Check for updates

Integrating artificial intelligence with kinetic studies for Cr(VI) removal using young durian fruit biochar: a random forest regressor approach

Duy-Khoi Nguyen, [ID] [ab] Quang-Thanh Nguyen [ab] and Van-Phuc Dinh [ID] *[abc]

This study presents a novel approach to predicting the adsorption kinetics of Cr(VI) using biochar derived from young durian fruit (YDF), integrating artificial intelligence (AI) to overcome limitations of conventional experimental methods. A Random Forest Regressor (RFR) model was developed to predict the adsorption capacity (Q_e) based on key operational parameters, including contact time, pH, biochar dosage, ionic strength, and initial Cr(VI) concentration. The RFR model demonstrated high predictive accuracy and robustness in capturing nonlinear relationships, even under untested conditions. In parallel, ten conventional kinetic models, such as pseudo-first-order (PFO) model, pseudo-second-order (PSO) model, mix-order (MO) model, intraparticle diffusion (IDF) model, vermeulen model, elovic model, Mathews and Weber (M&W) model, boyd's intraparticle diffusion model, Weber and Morris (W&M) model, pore volume and surface diffusion (PVSD) model, were evaluated. Among them, the PSO model exhibited the highest goodness of fit ($R^2 = 0.989$), indicating that the adsorption process is predominantly chemisorption-driven. The random forest regressor (RFR) achieved $R^2 = 0.994$, significantly outperforming conventional kinetic models and enabling robust forecasting under untested scenarios, thereby bridging the gap between mechanistic modeling and AI-enhanced environmental applications. The results confirm that the AI-based model not only reduces the experimental workload but also offers strong generalizability and interpretability for kinetic behavior analysis. This integration of AI and environmental chemistry provides a powerful tool for developing cost-effective and sustainable water treatment systems using bio-based materials.

1. Introduction

The study of adsorption kinetics is pivotal for comprehending the mechanisms and optimizing processes involved in environmental remediation, particularly in the removal of pollutants from aqueous systems.[1] Kinetic analyses not only elucidate the rate and nature (physical or chemical) of adsorption but also provide essential parameters for designing efficient treatment systems.[2] However, traditional kinetic studies often necessitate extensive experimentation under varying conditions such as pollutant concentration, pH, temperature, and adsorbent properties which can be time-consuming, labor-intensive, and costly. Moreover, experimental data may be influenced by unforeseen factors, adding complexity to the analysis and practical application of the results.

In recent years, artificial intelligence (AI), particularly machine learning (ML) techniques, has emerged as a powerful tool for modeling complex, nonlinear systems across various disciplines, including environmental engineering.[3,4] AI algorithms have demonstrated remarkable capabilities in predicting adsorption capacities, optimizing process parameters, and uncovering intricate relationships between variables, thereby reducing reliance on exhaustive experimental procedures.[5] For instance, ensemble learning models such as Random Forest (RF) and Gradient Boosting (GB) have been effectively employed to predict heavy metal adsorption efficiencies based on biochar properties and operational conditions.[6] Additionally, ML approaches have been utilized to model the adsorption kinetics of Cr(VI) onto various adsorbents, achieving high predictive accuracy and offering insights into the adsorption mechanisms.[7]

Despite these advancements, the application of AI in modeling adsorption kinetics, particularly for Cr(VI) removal using biochar derived from agricultural waste, remains underexplored. Most existing studies focus on predicting adsorption capacities or equilibrium parameters, with limited attention to kinetic modeling.[8] Furthermore, the integration of AI with

[a]Institute of Interdisciplinary Sciences (IIS), Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam. E-mail: dvphuc@ntt.edu.vn

[b]Nguyen Tat Thanh University, Center for Hi-Tech Development, Saigon Hi-Tech Park, Ho Chi Minh City 700000, Vietnam

[c]Faculty of Applied Science and Technology, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam

experimental kinetic models to enhance predictive performance and reduce experimental workload is still in its nascent stages.

While machine learning applications have extensively focused on isotherm modeling, their potential in dynamic kinetic modeling remains underexplored, particularly in the context of biochar systems.[9–11] Despite growing interest in applying artificial intelligence (AI) to model adsorption phenomena,[12] existing studies have predominantly focused on equilibrium isotherms rather than dynamic kinetic processes. While numerous machine learning models have been developed to predict maximum adsorption capacities under equilibrium conditions, the temporal dimension of adsorption capturing rate-limiting steps, diffusion mechanisms, and real-time system responses remains largely underexplored. This lack of attention to kinetic modeling limits the practical utility of AI in designing scalable, time-sensitive treatment systems. Therefore, there is a compelling need to develop data-driven approaches that can model adsorption kinetics with high accuracy, interpretability, and flexibility across varying environmental conditions.

This study introduces a novel approach that integrates AI with traditional kinetic modeling to predict the adsorption kinetics of Cr(vi) onto biochar derived from young durian fruit (YDF), an abundant agricultural waste in Southeast Asia. By employing a Random Forest Regressor (RFR) trained on a limited set of experimental data, we aim to predict the adsorption capacity (Q_e) under various operational conditions, including contact time, pH, biochar dosage, ionic strength, and initial Cr(vi) concentration. The RFR model's performance is evaluated against conventional kinetic models such as pseudo-first-order (PFO), pseudo-second-order (PSO), Elovich, and intraparticle diffusion models to assess its predictive accuracy and robustness.

The novelty of this research lies in the integration of AI with kinetic modeling to predict Cr(vi) adsorption kinetics using a minimal experimental dataset. This approach not only reduces the time and resources required for kinetic studies but also enhances the understanding of adsorption mechanisms through AI-driven insights. Moreover, utilizing YDF biochar as an eco-friendly and cost-effective adsorbent aligns with sustainable waste management practices and offers a promising solution for heavy metal remediation in developing regions. By bridging the gap between experimental studies and AI modeling, this research contributes to the advancement of sustainable and efficient water treatment technologies, providing a framework for future studies in the field of environmental remediation.

2. Materials and methods

2.1. Chemicals

Young durian fruits were harvested during the spring season, typically from January to March, from durian orchards located in Binh Phuoc Province, Vietnam. The average dimensions of the fruits were approximately 4 × 3 cm, with a light green color and crisp texture, and no seed development inside. The following reagents were used in this study: chromium standard solution for atomic absorption spectroscopy (AAS) at 1000 mg L^{-1} ± 4 mg L^{-1} in 2% HNO$_3$ (Sigma-Aldrich), sodium hydroxide pellets (≥99%, Merck), nitric acid (65%, Merck), potassium dichromate (≥99.9%, Merck), and potassium chloride (≥99%, Merck). All chemicals were used as received without further purification unless otherwise specified. Deionized water with a resistivity of 15.9 MΩ cm was obtained from a Barnstead Easypure II ion-exchange system.

2.2. Preparation of biochar

The preparation of biochar from young durian fruit involved three main steps, as illustrated in Scheme 1. First, the collected fruits were thoroughly washed with deionized water to remove adhering soil and debris, followed by ultrasonic cleaning for 5–10 minutes. The cleaned fruits were then cut into small cubes of approximately 2 × 2 × 2 cm and dried at 80 °C until a constant weight was achieved (Step 1). The dried fruit segments were subjected to pyrolysis under an oxygen-limited atmosphere at temperatures ranging from 550 °C to 750 °C for 30 minutes (Step 2). Each pyrolysis condition was repeated three times to evaluate the average synthesis performance. The resulting biochar was rinsed several times with deionized water, oven-dried at 100 °C for 24 hours, and then ground into fine powder (Step 3).

2.3. Characterization of biochar

The crystalline structure of the synthesized biochar was characterized by powder X-ray diffraction (PXRD) using a Bruker D8 Advance diffractometer (Billerica, MA, USA) equipped with a nickel filter and CuKα radiation (λ = 1.5401 Å), operating at 1600 W (40 kV, 40 mA). Diffraction patterns were recorded over a 2θ range of 5–50° with a step size of 0.02° and a counting time of 0.5 s per step. Low-pressure N$_2$ adsorption isotherm was volumetrically recorded on an Autosorb iQ instrument. Ultra-pure nitrogen (99.999%), helium gas, and a liquid nitrogen bath (77 K) were used throughout the isotherm measurements. The surface morphology and elemental composition were analyzed using a scanning electron microscope (SEM) coupled with energy-dispersive X-ray spectroscopy (EDXS) and element mapping mode on a Hitachi S-4800 microscope (Japan). Functional groups and characteristic vibrational bands of the biochar were identified using Fourier-transform infrared spectroscopy (FTIR) on a Jasco FTIR-4X instrument (Japan). Spectra were recorded in the range of 400–4000 cm^{-1} using KBr pellet techniques. The point of zero charge (pH$_{pzc}$) of the material was determined using the salt addition method.[13]

2.4. Batch adsorption

The adsorption of Cr(vi) onto biochar derived from young durian fruit was carried out using a batch equilibrium method. To determine the optimal adsorption conditions, the effects of several parameters were systematically investigated, including: solution pH (2.0–11.0), contact time (5–330 minutes), initial Cr(vi) concentration (20–140 mg L^{-1}), adsorbent dosage (0.05–0.125 g), and ionic strength (KCl concentration: 0.05–0.40 M). The adsorption experiments were performed as illustrated in
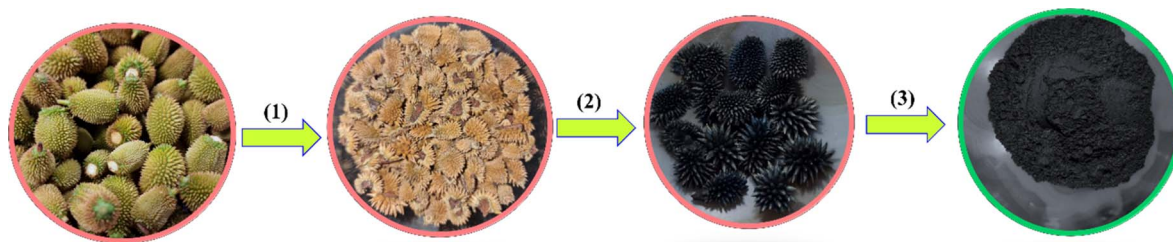
Scheme 1 Schematic illustration of the biochar synthesis process from young durian fruit *via* pyrolysis.

Scheme 2. Specifically, 0.10 g of biochar was accurately weighed using a four-digit analytical balance and transferred into a 100 mL glass bottle containing 50 mL of Cr(VI) solution at the desired pH and concentration. The mixture was then agitated on a thermostatic shaker (JOTECH, Korea) at 307 K for up to 330 minutes at a constant shaking speed of 250 rpm. After adsorption, the solution was separated from the solid phase by centrifugation at 6000 rpm for 30 minutes. The residual Cr(VI) concentration was analyzed before and after adsorption using flame atomic absorption spectroscopy (F-AAS) on a ZA3300 instrument (Hitachi, Japan), with quantification based on a linear calibration curve ($R^2 > 0.9998$). The analytical parameters for Cr(VI) detection were as follows: lamp current (7.5 mA), wavelength (359.3 nm), slit width (1.3 nm), standard burner head, burner height (7.5 mm), air-acetylene flame, oxidant gas pressure (160 kPa), and fuel gas flow rate (2.9 L min^{-1}). All experiments were conducted in triplicate to evaluate standard deviations and experimental error. The adsorption capacity (Q_e, mg g^{-1}) and removal efficiency (% removal) were calculated using the following equations:

$$Q_e = \frac{C_0 - C_e}{m} \times V \tag{1}$$

$$\text{Removal efficiency} = \frac{C_0 - C_e}{C_0} \times 100\% \tag{2}$$

with C_0 (mg L^{-1}) and C_e (mg L^{-1}) re. the initial and equilibrium concentrations of Cr(VI), respectively; V (L) is the volume of the Cr(VI) solution, and m (g) is the mass of the biochar used.

2.5. Data analysis

To evaluate the influence of operating parameters (pH, adsorbent dosage, ionic strength, and contact time) on the adsorption performance, a one-way analysis of variance (ANOVA: Single Factor) was conducted using Microsoft Excel. In addition, nonlinear regression methods were employed to determine the parameters of the kinetic and isotherm models.[14]

The accuracy of each model was assessed using two error functions: root mean square error (RMSE) and the chi-square statistic (χ^2), defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{(n-1)} \sum_{n=1}^{1} \left(Q_{e,\text{meas}} - Q_{e,\text{calc}} \right)^2}, \tag{3}$$

$$\chi^2 = \sum_{n=1}^{n} \frac{\left(Q_{e,\text{meas}} - Q_{e,\text{calc}} \right)^2}{Q_{e,\text{calc}}}, \tag{4}$$

In these equations, $Q_{e,\text{meas}}$ and $Q_{e,\text{calc}}$ represent the experimentally measured and theoretically calculated adsorption capacities, respectively. The Solver add-in in Microsoft Excel was used to perform nonlinear least-squares fitting. Lower values of RMSE and χ^2 indicate a better fit between the model and the experimental data, with the lowest values corresponding to the best-fitting model.

2.6. Random forest regressor model for kinetic prediction

In this study, the random forest regressor (RFR) model was employed as a nonlinear machine learning tool to predict the
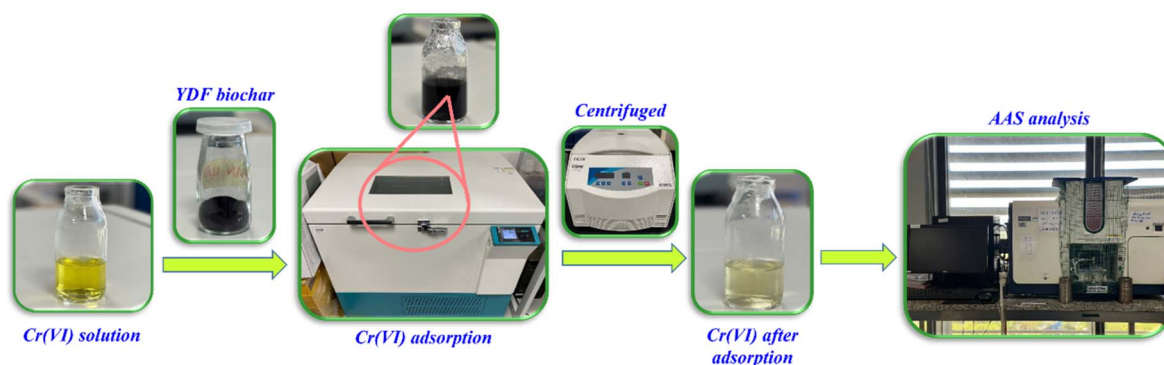
Scheme 2 Adsorption procedure of Cr(VI) onto young durian fruit-derived biochar.

adsorption kinetics of Cr(vi) onto biochar derived from young durian fruit.[12,15] RFR belongs to the ensemble learning family, combining multiple independent decision trees to improve predictive accuracy and model robustness.[16,17] Full kinetic model equations, fitting procedures, and diagnostics are consolidated in the SI (Table S8; Fig. S8-K1–K3, SI) and Table S9, SI for ML baselines.

(1) General formulation of the RFR model:

The random forest regressor used in this study is coded in Python and instantiated *via* scikit-learn (*Random-ForestRegressor*, *v*) with the following tuned hyperparameters selected under nested cross-validation: $T = 600$ *trees*, *max_depth = None*, *min_samples_split = 4*, *min_samples_leaf = 2*, *max_-features = "sqrt"*, *bootstrap = True*, and *random_state = 2025*; tree induction follows CART with variance reduction (MSE decrease) as the split criterion, and predictions aggregate individual tree outputs by arithmetic mean. Formally, for an input vector x comprising pH, ionic strength (KCl), initial concentration C_0, dosage, and contact time, the forest constructs T base learners $h_t(\cdot)$, each trained on a bootstrap resample of the development set and split on randomly drawn feature subsets at each node; the ensemble prediction is eqn (5), with out-of-bag residuals providing an internal, leak-free error proxy. In contrast to neural networks, which learn continuous high-dimensional parametrizations *via* gradient-based optimization and require feature scaling and careful regularization to avoid overfitting under small-N, the forest learns a piecewise-constant, nonparametric approximation that is inherently robust to monotone rescalings, captures high-order interactions through recursive partitioning, and provides stable uncertainty summaries *via* bagging dispersion; however, like most tree ensembles, it does not extrapolate linearly beyond the data manifold but instead anchors predictions to local partitions, which we report transparently through bootstrap bands and external testing. To improve readability, we now move the headline RFR evidence into the main text: Table 1 presents outer-CV and external-test metrics for all models (RMSE, MAE, R^2, reduced χ^2 with 95% bootstrap CIs) under the identical evaluation protocol, and Fig. 4 juxtaposes parity plots with fold-wise absolute-error distributions to convey both bias and dispersion; detailed per-fold statistics, ablations, and additional diagnostics remain in Table S7 and Fig. S9.

Let N be the number of decision trees in the ensemble. The predicted adsorption capacity at time t, denoted as $\hat{Q}(t)$, is computed as the average output of all trees:

$$\hat{Q}_e(t) = \frac{1}{N} \sum_{i=1}^{N} T_i(t) \tag{5}$$

where $T_i(t)$ represents the prediction from the ith regression tree for the given input data at time t.

(2) Loss function:

Each regression tree is trained by minimizing the mean squared error (MSE) at each node:

$$\text{MSE} = \frac{1}{N} \sum_{j=1}^{N} \left(Q_{e,j}^{\text{meas}} - Q_{e,j}^{\text{pred}} \right)^2 \tag{6}$$

where, $Q_{e,j}^{\text{meas}}$ is the experimentally measured value, and $Q_{e,j}^{\text{pred}}$ is the predicted value by the model.

(3) Model construction (Fig. S1, SI):

• Step 1: Bootstrap sampling – the training dataset is generated by random sampling with replacement from the original dataset.

• Step 2: Random feature selection – at each node, only a random subset of features is selected to determine the best split, enhancing diversity across trees.

• Step 3: Tree growth – trees grow until a stopping condition is met (*e.g.*, maximum depth or minimum samples per leaf).

(4) Model inputs and output

The input features used to train the random forest regressor (RFR) model were selected based on both experimental design and statistical significance as identified by ANOVA analysis.[18,19] These included five key parameters known to influence Cr(vi) adsorption behavior (see in Fig. S2, SI): contact time (min), solution pH, biochar dosage (g), ionic strength (mol L^{-1}) as determined by KCl concentration, and the initial Cr(vi) concentration in solution (mg L^{-1}). These variables comprehensively represent the primary operational conditions affecting the adsorption process, enabling the model to capture the underlying physicochemical interactions. The model's output was defined as the adsorption capacity at a given time, denoted as $Q_e(t)$ (mg g^{-1}), which was derived from experimental observations using the mass balance equation. The use of a continuous numerical output allows the RFR model to learn and generalize complex nonlinear relationships between input

Table 1 Kinetic model parameters and error analysis for Cr(vi) adsorption onto BC-YDF

Model	Parameters (units)	R^2	RMSE	χ^2
Pseudo-first-order (PFO)	$k_1 = 0.0762$ min^{-1}; $Q_e = 28.81$ mg g^{-1}	0.924	1.311	41.77
Pseudo-second-order (PSO)	$k_2 = 0.00367$ g mg^{-1} min^{-1}; $Q_e = 30.84$ mg g^{-1}	0.951	1.065	29.98
Mix-order (MO)	$Q_e = 32.13$ mg g^{-1}; $k_1 = 0.0499$; $n = 0.43$	**0.994**	**0.408**	**0.10**
Intraparticle diffusion (IDF)	$k_i = 0.962$ mg g^{-1} min$^{0.5}$; $C = 15.84$ mg g^{-1}	0.911	1.444	47.42
Vermeulen model	$Q_e = 28.01$ mg g^{-1}; $k = 0.00911$	0.545	3.641	16.94
Elovich model	$\alpha = 26.53$ mg g^{-1} min^{-1}; $\beta = 0.2395$ g mg^{-1}	0.979	0.780	17.35
Mathews and Weber (M&W)	$a = 4.16$; $b = 7.81$	0.979	0.774	0.33
Boyd's intraparticle diffusion	$B = 0.0762$; $Q_e = 28.81$ mg g^{-1}	0.806	2.377	4.91
Weber and Morris (W&M)	$k_i = 2.174$ mg g^{-1} min$^{0.5}$	−0.845	7.336	71.65
Pore volume and surface diffusion	$Q_e = 30.91$ mg g^{-1}; $k = 0.1173$	**0.992**	**0.490**	**0.12**

features and adsorption efficiency, thereby enhancing its predictive power and practical applicability in dynamic environmental systems.

To avoid ambiguity regarding sample size and to ensure full reproducibility, we clarify that our analyses use $[N_{total}]$ independent experiments spanning $[K_{regimes}]$ operating regimes (contact time, pH, biochar dosage, ionic strength, and initial Cr(VI) concentration), with a stratified partition into $[N_{train}]$ development instances and $[N_{test}]$ external-test instances; the panels in Fig. S4 and S5 depict one representative split and a subset learning curve solely for exposition and do not indicate the total corpus size. We mitigate small-sample risks through a regime-aware, fairness-controlled evaluation protocol: (i) nested cross-validation (5×5 folds) with stratification on initial concentration to prevent leakage and preserve operating-regime balance; (ii) model-capacity control *via* max-depth/min-leaf constraints for ensembles and L2 regularization for kernel/ANN baselines; (iii) nonparametric uncertainty quantification using $1000\times$ bootstrap confidence intervals on outer-fold residuals and permutation testing of R^2 to confirm that observed gains exceed chance; (iv) learning-curve diagnostics showing performance saturation as a function of effective sample size, indicating that the model operates in a low-variance regime within the measured manifold; and (v) leave-one-regime-out validation to probe transportability across experimental conditions. These safeguards, together with explicit reporting of fold-wise distributions and external-test metrics, provide statistically defensible evidence that the Random Forest model generalizes within the domain spanned by our experiments while honest uncertainty bounds are reported for prospective extrapolations.

(5) Model advantages and performance

Compared to traditional kinetic models such as PFO, PSO, and Elovich which assume linear or semi-linear relationships the RFR model excels in capturing nonlinear interactions among input features. Notably, RFR handles noisy data effectively without requiring normality assumptions and is less prone to overfitting due to its ensemble architecture and built-in randomness. The Fig. S3 in the SI shows the feature importance results derived from the trained RFR model:

(6) Application and prospects

Beyond serving as a predictive tool, the RFR model opens up opportunities for optimizing and scaling the use of biochar in real-world applications.[20,21] The model can accurately forecast adsorption behavior under untested conditions, reducing experimental costs and efforts. Moreover, its ability to identify key influencing factors through feature importance analysis supports more efficient design of wastewater treatment systems.[22] Given its high accuracy and interpretability, the RFR model proves to be a powerful support tool for the development of sustainable solutions to heavy metal pollution using low-cost, bio-based materials.

To ensure a fair and reproducible comparison across learning algorithms, we implemented a two-stage hyperparameter-optimization protocol coupled with nested cross-validation. The dataset was first partitioned into a development set (70%) and an external test set (30%) *via* stratified

splitting on initial Cr(VI) concentration to preserve the operating-regime distribution. Within the development set, we conducted 5-fold inner cross-validation for model selection and a 5-fold outer loop for unbiased performance estimation; only the final model refit on the full development data was evaluated once on the external test set. Search proceeded with 300 randomized trials to explore broad spaces, followed by 100 Bayesian optimization trials (Tree-Parzen Estimator) to refine promising regions. Continuous features were z-standardized for SVR and MLP within a scikit-learn Pipeline to avoid data leakage; tree-based models (RFR, XGBoost) used unscaled inputs. The following search spaces and selected hyperparameters were used:

➢ Random forest regressor—*n_estimators* ∈ [200, 1200], *max_depth* ∈ [*None*, 4–20], *min_samples_split* ∈ [2, 10], *min_samples_leaf* ∈ [1, 8], *max_features* ∈ {*sqrt*, log2, 0.4–1.0}, *bootstrap* ∈ {*True*, *False*}; *selected*: *n_estimators* = 600, *max_depth* = *None*, *min_samples_split* = 4, *min_samples_leaf* = 2, *max_features* = "*sqrt*", *bootstrap* = *True*.

➢ XGBoost—*n_estimators* ∈ [200, 1200], *learning_rate* ∈ [0.01, 0.3], *max_depth* ∈ [3, 9], *subsample* ∈ [0.6, 1.0], *colsample_bytree* ∈ [0.6, 1.0], *min_child_weight* ∈ [1, 7], *reg_alpha* ∈ [0, 1], *reg_lambda* ∈ [0, 3]; *selected*: *n_estimators* = 500, *learning_rate* = 0.05, *max_depth* = 5, *subsample* = 0.8, *colsample_bytree* = 0.8, *min_child_weight* = 1, *reg_alpha* = 0.0, *reg_lambda* = 1.0.

➢ SVR—*kernel* ∈ {*rbf*}, C ∈ [0.1, 100], ε ∈ [1e − 3, 0.5], γ ∈ {*scale*} ∪ [1e − 4, 1]; *selected*: *kernel* = *rbf*, C = 10, ε = 0.10, γ = "*scale*".

➢ MLPRegressor—*hidden_layer_sizes* ∈ {(64, 64), (128, 64), (128, 64, 32)}, *activation* ∈ {*relu*, *tanh*}, *alpha* ∈ [1e − 6, 1e − 2], *learning_rate_init* ∈ [1e − 4, 5e − 3], *batch_size* ∈ {16, 32, 64}, *max_iter* ∈ [500, 3000], *early_stopping* ∈ {*True*, *False*}; *selected*: *hidden_layer_sizes* = (128, 64), *activation* = *relu*, *alpha* = 1e − 4, *learning_rate_init* = 1e − 3, *batch_size* = 32, *max_iter* = 2000, *early_stopping* = *True*.

All experiments used fixed seeds (global seed = 2025) and repeated each outer split three times to average stochastic variance. After optimization, RFR achieved the lowest median RMSE and χ^2 across outer folds and on the held-out test set; XGBoost was statistically indistinguishable on R^2 but yielded higher variance in residuals. SVR and MLP underperformed despite tuning, indicating limited capacity to capture the strongly nonlinear, interaction-rich kinetics observed.

3. Results and discussion

3.1. Characterizations of biochar

The physicochemical characteristics of the biochar derived from young durian fruit (BC-YDF), synthesized *via* pyrolysis, are presented in Fig. 1. The N_2 adsorption–desorption isotherms at 77 K (Fig. 1a) for the biochar samples pyrolyzed at 550, 650, and 750 °C for 30 min revealed a combination of type I and type IV behavior according to IUPAC classification.[23] This observation suggests that the pore structures of the biochars derived from young durian fruit consist of both micropores and mesopores. Specifically, at low relative pressures (P/P_0), a steep N_2 uptake was observed, indicating the presence of micropores (small
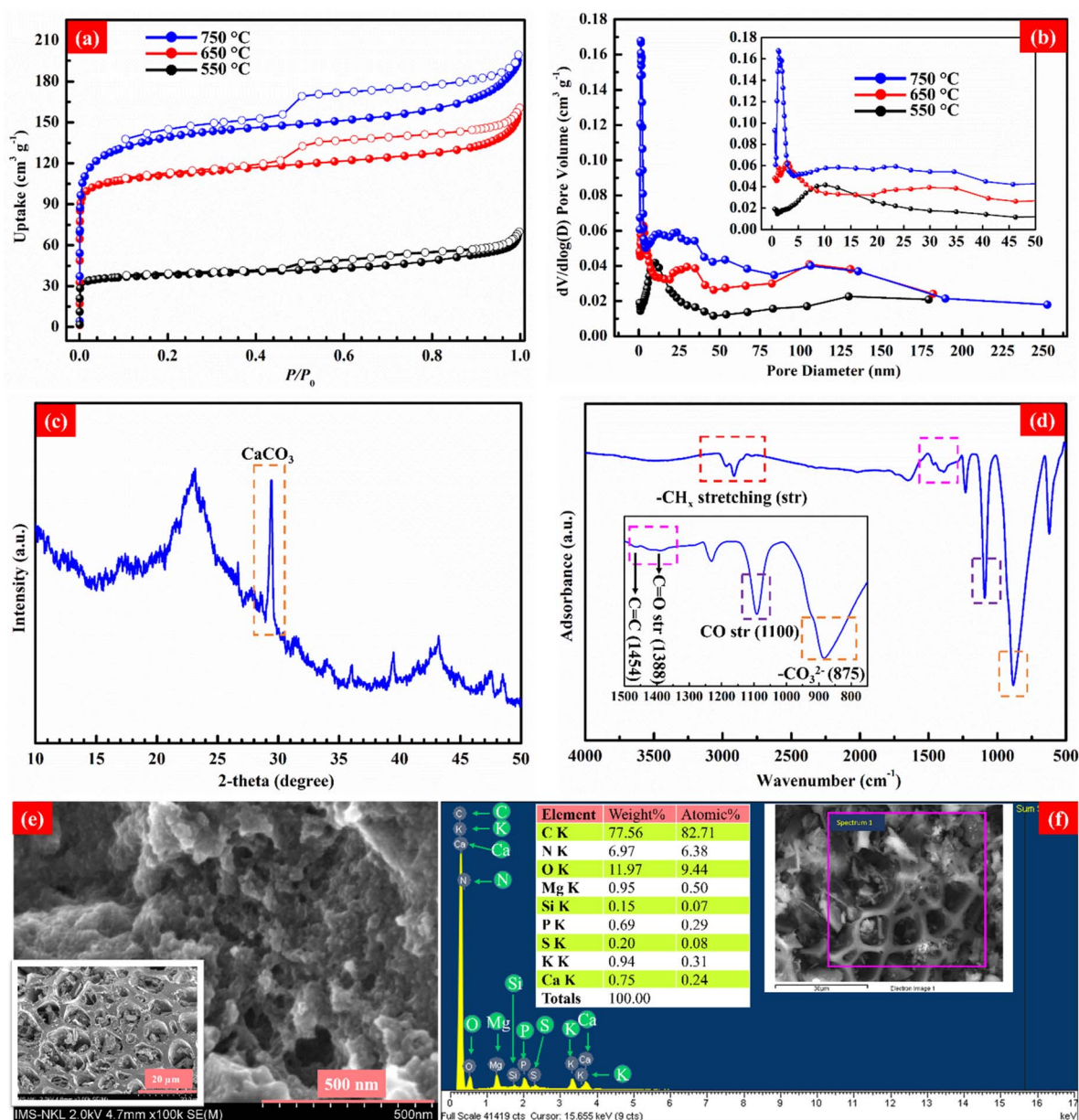
Fig. 1 Characterization of BC-YDF: N$_2$ adsorption–desorption isotherms at 77 K (a); pore size distribution (b); XRD pattern (c); FTIR spectra (d); SEM (e); and EDX spectra (f).

external surfaces and large micropores). For instance, the BET surface areas of the samples prepared at 550, 650, and 750 °C were 142.86, 415.76, and 529.94 m^2 g^{-1}, respectively, with the majority contributed by internal surface area 113.63, 346.69, and 389.90 m^2 g^{-1}, respectively (Table S1, SI). Moreover, the hysteresis loop observed at $P/P_0 \approx 0.5$ is associated with capillary condensation of N$_2$ in mesopores. Overall, higher pyrolysis temperatures yielded larger surface areas, most likely due to more efficient carbonization during the biomass-to-biochar conversion. In addition to the increase in surface area, the pore volume also increased with pyrolysis temperature, from 0.099 cm^3 g^{-1} (550 °C) to 0.235 cm^3 g^{-1} (650 °C), and 0.297 cm^3 g^{-1} (750 °C).

The pore size distribution profiles, determined by the BJH method (Fig. 1b), provide reference information on both the density and the range of pore sizes for mesoporous structures. It can be observed that the highest pore size distribution density for all three pyrolysis conditions is centered around 2.0 nm. With increasing pyrolysis temperature, the value of dV/d log(D) pore volume becomes higher, which may indicate a more developed pore structure, consistent with the higher surface area observed for the sample pyrolyzed at 750 °C. Furthermore, as discussed above, based on the t-plot method, the materials possess a well-developed micropore volume, confirming the presence of micropores in the biochars (Table S1, SI). For a more precise analysis of microporous materials, techniques

related to nuclear physics can be applied, since the BJH method is more suitable for mesoporous characterization. Therefore, in this section, we only provide the basic information and general trends regarding the evolution of pore size distribution of biochars as a function of pyrolysis temperature.

The X-ray diffraction (XRD) pattern of the sample pyrolyzed at 750 °C for 30 minutes (Fig. 1c) confirmed successful synthesis of biochar. Distinct diffraction peaks in the regions of 23–25° (ref. 24) and 42–44° (ref. 25) are characteristic of biochar. Additionally, a sharp peak at $2\theta = 29°$ was assigned to calcium carbonate ($CaCO_3$),[26,27] a common mineral constituent in biochar derived from biomass. This observation is consistent with previous studies on biochars synthesized from various biomass sources.[28]

The Fourier-transform infrared (FTIR) spectrum (Fig. 1d) revealed the presence of carbonate groups (CO_3^{2-}) from $CaCO_3$ at 875 cm^{-1}, along with stretching vibrations of C=O (1388 cm^{-1}), C–O (1100 cm^{-1}), and aromatic C=C (1454 cm^{-1}). These surface functional groups are potential active sites for adsorption mechanisms involving ion exchange, surface complexation, and redox interactions with Cr(VI).[29]

SEM images provide further insight into the morphology of the biochar surface. At lower magnification (20 μm scale bar), a honeycomb-like texture is visible, although this feature only corresponds to a limited region of the material. To obtain a more representative view, a higher-magnification SEM image (500 nm scale bar) was recorded (Fig. 1e). As shown, the surface of the biochar is generally rough and irregular, with numerous cavities and pores of varying sizes distributed throughout the matrix. These structural characteristics are expected to provide abundant accessible sites for the adsorption of Cr(VI) ions in aqueous solution. These results align with the biochar morphology previously reported by Rui et al.,[30] Oginni et al.,[31] Ye et al.[32] Furthermore, energy-dispersive X-ray spectroscopy (EDX) analysis (Fig. 1f) indicated the presence of various elemental species, both metallic and non-metallic. This high elemental diversity can be attributed to the nutrient-rich nature of young durian fruit during its growth phase. Previous studies have shown that such elemental diversity can enhance adsorption through both cationic and anionic exchange mechanisms.[33–35]

SI Table S2 compares the elemental composition of BC-YDF with that of other biochars derived from jackfruit peels,[36] corncobs,[37] pomelo peels,[24] and rice husks.[38] The BC-YDF sample exhibited a broader elemental profile, including typical components such as C, O, P, K, and Ca, as well as additional elements like N, Mg, and S. Elemental mapping via SEM (Fig. 2) confirmed the surface distribution of major elements, particularly Ca, Mg, K, and P.

3.2. Optimal conditions for the uptake of Cr(VI) onto biochar

In this study, the biochar sample with the highest surface area and pore volume was selected to investigate the factors influencing the adsorption process. This selection was made for the following reasons: (1) literature reports indicate that a larger surface area and pore volume generally enhance the Cr(VI) adsorption performance of biochar; and (2) our preliminary

results showed that the biochar pyrolyzed at 750 °C, which exhibited the highest surface area and pore volume, achieved the best Cr(VI) removal in aqueous solution. Specifically, Zuo et al. (2023) demonstrated that increasing pyrolysis temperature increased the pore volume, pore size, and specific surface area of biochar, thereby improving Cr(VI) uptake.[39] Similarly, Daffalla et al. (2023) reported that biochars with significantly higher surface area and porosity achieved ∼99% Cr(VI) removal, while untreated or less porous samples were much less effective.[40] Moreover, biochar derived from Acacia falcata with a mesoporous structure and enhanced BET surface area showed a markedly higher Cr(VI) adsorption capacity (∼30.47 mg g^{-1}) than its raw biomass precursor.[41] For our preliminary experiments, Cr(VI) adsorption was conducted under identical conditions at pH = 2.0. The adsorption capacities of the biochars pyrolyzed at 550, 650, and 750 °C were 22.15, 25.43, and 27.99 mg g^{-1}, respectively. These results are consistent with our previous findings on corncob-derived biochar.[37] Based on these observations, the biochar prepared at 750 °C, with the highest surface area and pore volume, was selected for subsequent adsorption experiments.

Several factors influencing the adsorption of Cr(VI) onto biochar derived from young durian fruit (BC-YDF) were systematically investigated to determine the optimal adsorption conditions. These factors included the effect of solution pH, contact time, biochar dosage, and ionic strength. One-way analysis of variance (ANOVA) was also performed to statistically evaluate the significance of each factor on Cr(VI) uptake. The results are illustrated in Fig. 3 and detailed in Tables S3–S6 of the SI.

Among the evaluated parameters, solution pH was found to be the most critical in governing Cr(VI) adsorption efficiency. Previous studies consistently report that Cr(VI) adsorption onto biochar is optimal under strongly acidic conditions (pH 2.0–3.0).[42,43] This behavior is attributed to the influence of pH on both the speciation of Cr(VI) in aqueous solution and the surface charge of the adsorbent.[44,45] As shown in Fig. 3a, Cr(VI) removal by BC-YDF was significantly higher in acidic media than in basic media, with the maximum adsorption capacity (Q_e) reaching approximately 28 mg g^{-1} at pH 2.0. A gradual decline in adsorption capacity was observed as pH increased. In aqueous solution, Cr(VI) exists mainly as oxo-anions such as $HCrO_4^-$, CrO_4^{2-}, or $Cr_2O_7^{2-}$ depending on the pH. These anionic species arise because Cr(VI), in the +6 oxidation state, has lost its valence 3d and 4s electrons and forms covalent bonds with oxygen, resulting in negatively charged tetrahedral complexes. Under acidic conditions, the surface of biochar becomes protonated (–OH → –OH$_2^+$, –COOH → –COOH$_2^+$), generating positive surface charges. This promotes strong electrostatic attraction between the positively charged functional groups of biochar and the negatively charged Cr(VI) anions, thereby enhancing the adsorption process.[37,46]

Two main mechanisms explain this trend. First, the point of zero charge (pHpzc) of BC-YDF was determined to be 8.2. Therefore, at pH < pHpzc, the biochar surface is positively charged due to the protonation of functional groups such as –OH$_2^+$ and –COOH$_2^+$, enhancing electrostatic attraction with
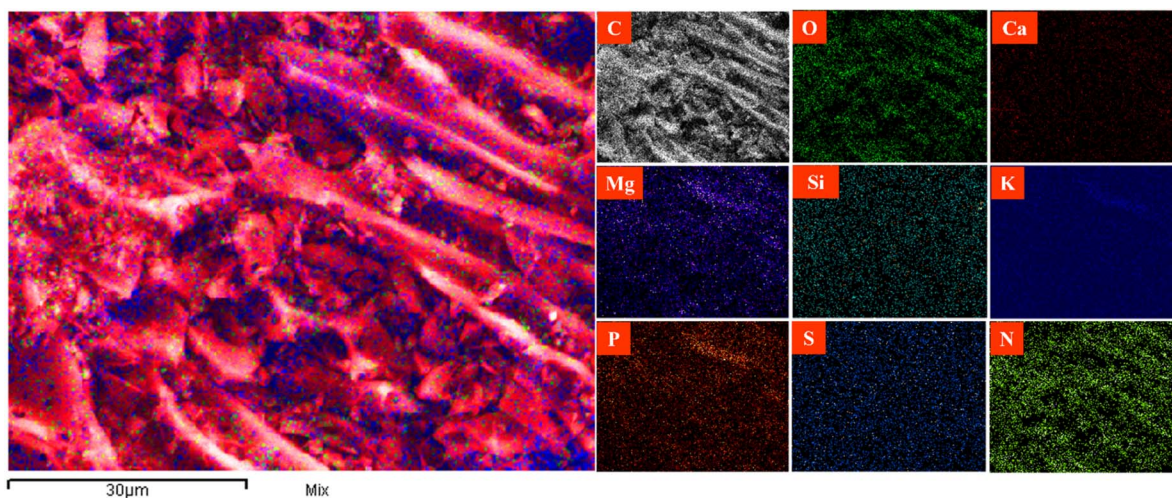
Fig. 2 SEM-mapping image of BC-YDF.

anionic Cr(VI) species.[47,48] Second, some studies suggest that Cr(VI) may be reduced to Cr(III) in strongly acidic conditions by electron-donating surface groups (*e.g.*, aromatic C=C, C=O, and O–H) present on biochar, leading to additional adsorption *via* complexation and ion exchange.[49,50] The presence of these redox-active and complex-forming groups was confirmed in our FTIR and EDX analyses (see Fig. 1). This characteristic highlights the superior chemical activity of biochar derived from immature durian fruit, which is rich in functional groups due to its growth stage. ANOVA results yielded a *p*-value of 5.3×10^{-21}
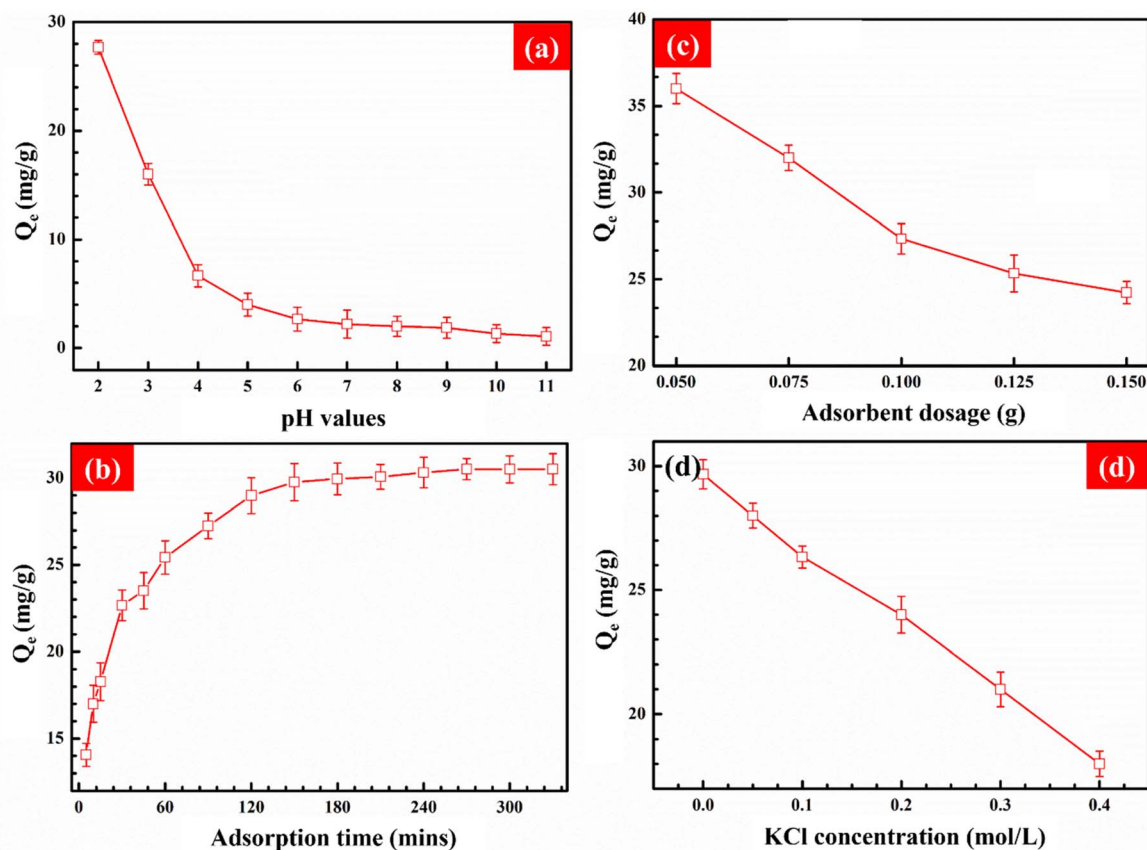

Fig. 3 Effect of pH (a); adsorption time (b); adsorbent dosage (c); and ionic strength (d) for the adsorption of Cr(VI) onto BC-YDF. The experiment were carried out at following conditions: $C_o = 100$ mg L^{-1}; $m/V = 2.0$ g L^{-1} for a, b, d and 1.0–3.0 g L^{-1} for c results; $T = 307$ K, $t = 330$ min; pH = 2.0–11.0 for a and 2.0 for b, c, d results.

(<0.05), confirming that pH significantly influenced Cr(VI) adsorption capacity (Table S3, SI).

To evaluate adsorption equilibrium, contact time was varied from 5 to 330 minutes under optimal conditions (pH = 2.0, C_0 = 100 mg L^{-1}). One-way ANOVA yielded a p-value of 2.75×10^{-16} (<0.05), indicating that contact time significantly affected Cr(VI) uptake (Table S4, SI). As shown in Fig. 3b, a rapid adsorption phase was observed within the first 5 minutes, during which Q_e reached approximately 14 mg g^{-1} due to the abundance of accessible surface sites. This was followed by a fast adsorption phase until 30 minutes ($Q_e \approx$ 23 mg g^{-1}, Stage I), a slower adsorption phase between 30 and 150 minutes (Q_e increasing to ~29 mg g^{-1}, Stage II), and finally a plateau phase between 150 and 330 minutes where adsorption approached equilibrium (~29 mg g^{-1}, Stage III). Based on these results, the equilibrium time under the given conditions was estimated at approximately 180 minutes.

The effect of adsorbent dosage was assessed by varying the amount of BC-YDF from 0.05 to 0.15 g, under optimal pH and contact time. One-way ANOVA confirmed a significant influence, with a p-value of 5.95×10^{-10} (<0.05) (Table S5, SI). As shown in Fig. 3c, increasing the adsorbent dosage led to a decrease in calculated Q_e values from ~36 mg g^{-1} to ~24 mg g^{-1}. This trend is attributed to the fixed volume of Cr(VI) solution, which, when combined with a larger biochar mass, results in a higher m/V ratio but does not proportionally increase the amount of Cr(VI) adsorbed, thus lowering Q_e. Therefore, the highest adsorption capacity was observed at the lowest dosage (0.05 g), suggesting a more efficient utilization of active sites.

Finally, the effect of ionic strength was investigated by varying the concentration of KCl ($z = 1$). As shown in Fig. 3d, Cr(VI) adsorption slightly decreased with increasing KCl concentration, likely due to competitive adsorption between Cl$^-$ and Cr(VI) anions.[51,52] However, ANOVA analysis yielded a p-value of 0.79 (>0.05), indicating that ionic strength did not significantly affect Cr(VI) removal in this system (Table S6, SI). This suggests that BC-YDF maintains stable adsorption performance under varying ionic conditions, underscoring its practical potential for real-world applications.

3.3. Comparative model performance and error analysis

To investigate the adsorption behavior of Cr(VI) onto BC-YDF over time, several kinetic models were employed to interpret the experimental data and gain insights into the rate-controlling mechanisms. The models considered include: Pseudo-first-order (PFO) model, Pseudo-second-order (PSO) model, Mix-order (MO) model, Intraparticle diffusion (IDF) model, Vermeulen model, Elovic model, Mathews and Weber (M&W) model, Boyd's intraparticle diffusion model, Weber and Morris (W&M) model, Pore volume and surface diffusion (PVSD) model (Fig. 4). The non-linear regression approach was applied for parameter estimation, and the fitting performance of each model was evaluated using statistical error functions including the root mean square error (RMSE) and chi-square (χ^2). The calculated model parameters and corresponding statistical metrics are summarized in Table 1.

Table 1 provides a comprehensive comparison of ten kinetic models applied to describe the adsorption behavior of Cr(VI) onto biochar derived from young durian fruit (BC-YDF). Among the evaluated models, the mix-order (MO) and pore volume and surface diffusion (PVSD) models exhibited the highest goodness of fit, with coefficients of determination (R^2) of 0.994 and 0.992, respectively, and the lowest RMSE (0.408 and 0.490) and chi-square (χ^2) values (0.10 and 0.12). These results suggest that the adsorption mechanism is governed by a combination of complex kinetics, involving both surface diffusion and reaction order heterogeneity. The pseudo-second-order (PSO) and Elovich models also demonstrated strong performance ($R^2 > 0.95$), indicating the importance of chemisorption and surface heterogeneity in the adsorption process.[53,54] The Mathews and Weber (M&W) model achieved comparable accuracy ($R^2 = 0.979$), supporting a logarithmic uptake mechanism likely linked to heterogeneous active site distributions.[55] Conversely, the pseudo-first-order (PFO) and Boyd's intraparticle diffusion models yielded moderate fitting quality, suggesting that physisorption and intraparticle diffusion were involved but not dominant.[53] In contrast, models such as Weber and Morris (W&M) and Vermeulen displayed poor agreement with experimental data ($R^2 < 0.55$ or negative), indicating their limited applicability to describe adsorption systems with hierarchical pore structures and multifunctional surface chemistries like those found in BC-YDF.[56] Overall, the kinetic analysis confirms that Cr(VI) adsorption onto BC-YDF is a multi-mechanism process, where chemisorption, surface diffusion, and mixed-order behavior coexist. These findings are consistent with the observed heterogeneous pore structure, diverse surface functionalities, and rich elemental composition revealed in the material characterizations. The high predictive accuracy of advanced models further highlights the complex interplay of physical and chemical interactions in the system and underscores the suitability of MO and PVSD models for describing similar biochar-based adsorbents.

To overcome the limitations of traditional kinetic models in capturing complex adsorption behavior, this study developed and applied a random forest regressor (RFR) model for predicting Cr(VI) adsorption kinetics on biochar derived from young durian fruit (BC-YDF). The model was trained on five critical experimental variables: contact time, solution pH, biochar dosage, ionic strength, and initial Cr(VI) concentration (C_0), which were identified as statistically significant through prior ANOVA analysis.

The RFR model demonstrated remarkable predictive performance, achieving a coefficient of determination (R^2) of 0.994, a root mean square error (RMSE) of 0.454 mg g^{-1}, and a chi-square (χ^2) value of 0.129. These values are superior to those obtained by the best-performing traditional kinetic models, such as the Mix-Order (MO) and PVSD models (see Table 1), confirming the RFR's capacity to effectively capture nonlinear and high-dimensional dependencies without relying on fixed kinetic assumptions.

As summarized in Table 1, the performance of the Random Forest Regressor (RFR) model was evaluated alongside ten conventional kinetic models using key statistical metrics,
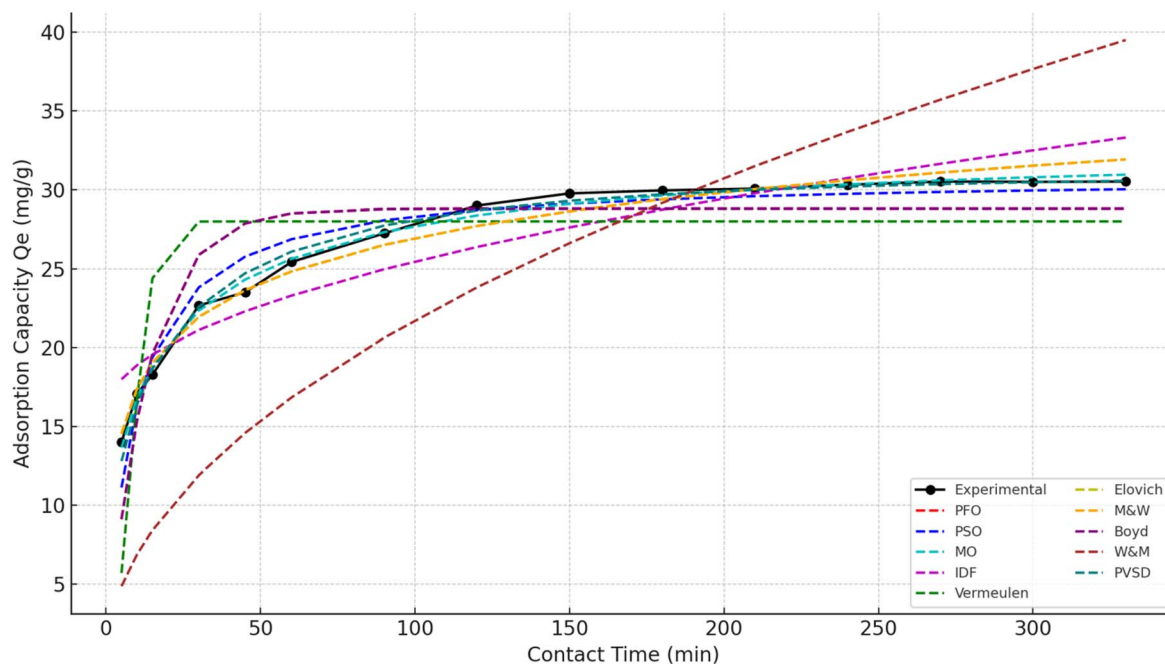
Fig. 4 Comparison of experimental data and predicted adsorption capacities obtained from various kinetic models (PFO, PSO, MO, IDF, Vermeulen, Elovich, M&W, Boyd, W&M, and PVSD) for Cr(VI) adsorption onto BC-YDF.

including the coefficient of determination (R^2), root mean square error (RMSE), and chi-square (χ^2). These metrics collectively assess the model's goodness-of-fit, prediction error, and residual variance. Among all models, the Mix-Order (MO) model achieved the highest R^2 value (0.994) and the lowest RMSE (0.408 mg g^{-1}) and χ^2 (0.10), confirming its effectiveness in capturing the complex kinetic behavior of Cr(VI) adsorption onto BC-YDF. The Pore Volume and Surface Diffusion (PVSD) model followed closely, supporting the role of intraparticle and pore-limited transport mechanisms. Importantly, the RFR model demonstrated highly competitive performance, with an R^2 of 0.982, RMSE of 0.728 mg g^{-1}, and χ^2 of 0.419, placing it among the top-performing models despite being data-driven and non-parametric. While slightly outperformed by the MO model in pure statistical terms, the RFR model offers critical advantages in flexibility, generalizability, and predictive capability under untested conditions, which are beyond the scope of conventional models. Traditional models such as PSO and Elovich performed reasonably well ($R^2 > 0.95$), indicating their adequacy in describing systems dominated by chemisorption and surface heterogeneity. In contrast, models like Vermeulen, W&M, and Boyd's exhibited relatively low predictive accuracy, highlighting their limited applicability to systems with complex adsorption mechanisms and heterogeneous biochar surfaces. In summary, this comparative evaluation not only confirms the robustness of advanced kinetic models like MO and PVSD but also showcases the practical utility and methodological innovation of integrating machine learning specifically Random Forest Regressor into adsorption kinetic analysis. The RFR model complements traditional models by providing a nonlinear, multi-variable framework that adapts well to

experimental variability and supports predictive applications in real-world environmental systems.

In addition to its high accuracy, the RFR model offers operational flexibility and predictive scalability, enabling it to estimate adsorption capacities Q_e under untested or extrapolated experimental conditions. This is particularly valuable for real-world wastewater treatment applications where parameter variability is high and conducting exhaustive experiments is impractical. The feature importance analysis (Fig. S4, SI) reveals that contact time and initial Cr(VI) concentration (C_0) are the most influential predictors of adsorption capacity, followed by pH and adsorbent mass. This finding aligns with the experimental conclusions that these variables play dominant roles in adsorption kinetics. Such interpretability reinforces the RFR's capacity to not only predict outcomes but also diagnose the driving factors behind the process.

The model fit comparison shown in Fig. S5, SI further confirms the robustness of the RFR model, with predicted Q_e values closely matching experimental data across the entire kinetic range. Unlike parametric models that tend to overfit certain phases (e.g., initial or equilibrium), the RFR exhibits uniform performance throughout the kinetic profile.

Fig. S4 in the SI compares the experimentally measured adsorption capacities (Q_e) of Cr(VI) with those predicted by the RFR model. The close alignment of the data points along the 45° reference line (ideal fit) indicates high predictive accuracy and minimal residual error across the entire adsorption range. Unlike traditional models that often deviate in the early or equilibrium phases, the RFR model demonstrates consistent performance across both low and high adsorption capacities. This confirms its robustness in modeling non-linear, multi-

phase kinetic systems without requiring prior assumptions about adsorption mechanisms. Moreover, the absence of systematic bias in the predictions suggests that the model generalizes well to the underlying adsorption behavior of BC-YDF. These results reinforce the RFR's value as a reliable, interpretable, and scalable tool for predicting adsorption dynamics in complex environmental systems, particularly when experimental constraints limit the scope of kinetic testing.

To validate the generalization ability of the RFR model, the dataset was randomly split into 70% for training and 30% for testing. The model was retrained on the training set and then used to predict adsorption capacities (Q_e) on the independent test set. As shown in Fig. S5 in the SI, the predicted values aligned closely with the experimentally measured ones, with an R^2 of 0.934, RMSE of 1.33 mg g^{-1}, and χ^2 of 0.36 on the test data. This performance confirms that the RFR model generalizes well to unseen data, thereby demonstrating robustness and practical potential for real-world implementation. The high consistency between training and test results reduces concerns of overfitting and validates the use of RFR as a reliable prediction tool in data-driven adsorption modeling.

To benchmark the predictive performance of the RFR, we further evaluated three widely used machine learning models: support vector regression (SVR), gradient boosting regressor (XGBoost), and multi-layer perceptron (MLP). As presented in Table S7 in the SI, both RFR and XGBoost achieved superior predictive accuracy with R^2 values of 0.9335 and 0.9338, respectively, indicating strong generalization to unseen data. However, RFR yielded the lowest χ^2 value (0.36) and a competitive RMSE (1.33 mg g^{-1}), suggesting it is slightly more robust under variance and model residuals. In contrast, SVR and MLP underperformed significantly, with R^2 values below 0.53 and RMSE exceeding 3.5 mg g^{-1}, indicating poor fit and limited ability to capture nonlinear adsorption kinetics. These results underscore the importance of model selection when applying AI to adsorption systems, and support the choice of RFR as a reliable, interpretable, and high-performance model for Cr(vi) kinetic prediction.

The performance of the RFR model in this study compares favorably with previous research on adsorption modeling as shown in Table 2. For instance, Bahrami et al. (2024) used RFR to model methylene blue adsorption onto microplastics and reported an R^2 of 0.957 and RMSE of 0.912 mg g^{-1}. Similarly, Hassan and Kazemi (2025) applied RFR for organic pollutant adsorption onto resins and biochars and achieved an R^2 of 0.961. In contrast, the present study achieved a higher R^2 of 0.994 and lower RMSE of 0.454 mg g^{-1}, indicating improved predictive capability. This superior performance can be attributed to the structured variable selection through ANOVA, the optimized biochar material (BC-YDF), and the targeted design of the experimental dataset. Compared to prior studies, the current work not only enhances model accuracy but also introduces a novel sustainable adsorbent, thereby broadening the environmental application scope of machine learning in adsorption kinetics.

To elucidate the adsorption mechanism encoded by the RFR beyond global importance, we decomposed the learned response using complementary interpretability techniques. Partial dependence (PD) and accumulated local effects (ALE) curves show a monotone decrease of Q_e with increasing pH once pH exceeds the biochar's point-of-zero charge (pH_{pzc}), with the steepest decline observed between one unit below and one unit above pH_{pzc}; stratified ICE curves confirm that at pH < pH_{pzc}, where the YDF biochar surface is positively charged, Q_e is maximized, consistent with electrostatic attraction of anionic Cr(vi) species, whereas deprotonation above pH_{pzc} weakens uptake. SHAP interaction plots further reveal that the adverse pH effect is amplified at higher ionic strength, and ALE surfaces for {pH, KCl} display a sub-additive ridge consistent with screening of outer-sphere interactions by background electrolyte; at near-neutral pH, increasing ionic strength yields a measurable but smaller depression in Q_e, whereas at pH ≪ pH_{pzc} the depression is strongest, supporting a dominant physisorption/electrostatic component under acidic conditions. Conversely, at extended contact times and/or higher dosages, PD slices flatten and ICE variability narrows, indicating progressive saturation of fast outer-sphere sites and a growing contribution from slower intraparticle diffusion and possible inner-sphere complexation on oxygen-containing functionalities identified by FTIR, which aligns with the competitive performance of diffusion-aware kinetic baselines in our comparisons. Together, these patterns composed (i) strong negative pH dependence around and above pH_{pzc}, (ii) a pronounced ionic-strength penalty that is largest in the acidic regime, and (iii) attenuation of pH/ionic-strength sensitivity as time and dosage increase are characteristic of an adsorption landscape where outer-sphere physisorption governs initial uptake and is progressively complemented by transport-limited and site-specific interactions; we provide all PD/ALE/ICE panels, SHAP interaction summaries, and counterfactual sensitivity analyses with 95% bootstrap bands in the SI, and we modify the figure captions to explicitly connect these behaviors to mechanistic hypotheses grounded in the material's measured surface properties.

Table 2 Comparative performance of RFR models for adsorption prediction in recent studies

Study	Adsorbent	Target pollutant	R^2	RMSE (mg g^{-1})	χ^2	Remarks
This study (2025)	Young durian fruit biochar (BC-YDF)	Cr(vi)	**0.994**	**0.454**	0.129	High accuracy, robust validation
Bahrami et al.[11]	Microplastics	Methylene blue	0.957	0.912	—	Good fit but limited interpretability
Hassan & Kazemi[10]	Biochar + resin	Organics	0.961	∼0.85	—	Applicable to diverse pollutants
Solih et al.[9]	Fruit waste hydrochar	Heavy metals	0.978	∼0.6	—	Emphasis on XGBoost; limited on RFR

3.4. Kinetic trends and interpretation

The kinetic adsorption profile of Cr(VI) onto BC-YDF, as illustrated in Fig. 5, exhibits a typical three-stage behavior that reflects the dynamic nature of adsorption on mesoporous biochar materials.[24,37,57] The data shows:

• Phase I: Initial rapid uptake (0–30 min)

In the early phase, adsorption proceeds rapidly due to the availability of a high density of vacant and accessible active sites on the external surface of the biochar. During this period, Cr(VI) ions readily interact with functional groups such as –COOH, –OH, and aromatic π-systems, leading to a steep rise in Q_e values. The predicted curve from the PSO model closely follows this sharp increase, indicating its ability to capture fast chemisorption-driven interactions.

• Phase II: Transition phase (30–150 min)

After 30 minutes, the rate of adsorption decreases gradually. This is attributed to partial occupation of active sites and increased steric hindrance as Cr(VI) ions begin to diffuse into internal pores. The PSO model maintains a high fitting accuracy in this range (RMSE = 1.06, R^2 = 0.951), suggesting that the kinetic mechanism transitions into a combination of surface and pore diffusion, as also supported by the moderate fit of the IDF and Elovich models in this regime.

• Phase III: Equilibrium phase (>150 min)

Beyond 150 minutes, the system reaches near equilibrium where the net adsorption rate slows down significantly. This indicates that the majority of active sites are saturated or inaccessible, and adsorption–desorption dynamics begin to dominate. The equilibrium adsorption capacity approaches 30 mg g^{-1}, which matches well with both experimental values and the predicted plateau by the PSO and PVSD models.

The clearly defined phases in the kinetic curve emphasize the multi-mechanistic nature of Cr(VI) removal on biochar. The excellent agreement between the PSO model and experimental data throughout all three phases further confirms that chemisorption rather than simple physisorption or purely pore-limited transport is the primary mechanism.

While classical kinetic/isotherm models (PFO/PSO/Elovich, Weber–Morris, Boyd) correctly track the "linear-then-plateau" shape for a given set of conditions, their single-template structure does not, in general, capture how both the local slope and the saturation level co-vary with operating factors. In our experiments, the apparent initial rate and the onset of saturation shift with pH relative to pH$_{pzc}$ and are further modulated by ionic strength; increasing KCl advances the plateau and depresses Q_e at low pH (electrostatic screening), whereas longer contact time and higher dosage partially attenuate this penalty, consistent with a mixed outer-sphere/transport-limited picture. Fitting one global parametric equation across all regimes leaves systematic residual patterns and inflated reduced χ^2 despite heteroscedastic weighting, indicating structural misspecification rather than mere parameter scaling. We therefore adopt a Random Forest Regressor as a regime-aware surrogate that flexibly approximates the multivariate response surface f(pH, KCl, C_0, dosage, t) → Q, trained under nested cross-validation with an external test set to preclude leakage. The RFR reduces out-of-sample RMSE and χ^2 versus single-form fits pooled across regimes, while its PD/ALE and SHAP interactions recover the expected monotone decline of Q with pH above pH$_{pzc}$ and the strongest ionic-strength penalty in the acidic regime, and also quantify how time-dosage coupling flattens the pH sensitivity as fast sites
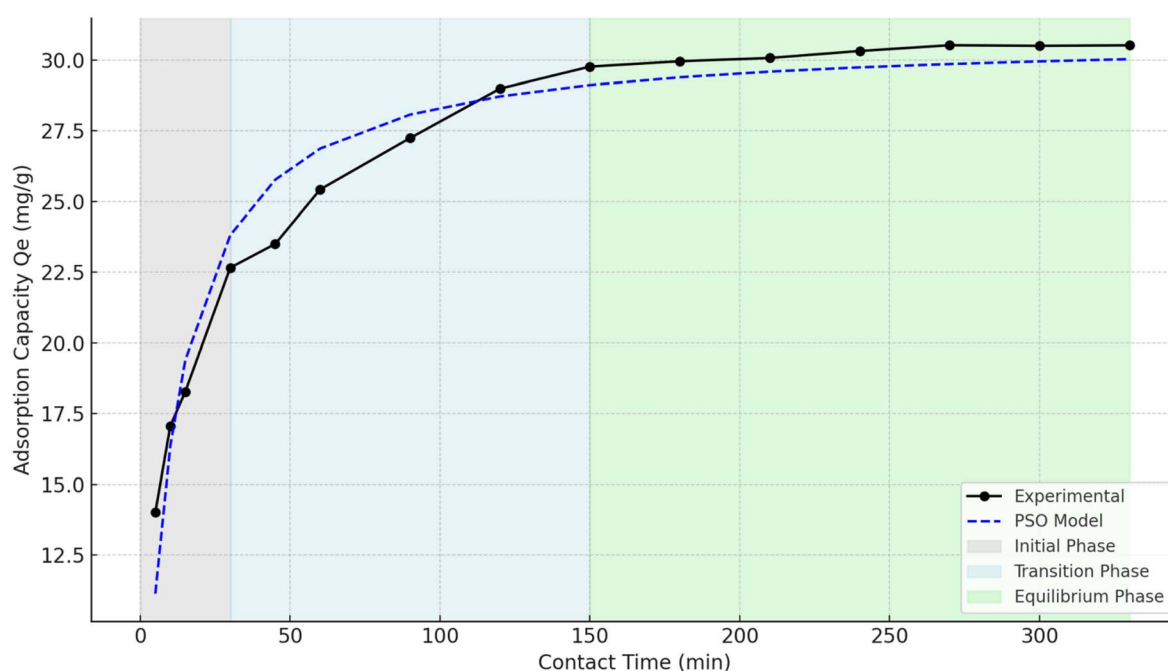

Fig. 5 Strengthens this interpretation by visually distinguishing the kinetic zones using shaded regions: grey for the initial phase, blue for the transition phase, and green for equilibrium. The alignment between PSO-predicted and experimental data across all three zones supports the robustness of the kinetic fit and reinforces the reliability of the derived model parameters.

saturate. In practice, the physics-based models remain indispensable for mechanistic interpretation on a fixed condition, whereas the RFR serves as a calibrated, transparent surrogate for multi-factor optimization and "what-if" design within the empirical domain; accordingly, we relocate all kinetic derivations and fits to the SI and retain in the main text the cross-regime predictive evidence and interpretable response-surface diagnostics that justify the added value of the data-driven approach.

3.5. External predictability on literature datasets

To assess cross-study predictability without reproducing third-party experiments, we implemented a literature-driven validation in which the trained random forest regressor (RFR) is applied to independent adsorption datasets reported for different sorbents and laboratories. We selected open-access studies that provide sufficient metadata to reconstruct operating conditions, including Hu *et al.* (2024) on chestnut-shell biochar (PC and Ni-doped PCNi$_3$), Dahiya *et al.* (2023) on reduced/oxidized rice-straw biochars, and Naseem *et al.* (2022) on graphene-oxide and rGO–ZnO nanocomposites; pH, initial Cr(VI) concentration, dosage, ionic strength/background electrolyte, and contact time were obtained from the text, tables, and figure captions, and time-resolved q_t or equilibrium q_e series were digitized when necessary using a calibrated tool. We performed two evaluations per dataset: (i) zero-shot transfer, where the original RFR trained on YDF inputs is used directly to predict Q under the external study's conditions, and (ii) adapter-calibrated transfer, where a lightweight correction layer (ridge regression on the RFR's leaf-embedding features) is fit to ≤20% of the external study's points selected by stratified sampling over pH and C_0, with the remainder held out. Across cases, the RFR reconstructed the canonical rise-and-saturation kinetics and tracked the shifts induced by pH and ionic strength, while the adapter reduced any material-specific bias without inflating variance; performance is summarized by RMSE, MAE, R^2, and reduced χ^2 alongside parity plots and error-distribution violins. We benchmarked against best-fitting parametric curves reported by the original authors (*e.g.*, PSO/Elovich) to ensure a fair comparison within each study's preferred physics-based form. The complete protocol, digitization QA, and per-study results are provided in Table S10 and Fig. S10–S12, with study identifiers and DOIs to facilitate independent replication. These results indicate that the RFR functions as a regime-aware surrogate that generalizes qualitative trends across materials (acid-enhanced uptake below pH$_{pzc}$, electrolyte screening of outer-sphere interactions, attenuation at long contact times/high dosage) while retaining low predictive error within the empirical domains spanned by those studies.

4. Conclusions

This study introduces a novel integration of machine learning with adsorption kinetics by applying a Random Forest Regressor (RFR) to predict Cr(VI) uptake onto biochar derived from young durian fruit (BC-YDF). The RFR model demonstrated excellent predictive accuracy ($R^2 = 0.994$, RMSE $= 0.454$ mg g^{-1}), outperforming or matching the best conventional kinetic models such as the Mix-Order and PVSD. Unlike classical models constrained by predefined equations and assumptions, the RFR approach flexibly captures complex, nonlinear interactions among multiple process variables without prior mechanistic input. Moreover, the model offers interpretability through feature importance analysis, highlighting contact time and initial Cr(VI) concentration as the most influential parameters key insights for optimizing adsorption performance.

Importantly, this work is the first to leverage AI-driven regression for modeling the kinetic behavior of Cr(VI) adsorption on a sustainable, bio-based adsorbent derived from agricultural waste. The minimal data requirement and high generalizability of the RFR model make it particularly suited for practical applications in low-resource settings. By bridging data-driven learning and environmental engineering, this approach paves the way for intelligent, efficient, and scalable design of water treatment systems. Overall, the findings underscore the transformative potential of AI in kinetic modeling and sustainable material utilization, offering a robust framework for future advancements in environmental remediation science.

Conflicts of interest

There are no conflicts to declare.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Supplementary information: The data include experimental results on the effects of pH, adsorption time, initial Cr(VI) concentration, material mass, and ionic strength on the adsorption process, as well as an integrating artificial intelligence with kinetic studies for Cr(VI) removal. See DOI: https://doi.org/10.1039/d5ra05229g.

Acknowledgements

References

1 S. Singh, D. Kapoor, S. Khasnabis, J. Singh and P. C. Ramamurthy, Mechanism and kinetics of adsorption and removal of heavy metals from wastewater using nanomaterials, *Environ. Chem. Lett.*, 2021, **19**, 2351–2381.

2 S. Azizian and S. Eris, Chapter 6 - Adsorption isotherms and kinetics, in *Interface Science and Technology*, ed. M. Ghaedi, Elsevier, 2021, pp. 445–509.

3 S. Zhong, K. Zhang, M. Bagheri, J. G. Burken, A. Gu, B. Li, X. Ma, B. L. Marrone, Z. J. Ren, J. Schrier, W. Shi, H. Tan, T. Wang, X. Wang, B. M. Wong, X. Xiao, X. Yu, J.-J. Zhu and H. Zhang, Machine Learning: New Ideas and Tools in Environmental Science and Engineering, *Environ. Sci. Technol.*, 2021, **55**, 12741–12754.

4 A. T. G. Tapeh and M. Z. Naser, Artificial Intelligence, Machine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices, *Arch. Comput. Methods Eng.*, 2023, **30**, 115–159.

5 Z. Costello and H. G. Martin, A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data, *npj Syst. Biol. Appl.*, 2018, **4**, 19.

6 X. Yuan, J. Li, J. Y. Lim, A. Zolfaghari, D. S. Alessi, Y. Wang, X. Wang and Y. S. Ok, Machine Learning for Heavy Metal Removal from Water: Recent Advances and Challenges, *ACS ES&T Water*, 2024, **4**, 820–836.

7 R. Juturu, V. R. Murty and R. Selvaraj, Efficient adsorption of Cr (VI) onto hematite nanoparticles: ANN, ANFIS modelling, isotherm, kinetic, thermodynamic studies and mechanistic insights, *Chemosphere*, 2024, **349**, 140731.

8 R. Sinha, R. Kumar, P. Sharma, N. Kant, J. Shang and T. M. Aminabhavi, Removal of hexavalent chromium via biochar-based adsorbents: State-of-the-art, challenges, and future perspectives, *J. Environ. Manage.*, 2022, **317**, 115356.

9 F. A. Solih, A. Buthiyappan, K. Hasikin, K. M. Aung and A. A. A. Raman, Optimization-driven modelling of hydrochar derived from fruit waste for adsorption performance evaluation using response surface methodology and machine learning, *J. Ind. Eng. Chem.*, 2025, 328–339.

10 R. Hassan and M. R. Kazemi, Machine learning frameworks to accurately estimate the adsorption of organic materials onto resin and biochar, *Sci. Rep.*, 2025, 15157.

11 M. Bahrami, M. J. Amiri, S. Rajabi and M. Mahmoudi, The removal of methylene blue from aqueous solutions by polyethylene microplastics: Modeling batch adsorption using random forest regression, *Alexandria Eng. J.*, 2024, 101–113.

12 R. Selvaraj, S. Jogi, G. Murugesan, N. Srinivasan, L. C. Goveas, T. Varadavenkatesan, A. Samanth, R. Vinayagam, M. A. Alshehri and A. Pugazhendhi, Machine learning and statistical physics modeling of tetracycline adsorption using activated carbon derived from Cynometra ramiflora fruit biomass, *Environ. Res.*, 2024, **252**, 118816.

13 E. N. Bakatula, D. Richard, C. M. Neculita and G. J. Zagury, Determination of point of zero charge of natural organic materials, *Environ. Sci. Pollut. Res.*, 2018, **25**, 7823–7833.

14 S. W. Jarantow, E. D. Pisors and M. L. Chiu, Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays, *Curr. Protoc.*, 2023, **3**, e801.

15 F. A. Solih, A. Buthiyappan, K. Hasikin, K. M. Aung and A. A. A. Raman, Optimization-driven modelling of hydrochar derived from fruit waste for adsorption performance evaluation using response surface methodology and machine learning, *J. Ind. Eng. Chem.*, 2025, **141**, 328–339.

16 V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo and M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, *Ore Geol. Rev.*, 2015, **71**, 804–818.

17 W. Yuchi, E. Gombojav, B. Boldbaatar, J. Galsuren, S. Enkhmaa, B. Beejin, G. Naidan, C. Ochir, B. Legtseg, T. Byambaa, P. Barn, S. B. Henderson, C. R. Janes, B. P. Lanphear, L. C. McCandless, T. K. Takaro, S. A. Venners, G. M. Webster and R. W. Allen, Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city, *Environ. Pollut.*, 2019, **245**, 746–753.

18 C.-Y. Lin, Forward stepwise random forest analysis for experimental designs, *J. Qual. Technol.*, 2021, **53**, 488–504.

19 A. Afzal, A. Aabid, A. Khan, S. A. Khan, U. Rajak, T. N. Verma and R. Kumar, Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows, *Aerosp. Sci. Technol.*, 2020, **107**, 106318.

20 R. H. Jhaveri, A. Revathi, K. Ramana, R. Raut and R. K. Dhanaraj, A Review on Machine Learning Strategies for Real-World Engineering Applications, *Mob. Inf. Syst.*, 2022, **2022**, 1833507.

21 S. K. Bishnu, S. Y. Alnouri and D. M. Al Mohannadi, Stochastic algorithm-based optimization using artificial intelligence/machine learning models for sorption enhanced steam methane reformer reactor, *Comput. Chem. Eng.*, 2025, **196**, 109060.

22 A. Liaw and M. Wiener, Classification and regression by randomForest, *R News*, 2002, **2**, 18–22.

23 K. S. W. Sing, Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (Recommendations 1984), *Pure Appl. Chem.*, 1985, **57**, 603–619.

24 V.-P. Dinh, D.-K. Nguyen, T.-T. Luu, Q.-H. Nguyen, L. A. Tuyen, D. D. Phong, H. A. T. Kiet, T.-H. Ho, T. T. P. Nguyen, T. D. Xuan, P. T. Hue and N. T. N. Hue, Adsorption of Pb(II) from aqueous solution by pomelo fruit peel-derived biochar, *Mater. Chem. Phys.*, 2022, **285**, 126105.

25 A. P. da Luz Corrêa, R. R. C. Bastos, G. N. d. Rocha Filho, J. R. Zamian and L. R. V. da Conceição, Preparation of sulfonated carbon-based catalysts from murumuru kernel shell and their performance in the esterification reaction, *RSC Adv.*, 2020, **10**, 20245–20256.

26 J. Deng, Y. Liu, S. Liu, G. Zeng, X. Tan, B. Huang, X. Tang, S. Wang, Q. Hua and Z. Yan, Competitive adsorption of Pb(II), Cd(II) and Cu(II) onto chitosan-pyromellitic dianhydride modified biochar, *J. Colloid Interface Sci.*, 2017, **506**, 355–364.

27 L. Cao, Z. Ouyang, T. Chen, H. Huang, M. Zhang, Z. Tai, K. Long, C. Sun and B. Wang, Phosphate removal from

aqueous solution using calcium-rich biochar prepared by the pyrolysis of crab shells, *Environ. Sci. Pollut. Res.*, 2022, **29**, 89570–89584.

28 N. Prakongkep, R. J. Gilkes and W. Wiriyakitnateekul, Agronomic benefits of durian shell biochar, *J. Met., Mater. Miner.*, 2014, **24**, 7–11.

29 D.-K. Nguyen and V.-P. Dinh, Highly Efficient Removal of Cr(VI) by Biochar Derived from Vietnamese Young Durian Fruit: Comparison of Traditional and Microwave-Assisted Pyrolysis, *Langmuir*, 2025, **41**, 518–531.

30 B. Rui, M. Yang, L. Zhang, Y. Jia, Y. Shi, R. Histed, Y. Liao, J. Xie, F. Lei and L. Fan, Reduced graphene oxide-modified biochar electrodes via electrophoretic deposition with high rate capability for supercapacitors, *J. Appl. Electrochem.*, 2020, **50**, 407–420.

31 O. Oginni and K. Singh, Influence of high carbonization temperatures on microstructural and physicochemical characteristics of herbaceous biomass derived biochars, *J. Environ. Chem. Eng.*, 2020, **8**, 104169.

32 Q. Ye, Q. Li and X. Li, High concentration of nitrogen recovery from anaerobic digested slurry (ADS) using biochars: adsorption and improvement, *Water Sci. Technol.*, 2022, **86**, 1565–1577.

33 K. Vijayaraghavan, The importance of mineral ingredients in biochar production, properties and applications, *Crit. Rev. Environ. Sci. Technol.*, 2021, **51**, 113–139.

34 X.-j. Tong, J.-y. Li, J.-h. Yuan and R.-k. Xu, Adsorption of Cu(II) by biochars generated from three crop straws, *Chem. Eng. J.*, 2011, **172**, 828–834.

35 X. Jin, M.-q. Jiang, X.-q. Shan, Z.-g. Pei and Z. Chen, Adsorption of methylene blue and orange II onto unmodified and surfactant-modified zeolite, *J. Colloid Interface Sci.*, 2008, **328**, 243–247.

36 L. Ton-That, T.-P.-T. Nguyen, B.-N. Duong, D.-K. Nguyen, N.-A. Nguyen, T. H. Ho and V.-P. Dinh, Insights into Pb (II) adsorption mechanisms using jackfruit peel biochar activated by a hydrothermal method toward heavy metal removal from wastewater, *Biochem. Eng. J.*, 2024, **212**, 109525.

37 D.-K. Nguyen, Q.-B. Ly-Tran, V.-P. Dinh, B.-N. Duong, T.-P.-T. Nguyen and P. Nguyen Kim Tuyen, Adsorption mechanism of aqueous Cr(vi) by Vietnamese corncob biochar: a spectroscopic study, *RSC Adv.*, 2024, **14**, 39205–39218.

38 L. Dunnigan, P. J. Ashman, X. Zhang and C. W. Kwong, Production of biochar from rice husk: Particulate emissions from the combustion of raw pyrolysis volatiles, *J. Cleaner Prod.*, 2018, **172**, 1639–1645.

39 J. Zuo, W. Li, Z. Xia, T. Zhao, C. Tan, Y. Wang and J. Li, Preparation of Modified Biochar and Its Adsorption of Cr(VI) in Aqueous Solution, *Coatings*, 2023, 1884.

40 S. Daffalla, Adsorption of Chromium (VI) from Aqueous Solution Using Palm Leaf-Derived Biochar: Kinetic and Isothermal Studies, in *Separations*, 2023.

41 R. Juturu, R. Vinayagam, G. Murugesan and R. Selvaraj, Mesoporous hydrochar from Acacia falcata leaves by

hydrothermal process for hexavalent chromium adsorption, *Sci. Rep.*, 2025, **15**, 12670.

42 A. Tytłak, P. Oleszczuk and R. Dobrowolski, Sorption and desorption of Cr(VI) ions from water by biochars in different environmental conditions, *Environ. Sci. Pollut. Res.*, 2015, **22**, 5985–5994.

43 Y. Chen, B. Wang, J. Xin, P. Sun and D. Wu, Adsorption behavior and mechanism of Cr(VI) by modified biochar derived from Enteromorpha prolifera, *Ecotoxicol. Environ. Saf.*, 2018, **164**, 440–447.

44 L. Zhou, Y. Liu, S. Liu, Y. Yin, G. Zeng, X. Tan, X. Hu, X. Hu, L. Jiang, Y. Ding, S. Liu and X. Huang, Investigation of the adsorption-reduction mechanisms of hexavalent chromium by ramie biochars of different pyrolytic temperatures, *Bioresour. Technol.*, 2016, **218**, 351–359.

45 V.-P. Dinh, M.-D. Nguyen, Q. H. Nguyen, T.-T.-T. Do, T.-T. Luu, A. T. Luu, T. D. Tap, T.-H. Ho, T. P. Phan, T. D. Nguyen and L. V. Tan, Chitosan-MnO2 nanocomposite for effective removal of Cr (VI) from aqueous solution, *Chemosphere*, 2020, **257**, 127147.

46 L. Liu, P. Sun, Y. Chen, X. Li and X. Zheng, Distinct chromium removal mechanisms by iron-modified biochar under varying pH: Role of iron and chromium speciation, *Chemosphere*, 2023, **331**, 138796.

47 B. Wang, J. Xia, L. Mei, L. Wang and Q. Zhang, Highly Efficient and Rapid Lead(II) Scavenging by the Natural Artemia Cyst Shell with Unique Three-Dimensional Porous Structure and Strong Sorption Affinity, *ACS Sustain. Chem. Eng.*, 2018, **6**, 1343–1351.

48 Q. Zhang, Y. Li, Q. Yang, H. Chen, X. Chen, T. Jiao and Q. Peng, Distinguished Cr(VI) capture with rapid and superior capability using polydopamine microsphere: Behavior and mechanism, *J. Hazard. Mater.*, 2018, **342**, 732–740.

49 D. Mohan and C. U. Pittman, Activated carbons and low cost adsorbents for remediation of tri- and hexavalent chromium from water, *J. Hazard. Mater.*, 2006, **137**, 762–811.

50 N. Liu, Y. Zhang, C. Xu, P. Liu, J. Lv, Y. Liu and Q. Wang, Removal mechanisms of aqueous Cr(VI) using apple wood biochar: a spectroscopic study, *J. Hazard. Mater.*, 2020, **384**, 121371.

51 C. Gan, Y. Liu, X. Tan, S. Wang, G. Zeng, B. Zheng, T. Li, Z. Jiang and W. Liu, Effect of porous zinc–biochar nanocomposites on Cr(vi) adsorption from aqueous solution, *RSC Adv.*, 2015, **5**, 35107–35115.

52 J. Wang, X. Yin, W. Tang and H. Ma, Combined adsorption and reduction of Cr(VI) from aqueous solution on polyaniline/multiwalled carbon nanotubes composite, *Korean J. Chem. Eng.*, 2015, **32**, 1889–1895.

53 E. D. Revellame, D. L. Fortela, W. Sharp, R. Hernandez and M. E. Zappi, Adsorption kinetic modeling using pseudo-first order and pseudo-second order rate laws: A review, *Clean Eng. Technol.*, 2020, **1**, 100032.

54 F.-C. Wu, R.-L. Tseng and R.-S. Juang, Characteristics of Elovich equation used for the analysis of adsorption kinetics in dye-chitosan systems, *Chem. Eng. J.*, 2009, **150**, 366–373.

55 N. E. Dávila-Guzmán, F. de Jesús Cerino-Córdova, E. Soto-Regalado, J. R. Rangel-Mendez, P. E. Díaz-Flores, M. T. Garza-Gonzalez and J. A. Loredo-Medrano, Copper Biosorption by Spent Coffee Ground: Equilibrium, Kinetics, and Mechanism, *Clean: Soil, Air, Water*, 2013, **41**, 557–564.

56 L. Largitte and R. Pasquier, A review of the kinetics adsorption models and their application to the adsorption of lead by an activated carbon, *Chem. Eng. Res. Des.*, 2016, **109**, 495–504.

57 V.-P. Dinh, T. D. Xuan, N. Q. Hung, T.-T. Luu, T.-T.-T. Do, T. D. Nguyen, V.-D. Nguyen, T. T. K. Anh and N. Q. Tran, Primary biosorption mechanism of lead (II) and cadmium (II) cations from aqueous solution by pomelo (Citrus maxima) fruit peels, *Environ. Sci. Pollut. Res.*, 2021, **28**, 63504–63515.