

ULRR

Employing supervised algorithms for the prediction of nanomaterial's antioxidant efficiency

Item Type	Article
Authors	Mirzaei, Mahsa;Furxhi, Irini;Murphy, Finbarr;Mullins, Martin
Citation	International Journal of Molecular Sciences, 24, 2792
Download date	2026-05-11 23:44:29
Item License	https://creativecommons.org/licenses/by-nc-sa/4.0/
Link to Item	https://doi.org/10.34961/researchrepository-ul.24049893



Article

Employing Supervised Algorithms for the Prediction of Nanomaterial's Antioxidant Efficiency

Mahsa Mirzaei ¹, Irini Furxhi ^{1,2,*}, Finbarr Murphy ¹ and Martin Mullins ¹

¹ Department of Accounting and Finance, Kemmy Business School, University of Limerick, V94PH93 Limerick, Ireland

² Transgero Limited, Newcastle West, V42V384 Limerick, Ireland

* Correspondence: irini.furxhi@ul.ie; Tel.: +353-85-106-9771

Abstract: Reactive oxygen species (ROS) are compounds that readily transform into free radicals. Excessive exposure to ROS depletes antioxidant enzymes that protect cells, leading to oxidative stress and cellular damage. Nanomaterials (NMs) exhibit free radical scavenging efficiency representing a potential solution for oxidative stress-induced disorders. This study aims to demonstrate the application of machine learning (ML) algorithms for predicting the antioxidant efficiency of NMs. We manually compiled a comprehensive dataset based on a literature review of 62 in vitro studies. We extracted NMs' physico-chemical (P-chem) properties, the NMs' synthesis technique and various experimental conditions as input features to predict the antioxidant efficiency measured by a 2,2-diphenyl-1-picrylhydrazyl (DPPH) assay. Following data pre-processing, various regression models were trained and validated. The random forest model showed the highest predictive performance reaching an $R^2 = 0.83$. The attribute importance analysis revealed that the NM's type, core-size and dosage are the most important attributes influencing the prediction. Our findings corroborate with those of the prior research landscape regarding the importance of P-chem characteristics. This study expands the application of ML in the nano-domain beyond safety-related outcomes by capturing the functional performance. Accordingly, this study has two objectives: (1) to develop a model to forecast the antioxidant efficiency of NMs to complement conventional in vitro assays and (2) to underline the lack of a comprehensive database and the scarcity of relevant data and/or data management practices in the nanotechnology field, especially with regards to functionality assessments.

Keywords: nanomaterials; nanoparticles; machine learning; antioxidant; reactive oxygen species



Citation: Mirzaei, M.; Furxhi, I.; Murphy, F.; Mullins, M. Employing Supervised Algorithms for the Prediction of Nanomaterial's Antioxidant Efficiency. *Int. J. Mol. Sci.* **2023**, *24*, 2792. <https://doi.org/10.3390/ijms24032792>

Academic Editors: Ilaria Rea, Nicola Borbone, Monica Terracciano and Chiara Tramontano

Received: 15 December 2022

Revised: 25 January 2023

Accepted: 29 January 2023

Published: 1 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. 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/>).

1. Introduction

Reactive Oxygen Species (ROS) are molecules that contain one or more unpaired electrons in their orbitals and can lead to the formation of free radicals, e.g., superoxide ($O_2^{\bullet-}$), hydroxyl ($\bullet OH$), peroxy ($ROO\bullet$) or hydrogen peroxide H_2O_2 . ROS are generated during normal cellular functions as a by-product of aerobic metabolism and are essential to regulate physiological functions in biological systems [1]. However, exposure to exogenous factors such as heavy metals or ionizing radiation can trigger the excessive generation of ROS. The disruption of the equilibrium of ROS leads to oxidative modifications of biological systems at a molecular level, causing damage to cellular components, such as DNA, lipids and proteins. Moreover, it leads to the acceleration of cellular death which underlies several diseases [2]. Therefore, it is necessary to reduce or eliminate the excess of free radicals. Free radicals can be eliminated or detoxified by a selection of antioxidative enzymes or antioxidants [3]. Therapeutic approaches for preventing oxidative damage are currently at the core of medical research, including the use of antioxidants [4].

The antioxidant efficiency (anti-ox) is defined as the ability of molecules to scavenge the free radicals. Antioxidant compounds (natural or synthetic) control autoxidation by preventing the formation of free radicals or interrupting their spread by neutralizing them

which delays or inhibits the damage to the target cells, decreasing oxidative stress, enhancing immune function and increasing longevity [5,6]. Numerous studies have investigated the anti-ox properties of various materials [7–9]. Conventional antioxidants are designed to reduce ROS in biological systems [10]. However, clinical investigations have shown that traditional molecules provide little to no protection or even, in some instances, increase the risk of mortality [11]. Ghorbani [12] demonstrated that antioxidants show inadequate effects due to their unspecific distribution and poor retention in disease sites. Therefore, the need for an efficient, safe yet more specific antioxidant is undeniable.

Nanotechnology has gained tremendous attention due to its potential to treat diseases and it is growing rapidly within the medicinal disciplines [13–15]. Nanomaterials (NMs) are particles with unique features such as a high surface-to-volume ratio, biocompatibility, redox and catalytic properties and are used in many fields such as the pharmaceutical industry, textile, cosmetics, food, water treatment, and particularly in health-related biological fields [16–19]. Nanomedicines overcome limitations such as non-specific distribution, high renal clearance, low delivery efficiency and inadequate in vivo bioavailability [20,21].

There are numerous organic and inorganic NMs exhibiting strong ROS-scavenging activities [22,23]. Studies have demonstrated promising results in developing a novel class of antioxidants (nano-antioxidants) to overcome the challenges associated with conventional antioxidants [12]. The potential use of NMs to reduce damages related to oxidative stress and prevent diseases is the subject of intense scientific research [20,24]. The potential use of NMs in treatments of degenerative Alzheimer's disease, cancer, metabolic diseases and chronic inflammatory diseases has been reported [25,26]. Ghorbani [12] demonstrated that NMs have stable anti-ox activity with specific pharmacokinetic actions, intrinsic ROS-scavenging properties and fewer adverse effects.

In recent decades, there has been considerable interest in digitising the information in medical and pharmaceutical disciplines and the use of such data to explain biological phenomena [27]. As more data becomes available, the application of artificial intelligence (AI) becomes increasingly valuable for analysing the data. AI technology, implemented in various fields, e.g., molecular medicine, drug development and nanotechnology, provides great reproducibility and timeliness to accomplish a variety of tasks [28–30]. This approach is data intensive [31] and has been applied in fields of medicine including disease prediction [32], the discovery of new medicine [33], the prediction of drug responses [34], medicine's toxicity [35], nanotoxicity [36], antimicrobial resistance [37], antibiotic resistance [38] and the antibacterial capacity of NMs [39], to mention a few. In the era of data, machine learning (ML) models, a sub-field of AI, can improve the predictions of toxicological and functional effects and provide insights into the design of NMs and/or strategies to eliminate exposure and minimize risks, contributing to sustainable and safe-by-design approaches (SSbD) [31,40,41]. The ML approach is a promising tool as it is time- and cost-effective reducing the number of tests required while allowing the researcher to be more effective by providing them precise forecasts. Furthermore, the traditional risk assessment is transforming into digital frameworks consisting of new approach methodologies (NAMs) without the need for animal testing [42,43].

ML models have been used to predict anti-ox activities, for example, Karydas [44] successfully developed an ML model to assess cherries' free radical scavenging capacity where in their analysis they utilised the 2,2-diphenyl-1-picrylhydrazyl (DPPH) and folin-ciocalteu (FCR) lowering capacity as outputs while sensing indices, soil attributes, climatic, topographic and hydrographic data were used as input features. They employed and achieved an extreme gradient boosting (XGBoost) algorithm with an accuracy of $R^2 = 0.89$. They employed important attribute analysis using the F-score method and determined the variables that affect the antioxidant capacity of cherries. Another example is from Idowu and Fatokun [45], which examined supervised and unsupervised algorithms to correlate polyphenol's antioxidant activity (output feature) to molecular descriptors, with parameters regulating in vivo antioxidant behaviour (input features). Olsen [46] demonstrated the use of a deep convolutional neural network (CNN) to predict the anti-ox efficiency measured

by a DPPH assay based on the properties of 687 peptides. They demonstrated that the model can be used to estimate the potential of a protein's peptides to scavenge free radicals.

Given the time-consuming, expensive and work-intensive nature of NMs' synthesis, having a computational method that can predict the anti-ox properties is beneficial. In addition, due to the large number of NMs' P-chem possible combinations and the impact of experimental variables, it is not feasible to evaluate all of them using conventional laboratory tests to determine the ones with the highest capacities. In this work, we manually compiled a dataset derived from a literature review of in vitro studies in different mediums that can be used by other researchers to train models, augment the dataset with additional features, fill data gaps through grouping approaches and predict the efficiency of NMs to scavenge free radicals. Despite the limited data available [47], we explored ML tools and we were able to predict the anti-ox efficiency of NMs and we also demonstrated which attributes have a substantial impact on the predictions. To the best of our knowledge, no study has been undertaken on the use of ML tools to predict the scavenging efficiency of NMs. This novel study contributes significantly to the existing research by proposing a model that can fit the SSbD framework, in the functionality dimension [31]. This tool utilizes P-chem properties, exposure conditions and the method of NMs' synthesis preparations. A random forest algorithm demonstrated the highest accuracy, reaching an $R^2 = 0.83$.

2. Results

2.1. Data Pre-Processing

Due to high missing values among the input-related data collected, we selected NMs' type, core size, shape, dosage, coating as well as the synthesis process, medium used, absorbance and duration as the input features for the model. This data cleansing resulted in a final dataset consisting of 10 columns and 1027 rows. Table 1 shows the percentages of missing values among the initial input variables: core size (3%), HDS (78.30%), shape (15.68%), EE (90.34%), ZP (65.09%), PDI (82.05%), SA (99.51), dosage (9.86%), temperature (76.33%), pH (98.62%) and thermal stability (86.59%).

Table 1. The primary and final input variables in dataset I and dataset II.

Category	Variables	Type	DATASET I		DATASET II		
			Min-Max or Labels	Data Transformation	Methods of Determination	Missing Values %	
Physico-chem properties	Core Size	Numeric	3.7–313 (nm), NaN	Selected	3.7–313 (nm), NaN	SEM, EDX, TEM, DLS, etc.,	3.06
	Hydrodynamic Size		32–752 (nm), NaN		–	DLS	78.30
	Encapsulation Efficiency		59.7–237 (%), NaN	Eliminated due to high NaN	–	DLS	90.34
	Zeta Potential		–20.6–46.8 (mV), NaN		–	DLS	65.09
	Polydispersity Index		0.145–1.76, NaN		–	DLS	82.05
	Surface Area		52.2–135.5 (m ² /g), NaN	–	–	–	99.51
	Dosage		0–200,000 (µg/mL), NaN	Selected	0–200,000 (µg/mL), NaN	–	9.86
	Synthesis Technique	Green, Chemical	Selected	Green, Chemical	–	–	
	Shape	Nominal	Spherical, Hexagonal, Rod, Spherical/Rod, Dendrimer, Triangular, Cubic	Selected	Spherical, Hexagonal, Rod, Spherical/Rod, Dendrimer, Triangular, Cubic	TEM, HRTEM, SEM, FESEM, EDX, XRD	15.68
	NMs' Type		Ag, ZnTe, CdTe, Chs, Se, Nd/ZnO, Au, Cu, CdSe, Chs/AL, AL, NaCas/GA, TiO ₂ , MgO, etc.,	Selected	Ag, ZnTe, CdTe, Chs, Se, Nd/ZnO, Au, Cu, CdSe, Chs/AL, etc.,	XRD, EDX, SEM, FESEM, TEM, HRTEM, etc.,	–
Coating		Allium sativum, areca catechu nut, Carboxymethyl cellulose (CMC), carboxymethyl cellulose (CMC)/curcumin, CCS, etc.,	Simplified: Transformed into Binary	Coated, Uncoated	SEM, EDX, TEM, DLS, XRD, HRTEM, STEM, FTIR	–	

Table 1. Cont.

Category	Variables	Type	DATASET I		DATASET II		
			Min-Max or Labels	Data Transformation	Methods of Determination	Missing Values %	
Exposure	Medium Used for Antioxidant Evaluation	Nominal	Extract, Solution, Emulsion, Food, Nanocomposites, Coating films	Selected	Extract, Solution, Emulsion, Food, Nanocomposites, Coating films	–	–
	Temperature	Numeric	0–37 (°C), NaN	Eliminated due to high NaN	–	–	76.33
	pH		7–7.4, NaN		–	–	98.62
	Thermal Stability		12–600, NA	–	TGA	86.59	
	Spectrophotometer Absorbance		517–540, NaN	Selected	517–540, NaN	–	–
	Duration		0–2190 (h), NaN	Selected	0–2190 (h), NaN	–	–

The HDI, EE, PDI, SA, ZP, temperature, pH and thermal stability columns were excluded due to the presence of numerous missing values. Null data in core size, dosage and shape were replaced with the mean (for numeric features) or mode value (for nominal features) [48]. Additionally, to avoid overfitting the model, the coating was transformed to a binary feature (coated and uncoated).

2.2. Model Validation

Following data homogenization and pre-processing, we trained and compared several regression models. The models are trained with the training dataset comprising 70% of the data (719 rows). Internal 10-fold cross validation is performed to optimize the hyper parameters and to ensure the models' robustness and goodness-of-fit. Following the model training, we performed model validation using 308 rows out of 1027 (30% of data) as a test set (external validation) to evaluate the final tuned predictivity. As shown Figure 1, the external validation of the RF model outperformed the other regression models exhibiting the lowest error with MAE at 7.8, RMSE at 12.7 and the highest accuracy displaying a $R^2 = 0.83$.

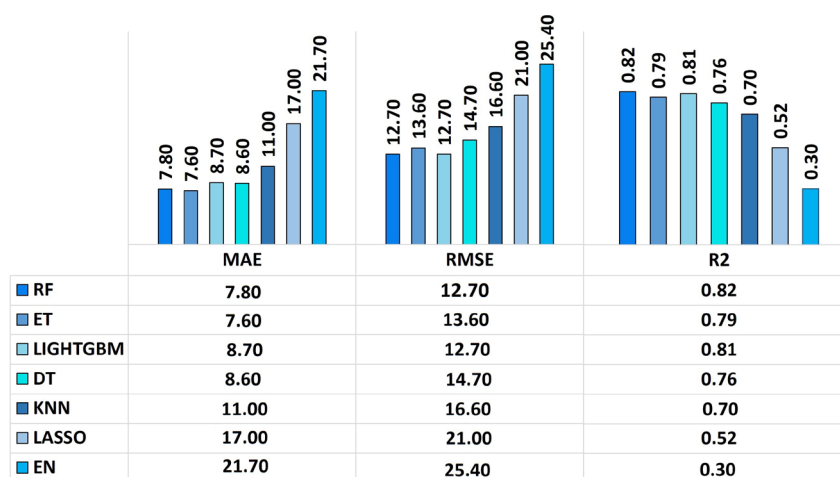


Figure 1. External models' validation metrics from the test dataset (30%). Comparison of the seven regression models (columns used for this evaluation includes type of NMs, medium used, core-size, shape, synthesis techniques, coating, dosage, absorbance, duration).

Nonlinear regression models (RF and ET) outperform linear models (LASSO and EN) because they exhibit more parameters (properties of training data learned during the learning process). RF is excellent at processing categorical, binary and numeric features, making it ideal for working with high-dimensional data with minimal pre-processing requirements.

Figure 2 presents the validation curve of the RF algorithm derived from the internal validation dataset (70%), the values of the training and cross-validation score are close, indicating the model goodness of fit and robustness of the training model.

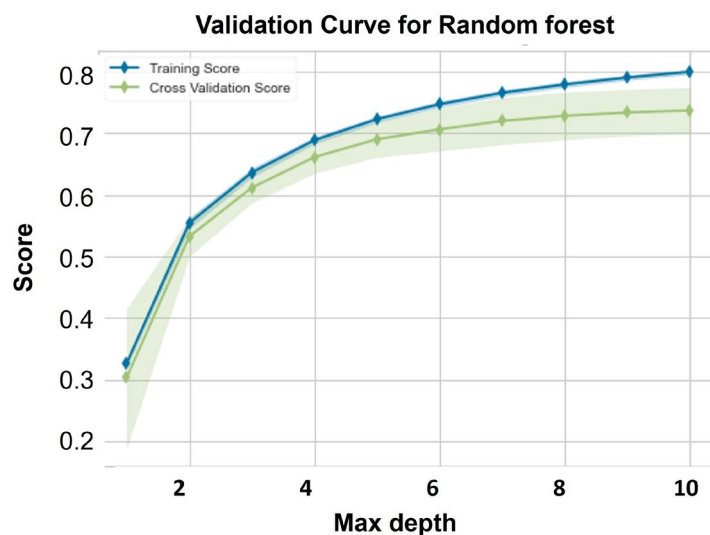


Figure 2. The internal model validation curve of RF model for the training score and the cross-validation score. x-axis represents the hyperparameter of the tree depth, y-axis represents the score (accuracy) of the model.

Depicted in Figure 3 are the prediction error of the predicted values against the actual values. It shows that the predicted values produced by our model (0.83) are close to the observed targets from the dataset. In regression models, prediction error measures how effectively the model predicts the response variable and as can be seen, the predicted values are relatively close to the actual values.

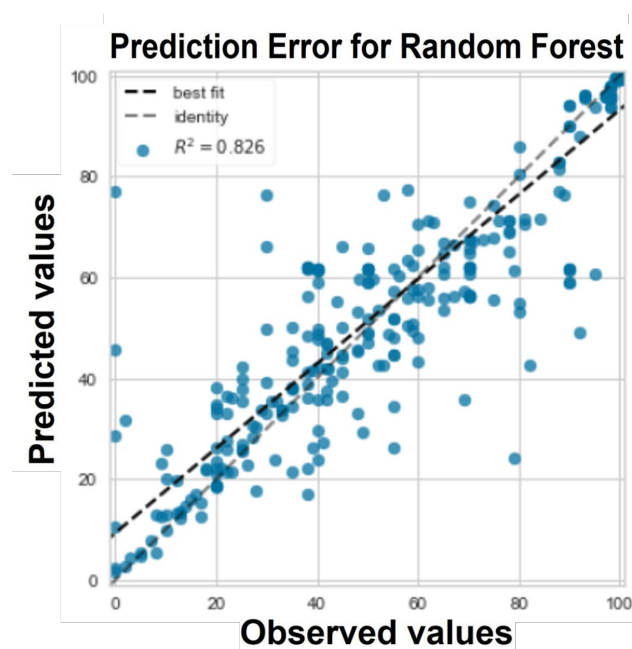


Figure 3. The prediction error by the RF regression model, the x-axis represents the actual (observed) values and the y-axis represents the predicted values.

2.3. Model Prediction

Attribute Importance

The feature importance in the RF model is calculated using the impurity Gini-based feature importance (or Mean Decrease in Impurity). Figure 4 presents the attribute importance analysis performed by RF. As illustrated, NMs' type, core-size and dosage appeared as the top three most important attributes influencing the prediction of the outcome.

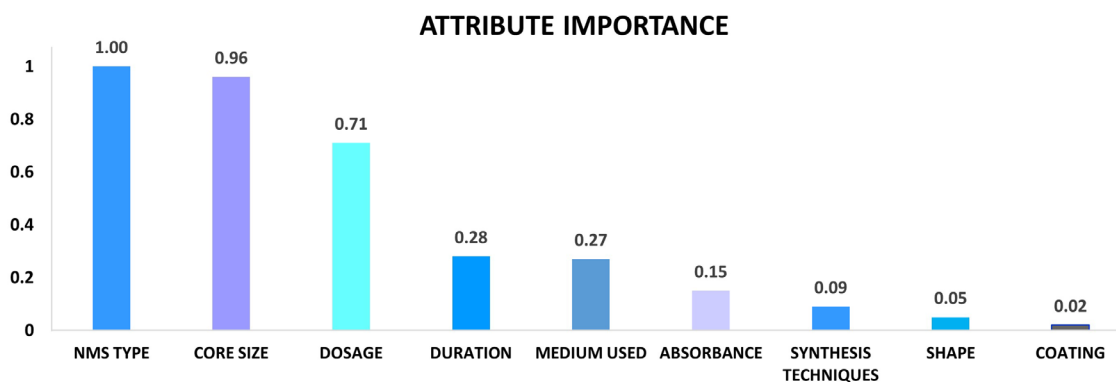


Figure 4. Attribute importance generated by random forest model, y-axis represents the variable importance from 0 to 1, x-axis represents the attributes.

The duration of exposure, the medium used and the absorbance are relatively important attributes that affect the prediction. The synthesis techniques, shape and coating (presence or absence) have relatively less effect on the prediction.

3. Discussion

The purpose of this research was initially to (i) gather data from the literature related to the anti-ox efficiency of NMs, (ii) to present the landscape of data gaps related to functionality assessments, (iii) to apply a tool to demonstrate the feasibility of ML for this outcome and iv) to identify significant factors influencing the prediction of the outcome.

Tremendous efforts are being put together to create databases with complete datasets and also many initiatives to promote the FAIR (findable, accessible, interoperable and reusable) principles are being created in the nano-domain [49,50]. To the best of our knowledge, there is no publicly accessible database that collects and summarizes data on the P-chem characteristics, functionality, application and production of NMs and at the same time links this information to safety aspects. Such a database could enhance the correlation between functionality and safety and thus support the SSbD paradigms [51]. In this study, we manually built a dataset that adheres to the FAIR principles: it is findable (10.5281/zenodo.6584826, accessed on 26 May 2022), accessible (open access), interoperable (containing the methodology and source of data as well as experimental parameters) and reusable in order to aid data scientists to accelerate the nano-informatics field. We collected data on P-chem experimental settings, such as TEM, SEM, FESEM, HRTEM, XRD, EDX, UV-vis spectroscopy and FTIR, and employed them to measure and evaluate intrinsic properties such as size, morphology, chemical composition and extrinsic properties such as encapsulation efficiency, polydispersity index and zeta potential. The various methodologies have an impact on the features' quality and homogeneity of information. However, due to the lack of data, we utilized all the available information. When more data becomes available and FAIR, attributes generated through one technique will reflect more scientifically sound choices for these modeling purposes.

To evaluate the free radical scavenging capacity, we gathered data from six different methods such as DPPH, ABTS, NO, H₂O₂, OD and metal chelating. However, we selected the DPPH assay as the outcome due to substantial missing values among the other techniques. According to the literature, ABTS and DPPH assays are recommended with

ABTS considered lengthy and expensive. It is challenging to conclude which approach is the most suitable for assessing the anti-ox activity. Bedlovičová [52] examined DPPH and ABTS assays, and they concluded that DPPH assay is inexpensive and simpler. When more information becomes available it could be feasible to utilize all the assays under one dataset if the parameters affecting the measurements become transparent with an increased variability due to data size. Another factor regarding the quality of data is the comparable measurements (unit metrics), which makes it difficult to merge a homogeneous dataset that contains all the knowledge gained through valuable experimental research. This may obstruct the advancement of a data-driven research. Thus, our work points out the shortcoming in current data management practices and reporting schemes in the field of nanotechnology. A uniform and homogenous system for reporting data on properties and experimental conditions will help to improve the quality of data and completeness.

Regarding P-chem characteristics, only 35%, 22%, 10%, 18% and 1% of the selected studies reported ZP, HDS, EE, PDI and SA of NMs, respectively. Therefore, due to a substantial number of missing values, we were unable to incorporate all these data. Therefore, for this study we employed the type of NM, medium used, core-size, shape, synthesis techniques, coating, dosage, absorbance, duration and DPPH variables. However, the performance of the models was compared by including ZP values computed with an ML missing values imputation methodology, and/or including shape (Supplementary Materials, Tables S1–S3). These properties are important in determining the activity of NMs. Regarding the core size, dosage and shape, the missing values were replaced with the mean or mode value. Thus, we stress that future research should report as many P-chem properties as feasible to allow the usage of intrinsic and/or extrinsic properties information. For experimental settings, we obtained data on the medium used, temperature, pH, thermal stability, spectrometers absorbance and exposure duration. However, the majority of them were excluded due to substantial number of missing values.

We utilized a model that adheres to OECD (organization for economic co-operation and development) principles (https://github.com/mahsa-mirzaei/NMAOP_M-M_wo-ZP-/blob/main/AOP_M%26M_wo%20ZP%20.ipynb, accessed on 18 January 2023), has a well-defined approach and is validated and built to an explicit endpoint derived from one assay [53]. The RF model surpasses all other models, with the highest R^2 value, $R^2 = 0.83$. In the nano domain, ML tools have been implemented to predict and assess the toxicity and functionality of NMs [54]. Furxhi [55] investigated NF's toxicity, whereas Mirzaei [39] and Mirzaei [56] assessed the antibacterial properties of NMs as such or coated on textiles, respectively. However, to our knowledge, no similar studies that predict the anti-ox capacity of NMs have been published. This approach is the initial step toward assisting researchers in discovering NMs with a high capacity and determining their P-chem properties and/or experimental settings. Despite these limitations, our findings are congruent with the previous literature [57–59]. The RF model's attribute importance analysis showed P-chem characteristics including type and core-size as influential variables on the prediction of the outcome, followed by shape, zeta potential and synthesis techniques. Researchers have investigated the anti-ox efficiency of numerous NMs and discovered that properties including particle core size, chemical composition, surface charge and surface coating have an impact on their anti-ox capacity [60–63]. According to Sharpe [57], the type of NM, core-size and dosage are among the most pivotal components. Fafal [59] stated that different synthesis techniques can affect the anti-ox capacity of NMs, they demonstrated that green techniques could have relatively higher activity. Nevertheless, the other characteristics such as ZP, EE and PDI can be equally valuable and play an important role. A study by Ruktanonchai [64], indicated that the ZP is a useful tool to predict the physical stability of NMs, and the PDI can show the size distributions. Kim [58] identified that particle-size, ZP, EE and PDI can enhance the release property and as a result improve the activity. One limitation of our study is the categorical representation of NMs in our dataset which are converted into binary representation through one-hot encoding. The categorical attributes can be changed only when the chemical composition of the NMs is available or swapped

with theoretical descriptors. Future studies should focus on the accurate representation of the intrinsic properties related to the composition, in order to be used with ML applications.

Significant numbers of NMs have been evaluated to date, resulting in a large pool of information scattered through the literature; it becomes increasingly challenging to analyze all NMs using traditional test methods. Accordingly, the demand for computational methods increases. We investigated inorganic and organic NMs to develop a model with good predictability ($R^2 = 0.83$). In addition to complementing conventional *in vitro* testing, this approach is substantially more cost-effective and time-efficient. The capacities of algorithms can be exploited to fulfill new approach methodologies [65]. Once safety can be integrated with the functionality of NMs, the modeling tools will allow the development and optimization of NMs with desired properties while ensuring worker, public and environmental safety, and functionality.

4. Materials and Methods

Figure 5 depicts an overview of the workflow followed. First, a literature review of studies investigating the efficiency of NMs via *in vitro* experiments was performed. P-chem properties of NMs, exposure conditions, measurement equipment-related features, characterization methodologies and assays used for assessing the anti-ox efficiency were extracted. Following an investigation of the dataset's completeness, data pre-processing was carried out, including normalization [66], one-hot encoding [67] and data split [68]. Several regression models were trained and validated using a variety of performance parameters in order to identify a model with high predictability. To train ML models, we used Scikit-learn and PyCaret library [69]. Finally, importance attribute analysis, a supervised technique that identifies and ranks the variables was used [70].

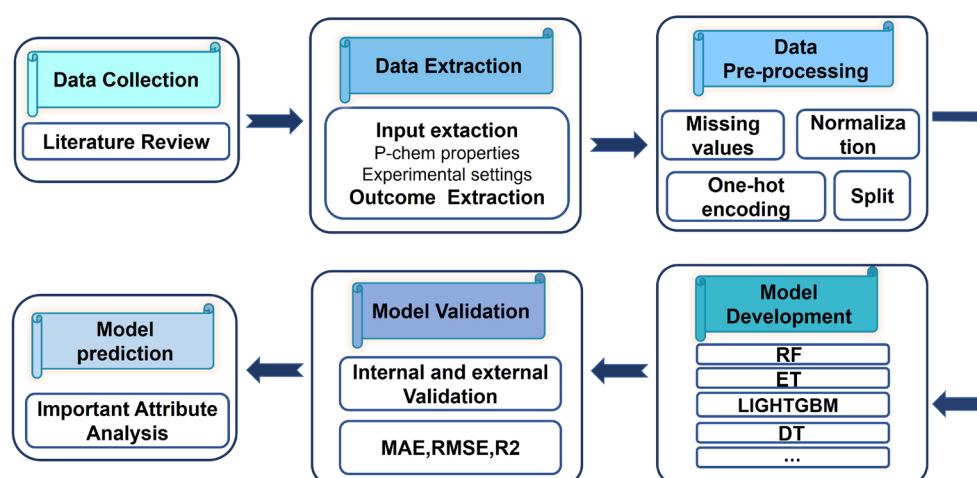


Figure 5. Model development workflow.

4.1. Data Collection

In order to compose a high-quality dataset, a great amount of data must be integrated in a harmonized fashion which is the most challenging aspect of the ML workflow. A literature search was conducted for studies that investigated NMs' antioxidant efficiency with quantitative results from 2013 to 2021. Following keywords were used 'antioxidants', 'free radicals', 'ROS', 'Oxidative stress', 'RNS', 'nanoparticles', 'nanofoms' and 'nanomaterials'. We selected 73 studies that investigated different organic and inorganic NMs with *in vitro* experiments. We excluded 11 studies due to binary or qualitative results. From the studies gathered, the following information was manually extracted:

Input-related collection: (1) information related to NMs' P-chem properties such as the NM type, core-size, coating, hydrodynamic size (HDS) and its polydispersity index (PDI), shape, surface area (SA), zeta potential (ZP) and encapsulation efficiency (EE); (2) ex-

perimental conditions such as the methods of NMs' synthesis such as chemical (top-down and bottom-up approach) and green (bottom down approach employing natural extracts or microorganisms and biological components as reducing agents), the culture medium for the anti-ox activity evaluation, information related to the application of the NF (pharmaceutical and food packaging, nanomedicine, etc.), pH, temperature, thermal stability, absorbance and duration of the test. In addition, the characterization methodologies used, i.e., scanning electron microscopy (SEM), field emission scanning electron microscopy (FESEM), transmission electron microscopy (TEM), high-resolution transmission electron microscopy (HRTEM), X-ray Diffraction (XRD), energy dispersive X-Ray (EDX), Fourier transform infrared spectroscopy (FTIR), dynamic light scattering (DLS) and Ultraviolet-visible spectroscopy (UV-vis spectroscopy) to evaluate the P-chem properties were collected.

Output-related collection: various assays can be used to assess the anti-ox efficiency of NPs including 2,2-diphenyl-1-picrylhydrazyl (DPPH), 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid (ABTS), nitric oxide (NO) or total radical-trapping antioxidant potential (TRAP). After collecting the above information, we selected the DPPH assay that measures the anti-ox efficiency of NMs as the output to be predicted, since it is the most popular, rapid, inexpensive and simple test [69]. In addition, the DPPH assay appeared as the most frequent assay (resulting in sufficient data points for modelling purposes) in the literature search and by choosing one assay we reflect the homogeneity of the experimental data measurements.

4.2. Model Development

Data pre-processing: PyCaret executes missing value imputation and categorical one-hot encoding. The missing values in the numerical features are replaced with the mean value of the features in the training dataset, whereas the missing values in categorical features are imputed with the most frequent or the mode value [70]. Rafsunjani [48] compared several imputation techniques and found that mean imputation performed the best for handling missing values. One-hot encoding was performed since regression models cannot handle categorical features. Normalization was used to rescale numerical values to reduce the impact of magnitude in the deviation without altering the differences in the range of values. The dataset was split into training (internal, 70% of the dataset to fit the model and to test goodness of fit) and test set (external, 30% of the dataset for evaluation of predictability).

Model development: We trained several regression models including Random Forest (RF), Light Gradient Boosting Machine (LIGHTGBM), Extra Trees (ET), Decision Tree (DT), K-Nearest Neighbors (KNN), Least Absolute Shrinkage And Selection Operator (LASSO), Elastic Net (EN) Least Angle Regression (LAR), passive aggressive regressor (PAR), Lasso Least Angle Regression (LLAR), Huber Regressor (HUBER), Adaboost Regressor (ADA), Bayesian Ridge (BR), Ridge Regression (RIDGE), Gradient Boosting Machine (GBM) and Orthogonal Matching Pursuit (OMP). All models were evaluated and compared to find the best model with good predictivity. The following seven models are selected for reporting in this study: RF is a supervised algorithm that employs both decision trees and bagging techniques. These decision trees organize attributes by arranging them by their values. RF executes feature selection, each tree calculates the significance of a feature based on its ability to improve the integrity of the leaves [71]. LIGHTGBM is a gradient boosting algorithm that uses decision trees. Different data points play different roles in the computation of the information gain. Higher gradient data points will have a greater influence on the information acquisition [49]. ET algorithm generates a large number of the unpruned decision or regression trees from the training dataset and then combines the predictions of multiple decision trees. ET possess similar properties as RF, randomly samples the features at each split point. Nevertheless, differs from RF, where it splits nodes by selecting cut-points entirely at random and fits each decision tree to the entire training dataset [50]. DT is a decision support technique that employs a binary model to display the hidden association between the input and outputs, also known as a pattern recognition capable of categorizing input variables into categories. It has some limitations such as failing to recognize the

impact of weak attributes [51]. The KNN technique works on the principle that comparable/similar samples exist in the vicinity and have a higher probability. KNN approach finds distances between a new data point and all existing training data points, then selects the nearest data points (k) to the query [52]. LASSO is used for the regularization of data models. It conducts feature selection and can minimize model complexity while preventing overfitting. It executes L1 penalty, providing a type of automatic feature selection that has the effect of shrinkage of coefficients and thus input variables that do not provide much to the prediction set to zero [53]. EN is a standardized or penalized linear regression approach that is commonly used to solve high-dimensional feature space regression problems. EN automatically chooses variables, promoting continuous shrinkage [54].

Model validation: 10-fold cross-validation was performed on the training set (70%) to optimize the hyperparameters and to ensure robustness. The mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2) were employed to assess the performance of the models. In addition, the external data (30% of data) was used to assess the predictability of the models.

Attributes importance: A supervised approach that ranks features that are important for predicting the outcome was performed once the best model was selected [55]. The value of attributes varies from 0 to 1, with 1 representing the most information gain. In a regression model, the concept of “variance reduction” is used to guide the feature selection for internal nodes. Thus, we can evaluate how each feature reduces the impurity of the split (the feature with the highest reduction is chosen for the internal node). We measure the average impurity reduction for each feature across all trees in the RF model representing the feature relevance:

$$I_g(f) = \sum_M w_{p,n} \sum_M \Delta i_f(\tau, M)$$

where I_g is the Gini importance which identifies the frequency of a particular feature (f) chosen for the split and the importance of the feature's value. Gini impurity reduction resulting from any peak split $\Delta i_f(\tau, M)$ is assembled individually for each feature nodes τ in the M number of weighted trees in the forest [71].

5. Conclusions

Researchers from multiple domains are interested in the application of nano-antioxidants for the prevention and treatment of oxidative stress-mediated disorders. We demonstrate the application of an RF algorithm to predict the efficiency of nano-antioxidants. The attribute importance analysis demonstrated the NMs' P-chem characteristics such as type and core-size to have a significant impact on the predictions along with the exposure concentration (dosage). Furthermore, this paper emphasizes the significance of a consistent and systematic reporting structure and comparable measurement techniques indicating that their absence can result in an obstacle for modelling purposes. This is an important finding for policymakers and research funding bodies alike. The ability to predict functionalities represents a great opportunity in the healthcare field. Additionally, it will assist insurers and other stakeholders in the healthcare ecosystem to determine whether it is safe and beneficial (profitable) to invest in the development of nano-antioxidant therapies. Future research should therefore attempt to address this challenge through data capturing templates from empirical researchers that facilitate the construction of a comprehensive dataset.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24032792/s1>.

Author Contributions: Conceptualization, M.M. (Mahsa Mirzaei) and I.F.; methodology, M.M. (Mahsa Mirzaei); formal analysis, M.M. (Mahsa Mirzaei); data curation, M.M. (Mahsa Mirzaei); writing—original draft preparation M.M. (Mahsa Mirzaei); writing—review and editing, F.M. and M.M. (Martin Mullins); visualization, M.M. (Mahsa Mirzaei) and I.F.; supervision, I.F., F.M. and M.M. (Martin Mullins); funding acquisition: I.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 862444.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available at: 10.5281/zenodo.6584826, accessed on 26 May 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mittler, R. ROS are good. *Trends Plant Sci.* **2017**, *22*, 11–19. [[CrossRef](#)] [[PubMed](#)]
2. Sies, H.; Jones, D.P. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 363–383. [[CrossRef](#)] [[PubMed](#)]
3. Di Meo, S.; Venditti, P. Evolution of the knowledge of free radicals and other oxidants. *Oxidative Med. Cell Longev.* **2020**, *2020*, 9829176. [[CrossRef](#)]
4. Alkadi, H. A review on free radicals and antioxidants. *Infect. Disord. -Drug Targets* **2020**, *20*, 16–26.
5. Gulcin, I. Antioxidants and antioxidant methods: An updated overview. *Arch. Toxicol.* **2020**, *94*, 651–715. [[CrossRef](#)] [[PubMed](#)]
6. Mirzaei, A.; Mirzaei, M.; Mirzaei, N. Antioxidant activity of the Pistacia atlantica in carbon tetrachloride intoxicated rats in Iran. *Clin. Biochem.* **2011**, *44*, S329–S330. [[CrossRef](#)]
7. Mirzaei, A.; Mohammadi, J.; Mirzaei, N.; Mirzaei, M. The Antioxidant Capacities and Total Phenolic Contents of Some Medicinal Plants in Iran. *J. Adv. Biomed. Sci.* **2011**, *1*, 160–167.
8. Mirzaei, A.; Toori, M.A.; Mirzaei, N.; Shirazi, R.G. Antioxidant, antimicrobial and antimutogenic potential of 4 Iranian medicinal plants. *Life Sci. J.* **2013**, *10*, 1085–1091.
9. Toori, M.A.; Mirzaei, M.; Mirzaei, N.; Lamrood, P.; Mirzaei, A. Antioxidant and hepatoprotective effects of the internal layer of oak fruit (Jaft). *J. Med. Plants Res.* **2013**, *7*, 24–28.
10. Poprac, P.; Jomova, K.; Simunkova, M.; Kollar, V.; Rhodes, C.J.; Valko, M. Targeting Free Radicals in Oxidative Stress-Related Human Diseases. *Trends Pharm. Sci.* **2017**, *38*, 592–607. [[CrossRef](#)]
11. Morry, J.; Ngamcherdtrakul, W.; Yantasee, W. Oxidative stress in cancer and fibrosis: Opportunity for therapeutic intervention with antioxidant compounds, enzymes, and nanoparticles. *Redox Biol.* **2017**, *11*, 240–253. [[CrossRef](#)] [[PubMed](#)]
12. Ghorbani, M.; Derakhshankhah, H.; Jafari, S.; Salatin, S.; Dehghanian, M.; Falahati, M.; Ansari, A. Nanozyme antioxidants as emerging alternatives for natural antioxidants: Achievements and challenges in perspective. *Nano Today* **2019**, *29*, 100775. [[CrossRef](#)]
13. Zhou, J.; Kroll, A.V.; Holay, M.; Fang, R.H.; Zhang, L. Biomimetic nanotechnology toward personalized vaccines. *Adv. Mater.* **2020**, *32*, 1901255. [[CrossRef](#)] [[PubMed](#)]
14. Doroudian, M.; MacLoughlin, R.; Poynton, F.; Prina-Mello, A.; Donnelly, S.C. Nanotechnology based therapeutics for lung disease. *Thorax* **2019**, *74*, 965–976. [[CrossRef](#)]
15. Grodzinski, P.; Kircher, M.; Goldberg, M.; Gabizon, A. *Integrating Nanotechnology into Cancer Care*; ACS Nano, National Cancer Institute, National Institutes of Health: Rockville, MD, USA, 2019; Volume 13, pp. 7370–7376.
16. Ijaz, I.; Gilani, E.; Nazir, A.; Bukhari, A. Detail review on chemical, physical and green synthesis, classification, characterizations and applications of nanoparticles. *Green Chem. Lett. Rev.* **2020**, *13*, 223–245. [[CrossRef](#)]
17. Liandi, A.R.; Cahyana, A.H.; Yunarti, R.T.; Wendari, T.P. Facile synthesis of magnetic Fe₃O₄@Chitosan nanocomposite as environmentally green catalyst in multicomponent Knoevenagel-Michael domino reaction. *Ceram. Int.* **2022**, *48*, 20266–20274. [[CrossRef](#)]
18. Nelson, B.C.; Johnson, M.E.; Walker, M.L.; Riley, K.R.; Sims, C.M. Antioxidant Cerium Oxide Nanoparticles in Biology and Medicine. *Antioxidants* **2016**, *5*, 15. [[CrossRef](#)]
19. Fadaka, A.; Aluko, O.; Awawu, S.; Theledi, K. Green Synthesis of Gold Nanoparticles using Pimenta dioica Leaves Aqueous Extract and Their Application as Photocatalyst, Antioxidant, and Antibacterial Agents. *J. Multidiscip. Appl. Nat. Sci.* **2021**, *1*, 78–88. [[CrossRef](#)]
20. Kumar, H.; Bhardwaj, K.; Nepovimova, E.; Kuča, K.; Singh Dhanjal, D.; Bhardwaj, S.; Bhatia, S.K.; Verma, R.; Kumar, D. Antioxidant functionalized nanoparticles: A combat against oxidative stress. *Nanomaterials* **2020**, *10*, 1334. [[CrossRef](#)]
21. Khalil, I.; Yehye, W.A.; Etxeberria, A.E.; Alhadi, A.A.; Dezfooli, S.M.; Julkapli, N.B.M.; Basirun, W.J.; Seyfoddin, A. Nanoantioxidants: Recent Trends in Antioxidant Delivery Applications. *Antioxid.* **2019**, *9*, 24. [[CrossRef](#)] [[PubMed](#)]
22. Valgimigli, L.; Baschieri, A.; Amorati, R. Antioxidant activity of nanomaterials. *J. Mater. Chem. B* **2018**, *6*, 2036–2051. [[CrossRef](#)]
23. Saravanakumar, G.; Kim, J.; Kim, W.J. Reactive-Oxygen-Species-Responsive Drug Delivery Systems: Promises and Challenges. *Adv. Sci.* **2017**, *4*, 1600124. [[CrossRef](#)]
24. Jampilek, J.; Kralova, K. Potential of nanonutraceuticals in increasing immunity. *Nanomaterials* **2020**, *10*, 2224. [[CrossRef](#)] [[PubMed](#)]
25. Rehman, A.; John, P.; Bhatti, A. Biogenic Selenium Nanoparticles: Potential Solution to Oxidative Stress Mediated Inflammation in Rheumatoid Arthritis and Associated Complications. *Nanomaterials* **2021**, *11*, 2005. [[CrossRef](#)] [[PubMed](#)]

26. Chauhan, P.S.; Tayal, S.; Jha, N. Nano-Antioxidants, A Next-Generation Therapeutic Strategy. In *Antioxidant-Based Therapies for Disease Prevention and Management*; Apple Academic Press: New York, NY, USA, 2021; pp. 1–20.
27. Hamet, P.; Tremblay, J. Artificial intelligence in medicine. *Metabolism* **2017**, *69*, S36–S40. [[CrossRef](#)] [[PubMed](#)]
28. Ibrahim, I.; Abdulazeez, A. The role of machine learning algorithms for diagnosing diseases. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 10–19. [[CrossRef](#)]
29. Patel, V.; Shah, M. Artificial intelligence and machine learning in drug discovery and development. *Intell. Med.* **2022**, *2*, 134–140. [[CrossRef](#)]
30. OECD. Available online: <https://oecd.ai/en/dashboards/policy-areas/PA11> (accessed on 28 January 2023).
31. Furxhi, I.; Perucca, M.; Blosi, M.; Lopez de Ipiña, J.; Oliveira, J.; Murphy, F.; Costa, A.L. ASINA Project: Towards a Methodological Data-Driven Sustainable and Safe-by-Design Approach for the Development of Nanomaterials. *Front. Bioeng. Biotechnol.* **2022**, *9*, 805096. [[CrossRef](#)] [[PubMed](#)]
32. Dahiwade, D.; Patle, G.; Meshram, E. Designing Disease Prediction Model Using Machine Learning Approach. In Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 27–29 March 2019; pp. 1211–1215.
33. Carracedo-Reboredo, P.; Liñares-Blanco, J.; Rodríguez-Fernández, N.; Cedrón, F.; Novoa, F.J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C. A review on machine learning approaches and trends in drug discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4538–4558. [[CrossRef](#)] [[PubMed](#)]
34. Adam, G.; Rampášek, L.; Safikhani, Z.; Smirnov, P.; Haibe-Kains, B.; Goldenberg, A. Machine learning approaches to drug response prediction: Challenges and recent progress. *npj Precis. Oncol.* **2020**, *4*, 19. [[CrossRef](#)]
35. Pu, L.; Naderi, M.; Liu, T.; Wu, H.-C.; Mukhopadhyay, S.; Brylinski, M. eToxPred: A machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol. Toxicol.* **2019**, *20*, 2. [[CrossRef](#)] [[PubMed](#)]
36. Furxhi, I.; Murphy, F. Predicting in vitro neurotoxicity induced by nanoparticles using machine learning. *Int. J. Mol. Sci.* **2020**, *21*, 5280. [[CrossRef](#)] [[PubMed](#)]
37. Chowdhury, A.S.; Call, D.R.; Broschat, S.L. PARGT: A software tool for predicting antimicrobial resistance in bacteria. *Sci. Rep.* **2020**, *10*, 11033. [[CrossRef](#)]
38. Li, L.-G.; Yin, X.; Zhang, T. Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. *Microbiome* **2018**, *6*, 93. [[CrossRef](#)]
39. Mirzaei, M.; Furxhi, I.; Murphy, F.; Mullins, M. A machine learning tool to predict the antibacterial capacity of nanoparticles. *Nanomaterials* **2021**, *11*, 1774. [[CrossRef](#)] [[PubMed](#)]
40. Tsalidis, G.A.; Soeteman-Hernández, L.G.; Noorlander, C.W.; Saedy, S.; van Ommen, J.R.; Vijver, M.G.; Korevaar, G. Safe-and-Sustainable-by-Design Framework Based on a Prospective Life Cycle Assessment: Lessons Learned from a Nano-Titanium Dioxide Case Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4241. [[CrossRef](#)]
41. Gottardo, S.; Mech, A.; Drbohlovová, J.; Małyska, A.; Bøwadt, S.; Riego Sintes, J.; Rauscher, H. Towards safe and sustainable innovation in nanotechnology: State-of-play for smart nanomaterials. *NanoImpact* **2021**, *21*, 100297. [[CrossRef](#)]
42. Stucki, A.O.; Barton-Maclaren, T.S.; Bhuller, Y.; Henriquez, J.E.; Henry, T.R.; Hirn, C.; Miller-Holt, J.; Nagy, E.G.; Perron, M.M.; Ratzlaff, D.E.; et al. Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health. *Front. Toxicol.* **2022**, *4*, 964553. [[CrossRef](#)]
43. Ramanarayanan, T.; Szarka, A.; Flack, S.; Hinderliter, P.; Corley, R.; Charlton, A.; Pyles, S.; Wolf, D. Application of a new approach method (NAM) for inhalation risk assessment. *Regul. Toxicol. Pharmacol.* **2022**, *133*, 105216. [[CrossRef](#)]
44. Karydas, C.; Iatrou, M.; Kouretas, D.; Patouna, A.; Iatrou, G.; Lazos, N.; Gewehr, S.; Tseni, X.; Tekos, F.; Zartaloudis, Z.; et al. Prediction of antioxidant activity of cherry fruits from UAS multispectral imagery using machine learning. *Antioxidants* **2020**, *9*, 156. [[CrossRef](#)]
45. Idowu, S.O.; Fatokun, A.A. Artificial Intelligence (AI) to the Rescue: Deploying Machine Learning to Bridge the Biorelevance Gap in Antioxidant Assays. *SLAS Technol. Transl. Life Sci. Innov.* **2020**, *26*, 16–25. [[CrossRef](#)]
46. Olsen, T.H.; Yesiltas, B.; Marin, F.I.; Pertseva, M.; García-Moreno, P.J.; Gregersen, S.; Overgaard, M.T.; Jacobsen, C.; Lund, O.; Hansen, E.B.; et al. AnOxPePred: Using deep learning for the prediction of antioxidative properties of peptides. *Sci. Rep.* **2020**, *10*, 21471. [[CrossRef](#)] [[PubMed](#)]
47. Furxhi, I. Health and environmental safety of nanomaterials: O Data, Where Art Thou? *NanoImpact* **2022**, *25*, 100378. [[CrossRef](#)] [[PubMed](#)]
48. Rafsunjani, S.; Safa, R.S.; Al Imran, A.; Rahim, M.S.; Nandi, D. An empirical comparison of missing value imputation techniques on APS failure prediction. *Int. J. Inf. Technol. Comput. Sci* **2019**, *2*, 21–29. [[CrossRef](#)]
49. Thompson, M.; Burger, K.; Kaliyaperumal, R.; Roos, M.; da Silva Santos, L.O.B. Making FAIR easy with FAIR tools: From creolization to convergence. *Data Intell.* **2020**, *2*, 87–95. [[CrossRef](#)]
50. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]
51. Tavernaro, I.; Dekkers, S.; Soeteman-Hernández, L.G.; Herbeck-Engel, P.; Noorlander, C.; Kraegeloh, A. Safe-by-Design part II: A strategy for balancing safety and functionality in the different stages of the innovation process. *NanoImpact* **2021**, *24*, 100354. [[CrossRef](#)]

52. Bedlovicova, Z.; Strapac, I.; Balaz, M.; Salayova, A. A Brief Overview on Antioxidant Activity Determination of Silver Nanoparticles. *Molecules* **2020**, *25*, 3191. [[CrossRef](#)]
53. Gramatica, P. Principles of QSAR modeling: Comments and suggestions from personal experience. *Int. J. Quant. Struct.-Prop. Relatsh.* **2020**, *5*, 61–97. [[CrossRef](#)]
54. Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C.A. Practices and trends of machine learning application in nanotoxicology. *Nanomaterials* **2020**, *10*, 116. [[CrossRef](#)]
55. Furxhi, I.; Murphy, F.; Mullins, M.; Poland, C.A. Machine learning prediction of nanoparticle in vitro toxicity: A comparative study of classifiers and ensemble-classifiers using the Copeland Index. *Toxicol. Lett.* **2019**, *312*, 157–166. [[CrossRef](#)] [[PubMed](#)]
56. Mirzaei, M.; Furxhi, I.; Murphy, F.; Mullins, M. A Supervised Machine-Learning Prediction of Textile's Antimicrobial Capacity Coated with Nanomaterials. *Coatings* **2021**, *11*, 1532. [[CrossRef](#)]
57. Sharpe, E.; Andreescu, D.; Andreescu, S. Artificial Nanoparticle Antioxidants. In *Oxidative Stress: Diagnostics, Prevention, and Therapy*; American Chemical Society: Washington, DC, USA, 2011; Volume 1083, pp. 235–253.
58. Kim, E.S.; Baek, Y.; Yoo, H.J.; Lee, J.S.; Lee, H.G. Chitosan-Tripolyphosphate Nanoparticles Prepared by Ionic Gelation Improve the Antioxidant Activities of Astaxanthin in the In Vitro and In Vivo Model. *Antioxidants* **2022**, *11*, 479. [[CrossRef](#)] [[PubMed](#)]
59. Fafal, T.; Taştan, P.; Tüzün, B.; Ozyazici, M.; Kivcak, B. Synthesis, characterization and studies on antioxidant activity of silver nanoparticles using *Asphodelus aestivus* Brot. *aerial part extract*. *South Afr. J. Bot.* **2017**, *112*, 346–353. [[CrossRef](#)]
60. Sandhir, R.; Yadav, A.; Sunkaria, A.; Singhal, N. Nano-antioxidants: An emerging strategy for intervention against neurodegenerative conditions. *Neurochem. Int.* **2015**, *89*, 209–226. [[CrossRef](#)]
61. Cîrcu, M.; Nan, A.; Borodi, G.; Liebscher, J.; Turcu, R. Refinement of magnetite nanoparticles by coating with organic stabilizers. *Nanomaterials* **2016**, *6*, 228. [[CrossRef](#)]
62. Samuel, E.L.; Duong, M.T.; Bitner, B.R.; Marcano, D.C.; Tour, J.M.; Kent, T.A. Hydrophilic carbon clusters as therapeutic, high-capacity antioxidants. *Trends Biotechnol.* **2014**, *32*, 501–505. [[CrossRef](#)]
63. Verma, A.K. Anti-oxidant activities of biopolymeric nanoparticles: Boon or bane. *J. Pharm. Res.* **2014**, *8*, 871–876.
64. Ruktanonchai, U.; Bejrappa, P.; Sakulkhu, U.; Opanasopit, P.; Bunyapraphatsara, N.; Junyaprasert, V.; Puttipipatkachorn, S. Physicochemical Characteristics, Cytotoxicity, and Antioxidant Activity of Three Lipid Nanoparticulate Formulations of Alpha-lipoic Acid. *AAPS PharmSciTech* **2009**, *10*, 227. [[CrossRef](#)] [[PubMed](#)]
65. Ball, N.; Bars, R.; Botham, P.A.; Cuciureanu, A.; Cronin, M.T.D.; Doe, J.E.; Dudzina, T.; Gant, T.W.; Leist, M.; van Ravenzwaay, B. A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Arch. Toxicol.* **2022**, *96*, 743–766. [[CrossRef](#)]
66. Jo, J.-M. Effectiveness of normalization pre-processing of big data to the machine learning performance. *J. Korea Inst. Electron. Commun. Sci.* **2019**, *14*, 547–552. [[CrossRef](#)]
67. Seger, C. *An Investigation of Categorical Variable Encoding Techniques in Machine Learning: Binary Versus One-Hot and Feature Hashing*; EECS: Stockholm, Sweden, 2018; p. 34. Available online: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-237426> (accessed on 28 January 2023).
68. Brownlee, J. *Data Preparation for Machine Learning Data Cleaning, Feature Selection, and Data Transforms in Python*; Machine Learning Mastery: San Juan, Argentina, 2020.
69. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
70. Mishra, S.; Mallick, P.K.; Tripathy, H.K.; Bhoi, A.K.; González-Briones, A. Performance Evaluation of a Proposed Machine Learning Model for Chronic Disease Datasets Using an Integrated Attribute Evaluator and an Improved Decision Tree Classifier. *Appl. Sci.* **2020**, *10*, 8137. [[CrossRef](#)]
71. Disha, R.A.; Waheed, S. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. *Cybersecurity* **2022**, *5*, 1. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.