

Article

# A Generalized Method for Modeling the Adsorption of Heavy Metals with Machine Learning Algorithms

Noor Hafsa <sup>1,\*</sup>, Sayeed Rushd <sup>2</sup>, Mohammed Al-Yaari <sup>2</sup>  and Muhammad Rahman <sup>3</sup> 

<sup>1</sup> Department of Computer Science, College of Computer Science and Information Technology, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia

<sup>2</sup> Department of Chemical Engineering, College of Engineering, King Faisal University, P.O. Box 380, Al-Ahsa 31982, Saudi Arabia; mrushd@kfu.edu.sa (S.R.); malyaari@kfu.edu.sa (M.A.-Y.)

<sup>3</sup> Department of Civil Engineering, College of Engineering, King Faisal University, P.O. Box 380, Al-Ahsa 31982, Saudi Arabia; mrahman@kfu.edu.sa

\* Correspondence: nhafsa@kfu.edu.sa; Tel.: +966-13895-8097

Received: 16 November 2020; Accepted: 9 December 2020; Published: 11 December 2020



**Abstract:** Applications of machine learning algorithms (MLAs) to modeling the adsorption efficiencies of different heavy metals have been limited by the adsorbate–adsorbent pair and the selection of specific MLAs. In the current study, adsorption efficiencies of fourteen heavy metal–adsorbent (HM-AD) pairs were modeled with a variety of ML models such as support vector regression with polynomial and radial basis function kernels, random forest (RF), stochastic gradient boosting, and bayesian additive regression tree (BART). The wet experiment-based actual measurements were supplemented with synthetic data samples. The first batch of dry experiments was performed to model the removal efficiency of an HM with a specific AD. The ML modeling was then implemented on the whole dataset to develop a generalized model. A ten-fold cross-validation method was used for the model selection, while the comparative performance of the MLAs was evaluated with statistical metrics comprising Spearman’s rank correlation coefficient, coefficient of determination ( $R^2$ ), mean absolute error, and root-mean-squared-error. The regression tree methods, BART, and RF demonstrated the most robust and optimum performance with  $0.96 \leq R^2 \leq 0.99$ . The current study provides a generalized methodology to implement ML in modeling the efficiency of not only a specific adsorption process but also a group of comparable processes involving multiple HM-AD pairs.

**Keywords:** artificial intelligence; regression; statistical analysis; ten-fold-cross-validation; adsorbent; removal efficiency

## 1. Introduction

Heavy metals (HMs) are stable inorganic pollutants with a low level of biodegradability [1–7] and thus tend to accumulate in living organisms [8–11]. Unlike some other pollutants, HMs can cause severe complications even at low concentrations. The US Environmental Protection Agency (EPA) listed lead (Pb), arsenic (As), nickel (Ni), chromium (Cr), copper (Cu), zinc (Zn), cadmium (Cd), and mercury (Hg) among the most serious water pollutants [12]. The permissible limits of these HMs in the industrial wastewater suggested by US EPA were 0.1, 0.01, 0.2, 0.1, 0.25, 1, 0.01, and 0.05 mg/L, respectively. The existence of such toxic metals in wastewater produced from industrial and agricultural activities can result in severe health and environmental issues due to their toxicity and environmental persistence [13]. Researchers around the globe are working on developing a feasible solution to maintain the HM concentration in natural water bodies and wastewater below the standard limits.

Various chemical and physical treatments have been evaluated to remove HMs from water. These methods include, but are not limited to, membrane separation, filtration, ion exchange,

precipitation, coagulation, reverse osmosis, and adsorption [13]. The cost and efficiency of a technique should be evaluated and judged from an engineering perspective before selecting it. The adsorption method is sometimes more preferable compared to other methods due to its many beneficial advantages, including low cost, reusability of adsorbents (ADs), environmental friendliness, and ease of operation [14]. Various ADs, including clays [15,16], zeolites [17], activated carbons [18,19], carbon nanotubes (CNTs) [20], nano-composites [21–23], graphene [24], chemical composites [25], and bio-sorbents [26–29], have been used to remove HMs from contaminated aqueous solutions. Usually, the success of any AD is mainly attributed to its morphology (porous structure), functional groups or inorganic minerals contained [30].

Extensive experimental works on removing HMs using different ADs have been reported in the literature. In general, the research scope of the previous studies was to find the maximum adsorption capacity for a single or multiple HM(s). Experimental conditions including pH, time, initial concentration, adsorbent dosage, and temperature were optimized initially. Then the adsorption process was modeled to describe its nature quantitatively. The measured values of the independent parameters were considered as the inputs (IPs) for the model, while the output (OP) was calculated based on the measurements of the initial and final concentrations of the respective HM. In most cases, the OP was the removal efficiency (%):

$$\text{Removal efficiency (\%)} = \frac{\text{Initial conc. of HM} - \text{Final conc. of HM}}{\text{Initial conc. of HM}} \times 100 \quad (1)$$

The traditional way of correlating the OP to the IPs is by identifying the most suitable adsorption isotherm, which demonstrates the adsorption capacity ( $q_e$ , mg/g) as a function of the adsorbate concentration ( $C_e$ ) in equilibrium condition.

$$q_e = \frac{(C_o - C_e)V}{m} \quad (2)$$

In Equation (2),  $C_o$  (mg/L) is the initial concentration of adsorbate,  $V$  (L) is the total volume of the fluids, and  $m$  (g) is the mass of AD. A few examples of the isotherms used in the previous studies are as follows:

$$\text{Langmuir Isotherm (LI)} : \frac{1}{q_e} = \frac{1}{q_{max}} + \frac{1}{q_{max} \cdot k_L} \cdot \frac{1}{C_e} \quad (3)$$

$$\text{Freundlich Isotherm (FI)} : \log(q_e) = \log(k_f) + \frac{1}{n} \log(C_e) \quad (4)$$

$$\text{Temkin Isotherm (TI)} : q_e = \beta_T \ln(K_T) + \beta_T \ln(C_e) \quad (5)$$

$$\text{Dubinin–Radushkevich Isotherm (DRI)} : \ln(q_e) = \ln(q_{max}) - B_D \cdot \varepsilon_D^2 \quad (6)$$

In Equations (3)–(6),  $q_{max}$  (mg/g) is the maximum adsorption capacity;  $k_L$  (L/mg) is the Langmuir constant;  $k_f$  ((mg/g)/(mg/L)<sup>n</sup>) is the Freundlich constant;  $n$  (-) represents the non-linearity of the correlation;  $K_T$  (L/mg) and  $\beta_T$  (mg/g) are the TI specific constants;  $B_D$  (mol<sup>2</sup>/kJ<sup>2</sup>) is the activity coefficient; and  $\varepsilon_D$  (kJ<sup>2</sup>/mol<sup>2</sup>) is the Polnyi potential. The standard practice of identifying the best isotherm for an adsorption process is to estimate the appropriate values of the isotherm-specific constants with a trial and error procedure. As analyzing the complex relative impacts of the IPs on the OP was found to be difficult with a traditional isotherm model, different statistical methodologies were also employed to model the adsorption processes. The most commonly used statistical tool was the response surface method (RSM). The data required to apply RSM were generated by conducting wet experiments. This kind of experiment can be considered a simple batch process of adsorption. An AD was added to the sample containing HM by adjusting all IPs. The concentration of HM in the sample was measured before and after the experiment to appraise the OP. The values of the IPs considered to significantly affect the OP for a specific HM-AD pair were varied, while the other IPs were maintained at fixed values for the experiments. Usually, a quadratic correlation (Equation (7)) of

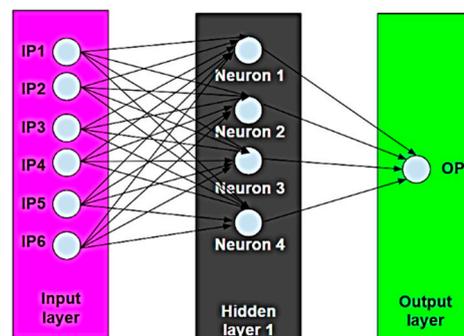
the OP to the variable IPs was developed by minimizing the difference between the predicted OP and its actual values.

$$OP = \beta_0 + \sum_{i=1}^k \beta_i IP_i + \sum_{i=1}^k \beta_{ii} IP_i^2 + \sum_{i=1}^k \sum_{j=1}^k \beta_{ij} IP_i IP_j + \varepsilon \quad (7)$$

In Equation (7),  $\beta$  and  $\varepsilon$  are the constants. An automated trial and error procedure was followed to determine the optimum values of these constants. Even though the RSM yielded acceptable predictions in most cases, it could not address the non-linearity of the correlation appropriately.

At present, artificial intelligence (AI) has been identified as a promising technique for modeling an adsorption process [16,19,22,23,25,27,29,31–34]. Compared to the traditional isotherms and statistical models, it has the advantage of directly predicting the impact of the IPs and AD-HM interaction on the adsorption process. Many AI-based machine learning algorithms (MLAs) have been employed to date [35]. The majority of these applications involved a specific algorithm, the artificial neural network (ANN). This correlates the IPs to the OP(s) using “neurons” or nodes arranged in hidden layers. As an example, a fully connected ANN architecture (6-4-1) with one input layer with six inputs, one hidden layer with four neurons, and one output layer with a single output is shown in Figure 1. Every node of each layer is connected with a weight to the nodes in the following layer. The arrangement is similar to the neurons in the animal brain. A non-linear activation function is activated for every neuron in the hidden layer to map the weighted inputs to the outputs of the neurons. The function used to predict the actual OP with an ANN can be expressed as follows:

$$\hat{y} = \sum_{i=1}^N w_i \varphi_i(x) + b_i \quad (8)$$



**Figure 1.** An example of artificial neural network (ANN) architecture (6-4-1).

In Equation (8),  $N$  is the number of neurons in the hidden layer,  $\varphi_i(x)$  is the non-linear activation function,  $w_i$  is the weighting coefficient, and  $b_i$  is the bias. Even though the non-linearity of a correlation can be addressed better by an ANN than an RSM or isotherm, its application usually suffers from several drawbacks [36]. It may experience the complication of over-fitting from a learning perspective if sufficient data are not used to train the model. Most of the previous studies on modeling the HM adsorption with ANN involved comparatively smaller datasets. It should be noted that this particular algorithm is usually applied using expensive commercial software, namely MATLAB.

Apart from ANN, other advanced MLAs, such as support vector regression (SVR), genetic algorithm (GA), genetic programming (GP), multiple linear regression (MLR), adaptive neural fuzzy interface (ANFIS), random forest (RF), stochastic gradient boosting (SGB), and Bayesian additive regression tree (BART), were also used to model various adsorption processes [35]. Instead of depending on specific commercial software, most of these algorithms can be applied using open-source statistical and data mining software, such as R. Earlier, Hafsa et al. [37] investigated the predicting performance of the non-ANN algorithms on modeling the adsorption efficiency of As in the oxidation state of As<sup>3+</sup>.

In the current study, the scope of the application is expanded further by investigating the regression performance of a set of similar models (SVR with polynomial and RBF kernels, RF, BART, and SGB) in predicting the adsorption efficiencies of five toxic metals (Pd, Hg, Cd, Cr, and As) in different oxidation states ( $\text{Pb}^{2+}$ ,  $\text{Hg}^{2+}$ ,  $\text{Cd}^{2+}$ ,  $\text{Cr}^{6+}$ , and  $\text{As}^{3+}$ ). The data required for the investigation were extracted from the literature. In addition to developing HM-AD-specific individual models, attempts were made to advance a generalized model that can predict the adsorption efficiency of multiple HM-AD combinations based on a single learning framework.

## 2. Materials and Methods

### 2.1. Regression Algorithms

A diverse variety of regression algorithms, including parametric, non-parametric, and Bayesian models were selected for the current study to model the removal efficiency of the toxic heavy metals. The list of the algorithms is as follows:

- (i). support vector regression with radial basis function (SVR-RBF) kernel
- (ii). support vector regression with polynomial (SVR-poly) kernel
- (iii). random forest (RF) regression
- (iv). stochastic gradient boosting (SGB) regression
- (v). Bayesian additive regression tree (BART)

All of these ML models are briefly discussed below, underscoring the respective statistical formulations that correlate the response with the inputs [38–45].

#### 2.1.1. SVR-RBF

The objective of SVR is to devise an as flat as possible hypothesis function,  $f(x)$ , that is insensitive to most of the  $\epsilon$ -deviations calculated from the measured responses in the training data. For a training dataset,  $(X, Y) = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ , the predicted response ( $\hat{y}$ ) or the output of  $f(x)$  can be expressed as follows:

$$\hat{y} = \langle w, \varphi(x) \rangle + b = \sum_{i=1}^N (\alpha_i k(x_i, x)) + b \quad (9)$$

where  $\langle w, \varphi(x) \rangle$  is the dot product of the weight vector ( $w$ ) and mapped feature vector with a non-linear transformation function  $\varphi(x)$ ,  $\alpha_i$  coefficients are the support vectors,  $k(x_i, x)$  is a suitable kernel function for non-linear feature mapping, and  $b$  is a constant term. The RBF kernel,  $K(x, x')$  for the feature vectors,  $(x, x')$  can be presented with the following equation:

$$K(x, x') = \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|^2\right) \quad (10)$$

where  $\sigma$  is the radius or width of the RBF kernel. It is a tuning parameter that indicates the influence of  $(x, x')$  on the model. Equation (10) shows the similitude between  $x$  and  $x'$  as a decaying function in squared form. A smaller value of the kernel indicates a higher similarity between the features. In addition to  $\sigma$ , there is another important tuning parameter,  $C$ . It is the regularization parameter of the objective loss function defined as the difference between the measured and predicted values. The ultimate goal of this ML model is to minimize the objective loss function for the training data.

#### 2.1.2. SVR-Poly

A polynomial kernel function is used in the SVR-polynomial algorithms to learn non-linear feature interactions. It compares two column vectors under the objective function framework using a degree- $d$  polynomial equation:

$$K(x, x') = (\gamma x^T x' + c)^d \quad (11)$$

where  $x$  and  $x'$  are two column vectors representing the feature vectors,  $\gamma$  is a scalar parameter,  $c$  is a constant, and  $d$  is the kernel degree. Combining Equation (11) with Equation (9), the following expression can be obtained for the response ( $\hat{y}$ ):

$$\hat{y} = \sum_{i=1}^N \alpha_i (\gamma x^T x_i + c)^d + b \quad (12)$$

The tuning parameters for SVR-Polynomial are  $d$ ,  $\gamma$ , and  $C$ .

### 2.1.3. RF Regression

RF is an ensemble learning method. It uses a multitude of regression trees during the training period. The variable splits at each node are selected based on the randomization. Using bootstrap sampling, the algorithm initiates the learning by dividing the training data into  $M$  subsets. Next, an individual regression tree ( $T_i$ ) is set for every subset by utilizing a randomly selected subset of features. This process of splitting the nodes results in a forest of  $M$  regression trees. After fitting the model to the entire training set, the response ( $\hat{y}$ ) is usually predicted for a test dataset ( $x'$ ) by averaging the individual predictions as follows:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^M T_i(x') \quad (13)$$

where  $M$  is the total number of regression trees and  $T_i(x')$  is a regression tree output. The hyper-parameter used for RF optimization is *mtry*. It designates the sum total of predictor variables that are randomly specified as nominees in the course of splitting the tree.

### 2.1.4. SGB Regression

An additive regression model is developed with an SGB algorithm by successively fitting a simple parameterized function as a weak base learner to the declivity of the difference between the measured value and model response. A random sample of the complete dataset is used to reduce the disagreement stochastically at each iteration.

In the gradient boosting algorithm, a simple regression tree can be used as a weak prediction model, and the final prediction model can then be produced in the form of an ensemble of such weak learners. The functional form of the gradient boosting-based approximation of the predicted response,  $\hat{y}$ , for each data point,  $x$  can be described as follows:

$$\hat{y} = \sum_{m=0}^M \beta_m f(x; a_m) \quad (14)$$

where the functions  $f(x; a_m)$  represent the weak learners that are simple functions of  $x$  involving weighting parameters  $a = \{a_1, a_2, \dots, a_m\}$ , and expansion coefficients as  $\beta = \{\beta_1, \beta_2, \dots, \beta_m\}$ . Both  $a$  and  $\beta$  are jointly fit to the training data. These parameters also define the split points for the base regression tree [20]. In SGB, the tuning parameters are the number of regression trees ( $m$ ) and the number of splits to be performed at each node, i.e., the maximum nodes for each tree.

### 2.1.5. BART

The BART is a Bayesian approach. It uses a sum-of-trees model to approximate the hypothesis function [30]. The BART algorithmic concept builds on enhancing the additive trees model by introducing a prior regularization term that attempts to fit the model by moderating the effect of the

individual regression tree. Consequently, each regression tree in the BART becomes a weak prediction model, explaining only a smaller portion of the training data. BART conveniently uses the additive representation of multiple regression trees to produce the final prediction model instead of constructing a single dominant large tree. The predicted response for a set of feature variables  $(x_1, x_2, \dots, x_n)$  associated with a single data point,  $x$ , could be formulated according to the sum-of-trees model, which is shown as follows:

$$\hat{y} = \sum_{j=1}^m g(x, T_j) \quad (15)$$

where  $T_j$  is a single binary regression tree and  $m$  is the total number of regression trees. Each tree  $T_j$  consists of a set of interior node decision rules and a set of prior regularized terminal nodes. For BART modeling, the tuning parameter is  $m$  or the number of trees used in the sum-of-trees model.

## 2.2. Evaluation Metrics

The evaluation metrics used to appraise the performance of the regression models consisted of Spearman's rank correlation coefficient (SPCC), the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE). The statistical parameters are briefly described as follows.

### 2.2.1. Spearman's Rank Correlation Coefficient (SPcorr)

The SRCC is a technique that summarizes the strength and the direction (positive or negative) of a relationship between actual ( $y$ ) and predicted ( $\hat{y}$ ) response variables. The formula to calculate SRCC adopts the following mathematical notation:

$$SRCC = 1 - \left( \frac{6 \sum d^2}{n^3 - n} \right) \quad (16)$$

where  $d$  = difference in the ranks between two variable sets and  $n$  = number of samples. The SRCC values range between  $-1$  and  $+1$ . The closer SRCC is to  $+1$  or  $-1$ , the stronger the probable correlation. In the case of SRCC, a perfect positive correlation is  $+1$  and a perfect negative correlation is  $-1$ .

### 2.2.2. Coefficient of Determination ( $R^2$ )

The  $R^2$  is a statistical measurement of how well the results of a regression model fit the actual measurements. It quantifies the fraction of the variation in outputs explained by the model. The equation for  $R^2$  can be described as follows:

$$R^2 = \frac{\text{Explained variations}}{\text{Total variations}} \quad (17)$$

The maximum and minimum values of  $R^2$  are 1 and 0, respectively. Generally, a higher value of  $R^2$  indicates a better model.

### 2.2.3. Mean Absolute Error (MAE)

The MAE is defined as an average of absolute differences between the model outputs and the actual measurements. The MAE can be calculated using the following formula:

$$MAE = |\hat{y} - y| \quad (18)$$

where  $\hat{y}$  and  $y$  are the predicted and actual responses respectively.

### 2.2.4. Root Mean Squared Error (RMSE)

The *RMSE* is a statistical measure of the dispersion of the prediction errors. It is a popular parameter to present the overall error of a model, as it can provide a relative perception of how concentrated the model outputs are around the best fit curve. The formula for *RMSE* can be written as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y} - y_i)^2} \quad (19)$$

where  $y$  and  $\hat{y}$  are actual and predicted responses, respectively, and  $N$  is the total number of samples.

### 2.3. Dataset

The experimental data used for the current study were collected from eight independent sources as presented in Tables 1 and 2. All of the experiments were conducted to test the efficiencies of different adsorbents in removing toxic heavy metals such as Cr, Pd, Hg, Cd, and As from polluted water. The ADs used for the experimental studies were as follows:

- AD1: Superheated steam-activated granular carbon
- AD2: Ragi husk powder (bio-sorbent)
- AD3: Antep pistachio or *Pistacia vera* L. (bio-sorbent)
- AD4: Red mud
- AD5: Synthesized functional polydopamine@Fe<sub>3</sub>O<sub>4</sub> nanocomposite (PDA@Fe<sub>3</sub>O<sub>4</sub>)
- AD6: Eucalyptus leaves (bio-sorbent)
- AD7: *Spirulina* (*Arthospira*) *maxima* (bio-sorbent)
- AD8: *Spirulina* (*Arthospira*) *indica* (bio-sorbent)
- AD9: *Spirulina* (*Arthospira*) *platensis* (bio-sorbent)
- AD10: Reduced graphene oxide-supported nanoscale zero-valent iron (nZVI/rGO) composites
- AD11: Cupric oxide nanoparticles (CuONPs) prepared with *Tamarindus indica* pulp extract
- AD12: Cerium hydroxylamine hydrochloride (Ce-HAHCl)

The following input and output parameters were analyzed in the course of the experiments:

- IP1: Operating temperature,  $T$  (°C)
- IP2: Initial  $p^H$  (-)
- IP3: Initial concentration (mg/L)
- IP4: Contact time (min)
- IP5: Adsorbent dosage (mg)
- IP6: Agitator speed (rpm)
- OP: Removal efficiency (%)

Table 1. Summary of the experimental studies.

Reference	HM	AD	Experimental Parameters				Modeling Methodology
			Variable Inputs	Fixed Inputs	Output	Data Points	
[19]	Cr(VI)	AD1	IP1 IP2 IP3 IP4 IP5	IP6		36	- RSM: $R^2 = 0.9986$ - ANN: $R^2 = 0.9911$
[27]	Cr(VI)	AD2	IP2 IP3 IP5	IP1 IP4 IP6	OP	16	- ANN: $R^2 = 0.996$ - RSM: $R^2 = 0.993$
[46]	Pb(II)	AD3	IP2 IP3 IP4	IP1 IP5 IP6		17	- RSM: $R^2 = 0.98383$
[16]	Pb(II)	AD4	IP2 IP4 IP5	IP1 IP3 IP6		15	- ANN: $R^2 = 0.898$ - RSM: $R^2 = 0.672$
[21]	Hg(II)	AD5	IP2 IP3 IP4	IP1 IP5 IP6		15	- LI: $R^2 = 0.991$ - FI: $R^2 = 0.989$ - RSM: $R^2 = 0.9871$
[28]	Hg(II)	AD6	IP2 IP3 IP4 IP5	IP1 IP6		30	- RSM: $R^2 = 0.984$ - FI: $R^2 = 0.9849$ - LI: $R^2 = 0.9802$ - DRI: $R^2 = 0.9293$ - TI: $R^2 = 0.8769$
	Cd(II)	AD7	IP2 IP3 IP5 IP6	IP1 IP4		27	- FI: $R^2 = 0.998$ - LI: $R^2 = 0.969$ - ANN: $R^2 = 0.965$ - RSM: $R^2 = 0.760$
[29]	Cd(II)	AD8				27	- FI: $R^2 = 0.994$ - ANN: $R^2 = 0.967$ - RSM: $R^2 = 0.962$ - LI: $R^2 = 0.953$
	Cd(II)	AD9			OP	27	- ANN: $R^2 = 0.9955$ - FI: $R^2 = 0.979$ - RSM: $R^2 = 0.974$ - LI: $R^2 = 0.967$
[22]	Cd(II)	AD10	IP1 IP2 IP3 IP4	IP5 IP6		29	- ANN: $R^2 = 0.9999$ - LI: $R^2 = 0.9909$ - FI: $R^2 = 0.9852$ - RSM: $R^2 = 0.9826$ - DRI: $R^2 = 0.8226$
[23]	As(III)	AD11	IP1 IP2 IP3 IP5	IP4 IP6		31	- ANN: $R^2 = 0.9994$ - LI: $R^2 = 0.997$ - FI: $R^2 = 0.805$
[25]	As(III)	AD12	IP1 IP2 IP3 IP4 IP5 IP6	-		105	- ANN: $R^2 = 0.975$

**Table 2.** Statistical presentation of the data.

Parameter (Unit)	Average	Maximum	Minimum	Standard Deviation	HM-AD
IP1 (°C)	25.0	48.8	1.2	9.4	
IP2 (-)	6.0	10.8	1.2	1.9	
IP3 (mg/L)	150.0	268.9	31.1	47.0	
IP4 (min)	50.0	73.8	26.2	9.4	Cr(VI)-AD1
IP5 (mg)	1.2	2.2	0.3	0.4	
IP6 (rpm)	150.0	150.0	150.0	0.0	
OP (%)	71.2	96.3	39.7	10.6	
IP1 (°C)	25.0	25.0	25.0	0.0	
IP2 (-)	2.0	3.0	1.0	0.8	
IP3 (mg/L)	19.3	25.0	2.0	4.9	
IP4 (min)	120.0	120.0	120.0	0.0	Cr(VI)-AD2
IP5 (mg)	3.9	60.9	1.6	10.6	
IP6 (rpm)	180.0	180.0	180.0	0.0	
OP (%)	67.0	72.7	59.2	4.0	
IP1 (°C)	30.0	30.0	30.0	0.0	
IP2 (-)	3.8	5.5	2.0	1.2	
IP3 (mg/L)	27.5	50.0	5.0	17.8	
IP4 (min)	62.5	120.0	5.0	45.5	Pb(II)-AD3
IP5 (mg)	1000.0	1000.0	1000.0	0.0	
IP6 (rpm)	250.0	250.0	250.0	0.0	
OP (%)	76.0	97.3	26.5	22.5	
IP1 (°C)	23.0	23.0	23.0	0.0	
IP2 (-)	5.0	7.0	3.0	1.5	
IP3 (mg/L)	32.1	32.1	32.1	0.0	
IP4 (min)	32.5	60.0	5.0	20.8	Pb(II)-AD4
IP5 (mg)	5.6	10.0	1.3	3.3	
IP6 (rpm)	150.0	150.0	150.0	0.0	
OP (%)	80.6	96.8	38.8	20.9	
IP1 (°C)	20.0	20.0	20.0	0.0	
IP2 (-)	4.0	7.0	1.0	2.3	
IP3 (mg/L)	60.0	100.0	20.0	30.2	
IP4 (min)	240.0	420.0	60.0	136.1	Hg(II)-AD5
IP5 (mg)	10.0	10.0	10.0	0.0	
IP6 (rpm)	400.0	400.0	400.0	0.0	
OP (%)	32.7	41.0	20.5	6.3	

Table 2. Cont.

Parameter (Unit)	Average	Maximum	Minimum	Standard Deviation	HM-AD
IP1 (°C)	25.0	25.0	25.0	0.0	
IP2 (-)	6.0	9.0	3.0	1.1	
IP3 (mg/L)	2.7	3.9	0.5	0.5	
IP4 (min)	47.5	90.0	5.0	15.8	Hg(II)-AD6
IP5 (mg)	1.5	2.5	0.5	0.3	
IP6 (rpm)	120.0	120.0	120.0	0.0	
OP (%)	92.6	94.7	78.5	4.2	
IP1 (°C)	25.0	25.0	25.0	0.0	
IP2 (-)	7.0	8.0	6.0	0.7	
IP3 (mg/L)	0.0	0.0	0.0	0.0	
IP4 (min)	6.0	6.0	6.0	0.0	Cd(II)-AD7
IP5 (mg)	0.2	0.2	0.1	0.0	
IP6 (rpm)	14.0	16.0	12.0	1.4	
OP (%)	62.3	73.3	56.6	3.8	
IP1 (°C)	25.0	25.0	25.0	0.0	
IP2 (-)	7.0	8.0	6.0	0.7	
IP3 (mg/L)	0.0	0.0	0.0	0.0	
IP4 (min)	6.0	6.0	6.0	0.0	Cd(II)-AD8
IP5 (mg)	0.2	0.2	0.1	0.0	
IP6 (rpm)	14.0	16.0	12.0	1.4	
OP (%)	66.2	79.2	58.2	5.7	
IP1 (°C)	25.0	25.0	25.0	0.0	
IP2 (-)	7.0	8.0	6.0	0.7	
IP3 (mg/L)	0.0	0.0	0.0	0.0	
IP4 (min)	6.0	6.0	6.0	0.0	Cd(II)-AD9
IP5 (mg)	0.2	0.2	0.1	0.0	
IP6 (rpm)	14.0	16.0	12.0	1.4	
OP (%)	69.9	82.5	61.8	5.6	
IP1 (°C)	30.0	40.0	20.0	6.5	
IP2 (-)	6.0	7.0	5.0	0.7	
IP3 (mg/L)	30.0	40.0	20.0	6.5	
IP4 (min)	20.0	30.0	10.0	6.5	Cd(II)-AD10
IP5 (mg)	30.0	30.0	30.0	0.0	
IP6 (rpm)	200.0	200.0	200.0	0.0	
OP (%)	60.1	77.3	44.3	8.7	

Table 2. Cont.

Parameter (Unit)	Average	Maximum	Minimum	Standard Deviation	HM-AD
IP1 (°C)	40.0	60.0	20.0	8.9	
IP2 (-)	7.0	12.0	2.0	2.2	
IP3 (mg/L)	1000.0	1900.0	100.0	402.5	
IP4 (min)	270.0	270.0	270.0	0.0	As(III)-AD11
IP5 (mg)	75.0	135.0	15.0	26.8	
IP6 (rpm)	100.0	100.0	100.0	0.0	
OP (%)	76.2	92.7	48.2	12.3	
IP1 (°C)	38.5	60.0	20.0	16.3	
IP2 (-)	7.5	10.0	4.0	2.4	
IP3 (mg/L)	23.2	50.0	10.0	15.7	
IP4 (min)	62.3	90.0	30.0	23.4	As(III)-AD12
IP5 (mg)	7733.3	10,000.0	6000.0	1761.0	
IP6 (rpm)	162.1	180.0	120.0	23.8	
OP (%)	76.6	98.9	50.0	13.9	
Overall statistics					
IP1 (°C)	30.0	60.0	1.2	11.9	Cr(VI)-AD1
IP2 (-)	6.0	12.0	1.0	2.3	Cr(VI)-AD2
IP3 (mg/L)	102.6	1900.0	0.0	261.1	Pb(II)-AD3
IP4 (min)	78.7	420.0	5.0	78.9	Pb(II)-AD4
IP5 (mg)	1737.0	10,000.0	0.0	3281.4	Hg(II)-AD5
IP6 (rpm)	178.7	800.0	12.0	178.1	Hg(II)-AD6
OP (%)	68.1	98.9	0.9	21.3	Cd(II)-AD7
					Cd(II)-AD8
					Cd(II)-AD9
					Cd(II)-AD10
					As(II)-AD11
					As(II)-AD12

The IPs were measured, while the OP was calculated based on the measured values of initial and final concentrations of the respective HM.

#### 2.4. MLA Modeling

As mentioned earlier, the predictive power of MLA was assessed in the current study by training and testing five regression models to estimate the removal efficiency of five heavy metals using six experimental parameters. The MLA modeling involved interpolating the experimental parameters to produce synthetic data and developing models for the datasets individually as well as comprehensively. For this purpose, the following steps were executed in successive order: (i) interpolation of the experimental data; (ii) parameter optimization and model selection for individual heavy metals; (iii) parameter optimization and model selection for the comprehensive dataset.

##### 2.4.1. Data Interpolation

As can be observed from Table 1, the ten datasets used in the present study consist of a relatively lower number of actual measurements. Therefore, the training set composed of actual measurements is not necessarily large enough, as the MLAs are data-driven and demand a reasonable quantity of

data for optimizing the parameters and training the models reliably. To resolve this issue, a data augmentation technique is necessary for increasing the number of data points in the training set. Earlier, Podder et al. [47] used a cubic spline function for generating interpolated data points for their ANN-based modeling of As adsorption efficiency. Cubic spline or piecewise cubic interpolation can be categorized as an exact point interpolation method in the family of spatial interpolation techniques. This piecewise polynomial interpolation method, unlike its polynomial analogs, is capable of finding a continuous second derivative at all data points by minimizing the interpolation errors and produces a smoother distribution of interpolated data points within a certain range [48,49]. The cubic spline interpolation also mitigates the distortion issues in boundary regions observed in least-squares interpolation [47]. Considering the advantages, a piecewise cubic interpolation method was adopted to interpolate the original data points in the current study. The process of interpolation was initiated by interpolating the output column based on the first predictor. It was then extended to the predictor columns using the previously interpolated output values. On average, 250 data points were interpolated for each data set. All interpolations were performed using the ‘spline’ function of the ‘stats’ package in R [50].

#### 2.4.2. Parameter Optimization and Model Selection

##### Individual Metal

The ML modeling to predict the removal efficiency involved associated parameter optimization on a training set and the final model selection based on the validation using the independent test data points. As a first step, the original experimental datasets extracted from the literature were interpolated using a natural cubic spline technique and merged. Next, the merged dataset was split into a training (80%) and a test (20%) subset using random sampling. The training of five ML models was carried out using the training data points (80%). The test dataset (20%), which can be considered as an independent test set, was withheld for model verification. The optimized parameters for the ML models and, also, the best performing model were selected based on a repeated k-fold cross-validation (CV) technique integrated with a grid search [51]. The hyper-parameters of the ML models, as presented in Table 3, were optimized by minimizing the average prediction error in training data.

**Table 3.** The machine learning (ML) models, hyperparameter names, and corresponding optimized values after cross-validated training (the last column includes R packages used for different ML models).

Model	Hyperparameter Names	R Package
Random Forest	[mtry]	randomForest
SVR–RBF Kernel	[sigma, C]	kernlab
SVR–Polynomial Kernel	[degree, scale, C]	kernlab
Stochastic Gradient Boosting	[n.trees, interaction.depth]	gbm
Bayesian Additive Regression	[num_trees]	bartMachine

At every fold of the CV, (k-1) subsets of the data were used to train a model. The hold-out dataset was then used to validate the trained model by calculating the *RMSE*. This process was repeated *k*-times with the desired number of iterations to identify the best performing model with the optimal values of the associated parameters that minimized the average *RMSE* across the repeated CV. From a mathematical perspective, the following optimization problem was attempted to be minimized:

$$\min_{params} \left( \sum_{i=1}^n \sum_{j=1}^k RMSE_{i,j} \right) \quad (20)$$

where *n* indicates the number of iterations, *k* is the number of folds in CV, and *RMSE<sub>i,j</sub>* is the root mean square deviation of the predicted response from the actual response for the *j*-th fold in the *i*-th iteration

of the repeated CV. The algorithm used for model selection and parameter optimization is presented in Figure 2. Both the number of folds ( $k$ ) used in the CV and the repetition times ( $N$ ) were considered as 10 in the present study.

<b>Pseudocode: Model Selection Using Repeated 10-Fold Cross-Validation</b>	
Input	Learning model and Training set divided into equal-sized folds $F_i$
Output	Final best model with optimal parameters
1.	Repeat
2.	Divide the training dataset into 10 folds
3.	for each fold $F_i$
4.	Select the $F_i$ as the validation dataset
5.	Merge the remaining folds to use as the training dataset
6.	Tune the model parameters on the training dataset
7.	Use the model to predict on the validation dataset
8.	Calculate the prediction error (e.g., RMSE)
9.	End
10.	Calculate the average prediction error over 10 folds
11.	until $N = 10$
12.	Calculate the average prediction error over 10 repetitions
13.	Select the model with the lowest average error across the repeated CV
14.	Return the final chosen model with corresponding optimal parameters

**Figure 2.** A pseudocode representation of the repeated 10-fold cross-validation (CV) process used for parameter optimization and model selection of ML algorithms.

### Comprehensive Dataset

We performed MLA modeling for a comprehensive dataset that includes the data related to all metals. The objective was to produce a generalized model for the five toxic metals (As, Cd, Cr, Hg, and Pb) and 10 adsorbents (see Tables 1 and 2) used in this study. The dataset contained all the merged (original and interpolated) data points associated with five toxic metals. The statistics for this comprehensive dataset are presented in Table 4. There were two modifications in the feature description in this combined dataset. As shown in Table 1, six experimental parameters were available for any dataset including the variable and fixed inputs such as operating temperature, pH, initial concentration, contact time, sorbent dosage, and agitator speed. In the combined dataset, all six of these experimental parameters were included in the feature description. In addition to that, the types of metal and the specific adsorbent used in the experiment were also incorporated as features. Therefore, a total of eight (8) feature variables, as opposed to six (6) parameters in individual modeling, were used for the MLA modeling of removal efficiency on a total of 2476 (80%) training data points in the combined dataset. All five MLAs previously used for individual metal removal efficiency modeling were utilized to develop these generalized models. A similar strategy of repeated 10-fold cross-validation was performed to optimize the different parameters of the MLA models on this combined dataset. Finally, the parameter optimization and model selection steps gave us five optimal models (from using five different ML algorithms) chosen based on the lowest average RMSE observed on the validation data points during the cross-validated training. These models were then used to predict the removal efficiency on 619 (20%) independent test data points from the combined dataset.

**Table 4.** Data set statistics (Combined dataset).

Combined Dataset (Five Metals)	Percentage	No. Data Points
Training	80%	2476
Test	20%	619
Total	100%	3095

### 2.5. Computing Framework

A computer with the configuration of Intel CORE i7 8th Gen, 2GHz processor with 8 GB of RAM was used to perform the dry experiments and statistical analysis. The hardware was operated with a 64-bit Windows 10 operating system. The open-source statistical computing framework, R (4.0.2) was used for conducting all ML experiments and performing the statistical analysis. We chose R because it can be generally used in any platform. It can also provide all necessary packages and library functions for ML model training, hyper-parameters tuning, visualizing and data preprocessing, performing post-prediction related statistical analyses, and plotting graphs and trends of the results. R is considered a standard industrial choice for exploratory data analysis. The time requirement of prediction for a typical data matrix of  $200 \times 6$  dimensions was approximately 10 s.

## 3. Results

### 3.1. ML Model Evaluation for Individual Dataset

The performances of five MLAs in predicting the efficiencies of absorbing five toxic metals (As, Cd, Cr, Hg, and Pb) by different adsorbents are shown in Tables 5–9. For each metal, the prediction results are reported for two separate adsorbents. The outcomes are evaluated with a statistical metric comprising MAE, RMSE, SPcorr, and  $R^2$ .

**Table 5.** The performances of five (5) machine learning algorithms (MLA) models on independent test data (20%) of As (III) datasets.

Metal	Algorithm	Performance			
		MAE	RMSE	SPcorr	$R^2$
As (III) 1	SVR-Poly	2.42	5.43	0.91	0.84
	Stochastic Gradient Boosting	1.51	3.13	0.97	0.93
	SVR-RBF	2.41	5.30	0.92	0.84
	Random Forest	1.36	3.53	0.96	0.93
	<b>Bayesian Additive Regression Tree</b>	<b>1.33</b>	<b>4.18</b>	<b>0.98</b>	<b>0.97</b>
As (III) 2	SVR-Poly	3.32	6.08	0.89	0.80
	<b>Stochastic Gradient Boosting</b>	<b>2.71</b>	<b>5.67</b>	<b>0.90</b>	<b>0.81</b>
	SVR-RBF	3.38	5.89	0.89	0.80
	Random Forest	2.72	5.92	0.89	0.80
	Bayesian Additive Regression Tree	2.57	5.83	0.89	0.79

**Table 6.** The performances of five (5) MLA models on independent test data (20%) of Cr(IV) datasets.

Metal	Algorithm	Performance			
		MAE	RMSE	SPcorr	R <sup>2</sup>
Cr(IV) 1	SVR-Poly	0.38	1.08	0.94	0.89
	Stochastic Gradient Boosting	1.51	3.13	0.97	0.93
	SVR-RBF	0.49	1.14	0.94	0.89
	Random Forest	1.36	3.53	0.96	0.93
	<b>Bayesian Additive Regression Tree</b>	<b>0.10</b>	<b>0.15</b>	<b>0.99</b>	<b>0.99</b>
Cr (IV) 2	SVR-Poly	2.16	3.84	0.97	0.95
	Stochastic Gradient Boosting	2.04	4.80	0.96	0.92
	<b>SVR-RBF</b>	<b>1.62</b>	<b>3.04</b>	<b>0.98</b>	<b>0.96</b>
	Random Forest	1.60	4.65	0.96	0.92
	Bayesian Additive Regression Tree	1.21	4.0	0.97	0.94

**Table 7.** The performances of five (5) MLA models on independent test data (20%) of Cd(II) datasets.

Metal	Algorithm	Performance			
		MAE	RMSE	SPcorr	R <sup>2</sup>
Cd (II) 1	SVR-Poly	1.06	1.77	0.97	0.95
	Stochastic Gradient Boosting	0.58	1.32	0.98	0.97
	SVR-RBF	0.95	1.39	0.98	0.97
	Random Forest	0.66	2.00	0.96	0.92
	<b>Bayesian Additive Regression Tree</b>	<b>0.65</b>	<b>1.60</b>	<b>0.99</b>	<b>0.98</b>
Cd (II) 2	SVR-Poly	2.44	5.42	0.96	0.92
	Stochastic Gradient Boosting	2.05	5.07	0.96	0.93
	SVR-RBF	2.0	3.59	0.98	0.97
	Random Forest	1.63	5.18	0.96	0.92
	<b>Bayesian Additive Regression Tree</b>	<b>1.16</b>	<b>3.22</b>	<b>0.98</b>	<b>0.97</b>

**Table 8.** The performances of five (5) MLA models on independent test data (20%) of Hg(II) datasets.

Metal	Algorithm	Performance			
		MAE	RMSE	SPcorr	R <sup>2</sup>
Hg (II) 1	SVR-Poly	0.54	0.95	0.97	0.95
	Stochastic Gradient Boosting	0.29	0.61	0.99	0.98
	SVR-RBF	0.42	0.90	0.98	0.96
	<b>Random Forest</b>	<b>0.11</b>	<b>0.38</b>	<b>0.99</b>	<b>0.99</b>
	Bayesian Additive Regression Tree	0.24	0.78	0.99	0.98
Hg (II) 2	SVR-Poly	0.61	1.67	0.94	0.88
	Stochastic Gradient Boosting	0.26	0.75	0.98	0.97
	SVR-RBF	1.13	1.99	0.95	0.91
	Random Forest	0.23	0.85	0.95	0.97
	<b>Bayesian Additive Regression Tree</b>	<b>0.14</b>	<b>0.30</b>	<b>0.99</b>	<b>0.99</b>

**Table 9.** The performances of five (5) MLA models on independent test data (20%) of Pb(II) datasets.

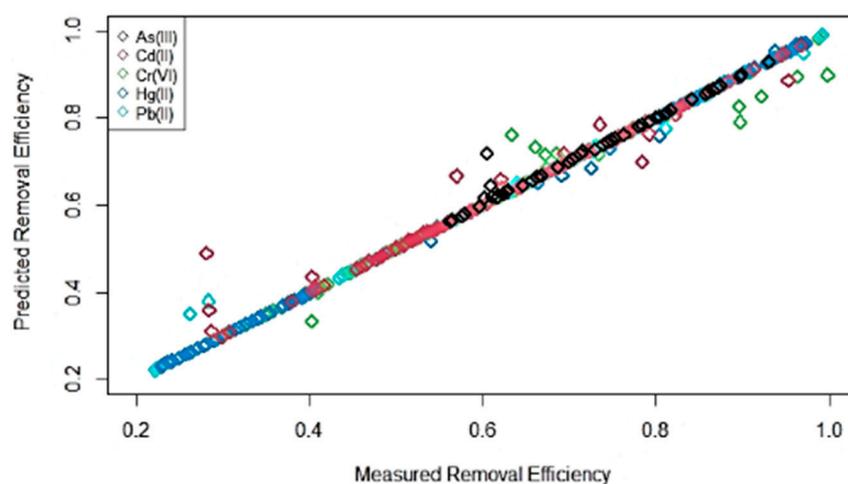
Metal	Algorithm	Performance			
		MAE	RMSE	SPcorr	R <sup>2</sup>
Pb (II) 1	SVR-Poly	2.29	3.47	0.98	0.97
	Stochastic Gradient Boosting	1.46	1.37	0.98	0.96
	SVR-RBF	1.96	3.59	0.98	0.97
	Random Forest	0.92	3.14	0.98	0.96
	<b>Bayesian Additive Regression Tree</b>	<b>0.61</b>	<b>1.37</b>	<b>0.99</b>	<b>0.99</b>
Pb (II) 2	<b>SVR-Poly</b>	<b>1.13</b>	<b>1.99</b>	<b>1.0</b>	<b>1.0</b>
	Stochastic Gradient Boosting	0.90	2.21	0.99	0.99
	SVR-RBF	2.29	3.47	1.0	1.0
	Random Forest	0.18	0.42	0.99	0.99
	Bayesian Additive Regression Tree	0.69	2.78	0.99	0.99

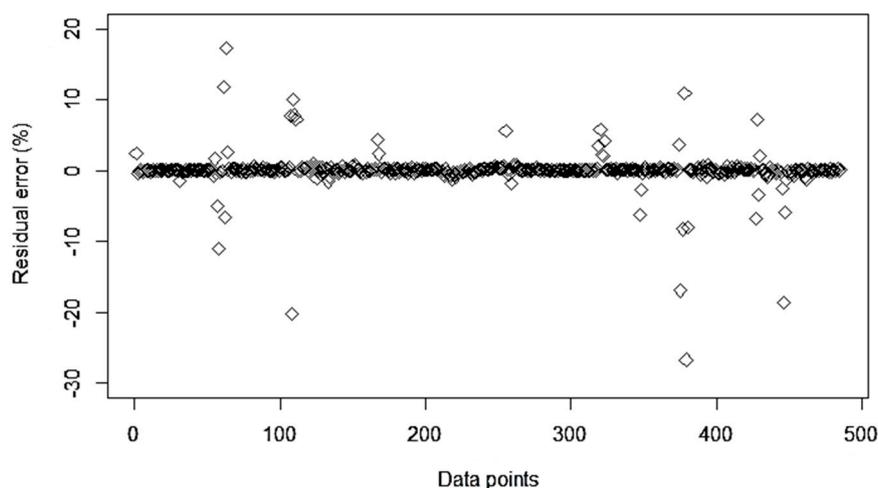
### 3.2. ML Model Evaluation for Combined Dataset

The results of the performance evaluations of the five generalized ML models on the independent test dataset are described in Table 10. The performance of the best model (RF) is presented in Figure 3, with a graph depicting the predicted values of the removal efficiencies as the function of the measured values for different HMs. In addition, the residual percentile error plot for the RF model is depicted in Figure 4 for independent test data.

**Table 10.** Training and independent test performances of five generalized models.

Model	Train				Test			
	MAE	RMSE	SPCC	R <sup>2</sup>	MAE	RMSE	SPCC	R <sup>2</sup>
SVR-Poly	0.0276	0.046	0.977	0.976	0.0278	0.052	0.972	0.970
SGB	0.0247	0.043	0.981	0.979	0.249	0.047	0.979	0.976
SVR-RBF	0.0267	0.043	0.981	0.978	0.0273	0.050	0.976	0.973
<b>RF</b>	<b>0.004</b>	<b>0.015</b>	<b>0.997</b>	<b>0.997</b>	<b>0.007</b>	<b>0.033</b>	<b>0.989</b>	<b>0.988</b>
BART	0.023	0.048	0.990	0.974	0.025	0.054	0.983	0.969

**Figure 3.** The predictions of removal efficiencies for the independent test data using the generalized RF model for different metals.



**Figure 4.** The percentage residual error plot for the generalized RF model estimated removal efficiency for the independent test data.

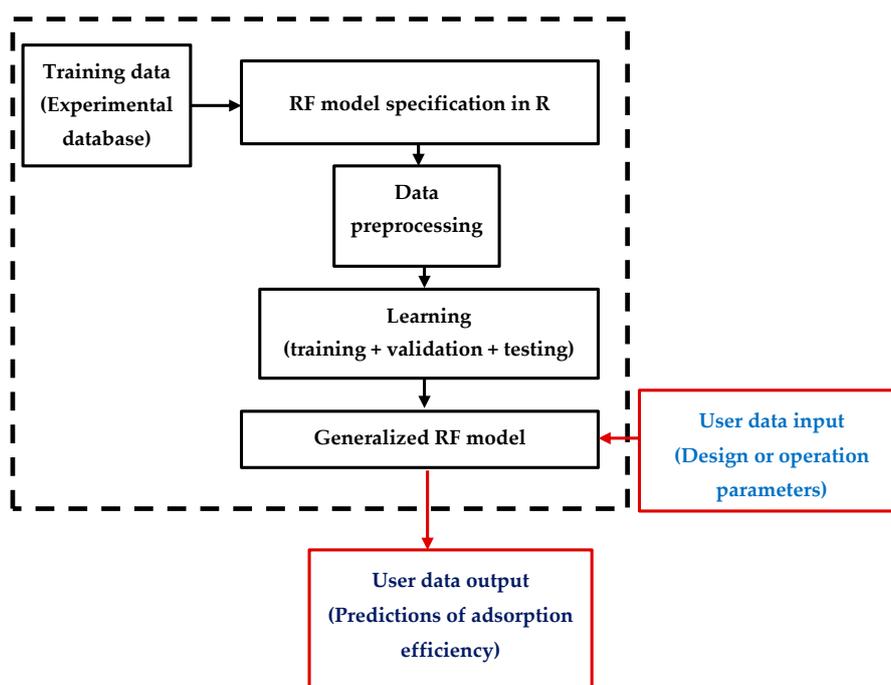
#### 4. Discussion

The current study presents a comprehensive approach to modeling adsorption efficiency. A wide range of ML models was applied to model the experimental adsorption of five toxic heavy metals with ten different adsorbents. As the modeling of an adsorption process involves non-linear feature interactions, the utility of the non-linear parametric regression models, such as SVR with polynomial and RBF kernels, RF, and SGB, including a Bayesian regression approach called BART, were examined in the current study. The RF and SGB were selected as the bagging and boosting algorithms, respectively. Both RBF and polynomial kernels in the SVR algorithm perform mapping of the input space to higher dimensional feature space, and, subsequently, the data points become linearly separable into that higher feature dimension. Similarly, three different variations of regression trees used in the current study are suitable for non-linear regression tasks. For each toxic metal, two datasets using two different adsorbents were considered, resulting in a total of 10 datasets for the ML experiments. Note that each of these datasets consists of both original and interpolated data points, which were split into an 20 to 80% ratio of training and test data, respectively. Tables 5–9 report the results of ML modeling of the selected regression models on 20% independent test data points for each of these 10 datasets. Interestingly, a single learning algorithm did not stand alone for all ten datasets when evaluated with the independent test points (see Tables 5–9). However, the BART algorithm showed the optimum performance compared to other models for all data. The average  $R^2$  value was 96%. The other two regression tree approaches, SGB and RF, demonstrated the next best performances with average  $R^2$  values of 94% and 93%, respectively. In the case of SVR, the models with the RBF kernel demonstrated slightly better performance ( $R^2 = 93\%$ ) than its polynomial counterpart ( $R^2 = 91\%$ ). However, an extensive comparative analysis (e.g., finding min, max, and standard deviation) of the performance of these 10 individual models may not be appropriate here, as the 10 datasets used were collected under different experimental setups using 12 different adsorbents and five different metals.

Since a generalized ML model applicable to different adsorption processes does not exist in the literature, we performed the modeling based on the strategy that combines diverse datasets in a single learning framework to which different ML algorithms can be applied. This effort provided insights about the generalized predictive power of the ML algorithms for estimating adsorption efficiency irrespective of the HM-AD combinations and the reliability of the prediction made by the generalized models in the case of different toxic metals. It also made the comparative analysis of the performances of ML algorithms more meaningful as all variations in the experimental setup, metal, and adsorbent types were brought under a single learning framework of model development using a specific algorithm and all five algorithms underwent the training on the same set of data points.

The evaluation of the generalized models, as presented in Table 10, shows that all of those demonstrate consistent and comparable performances for training and test datasets. The SVR-polynomial kernel performs almost identically to its RBF kernel counterpart. Among these methods, the RF model yielded the best scores in terms of all evaluation metrics (SPCC = 0.989,  $R^2 = 0.988$ , MAE = 0.007, and RMSE = 0.033). It is important to observe that both bagging- (RF) and boosting (SGB)-based regression tree algorithms with stochastic components were found to perform better by choosing the best possible random set of predictors (RF) or observations (SGB) for splitting at each node of the regression tree and several iterations for parameter optimization. Both regression tree models were able to capture the non-linearity of the data accurately in estimating the response variable. The BART was able to achieve one of the best correlations (SPCC = 0.983 and  $R^2 = 0.969$ ) by imposing regularization on each tree while fitting to a small portion of the training data, leading to a bias-free prediction when several trees were fitted to the complete set of training samples. The measured removal efficiencies for the test dataset are shown against the predicted values by the best performing RF model in Figure 3. Compared to the metal-specific predictions shown in Figure 4, the RF model is evidently accurate in predicting the removal efficiencies for all different types of metals, irrespective of the adsorbent type used for the adsorption experiments. The residual error analysis of the RF model is presented in Figure 4 with the range of errors in the percentile level. More than 98% of test data lie within a  $\pm 10\%$  error limit.

A methodology to implement the best performing RF model is outlined in Figure 5, with a block diagram. The model in its current form is directly applicable to predict the adsorption efficiency for a given set of process conditions. It requires only the design or operating parameters (IPs) as inputs from the user. These input parameters are to be treated as the predictor variables to provide the output of adsorption efficiency. In the case of using the current database, the predictions would be limited to the twelve HM-AD pairs used for this study. However, the database can be enriched further by adding new experimental measurements for different HM-AD pairs. That will help to extend the predicting scope of the current model. The AI-based automated methodology is expected to replace the traditional modeling approach that requires indefinite iterations to figure out the appropriate model with the optimized values of the coefficients. It will be significantly beneficial for the general users, including design and operating engineers, as well as management and research personnel.



**Figure 5.** Presentation of the automated prediction methodology by using the current ML model.

## 5. Conclusions

State-of-the-art ML algorithms were applied to model the sorption efficiencies of different adsorbents in removing toxic heavy metals (As, Cd, Hg, Cr, and Pb) for the current study. Specifically, the predictive power of the non-ANN approach was analyzed in an open-source statistical computing framework, R. Probably the most significant contribution of the current study is a generalized ML model that can be used to predict the removal efficiency of five toxic metals using twelve different adsorbents. All ML models were developed using original and synthetic data produced by interpolating the original data using a cubic spline interpolation algorithm. Model assessments using standard evaluation metrics show an excellent agreement between the actual and predicted removal efficiency for both individual ( $R^2 = 96\%$ ) and generalized ( $R^2 = 98.8\%$ ) predictive models. The present work provides important insights about the predictive power of non-ANN ML approaches for both metal- and adsorbent-specific individual learning models, and the models in which all data are combined into a single learning framework. With the superior performances and beneficial attributes of the generalized models, the proposed system has a high potential to be employed and used for the industrial production system. Although the current approach was successfully tested for a set of adsorption systems comprising five toxic heavy metals and twelve varieties of adsorbents, it should be implemented further to a larger dataset to develop a universal model.

**Author Contributions:** Conceptualization, N.H. and S.R.; methodology, N.H.; software, N.H.; validation, N.H. and S.R.; formal analysis, N.H. and S.R.; investigation, N.H. and S.R.; resources, N.H. and S.R.; data curation, S.R.; writing—original draft preparation, N.H., S.R., and M.A.-Y.; writing—review and editing, N.H., S.R., M.A.-Y., and M.R.; visualization, N.H. and S.R.; supervision, N.H. and S.R.; project administration, N.H. and S.R.; funding acquisition, N.H., S.R., M.A.-Y., and M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number IFT20063. The authors also acknowledge the Deanship of Scientific Research at King Faisal University for their kind assistance.

**Acknowledgments:** The authors extend their appreciation to the Deanship of Scientific Research, College of Computer Science and Information Technology, and College of Engineering at King Faisal University, Saudi Arabia.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Hegazi, H.A. Removal of heavy metals from wastewater using agricultural and industrial wastes as adsorbents. *HBRC J.* **2013**, *9*, 276–282. [[CrossRef](#)]
2. Gupta, V.K.; Gupta, M.; Sharma, S. Process development for the removal of lead and chromium from aqueous solutions using red mud—An aluminium industry waste. *Water Res.* **2001**, *35*, 1125–1134. [[CrossRef](#)]
3. Kumar, P.S.; Saravanan, A. Sustainable wastewater treatments in textile sector. In *Sustainable Fibres and Textiles*; Muthu, S.S., Ed.; Woodhead Publishing: Cambridge, UK, 2017; pp. 323–346. [[CrossRef](#)]
4. Peng, B.; Fang, S.; Tang, L.; Ouyang, X.; Zeng, G. Nanohybrid Materials Based Biosensors for Heavy Metal Detection. In *Micro and Nano Technologies, Nanohybrid and Nanoporous Materials for Aquatic Pollution Control*; Tang, L., Deng, Y., Wang, J., Wang, J., Zeng, G., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 233–264. [[CrossRef](#)]
5. Tasharrofi, S.; Hassani, S.S.; Taghdisian, H.; Sobat, Z. Environmentally friendly stabilized nZVI-composite for removal of heavy metals. In *New Polymer Nanocomposites for Environmental Remediation*; Hussain, C.M., Mishra, A.K., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 623–642. [[CrossRef](#)]
6. Rhouati, A.; Marty, J.L.; Vasilescu, A. Metal Nanomaterial-Assisted Aptasensors for Emerging Pollutants Detection. In *Advanced Nanomaterials*; Nikolelis, D.P., Nikoleli, G.P., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 193–231. [[CrossRef](#)]
7. Atieh, M.A.; Ji, Y.; Kochkodan, V. Metals in the Environment: Toxic Metals Removal. *Bioinorg. Chem. Appl.* **2017**, *2017*, 4309198. [[CrossRef](#)] [[PubMed](#)]

8. Jin, L.; Zhang, G.; Tian, H. Current state of sewage treatment in China. *Water Res.* **2014**, *66*, 85–98. [[CrossRef](#)] [[PubMed](#)]
9. Lau, Y.J.; Khan, F.S.A.; Mubarak, N.M.; Lau, S.Y.; Chua, H.B.; Khalid, M.; Abdullah, E.C. Functionalized carbon nanomaterials for wastewater treatment. In *Micro and Nano Technologies, Industrial Applications of Nanomaterials*; Thomas, S., Grohens, Y., Pottathara, Y.B., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 283–311. [[CrossRef](#)]
10. Järup, L. Hazards of heavy metal contamination. *Br. Med. Bull.* **2003**, *68*, 167–182. [[CrossRef](#)] [[PubMed](#)]
11. Khan, S.; Cao, Q.; Zheng, Y.M.; Huang, Y.Z.; Zhu, Y.G. Health risks of heavy metals in contaminated soils and food crops irrigated with wastewater in Beijing, China. *Environ. Pollut.* **2008**, *152*, 686–692. [[CrossRef](#)] [[PubMed](#)]
12. Schmidt, S.A.; Gukelberger, E.; Hermann, M.; Fiedler, F.; Großmann, B.; Hoinkis, J.; Ghosh, A.; Chatterjee, D.; Bundschuh, J. Pilot study on arsenic removal from groundwater using a small-scale reverse osmosis system towards sustainable drinking water production. *J. Hazard. Mater.* **2016**, *318*, 671–678. [[CrossRef](#)]
13. Fu, F.; Wang, Q. Removal of heavy metal ions from wastewaters: A review. *J. Environ. Manage.* **2011**, *92*, 407–418. [[CrossRef](#)]
14. Saleh, T.A.; Sari, A.; Tuzen, M. Optimization of parameters with experimental design for the adsorption of mercury using polyethylenimine modified activated carbon. *J. Environ. Chem. Eng.* **2017**, *5*, 1079–1088. [[CrossRef](#)]
15. Benhammou, A.; Yaacoubi, A.; Nibou, L.; Tanouti, B. Adsorption of metal ions onto Moroccan stevensite: Kinetic and isotherm studies. *J. Colloid Interface Sci.* **2005**, *282*, 320–326. [[CrossRef](#)]
16. Geyikçi, F.; Kılıç, E.; Çoruh, S.; Eleveli, S. Modelling of lead adsorption from industrial sludge leachate on red mud by using RSM and ANN. *Chem. Eng. J.* **2012**, *183*, 53–59. [[CrossRef](#)]
17. Wang, S.; Peng, Y. Natural zeolites as effective adsorbents in water and wastewater treatment. *Chem. Eng. J.* **2010**, *156*, 11–24. [[CrossRef](#)]
18. Perrich, J.R. *Activated Carbon Adsorption for Wastewater Treatment*; Fla: Boca Raton, FL, USA; CRC Press: Chicago, IL, USA, 2018. [[CrossRef](#)]
19. Halder, G.; Dhawane, S.; Barai, P.K.; Das, A. Optimizing chromium (VI) adsorption onto superheated steam activated granular carbon through response surface methodology and artificial neural network. *Environ. Prog. Sustain.* **2015**, *34*, 638–647. [[CrossRef](#)]
20. Abbas, A.; Al-Amer, A.M.; Laoui, T.; Al-Marri, M.J.; Nasser, M.S.; Khraisheh, M.; Atieh, M.A. Heavy metal removal from aqueous solution by advanced carbon nanotubes: Critical review of adsorption applications. *Sep. Purif. Technol.* **2016**, *157*, 141–161. [[CrossRef](#)]
21. Davodi, B.; Ghorbani, M.; Jahangiri, M. Adsorption of mercury from aqueous solution on synthetic polydopamine nanocomposite based on magnetic nanoparticles using Box–Behnken design. *J. Taiwan Inst. Chem. Engrs.* **2017**, *80*, 363–378. [[CrossRef](#)]
22. Fan, M.; Li, T.; Hu, J.; Cao, R.; Wei, X.; Shi, X.; Ruan, W. Artificial neural network modeling and genetic algorithm optimization for cadmium removal from aqueous solutions by reduced graphene oxide-supported nanoscale zero-valent iron (nZVI/rGO) composites. *Materials* **2017**, *10*, 544. [[CrossRef](#)]
23. Singh, D.K.; Verma, D.K.; Singh, Y.; Hasan, S.H. Preparation of CuO nanoparticles using Tamarindus indica pulp extract for removal of As (III): Optimization of adsorption process by ANN-GA. *J. Environ. Chem. Eng.* **2017**, *5*, 1302–1318. [[CrossRef](#)]
24. Peng, W.; Li, H.; Liu, Y.; Song, S. A review on heavy metal ions adsorption from water by graphene oxide and its composites. *J. Mol. Liq.* **2017**, *230*, 496–504. [[CrossRef](#)]
25. Mandal, S.; Mahapatra, S.S.; Sahu, M.K.; Patel, R.K. Artificial neural network modelling of As (III) removal from water by novel hybrid material. *Process Saf. Environ. Prot.* **2015**, *93*, 249–264. [[CrossRef](#)]
26. Minamisawa, M.; Minamisawa, H.; Yoshida, S.; Takai, N. Adsorption behavior of heavy metals on biomaterials. *J. Agric. Food Chem.* **2004**, *52*, 5606–5611. [[CrossRef](#)]
27. Krishna, D.; Sree, R.P. Artificial neural network and response surface methodology approach for modeling and optimization of chromium (VI) adsorption from waste water using Ragi husk powder. *Indian Chem. Eng.* **2013**, *55*, 200–222. [[CrossRef](#)]
28. Alimohammadi, M.; Saeedi, Z.; Akbarpour, B.; Rasoulzadeh, H.; Yetilmesoy, K.; Al-Ghouti, M.A.; Khraisheh, M.; McKay, G. Adsorptive removal of arsenic and mercury from aqueous solutions by eucalyptus leaves. *Water Air Soil Pollut.* **2017**, *228*, 429. [[CrossRef](#)]

29. Kiran, R.S.; Madhu, G.M.; Satyanarayana, S.V.; Kalpana, P.; Rangaiah, G.S. Applications of Box–Behnken experimental design coupled with artificial neural networks for biosorption of low concentrations of cadmium using *Spirulina (Arthrospira) spp.* *Resour. Effic. Technol.* **2017**, *3*, 113–123. [[CrossRef](#)]
30. Inyang, M.I.; Gao, B.; Yao, Y.; Xue, Y.; Zimmerman, A.; Mosa, A.; Pullammanappallil, P.; Ok, Y.S.; Cao, X. A review of biochar as a low-cost adsorbent for aqueous heavy metal removal. *Crit. Rev. Environ. Sci. Technol.* **2016**, *46*, 406–433. [[CrossRef](#)]
31. Zhu, X.; Wang, X.; Ok, Y.S. The application of machine learning methods for prediction of metal sorption onto biochars. *J. Hazard. Mater.* **2019**, *378*, 120727. [[CrossRef](#)] [[PubMed](#)]
32. Emigdio, Z.; Abatal, M.; Bassam, A.; Trujillo, L.; Juarez-Smith, P.; El Hamzaoui, Y. Modeling the adsorption of phenols and nitrophenols by activated carbon using genetic programming. *J. Clean. Prod.* **2017**, *161*, 860–870. [[CrossRef](#)]
33. Febrianto, J.; Kosasih, A.N.; Sunarso, J.; Ju, Y.; Indraswati, N.; Ismadji, S. Equilibrium and kinetic studies in adsorption of heavy metals using biosorbent: A summary of recent studies. *J. Hazard. Mater.* **2009**, *162*, 616–645. [[CrossRef](#)]
34. Vithanage, M.; Rajapaksha, A.U.; Dou, X.; Bolan, N.S.; Yang, J.E.; Ok, Y.S. Surface complexation modeling and spectroscopic evidence of antimony adsorption on ironoxide-rich red earth soils. *J. Colloid Interface Sci.* **2013**, *406*, 217–224. [[CrossRef](#)]
35. Bhagat, S.K.; Tung, T.M.; Yaseen, Z.M. Development of artificial intelligence for modeling wastewater heavy metal removal: State of the art, application assessment and possible future research. *J. Clean. Prod.* **2020**, *250*, 119473. [[CrossRef](#)]
36. Sakizadeh, M. Artificial intelligence for the prediction of water quality index in groundwater systems. *Model. Earth Syst. Environ.* **2016**, *2*, 8. [[CrossRef](#)]
37. Hafsa, N.; Al-Yaari, M.; Rushd, S. Prediction of arsenic removal in aqueous solutions with non-neural network algorithms. *Can. J. Chem. Eng.* **2020**, in press. [[CrossRef](#)]
38. Ahmadi, M.; Chen, Z. Machine learning models to predict bottom hole pressure in multi-phase flow in vertical oil production wells. *Can. J. Chem. Eng.* **2019**, *97*, 2928–2940. [[CrossRef](#)]
39. Guo, Y.; Bartlett, P.; Shawe-Taylor, J.; Williamson, R. Covering numbers for support vector machines. *IEEE Trans. Inf. Theory* **2002**, *48*, 239–250. [[CrossRef](#)]
40. Durbha, S.; King, R.; Younan, N. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sens. Environ.* **2007**, *107*, 348–361. [[CrossRef](#)]
41. Omer, G.; Mutanga, O.; Abdel-Rahman, E.; Adam, E. Performance of support vector machines and artificial neural network for mapping endangered tree species using WorldView-2 data in Dukuduku Forest, South Africa. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2015**, *8*, 4825–4884. [[CrossRef](#)]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
43. Čeh, M.; Kilibarda, M.; Lisec, A.; Bajat, B. Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 168. [[CrossRef](#)]
44. Wei, L.; Yuan, Z.; Zhong, Y.; Yang, L.; Hu, X.; Zhang, Y. An improved gradient boosting regression tree estimation model for soil heavy metal (arsenic) pollution monitoring using hyperspectral remote sensing. *Appl. Sci.* **2019**, *9*, 1943. [[CrossRef](#)]
45. Cha, Y.; Kim, Y.; Choi, J.; Sthiannopkao, S.; Cho, K. Bayesian modeling approach for characterizing groundwater arsenic contamination in the Mekong River basin. *Chemosphere* **2016**, *143*, 50–56. [[CrossRef](#)]
46. Yetilmezsoy, K.; Demirel, S.; Vanderbei, R.J. Response surface modeling of Pb (II) removal from aqueous solution by *Pistacia vera* L.: Box–Behnken experimental design. *J. Hazard. Mater.* **2009**, *171*, 551–562. [[CrossRef](#)]
47. Podder, M.S.; Majumder, C.B. The use of artificial neural network for modelling of phycoremediation of toxic elements As (III) and As (V) from wastewater using *Botryococcus braunii*. *Spectrochim. Acta A* **2016**, *155*, 130–145. [[CrossRef](#)]
48. Won, W.; Lee, K. Adaptive predictive collocation with a cubic spline interpolation function for convection-dominant fixed-bed processes: Application to a fixed-bed adsorption process. *Chem. Eng. J.* **2011**, *166*, 240–248. [[CrossRef](#)]
49. Aguilera, A.; Morillo, A. Comparative study of different B-spline approaches for functional data. *Math. Comput. Model.* **2013**, *58*, 1568–1579. [[CrossRef](#)]

50. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.R-project.org/> (accessed on 20 September 2020).
51. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE PAMI* **2009**, *32*, 569–575. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).