The nano-SPR model for predicting dissolution rate of metal and metal oxide nanomaterials in aqueous environment

*Michal Kalapus, Tomasz Puzyn**

Laboratory of Environmental Chemoinformatics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland

1. Applicability Domain

Initially, multiple subclassifiers are constructed using ensemble learning techniques, wherein the training data undergoes random variations in X variables and samples. Classification errors are calculated for each subclassifier. For new data points, the subclassifiers predict class values, and the final predicted class is determined based on the majority vote. Performance Ratios (PR) such as $RT_{!} = \frac{N_{subclassfiers}}{k}$ and $RT_{-1} = 1 - RT_{1}$ (Eq.2), where $N_{subclassifiers}$ is the number of subclassifiers which predictions is equal to 1, and k is the number subclassifier models. PR representing the ratios of predicted class values, are utilized to assess prediction reliability. In parallel, data density is utilized as an additional PR for AD. By evaluating the number of neighboring training data points, data density provides insight into the reliability of predictions. If the density is high, indicating a significant number of neighboring data points, the prediction result is considered reliable. To establish the AD, a combined approach is proposed. First, a threshold for data density is set. If the data density of a new data point falls below this threshold, the prediction result is deemed unreliable. However, for data points with a density exceeding the threshold, the PR based on ensemble learning is used to evaluate the prediction result's reliability. By integrating ensemble learning and data density, this methodology offers a

comprehensive framework for defining the AD of classifiers, ensuring reliable predictions within the established boundaries.

The second method described by Roy *et al*. to check whether an object lies within the applicability domain of our model, we assess the descriptors of the training set under the assumption that they ideally follow a normal distribution. Based on the characteristics of the normal distribution, approximately 99.7% of data points are expected to fall within three standard deviations (±3 SD) from the mean. This range, therefore, represents where the majority of the training compounds are located. For any given compound *k*, if the standardized value for any descriptor *i* (denoted as S_{ki}) exceeds 3, the compound is considered an X-outlier (if it belongs to the training set) or is outside the AD (if it belongs to the test set) with respect to that descriptor. If some S_{ki} values exceed 3 while others do not, the compound is similar in certain descriptors and dissimilar in others, necessitating a criterion for assessment. To address this scenario, method utilize the standard score corresponding to a cumulative probability of 90%, which is Z=1.28 in a standard normal distribution. Then compute a new statistic, S_{new}, defined as the mean of the S_{ki} values plus 1.28 times their standard deviation. If $S_{new} < 3$, there is a 90% probability that the S_{ki} values for that object are below 3. In such cases, the compound is considered not an X-outlier or within the AD.

2. Evaluation metrics

Detailed information about evaluation metrics can be found in Supplementary Information. The accuracy of a model is determined by calculating the ratio of correctly classified instances to the total number of instances. This metric is particularly useful when the dataset is balanced, meaning that all classes are represented equally. In such cases, it provides a quick and general assessment of performance. However, it is important to exercise caution when interpreting accuracy in cases where the classes are imbalanced.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

The precision of a prediction is evaluated by calculating the ratio of true positives to the total number of predicted positives. By prioritising precision, we can ensure that the classifier makes fewer incorrect positive predictions, which is of paramount importance when we require a high level of confidence in each positive classification.

58 $$Precision = \frac{TP}{(TP + FP)}$$

59 The recall, or sensitivity, of a model represents its capacity to identify all actual positive
60 instances. A high recall rate indicates that the model is able to effectively identify the majority
61 of true positive cases, making it an appropriate choice for scenarios where the cost of failing to
62 detect positive cases is greater than that of incorrectly flagging negatives as positives.

63 $$Recall = \frac{TP}{(TP + FN)}$$

64 Kappa is a statistical measure that assesses the agreement between a model's predictions and
65 the actual labels, taking into account the probability of such an agreement occurring by chance.
66 This metric is advantageous for comparing the performance of classifiers across datasets with
67 disparate distributions or class imbalances. The incorporation of random chance into the Kappa
68 metric provides a more robust measure than accuracy in scenarios with unbalanced classes.
69 This allows for the identification of whether the model is genuinely effective or merely
70 exploiting the dataset's composition.

71 $$Kappa = \frac{Accuracy - Pe}{1 - Pe}, where\ Pe = \frac{(TP + FP)(TP + FN)}{(TP + TN + FP + FN)^2} + \frac{(TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}$$

72 The F1 Score is a metric that balances precision and recall, providing a single score that reflects
73 both the correctness of positive predictions and the ability to capture all positives. Formally,
74 the F1 Score is the harmonic mean of precision and recall. The F1 score is particularly
75 advantageous in scenarios where the class distribution is uneven or when the costs associated
76 with false positives and false negatives are significant. The F1 score offers a more
77 comprehensive evaluation of a classifier's performance than precision or recall alone, rendering
78 it an appropriate choice when an overall metric that considers both is desired.

79 $$F1 = 2 \cdot \frac{(Precision \cdot Recall)}{(Precision + Recall)}$$

80 The MCC (Matthews Correlation Coefficient) is a statistical measure that takes into account
81 both true and false positives and negatives, making it particularly informative in datasets with
82 imbalanced classes. The coefficient ranges from -1 to +1, with +1 indicating a perfect
83 prediction, 0 indicating no better than random guessing, and -1 indicating total disagreement.
84 In scientific contexts where all errors are equally costly and the dataset may not have a balanced

class distribution, the MCC is of great value as it provides a reliable measure even when classes are unevenly distributed.

$$MCC = \frac{(TP \cdot TN - FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Balanced Accuracy is a method of accounting for class imbalances in a dataset by averaging the recall values of each class, thereby providing a more equitable view of the model's effectiveness across classes. It is of paramount importance when working with imbalanced datasets, as it prevents a situation whereby high accuracy in one class would result in poor performance in another being overlooked. The use of balanced accuracy is beneficial when assessing the performance of models on datasets comprising imbalanced classes. This approach ensures that the detection rate for each class is represented equally in the performance metric, providing a more accurate evaluation of the model's effectiveness across different class distributions.

$$Balanced\ Accuracy = \frac{1}{2} \cdot \left(\frac{TP}{(TP + FN)} + \frac{TN}{(TN + FP)} \right)$$

where TP stands for True Positive, TN True Negative, FP False Positive, FN False Negative.

3. Descriptor calculations

As stated in the "Descriptor calculations" section in main text, the modelling did not incorporate some of the periodic table descriptors. This is attributable to the impracticability of calculating the values of certain descriptors, given the presence of metals in the dataset (initially, these descriptors were prepared for utilisation with metal oxides in mind). The subsequent section provides a detailed rationale for the exclusion of specific descriptors during the modelling stage.

1. *Total metal electronegativity in a specific metal oxide* – all metals and metals oxides in dataset contains only one metal atom in formula, so this descriptor have the same value as *electroneativity of metal in metal oxide.*
2. *Total metal electronegativity in a specific metal oxide relative to the number of oxygen atoms* – it requiers dividing by number of oxygen atoms. For metals there is 0 oxygen atoms, so to avoid dividing by 0 this descriptor was not introduce to the manuscript.
3. $\sum \alpha_{metal} = \alpha_{metal} \cdot N_{metal}$ – as N_{metal} in dataset is equal to 1, sum is eqaul to α_{metal}.

4. $\sum \alpha_{oxy} = N_{oxy} \cdot 0.33$ – due to the low variance in N_{oxy} (only three values) it was not included into dataset.
5. *Square of the summation of alpha, gives a measure of the molecular bulk* – after autoscailing value for this descriptor is the same as for alpha.
6. *Squared sum epsilon relative to the number of atoms* – after autoscailing value for this descriptor is the same as for epsilon.
7. *Valence of metal* – is not clear how this is described in orginal manuscript. Metals used in modeling can have different valence state, so due to this reason this descriptor was neglected.

Rest of computed descriptors (8) was added to the dataset, and one more time feature selection were performed. All of added descriptors were not choose to the final model, as their variance and so also impact on modeling were low.

4. Splitting algorithm

Table S1. Distriubution on object in between different class depends on spliting data algorithm.

	1:X		**KS**		**Random selection 1**		**Random selection 2**		**Random selection 3**	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Very slow dissolution	5	3	7	1	3	5	5	3	6	2
Partail dissolution	23	11	30	4	20	14	25	9	21	13
Quick dissolution	48	24	39	33	53	19	46	26	49	23

Table S2. Evaluation metrics based on splitting method use.

	1:X		**KS**		**Random split (k=3)**	
	Train	Test	Train	Test	Train	Test
Accuracy	0.96	0.97	0.97	0.97	0.96	0.94
Recall	0.97	0.97	0.97	0.97	0.96	0.94
Precision	0.96	0.97	0.98	0.95	0.97	0.97
F1	0.96	0.97	0.97	0.96	0.97	0.95
MCC	0.92	0.95	0.96	0.89	0.93	0.89
Kappa	0.92	0.95	0.95	0.89	0.93	0.89
CV	0.91		0.95		0.95	

BalAcc	0.95	0.95	0.98	0.67	0.97	0.94

Results

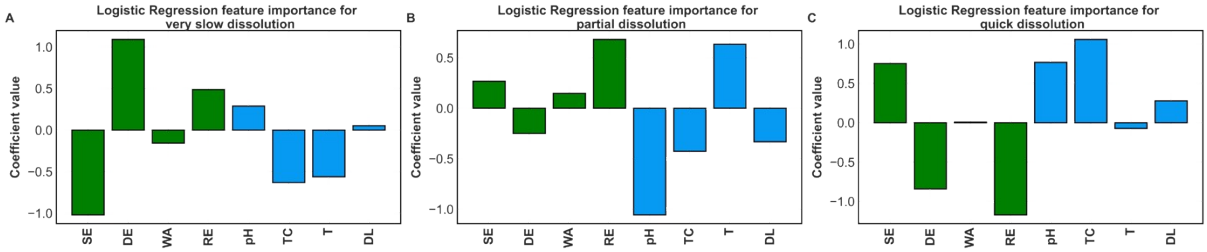

Figure 1. Feature importance for three dissolution classes by Logistic Regression model. A) very slow dissolution, B) partial dissolution, C) quick dissolution.

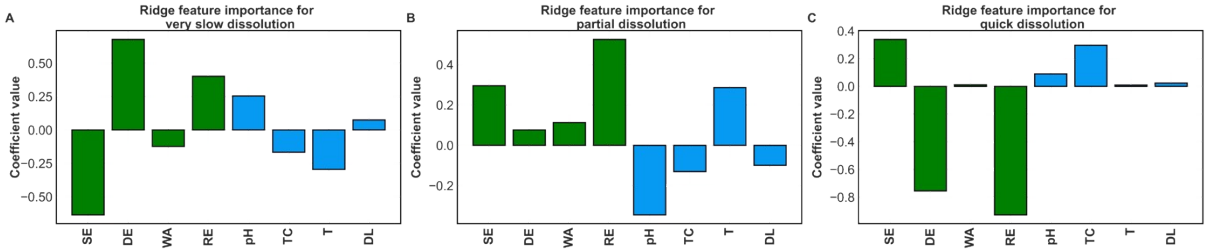

Figure 2. Feature importance for three dissolution classes by the Ridge classification model. A) very slow dissolution, B) partial dissolution, C) quick dissolution.

The models presented demonstrate the highest evaluation metrics subsequent to the SVC model. As illustrated in Figures 1 and 2, the overall trends in model behaviour are consistent for both Logistic Regression and Ridge, in a manner analogous to that observed for SVC. However, a discrepancy emerges with regard to the concentration of total ENMs in the medium.

QMRF 2.1

	Element	Explanation
1.	**QSAR identifier**	
1.1.	QSAR identifier (title)	Dissolution Rate of Engineered Nanomaterials for environmental related waters (Python 3.12, sklearn).
1.2	Other related models	No other related models.
1.3.	Software coding the model	Python 3.12, sklearn, xgboost, shap, lighgbm, pandas.
2.	**General information**	

2.0	Abstract	Dissolution plays a significant role in determining both aspects. However, understanding and predicting the dissolution rate is a complex process influenced by various factors, including the nanoparticles' properties and the surrounding environment's characteristics. This study aimed to develop a novel structure-property relationship (nano-SPR) classification model to predict the dissolution rate of metal and metal oxide ENMs by considering both the nanoparticle properties and the characteristics of the environment. The model assigns dissolution rate to one of three classes, dependently on the way of defining dissolution rate threshold. The developed models exhibited good overall quality, with balanced accuracies ranging above 0.9 depending on the used model type. Through the analysis, we identified several important factors that significantly influence the solubility of studied ENMs. These factors include bond dissociation enthalpy, solvation enthalpy, primary size, valence electrons to core electrons ratio in metal, pH of the medium, presence of light, temperature, and initial concentration of the ENMs. The results provide valuable insights for assessing the environmental transport and fate, predicting (eco)toxicity, and grouping ENMs.
2.1.	Date of QMRF	7 December 2024
2.2.	QMRF author(s) and contact details	Michał Kałapus, mail: michal.kalapus@ug.edu.pl, tel: +48 58 523 52 48 Tomasz Puzyn, mail: tomasz.puzyn@ug.edu.pl, tel: +48 58 523 52 48
2.3.	Date of QMRF update(s)	N/A
2.4.	QMRF update(s)	N/A
2.5.	Model developer(s) and contact details	Michał Kałapus, mail: michal.kalapus@ug.edu.pl, tel: +48 58 523 52 48 Tomasz Puzyn, mail: tomasz.puzyn@ug.edu.pl, tel: +48 58 523 52 48
2.6.	Date of model development and/or publication	2025
2.7.	Reference(s) to main scientific papers and/or software package	1. Cortes, C.; Vapnik, V. Support-Vector Networks. Machine Learning 1995 20:3 1995, 20 (3), 273–297. https://doi.org/10.1007/BF00994018. 2. Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016. 3. Python Software Foundation. Python Language Reference, version 3.12. Available at: https://www.python.org/ 4. Cross, R. K.; Spurgeon, D.; Svendsen, C.; Lahive, E.; Little, S.; von der Kammer, F.; Loosli, F.; Matzke, M.; Fernandes, T. F.; Stone, V.; Peijnenburg, W. J. G. M.; Bleeker, E. A. J. An Integrated Approach to

		Testing and Assessment (IATA) to Support Grouping and Read-across of Nanomaterials in Aquatic Systems. Nano Today 2024, 54, 102065. https://doi.org/10.1016/J.NANTOD.2023.102065.
2.8.	Availability of information about the model	The model is non-proprietary: full description of the model algorithm is available, training and test sets are available as supplementary material of original research article.
2.9.	Availability of another QMRF for exactly the same model	N/A
3	**Defining the endpoint - OECD Principle 1: "A DEFINED ENDPOINT"**	**PRINCIPLE 1: "A DEFINED ENDPOINT". ENDPOINT refers to any physicochemical, biological, or environmental property/activity/effect that can be measured and therefore modelled. The intent of PRINCIPLE 1 (a (Q)SAR should be associated with a defined endpoint) is to ensure clarity in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions. It is therefore important to identify the experimental system and test conditions that is being modelled by the Q)SAR.**
3.1.	Species	N/A
3.2.	Endpoint	Endpoint: dissolution rate (with assumption of first order or pseudo-first order kinetics of dissolution) of Engineered Nanomaterials and Nanoforms in water, unit: $1*s^{-1}$. Endpoint was categorical value of dissolution rate. Three class was established, and they are shown in Table below: Dissolution rate [$1 \cdot s^{-1}$] Half time [hours] Quick $k \geq 2.75 \cdot 10\text{-}5$ $T1/2 \leq 7.2$ Partial $2.75 \cdot 10\text{-}5 > k > 3.8 \cdot 10\text{-}7$ $507 > T1/2 > 7.2$ Very Slow $k \leq 3.8 \cdot 10\text{-}7$ $T1/2 \geq 507$ Predicted endpoint was dissolution rate class.
3.3	Comment on endpoint	Endpoint was defined accordingly to OECD Series on Testing and Assessment. Guidance Document for the Testing of Dissolution and Dispersion Stability of Nanomaterials, and the Use of the Data for Further Environmental Testing and Assessment. Dissolution rate was only gatherd from studies which follow Static Batch Test reccomended by OECD. Dissolution rate was measured in natural waters or waters which simulated natural waters.
3.4.	Endpoint units	$1*s^{-1}$, where s stand for seconds
3.5.	Dependent variable	All dependent variables were autoscaled according to equation: $X_Z = \frac{X - \hat{X}}{S}$, where X_z is transformed value, X orginal value for descriptor, $\hat{X}$ mean value for descriptor and S standard deviation.
3.6.	Experimental protocol	OECD Series on Testing and Assessment. Guidance Document for the Testing of Dissolution and Dispersion Stability of Nanomaterials, and the Use of the Data for Further Environmental Testing and Assessment.
3.7.	Endpoint data quality and variability	Endpoint data quality and variability: provide available information about the experimental test data quality selection and evaluation and include a description of the data quality used to develop the model. This includes provision of information about in terms of the known variability of the test data, i.e. repeatability (variability over time) and reproducibility (variability between laboratories) and sources of error (confounding factors which may influence testing results) etc.. Please also as far as possible provide information about test chemical purity. Ideally, (Q)SARs should be based on experimental tests performed with test chemical of high purity to assure good correlation between structures and

		effect. Test chemical purity should preferably be provided for the individual substances used in the training and validation sets. The data curation procedure and its effect on data quality should also be described here.
4	**Defining the algorithm - OECD Principle 2 : "AN UNAMBIGUOUS ALGORITHM"**	**PRINCIPLE 2: "AN UNAMBIGUOUS ALGORITHM". The (Q)SAR estimate of an endpoint is the result of applying an ALGORITHM to a set of structural parameters which describe the chemical structure. The intent of PRINCIPLE 2 (a (Q)SAR should be associated with an unambiguous algorithm) is to ensure transparency in the model algorithm that generates predictions of an endpoint from information on chemical structure and/or physicochemical properties. In this context, algorithm refers to any mathematical equation, decision rule or output approach.**
4.1.	Type of model	A Support Vector Classifier is described as a data-driven, statistical machine learning model. It uses training data to find an optimal decision boundary—relying on mathematical optimization and statistical principles rather than predefined rules or domain-specific alerts—to classify new examples.
4.2.	Explicit algorithm	Algorithm is presented in manuscript and also it is available in original manuscript: Cortes, C.; Vapnik, V. Support-Vector Networks. Machine Learning 1995 20:3 1995, 20 (3), 273–297. https://doi.org/10.1007/BF00994018.
4.3.	Descriptors in the model	8 descriptors were utilized in developing model. Solvation enthalpy [kcal/mol], bond dissociation enthalpy [kcal/mol], primary size of ENM [nm], ratio of valence to core electron in metal [-], pH, total concentration of ENM in water [mg/L], temperature [C], presence of light (binary) [-]
4.4.	Descriptor selection	The descriptors employed in the models were selected using a meta-transformer (SFM), based on the relative importance of the features (1). Alternative approaches, such as a genetic algorithm and sequential feature selection, yielded inferior results in comparison to those based on feature importance. In order to select the most appropriate descriptors, SFM makes use of an external machine learning algorithm. All methodologies employed to predict the class of dissolution were utilised within the SFM framework. Based on those findings we select four descriptors from both groups, which gives total number of 8 descriptors. The intrinsic descriptors occurring most frequently are: solvation enthalpy (SE), bond dissociation enthalpy (DE), enthalpy of water adsorption on surface (WA), valence to core electron ratio in metal (RE). For extrinsic descriptors se select pH, total concentration (TC), temperature (T) and presence of light (DL). What important, primary size of ENMs which is a only distinctive factor between nanoforms was choose to the final number of descriptors replacing WA.
4.5.	Algorithm and descriptor generation	RE was computed by approach proposed by Kar et al. (2) and next developed by De et al. (3) is based at the initial stage on values directly extracted from the periodic table (1st generation) and derived rest of the descriptors (2nd generation) from previously collected data. The authors of the original paper present 23 descriptors, but some of them apply only to metal oxides, while the data set used for this research also includes metals, so part of them were omitted and in the end, only 9 were used. A list of descriptors with details can be found in Supplementary Information 1. This method offers the advantage of easy determination of descriptors by using only the periodic table.

		The second class of descriptors includes thermodynamic functions related to the dissolution process of metals and metal oxides, the enthalpy of dissociation of the metal-metal and metal-oxygen bond, the enthalpy of solvation of ENM and enthalpy of adsorption water on the surface. To calculate thermodynamic descriptors, we first built molecular models of nanoparticles without coating, then performed geometry optimization using the B3LYP method and the cc-pVTZ basis set for particles composed with elements with atomic numbers less than 36, and the aug-cc-pVTZ-PP - augmented mixed function method along with a pseudopotential, for compounds with atomic number greater than 36. Rest of descriptors are in fact characterization of the medium parameters such as pH, T, TC and DL.
4.6.	Software name and version for descriptor generation	Gaussian 16
4.7.	Chemicals/Descriptors ratio	76/8 (utilized model) 76/20 (all descriptors)
5	**Defining the applicability domain - OECD Principle 3: "A DEFINED DOMAIN OF APPLICABILITY"**	**PRINCIPLE 3: "A DEFINED DOMAIN OF APPLICABILITY". APPLICABILITY DOMAIN refers to the response and chemical structure space in which the model makes predictions with a given reliability. Ideally the applicability domain should express the structural, physicochemical and response space of the model. The CHEMICAL STRUCTURE (x variable) space can be expressed by information on physicochemical properties and/or structural fragments. The RESPONSE (y variable) can be any physicochemical, biological or environmental effect that is being predicted. According to PRINCIPLE 3 a (Q)SAR should be associated with a defined domain of applicability. Section 5 can be repeated (e.g., 5.a, 5.b, 5.c, etc) as many times as necessary if more than one method has been used to assess the applicability domain.**
5.1.	Description of the applicability domain of the model	AD is based on probabilistic boundaries. AD is described in structural feature, descriptors and parameters of medium space. Model is applicable for nano metals and metals oxides. Model is applicable to normal natural waters or simulated natural waters.
5.2.	Method used to assess the applicability domain	Method described by Roy et al. (4) to check whether an object lies within the applicability domain of our model, we assess the descriptors of the training set under the assumption that they ideally follow a normal distribution. Based on the characteristics of the normal distribution, approximately 99.7% of data points are expected to fall within three standard deviations (±3 SD) from the mean. This range, therefore, represents where the majority of the training compounds are located. For any given compound k, if the standardized value for any descriptor i (denoted as Ski) exceeds 3, the compound is considered an X-outlier (if it belongs to the training set) or is outside the AD (if it belongs to the test set) with respect to that descriptor. If some Ski values exceed 3 while others do not, the compound is similar in certain descriptors and dissimilar in others, necessitating a criterion for assessment. To address this scenario, method utilize the standard score corresponding to a cumulative probability of 90%, which is Z=1.28 in a standard normal distribution. Then compute a new

		statistic, Snew, defined as the mean of the Ski values plus 1.28 times their standard deviation. If Snew < 3, there is a 90% probability that the Ski values for that object are below 3. In such cases, the compound is considered not an X-outlier or within the AD.
5.3.	Software name and version for applicability domain assessment	AD has been established based on method describe by Roy et al. The program for determining the domain was prepared by hand in the Python programming language (Python 3.12).
5.4.	Limits of applicability	AD is limited to normal environmental condition. For example, water with pH 2 or 12 is not recognized as normal environmental condition. Also there is limitation in types of nanomaterials, as only metals and metal oxide nanomaterials are consider in AD.
6	**Defining goodness-of-fit and robustness (internal validation) – OECD Principle 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY"**	**PRINCIPLE 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY". PRINCIPLE 4 expresses the need to perform validation to establish the performance of the model. GOODNESS-OF-FIT and ROBUSTNESS refer to the internal model performance.**
6.1.	Availability of the training set	It is available and attached. Training data are available at: **doi do danych**
6.2.	Available information for the training set	Data only consist different nanomaterials and nanoforms in different environmental waters.
6.3.	Data for each descriptor variable for the training set	It is available and attached.
6.4.	Data for the dependent variable for the training set	It is available and attached.
6.5.	Other information about the training set	N/A
6.6.	Pre-processing of data before modelling	Data were autoscaled before analysis.
6.7.	Statistics for goodness-of-fit	Train Acc Precisio Recall F1 Kappa MCC BalAcc DT 0.99 0.99 0.99 0.99 0.97 0.97 0.99 RF 0.97 0.98 0.97 0.97 0.95 0.95 0.97 GB 0.97 0.97 0.97 0.97 0.95 0.95 0.92 ET 0.99 0.99 0.99 0.99 0.97 0.97 0.99 LR 0.96 0.96 0.96 0.96 0.92 0.92 0.90 **SVC 0.96 0.97 0.96 0.96 0.92 0.92 0.96** Ridge 0.97 0.97 0.97 0.97 0.95 0.95 0.92 Ada 0.75 0.94 0.75 0.74 0.55 0.62 0.72 LDA 0.97 0.98 0.97 0.97 0.95 0.95 0.87 Cat 0.96 0.96 0.96 0.96 0.92 0.92 0.85 LGB 0.99 0.99 0.99 0.99 0.97 0.97 0.93 XGB 0.99 0.99 0.99 0.99 0.97 0.97 0.93
6.8.	Robustness - Statistics obtained by leave-one-out cross-validation	N/A

<table>
<tr><td>6.9.</td><td>Robustness - Statistics obtained by leave-many-out cross-validation</td><td>
<table>
<tr><td>Train</td><td>CV</td></tr>
<tr><td>DT</td><td>0.96</td></tr>
<tr><td>RF</td><td>0.94</td></tr>
<tr><td>GB</td><td>0.91</td></tr>
<tr><td>ET</td><td>0.96</td></tr>
<tr><td>LR</td><td>0.92</td></tr>
<tr><td>SVC</td><td>0.95</td></tr>
<tr><td>Ridge</td><td>0.91</td></tr>
<tr><td>Ada</td><td>0.93</td></tr>
<tr><td>LDA</td><td>0.92</td></tr>
<tr><td>Cat</td><td>0.94</td></tr>
<tr><td>LGB</td><td>0.92</td></tr>
<tr><td>XGB</td><td>0.95</td></tr>
</table>
</td></tr>
<tr><td>6.10.</td><td>Robustness - Statistics obtained by Y-scrambling</td><td>100 iterations
Mean accuracy for SVC = 0.34
Mean balanced accuracy for SVC =</td></tr>
<tr><td>6.11.</td><td>Robustness - Statistics obtained by bootstrap</td><td>N/A</td></tr>
<tr><td>6.12.</td><td>Robustness - Statistics obtained by other methods</td><td>Accuracy from Dummy Classifier is equal to 0.4</td></tr>
<tr><td>7</td><td>Defining predictivity (external validation) – OECD Principle 4: “APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY”</td><td>PRINCIPLE 4: “APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY”. PRINCIPLE 4 expresses the need to perform validation to establish the performance of the model. PREDICTIVITY refers to the external model validation. Section 7 can be repeated (e.g., 7.a, 7.b, 7.c, etc) as many times as necessary if more validation studies need to be reported in the QMRF.</td></tr>
<tr><td>7.1.</td><td>Availability of the external validation set</td><td>It is available and attached.</td></tr>
<tr><td>7.2.</td><td>Available information for the external validation set</td><td>Data only consist different nanomaterials and nanoforms in different environmental waters.</td></tr>
<tr><td>7.3.</td><td>Data for each descriptor variable for the external validation set</td><td>It is available and attached.</td></tr>
<tr><td>7.4.</td><td>Data for the dependent variable for the external validation set</td><td>It is available and attached.</td></tr>
<tr><td>7.5.</td><td>Other information about the external validation set</td><td>External validation set with 38 nanomaterials and nanoforms in different environmental conditions.</td></tr>
<tr><td>7.6.</td><td>Experimental design of test set</td><td>By choosing every third (1:X method) NM or NF from whole dataset.</td></tr>
<tr><td>7.7.</td><td>Predictivity - Statistics obtained by external validation</td><td>
<table>
<tr><td>Train</td><td>CV</td></tr>
<tr><td>DT</td><td>0.94</td></tr>
<tr><td>RF</td><td>0.90</td></tr>
<tr><td>GB</td><td>0.95</td></tr>
<tr><td>ET</td><td>0.94</td></tr>
</table>
</td></tr>
</table>

		LR 0.93 **SVC 0.95** Ridge 0.91 Ada 0.90 LDA 0.95 Cat 0.93 LGB 0.95 XGB 0.94
7.8.	Predictivity - Assessment of the external validation set	See table below
7.9.	Comments on the external validation of the model	N/A
8	**Providing a mechanistic interpretation - OECD Principle 5: "A MECHANISTIC INTERPRETATION, IF POSSIBLE"**	**PRINCIPLE 5: "A MECHANISTIC INTERPRETATION, IF POSSIBLE". According to PRINCIPLE 5, a (Q)SAR should be associated with a mechanistic interpretation, if possible.**
8.1.	Mechanistic basis of the model	The fundamental premise was to identify descriptors for the model that would consider both the structural characteristics of the ENM and the properties of the medium that influence the dissolution of nano metals and metal oxides. In order to achieve this, descriptors were prepared based on the enthalpy of processes that are potentially relevant to the kinetics of dissolution. These processes include the enthalpy of solvation and the enthalpy of dissociation of bonds between atoms in the metal and metal oxide crystals. A descriptor indicating the ratio of electrons in the core of the metal atom to valence electrons was employed to distinguish between ENMs, while the original size of the nanomaterial was used to distinguish between nanoforms. Subsequently, medium parameters, including initial concentration, pH, temperature and the presence of light, were considered. Experimental studies have demonstrated that all of these properties exert an influence on the dissolution kinetics of ENMs.
8.2.	A priori or a posteriori mechanistic interpretation	Posteriori, description is available in attached manuscript.

Validation	Accuracy	Precision	Recall	F1	Kappa	MCC	Bal. Acc
DT	0.95	0.95	0.95	0.95	0.89	0.90	0.86
RF	0.92	0.93	0.92	0.91	0.84	0.84	0.75
GB	0.92	0.93	0.92	0.91	0.84	0.84	0.75
ET	0.92	0.93	0.92	0.91	0.84	0.84	0.75
LR	0.95	0.95	0.95	0.95	0.90	0.90	0.88
SVC	**0.97**	**0.97**	**0.97**	**0.97**	**0.95**	**0.95**	**0.97**
Ridge	0.97	0.98	0.97	0.97	0.95	0.95	0.89
Ada	0.87	0.89	0.87	0.87	0.74	0.75	0.77
LDA	0.92	0.93	0.92	0.91	0.84	0.84	0.75
Cat	0.92	0.93	0.92	0.91	0.84	0.84	0.75
LGB	0.92	0.86	0.92	0.89	0.84	0.85	0.67

Descriptors values range is similar in training and validation set. The descriptor and response range for the validation test set is compared with that for the training set.

8.3.	Other information about the mechanistic interpretation	N/A
9	**Miscellaneous information**	
9.1.	Comments	The proposed model can be used in a framework for nano metals and metal oxides in a "OECD Series on Testing and Assessment. Guidance Document for the Testing of Dissolution and Dispersion Stability of Nanomaterials, and the Use of the Data for Further Environmental Testing and Assessment". Model can be also used in IATA proposed by Cross et al. (5)
9.2.	Bibliography	1. Ferri, F. J.; Pudil, P.; Hatef, M.; Kittler, J. Comparative Study of Techniques for Large-Scale Feature Selection. Machine Intelligence and Pattern Recognition 1994, 16 (C), 403–413. https://doi.org/10.1016/B978-0-444-81892-8.50040-7 2. Kar, S.; Gajewicz, A.; Puzyn, T.; Roy, K.; Leszczynski, J. Periodic Table-Based Descriptors to Encode Cytotoxicity Profile of Metal Oxide Nanoparticles: A Mechanistic QSTR Approach. Ecotoxicol Environ Saf 2014, 107, 162–169. https://doi.org/10.1016/J.ECOENV.2014.05.026. 3. De, P.; Kar, S.; Roy, K.; Leszczynski, J. Second Generation Periodic Table-Based Descriptors to Encode Toxicity of Metal Oxide Nanoparticles to Multiple Species: QSTR Modeling for Exploration of Toxicity Mechanisms. Environ Sci Nano 2018, 5 (11), 2742–2760. https://doi.org/10.1039/C8EN00809D. 4. Roy, K.; Kar, S.; Ambure, P. On a Simple Approach for Determining Applicability Domain of QSAR Models. Chemometrics and Intelligent Laboratory Systems 2015, 145, 22–29. https://doi.org/10.1016/J.CHEMOLAB.2015.04.013. 5. Cross, R. K.; Spurgeon, D.; Svendsen, C.; Lahive, E.; Little, S.; von der Kammer, F.; Loosli, F.; Matzke, M.; Fernandes, T. F.; Stone, V.; Peijnenburg, W. J. G. M.; Bleeker, E. A. J. An Integrated Approach to Testing and Assessment (IATA) to Support Grouping and Read-across of Nanomaterials in Aquatic Systems. Nano Today 2024, 54, 102065. https://doi.org/10.1016/J.NANTOD.2023.102065.
9.3	Supporting information	All data use in developed models are available and attached to this work. Training and test set are submitted in .xlsx format. Predictions are included in the Supplementary Information.