Arabian Journal of Chemistry (2022) **15**, 104261

King Saud University

Arabian Journal of Chemistry

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Computational modeling of Hg/Ni ions separation via MOF/LDH nanocomposite: Machine learning based modeling

Munjed M. Ibrahim [a], Maha Abdallah Alnuwaiser [b], Eslam B. Elkaeed [c], Hossam Kotb [d], Sameer Alshehri [e], Mohammed A.S. Abourehab [f,g,*]

[a] *Department of Pharmaceutical Chemistry, College of Pharmacy, Umm Al-Qura University, 21955 Makkah, Saudi Arabia*
[b] *Department of Chemistry, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[c] *Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Riyadh 13713, Saudi Arabia*
[d] *Department of Electrical Power and Machines, Faculty of Engineering, Alexandria University, Alexandria, Egypt*
[e] *Department of Pharmaceutics and Industrial Pharmacy, College of Pharmacy, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*
[f] *Department of Pharmaceutics, College of Pharmacy, Umm Al-Qura University, Makkah 21955, Saudi Arabia*
[g] *Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Minia University, Minia 61519, Egypt*

Received 1 July 2022; accepted 11 September 2022
Available online 16 September 2022

KEYWORDS

Water treatment;
Nanocomposite material;
Artificial intelligence;
Common pollutants

Abstract Nowadays, sustainable supplement of water has recently been identified as a vital necessity due to the existence of limited drinkable water sources. To do this, various techniques are being developed to remove various types of pollutants from water/wastewater sources. Adsorption of common water pollutants using nanocomposite materials has been of great popularity in recent years due to its high efficiency. This paper aims to develop various models based on machine learning approach to study their efficiency on predicting the experimentally measured results of Hg/Ni ions removal from water sources. To do this, this study attempts regression on a small data set using two parameters as inputs and two parameters as outputs. In this dataset, the inputs are Ion and C0, and the outputs are Ce and Qe. AdaBoost (Adaptive Boosting), a well-known ensemble method, was applied on top of three different models, including Decision Tree Regression (DT), Gaussian Process Regression (GPR), and Linear Regression (LR). After fine-tuning their hyper-

* Corresponding authors at: Department of Pharmaceutics, College of Pharmacy, Umm Al-Qura University, Makkah 21955, Saudi Arabia (M.A.S. Abourehab).
E-mail address: maabourehab@uqu.edu.sa (M.A.S. Abourehab).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.arabjc.2022.104261

parameters, the optimized model was evaluated through various metrics. For example, the R^2 for ADA + GPR model has a score of 0.998 for Ce and 0.999 for Qe as the best model among these three models. This model in RMSE is the best and illustrates 0.1512 and 1.490 for Ce and Qe as error. Eventually, ADA + GPR has been selected as the optimized model with optimized dataset: (Ion = Ni, C0 = 250, Ce = 206.0). But for Qe, different amounts are illustrated: (Ion = Hg, C0 = 106.7, Ce = 577.35)

1. Introduction

The growth of population has exacerbated the water scarcity in developing countries (Pendergast and Hoek, 2011; Azmil, et al., 2022). Sustainable supplement of water has become a serious global challenge in the next forty years. Upon the anticipation of the World Water Council, by 2030, more than 3.9 billion of the world will face with a serious water stress situation and water-scarce countries (Urgency, 2007). Molecular-based separation of ions is of great importance in disparate industrial-based activities for the pretreatment of raw materials and purification of products (Lyu, 2018; Sharif, 2021). Relying on water resources, disparate pollutants such as organic substances, pharmaceutical agents, dyes and heavy metals may be available in water and wastewater streams, which their efficient removal can be an important activity (Mengting, 2019; Qin, 2020).

In various industrial topics such as gas sweetening, water/wastewater treatment, adsorption and pharmaceutics, the separation process possesses great importance to be the subject of interest (Nagase, 2020; Memon, 2022; Chen, 2021). Adsorption is known as one of the most well-mature and promising techniques for molecular-base removal of pollutants/impurities from water-/wastewater sources (Li, 2022; Nayak et al., 2022). The adsorbent structure is an important parameter for the separation of contaminants, which must be designed to obtain to hit a satisfactory specific point (Yang, 2021).

To enhance the adsorption performance of substances, the usage of nanocomposite materials would be an appropriate option (Yin and Deng, 2015). Considering high dependency of the adsorption capacity on the absorbent's surface area/chemistry, these parameters have suitable potential of reinforcement by fabricating a nanocomposite structure like metal organic framework (MOF)/Layered double hydroxide (MOF/LDH). In recent years, the synthesis and use of MOF/LDH composite materials have been prevalent to increase the removal efficiency of ions/organic molecules from water (Mohd Sidek, 2017; Abasi, 2022).

Over the past years, artificial intelligence (AI) has opened new horizons towards predicting the experimental measured results in various industries like membrane-based separation, adsorption, pharmaceutics, and chemical reactions (Gaudio, 2021; Elbadawi, 2021; Mahmoodi et al., 2018). In various scientific disciplines, machine learning (ML) methodologies are progressively taking the role of classical computing methods. Neural Network, Ensemble, and Tree-based models are examples of these techniques to solve problems. Many problems with a few input attributes and multiple output values may now be studied using machine learning models. These models derive some correlations between inputs and outputs (Bishop and Nasrabadi, 2006; Rodriguez-Galiano, 2015).

Ensemble methods are roughly the models that use several base models to make more general and robust predictors (Dietterich, 2000). This paper aims to develop various novel models based on machine learning approach to study their efficiency on predicting the experimentally measured results of Hg/Ni ions removal from water sources. In this research, we used a well-known ensemble method, AdaBoost (Adaptive Boosting), on the top of three different models, including Linear Regression (LR), Decision Tree Regression (DT), and Gaussian Process Regression (GPR).

As one of the most popular non-parametric models, Decision trees (DT) are straightforward, hierarchical models that employ recursive splits on observed examples to accumulate close examples in minor portions of the input space (recursive-partitioning regression) (Alpaydin, 2020; Quinlan, 1996; Hassan, 2017). Furthermore, Random Forest, Extra Trees, and Gradient Boosting Trees are applicable ensemble models based on DT used in this research.

GPR is more popular because it can be connected to Bayesian non-parametric statistics, infinite neural networks, kernel approaches in machine learning, and spatial statistics (Rasmussen, 2003; Shi and Choi, 2011). Also, linear regression (LR) is a straightforward approach to supervised learning. Linear regression, in particular, is a useful method to predict a quantitative response.

2. Data set

Ce and Qe outputs have to be analyzed using a small dataset provided by (Syah, 2022). Ion is a categorical input, and C0 is a real-valued. The dataset including 18 items which is shown in Table 1 (Hu, 2022).

3. Methodology

3.1. Base models

The decision tree (DT) is a well-known model inspired by nature that can be used to solve various problems. A decision tree is composed of three parts: a root (start) node; multiple decision nodes to check some queries on data (internal); and multiple leaf nodes that are final predictions (terminal). The output of the model is reflected by leaf (terminal) nodes, whilst incoming data enters the network via the root node. Decision nodes are located between the root and the terminal nodes. In general, incoming data begins to flow through branches from the root and passes to sub-trees based on the query of each node to reach terminal nodes. When the algorithm receives the input data, it begins to create the tree by splitting, pruning, and stopping (Quinlan, 1996; Mathuria, 2013; Xu, 2005; Yang, 2017; Song and Ying, 2015). These behaviors begin at the root node and progress until a specific condition is met. A simple decision tree is depicted schematically in Fig. 1 (Hu, 2022).

Nonlinear regression approach which utilizes a probabilistic regression framework to improve the errors but does not utilize parametric models, such as the Gaussian Process Regression (GPR) (Grbić et al., 2013; Alqarni, 2022). Here, the presumption is that y as output calculations is taken below method:

$$y = f(\mathbf{x}(k)) + \varepsilon$$

Table 1 The whole data set (Syah, 2022; Hu, 2022).

No.	X_1 = Ion	X_2 = C0	Y_2 = Qe	Y_1 = Ce
1	Ni	0.5	5	0
2	Ni	2	20	0
3	Ni	5	50	0
4	Ni	10	95	0.5
5	Ni	25	230	2
6	Ni	50	380	12
7	Ni	100	426	57
8	Ni	160	433	116
9	Ni	250	439	206
10	Hg	0.5	5	0
11	Hg	2	20	0
12	Hg	5	50	0
13	Hg	10	99	0.1
14	Hg	25	247	0.3
15	Hg	50	445	5
16	Hg	100	500	50
17	Hg	160	518	108
18	Hg	250	527	197

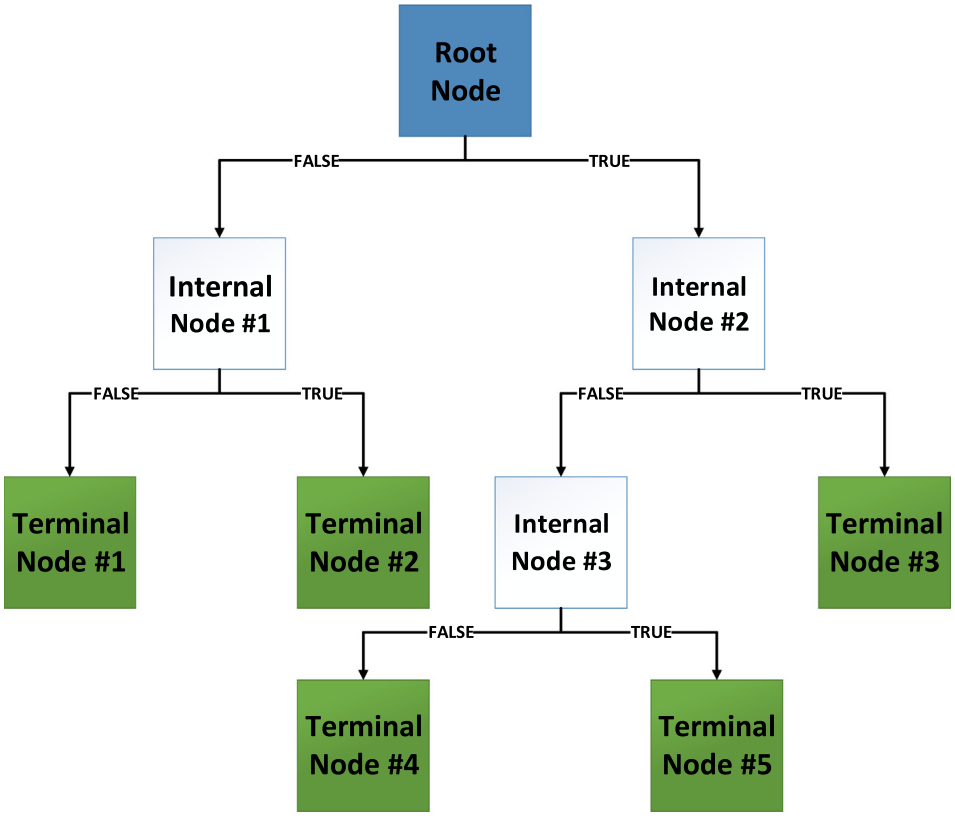

Fig. 1 General structure of Decision Tree.

In the above equation, x reflects a quantification of input value, f demonstrates the indistinct functional interdependence, then ε notes Gaussian noise σ_n^2 denotes the variance of Gaussian noise. Instead of parameterizing function f, the previous probability is expressed using the Gaussian process (Rasmussen, 2003).

The Gaussian process's covariance function $cov(x,x')$ and mean $m(x)$ carry ideas about the data's underlying mechanism. After defining the covariance function and mean, the result is determined with regard to the Gaussian distribution $p\ (y^*|X, y, x^*\ )$ with:

$$\hat{y}_* = m(\mathbf{x}_*) + \mathbf{k}_*^{\mathrm{T}}\left(\mathbf{K} + \sigma_n^2\mathbf{I}\right)^{-1}(\mathbf{y} - m(\mathbf{x}_*)),$$
$$\sigma_{y_*}^2 = k_* + \sigma_n^2 - \mathbf{k}_*^{\mathrm{T}}\left(\mathbf{K} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{k}_*,$$

The prediction is constructed on the training vector X, y, as illustrated below. Instead, then relying exclusively on the parameters to make the forecast, this approach departs from typical regression methods (Jin, 2022).

Variable K refers to a covariance matrix by the elements $K_{i,j} = cov(x_i, x_j)$ and k* which illustrate in the following formulation (Jin, 2022):

$$[k*]i = cov(xi, x*) and k* = cov(x*, x*)$$

The covariance function and mean parameters must be obtained using the dataset in order to generate reliable predictions. Because they describe parts of the predictive probability distribution, these variables are called hyper-parameters. Values of hyper-parameters are primarily arranged through optimizing the log-likelihood estimate of the train dataset (Rasmussen and Nickisch, 2010):

$$\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2}\mathbf{y}^{\mathrm{T}}\left(\mathbf{K} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{y} - \frac{1}{2}\log\left(|\mathbf{K} + \sigma_n^2\mathbf{I}|\right) - \frac{n}{2}\log(2\pi)$$

n illustrates the quantity of training subset.

3.2. Adaboost

An ensemble learning-based model can exceed a single predictor in terms of performance by merging numerous base learners. As the AdaBoost method, it serves as an ensemble learning method for improving the robustness of base estimators by altering the weights of samples (Freund and Schapire, 1997). This approach has increased in prominence as a result of its capabilities (Schapire, 2003; Ying, 2013).

To deal with more complex challenges, these adaptively improved base models might be used. In addition, this is since basic models have high generalization properties, which makes them appealing. Even though they can be used in practical situations, they are unable to handle complex tasks due to their design bias. AdaBoost is one of the solutions to improve simple models.

This method starts with a base predictor (the weak learner), and then the base predictor and the other predictors are continuously and consistently combined to build a trustworthy system that can deal with complex scenarios (Lemaître et al., 2017; Drucker, 1997). These steps summarize the general AdaBoost algorithm (Hastie et al., 2001; Bishop, 2006):

Initialize the weights and number of base learners (M):

$$w_i = \frac{1}{\mathrm{N}}, \quad i = 1, \cdots, N$$

For k ∈ {1, …, M}:

i. Build an estimator $\mathrm{G_k(x)}$ using the weights w_i.
ii. $err_k = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_k(x_i))}{\sum_{i=1}^{n} w_i}$.
iii. $\alpha_k = \log(\frac{1-err_k}{err_k})$.
iv. $w_i \leftarrow w_i . \exp\left(\alpha_k . I(y_i \neq G_k(x_i))\right), i = 1, \cdots, N$.

Final Output:

$$\mathrm{G(x)} = sign\left(\sum_{k=1}^{M} \alpha_k G_k(x)\right)$$

In this pseudocode, the quantity of the data points and the learners are shown with N and M. The predictor employing b over the data points is $\mathrm{G_b(x)}$ (Hastie, 2009; Berk, 2006; Ouyang et al., 2021).

4. Results and discussion

The three new models were each evaluated according to the criteria that are typically used, and then their hyper-parameters were tweaked using the test data. These requirements are outlined in the following paragraphs.

R^2 score or coefficient of determination is, without a doubt, the most used criteria for assessing the efficiency of prediction results. It reveals how well the patterns of the projected findings mirror the trends of observations (Gouda, 2019).

$$R^2 = 1 - \frac{\sum (\mathrm{y_i} - \mathrm{x_i})^2}{\sum (\mathrm{x_i} - \bar{\mathrm{x}}_\mathrm{i})^2}$$

The RMSE is calculated using the sample standard deviation of the distinction among the estimated and predicted targets. It provides an accuracy measure for the predicted values (Zang, 2020).

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{\mathrm{n}} (\mathrm{y_i} - \mathrm{x_i})^2}$$

y_i and x_i reflect predicted and actual (observed) outputs, respectively. n stands for the size of dataset and $\underline{\mathrm{x}}_\mathrm{i}$ is the mean of observed data. Tables 2 and 3 compare the obtained results of different models. As can be seen, despite the satisfactory performance of ADA-DT and ADA-LR models to predict the experimental measured results, the use of ADA-GPR has illustrated better agreement between the model outcomes and measured values with great R^2. The calculated R^2 for the prediction of both Ce and Qe are higher than 0.999, which corroborates that the ADA-GPR model is significantly valid to predict the values in the simulation process of adsorption isotherms and equilibrium data. Considering Tables 2 and 3, we can conclude that the Boosted GPR procedure is the most general of the available models. In the end, the GPR model, which is depicted in Figs. 2 and 3 of the model surfaces, is chosen according to the facts in the preceding tables.

Table 2 Final outputs for Ce.

Models	RMSE	R^2
ADA-DT	0.2089	0.946
ADA-GPR	0.1512	0.998
ADA-LR	0.2854	0.941

Table 3 Final model results for Qe.

Models	RMSE	R^2
ADA-DT	2.893	0.957
ADA-GPR	1.490	0.999
ADA-LR	3.162	0.948

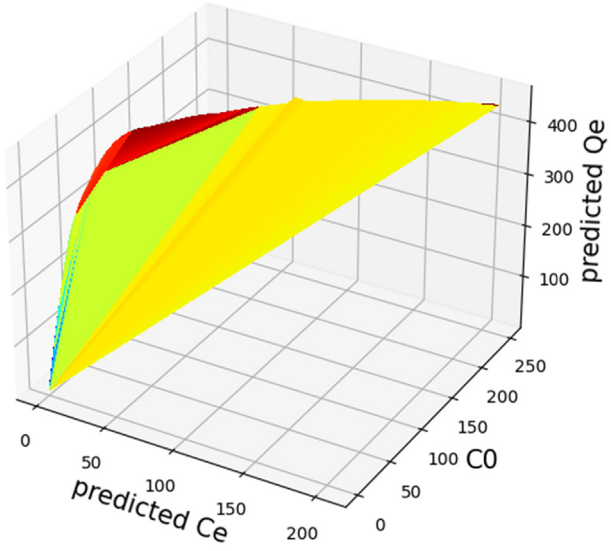

Fig. 2 Simulated surface for Ni.

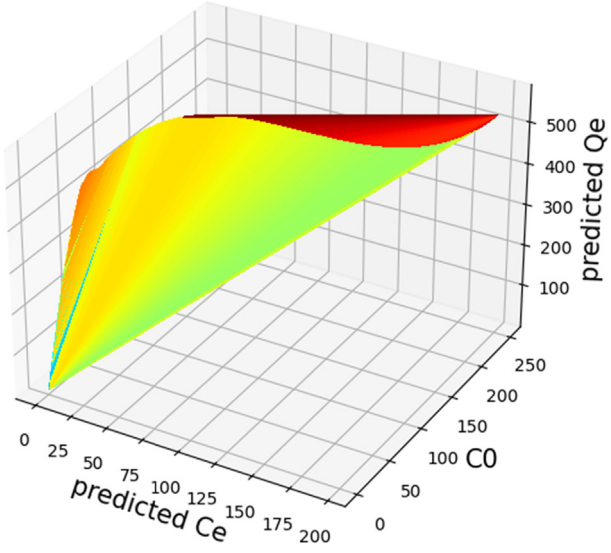

Fig. 3 Simulated surface for Hg.

After ensuring the validation of developed models, the validated ML model was applied to achieve the adsorption isotherm for Ni and Hg ions. Figs. 4 and 5 present three-dimensional simulated surface plot about the adsorption of both Ni and Hg ions. As demonstrated, presentation of the

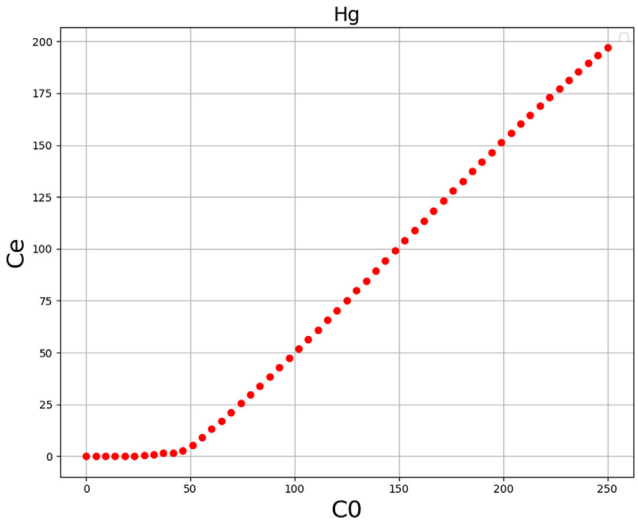

Fig. 4 2D Plot of Ce applying Ada-GPR through Hg Ion.

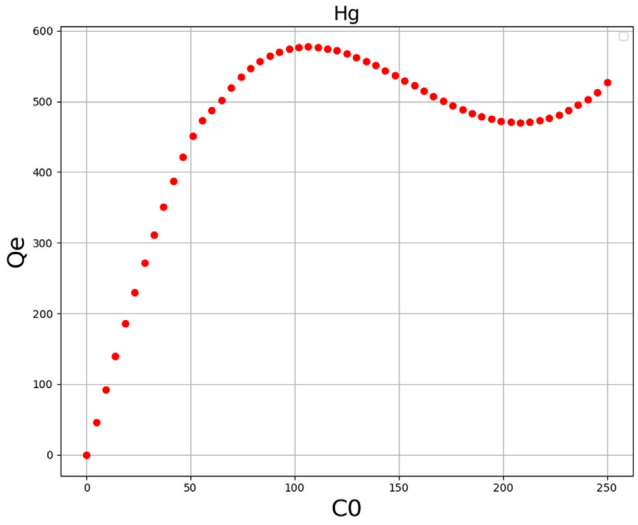

Fig. 5 6-2D Plot of Qe applying Ada-GPR through Hg Ion.

adsorption isotherms of Hg and Ni ions are based on adsorption capacity versus equilibrium concentration. According to the results, it can be perceived the amount of Hg adsorption on the surface of MOF/LDH structure is more than Ni because of the interaction of Hg with the surface of nanocomposite. It is worth pointing out that the increment of the solutes' initial concentration in the solution resulted in a substantial improvement in the estimated adsorption capacity (Qe). Increase in the solutes' initial concentration in the solution can be justified due to enhancing the driving force for mass transfer among the bulk of solution and the adsorbent surface. The obtained results can be of great importance to present the two-dimensional demonstrations of the adsorption capacity versus some momentous parameters such as initial concentration. Figs. 5–8 demonstrate the two-dimensional plots of Ce and Qe using Ada-GPR models for Hg and Ni ions. According to the simulation outcomes, it is perceived with the aim of improving the adsorption removal, the

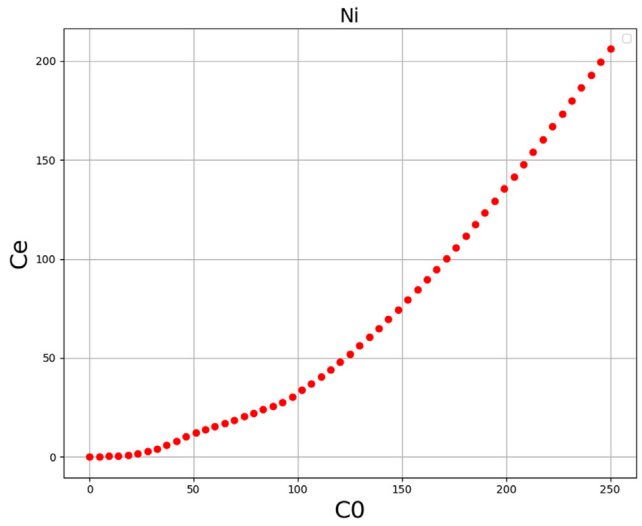

Fig. 6 2D Plot of Ce applying Ada-GPR applying Ni Ion.

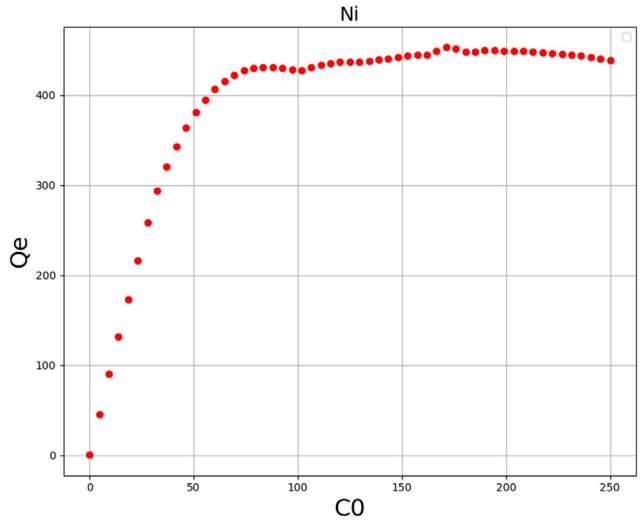

Fig. 7 2D Plot of Qe applying Ada-GPR through Ni Ion.

Table 4 Optimized values.

Output	Ion	C0	Output Value
Ce	Ni	250	205.93
Qe	Hg	106.7	577.35

adsorption dosage must be maintained at its maximum values. Despite increasing the separation costs, achieving the optimized value of adsorbent is of great importance (Yang, 2021).

Table 4 shows the optimal values for both outputs separately, with the optimal values for each output shown in parentheses.

5. Conclusion

Finally, a plenary modeling and simulation through machine learning approach have been developed to investigate the adsorption process of Hg and Ni ions from solutions using MOF/LDH nanocomposite material. In order to develop the ideal model, ions and the primary amount of the concentration of solute have been selected as input, then, the training process of network has been implemented applying the experimentally calculated value achieved form literature. Using only two inputs and two outputs, this study attempts regression. Ce and Qe are the outputs of this dataset, which takes as inputs Ion and C0. On top of three models, including DT, GPR, and LR, the well-known ensemble method, AdaBoost, was applied. Through modifying their hyper-parameters, the models were performed and evaluated applying a variety of metrics. According to the metric, ADA + GPR has the top R-squared score of the three models, with a score of 0.998 for Ce and 0.999 for Qe. For Ce and Qe, this model in RMSE has an error rate of 0.1512 and 1.490. Finally, the ADA + GPR model's optimal values (Ion = Ni, C0 = 250, Ce = 206.0) are identical to the dataset's optimal values. Although it is equal to (Ion = Hg, C0 = 106.7, and Ce = 577.35) for Qe.

Data Availability:

All data are available within the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to acknowledge the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant code: 22UQU4310259DSR01.

This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R186), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

Abasi, C. et al, 2022. Postsynthetic Modification of Metal Organic Frameworks with Substituted Layered Double Hydroxides for Adsorption of Lead Ions. Compos. Nano Eng. 1 (1), 1–15.

Alpaydin, E., 2020. Introduction to Machine Learning. MIT Press.

Alqarni, M. et al, 2022. Solubility Optimization of Loxoprofen as a Nonsteroidal Anti-Inflammatory Drug: Statistical Modeling and Optimization. Molecules 27 (14), 4357.

Azmil, A.B.A., et al., 2022. Short Review: Advanced Water Treatment Technology.

Berk, R.A., 2006. An Introduction to Ensemble Methods for Data Analysis. Sociol. Methods Res. 34 (3), 263–295.

Bishop, C.M., 2006. Pattern recognition. Mach. Learn. 128, 9.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning, vol. 4. Springer.

Chen, T. et al, 2021. Mechanistic study on boron adsorption and isotopic separation with magnetic magnetite nanoparticles. J. Mater. Sci. 56 (7), 4624–4640.

Dietterich, T.G., 2000. Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems. Springer.

Drucker, H., 1997. Improving regressors using boosting techniques. In: ICML. Citeseer.
Elbadawi, M. et al, 2021. Harnessing artificial intelligence for the next generation of 3D printed medicines. Adv. Drug Deliv. Rev. 175, 113805.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.
Gaudio, M.T. et al, 2021. Artificial Intelligence-Based Optimization of Industrial Membrane Processes. Earth Syst. Environ. 5 (2), 385–398.
Gouda, S.G. et al, 2019. Model selection for accurate daily global solar radiation prediction in China. J. Cleaner Prod. 221, 132–144.
Grbić, R., Kurtagić, D., Slišković, D., 2013. Stream water temperature prediction based on Gaussian process regression. Expert Syst. Appl. 40 (18), 7407–7414.
Hassan, M.A. et al, 2017. Exploring the potential of tree-based ensemble methods in solar radiation modeling. Appl. Energy 203, 897–916.
Hastie, T. et al, 2009. Multi-class adaboost. Stat. Interf. 2 (3), 349–360.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. Springer Series in Statistics. Springer.
Hu, X. et al, 2022. Predictive modeling and computational machine learning simulation of adsorption separation using advanced nanocomposite materials. Arab. J. Chem. 15, (9) 104062.
Jin, H. et al, 2022. Computational simulation using machine learning models in prediction of CO2 absorption in environmental applications. J. Mol. Liq. 358, 119159.
Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res. 18 (1), 559–563.
Li, A. et al, 2022. Removal of heavy metals from wastewaters with biochar pyrolyzed from MgAl-layered double hydroxide-coated rice husk: Mechanism and application. Bioresour. Technol. 347, 126425.
Lyu, J. et al, 2018. Separation and purification using GO and r-GO membranes. RSC Adv. 8 (41), 23130–23151.
Mahmoodi, F., Darvishi, P., Vaferi, B., 2018. Prediction of coefficients of the Langmuir adsorption isotherm using various artificial intelligence (AI) techniques. J. Iran. Chem. Soc. 15 (12), 2747–2757.
Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 3 (6).
Memon, F.H. et al, 2022. Transition metal dichalcogenide-based membranes for water desalination, gas separation, and energy storage. Sep. Purif. Rev., 1–15
Mengting, Z. et al, 2019. Applicability of BaTiO3/graphene oxide (GO) composite for enhanced photodegradation of methylene blue (MB) in synthetic wastewater under UV–vis irradiation. Environ. Pollut. 255, 113182.
Mohd Sidek, H.B. et al, 2017. Enhancement of the Water Adsorptivity of Metal-Organic Frameworks upon Hybridization with Layered Double Hydroxide Nanosheets. J. Phys. Chem. C 121 (28), 15008–15016.
Nagase, K. et al, 2020. Antibody drug separation using thermoresponsive anionic polymer brush modified beads with optimised electrostatic and hydrophobic interactions. Sci. Rep. 10 (1), 1–13.
Nayak, V., Cuhorka, J., Mikulášek, P., 2022. Separation of Drugs by Commercial Nanofiltration Membranes and Their Modelling. Membranes 12 (5), 528.
Ouyang, Z., Ravier, P., Jabloun, M., 2021. STL Decomposition of Time Series Can Benefit Forecasting Done by Statistical Methods but Not by Machine Learning Ones. Eng. Proc. 5 (1).
Pendergast, M.M., Hoek, E.M., 2011. A review of water treatment membrane nanotechnologies. Energy Environ. Sci. 4 (6), 1946–1971.
Qin, H. et al, 2020. The improved methods of heavy metals removal by biosorbents: A review. Environ. Pollut. 258, 113777.
Quinlan, J.R., 1996. Learning decision tree classifiers. ACM Comput. Surv. (CSUR) 28 (1), 71–72.
Rasmussen, C.E., 2003. Gaussian processes in machine learning. Summer School on Machine Learning. Springer.
Rasmussen, C.E., Nickisch, H., 2010. Gaussian processes for machine learning (GPML) toolbox. J. Mach. Learn. Res. 11, 3011–3015.
Rodriguez-Galiano, V. et al, 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818.
Schapire, R.E., 2003. The boosting approach to machine learning: An overview. Nonlinear Estimat. Classif., 149–171
Sharif, S. et al, 2021. Two-dimensional graphene oxide based membranes for ionic and molecular separation: Current status and challenges. J. Environ. Chem. Eng. 9, (4) 105605.
Shi, J.Q., Choi, T., 2011. Gaussian Process Regression Analysis for Functional Data. CRC Press.
Song, Y.-Y., Ying, L., 2015. Decision tree methods: applications for classification and prediction. Shanghai Arch. Psyc. 27 (2), 130.
Syah, R. et al, 2022. Artificial Intelligence simulation of water treatment using nanostructure composite ordered materials. J. Mol. Liq. 345, 117046.
Urgency, U., 2007. Water Caucus Summary. World Water Council (WWC), Marseille, France, pp. 59.
Xu, M. et al, 2005. Decision tree regression for soft classification of remote sensing data. Remote Sens. Environ. 97 (3), 322–336.
Yang, L. et al, 2017. A regression tree approach using mathematical programming. Expert Syst. Appl. 78, 347–357.
Yang, J. et al, 2021. Artificial intelligence simulation of water treatment using a novel bimodal micromesoporous nanocomposite. J. Mol. Liq. 340, 117296.
Yin, J., Deng, B., 2015. Polymer-matrix nanocomposite membranes for water treatment. J. Membr. Sci. 479, 256–275.
Ying, C. et al, 2013. Advance and prospects of AdaBoost algorithm. Acta Autom. Sin. 39 (6), 745–758.
Zang, H. et al, 2020. Application of functional deep belief network for estimating daily global solar radiation: A case study in China. Energy 191, 116502.