Linear and non-linear modelling of the cytotoxicity of TiO\textsubscript{2} and ZnO nanoparticles by empirical descriptors

E. Papa\textsuperscript{a,b,}\textsuperscript{*}, J.P. Doucet\textsuperscript{b} and A. Doucet-Panayeb

\textsuperscript{a}QSAR Research Unit in Environmental Chemistry and Ecotoxicology, University of Insubria, Varese, Italy; \textsuperscript{b}Laboratoire ITODYS, UMR 7086, Université Paris Diderot, Paris, France

(Received 27 June 2015; in final form 1 August 2015)

Titanium oxide (TiO\textsubscript{2}) and zinc oxide (ZnO) nanoparticles are among the most widely used in different applications in daily life. In this study, local regression and classification models were developed for a set of ZnO and TiO\textsubscript{2} nanoparticles tested at different concentrations for their ability to disrupt the lipid membrane in cells. Different regression techniques were applied and compared by checking the robustness of the models and their external predictive ability. Additionally, a simple classification model was developed, which predicts the potential for disruption of the studied nanoparticles with good accuracy (overall accuracy, specificity, and sensitivity >80%) on the basis of two empirical descriptors. The present study demonstrates that empirical descriptors, such as experimentally determined size and tested concentrations, are relevant to modelling the activity of nanoparticles. This information may be useful to screen the potential for harmful effect of nanoparticles in different experimental conditions and to optimize the design of toxicological tests. Results from the present study are useful to support and refine the future application of in silico tools to nanoparticles, for research and regulatory purposes.

Keywords: QSAR; TiO\textsubscript{2} nanoparticles; ZnO nanoparticles; in silico models; membrane disruption; cytotoxicity

1. Introduction

The use of nanomaterials for industrial and commercial applications has become a reason for concern in relation to their possible effects on humans and the environment. In addition to possible exposure in working places, intentional and unintentional release of such materials into the environment can generate different exposure scenarios, which may be a cause of risk for humans and other living organisms [1,2]. Experimental evidence, reporting in vivo in different organisms, and in vitro in different cellular systems, links different toxicological effects of various extent to exposure to different nanomaterials. This evidence highlights the need for a more clear comprehension of the toxicological and eco-toxicological behaviour of such materials, as well as for a clear regulation of their marketing and use [1,3–7]. From a regulatory point of view, nanomaterials may fall within the REACH regulation obligations, and the consequent need to provide consistent data for the evaluation of their safety [7,8]. However, as for the many other substances that need to comply with REACH or other regulations (e.g. the cosmetic directive [9]) the lack of experimental information related mainly to toxicological and
eco-toxicological information calls for the application of alternative methods to animal testing, such as in silico models based on quantitative structure-activity relationships [7,9]. Despite an increase of attention to this field, with the increasing availability of tools useful for the screening and the prediction of properties and activities of 'conventional' substances, the development of QSAR-like models for nanoparticles (NPs) is still in progress, and facing many difficulties [10,11]. In fact, despite the efforts currently made by theoretical and experimental scientists, different issues limit the further development of QSAR models for NPs [10–15], such as scattered, and often insufficient, information regarding the characterization of chemico-physical, toxicological, and fate-related properties, the lack of harmonized testing strategies, and the limited availability of suitable theoretical descriptors. Furthermore, the currently available experimental data are generally sufficient to generate very simple QSARs, with limited possibilities for performing external validation, and small applicability domains.

Titanium oxide (TiO₂) and zinc oxide (ZnO) nanoparticles are among the most widely studied due to their large utilization in different commercial applications, and for their

Figure 1. Plot of experimental lactate dehydrogenase release (LDH) values vs. predicted values calculated by Equation 1. Black circles = training. Descriptors: X0, engineered size; X2, size in PBS; X4, concentrations of particle suspensions in ultrapure water.
association to potential toxicological and eco-toxicological effects [15–21]. Recent studies report effects of TiO$_2$ and ZnO NPs in different in vitro and in vivo systems [15–20]. Moreover, additional evidence of toxicity through the respiratory system recently supported the opinion of the Scientific Commission on Consumer Safety (SCCS) of the European Commission not to recommend the use of TiO$_2$ and ZnO nanoparticles in spray products that could expose consumers by inhalation [21]. All of these studies highlight the need for a better evaluation of the potential toxicity of TiO$_2$ and ZnO NPs and, as mentioned above, the QSAR approach could be used to generate quantitative models that are possibly applicable for calculating reliable predictions. A recent study by Sayes and Ivanov [15] investigated the potential of ZnO and TiO$_2$ NPs to induce the release of lactate dehydrogenase (LDH) in rat lung cells. The induction of LDH release has been used to evaluate the level of membrane damage after exposure to NPs in different test systems [15,16,19,22]. In addition to the toxicological response, Sayes and Ivanov’s study [15] also provided values for five descriptors related to the experimental conditions (i.e. concentration, engineered size, size in water, size in phosphate buffered saline (PBS), and zeta potential) to be used for QSAR-like purposes. Results reported by Sayes and Ivanov for the development of predictive models were satisfactory in classification, but no successful examples of linear regression models were described. Particularly, Sayes and Ivanov’s study did not quantify the robustness of the models or their potential predictivity (absence of external validation), and the evaluation of the applicability domain was not performed.

However, a more recent publication by Toropova et al. [23], successfully addressed the linear regression of the LDH induction performed for ZnO and TiO$_2$ NPs by using the Correlation And Logic (CORAL) approach, i.e. on the basis of numerical descriptors (optimal descriptors) derived from a weighted codification of the simplified molecular-input line-entry system (SMILES) code of a chemical structure. Even though the CORAL approach provided good results, the model generation and interpretation was not straightforward, and the application of the model for prediction could only be performed through the use of CORAL software.

In this paper we address the modelling of membrane disruption mediated by ZnO and TiO$_2$ NPs by using the original experimental descriptors published by Sayes and Ivanov [15] and comparing different linear and non-linear techniques [24–27]. Unlike the existing approaches, we studied TiO$_2$ and ZnO NPs both separately and as a unique dataset in order to propose models based on an ‘as large as possible’ structural domain, and checked by robust validations [24]. The aim of the study was to provide tools that were easily applicable and interpretable, such as linear models based on multiple linear regression (MLR) [24,25], or classification trees [28]. Additionally, we performed a comparative analysis to examine if (non-linear) machine learning approaches would be useful to complement or replace MLR models in the studied scenario. Indeed, non-linear methods have been increasingly used in SAR and QSAR modelling studies in pharmacology and toxicology, where their flexibility to adapt to experimental data and their capability to cope with complex mechanisms are crucial [26,27,29,30]. So far, machine learning approaches have not been extensively used in NPs modelling [13,31–35], where most of the scarce correlations available rely on MLR [13,34], probably because they are generally applied to deal with larger datasets. However, successful applications of machine learning approaches, also to small sized datasets of NPs, are reported in the literature [13,31–35].

The generation of QSAR-like models based on descriptors related to the experimental conditions (e.g. size in different media, concentrations, etc.), such as those generated in this study, represents a new way to maximize the information content of the available experimental
data, and has been reported in the literature [35,36]. Results from these approaches are mechanistically relevant to highlight the experimental factors with higher influence on the tested activity. This gives the possibility to investigate possible scenarios of toxicity of a specific nano-form, by simulation of the variation of the most influential experimental parameters selected in a QSAR-like model, without the need to perform additional experimental measures.

In the present study, we intentionally compared different possible linear and non-linear approaches in order to identify which statistical methods might be more suitable to model the studied NPs, and to address some ‘classic’ problems arising when dealing with small datasets (such as lack of stability and robustness of the models, presence of collinearity, and instability of external predictions).

Finally, it is worth mentioning that QSAR-like approaches fall within the purposes of the REACH regulation [7], which supports the use of in silico methodologies to reduce experimental costs and the number of animals tested. This paper is not specifically aimed at proposing models for application within REACH, due to the small size of the available experimental data and the consequent limited generalizability of the models. However, we believe that our results will be useful to support and refine the future application of in silico tools for NPs for research and regulatory purposes.

2. Methods

2.1 Experimental data set and descriptors

Experimental data for the cytotoxic response ‘lactate dehydrogenase release (LDH)’, i.e. membrane damage, were homogeneously measured in vitro by Sayes and Ivanov [15]. Data were measured in co-cultures of immortalized rat L2 lung epithelial cells and rat lung alveolar macrophages. Data were measured for differently sized ZnO and TiO₂ NPs for a total of 42 different nano-forms (24 nano-forms of TiO₂ and 18 of ZnO). The LDH data were normalized to unexposed control cell population, and expressed in units/L. Additionally, to facilitate the use of these experimental data to build regression as well as classification models, continuous LDH release values were associated by Sayes and Ivanov to different levels of membrane damage from dense to disrupted according to the following cut-off values: LDH < 0.99 = dense; 0.99 < LDH < 1.09 = normal; 1.09 < LDH < 1.25 = leaky; LDH > 1.25 = disrupted (cell death). For simplification in this paper we only used one cut-off value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs.

Values for five different experimental descriptors were quantified for all of the 42 nano-forms (i.e. X0: engineered size (nm); X1: size in water (ultrapure water stock suspensions, nm); X2: size in phosphate buffered saline (PBS solution, sterile filtered, nm); X4: concentrations of particle suspensions in ultrapure water (mg/L); X5: zeta potential (mV)) were used to generate classification and regression models. Experimental LDH and descriptors values are available in the original literature and as a Supplemental file (Table S1). These data were analysed and commented upon in Section 3.1.1 Data analysis and models for TiO₂ and ZnO nanoparticles and in Supplemental material (Figures S1, S2). Additionally, the variable Agglomeration Rate was calculated as the ratio among size in water (nm) and engineered size (nm). The additional analysis performed for this variable was reported in Supplemental material, Figure S3 and related text.
2.2 Modelling and validation

Regression and classification models were generated to model the toxicity of TiO$_2$ NPs and ZnO NPs separately and in combination, by using different linear and non-linear methods. Particular attention was dedicated to the validation of the robustness and of the predictivity of the models [24,37], which were performed compatibly with the available experimental data [15].

2.2.1 Multiple linear regression models

The best modelling variables were selected by exploring the statistical quality of all the possible combinations of the available experimental descriptors, by applying multiple linear regression based on ordinary least squares (MLR-OLS) in the QSARINS software [25]. This ‘variable selection’ procedure generates a ‘population’ of models, ranked according to decreasing $r^2$ values. The best models were chosen by using $Q^2$ leave-one-out ($Q^2_{\text{loo}}$) as the optimization value, and taking into account the parsimony principle regarding the complexity of the models, which should be as small as possible. For this reason, only up to three descriptors were included in the QSARs generated in this study. Furthermore, the correlation between the modelling descriptors and the modelled response was checked by the QUIK rule (Q under influence of K), to exclude models with high predictor co-linearity and exclude chance correlation [37]. Additionally, Y-scrambling was also applied to verify that the models are not based on a chance correlation of descriptors with the response. Low $r^2_{\text{SY}}$ values of the models, which were calculated on scrambled responses, confirmed the absence of chance correlation in the original model. Moreover, the robustness of the models was evaluated by applying the leave-many-out (30%) procedure $Q^2_{\text{lmo}}$ (2000 iterations).

Standardized residuals were calculated to identify outliers for the response (chemicals with cross-validated standardized residuals greater than 2.5 standard deviation units).

The quantification of the external predictivity, i.e. the ability of the model to provide correct prediction of the activity of NPs not used to train the model (prediction set), was performed by comparing different parameters, such as different measures of external $Q^2$ (i.e. $Q^2_{\text{ext}}$), and the concordance correlation coefficient [24,38,39]. In addition, the root mean square error (RMSE), was used to measure and compare prediction accuracy in the training (RMSE$_{\text{tr}}$) and in the prediction (RMSE$_{\text{PRED}}$) sets.

The external predictivity of models was evaluated on multiple prediction sets (10 for each data set) manually generated by unbiased random splitting (without taking into consideration response or descriptor distributions), and leaving out about 20% of the original data. An additional splitting was performed on the largest training set available (i.e. including both TiO$_2$ and ZnO NPs) in order to test the external predictivity of the best models by leaving out 50% of the training set compounds. The variable selection procedure was performed independently for each training set, and the resulting population of models was tested on the respective prediction set (external validation). Columns of the random splitting used for each dataset are reported in Supplemental material Tables S2–S4.

Finally, the equation for the best model among the split models generated for each training set was newly calibrated on all of the available data for each dataset (full models).

2.2.2 Non-linear regression models

Three approaches were used to develop nonlinear models: support vector machine (SVM), radial basis function neural networks (RBFNN), and general regression neural networks
(GRNN). They present the advantages of easy setting, rapid training, and a guarantee of finding the global minimum on the error surface. We indicate here just the basic principle of these approaches. More information may be found in the literature [26–32,40–47].

SVM [26,27,29,31,41] is rooted in two key ideas: the first is to privilege robustness of the model over the search for an optimal recall of the data in order to get more generalization ability. The second one is to project data, by a kernel function, in a higher dimensional space where it may be expected that a linear representation would work better than in the initial descriptor space. The trick is that, with a large variety of kernel functions, calculations can still be made in the original descriptor space, considerably limiting the complexity of the mathematical treatment. In this study we used the two more generally used kernel functions (linear and Gaussian).

RBFNN is formed from three levels. For each pattern, the first level (input) is fed with the associated structural descriptors. This level does not process the data but only transmits information to the second layer. This hidden layer consists of a limited number of nodes (centres) acting as locally tuned processors. Weights of these centres are computed from the distance between the investigated input pattern and centre descriptors, weighted by a non-linear transfer function (generally a Gaussian). Bias may be added. Once these weights are set, the output value (third layer) is calculated as a linear combination of the weighted hidden centres [43].

Optimization of a RBFNN involves the location of the hidden centres and the determination of the common width of the Gaussian function. Orr’s forward subset selection method [44,45] (which was used in this study) simultaneously selects the number and the hidden centres chosen among the data samples. The weight adjustments between hidden and output layers are performed by a least squares method, the best width of the Gaussian transfer function being determined from a number of trials looking at the minimum error on a learning set in leave-one-out cross-validation.

GRNN are non-parametric estimators that calculate a weighted average of the target value for the training patterns. This is determined by a probability density function using Parzen’s non-parametric estimator [30,42]. Each training pattern is weighted exponentially according to its distance to the unknown pattern under investigation. The common width of this Gaussian weighting function is determined for the training set in leave-one-out cross-validation.

RBFNN and GRNN rely on proximity between samples, Euclidian distances being calculated on descriptor values. These processes bear some analogy with the \( k \)-nearest neighbour method. However in \( k \)-nn the (active) neighbours change depending on every examined pattern, whereas in RBFNN, for a given dataset, fixed centres (limited in number) are selected, and in GRNN all samples are considered as centres.

It seems important to stress here that in this study, for a given data set, the approaches we used are strictly reproducible, and the settings that were applied to ensure reproducibility were specified.

It should be observed that the number and location of the centres in RBFNN were automatically determined by MATLAB routines. The common width of the Gaussian function was therefore the only setting to be fixed, which was obtained directly and unambiguously by performing leave-one-out on the training set.

Similarly, GRNN also needed the setting of a single parameter, which was the common width of the weighting Gaussian, derived by leave-one-out training assay.

The unique parameter to be fixed in the SVM approach based on linear or Gaussian kernels was the regularization constant (C); the width (sigma in the usual terminology) of the Gaussian kernel was automatically fixed by the software, using Caputo’s method. The ‘best’
C value was determined on a scale of pre-defined C values (10 values were used in this study), looking at the best mean results calculated on the basis of various possible options for the training sets (x-fold cross-validation, possibly repeated, leave-one-out, etc.). Here we chose a five-fold cross-validation repeated three times on the training set. Furthermore, although of marginal importance on the final results, the ‘seed’ value used in this study, which is required to start the initial partitions and ensure perfect reproducibility, was set at 2.

The best combinations of descriptors identified in this study by MLR-OLS approach were used as the input for the development of SVM, RBFNN, and GRNN models. SVM calculations were performed using the Caret package [46] of the Cran-R software [47]; MATLAB routines were used to apply RBFNN [45] and GRNN [27] approaches. The quality of the models was evaluated by quantification of internal fitting ($r^2$) and predictivity ($Q^2_{\text{loo}}$). The external predictivity of the models was evaluated on the same external prediction sets used for MLR models described in Section 2.3.1 and quantified by $r^2$. Additionally, the root mean squared error (RMSE) was used to measure and compare prediction accuracy in the training (RMSE$_{\text{tr}}$) and prediction (RMSE$_{\text{pred}}$) sets.

All of these parameters, obtained in comparable conditions, were used for comparing the quality of the predictions using linear or non-linear models.

2.2.3 Classification models

_A priori_ categorization and the following classification were performed based on the experimental cytotoxic activity of the studied NPs. As described above, the LDH value 1.09 was used as the cut-off to distinguish among active and inactive NPs, able to disrupt (Active) or to not affect (Inactive) the cellular membrane, respectively. A linear classification model was generated on the _a priori_ classes by applying the J48 method in WEKA software [48], after random splitting of the complete dataset set composed of 42 compounds, into a training set and a prediction set (30% in the prediction set). The best combinations of variables selected in regression were manually verified in WEKA and the final model was chosen as the best combination of the parameters Accuracy%, Sensitivity% (Sn), and Specificity% (Sp), which were calculated with the following formulas [37]:

$$\text{Accuracy\%} = \frac{TP + TN}{TP + TN + FN + FP} \times 100$$

$$\text{Sn} = \frac{TP}{TP + FN} \times 100$$

$$\text{Sp} = \frac{TN}{TN + FP} \times 100$$

where TP (true positive = Active) and TN (true negative = Inactive) are the number of compounds correctly classified as active and inactive, respectively. FP are negatives predicted as positives, and FN are positives predicted as negatives.

2.3 Applicability domain

The structural applicability domain (AD) of MLR models was quantified by the leverage approach [25,37] in order to verify the presence of influential objects (i.e. NPs) in the training set, and to verify the reliability of predictions for objects not included in the training set.
(i.e. reliable predictions should fall within the AD of a model). The leverage matrix \( H \) is calculated from the \( X \) matrix, which includes \( n \) training set samples and \( p \) modelling descriptors [24,37]. Diagonal elements of the \( H \) matrix quantify the leverage of each object \((h)\) in the space of the model, i.e. the influence of each object on the regression results. The limit of a model domain is quantified by the value \( h^* \), which is calculated as \( 3(p+1)/n \), where \( p = \) number of variables in the model and \( n = \) number of compounds in the training set. Leverage values greater than \( h^* \) characterize structurally influential compounds, which ‘influence’ the mathematical structure of the model, and fall outside of the model’s structural AD. Predictions calculated for high leverage chemicals in the prediction set should be considered as less reliable (i.e. extrapolated values) [24].

The applicability domain of MLR models was graphically inspected by using the Williams plot. The Williams graph is the plot of hat diagonal values vs. standardized residuals and gives an immediate view of NPs falling within the structural AD of the models (i.e. \( h < h^* \)), and of response outliers that are characterized by standardized residuals larger than 2.5. The applicability domain of the J48 classification tree and of non-linear models was checked by principal component analysis (PCA). PCA was calculated from the correlation matrix of the descriptors included in the QSAR models by SCAN software [49,50].

3. Results and discussion

3.1 Linear regression models

3.1.1 Data analysis and models for TiO\(_2\) and ZnO nanoparticles

Different regression models were developed based on both TiO\(_2\) and ZnO NPs combined in a single dataset or in two separated datasets. The best models developed on the combined dataset were generated after exclusion of two TiO\(_2\) NPs and the largest ZnO NPs. In fact, a first attempt to model all the NPs together by using all the possible combinations of the five available variables led to poor models (maximum \( r^2 = 0.4 \)) that apparently confirmed the observations reported by Sayes and Ivanov [15], who were unable to provide satisfactory results by treating the studied dataset by regression. The first element that we considered as the potential cause of the poor performances of the models was the range of size of the studied NPs. We observed that most of the experiments were conducted for NPs with an engineered size between 30 nm and 70 nm, but nine experiments were conducted for NPs larger than 1000 nm (i.e. ZnO NPs). The PCA analysis [49,50] performed on the available descriptors measured for the 42 NPs showed that large ZnO NPs were isolated from the other NPs (Figure S1, Supplemental material). Furthermore, as shown in Figure S2 (Supplemental material), we explored the behaviour of the studied data by comparing results tested at different concentrations (i.e. 25 mg/L, 50 mg/L, and 100 mg/L). We identified two data points (i.e. TiO\(_2\) 45 nm, 100 mg/L, ID 13 and 14 respectively) with toxicity values that were anomalous when considering the global trend of the three series of data. We also observed an anomalous trend of toxicity values for ZnO NPs sized >1000 nm tested at 50 mg/L (inverse trend of cytotoxicity, compared to values measured at 25 mg/L and 100 mg/L). Finally, we performed an additional analysis of the structure-activity relationship among agglomeration rate (AgR) and the response of toxicity LDH (results reported in Supplemental material Figures S3A, S3B and related text). This analysis highlighted an anomaly in the aggregation rate of NPs 13 and 14 compared to the other NPs; however, no additional experimental details were available on factors that may be responsible for changes in aggregation (such as pH) to justify the large AgR values calculated for these two NPs. In general, we observed that AgR played a clear...
Table 1. Performances of the MLR models developed for the complete dataset (31 nanoparticles).

| Model | \( n \) | \( r^2 \) | \( Q^2_{\text{loo}} \) | RMSE\(_{\text{tr}} \) | CCC\(_{\text{tr}} \)* | \( s \) | \( F \) | RMSE\(_{\text{cv}} \) | CCC\(_{\text{cv}} \)* | \( r^2_{\text{Ycvr}} \) | RMSE\(_{\text{ext}} \) | \( r^2_{\text{ext}} \) | \( Q^2_{\text{ext}} \) | \( Q^2_{\text{ext}} \) | \( Q^2_{\text{ext}} \) | CCC\(_{\text{ext}} \)* |
|-------|--------|--------|---------------|-------------|----------------|-----|----|-------------|---------------|----------------|-------------|---------------|-------------|----------------|----------------|----------------|----------------|
| Random 1 | 27 | 0.81 | 0.73 | 0.11 | 0.90 | 0.12 | 30.08 | 0.13 | 0.85 | 0.71 | 0.13 | 0.08 | 0.98 | 0.80 | 0.79 | 0.89 |
| Random 2 | 27 | 0.78 | 0.69 | 0.11 | 0.88 | 0.12 | 25.16 | 0.13 | 0.83 | 0.67 | 0.12 | 0.08 | 0.92 | 0.92 | 0.92 | 0.89 | 0.96 |
| Random 3 | 27 | 0.81 | 0.72 | 0.10 | 0.89 | 0.11 | 29.49 | 0.12 | 0.85 | 0.70 | 0.12 | 0.11 | 0.88 | 0.81 | 0.80 | 0.77 | 0.87 |
| Random 4 | 27 | 0.81 | 0.74 | 0.10 | 0.90 | 0.11 | 30.04 | 0.12 | 0.86 | 0.72 | 0.12 | 0.12 | 0.85 | 0.80 | 0.79 | 0.75 | 0.91 |
| Random 5 | 27 | 0.82 | 0.75 | 0.10 | 0.90 | 0.11 | 32.02 | 0.11 | 0.86 | 0.73 | 0.13 | 0.14 | 0.83 | 0.76 | 0.72 | 0.62 | 0.89 |
| Random 6 | 27 | 0.85 | 0.80 | 0.09 | 0.92 | 0.10 | 41.21 | 0.10 | 0.89 | 0.77 | 0.12 | 0.17 | 0.65 | 0.62 | 0.62 | 0.49 | 0.73 |
| Random 7 | 27 | 0.85 | 0.80 | 0.09 | 0.92 | 0.10 | 41.21 | 0.10 | 0.89 | 0.77 | 0.12 | 0.17 | 0.65 | 0.62 | 0.62 | 0.49 | 0.73 |
| Random 8 | 27 | 0.85 | 0.79 | 0.09 | 0.92 | 0.10 | 40.52 | 0.11 | 0.89 | 0.78 | 0.13 | 0.15 | 0.70 | 0.61 | 0.61 | 0.61 | 0.72 |
| Random 9 | 27 | 0.78 | 0.70 | 0.11 | 0.88 | 0.11 | 25.19 | 0.12 | 0.83 | 0.68 | 0.13 | 0.10 | 0.86 | 0.89 | 0.86 | 0.81 | 0.93 |
| Random 10 | 27 | 0.81 | 0.73 | 0.10 | 0.89 | 0.11 | 29.02 | 0.12 | 0.85 | 0.71 | 0.12 | 0.10 | 0.83 | 0.84 | 0.81 | 0.82 | 0.91 |
| Random 11 | 15 | 0.80 | 0.64 | 0.11 | 0.89 | 0.12 | 15.59 | 0.14 | 0.80 | 0.61 | 0.20 | 0.10 | 0.83 | 0.82 | 0.82 | 0.81 | 0.91 |
| Full | 31 | 0.82 | 0.76 | 0.10 | 0.90 | 0.11 | 39.87 | 0.12 | 0.87 | 0.74 | 0.10 | N.A. | N.A. | N.A. | N.A. | N.A. | N.A. |

*For a detailed explanation of these statistical parameters see Chirico and Gramatica [38].
role in defining the ability of the studied NPs to disrupt cellular membranes, since less aggregated and larger NPs within each concentration group had the largest LDH values (Figures S3A).

On the basis of the results reported above, we decided to exclude the largest ZnO NPs and the two TiO₂ data points highlighted above from the studied dataset. Therefore, the final dataset was composed of data for 31 NPs. Ten different random splitting of 20% out, and one 50% out were generated in order to identify the most predictive and stable model among the possible combinations of the available descriptors.

The all-subset variable selection procedure was applied to the 31 TiO₂ and ZnO NPs, and different populations of models were generated for each splitting. The combination engineered size (X₀), size in PBS (X₂), and concentration (X₄) was selected as the best model, according to internal fitting and $Q^2_{\text{ext}}$, in 10 out of 11 populations (including the splitting of 50% out), and as the second best in the remaining one.

![Figure 2](image-url)

Figure 2. Plot of hat diagonal elements of the hat matrix vs. standardized residuals calculated for Equation 1 ($h^\#$ value = 0.387). Black circles = training. Descriptors: X₀, engineered size; X₂, size in PBS; X₄, concentrations of particle suspensions in ultrapure water.
Statistical results calculated for this model in the populations generated for each splitting are reported in Table 1.

High internal fitting, robustness, and external predictivity characterized the models, as demonstrated by the multiple statistical parameters reported in Table 1. It should also be noted that the exclusion of the TiO₂ and ZnO outliers previously identified led to a large increase in the predictive performance of the models compared to the first attempt of modelling commented above (i.e. delta of $Q²_{\text{loo}}$ and $r²$, calculated after and before the cleaning procedure, were 0.52 and 0.41, respectively). Therefore, considering the experimental anomalies highlighted for TiO₂ and ZnO NPs, we suggest the repetition of these experiments to confirm the nature of these outliers.

Since all of the populations of models examined to choose the best combination of variables were evolved separately on the basis of independent training sets, the results reported in Table 1 also demonstrate that the best models performed consistently well in prediction, independently of the distribution of the NPs into training and prediction sets.

The variables concentration (X4) and engineered size (X0) were the most important descriptors according to standardized coefficients, and had a positive sign. Therefore, the cytotoxicity of TiO₂ NPs and of the ZnO NPs of smaller dimensions in the dataset increased for increasing size and concentration. Moreover, the less important variable in the model, i.e. size in PBS (X2), had a negative sign and seemed to be relevant to model the internal trend of toxicity of the small group of ZnO NPs included in the dataset.

Having verified the statistical robustness of the proposed variables the equation of the full model, newly calibrated on all of the available experimental data (to ensure higher statistical robustness and a larger applicability domain), was:

$$
\text{LDH}_{(\text{TiO}_2 + \text{ZnO})} = 0.66 + 0.003 \times X_4 + 0.005 \times X_0 - 4.46E - 5X_2
$$

$$
n = 31, \ r^2 = 0.82, \ Q^2_{\text{loo}} = 0.76, \ Q^2_{\text{lmo30%}} = 0.74, \ r^2_{\text{YS}} = 0.10, \ s = 0.11, \ F = 40
$$

Figures 1 and 2 plot the experimental vs. predicted values of the full model and the standardized residuals vs. the diagonal elements of the hat leverage matrix (Williams plot for the applicability domain), respectively. Figure 2 in particular shows the applicability domain of the model, which did not present any anomalies (all of the NPs fall within the cut-off value $h^* = 0.387$, and no NPs had standardized residuals exceeding 2.5σ).

### 3.1.2 Models for TiO₂ nanoparticles

As anticipated above, we additionally decided to model the two sets of TiO₂ and ZnO NPs separately. Ten different training/prediction sets were identified for the TiO₂ dataset to test the external predictivity of the models, i.e. 20% of the NPs were used as prediction sets. The summary of the performances of the best models derived separately for TiO₂ are reported in Table S5 (Supplemental material).

The best combination of variables selected to model TiO₂ NPs was engineered size (X0) and concentration (X4). This combination was consistent with descriptors selected to model the largest dataset (Table 1), and indicated that the increase of membrane disruption caused by exposure to TiO₂ NPs was concentration-dependent as well as size-dependent.

The statistical parameters calculated for the 10 split models reported in Table S5 showed globally similar performances but a higher instability of the external predictive power compared to the models reported in Table 1 (Table S5: range RMSE<sub>ext</sub> 0.09–0.19; range $r^2_{\text{ext}}$: 0.32–0.98). These results suggested a higher dependency of the performances of models
reported in Table S5 on the distribution of the NPs in the training/prediction sets, which was justified by the smaller size of the TiO₂ training set compared to the complete data set. However, the little differences in RMSE values calculated for the training and the prediction sets (range of RMSEₜ_r: 0.09–0.12; range RMSE_PRED: 0.09–0.19) confirmed the consistency of the results and the validity of the models.

The equation of the model recalibrated on the full TiO₂ dataset (22 NPs) was:

\[
LDH_{(TiO2)} = 0.599 + 0.003X4 + 0.004X0
\]

\[n = 22, \ r^2 = 0.84, \ Q^2_{loo} = 0.79, \ Q^2_{lmo30\%} = 0.78, \ r^2_{YS} = 0.10, \ s = 0.12, \ F = 48\]

As explained above, the new calibration was performed in order to propose a statistically more robust model, based on a larger amount of experimental data, and characterized by a possibly larger AD than split models.

3.1.3 Models for ZnO nanoparticles

The modelling of the ZnO dataset required the inclusion of three variables in order to obtain stable results in fitting and external prediction. In fact, unsatisfactory performances were obtained for models with lower complexity (i.e. developed including up to two variables), which had values of \(Q^2\) up to 0.58 and delta \(r^2 - Q^2\) up to 0.3 (calculated for models with \(r^2 > 0.7\)). Results reported in Table S6 were particularly satisfying considering the structural heterogeneity of the studied ZnO NPs, and the limited amount of data available to build the models (range of \(r^2 = 0.89–0.92\); range of \(Q^2_{loo} = 0.66–0.94\); range of \(r^2_{ext} = 0.45–0.99\)). However, also in this case, the statistical parameters, which were calculated for the 10 split models (Table S6), showed a higher instability of the external predictive power of ZnO NPs models compared to the models developed for the complete dataset (Table 1). Moreover, different from the models reported in Table 1 and from those developed for TiO₂ NPs, the best models developed for ZnO NPs were based on size in water (X1), size in PBS (X2) and concentration (X4), where X4 was the least important variable. In fact, according to the standardized coefficients calculated for the descriptors, size-related descriptors were the dominant factors to describe the cytotoxicity of ZnO NPs, compared to the concentration. However, the two descriptors of size were oppositely correlated with the response, i.e. a large size in water increased membrane disruption, while a large size in PBS decreased this effect (in particular for larger NPs). This inverse correlation between size in PBS and cytotoxicity was observed before, in the model developed for both the studied species of NPs, and associated to the presence of ZnO NPs in the training set. Therefore, the selection of X2 with a negative sign in the model specific for ZnO was consistent with the observations reported above.

Additionally, but of secondary importance according to the standardized residuals, membrane disruption increased at increasing concentrations of the NPs, and the effect was more evident for ZnO NPs of smaller dimensions.

The complexity of the split models reported in Table S6 was further analysed. In these models the ratio of molecules/number of descriptors was slightly below the cut-off value of 5 (Topliss ratio), recommended in the regulations to avoid excess of complexity and overfitting, according to the parsimony principle [37]. However, we would like to stress that for every population of models we verified the absence of overfitting by evaluating whether the inclusion of a new variable would increase the internal and external predictive power of the models. Additionally, the Y-scrambling procedure confirmed the absence of a chance correlation between the modelling descriptors and the response. Furthermore, the Topliss ratio [37] was
respected when the model was newly calibrated on all of the available data (i.e. 15 NPs, 3 descriptors). The equation for the full model was:

$$\text{LDH}_{\text{ZnO}} = 1.041 + 0.001 \times X1 - 0.001 \times X2 + 0.001 \times X4$$

(3)

$$n = 15, \; r^2 = 0.91, \; Q^2_{\text{loo}} = 0.80, \; Q^2_{\text{imo30\%}} = 0.76, \; r^2_{\text{YS}} = 0.22, \; s = 0.08, \; F = 35$$

The analysis of the applicability domains of the models reported in Table S5 and S6 always reported all of the NPs within the respective models’ ADs. Few NPs were occasionally detected as outliers in the prediction sets of the split models (i.e. ZnO 60 nm, 1000 nm and 1500 nm measured at 100 mg/L), while they were always correctly fitted if used as a training set. However, the models were particularly sensitive to the inclusion of ZnO 60 nm and 1500 nm in the prediction set, which caused the worst external performance in split number 4. This fact highlighted the importance of a balanced partitioning of data among training and prediction sets for the external validation of models based on small datasets, to avoid the risk of excluding relevant information for the correct development of the models. No anomalies of the applicability domain were detected in the full models (Figures S4 and S5, Supplemental material).

3.2 Non-linear regression models

In order to explore the possibility of improving the quality of the proposed models, the best modelling variables selected by MLR were used to develop non-linear regression models by using different techniques.

Results from the applications of these additional approaches in comparison with MLR have been reported in Tables S7–S9 (Supplemental material).

A first analysis of the statistical parameters generated for models developed for the complete dataset (TiO$_2$ + ZnO NPs) showed that all the non-linear techniques provided satisfactory results when checked on different prediction sets (Table S7). RMSE$_{\text{tr}}$ values ranged from 0.2–0.12 and were comparable to MLR models (0.09–0.12). However, a more in-depth analysis of the results (Figures S6 A–D, Supplemental material) showed that among the non-linear techniques, SVM linear and RBFNN had a more balanced behaviour considering the statistical parameters calculated for the split and full models compared to the other methods. Figure S6 A–D clearly showed that SVM radial and GRNN had the best fitting performances but provided unstable results when the datasets were perturbed by cross-validation and external validation. SVM linear, RBFNN, and MLR linear performed similarly, based on both the external and the internal validation parameters.

A different situation was reported in Table S8 regarding the models developed for TiO$_2$ NPs. In fact, as also clearly appeared in Figure S7 A–D (Supplemental material), SVM radial outperformed the other methods, and represented a valid alternative to the MLR model.

Finally, considering the models developed for the small ZnO dataset (Table S9 and Figure S8 A–D, Supplemental material), SVM radial and GRNN were those with best performances for fitting and internal robustness, and clearly outperformed the MLR model.

The principal component analysis performed on the modelling variables was used to verify the structural space of the non-linear models (Figures S9–S11, Supplemental material). Results from PCA did not highlight particular anomalies, such as clusters or single points isolated from the NPs in the respective training sets.

The results proposed here show that non-linear models are successful in modelling the datasets studied here, and may be used as valid alternatives to, or in combination with, MLR
models to address and improve the predictive accuracy of individual models. Additionally, the efficiency of non-linear methods on small datasets may be of help to address statistical issues, which may affect MLR models (such as model complexity, correlated variables, etc).

3.3 Classification models

The classifier J48 from the WEKA software was used to generate a simple regression tree model (unpruned, minimum number of objects assigned to each class = 2) for the complete set of NPs \( (n = 42) \). The 30% of the available data was not used to train the model, but only to perform the external validation (28 NPs were included in the training set, and 14 in the prediction set). Different combinations of the available variables were manually, one by one, verified in the WEKA software, and the statistical performances calculated for the best model (percentages of internal and external sensitivity (Sn), specificity (Sp) and overall accuracy (Acc), are reported in Table 2).

The parameters reported in Table 2 showed that the proposed classification tree had high specificity, sensitivity, and global overall accuracy, quantified both in fitting and in cross-validation. Additionally, the model performed very well when tested on the external prediction set, and only one NP (TiO\(_2\) 30 nm (no. 6)) was overestimated from class 2 (inactive) to class 1 (active). Misclassifications in fitting and in cross-validation were mainly associated with TiO\(_2\) species (i.e. no. 5 and no. 20 in fitting, in addition to no. 10 and no. 17 in cross-validation), and to one ZnO NP (i.e. no. 25).

These errors were associated with the borderline nature of the respective NPs, which had experimental measures falling close to the cut-off value 1.09, and, additionally, values of the experimental descriptors similar to NPs of the opposite \( a \text{ priori} \) class (Figure S12, Supplementary material).

\( a \text{ priori} \) and fitted classes are reported in Table S1, together with the splitting column used in classification. Figure 3 shows the J48 model generated for the studied dataset.

This model was composed of three knots and four leaves and easily classified NPs according to their ability to disrupt cellular membrane on the basis of size in water and concentration features. The model showed that high concentrations of the NPs (i.e. >50 mg/L) caused membrane disruption (class 1), while the size in ultrapure water regulated the effect at small concentrations with disruption associated with a larger dimension of the NPs (size in water >129 nm). It was interesting to note that the descriptor size in water was selected in the regression models developed for ZnO NPs, which suggested a higher relevance of this size factor for ZnO compared to TiO\(_2\) NPs.

Table 2. Confusion matrix and performances of the J48 classification tree.

|          | Active | Inactive | Sn  | Sp  | Acc |
|----------|--------|----------|-----|-----|-----|
| Training |        |          |     |     |     |
| Active   | 14     | 2        | 88% | 92% | 89% |
| Inactive | 1      | 11       |     |     |     |
| Cross-validation (10 cross-validation groups) |          |           |     |     |     |
| Active   | 13     | 3        | 81% | 83% | 82% |
| Inactive | 2      | 10       |     |     |     |
| Prediction |        |          |     |     |     |
| Active   | 7      | 0        | 100%| 86% | 93% |
| Inactive | 1      | 6        |     |     |     |
PCA analysis was performed to evaluate the distribution of the studied NPs in the space of the modelling descriptors. Figure S12 clearly showed the discriminant power of the modelling variables (good separation between active (red) and inactive (green) NPs, moving from the right hand side to the left hand side of the PCA plot), as well as the presence of areas of partial overlapping among the two activity classes, where the misclassified NPs were (red circles in Figure S12).

Unfortunately, it was not possible to perform a direct comparison with the classification proposed by Sayes and Ivanov [15], which modelled TiO₂ and ZnO NPs separately, and by using a different classifier (i.e. linear discriminant analysis (LDA)). However, the combination of descriptors, concentration in ultrapure water and size in ultrapure water, was highlighted by Sayes and Ivanov [15] among the most important features to classify between levels of disruption of the cytoplasmatic membrane, which was consistent with the results proposed here. Moreover, unlike the results reported by Sayes and Ivanov [15], the descriptor zeta potential was not relevant for the generation of any of the classification models proposed here.

4. Conclusions

The *in silico* modelling of the toxicological properties of NPs is a challenging topic due to the limited availability of data and descriptors useful for capturing quantitatively relevant structure-activity relationships and to perform proper validation of the mathematical models.

In this study we demonstrated the possibility of generating MLR-OLS models for a dataset that presented issues related to structural heterogeneity, and to the presence of possibly inaccurate experimental values, which were highlighted for further investigation.

Figure 3. J48 classification tree calculated by the WEKA software (class 1 = active; class 2 = inactive). The number of nanoparticles classified at each node/the number of misclassifications (if any) are reported in brackets.
Unlike former work in the literature, we were able to generate simple QSAR-like models based on optimal selections of the original, experimentally determined, descriptors related to NPs’ size and experimental conditions, for three different datasets, thereby addressing the prediction of the cytotoxicity of ZnO and TiO₂ NPs taken singularly or in combination.

Results demonstrated the relevance of the descriptors of size and concentration, for both ZnO and TiO₂ NPs. In particular, the engineered size was relevant for TiO₂ NPs, while the size in water was selected to model ZnO NPs. The statistical quality of the MLR models was thoroughly verified by internal and external validations. Three MLR models were proposed as tools to predict the cytotoxic behaviour of ZnO and TiO₂ NPs. These models will enable the screening (within their applicability domain) of the potential toxicity of virtual experimental scenarios designed for new and existing ZnO and TiO₂ nano-forms, on the basis of up to three experimental parameters.

Additionally, the possibility of improving the fitting and predictive ability of the proposed QSARs was explored by applying non-linear methods. Results generated by non-linear modelling demonstrated the better suitability of SVM linear and RBFNN on the complete dataset, and of SVM radial and GRNN on the smaller datasets. Moreover, we proved the utility of these non-linear techniques as valid alternatives to MLR models generated on small datasets.

Finally, we proposed a J48 classification tree model as a simple and intuitive tool (based on two descriptors) easily applicable to screen the potential for membrane disruption of TiO₂ and ZnO NPs.

We believe that our results are relevant to evaluating the potential cytotoxicity of existing and new NPs, and to optimize the design of new experiments. However, we are aware that the proposed models may have limitations mainly related to the size of the available experimental dataset, and therefore their application is not recommended for NPs different than ZnO and TiO₂, and outside the applicability domain.

We also want to highlight the relevance of size descriptors, highlighted in this study, as an important factor to consider while performing categorization approaches to group NPs, and read across procedures. In fact, the disregard of size factors may cause wrong estimations of the potential activity of differently sized nano-forms of a specific NP, when estimation techniques based on similarity are applied.

Supplemental material
Additional tables and figures are reported as Supplemental material (i.e. Supplemental material_Figures and Supplemental Material_Tables). In particular, Table S1 in Supplemental Material_Tables lists ID numbers, names, values of the experimental descriptors and of the experimental response (LDH), classes of activities (class 1 = active, class 2 = not active), and predictions calculated by Equations 1–3 (MLR models) and by the J48 classification tree for the 42 studied NPs.

Acknowledgement
We would like to thank Professor Paola Gramatica from the University of Insubria, Italy, for providing fruitful comments on the manuscript.

Disclosure statement
No potential conflict of interest was reported by the authors.
References

[1] A. Nel, T. Xia, L. Mädler, and N. Li, Toxic potential of materials at the nanolevel, Science 311 (2006), pp. 622–627.
[2] S.M. Louie, R. Ma, and G.V. Lowry, Transformations of nanomaterials in the environment, Front. Nanosci. 7 (2014), pp. 55–87.
[3] H.A. Jeng and J. Swanson, Toxicity of metal oxide nanoparticles in mammalian cells, J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng. 41 (2006), pp. 2699–2711.
[4] X. Zhu, L. Zhu, Z. Duan, R. Qi, Y. Li, and Y. Lang, Comparative toxicity of several metal oxide nanoparticle aqueous suspensions to Zebrafish (Danio rerio) early developmental stage, J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng. 43 (2008), pp. 278–284.
[5] X. Hu, S. Cook, P. Wang, and H.M. Hwang, In vitro evaluation of cytotoxicity of engineered metal oxide nanoparticles, Sci. Total Environ. 407 (2009), pp. 3070–3072.
[6] M. Horie, H. Kato, K. Fujita, S. Endoh, and H. Iwahashi, In vitro evaluation of cellular response induced by manufactured nanoparticles, Chem. Res. Toxicol. 25 (2011), pp. 605–619.
[7] European Parliament and Council of the European Union, REACH – Registration, Evaluation, Authorization and Restriction of Chemicals, Off. J. Eur. Union L396 (2006), pp. 1–849. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:396:0001:0849:EN:PDF.
[8] S.M Hankin, S.A.K. Peters, C.A. Poland, H.S. Foss, J. Holmqvist, B.L. Ross, J. Varet, and R.J. Aitken, Specific advice on fulfilling information requirements for nanomaterials under REACH (RIP-oN 2) – Final project report, RNC/RIP-oN2/FPR/1/FINAL, 2011. Available at http://nanotechia.org/sites/default/files/files/Published_RIPoN2.pdf.
[9] European Parliament and Council of the European Union, Regulation on cosmetic products, Regulation (EC) No 1223/2009, 2009. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009R1223.
[10] T. Puzyn, D. Leszczynska, and J. Leszczynski, Toward the development of “Nano-QSARs”: Advances and challenges, Small 5 (2009), pp. 2494–2509.
[11] D. Winkler, E. Mombelli, A. Pietroiusti, L. Tran, A. Worth, B. Fadeel, and M.J. McCall, Applying quantitative structure-activity relationship approaches to nanotoxicology: Current status and future potential, Toxicology 313 (2013), pp. 15–23.
[12] L. Lubinski, P. Urbanzak, A. Gajewicz, M.T.D. Cronin, S.J. Enoch, J.C. Madden, D. Leszczynska, J. Leszczynski, and T. Puzyn, Evaluation criteria for the quality of published experimental data on nanomaterials and their usefulness for QSAR modelling, SAR QSAR Environ. Res. 24 (2013), pp. 995–1008.
[13] C. Oksel, C.Y. Ma, and X.Z. Wang, Current situation on the availability of nanostructure–biological activity data, SAR QSAR Environ. Res. 26 (2015), pp. 79–94.
[14] C. Oksel, C.Y. Ma, J.J. Liu, T. Wilkins, and X.Z. Wang, (Q)SAR modelling of nanomaterial toxicity: A critical review, Particulometry (2015), pp. 1–19.
[15] C. Sayes and I. Ivanov, Comparative study of predictive computational models for nanoparticle-induced cytotoxicity, Risk Anal. 30 (2010), pp. 1723–1734.
[16] H. Shi, R. Magaye, V. Castranova, and J. Zhao, Titanium dioxide nanoparticles: A review of current toxicological data, Part. Fibre Toxicol. 10 (2013), p. 15.
[17] A. Menard, D. Drobné, and A. Jemec, Ecotoxicity of nanosized TiO2. Review of in vivo data, Environ. Pollut. 159 (2011), pp. 677–684.
[18] T.G. Smijs and S. Pavel, Titanium dioxide and zinc oxide nanoparticles in sunscreens: Focus on their safety and effectiveness, Nanotechnol. Sci. Appl. 4 (2011), pp. 95–112.
[19] J. Zhang, W. Song, J. Guo, J. Zhang, Z. Sun, F. Ding, and M. Gao, Toxic effect of different ZnO particles on mouse alveolar macrophages, J. Hazard. Mater. 219–220 (2012), pp. 148–155.
[20] H. Ma, P.L. Williams, and S.A. Diamond, Ecotoxicity of manufactured ZnO nanoparticles – A review, Environ. Pollut. 172 (2013), pp. 76–85.
[21] European Commission, Scientific Commission on Consumer Safety, Opinion for clarification of the meaning of the term "sprayable applications/products" for the nano forms of Carbon Black CI
[22] H.A. Jeng and J. Swanson, Toxicity of metal oxide nanoparticles in mammalian cells, J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng. 41 (2006), pp. 2699–2711.

[23] A.P. Toropova, A.A. Toropov, E. Benfenati, T. Puzyn, D. Leszcynska, and J. Leszczynski, Optimal descriptor as a translator of eclectic information into the prediction of membrane damage: The case of a group of ZnO and TiO2 nanoparticles, Ecotoxicol. Environ. Saf. 108 (2014), pp. 203–209.

[24] P. Gramatica, Principles of QSAR models validation: Internal and external, QSAR Comb. Sci. 26 (2007), pp. 694–701.

[25] P. Gramatica, N. Chirico, E. Papa, S. Cassani, and S. Kovarich, QSARINS: A new software for the development, analysis, and validation of QSAR MLR models, J. Comput. Chem. 34 (2013), pp. 2121–2132.

[26] J.P. Doucet and A. Panaye Three-dimensional QSAR. Applications in Pharmacology and Toxicology. CRC Press, BocaRaton, FL, 2010.

[27] A. Panaye, B.T. Fan, J.P. Doucet, X.J. Yao, R.S. Zhang, M.C. Liu, and Z.D. Hu, Quantitative structure-toxicity relationships (QSTRs): A comparative study of various non linear methods. General regression neural network, radial basis function neural network and support vector machine in predicting toxicity of nitro- and cyano- aromatics to Tetrahymena pyriformis, SAR QSAR Environ. Res. 17 (2006), pp. 75–91.

[28] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[29] X.J. Yao, A. Panaye, J.P. Doucet, R.S. Zhang, H.F. Chen, M.C. Liu, Z.D. Hu, and B.T. Fan, Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression, J. Chem. Inf. Comput. Sci. 44 (2004), pp. 1257–1266.

[30] T. Niwa, Using general regression and probabilistic neural networks to predict human intestinal absorption with topological descriptors derived from two-dimensional chemical structures, J. Chem. Inf. Comput. Sci. 43 (2003), pp. 113–119.

[31] D. Fourches, D. Pu, C. Tassa, R. Weissleder, S.Y. Shaw, R.J. Mumper, and A. Tropsha, Quantitative nanostructure – Activity relationship modeling, ACS Nano. 4 (2010), pp. 5703–5712.

[32] D.A. Winkler, F.R. Burden, B. Yan, R. Weissleder, C. Tassa, S. Shaw, and V.C. Epa, Modelling and predicting the biological effects of nanomaterials, SAR QSAR Environ. Res. 25 (2014), pp. 161–172.

[33] Y.T. Chau and C.W. Yap, Quantitative nanostructure–activity relationship modelling of nanoparticles, RSC Advances 2 (2012), pp. 8489–8496.

[34] Q. Mu, G. Jiang, L. Chen, H. Zhou, D. Fourches, A. Tropsha, and B. Yan, Chemical basis of interactions between engineered nanoparticles and biological systems, Chem. Rev. 114 (2014), pp. 7740–7781.

[35] R. Liu, R. Rallo, S. George, Z. Ji, S. Nair, A.E. Nel, and Y. Cohen, Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles, Small. 7 (2011), pp. 1118–1126.

[36] V.V. Kleandrova, F. Luan, H. Gonza, J.M. Ruso, A. Speck-Planche, M. Nata, and D.S. Cordeiro, Computational tool for risk assessment of nanomaterials: Novel QSTR-perturbation model for simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under multiple experimental conditions, Environ. Sci. Technol. 48 (2014), pp. 14686–14694.

[37] Organisation for Economic Co-operation and Development, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models, ENV/JM/MONO(2007)2, 2007. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2.

[38] N. Chirico and P. Gramatica, Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, J. Chem. Inf. Model. 51 (2011), pp. 2320–2335.
[39] N. Chirico and P. Gramatica, Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection, J. Chem. Inf. Model. 52 (2012), pp. 2044–2058.

[40] I.V. Tetko, V.P. Solovev, A.V. Antonov, X. Yao, J.P. Doucet, B. Fan, F. Hoonakker, D. Fourches, P. Jost, N. Lachiche, and A. Varnek, Benchmarking of linear and nonlinear approaches for quantitative structure–property relationship studies of metal complexation with ionophores, J. Chem. Inf. Model 46 (2006), pp. 808–819.

[41] J.P. Doucet, F. Barbault, H. Xia, A. Panaye, and B.T. Fan, Nonlinear SVM approaches to QSPR/QSAR studies and drug design, Curr. Comput.-Aided Drug Des. 3 (2007), pp. 263–289.

[42] D.F. Specht, A general regression neural network, IEEE Trans. Neural Netw. 2 (1991), pp. 568–576.

[43] B. Waleczak and D.L. Massart, Local modelling with radial basis function networks, Chemom. Intell. Lab. 50 (2000), pp. 179–198.

[44] M.J.L. Orr, Introduction to Radial Basis Function Networks, Edinburgh University, UK, Centre for Cognitive Science, 1996.

[45] M.J.L. Orr, MATLAB Routines for Subset Selection and Ridge Regression in Linear Neural Networks, Centre for Cognitive Science, Edinburgh University, UK, 1996.

[46] M. Kuhn, Building predictive models in R using the Caret package, J. Stat. Soft. 28 (2008), pp. 1–26.

[47] R.D.C. Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Software caret available at: http://www.R-project.org.

[48] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, The WEKA data mining software: An update, SIGKDD Exploration 11 (2009), pp. 1–9.

[49] S. Wold, K. Esbensen, and P. Geladi, Principal component analysis, Chemom. Intell. Lab. 2 (1987), pp. 37–52.

[50] SCAN, Software for Chemometric Analysis, rel. 1.1 for Windows, Minitab, USA, 1995.