Using machine learning to predict adverse effects of metallic nanomaterials to various aquatic organisms

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

Environmental Science & Technology

pubs.acs.org/est

Article

Using Machine Learning to Predict Adverse Effects of Metallic Nanomaterials to Various Aquatic Organisms

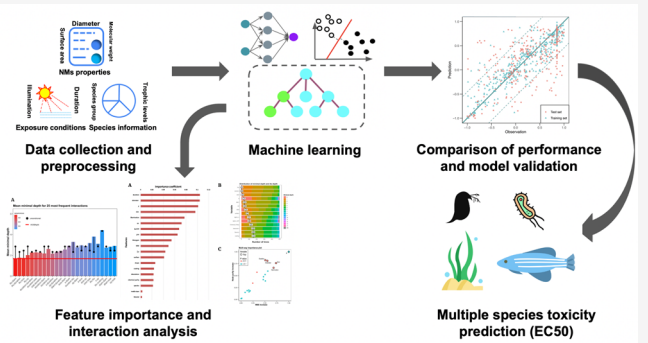
Yunchi Zhou, Ying Wang,* Willie Peijnenburg, Martina G. Vijver, Surendra Balraadjsing, and Wenhong Fan*

Cite This: *Environ. Sci. Technol.* 2023, 57, 17786−17795

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The wide production and use of metallic nanomaterials (MNMs) leads to increased emissions into the aquatic environments and induces high potential risks. Experimentally evaluating the (eco)toxicity of MNMs is time-consuming and expensive due to the multiple environmental factors, the complexity of material properties, and the species diversity. Machine learning (ML) models provide an option to deal with heterogeneous data sets and complex relationships. The present study established an *in silico* model based on a machine learning properties-environmental conditions-multi species-toxicity prediction model (ML-PEMST) that can be applied to predict the toxicity of different MNMs toward multiple aquatic species. Feature importance and interaction analysis based on the random forest method indicated that exposure duration, illumination, primary size, and hydrodynamic diameter were the main factors affecting the ecotoxicity of MNMs to a variety of aquatic organisms. Illumination was demonstrated to have the most interaction with the other features. Moreover, incorporating additional detailed information on the ecological traits of the test species will allow us to further optimize and improve the predictive performance of the model. This study provides a new approach for ecotoxicity predictions for organisms in the aquatic environment and will help us to further explore exposure pathways and the risk assessment of MNMs.



KEYWORDS: *nanomaterials, toxicity, machine learning, prediction, exposure conditions, aquatic organisms*

INTRODUCTION

Metallic nanomaterials (MNMs) are widely used in many fields because of their outstanding physical and chemical properties.[1] The increasing applications of MNMs in industries and society have caused a rapid growth of the production volume of NMs,[2] of which 80% are metals or metal oxides with an annual production of approximately 242,000 tons.[2,3] MNMs released into the aquatic environment will pose risks to aquatic organisms like fish, crustacean, algae, bacteria.[4,5] The exposure to MNMs has been proved to cause the growth inhibition of algae,[6] the immobilization and genotoxic effect of *Daphnia magna* (*D. magna*),[7] and the cytotoxicity of fish cells.[8] Numerous studies proved that the toxicity varies with properties such as size,[9] shape,[10] surface area, and surface composition.[11] Moreover, due to the complex properties of different MNMs, a slight change in shape and size may significantly affect their toxic effect in aquatic environment.[12,13] Given the complexity of the aquatic environment and the sophisticated properties of MNMs, laboratory experiments take a long time and each tested condition can give different response results,[14,15] making process-based predictions challenging to the plethora of organisms present.[16−18] It is essential to develop *in silico* models to prioritize experimental testing efforts to understand exposure and process-based toxicity as well as to identify the main descriptors for toxicity which allow for scenario modeling to predict MNMs toxicity for a variety of organisms and environmental settings.

As a computational *in silico* modeling method, nano-quantitative structure−activity relationships (nano-QSARs) provide us with an insight of the relationships between physicochemical properties of MNMs and their toxicity profile. In the study of Mu et al.,[3] relationships between 26 physicochemical properties of 51 metal oxide nanoparticles (MeONPs) and their cytotoxic effects in *Escherichia coli* (*E. coli*) were tested. The study showed that enthalpy of formation of a gaseous cation (ΔH_{me^+}) and polarization force (Z/r) were

Special Issue: Data Science for Advancing Environmental Science, Engineering, and Technology

Received: September 25, 2022
Revised: January 20, 2023
Accepted: January 20, 2023
Published: February 2, 2023

important modulators to the toxicity of MeONPs. In another study, nano-QSARs models were used to predict the cytotoxicity of 17 MeONPs to *E. coli*. Linear regression models were created and the R^2-value reached 0.86 for the test set.[9] However, nano-QSARs models focused only on the physicochemical properties of materials, ignoring other significant factors that contribute to the toxicity mechanism.[19,20]

Most importantly, the models built in previous studies mainly focused on a single organism. Large ecological differences between species result in significantly different toxic effects following exposure to different substances. Ye et al.[21] reported that zinc oxide nanoparticles (ZnO NPs) were less toxic to algae than to daphnids, and the EC_{50} values of *Danio rerio* were more than 2 orders of magnitude higher than the EC_{50} values of daphnids. This finding showed species at the highest trophic level are among the least sensitive species.[21] However, in another study, the acute toxicity of ZnO NPs to various invertebrates and fish has been reported from 40 μg/L to 58 mg/L. Therefore, it is necessary to reveal the sensitivity of different species to MNMs in order to obtain a broader view on species sensitivity to metallic nanomaterials. The main challenge of developing a multispecies model lies in the difficulty to find features which can describe the differences between organisms, while at the same time being easy to be encoded.

The coupled influence of multiple factors on nanotoxicity calls for new *in silico* methods. As a newly developing robust nonparametric approach, machine learning (ML) methods provide an alternative to experimental approaches, especially as ML methods are efficient in dealing with heterogeneous data and in finding relationships between different complex factors.[22] ML methods can also be applied to follow the 3R principles (Replacement, Reduction, Refinement).[23,24] Moreover, with the help of analytical tools, important features and their interaction with each other can be identified, thus giving solid support to the risk assessment of MNMs.

The objective of this study was to develop a regression model for aquatic toxicity prediction that takes into account physicochemical properties of MNMs, environmental factors, and different organisms with their own traits and exposure conditions covering different trophic levels. To achieve this, we (1) built a model of 14 different MNMs against 51 species based on published data sets,[25−27] (2) collected recently published literature for the external validation, and (3) analyzed the importance and interaction between physicochemical properties, environmental factors and species. The model developed has the acronym ML-PEMST for a properties−environmental conditions−multispecies-toxicity prediction model. Based on the results of this study, we outline the role of environmental factors and physicochemical properties of MNMs on the toxicity. This *in silico* method is proved to be effective for multi species toxicity prediction in the aquatic environment and to provide guidance for risk assessment of MNMs.

MATERIALS AND METHODS

Data Sets. Toxicity data of various species were collected from three available databases, including Nano E-Tox[25] (https://cfpub.epa.gov/ecotox/), the database reported by Chen et al.,[26] and the database reported by Bunmahotama et al.[27] Quantitative information on the toxicity of MNMs (these include MeONPs) is collected in these databases from published articles up to 16 September 2020. Given that the descriptors are not identical for three databases, a selection of the descriptors was performed to merge the data from different sources (Text S1, Figure S1). The effective concentration affecting 50% of individuals (EC_{50}, in mg/L) was selected as the acute toxicity endpoint, which was chosen as the label of the model.

Four kinds of independent variables were collected to derive the ML models: (1) seven nanospecific physicochemical properties, i.e., primary size, hydrodynamic diameter, shape, surface area, purity, dissolution (yes/no) and presence of coating (yes/no); (2) the information on the species tested, i.e., group of species and trophic level; (3) four characteristics of the exposure conditions used during toxicity testing, i.e., zeta potential, exposure duration, illumination and media; and (4) seven molecular descriptors of MNMs, as proposed by Kar et al.,[28] i.e., molecular weight (MW), the number of metal atoms of each molecular (N_{metal}), the number of oxygen atoms of each molecular (M_{oxygen}) (for example, the N_{metal} of Al_2O_3 is two, and the M_{oxygen} of Al_2O_3 is three), metal electronegativity (χ), sum of metal electronegativity for an individual metal oxide ($\sum\chi$), the charge of the metal cation (χ_{ox}), and the sum of metal electronegativity for an individual metal oxide divided by the number of oxygen atoms present in a particular metal oxide ($\sum\chi$/nO). These properties are simple to calculate but critical for characterization of MNMs (Table S1).

Preprocessing Data. A series of data cleaning and preprocessing tasks were performed to make our data set suitable for modeling.

First, data with more than 33% missing values (that means the data space is blank, or recorded as "#N/A") were eliminated because of their poor information content. Second, for the numerical descriptors (diameter, size, surface area, purity, zeta potential), values noted by range and possibilities were replaced by mean values, and missing values of each feature were filled by the corresponding mean value. Third, for the characteristic values (i.e., dissolution, coating, shape, media, species, illumination), digitization was performed to make the data recognizable for the ML software. One-hot encoding, a general encoding method that converts categorical variables into binary vectors, was applied in order to convert these features into numerical values. The data for illumination were converted into the percentage of light during 24 h (Table 1). Finally, for the species, one-hot encoding is not suitable as there are 51 different species, which together with the diversity of NMs types will cause an unnecessary increase of data dimension. Therefore, a biological grade index based on trophic level was created as the indicator to classify different species, as the trophic level often represents the relationship of different species in the food chain and the accumulations of NMs in highly trophic-grade organisms often increases by feeding. Besides, unlike for fish, whose trophic levels are measured precisely by experiments and available in "Fishbase",[29] species like bacteria are grouped as primary producers or consumers because they are at the bottom of food chain. Therefore, the classification of bacteria is based on whether they are autotrophic or heterotrophic.

Furthermore, in order to prevent the overcontribution and biases, the normalization method was applied to certain features before modeling (i.e., diameter, size, surface area, purity, zeta potential) (eq 1)

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

Table 1. Data Pretreatment Method

Feature	Treatment method	Detailed information	Missing values treatment
Diameter	Normalization	Z-score	Mean value filling
Size	Normalization	Z-score	Mean value filling
Surface area	Normalization	Z-score	Mean value filling
Purity	Normalization	Z-score	Mean value filling
Zeta potential	Normalization	Z-score	Mean value filling
Dissolution	One-hot coding	Soluble (1), not soluble (0)	Filled by 2
Coating	One-hot coding	With coating (1), without coating (0)	Filled by 2
Shape	One-hot coding	Spherical shape (1), other shape (0)	Filled by 2
Media	One-hot coding	Seawater (1), freshwater (0)	Filled by 2
Illumination	Numericalization	Proportion of light during 24 h	
Species (fish)	Classified by trophic levels	Data according to "Fishbase"	
Species (plants, crustacean, bacteria and others)	Classified by feeding types	Autotrophic (1), heterotrophic (2)	

$$x_i' = \frac{(x_i - \mu_i)}{\sigma_i} \quad (1)$$

where x_i is the ith feature value, μ_i is the mean value of the ith feature, and σ_i is the standard deviation (SD) of the ith feature, x_i' is the ith feature value after normalization.

Development of the Multispecies-Toxicity Prediction Model Based on Machine Learning (ML-PEMST). Three different supervised machine learning algorithms were applied in the present study: random forest (RF), support vector machine (SVM), and artificial neural network (ANN). Each of these three algorithms has been shown to possess strong abilities in classification and regression, and researchers often use them for *in silico* pollutant toxicity prediction research.[16,30]

The RF algorithm shows good performance in classification and regression tasks by constructing a multitude of decision trees. In fact, it is widely used *in silico* modeling due to its outstanding ability to deal with heterogeneous data and resistance toward overfitting. The RF algorithm was applied to study the cytotoxicity of MeONPs and showed good specificity (75% for the training set) and accuracy.[31] In the present study, the RF model used 500 random decision trees, and at each node five random features were selected.

SVM is good at handling high-dimensional problems; it projects data as points into a high-dimensional space and then searches for the hyperplane that can best separate the data according to our needs.[32] SVM has been used for the building of a regression model for prediction of cytotoxicity of ZnO and TiO_2 nanoparticles.[33] By choosing the appropriate kernel function, SVM shows a high accuracy and also a resistance to overfitting.[34] In our research, the radial basis function (RBF) was chosen as the kernel because it has a localized and finite response along the entire x-axis.

ANN is a group of nodes which are interconnected and divided by multiple layers (e.g., input layers, hidden layers), the node of each layer works by receiving and processing signals and sends the signals to the next layer. ANN is known for its ability to process large and complex data; it has been used for building QSAR models to predict the cytotoxicity of metal oxide nanoparticles to *Escherichia coli* (*E. coli*).[35]

However, it is also worth noting that ANN is memory intensive and easy to get overfitted, whereas the interpretability of the results is also a shortcoming of ANN. In our model, 25% of the data in the training set were split randomly as validation data. "Adam" was used as the optimizer, the learning rate was set to 0.1, 38 units in the hidden layers were adopted, and the decay of learning rate was set as 0 to prevent overfitting during the training.

The R^2 value and root-mean-square error (RMSE) were collected as the criteria for the goodness-of-fit of the model, as a high R^2 value and a small RMSE value indicate a credible result. RF, SVM, and ANN modeling used in the present study were all performed by Python 4.0, mainly using scikit-learn and Keras packages.

Internal Validation. As a common tool of model evaluation, the k-fold cross validation was used in order to avoid the overfitting and to evaluate the performance of model.[36] In the present study, 10-fold cross validation was used for internal validation of models, which means that 90% of the samples in each subset were chosen as the training set and 10% as the test set.[11] 20, 40, 60, 80, and 100% of the label values in the training set were replaced by random values within the original label range during each process of 10-fold cross-validation. The corresponding cross-validation coefficients (Q^2) were used as indexes to see whether the model is overfitted (eq 2)

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \quad (2)$$

where y_i is the observed label value, $\hat{y}$ is the predicted label value, and $\overline{y}$ is the average of the observed label values.

The permutation test, a method that uses the random arrangement of sample data to make statistical inference, was performed 10 times for each ratio, with 5 ratios in 10-fold giving 500 Q^2 values. Linear regressions were then performed on the Q^2 values and the correlation coefficients between the original labels and the permutation labels. We can infer from the intercept of the regression result on the y axis, i.e., if the intercept is less than 0.05, that proves the model is not overfitting.[37]

External Validation. A robust and reliable prediction ability is based on good results of internal and external validation.[38] The validation set did not participate in the construction of the model in order to ensure that the model did not learn from these data. In order to further test the performance of the model, an additional data set was collected from recent published articles from 1 January 2020 to the present using Web of Science, Pubmed, and Scopus's advanced search, with the search keywords: **Nano** AND **Toxicity** AND **Metal**. The acquired articles were filtered by the following rules: (I) the topic should relate to toxic effects of NMs on aquatic species; (II) the materials should be metal or metal oxide, and the basic material information needed to be available; (III) the endpoint of toxicity should be EC_{50}; (IV) no duplicate data were allowed to be present in the test data set and in the external validation data set. The feature selection and data processing remain the same with the training and test

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

data set. The external validation data set is visualized and arranged in descending order of size in Figure S2.

The externally validated determination coefficient (Q^2_{ext}) and the root mean square error of prediction ($RMSE_{ext}$) were used to evaluate the model performance (eq 3), and a small value of $RMSE_{ext}$ indicates a good performance of model in external validation[39]

$$RMSE_{ext} = \sqrt{\frac{\sum_{i=1}^{m} (y_i - \hat{y})^2}{m}} \quad (3)$$

where y_i is the observed label value, $\hat{y}$ is the predicted label value, and m is the amount of data in the validation set.

Feature Selection and Importance Analysis. To fully understand the model and improve its interpretability, feature analysis is an indispensable step to find which features have a bigger influence on the toxicity of MNMs. The importance of features is usually evaluated by importance ranking in the RF model.[40] In the present study, a combination of four indicators, including mean minimal depth, P value, node purity increase, and MSE increase, were adopted to avoid the bias caused by a single indicator. The mean minimal depth focuses on the structure of the random forest. During training, at each node the feature which contributed the most to the overall split among the five was retained at the node. This means that a smaller depth represents more importance.[41] The P value is based on the one-sided binomial test to evaluate the significance of feature importance, and a small value of P indicates that it is less likely to observe an extreme value under the null hypothesis.[42] Node purity increase looks at the changes in node purity after splits on the variable,[43] while the MSE increase is based on the decrease in predictive accuracy of the forest after perturbation of the variable.[44] If all the four indicators show that a feature is important, then we recognize the importance of that feature. The feature importance analysis was conducted by R 4.1.2, with the help of the "randomForest" and "randomForestExplainer" packages. The model and the analysis of the result are based on the software proposed by Yu et al.[11]

RESULTS AND DISCUSSION

Description and Pretreatment of Data. After selection and screening, we obtained a total of 684 samples for 14 different MNMs including nine metal oxide nanoparticles and five metallic nanoparticles. The set contained 19 features for four classes; i.e., nanospecific physicochemical properties, information on the species tested, exposure conditions, molecular descriptors, and corresponding toxicity data were obtained (Table 2). Overall, this comprehensively covers the main information in the toxicity analysis.

Table 2. Selected Features for Model Building in Order of Importance

Categories	Features
Nanospecific physicochemical properties	Size, surface area, shape, coating, purity
Species information	Biological index based on trophic level
Exposure condition	Duration, hydrodynamic diameter, illumination, zeta potential, dissolution, media
Molecular descriptor	χ, $\sum\chi$/nO, χox, MW, Moxygen, $\sum\chi$, Nmetal

The distribution of 13 numeric variables was described by mathematical statistics (mean, SD, median, and distribution range), while the six characteristic variables (type, coating, media, species, shape, and dissolution) were described by the reported frequency, which was visualized and arranged in descending order of size in Figure 1A. As we can infer from Figure 1A, the heterogeneity of the toxicity analysis data set was mainly caused by MNMs diversity and species tested. The primary size of the particles ranges from 1 to 300 nm, and the surface area ranges from 4 to 325 m^2/g. Among the characteristic variables, "coating", "shape", and "dissolution" have the most missing values, which may allow the model to misestimate their importance. For example, for the coating, 61% of the MNMs were noted as value missed. Therefore, one-hot encoding of these discrete NMs types was performed to reduce biases caused by their imbalance.

Fifty-one species were included in our data set. Among them, bacteria account for 19%, algae and plant accounted for 15%, fish accounted for 6%, and other species accounted for 7%. The database included 53% crustacean data, of which *D. magna* accounted for 43% of the full data set. Their results in toxicity varied from 0.2 μg/L to (note of authors almost unrealistic high) 5000 mg/L in our data set (Figure 1B). From the perspective of MNMs, it is noticed that titanium dioxide nanoparticles (TiO_2NPs) and silver nanoparticles (AgNPs) accounted for 24.7% and 29.6% of the data in the toxicity data set. The range of EC_{50} values for TiO_2NPs and AgNPs with different properties to different species varies from 0.01 mg/L to 5000 mg/L and 0.02 μg/L to 571 mg/L, respectively.

The toxicity distribution of species for different MNMs is visualized in Figure 1C, and the AgNPs showed in general the lowest EC_{50} values for species marked by 2 and ≥3, which indicates that AgNPs are more toxic for aquatic organisms than other MNMs. The reason for the high toxicity of AgNPs is closely related to its surface oxidation, release of silver ions, and interaction with biological macromolecules.[45] The Fe shows low toxicity for species marked by 2; it is unusual because Fe is very reactive in the water. The reason could be attributed to the change of morphology of Fe in the water. Generally, the toxicity thresholds for species marked by 1 (algae) and 2 (mainly crustacean and bacteria) are less than for other species, which means that these species are more sensitive to MNMs in the aquatic environment. Various reasons can be found for this; e.g., it is likely that these organisms adsorb relatively a lot because of their large membrane surface area compared to their body volume and/or the fact that bacteria and algae have a worst case exposure because they are mixed with the NMs in a shaking exposure setting in the lab,[46] and thus stochastic exposure will give them higher uptake for a daphnia that is swimming but with a static exposure in the lab.[47] In addition, the AgNPs, TiO_2NPs, and ZnONPs showed the largest variation in EC_{50} values; the reason is not only their large proportion in the data set but also the impact of environmental factors, different species, and materials.

Development of ML-PEMST Models. Three ML models were used all with assumptions that they take all features into consideration and search for several important features for various MNMs at the same time. The (PEMST) prediction model of MNMs was developed using RF, SVM, and ANN methods, respectively. The results of all models show a good fit to the data (all values of $R^2 > 0.7$, Figure 2A). RF has a better performance ($R^2 = 0.82 \pm 0.04$) than SVM ($R^2 = 0.75 \pm 0.05$)

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

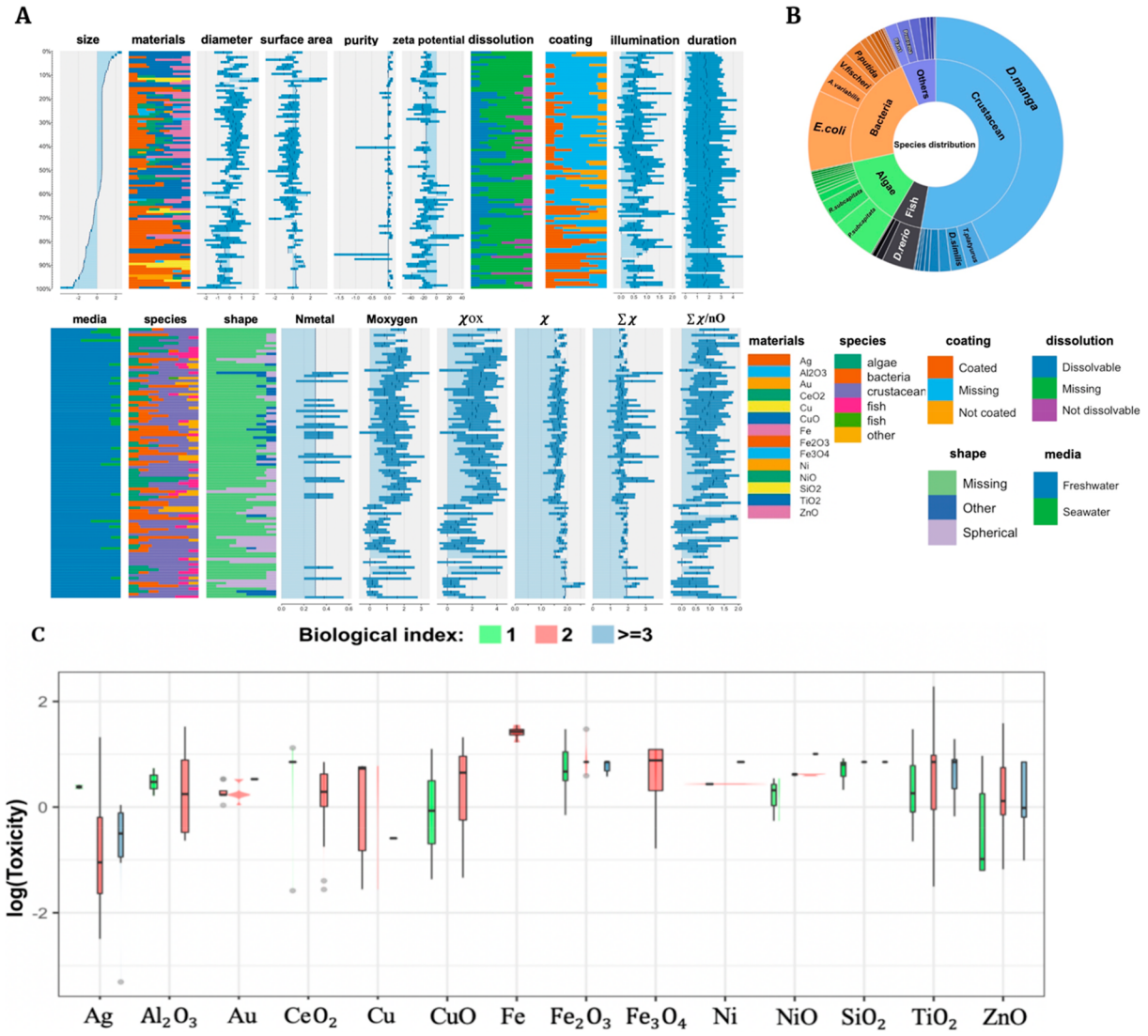


Figure 1. Descriptive statistics of the modeling data set. (A) Visualization of the distribution of the raw data set using the "tabplot" package in R software. (B) Distribution of species in the data set using the "highcharts" platform (https://www.hcharts.cn). (C) Toxicity distribution of biological indexes for different MNMs, in which the biological indexes (1, 2, and 3) mean the trophic level of test organisms in the data set.

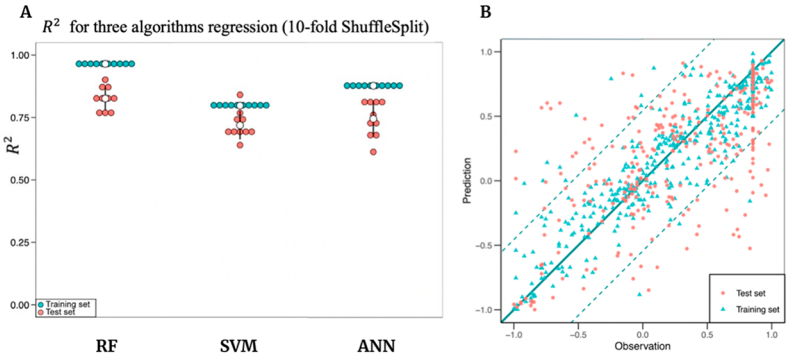


Figure 2. ML-PEMST modeling. (A) R^2 distribution of the RF, SVM and ANN regression (10-fold ShuffleSplit cross-validation). (B) Prediction performance of toxicity concentration using RF models. The slope of the solid lines is 1, and the dotted lines represent the intercepts ± RMSE.

and ANN (R^2 = 0.74 ± 0.06). In addition, the predictive performance of the RF model indicates that the difference between the observed value and the predicted value is small, as most of the data lie in the range of ± RMSE (Figure 2B).

The good performance of RF in nanotoxicity prediction is not an isolated phenomenon. The RF model also achieved better performance than CORAL models in the prediction of *in vitro* cytotoxicity of silica nanoparticles.[48] Moreover, Mirzaei

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

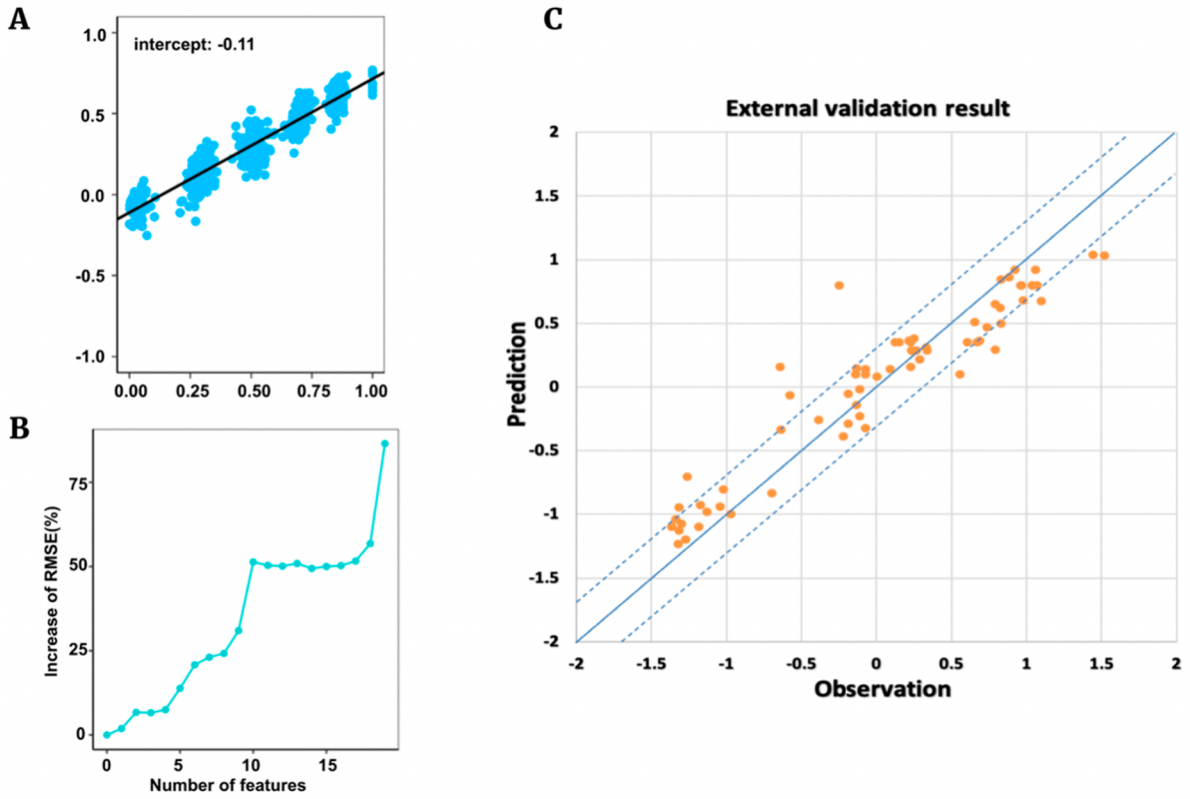


Figure 3. Validation of the RF model. (A) Permutation test for overfitting. The *X*-axis represents the correlation coefficients of the raw label values and the permutated label values, and the *Y*-axis represents the Q^2 values. The permutation test results showed that the model did not overfitting (intercept less than 0.05). (B) Feature value shuffling. The model loses its ability of prediction after feature value shuffling. (C) Prediction performance of the model on external validation set. The slope of the solid line is 1, and the dotted lines represent the intercepts ± RMSE.

et al.[49] also reported in their study that an RF model exhibited the lowest error and the highest R^2 score compared to other methods (i.e., SVM, Elastic Net Regression, Ridge Regression). The reason might be that the RF method is better at determining nonlinear relationships and at handling numeric and categorical data simultaneously. Therefore, RF could be considered as a better approach when dealing with similar problems in the future.

A reliable prediction can be generated only from models that have been validated internally and externally.[38] Cross-validation and value shuffling were performed as internal validation to clarify that the model did not overfit and that the features provided useful information. The result of permutation test showed that the intercept of the regression result on the *y* axis was less than 0.05 (Figure 3A), meaning that the model did not overfit.[50] Furthermore, the performance after feature values shuffling indicated that the prediction was totally disturbed, which proved the credibility of the model (Figure 3B). Aside from internal validation, a new external validation data set was built to evaluate the accuracy and robustness of the model. The results indicated that the RF model achieved a great performance on the external validation set (Q^2_{ext} = 0.87, $RMSE_{ext}$ = 0.3). 95% of the difference between observed and measured values was within 1 order of magnitude (Figure 3C). The results of internal and external validation illustrated that the RF-based PEMST model is robust and sufficiently reliable to be used for further toxicity data prediction studies of aquatic species.

Features Importance Analysis. Based on the results of the RF model, multiple indicators, including MSE increase, node purity increase, *P* value and mean minimal depth, were applied together to evaluate the importance of features in order to avoid the biases (Figure 4). The result given by the MSE increase in Figure 4A shows that test duration, particle diameter, electronegativity, and size make up the first echelon.

The physicochemical properties of MNMs have been proven to be crucial to the toxicity mechanism, which is in line with previous research findings.[49,51,] Notably, in our study the hydrodynamic diameter and primary size of MNMs show similar higher importance for different indicators.[52] The hydrodynamic diameter was found to be more important than primary size, as it reflects the actual size of MNMs in the media when they interact with a test organism.[53]

Figure 4B suggests that environmental factors may have a significant impact in MNM toxicity prediction. The primary role of duration seems to be reasonable, as a longer exposure time will definitely cause more damage to a test organism at the same chemical concentration. Moreover, the importance of exposure duration was already confirmed by Choi et al.[16] In fact, exposure duration has been demonstrated to correlate directly with toxicity, such as Connell et al.[54] have proposed by means of the Reduced Life Expectancy (RLE) model, which describes a negative correlation between exposure duration and LC_{50}. With regard to the environmental factors, it is worth noting that the illumination appeared most frequently near the roots of the model than any of the other features. There are several research projects that have investigated the relation between illumination and toxicity of MNMs. For example, Yang et al.[17] have reported that the LC_{50} of TiO_2 NPs to zebrafish was lower under simulated sunlight illumination than

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

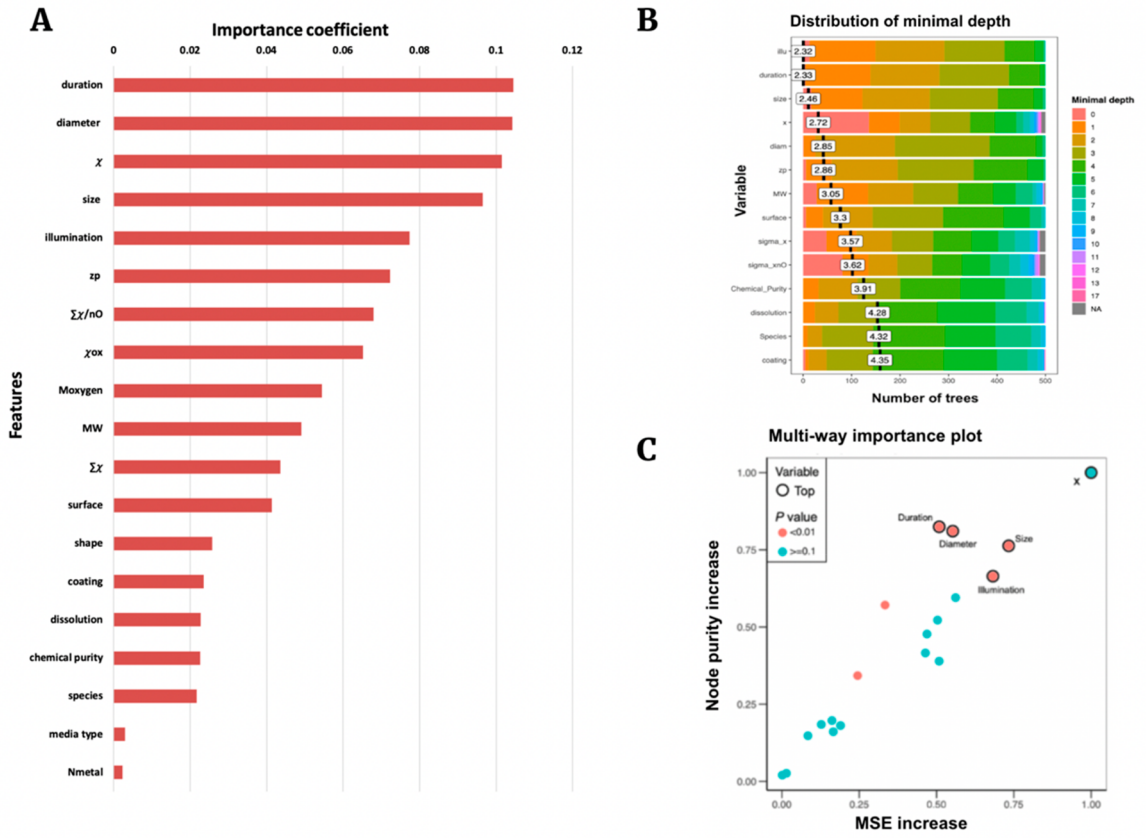


Figure 4. Feature importance analysis of model. (A) Importance rank of different descriptors. The *X*-axis represents the importance coefficients measured by MSE increase, and the *Y*-axis displays the model features. (B) Distribution of the features' mean minimal depth, in which illumination, exposure duration, and primary size are closer to the rood of the trees than other features. (C) Multiway feature importance analysis of model, combining the MSE increase, node purity increase, and *P* values of the features.

in the dark. Mittelman et al.[18] have proven that the exposure to ultraviolet (UV) light will significantly enhance AgNPs retention and dissolution, which is crucial to the toxicity activities of AgNPs. Souza et al.[55] have also reported that visible light will increase the aggregation of uncoated AgNPs and polyvinylpyrrolidone (PVP) coated AgNPs, thus reducing their toxicity. The reason could be the UV irradiation disturbed electrons at the surface of AgNPs, in turn driving dipole−dipole interactions and the formation of aggregates, these same surface energy phenomena also disturb the equilibrium of silver in the oxide layer, leading to enhanced silver dissolution.[18]

The electronegativity was considered to be important according to the MSE increase and the node purity increase. However, electronegativity was not found to be the main factor in the toxicity models since its *P* value exceeds 0.1, meaning that its importance was not statistically significant (Figure 4C), which is consistent with our previous study.[56] Moreover, combining different indicators together overcomes the biases caused by a single indicator and identifies that the physicochemical properties of MNMs and the exposure conditions are both important for the toxicity assessment. These features are thus of crucial importance for experimentalists when they report their findings.

It is noticeable that dissolution is not considered as very important in the results, given that we know the most toxic MNMs are those that release ions. The reason can be attributed to the quality of the original data set and the binary representation: the large amount of missing data makes the model underestimate the importance of dissolution. Also, the method for noting dissolution data is mixed by percentage and mass/volume, making it difficult to unify them. Therefore, binary was applied to convert dissolution, which may cause a loss of information. For the coating, the problem is similar; the missing value and binary make the model unable to recognize the importance of coating correctly. A better approach for dealing these data, such as unifying the record of dissolution, and classifying the different coatings by their functions, is needed for data preprocessing in the future.

The aquatic test species in our model was found not to be crucial for prediction of MNM toxicity. This was surprising (and unexpected) to us, and we try to unravel here some reasoning for this finding. First of all, compared to the trophic levels, it is more difficult to know what the organisms eat and their feeding behavior, which does make the trophic levels not descriptive enough to explain the exposure pathway. Second, the single feature does not fully reveal the difference between species. It has been proved that the responses of microcrustaceans to MNMs also rely on their morphology (body size and shape), ecological traits (feeding mechanisms, life cycles), and intrinsic sentitivities.[57]

Although the method of coding species information can still be optimized here, but it does reveal the difference of sensitivity between different trophic levels. To confirm this conclusion, there are still work to do in further research. For example, more detailed trait information, such as surface area-

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

to-body length, body size and weight, life-span, cell membranes, mode of internalizing the MNMs, etc., which have been proven to be linked with species sensitivity,[58] need to be collected and considered. In addition, the development of new descriptors that can fully reveal the difference between species with regard to toxicity mechanism is urgently needed for multispecies modeling work.

Finally, the unbalanced structure of the data set may weaken its representativeness, as 77% of the species are marked by two, within which *D. magna* makes a large contribution (accounted for 43% of the whole data set). To optimize the predictive power of models and make them representative for a larger variety of organisms, we look for more experimental data that fits the trait category "1" and "3".

Feature Interaction analysis. Understanding how the features in our models interact with each other is critical for identifying the toxicity mechanisms of MNMs. The conditional minimal depth was applied to further explore the order of the interaction strength between features (Figure S3). The strongest four feature interaction relationships occur between those that are already important; i.e., illumination interacts with exposure duration, hydrodynamic diameter, primary size, and zeta potential. These four relationships are displayed by a double-variable partial dependence analysis (Figure S4). The results indicated that environmental factors change the morphology of MNMs present in water. Although the specific effects of illumination on MNMs toxicity still need a case-by-case investigation, Yang et al. have reported that the increase of illumination time will enhance the toxicity of TiO_2NPs.[17] The mechanism of this relationship can be attributed to the production of ROS by photoinduced electron hole pairs: the holes lead to abstraction of electrons from water and hydroxyl anions to form hydroxyl radicals ($^{\bullet}OH$) whereas the electrons react with molecular oxygen to produce superoxide anions ($O_2^{\bullet -}$).[59] Moreover, the interaction between zeta potential and illumination is also worth noting in future toxicity assessments, as Mittelman et al. reported a 10−15 mV increase in zeta potential and a 5-fold increase in diameter of Ag particles after a 3-day UV exposure.[18] The mechanism could be that the UV irradiation disturbed electrons at the particle surface, thus driving dipole−dipole interaction,[60] and the zeta potential's variation will further affect the toxicity. As reported by Nedyalkova et al. there is a negative correlation between zeta potential and its toxicity for AgNPs in an aquatic environment.[61]

Implications and Future Recommendations. In this study, three ML regression models, i.e., RF, SVM, and ANN models, were built to investigate the ecotoxicity of MNMs to various species in the aquatic environment. During the selection of features, we tried to strike a balance between the selection of representative features and the feasibility of modeling. The RF model gives results which are consistent with the experiments, further improves the interpretability of ML, and discovers hidden feature interactions that are difficult to discover by means of other ML methods (e.g., SVM and ANN). However, when we apply multiple ML approaches, the consensus prediction is worth trying because it overcomes the limit caused by one single algorithm and provides more accurate prediction results. The exposure duration and the hydrodynamic diameter of the particles were considered to play a critical role, and the specific role of illumination was also discussed. It was concluded that parameters related to exposure pathway are important to consider; more detailed traits of the parameters on how organisms deal with internalized MNMs were not accounted for because of a lack of data and scientific knowledge on that. It is, however, most likely that behavior of MNMs within organism's bodies, biodistribution dynamics, and effect targets will be information that allows optimization on predictive power for those models.

It is important to acknowledge that the different experiments reported in the database follow different guidelines, resulting in inconsistent and incomplete data, as reflected in the model by a large amount of missing data. Therefore, better approaches should be developed to deal with missing data, such as the k-nearest neighbors (KNN) approach. Also, the unbalanced structure of the data set affects the accuracy of prediction. For example, TiO_2NPs and AgNPs accounted for more than half of the data set in our data set, and all these particles are sensitive to the light conditions. The importance of illumination could thus be overestimated, and a more balanced data set is needed to make the model more representative.

The last consideration is the applicability domain (AD), which defines the ranges within which the model can make reliable predictions.[62,63] In our model, 14 different metal/metal oxide were studied against 51 different species. According to the results of external validation, when the model meets new material and/or new species, it cannot make reliable predictions. Therefore, a well-defined AD is needed to make the model predictions more reliable.

Nonetheless, in general, the ML-PEMST model provides a new approach to handle the problem of multiple species and take environmental factors into account. This will help us to further explore the mechanisms of ecotoxicity of MNMs in the aquatic environment and to provide guidance for their risk assessment.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c07039.

Material and methods for original data set, descriptive statistics of the external validation data set, detailed information about descriptors of modeling, mean minimal depth for 25 most frequent interactions, double-variable partial dependence on the four strongest feature interactions (PDF)

Original data set for model development (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Ying Wang − *School of Space and Environment, Beihang University, Beijing 100191, China;* orcid.org/0000-0002-1424-0422; Email: yingw@buaa.edu.cn

Wenhong Fan − *School of Space and Environment and 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 Space and Environment, Beihang University, Beijing 100191, China;* orcid.org/0000-0001-6603-2177

Willie Peijnenburg − *Institute of Environmental Science (CML), Leiden University, Leiden 2300 RA, The*

https://doi.org/10.1021/acs.est.2c07039
Environ. Sci. Technol. 2023, 57, 17786−17795

Netherlands; Center for Safety of Substances and Products, National Institute of Public Health and the Environment (RIVM), Bilthoven 3720 BA, The Netherlands

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

Surendra Balraadjsing − *Institute of Environmental Science (CML), Leiden University, Leiden 2300 RA, The Netherlands*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.2c07039

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 42177240 and 41977352), the National Key R&D Program of China (No. 2022YFC3204800) and the Fundamental Research Funds for the Central Universities. The research described in this work was also supported by the European Union's Horizon 2020 research and innovation program via the project "NanoinformaTIX" (grant number 814426). MGV is grateful to the ERC-C EcoWizard support (grant number 101002123).

■ 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) Giese, B.; Klaessig, F.; Park, B.; Kaegi, R.; Steinfeldt, M.; Wigger, H.; von Gleich, A.; Gottschalk, F. Risks, Release and Concentrations of Engineered Nanomaterial in the Environment. *Sci. Rep.* **2018**, *8* (1), 1565.

(3) Mu, Y.; Wu, F.; Zhao, Q.; Ji, R.; Qie, Y.; Zhou, Y.; Hu, Y.; Pang, C.; Hristozov, D.; Giesy, J. P.; Xing, B. Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. *Nanotoxicology* **2016**, *10* (9), 1207−1214.

(4) 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.

(5) 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.

(6) 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.

(7) 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.

(8) 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.

(9) Pan, Y.; Li, T.; Cheng, J.; Telesca, D.; Zink, J. I.; Jiang, J. C. Nano-QSAR modeling for predicting the cytotoxicity of metal oxide nanoparticles using novel descriptors. *RSC Adv.* **2016**, *6* (31), 25766−25775.

(10) 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.

(11) 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). DOI: DOI: 10.1126/sciadv.abf4130.

(12) 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.

(13) Pikula, K.; Zakharenko, A.; Chaika, V.; Kirichenko, K.; Tsatsakis, A.; Golokhvast, K. Risk assessments in nanotoxicology: Bioinformatics and computational approaches. *Current Opinion in Toxicology* **2020**, *19*, 1−6.

(14) Oh, E.; Liu, R.; Nel, A.; Gemill, K. B.; Bilal, M.; Cohen, Y.; Medintz, I. L. Meta-analysis of cellular toxicity for cadmium-containing quantum dots. *Nat. Nanotechnol.* **2016**, *11* (5), 479.

(15) Pun, G. P. P.; Batra, R.; Ramprasad, R.; Mishin, Y. Physically informed artificial neural networks for atomistic modeling of materials. *Nat. Commun.* **2019**, *10*, 2339 DOI: 10.1038/s41467-019-10343-5.

(16) 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), 1−10.

(17) Yang, S. P.; Bar-Ilan, O.; Peterson, R. E.; Heideman, W.; Hamers, R. J.; Pedersen, J. A. Influence of Humic Acid on Titanium Dioxide Nanoparticle Toxicity to Developing Zebrafish. *Environ. Sci. Technol.* **2013**, *47* (9), 4718−4725.

(18) 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.

(19) Singh, A. V.; Maharjan, R. S.; Kanase, A.; Siewert, K.; Rosenkranz, D.; Singh, R.; Laux, P.; Luch, A. Machine-Learning-Based Approach to Decode the Influence of Nanomaterial Properties on Their Interaction with Cells. *ACS Appl. Mater. Interfaces* **2021**, *13* (1), 1943−1955.

(20) Trinh, T. X.; Seo, M.; Yoon, T. H.; Kim, J. Developing random forest based QSAR models for predicting the mixture toxicity of TiO2 based nano-mixtures to Daphnia magna. *NanoImpact* **2022**, *25*, 100383.

(21) Ye, N.; Wang, Z.; Wang, S.; Peijnenburg, W. J. G. M. Toxicity of mixtures of zinc oxide and graphene oxide nanoparticles to aquatic organisms of different trophic level: particles outperform dissolved ions. *Nanotoxicology* **2018**, *12* (5), 423−438.

(22) 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.

(23) Chen, G. C.; Peijnenburg, W. J.; Kovalishyn, V.; Vijver, M. G. Development of nanostructure−activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. *RSC Adv.* **2016**, *6* (57), 52227−52235.

(24) 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.

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

(26) Chen, G. C.; Peijnenburg, W. J. G. M.; Xiao, Y. L.; Vijver, M. G. Developing species sensitivity distributions for metallic nanomaterials considering the characteristics of nanomaterials, experimental conditions, and different types of endpoints. *Food Chem. Toxicol.* **2018**, *112*, 563−570.

(27) Bunmahotama, W.; Vijver, M. G.; Peijnenburg, W. Development of a Quasi−Quantitative Structure−Activity Relationship Model for Prediction of the Immobilization Response of Daphnia magna

Exposed to Metal-Based Nanomaterials. *Environ. Toxicol. Chem.* **2022**, *41* (6), 1439−1450.
(28) Kar, S.; Gajewicz, A.; Puzyn, T.; Roy, K. Nano-quantitative structure-activity relationship modeling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells. *Toxicol In Vitro* **2014**, *28* (4), 600−606.
(29) Forese, R. a. D. P. E.*FishBase*; Forese, R. a. D. P., Ed.; World Wide Web electronic publication, 2022.
(30) Ji, Z. W.; Guo, W. J.; Wood, E. L.; Liu, J.; Sakkiah, S.; Xu, X. M.; Patterson, T. A.; Hong, H. X. Machine Learning Models for Predicting Cytotoxicity of Nanomaterials. *Chem. Res. Toxicol.* **2022**, *35* (2), 125−139.
(31) 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.
(32) 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.
(33) Papa, E.; Doucet, J. P.; Doucet-Panaye, A. Linear and nonlinear modelling of the cytotoxicity of TiO2 and ZnO nanoparticles by empirical descriptors. *SAR and QSAR in Environmental Research* **2015**, *26* (7−9), 647−665.
(34) Wilson, M. D.Support Vector Machines. In *Encyclopedia of Ecology*; Jørgensen, S. E., Fath, B. D., Eds.; Academic Press, 2008; pp 3431−3437.
(35) 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.
(36) Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Statistics Surveys* **2010**, *4* (none), 40−79.
(37) Eriksson, L.; Byrne, T.; Johansson, E.; Trygg, J.; Vikström, C.*Multi-and megavariate data analysis basic principles and applications*; Umetrics Academy, 2013.
(38) 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.
(39) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science* **2003**, *22* (1), 69−77.
(40) Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* **2007**, *8* (1), 25.
(41) Cafri, G.; Li, L.; Paxton, E. W.; Fan, J. Predicting risk for adverse health events using random forest. *Journal of Applied Statistics* **2018**, *45* (12), 2279−2294.
(42) Thiese, M. S.; Ronna, B.; Ott, U. P value interpretations and considerations. *J. Thorac Dis* **2016**, *8* (9), E928−e931.
(43) Steidinger, B. S.; Crowther, T. W.; Liang, J.; Van Nuland, M. E.; Werner, G. D. A.; Reich, P. B.; Nabuurs, G. J.; de-Miguel, S.; Zhou, M.; Picard, N.; et al. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **2019**, *569* (7756), 404−408.
(44) Kamble, V.; Deshmukh, S. Comparision between accuracy and MSE, RMSE by using proposed method with imputation technique. *Oriental journal of computer science and technology* **2017**, *10* (4), 773−779.
(45) McShan, D.; Ray, P. C.; Yu, H. Molecular toxicity mechanism of nanosilver. *Journal of Food and Drug Analysis* **2014**, *22* (1), 116−127.
(46) Skjolding, L. M.; Kruse, S.; Sørensen, S. Nør.; Hjorth, R.; Baun, A. A Small-Scale Setup for Algal Toxicity Testing of Nanomaterials and Other Difficult Substances. *JoVE* **2020**, No. 164, e61209.
(47) Abdolahpur Monikh, F.; Chupani, L.; Guo, Z.; Zhang, P.; Darbha, G. K.; Vijver, M. G.; Valsami-Jones, E.; Peijnenburg, W. The stochastic association of nanoparticles with algae at the cellular level: Effects of NOM, particle size and particle shape. *Ecotoxicol Environ. Saf* **2021**, *218*, 112280.
(48) Cassano, A.; Robinson, R. L. M.; Palczewska, A.; Puzyn, T.; Gajewicz, A.; Tran, L.; Manganelli, S.; Cronin, M. T. D. Comparing the CORAL and Random Forest Approaches for Modelling the In Vitro Cytotoxicity of Silica Nanomaterials. *Alternatives to Laboratory Animals* **2016**, *44* (6), 533−556.
(49) Mirzaei, M.; Furxhi, I.; Murphy, F.; Mullins, M. A Machine Learning Tool to Predict the Antibacterial Capacity of Nanoparticles. *Nanomaterials* **2021**, *11* (7), 1774.
(50) Eriksson, L.; Kettanehwold, N.; Trygg, J.; Wikström, C.; Wold, S.*Multi- and Megavariate Data Analysis: Part I: Basic Principles and Applications*; Umetrics, Inc., 2006.
(51) Toropova, A. P.; Toropov, A. A.; Leszczynski, J.; Sizochenko, N. Using quasi-SMILES for the predictive modeling of the safety of 574 metal oxide nanoparticles measured in different experimental conditions. *Environmental Toxicology and Pharmacology* **2021**, *86*, 103665.
(52) Liang, D. Y.; Wang, X. R.; Fan, W. H. Potential application of Au core labeling for tracking Ag nanoparticles in the aquatic and biological system. *Water Res.* **2022**, *215*, 118280.
(53) Oksel, C.; Ma, C. Y.; Liu, J. J.; Wilkins, T.; Wang, X. Z. (Q)SAR modelling of nanomaterial toxicity: A critical review. *Particuology* **2015**, *21*, 1−19.
(54) Connell, D. W.; Yu, Q. J.; Verma, V. Influence of exposure time on toxicity—An overview. *Toxicology* **2016**, *355*−*356*, 49−53.
(55) 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.
(56) Wang, Y.; Wu, F. C.; Mu, Y. S.; Zeng, E. Y.; Meng, W.; Zhao, X. L.; Giesy, J. P.; Feng, C. L.; Wang, P. F.; Liao, H. Q.; et al. Directly Predicting Water Quality Criteria from Physicochemical Properties of Transition Metals. *Sci. Rep.* **2016**, *6* (1), 22515.
(57) Gutierrez, M. F.; Ale, A.; Andrade, V.; Bacchetta, C.; Rossi, A.; Cazenave, J. Metallic, metal oxide, and metalloid nanoparticles toxic effects on freshwater microcrustaceans: An update and basis for the use of new test species. *Water Environment Research* **2021**, *93* (11), 2505−2526.
(58) Spurgeon, D.; Lahive, E.; Robinson, A.; Short, S.; Kille, P.Species Sensitivity to Toxic Substances: Evolution, Ecology and Applications. *Frontiers in Environmental Science***2020**, *8*. DOI: DOI: 10.3389/fenvs.2020.588380.
(59) Hoffmann, M. R.; Martin, S. T.; Choi, W.; Bahnemann, D. W. Environmental Applications of Semiconductor Photocatalysis. *Chem. Rev.* **1995**, *95*, 69−96.
(60) Lu, X.; Rycenga, M.; Skrabalak, S. E.; Wiley, B.; Xia, Y. Chemical synthesis of novel plasmonic nanoparticles. *Annu. Rev. Phys. Chem.* **2009**, *60*, 167−192. From NLM.
(61) Nedyalkova, M.; Dimitrov, D.; Donkova, B.; Simeonov, V. Chemometric Evaluation of the Link between Acute Toxicity, Health Issues and Physicochemical Properties of Silver Nanoparticles. *Symmetry* **2019**, *11* (9), 1159.
(62) 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.
(63) 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.