

RESEARCH

Open Access



A Nano-QSTR model to predict nano-cytotoxicity: an approach using human lung cells data

João Meneses¹, Michael González-Durruthy¹, Eli Fernandez-de-Gortari¹, Alla P. Toropova², Andrey A. Toropov² and Ernesto Alfaro-Moreno^{1*}

Abstract

Background The widespread use of new engineered nanomaterials (ENMs) in industries such as cosmetics, electronics, and diagnostic nanodevices, has been revolutionizing our society. However, emerging studies suggest that ENMs present potentially toxic effects on the human lung. In this regard, we developed a machine learning (ML) nano-quantitative-structure-toxicity relationship (QSTR) model to predict the potential human lung nano-cytotoxicity induced by exposure to ENMs based on metal oxide nanoparticles.

Results Tree-based learning algorithms (e.g., decision tree (DT), random forest (RF), and extra-trees (ET)) were able to predict ENMs' cytotoxic risk in an efficient, robust, and interpretable way. The best-ranked ET nano-QSTR model showed excellent statistical performance with R^2 and Q^2 -based metrics of 0.95, 0.80, and 0.79 for training, internal validation, and external validation subsets, respectively. Several nano-descriptors linked to the core-type and surface coating reactivity properties were identified as the most relevant characteristics to predict human lung nano-cytotoxicity.

Conclusions The proposed model suggests that a decrease in the ENMs diameter could significantly increase their potential ability to access lung subcellular compartments (e.g., mitochondria and nuclei), promoting strong nano-cytotoxicity and epithelial barrier dysfunction. Additionally, the presence of polyethylene glycol (PEG) as a surface coating could prevent the potential release of cytotoxic metal ions, promoting lung cytoprotection. Overall, the current work could pave the way for efficient decision-making, prediction, and mitigation of the potential occupational and environmental ENMs risks.

Keywords Engineered nanomaterials, Computational nanotoxicology, Machine learning, Nano-QSTR, Lung nano-cytotoxicity, A549 cell line

*Correspondence:

Ernesto Alfaro-Moreno
ernesto.alfaro@inl.int

¹NanoSafety Group, International Iberian Nanotechnology Laboratory,
Braga 4715-330, Portugal

²Instituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri
2, Milano 20156, Italy



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Engineered nanomaterials (ENMs) based on metal oxide nanoparticles offer a wide range of promising applications, including cosmetics, electronics, sunscreens, textiles, biomedical products, and diagnostic nanodevices, among others [1–5]. Although inorganic ENMs offer multiple technological advantages and reveal exciting physicochemical properties, understanding their interaction with a biological environment is still challenging. Growing evidence demonstrates that some inorganic ENMs (e.g., CuO, ZnO, Fe₂O₃, CeO₂, Ag, Au, and TiO₂) could be potentially more toxic than their organic counterparts, such as carbon-based ENMs [2]. It is well-recognized that, from the occupational and molecular epidemiology point of view, engineered inorganic ENMs present a higher potential to induce several human lung epithelial perturbations mainly based on the intracellular increase of reactive oxygen species (ROS) [6, 7], which usually play an important role in the high prevalence of human lung nano-cytotoxicity of ENMs at the molecular, cellular, and subcellular levels (e.g., mitochondria, lysosomes, and nuclei) [7, 8].

Over the last few decades, various sampling strategies have been used to determine the ENMs occupational exposure. However, there is still no international consensus on measurement strategies, metrics, or exposure limits, as toxicity studies of ENMs have generally been conducted in non-human *in vitro* cell-based models. The assessment of individual human exposure to ENMs remains a critical issue despite recent innovative developments in personal measurement nanodevices [9]. In this regard, most of the research institutes that synthesize and manufacture ENMs, manage detailed action plans to mitigate the personal nano-exposure of workers, mainly by the respiratory tract [9–13]. Herein, the current nanorisk assessment paradigm was developed by the U.S. National Academy of Sciences and the Federal Government by considering four critical steps: (i) hazard identification, (ii) dose-response assessment, (iii) exposure assessment, and (iv) nano-risk characterization [14, 15]. Despite the numerous *in vitro* and *in vivo* studies to tackle the relationship between the physicochemical properties of ENMs and their nanotoxicological responses, the obtained evidence is often contradictory or nonconclusive [9].

Recently, there has been unprecedented global interest in improving human nanosafety relevance, which is possible through *in silico* models [16]. This interest has been significantly supported by increased top-down investment from central sources such as the EPA-Tox21 Consortium [17], the National Institute of Health [18], the International Organization for Standardization (ISO) [19], the European Commission through the Horizon 2020 Initiative, and the Organisation for Economic

Co-operation and Development [20]. The efforts include the development of several computational models for nanotoxicology predictions [16]. Overall, *in silico* approaches are greatly beneficial to address the current concerns on ENMs nanotoxicity, as they introduce predictive animal-free technologies based on state-of-the-art machine learning (ML) methods. Newly published in Nano Today journal, Burden et al. [21] strongly suggest that by using sophisticated ML-based methodologies, which rigorously follow the 3Rs ethical principles adapted to nanosafety, it is possible to extrapolate *in vitro* exposure effects to explain human exposure [22, 23].

Following this idea, predictive *in silico* approaches could efficiently address Nano-Quantitative Structure-Activity/Property/Toxicity Relationships (nano-QSAR/QSPR/QSTR) of ENMs based on metal oxide nanoparticles to prevent potential human lung nano-cytotoxicity. Although the classical term to address this type of problem is QSAR, which gives a broad framework capable of integrating the most up-to-date models, in this work, the fundamental concept is nano-QSTR. Such a term derives from addressing ENMs from a toxicological point of view [24–26]. Given the complexity of the ENMs, the nano-QSTR model application is restricted by the use of physicochemical properties as nano-descriptors. Nevertheless, with the advancement of Artificial intelligence (AI) and Data Science in the last years, establishing relationships between the physicochemical properties (e.g., electronegativity, ionization potential, van der Waals radius, among others) of a complex system such as ENMs and their nanotoxicity is reasonable [27–29].

In light of the basic concept underlying the implementation of nano-QSTR predictive models, it assumes that a given set of similar structures (e.g., ENMs) have an equivalent toxicological behaviour. Thus, a subtle structural change in the ENMs composition, such as different crystallographic cores, the presence or absence of doping agents, or surface coatings, among others, should lead to a slight divergence in the toxicological paradigm. In contrast, the advances in AI and ML continue to provide large opportunities to move forward in the mechanistic understanding of nanotoxicology responses [30]. In this regard, there are several examples of transparent and understandable techniques, including multiple linear regression (MLR), partial least squares (PLS) regression, decision tree (DT), and random forest (RF), among others [31]. Besides, such algorithms should be supported by strategies that present a visual glance of the diversity and homogeneity of the data distributions and, together with the linear and non-linear correlation among the ENMs nano-descriptors, allow the exclusion of redundant information.

Therefore, this work aims to develop a novel and robust ML nano-QSTR model to predict the potential human

lung nano-cytotoxicity induced by ENMs based on metal oxide nanoparticles. Moreover, the present work is an effort to pave the way for using *in silico* tools to efficiently predict ENMs potential occupational risks and make regulatory decisions in nanotoxicology and environmental health.

Materials and methods

Dataset

The data of 16 ENMs for human lung carcinoma cell line A549 have been taken from the literature [32] and recently the same data set was used by us using a quasi-SMILES approach [33]. The dataset contains 377 observations ($N=377$) on cell viability (%), and covers several relevant assay conditions, such as the different composition of the core, doping, surface coating, diameter (nm), and concentrations ($\mu\text{g/ml}$) of the ENMs. An overview of the physicochemical composition of the ENMs is presented in Table 1; Fig. 1. Complete data of this subsection is available in the Supplementary Information (see Additional File 1, Table S1).

Nano-descriptors calculation

A crucial step in developing ML nano-QSTR models is calculating the adequate nano-descriptors. Nano-descriptors are numerical forms of nanoparticle properties (e.g., electronic, physicochemical, structural, and topological) that could be used to represent ENMs and might be quantitatively associated with cytotoxicity. Moreover, nano-descriptors contain relevant information on metal, non-metal, and semimetals obtained from the periodic table and other literature sources [34]. A set of 31 nano-descriptors (e.g., Van der Waals radius of the active metal, the number of metallic elements, and the number of electrons of the active metal) was calculated using the Elemental-Descriptor (software version 1.0, Gdansk, Poland). Complete data of this subsection

is available in the Supplementary Information (see Additional file 1, Table S2 and Table S3).

Dataset pre-processing

In the context of ML nano-QSTR models, applying data pre-processing techniques represents a step that efficiently helps improve data quality by extracting relevant features from the raw data. Overall, the data pre-processing includes several procedures, such as cleaning, organizing, and structuring the data into an understandable and readable input format for building ML nano-QSTR models to predict a given biological response (i.e., cytotoxicity-induce by ENMs).

In the present study, we carried out the following pre-processing steps: (i) filter the ENMs with a diameter of less than 200 nm, reducing the number of available ENMs to 11 from the original amount of 16 ENMs, and the number of observations to $N=333$ from the initial amount of $N=377$; (ii) encode the categorical nano-descriptors (core, doping, and surface coating) into numerical readable inputs by using a One-Hot Encoding procedure [35]; (iii) application of a standardization procedure based on the Z-score normalization method, where the values are centered on the mean with a unit standard deviation. The standardization procedure is represented by Eq. (1), where μ is the mean of the ENMs nano-descriptor values and σ its standard deviation:

$$X' = \frac{X - \mu}{\sigma} \quad (1)$$

Lastly, the iv) pre-processing step was to approximate the shape of the distribution of each numerical nano-descriptor to a Gaussian distribution by applying the Yeo-Johnson transformation [36]. Complete data of this subsection is available in the Supplementary Information (see Additional file 2, Figure S1, Figure S2 and Figure S3).

Table 1 A short version of the original dataset covering relevant assay conditions, such as the different physicochemical nature of the core, doping, surface coating, diameter (nm), and concentrations ($\mu\text{g/ml}$) of the ENMs along with the cell viability (%) data

| ENMs ID | Core | Doping | Surface Coating | Diameter (nm) | Concentration ($\mu\text{g/ml}$) | Cell Viability (%) |
|---------|------|-----------|-----------------|---------------|------------------------------------|--------------------|
| 1 | ZnO | ND | NSC | 68.90 | 0.01 | 100.00 |
| 2 | ZnO | ND | NSC | 68.90 | 1.56 | 104.79 |
| 3 | ZnO | ND | NSC | 68.90 | 3.13 | 109.64 |
| 4 | ZnO | ND | NSC | 68.90 | 6.25 | 112.06 |
| 5 | ZnO | ND | NSC | 68.90 | 12.50 | 120.16 |
| 6 | ZnO | ND | NSC | 68.90 | 25.00 | 47.35 |
| 7 | ZnO | Na (1.5%) | NSC | 5.50 | 0.01 | 100.00 |
| 8 | ZnO | Na (1.5%) | NSC | 5.50 | 6.25 | 94.61 |
| 9 | ZnO | Na (1.5%) | NSC | 5.50 | 12.50 | 93.56 |
| 10 | ZnO | Na (1.5%) | NSC | 5.50 | 25.00 | 93.57 |

ND: No doping; NSC: No surface coating

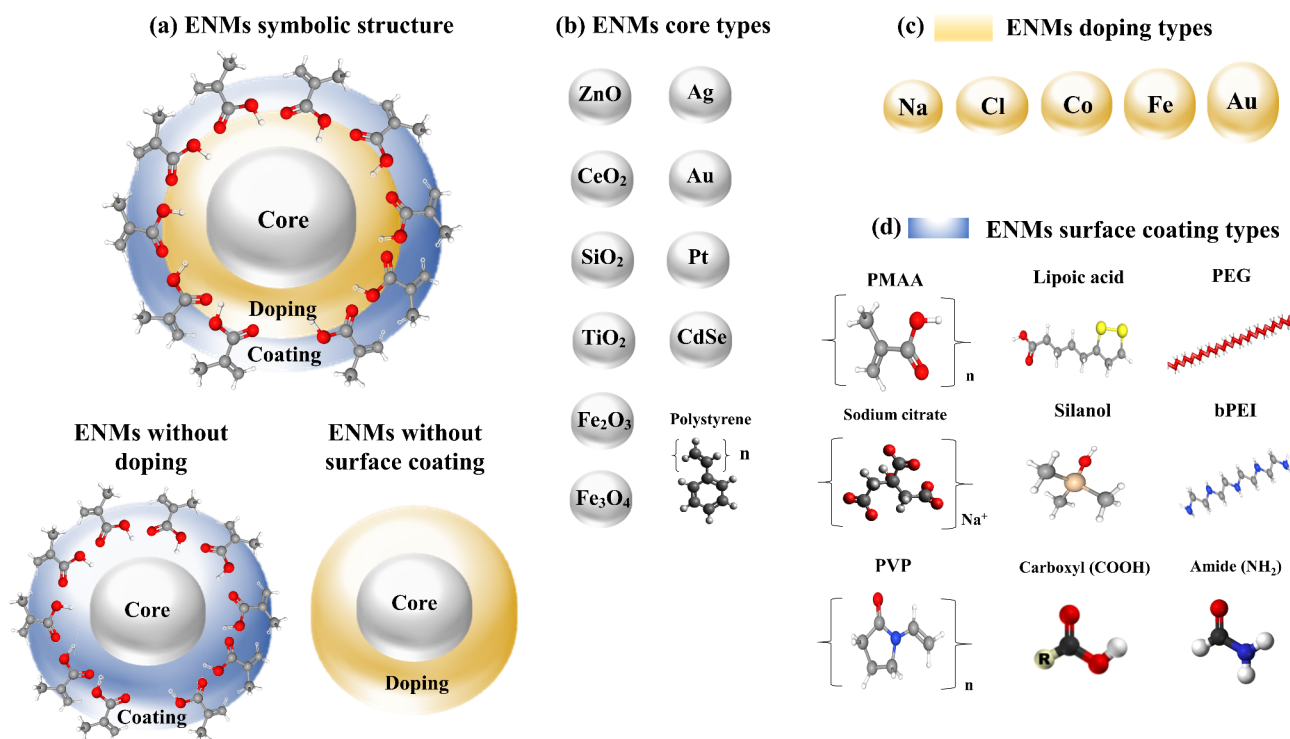


Fig. 1 General representation of the structure and chemical composition of the engineered nanomaterials (ENMs), evaluated in this work. **(a)** Depiction of the whole composition of an ENM with the corresponding counterparts without doping and surface coating shown below; **(b)** Representation of the different ENMs core compositions; **(c)** and **(d)** Ball-and-stick model representation of the different doping and coating types evaluated for the ENMs, respectively

Dataset splitting

The first step was to withhold a random sample of $N=33$ from a total of $N=333$ to simulate an unseen dataset. Another way to think about this step is that 33 observations were unavailable to train and evaluate the ML nano-QSTR models. Afterward, the 300 observations in the dataset were randomly divided into a training subset of $N=210$ (70%) and a test subset of $N=90$ (30%), respectively [37]. The training subset was used to train the ML nano-QSTR models, whereas the test subset was employed to evaluate its predictive performance. Complete data of this subsection is available in the Supplementary Information (see Additional file 1, Table S4 and Table S5).

ML nano-QSTR model development

Considering the trade-off between ML nano-QSTR model performance and its interpretability, the ENMs nano-descriptors were combined by simple arithmetical operations and converted into new statistically significant nano-descriptors. For instance, two ENMs nano-descriptors were multiplied to be more preponderant in explaining the data variance than the same two ENMs nano-descriptors separately. The next step was to reduce the dimensionality of the data, reduce the computational cost, and minimize the redundancy between the ENMs

nano-descriptors. Three key operations were performed, namely (i) remove the ENMs nano-descriptors with a low variance; (ii) remove highly inter-correlated ENMs nano-descriptors; (iii) implement a combination of various permutation importance techniques to achieve the final subset of ENMs nano-descriptors [38], such as Shapley additive explanations (SHAP), which explains the contribution of each ENMs nano-descriptor to the ML nano-QSTR model [39]. Afterward, several tree-based algorithms were used to make the ML pipeline more transparent and interpretable, such as decision tree (DT), random forest (RF), and extra-trees regressor (ET) [31]. Complete data of this subsection is available in the Supplementary Information (see Additional file 1, Table S6).

ML nano-QSTR model validation

In the present study, we followed the principles of the Organization for Economic Co-operation and Development (OECD) concerning model validation, which establishes that a reliable model should present appropriate goodness-of-fit measures, robustness, and predictivity performance [40]. In this regard, there are two methods to evaluate the goodness of a model: (i) internal and (ii) external validation. The (i) internal validation evaluates the fitting of the model on existing data (training set); the (ii) external validation evaluates future data,

i.e., how reliable the model can predict new data (test set and unseen set). Here, the (i) internal validation of the regression-based models was determined based on several statistical metrics such as determination coefficient (R^2), determination coefficient based-metrics (Q^2_{LOO}), root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of concordance (CCC). Moreover, the robustness of the model was represented by a 5-fold cross-validation process [41]. The (ii) external validation of the models was determined using similar statistical parameters, such as R^2_{ext} , Q^2_{F1} , Q^2_{F2} , RMSE, MSE, and CCC. All statistical metrics were computed via DTC Lab Xternal Validation Plus (software version 1.2, India) [42].

Applicability domain

According to the OECD third principle [40], a QSAR model to predict a given biological response (i.e., ENMs cytotoxicity) should be associated with a defined applicability domain (AD). The AD is a theoretical region of the chemical space that contains both model nano-descriptors and modelled responses, in which the model makes predictions with a given reliability [43]. Herein, the AD was calculated using a standardization approach and retrieved via DTC Lab Applicability Domain Calculator (software version 1.0, India). For calculating the AD, Eq. (2) was used:

$$S_{\text{new}(k)} = \overline{S}_k + 1.28 \times \sigma_{S_k} \quad (2)$$

Where $S_{\text{new}(k)}$ is S_{new} value for ENM_k , \overline{S}_k is the mean of the standardized nano-descriptors for ENM_k (from the training, test, or unseen set), and σ_{S_k} is the standard deviation of standardized nano-descriptors for ENM_k (from the training, test, or unseen set). Overall if the $S_{\text{new}(k)}$ is lower or equal to 3, then the ENM_k is not an outlier (if in the training set) or is within the AD (if in the test or unseen set) [43].

Summary

An overview of the implemented data-driven approach is presented in Fig. 2. All algorithms were implemented in Python (software version 3.9), using libraries such as Pycaret (software version 2.3.8) and scikit learn (software version 0.23.2). Altogether, the experiments allowed the extraction of valuable insights from the dataset and provided the baseline to construct the ML nano-QSTR models. The definition of AD made it possible to understand the limitations and boundaries where the predicted values can be trusted with confidence.

Results

Nano-descriptors distribution and diversity

Based on the data of 16 engineered nanomaterials (ENMs), under different experimental conditions, a ML

nano-QSTR model with cell viability endpoint as the dependent variable was established. As previously mentioned, the ultimate goal of the developed model was to predict the potential human lung nano-cytotoxicity.

As the application of any data-driven algorithm requires a comprehensive understanding of the data, one of the first concerns was to consider the differences in the structure and chemical composition of the ENMs, extract valuable insights from the dataset and provide a baseline to construct the ML nano-QSTR model. In this regard, Fig. 3 presents an overview of the dataset characteristics, such as the endpoint frequency distribution, and the diversity of core, doping, and surface coating nanomaterials composition.

It is not uncommon to find a skewed frequency distribution for ENMs nano-descriptors from nanotoxicological experimental data, i.e., some types of structural attributes appear much more frequently [44]. In this regard, the evaluated endpoint (cell viability) follows a normal distribution, predominantly ranging from 75 to 125%, with a mean value of 96% and a standard deviation of 23% (Fig. 3(a)). Regarding the composition of the ENMs, based on their frequency distribution, the core and the surface coating features present a qualitative and quantitative diversity, with the most prevalent core types being Zn, Ag, SiO_2 , and surface coating compositions being the ENMs without surface coating, and the ENMs coated with PMAA, and sodium citrate (Fig. 3(b) and Fig. 3(d)). The frequency distribution of the different doping types is predominantly represented by the absence of doping in the ENMs and may seem skew toward that direction (Fig. 3(c1)). However, it is important to note that for ENMs with and without doping, the frequency distribution of both types is much more balanced (Fig. 3(c2)). To complement this analysis, Fig. 3(e) to Fig. 3(g) present the trend of the cell viability variation with representative examples of each nano-descriptor. Despite the existence of specific ENMs compositions that significantly vary the cell viability, the tendency of variation agrees with the mean and standard deviation previously mentioned. A similar frequency distribution analysis was performed on the set of 31 ENMs nano-descriptors, such as the number of metallic elements, the Van der Waals radius of the active metal, and the number of electrons of the active metal, among other examples (see Additional file 2, Figure S4). Overall, each nano-descriptor depicts a high degree of diversity by presenting several attributes that can span the ENMs structural space.

ML nano-QSTR model performance and validation

Through the dataset exploration and characterization conducted in the previous section, the richness of the dataset was assured, providing the basis for the

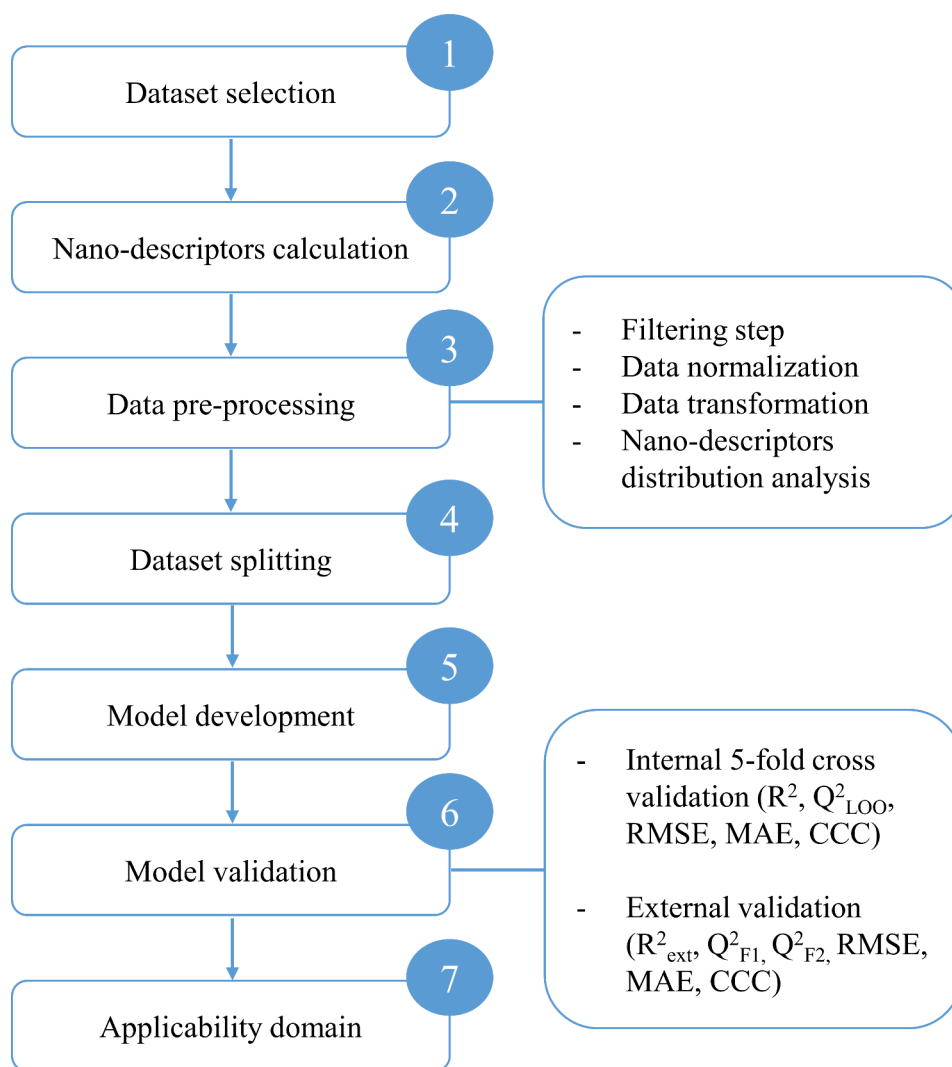


Fig. 2 Overview of ML nano-QSTR model approach to predict the potential human lung nano-cytotoxicity induced by ENMs based on metal oxide nanoparticles

development of the ML nano-QSTR model. In this particular problem, three interpretable learning algorithms, including decision tree (DT), random forest (RF), and extra-trees regressor (ET) are presented. Although DT, RF, and ET belong to the same family of learning algorithms, i.e., tree-based models that use conditional statements to make predictions, DT is the simplest. Therefore, it is expected that DT presents slightly lower statistical metrics than RF and ET. The use of such an algorithmic family is in agreement with recent studies that point out that an interpretable learning algorithm is more valuable to experimentalists than a highly predictive but black-box model since its interpretation is complex and non-trivial [31, 45]. To verify the reliability and robustness of the developed models, Table 2 presents an overview of the statistical metrics for training, validation, and test sets.

Regarding the evaluation of the developed models on existing data, i.e., training subset, DT, RF, and ET

present an R^2 between 0.95 and 0.96, highlighting their high statistical performance in learning the behavior of the training ENMs. Concerning DT, RF, and ET internal validation, the 5-fold cross-validation process enhanced their robustness as R^2 for each model is between 0.7 and 0.8. As previously mentioned, with the slight increase in model complexity, there is an increment in the statistical performance as DT presents an R^2 of 0.73, RF of 0.76, and ET of 0.79. Overall, the internal validation of each model is guaranteed by R^2 and Q^2_{LOO} to be in the same order of magnitude. As for the evaluation of the developed models on new data, i.e., test subset, DT presents an R^2_{ext} of 0.76, while RF and ET present an R^2_{ext} of 0.79. This statistical parameter is complemented by the determination coefficient based-metrics (Q^2_{F1} and Q^2_{F2}), which depict similar values. Altogether, both internal and external validation confirms that the developed models have the potential to reliably predict A549 cell line viability.

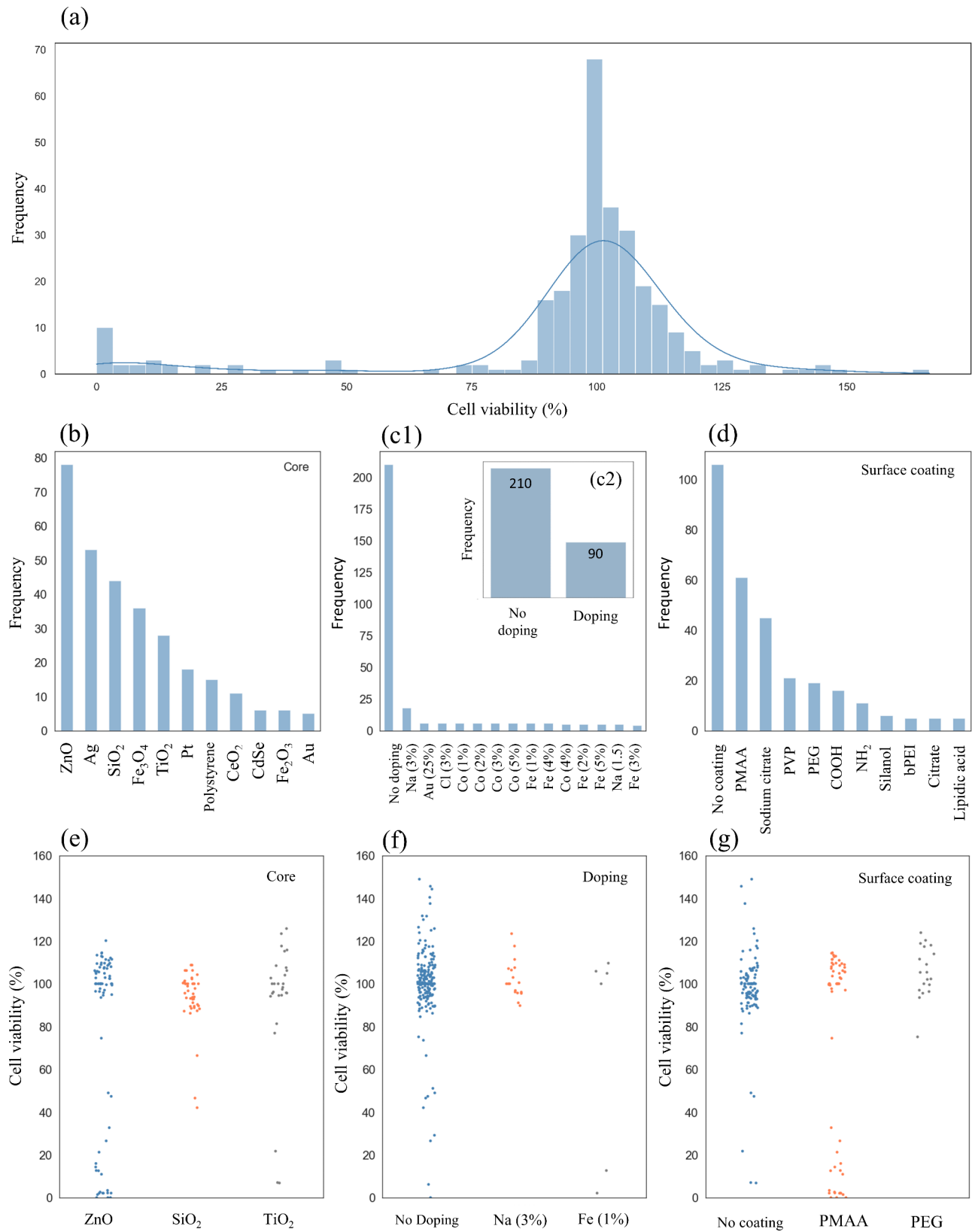


Fig. 3 Dataset overview. **(a)** Cell viability (%) frequency distribution; **(b)**, **(c1)**, **(c2)**, and **(d)** Diversity of core, doping, and surface coating nanomaterials composition; **(e)**, **(f)**, and **(g)** Cell viability (%) variation trend with representative examples of each nano-descriptor

Table 2 ML nano-QSTR models performance for training, validation, and test sets

| Model | Subset | R ² | R ² _{ext} | Q ² _{LOO} | Q ² _{F1} | Q ² _{F2} | RMSE | MAE | CCC |
|-------|------------|----------------|-------------------------------|-------------------------------|------------------------------|------------------------------|--------|-------|-------|
| DT | Training | 0.953 | - | - | - | - | 6.147 | 3.840 | 0.976 |
| | Validation | 0.733 | - | 0.730 | - | - | 14.103 | 8.441 | 0.851 |
| | Test | - | 0.765 | - | 0.765 | 0.765 | 13.264 | 7.139 | 0.868 |
| RF | Training | 0.965 | - | - | - | - | 5.299 | 2.963 | 0.981 |
| | Validation | 0.768 | - | 0.767 | - | - | 13.266 | 7.908 | 0.866 |
| | Test | - | 0.790 | - | 0.789 | 0.789 | 12.573 | 6.879 | 0.881 |
| ET | Training | 0.953 | - | - | - | - | 6.158 | 3.912 | 0.975 |
| | Validation | 0.798 | - | 0.797 | - | - | 12.054 | 7.294 | 0.889 |
| | Test | - | 0.788 | - | 0.788 | 0.788 | 12.580 | 6.603 | 0.883 |

DT: decision tree; RF: random forest; ET: extra-trees regressor; R²: determination coefficient (R²_{ext} for external validation); Q: determination coefficient based-metrics (Q²_{LOO} for internal validation; Q²_{F1} and Q²_{F2} for external validation); RMSE: root-mean-square error; MAE: mean absolute error; CCC: coefficient of concordance.

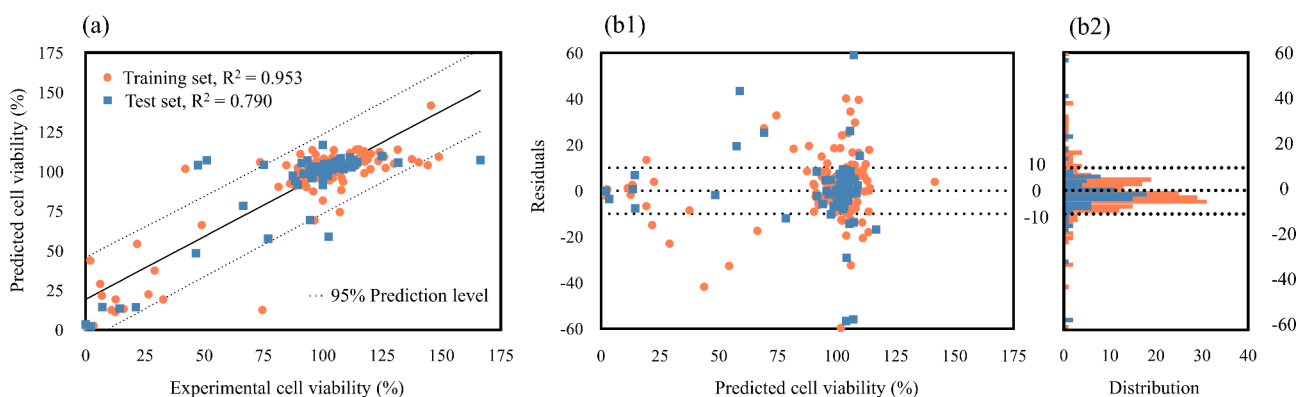


Fig. 4 Overview of experimental and predicted cell viability (%) by ET nano-QSTR model for training and test sets. **(a)** Scatter plot representing the experimental cell viability (%) as a function of the predicted cell viability (%). The straight line illustrates the perfect agreement between experimental and calculated values. The dashed lines represent the 95% prediction level. **(b1)** Scatter plot representing the predicted cell viability (%) as a function of the residuals; **(b2)** Frequency distribution of the predicted cell viability (%). The dashed lines represent the residuals that are between -10 and 10

From a general point of view, the key performance indicator in selecting the model to be used in the final stage of prediction was the internal validation subset. Such a decision relies on ET presenting R² and Q²-based metrics higher than DT in 8% and RF in 4%, while the training and test set statistical parameters are in the same order of magnitude. To corroborate this analysis, Additional file 2, Figure S5 presents the learning curves for the ET nano-QSTR model for training and validation sets, highlighting a slight trade-off between bias and variance, which could be adjusted by having more training instances. Thus, Fig. 4 presents an overview of the agreement between experimental and predicted cell viability values by the ET model for training and test sets.

Through the exploration of Fig. 4(a), it is possible to quantitative and qualitatively highlight the strong correlation between the observed and predicted cell viability values, as most ENMs are within the 95% prediction level range. Such analysis is complemented by Fig. 4(b), which enhances that training and test data points approximately follow a symmetrical distribution, tending to cluster towards the middle of the plot, around lower values of

the y-axis (e.g., most of the residuals are between -10 and 10).

Nano-descriptors interpretation

Bearing in mind that the final aim of the present work is to predict the potential nano-cytotoxicity induced by ENMs on human lung carcinoma cells, it is fundamental to understand the correlation between the nano-descriptors and the cell viability. Moreover, it is noteworthy to highlight the most significant nano-descriptors in the ET nano-QSTR model performance. Thus, Fig. 5 presents an overview of some representative nano-descriptors and their correlation with cell viability.

Through the exploration of Fig. 5(a), it is possible to identify two distinct groups of nano-descriptors that present a negative correlation with cell viability. One group depicts nano-descriptors retrieved from the original dataset, such as PMAA surface coating, Cl (3%) doping, and ENMs concentration. The other group describes a set of nano-descriptors that were calculated to become more preponderant in explaining variances in the cell viability data, such as the mathematical combination of ENMs concentration and PMAA surface, ENMs

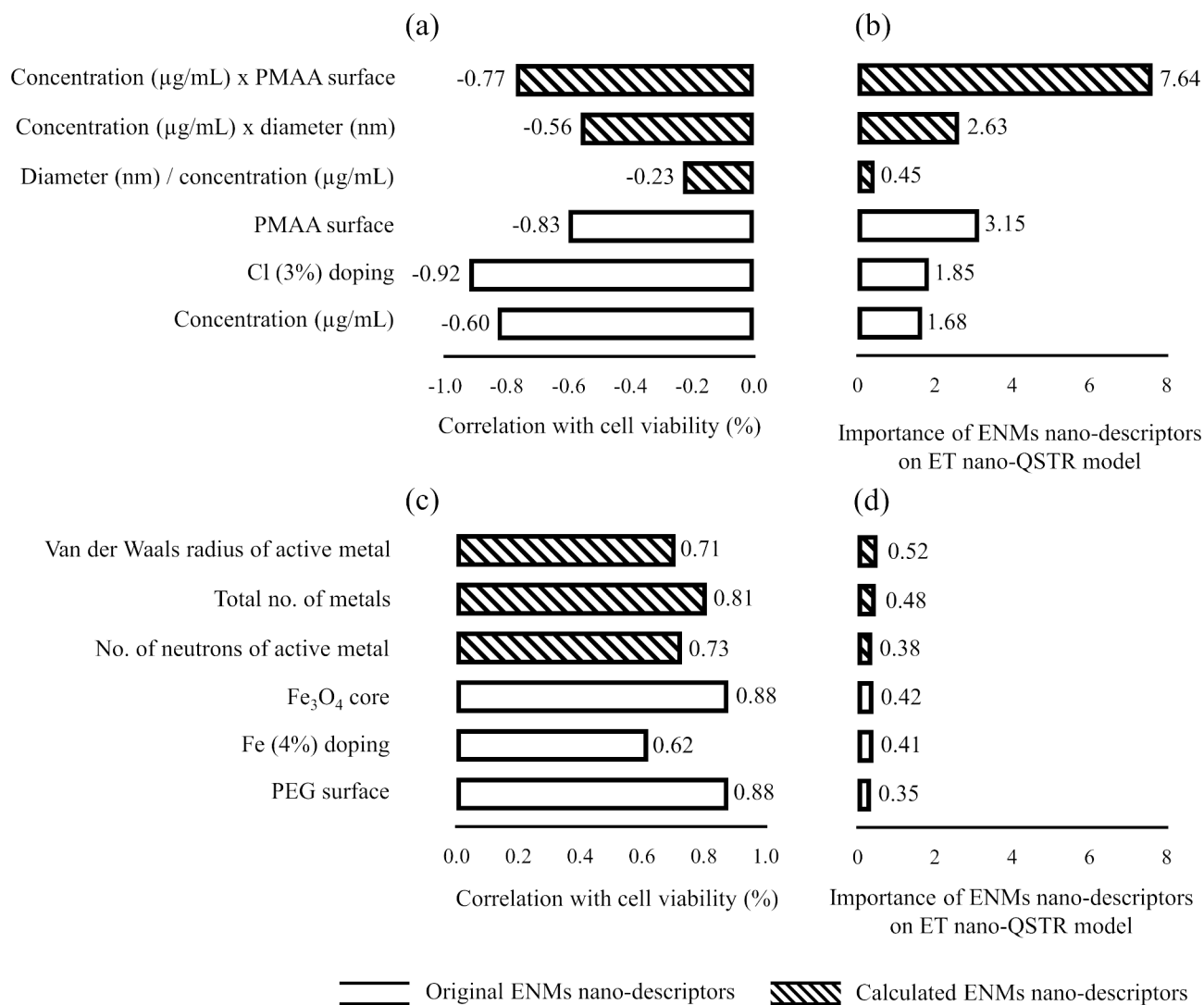


Fig. 5 Summary of representative nano-descriptors. (a) and (c) Negatively and positively correlated nano-descriptors with cell viability (%); (b) and (d) Nano-descriptors effect on ET nano-QSTR model performance

concentration, and ENMs diameter. Overall, Fig. 5(a) highlights nano-descriptors that tend to decrease cell viability and might promote cytotoxicity at the cellular level. As for nano-descriptors that present a positive correlation with the endpoint, Fig. 5(c) highlights Fe_3O_4 core, Fe (4%) doping, and polyethylene glycol (PEG) surface coating as the group of nano-descriptors retrieved from the original dataset. In a similar analysis, the other group describes a set of nano-descriptors obtained from the Elemental descriptor calculator, such as the Van der Waals radius of the active metal, the number of metallic elements, and the number of electrons of the active metal. Comprehensively, Fig. 5(c) enhances nano-descriptors that tend to increase cell viability and might avoid cytotoxicity at the cellular level.

From a modelling point of view, Fig. 5(b) and Fig. 5(d) represent a fundamental aspect as both figures highlight

the set of nano-descriptors that contributed the most to the model performance. Interestingly, the nano-descriptors that present a negative correlation with the endpoint are the most relevant from the entire set. An in-depth analysis of Fig. 5(b) shows that the nano-descriptor with the highest contribution on the model performance results from the mathematical combination of the ENMs concentration and PMAA surface coating.

Considering that ENMs surface area is described as a central factor related to the toxic potential of particulate matter, we did explore if this descriptor could play a relevant role in our model. As detailed above, several nano-descriptors linked to the surface coating reactivity properties and diameter of the ENMs were identified as some of the most relevant characteristics to predict human lung nano-cytotoxicity. To mathematically calculate the surface area of the ENMs, it is necessary to

perform some assumptions, namely (i) the ENM is a perfect sphere, and (ii) the ENM size is the diameter of the sphere. Therefore, even though indirectly, the model does consider the surface area of the ENMs as a driving factor for nano-cytotoxicity. All the assumptions and mathematical equations to calculate the surface area are explained in detail by Shin et al. [32]. Complete data on the direct influence of ENMs surface area on model performance is available in the Supplementary Information (see Additional file 1, Table S7, and Additional file 2, Table S1).

Taking advantage of such knowledge, the ET nano-QSTR model was retrained with the ENMs nano-descriptors identified in Fig. 5 and applied to new data, which was not used to develop and validate the model. Table 3 presents an overview of the statistical metrics of ET nano-QSTR model when applied to the unseen subset.

Table 3 shows a significant increase in ET nano-QSTR model performance as the R^2_{ext} increased from 0.79 to 0.93, representing an increase of 18%. Another interesting observation is the similar order of magnitude between the R^2_{ext} and Q^2 -based metrics. Additionally, the RMSE, MAE, and CCC performance metrics increased from 12.58 to 4.37, 6.60 to 3.52, and 0.883 to 0.96, representing an increase of 65%, 47%, and 8%, respectively. Overall the developed model presents a strong, reliable, and robust performance in predicting cell viability.

Applicability domain of the proposed model

However, it is paramount to identify the border of the optimum prediction space, i.e., applicability domain (AD). From a general point of view, the developed ET nano-QSTR model is valid in a chemical space where the ENMs possess structural and physicochemical properties similar to the ENMs used to train and validate the model. Otherwise, the ENMs might be considered outliers or even out of the AD. It is significant to mention that all the studied ENMs are within the AD. Detailed values of the AD of the ENMs are depicted in the Supplementary Information (see Additional file 1, Table S8).

Discussion

Although human lung nano-cytotoxicity induced by ENMs based on metal oxide nanoparticles is among the current occupational and environmental concerns, it remains unexplored and under-researched. The current standard in vitro and in vivo models used to evaluate

such a type of nano-cytotoxicity are time-consuming, costly, and could involve many ethical concerns in animal experimentation. In this sense, this work presents an ET nano-QSTR model to assist experimental scientists by providing a mechanistic interpretation learned from data of 16 ENMs for the A549 cell line [40, 46].

The mechanistic interpretation regards the set of optimal nano-descriptors, i.e., the most significant nano-descriptors in the model performance, and considers if the nano-descriptors are (i) negatively or (ii) positively correlated with the cell viability [47, 48]. Then, three well-recognized nanotoxicological mechanisms are used to complement such interpretation, including (i) the ENMs core type- and diameter-dependent release of cytotoxic metal ions (e.g., Fe^{2+} , Fe^{3+} , Ag^+ , Au^+ , Ti^{2+} , Cd^{2+}) from the ENMs core reactive surface, which could promote redox-homeostasis perturbations, (ii) the physio-pathological increase of intracellular reactive oxygen species (ROS) levels by mitochondrial dysfunction, and (iii) the nano-bio interaction with binding sites of key molecular targets, such as the human lung epithelial proteins and multiprotein junctional complexes that form the selective permeability barrier of the human lung epithelial, which may induce barrier dysfunction [7, 32, 49, 50].

An in-depth analysis of Fig. 5, which identifies six negatively and positively correlated nano-descriptors with cell viability, suggests that the presence of PEG as a surface coating of a Fe_3O_4 core significantly enhances cell viability and inversely attenuates human lung nano-cytotoxicity. As PEG presents excellent pharmacokinetic properties based on absorption, distribution, metabolism, and excretion (ADME), its presence as a surface coating could significantly reduce the potential release of cytotoxic ions (Fe^{2+}) from the Fe_3O_4 core. Overall, such behavior is congruent with the literature, as the surface-dependent release of cytotoxic metal ions has been well documented in previous experimental works [51–54].

In this regard, the generated divalent Fe^{2+} ions from the ENMs based on metal oxide nanoparticles tend to increase the intracellular concentration of oxygen-free-radical groups (e.g., hydroxyl radical). Such mechanism is explained due to the occurrence of the Fenton-Haber-Weiss reaction at the subcellular level, which is directly associated with molecular lung cytotoxic mechanisms (e.g., mitochondria dysfunction promoted by Fe^{2+} ions) [55–59]. Therefore, the presence of PEG as a surface coating significantly contributes to the inhibition of these cytotoxic signaling pathways.

Table 3 ET Nano-QSTR model performance for unseen subset

| Model | Subset | R^2_{ext} | Q^2_{F1} | Q^2_{F2} | RMSE | MAE | CCC |
|-------|--------|--------------------|------------|------------|-------|-------|-------|
| ET | Unseen | 0.927 | 0.929 | 0.925 | 4.375 | 3.528 | 0.961 |

ET: extra-trees regressor; R^2_{ext} : determination coefficient for external validation; Q^2_{F1} and Q^2_{F2} : determination coefficient based-metrics for external validation; RMSE: root-mean-square error; MAE: mean absolute error; CCC: coefficient of concordance.

In opposition to the previously described behavior of PEG, the presence of PMAA as a surface coating or the Cl^- (3%) as a doping condition tend to decrease cell viability, as they intensify the potential release of cytotoxic metal ions (e.g., Fe^{2+} , Fe^{3+} , Ag^+ , among others) from the inorganic ENMs core or doping composition [51–54, 60]. Such evidence is corroborated by the ET nano-QSTR model, as the most influential nano-descriptor to the model performance derives from the arithmetic combination of the ENMs concentration and PMMA surface coating. Furthermore, the nano-descriptor also presents a high negative correlation ($R^2 = -0.83$) with cell viability and may be inversely linked with human lung nano-cytotoxicity.

Besides, all the core-based ENMs nano-descriptors, such as the number of metallic elements ($R^2 = 0.81$), the Van der Waals radius of the active metal ($R^2 = 0.71$), and the number of electrons of the active metal ($R^2 = 0.73$), are positively correlated with the cell viability. Such structural attributes are directly associated with the metal core-based reactive properties when the ENMs are found in their pristine form and have an intrinsic cytotoxic potential according to the diameter and charge of the inorganic core. Nonetheless, such nano-descriptors do not contribute to intensifying the cytotoxicity from the *in silico* point of view. A reasonable explanation behind this behavior focuses on the presence of a PEG surface coating that stabilizes the ENMs' chemical surface reactivity and avoids the occurrence of direct nano-bio interactions of the metal oxide core with binding sites of the target proteins forming the human lung epithelial (i.e., A549 cells target proteins).

It is well-established that a decrease in the diameter of the metal oxide nanoparticles contributes to a significant increase in their ability to access lung cell compartments, including mitochondria, lysosomes, and nuclei. Therefore, the concentration of metal oxide nanoparticles in such organelles will increase and play a fundamental role in lung epithelial barrier dysfunction [7]. More importantly, the diameter decrease of the evaluated metal oxide nanoparticles could contribute to an exponential increase in the number of reactive atoms expressed on the face-based crystallographic planes of the ENMs core. Such a decrease simultaneously promotes the core chemical reactivity and its potential capacity to interact with relevant lung tissues and cells [7, 61, 62]. The proposed ML nano-QSTR model corroborates this information through the negatively correlated nano-descriptors that result from the mathematical combination of ENMs concentration and ENMs diameter.

Overall, the current mechanistic interpretation brings a novel contribution to assist experimental scientists in understanding and analyzing nanotoxicological data. Nonetheless, achieving a broader domain of applicability

and trustability in nano-QSTR models is still a challenge. Therefore, it is necessary to start or continue to collect human lung experimental data standardly, i.e., to implement the findability, accessibility, interoperability, and reusability (FAIR) data principles [63]. In this regard, it is crucial to highlight that data concerning (i) the potential release of cytotoxic metal ions through time and consequently, (ii) the actual concentration of ENMs that reach the cells were not included in the modeling procedure due to the unavailability of such data for human lung cells.

Indeed, using advanced computational models to predict the effective cellular dose is fundamental to understanding the interaction of submerged materials with biological systems. Recently, DeLoid et al. [64] explored both three-dimensional computational fluid dynamics (CFD) and a newly-developed one-dimensional Distorted Grid (DG) model to predict the delivered dose metrics for submerged ENMs in culture media. The last model was later used and validated in a study by Kowoll et al. [65] to predict the deposition of particles on cellular and intercellular human lung surfaces. Interestingly, the authors highlight both model capabilities and limitations, specifically regarding the spatial distribution of particles on heterogeneous surfaces, which is the case in our study. Such considerations are even more relevant since Kowoll et al. performed the experiments in a human lung cell line (i.e., A549 cells) that fit with the same biological model considered in our *in silico* study.

Therefore, in future works, we plan to address these limitations by incorporating computational particokinetics models to estimate the relationship between the release of cytotoxic metal ions through time, the relevant *in vitro* dose criteria for the dosimetry of ENMs, and their influence on the shape of the dose-response curve [66]. Even with all the challenges, a prospective nano-QSTR model should be performed in a useful way that could lead and orient experimental scientists to decision-making processes about nanotoxicological data.

Conclusions

An ML nano-QSTR model was successfully developed to predict the potential human lung nano-cytotoxicity induced by ENMs based on metal oxide nanoparticles. Results demonstrated that using tree-based learning algorithms (e.g., extra-trees regressor) allowed the development of a simple, interpretable, and robust nano-QSTR model, as ET presented R^2 and Q^2 -based metrics of 0.95, 0.80, and 0.79 for training, internal validation, and external validation subsets. By taking advantage of the advances in AI and ML, which continue to provide opportunities to move forward in the mechanistic understanding of nanotoxicology responses, we could identify the six most significant nano-descriptors in the

model performance. Therefore, we could understand if the nano-descriptors are (i) negatively or (ii) positively correlated with the cell viability. As for the (i) negatively correlated, a decrease in the diameter of the metal oxide nanoparticles contributes to a significant increase in their ability to access lung cell compartments, thus promoting lung epithelial barrier dysfunction. As for the (ii) positively correlated, the presence of PEG as a surface coating significantly stabilizes the ENMs' chemical surface reactivity and avoids the potential release of cytotoxic ions, promoting cell viability and inversely attenuating human lung nano-cytotoxicity. By exploiting such knowledge, the ML nano-QSTR model was retrained with the most significant nano-descriptors and applied to new data (e.g., unseen subset), allowing the increase of R^2 and Q^2 -based metrics from 0.79 to 0.92. Based on these findings, the present work may pave the way to possibly predict ENMs' potential occupational risks and make regulatory decisions in nanotoxicology and environmental health.

List of abbreviations

| | |
|-------|---|
| AI | Artificial Intelligence |
| AD | Applicability Domain |
| ADME | Absorption, Distribution, Metabolism, and Excretion |
| CCC | Coefficient of Concordance |
| DT | Decision Tree |
| ET | Extra-trees |
| ENMs | Engineered Nanomaterials |
| FAIR | Findability, Accessibility, Interoperability, and Reusability |
| ISO | International Organization for Standardization |
| ML | Machine Learning |
| MLR | Multiple Linear Regression |
| OECD | Organization for Economic Co-operation and Development |
| PEG | Polyethylene Glycol |
| PLS | Partial Least Squares |
| Q^2 | Determination Coefficient Based-Metrics |
| QSAR | Quantitative-Structure-Activity Relationship |
| QSPR | Quantitative-Structure-Property Relationship |
| QSTR | Quantitative-Structure-Toxicity Relationship |
| RF | Random Forest |
| R^2 | Determination Coefficient |
| ROS | Reactive Oxygen Species |
| RMSE | Root-Mean-Square Error |
| SHAP | Shapley Additive Explanations |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12989-023-00530-0>.

Additional file 1. Table S1 Original dataset. **Table S2** List and description of calculated nano-descriptors. **Table S3** Dataset with calculated nano-descriptors. **Table S4** Dataset for modelling after the filtering step. **Table S5** Unseen subset after the filtering step. **Table S6** Training and test data after nano-descriptors selection. **Table S7** Dataset assuming data normalization by ENMs surface area. **Table S8** Applicability domain of the proposed ML nano-QSTR model.

Additional file 2. Figure S1 ENMs diameter (nm) frequency distribution (a) before and (b) after the filtering step. **Figure S2** Cell viability (%) frequency distribution (a) before and (b) after the filtering step. **Figure S3** General representation of data normalization and transformation pre-processing steps for ENMs diameter and concentration. **Figure S4** Dataset overview. (a), (b), and (c) Diversity of a set of nano-descriptors obtained

from the Elemental descriptor calculator; (e), (f), and (g) Cell Viability (%) variation trend with representative examples of each nano-descriptor.

Figure S5 Learning curves for ET nano-QSTR model for training and validation sets according to (a) determination coefficient (R^2) and (b) root-mean-square error (RMSE). **Table S1** ET nano-QSTR model performance for training, validation, test, and unseen sets assuming data normalization by ENMs surface area.

Acknowledgements

Not applicable.

Authors' contributions

JM was involved in conceptualization, writing the original draft, model development and evaluation, data analysis and visualization, review, and editing. MGD was involved in data analysis and visualization, review, and editing. EFG was involved in writing the original draft, review, and editing. APT was involved in discussion of the concept and type of datasets to evaluate, review and editing. AAT was involved in discussion of the concept and type of datasets to evaluate, review and editing. EAM was involved in conceptualization, writing the original draft, review, and editing, obtaining the funding to support this research. All authors read and approved the final manuscript.

Funding

This research was funded by the European Union's H2020 project Sinfonia (N.857253). SbDToolBox, with reference NORTE-01-0145-FEDER-000047, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund.

Data Availability

The datasets supporting the conclusions of this article are included in the Supplementary Information. The python script can be retrieved from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Received: 2 February 2023 / Accepted: 1 May 2023

References

- Mu Y, Wu F, Zhao Q, Ji R, Qie Y, Zhou Y, et al. Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. *Nanotoxicology*. 2016;10:1207–14.
- Karlsson HL, Cronholm P, Gustafsson J, Möller L. Copper oxide nanoparticles are highly toxic: a comparison between Metal Oxide Nanoparticles and Carbon Nanotubes. *Chem Res Toxicol*. 2008;21:1726–32.
- De M, Ghosh PS, Rotello VM. Applications of Nanoparticles in Biology. *Adv Mater*. 2008;20:4225–41.
- Lebre F, Chatterjee N, Costa S, Fernández-De-gortari E, Lopes C, Meneses J et al. Nanosafety: An Evolving Concept to Bring the Safest Possible Nanomaterials to Society and Environment. *J Nanomater*. 2022;12.
- Yin N, Liu Q, Liu J, He B, Cui L, Li Z, et al. Silver nanoparticle exposure attenuates the viability of Rat Cerebellum Granule cells through apoptosis coupled to oxidative stress. *Small*. 2013;9:1831–41.
- Wilson MR, Lightbody JH, Donaldson K, Sales J, Stone V. Interactions between ultrafine particles and transition metals in vivo and in vitro. *Toxicol Appl Pharmacol*. 2002;184:172–9.

7. Garcés M, Magnani ND, Pecorelli A, Calabró V, Marchini T, Cáceres L, et al. Alterations in oxygen metabolism are associated to lung toxicity triggered by silver nanoparticles exposure. *Free Radic Biol Med*. 2021;166:324–36.
8. Brown DM, Wilson MR, MacNee W, Stone V, Donaldson K. Size-dependent Proinflammatory Effects of Ultrafine polystyrene particles: a role for Surface Area and oxidative stress in the enhanced activity of Ultrafines. *Toxicol Appl Pharmacol*. 2001;175:191–9.
9. Iavicoli I, Fontana L, Pingue P, Todea AM, Asbach C. Assessment of occupational exposure to engineered nanomaterials in research laboratories using personal monitors. *Sci Total Environ*. 2018;627:689–702.
10. Hischer R, Walsler T. Life cycle assessment of engineered nanomaterials: state of the art and strategies to overcome existing gaps. *Sci Total Environ*. 2012;425:271–82.
11. Kuhlbusch TA, Asbach C, Fissan H, Göhler D, Stintz M. Nanoparticle exposure at nanotechnology workplaces: a review. *Part Fibre Toxicol*. 2011;8:22.
12. Hodson L, Methner M, Zumwalde RD. Approaches to safe nanotechnology; managing the health and safety concerns associated with engineered nanomaterials. 2009.
13. Sahu M, Biswas P. Size distributions of aerosols in an indoor environment with engineered nanoparticle synthesis reactors operating under different scenarios. *J Nanopart Res*. 2010;12:1055–64.
14. Council NR. Risk assessment in the Federal Government: managing the process. 1st ed. Washington, DC: The National Academies Press; 1983.
15. Savolainen K, Alenius H, Norppa H, Pylkkänen L, Tuomi T, Kasper G. Risk assessment of engineered nanomaterials and nanotechnologies-A review. *Toxicology*. 2010;269:92–104.
16. von Ranke NL, Geraldo RB, Lima dos Santos A, Evangelho VGO, Flammini F, Cabral LM et al. Applying *in silico* approaches to nanotoxicology: Current status and future potential. *Comput Toxicol*. 2022;22.
17. Environmental Protection Agency E. ToxCast Owner's Manual-Guidance for Exploring Data. 2018. <https://www.epa.gov/sites/default/files/2018-04/documents/toxcastownermanual4252018.pdf> Accessed on 6 Nov 2022.
18. EU-ToxRisk-About. EU-ToxRisk. <https://www.eu-toxrisk.eu/page/en/about-eu-toxrisk.php>. Accessed 6 Nov 2022.
19. International Organization for Standardization I. ISO - ISO/TC 229 - Nanotechnologies. 2011. <https://www.iso.org/committee/381983.html>. Accessed 6 Nov 2022.
20. Organisation for Economic Co-operation and Development O. AOP knowledge base 2021. <https://aopkb.oecd.org/>. Accessed 6 Nov 2022.
21. Burden N, Aschberger K, Chaudhry Q, Clift MJD, Doak SH, Fowler P, et al. The 3Rs as a framework to support a 21st century approach for nanosafety assessment. *Nano Today*. 2017;12:10–3.
22. Ram RN, Gadaleta D, Allen TEH. The role of 'big data' and 'in silico' New Approach Methodologies (NAMs) in ending animal use – A commentary on progress. *Comput Toxicol*. 2022;23.
23. Schwarz-Plaschg C, Kallhoff A, Eisenberger I. Making Nanomaterials Safer by Design? *NanoEthics*. 2017;1:277–81.
24. Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari TP, Michalkova A, et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat Nanotechnol*. 2011;6:175–8.
25. González-Durruthy M, Werhli AV, Seus V, Machado KS, Pazos A, Munteanu CR et al. Decrypting Strong and Weak Single-Walled Carbon Nanotubes Interactions with Mitochondrial Voltage-Dependent Anion Channels Using Molecular Docking and Perturbation Theory. *Sci Rep*. 2017;7.
26. Toropova AP, Toropov AA. Nanomaterials. Quasi-SMILES as a flexible basis for regulation and environmental risk assessment. *Sci Total Environ*. 2022;823.
27. Gajewicz A, Puzyn T. Computational, Nanotoxicology. Challenges and Perspectives. CRC Press; 2019.
28. Kar S, Gajewicz A, Puzyn T, Roy K, Leszczynski J. Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: a mechanistic QSTR approach. *Ecotoxicol Environ Saf*. 2014;107:162–9.
29. Furchi I, Murphy F, Mullins M, Arvanitis A, Poland CA. Practices and trends of machine learning application in nanotoxicology. *J Nanomater*. 2020;10.
30. Singh AV, Ansari MHD, Rosenkranz D, Maharjan RS, Kriegl FL, Gandhi K, et al. Artificial Intelligence and Machine Learning in Computational Nanotoxicology: Unlocking and Empowering Nanomedicine. *Adv Healthc Mater*. 2020;9:1901862.
31. Roy K. Advances in QSAR modeling. Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences. Volume 555, 1st ed. Cham, Switzerland: Springer; 2017. p. 39.
32. Shin HK, Kim S, Yoon S. Use of size-dependent electron configuration fingerprint to develop general prediction models for nanomaterials. *NanoImpact*. 2021;21.
33. Toropova AP, Meneses J, Alfaro-Moreno E, Toropov AA. The system of self-consistent models based on quasi-SMILES as a tool to predict the potential of Nano-inhibitors of human lung carcinoma cell line A549 for different experimental conditions. *Drug Chem Toxicol*. 2023;1–8.
34. Ambure P, Balasaheb Aher R, Puzyn T, Roy K. NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling. *Chemometr Intell Lab Syst*. 2015;147:1–13.
35. Brownlee J. Data Preparation for Machine learning: data cleaning, feature selection, and Data Transforms in Python. 1st ed. Machine Learning Mastery; 2020.
36. Yeo I-k, Johnson RA. A new family of power transformations to improve normality or symmetry. *Biometrika*. 2000;87:954–9.
37. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci*; 2007.
38. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. *Neurocomputing*. 2018;300:70–9.
39. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30.
40. Directorate OE, ENV/JM/MONO. (2007)2 2 OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69 GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP [(Q)SAR] MODELS. 2007.
41. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data Mining, Inference, and Prediction. 2nd ed. Springer: New York; 2009.
42. Roy K, Das RN, Ambure P, Aher RB. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemometr Intell Lab Syst*. 2016;152:18–33.
43. Roy K, Kar S, Ambure P. On a simple approach for determining applicability domain of QSAR models. *Chemometr Intell Lab Syst*. 2015;145:22–9.
44. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell*. 2016;5:221–32.
45. Linardatos P, Papastefanopoulos V, Kotsiantis S, Explainable AI. A review of machine learning interpretability methods. *Entropy*. 2020;23:18.
46. Fjodorova N, Novič M. Integration of QSAR and SAR methods for the mechanistic interpretation of predictive models for carcinogenicity. *Comput Struct Biotechnol J*. 2012;1:e201207003.
47. De P, Kar S, Ambure P, Roy K. Prediction reliability of QSAR models: an overview of various validation tools. *Arch Toxicol*. 2022;96:1279–95.
48. Roy J, Roy K. Risk assessment and data gap filling of toxicity of metal oxide nanoparticles (MeOx NPs) used in nanomedicines: a mechanistic QSAR approach. *Environ Sci Nano*. 2022;9:3456–70.
49. Subramanian NA, Palaniappan A, NanoTox. Development of a parsimonious *In Silico* Model for Toxicity Assessment of Metal-Oxide Nanoparticles using Physicochemical features. *ACS Omega*. 2021;6:11729–39.
50. Thwala MM, Afantitis A, Papadiamantis AG, Tsoumanis A, Melagraki G, Dlamini LN, et al. Using the Isalos platform to develop a (Q)SAR model that predicts metal oxide toxicity utilizing facet-based electronic, image analysis-based, and periodic table derived properties as descriptors. *Struct Chem*. 2022;33:527–38.
51. Hahn A, Fuhrrott J, Loos A, Barcikowski S. Cytotoxicity and ion release of alloy nanoparticles. *J Nanopart Res*. 2012;14.
52. Wang Y, Cai R, Chen C. The Nano-Bio interactions of nanomedicines: understanding the biochemical driving forces and redox reactions. *Acc Chem Res*. 2019;52:1507–18.
53. Wang X, Cui X, Zhao Y, Chen C. Nano-bio interactions: the implication of size-dependent biological effects of nanomaterials. *Sci China Life Sci*. 2020;63:1168–82.
54. Batool F, Iqbal MS, Khan SUD, Khan J, Ahmed B, Qadir MI. Biologically synthesized iron nanoparticles (FeNPs) from *Phoenix dactylifera* have anti-bacterial activities. *Sci Rep*. 2021;11.
55. González-Durruthy M, Castro M, Nunes SM, Ventura-Lima J, Alberici LC, Naal Z, et al. QSPR/QSAR-based Perturbation Theory approach and mechanistic electrochemical assays on carbon nanotubes with optimal properties against mitochondrial Fenton reaction experimentally induced by Fe²⁺-overload. *Carbon*. 2017;115:312–30.
56. Toyokuni S. Iron-induced carcinogenesis: the role of redox regulation. *Free Radic Biol Med*. 1996;20:553–66.
57. Toyokuni S. Iron as a target of chemoprevention for longevity in humans. *Free Radic Res*. 2011;45:906–17.

58. Kuban-Jankowska A, Gorska M, Jaremko L, Jaremko M, Tuszyński JA, Wozniak M. The physiological concentration of ferrous iron (II) alters the inhibitory effect of hydrogen peroxide on CD45, LAR and PTP1B phosphatases. *Biometals*. 2015;28:975–86.
59. Uchiyama A, Kim JS, Kon K, Jaeschke H, Ikejima K, Watanabe S, et al. Translocation of iron from lysosomes into mitochondria is a key event during oxidative stress-induced hepatocellular injury. *J Hepatol*. 2008;48:1644–54.
60. Srivastava S, Kumar A. Comparative cytotoxicity of nanoparticles and ions to *Escherichia coli* in binary mixtures. *Res J Environ Sci*. 2017;55:11–9.
61. Sengul AB, Asmatulu E. Toxicity of metal and metal oxide nanoparticles: a review. *Environ Chem Lett*. 2020;18:1659–83.
62. Gliga AR, Skoglund S, Odnevall Wallinder I, Fadeel B, Karlsson HL. Size-dependent cytotoxicity of silver nanoparticles in human lung cells: The role of cellular uptake, agglomeration and Ag release. *Part Fibre Toxicol*. 2014;11.
63. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018.
64. DeLoid GM, Cohen JM, Pyrgiotakis G, Pirela SV, Pal A, Liu J, et al. Advanced computational modeling for in vitro nanomaterial dosimetry. *Part Fibre Toxicol*. 2015;12(1):1–20.
65. Kowoll T, Fritsch-Decker S, Diabaté S, Nienhaus GU, Gerthsen D, Weiss C. Assessment of in vitro particle dosimetry models at the single cell and particle level by scanning electron microscopy. *J Nanobiotechnol*. 2018;16:1–5.
66. Cohen JM, Teeguarden JG, Demokritou P. An integrated approach for the in vitro dosimetry of engineered nanomaterials. *Part Fibre Toxicol*. 2014;11:1–12.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.