hoecke, K. v.; Quik, J. T. K.; Mankiewicz-Boczek, J.; Schamphelaere, K. A. C. d.; Elsaesser, A.; Meeren, P. v. d.; Barnes, C.; McKerr, G.; Howard, C. V.; Van De Meent, D.; et al. Fate and Effects of $CeO_2$ Nanoparticles in Aquatic Ecotoxicity Tests. *Environ. Sci. Technol.* **2009**, *43* (12), 4537−4546.
(68) Mikolajczyk, A.; Gajewicz, A.; Rasulev, B.; Schaeublin, N.; Maurer-Gardner, E.; Hussain, S.; Leszczynski, J.; Puzyn, T. Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure-Property Relationship Approach. *Chem. Mater.* **2015**, *27* (7), 2400−2407.
(69) Jo, H. J.; Choi, J. W.; Lee, S. H.; Hong, S. W. Acute toxicity of Ag and CuO nanoparticle suspensions against Daphnia magna: The importance of their dissolved fraction varying with preparation methods. *Journal of Hazardous Materials* **2012**, *227*, 301−308.
(70) Stewart, J. J. P. *MOPAC2016*; Stewart Computational Chemistry: Colorado Springs,CO, 2016.
(71) Sushko, I.; Novotarskyi, S.; Koerner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y.; et al. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *Journal of Computer-Aided Molecular Design* **2011**, *25* (6), 533−554.
(72) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31−36.
(73) Leone, C.; Bertuzzi, E. E.; Toropova, A. P.; Toropov, A. A.; Benfenati, E. CORAL: Predictive models for cytotoxicity of functionalized nanozeolites based on quasi-SMILES. *Chemosphere* **2018**, *210*, 52−56.
(74) Roy, J.; Roy, K. Evaluating metal oxide nanoparticle (MeOx NP) toxicity with different types of nano descriptors mainly focusing on simple periodic table-based descriptors: a mini-review. *Environmental Science: Nano* **2023**, *10* (11), 2989−3011.
(75) Huang, Y.; Li, X.; Cao, J.; Wei, X.; Li, Y.; Wang, Z.; Cai, X.; Li, R.; Chen, J. Use of dissociation degree in lysosomes to predict metal oxide nanoparticle toxicity in immune cells: Machine learning boosts nano-safety assessment. *Environ. Int.* **2022**, *164*, 107258.
(76) Huang, Y.; Li, X.; Xu, S.; Zheng, H.; Zhang, L.; Chen, J.; Hong, H.; Kusko, R.; Li, R. Quantitative Structure-Activity Relationship Models for Predicting Inflammatory Potential of Metal Oxide Nanoparticles. *Environ. Health Perspect.* **2020**, *128* (6) DOI: 10.1289/EHP6508.
(77) Regonia, P. R.; Olorocisimo, J. P.; De los Reyes, F.; Ikeda, K.; Pelicano, C. M. Machine learning-enabled nanosafety assessment of multi-metallic alloy nanoparticles modified $TiO_2$ system. *NanoImpact* **2022**, *28*, 100442.
(78) Racz, A.; Dunn, T. B.; Bajusz, D.; Kim, T. D.; Miranda-Quintana, R. A.; Heberger, K. Extended continuous similarity indices: theory and application for QSAR descriptor selection. *Journal of Computer-Aided Molecular Design* **2022**, *36* (3), 157−173.
(79) Yan, X.; Yue, T.; Zhu, H.; Yan, B. Bridging the Gap Between Nanotoxicological Data and the Critical Structure-Activity Relationships. In *Advances in Toxicology and Risk Assessment of Nanomaterials and Emerging Contaminants*, Guo, L.-H., Mortimer, M., Eds.; Springer Singapore, 2022; pp 161−183.
(80) Trinh, T. X.; Seo, M.; Yoon, T. H.; Kim, J. Developing random forest based QSAR models for predicting the mixture toxicity of $TiO_2$ based nano-mixtures to Daphnia magna. *Nanoimpact* **2022**, *25*, 100383.
(81) Svendsen, C.; Walker, L. A.; Matzke, M.; Lahive, E.; Harrison, S.; Crossley, A.; Park, B.; Lofts, S.; Lynch, I.; Vázquez-Campos, S.; et al. Key principles and operational practices for improved nanotechnology environmental exposure assessment. *Nat. Nanotechnol.* **2020**, *15* (9), 731−742.
(82) Lynch, I.; Afantitis, A.; Leonis, G.; Melagraki, G.; Valsami-Jones, E. Strategy for Identification of Nanomaterials' Critical Properties Linked to Biological Impacts: Interlinking of Experimental and Computational Approaches. In *Advances in QSAR Modeling: Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences*, Roy, K., Ed.; Springer International Publishing, 2017; pp 385−424.
(83) Lynch, I.; Weiss, C.; Valsami-Jones, E. A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. *Nano Today* **2014**, *9* (3), 266−270.
(84) Sizochenko, N.; Gajewicz, A.; Leszczynski, J.; Puzyn, T. Causation or only correlation? Application of causal inference graphs for evaluating causality in nano-QSAR models. *Nanoscale* **2016**, *8* (13), 7203−7208.
(85) Varsou, D.-D.; Ellis, L.-J. A.; Afantitis, A.; Melagraki, G.; Lynch, I. Ecotoxicological read-across models for predicting acute toxicity of freshly dispersed versus medium-aged NMs to Daphnia magna. *Chemosphere* **2021**, *285*, 131452.
(86) Fan, W.; Shi, Z.; Yang, X.; Cui, M.; Wang, X.; Zhang, D.; Liu, H.; Guo, L. Bioaccumulation and biomarker responses of cubic and octahedral Cu2O micro/nanocrystals in Daphnia magna. *Water Res.* **2012**, *46* (18), 5981−5988.
(87) Fan, W.; Wang, X.; Cui, M.; Zhang, D.; Zhang, Y.; Yu, T.; Guo, L. Differential Oxidative Stress of Octahedral and Cubic Cu2O Micro/Nanocrystals to Daphnia magna. *Environ. Sci. Technol.* **2012**, *46* (18), 10255−10262.
(88) Fan, W.; Lu, H.; Wang, W.-X. Aging Influences on the Biokinetics of Functional $TiO_2$ Nanoparticles with Different Surface Chemistries in Daphnia magna. *Environ. Sci. Technol.* **2018**, *52* (14), 7901−7909.
(89) Na, M.; Nam, S. H.; Moon, K.; Kim, J. Development of a nano-QSAR model for predicting the toxicity of nano-metal oxide mixtures to Aliivibrio fischeri. *Environmental Science-Nano* **2023**, *10* (1), 325−337.
(90) Connell, D. W.; Yu, Q. J.; Verma, V. Influence of exposure time on toxicity—An overview. *Toxicology* **2016**, *355*−*356*, 49−53.
(91) Lu, H. T.; Dong, H. F.; Fan, W. H.; Zuo, J. X.; Li, X. M. Aging and behavior of functional $TiO_2$ nanoparticles in aqueous environment. *Journal of Hazardous Materials* **2017**, *325*, 113−119.
(92) Sengottiyan, S.; Mikolajczyk, A.; Jagiełło, K.; Swirog, M.; Puzyn, T. Core, Coating, or Corona? The Importance of Considering Protein Coronas in nano-QSPR Modeling of Zeta Potential. *ACS Nano* **2023**, *17* (3), 1989−1997.
(93) Mittelman, A. M.; Fortner, J. D.; Pennell, K. D. Effects of ultraviolet light on silver nanoparticle mobility and dissolution. *Environmental Science: Nano* **2015**, *2* (6), 683−691.
(94) Souza, I. R.; Silva, L. R.; Fernandes, L. S. P.; Salgado, L. D.; Silva de Assis, H. C.; Firak, D. S.; Bach, L.; Santos-Filho, R.; Voigt, C. L.; Barros, A. C.; et al. Visible-light reduced silver nanoparticles' toxicity in Allium cepa test system. *Environ. Pollut.* **2020**, *257*, 113551.
(95) Lu, H. T.; Fan, W. H.; Dong, H. F.; Liu, L. L. Dependence of the irradiation conditions and crystalline phases of $TiO_2$ nanoparticles on their toxicity to Daphnia magna. *Environmental Science-Nano* **2017**, *4* (2), 406−414.

(96) Johnson, T. N.; Ke, A. B. Physiologically Based Pharmacokinetic Modeling and Allometric Scaling in Pediatric Drug Development: Where Do We Draw the Line? *J. Clin Pharmacol* **2021**, *61* (S1), S83−S93.
(97) Tsiros, P.; Cheimarios, N.; Tsoumanis, A.; Jensen, A. C. Ø.; Melagraki, G.; Lynch, I.; Sarimveis, H.; Afantitis, A. Towards an in silico integrated approach for testing and assessment of nanomaterials: from predicted indoor air concentrations to lung dose and biodistribution. *Environmental Science: Nano* **2022**, *9* (4), 1282−1297.
(98) Song, L.; Vijver, M. G.; de Snoo, G. R.; Peijnenburg, W. J. G. M. Assessing toxicity of copper nanoparticles across five cladoceran species. *Environ. Toxicol. Chem.* **2015**, *34* (8), 1863−1869.
(99) del Giudice, G.; Serra, A.; Saarimäki, L. A.; Kotsis, K.; Rouse, I.; Colibaba, S. A.; Jagiello, K.; Mikolajczyk, A.; Fratello, M.; Papadiamantis, A. G.; et al. An ancestral molecular response to nanomaterial particulates. *Nat. Nanotechnol.* **2023**, *18* (8), 957−966.
(100) Dixit, R.; Wasiullah; Malaviya, D.; Pandiyan, K.; Singh, U. B.; Sahu, A.; Shukla, R.; Singh, B. P.; Rai, J. P.; Sharma, P. K.; et al. Bioremediation of Heavy Metals from Soil and Aquatic Environment: An Overview of Principles and Criteria of Fundamental Processes. *Sustainability* **2015**, *7* (2), 2189−2212.
(101) Alexander, D. L.; Tropsha, A.; Winkler, D. A. Beware of R(2): Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf Model* **2015**, *55* (7), 1316−1322.
(102) Cho, W.-S.; Thielbeer, F.; Duffin, R.; Johansson, E. M. V.; Megson, I. L.; MacNee, W.; Bradley, M.; Donaldson, K. Surface functionalization affects the zeta potential, coronal stability and membranolytic activity of polymeric nanoparticles. *Nanotoxicology* **2014**, *8* (2), 202−211.
(103) Zhang, F.; Wang, Z.; Peijnenburg, W. J. G. M.; Vijver, M. G. Review and Prospects on the Ecotoxicity of Mixtures of Nanoparticles and Hybrid Nanomaterials. *Environ. Sci. Technol.* **2022**, *56* (22), 15238−15250.
(104) Sang, L.; Wang, Y.; Zong, C.; Wang, P.; Zhang, H.; Guo, D.; Yuan, B.; Pan, Y. Machine Learning for Evaluating the Cytotoxicity of Mixtures of Nano-$TiO_2$ and Heavy Metals: QSAR Model Apply Random Forest Algorithm after Clustering Analysis. *Molecules* **2022**, *27* (18), 6125.
(105) Li, S.; Barnard, A. S. Safety-by-design using forward and inverse multi-target machine learning. *Chemosphere* **2022**, *303*, 135033.
(106) Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C. A. Practices and Trends of Machine Learning Application in Nanotoxicology. *Nanomaterials* **2020**, *10* (1), 116.
(107) Wilson, M. D. Support Vector Machines. In *Encyclopedia of Ecology*, Jørgensen, S. E., Fath, B. D., Eds.; Academic Press, 2008; pp 3431−3437.
(108) Sizochenko, N.; Mikolajczyk, A.; Syzochenko, M.; Puzyn, T.; Leszczynski, J. Zeta potentials ($\zeta$) of metal oxide nanoparticles: A meta-analysis of experimental data and a predictive neural networks modeling. *NanoImpact* **2021**, *22*, 100317.
(109) Xu, Y.; Pei, J.; Lai, L. Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J. Chem. Inf. Model.* **2017**, *57* (11), 2672−2685.
(110) Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: Toxicity Prediction using Deep Learning. *Frontiers in Environmental Science* **2016**, *3*, DOI: 10.3389/fenvs.2015.00080.
(111) Luan, F.; Kleandrova, V. V.; Gonzalez-Diaz, H.; Ruso, J. M.; Melo, A.; Speck-Planche, A.; Cordeiro, M. Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale* **2014**, *6* (18), 10623−10630.
(112) Labouta, H. I.; Asgarian, N.; Rinker, K.; Cramb, D. T. Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature. *ACS Nano* **2019**, *13* (2), 1583−1594.
(113) Mikolajczyk, A.; Sizochenko, N.; Mulkiewicz, E.; Malankowska, A.; Rasulev, B.; Puzyn, T. A chemoinformatics approach for the characterization of hybrid nanomaterials: safer and efficient design perspective. *Nanoscale* **2019**, *11* (24), 11808−11818.
(114) Melagraki, G.; Afantitis, A. Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles. *RSC Adv.* **2014**, *4* (92), 50713−50725.
(115) Webb, G. I.; Keogh, E.; Miikkulainen, R. Naïve Bayes. *Encyclopedia of machine learning* **2011**, *15* (1), 713−714.
(116) Simeone, F. C.; Costa, A. L. Assessment of cytotoxicity of metal oxide nanoparticles on the basis of fundamental physical-chemical parameters: a robust approach to grouping. *Environmental Science-Nano* **2019**, *6* (10), 3102−3112.
(117) Zhou, Z.; Tang, X.; Dai, W.; Shi, J.; Chen, H. Nano-QSAR models for predicting cytotoxicity of metal oxide nanoparticles (MONPs) to E. coli. *Can. J. Chem.* **2017**, *95* (8), 863−866.
(118) Roy, J.; Roy, K. Assessment of toxicity of metal oxide and hydroxide nanoparticles using the QSAR modeling approach. *Environmental Science-Nano* **2021**, *8* (11), 3395−3407.
(119) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58* (2), 109−130.
(120) Buglak, A. A.; Zherdev, A. V.; Dzantiev, B. B. Nano-(Q)SAR for Cytotoxicity Prediction of Engineered Nanomaterials. *Molecules* **2019**, *24* (24), 4537.
(121) Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.
(122) Slavakis, K.; Bouboulis, P.; Theodoridis, S. Online Learning in Reproducing Kernel Hilbert Spaces. In *Academic Press Library in Signal Processing*; Diniz, P. S. R., Suykens, J. A. K., Chellappa, R., Theodoridis, S., Eds.; Elsevier, 2014; Vol. *1*, pp 883−987.
(123) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; The MIT Press, 2005. DOI: 10.7551/mitpress/3206.001.0001.
(124) Russo, D. P.; Yan, X.; Shende, S.; Huang, H.; Yan, B.; Zhu, H. Virtual Molecular Projections and Convolutional Neural Networks for the End-to-End Modeling of Nanoparticle Activities and Properties. *Anal. Chem.* **2020**, *92* (20), 13971−13979.
(125) Du, M. N.; Liu, N. H.; Hu, X. Techniques for Interpretable Machine Learning. *Communications of the ACM* **2019**, *63* (1), 68−77.
(126) Olden, J. D.; Jackson, D. A. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling* **2002**, *154* (1−2), 135−150.
(127) Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proceedings of the Ieee* **2021**, *109* (1), 43−76.
(128) Zhong, S.; Zhang, Y.; Zhang, H. Machine Learning-Assisted QSAR Models on Contaminant Reactivity Toward Four Oxidants: Combining Small Data Sets and Knowledge Transfer. *Environ. Sci. Technol.* **2022**, *56* (1), 681−692.
(129) Kotzabasaki, M.; Sotiropoulos, I.; Charitidis, C.; Sarimveis, H. Machine learning methods for multi-walled carbon nanotubes (MWCNT) genotoxicity prediction. *Nanoscale Advances* **2021**, *3* (11), 3167−3176.
(130) Kar, S.; Roy, K.; Leszczynski, J. Applicability domain: a step toward confident predictions and decidability for QSAR modeling. *Computational toxicology: Methods and protocols* **2018**, *1800*, 141−169.
(131) Kotzabasaki, M. I.; Sotiropoulos, I.; Sarimveis, H. QSAR modeling of the toxicity classification of superparamagnetic iron oxide nanoparticles (SPIONs) in stem-cell monitoring applications: an integrated study from data curation to model development. *Rsc Advances* **2020**, *10* (9), 5385−5391.
(132) Gadaleta, D.; Mangiatordi, G. F.; Catto, M.; Carotti, A.; Nicolotti, O. Applicability domain for QSAR models: where theory meets reality. *International journal of quantitative structure-property relationships (IJQSPR)* **2016**, *1* (1), 45−63.
(133) Wang, Y.; Xu, Z. H.; Rume, T.; Li, X. M.; Fan, W. H. Predicting and comparing chronic water quality criteria from

physicochemical properties of transition metals. *Chemosphere* **2020**, *244*, 125465.
(134) Wang, Y.; Wu, F.; Mu, Y.; Zeng, E. Y.; Meng, W.; Zhao, X.; Giesy, J. P.; Feng, C.; Wang, P.; Liao, H.; et al. Directly Predicting Water Quality Criteria from Physicochemical Properties of Transition Metals. *Sci. Rep.* **2016**, *6* (1), 22515.
(135) Sheridan, R. P.; Feuston, B. P.; Maiorov, V. N.; Kearsley, S. K. Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (6), 1912−1928.
(136) Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *Alternatives to laboratory animals* **2005**, *33* (5), 445−459.
(137) Devillers, O. An introduction to randomization in computational geometry. *Theoretical Computer Science* **1996**, *157* (1), 35−52.
(138) Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems* **2015**, *145*, 22−29.
(139) Shin, H. K.; Kim, S.; Yoon, S. Use of size-dependent electron configuration fingerprint to develop general prediction models for nanomaterials. *Nanoimpact* **2021**, *21*, 100298.
(140) Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Statistics Surveys* **2010**, *4* (none), 40−79.
(141) Rücker, C.; Rücker, G.; Meringer, M. y-Randomization and Its Variants in QSPR/QSAR. *J. Chem. Inf. Model.* **2007**, *47* (6), 2345−2357.
(142) Bleeker, S. E.; Moll, H. A.; Steyerberg, E. W.; Donders, A. R. T.; Derksen-Lubsen, G.; Grobbee, D. E.; Moons, K. G. M. External validation is necessary in prediction research:: A clinical example. *Journal of Clinical Epidemiology* **2003**, *56* (9), 826−832.
(143) Wang, Y.; Wu, F.; Liu, Y.; Mu, Y.; Giesy, J. P.; Meng, W.; Hu, Q.; Liu, J.; Dang, Z. Effect doses for protection of human health predicted from physicochemical properties of metals/metalloids. *Environ. Pollut.* **2018**, *232*, 458−466.
(144) Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405* (2), 442−451.
(145) Osei-Bryson, K.-M. Evaluation of decision trees: a multi-criteria approach. *Computers & Operations Research* **2004**, *31* (11), 1933−1945.
(146) Pfaff, T. J. Area Under Curves. In *Applied Calculus with R;* Pfaff, T. J., Ed.; Springer International Publishing, 2023; pp 317−329.
(147) Visani, G.; Bagli, E.; Chesani, F.; Poluzzi, A.; Capuzzo, D. Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. *Journal of the Operational Research Society* **2022**, *73* (1), 91−101.
(148) Ekanayake, I. U.; Meddage, D. P. P.; Rathnayake, U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Studies in Construction Materials* **2022**, *16*, e01059.
(149) Mangalathu, S.; Hwang, S.-H.; Jeon, J.-S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures* **2020**, *219*, 110927.
(150) Greenwell, B. M. pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal* **2017**, *9*, 421−436, DOI: 10.32614/RJ-2017-016.
(151) Palatnik de Sousa, I.; Maria Bernardes Rebuzzi Vellasco, M.; Costa da Silva, E. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. *Sensors* **2019**, *19* (13), 2969.
(152) Liu, Y.; Vijver, M. G.; Qiu, H.; Baas, J.; Peijnenburg, W. J. G. M. Statistically significant deviations from additivity: What do they mean in assessing toxicity of mixtures? *Ecotoxicology and Environmental Safety* **2015**, *122*, 37−44.
(153) Limbach, L. K.; Wick, P.; Manser, P.; Grass, R. N.; Bruinink, A.; Stark, W. J. Exposure of engineered nanoparticles to human lung epithelial cells: Influence of chemical composition and catalytic activity on oxidative stress. *Environ. Sci. Technol.* **2007**, *41* (11), 4158−4163.
(154) Ramchandran, V.; Gernand, J. M. Examining the in vivo pulmonary toxicity of engineered metal oxide nanomaterials using a genetic algorithm-based dose-response-recovery clustering model. *Computational Toxicology* **2020**, *13*, 100113.
(155) Rallo, R.; France, B.; Liu, R.; Nair, S.; George, S.; Damoiseaux, R.; Giralt, F.; Nel, A.; Bradley, K.; Cohen, Y. Self-organizing map analysis of toxicity-related cell signaling pathways for metal and metal oxide nanoparticles. *Environ. Sci. Technol.* **2011**, *45* (4), 1695−1702.
(156) Ivask, A.; Titma, T.; Visnapuu, M.; Vija, H.; Kaekinen, A.; Sihtmae, M.; Pokhrel, S.; Maedler, L.; Heinlaan, M.; Kisand, V.; et al. Toxicity of 11 Metal Oxide Nanoparticles to Three Mammalian Cell Types In Vitro. *Current Topics in Medicinal Chemistry* **2015**, *15* (18), 1914−1929.
(157) George, S.; Xia, T.; Rallo, R.; Zhao, Y.; Ji, Z.; Lin, S.; Wang, X.; Zhang, H.; France, B.; Schoenfeld, D.; et al. Use of a High-Throughput Screening Approach Coupled with In Vivo Zebrafish Embryo Screening To Develop Hazard Ranking for Engineered Nanomaterials. *ACS Nano* **2011**, *5* (3), 1805−1817.
(158) Zhang, H.; Ji, Z.; Xia, T.; Meng, H.; Low-Kam, C.; Liu, R.; Pokhrel, S.; Lin, S.; Wang, X.; Liao, Y.-P.; et al. Use of Metal Oxide Nanoparticle Band Gap To Develop a Predictive Paradigm for Oxidative Stress and Acute Pulmonary Inflammation. *ACS Nano* **2012**, *6* (5), 4349−4368.
(159) Walczyk, D.; Bombelli, F. B.; Monopoli, M. P.; Lynch, I.; Dawson, K. A. What the Cell "Sees" in Bionanoscience. *J. Am. Chem. Soc.* **2010**, *132* (16), 5761−5768.
(160) Wheeler, K. E.; Chetwynd, A. J.; Fahy, K. M.; Hong, B. S.; Tochihuitl, J. A.; Foster, L. A.; Lynch, I. Environmental dimensions of the protein corona. *Nat. Nanotechnol.* **2021**, *16* (6), 617−629.
(161) Xin, X.; Huang, G.; Zhang, B. Review of aquatic toxicity of pharmaceuticals and personal care products to algae. *Journal of Hazardous Materials* **2021**, *410*, 124619.
(162) Bar-Ilan, O.; Louis, K. M.; Yang, S. P.; Pedersen, J. A.; Hamers, R. J.; Peterson, R. E.; Heideman, W. Titanium dioxide nanoparticles produce phototoxicity in the developing zebrafish. *Nanotoxicology* **2012**, *6* (6), 670−679.
(163) Mosquera, J.; García, I.; Liz-Marzán, L. M. Cellular Uptake of Nanoparticles versus Small Molecules: A Matter of Size. *Acc. Chem. Res.* **2018**, *51* (9), 2305−2313.
(164) Zhang, M.; Xu, L. H. Transport of micro- and nanoplastics in the environment: Trojan-Horse effect for organic contaminants. *Critical Reviews in Environmental Science and Technology* **2022**, *52* (5), 810−846.
(165) Hsiao, I. L.; Hsieh, Y.-K.; Wang, C.-F.; Chen, I. C.; Huang, Y.-J. Trojan-Horse Mechanism in the Cellular Uptake of Silver Nanoparticles Verified by Direct Intra- and Extracellular Silver Speciation Analysis. *Environ. Sci. Technol.* **2015**, *49* (6), 3813−3821.
(166) Wang, Y.; Gao, X.; Cheng, Y. H.; Peijnenburg, W.; Dong, Z. M.; Fan, W. H. Nano-$TiO_2$ modifies heavy metal bioaccumulation in Daphnia magna: A model study. *Chemosphere* **2023**, *312*, 137263.
(167) Vo, A. H.; Van Vleet, T. R.; Gupta, R. R.; Liguori, M. J.; Rao, M. S. An Overview of Machine Learning and Big Data for Drug Toxicity Evaluation. *Chem. Res. Toxicol.* **2020**, *33* (1), 20−37.
(168) Wu, X.; Zhou, Q.; Mu, L.; Hu, X. Machine learning in the identification, prediction and exploration of environmental toxicology: Challenges and perspectives. *Journal of Hazardous Materials* **2022**, *438*, 129487.
(169) Kim, M. T.; Wang, W.; Sedykh, A.; Zhu, H. Curating and Preparing High-Throughput Screening Data for Quantitative Structure-Activity Relationship Modeling. *Methods Mol. Biol.* **2016**, *1473*, 161−172.
(170) Yan, X. L.; Sedykh, A.; Wang, W. Y.; Zhao, X. L.; Yan, B.; Zhu, H. In silico profiling nanoparticles: predictive nanomodeling using universal nanodescriptors and various machine learning approaches. *Nanoscale* **2019**, *11* (17), 8352−8362.

(171) Helma, C.; Rautenberg, M.; Gebele, D. Nano-Lazar: Read across Predictions for Nanoparticle Toxicities with Calculated and Measured Properties. *Frontiers in Pharmacology* **2017**, *8*. DOI: 10.3389/fphar.2017.00377.
(172) Furxhi, I.; Murphy, F.; Mullins, M.; Poland, C. A. Machine learning prediction of nanoparticle in vitro toxicity: A comparative study of classifiers and ensemble-classifiers using the Copeland Index. *Toxicol. Lett.* **2019**, *312*, 157−166.
(173) Azodi, C. B.; Tang, J.; Shiu, S. H. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet* **2020**, *36* (6), 442−455.
(174) Sheridan, R. P. Interpretation of QSAR Models by Coloring Atoms According to Changes in Predicted Activity: How Robust Is It? *J. Chem. Inf. Model.* **2019**, *59* (4), 1324−1337.
(175) Du, M.; Liu, N.; Hu, X. Techniques for interpretable machine learning. *Commun. ACM* **2019**, *63* (1), 68−77.
(176) Murdoch, W. J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (44), 22071−22080.
(177) Scott-Fordsmand, J. J.; Amorim, M. J. B. Using Machine Learning to make nanomaterials sustainable. *Sci. Total Environ.* **2023**, *859*, 160303.
(178) Greden, J. F.; Parikh, S. V.; Rothschild, A. J.; Thase, M. E.; Dunlop, B. W.; DeBattista, C.; Conway, C. R.; Forester, B. P.; Mondimore, F. M.; Shelton, R. C.; et al. Impact of pharmacogenomics on clinical outcomes in major depressive disorder in the GUIDED trial: A large, patient- and rater-blinded, randomized, controlled study. *J. Psychiatr Res.* **2019**, *111*, 59−67.
(179) Rahman, S.; Khan, S.; Porikli, F. A Unified Approach for Conventional Zero-Shot, Generalized Zero-Shot, and Few-Shot Learning. *Ieee Transactions on Image Processing* **2018**, *27* (11), 5652−5667.
(180) Lynch, I. A. Antreas. WorldFAIR (D4.2) FAIRification of nanoinformatics tools and models recommendations. In *Zenodo*, 2024, DOI: 10.5281/zenodo.10629631.
(181) Olker, J. H.; Elonen, C. M.; Pilli, A.; Anderson, A.; Kinziger, B.; Erickson, S.; Skopinski, M.; Pomplun, A.; LaLone, C. A.; Russom, C. L.; et al. The ECOTOXicology Knowledgebase: A Curated Database of Ecologically Relevant Toxicity Tests to Support Environmental Research and Risk Assessment. *Environ. Toxicol. Chem.* **2022**, *41* (6), 1520−1539.
(182) Maier, D.; Exner, T. E.; Papadiamantis, A. G.; Ammar, A.; Tsoumanis, A.; Doganis, P.; Rouse, I.; Slater, L. T.; Gkoutos, G. V.; Jeliazkova, N. et al.et al. Harmonising knowledge for safer materials via the "NanoCommons" Knowledge Base. *Frontiers in Physics* **2023**, *11*, Original Research. DOI: 10.3389/fphy.2023.1271842.


https://doi.org/10.1021/acs.est.4c03328
*Environ. Sci. Technol.* 2024, 58, 14973−14993