



RESEARCH

Open Access



Machine learning algorithm and neural network architecture for optimization of pharmaceutical and drug manufacturing industrial effluent treatment using activated carbon derived from breadfruit (*Treculia africana*)

Prosper Eguono Ovuoraye^{1,2*} , Victor Ifeanyi Ugonabo¹, Endrit Fetahi^{3*}, Ahmad Chowdhury⁴, Mohammad Abdullah Tahir⁵, Chinenye Adaobi Igwegbe^{1,6}  and Mohammad Hadi Dehghani^{7,8,9}

*Correspondence:
pe.ovuoraye@gmail.com; endrit.fetahi@uni-prizren.com

¹ Department of Chemical Engineering, Nnamdi Azikiwe University, P.M.B. 5025, Awka 420218, Nigeria

³ Faculty of Computer Science, University of Prizren "Ukshin Hoti", Prizren, Kosovo
Full list of author information is available at the end of the article

Abstract

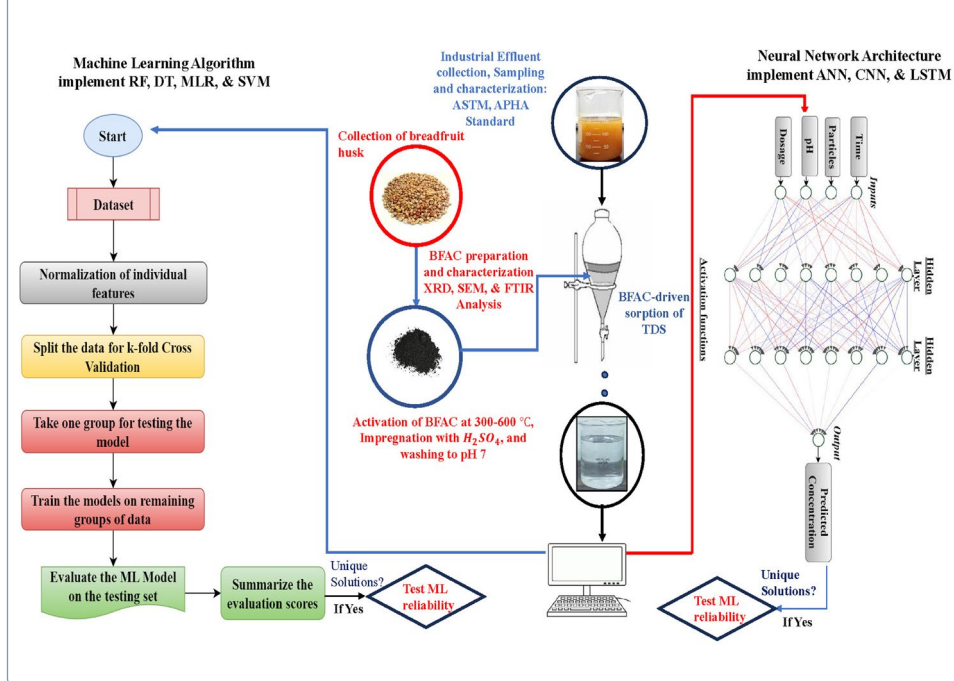
In a recent development, attention has shifted to the application of artificial intelligence for the optimization of wastewater treatment processes. This research compared the performances of the machine learning (ML) model: random forest, decision tree, support vector machine, artificial neural network, convolutional neural network, long-short term memory, and multiple linear regressors for optimization in effluent treatment. The training, testing, and validation datasets were obtained via the design of an experiment conducted on the removal of total dissolved solids (TDS) from pharmaceutical effluent. The breadfruit-activated carbon (BFAC) adsorbent was characterized using scanning electron microscopy and X-ray diffraction techniques. The predictive capacity of an ML algorithm, and neural network architecture implemented to optimize the treatment process using statistical metrics. The results showed that $MSE \leq 1.68$, $MAE \leq 0.95$, and $\text{predicted-}R^2 \geq 0.9035$ were recorded across all ML. The ML output with minimum error functions that satisfied the criterion for clean discharge was adopted. The predicted optimum conditions correspond to BFAC dosage, contact time, particle size, and pH of 2.5 mg/L, 10 min, 0.60 mm, and 6, respectively. The optimum transcends to a reduction in TDS concentration from 450 mg/L to a residual ≤ 40 mg/L and corresponds to 90% removal efficiency, indicating ± 1.01 standard deviation from the actual observation practicable. The findings established the ML model outperformed the neural network architecture and affirmed validation for the optimization of the adsorption treatment in the pharmaceutical effluent domain. Results demonstrated the reliability of the selected ML algorithm and the feasibility of BFAC for use in broad-scale effluent treatment.

Highlights

- ML/RSM/artificial intelligence for adsorptive uptake interpretation of TDS from pharmaceutical effluent
- The performance of various machine learning models (RF, DT, SVM, ANN, CNN, LSTM, and MLR) for effluent treatment optimization
- Removal of TDS from pharmaceutical effluent using breadfruit-activated carbon as an adsorbent

Keywords: Pharmaceutical and drug manufacturing effluent, RSM optimization, Breadfruit activated carbon, Machine learning algorithm, Convolutional and artificial neural network

Graphical Abstract



Introduction

The wake of the COVID-19 pandemic has seen a wide range of improvements and development leading to the production of vaccines and drugs for health care, sanitizing materials, and veterinary chemicals amongst other biological compounds with the need to maintain global health demands. Pharmaceutical and drug manufacturing industry wastewater is not just one of the main causes of irreversible damage to the environmental balances [1] but also serves as contributing source to the depletion of freshwater reserves and groundwater across the globe. The presence of a high concentration of TDS in pharmaceutical effluent discharge is toxic to aquatic life and causes digestive disorders and cancer in humans amongst other aesthetic problems [2]. High levels of total dissolved solids (TDS) over a long period in high concentrations will expose the body to chemicals that can cause chronic health conditions like kidney and liver infections and even cancer [3, 4]. Groundwater can become saline on contamination with high TDS concentration and may become problematic to agriculture

and irrigation [2, 5]. Consequently, TDS content in pharmaceutical wastewater must be properly reduced to acceptable levels in the industrial effluent before its discharge into the water bodies [4]. The adsorption process has been considered an effective method for the removal of TDS from pharmaceutical effluent to guarantee environmental sustainability [6].

The applicability of activated carbon derived from breadfruit shells for TDS removal from wastewater has been reported in the literature [7]. However, the wastewater treatment process must be optimized to ascertain the cost economics, dosage consumption, and feasibility of the treatment process [6, 8–10]. The application of artificial intelligence which cuts across frequently used ML algorithms and neural network architectures including support vector machine (SVM), random forest (RF), decision tree (DT), and multiple linear regression (MLR) has been reported in previous works of literature [6]. The ANN and RF model has been frequently applied in the remediation of water [9, 10]. The effective application of the long short-term memory (LSTM) network has also been reported as effective in optimizing the water treatment process [8]. The SVM have been reported to record 99.4% prediction accuracy following the treatment of effluent wastewater [10, 11]. However, the integrations of deep learning architectures such as long short-term memory networks (LSTN), convolutional neural networks (CNN), and decision trees (DT) for optimizing the removal of TDS in wastewater treatment operations are limited in literature [6, 11]. The reliability of the ML algorithm and neural network architecture for predicting residual TDS in pharmaceutical effluent is not fully explored, although the current trend in adsorption treatment is driven towards the application of low-cost biomass produced for adsorption processes. The selectivity of activated carbon derived from breadfruit husk has been explored, and detailed ML model comparison, selection, and application involving choice ML algorithm and neural network architecture where breadfruit sorbent is the adsorbent are yet to be investigated. We have previously examined the performance of the RSM approach in modeling and optimizing the removal of TDS from pharmaceutical effluent reported in the literature [7]. There were a few advantages of the RSM in the prediction of the treatment processes taking into consideration the cost-economics of managing optimization solutions with limited datasets compared to the ML model employed in this study. Previous findings established that the RSM approach predicts optimal that create a better understanding of the effect of independent factors on the response [7, 12, 13]. The methodology was simpler in approach [13, 14], requiring less computation power [13, 14], and model performances were largely dependent on the level of control administered in the design space [12, 14]. Another advantage we explored for the RSM can accommodate fewer input parameters than ML.

Consequently, the current research investigated the prediction capacities of RF, DT, MLR, SVM, ANN, CNN, and LSTM for the optimization of pharmaceutical and drug manufacturing industrial effluent treatment process. The output of the ML algorithm and neural network architecture was designed to minimize the residual TDS concentration and determine optimum operating conditions following the effluent treatment process. The objective of employing the ML algorithm and neural network architecture is to compare the performances of the various AI model predictions of the optimum

conditions and to identify the most suitable approach to optimization of the effluent treatment process.

In this work, the pharmaceutical and drug manufacturing industrial effluent characteristics were determined following ASTM standards. The BFAC used for the treatment of the effluent was characterized using FTIR, SEM, and XRD techniques. The adsorptive uptake of TDS onto BFAC was monitored via batch-scale experimentation to obtain the research data stored in a repository for the ML optimization procedures. The reliability of the AI models was validated following response surface methodology (RSM), and the outputs were compared based on statistical error functions and following established standards for effluent discharge. By leveraging the power of the selected ML model, we aimed to enhance our understanding of the predictive capacities of the ML algorithm and neural network architecture in this domain. The findings will provide information on the feasibility of the biochar on specific contaminant removal and provide direction for further research on the application of the selected ML models for optimization of pilot and industrial scale pharmaceutical effluent water treatment processes.

Methods

Field sampling

In this research experiment, the pharmaceutical and drug manufacturing industrial effluent sample was collected from outlet drain of an industrial facility located at Awka, southeastern Nigeria. The sample was preserved and characterized following ASTM standards for water examination. The wastewater sampling and analysis were carried out to determine its initial properties before formal analysis for datasets was obtained via batch mode adsorption design of experiment. Adsorption experiments were carried out on treatment of fresh industrial effluent sample using locally developed green biochar. The procedures involve sample collection, batch mode treatment of sample, data collection, data training and testing, ML prediction and reliability, optimization, and validation analysis. All reagents used were of analytical grade and complied with the guidelines recommended by the Department of Chemical Engineering, Nnamdi Azikiwe University Awka, Nigeria [15].

The African breadfruit (*Treculia africana*) samples used for the preparation of activated carbon were obtained from local farmers in Agwuaka in Anambra State, Nigeria. The samples of breadfruit husks were separated from the fruit and dried in an oven set at 80 °C until a constant mass was recorded. The dried husks were pyrolyzed in earthenware and placed inside a furnace set at a temperature of 300–600 °C. The carbonization process was monitored for 2-h holding time; the sample was allowed to cool. The corresponding weights before and after the pyrolysis were recorded to determine the weight loss. The pulverized samples were crushed into powdered form using a mechanical grinder to increase their surface area. The sample was soaked in a 60 wt% solution of H₂SO₄ in a ratio of 1:1. The impregnated sample was left at room temperature for 24 h. After impregnation, the excess solution was filtered off, and the sample was allowed to stabilize by washing with distilled water until the pH reached approximately 7. After stabilization, the samples were dried in the oven at a temperature of 100 °C for 2 h and allowed to cool at room temperature and stored in a desiccator. A 40% yield of the initial mass was obtained before the sieved sample was passed through screens to

retain varying fractions of particle sizes from 0.3 to 0.15 mm used for the experiment. The functional characteristics of the BFAC produced and used for the treatments of the pharmaceutical wastewater were determined by FTIR using diffuse reflectance infrared Fourier transform spectrometer (DRIFT, PerkinElmer, model Spectrum One, USA). The SEM examination and XRD analysis were also performed on the BFAC to determine surface morphology and physiochemical constituents present in the biochar. The summary of the experimental procedure which is illustrated in Fig. 1 shows the schematic illustration for the experimental and data train-test methodology.

Data collection procedure

To obtain data samples for experimental ML training, testing, and validation, batch mode adsorption experimentation on the industrial effluent sample using the breadfruit activated carbon (BFAC) developed. The breadfruit was activated to increase surface area and pyrolyzed in an oven. The biomass was impregnated using 2 M Conc. H₂SO₄ and thoroughly screened to obtain varying particle sizes (0.15–1.0 mm) to be used for the treatment of pharmaceutical effluent. Scanning electron microscopy (SEM) was done using a scanning electron probe micro analyzer (model Jeol-JXA 840A, Japan), and X-ray diffraction analysis (XRD) on the BFAC sample was done using a diffraction analyzer (XPRT). Fourier-transform infrared spectroscopy (FT-IR) of the adsorbent was done using diffuse reflectance infrared Fourier transform spectrometer (DRIFT, PerkinElmer, model Spectrum One, USA), in the range of 400–4000 cm⁻¹ to determine the functional groups present in the BFAC.

The experiment carried out was conducted at varying effects of pH, BFAC dosage, contact time, and particle sizes of the adsorbent for the adsorption treatments of the pharmaceutical wastewater. The adsorption experiments were conducted in batch mode using dosage (0.5–1.5 mg/L). The initial and residual TDS concentrations in the water medium were analyzed at varying contact times (10–60 min) using a UV–visible recording spectrophotometer (Shimadzu Model, CE-1021-UK). The pH (4–10) of the solution was measured using a MIT65 pH meter. The water samples were introduced into five different tubes fitted with cottonwool at one end to prevent adsorbent leakage and to aid in easy recovery of the spent adsorbent. The adsorption treatment in the column was monitored at pH 4–10, and the filtrates were collected into separate conical flasks at the burette point of discharge. The treated samples from each column outlet were

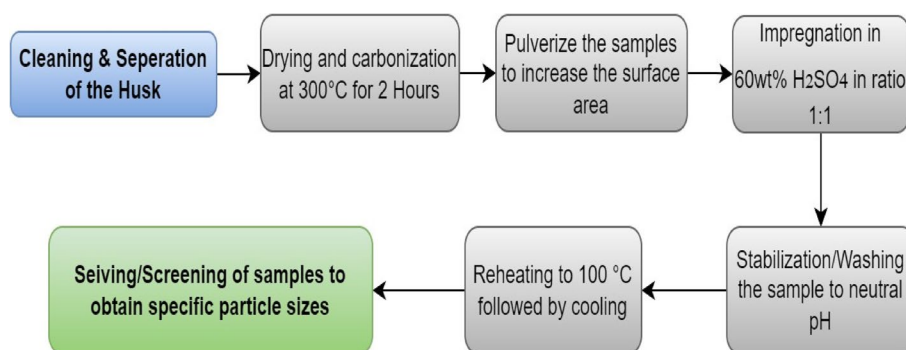


Fig. 1 Schematic illustration of the synthesis of the breadfruit activated carbon (BFAC) adsorbent

analyzed for absorbance using the photoelectric colorimeter, and pH was measured using a pH meter. The procedure was monitored at varying contact times of 10, 20, 30, 40, and 50 min. The procedure was repeated for other particle sizes (0.5, 1.0, 1.5, 2.0, and 2.5 mm). The pH was varied from 4 to 10. The resulting changes in TDS concentration data were recorded for various particle sizes, pH, dosage, and contact time in the form of a CCD matrix. The data obtained were stored in a repository for screening, cleaning, training, and testing following the implementation of a machine learning algorithm and artificial neural network architecture used for modeling and optimization of the treatment process Fig. 2.

Development of machine learning algorithm and neural network architecture

In this research work, the data sampling consisted of a set of 625 experimental runs obtained from the design of the experiment conducted following the adsorption treatment procedure described in the “[Data collection procedure](#)” section. The optimization setup involves the utilization of various ML models and neural network which includes the following: random forest model (RF), support vector machine (SVM), multiple linear regression (MLR), decision tree model (DTM), artificial neural network (ANN), convolutional neural network (CNN), and long short-term memory (LSTM) network, respectively. These models were selected following their proven effectiveness in capturing complex patterns [2, 6, 8], and their relationship with capture data underreported in the literature [2, 5, 9, 16]. The optimization setup experiment began with identifying candidate variants, imported libraries and modules for nonlinear regression, ML algorithm, and neural network architecture using the scikit-learn library and kernel with Python program [10]. The Python code for the machine learning was executed using a 64-bit system requiring a Core i5 processor with 8 GB of RAM. The data matrix was loaded from the repository file, followed by the sequence of sorting and splitting of data into subsets (testing and training sets). The subsets were employed separately for training ML and for testing the model predictions, whereby several rounds of cross-validation of the dataset will be employed. This approach will minimize overfit and ensure the selected model generalizes the datasets rather than memorizing the data. A grid search method was employed to execute the ML optimization taking the factors (pH, particle size, dosage, and contact time) from the treatment process as inputs, with the set objectives to minimize the response (residual TDS concentration in the effluent), and to predict the optimum operating conditions as response parameter outputs. The performance output of each ML model and neural network was compared by employing the criterion of minimal error functions (MSE, MAE, RMSE, and MAD) determined through a grid search. This research protocol is based on the significance of the selected model validation with the output from the RSM validation matrix using a nonlinear design space and statistical evaluation metrics (standard deviation and predicted R^2). This approach will also guarantee a new dataset that will be used to quantify the selected ML performance accuracy and degree of precision and identify areas for model improvement.

Response surface design of experiment on validation dataset

Response surface methodology (RSM) is an established statistical optimization methodology in experimental design [12, 13]. It is necessary to validate the predictive capacity of

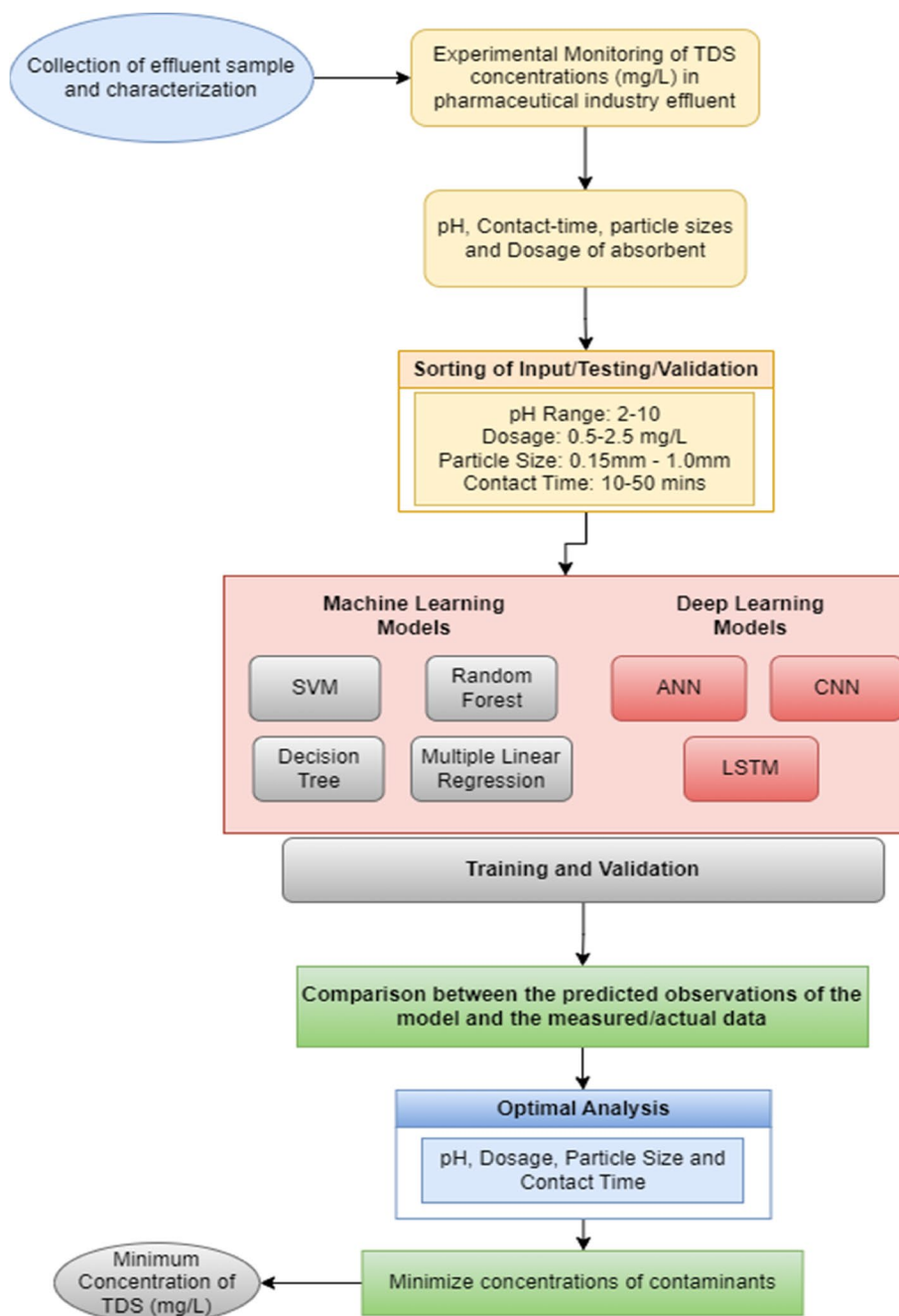


Fig. 2 Sequence of workflow of the experimental methodology and the machine learning (ML) optimization process

the selected ML algorithm to ascertain for reliability and dynamics of the selected model for this domain. We believe it is necessary to monitor and update the selected ML for consistency and reliability as new datasets become available. In this research, the new validation data sets used to test the reliability of the selected ML algorithm were based on the central composite design (CCD) of the RSM. Bearing in mind the running cost of experimentation, the CCD was developed based on 20 experimental runs conducted

at pH (6–8) which satisfied the criterion for industrial effluent discharge [9], at varying particle size (0.5–2.5), time (10–50 min), and adsorbent dosage (0.5–1.0 mg/L) to minimize operation cost. The CCD space is expressed in terms of the actual values of the experimental factor under investigation as input. Table 1 shows the CCD validation matrix in terms of actual values. The validation set obtained from the ML algorithm will be compared with the predicted outputs from the RSM, and the finding will be used to attest to the reliability of the ML and subsequent interpretation of the optimization of the BFAC-driven adsorption treatment of pharmaceutical and drug manufacturing industrial effluent.

Results and discussion

Physiochemical composition of the industrial effluent

The physiochemical characteristic results of the pharmaceutical and drug manufacturing industry effluent are presented in Table 2. The characterization result shows that at an initial temperature of 27.1, the effluent appears brownish with a total solid (TS) concentration of 1100 mg/L. This characteristic is an indication of the likely presence of fluorescence and colored dissolved organic matter in the effluent. The high volume of TS confirms a strong presence of TDS (220 mg/L) and suspended solids (TSS) amounting to 880 mg/L in the industrial effluent. The concentrations of $TSS > 30$ mg/L and $TDS > 50$ mg/L are obvious indications that the contaminant levels do not satisfy the criteria for clean discharge [13]. However, further investigation into the chemical oxygen demand (COD) and biological oxygen demand (BOD) content of the industrial effluent indicates concentrations corresponding to 26.5 mg/L and 10 mg/L, respectively. The

Table 1 Experimental design variables in terms of actual and coded variables

Factor	Name	Units	Coded low	Coded high	Mean	Std. dev
A	Contact time	Minutes	- 1 ↔ 10.00	+ 1 ↔ 50.00	28.10	17.78
B	Dosage	mg/L	- 1 ↔ 0.50	+ 1 ↔ 2.50	1.50	0.8263
C	pH	-	- 1 ↔ 6.00	+ 1 ↔ 8.00	7.00	1.67
D	Particle size	mm	- 1 ↔ 0.15	+ 1 ↔ 1.00	0.5386	0.3634

Table 2 Pharmaceutical and drug manufacturing industrial effluent characteristics

Parameter	Unit	Method of determination	Measured value
pH		APHA 4500-H (APHA 1992)	4.92
Temperature		APHA 2550-A (APHA 1992)	27.1
Appearance	-	-	Brownish
Biochemical oxygen demand (BOD)	mg/L	APHA 5210-B (APHA 1992)	26.7
Chemical oxygen demand (COD)	mg/L	APHA 5220-D (APHA 1998)	10
Biodegradability index (BI)	-	-	2.56
Total solids (TS)	mg/L	APHA 2540 B (APHA 1998)	1100
Total suspended solids (TSS)	mg/L	APHA 2540 D (APHA 1998)	650
TDS	mg/L	APHA 2540 C (APHA 1998)	450
Electrical conductivity (EC)	μS/cm	APHA, 2510B (APHA 1998)	0.294
Nitrates	mg/L	APHA 4110B (APHA 1992)	50
Total iron	mg/L	APHA 3112B (APHA 1995)	1.49

measured values of the BOD-COD ratio yield a biodegradability index of 2.65 and confirmed the presence of a low amount of organic matter. This outcome proved that BOD/COD ratio is >0.1 , suggesting that the degree of biodegradation of the industrial effluent is only compromised [10, 16] and relatively insufficient to ensure stability. It can be concluded from the characterization result that the industrial effluent requires further treatment to drive suspended and dissolved solids to stability, rendering the colored dissolved organic matter susceptible to biological oxidation.

Characterization of the BFAC

The SEM image, IR spectroscopy, and X-ray diffraction results that describe the active sites and sorption potentials of the BFAC are shown in Fig. 3a–c. The characteristics of the BFAC prepared showed strong morphological and surface properties well adapted for the sorption of contaminants. The SEM image in 3a Fig. 3 shows a high surface with active sites and irregular surfaces with a network of pores well configured for the sorption of dissolved organics from the effluent medium [7, 14]. The SEM image at 500 magnifications shows irregular and heterogeneous surface morphology and complex network of pores. The FTIR characteristics in 3b of Fig. 3 also confirmed the presence of Si–O and Si–X at IR peaks 913.32 cm^{-1} , 1878.73 cm^{-1} , and 2351.50 cm^{-1} confirming the presence of quartz with C–H and O–H stretching [7] in the activated carbon structure of the biochar produced. Si–O, O–H, Si–O–Si, and S–O–X groups resulting from acid activation. The XRD in 3c in Fig. 3 shows diffraction scattering and interfacial planes at values of 2θ between 20.8 and 68.2° . Peaks at $2\theta = 20.8^\circ$, 26.6° , 42.4° , 50.1° , 59.9° , and 68.2° at $1.22 \leq d\text{-spacing} [\text{Å}] \leq 4.0$. The findings confirm the presence of silicates and metallic constituents well structure to attract negative charge ions and dissolved organic matter from the effluent medium. Interspatial planes between atomic lattices with varying peaks showed least presence of quartz and silicate mineral, and metal-halogen bond characterizes the morphology of BFAC.

Machine learning validation statistics and evaluation metrics

Table 3 shows the summary of the selected model validation metrics and statistical error function recorded across the ML algorithm in terms of the predicted R^2 , RMSE, MSE, MAD, and MAE. The MSE and MSE functions are a measure to differentiate the performances of the ML algorithm and the neural network architecture. Although both metrics are easy to interpret, they are both sensitive to outliers. The evaluation result in terms of mean-squared-error metrics shows that the ML models and neural networks yielded MSE output across all ML and supervised learning models following the order RF (1.03), DT (1.06), MLR (1.15), SVM (1.16), ANN (1.27), LSTM (1.29), and CNN (1.68). The result proved that minimum MSE was recorded across all models. Lower MSE is considered good; in this case, the ML predictions at MSE probably violated probability and assumptions of the interpretation of the optimal output. The higher MSE outputs ≥ 1.27 recorded for the ANN, LSTM, and CNN indicating a lower adequacy of prediction were associated with the neural network models probably due to noise interference [10, 14]. The RF model's predictions were more precise than the ANN, CNN, and LSTM suggesting that the neural network architectures fail

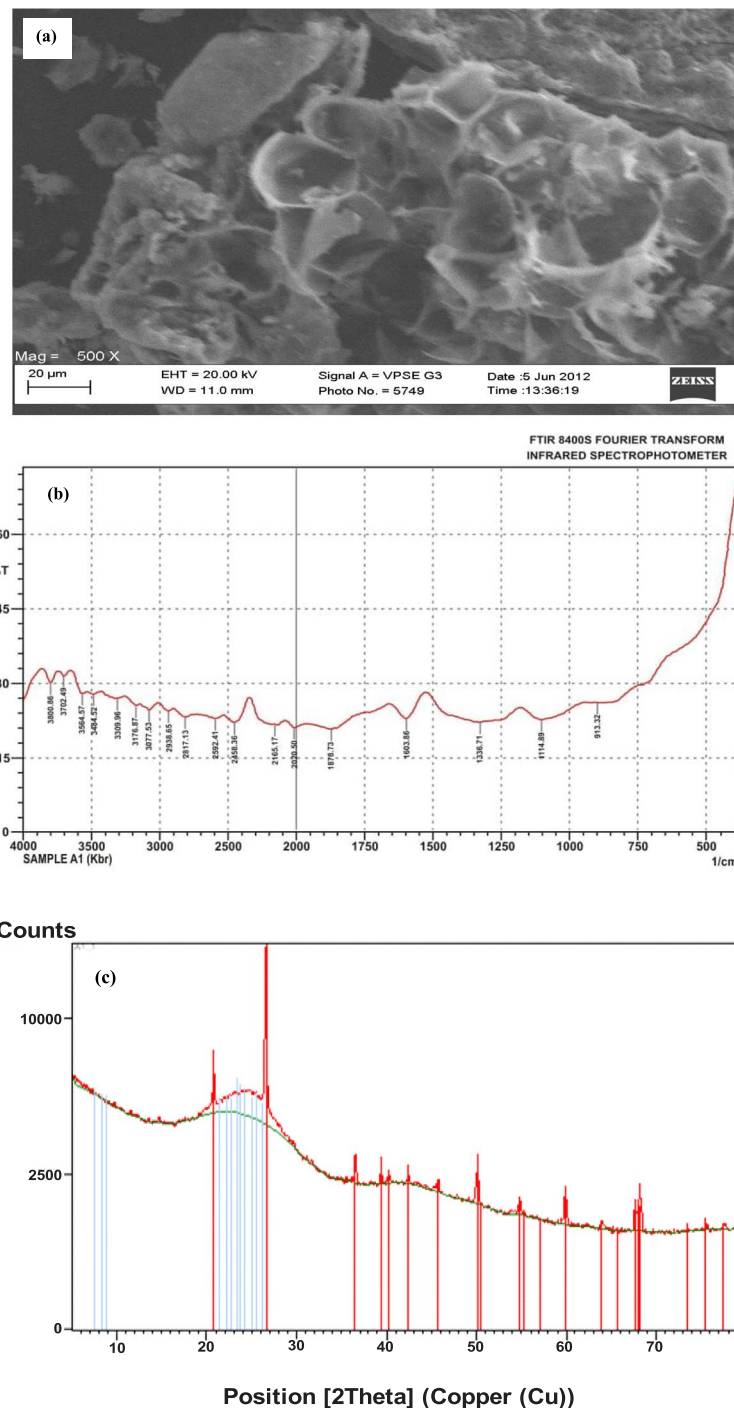


Fig. 3 Characterization results of breadfruit activated carbon adsorbent (BFAC) showing **a** SEM image, **b** FTIR spectroscopy, and **c** XRD profile of the adsorbent

to account for more important predictive features [8, 10, 14]. This was likely due to the computational complexity features and characteristics associated with neural network architecture compared to the simpler ML models [6, 17].

Table 3 Machine learning and deep learning model fit error statistical and evaluation metrics

	SVM	MLR	RFM	DTM	ANN	CNN	LSTM
MAE	0.8084	0.8080	0.7601	0.7692	0.7957	0.9142	0.9373
MSE	1.1612	1.1519	1.0272	1.0616	1.2732	1.6813	1.2905
RMSE	1.0776	1.0733	1.0135	1.0303	1.1283	1.2966	1.1360
Predicted R ²	0.9401	0.9406	0.9411	0.9391	0.9269	0.9035	0.9335
Mean ave	0.8084	0.8080	0.7601	0.7692	0.7952	0.9243	0.9448
Std. dev	1.0776	1.0733	1.0109	1.0286	1.1181	1.3086	1.0596

MSE output: RF > DT > MLR > SVM > ANN > LSTM > CNN

However, the prediction performance of the ML and neural network in terms of the RMSE outputs which showed the respective values of RF (1.01), DT (1.03), SVM (1.07), MLR (1.08), ANN (1.12), LSTM (1.14), and CNN (1.20) was recorded. The output confirmed an $RMSE \leq 1.20$ across all models under investigation. The result proved that the RF yielded the least RMSE output, while the highest RMSE was recorded with the CNN. The lower RMSE performance of the RF model to a collection of bootstrapping aggregation of multiple trees fits the data with lower variance compared to the single decision tree. The inferior error metrics of the neural network architectures were probably due to the complex interconnected nodes used to learn the pattern [14, 17] and capture the nonlinear relationship that might be neglected compared to the RF, DT, and other simpler models [5, 10]. The comparative analysis of the error functions across all models and network established that the order of significance of the optimization and interpretation of the adsorptive uptake of TDS from the industrial effluent is as follows:

RMSE output: RF > DT > SVM > MLR > ANN > LSTM > CNN

Although MAE is usually less sensitive to outliers, the evaluation metric will help treat the associated error outcome linearly [10, 18], while the MAD metric will function as a robust measure of the variability of the ML predictions taking into consideration the extreme values of the observed dataset [17, 18]. The respective MAE values recorded across the outputs of the ML algorithm and neural network architecture correspond to RF (0.76), DT (0.77), ANN (0.78), MLR (0.81), SVM (0.82), CNN (0.91), and LSTM (0.94). The output confirmed MAE function ≤ 0.94 , was recorded for the ML interpretation of the sorption process. The MAE values across all models are > 0 suggesting that overfit is minimal to noise ratio [17, 19]. A higher MAE output associated with the RF outputs indicates that the prediction and modeled assumptions are significant and well-suited for the optimization analysis compared to other ML models. The highest value of MAE recorded for the LSTM network which indicates a lesser prediction accuracy is associated with the modeled modified neural network. This outcome suggests the LSTM output deviates slightly from the actual observation.

MAE output: RF > DT > ANN > MLR > SVM > CNN > LSTM

The performance evaluation of the ML models in terms of mean average deviation (MAD) outputs of the ML algorithm corresponds to RF (0.76), DT (0.77), ANN (0.79),

MLR (0.81), SVM (0.81), CNN (0.92), and LSTM (0.94), respectively. The higher MAD output is an indication that modeled output from the ML deviates from the actual observation practicable. In this case, the RF yielded the least MAD error function suggesting the model is well suited for the optimization of the effluent treatment process. The representativeness of the level of importance of every input dataset via its many trees contributed to the better performance of the RF model compared to other ML and the neural network. The performance evaluation of the ML follows the order:

MAD output: RF > DT > ANN > MLR > SVM > CNN > LSTM

It can be observed from the outline of the validation plots in Fig. 4 a–g that the RF, DT, ANN, MLR, SVM, CNN, and LSTM yielded predicted R^2 corresponding to 0.9411, 0.9405, 0.9401, 0.9391, 0.9335, 0.9269, and 0.9035. This output confirmed the values of predicted R^2 in the range of $0.9035 \leq R^2 \leq 0.9411$ which was consistent across the ML models and the neural network architectures under investigation. This outcome suggests the outputs from the ML models, and neural networks are well correlated with the observed data [20]. The highest R^2 (0.9411) was recorded for the random forest (RF), and the lowest correlation value of 0.9035 was recorded for the CNN. The findings established in the predicted R^2 across all ML is closest unit, suggesting the model assumptions are reliable [10, 20], and the predicted point is well correlated with the actual observation that is practicable. This outcome is well illustrated by the orientation of scattered points of the predicted around the regression line as illustrated in 4c in Fig. 4. The levels of significance of the ML models in terms of predicted r-squared follow the order:

Predicted- R^2 : RF > DT > ANN > MLR > SVM > CNN > LSTM

The standard deviation (Std dev) outputs across the ML statistical metrics will help check the variability of the observations and measure the extent the predicted output deviates from the mean. The standard deviation of the ML will also cross-check sensitivity to outliers and the normality of the distribution. The Std dev evaluation metrics across the ML transcends to RF (1.01), DT (1.02), LSTM (1.06), MLR (1.07), SVM (1.08), ANN (1.12), and CNN (1.31), respectively. The lower values of standard deviation approximately ≤ 1.0 were recorded for the DT, and RF suggesting their predicted outputs are well clustered around the mean value of the observation [10, 17]. Higher Std dev outputs across other ML indicate predicted points are spread out from the observed values from the experimentation [20].

Overall, the ML model with minimal error functions is considered a good and more precise prediction. This outcome was largely due to the efficiency associated with non-linear regression models concerning limited and smaller datasets in size [18, 21]. It can also be concluded from the analysis of the error function that amongst the evaluated ML models and neural network, the RF model revealed the least error evaluation functions, closely followed by the DT and then ANN. This outcome suggests that the RF and DT models produced the most significant assumptions with the highest predictive efficiency. In terms of the evaluation of the statistical metrics and error function outputs across all ML, the RF stood out from the DT, ANN, CNN, MLR, SVM, and LSTM networks.

The research findings proved that, despite the demerit of a higher training time required of the RF and DT models, their evaluation outputs confirmed both models

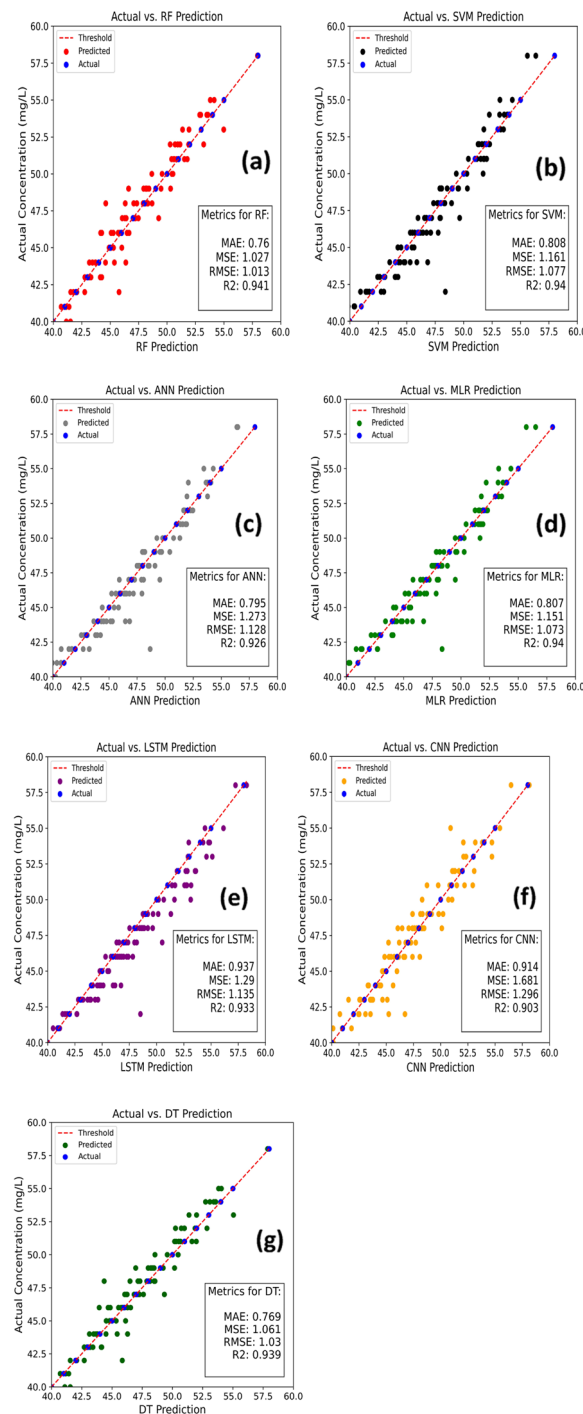


Fig. 4 Plots showing actual observation versus predicted outputs for **a** RF model, **b** SVM, **c** ANN **d** MLR, **e** LSTM, **f** CNN, and **g** DT model

handle the categoric and numerical datasets efficiently, suggesting the ML algorithm is well-tuned to minimize overfitting. Although metrics of the ANN, MLR, and SVM model were varying at $\leq \pm 0.5$ from the actual observation, the results established that the performance of the ANN was limited to the availability of sufficient data size to suit its good generalization power and flexibility to outliers [14, 17]. The

slight variation of the standard deviation from the MLR and SVM suggests a nonlinear relationship across the categorical and numerical data. Consequently, we reasoned that the dynamics of the SVM and MLR were requiring a higher-dimensional space. The performance of the MLR was prone to multicollinearity [10], thereby limiting its flexibility. Despite the merits of capturing spatial hierarchies and abilities to extract meaningful features, the inferior performance of the CNN and LSTM network compared to the DT and RF was possibly due to the high sensitivity of the neural network architecture to outliers in contrast to the more robust ML algorithm in for this domain [10, 17]. The neural network architecture encountered difficulties in interpretability and understanding the learned features. The LSTM recorded the lowest performance significantly due to limited data size to match its higher computational complexity for proper learning and variable-length sequences [6, 8]. In summary, RF and DT models have performed better overall with the most accuracy and consistency. Proceeding, prediction outputs of the RF and DT algorithms were subsequently adopted for the optimization modeling, while other model assumptions and the probability of the neural network architectures were ignored [20].

Optimization solution of the effluent treatment process

It is difficult to ascertain whether the selected model will be making random predictions or incorrect metrics that can lead to unreliable optimization outputs. Considering the good overall performance metrics of the selected RF and DT models, it is imperative to attest to the reliability of the selected ML algorithm for consistency of prediction output using a new testing dataset after training with previously available datasets. The RSM via CCD design matrix spanning 20 experimental runs was applied to validate the reliability of the selected ML algorithm. The design matrix was drafted into the selected ML algorithm for the computation of the optimum objectives.

The analysis of variance (ANOVA) report presented in Table 4 confirmed the output of the RSM (predicted $R^2=0.9876$, $MAD=1.09$, $MAE=0.00437$, and $Std\ dev=\pm 0.54$) obtained from the CCD. The result confirmed that the RF model yielded relatively lower error functions and statistical values compared to the DT. The comparative optimization

Table 4 ANOVA of response surface design of the treatment process

Source	Sum of squares	df	Mean square	F-value	p-value	
Model	1201.42	8	150.18	44.62	< 0.0001	Significant
A—Contact time	142.96	1	142.96	42.48	< 0.0001	
B—Dosage	100.39	1	100.39	29.83	0.0001	
C—pH	1.33	1	1.33	0.3946	0.5416	
D—Particle size	21.82	1	21.82	6.48	0.0256	
AB	11.92	1	11.92	3.54	0.0844	
AC	49.29	1	49.29	14.64	0.0024	
BC	40.40	1	40.40	12.00	0.0047	
A ²	53.28	1	53.28	15.83	0.0018	
Residual	40.39	12	3.37			
Lack of fit	15.89	9	1.77	0.2162	0.9680	Not significant
Pure error	24.50	3	8.17			
Cor. total	1241.81	20				

statistical output recorded for the selected ML shows that the DT ($R^2=0.5000$, $MAD=2.71$, $Std\ dev = \pm 3.58$) and RF ($R^2=0.8025$, $MAD=1.10$, and $Std\ dev = \pm 2.50$) were recorded for the ML algorithm. This output indicated that the predictive results from the machine learning algorithm with low adequacy of precision due to noise, indicating the ML algorithm, were associated with low stability [6, 10]. It can be concluded that while the RF outperformed the DT model, the RSM yield better model statistical metrics. The variation in the statistical metrics of the ML and the RSM confirmed that the smaller sizes of the RSM validation datasets (20 runs) had a direct bearing on the overall inferior performances of the ML algorithm. The reliability statistics provide the need for an increased CCD design space with possible replication to accommodate a minimum of 75 test samples to suit the complexity of the ML domain. The findings confirmed that the smaller data size increases the degree of error functions, the chance of model overfits, and the sensitivity to outliers associated with the complexity of the ML algorithm. The disparity between the RSM and ML models is evident. The results established that the RF yielded outputs with good adequacy of precision, resulting in minimal errors and better validation output than the DT. The predictive outcome from the RSM confirmed that the ML model assumptions and probabilities were compromised. The RSM produced better adequacy of the signal-to-noise ratio [10, 17, 19], with the lesser model constraint [14, 21].

The predicted optimum operating conditions based on the correlations of the minimal residual concentrations output that defines the BFAC-driven adsorption treatment of the pharmaceutical industry effluent are represented in bold text as shown in Table 5. The performance optimization output across the ML algorithm established that the predicted optimum operating conditions for minimizing the residual TDS concentration following the BFAC-driven sorption treatment of the industrial effluent correspond to particle sizes (0.6 mm), pH (6), contact time of 10 min, and required dosage of 2.5 mg/L. For the RF model, the predicted optimum output translates to ≤ 40 mg/L residual TDS concentration present in the effluent after treatment and corresponds to 90% TDS removal efficiency. When compared to the RSM, the DT model output translates to a predicted residual TDS concentration ≥ 41 mg/L under the optimum conditions and corresponds to less than 90% removal. However, the RSM result showed that at similar optimum conditions, the optimal result transcends to a residual TDS concentration ≤ 38 mg/L which is confirmed by the flag point on the 3D surface plot in 5b and c of Fig. 5. This outcome corresponds to the predicted TDS removal efficiency of 91% and transcends to a standard deviation of ± 0.54 from the actual observation practicable. The predicted versus actual plot in 5a in Fig. 5 confirmed the output validation metrics correspond to adjusted R^2 (0.9836) with an adequacy of precision and model- f -value of 44.64, and a p -value of 0.0001 which confirmed that the modeled output was significant [11, 12, 14].

In summary, the predicted outputs from the ML algorithm were significant at a standard deviation output $\leq \pm 2.05$ from the actual values. Although both ML and RSM predicted similar optimum conditions to minimize the residual concentration of TDS present in the water after treatment, the RSM yielded the most significant statistical output at standard deviation output of ± 0.54 from the actual value practicable. The outcome indicates a $\leq 1\%$ difference from the outputs from the ML algorithm. The finding

Table 5 Predicted optimum based on the ML performance and response surface design space

Contact time	Dosage	pH	Particle size	Actual observation	RSM prediction	Predicted DT	Predicted RF
Minute	mg/L		mm	mg/L	mg/L	mg/L	mg/L
10	2.5	6	0.60	42	42.42	42.57	41.61
10	2.5	8	0.15	41	39.69	41.55	41.41
50	0.5	6	0.30	54	54.46	52.65	52.95
10	2.5	8	0.30	42	40.38	42.88	42.69
10	0.5	6	0.85	57	56.99	45.77	45.72
10	2.5	10	0.60	42	41.16	42.89	42.80
50	0.5	8	0.60	58	57.21	57.93	58.08
50	0.5	10	0.85	59	62.12	58.48	58.45
10	2.5	6	0.15	38	40.32	41.74	41.23
50	0.5	8	1.00	59	59.08	57.96	58.04
10	2.5	10	0.15	41	39.06	41.73	41.47
30	0.5	6	0.85	51	52.04	51.58	51.76
30	0.5	8	0.15	48	47.20	48.06	48.20
50	0.5	6	0.85	58	57.03	57.28	57.27
10	0.5	10	0.30	45	44.61	44.74	44.75
50	2.5	6	0.15	45	45.41	46.11	46.18
50	0.5	8	1.00	59	59.08	46.19	45.93
30	2.5	10	0.60	46	45.19	46.19	41.74
10	2.5	6	0.15	40	37.76	41.74	40.23
30	0.5	8	1.00	54	54.32	52.73	52.87
10	2.5	10	0.15	34	39.06	52.74	41.74

ML predicted optimum, pH of 6, dosage of 2.5 mg/L, contact time of 10 min, and particle size of 0.15 mm.

RF reliability metrics, Adj-R² = 0.8532, MSE = 2.06, RMSE = 2.7, and Std dev = ± 1.01.

DT reliability metrics: Adj-R² = 0.5000, MSE = 3.88, RMSE = 5.02, and Std dev. = ± 2.05.

ML predicted efficiency = 90%, residual TDS concentration ≤ 40 mg/L, initial TDS concentration of 450 mg/L.

RSM validation metrics: Adj-R² = 0.9585, MSE = 0.00388, RMSE = 0.00627, Std dev = ± 0.54.

RSM predicted efficiency = 91%, residual TDS concentration ≤ 38 mg/L. Initial TDS concentration of 450 mg/L

from the RSM output confirmed that the BFAC-driven adsorption variables such as BFAC dosage and particle size have a significant antagonistic effect on the efficiency of the treatment process at p -values ≤ 0.005, while the pH at a p -value of 0.5416 does not significantly impact on the BFAC-driven sorption of TDS from the industrial effluent.

The RSM result (Table 4) also established that the contact time and pH of the solution (A × C), BFAC dosage, and contact time (B × C) in the combined system have a significant synergetic effect on the removal of TDS from the pharmaceutical and drug manufacturing effluent treatment process at the p -values of 0.0024 and 0.0047, respectively. These p -values are less than 0.05 and confirmed the selected model terms following the CCD have a significant effect on the performance of the treatment process [11, 12]. The synergetic effect of these variables is illustrated by the 3D surfaces presented in 5 c–d in Fig. 5. The synergetic effect of BFAC dosage and contact time is illustrated by the curvature of the 3D surface of 5a in Fig. 5. The result proved that the TDS reduction reached its optimum at a lower contact time and at a high dosage of 2.5 mg/L, which drives the residual TDS concentration lower than 40 mg/L. The finding indicated that the BFAC sorption propensity occurred rapidly at a sufficient time of 10 min to reach equilibrium and becomes saturated as time increased beyond 10 min. The higher dosage suggests an

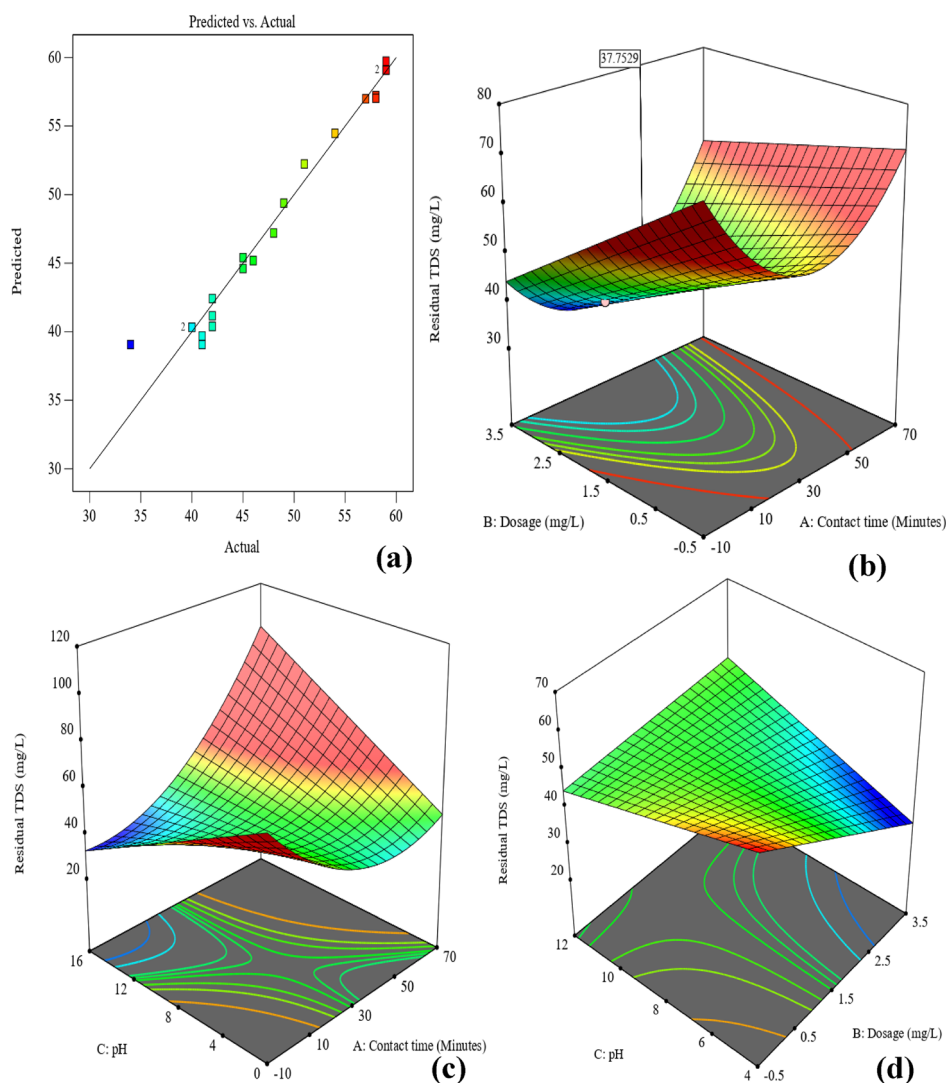


Fig. 5 RSM optimization showing the plot of **a** actual observation vs predicted output and 3D-surface visualization of synergetic effect, **b** dosage and pH, **(c)** pH and contact time, and **d** pH and dosage on performance of BFAC

increase in surface area available as time increased; beyond 10 min, the decrease in TDS concentration may not be as significant at higher contact time.

The curvature of the 3D surfaces in Fig. 5d proved that as the combined effect of contact time and dosage increased, the concentration of TDS decreased intermittently. The outline is indicated by the red color gradient on the base of the surfaces, which translates to a region of lower residuals of TDS following the treatment of the wastewater. The bluish-green color interface on the 3-D surface corresponds to areas of poor adsorptive uptake of the TDS and transcends to a higher residual contamination level. The findings confirmed that the adsorption uptake of TDS from the pharmaceutical effect increased significantly as the dosage increased from 0.5 to 2.5 mg/L, as the pH increased from 2 to 6. This outcome corresponds to a significant reduction in the residual TDS concentration to 60 mg/L. As the pH of the solution increased from 8 to 10, residual TDS concentration raised from 43 to 55 mg/L. The findings suggest that the pH of the solution

affects the surface charge of biosorbent and the TDS molecules present in the solution. At a lower pH window (4–6), the surface charge on the BFAC becomes more positive which drove the affinity of the TDS molecules to adsorb onto the biochar surface leading to a reduction in their concentration in the medium. As the pH of the solution rises from 6 to 8, surface charge on adsorbent became less positive, decreasing its affinity for TDS molecules. However, as the dosage of adsorbent increases from 0.5 to 2.5 mg/L, the number of BFAC particles available for sorption of TDS molecules increases which in turn drives the reduction of TDS to an optimum 40 mg/L in the effluent medium.

Physiochemical property of the treated effluent after treatment

In this research work, we investigated the pharmaceutical and drug manufacturing effluent medium on the characterization of the treated sample to ensure effluent quality satisfied the World Health Organization (WHO) standard for clean discharge. The result is presented in Table 6. It can also be observed from the characterization result in Table 6 that the biosorbent raised the BOD of the finished water from an initial 10 mg/L by 18.2 mg/L with a corresponding rise in pH from 4.92 to 5.5 and temperature from 27.1 to 27.5 respectively. This outcome indicates that the finished effluent has become more alkaline by a factor of 0.56 suggesting the effluent satisfied a stability index of an acceptable standard ($5.5 \leq \text{pH} \leq 8.5$) [10, 13]. The slight increase in BOD concentration indicates a significant decrease in the efficiency of the BFAC-driven sorption process, suggesting the treatment method used may not be specifically designed to target BOD remediation [7, 22]. The increase in BOD by 8.2 mg/L can be attributed to the impact of the higher dosage of the BFAC adsorbent used. However, it is worth noting that the pH, BOD, and temperature contents of the finished effluent satisfied the maximum permissible limits for industrial effluent discharge [13].

The characterization result in Table 6 confirmed the BFAC-driven sorption treatment of the pharmaceutical industry effluent reduced the COD, TSS, total iron, TDS, nitrate, and odor content in the industrial effluent to minimum thresholds. The TDS and TSS concentrations of the finished water after BFAC-driven sorption treatment satisfied the WHO acceptable standard for effluent discharge at the optimum condition. This

Table 6 Pharmaceutical effluent characteristic after the BFAC-driven sorption treatment

Effluent parameter	WHO (max. permissible limit)	Residual after treatment
Temperature	Nil	27.5 °C
BOD ₅	5 mgL ⁻¹	18.2 mgL ⁻¹
Biodegradability index (BI)	NA	0.5
Total suspended solid (TSS)	30 mgL ⁻¹	4.2 mgL ⁻¹
COD	10 mgL ⁻¹	0.001 mgL ⁻¹
Conductivity	1200 μS cm ⁻¹	1210 μS cm ⁻¹
Total iron (Fe)	3.0 mgL ⁻¹	0.09 mgL ⁻¹
pH	6.59–9.50	5.5
Color	Nil	Colorless
Nitrate	50 mg/L	12
TDS	50 mgL ⁻¹	40 mgL ⁻¹
Odor	Nil	Odorless
E-coil	Negative	Negative

outcome indicates that the biochar is effective for the removal of suspended and dissolved solids and colored and fluorescent dissolved organic matter present in the effluent medium. The treatment method was successful in reducing the TDS parameter to a desirable level that satisfied the WHO standard for industrial effluent discharge.

Conclusions

This research work is focused on the ML optimization treatment of pharmaceutical and drug manufacturing effluent using activated carbon derived from breadfruit husk. The implementation of 625 datasets following 80% testing and 20% validation on selected ML algorithms and neural network architecture for optimization of the removal of TDS from pharmaceutical and drug manufacturing effluent have been investigated. The SVM, DT, RF, MLR, CNN, ANN, and LSTM networks were applied to minimize the concentration of residual TDS present in the effluent following BFAC-driven sorption treatment. The results proved that the RF yielded the best evaluation and validation metrics for the optimization of the treatment process. The selected RF model predicted optimum transcends to an optimum residual ≤ 40 mg/L, corresponding to $\geq 90\%$ efficiency of removal and translates to ± 1.01 standard deviation from actual values practicable. The optimum conditions transcend to contact time of 10 min, pH 6, dosage (2.5 mg/L), and particle size of 0.6 mm, respectively. The ML algorithm outperformed the neural network architecture. The performance of the ML proved that a large dataset size (125) suits the test model evaluation metrics with a lower probability of error functions. The research outcome confirmed that the higher the data size, the smaller the MSE and RMSE error outputs and therefore recommends that new test datasets different from the trained and learned dataset should be used to validate selected ML models in other to confirm their reliability. Overall the adsorption treatment method used was effective in significantly reducing the concentrations of these contaminants (COD, color, TSS, nitrates, and TDS) present in the pharmaceutical effluent. It also recommends further investigation of the reasons behind the slight increases in temperature, BOD concentration, and pH of the solution at the optimum to ensure proper treatment of these parameters for future efforts. The finding recommends BFAC as an active sorbent for the remediation of TDS, TSS, and various other contaminants from the effluent medium.

Abbreviations

BFAC	Breadfruit activated carbon
TDS	Total dissolved solids
TSS	Total suspended solids
CNN	Convolutional neural network
ANN	Artificial neural network
DTM	Decision tree model
RFM	Random forest model
LSTM	Long short-term memory
SVM	Support vector machine
RSM	Response surface methodology
ML	Machine learning
MLR	Multiple linear regression
COD	Chemical oxygen demand
BOD	Biological oxygen demand
RMSE	Root-mean-square error
MSE	Mean square error
MAE	Mean absolute error
MAD	Mean absolute deviation
Std	Standard deviation

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

OPE, validity tests, data curation, and writing—original draft preparation. UVI, conceptualization, methodology, data curation, and project administration and supervision. EF, software, validation, and writing—reviewing and editing. AC, visualization, investigation, and software. TAM, software, validation, and writing—reviewing and editing. ICA, supervision, writing—review, and project administration. MHD, validation and writing—review and editing. All authors have read and approved the manuscript for publication.

Funding

No financial support was received for this work.

Availability of data and materials

The authors declare that data will be made available prior to request to the corresponding authors.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Chemical Engineering, Nnamdi Azikiwe University, P.M.B. 5025, Awka 420218, Nigeria. ²Department of Chemical Engineering, Federal University of Petroleum Resources, P.M.B. 1221, Effurun, Nigeria. ³Faculty of Computer Science, University of Prizren "Ukshin Hoti", Prizren, Kosovo. ⁴Technical Services, LLC, Dayton, NJ, USA. ⁵Department of Multimedia, UFR STGI-Universite de Franche-Comte, 4 Pl, Lucien Tharradin, Montbeliard 25200, France. ⁶Department of Applied Bioeconomy, Wroclaw University of Environmental and Life Sciences, Wroclaw 51-630, Poland. ⁷Department of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. ⁸Center for Water Quality Research, Institute for Environmental Research, Tehran University of Medical Sciences, Tehran, Iran. ⁹Center for Solid Waste Research, Institute for Environmental Research, Tehran University of Medical Sciences, Tehran, Iran.

Received: 15 August 2023 Accepted: 20 October 2023

Published online: 14 November 2023

References

1. Karimi-Maleh H, Ranjbari S, Tanhaei B, Ayati A, Orooji Y, Alizadeh M, Karimi F, Salmanpour S, Rouhi J, Sillanpää M (2021) Novel 1-butyl-3-methylimidazolium bromide impregnated chitosan hydrogel beads nanostructure as an efficient nanobio-adsorbent for cationic dye removal: kinetic study. *Environ Res* 195:110809
2. Hijji, M., Chen, T., Ayaz, M., Abosinnee, A. S., Muda, I., Razoumny, Y., & Hatamiakouei, J. Optimization of state of the art fuzzy-based machine learning techniques for total dissolved solids prediction. *Sustainability*, 15(8), 2023.7016 <https://doi.org/10.3390/su15087016>
3. Saravanan, A., P.S. Kumar, S. Jeevanantham, M. Anubha, and S. Jayashree, Degradation of toxic agrochemicals and pharmaceutical pollutants: effective and alternative approaches toward photocatalysis. *Environmental Pollution*, 2022: p. 118844.
4. Majumder, A., B. Gupta, and A.K. Gupta, Pharmaceutically active compounds in aqueous environment: a status, toxicity and insights of remediation. *Environmental research*, 2019. 176: Pp.108542.
5. Barzegari Banadkooki, F., Ehteram, M., Panahi, F., Sh. Sammen, S., Binti Othman, F., & EL-Shafie, A. Estimation of total dissolved solids (TDS) using new hybrid machine learning models. *Journal of Hydrology*, 2020. 124989: doi:<https://doi.org/10.1016/j.jhydrol.2020.124989>
6. Ewusi A, Ahenkorah I, Aikins D (2021) Modelling of total dissolved solids in water supply systems using regression and supervised machine learning approaches. *Appl Water Sci* 11:13. <https://doi.org/10.1007/s13201-020-01352-7>
7. Ugonabo VI, Ezech EM, Onukwuli OD et al (2023) Remediation of pharmaceutical industrial wastewater using activated carbon from seeds of *Mangifera indica* and husks of *Treculia Africana*: optimization, kinetic, thermodynamic and adsorption studies. *Chemistry Africa* 6:683–698. <https://doi.org/10.1007/s42250-022-00450-0>
8. Yaqub M, Asif H, Kim S, Lee W (2020) Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network. *J of Water Process Eng* 37:101388. <https://doi.org/10.1016/j.jwpe.2020.101388>
9. Granata F, Papirio S, Esposito G, Gargano R, De Marinis G (2017) Machine learning algorithms for the forecasting of wastewater quality indicators. *Water* 9(2):105. <https://doi.org/10.3390/w9020105>
10. Ugonabo VI, Ovuoraye PE, Chowdhury A et al (2022) Machine learning model for the optimization and kinetics of petroleum industry effluent treatment using aluminum sulfate. *J Eng Appl Sci* 69:108. <https://doi.org/10.1186/s44147-022-00164-7>
11. Barzegari Banadkooki, F., Ehteram, M., Panahi, F., Sh. Sammen, S., Binti Othman, F., & EL-Shafie, A. Estimation of total dissolved solids (tds) using new hybrid machine learning models. *Journal of Hydrology*, 2020. 124989: <https://doi.org/10.1016/j.jhydrol.2020.124989>
12. Enyoh, C. E., Wang, Q., & Ovuoraye, P. E. (2022). Response surface methodology for modeling the adsorptive uptake of phenol from aqueous solution using adsorbent polyethylene terephthalate microplastics. *Chemical Engineering Journal Advances*, 12, 100370. <http://doi.org/10.1016/j.cea.2022.100370>

13. Ovuoraye PE, Okpala LC, Ugonabo VI et al (2021) Clarification efficacy of eggshell and aluminum base coagulant for the removal of total suspended solids (TSS) from cosmetics wastewater by coag-flocculation. *Chem Pap* 75:4759–4777. <https://doi.org/10.1007/s11696-021-01703-x>
14. Igwegbe CA, Mohammadi L, Ahmadi S, Rahdar A, Khadkhodayi D, Dehghani R, Rahdar S (2019) Modeling of adsorption of methylene blue dye on Ho-CaWO₄ nanoparticles using response surface methodology (RSM) and artificial neural network (ANN) techniques. *MethodsX* 6:1779–1797
15. Igwegbe CA, Ovuoraye PE, Białowiec A, Okpala CO, Onukwuli OD, Dehghani MH (2022) Purification of aquaculture effluent using *Picralima nitida* seeds. *Sci Rep* 12(1):1–19. <https://doi.org/10.1038/s41598-022-26044-x>
16. Mangkoedihardjo S (2006) Biodegradability improvement of industrial wastewater using hyacinth. *J Appl Sci* 6:1409–1414
17. Ovuoraye PE, Ugonabo VI, Tahir MA, Balogun PA (2022) Kinetics-driven coagulation treatment of petroleum refinery effluent using land snail shells: an empirical approach to environmental sustainability. *Cleaner Chem Eng* 4:100084. <https://doi.org/10.1016/j.clce.2022.100084>
18. Zhang, Y. and Y. Wu, Introducing machine learning models to response surface methodologies, in *Response Surface Methodology in Engineering Science*. 2021, IntechOpen.
19. Guo H, Jeong K, Lim J, Jo J, Kim YM, Park J-P, Kim JH, Cho KH (2015) Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *J Environ Sci* 32:90–101
20. Ebere, E., Ovuoraye, P., Isiuku, O., & Igwegbe, C. (2023). Artificial neural network and response surface design for modeling the competitive biosorption of pentachlorophenol and 2,4,6-trichlorophenol to *Canna indica* L. in Aquaponia. *Analytical Methods in Environmental Chemistry Journal*, 6(01), 79–99. <https://doi.org/10.24200/amecj.v6.i01.228>
21. Wang D, Thunéll S, Lindberg U, Jiang L, Trygg J, Tysklind M, Souihi N (2021) A machine learning framework to improve effluent quality control in wastewater treatment plants. *Sci Total Environ* 784:147138
22. Majumder A, Gupta B, Gupta AK (2019) Pharmaceutically active compounds in aqueous environment: a status, toxicity and insights of remediation. *Environ Res* 176:108542

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
