


Research Article

Gaussian Process Regression and Machine Learning Methods for Carbon-Based Material Adsorption

Manar Ahmed Hamza ¹, **Maha M. Althobaiti**,² **Fahd N. Al-Wesabi**,^{3,4} **Rana Alabdan**,⁵ **Hany Mahgoub**,^{3,6} **Anwer Mustafa Hilal**,¹ **Abdelwahed Motwakel**,¹ and **Mesfer Al Duhayyim**⁷

¹Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

²Department of Computer Science, College of Computing and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

³Department of Computer Science, College of Science & Arts at Muhayil, King Khalid University, Saudi Arabia

⁴Faculty of Computer and IT, Sana'a University, Yemen

⁵Department of Information Systems, College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia

⁶Faculty of Computers and Information, Computer Science Department, Menoufia University, Egypt

⁷Department of Natural and Applied Sciences, College of Community-Aflaj, Prince Sattam Bin Abdulaziz University, Saudi Arabia

Correspondence should be addressed to Manar Ahmed Hamza; ma.hamza@psau.edu.sa

Received 18 November 2021; Accepted 17 December 2021; Published 18 January 2022

Academic Editor: Lakshmiathy R

Copyright © 2022 Manar Ahmed Hamza 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.

Antibiotics have received a lot of attention as promising contaminants because of their ecotoxicological and long-term chemical stability in the atmosphere. Antibiotic adsorption on carbon-based materials (CBMs) such as charcoal and activated carbon has been identified as mainly effective for treating the wastewater strategies. Machine learning (ML) approaches were used to create generalized computation methods for tetracycline (TC) and sulfamethoxazole (SMX) adsorption in CBMs in this investigation. In the existing system, random forest and ANN methods were used for TC and SMX for predicting the quantities of antibiotics in the CBMs. For reducing the antibiotics from the industrial wastewater, the broadcast efforts of the experiments are a little complicated. In the proposed method, Gaussian process regression (GPR), active learning (AL), and ANN are used for predicting the antibiotic levels in the industrial wastewater. Below a variety of environmental parameters (e.g., warmth, solution pH) and adsorbent varieties, the created ML algorithms outperformed classic isotherm models in conditions of generalisation. To evaluate TC and SMX adsorption on CBMs, we used comparative significance investigation and partial trust plots based on ML models. The proposed GPR reduces the antibiotics in wastewater; minimal experimental screening and the comparative significance and partial trust plot help in the treatment of wastewater.

1. Introduction

Antibiotics are a very well category of antipathogen medications that were broadly utilized in a variety of disciplines, increasing the risk of unintended discharge to the surroundings. Antibiotics are found in urban sewage stream, rainwa-

ter, underground, and soil pollution, among other complex compartments [1]. Antibiotics are recognised as growing contaminants in the local habitat, with the capacity to create antimicrobial drug microorganisms and antimicrobial agents [2, 3]. Adsorption is considered a strong technique for dealing with antibiotic-rich manufacturing effluents

compared to covering separation, electrochemical oxidation, and biosorption [4–6] because of its tall extraction effectiveness, cheap price, and ease of process [7].

The cost-effective CBMs made by chemical by-products and biological trash, such as activated charcoal (AC) and bio-char (BC), can achieve squander minimisation, material recovery, and valuation uses all at the same time [8]. Because of its well-organized architectures and customised outside composition, carbon nanotubes and graphene-based materials also performed well enough in sewage disposal. Furthermore, standard adsorption process investigations are energy and cash costly. As a result, developing universal forecasting replica for assessing antimicrobial adsorption and gaining additional nearby into the perceived significance of every important component and the groups are more likely on antibiotic sorption would have been beneficial [9–11].

We offer an AL strategy to balancing modelling correctness rate through the amount of training information necessary to get suitable facts placed in this paper. The process is based on Gaussian process regression (GPR), which involves fitting a conducted previous. The GPR model generates a forecast average and a forecast norm deduction (improbability), with the latter being utilized to choose after that single experiment to be run. For discrete hysteresis loop in the thermal region, we illustrate this method by modelling adsorption of individual constituent's gas and atmospheric CO₂ in Cu-BTC. They describe the technique and demonstrate how it can reduce the amount of iterations needed to precisely measure the adsorption by an absolute scale [12].

The explosive growth of easily understood machine learning (e.g., decision tree, random forest, and gradient boosting trees) and analysis procedures for ML models (e.g., local interpretable prototype understandings) has made machine learning (ML) a useful implement for elaborating complex multivariety interactions [13, 14]. In this paper, the Gaussian process regression (GPR) and Artificial Neural Networks (ANN) are used for predicting the antibiotics which are placed in the carbon-based materials (CBMs). The tetracycline (TC) and sulfamethoxazole (SMX) have many ecological problems; it is extensively extended in the surroundings and also has a much unique quality which is studied in this paper. The research helps to identify the solutions for the following:

- (1) How the ML methods predict the quantity of adsorption in TC and SMX?
- (2) How the Gaussian process regression works in antibiotic adsorption?
- (3) How the experimental broadcast is reduced?

The proposed Gaussian process regression and ANN help to predict the antibiotic adsorption in carbon-based materials. The major contribution of this study is given:

- (1) To create general machine learning models for estimating TC/SMX antibiotic adsorption capability on CBMs based on substance parameters and adsorption circumstances

- (2) To examine the quality of ML methods created with the CPR and ANN algorithms
- (3) Evaluate the comparative meaning and power of every substance attribute and adsorption situation on the TC and SMX adsorption ability, as well as the synergy between the elements

The remaining part of our research is written as follows: Section 2 consists of a brief study of existing adsorption on carbon-based materials, Artificial Neural Networks (ANN), and machine learning (ML). Section 3 describes the working principal of the proposed model. Section 4 evaluates the result and gives a comparison of different algorithms. Section 5 concludes the research work.

2. Related Works

Ensemble learning takes the information used to train a model and uses it to create a novel method that can be utilized in a different environment with substantially less input. At 243 K and 100 pressure, researchers developed convolution neural networks for gas adsorption. We subsequently used this as a source domain, with all of the DNN levels remaining unchanged only the last level being suitable for a different target job. Adsorption of hydrogen and methane at room temperatures was one of the new objective tasks. Surprisingly, despite using an enormous amount of fewer data, the transfer's adaptive learning had greater precision than direct retraining. Conveying information from phase equilibrium adsorption of hydrogen or methane to Xe breakups, on the other hand, is a challenge.

From the production of innovative classifiers that captured the significant features for applications of attention, these ML methods contain yielded substantial objective understanding. Substitute systems performing computation magnitudes quicker than the molecular computations to facilitate dependence on it for information have indeed emerged from ML research [15–22]. Large datasets are required for the appropriate guidance and application of numerous ML methods, which poses a hurdle and barrier for processes that depend on ML for forecasts. The possibility of deep learning systems and processes is extremely incomplete in situations when getting information is hard or instance-intensive [23].

AL was recently developed by Uteva and colleagues on antiparticle entropy interfaces, and it outperformed power techniques [24]. Meanwhile, AL was employed to counteract the use of molecular mechanics computations to generate configurations. It [25] provides a modern instance of biomaterials in which they attempt to link disparate dimension and instance scales. They need costly numerical simulations (MD) and calculations to accomplish. So researchers employed AL and a question method to choose experiments depending on replica errors. Researchers demonstrate that building this information set requires an enormous amount of fewer trials [26].

Even without presumptions of typical isotherm, the unique ML models were built straight using existing

experimental published results (such as monolayer adsorption of Langmuir model). As a result, ML models may have a superior classification performance for a wide range of application settings and CBM adsorbent materials [27], although the reliability of model prediction is dependent on the availability and quantity of entering data. The given frame of produced programs may considerably reduce the moment and expensive experimental treatments for the medicine adsorption system, thanks to the specialised design of machine learning algorithms for comprehending the antibiotic adsorption rate.

Since SMX has a relatively low molecular size than TC, SBET of CBMs may have a more essential role in SMX adsorption via a porous process [28], because SMX could infiltrate into tiny micropores that really are unavailable to TC. The Pearson correlation coefficient among SMX and SBET adsorption reached up to 0.951 in a prior study [29]. Alternative methods use an active learning (AL) technique, also called as process is crucial, to assist in combining the prediction model's quality with the number of observations to be collected. This is especially appealing when the feature space is extensive (adsorption under various thermal settings) and/or time-consuming or asset tests or models are required. In the development of molecular modelling, these methods are becoming incredibly common. AL was recently developed by Uteva and colleagues for intermolecular possible energy interfaces, and it outperformed grid-based techniques [30].

Furthermore, we extract the logarithm (base 10) of all of the information (altitude, heat, and adsorbent loads) and standardise the input parameters (force and heat) before running this through the GP procedure to guarantee that the GP is properly fitted. The acquiring mechanism, and how to select a new simulator, is yet another crucial part of AL methods. The goal of this research is to quickly and precisely determine the pressure parameters for a particular adsorbate [31, 32].

Antibiotics have received a lot of attention as promising contaminants because of their ecotoxicological and long-term chemical stability in the atmosphere. Antibiotic adsorption on carbon-based materials (CBMs) such as charcoal and activated carbon has been identified as mainly effective for treating the wastewater strategies. Machine learning (ML) approaches were used to create generalized computation methods for tetracycline (TC) and sulfamethoxazole (SMX) adsorption in CBMs in this investigation. In the existing system, random forest and ANN methods were used for TC and SMX for predicting the quantities of antibiotics in the CBMs.

3. Proposed Methodologies

To predict the adsorption quantity of antibiotics in carbon-based materials, the proposed method uses Gaussian process regression, active learning (AL), and ANN. It evaluates the adsorption in tetracycline (TC) and sulfamethoxazole (SMX) in CBMs. Also, it predicts the various properties according to their molecular organization, practical principles, and also the values of TC and SMX. The comparative

relevance and biased dependency graph assessment may lead to reasonable uses of CBMs for antimicrobial treating wastewater, whereas reliable ML prediction methods with generalisation capability are helpful for constructing effective CBMs with little practical testing. Figure 1 shows the architecture of the proposed system.

Figure 1 shows the overall process of the proposed system. Initially, it collects the data from for antibiotic adsorption and then, it is compared with the physicochemical property of CBMs. The data is preprocessed with z -score standardization, and then, the machine learning methods such as Gaussian process regression (GPR), active learning (AL), and ANN were used.

3.1. CBM-Based Data Collection. The evaluation information for tetracycline (TC) and sulfamethoxazole (SMX) adsorption on carbon-based materials (CBMs) was gathered from an entirety of 40 technical journals published in the final place. The additional information summarises the comprehensive suggestion for related TC and SMX adsorption which is taken from article [24]. In research article, lists of molecular arrangement and physicochemical parameters of TC and SMX are studied. With Plot Digitizer 2.6.8, the relevant data of CBM characteristics and associated antibiotic adsorption ability were straightly obtained from tables or extracted from adsorption isotherms. There are many CBMs collected in this process. Such waste materials are processed with thermal decomposition and/or mechanochemical grinding, resulting in valuation porosity CBMs that can be used as an expense and ecologically friendly resource recovery approach.

For predicting the quantity of adsorption in CBM, several properties are used to evaluate the amount of adsorption in TC and SMX. The overall percentage of carbon (C, wt.%), mass ratio of hydrogen to carbon (H/C), mass ratio of oxygen to carbon (O/C), mass ratio of nitrogen and oxygen to carbon $[(O + N)/C]$ (depicting adsorbent polarisation), ash substance (ash, wt.%), Brunauer-Emmett-Teller size distribution (SBET, m^2/g), and figure of low potential (pHpz) [1]. It includes the temperature and pH of solution.

3.2. Comparing the Properties of Physicochemical CBMs. The physicochemical characteristics of CBMs that were evaluated between some of the six categories of CBMs by quantitative tests had a substantial impact on antibiotic adsorption ability on CBMs. The pattern of H/C versus O/C mass fraction, which reflected the level of carbon and hydrophilic of CBMs, was visualised using a Van Krevelen chart. H/C molecular proportions under 0.3 indicate a substantially carbonised architecture inside the CBM architecture, whereas H/C molecular proportions over 0.7 indicate an incompletely carbonised organization. The carbonization level and mineral composition of various CBMs too were evaluated using the C wt. percent and ash wt. percent. Boxplots were used to obtain the probability parameters of SBET and pH pzc of several CBMs. The simple correlations among any pair AC and BC parameters too were investigated. The Pearson correlation coefficient was used to create the graph.

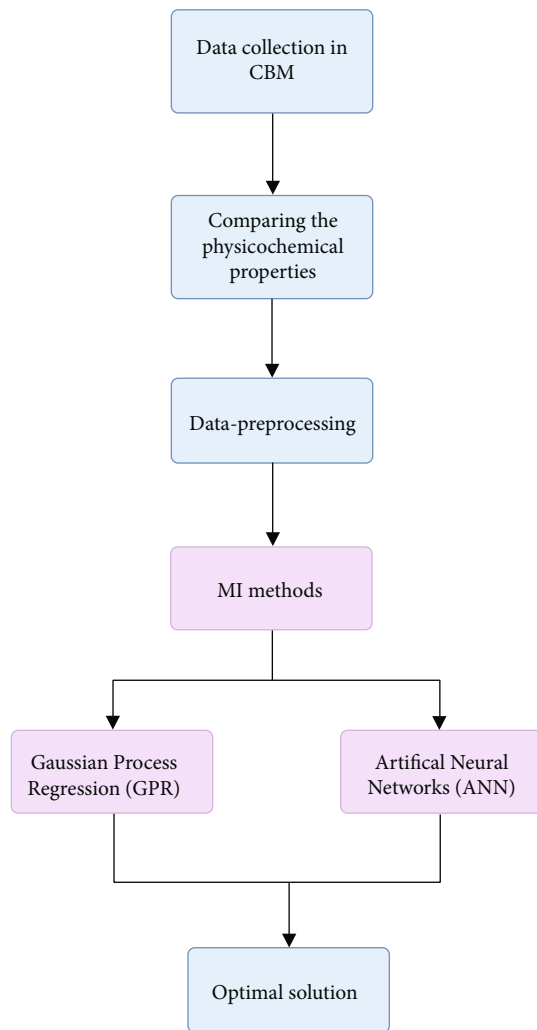


FIGURE 1: Architecture of the proposed system.

3.3. Preprocessing of Data and Machine Learning Methods. While developing any machine learning algorithm, all information on contextual variables as inputs was already adjusted with Z-score standardisation by

$$X_I^* = \frac{(X_I - \mu)}{\sigma}. \quad (1)$$

Here, X_I^* represents the normalization range and X_I represented as the new standards of input variables, where μ and σ are denoted as mean and standard deviation for calculating the variables.

The biological neural network, which comprised of a nonlinear mapping structure with an input layer, hidden layer(s), and output layer, was the inspiration for the ANN method. A feed forward error-back propagation strategy was used to fit the classifier. The signals (e.g., sigmoid and ReLu) and the connection factors among levels have been used to generate the node values in the output layer that were then matched to the goal result. The parameters of the perceptron were tweaked till the difference among anticipated and definite principles was as little as possible.

The records were being arbitrarily separated between trained testing data groupings with just an 80:20 ratio, having various CBM types being good dispersion in each category. For comparing the estimated results from each ML model, the very same divides were utilized. With the learning algorithm, fivefold cross-validation has been used to pick the best high energy, whereas the leftover test dataset was utilized to evaluate the performance of the model as a social approval.

$$R^2 - 1 = \frac{\sum_{I=1}^N (y_I^{\text{exp}} - y_I^{\text{prediction}})^2}{\sum_{I=1}^N (y_I^{\text{exp}} - y_{\text{ave}}^{\text{exp}})^2}, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{I=1}^N (y_I^{\text{exp}} - y_I^{\text{prediction}})^2}.$$

Here, y_I^{exp} and $y_I^{\text{prediction}}$ are the predicted standards and $y_{\text{ave}}^{\text{exp}}$ is represented as definite tentative principles.

3.4. Proposed Gaussian Process Regression for Antibiotic Adsorption in CBMs. The method works cleverly and chooses the adsorption imitation, which is worked and used for training the exact Gaussian process (GP) substitute replica. It is a one of the nonparametric ML methods. It helps to process the functions $f(X)$, and $\mu(X_I)$ is defined as a mean, and $Kv(X, X')$ is described as a covariance of the function.

$$f(X) \sim n\left(\mu(X), Kv(X, X')\right). \quad (3)$$

There are numerous options for $Kv(X, X')$. The reasonable cubic core was chosen since it was previously being utilized to represent adsorption pressure in CBMs.

$$Kv(X, X') = \left(1 + \frac{D(X, X')^2}{2AL^2}\right)^{-\alpha}. \quad (4)$$

Here, $D(X, X')^2$ is defined as a Euclidean distance among the X and X' ; here, L represented as kernel's size of length, and α is defined as the size combination constraint. For discovering the force and warmth state rapidly and precisely for the adsorption progress, Greedy method is used to discover the new break. GRP prediction uses some iteration.

$$X_{n+1} \text{ Argmax}_{X \in \mathcal{X}} \sigma_N^2(X). \quad (5)$$

Here, $\sigma_N^2(X)$ represented as a threshold; the GP is repeated until the new function formed. We ran an approximation experiment together for TC and SMX at the lower area (10-6 to 1 bar) when using this regularly aligned matrix requirements for evaluating and developing the AL system. We used 50 Cartesian coordinates spread inside the log likelihood for this experiment to see how well the complete GP regression performed following AL. This extrapolation test

was only done for the low-pressure region, and the higher high temperature was maintained similar as the Xtest.

This must have been done to develop an atmosphere where a user may assess overall effectiveness of a finalized AL fit theory that was absolutely oblivious to an augmentation testing data. Inside the findings section, overall AL efficiency is provided for both AL original test sets (i.e., on Xtest) as well as the low-flow extrapolation testing. In addition, to produce the depth data, all trials were subjected to a series of GCMC simulations. Within next portions, the specifications of GCMC simulation ambiguity (GCMC) for both Xtest and the low-flow extrapolation testing were provided in the corresponding statistics.

3.4.1. *Calculation of Error.* Error calculation has three following criteria in AL structure:

(1) Relative error for GP-predicted

This would be the proportion of GP-predicted ambiguity to GP-predicted adsorption at a given site. Keep in mind that the goal of an AL method would be to keep the GP-predicted mistake percentage within that range.

$$\text{Error rate of GP in\%} = \frac{\sigma_{\text{GP}} \cdot \text{prediction}(X)}{y_{\text{GP-prediction}(X)}} \times 100. \quad (6)$$

(2) Relative error

Thus, the percentage of such dissimilarity between both the GP-predicted adsorption and the GCMC-calculated regression coefficient adsorption is calculated.

$$\text{Relative error in\%} = \left| \frac{y_{\text{GP-prediction}(X)} y_{\text{Gcmc}(X)}}{y_{\text{Gcmc}(X)}} \right| \times 100. \quad (7)$$

(3) Mean for relative error (MRE)

This is derived as the average comparative inaccuracy throughout all AL iterations. We contrast the mistake towards the greatest GP absolute errors to determine the AL protocol's rate of junction. MRE also acts as a measure to gauge the efficiency of an AL representation because it combines GP-predicted adsorption with underlying data depending on GCMC simulation.

$$\text{MRE in\%} = \left(\sum_I^n \left| \frac{y_{\text{GP-prediction}(X_I)} y_{\text{Gcmc}(X_I)}}{y_{\text{Gcmc}(X_I)}} \right| \right) \times \frac{100}{N}. \quad (8)$$

4. Result Analysis

The proposed Gaussian process regression (GPR), active learning (AL), and Artificial Neural Network (ANN) help to predict the antibiotic adsorption in carbon-based mate-

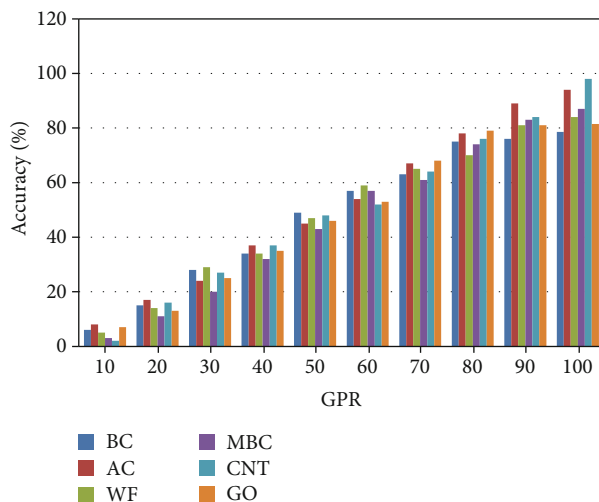


FIGURE 2: Accuracy of CBMs.

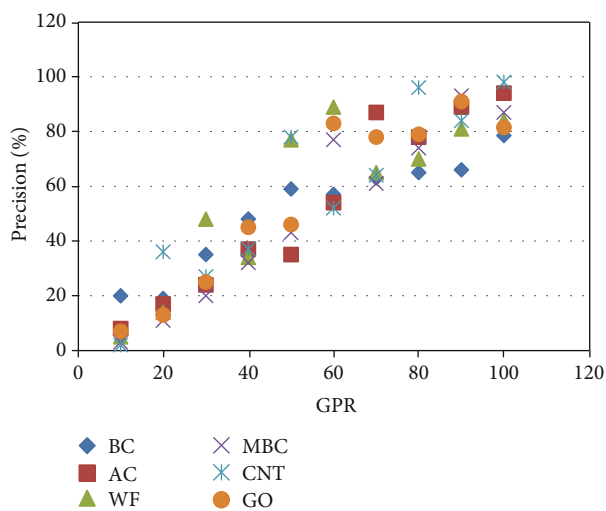


FIGURE 3: Precision rate of CBMs.

rials (CBMs). The proposed work uses several parameters such as biochar (BC), activated carbon (AC), waste feed (WF), magnetic modified biochar (MBC), carbon nanotube (CNT), and graphene oxide (GO). It measures the accuracy, recall, precision rate, false-positive rate (FPR), and false-negative rate (FNR).

4.1. *Accuracy.* It is used to evaluate the classification of carbon levels in the CBMs accurately. It uses biochar (BC, 65 items), activated carbon (AC, 65 items), waste feed (WF, 4 items), magnetic modified biochar (MBC, 12 items), carbon nanotube (CNT, 3 items), and graphene oxide (GO, 2 items).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100. \quad (9)$$

Figure 2 shows that the proposed method classifies the data accurately; it classifies biochar 78.55% and the other materials also classified better.

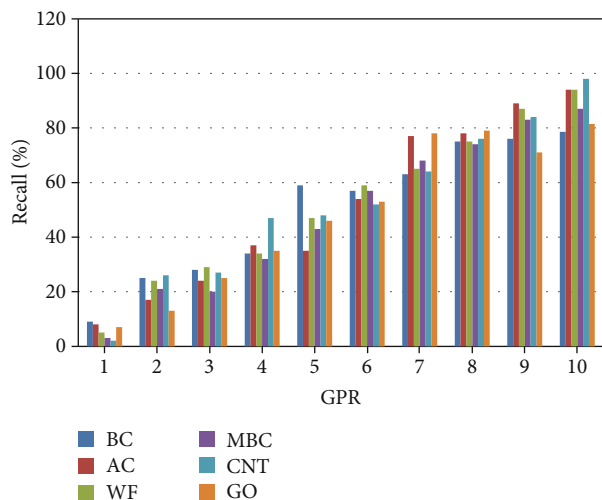


FIGURE 4: Recalls of CBMs.

4.2. *Precision*. It precisely classifies the CBMs such as biochar (BC, 65 items), activated carbon (AC, 65 items), waste feed (WF, 4 items), magnetic modified biochar (MBC, 12 items), carbon nanotube (CNT, 3 items), and graphene oxide (GO, 2 items) in the dataset, given by the following equation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100. \quad (10)$$

Figure 3 shows the precision rate of carbon-based materials. The proposed GRP outperforms better compared with the existing random forest methods.

4.3. *Recall*. It correctly classifies the positive values from the dataset. It uses various parameters for the classification. The following equation was used for evaluating the values.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (11)$$

Figure 4 shows the recall rate of the proposed method. The proposed method outperforms well.

4.4. *False-Positive Rate (FPR)*. It is used to evaluate the ratio between correctly identified CBMs to wrongly identified CBMs using

$$\text{FPR} = 1 - \text{specificity}. \quad (12)$$

4.5. *False-Negative Rate (FNR)*. It is used to evaluate the positive proportion value but negative CBMs value is identified using

$$\text{FNR} = 1 - \text{sensitivity}. \quad (13)$$

5. Conclusion

Antibiotic adsorption on carbon-based materials (CBMs) such as charcoal and activated carbon has been identified as mainly effective for treating the wastewater strategies. In

the existing system, random forest and ANN methods were used for TC and SMX for predicting the quantities of antibiotics in the CBMs. For reducing the antibiotics from the industrial wastewater, the broadcast efforts of the experiments are a little complicated. In the proposed method, Gaussian process regression (GPR), active learning (AL), and ANN are used for predicting the antibiotic levels in the industrial wastewater. The proposed work uses several parameters such as biochar (BC), activated carbon (AC), waste feed (WF), magnetic modified biochar (MBC), carbon nanotube (CNT), and graphene oxide (GO). It measures the accuracy, recall, precision rate, false-positive rate (FPR), and false-negative rate (FNR). It predicts accurately and also helps to remove the antibiotics from the wastewater; also, it reduces the experimental screening.

Data Availability

All the required data is available in the article itself.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number RGP 2/209/42. We deeply acknowledge Taif University for supporting this research through Taif University Researchers Supporting Project Number TURSP-2020/328, Taif University, Taif, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project Number R-2021-314.

References

- [1] X. Zhu, Z. Wan, D. C. W. Tsang et al., "Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption," *Chemical Engineering Journal*, vol. 406, article 126782, 2021.
- [2] Q. Q. Zhang, G. G. Ying, C. G. Pan, Y. S. Liu, and J. L. Zhao, "Comprehensive evaluation of antibiotics emission and fate in the river basins of China: source analysis, multi-media modeling, and linkage to bacterial resistance," *Environmental Science & Technology*, vol. 49, no. 11, pp. 6772–6782, 2015.
- [3] J. Sun, Q. Zeng, D. C. Tsang, L. Zhu, and X. Li, "Antibiotics in the agricultural soils from the Yangtze River Delta, China," *Chemosphere*, vol. 189, pp. 301–308, 2017.
- [4] O. M. Rodriguez-Narvaez, J. M. Peralta-Hernandez, A. Goonetilleke, and E. R. Bandala, "Treatment technologies for emerging contaminants in water: a review," *Chemical Engineering Journal*, vol. 323, pp. 361–380, 2017.
- [5] B. L. Phoon, C. C. Ong, M. S. Mohamed Saheed et al., "Conventional and emerging technologies for removal of antibiotics from wastewater," *Journal of Hazardous Materials*, vol. 400, article 122961, 2020.
- [6] Y. Sun, D.-W. Cho, N. J. Graham et al., "Degradation of antibiotics by modified vacuum-UV based processes: mechanistic

- consequences of H_2O_2 and $\text{K}_2\text{S}_2\text{O}_8$ in the presence of halide ions,” *Science of The Total Environment*, vol. 664, pp. 312–321, 2019.
- [7] K. Premarathna, A. U. Rajapaksha, B. Sarkar et al., “Biochar-based engineered composites for sorptive decontamination of water: a review,” *Chemical Engineering Journal*, vol. 372, pp. 536–550, 2019.
- [8] M. R. Berber, “Current advances of polymer composites for water treatment and desalination,” *Journal of Chemistry*, vol. 2020, Article ID 7608423, 19 pages, 2020.
- [9] S. Zhang, B. Li, X. Wang et al., “Recent developments of two-dimensional graphene-based composites in visible- light photocatalysis for eliminating persistent organic pollutants from wastewater,” *Chemical Engineering Journal*, vol. 390, article 124642, 2020.
- [10] W. Xiong, G. Zeng, Z. Yang et al., “Adsorption of tetracycline antibiotics from aqueous solutions on nanocomposite multi-walled carbon nanotube functionalized MIL-53(Fe) as new adsorbent,” *International Journal of Pharmacy and Technology*, vol. 627, pp. 235–244, 2018.
- [11] L. Yi, L. Zuo, C. Wei et al., “Enhanced adsorption of bisphenol A, tylosin, and tetracycline from aqueous solution to nitrogen-doped multiwall carbon nanotubes via cation- π and π - π electron-donor-acceptor (EDA) interactions,” *Science of The Total Environment*, vol. 719, article 137389, 2020.
- [12] K. Mukherjee, A. W. Dowling, and Y. Colón, “Sequential design of adsorption simulations in metal-organic frameworks,” <https://arxiv.org/abs/2110.00069>.
- [13] C. Molnar, “Interpretable machine learning,” in *A Guide for Making Black Box Models Explainable* Lulu. com.
- [14] X. Zhu, D. C. W. Tsang, L. Wang et al., “Machine learning exploration of the critical factors for CO_2 adsorption capacity on porous carbon materials at different pressures,” *Journal of Cleaner Production*, vol. 273, article 122915, 2020.
- [15] X. Zhu, X. Wang, and Y. S. Ok, “The application of machine learning methods for prediction of metal sorption onto biochars,” *Journal of Hazardous Materials*, vol. 378, article 120727, 2019.
- [16] M. Kumar, X. Xiong, Z. Wan et al., “Ball milling as a mechanochemical technology for fabrication of novel biochar nanomaterials,” *Bioresource Technology*, vol. 312, article 123613, 2020.
- [17] Y. Wang and R. Liu, “Comparison of characteristics of twenty-one types of biochar and their ability to remove multi-heavy metals and methylene blue in solution,” *Fuel Processing Technology*, vol. 160, pp. 55–63, 2017.
- [18] I. López-Cano, M. Cayuela, C. Mondini, C. Takaya, A. Ross, and M. Sánchez-Monedero, “Suitability of different agricultural and urban organic wastes as feedstocks for the production of biochar—part 1: physicochemical characterisation,” *Sustainability*, vol. 10, no. 7, 2018.
- [19] M. Fernandez and A. S. Barnard, “Geometrical Properties Can Predict CO_2 and N_2 Adsorption Performance of Metal-Organic Frameworks (MOFs) at Low Pressure,” *ACS Combinatorial Science*, vol. 18, no. 5, pp. 243–252, 2016.
- [20] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, and T. K. Woo, “Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO_2 Capture,” *The Journal of Physical Chemistry Letters*, vol. 5, no. 17, pp. 3056–3060, 2014.
- [21] B. J. Bucior, N. S. Bobbitt, T. Islamoglu et al., “Energy-based descriptors to rapidly predict hydrogen storage in metal-organic frameworks,” *Molecular Systems Design & Engineering*, vol. 4, no. 1, pp. 162–174, 2019.
- [22] A. Sturluson, M. T. Huynh, A. H. P. York, and C. M. Simon, “Eigencages: Learning a Latent Space of Porous Cage Molecules,” *ACS Central Science*, vol. 4, no. 12, pp. 1663–1676, 2018.
- [23] B. J. Befort, R. S. DeFever, G. M. Tow, A. W. Dowling, and E. J. Maginn, “Machine learning directed optimization of classical molecular modeling force fields,” 2021, <https://arxiv.org/abs/2103.03208>.
- [24] E. Uteva, R. S. Graham, R. D. Wilkinson, and R. J. Wheatley, “Active learning in Gaussian process interpolation of potential energy surfaces,” *The Journal of Chemical Physics*, vol. 149, no. 17, article 174114, 2018.
- [25] J. Vandermause, S. B. Torrisi, S. Batzner et al., “On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events,” *NPJ Computational Materials*, vol. 6, 2019.
- [26] J. E. Santos, M. Mehana, H. Wu et al., “Modeling Nanoconfinement Effects Using Active Learning,” *The Journal of Physical Chemistry C*, vol. 124, no. 40, pp. 22200–22211, 2020.
- [27] B. Kim, S. Lee, and J. Kim, “Inverse design of porous materials using artificial neural networks,” *Science Advances*, vol. 6, no. 1, 2020.
- [28] C. Peiris, S. R. Gunatilake, T. E. Mlsna, D. Mohan, and M. Vithanage, “Biochar based removal of antibiotic sulfonamides and tetracyclines in aquatic environments: A critical review,” *Bioresource Technology*, vol. 246, pp. 150–159, 2017.
- [29] Y. Shi, G. Liu, L. Wang, and H. Zhang, “Activated carbons derived from hydrothermal impregnation of sucrose with phosphoric acid: remarkable adsorbents for sulfamethoxazole removal,” *RSC Advances*, vol. 9, no. 31, pp. 17841–17851, 2019.
- [30] A. Gopalan, B. Bucior, N. Bobbitt, and R. Snurr, “Prediction of hydrogen adsorption in nanoporous materials from the energy distribution of adsorption sites,” *Molecular Physics*, vol. 117, no. 23–24, pp. 3683–3694, 2019.
- [31] D. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Mathematical Programming*, vol. 45, no. 1–3, pp. 503–528, 1989.
- [32] F. Pedregosa, G. Varoquaux, A. Gramfort et al., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.