

Article

Feedforward and Recurrent Neural Networks for Reliability-Aware Modeling of Chromium Breakthrough in Fixed-Bed Biosorption Columns

Alma Rosa Netzahuatl-Muñoz ^{1,*} , Juan Crescenciano Cruz-Victoria ² , Erick Aranda-García ³  and Eliseo Cristiani-Urbina ³ 

- ¹ Programa Académico de Ingeniería en Biotecnología, Universidad Politécnica de Tlaxcala, Avenida Universidad Politécnica No. 1, San Pedro Xalcaltzinco, Tepeyanco 90180, Tlaxcala, Mexico
- ² Programa Académico de Ingeniería Mecatrónica, Universidad Politécnica de Tlaxcala, Avenida Universidad Politécnica No. 1, San Pedro Xalcaltzinco, Tepeyanco 90180, Tlaxcala, Mexico; juancrescenciano.cruz@uptlax.edu.mx
- ³ Instituto Politécnico Nacional, Escuela Nacional de Ciencias Biológicas, Departamento de Ingeniería Bioquímica, Avenida Wilfrido Massieu s/n, Unidad Profesional Adolfo López Mateos, Alcaldía Gustavo A. Madero, Ciudad de México 07738, Mexico; earanda@ipn.mx (E.A.-G.); ecristiani@ipn.mx (E.C.-U.)
- * Correspondence: almarosa.netzahuatl@uptlax.edu.mx

Abstract

Breakthrough-curve prediction is central to fixed-bed biosorption column design, but data-driven modeling remains challenging with small, unevenly distributed datasets. This study compares a feedforward multilayer perceptron (MLP) with recurrent long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) networks for simultaneous prediction of hexavalent chromium [Cr(VI)] and total chromium breakthrough curves in columns packed with *Quercus crassipes* acorn shell. A dataset of 14 experimental curves was used, covering variations in pH, influent concentration, flow rate, bed height, and biosorbent mass. Model performance was assessed using Mahalanobis distance (D_M)-based holdout selection, curve-shape evaluation, inter-model coefficient of variation (CoV), and SHAP attribution. All architectures achieved high internal cross-validation accuracy ($R^2 = 0.9879$ – 0.9979 ; RMSE = 0.0149 – 0.0309). However, complete-curve holdout performance depended on the excluded condition and chromium response. Under the central holdout condition ($D_M = 0.317$), R^2 remained high for both outputs (0.954 – 0.990). In non-central holdouts, errors also affected curve shape, including breakthrough timing and the approach to saturation. The largest deterioration occurred for Cr(VI) under the peripheral low-flow condition, where the MLP decreased to $R^2 = 0.325$, while LSTM and Bi-LSTM retained higher agreement. SHAP attribution confirmed process-consistent effects, with influent concentration and flow rate as dominant predictors and pH contributing more strongly to Cr(VI). These findings show that recurrent architectures can provide more robust representations of complex breakthrough profiles, while CoV– D_M analysis helps identify regions where additional experiments are needed.

Keywords: Cr(VI) removal; total chromium removal; biosorption capacity; Mahalanobis distance; model consensus; SHAP



Academic Editors: Adolph Anga Muleja, Mukuna Patrick Mubiayi, Gcina Mamba, Adewale Adewuyi and Charmaine Sesethu Tshangana

Received: 15 June 2026

Revised: 16 July 2026

Accepted: 24 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Adsorption has become one of the most versatile and effective technologies for removing a wide spectrum of pollutants from water and wastewater, including heavy metals, dyes,

pharmaceuticals, pesticides, and emerging organic contaminants [1–3]. Its widespread adoption stems from operational simplicity, effective removal across a broad range of pollutant concentrations, and the potential for adsorbent regeneration [4]. Biosorption—the use of biological materials as adsorbents—has attracted considerable interest as a sustainable alternative to conventional adsorbents such as activated carbon, zeolites, and clay minerals [5]. Agricultural residues, forestry byproducts, and microbial biomass provide abundant, renewable, and low-cost sources of biosorbents, whose surface functional groups confer high affinity for diverse pollutants [6–8]. The valorization of these waste materials into effective adsorbents addresses waste management challenges while reducing treatment costs, thereby aligning with circular economy principles [7].

Fixed-bed column adsorption is the preferred configuration for large-scale, continuous wastewater treatment. The successful design and scale-up of such systems rely on accurate prediction of breakthrough curves, which characterize the dynamic response of the column and determine critical operational parameters such as service time, regeneration frequency, and bed utilization efficiency [9,10]. Conventional modeling approaches have relied predominantly on semi-empirical models, including the Thomas, Bohart–Adams, Yoon–Nelson, and dose–response models [11,12]. These models are useful for describing breakthrough behavior under specific conditions and for preliminary design purposes, offering mathematical simplicity and physically interpretable parameters [13]. However, their applicability is constrained by inherent assumptions. They typically postulate a single rate-controlling mechanism governing the entire process, neglecting the interplay among film diffusion, pore diffusion, and surface reaction that occurs in real systems [9]. Moreover, they impose symmetric sigmoidal shapes that may fail to describe the asymmetric breakthrough profiles often observed experimentally, particularly near saturation [14].

To address these limitations, two complementary strategies have been pursued. The first involves mechanistically enriched transport models that explicitly account for intra-particle diffusion and adsorption kinetics, including approaches based on fractional-order differential equations that capture anomalous diffusion behavior in porous adsorbents [15]. The second relies on data-driven methods based on machine learning (ML), particularly artificial neural networks (ANNs), which can model complex nonlinear relationships without requiring explicit mechanistic assumptions [16–18].

Among approaches for fixed-bed column modeling, feedforward ANNs—particularly the multilayer perceptron (MLP)—have been the predominant architecture. MLP models have been applied to predict breakthrough curves for a variety of pollutants, including dyes [19], metal ions such as Cu(II) and Zn(II) [20], hexavalent Cr [21–23], fluoride [24], pharmaceutical compounds [25], and organic micropollutants [16]. Ensemble methods, such as random forest and gradient boosting algorithms, have also been used for breakthrough curve prediction in systems involving selenite [26], herbicides [27], and organic dyes [28]. However, these approaches share a common limitation: breakthrough data are typically treated as independent observations, overlooking the inherently sequential nature of the adsorption process, in which the effluent concentration at a given time depends on the preceding loading history of the bed. Recurrent architectures, such as long short-term memory (LSTM) and bidirectional long short-term memory (Bi-LSTM) networks, are specifically designed to capture temporal dependencies in sequential data [29,30]. Therefore, they provide a relevant but still underexplored alternative for breakthrough-curve modeling in fixed-bed adsorption and biosorption systems.

Another practical limitation in this field is the small size of most fixed-bed experimental datasets. Complete breakthrough experiments require long operation times and multiple sampling points, which often restricts the number of operating conditions available for model development. Under these conditions, high validation accuracy may not

be sufficient to establish model reliability, particularly when predictions are required for operating conditions that are distant from those used for training. This issue is relevant for neural-network models because their extrapolation behavior can deteriorate outside the training distribution [31,32]. Accordingly, model evaluation should consider not only global accuracy metrics but also the response of the model under progressively less represented operating conditions.

Hexavalent chromium [Cr(VI)] represents a particularly challenging system for breakthrough curve modeling because Cr(VI) biosorption involves both adsorption and partial reduction to trivalent chromium [Cr(III)] at the biosorbent surface. The extent of reduction depends on factors such as pH, contact time, and the nature of surface functional groups [33]. The lignocellulosic biomass of *Quercus crassipes* acorn shell, which is rich in hydroxyl and carboxyl groups capable of participating in both processes, exemplifies this complexity [34]. Cr biosorption breakthrough curves exhibit shapes that vary with operating conditions, and conventional semi-empirical models have shown variable predictive performance when applied to this system, illustrating their limitations in the presence of coupled removal mechanisms [35].

This study aims to develop and compare multivariate feedforward and recurrent neural network architectures for the simultaneous prediction of Cr(VI) and total Cr breakthrough curves in fixed-bed columns packed with *Quercus crassipes* acorn shell, using the experimental dataset reported by Aranda-García and Cristiani-Urbina [35]. The specific objectives are:

- i. to implement a Mahalanobis distance (D_M)-based holdout strategy to select curves located at progressively different regions of the experimental space for systematic model evaluation;
- ii. to optimize and validate MLP, LSTM, and Bi-LSTM architectures through Bayesian hyperparameter tuning and cross-validation within each holdout scenario;
- iii. to assess inter-model consensus on predicted biosorption capacity across the operational space by combining the coefficient of variation (CoV) with D_M as a descriptor of the relative location of operating conditions;
- iv. to apply SHapley Additive exPlanations (SHAP) analysis to identify the key variables governing model predictions and to examine their temporal evolution throughout the breakthrough process.

2. Materials and Methods

2.1. Experimental Data

The experimental data used in this study were obtained from fixed-bed column biosorption experiments previously reported by Aranda-García and Cristiani-Urbina [35]. A total of 14 breakthrough curves were generated under varying operational conditions using a one-factor-at-a-time (OFAT) approach. Each breakthrough curve was characterized by the normalized effluent concentration as a function of time (C_t/C_0), where C_t is the effluent concentration at time t and C_0 is the influent concentration. This ratio was determined independently for Cr(VI) and total Cr, yielding two output variables per time point. Table 1 summarizes the input variables and their respective ranges.

Although bed height (H) and biosorbent mass (m) are related through column geometry and packing density, both variables were included as model inputs to capture potential independent effects on breakthrough behavior.

Because the dataset consisted of a limited number of complete breakthrough curves generated under an OFAT design, the modeling strategy emphasized architecture comparison, holdout-based generalization assessment, and diagnostic analysis of prediction agreement rather than broad extrapolation beyond the experimental domain.

Table 1. Input variables and their ranges for breakthrough-curve modeling.

Variable	Symbol	Min	Max	Units
Operation time ¹	t	2.0	218.0	h
Influent solution pH	pH	1.0	3.0	—
Influent Cr(VI) concentration	C_0	50.0	400.0	mg L ^{−1}
Influent flow rate	Q	0.25	1.0	mL min ^{−1}
Bed height	H	1.7	6.5	cm
Biosorbent mass	m	1	4	g

¹ Maximum operation time corresponds to the longest curve; terminal times were curve-specific.

2.2. Data Preprocessing

2.2.1. Temporal Resampling by PCHIP Interpolation

The original experimental data consisted of sparsely sampled time points (21–31 per curve). A piecewise cubic Hermite interpolating polynomial (PCHIP) was used as a shape-preserving temporal resampling procedure to place all breakthrough curves on a common 64-point temporal grid while preserving the monotonicity and shape characteristics of the original profiles [36,37]. The original experimental points were retained, and interpolation was performed independently for each curve. This procedure was used only to standardize sequence length across curves for neural-network training and comparison; it did not increase the number of independent experimental breakthrough curves.

2.2.2. Data Stratification by Mahalanobis Distance

To characterize the relative location of each breakthrough curve within the multidimensional operational space, the 14 curves were classified using D_M [38] calculated from the operating variables pH, C_0 , Q , and m :

$$D_M(x) = \left[(x - \mu)^T \Sigma^{-1} (x - \mu) \right]^{1/2} \quad (1)$$

where x is the vector of operational conditions for each curve, μ is the centroid of all experimental conditions, and Σ is the covariance matrix. Operation time was not included in the D_M calculation because D_M was intended to describe the position of each experimental condition, not the temporal progression within each curve. Bed height was also excluded from the D_M calculation because it is directly related to biosorbent mass in the experimental configuration; therefore, including both variables would introduce redundant geometric information into the distance calculation.

Four independent training–evaluation scenarios were defined based on the relative position of selected curves within the operational space: Scenario A, representing the most central condition; Scenarios B1 and B2, representing intermediate conditions; and Scenario C, representing the most peripheral condition. In each scenario, one complete breakthrough curve was held out, and the remaining 13 curves were used for model development. The holdout curve was excluded before hyperparameter optimization, cross-validation, scaler fitting, and final model training.

2.2.3. Feature Normalization

For each holdout scenario, input features were normalized using a StandardScaler fitted exclusively on the corresponding training set and then applied to the validation and holdout data. This prevented information from the external holdout curve from influencing preprocessing.

2.3. Neural Network Architectures

2.3.1. Feedforward MLP Architecture

The MLP is a neural network architecture widely used for regression tasks due to its ability to learn complex nonlinear mappings between input variables and continuous outputs directly from data [39]. In this study, the MLP was implemented for point-wise breakthrough-curve prediction, in which each time point is treated as an independent observation. The model comprised a three-layer network arranged in a pyramidal structure, with information flowing unidirectionally from input to output. Rectified Linear Unit (ReLU) activation functions were applied in each hidden layer. To mitigate overfitting, two complementary regularization strategies were employed: dropout, which randomly deactivates a fraction of neurons during training, and L2 weight decay, which penalizes large parameter values [40]. The output layer used a sigmoid activation function, producing two continuous values constrained to the interval $[0, 1]$, corresponding to the physical range of the normalized effluent concentration C_t/C_0 for Cr(VI) and total Cr.

2.3.2. Recurrent LSTM and Bi-LSTM Architectures

The LSTM and Bi-LSTM models were implemented as recurrent neural network architectures designed to process breakthrough curves as ordered sequences. Unlike the MLP, which treats each time point as an independent observation, recurrent models update an internal hidden state along the sequence, allowing information from previous time steps to contribute to the prediction at a given time. This structure is particularly suitable for breakthrough-curve modeling, where the effluent concentration at each time reflects the preceding loading history and progressive saturation of the fixed bed.

The LSTM network processes the sequence in the forward temporal direction. To overcome the vanishing-gradient problem that limits conventional recurrent networks in capturing long-range dependencies, the LSTM employs a gating mechanism—comprising input, forget, and output gates—that enables the network to selectively retain or discard information over long sequences [29]. The model implemented in this study consisted of two stacked unidirectional LSTM layers arranged in a pyramidal structure.

The Bi-LSTM model extends the unidirectional LSTM by processing each sequence simultaneously in both temporal directions—past to future and future to past [30]. Unlike the unidirectional LSTM, which accumulates information only from previous observations, the Bi-LSTM incorporates both preceding and subsequent temporal context at each time step through the concatenation of forward and backward hidden states. This bidirectional processing has been shown to improve predictive performance over unidirectional architectures in time-series regression tasks [41]. The model implemented in this study consisted of two stacked bidirectional LSTM layers arranged in a pyramidal structure.

For both recurrent architectures, recurrent dropout was applied to the hidden state between consecutive time steps, standard dropout was applied between layers, and L2 weight decay was incorporated to mitigate overfitting. The output layer used a sigmoid activation function, constraining predictions to the interval $[0, 1]$, corresponding to the physical range of the normalized effluent concentration C_t/C_0 . This constraint was introduced to prevent nonphysical negative predictions, particularly at early time steps, where recurrent hidden states have limited prior sequence information.

2.4. Hyperparameter Optimization

Hyperparameter optimization was performed using Bayesian search with Optuna and the Tree-structured Parzen Estimator (TPE) algorithm. The TPE sampler was initialized with a fixed seed of 42, including the initial random exploration trials. To ensure comparability across architectures, a deterministic 80/20 validation scheme was implemented at the

time-point level within each curve: validation points were selected by uniformly spaced sampling along the temporally ordered breakthrough profile, and the remaining points were used for trial-level training. This validation split ensured representation of the initial, breakthrough, and saturation regions and was used only during hyperparameter optimization to monitor point-wise prediction error. External generalization was assessed separately using complete holdout curves, as described in Section 2.5.2.

For recurrent networks, batch size operates at the sequence level, meaning that each batch contains complete breakthrough curves rather than individual time points. Table 2 summarizes the hyperparameter search space for each architecture.

Table 2. Hyperparameter search space for feedforward and recurrent network architectures..

Category	Hyperparameter	MLP	LSTM/Bi-LSTM
Architecture	Number of layers	3	2
	Layer 1 size (n_1, h_1)	16–128, step 16	16–128, step 16
	Layer 2 size (n_2, h_2)	16– n_1 , step 16	16– h_1 , step 16
	Layer 3 size (n_3)	16– n_2 , step 16	—
Regularization	Standard dropout rate	0.0–0.5, step 0.01	0.0–0.2, step 0.01
	Recurrent dropout rate	—	0.0–0.5, step 0.01
	L2 weight decay	10^{-6} – 10^{-2}	10^{-6} – 10^{-3}
Monotonicity constraint	λ_{mono}	—	0.8–2.0
Training	Batch size	16, 32, 64	1, 2

MLP, multilayer perceptron; LSTM, long short-term memory; Bi-LSTM, bidirectional long short-term memory.

Each optimization run comprised 100 trials, including 10 initial random exploration trials before activation of the Bayesian algorithm. Models were trained for 50 epochs per trial with a fixed learning rate of 10^{-3} using the Adam optimizer. The objective function minimized the mean squared error (MSE) on the validation set:

$$\text{MSE} = \frac{1}{N_v \cdot n_o} \sum_{i=1}^{N_v} \sum_{j=1}^{n_o} (y_{i,j} - \hat{y}_{i,j})^2 \quad (2)$$

where $y_{i,j}$ and $\hat{y}_{i,j}$ are the observed and predicted values for output j at sample i , N_v is the number of validation samples, and $n_o = 2$ is the number of output variables (C_t/C_0 for Cr(VI) and total Cr).

The optimization progress was inspected using the trial-wise validation MSE and the best-so-far validation MSE curves. In all cases, the best validation loss stabilized before completion of the 100 trials, and no systematic improvement was observed during the final trials; therefore, the selected hyperparameters were considered adequate for scenario-specific model comparison.

Monotonicity Constraint for Recurrent Models

For the LSTM and Bi-LSTM architectures, a monotonicity penalty was incorporated into the loss function to discourage decreases in the predicted breakthrough profiles and to promote the physically expected non-decreasing behavior of C_t/C_0 . This penalty complements the sigmoid output activation: while the sigmoid constrains predictions to the physical interval $[0, 1]$, the monotonicity term penalizes local violations of the expected temporal trend. The total loss function is defined as:

$$\mathcal{L} = \text{MSE} + \lambda_{\text{mono}} \cdot P_{\text{mono}} \quad (3)$$

where λ_{mono} is the penalty weight (optimized as a hyperparameter; Table 2) and P_{mono} penalizes violations of monotonicity:

$$P_{\text{mono}} = \frac{1}{N_s} \sum_i \sum_t \max(0, -(\hat{y}_{i,t+1} - \hat{y}_{i,t})) \quad (4)$$

where N_s is the number of sequences, the index i runs over sequences, t runs over consecutive time steps within each sequence, and $\hat{y}_{i,t}$ is the predicted C_t/C_0 for sequence i at time step t . This constraint was not applied to the MLP model, which processes each time point independently and therefore lacks the sequential structure required to enforce temporal monotonicity.

2.5. Model Development and Evaluation Pipeline

The model development and evaluation pipeline was executed independently for each of the four holdout scenarios (Scenarios A, B1, B2 and C) and for each of the three architectures (MLP, LSTM, and Bi-LSTM), yielding twelve independently optimized models. In each scenario, the holdout curve was excluded before hyperparameter optimization, cross-validation, scaler fitting, and final model training. Thus, holdout evaluation was performed at the complete-curve level rather than at the interpolated-point level. Figure 1 summarizes the pipeline, which comprises cross-validation with early stopping, final model training, and evaluation on the holdout curve.

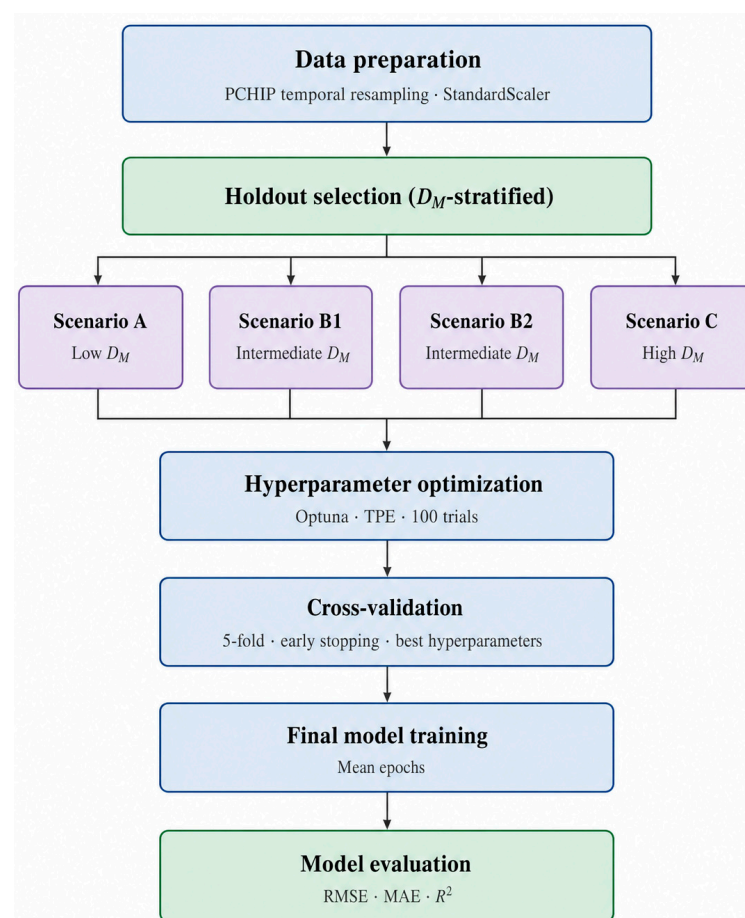


Figure 1. Model development and evaluation pipeline executed independently for each of the three architectures (MLP, LSTM, Bi-LSTM). PCHIP, piecewise cubic Hermite interpolating polynomial; D_M , Mahalanobis distance; TPE, Tree-structured Parzen Estimator.

2.5.1. Cross-Validation with Early Stopping

Five-fold cross-validation (CV5) was performed at the time-point level within the training set of each holdout scenario, using the optimal hyperparameters identified in Section 2.4. For each fold, 20% of the points from every training curve were assigned to validation and the remaining 80% to training. Point indices within each training curve were randomly permuted using a fixed seed of 42 and then assigned to the five folds in a round-robin manner. The same fold-construction rule was applied to all architectures within each holdout scenario, ensuring reproducible internal validation folds with validation points distributed along each breakthrough profile.

This internal CV5 procedure was used for early stopping, epoch selection, and assessment of training stability, rather than as an independent test of generalization to new experimental curves.

Models were trained with early stopping based on validation loss, for a maximum of 500 epochs, with a patience of 25 epochs for the LSTM and Bi-LSTM models or 30 epochs for the MLP. If the validation loss did not improve by at least 10^{-6} within the specified patience interval, training was terminated and the model instance with the lowest validation loss was retained. Model stability across folds was assessed using the root mean square error (RMSE) and CoV:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

$$\text{CoV} = \frac{\sigma}{\bar{x}} \times 100 \quad (6)$$

where y_i and $\hat{y}_i$ are the observed and predicted values, respectively; N is the number of samples; and σ and $\bar{x}$ are the standard deviation and mean of the variable of interest, respectively. In this context, CoV was calculated from the RMSE values across the five folds to quantify inter-fold variability.

2.5.2. Final Model Training

After CV5, the mean number of optimal epochs across the five folds was calculated and used as the training duration for the final model. This model was then trained on the entire training dataset, without an additional validation split, using the optimized hyperparameters. The external holdout curve remained excluded from scaler fitting, hyperparameter optimization, CV5, and final model training.

2.5.3. Model Evaluation

Performance metrics were calculated independently for each output variable (C_t/C_0) for Cr(VI) and total Cr, including RMSE (Equation (5)), Mean Absolute Error (MAE), and coefficient of determination (R^2):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (8)$$

where $\bar{y}$ is the mean of the observed values.

2.6. Model Consensus on Biosorption Chromium Capacity

2.6.1. Biosorption Capacity Calculation

Biosorption capacity was evaluated based on the predicted C_t/C_0 curves for total Cr, which represent the overall removal achieved by the biosorbent under a given scenario. For each set of operational conditions, the biosorption capacity q (mg g^{-1}) was calculated by integrating the predicted breakthrough curve:

$$q = \frac{Q \cdot C_0 \cdot 60}{m \cdot 1000} \int_0^{t_f} \left(1 - \frac{C_t}{C_0}\right) dt \quad (9)$$

where Q is the flow rate (mL min^{-1}), C_0 is the influent concentration (mg L^{-1}), m is the biosorbent mass (g), and t_f is the evaluation time (h). The factor 60 converts hours to minutes to match the flow-rate units, and 1000 converts milliliters to liters. The integral was computed numerically using the trapezoidal rule. Capacity was evaluated at $t_f = 144$ h (extended operation approaching saturation).

2.6.2. Model Consensus Analysis

Response surfaces were generated by systematically varying two operational variables while holding the remaining variables at reference values ($\text{pH} = 2$, $C_0 = 200 \text{ mg L}^{-1}$, $Q = 0.5 \text{ mL min}^{-1}$, $m = 2 \text{ g}$). Bed height H was derived from m by direct proportionality. A 100×100 grid was constructed for each variable pair.

Inter-model consensus was quantified using the CoV (Equation (6)), applied here to the predicted capacity q across the three architectures, where σ and $\bar{x}$ represent the standard deviation and mean of q at each grid point. To assess the distance of each grid point from the training distribution, D_M (Equation (1)) was calculated in the same four-variable operational space used for holdout stratification (pH , C_0 , Q , and m), using the covariance matrix of the 13 training curves. D_M contours were overlaid on the CoV response surfaces to relate inter-model agreement to the relative location of each simulated condition within the training domain.

2.7. Feature Importance Analysis

2.7.1. SHAP Analysis

SHAP analysis was conducted to quantify the contribution of each input feature to model predictions and to determine its directional effect [42]. The resulting SHAP values were interpreted as model-attribution measures rather than direct mechanistic parameters. The KernelExplainer was employed as a model-agnostic method, enabling consistent comparison across architectures. For the MLP model, SHAP values were computed for all operational features (pH , C_0 , Q , m , H , and t), using a background dataset of 100 randomly sampled training points and 200 synthetic conditions generated via Latin Hypercube Sampling (LHS).

For the LSTM and Bi-LSTM models, the analysis was restricted to the five operational features (pH , C_0 , Q , m , and H), excluding t because recurrent architectures represent temporal progression through the ordered sequence rather than treating time as an independent point-wise predictor. The background dataset comprised the first time step of each of the 13 training curves. For each synthetic condition, a complete 64-step sequence was constructed using the time vector of the first training curve as a fixed template, and the prediction function returned the mean predicted C_t/C_0 across all time steps.

2.7.2. Time-Period SHAP Analysis

To investigate how feature influence evolves throughout the breakthrough process, SHAP analysis was performed separately for five time intervals: 2–12, 12–24, 24–48, 48–72,

and 72–144 h. These intervals were used to summarize model attributions during successive stages of the breakthrough curve, from early operation to extended operation approaching saturation.

For the MLP model, synthetic conditions were generated within each interval using LHS, with the time feature constrained to the corresponding range. Background points were drawn from training data within the same interval.

For both LSTM and Bi-LSTM models, the complete 64-point sequence was provided to the model to preserve recurrent-state continuity. The prediction function then returned the mean prediction only for the time steps belonging to the interval being analyzed. Background and synthetic conditions followed the same strategy described in Section 2.7.1.

2.8. Computational Framework

All data processing, model development, and analysis were performed in PyCharm 2024.3.1.1 (JetBrains s.r.o., Prague, Czech Republic) using Python 3.10 (Python Software Foundation, Beaverton, OR, USA). Neural network models were implemented in PyTorch 2.9.1 (PyTorch Foundation, Linux Foundation, San Francisco, CA, USA), hyperparameter optimization was conducted using Optuna 4.4.0 (Preferred Networks, Inc., Tokyo, Japan), and interpretability analysis was performed using SHAP 0.46.0 (Microsoft Research, Redmond, WA, USA). Code development was assisted by ClaudeOpus 4.5 and Opus 4.6 (Anthropic, PBC, San Francisco, CA, USA) for selected scripting tasks; all code was reviewed, tested, and validated by the authors prior to use.

3. Results and Discussion

3.1. Data Stratification by Mahalanobis Distance and Holdout Selection

The D_M was calculated for each of the 14 experimental breakthrough curves relative to the centroid of the operational-variable space defined by pH, C_0 , Q , and m (Table 3). D_M values ranged from 0.317 (K03) to 2.802 (K08 and K09), reflecting substantial variability in the distribution of experimental conditions across the input space.

Table 3. Mahalanobis distance-based stratification and holdout assignment of experimental breakthrough curves.

Curve ID	pH	C_0 (mg L ^{−1})	Q (mL min ^{−1})	m (g)	D_M
K01	1.0	200	0.75	2	2.302
K02	1.5	200	0.75	2	1.183
K03	2.0	200	0.75	2	0.317
K04	2.5	200	0.75	2	1.183
K05	3.0	200	0.75	2	2.302
K06	2.0	200	0.75	1	1.751
K07	2.0	200	0.75	3	1.302
K08	2.0	200	0.75	4	2.802
K09	2.0	200	0.25	2	2.802
K10	2.0	200	0.50	2	1.302
K11	2.0	200	1.00	2	1.751
K12	2.0	50	0.75	2	1.983
K13	2.0	100	0.75	2	1.325
K14	2.0	400	0.75	2	2.752

As D_M accounts for the covariance structure among variables, it captures not only whether individual variables assume extreme values but also whether a given combination of conditions is well represented in the dataset. A curve with a low D_M is closer to the centroid of the experimental design and is therefore expected to represent a more central

operating condition. In contrast, a curve with a high D_M is more distant from the centroid and may represent a less covered combination of operating variables. Accordingly, the selected holdout scenarios provide a practical way to examine model behavior from central to more peripheral regions of the available experimental space.

Curve K03 (pH 2.0, $C_0 = 200 \text{ mg L}^{-1}$, $Q = 0.75 \text{ mL min}^{-1}$, $m = 2 \text{ g}$) exhibited the lowest D_M value, indicating that its operational conditions most closely approximate the centroid of the experimental design and are therefore the most representative of the dataset. Curves K02 and K04 share the second and third positions ($D_M = 1.183$), representing symmetric variants of K03 along the pH dimension (pH 1.5 and 2.5, respectively). Most curves cluster within the low-to-intermediate distance range ($D_M < 2.1$).

Curves K01 and K05 have identical D_M values (2.302), as they are symmetric about the centroid along the pH axis (pH 1.0 and 3.0, respectively). The most peripheral curves were K08 ($D_M = 2.802$), operated at the maximum biosorbent mass ($m = 4 \text{ g}$), and K09 ($D_M = 2.802$), operated at the minimum flow rate ($Q = 0.25 \text{ mL min}^{-1}$).

Four external holdout scenarios were defined using D_M to evaluate model behavior across different regions of the experimental space. Scenario A uses K03 ($D_M = 0.317$), the most central curve; Scenario B1 uses K01 ($D_M = 2.302$), corresponding to the lowest pH tested; Scenario B2 uses K06 ($D_M = 1.751$), corresponding to the lowest biosorbent mass; and Scenario C uses K09 ($D_M = 2.802$), corresponding to the minimum flow rate and one of the most peripheral conditions. In each scenario, the remaining 13 curves constituted the training set, and the holdout curve was excluded entirely from both hyperparameter optimization and model training.

3.2. Hyperparameter Optimization Outcomes

Bayesian hyperparameter optimization was performed independently for each architecture and holdout scenario, yielding twelve separately optimized models. The selected hyperparameters varied across scenarios, indicating that the optimal model configuration depended on both the architecture and the composition of the training set (Tables S1–S3).

For the MLP, the selected architectures showed scenario-dependent changes in layer size and regularization (Table S1). Scenarios B1 and B2 selected the largest network configuration within the final optimization results (112–96–80), whereas Scenario C selected the smallest configuration (80–64–48). Dropout was generally low, except for the third hidden layer in Scenario A, and the batch size remained constant at 16 across all scenarios. The best validation MSE values were in the order of 10^{-4} .

For the LSTM, the selected configurations generally favored high recurrent-layer capacity (Table S2). Scenarios A, B1, and C selected 128 cells in both recurrent layers, whereas Scenario B2 selected a smaller 112–96 configuration. The monotonicity penalty varied across scenarios, with the highest value selected for Scenario B2. Recurrent dropout also varied substantially, indicating scenario-dependent regularization requirements. The best validation MSE values were in the order of 10^{-3} .

For the Bi-LSTM, the selected recurrent-layer sizes were smaller than those of the unidirectional LSTM across all scenarios (Table S3). The largest Bi-LSTM configuration was selected in Scenario B1, whereas Scenario C selected the smallest one. Standard dropout was applied mainly in the first layer, while recurrent dropout varied among scenarios. The monotonicity penalty showed a narrower range than in the LSTM, and the best validation MSE values were in the order of 10^{-4} .

Figure 2 summarizes the main hyperparameter patterns across architectures and holdout scenarios. Overall, the optimization results show that no single architecture or hyperparameter configuration was uniformly optimal across all scenarios. This supports the use of

independent optimization within each holdout scenario and provides the basis for comparing feedforward and recurrent architectures under equivalent model-selection procedures.

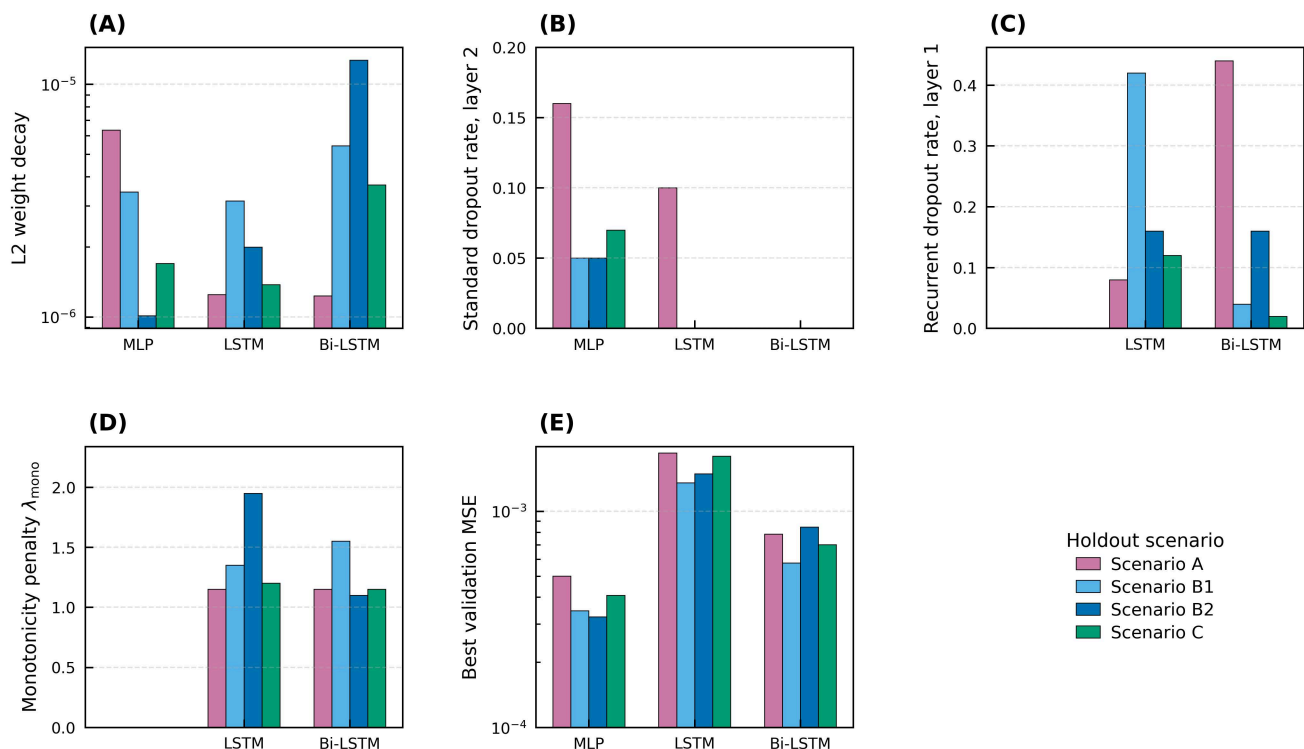


Figure 2. Optimized regularization and selected training hyperparameters across architectures and holdout scenarios: (A) L2 weight decay, (B) standard dropout rate, (C) recurrent dropout rate, (D) monotonicity penalty λ_{mono} , and (E) best validation MSE from Optuna optimization.

3.3. Cross-Validation Performance

Five-fold cross-validation results are summarized in Table 4. Across all architectures and scenarios, validation RMSE ranged from 0.0149 to 0.0309 and R^2 from 0.9879 to 0.9979, indicating that all models accurately reproduced the overall shape of the breakthrough curves within the scenario-specific training sets. At the global level, the MLP achieved the lowest validation RMSE values across scenarios (0.0149–0.0173), followed by the Bi-LSTM (0.0196–0.0212) and the LSTM (0.0229–0.0296). The Bi-LSTM showed a narrow RMSE range across the four scenarios, whereas the LSTM consistently exhibited the highest internal validation errors.

When metrics are disaggregated by output variable, both the MLP and Bi-LSTM consistently showed slightly higher RMSE for Cr(VI) than for total Cr across all scenarios, likely reflecting the additional complexity associated with the simultaneous reduction of Cr(VI) to Cr(III) within the column [35]. This pattern was less consistent for the LSTM. In Scenario A, the LSTM exhibited higher error for total Cr than for Cr(VI) (RMSE = 0.0309 vs. 0.0281; R^2 = 0.9879 vs. 0.9920), indicating that the relative difficulty of predicting each output depended on both the architecture and the training-set composition.

The holdout scenario also affected inter-fold stability. For the MLP, global CoV values ranged from 7.41% to 16.7%, with the lowest variability in Scenario C. For the Bi-LSTM, global CoV values ranged from 6.6% to 16.4%, with the most stable result in Scenario B1. The LSTM showed the highest inter-fold variability overall, with global CoV values of 16.5–25.0% and the poorest stability in Scenario B2. These results indicate that internal validation stability was influenced by both architecture and the composition of the scenario-specific training set.

Table 4. Five-fold cross-validation results for the MLP, LSTM, and Bi-LSTM across holdout scenarios.

Architecture	Scenario	Train Epochs	Output	RMSE	R ²	CoV RMSE (%)	CoV R ² (%)
MLP	A	79	Cr(VI)	0.0149	0.9979	10.3	0.05
			Total Cr	0.0149	0.9973	13.1	0.07
			Global	0.0149	0.9977	11.1	0.05
	B1	66	Cr(VI)	0.0169	0.9972	18.7	0.11
			Total Cr	0.0155	0.9971	14.8	0.10
			Global	0.0162	0.9972	16.7	0.10
	B2	141	Cr(VI)	0.0180	0.9968	14.6	0.10
			Total Cr	0.0166	0.9966	19.2	0.14
			Global	0.0173	0.9968	16.7	0.11
	C	132	Cr(VI)	0.0161	0.9973	7.82	0.04
			Total Cr	0.0150	0.9970	8.72	0.06
			Global	0.0156	0.9972	7.41	0.05
LSTM	A	106	Cr(VI)	0.0281	0.9920	21.0	0.43
			Total Cr	0.0309	0.9879	15.7	0.50
			Global	0.0296	0.9904	17.5	0.43
	B1	128	Cr(VI)	0.0241	0.9943	19.7	0.50
			Total Cr	0.0215	0.9943	18.1	0.43
			Global	0.0229	0.9944	18.2	0.25
	B2	121	Cr(VI)	0.0241	0.9939	28.8	0.26
			Total Cr	0.0245	0.9921	23.4	0.24
			Global	0.0243	0.9933	25.0	0.41
	C	122	Cr(VI)	0.0263	0.9925	16.7	0.26
			Total Cr	0.0272	0.9898	18.8	0.42
			Global	0.0268	0.9915	16.5	0.30
Bi-LSTM	A	131	Cr(VI)	0.0216	0.9954	13.7	0.44
			Total Cr	0.0198	0.9951	16.7	0.40
			Global	0.0207	0.9954	14.9	0.15
	B1	104	Cr(VI)	0.0205	0.9960	7.7	0.16
			Total Cr	0.0186	0.9959	7.5	0.08
			Global	0.0196	0.9960	6.6	0.07
	B2	126	Cr(VI)	0.0217	0.9954	14.1	0.15
			Total Cr	0.0208	0.9946	15.0	0.19
			Global	0.0212	0.9951	14.4	0.16
	C	132	Cr(VI)	0.0207	0.9954	13.9	0.16
			Total Cr	0.0188	0.9951	20.1	0.22
			Global	0.0198	0.9954	16.4	0.18

Train epochs, RMSE, and R² are reported as the mean across cross-validation folds. CoV, coefficient of variation.

Overall, the CV5 results show that all three architectures achieved high internal accuracy after hyperparameter optimization. The MLP yielded the lowest point-wise validation errors, whereas the Bi-LSTM provided the most consistent recurrent-model performance. Because CV5 was performed at the time-point level within the training curves, these results are interpreted as evidence of internal model stability rather than external complete-curve generalization. The latter is evaluated separately in Section 3.4 using the holdout curves excluded from model development.

3.4. External Holdout Evaluation of Final Models

The complete-curve predictive performance of the three neural network architectures was evaluated using an external holdout curve excluded from model development in each scenario. Table 5 summarizes the RMSE, MAE, and R² values obtained for both the training and holdout sets for the two breakthrough responses, C_t/C_0 of Cr(VI) and C_t/C_0 of total Cr.

Table 5. Training and holdout performance metrics of the MLP, LSTM, and Bi-LSTM final models across the four experimental scenarios for the Cr(VI) and total chromium C_t/C_0 outputs.

Scenario	Architecture	Dataset	Cr(VI)			Total Chromium		
			RMSE	MAE	R ²	RMSE	MAE	R ²
A	MLP	Train	0.0244	0.0189	0.9949	0.0233	0.0183	0.9943
		Holdout	0.0729	0.0491	0.9537	0.0658	0.0543	0.9541
	LSTM	Train	0.0365	0.0280	0.9885	0.0400	0.0300	0.9831
		Holdout	0.0417	0.0329	0.9848	0.0568	0.0494	0.9659
	Bi-LSTM	Train	0.0237	0.0171	0.9952	0.0214	0.0160	0.9952
		Holdout	0.0345	0.0301	0.9896	0.0337	0.0264	0.9880
B1	MLP	Train	0.0176	0.0138	0.9974	0.0159	0.0122	0.9974
		Holdout	0.1505	0.1195	0.7277	0.0527	0.0414	0.9521
	LSTM	Train	0.0316	0.0246	0.9915	0.0320	0.0234	0.9894
		Holdout	0.0505	0.0431	0.9693	0.1139	0.0894	0.7767
	Bi-LSTM	Train	0.0254	0.0192	0.9945	0.0218	0.0163	0.9951
		Holdout	0.0870	0.0805	0.9090	0.0975	0.0709	0.8363
B2	MLP	Train	0.0154	0.0120	0.9980	0.0166	0.0133	0.9971
		Holdout	0.1239	0.0798	0.7635	0.1205	0.0739	0.7522
	LSTM	Train	0.0258	0.0201	0.9942	0.0283	0.0183	0.9915
		Holdout	0.1062	0.0837	0.8262	0.1262	0.0969	0.7282
	Bi-LSTM	Train	0.0223	0.0179	0.9957	0.0203	0.0159	0.9956
		Holdout	0.1137	0.0874	0.8011	0.1423	0.0920	0.6543
C	MLP	Train	0.0183	0.0140	0.9970	0.0190	0.0143	0.9960
		Holdout	0.1437	0.1128	0.3251	0.0837	0.0676	0.8266
	LSTM	Train	0.0265	0.0208	0.9937	0.0295	0.0207	0.9904
		Holdout	0.0790	0.0514	0.7960	0.0390	0.0322	0.9624
	Bi-LSTM	Train	0.0236	0.0187	0.9950	0.0213	0.0159	0.9950
		Holdout	0.1068	0.0876	0.6274	0.0691	0.0589	0.8818

All three architectures achieved good fits on the training set across all scenarios, with R² values consistently above 0.98 and RMSE below 0.041 for both outputs, confirming that each model effectively reproduced the sigmoidal dynamics of the breakthrough curves during training. Holdout performance, however, varied with the relative position of the excluded curve within the experimental space, as described by D_M (Figure 3).

For the representative holdout (Scenario A), training and holdout metrics were of comparable magnitude across all architectures. The Bi-LSTM achieved the highest holdout performance, with R² = 0.9896 and 0.9880 for C_t/C_0 of Cr(VI) and total chromium, respectively. The LSTM also maintained high agreement for both outputs, while the MLP showed slightly larger deviations but still acceptable holdout performance. As shown in Figure 3A,E, all three models accurately reproduced the initial rise, the inflection region, and the saturation plateau of the experimental curve over the full 96 h observation period, without appreciable shape distortion. This behavior is consistent with prediction near the centroid of the experimental design space.

Performance decreased noticeably in Scenario B1, where the holdout corresponded to the lowest pH tested (pH 1.0), representing an acidic boundary condition of the experimental space. The effect was output-dependent: for C_t/C_0 of Cr(VI), the LSTM showed the highest holdout R² (0.9693), followed by the Bi-LSTM (0.9090), whereas the MLP showed the largest deviation (R² = 0.7277). Figure 3B shows that the MLP underestimated the late-stage Cr(VI) profile, while the recurrent models better followed the overall experimental trend. In contrast, for total chromium, the MLP achieved the highest R² (0.9521), whereas the recurrent models showed larger deviations, mainly because they overestimated the early increase in C_t/C_0 (Figure 3F). These differences are consistent with the limited repre-

sensation of pH 1.0 in the training set, a condition under which Cr(VI) reduction to Cr(III) is more pronounced and modifies the relative behavior of the two outputs, as previously reported for this biosorbent system by Aranda-García et al. [34] and Aranda-García and Cristiani-Urbina [35].

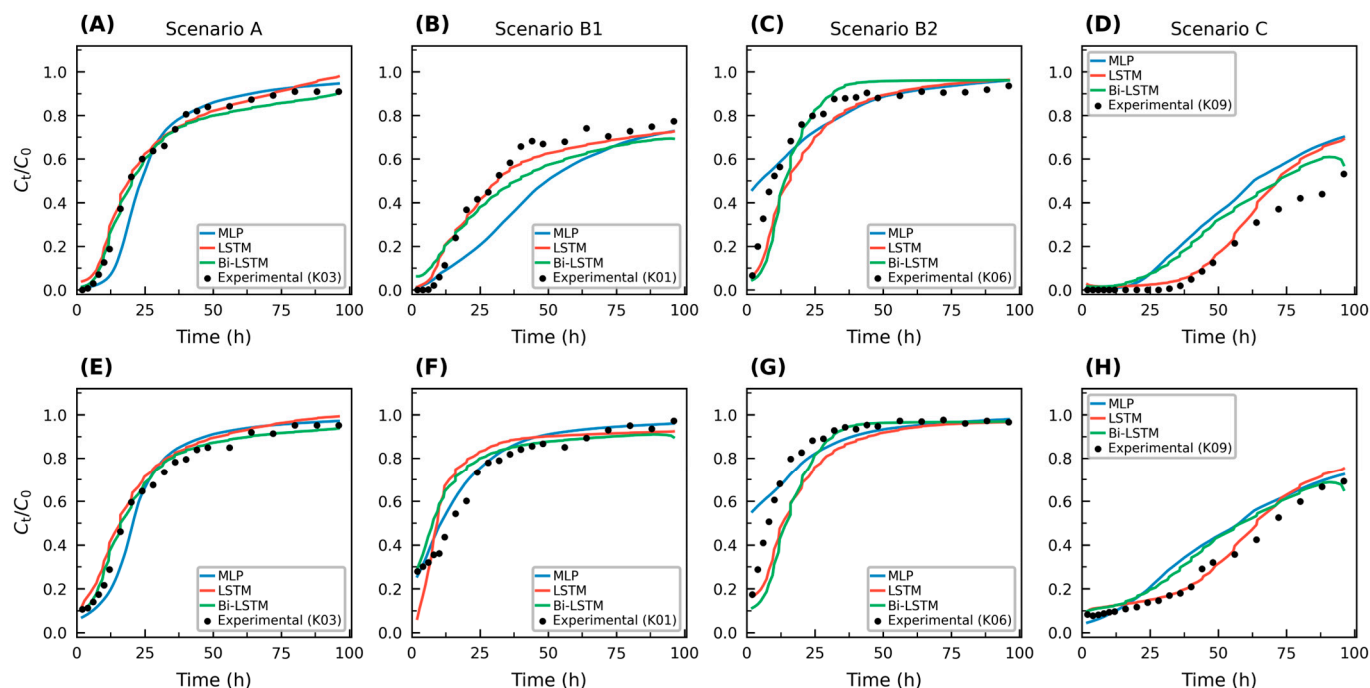


Figure 3. Predicted versus experimental C_t/C_0 breakthrough curves for Cr(VI) (A–D) and total chromium (E–H) under holdout Scenarios A (K03, $D_M = 0.317$; pH = 2.0, $C_0 = 200 \text{ mg L}^{-1}$, $Q = 0.75 \text{ mL min}^{-1}$, $m = 2 \text{ g}$, $H = 3.5 \text{ cm}$), B1 (K01, $D_M = 2.302$; pH = 1.0, $C_0 = 200 \text{ mg L}^{-1}$, $Q = 0.75 \text{ mL min}^{-1}$, $m = 2 \text{ g}$, $H = 3.5 \text{ cm}$), B2 (K06, $D_M = 1.751$; pH = 2.0, $C_0 = 200 \text{ mg L}^{-1}$, $Q = 0.75 \text{ mL min}^{-1}$, $m = 1 \text{ g}$, $H = 1.7 \text{ cm}$), and C (K09, $D_M = 2.802$; pH = 2.0, $C_0 = 200 \text{ mg L}^{-1}$, $Q = 0.25 \text{ mL min}^{-1}$, $m = 2 \text{ g}$, $H = 3.5 \text{ cm}$).

Scenario B2 corresponded to the lowest biosorbent mass ($m = 1 \text{ g}$). Under this condition, the three architectures showed reduced holdout performance for both outputs. For Cr(VI), R^2 values ranged from 0.7635 to 0.8262, whereas for total chromium they ranged from 0.6543 to 0.7522. Figure 3C,G shows that all models captured the rapid approach to saturation, but they differed in the early breakthrough region. The MLP predicted a higher initial C_t/C_0 , whereas the recurrent models delayed the early rise and then approached the experimental plateau. These results indicate that the low-mass condition was difficult to predict despite its intermediate D_M value, suggesting that local data support and process sensitivity also influenced complete-curve generalization.

Scenario C corresponded to K09, operated at the minimum flow rate ($Q = 0.25 \text{ mL min}^{-1}$) and represented one of the most peripheral conditions according to D_M . Under this condition, prediction of Cr(VI) was markedly affected, especially for the MLP ($R^2 = 0.3251$), while the LSTM and Bi-LSTM achieved higher but still reduced agreement ($R^2 = 0.7960$ and 0.6274 , respectively). As shown in Figure 3D, all models anticipated the onset of Cr(VI) breakthrough and overestimated C_t/C_0 during the late stage, with the largest deviation observed for the MLP. In contrast, total chromium was predicted more accurately, particularly by the LSTM ($R^2 = 0.9624$). Figure 3H shows that the models reproduced the gradual increase in total chromium C_t/C_0 more closely than the corresponding Cr(VI) profile. Thus, the peripheral low-flow condition affected Cr(VI) prediction more strongly than total chromium prediction.

The comparison of experimental and predicted total chromium capacities (Table 6) provides a complementary evaluation of the holdout predictions. In Scenario A, the three neural-network models produced integrated capacities close to the experimental value. In Scenario B1, the models showed larger deviations but remained closer to the experimental capacity than the fitted Thomas and Yan estimates. In Scenario B2, the MLP remained close to the experimental capacity, whereas the LSTM and Bi-LSTM substantially overestimated q . In Scenario C, the three neural-network estimates were close to the experimental capacity, which is consistent with the comparatively better holdout metrics obtained for the total chromium C_t/C_0 curves under this condition.

Table 6. Comparison of experimental, multivariate modeling, and kinetic modeling total chromium biosorption capacities for holdout scenarios.

Scenario	Model	q (mg g ⁻¹)	Evaluation Time (h)	Method	Reference
A	Experimental	107.86	96	Integrated	This study
	Thomas	98.20	—	Fitted parameter	[35]
	Yan	83.79	—	Fitted parameter	[35]
	MLP	103.69	96	Integrated	This study
	LSTM	94.91	96	Integrated	This study
	Bi-LSTM	105.04	96	Integrated	This study
B1	Experimental	91.38	96	Integrated	This study
	Thomas	71.47	—	Fitted parameter	[35]
	Yan	55.48	—	Fitted parameter	[35]
	MLP	76.36	96	Integrated	This study
	LSTM	83.45	96	Integrated	This study
	Bi-LSTM	79.08	96	Integrated	This study
B2	Experimental	100.00	96	Integrated	This study
	Thomas	76.98	—	Fitted parameter	[35]
	Yan	65.53	—	Fitted parameter	[35]
	MLP	105.40	96	Integrated	This study
	LSTM	176.28	96	Integrated	This study
	Bi-LSTM	151.36	96	Integrated	This study
C	Experimental	93.31	96	Integrated	This study
	Thomas	109.90	—	Fitted parameter	[35]
	Yan	97.81	—	Fitted parameter	[35]
	MLP	82.79	96	Integrated	This study
	LSTM	88.62	96	Integrated	This study
	Bi-LSTM	84.10	96	Integrated	This study

Experimental capacities were calculated by integration on the 64-point resampled time grid; ML-based capacities were integrated over the same interval.

Overall, the three architectures exhibited comparable performance under the central holdout condition, whereas differences became more pronounced under boundary, low-mass, and peripheral low-flow conditions. This pattern is consistent with the known tendency of neural networks to perform well within the training distribution but degrade under poorly represented or extrapolative conditions, as discussed by Chen et al. [31] and Xu et al. [32]. The largest deterioration was observed for the Cr(VI) output, particularly for the MLP under Scenario C, where the holdout R^2 decreased to 0.3251. This output-dependent behavior suggests that the point-wise feedforward model was less effective for the more complex Cr(VI) response, which involves both breakthrough dynamics and Cr(VI) reduction to Cr(III). The advantage observed for the recurrent models is consistent with their ability to process ordered sequences and retain temporal information, a feature widely recognized in time-series prediction and recently applied to adsorption–desorption

dynamics in activated carbon columns [43]. In this study, this sequence-level representation, together with monotonicity regularization, likely contributed to the more consistent Cr(VI) profiles obtained with the LSTM and Bi-LSTM models under the most challenging holdout conditions.

3.5. Biosorption Capacity and Inter-Model Consensus

For the response-surface analysis, chromium biosorption capacity was estimated by numerically integrating the predicted total chromium C_t/C_0 breakthrough curves over the 2–144 h time window at each point in the operational-condition grid. All three models correspond to Scenario A (holdout K03), ensuring that the consensus analysis reflects architectural differences under a common training set rather than differences arising from holdout partitioning. The CoV across the three architectures was used to quantify inter-model consensus: low CoV values indicate strong agreement among architectures, whereas high CoV values indicate increased model divergence and therefore lower confidence in architecture-independent predictions. The surfaces presented below are two-dimensional projections of the multidimensional design space, with the remaining variables fixed at their reference values; they therefore illustrate representative patterns along selected variable pairs.

The predicted mean capacity ($\bar{q}$) and CoV surfaces at $t_f = 144$ h are shown in Figure 4. In the C_0 – Q plane (Figure 4A,B), $\bar{q}$ increased with influent concentration and was favored at lower-to-intermediate flow rates. Within the region enclosed approximately by the $D_M = 2$ contour, CoV values were mostly low to moderate, generally below approximately 12%, indicating stronger agreement among architectures in the better-represented portion of the operating space. CoV increased toward peripheral combinations, particularly near high C_0 and high Q , where D_M approached or exceeded 3, indicating reduced consensus under more demanding loading and hydraulic conditions.

In the m – Q plane (Figure 4C,D), $\bar{q}$ increased with biosorbent mass and decreased with flow rate within the interpolation region. Low CoV values were observed mainly within the $D_M = 2$ contour, although some low-disagreement zones extended beyond this boundary. In contrast, CoV increased markedly at low biosorbent mass combined with high flow rate, reaching values above 20%. This localized divergence is consistent with the reduced holdout performance observed for Scenario B2 and indicates that model disagreement was controlled not only by geometric distance from the training centroid, but also by local data support in process-relevant regions.

In the C_0 –pH plane (Figure 4E,F), $\bar{q}$ increased with inlet concentration and was favored under acidic conditions. Within the central region defined approximately by $D_M < 2$, CoV values remained mostly below approximately 12%. In contrast, CoV increased toward high C_0 combined with pH values near the edges of the evaluated range, particularly in regions approaching $D_M = 3$. These zones indicate operating conditions where capacity predictions became more dependent on model architecture.

Across the three variable pairs, the combined interpretation of $\bar{q}$, CoV, and D_M shows that inter-model agreement was generally stronger in regions closer to the training distribution and decreased in selected peripheral or poorly supported regions. However, CoV and D_M did not coincide perfectly. Some conditions with moderate D_M exhibited elevated CoV, such as the low-mass region in the m – Q surface, whereas some more distant conditions showed only moderate architectural disagreement. This indicates that D_M and CoV capture related but distinct aspects of the prediction problem.

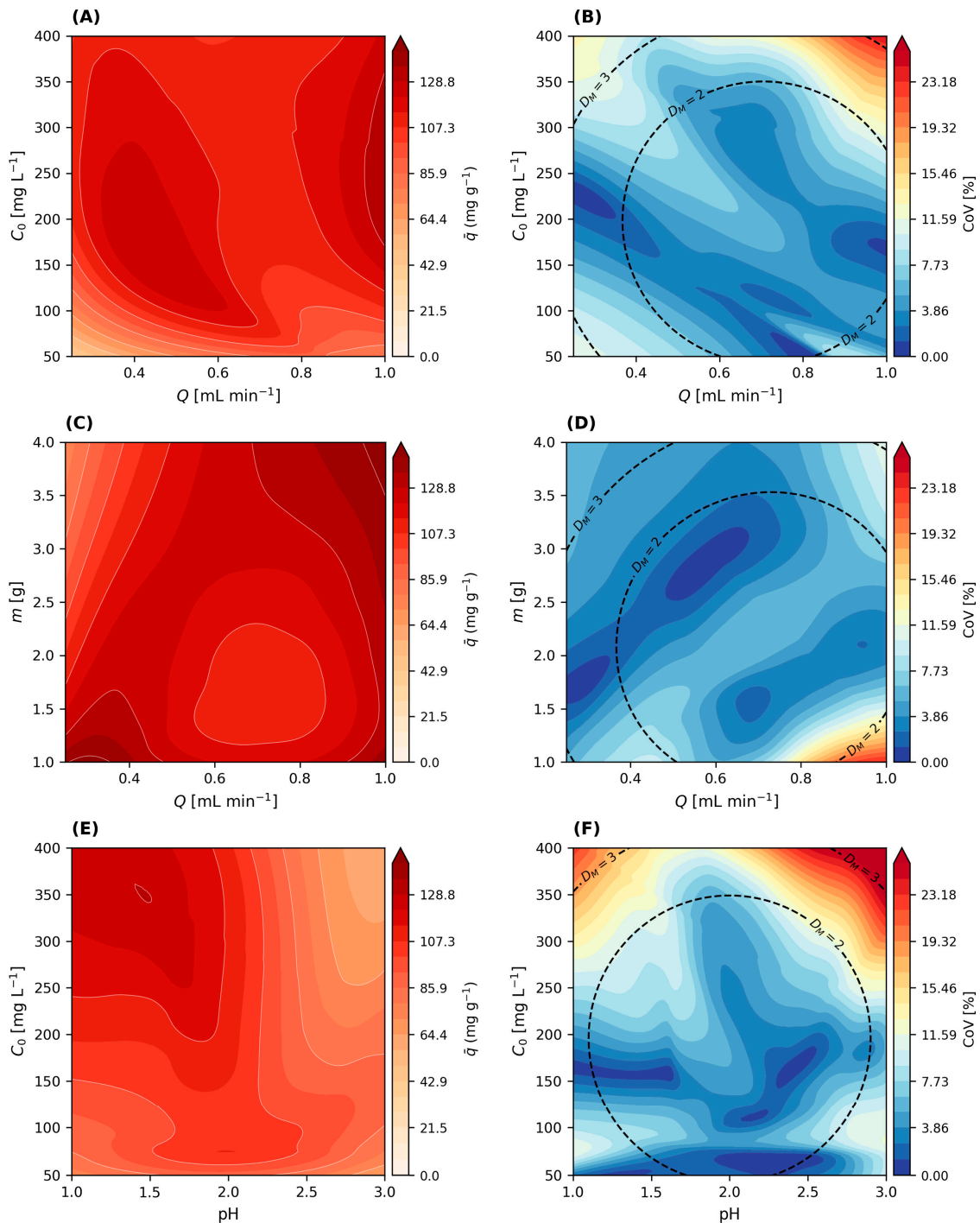


Figure 4. Mean predicted chromium biosorption capacity ($\bar{q}$) and inter-model coefficient of variation (CoV) for three operational variable pairs: C_0 – Q (A,B), m – Q (C,D), and C_0 – pH (E,F). $\bar{q}$ surfaces are shown in (A,C,E), and CoV in (B,D,F). Low CoV values indicate stronger agreement among the three architectures, whereas high CoV values indicate increased inter-model divergence. D_M contours indicate the relative distance of each simulated condition from the training distribution. All surfaces were generated using models from Scenario A.

D_M is a geometric descriptor of the input space that accounts for the covariance structure of the operational variables [44]. It is independent of model architecture and therefore indicates how far a simulated condition lies from the center of the training distribution. CoV, in contrast, measures the actual disagreement among the MLP, LSTM, and Bi-LSTM predictions at each grid point. Although the models used here do not

constitute a formal deep ensemble, using disagreement among independently trained predictors as an empirical indicator of predictive uncertainty is consistent with the rationale of ensemble-based uncertainty estimation, as discussed by Lakshminarayanan et al. [45]. In this study, the use of architecturally distinct models allowed CoV to reflect both data coverage and structural differences among feedforward and recurrent formulations.

In this dataset, $D_M \approx 2$ acted as an approximate empirical boundary between regions of generally low and increased inter-model divergence. However, this value should not be interpreted as a transferable threshold. Its meaning depends on the number and distribution of experimental conditions, the dimensionality of the operational space, and the covariance structure of the input variables.

For process analysis, the combined use of $\bar{q}$, CoV, and D_M provides a practical diagnostic framework for differentiating regions where the architectures give similar capacity predictions from regions where additional experimental information would be valuable. Within low-CoV regions, predicted capacity trends were consistent across the three neural-network formulations. In high-CoV regions, particularly at low biosorbent mass, the architectures produced more divergent capacity estimates, indicating that predictions in these zones are more sensitive to model structure and training-set coverage.

Inter-model CoV should therefore be interpreted as an empirical measure of model disagreement, not as a formally calibrated uncertainty interval. Nevertheless, disagreement among independently trained models can serve as a practical indicator of predictive uncertainty [45], and uncertainty-guided sampling is commonly used in active-learning workflows [46,47]. Future fixed-bed experiments could therefore focus on high-divergence regions, followed by model retraining to evaluate whether architecture-dependent predictions decrease.

3.6. SHAP-Based Feature Attribution Analysis

The global SHAP beeswarm plots (Figure 5) illustrate the direction and relative magnitude of each feature's contribution to the predicted C_t/C_0 for each architecture and output. Because SHAP values quantify feature contributions to model outputs, they were interpreted as model-attribution measures rather than direct mechanistic parameters [42]. To allow comparison across architectures, the analysis reported here focuses on the five operational variables common to all models: pH, C_0 , Q , m , and H .

Among the operational variables, influent concentration and flow rate exhibited the largest SHAP magnitudes across all architectures and both outputs. High C_0 values were associated with large positive SHAP contributions, indicating accelerated breakthrough due to increased chromium mass loading per unit time. Similarly, high Q values produced positive SHAP contributions, reflecting reduced residence time in the bed and a corresponding increase in C_t/C_0 at a given time. Conversely, low C_0 and low Q delayed breakthrough and extended bed service time.

Biosorbent mass and bed height generally showed negative SHAP contributions, consistent with delayed breakthrough at larger bed loadings and longer beds [35]. However, their relative importance varied among architectures and outputs. This behavior should be interpreted cautiously because H and m are correlated in the OFAT dataset, so part of their attribution may be redistributed differently depending on the model architecture.

The most chemically informative attribution pattern was the differential contribution of pH to the two outputs. For Cr(VI), pH showed a relevant contribution in the recurrent models. Low pH values were associated with negative SHAP contributions, indicating lower predicted Cr(VI) breakthrough, whereas higher pH values yielded positive SHAP contributions, indicating greater Cr(VI) passage through the column. This pattern is consistent with the pH-dependent mechanisms previously reported for this biosorbent system:

under strongly acidic conditions, protonation of biosorbent surface groups enhances electrostatic attraction toward anionic Cr(VI) species (HCrO_4^- , $\text{Cr}_2\text{O}_7^{2-}$) while simultaneously promoting Cr(VI) bioreduction to Cr(III) [34,35]. As pH increases, both effects weaken, reducing Cr(VI) uptake and allowing greater Cr(VI) passage through the column.

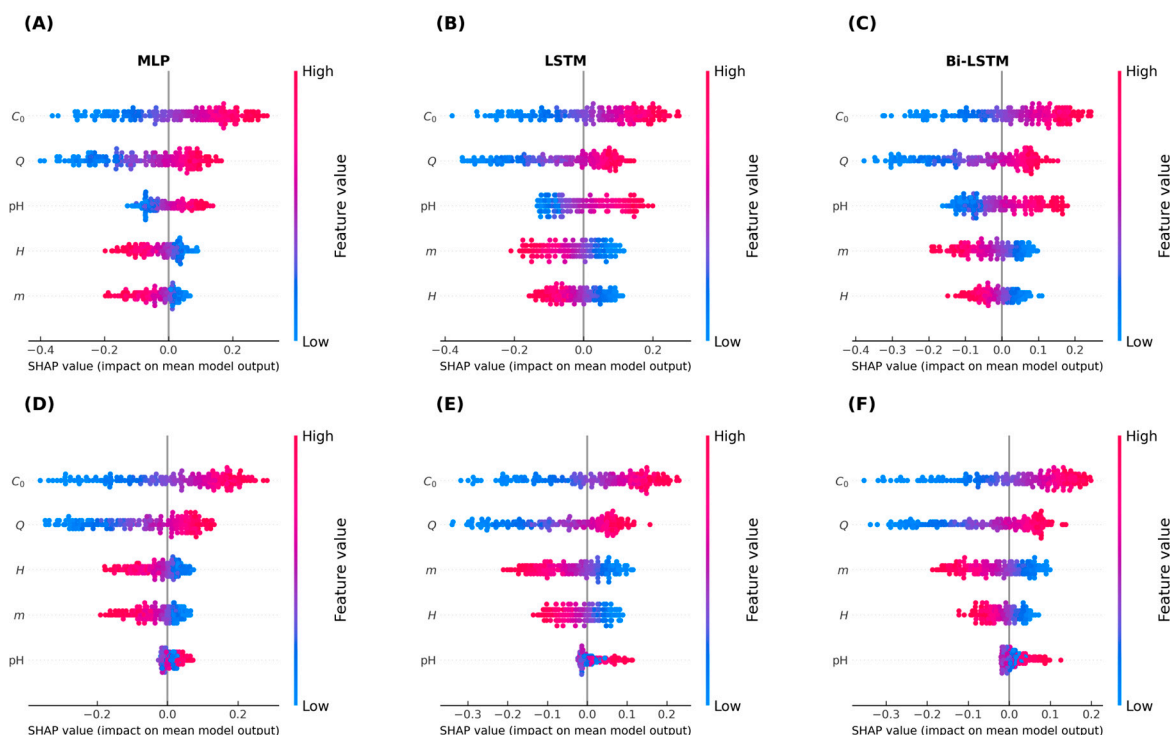


Figure 5. SHapley Additive exPlanations (SHAP) feature importance for Cr(VI) (A–C) and total chromium (D–F) removal, as predicted by the MLP (A,D), LSTM (B,E), and Bi-LSTM (C,F) models for Scenario A.

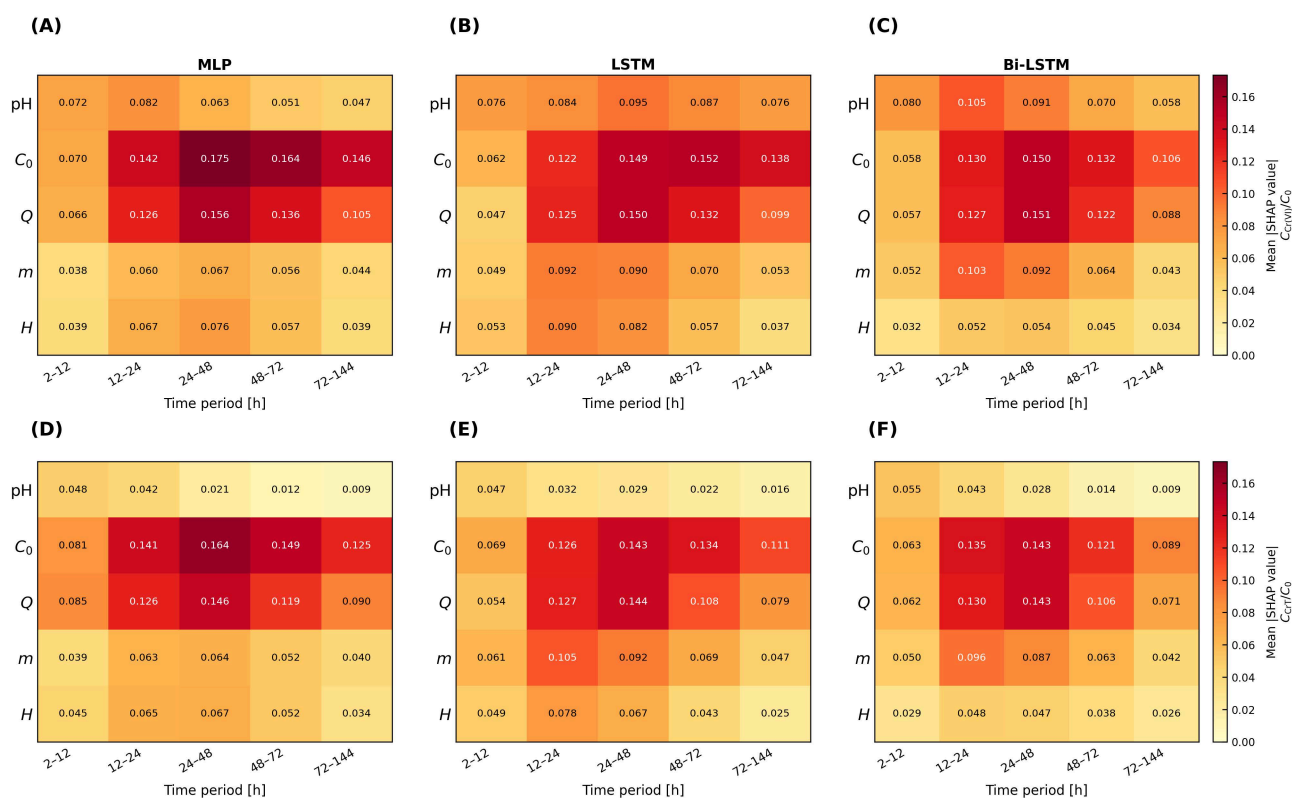
For total chromium, low pH values produced small SHAP contributions, suggesting a limited effect on the predicted total chromium C_t/C_0 under strongly acidic conditions. In contrast, higher pH values produced notable positive SHAP contributions, indicating increased predicted total chromium breakthrough. This pattern is consistent with a reduced affinity of the biosorbent for Cr(III) cations as surface deprotonation progresses. The resulting asymmetry reflects the distinct governing mechanisms for each output: Cr(VI) breakthrough is sensitive to pH across the evaluated range because low pH favors electrostatic attraction and bioreduction, whereas total chromium breakthrough is affected more strongly at higher pH, where diminished cation-binding capacity increases chromium passage through the column.

The global mean $|\text{SHAP}|$ values reported in Table 7 confirm that C_0 and Q were the two most influential operational features across all architectures and both outputs. For Cr(VI), pH showed a higher contribution than m in the recurrent models (0.084 and 0.076 for LSTM and Bi-LSTM, respectively), whereas the MLP assigned similar importance to pH, H , and m . For total chromium, pH ranked lowest across all three architectures (0.022–0.026), below C_0 , Q , m , and H . This result supports the interpretation that pH had a stronger model-attribution role for Cr(VI) breakthrough than for total chromium breakthrough.

Table 7. Global SHAP feature importance (mean |SHAP| values) for the MLP, LSTM, and Bi-LSTM models (Scenario A) for the Cr(VI) and total chromium C_t/C_0 outputs.

Feature	Cr(VI)			Total Chromium		
	MLP	LSTM	Bi-LSTM	MLP	LSTM	Bi-LSTM
C_0	0.144	0.129	0.120	0.136	0.121	0.112
Q	0.122	0.116	0.114	0.116	0.104	0.101
m	0.052	0.071	0.069	0.052	0.073	0.067
pH	0.058	0.084	0.076	0.022	0.026	0.024
H	0.057	0.061	0.044	0.053	0.051	0.037

Although the global mean |SHAP| values in Table 7 summarize feature contributions across the full breakthrough curve, they may mask changes in attribution during different stages of column operation. The time-resolved analysis (Figure 6) addresses this limitation by partitioning the breakthrough process into five consecutive windows (2–12, 12–24, 24–48, 48–72, and 72–144 h), thereby revealing how feature importance evolves as the column transitions from initial uptake to saturation. This windowed analysis describes how the attribution of the five operational variables evolves across successive stages of the breakthrough process.

**Figure 6.** Time-resolved mean |SHAP| values for the MLP (A,D), LSTM (B,E), and Bi-LSTM (C,F) across the Cr(VI) (A–C) and total chromium (D–F) outputs.

The time-resolved analysis confirmed that C_0 and Q were the dominant contributors during the active breakthrough region, although their exact temporal maxima varied slightly among architectures and outputs. For most cases, the largest or near-largest mean |SHAP| values for C_0 and Q occurred during the 24–48 h window, coinciding with the steepest part of the breakthrough profiles and the greatest sensitivity to influent loading and residence time. In the LSTM model for Cr(VI), C_0 importance remained slightly higher

during 48–72 h than during 24–48 h, indicating a more sustained contribution of influent concentration into the later breakthrough region.

The temporal pattern of biosorbent mass was more architecture-dependent. For Cr(VI), m reached its maximum contribution during 24–48 h in the MLP, whereas in the LSTM and Bi-LSTM models, its maximum occurred earlier, during 12–24 h. For total chromium, m also contributed most strongly during the early-to-active breakthrough period, with the recurrent models showing their highest m attribution at 12–24 h. These patterns are consistent with the role of bed capacity in delaying the onset of saturation but also indicate that the timing of this contribution depended on how each architecture represented the breakthrough sequence.

The temporal evolution of pH importance reinforced the output-dependent pattern observed in the global analysis. For Cr(VI), pH showed its highest attribution during the early-to-mid breakthrough stages, with maxima occurring between 12 and 48 h depending on the architecture. This trend is consistent with the pH-dependent uptake and reduction of Cr(VI), particularly during the period when active sites are still available. For total chromium, pH attribution remained lower across all time windows and decreased toward the saturation region, supporting the interpretation that pH had a stronger model-attribution role for Cr(VI) than for total chromium.

The heatmaps also revealed architecture-dependent differences that were partially obscured in the global rankings. During the early phase (2–12 h), the MLP assigned comparable contributions to pH, C_0 , and Q for Cr(VI), whereas C_0 and Q were more prominent for total chromium. In the recurrent models, pH was the dominant early feature for Cr(VI), while early total chromium predictions were influenced more evenly by C_0 , Q , and bed-related variables.

Between the two recurrent architectures, additional temporal distinctions were evident. For Cr(VI), the LSTM sustained higher C_0 importance into the late phase than the Bi-LSTM, with mean $|\text{SHAP}|$ values of 0.138 and 0.106, respectively, at 72–144 h. The LSTM also attributed greater importance to H in most time windows and outputs, whereas the Bi-LSTM concentrated the importance of C_0 and Q more sharply within the active breakthrough period and showed a more pronounced decline in pH importance as saturation approached. These differences indicate that, although the LSTM and Bi-LSTM produced similar global feature rankings, they distributed feature contributions differently along the temporal axis—a distinction that becomes apparent only through period-level decomposition.

These temporal attribution patterns complement the holdout and consensus analyses. The recurrent architectures distributed feature contributions across the breakthrough sequence and produced more consistent Cr(VI) profiles under the most challenging holdout conditions. In contrast, the MLP treated each time point independently, which likely contributed to its lower robustness under peripheral or sparsely represented conditions, particularly for the Cr(VI) output. The convergence of holdout metrics, curve-shape assessment, inter-model consensus, and SHAP-based attribution supports the use of multiple diagnostic criteria when evaluating neural-network models for data-limited biosorption systems.

The main limitation of this study is the narrow range of fixed-bed configurations represented in the dataset. The experiments were conducted in relatively short laboratory-scale beds (1.7–6.5 cm) and under a limited range of hydraulic residence conditions, which may have restricted the development and representation of the mass-transfer zone. Accordingly, the predictive robustness demonstrated by the present models should be interpreted only within the experimental domain covered by the dataset, and the results should not be extrapolated to taller beds, longer residence times, or different hydrodynamic regimes. Broader applicability would require additional experiments covering a wider range of bed geometries and hydraulic conditions, followed by model retraining and external validation.

Within the current experimental domain, however, model improvement could be pursued through targeted expansion of the dataset. In the present study, the restricted size and OFAT structure of the dataset left some portions of the operational space sparsely represented. The architecture- and output-dependent differences observed in the holdout evaluation, together with the high-divergence regions identified by the CoV- D_M analysis, illustrate how future experiments could be prioritized in these regions. Such targeted sampling would allow assessment of whether additional data reduce inter-model disagreement and improve complete-curve prediction. If a specific objective function were defined, the framework could also support Bayesian optimization or active-learning-based experimental design.

A subsequent extension would be online breakthrough prediction, in which recurrent models are updated or corrected as effluent concentration measurements become available during column operation. However, full feedback-control implementation should be considered a longer-term application because it would require real-time sensing, dynamic validation, and control-system integration. Overall, these extensions would move the proposed framework from retrospective model evaluation toward progressively more reliable, experimentally guided decision support for fixed-bed biosorption processes.

4. Conclusions

MLP, LSTM, and Bi-LSTM neural networks were developed and compared for the simultaneous prediction of Cr(VI) and total chromium breakthrough curves in fixed-bed columns packed with *Quercus crassipes* acorn shell. In the internal five-fold cross-validation of the scenario-specific training sets, all architectures achieved high point-wise accuracy, with RMSE values between 0.0149 and 0.0309 and R^2 values between 0.9879 and 0.9979. The MLP yielded the lowest point-wise validation errors, whereas the Bi-LSTM provided the most consistent performance among the recurrent models. The LSTM exhibited the highest inter-fold variability, with global CoV values reaching 25.0%.

External holdout evaluation showed that complete-curve prediction depended strongly on the selected holdout condition and on the response variable. Under the central condition (Scenario A, K03, $D_M = 0.317$), all architectures produced high holdout performance, with R^2 values ranging from 0.9537 to 0.9896 for Cr(VI) and from 0.9541 to 0.9880 for total chromium. Performance decreased under the acidic boundary condition (Scenario B1, K01, $D_M = 2.302$) and under the low-biosorbent-mass condition (Scenario B2, K06, $D_M = 1.751$). The most pronounced deterioration occurred for Cr(VI) under the peripheral low-flow condition (Scenario C, K09, $D_M = 2.802$), particularly for the MLP, whose holdout R^2 decreased to 0.3251. In that scenario, total chromium was less affected than Cr(VI), with holdout R^2 values between 0.8266 and 0.9624.

These results indicate that high internal validation accuracy did not translate uniformly to complete-curve generalization. The MLP remained competitive under central or better-represented conditions but was less effective for the more complex Cr(VI) response under the most challenging holdout case. The recurrent models did not eliminate prediction errors, but their sequence-level representation helped preserve more consistent Cr(VI) breakthrough profiles under poorly represented conditions.

The combined use of Mahalanobis distance and inter-model CoV provided a practical diagnostic framework for analyzing prediction behavior across the operational space. For $D_M < 2$, the three architectures produced similar total chromium capacity predictions, with CoV values generally below 12%, but localized divergence zones above 20% were identified, particularly at low biosorbent mass combined with high flow rate and at selected peripheral combinations. These results show that D_M and CoV provide complementary

information: D_M describes geometric distance from the training distribution, whereas CoV reflects architecture-dependent disagreement.

SHAP-based attribution analysis identified influent concentration and flow rate as the dominant operational variables across architectures and outputs. For Cr(VI), C_0 and Q showed the highest global mean $|SHAP|$ values (0.120–0.144 and 0.114–0.122, respectively), while pH was especially relevant in the recurrent models (0.076–0.084). In contrast, pH had lower attribution values for total chromium (0.022–0.026), supporting its stronger role in Cr(VI) breakthrough. Time-resolved SHAP further showed that C_0 and Q contributed most strongly during the 24–48 h active breakthrough interval.

Overall, the results support the use of recurrent neural networks for reliability-aware breakthrough-curve modeling, particularly when sequential information and curve-shape consistency are relevant. However, the predictive robustness demonstrated in this study is restricted to the laboratory-scale column geometries, hydraulic residence conditions, and operating ranges represented in the training dataset. Compared with the feedforward MLP, the recurrent architectures provided a more suitable representation of complete breakthrough profiles under challenging holdout conditions, although no single model was uniformly superior across all scenarios and outputs. Future work should prioritize targeted experiments in high-divergence regions, followed by model retraining, to reduce architecture-dependent disagreement and expand the reliable operating domain for column-design analysis, scenario evaluation, and decision support. This strategy would provide a practical transition from retrospective model assessment toward experimentally guided, adaptive modeling of fixed-bed biosorption processes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr14152444/s1>, Table S1: Optimal hyperparameters for the multilayer perceptron (MLP), identified by Bayesian optimization for each holdout scenario; Table S2: Optimal hyperparameters for the long short-term memory (LSTM) architecture, identified by Bayesian optimization for each holdout scenario; Table S3: Optimal hyperparameters for the bidirectional long short-term memory (Bi-LSTM) architecture, identified by Bayesian optimization for each holdout scenario.

Author Contributions: Conceptualization, A.R.N.-M., J.C.C.-V. and E.C.-U.; methodology, A.R.N.-M. and E.C.-U.; software, A.R.N.-M. and J.C.C.-V.; validation, J.C.C.-V.; formal analysis, A.R.N.-M. and J.C.C.-V.; investigation, A.R.N.-M. and E.A.-G.; resources, E.C.-U.; data curation, J.C.C.-V. and E.A.-G.; writing—original draft preparation, A.R.N.-M.; writing—review and editing, A.R.N.-M., J.C.C.-V., E.A.-G. and E.C.-U.; visualization, A.R.N.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the findings of this study are openly available in Figshare at <https://doi.org/10.6084/m9.figshare.32049024>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kumar, D.; Rani, J.; Selwal, K.K. Low-cost adsorbents: A sustainable strategy for emerging contaminant mitigation. *Water Air Soil Pollut.* **2025**, *236*, 947. [\[CrossRef\]](#)
2. Al-Ghouti, M.A.; Ashfaq, M.Y.; Khan, M.; Al Disi, Z.; Da'na, D.A.; Shoshaa, R. State-of-the-art adsorption and adsorptive filtration based technologies for the removal of trace elements: A critical review. *Sci. Total Environ.* **2023**, *895*, 164854. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Topare, N.S.; Wadgaonkar, V.S. A review on application of low-cost adsorbents for heavy metals removal from wastewater. *Mater. Today Proc.* **2023**, *77*, 8–18. [\[CrossRef\]](#)
4. Satyam, S.; Patra, S. Innovations and challenges in adsorption-based wastewater remediation: A comprehensive review. *Heliyon* **2024**, *10*, e29573. [\[CrossRef\]](#) [\[PubMed\]](#)

5. Yaashikaa, P.R.; Palanivelu, J.; Hemavathy, R.V. Sustainable approaches for removing toxic heavy metal from contaminated water: A comprehensive review of bioremediation and biosorption techniques. *Chemosphere* **2024**, *357*, 141933. [[CrossRef](#)] [[PubMed](#)]
6. Agarwal, A.; Upadhyay, U.; Sreedhar, I.; Singh, S.A.; Patel, C.M. A review on valorization of biomass in heavy metal removal from wastewater. *J. Water Process Eng.* **2020**, *38*, 101602. [[CrossRef](#)]
7. Kainth, S.; Sharma, P.; Pandey, O.P. Green sorbents from agricultural wastes: A review of sustainable adsorption materials. *Appl. Surf. Sci. Adv.* **2024**, *19*, 100562. [[CrossRef](#)]
8. Karić, N.; Maia, A.S.; Teodorović, A.; Atanasova, N.; Langergraber, G.; Crini, G.; Ribeiro, A.R.L.; Đolić, M. Bio-waste valorisation: Agricultural wastes as biosorbents for removal of (in)organic pollutants in wastewater treatment. *Chem. Eng. J. Adv.* **2022**, *9*, 100239. [[CrossRef](#)]
9. Chu, K.H. Fixed bed adsorption of water contaminants: A cautionary guide to simple analytical models and modeling misconceptions. *Sep. Purif. Rev.* **2023**, *52*, 75–97. [[CrossRef](#)]
10. Thirunavukkarasu, A.; Nithya, R.; Sivashankar, R. Continuous fixed-bed biosorption process: A review. *Chem. Eng. J. Adv.* **2021**, *8*, 100188. [[CrossRef](#)]
11. Hu, Q.; Yang, X.; Huang, L.; Li, Y.; Hao, L.; Pei, Q.; Pei, X. A critical review of breakthrough models with analytical solutions in a fixed-bed column. *J. Water Process Eng.* **2024**, *59*, 105065. [[CrossRef](#)]
12. Basu, A.; Ali, S.S.; Hossain, S.S.; Asif, M. A review of the dynamic mathematical modeling of heavy metal removal with the biosorption process. *Processes* **2022**, *10*, 1154. [[CrossRef](#)]
13. Zubir, A.; Normaya, E.; Danial, W.H.; Goh, P.S.; Mat Piah, M.B.; Show, P.L.; Ismail, A.F.; Ahmad, M.N. State-of-the-art modification, mechanistic insight and breakthrough curve analysis of chitosan scaffolds for sustainable heavy metal adsorbent. *Int. J. Environ. Sci. Technol.* **2026**, *23*, 195. [[CrossRef](#)]
14. Kimani, P.K. Asymmetrical fixed-bed breakthrough curve modelling: Comparing simplistic, log-modified, fractal-like, and probability distribution function models. *Chem. Eng. Res. Des.* **2024**, *201*, 446–456. [[CrossRef](#)]
15. Shen, X.; Thibault, J. Modelling adsorption breakthrough curves. *Separations* **2026**, *13*, 100. [[CrossRef](#)]
16. Koyama, Y.; Fasaee, M.A.K.; Berglund, E.Z.; Knappe, D.R.U. Machine learning models to predict early breakthrough of recalcitrant organic micropollutants in granular activated carbon adsorbents. *Environ. Sci. Technol.* **2024**, *58*, 17114–17124. [[CrossRef](#)] [[PubMed](#)]
17. Reynel-Ávila, H.E.; Aguayo-Villarreal, I.A.; Diaz-Muñoz, L.L.; Moreno-Pérez, J.; Sánchez-Ruiz, F.J.; Rojas-Mayorga, C.K.; Mendoza-Castillo, D.I.; Bonilla-Petriciolet, A. A review of the modeling of adsorption of organic and inorganic pollutants from water using artificial neural networks. *Adsorpt. Sci. Technol.* **2022**, *2022*, 9384871. [[CrossRef](#)]
18. Lowe, M.; Qin, R.; Mao, X. A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. *Water* **2022**, *14*, 1384. [[CrossRef](#)]
19. Balci, B.; Keskinan, O.; Avci, M. Use of BDST and an ANN model for prediction of dye adsorption efficiency of *Eucalyptus camaldulensis* barks in fixed-bed system. *Expert Syst. Appl.* **2011**, *38*, 949–956. [[CrossRef](#)]
20. Tomczak, E. Application of ANN and EA for description of metal ions sorption on chitosan foamed structure—Equilibrium and dynamics of packed column. *Comput. Chem. Eng.* **2011**, *35*, 226–235. [[CrossRef](#)]
21. Banerjee, M.; Bar, N.; Basu, R.K.; Das, S.K. Comparative study of adsorptive removal of Cr(VI) ion from aqueous solution in fixed bed column by peanut shell and almond shell using empirical models and ANN. *Environ. Sci. Pollut. Res.* **2017**, *24*, 10604–10620. [[CrossRef](#)] [[PubMed](#)]
22. Banerjee, M.; Bar, N.; Basu, R.K.; Das, S.K. Removal of Cr(VI) from its aqueous solution using green adsorbent pistachio shell: A fixed bed column study and GA-ANN modeling. *Water Conserv. Sci. Eng.* **2018**, *3*, 19–31. [[CrossRef](#)]
23. Nag, S.; Bar, N.; Das, S.K. Cr(VI) removal from aqueous solution using green adsorbents in continuous bed column—Statistical and GA-ANN hybrid modelling. *Chem. Eng. Sci.* **2020**, *226*, 115904. [[CrossRef](#)]
24. Bai, S.; Ya, R.; Ding, W.; Xie, X.; Zhou, M.; Zhao, G.; Yang, T. Continuous removal of fluoride by Fe/Al-anchored bamboo charcoal in fixed bed column and prediction of breakthrough curve using artificial neural network model. *Water Air Soil Pollut.* **2026**, *237*, 26. [[CrossRef](#)]
25. Costa, H.P.S.; Oliveira, M.G.; Duarte, E.D.V.; Gomes, L.; Selvasembian, R.; da Silva, M.G.C.; Vieira, M.G.A. Multi-walled carbon nanotubes green-functionalized with iron nanoparticles for continuous removal of pharmaceutical pollutants through fixed-bed adsorption: Integrated experimental and machine learning approaches. *Environ. Sci. Pollut. Res.* **2025**, *32*, 18058–18075. [[CrossRef](#)] [[PubMed](#)]
26. Halalsheh, N.; Alshboul, O.; Shehadeh, A.; Al Mamlook, R.E.; Al-Othman, A.; Tawalbeh, M.; Saeed Almuflih, A.; Papelis, C. Breakthrough curves prediction of selenite adsorption on chemically modified zeolite using boosted decision tree algorithms for water treatment applications. *Water* **2022**, *14*, 2519. [[CrossRef](#)]
27. Sridevi, H.; Bhat, M.R.; Selvaraj, R. Machine learning predictions for 2,4-dichlorophenoxyacetic acid removal: Breakthrough analysis using Fe₂O₃@PPy nanocomposite columns. *J. Ind. Eng. Chem.* **2025**, *144*, 414–424. [[CrossRef](#)]
28. Sudarsan, S.; Srinivasan, N.R.; Vinayagam, R.; Selvaraj, R. Machine learning-enhanced breakthrough modeling of malachite green adsorption onto superparamagnetic activated carbon. *Chem. Eng. J. Adv.* **2025**, *24*, 100932. [[CrossRef](#)]

29. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
30. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [[CrossRef](#)] [[PubMed](#)]
31. Chen, S.Z.; Wang, X.Y.; Ma, X.R.; Wang, T.; Askari, M.; Taciroglu, E. Transforming regression to classification for enhancing extrapolation capacity of data-driven structural behavior prediction. *Expert Syst. Appl.* **2025**, *277*, 127302. [[CrossRef](#)]
32. Xu, K.; Zhang, M.; Li, J.; Du, S.S.; Kawarabayashi, K.; Jegelka, S. How neural networks extrapolate: From feedforward to graph neural networks. *arXiv* **2021**. arXiv:2009.11848.
33. Gkika, D.A.; Tolkou, A.K.; Katsoyiannis, I.A.; Kyzas, G.Z. The adsorption-desorption-regeneration pathway to a circular economy: The role of waste-derived adsorbents on chromium removal. *Sep. Purif. Technol.* **2025**, *368*, 132996. [[CrossRef](#)]
34. Aranda-García, E.; Morales-Barrera, L.; Pineda-Camacho, G.; Cristiani-Urbina, E. Effect of pH, ionic strength, and background electrolytes on Cr(VI) and total chromium removal by acorn shell of *Quercus crassipes* Humb. & Bonpl. *Environ. Monit. Assess.* **2014**, *186*, 6207–6221. [[CrossRef](#)] [[PubMed](#)]
35. Aranda-García, E.; Cristiani-Urbina, E. Hexavalent chromium removal and total chromium biosorption from aqueous solution by *Quercus crassipes* acorn shell in a continuous up-flow fixed-bed column: Influencing parameters, kinetics, and mechanism. *PLoS ONE* **2020**, *15*, e0227953. [[CrossRef](#)] [[PubMed](#)]
36. Fritsch, F.N.; Butland, J. A method for constructing local monotone piecewise cubic interpolants. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 300–304. [[CrossRef](#)]
37. Vijay, J.; Vaishnavi, S.; Prabhakar, V. Enhancing wave parameters prediction: Machine learning models combined with PCHIP. *Ocean Dyn.* **2025**, *75*, 101. [[CrossRef](#)]
38. Mahalanobis, P.C. On the generalized distance in statistics. *Sankhyā A* **2018**, *80*, S1–S7. [[CrossRef](#)]
39. Razavi, S. Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. *Environ. Model. Softw.* **2021**, *144*, 105159. [[CrossRef](#)]
40. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: <https://www.deeplearningbook.org/> (accessed on 25 March 2026).
41. da Silva, D.G.; Meneses, A.A.M. Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction. *Energy Rep.* **2023**, *10*, 3315–3334. [[CrossRef](#)]
42. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 4766–4777. Available online: <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html> (accessed on 10 March 2026).
43. Nurhayati, M.; Lee, H.-J.; Park, J.E.; Hong, B.U.; Kang, H.G.; Kim, S.; Lee, S. Prediction of hydrocarbon adsorption-desorption dynamics in activated carbon columns using long short-term memory networks for intelligent removal systems. *Eng. Appl. Artif. Intell.* **2026**, *163*, 113005. [[CrossRef](#)]
44. De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D.L. The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 1–18. [[CrossRef](#)]
45. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 6402–6413. Available online: <https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html> (accessed on 10 February 2026).
46. Na, G.S.; Kim, H.W. Metaheuristics-guided active learning for optimizing reaction conditions of high-performance methane conversion. *Appl. Soft Comput.* **2024**, *164*, 111935. [[CrossRef](#)]
47. Lookman, T.; Balachandran, P.V.; Xue, D.; Yuan, R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput. Mater.* **2019**, *5*, 21. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.