

Article

Long Short-Term Memory and Bidirectional Long Short-Term Memory Modeling and Prediction of Hexavalent and Total Chromium Removal Capacity Kinetics of *Cupressus lusitanica* Bark

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

¹ 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

² Programa Académico de Ingeniería en Biotecnología, Universidad Politécnica de Tlaxcala, Avenida Universidad Politécnica No. 1, San Pedro Xalcaltzinco, Tepeyanco 90180, Tlaxcala, Mexico; almarosa.netzahuatl@uptlax.edu.mx

³ Departamento de Ingeniería Bioquímica, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Unidad Profesional Adolfo López Mateos, Avenida Wilfrido Massieu s/n, Delegación Gustavo A. Madero, Ciudad de México 07738, Mexico

* Correspondence: ecristiani@ipn.mx or ecristianiu@yahoo.com.mx

Abstract: Hexavalent chromium [Cr(VI)] is a high-priority environmental pollutant because of its toxicity and potential to contaminate water sources. Biosorption, using low-cost biomaterials, is an emerging technology for removing pollutants from water. In this study, Long Short-Term Memory (LSTM) and bidirectional LSTM (Bi-LSTM) neural networks were used to model and predict the kinetics of the removal capacity of Cr(VI) and total chromium [Cr(T)] using *Cupressus lusitanica* bark (CLB) particles. The models were developed using 34 experimental kinetics datasets under various temperature, pH, particle size, and initial Cr(VI) concentration conditions. Data preprocessing via interpolation was implemented to augment the sparse time-series data. Early stopping regularization prevented overfitting, and dropout techniques enhanced model robustness. The Bi-LSTM models demonstrated a superior performance compared to the LSTM models. The inherent complexities of the process and data limitations resulted in a heavy-tailed and left-skewed residual distribution, indicating occasional deviations in the predictions of capacities obtained under extreme conditions. K-fold cross-validation demonstrated the stability of Bi-LSTM models 38 and 43, while response surfaces and validation with unseen datasets assessed their predictive accuracy and generalization capabilities. Shapley additive explanations analysis (SHAP) identified the initial Cr(VI) concentration and time as the most influential input features for the models. This study highlights the capabilities of deep recurrent neural networks in comprehending and predicting complex pollutant removal kinetic phenomena for environmental applications.

Keywords: LSTM; Bi-LSTM; chromium; SHAP; modeling



Citation: Cruz-Victoria, J.C.; Netzahuatl-Muñoz, A.R.; Cristiani-Urbina, E. Long Short-Term Memory and Bidirectional Long Short-Term Memory Modeling and Prediction of Hexavalent and Total Chromium Removal Capacity Kinetics of *Cupressus lusitanica* Bark. *Sustainability* **2024**, *16*, 2874. <https://doi.org/10.3390/su16072874>

Academic Editors: Xiaofei Tan and Jia Wen

Received: 8 February 2024

Revised: 27 March 2024

Accepted: 28 March 2024

Published: 29 March 2024



Copyright: © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chemical pollution of the environment has been identified as one of nine critical frontiers that should not be exceeded, in order to avoid catastrophic impacts on the Earth's balance and human well-being [1]. Metals are globally dispersed because of their persistence, high production, global trade, and widespread use in multiple applications [2].

Chromium is a transition metal that is extensively utilized in various industrial processes such as metal plating, wood treatment, metal refining, stainless steel production, leather tanning, and chemical dye production [3]. However, one of its most stable forms, Cr(VI), is considered a high-priority pollutant because of its high toxicity and ability to contaminate soil and drinking water sources [4].

The global dispersal of Cr(VI) and other metals seriously threatens food security, because metals can accumulate in the food chain and reach hazardous levels in agricultural products [5]. Exposure to Cr(VI) can have detrimental health consequences for humans and is classified as carcinogenic by the International Agency for Research on Cancer [6]. It can also cause damage to the skin, mucous membranes, skin allergies, respiratory problems, stomach ulcers, and harm to the kidneys and liver, depending on the form of exposure [3,7]. Controlling and mitigating chemical pollution, especially metal-related pollution, is essential for maintaining environmental sustainability and protecting human health.

Biosorption, a fundamental technology for treating metal-contaminated water, is a process in which biological materials, notably agricultural and forestry waste products, operate as effective adsorbents to remove metal ions from contaminated aqueous solutions. This process relies on intricate interactions, including ion exchange, complexation, and surface adsorption, to capture and retain metal ions [8]. Biosorbents are favored in this endeavor because of their cost-effectiveness, wide availability, and remarkable adsorption capacity.

The study of metal biosorption involves extensive research, primarily on kinetic and equilibrium studies. Typically, univariate analysis is employed, in which one variable is studied independently, while keeping the others constant, often fitting different mechanistic or empirical models [9,10]. This approach allows for a precise understanding of the influence of each variable on the metal biosorption capacity of a given material. The key variables that have garnered significant attention include the initial metal concentration, pH, and temperature of the solution, as well as the particle size and concentration of the biomaterial [11,12]. Multivariate analyses investigating the interrelationships between various variables and their impact on biosorption have been explored to a lesser extent. These techniques delve into how different variables interact and how their combinations influence the adsorption capacity. The most frequently used tools for this analysis are the response surface technique, machine learning (ML), and deep learning (DL) [13–16].

ML involves the development of algorithms and analytical models that can learn from data, without explicitly relying on programming rules. These models automatically identify patterns in input data and use them to perform tasks such as classification, regression, clustering, and association [17]. Artificial neural networks (ANNs), decision trees, support vector machines, and reinforcement learning are among the most popular machine learning approaches [18].

DL refers to artificial neural networks with multiple layers that learn hierarchical representations from data [19]. Recurrent neural networks (RNNs) are powerful deep learning models for sequential data. A key feature of the RNN architecture is its cyclic connections, which allow the RNN to update its current state based on past states and current input data [20]. However, a standard RNN can struggle with long-term dependencies because of the vanishing gradient problem. Long Short-Term Memory (LSTM) networks were designed to overcome this limitation by introducing input, output, and forget gates [21,22]. The bidirectional LSTM (Bi-LSTM) further enhances sequence learning by processing data in both temporal directions [23].

RNNs are uniquely adaptable to effectively utilize missing value patterns, time intervals, and intricate temporal dependencies in irregular univariate and multivariate time-series data [24,25]. The architectures of LSTM and Bi-LSTM networks are especially well suited for addressing complex dynamics and nonlinear relationships in temporal data. This suitability is attributed to their sophisticated gate structure, which allows precise information flow control and enables them to effectively capture long-term dependencies [26].

When complex temporal dynamics are crucial, LSTM and Bi-LSTM networks may offer superior modeling capability and predictive performance compared to other architectures, such as GRU and standard RNNs [27]. Their versatility positions LSTM and Bi-LSTM networks as efficient tools for tackling real-world problems, making them robust and adaptable models for analyzing and predicting complex processes across various domains. These applications include contaminant removal through adsorption [14], biomass pyrolysis [28], constructive peptide design [29], and blood glucose prediction [30].

Cupressus lusitanica bark (CLB) was highly effective in removing Cr(VI) and Cr(T) from aqueous solutions. The removal of Cr(VI) using CLB is a complex phenomenon that involves both biosorption and Cr(VI) reduction. Biosorption adheres to pseudo-second-order kinetics, indicating that chemisorption is the rate-limiting step [31,32].

The biosorption of chromium from Cr(VI) solutions using CLB involves four complex reaction steps. This process begins with the formation of Cr(VI) complexes through interactions between Cr(VI) ions and oxygen-containing groups, resulting in the adsorption of Cr(VI) oxyanions. In the second step, Cr(VI) is reduced to trivalent chromium [Cr(III)]. The third step involves forming carboxyl groups through the oxidation of oxygen-containing groups. The last step involves the interaction of Cr(III) with the carboxyl groups, forming Cr(III)–carboxylate complexes. These successive reactions collectively contribute to the complex yet effective removal of chromium from Cr(VI) solutions using CLB [32].

The Cr(III) formed can either remain adsorbed or transition into the aqueous phase, suggesting that the biomaterial may have a higher capacity for removing Cr(VI) from the solution than Cr(T). Understanding this distinction is crucial for accurately assessing the efficiency of a biomaterial in removing different chromium species (Cr(VI) and/or Cr(III)). Furthermore, the literature contains a diverse range of biomaterials that exhibit mechanisms like those of CLB [33,34]. These processes are intricately dependent on operational conditions and time, posing a significant challenge in comprehending the kinetics of this phenomenon.

Prior studies have explored the application of ML techniques, such as ANNs, to model and predict the removal of Cr(VI) or Cr(T) using various biomaterials. These investigations encompassed a range of approaches, including feed-forward neural networks with backpropagation algorithms [35,36], multilayer perceptrons [37–39], support vector machines [40], hybrid models combining genetic algorithms with ANNs [41], ANNs paired with Particle Swarm Optimization [42], Adaptive Neuro-Fuzzy Inference Systems [43,44], and Random Forest algorithms [45]. However, most of these methods have primarily focused on predicting removal efficiency and adsorption capacity in equilibrium studies, with limited emphasis on modeling continuous and batch adsorption kinetics. Notably, the complex kinetic removal of Cr(VI) using these biomaterials, considering both Cr(VI) and Cr(T), has not been previously modeled using ML or DL techniques [46,47]. Understanding this kinetic complexity is crucial for accurately assessing the material's efficiency in removing the different chromium species and optimizing process conditions.

LSTM and Bi-LSTM neural networks were chosen for their robustness in handling incomplete data and their superior performance in modeling extended temporal sequences. Their capability to capture long-term dependencies positions them as optimal choices for the modeling objectives, particularly in predicting the dynamic adsorption capacities of CLB for Cr(VI) and Cr(T). Notably, this study represents a pioneering effort and is the first to consider the removal of Cr(VI) and Cr(T) through deep-learning-based modeling. This groundbreaking approach advances our comprehension of CLB's efficacy in Cr(VI) removal and is a compelling example of applying ML and DL techniques in environmental engineering research

2. Materials and Methods

2.1. Experimental Data

The experimental data utilized in this study originated from various univariate kinetic investigations involving the removal of Cr(VI) and Cr(T) using CLB. The methodology for data collection was previously documented by Netzahuatl-Muñoz et al. [31,32]. Unpublished kinetic datasets were integrated to expand the scope of the conditions under examination. In total, 37 kinetic datasets were employed in this analysis, covering a range of variables, including pH, temperature (T), initial Cr(VI) concentration (C_0), smallest particle size (SPS), and largest particle size (LPS). Each dataset comprised a sequence of data points that recorded the removal capacities of Cr(VI) [$q_{Cr(VI)}$] and Cr(T) [$q_{Cr(T)}$], using CLB, as a function of time.

2.2. Data Preprocessing

Data preprocessing was critical because of missing and incomplete time-series data. The original datasets had fewer than 25 data points per time series with irregular sampling intervals, making measurements unavailable at all possible time points across all the kinetics. This data sparsity and irregular temporal coverage pose challenges, as inadequate time-series data can negatively impact model performance and accuracy when training ML models [48]. A Python linear interpolation algorithm was implemented to augment the data to address this data density issue. By interpolating existing data points in time, missing values were filled and the total number of data points for 34 kinetics was significantly increased, achieving a 15-fold expansion to 7429 data points. Interpolation-based augmentation enhances information density, while preserving the overall trends and patterns of the original sparse datasets.

2.3. Data Partition and Input–Output Variable Assessment

The dataset was divided into training and testing sets, with 80% of the data allocated for training the DL algorithms and the remaining 20% was reserved for evaluating the model performance. The specific maximum and minimum values of each variable under consideration are listed in Table 1. It is important to note that this analysis collectively considered both output variables, as follows: the removal capacity of Cr(VI) and Cr(T) using CLB.

Table 1. Ranges of variables for network modeling of Cr(VI) and Cr(T) removal capacity using CLB.

Layer	Variable	Variable Code	Minimum Value	Maximum Value
Input	Temperature (°C)	T	15.0	45.0
	pH	pH	1.0	4.0
	Initial Cr(VI) concentration (mg L ⁻¹)	Co	10.6	1035
	Smallest particle size (µm)	SPS	150	1410
	Largest particle size (µm)	LPS	180	1700
	Time (h)	time	0	144
Output	Cr(VI) removal capacity (mg g ⁻¹)	qCr(VI)	0	611
	Cr(T) removal capacity (mg g ⁻¹)	qCr(T)	0	279

2.4. Neural Network Architecture and Regularization Techniques

The architecture employed in this study is based on two variants of recurrent neural networks, LSTM and Bi-LSTM. These networks belong to a category recognized for their ability to capture long-term dependencies in sequential data. A distinctive feature of LSTM networks is their ability to address the vanishing gradient problem through an intelligent cell design incorporating interactive gates to regulate the information flow.

2.4.1. LSTM Cell Mechanism

The components of an LSTM cell play a crucial role in managing the flow of information within a network, as illustrated in Figure 1. The LSTM cell has three fundamental gates, as follows: input, forget, and output [49]. The input gate regulates the integration of new information into the cell, the forget gate controls the specific removal of previous information, and the output gate determines the information emitted by the cell [20].

The key equations defining the LSTM cell function are as follows [20]:

$$f_t = \sigma(W_f \cdot [h_{t-1}; x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}; x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}; x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}; x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

where f_t , i_t , and o_t are the activation vectors of the forget, input, and output gates at time t , respectively; x_t denotes the input at time t ; \tilde{C}_t is the cell input activation vector, c_t and c_{t-1} are the cell state at the current and previous time step, respectively; h_t and h_{t-1} represent the hidden state at t and from the previous time step, respectively; σ denotes the sigmoid function, which bounds the gate outputs to a range between 0 and 1; \tanh represents the hyperbolic tangent function, ensuring cell state values remain within the interval $[-1, 1]$; W and b symbolize the weight matrices and bias vectors, respectively, optimized during the learning process; concatenation is represented by $[\cdot]$; and $[\cdot]$ signifies element-wise multiplication, known as the Hadamard product.

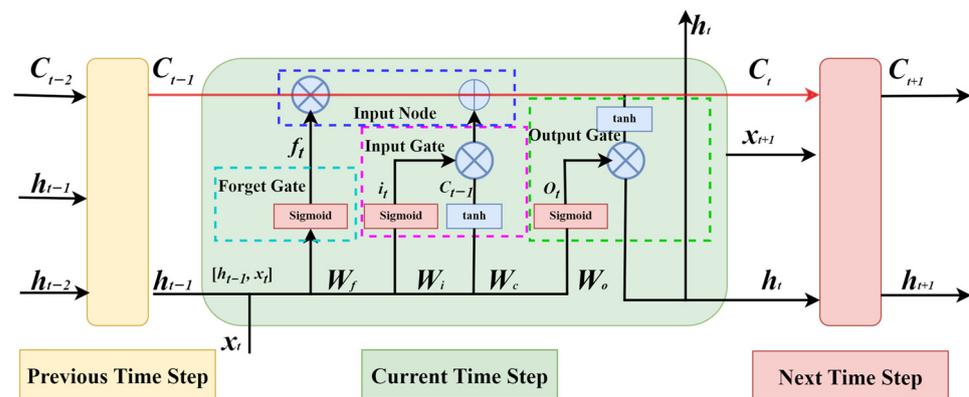


Figure 1. Schematic representation of the internal architecture and operations of an LSTM cell.

The conventional \tanh activation function in an LSTM cell can be substituted by alternative non-linear functions. The Exponential Linear Unit (ELU) and Rectified Linear Unit (ReLU) activation functions were used. The ELU provides a smoother activation for negative values, whereas the ReLU was selected for its computational efficiency and capability to expedite convergence during network training [50,51].

2.4.2. LSTM and Bi-LSTM Networks

The LSTM network represents a specialized architecture design for processing sequences in a singular temporal direction, typically forward. This unidirectional approach allows the network to capture the past context, fostering a cumulative understanding of sequential data. The unidirectional LSTM network performs a specific calculation to determine the final output at each time step (y_t). The following equation can succinctly express this process as follows:

$$y_t = \text{activation}(W_{hy}h_t + b_y) \quad (7)$$

where W_{hy} is the weight matrix connecting the output of the LSTM cell to y_t ; b_y is the bias vector; and activation represents the activation function.

The Bi-LSTM network is an extension of the LSTM architecture that introduces a nuanced approach to sequence processing [23]. As highlighted in the visual representation depicted in Figure 2, the Bi-LSTM network operates simultaneously in both the forward and backward temporal directions [52]. This bidirectional configuration enables the network to incorporate past and future information, providing a more comprehensive understanding of the data.

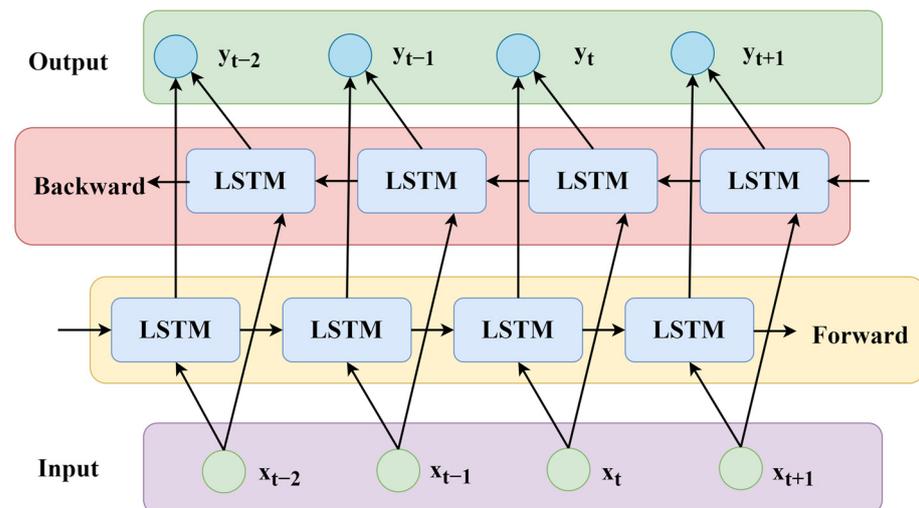


Figure 2. Integration of LSTM cells within the overall information flow in a Bi-LSTM network with a hidden layer.

In a bidirectional LSTM, implemented in TensorFlow Keras, y_t was computed by concatenating the hidden states from both the forward (h_t^f) and backward (h_t^b) directions [53]. Subsequently, a dense layer with an activation function was applied, as expressed by the following equation:

$$y_t = \text{activation}\left(W_{hy} \cdot \left[h_t^f; h_t^b\right] + b_y\right) \quad (8)$$

where the weights W_{hy} and bias b_y constitute the dense layer, following the Bi-LSTM layers. The choice of activation function (*activation*) in Equations (7) and (8) is determined by the specific nature of the problem and the desired interpretation of the model outputs.

2.4.3. LSTM and Bi-LSTM Network Configuration

The architecture consisted of two hidden layers, each with the same number of units, 12, 25, or 50. The activation functions used were ‘elu’ for the first hidden layer, ‘relu’ for the second hidden layer, and ‘softplus’ for the output layer. The recurrent activation function for the two hidden layers was ‘sigmoid.’ The Adam optimization algorithm was applied to adjust the network weights, with the mean square error (MSE) serving as the loss function (Equation (9)). It quantifies the average of the squared differences between the actual values (Y_i) and the predicted values (\hat{Y}_i) of i -th sample for the number of data points in the dataset (N).

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (9)$$

The hyperparameter tuning process involved experiments with different dropout and recurrent dropout value combinations in the second hidden layer. Dropout is a regularization technique that helps mitigate overfitting by randomly deactivating and activating neurons or neuron groups during training [54]. Recurrent dropout, however, regulates recurrent connections within the LSTM units. The objective was to determine the optimal configuration to achieve the best predictive performance. The network configurations of the hyperparameters are summarized in Table 2.

Table 2. Hyperparameters for LSTM and Bi-LSTM neural networks in kinetics analysis.

Layer	Hyperparameter	Value
Input	Number of neurons	6
	Number of cells LSTM	12, 25, 50
First hidden	Activation function	elu
	Recurrent activation function	sigmoid
	Activation function dropout	0
	Recurrent activation dropout	0
	Return sequences	true
Second hidden	Number of cells LSTM	12, 25, 50
	Activation function	relu
	Recurrent activation function	sigmoid
	Activation function dropout	0, 0.3
	Recurrent activation dropout	0, 0.3, 0.6, 0.9
Output	Return sequences	false
	Number of neurons	2
	Activation function	softplus

Subsequently, the LSTM network was trained using the training dataset and predictions were made using the test dataset. A methodology known as Early Stopping was implemented to ensure the effectiveness of the LSTM and Bi-LSTM models in predicting removal kinetics. Early Stopping is a regularization technique used during the training of neural networks to prevent overfitting and to achieve a more generalizable model [55]. The MSE of the validation set was used as an evaluation metric and a stopping threshold of 10 epochs, without improvement in the MSE of the validation set, was defined.

2.5. Implementation of Deep Learning Algorithm

The Keras library in Python was selected for implementing the proposed neural networks because of its versatility and efficiency in developing neural network models. This process was performed in a Visual Studio Code (VS Code, Microsoft, Redmond, WA, USA) development environment, providing a highly versatile programming and code development environment. In addition, Google Colab, a cloud-based platform based on Jupyter Notebooks, was used to leverage high-performance computational resources, and facilitate data analysis and deep learning algorithm implementation.

2.6. Model Fitting Evaluation

Various analyses were conducted to assess the quality of the tested LSTM and Bi-LSTM models to ensure their accuracy and reliability. The coefficient of determination (R^2) provides insight into how the model explains data variability. This value was calculated using the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y}_i)^2} \quad (10)$$

where \bar{Y}_i represents the mean of the observations. R^2 was calculated using the test data to evaluate the model's goodness of fit.

2.7. Residual Analysis

The residuals (r_i) are the difference between Y_i and \hat{Y}_i for the test data, representing the discrepancy between the observed and predicted values. Quantile–quantile (Q-Q) plots were generated to examine the distribution of residuals in the test dataset. Furthermore, skewness and kurtosis were computed to assess the distribution of the residuals and their deviations from a normal distribution.

The Fisher–Pearson coefficient of skewness (g_1), which measures the asymmetry of distribution, was calculated as follows:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^3}{\left[\frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2 \right]^{3/2}} \quad (11)$$

where \bar{r} is the mean of the residuals, m_3 is the third momentum, and m_2 is the second momentum.

The kurtosis, which quantifies the shape of the residual distribution, was calculated using Fisher's definition of kurtosis (g_2).

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^4}{\left[\frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2 \right]^2} - 3 \quad (12)$$

where m_4 denotes the fourth momentum. Collectively, these measures assist in evaluating the extent to which the residuals conform to a normal distribution.

Plots and calculations were conducted using the 'stats' module from the SciPy Python library.

2.8. Post-Training Validation of Selected Models

2.8.1. Response Surface Validation

As a validation method, the predicted response surface of qCr(VI) and qCr(T) was obtained as a function of pH and time, under previously studied conditions ($T = 28 \text{ }^\circ\text{C}$, $\text{Co} = 100 \text{ mg g}^{-1}$, $\text{SPS} = 400 \text{ }\mu\text{m}$, $\text{LPS} = 500 \text{ }\mu\text{m}$). The generated response surfaces were compared with the expected trends and behaviors, based on the understanding of the chromium removal process using CLB.

2.8.2. Validation with Unseen Kinetic Data

To further validate the robustness and generalization capabilities of the developed DL models, they were evaluated against experimental kinetic data from three independent studies of Cr(VI) and Cr(T) removal using CLB. These studies were conducted under specific conditions not included in the training and testing data to develop the deep learning model. The experimental kinetic data were compared with the model predictions. The comparison between the experimental time series and model predictions allowed the evaluation of how the selected models generalized and performed in previously unseen situations.

2.8.3. K-Fold Cross-Validation

The k-fold cross-validation technique was used to ensure the stability and robustness of the selected LSTM and Bi-LSTM models. This technique validates the model performance by dividing the data into k subsets or folds, iteratively using one subset for testing and the remaining for training, which provides a comprehensive evaluation of the model's ability to generalize to unseen data and maintain consistent performance across different subsets of the dataset [56]. By exposing the models to a diverse range of training scenarios, this approach helps to identify and correct potential weaknesses [57].

Using the scikit-learn library in Python, the dataset was divided into five distinct subsets or folds. The 5-fold cross-validation process was then applied to each of the selected LSTM and Bi-LSTM models. For each fold, key metrics such as R^2 , MAE, and root-mean-squared error ($\text{RMSE} = \text{MSE}^{1/2}$) were obtained for the predicted qCr(VI) and qCr(T). The coefficient of variation (CoV) for each metric was subsequently calculated using the following formula:

$$\text{CoV} = \frac{\sigma}{\mu} \times 100 \quad (13)$$

where σ and μ ($\mu \neq 0$) are the metric's standard deviation and mean, respectively.

A feature with a CoV larger than 33% falls under the weak-to-high inconsistency category [58].

2.9. SHapley Additive exPlanations (SHAP) Analysis

The SHAP value is derived from game theory, specifically the Shapley values. To calculate the contribution of a specific feature, a permutation of that feature is performed while keeping the others constant, assessing how this permutation impacted the model's prediction. The marginal contribution of the feature is determined by observing the difference between the prediction with the permuted feature and the average prediction without permutation [59].

SHAP analysis was executed using Python within a vs. Code development environment alongside the selected models. To evaluate the influence of each variable on model predictions, a random sample of 300 data points was employed to ensure a comprehensive assessment that captured the diverse scenarios present in the dataset. This process involved generating summary plots and calculating the relative importance of each model output. This analytical approach is particularly valuable for discerning the most influential variables and gaining insights into their contributions to the Cr(VI) and Cr(T) removal processes using CLB, as predicted by the selected models.

3. Results and Discussion

3.1. Model Performance

The results obtained from training the different LSTM and Bi-LSTM neural networks proposed for predicting and modeling the removal kinetics of Cr(VI) and Cr(T) using CLB are provided in Tables 3 and 4. The training process results were assessed using key metrics, including the MSE and R^2 for qCr(VI) and qCr(T), considering the network configuration characteristics.

Table 3. LSTM model configurations and performance.

Model	LSTM Cells	Second Hidden Layer		Epochs	MSE Train	MSE Test	R^2 qCr(VI)	R^2 qCr(T)
		Dropout	Recurrent Dropout					
1	12	0	0	307	23.77	23.87	0.9969	0.9955
2	12	0	0.3	226	44.00	41.94	0.9947	0.9916
3	12	0	0.6	297	20.26	21.01	0.9973	0.9960
4	12	0	0.9	343	28.92	29.10	0.9962	0.9945
5	12	0.3	0	136	94.52	101.82	0.9870	0.9802
6	12	0.3	0.3	186	43.54	41.74	0.9953	0.9903
7	12	0.3	0.6	155	54.25	55.30	0.9937	0.9873
8	12	0.3	0.9	92	118.36	126.02	0.9832	0.9772
9	25	0	0	193	49.98	50.60	0.9931	0.9912
10	25	0	0.3	188	27.78	27.95	0.9962	0.9951
11	25	0	0.6	228	28.70	29.71	0.9962	0.9943
12	25	0	0.9	203	31.44	31.54	0.9961	0.9935
13	25	0.3	0	92	75.61	78.40	0.9894	0.9861
14	25	0.3	0.3	125	35.74	34.84	0.9960	0.9921
15	25	0.3	0.6	158	31.52	30.57	0.9967	0.9925
16	25	0.3	0.9	121	38.91	38.09	0.9957	0.9912
17	50	0	0	213	18.77	18.67	0.9976	0.9964
18	50	0	0.3	121	28.05	28.51	0.9962	0.9947
19	50	0	0.6	126	35.58	34.22	0.9957	0.9933
20	50	0	0.9	117	28.03	27.70	0.9968	0.9939
21	50	0.3	0	79	54.07	54.16	0.9934	0.9887
22	50	0.3	0.3	117	38.73	36.56	0.9957	0.9921
23	50	0.3	0.6	85	54.56	54.96	0.9931	0.9890
24	50	0.3	0.9	115	30.46	28.25	0.9967	0.9939

Table 4. Bi-LSTM model configurations and performance.

Model	LSTM Cells	Second Hidden Layer		Epochs	MSE Train	MSE Test	R^2 qCr(VI)	R^2 qCr(T)
		Dropout	Recurrent Dropout					
25	12	0	0	188	21.98	22.06	0.9971	0.9959
26	12	0	0.3	235	30.41	30.72	0.9960	0.9942
27	12	0	0.6	199	33.83	33.87	0.9958	0.9932
28	12	0	0.9	175	34.25	34.10	0.9958	0.9929
29	12	0.3	0	153	48.89	48.61	0.9939	0.9902
30	12	0.3	0.3	147	35.33	33.70	0.9961	0.9925
31	12	0.3	0.6	112	42.84	42.52	0.9951	0.9903
32	12	0.3	0.9	156	51.35	50.86	0.9934	0.9903
33	25	0	0	120	36.37	37.36	0.9954	0.9922
34	25	0	0.3	246	12.94	12.99	0.9982	0.9977
35	25	0	0.6	147	16.98	17.14	0.9978	0.9967
36	25	0	0.9	131	23.03	24.70	0.9970	0.9949
37	25	0.3	0	155	25.09	23.72	0.9970	0.9954
38	25	0.3	0.3	139	25.52	24.27	0.9970	0.9950
39	25	0.3	0.6	160	19.59	19.11	0.9978	0.9958
40	25	0.3	0.9	118	27.96	27.05	0.9968	0.9942
41	50	0	0	144	21.00	20.76	0.9973	0.9962
42	50	0	0.3	175	19.14	19.68	0.9973	0.9967
43	50	0	0.6	152	11.45	11.82	0.9985	0.9977
44	50	0	0.9	104	21.33	20.69	0.9973	0.9960
45	50	0.3	0	83	27.72	27.87	0.9965	0.9943
46	50	0.3	0.3	116	18.81	18.38	0.9976	0.9965
47	50	0.3	0.6	123	17.16	16.27	0.9980	0.9966
48	50	0.3	0.9	132	19.31	18.73	0.9975	0.9965

The training process revealed notable variations in the number of epochs required for the different network architectures. Networks with increased complexity, characterized by a larger number of cells (25 and 50) and the incorporation of dropout techniques generally exhibited faster convergence, achieving optimal performance within a reduced number of epochs.

Furthermore, the successful use of Early Stopping demonstrated minimal differences between the MSE values of the training and testing datasets, indicating a reliable performance; this suggests that the models achieved a commendable balance, effectively fitting the training data while generalizing well to the unseen data, demonstrating a robust control mechanism and effective measures to prevent overfitting.

In the comparative analysis between the LSTM and Bi-LSTM networks with similar cell counts and dropout-related hyperparameters, the Bi-LSTM networks outperformed the LSTM networks in 21 of the 24 cases. These Bi-LSTM networks showed a superior fit with higher R^2 and lower MSE values, attributed to their advanced capabilities in capturing complex temporal dependencies through bidirectional data processing; this facilitated more effective modeling of the removal kinetics of both Cr(VI) and Cr(T).

The results of MSE and R^2 for different configurations of the LSTM networks showed significant variability. There was no clear optimal configuration regarding the number of cells, dropout, or recurrent dropout, which consistently improved the predictive capacity of the LSTM networks. This lack of consistency suggests the complexity and sensitivity of these models to different configurations, which may depend on the specific characteristics of the dataset and the nature of the chromium removal process.

The analysis of the Bi-LSTM networks showed that those with 12 cells in their hidden layers, subjected to dropout and recurrent dropout, exhibited higher MSE and lower R^2 values than configurations with model 25 without this regularization; this suggests that the introduction of dropouts, causing a deliberate loss of information during training, affected the ability of these models to fully adjust to the training data, potentially hindering their capacity to capture trends in chromium removal kinetics.

In contrast, networks with 25 and 50 cells in the hidden layers consistently displayed low MSE and high R^2 values, regardless of the dropout and recurrent dropout hyperparameters variations in the second hidden layer. Despite the deliberate loss of information through regularization, these models balanced between minimizing information loss and optimizing overall performance in capturing the complexities of chromium removal kinetics. The unique architecture of the Bi-LSTM networks with 25 and 50 cells enabled them to effectively adapt to diverse conditions and comprehend the intricate temporal relationships in the data, making them less susceptible to changes in dropout or recurrent dropout configurations.

3.2. Residuals Analysis

Tables 5 and 6 present the skewness (g_1) and kurtosis (g_2) results. Most models show negative skewness ($g_1 < 0$), indicating left-skewed residuals and a tendency to predict values higher than the experimental data. Approaching the g_1 to zero signal, the predictions resembled a normal distribution. Models 1, 8, 31, 38, 43, and 46 displayed low g_1 absolute values for qCr(VI) residuals, with p -values exceeding 0.05, indicating a satisfactory fit to symmetric residuals, comparable to a normal distribution. For the qCr(T) residuals, only models 14 and 31 had g_1 values close to zero and the p -values surpassed 0.05, suggesting symmetric residuals.

Table 5. Skewness and kurtosis residual qCr(VI) and qCr(T) analysis for LSTM models.

Model	g_1 Test qCr(VI)	g_1 Test qCr(T)	g_2 Test qCr(VI)	g_2 Test qCr(T)
1	-0.0455	-0.5549	2.2219	2.2004
2	-0.2822	-0.7838	3.3552	1.6272
3	-0.6210	-0.8913	3.9428	3.3062
4	-0.5061	-0.8911	2.8619	2.8945
5	0.4543	-0.2875	5.1015	2.1279
6	-0.1317	-0.1288	3.4923	1.3890
7	-0.6835	-0.2361	5.9065	2.5282
8	0.0472	-0.5873	4.8401	3.1900
9	0.3781	-0.6285	8.9449	3.0856
10	-0.6494	-1.0237	8.4897	5.4947
11	-0.6510	-1.6155	11.2792	8.1413
12	-0.4626	-0.9209	4.8655	2.8924
13	0.7905	-0.3265	6.2601	2.2734
14	-0.3529	-0.0508	4.6765	1.2591
15	-0.9428	-0.5088	5.3820	2.1579
16	-0.6515	-0.2183	5.2611	1.4242
17	-0.6180	-1.1433	8.0424	4.8109
18	-0.2996	-0.9488	5.7356	4.0755
19	-1.2285	-1.5220	9.3136	6.4644
20	-0.7954	-0.7378	6.3935	3.6898
21	-0.7349	-0.9557	9.0527	4.4115
22	-0.8613	-0.7387	5.1087	2.4119
23	-0.1437	-0.8407	7.0884	3.4345
24	-0.6956	-0.6577	4.4262	1.9243

Table 6. Skewness and kurtosis residual qCr(VI) and qCr(T) analysis for Bi-LSTM models.

Model	g_1 Test qCr(VI)	g_1 Test qCr(T)	g_2 Test qCr(VI)	g_2 Test qCr(T)
25	-0.6114	-1.1884	8.9240	5.5195
26	-0.1719	-1.0423	10.7687	3.9308
27	-1.2914	-1.2329	-1.2329	6.3859
28	-0.7154	-0.9585	5.3889	3.7935

Table 6. Cont.

Model	g_1 Test qCr(VI)	g_1 Test qCr(T)	g_2 Test qCr(VI)	g_2 Test qCr(T)
29	0.6143	−0.3772	6.8712	1.6905
30	−0.9709	−0.3189	4.0618	1.2682
31	−0.0268	−0.0815	3.0070	1.5803
32	0.4572	−0.7314	6.2135	1.7421
33	−1.1742	−1.4410	10.6901	6.2448
34	−0.3874	−0.4980	3.7380	3.4149
35	0.3038	−0.7046	7.0371	4.8996
36	0.8650	−0.9404	18.2947	4.5444
37	0.3468	−0.5895	4.2708	3.1144
38	−0.0769	−0.1377	2.8427	2.6233
39	−0.8726	−0.4725	5.3467	3.6010
40	−0.5931	−0.5960	4.1516	2.3497
41	−0.2309	−1.0816	7.3694	5.1950
42	0.1425	−0.9537	5.5784	3.9419
43	−0.0485	−0.4250	8.0420	4.8184
44	−1.3757	−1.9055	12.2041	10.0599
45	−0.9037	−0.8925	7.3314	3.3707
46	0.0719	−0.4060	7.6136	3.0134
47	−0.5985	−0.5312	5.2234	3.5258
48	−1.5367	−0.8365	5.9042	3.4500

The kurtosis values displayed significant variations among the models, with 47 of the 48 presenting positive kurtosis values for qCr(VI). Positive kurtosis, also known as leptokurtosis, suggests that the residual distributions have heavier tails and a more peaked shape compared to a normal distribution. This can be attributed to the presence of outliers or extreme values in the data [60]. A g_2 value of zero suggests a normal distribution. The p -values associated with kurtosis consistently remained below the 0.05 threshold, indicating that none of the models had residuals conforming to a normal distribution, due to outliers in the tails.

Figures 3 and 4 display Q-Q plots for a subset of nine selected models (1, 6, 8, 14, 31, 38, 42, 43, and 46), characterized by low absolute g_1 values ($-0.143 < g_1 < 0.143$) for qCr(VI) or qCr(T) residuals. In these plots, it was observed that the residuals closely followed the diagonal line in the central region, indicating a good fit for most models within the theoretical quantiles range of -2 to 2 . However, as expected, the high g_2 values made noticeable deviations at the extremes. For qCr(VI), the Q-Q plots revealed deviations in both the negative and positive quantiles, suggesting that the residuals can vary significantly both below and above expectations, compared to a normal distribution; this implies potential overestimation and underestimation in different scenarios.

For qCr(T), the Q-Q plots present a more pronounced deviation in the negative quantiles, suggesting that, in some cases, the models may overestimate the observed values of qCr(T) under specific conditions. These deviations are represented as data points falling below the diagonal line in the Q-Q plots, indicating that the residuals tended to have lower values than expected, compared with a normal distribution in that region.

Additional analyses were performed to clarify whether the observed kurtosis and Q-Q plot trends are due to outliers or extreme values associated with the inherent complexity of the chromium removal process. The conditions consistently producing the highest residual values in the nine selected models were identified. These corresponded to two kinetics obtained at a Co of 805 mg L^{-1} , but at different temperatures (35 and $45 \text{ }^\circ\text{C}$). These were the conditions in which the highest removal capacities were reached from the first hours of the process for the entire dataset because temperature increases the initial removal rate of Cr(VI) and Cr(T) using CLB [31].

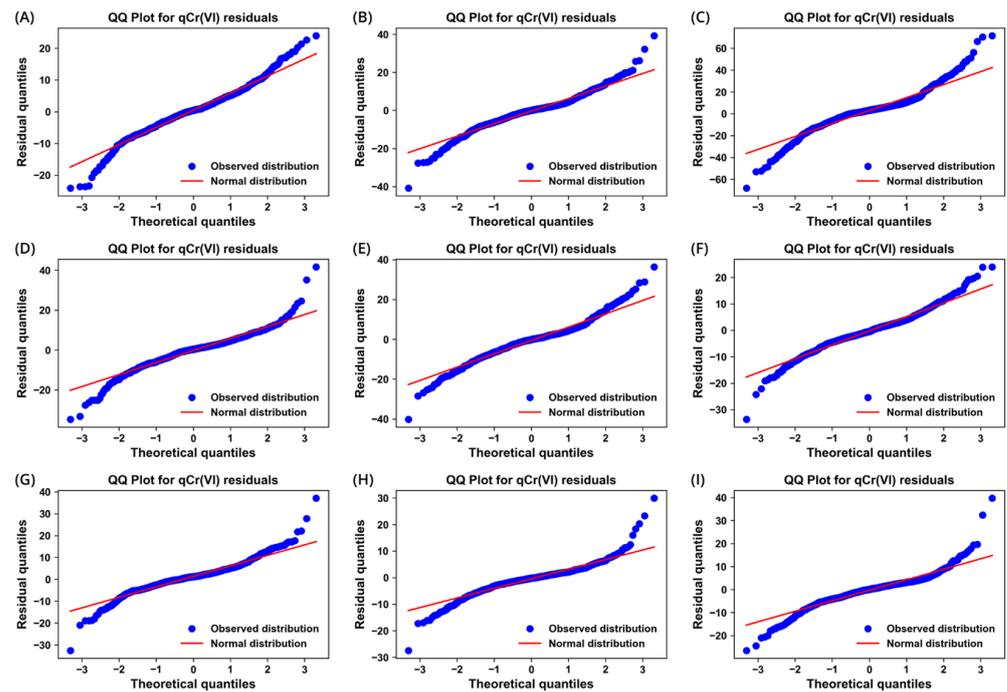


Figure 3. Quantile–quantile plots for qCr(VI) residuals for LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I).

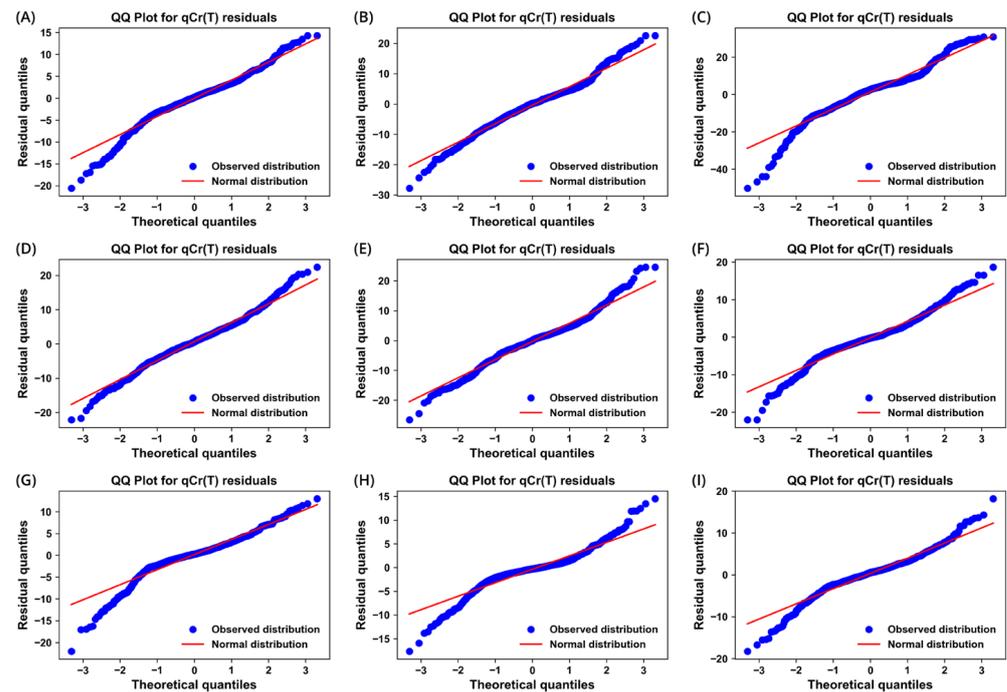


Figure 4. Quantile–quantile plots for qCr(T) residuals for LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I).

Once the conditions were identified, additional experiments were conducted by training the nine selected models without considering these seemingly outlying data points, and the g_1 and g_2 values were calculated for the residuals generated in the test set. The results are presented in Table 7. The results showed g_1 values further from zero, except for model 42. The g_2 values were also higher in most cases; some models showed improvement in this aspect, but not enough to be considered statistically normal, with respect to the tails. This suggests

that including these data points, rather than hindering model performance, enhances their ability to capture the variability and complex relationships present in the process.

Table 7. Skewness and kurtosis residual qCr(VI) and qCr(T) analysis for LSTM and Bi-LSTM selected models without seemingly outlying data.

Model	g_1 Test qCr(VI)	g_1 Test qCr(T)	g_2 Test qCr(VI)	g_2 Test qCr(T)
1	−0.3647	−1.0891	3.9312	3.6341
6	1.5624	−0.3739	10.9953	3.7768
8	1.9569	−0.6549	14.7837	2.9181
14	−0.7895	−0.9101	4.2179	2.4109
31	−0.7046	−0.7783	4.3127	2.6463
38	−0.7653	−1.006	5.3230	4.1660
42	0.0421	−0.8881	8.0993	6.8762
43	−1.1964	−1.5219	5.5016	7.2949
46	−0.7880	−1.2769	4.5748	3.6477

It is important to highlight that, first, the seemingly outlying data points represent valid but extreme experimental conditions that reflect the response of the chromium removal system under specific circumstances. Second, including these data increases the diversity and size of the training set, benefiting the models' learning. A more diverse and representative dataset allows the models to better capture the subtleties and nonlinear relationships present in the system, especially when working with deep learning techniques like LSTM and Bi-LSTM networks, which benefit from large amounts of data to learn meaningful and generalizable representations [61].

Regarding the experimental design and data collection, it is crucial to emphasize the importance of a well-planned approach that covers a wide range of conditions and ensures representative data. In future studies, expanding the experimental design to include more extreme conditions, diverse scenarios, and a more comprehensive range of water chemistry parameters, such as the presence of co-existing ions or contaminants in the solution, will enhance the robustness and generalizability of the developed models [62]. This comprehensive data collection will allow for establishing the reliability and applicability of the models in practical metal removal systems, enabling their effective translation to real-world environmental engineering applications.

3.3. Post-Training Validation of Selected Models

Figures 5 and 6 show the response surface plots the nine selected models predicted, illustrating their relationships with pH and time. These predictions were made under the same temperature, pH, and initial Cr(VI) concentration conditions as those reported previously by Netzahuatl-Muñoz et al. [32]. This prior research served as a valuable reference for assessing how well each model replicated the expected behaviors.

All nine models generally generated response surface plots that conformed to the most anticipated characteristics. A noteworthy discovery is that all the models accurately predicted values for qCr(VI) and qCr(T) well within the feasible range; specifically, less than 100 mg g^{-1} . They also demonstrated proficiency in predicting variations in the initial rates of removal capacity based on pH. The response surfaces exhibited lower adsorption capacity during the initial hours at high pH values than at low pH values. The model's success in capturing this relationship validates its ability to represent the expected behavior concerning pH, which is vital for comprehending how changes in conditions influence the kinetics of chromium removal.

However, regarding qCr(VI), most models predicted negative slopes in the later stages of kinetics; this contradicted the well-documented irreversible nature of Cr(VI) reduction using biomaterials under acidic conditions [63]. This issue could be related to the provided

data, as it is incomplete and not all data points cover the same time, particularly for extended contact times.

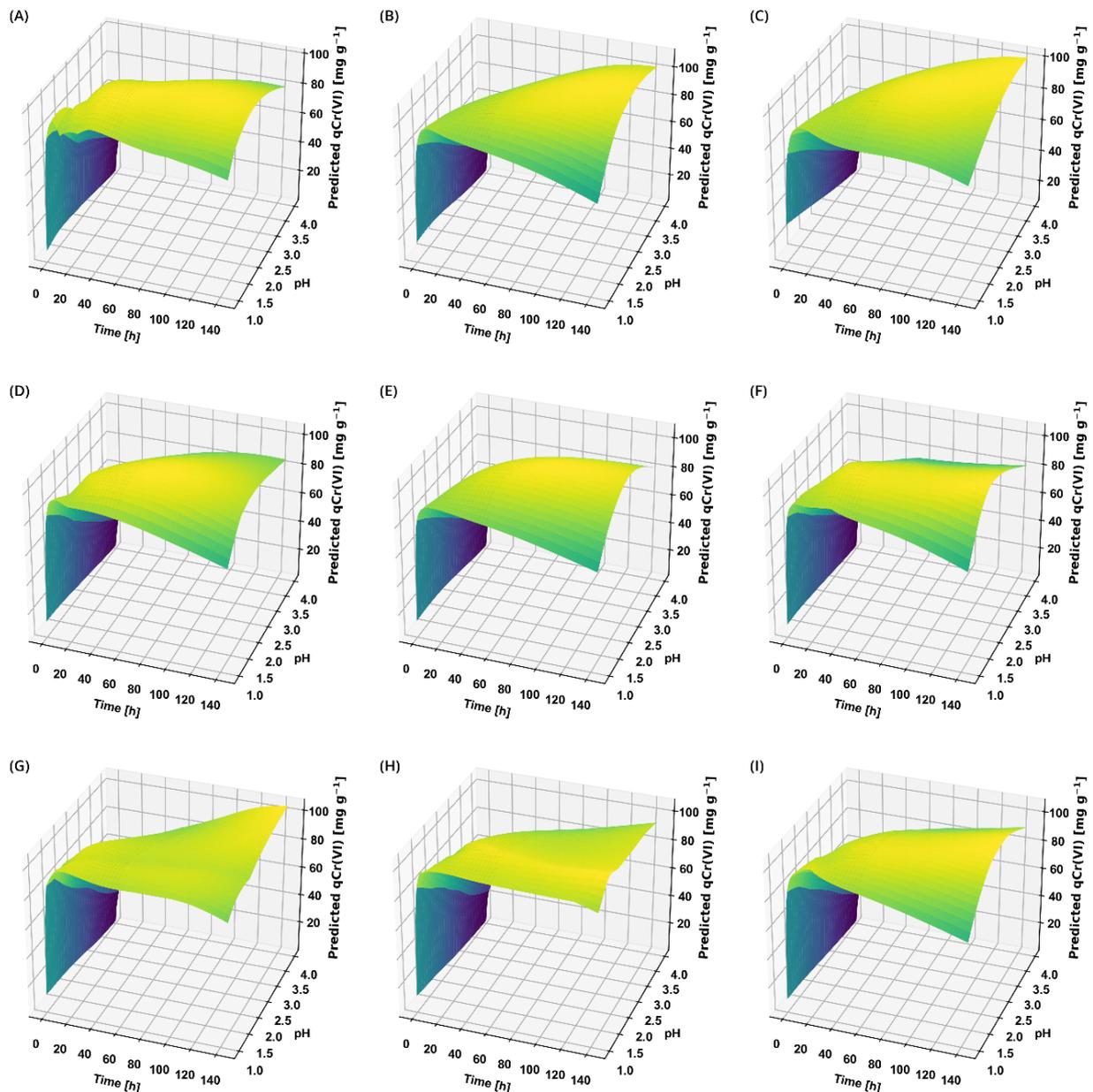


Figure 5. Surface response for $q_{Cr(VI)}$ as a function of pH and time for LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I). Conditions: $C_o = 100 \text{ mg L}^{-1}$, $SPS = 420 \text{ }\mu\text{m}$, $LSP = 500 \text{ }\mu\text{m}$, $T = 28 \text{ }^\circ\text{C}$.

In contrast, for $q_{Cr(T)}$, a consistent transition from a positive to negative slope over time, particularly at lower pH values, was observed for all response surfaces. This change was expected and was attributed to the desorption of chromium in its trivalent form.

Furthermore, some models displayed slight irregularities in their response surface plots, including changes in slope, rather than adhering to the anticipated systematic trends in the behavior of the studied phenomenon. These peculiar behaviors appear to occur in specific regions of the experimental space and may be attributed to potential overfitting. In these cases, the models likely captured certain peculiarities or random fluctuations in the training dataset that did not precisely reflect the actual phenomenon. Models 1 and 14 exhibited more pronounced overfitting tendencies, as shown in Figures 5A,D and 6A,D.

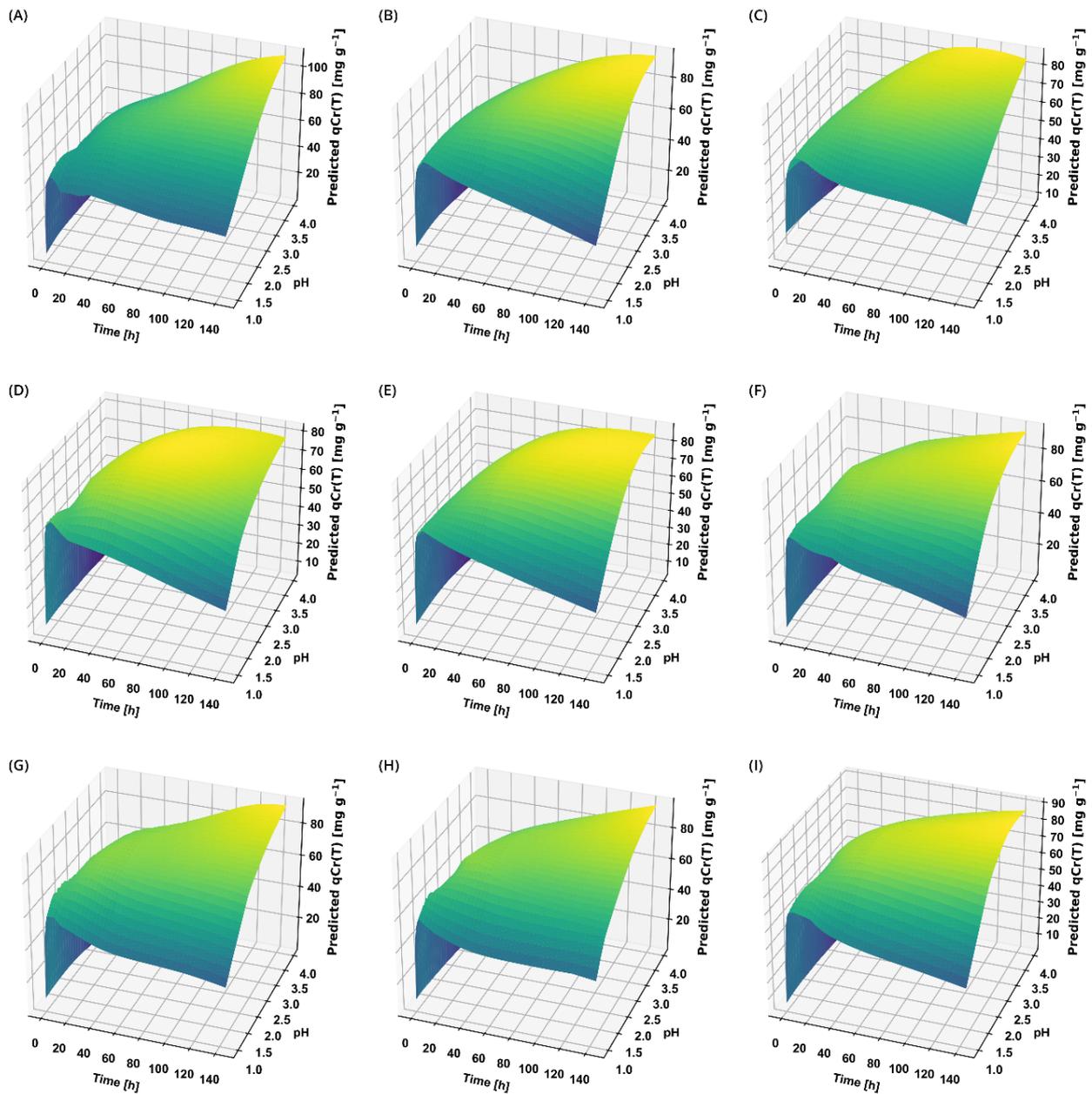


Figure 6. Surface response for $q_{Cr}(T)$ as a function of pH and time for LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I). Conditions: $C_o = 100 \text{ mg L}^{-1}$, $SPS = 420 \text{ }\mu\text{m}$, $LSP = 500 \text{ }\mu\text{m}$, $T = 28 \text{ }^\circ\text{C}$.

In Figures 7–9, the nine selected LSTM and Bi-LSTM models present predictions for the three kinetics not included in the training data. Overall, the models showed a robust fit, with only slight variations in the predictions of the specific models. For instance, in kinetic 1, model 8 (Figure 7C) tended to slightly underestimate $q_{Cr}(T)$ at longer contact times. Additionally, inconsistent changes in slope were observed in models 1, 14, and 46, particularly in the predictions for kinetic 1, suggesting a potential overfitting concern. Moreover, for kinetic 3, model 42 (Figure 9G) tended to underestimate $q_{Cr}(VI)$, whereas models 8, 14, 31, and 42 (Figure 9C–E,G) tended to underestimate $q_{Cr}(T)$. In contrast, model 6 was the only model that predicted $q_{Cr}(VI)$ values higher than the expected maximum limit of 102 mg g^{-1} .

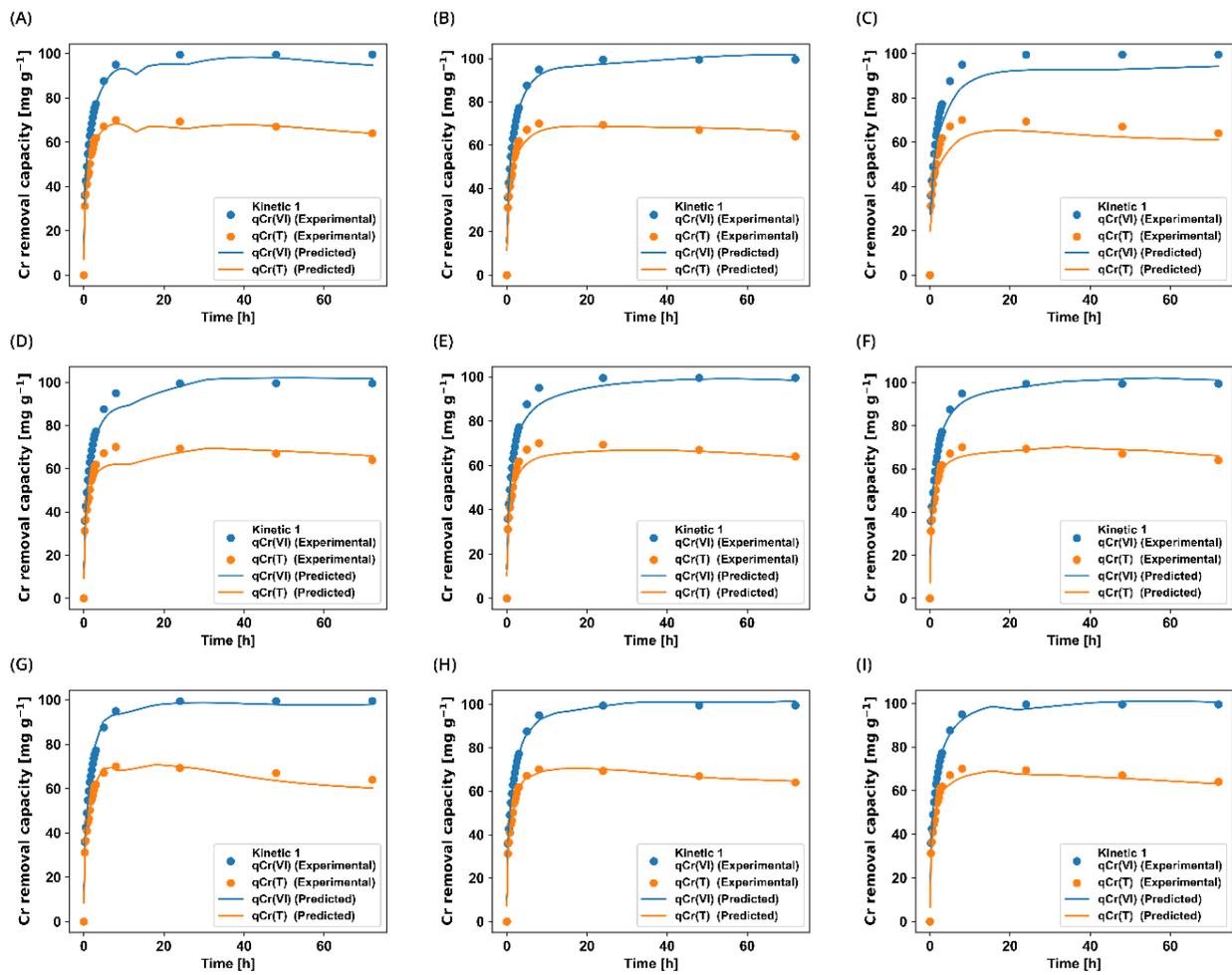


Figure 7. Experimental and predicted capacity removal kinetics of Cr(VI) and Cr(T) using LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I). Conditions: pH = 1.5, T = 28 °C, Co = 99.5 mg L⁻¹, SPS = 420 μm, LPS = 500 μm.

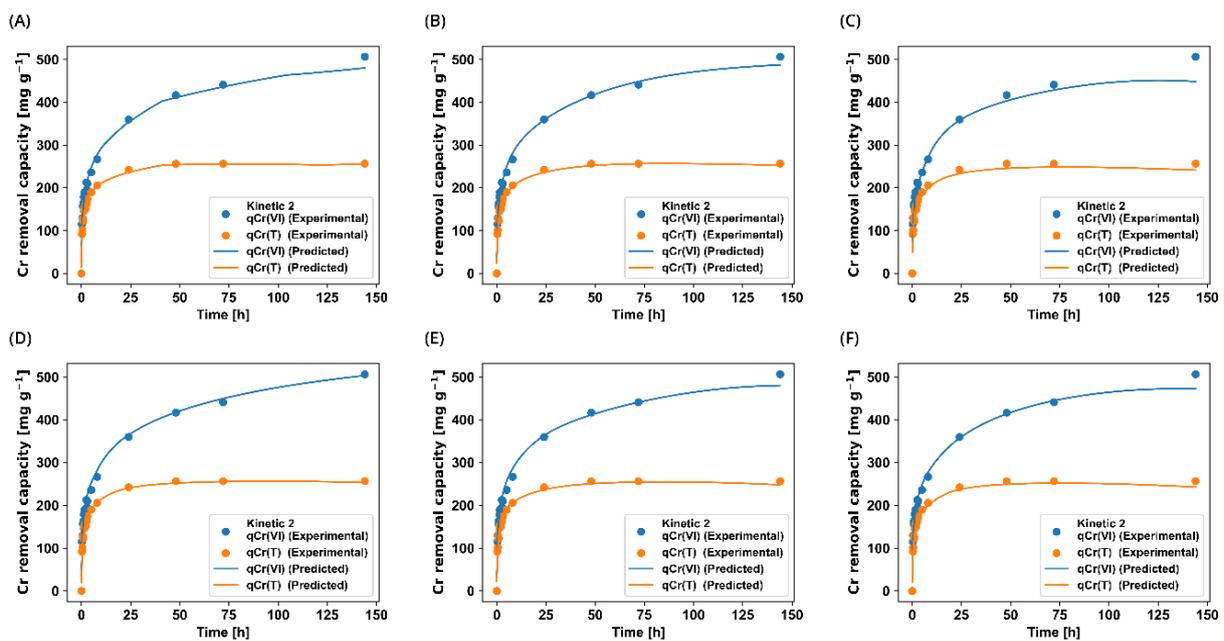


Figure 8. Cont.

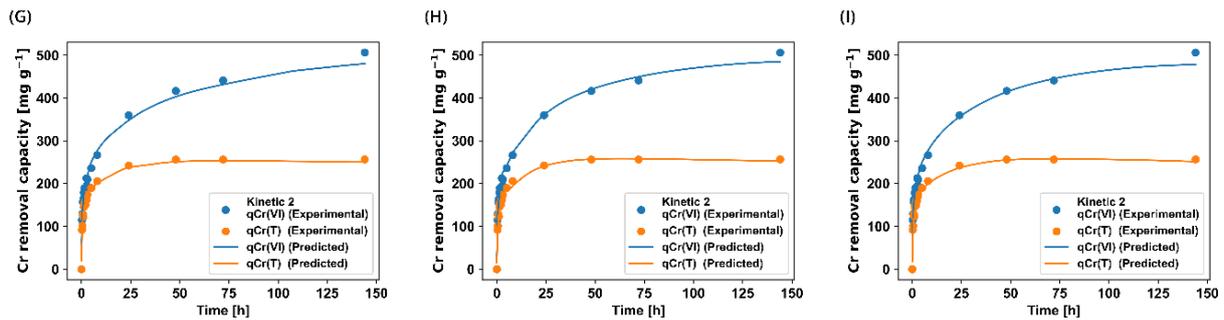


Figure 8. Experimental and predicted capacity removal kinetics of Cr(VI) and Cr(T) using LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I). Conditions: pH = 1.5, T = 28 °C, Co = 805 mg L⁻¹, SPS = 420 μm, LPS = 500 μm.

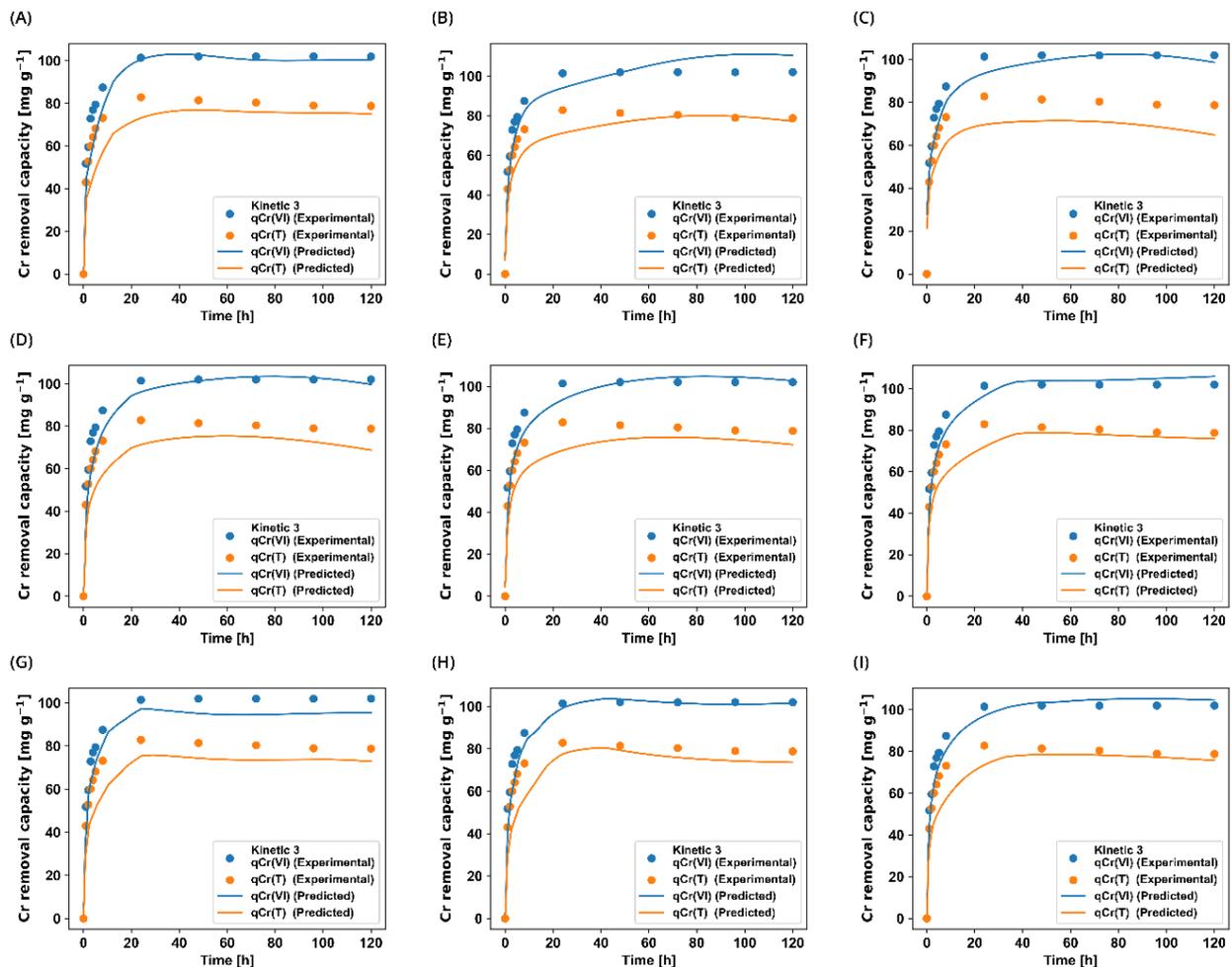


Figure 9. Experimental and predicted capacity removal kinetics of Cr(VI) and Cr(T) using LSTM and Bi-LSTM models: 1 (A), 6 (B), 8 (C), 14 (D), 31 (E), 38 (F), 42 (G), 43 (H), and 46 (I). Conditions: pH = 2.0, T = 28 °C, Co = 98.8 mg L⁻¹, SPS = 420 μm, LPS = 500 μm.

These discrepancies could arise because the training data were not specifically tailored for this analysis and originated from univariate studies; thus, they were not randomly distributed in the sample space. The lack of a balanced and representative distribution in the input variable space may introduce challenges in capturing the full complexity of the underlying phenomena. Despite these challenges, the models exhibited a strong ability to predict trends in these unseen kinetics. The models that performed best for the three

post-training validation kinetics were Bi-LSTM models 38, 43, and 46, with g_1 values close to zero, especially for qCr(VI), and high R^2 values for both qCr(VI) and qCr(T).

A 5-fold cross-validation technique was used to assess the precision and stability of the selected LSTM and Bi-LSTM models in predicting chromium removal capacity. Table 8 reports CoV calculated for R^2 , RMSE, and MAE metrics to assess the consistency of model performance across different data subsets.

Table 8. Coefficient of variation for k-fold cross-validation metrics.

Model	CoV R^2 (%)		CoV RMSE (%)		CoV MAE (%)	
	qCr(VI)	qCr(T)	qCr(VI)	qCr(T)	qCr(VI)	qCr(T)
1	0.2904	0.4501	27.22	28.10	21.56	25.17
6	0.5579	0.2700	20.23	7.72	15.12	7.44
8	1.1463	0.4794	37.88	16.93	29.31	16.12
14	0.2063	0.4036	17.89	21.54	17.64	21.40
31	0.1889	0.2553	19.88	15.93	16.51	14.62
38	0.0265	0.0398	5.62	5.01	6.11	4.54
42	0.1120	0.1625	19.75	20.67	16.85	18.64
43	0.0665	0.0736	10.58	7.71	8.06	7.25
46	0.0671	0.0992	15.35	16.15	16.78	16.67

The calculated CoV values for the R^2 metric were notably low, below 1.2%. Low CoV values for R^2 indicate that the models are highly stable. Stability, in this context, refers to the consistency of a model's predictions when evaluated on different subsets of data. Notably, the Bi-LSTM models 38, 43, and 46 demonstrated outstanding stability, with R^2 CoV values below 0.1% for qCr(VI) and qCr(T). It is worth emphasizing that the R^2 values were high in all tests. For the LSTM models, the R^2 values were higher than 0.9667 and 0.9800 for qCr(VI) and qCr(T), respectively, while for the Bi-LSTM models, the R^2 values were higher than 0.9752 for qCr(VI) and 0.9625 for qCr(T). The combination of low CoV values for R^2 and high R^2 values across all folds provides strong evidence for the stability and reliability of the developed models.

Precision relates to how close the model's predictions are to the actual values. A model is considered more precise if it has lower CoV values for RMSE and MAE, with values below 10% suggesting good precision [64]. Bi-LSTM model 38 exhibited the best precision for the qCr(VI) target, with the lowest CoV values for RMSE (5.62%) and MAE (6.11%). For the qCr(T) target, LSTM model 6 and Bi-LSTM models 38 and 43 showed higher precision regarding RMSE and MAE, with CoV values below 8%.

A robust model has few performance variations when faced with data changes. RMSE is sensitive to outliers, so a low CoV in RMSE suggests robustness. Bi-LSTM models 38 and 43 demonstrated the most robust performance among the selected models. Bi-LSTM model 38 exhibited a CoV of 5.62% and 5.01% for qCr(VI) and qCr(T), respectively, while Bi-LSTM model 43 showed a CoV of 10.58% for qCr(VI) and 7.71% for qCr(T). In contrast, most of the remaining models presented CoV values for RMSE above 15% and this observation is consistent with the findings from the residual analysis, which revealed heavy-tailed, left-skewed distributions, indicating the occasional overprediction of extreme values.

The robustness of Bi-LSTM models 38 and 43 is further supported by their ability to generate accurate response surfaces and predict unseen kinetic data. These models demonstrate a remarkable capacity to capture the complex relationships in the data and generalize well to new scenarios. Their consistent performance across various evaluation metrics and validation techniques underscores their reliability and potential for practical application in chromium removal using biosorption techniques.

3.4. SHAP Analysis

SHAP studies play a crucial role in interpreting machine learning models by providing a detailed understanding of how each feature or variable influences model predictions. This

methodology is essential for identifying the most relevant variables and understanding their contributions to the prediction process. SHAP analysis was performed on nine selected models.

Figure 10 presents the SHAP values for 300 random samples used in the analysis of LSTM model 1. The baseline, represented by the average values of $q_{Cr(VI)}$ and $q_{Cr(T)}$, served as a reference for understanding the deviations caused by changes in the variable values. Notably, Co, time, and T emerged as the three variables with the most significant impacts on predicting $q_{Cr(VI)}$, whereas, for $q_{Cr(T)}$, the influential variables were Co, time, and pH. Higher Co, time, and T values led to increased predictions for both $q_{Cr(VI)}$ and $q_{Cr(T)}$, whereas elevated pH values had an inverse effect, reducing the model outputs. The impact of these principal variables was more pronounced for $q_{Cr(VI)}$ than for $q_{Cr(T)}$, consistent with the behavior of the chromium removal process in the chosen material. In this context, $q_{Cr(T)}$ is constrained by the available sites for chromium biosorption, which have been reported to be significantly lower than its capacity for Cr(VI) reduction.

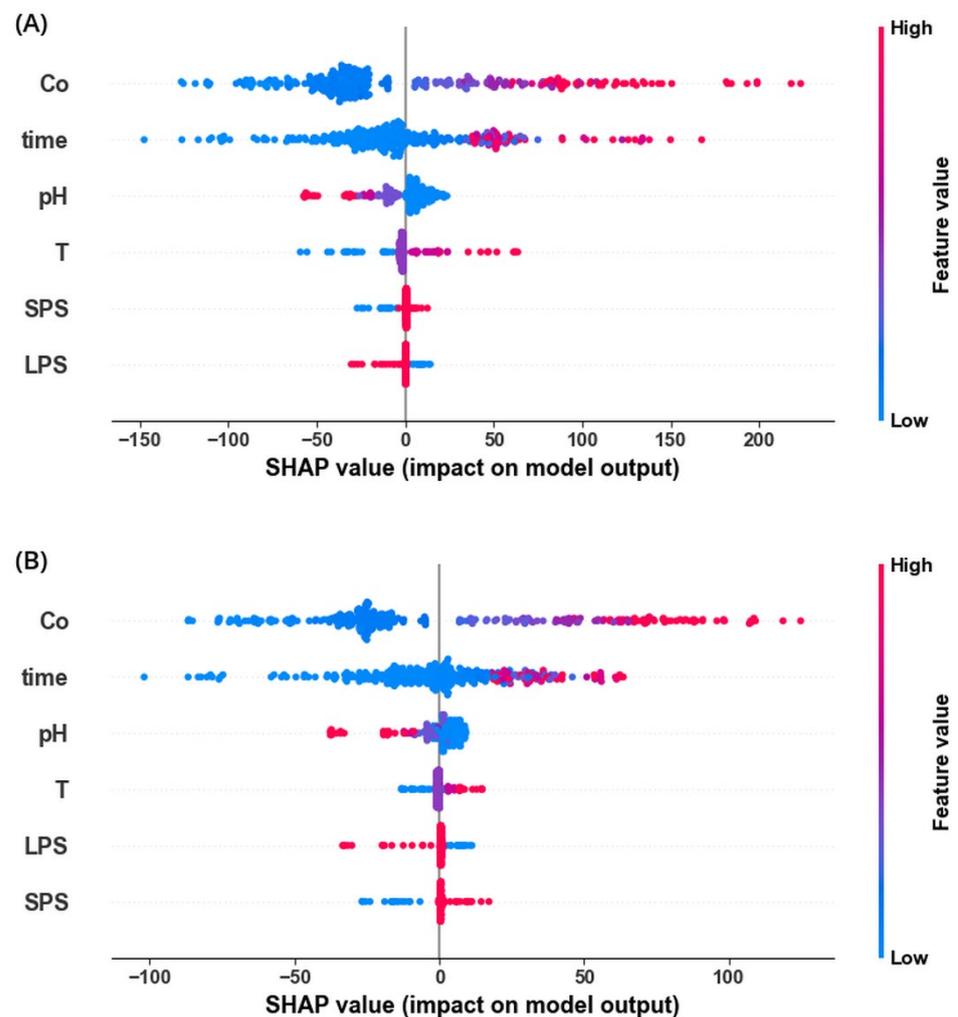


Figure 10. SHAP Analysis: Variable impact on $q_{Cr(VI)}$ (A) and $q_{Cr(T)}$. (B) Predictions for model 1. Baseline: Average $q_{Cr(VI)}$ and $q_{Cr(T)}$ values.

Consequently, under saturation conditions, both Co and time cease to influence $q_{Cr(T)}$ [31]. Regarding the impact of SPS and LPS, it was observed that they had minor and similar effects for both $q_{Cr(VI)}$ and $q_{Cr(T)}$ within the employed particle size range. Most SHAP values for these characteristics were concentrated around the baseline.

Table 9 reports the SHAP relative importance values, providing insight into the average impact of each variable on the model outputs. As observed in model 1, the key drivers influencing the predictions across all nine models for $q_{Cr(VI)}$ and $q_{Cr(T)}$ were Co, time, pH,

and T, showing marginal differences in their calculated values. For instance, the most crucial variable, Co, exhibited relative importance values ranging from 53.45 mg g⁻¹ to 60.65 mg g⁻¹ for qCr(VI), while its contributions to qCr(T) ranged from 39.25 mg g⁻¹ to 44.06 mg g⁻¹.

Table 9. Relative importance of SHAP features on model outputs.

Model	Feature Relative Importance on qCr(VI) (mg g ⁻¹)					
	Time	pH	T	Co	SPS	LPS
1	34.91	11.84	6.83	56.32	1.53	1.43
6	30.69	11.04	8.13	55.41	1.44	1.51
8	33.24	10.44	5.69	54.96	0.61	0.25
14	36.35	12.64	6.95	60.65	2.09	1.31
31	34.41	7.86	6.68	57.59	1.60	2.26
38	36.42	11.09	6.97	59.47	1.12	1.19
42	32.99	10.33	5.06	57.59	1.34	1.30
43	34.82	12.53	6.27	53.45	2.05	0.70
46	39.05	11.80	5.48	59.34	0.97	2.02
Model	Feature Relative Importance on qCr(T) (mg g ⁻¹)					
	Time	pH	T	Co	SPS	LPS
1	21.14	5.81	1.88	41.13	1.79	1.81
6	18.70	5.15	2.74	42.43	1.13	1.23
8	19.43	5.10	2.12	42.67	0.32	0.17
14	20.57	6.34	2.01	44.06	1.66	0.92
31	20.86	3.72	2.23	43.54	1.42	1.70
38	20.86	5.05	2.06	43.50	1.34	0.83
42	19.52	5.39	1.65	43.49	1.62	0.83
43	20.36	6.12	1.76	39.25	2.01	1.04
46	22.50	5.20	1.64	43.97	1.27	1.37

Consistent with model 1, the contributions of the LPS and SPS to qCr(VI) and qCr(T) were minimal; this aligns with the findings of similar phenomena, where the particle size has a more pronounced effect on the process rate than on the chromium removal capacity of a lignocellulosic material [65]. The SHAP analysis reflected this in the models, assigning higher importance to time than to SPS and LPS.

While there may be similarities in many aspects of the SHAP analysis, the observed differences in the relative importance of variables among the models suggest that they are not entirely equivalent in terms of how they interpret and utilize these variables to make predictions, as evidenced by the predictions for unseen validation kinetics. This highlights the need for a nuanced understanding of the strengths and limitations of each modeling approach and their complementary roles in advancing the field.

Integrating ML and DL techniques in metal biosorption studies offers multifaceted benefits. From an economic perspective, these advanced modeling approaches optimize experimental design, reducing costs associated with extensive laboratory experiment, while enhancing our understanding of the intricate biosorption processes [11,33]. The optimization of biosorption processes through ML and DL can generate significant savings in operating and capital costs for industries and wastewater treatment facilities [66,67]. Furthermore, these techniques can support the transition towards a circular economy by optimizing recycling systems, predicting demand for recovered materials, and facilitating industrial symbiosis [68].

From a social standpoint, using ML and DL to accurately predict contaminant removal can reduce human exposure to toxic substances, prevent diseases, and promote equity in access to clean and safe water [69,70]. The application of ML and DL aligns seamlessly with the global shift towards sustainable practices [61]. The remarkable ability of these techniques to extract insights from complex datasets positions researchers at the forefront

of innovation, fostering a culture of continuous improvement and adaptive strategies in the pursuit of sustainable environmental solutions [46].

However, it is essential to acknowledge some limitations and opportunities for future research. A more comprehensive evaluation of the adaptability and universality of the models under different real-world environmental conditions is required [47]. Moreover, the scalability of the models and their potential application in removing various contaminants besides chromium require further exploration [71]. Integrating multimodal data, such as microscopic and spectroscopic images [72], along with kinetic and equilibrium data, could provide a more complete understanding of the biosorption mechanisms and further improve the accuracy of the predictive models.

4. Conclusions

The trained LSTM and Bi-LSTM models successfully captured the complex mechanisms of Cr(VI) reduction, Cr(III) desorption, and simultaneous biosorption, which characterize the kinetic phenomena. The combined use of Early Stopping, dropout, and recurrent dropout methodologies effectively prevented overfitting and maintained a commendable equilibrium between fitting the training data and generalizing it to unseen data.

The Bi-LSTM models set up with 25 and 50 cells in the hidden layers proved highly effective for modeling and predicting the complexities of qCr(VI) and qCr(T) removal kinetics. Their bidirectional architecture and resilience to dropouts improve their ability to learn over time and make predictions, all accomplished within short training periods.

Residual analysis employing skewness, kurtosis, and Q-Q plots revealed heavy-tailed, left-skewed distributions, indicating the occasional overprediction of extreme values. However, this behavior is likely tied to the inherent complexities of the process and data limitations.

Validation with unseen kinetic data and pH-time response surfaces demonstrated the models' proficiency in replicating the anticipated trends of initial rate changes with pH and irreversible Cr(VI) reduction. Limitations related to overfitting in specific regions suggest opportunities to enhance diversity during training.

According to the k-fold cross-validation, the Bi-LSTM models exhibited the best characteristics in terms of stability, precision, and robustness. Specifically, models 38 and 43 demonstrated remarkable consistency, as evidenced by their low coefficients of variation of R^2 , RMSE, and MAE. These findings highlight the suitability of Bi-LSTM models for reliable and accurate predictions of chromium removal capacity.

The SHAP method underscores the dominance of the initial Cr(VI) concentration and time as drivers of model outputs. This analysis provides interpretable insights into variable contributions, enhances reliability, and reveals the differences in how models weigh inputs for predictions.

The inclusion of extreme conditions and diverse scenarios in the training data improved the models' robustness and generalization capability. This information can guide the planning of future experiments, ensuring that a wider range of operating conditions is covered and representative data of the inherent variability in the biosorption process is generated.

This work highlights the potential of recurrent deep learning, showcasing the LSTM and Bi-LSTM models as powerful tools for deriving kinetic predictors from experimental data. Their remarkable ability to capture complex relationships and generalize to new conditions holds great promise for advancing environmental engineering, especially for removing toxic compounds from polluted waters. Developing these promising deep-learning models necessitates a rigorous approach that prioritizes data quality, carefully selects modeling techniques aligned with research objectives, and thoroughly evaluates and validates models using diverse techniques. These practices are essential to ensure the developed models' generalization capabilities, robustness, and enhanced applicability in real-world conditions.

As the world faces increasingly complex environmental challenges, adopting advanced modeling approaches will be crucial in developing sustainable and resilient solutions. Integrating deep learning into metal biosorption studies drives innovation in the field and

contributes to global efforts to safeguard aquatic ecosystems, ensure human health, and promote sustainable development.

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

Funding: This research was funded by the Instituto Politécnico Nacional (IPN), Secretaría de Investigación y Posgrado, grant number SIP20242007.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are within the paper.

Acknowledgments: E.C.-U. holds grants from EDI-IPN, COFAA-IPN, and SNI-CONAHCYT.

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

References

1. Rockström, J.; Steffen, W.; Noone, K.; Persson, A.; Chapin, F.S.; Lambin, E.F.; Lenton, T.M.; Scheffer, M.; Folke, C.; Schellnhuber, H.J.; et al. A safe operating space for humanity. *Nature* **2009**, *461*, 472–475. [[CrossRef](#)] [[PubMed](#)]
2. Naidu, R.; Biswas, B.; Willett, I.R.; Cribb, J.; Singh, B.K.; Nathanail, C.P.; Coulon, F.; Semple, K.T.; Jones, K.C.; Barclay, A.; et al. Chemical pollution: A growing peril and potential catastrophic risk to humanity. *Environ. Int.* **2021**, *156*, 106616. [[CrossRef](#)] [[PubMed](#)]
3. Saha, R.; Nandi, R.; Saha, B. Sources and toxicity of hexavalent chromium. *J. Coord. Chem.* **2011**, *64*, 1782–1806. [[CrossRef](#)]
4. Singh, S.; Kumar Naik, T.S.S.; Chauhan, V.; Shehata, N.; Kaur, H.; Dhanjal, D.S.; Marcelino, L.A.; Bhati, S.; Subramanian, S.; Singh, J.; et al. Ecological effects, remediation, distribution, and sensing techniques of chromium. *Chemosphere* **2022**, *307*, 135804. [[CrossRef](#)] [[PubMed](#)]
5. Kormoker, T.; Proshad, R.; Islam, M.S.; Tusher, T.R.; Uddin, M.; Khadka, S.; Chandra, K.; Sayeed, A. Presence of toxic metals in rice with human health hazards in Tangail district of Bangladesh. *Int. J. Environ. Health Res.* **2022**, *32*, 40–60. [[CrossRef](#)] [[PubMed](#)]
6. IARC. *A Review of Human Carcinogens—Part C: Arsenic, Metals, Fibres and Dusts*; IARC Monographs: Lyon, France, 2012; Volume 100, pp. 147–168.
7. DesMarais, T.L.; Costa, M. Mechanisms of Chromium-Induced Toxicity. *Curr. Opin. Toxicol.* **2019**, *14*, 1–7. [[CrossRef](#)] [[PubMed](#)]
8. Srivastava, S.; Agrawal, S.B.; Mondal, M.K. A review on progress of heavy metal removal using adsorbents of microbial and plant origin. *Environ. Sci. Pollut. Res.* **2015**, *22*, 15386–15415. [[CrossRef](#)] [[PubMed](#)]
9. Chen, X.; Hossain, M.F.; Duan, C.; Lu, J.; Tsang, Y.F.; Islam, M.S.; Zhou, Y. Isotherm models for adsorption of heavy metals from water—A review. *Chemosphere* **2022**, *307*, 135545. [[CrossRef](#)] [[PubMed](#)]
10. Febrianto, J.; Kosasih, A.N.; Sunarso, J.; Ju, Y.H.; Indraswati, N.; Ismadji, S. Equilibrium and kinetic studies in adsorption of heavy metals using biosorbent: A summary of recent studies. *J. Hazard. Mater.* **2009**, *162*, 616–645. [[CrossRef](#)] [[PubMed](#)]
11. Nathan, R.J.; Jain, A.K.; Rosengren, R.J. Biosorption of heavy metals from water: Mechanism, critical evaluation and translatability of methodology. *Environ. Technol. Rev.* **2022**, *11*, 91–117. [[CrossRef](#)]
12. Razzak, S.A.; Faruque, M.O.; Alsheikh, Z.; Alsheikhmohamad, L.; Alkuroud, D.; Alfayez, A.; Hossain, S.M.Z.; Hossain, M.M. A comprehensive review on conventional and biological-driven heavy metals removal from industrial wastewater. *Environ. Adv.* **2022**, *7*, 100168. [[CrossRef](#)]
13. Alam, G.; Ihsanullah, I.; Naushad, M.; Sillanpää, M. Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes: Recent advances and prospects. *Chem. Eng. J.* **2022**, *427*, 130011. [[CrossRef](#)]
14. Skrobek, D.; Krzywanski, J.; Sosnowski, M.; Kulakowska, A.; Zylka, A.; Grabowska, K.; Ciesielska, K.; Nowak, W. Implementation of deep learning methods in prediction of adsorption processes. *Adv. Eng. Softw.* **2022**, *173*, 103190. [[CrossRef](#)]
15. Taoufik, N.; Boumya, W.; Achak, M.; Chennouk, H.; Dewil, R.; Barka, N. The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Sci. Total Environ.* **2022**, *807*, 150554. [[CrossRef](#)] [[PubMed](#)]
16. Zhu, X.; Wang, X.; Ok, Y.S. The application of machine learning methods for prediction of metal sorption onto biochars. *J. Hazard. Mater.* **2019**, *378*, 120727. [[CrossRef](#)] [[PubMed](#)]
17. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)] [[PubMed](#)]
18. Belle, V.; Papantonis, I. Principles and Practice of Explainable Machine Learning. *Front. Big Data* **2021**, *4*, 688969. [[CrossRef](#)] [[PubMed](#)]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]

20. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [[CrossRef](#)] [[PubMed](#)]
21. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
22. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. In Proceedings of the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99 (Conf. Publ. No. 470), Edinburgh, UK, 7–10 September 1999; Volume 2, pp. 850–855.
23. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [[CrossRef](#)]
24. Weerakody, P.B.; Wong, K.W.; Wang, G.; Ela, W. A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing* **2021**, *441*, 161–178. [[CrossRef](#)]
25. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [[CrossRef](#)]
26. Lipton, Z.C.; Berkowitz, J.; Elkan, C.A. Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv* **2015**, arXiv:1506.00019v4.
27. Faridi, I.K.; Tsotsas, E.; Heineken, W.; Koegler, M.; Kharaghani, A. Spatio-temporal prediction of temperature in fluidized bed biomass gasifier using dynamic recurrent neural network method. *Appl. Therm. Eng.* **2023**, *219*, 119334. [[CrossRef](#)]
28. Ozcan, A.; Kasif, A.; Sezgin, I.V.; Catal, C.; Sanwal, M.; Merdun, H. Deep learning-based modelling of pyrolysis. *Clust. Comput.* **2024**, *27*, 1089–1108. [[CrossRef](#)]
29. Müller, A.T.; Hiss, J.A.; Schneider, G. Recurrent Neural Network Model for Constructive Peptide Design. *J. Chem. Inf. Model.* **2018**, *58*, 472–479. [[CrossRef](#)] [[PubMed](#)]
30. Rabby, M.F.; Tu, Y.; Hossen, M.I.; Lee, I.; Maida, A.S.; Hei, X. Stacked LSTM based deep recurrent neural network with kalman smoothing for blood glucose prediction. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 101. [[CrossRef](#)] [[PubMed](#)]
31. Netzahuatl-Muñoz, A.R.; Cristiani-Urbina, M.D.C.; Cristiani-Urbina, E. Chromium Biosorption from Cr(VI) Aqueous Solutions by *Cupressus lusitanica* Bark: Kinetics, Equilibrium and Thermodynamic Studies. *PLoS ONE* **2015**, *10*, e0137086.
32. Netzahuatl-Muñoz, A.R.; Guillén-Jiménez, F.D.M.; Chávez-Gómez, B.; Villegas-Garrido, T.L.; Cristiani-Urbina, E. Kinetic Study of the Effect of pH on Hexavalent and Trivalent Chromium Removal from Aqueous Solution by *Cupressus lusitanica* Bark. *Water Air Soil Pollut.* **2012**, *223*, 625–641. [[CrossRef](#)]
33. Islam, M.d.A.; Angove, M.J.; Morton, D.W. Recent innovative research on chromium (VI) adsorption mechanism. *Environ. Nanotechnol. Monit. Manag.* **2019**, *12*, 100267. [[CrossRef](#)]
34. Rajapaksha, A.U.; Selvasembian, R.; Ashiq, A.; Gunarathne, V.; Ekanayake, A.; Perera, V.O.; Wijesekera, H.; Mia, S.; Ahmad, M.; Vithanage, M.; et al. A systematic review on adsorptive removal of hexavalent chromium from aqueous solutions: Recent advances. *Sci. Total Environ.* **2022**, *809*, 152055. [[CrossRef](#)] [[PubMed](#)]
35. Anupam, K.; Dutta, S.; Bhattacharjee, C.; Datta, S. Artificial neural network modelling for removal of chromium (VI) from wastewater using physisorption onto powdered activated carbon. *Desalin. Water Treat.* **2016**, *57*, 3632–3641. [[CrossRef](#)]
36. Singha, B.; Bar, N.; Das, S.K. The use of artificial neural networks (ANN) for modeling of adsorption of Cr(VI) ions. *Desalin. Water Treat.* **2014**, *52*, 415–425. [[CrossRef](#)]
37. 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](#)]
38. Saber, W.I.A.; El-Naggar, N.E.A.; El-Hersh, M.S.; El-khateeb, A.Y.; Elsayed, A.; Eldadamony, N.M.; Ghoniem, A.A. Rotatable central composite design versus artificial neural network for modeling biosorption of Cr⁶⁺ by the immobilized *Pseudomonas alcaliphila* NEWG-2. *Sci. Rep.* **2021**, *11*, 1717. [[CrossRef](#)] [[PubMed](#)]
39. Shanmugaprakash, M.; Sivakumar, V. Development of experimental design approach and ANN-based models for determination of Cr(VI) ions uptake rate from aqueous solution onto the solid biodiesel waste residue. *Bioresour. Technol.* **2013**, *148*, 550–559. [[CrossRef](#)] [[PubMed](#)]
40. Parveen, N.; Zaidi, S.; Danish, M. Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr(VI). *Process Saf. Environ. Prot.* **2017**, *107*, 428–437. [[CrossRef](#)]
41. 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](#)]
42. Khan, H.; Hussain, S.; Hussain, S.F.; Gul, S.; Ahmad, A.; Ullah, S. Multivariate modeling and optimization of Cr(VI) adsorption onto carbonaceous material via response surface models assisted with multiple regression analysis and particle swarm embedded neural network. *Environ. Technol. Innov.* **2021**, *24*, 101952. [[CrossRef](#)]
43. Banza, M.; Seodigeng, T.; Rutto, H. Comparison Study of ANFIS, ANN, and RSM and Mechanistic Modeling for Chromium(VI) Removal Using Modified Cellulose Nanocrystals–Sodium Alginate (CNC–Alg). *Arab. J. Sci. Eng.* **2023**, *48*, 16067–16085. [[CrossRef](#)]
44. Zafar, M.; Aggarwal, A.; Rene, E.R.; Barbusiński, K.; Mahanty, B.; Behera, S.K. Data-Driven Machine Learning Intelligent Tools for Predicting Chromium Removal in an Adsorption System. *Processes* **2022**, *10*, 447. [[CrossRef](#)]
45. Zhu, X.; Xu, Z.; You, S.; Komárek, M.; Alessi, D.S.; Yuan, X.; Palansooriya, K.N.; Ok, Y.S.; Tsang, D.C. Machine learning exploration of the direct and indirect roles of Fe impregnation on Cr(VI) removal by engineered biochar. *Chem. Eng. J.* **2022**, *428*, 131967. [[CrossRef](#)]

46. Nighojkar, A.; Zimmermann, K.; Ateia, M.; Barbeau, B.; Mohseni, M.; Krishnamurthy, S.; Dixit, F.; Kandasubramanian, B. Application of neural network in metal adsorption using biomaterials (BMs): A review. *Environ. Sci. Adv.* **2023**, *2*, 11–38. [CrossRef] [PubMed]
47. Yaseen, Z.M. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere* **2021**, *277*, 130126. [CrossRef] [PubMed]
48. Oh, C.; Han, S.; Jeong, J. Time-Series Data Augmentation based on Interpolation. *Procedia Comput. Sci.* **2020**, *175*, 64–71. [CrossRef]
49. Xing, Y.; Yue, J.; Chen, C. Interval Estimation of Landslide Displacement Prediction Based on Time Series Decomposition and Long Short-Term Memory Network. *IEEE Access* **2020**, *8*, 3187–3196. [CrossRef]
50. Clevert, D.A.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv* **2016**, arXiv:1511.07289v5.
51. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* **2022**, *503*, 92–108. [CrossRef]
52. Rahman, M.M.; Watanobe, Y.; Nakamura, K.A. Bidirectional LSTM Language Model for Code Evaluation and Repair. *Symmetry* **2021**, *13*, 247. [CrossRef]
53. tf.keras.layers.Bidirectional | TensorFlow v2.15.0.post1. Available online: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Bidirectional (accessed on 25 January 2024).
54. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *JMLR* **2014**, *15*, 1929–1958.
55. Razavi, S. Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. *Environ. Model. Softw.* **2021**, *144*, 105159. [CrossRef]
56. Bengio, Y.; Grandvalet, Y. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. *J. Mach. Learn. Res.* **2004**, *5*, 1089–1105.
57. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Resampling Methods. In *An Introduction to Statistical Learning: With Applications in R*; James, G., Witten, D., Hastie, T., Tibshirani, R., Eds.; Springer: New York, NY, USA, 2013; pp. 175–201. ISBN 978-1-4614-7138-7-5.
58. Bindu, K.H.; Morusupalli, R.; Dey, N.; Rao, C.R. *Coefficient of Variation and Machine Learning Applications*; CRC Press: New York, NY, USA, 2019; pp. 1–26. ISBN 978-0-367-27328-6.
59. Jeong, N.; Chung, T.H.; Tong, T. Predicting Micropollutant Removal by Reverse Osmosis and Nanofiltration Membranes: Is Machine Learning Viable? *Environ. Sci. Technol.* **2021**, *55*, 11348–11359. [CrossRef] [PubMed]
60. Livesey, J.H. Kurtosis provides a good omnibus test for outliers in small samples. *Clin. Biochem.* **2007**, *40*, 1032–1036. [CrossRef] [PubMed]
61. Wang, H.S.H.; Yao, Y. Machine learning for sustainable development and applications of biomass and biomass-derived carbonaceous materials in water and agricultural systems: A review. *Resour. Conserv. Recycl.* **2023**, *190*, 106847. [CrossRef]
62. Huang, R.; Ma, C.; Ma, J.; Huangfu, X.; He, Q. Machine learning in natural and engineered water systems. *Water Res.* **2021**, *205*, 117666. [CrossRef] [PubMed]
63. Park, D.; Yun, Y.S.; Ahn, C.K.; Park, J.M. Kinetics of the reduction of hexavalent chromium with the brown seaweed *Ecklonia* biomass. *Chemosphere* **2007**, *66*, 939–946. [CrossRef] [PubMed]
64. He, T.; Niu, D.; Chen, G.; Wu, F.; Chen, Y. Exploring Key Components of Municipal Solid Waste in Prediction of Moisture Content in Different Functional Areas Using Artificial Neural Network. *Sustainability* **2022**, *14*, 15544. [CrossRef]
65. Lopez-Nunez, P.V.; Aranda-Garcia, E.; Cristiani-Urbina, M.D.C.; Morales-Barrera, L.; Cristiani-Urbina, E. Removal of hexavalent and total chromium from aqueous solutions by plum (*P. domestica* L.) tree bark. *Environ. Eng. Manag. J.* **2014**, *13*, 1927–1938.
66. Alvi, M.; Batstone, D.; Mbamba, C.K.; Keymer, P.; French, T.; Ward, A.; Dwyer, J.; Cardell-Oliver, R. Deep learning in wastewater treatment: A critical review. *Water Res.* **2023**, *245*, 120518. [CrossRef] [PubMed]
67. 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]
68. Ellen MacArthur Foundation. Artificial Intelligence and the Circular Economy: AI as a Tool to Accelerate the Transition. 2019. Available online: <https://www.ellenmacarthurfoundation.org/artificial-intelligence-and-the-circular-economy> (accessed on 18 March 2024).
69. Yuan, X.; Li, J.; Lim, J.Y.; Zolfaghari, A.; Alessi, D.S.; Wang, Y.; Wang, X.; Ok, Y.S. Machine Learning for Heavy Metal Removal from Water: Recent Advances and Challenges. *ACS EST Water* **2024**, *4*, 820–836. [CrossRef]
70. Mondejar, M.E.; Avtar, R.; Diaz, H.L.B.; Dubey, R.K.; Esteban, J.; Gómez-Morales, A.; Hallam, B.; Mbungu, N.T.; Okolo, C.C.; Prasad, K.A.; et al. Digitalization to achieve sustainable development goals: Steps towards a Smart Green Planet. *Sci. Total Environ.* **2021**, *794*, 148539. [CrossRef] [PubMed]
71. Fan, M.; Hu, J.; Cao, R.; Ruan, W.; Wei, X. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. *Chemosphere* **2018**, *200*, 330–343. [CrossRef] [PubMed]
72. Dico, G.L.; Peña Nuñez, Á.; Carcelén, V.; Haranczyk, M. Machine-learning-accelerated multimodal characterization and multiobjective design optimization of natural porous materials. *Chem. Sci.* **2021**, *12*, 9309–9317. [CrossRef] [PubMed]

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.