












Utilizing Machine Learning Algorithms for Accurate Prediction of Nanomaterials Cytotoxicity

Aries P. Valeriano ^{1,3} , Farley G. Bondaug ^{1,2}, Irelie P. Ebarido ^{1,2}, Marybeth Hope T. Banda ^{1,2} ,
Penelope P. Almonte ¹ , Michelle Amor P. Sabugaa ¹ , John Riz V. Bagnol ³ ,
Mary Joy R. Latayada ⁴ , Jay Michael R. Macalalag ⁴ , Brian Dominic Paradero ⁵,
Maricris L. Mayes ⁶ , Mannix P. Balanay ⁷ , Arnold C. Alguno ⁸ , Rey Y. Capangpangan ^{1,9,*} 

¹ Research on Environment and Nanotechnology Laboratories, Research Division, Mindanao State University at Naawan, Naawan, 9023, Misamis Oriental, Philippines; penelope.almonte@msunaawan.edu.ph (P.A.), michelleamor.sabugaa@msunaawan.edu.ph (M.A.S.);

² Department of Science and Technology - Science Education Institute, Taguig City, 1631, Metro Manila, Philippines; farley.bondaug@msunaawan.edu.ph (F.B.); irelie.ebarido@msunaawan.edu.ph (I.E.); marybethhope.banda@msunaawan.edu.ph (M.H.B.);

³ Department of Mathematics and Statistics, University of Southeastern Philippines, Davao City, 8000, Davao del Sur, Philippines; apvaleriano@usep.edu.ph (A.V.); jrbagnol@usep.edu.ph (J.R.B.);

⁴ Department of Mathematics, Caraga State University, Butuan City, 8600, Agusan del Norte, Philippines; mrlatayada@carsu.edu.ph (M.J.L.); jrmacalalag@carsu.edu.ph (J.M.M.);

⁵ Information, Communication and Technology Center, Mindanao State University at Naawan, Naawan, 9023, Misamis Oriental, Philippines; brian.paradero@msunaawan.edu.ph;

⁶ Department of Chemistry and Biochemistry, University of Massachusetts, Dartmouth, 02748, Massachusetts, USA; maricris.mayes@umassd.edu;

⁷ Department of Chemistry, Nazarbayev University, Astana, Kazakhstan; mannix.balanay@nu.edu.kz;

⁸ Department of Physics, Mindanao State University-Iligan Institute of Technology, Iligan City, 9200, Lanao del Norte, Philippines; arnold.alguno@g.msuiit.edu.ph;

⁹ Department of Physical Sciences and Mathematics, Mindanao State University at Naawan, Naawan, 9023, Misamis Oriental, Philippines; rey.capangpangan@msunaawan.edu.ph;

* Correspondence: rey.capangpangan@msunaawan.edu.ph

Received: 12.02.2024; Accepted: 30.06.2024; Published: 6.09.2025

Abstract: Nanomaterials (NMs) have been widely used in various sectors in recent years, which has prompted efforts to solve the difficulties in synthesizing safe-by-design NMs. The toxicity of NMs can be investigated by *in silico* approaches, which employ machine learning (ML) algorithms to develop quantitative structure-activity relationship (QSAR) models. If built correctly, these models can predict NMs' toxicity with the highest accuracy and reliability. As such, this study aimed to develop several QSAR models that predict the cytotoxicity level of various engineered NMs. Partial least squares - discriminant analysis (PLS-DA), random forest (RF), decision trees (DT), and k-nearest neighbors (kNN) were used to develop the QSAR models. Consequently, during internal and external validation, the models achieved F1 scores ranging from 91.03–95.23% and 90.52–95.19%. Additionally, all models identified exposure time and concentration as highly influential descriptors. Moreover, data clustering based on the most significant descriptors further enhanced the performance of the models, particularly when clustered based on the condition “concentration < 31 µg/ml”. As a result, all models achieved slightly over 95% F1 scores upon internal and external validation. These results imply that the developed QSAR models are highly accurate and reliable.

Keywords: *in silico*; machine learning; QSAR; cytotoxicity; nanomaterials; clustering.

© 2025 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The authors retain copyright of their work, and no permission is required from the authors or the publisher to reuse or distribute this article, as long as proper attribution is given to the original source.

1. Introduction

Nanomaterials (NMs) possess distinctive attributes, including their diminutive size and elevated surface-to-volume ratio, which render them advantageous yet possibly hazardous to biological entities. In the present times, there is a concerning rise in the number of diseases that stem from daily exposure to potentially harmful chemicals or materials [1, 2]. Despite these potential drawbacks, due to their convenience and efficacy, the widespread utilization of NMs is prevalent across diverse industries, including but not limited to agriculture, business, medicine, and public health [3, 4]. Consequently, the growing concern for safety has led to a rapid increase in research endeavors aimed at comprehending the toxicity patterns of NMs. To understand the hazard potential of NMs, researchers have found computational methods in silico to be an effective approach [5]. In silico approaches utilized machine learning (ML) algorithms to develop quantitative structure-activity relationship (QSAR) models for predicting NMs' toxicity. In this study, cytotoxicity was the focus of various toxicity assessments, mainly because it is highly responsive to different treatments and doesn't involve complex or time-consuming procedures. [6]. QSAR models can identify the descriptors that lead to the toxic effects of NMs on living cells. For instance, descriptors such as NMs' physicochemical properties, cell exposure time, and/or chemical composition, among others. In recent years, several studies have been conducted to develop a reliable and accurate QSAR model for NM cytotoxicity evaluation. For example, Sang *et al.* developed Partial least squares (PLS) and random forest (RF) models that reliably predict the cytotoxicity of a mixture of nano-metal oxides and heavy metals to human renal cortex proximal tubule epithelial (HK-2) cells [7]. Yuan *et al.* presented PLS and RF regression models that predict well the viability of HK-2 cells exposed to a mixture containing nano-TiO₂ and heavy metals [8]. Qi *et al.* have built a highly accurate Monte Carlo – PLS model in the prediction of 34 modified TiO₂-based nanoparticles' cytotoxicity towards the Chinese hamster ovary cell line, where Ag and Brunauer-Emmett-Teller (BET) surface area were identified as the most significant properties [9]. Shi *et al.* [7] proposed an RF model that could successfully describe the cytotoxicity of metal oxide nanoparticles to pancreatic cancer (PaCa2) cells [10]. Desai *et al.* suggested the use of their developed decision trees (DT) and RF models to enhance the synthesis of Ag nanoparticles for various applications, particularly in drug delivery and cancer treatments [3]. Bilgi and Karakus developed machine learning models, including decision tree (DT) and artificial neural network (ANN) models, to predict nanosilver toxicity. Their findings highlight factors such as exposure concentration, duration, zeta potential, particle size, and coating with the greatest impact on nanotoxicity [11]. Furthermore, Labouta *et al.* have developed DT models for inorganic and organic NPs cytotoxicity, which were said to be simple, accurate, and have broad applicability [12]. All these studies have shown that computational methods in silico are effective in assessing the toxic effects of NMs on living cells, which can improve the safety of NMs used in various applications. However, except for the DT models developed by Labouta *et al.*, all other models were built with a specific domain. Some of them apply only to metal oxide, TiO₂-based, or Ag NMs. Also, these NMs were exposed only to specific cells to measure cytotoxicity.

As such, this study aims to build several QSAR models to predict the cytotoxicity of a more comprehensive type of engineered NMs. The NMs can be inorganic, organic, or even carbon-based. This study ensures that the models are highly accurate and reliable. Hence, the dataset used in this study, which consisted of 4863 data points and 22 variables, has undergone

exploration and preprocessing before modeling. In this study, a data point is defined as a single observation with defined variables and endpoint, which, in our study, is % cell viability. In addition, the QSAR models based on the following ML algorithms, PLS-DA, RF, DT, and k-nearest neighbors (kNN), must follow the five principles of Organization for Economic Cooperation and Development (OECD) guidelines. These principles are the following: (i) a defined endpoint, (ii) an unambiguous algorithm, (iii) an applicability domain, (iv) an appropriate measure to evaluate the goodness of fit, and (v) a mechanistic interpretation.

The results of this study will provide more valuable insights into the cytotoxicity of diverse engineered NMs and can be used to improve the safety of their use in various applications.

2. Materials and Methods

2.1. Dataset.

This study involves a dataset extracted from the literature published between the years 2010 and 2020, with an update for the years 2021 and 2022 [13, 14]. The dataset consists of 4863 observations and 26 variables. An observation is an experimental setup with defined variables and a result or endpoint. One of the variables is the endpoint, % cell viability, which was transformed into the NMs cytotoxicity variable. Cell viability $\leq 50\%$ was labeled as high, otherwise low. The remaining 25 variables serve as descriptors for NMs cytotoxicity, and they are as follows: diameter (nm), size in medium (nm), size in water (nm), zeta in medium (mV), zeta in water (mV), number of cells (cells/well), exposure time (hr), concentration ($\mu\text{g/ml}$), aspect ratio, polydispersity index (PDI), core material type, shape, coat or functional group, synthesis method, surface charge, cell type, subject (human or animal), cell source, cell tissue, cell morphology, cell age, cell line or primary cell, cell viability test, and test indicator. However, 4 out of the 25 descriptors, namely aspect ratio, polydispersity index (PDI), zeta in medium (mV), and size in medium (nm), each have more than 50% missing values. The inclusion of these descriptors with such a large proportion of missing values would only introduce bias to the models developed. Therefore, they were excluded from this study. What remained in the dataset was then used to develop highly accurate and reliable predictive models using the caret package in R software.

2.2. Data partitioning.

This study employed a typical split ratio of 80:20 to partition the dataset. Accordingly, 80% of the data points were assigned to the training set for constructing QSAR models, while the remaining 20% constituted the testing set for external validation [15]. The selection of data points for the training set was carried out using simple random sampling, ensuring that each data point in the dataset had an equal chance of being chosen.

2.3. Assessing data imbalance.

In a classification task, particularly with an endpoint of two outcomes. It is crucial to assess data imbalance, as failure to do so could lead to incorrect conclusions. During this assessment, it is important to observe a high imbalance between the two classes of the endpoint. This occurs when one class has significantly more occurrences than the other. In such cases, the model's prediction of the minority class often becomes the class of interest. Therefore,

accurately classifying instances of this minority class holds great importance. In this study, by using the F1 score as an appropriate measure of the QSAR model's performance, the challenges posed by highly imbalanced data can be addressed [16]. The calculation of the F1 score considers both correctly and incorrectly classified outcomes. In contrast, accuracy only considers the correctly classified outcomes, which potentially leads to misleading results.

2.4. Data preprocessing.

Data preprocessing is an essential step to prepare the data for QSAR modeling. It involves cleaning, transforming, and organizing data. In this study, several data preprocessing techniques were employed. Each technique was implemented independently of training and testing sets to avoid data leakage, which could result in overfitting. Overfitting is when the model fits too well with the training set but fails to predict accurately when presented with the testing set [15].

One of the preprocessing techniques used in this study was kNN imputation. kNN imputation estimates the missing values in the dataset through the calculation of the weighted mean of k candidate samples, referred to as neighbors. These neighbors were selected based on a distance measure. In this study, a combination of qualitative and quantitative descriptors was involved; hence, Gower's distance measure was utilized [17].

Another technique used in this study was Z transformation. Z transformation standardizes each NMs cytotoxicity numerical descriptor so that it can have a mean of 0 and a standard deviation of 1. This technique is useful when the variables are measured in different units or have different ranges. It helps to reduce the impact of outliers and enables the comparison of different descriptors on the same scale, such as diameter (nm), exposure time (hr), and concentration ($\mu\text{g/ml}$), among others [18].

Moreover, categorical descriptors of NMs cytotoxicity that lack any inherent order, such as subject (human or animal) and cytotoxicity test (MTT, LDH, XTT, etc.), among others, were transformed into numeric variables through dummy encoding. Dummy encoding transforms a categorical descriptor with a C number of levels into a C-1 number of binary dummy descriptors [19].

Finally, to improve the models' performance, LASSO regression was used as a variable selection method. This technique removes insignificant descriptors that do not contribute to the variation of NPs cytotoxicity by analysing the regression coefficients, β . If $\beta_i = 0$, then the i^{th} variable is deemed irrelevant to the NP's cytotoxicity variation. This method helps to reduce the noise created by irrelevant descriptors during the modeling process and focuses only on the ones that improve the model's performance [20].

Overall, the data preprocessing techniques used in this study help to ensure that the NMs cytotoxicity dataset is clean, consistent, and suitable for QSAR modeling. They provide a solid foundation for the development of accurate and reliable QSAR models for NMs' cytotoxicity.

2.5. Model building.

To develop QSAR models for the prediction of various engineered NMs cytotoxicity, this study employed four ML algorithms, namely PLS-DA, kNN, DT, and RF. PLS-DA is a classification method that distinguishes the two outcomes of an endpoint by identifying the latent variables in the feature space that have the highest covariance with the descriptors.

Consequently, this improves the accuracy and interpretability of the model [21, 22]. kNN assigns a class label (low or high cytotoxicity) to a data point based on the majority vote of the k nearest neighbors [21, 23]. DT classifies data by recursively splitting it based on descriptors (e.g., concentration, test, exposure time). This process creates a tree structure with leaf nodes that represent class labels. When new instances occur, the tree is traversed, and the corresponding leaf node will provide the classification. Lastly, RF classifies new data points by finding the majority class of multiple DTs' predictions.

During the model-building process for each algorithm, internal validation was made through the implementation of 5-fold cross-validation. 5-fold cross-validation divides the training set into five subsets. Each model was trained on four of the subsets while using the remaining subset as the validation set. This process was repeated five times, with a different subset used for validation in each iteration. The final model's performance was determined by averaging the results of these iterations [15]. By using different validation sets, cross-validation provides a reliable estimate of the model's performance across diverse subsets of the data. Moreover, the importance of the descriptor for each of the QSAR models built was assessed using the varImp function in R [24].

2.6. Model validation.

The validity of the developed QSAR models was ensured by following the guidelines established by the Organization for Economic Co-operation and Development (OECD). The guidelines specify that the models should be associated with (i) a defined endpoint, (ii) an unambiguous algorithm, (iii) a domain of applicability, (iv) an appropriate measure of the model's goodness of fit, robustness, and predictivity, and (v) a mechanistic interpretation [25].

In this study, the defined endpoint was the various engineered NMs cytotoxicity. The unambiguous algorithms were the PLS-DA, RF, DT, and kNN. The applicability domain was established, which involved varying descriptors, such as physicochemical, test-related, and cell-related properties.

Moreover, the applicability domain was analyzed through a principal components analysis (PCA) bounding box. The PCA bounding box provides a way to assess the applicability domain (AD) of the models developed. This approach represents data points as PCA scores and labels them as training and testing sets in a 2D space. The x and y axes of this space are the principal components (PCs) that capture the most variation in the original dataset. The AD of the models was determined by the boundary space created by the data points labeled as training sets. Consequently, data points labeled as testing sets were considered within the model's AD if they fell within this boundary space [26].

The F1 score was used as an appropriate measure of the model's goodness of fit, robustness, and predictivity during internal and external validation [16]. Finally, a mechanistic interpretation was provided to describe the potential reason why certain descriptors were determined to be the most important in predicting NMs cytotoxicity levels.

2.7. Flow diagram of the procedure.

Figure 1 depicts the workflow of the study. The initial step involves utilizing a dataset comprising an endpoint variable representing nanomaterials (NMs) cytotoxicity (along with multiple descriptor variables). The cytotoxicity variable is characterized by two distinct classes, namely high and low, and an evaluation was conducted to determine their respective

distributions. Subsequently, the dataset undergoes preprocessing to enhance its quality. Using the preprocessed data, predictive models were constructed. Finally, the built models were validated using the testing set.

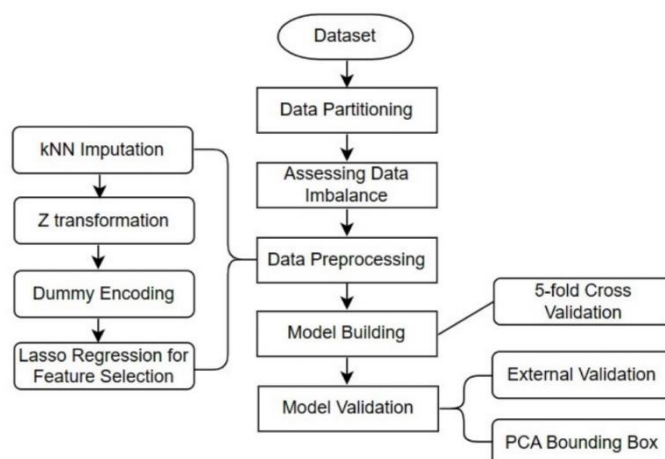


Figure 1. Flow chart of the procedure.

2.8. Data clustering.

The dataset utilized in this study is diverse and extensive. Hence, more accurate QSAR models can potentially be built by performing data clustering. Data clustering can be done by grouping similar data points, often through the implementation of statistical methodologies. However, in this study, an intuitive approach was used to cluster the dataset. The clustering was based on the most important descriptors determined across all models. For instance, various QSAR models such as PLS-DA, kNN, DT, and RF have determined that the synthesis method ‘commercial’ is one of the most important descriptors of NMs cytotoxicity. Consequently, the observations in the original dataset can be divided into two clusters: observations with the synthesis method ‘commercial’ and observations with the synthesis method ‘non-commercial’.

Moreover, if the most influential descriptor determined across all models is quantitative, such as concentration ($\mu\text{g/ml}$). Clustering will involve finding its central value. This can be achieved by taking the mean if there are no outliers or by using the median to obtain a central value robust to outliers. Subsequently, clusters can be formed by grouping observations from the original dataset based on concentration $<$ central value or concentration \geq central value. It is important to note that every time a descriptor is used as a decision rule for clustering, that descriptor is excluded from the created cluster of data.

For each of the clusters, the following steps in the flowchart were repeated: data preprocessing, model building, and, lastly, model validation.

3. Results and Discussion

3.1. Model performance evaluation.

The cytotoxicity endpoint variable in the dataset exhibits a highly imbalanced distribution, with 81% of the observations classified as low and the remaining 19% classified as high. Given this significant class imbalance, the F1 score is deemed the appropriate performance metric for evaluating the models built in this study. The F1 scores were denoted by $F1_{tr}$ and $F1_{te}$ during internal and external validation, respectively.

In Table 1, the performance of each model developed is displayed. It can be noticed that the four QSAR models, PLS-DA, RF, DT, and kNN, exhibit high-performance power that ranges from $F1_{tr} = 91.03$ to $F1_{tr} = 95.23\%$, and from $F1_{te} = 90.52$ to $F1_{te} = 95.19\%$ (see more details in Table S1). This indicates that the models are highly accurate in predicting NMs cytotoxicity levels. Desai *et al.* [3] also used F1 scores to evaluate the performance of DT and RF algorithms to enhance the synthesis of Ag nanoparticles for various applications. In addition, the small difference in the model's performance between internal and external validation suggests that each model generalizes well to new data points. Furthermore, among the four QSAR models built, RF has shown the highest predictive performance, with $F1_{tr} = 95.23\%$ and $F1_{te} = 95.19\%$. Hence, RF is the optimum QSAR model for the given dataset.

Table 1. Models' performance evaluation.

Model	% $F1_{tr}$	% $F1_{te}$
PLS-DA	91.03	90.52
RF	95.23	95.19
DT	91.59	91.35
KNN	91.79	91.71

The QSAR models built in this study can be compared to the four DT (DT1, DT2, DT3, and DT4) models developed by [12]. These DT models were built from a dataset with most of the descriptors similar to those used in this study but with fewer observations. Internal validation had shown high predictive performance, with DT1, DT2, DT3, and DT4 achieving accuracy rates of 87.9%, 90%, 88.2%, and 91.8%, respectively. The DT4 model may appear to perform slightly better than the PLS-DA, DT, and kNN models developed in this study. However, it is important to note that the four DT models did not undergo external validation. Therefore, the PLS-DA, DT, and kNN models can be considered more reliable than DT4.

3.2. Important descriptors.

Figure 2 illustrates the top 10 most important descriptors determined by QSAR models: DT, kNN, PLS-DA, and RF (see more details in Table S3). The importance measure of each descriptor is based on the `varImp` function in R, with values scaled from 0 to 100. A higher value indicates a greater influence of the descriptor on NMs cytotoxicity. Figure 2 shows that three descriptors are common across all models, which can be identified through '***' before their names. These descriptors include the synthesis method 'commercial', concentration, and exposure time. Among these three descriptors, exposure time and concentration often rank the highest by the built models. The importance of these descriptors aligns with the findings of the four DT models developed by [12].

Many known toxic substances have a direct relationship with exposure time and concentration. At elevated concentrations, nanomaterials exhibit an enhanced probability of engaging with cellular constituents, resulting in increased uptake and the possibility of unfavorable consequences. The concentration of nanomaterials plays a crucial role in determining the extent of their interactions with cells, tissues, or organisms [27]. Elevated concentrations have the potential to overpower cellular defense mechanisms, interfere with cellular processes, and trigger cellular stress, ultimately resulting in cytotoxicity. The exposure time is equally critical because it determines the duration of interaction between nanomaterials and biological systems [27]. Prolonged exposure to nanomaterials can allow for more significant cellular uptake and internalization, increasing the likelihood of adverse effects.

Additionally, extended exposure may lead to the accumulation of nanomaterials within cells or tissues, potentially affecting cellular functions and causing toxicity over time.

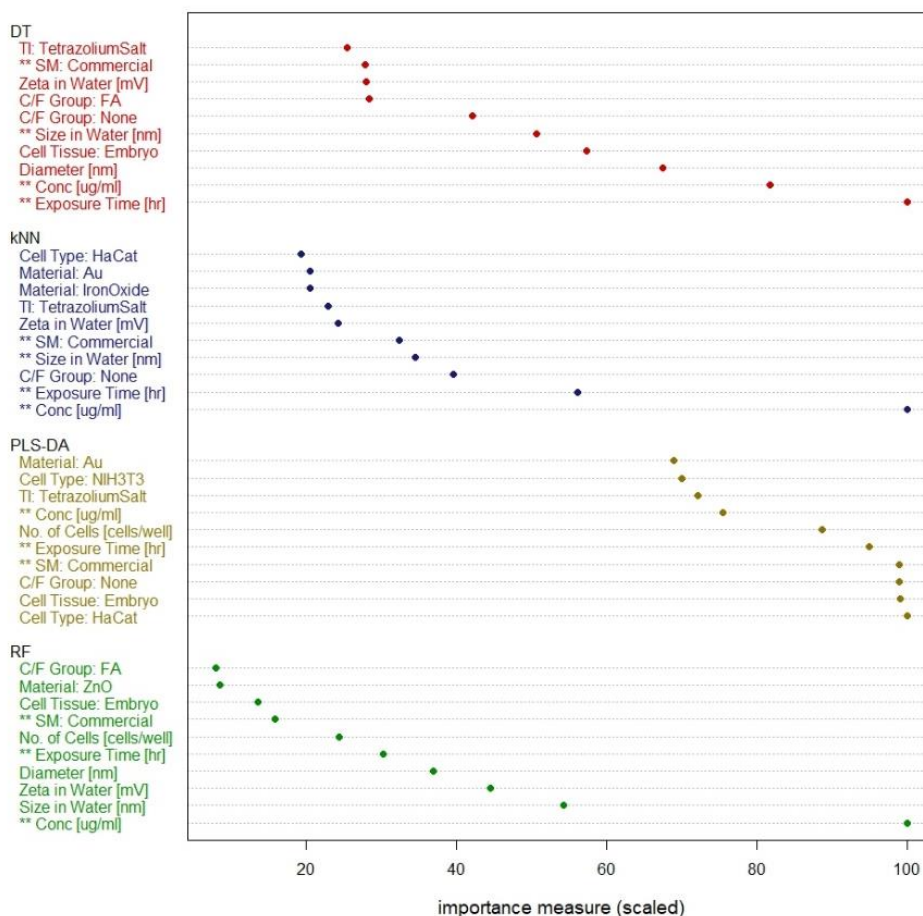


Figure 2. Top 10 descriptors identified by each of the models built from the original dataset. ‘**’ = common across all models, Conc = Concentration, SM = Synthesis Method, C/F = Coat/Functional, TI = Test Indicator, Material = Core Material.

Liu *et al.* [28], in their comprehensive review, discussed the applications of carbon nanotubes (CNTs) in biology and medicine. They addressed the importance of considering the concentration and exposure time of CNTs in understanding their cytotoxicity and their potential as drug delivery vehicles. Moreover, the US Department of Labor & Industries – WISHA Division has set Permissible Exposure Limits (PELs) for about 600 chemicals. The three PELs set by the said department are 8-hour time-weighted average (TWA), ceiling limit, and short-term exposure Limit (STEL). The study by [29] on the interactions of AuNPs with human dermal fibroblasts shows that in the presence of the 45 nm AuNPs, there is a higher rate of apoptosis, regardless of longer exposure or higher particle concentration, than that of the 13 nm AuNPs. The data presented suggests that AuNPs exhibit toxicity towards human dermal fibroblasts, with the extent of toxicity being dependent on the concentration and size of the nanoparticles, as well as the duration of exposure.

The prediction of cytotoxicity in nanomaterials can be greatly influenced by the commercial synthesis methods employed [30, 31]. This is because the chosen synthesis route substantially impacts the physicochemical characteristics of the nanomaterials, which subsequently affect their biological interactions and potential toxicity. Various synthesis methods can lead to differences in the dimensions, morphology, chemical composition, electrical charge, structural integrity, and presence of impurities in nanomaterials [32]. The aforementioned factors have the potential to exert an influence on the stability, dissolution rate,

agglomeration/aggregation behavior, cellular uptake, and subsequent biological responses of nanomaterials [33]. Pan *et al.* [31] investigated the cytotoxic effects of gold nanoparticles synthesized via different methods. The findings of their study indicate that the method employed for synthesis substantially impacts the size, surface chemistry, and cytotoxicity of the nanoparticles.

3.3. Model's performance evaluation per cluster.

Previously, in Figure 2, it was found that synthesis method 'commercial', concentration, and exposure time, were commonly identified as the most important descriptors across all QSAR models. By employing data clustering based on each of these descriptors, a more accurate model can potentially be built due to the dataset being varied and large in nature. The clustering resulted in six different datasets. These are as follows: observations with synthesis method 'non-commercial' ($n = 2365$), synthesis method 'commercial' ($n = 2498$), exposure time < 32.10 hrs ($n = 3499$), exposure time ≥ 32.10 hrs ($n = 1364$), concentration $< 31 \mu\text{g/ml}$ ($n = 2330$), and concentration $\geq 31 \mu\text{g/ml}$ ($n = 2346$). Here, n represents the number of observations per cluster, and the threshold values were determined by mean or median.

Figure 3 displays the performance of QSAR models built for each cluster. The models exhibit varying levels of F1 scores. In contrast to the predictive performance of the models built from the original dataset, clusters based on the synthesis method 'commercial', exposure time ≥ 32.10 hrs, and concentration $\geq 31 \mu\text{g/ml}$ have resulted in models with lower F1 scores. On the other hand, the remaining clusters, which are based on the synthesis method 'non-commercial', exposure time < 32.10 hrs, and concentration $< 31 \mu\text{g/ml}$, have produced models with higher F1 scores. Additionally, among these clusters, the cluster based on concentration $< 31 \mu\text{g/ml}$ appears to optimize the performance of PLS-DA, DT, and kNN models, while RF performance is consistent, yielding all F1 scores of over 95% during internal and external validation (see more details in Table S2). Hence, these QSAR models need to be further assessed.

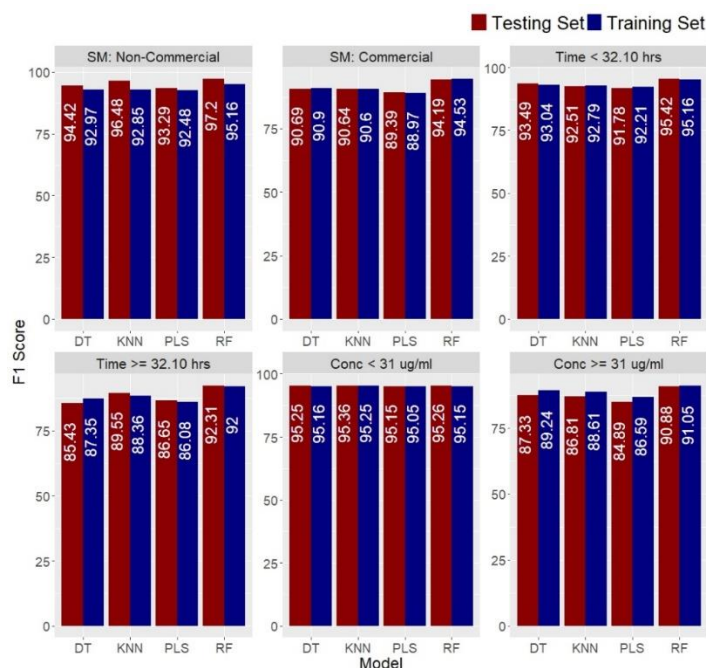


Figure 3. Model's performance from clustered data based on the prevalent most important descriptors. SM = Synthesis Method, Conc. = Concentration, SW = Size in Water, Time = Exposure Time.

3.4. Important descriptors for the cluster based on 'concentration < 31 µg/ml'.

Figure 4 shows the top 10 most influential descriptors of NMs cytotoxicity, as determined by each of the QSAR models built from the cluster based on concentration < 31 µg/ml (see more details in Table S4). It is evident that exposure time, material 'chitosan', and cell type 'k562' are identified as the most important descriptors across all models. K-562 cells are a type of lymphoblast that was extracted from the bone marrow of a 53-year-old individual who was diagnosed with chronic myelogenous leukemia [34]. Chitosan is a natural biopolymer derived from chitin, which is found in the shells of crustaceans such as shrimp and crabs. It is known for its biocompatibility, biodegradability, and non-toxic nature [35]. Chitosan has gained attention in the field of nanomaterials due to its ability to modify the surface properties and enhance the stability of nanoparticles, as well as its potential to mitigate nanomaterial toxicity [36]. Cinteza *et al.* [37] demonstrated the use of chitosan as a capping agent for the synthesis of silver nanoparticles. They highlighted the role of chitosan in stabilizing and modifying the surface properties of nanoparticles, which can influence their toxicity. Lastly, exposure time is ranked within the top three most important descriptors across all models, consistent with the top 10 most important descriptors determined by QSAR models DT, kNN, PLS-DA, and RF developed before data clustering. As discussed previously, exposure time can be an important predictor of toxicity. Exposure time influences the duration of interaction between nanomaterials and biological systems, which can significantly impact the cellular response and potential adverse effects. The duration of exposure determines the extent of cellular uptake, intracellular accumulation, and subsequent biological interactions, which can ultimately determine the toxicity of nanomaterials.

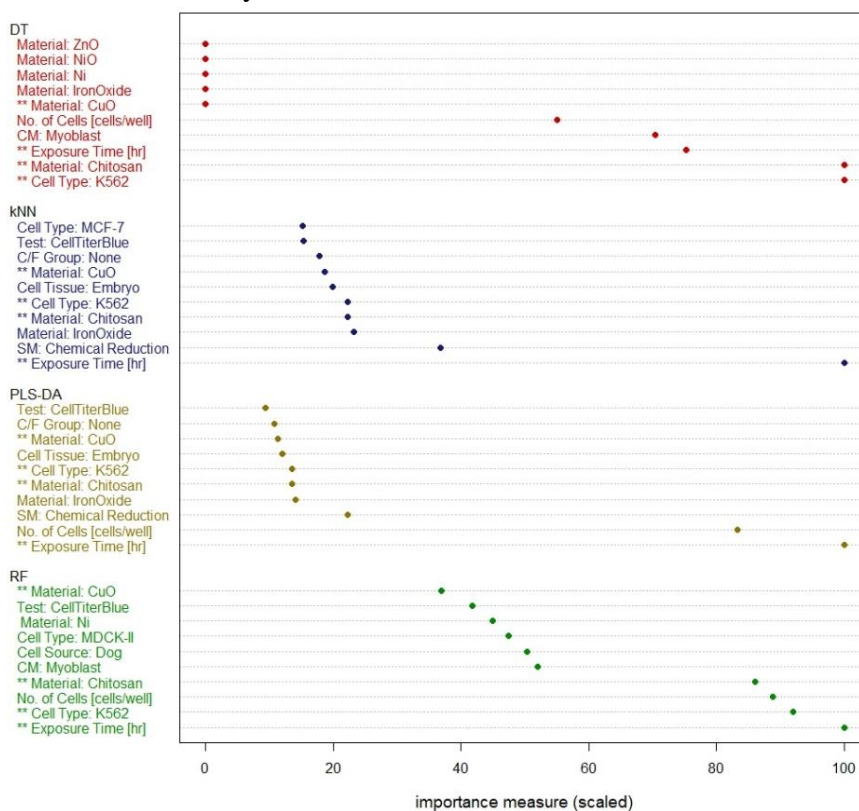


Figure 4. The top 10 descriptors were determined by each of the models developed from clustered data based on 'concentration < 31 µg/ml'. '**' = common across all models, SM = Synthesis Method, C/F = Coat/Functional Group, TI = Test Indicator, Time = Exposure Time, CM = Cell Morphology.

3.5. Applicability domain assessment.

The OECD guidelines require quantitative structure-activity relationship (QSAR) models to have a defined domain of applicability. In this study, the applicability domain was analyzed using PCA Bounding Box. Figure 5 displays the data points from both the training and testing sets of the original dataset and the clustered data based on concentration < 31 $\mu\text{g/ml}$. It is important to recall that the training set is where the predictive models were built, while the testing set is used for external validation. As shown in Figures 5a and 5b, all the testing data points fall within the boundary created by the training data points, indicating that they are within the applicability domain of the QSAR models developed.

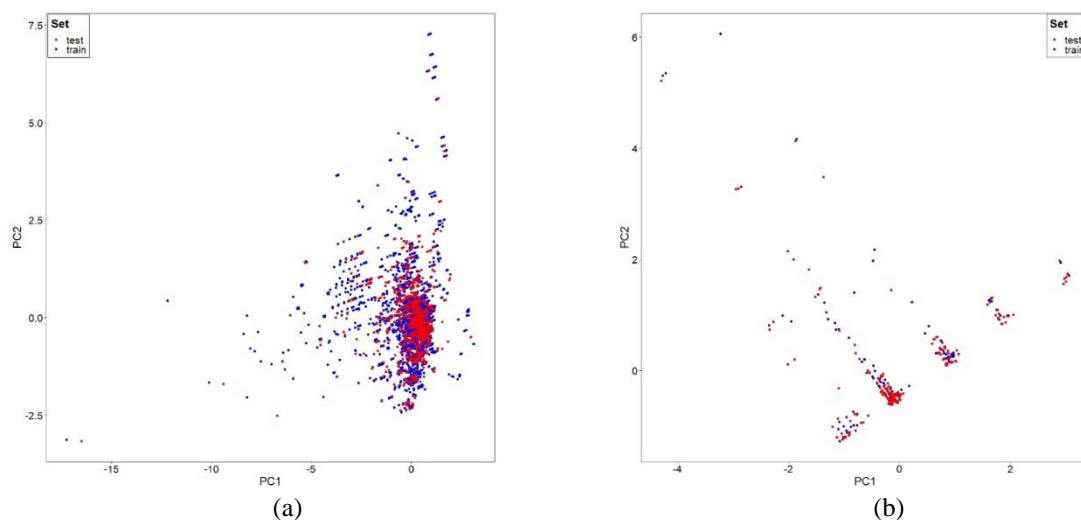


Figure 5. Applicability domain via PCA bounding box, where (a) is for the models built from the original dataset; (b) for the models developed from the cluster based on concentration < $\mu\text{g/ml}$.

4. Conclusions

This study set to develop QSAR models to evaluate the cytotoxicity level of a more comprehensive type of engineered NMs. The models built, namely PLS-DA, kNN, RF, and DT, were first trained and externally validated using the original dataset, which contained 4863 data points. As a result, the models demonstrated a predictive performance of over 90%. Among the descriptors, synthesis method ‘commercial’, concentration, and exposure time were determined to be the most important in the prediction of NMs cytotoxicity levels across all models.

Considering the diversity and extent of the original dataset, an intuitive data clustering approach was employed based on the most important descriptors across the four models. And QSAR models were further built for each cluster accordingly. Notably, the models developed using the cluster based on a ‘concentration < 31 $\mu\text{g/ml}$ ’, which contained 2339 data points, significantly improved the performance of PLS-DA, kNN, and DT models, matching RF with F1 scores no less than 95%. Additionally, exposure time, material ‘chitosan’, and cell type ‘k562’, are the descriptors identified to be the most influential across all models.

Hence, the study successfully built several highly accurate QSAR models that can effectively generalize to new data points. The results of this study complement the existing literature and could help assist with the safe design of various types of NMs. Moreover, further investigation into the specific mechanism of how the identified important descriptors influence NMs cytotoxicity is recommended.

Author Contributions

All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data supporting the findings of this study are available upon reasonable request from the corresponding author.

Funding

This research was funded by the Department of Science and Technology – Philippine Council for Industry, Energy, and Emerging Technology Research and Development (DOST-PCIEERD) under grant number 10185.

Acknowledgments

The authors are grateful for the support of the Department of Science and Technology - Philippine Council for Industry, Energy, and Emerging Technology Research and Development with grant no. 10185. This study was done at the Research on Environment and Nanotechnology Laboratories, Research Division, Mindanao State University at Naawan.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Singh, A.V.; Varma, M.; Laux, P.; Choudhary, S.; Datusalia, A.K.; Gupta, N.; Luch, A.; Gandhi, A.; Kulkarni, P.; Nath, B. Artificial intelligence and machine learning disciplines with the potential to improve the nanotoxicology and nanomedicine fields: a comprehensive review. *Arch Toxicol.* **2023**, *97*, 963–979. <https://doi.org/10.1007/s00204-023-03471-x>.
2. Domingues, C.; Santos, A.; Alvarez-Lorenzo, C.; Concheiro, A.; Jarak, I.; Veiga, F.; Barbosa, I.; Dourado, M.; Figueiras, A. Where Is Nano Today and Where Is It Headed? A Review of Nanomedicine and the Dilemma of Nanotoxicology. *ACS Nano* **2022**, *16*, 9994–10041, <https://doi.org/10.1021/acsnano.2c00128>.
3. Desai, A.; Ashok, A.; Edis, Z.; Bloukh, S.; Gaikwad, M.; Patil, R.; Pandey, B.; Bhagat, N. Meta-Analysis of Cytotoxicity Studies Using Machine Learning Models on Physical Properties of Plant Extract-Derived Silver Nanoparticles. *Int J Mol Sci.* **2023**, *24*, 4220, <https://doi.org/10.3390/ijms24044220>.
4. Ray, P.C.; Yu, H.; Fu, P.P. Toxicity and Environmental Risks of Nanomaterials: Challenges and Future Needs. *Journal of Environmental Science and Health Part C* **2009**, *27*, 1–35, <https://doi.org/10.1080/10590500802708267>.
5. Forest, V.; Hochepped, J.-F.; Pourchez, J. Importance of Choosing Relevant Biological End Points To Predict Nanoparticle Toxicity with Computational Approaches for Human Health Risk Assessment. *Chem Res Toxicol.* **2019**, *32*, 1320–1326, <https://doi.org/10.1021/acs.chemrestox.9b00022>.

6. Ji, Z.; Guo, W.; Wood, E.L.; Liu, J.; Sakkiyah, S.; Xu, X.; Patterson, T.A.; Hong, H. Machine Learning Models for Predicting Cytotoxicity of Nanomaterials. *Chem Res Toxicol* **2022**, *35*, 125–139, <https://doi.org/10.1021/acs.chemrestox.1c00310>.
7. Sang, L.; Wang, Y.; Zong, C.; Wang, P.; Zhang, H.; Guo, D.; Yuan, B.; Pan, Y. Machine Learning for Evaluating the Cytotoxicity of Mixtures of Nano-TiO₂ and Heavy Metals: QSAR Model Apply Random Forest Algorithm after Clustering Analysis. *Molecules* **2022**, *27*, 6125, <https://doi.org/10.3390/molecules27186125>.
8. Yuan, B.; Wang, P.; Sang, L.; Gong, J.; Pan, Y.; Hu, Y. QNAR modeling of cytotoxicity of mixing nano-TiO₂ and heavy metals. *Ecotoxicol Environ Saf.* **2021**, *208*, 111634, <https://doi.org/10.1016/j.ecoenv.2020.111634>.
9. Qi, R.; Pan, Y.; Cao, J.; Yuan, B.; Wang, Y.; Jiang, J. Toward comprehension of the cytotoxicity of heterogeneous TiO₂-based engineered nanoparticles: a nano-QSAR approach. *Environ Sci Nano.* **2021**, *8*, 927–936, <https://doi.org/10.1039/D0EN01266A>.
10. Shi, H.; Pan, Y.; Yang, F.; Cao, J.; Tan, X.; Yuan, B.; Jiang, J. Nano-SAR Modeling for Predicting the Cytotoxicity of Metal Oxide Nanoparticles to PaCa₂. *Molecules* **2021**, *26*, 2188, <https://doi.org/10.3390/molecules26082188>.
11. Bilgi, E.; Karakus, C.O. Machine learning-assisted prediction of the toxicity of silver nanoparticles: a meta-analysis. *Journal of Nanoparticle Research* **2023**, *25*, 157, <https://doi.org/10.1007/s11051-023-05806-2>.
12. Labouta, H.I.; Asgarian, N.; Rinker, K.; Cramb, D.T. Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature. *ACS Nano.* **2019**, *13*, 1583–1594, <https://doi.org/10.1021/acsnano.8b07562>.
13. Gul, G.; Yildirim, R.; Ileri-Ercan, N. Cytotoxicity analysis of nanoparticles by association rule mining. *Environ. Sci. Nano.* **2021**, *8*, 937–949, <https://doi.org/10.1039/D0EN01240H>.
14. Valeriano, A.; Bondaug, F.; Ebarido, I.; Almonte, P.; Sabugaa, M.A.; Bagnol, J.R.; Latayada, M.J.; Macalalag, J.M.; Paradero, B.D.; Mayes, M.; Balanay, M.; Alguno, A.; Capangpangan, R. Predicting cytotoxicity of engineered nanoparticles using regularized regression models: an in silico approach. *SAR QSAR Environ Res.* **2023**, *34*, 591–604, <https://doi.org/10.1080/1062936X.2023.2242785>.
15. Boehmke, B.; Greenwell, B.M. Hands-on machine learning with R. *CRC Press* **2019**, 484, <https://doi.org/10.1201/9780367816377>.
16. Lee, J.; Park, K. GAN-based imbalanced data intrusion detection system. *Pers Ubiquitous Comput.* **2021**, *25*, 121–128, <https://doi.org/10.1007/s00779-019-01332-y>.
17. Faisal, S.; Tutz, G. Imputation methods for high-dimensional mixed-type datasets by nearest neighbors. *Comput Biol Med.* **2021**, *135*, 104577, <https://doi.org/10.1016/j.compbimed.2021.104577>.
18. Sree, K.D.; Bindu, C.S. Data Analytics: Why Data Normalization. *International Journal of Engineering & Technology* **2018**, *7*, 209–213, <https://doi.org/10.14419/ijet.v7i4.6.20464>.
19. Kuhn, M.; Johnson, K. Feature Engineering and Selection: A Practical Approach for Predictive Models. *Chapman and Hall/CRC* **2019**, <https://doi.org/10.1080/00031305.2020.1790217>.
20. Muthukrishnan, R.; Rohini, R. LASSO: A feature selection technique in predictive modeling for machine learning. In *IEEE International Conference on Advances in Computer Applications (ICACA)* **2016**, 18–20, <https://doi.org/10.1109/ICACA.2016.7887916>.
21. Zheng, W.; Fu, X.; Ying, Y. Spectroscopy-based food classification with extreme learning machine. *Chemometrics and Intelligent Laboratory Systems.* **2014**, *139*, 42–47, <https://doi.org/10.1016/j.chemolab.2014.09.015>.
22. Brereton, R.G.; Lloyd, G.R. Partial least squares discriminant analysis: taking the magic away. *J. Chemom.* **2014**, *28*, 213–225, <https://doi.org/10.1002/cem.2609>.
23. Ruiz, J.-R.; Canals, T.; Cantero Gomez, R. Comparative Study of Multivariate Methods to Identify Paper Finishes Using Infrared Spectroscopy. *IEEE Trans Instrum Meas.* **2012**, *61*, 1029–1036, <https://doi.org/10.1109/TIM.2011.2173048>.
24. Kuhn, M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* **2008**, *28*, <https://doi.org/10.18637/jss.v028.i05>.
25. [OECD] Organisation for Economic Co-operation and Development: Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. *OECD* **2014**, <https://doi.org/10.1787/9789264085442-en>.
26. Phanus-umporn, C.; Shoombuatong, W.; Prachayasittikul, V.; Anuwongcharoen, N.; Nantasenamat, C. Privileged substructures for anti-sickling activity via cheminformatic analysis. *RSC Adv.* **2018**, *8*, 5920–5935, <https://doi.org/10.1039/C7RA12079F>.

27. Fröhlich, E. The role of surface charge in cellular uptake and cytotoxicity of medical nanoparticles. *Int J Nanomedicine* **2012**, *2012*, 5577-5591, <https://doi.org/10.2147/IJN.S36111>
28. Liu, Z.; Tabakman, S.; Welsher, K.; Dai, H. Carbon nanotubes in biology and medicine: In vitro and in vivo detection, imaging and drug delivery. *Nano Res.* **2009**, *2*, 85–120, <https://doi.org/10.1007/s12274-009-9009-8>.
29. Mironava, T.; Hadjiargyrou, M.; Simon, M.; Jurukovski, V.; Rafailovich, M.H. Gold nanoparticles cellular toxicity and recovery: Effect of size, concentration and exposure time. *Nanotoxicology* **2010**, *4*, 120–137, <https://doi.org/10.3109/17435390903471463>.
30. Nel, A.; Xia, T.; Mädler, L.; Li, N. Toxic Potential of Materials at the Nanolevel. *Science* **2006**, *311*, 622–627, <https://doi.org/10.1126/science.1114397>.
31. Pan, Y.; Leifert, A.; Ruau, D.; Neuss, S.; Bornemann, J.; Schmid, G.; Brandau, W.; Simon, U.; Jahn-Dechent, W. Gold Nanoparticles of Diameter 1.4 nm Trigger Necrosis by Oxidative Stress and Mitochondrial Damage. *Small* **2009**, *5*, 2067–2076, <https://doi.org/10.1002/sml.200900466>.
32. Li, X.; Xu, H.; Chen, Z.-S.; Chen, G. Biosynthesis of Nanoparticles by Microorganisms and Their Applications. *J Nanomater.* **2011**, *2011*, 1–16, <https://doi.org/10.1155/2011/270974>.
33. Chen, X.; Mao, S.S. Titanium Dioxide Nanomaterials: Synthesis, Properties, Modifications, and Applications. *Chem Rev.* **2007**, *107*, 2891–2959, <https://doi.org/10.1021/cr0500535>.
34. Skopek, R.; Palusińska, M.; Kaczor-Keller, K.; Pingwara, R.; Papierniak-Wyglądała, A.; Schenk, T.; Lewicki, S.; Zelent, A.; Szymański, Ł. Choosing the Right Cell Line for Acute Myeloid Leukemia (AML) Research. *International Journal of Molecular Sciences.* **2023**, *24*, 5377. <https://doi.org/10.3390/ijms24065377>.
35. Ahmed, S.; Ikram, S. Chitosan Based Scaffolds and Their Applications in Wound Healing. *Achievements in the Life Sciences* **2016**, *10*, 27–37, <https://doi.org/10.1016/j.als.2016.04.001>.
36. Rizeq, B.R.; Younes, N.N.; Rasool, K.; Nasrallah, G.K. Synthesis, Bioapplications, and Toxicity Evaluation of Chitosan-Based Nanoparticles. *Int J Mol Sci.* **2019**, *20*, 5776, <https://doi.org/10.3390/ijms20225776>.
37. Cinteza, L.; Scamorosenco, C.; Voicu, S.; Nistor, C.; Nitu, S.; Trica, B.; Jecu, M.-L.; Petcu, C. Chitosan-Stabilized Ag Nanoparticles with Superior Biocompatibility and Their Synergistic Antibacterial Effect in Mixtures with Essential Oils. *Nanomaterials* **2018**, *8*, 826, <https://doi.org/10.3390/nano8100826>.

Supplementary Materials

Table S1. Hyperparameter Tuning of Each Model During Model Building/Training Using the Original Dataset.

DT		kNN		PLS-DA		RF	
Maxdepth	%F1 _{tr}	k	%F1 _{tr}	ncomp	%F1 _{tr}	mtry	%F1 _{tr}
5	90.93	5	91.79	1	89.74	2	89.88
6	90.88	7	91.71	2	90.01	108	95.21
7	90.87	9	91.47	3	90.67	215	95.23
10	91.06	11	91.32	4	90.97		
11	91.11	13	91.15	5	90.94		
13	91.16	15	91.09	6	91.03		
18	91.56	17	91.25	7	90.95		
20	91.59	19	91.05	8	90.84		
25	91.59	21	91.11	9	90.93		
27	91.59	23	91.03	10	90.79		
30	91.59	25	91.06				
		27	90.90				
		29	90.81				
		31	90.84				
		33	90.78				
		35	90.71				
		37	90.65				
		39	90.59				
		41	90.61				
		43	90.50				

Highlighted Cell = optimum hyperparameter; Maxdepth = maximum depth of the tree.; k = number of nearest neighbors.; ncomp = number of latent variables.; mtry = number of features to consider when splitting a node.

Table S2. Hyperparameter Tuning of Each Model During Model Building/Training Using the Dataset Clustered Based on ‘Concentration < 31 µg/ml’.

DT		kNN		PLS-DA		RF	
Maxdepth	%F1 _{tr}	k	%F1 _{tr}	ncomp	%F1 _{tr}	mtry	%F1 _{tr}
1	95.16	5	94.35	1	0.9505	2	0.9515
2	95.15	7	94.70	2	0.9499	23	0.9511
5	95.00	9	94.96	3	0.9505	44	0.9499
9	95.00	11	95.25				
		13	95.25				
		15	95.15				
		17	95.15				
		19	95.10				
		21	95.07				
		23	95.05				
		25	95.07				
		27	95.05				
		29	95.05				
		31	95.05				
		33	95.05				
		35	95.05				
		37	95.05				
		39	95.05				
		41	95.05				
		43	95.05				

Table S3. Ranking of Descriptors for Each Model Built from the Original Dataset.

Rank	DT		kNN		PLS-DA		RF	
	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure
1	Exposure Time [hr]	100	Concentration [ug/ml]	100	Cell Type: HaCat	100	Concentration [ug/ml]	100
2	Concentration [ug/ml]	81.79	Exposure Time [hr]	56.16	Cell Tissue: Embryo	99.06	Size in Water [nm]	54.26
3	Diameter [nm]	67.47	Coat/Functional Group: None	39.64	Coat/Functional Group: None	98.98	Zeta in Water [mV]	44.61
4	Cell Tissue: Embryo	57.34	Size in Water. [nm]	34.57	Synthesis Method: Commercial	98.96	Diameter [nm]	36.89

Rank	DT		kNN		PLS-DA		RF	
	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure
5	Size in Water [nm]	50.69	Synthesis Method: Commercial	32.41	Exposure Time [hr]	94.98	Exposure Time [hr]	30.34
6	Coat/Functional Group: None	42.09	Zeta in Water [mV]	24.31	No. of Cells [cells/well]	88.69	No. of Cells [cells/well]	24.38
7	Coat Functional Group: FA	28.44	Test Indicator: Tetrazolium Salt	22.90	Concentration [ug/ml]	75.52	Synthesis Method: Commercial	15.93
8	Zeta in Water. [mV]	27.98	Material: Iron Oxide	20.59	Test Indicator: Tetrazolium Salt	72.16	Cell Tissue: Embryo	13.62
9	Synthesis Method: Commercial	27.90	Material: Au	20.58	Cell Type: NIH3T3	69.94	Material: ZnO	8.58
10	Test Indicator: Tetrazolium Salt	25.48	Cell Type: HaCat	19.39	Material: Au	68.89	Coat/Functional Group: FA	7.99
11	Cell Type: NHDF	22.27	Cell Tissue: Embryo	17.51	Cell Tissue: Cervix	68.39	Material: CuO	7.90
12	No. of Cells. [cells/well]	22.17	Cell Tissue: Cervix	16.78	Cell Type: HeLa	62.56	Cell Type: A549	6.48
13	Cell Type: LNCaP	20.47	Cell Type: HeLa	15.61	Zeta in Water [mV]	57.24	Cell Morphology: Myoblast	5.65
14	Cell Type: HaCat	18.14	Cell Tissue: Bone Marrow	12.89	Material: CuO	57.15	Cell/Type: HaCat	5.52
15	Cell Type: NIH3T3	18.00	Cell Type: NIH3T3	11.79	Size in Water [nm]	53.39	Coat/Functional Group: None	4.95
16	Material: ZnO	16.75	Test: MTS	10.99	Synthesis Method: Chemical Reduction	52.79	Test Indicator: Tetrazolium Salt	4.84
17	Cell Tissue: Cervix	14.42	Cell Type: SHSY5Y	10.20	Cell Age: Fetus	52.51	Test: CCK.8	4.62
18	Test: NRU	14.35	Material: ZnO	9.93	Material: ZnO	51.22	Synthesis Method: Multiple Emulsion Method	3.65
19	Cell Type: NR8383	13.81	Type: O	9.85	Material: PLGA	48.95	Synthesis Method: Green Synthesis	3.41
20	Cell Morphology: Myoblast	10.58	Shape: Irregular	9.27	Synthesis Method: Emulsion Solvent Evaporation.	48.23	Cell Type: MCF.7	3.21

Table S4. Ranking of Descriptors for Each Model Built from Clustered Data Based on ‘Concentration < 31 µg/ml’.

Rank	DT		kNN		PLS-DA		RF	
	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure
1	Cell Type: K562	100	Exposure time [hr]	100	Exposure time [hr]	100	Exposure time [hr]	100
2	Material: Chitosan	100	Synthesis Method: Chemical Reduction	36.74	No. of Cells [cells/well]	83.31	Cell Type: K562	91.96
3	Exposure time [hr]	75.24	Material: Iron Oxide	23.26	Synthesis Method:	22.28	No. of Cells [cells/well]	88.80

Rank	DT		kNN		PLS-DA		RF	
	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure	Descriptors	Importance Measure
					Chemical Reduction			
4	Cell Morphology: Myoblast	70.37	Material: Chitosan	22.23	Material: Iron Oxide	14.11	Material: Chitosan	86.06
5	No. of Cells [cells/well]	55.07	Cell Type: K562	22.23	Material: Chitosan	13.48	Cell Morphology: Myoblast	52.02
6	Material: CuO	0.00	Cell Tissue: Embryo	19.89	Cell Type: K562	13.48	Cell Source: Dog	50.31
7	Material: Iron Oxide	0.00	Material: CuO	18.63	Cell Tissue: Embryo	12.06	Cell Type: MDCK.II	47.39
8	Material: Ni	0.00	Coat/Functional Group: None	17.76	Material: CuO	11.30	Material: Ni	45.01
9	Material: NiO	0.00	Test Cell: Titer Blue	15.38	Coat/Functional Group: None	10.77	Test Cell: TiterBlue	41.73
10	Material: ZnO	0.00	Cell Type: MCF.7	15.14	Test Cell: Titer Blue	9.32	Material: CuO	36.96
11	Shape: Star	0.00	Cell Tissue: Colon	14.36	Cell Type: MCF.7	9.18	Cell Type: MCF.7	36.73
12	Coat/Functional Group: CKK	0.00	Cell Morphology: Myoblast	13.63	Cell Tissue: Colon	8.70	Coat/Functional Group: CKK	33.94
13	Coat/Functional Group: GSH	0.00	Shape: Star	12.37	Cell Morphology: Myoblast	8.27	Coat/Functional Group: MSA	32.65
14	Coat/Functional Group: Liposome	0.00	Coat/Functional Group: PVP	12.22	Shape: Star	7.50	Cell Type: L6	30.05
15	Coat/Functional Group: MSA	0.00	Coat/Functional Group: CKK	11.93	Coat/Functional Group: PVP	7.41	Synthesis Method: Chemical Reduction	25.49
16	Coat/Functional Group: None	0.00	Synthesis Method: Commercial	11.45	Coat/Functional Group: CKK	7.23	Cell Type: hPDLF	22.33
17	Coat/Functional Group: OH	0.00	Cell Type: MDCK.II	10.81	Synthesis Method: Commercial	6.94	Cell Tissue: Embryo	21.46
18	Coat/Functional Group: PEI	0.00	Cell Source: Dog	10.81	Cell Type: MDCK.II	6.56	Cell Tissue: Muscle Tissue	20.51
19	Coat/Functional Group: PVP	0.00	No. of Cells [cells/well]	10.53	Cell Source: Dog	6.56	Cell Tissue: Teeth	20.27
20	Synthesis Method: Chemical Reduction	0.00	Cell Type: hPDLF	10.47	Cell Type: hPDLF	6.35	Coat/Functional Group: PEI	20.11

Publisher's Note & Disclaimer

The statements, opinions, and data presented in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect the views of the publisher and/or the editor(s). The publisher and/or the editor(s) disclaim any responsibility for the accuracy, completeness, or reliability of the content. Neither the publisher nor the editor(s) assume any legal liability for any errors, omissions, or consequences arising from the use of the information presented in this publication. Furthermore, the publisher and/or the editor(s) disclaim any liability for any injury, damage, or loss to persons or property that may result from the use of any ideas, methods, instructions, or products mentioned in the content. Readers are encouraged to independently verify any information before relying on it, and the publisher assumes no responsibility for any consequences arising from the use of materials contained in this publication.