Zetapotential readacross model utilizing nanodescriptors extracted via the nanoxtract image analysis tool available on the Enalos Nanoinformatics Cloud platform

Lynch, Iseult; Valsami-Jones, Eva; Papadiamantis, Tassos

DOI:
10.1002/smll.201906588

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Lynch, I, Valsami-Jones, E & Papadiamantis, T 2020, 'Zetapotential readacross model utilizing nanodescriptors extracted via the nanoxtract image analysis tool available on the Enalos Nanoinformatics Cloud platform', *Small*. https://doi.org/10.1002/smll.201906588

NANO · MICRO small
www.small-journal.com

Zeta-Potential Read-Across Model Utilizing Nanodescriptors Extracted via the NanoXtract Image Analysis Tool Available on the Enalos Nanoinformatics Cloud Platform

Dimitra-Danai Varsou, Antreas Afantitis, Andreas Tsoumanis, Anastasios Papadiamantis, Eugenia Valsami-Jones, Iseult Lynch, and Georgia Melagraki**

Zeta potential is one of the most critical properties of nanomaterials (NMs) which provides an estimation of the surface charge, and therefore electrostatic stability in medium and, in practical terms, influences the NM's tendency to form agglomerates and to interact with cellular membranes. This paper describes a robust and accurate read-across model to predict NM zeta potential utilizing as the input data a set of image descriptors derived from transmission electron microscopy (TEM) images of the NMs. The image descriptors are calculated using NanoXtract (http://enaloscloud.novamechanics.com/EnalosWebApps/NanoXtract/), a unique online tool that generates 18 image descriptors from the TEM images, which can then be explored by modeling to identify those most predictive of NM behavior and biological effects. NM TEM images are used to develop a model for prediction of zeta potential based on grouping of the NMs according to their nearest neighbors. The model provides interesting insights regarding the most important similarity features between NMs—in addition to core composition the main elongation emerged, which links to key drivers of NM toxicity such as aspect ratio. Both the NanoXtract image analysis tool and the validated model for zeta potential (http://enaloscloud.novamechanics.com/EnalosWebApps/ZetaPotential/) are freely available online through the Enalos Nanoinformatics platform.

1. Introduction

The unique properties of nanomaterials (NMs) in comparison to their bulk counterparts have led to their extensive use in a wide range of applications and consumer products.[1,2] In parallel to their widespread use, concerns arise regarding possible risks from NMs for human health and the environment including cytotoxic effects, oxidative stress, genotoxicity, cell apoptosis, accumulation in different organs, Trojan horse effects, etc.[3–5] To assess these risks, a robust NMs risk assessment framework is highly desired to reveal, among others, the relationship between NMs physicochemical properties and their diverse effects on cells, organisms, and eventually humans. Toward this goal, NMs characterization and experimental biological and environmental testing (release from products, exposure, and effects) is performed.

NMs characterization immediately following synthesis is routinely performed to confirm that the desired product was produced. The gold-standard characterization method is transmission electron microscopy (TEM) whereby images are obtained to determine quantitatively the size (by measuring large numbers of particles) and qualitatively the shape of the NMs and other morphological characteristics.[6] Typically, 500 individual NMs are assessed across several images to get robust statistics in terms of the particle size and size distribution. Dudkiewicz et al.[7]

D.-D. Varsou Dr. A. Afantitis, Dr. A. Tsoumanis, Dr. G. Melagraki
Nanoinformatics Department
NovaMechanics Ltd.
Nicosia 1065, Cyprus
E-mail: afantitis@novamechanics.com; melagraki@novamechanics.com

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smll.201906588.

DOI: 10.1002/smll.201906588

D.-D. Varsou
School of Chemical Engineering
National Technical University of Athens
Athens 15780, Greece

Dr. A. Papadiamantis, Prof. E. Valsami-Jones, Prof. I. Lynch
School of Geography
Earth and Environmental Sciences
University of Birmingham
B152TT Birmingham, UK

have pointed out that the number of particles that need to be counted depends on the particle size distribution, although the contribution of matrix effects to measurement uncertainty is much higher.

TEM images are an invaluable source of information from an in silico point of view, since a range of different variables, the so-called image descriptors, can be extracted from a given NM TEM image with the use of appropriate software. These image descriptors can then be used to develop meaningful correlations with an activity or property of interest. Past examples of such efforts, where image descriptors have been calculated and then used to derive quantitative nanostructure–activity/property relationships, have been reported in the literature.[8,9] However, these had not been integrated with a web service to allow widespread utilization by the nanosafety and NMs synthesis and modeling communities, nor integrated with modeling approaches that allow direct integration of the calculated image descriptors into predictive models. This can be easily facilitated by a web service dedicated to the calculation of such image descriptors for NMs, as demonstrated here. To date the existing online software for NMs is in a preliminary stage and thus we present here the design, development, and implementation of a web service that would allow, in a minimum number of steps, the online extraction of image descriptors from a given NM TEM image that can be directly uploaded by the user.

Beyond the characterization of the pristine NMs, it is also vital to understand their behavior under the relevant exposure conditions, e.g., dispersed in the relevant Organisation for Economic Co-operation and Development (OECD) or other test medium. Here, important aspects that are characterized include the NM zeta potential, which provides an estimation of the surface charge, and therefore electrostatic stability, in medium and, in practical terms, controls the NM's tendency to form agglomerates.[8,10] Thus, zeta-potential values of ±30 mV are often considered to denote good electrostatic stabilization.[11] Note, however, that many NMs may also have a strong contribution from steric stabilization, so close to neutral zeta potentials cannot always be used to predict instability and agglomeration potential.[12] Thus, models extracting descriptors from TEM images, usually prepared in simple medium, need to have input information on the surface charge and the type of capping (e.g., small molecules such as citrate) or coating (typically larger polymers such as PEG, PVP, etc.) molecules present on the NMs.

As experimental approaches are often costly and time consuming, computational approaches can provide significant aid to prioritization of NMs for experimental testing, and indeed to prioritize which characterization endpoints, from the quite extensive lists of minimal characterization needs,[13] are most useful for correlating with toxicity and other biological or environmental effects.[14] Predictive models based on experimentally measured or theoretically calculated descriptors that encode NMs structural characteristics, can be built to predict a property or activity of interest.[3,15,16] However, only a few such models and tools have been proposed to date in the nanoinformatics field for the prediction of properties or activities of NMs.[8,16–21]

One promising approach is to develop computational tools that extract additional information from existing experimental datasets, i.e., to enrich the experimental datasets with computationally determined descriptors, thus maximizing the utility of the experimental datasets. For example, scanning electron microscopy (SEM) or TEM images are currently utilized by experimentalists to determine size and size distribution (by counting particles manually or automatically in ImageJ[22]), and occasionally for characterization of shape, aspect ratio, or other morphological parameters, although there are no agreed methodologies or reporting conventions for these, and it is very challenging to extract NMs morphological data directly from such images. Tools and workflows for extracting image descriptors from high-throughput fluorescence images of cells, such as global intensity level, cell count, cell shape, cellular and subcellular constellations, colocalization information, etc., already exist in the field of biology[23] and can also be used for the analysis of NMs microscopy images with only a few modifications. Currently, many efforts for the extraction of image analysis descriptors have been presented, including the open-source ImageJ[24] tool and similar tools like Fiji,[4] or implemented in programming environments like MATLAB,[4,6] to mathematically describe and "quantify" the different shapes of the NMs and utilize these image descriptors as additional input information for predictive models of NMs toxicity.[4,25]

In the present work, we first developed **NanoXtract**, an automated online tool for the extraction of NM image descriptors from TEM images to enrich the value of information available from the images, and in a second step to integrate this tool within a novel nanoinformatics workflow to develop a fully validated predictive model using these new computational descriptors for the prediction of NMs properties, such as zeta potential, based only on the calculated descriptors from the TEM images. A discussion of the selected descriptors is included to highlight the influence of each of the descriptors on the selected property (i.e., zeta potential in this case). Both the **NanoXtract** image analysis tool and the validated predictive model for zeta potential were made available as web applications through the Enalos nanoinformatics Cloud Platform.

2. Experimental Section

2.1. NanoXtract: Enalos Image Descriptors Web Service

In this work, we have designed, developed, and implemented an online tool, based on ImageJ features,[26] with the ability to process any given NM image of interest (typically a TEM image) to extract image descriptors for NMs. This tool, called **NanoXtract**, is available online: http://enaloscloud.novamechanics.com/EnalosWebApps/NanoXtract/.

Within a user-friendly interface, the TEM image is uploaded by a user, and the analysis is completed in four steps for parameter tuning (e.g., indication of the scale from the TEM image, removal of any NMs partially outside the field of view, etc.). As the output, a set of 18 image descriptors are calculated and displayed on the screen and can be downloaded as a .csv file. More details on the web interface and the tuning steps are given in the Results and Discussion section and in the **NanoXtract** tutorial included in the Supporting Information.

2.2. Nanomaterials Characteristics

In this study, TEM images from 37 diverse NMs, selected and characterized extensively in the EC-funded project NanoMILE,[27] were analyzed. Available data include information on the NM core composition and the experimental zeta-potential values (Table S1, Supporting Information). Zeta potential was used in our demonstration of the predictive model as the analysis endpoint since it is easy to obtain experimentally (although it is noteworthy that the meaningfulness of zeta-potential measurements on nonspherical particles is debatable, given that the underpinning mathematical assumptions assume spherical particles),[10,28] and thus can be utilized to validate the overall robustness and predictive power of the model. Being able to predict zeta potentials for nonspherical NMs from TEM images would thus improve the meaningfulness of this datapoint for prediction of NMs behavior also. Moreover, during the characterization process, TEM images were taken to observe the NMs morphology and measure their size characteristics (e.g., diameter). These TEM images were first used to extract quantified information and derive useful nanodescriptors for the prediction of zeta-potential values for untested NMs.

All available characterization data were selected to develop our modeling workflow. NMs included in the dataset have different core compositions (pure metal/metal oxide), coatings (uncoated/anionic/cationic/neutral coatings), and shapes (circular for spheres, rods, and plates). In **Figure 1**, information about the dataset classification is presented. The full dataset and representative TEM images for all 37 NMs are included in Table S1 and Figure S1 (Supporting Information).

2.3. Nanoinformatics Workflow Development

To initiate the study, the Konstanz Information Miner (KNIME) platform was used, which is a user-friendly and open source software for data integration and analysis.[17] The KNIME platform enables the user to create visual data flows consisting of nodes and connections between them. The nodes represent concrete steps of the data processing and can be selectively or entirely executed, giving users the flexibility to experiment easily between different methodologies, compare the results among different tuning parameters,[29] as well as to employ modules and tools from different analysis suites (e.g., CDK, RDKit, ChEMBL, WEKA, Enalos+, etc.). Especially for image analysis, KNIME made it possible to read images and apply preprocessing, segmentation, feature extraction and classification methods on NM TEM images, and—in addition to the available relevant KNIME nodes—it offers a selection of different image analysis tools (e.g., ImageJ, CellProfiler, etc.).[30]

In the present work, the KNIME platform was used to perform all the various components of our nanoinformatics analysis under a common interface, including image processing, model development and validation with the aid of our in-house Enalos+ KNIME nodes.[31] We employed different nodes for preprocessing and segmentation of microscopy images, feature extraction, and modeling. In this way, we ensured a systematic approach for the analysis and a complete supervision of the workflow. Users can apply some or all the steps, i.e., can stop after extraction of data from their TEM image, or can proceed to modeling.

2.4. Image Descriptors Calculation

Effort focused initially on the extraction of quantitative information from the NMs 2D projection in the TEM microscopy images that could serve as nanodescriptors in a predictive modeling framework. A key step of the workflow development was to extract image descriptors from the TEM images of NMs and to provide this procedure as a web service to facilitate future needs within nanoinformatics modeling.

In a later stage, the development of significant correlations among the calculated image nanodescriptors and the experimentally determined zeta potential,[32] which has been demonstrated

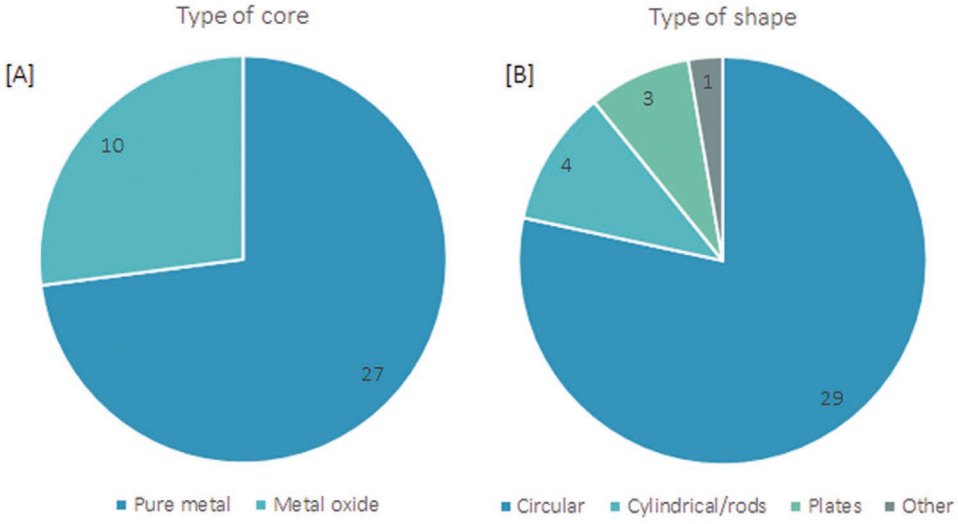

Figure 1. Dataset classification. A) Different cores of the NMs in the sample. B) Different shapes of the NMs in the sample.

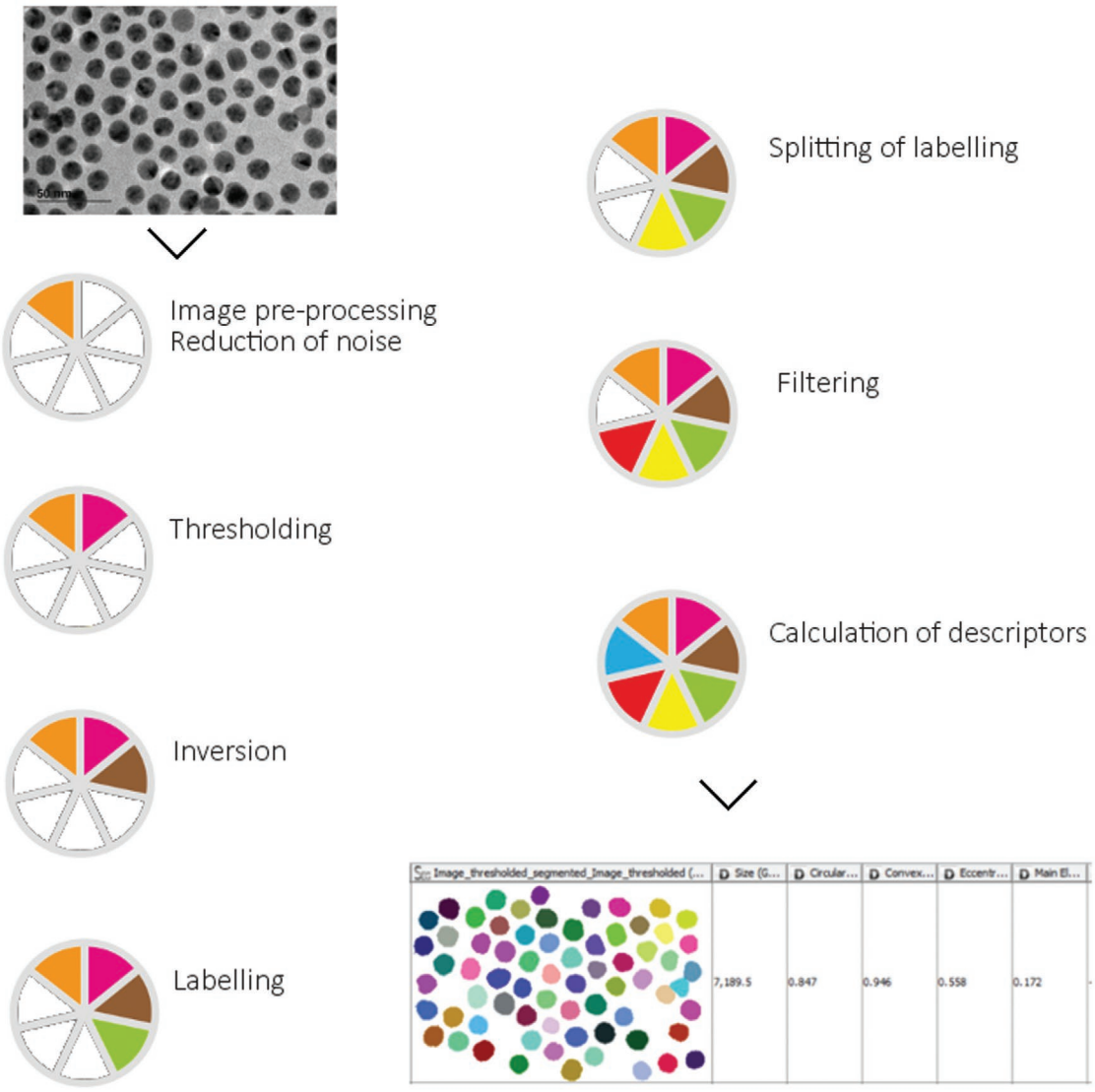

Figure 2. Schematic workflow of the image processing steps.

previously to show some predictivity for NM cytotoxicity,[8] was explored. The extraction of image nanodescriptors can be a complex process due to the wide variety of microscopy images with varying resolution, mixed sizes and shapes, as well as the agglomeration and aggregation of the NMs depicted within a TEM image, which can either be a drying artifact or indicative of the presence of agglomerates/aggregates in the sample prior to deposition on the TEM grid.[33] In this work, a methodical KNIME nanoinformatics workflow was built for the automated and time-effective extraction of nanodescriptors from a dataset of 68 TEM microscopy images of 37 NMs. We need to underline that the same KNIME workflow was used for the extraction of the descriptors from all available images, despite the different shapes of NMs that existed in our dataset, and indeed the same approach is equally applicable for complex NM structures, such as sea-urchin-shaped NMs that result from NM ageing in the environment (see Figure S1, Supporting Information, for representative examples). A schematic workflow of the image processing steps that are performed "behind" the **NanoXtract** interface is presented in **Figure 2**. **NanoXtract** was developed entirely based on the KNIME platform.

After importing image files into the KNIME platform, a preprocessing step was performed in order to improve image quality and reduce noise. For this purpose, the Gaussian blur was used; the details (noise) were eliminated and the depicted components (particles) were enhanced. In a later step, images were converted to binary images, where their background and their foreground were totally separated based on the grayscale distribution of the initial image and a predefined threshold.[23] In that way, the NMs were defined as groups of pixels that form meaningful components and were separated from the medium matrix (note that this can also be applied to more complex media, containing proteins, salts, etc., indicating that the approach is also suitable for NMs dispersed in biological or environmental medium). The images were inverted in order to proceed with the labeling of the connected components (segments derived by the thresholding process). If needed, the

Table 1. Calculated image descriptors obtained from the KNIME workflow, and a brief explanation of their physical meanings, value ranges, and units.[35–37]

Image descriptor	Brief meaning	Range of values
Area	The area of the NM	>0 [nm^2]
Boundary size	Total length of the NM's boundary (the perimeter calculated by a different method)	>0 [nm]
Boxivity	The extent to which a NM approaches a rectangle	0–1 [unitless]
Circularity[a)]	The degree to which a NM approaches a perfect circle	0–1 [unitless]
Convexity[b)]	The NM's edge roughness	0–1 [unitless]
Diameter	The NM's diameter	>0 [nm]
Eccentricity	The measure of how much the NM deviates from being circular	0–1 [unitless]
Extent	The boxivity calculated using different method	0–1 [unitless]
Main elongation	The lengthening of the NM	0–1 [unitless]
Major axis	The longest diameter of the best fitting ellipse to the NM	>0 [nm]
Maximum Feret's diameter	The longest distance between any two points along the selection boundary (caliper diameter)	>0 [nm]
Minimum Feret's diameter	The shortest distance between any two points along the selection boundary	>0 [nm]
Minor axis	The shortest diameter of the best fitting ellipse to the NM	>0 [nm]
Perimeter	Total length of the NM's boundary	>0 [nm]
Roundness	Compares the surface of the NM to the surface of the disc of diameter equal to the major axis	0–1 [unitless]
Solidity	The degree of the overall concavity or convexity of a NM	0–1 [unitless]

[a)] **NanoXtract** produces two circularity values, calculated by different KNIME nodes; [b)] **NanoXtract** produces two convexity values, calculated by different KNIME nodes.

segmentation was improved using the *Waehlby Clump Cell Splitter* node that splits and merges by shape the objects within a specific label (enumerated segment). In cases where segmentation was unsatisfactory (e.g., large segments that represented large aggregates still existed or background noise produced small segments), a minimum and maximum pixel-size cutoff were set that excluded the outlier segments. After satisfactory segmentation, 18 image nanodescriptors were calculated for all segments (**Table 1**) and their mean values and their standard deviations were the final descriptors produced by the image analysis workflow. The user can have complete supervision and can "interfere" in this workflow by tuning the parameters for the thresholding and the filtering.

The calculated image nanodescriptors had substantially different numerical ranges, and therefore, in order to force them to contribute equally to the rest of the analysis, we performed Gaussian (z-score) normalization.[34]

Each descriptor can give specific information about the NMs, however their combination can also give information about a dataset of images, especially when NMs of the same composition exhibit different behaviors; for example, if the geometric proportion is a factor that controls a specific behavior, by comparing the values of circularity and boxivity, we can have an initial picture of the proportion of spherical and cubic particles in the dataset. An emerging concept in nanomedicine is the so-called off-target effect, whereby a percentage of the dose never reaches the target site but accumulates elsewhere—detailed descriptors such as the ones described here might also be able, in the future, to provide insights as to which subsets of a NM batch size/shape distribution are likely to have off-target effects, allowing synthesis or separation procedures to remove these "rogue" NMs. These factors could also be utilized in overall NMs production quality control procedures, for analysis of batch-to-batch variability,[38] and potentially to help identify the sources of such variability through analysis of changes in the various ratios as synthesis parameters are varied.[38]

2.5. Variable Selection

The set of image descriptors was augmented by two additional descriptors, which are relevant to the key intrinsic properties of a NM, namely, the type of core (pure metal or metal oxide) and the pH where the zeta potential was measured. These two descriptors were part of the full physicochemical characterization dataset of the NMs used.

During the modeling process and for validation purposes, multiple splitting of the initial dataset was carried out in order to eliminate the possible influence of the splitting on the modeling results. In this process, Best First (*BestFirst*) variable selection was performed using the CFS Subset Evaluator (*CfsSubsetEval*) included in WEKA, to select the attributes that are the most relevant to the specific endpoint (zeta potential in our case). *CfsSubsetEval* is a correlation-based attribute subset evaluator that takes into account subsets of uncorrelated features but that are highly correlated with the predicted endpoint. *BestFirst* search method searches the space of attribute subsets by greedy hill-climbing with backtracking facility.[39,40] By applying variable selection, noisy attributes were excluded, the algorithm's performance was greatly improved and overfitting of the model to the dataset was avoided.[41]

2.6. *k*-Nearest Neighbors (*k*NN)/Read-Across Model Development

After the extraction and normalization of the image nanodescriptors, and the variable selection process, predictive modeling was applied, with the aim of correlating experimental data to the calculated image and structural descriptors. Zeta potential was selected as the comparative set of experimental data. This was for practical purposes, as zeta potential is easy to measure, but also since previous work has shown this property to correlate with NM behavior and cytotoxicity. To develop the model, the *k*NN method was incorporated into KNIME (via the *EnaloskNN* KNIME node)[42] that was, among the tested methodologies, the one that produced the best correlation between input variables and the endpoint. An additional advantage of

the *k*NN method is that it is actually a read-across strategy,[43] as it requires experimental observations of only a few neighbors (similar NMs) to the query NM, in order to compute the end-point prediction. The *k*NN methodology is a lazy learning technique that classifies an instance based on the majority vote of the k closest training examples (neighbors). Here, since the endpoint had a numeric class, the prediction was the distance weighted average of the endpoint of the selected neighbors. An optimal k value was selected based on the calculated Euclidean distance between all instances and used as weighting factors of the inversed distance.[39,41] Another important aspect of the analysis—apart from the simple endpoint prediction—was to clearly define and present the groups of k neighbors of each test NM, and therefore to specify and map the analogous space, which is a prerequisite of the read-across framework.[44]

2.6.1. Model Validation

In order to fully validate the proposed model, external validation was performed by separating the initial dataset into training and test sets, with the test set left out of modeling and used subsequently for validation purposes. Nevertheless, as the initial dataset was limited, we performed multiple divisions in order to eliminate the possible bias of the splitting on the predictive accuracy. Various random—but stratified—splits were performed, keeping the previous proportion between training and test sets, and the results of modeling based on the different training sets were compared to each other, until it was ensured that the model is sufficiently robust.

To evaluate the models' performance, the goodness-of-fit on the test data was measured, using the coefficient of multiple determination (R^2_{pred}, Equation (1))[45] and the following statistical indices, as proposed by Tropsha, were used to assess the predictive power of regression predictive models (Equations (2)–(4))[46]

$$R^2_{\text{pred}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{1}$$

$$R^2_{\text{ext}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_{\text{tr}})^2} \tag{2}$$

$$k = \frac{\sum_{i=1}^{n} y_i \tilde{y}_i}{\sum_{i=1}^{n} \tilde{y}_i^2} \tag{3a}$$

$$k' = \frac{\sum_{i=1}^{n} y_i \tilde{y}_i}{\sum_{i=1}^{n} y_i^2} \tag{3b}$$

$$R^2_{\text{o}} = 1 - \frac{\sum_{i=1}^{n}\left(\tilde{y}_i - \tilde{y}_i^{\text{ro}}\right)^2}{\sum_{i=1}^{n}\left(\tilde{y}_i - \overline{\tilde{y}}_i\right)^2},\ \tilde{y}_i^{\text{ro}} = k y_i \tag{4a}$$

$$R'^2_{\text{o}} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \tilde{y}_i^{\text{ro}}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \tag{4b}$$

where $\tilde{y}_i$ is the predicted endpoint value for the i^{th} NM, y_i is the experimental (observed) endpoint value for the i^{th} NM, $\overline{y}_i$ is the average value of the endpoint in the test set, $\overline{y}_{\text{tr}}$ is the average value of the endpoint in the training set, n is the number of NMs that constitute the test dataset, and $\overline{\tilde{y}}_i$ is the average over all $\tilde{y}_i$ ($i = 1, \ldots, n$).

A further set of conditions were also tested, in accordance with the approach of Tropsha et al.[46,47] who considered that a regression model is predictive if the following conditions (Equation (5)–(8)) are satisfied

$$R^2_{\text{pred}} > 0.6 \tag{5}$$

$$R^2_{\text{ext}} > 0.5 \tag{6}$$

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \text{ or } \frac{R^2 - R_{\text{o}}'^2}{R^2} < 0.1 \tag{7}$$

$$0.85 < k < 1.15 \text{ or } 0.85 < k' < 1.15 \tag{8}$$

Finally, as an additional test of the robustness of the proposed model, the Y-randomization test (Y-scrambling) was performed. In this technique, all modeling procedures are repeated several times using the original values of the independent variables, but scrambled values for the endpoint variable. In case that these new models have statistically lower predictive power than the models built with the original endpoints, then the initial models are considered reliable, because the possibility of chance correlation is eliminated.[46]

2.6.2. Domain of Applicability

An important step after the proposition of a validated model is to ensure the necessary conditions in order to strengthen users' confidence in, and acceptance of, the model, and its practical use. The above step can be achieved by defining the domain of applicability of the model and therefore allow its use in real-life applications.[48] When the model is used to make predictions of the properties of novel NMs, for example, in a safety-by-design framework, it is crucial to denote whether these predictions can be considered reliable or not; this can be achieved by comparing the relative "position" of the NMs to the space limits framed by the NMs of the training set.

Considering that in this study a local (read-across) methodology was applied, the domain of applicability could not be defined using all the samples of the training set; therefore, for each query NM, the domain of applicability was defined using similarity measurements based on the Euclidean distance among the k selected neighbors of the training set and the test NM. The distance of a test NM to its nearest neighbor in the training set is compared to the predefined applicability domain (PAD) threshold (Equation (9)) and its prediction is considered unreliable if the distance exceeds this PAD limit.[17,29,39] The assessment of the domain of applicability of the proposed model was introduced into a KNIME workflow, using the *Enalos+ Domain–AD* KNIME node that executes the following procedure[49,50]

$$\mathrm{PAD} = d + Z\sigma \tag{9}$$

Euclidean distances (D) between all samples in the training set are calculated. d and σ are the mean value and the standard deviation, respectively, of the distances between members of the training subset that includes samples with a distance lower than the mean value of distances (D). Z, is an empirical cutoff value (in this case was set equal to 0.5).

3. Enalos Cloud Platform

A crucial step that should be performed after the development of a predictive model, is to disseminate the model's results and facilitate its use by all interested stakeholders in real-life applications. It is also important to simplify the different technical aspects of the model and present its important features and results through a user-friendly environment, especially designed for noninformatics experts. The Enalos Cloud platform,[51] developed by NovaMechanics Ltd., is an online, freely available toxicity and drug discovery platform that hosts predictive models released as web services, which aim to address the need to reduce the amount of time and cost spent on experimental testing during the drug discovery and risk assessment procedures for small molecules and NMs. Several predictive models, based on open source and in house algorithms and software, are already available within the Enalos Cloud platform, including models for NMs toxicity, biological activity, and properties evaluation.[51]

4. Results and Discussion

One of the main purposes of the present study was to build a framework for extracting and assessing nanodescriptors from TEM images of NMs that could be used to build validated models for the prediction of NM properties (and eventually toxicity and/or adverse outcomes).

Toward this goal, the Enalos Image Nanodescriptors web service (**NanoXtract**) was built that allows the calculation of a set of meaningful image descriptors for further in silico exploitation. Additionally, recommendations regarding some easy checks for image quality are included in the user notes accompanying the web service.

The steps required for this are as follows: First, a TEM NM image with an embedded scale bar is uploaded as a .png or .tiff file as shown in **Figure 3**. The .png format allows not only to use raw TEM image files, but also to use images extracted from publications using, for instance, a snipping tool.

When uploaded, the image is shown in the upper right part of the page. The user then activates the capture tool in order to draw a line on the scale bar of the image (to calculate the number of pixels) and provides as input in the corresponding box the length of the scalebar in nm. The user also selects from

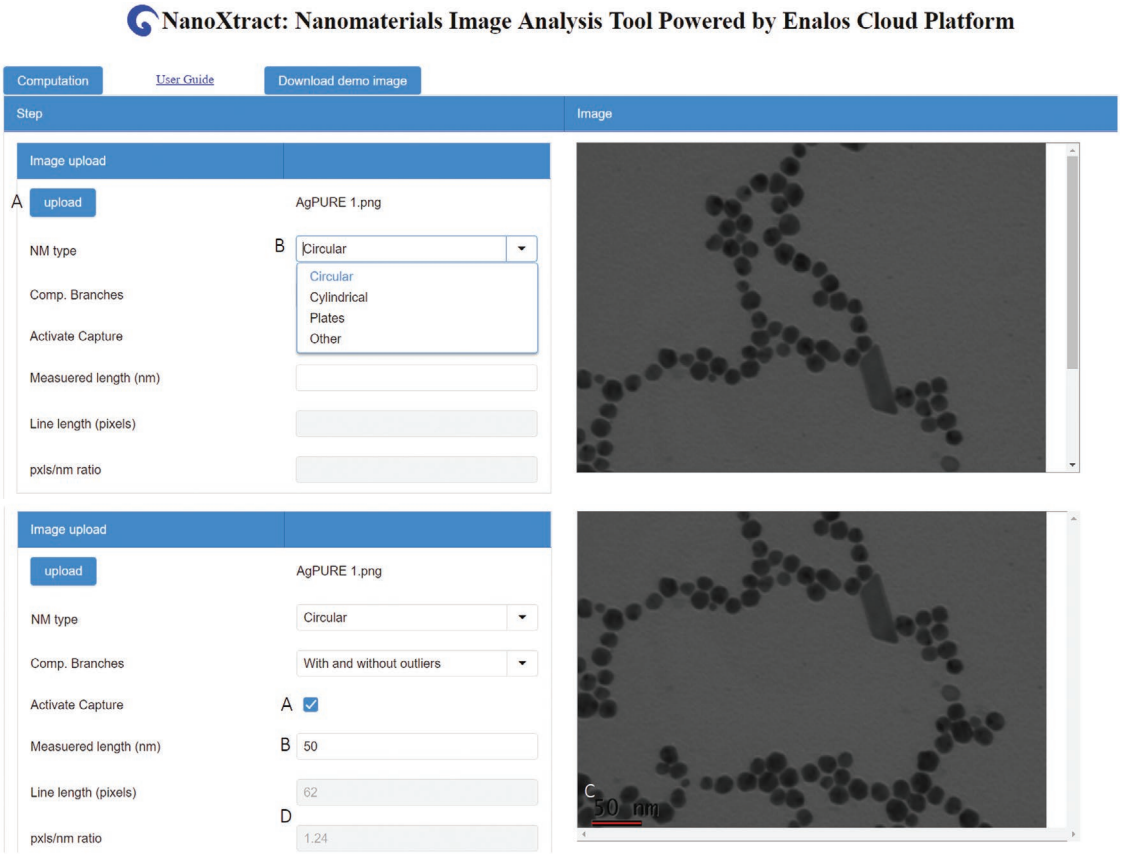

Figure 3. Screenshot of TEM image upload, parameter setting, and submission.

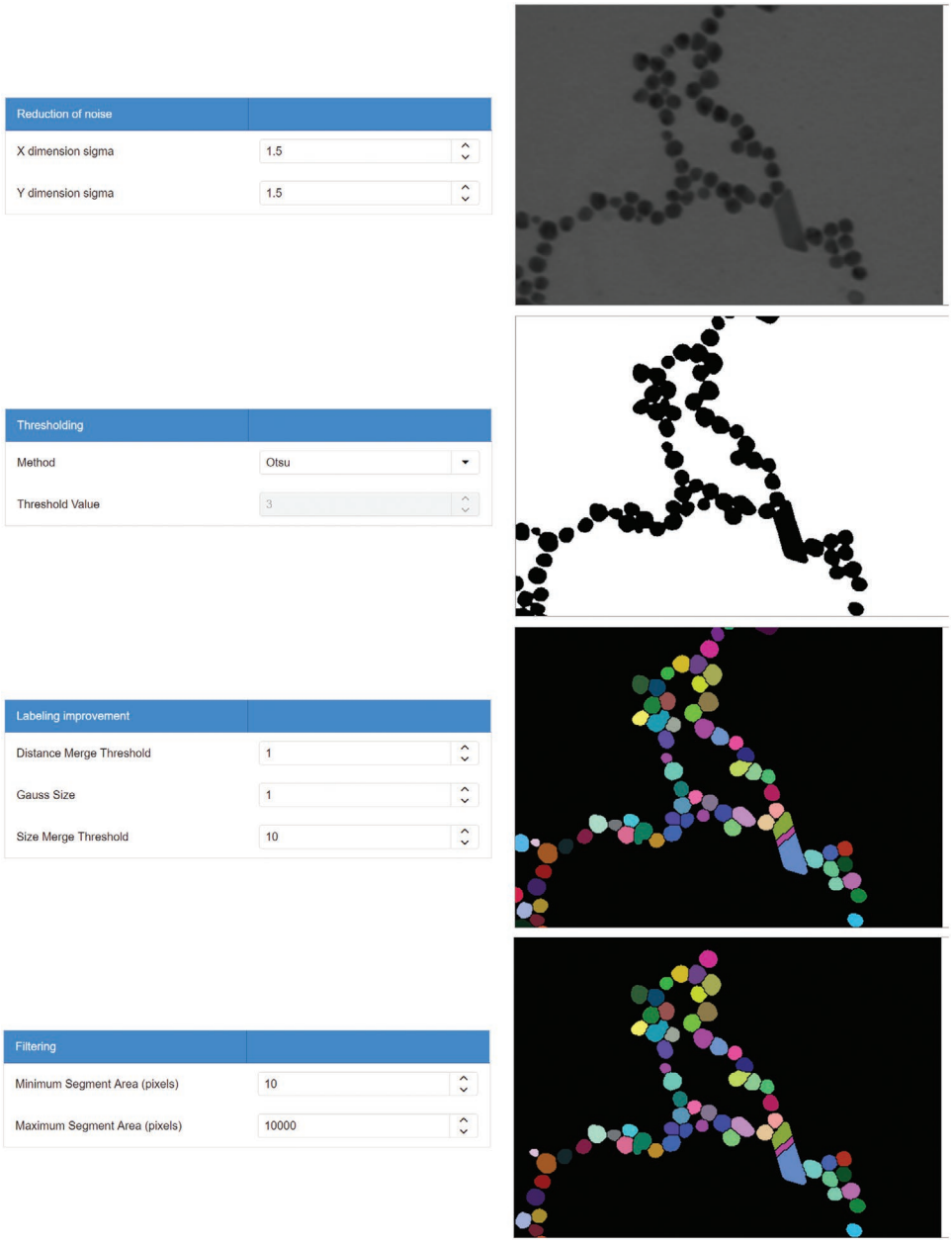

Figure 4. Screenshot of the four steps for parameter tuning (reduction of noise, thresholding, labeling improvement, filtering) with the default values shown. Users can tune these values, as per their needs and expertise level.

the NM type dropdown menu the shape (circular/cylindrical/plates/other) of the depicted particles in the uploaded image.

Four subsequent steps are included for extra parameter tuning (reduction of noise, thresholding, labeling improvement, filtering) if desired by the user, or alternatively the default parameters are applied. In each step, the processed image appears on the right side for visualization of the results, as shown in **Figure 4**.

The computation button is used to submit the processed image and the results appear at the end of the page, as shown in **Figure 5**. The 18 image nanodescriptors corresponding to the submitted image file can be downloaded in a .csv file for further exploitation as shown in **Figure 6**. Both descriptors per NM and mean values are provided. For the subsequent illustration, the mean values file is used, as a single endpoint per sample (and not for each individual NM) is typical from many (ensemble) experimental measurements.

The NM image descriptor calculation tool presented above is available as a web service to facilitate the extraction of useful information from NM TEM images that could subsequently be correlated with a biological endpoint or a physicochemical characteristic of interest. This tool will be one of the first online tools available to extract image descriptors from NM TEM images that could then be explored within an in silico workflow to develop robust and predictive models for NM properties and biological effects.

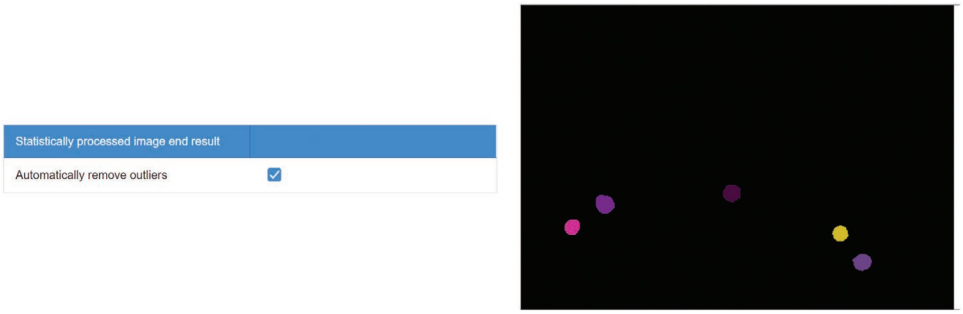

Figure 5. Screenshot of the final statistically processed image, utilized for the computation of the nanodescriptors. Further statistical filtering may be applied.

An important goal within this study was to integrate all the required steps for predictive model development (including image analysis, variable selection, modeling, and validation) based solely on open source and freely available software (KNIME, WEKA, etc.) and in-house proprietary software (Enalos+) and then release the validated nanoinformatics workflow as an open web service with a user-friendly interface. For this purpose, the KNIME Analytics platform was employed, which also gave the flexibility to bridge different techniques, experiment between different scenarios, and have overall supervision of the analysis. The workflow is integrated within the Enalos Cloud Platform to support its further uptake and adoption by other modelers/tool developers.

The initial dataset consisted of a set of 68 TEM microscopy images of 37 NMs that included different cores and coatings and had measured experimental values for their zeta potentials. An important aspect in NMs toxicity assessment is the study of their extrinsic properties, such as the agglomeration of the NMs under certain conditions, taking into consideration that a large agglomerate of NMs may dissociate or break up in a cellular environment and later release smaller (and potentially more bioavailable) particles in the body.[8,52] The agglomeration phenomena are greatly affected by the surface charge of NMs, as encoded by the zeta-potential index; high zeta-potential values either negative or positive, produce stable NM suspensions, whereas NMs with low zeta-potential values tend to form agglomerates in the absence of either steric stabilization resulting from polymer coatings or association of biomolecules with the NMs via the formation of an acquired corona.[53] Thus, the zeta-potential index is a critical factor in NMs characterization and the study or prediction of their toxicity[8,54,55] and, as such, this property was used as the endpoint of the analysis.

Descriptors mean values | Descriptors per component

Description	Values	Standard Deviation
row ID	Mean	Std. deviation
Circularity	0.815304	0.026804
Perimeter [nm]	65.854691	5.372704
Convexity	0.969520	0.011609
Extend	0.751259	0.051784
Diameter [nm]	21.507682	1.847073
Size [nm2]	260.634622	45.008274
Circularity #2	0.860098	0.031631
Convexity #2	0.963618	0.009613
Eccentricity	0.384856	0.098693
Main Elongation	0.037501	0.035473
Minimum Ferets Diameter [nm]	17.694778	1.592952
Maximum Ferets Diameter [nm]	19.977970	1.852773
Major Axis [nm]	18.970852	1.743344
Minor Axis [nm]	17.390103	1.439849
Boundary Size [nm]	61.549123	5.382984
Boxivity	0.785095	0.029089
Roundness	0.917831	0.036212
Solidity	0.966471	0.012559

Figure 6. Screenshot of the output (mean values) of the image nanodescriptors calculated from the processed TEM image.

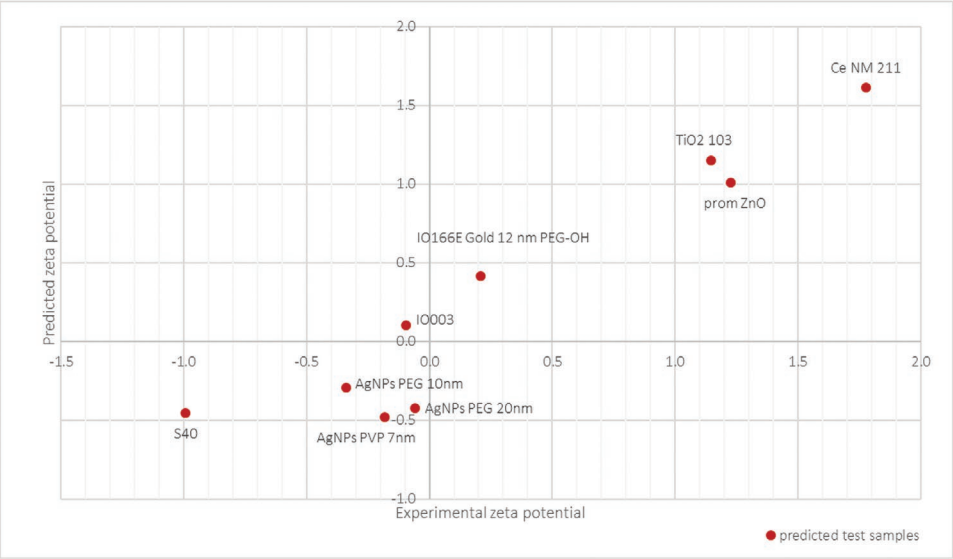

Figure 7. Predicted zeta-potential values (normalized) using the proposed model for the test set.

From the corresponding images of each NM, 18 image descriptors were extracted that describe shape and geometrical characteristics of the particles, using a process workflow entirely built in the KNIME platform, as described in the Experimental Section. The descriptor values were later normalized, in order to be comparable and to contribute equally to the analysis. In order to strengthen the group of descriptors two categorical ones were added that include information about the core of the NMs (pure metal or metal oxide) and the pH of the zeta-potential measurement (as pH affects the degree of charge neutralization).

In a next step, the whole dataset (18 parameters extracted from each of the 68 TEM images and 2 additional data points (core and pH) for each of the 37 NMs, in total 740 data points) was partitioned randomly for external validation purposes into training and test sets in the proportion 75:25. The training set was used in the model development and the test set was not involved in this process but was kept as a blank set for subsequent model validation. The *BestFirst* variable selection along with the *CfsSubsetEval* evaluator were applied to the training set, in order to select the most significant of the 20 nanodescriptors.

Using the KNIME platform gives the flexibility not only to experiment with different modeling techniques included in the WEKA suite, but also to implement bespoke approaches for the *k*NN modeling. In-house *Enalos+ kNN* KNIME node, with enriched read-across functionalities, was used. One of the major advantages of this node is that it presents—along with the predicted values—the neighboring training samples for each test NM, and thus it was possible to gain insight into the overall samples' space and proceed with additional interpretation of the results in terms of categorization and grouping of NMs. The *k*NN methodology, with the optimal value of $k = 7$ neighbors, has emerged as the one that produced both satisfactory and reliable predictions, as it was successful in all Tropsha's[46,47] recommended tests (Equations (5) to (8)) to assess the predictive ability of developed models

$$R^2_{\text{pred}} = 0.898 > 0.6 \tag{10}$$

$$R^2_{\text{ext}} = 0.907 > 0.5 \tag{11}$$

$$\frac{R^2 - R^2_o}{R^2} = 0.003 < 0.1 \tag{12}$$

$$0.85 < k = 1.056 < 1.15 \tag{13}$$

In **Figure 7**, the predicted zeta-potential values for the test NMs set are presented along with the corresponding experimentally measured values for the test NMs set.

Finally, the Y-randomization test was performed, which confirmed that the proposed predictive model is robust. Five random shuffles of the endpoint variable vector (zeta potential) were performed, whereas the descriptor matrix remained intact. The correlation coefficient (R^2) was tracked, as can be seen in **Table 2**, and the models presented have statistically lower predictive power than the initial one, therefore it can be considered that the accuracy of the proposed model is true, and is not due to chance correlation.

For the calculation of the applicability domain (APD) cut-off limit, the type of core and the pH were excluded as categorical variables. However, given that in both training and test sets these

Table 2. Correlation coefficient and satisfied tests between different random shuffles in a Y-randomization test.

	R^2_{pred}	Satisfied criteria
Initial	0.898	4
Rand-1	0.015	0
Rand-2	0.098	0
Rand-3	0.317	0
Rand-4	0.416	1
Rand-5	0.386	1

variables had the same possible values, the calculation of the APD was not affected by this exclusion. The APD threshold was calculated for each test NM according to its seven selected neighbors in the training set. As can be seen in Table S2 (Supporting Information), the test NMs with domain values higher than their corresponding PAD threshold are considered unreliable.

As described above, for validation purposes, various random, stratified (regarding the type of core) partitions with the same proportion (75:25) were performed, to assess the predictive power of the approach independently of the data partitioning. In every case, variable selection (as described above) was performed in order to clearly define the image nanodescriptors space among the initial set of 18 descriptors. All models were successful in Tropsha's[46,47] recommended tests and the results for the squared correlation for all different splits are presented in Table S3 (Supporting Information).

4.1. The Descriptors Space

As a next step, after extracting image nanodescriptors, building and validating the predictive model, interpretation of the variable selection results, and clear definition of the variables that emerged as important for modeling the zeta-potential endpoint are provided. In all partitions, the type of core of the NMs and their main elongation emerged as important variables. Here, the information encoded by these descriptors is analyzed in order to understand how they affect the zeta-potential value for a specific NM. The predominant variables of the initial model are presented next in order of significance.

4.1.1. Main Elongation

The elongation (Equation (14)) is calculated using the parameters of the minimum bounding box; the larger side (L) and the shortest side (S) of the minimum bounding rectangle. The smallest enclosing rectangle (or minimum bounding box, **Figure 8**) is the smallest rectangle that contains every point of the particle. The main elongation variable expresses the lengthening of the particle and is similar to aspect ratio descriptor (L/S)[56]

$$\text{Main elongation} = 1 - \frac{S}{L} \tag{14}$$

The fact that this parameter has a high appearance rate for zeta potential is interesting for two reasons: First, there are significant questions over the reliability of experimental measurements for zeta potential as NMs become less spherical and/or agglomerated, since the underpinning mathematical models for zeta potential are based on Stoke's law and assume spherical particles.[28] Second, there is a well-established paradigm for increasing toxicity of NMs with increasing aspect ratio, resulting in the definition of a subgroup of NMs described as high aspect ratio NMs (HARNs), which are perceived as more hazardous, especially when combined with rigidity and persistence.[57]

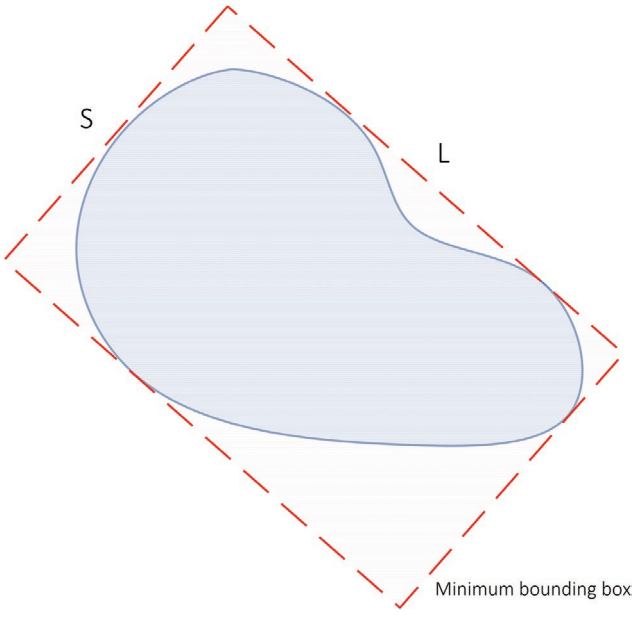

Figure 8. Minimum bounding box (or rectangle) of a particle.

4.1.2. Type of Core

The inner material of a NM, its core, is one of the principal factors that define their behavior during production and processing and their interaction with the environment and humans. The core of the NMs is also responsible for their main intrinsic physicochemical properties, such as electrical, magnetic, and catalytic properties, selectivity, solubility, etc.[58] For example, some NMs (e.g., Ag, Cu/CuO, ZnO) may dissolve quickly in a medium, while others—like cerium dioxide or titanium dioxide—dissolve at a slower rate.[3] Finally, the core of the NMs consists of the matrix for the application of the coating that—as mentioned before—can alter their properties.[58] One of the major paradigms suggested as being predictive of metal/metal oxide NMs toxicity is bandgap,[59,60] i.e., the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) which is linked to core composition and crystal structure, and thus the appearance of core composition as one of the main nanodescriptors is not surprising.

4.2. The Neighboring Space—Read-Across

Another important aspect that has been considered was to "unbox" the *k*NN algorithm and study the selected training neighbors for each test NM. In that way, it was possible to search for patterns and similarities in the neighborhood space and to do a preliminary grouping of the NMs as can be seen in the qualitative illustration of the neighboring relationships in **Figure 9**. In Table S4 (Supporting Information), the selected training neighbors for each of the NMs in the initial test set are presented.

During the study of the neighborhood (Figure 9 and Table S4, Supporting Information), it was expected that some patterns in

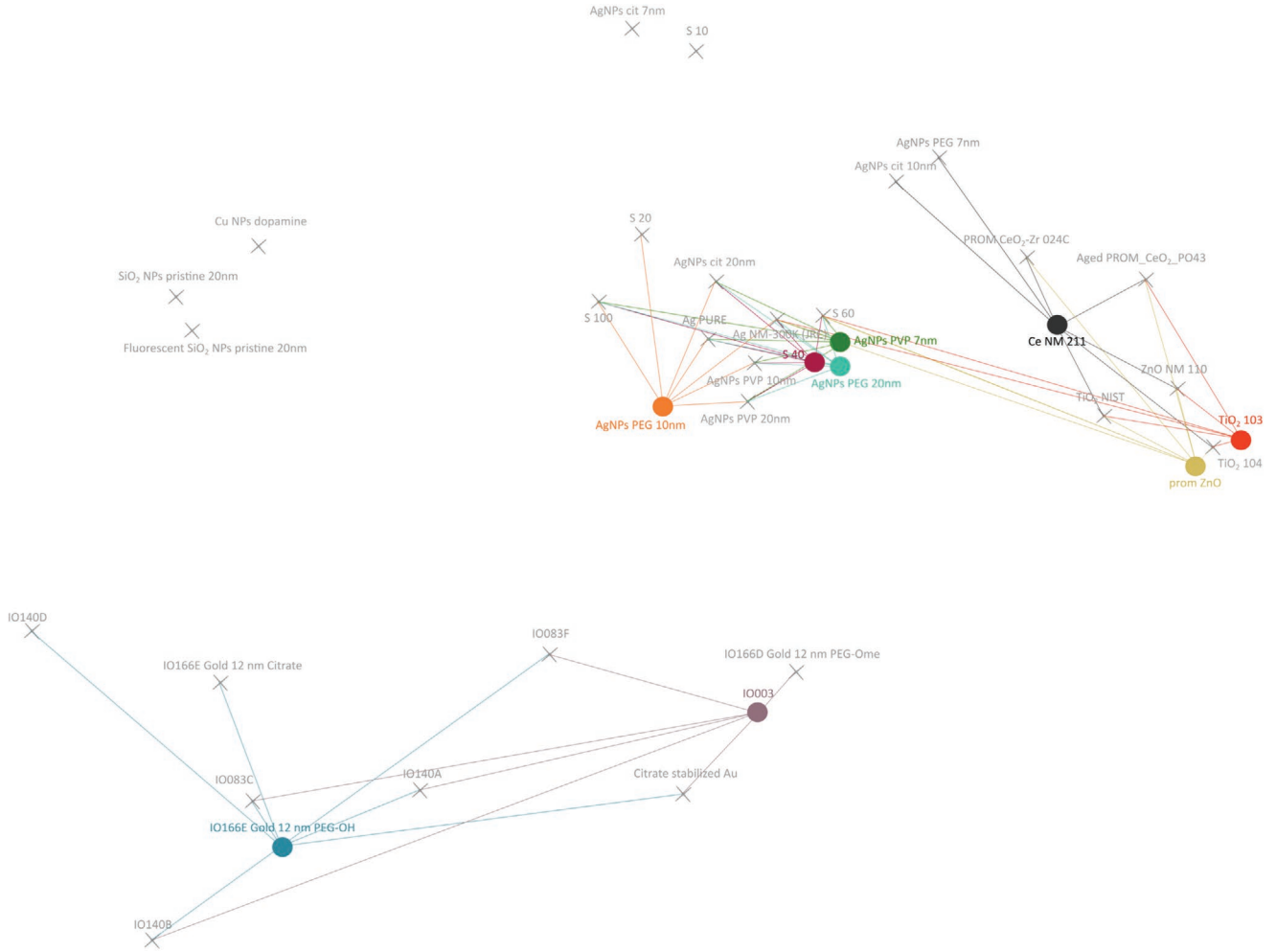

Figure 9. A qualitative representation of the neighboring space of the training and the test NM sets. Test NMs are depicted with colored circles, whereas training NMs are illustrated with gray crosses. The seven selected-closest neighbors for each test NM are clearly defined via lines.

the selection of neighbors would be recognized, due to obvious characteristics such as having the same core, coating, production method, etc., of the NMs. At a first glance this is clearly the case for the samples AgNPs PEG 10 nm, AgNPs PEG 20 nm, AgNPs PVP 7 nm and S40, that have neighbors with the same core (silver) and coating, and the case of samples IO003 and IO166E Gold 12 nm PEG-OH which have neighbors with the same core (gold). For the rest of the samples (Ce NM 211, Prom ZnO, and TiO_2 103) there are not enough (at least seven) NMs with the same core in the training set to be selected as neighbors, therefore the furthest neighbors are training samples with different core compositions.

It could be assumed in this case study that the type of core could be (along with the medium pH variable) the only variable that controls the whole grouping and prediction of zeta-potential process. However, the role of the main elongation variable must not be underestimated; its values have a balancing function in the selection of similar NMs (e.g., in the case where not enough training NMs with the same core are available). In the methodology of the neighbors' selection, the categorical values have either a great (equal to 1, when the compared attributes have different values) or a null participation (equal to 0, when the compared attributes have the same values) in the distance calculation. The rest of the nanodescriptors which are numeric values, have a tuning role in the calculation of the distance, as they encode the subtle similarities or dissimilarities between the NMs, and they finally contribute to a better selection of neighbors. What's more their participation in the calculation of distance (and thereby in the calculation of weighting factors) leads to more accurate predictions.

It can therefore be concluded that the selected variables are indeed the appropriate ones, among the initial set of nanodescriptors, for modeling the zeta-potential index. To demonstrate this, the modeling process was repeated using as input descriptors only the core of the NMs (and the medium pH). This model presented lower predictive power than our initial one ($R^2_{\text{pred}} = 0.85$ and passed only three out of the four validation criteria[46,47]); thus, its predictive ability cannot be considered as high as the predictive ability of the initial model utilizing the core, the pH of the medium, and the NM main elongation extracted from the NMs TEM images. It is worth noting that the used dataset was rather small and heterogeneous (some training NMs are not selected as neighbors); as the dataset becomes more complete in size it is expected to achieve a better neighbor tuning and better prediction results.

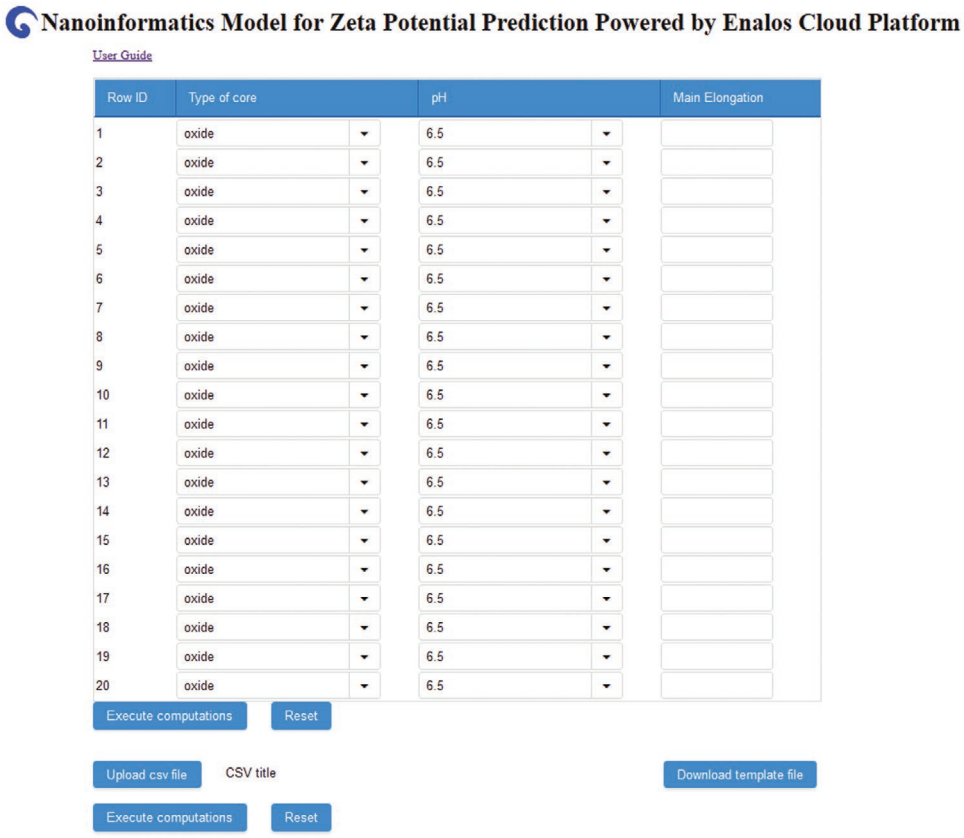

Figure 10. Enalos Zeta Potential Prediction platform. Users can either input data via the input form or by clicking on Upload csv file button at the bottom of the form, the user can import a .csv file with all the required properties.

4.3. Enalos Zeta Potential Web Tool

As noted earlier, it is very important to maximize the impact of the developed and validated model, by providing a user-friendly environment that will facilitate its direct use by interested users (nanoinformatics experts or not) and will serve as an important tool in future computer-aided NM design and quality control.

Within this work the NM image descriptors calculation tool has been integrated within a nanoinformatics workflow to identify the important descriptors among the available pool of NM image descriptors to develop a predictive model for zeta potential, as a demonstration case. Other physiochemical properties, behaviors, or biological effects could also be explored and modeled using the image-derived nanodescriptors. Based on the selected set of significant descriptors (type of core and main elongation) and the pH of interest a predictive model was developed and then released through the Enalos Cloud platform. This web service can be easily accessed (http://enaloscloud.novamechanics.com/EnalosWebApps/ZetaPotential/) and explored, within minimum steps required, by anyone interested in NM design, without any need for prior programming skills.

Users should use the Enalos **NanoXtract** tool or an image analysis tool of their choice, to provide the requested properties required by the grouping model (main elongation) and the type of core and the pH of interest. In order to introduce the above information to the model there are two different options provided: i) users can either enter manually the three required parameters using the form given in the website (advisable for small NM sets) or ii) users can import a file in .csv format, containing the NM samples and their properties (**Figure 10**). After submitting the required information, predictions are produced and presented within seconds in two different formats: a summary of the results in a new html page, or a file in .csv format, containing all the available information for further analysis. The results include the predicted zeta-potential values for each included NM as well as a warning on the prediction reliability according to the domain of applicability limits. Users can import different datasets with NMs of interest and study the effects of different inputs on the zeta-potential value, a decisive step during a safety-by-design process.

5. Conclusions

In silico assessment of various NM properties and biological effects prior to their use or even prior to actual synthesis is significantly contributing to a reduction of the cost and time required for experimental procedures required to generate, for example, regulatory dossiers. NM specific descriptors are thus highly desired to develop significant correlations between NM descriptors and NM properties and biological effects. Image descriptors derived from NM TEM images are emerging as an important source of additional information providing an

enriched parameter space to explore in order to find correlations between NMs properties and their effects, although to date these have not been extensively studied in part because of the lack of tools available to extract additional nanodescriptors from TEM images.

To facilitate NM image descriptors generation, a unique tool dedicated to extraction of image descriptors from NM TEM images has been developed and is presented here. **NanoXtract** thus fulfills the unmet need for the generation of larger sets of nano-specific descriptors with minimal experimental data requirements. Within a simple and user-friendly interface, the user can upload a single TEM image of a specific NM and with just a few clicks to select the core and coating descriptors, obtain a set of NM image descriptors averaged across all NMs present in the image or on a particle by particle basis. The image descriptors can then be used to develop robust and accurate predictive models.

Based on this tool for image descriptors calculation a workflow to demonstrate the utility of the extracted image nanodescriptors for prediction of NMs physicochemical properties was implemented, utilizing zeta potential as a first example as experimental values were available for the 37 NMs included in the dataset. The predictive model for NM zeta potential was based on grouping of the NMs according to their nearest neighbors and provided some interesting insights into the most important similarity features between the NMs and their nearest neighbors. Thus, in addition to grouping based on the NMs core composition (e.g., CeO_2, Ag, TiO_2, etc.), the main elongation emerged as an important grouping parameter, which links to key drivers of NM toxicity, such as aspect ratio. Importantly, the ability to predict zeta-potentials values for non-spherical NMs fills a gap where experimental measurement reliability and meaningfulness is poor and thus the image-based nanodescriptors and predicted zeta potentials can be used to improve subsequent predictions of NMs toxicity and adverse outcomes.

To ensure their accessibility to the wider community, both the **NanoXtract** image analysis tool and the zeta-potential predictive model have been made publicly available as web services through the Enalos Cloud Platform, enabling future in silico exploitation of NM properties and behavior based on the extracted image descriptors.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Data were generated within EU FP7 NanoMILE (310451), analysis and development of the **NanoXtract** webtool was funded by EU H2020 research infrastructure NanoCommons (731032) and via secondments of I.L. and E.V.-J. to NovaMechanics, and D.-D.V. to UoB supported by the EU H2020 RISE project NanoGenTools (691095). The zeta potential predictive model was developed via H2020 project NanoSolveIT (814572). D.-D.V. acknowledges funding from the Onassis Foundation for her Ph.D. studies (G ZN 008-1/2017-2018).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

image nanodescriptors, nanoinformatics, read-across, zeta potential

Received: November 13, 2019
Revised: February 7, 2020
Published online:

[1] M. Fojtů, W. Z. Teo, M. Pumera, *Environ. Sci.: Nano* **2017**, *4*, 1617.
[2] Y. Robert, R. MacPhail, *J. Occup. Med. Toxicol.* **2011**, *6*, 7.
[3] E. Valsami-Jones, I. Lynch, *Science* **2015**, *350*, 388.
[4] A. Hughes, Z. Liu, M. Raftari, M. E. Reeves, *PeerJ Prepr.* **2014**, *2*, e671v2.
[5] G. Lalwani, M. D'Agati, A. M. Khan, B. Sitharaman, *Adv. Drug Delivery Rev.* **2016**, *105*, 109.
[6] S. Mondini, A. M. Ferretti, A. Puglisi, A. Ponti, *Nanoscale* **2012**, *4*, 5356.
[7] A. Dudkiewicz, A. B. A. Boxall, Q. Chaudhry, K. Mølhave, K. Tiede, P. Hofmann, T. P. J. Linsinger, *Food Chem.* **2015**, *176*, 472.
[8] A. Mikolajczyk, A. Gajewicz, B. Rasulev, N. Schaeublin, E. Maurer-Gardner, S. Hussain, J. Leszczynski, T. Puzyn, *Chem. Mater.* **2015**, *27*, 2400.
[9] P. Toth, J. K. Farrer, A. B. Palotas, J. S. Lighty, E. G. Eddings, *Ultramicroscopy* **2013**, *129*, 53.
[10] G. V. Lowry, R. J. Hill, S. Harper, A. F. Rawle, C. O. Hendren, F. Klaessig, U. Nobbmann, P. Sayre, J. Rumble, *Environ. Sci.: Nano* **2016**, *3*, 953.
[11] R. Vogel, A. K. Pal, S. Jambhrunkar, P. Patel, S. S. Thakur, E. Reátegui, H. S. Parekh, P. Saá, A. Stassinopoulos, M. F. Broom, *Sci. Rep.* **2017**, *7*, 17479.
[12] S. M. Briffa, F. Nasser, E. Valsami-Jones, I. Lynch, *Environ. Sci.: Nano* **2018**, *5*, 1745.
[13] A. B. Stefaniak, V. A. Hackley, G. Roebben, K. Ehara, S. Hankin, M. T. Postek, I. Lynch, W. E. Fu, T. P. J. Linsinger, A. F. Thünemann, *Nanotoxicology* **2013**, *7*, 1325.
[14] A. Afantitis, G. Melagraki, A. Tsoumanis, E. Valsami-Jones, I. Lynch, *Nanotoxicology* **2018**, *12*, https://doi.org/10.1080/17435390.2018.1504998.
[15] A. Bigdeli, M. R. Hormozi-Nezhad, M. Jalali-Heravi, M. R. Abedini, F. Sharif-Bakhtiar, *RSC Adv.* **2014**, *4*, 60135.
[16] A. Gajewicz, T. Puzyn, K. Odziomek, P. Urbaszek, A. Haase, C. Riebeling, A. Luch, M. A. Irfan, R. Landsiedel, M. van der Zande, H. Bouwmeester, *Nanotoxicology* **2018**, *12*, 1.
[17] G. Melagraki, A. Afantitis, *Curr. Top. Med. Chem.* **2015**, *15*, 1827.
[18] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H. M. Hwang, A. Toropov, D. Leszczynska, J. Leszczynski, *Nat. Nanotechnol.* **2011**, *6*, 175.
[19] D. D. Varsou, G. Tsiliki, P. Nymark, P. Kohonen, R. Grafstrom, H. Sarimveis, *J. Chem. Inf. Model.* **2018**, *58*, 543.
[20] A. Toropov, N. Sizochenko, A. Toropova, J. Leszczynski, *Nanomaterials* **2018**, *8*, 243.
[21] G. Melagraki, A. Afantitis, *Food Chem. Toxicol.* **2018**, *112*, 476.
[22] S. B. Rice, C. Chan, S. C. Brown, P. Eschbach, L. Han, D. S. Ensor, A. B. Stefaniak, J. Bonevich, A. E. Vladár, A. R. H. Walker, J. Zheng, C. Starnes, A. Stromberg, J. Ye, E. A. Grulke, *Metrologia* **2013**, *50*, 663.
[23] T. Wollmann, H. Erfle, R. Eils, K. Rohr, M. Gunkel, *J. Biotechnol.* **2017**, *261*, 70.

[24] C. Chomenidis, G. Drakakis, G. Tsiliki, E. Anagnostopoulou, A. Valsamis, P. Doganis, P. Sopasakis, H. Sarimveis, *J. Chem. Inf. Model.* **2017**, *57*, 2161.
[25] K. Odziomek, D. Ushizima, T. Puzyn, M. Haranczyk, *Proc. Am. Chem. Soc.* **2014**.
[26] M. D. Abràmoff, P. J. Magalhães, S. J. Ram, *Biophotonics Int.* **2004**, *11*, 36.
[27] "NanoMILE," http://nanomile.eu-vri.eu/ (accessed: April 2018).
[28] R. R. R. Marín, F. Babick, L. Hillemann, *Colloids Surf., A* **2017**, *532*, 516.
[29] D.-D. Varsou, G. Melagraki, H. Sarimveis, A. Afantitis, *Food Chem. Toxicol.* **2017**, *110*, 83.
[30] KNIME, "KNIME Image Processing," https://www.knime.com/community/image-processing (accessed: January 2018).
[31] D.-D. Varsou, S. Nikolakopoulos, A. Tsoumanis, G. Melagraki, A. Afantitis, in *Rational Drug Design: Methods and Protocols, Methods in Molecular Biology*, Vol. (Eds: T. Mavromoustakos, T. F. Kellici), Humana, New York **2018**, pp. 113–138.
[32] I. Hansjosten, J. Rapp, L. Reiner, R. Vatter, S. Fritsch-Decker, R. Peravali, T. Palosaari, E. Joossens, K. Gerloff, P. Macko, M. Whelan, D. Gilliland, I. Ojea-Jimenez, M. P. Monopoli, L. Rocks, D. Garry, K. Dawson, P. J. F. Röttgermann, A. Murschhauser, J. O. Rädler, S. V. Y. Tang, P. Gooden, M. F. A. Belinga-Desaunay, A. O. Khan, S. Briffa, E. Guggenheim, A. Papadiamantis, I. Lynch, E. Valsami-Jones, S. Diabaté, C. Weiss, *Arch. Toxicol.* **2018**, *92*, 633.
[33] B. Michen, C. Geers, D. Vanhecke, C. Endes, B. Rothen-Rutishauser, S. Balog, A. Petri-Fink, *Sci. Rep.* **2015**, *5*, 9793.
[34] A. R. Leach, V. J. Gillet, *An Introduction to Chemoinformatics*, Springer, Berlin **2007**.
[35] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nat. Methods* **2012**, *9*, 676.
[36] "ImageJ," https://github.com/imagej/imagej-ops/find/master (accessed: January 2018).
[37] M. N. Pons, H. Vivier, K. Belaroui, B. Bernard-Michel, F. Cordier, D. Oulhana, J. A. Dodds, *Powder Technol.* **1999**, *103*, 44.
[38] S. Mülhopt, S. Diabaté, M. Dilger, C. Adelhelm, C. Anderlohr, T. Bergfeldt, J. Gómez de la Torre, Y. Jiang, E. Valsami-Jones, D. Langevin, I. Lynch, E. Mahon, I. Nelissen, J. Piella, V. Puntes, S. Ray, R. Schneider, T. Wilkins, C. Weiss, H.-R. Paur, *Nanomaterials* **2018**, *8*, 311.
[39] G. Melagraki, A. Afantitis, *RSC Adv.* **2014**, *4*, 50713.
[40] I. H. Witten, E. Frank, M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann, Burlington, MA **2011**.
[41] C. J. Pal, I. H. Witten, E. Frank, M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlington, MA **2016**.
[42] NovaMechanics Ltd, "Enalos+ Modelling nodes," http://www.enalosplus.novamechanics.com/index.php/enalosplusnodes/modelling/ (accessed: 2020).
[43] R. Huluban, *Practical Guide: How to Use and Report (Q)SARs*, ECHA, Helsinki, Finland **2016**.
[44] *Read-Across Assessment Framework (RAAF)*, ECHA, Helsinki, Finland **2017**.
[45] Environment Directorate, *OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69 Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, **2007**.
[46] A. Tropsha, *Mol. Inf.* **2010**, *29*, 476.
[47] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2003**, *22*, 69.
[48] D. Gadaleta, G. F. Mangiatordi, M. Catto, A. Carotti, O. Nicolotti, *Int. J. Quant. Struct.-Prop. Relat.* **2016**, *1*, 45.
[49] G. Leonis, G. Melagraki, A. Afantitis, in *Handbook of Computational Chemistry* (Ed: J. Leszczynski), Springer, Berlin **2016**.
[50] NovaMechanics Ltd., "Enalos+ KNIME nodes," http://enalosplus.novamechanics.com/ (accessed: January 2020).
[51] NovaMechanics Ltd., "Enalos Cloud Platform," http://www.insilicotox.com/index.php/products/predictive-models-web-services/ (accessed: February 2018).
[52] A. R. Gliga, S. Skoglund, I. O. Wallinder, B. Fadeel, H. L. Karlsson, *Part. Fibre Toxicol.* **2014**, *11*, 11.
[53] D. Lin, X. Tian, F. Wu, B. Xing, *J. Environ. Qual.* **2010**, *39*, 1896.
[54] W. S. Cho, R. Duffin, F. Thielbeer, M. Bradley, I. L. Megson, W. MacNee, C. A. Poland, C. L. Tran, K. Donaldson, *Toxicol. Sci.* **2012**, *126*, 469.
[55] J. M. Berg, A. Romoser, N. Banerjee, R. Zebda, C. M. Sayes, *Nanotoxicology* **2009**, *3*, 276.
[56] "ImageJ Shape Filter," https://imagej.net/Shape_Filter (accessed: October 2018).
[57] Z. Ji, X. Wang, H. Zhang, S. Lin, H. Meng, B. Sun, S. George, T. Xia, A. E. Nel, J. I. Zink, *ACS Nano* **2012**, *6*, 5366.
[58] R. G. Chaudhuri, S. Paria, *Chem. Rev.* **2012**, *112*, 2373.
[59] S. Noventa, C. Hacker, D. Rowe, C. Elgy, T. Galloway, *Nanotoxicology* **2018**, *12*, 63.
[60] H. Zhang, Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y.-P. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Mädler, Y. Cohen, J. I. Zink, A. E. Nel, *ACS Nano* **2012**, *6*, 4349.