

Research Article

Prediction of the Antibacterial Activity of the Green Synthesized Silver Nanoparticles against Gram Negative and Positive Bacteria by using Machine Learning Algorithms

Afshin Saadat ¹, Asma Dehghani Varniab ², and Seyyed Mohammad Madani ²

¹Department of Chemistry, Germe Branch, Islamic Azad University, Germe, Iran

²Department of Computer Engineering, University of Mohaghegh Ardabili, Ardabil, Iran

Correspondence should be addressed to Afshin Saadat; afshin.saadat@iau.ac.ir

Received 24 June 2022; Revised 17 October 2022; Accepted 27 October 2022; Published 6 December 2022

Academic Editor: Angelo Taglietti

Copyright © 2022 Afshin Saadat et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the appearance and growth of microbial organisms resistant to various antibiotics, as well as the need to reduce the cost of care of health, the production of antimicrobials at lower costs has become an inescapable necessity for today's human societies. Recently, the interdisciplinary field of nanotechnology has developed widely. One of the applications of nanobiotechnology is the use of silver nanoparticles (AgNPs) for new solutions in the treatment of microbial infections. AgNPs have unique properties which help in molecular diagnostics, therapies, and also in devices that are used in several medical procedures. In this field, machine learning algorithms have been used with hopeful results. One of the branches of artificial intelligence (AI) is machine learning (ML) that focuses on data and shows the power of the data. Machine learning techniques are taking considerable attention because of their obvious successes in a broad range of predictive tasks. In this work, we studied machine learning technique to predict the antibacterial activity of AgNPs against *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Klebsiella pneumoniae*. Here, we reviewed 100 articles for completing the data, highlighting the recently used different plants for the synthesis of highly efficient antimicrobial green AgNPs, which consist of key experimental conditions (amount of plant extract, volume of plant extract, volume of solvent, volume of AgNO₃ solution, reaction temperature, reaction time, concentration of precursors, and nanoparticle size). The results showed that nanoparticles size and concentration of AgNPs are key factors in predicting the antibacterial effect of AgNPs.

1. Introduction

Resistance to human pathogens is a major challenge in the field of pharmaceutical and medical science. This antibiotic resistance and continuous use of chemical compounds have caused an important phenomenon of resistance in microorganisms [1]. This phenomenon has weakened or neutralized the effects of medicines and eventually has led to an increase in drug use. It has also led to the willingness to use the compounds with newer and stronger formulation [2, 3]. Another disadvantage of using these drugs is their increased side effects leading to the reemergence of multidrug resistant pathogens and parasites, which create diseases that are more dangerous than primary disease [4]. Therefore, developing and making changes in antibacterial compounds for

enhancing their antibacterial potential of them have been an important part of research in recent years [5]. Nanotechnology as the most advanced technology of the present century has been able to penetrate all parts and angles of human life, animals, and plants. It has also affected its current and future status with its innovation [6–8]. In this field, it is a well-known fact that silver nanoparticles are highly toxic to microorganisms [9]. Silver nanoparticles technology which caused a dramatic revolution in the antibacterial materials is the main orientation for developing nanosilver products and has many advantages over chemical materials [10, 11]. Antibacterial effects of silver nanoparticles were investigated by many researchers and their potential effects against a wide range of microbes including antibiotic-resistant bacteria have been proved [12–17]. Many studies have been done on the

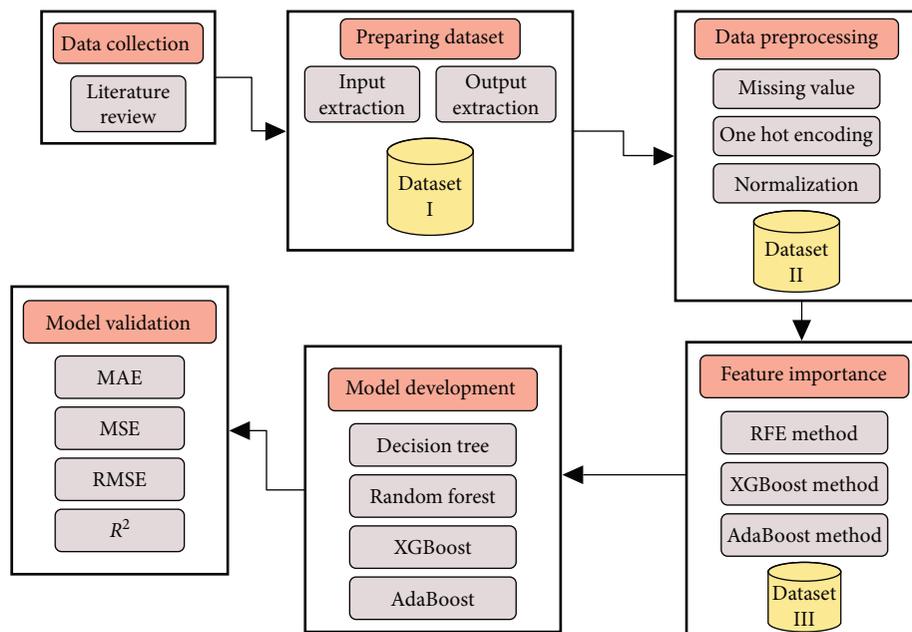


FIGURE 1: Workflow of data preprocessing.

possible relationship between physicochemical properties and toxicity of silver nanoparticles [18]. The shapes of the silver nanoparticles such as spherical, rods, beads, nanoprisms, and sheets are being developed and studied for their specific antibacterial effects [19–23]. Recently, studies have shown that surface coating and size of silver nanoparticles play an important role in antibacterial activity. They indicated that smaller nanoparticles deliver higher magnitude of toxicity [19, 24, 25]. Two main factors determine antibacterial activity of nanoparticles: (1) physicochemical properties, such as natural properties, surface modification, and composition and (2) the type of bacteria [26–28]. Computational tools can decrease the design space by predicting the characteristics of nanoparticles before synthesis, which helps to the reduction of experimental trial and error work. Tools are used for the prediction of the three-dimensional structure of metallic nanoparticles [29, 30], and also used to characterize the functional composition of the protein corona of nanoparticles [31]. One of the branches of computer science is artificial intelligence (AI). AI allows software applications to become more powerful and precise at predicting the result without programming explicitly. Because of this property, AI has attracted much attention in different fields. Machine learning (ML) is one of the subsets of AI and historical data are used by ML algorithms as input to predict the new output. Also, they apply information and data for their output estimation and do not use physical test materials, so they are inexpensive [32, 33]. Antimicrobial resistance for specific bacteria has been illustrated by ML. For instance, Tiihonen et al. [34] studied the antimicrobial activity of conjugated oligoelectrolyte molecules against *Escherichia coli*. Söylemez et al. [35] also utilized a ML for prediction of antimicrobial activity of peptides against gram negative and positive bacteria. Her and Wu [36] explored using a pan-genome-based ML method for prediction of the

antimicrobial resistance activities of *E. coli*. In this work, we used machine learning techniques to predict the antibacterial activity of AgNPs against *E. coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Klebsiella pneumoniae*. The antibacterial activity was predicted by studying experimental conditions (amount of plant extract, volume of plant extract, volume of solvent, volume of AgNO₃ solution, reaction temperature, reaction time, concentration of precursors, and nanoparticle size). This method helps in the prediction of antibacterial activity of AgNPs by saving time and decreasing costs. It also can be useful to researchers for preventing the growth of bacteria. As we know, bacterial growth is harmful for environmental and industrial components, especially for human health [37–40].

2. Materials and Methods

2.1. Approach. In Figure 1, we show workflow for our model step-by-step. After collecting data, we prepared and identified features and outcomes. Then, we created our dataset I and preprocessing phase started. In this step, we handled categorical and missing values and finally normalized them and created dataset II. In the third step, before training regression models, important features were selected by ML algorithms and feature selection algorithms. Then, dataset III was created based on effective features and outcomes. To find a suitable model, we trained some of regression algorithms and finally, we evaluated our models with different evaluation metrics.

2.2. Dataset. In this study, we studied researches that have been done on green synthesis and antibacterial activity of AgNPs. These studies contained different keywords, such as “silver nanoparticles,” “antimicrobial,” “green synthesis,” and “antibacterial” results. AgNPs are well known because of

their high stability, low cost, inhibition result against a wide range of bacteria, and their applications in food industry [41–43]. Therefore, AgNPs were selected as target antibacterial agents against gram-negative bacteria including *E. coli*, *P. aeruginosa*, and *K. pneumoniae*, and gram-positive bacteria including *S. aureus*. In this study, we reviewed 100 articles by concentrating on the amount of plant extract, volume of plant extract, volume of solvent, volume of AgNO₃ solution, reaction temperature, reaction time, concentration of precursors, and nanoparticle size. These variables were acquired as input attributes for the prediction of the antibacterial activity of the AgNPs. The outcomes were documented based on the antibacterial measurements such as zone of inhibition.

2.3. Data Preprocessing. The data preprocessing is one of the main stages in ML. It includes data cleaning, normalization, and transformation. It has considerable effect on generalization performance of supervised ML algorithms [44]. Missing data (or missing values) are defined as the data value that is not stored for a variable in the observation. Lack of values can cause problems and has a considerable effect on the results. Because of that, handling them is very important rather than ignoring or omitting them. There are some techniques that researchers can use and tackle them. In the primary dataset (dataset I), there are some missing values in features selection and since these missing values affect results in our regression model, we choose impute method for handling them [45]. Categorical variables take on values that are names or labels. ML algorithms cannot operate on label data directly. So, categorical variables in the regression model must be turned into the numerical form [46]. There are some methods for converting them to numerical form [47]. The process of converting categorical variable to numerical form is called “one hot encoding.” We utilized this method due to its high accuracy. By this process, machine learning algorithms can do a better job in prediction. Sometimes features in machine learning models have different ranges and they are not on the similar scale. In this case, normalization reveals its power. The goal of normalization is to transform the numeric columns in the dataset to the similar and common scale. Note that every dataset does not need normalization. Normalization is required when feature columns have various ranges. Normalization can be achieved by several methods. The most common methods are z-score and the min–max [44]. In this study, normalization was done by applying the min–max method. This method rescales the features in any range to a new range and this range usually is between (0, 1) and (−1, +1) [48].

$$\text{Normalized value } x' = \frac{x - \text{Minimum value of } x}{\text{Maximum value of } x - \text{Minimum value of } x} \quad (1)$$

Data splitting is a noticeable aspect of data science, especially for working with models based on the data. Data splitting means dividing dataset into two subsets and its goal is training and developing models. The first subset is for training dataset and the second one is for testing dataset. Train dataset is used to fit the ML models and test dataset is used after training is done. In this paper, we split our dataset into two subsets randomly. Train dataset includes 70% of our dataset and the rest 30% remains for testing the model.

2.4. Feature Importance. In the fields of ML and data mining, analyzing high-dimensional data is challenging for researchers and engineers. In such cases, feature selection methods provide an impressive way to tackle this problem by removing irrelevant and redundant data. By these methods, computation time is decreased while learning accuracy is increased. So, suitable feature selection can eliminate irrelevant or redundant columns from our dataset without sacrificing accuracy and creates a better understanding of the learning model or data. It also reduces the chance of overfitting which is very important for the future prediction. On the other hand, good feature selection decreases the amount of time needed to train a model. Features importance shows which features have the biggest impact on model prediction and which ones are not impressive [49]. There are some methods for choosing important features such as: filter methods, wrapper methods, and embedded methods [50, 51]. Researchers and engineers choose one of these models based on their datasets and models. In this study, feature selection was done by wrapper method (recursive feature elimination (RFE)) and boosting method (XGBoost, Ada Boost) and for each method we have recorded the scores assigned to each feature.

2.4.1. RFE Method. A good feature ranking criterion is not necessarily a good feature subset ranking criterion. RFE estimates the effect of removing one feature at a time on the objective function. They become very suboptimal when it comes to removing several features at a time, which is necessary to obtain a small feature subset [52]. Among different feature selection methods, RFE is effective in selecting features and it searches to progress generalization performance by deleting the least important features whose elimination will have the least effect on training errors [53].

2.4.2. Boosting Method. (1) XGBoost. One of the advantages of using ensembles of decision tree methods such as gradient boosting is that they can estimate feature importance automatically from training set. The more an attribute is used to make key decisions with decision trees, the higher is its relative importance. A well-known computational speed and model performance is XGBoost. This method is scalable and end-to-end tree-boosting system [54]. So, this algorithm was selected in this paper based on the mentioned properties. **(2) AdaBoost.** Boosting is a general method for improving the accuracy of any given learning algorithm. One of the most popular ensemble boosting regressors is adaptive boosting known as AdaBoost. This algorithm was introduced for the

first time in 1995 by Freund and Schapire [55, 56]. AdaBoost tries to combine multiple weak subsets of features into a strong subset to improve the performance of future models.

2.4.3. Feature Selection Based on Feature Importance. In this study, for each tested feature selection algorithm, we have recorded the scores. Based on this feature selection result, we observed that nanoparticle size, extract mass, solvent volume, extract volume, AgNO₃ volume, AgNO₃ concentration, reaction temperature, reaction time, diameters of disks and wells, incubation temperature, incubation time, bacteria concentration, and nanoparticles concentration are more crucial and effective features than others, so we created dataset III with most important features.

2.5. Model Development or Regression Models. ML is a subset of AI. ML enables the machine to learn from data, use past experiences to improve performance, and make predictions [57]. Based on data and the way of learning, there are mainly four types of learning: supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning. Supervised learning, as the name shows is an approach where algorithms are trained on labeled datasets [58]. In a labeled dataset, there are both input and output. ML algorithms train on this kind of dataset until they detect special relationship between input and output. The supervised learning itself is divided into two main groups: classification and regression. Classification is the process of predicting the class of given data points. In this method, there are some classes, and scientists try to categorize datasets based on those classes. Regression analysis is a subset of supervised learning wherein the algorithms are trained and learned with both input features and output labels. Regression models seek to estimate a mapping function from inputs to a continuous output [59]. In this study, the desired dataset is labeled and both input and output are in the dataset. There is also no intention to categorize the data and we want to predict the output. So, regression method in supervised learning is selected for this study. ML algorithms try to map functions from AgNPs features to inhibition of bacteria and predict the antibacterial activity of AgNPs. Several supervised regression algorithms are used for developing our model to find out which model can provide the most precise prediction. The decision tree, random forest, and two boosting algorithms: XGBoost and AdaBoost are examined in this study.

2.5.1. Decision Tree. It is one of the most commonly used algorithms in supervised learning and can be used for both classification and regression problems. The name of decision tree suggests that it uses a tree structure to present the predictions resulting from a series of feature-based splits. It divides the dataset into smaller subsets. In other words, it starts with a root node and finishes with a decision made by leaves. Decision tree has some advantages such as the flexibility to handle a broad range of response types, including numeric, categorical, ratings, and survival data. Besides, it is simple to understand and interpret [60].

2.5.2. Random Forest. Tree-based algorithms are attractive due to their high execution speed. However, random forest

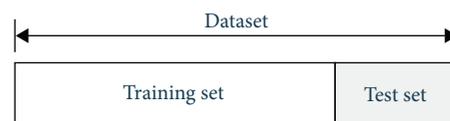


FIGURE 2: Cross-validation technique.

is an algorithm which combines the results of many different decision trees to make the best possible decisions. In this algorithm after a large number of trees are generated, they vote for the most popular class in classification problems or average in the case of regression. Like decision tree, random forest can work with both classification and regression. It also can work with thousands of variables without deletion or reduction in accuracy, while preventing overfitting [61].

2.5.3. XGBoost. eXtreme Gradient Boosting (XGBoost) is an ensemble ML algorithm for tree boosting and uses a gradient boosting framework, which is designed for speed and performance [54]. Ensemble learning consists of a collection of predictors, which are multiple models to provide better prediction accuracy [62]. The main idea of this algorithm is that it builds N regression trees one-by-one, so each subsequent tree is trained using the residuals of the previous tree [63]. Weights play a significant role in XGBoost. All independent variables have weights, these weights are then fed into the decision tree which predicts results. The weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. There are two main reasons why we use this method: execution speed and model performance [54].

2.5.4. AdaBoost. The term “Boosting” is related to algorithms, which convert weak and low-accurate learner to a strong one. The significant goal of using boosting is to decrease bias as well as variance for supervised learning [64]. The AdaBoost algorithm, short for adaptive boosting is one of them and is utilized as an ensemble method. This algorithm starts to predict on original dataset and then it gives same weight to each observation. If the first prediction, which is made by first learner, is incorrect, it dedicates the higher importance to the incorrectly predicted statement and an iterative process. It continues to add new learners until the limit is reached in the model [56]. In real-world data include some patterns of which some of them are linear and some not. Finding these kinds of relationships can be challenging. So, utilizing ensembles like AdaBoost allows us to capture many of nonlinear relationships and makes accurate predictions from them [55].

2.6. Model Validation. Algorithms mentioned in the previous section were applied for predicting antibacterial activity of AgNPs in this study. The main goal of ML is to produce an effective computational model with high accuracy. In addition, ML tries to avoid overfitting and underfitting [65]. Cross-validation is a useful technique to evaluate the effectiveness of a model, especially in cases where we need to mitigate overfitting. The basic idea for this technique is

TABLE 1: Structure of primary dataset (dataset I).

	Features	Unit	Min–max of label	Missing values (%)
Numerical	AgNPs size	nm	3.0–183.2	0
	Extract mass	g	0.05–100.0	4.7
	Solvent volume	ml	2.0–1,000.0	6
	Extract volume	ml	0.5–250.0	0
	AgNO ₃ volume	ml	1.0–1,800.0	0
	AgNO ₃ concentration	mM	0.01–20.0	0
	Reaction temperature	°C	24.0–100.0	0
	Reaction time	min	10–10,080.0	3.5
	Diameters of disks and wells	mm	4.0–10.0	0
	Incubation temperature	°C	30–37	0
	Incubation time	hr	12–48	0
	Bacteria concentration	CFU/ml	1,00,000–1,50,000,000	0
	AgNPs concentration	mg/ml	0.05–50,000.0	0
Categorical	Method	–	Well/disk	37.6

splitting the dataset into two parts, part one for training and the other one for testing model (Figure 2).

In the present study, we dedicated 70% of our dataset for training set and 30% for testing phase. We evaluated the performance of our models using different performance evaluation metrics such as mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination or R -squared (R^2).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2)$$

where y_i is the actual value, \hat{y}_i is the predicted value, and n is the number of observations/rows.

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

$$\text{MSE} = \frac{\sum (y_i - \hat{y}_i)^2}{n}, \quad (3)$$

where y_i is the actual value, \hat{y}_i is the predicted value, and n is the number of observations/rows.

RMSE is an extension of the mean squared error and it can be calculated through the $\sqrt{\text{MSE}}$. For MAE, MSE and RMSE the smaller the value, the better the predictive. The coefficient of determination or R -squared (R^2) is an important evaluation measure for regression-based algorithms and it indicates how much variation of the output is supported by the inputs and it varies between 0% and 100%. Unlike the other three metrics for R^2 metric the higher the value, the better the predictive.

3. Result

3.1. Data Preprocessing. In this study, from 100 studies researching the antibacterial activity of AgNPs, primary dataset I was selected, which consisted of 14 columns and

200 rows. But this dataset was not suitable for ML algorithms because it had some missing values. It also had some categorical features, which ML algorithms cannot operate with this kind of variables correctly. Besides, other features in this dataset were not in the same range and had different scales. So, we utilized impute method for solving missing values problem. Then one hot encoding was used for transforming categorical features to numerical. Finally, we normalized our dataset with the min–max normalization technique and created dataset II. It had 18 columns (14 features, 4 results). The features comprised of nanoparticle size (nm), extract mass (g), solvent volume (ml), extract volume (ml), AgNO₃ volume (ml), AgNO₃ concentration (mM), reaction temperature (°C), reaction time (min), diameters of disks and wells (mm), incubation temperature (°C), incubation time (hr), bacteria concentration (CFU/ml), and nanoparticles concentration ($\mu\text{g/ml}$). Table 1 introduces the structure of primary dataset. The nanoparticle size, extract volume, AgNO₃ volume, AgNO₃ concentration, reaction temperature, diameters of disks and wells, incubation temperature, incubation time, bacteria concentration, and nanoparticles concentration had no missing values. The extract mass, solvent volume, reaction time, and method had 4.7%, 6%, 3.5%, and 37.6% missing values, respectively. After preprocessing of the dataset I, there was no missing or categorical variable and all of the features were in the same range (0, 1). Finally, the second dataset (dataset II) was created with the changes that were applied to the primary dataset (dataset I).

3.2. Feature Scoring and Feature Ranking. In developing ML model, only a few features in dataset are effective for building the model, and the rest variables are either irrelevant or redundant. We analyzed the feature scores and observed that the most important features were nanoparticle size and concentration. These features were more crucial and effective features than others. All of the feature selection algorithms selected them. The extract mass, reaction time, and temperature are known as comparatively important. Figures 3–6 show the results of selecting the important

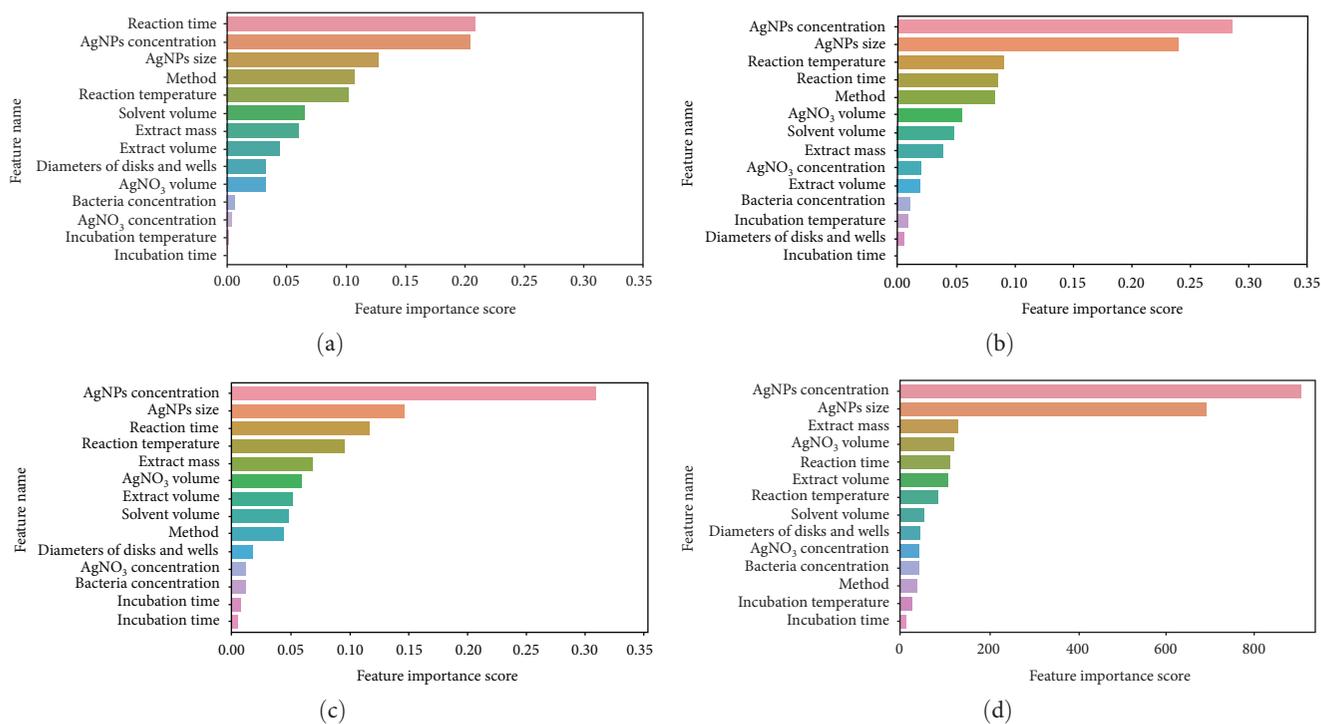


FIGURE 3: Feature selection for *Escherichia coli*: (a) AdaBoost; (b) decision tree; (c) random forest; (d) XGBoost.

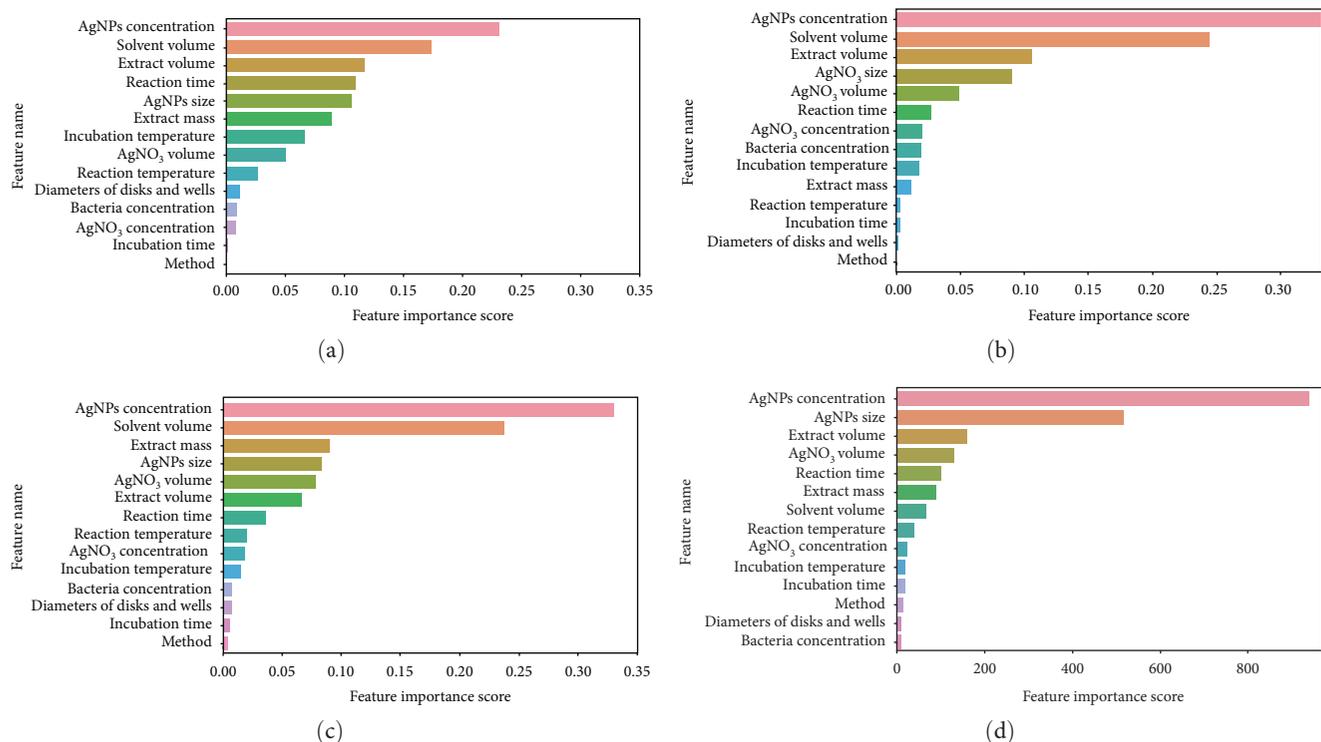


FIGURE 4: Feature selection for *Klebsiella pneumoniae*: (a) AdaBoost; (b) decision tree; (c) random forest; (d) XGBoost.

feature with four algorithms for four bacteria: *E. coli*, *P. aeruginosa*, *S. aureus*, and *K. pneumoniae*. More data analyses such as correlation values are given in Supplementary Materials (S1–S4).

3.3. Model Validation. In this study, based on important and selected feature, different ML methods were used to learn whether AgNPs have antibacterial activity or not. We have applied methods such as decision tree, random forest,

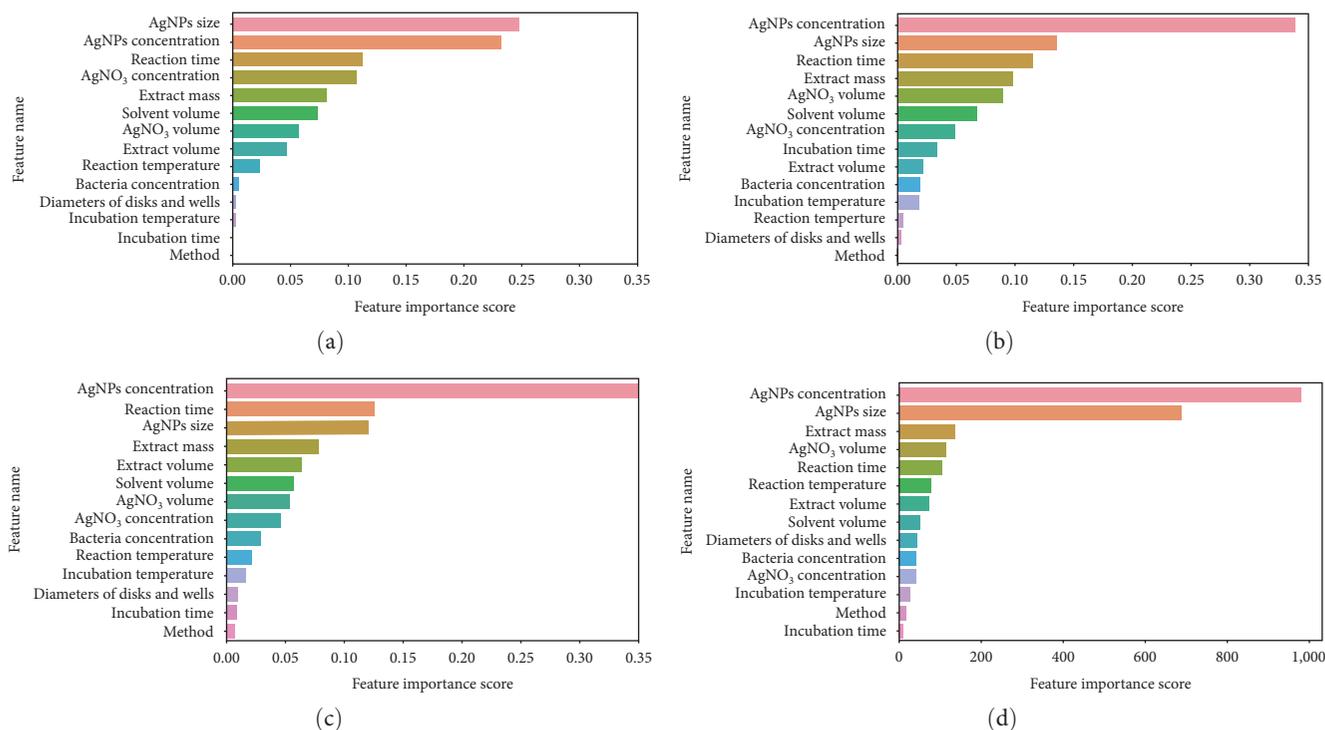


FIGURE 5: Feature selection for *Pseudomonas aeruginosa*: (a) AdaBoost; (b) decision tree; (c) random forest; (d) XGBoost.

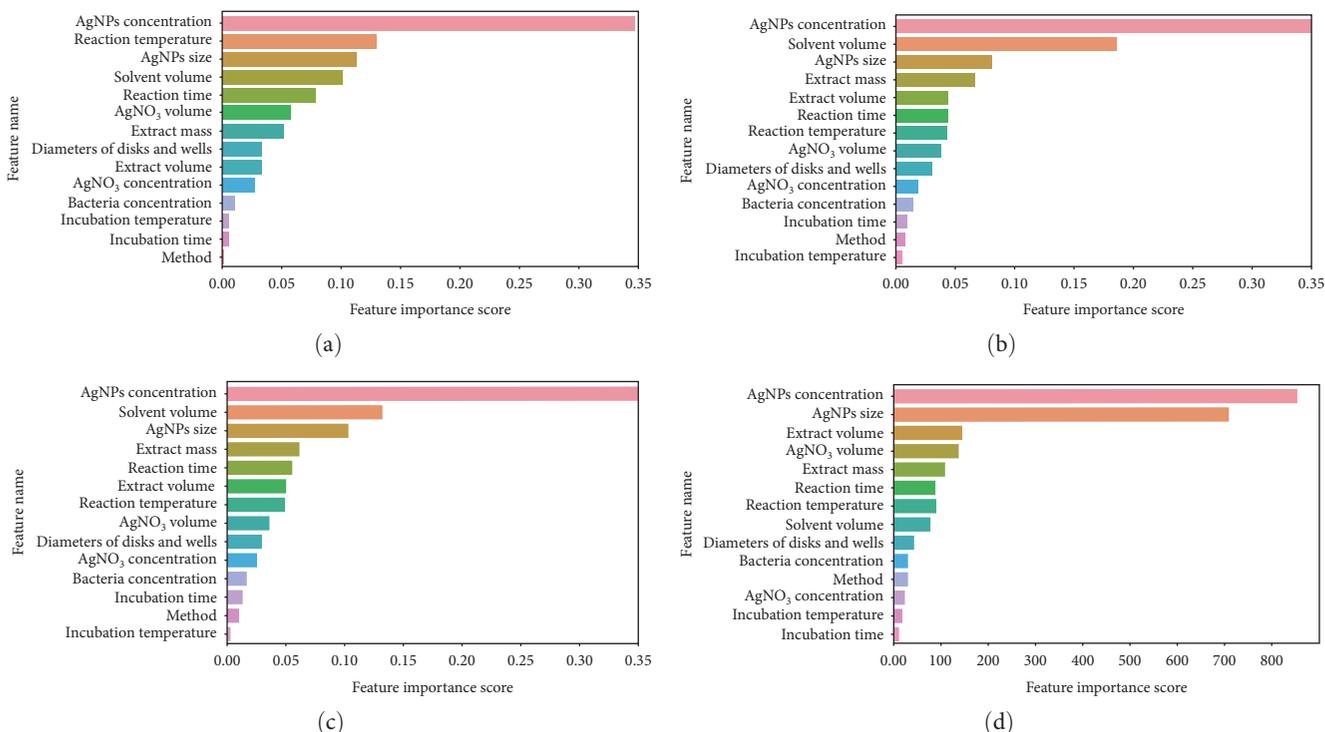


FIGURE 6: Feature selection for *Staphylococcus aureus*: (a) AdaBoost; (b) decision tree; (c) random forest; (d) XGBoost.

XGBoost, and AdaBoost (Supplementary Materials S5–S8). Finally, these models were evaluated by four evaluation metrics (MAE, MSE, RMSE, and R^2) and the results of evaluation are shown in Table 2 for four bacteria *E. coli*,

P. aeruginosa, *S. aureus*, and *K. pneumoniae*. The closer the value of the first three criteria is to zero, the more valuable it is while the higher value of R^2 is important. According to these expectations, XGBoost gives the highest R^2 (0.84) for

TABLE 2: Comparing ML algorithms according to evaluation metrics.

Bacteria	Model	MAE	MSE	RMSE	R^2
<i>Escherichia coli</i>	Decision tree	2.36	10.32	3.21	0.73
	Random forest	2.61	10.53	3.24	0.73
	XGBoost	2.42	9.51	3.08	0.77
	AdaBoost	2.51	10.40	3.22	0.70
<i>Pseudomonas aeruginosa</i>	Decision tree	1.81	5.32	2.30	0.76
	Random forest	1.89	5.80	2.40	0.70
	XGBoost	1.51	3.90	1.97	0.84
	AdaBoost	2.56	11.37	3.37	0.57
<i>Staphylococcus aureus</i>	Decision tree	2.04	7.48	2.73	0.72
	Random forest	1.84	5.82	2.41	0.73
	XGBoost	2.18	8.73	2.95	0.71
	AdaBoost	2.81	11.97	3.46	0.52
<i>Klebsiella pneumoniae</i>	Decision tree	2.67	13.61	3.69	0.73
	Random forest	1.76	4.97	2.23	0.81
	XGBoost	2.40	8.99	2.99	0.73
	AdaBoost	1.97	7.15	2.67	0.70

P. aeruginosa and the lowest error compared with other algorithms for these bacteria while AdaBoost model has the lowest R^2 (0.52) and the highest error for *S. aureus*.

4. Conclusions

In this study, we used ML algorithms to predict the antibacterial activity of AgNPs against *E. coli*, *P. aeruginosa*, *S. aureus*, and *K. pneumoniae*. This method helps in the prediction of the antibacterial activity of AgNPs by saving time and decreasing costs. For achieving this goal, first, we needed to collect data. As mentioned above, without appropriate dataset, using ML algorithms would not be possible so we prepared our dataset by reviewing already published works. In the preprocessing step, various types of processing such as one hot encoding and normalization were performed on the dataset. By feature selection methods, we can decrease computation time and eliminate irrelevant or redundant columns from our dataset without sacrificing accuracy. Thus, good feature selection decreases the amount of time needed to train a model. Features importance shows which features have the biggest impact on the model prediction and which ones are not impressive. For selecting significant features, we utilized different algorithms like decision tree, random forest, and two boosting algorithms: XGBoost and AdaBoost. The results showed that the most important features were nanoparticle size and concentration. These features were more crucial and effective features than others. The XGBoost gave the highest R^2 (0.84) for *P. aeruginosa* and AdaBoost model has the lowest R^2 (0.52) for *S. aureus*.

Data Availability

All the steps mentioned in the article, which include data preprocessing to the implementation of algorithms and

their evaluation, are available on <https://github.com/asma-dehghani/AntiBacterialLearning.git>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

Figures S1–S4 show pairplot correlation analysis of input variables with the outcome. Figures S5–S8 are the results of modeling based on the algorithms presented in this study. The x -axis shows the data points that each study provided and y -axis demonstrates the predicted and actual value. (*Supplementary Materials*)

References

- [1] E. Peterson and P. Kaur, "Antibiotic resistance mechanisms in bacteria: relationships between resistance determinants of antibiotic producers, environmental bacteria, and clinical pathogens," *Frontiers in Microbiology*, vol. 9, Article ID 2928, 2018.
- [2] F. C. Tenover, "Mechanisms of antimicrobial resistance in bacteria," *The American Journal of Medicine*, vol. 119, no. 6, pp. S3–S10, 2006.
- [3] M. R. Yeaman and N. Y. Yount, "Mechanisms of antimicrobial peptide action and resistance," *Pharmacological Reviews*, vol. 55, no. 1, pp. 27–55, 2003.
- [4] H. Nikaido, "Multidrug resistance in bacteria," *Annual Review of Biochemistry*, vol. 78, no. 1, pp. 119–146, 2009.
- [5] J. Vila, J. Moreno-Morales, and C. Ballesté-Delpierre, "Current landscape in the discovery of novel antibacterial agents," *Clinical Microbiology and Infection*, vol. 26, no. 5, pp. 596–603, 2020.
- [6] P. K. Singh, G. Jairath, and S. S. Ahlawat, "Nanotechnology: a future tool to improve quality and safety in meat industry," *Journal of Food Science and Technology*, vol. 53, pp. 1739–1749, 2016.

- [7] Md A. Subhan, K. P. Choudhury, and N. Neogi, "Advances with molecular nanomaterials in industrial manufacturing applications," *Nanomanufacturing*, vol. 1, no. 2, pp. 75–97, 2021.
- [8] S. Huang, L. Wang, L. Liu, Y. Hou, and L. Li, "Nanotechnology in agriculture, livestock, and aquaculture in China. A review," *Agronomy for Sustainable Development*, vol. 35, pp. 369–400, 2015.
- [9] R. K. Rajendranqz and C.-C. Lin, "Stability and microbial toxicity of silver nanoparticles under denitrifying conditions," *ACS Applied Materials & Interfaces*, vol. 13, no. 39, pp. 46233–46246, 2021.
- [10] C. Liao, Y. Li, and S. C. Tjong, "Bactericidal and cytotoxic properties of silver nanoparticles," *International Journal of Molecular Sciences*, vol. 20, no. 2, Article ID 449, 2019.
- [11] S. Prabhu and E. K. Poulouse, "Silver nanoparticles: mechanism of antimicrobial action, synthesis, medical applications, and toxicity effects," *International Nano Letters*, vol. 2, Article ID 32, 2012.
- [12] K. M. M. Abou El-Nour, A. Eftaiha, A. Al-Warthan, and R. A. A. Ammar, "Synthesis and applications of silver nanoparticles," *Arabian Journal of Chemistry*, vol. 3, no. 3, pp. 135–140, 2010.
- [13] I.-M. Chung, I. Park, K. Seung-Hyun, M. Thiruvengadam, and G. Rajakumar, "Plant-mediated synthesis of silver nanoparticles: their characteristic properties and therapeutic applications," *Nanoscale Research Letters*, vol. 11, Article ID 40, 2016.
- [14] S. M. Alsharif, S. S. Salem, M. A. Abdel-Rahman et al., "Multifunctional properties of spherical silver nanoparticles fabricated by different microbial taxa," *Heliyon*, vol. 6, no. 5, Article ID e03943, 2020.
- [15] M. Saravanan, S. K. Barik, D. MubarakAli, P. Prakash, and A. Pugazhendhi, "Synthesis of silver nanoparticles from *Bacillus brevis* (NCIM 2533) and their antibacterial activity against pathogenic bacteria," *Microbial Pathogenesis*, vol. 116, pp. 221–226, 2018.
- [16] M. S. Samuel, S. Jose, E. Selvarajan, T. Mathimani, and A. Pugazhendhi, "Biosynthesized silver nanoparticles using *Bacillus amyloliquefaciens*; application for cytotoxicity effect on A549 cell line and photocatalytic degradation of p-nitrophenol," *Journal of Photochemistry and Photobiology B: Biology*, vol. 202, Article ID 111642, 2020.
- [17] A. Roy, O. Bulut, S. Some, A. K. Mandal, and M. D. Yilmaz, "Green synthesis of silver nanoparticles: biomolecule-nanoparticle organizations targeting antimicrobial activity," *RSC Advances*, vol. 9, no. 5, pp. 2673–2702, 2019.
- [18] K. B. Riaz Ahmed, A. M. Nagy, R. P. Brown, Q. Zhang, S. G. Malghan, and P. L. Goering, "Silver nanoparticles: significance of physicochemical properties and assay interference on the interpretation of *in vitro* cytotoxicity studies," *Toxicology in Vitro*, vol. 38, pp. 179–192, 2017.
- [19] H. B. Fredj, S. Helali, C. Esseghaier, L. Vonna, L. Vidal, and A. Abdelghani, "Labeled magnetic nanoparticles assembly on polypyrrole film for biosensor applications," *Talanta*, vol. 75, no. 3, pp. 740–747, 2008.
- [20] Q. Zhang and Y. Yin, "Beyond spheres: Murphy's silver nanorods and nanowires," *Chemical Communications*, vol. 49, no. 3, pp. 215–217, 2013.
- [21] A. M. Abdelgawad, S. M. Hudson, and O. J. Rojas, "Antimicrobial wound dressing nanofiber mats from multicomponent (chitosan/silver-NPs/polyvinyl alcohol) systems," *Carbohydrate Polymers*, vol. 100, pp. 166–178, 2014.
- [22] Z. Zhang, J. Zhang, B. Zhang, and J. Tang, "Mussel-inspired functionalization of graphene for synthesizing Ag-polydopamine-graphenenanosheets as antibacterial materials," *Nanoscale*, vol. 5, no. 1, pp. 118–123, 2013.
- [23] J. E. Millstone, S. J. Hurst, G. S. Métraux, J. I. Cutler, and C. A. Mirkin, "Colloidal gold and silver triangular nanoprisms," *Small*, vol. 5, no. 6, pp. 646–664, 2009.
- [24] Y. Zhou, Y. Kong, S. Kundu, J. D. Cirillo, and H. Liang, "Antibacterial activities of gold and silver nanoparticles against *Escherichia coli* and bacillus Calmette-Guérin," *Journal of Nanobiotechnology*, vol. 10, Article ID 19, 2012.
- [25] S. Chernousova and M. Epple, "Silver as antibacterial agent: ion, nanoparticle, and metal," *Angewandte Chemie International Edition*, vol. 52, no. 6, pp. 1636–1653, 2013.
- [26] S. J. Soenen, P. Rivera-Gil, J.-M. Montenegro, W. J. Parak, S. C. De Smedt, and K. Braeckmans, "Cellular toxicity of inorganic nanoparticles: common aspects and guidelines for improved nanotoxicity evaluation," *Nano Today*, vol. 6, no. 5, pp. 446–465, 2011.
- [27] A. Nel, "Air pollution-related illness: effects of particles," *Science*, vol. 308, no. 5723, pp. 804–806, 2005.
- [28] A. E. Nel, L. Mädler, D. Velegol et al., "Understanding biophysicochemical interactions at the nano-bio interface," *Nature Materials*, vol. 8, pp. 543–557, 2009.
- [29] L. Lu, X. Guo, and J. Zhao, "A unified nonlocal strain gradient model for nanobeams and the importance of higher order terms," *International Journal of Engineering Science*, vol. 119, pp. 265–277, 2017.
- [30] J. Timoshenko, D. Lu, Y. Lin, and A. I. Frenkel, "Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles," *The Journal of Physical Chemistry Letters*, vol. 8, no. 20, pp. 5091–5098, 2017.
- [31] Z. Ban, P. Yuan, F. Yu, T. Peng, Q. Zhou, and X. Hu, "Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles," in *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10492–10499, 2020.
- [32] I. Fuxrhi, F. Murphy, M. Mullins, A. Arvanitis, and C. A. Poland, "Nanotoxicology data for in silico tools: a literature review," *Nanotoxicology*, vol. 14, no. 5, pp. 612–637, 2020.
- [33] I. Fuxrhi, F. Murphy, C. A. Poland, B. Sheehan, M. Mullins, and P. Mantecca, "Application of Bayesian networks in determining nanoparticle-induced cellular outcomes using transcriptomics," *Nanotoxicology*, vol. 13, no. 6, pp. 827–848, 2019.
- [34] A. Tiihonen, S. J. Cox-Vazquez, Q. Liang et al., "Predicting antimicrobial activity of conjugated oligoelectrolyte molecules via machine learning," *Journal of the American Chemical Society*, vol. 143, no. 45, pp. 18917–18931, 2021.
- [35] Ü. G. Söylemez, M. Yousef, Z. Kesmen, M. E. Büyükkiraz, and B. Bakir-Gungor, "Prediction of linear cationic antimicrobial peptides active against gram-negative and gram-positive bacteria based on machine learning models," *Applied Sciences*, vol. 12, no. 7, Article ID 3631, 2022.
- [36] H.-L. Her and Y.-W. Wu, "A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the *Escherichia coli* strains," *Bioinformatics*, vol. 34, no. 13, pp. i89–i95, 2018.
- [37] C. G. Kumar and S. K. Anand, "Significance of microbial biofilms in food industry: a review," *International Journal of Food Microbiology*, vol. 42, no. 1-2, pp. 9–27, 1998.
- [38] T. Mattila-Sandholm and G. Wirtanen, "Biofilm formation in the industry: a review," *Food Reviews International*, vol. 8, no. 4, pp. 573–603, 1992.
- [39] A. C. Lee Wong, "Biofilms in food processing environments," *Journal of Dairy Science*, vol. 81, no. 10, pp. 2765–2770, 1998.

- [40] B. Carpentier and O. Cerf, "Biofilms and their consequences with particular references to hygiene in the food industry," *Journal of Applied Microbiology*, vol. 75, pp. 499–511, 1993.
- [41] P. Rani, V. Kumar, P. P. Singh et al., "Highly stable AgNPs prepared via a novel green approach for catalytic and photocatalytic removal of biological and non-biological pollutants," *Environment International*, vol. 143, Article ID 105924, 2020.
- [42] T. C. Dakal, A. Kumar, R. S. Majumdar, and V. Yadav, "Mechanistic basis of antimicrobial actions of silver nanoparticles," *Frontiers in Microbiology*, vol. 7, Article ID 1831, 2016.
- [43] S. Gaillet and J.-M. Rouanet, "Silver nanoparticles: their potential toxic effects after oral exposure and underlying mechanisms—a review," *Food and Chemical Toxicology*, vol. 77, pp. 58–63, 2015.
- [44] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised learning," *International Journal of Computer Science*, vol. 1, pp. 111–117, 2006.
- [45] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *Journal of Big Data*, vol. 8, no. 1, pp. 1–37, 2021.
- [46] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *Journal of Big Data*, vol. 7, Article ID 28, 2020.
- [47] K. Potdar, T. S. Pardawala, and C. D. Pai, "A comparative study of categorical variable encoding techniques for neural network classifiers," *International Journal of Computer Applications*, vol. 175, no. 4, pp. 7–9, 2017.
- [48] G. Aksu, C. O. Güzeller, and M. T. Eser, "The effect of the normalization method used in different sample sizes on the success of artificial neural network model," *International Journal of Assessment Tools in Education*, vol. 6, no. 2, pp. 170–192, 2019.
- [49] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*, Springer, 2008.
- [50] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [51] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.
- [52] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 83–90, 2006.
- [53] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002.
- [54] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, 2016.
- [55] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [56] Y. Freund and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, no. 5, pp. 771–780, 1999.
- [57] T. M. Mitchell, *Machine Learning*, McGraw-Hill Science, New York, 1997.
- [58] S. J. Russell and P. Norvig, *Artificial Intelligence a Modern Approach*, Pearson Education, 2010.
- [59] L. Fahrmeir, T. Kneib, S. Lang, and B. Marx, *Regression Models, Methods and Applications*, Springer, Berlin, Heidelberg, 2021.
- [60] G. De'ath and K. E. Fabricius, "Classification and regression trees: a powerful yet simple technique for ecological data analysis," *Ecology*, vol. 81, no. 11, pp. 3178–3192, 2000.
- [61] D. Opitz and R. Maclin, "Popular ensemble methods: an empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 1999.
- [62] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [63] J. Pesantez-Narvaez, M. Guillen, and M. Alcañiz, "Predicting motor insurance claims using telematics data—XGBoost versus logistic regression," *Risks*, vol. 7, no. 2, Article ID 70, 2019.
- [64] L. B. Breiman, *Variance, and Arcing Classifiers*, University of California, Berkeley, CA, USA, 1996.
- [65] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds., pp. 532–538, Springer, Boston, MA, 2009.