Article

A Machine Learning Tool to Predict the Antibacterial Capacity of Nanoparticles

Mahsa Mirzaei [1], **Irini Furxhi** [1,2,*], **Finbarr Murphy** [1,2] **and Martin Mullins** [1]

1 Department of Accounting and Finance, Kemmy Business School, University of Limerick, V94PH93 Limerick, Ireland; mahsa.mirzaei@ul.ie (M.M.); finbarr.murphy@ul.ie (F.M.); martin.mullins@ul.ie (M.M.)
2 Transgero Limited, Cullinagh, Newcastle West, V42V384 Limerick, Ireland
* Correspondence: irini.furxhi@ul.ie; Tel.: +353-85-106-9771

Abstract: The emergence and rapid spread of multidrug-resistant bacteria strains are a public health concern. This emergence is caused by the overuse and misuse of antibiotics leading to the evolution of antibiotic-resistant strains. Nanoparticles (NPs) are objects with all three external dimensions in the nanoscale that varies from 1 to 100 nm. Research on NPs with enhanced antimicrobial activity as alternatives to antibiotics has grown due to the increased incidence of nosocomial and community acquired infections caused by pathogens. Machine learning (ML) tools have been used in the field of nanoinformatics with promising results. As a consequence of evident achievements on a wide range of predictive tasks, ML techniques are attracting significant interest across a variety of stakeholders. In this article, we present an ML tool that successfully predicts the antibacterial capacity of NPs while the model's validation demonstrates encouraging results ($R^2 = 0.78$). The data were compiled after a literature review of 60 articles and consist of key physico-chemical (p-chem) properties and experimental conditions (exposure variables and bacterial clustering) from in vitro studies. Following data homogenization and pre-processing, we trained various regression algorithms and we validated them using diverse performance metrics. Finally, an important attribute evaluation, which ranks the attributes that are most important in predicting the outcome, was performed. The attribute importance revealed that NP core size, the exposure dose, and the species of bacterium are key variables in predicting the antibacterial effect of NPs. This tool assists various stakeholders and scientists in predicting the antibacterial effects of NPs based on their p-chem properties and diverse exposure settings. This concept also aids the safe-by-design paradigm by incorporating functionality tools.

Keywords: nanoparticles; antibacterial effect; antimicrobial capacity; biofilm; machine learning

1. Introduction

Antibiotic resistance is increasing to alarmingly high levels, the resistance mechanisms threatening our ability to treat common infectious diseases which leads to a global health risk [1]. Antibacterial agents are compounds that can be classified as either bactericidal, completely inhibiting and eradicating bacteria, or bacteriostatic, which inhibits bacterial growth [2]. However, several factors may influence this classification, including growth conditions, bacterial density or test duration [3]. More importantly, the effectiveness of most compounds depends on the type of bacteria (Gram-positive and Gram-negative bacteria) exposed to [2,4]. The majority of existing antibacterial agents are chemically modified natural compounds, e.g., β-lactamines (i.e., penicillin), cephalosporins or carbapenems; or purely natural products (i.e., aminoglycosides), and purely synthetic antibiotics, such as sulfonamides [2,5]. As a result of the recurrence of infections, the microorganisms develop resistance due to inherent genetic changes [6,7]. With the excessive use or misuse of antibacterial agents, the emergence of resistance to antibacterial drugs has become one of the most significant public health challenges [8,9].

While bacteria are normally found in nature in the form of individual cells, they may also develop multicellular structures called biofilms; densely packed groups of bacteria that contain one or more species of bacteria [10]. The biofilm provides mechanical stability and adhesion to a wide variety of surfaces both biotic and abiotic, including human tissues, surgical devices, implants, and industrial equipment [11,12]. Biofilms are a major issue in almost all industries, contaminating equipment and the surrounding environment, resulting in reduced quality of products and economic losses [13,14]. Bacterial biofilms contribute to microbial resistance and therefore play a significant role in therapeutic failure, resulting in chronic bacterial infections [15,16]. Considering the role of biofilm in antibiotic resistance, the elimination of bacteria needs multiple drugs with potential side effects in humans and environments, as a result of the need for high doses of common disinfectants and antibiotics there is the increase in toxicity, cost and duration of therapy; therefore, new treatments are necessary to eliminate bacteria.

Nanoparticles (NPs) are widely used due to their unique and size-dependent physical and chemical (p-chem) properties. They exhibit enhanced antimicrobial capacities [17], making them a suitable alternative to antibiotics. NPs have been studied for their capacity to inhibit microbial infections [18] and prevent bacterial colonization on various surface devices such as catheters [19] and prostheses [20] by eradicating biofilms [21,22]. Research into NPs is of great interest as they can be applied in various fields such as medicine, food industry, and manufacturing, while retaining their original unique functions [23–26]. Over the past few decades, the search for new antimicrobial substances has been central to many research areas, both in public and private research centers, for the reduction of nosocomial and foodborne infections [27–29].

Metal and metal oxide NPs (MO-NPs) are promising agents against a broad spectrum of microorganisms including drug-resistant strains [17,30,31]. The exact mechanisms of NP toxicity against different bacteria depends on surface modification, intrinsic properties, composition, and the bacterial species. The main mechanisms of the antibacterial effects of NPs are: (1) mechanical damage to the cell wall through electrostatic interaction; (2) oxidative stress by means of the generation of reactive oxygen species (ROS); and (3) disruption of cell and protein structures as a result of metal cation release [32]. Among MO-NPs, the most promising and widely studied ones are Fe_3O_4 and ZnO [33]. Fe_3O_4 NPs release Fe^{2+} ions, which cause the generation of ROS after a reaction with hydrogen peroxide (H_2O_2) and induces oxidative stress in the cell, as a result of which the bacteria cell dies [34,35]. ZnO NPs produce H_2O_2 and hydroxyl radicals (OH^-), but not superoxide (O_2^-) and have weak mutagenic capability that induces frameshift mutation in bacterium [36–38]. Metal NPs such as AgNPs have an oligodynamic effect (the biocidal effects of metals) due to their large surface areas and have the ability to accumulate at the cell wall and bind with bacterial biomolecules [39] and penetrate the cells [40], generating ROS and free radicals, and act as modulators in the signal transduction pathways of microorganisms [41–43].

Antibacterial activities of NPs depend on two main factors: (i) p-chem properties, such as composition, surface modification and intrinsic properties, and (ii) the type of bacteria species [2,44–46]. For a better understanding of their properties and effects, a computational tool can assist in reducing the design space by predicting the characteristics of desired NPs before synthesis, which helps to decrease the experimental trial and error work. For example, tools have been employed to predict the three-dimensional structure of metallic nanoparticles [47,48] in order to determine the functional composition of the protein corona of NPs [49].

Artificial intelligence (AI) is a branch of computer science that has attracted much interest in many fields due to its problem-solving, decision-making and trend recognition capabilities. Machine learning (ML), a subset of AI, focuses on the ability of algorithms to learn from data while organizing the information they process. ML is a method of data analysis that automates model building while not requiring deterministic insights, bypassing in-depth comprehension and bridging input data directly to the outcome [50].

Moreover, these tools are fast and inexpensive, and they rely on information inputs rather than physical test materials. In addition, they can predict the impact of materials not yet synthesized, thereby contributing to safe-by-design approaches [51]. ML has been effectively employed for the prediction of toxicity profiles of NPs [52–56] and for the development of new antibiotics [57,58]. Furthermore, models for the prediction of the antimicrobial resistance for specific bacteria have been demonstrated [59–61]. For example, Khaledi, Weimann et al. [60] investigated the antimicrobial susceptibility of *Pseudomonas aeruginosa* predicted by genomic and transcriptomic markers. Yang, Niehaus et al. [62] employed algorithms for the identification of *Mycobacterium tuberculosis* resistance against several tuberculosis drugs. Her and Wu [59] demonstrated the prediction of antimicrobial susceptibility of *Escherichia coli* by using the pangenome-based ML approach.

To the best of our knowledge, there is no study that uses ML to predict the antibacterial effects of NPs. We propose an ML tool predicting the antibacterial activity of NPs against a vast range of Gram-positive and Gram-negative bacteria. This tool predicts the antibacterial effect by exploring p-chem properties, experimental exposure conditions and bacteria characteristics as inputs. We gathered in vitro experimental data from literature studies and structured them into a comprehensive dataset (Supplementary Material S1). The present approach allows the screening of NPs, predicting their capacity to eradicate bacteria, saving time and reducing costs by reducing the amount of trial and error in the lab. Such an approach would help scientists to prevent bacterial growth that can be harmful to human health, environments and industrial components that are subjected to biofilm formation and bacterial growth [63].

2. Materials and Methods

Approach

Figure 1 demonstrates the roadmap followed for the model implementation. Initially, studies related to the antibacterial effect of NPs are collected and data extraction is performed relating to p-chem properties of NPs, exposure conditions and information about the exposed bacteria. The original dataset was evaluated for completeness. Data pre-processing followed, including standardization [64], one hot encoding and one data split [65]. To find a model with good predictivity, we trained and validated several regression models. Finally, an analysis of attribute importance [66,67] was conducted to reveal the attributes that most influence the prediction of the results.
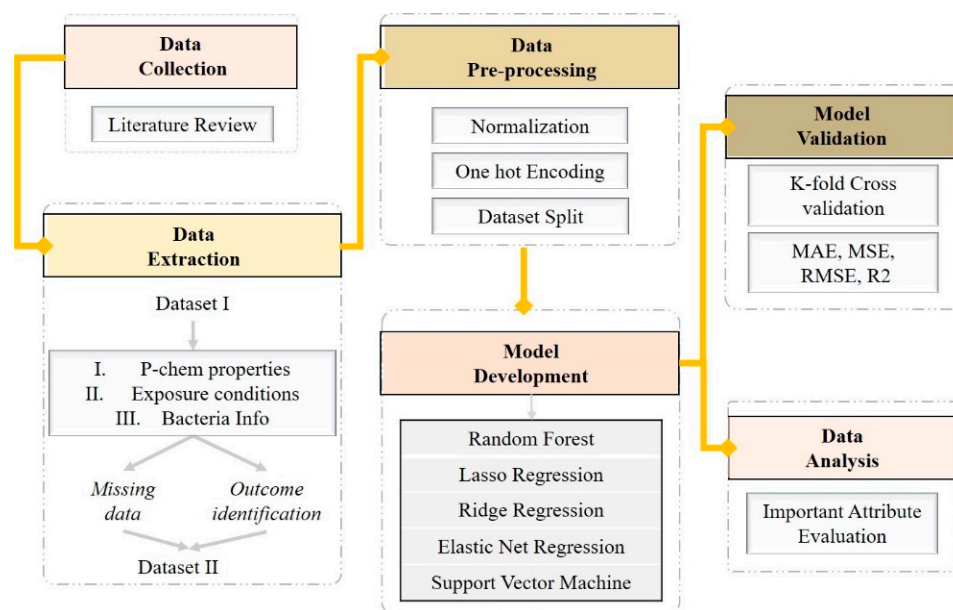

Figure 1. Model development workflow.

2.1. Data Collection

A literature search was carried out in January 2021 for studies that investigated the impact of NPs on the elimination or inhibition of bacteria and consequently the degradation of biofilm. The articles collected were published between 2010 and 2020 including different keywords, such as "antibacterial", "antimicrobial", "bactericidal", "biofilm" and "antifungal" effects. We determined to evaluate silver (Ag) [68], iron oxide (IO) and zinc oxide (ZnO) NPs as they are widely used as bactericidal agents [17,69]. These NPs are promising candidates, since they demonstrate greater sustainability, reduced toxicity, greater stability and selectivity compared to organic NPs [70,71]. Their low toxicity against human cells [72,73], low cost [74], inhibition effect against a broad range of bacteria and inhibition of biofilm formation [75] makes them fitting for application as antibacterial agents in biomedical industries [76], food additives [77], fabric [75], and skincare products [78]. Studies using both chemical and green synthesis of NPs have been included. The green synthesis is a growing domain of bio-nanotechnology due to its low cost and non-toxic nature [69,79,80].

- Inclusion criteria for the studies include English language, original studies focusing on the antibacterial properties of NPs, published in the last decade, and in vitro studies.
- Exclusion criteria include reviews, case reports, studies with binary results, studies that demonstrated results only in figures.

A total of 85 papers were selected and 60 articles were deemed relevant to this study. We concentrated on in vitro studies due to the significant benefits they offer in agreement with the three R's movement (replace, reduce and refine the animal experiments), reduced costs and allowing direct evaluation without the influence of pharmacokinetic variables [81].

2.2. Data Extraction

2.2.1. Input Extraction

Each paper was reviewed with a focus on (i) the type of NPs (IONPs, AgNPs, ZnONPs) [2,82]; (ii) the nano-specific descriptors (core size, shape, zeta potential, surface area, etc.) [83]; and (iii) the study design experimental parameters (exposure conditions and bacteria characteristics) [84]. The above variables were acquired as input attributes for the prediction of the antibacterial efficiency of the investigated NPs.

2.2.2. Outcome Extraction

For the evaluation of the antibacterial efficacy, studies reported different assays and techniques. Several outcomes based on antibacterial measurements were documented, such as bacteria viability, zone of inhibition (ZOI), minimum inhibitory concentration (MIC), minimum bactericidal concentration (MBC), colony-forming unit (CFU), optical density (OD), and inhibition in biofilm formation. Different metrics and expressions were shown in output values which stressed the fact that a standardized method and reporting of scientific data is required.

2.3. Data Pre-Processing

2.3.1. Missing Values

In the initial raw dataset (Dataset I), missing data occurred in all the extracted attributes. Following the selection of the outcome, we created the final dataset (Dataset II). Our final data had few missing values among the inputs. As regression models cannot perform well with null data, we deleted the rows with missing values [85].

2.3.2. One Hot Encoding

In regression models, categorical variables (nominal attributes) must be converted into integers (numerical dummy variables). There are several conversion methods [65]; we created dummy variables for each of the categorical columns and integrated the new

columns into the main data frame. The value 0 or 1 was used to denote the absence or presence of the original attributes.

2.3.3. Normalization

Data normalization was conducted on the numeric inputs to enhance model performance. Normalization is achieved by different techniques such as Z-score, min-max, mean and median absolute deviation scaling [86]. Normalization was done by applying a z-score that standardizes a feature to have zero mean unit variance [87].

$$x'i, n = (x(i, n) - \mu i)/\sigma \iota \tag{1}$$

where x represents the features in the dataset, μ the mean and σ the standard deviation of the i, n features.

2.3.4. Data Split

For a supervised computational algorithm to predict outputs of an unknown target function, a training set is provided initially. We randomly split the data into two sets, one to train the model (training set) containing 70% of the dataset and the rest ones (30%) for testing the performance (test set) [88].

2.4. Regression Models

The regression technique constructs a model with the ability to predict new numeric values from the input variables. Regression modeling is the task of approximating a mapping function from inputs to a continuous output [89,90]. The ML algorithm maps functions from NP's p-chem properties and experimental conditions to the inhibition of bacteria and enables the prediction of the antibacterial capacity of NPs. We used several supervised regression algorithms as potential candidates for developing our model to explore which model can provide the most accurate prediction. The Least Absolute Shrinkage and Selection Operator (LASSO) Regression, Ridge Regression (RR), Elastic Net Regression (ENR), Random Forest (RF) and Support Vector Machine (SVM) are examined in this study. Models were built in Python version 3.7.6, Scikit-learn version 0.24.1.

LASSO regression is a popular variable selection and shrinkage estimation method [91] which finds the variables and corresponding regression coefficients that lead to a model with higher accuracy. This is accomplished by imposing a restriction on model parameters, which then 'shrinks' the regression coefficients close to zero. Variables with a regression coefficient of zero after shrinkage are excluded from the model [92].

RR is a simple approach [93] that addresses the collinearity challenge arising in multiple linear regression [94–96]. When the covariates are super-collinear, two or multiple covariates are strongly related. When there is multicollinearity, least squares estimates are unbiased and their differences are larger, so they may be far from the real value. By adding a scale of bias to the estimates, RR decreases the standard errors. The RR model gives different importance weights to the features but does not drop unimportant features in comparison with LASSO [97].

ENR applies the penalties from both the LASSO and RR methods to regularize regression models [98]. ENR often outperforms LASSO, which is particularly useful when the number of predictors is much bigger than the number of observations [98]. This method aims to improve predictions by performing variable selection by forcing the coefficients of "non-significant" variables towards zero (shrinkage) [92].

RF comprises various decision trees that are trained independently on a random subset of data. RF can work with thousands of variables without deletion or reduction in accuracy, while preventing overfitting [99,100]. As a classifier, RF performs an implicit feature selection, using a small subset of "strong variables" for the classification, which is the reason for its superior performance on high dimensional data [101].

SVM learns by assigning labels to objects and is widely used in biological fields. It is used for both classification and regression challenges [102,103]. In SVM, each data item

is plot as a point in an n-dimensional space with the value of each feature related to the value of a specific coordinate. It performs classification by finding the hyper-plane that differentiates the two classes.

2.5. Model Validation

The primary objective of ML is to generate an effective computational model with a high predictive capacity. Cross-validation is used to guarantee a stable assessment of the model performance and to avoid overfitting [104,105]. In cross-validation, the model is trained using parts of the training set by leaving one subset for later testing.

To produce an optimal model, a balance to avoid both underfitting and overfitting by adjusting hyperparameters is critical. In LASSO, RIDGE, and EN models, several statistical techniques were used to evaluate the model (data not shown) by using different sets of alpha values. To tune LASSO, RR and ENR models, we performed a "grid search" approach. The approach reduces the model's complexity by keeping the most important features. The higher the alpha value, the more the regularization parameter influences the final model, hence decreasing the error due to variance (overfit). Alpha in regression models takes various values, however, when $\alpha = 0$, the model gets same coefficients as in simple linear regression (no regularization).

The models were evaluated by mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and Coefficient of Determination or R-squared (R^2).

MAE is a popular metric, calculated as follows:

$$\text{MAE} = 1/n \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \tag{2}$$

where y_i is the i'th expected value in the dataset, $\hat{y}_i$ is the i'th predicted value.

MSE is a standard and common error metric for regression model problems. The MSE is analyzed as the mean or average of the squared differences between predicted and expected target values in a dataset. It can be calculated by

$$\text{MSE} = 1/n \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{3}$$

where y_i is the i'th expected value in the dataset and $\hat{y}_i$ is the i'th predicted value.

The RMSE of expected and predicted values can be calculated through the $\sqrt{\text{MSE}}$. Although a good RMSE value is relative to a dataset, the smaller the value, the better the predictive model.

R^2 is a statistical measure of fit that suggests how much variation of the output is supported by the inputs. R^2 explains to what degree the variance of one variable describes the variance of the second variable. For instance, if R^2 is 0.70, then approximately 70 percent of the observed variation can be explained by the model's inputs, and the greater the R^2 value, the better is the model.

2.6. Important Attribute Analysis

Attribute importance is a supervised event that distinguishes and ranks the attributes that are most important in predicting the outcome in a relative manner [106]. Attribute importance was derived through random forest optimization (built-in function). The analysis is based on the Gini importance, an all-nodes accumulating quantity that indicates how often a particular attribute was selected for a split, and how large its overall discriminative effect was in the regression [100]. Information values range from 0 to 1, with 1 representing maximum information gain.

3. Results

3.1. Data Pre-Processing

The primary dataset comprised 1176 rows and 18 columns (11 inputs, 7 outputs) extracted from 60 studies investigating the antibacterial properties of the NPs.

Input selection and transformation: The input data consisted of specific surface area (m^2/g), hydrodynamic size (nm), zeta potential (measured in water and medium) (mV), core size (nm), exposure dose (μg/mL), and duration (h) reported in numeric values. Variables with nominal values included shape, type of NPs, coating, bacterium and aggregation as summarized in Table 1. As can be seen from Figure 2 (left), specific surface area, hydrodynamic size and zeta potential had approximately more than 90% missing values and were therefore excluded. The aggregation potential had 39% missing data and therefore was discharged. The different types of NPs, coating, duration, and bacteria had no missing values. The dose, shape, and core size had 17.3%, 13.5%, and 9.6% missing values, respectively.

Table 1. The primary and final Input variables in Dataset I and Dataset II.

Category	Variables	Type	DATASET I	Data Transformation	DATASET II
			Min Max or Labels		
P-chem properties	Sp_Surf_Area	Numeric	1.2–96 (m^2/g), NA	Eliminated due to high NA	–
	Hydro_size		11.5–993 nm, NA		–
	Zeta_Medium		−40–90 (mV), NA		–
	Zeta_Water		−40–80 (mV), NA		–
	Core_size		2–1000 (nm), NA	Selected	4–546 (nm), NA
	Aggregation	Nominal	Yes, None, NA	Eliminated due to high NA	–
	Shape		Spherical, Hexagonal, Rod, Spindle, Disc, Cubic, NA	Selected	Spherical, Hexagonal, Rod, Cubic, NA
	NPs type		AgNPs, Fe_3O_4, ZnO		AgNPs, Fe_3O_4, ZnO
	Coating		Iron Oxide, dextran, pullulan, Taraxacum officinale, Aspergillus, Emericella nidulans, Tannic acid, quercetin, TXT_100, SDBD, SDS, Tween 81, PEG, PVP, Crataeva nurvala, PMC, PG, Moringa oleifere, Oleic acid, Zinc oxide, Gold, Chitosan, APTES, Flaxseed oil, silver, CES, alginate, PVA, Carbon, Alow vera, Titanium, SiO_2, starch, Magnesium	Simplified: Transformed into Binary	Coated, Uncoated
Exposure	Dose	Numeric	0.01–10.000 (μg/mL), NA	Selected	0.01–10.000 (μg/mL), Na
	Duration		17–1440 (h)		17–1440 (h)

Table 1. *Cont.*

In vitro Info.	Bacteria	Nominal	*Acetomicrobium faecale, Acidaminococcus fermentans, Actinomyces denticolens, Aspergilus (niger, terreus strain), Bacillus brevis, Bacillus cereus, Bacillus subtilis, Bacteroides (eggerthii, stercoris, thetaiotaomicron, uniformis, vulgatus, xylanolyticus), Bifidobacterium (adolescentis, bifidum, longum, suis, thermophilum), Campylobacter jejuni, Candida (albicans, parapsilosis, tropicalis), Citrobacter freundii, Clostridium (butyicum, cellulovorans, coccoides, histolyticum, leptum, perfringens, thermocellum), Corynebacterium glutamicum, Enterobacter (aerogenes, cloacae), Enterococcus (cecorum, durans, faecalis, faecium, hirae), Escherichia coli, Eubacterium eligens, Fusarium solani, Ganoderma, Klebsiella (aerogenes, oxytoca, pneumoniae), Lactobacillus (acidophilus, amylovorus, casei, fermentum, johnsonii, plantarum, reuteri, salivarius), Leuconostoc (citreum, fallax, lactis, mesenteroides), Listeria monocytogenes, Microbacterium hominis, Neisseria canis, Olsenella (profusa, uli), Proteus (mirabilis, vulgaris), Pseudomonas aeruginosa, Putida vulgaris, Ralstonia solanacearum, Salmonella (Enteritidis, paratyphi, typhi, typhimurium), Serratia marcescens, Shigella (dysenteriae, sonnei), Staphylococcus (aureus, epidermidis), Streptococcus (aureus, epidermidis, bovis, gallolyticus, hyointestinalis, porcinus, pyogenes, salivarius), Veillonella ratti, Vibrio cholerae, Weissella (confusa, hellenica), Xanthomonas oryzae*	Simplified: Data transformed into general Species categories	*A. niger, A. terreus, Aspergillus, B. brevis, B. cereus, B. licheniformis, B. subtilis, C. albicans, C. tropicalis, E. aerogenes, E. coli, E. faecalis, Enterococcus, F. solani, Fusarium, Ganoderma, K. pneumoniae, L. monocytogenes, P. aeruginosa, P. mirabilis, P. multocida, P. putida, P. vulgaris, Penicillium, S. aureus, S. dysenteriae, S. epidermidis, S. marcescens, S. paratyphi, S. typhi, Salmonella, Scedosporium, Shigella, V. cholerae, X. oryzae, Xanthomonas*

The coating and bacteria variables were very dispersed (as illustrated in Figure 3). To prevent model overfitting, we transformed coating into coated and uncoated (binary format). The studies were conducted on several strains of bacteria for which we collected the class, family, and species information. To avoid overloading the model we only kept species as a subcategory of class and family, grasping general (Gram-positive or negative bacteria) and specific information. In the final dataset (Dataset II), we included shape with 11%, dose with 9.8% and core-size with 5% missing values. The other variables had no missing values.
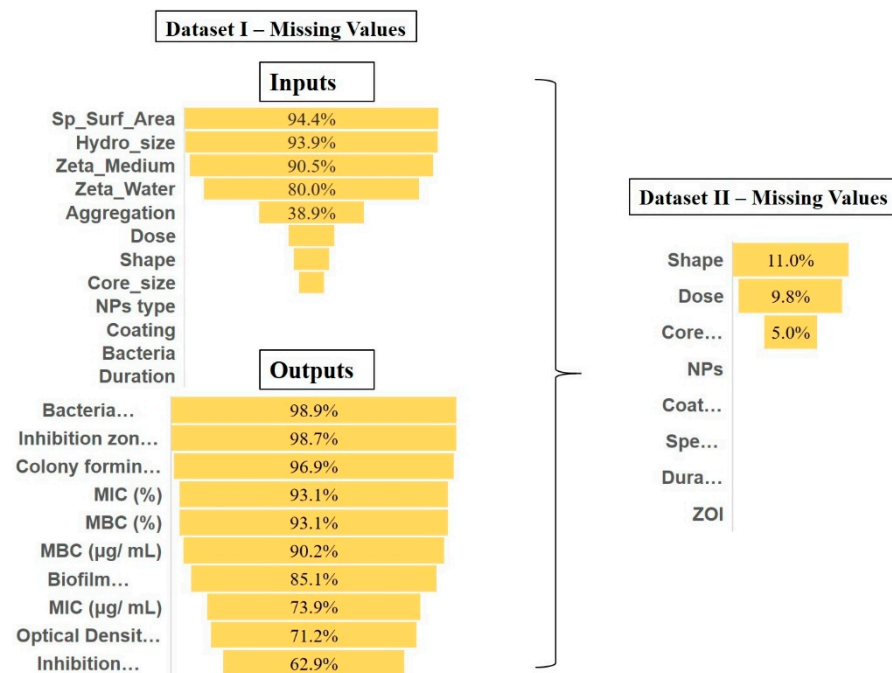
Figure 2. Dataset I. Missing values (percentages) of input and output parameters (left). Dataset II missing values of inputs and one outcome, ZOI (right).
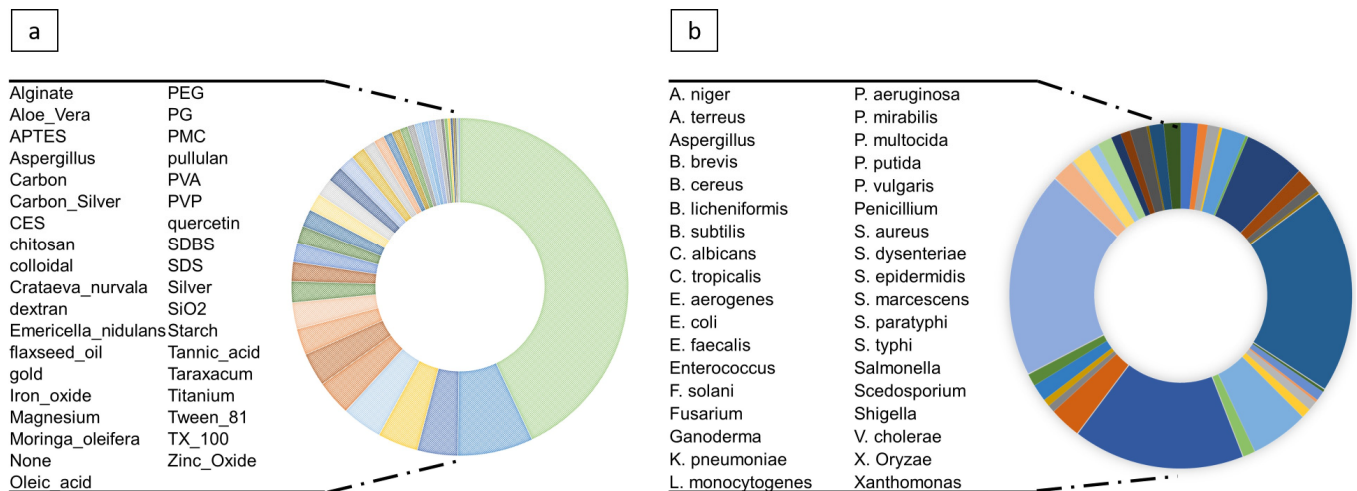

Figure 3. Coating information of NP, the coating variables are all 57% and the un-coated is 43% (**a**). Species of several investigated foodborne and environmental bacterium (**b**).

Outcome Selection: We gathered all the evaluations used to determine the antibacterial efficacy of the NPs. Each of the outcomes has different unit metrics. For example, the bacteria viability/growth is expressed in the number of bacteria cells and cell viability is reported as a percentage of live bacteria. ZOI is given in mm, representing the diameter of the area of media where bacteria are unable to grow [107]. MIC is extracted in μg/mL as the minimum concentration of NPs that inhibit the growth of bacterium [107]. MBC is expressed in μg/mL as the lowest concentration of antibacterial agent required to kill a bacterium [108]. OD measurements represent growth analysis by measuring the optical density at different settings of 580, 600 and 572 nm. The biofilm formation is reported in a percentage and change in biofilm growth is reported in colony-formed unit (CFU) [109].

Due to the high occurrence of missing data and diversity in the outcomes (as shown in Figure 2), we chose the outcome with the least missing value, the inhibition zone

measurement (ZOI) in mm, with 60% missing values. Hence, the rest of the outcomes were dismissed in Dataset II. ZOI testing, also known as the disk diffusion method (DDM), is a fast and inexpensive assay compared to the other laboratory tests [110,111]. The diameter of the ZOI illustrates the antimicrobial activity present in the sample or product—a larger zone of inhibition means that the antimicrobial is more potent. In summary, the final dataset of 436 rows consists of seven inputs (shape, dose, size, coating, type of NPs, bacteria species, and duration as inputs) and one output (ZOI) derived from 60 studies.

3.2. Validation of Models and Attribute Analysis

Following data homogenization and pre-processing, we trained various regression models. Model performance results are presented in Figure 4a. The results suggest that the RF model exhibits the lowest error and the highest R^2 score compared to the other algorithms employed (LASSO, RR, ENR, SVM), with R^2, RMSE, MAE and MSE of 0.78, 4.30, 2.78 and 18.56 values, respectively. The outcome of attribute importance analysis is shown in Figure 4b. Core-size is the most important attribute that determines the efficacy of the antibacterial effect of NPs. The dose, species, and type of NPs are identified as comparatively important, followed by coating, shape, and duration. Further data analysis such as correlation and batch effects are presented in Supplementary Materials (S2).
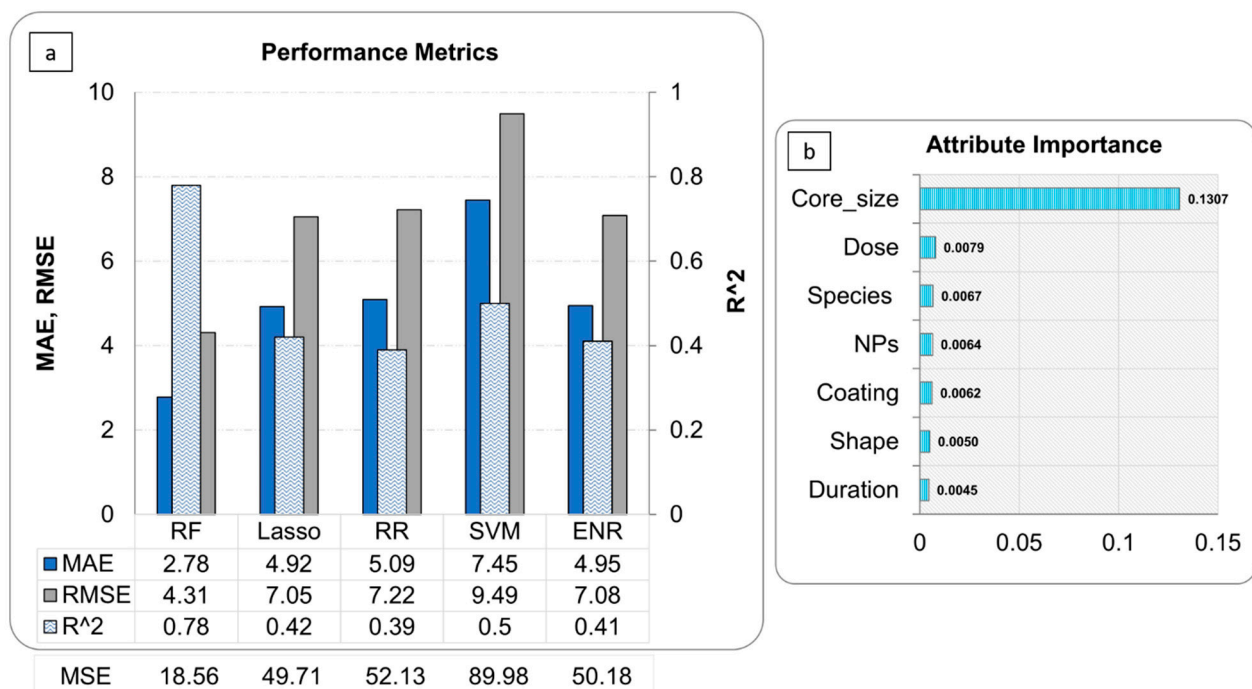

Figure 4. Performance metrics (**a**) and Random Forest Attribute Importance analysis results (**b**).

4. Discussion

In the present study, we implemented an ML tool from data collection to model validation, to predict the bactericidal effects induced by NPs in in-vitro systems. The model is consistent with the OECD principles [112] addressing the selection of (i) a defined endpoint (zone of inhibition as a metric to evaluate the susceptibility of the bacteria to NPs); (ii) an explicit algorithm (RF, https://github.com/mahsa-mirzaei/RFR_ABA.git, accessed on 6 July 2021); (iii) a well-defined domain of applicability (data ranges and nominal categories are provided, Table 1); and (iv) appropriate measures of goodness-of-fit; robustness and predictability.

Recent studies have demonstrated that p-chem properties such as core-size [113], shape [114], surface area [115], zeta potential [116], aggregation [117], and hydrodynamic size [115] play an important role in determining the antibacterial activity of NPs [116,118].

Exposure conditions such as dose play an important role as well [119]. For the above reasons, we gathered information regarding the p-chem properties of NPs and exposure conditions. However, the appearance of missing data among our input was significant. For instance, specific surface area, hydrodynamic size, and zeta potential were absent in more than 90% of our data, aggregation information was 39% missing. Although this is essential information, regression tools do not perform well with missing data. P-chem properties are important factors and should be mentioned in future studies. Another point is the appearance of multiple diverse coatings found in the literature. In our dataset we found thirty-seven coatings (less than 4% each and 43% uncoated). For the moment, the dataset is not sufficiently large to distinguish the influence and variance of each coating to the antibacterial capacity since the models overfit (reduction in predictive power).

In the studies reviewed, different methods to determine the NP size and morphology, such as Transmission Electron Microscopy (TEM), Scanning Electron Microscopy (SEM), Differential Mobility Analyzer (DMA), Dynamic Light Scattering (DLS), X-ray scattering, and UV–vis absorption spectrum, were reported. The presence of coating was assessed using electron microscopy combined with X-ray (energy-dispersive X-ray spectroscopy, EDX) or X-ray fluorescence spectroscopy (XRF) measurements. The specific surface area was measured by N2-BET, Ultra X-ray photoelectron spectroscopy technique, and NMR. The zeta potential was measured in different media (water and culture media) by electrophoretic mobility using Henry's Equation and the Schmolukowski approximation. The above summary signifies the need for a standardized characterization workflow to obtain homogenized data across different studies.

The reviewed studies captured the antibacterial effects of NPs by using different protocols. To create a consistent dataset, the experimental data should be generated by a single protocol, but this is impractical. For the outcome we focused on one antibacterial evaluation method to obtain uniform data, which represented only 40% of the gathered data. Subsequently, it demonstrates the need for harmonized and rigorous methods used to evaluate the antibacterial activity of NPs to accomplish reproducibility as well as reliability.

RF has been effectively utilized in various domains, e.g., in microbiology and genetics, and has become a major data analysis tool due to its superior performance on high dimensional data [101,120,121]. Our results show that RF predicts the antibacterial effect more acutely compared to other models due to the determination of the non-linear relationship between input and output variables. The second-best model was LASSO. The key challenge with LASSO is correlated variables, in that it retains one variable and sets the other to zero. This will lead to some loss of information resulting in lower accuracy. In order to evaluate the reliability and performance of the resulting models, we assessed the goodness-of-fit, robustness and predictivity by MSE, MAE, RMSE, and R^2 statical methods. The closer the value of R^2 (measures of goodness-of-fit) to 1, the better the model is fitted [122]; smaller values of MSE, RMSE, and MAE verified model performance [123,124].

Several techniques exist for pre-processing data to make them suitable for use in computational tools, such as normalization, one hot encoding, and feature selection handling of missing values. One technique worth noting is the description of molecular structures [125]. The most common methods to codify structures are (i) the chemical graph; (ii) the notations as Simplified Molecular Input-Line Entry System (SMILES); and iii) the de-facto standard chemical formats. Experimental and exposure conditions are vital variables in the representation of antibacterial capacity since the same type of NPs may exhibit diverse effects in different experimental conditions. This makes the development of classic QSAR difficult [126]. Toropova et al. [127] suggested a quasi-SMILES approach to represent molecular structures, p-chem properties, and experimental conditions (eclectic data) with NPs [126,128].

There was no ML study to compare with as to which parameters are the most important to predict the antibacterial activity of NPs on the model performance. We based our evaluation on the attributes that are investigated by researchers in the lab [129]. For

example, according to the literature, various parameters such as core size, dose, shape, and special surface area of the NPs affect the antibacterial activity of NPs [118,130].

- Size: The smaller the particle size, the higher the antibacterial activity [118]. This can be explained by the fact that NPs can easily cross the bacteria membrane and reach the nuclear; secondly, because of a larger surface/volume ratio [131–134].
- Species: Type of bacteria is important in determining the antibacterial activity of the NPs [135–137]. Depending on their cell wall composition, bacteria are divided into two groups: Gram-negative and Gram-positive [138]. Various NPs with different surface charges can act distinctly depending on what the differentiation is in the bacteria cell wall composition [132,135].
- Dose: A dose-dependent reduction of bacterial growth and biofilm biomass is observed following exposure to metal and metal oxide NPs [139,140]. Remarkably, our findings, according to the attribute important analysis, confirm that the core size, dose, and bacteria species are the most important attributes affecting the prediction of the antibacterial activity of NPs.

More systematic data are needed to enable building models accounting accurately for the all the possible in vitro determinant combinations. Without agreement on standard characterization workflow of NPs or defined key properties that define their efficacy and a lack of reference bacteria and defined assays, there is still a long way to go to unravel systematically the antibacterial properties of NPs. In addition to precise protocols and standardization of methods, there should be further harmonized outlines in how to report p-chem properties of NPs or experimental conditions and to make those measurements more comparable to improve the reporting data. The absence of comprehensive metadata description for related bioassays may have an impact on the clarity and comprehension and therefore the quality of the results [141]. Several different initiatives are currently working on defining frameworks, methods, and criteria for evaluating the quality of the reported data based on the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [142].

5. Conclusions

Antibiotic resistance of bacteria has become one of the major concerns in human healthcare. NPs represent a valuable and innovative technology to build alternatives to antibiotics. In this study, we investigate the performance of various ML tools to predict the effects of NPs as an antibacterial agent on vast groups of Gram-positive and Gram-negative bacteria, with RF being the best model. This study is a first step and the first tool to assist researchers towards screening NPs with potentially high antibacterial effects and could help fine-tune their properties. With this interdisciplinary approach we combine knowledge of the underlying science with computer science tools. This is a valuable activity as it allows those working in the laboratories to leverage development in the AI space and thus improve timeliness, reducing the number of experiments performed and the costs involved. Due to the inconsistency of reporting NP p-chem properties in antibacterial studies, the resulting dataset has large data gaps. We highlight the need for standardized measurements to evaluate the properties of NPs, allowing more consistent metadata. The majority of data in the literature revolves around the toxicity of NPs. This paper stresses the need for more data, raising awareness to the scientific community of the lack of comprehensive datasets regarding the antimicrobial capacity of NPs.

points that each study provided. The y axis demonstrates the actual and predicted zone of inhibition values. Experiment with Data points < 10.

Data Availability Statement: Data can be found in Supplementary Material S1.

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

Abbreviations

NPs	Nanoparticles
NMs	Nanomaterials
AL	Artificial Intelligence
ML	Machine Learning
MO NPs	Metal Oxide nanoparticles
ZnO NPs	Zinc Oxide nanoparticles
Fe_2O_3 NPs	Iron Oxide nanoparticles
Ag NPs	Silver nanoparticles
ROS	Reactive Oxygen Species
ECP	ExtraCellular Polymers
OD	Optical Density
ZOI	Zone Of Inhibition
MIC	Minimum Inhibitory Concentration
MBC	Minimum bactericidal Concentration
CFU	Colony Forming Unit
RF	Random Forest
ENR	Elastic Net Regression
SVM	Supervised Vector Machine
LASSO	Least Absolute Shrinkage and Selection Operator
RR	Ridge Regression

References

1. Cantón, R.; Morosini, M.-I. Emergence and spread of antibiotic resistance following exposure to antibiotics. *FEMS Microbiol. Rev.* **2011**, *35*, 977–991. [CrossRef]
2. Hajipour, M.J.; Fromm, K.M.; Ashkarran, A.A.; de Aberasturi, D.J.; de Larramendi, I.R.; Rojo, T.; Serpooshan, V.; Parak, W.J.; Mahmoudi, M. Antibacterial properties of nanoparticles. *Trends Biotechnol.* **2012**, *30*, 499–511. [CrossRef] [PubMed]
3. Nemeth, J.; Oesch, G.; Kuster, S.P. Bacteriostatic versus bactericidal antibiotics for patients with serious bacterial infections: Systematic review and meta-analysis. *J. Antimicrob. Chemother.* **2015**, *70*, 382–395. [CrossRef]
4. Pankey, G.A.; Sabath, L.D. Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of Gram-positive bacterial infections. *Clin. Infect. Dis.* **2004**, *38*, 864–870. [CrossRef] [PubMed]
5. von Nussbaum, F.; Brands, M.; Hinzen, B.; Weigand, S.; Häbich, D. Antibacterial natural products in medicinal chemistry–exodus or revival? *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 5072–5129. [CrossRef]
6. Witte, W. International dissemination of antibiotic resistant strains of bacterial pathogens. *Infect. Genet. Evol.* **2004**, *4*, 187–191. [CrossRef]
7. Vimbela, G.V.; Ngo, S.M.; Fraze, C.; Yang, L.; Stout, D.A. Antibacterial properties and toxicity from metallic nanomaterials. *Int. J. Nanomed.* **2017**, *12*, 3941–3965. [CrossRef]
8. Agarwal, S.; Yewale, V.N.; Dharmapalan, D. Antibiotics use and misuse in children: A knowledge, attitude and practice survey of parents in India. *J. Clin. Diagn. Res. JCDR* **2015**, *9*, SC21. [CrossRef]
9. Tangcharoensathien, V.; Chanvatik, S.; Sommanustweechai, A. Complex determinants of inappropriate use of antibiotics. *Bull. World Health Organ.* **2018**, *96*, 141–144. [CrossRef] [PubMed]
10. Nazar, C.J. Biofilms bacterianos. *Rev. Otorrinolaringol. Cirugía Cabeza Cuello* **2007**, *67*, 161–172. [CrossRef]

11. Jamal, M.; Ahmad, W.; Andleeb, S.; Jalil, F.; Imran, M.; Nawaz, M.A.; Hussain, T.; Ali, M.; Rafiq, M.; Kamil, M.A. Bacterial biofilm and associated infections. *J. Chin. Med. Assoc.* **2018**, *81*, 7–11. [CrossRef]

12. González, S.; Fernández, L.; Campelo, A.B.; Gutiérrez, D.; Martínez, B.; Rodríguez, A.; García, P. The behavior of Staphylococcus aureus dual-species biofilms treated with bacteriophage phiIPLA-RODI depends on the accompanying microorganism. *Appl. Environ. Microbiol.* **2017**, *83*, e02821-16. [CrossRef]

13. Di Pippo, F.; Di Gregorio, L.; Congestri, R.; Tandoi, V.; Rossetti, S. Biofilm growth and control in cooling water industrial systems. *FEMS Microbiol. Ecol.* **2018**, *94*. [CrossRef] [PubMed]

14. Liu, S.; Gunawan, C.; Barraud, N.; Rice, S.A.; Harry, E.J.; Amal, R. Understanding, monitoring, and controlling biofilm Growth in drinking water distribution systems. *Environ. Sci. Technol.* **2016**, *50*, 8954–8976. [CrossRef] [PubMed]

15. Stewart, P.S.; Costerton, J.W. Antibiotic resistance of bacteria in biofilms. *Lancet* **2001**, *358*, 135–138. [CrossRef]

16. Sharma, D.; Misba, L.; Khan, A.U. Antibiotics versus biofilm: An emerging battleground in microbial communities. *Antimicrob. Resist. Infect. Control* **2019**, *8*, 76. [CrossRef]

17. Alavi, M.; Rai, M. Recent advances in antibacterial applications of metal nanoparticles (MNPs) and metal nanocomposites (MNCs) against multidrug-resistant (MDR) bacteria. *Expert Rev. Anti-Infect. Ther.* **2019**, *17*, 419–428. [CrossRef] [PubMed]

18. Paddle-Ledinek, J.E.; Nasa, Z.; Cleland, H.J. Effect of different wound dressings on cell viability and proliferation. *Plast. Reconstr. Surg.* **2006**, *117*, 110S–118S. [CrossRef] [PubMed]

19. Samuel, U.; Guggenbichler, J. Prevention of catheter-related infections: The potential of a new nano-silver impregnated catheter. *Int. J. Antimicrob. Agents* **2004**, *23*, 75–78. [CrossRef]

20. Gosheger, G.; Hardes, J.; Ahrens, H.; Streitburger, A.; Buerger, H.; Erren, M.; Gunsel, A.; Kemper, F.H.; Winkelmann, W.; Von Eiff, C. Silver-coated megaendoprostheses in a rabbit model—An analysis of the infection rate and toxicological side effects. *Biomaterials* **2004**, *25*, 5547–5556. [CrossRef]

21. Han, C.; Romero, N.; Fischer, S.; Dookran, J.; Berger, A.; Doiron, A.L. Recent developments in the use of nanoparticles for treatment of biofilms. *Nanotechnol. Rev.* **2017**, *6*, 383–404. [CrossRef]

22. Leid, J.G.; Ditto, A.J.; Knapp, A.; Shah, P.N.; Wright, B.D.; Blust, R.; Christensen, L.; Clemons, C.B.; Wilber, J.P.; Young, G.W.; et al. In vitro antimicrobial studies of silver carbene complexes: Activity of free and nanoparticle carbene formulations against clinical isolates of pathogenic bacteria. *J. Antimicrob. Chemother.* **2012**, *67*, 138–148. [CrossRef]

23. Hariharan, H.; Al-Harbi, N.; Karuppiah, P.; Rajaram, S. Microbial synthesis of selenium nanocomposite using Saccharomyces cerevisiae and its antimicrobial activity against pathogens causing nosocomial infection. *Chalcogenide Lett.* **2012**, *9*, 509–515.

24. Saqib, S.; Zaman, W.; Ullah, F.; Majeed, I.; Ayaz, A.; Hussain Munis, M.F. Organometallic assembling of chitosan-Iron oxide nanoparticles with their antifungal evaluation against Rhizopus oryzae. *Appl. Organomet. Chem.* **2019**, *33*, e5190. [CrossRef]

25. Asghar, M.; Habib, S.; Zaman, W.; Hussain, S.; Ali, H.; Saqib, S. Synthesis and characterization of microbial mediated cadmium oxide nanoparticles. *Microsc. Res. Tech.* **2020**, *83*, 1574–1584. [CrossRef]

26. Bhati-Kushwaha, H.; Malik, C. Assessment of antibacterial and antifungal activities of silver nanoparticles obtained from the callus extracts (stem and leaf) of *Tridax procumbens* L. *Indian J. Biotechnol.* **2014**, *13*, 114–120.

27. Bhavnani, S.M.; Krause, K.M.; Ambrose, P.G. A broken antibiotic market: Review of strategies to incentivize drug development. *Open Forum Infect. Dis.* **2020**, *7*, ofaa083. [CrossRef] [PubMed]

28. Bush, K.; Pucci, M.J. New antimicrobial agents on the horizon. *Biochem. Pharmacol.* **2011**, *82*, 1528–1539. [CrossRef]

29. Cheesman, M.J.; Ilanko, A.; Blonk, B.; Cock, I.E. Developing new antimicrobial therapies: Are synergistic combinations of plant extracts/compounds with conventional antibiotics the solution? *Pharmacogn. Rev.* **2017**, *11*, 57. [PubMed]

30. Dizaj, S.M.; Lotfipour, F.; Barzegar-Jalali, M.; Zarrintan, M.H.; Adibkia, K. Antimicrobial activity of the metals and metal oxide nanoparticles. *Mater. Sci. Eng. C* **2014**, *44*, 278–284. [CrossRef]

31. Niño-Martínez, N.; Salas Orozco, M.F.; Martínez-Castañón, G.-A.; Torres Méndez, F.; Ruiz, F. Molecular Mechanisms of Bacterial Resistance to Metal and Metal Oxide Nanoparticles. *Int. J. Mol. Sci.* **2019**, *20*, 2808. [CrossRef]

32. Wang, L.; Hu, C.; Shao, L. The antimicrobial activity of nanoparticles: Present situation and prospects for the future. *Int. J. Nanomed.* **2017**, *12*, 1227–1249. [CrossRef]

33. Hu, L. The Use of Nanoparticles to Prevent and Eliminate Bacterial Biofilms. Ph. D. Thesis, Laboratory of Enteric and Sexually Transmitted Diseases, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD, USA, 2017.

34. Tran, N.; Mir, A.; Mallik, D.; Sinha, A.; Nayar, S.; Webster, T.J. Bactericidal effect of iron oxide nanoparticles on Staphylococcus aureus. *Int. J. Nanomed.* **2010**, *5*, 277–283. [CrossRef]

35. Hayat, S.; Muzammil, S.; Aslam, B.; Siddique, M.H.; Saqalein, M.; Nisar, M.A. Quorum quenching: Role of nanoparticles as signal jammers in Gram-negative bacteria. *Future Microbiol.* **2019**, *14*, 61–72. [CrossRef]

36. Szabó, T.; Németh, J.; Dékány, I. Zinc oxide nanoparticles incorporated in ultrathin layer silicate films and their photocatalytic properties. *Colloids Surf. A Physicochem. Eng. Asp.* **2003**, *230*, 23–35. [CrossRef]

37. Baek, Y.W.; An, Y.J. Microbial toxicity of metal oxide nanoparticles (CuO, NiO, ZnO, and Sb_2O_3) to Escherichia coli, Bacillus subtilis, and Streptococcus aureus. *Sci. Total Environ.* **2011**, *409*, 1603–1608. [CrossRef]

38. Pan, X.; Redding, J.E.; Wiley, P.A.; Wen, L.; McConnell, J.S.; Zhang, B. Mutagenicity evaluation of metal oxide nanoparticles by the bacterial reverse mutation assay. *Chemosphere* **2010**, *79*, 113–116. [CrossRef] [PubMed]

39. Thill, A.; Zeyons, O.; Spalla, O.; Chauvat, F.; Rose, J.; Auffan, M.; Flank, A.M. Cytotoxicity of CeO2 nanoparticles for Escherichia coli. Physico-chemical insight of the cytotoxicity mechanism. *Environ. Sci. Technol.* **2006**, *40*, 6151–6156. [CrossRef] [PubMed]

40. Manivasagan, P.; Venkatesan, J.; Senthilkumar, K.; Sivakumar, K.; Kim, S.-K. Biosynthesis, antimicrobial and cytotoxic effect of silver nanoparticles using a novel Nocardiopsis sp. MBRC-1. *BioMed. Res. Int.* **2013**, *2013*, 287638. [CrossRef]
41. Sondi, I.; Salopek-Sondi, B. Silver nanoparticles as antimicrobial agent: A case study on E. coli as a model for Gram-negative bacteria. *J. Colloid Interface Sci.* **2004**, *275*, 177–182. [CrossRef] [PubMed]
42. Azam, Z.; Ayaz, A.; Younas, M.; Qureshi, Z.; Arshad, B.; Zaman, W.; Ullah, F.; Nasar, M.Q.; Bahadur, S.; Irfan, M.M.; et al. Microbial synthesized cadmium oxide nanoparticles induce oxidative stress and protein leakage in bacterial cells. *Microb. Pathog.* **2020**, *144*, 104188. [CrossRef] [PubMed]
43. Prasher, P.; Singh, M.; Mudila, H. Oligodynamic effect of silver nanoparticles: A review. *BioNanoScience* **2018**, *8*, 951–962. [CrossRef]
44. Nel, A. Atmosphere. Air pollution-related illness: Effects of particles. *Science* **2005**, *308*, 804–806. [CrossRef] [PubMed]
45. Soenen, S.J.; Rivera-Gil, P.; Montenegro, J.-M.; Parak, W.J.; De Smedt, S.C.; Braeckmans, K. Cellular toxicity of inorganic nanoparticles: Common aspects and guidelines for improved nanotoxicity evaluation. *Nano Today* **2011**, *6*, 446–465. [CrossRef]
46. Nel, A.E.; Mädler, L.; Velegol, D.; Xia, T.; Hoek, E.M.; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M. Understanding biophysicochemical interactions at the nano–bio interface. *Nat. Mater.* **2009**, *8*, 543–557. [CrossRef] [PubMed]
47. Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A.I. Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles. *J. Phys. Chem. Lett.* **2017**, *8*, 5091–5098. [CrossRef] [PubMed]
48. Lu, L.; Guo, X.; Zhao, J. A unified nonlocal strain gradient model for nanobeams and the importance of higher order terms. *Int. J. Eng. Sci.* **2017**, *119*, 265–277. [CrossRef]
49. Ban, Z.; Yuan, P.; Yu, F.; Peng, T.; Zhou, Q.; Hu, X. Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 10492. [CrossRef]
50. Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C.A. Nanotoxicology data for in silico tools: A literature review. *Nanotoxicology* **2020**, *14*, 612–637. [CrossRef]
51. Furxhi, I.; Murphy, F.; Poland, C.A.; Sheehan, B.; Mullins, M.; Mantecca, P. Application of Bayesian networks in determining nanoparticle-induced cellular outcomes using transcriptomics. *Nanotoxicology* **2019**, *13*, 827–848. [CrossRef]
52. Luan, F.; Kleandrova, V.V.; González-Díaz, H.; Ruso, J.M.; Melo, A.; Speck-Planche, A.; Cordeiro, M.N.D. Computer-aided nanotoxicology: Assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale* **2014**, *6*, 10623–10630. [CrossRef]
53. Furxhi, I.; Murphy, F. Predicting in vitro neurotoxicity induced by nanoparticles using machine learning. *Int. J. Mol. Sci.* **2020**, *21*, 5280. [CrossRef] [PubMed]
54. 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]
55. Kleandrova, V.V.; Luan, F.; González-Díaz, H.; Ruso, J.M.; Speck-Planche, A.; Cordeiro, M.N.l.D. Computational tool for risk assessment of nanomaterials: Novel QSTR-perturbation model for simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under multiple experimental conditions. *Environ. Sci. Technol.* **2014**, *48*, 14686–14694. [CrossRef] [PubMed]
56. Kleandrova, V.V.; Luan, F.; González-Díaz, H.; Ruso, J.M.; Melo, A.; Speck-Planche, A.; Cordeiro, M.N.D. Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. *Environ. Int.* **2014**, *73*, 288–294. [CrossRef] [PubMed]
57. Serafim, M.S.M.; Kronenberger, T.; Oliveira, P.R.; Poso, A.; Honório, K.M.; Mota, B.E.F.; Maltarollo, V.G. The application of machine learning techniques to innovative antibacterial discovery and development. *Expert Opin. Drug Discov.* **2020**, *15*, 1165–1180. [CrossRef]
58. Durrant, J.D.; Amaro, R.E. Machine-learning techniques applied to antibacterial drug discovery. *Chem. Biol. Drug Des.* **2015**, *85*, 14–21. [CrossRef]
59. Her, H.-L.; Wu, Y.-W. A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains. *Bioinformatics* **2018**, *34*, i89–i95. [CrossRef]
60. Khaledi, A.; Weimann, A.; Schniederjans, M.; Asgari, E.; Kuo, T.H.; Oliver, A.; Cabot, G.; Kola, A.; Gastmeier, P.; Hogardt, M. Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics. *EMBO Mol. Med.* **2020**, *12*, e10264. [CrossRef]
61. Liu, Z.; Deng, D.; Lu, H.; Sun, J.; Lv, L.; Li, S.; Peng, G.; Ma, X.; Li, J.; Li, Z. Evaluation of machine learning models for predicting antimicrobial resistance of Actinobacillus pleuropneumoniae from whole genome sequences. *Front. Microbiol.* **2020**, *11*, 48. [CrossRef]
62. Yang, Y.; Niehaus, K.E.; Walker, T.M.; Iqbal, Z.; Walker, A.S.; Wilson, D.J.; Peto, T.E.; Crook, D.W.; Smith, E.G.; Zhu, T. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. *Bioinformatics* **2018**, *34*, 1666–1671. [CrossRef]
63. Vishwakarma, V. Impact of environmental biofilms: Industrial components and its remediation. *J. Basic Microbiol.* **2020**, *60*, 198–206. [CrossRef]
64. Gal, M.S.; Rubinfeld, D.L. Data standardization. *NYUL Rev.* **2019**, *94*, 737. [CrossRef]
65. Potdar, K.; Pardawala, T.S.; Pai, C. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. *Int. J. Comput. Appl.* **2017**, *175*, 7–9. [CrossRef]
66. Jakulin, A. Machine Learning Based on Attribute Interactions. Ph.D. Thesis, Univerza v Ljubljani, Ljubljana, Slovenia, 2005.

67. Jakulin, A.; Bratko, I.; Smrke, D.; Demšar, J.; Zupan, B. Attribute interactions in medical data analysis. In Proceedings of the Conference on Artificial Intelligence in Medicine in Europe, Protaras, Cyprus, 18–22 October 2003; Springer: Berlin/Heidelberg, Germany, 2003; pp. 229–238.

68. Osibe, D.A.; Chiejina, N.V.; Ogawa, K.; Aoyagi, H. Stable antibacterial silver nanoparticles produced with seed-derived callus extract of Catharanthus roseus. *Artif. Cells Nanomed. Biotechnol.* **2018**, *46*, 1266–1273. [CrossRef]

69. Singh, A.; Gautam, P.K.; Verma, A.; Singh, V.; Shivapriya, P.M.; Shivalkar, S.; Sahoo, A.K.; Samanta, S.K. Green synthesis of metallic nanoparticles as effective alternatives to treat antibiotics resistant bacterial infections: A review. *Biotechnol. Rep.* **2020**, *25*, e00427. [CrossRef] [PubMed]

70. Stankic, S.; Suman, S.; Haque, F.; Vidic, J. Pure and multi metal oxide nanoparticles: Synthesis, antibacterial and cytotoxic properties. *J. Nanobiotechnol.* **2016**, *14*, 73. [CrossRef]

71. Li, Y.; Zhang, W.; Niu, J.; Chen, Y. Mechanism of photogenerated reactive oxygen species and correlation with the antibacterial properties of engineered metal-oxide nanoparticles. *ACS Nano* **2012**, *6*, 5164–5173. [CrossRef] [PubMed]

72. Reddy, K.M.; Feris, K.; Bell, J.; Wingett, D.G.; Hanley, C.; Punnoose, A. Selective toxicity of zinc oxide nanoparticles to prokaryotic and eukaryotic systems. *Appl. Phys. Lett.* **2007**, *90*, 213902. [CrossRef] [PubMed]

73. Li, R.; Chen, Z.; Ren, N.; Wang, Y.; Wang, Y.; Yu, F. Biosynthesis of silver oxide nanoparticles and their photocatalytic and antimicrobial activity evaluation for wound healing applications in nursing care. *J. Photochem. Photobiol. B Biol.* **2019**, *199*, 111593. [CrossRef] [PubMed]

74. Huh, A.J.; Kwon, Y.J. "Nanoantibiotics": A new paradigm for treating infectious diseases using nanomaterials in the antibiotics resistant era. *J. Control. Release* **2011**, *156*, 128–145. [CrossRef]

75. Dastjerdi, R.; Montazer, M. A review on the application of inorganic nano-structured materials in the modification of textiles: Focus on anti-microbial properties. *Colloids Surf. B Biointerfaces* **2010**, *79*, 5–18. [CrossRef] [PubMed]

76. Applerot, G.; Lellouche, J.; Perkas, N.; Nitzan, Y.; Gedanken, A.; Banin, E. ZnO nanoparticle-coated surfaces inhibit bacterial biofilm formation and increase antibiotic susceptibility. *Rsc Adv.* **2012**, *2*, 2314–2321. [CrossRef]

77. Blecher, K.; Nasir, A.; Friedman, A. The growing role of nanotechnology in combating infectious disease. *Virulence* **2011**, *2*, 395–401. [CrossRef]

78. Nohynek, G.J.; Lademann, J.; Ribaud, C.; Roberts, M.S. Grey goo on the skin? Nanotechnology, cosmetic and sunscreen safety. *Crit. Rev. Toxicol.* **2007**, *37*, 251–277. [CrossRef]

79. Pradheesh, G.; Suresh, S.; Suresh, J.; Alexramani, V. Antimicrobial and anticancer activity studies on green synthesized silver oxide nanoparticles from the medicinal plant cyathea nilgiriensis holttum. *Int. J. Pharm. Investig.* **2020**, *10*, 146–150. [CrossRef]

80. Agarwal, H.; Gayathri, M. Biological synthesis of nanoparticles from medicinal plants and its uses in inhibiting biofilm formation. *Asian J. Pharm. Clin. Res.* **2017**, *10*, 64–68. [CrossRef]

81. Polli, J.E. In vitro studies are sometimes better than conventional human pharmacokinetic in vivo studies in assessing bioequivalence of immediate-release solid oral dosage forms. *AAPS J.* **2008**, *10*, 289–299. [CrossRef]

82. Dreaden, E.C.; Alkilany, A.M.; Huang, X.; Murphy, C.J.; El-Sayed, M.A. The golden age: Gold nanoparticles for biomedicine. *Chem. Soc. Rev.* **2012**, *41*, 2740–2779. [CrossRef] [PubMed]

83. Khan, I.; Saeed, K.; Khan, I. Nanoparticles: Properties, applications and toxicities. *Arab. J. Chem.* **2019**, *12*, 908–931. [CrossRef]

84. Djurišić, A.B.; Leung, Y.H.; Ng, A.M.C.; Xu, X.Y.; Lee, P.K.H.; Degger, N.; Wu, R.S.S. Toxicity of metal oxide nanoparticles: Mechanisms, characterization, and avoiding experimental artefacts. *Small* **2015**, *11*, 26–44. [CrossRef] [PubMed]

85. Beale, E.M.; Little, R.J. Missing values in multivariate analysis. *J. R. Stat. Soc. Ser. B* **1975**, *37*, 129–145. [CrossRef]

86. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [CrossRef]

87. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Elsevier: Amsterdam, The Netherlands, 2013.

88. Reitermanova, Z. Data splitting. In Proceedings of the 19th Annual Conference of Doctoral Students—WDS 2010, Prague, Czech Republic, 1–4 June 2010; Part I. pp. 31–36.

89. Doan, T.; Kalita, J. Selecting machine learning algorithms using regression models. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14–17 November 2015; pp. 1498–1505.

90. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. Regression models. In *Regression*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 21–72.

91. Meier, L.; Van De Geer, S.; Bühlmann, P. The group lasso for logistic regression. *J. R. Stat. Soc. Ser. B* **2008**, *70*, 53–71. [CrossRef]

92. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]

93. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]

94. García, J.; Salmerón, R.; García, C.; López Martín, M.d.M. Standardization of Variables and Collinearity Diagnostic in Ridge Regression. *Int. Stat. Rev.* **2016**, *84*, 245–266. [CrossRef]

95. Curto, J.D.; Pinto, J.C. New multicollinearity indicators in linear regression models. *Int. Stat. Rev./Rev. Int. De Stat.* **2007**, *75*, 114–121. [CrossRef]

96. Assaf, A.G.; Tsionas, M.; Tasiopoulos, A. Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression. *Tour. Manag.* **2019**, *71*, 1–8. [CrossRef]

97. McDonald, G.C. Ridge regression. *WIREs Comput. Stat.* **2009**, *1*, 93–100. [CrossRef]

98. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [CrossRef]

99. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

100. Menze, B.H.; Kelm, B.M.; Masuch, R.; Himmelreich, U.; Bachert, P.; Petrich, W.; Hamprecht, F.A. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinform.* **2009**, *10*, 213. [CrossRef] [PubMed]

101. Breiman, L. *Consistency for a Simple Model of Random Forests, Technical Report 670*; Statistics Department, University of California at Berkeley: Berkeley, CA, USA, 2004.

102. Satapathy, S.K.; Dehuri, S.; Jagadev, A.K.; Mishra, S. Chapter 1—Introduction. In *EEG Brain Signal Classification for Epileptic Seizure Disorder Detection*; Satapathy, S.K., Dehuri, S., Jagadev, A.K., Mishra, S., Eds.; Academic Press: Cambridge, MA, USA, 2019; pp. 1–25. [CrossRef]

103. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.

104. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-validation. *Encycl. Database Syst.* **2009**, *5*, 532–538.

105. Picard, R.R.; Cook, R.D. Cross-validation of regression models. *J. Am. Stat. Assoc.* **1984**, *79*, 575–583. [CrossRef]

106. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef]

107. Ericsson, B.H.; Tunevall, G.; Wickman, K. The Paper Disc Method for Determination of Bacterial Sensitivity to Antibiotics: Relationship between the Diameter of the Zone of Inhibition and the Minimum Inhibitory Concentration. *Scand. J. Clin. Lab. Investig.* **1960**, *12*, 414–422. [CrossRef] [PubMed]

108. Taylor, P.C.; Schoenknecht, F.D.; Sherris, J.C.; Linner, E.C. Determination of minimum bactericidal concentrations of oxacillin for Staphylococcus aureus: Influence and significance of technical factors. *Antimicrob. Agents Chemother.* **1983**, *23*, 142–150. [CrossRef] [PubMed]

109. Sutton, S. Measurement of microbial cells by optical density. *J. Valid. Technol.* **2011**, *17*, 46–49.

110. Hudzicki, J. *Kirby-Bauer Disk Diffusion Susceptibility Test Protocol*; American Society for Microbiology: Washington, DC, USA, 2009.

111. Barnard, R.T. The Zone of Inhibition. *Clin. Chem.* **2019**, *65*, 819. [CrossRef]

112. OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models No. 69*; OECD Publishing: Paris, France, 2014. [CrossRef]

113. Lallo da Silva, B.; Caetano, B.L.; Chiari-Andréo, B.G.; Pietro, R.C.L.R.; Chiavacci, L.A. Increased antibacterial activity of ZnO nanoparticles: Influence of size and surface modification. *Colloids Surf. B Biointerfaces* **2019**, *177*, 440–447. [CrossRef]

114. Pal, S.; Tak, Y.K.; Song, J.M. Does the antibacterial activity of silver nanoparticles depend on the shape of the nanoparticle? A study of the gram-negative bacterium *Escherichia coli*. *Appl. Environ. Microbiol.* **2007**, *73*, 1712. [CrossRef]

115. Suttiponparnit, K.; Jiang, J.; Sahu, M.; Suvachittanont, S.; Charinpanitkul, T.; Biswas, P. Role of surface area, primary particle size, and crystal phase on titanium dioxide nanoparticle dispersion properties. *Nanoscale Res. Lett.* **2011**, *6*, 1–8. [CrossRef] [PubMed]

116. Pan, X.; Wang, Y.; Chen, Z.; Pan, D.; Cheng, Y.; Liu, Z.; Lin, Z.; Guan, X. Investigation of antibacterial activity and related mechanism of a series of nano-Mg(OH)$_2$. *ACS Appl. Mater. Interfaces* **2013**, *5*, 1137–1142. [CrossRef]

117. Le Ouay, B.; Stellacci, F. Antibacterial activity of silver nanoparticles: A surface science insight. *Nano Today* **2015**, *10*, 339–354. [CrossRef]

118. Tang, S.; Zheng, J. Antibacterial activity of silver nanoparticles: Structural effects. *Adv. Healthc. Mater.* **2018**, *7*, 1701503. [CrossRef] [PubMed]

119. Pazos-Ortiz, E.; Roque-Ruiz, J.H.; Hinojos-Márquez, E.A.; López-Esparza, J.; Donohué-Cornejo, A.; Cuevas-González, J.C.; Espinosa-Cristóbal, L.F.; Reyes-López, S.Y. Dose-dependent antimicrobial activity of silver nanoparticles on polycaprolactone fibers against gram-positive and gram-negative bacteria. *J. Nanomater.* **2017**, *2017*, 4752314. [CrossRef]

120. Heidema, A.G.; Boer, J.M.A.; Nagelkerke, N.; Mariman, E.C.M.; Feskens, E.J.M. The challenge for genetic epidemiologists: How to analyze large numbers of SNPs in relation to complex diseases. *BMC Genet.* **2006**, *7*, 23. [CrossRef]

121. Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*, 3. [CrossRef] [PubMed]

122. Vidyullatha, P.; Rao, D.R. Machine learning techniques on multidimensional curve fitting data based on R-square and chi-square methods. *Int. J. Electr. Comput. Eng.* **2016**, *6*, 974.

123. Yamaç, S.S.; Todorovic, M. Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agric. Water Manag.* **2020**, *228*, 105875. [CrossRef]

124. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

125. 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] [PubMed]

126. Choi, J.-S.; Trinh, T.X.; Yoon, T.-H.; Kim, J.; Byun, H.-G. Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. *Chemosphere* **2019**, *217*, 243–249. [CrossRef]

127. Toropova, A.P.; Toropov, A.A.; Benfenati, E. A quasi-QSPR modelling for the photocatalytic decolourization rate constants and cellular viability (CV%) of nanoparticles by CORAL. *SAR QSAR Environ. Res.* **2015**, *26*, 29–40. [CrossRef] [PubMed]

128. Trinh, T.X.; Choi, J.-S.; Jeon, H.; Byun, H.-G.; Yoon, T.-H.; Kim, J. Quasi-SMILES-based nano-quantitative structure–activity relationship model to predict the cytotoxicity of multiwalled carbon nanotubes to human lung cells. *Chem. Res. Toxicol.* **2018**, *31*, 183–190. [CrossRef]

129. Slavin, Y.N.; Asnis, J.; Häfeli, U.O.; Bach, H. Metal nanoparticles: Understanding the mechanisms behind antibacterial activity. *J. Nanobiotechnol.* **2017**, *15*, 1–20. [CrossRef] [PubMed]

130. Shrivastava, S.; Bera, T.; Roy, A.; Singh, G.; Ramachandrarao, P.; Dash, D. Characterization of enhanced antibacterial effects of novel silver nanoparticles. *Nanotechnology* **2007**, *18*, 225103. [CrossRef]

131. Martínez-Castañon, G.-A.; Nino-Martinez, N.; Martinez-Gutierrez, F.; Martinez-Mendoza, J.; Ruiz, F. Synthesis and antibacterial activity of silver nanoparticles with different sizes. *J. Nanoparticle Res.* **2008**, *10*, 1343–1348. [CrossRef]

132. Abbaszadegan, A.; Ghahramani, Y.; Gholami, A.; Hemmateenejad, B.; Dorostkar, S.; Nabavizadeh, M.; Sharghi, H. The effect of charge at the surface of silver nanoparticles on antimicrobial activity against gram-positive and gram-negative bacteria: A preliminary study. *J. Nanomater.* **2015**, *2015*, 720654. [CrossRef]

133. Xiu, Z.-M.; Zhang, Q.-B.; Puppala, H.L.; Colvin, V.L.; Alvarez, P.J. Negligible particle-specific antibacterial activity of silver nanoparticles. *Nano Lett.* **2012**, *12*, 4271–4275. [CrossRef] [PubMed]

134. Inphonlek, S.; Pimpha, N.; Sunintaboon, P. Synthesis of poly (methyl methacrylate) core/chitosan-mixed-polyethyleneimine shell nanoparticles and their antibacterial property. *Colloids Surf. B Biointerfaces* **2010**, *77*, 219–226. [CrossRef]

135. Pajerski, W.; Ochonska, D.; Brzychczy-Wloch, M.; Indyka, P.; Jarosz, M.; Golda-Cepa, M.; Sojka, Z.; Kotarba, A. Attachment efficiency of gold nanoparticles by Gram-positive and Gram-negative bacterial strains governed by surface charges. *J. Nanoparticle Res.* **2019**, *21*, 186. [CrossRef]

136. Azam, A.; Ahmed, A.S.; Oves, M.; Khan, M.S.; Habib, S.S.; Memic, A. Antimicrobial activity of metal oxide nanoparticles against Gram-positive and Gram-negative bacteria: A comparative study. *Int. J. Nanomed.* **2012**, *7*, 6003–6009. [CrossRef]

137. Emami-Karvani, Z.; Chehrazi, P. Antibacterial activity of ZnO nanoparticle on gram-positive and gram-negative bacteria. *Afr. J. Microbiol. Res.* **2011**, *5*, 1368–1373.

138. Silhavy, T.J.; Kahne, D.; Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000414. [CrossRef] [PubMed]

139. Tiwari, D.K.; Behari, J.; Sen, P. Time and dose-dependent antimicrobial potential of Ag nanoparticles synthesized by top-down approach. *Curr. Sci.* **2008**, *95*, 647–655.

140. Awasthi, A.; Sharma, P.; Jangir, L.; Awasthi, G.; Awasthi, K.K.; Awasthi, K. Dose dependent enhanced antibacterial effects and reduced biofilm activity against Bacillus subtilis in presence of ZnO nanoparticles. *Mater. Sci. Eng. C* **2020**, *113*, 111021. [CrossRef]

141. Klimisch, H.-J.; Andreae, M.; Tillmann, U. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul. Toxicol. Pharmacol.* **1997**, *25*, 1–5. [CrossRef]

142. Thompson, M.; Burger, K.; Kaliyaperumal, R.; Roos, M.; Santos, L.O.B.d.S. Making FAIR Easy with FAIR Tools: From Creolization to Convergence. *Data Intell.* **2020**, *2*, 87–95. [CrossRef]