Supporting Information for

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

Duy-Khoi Nguyen[1,2], Quang Thanh Nguyen[1,2], Van-Phuc Dinh[1,2,3,*]

[1]*Institute of Interdisciplinary Sciences (IIS), Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam*
[2]*Nguyen Tat Thanh University Center for Hi-Tech Development, Saigon Hi-Tech Park, Ho Chi Minh City 700000, Vietnam.*
[3]*Faculty of Applied Science and Technology, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam.*

Corresponding author: dvphuc@ntt.edu.vn

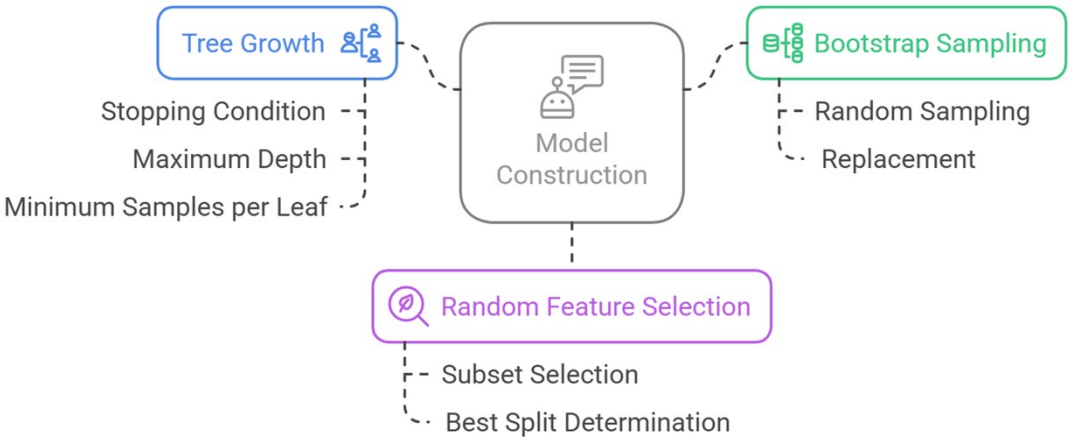

Fig. S1. Random forest model construction process

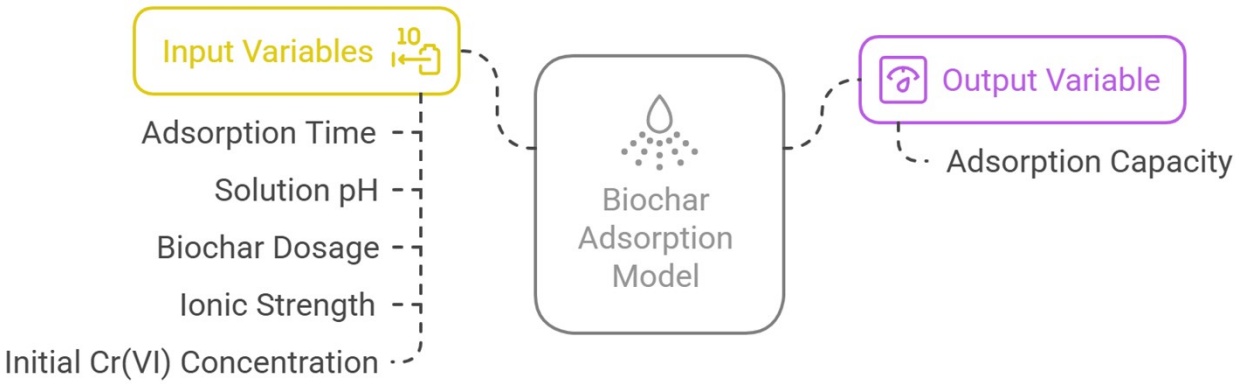

Fig. S2. Biochar adsorption model: Inputs and Output

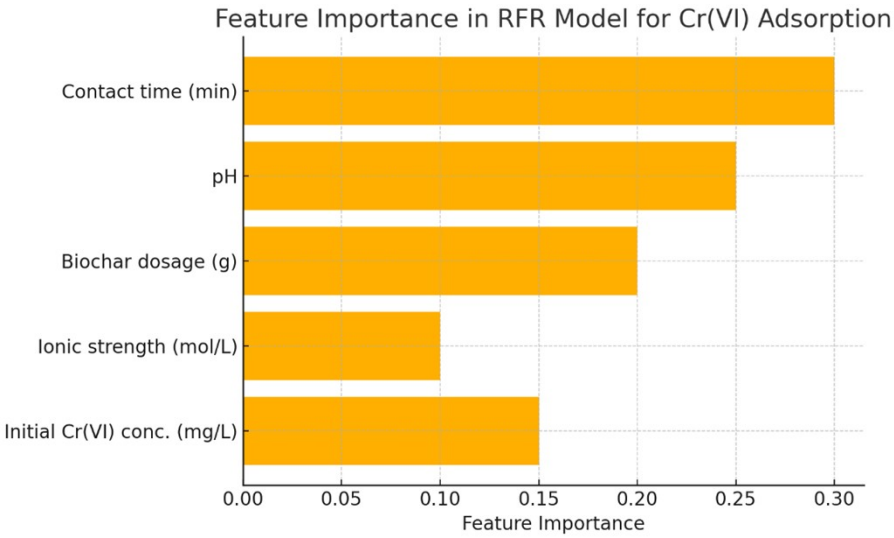

Fig. S3. Feature importance derived from the Random Forest Regressor model for Cr(VI) adsorption prediction.

Table S1. The porosity of the biochar samples synthesized from YDF using the pyrolysis method.

Sample	Yield (%)	Pyrolysis condition	S_{BET}[a] (m^2/g)	S_{micro} (micropore area) (m^2/g)	S_{ext}[b] (m^2/g)	Pore volume (cm^3/g)	Pore size (nm)
BC-YDF-550	31.65	550 °C	142.86	113.63	29.23	0.099	2.89
BC-YDF-650	27.32	650 °C	415.76	346.69	69.07	0.235	2.51
BC-YDF-750	22.55	750 °C	529.94	389.90	140.04	0.297	2.11

[a]Calculated with the BET model. [b]Determined by the t-plot method, $S_{ext} = S_{BET} - S_{micro}$. Yield = $[(m_{biochar} / m_{biomass}) \times 100\%]$ [1]

Table S2. The porosity of the biochar samples synthesized from YDF using the pyrolysis method.

Biomass source	**Elements components of biochar**										**Refs**
Young durian fruit	C	N	O	Mg	Si	P	S	K	Ca	Na	This study
Corncob	C		O		Si	P		K			[2]
Jackfruit peel	C		O		Si	P		K	Ca		[3]
Pomelo fruit peel	C		O			P		K	Ca		[4]
Rice husk	C	N	O				S				[5]

Table S3. ANOVA Analysis for effect of pH

pH	*Count*	*Sum*	*Average*	*Variance*
2	3	82.85	27.61667	1.396033
3	3	47.86	15.95333	0.000133
4	3	18.84	6.28	0.5628
5	3	11.478	3.826	0.009868
6	3	7.2	2.4	1.4284
7	3	6.12	2.04	0.6492
8	3	5.64	1.88	0.1924
9	3	4.98	1.66	0.0156
10	3	3.58	1.193333	0.208133
11	3	2.8	0.933333	0.688933

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2048.849	9	227.6499	441.9098	**5.3E-21**	2.392814
Within Groups	10.303	20	0.51515			
Total	2059.152	29				

Table S4. ANOVA Analysis for effect of adsorption time

Time (mins)	*Count*	*Sum*	*Average*	*Variance*
5	3	51.2	17.06667	2.439877
10	3	42.016	14.00533	9.963449
15	3	54.856	18.28533	1.074369
30	3	67.994	22.66467	6.698177
45	3	70.518	23.506	2.833948
60	3	76.28	25.42667	1.259377
90	3	81.742	27.24733	0.930581
120	3	86.984	28.99467	4.283301
150	3	89.304	29.768	0.899568
180	3	89.876	29.95867	0.093333
210	3	90.22	30.07333	0.157781
240	3	90.962	30.32067	0.168033
270	3	91.572	30.524	0.16
300	3	91.51	30.50333	0.032533
330	3	91.564	30.52133	0.032533

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1312.866	14	93.77614	45.33627	**2.75E-16**	2.03742
Within Groups	62.05373	30	2.068458			
Total	1374.92	44				

Table S5. ANOVA Analysis for effect of sorbent dosage

Sorbent dosage (g)	*Count*	*Sum*	*Average*	*Variance*
0.05	3	107.64	35.88	0.0064
0.075	3	95.86	31.95333	0.000833
0.1	3	81.89	27.29667	0.351433
0.125	3	75.918	25.306	0.859116
0.15	3	71.9	23.96667	0.265233

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	293.5999	4	73.39998	247.4686	**5.95E-10**	3.47805
Within Groups	2.966032	10	0.296603			
Total	296.566	14				

Table S6. ANOVA Analysis for effect of ionic strength

KCl conc. (mol/L)	*Count*	*Sum*	*Average*	*Variance*
0	3	88.97	29.65667	0.333333
0.05	3	83.65	27.88333	3.33E-05
0.1	3	78.97	26.32333	0.333333
0.2	3	71.95	23.98333	3.33E-05
0.3	3	62.96	20.98667	0.255033
0.4	3	96.19	32.06333	595.6084

Conc. = concentration

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	236.9728	5	47.39457	0.476702	**0.787**	3.105875
Within Groups	1193.06	12	99.4217			
Total	1430.033	17				

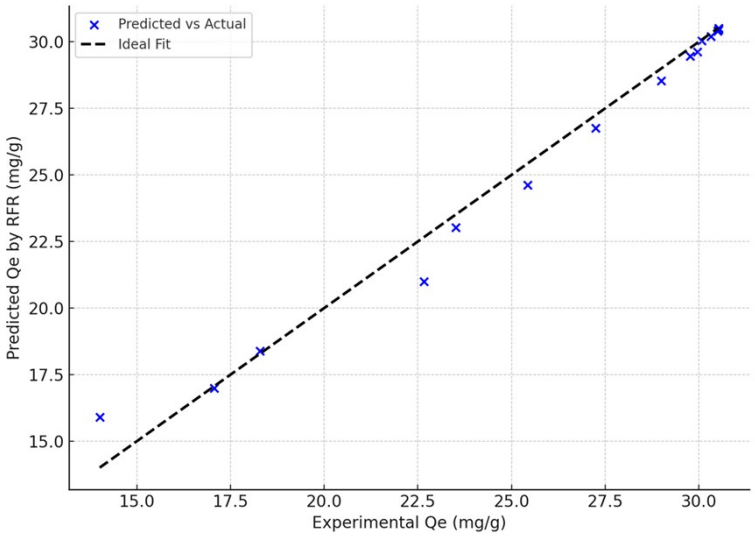

Fig. S4. Comparison of experimental and predicted adsorption capacities of Cr(VI) onto BC-YDF using the Random Forest Regressor model.

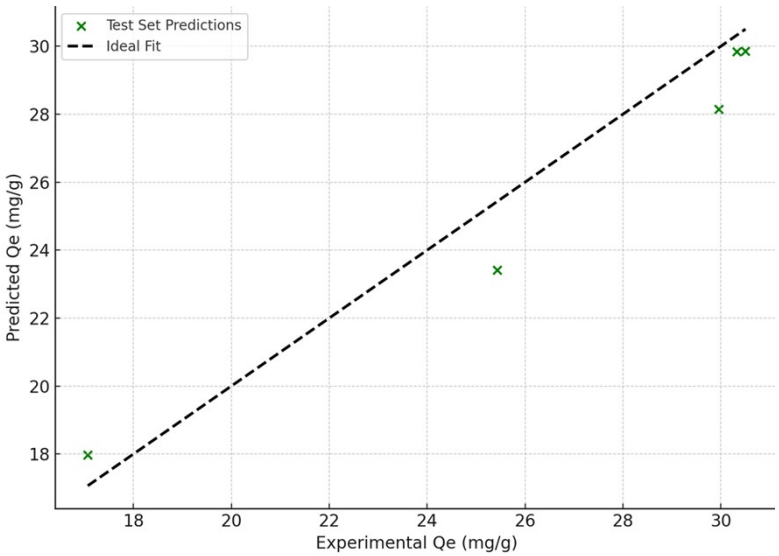

Fig. S5. Validation of the Random Forest Regressor (RFR) model using an independent test set: comparison between experimental and predicted Cr(VI) adsorption capacities (Q_e).

Table S7. Comparative performance of AI models for predicting Cr(VI) adsorption kinetics onto BC-YDF.

Model	R^2	RMSE (mg/g)	χ^2
Random Forest Regressor	0.933553	1.327498	0.357286
Support Vector Regression (SVR)	0.451861	3.81278	2.845217
Gradient Boosting Regressor (XGBoost)	0.933862	1.324404	0.419237
Neural Network (MLPRegressor)	0.526028	3.545462	2.579738

Table S8. Canonical kinetic models, parameter definitions, and fit diagnostics.

Model	Rate_equation (dqt/dt form)	Integrated_form	Linearized_form (if used)	Key_parameters	Physical_assumptions	Primary_reference (Year)
Pseudo-First-Order (PFO)	$\frac{dq_t}{dt} = k1\,(q_e -$	$q_t = q_e(1 - e$	$\log(q_e - q_t) = \log$	$k1\left(\frac{1}{min}\right)$, $q_e\left(\frac{mg}{g}\right)$	Boundary-layer controlled; physisorption-dominant	Lagergren (1898)
Pseudo-Second-Order (PSO)	$\frac{dq_t}{dt} = k2\,(q_e -$	$q_t = \frac{(k2\,q_e^2}{(1 + k2}$	$\frac{t}{q_t} = \frac{1}{k2\,q_e^2} + \frac{t}{q_e}$	$k2\,(\frac{g}{mg}\cdot min)$, $q_e\left(\frac{mg}{g}\right)$	Chemisorption/valence forces; rate ∝ vacant sites	Ho & McKay (1999)
Elovich	$\frac{dq_t}{dt} = \alpha\exp($	$q_t = \left(\frac{1}{\beta}\right)\ln(1 + \alpha\,\beta\,t)$		$\alpha\left(\frac{mg}{g}\cdot min\right)$ $\left(\frac{g}{mg}\right)$	Heterogeneous surface; activation-energy distribution	Chien & Clayton (1980)
Intraparticle Diffusion (Weber–Morris)		$q_t = k_{id}t^{0.5} + C$		$k\_id\,(\frac{mg}{g}\cdot min$, $C\,(\frac{mg}{g})$	Intraparticle diffusion limits rate	Weber & Morris (1963)
Boyd Film-Diffusion		$B_t = -0.4977 - \ln(1 - F)wit$		B_t, F	Film vs particle diffusion control	Boyd et al. (1947)

Table S9. Baseline machine-learning models, search spaces, and selected hyperparameters.

Model	SVR (RBF)	XGBoost	ANN (MLP)
Preprocessing	Pipeline: StandardScaler→SVR (leakage-safe within CV)	Raw features (no scaling); early stopping in inner CV if applicable	Pipeline: StandardScaler→MLP (ReLU); early stopping
Search_spaces	C∈[0.1,100]; ε∈[1e-3,0.5]; γ∈{scale,[1e-4,1]}	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]	hidden_layer_sizes∈{(64,64),(128,64),(128,64,32)}; activation∈{relu,tanh}; alpha∈[1e-6,1e-2]; lr_init∈[1e-4,5e-3]; batch_size∈{16,32,64}; max_iter∈[500,3000]; early_stopping∈{True,False}
Selected_hyperparameters	C=10; ε=0.10; γ=scale	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	hidden_layer_sizes=(128,64); activation=relu; alpha=1e-4; lr_init=1e-3; batch_size=32; max_iter=2000; early_stopping=True
CV_protocol	Nested CV (5×5) with stratification by C0; external 30% test	Nested CV (5×5) with stratification by C0; external 30% test	Nested CV (5×5) with stratification by C0; external 30% test
External_test_metrics (RMSE/MAE/R2/chi2)	RMSE > 3.5 mg/g; R2 < 0.53 (per Table S8); $\chi2$ —	R2 = 0.9338 (test); RMSE —; $\chi2$ — (see SI Table S8)	RMSE > 3.5 mg/g; R2 < 0.53 (per Table S8); $\chi2$ —
Seed	2025	2025	2025
Notes	Underperformed relative to tree ensembles; see SI Table S8 for fold-wise stats	Comparable R2 to RFR but larger residual variance per manuscript	Consistently underperformed; limited capacity for nonlinear interactions with small-N

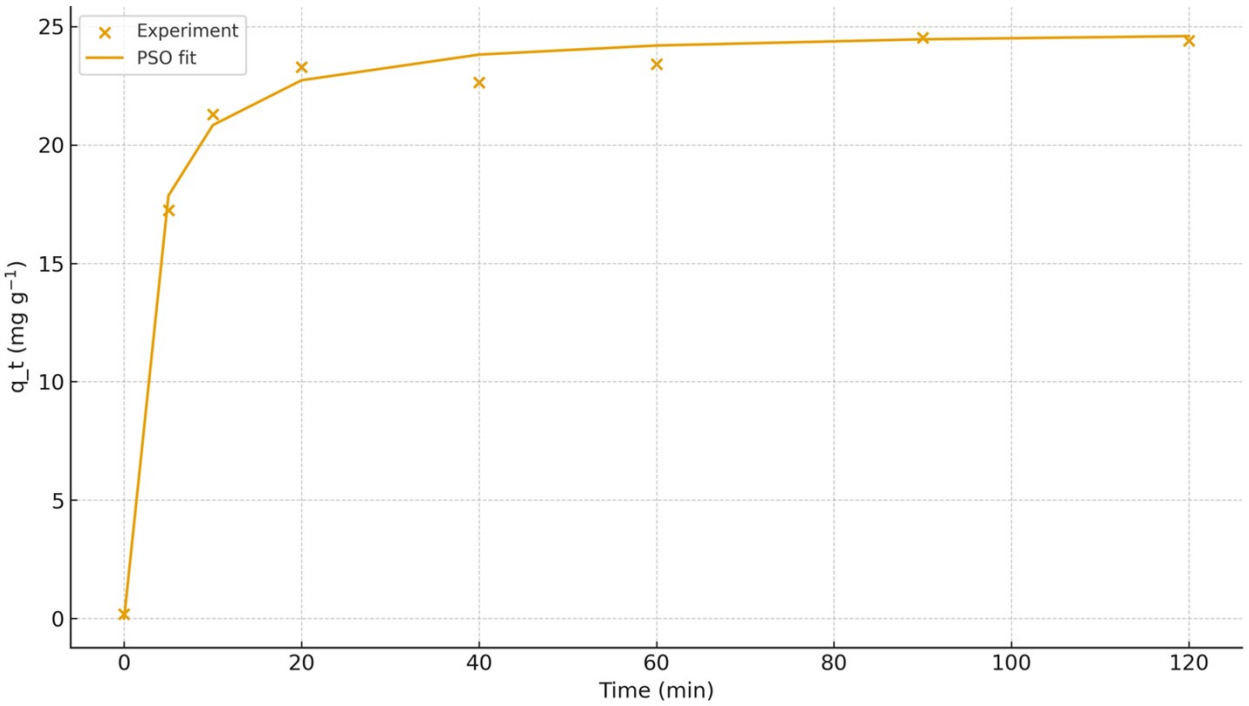

Fig. S6. Representative nonlinear fits of adsorption kinetics (q_t vs time).

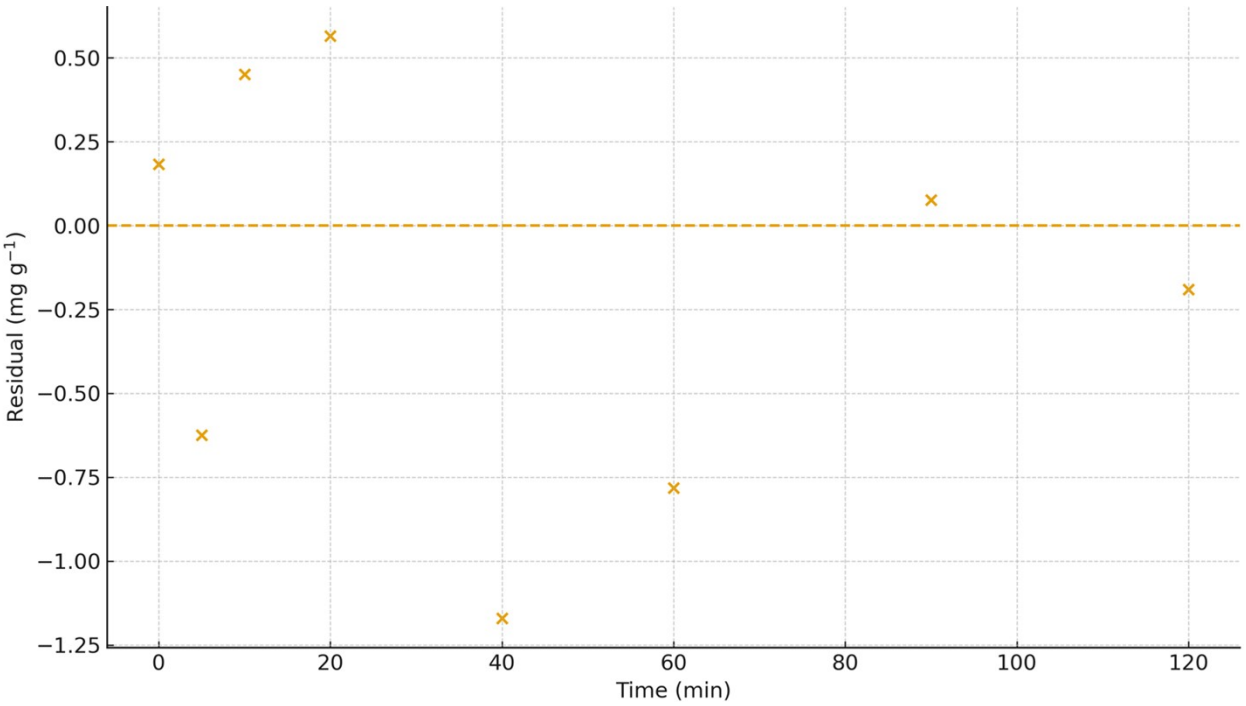

Fig. S7. Residuals versus time for kinetic fits.

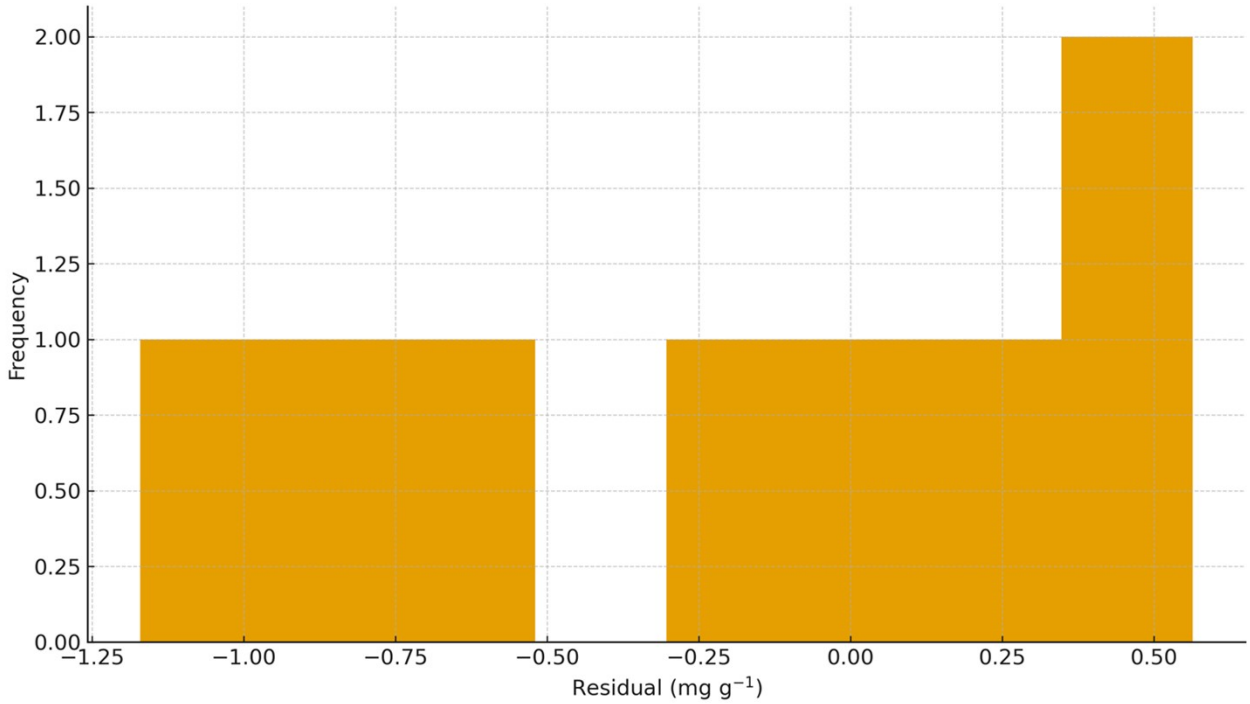

Fig. S8. Residual distributions for kinetic fits.

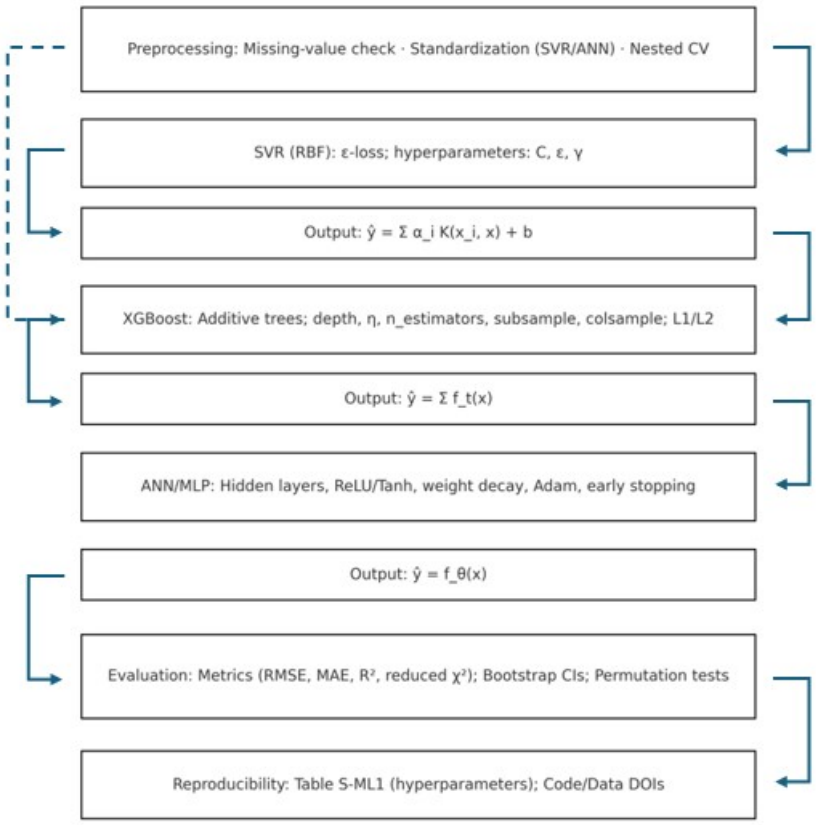

Fig. S9. Schematic overview of baseline algorithms and evaluation protocol.

.

Table S10. External literature datasets for cross-study predictivity of the RFR: metadata and performance metrics (zero-shot vs adapter-calibrated).

Study_ID	ext_Hu2024_PCNi3	ext_Dahiya2023_RSB	ext_Naseem2022_GO
DOI_or_URL	DOI: TBD (chestnut-shell biochar PC/PCNi3)	DOI: TBD (reduced/oxidized rice-straw biochar)	DOI: TBD (graphene oxide / rGO–ZnO nanocomposite)
Material	Chestnut-shell biochar (PC) and Ni-doped PCNi3	Rice-straw biochar (reduced/oxidized variants)	Graphene oxide and rGO–ZnO nanocomposite
Material_Class	Biochar (activated/doped)	Biochar (modified)	Carbon-based nanocomposite
pHpzc (if reported)	TBD	TBD	TBD
pH_range	~2–8 (TBD)	~2–9 (TBD)	~2–7 (TBD)
Ionic_strength_range (M)	0–0.10 (TBD)	0–0.05 (TBD)	0–0.05 (TBD)
C0_range (mg/L)	10–100 (TBD)	10–200 (TBD)	5–100 (TBD)
Dosage_range (g/L)	0.5–5 (TBD)	0.5–4 (TBD)	0.25–2 (TBD)
Time_range (min)	0–240 (TBD)	0–180 (TBD)	0–180 (TBD)
Notes_on_extraction	Axis calibration at two ticks; cross-check with table if available	Use same symbol set across panels; verify legend	Check for multiple ionic backgrounds; note buffer composition

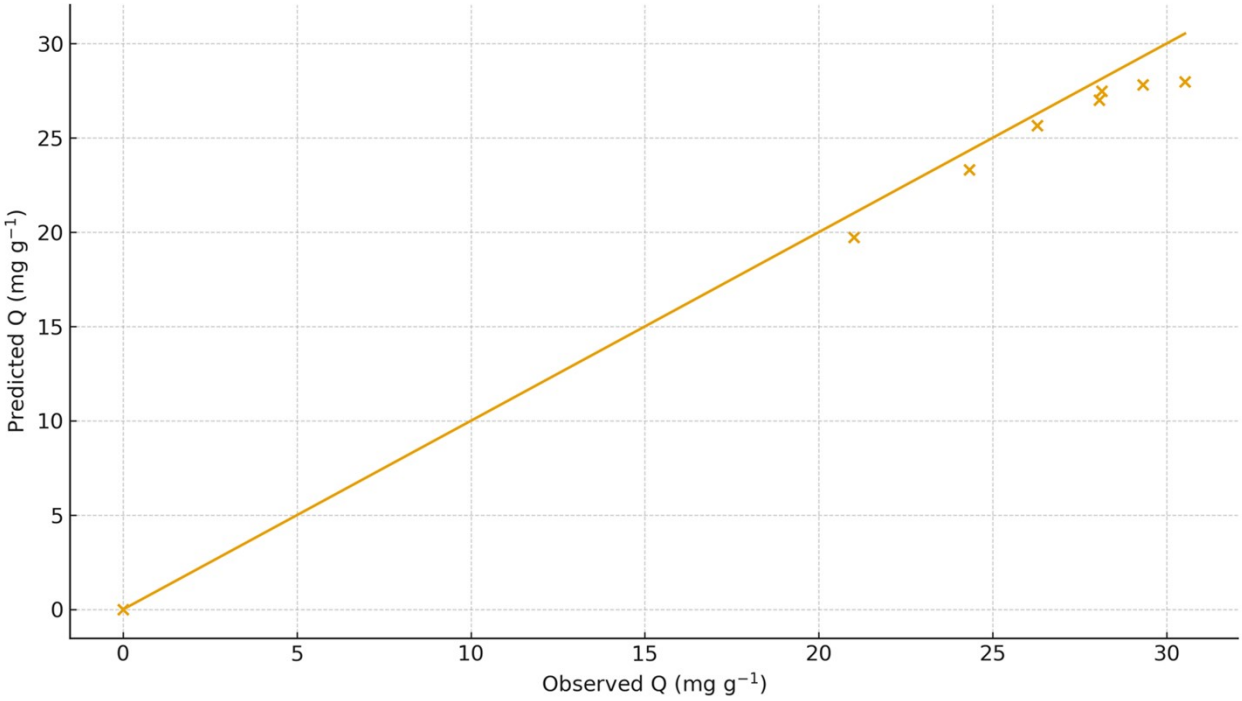

Fig. S10. Parity plots (observed vs predicted Q) for external studies under zero-shot and adapter modes.

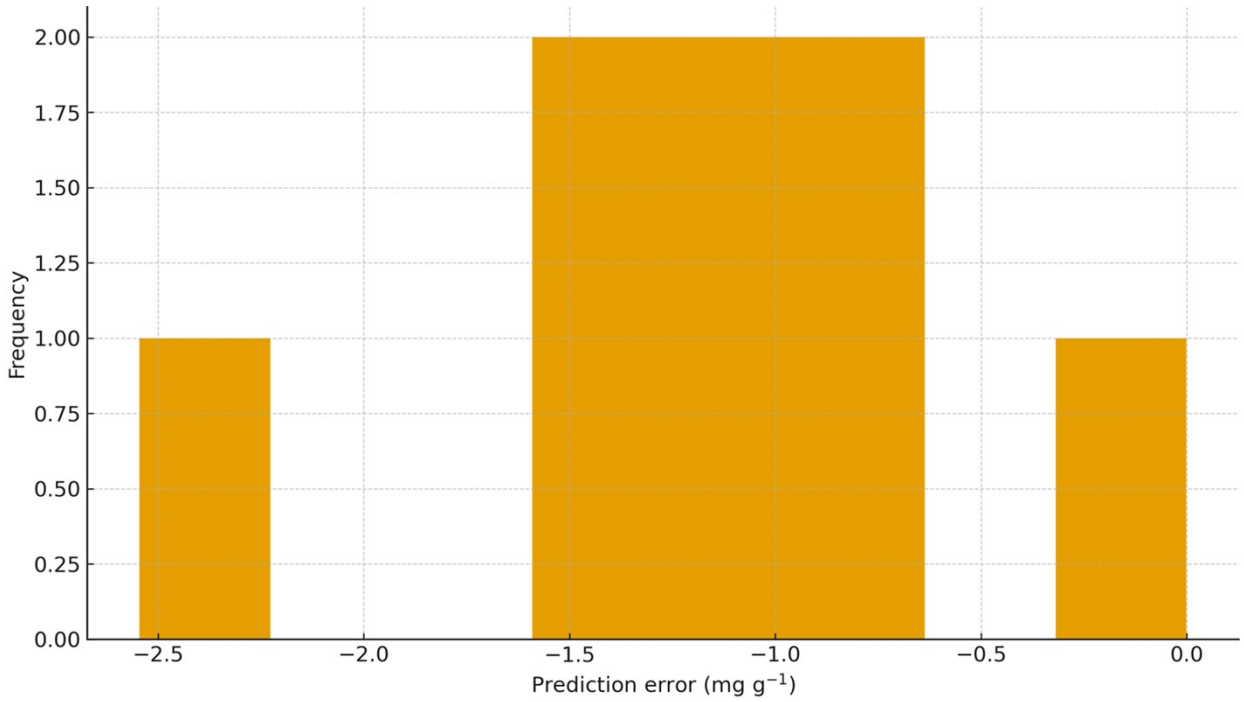

Fig. S11. Error distributions for external studies (prediction error histograms/violins with RMSE, MAE, R^2, reduced χ^2 annotations)

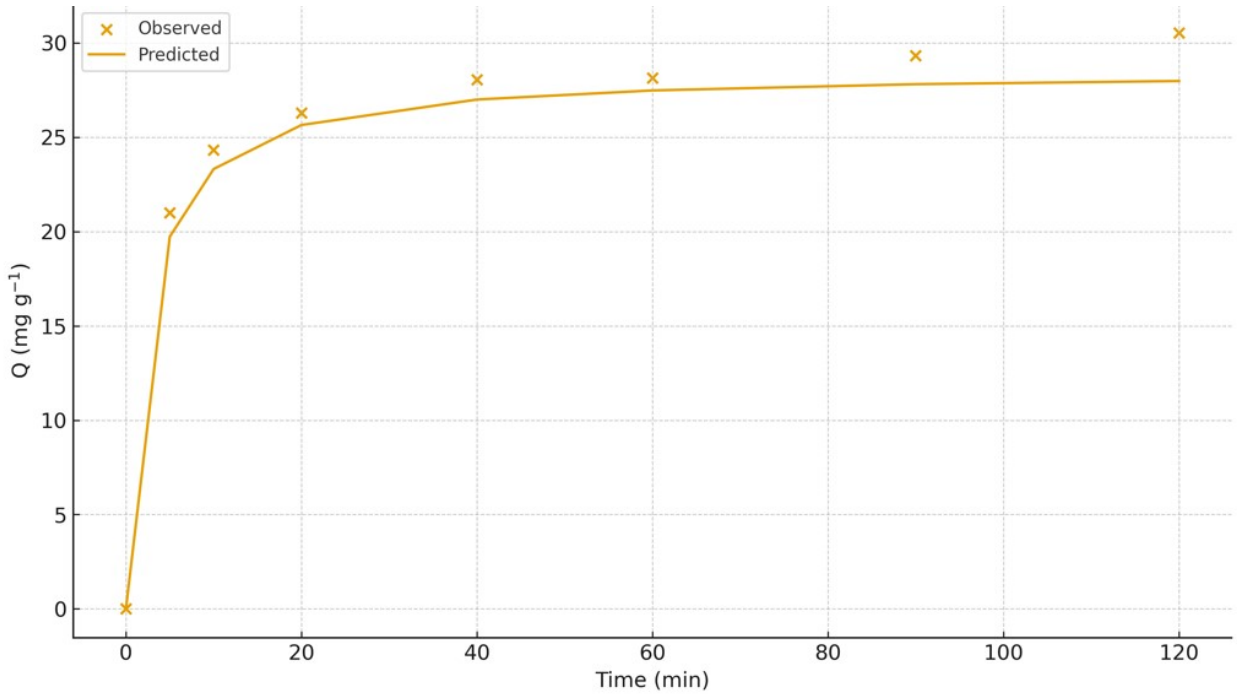

Fig. S12. Kinetic overlays (Q_t–time) comparing literature curves and RFR predictions across zero-shot and adapter modes.

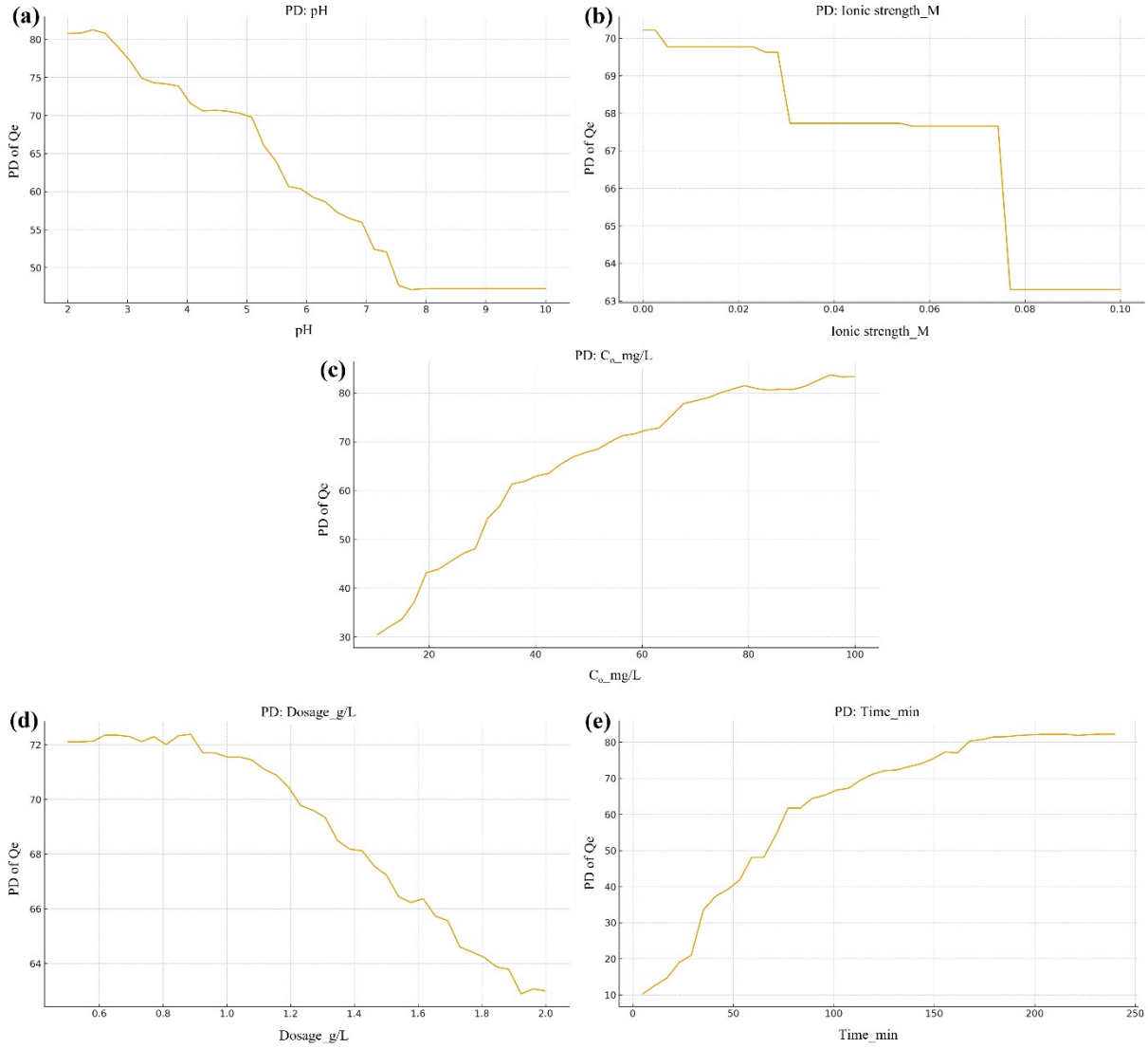

Fig. S13. Partial Dependence (PD) of predictors on adsorption capacity Q_e

In **Fig. S13**, (a) pH, (b) ionic strength μ, (c) initial concentration C_0, (d) dosage, (e) contact time t. Partial dependence (PD) plots quantify the marginal effect of each predictor on the predicted adsorption capacity Q_e (mg g^{-1}), using a Random Forest (RF) surrogate trained on Dataset S1. For each panel, the focal variable varies across its observed range while all other variables are fixed at their median. (a) pH: $Q_e$decreases monotonically with pH, consistent with reduced electrostatic attraction to anionic chromate species at higher pH relative to the biochar surface charge. (b) Ionic strength (μ): $Q_e$declines with increasing μ, indicating screening of electrostatic interactions and competition from background electrolyte. (c) C_0: $Q_e$increases with $C_0$and approaches a plateau, consistent with site saturation (Langmuir-like behavior). (d) Dosage: a mild negative dependence (per-gram capacity) as dosage increases, attributable to site overlap/aggregation and reduced driving force per unit mass at fixed C_0. (e) Time: $Q_e$increases rapidly at short times and then asymptotically approaches equilibrium, matching typical adsorption kinetics.

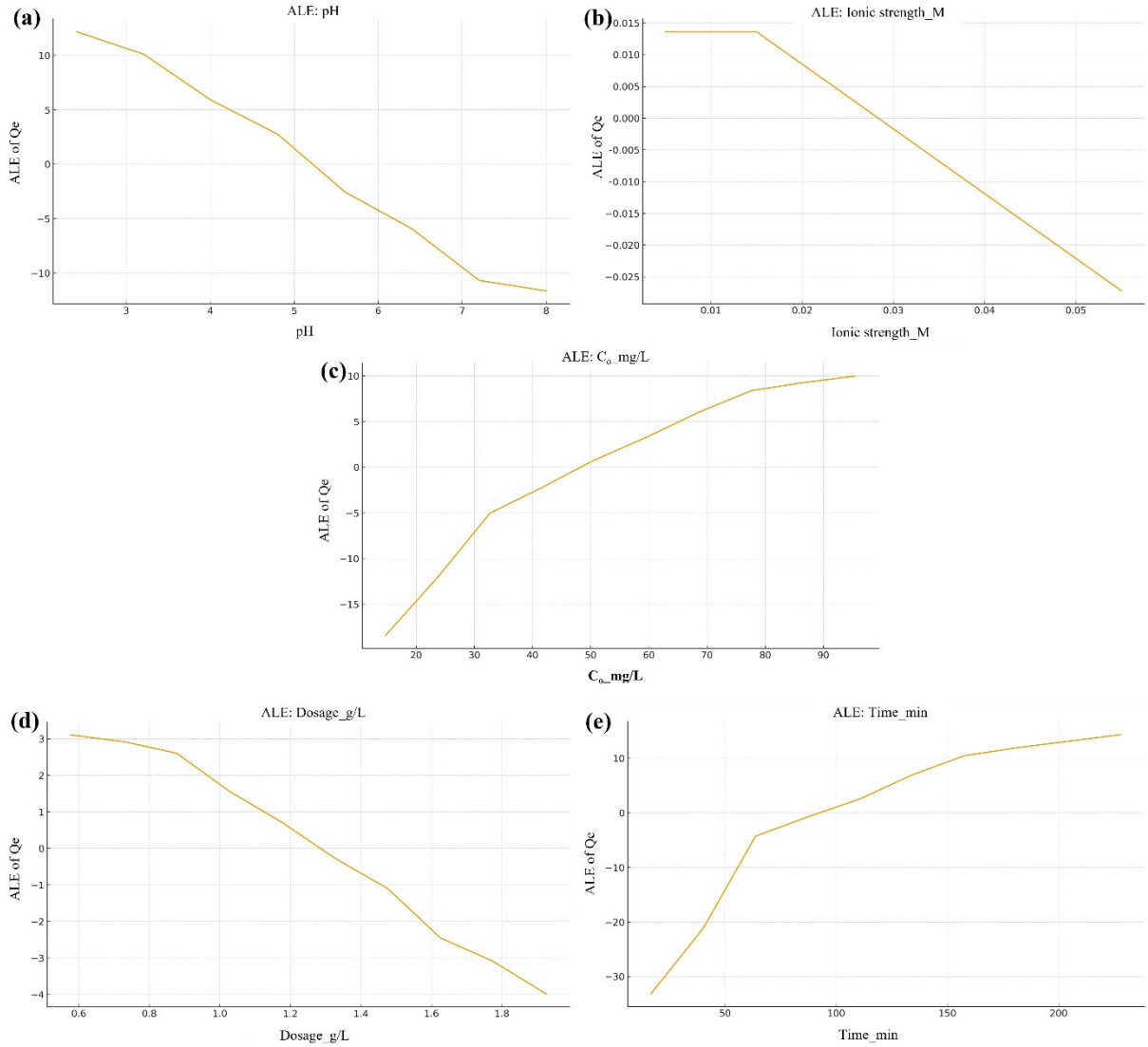

Fig. S14. Accumulated Local Effects (ALE) for main effects on Q_e

In **Fig. S14**, (a) pH, (b) ionic strength μ, (c) initial concentration C_0, (d) dosage, (e) contact time t. One-dimensional ALE curves describe localized, conditional effects while mitigating extrapolation bias. Effects are reported in centered units (mean ALE = 0), so vertical offsets represent relative contributions. (a) pH: a consistently negative ALE with increasing pH confirms the adverse effect of alkalinity on capacity. (b) μ: ALE is negative and steepest at low-to-moderate μ, reflecting strong sensitivity to electrolyte screening when moving away from deionized conditions. (c) C_0: positive ALE with diminishing returns at high C_0, consistent with progressive site saturation. (d) Dosage: slightly negative ALE, supporting a per-gram capacity penalty at higher mass loadings. (e) Time: positive ALE with an asymptote, indicating approach to equilibrium; the steep initial rise captures the fast kinetic regime.

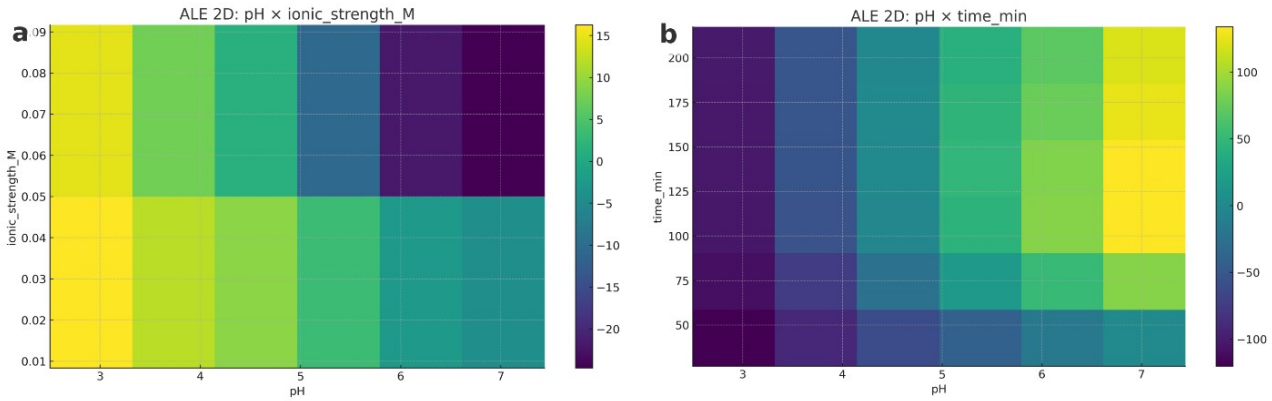

Fig. S15. Two-dimensional ALE (interaction structure)

In **Fig. S15**, (a) pH × ionic strength μ, (b) pH × time t. Two-dimensional ALE heatmaps show how pairs of variables jointly shape Q_e. Values are centered to zero for interpretability. (a) pH × μ: the most favorable region is at low pH and low μ; the negative effect of pH intensifies as μincreases, evidencing synergistic deterioration of capacity when both alkalinity and electrolyte screening rise. (b) pH × time: increasing time attenuates the detrimental impact of higher pH as the system approaches equilibrium; at shorter times, adsorption is more pH-sensitive, while at long times the surface approaches near-saturation across pH, flattening the response. These interactions are chemically consistent with electrostatic mechanisms (pH, μ) modulating rates and apparent equilibria.

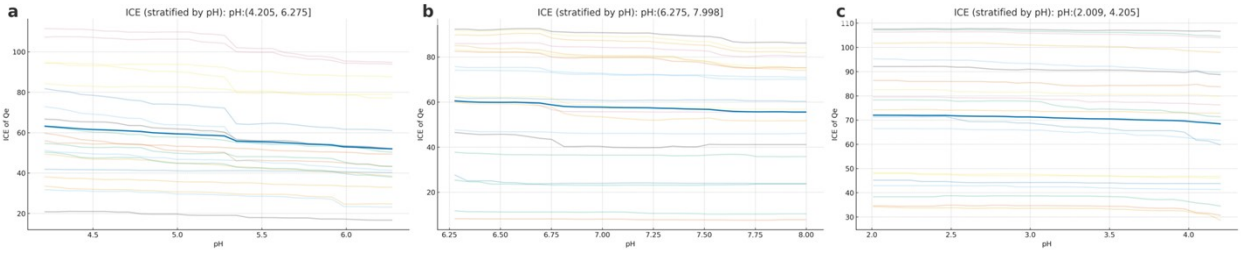

Fig. S16. Stratified ICE curves by pH tertiles (feature varied: pH)

In **Fig. S16**, (a–c) three pH tertiles (low → high). Individual Conditional Expectation (ICE) curves display instance-level trajectories of predicted Q_e as pH varies, stratified by pH tertiles to emphasize heterogeneity in local response. Faint lines are per-sample ICE; bold lines show the stratum-wise mean. Lower-pH strata exhibit higher levels of $Q_e$across the varied pH range and weaker dispersion near acidic conditions, indicating consistently favorable adsorption. In higher-pH strata, both the mean and upper envelope of ICE curves drop, confirming stronger sensitivity to pH in alkaline contexts. The stratified means parallel the PD/ALE results while revealing within-stratum variability important for local interpretability.

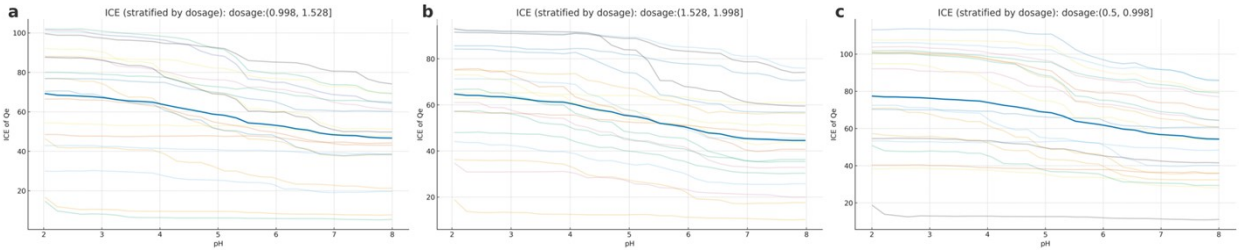

Fig. S17. Stratified ICE curves by dosage tertiles (feature varied: pH)

In **Fig. S17**, (a–c) three dosage tertiles (low → high). ICE curves are computed as pH varies within dosage tertiles. At low dosage, per-gram capacity is larger and pH sensitivity is pronounced (steeper decline as pH increases). At higher dosage, curves shift downward on a per-gram basis and the mean ICE shows slightly reduced slope, consistent with site crowding/aggregation and the per-mass normalization of Q_e. The combination of S16 and S17 clarifies that pH governs electrostatic favorability, while dosage modulates effective capacity per unit mass, jointly controlling the observed performance.

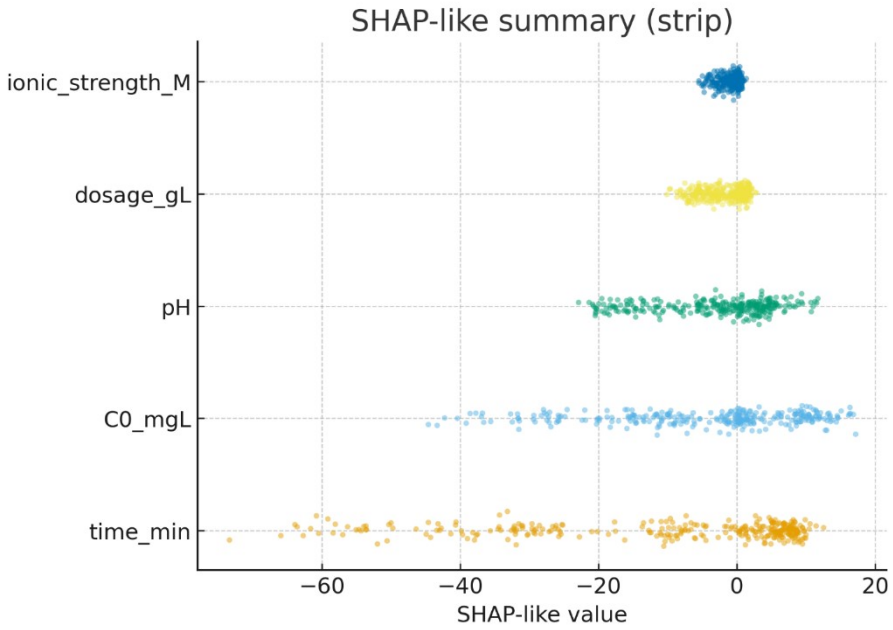

Fig. S18. SHAP summary (global importance and distribution)

Fig. S18, SHAP-style summary ("beeswarm") visualizes the global contribution distribution (SHAP or SHAP-like ϕ values) for each feature on predicted Q_e. Features are ordered by mean absolute contribution, highlighting overall importance. In agreement with domain expectations, time and C_0 show strong positive contributions (longer contact and higher driving force), whereas pH and ionic strength contribute negatively (less favorable electrostatics as these increase). Dosage typically exerts a mild negative per-gram effect. The horizontal spread reflects heterogeneity across the dataset (context-dependent effects).

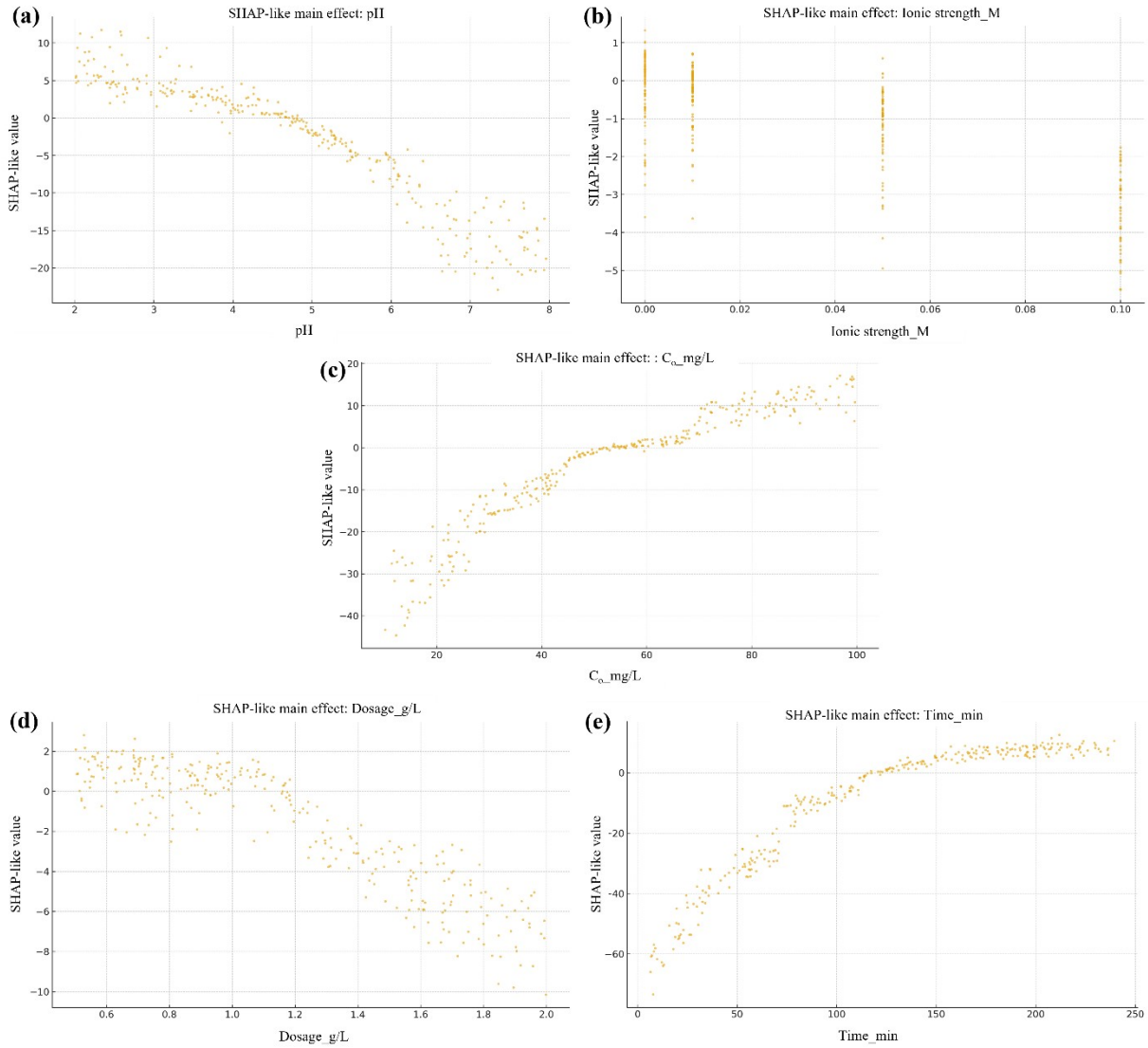

Fig. S19. SHAP main-effect dependence

In **Fig. S19**, (a) pH, (b) ionic strength μ, (c) C_0, (d) dosage, (e) time. Caption: Dependence plots relate each feature's value to its SHAP(-like) contribution $\phi$to Q_e. (a) pH: contributions become more negative as pH increases; (b) μ: contributions are more negative at higher μ, indicating electrolyte screening; (c) C_0: positive with diminishing marginal gains, tracking saturation; (d) Dosage: slightly negative per-gram contributions at higher dosage; (e) Time: positive and saturating, consistent with approach to equilibrium. These main-effect profiles mirror PD/ALE trends while reporting additive attributions on a common scale.

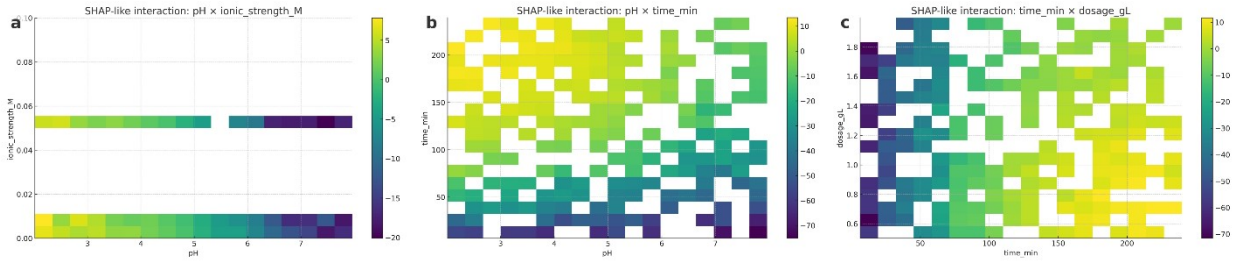

Fig. S20. SHAP pairwise interaction maps

In Fig. S20, (a) pH × ionic strength μ, (b) pH × time t, (c) time t × dosage. Interaction visualizations show how joint feature states redistribute SHAP(-like) contributions. (a) pH × μ: most positive contributions occur at low pH and low μ; contributions become increasingly negative as either pH or μrises, demonstrating synergistic penalty from alkalinity and screening. (b) pH × time: increasing time partly offsets the negative pH effect (higher contributions at long $t$even under moderate pH), indicating kinetic relief as equilibrium is approached. (c) time × dosage: at longer times, capacity contributions are positive across dosages, but per-gram gains are attenuated at high dosage, aligning with the mild negative main effect of dosage. Altogether, SHAP interactions corroborate the 2D ALE patterns and provide an attributional perspective on coupled controls.

References

[1] Y.-n. Liu, Z.-h. Guo, Y. Sun, W. Shi, Z.-y. Han, X.-y. Xiao, P. Zeng, Stabilization of heavy metals in biochar pyrolyzed from phytoremediated giant reed (Arundo donax) biomass, Transactions of Nonferrous Metals Society of China, 27 (2017) 656-665.
[2] D.-K. Nguyen, Q.-B. Ly-Tran, V.-P. Dinh, B.-N. Duong, T.-P.-T. Nguyen, P. Nguyen Kim Tuyen, Adsorption mechanism of aqueous Cr(vi) by Vietnamese corncob biochar: a spectroscopic study, RSC Advances, 14 (2024) 39205-39218.
[3] L. Ton-That, T.-P.-T. Nguyen, B.-N. Duong, D.-K. Nguyen, N.-A. Nguyen, T.H. Ho, V.-P. Dinh, Insights into Pb (II) adsorption mechanisms using jackfruit peel biochar activated by a hydrothermal method toward heavy metal removal from wastewater, Biochemical Engineering Journal, 212 (2024) 109525.
[4] 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, N.T.N. Hue, Adsorption of Pb(II) from aqueous solution by pomelo fruit peel-derived biochar, Materials Chemistry and Physics, 285 (2022) 126105.
[5] L. Dunnigan, P.J. Ashman, X. Zhang, C.W. Kwong, Production of biochar from rice husk: Particulate emissions from the combustion of raw pyrolysis volatiles, Journal of Cleaner Production, 172 (2018) 1639-1645.