

# NanoTox: Development of a parsimonious *in silico* model for toxicity assessment of metal-oxide nanoparticles using physicochemical features

Nilesh AnanthaSubramanian<sup>1,2</sup>, Ashok Palaniappan<sup>3\*</sup>

<sup>1</sup>Department of Medical Nanotechnology, School of Chemical and BioTechnology, SASTRA Deemed University, <sup>2</sup>Department of Computer Science and Engineering, IIT Madras, Chennai, <sup>3</sup>Department of Bioinformatics, School of Chemical and BioTechnology, SASTRA Deemed University, Thanjavur 613401. India

\*Corresponding author: [apalania@scbt.sastra.edu](mailto:apalania@scbt.sastra.edu)

## Abstract:

Metal-oxide nanoparticles find widespread applications in mundane life today, and cost-effective evaluation of their cytotoxicity and ecotoxicity is essential for sustainable progress. Machine learning models use existing experimental data, and learn the relationship of various features to nanoparticle cytotoxicity to generate predictive models. In this work, we adopted a principled approach to this problem by formulating a feature space based on intrinsic and extrinsic physico-chemical properties, but exclusive of any *in vitro* characteristics such as cell line, cell type, and assay method. A minimal set of features was developed by applying variance inflation analysis to the correlation structure of the feature space. Using a balanced dataset, a mapping was then obtained from the normalized feature space to the toxicity class using various hyperparameter-tuned machine learning models. Evaluation on an unseen test set yielded > 96% balanced accuracy for both the random forest model, and neural network with one hidden layer model. The obtained cytotoxicity models are parsimonious, with intelligible inputs, and include an applicability check. Interpretability investigations of the models yielded the key predictor variables of metal-oxide nanoparticle cytotoxicity. Our models could be applied on new, untested oxides, using a majority-voting ensemble classifier, NanoTox, that incorporates the neural network, random forest, support vector machine, and logistic regression models. NanoTox is the very first predictive nanotoxicology pipeline made freely available under the GNU General Public License (<https://github.com/NanoTox>).

## INTRODUCTION

Nanotechnology has delivered the promise of ‘plenty of room at the bottom’ with transformative applications for human welfare [1]. The distinctive properties of nanoscale materials have been indispensable in industrial and medical applications, including the delivery of biologically active molecules, and development of biosensors for human health and disease [2]. Engineered metal-oxide nanoparticles are characterized by a concentration of sharp edges, and lend themselves to a variety of uses (for e.g. [3]). However there is a potential caveat to nanobiotechnology: the differential nanoscale behaviour of nanomaterials also obtains emergent toxic side-effects in the biological domain and ecological realm [4-7]. These hazards are related to the capacity of nanomaterials to engender free radicals in the cellular milieu, which inflict damaging oxidative stress. Such events could trigger inflammatory responses, which could balloon out of control leading to apoptosis and cytotoxicity [8-11], and genotoxicity [12].

The mundane use of nanoparticles has necessitated vigorous safety assessment of toxicity, in the interests of sustainable progress [13-16]. Such methods could also help discern safe-by-design principles that could guide adjustments to the nanoparticle formulation, and thereby mitigate adverse effects at the source. Intelligent and alternative testing strategies could accelerate rational design of nanoparticles for optimal functionality and minimal toxicity [17-20]. Various computational methods have been applied to predicting toxicity of engineered nanomaterials [21-30], but with the accumulation of high-quality data, machine learning methods have shown the most promise [31]. Such techniques provide a non-invasive 'instantaneous' readout of nanoparticle toxicity [32-34], and originate from the evolution of QSAR models [35]. Machine learning models of nanoparticle toxicity have tended to be either generalized [36] or tissue-specific [37,38], and are built from experimental toxicity data that have been scored, standardised and curated into databases like the Safe and Sustainable Nanotechnology db (S2NANO) [39-41].

Earlier studies have tended to neglect systematic multicollinearity among the predictor variables, which would lead to confounding and data snooping. Secondly, gross imbalance between the numbers of nontoxic and toxic instances usually exists, which could lead to overfitting to the 'nontoxic' class [42]. Third, we were motivated to develop a model that would be agnostic of *in*

*in vitro* characteristics, such as cell line, cell type, and assay method. A truly general model of nanoparticle cytotoxicity, independent of *in vitro* factors, would lead to significantly broader interpretability and wider applicability [43]. Our study departs also from the notion that tissue-specific models are superior to generalized models [38], and demonstrates that model interpretability is best achieved using a minimal non-redundant feature space, consistent with Occam's parsimony. Further, with a view to increasing reliability, we have deployed the insights from our study into a majority-voting ensemble classifier. Finally, the end-to-end pipeline of our work, including the ensemble classifier, is made freely available as a user-friendly open-source nanosafety prediction system, NanoTox, under GNU GPL (<https://github.com/NanoTox>). All implementations were carried out in R ([www.r-project.org](http://www.r-project.org)).

## METHODS

### Problem and dataset

*In vitro* parameters such as cell-type, cell line, cell origin, cell species, and type of assay, could be extraneous to modelling the intrinsic hazard posed by a nanoparticle to cellular viability and the environment. This motivated us to formulate the problem in a feature space devoid of biological predictors. The machine learning task is stated as: given a certain nanoparticle at a certain dose for a certain duration, would its administration prove cytotoxic? To address this problem, we used a hybrid dataset building on the physico-chemical descriptors and toxicity data found in Choi *et al's* study [36]. All *in vitro* features were removed from the dataset, as noted above. Extrinsic physico-chemical properties, namely dosage and exposure duration were retained [44]. The periodic table properties of metal-oxide nanoparticles published in Kar *et al* [45] were used to augment the dataset. Only complete cases were considered in the process of matching the two datasets. This process yielded a final dataset of 19 features of five metal-oxide nanoparticles: Al<sub>2</sub>O<sub>3</sub>, CuO, Fe<sub>2</sub>O<sub>3</sub>, TiO<sub>2</sub>, and ZnO (Table 1). Cytotoxicity was used as the outcome variable, encoded as '1' (true) if measured cell viability was < 50% with respect to the control, and '0' (false) otherwise. The dataset is available on NanoTox.

### Elimination of multicollinearity:

A simple inspection of the properties in Table 1 suggested the existence of correlated features. Correlated features would adversely impact model performance as well as complicate model

interpretation. Multicollinearity is an even deeper problem in the pursuit of a non-redundant feature space [46]. The existence of highly correlated (Pearson's  $\rho \geq 0.9$ ) variables was first ascertained. To address multicollinearity, we used a systematic variance inflation factor (vif) analysis. Each independent variable was regressed on all the other independent variables in turn, and the goodness-of-fit of the models (fraction of variance explained;  $R^2$ ) were estimated. The vif-score for each independent variable was then calculated using eqn. (1). In each iteration of the vif analysis, the variable in the current set that had the largest vif score when regressed on all the other variables was eliminated. This process was continued until a set of variables all of whose vif scores  $< 5.0$  was obtained. Note that a vif score of 1.0 is possible only when a variable is perfectly independent of all other variables (all pairwise Pearson's  $\rho$  identically zero).

$$\text{vif} = 1 / (1 - R^2) \quad \dots (1)$$

## Feature transformation

The feature space could be vulnerable to heteroscedastic effects, given the varying scales for the variables. It is necessary to pre-process and prevent features with large variances from swamping out the rest. Positively skewed features could be stabilized using the log transformation. Ec values, which are negative, were first offset by +6.17, then log-transformed. Dosage spanned many orders-of-ten, and was  $\log_{10}$ -transformed. Exposure time spanned orders-of-two magnitude, so we performed a  $\log_2$  transformation. Surface charge whose values could be either positive or negative, was standardized (i.e., Z-transformed). All the other features were log-transformed (to the base  $e$ ).

## Class rebalancing

The cost of missing a toxic instance is manifold higher than the cost of missing a non-toxic instance, and the imbalance between toxic vs nontoxic instances could exacerbate this problem. In such situations, where the essential problem is to learn the minority outcome class effectively, resampling techniques could be useful [47]. We addressed the class skew problem using Synthetic Minority Over-Sampling TEchnique (SMOTE) [48]. SMOTE synthesises new minority samples from the existing ones, without influencing the instances of the majority class,

thereby increasing the number of “toxic” instances relative to the number of “non-toxic” instances. Balancing the dataset thus would normalise the learning bias arising from unequal representation of the outcome classes.

**Table 1.** The 19 physico-chemical features of MeOx nanoparticles considered in our study.

SNo	Type of feature	Feature	Shorthand
1	Intrinsic physico-chemical properties	Core size	CoreSize
2		Hydrodynamic size	HydroSize
3		Surface charge	SurfCharge
4		Surface area	SurfArea
5		Conduction band energy	Ec
6		Valence band energy	Ev
7		Standard enthalpy of formation	Hsf
8		Mulliken Electronegativity	MeO
9		Enthalpy of formation of cation	enthalpy
10		Polarization ratio	ratio
11	Periodic table properties	Pauling electronegativity	Eneg
12		Summation of electronegativity	esum
13		Molecular weight	MW
14		Number of oxygen atoms	Noxygen
15		Number of metal atoms	NMetal
16		Ratio of esum to Noxygen	esumbyo
17		Oxidation state	ox
18	Extrinsic physico-chemical properties	Exposure time	Time
19		Dosage	Dosage

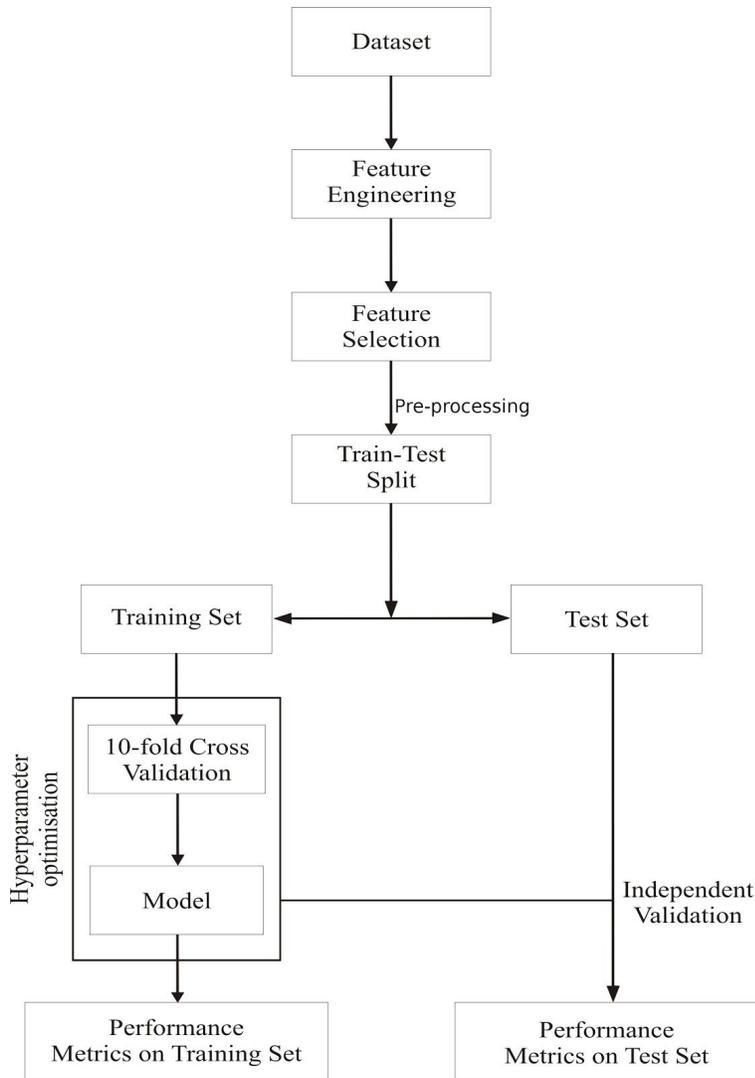
## Predictive Modelling

The overall workflow of our approach is summarized in Figure 1. The normalized, balanced dataset was randomly split into an 70:30 train:test ratio stratified on the outcome variable [49]. Towards this, a variety of classification algorithms were tried and tested, namely logistic regression [50], random forests [51], SVMs [52], and neural networks [53,54]. Table 2 shows these details and the classifier-wise hyperparameters considered in our work. The optimal values of the hyperparameters were found using 10-fold internal cross-validation [55]. The performance of each optimized model was evaluated on the unseen test set. In order to penalize false positives and false negatives equally, we used an objective measure of performance:

$$\text{Balanced accuracy} = (\text{Specificity} + \text{Sensitivity})/2 \quad \dots (2)$$

**Table 2.** Classifiers used in our study and their respective hyperparameters.  $m_{\text{try}}$  represents the number of features used for each split in the random forest model.

No.	Classifier	Type / Basis	Package / function	Hyperparameters	Optimisation
1	Logistic regression	Algebraic	glm	Threshold (= 0.5)	n/a
2	Random forest	Rule-based	randomForest	1. #trees (= 500) 2. $m_{\text{try}}$	caret::train
3	Support vector machine	Geometric	e1071	1. Kernels (linear, radial, polynomial) 2. Cost 3. Gamma 4. Degree	e1071::tune
4	Neural networks	Connectionist	RSNNS	1. #hidden layers = 1,2 2. Size of each hidden layer 3. Decay rate	caret::train, caret::mlpML



**Figure 1.** The workflow of the study upto predictive modelling. Pre-processing refers to both normalization and class balancing.

### Applicability domain

The specification of the applicability boundaries of machine learning models would increase their reliability and utility [43]. This would define the perimeter of model generalization to new instances, and safeguard against application to atypical data. We used a Euclidean nearest-neighbor approach to define the applicability domain (AD) of the machine learning models [56]. For each instance in the training set, its distances to all the other training instances were found. The nearest neighbours of each instance are then defined as the k smallest values

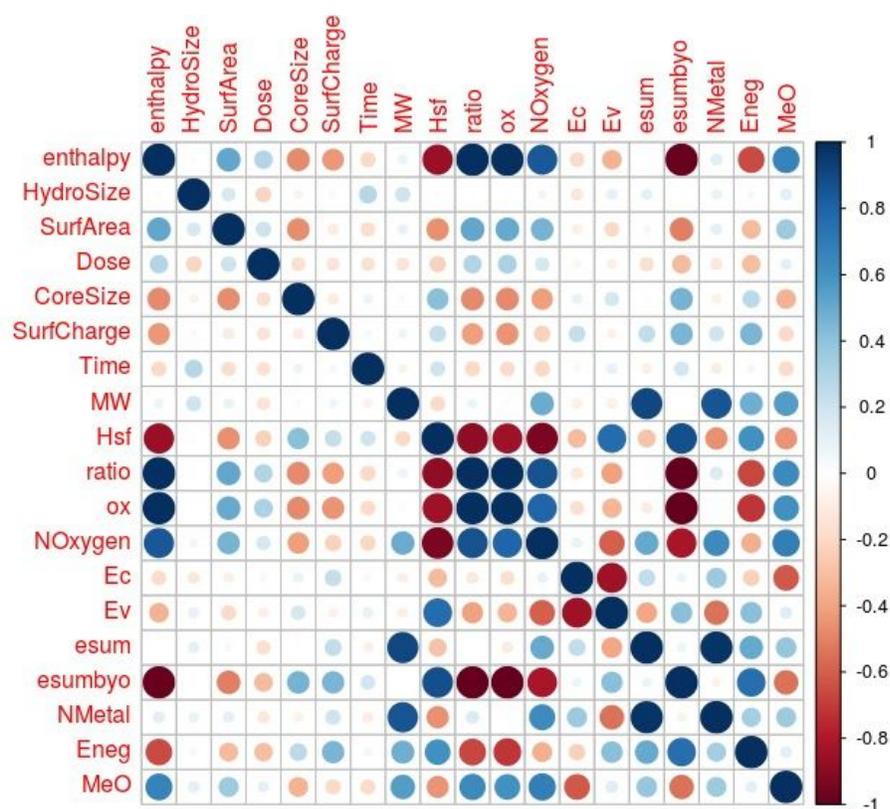
from this set, where  $k$  is an integer parameter set to the square-root of the number of instances in the training set. The mean distance of an instance to its  $k$ -nearest neighbours is found, and this process is repeated for all instances to yield the sampling distribution of these mean distances. The mean and standard deviation of this sampling distribution were designated as  $\mu_k$  and  $\sigma_k$ , respectively. The applicability domain is then defined as follows:

$$AD = \mu_k + z * \sigma_k \quad \dots(3)$$

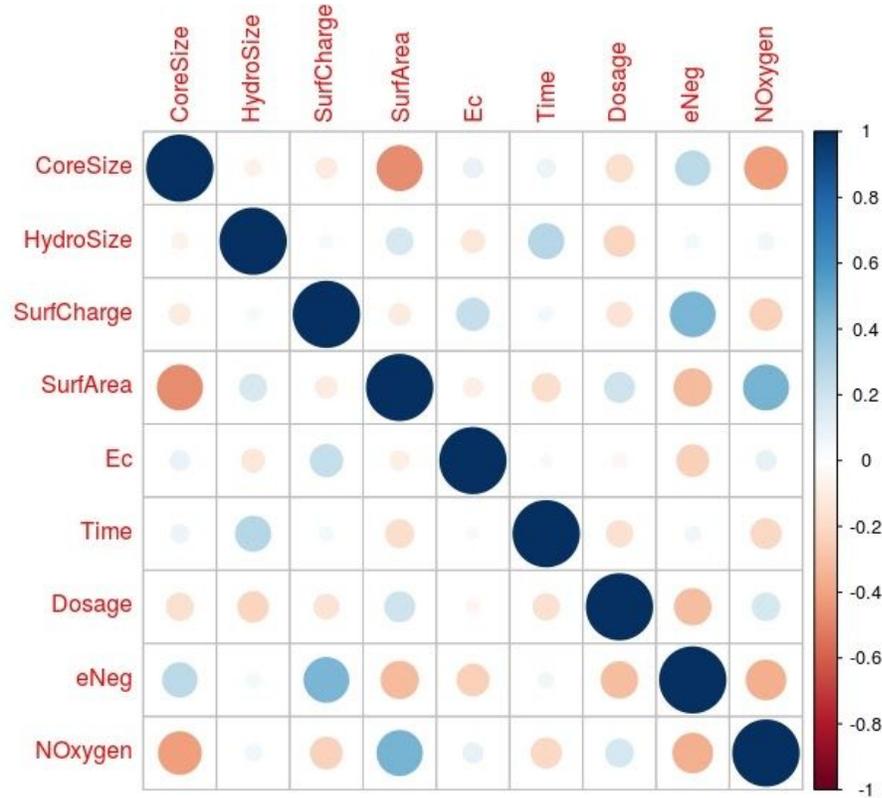
where  $z$  is an empirical parameter (related to the  $z$ -distribution) that characterises the width of belief in the model, and here set to 1.96.

## RESULTS

Our dataset consisted of 483 instances of the five metal-oxide nanoparticles with 19 features and one outcome variable. Correlogram plots identified the existence of high correlation among these 19 variables (Figure 2; see Figure S1 for a zoomed correlogram of the periodic table properties). Three clusters of high correlation were revealed: one cluster of enthalpy, Hsf, ratio, ox, NOxygen, and esumbyo; a second cluster of Ec and Ev; and a third cluster of esum, NMetal and MW. Based on the vif analysis, we were able to obtain a feature space of just nine uncorrelated non-redundant variables (Table 3). The highest vif of any variable in this feature space was < 2.02, indicating little residual multicollinearity (Figure 3). This optimal feature space included two periodic table properties (Eneq, NOxygen), five other intrinsic physico-chemical properties (CoreSize, HydroSize, SurfArea, SurfCharge, Ec), and both the extrinsic physico-chemical properties (Dose, Time). This final dataset of 483 instances with nine features and one outcome variable is available at NanoTox.



**Figure 2.** Correlogram of the 19 features. The correlation between a row feature and a column feature is shown by a dot in the corresponding cell. The size of the dot represents the magnitude of the correlation, and colour represents the sign of the correlation – blue: positive, and red: negative. White indicates a value near 0, i.e., independence.



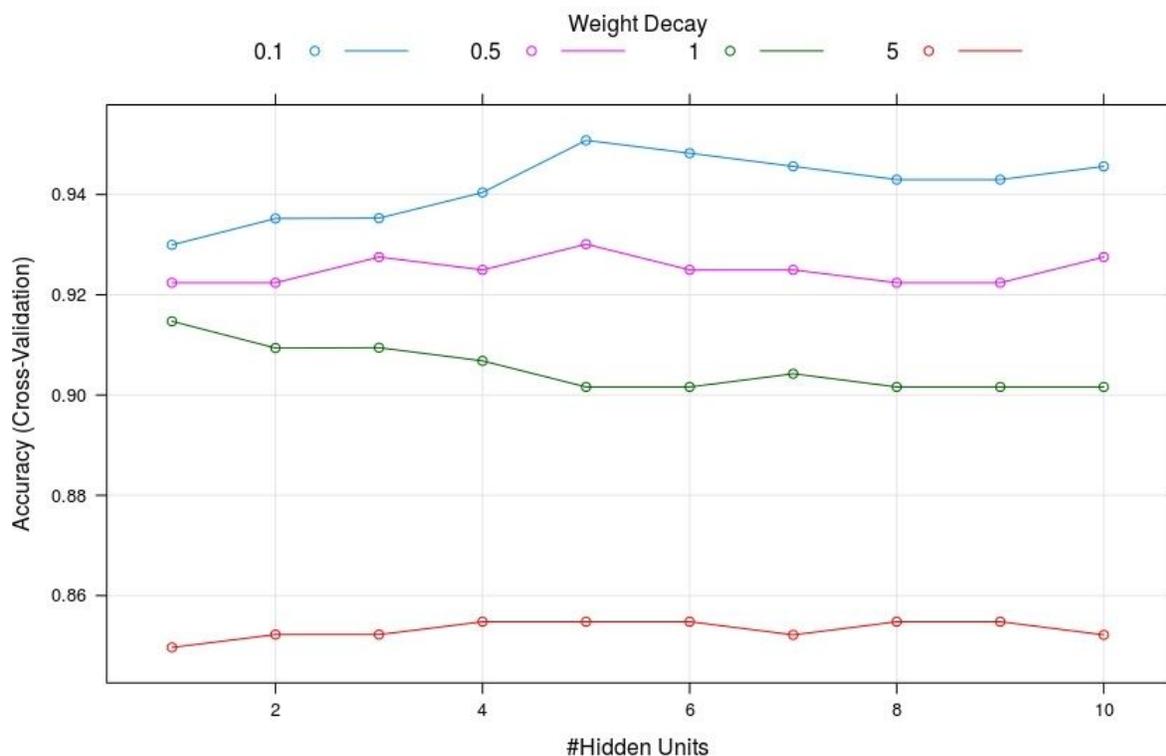
**Figure 3.** Correlogram of the optimised feature space. No subset of variables in this set is multicollinear.

**Table 3.** vif scores for the features in the final reduced set. The maximum vif score is  $\sim 2.0$ , corresponding to maximum  $R^2 \sim 0.5$  (cf. Eqn.(1)).

S.No	Feature	Variance inflation factor
1	CoreSize	1.65
2	HydroSize	1.24
3	SurfCharge	1.85
4	SurfArea	1.58
5	Ec	1.50
6	Time	1.19
7	Dose	1.21
8	Eneg	2.02
9	NOxygen	1.60

**Table 4.** Dataset normalization. Log-transformation was performed to the base  $e$ . Skewness was controlled, and the range of all predictors was brought into the same order of magnitude.

Features	Skewness before	Type of Normalization	Skewness after	Range (min-max)
CoreSize	0.92	Log	-0.2	2.01 - 4.82
HydroSize	1.76	Log	0.14	4.30 - 7.52
SurfCharge	0.45	z-score	0.45	-1.62 - +1.98
SurfArea	2.14	Log	-0.23	1.95 - 5.35
Ec	2.68	Log, with offset	-0.23	0.00 - 1.54
Time	1.36	Rescale, log	-0.48	0.00 - 4.58
Dosage	1.74	Log10	-1.5	-5.00 - +2.48
Eneg	1.46	Log	1.26	0.43 - 0.64
Noxygen	0.66	none	0.66	1 - 3



**Figure 4.** Hyperparameter tuning, for the neural network - 1 layer model. It is seen that the

cross-validation accuracy is sensitive to the choice of the set of hyperparameters.

The nine features were normalized, producing acceptable skew values for HydroSize, SurfArea, Ec, and Time (Table 4). The normalized dataset was partitioned using a random 70:30 split stratified on the outcome variable, providing a training dataset of 339 instances (with 55 'toxic' instances), and an independent test dataset of 144 instances (with 23 'toxic' instances). The training dataset (and not the test dataset) was balanced for the minority “toxic” instances using SMOTE resampling, yielding 165 'toxic' and 220 'non-toxic' instances, for a training dataset of 385 instances. This normalised and balanced dataset was used to train the various classifiers. The optimal hyperparameters of each classifier were determined using the R `e1071` package for SVMs (Figure S2), and the R `caret` package for the neural networks, both one-layer (Figure 4) and two-layers (Figure S2). The full set of model-wise optimal hyperparameters could be found in Table S1. The trained, optimised classifiers were then evaluated on the unseen test dataset. All the models, except the SVM with polynomial kernel, achieved perfect sensitivity to the 'toxic' instances, i.e. all cytotoxic nanoparticles were classified correctly. The models were not perfectly specific to the non-toxic instances, however. On this basis, the random forest and neural network – one layer models outperformed all the others. They were each frustrated by eight false positives, yielding a balanced accuracy of 96.69%. All the classifiers achieved balanced accuracy > 90%. Table 5 summarises the performance of all the models on the test set. Five 'non-toxic' instances were classified incorrectly by all the models, representing refractory instances and constituting a challenge to perfect learning. One of these instances was only marginally viable (0.52), indicating the possible source of refractoriness.

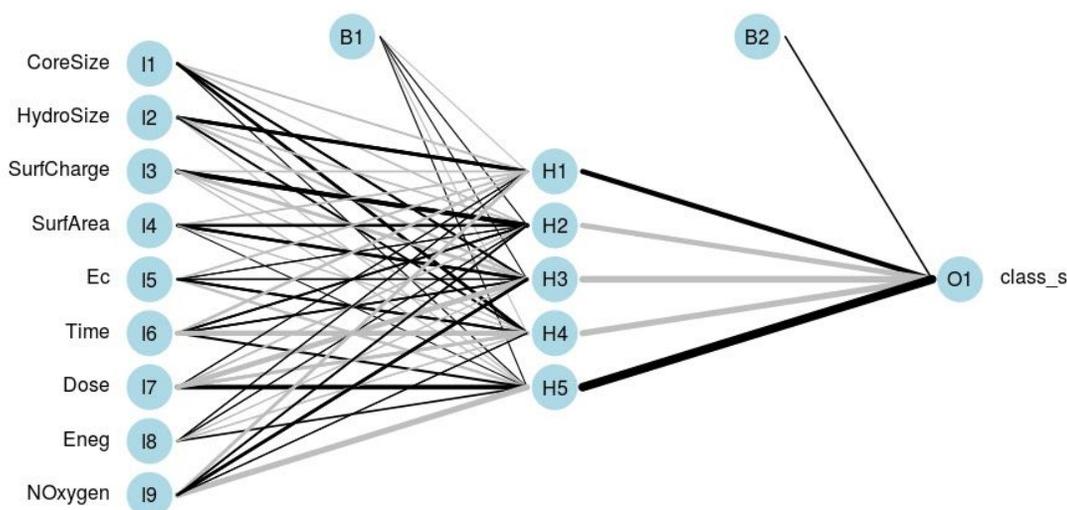
**Table 5.** Performance of the various models. Models with balanced accuracy > 94% are highlighted.

ID	Classifier	Train set			Test set	
		Accuracy	Balanced accuracy	Cross-valid accuracy	Accuracy	Balanced accuracy
Model_1	Logistic regression	0.94	0.94	0.93	0.91	<b>0.95</b>
Model_2	Random forest	0.98	0.98	0.94	0.94	<b>0.97</b>
Model_3a	SVM-Linear	0.94	0.95	1	0.9	0.94
Model_3b	SVM-Radial	0.94	0.94	1	0.86	0.92
Model_3c	SVM-Poly	0.98	0.98	1	0.84	0.85
Model_4a	Neural network – 1L	0.96	0.96	0.94	0.94	<b>0.97</b>
Model_4b	Neural network – 2L	0.96	0.95	0.95	0.91	<b>0.95</b>

### Deployment:

The applicability domain was calculated with the normalised train data, prior to SMOTE balancing. Substituting  $k = 19$  and  $z = 1.96$  in eqn. (2) yielded the AD threshold = 2.23. About 95% of the test instances (i.e, 137 / 144 instances) were located within the AD radius. It must be noted that the misclassified instances did not coincide with these outliers. We have provided a workflow, deployment.R (available at NanoTox), for prediction on new, untested oxides. The prediction is executed by a majority-voting ensemble classifier [57], since bagging the predictions of the best models on the test set improved the performance to just five false positives (~ 98% balanced accuracy). Any new instance for classification supplied by the user is pre-processed (normalised), and its 'typicality' determined by calculating its distances to the instances in the original train data, and finding the mean,  $D_i$ , of the 19 closest distances. If the  $D_i$  is greater than the AD threshold, then the instance is deemed atypical for requesting the

ensemble model. Predictions are obtained using the top two models – the random forest and the neural network - one layer, and a consensus prediction is sought. In the absence of a consensus, an ensemble of the top five classifiers, all with balanced accuracy > 94% (highlighted in Table 5), is used. In the end, the majority prediction of the ensemble classifier is the predicted cytotoxicity of the given instance. Deployment.R automates this pipeline for a batch of new, untested oxides of any size. Furthermore, the RDS images of all the models trained in our study are provided on NanoTox, for the interested scientist.



**Figure 5.** A schematic of the trained neural network- 1 layer model, with the weights of the connections indicated by the linewidth. Black lines indicate positive weights, and gray lines indicate negative weights. Two bias units are seen, one for the hidden layer, and one for the output layer.

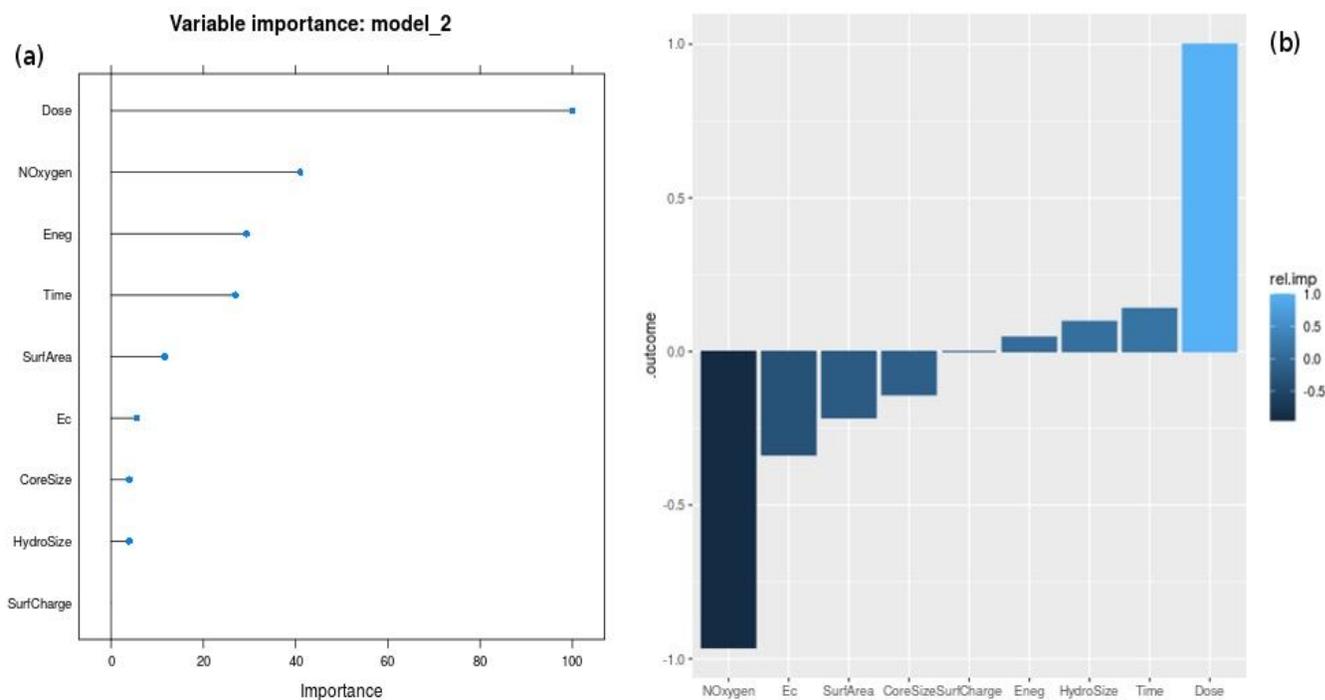
## DISCUSSION

It is clear that SMOTE balancing made a difference in the ability of the classifiers to detect the under-represented 'toxic' instances. Filtering based on applicability domain and use of an

ensemble classification strategy further mitigate model uncertainty given the 'no free lunch' theory. Benchmarking our results with Choi *et al.* [36], we see that the best model in each classifier from our work outperformed the corresponding best models of their work (Table 6). The overall best models in our work (random forest and neural network 1-layer) yielded a balanced accuracy of ~97% compared to 93% for their best overall model ('neural networks'). All the five models from this work with balanced accuracy > 93% are deployed in an ensemble classifier to further mitigate uncertainty in prediction.

**Table 6.** Benchmarking. SVM (a), (b), and (c) correspond to linear, radial, and polynomial kernels, respectively. Neural networks (a) and (b) refer to one and two hidden layer(s), respectively. No information regarding model hyperparameters were available in Choi et al. [35]. The best-performing models from our work are highlighted.

Model	Balanced accuracy (%)	
	Choi et al. [35]	Present work
Logistic regression	92	94.63
Random forest	91	<b>96.69</b>
SVM	91	(a) 94.21
		(b) 91.74
		(c) 85.21
Neural networks	93	(a) <b>96.69</b>
		(b) 94.63



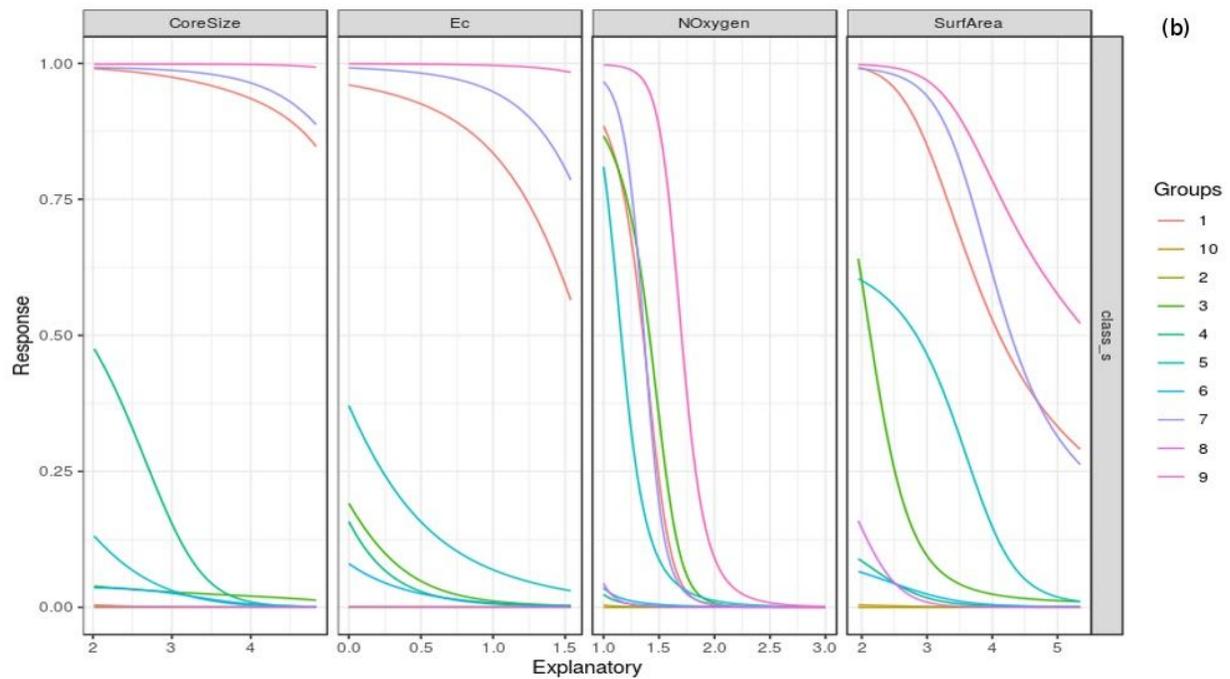
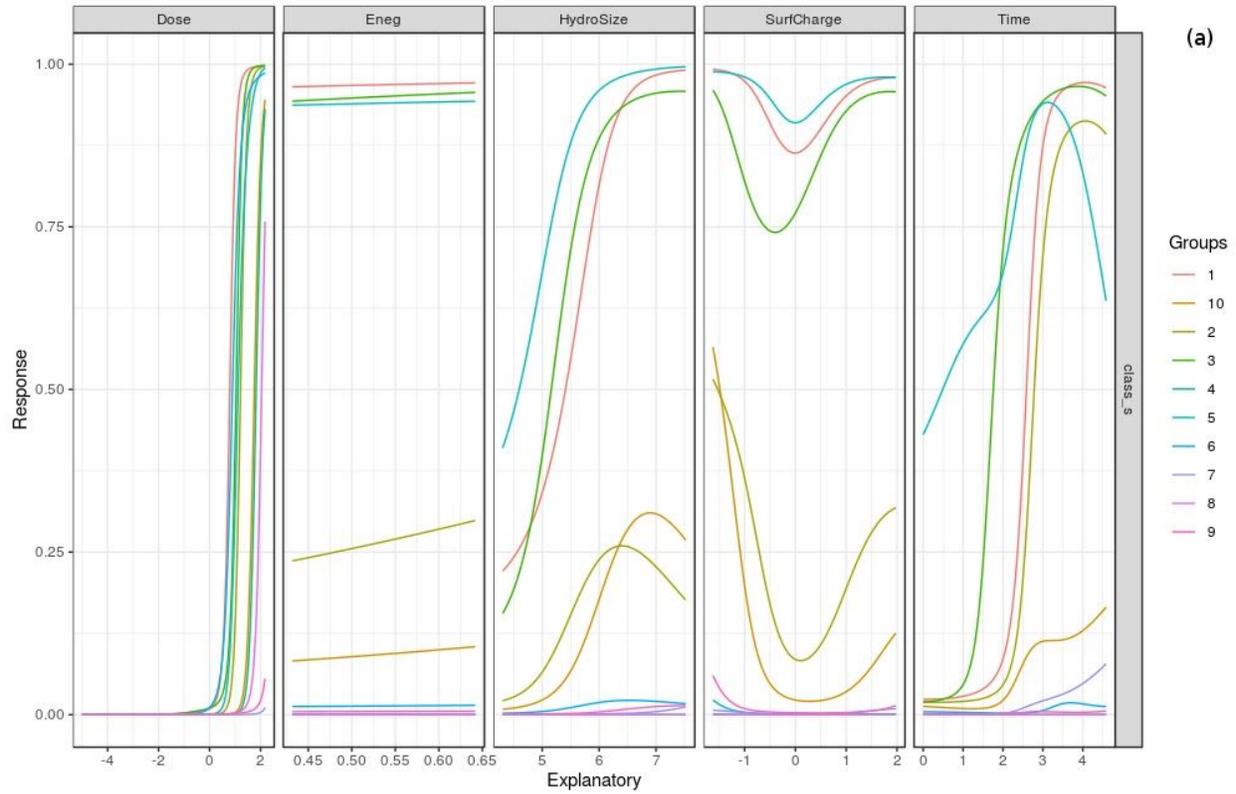
**Figure 6.** (a) Normalized variable importance for the Random Forest model computed with caret. Dose is by and far the attribute with the greatest effect on the toxicity in the Random Forest model. (b) Relative importance plot for the NeuralNet-1L. Positive values correspond to the ‘true’ (i.e, ‘toxic’) class, and negative values correspond to the ‘non-toxic’ class. It is seen that Dose and NOxygen exert the maximum importance on the outcome class, though in opposite directions.

Figure 5 shows the best-performing neural network model, with connections weighted by their importance [58]. A variable importance plot of the random forest shows Dose as the most important predictor, followed by NOxygen, Eneg and Time (Figure 6(a)) [59]. A relative importance analysis of the neural network – 1-layer model obtains concurrence to these findings, and adds a direction to the favoured binary outcome (Figure 6(b)) [60,61]. Dose emerges as the key variable determining nanoparticle toxicity, and Time, HydroSize and Eneg are the other variables influencing the 'toxic' prediction. NOxygen emerges as the key predictor influencing the 'nontoxic' prediction, and SurfArea, Ec and CoreSize are the other predictors in this category. Logistic regression provides us with not just the effect size (coefficients) of the individual variables but also an estimate of their significance (in terms of the p-value of the coefficients) (Table S2). It is significant that the logistic regression model supports these findings, with few

inconsistencies. Further the sign of the coefficient of each variable matches its direction of influence presented in Figures S3, S4. It is noteworthy that the two 'periodic table' properties (E<sub>neg</sub>, Noxygen), and the quantum chemical property, E<sub>c</sub>, show large effect sizes (coefficients) but poor significance, when all the other variables remain extremely significant.

Consensus among the models is necessary for explainable AI [62], and in this direction, we performed a Lek sensitivity analysis with the neural network 1-layer model [63]. How does the response variable change with changes in a given explanatory variable, given the context of the other explanatory variables? In investigating the effect of one explanatory variable, all the other explanatory variables are clustered into a specified number of lakes with like characteristics. While the unevaluated explanatory variables are held constant at the centroid of one lake cluster, the explanatory variable of interest is sequenced from minimum to maximum in 100 quantile steps, with the response variable predicted at each step, yielding a sensitivity curve. This process is iterated for each lake of the unevaluated explanatory variables, yielding the sensitivity profile of the response variable with respect to the specific explanatory variable in the context of the unevaluated explanatory variables. We set the number of clusters to ten, to visualize a sufficient number of the response curves for each explanatory variable. In this way, the sensitivity profiles of the response variable are obtained for each predictor (Figure 7).

The two input variables that decisively differentiate the outcome are Dose and Noxygen. Dose appears to exert a nearly thresholding effect on the toxic class. The consistent sigmoidal effect seen in the 'dose-response' curve, independent of the lake of unevaluated explanatory variables, echoes the maxim attributed to Paracelsus, 'The dose makes the poison.' The attributes influencing toxicity also included: (i) Time, with a pronounced effect depending on the lakes of the unevaluated variables; and (ii) HydroSize, with a steady non-linear effect on toxicity that is also sensitive to the context of the unevaluated explanatory variables. The response profile for E<sub>neg</sub> is almost flat at all lakes, indicating little to no effect in changing the outcome. The interpretation of the response with respect to SurfCharge remained obscure. NOxygen emerged as the attribute with the clearest inverse effect on toxicity, with a response profile displaying a tipping point to non-toxic class at most, but not all, of the centroids. Other attributes seen to dial down the toxicity include SurfArea, CoreSize, and E<sub>c</sub>. These observations of effect size may be tempered with a significance analysis for deeper understanding.



**Figure 7.** (a) Lek sensitivity analysis of attributes with positive effect on the outcome class. The steep effect of Dose is evident, with the location of the tipping point moving slightly with the cluster of the unevaluated variables. Increasing exposure times and HydroSize are also seen to tip to toxicity. (b) Lek sensitivity analysis of attributes with relatively consistent negative effect on the outcome class: CoreSize, Ec, NOxygen, and SurfArea. The number of lakes of the unevaluated variables is set to 10 in both the cases.

In summary, the ML models of our work are represented by a purely numeric feature space of just nine predictors, and it is possible to consider them in their entirety, similar to the interpretability of a QSAR model. The models conform to the Findable, Accessible, Interoperable, Reusable (FAIR) principles, and are presented in a unified ensemble prediction engine, NanoTox (<https://github.com/NanoTox>). The scripts necessary for normalization and reproducible research are also available at NanoTox. Our methods may be extendable to other classes of engineered nanomaterials requiring urgent, sustainable, and rapid hazard classification prior to regular use [64-67].

## CONCLUSION

We have optimized the problem formulation of cytotoxicity modelling of nanoparticles using a principled approach agnostic of *in vitro* characteristics. The feature space is trimmed for multicollinearity, tunable hyperparameters were optimized, and the training data corrected for class imbalance. These steps further improved the performance of the parsimonious ML models to >96% balanced accuracy. The benefits of a parsimonious approach to modelling nanoparticle toxicity include enhanced model interpretability and generalizability. We have embedded our models into an unambiguous ensemble classifier that surpasses ~98% balanced accuracy. Our entire workflow is available as a free open-source resource for use and enhancement by the scientific community towards proactive non-invasive testing and design of nanoparticles for varied applications.

## Acknowledgements

We would like to thank the School of Chemical and BioTechnology, SASTRA Deemed University for infrastructure and computing support. This work was supported in part by the

SASTRA TRR grant to A.P.

### **Author contributions**

A.P. conceived, designed and supervised the work. N.A. and A.P. performed research, and analyzed the results; A.P. wrote the paper.

### **REFERENCES**

1. Sanvicens N and Marco M. P. (2008). Multifunctional nanoparticles—properties and prospects for their use in human medicine. *Trends in biotechnology*, 26(8), 425-433.
2. Ogale, Satishchandra Balkrishna, T. Venky Venkatesan, and Mark Blamire, eds. *Functional metal oxides: new science and novel applications*. John Wiley & Sons, 2013.
3. Subbiah DK, Babu KJ, Das A, and Rayappan JBB (2009). NiO<sub>x</sub> Nanoflower Modified Cotton Fabric for UV Filter and Gas Sensing Applications. *ACS Applied Materials & Interfaces* 11(22): 20045-20055. DOI: 10.1021/acsami.9b04682
4. Monteiro-Riviere N, A and Tran, C. L. (Eds.) (2007). *Nanotoxicology: characterization, dosing and health effects*. CRC Press.
5. Aillon, Kristin L., et al. "Effects of nanomaterial physicochemical properties on in vivo toxicity." *Advanced drug delivery reviews* 61.6 (2009): 457-466.
6. Buzea, Cristina, Ivan I. Pacheco, and Kevin Robbie. "Nanomaterials and nanoparticles: sources and toxicity." *Biointerphases* 2.4 (2007): MR17-MR71.
7. Jeng H A, & Swanson J. (2006). Toxicity of metal oxide nanoparticles in mammalian cells. *Journal of Environmental Science and Health Part A*, 41(12), 2699-2711.
8. Fu, P.P.; Xia, Q.; Hwang, H.-M.; Ray, P.C.; Yu, H. Mechanisms of nanotoxicity: Generation of reactive oxygen species. *J. Food Drug Anal.* 2014, 22, 64–75.
9. Sarkar A, Ghosh M and Sil, P. C. (2014). Nanotoxicity: oxidative stress mediated toxicity of metal and metal oxide nanoparticles. *Journal of nanoscience and nanotechnology*, 14(1), 730-743.
10. Abdal Dayem, A.; Hossain, M.K.; Lee, S.B.; Kim, K.; Saha, S.K.; Yang, G.-M.; Choi,

- H.Y.; Cho, S.-G. The Role of Reactive Oxygen Species (ROS) in the Biological Activities of Metallic Nanoparticles. *Int. J. Mol. Sci.* 2017, 18, 120.
11. Shah, S.A.; Yoon, G.H.; Ahmad, A.; Ullah, F.; Amin, F.U.; Kim, M.O. Nanoscale-alumina induces oxidative stress and accelerates amyloid beta (A $\beta$ ) production in ICR female mice. *Nanoscale* 2015, 7, 15225–15237.
  12. Golbamaki, Nazanin, et al. "Genotoxicity of metal oxide nanomaterials: review of recent data and discussion of possible mechanisms." *Nanoscale* 7.6 (2015): 2154-2198.
  13. Wang D, Lin Z, Wang T, Yao Z, Qin M, Zheng S and Lu W. (2016). Where does the toxicity of metal oxide nanoparticles come from: the nanoparticles, the ions, or a combination of both?. *Journal of hazardous materials*, 308, 328-334.
  14. Ai, J.; Biazar, E.; Jafarpour, M.; Montazeri, M.; Majdi, A.; Aminifard, S.; Zafari, M.; Akbari, H.R.; Rad, H.G. Nanotoxicology and nanoparticle safety in biomedical designs. *Int. J. Nanomed.* 2011, 6, 1117–1127.
  15. Singh, A.V.; Laux, P.; Luch, A.; Sudrik, C.; Wiehr, S.; Wild, A.-M.; Santomauro, G.; Bill, J.; Sitti, M. Review of emerging concepts in nanotoxicology: Opportunities and challenges for safer nanomaterial design. *Toxicol. Mech. Methods* 2019, 29, 378–387.
  16. Fadeel, Bengt, ed. *Handbook of safety assessment of nanomaterials: from toxicological testing to personalized medicine*. Vol. 5. CRC Press, 2014.
  17. Chen, R.; Qiao, J.; Bai, R.; Zhao, Y.; Chen, C. Intelligent testing strategy and analytical techniques for the safety assessment of nanomaterials. *Anal. Bioanal. Chem.* 2018, 410, 6051–6066.
  18. ECHA. *The use of alternatives to testing on animals for the REACH Regulation*; ECHA: Helsinki, Finland, 2017; p. 103.
  19. Gellatly, N.; Sewell, F. Regulatory acceptance of in silico approaches for the safety assessment of cosmetic-related substances. *Comput. Toxicol.* 2019, 11, 82–89.
  20. Burgdorf, T.; Piersma, A.H.; Landsiedel, R.; Clewell, R.; Kleinstreuer, N.; Oelgeschläger, M.; Desprez, B.; Kienhuis, A.; Bos, P.; de Vries, R.; et al. Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory

toxicology—Evolution versus revolution. *Toxicol. Vitr.* 2019, 59, 1–11.

21. Raies A B and Bajic V. B. (2016). In silico toxicology: computational methods for the prediction of chemical toxicity. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 6(2), 147-172.
22. Le, T.C.; Yin, H.; Chen, R.; Chen, Y.; Zhao, L.; Casey, P.S.; Chen, C.; Winkler, D.A. An Experimental and Computational Approach to the Development of ZnO Nanoparticles that are Safe by Design. *Small (Weinh. Bergstr. Ger.)* 2016, 12, 3568–3577
23. Knudsen T B, Keller D A, Sander M, Carney E W, Doerrer N G, Eaton D L and Watkins, P. B. (2015). FutureTox II: in vitro data and in silico models for predictive toxicology. *Toxicological Sciences*, 143(2), 256-267.
24. Furxhi, I.; Murphy, F.; Poland, C.A.; Sheehan, B.; Mullins, M.; Mantecca, P. Application of Bayesian networks in determining nanoparticle-induced cellular outcomes using transcriptomics. *Nanotoxicology* 2019, 13, 827–848.
25. Labouta, H.I.; Asgarian, N.; Rinker, K.; Cramb, D.T. Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature. *ACS Nano* 2019, 13, 1583–1594.
26. Bilal, M.; Oh, E.; Liu, R.; Breger, J.C.; Medintz, I.L.; Cohen, Y. Bayesian Network Resource for Meta-Analysis: Cellular Toxicity of Quantum Dots. *Small (Weinh. Bergstr. Ger.)* 2019, 15, 1900510.
27. Marvin, H.J.P.; Bouzembrak, Y.; Janssen, E.M.; van der Zande, M.; Murphy, F.; Sheehan, B.; Mullins, M.; Bouwmeester, H. Application of Bayesian networks for hazard ranking of nanomaterials to support human health risk assessment. *Nanotoxicology* 2017, 11, 123–133.
28. Luan, F.; Kleandrova, V.V.; González-Díaz, H.; Ruso, J.M.; Melo, A.; Speck-Planche, A.; Cordeiro, M.N.D.S. Computer-aided nanotoxicology: Assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale* 2014, 6, 10623–10630.
29. Shaw, S. Y. et al. Perturbational profiling of nanomaterial biologic activity. *Proc. Natl. Acad. Sci.* 105, 7387–7392 (2008).

30. Concu, R.; Kleandrova, V.V.; Speck-Planche, A.; Cordeiro, M.N.D.S. Probing the toxicity of nanoparticles: A unified in silico machine learning model based on perturbation theory. *Nanotoxicology* 2017, 11, 891–906.
31. Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C.A. Practices and Trends of Machine Learning Application in Nanotoxicology. *Nanomaterials* 2020, 10, 116.
32. Puzyn, T., Rasulev, B., Gajewicz, A. et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotech* 6, 175–178 (2011). <https://doi.org/10.1038/nnano.2011.10>
33. Pan Y, Li, T, Cheng J, Telesca D, Zink J I and Jiang J. (2016). Nano-QSAR modeling for predicting the cytotoxicity of metal oxide nanoparticles using novel descriptors. *RSC Advances*, 6(31), 25766-25775.
34. Mu Y, Wu F, Zhao Q, Ji R, Qie Y, Zhou Y and Xing B. (2016). Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. *Nanotoxicology*, 10(9), 1207-1214.
35. Randić, M. Generalized molecular descriptors. *J Math Chem* 7, 155–168 (1991). <https://doi.org/10.1007/BF01200821>
36. Choi J S, Ha M K, Trinh T X, Yoon T H and Byun, H. G. (2018). Towards a generalized toxicity prediction model for oxide nanomaterials using integrated data from different sources. *Scientific reports*, 8(1), 6110.
37. Choi J S, Trinh T, X Yoon, T H, Kim J and Byun, H. G. (2019). Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. *Chemosphere*, 217, 243-249.
38. Furxhi I, Murphy F. (2020) Predicting In Vitro Neurotoxicity Induced by Nanoparticles Using Machine Learning. *Int J Mol Sci*, 21(15):5280. doi: 10.3390/ijms21155280.
39. Trinh, T.X.; Ha, M.K.; Choi, J.S.; Byun, H.G.; Yoon, T.H. Curation of datasets, assessment of their quality and completeness, and nanoSAR classification model development for metallic nanoparticles. *Environ. Sci. Nano* 2018, 5, 1902–1910.
40. Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, A.C. Nanotoxicology data for in

silico tools. A literature review. *Nanotoxicology* 2020, 14, 612–637.

41. Ha, M.K.; Trinh, T.X.; Choi, J.S.; Maulina, D.; Byun, H.G.; Yoon, T.H. Toxicity Classification of Oxide Nanomaterials: Effects of Data Gap Filling and PChem Score-based Screening Approaches. *Sci. Rep.* 2018, 8, 3141.
42. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. *Intelligent data analysis.* 2002; 6(5):429–49.
43. OECD. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models; OECD Publishing: Paris, France, 2014
44. Jha S K, Yoon T H and Pan Z. (2018). Multivariate statistical analysis for selecting optimal descriptors in the toxicity modeling of nanomaterials. *Computers in biology and medicine*, 99, 161-172
45. Kar S, Gajewicz A, Puzyn T, Roy K and Leszczynski J. (2014). Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and environmental safety*, 107, 162-169.
46. Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of machine learning research.* 2003; 3(Mar):1157–82.
47. Singh ND, Dhall A. Clustering and learning from imbalanced data. *arXiv preprint arXiv:181100972.* 2018.
48. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research.* 2002; 16:321–57.
49. Faraway, J. J. Data splitting strategies for reducing the effect of model selection on inference. *Comput. Sci. Stat.* 30, 332–341 (1998)
50. Hosmer Jr, David W, Stanley Lemeshow and Rodney X. Sturdivant. *Applied logistic regression.* Vol. 398. John Wiley & Sons, 2013.
51. Breiman, L. (2001). Random forests. *Machine Learning*, 45: 5-32.
52. Chang, C-C and Lin, C-J. LIBSVM : a library for support vector machines. *ACM*

Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.

53. Rumelhart DE, Hinton GE, Williams RJ (1986). "Learning Representations by BackPropagating Errors." *Nature*, 323(6088), 533–536. doi:10.1038/323533a0.
54. Bergmeir C, Benítez JM (2012). "Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS." *Journal of Statistical Software*, 46(7), 1–26. doi:10.18637/jss.v046.i07.
55. Kuhn M (2008). "Building Predictive Models in R Using the caret Package." *Journal of Statistical Software*, 28(5), 1–26. doi:10.18637/jss.v028.i05
56. Sahigara, F.; Kamel, M.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules* 2012, 17, 4791–4810.
57. Chen, J. J., et al. "Classification ensembles for unbalanced class sizes in predictive toxicology." *SAR and QSAR in Environmental Research* 16.6 (2005): 517-529.
58. Beck, M. (2018). *NeuralNetTools: Visualization and Analysis Tools for Neural Networks*. *Journal of Statistical Software*, 85(11), 1 - 20. doi:<http://dx.doi.org/10.18637/jss.v085.i11>
59. Garson GD (1991). "Interpreting Neural Network Connection Weights." *Artificial Intelligence Expert*, 6(4), 46–51.
60. Olden JD, Jackson DA. 2002. Illuminating the 'black box': A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*. 154:135-150.
61. Olden JD, Joy MK, Death RG. 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*. 2004. 389-397.
62. Lipton ZC. (2017) *The Mythos of Model Interpretability*. arXiv:1606.03490
63. Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S (1996). "Application of Neural Networks to Modelling Nonlinear Relationships in Ecology." *Ecological Modelling*, 90(1), 39–52. doi:10.1016/0304-3800(95)00142-5.

64. Kleandrova, V.V.; Luan, F.; González-Díaz, H.; Ruso, J.M.; Melo, A.; Speck-Planche, A.; Cordeiro, M.N.D.S. Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. *Environ. Int.* 2014, 73, 288–294.
65. Trinh, T.X.; Choi, J.-S.; Jeon, H.; Byun, H.-G.; Yoon, T.-H.; Kim, J. Quasi-SMILES-Based Nano-Quantitative Structure–Activity Relationship Model to Predict the Cytotoxicity of Multiwalled Carbon Nanotubes to Human Lung Cells. *Chem. Res. Toxicol.* 2018, 31, 183–190.
66. Oh, E.; Liu, R.; Nel, A.; Gemill, K.B.; Bilal, M.; Cohen, Y.; Medintz, I.L. Meta-analysis of cellular toxicity for cadmium-containing quantum dots. *Nat. Nanotechnol.* 2016, 11, 479.
67. Horev-Azaria, L.; Kirkpatrick, C.J.; Korenstein, R.; Marche, P.N.; Maimon, O.; Ponti, J.; Romano, R.; Rossi, F.; Golla-Schindler, U.; Sommer, D.; et al. Predictive Toxicology of Cobalt Nanoparticles and Ions: Comparative In Vitro Study of Different Cellular Models Using Methods of Knowledge Discovery from Data. *Toxicol. Sci.* 2011, 122, 489–501.
68. R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>