

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/181101/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zhang, Jingkang, Wang, Xingjie, Ma, Liyuan, Fu, Yao, Liu, Peng, Wang, Hongmei and Zhou, Jianwei 2025. Critical operational parameters for metal removal efficiency in acid mine drainage treated by constructed wetlands: An explainable machine learning approach. *Process Safety and Environmental Protection* 202 (Part B) , 107850. 10.1016/j.psep.2025.107850

Publishers page: <https://doi.org/10.1016/j.psep.2025.107850>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Critical operational parameters for metal removal efficiency in acid mine drainage treated by constructed wetlands: An explainable machine learning approach

Jingkang Zhang ^{a, b}, Xingjie Wang ^{b, *}, Liyuan Ma ^{a, b, c, *}, Yao Fu ^c, Peng Liu ^a, Hongmei Wang ^a, Jianwei Zhou ^{a, b}

^a *School of Environmental Studies, China University of Geosciences, Wuhan 430074, China*

^b *Institute of Geological Survey, China University of Geosciences, Wuhan 430074, Hubei, China*

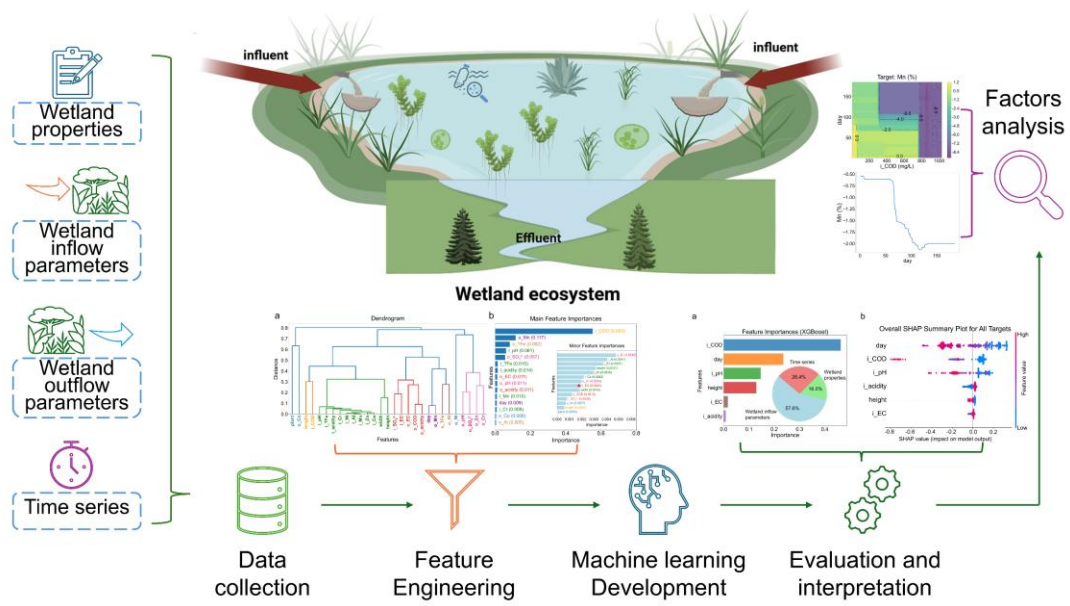
^c *School of Engineering, Cardiff University, Cardiff CF243AA, United Kingdom*

***Corresponding author at:** School of Environmental Studies, China University of Geosciences, Wuhan 430074, China. *E-mail address:* wangxingjie@cug.edu.cn (X. Wang), maly@cug.edu.cn (L. Ma)

Abstract

Constructed wetlands have long been recognized as a sustainable, effective and economical approach for treating acid mine drainage (AMD). The varying components of AMD at different locations impose significant site-specific constraints on the construction and maintenance of these wetlands. Herein, machine learning (ML) was utilized to predict and analyze multi-metal removal efficiencies, and address the complex interactions in constructed wetlands. Five ML models were developed, among which the XGBoost model achieved high apparent accuracy ($R^2 > 0.8$) for the removal efficiency of total iron, manganese, aluminum and zinc in the main pipeline. While model performance generally declined (R^2 decreased by approximately 0.2 overall) under leakage-safe out-of-fold (OOF) evaluation and forward-chaining time-series tests with naive baselines, tree-based models remained dominant, providing conservative estimates. Detailed feature and sensitivity analyses identified operation Days and inflow Chemical Oxygen Demand as significant predictors of metal removal efficiency. Furthermore, the empirical categories for metal removal, ranked by importance, were inflow parameters in first place, followed by time series, and wetland properties in last place. Partial dependence plots revealed certain ranges of the significant predictors and systematically illustrated their interactions and contributions to the metal removal efficiencies. These findings support near-real-time monitoring and short-horizon operational decisions.

Keywords: Acid mine drainage; Machine learning; Constructed wetland; Metals removal efficiency; Feature engineering



1. Introduction

Acid mine drainage (AMD) has been identified as a global environmental challenge (Vasquez et al., 2022). It arises from mining activities and is distinguished by high metal concentrations and low pH, carrying a risk of contaminating both surface and groundwater sources (Daraz et al., 2023; She et al., 2023; Stumm and Morgan, 2013). AMD is generated when sulfide minerals in mine tailings undergo oxidation due to exposure to water and atmospheric oxygen (Rodríguez-Galán et al., 2019; Wang et al., 2019), releasing dissolved ferrous iron and acidity into the water, consequently mobilizing additional metal ions (Naidu et al., 2019; Tabelin et al., 2020). The metal concentrations in site-specific AMD vary significantly (Nieva et al., 2018; Schaidler et al., 2014). Conventional chemical remediation techniques, such as the addition of limestone, sodium hydroxide, or other alkaline substances, have been employed to treat existing AMD by increasing its pH and thereby reducing its metal and sulfate concentrations (Masindi et al., 2017; Masindi et al., 2018; Masindi et al., 2022). However, the limitations and shortcomings of these traditional methods have promoted a growing body of research and practical efforts aimed at developing novel, efficient and cost-effective technologies for AMD remediation.

Constructed Wetlands, as a passive treatment comprises natural and biological processes to remediate AMD pollution (Jouini et al., 2020; Villegas-Plazas et al., 2022), offer an alternative to conventional chemical methods. The majority of constructed wetlands are largely shallow-water bodies (<30 cm in depth) rich in limestone gravel, soil and organic matter (Nguegang et al., 2022). Metal removal in these systems primarily occurs through oxidation, adsorption, and co-precipitation processes (Batty and Younger, 2002; Sheoran and Sheoran, 2006; Younger et al., 2002). For example, Fe and Mn are commonly removed as hydroxides or sulfides, while Al and Zn tend to form complexes with organic matter. These mechanisms have been shown to significantly improve water quality by reducing metal concentrations and acidity (Chen et al., 2021b; Wang et al., 2021; Wu et al., 2013b). Therefore, constructed wetlands have been noted for their cost-effectiveness, long-term stability and overall

effectiveness in removing contaminants from AMD, while also providing ecological benefits (Irshad et al., 2021). However, some studies have revealed that the treatment outcomes in constructed wetlands for AMD can be unpredictable, influenced by a variety of factors (Mitsch and Wise, 1998).

Traditional statistical methods typically establish linear or quadratic relationships between individual factors and a target outcome but often fail to capture the complex multivariate interactions inherent in such systems. In contrast, machine learning (ML) techniques offer a robust alternative to address this complexity by considering a wider array of influential factors and identifying intricate relationships, both linear and nonlinear, between features and targets (Palansooriya et al., 2022). Data-driven ML approaches are effective in dealing with multivariate complexity in environmental remediation, relying on extensive data retrieved from literatures. Various ML methods, such as Random Forest (RF), Extreme Gradient Boosting (XGB), k-Nearest Neighbors (kNN) and Neural Networks (NN), have been utilized to monitor and map contaminants in soil (Wu et al., 2013a) and groundwater (Lopez et al., 2021). Furthermore, some studies have used ML models to create risk assessment frameworks for groundwater pollution (Sajedi-Hosseini et al., 2018). In the context of constructed wetlands, the complex interactions and a lack of systematic datasets, have limited the success of traditional statistical methods in predicting metal removal efficiency. These challenges have driven the adoption of ML techniques (Hong et al., 2024a; Song et al., 2022; Zou et al., 2023).

Herein, five ML models, RF, XGB, Support Vector Regression (SVR), kNN and Artificial Neural Networks (ANN), were developed to predict metal removal efficiencies in AMD treatment by constructed wetlands. Detailed descriptions of these Traditional models, including their advantages and disadvantages, can be found in the supplementary material (Text S4: Traditional Machine Learning Models). As the quality of the dataset brought into a model profoundly affects the validity of the model (Briscoe and Marin, 2020; Kim et al., 2022), it was crucial to ensure the robustness of the initial datasets. Therefore, we devoted much effort to the construction of the dataset. The complete dataset was defined

as dataset A, and based on feature importance and hierarchical clustering, crucial factors were extracted as dataset B. Subsequently, further feature engineering was performed considering practical requirements (Briscoe and Marin, 2020) and tractability, to obtain a practice-oriented dataset C. The specific tasks are: (1) to select the optimal model by evaluating and comparing the models using multiple performance metrics; (2) to identify significant predictors by conducting feature importance analysis and Shapley additive explanation (SHAP) analysis based on the optimal model and dataset C; (3) to quantify the importance of empirical categories through comprehensive analysis; and (4) to reveal the interactional impact between significant predictors on predicting multi-metal removal efficiency using partial dependence plots. This data-driven ML approach elucidated the complex interactions in constructed wetlands, providing a deeper understanding of how varying parameters affect the removal efficiency of metals in AMD treatment.

2. Materials and methods

2.1 Data collection and data preprocessing

Data for this study were collected from published papers (from 2006 to 2023) by searching through Google Scholar databases and Web of Science using “constructed wetlands” and “acid mine drainage” as keywords. Initially, 31 studies were that pertained to the treatment of AMD using constructed wetlands and included available monitoring data. However, only a subset of these studies reported comprehensive and standardized monitoring data, which are essential for ML model development. To ensure data consistency and comparability, the largest subset of studies (five wetlands and the detailed attributes are shown in Table S1) that shared common monitoring parameters (e.g., pH, COD, metal concentrations) was ultimately selected (Chen et al., 2021b; Singh and Chakraborty, 2020; Singh and Chakraborty, 2021). This approach allowed us to construct a robust dataset for model training and analysis, despite the inherent variability in data reporting across the literature. 31 parameters from five wetlands were obtained

from tables or extracted from figures in the papers as dataset variables using Origin 2021 (OriginLab Corporation, 2021), resulting in 354 data points for ML exploration. The features in datasets for ML models development include 4 categories: wetland properties (length, width, height, plant), inflow water parameters (pH, COD, acidity, TDS, EC, SO_4^{2-} , concentrations of Total Fe, Mn, Al, Zn, Ni, Co and Cr), outflow water parameters corresponding the inflow ones, and time series (Table S2). The “Day” parameter represents the continuous operation time (in days) since the initiation of the constructed wetland, with all data aligned to this timeline to ensure consistency across different studies. Given the inherent one-to-one relationship between inflow and outflow water parameters, the dataset utilizes the notation “i_X” to denote inflow parameters and “o_X” to denote outflow parameters. The metal removal efficiency were calculated as a percentage using eq 1.

$$\text{Metal removal efficiency (100\%)} = \frac{(C_{i_X} - C_{o_X})}{(C_{i_X})} \times (100\%) \quad (1)$$

where C_{i_X} refers to the inflow concentration of a certain metal, and C_{o_X} is the outflow concentration of the same metal. The concentrations are expressed in $\text{mg} \cdot \text{L}^{-1}$.

Metals (Total Fe, Mn, Al, Zn, Ni, Co and Cr) removal efficiencies was considered as the target variables. The detailed process of ML exploration associated with the metal removal efficiencies during the treatment of AMD using constructed wetlands is illustrated in Figure 1. In the main pipeline, RF was identified as the primary imputation method by comparing the performance of different imputation methods (RF, Bagging, Histogram Gradient Boosting Regression (HGBR), DecisionTree, RegressionTree, AdaBoost, MICE, KNN, MLP, SVR) in a five-fold cross-validation for missing data in the initial dataset (Figure S1 and Table S3) (Liu et al., 2022). HGBR (see Text S4 for details) and Hot deck imputation (see Text S1 for details) were applied to specific variables. Due to the high proportion of missing data (29.7%) (Zhu et al., 2023), outflow TDS was removed after a series of attempts, along with its corresponding inflow TDS. After pre-processing, the final dataset contained 29 features, 7 targets and 354 data points. Leakage control for imputation, including the leakage-safe out-of-fold (OOF)

pipeline and sensitivity analysis of the imputation strategies are described in Section 2.3.2 *Controls for data leakage*.

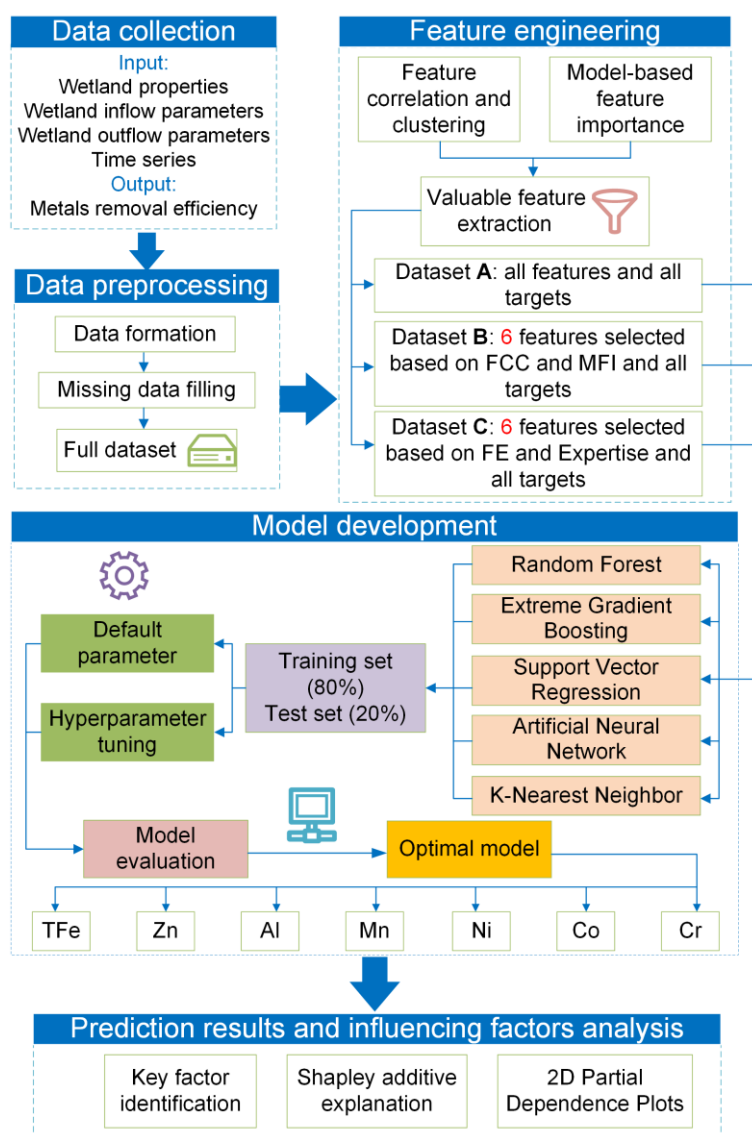


Figure 1. The flowchart provides a detailed overview of the strategy employed for predicting the efficiencies of metals removal in AMD treated by constructed wetlands using a machine learning framework. During the first step, data were collected from the literature and subjected to data preprocessing. Subsequently, feature engineering was applied to extract and optimize features from the full dataset, which was then utilized for model development and comparative evaluation. Finally, impact factor analysis was conducted based on the optimal model and dataset to investigate the influence of different input features on the prediction target variables. Note: FCC: feature correlation and clustering; MFI: model-based feature importance; FE: feature engineering; TFe: total Fe.

2.2 ML-based feature engineering

Principal component analysis (PCA) was first performed to visualize and assess the diversity and

distribution of input features across all wetland samples. To simplify the ML model and improve its performance, feature filtering was performed based on feature correlation and ML-based feature importance analysis (Palansooriya et al., 2022). The correlation analyses between different variables were conducted using Pearson correlation coefficient (PCC), and the results were visualized in the form of a heatmap (Figure S2). Hierarchical clustering was then implemented, utilizing Pearson rank-order correlations to group highly correlated features based on calculated similarities or distances between data points (Johnson, 1967). Further, feature importance analysis was conducted with ML-based model to determine the significance of each feature in predicting the target variable (Zhu et al., 2019a). By integrating results from feature importance and correlation analysis, the most important feature within a cluster was selected as input features to simplify the dataset and reduce the complexity of model execution.

To compare the predictive performance of the same model when facing different input datasets, three distinct datasets, A, B and C, were delineated based on domain expertise and results of feature engineering. Dataset A comprised all features and all target variables from the full dataset. Dataset B was determined by the evaluation of feature importance and correlation, resulting in the selection of the six most important features and all target variables. Dataset C was crafted with consideration of monitoring difficulty and cost, as well as domain expertise (Daraz et al., 2023; Mayes et al., 2009; Pat-Espadas et al., 2018; Sheoran and Sheoran, 2006), alongside the assessment of feature importance and correlation, leading to the selection of six features (inflow COD: selected based on feature engineering; inflow pH: selected based on feature engineering; inflow acidity: selected based on feature engineering and monitoring difficulty; inflow EC: selected based on feature engineering and monitoring difficulty; Day: a monitoring indicator that is considered by almost all wetlands; Height: considering the importance of wetland properties for selection). After finalizing the selection of six key input features for dataset C, boxplots were generated to visualize their distributions among different wetlands.

2.3 Model development and evaluation

2.3.1 Model development of the main pipeline

To improve the training process of ML models for rapid convergence, the input features were standardized using StandardScaler in Scikit-Learn (version 1.4.1.post1) (Pedregosa et al.) with Python 3.9.7 to obtain a similar scale and approximate a normal distribution. Following data standardisation, 80% of data were randomly extracted from each input dataset and used for model training, while the remaining 20% were used for testing in the main pipeline (Yin et al., 2024; Zhang et al., 2023b). Based on the dataset size, data type and existing research, five widely applied models, namely RF (Zhao et al., 2023), Extreme Gradient Boosting (XGB) (Sun et al., 2022), Support Vector Regression (SVR) (Palansooriya et al., 2022; Zhang et al., 2020), k-Nearest Neighbors (kNN) (Yin et al., 2024) and Artificial Neural Network (ANN) (Zhang et al., 2020), were selected for this study. The method of grid search with cross-validation was employed during the initial training process to conduct hyperparameter tuning, aiming to enhance the accuracy of model learning and generalization (Bergstra and Bengio, 2012; Zhu et al., 2023; Zhu et al., 2019b). It is implied that each input dataset will be fed into the five models, with each model being run once using default parameters and once using optimized parameters for comparative analysis. All models were validated using 5-fold cross-validation in the training process to achieve more stable predictive performance and mitigate overfitting (Yan et al., 2021).

The coefficient of determination (R^2), root-mean-square error (RMSE) and mean absolute error (MAE) were utilized to compare the prediction accuracy and quantify the prediction performance (Hu et al., 2022). R^2 , RMSE and MAE values were calculated by using eqs 2 and 3, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i^a - y_i^p)^2}{\sum_{i=1}^n (y_i^a - \bar{y}_i^a)^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i^a - y_i^p)^2}{n}} \quad (3)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4) \text{ where } y_i^a \text{ is the predicted value of the output, } y_i^p \text{ is the true value}$$

of the output collected from the literature on experimental research, $\bar{y}_i^a$ is the mean value of all output

values, y_i is the true value of the i th sample, $\hat{y}_i$ is the predicted value of the i th sample, and n is the number of data samples in the training or testing datasets.

2.3.2 Controls for data leakage and generalization stability evaluation

To eliminate imputation-order leakage (i.e., imputing before splitting), we also conducted a OOF evaluation in addition to the main pipeline, where the imputation and all transformers were fitted only on training folds within a scikit-learn pipeline, and the OOF predictions were then concatenated for evaluation (Lones, 2024). In addition, we further performed sensitivity checks using global, group-wise, and group-wise K-fold imputations. To assess temporal leakage and genuine forecasting skill, expanding-window (forward-chaining) TimeseriesSplit and naive persistence baselines (last-value and short moving-average) were employed (details in Text S2) (Megahed et al., 2024; Schroer and Just, 2024). Leave-one-wetland-out (LOGO) cross-validation further rigorously assessed the generalization of models across different wetland systems. In addition, to evaluate the engineering reliability and robustness of the optimal model, a systematic $\pm 10\%$ input perturbation sensitivity analysis was conducted on each key input feature in dataset C.

2.4 Influential factors analysis

Two types of feature analysis methods were employed to assess the importance of features and the correlation with metal removal efficiencies. One analysis method was conducted based on the optimal model of dataset C to analyze feature importance. The other analysis method utilized the SHAP method, which is widely employed in ML models explanation and feature analysis (Li et al., 2020; Li et al., 2022). The marginal effects of each feature on predicted outputs were determined by using ML models and the relevance between input features and target variables (such as linear, monotonic, or even more complex relationships) (Palansooriya et al., 2022). Two-dimensional (feature interaction) partial dependence plots were utilized to visually interpret the optimal model, and systematically expressing the correlation between the interaction of two features and the target variables.

3. Results and discussion

3.1 Dataset formation and feature engineering

Across the initial dataset, missing data were identified for 20 variables. Subsequent RF model-based imputation successfully addressed a majority of these missing data (Figure 2a, 2b). However, it is important to note that imputation may introduce some uncertainty into the dataset. Nevertheless, cross-validation results demonstrated the robustness of RF model to potential noise, with R^2 fluctuations remaining below 5% across different imputation scenarios. In addition, three variables (outflow COD, outflow TDS and outflow acidity) persisted with suboptimal performance in missing data prediction. Following a second round of imputation, where the HGBR model was selected for its precise in imputing missing data for outflow acidity (Figure 2c). Upon examining the proportion of missing data for the remaining two variables (4.5% for outflow COD and 29.7% for outflow TDS), hot-deck imputation was applied to outflow COD. The Kolmogorov-Smirnov test(KS test) (Berger and Zhou, 2014; Kini et al., 2024) was used to compare the cumulative distribution functions (CDFs) (Hong et al., 2024b; Ransom et al., 2017) before and after imputation, validating the imputation effectiveness. According to the KS test evaluation results for this hot-deck imputation, the KS statistic is 0.02, and the p-value is 0.99. Since the p-value is much greater than the commonly used significance level (0.05), indicating that the distributions of the data before and after imputation are consistent. This suggests that there is no significant difference in distribution between the data imputed using hot-deck imputation and the original data, demonstrating the effectiveness of the imputation method in this aspect (Figure 2d). Due to the high proportion of missing values for outflow TDS (exceeding 10%), which rendered the variable unsuitable for imputation methods and resulted in suboptimal model fitting and prediction, the decision was made to remove this variable alongside its corresponding inflow TDS variable. Although TDS is an important parameter in water quality monitoring, its exclusion was necessary to avoid introducing significant uncertainty into the model.

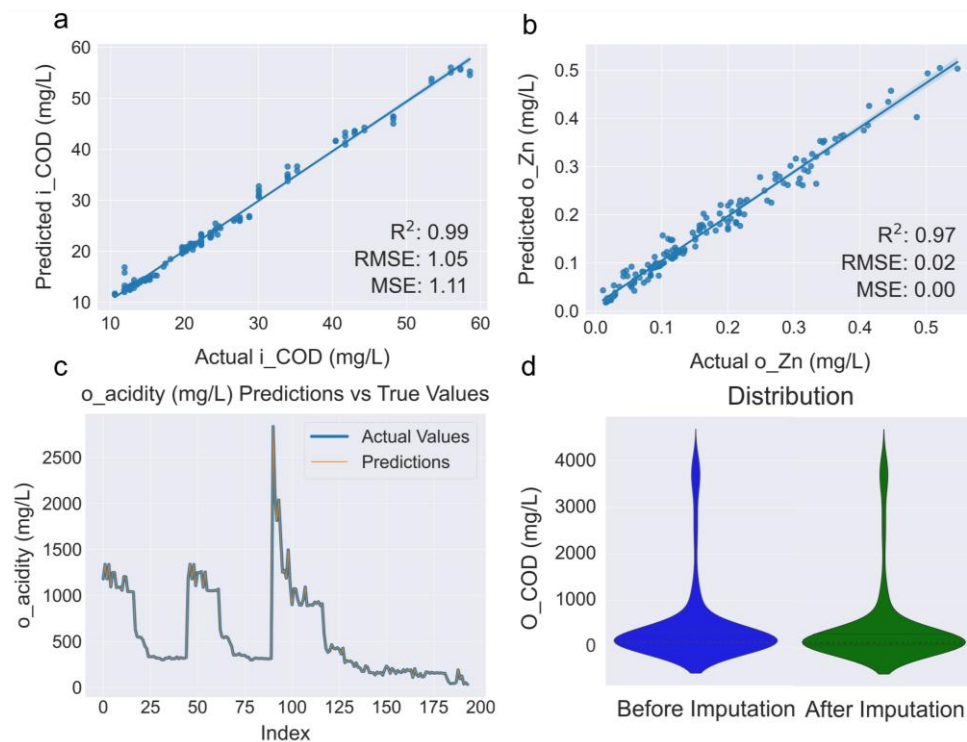


Figure 2. Results of (a) the RF model for predicting missing data in inflow COD, (b) the RF model for predicting missing data in outflow Zn concentrations, (c) the actual values and predictions of outflow acidity by the HGBR model, and (d) the comparison of the frequency distribution of outflow COD before and after hot-deck imputation.

Following the completion of dataset filling, PCA was performed to visualize the distribution and heterogeneity of the input features of the five wetlands (Figure S3) (Wang et al., 2025). The principal component analysis results show that samples from the five wetlands form distinct clusters in the principal component space. Although some overlap exists, the overall pattern highlights substantial inter-wetland differences as well as partial shared variability within the dataset. Subsequently, an extensive feature analysis ensued, which included the generation of a heatmap illustrating feature correlations based on PCC (Figure S2). Hierarchical clustering was performed by calculating the similarity between inputs and converting it into distances. Features with distances below a threshold (0.4) were divided into one cluster, with closer distances indicating closer relationships in the dendrogram (Figure 3a). Figure 3b displays the importance of each feature for predicting the target variables, computed using the RF model. Inflow COD was discovered to be the most important feature for predicting metal removal

efficiencies. Figure 3a delineates distinct clusters of features using varied colors, yet it is noteworthy that the trio of blue features doesn't constitute a single cluster but rather represents three distinct clusters. To refine the dataset and enhance model generalization while reducing computational complexity, a feature filtering process which involved integrating the outcomes of hierarchical clustering with feature importance assessments was conducted.

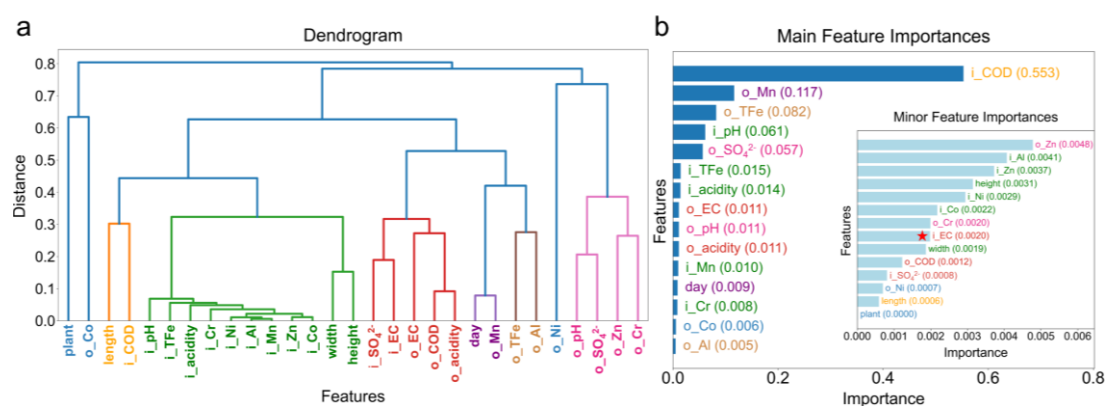


Figure 3. Input feature analysis: (a) hierarchical clustering and (b) feature importance from the XGBoost model. A total of 354 data points were utilized for model development. In panel (a), clusters are shown in different colors and hierarchical levels, derived from the Pearson rank-order correlations depicted in Figure S2 using a hierarchical clustering algorithm. A distance threshold of 0.4 was chosen to evaluate feature similarity. Features converging into the same cluster at this threshold indicate high similarity, simplifying the model by postextraction of important features. Note: the prefix “i_” represents inflow parameters, while the prefix “o_” represents outflow parameters.

Figure 3a illustrates that length and inflow COD were clustered together, and according to the feature importance ranking (Figure 3b), inflow COD is the more critical feature. Hence, inflow COD was selected as the representative feature for this cluster input into the model. Furthermore, the clustering identified distinct groupings: the green cluster is characterized by the inflow pH; the red cluster embodies outflow EC; the purple cluster represents outflow Mn concentration; the brown cluster encapsulates outflow Total Fe (TFe) concentration, and the pink cluster epitomizes outflow SO₄²⁻. Additionally, due to their lower importance, features from the remaining three blue clusters were omitted. Consequently, dataset B (inflow COD, inflow pH, outflow EC, outflow Mn concentration, outflow TFe concentration

and outflow SO_4^{2-}) was constructed for model development.

Despite the apparent optimality of dataset B from an ML perspective, it is crucial to acknowledge that the selection of input features based solely on their correlation and importance may not consistently adhere to domain expertise and real-world necessities. The features within dataset B comprise two inflow parameters and four outflow parameters, eschewing wetland and time series variables. Upon scrutiny, it becomes evident that monitoring the outflow concentrations of TFe and Mn poses considerable challenges; therefore, non-metallic parameters emerge as a more pragmatic choice. Similarly, focusing solely on either inflow or outflow parameters can streamline monitoring efforts and mitigate associated costs. However, from a pragmatic perspective, integrating wetland and time series variables into the predictive framework offers superior guidance for developing constructed wetlands and managing AMD. Through the integration of particle experiment conditions and feature analysis, inflow COD, inflow pH, inflow acidity, inflow EC, wetland Height and operation Day as inputs for dataset C in model development. Boxplots of the six selected input features for dataset C (Figure S4) reveal distinct distribution patterns among the five wetlands, further supporting the presence of inter-wetland heterogeneity in the final modeling dataset. Notably, Wetlands 1 and 2 exhibit much higher influent pH and COD values, while Wetlands 3, 4, and 5 have higher influent acidity and electrical conductivity. There are also significant differences in wetland height among the groups. These observations highlight the necessity of considering wetland-specific input conditions when developing and deploying predictive models (Reed et al., 2020). Finally, in assessing the fundamental predictive performance of different models, dataset A, encompassing all inputs from the full dataset was utilized. This strategy aimed to mitigate the potential reduction in predictive accuracy resulting from the exclusion of valuable features by datasets B and C. However, given that the target (metal removal efficiency) is calculated according to eq 1, datasets A and B containing outflow variables have the risk of target leakage. Therefore, the model performance of these two datasets is only used as a theoretical upper limit reference, and the final model

evaluation and interpretation are based on dataset C.

3.2 ML model development and evaluation

3.2.1 Model development and comparison

In the main pipeline, five established and widely utilized ML models (RF, XGB, SVR, ANN and kNN) were employed. Each model underwent a comparison between two configurations: default parameters and optimized parameters from GridSearchCV (Table S4). 5-fold cross-validation was employed to enhance the generalization ability of model, reduce the risk of overfitting, and obtain robust model, while the test dataset was utilized to test the model performance. Following the model development, 30 prediction results were obtained (Table S5 and 6), and model comparison was conducted using R^2 as the main evaluation metric.

In accordance with previously published research, both Fe and Mn were identified as two key heavy metals requiring particular attention in the treatment of AMD using constructed wetlands (Chen et al., 2023; Singh and Chakraborty, 2020; Wang et al., 2021). They were commonly encountered in AMD and were frequently found to exceed standard concentration levels. During the process of Fe conversion to hydroxides, the transformation of aluminum often accompanies (Singh and Chakraborty, 2020). Additionally, research indicated that aluminum played a significant role in plant growth and can mitigate the toxicity of metals such as Fe, Mn and H^+ in acidic soils (Muhammad et al., 2019; Nguegang et al., 2022). Furthermore, it was found that the concentration of Zn significantly exceeds the standard limits. Therefore, this study focuses on predicting the removal efficiency and analyzing the influential factors of TFe, Mn, Al and Zn. This aligns with the emphasis on key pollutant metals in relevant published studies (Daraz et al., 2023; Singh and Chakraborty, 2020; Wang et al., 2021).

Figure 4 depicts the apparent predictive performance of three models (RF, XGB and kNN) concerning the removal efficiency of four different metals (TFe, Mn, Al and Zn) when faced with varying datasets (A, B and C). SVR and ANN were excluded due to their relatively lower apparent predictive

performance compared to the other tree-based models. The decision to focus on the three better-performing models was made to emphasize the most effective approaches. The detailed performance metrics for SVR and ANN are provided in the supplementary material (Tables S4 and S5) for reference. The exceptional analytical capabilities were demonstrated by all three models on dataset A, yielding precise apparent predictive results ($R^2 = 0.94$ -1.00). This aligns with the diverse and comprehensive feature characteristics of dataset A, indicating that the abundant relevant features and sample data can effectively predict the removal efficiency. Compared to dataset A, the apparent predictive performance of the three models deteriorated when applied to datasets B ($R^2 = 0.80$ -1.00) and C ($R^2 = 0.53$ -0.99). This decline can be attributed to the feature selection process applied to datasets B and C, resulting in the loss of some valuable information. In predicting the removal efficiencies of TFe, Mn and Al, datasets B and C exhibit similar performances across different models. However, in predicting the removal efficiency of Zn, dataset B outperforms dataset C in the RF and kNN models, whereas dataset C performs better in the XGB model. Overall, dataset C shows a certain decline in performance compared to dataset B, which is consistent with the feature selection criteria applied to these two datasets.

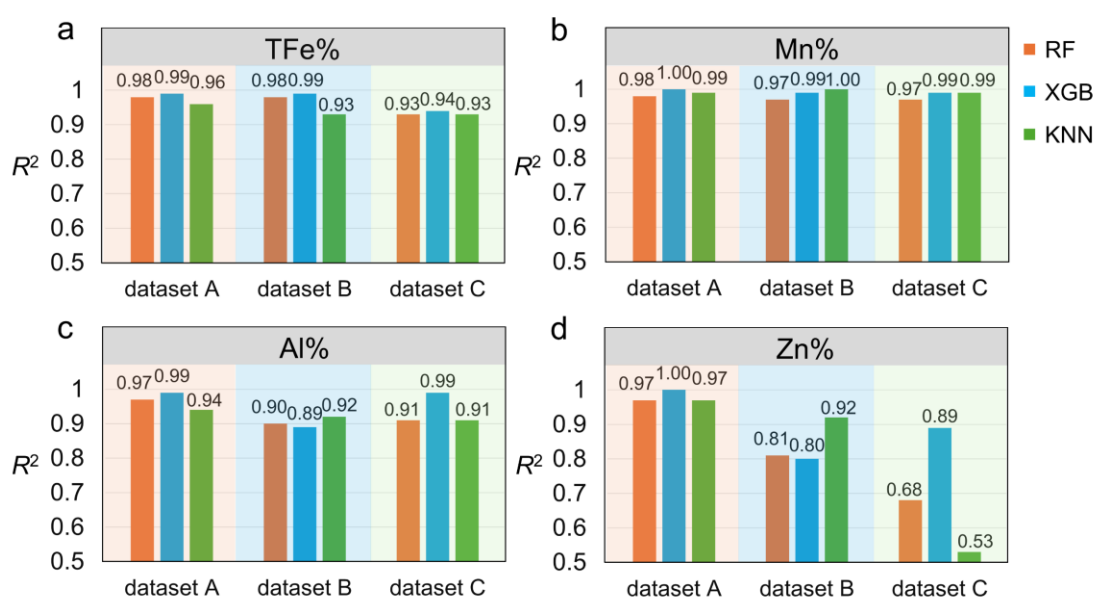


Figure 4. The comparative predictive performance of different models across varied input datasets for

(a) TFe%, (b) Mn%, (c) Al% and (d) Zn%, assessed using R^2 as the evaluation metric.

Compared to the feature selection in dataset B, the slight performance decrease observed in dataset C is acceptable because it includes less important but more easily monitored features. Therefore, model evaluation based on dataset C identified the optimized model as the XGB model with tuned hyperparameters, and the optimal parameters settings were shown in Figure S5. Figure 5 illustrates the predictive results and performance of the optimal model XGB based on dataset C for removal efficiencies of TFe, Mn, Al and Zn. The prediction results and model evaluation metrics for Ni, Co and Cr are presented in Figure S6. The XGB model demonstrated excellent predictive performance on all target variables in the training set ($R^2 > 0.93$). However, there was a general decline in performance on the test set, particularly for the prediction of Zn removal efficiency ($R^2 = 0.87$). Nonetheless, this level of accuracy is deemed acceptable.

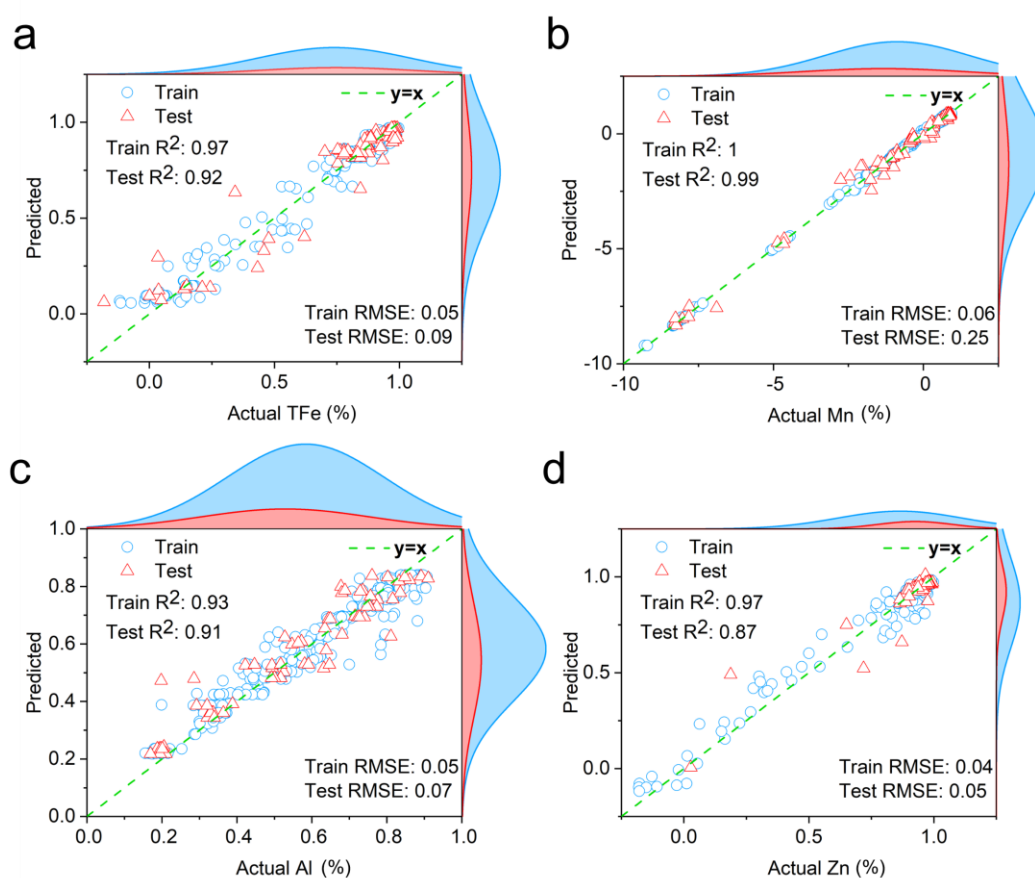


Figure 5. The predictive performance demonstration of the optimal model XGB for (a) TFe removal

efficiency, (b) Mn removal efficiency, (c) Al removal efficiency, and (d) Zn removal efficiency based on dataset C, evaluated using R^2 and RMSE as assessment metrics. RMSE = root-mean-square error.

3.2.2 Controls for data leakage and generalization stability evaluation

To quantify the optimism that can arise from imputing on the full table before splitting, we re-evaluated the pipeline using the OOF protocol. Relative to the main pipeline, OOF scores decreased as expected (median $\Delta R^2 = -0.163$ and mean = -0.198), but the relative model ranking was preserved (Figure S7a and b). In particular, RF achieves the highest mean R^2 across models and wins 6/7 targets (kNN wins Ni once), XGB ranks next, SVR followed, whereas ANN was unstable (Figure S7c and d). This indicates that while absolute scores are lower under OOF (as expected), the substantive conclusion that tree-based models are best for this task is robust. For completeness, sensitivity analyses including global imputation and group-wise imputation (with/without K-fold) were conducted. The results showed a conservative downward shift in absolute R^2 (down 0.064 ~ 0.309 compared to the main pipeline), but do not change the qualitative ranking (tree-based models are the strongest). Bland-Altman views indicate penalties concentrate on lower-baseline metals, with less detrimental bias at higher performance levels (Figure S8). Table S7 lists the best per-metal model under each strategy. With forward-chaining TimeseriesSplit and naive baselines, learned models won 5/7 metals for one-step-ahead, but only 3/7 under a ~10-sample window, indicating discernible short-term skill but limited long-horizon skill (Table S8 and 9). Moreover, as shown in Figure S9, the multi-horizon fold-median error analysis showed similar findings, with varying variability for each metal. Together, these results outline the boundaries of generalization ability and reinforce that our primary contribution relates to static conditional response estimation and near-real-time operational support, rather than long-term forecasting.

Furthermore, to evaluate the representativeness and diversity of the full dataset and its impact on model generalizability, LOGO cross-validation was performed using all available input features. The LOGO results (Table S10) indicate that the model achieves R^2 values above 0.5 for most target metals and wetlands, with a few cases of reduced performance reflecting the high heterogeneity among wetlands. These findings confirm that, while high predictive accuracy can be achieved in most scenarios, cross-wetland prediction remains challenging for certain systems, underscoring the importance of dataset diversity and realistic model evaluation in environmental applications. To further assess the engineering reliability and robustness of the optimal model, we conducted a systematic $\pm 10\%$ perturbation sensitivity analysis for each input feature in dataset C (Figure S10). The results demonstrated that model performance for most metals removal efficiency was relatively robust to moderate input fluctuations, with only a small number of cases (e.g., perturbations to i_COD, i_acidity, or height for certain metals) resulting in notable declines in prediction accuracy. This highlights both the practical stability of the modeling approach and the critical importance of certain key input features in real-world monitoring scenarios.

Overall, these results suggest that the removal of TFe, Mn, Al and Zn in the treatment of AMD by constructed wetlands can be effectively predicted using a basic ML model with well-designed feature engineering. The reduction in input features not only lowered the computational and time costs of model development but also accelerated the analysis process. By selecting input features that were more relevant in practical terms while maintaining a high level of predictive accuracy, this approach holds valuable implications for the monitoring and management of constructed wetlands. However, it is also necessary to strictly control the risk of data leakage and improve the generalization ability and actual engineering reliability of the model.

3.3 Key influential factor analysis

To quantitatively decipher the factors influencing the prediction of metal removal efficiencies, we

employed the SHAP analysis on optimal model to reflect the importance of these factors (Figure 6a, b). The comparison of feature importance rankings between the two analysis methods reveals discrepancies. While the XGB-based feature importance analysis indicated that inflow COD held the highest significance, followed by operation Day, the SHAP explanation method suggested that operation Day took precedence, followed by inflow COD. Furthermore, to ensure the robustness of our interpretability results, three other mainstream feature importance analyses (permutation importance, feature ablation, and LIME) were performed (Altmann et al., 2010; Garreau and Luxburg, 2020; Ribeiro et al., 2016). All three methods consistently identified i_COD and day as the most important features, further supporting the reliability of our main conclusions (Figure S11). Feature importance and cross-fold stability of response shapes were verified under leak safety assessment using XGB in the same OOF protocol. The results of permutation importance (top-ranks) showed that day is the most stable driver (Top 1 in 4/7 targets and Top 3 in 7/7), followed by i_COD (Top 1 in 2/7 and Top 3 in 5/7) and height (Top 1 in 1/7 and Top 3 in 4/7), i_pH and i_EC are secondary, and i_acidity rarely makes it into the top three (Figure S12a and b). Results from PDP cross-fold similarity indicated that the dominant drivers (day and i_COD) exhibit stable response shapes across folds (Figure S12c). The results of SHAP with permutation agreement confirmed high concordance between two explanation methods on the key drivers (Figure S12d). Together, these results show that the main interpretation (day and i_COD as primary drivers) is stable under a leakage-safe evaluation, reinforcing our original mechanistic reading while providing more conservative performance estimates.

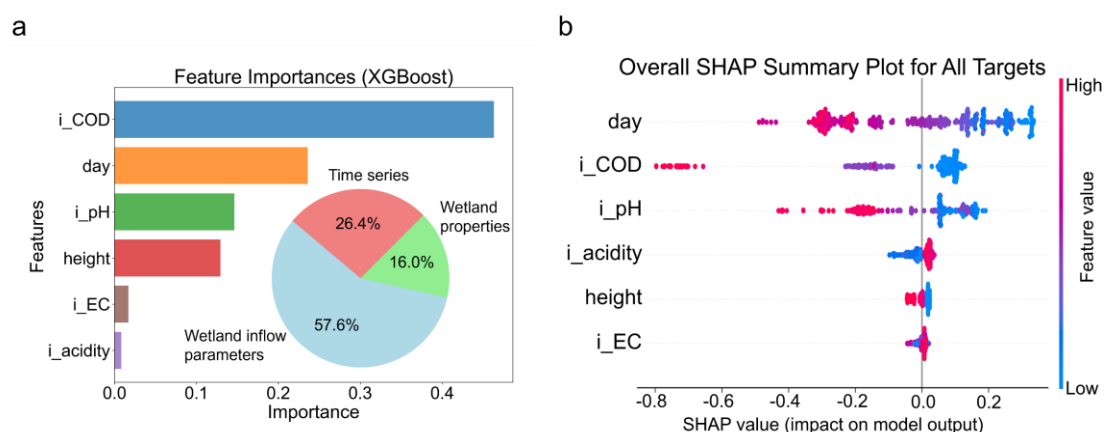


Figure 6. Influential factors analysis based on dataset C and the optimal XGB model: (a) feature importance assessment based on XGB model and (b) Shapley additive explanation method.

Afterwards, each feature in dataset C was analyzed separately for its impact on the removal efficiency of different metals (Figure S13). Figure S13 illustrates that the removal efficiency of different metals decreases with the increase in AMD entering the wetland, and the timing of the decline varies for each metal (approximately 85 days for TFe, 65 days for Mn, 40 days for Al, and 100 days for Zn). Similarly, the removal efficiency of different metals decreases with the increase in inflow COD. This could be attributed to the prioritization of microbial degradation and oxidation reactions of organic compounds in wastewater over the oxidation-reduction of metals when the COD in the influent water is high (Li et al., 2015; Lu et al., 2016). Elevated levels of COD may also affect the normal absorption and transport functions of plant roots, reducing their ability to absorb metals and thus impacting metal removal efficiency (Sharma et al., 2021; Zhang et al., 2023a). After a period of time following the entry of AMD which exhibits high COD concentration into the constructed wetland, as microbial populations increase and organic matter is removed, the COD concentration in the wetland water decreases, and the functionality of plants becomes more active, facilitating the oxidation, precipitation and absorption of metals, thereby enhancing metal removal efficiency (Chen et al., 2021b).

However, the threshold points of inflow COD that affect the removal efficiency of the four metals are not the same (approximately 300 mg/L for TFe, 650 mg/L for Mn, 60 mg/L for Al and 300 mg/L for

Zn), indicating that in the operation and management of the wetland, COD regulation needs to be comprehensively determined considering the influence of other input features. In general, the COD concentration in typical AMD is relatively low (below 50 mg/L). However, three constructed wetlands have added domestic wastewater or plant litter leachate to increase the inflow COD, thereby enhancing microbial activity and promoting plant growth (Chen et al., 2021b). Additionally, a layer of organic substrate is incorporated into the wetland bed to promote plant growth, provide a carbon and nitrogen source for microorganisms, and improve heavy metals retention (Choudhary and Sheoran, 2012).

The influence of inflow pH on the removal efficiency of the four metals also follows a similar trend, decreasing with increasing inflow pH (TFe is approximately 2.8 and 5.7, Mn is approximately 2.9 and 4, Al is 3, and Zn is approximately 3.5 and 4). The presence of different inflow pH threshold points for the removal efficiency of a metal may indicate varying removal mechanisms for that metal. For example, Fe may be removed as sulfides or hydroxides (Oldham et al., 2019), or it may be adsorbed and enriched by plants and constructed wetland substrates (Singh and Chakraborty, 2020). Changes in pH can also affect the microbial diversity in wetlands, thereby influencing the removal of metals (She et al., 2021; Sun et al., 2020). In general, as acidic AMD enters the constructed wetland and pH gradually increases over time, the removal efficiencies of different metals tend to decrease. This is due to the increased pH causing some metal precipitates to become less stable, resulting in re-dissolution or further dissolution (Rodrigues et al., 2019). Additionally, plants may experience toxic effects after absorbing significant amounts of metals, leading to a decrease in their ability to absorb heavy metals. As pH increases, the solubility of some metal ions may decrease, making them more difficult to be captured and removed by adsorbents or precipitates in the wetland (Sheoran et al., 2010). The effects of these three features on the removal efficiency of the four metals vary, indicating differences in their underlying mechanisms. Therefore, we further analyze the interactions between features to understand their combined impact on the removal efficiency of different metals.

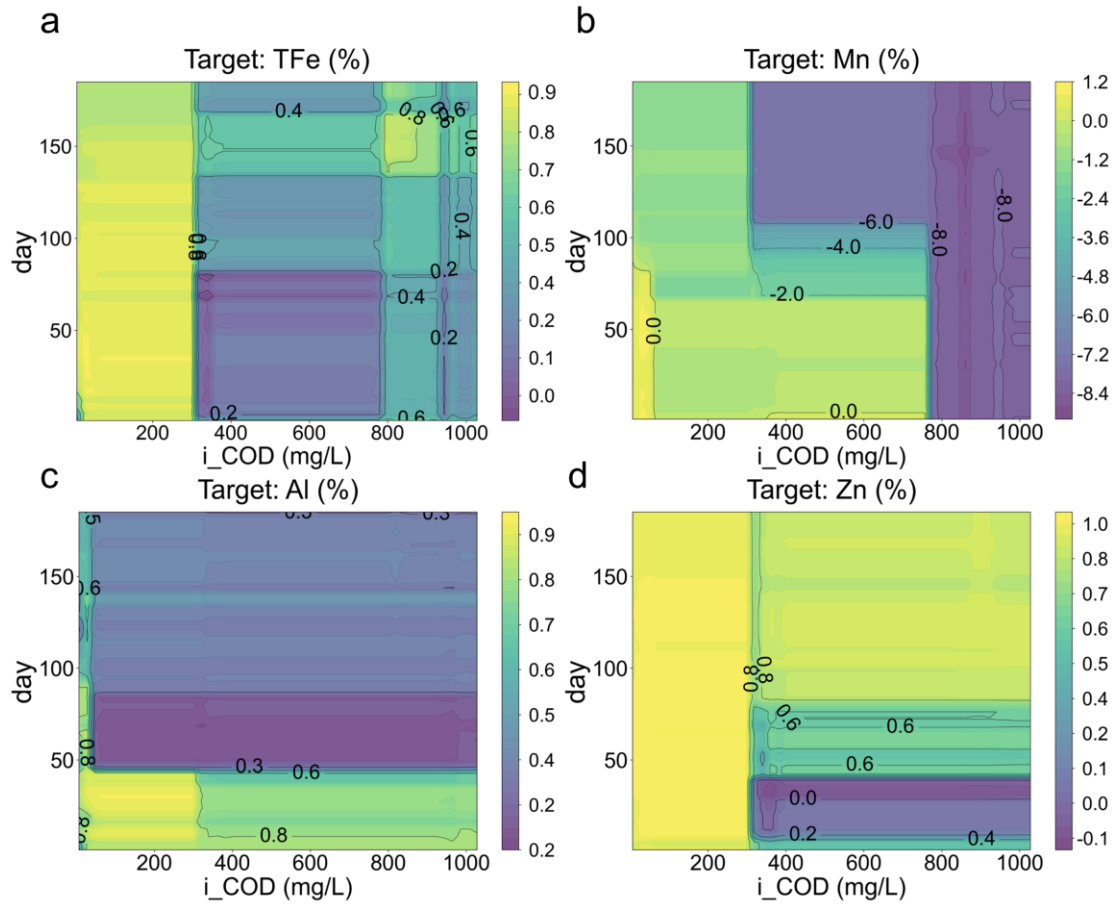


Figure 7. The interaction between inflow COD and operation Day was analyzed to assess its impact on the (a) TFe removal efficiency, (b) Mn removal efficiency, (c) Al removal efficiency, and (d) Zn removal efficiency.

The influence of the interactions between these variables (inflow COD and operation Day) on predicting the removal efficiencies of TFe, Mn, Al and Zn were depicted in Figure 7 (Ni, Co and Cr in Figure S14). For TFe removal, three distinct regions were delineated in Figure 7a by the inflow COD. When the inflow COD was below 300 mg/L, the TFe removal efficiency was close to 100%, with little impact by the operation Day. In the range of inflow COD between 300-800 mg/L, there was a significant decline in TFe removal efficiency, reaching its lowest point (below 20%) between 0-80 days. In the region where inflow COD exceeded 800 mg/L, the TFe removal efficiency fluctuated between 20%-80%. The results suggest that inflow COD is a more significant factor than operation Day in influencing TFe removal efficiency.

For Mn removal, the interaction between inflow COD and operation Day was more pronounced.

When inflow COD was below 300 mg/L, Mn removal efficiency was generally observed to be close to 120%, with a peak noted when inflow COD was below 50 mg/L in the 0-80 days. This specific region likely corresponds to the changes in Mn removal efficiency in wetlands where the inflow COD concentration was not altered. In the range of inflow COD between 300-800 mg/L, significant declines in Mn removal were observed over time, with Mn removal efficiency dropping below -600% after 100 days. The occurrence of negative Mn removal efficiency has been reported in many previous studies (Wang et al., 2021), possibly due to manganese remaining in the Mn^{2+} state under anaerobic/oxic conditions, making it more soluble and difficult to remove (Singh and Chakraborty, 2020). Additionally, the formation of insoluble manganese sulfide precipitates is challenging, making it prone to release.

For Al removal efficiency, the influence of operation Day was more pronounced, with 50 days and 80 days serving as clear boundaries. From 0 to 50 days, Al removal efficiency was controlled by inflow COD, with the highest efficiency observed when inflow COD was below 300 mg/L (exceeding 90%). From 50 to 80 days, Al removal efficiency was generally low (less than 20%), but exceeded 80% in areas where inflow COD was below 30 mg/L. Similar trends were observed after 80 days, with overall Al removal efficiency below 50%, but exceeding 50% in areas where inflow COD was below 30 mg/L. Under conditions of low COD concentration, microbial activity might be reduced, leading to a greater reliance on physical and chemical processes for Al removal. Simultaneously, high removal efficiency of TFe and Mn was noted, with the hydroxides and oxides generated potentially serving as secondary adsorbents for co-precipitation with Al, thereby facilitating its removal (Singh and Chakraborty, 2020). Furthermore, robust growth and development of plants and root systems were observed, favoring Al absorption and transportation by plants, consequently enhancing Al removal (Nguegang et al., 2022).

The higher removal efficiencies of Mn and Al were both exhibited in a low concentration (below 50 mg/L) inflow COD region (Figure 7b, c). However, in the low inflow COD region, the Mn removal efficiency was optimal, whereas the Al removal efficiency did not reach its peak. This suggests that

increasing inflow COD is not effective for Mn removal but has a positive effect on Al removal.

Under conditions where inflow COD was less than 300 mg/L, the influence of operation Day was negligible, with Zn removal efficiency remaining close to 100% throughout the entire period, decreasing to approximately 80% after 150 days. Conversely, when inflow COD exceeded 300 mg/L, the effect of operation Day became more pronounced. In the initial 0-5 day, Zn removal efficiency was lower than 60% within 75 days, and subsequently increasing to 70%-80% beyond 75 days. Zn exhibits elevated mobility and potential bioavailability, predominantly occupying the exchangeable fraction. Moreover, Zn demonstrates similar coprecipitation and adsorption/complexation interactions with iron-manganese oxides/hydroxides as Al (Chen et al., 2021b).

3.4 Environmental implications

This study highlights the potential of ML to offer more accurate predictions and interpretable understanding of multi-metal removal efficiencies and their determinants in constructed wetlands. Within the leakage-aware and interpretable ML pipelines, imputation is coupled with learning, models are evaluated under out-of-fold and forward-chaining protocols, and explanatory tools are cross-validated, thereby turning heterogeneous monitoring records into actionable condition response insights. The approach quantifies relative importance and interactions among routinely measured parameters, with consistent patterns across methods, thereby informing near-real-time monitoring, anomaly screening, and operational tuning. Taken together, the framework underscores the value of embedding advanced data analytics in environmental research and provides a basic pipeline for future studies.

However, due to limitations imposed by both the quality and quantity of published literature, this study was subject to certain constraints. Owing to the diversity in research objectives, methodologies and experimental settings across different wetlands, the data types for certain input features and prediction targets were inconsistent. Such variations could introduce uncertainties into prediction outcomes and may not precisely reflect real-world scenarios. To address these challenges, future research should aim

to utilize a more extensive and standardized database to develop ML models. This database should encompass studies with clearly defined scientific objectives and employ similar methodologies under standardized experimental conditions. Particularly noteworthy are the absence of certain data types in this study, such as quantified parameters of wetland substrates (Singh et al., 2023), quantitative parameters of wetland microorganisms (Chen et al., 2020; Chen et al., 2021a) and plants (Shen et al., 2023), and water quality parameters like dissolved oxygen (DO) (Oldham et al., 2019) and hydraulic retention times (HRTs) (Demir et al., 2021). These types of data are crucial for a comprehensive analysis of the removal mechanisms of metals in wetlands. Numerous studies indicate that the physical adsorption of metals by wetland substrates is a highly significant removal pathway (Lizama-Allende et al., 2021; Nguyen et al., 2021). Incorporating more valuable features would enhance accuracy and provide deeper insights into the mechanisms and pathways of metal removal in constructed wetlands from a data analytics perspective.

4. Conclusion

In this study, we utilized ML to predict and analyze multi-metal removal efficiencies in constructed wetlands treating AMD and systematically audited risks of data leakage. The main findings are summarized as follows:

- Five ML models were developed in the main pipeline, with the XGBoost model emerging as the most effective, achieving high apparent predictive accuracy ($R^2 > 0.8$) for the removal efficiency of total iron, manganese, aluminum and zinc.

- Under leakage-safe out-of-fold evaluation, absolute scores decreased as expected (median $\Delta R^2 \approx -0.16$), yet tree-based models remained the most reliable class, with RF strongest overall. Forward-chaining, multi-horizon tests against naive baselines showed discernible short-term skill (one-step wins in 5/7 metals) but limited long-horizon skill (~10-sample windows win in 3/7), defining a realistic

boundary of generalizability.

- Detailed feature analysis using the optimal model combined with fold aggregation and dual explainers under the leakage-safe out-of-fold protocol identified operation Days (1-185) and inflow COD (6.523-1027.631 mg/L) as stable and dominant predictors of metal removal efficiency.

- Leave-one-wetland-out validation indicated moderate cross-wetland generalization ($R^2 > 0.5$ in most metal–wetland pairs) with a few degraded cases reflecting site heterogeneity. The $\pm 10\%$ input-perturbation test on dataset C features showed predictions were broadly stable (occasional sensitivity to i_COD , $i_acidity$, height), evidencing engineering robustness.

- Partial dependence plots elucidated the non-linear relationships between key predictors and metal removal efficiencies. This analysis revealed that specific ranges of operation Days and inflow COD levels are critical for optimizing the removal processes, providing actionable insights for the monitoring and management of constructed wetlands.

This work establishes a leakage-aware and interpretable machine-learning framework, coupling imputation with learning and systematically documenting sensitivity to methodological choices. Overall, the framework and findings offer a foundation for further research and practice, while warranting attention to temporal-scale bounds, cross-site extrapolation limits, and constraints of data and feature.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Raw data and all relevant code fully available on GitHub (<https://github.com/twelvevminusone/ML-AMD-CWs>) for open access.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2023YFC3710000) and the National Natural Science Foundation of China (42277193).

References

- Altmann, A., Tolosi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26, 1340-1347.
- Batty, L.C., Younger, P.L., 2002. Critical role of macrophytes in achieving low iron concentrations in mine water treatment wetlands. *Environ. Sci. Technol.* 36, 3997-4002.
- Berger, V.W., Zhou, Y., 2014. Kolmogorov–Smirnov Test: Overview, in: Kenett, R.S., Longford, N.T., Piegorsch, W.W., Ruggeri, F. (Eds.), *Wiley StatsRef: Statistics Reference Online*, 1 ed. Wiley.
- Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* 13, 281-305.
- Briscoe, J., Marin, O., 2020. Looking at neurodevelopment through a big data lens. *Science* 369.
- Chen, D., Wang, G., Chen, C., Feng, Z., Jiang, Y., Yu, H., Li, M., Chao, Y., Tang, Y., Wang, S., Qiu, R., 2023. The interplay between microalgae and toxic metal(loid)s: mechanisms and implications in AMD phycoremediation coupled with Fe/Mn mineralization. *J Hazard Mater* 454, 131498.
- Chen, H., Xiao, T., Ning, Z., Li, Q., Xiao, E., Liu, Y., Xiao, Q., Lan, X., Ma, L., Lu, F., 2020. In-situ remediation of acid mine drainage from abandoned coal mine by filed pilot-scale passive treatment system: Performance and response of microbial communities to low pH and elevated Fe. *Bioresource Technol* 317, 123985.
- Chen, J., Deng, S., Jia, W., Li, X., Chang, J., 2021a. Removal of multiple heavy metals from mining-impacted water by biochar-filled constructed wetlands: Adsorption and biotic removal routes. *Bioresource Technol* 331, 125061.

- Chen, J., Li, X., Jia, W., Shen, S., Deng, S., Ji, B., Chang, J., 2021b. Promotion of bioremediation performance in constructed wetland microcosms for acid mine drainage treatment by using organic substrates and supplementing domestic wastewater and plant litter broth. *J Hazard Mater* 404, 124125.
- Choudhary, R.P., Sheoran, A.S., 2012. Performance of single substrate in sulphate reducing bioreactor for the treatment of acid mine drainage. *Minerals Engineering* 39, 29-35.
- Daraz, U., Li, Y., Ahmad, I., Iqbal, R., Ditta, A., 2023. Remediation technologies for acid mine drainage: Recent trends and future perspectives. *Chemosphere* 311, 137089.
- Demir, E.K., Yaman, B.N., Celik, P.A., Puhakka, J.A., Sahinkaya, E., 2021. Simulated acid mine drainage treatment in iron oxidizing ceramic membrane bioreactor with subsequent co-precipitation of iron and arsenic. *Water Res* 201, 117297.
- Garreau, D., Luxburg, U., 2020. Explaining the Explainer: A First Theoretical Analysis of LIME, in: Silvia, C., Roberto, C. (Eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR, *Proceedings of Machine Learning Research*, pp. 1287-1296.
- Hong, M., Wang, J., Yang, B., Liu, Y., Sun, X., Li, L., Yu, S., Liu, S., Kang, Y., Wang, W., Qiu, G., 2024a. Inhibition of pyrite oxidation through forming biogenic K-jarosite coatings to prevent acid mine drainage production. *Water Res* 252, 121221.
- Hong, S.M., Morgan, B.J., Stocker, M.D., Smith, J.E., Kim, M.S., Cho, K.H., Pachepsky, Y.A., 2024b. Using machine learning models to estimate *Escherichia coli* concentration in an irrigation pond from water quality and drone-based RGB imagery data. *Water Res* 260, 121861.
- Hu, S., Liu, G., Zhang, J., Yan, J., Zhou, H., Yan, X., 2022. Linking electron ionization mass spectra of organic chemicals to toxicity endpoints through machine learning and experimentation. *J Hazard Mater* 431, 128558.

- Irshad, S., Xie, Z., Mehmood, S., Nawaz, A., Ditta, A., Mahmood, Q., 2021. Insights into conventional and recent technologies for arsenic bioremediation: A systematic review. *Environ Sci Pollut Res Int* 28, 18870-18892.
- Johnson, S.C., 1967. Hierarchical clustering schemes. *Psychometrika* 32, 241-254.
- Jouini, M., Benzaazoua, M., Neculita, C.M., Genty, T., 2020. Performances of stabilization/solidification process of acid mine drainage passive treatment residues: Assessment of the environmental and mechanical behaviors. *J Environ Manage* 269, 110764.
- Kim, T., Shin, J., Lee, D., Kim, Y., Na, E., Park, J.H., Lim, C., Cha, Y., 2022. Simultaneous feature engineering and interpretation: Forecasting harmful algal blooms using a deep learning approach. *Water Res* 215, 118289.
- Kini, K.R., Harrou, F., Madakyaru, M., Sun, Y., 2024. Enhanced data-driven monitoring of wastewater treatment plants using the Kolmogorov–Smirnov test. *Environ. Sci. Water Res. Technol.* 10, 1464-1480.
- Li, J., Pan, L., Suvarna, M., Tong, Y.W., Wang, X., 2020. Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. *Appl. Energ.* 269, 115166.
- Li, J., Zhang, L., Li, C., Tian, H., Ning, J., Zhang, J., Tong, Y.W., Wang, X., 2022. Data-Driven Based In-Depth Interpretation and Inverse Design of Anaerobic Digestion for CH₄-Rich Biogas Production. *ACS EST Eng.* 2, 642-652.
- Li, W., Niu, Q., Zhang, H., Tian, Z., Zhang, Y., Gao, Y., Li, Y.-Y., Nishimura, O., Yang, M., 2015. UASB treatment of chemical synthesis-based pharmaceutical wastewater containing rich organic sulfur compounds and sulfate and associated microbial characteristics. *Chem. Eng. J.* 260, 55-63.
- Liu, X., Lu, D., Zhang, A., Liu, Q., Jiang, G., 2022. Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. *Environ. Sci. Technol.* 56, 2124-2133.
- Lizama-Allende, K., Ayala, J., Jaque, I., Echeverria, P., 2021. The removal of arsenic and metals from

- highly acidic water in horizontal subsurface flow constructed wetlands with alternative supporting media. *J Hazard Mater* 408, 124832.
- Lones, M.A., 2024. Avoiding common machine learning pitfalls. *Patterns (N Y)* 5, 101046.
- Lopez, A.M., Wells, A., Fendorf, S., 2021. Soil and Aquifer Properties Combine as Predictors of Groundwater Uranium Concentrations within the Central Valley, California. *Environ. Sci. Technol.* 55, 352-361.
- Lu, X., Zhen, G., Ni, J., Hojo, T., Kubota, K., Li, Y.Y., 2016. Effect of influent COD/SO₄(²⁻) ratios on biodegradation behaviors of starch wastewater in an upflow anaerobic sludge blanket (UASB) reactor. *Bioresource Technol* 214, 175-183.
- Masindi, V., Akinwekomi, V., Maree, J.P., Muedi, K.L., 2017. Comparison of mine water neutralisation efficiencies of different alkaline generating agents. *Journal of Environmental Chemical Engineering* 5, 3903-3913.
- Masindi, V., Chatzisymeon, E., Kortidis, I., Foteinis, S., 2018. Assessing the sustainability of acid mine drainage (AMD) treatment in South Africa. *Sci Total Environ* 635, 793-802.
- Masindi, V., Fosso-Kankeu, E., Mamakoa, E., Nkambule, T.T.I., Mamba, B.B., Naushad, M., Pandey, S., 2022. Emerging remediation potentiality of struvite developed from municipal wastewater for the treatment of acid mine drainage. *Environ Res* 210, 112944.
- Mayes, W.M., Batty, L.C., Younger, P.L., Jarvis, A.P., Koiv, M., Vohla, C., Mander, U., 2009. Wetland treatment at extremes of pH: a review. *Sci Total Environ* 407, 3944-3957.
- Megahed, F.M., Chen, Y.J., Jones-Farmer, L.A., Rigdon, S.E., Krzywinski, M., Altman, N., 2024. Comparing classifier performance with baselines. *Nat Methods* 21, 546-548.
- Mitsch, W.J., Wise, K.M., 1998. Water quality, fate of metals, and predictive model validation of a constructed wetland treating acid mine drainage. *Water Research* 32, 1888-1900.
- Muhammad, N., Zvobgo, G., Zhang, G.-p., 2019. A review: The beneficial effects and possible

- mechanisms of aluminum on plant growth in acidic soil. *J. Integr. Agric.* 18, 1518-1528.
- Naidu, G., Ryu, S., Thiruvengkatachari, R., Choi, Y., Jeong, S., Vigneswaran, S., 2019. A critical review on remediation, reuse, and resource recovery from acid mine drainage. *Environmental Pollution* 247, 1110-1124.
- Nguegang, B., Masindi, V., Msagati Makudali, T.A., Tekere, M., 2022. Effective treatment of acid mine drainage using a combination of MgO-nanoparticles and a series of constructed wetlands planted with *Vetiveria zizanioides*: A hybrid and stepwise approach. *J Environ Manage* 310, 114751.
- Nguyen, X.C., Ly, Q.V., Peng, W., Nguyen, V.H., Nguyen, D.D., Tran, Q.B., Huyen Nguyen, T.T., Sonne, C., Lam, S.S., Ngo, H.H., Goethals, P., Le, Q.V., 2021. Vertical flow constructed wetlands using expanded clay and biochar for wastewater remediation: A comparative study and prediction of effluents using machine learning. *J Hazard Mater* 413, 125426.
- Nieva, N.E., Borgnino, L., Garcia, M.G., 2018. Long term metal release and acid generation in abandoned mine wastes containing metal-sulphides. *Environ Pollut* 242, 264-276.
- Oldham, C., Beer, J., Blodau, C., Fleckenstein, J., Jones, L., Neumann, C., Peiffer, S., 2019. Controls on iron(II) fluxes into waterways impacted by acid mine drainage: A Damkohler analysis of groundwater seepage and iron kinetics. *Water Res* 153, 11-20.
- Origin. OriginLab Corporation, Northampton, MA.
- Palansooriya, K.N., Li, J., Dissanayake, P.D., Suvarna, M., Li, L., Yuan, X., Sarkar, B., Tsang, D.C.W., Rinklebe, J., Wang, X., Ok, Y.S., 2022. Prediction of Soil Heavy Metal Immobilization by Biochar Using Machine Learning. *Environ. Sci. Technol.* 56, 4187-4198.
- Pat-Espadas, A.M., Loredó Portales, R., Amabilis-Sosa, L.E., Gómez, G., Vidal, G., 2018. Review of Constructed Wetlands for Acid Mine Drainage Treatment. *Water* 10, 1685.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Scikit-learn: Machine

Learning in Python. MACHINE LEARNING IN PYTHON.

- Ransom, K.M., Nolan, B.T., J, A.T., Faunt, C.C., Bell, A.M., Gronberg, J.A.M., Wheeler, D.C., C, Z.R., Jurgens, B., Schwarz, G.E., Belitz, K., S, M.E., Kourakos, G., Harter, T., 2017. A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. *Sci Total Environ* 601-602, 1160-1172.
- Reed, D., Wang, Y., Meselhe, E., White, E., 2020. Modeling wetland transitions and loss in coastal Louisiana under scenarios of future relative sea-level rise. *Geomorphology* 352, 106991.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, San Francisco, California, USA, pp. 1135-1144.
- Rodrigues, C., Nunez-Gomez, D., Silveira, D.D., Lapolli, F.R., Lobo-Recio, M.A., 2019. Chitin as a substrate for the biostimulation of sulfate-reducing bacteria in the treatment of mine-impacted water (MIW). *J Hazard Mater* 375, 330-338.
- Rodríguez-Galán, M., Baena-Moreno, F.M., Vázquez, S., Arroyo-Torralvo, F., Vilches, L.F., Zhang, Z., 2019. Remediation of acid mine drainage. *Environmental Chemistry Letters* 17, 1529-1538.
- Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., Pradhan, B., 2018. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci Total Environ* 644, 954-962.
- Schaider, L.A., Senn, D.B., Estes, E.R., Brabander, D.J., Shine, J.P., 2014. Sources and fates of heavy metals in a mining-impacted stream: temporal variability and the role of iron oxides. *Sci Total Environ* 490, 456-466.
- Schroer, H.W., Just, C.L., 2024. Feature Engineering and Supervised Machine Learning to Forecast Biogas Production during Municipal Anaerobic Co-Digestion. *ACS ES T Eng* 4, 660-672.
- Sharma, P., Tripathi, S., Sirohi, R., Kim, S.H., Ngo, H.H., Pandey, A., 2021. Uptake and mobilization of

- heavy metals through phytoremediation process from native plants species growing on complex pollutants: Antioxidant enzymes and photosynthetic pigments response. *Environ Technol Innov* 23, 101629.
- She, Z., Pan, X., Wang, J., Shao, R., Wang, G., Wang, S., Yue, Z., 2021. Vertical environmental gradient drives prokaryotic microbial community assembly and species coexistence in a stratified acid mine drainage lake. *Water Res* 206, 117739.
- She, Z., Wang, J., Pan, X., Ma, D., Gao, Y., Wang, S., Chuai, X., Yue, Z., 2023. Decadal evolution of an acidic pit lake: Insights into the biogeochemical impacts of microbial community succession. *Water Res* 243, 120415.
- Shen, C., Su, L., Zhao, Y., Liu, W., Liu, R., Zhang, F., Shi, Y., Wang, J., Tang, Q., Yang, Y., Bon Man, Y., Zhang, J., 2023. Plants boost pyrrhotite-driven nitrogen removal in constructed wetlands. *Bioresource Technol* 367, 128240.
- Sheoran, A.S., Sheoran, V., 2006. Heavy metal removal mechanism of acid mine drainage in wetlands: A critical review. *Minerals Engineering* 19, 105-116.
- Sheoran, A.S., Sheoran, V., Choudhary, R.P., 2010. Bioremediation of acid-rock drainage by sulphate-reducing prokaryotes: A review. *Minerals Engineering* 23, 1073-1100.
- Singh, N.K., Yadav, M., Singh, V., Padhiyar, H., Kumar, V., Bhatia, S.K., Show, P.L., 2023. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresource Technol* 369, 128486.
- Singh, S., Chakraborty, S., 2020. Performance of organic substrate amended constructed wetland treating acid mine drainage (AMD) of North-Eastern India. *J Hazard Mater* 397, 122719.
- Singh, S., Chakraborty, S., 2021. Bioremediation of acid mine drainage in constructed wetlands: Aspect of vegetation (*Typha latifolia*), loading rate and metal recovery. *Miner. Eng.* 171, 107083.
- Song, Y., Guo, Z., Wang, R., Yang, L., Cao, Y., Wang, H., 2022. A novel approach for treating acid mine

- drainage by forming schwertmannite driven by a combination of biooxidation and electroreduction before lime neutralization. *Water Res* 221, 118748.
- Stumm, W., Morgan, J.J., 2013. *Aquatic chemistry: chemical equilibria and rates in natural waters*. John Wiley & Sons.
- Sun, R., Zhang, L., Wang, X., Ou, C., Lin, N., Xu, S., Qiu, Y.Y., Jiang, F., 2020. Elemental sulfur-driven sulfidogenic process under highly acidic conditions for sulfate-rich acid mine drainage treatment: Performance and microbial community analysis. *Water Res* 185, 116230.
- Sun, Z., Feng, L., Li, Y., Han, Y., Zhou, H., Pan, J., 2022. The role of electrochemical properties of biochar to promote methane production in anaerobic digestion. *Journal of Cleaner Production* 362, 132296.
- Tabelin, C.B., Corpuz, R.D., Igarashi, T., Villacorte-Tabelin, M., Alorro, R.D., Yoo, K., Raval, S., Ito, M., Hiroyoshi, N., 2020. Acid mine drainage formation and arsenic mobility under strongly acidic conditions: Importance of soluble phases, iron oxyhydroxides/oxides and nature of oxidation layer on pyrite. *J Hazard Mater* 399, 122844.
- Vasquez, Y., Neculita, C.M., Caicedo, G., Cubillos, J., Franco, J., Vasquez, M., Hernandez, A., Roldan, F., 2022. Passive multi-unit field-pilot for acid mine drainage remediation: Performance and environmental assessment of post-treatment solid waste. *Chemosphere* 291, 133051.
- Villegas-Plazas, M., Sanabria, J., Arbeli, Z., Vasquez, Y., Roldan, F., Junca, H., 2022. Metagenomic Analysis of Biochemical Passive Reactors During Acid Mine Drainage Bioremediation Reveals Key Co-selected Metabolic Functions. *Microb Ecol* 84, 465-472.
- Wang, H., Zhang, M., Xue, J., Lv, Q., Yang, J., Han, X., 2021. Performance and microbial response in a multi-stage constructed wetland microcosm co-treating acid mine drainage and domestic wastewater. *J. Environ. Chem. Eng.* 9, 106786.
- Wang, X., Jiang, H., Fang, D., Liang, J., Zhou, L., 2019. A novel approach to rapidly purify acid mine

- drainage through chemically forming schwertmannite followed by lime neutralization. *Water Res* 151, 515-522.
- Wang, Y.Q., Zhou, H.B., Luo, X.Q., Deng, S.W., Xu, H.R., Song, Y.P., Chen, J.J., Yin, W.X., Cheng, H.Y., Wang, A.J., Wang, H.C., 2025. Machine Learning-Driven Dynamic Measurement of Environmental Indicators in Multiple Scenes and Multiple Disturbances. *Environ. Sci. Technol.* 59, 15877-15889.
- Wu, G., Kechavarzi, C., Li, X., Wu, S., Pollard, S.J.T., Sui, H., Coulon, F., 2013a. Machine learning models for predicting PAHs bioavailability in compost amended soils. *Chemical Engineering Journal* 223, 747-754.
- Wu, S., Kuschik, P., Wiessner, A., Müller, J., Saad, R.A.B., Dong, R., 2013b. Sulphur transformations in constructed wetlands for wastewater treatment: A review. *Ecological Engineering* 52, 278-289.
- Yan, J., Yan, X., Hu, S., Zhu, H., Yan, B., 2021. Comprehensive Interrogation on Acetylcholinesterase Inhibition by Ionic Liquids Using Machine Learning and Molecular Modeling. *Environ. Sci. Technol.* 55, 14720-14731.
- Yin, M., Zhang, X., Li, F., Yan, X., Zhou, X., Ran, Q., Jiang, K., Borch, T., Fang, L., 2024. Multitask Deep Learning Enabling a Synergy for Cadmium and Methane Mitigation with Biochar Amendments in Paddy Soils. *Environ. Sci. Technol.* 58, 1771-1782.
- Younger, P.L., Banwart, S.A., Hedin, R.S., Younger, P.L., Banwart, S.A., Hedin, R.S., 2002. Mine water hydrology. Springer.
- Zhang, H., Yin, S., Chen, Y., Shao, S., Wu, J., Fan, M., Chen, F., Gao, C., 2020. Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. *Journal of Cleaner Production* 273, 122858.
- Zhang, S., Zheng, H., Tratnyek, P.G., 2023a. Advanced redox processes for sustainable water treatment. *Nat Water* 1, 666-681.
- Zhang, Y., Feng, Y., Ren, Z., Zuo, R., Zhang, T., Li, Y., Wang, Y., Liu, Z., Sun, Z., Han, Y., Feng, L.,

- Aghbashlo, M., Tabatabaei, M., Pan, J., 2023b. Tree-based machine learning model for visualizing complex relationships between biochar properties and anaerobic digestion. *Bioresource Technol* 374, 128746.
- Zhao, W., Ma, J., Liu, Q., Dou, L., Qu, Y., Shi, H., Sun, Y., Chen, H., Tian, Y., Wu, F., 2023. Accurate Prediction of Soil Heavy Metal Pollution Using an Improved Machine Learning Method: A Case Study in the Pearl River Delta, China. *Environ. Sci. Technol.* 57, 17751-17761.
- Zhu, J.J., Yang, M., Ren, Z.J., 2023. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environ. Sci. Technol.* 57, 17671-17689.
- Zhu, X., Li, Y., Wang, X., 2019a. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresource Technol* 288, 121527.
- Zhu, X., Wang, X., Ok, Y.S., 2019b. The application of machine learning methods for prediction of metal sorption onto biochars. *J Hazard Mater* 378, 120727.
- Zou, J., Qiu, Y.Y., Li, H., Jiang, F., 2023. Sulfur disproportionation realizes an organic-free sulfidogenic process for sustainable treatment of acid mine drainage. *Water Res* 232, 119647.