Separation and Purification Technology 406 (2026) 139012

Contents lists available at ScienceDirect

Separation and Purification Technology

journal homepage: www.elsevier.com/locate/seppur

Optimized adsorption removal and capacity prediction of anionic pollutants using a hybrid strategy of machine learning algorithms☆

Hamza Ul Haq [a], Muhammad Yasir [b,*], Muhammad Nouman Aslam Khan [a], Jawad Gul [a], Mukarram Zubair [c], Hassan Ali [b], Vladimir Sedlarik [b], Nasir.M. Ahmad [a,*]

[a] Multiscale & Intelligence for Materials Innovation and Chemical Engineering, School of Chemical and Materials Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan
[b] Centre of Polymer Systems, Tomas Bata University in Zlin, Tr. T. Bati 5678, 76001 Zlin, Czech Republic
[c] Environmental Engineering Department, College of Engineering A13, Imam Abdulrahman Bin Faisal University, Main Campus, P.O. Box 1982, Dammam 34212, Saudi Arabia

ARTICLE INFO

Editor Name: B. Van der Bruggen

Keywords:
Machine learning
Adsorption
Anionic pollutant
Wastewater treatment
Adsorbent materials

ABSTRACT

Accurate prediction of adsorption performance is crucial for optimizing wastewater treatment systems; however, the complex interactions among operational variables and adsorbent properties often limit conventional modeling approaches. In this study, a machine learning framework was developed to predict the adsorption removal efficiency and kinetic capacity of anionic pollutants in aqueous systems. A comprehensive experimental dataset was generated using four representative pollutants, i.e., bovine serum albumin, methyl orange, sulfate, and nitrate, and four adsorbent materials, including powdered activated carbon (PAC), thermally modified PAC, chemically modified PAC, and ion-exchange chitosan beads. Key operational parameters, including pH, contact time, adsorbent dosage, BET surface area, solution volume, and pollutant concentration, were used as input features. Four ML algorithms, i.e., Decision Tree (DT), Gaussian Process Regression (GPR), Support Vector Machine (SVM), and Ensemble Learning Tree (ELT), were developed and further optimized using Bayesian optimization to improve predictive performance. Among the evaluated models, the optimized ELT model demonstrated the highest predictive accuracy with a coefficient of determination (R^2) of 0.78, indicating its strong capability in capturing nonlinear adsorption behavior. Model interpretation through partial dependence plots revealed significant influences of pH, adsorbent dosage, BET surface area, and initial pollutant concentration on adsorption performance, while Sobol sensitivity analysis confirmed the dominant role of initial concentration. Experimental validation using jojoba-derived biochar for the removal of methyl orange and Eriochrome Black T dyes showed strong agreement with model predictions. The developed ML models provide a reliable tool for predicting adsorption performance and designing efficient adsorbent-based wastewater treatment systems.

1. Introduction

Anthropogenic activities over the last century have resulted in the release of various pollutants into the environment, greatly affecting global biodiversity and humans' health alike. These pollutants, originating primarily from industrial processes, energy production, and transportation, pose grave risks to the planet's ecosystems. Water pollution, which arises mainly from the discharge of untreated industrial effluents, agricultural runoff containing pesticides and fertilizers, and domestic sewage, degrades aquatic habitats and threatens biodiversity [1]. Industries often release toxic chemicals, heavy metals, and dyes into nearby water bodies without adequate treatment, while agricultural activities contribute pollutants such as fertilizers, pesticides, and animal waste, which can lead to eutrophication [2,3]. This process causes excessive algal growth, depletes oxygen levels, and results in the death of aquatic organisms. This uncontrolled level of environmental pollution threatens environmental sustainability; therefore, tackling global environmental pollution has become one of the primary goals to improve the

☆ This article is part of a Special issue entitled: 'Innovative Materials' published in Separation and Purification Technology.

* Corresponding authors.
E-mail addresses: yasir@utb.cz (M. Yasir), nasir.ahmad@scme.nust.edu.pk (Nasir.M. Ahmad).

https://doi.org/10.1016/j.seppur.2026.139012
Received 20 April 2026; Received in revised form 16 June 2026; Accepted 18 June 2026
Available online 23 June 2026
1383-5866/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

quality of life for the entire global biodiversity of the planet [4]. Therefore, stringent regulations have been implemented to prevent the unregulated release of environmental toxicants. For example, the UN's Sustainable Development Goals No. 6 and 14 directly address this issue by providing guidance on how to tackle it [5]. To date, various technologies have been explored to address water pollution, including sedimentation [6], reverse osmosis, ozonation, ion-exchange resins, membrane filtration [7], adsorption, biological treatment, and advanced oxidation processes (AOPs) [8–10]. Each method provides different rates of pollutant removal efficiency and flexibility. Among these techniques, adsorption is widely regarded as one of the most effective and versatile technologies for removing pollutants from water, offering several advantages over other treatment methods. Unlike chemical or biological treatments that may require complex reactions or specific environmental conditions, adsorption is simple, efficient, and easy to operate, as it requires mainly a suitable adsorbent with high affinity to adsorb pollutants to the surface, such as activated carbon [11], biochar [12,13], zeolites [14], or nanomaterials, without altering the chemical structure of the pollutants [15]. Moreover, adsorption systems generally utilize low-cost feedstock adsorbents [16], which produce minimal secondary waste, unlike other methods, which mostly generate secondary pollutants after pollutant degradation or decomposition [17], such as from ozonation or AOPs [18]. To date, research in this field is primarily focused on developing highly porous, high-surface-area materials with tailored surface chemistry to improve adsorption, adsorption capacity, selectivity, kinetics, and regeneration [19]. These materials include carbon-based materials, such as carbon quantum dots and metal-organic frameworks, as well as bio-based materials [20,21]. In addition to structurally engineered adsorptive materials for pollutant removal based on the preceding strategies, various operational parameters also affect the rate and extent of adsorptive removal. These factors include contact time, pollutant concentration, pH, solution volume, surface area, adsorbent dosage, and others [22,23]. These factors, along with others, greatly influence and control the adsorptive performance of the adsorbent material. However, it is challenging to analyze the influence of each factor on determining the adsorptive rate of pollutant removal, as it proceeds through multivariable interactions. In this regard, an investigation into the influence of each operational parameter on the adsorptive removal of pollutants can provide valuable insights into achieving highly cost-effective, efficient pollutant removal with adsorbent materials.

Fortunately, owing to recent advances in computer modeling and the implementation of machine learning (ML), the complex interplay between various operational parameters can be elucidated using ML. In ML, various algorithms are applied to a collected dataset to analyze patterns and draw predictions [24]. To obtain highly accurate ML models that can predict required output values, it is necessary to acquire a large dataset, select suitable ML models, train them, and validate them by testing them on randomly selected parameter values. So far, ML has been successfully implemented in various fields, including chemistry, for materials discovery [25], prediction of chemical reactions [26], optimization of reactions [27], photocatalytic dye removal optimization [28], and adsorptive removal of pollutants [29]. However, the accuracy and feasibility of each ML is different from application to application, with each model having its own degree of applicability, versatility, and suitability. Some common ML models investigated for various use cases include decision trees (DT) [24], random forests (RF) [25], support vector machines (SVM) [26], and linear regression (LR) [27]. As in the case of other chemical science-based applications, the literature survey suggests that ML has been implemented to investigate various aspects of the adsorptive removal of pollutants. For instance, the Venegas group reported on utilizing an ML approach to predict adsorptive removal of ciprofloxacin, a highly bioactive drug detected in high concentrations in wastewaters, using sugarcane bagasse biomass. Based on their experimental data, the developed ML models achieved great prediction capability at both laboratory and pilot scale experiments. The authors reported optimized adsorptive capacities of 2.19 mg/g and 2.53 mg/g, for laboratory and pilot scale experiments, respectively, by which optimized operational parameters of both different experiments were obtained [30]. Similarly, other groups have reported applying ML models to predict adsorptive removal of pollutants such as Lindane [31], organic pollutants [32], and selenium [33]. However, the preceding studies, while novel themselves, have been conducted on a limited scale and with a limited variety of pollutants. Therefore, considering the current state of the art in this field, we report on the implementation of ML models on a large dataset to remove various anionic pollutants using adsorptive materials. Four commonly encountered anionic pollutants in wastewater, i.e., bovine serum albumin (BVA), methyl orange (MO), SO_4^{2-}, and NO_3^{1-} were selected, while four different adsorptive materials, i.e., powdered activated carbon (PAC), thermally modified PAC (TPAC), chemically modified PAC (CPAC), and ion exchange chitosan beads (IECBs), were chosen to obtain a reliable dataset. To leverage ML models, various operational influencing parameters were selected, and four ML models (DT, GPR, SVM, and ELT) were applied to establish intricate relationships between the adsorptive pollutants' removal degree and these parameters. The developed and trained ML models were further optimized using the Bayesian optimization technique, enabling more accurate prediction of the adsorptive capacity of selected adsorptive materials by considering the influence of operational parameters. The developed ML models were further experimentally validated to assess their reliability by conducting adsorptive removal of pollutants using randomly selected input features, yielding high predictive accuracy.

2. Methodology

2.1. Data collection and pre-processing

For training and testing ML models, adsorptive removal data for various adsorbents were obtained from our group's laboratory experiments. Diverse functional materials with key morphologies were synthesized and tested for the remediation of anionic pollutants from aqueous media. Distinct physicochemical characteristics such as surface area, surface chemistry, functional groups, and pore architecture resulted in a wide range of adsorption performances of materials. Experiments were conducted by varying selected parameters to obtain a diverse range of data points suitable for enhancing the model's training and predictive capabilities. These experimental data points in the form of performance at various operational conditions were then unified into a consolidated dataset. A tabulated list of all data points, together with their associated parameters, is provided in Table S1. Relevant descriptors affecting the adsorption of the target pollutants were selected as input features for the development of the predictive model. The adsorptive experiments were conducted in sonication and non-sonication (batch) modes, which can also influence the adsorptive removal capacity. To keep the study's data points consistent, only commonly encountered anionic pollutants were used in all experiments (Fig. 1).

2.2. Bibliometric analysis

The bibliometric analysis reveals a steadily expanding research landscape at the intersection of ML, adsorption processes, and anionic pollutants removal from wastewater systems between 2016 and 2026. As illustrated in Fig. S1a, publication output remained modest during the early years (2016–2018) but began to increase significantly after 2020, with a pronounced rise in both publications and citations between 2022 and 2025. The surge in citations in later years indicates growing scholarly recognition and the consolidation of foundational studies in this domain. Although the reported annual growth rate appears statistically neutral, the visual trend clearly shows accelerated research activity in recent years, suggesting that ML-assisted adsorption research is

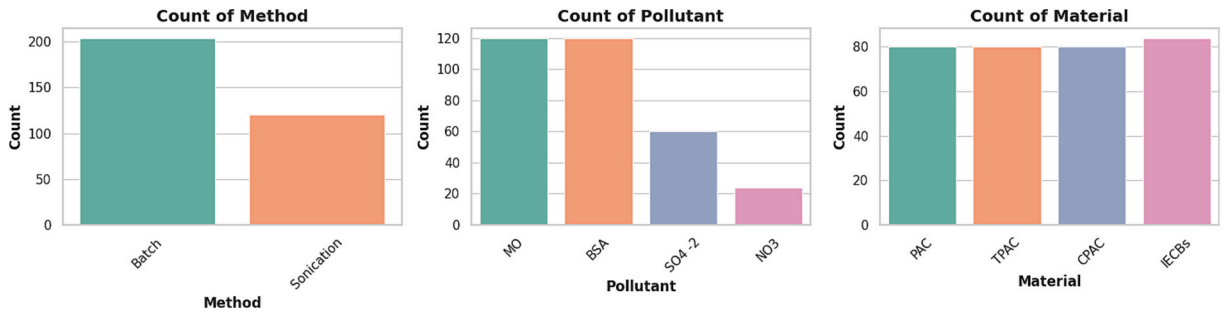


Fig. 1. Number of counts of adsorptive removal method, pollutant, and material type utilized for training ML models.

transitioning from exploratory investigations to a more mature and structured field. The distribution of subject areas (Fig. S1b) further highlights the field's interdisciplinary nature. Engineering, Chemical, and Environmental Sciences constitute the dominant categories, reflecting the strong process-engineering orientation of adsorption-based wastewater treatment studies. Concurrently, substantial contributions from Chemistry-related disciplines underscore the importance of material design, adsorbent characterization, and physicochemical interaction mechanisms. The presence of Water Resources and Engineering Environmental categories confirms that research efforts are closely aligned with sustainable water management and environmental remediation objectives.

The thematic and collaboration network depicted in Fig. S1c provides additional insight into the intellectual structure of the field. "Adsorption," "machine learning," and "artificial neural network" emerge as central conceptual nodes, indicating that predictive modeling and optimization techniques are increasingly integrated into adsorption performance evaluation. Countries such as India, Egypt, Saudi Arabia, and China appear prominently in the collaboration network, suggesting that emerging economies are actively contributing to advancements in ML-driven wastewater treatment research. Furthermore, the concentration of publications in specialized journals related to separation technology, environmental engineering, and chemical processing reflects the applied and solution-oriented nature of this research area. Collectively, these findings indicate a rapidly evolving, multidisciplinary field characterized by increasing research impact, collaborative engagement, and methodological integration of artificial intelligence within environmental engineering frameworks.

The keyword co-occurrence analysis provides insight into the intellectual structure and thematic evolution of ML-assisted adsorption research for anionic pollutants removal in wastewater systems. As illustrated in Fig. S2a, adsorption emerges as the most dominant and central keyword, confirming its foundational role within this research domain. Closely associated terms such as removal, aqueous solution, adsorbent, and activated carbon indicate that experimental adsorption studies remain the primary research backbone. The dense interconnections among these terms highlight the continued emphasis on pollutant removal from aqueous media using engineered adsorbent materials. Importantly, ML-related terms, including machine learning, artificial neural network (ANN), and optimization, are strongly integrated within the network rather than appearing as isolated clusters (Fig. S2a). This suggests that predictive modeling approaches are increasingly embedded within adsorption performance evaluation frameworks. The presence of modeling-associated keywords such as response surface methodology (RSM), kinetics, and isotherm further indicates the convergence of statistical, mechanistic, and AI-based methods. Additionally, the appearance of material-oriented terms such as nanoparticles, graphene oxide, biochar, and composites reveals ongoing innovation in adsorbent design, demonstrating that computational intelligence is being applied alongside material-engineering advancements. The thematic map shown in Fig. S2b further categorizes research themes according to their centrality and development density. Core topics such as adsorption, aqueous solutions, and removal are positioned within the motor theme quadrant, indicating that they are well developed and highly central to the field. These themes represent the driving intellectual forces of the research area. In contrast, foundational yet still evolving topics such as wastewater treatment, performance, and adsorption mechanisms fall within the basic theme quadrant, suggesting their essential yet developing role in sustaining research growth. Methodological themes, including response surface methodology and artificial neural networks, appear in niche or transitional areas, suggesting that although these approaches are technically well developed, their broader conceptual integration continues to expand. The placement of certain AI-related terms in emerging or transitional zones suggests a dynamic shift toward more advanced or hybrid modeling frameworks.

The quantitative keyword distribution presented in Fig. S2c corroborates these observations. The tree map clearly demonstrates that adsorption accounts for the highest frequency of occurrences, followed by aqueous solution, removal, and activated carbon. The notable presence of ML, ANN, and optimization confirms the increasing prominence of computational modeling within adsorption research. Moreover, frequent references to model pollutants such as methylene blue, Congo red, and anionic dyes indicate that dye-removal studies remain a common benchmark for evaluating adsorbent performance. This pattern suggests that while artificial intelligence techniques are gaining importance, the field remains application-oriented and strongly driven by experimental performance assessment.

2.3. *Data representation*

The distributions of the input and output variables used to train the machine learning models are illustrated in Fig. S3 using violin plots. These plots combine kernel density estimation with boxplots, providing insight into the data's spread, central tendency, and density. The adsorption efficiency exhibits a broad distribution with higher data density at moderate to high removal efficiencies, indicating that most reported systems achieve effective adsorption under optimized conditions. BET surface area shows a broad, partially bimodal distribution, indicating the use of both low- and high-porosity adsorbents across the collected studies. Experimental parameters such as adsorbent dosage, solution volume, and contact time are clustered within commonly adopted ranges, while sufficient variability is retained to ensure robust model learning. The initial pollutant concentration displays a pronounced right-skewed distribution, spanning dilute to highly concentrated systems, whereas equilibrium concentration is predominantly skewed toward lower values, consistent with efficient contaminant removal. The pH distribution is multimodal, with most experiments conducted under acidic to near-neutral conditions, highlighting the strong role of pH in adsorption processes. Overall, the broad yet physically meaningful distributions of these variables obtained in our lab confirm the dataset's diversity and representativeness, which are essential for developing reliable and generalizable machine learning models.

Fig. 2 presents the combined correlation heatmap (lower triangle) and coin plot (upper triangle) illustrating the pairwise relationships among the selected variables utilized for training the ML models. From the figure, the adsorption efficiency shows moderate positive correlations with contact time and initial pollutant concentration, indicating enhanced removal performance at longer interaction times and higher driving force for mass transfer, while it is negatively correlated with equilibrium concentration, consistent with effective adsorption leading to lower residual concentrations. Kinetic adsorption capacity (Q_e) exhibits a strong positive correlation with the initial concentration and a negative correlation with equilibrium concentration, implying typical adsorption behavior governed by concentration gradients. Contact time is positively correlated with initial concentration but negatively correlated with equilibrium concentration, suggesting improved pollutant uptake as exposure time increases. BET surface area and solution volume display strong intercorrelation, reflecting their combined influence on available adsorption sites and mass transfer conditions, while both variables show moderate correlations with adsorption efficiency and kinetic capacity. Adsorbent dosage is moderately negatively correlated with equilibrium concentration and positively correlated with solution volume, indicating that higher adsorbent loading promotes greater pollutant removal. Overall, the correlation analysis reveals physically meaningful interdependencies among operational and structural parameters, supporting their relevance as input features for ML models development.

2.4. Machine learning models

In this study, multiple machine learning algorithms were employed to develop a hybrid predictive framework for modeling the adsorption efficiency and capacity of anionic dyes and proteins on functionalized biochar surfaces. The integration of diverse models, such as Decision Tree (DT), Ensemble Learning Tree (ELT), Gaussian Process Regression (GPR), and Support Vector Machine (SVM), enables a comprehensive learning of both linear and nonlinear adsorption behavior [34,35]. Each algorithm contributed unique computational strengths, collectively improving prediction reliability and interpretability of adsorption phenomena. Each algorithm contributed unique computational strengths, collectively improving prediction reliability and interpretability of adsorption phenomena. Overfitting was mitigated by scrutinizing feature selection followed by hyperparameter optimization. Moreover, the testing data for the developed surrogate model were kept independent from the training dataset with 5-fold cross-validation.

2.4.1. Decision tree (DT)

The decision tree algorithm was employed for its ability to provide clear interpretability and variable importance analysis. DT works by recursively partitioning the data based on decision rules that minimize prediction error, allowing straightforward identification of critical factors influencing adsorption removal and capacity. In adsorption-related research, DT is particularly useful for revealing threshold-dependent behavior [36], for instance, determining the optimal pH or adsorbent dosage beyond which removal efficiency plateaus. Its transparency makes it ideal for understanding mechanistic trends in adsorption systems. To determine estimated outputs, we usually average the true outputs of training samples within a terminal node [37]:

$$\widehat{y}_j = \frac{\sum_{i \in t_j} Y_j}{|t_j|} \tag{1}$$

where t_j = leaf node j, and $|t_j|$ = total number of datapoints/samples in the leaf node j. The split criteria of the decision tree are based on the impurity measure with the least squares deviation [38].

$$I(t_j) = \sum_{i \in t_j} \left(y_i - \widehat{y}_j\right)^2 \tag{2}$$

where $I(t_j)$ = the impurity measure at node j. Using the least square deviation, the splitting criteria are calculated as follows:

$$\Delta I = I(t_p) - p_l I(t_l) - P_r I(t_r) \tag{3}$$

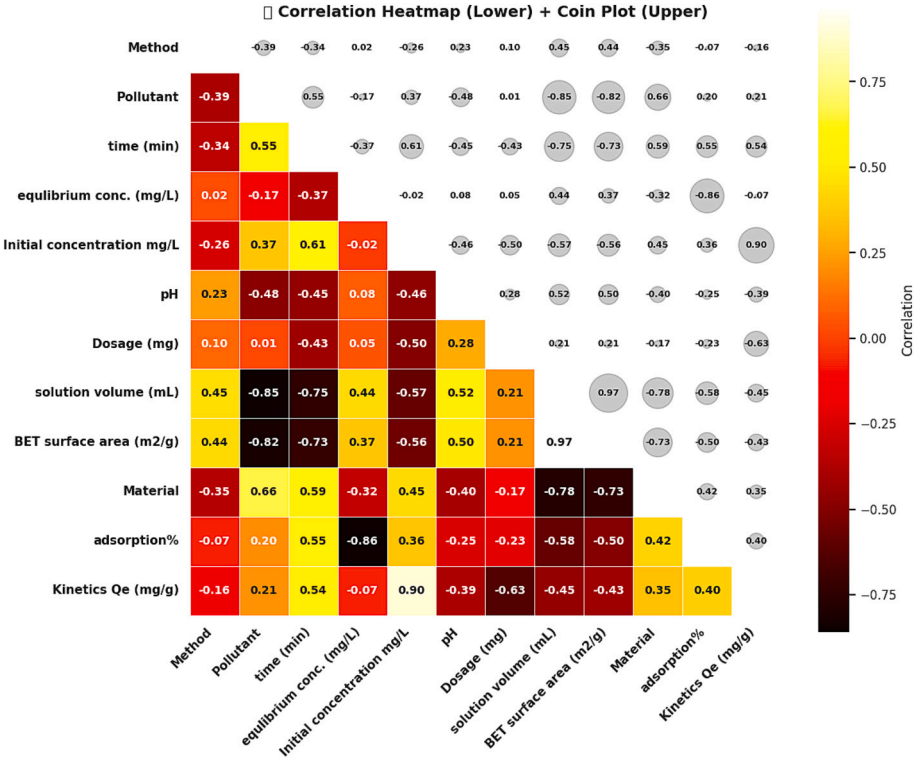


Fig. 2. Pearson correlation and coin plot between adsorption removal and experimental conditions.

where t_l and t_r = two child nodes of t_p = parent node. The left child node assigned with proportions of data is represented by P_l, and the right child node assigned with proportions of data is represented by P_r.

2.4.2. Ensemble learning tree (ELT)

The Ensemble Learning Tree approach enhances predictive robustness by aggregating multiple weak learners to form a stronger composite model [39]. This ensemble strategy reduces overfitting, mitigates noise sensitivity, and improves overall model generalization, an essential feature when dealing with complex, nonlinear adsorption datasets [40]. In the current study, ELT was adopted to improve accuracy and stability in predicting adsorption performance across varying conditions. By averaging or combining outputs from several individual models, the ensemble method captures diverse feature interactions, making it effective for systems where adsorption behavior depends on multiple correlated factors such as ionic strength and molecular size. ELT offers a competitive, superior predictive performance edge over the modern ensemble algorithms being applied for environmental modeling. Conventional Random Forest models lack the node splitting accompanied by additional randomness, which are primary factors for improved model generalizability and reduced variance. ELT offers lower computational complexity alongside effective capture of the nonlinear behavior of adsorption parameters [41].

2.4.3. Gaussian process regression (GPR)

Gaussian Process Regression was integrated into the hybrid model due to its probabilistic and non-parametric nature, allowing it to represent complex functional relationships between input variables and adsorption outcomes [42]. GPR not only provides accurate predictions but also quantifies uncertainty [43], which is particularly valuable for experimental adsorption data that may exhibit variability due to surface heterogeneity or adsorbate–adsorbent interactions. Its kernel-based approach enables modeling of nonlinear adsorption dynamics, making it a strong candidate for capturing the underlying physicochemical processes in biochar-based adsorption systems for dyes and proteins. The Gaussian process t(u) is parameterized by two independent variables, the mean function m(u) and the kernel function, also known as the covariance function d (u, u'), which are each evaluated for the instances u and u', respectively (u) [44]. The equations below provide definitions for these two functions.

$$m(u) = E(t(u)) \tag{4}$$

$$\mathrm{Cov}(t(u), t(u')) = d(u, u'; \sigma) = E((t(u) - m(u))(t(u') - m(u'))) \tag{5}$$

where, σ = model hyperparameters that must be adjusted. Mathematically Gaussian process t(u) is represented as

$$t(u) \sim \mathrm{GP}(m(u), k(u, u')) \tag{6}$$

where GP denotes the Gaussian process model This implies that t(u) is a probabilistic function where, m(u) = Gaussian mean and d (u, u') = covariance function.

2.4.4. Support vector machine (SVM)

Support Vector Machine was utilized for its superior performance in handling nonlinear regression tasks and its ability to model high-dimensional feature spaces effectively [45]. Through kernel transformations, SVM can map input parameters, such as solution chemistry, adsorbent properties, and operational conditions, into higher dimensions where linear separation or regression becomes feasible. In the context of adsorption, SVM excels at modeling complex sorption mechanisms and capturing multivariate dependencies between process parameters and adsorption capacity [46]. Its robustness against overfitting and strong generalization ability makes it a key component of the hybrid framework. Enhancing SVMs with kernels improves their flexibility and simplifies handling nonlinear tasks. The equations defining the prediction function for the target output differ between a linear support vector machine and a nonlinear SVM (including quadratic and cubic variations) [47].

$$y(x) = \sum_{u=1}^{U} (a_u - a_u^*)(x_u', x) + b \tag{7}$$

$$y(x) = \sum_{u=1}^{U} (a_u - a_u^*) P(x_u', x) + b \tag{8}$$

where, x_u = observed predictors, a_u and a_u^* = non-negative multiplier, b = random error term, P = Gram matrix of size u^*u that consists of elements $p_{i,j} = P(x_i, x_j)$.

2.5. Optimization strategies

To further refine model performance and identify optimal parameter configurations, Bayesian optimization (BO) was implemented as the optimization strategy. BO is a data-efficient, probabilistic approach that iteratively explores the hyperparameter space to locate the best-performing model settings with minimal computational cost [48]. Unlike exhaustive search methods such as grid or random search, BO constructs a surrogate model of the objective function and continuously updates it using prior evaluations [49]. This enables a balanced trade-off between exploration (testing new regions of the parameter space) and exploitation (refining known promising areas). In this study, BO was employed to systematically optimize each ML model's hyperparameters, ensuring convergence toward the global optimum for adsorption prediction accuracy. This strategy significantly improved predictive performance while maintaining computational efficiency, making it well-suited for complex, nonlinear adsorption datasets. The utilized hyper-tuning parameters for selected ML models are provided in Table S2.

2.6. Production of biochar

In this study, biochar was produced from jojoba residues at different temperatures and heating times to experimentally validate the developed ML models. The jojoba residues were obtained after pressing oil from jojoba seeds, by crushing and sieving through a 500-μm mesh. The residues were pyrolyzed using a muffle furnace at a temperature of (500–800) °C, heating time (2–4) h, and a fixed heating rate of 10 °C/min (Fig. 3). The produced biochar was then tested for the adsorption of anionic dyes.

3. Results and discussion

The dataset was used to train and evaluate the selected models, as shown in Fig. 4. Four machine learning models, namely DT, GPR, SVM, and ELT, were applied and fine-tuned using Bayesian Optimization, with the resulting performance metrics summarized in Table 1. Of the four models, DT performed worst, with poor predictive accuracy and limited explanatory power on both the training and test sets, suggesting underfitting. GPR fared somewhat better, though it still accounted for only around 33% of the variance in the target variable. SVM told a similar story, explaining roughly 22% of the variance while also showing signs of overfitting. It performed noticeably better on the training data than on unseen test data, suggesting it had not generalized well. The ELT model stood apart from the rest, achieving the highest R^2 on the test dataset and showing a more consistent performance gap between training and test results, indicating that overfitting was kept in check. ELT proved to be the most reliable of the four models, capturing the underlying patterns in the data more effectively and generalizing better to new observations. The broad range of evaluation metrics clearly indicates a generalized model development, limiting the risk of

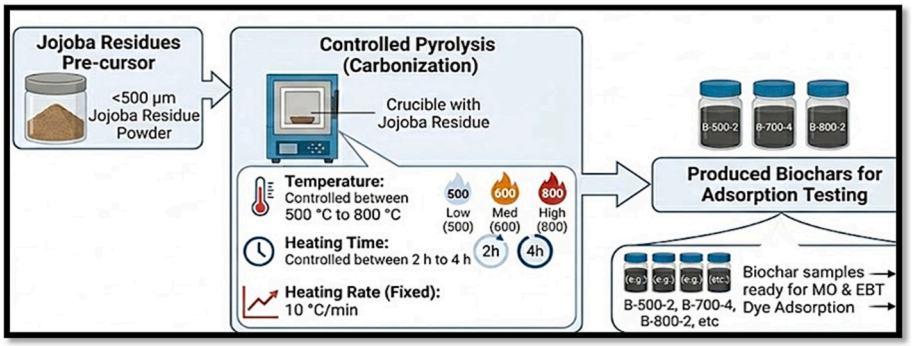


Fig. 3. Production of biochar from jojoba residues.

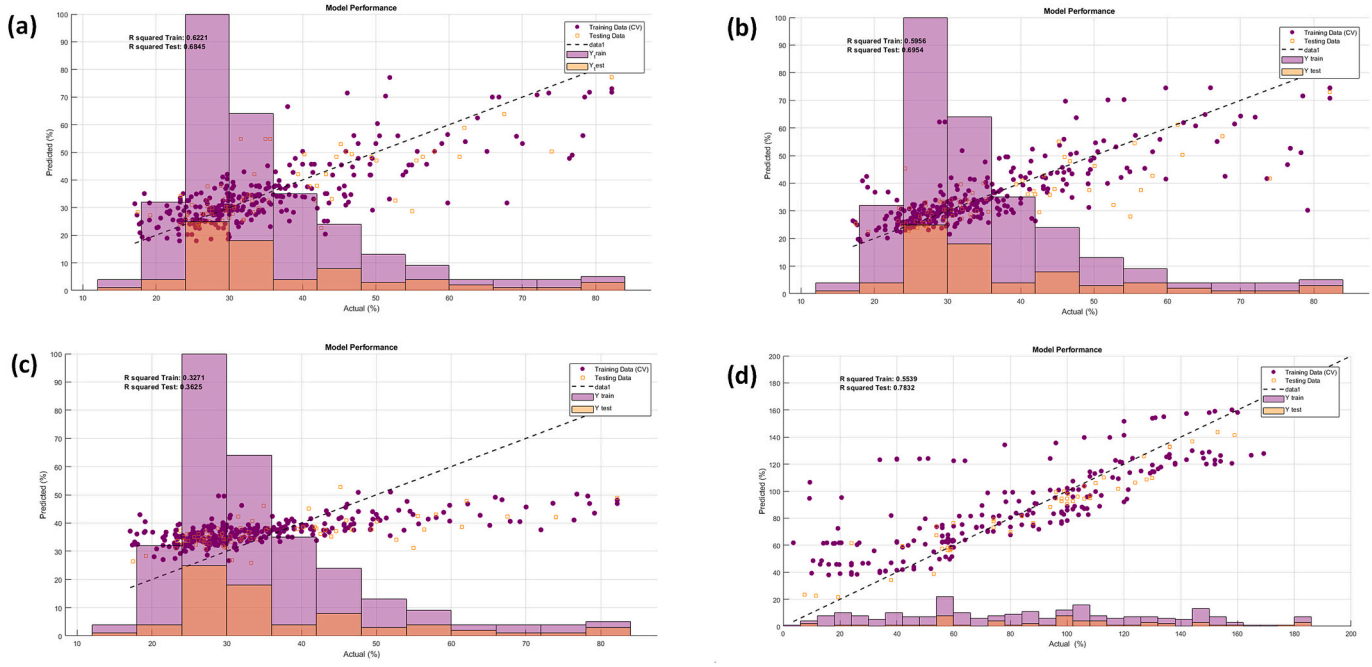


Fig. 4. Model training and testing dataset optimized utilizing four different ML models: (a) DT, (b) GPR, (c) SVM, and (d) ELT.

Table 1
Model training and testing dataset utilizing four different ML models.

Models		Before optimization		After optimization		Performance metrics			
		R^2	RMSE	R^2	RMSE	MAE	EVS	MAPE	RSR
DT	Train	0.1212	17.251	0.1816	11.608	7.9411	0.18161	22.867	0.90313
	Test	0.2175	18.214	0.1595	13.639	9.6844	0.1743	24.227	0.91055
GPR	Train	0.1254	15.257	0.3345	10.467	7.234	0.33479	20.822	0.81441
	Test	0.2954	17.8154	0.33377	12.143	8.4353	0.35124	21.125	0.81069
SVM	Train	0.1595	13.254	0.2236	11.306	8.5821	0.2364	26.324	0.87966
	Test	0.1621	17.215	0.22371	13.108	10.382	0.22526	27.624	0.8751
ELT	Train	0.4487	84.2147	0.56672	76.433	34.379	0.56672	48.528	0.65696
	Test	0.6421	45.251	0.78345	38.472	23.077	0.79312	21.882	0.4169

overfitting. 5-fold cross-validation ensured minimal risk of overfitting. Lower values of RMSE and MAE relate to the minimal error in predictions. While reliable performance was corroborated by higher values of EVS, $R^{2,}$ and MAPE [50].

To evaluate the predictive performance of the four machine learning models, a comprehensive set of metrics was employed, including MAE, RMSE, EVS, MAPE, and RSR. Among all models, ELT consistently returned the most favorable values across every metric. The low MAE and RMSE values recorded for ELT indicate that its predictions stayed close to the experimentally measured adsorption capacities, with minimal deviation, a quality that matters greatly when the goal is to reliably estimate real-world behavior. The EVS score approaching 1.0 further reinforces this, confirming that the model was able to account for the vast majority of variability present in the data rather than leaving it unexplained. When looking at MAPE, ELT's test value of 21.9% was the lowest of the group, meaning its percentage error on unseen data was not only acceptable but clearly better than what DT, GPR, or SVM could offer. Perhaps most telling is the RSR value of 0.42, falling comfortably below the 0.50 benchmark widely recognized in the modeling literature as the threshold for a good fit. Taken together, these metrics do not just

confirm ELT as the top-performing model in a statistical sense; they paint a consistent picture of a model that genuinely understands the patterns driving adsorption in this system and can translate that understanding into accurate, trustworthy predictions for conditions it has never encountered before.

3.1.1. 2D partial dependence plots

Machine Learning Models applied to the tabular dataset can be explained using Partial dependence plots (PDPs). They provide information on the impact of the input feature or features on the prediction of the output feature or features [51].

In this regard, 2D- PDPS of the input features (BET surface area, dosage of anionic dyes and proteins, equilibrium concentration, initial concentration, filtration material, pH of the solution, solution volume, and time) are developed using the ELT model against the output features (Kinetics Qe and adsorption Capacity) as shown in Fig. 5. A decreasing trend is observed in both adsorption Capacity and Kinetics Q_e with the increase in the BET surface area to about 600 m^2/g after which it starts to increase. Adsorption capacity of the adsorbent and kinetics Q_e of materials are not solely dependent on the BET surface area, but also depend on the surface chemistry, molecule size of the anionic dyes and proteins, and surface morphology of the adsorbent material. The U-shaped plot for this well describes the above phenomenon [52]. Adsorption capacity shows an oscillating trend with the increase in the dosage of the pollutants. Initially, at low pollutant dosage, there are abundant adsorption sites, but as the dosage increases, these sites become saturated with pollutants. The availability of the multilayer adsorbent enhances adsorption capacity at higher pollutant dosages, and similarly, Qe Kinetics is affected in the same way [53]. PDP generated using the ELT ML model describes the same behavior. A non-monotonic trend is observed in the PDP of the initial concentration against the kinetics Qe. Raising the equilibrium concentration increases Qe to about 20 mg/L, after which it declines due to the abundance of adsorbent sites [28]. An inverse relationship is observed between the adsorption capacity and the equilibrium concentration. A linear trend is observed in a PDP of initial concentration and Kinetics Qe aligned with the established understanding of enhanced mass transfer driving force and abundance of the dye molecules. While adsorption capacity shows oscillatory behavior, this arises from the interactions among the complex features captured by the model. This is the common trend that can be observed in water treatment studies where Kinetics Qe is enhanced by increasing the initial concentration and maintaining the adsorption capacity till saturation [53]. The dependence of adsorption behavior on solution pH is revealed by the PDP. At high acidic conditions (~pH 2–3), the model predicts higher Kinetics Qe values, indicating enhanced electrostatic attraction between the anionic dye molecules and the positively charged adsorbent surface. Further increasing the pH generates negative surface charges that repel the anionic dye ions. Oscillatory peaks in adsorption capacity around pH 4, 6, and 8 indicate optimal charge-solution conditions, followed by a decline under strong alkaline conditions [54]. The PDP observes a negative effect of increasing solution volume on adsorption capacity. Adsorption capacity starts just above 90% at 30 mL of the solution volume and decreases gradually to form a plateau around 60 to 75 mL of solution volume, and drops sharply to 65% capacity after 85 mL of Solution volume. Kinetics Qe behaves in a similar fashion, dropping from around 200 mg/g to 90 mg/g by increasing the solution volume from 30 mL to 100 mL. This is due to dilution-enhanced reduction in the driving force of mass transfer, leading to fewer dye molecules available for the per unit mass of the adsorbent.

It is worth noting that the oscillatory behavior and localized peaks observed in certain PDPs, most notably for dosage (Fig. 5b) and pH (Fig. 5e), are not physical anomalies but rather a mathematically predictable outcome of how the ELT model structures its predictions. As a tree-based ensemble, ELT partitions the feature space through a series of piecewise-constant decision boundaries. When the PDP marginalizes predictions across all other features at a fixed value of the variable of interest, the superposition of these boundaries produces constructive and destructive interference patterns, manifesting visually as peaks and oscillations in the averaged response curve. In the case of pH, this

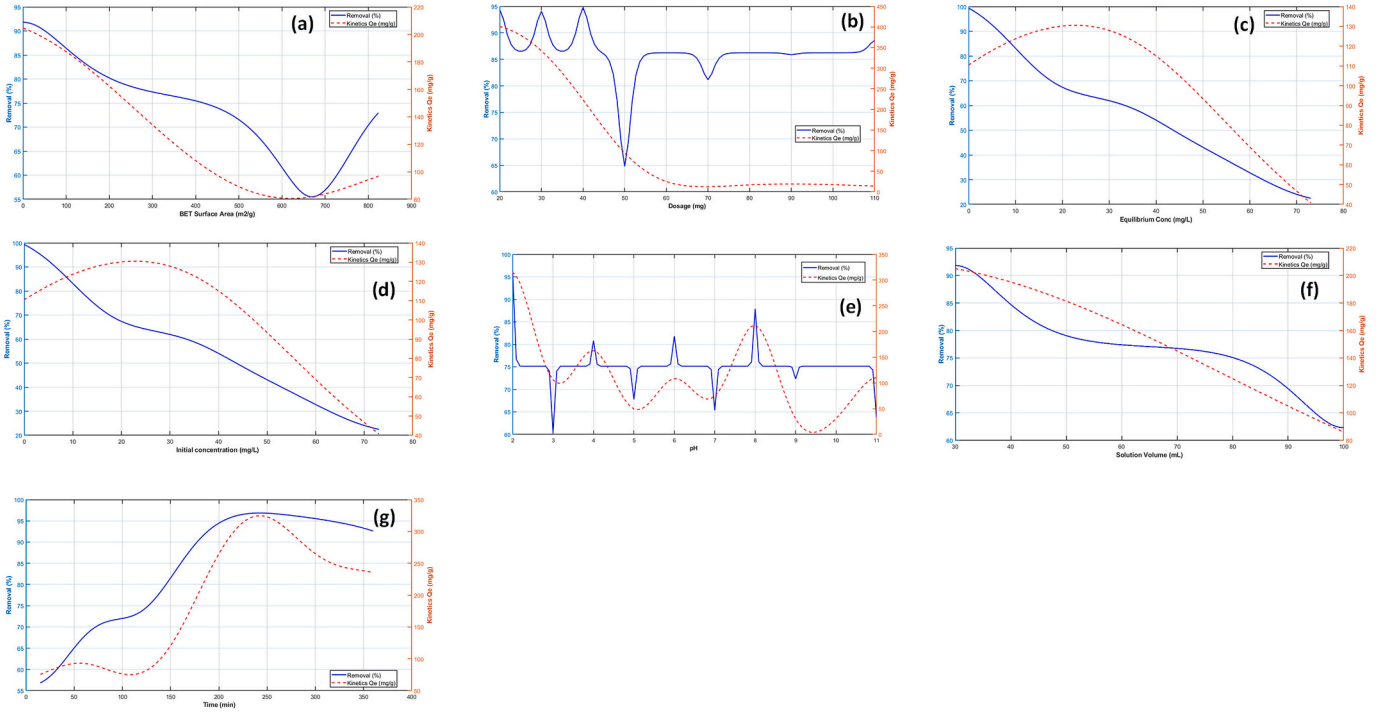


Fig. 5. 2D PDPs of adsorption removal vs experimental conditions: BET surface area (a), dosage (b), equilibrium concentration (c), initial concentration (d), pH (e), solution volume (f), and time (g).

behavior is further amplified by the inherently discontinuous nature of adsorption chemistry; protonation and deprotonation of biochar surface functional groups occur at discrete pH thresholds, introducing genuine abrupt transitions into the training data that the model faithfully reproduces. For dosage, the effect is driven by strong nonlinear cross-interactions with concentration variables; when these correlated features are marginalized simultaneously, the resulting average surface inherits local maxima and minima that reflect feature interdependence rather than noise. These findings are consistent with the theoretical limitations of PDPs under correlated feature conditions, as documented in the broader machine learning interpretability literature. Readers should therefore interpret these regions of the PDPs as indicators of complex, nonlinear feature interactions captured by the model rather than as evidence of instability or poor fit.

3.1.2. 3D partial dependence plots

Similarly, 3D PDP plots were also obtained as shown in Fig. S4 (a-f). From the results, it can be inferred that adsorption % is more sensitive to pH, adsorbent dosage, and BET surface area due to synergistic interactions. 3D PDPs show that the increasing trend of removal efficiency is observed with higher dosage because of the availability of more active sites, but as soon particle starts to saturate, diminishing returns [55]. Effect of pH on adsorption % and dosage was also studied using 3D PDP which reveal pH crucial role in optimizing electrostatic condition for enhanced removal even at moderate dosages while with poor pH selection could also results in degraded removal at higher dosages as well partial dependence of adsorption % on BET surface area and pH was developed, high surface area improves adsorption % due to excess binding sites and it could be enhanced using optimal pH window. Surface charge can be changed by maximizing the pH, which may render ineffective the high surface area [56].

3.1.3. Sensitivity

Sobol Global Sensitivity analyses were performed to assess the individual impact of each input feature on the adsorption capacity and kinetics Qe. Using this variance-based global sensitivity technique, the disintegration of the total variance of the output features can be evaluated as the individual and relative impacts of input features on the total variance, on a scale of 0 to 1. Direct influence of the input feature is the first-order Sobol index, whereas the total order refers to the interaction-based influence of the input features on the total variance of the Output [57].

Sensitivity patterns of all the input features with individual and relative Sobol index are shown in Fig. 6 (a -b) for adsorption capacity and Kinetics Qe. Initial Concentration is the dominant first-order feature for both adsorption capacity (Qe) and kinetics, suggesting that the availability of active sites, driven by the concentration gradient, controls adsorption efficiency independently. While the other input features show low first order compared to initial concentration, their total order reaches unity, indicating that their interaction with the other features strengthens their influence [58].

Model outputs are further tested for reliability and variability using kernel density estimates (KDEs) of adsorption capacity and kinetics. These KDEs are developed based on separation methods (batch and sonication) and the type of pollutant removed. While the pollutant distinctly affects the distribution of adsorption capacity and kinetics (Qe), the model depends on the type of pollutant being removed (Fig. S5a). Adsorption capacity and Kinetics distribution became narrower and higher for the sonication method, predicting model performance being more concise compared to the batch method (Fig. S5a) [59].

Monte Carlo simulations ($N = 1000$), as shown in Fig. S5b, are used to further quantify the prediction uncertainty by selecting input parameters randomly not exceeding their experimental limits and propagating through the optimized ML model. A narrow distribution (mean ≈ 90%, 95% CI ≈ 85–95%) of the predicted adsorption capacity validates the model's accuracy and reliability. On the other hand, kinetics Qe predictions are broader and right-skewed (mean ≈ 500 mg/g, 95% CI ≈ 100–2000 mg/g), reflecting the high dependence and sensitivity on the input fluctuations. These results suggest that more Kinetics Qe data is required to eliminate the uncertainty, along with affirmation of the prediction robustness of the ML model for adsorption capacity [60].

3.1.4. Experimental Validation

Table 2 shows the experimental and predicted results for the % removal and kinetic adsorption capacity from MO and EBT dyes removal using different jojoba-derived biochar produced using different temperatures (500–800 °C) and heating times (2–4 h). As shown, increasing the pyrolysis temperature from 500 °C to 800 °C results in a significant increase in % removal for both methyl orange (MO) and Eriochrome black T (EBT) dyes. In addition, a slight increase in adsorption performance was observed when the pyrolysis heating time increased from 2 h to 4 h. The increase in adsorption of dyes at higher biochar pyrolysis temperatures is mainly associated with the development of a porous surface structure and an increase in specific surface area, resulting from the complete degradation of jojoba residues [61]. The porous surface serves as active binding sites for dye molecules via pore-filling behavior [62]. Furthermore, the high-temperature biochar (700–800 °C) exhibited high aromaticity, which dominated pi-pi interactions between the biochar's carbon layers and the aromatic rings of the dyes. Specifically, EBT dye exhibits a higher adsorption capacity (Q_e) than MO due to its smaller molecular size, which allows it to penetrate biochar pores more easily via a pore-filling mechanism. Previous studies also indicated that the formation of porous and graphite structures at 700–800 °C promoted pore filling and hydrophobic interactions as dominant adsorption mechanisms [63]. The comparison between experimental data and ML predictions reveals a strong correlation, suggesting that the ML model

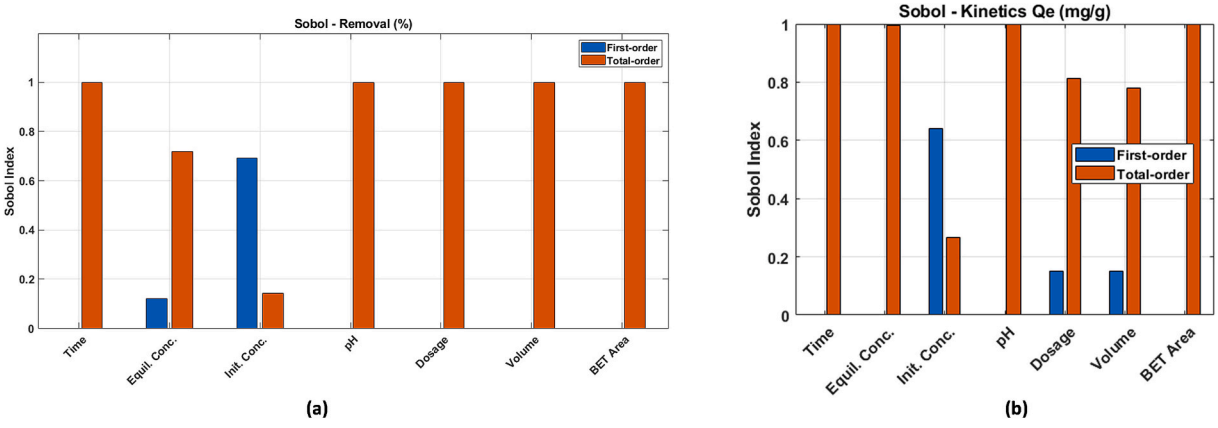


Fig. 6. Sobol sensitivity patterns of input features for adsorption removal capacity and kinetics Q_e.

Table 2
Comparison of ML-predicted results with experimental values at different conditions for removing anionic dyes after 6 h using different jojoba-derived biochars (fixed conditions: pH = 3, temperature = 25 °C, solution volume = 40 mL).

Pollutant	Material	Initial concentration (mg/L)	Dosage (mg)	Experimental		Predicted		% error	
				Removal %	Kinetics Qe (mg/g)	Removal (%)	Kinetics Qe (mg/g)	Removal (%)	Kinetics Qe (mg/g)
EBT	B-700-4	48.88	10.75	25.51	46.40	24.31	45.20	4.71	2.59
MO	B-700-4	38.46	10.71	21.48	30.85	23.51	32.18	9.47	4.31
MO	B-500-2	38.46	10.5	48.28	70.74	50.92	72.88	5.46	3.02
EBT	B-500-2	48.88	10.77	9.80	17.79	10.50	18.81	7.15	5.73
EBT	B-500-4	48.88	10.7	15.22	27.81	14.91	29.15	2.04	4.81
EBT	B-600-2	48.88	10.46	10.11	18.89	9.50	17.16	6.00	9.16
EBT	B-700-2	48.88	10.17	16.76	32.21	16.42	32.73	2.00	1.61
EBT	B-800-2	48.88	10.42	24.71	46.37	24.06	47.53	2.64	2.50
EBT	B-800-4	48.88	10.41	29.66	55.72	31.35	58.48	5.68	4.96
MO	B-500-4	38.46	10.69	24.05	34.61	26.37	31.91	9.64	7.81
MO	B-600-2	38.46	10.52	52.29	76.46	50.51	78.66	3.40	2.87
MO	B-600-4	38.46	10.09	22.05	33.62	23.31	35.15	5.72	4.56
MO	B-700-2	38.46	10.63	32.24	46.66	33.64	45.51	4.34	2.47
MO	B-800-2	38.46	10.18	38.85	58.70	40.12	57.32	3.28	2.36
MO	B-800-4	38.46	10.29	19.68	29.43	21.44	27.81	8.93	5.49

successfully captures the non-linear relationship between production parameters and adsorption efficiency. The ML model's ability to mirror the "plateau" or "steep increase" in experimental % removal indicates that it accurately weights pyrolysis temperature as the most critical feature relative to heating time. Overall, the integration of ML validates the experimental results, confirming that the pyrolysis parameters are reliable predictors for the design of high-performance biochars for effective treatment of dye-contaminated wastewater.

To facilitate the practical implementation of the developed machine learning framework, a graphical user interface (GUI) was designed using MATLAB R2021b, as shown in Fig. 7. The GUI integrates the optimized Ensemble Learning Tree (ELT) model, allowing users to rapidly predict adsorption removal efficiency and adsorption capacity under user-defined experimental conditions. The developed interface provides an accessible platform that eliminates the need for direct interaction with the underlying machine learning code, enabling researchers and engineers to perform rapid assessments of adsorption performance. The GUI was developed using MATLAB's App Designer, which supports interactive user input and real-time prediction using the trained ML model. The interface consists of several input panels corresponding to the key operational parameters used during model training. These parameters include the solution pH, adsorbent dosage, BET surface area of the adsorbent, initial pollutant concentration, equilibrium concentration, solution volume, contact time, and adsorbent material type. Each parameter can be entered manually within the range of values used during model training to ensure reliable predictions. Once the experimental parameters are specified, the GUI processes the input data and passes it to the trained ELT regression model, which computes the predicted adsorption performance. The interface subsequently displays two main outputs, i.e., adsorptive removal (%) and capacity (q_e, mg/g). These predictions are generated instantaneously, enabling users to evaluate how variations in operational parameters influence adsorption performance without conducting time-consuming laboratory experiments. The developed GUI demonstrates the potential of integrating machine learning models with user-friendly software tools to bridge the gap between data-driven modeling and practical environmental engineering applications.

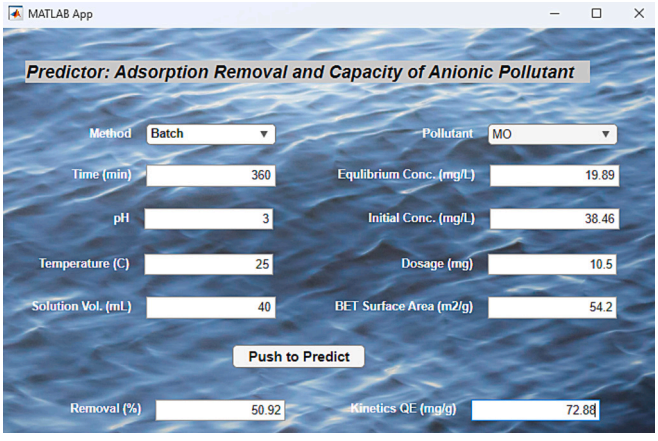


Fig. 7. Screenshot of the GUI program developed for predictive removal of anionic pollutants via selected carbonaceous adsorbents.

4. Conclusion

This study presents a machine learning-assisted strategy for predicting adsorption performance and optimizing operational conditions for the removal of anionic pollutants from aqueous systems. A comprehensive experimental dataset comprising multiple pollutants, adsorbent materials, and operational parameters was used to develop predictive models utilizing Decision Tree, Gaussian Process Regression, Support Vector Machine, and Ensemble Learning Tree algorithms. Bayesian optimization significantly improved model performance, with the optimized Ensemble Learning Tree demonstrating the highest predictive accuracy and the strongest capability to capture nonlinear adsorption behavior. A robust and generalized model architecture was systematically developed with feature selection and hyperparameter tuning. Independent test and training datasets yielded different evaluation metrics. Which ensured minimal overfitting on the consolidated dataset. Overfitting of the dataset in the developed model was minimized by applying 5-fold cross-validation. Model interpretation through partial dependence analysis and Sobol sensitivity assessment revealed that adsorption performance is primarily governed by initial pollutant concentration, pH, adsorbent dosage, and BET surface area. These analyses provide valuable insights into the complex interactions between operational conditions and adsorption efficiency. The reliability of the developed ML framework was further validated experimentally using jojoba-derived biochar for the removal of methyl orange and Eriochrome Black T dyes, where predicted results showed strong agreement with experimental observations. The integration of experimental data, advanced machine learning modeling, and uncertainty analysis demonstrates a robust predictive framework for adsorption-based water treatment systems. Furthermore, the development of a graphical user interface enables rapid estimation of adsorption performance under different conditions, offering a practical decision-support tool for process design. This work highlights the potential of data-driven

approaches to accelerate the development and optimization of efficient adsorbent materials for wastewater remediation.

CRediT authorship contribution statement

Hamza Ul Haq: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Muhammad Yasir:** Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation. **Muhammad Nouman Aslam Khan:** Writing – review & editing, Supervision, Software, Formal analysis, Data curation. **Jawad Gul:** Writing – review & editing, Data curation. **Mukarram Zubair:** Validation, Investigation, Formal analysis. **Hassan Ali:** Writing – review & editing, Investigation. **Vladimir Sedlarik:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Nasir.M. Ahmad:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Vladimir Sedlarik reports financial support was provided by Ministry of Education Youth and Sports of the Czech Republic. Vladimir Sedlarik reports financial support was provided by Ministry of the Environment of the Czech Republic. Hassan Ali reports financial support was provided by Ministry of Education Youth and Sports of the Czech Republic. Nasir M. Ahmad reports financial support was provided by Higher Education Commission Pakistan. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are thankful to the National University of Sciences and Technology (NUST) Research Directorate, HEC, and NRPU through Project No. 6020 for all the technical assistance and financial support. The authors also gratefully acknowledge the support from the European Just Transition Fund within the Operational Programme: Just Transition under the aegis of the Ministry of the Environment of the Czech Republic, project CirkArena number CZ.10.03.01/00/22_003/0000045 and the Ministry of Education Youth and Sports of the Czech Republic, Operational Programme Johannes Amos Comenius OP JAC "Application potential development in the field of polymer materials in the context of circular economy compliance (POCEK)", number CZ.02.01.01/00/23_021/0009004. The authors are further grateful for *co*-funding from the development process of the Centre of Polymer Systems, Tomas Bata University in Zlin, program DKRVO (RP/CPS/2024-28/002) and (RP/CPS/2024-28/007), supported by the Ministry of Education, Youth and Sports of the Czech Republic. Muhammad Yasir also expresses his gratitude for support within the "Creativity, Intelligence & Talent for the Zlin Region" (CIT - ZK) program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.seppur.2026.139012.

References

[1] P. Babuji, S. Thirumalaisamy, K. Duraisamy, G. Periyasamy, Human health risks due to exposure to water pollution: a review, Water 15 (2023) 2532, https://doi.org/10.3390/w15142532.
[2] A. Gezahegn, L.B. Merga, S. Mammo, Industrial effluents caused environmental pollution and its potential ecological and human health impacts in Ethiopia: a review, Waste Management Bulletin 3 (2025) 100240, https://doi.org/10.1016/j.wmb.2025.100240.
[3] K. Sathya, K. Nagarajan, G. Carlin Geor Malar, S. Rajalakshmi, P. Raja Lakshmi, A comprehensive review on comparison among effluent treatment methods and modern methods of treatment of industrial wastewater effluent from different sources, Appl Water Sci 12 (2022) 70, https://doi.org/10.1007/s13201-022-01594-7.
[4] G.M. Filippelli, M.P. Taylor, Addressing pollution-related global environmental health burdens, GeoHealth 2 (2018) 2–5, https://doi.org/10.1002/2017GH000119.
[5] G. Yumnam, Y. Gyanendra, C.I. Singh, A systematic bibliometric review of the global research dynamics of United Nations sustainable development goals 2030, Sustainable Futures 7 (2024) 100192, https://doi.org/10.1016/j.sftr.2024.100192.
[6] S. Duan, H. Dong, C. Jiang, H. Liang, L. Jiang, Q. Xu, X. Cheng, Z. Qiang, Removal of particulate matter and dissolved organic matter from sedimentation sludge water during pre-sedimentation process: performances and mechanisms, J. Environ. Sci. 148 (2025) 409–419, https://doi.org/10.1016/j.jes.2023.11.026.
[7] G. Zhao, H. Gao, Z. Qu, H. Fan, H. Meng, Anhydrous interfacial polymerization of sub-1 Å sieving polyamide membrane, Nat. Commun. 14 (2023) 7624, https://doi.org/10.1038/s41467-023-43291-2.
[8] S. Jamil, P. Loganathan, J. Kandasamy, A. Listowski, J.A. McDonald, S.J. Khan, S. Vigneswaran, Removal of organic matter from wastewater reverse osmosis concentrate using granular activated carbon and anion exchange resin adsorbent columns in sequence, Chemosphere 261 (2020) 127549, https://doi.org/10.1016/j.chemosphere.2020.127549.
[9] E. Topkaya, A. Arslan, S. Veli, S. Kuru, M. Sezer, Removal of microplastics by ozone oxidation from urban wastewater using Taguchi experimental design, Water Air Soil Pollut. 236 (2025) 606, https://doi.org/10.1007/s11270-025-08249-8.
[10] S. Chen, J. Deng, L. Wang, X. Ma, M. Xu, D. Xu, H. Wang, W. Shi, Comparison of vacuum ultraviolet-based oxidation as ultrafiltration pretreatment in real water: performance, risk assessment, and mechanistic insights, Water Res. 298 (2026) 125801, https://doi.org/10.1016/j.watres.2026.125801.
[11] J.-H. Joo, S.-H. Kim, J.H. Kim, H.-J. Kang, J.H. Lee, H.-J. Jeon, Y.H. Jang, J.-H. Lee, S.-Y. Lee, S.-J. Park, M.-K. Seo, Recent advances in activated carbon fibers for pollutant removal, Carbon Lett. 35 (2025) 21–44, https://doi.org/10.1007/s42823-024-00803-4.
[12] S.M. Chistie, S.U. Naik, P. Rajendra, R.K. Apeksha, G. Mishra, S. Albasher, G. P. Chinnam, Z. Jeppu, J. Hameed Arif, Production and characterization of magnetic biochar derived from pyrolysis of waste areca nut husk for removal of methylene blue dye from wastewater, Sci. Rep. 15 (2025) 23209, https://doi.org/10.1038/s41598-025-03359-z.
[13] X. Li, Y. Li, Y. Li, M. Zhang, J. Zhu, Synthesis of lignin-derived hierarchical porous carbon via hydrothermal–phosphoric acid synergistic activation for enhanced adsorption of tetracycline, Molecules 31 (2026) 447, https://doi.org/10.3390/molecules31030447.
[14] R. Shahi, M. Khatamian, Synthesis of magnetic borosilicate zeolite/graphene quantum dots nanocomposites for removal of nitrate and organic pollutants from water, Sci. Rep. 15 (2025) 20726, https://doi.org/10.1038/s41598-025-07746-4.
[15] L.N. Munuhe, E.S. Madivoli, D.M. Nzilu, P.N. Lemeitaron, P.K. Kimani, Advances in adsorbent materials for pharmaceutical pollutant removal: a review of occurrence, fate, and state-of-the-art remediation, J. Chemother. 2025 (2025) 4477822, https://doi.org/10.1155/joch/4477822.
[16] X. Li, R. Ran, W. Zeng, S. Ren, Y. Wang, A. Zhu, C. Zheng, Coffee shell as a green reductant for iron recovery from pyrite cinder: synergy of biomass-driven reduction and acid leaching, Miner. Eng. 234 (2025) 109747, https://doi.org/10.1016/j.mineng.2025.109747.
[17] D. Wang, W. Hu, Y. Zheng, Z. Xia, H. Cheng, X. Lu, Z. He, S. Giannakis, S. Song, J. Ma, Mechanistic insights into hexamethylphosphoramide degradation and phosphorus recovery via mZVI/O3: beyond conventional oxidation through iron-phosphorus speciation, Chem. Eng. J. 538 (2026) 176658, https://doi.org/10.1016/j.cej.2026.176658.
[18] R.B. Phillips, R.R. James, M.L. Magnuson, Functional categories of microbial toxicity resulting from three advanced oxidation process treatments during management and disposal of contaminated water, Chemosphere 238 (2020) 124550, https://doi.org/10.1016/j.chemosphere.2019.124550.
[19] L. Huang, X. Jiao, J. Lin, J. Hao, L. Tan, D. Zeng, S. Wang, Y. Lu, X. Shen, Z. Zhang, F. Yan, EGDE-induced crosslinking of PEI@D500 adsorbents for direct air capture: enhanced hydrothermal and oxidative stability, Chem. Eng. J. 537 (2026) 176243, https://doi.org/10.1016/j.cej.2026.176243.
[20] W. Ashraf, D. Sharma, A. Khan, M. Khanuja, Green carbon dots as adsorbents for removal of toxic chemicals, in: S.U. Islam, C.M. Hussain (Eds.), ACS Symposium Series, American Chemical Society, Washington, DC, 2025, pp. 141–160, https://doi.org/10.1021/bk-2025-1506.ch006.
[21] P. Waribam, T.R. Katugampalage, M. Ogawa, P. Opaprakasit, W. Chooaksorn, P. Sooksaen, P. Puathawee, P. Sreearunothai, Magnetic metal–organic frameworks (MOFs) from waste: a solvent-free rapid synthesis of green catalyst for environmental cleanup, ACS Sustain. Chem. Eng. 13 (2025) 9576–9587, https://doi.org/10.1021/acssuschemeng.5c02222.
[22] R. Natarajan, K. Saikia, S.K. Ponnusamy, A.K. Rathankumar, D.S. Rajendran, S. Venkataraman, D.B. Tannani, V. Arvind, T. Somanna, K. Banerjee, N. Mohideen, V.K. Vaidyanathan, Understanding the factors affecting adsorption of pharmaceuticals on different adsorbents – a critical literature update, Chemosphere 287 (2022) 131958, https://doi.org/10.1016/j.chemosphere.2021.131958.
[23] R. Li, X. Cao, X. Fan, J. Shi, B. Meng, J. Zhang, Y. Wang, J. Du, X. Deng, C. Zheng, Study of the factors influencing the adsorption of heavy metal pollutants in water by activated carbon gel particles, J. Environ. Eng. 150 (2024) 04024024, https://doi.org/10.1061/JOEEDU.EEENG-7606.

[24] P.B. Joshi, Navigating with chemometrics and machine learning in chemistry, Artif. Intell. Rev. 56 (2023) 9089–9114, https://doi.org/10.1007/s10462-023-10391-w.
[25] A.K. Cheetham, R. Seshadri, Artificial intelligence driving materials discovery? Perspective on the article: scaling deep learning for materials discovery, Chem. Mater. 36 (2024) 3490–3495, https://doi.org/10.1021/acs.chemmater.4c00643.
[26] N. Li, S. Girhe, M. Zhang, B. Chen, Y. Zhang, S. Liu, H. Pitsch, A machine learning method to predict rate constants for various reactions in combustion kinetic models, Combust. Flame 263 (2024) 113375, https://doi.org/10.1016/j.combustflame.2024.113375.
[27] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, A. Borodulin, Optimizing neural networks for chemical reaction prediction: insights from methylene blue reduction reactions, IJMS 25 (2024) 3860, https://doi.org/10.3390/ijms25073860.
[28] H. Ali, M. Yasir, H.U. Haq, A.C. Guler, M. Masar, M.N.A. Khan, M. Machovsky, V. Sedlarik, I. Kuritka, Machine learning approach for photocatalysis: an experimentally validated case study of photocatalytic dye degradation, J. Environ. Manag. 386 (2025) 125683, https://doi.org/10.1016/j.jenvman.2025.125683.
[29] H. Ullah, S. Khan, X. Zhu, B. Chen, Z. Rao, N. Wu, A.M. Idris, Machine learning-aided biochar design for the adsorptive removal of emerging inorganic pollutants in water, Sep. Purif. Technol. 362 (2025) 131421, https://doi.org/10.1016/j.seppur.2025.131421.
[30] M. Vera, J. Aguilar, S. Coronel, D. Juela, E. Vanegas, C. Cruzat, Machine learning for the adsorptive removal of ciprofloxacin using sugarcane bagasse as a low-cost biosorbent: comparison of analytic, mechanistic, and neural network modeling, Environ. Sci. Pollut. Res. 31 (2024) 48674–48686, https://doi.org/10.1007/s11356-024-34345-z.
[31] A.A. Akinpelu, M.K. Nazal, M. Shafiullah, M.K. Islam, M.M. Islam, A. Rahman, S. M. Rahman, M.M. Rahman, A multivariate machine learning model of adsorptive Lindane removal from contaminated water, Appl. Sci. 13 (2023) 7086, https://doi.org/10.3390/app13127086.
[32] J. Li, J. Wang, H. Mu, H. Hu, J. Wang, H. Ren, B. Wu, Prediction of adsorptive activities of MOFs for pollutants in aqueous phase based on machine learning, ACS EST Eng. 3 (2023) 1258–1266, https://doi.org/10.1021/acsestengg.3c00086.
[33] H. Ullah, S. Khan, B. Chen, A. Shahab, L. Riaz, L. Lun, N. Wu, Machine learning approach to predict adsorption capacity of Fe-modified biochar for selenium, Carbon Res. 2 (2023) 29, https://doi.org/10.1007/s44246-023-00061-5.
[34] E. Abbasi, M.R. Alavi Moghaddam, E. Kowsari, A systematic and critical review on development of machine learning based-ensemble models for prediction of adsorption process efficiency, J. Clean. Prod. 379 (2022) 134588, https://doi.org/10.1016/j.jclepro.2022.134588.
[35] A. Gupta, K. Ghanshala, R.C. Joshi, Machine learning classifier approach with Gaussian process, ensemble boosted trees, SVM, and linear regression for 5G signal coverage mapping, IJIMAI 6 (2021) 156–163, https://doi.org/10.9781/ijimai.2021.03.004.
[36] E. Syguła, K. Piasecka, J. Łyczko, A. Białowiec, The potential of decision tree application in threshold analysis of hazardous volatile organic compound release from biochar: implications for environmental risk assessment, Sci. Total Environ. 998 (2025) 180252, https://doi.org/10.1016/j.scitotenv.2025.180252.
[37] K. Kim, J. Hong, A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis, Pattern Recogn. Lett. 98 (2017) 39–45, https://doi.org/10.1016/j.patrec.2017.08.011.
[38] J. Li, W. Zhang, T. Liu, L. Yang, H. Li, H. Peng, S. Jiang, X. Wang, L. Leng, Machine learning aided bio-oil production with high energy recovery and low nitrogen content from hydrothermal liquefaction of biomass with experiment verification, Chem. Eng. J. 425 (2021) 130649, https://doi.org/10.1016/j.cej.2021.130649.
[39] I. Bibers, O. Arreche, W. Alayed, M. Abdallah, Ensemble-IDS: an ensemble learning framework for enhancing AI-based network intrusion detection tasks, Appl. Sci. 15 (2025) 10579, https://doi.org/10.3390/app151910579.
[40] D.K. Jha, Smart screening of hydrogen storage capacities in MOFs using a tailored machine learning, Next Energy 9 (2025) 100431, https://doi.org/10.1016/j.nxener.2025.100431.
[41] J. Gul, M.N.A. Khan, U. Sikander, A.H. Khoja, M. Kah, S.R. Naqvi, Machine learning optimization for algal biochar yield: integrating experimental validation and sensitivity analysis, Biochar 8 (2026) 8, https://doi.org/10.1007/s42773-025-00511-w.
[42] V.L. Deringer, A.P. Bartók, N. Bernstein, D.M. Wilkins, M. Ceriotti, G. Csányi, Gaussian process regression for materials and molecules, Chem. Rev. 121 (2021) 10073–10141, https://doi.org/10.1021/acs.chemrev.1c00022.
[43] Z.M. Yaseen, M.M. Hameed, Z.H. Doost, Validation of machine learning models for heavy metals bioavailability prediction: a comparative study, J. Environ. Chem. Eng. 13 (2025) 116749, https://doi.org/10.1016/j.jece.2025.116749.
[44] E. Momeni, M.B. Dowlatshahi, F. Omidinasab, H. Maizir, D.J. Armaghani, Gaussian process regression technique to estimate the pile bearing capacity, Arab. J. Sci. Eng. 45 (2020) 8255–8267, https://doi.org/10.1007/s13369-020-04683-4.
[45] X. Li, W. Sun, C. Qin, Y. Yan, L. Zhang, J. Tu, Evaluation of supervised machine learning regression models for CFD-based surrogate modelling in indoor airflow field reconstruction, Build. Environ. 267 (2025) 112173, https://doi.org/10.1016/j.buildenv.2024.112173.
[46] S. Emamgholizadeh, B. Mohammadi, New hybrid nature-based algorithm to integration support vector machine for prediction of soil cation exchange capacity, Soft. Comput. 25 (2021) 13451–13464, https://doi.org/10.1007/s00500-021-06095-4.
[47] A. Bahadar, R. Kanthasamy, H.H. Sait, M. Zwawi, M. Algarni, B.V. Ayodele, C. K. Cheng, L.J. Wei, Elucidating the effect of process parameters on the production of hydrogen-rich syngas by biomass and coal co-gasification techniques: a multi-criteria modeling approach, Chemosphere 287 (2022) 132052, https://doi.org/10.1016/j.chemosphere.2021.132052.
[48] R. Guay-Hottin, L. Kardassevitch, H. Pham, G. Lajoie, M. Bonizzato, Robust prior-biased acquisition function for human-in-the-loop Bayesian optimization, Knowl.-Based Syst. 311 (2025) 113039, https://doi.org/10.1016/j.knosys.2025.113039.
[49] B. Lei, T.Q. Kirk, A. Bhattacharya, D. Pati, X. Qian, R. Arroyave, B.K. Mallick, Bayesian optimization with adaptive surrogate models for automated experimental design, npj Comput. Mater. 7 (2021) 194, https://doi.org/10.1038/s41524-021-00662-x.
[50] M.A. Haider, J. Gul, M.N.A. Khan, D. Juchelkova, Sustainable valorization of olive pomace industrial side stream via explainable machine learning guided biochar pyrolysis, Process. Saf. Environ. Prot. 210 (2026) 108639, https://doi.org/10.1016/j.psep.2026.108639.
[51] D. Kerrigan, B. Barr, E. Bertini, PDPilot: exploring partial dependence plots through ranking, filtering, and clustering, IEEE Trans. Vis. Comput. Graph. 31 (2025) 7377–7390, https://doi.org/10.1109/TVCG.2025.3545025.
[52] I. Karume, S. Bbumba, S. Tewolde, I.Z.T. Mukasa, M. Ntale, Impact of carbonization conditions and adsorbate nature on the performance of activated carbon in water treatment, BMC Chem. 17 (2023) 162, https://doi.org/10.1186/s13065-023-01091-1.
[53] N. Hamri, A. Imessaoudene, A. Hadadi, S. Cheikh, A. Boukerroui, J.-C. Bollinger, A. Amrane, H. Tahraoui, H.N. Tran, A.O. Ezzat, H.A. Al-Lohedan, L. Mouni, Enhanced adsorption capacity of methylene blue dye onto kaolin through acid treatment: batch adsorption and machine learning studies, Water 16 (2024) 243, https://doi.org/10.3390/w16020243.
[54] W. Zhu, F. Chen, L. Ye, X. Yang, Y. Song, H. Wang, Methylene blue and acid red adsorption on biochar made from modified sugarcane bagasse: a dynamic, equilibrium, and thermodynamic investigation, Adsorpt. Sci. Technol. 42 (2024) 02636174241273522, https://doi.org/10.1177/02636174241273522.
[55] Y. Luo, Z. Zheng, P. Wu, Y. Wu, Effect of different direct revegetation strategies on the mobility of heavy metals in artificial zinc smelting waste slag: implications for phytoremediation, Chemosphere 286 (2022) 131678, https://doi.org/10.1016/j.chemosphere.2021.131678.
[56] M. Bellaj, K. Aziz, M. El Achaby, M. El Haddad, L. Gebrati, T.A. Kurniawan, Z. Chen, P.-S. Yap, F. Aziz, Cationic and anionic dyes adsorption from wastewater by clay-chitosan composite: an integrated experimental and modeling study, Chem. Eng. Sci. 285 (2024) 119615, https://doi.org/10.1016/j.ces.2023.119615.
[57] W. Zouhri, L. Homri, J.-Y. Dantan, Handling the impact of feature uncertainties on SVM: a robust approach based on Sobol sensitivity analysis, Expert Syst. Appl. 189 (2022) 115691, https://doi.org/10.1016/j.eswa.2021.115691.
[58] H. Zheng, X. Meng, Y. Yang, J. Chen, S. Huo, Bifunctional photocatalytic nanofiltration membranes with immobilized BaTiO3/Ti3C2T catalysts for the simultaneous separation and degradation of azo compounds, J. Environ. Chem. Eng. 11 (2023) 110064, https://doi.org/10.1016/j.jece.2023.110064.
[59] J. Latif, N. Chen, A. Saleem, K. Li, J. Qin, H. Yang, H. Jia, Machine learning for persistent free radicals in biochar: dual prediction of contents and types using regression and classification models, Carbon Res. 3 (2024) 39, https://doi.org/10.1007/s44246-024-00125-0.
[60] S. Chheda, W. Jeong, N. Hanikel, L. Gagliardi, J.I. Siepmann, Monte Carlo simulations of water adsorption in aluminum oxide rod-based metal–organic frameworks, J. Phys. Chem. C 127 (2023) 7837–7851, https://doi.org/10.1021/acs.jpcc.3c00354.
[61] M. Zubair, N.D. Mu'azu, N. Jarrah, N.I. Blaisi, H.A. Aziz, M.A. Al-Harthi, Adsorption behavior and mechanism of methylene blue, crystal violet, Eriochrome black T, and methyl orange dyes onto biochar-derived date palm fronds waste produced at different pyrolysis conditions, Water Air Soil Pollut. 231 (2020) 240, https://doi.org/10.1007/s11270-020-04595-x.
[62] M. Zeeshan, T. Javed, C. Kumari, A. Thumma, M. Wasim, M.B. Taj, I. Sharma, M. N. Haider, M. Batool, Investigating the interactions between dyes and porous/composite materials: a comprehensive study, Sustain. Chem. Environ. 9 (2025) 100217, https://doi.org/10.1016/j.scenv.2025.100217.
[63] S. Qiu, Q. Li, X. Li, J. Ma, L. Wu, X. Xie, L. Wu, S. Askari, N. Dewangan, J. Ashok, X. Gao, S. Kawi, Biomass-derived carbon materials for the adsorption of organic pollutants, Adv. Sustainable Syst. 8 (2024) 2300340, https://doi.org/10.1002/adsu.202300340.